

“Shifting Travel and Hospitality Consumer Groups”

Jorge LeonFrausto

TABLE OF CONTENTS

I. Introduction	3
II. Literature Review	5
III. Data Information and Exploration	9
a. Data Summary	9
b. June 2020 Data	9
c. December 2002 Data	14
IV. Machine Learning Models	17
a. June 2020 Consumer Clusters	17
b. Dec 2020 Consumer Clusters	21
V. Cluster Analysis	24
a. June Clusters	24
b. December Clusters	25
c. Clusters Comparison	27
VI. Discussion.	30
a. June Data	30
b. December Data	34
c. Going Forward	35

I. Introduction

Coronavirus disease, or COVID-19 is an infectious disease caused by a newly discovered coronavirus. (WHO, 2021) On March 11, 2020 WHO made the assessment that COVID-19 could be characterized as a pandemic. (WHO, 2021)

The actions that governments took to help prevent the spread of the disease affected the tourism and hospitality industries the most. Since the pandemic began, aviation passenger traffic declined by 60 percent in 2020. International tourism recorded its worst year ever on record – a decline of 74 percent (CCSA, 2021). This decline represents a loss of an estimated 1.3 trillion USD in international tourism expenditure, about 11 times the loss of the 2009 global crisis (CCSA, 2021). According to the AHLA (2021), Leisure and hospitality has lost 3.1 million jobs during the pandemic that have yet to return, representing more than a third of all unemployed persons in the United States.

On December 11, 2020, the U.S. Food and Drug Administration issued the first emergency use authorization (EUA) for a vaccine for the prevention of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in individuals 16 years of age and older.

From past literature, consumers seem place safety and risk perception as the most important factor in deciding to engage in public activities – including travel. As the economy “reopens” and customers and businesses embrace a new normal, it is key for travel and tourism

businesses – inclusive of travel, hospitality, and restaurant industries - to understand their consumers.

Related to travel and hospitality there are three periods that surround coronavirus. The “old normal” before COVID-19 was identified in China. The “new normal” period, which started with the full declaration of coronavirus as a pandemic. In the United States, lockdowns and travel restrictions were put into place. The next period would be the transition period into the “return to normal” stage which is characterized by the development of the coronavirus vaccine.

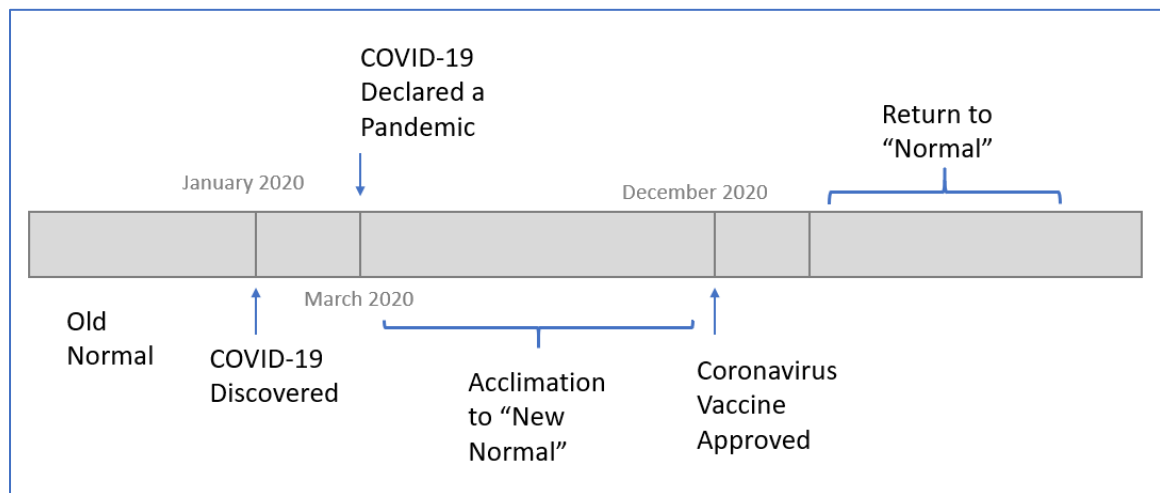


Figure 1: Coronavirus Time Periods

The data used for analysis of traveler sentiment and groups is gathered from two periods: June 2020, and December 2020. The June 2020 period allows for a three-month period of adaptation into the “new normal.” With the December 2020 period data, we can conduct an analysis of change in sentiment around two factors: an increase in number of cases; and, the development and rollout of the coronavirus vaccine.

The goals of the study are to identify: 1) Distinct consumer groups; 2) The factors that would best promote a sense of safety and engage travelers and tourists within each cluster; and, 3) Any shifts in cluster composition from the June 2020 to the December 2020 period.

II. Literature Review

Past literature studying the effect of coronavirus and changes in consumer perspectives have focused specific effects on business within the travel and hospitality industry. The institution who carries out any given study has a significant influence in the nature of the study or survey. This influence extends to the language and terminology used. As a result, terminology used tends to overlap with blurred lines as to what is the separation between “travel,” “hospitality,” and “tourism.” As a generality, there are three “industries” in the general classification: 1) Travel industry, which can include business and leisure travel; 2) Hotel industry, in some literature referred to as accommodation or lodging; 3) Restaurant industry.

a. Travel Industry

Closest to this project, Neuburger (2021) carried a study seeking to identify profiles of tourists during a pandemic outbreak at two different stages of its spread. The authors collected the data in the first week of March 202 prior to COVID-19 being declared a pandemic; and, late March after the declaration by the World Health Organization.

A total of four clusters were identified: Anxious, Nervous, Reserved, and Relaxed. Anxious cluster had the highest level of risk perception. Relaxed cluster had the lowest values of

risk perception. The study noted that after the pandemic was officially declared the ‘Relaxed’ cluster assignments to participants disappeared. Age was a key identified demographic variable. As age increase so did the travel risk perception of participants.

b. Hotel or Lodging Industry

Within the tourist sector, the hotel is one of the most vulnerable to the effects of the COVID-19 pandemic (Lopez, 2020). Lopez studied the necessary actions that could defuse the effects of the pandemic on the hotel industry. This study took place in Spain via an online survey between April and June 2020.

The study presented 17 actions that either hotels or guests could take. Participants were asked to rate their importance. The most important or those with which participants agreed the most were: 1) The hotel must present good hygiene conditions; 2) Adaptation to the recommendations of the World Health Organization; The social distance must be respected at the hotel; and, 3) Official information regarding the two actions above must available to guests prior to, and at, check-in.

In June 2020, The New York Times highlighted in an article cleaning-focused initiatives and programs started by hotels. (Firshein, 2020). In it, they point out that hotels had begun to adapt reopening practices for businesses – as there were no CDC-provided guidelines specific to hotels.

c. Restaurants

Byrd et al. (2020) studied consumers risk perceptions surrounding restaurant food. The data was collected through an online survey during the first week of May 2020. The article indicates that that 89% of consumers believed that food from grocery stores and home was safer than food from restaurants.

Consumers were more concerned about restaurant food and restaurant food packaging than the broader category of food in general. (Byrd et al., 2021) From the categories explored, the two highest categories which represented a concern to consumers were: 1) Food served in restaurants; and, 2) for cold, raw, or uncooked food from restaurants. The study found differences in four demographic areas: gender; people reported as being in a high-risk category of contracting COVID-19; race / ethnicity; and, household size.

An important highlight of the study is that almost twice as many participants indicated they were “very concerned” about contracting COVID-19 from food served in restaurants compared to restaurant food in general. (Byrd et al., 2020). This risk perception extends when comparing the food served in restaurants to take-out and delivered restaurant food. Even though, in theory, it would be the same food served.

