Copenhagen Business School
Department of Digitalisation
cand.merc.(it) (Data Science)
CDSCO4001E

Clustering COVID-19 Travellers' Perceptions and Behaviours for Tourism Recovery: Using K-Modes

Master Thesis in

Business Administration and Data Science

May 15th, 2023

By:
Prasoon Saxena, 142920

Supervisor:
Fumiko Kano Glückstad

Contract No: 23106

Number of Pages: 62

Number of Characters: 112332

Acknowledgement

I would like to express my sincere gratitude to my supervisor Fumiko Kano Glückstad for her guidance, patience, and invaluable feedback throughout this research journey. I am also thankful to her for providing me the dataset used in the project.

I would like to extend my thanks to my family and friends for their support, encouragement, and understanding during this time.

Thank you all for your contributions and support.

Abstract

The COVID-19 pandemic has had a significant impact on the travel industry, leading to a reassessment of travel behaviour and risk perception by individuals and businesses alike. Against this backdrop, this thesis examines travel risk behaviour and perception in a COVID-like situation by analysing survey responses from individuals in four countries: Japan, Italy, China, and Denmark. The study aims to provide a comprehensive understanding of how individuals perceive and respond to travel risks during such times, gathering quantitative data on individuals' travel experience, important factors for their next travel destination, travel behaviour and risk perception, and intentions to travel within one's respective country during the transitional phase. The study also investigates the impact of demographic factors such as age, gender, and geography on travel risk perception and behaviour. The research can provide valuable insights for policymakers and travel industry stakeholders on how to mitigate travel risks during pandemics and other similar situations, particularly in the context of COVID-19 tourism recovery. The study identified three clusters of respondents with different travel preferences and intentions termed as "Risk-Takers," "Risk-Averse," and "Risk-Neutral" groups, indicating that there is still high demand for domestic travel despite the pandemic. While the research was conducted during a transitional phase when societies were reopening with restrictions in place, it highlights the importance of measures that can influence travel demand.

Keywords: Cluster Analysis, K-mode Algorithm, Latent Class Analysis, Travel Risk Perceptions, COVID-19, Tourism Recovery

Table of Contents

1. Introduction

1.1 Background

The outbreak of COVID-19 began in Wuhan, China, in December 2019, and spread rapidly to become a global pandemic by the end of that month. The impact of the COVID-19 pandemic has been felt in almost every aspect of life, including the tourism industry. As countries have imposed travel restrictions and lockdowns (at local and national levels) because of Covid-19 the tourism sector got severely affected. Globally tourist arrival in the year 2020 has gone down considerably by around 73% as compared to pre-covid year 2019 which has caused significant economic losses to the countries heavily dependent on the tourism sector (UNWTO, 2021). This significant drop in tourism has caused job losses for the people involved in the tourism industry. One of the biggest change this Covid-19 have bought was the ban on not only international travel where the borders got closed but also the local and national level lockdown which made it difficult for people to travel internationally and nationally. Such restrictions have caused a decrease in demand for flights, ship cruises, train travel, accommodations, and tourist activities which had a noteworthy negative impact on the tourism industry. The international spread of COVID-19 was caused by a complex interplay of factors, including global travel and trade, asymptomatic spread, delayed response, and lack of international coordination. Among all the factors global travel is one of the main factors influencing the spread (Sudhvir et al.,2021). Tourism is negatively impacted by the risk associated with infectious diseases, one of the earlier studies shows that eradication of diseases like Malaria, Dengue, Yellow Fever, and Ebola from the affected countries could increase around 10 million tourists which eventually has increased tourism expenditure by 12 billion US dollars (Jaume et al., 2017).

Earlier studies conducted have mentioned that tourists' risk perception is complicated and depends on individual characteristics. (Roehl et al., 1992) the study categorised tourists into three groups based on their perceptions: Risk Neutral Group, Functional Risk group, and Place Risk group. The risk-neutral group was the group of younger and less educated tourists' who on one side recognizes the risks associated with travel but on the other side, they didn't let those risks deter their travel plans to a good extent. The functional-risk group is also the group of young tourists' but is more educated than the risk-neutral group who are motivated by destinations perceived to be risky like adventure travel or international travel. The place-risk tourist group consists of highly educated older tourists who either do not travel at all as they associate high levels of risks with travel or travel to destinations perceived to be safer.

The perception of risks is also influenced by nationality or cultural background. As per (Seddighi et al., 2001) cultural background plays a significant role in destination choice, with tourists from different cultural backgrounds exhibiting different preferences for destinations and these differences can be attributed to variations in cultural values, norms, and expectations.

Gender also plays an important role in behaviour toward travel risks. Mónica Ferrín in her study (Ferrin, 2022) identified that women are risk-averse compared to men even when the risks associated with COVID-19 are low. An earlier study also showcased that COVID-19 is considered a very serious health problem by more women than men (Pons et al.,2020).

A variety of risk factors involved in the tourism industry can influence tourists' decisions for any travel destination. These risk factors can be related to health, safety, security, transportation, legal, and financial. So, whenever an infectious outbreak occurs then tourists' travel perceptions can affect their travel decision which can range from cancellation or postponement of trips, avoidance of specific travel destinations, and behaviours like precautionary measures for safety and sanitation, confirmation about the disease control from authorities and social media. Pandemics like Ebola, Spanish Flu, Swine Flu, and SARS have shown us how risk perception influences travel behaviour.

Previous studies conducted in identifying the impact on tourism because of the outbreak of infectious diseases or pandemics like Spanish Flu, Swine Flu, Ebola, and SARS have stressed the fact that the tourism industry gets negatively impacted as tourist travel behaviours get influenced by their risk perceptions. However, the Covid-19 pandemic was beyond anyone's imagination and hence none of the states have any measures to control it and in turn would have stopped or controlled the negative impact on tourism. Thus, it becomes necessary to conduct new research to understand the situation.

This academic thesis seeks to contribute to this understanding by investigating tourists' perceptions and behaviour during the pandemic. The study will examine how tourists' attitudes and behaviours have changed since the outbreak of Covid-19 and how these changes affect their travel intentions. Moreover, the study will implement clustering and statistical methods to identify distinct groups of tourists with similar characteristics, preferences, behaviours, and intentions to travel during the transitional phase. The results of this study look forward to providing valuable insights into the post-Covid-19 tourism industry to help policymakers, tourism businesses, and other stakeholders to create effective strategies for the recovery and growth of tourism in situations like the pandemic.

1.2 Problem Statement

The Covid-19 pandemic has had a significant impact on the tourism industry worldwide, including domestic tourism. Many countries have imposed travel restrictions and safety measures to limit the spread of the virus, leading to a decline in domestic tourism. It is essential to understand the travel intentions of domestic tourists in the transitional phase to assist the recovery of domestic tourism as the situation evolves and restrictions are lifted.

The problem statement of the thesis is to use clustering analysis in exploring the factors that influence the travel intentions of domestic tourists in the transitional phase and how these factors affect their decision-making processes. This includes understanding their past travel experiences, preferred travel arrangements, preferred travel companions, important factors related to travel destinations, travel risks, behaviours, and perceptions. The thesis aims to provide insights into how the domestic tourism industry can adapt to the new normal and develop strategies to recover from the impact of the pandemic on domestic tourism.

1.3 Research Questions

The K-modes algorithm is a well-known unsupervised clustering algorithm utilized in data science for categorizing categorical variables. It is an expansion of the popular K-means algorithm that is utilized for numerical data. In this thesis, data from four countries will be utilized to establish a framework for future analyses, which may incorporate data from countries with distinct sociocultural backgrounds. Because a computationally efficient algorithm is required, this investigation will compare the K-modes algorithm to the Latent Component Analysis (LCA) statistical method, which is utilized to identify underlying factors implicated in the development of groups. LCA is computationally intense and may create issues for further research on large datasets, making it essential to discover an equally effective approach. Once the preferred clustering or grouping method is identified, this study aims to determine the factors that influence the formation of these groups. The findings of this analysis will provide insights into developing strategies to recover domestic tourism from the impact of the pandemic. Therefore, the two research questions are:

- Q1: What is the level of agreement between clusters formed by K-modes and groups identified by LCA in the context of clustering categorical data for tourism research?
- Q2: What are the factors that influence the formation of the identified clusters or groups and how can these insights be used to develop effective strategies to recover domestic tourism from the impact of the Covid-19 pandemic?

1.4 Thesis Structure

The purpose of this section is to provide the reader with an overview of the thesis structure and help them navigate through it. Once the *Introduction* chapter has been completed, the *Literature Review* chapter will explore the relevant research that pertains to understanding the various aspects of tourist behaviour, perception, attitudes, and intentions with regard to travelling in a pandemic scenario similar to the covid-19 outbreak. Moving forward, the *Theoretical Framework* section will provide an in-depth analysis of the clustering and statistical technique that are applicable to the study. This includes an extensive discussion on the k-modes algorithm and LCA, along with their respective benefits and limitations. In addition to this, the section will also address the measures that will be used in comparing clusters formed by LCA groups and k-modes clusters, the measure used in identifying the optimal number of clusters or groups of k-modes and LCA respectively, measure to test independence between categorical variables (Chi-square test), the measure of find the association between two categorical variables (Cramer's V test), and the post-hoc testing methods. Moreover, the *Methodology* section, which will be guided by the CRISP-DM framework, will outline the steps taken in understanding the business objectives, exploring data, and modelling and analysing the patterns that are identified in distinct clusters/groups. The *Results* section will then evaluate the models based on the relevant metrics mentioned in the *Theoretical Framework* section. The Discussion section will present contextualised results and answer the research questions, while also elaborating on the limitations, implications, future work, and learning reflections. Finally, the *Conclusion* chapter will summarize the thesis.

2 Literature Review

This chapter offers an overview of the previous studies to establish the groundwork on which the thesis is based.

2.1 LCA and K-Modes in Tourism

Lately, the clustering techniques have become increasingly popular in various social science research as they provide valuable insights of homogeneous groups or entities based on their shared characteristics. Two such methods that have gained good popularity are K-modes and Latent class analysis (LCA).

K-modes is a clustering algorithm that is specifically designed for categorical data. It works by assigning each observation to the mode of the cluster it is most similar to. In contrast, LCA is a statistical method that is used to identify unobserved or latent classes within a population. Both techniques have been widely used in various fields such as marketing, psychology, and tourism research.

Several studies have been conducted in the past using either LCA or K-modes separately. As an example (Parikh et al.,2018) proposed a mobile application using K-modes for tourist place recommendation and recognition system by taking the user's interest and recommending attractions, restaurants, and hotels. The study done by (Kucukefe & Kaya, 2023) was focussed on understanding Turkey's labour market during the Covid-19 pandemic, which had a profound impact on labour force involved in the tourism sector as well. Another study by (Mantouka et al., 2022) performed the study in understanding Athens's user perception and feelings for autonomous mobility on demand in the COVID-19 pandemic era using K-modes for categorical variables. Similarly, using LCA (Wang et al., 2022) studied the factors influencing holiday destination attractiveness during Covid-19 and found two main destination attributes in terms of accommodation type and crowdedness concluding the four groups of people as "social", "relax", "intellectual" and "mastery". A separate study (Aresi et al., 2022) studied the prosocial behaviours of Italian people under collective quarantine conditions during the 2020 COVID-19 lockdown. However, fewer studies have investigated the congruence between K-modes and LCA, and their findings have been positive as the results obtained by the two methods have shown substantial agreement between them. For instance, a study by (Papachristou et al.,2018) compared the two methods in the identification of oncology patients with distinct symptom experiences and found that the two techniques produced similar results with substantial agreement. In another study by (Daniel & Mateo, 2021) found that the two methods produced similar results when applied to analyse and cluster records from the Global Terrorism Database (GTD) referring to terrorist attacks belonging to the Islamic State. Hence, there is a need for further investigation into the congruence between these methods using the tourism related study so that either the method or the two methods in conjunction with each other can be used in future research. However, I will be using either of the approaches in further analysis of the clusters deeply if the approach finds considerable agreement.

2.2 Tourist's Perceptions, Attitudes, and Behaviour

The pandemic has caused a big impact on the tourism industry which has caused a decline in tourism for many places and countries. This has led to the need of understanding the various factors (including the tourist's behaviour, perceptions, and intentions to travel post-pandemic phase-1) which when addressed can help in the revival of the tourism industry.

Several studies have examined the impact of COVID-19 on tourist behaviour and attitudes. For instance, a study by (Perić et al., 2021) suggests that Serbian tourists' travel intentions during the COVID-19 pandemic are negatively affected by their risk perception, which includes health, psychological, financial, and destination risks. Additionally, travel risk has a negative impact on travel abroad, while health risk is on the verge of significance as a predictor of travel abroad during the pandemic. The respondents' monthly income was also identified as a significant predictor of travel abroad during the pandemic.

Covid-19 had a big impact on the mode of transport, in another study conducted by (Dias et al., 2020) they found that during the pandemic, there was a significant change in transportation mode choice from public transport to private or non-motorized modes. People were more concerned about pandemic-related issues when choosing a transportation mode. Factors such as gender, car ownership, employment status, travel distance, primary purpose of travel, and pandemic-related underlying factors were found to be significant predictors of mode choice during the pandemic.

Given the uncertainties and challenges associated with the COVID-19 pandemic, many experts predict that the revival of international tourism will be slow and gradual. Therefore, in the short term, the focus will likely be on the revival of domestic tourism. The study conducted by (Yuni, 2020) on Indonesians found that the majority of post-pandemic tourists were aged between 26-

45 years old, female, undergraduates, from Bali, and private employees. They made regular trips 1 to 5 times a year and preferred to travel immediately after the pandemic for 1-3 months. Finances and travel costs were important concerns, and they preferred to travel with a partner and by airplane. They also preferred cheap homestays and nature tourism and arranged their own trips.

A study by (Tiago, et al., 2021) investigated the impact of the pandemic on tourists' intention to travel and found that fear of infection, and travel restrictions have the greatest impact on tourist behaviour, followed by their use of social media and technology. Therefore, tourism and hospitality industry players should emphasize the importance of health and hygiene measures in their communication strategy and ensure the accuracy and reliability of information shared with tourists. Another study by (Moya et al.,2022) found that most respondents intended to travel for two nights with family members within the first six months after the lockdown. Safety and security were the primary factors influencing travel decisions, and protected areas and nature-based tourism were the preferred destinations. The study also identified the most visited protected areas in Costa Rica. Similarly, a study by (Orîndaru et al., 2021) found that the pandemic had led to a shift in tourists' preferences toward more sustainable and local tourism experiences.

As most of the studies conducted looked at any specific country, or specific sub-area within the country so there is a need of performing the study holistically by combining the data of different countries together. Moreover, there is still a need for research that examines how tourists cluster based on their behaviour and attitudes toward travel during the pandemic. Such research could provide valuable insights into the heterogeneity of tourists and help tourism businesses tailor their offerings to specific segments. Moreover, it could provide insights into the factors that influence tourist behaviour in the context of the pandemic, which could help destinations and tourism businesses develop effective strategies for recovery.

Overall, the literature suggests that there is a need for further research into the congruence between K-modes and LCA and how tourists cluster differ based on their behaviour and attitudes towards travel during the COVID-19 pandemic and finally understanding the tourist's intentions to travel during the transitional phase.

3. Theoretical Framework

3.1 Clustering

Clustering generally is a way to group the items or things who similar to each other. In data science, Clustering as a part unsupervised machine learning technique is used to divide unlabelled data points into different clusters based on some similarity metric. In other words, clustering divides an unlabelled dataset into clusters or groups, where data points in each group are more similar to one another than to data points in other groups (Michael et al., 2005). It is applied to perform various things like performing exploratory data analysis, pattern recognition, customer segmentation, anomaly detection, and various other applications.

The tourism sector can greatly benefit from clustering techniques. Clustering is a powerful tool that enables businesses and destinations to gain a deeper understanding of tourists' behaviour. By clustering tourists according to their characteristics, preferences, and behaviours, businesses and destinations can obtain valuable insights into the factors that drive their behaviour and decision-making related to travel.

One common use of clustering in tourism is to understand the motivations of different types of tourists. By clustering tourists based on their motivations for travel (Ramires et al.,2018), businesses, and destinations can gain insights into what drives their behavior and what types of activities and services are most likely to appeal to them. For example, some tourists may be motivated by adventure, while others may be motivated by relaxation or cultural experiences. By understanding these motivations, businesses and destinations can develop customized packages and services that cater to the needs and preferences of each segment.

Another use of clustering in tourism is to understand the decision-making process of tourists. By clustering tourists based on their decision-making processes, businesses, and destinations can gain insights into how tourists gather information, evaluate options, and make decisions related to travel. For example, some tourists may rely on recommendations from friends and family, while others may rely on online reviews or travel guides (Banasree & Mrinmoy, 2010). By understanding these decision-making processes, businesses and destinations can develop more effective marketing strategies and communication channels that resonate with different segments of tourists.

Clustering can also be used to understand the behaviour of tourists during different legs of the travel experience (Prebensen et al., 2012). For instance, we can capture the behaviours of tourists at various legs (before, during, and after) of their trip and use that to cluster them based on their behaviour in various legs then it will be helpful for businesses and destinations to gain insights into what types of information and services are most important to tourists at each leg. The information obtained can be used by the relevant stakeholders to develop and design effective marketing strategies and come up with relevant service offerings that serve the needs of tourists at every leg of their journey.

There are several types of clustering algorithms, and the choice of algorithm depends on the nature of the data and the problem being addressed. These algorithms include K-means clustering, Hierarchical clustering, K-modes clustering, and Density-based clustering. The most commonly used algorithm is K-means clustering, which partitions observations into k clusters based on the mean value of the observations. However, K-means clustering is sensitive to the initial centroid values and may not converge to the optimal solution. K-modes clustering is specifically designed for clustering categorical data based on the most common categories or modes. This algorithm calculates a distance metric between observations based on the number of categorical variables that differ between them and assigns each observation to the closest mode. Another algorithm that can be used for any type of data is Hierarchical clustering, which creates a hierarchy of clusters either by merging smaller ones into larger ones or by dividing larger ones into smaller ones (Huang, 1998).

3.2 K-Modes Algorithm

K-modes clustering is a clustering algorithm that is used in data science to group categorical data into clusters. The algorithm partitions the data into a predefined number of K clusters, where each cluster is represented by a centroid which is the mode of the categorical data. K-modes algorithm tries to minimize the cost function, which is the sum of the dissimilarity between each data point and the centroid of its assigned cluster. To measure the dissimilarity between two categorical values, simple matching techniques are used (Kaufman and Rousseeuw, 1990).

