TEAM PROJECT 2

Course: Customer Analysis

Dr. Alan Cooke

Date: February 3, 2025

Abstract

We analyzed both data sets and had insights on them. In the report, we will cluster consumers into different groups and choose one to target on.

Team3 – New Bees

Anni Liu, Lu Huang, Jiayin Yu, Ting Xia

Executive Summary

By performing PCA on the psychographic survey, we identified four key factors that capture a substantial portion of the data's variance. Based on these four key factors, we developed strategic recommendations to better serve our customers by combining marketing plans and product offerings with insights gained from each factor. Based on the second dataset, we group all the consumers into 6 clusters and analyze their general characteristics and their needs related to dining at restaurants. And we decided to target on cluster 4, which we name it “luxury gourmet”. We hope that we will make more profits by targeting high-end consumers.

Insights on Data set TP2_PCA

First, to get some basic information from the psychographic survey data, we use R to summary the mean and variation (SD). Based on the information in chart Figure 1 (view in Appendix), we notice:

- The mean of Q1-Q15 is almost equal to 4, and their median, although not included in the figure, is estimated to be almost equal to 4 based on Pctl. 25 and Pctl. 75. Since the mean is almost equal to the median, so the data may be symmetrically distributed (close to normal).
- The Min (range 1-2) and Max (range 6-7) of Q1-Q15 are basically symmetrical about the mean 4, so the data may be normally distributed.

Based on the above found, the data is normally distributed. However, to ensure the certainty of the data, we draw a normal distribution diagram (Figure 2), which confirms that the data is roughly normally distributed.

Next, we conduct a correlation analysis. According to the results in the chart (Figure 3 and Figure 4), from the result, we can get information:

- The correlation scores between Q4 (fresh ingredients), Q9 (special food), Q7 (diverse menu) and Q13 (limited time special dishes) are all positive and are mostly greater than 0.5, indicating that there is a strong correlation between these questions. From the strong correlation of this group of questions, we speculate that these questions may constitute a potential component. This may indicate that customers value fresh ingredients as much as a varied menu.
- The correlation scores between Q3 (dislike small portions), Q6 (good value) and Q12 (reasonable prices) are all positive and greater than 0.5, indicating that there is a strong correlation between them. These questions may constitute a potential component for customers to pursue a reasonable balance between food price and portion.
- There is a positive correlation between Q15 (attentive staff) and Q14 (servers make me feel special). Q1 (Prefer romantic, secluded seating) has a negative correlation with Q15 (attentive staff) and Q14 (servers make me feel special). They may form a potential component, where customers who value customer service like to be taken care of by servers, while customers who prefer secluded seating do not want to be disturbed by servers.

And we also get the PCA factor loading plot (Figure 5), which showed us very similar result we get from the correlation matrix. Two major problem categories are

- Q4, Q,7, Q9 and Q13, which shows that customers value food quality and food uniqueness.
- Q14, Q15, Q1, which indicates customers' views on atmosphere and service experience

Then we turned to Barlett’s Sphericity and KMO test. According to the Bartlett (Figure 6) and KMO test (Figure 7), the data are appropriate for PCA/FA. Based on the R output, Bartlett’s p-value is 2.22e-16 (<0.05), which indicates an extremely low probability that the null hypothesis is true; therefore, it is strongly proven that datasets are correlated enough to support factor analysis or PCA. And for the KMO test, an overall MSA is 0.66, which indicates moderate adequacy. It means the data is marginally acceptable for factor analysis. And for MSA for each item, Q4, Q7, Q9, Q10 and Q13 have a high MSA score that is bigger than 0.7, which shows that these variables contribute well to the factor analysis. And

for Q5, Q8 and Q11, since their MSA scores are less than 0.50, which indicate these variables may not fit well in the factor structure and could be candidates for removal.

Finally, we start with the PCA. First, we examined the eigenvalues (Figure 8) for each factor. The first four factors have eigenvalues of 3.42, 2.42, 2.14, and 1.44, all exceeding 1. This meets the Kaiser Criterion, which recommends retaining factors with eigenvalues greater than 1. Collectively, these four factors explain 50.7% of the total variance, indicating that they capture a substantial portion of the information in the data.

Next, we conducted a chi-square test to evaluate the model fit ($\chi 2(51) = 61.22$, $p=0.155$\chi^2(51) = 61.22, p = 0.155$\chi 2(51) = 61.22, p=0.155$). Since $p>0.05$p > 0.05$p>0.05$, we fail to reject the null hypothesis, suggesting that the 4-factor solution adequately represents the underlying structure of the data without significant information loss.

In conclusion, we should extract four factors based on the Kaiser Criterion and the chi-square test.

By comparing the component matrix and rotated component matrix (Figure 9), we found there are some similarities and differences between these two matrices:

- **Similarities:** In both the component matrix and rotated component matrix, many questions have very similar factor loadings, especially the loadings for each factor. For example, Q4 and Q7 both show high loadings on Factor 1 in both the component matrix and rotated matrix, indicating that these two questions strongly load onto Factor 1.
- **Differences:** In the component matrix, Q12 and Q15 have loadings on multiple factors, while in the rotated component matrix, their loadings are adjusted and concentrated on one factor. For example, Q12 has loadings on Factor 3 (0.779 and Factor 4 (-0.186) in the component matrix, but in the rotated matrix, Q12 has a clearer loading on Factor 3 (0.802). This shows that the rotation helps to more clearly assign Q12 to Factor 3. Other than that, Q15 only has a loading of 0.761 on Factor 4 in the rotated component matrix, but in the component matrix, Q15 loads on both Factor 2 (0.733) and Factor 4 (-0.152). The rotation makes its loading more concentrated on Factor 4.

Now we can explain which questions load on which particular factors in detail:

- **Factor 1:** Strongly associated with Q4, Q7, Q9, and Q13 — likely related to variety, specialty dishes, and freshness of ingredients.
- **Factor 2:** Dominated by Q14 and Q15, suggesting customer service and attentive staff.
- **Factor 3:** Includes strong loadings from Q3, Q6, and Q12, possibly associated with portion size, reflecting value and price.
- **Factor 4:** Includes Q2 and Q3, which might be linked to distance and portion size.

