

**Mall of America Case Study**

# *Title Page*

MIS 315

Professor Cao

**Group 2**

Chad Schaschwary

Brooke Downing

Dylan Schultz

Jeffery Morische

Kenzi Drummond

Yen Vo

Jessica Coffin

# **Table of Contents**

[**Title Page**](#_ey5ox9twmi0x) **1**

[**Table of Contents**](#_2qcuc7kwx93b) **2**

[**Case Background**](#_6mwcr1oe46m2) **2**

[**Problems Analyzed**](#_w2fg3dcd6me0) **3**

[**Evaluation of Conducted Procedures**](#_3n2fjsq4y7rq) **4**

[**Data Variables**](#_25qq0vdcr5hp) **5**

[**Data Cleaning**](#_lml2wxmm8ks) **6**

[**Data Analysis**](#_jk75gj1s4ddg) **6**

[Multiple Visits](#_vjwzbsd5uv0j) 7

[Shopping Habits](#_wbcf3irn49yk) 9

[Duration](#_f9uau8g7ix6a) 11

[**Usability of Research**](#_dcffnj8kccap) **12**

[**Conclusions**](#_yqwskwwn8u0y) **13**

[**Appendix**](#_jvi857gf6kik) **14**

# 

# **Case Background**

Mall of America (MOA) is one of America’s largest tourist attractions located in Bloomington, Minnesota. At over 5.5 million square feet, MOA is currently the largest mall in

America. MOA focuses on being a retail and entertainment venue. MOA is home to over 520 stores, a sea life aquarium, a transit station, and two hotels. It hosts over 400 events every year (Thomas and Bapna 2020).

The online shopping industry has radically improved the convenience of consumer shopping. To compete in the marketplace, MOA needed to adjust its operations to enrich the in-person shopping experience. Despite the significant investment cost, top management implemented Wi-Fi internet access throughout the entire mall. Wi-Fi helped consumers navigate the mall, helping them locate specific store locations. While Wi-Fi enhanced the consumer experience, it also represented an opportunity for management to capture new consumer information. Management saw an opportunity to use this information to improve consumer experiences to get them to stay longer and to come back more frequently. In this case study, we used the Wi-Fi data to better understand the shopping habits of MOA customers.

# **Problems Analyzed**

The research questions we decided to answer were categorized into three main categories: recurrence, shopping habits, and duration of mall visitors.

The recurrence category focused on defining what characteristics frequent shoppers have in common. We conducted a clustering analysis to determine the relationship between recurrent visitors and sections of the mall visited using recurrence data. We wanted to see if frequent visitors are returning to a specific section(s) of the mall. Also, we used clustering analysis to evaluate if frequent visitors visit on the weekend or during the week. Recurrent visitors are valuable to MOA as they represent more frequent shoppers. The more we understand recurrent visitors, the better MOA management can design their offerings and physical location to appeal to them.

The shopping habits category focused on significant shopping trends captured by Wi-Fi data. We looked for the busiest days and months for the shopping mall, both for the stores and the food courts. Also, we wanted to understand at what point consumers typically begin using the Wi-Fi during a visit, whether that’s inside or outside the mall. Visitors accessing the Wi-Fi for a long duration outside the mall may represent non-customers. Finally, we also looked at what time of the day the average customer arrives at the mall. A large number of customers arriving at opening can indicate the need to open earlier in the morning. Also, this information can help develop promotional efforts based on a time of day basis.

Finally, the duration section of our analysis included questions about customers who visit the longest. We want to know what floors and sections most customers spend the most amount of time on. These floors and sections can signify popular stores, and an adjustment to the location of these stores can be made for convenience. We also created visualizations to show what days customers stay at the mall the longest. When customers are at the mall longer, the higher the probability of purchasing. To entice customers to purchase, stores in MOA can create advertisements for days in which customers are likely to visit longer.

# **Evaluation of Conducted Procedures**

To answer the research questions discussed above, our group created visualizations in Tableau and conducted cluster analyses in RapidMiner using the given data. Tableau helped us visualize aggregated data to observe useful trends in customer shopping habits. We used cluster analysis to define customers into different segments based on their different characteristics. More specifics will be discussed in the data analysis section of the report.