3.2.1 Dissimilarity Measure

Consider two categorical objects X and Y with m categorical attributes. The dissimilarity measure between these two objects can be defined as the total number of mismatches in the corresponding attribute categories. If the number of mismatches is smaller, it implies that the two objects are more similar (Huang, 1998).

$$d_1(X, Y) = \sum_{j=1}^{m} \delta(x_j, y_j)$$

where

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases}$$

Source: Huang, Z. (1998)

3.2.2 Cost Function and Algorithm

The K-modes algorithm can be formulated as follows:

1. Begin by selecting k initial modes, one for each cluster.
2. Allocate each object to the cluster with the nearest mode and update the mode of each cluster after each allocation.
3. Once all objects have been allocated to clusters, test the dissimilarity of each object against the current modes. If any object is found to have a nearest mode that belongs to a different cluster than its current one, reallocate the object to that cluster and update the modes of both clusters.
4. Repeat step 3 until no object changes clusters after a full cycle test of the entire dataset.

The cost function for K-modes can be written as:

$$P(W, \mathbf{Q}) = \sum_{l=1}^{k} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{i,l} \delta(x_{i,j}, q_{l,j})$$

where $w_{i,l} \in W$ and $Q_l = [q_{l,1}, q_{l,2}, \ldots, q_{l,m}] \in \mathbf{Q}$.

Source: Huang, Z. (1998)

By minimizing the cost function, the K-modes algorithm effectively partitions the categorical data into K clusters such that the dissimilarity between the data points within each cluster and the centroid of that cluster is minimized. The resulting clusters can then be used for further analysis, such as identifying patterns or making predictions (Huang, 1998).

3.2.3 Benefits of K-modes

Here are few benefits of using the K-modes algorithm for clustering:

- Handling categorical data: This is a special algorithm designed to handle categorical data.
- Faster than other algorithms: K-modes algorithm is relatively fast and computationally efficient compared Latent Class Analysis (LCA) (Chaturvedi, et al., 2001).
- Can handle the large datasets typically found in survey research applications (Chaturvedi, et al., 2001).
- Simple to implement: The K-modes algorithm is relatively easy to understand and implement, as it only requires the calculation of mode and dissimilarity measures.
- Interpretable clusters: The clusters generated by K-modes algorithm are often interpretable because they are based on categorical variables, making it easier to understand and interpret the results.
- Robust to noise: K-modes algorithm is robust to noise and can handle missing values, making it a suitable algorithm for real-world datasets that may contain noise or missing values.

3.2.4 Drawbacks of K-modes

- Categorical data only: K-modes is designed to work with categorical data, which implies that they may not be as effective in handling continuous or mixed data.
- Initial centroids sensitivity: The K-modes algorithm is sensitive to the initial centroids selected for the clusters. If the initial centroids are not well-chosen, the algorithm may converge to a suboptimal solution. Hence, to increase the likelihood of finding the optimal solution, it's preferable to select multiple random starting seeds.
- K-modes algorithm only guarantees locally optimal solutions: Like K-means, K-modes clustering only guarantees locally optimal solutions. Hence, it's advised to use multiple starting seeds if wanted to improve the chances of finding the global optimal solution.
- Hard to get an optimal number of clusters: Similarly, to K-means clustering, finding the optimal number of clusters in K-modes clustering is difficult. The Silhouette Coefficient method to estimate the optimal number of clusters is used in various studies conducted earlier.
- Unbalanced cluster sizes: K-modes algorithm tends to produce clusters with unbalanced sizes if the data is highly skewed which can lead to the misrepresentation (by giving the under- representation or over-representation) of those categories in the resulting clusters.

3.3 Latent Class Analysis (LCA)

In addition to machine learning clustering techniques, another popular statistical technique for categorical variables used for identifying clusters or groups is called Latent Class Analysis (LCA). According to (McCutcheon and Hagenaars, 1997), Latent Class Analysis (LCA) is a versatile and effective method for analyzing categorical data. It is a form of model-based clustering that usually employs the expectation-maximization (EM) algorithm to estimate the model parameters. LCA aims to identify distinct groups or classes of individuals based on patterns of responses to categorical variables. The method assumes that individuals within each class share similar response patterns and that differences between classes are due to systematic variations in response probabilities. LCA is frequently employed across various fields, such as psychology, sociology, marketing, and public health, to detect hidden subgroups or "classes" within a larger population based on their responses to categorical variables. By analysing the patterns of behaviours, attitudes, or preferences, LCA can reveal underlying trends that may not be easily discerned using alternative analytical methods.

3.3.1 Mathematical Principles of LCA

Latent Class Analysis is based on the idea that the observed multivariate distribution results from a mixture of distributions created by hidden or latent classes. By using a set of observed indicators, LCA models aim to identify these classes as accurately as possible. The models use a maximum likelihood function to estimate the parameters of the model, which represents the most fundamental components of the model and determines its behaviour. It is critical to maximize the data to find solutions or estimates for the parameters that best reflect the actual data from which the model is created. Expectation-maximization (EM) algorithms are employed to fit LCA models since class membership is not directly observed. These algorithms offer a framework for computing likelihood estimates when data is missing. The mixture parameters are optimized iteratively until a global solution for the maximum likelihood estimate is identified. The maximum likelihood estimates of the parameters aid in establishing the clusters and their separation in the model. As a result, based on an individual's observed indicators, a probability of belonging to all classes in a model can be produced. For g classes with proportions given by the model can be expressed as

$$f(y) = \sum_{i=1}^{g} \pi_i \, fi(y|x' B_i)$$

The equation represents the probability of the i^{th} class (π_i) and the conditional probability density function of the response in the i^{th} class model ($f_i(\cdot)$).

One can estimate the probabilities for the latent classes by employing a multinomial model, which can be expressed as:

$$\pi_i = \frac{\exp(\gamma_i)}{\sum_{j=1}^{g} \exp(\gamma_i)}$$ Where γ_i is the linear prediction for the i^{th} latent class.

In order to obtain the maximum likelihood solution, the EM algorithm requires iterative back-and-forth calculations until the likelihood reaches its maximum value. However, in LCA, the EM algorithms are "greedy" and can be sensitive to the initial values used. This may result in a "local maximum" rather than the true maximum likelihood value. To prevent this, it is recommended to use multiple random starting values. If each starting value leads to the same maximum likelihood value, then it is likely that the true maximum has been identified. To ensure the consistency of model performance, it is advisable to replicate the models with the best fit (maximum log-likelihood) by increasing the number of starting values. This will help ensure that the maximum value has been correctly identified. If the model fails to converge repeatedly to the maximum log-likelihood value, it may suggest that the data does not support the number of latent classes included in the model (Sinha et. al, 2021).

3.3.2 Benefits of LCA

- LCA identifies hidden structures: Latent Class Analysis (LCA) can identify underlying or hidden structures in a population, revealing subgroups or segments that exist within it.
- LCA is flexible: LCA can handle various types of data, including categorical and continuous data, and can work with a mixture of data types, providing flexibility in data analysis.
- LCA supports improved decision-making: By identifying patterns and segments within a population, LCA can assist in developing targeted interventions or marketing strategies, resulting in improved decision-making.
- LCA reduces data complexity: LCA groups individuals with similar characteristics together, thereby reducing the complexity of data. This simplifies data interpretation and enhances its comprehensibility.

- LCA allows model evaluation: Statistical tests and model fit indices can be used to evaluate LCA models, enabling researchers to assess the model quality and make improvements if necessary.

3.3.3 Drawbacks of LCA

- Interpretation challenges: One drawback of LCA is that interpreting the latent classes can be difficult, as they are not directly observable. Researchers may struggle to understand what the classes represent and how they relate to the observed variables.
- Model selection difficulty: Selecting the appropriate number of latent classes can be challenging, as it relies on various information criteria such as BIC and AIC. The model with the lowest BIC is typically chosen, but this may result in too many classes, requiring researchers to use their own judgment and expertise.
- Assumption of conditional independence: LCA assumes that the observed variables are conditionally independent given the latent classes. However, this assumption may not always hold, leading to biased estimates.
- Computationally intensive: LCA can be computationally intensive, especially for large sample sizes or many observed variables. Running LCA models may require specialized software and significant computing resources.

3.4 Various Measures

3.4.1 Cohen’s Kappa

For categorical variables, Cohen's kappa coefficient (Cohen, 1960) is a statistical measure used to assess the reliability between multiple raters or a single rater at different times. Compared to a simple percent agreement calculation, it is considered a more reliable measure, as it considers the possibility of chance agreement. In the thesis, we will use this measure to find the agreement between the clusters created by the K-modes algorithm and groups identified by LCA (Cohen, 1960).

Formula is $k = (po – pe) / (1 – pe)$; where

po: Relative observed agreement among raters

pe: Hypothetical probability of chance agreement

3.4.2 Silhouette Coefficient (SC)

The Silhouette technique is utilized to assess and validate the consistency of clusters in a dataset. It determines the similarity of an object to its cluster (cohesion) versus other clusters (separation). The outcome ranges from -1 to +1, with higher values indicating that an object fits well within its cluster but not with nearby clusters. If the majority of objects have high Silhouette values, the clustering configuration is considered appropriate. However, if many points have low or negative values, the clustering configuration may have either too few or too many clusters. The Silhouette calculation can use any distance metric, including the Euclidean distance or the Manhattan distance. For the thesis, the Manhattan distance will be utilized in validating the optimal number of clusters.

Mathematical Representation

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ if } |C_I| > 1$$

and

$$s(i) = 0, \text{ if } |C_I| = 1$$

Which can be also written as:

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{if } a(i) > b(i) \end{cases}$$

From the above definition it is clear that

$$-1 \leq s(i) \leq 1$$

Source: (Rousseeuw, 1987)

Where s(i), is the silhouette (value) of one data point I, a(i) is the average dissimilarity of point i to all other objects within the cluster, and b(i)the minimum average distance between point i and all other clusters, where i is not a member of those clusters (Rousseeuw, 1987).

3.4.3 Akaike information criterion (AIC) and Bayesian information criterion (BIC)

There are various statistical measures to select the model among the set of available models. The two such statistical measures are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) that are used in the selection of model and this measure I will be using in finding the best model from two to ten LCA models. AIC is based on the likelihood function and the number of parameters in the model (Bozdogan, 1987) whereas BIC is a similar model based on the Bayesian principles on complexity measurements placing greater penalty on the models with more parameters.

The formula to obtain AIC value for a model is:

$$AIC = -2\log(L) + 2k$$

where k represents the number of parameters in the model and L represents the likelihood of the model. The model which will have the lowest AIC value among the available candidate's model will be chosen as the best one.

The formula to obtain BIC value for a model is:

$$BIC = -2\log(L) + k\log(n)$$

Here 'n' is the sample size. BIC puts a higher penalty on models with more parameters than AIC, making it more likely to select simpler models (Schwarz, 1978).

Even though both AIC and BIC can be used to compare and determine which model is the best fit for a given set of data but in generally the model which will give the lowest BIC is ideally chosen to be the best model, however, while selecting the model the model interpretability will also need to be considered.

3.4.4 Chi-Square Test of Independence

Once the homogeneous clusters will get created using the clustering technique, I would like to know that these clusters have a statistically significant relationship with the variables used in the clustering and to achieve that I will perform the chi-square test of independence.

Formula:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

Where X^2 is the chi-square test statistic, O and E are the observed frequency and the expected frequency respectively. The chi-square test performed will give p-value (for a significant level of 0.05) and if the p-value obtained is less than or equal to the significance level, then we can say that the two categories are statistically related to each other.

3.4.5 Cramer's V Test

Once using the chi-square test the relationship between the homogeneous clusters and the variables used in the creation of the model is identified I will use Cramer's V test, a statistical measure to identify the strength of the relationship between two categorical variables. That will help in understanding the scale of strength of the variables driving the cluster formation, which in turn will help the tourist department to formulate the strategies in the recovery of tourism. The Cramer's V value ranges from 0 to 1 (Lee, 2016) where a value of 0 indicates no association between the variables, while a value of 1 indicates a strong association.

Formula to calculate Cramer's V is:

$$V = \sqrt{\frac{\varphi^2}{\min(k-1, r-1)}} = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}}$$

Where,

- χ^2 is derived from Pearson's chi-squared test
- n is the grand total of observations and
- k being the number of columns.
- r being the number of rows.

Estimated values	Interpretation of association
0.00–0.10	Negligible
0.10–0.20	Weak
0.20–0.40	Moderate
0.40–0.60	Relatively strong
0.60–0.80	Strong
0.80–1.00	Very strong

Source: Wikipedia

Interpretation of Association

3.4.6 Post-Hoc Test

After determining the magnitude of the association between two categorical variables with Cramer's V test, we may want to identify which pairs of values from those variables are contributing to the strength. To accomplish this, we can use a post-hoc test. One possible post-hoc method is to conduct chi-square tests of independence on each pair of groups and apply a Bonferroni correction (Dayton & Schafer, 1973). This correction involves dividing the original significance level, typically .05, by the number of tests being performed.

4. Methodology

4.1 CRISP-DM Methodology

The CRISP-DM framework, depicted in Figure 1, will be adopted as the methodology for this data science project since it is a widely recognized framework applicable to any industry or tool. The reason behind this choice is its ability to incorporate both technical and business aspects, enabling the use of Unsupervised learning models to extract business insights in identifying factors responsible for homogeneous clusters of tourists. CRISP-DM provides the required structure and flexibility to develop high-performance data mining models. The thesis project will follow the six phases of CRISP-DM, namely business understanding, data understanding, data preparation, modelling, evaluation, and deployment, which will serve as a roadmap (Wirth & Hipp, 2000).

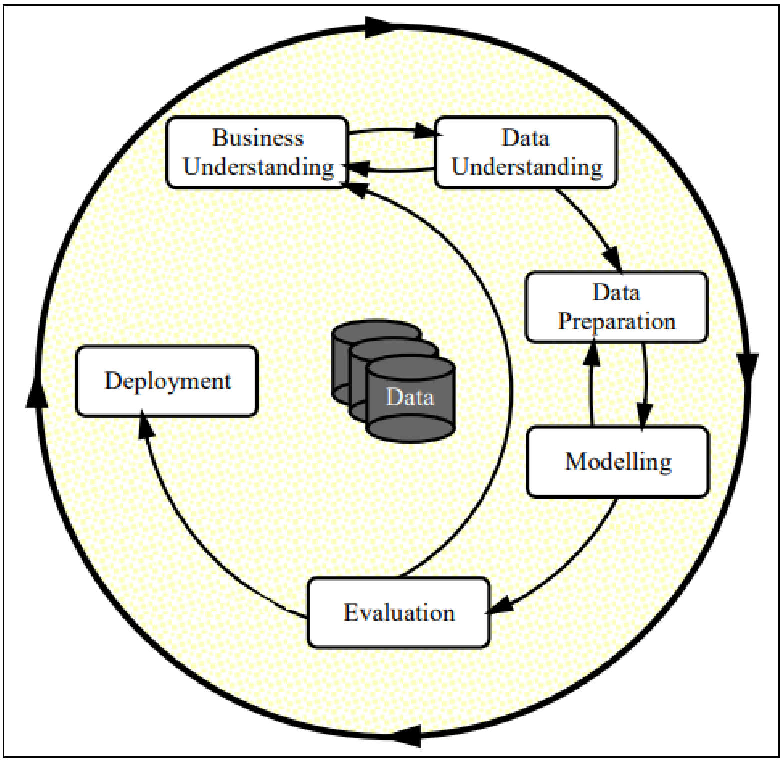


Figure 1: CRISP-DM Framework

4.2 Business Understanding

This stage is focused on building the essential foundation by analysing the objectives and requirements of the business value behind understanding the tourist's behaviours, perceptions, and intention to travel during the transition phase in a situation like Covid-19 pandemic (Wirth & Hipp, 2000).

4.2.1 Business Objectives

The timing of the survey for the thesis is critical as it was conducted during the time when the society is transitioning after the peak of the COVID-19 emergency, where it has reopened almost completely with restrictions in place to prevent the spread of the virus. However, there are still reported cases of infection and deaths, no cure or vaccine is available yet, and the risk of a second wave is acknowledged. This phase can be considered as *response* and *recovery* phase for tourism.

The goal of this thesis is to group tourists with similar risk perceptions and behaviours during a pandemic such as Covid-19, analyse the factors that create these clusters, and examine what influences their travel intentions during the transitional period. By investigating the motivational factors behind individuals' travel choices, the study will offer useful insights for devising plans to revive domestic tourism.

4.2.2 Situation Assessment

Situation assessment can be done based on the resources available for the study. The survey that was carried out covers various variables and questions that are not directly related to the specific perspective of this thesis study. However, this thesis aims to gather and analyse responses from all four countries, which represent diverse cultures and geographic locations, in order to gain insights into the global population's attitudes and behaviours related to COVID-19. By combining the data from these countries into one sample for cluster analysis, it is hoped that broader patterns and trends can be identified.

With a focus on the business goal of identifying relevant clusters of people who can help revive the tourism industry in the aftermath of the first phase of Covid-19, this study has selected a set of variables that includes preferred travel arrangements, preferred companions, factors that influence travel destination choices, and variables that relate to domestic leisure travel during the transitional period. Demographic variables have also been included in the analysis.

4.3 Data Understanding

The section dedicated to data understanding will provide an information about the data source and how it was collected. It will delve into the analysis and description of the secondary data, including its structure, data type, the variables that are encompassed within it, as well as an evaluation of the quality of the data (Wirth & Hipp, 2000).

4.3.1 Data Collection

The survey questionnaire, which aimed to examine the travel behaviour of Chinese, Danish, Italian, and Japanese travellers before and during the Covid-19 pandemic, was created by Associate Professor Prof. Fumiko Kano Glückstad of the Copenhagen Business School. Data for this study was obtained by the author of the survey from panels registered with YouGov survey company for Chinese, Danish, and Italian participants, and Cross Marketing Inc. for Japanese participants. Quota sampling was used to ensure representation based on age, gender, and geography in each country. A self-administered online questionnaire was created in English, translated into native languages, and sent via email to participants with a unique survey link. The questionnaire was mandatory, and participants were given the option to select "don't know" for uncertain responses. The survey was conducted between 10th July 2020 to 24th July 2020 both days inclusive. YouGov (an international research and data analytics organisation) provided the complete data of China (n = 1,019), Denmark (n = 1,028), and Italy (n = 1,014), while Cross Marketing provided complete data from Japan (n = 1,111) (Fumiko, 2022).