Based on the rotated component matrix, we found there are some negative loadings that need to be reversed:

- Q1 shows a negative loading on Factor2 (-0.675). This strong negative loading suggests that customers valuing romantic seating may oppose characteristics captured by Factor 2 (Customers who prefer secluded seating may not like being disturbed by waiters).
- Q10 also shows a negative loading on Factor 2(-0.468). The negative loading indicates that customers who prioritize scenic views may not value aspects linked to Factor 2.
- Q14 has negative loading on Factor1(-0.102). Although this is a small negative value, it may indicate subtle differences in customer preferences for attentiveness versus independence.

After running the rotated component matrix, we can clearly see which one question contributes to which factor, we now give each factor a meaningful interpretation to each factor based on the content of the questions.

- **Factor 1: Food Quality and Diversity.** This factor emphasizes the overall sophistication and quality of dining, including ingredient freshness, diverse menu options, and exclusive offerings. Customers valuing this factor seek elevated, high-quality dining experience.
- **Factor 2: Service and Atmosphere Preferences.** This factor focuses on service attentiveness and practical dining preferences. The negative loadings on Q1 and Q10 indicate that some respondents may prefer functional dining over a romantic or scenic experience.
- **Factor 3: Price and portion preferences.** This factor captures customers who are value-conscious and care about adequate portion sizes and fair pricing. These individuals want to feel they've received a good deal when dining out.
- **Factor 4: Convenience and Casual Atmosphere.** This factor highlights customer preferences for convenience and casual, relaxed dining environments

Based on the psychological factors we obtained, we think our restaurant should make the following improvements:

- Add more menu options while ensuring food quality. Seasonal or limited-time special dishes can be provided to enhance the customer's dining experience
- Train waiters to provide different services to customers who have different requirements for a dining atmosphere.
- Offer workday/weekend packages, loyalty programs or promotions to attract price-sensitive customers.
- Restaurants should not only create an overly refined dining environment but also provide customers with a more cozy and relaxing dining place.

Insights on Data set TP2_CAdata

The variables that we are going to enter into the cluster analysis are the ones that are directly related to consumers' needs and benefits. We standardize our data because variables are on different scales. For example, Lunch frequency is 1-5, but Atmosphere Importance is 1-7. The implications of standardizing data are equal weighting, improving comparability, enhancing Performance of Distance-Based Algorithms, normalizing units, and Improving Convergence.

Here is the list of variables and their variations (figure 20 and variable list) that we include in our cluster analysis:

> Based on the R output, we can see that Most variables have a wide range (from 1 to 7 or 8), indicating significant variability. AtmosImp and FoodImp have higher means, suggesting customers value ambiance and food highly. Lunch and Dinner frequencies are moderate, with DinnerFreq slightly higher. NumInParty shows variability in dining group sizes.

Model of Fit: We run the silhouette model on clusters of 5 and 6 to check the model of fit. According to the R outputs, the average silhouette for 5 clusters is 0.5 and the average silhouette score for 6 clusters is 0.62. We decided to choose 6 clusters because its silhouette score is the closest to 1. According to the silhouette graph with 6 clusters, cluster 5 has the highest score which exceeds 0.75 and shows a good fit.

Dendrogram: The higher the branch, the greater the dissimilarity between clusters being merged. The shorter the branch, the more similar the merged data points are. Based on the dendrogram, you can see the 6 clusters are formed at distance=4, and there is a huge distance of clusters jumping from 8 to 4. At distance 4, there are 6 lines. Therefore, the dendrogram indicates that 6 is an appropriate number of clusters. In addition, based on the elbow chart(figure11), the diminish of return on distance occurs after 6 clusters are formed.

Euclidean Distance: The Euclidean distance matrix provides a measure of how similar or dissimilar the observations are based on their features. According to the R output: there are smaller distances (e.g., 0.48, 0.85): These indicate that the corresponding points (rows) are closer together in the feature space,

meaning they are more similar. And there are larger distances (e.g., 3.7, 4.59): These suggest that the corresponding observations are far apart, meaning they differ significantly.

Here are the implications on clustering based on Euclidean Distance:

- Observations 1 and 2 have a very small distance (0.48), meaning they are likely in the same cluster.
- Observations 3 and 6 have much higher distances (e.g., 3.7, 4.59), indicating they belong to different clusters.
- Observations 4 and 5 have low distances (0.85), suggesting strong similarity.
- Some observations have consistently large distances (e.g., observation 6 with 3.5+ values), meaning that 6 may be a good number of clusters because the distance increases at cluster number =6.

Above all, **choosing 6 clusters results in clearer separation and higher clustering validity**, namely the following groups：(General Descriptions on each segment figure 17-18)

Cluster 1- Moderate Males
Summary: This cluster consists of slightly younger customers with a higher percentage of males, slightly higher income, moderate TV and web usage, but lower exposure to out-of-home ads, radio, and social media.
Cluster 2- TV Enthusiasts
Summary: This cluster has an equal gender distribution, significantly lower income, high TV watching hours, but low web usage, out-of-home ads exposure, radio usage, and social media usage.
Cluster 3 – Social Media Savvy
Summary: This cluster has a lower percentage of males, lower income, lower TV watching hours, and higher social media usage.
Cluster 4 – High Earners
Summary: This cluster has a lower percentage of males, significantly higher income, high web usage, out-of-home ads exposure, and radio usage.
Cluster 5 – Balanced Users
Summary: This cluster has an equal gender distribution, average income, moderate TV watching hours, and higher radio usage.
Cluster 6 – Selective Consumers
Summary: This cluster has a lower percentage of males, higher income, low TV watching hours, and higher out-of-home ads exposure.