In another study which focused on the in-restaurant experience, Tuzovic et al. (2021) studied consumer sentiment as applied to measures taken by restaurants for dine-in consumers. The study took place between May and June 2020 in Germany. The results showed that restaurants are expected to aid (e.g., instructions) their guests in minimizing their initial insecurity and uncertainty (Tuzovic, 2021). Specific actions that restaurants can take include: listing information regarding their cleaning, and safety precautions information on their website; having hand sanitizers and disinfectants at the entrance; and, employees must also adhere to new safety measures. Providing the cleaning information would allow patrons to have this

information when planning their visit. The availability of hygienic products was described as a must.

d. Post-Pandemic Travel (“The Road to Recovery”)

For consumers to feel confident enough to return to a semblance of normal physical interactions with retailers, hotels and other consumer-facing businesses, the first thing they’ll need is a solid assurance that places of business have a plan to make their customer experience as safe as possible. (PwC, 2020)

The McKinsey report (Krishnan et al., 2020) indicates that, as of August 2020, anxiety surrounding COVID-19 remains high. They suggest that authorities and destination managers must work to ensure travelers know about, and feel reassured by, protocols put in place for their protection (Krishnan et al., 2020). The level of information provided and communication from authorities is a key factor in bolstering demand.

A KEKST CNC poll from March 2021 looked at what consumers are “actually looking for and the factors that will define the choice of destination.” (KEKST CNC, 2021) This is an international poll from six countries: United Kingdom, United States, Germany, Sweden, Japan and France. Personal vaccination is the most important factor that people say would make them likely to travel globally. Older tourists insist that they have a vaccine before travel (47%), as well as people in higher income groups. The top-rated factors for consumers when deciding to go on vacation after a COVID-19 vaccine is available are: That they’ve had the vaccine (38%); falling cases of known infections numbers globally (30%); falling number of cases in the country that

they are traveling to (29%); and, that there is no quarantine when returning to their country (25%).

e. Conceptual Model

Consumers believe that precautions should be taken by hospitality businesses. In choosing to visit a hospitality business, consumers place a priority in having information as to what actions a particular business is taking or has taken to address health concerns. The next step or phase is the decision to visit the place based on what actions a business is taking to prevent potentially infected people from entering the business (“barriers to entry”); and finally, the actions taken inside the business to prevent, or diminish risk of contraction of COVID-19.

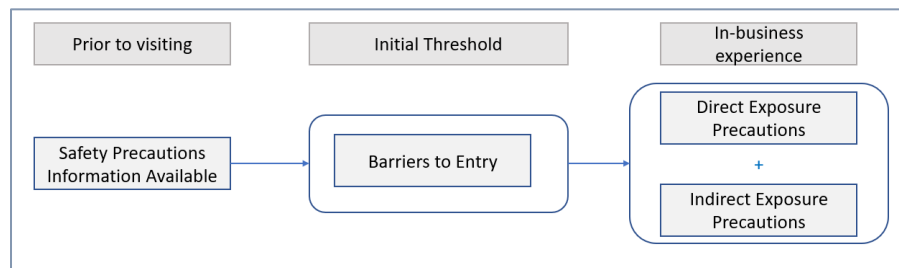


Figure 2: Factor categories influencing tourists decision making.

III. Data Information and Exploration

Data was compiled from three data sets: 1) Accenture Research data collected during June 2002; 2) Axios/Ipsos Coronavirus Survey data from June 12 – 22, 2020; and, 3) Axios/Ipsos Coronavirus Survey data from December 4 – 21, 2020. The Axios studies were conducted using the web-enabled KnowledgePanel service.

I analyzed the data from each time period independently. Therefore, the Accenture Research and the June Axios/Ipsos data was merged into a single data set for analysis. The December data in Axios/Ipsos data was retained for that time period.

After the data analysis, and data cleaning the June data set had 1,769 observations and the December data set 3,581.

a. Data Exploration

June 2020 Data

The June data set used for analysis was composed of 1,769 observations. The data included the following demographic variables: age, gender, employment status, level of education, household size, if respondent was a parent, and income. The initial exploration of the data began with some summary statistics. The frequencies of these features indicate the following:

- 48.62% of the respondents were women; 51.38% men.
- The breakdown by age shows that the largest individual group was respondents 56 to 69 years old (31.77% of observations).
- Most respondents had a level of education with a high school degree or less at 33.80%

- The percentage of respondents with a bachelor's degree or higher is 36.93%.
- The most common household size is of 2 persons (40.76%).
- Respondents making \$100,000 or more yearly account for 40.59% of the observations.

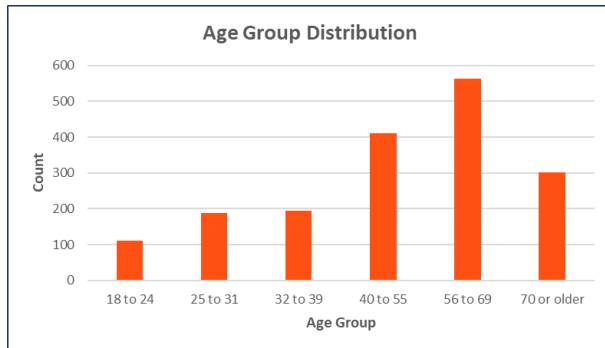


Figure D3: Age Group Distribution.

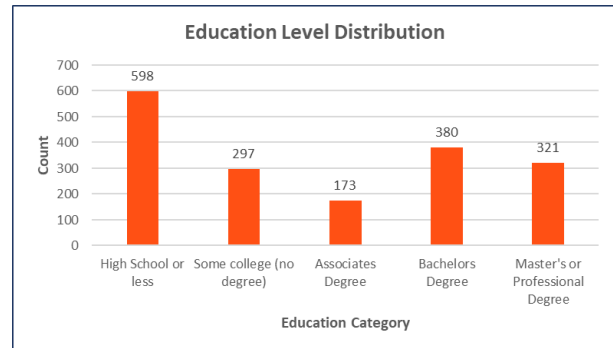


Figure D4: Education Level Distribution.

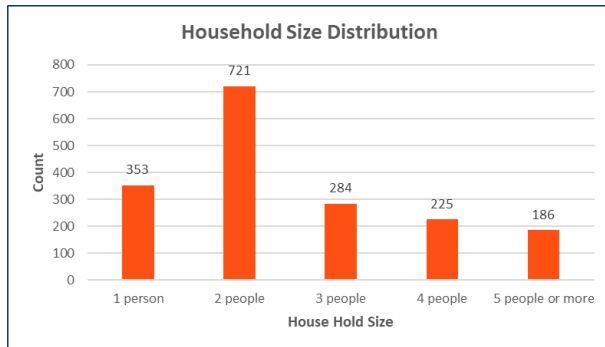


Figure D5: Distribution of Household Size

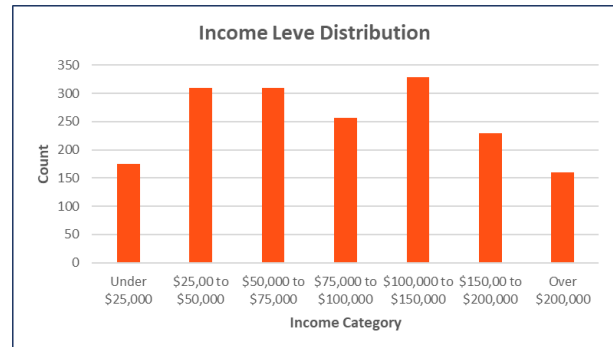


Figure D6: Distribution of Income Level Group.