4.3.2 Data Description

The responses to the survey questions were all categorical data. It consists of both nominal and ordinal type data. The responses to the survey questions were captured using various Likert scale levels, binary (yes, no), rank based to capture respondents' past travel (domestic and abroad) behaviour, attitudes, and behaviours in the Covid-19 situation. The variables will be discussed in the detail in data exploration section.

4.3.3 Data Exploration

The quota sampling method w.r.t country, gender, and age were used while collecting the data with the target group defined by gender who are above 18 years of age and travelled abroad or within their country for either business or leisure purposes. The sample size of n=4172 used in this study has almost equal participation from each country with Denmark (n=1028), Japan (n=1111), Italy (n=1014), and China (n=1019).

Figure 2 provides the demographic distribution of each country when compared with gender and age groups respectively.

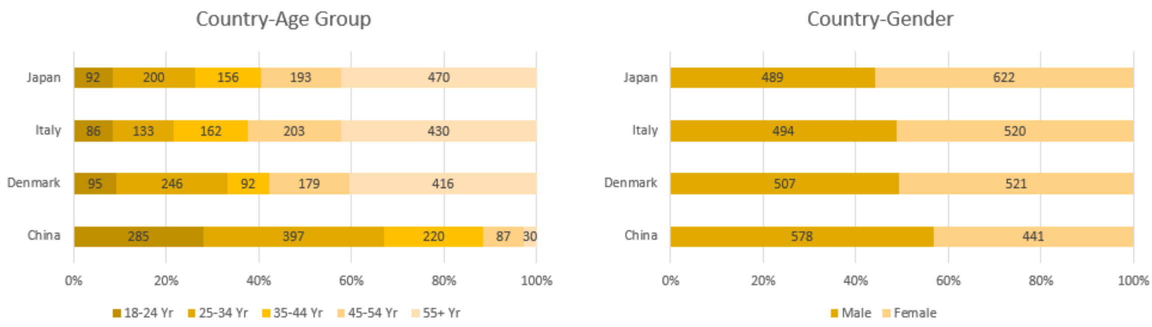


Figure 2: Age and Gender distribution

Source: Inspiration from (Fumiko et al.,2021)

The sample overall has almost equal participation of Males (49.6%) and Females (50.4%), however, if we compare it with the country-specific distribution then the male participation (56.7%) in China is considerably higher than the female participation whereas in Japan the female participation (56%) is considerably higher than the male participation. The gender distribution of Italy and Denmark is equivalent to the overall sample gender distribution. In the overall sample, the respondents in the age group of 55+ years are highest with 32.3%, followed by the age group 25-34 years (23.4%) with other three age groups with almost equal presentation (around 15%), however, this overall sample doesn't represent evenly when we look at the country-specific distribution. Chinese respondents are majorly young with around 67% of respondents under the age of 35 years with merely 3% of respondents in the 55+ age group. Denmark, Italy, and Japan have almost equivalent participation of people above 55 years of age (around 40-42%).

The survey conducted had many variables, however, for this study I have chosen 43 variables. Apart from demographics (country, age; and gender), the variables include past travel experience (foreign and domestic), travel arrangement preference, travel companion preferences, expectations from the next travel destinations, attitudes towards self-protective and responsible behaviour, and of public, travel intentions during the transitional phase and attitude and assumptions during the travel during the transitional phase. Using these variables in a COVID-19 tourism recovery study can provide valuable insights for businesses, policymakers, and researchers to develop effective strategies to promote tourism recovery in the situations like Covid-19. Each variable is a question from the survey on which the respondents provided their responses. All the 40 questions have been categorised into 12 categories as mentioned below:

Travel Experience: The purpose of gathering travel experience information from respondents is to determine the frequency of their travel in the last two years (2018 and 2019), whether for business or leisure purposes abroad or for leisure domestically. By analysing this information in conjunction with other factors, it is possible to anticipate a similar pattern of demand for domestic travel after the first phase of Covid.

Table 1: Travel Experience (% of sample)

Questions	Not at all	1-3 times	4-6 times	7-12 times	13+ times	Don't know
How many times have you travelled overseas for business purposes, within the last 2 years?	68.2	21.1	6.7	1.9	1.4	0.7
How many times have you travelled overseas for leisure purposes, within the last 2 years?	8.8	68.2	16.3	4.9	1.4	0.3
How many times have you travelled domestically for leisure purposes (with an overnight stay), within the last 2 years?	12.5	47.2	25.8	9.0	4.2	1.3

Source: Own elaboration

Table 1 shows that around 68% of the respondents have not travelled overseas for business purposes within the last two years of the time when the survey was conducted, and assuming that mostly the business travel is sponsored by the companies hence we see the low participation here. However, if we look at overseas travel for leisure purposes then it is very clear that the respondents have an appetite for tourism as around 91% of the respondents at least travelled between one to three times. A similar trend is reflected in domestic travel for leisure purposes where around 87% of the respondents travelled at least once for leisure purposes. This shows a positive attitude and willingness to travel for vacation and leisure activities among the population. Overall, the table shows that around 31 % of respondents have travelled at least once in the last two years for business purposes, and around 90% travelled overseas at least once for leisure purposes however around 86% of respondents travelled domestically for leisure purposes. In situations like COVID-19 where overseas travel gets banned, the tourism economy gets open with domestic travel first and the survey asked questions about respondents' intention to travel during the transitional phase (it will be discussed later in this chapter).

Preferred Travel Arrangement: Understanding the preferred travel arrangement of tourists, whether it is through a third-party arrangement (such as travel agencies or tour operators) or self-planned, can help in the recovery of tourism after the first phase of Covid-19.

If tourists exhibit a preference for third-party arrangements, it may indicate their inclination towards a more convenient and stress-free travel experience. This trend could provide travel agencies and tour operators with an opportunity to offer tailored packages that meet the specific needs of post-Covid tourists, who may be seeking such services. Conversely, if tourists tend to opt for self-planned travel arrangements, it may reflect their need for greater control over their travel experience, leading to an increase in demand for rental accommodations, car rentals, and other services that facilitate independent travel. This could present an opportunity for businesses operating in these sectors to cater to the needs of such tourists and develop strategies that align with their preferences.

Table 2: Preferred Travel Arrangement (% of sample)

Rank/Travel Arrangement	Arranged by a third party	Self-planned
First Rank	22.0	78.0
Second Rank	78.0	22.0

Source: Own elaboration

Upon analysing Table 2, it is evident that a significant majority of tourists (78%) have a preference for self-planning their travel arrangements.

Preferred Travel Companion: Understanding the preferred travel companion, such as "Travelling with a larger group (above 8 people)", "Travelling with closest family or friends" and "Travelling alone or with a significant other", can provide insights into the social aspect of travel and help in the recovery of tourism after the first phase of COVID-19.

Table 3: Preferred Travel Companion (% of sample)

Rank/Companion	Traveling with a larger group (above 8 people)	Traveling with closest family or friends	Traveling alone or with a significant other
First Rank	6.2	28.9	64.9
Second Rank	41.7	39.4	19.0
Third Rank	52.2	31.7	16.1

Source: Own elaboration

For example, if travelling with larger groups is still perceived as a risk due to the virus, promoting destinations and activities that are suitable for small groups or individuals can help in attracting tourists who prefer travelling alone or with a significant other. Alternatively, if travelling with family or friends is deemed a safer option, destinations and activities that cater to larger groups can be emphasized. By understanding the preferred travel companions, tourism stakeholders can tailor their marketing and promotional strategies to cater to the changing preferences and needs of tourists post-COVID-19. Table 3 indicates that approximately 65% of respondents prefer to travel either alone or with a significant other as their first choice, while the preference of 29% is to travel with family or friends.

Factors Impacting Travel Destination Choice: Understanding the importance of these factors can provide insights into the changing preferences of tourists after the first phase of Covid-19.

Table 4: Importance of Next Travel Destinations (% of sample)

Question	Much less important	Less important	Neither important nor unimportant	More important	Much more important	Don't know
A destination where I have previous travel experiences	8.2	19.6	38.2	18.6	8.8	6.5
A safe travel destination	1.6	2.7	14	33.4	44.2	4.2
A destination with good shopping possibilities	11	17.3	34.2	21.9	11	4.7
A destination offering possibilities to visit museums, exhibitions, historical attractions	4.9	12	33.4	29.7	14.6	5.4
A destination offering possibilities to visit amusement parks, zoos, water parks and so on	13.6	19.3	31.5	20.4	9.8	5.2
A destination with good restaurants, cafes, bars and so on	5.4	12.7	32.8	30.2	13.8	5.1
A destination with good hygiene which minimises risks of spreading infectious diseases	1.7	2.9	14.3	33.7	43.7	3.8
Peace and quiet - a destination with less tourists	2.3	4.5	24.4	38	27	3.8
A destination close to a beach, harbour and coast line	4.9	9.4	33	29.7	17.5	5.5
A destination with forests and nature	3	6.6	28.3	35.7	22.2	4.2
Children friendly destination	20.6	14.2	29	18.4	11.1	6.7
A destination friendly for senior citizens	16	15.2	34.9	18	9.3	6.5
Clean destination (no trash, clean beach and air)	2	3.2	18.8	38.4	33.3	4.3

Source: Own elaboration

For example, if safety and hygiene are now deemed more important than previous travel experiences, it suggests that tourists are more cautious and risk-averse due to the pandemic. Similarly, if destinations with fewer tourists and natural attractions are now preferred, it suggests a shift towards more secluded and nature-based destinations. Understanding these changing preferences can help tourism businesses and destinations adapt and cater to the needs and wants of tourists in the post-Covid era. Table 4 reveals that tourists prioritize choosing a safe destination (approximately 78%), a destination with good hygiene to minimize the spread of infectious diseases (78%), and a clean destination (with clean beach and air and no trash) (approximately 71%). Nature and forests are also important to around 58% of respondents, while around 47% consider proximity to beaches, harbours, and coastlines significant. Tourists value peace and quiet, with 65% opting for less touristy destinations. Interestingly, roughly the same percentage of respondents (around 30%) consider child-friendly and senior-friendly destinations important. Additionally, around 38% of respondents stated that their past travel experiences did not factor into their decision-making when selecting a holiday destination this year compared to last year.

Attitudes to Public Behaving Responsible: The attitude and behaviour of individuals towards protecting themselves and society have a significant impact on the spread of infections that can affect tourism. According to the survey results (Table 5), 82% of respondents strongly believe that their behaviour in public can play a crucial role in reducing the risk of infection.

Table 5: Attitudes to Public Behaving Responsible (% of sample)

Question	Strongly Disagree	Disagree	Somewhat Disagree	Neither Disagree nor Agree	Somewhat Agree	Agree	Strongly Agree	Don’t Know
It is important that individuals contribute to minimise the risk of spreading infectious diseases in public spaces	0.7	1.6	2.7	9.8	13.2	24.6	45	2.3
I feel safe and comfortable if staffs in hotels, airlines, restaurants etc. wear a mask	2.1	3.5	5.3	18.9	20.7	23.2	23	3.3

Source: Own elaboration

Additionally, 68% of respondents feel safer and more comfortable when staff members in hotels, airlines, and restaurants wear masks.

Attitudes to Self-Protective Behaviours: Almost every country stressed the measures individuals must take to restrict the spread of the virus and it is reflected in the Table 6 where around 74% of respondents agreed with using social distancing in public spaces, around 64% of respondents mentioned that they use disinfectant to clean hands while shopping and around 60% mentioned that they wear the masks to feel safe.

Table 6: Attitudes to Self-Protective Behaviours (% of sample)

Question	Strongly Disagree	Disagree	Somewhat Disagree	Neither Disagree nor Agree	Somewhat Agree	Agree	Strongly Agree	Don't Know
I am keeping social distances in public spaces. If it is not possible, I will leave that place	1.4	2.6	5.2	14.5	22.2	26.6	25.5	2.1
I carry and use disinfectant to clean my hand after touching items in shops to make me feel clean and safe	4.2	6.8	7.2	15.7	15.8	22	26	2.3
I wear a mask to make me feel safe	9.8	7.4	5.4	14.3	14.5	19.7	26.2	2.6

Source: Own elaboration

Overall responses reflect that the tourists are concerned for their safety, and they are adhering to the preventive measures to reduce the risk of infection.

Attitudes to Responsible Behaviours: Table 7 shows that around 68% of respondent agree about cleaning public spaces like toilets after use for subsequent users. Around 63% of respondents agree about using disinfectant to clean their hands before touching things in the shop while around 65% of respondents say that they use masks to make feel safe and comfortable around them.

Table 7: Attitudes to Responsible Behaviours (% of sample)

Question	Strongly Disagree	Disagree	Somewhat Disagree	Neither Disagree nor Agree	Somewhat Agree	Agree	Strongly Agree	Don't Know
I clean up a public space (e.g. Toilet) after I use it so that people who use it after me feel clean and safe	2	4.2	5.1	17.8	19.1	26.6	21.8	3.4
I carry and use disinfectant to clean my hands before touching items in shops so that other people who touch after me feel clean and safe	4	6.2	8.2	16.1	16.9	22.3	24	2.3
I wear a mask to keep those around me safe and comfortable	8.7	7.7	4	12.3	14.1	22.2	28.4	2.7

Source: Own elaboration

The overall responses indicate that people take individual responsibility to prevent infectious diseases. They are willing to take necessary actions to keep public spaces and people safe. This information can be useful in developing public health campaigns to promote responsible behaviour and in planning measures to prevent infectious diseases.

Transitional Phase: The survey in question was conducted during the time when the peak caused by Covid-19 in the surveyed countries is getting over. The author of the survey called this phase the *transitional phase*. This phase is characterized as the phase where the societies are re-opened almost completely, with various restrictions in place like social distancing, use of disinfectants, and wearing masks in public places to prevent the further spread of the virus. During this phase the risk of Covid-19 second wave is acknowledged, the infection cases and few deaths because of the Covid-19 virus are still being reported and there was no confirmed cure or vaccine available. So, to gauge respondents' intention to travel during this phase in the next six months, the author of the survey asked one direct question about their intentions to travel with other questions that might impact their intentions to travel.

Attitudes to Travel within the Country: The respondents were asked to rate their feeling about various factors on a scale of 1-7. The two feelings as shown in (Table 8) were mentioned at the extreme ends of the scale. On average around 44% of respondents felt that domestic travel during the transitional phase (for the next six months) will be safe (44%), enjoyable (47%), effortless (40%), and beneficial (43%) while on average around 29% of respondents find such travel to be dangerous (30%), unenjoyable (30%), onerous (34%), and harmful (25%). While around 26% of respondents demonstrated the neutral stance towards all the categories.

Table 8: Attitudes to Travel within the Country (% of sample)

Question	1	2	3	4	5	6	7
left="Dangerous" right="Safe"	6.4	9.3	14.2	26	19.7	17.5	6.9
left="Unenjoyable"right="Enjoyable"	6.2	8.4	12.3	26.1	20.8	18.6	7.6
left="Onerous" right="Effortless	6.7	10.6	16.3	26.2	17.3	16	7
left="Harmful" right="Beneficial	5.2	6.7	13.1	32.3	19.4	16	7.4

Source: Own elaboration

These insights can be used to better understand the overall sentiment towards domestic travel and to develop targeted strategies to address the concerns of those who feel negatively about it.

Intention to Travel within country for Leisure: Around 49% of respondents (Table 9) agreed to travel in the transitional phase for leisure purpose. However, there was an almost equal percentage 25% of respondents who are neutral towards the travel decision with almost similar percentage of respondents disagreeing to travel during the transition phase within the country for leisure purposes.

Table 9: Intention to Travel within country for Pleasure (% of sample)

Question	Strongly disagree	Disagree	Somewhat disagree	Neither disagree nor agree	Somewhat agree	Agree	Strongly agree
I intend to travel for pleasure within Country in the transitional phase	7.2	8.1	10.5	25.5	21.1	18.5	9.1

Source: Own elaboration

The raw information suggests that there is a considerable demand for domestic tourism even during the pandemic, providing opportunities for travel businesses to offer services that cater to the needs of post-COVID tourists. However, it also implies that attracting tourists during this phase may be challenging, and specific efforts may be required to address the concerns of those hesitant to travel.

Assumption to My Travel by Known People: Individuals do value the perceptions of their known ones towards their actions. Table 10 shows around 54% of respondents believe that people who are important to them will support their travel during the transitional phase as they will believe that the travel decision would have been given proper thought keeping in might the safety of the respondent, as well as others, and around 53% respondents, believe that their known ones hence approve their decision of travel during the transitional phase.

Table 10: Assumption to My Travel by Known People (% of sample)

Question	Strongly disagree	Disagree	Somewhat disagree	Neither disagree nor agree	Somewhat agree	Agree	Strongly agree
Most people who are important to me think that my traveling within Country in the transitional phase will be thoughtful and respectful of their and my safety	3.9	5.7	8.9	28.5	21.7	21.9	9.6
Most people whose opinion I value would approve my traveling within Country in the transitional phase	3.5	6.5	9.9	30.8	20.7	19.3	9.2
Most people I respect and admire will be traveling within Country in the transitional phase	5.1	8.7	10.8	33.8	20	16.4	5.3

Source: Own elaboration

The information reflects that social influence is an important factor in an individual's decision to travel during the transitional phase. This suggests that people who will be willing travel can influence the people who matter to them or people who are close to them.

Resources (Media, Money & Time) Role during the Transitional phase: In the age of the internet the role of mass and social media has become a very important factor in influencing the travel plan of people, around 51% of respondents (Table 11) in the survey agreed that it will encourage their travel plans. Also, travel and time are other two factors that also influence the travel plan, and this is also evident when around 57% of respondents agree on the time and 54% of respondents agree on the money factor as well which will influence their travel plan in the next six months of transitional phase.

Table 11: Resources (Media, Money & Time) Role during the Transitional phase (% of sample)

Question	Strongly disagree	Disagree	Somewhat disagree	Neither disagree nor agree	Somewhat agree	Agree	Strongly agree
Mass and social media will encourage traveling activities within Country in the transitional phase	2.9	4.9	9.7	31.2	24.6	19.6	7.1
If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available financial resources to do it	4.4	5.5	9.5	27	23.6	20.4	9.6
If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available time to do it	2.9	4.9	9.5	26.1	25.4	22.7	8.4

Source: Own elaboration

These findings suggest that travel businesses may need to focus on social media marketing to attract potential customers. They may also need to offer affordable and flexible travel packages to cater to the budget and time constraints of travellers during the transitional phase.