Insights on R Output - Profile by variables that are related to restaurants (Needs in the product categories – figure 19)：

Cluster 1: "Service Seekers" -- They value good service but don't prioritize food or atmosphere
Description: This cluster consists of customers who place moderate importance on service but lower importance on atmosphere and food. They tend to have below-average spending and visit frequencies, both for lunch and dinner. They sure don't want to encounter the situation when they complain to the waiter about flies in their soups and the server replies back: "Don't worry, folks. The spiders in your salad will take care of it!"
Cluster 2: "Casual Diners" —— Dining is not a big deal for them; they eat out occasionally with minimal spending
Description: This cluster includes customers who place low importance on all aspects (atmosphere, food, and service) and have low visit frequencies for both lunch and dinner. They also tend to spend less and dine in smaller groups.
Cluster 3: "Social Foodies" --- They love eating out in groups and enjoy the experience more than the food itself

Description: This cluster consists of customers who place above-average importance on atmosphere and have moderate visit frequencies for lunch and dinner. They tend to spend slightly less and dine in larger groups.

Cluster 4: "Luxury Gourmets" —These are high-end diners who want the best of everything—food, service, and ambiance

Description: This cluster includes customers who place high importance on all aspects (atmosphere, food, and service) and have high visit frequencies for both lunch and dinner. They tend to spend more and dine in larger groups.

Cluster 5: "Selective Tasters" —— They are food connoisseurs, picky about quality but don't dine out often

Description: This cluster consists of customers who place high importance on food and atmosphere but have low visit frequencies for both lunch and dinner. They tend to spend more but dine in smaller groups.

Cluster 6: "Fast & Frequent" —— They eat out a lot but mainly for convenience, not for the experience

Description: This cluster includes customers who have high visit frequencies for both lunch and dinner but place low importance on atmosphere, food, and service. They tend to spend less and dine in smaller groups.

Targeting Strategy

We decide our Target Market is Cluster 4 - "Luxury Gourmets". This is the ideal target, representing **20% of the total market (40 out of 200 customers)** with **high spending power and frequent dining habits**. This group prioritizes **premium food (Avg_FoodImp: 1.21), ambiance (Avg_Atoms_Imp: 0.79), and service quality (Avg_ServIMP: 1.28)**, making them ideal customers for a **high-end dining experience**.

There is a joke about a diner who is a Luxury Gourmet ordering dry-aged Wagyu steak, seared with Himalayan salt and topped with white truffle shavings. When the waiter asks how they'd like it cooked, the diner simply replies, "Expensive." This perfectly illustrates how luxury gourmets willingly pay more for perceived exclusivity, proving that with the right branding, the same dish can command a significantly higher price.

To focus on this segment, our restaurant should be positioned as a **luxury fine-dining destination** offering **gourmet dishes, premium wine pairings, and an elegant ambiance**. Exclusive **private dining, chef's table experiences, and personalized service** will enhance appeal.

We will adopt the following strategy:

- VIP Memberships & Loyalty Programs → **Exclusive dining perks for high-value customers.**
- Private Events & Tastings → **Michelin-style menus, wine pairings, and invitation-only experiences.**
- Luxury Brand Collaborations → **Partner with premium brands for co-branded promotions.**
- Targeted Digital & Influencer Marketing → **Engage** food critics, lifestyle influencers, and high-net-worth customers **through curated content.**

Conclusion

We group all customers into 6 clusters, according to their needs and behavior named as "Service Seekers", "Casual Diners", "Social Foodies", "Luxury Gourmets", "Selective Tasters" and "Fast & Frequent".

Luxury Gourmets are the best segment for maximizing **revenue, exclusivity, and brand prestige**. By focusing on **high-quality experiences and strategic marketing**, the restaurant can establish itself as a premier fine-dining destination.

Appendix

Segments Profile:

Insights on R output - General Descriptions on each segment: (figure 17-18)

Cluster 1- Moderate Males

- **Gender Distribution**: 54% Male
- **Avg_Income**: Slightly above the overall mean income
- **Avg_TV_Hours**: Slightly above the overall mean TV watching hours
- **Avg_WEB**: Slightly above the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Slightly below the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Below the overall mean radio usage frequency
- **Avg_social_media_freq**: Significantly below the overall mean social media usage frequency
- **Avg_Numkids**: Slightly above the overall mean number of kids
- **Avg_race**: Average race value (2.50)

Summary: This cluster consists of slightly younger customers with a higher percentage of males, slightly higher income, moderate TV and web usage, but lower exposure to out-of-home ads, radio, and social media.

Cluster 2- TV Enthusiasts

- **Gender Distribution**: 50% Male
- **Avg_Income**: Significantly below the overall mean income
- **Avg_TV_Hours**: Significantly above the overall mean TV watching hours
- **Avg_WEB**: Significantly below the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Significantly below the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Slightly below the overall mean radio usage frequency
- **Avg_social_media_freq**: Significantly below the overall mean social media usage frequency
- **Avg_Numkids**: Slightly below the overall mean number of kids
- **Avg_race**: Average race value (2.45)

Summary: This cluster has an equal gender distribution, significantly lower income, high TV watching hours, but low web usage, out-of-home ads exposure, radio usage, and social media usage.

Cluster 3 – Social Media Savvy

- **Gender Distribution**: 40% Male
- **Avg_Income**: Below the overall mean income
- **Avg_TV_Hours**: Below the overall mean TV watching hours
- **Avg_WEB**: Slightly below the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Below the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Below the overall mean radio usage frequency
- **Avg_social_media_freq**: Significantly above the overall mean social media usage frequency
- **Avg_Numkids**: Below the overall mean number of kids
- **Avg_race**: Average race value (2.45)

Summary: This cluster has a lower percentage of males, lower income, lower TV watching hours, and higher social media usage.

Cluster 4 – High Earners

- **Gender Distribution**: 37.5% Male
- **Avg_Income**: Significantly above the overall mean income
- **Avg_TV_Hours**: Below the overall mean TV watching hours
- **Avg_WEB**: Significantly above the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Significantly above the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Significantly above the overall mean radio usage frequency
- **Avg_social_media_freq**: Slightly below the overall mean social media usage frequency
- **Avg_Numkids**: Slightly above the overall mean number of kids
- **Avg_race**: Average race value (2.45)

Summary: This cluster has a lower percentage of males, significantly higher income, high web usage, out-of-home ads exposure, and radio usage.