# **Data Variables**

The first table that was used for analysis was the moa\_analytics table. The table included information about 107 attributes for 12,621 data records. The most important data variables used are defined below:

**Device**: MAC address of the device

**Session\_date**: Date of visit

**entry/exit\_time: Time of first/last session update with mall Wi-Fi**

**Duration:** Sum of duration of all session updates (in seconds)

**Distinct\_sections\_visited:** Totaldistinct level-sections of the mall visited

**Distinct\_levels\_visited:** Total distinct levels of the mall visited

**Weekday\_flag:** Visit on weekday or weekend, 1 stands for weekend visit

**day\_of\_week:** Day of the week (Sunday=1, Friday =6)

**time\_0-4:** Time spend on levels 0-4

**Time\_0-4\_section:** Times spent on level and section (level 0 South)

**time\_outside:** Total time spent outside the mall building

**Time\_entrance:** Total time spent at entrances

The second table that was used for analysis was the moa\_recurrence table. The table included information about four attributes for 12,577 data records. Each attribute for the dataset is described below:

**Device:** MAC address of the device

**Session\_date:** Date of visit

**In\_30days:** Number of visits within the past 30 days

**In\_60days:** Number of visits within the past 60 days

# **Data Cleaning**

The data used for analysis did not include any null or missing information. We combined the recurrence and analytical datasets to analyze the data together.

We created an “ID” field within the recurrence dataset as a first step to combining both datasets. The ID field, numbered 1-12,577, established a unique identity for every record. Since both datasets have “device” and “session\_date,” we created a concatenated primary key using the two values for every record in both datasets. Concatenation takes the text of both attributes and combines them into a singular attribute. Using the new concatenated key, device\_session\_date, we used the Excel function, VLOOKUP, to match the concatenated keys of the analytics and recurrence tables. We returned the ID field from the recurrence data onto the analytics data. Using another VLOOKUP function, we used the ID field from the analytics table to return the appropriate in\_30days and in\_60days recurrence values from the recurrence dataset. Essentially, we have now merged the two tables. However, since the analytics table included more records than the recurrence table, some of the included records did not have a valid ID or any recurrence data. To visualize the datasets in Tableau and conduct clustering analysis in RapidMiner, we excluded the values from Tableau and replaced the missing values with an average of their respective attributes in RapidMiner.

# **Data Analysis**

This section will discuss information we learned pertaining to the case study questions disclosed in the “Problems Analyzed” section. The research questions will be organized into their three categories: multiple visits, shopping habits, and duration. The visualizations can be found in the Appendix section of the report.

## *Multiple Visits*

Q1: What is the relationship between floor levels visited and recurrent visitors?

To answer this question, we conducted a cluster analysis in RapidMiner using recurrence data (in\_30days, in\_60days), time spent on each floor, and time spent at the food courts. The goal of the clustering analysis was to analyze the locations that recurring visitors are more likely to visit. We used k-means clustering, and based on the Davies Bouldin performance indicator, we determined that k=3 clusters existed in the model. *Figure 7.1* shows the size of each cluster generated. We noticed how cluster 0 is significantly larger than clusters 1 and 2. However, we were still interested in the insights that can be drawn from the smaller clusters. *Figure 7.2* includes the centroid table that describes the basic characteristics of each cluster.

**Cluster 0**: Spends the least amount of time at the food courts and restaurants. Spends the most amount of time on floors 1-2. These visitors are least likely to visit again, having recurrence values of 3.2 in\_30days and 4.8 in\_60days. These visitors are more likely to be customers since their recurrence values score the lowest.

**Cluster 1**: Spends the least amount of time at the south restaurant. Spends the most amount of time on the main floor (floor 0), while spending the least amount of time at every other floor location. These visitors have the highest recurrence rates, with values of 14.6 and 23.6 for 30 and 60 days, respectively. They spend most of their time on floors 0, 1 and 3, so it is paramount to consider the store locations on these floors because of their popularity with the highest recurring customers.

**Cluster 2**: Spends the most amount of time at the northern and southern food courts. Additionally, they spend the most amount of time at the south restaurant. Considering they spend, on average, about 2.5 hours at the south restaurant, this group could mainly consist of employees at that particular restaurant. They spend a significant amount of time on floor 3, around 6.6 hours for the average individual. This cluster is highly differentiated and could represent employees because of their medium recurrence scores and the amount of time spent on floor 3 and the south restaurant.

Q2: Are recurrent visitors more likely to visit on the weekend? How does recurrence impact duration?

To answer these questions, we again used RapidMiner cluster analysis to broadly define different customer segments. In this analysis, we used weekend\_flag (1=weekend, 0=weekday), duration, and recurrence data as inputs for the cluster analysis. Based on the Davies Bouldin measure, k=5 clusters is the best performing model. Our cluster size details are noted in *Figure 8.1.* We noticed that the cluster sizes are more evenly distributed than the first cluster analysis we performed using the floor level details. The centroid table can be found in *Figure 8.2*.