COVID-19 Perception during Transitional Phase: Tourism revival in situations like the Covid-19 pandemic will be possible slowly. Moving out and travelling for pleasure purposes will be rather slow as compared to moving out because of work. However, the response to the survey (Table 12) is quite encouraging where around 42% of respondents mentioned that they might not contribute to the spread of Covid-19 if they travel during the transitional phase as compared to 32% of respondents who believe that they might contribute to the spread of Covid-19 disease. Also, the respondents (38%) who believe that they won't get infected by Covid-19 are 25% more than the respondent (30%) who feel that they might get infected by the virus if they travel during the transitional phase. This difference is around 57% among the respondents (48%) who believe that they might not infect others in comparison to the respondents (28%) who believe that they might infect others with Covid-19 because of their travel during the transitional phase.

Table 12: COVID-19 Perception during Transitional Phase (% of sample)

Question	Very unlikely	Unlikely	Somewhat unlikely	Neither unlikely nor likely	Somewhat likely	Likely	Very likely
If I travel within Country in the transitional phase, I will contribute to the spread of Covid-19 at the destination	12.7	12.4	16.4	27	18.5	8.7	4.3
If I travel within Country in the transitional phase, I will get infected with Covid-19	8.4	12.1	17.8	31.3	19.1	7.8	3.5
If I travel within Country in the transitional phase, I will infect others (relatives, friends, colleagues etc.) with Covid-19	13	13.8	17	28.3	17.1	7.1	3.8

Source: Own elaboration

Overall, the responses suggest that there is a level of confidence among the respondents regarding the safety of travel during the transitional phase, which could potentially support the recovery of the travel industry.

4.4 Data Preparation

The thesis aims to accomplish two primary objectives. Firstly, it attempts to establish the equivalence of the unsupervised machine learning algorithm K-modes clustering method (used for categorical data) with the statistical method LCA for finding the group or cluster using the same variables. Secondly, the thesis aims to identify the homogeneous clusters and the factors responsible for such clusters, which can be utilized in the recovery of tourism during a Covid-19-like situation. The thesis adopts a broader perspective to gain insights into the recovery of tourism. In this regard, the combined data of four countries are considered as a representative sample of the overall population, assuming that similar response patterns will be observed in other countries of Europe and Asia.

4.4.1 Variables Selection

Recovering the tourism industry during a pandemic like Covid-19 can be a daunting task, as people may be apprehensive about leaving their homes and risking infection. However, they may also feel tired of their mundane routine due to lockdown restrictions and want to explore the outdoors. Therefore, identifying factors such as preferred travel arrangements, travel companions, and important vacation factors for travellers can help tourism operators promote relevant options to attract travellers. Furthermore, combining these variables with perceptions and behaviours during the transition phase can aid in the development of tourism recovery strategies by tourism-related sectors such as travel operators, hotels, and government officials.

With this perspective in mind, the thesis uses 40 variables in clustering, including demographic variables such as age, sex, and country, a variable for the first-ranked travel arrangement, a variable for the first-ranked travel companionship, 13 variables for holiday destination factors, two variables for attitudes towards responsible public behaviour, three variables for attitudes towards self-protective behaviour, three variables for attitudes towards responsible behaviour, and 14 variables related to the transition period, including the intention to travel within the country (in next six months) for pleasure purposes.

4.4.2 Pre-processing

Since the Likert responses capture the three dimensions of negative, neutral, and positive feelings or opinions, the brute force method of manually combining the scales is utilized. This approach may cause a loss of information in extreme cases but as I am working on the broader perspective of identifying the factors broadly which will help in the recovery of the tourism industry, I assume that the approach is good. Additionally, the scales will be cross-culturally motivated as the people from different countries might be using the different ends of the scales, i.e., the people from one country generally would like to use the extreme end of the scales while some would hesitate to use those extreme end values. By combining the scales, the thesis aims to minimize this heterogeneity (Lipovetsky & Conklin, 2018) (Rossi et al., 2001). Finally, it will help in making a meaningful interpretation and explanation from the output achieved to form the generalized view.

The survey responses measured on the Likert scale of 1-8 where merged to into the three responses “Disagree”, “Neutral”, and “Agree” where “1-Strongly disagree”, “2-Disagree", and “3-Somewhat disagree” are merged to term it as “Disagree” with value 1; “4-Neither disagree nor agree” and “8-Don’t know” are merged and termed as “Neutral” with value 2; “5-Somewhat agree", “6-Agree", and “7-Strongly agree" are merged and termed as “Agree” with value 3. The values of the 6-point Likert scale where 1 is “Much less important”, 2 is “Less important” are merged together to term it as “Unimportant” with value 1; 3 is “Neither important nor unimportant” and 977 as “Don’t Know” are merged together to be termed as “Neutral with value 2; 4 is “More important” and 5 is “Much more important” are merged to be termed as “Important” with value 3. The values of the 7-point Likert scale are also manually merged; here 1 representing “very unlikely”, 2 representing “unlikely”, 3 representing “somewhat unlikely” are merged as “unlikely” with value 1; 4 representing “neither likely nor unlikely” is termed as “neutral” with value 2; and 5 representing “somewhat likely”, 6 representing “likely”, 7 representing “very likely” are merged as “likely” with the value 3. Finally, the values of the 7-point Likert scale with values represented at extreme ends are also manually combined (e.g. the response “dangerous” mentioned at the left end showing some sort of danger with values 1,2,3 are merged with values 1 and termed as “dangerous”; value 4 is considered as “neutral”; and values 5,6,7 are combined together to be termed as “safe” with value 3 which on the survey is mentioned at extreme right).

Likert Response with Eight values

Original	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither disagree nor agree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Manually Combined	Disagree (1)			Neutral (2)	Agree (3)		

Likert Response with seven values

Original	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither disagree nor agree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Manually Combined	Disagree (1)			Neutral (2)	Agree (3)		

Likert Response with values at extreme level with seven values

Original	Dangerous (1)	2	3	4	5	6	Safe (7)
Manually Combined	Dangerous (1)			Neutral (2)	Safe (3)		

In the software R-studio the number 0 is not computed for Latent Class Analysis (LCA) hence the value 0 for the country variable is changed to 1, however, before that value 1 was changed to 2, 3 to 4, and 4 to 5. Also, keeping value 977 is computationally inaccurate because when running the LCA model in R-studio it assumes all the valid values until 977, so to avoid that initially 977 was converted to the next available highest number of the respective variable, so if the highest number used for the variable was 7 (apart from 977) then 977 was changed to 8.

Tools used in the data analysis, data preparation, and modelling are R-studio and Excel.

4.5 Modelling

In this study, as the data is categorical hence, we will perform clustering using two methods one unsupervised machine learning algorithm K-Modes Analysis and another statistical method called Latent Class Analysis (LCA).

4.5.1 K-Modes Analysis

The K-modes algorithm is a clustering algorithm that is used to cluster categorical data. It is an extension of the K-means algorithm, which is designed for continuous numerical data. K-modes replace the means of clusters with modes and use a frequency-based method to update the modes in the clustering process (Huang, 1998). The K-modes algorithm employs a simple matching dissimilarity measure to handle categorical objects, which counts the number of features that differ between two data points. The algorithm then uses this measure to compute the similarity between data points, rather than using Euclidean distance as is commonly done in K-means. The K-modes algorithm aims to minimize the clustering cost function by iteratively updating the modes and re-assigning data points to the cluster with the closest mode. This process continues until convergence is reached, and the final clusters represent the groups of data points that are most like each other in terms of their categorical features.

The K-modes analysis is carried out using the software R with klaR (R: K-Modes Clustering, n.d.) library. The optimal number of clusters was identified using the Silhouette Coefficient (SC) (Rousseeuw, 1987). The Silhouette Coefficient (SC) is a metric used to assess the quality of a clustering solution. It measures how well each data point (e.g., survey respondent) fits into its assigned cluster, as well as how appropriate the assignment is relative to other clusters. The Silhouette Index (SI) is the average SC across all data points and provides an overall measure of how well the data is separated into clusters. To calculate the SC, the distance between each data point and the other point within its assigned cluster is compared to the distance between that data point and the nearest neighbouring cluster. The SC ranges from -1 to 1, with values near -1 indicating inappropriate clustering and values near 1 indicating highly compact clustering (Rousseeuw, 1987). An SC near zero suggests that a data point may be assigned to overlapping clusters. In general, a high average SI indicates that the clusters are dense and well-separated.

Optimal no of K-modes: To find the number of optimal k-modes on the given data the Silhouette Index (SI) values were calculated for the k-modes algorithm for two to ten clusters model. It gave the three-cluster model (Figure 3) as the best model.

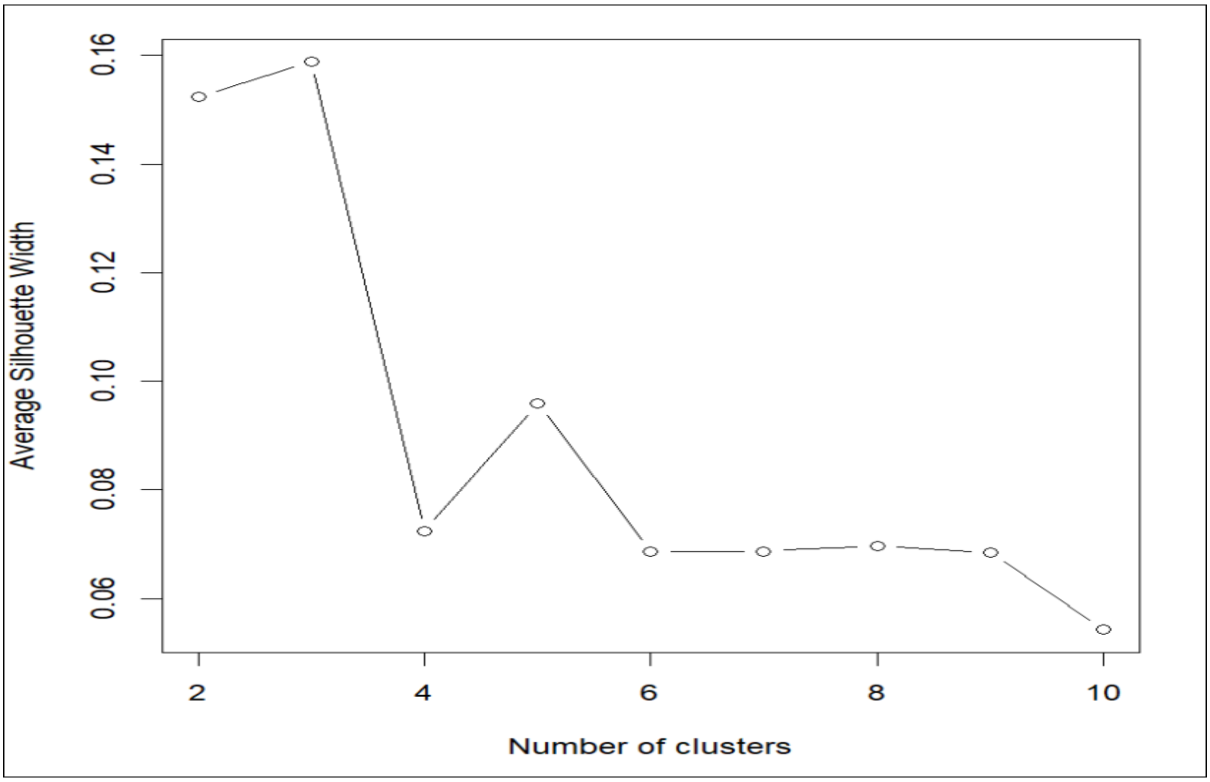


Figure 3: No of K-modes cluster

Source: Own elaboration

The Figure 3 shows clearly that the three k-modes are the optimal number of k-modes for the given data.

Table 13: K-Modes-Silhouette Indices (SI) for Two-Through Four-Class Solution

Model	Silhouette Index
Two cluster	0.152
Threee cluster	0.159
Four cluster	0.072

Source: Own elaboration

This is also supported by the Silhouette Index (SI) values (Table 13), where the three-cluster model has the highest SI (0.159) value.

4.5.2 Latent Class Analysis (LCA)

Latent Class Analysis (LCA) is a statistical method used to identify unobserved subgroups or clusters, or "latent classes," within a population based on observed categorical data. LCA is a model-based clustering approach that seeks to identify subgroups of individuals who share similar characteristics or behaviours but differ from individuals in other subgroups (Collins &

Lanza, 2010). In LCA, researchers typically start by specifying a hypothesized number of latent classes (i.e., the number of subgroups they expect to exist in the population). The method then estimates the probabilities of each individual belonging to each of the latent classes, based on their responses to a set of categorical variables. These variables could be, for example, responses to survey questions, diagnostic criteria, or demographic variables. It is a probabilistic model, which means that it provides estimates of the probability of each individual belonging to each latent class (Collins & Lanza, 2010).

The Latent Class Analysis is carried out using the software R with poLCA (poLCA Function - RDocumentation, n.d.) library. For the relative model fit the optimal number of latent classes for this LCA was selected based on the Akaike information criterion (AIC; Akaike, 1987) and the Bayesian information criterion (BIC; Schwartz, 1978). Theoretically, the smaller value of information criteria represents a more optimal balance of model fit and parsimony; hence the model with the minimum BIS will be chosen.

Optimal number of Classes: While there is no unanimous agreement on the ideal criteria for comparing latent class solutions, it is generally accepted that when choosing a final model, multiple fit statistics should be employed (or at least disclosed), and the Bayesian information criterion (BIC) might be the most dependable fit statistic that should be reported consistently. Moreover, theoretical interpretability must be considered while selecting a solution. Considering my study, I ran the LCA model for two classes to ten classes and the smallest BIC is for ten class model. However, considering that I will be comparing the K-modes analysis and LCA, I am displaying the information criteria for two to four classes LCA model.

Table 14: AIC and BIC for Two-Through Four-Class Solution

Model	AIC	Diff AIC	BIC	Diff BIC
Two class	342023	0	343258	0
Threee class	330529	11494	332385	10873
Four class	323387	7142	325865	6520

Source: Own elaboration

We can notice from Table 14 that BIC for four class model is smaller than the three-class model which suggests that the four-class model is better than the three-class model. However, considering the study where I want to compare whether the results obtained by the two models are comparable or not, I will still go with the three-class model which is suggested as the optimal number of modes by the k-modes algorithm.

4.6 Evaluation

4.6.1 Agreement Between Three Clusters Identified (K-mode and LCA).

Based on statistics used for model evaluation we selected the three cluster K-modes and LCA respectively. These three clusters identified will also be interpretable. Table 15 shows that the number of respondents in the three sub-groups is proportionate in both K-modes and LCA.

Table 15: Cluster Distribution for K-Modes and LCA

Cluster Distribution for K-Modes and LCA			
	Cluster 1 "Risk-Takers"	Cluster 2 "Risk-Averse"	Cluster 3 "Risk-Neutral"
K-modes	2154 (52%)	1310 (31%)	708 (17%)
LCA	2438 (58%)	962 (23 %)	772 (19%)

Source: Own elaboration

4.6.2 Pairwise Agreement between LCA and K-Modes Analysis

The pairwise agreement analysis from Table 16 signifies that the observed agreement among the three clusters is 80%, and the expected agreement is 20%. The “Risk-Takers” cluster of LCA and K-modes analysis have 47% agreement, The “Risk-Averse” cluster has 19% agreement, and the “Risk-Neutral” have 14% agreement among each other. The noticeable disagreement between the LCA and K-modes cluster arises is Risk-Averse of (K-modes) with Risk-Takers of LCA which is 11% while only slight disagreement between the LCA and K-modes cluster arises is i.) The Risk-Takers from (K-modes) group and Risk-Averse and Risk-Neutral from (LCA) group ii.) The Risk-Neutral from (K-modes) group and Risk-Averse from LCA (group).

Table 16: Pairwise Agreement among the clusters identified using K-Modes and LCA

Pairwise Agreement Among the Respondent Clusters using K-Modes Analysis and Latent Class Analysis

Pairwise Agreement among the Respondent Clusters		LCA Risk-Takers n^{a} (%[c])	LCA Risk-Averse n^{a} (%[c])	LCA Risk-Neutral n^{a} (%[c])	Total n (%)
K-Modes	Risk Takers n^{b} (%[c])	1942 (47)	84 (2)	128 (3)	2154 (52)
	Risk-Averse n^{b} (%[c])	457 (11)	785 (19)	68 (2)	1310 (32)
	Risk-Neutral n^{b} (%[c])	39 (0)	93 (2)	576 (14)	708 (16)
	Total n (%)	2438 (58)	962 (23)	772 (19)	4172 (100)

[a] *for LCA: Risk-Takers (n=2154, 52%), Risk-Averse (n=1310, 31%); Risk-Neutral (n=708, 17%)*
[b] *for K-modes: Risk-Takers (n=2438, 58%), Risk-Averse (n=962, 23%); Risk-Neutral (n=772, 19%)*
[c] *percentage of respondents from the total sample of 4172 respondents.*

Source: Own elaboration

To make sure that the agreement between the clusters of K-modes analysis and latent component analysis obtained is not just by chance the inter-rater reliability measure Cohen's Kappa is performed that gave the estimated value of 0.65 (Table 17) which on the basis of Cohen's Kappa table falls in the range (0.61 – 0.80), and it means that there is substantial agreement between the K-modes clusters and LCA classes. Hence, either of the two methods can be used for such studies to understand the behaviours of the sub-groups. I will be using the K-modes analysis for my study to further analyse the clusters.

Table 17: Cohen's Kappa Coefficient

Cohen Kappa and Confidence Boundaries	Lower	Estimate	Upper
Unweighted Kappa	0.63	**0.65**	0.67

Source: Own elaboration

4.6.3 Analysis of the Factors Influencing the Three clusters using K-Modes Analysis

Table 18 will be used as a basis for three clusters analysis descriptions.

Cluster-1: "Risk-Takers"

This is the largest group in the sample accounting for 52% or 2154 of the respondents of the sample size. This group is composed of respondents from Denmark (35%), Italy (33%), and

China (25%), with a minority of Japanese individuals representing only 7% of the group. The majority of the population of this group belongs to two age distribution one with 35% of respondents aged 55 years and above and another 21% of respondents aged between 25 and 34 years old. Males have the majority representing 56% of the respondents of this group, while females make up 44%.

Upon analysing the group's travel experiences in the last two years, it was observed that the majority of respondents (68%) did not engage in international travel for business purposes. On the other hand, approximately 95% of respondents travelled outside their country of residence for leisure purposes at least once to three times. Similarly, around 86% of respondents travelled within their country of residence for leisure purposes at least once to three times, with 38% of respondents undertaking at least four to six trips in the last two years.