Cluster 5 – Balanced Users

- **Gender Distribution**: 50% Male
- **Avg_Income**: Close to the overall mean income
- **Avg_TV_Hours**: Slightly above the overall mean TV watching hours
- **Avg_WEB**: Below the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Slightly below the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Significantly above the overall mean radio usage frequency
- **Avg_social_media_freq**: Slightly above the overall mean social media usage frequency
- **Avg_Numkids**: Slightly below the overall mean number of kids
- **Avg_race**: Average race value (2.45)

Summary: This cluster has an equal gender distribution, average income, moderate TV watching hours, and higher radio usage.

Cluster 6 – Selective Consumers

- **Gender Distribution**: 40% Male
- **Avg_Income**: Above the overall mean income
- **Avg_TV_Hours**: Significantly below the overall mean TV watching hours
- **Avg_WEB**: Slightly below the overall mean web usage frequency
- **Avg_Out_of_Home_ads**: Above the overall mean exposure to out-of-home ads
- **Avg_Radio_freq**: Slightly below the overall mean radio usage frequency
- **Avg_social_media_freq**: Significantly below the overall mean social media usage frequency
- **Avg_Numkids**: Below the overall mean number of kids
- **Avg_race**: Average race value (2.30)

Summary: This cluster has a lower percentage of males, higher income, low TV watching hours, and higher out-of-home ads exposure.

Insights on R Output - Profile by variables that are related to restaurants (Needs in the product categories – figure 19)

Cluster 1: "Service Seekers" -- They value good service but don't prioritize food or atmosphere

- **Avg_Atoms_Imp**: -1.1026235 (Below average importance on atmosphere)
- **Avg_Cost**: -0.7319202 (Below average cost)
- **Avg_LunchFreq**: -1.0051231 (Below average lunch frequency)
- **Avg_DinnerFreq**: -0.09432110 (Slightly below average dinner frequency)
- **Avg_FoodImp**: -0.2300342 (Slightly below average importance on food)
- **Avg_ServIMP**: 0.6483539 (Above average importance on service)
- **Avg_NumInParty**: -0.05072167 (Slightly below average number in party)

Description: This cluster consists of customers who place moderate importance on service but lower importance on atmosphere and food. They tend to have below-average spending and visit frequencies, both for lunch and dinner. They sure don't want to encounter the situation when they complain to the waiter about flies in their soups and the server replies back: "Don't worry, folks. The spiders in your salad will take care of it!"

Cluster 2: "Casual Diners" —— Dining is not a big deal for them; they eat out occasionally with minimal spending

- **Avg_Atoms_Imp**: -1.4459843 (Significantly below average importance on atmosphere)
- **Avg_Cost**: -0.6719969 (Below average cost)
- **Avg_LunchFreq**: -0.9333286 (Below average lunch frequency)
- **Avg_DinnerFreq**: -1.81195804 (Significantly below average dinner frequency)
- **Avg_FoodImp**: -2.0922161 (Significantly below average importance on food)
- **Avg_ServIMP**: -1.8271791 (Significantly below average importance on service)
- **Avg_NumInParty**: -1.24630381 (Significantly below average number in party)

Description: This cluster includes customers who place low importance on all aspects (atmosphere, food, and service) and have low visit frequencies for both lunch and dinner. They also tend to spend less and dine in smaller groups.

Cluster 3: "Social Foodies" --- They love eating out in groups and enjoy the experience more than the food itself

- **Avg_Atoms_Imp**: 0.7947271 (Above average importance on atmosphere)
- **Avg_Cost**: -0.3010432 (Slightly below average cost)
- **Avg_LunchFreq**: 0.4626757 (Above average lunch frequency)
- **Avg_DinnerFreq**: 0.04136891 (Slightly above average dinner frequency)
- **Avg_FoodImp**: -0.2300342 (Slightly below average importance on food)
- **Avg_ServIMP**: -0.6265238 (Below average importance on service)
- **Avg_NumInParty**: 0.80671603 (Above average number in party)

Description: This cluster consists of customers who place above-average importance on atmosphere and have moderate visit frequencies for lunch and dinner. They tend to spend slightly less and dine in larger groups.

Cluster 4: "Luxury Gourmets" —These are high-end diners who want the best of everything—food, service, and ambiance

- **Avg_Atoms_Imp**: 0.7906971 (Above average importance on atmosphere)

- **Avg_Cost**: 1.4681207 (Significantly above average cost)
- **Avg_LunchFreq**: 1.2205066 (Significantly above average lunch frequency)
- **Avg_DinnerFreq**: 1.06731775 (Significantly above average dinner frequency)
- **Avg_FoodImp**: 1.2122439 (Significantly above average importance on food)
- **Avg_ServIMP**: 1.2836097 (Significantly above average importance on service)
- **Avg_NumInParty**: 0.69198845 (Above average number in party)

Description: This cluster includes customers who place high importance on all aspects (atmosphere, food, and service) and have high visit frequencies for both lunch and dinner. They tend to spend more and dine in larger groups.

Cluster 5: "Selective Tasters" —— They are food connoisseurs, picky about quality but don't dine out often

- **Avg_Atoms_Imp**: 0.7786069 (Above average importance on atmosphere)
- **Avg_Cost**: 0.7832831 (Above average cost)
- **Avg_LunchFreq**: -1.0131003 (Below average lunch frequency)
- **Avg_DinnerFreq**: -1.06731775 (Below average dinner frequency)
- **Avg_FoodImp**: 1.1209604 (Significantly above average importance on food)
- **Avg_ServIMP**: -0.0916864 (Slightly below average importance on service)
- **Avg_NumInParty**: -1.46368238 (Significantly below average number in party)

Description: This cluster consists of customers who place high importance on food and atmosphere but have low visit frequencies for both lunch and dinner. They tend to spend more but dine in smaller groups.