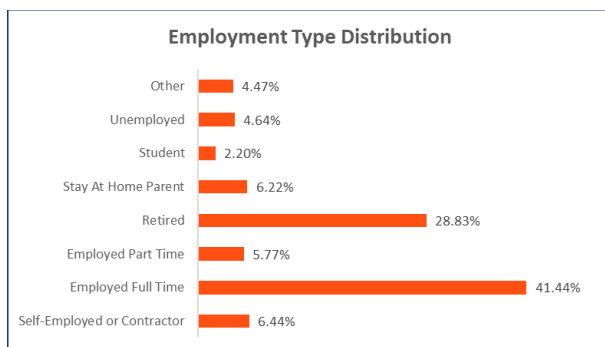


Figure D7: Employment Type Distribution

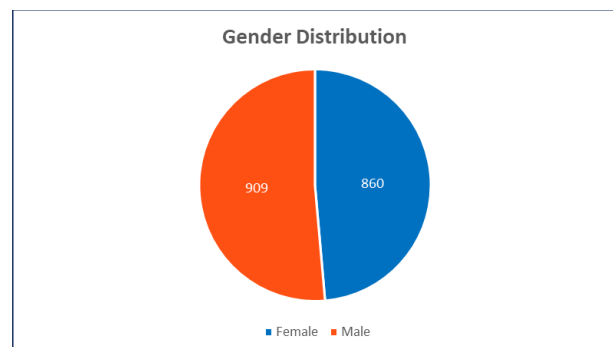


Figure D8: Gender Distribution

Of interest to the analysis is the perceived risk of particular actions and concern about the coronavirus outbreak. The respondents were asked to indicate concern for the coronavirus outbreak from the following options: 1) Not at all concerned; 2) not very concerned; 3) somewhat concerned; 4) very concerned; or 5) extremely concerned.

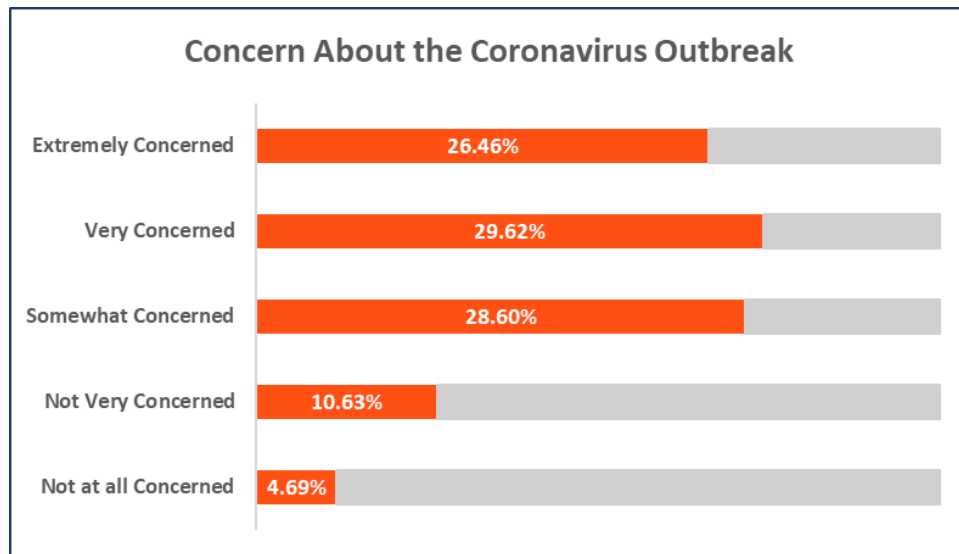


Figure 9: Concern for Coronavirus Distribution

56.08% of respondents were very or extremely concerned by the coronavirus outbreak. The table below shows the distribution of these options broken down by the various demographics in the dataset.

The combinations which show an increase compared to the overall dataset values for the “Not at All Concerned” about coronavirus are: those of age 18 to 24 years old; those living in households with 5 or more persons; those with income under \$25,000; and, those who identified as unemployed, or “other” in the employment type. These last two demographics are also those that have a significant increase in the “Extremely Concerned” category.

Variable	Variable Values	Not at all Concerned	Not Very Concerned	Somewhat Concerned	Very Concerned	Extremely Concerned
Gender	Female	3.37%	8.37%	26.51%	30.47%	31.28%
	Male	5.94%	12.76%	30.58%	28.82%	21.89%
Age Group	18 to 24	10.81%	11.71%	30.63%	26.13%	20.72%
	25 to 31	5.82%	7.94%	34.92%	28.57%	22.75%
	32 to 39	5.13%	8.72%	29.74%	29.23%	27.18%
	40 to 55	5.37%	13.66%	26.83%	28.78%	25.37%
	56 to 69	4.09%	10.85%	27.05%	29.00%	29.00%
	70 or older	1.66%	8.61%	28.48%	34.11%	27.15%
Education	High School or less	6.52%	9.87%	31.27%	25.42%	26.92%
	Some college (no degree)	5.39%	12.79%	27.95%	29.63%	24.24%
	Associates Degree	5.20%	8.09%	32.37%	30.64%	23.70%
	Bachelor's Degree	2.37%	14.21%	27.11%	29.47%	26.84%
	Master's or Professional Degree	3.12%	7.17%	23.99%	37.07%	28.66%
Household Size	1 Person	6.52%	8.22%	28.90%	26.06%	30.31%
	2 Persons	2.08%	10.26%	27.60%	32.87%	27.18%
	3 Persons	5.99%	10.21%	29.58%	29.23%	25.00%
	4 Persons	4.44%	12.89%	28.89%	27.11%	26.67%
	5 or more persons	9.68%	14.52%	30.11%	27.42%	18.28%
Income Level	Under \$25,000	8.00%	8.00%	22.29%	26.29%	35.43%
	\$25,00 to \$50,000	3.87%	10.00%	30.65%	26.45%	29.03%
	\$50,000 to \$75,000	3.55%	10.65%	29.35%	30.00%	26.45%
	\$75,000 to \$100,000	6.25%	12.11%	26.17%	34.77%	20.70%
	\$100,000 to \$150,000	3.95%	10.64%	28.57%	28.27%	28.57%
	\$150,00 to \$200,000	4.37%	11.35%	31.00%	32.31%	20.96%
	Over \$200,000	4.38%	11.25%	30.63%	29.38%	24.38%
Employment	Self-Employed or Contractor	7.89%	9.65%	31.58%	28.07%	22.81%
	Employed Full Time	4.64%	12.82%	28.51%	29.33%	24.69%
	Employed Part Time	2.94%	7.84%	29.41%	28.43%	31.37%
	Retired	2.94%	8.82%	27.45%	32.55%	28.24%
	Stay At Home Parent	5.45%	14.55%	32.73%	23.64%	23.64%
	Student	5.13%	15.38%	33.33%	30.77%	15.38%
	Unemployed	8.54%	6.10%	26.83%	35.37%	23.17%
	Other	8.86%	3.80%	25.32%	18.99%	43.04%
TOTAL		4.69%	10.63%	28.60%	29.62%	26.46%

Table 10: Distribution of Coronavirus outbreak concern by demographic.

December 2020 Data

The December data set used for the second time frame in the analysis was composed of 3,581 observations. It included the same demographic data as the June period. The data has a similar distribution of demographics as the data for June.

- 48.37% of the respondents were women; 51.63% men.
- The breakdown by age shows that the largest individual group was respondents 56 to 69 years old (30.77% of observations).
- 32.53% of respondents had a level of education with a high school degree or less, while respondents with a bachelor's degree or higher is 40.83%.
- The most common household size is of 2 persons (42.28%).
- Respondents making \$100,000 or more yearly account for 43.42% of the observations.