Table 18: Three Clusters with their Distribution

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics
Demographics	Country	**Denmark(757, 35%)** Italy(698,33%) China(547,25%) Japan(152,7%)	**Japan(640,49%)** China(354,27%) Italy(198,15%) Denmark(118, 9%)	**Japan(319,45%)** China(118,17%) Denmark(153, 21%) Italy(118,17%)	X^2=1014.7 ;p=2.2e-16 Cramers_v=0.35
	Profile_Age	**55+ Years (760,35%)** 25-34 Years (448,21%) rest are almost equal with around 14%	**55+ Years (436,33%)** 25-34 Years (312,24%) rest are almost equal with around 14%	**25-34 Years (216,31%)** 55+ Years (150,21%) rest are almost equal with around 16%	X^2 = 70.16 p = 4.562e-12 Cramers_v=0.09
	Gender	**Male (1207,56%)** Female (947,44%)	Male (457, 35%) **Female (853, 65%)**	**Male (404, 57%)** Female (304, 43%)	X^2 = 164.92 p < 2.2e-16 Cramers_v=0.20
Travel Experience	q1_1 (overseas for business in last 2 years)	Not at all (1463, 68%) 1-3 times (463, 22%)	Not at all (940, 72%) 1-3 times (264, 20%)	Not at all (471, 67%) 1-3 times (153, 22%) 4-6 times (53, 9%)	X^2=15.08 p=0.0576 Cramers_v=0.03
	q1_2 (overseas for leisure in last 2 years)	Not at all (112, 5%) **1-3 times (1430, 66%)** 4-6 times (498, 20%)	Not at all (166, 13%) **1-3 times (847, 73%)** 4-6 times (139, 11%)	Not at all (101, 14%) **1-3 times (467, 66%)** 4-6 times (103, 15%)	X^2=154.1 p < 2.2e-16 Cramers_v=0.13
	q1_3 (domestic for business in last 2 years)	Not at all (298, 14%) 1-3 times (936, 44%) 4-6 times (602, 28%) 7-12 times (223, 10%)	Not at all (140, 11%) 1-3 times (680, 52%) 4-6 times (318, 24%) 7-12 times (122, 9%)	Not at all (138, 19%) 1-3 times (354, 50%) 4-6 times (157, 21%)	X^2=71.263 p=2.753e-12 Cramers_v=0.09
Preferred Travel Arrangement	q3_1 (Preferred Travel Arrangement)	Arrange by Third Party (401, 19%) **Self-Planned (1753, 81%)**	Arrange by Third Party (330, 25%) **Self-Planned (980, 75%)**	Arrange by Third Party (188, 27%) **Self-Planned (520, 73%)**	X^2=30.668 p=2.19e-07 Cramers_v=0.08
Preferred Travel Companionship	q4_1 (preferred travel companionship)	Male (1207,56%) Female (947,44%)	Larger group (above 8 people) (65, 5%) Closest family or friends **(566, 43%)** Alone or with a significant other **(679, 52%)**	Larger group (above 8 people) (65, 9%) Closest family or friends (256, 37%) Alone or with a significant other **(387, 55%)**	X^2=298.37 p < 2.2e-16 Cramers_v=0.19

Table…continued

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics
How Important the factors for next travel	q11_5 (A destination where I have previous travel experiences)	Unimportant (720, 33%) **Neutral (816, 38%)** Important (618, 29%)	Unimportant (339, 26%) **Neutral (551, 42%)** Important (420, 32%)	Unimportant (99, 14%) **Neutral (500, 71%)** Important (109, 15%)	X^2=254.47 p < 2.2e-16 Cramers_v=0.17
	q11_6 (A safe travel destination)	Unimportant (89, 4%) Neutral (229, 11%) **Important (1836, 85%)**	Unimportant (45, 3%) Neutral (116, 9%) **Important (1149, 88%)**	Unimportant (43, 6%) **Neutral (413, 58%)** Important (252, 36%)	X^2=960.66 p < 2.2e-16 Cramers_v=0.34
	q11_7 (A destination with good shopping possibilities)	**Unimportant (754, 35%)** Neutral (704, 33%) Important (696, 32%)	Unimportant (337, 26%) Neutral (408, 31%) **Important (565, 43%)**	Unimportant (89, 13%) **Neutral (509, 72%)** Important (110, 15%)	X^2=446.18 p < 2.2e-16 Cramers_v=0.23
	q11_10 (A destination offering possibilities to visit museums, exhibitions, historical attractions)	Unimportant (405, 19%) Neutral (626, 29%) **Important (1123, 52%)**	Unimportant (224, 17%) Neutral (474, 36%) **Important (612, 47%)**	Unimportant (76, 11%) **Neutral (520, 73%)** Important (112, 16%)	X^2=455.67 p < 2.2e-16 Cramers_v=0.23
	q11_11 (A destination offering possibilities to visit amusement parks, zoos, water parks and so on)	**Unimportant (911, 42%)** Neutral (630, 29%) Important (613, 29%)	Unimportant (365, 28%) Neutral (400, 30%) **Important (545, 42%)**	Unimportant (100, 14%) **Neutral (505, 71%)** Important (103, 15%)	X^2=536.05 p < 2.2e-16 Cramers_v=0.25
	q11_12 (A destination with good restaurants, cafes, bars and so on)	Unimportant (386, 18%) Neutral (656, 30%) **Important (1112, 52%)**	Unimportant (305, 23%) Neutral (411, 31%) **Important (594, 46%)**	Unimportant (65, 9%) **Neutral (514, 73%)** Important (129, 18%)	X^2=457.7 p < 2.2e-16 Cramers_v=0.23
	q11_13 (A destination with good hygiene which minimises risks of spreading infectious diseases)	Unimportant (98, 5%) Neutral (217, 10%) **Important (1839, 85%)**	Unimportant (50, 3%) Neutral (124, 10%) **Important (1136, 87%)**	Unimportant (43, 6%) **Neutral (411, 58%)** Important (254, 36%)	X^2=953.31 p < 2.2e-16 Cramers_v=0.34
	q11_14 (Peace and quiet - a destination with less tourists)	Unimportant (156, 7%) Neutral (494, 23%) **Important (1504, 70%)**	Unimportant (81, 6%) Neutral (222, 17%) **Important (1007, 77%)**	Unimportant (43, 6%) **Neutral (464, 66%)** Important (201, 28%)	X^2=612.68 p < 2.2e-16 Cramers_v=0.27
	q11_15 (A destination close to a beach, harbour and coast line)	Unimportant (328, 15%) Neutral (606, 28%) **Important (1220, 57%)**	Unimportant (209, 16%) Neutral (489, 37%) **Important (612, 47%)**	Unimportant (59, 8%) **Neutral (512, 73%)** Important (137, 19%)	X^2=449.28 p < 2.2e-16 Cramers_v=0.23
	q11_16 (A destination with forests and nature)	Unimportant (231, 11%) Neutral (552, 26%) **Important (1371, 63%)**	Unimportant (101, 8%) Neutral (320, 24%) **Important (889, 68%)**	Unimportant (67, 10%) **Neutral (483, 68%)** Important (158, 22%)	X^2=532.9 p < 2.2e-16 Cramers_v=0.25
	q11_18 (Children friendly destination)	**Unimportant (972, 45%)** Neutral (575, 27%) Important (607, 28%)	Unimportant (367, 28%) Neutral (431, 33%) **Important (512, 39%)**	Unimportant (113, 16%) **Neutral (485, 68%)** Important (110, 16%)	X^2=509.41 p < 2.2e-16 Cramers_v=0.25
	q11_19 (A destination friendly for senior citizens)	**Unimportant (868, 40%)** Neutral (763, 35%) Important (523, 25%)	Unimportant (334, 26%) Neutral (446, 34%) **Important (530, 40%)**	Unimportant (101, 14%) **Neutral (518, 73%)** Important (89, 13%)	X^2=490.14 p < 2.2e-16 Cramers_v=0.24
	q11_20 (Clean destination (no trash, clean beach and air))	Unimportant (110, 5%) Neutral (302, 14%) **Important (1742, 81%)**	Unimportant (63, 5%) Neutral (210, 16%) **Important (1037, 79%)**	Unimportant (46, 7%) **Neutral (451, 64%)** Important (211, 29%)	X^2=822.45 p < 2.2e-16 Cramers_v=0.31

Table…continued

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics
Attitudes to public behaving responsible	q19_1 (It is important that individuals contribute to minimise the risk of spreading infectious diseases in public spaces)	Disagree (80, 4%) Neutral (88, 4%) **Agree (1986, 92%)**	Disagree (45, 3%) Neutral (63, 5%) **Agree (1202, 92%)**	Disagree (86, 12%) **Neutral (355, 50%)** Agree (267, 38%)	X^2=1310.6 $p < 2.2e-16$ Cramers_v=0.40
	q19_8 (I feel safe and comfortable if staffs in hotels, airlines, restaurants etc. wear a mask)	Disagree (259, 12%) Neutral (352, 16%) **Agree (1543, 72%)**	Disagree (101, 8%) Neutral (154, 12%) **Agree (1055, 80%)**	Disagree (93, 13%) **Neutral (422, 60%)** Agree (193, 27%)	X^2=770.75 $p < 2.2e-16$ Cramers_v=0.30
Attitudes to self-protective behaviors	q19_2 (I am keeping social distances in public spaces. If it is not possible, I will leave that place)	Disagree (206, 10%) Neutral (183, 9%) **Agree (1765, 82%)**	Disagree (69, 5%) Neutral (117, 9%) **Agree (1124, 86%)**	Disagree (108, 15%) **Neutral (392, 55%)** Agree (208, 30%)	X^2=1056.6, $p < 2.2e-16$ Cramers_v=0.36
	q19_4 (I carry and use disinfectant to clean my hand after touching items in shops to make me feel clean and safe)	Disagree (365, 17%) Neutral (231, 11%) **Agree (1558, 72%)**	Disagree (235, 18%) Neutral (150, 12%) **Agree (925, 70%)**	Disagree (158, 22%) **Neutral (371, 53%)** Agree (179, 25%)	X^2=766.5 $p < 2.2e-16$ Cramers_v=0.30
	q19_6 (I wear a mask to make me feel safe)	Disagree (682, 32%) Neutral (258, 12%) **Agree (1214, 56%)**	Disagree (117, 9%) Neutral (88, 7%) **Agree (1105, 84%)**	Disagree (146, 20%) **Neutral (359, 51%)** Agree (203, 29%)	X^2=1025.5 $p < 2.2e-16$ Cramers_v=0.35
Attitudes to responsible behaviors	q19_3 (I clean up a public space (e.g. Toilet) after I use it so that people who use it after me feel clean and safe)	Disagree (282, 13%) Neutral (338, 16%) **Agree (1534, 71%)**	Disagree (92, 7%) Neutral (130, 10%) **Agree (1088, 83%)**	Disagree (98, 14%) **Neutral (414, 59%)** Agree (196, 27%)	X^2=823.83 $p < 2.2e-16$ Cramers_v=0.31
	q19_5 (I carry and use disinfectant to clean my hands before touching items in shops so that other people who touch after me feel clean and safe)	Disagree (371, 17%) Neutral (236, 11%) **Agree (1547, 72%)**	Disagree (244, 19%) Neutral (157, 12%) **Agree (909, 69%)**	Disagree (151, 21%) **Neutral (373, 53%)** Agree (184, 26%)	X^2=740.7 $p < 2.2e-16$ Cramers_v=0.30
	q19_7 (I wear a mask to keep those around me safe and comfortable)	Disagree (616, 29%) Neutral (205, 10%) **Agree (1333, 61%)**	Disagree (103, 8%) Neutral (72, 6%) **Agree (1135, 86%)**	Disagree (129, 18%) **Neutral (349, 49%)** Agree (230, 33%)	X^2=1064.9 $p < 2.2e-16$ Cramers_v=0.36

Table…continued

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics
Attitudes to traveling within the country in Transitional Phase	q22_1A	Dangerous (239, 11%) Neutral (472, 22%) **Safe (1443, 67%)**	**Dangerous (850, 65%)** Neutral (265, 20%) Safe (195, 15%)	Dangerous (158, 22%) **Neutral (348, 49%)** Safe (202, 29%)	X^2=1529.8 $p < 2.2e-16$ Cramers_v=0.43
	q22_1B	Unenjoyable (191, 9%) Neutral (438, 20%) **Enjoyable (1525, 71%)**	**Unenjoyable (787, 60%)** Neutral (284, 22%) Enjoyable (239, 18%)	Unenjoyable (145, 20%) **Neutral (365, 52%)** Enjoyable (198, 28%)	X^2=1564.4 $p < 2.2e-16$ Cramers_v=0.43
	q22_1C	Onerous (321, 15%) Neutral (510, 24%) **Effortless (1323, 61%)**	**Onerous (913, 70%)** Neutral (230, 17%) Effortless (167, 13%)	Onerous (171, 24%) **Neutral (351, 50%)** Effortless (186, 26%)	X^2=1461.1 $p < 2.2e-16$ Cramers_v=0.42
	q22_1D	Harmful (187, 9%) Neutral (618, 29%) **Beneficial (1349, 62%)**	**Harmful (724, 55%)** Neutral (370, 28%) Beneficial (216, 17%)	Harmful (130, 18%) **Neutral (358, 51%)** Beneficial (220, 31%)	X^2=1244.6 $p < 2.2e-16$ Cramers_v=0.39
Intention to Travel within country for Pleasure	Q22_2_1 (I intend to travel for pleasure within Country in the transitional phase)	Disagree (224, 11%) Neutral (343, 16%) **Agree (1567, 73%)**	**Disagree (671, 51%)** Neutral (277, 21%) Agree (362, 28%)	Disagree (164, 23%) **Neutral (444, 63%)** Agree (100, 14%)	X^2=1525 $p < 2.2e-16$ Cramers_v=0.43
Assumptions of Known People Towards Respondents Travel during Transitional Phase	Q22_2_2 (Most people who are important to me think that my traveling within Country in the transitional phase will be thoughtful and respectful of their and my safety)	Disagree (170, 8%) Neutral (398, 18%) **Agree (1586, 74%)**	**Disagree (478, 37%)** Neutral (310, 24%) **Agree (522, 39%)**	Disagree (122, 17%) **Neutral (479, 68%)** Agree (107, 15%)	X^2=1236.7 $p < 2.2e-16$ Cramers_v=0.38
	Q22_2_3 (Most people whose opinion I value would approve my traveling within Country in the transitional phase)	Disagree (160, 7%) Neutral (412, 19%) **Agree (1582, 74%)**	**Disagree (545, 42%)** Neutral (404, 31%) Agree (361, 27%)	Disagree (126, 18%) **Neutral (471, 67%)** Agree (111, 15%)	X^2=1411 $p < 2.2e-16$ Cramers_v=0.41
	Q22_2_4 (Most people I respect and admire will be traveling within Country in the transitional phase)	Disagree (210, 10%) Neutral (608, 28%) **Agree (1336, 62%)**	**Disagree (656, 50%)** Neutral (322, 25%) Agree (332, 25%)	Disagree (156, 22%) **Neutral (480, 68%)** Agree (72, 10%)	X^2=1303.9 $p < 2.2e-16$ Cramers_v=0.39
Resources (Media, Money & Time) Role during transitional phase	Q22_2_5 (Mass and social media will encourage traveling activities within Country in the transitional phase)	Disagree (240, 11%) Neutral (419, 20%) **Agree (1495, 69%)**	Disagree (355, 27%) Neutral (409, 31%) **Agree (546, 42%)**	Disagree (131, 19%) **Neutral (475, 67%)** Agree (102, 14%)	X^2=855.65 $p < 2.2e-16$ Cramers_v=0.32
	Q22_2_6 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available financial resources to do it)	Disagree (294, 13%) Neutral (357, 17%) **Agree (1503, 68%)**	Disagree (372, 28%) Neutral (326, 25%) **Agree (612, 47%)**	Disagree (142, 20%) **Neutral (443, 63%)** Agree (123, 17%)	X^2=802.39 $p < 2.2e-16$ Cramers_v=0.31
	Q22_2_7 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available time to do it)	Disagree (196, 9%) Neutral (344, 16%) **Agree (1614, 75%)**	Disagree (379, 29%) Neutral (310, 24%) **Agree (621, 47%)**	Disagree (147, 21%) **Neutral (436, 62%)** Agree (125, 17%)	X^2=957.11 $p < 2.2e-16$ Cramers_v=0.34

Table……continued

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics
COVID-19 perception during transitional phase	Q22_3_1 (If I travel within Country in the transitional phase, I will contribute to the spread of Covid-19 at the destination)	**Disagree (1385, 64%)** Neutral (438, 20%) Agree (331, 16%)	Disagree (185, 14%) Neutral (249, 19%) **Agree (876, 67%)**	Disagree (162, 23%) **Neutral (439, 62%)** Agree (107, 15%)	X^2=1712 p < 2.2e-16 Cramers_v=0.45
	Q22_3_2 (If I travel within Country in the transitional phase, I will get infected with Covid-19)	**Disagree (1294, 60%)** Neutral (552, 26%) Agree (308, 14%)	Disagree (149, 11%) Neutral (308, 24%) **Agree (853, 65%)**	Disagree (155, 22%) **Neutral (444, 63%)** Agree (109, 15%)	X^2=1589.9 p < 2.2e-16 Cramers_v=0.44
	Q22_3_3 (If I travel within Country in the transitional phase, I will infect others (relatives, friends, colleagues etc.) with Covid-19)	**Disagree (1451, 67%)** Neutral (421, 20%) Agree (282, 13%)	Disagree (213, 16%) Neutral (319, 24%) **Agree (778, 60%)**	Disagree (159, 23%) **Neutral (442, 62%)** Agree (107, 15%)	X^2=1606.1 p < 2.2e-16 Cramers_v=0.44

Source: Own elaboration

Out of the respondents in this sub-group, 81% reported that they prefer to plan their travel arrangements, while the remaining percentage prefers to have a third party handle it. Additionally, 76% of the respondents stated that they prefer to travel alone or with their significant other as their top choice of travel companionship, followed by their closest family or friends, which was the first preference for 18% of the respondents.

When asked to compare the importance of factors for their next travel destination to the previous year before the Covid-19 pandemic, the majority of respondents in this sub-group prioritized a clean destination (81%), a destination with good hygiene to minimize infectious disease risks (85%), a safe travel destination (85%), and a peaceful and quiet destination with fewer tourists (70%). For a subsection of this group, a destination close to a beach, harbor, or coastline was important to 57% of respondents, while 63% preferred a destination with forests and nature, and approximately 52% gave importance to destinations that offer good restaurants, cafes, bars, and similar amenities. Child and senior-friendly places and previously visited destinations were not considered important by this group.