Cluster 6: "Fast & Frequent" —— They eat out a lot but mainly for convenience, not for the experience

- **Avg_Atoms_Imp**: -1.0832792 (Below average importance on atmosphere)
- **Avg_Cost**: -0.6291946 (Below average cost)
- **Avg_LunchFreq**: 1.2603925 (Significantly above average lunch frequency)
- **Avg_DinnerFreq**: 1.71267267 (Significantly above average dinner frequency)
- **Avg_FoodImp**: -0.3760877 (Below average importance on food)
- **Avg_ServIMP**: -0.7793344 (Below average importance on service)
- **Avg_NumInParty**: -1.93466928 (Significantly below average number in party)

Description: This cluster includes customers who have high visit frequencies for both lunch and dinner but place low importance on atmosphere, food, and service. They tend to spend less and dine in smaller groups.

sumtable {vtable}

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Q1	100	4	1.1	2	3	5	7
Q2	100	3.9	2.1	1	2	6	7
Q3	100	4.1	1	2	3	5	7
Q4	100	4	1.2	1	3	5	6
Q5	100	4	2.1	1	2	6	7
Q6	100	4.1	1	1	4	5	7
Q7	100	4	1.3	1	3	5	7
Q8	100	4	2.1	1	2	6	7
Q9	100	4	1.2	1	3	5	7
Q10	100	4	1.2	2	3	5	7
Q11	100	3.9	1.9	1	3	6	7
Q12	100	4	1.1	1	3	5	6
Q13	100	4	1.1	2	3	5	7
Q14	100	3.9	1.1	1	3	5	6
Q15	100	3.9	1	1	3	5	6