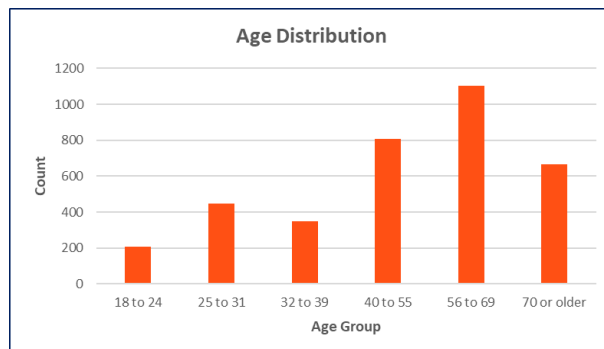


Figure D11: Age Distribution

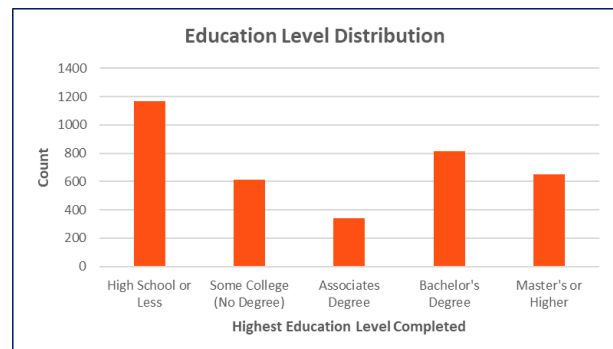


Figure D12: Education Level Distribution

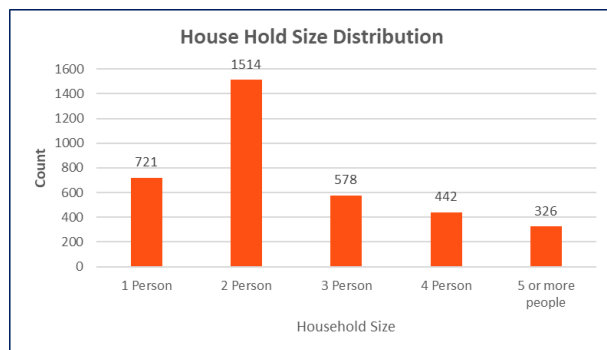


Figure D13: Household Size Distribution

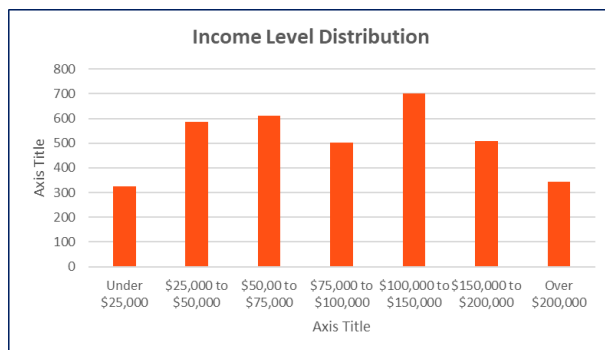


Figure D14: Income Level Distribution

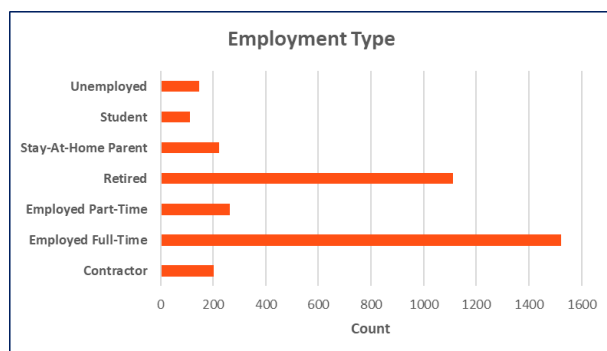


Figure D15: Employment Type Distribution

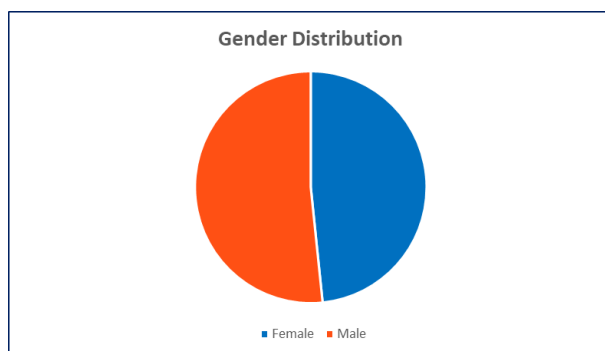


Figure D16: Gender Distribution

The reported concern levels for coronavirus reflect a shift towards a higher degree of concern for coronavirus in the December period compared to the June data. For the December period the percentage of respondents who were very or extremely concerned about Coronavirus increased from 56% to about 64%.

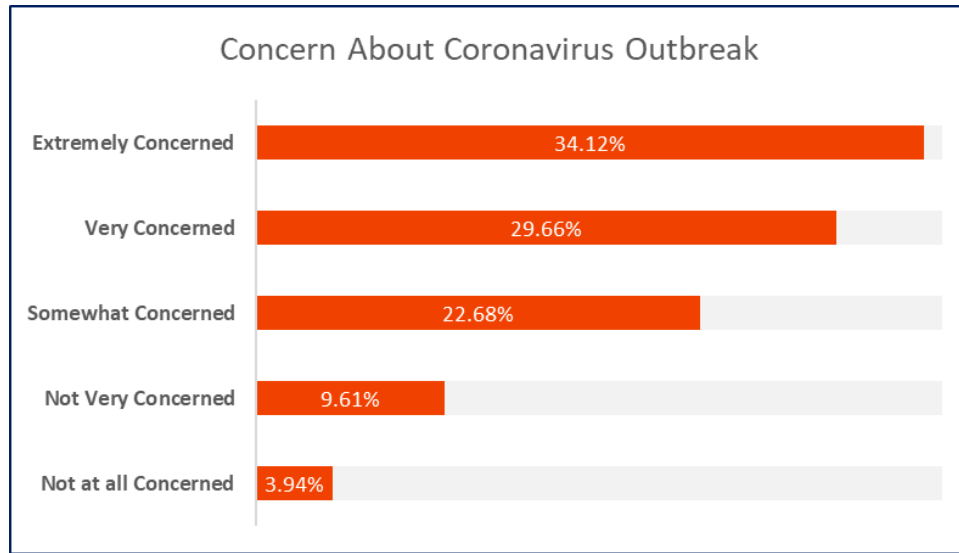


Figure 17: Reported concern about coronavirus distribution

The table below shows the same breakdown for the percentages of how concerned respondents indicated they were about coronavirus. The groups which had the highest reported percentage of respondents indicating they were “Extremely Concerned” were: women; those making under \$25,000; and, those retired. The demographics which had the highest percentage of respondents “Not at all Concerned” were: those 40 to 55 years old; contractors; and those living in household sizes of 5 or higher.