Within this specific sub-group, almost 92% of the participants acknowledged the importance of taking measures to minimize the transmission of infectious diseases in public areas. Moreover, about 72% of the respondents reported feeling secure and at ease when the staff at hotels, airlines, eateries, and comparable establishments wore masks. This particular cluster of individuals exhibits a favourable disposition towards maintaining personal hygiene. Roughly

82% of respondents mentioned that they practice social distancing in public places and leave if distancing is not possible. Around 72% of respondents mentioned that they use disinfectants to clean their hands before touching items in public domains or shops to keep themselves safe. However, there is room for improvement regarding mask-wearing, as only 56% of respondents mentioned that they wear a mask. Around 71% of respondents in this sub-group displayed a positive disposition towards the public and public spaces by cleaning public areas such as toilets to ensure the safety of future users. Similarly, a similar percentage of 72% of respondents stated that they use disinfectants to clean their hands before touching anything in public spaces, with the aim of ensuring the safety of future users. Additionally, around 61% of respondents mentioned that they wear masks in public areas to make those around them feel safe and comfortable.

During the transitional phase, which involves society reopening almost entirely with virus prevention measures in place, a significant majority of the respondents (73% or 1567 individuals) agreed to travel domestically for leisure within the next six months. Although some infection cases and a small number of deaths are still being reported, there is currently no cure or vaccine available, and the possibility of a second wave of infections is acknowledged. The individuals within this sub-group seem to have a positive outlook regarding travelling during the transition phase. Around 67% of respondents believe that travelling during this period will be safe, while 71% believe it will be enjoyable. Additionally, 62% of respondents view it as beneficial, and 61% consider it to be effortless. However, approximately 74% of the respondents feel that their decision to travel during this phase is a thoughtful and respectful choice for their own and others' safety, according to people known to them. Furthermore, about 74% of the respondents believe that the people whose opinions they value the most would approve of their domestic travel plans during the transitional phase, while around 62% of this sub-group believe that the people they respect and admire the most will also be travelling domestically during this period. Around 69% of the participants expressed their belief that mass and social media will play a crucial role in promoting domestic travel during the transitional phase when asked about the impact of resources such as time, money, and social media on their leisure travel decisions. Additionally, 68% of the respondents expressed confidence in their financial capacity to travel during the transitional phase, while 75% felt confident that they would have adequate time to do so. In this sub-group, approximately 64% of the respondents hold the perception that their travel during the transitional phase will not contribute to the spread of Covid-19 at their destination. Also, a similar percentage (60%) of respondents in the sub-group believed that they will not be infected by Covid-19 while travelling, and around 67% believe that inadvertently they will not infect their loved ones, friends, colleagues, or other individuals.

Cluster-2: “Risk-Averse”

This sub-group accounts for 31% of the sample and is the second largest. The majority of individuals in this group are Japanese nationals, representing 49%, followed by individuals from China (27%) and Italy (15%). In contrast, only 9% of individuals in this sub-group are from Denmark. Among this sub-group, individuals aged 55 years and above constitute the largest group at 33%, followed by respondents aged 25 to 34 years, who make up 24% of this sub-group. Females dominate this sub-group, representing 65%, while males constitute only 35%.

According to the travel patterns of this particular subgroup, it appears that around 28% of the respondents have travelled internationally for business reasons at least one to three times in the last two years. In contrast, a significant 87% of respondents from this subgroup have travelled internationally for leisure purposes at least one to three times during the same period, with a similar 89% of respondents travelling domestically for leisure.

Approximately 75% of respondents in this group indicated a preference for self-planning their travel arrangements instead of relying on a third party. While 52% of the respondents prefer to travel alone or with their significant other, approximately 43% of the respondents expressed a preference for travelling with their closest family members or friends.

When asked about the importance of different factors when choosing a holiday destination this year compared to the last year, 88% of the respondents in this sub-group considered a safe destination to be important, with a similar percentage (87%) indicating the importance of a destination with good hygiene to minimize the risk of spreading infectious diseases. Additionally, 81% of the respondents gave importance to a clean destination, while 70% mentioned peace, and the importance of a destination with fewer tourists. This sub-group also displayed a preference for destinations that are child-friendly (39%) and senior-friendly (40%), indicating a desire to travel with family or friends who have children or with elderly parents. Approximately 68% of the sub-group was interested in destinations with forests and nature, while 47% gave importance to a destination close to a beach, harbour, or coastline. Similarly, around 47% of the respondents expressed an interest in destinations that offer possibilities to visit museums, exhibitions, or historical attractions, while 42% were interested in amusement parks, zoos, water parks, and so on. Furthermore, this group showed a preference for destinations with good shopping opportunities (43%) and good restaurants, cafes, bars, and the like (46%). However, this group did not express a preference for previously travelled destinations, with 42% of the respondents indicating a neutral response.

Regarding self-protective measures against Covid-19, a significant proportion (86%) of individuals within this sub-group reported that they maintain social distance in public spaces, and if it is not feasible, they would leave the area. Similarly, approximately 80% of this sub-group stated that they wear masks to protect themselves. However, in contrast, only 70% of them mentioned that they carry and use disinfectant to clean their hands after touching items in shops, in order to keep themselves clean and safe.

An overwhelming majority (92%) of individuals within this sub-group hold the belief that each person must contribute to reducing the spread of infectious diseases in public spaces. Furthermore, 80% of respondents in this group reported that they would feel secure and at ease if staff in hotels, airlines, restaurants, and other establishments wear masks.

Regarding the behaviour of individual respondents, around 83% of individuals within this sub-group reported that they clean up a public space, such as a toilet, after using it, to ensure that others who use it after them feel clean and safe. Approximately 86% of respondents mentioned that they wear a mask to keep those around them secure and comfortable. However, only 69% of respondents reported that they carry and use disinfectant to clean their hands before touching items in shops, with the intention of making it safe and clean for others who touch those items later.

The responses of this sub-group reveal their cautious approach towards domestic leisure travel during the next six months, referred to as the transition period. The majority (51%) indicated that they do not want to travel during this period, as they believe that the travel will be dangerous (65%), unenjoyable (60%), difficult (70%), and harmful (55%). Additionally, the group has mixed responses towards how their know people will consider the travel during this phase 39% believe that they know people with consider it as a thoughtful and respectful however almost similar 37% disagree with that and they feel the travel will be looked upon as during the transition period as unthoughtful and disrespectful towards their and knowns people safety. Also 42% believe their travel decisions will not be approved by those whose opinions matter to them. Furthermore, 50% of respondents feel that people they admire, and respect will not travel during this period. Regarding the influence of resources, around 42% of the sub-group was neutral towards the role of mass and social media. However, an equal percentage (47%) indicated that they will not have enough financial resources and time to travel within the country during the transition period. Moreover, the majority (67%) believe that travelling within the country during the transition period will contribute to the spread of Covid-19 at their destination, with 65% believing they will get infected with the virus due to travel. Additionally, 60% of respondents believe that if they get infected, they will in turn infect their relatives, friends, colleagues, and others with Covid-19. This cautious behaviour illustrates the risk-averse nature of this sub-group toward the prevention of the spread of Covid-19.

Cluster-3: "Risk-Neutral"

This sub-group constitutes only 17% or 708 individuals of the total sample size. Among them, Japanese respondents comprise the majority at approximately 45%, while Italy and China are equally represented at around 17% each with Denmark representing 21% of the respondents. This sub-group is primarily represented by younger people with 31% of the respondents falling in the age bracket of 25-34 years. Male respondents constitute the majority at 57%, while women respondents comprise 43% of the sub-group.

Upon examining the travel responses of this particular subset over the past two years, it becomes apparent that, like two other subsets, the majority (67%) did not venture abroad for business purposes. Nonetheless, roughly 9% of this group travelled overseas for business purposes between four and six times. Concerning leisurely overseas travel, approximately 86% of respondents journeyed abroad between one to three times, while 81% of them opted for domestic travel. In summary, these respondents had a favourable leisure travel experience over the past two years.

Similar to the other two sub-groups, 73% of the respondents in this sub-group also prefer to plan their own travel arrangements. While 55% of the respondents prefer to travel alone or with their significant other, a significant portion of 37% like to travel with their closest family or friends as their second preference for travel companionship.

This sub-group demonstrated a neutral stance when asked about factors important to their travel destination when planning their vacation for the current year compared to the previous year. Around 58% of respondents remained neutral towards whether the destination is safe or hygienic, while 64% of respondents showed a neutral stance toward a clean destination.

Moreover, 66% of respondents displayed a neutral attitude towards a peaceful destination. Similarly, a majority of respondents had a neutral stance towards offerings such as good shopping facilities (72%) and good restaurants, food, or cafes (73%) in the travel destination. Furthermore, the sub-group also remained neutral towards children or family-friendly places, with an average of 70% of respondents showing a neutral stance towards them. Similarly, approximately 70% of respondents demonstrated a neutral stance towards destinations offering possibilities to visit museums, exhibitions, historical attractions, amusement parks, zoos, water parks, destinations close to a beach, harbour, and coastline, or destinations with forests and nature.

Most of the individuals in this particular subgroup held a neutral position regarding protecting themselves. Approximately 55% of the participants responded neutrally when asked if they practice social distancing in public places and leave if it's not feasible. Around 53% of the respondents had a neutral response regarding the use of disinfectants to clean their hands after touching things in shops to maintain hygiene. Similarly, approximately 51% of the participants had a neutral attitude toward wearing masks to ensure their safety.

This subgroup displayed a neutral attitude toward protecting individuals in public spaces. About 59% of the participants held a neutral stance on cleaning public facilities, such as toilets, after use to ensure cleanliness and safety for others. Similarly, 53% of the respondents maintained a neutral stance on using disinfectants to clean their hands before touching items in shops or other public areas. Additionally, 49% of the participants had a neutral attitude towards wearing masks in public to make others feel safe and secure. The subgroup also held a similarly neutral stance when asked about public responsibility. 50% of the participants responded neutrally to the question of whether individuals should contribute to minimizing the risk of spreading infectious diseases in public spaces, and 60% chose to remain neutral towards the question of feeling safe and comfortable if staff in hotels, airlines, and restaurants wore masks.

The sub-group in question has a neutral stance on the idea of travelling within their own country for leisure purposes during the transitional phase with 63% of the respondents choosing the neutral option for this straight question. On average around 50% of the respondents in this sub-group have chosen this neutral option when they were asked that travel during this transitional phase will be with dangerous or safe, unenjoyable or enjoyable, onerous or effortless, harmful or beneficial. Around 68% of the respondents of this sub-group remained neutral on the question of whether their loved ones will consider the travel as respectful and thoughtful or approve such decision of travel or their loved ones themselves will travel during the next six months. Similarly, a comparable percentage of respondents have expressed a neutral stance on the resources required for travelling during the transitional phase, including time, money, and media coverage. Finally, when asked about their Covid-19 perception and how it relates to travel within the country during the transitional phase, the 62% respondents on average once again showed a neutral approach towards questions pertaining to the potential spread of Covid-19, personal infection, and infecting others.

4.7 Deployment

Deployment is the final stage of CRISP-DM, where the outcomes are presented to the client and transferred to the production environment after approval. The model then goes through a testing phase to evaluate its performance in real-time. As this thesis is focused on an academic perspective, the deployment process is beyond the scope of this study.

5. Results

The evaluation or the results of the modelling are discussed in detail in the *evaluation* section of the CRISP-DM methodology, however, here I will discuss the results briefly using only the empirical information.

5.1 Agreement between LCA and K-modes

In this thesis, I was trying to find an alternative machine learning data science method that can be used for clustering the survey data that can perform comparably to the statistical method LCA. The two models K-Modes algorithm and LCA were run using the same set of categorical variables and their agreement was compared using Cohen's kappa score.

Cohen Kappa and Confidence Boundaries
LCA and K-Modes

	Lower	Estimate	Upper
Unweighted Kappa	0.63	**0.65**	0.67

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

The Cohen's kappa score of 0.65 was achieved when comparing the three clusters with three groups from LCA and using the Cohen's kappa score table it shows that the two methods have a substantial agreement removing any apprehension that the agreement achieved is not just by chance. Based on the result I have used the K-Modes cluster to the further analysis of clusters.

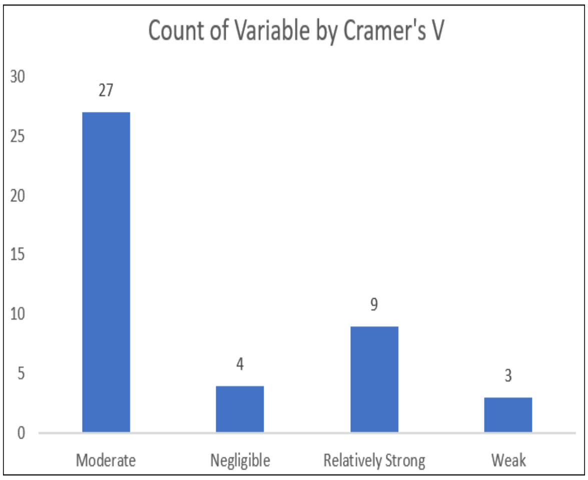
5.2 Homogeneous Clusters

5.2.1 Test of Independence (Chi-Square Test)

The three clusters were identified from the K-modes algorithm. As all the variables are categorical in nature, hence, to identify the relationship between the tourist's cluster variable and the variables used in modelling, I have performed the chi-square test of independence with alpha-value=0.05 considering the null hypothesis that all the variables are independent of the tourist's cluster variable. All variables, except for "overseas travel for business in last 2 years," had a p-value lower than the threshold value of 0.05, indicating that the null hypothesis is rejected, and the alternative hypothesis is accepted, which states that all variables are associated with the tourist's cluster variable.

5.2.2 Test of Strength (Cramer's V Test)

This chi-square test with a significance level of 0.05 do give that the variables are related, however, it doesn't tell how strong or weak the relationship is between the variables. So, to identify the strength of the variables with the cluster variables the Cramer's V test is performed (Table 18 of the Evaluation section). The total of nine variables had a relatively strong relationship with the cluster variable.



Estimated values	Interpretation of association
0.00–0.10	Negligible
0.10–0.20	Weak
0.20–0.40	Moderate
0.40–0.60	Relatively strong
0.60–0.80	Strong
0.80–1.00	Very strong

Out of the nine variables analysed, eight of them were related to the transition period, including a variable regarding the intention to travel for pleasure within the country during the transitional phase. This variable exhibited a moderate relationship with a Cramer's V strength value of 0.43. The three variables related to the perception of COVID-19 health concerns regarding travel during the transition period displayed a relatively strong relationship with the cluster variable, with Cramer's V values of 0.44 and 0.45, which is consistent with the results of the chi-square test. Variables related to resources, such as mass media, time, and money, displayed a moderate relationship with the cluster variable, with Cramer's V values ranging between 0.31 and 0.34. Similarly, variables related to assumptions of known people towards respondents' travel during the transitional phase showcased a moderate relationship with Cramer's V values of 0.38 and 0.39. Regarding the next travel destination, variables related to safe, hygienic, and clean destinations displayed a moderate relationship with Cramer's V values of 0.34, 0.34, and 0.31, respectively. Other destination variables also displayed a moderate relationship with Cramer's V values ranging between 0.23 and 0.27, except for the variable

Category	Variable	Cramer's V
Demographics	Country	Moderate
	Profile_Age	Negligible
	Gender	Moderate
Travel Experience	q1_1 (overseas for business in last 2 years)	Negligible
	q1_2 (overseas for leisure in last 2 years)	Weak
	q1_3 (domestic for business in last 2 years)	Negligible
Preferred Travel Arrangement	q3_1 (Preferred Travel Arrangement)	Negligible
Preferred Travel Companionship	q4_1 (preferred travel companionship)	Weak

Category	Variable	Cramer's V
How Important the factors for next travel	q11_5 (A destination where I have previous travel experiences)	Weak
	q11_6 (A safe travel destination)	Moderate
	q11_7 (A destination with good shopping possibilities)	Moderate
	q11_10 (A destination offering possibilities to visit museums, exhibitions, historical attractions)	Moderate
	q11_11 (A destination offering possibilities to visit amusement parks, zoos, water parks and so on)	Moderate
	q11_12 (A destination with good restaurants, cafes, bars and so on)	Moderate
	q11_13 (A destination with good hygiene which minimises risks of spreading infectious diseases)	Moderate
	q11_14 (Peace and quiet - a destination with less tourists)	Moderate
	q11_15 (A destination close to a beach, harbour and coast line)	Moderate
	q11_16 (A destination with forests and nature)	Moderate
	q11_18 (Children friendly destination)	Moderate
	q11_19 (A destination friendly for senior citizens)	Moderate
	q11_20 (Clean destination (no trash, clean beach and air))	Moderate

Category	Variable	Cramer's V
Attitudes to public behaving responsible	q19_1 (It is important that individuals contribute to minimise the risk of spreading infectious diseases in public spaces)	Relatively Strong
	q19_8 (I feel safe and comfortable if staffs in hotels, airlines, restaurants etc. wear a mask)	Moderate
Attitudes to self-protective behaviors	q19_2 (I am keeping social distances in public spaces. If it is not possible, I will leave that place)	Moderate
	q19_4 (I carry and use disinfectant to clean my hand after touching items in shops to make me feel clean and safe)	Moderate
	q19_6 (I wear a mask to make me feel safe)	Moderate
Attitudes to responsible behaviors	q19_3 (I clean up a public space (e.g. Toilet) after I use it so that people who use it after me feel clean and safe)	Moderate
	q19_5 (I carry and use disinfectant to clean my hands before touching items in shops so that other people who touch after me feel clean and safe)	Moderate
	q19_7 (I wear a mask to keep those around me safe and comfortable)	Moderate

regarding the last destination travelled, which had a weak relationship with the cluster variable.