Figure 1

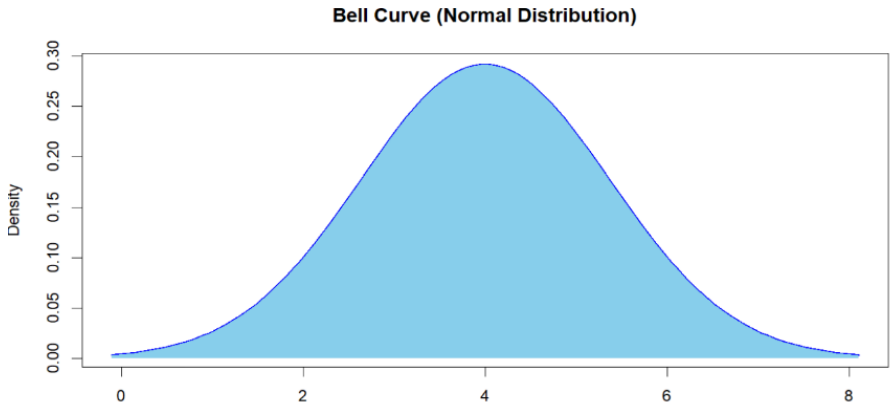

Figure 2

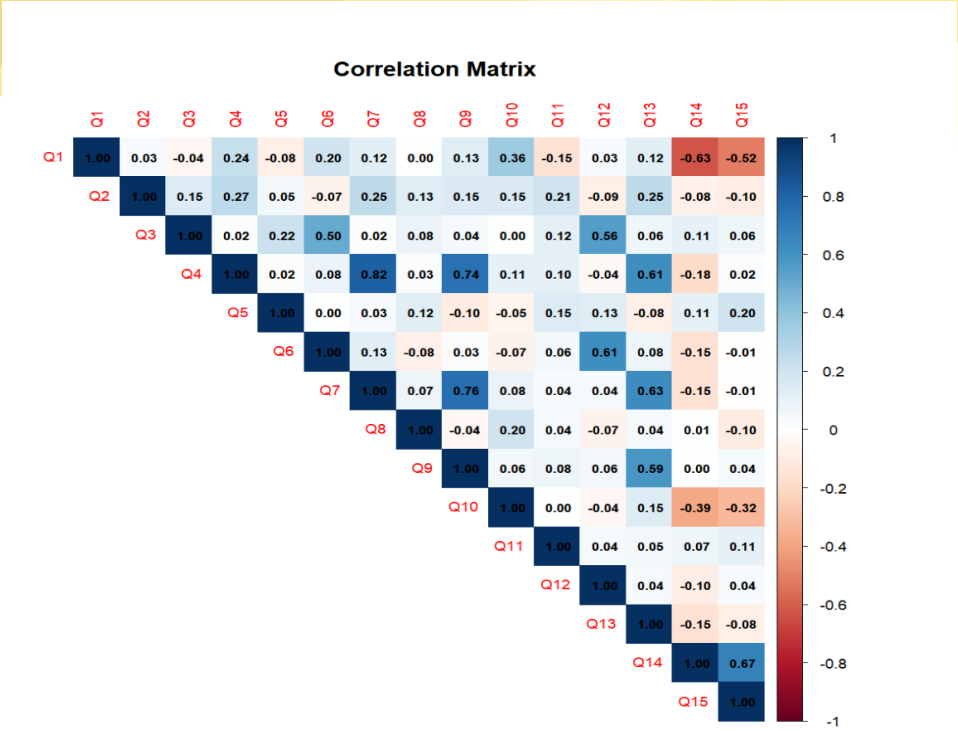

Figure 3

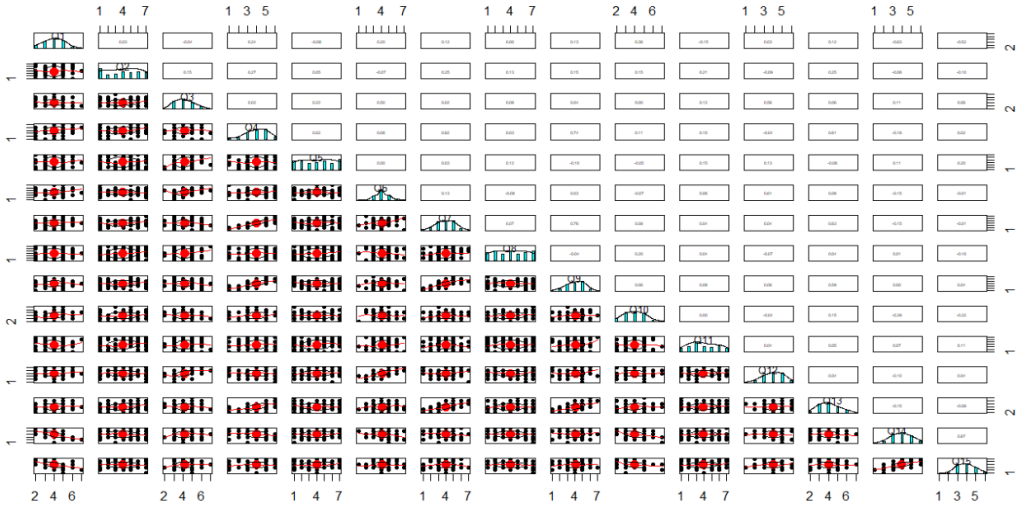

Figure 4

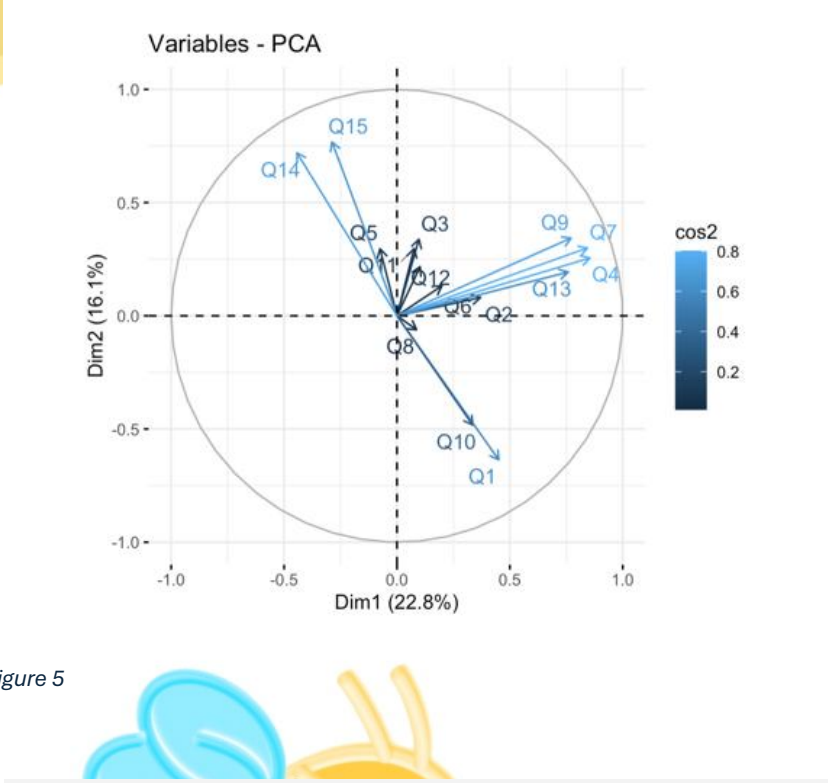

Figure 5

```
        Bartlett test of homogeneity of variances

data:  question_factors
Bartlett's K-squared = 247.07, df = 14, p-value < 2.2e-16
```

Figure 6

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cor(question_data))
Overall MSA =  0.66
MSA for each item =
  Q1   Q2   Q3   Q4   Q5   Q6   Q7   Q8   Q9  Q10  Q11  Q12
Q13  Q14
0.65 0.66 0.57 0.75 0.38 0.54 0.71 0.46 0.70 0.75 0.45 0.58
0.92 0.61
 Q15
0.63
```

Figure 7

```r
## Determine Number of Factors
question_factors <- select(data, -CustID)
eigen(cor(question_factors))$values
```

```
 [1] 3.4217043 2.4180019 2.1405423 1.4362091 0.9707848
0.8914474
 [7] 0.7524339 0.6824267 0.5345770 0.4520333 0.4138935
0.3359166
[13] 0.2632684 0.1509373 0.1358237
```

Figure 8

```
Call:
factanal(x = question_factors, factors = 4)

Uniquenesses:
   Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    Q9   Q10
Q11
0.502 0.695 0.275 0.178 0.912 0.395 0.179 0.863 0.294 0.725
0.931
  Q12   Q13   Q14   Q15
0.353 0.513 0.164 0.415

Loadings:
    Factor1 Factor2 Factor3 Factor4
Q1   0.149  -0.675          -0.117
Q2   0.227  -0.130           0.482
Q3                  0.727   0.433
Q4   0.895  -0.108
Q5           0.130   0.140   0.226
Q6                  0.760  -0.111
Q7   0.900
Q8                          0.354
Q9   0.837
Q10         -0.468          0.219
Q11                         0.229
Q12                 0.802
Q13  0.672  -0.119          0.138
Q14 -0.102   0.895          0.142
Q15          0.761

               Factor1 Factor2 Factor3 Factor4
SS loadings      2.870   2.152   1.805   0.780
Proportion Var   0.191   0.143   0.120   0.052
Cumulative Var   0.191   0.335   0.455   0.507

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 61.22 on 51 degrees of freedom.
The p-value is 0.155
```

```
Loadings:
    Factor1 Factor2 Factor3 Factor4
Q1   0.330  -0.622
Q2   0.284                  0.473
Q3           0.136   0.789   0.286
Q4   0.891   0.154
Q5           0.151   0.180   0.179
Q6   0.157  -0.102   0.719  -0.231
Q7   0.885   0.188
Q8                          0.367
Q9   0.788   0.281
Q10  0.199  -0.389           0.286
Q11          0.142           0.191
Q12                  0.779  -0.186
Q13  0.688
Q14 -0.343   0.847
Q15 -0.152   0.733          -0.150