Variable	Variable Values	Not at all Concerned	Not Very Concerned	Somewhat Concerned	Very Concerned	Extremely Concerned
Gender	Female	2.66%	8.20%	20.73%	29.56%	38.86%
	Male	5.14%	10.92%	24.50%	29.75%	29.69%
Age Group	18 to 24	5.34%	10.68%	26.70%	30.58%	26.70%
	25 to 31	4.91%	10.04%	27.23%	28.35%	29.46%
	32 to 39	5.14%	12.00%	20.86%	32.57%	29.43%
	40 to 55	6.07%	10.16%	24.04%	28.62%	31.10%
	56 to 69	2.36%	10.16%	21.78%	27.77%	37.93%
	70 or older	2.25%	6.14%	19.16%	33.08%	39.37%
Education	High School or less	4.66%	11.08%	22.74%	30.03%	31.49%
	Some college (no degree)	3.33%	9.85%	22.04%	32.39%	32.39%
	Associates Degree	4.98%	10.82%	25.84%	24.03%	34.33%
	Bachelor's Degree	2.15%	5.54%	17.38%	35.54%	39.38%
	Master's or Professional Degree	4.26%	10.47%	23.08%	30.28%	31.91%
Household Size	1 Person	4.02%	8.60%	19.69%	33.01%	34.67%
	2 Persons	3.57%	8.72%	22.79%	27.94%	36.99%
	3 Persons	3.29%	9.17%	23.88%	30.62%	33.04%
	4 Persons	4.07%	11.09%	25.79%	30.32%	28.73%
	5 or more persons	6.44%	14.72%	22.39%	27.61%	28.83%
Income Level	Under \$25,000	4.31%	8.31%	19.38%	24.62%	43.38%
	\$25,00 to \$50,000	3.41%	9.37%	21.47%	27.60%	38.16%
	\$50,000 to \$75,000	4.75%	10.15%	22.42%	29.13%	33.55%
	\$75,000 to \$100,000	4.57%	11.13%	23.66%	31.01%	29.62%
	\$100,000 to \$150,000	3.56%	11.11%	25.78%	29.34%	30.20%
	\$150,00 to \$200,000	4.32%	7.27%	22.79%	29.67%	35.95%
	Over \$200,000	2.33%	8.43%	20.35%	37.50%	31.40%
Employment	Self-Employed or Contractor	5.91%	16.75%	24.63%	23.15%	29.56%
	Employed Full Time	4.93%	10.18%	23.72%	29.63%	31.54%
	Employed Part Time	1.89%	8.71%	27.27%	29.17%	32.95%
	Retired	2.70%	7.83%	19.53%	30.78%	39.15%
	Stay At Home Parent	3.60%	10.36%	20.27%	32.43%	33.33%
	Student	4.46%	7.14%	29.46%	30.36%	28.57%
	Unemployed	4.08%	9.52%	23.13%	26.53%	36.73%
TOTAL		3.94%	9.61%	22.68%	29.66%	34.12%

Table 18: Concern levels about coronavirus across demographics

IV. Machine Learning Models

For applicability to hospitality and tourism industry, businesses would most be interested in what actions they can take to attract customers. I created a clustering model based on the risk perception, known behaviors, and age attributes to determine clusters (if any) could be created. A different set of clusters was created for each time period.

a. Creating Consumer Clusters

The clustering algorithm requests two inputs from the user: the variables to be used; and the number (or range of numbers of) clusters. Using the variables that addressed: present state of wellbeing of respondents; perceived risks; changes in lifestyle; concern for effects of coronavirus; and demographics I used various combinations to create clusters and analyzed the clusters returned by the algorithm for each period.

b. June 2020 Consumer Clusters

After reviewing the various cluster “types” as defined by which variables were used and the number of clusters assigned to each model, a single cluster model was chosen. The final clustering model was chosen to identify five different clusters. The significant features that formed part of this model were:

- Risk Perception Activities
 - Level of risk perception for visiting grocery stores
 - Perceived risk of dining at restaurants
 - Perceived risk of attending family or friends’ gatherings
 - Perceived risk of going to a salon / barbershop
 - Perceived risk of taking a vacation

- Changes in Lifestyle Variables
 - Change in savings
 - Change in spending
 - Change in household debt
- Coronavirus Health Factors
 - Knowing someone who tested positive for coronavirus
 - Knowing someone who died from coronavirus
 - Having personally tested positive for coronavirus

Understanding that demographics were not used, these groups show separation based almost exclusively on age and employment type. The analysis post-creation of clusters showed that gender and income level were not significant factors in the composition of the clusters. I will next review the underlying conditions assumed based on how the groups were formed.

Cluster	Description
Cluster 1: “Controllers”	<ul style="list-style-type: none"> • 90% of respondents in this cluster are ages 32 to 55. • Much more likely to be stay-at-home parent. • Much more likely to be a contractor. • 2nd least concerned group about Covid.
Cluster 2: “Relaxed”	<ul style="list-style-type: none"> • Least concerned about coronavirus and risk from activities. • Higher proportion are retired.
Cluster 3: “At-Risk”	<ul style="list-style-type: none"> • Group most concerned about coronavirus • Highest concentration of people 40 years of age or older than other clusters. • 70% of all respondents over 70 are in this group. • Highest proportion of retired people (attributed to age effect above)

Cluster 4: “Known Risk”	<ul style="list-style-type: none"> • Smallest group. • All respondents tested positive for coronavirus.
Cluster 5: “Exposed”	<ul style="list-style-type: none"> • The cluster most concerned about coronavirus. • Higher proportion are employed full-time and part-time. • About half of respondents 32 to 55 in data belong in this group.

Table 19: June 2020 Consumer Cluster Descriptions

Cluster 3 (“At-Risk”) is the cluster with the highest reported concern for coronavirus. 75.60% of respondents answered their concern as “Very Concerned” or “Extremely Concerned.” By looking at the age distribution, it became clear that their concern was based on the fact that they’re recognized as an at-risk population. This group, however, is not exclusive to the oldest category used. Across all age groups, the proportions that make up this cluster are higher than average for the entire data for ages 40 and older.

Cluster 4 “Known Risk” is the smallest group with 9 observations. This cluster is separated from others based on respondents having tested positively to a coronavirus test.

Cluster 5, “The Exposed”, is the most concerned about the coronavirus outbreak. These respondents are the most likely to be employed full-time or part-time. From the data set, around 55% of respondents ages 32 to 55 belong to this cluster. The high percentage, combined with the employment status, led me to believe that their perceived level of risk is raised by the nature and exposure to others at work.

Cluster 1, “The Controllers”, is also highly concentrated on the “middle age” group. 90% of its members are between the ages 32 and 55. However, this group has the highest proportion of contractors, or self-employed people. This includes a higher proportion of stay-at-home parents. Compared to Cluster 5, I assume that the lower risk relates to the nature of work. Yet, 76% of contractors stated they did not start working from home during the survey period. I

conclude that this is related to work nature that does not takes place in a single place, e.g. plumbers, a/c technicians, etc. For stay-at-home parents, diminished exposure to others seems to be the underlying factor. Clusters 5 and 1 were the most like each other. The nature of their work, as designated in the survey, appears to be the most significant separator.

Cluster 2, “The Relaxed”, is the group which showed the least concern for the coronavirus outbreak. It had a higher proportion of retirees compared to the entire data. This group has the lowest proportion of respondent who have social distanced when going outdoors; and, the highest proportion of respondents who’ve dined in a restaurant.