Category	Variable	Cramer's V
Attitudes to traveling within the country in Transitional Phase	q22_1A	Relatively Strong
	q22_1B	Relatively Strong
	q22_1C	Relatively Strong
	q22_1D	Moderate
Intention to Travel within country for Pleasure	Q22_2_1 (I intend to travel for pleasure within Country in the transitional phase)	Relatively Strong

Category	Variable	Cramer's V
Resources (Media, Money & Time) Role during transitional phase	Q22_2_5 (Mass and social media will encourage traveling activities within Country in the transitional phase)	Moderate
	Q22_2_6 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available financial resources to do it)	Moderate
	Q22_2_7 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available time to do it)	Moderate

Category	Variable	Cramer's V
Assumptions of Known People Towards Respondents Travel during Transitional Phase	Q22_2_2 (Most people who are important to me think that my traveling within Country in the transitional phase will be thoughtful and respectful of their and my safety)	Moderate
	Q22_2_3 (Most people whose opinion I value would approve my traveling within Country in the transitional phase)	Relatively Strong
	Q22_2_4 (Most people I respect and admire will be traveling within Country in the transitional phase)	Moderate

Category	Variable	Cramer's V
19 perception during transitio	Q22_3_1 (If I travel within Country in the transitional phase, I will contribute to the spread of Covid-19 at the destination)	Relatively Strong
	Q22_3_2 (If I travel within Country in the transitional phase, I will get infected with Covid-19)	Relatively Strong
	Q22_3_3 (If I travel within Country in the transitional phase, I will infect others (relatives, friends, colleagues etc.) with Covid-19)	Relatively Strong

Similarly, the variables related to the attitude to public behaving properly, attitude to self-protective behaviour, and attitude to responsible behaviour towards covid-19 also had a moderate relationship with the cluster variable and the Cramer’s V value range between 0.30 to 0.36. Among the demographics, the country (0.35) and gender (0.20) variables showcased a moderate relationship with the clusters however the profile age showcased a negligible relationship with Cramer’s V value of 0.09.

5.2.3 Post-Hoc Test (using Bonferroni correction)

Once the relationship between the variables and their respective strength with cluster variables has been determined using the chi-square test and Cramer's V test, the focus shifts to identifying the specific pairwise groups that are significantly associated with the variable's strength in each level of the cluster variable. By conducting post hoc comparisons between each level of variables used in the modelling with each cluster level of cluster variable in a way that avoids excessive type-1 error, in other words, avoiding rejecting the null hypothesis when the null hypothesis is true, I will be much better to able to appropriately describe which levels of each variable is different from the others. If I reject the null hypothesis, I need to perform comparisons for each pair of cluster levels across the level of each variable. For instance, suppose we need to compare the variable Q22_2_1 (I intend to travel for pleasure within the Country in the transitional phase), which has three levels (disagree, neutral, and agree), with a cluster variable that has three levels representing three clusters. In such a case, I need to perform nine pairwise comparisons. To avoid type-1 errors in the chi-square test, I will apply the Bonferroni adjustment post-hoc approach. This approach aims to control the familywise error rate, also called maximum overall type-1 error rates, so that it can determined which pairs of Q22_2_1 dependence rate differ from each other. Precisely, I will perform all nine paired comparisons, but instead of using a significance level of 0.05, I will adjust the significance to reduce the chance of rejecting the null hypothesis. To calculate the adjusted significance value, I will divide the original significance value of 0.05 by the number of comparisons I plan to make, which is nine in this case. Therefore, I will only reject the null hypothesis if the adjusted significance value is 0.0056 or lower (as 0.05 divided by 9 equals 0.0056). However, before comparison, I need to perform the nine chi-square test comparisons and get the standardized residuals of the pairs by using the command below.

```
chisq.test(table(KMD3table$`M3$cluster`, KMD3table$Q22_2_1))$stdres
adjpval=0.05/9
qnorm(adjpval/2)
```

This gave the contingency table as mentioned below

```
> chisq.test(table(KMD3table$`M3$cluster`, KMD3table$Q22_2_1))$stdres

            1          2          3
  1 -22.151189 -14.665941  32.196387
  2  25.306949  -4.369430 -18.360639
  3  -1.799927  24.927291 -20.161838
```

and now to see if any of these are significant enough, we need to perform an inverse normal distribution standard normal (qnorm) as mentioned in the code above which will give the critical value of -2.77. So, any value which is either <= -2.77 or >= 2.77 would be significant.

The table that shows the results of the post-hoc test of each variable is mentioned in the appendix section, however, here are the most relevant factors.

The significant factors associated with the formation of cluster-1, referred to as "Risk-Takers," are primarily older males who have previously travelled domestically for leisure purposes several times (4-6 times or 7-12 times). They prefer to plan their own travel arrangements and enjoy travelling alone, with a significant other, or with close family and friends. Safe, hygienic, and clean destinations are important considerations for this sub-group when choosing their next travel destination, as are places with good restaurants, bars, and other amenities. They also prefer open, nature-specific places like beaches and forests, as well as museums and amusement parks with large spaces. Risk-takers are optimistic about travel during the transitional phase and believe that the people who matter to them will also travel during the same timeframe. They are confident that they will neither contribute to nor be infected by COVID-19 during their travels, and they have the necessary time and financial resources to support their travel plans.

The group named "Risk-Averse" is primarily composed of females who have previously travelled for leisure purposes domestically 1-3 times and show a positive significance level for this variable. They prefer to have their travel plans arranged by a third party and would like to travel with close family or friends. The group places importance on previously visited destinations and prefers destinations that are child and senior-friendly. The next travel destination should be safe, hygienic, and clean, offering shopping opportunities, and they are also open to places like museums, theme parks, zoos, or places that offer natural elements such as forests. Unlike the "Risk-Takers" group, this group does not consider places like beaches, harbours, or coastlines as important nor places that offer good restaurants, bars, or cafes. The tourists of this group do not prefer to travel during the transitional phase as they feel that the travel during this phase will be unsafe for themselves and their loved ones, also they feel that if they travel during this transitional phase then there is a good chance that they will catch covid-19 infection and in turn, also infect their family members and people whom they come in close contact. Additionally, they believe that they will not have sufficient money or time for travel.

The third and the smallest group, termed as 'Risk-Neutral', have more representation from male and consists of individuals aged between 25 and 34 years. This group is not particularly inclined towards travel, as indicated by the significant negative level for the option 'not at all' when asked about their past travel experience for domestic leisure purposes. However, they do have a preference for travelling in larger groups with arrangements made by a third party. Unlike the other two groups, they do not have any particular preferences for factors related to their next travel destination, compared to their previous ones. This group takes a neutral stance towards travelling during the transitional phase and does not consider any specific factors related to health or resources (money, time) for travel during the next six months.

6. Discussion

6.1 Research Question Answers

The purpose of this study was twofold. Firstly, it aimed to determine the level of agreement between Latent Class Analysis (LCA) and K-Modes Analysis in identifying distinct sub-groups of tourists based on their perceived risk associated with Covid-19. Secondly, the study aimed to identify the factors that corresponded to each sub-group. The results indicated a substantial agreement between the sub-groups identified using LCA and K-Modes Analysis, as demonstrated by a Cohen's kappa value of 0.65. This finding supported the assumption that both methods would yield similar results, suggesting that either approach could be used for such studies. However, it was observed that the optimal number of clusters differed between the two methods, with K-Modes Analysis identifying three clusters, while LCA was unable to determine the optimal number of classes based on the information criteria.

The study further identified three distinct sub-groups of tourists based on their behaviours, past travel experience, risk perception associated with Covid-19, and finally the intention to travel during the transitional phase. The largest group consisted of people who were willing to take risks and intended to travel for leisure within their home country in the next six months. These tourists would like to plan their own travel and like to either travel alone or with their significant other. They look forward to a place to travel that can offer safety, sanitation, hygiene, cleanliness, and peacefulness, like beaches, harbours, forests, and museums, and have good restaurants, café, and bars. They believed that their travel decision during the transitional period would be considered thoughtful and respectful and that they would take proper social distancing measures, wear masks in public, and maintain hygiene to prevent the spread of Covid-19. They also believed that their travel would not contribute to the spread of the disease, nor would they get infected or infect others. This sub-group considered mass and social media to have a positive impact on their decision to travel and believed they had sufficient financial resources and time for such travel.

The second largest group obtained from the k-mode model consisted of tourists who like to get their travel plans arranged by third parties like travel agents, tour and travel agencies and they would like to travel with close family and friends. For them, it is important that travel destinations offer safety, cleanliness, and hygiene. It will be safe to that this group is a family-oriented group as they prefer places that are children-friendly and elderly-friendly like amusement parks, zoos, and water parks. They strongly believe that it's both individual's and the public duty to keep common places clean and hygienic so that it's safe for everyone to use. These set of people are particular about social distancing and the usage of disinfectants and advocates the usage of mask in public places. However, tourists in this group were termed risk-averse people as they don't want to during the transitional period. They believe that the travel during the transitional phase would be dangerous, unenjoyable, onerous, and harmful to themselves and their loved ones. They believe that their decision to travel during this phase would be considered unthoughtful and disrespectful to their safety as well as to the safety of their know ones. These tourists also assume that their loved ones would also not be travelling during this time. They further believe that they did not have sufficient money or time for such travel and perceived the travel as a threat not only to their own health, but also to the health of their families, friends, and neighbours.

The smallest sub-group consisted of people who were risk-neutral and remained neutral towards travelling during the transitional phase. They have a neutral stance towards the health-related risks, and resources like time and money associated with travel during the transitional period. However, the optimistic part about the respondents of this group is that they could move toward either of the other two sub-groups based on their overall feeling in the future.

6.2 Limitations

Firstly, this thesis has limited data of only four countries from two continents, Europe, and Asia, and hence using them to form the generalized view may not provide good results. So, while looking at the results of this thesis, it should be looked at as getting a framework for further research on similar lines with data collected from different countries. Also, the thesis is only concentrated on the recovery of domestic tourism for leisure purposes hence the results may not be applicable to other kinds of travel like business travel or international travel for leisure purposes.

Secondly, by manually merging the Likert scale data to represent primarily three feelings negative, positive, and neutral will lose some information like the extreme responses which might be inspired by the socio-cultural background of the respondents. It won't provide the further sub-groups of those three feelings which can help in targeting the very specific sub-group for target campaigns and promotions by travel agencies and or tourism departments.

Finally, as the results obtained by K-modes are dependent on the initial centroids chosen for modelling, the results obtained from the selected model may not be the best ones. Hence, future research may consider alternative machine-learning techniques that can be applied to categorical data.

7. Conclusion

In conclusion, this academic thesis aimed to identify the factors that corresponded to each sub-group of tourists in the context of the COVID-19 pandemic. The three distinct sub-groups of tourists were identified: the “Risk-Takers”, the “Risk-Averse”, and the “Risk-Neutral” as a result of the thesis.

The “Risk-Takers” were the largest sub-group identified from our k-mode model and the respondents of this group were willing to take risks and travel for leisure within their home country in the next six months. They preferred to plan their own travel and either travel alone or with their significant other. They looked for destinations that offered safety, sanitation, hygiene, and peacefulness, such as beaches, harbours, forests, and museums. This sub-group considered mass and social media to have a positive impact on their decision to travel and believed they had sufficient financial resources and time for such travel. For this sub-group, it is crucial to promote destinations that offer safety, sanitation, hygiene, cleanliness, and peacefulness. Tourist hotspots such as beaches, harbours, forests, and museums should be marketed as safe and secure places to travel. Promoting responsible tourism practices, including social distancing, the use of masks in public, and maintaining hygiene, can help assure this group that their travel is safe and does not contribute to the spread of Covid-19.

The second largest sub-group was the Risk-Averse, who preferred to get their travel plans arranged by third parties and travel with close family and friends. They looked for destinations that offered safety, cleanliness, and hygiene and were particularly interested in places that were child and elderly friendly like amusement parks, zoos, and water parks. This sub-group strongly believed in social distancing and the use of disinfectants and advocated the use of masks in public places. Tourists in this group were not in favour of travelling during the transitional period due to their risk-averse nature they feel that the travel during this phase is not safe for themselves and their loves ones. It is necessary to promote and emphasize the importance of keeping common places clean and hygienic, both as an individual and public duty. Campaigns and policies focusing on social distancing and the use of disinfectants and masks in public can help reassure this group that their travel is safe and responsible.

The smallest sub-group was the Risk-Neutral, who remained neutral toward travelling during the transitional period and did not have a strong opinion on the risks associated with COVID-19 during this period. These individuals can move towards either of the other two sub-groups based on their overall feeling in the future. It is important to keep them engaged and informed through promotional campaigns that highlight safe travel practices and the benefits of tourism for the local economy.

These findings have significant implications for the formulation of policies, campaigns, and business strategies to promote tourism and revive the industry in the context of COVID-19. The Risk-Takers could be targeted with promotional campaigns highlighting destinations that offer safety, sanitation, and hygiene, while the Risk-Averse could be targeted with campaigns that emphasize the importance of maintaining cleanliness and hygiene in common places.

In addition, the travel industry could prioritize offering flexible booking policies and travel insurance to address concerns about the uncertainty of the pandemic. Educational campaigns could also be launched to increase public awareness about the importance of social distancing, mask-wearing, and personal hygiene while travelling.

Overall, this study provides insights into tourists' behaviour and perceptions in situations like the Covid-19 pandemic. Specifically, it provided insights about the group who are willing to take risks and travel during the transitional phase, this will help the travel industry and policymakers to develop effective strategies in attracting tourists towards the travel destination by doing the right kind of campaigns and promotions. This study also adds to the previous studies conducted by scholars in understanding the tourist's risk perceptions and travel intentions of different sub-groups of tourists. The findings of this thesis can be used as a foundation for future research in forming a generalised view of tourists in the field of tourism and hospitality during the pandemic-like situation.

References

UNWTO (Ed.). (2021, January 28). 2020: WORST YEAR IN TOURISM HISTORY WITH 1 BILLION FEWER INTERNATIONAL ARRIVALS. UNWTO. https://www.unwto.org/news/2020-worst-year-in-tourism-history-with-1-billion-fewer-international-arrivals

Singh, S., McNab, C., Olson, R. M., Bristol, N., Nolan, C., Bergstrøm, E., Bartos, M., Mabuchi, S., Panjabi, R., Karan, A., Abdalla, S. M., Bonk, M., Jamieson, M., Werner, G. K., Nordström, A., Legido-Quigley, H., & Phelan, A. (2021). How an outbreak became a pandemic: a chronological analysis of crucial junctures and international obligations in the early months of the COVID-19 pandemic. Lancet (London, England), 398(10316), 2109–2124. https://doi.org/10.1016/S0140-6736(21)01897-3

Banda, J., Dube, A. N., Brumfield, S., Amoah, A. S., Reniers, G., Crampin, A. C., & Helleringer, S. (2021). Knowledge, risk perceptions, and behaviors related to the COVID-19 pandemic in Malawi. *Demographic Research*, *44*, 459–480. https://www.jstor.org/stable/27032922

Jaume Rosselló, Maria Santana-Gallego, Waqas Awan, Infectious disease risk and international tourism demand, Health Policy and Planning, Volume 32, Issue 4, May 2017, Pages 538–548, https://doi.org/10.1093/heapol/czw177

Roehl, W. S., & Fesenmaier, D. R. (1992). Risk Perceptions and Pleasure Travel: An Exploratory Analysis. Journal of Travel Research, 30(4), 17–26. https://doi.org/10.1177/004728759203000403

Seddighi, H. ., Nuttall, M. ., & Theocharous, A. . (2001). Does cultural background of tourists influence the destination choice? an empirical study with special reference to political instability. Tourism Management (1982), 22(2), 181–191. https://doi.org/10.1016/S0261-5177(00)00046-7

Reassessing Gender Differences in COVID-19 Risk Perception and Behavior. Social Science Quarterly, 103(1), 31–41. https://doi.org/10.1111/ssqu.13116

Pons, V., Profeta, P., Foucault, M., Brouard, S., Galasso, V., & Becher, M. (2020). Gender Differences in COVID-19 Related Attitudes and Behavior: Evidence from a Panel Survey in Eight OECD Countries. https://doi.org/10.3386/w27359

Parikh, V., Keskar, M., Dharia, D., & Gotmare, P. (2018). A Tourist Place Recommendation and Recognition System. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 218–222. https://doi.org/10.1109/ICICCT.2018.8473077

Kucukefe, B. & Kaya Kanlı, N. (2023). Turkey's Labor Market During Covid-19 Pandemic: A K-Modes Analysis . İzmir İktisat Dergisi , 38 (1) , 142-157 . DOI: 10.24988/ije.1096088

Mantouka, E. G., Fafoutellis, P., Vlahogianni, E. I., & Oprea, G.-M. (2022). Understanding user perception and feelings for autonomous mobility on demand in the COVID-19 pandemic era. Transportation Research Interdisciplinary Perspectives, 16, 100692–. https://doi.org/10.1016/j.trip.2022.100692

Wang, D., Kotsi, F., Mathmann, F., Yao, J., & Pike, S. (2022). Short break drive holiday destination attractiveness during COVID-19 border closures. Journal of Hospitality and Tourism Management, 51, 568–577. https://doi.org/10.1016/j.jhtm.2022.05.013

Aresi, G., Procentese, F., Gattino, S., Tzankova, I., Gatti, F., Compare, C., Marzana, D., Mannarini, T., Fedi, A., Marta, E., & Guarino, A. (2022). Prosocial behaviours under collective quarantine conditions. A latent class analysis study during the 2020 COVID-19 lockdown in Italy. Journal of Community & Applied Social Psychology, 32(3), 490–506. https://doi.org/10.1002/casp.2571

Papachristou, N., Barnaghi, P., Cooper, B. A., Hu, X., Maguire, R., Apostolidis, K., Armes, J., Conley, Y. P., Hammer, M., Katsaragakis, S., Kober, K. M., Levine, J. D., McCann, L., Patiraki, E., Paul, S. M., Ream, E., Wright, F., & Miaskowski, C. (2018). Congruence Between Latent Class and K-Modes Analyses in the Identification of Oncology Patients With Distinct Symptom Experiences. Journal of Pain and Symptom Management, 55(2), 318–333.e4. https://doi.org/10.1016/j.jpainsymman.2017.08.020

Perić, G. ., Dramićanin, S. ., & Conić, M. . (2021). The impact of Serbian tourists' risk perception on their travel intentions during the COVID-19 pandemic. European Journal of Tourism Research, 27, 2705. https://doi.org/10.54055/ejtr.v27i.2125

Abdullah, M., Dias, C., Muley, D., & Shahin, M. (2020). Exploring the impacts of COVID-19 on travel behavior and mode preferences. Transportation Research Interdisciplinary Perspectives, 8, 100255–100255. https://doi.org/10.1016/j.trip.2020.100255

Yuni, L. H. K. (2020). Analysis of domestic tourist travel preferences post-Covid-19 pandemic. Journal of Applied Sciences in Travel and Hospitality, 3(2), 80-88.

Borges-Tiago, T., Silva, S., Avelar, S., Couto, J. P., Mendes-Filho, L., & Tiago, F. (2021). Tourism and COVID-19: The show must go on. Sustainability, 13(22), 12471.