               Factor1 Factor2 Factor3 Factor4
SS loadings      3.078   2.012   1.801   0.716
Proportion Var   0.205   0.134   0.120   0.048
Cumulative Var   0.205   0.339   0.459   0.507
```

Figure 9- left: with rotation; right: without rotation

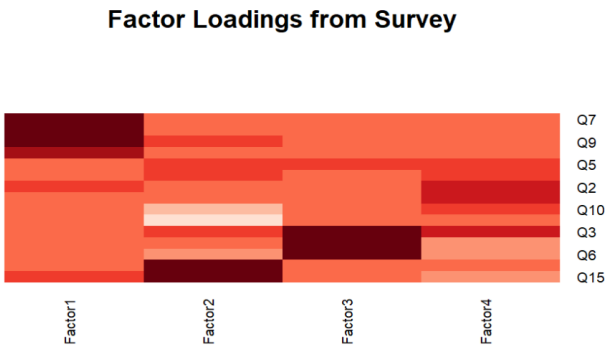

Figure10

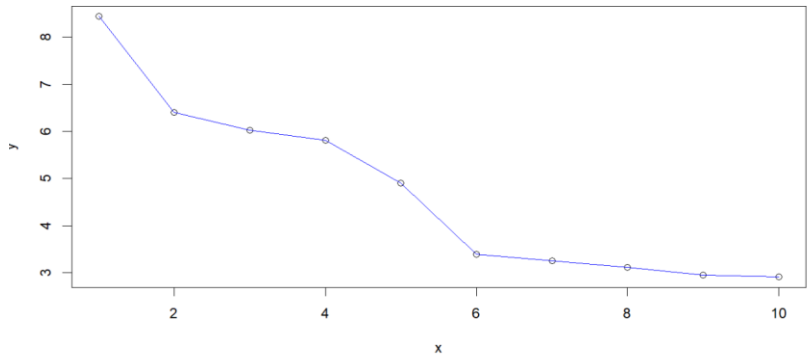

Figure 11

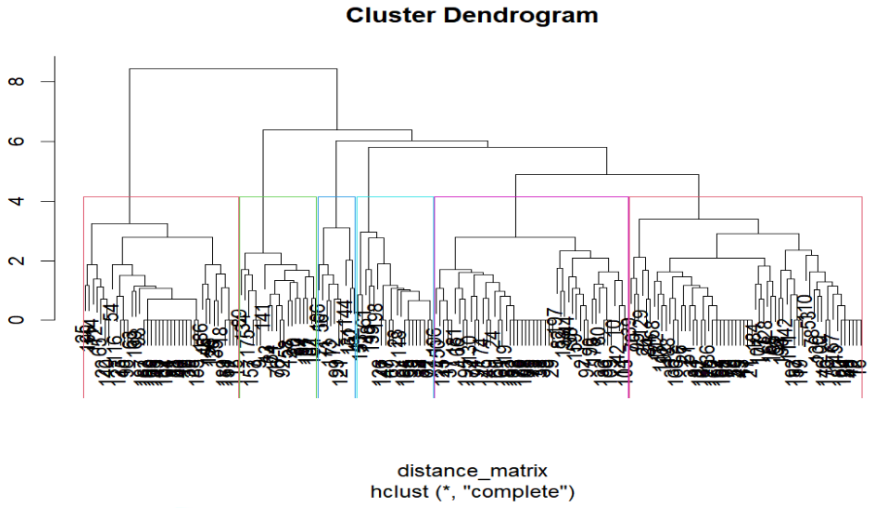

Figure 12

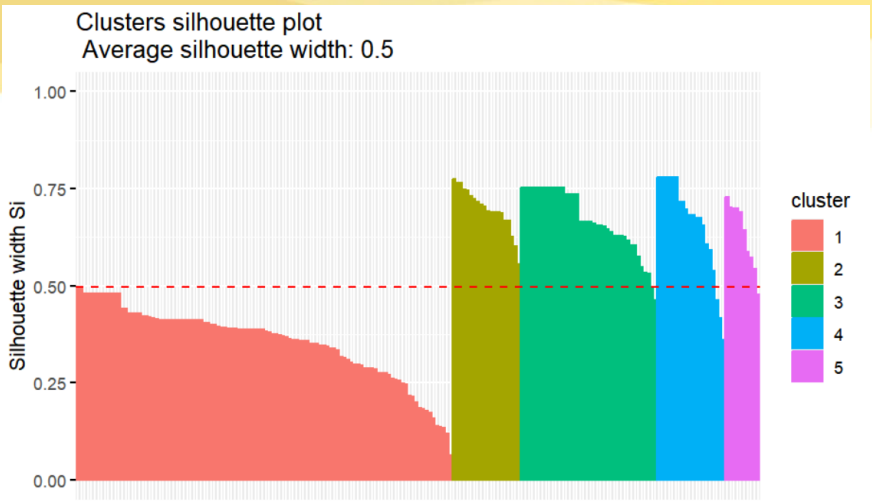

Figure 13

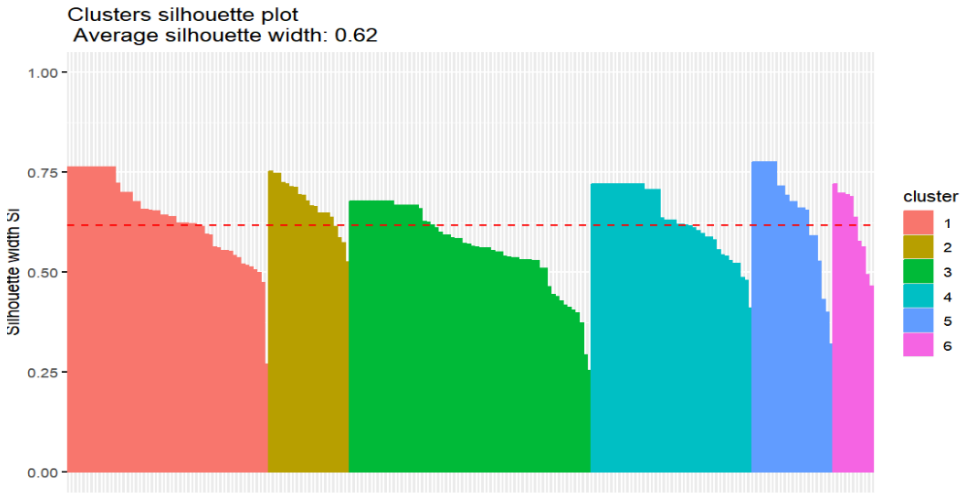

Figure 14

	1	2	3	4	5	6	7	8
2	0.4836068							
3	3.7080664	3.6122196						
4	1.7790845	1.7120941	3.6122196					
5	0.9832050	0.8560471	3.5093181	0.8560471				
6	3.2871711	3.5274101	4.5971635	3.5274101	3.4219594			

	9	10	11	12	13	14	15	16
2								
3								
4								
5								
6								

	17	18	19	20	21	22	23	24
2								
3								
4								
5								
6								

	25	26	27	28	29	30	31	32
2								

Figure 15