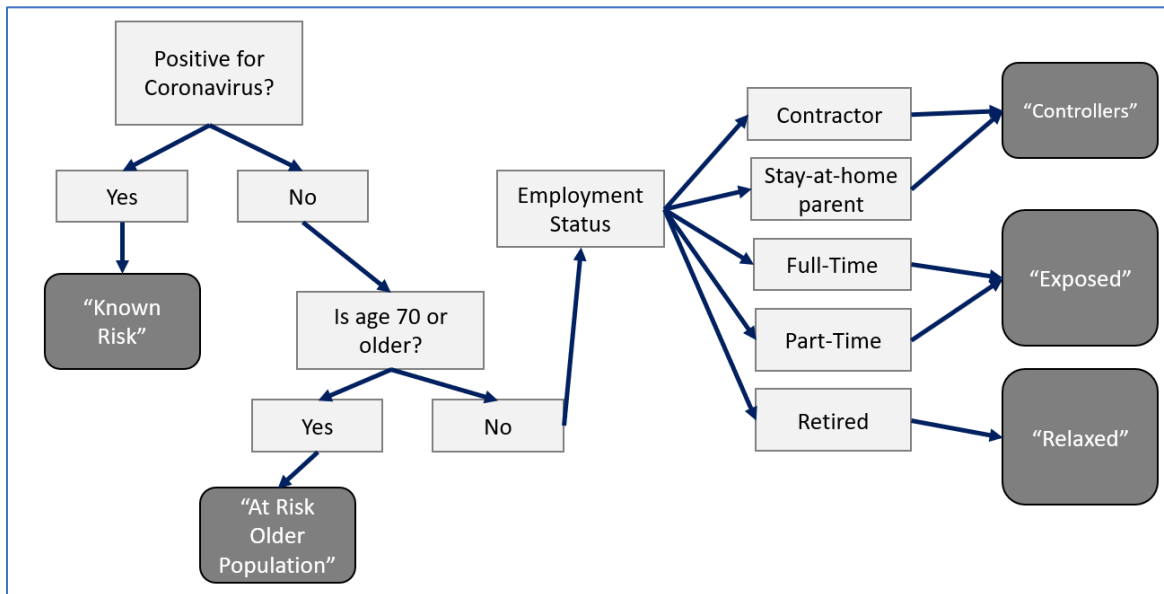


Figure 20: Pre-liminary flow to assign respondents to clusters

c. December 2020 Consumer Clusters

The variables selected for the algorithm in the December data were the same as those which were used in the final model for the June Data. I reviewed the algorithm results and settled on the same number of clusters as for the June data, five. As with the June data, the fact that a person tested positive for coronavirus affected their responses and all belong to one cluster.

For these clusters, age and familiarity with someone who had died due to coronavirus are the two most influential factors in determining to which cluster one person belongs.

Cluster	Description
Cluster 1: “Older Relaxed”	<ul style="list-style-type: none"> • Much more likely to be over age 56 • Much more likely to be a parent. • Least likely to be employed full time or part time. • 38% of unemployed respondents fall in this group.
Cluster 2: “Known Risk”	<ul style="list-style-type: none"> • Tested positive for Coronavirus. • Much more likely to know someone who tested positive for coronavirus (other than themselves).
Cluster 3: “Younger Relaxed”	<ul style="list-style-type: none"> • Much more likely to be a parent. • About 49% of respondents ages 18 to 24 belong to this group. • Least concerned about coronavirus
Cluster 4: “Younger Concerned”	<ul style="list-style-type: none"> • About 47% of respondents age 18 to 31 belong to this group. • Second most concerned cluster about coronavirus. • More likely to know someone who has died of coronavirus.
Cluster 5: “Older Concerned”	<ul style="list-style-type: none"> • Much more likely to be over age 56. • Much more likely to be a parent. • Highest concern for coronavirus. • More likely to know someone who has died of coronavirus.

Table 21: December Consumer Cluster Descriptions

For cluster 1, “Older Average Concern”, 85% of respondent in this cluster are 56 or older. At 23%, this cluster has the lowest level of respondents who are employed either full- or part-time. The concerned levels for coronavirus are in the middle of the pack, about 30% of respondents in this cluster indicated high concern for coronavirus. 43% indicated “somewhat” concerned.

Cluster 2 “Known Risk” is exclusively made up of people who tested positive for coronavirus. This group is also more likely to know someone who has tested positive than any other cluster.

Cluster 3 is the cluster least concerned about coronavirus. The “Younger Relaxed” was labeled as such based-on employment and household characteristics. This cluster has the highest concentration of households with 4 or more people at 38%. About 47% of respondents who are students fall under this cluster. Only 20% of respondents in this group stated they had a high level of concern about coronavirus.

In cluster 4, “Younger Concerned”, 61% of its respondents are between the ages 32 and 55. About 0.6% of respondents ages 56 or older belong to this group. On the other side of this separation point, 40% of those ages 40 to 55 belong to this group as well as around 48% of those 39 and younger. In the education demographic, this group has a higher concentration for people with Bachelor’s degrees or higher – 55% vs 41% for the entire data set. This is the second most concerned cluster about coronavirus – 91% of respondents stated they were “Very” or “Extremely” concerned about coronavirus.

Cluster 5 “Older and Concerned” 79% of people in this cluster are 56 years or older. As a result of this its proportion of retirees is the highest at 58% compared to 31% for the entire data

set. This group has the highest reported concern for coronavirus. More than 95% of respondents in this cluster indicated they were either “Very Concerned” or “Extremely Concerned” about coronavirus. This is also the largest cluster with 1100 observations or about 31% of the dataset.

For this period, and unlike the June 2020 clusters, an assignment based on demographics would not yield an assignment to five clusters. Instead, it would create only three: 1) “Known Risk” would remain as is based on coronavirus test results. Clusters 3 and 4, the “younger” clusters would merge. As would Clusters 1 and 5, the “older” clusters. The true separator between these four clusters is if they know someone who has died of coronavirus. Which I believe is the underlying factor in increasing the perception of risk.

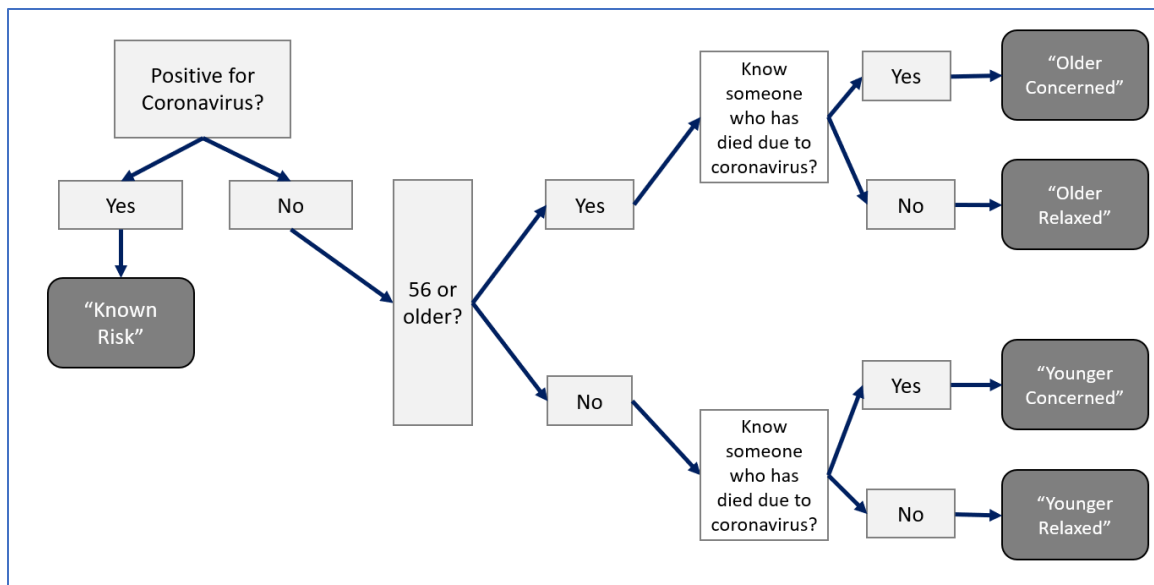


Figure 22: Concept Decision Tree for December Cluster Assignment

V. Cluster Analysis

a. June Clusters

Overall, going to the grocery store was the activity marked as the least risky in the survey data. This held true for the “Exposed” cluster. However, the portion of respondents who yet identified it as moderate or high risk was 69% compared to 45% for the entire dataset. For all other activities, at least 82% of respondents labeled them as moderately risky or highly risky.