Orîndaru, A., Popescu, M. F., Alexoaei, A. P., Căescu, Ș. C., Florescu, M. S., & Orzan, A. O. (2021). Tourism in a post-COVID-19 era: Sustainable strategies for industry's recovery. Sustainability, 13(12), 6781.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to data mining. Pearson Addison Wesley

.Ramires, A., Brandão, F., & Sousa, A. C. (2018). Motivation-based cluster analysis of international tourists visiting a World Heritage City: The case of Porto, Portugal. Journal of Destination Marketing & Management, 8, 49–60. https://doi.org/10.1016/j.jdmm.2016.12.001

Dey, B., & Sarma, M. K. (2010). Information source usage among motive-based segments of travelers to newly emerging tourist destinations. Tourism Management (1982), 31(3), 341–344. https://doi.org/10.1016/j.tourman.2009.03.015

Prebensen, N. K., Woo, E., Chen, J. S., & Uysal, M. (2012). Experience quality in the different phases of a tourist vacation: A case of northern Norway. Tourism Analysis, 17(5), 617-627.

Chaturvedi, A., Green, P., & Carroll, J. (2001). K-modes clustering. Journal of Classification, 18(1), 35–55. https://doi.org/10.1007/s00357-001-0004-3

Arimond, G., & Elfessi, A. (2001). A Clustering Method for Categorical Data in Tourism Market Segmentation Research. Journal of Travel Research, 39(4), 391–397. https://doi.org/10.1177/004728750103900405

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery, 2(3), 283–. https://doi.org/10.1023/A:1009769707641

Kaufman, L., Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis.. John Wiley. ISBN: 978-0-47031680-1

CHATURVEDI, A., FOODS, K., GREEN, P. E., & DOUGLAS CARROLL, J. (2001). K-modes clustering. Journal of Classification, 18(1), 35–55. https://doi.org/10.1007/s00357-001-0004-3

McCutcheon, A. L. (1987). Latent class analysis. SAGE.

Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. Critical care medicine, 49(1), e63–e79. https://doi.org/10.1097/CCM.0000000000004710

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136

Dayton, C. M., & Schafer, W. D. (1973). Extended Tables of t and Chi Square for Bonferroni Tests with Unequal Error Allocation. Journal of the American Statistical Association, 68(341), 78–83. https://doi.org/10.1080/01621459.1973.10481337

R: K-Modes Clustering. (n.d.-b). https://search.r-project.org/CRAN/refmans/klaR/html/kmodes.html

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis : with applications in the social behavioral, and health sciences. Wiley.

Glückstad, Fumiko Kano et al. “Cross-Cultural Bayesian Network Analysis of Factors Affecting Residents' Concerns About the Spread of an Infectious Disease Caused by Tourism.” Frontiers in psychology vol. 12 635110. 7 Jun. 2021, doi:10.3389/fpsyg.2021.635110

Fumiko Kano Glückstad. (2022). Risk Perceptions, Attitudes, and Behavioral Intentions to Spend on Experiences in the post-Corona Crisis - Data from Italy, Denmark, China, and Japan.

poLCA function - RDocumentation. (n.d.). https://www.rdocumentation.org/packages/poLCA/versions/1.6.0.1/topics/poLCA

H. Bozdogan. Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52:345-370, 1987

K. Larntz. Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. Journal of the American Statistical Association, 13:253-263, 1978.

Lipovetsky S, Conklin M. Decreasing Respondent Heterogeneity by Likert Scales Adjustment via Multipoles. Stats. 2018; 1(1):169-175. https://doi.org/10.3390/stats1010012

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach. Journal of the American Statistical Association, 96(453), 20–31. https://doi.org/10.1198/016214501750332668

Lee, Dong. (2016). Alternatives to P value: Confidence interval and effect size. Korean Journal of Anesthesiology. 69. 555. 10.4097/kjae.2016.69.6.555.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-39).

Varona Henao, D., & Toquica Arango, M. (2021). Exploring the Historical Behavior of Islamic State Groups: Latent Class Analysis and K-Modes Clustering Approach.

Moya Calderón, M., Chavarría Esquivel, K., Arrieta García, M. M., & Lozano, C. B. (2022). Tourist behaviour and dynamics of domestic tourism in times of COVID-19. Current Issues in Tourism, 25(14), 2207-2211.

Appendix

Table: Post-Hoc Analysis

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics (Critical Value) qnorm(adj p-val/2)
Demographics	Country	China(1.51) Denmark(16.27) Italy(12.60) Japan(-29.54)	China(2.64) Denmark(-15.85) Italy(-9.36) Japan(21.97)	China(-5.27) Denmark(-2.05) Italy(-5.20) Japan(12.17)	**±2.87**
	Profile_Age	18-24 Year (0.26) 25-34 Year (-4.09) 35-44 Year (-2.09) 45-54 Year (1.04) +55 Year (4.31)	18-24 Year (0.47) 25-34 Year (0.44) 35-44 Year (0.95) 45-54 Year (-3.09) +55 Year (0.95)	18-24 Year (-0.93) 25-34 Year (4.91) 35-44 Year (1.62) 45-54 Year (2.44) +55 Year (-6.91)	**±2.94**
	Gender	Male (8.63) Female (-8.63)	Male (-12.83) Female (12.83)	Male (4.38) Female (-4.38)	**±2.64**
Travel Experience	q1_1 (overseas for business in last 2 years)				No significant relation with clusters
	q1_2 (overseas for leisure in last 2 years)	Not at all (-9.02) 1-3 times (-2.66) 4-6 times (7.36) 7-12 times (4.81) 13+ times (0.54)	Not at all (5.45) 1-3 times (4.02) 4-6 times (-6.78) 7-12 times (-4.26) 13+ times (-0.06)	Not at all (5.26) 1-3 times (-1.43) 4-6 times (-1.42) 7-12 times (-1.13) 13+ times (-0.65)	**±2.94**
	q1_3 (domestic for business in last 2 years)	Not at all (0.05) 1-3 times (-5.03) 4-6 times (3.25) 7-12 times (3.18) 13+ times (0.80)	Not at all (-3.95) 1-3 times (4.10) 4-6 times (-1.53) 7-12 times (0.50) 13+ times (-0.77)	Not at all (4.81) 1-3 times (1.63) 4-6 times (-2.43) 7-12 times (-4.85) 13+ times (-0.11)	**±2.94**
Preferred Travel Arrangement	q3_1 (Preferred Travel Arrangement)	Arrange by Third Party (-5.49) Self-Planned (5.49)	Arrange by Third Party (3.34) Self-Planned (-3.34)	Arrange by Third Party (3.19) Self-Planned (-3.19)	**±2.64**
Preferred Travel Companionship	q4_1 (preferred travel companionship)	Larger group (above 8 people) (-0.73) Closest family or friends (-16.27) Alone or with a significant other	Larger group (above 8 people) (-2.18) Closest family or friends (13.76) Alone or with a significant other (-	Larger group (above 8 people) (3.67) Closest family or friends (4.65) Alone or with a significant other (-	**±2.77**

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics (Critical Value) qnorm(adj p-val/2)
How Important the factors for next travel	q11_5 (A destination where I have previous travel experiences)	Unimportant (8.45) Neutral (-9.27) Important (1.79)	Unimportant (-1.83) Neutral (-2.36) Important (4.47)	Unimportant (-8.98) Neutral (15.19) Important (-7.91)	**±2.77**
	q11_6 (A safe travel destination)	Unimportant (-0.37) Neutral (-13.04) Important (12.24)	Unimportant (-1.75) Neutral (-10.56) Important (10.61)	Unimportant (2.65) Neutral (30.42) Important (-29.41)	**±2.77**
	q11_7 (A destination with good shopping possibilities)	Unimportant (9.96) Neutral (-8.45) Important (-0.78)	Unimportant (-2.48) Neutral (-6.91) Important (9.55)	Unimportant (-10.19) Neutral (19.79)	**±2.77**
	q11_10 (A destination offering possibilities to visit museums, exhibitions, historical attractions)	Unimportant (3.39) Neutral (-13.37) Important (10.57)	Unimportant (0.23) Neutral (-2.37) Important (2.15)	Unimportant (-4.80) Neutral (20.74) Important (-16.73)	**±2.77**
	q11_11 (A destination offering possibilities to visit amusement parks, zoos, water parks and so on)	Unimportant (13.22) Neutral (-10.44) Important (-2.57)	Unimportant (-4.76) Neutral (-5.67) Important (10.83)	Unimportant (-11.71) Neutral (20.91) Important (-9.96)	**±2.77**
	q11_12 (A destination with good restaurants, cafes, bars and so on)	Unimportant (-0.35) Neutral (-10.23) Important (10.27)	Unimportant (5.85) Neutral (-5.87) Important (1.2)	Unimportant (-6.78) Neutral (20.89) Important (-15.16)	**±2.77**
	q11_13 (A destination with good hygiene which minimises risks of spreading infectious diseases)	Unimportant (-0.09) Neutral (-13.80) Important (12.73)	Unimportant (-1.59) Neutral (-9.73) Important (9.73)	Unimportant (2.09) Neutral (30.41) Important (-28.99)	**±2.77**
	q11_14 (Peace and quiet - a destination with less tourists)	Unimportant (1.42) Neutral (-7.93) Important (6.74)	Unimportant (-0.92) Neutral (-11.00) Important (10.87)	Unimportant (-0.74) Neutral (24.15) Important (-22.42)	**±2.77**
	q11_15 (A destination close to a beach, harbour and coast line)	Unimportant (1.79) Neutral (-14.24) Important (12.62)	Unimportant (2.08) Neutral (-1.07) Important (-0.42)	Unimportant (-4.97) Neutral (20.28) Important (-16.29)	**±2.77**
	q11_16 (A destination with forests and nature)	Unimportant (2.63) Neutral (-9.76) Important (7.69)	Unimportant (-2.75) Neutral (-7.51) Important (8.77)	Unimportant (-0.10) Neutral (22.29) Important (-21.08)	**±2.77**
	q11_18 (Children friendly destination)	Unimportant (14.46) Neutral (-12.59) Important (-1.87)	Unimportant (-6.23) Neutral (-2.59) Important (9.23)	Unimportant (-11.55) Neutral (19.96)	**±2.77**
	q11_19 (A destination friendly for senior citizens)	Unimportant (13.05) Neutral (-8.09) Important (-4.63)	Unimportant (-5.41) Neutral (-6.52) Important (12.82)	Unimportant (-10.69) Neutral (18.83)	**±2.77**
	q11_20 (Clean destination (no trash, clean beach and air))	Unimportant (-0.43) Neutral (-14.35) Important (13.63)	Unimportant (-0.86) Neutral (-7.31) Important (7.26)	Unimportant (1.63) Neutral (28.15) Important (-27.13)	**±2.77**

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics (Critical Value) qnorm(adj p-val/2)
Attitudes to public behaving responsible	q19_1 (It is important that individuals contribute to minimise the risk of spreading infectious diseases in public spaces)	Disagree (-4.09) Neutral (-16.44) Agree (16.60)	Disagree (-3.24) Neutral (-9.80) Agree (10.36)	Disagree (9.45) Neutral (34.00) Agree (-34.91)	**±2.77**
	q19_8 (I feel safe and comfortable if staffs in hotels, airlines, restaurants etc. wear a mask)	Disagree (2.50) Neutral (-9.47) Agree (6.72)	Disagree (-4.43) Neutral (-11.02) Agree (12.66)	Disagree (2.14) Neutral (26.23) Agree (-24.60)	**±2.77**
Attitudes to self-protective behaviors	q19_2 (I am keeping social distances in public spaces. If it is not possible, I will leave that place)	Disagree (0.89) Neutral (-14.52) Agree (11.76)	Disagree (-5.92) Neutral (-8.99) Agree (11.56)	Disagree (6.14) Neutral (30.44) Agree (-29.95)	**±2.77**
	q19_4 (I carry and use disinfectant to clean my hand after touching items in shops to make me feel clean and safe)	Disagree (-2.12) Neutral (-12.67) Agree (11.84)	Disagree (-0.26) Neutral (-7.47) Agree (6.19)	Disagree (3.14) Neutral (26.11) Agree (-23.41)	**±2.77**
	q19_6 (I wear a mask to make me feel safe)	Disagree (14.37) Neutral (-8.76) Agree (-5.82)	Disagree (-14.32) Neutral (-11.87) Agree (21.36)	Disagree (-1.42) Neutral (26.34) Agree (-18.98)	**±2.77**
Attitudes to responsible behaviors	q19_3 (I clean up a public space (e.g. Toilet) after I use it so that people who use it after me feel clean and safe)	Disagree (3.74) Neutral (-8.91) Agree (5.23)	Disagree (-5.92) Neutral (-12.00) Agree (14.47)	Disagree (2.33) Neutral (26.70) Agree (-24.86)	**±2.77**
	q19_5 (I carry and use disinfectant to clean my hands before touching items in shops so that other people who touch after me feel clean and safe)	Disagree (-1.96) Neutral (-12.76) Agree (11.82)	Disagree (0.30) Neutral (-7.20) Agree (5.54)	Disagree (2.24) Neutral (25.89) Agree (-22.59)	**±2.77**
	q19_7 (I wear a mask to keep those around me safe and comfortable)	Disagree (13.72) Neutral (-10.25) Agree (-3.89)	Disagree (-13.53) Neutral (-11.64) Agree (20.09)	Disagree (-1.53) Neutral (28.04) Agree (-19.66)	**±2.77**

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics (Critical Value) qnorm(adj p-val/2)
Attitudes to traveling within the country in Transitional Phase	q22_1A	Dangerous (-27.39) Neutral (-6.23) Safe (30.76)	Dangerous (33.41) Neutral (-5.76) Safe (-25.72)	Dangerous (-4.83) Neutral (15.41) Safe (-9.16)	**±2.77**
	q22_1B	Unenjoyable (-27.16) Neutral (-8.70) Enjoyable (31.78)	Unenjoyable (32.67) Neutral (-4.36) Enjoyable (-25.20)	Unenjoyable (-4.24) Neutral (16.96) Enjoyable (-11.15)	**±2.77**
	q22_1C	Onerous (-26.51) Neutral (-3.76) Effortless (28.92)	Onerous (33.30) Neutral (-8.54) Effortless (-24.45)	Onerous (-5.88) Neutral (15.57) Effortless (-8.28)	**±2.77**
	q22_1D	Harmful (-25.09) Neutral (-5.10) Beneficial (26.76)	Harmful (30.61) Neutral (-3.76) Beneficial (-23.23)	Harmful (-4.45) Neutral (11.43) Beneficial (-6.91)	**±2.77**
Intention to Travel within country for Pleasure	Q22_2_1 (I intend to travel for pleasure within Country in the transitional phase)	Disagree (-22.15) Neutral (-14.67) Agree (32.20)	Disagree (25.31) Neutral (-4.37) Agree (-18.36)	Disagree (-1.80) Neutral (24.93) Agree (-20.16)	**±2.77**
Assumptions of Known People Towards Respondents Travel during Transitional Phase	Q22_2_2 (Most people who are important to me think that my traveling within Country in the transitional phase will be thoughtful and respectful of their and my safety)	Disagree (-18.17) Neutral (-14.75) Agree (27.46)	Disagree (20.31) Neutral (-4.64) Agree (-11.60)	Disagree (-0.92) Neutral (25.37) Agree (-22.22)	**±2.77**
	Q22_2_3 (Most people whose opinion I value would approve my traveling within Country in the transitional phase)	Disagree (-20.87) Neutral (-16.94) Agree (32.32)	Disagree (23.73) Neutral (-0.01) Agree (-18.95)	Disagree (-1.55) Neutral (22.56) Agree (-19.60)	**±2.77**
	Q22_2_4 (Most people I respect and admire will be traveling within Country in the transitional phase)	Disagree (-22.88) Neutral (-7.86) Agree (27.50)	Disagree (25.99) Neutral (-8.51) Agree (-14.50)	Disagree (-1.67) Neutral (20.99) Agree (-18.68)	**±2.77**
Resources (Media, Money & Time) Role during transitional phase	Q22_2_5 (Mass and social media will encourage traveling activities within Country in the transitional phase)	Disagree (-11.02) Neutral (-16.96) Agree (24.09)	Disagree (11.18) Neutral (-0.01) Agree (-8.47)	Disagree (0.85) Neutral (22.59) Agree (-21.59)	**±2.77**
	Q22_2_6 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available financial resources to do it)	Disagree (-9.66) Neutral (-15.66) Agree (21.59)	Disagree (9.99) Neutral (-2.07) Agree (-6.07)	Disagree (0.51) Neutral (23.41) Agree (-21.24)	**±2.77**
	Q22_2_7 (If I wanted to travel within Country in the transitional phase, I am confident that I will have enough available time to do it)	Disagree (-14.48) Neutral (-15.43) Agree (24.72)	Disagree (13.43) Neutral (-2.45) Agree (-8.08)	Disagree (2.67) Neutral (23.57) Agree (-22.92)	**±2.77**

Category	Variable	Cluster-1 "Risk-Takers" (2154, 52%)	Cluster-2 "Risk-Averse" (1310, 31%)	Custer-3 "Risk-Neutral" (708,17%)	Statistics (Critical Value) qnorm(adj p-val/2)
COVID-19 perception during transitional phase	Q22_3_1 (If I travel within Country in the transitional phase, I will contribute to the spread of Covid-19 at the destination)	Disagree (30.85) Neutral (-10.00) Agree (-23.17)	Disagree (-24.29) Neutral (-7.86) Agree (33.28)	Disagree (-11.04) Neutral (23.03) Agree (-10.30)	**±2.77**
	Q22_3_2 (If I travel within Country in the transitional phase, I will get infected with Covid-19)	Disagree (29.89) Neutral (-8.10) Agree (-23.41)	Disagree (-24.21) Neutral (-7.30) Agree (32.93)	Disagree (-9.86) Neutral (19.82) Agree (-9.55)	**±2.77**
	Q22_3_3 (If I travel within Country in the transitional phase, I will infect others (relatives, friends, colleagues etc.) with Covid-19)	Disagree (31.84) Neutral (-13.01) Agree (-22.12)	Disagree (-24.17) Neutral (-3.86) Agree (30.59)	Disagree (-12.50) Neutral (22.10) Agree (-8.37)	**±2.77**

adj p-val is bonferroni adjustment of p-value calculated for each variable, based on the number of comparisons performed for that variable with cluster variable.

Eg: p-value for Q22_3_1 is 0.05, no.of pairwise comparisons=9, so "adj p-val" =0.05/9=0.0056

Note: Only the significant values are highlighted.

Green are the significant is positive direction Red are the significant in negative direction