```
clusters_hclust
 1  2  3  4  5  6
50 20 60 40 20 10
```

Figure 16

A tibble: 6 × 11

Cluster <int>	Avg_Age <dbl>	Gender_Distribution <chr>	Avg_Income <dbl>	Avg_TV_Hours <dbl>	Avg_WEB <dbl>	Avg_Out_of_Home_ads <dbl>	Avg_Radio_freq <dbl>	Avg_social_media_freq <dbl>	Avg_Numkids <dbl>
1	-0.38598154	54% Male	0.35555233	0.1498545	0.2420435	-0.3460703	-0.4287084	-0.6411446	1.0841418
2	-0.04062964	50% Male	-1.77776165	2.1566020	-1.9221102	-1.2316033	-0.3557367	-1.3899265	-0.3554563
3	0.16251854	40% Male	-0.64902409	-0.3844094	-0.1977480	-0.4342844	-0.5077610	1.2152107	-0.7849661
4	0.89893070	37.5% Male	1.26982975	-0.4495635	1.3051366	1.6947268	1.0580887	-0.2433541	0.3332403
5	-1.36109280	50% Male	-0.01693106	0.1368237	-0.5932439	-0.3155347	1.0124815	0.3884306	-0.3110243
6	0.16251854	40% Male	0.62644934	-1.2314132	-0.2135678	0.6514265	-0.3557367	-1.1091333	-0.7109127

6 rows | 1-10 of 11 columns

Figure 17

A tibble: 6 × 11

Avg_Age <dbl>	Gender_Distribution <chr>	Avg_Income <dbl>	Avg_TV_Hours <dbl>	Avg_WEB <dbl>	Avg_Out_of_Home_ads <dbl>	Avg_Radio_freq <dbl>	Avg_social_media_freq <dbl>	Avg_Numkids <dbl>	Avg_race <dbl>
-0.38598154	54% Male	0.35555233	0.1498545	0.2420435	-0.3460703	-0.4287084	-0.6411446	1.0841418	2.50
-0.04062964	50% Male	-1.77776165	2.1566020	-1.9221102	-1.2316033	-0.3557367	-1.3899265	-0.3554563	2.45
0.16251854	40% Male	-0.64902409	-0.3844094	-0.1977480	-0.4342844	-0.5077610	1.2152107	-0.7849661	2.45
0.89893070	37.5% Male	1.26982975	-0.4495635	1.3051366	1.6947268	1.0580887	-0.2433541	0.3332403	2.45
-1.36109280	50% Male	-0.01693106	0.1368237	-0.5932439	-0.3155347	1.0124815	0.3884306	-0.3110243	2.45
0.16251854	40% Male	0.62644934	-1.2314132	-0.2135678	0.6514265	-0.3557367	-1.1091333	-0.7109127	2.30

6 rows | 2-11 of 11 columns

Figure 18

A tibble: 6 × 8

Cluster <int>	Avg_Atoms_Imp <dbl>	Avg_Cost <dbl>	Avg_LunchFreq <dbl>	Avg_DinnerFreq <dbl>	Avg_FoodImp <dbl>	Avg_ServIMP <dbl>	Avg_NumInParty <dbl>
1	-1.1026235	-0.7319202	-1.0051231	-0.09432110	-0.2300342	0.6483539	-0.05072167
2	-1.4459843	-0.6719969	-0.9333286	-1.81195804	-2.0922161	-1.8271791	-1.24630381
3	0.7947271	-0.3010432	0.4626757	0.04136891	-0.2300342	-0.6265238	0.80671603
4	0.7906971	1.4681207	1.2205066	1.06731775	1.2122439	1.2836097	0.69198845
5	0.7786069	0.7832831	-1.0131003	-1.06731775	1.1209604	-0.0916864	-1.46368238
6	-1.0832792	-0.6291946	1.2603925	1.71267267	-0.3760877	-0.7793344	-1.93466928

6 rows

Figure 19

```
   AtmosImp        AvgCost         LunchFreq       DinnerFreq      FoodImp
Min.   :1.00   Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
1st Qu.:2.00   1st Qu.:2.000   1st Qu.:1.00   1st Qu.:3.000   1st Qu.:4.000
Median :6.00   Median :2.000   Median :3.00   Median :3.000   Median :4.000
Mean   :4.34   Mean   :2.835   Mean   :2.42   Mean   :3.075   Mean   :4.315
3rd Qu.:6.00   3rd Qu.:4.000   3rd Qu.:4.00   3rd Qu.:4.000   3rd Qu.:6.000
Max.   :7.00   Max.   :7.000   Max.   :5.00   Max.   :5.000   Max.   :7.000
   ServImp        NumInParty
Min.   :1.00   Min.   :1.00
1st Qu.:3.00   1st Qu.:3.00
Median :4.00   Median :4.00
Mean   :3.99   Mean   :4.07
3rd Qu.:5.00   3rd Qu.:5.00
Max.   :7.00   Max.   :8.00
```

Figure 20

Variable list:

AtmosImp (Atmosphere Importance): Mean: 4.34

- Variation: The range is from 1 to 7, with a standard distribution as seen in the quartiles. The mean is closer to the upper half, indicating that most people rate atmosphere importance moderately to high.

AvgCost (Average Cost): Mean: 2.835

- Variation: Ranges from 1 to 7, but the mean suggests a slight skew towards lower costs. The spread between quartiles is also moderate.

LunchFreq (Lunch Frequency): Mean: 2.42

- Variation: Ranges from 1 to 5, meaning people have a variety of lunch frequency habits, but most are on the lower end.

DinnerFreq (Dinner Frequency): Mean: 3.075

- Variation: Similar to LunchFreq but slightly higher, suggesting that respondents might dine out more for dinner than lunch.

FoodImp (Food Importance): Mean: 4.315

- Variation: Ranges from 1 to 7, with the mean on the higher side, indicating that food quality is considered relatively important.

ServImp (Service Importance): Mean: 3.99

- Variation: The range is 1 to 7, with the mean around the middle, suggesting mixed perceptions about service importance.

NumInParty (Number in Party): Mean: 4.07

- Variation: Ranges from 1 to 8, showing a wide spread in group sizes, with a slightly larger means indicating moderate group dining.