The “Relaxed” cluster considered taking a vacation as the riskiest activity with 24.43% labeling it as representing moderate or high risk. However, this is far lower than the entire data set’s 71%.

For the other three clusters, the proportions were similar to the values for the entire data set.

Activities Represent Moderate or High Risk To Health						
	Controllers	Relaxed	At-Risk	Known Risk	Exposed	Total
Restaurants	70.49%	12.57%	86.18%	77.77%	90.82%	65.29%
Friends	60.66%	11.81%	78.49%	77.78%	82.65%	59.47%
Grocery	50.82%	6.29%	53.58%	66.67%	69.56%	44.83%
Retail Store	60.66%	8.57%	78.84%	66.67%	83.16%	58.73%
Salons / Barber	67.21%	10.67%	75.26%	66.67%	84.86%	58.96%
Vacation	77.05%	27.43%	88.91%	77.78%	91.67%	71.11%

Table 23: Activity Risk Levels in June Data

While the “Controllers” were the least concerned about coronavirus, their data indicates they’re the least comfortable engaging in certain activities. All other clusters were similar to the values for the entire data set.

Uncomfortable Engaging In These Activities						
	Controllers	Relaxed	At-Risk	Known Risk	Exposed	Total
Sports Event or Concert	70.98%	50.11%	72.45%	46.88%	37.17%	53.91%
Public Transportation	77.45%	52.63%	72.45%	53.44%	40.66%	56.07%
Visiting a Restaurant	54.10%	35.06%	47.81%	35.22%	28.72%	37.83%
Visiting Grocer	37.87%	18.27%	14.17%	20.33%	21.21%	18.58%
Visiting Friends	74.52%	26.49%	20.50%	31.33%	28.88%	26.98%
Going to Pub or Bar	76.29%	48.56%	69.13%	45.67%	36.79%	52.40%

Table 24: Cluster Uncomfortable Levels in Public Activities

b. December Clusters

When analyzing how risky the activities mentioned in the December survey, it became apparent that the two “concerned” clusters had values that separated them heavily from the other clusters. By comparison, the “Known Risk” cluster, composed of people who have tested positive for coronavirus has a mid-level concern. The “relaxed” clusters show the lowest concern levels. The “Younger Relaxed” has the lowest perceived level of risk across the activities of all clusters.

Activites Which Represent Moderate or High Risk to Health						
	Older Relaxed	Known Risk	Younger Relaxed	Younger Concerned	Older Concerned	Total
Restaurant	40.17%	61.54%	39.59%	94.56%	96.09%	70.65%
Friends	50.73%	63.64%	44.83%	96.01%	97.55%	74.95%
Normal Pre-Covid Life	47.82%	65.73%	41.92%	96.74%	94.36%	0.73%
Vaccine	55.46%	68.53%	47.16%	95.77%	98.36%	76.88%

Table 25: Cluster Perceived Risk of Activities

The survey asked respondents to indicate which activities they'd engaged in the previous week. Regarding 'preventive' actions, two activities were presented: self-quarantining, and social distancing while going outside. As with the perceived risk, the "concerned" clusters show a much higher probability of having engaged in social distancing. The "Known" risk is the cluster that is most likely to have self-quarantined. Two 'engaging with others' activities were presented by the survey: going out to eat; and, visiting friends. The three-way separation appears here again with the "relaxed" clusters more likely to report having engaged in either activity. And, the "concerned" clusters having lower participation rates.

Have Engaged in Activity in the Last Week						
	Older Relaxed	Known Risk	Younger Relaxed	Younger Concerned	Older Concerned	Total
Gone Out To Eat	39.44%	32.87%	48.33%	19.23%	15.18%	28.76%
Visited Friends	46.97%	38.46%	57.35%	28.90%	22.36%	36.89%
Self – Quarantined	14.32%	40.56%	10.19%	18.74%	25.18%	18.93%
Social Distanced	67.48%	79.02%	58.37%	92.74%	94.27%	80.26%

Table 26: Past Activities by Cluster

c. June and December Clusters Compared

The third goal of the study was to determine if there were shifts in the composition of the clusters between the two periods. While some demographic characteristics changed in importance for the clusters, overall, the clusters seem to be highly similar between June and December. Some characteristics were minimized, while others increased within each cluster. Four clusters appear to be closely related and have morphed by shifting characteristics. In particular for December, knowing someone who had died from coronavirus.

- The “Known Risk” cluster carried over. A higher portion of respondents of the data set belong to this group in the December data.
- The “At-Risk” became the “Older Concerned” with both groups characterized by older respondents.
- The “Exposed” became “Younger Concerned.” Employment type was eliminated as a key factor in cluster assignment.
- The “Relaxed” cluster in June appears to have become the “Older Relaxed.” The June group had a lot of members who were retired, and this is highly correlated with older respondents.
- The “Controllers” cluster from June appears to have morphed into “Younger Relaxed” while losing the employment type as a strong predictor of belonging to it. The June group had 90% of respondents between ages 30 and 55. The December cluster had a drop to about 60% within the same age range.

June Clusters	Portion	December Clusters	Portion
At-Risk	33.13%	Older Concerned	30.72%
Exposed	33.24%	Younger Concerned	23.09%
Relaxed	29.68%	Older Relaxed	23.01%
Controllers	3.45%	Younger Relaxed	19.18%
Known Risk	0.51%	Known Risk	3.99%

Table 27: Cluster Portions of Data Set

The June data showed that the clusters had two groups who were more concerned about coronavirus - “At-Risk” and “Exposed -. “Relaxed” cluster that appeared to be in the middle-of-the-road concern levels; “Controllers” were middle-of-the-road to concerned; and, “Known Risk” were balanced through all reported levels.

June	Not at all Concerned	Not Very Concerned	Somewhat Concerned	Very Concerned	Extremely Concerned
At-Risk	1%	2%	22%	39%	36%
Exposed	1%	2%	22%	39%	37%
Relaxed	14%	30%	44%	9%	3%
Controllers	7%	7%	31%	21%	34%
Known Risk	22%	0%	22%	11%	44%
Grand Total	5%	11%	29%	30%	26%

Table 28: Coronavirus Concern Levels by Cluster - June

For December, the number of respondents that indicated they were ‘somewhat concerned’ about coronavirus decreased, while the ‘extremely concerned’ increased. This shift also occurs among the clusters. In this period, the contrast between the groups increased. Specifically, the two “concerned” clusters had higher concern levels. The two “relaxed” clusters moved to ‘middle-of-the-road’ levels. And the known risk remained somewhat the same.

December	Not at all Concerned	Not Very Concerned	Somewhat Concerned	Very Concerned	Extremely Concerned
Older Concerned	0%	0%	5%	33%	63%
Younger Concerned	0%	0%	8%	40%	52%
Older Relaxed	6%	19%	43%	25%	6%
Younger Relaxed	12%	24%	44%	17%	3%
Known Risk	3%	10%	28%	33%	25%
Grand Total	4%	10%	23%	30%	34%

Table 29: Coronavirus Concern Levels by Cluster - December

VI. Discussion

Moving forward past gaining an understanding the attributes of the clusters, the focus is to identify the initiatives or measures to which each cluster will respond to engage in travel. The June data focused on which activities travel providers could take to reduce concerns surrounding travel. The December data was collected after the initial rounds of COVID-19 vaccines. This data included questions to gain sentiment information about the vaccination status of travelers.

The study identified five distinct cluster in each time period. All clusters shared three primary factors as motivators or incentives to travel and visit public spaces: 1) physical space management; 2) a cleaning program; 3) require that staff and public wear masks. The December clusters all had an approval of the use of a Covid passport.

The analysis revealed that there was a shift in the clusters from one period to the other. Across the board concerned levels for coronavirus increased. The changes in each cluster reduced the effect or influence of employment type. Familiarity with someone who had died from coronavirus became as a strong factor in determining cluster assignment. The number of people who knew someone who died from coronavirus increased from 1.4% to about 30%. The hypothesis is that in June, those who had control over their exposure to others felt in control –

primarily in the “Controller” cluster – as the number of cases increased, so did a reduction of control.

a. June Data

As presented, the questions asked respondents to indicate what they considered an absolute requirement to visit a business, and which activities they considered the most important that travel providers could take to reduce the perceived risk of contracting coronavirus. The main four attributes asked were: 1) The elimination of specific set of services to minimize or eliminate staff-consumer interactions; 2) Require travelers to disclose personal information in order to conduct contact tracing; 3) Managing the physical space on flights, hotels, or large gatherings to facilitate social distancing; 4) Launch company-certified cleaning and sterilization program for plane, hotel room, or rental car.

a(i). Health-Related Services

When selecting the top health-related service a travel provider could offer, there is a clear separation between the top two and the bottom two. Around 67% respondents selected launching a cleaning program as the top offering. Around 32% selected managing the physical space. Only 9% of respondents approved of a requirements that travelers disclose of personal information for the purpose of contact tracing.

	Eliminate Services	Require Persons to Disclose Personal Information for Contact Tracing	Manage or Rearrange Physical Space	Launch Health & Safety Cleaning Program
Controllers	15.86%	1.20%	22.73%	60.22%
Relaxed	2.44%	9.24%	28.77%	59.55%
At-Risk	3.66%	7.83%	44.62%	43.89%
Known Risk	2.57%	9.25%	23.65%	64.53%
Exposed	0.87%	9.55%	22.04%	67.54%
Grand Total	2.78%	8.60%	31.77%	56.86%

Table 30: Top Actions That Businesses Can Take by Cluster – June

The “Controllers” cluster selected elimination of services at a significantly higher rate than the other clusters. I hypothesize that this related to the nature of their employment as the cluster has a high concentration of contractors. Additionally, they support the disclosure of personal information at a much lower rate (1.2% compared to 8,6% for the entire data). This may stem from a belief that the responsibility and control lies on the business.

The “Exposed” and the “Controllers” clusters favor a shift away from managing physical space in preference of a cleaning program.

The “At-Risk” cluster is more closely balanced between managing physical space and cleaning program as the most important offering.

a(ii). Business Initiatives to Make Consumers Feel Safe

Respondents were asked to indicate which, if any, initiatives they would label as critical in engaging in business. The question was presented with the explanation: “I would not go to public places without this.” The options were, thus, presented as what respondents considered absolute requirements.

The Controllers valued having hygiene products for public use, as well as requiring both staff and public to wear masks higher than other clusters. This cluster labeled as critical having dedicated hours for at-risk populations at a higher rate than any cluster.

The “At-Risk” cluster indicated that having dedicated business hours for at-risk population was critical for the lowest percentage of any cluster. Other than requiring Staff to wear masks, this cluster generally labeled the initiatives as critical at a lower level than those for all respondents. Additionally, while other clusters have close numbers in how critical they perceive requiring both public and staff to wear masks is, this cluster has a higher value for staff (61%) compared to the requirement for the public (50%).

Those in the “Known Risk” cluster generally had an above-average reported percentage of members who categorized the various initiatives as critical.

With the exception of requirements to wear masks – for both public and staff – the “Exposed” cluster had higher than average percentages across the various initiatives.

		Clusters					
		Controllers	Relaxed	At-Risk	Known Risk	Exposed	Grand Total
Initiatives	Broader Contact Tracing Strategies	34.56%	29.20%	20.78%	35.33%	35.09%	28.59%
	Creating Dedicated Hours for At-Risk Populations	50.33%	31.35%	28.06%	35.67%	35.44%	32.30%
	Availability of Hygiene Products for Public Use	64.95%	46.27%	42.45%	48.11%	49.51%	46.74%
	Incentives (promotions)	27.00%	19.13%	14.58%	20.78%	22.09%	18.89%
	Restricting Number of Total People Allowed In	38.16%	46.98%	45.98%	43.44%	45.86%	45.95%
	Restricting Opening Hours	21.54%	30.07%	23.70%	34.11%	35.82%	29.60%
	Public Required To Wear Masks/Gloves	67.08%	49.11%	50.48%	53.67%	49.82%	50.28%
	Staff Required to Wear Masks/Gloves	70.43%	55.19%	61.37%	53.22%	50.61%	56.23%
	Temperature Checks on Entry	27.10%	26.35%	21.59%	30.67%	31.19%	26.43%
	Widely-Available Vaccine	27.74%	35.21%	34.85%	32.44%	34.52%	34.59%

Table 31: Critical Actions Required by Consumers by Cluster - June

b. December Data

The December data introduced the question of approval of a Covid Passport for international travel. Additionally, respondents were asked to indicate if they had any travel plans in the next three months (from time of the survey).

When answering the question about likely hood to travel in the next three months, age is the differentiating factor. The “older” clusters had a higher proportion of respondents who answered they were “unlikely to travel” in the next three months. On the “Likely” and “Very Likely” answers, the “Younger” clusters had higher proportions than the entire data set.

All clusters indicated support for a Covid passport at around the same value of 70% for the entire data set.

	Unlikely to Travel	Likely To Travel	Very Likely	Trip Already Planned	Approve Covid Passport
Older Relaxed	55.69%	18.33%	20.99%	9.83%	68.92%
Known Risk	42.80%	23.27%	29.91%	13.94%	70.43%
Younger Relaxed	32.39%	26.59%	37.33%	17.24%	71.78%
Younger Concerned	32.30%	26.79%	37.36%	17.18%	71.91%
Older Concerned	56.58%	17.88%	20.61%	9.62%	69.49%
Grand Total	45.57%	21.93%	28.14%	13.05%	70.40%

Table 32: Travel Intentions and Approval of Covid Passport by Cluster

Going Forward

All consumer groups valued the implementation or creation of a cleaning and sterilization programs in businesses. With no other information available to travel and hospitality businesses, this would be the number one recommendation as to which actions, they can take. As seen in the June data, the second most common initiatives businesses could take were: restriction of people; and, requiring masks/gloves to be worn.

In June, the third factor that would make people feel safe when traveling or visiting public places was having a vaccine widely available. The creation, and use, of a Covid passport has an approval of 70% for a Covid passport in the December data; compared to only an 8.6% approval of requiring people to provide personal information for contact tracing in June. Thus, a requirement of proof of vaccination can increase traveler engagement.

Focusing on travel, the “younger” December clusters had a higher intention or reported confirmed travel than the “older” clusters. Even when considering whether they fell under the “relaxed” or the “concerned” clusters.

In May 2020, according to the CDC (2021), there were an average of 24,325 new reported cases per day. That number climbed to 211,178 during December 2020. A year later, for June 2021 the average reported of new daily cases was 12,773. With number of reported cases dropping, it would be valuable to re-analyze the data to gather consumer sentiment.

“Widely-Available Vaccine” was among the top three factors influencing traveler’s decisions for future travel. Future research can focus on how this sentiment has changed or remained now that the vaccine has been approved.