Based on the centroid table results, we found that recurring visitors are less likely to visit on the weekend. In clusters 0,3, and 4, the weekend flag is higher (indicating more weekend days), while the recurring visit values are lower. On the other hand, clusters 1 and 2 have lower weekend flag values (indicating more weekdays), while the recurring visit values for both 30 and 60 days are significantly higher that of the other clusters. We also discovered that the duration of visits tends to be higher during the weekdays. This was an unusual observation, as individuals tend to have more free time on the weekends. In cluster 1, the centroid value for the duration was unusually high, where the duration was for around 23 hours. This cluster can signify potential devices that are always located within the mall. There were only 148 data records classified within this cluster, so we can assure cluster 1 to be negligible. Cluster 2 has a duration centroid value of just over 10 hours. While this is still a large value, it is more feasible than 23 hours. This centroid can include both employees and customers. Their recurrence values are also substantive, with 11.0 and 16.7 visits in 30 and 60 days, respectively.

## *Shopping Habits*

Q1: What are the busiest months and days of the week?

We found the busiest months and days of the week for our dataset in Tableau using the session date and the number of devices connected. Our objective was to determine which months and days have the most customers at the mall. First, we used Tableau and created a graph showing the number of devices in the mall for each month. *Figure 9.1* illustrates the visits by month. We saw there was a large drop in devices following the month of March and into April. The number of users drops nearly 2,400 just from March to April. We also noticed the winter months, from November to March, have more Wi-Fi users.

Additionally, we created another Tableau graph using the number of devices and the day of the week. We noticed the weekend draws the most people to the mall. Also, we noticed the middle of the week, Tuesday to Thursday, has the lowest number of users. *Figure 9.2* shows these visits based on the day of the week.

Q2: When are restaurants and food courts busy?

The goal was to find food court areas that customers spend more time in along with how this has changed from year to year. We created a graph that compares time spent in seconds with the food court locations each month. We noticed a few patterns in the data from this graph. The first pattern noticed was customers spend significantly more time in the South restaurant. We also realized the busiest month for food courts is November. The graph of this can be seen in *Figure 10.1*.

The second graph we created was to compare 2015 food court data with 2016 food court and restaurant data to see changes in tendencies. We noticed both food court sections had higher numbers of visitors in 2016 than in 2015. However, we noticed that the visitors for 2015 are significantly higher than in 2016 for the south restaurant. In fact, every day of the week has dropped in visitors since 2015. The graph between food court and restaurants per year is shown in *Figure 10.2*.

Q4: What time, on average, do customers arrive at the mall? Does this differ by day of week?

To answer these questions, we created a Tableau visualization that highlights the entry time for every hour during each day of the week. Based on the visualization in *Figure 10.3*, the peak entry time for customers tends to between 11am-12pm. The day that stands out the most is Saturday, as many customers continue to enter past 4pm. While the number of connected devices appears to plateau throughout the day Monday-Friday, the growth in the number of visitors accessing the Wi-Fi throughout the early afternoon is much more noticeable on Saturday and Sunday.

Q3: Where do people begin using Wifi(entrance and outside)?

We implemented the K-mean algorithm in Rapid Minner to define the average customer experience. Based on *Figure 10.4*, Cluster 1 has the most observations and the information shows that the largest cluster of customers spend around half an hour at the entrance. Characteristics of this cluster are they may visit two to three distinct levels, spend time at North and South food courts, and visit the mall on Wednesday.

## *Duration*

Q1: At which sections of the mall are customers spending the most time?

We used Tableau to create bar graphs of the data using duration, section, floors, and the date. The data was used to determine if certain sections or floors were more popular than others. Our graphs measured the average duration for each floor section based on the day of the week and can be found in *Figure 11.1*. A few points stood out after creating these visualizations. The first was that the East, West, and North sections of the mall are frequented the most on the first floor. However, the third floor has the longest duration by far for the South section.

Q2: Is there any relationship between Apple users and duration of visits?

The number of Apple devices connected to Wi-Fi nearly doubled in 2016 (*Figure 11.2*). Refer to *Figure 11.3,* it showsa relationship between Apple and duration . It is more likely that Apple users have higher duration values. This segment is also another important segment to help the mall target their customers, they can estimate the spending and capacity of Apple users in each level, such as their favorite locations, restaurants in order to customize food chains and stores for this segment.

Q3: What is the average duration by day of week?

To answer this question, we used Tableau to create a time graph to highlight the days in which average duration values are higher than average. Our visualization can be found in *Figure 12.1* in the Appendix. Based on our results, duration is higher, on average, Tuesday (3) and Wednesday (4). This was contrary to our initial hypothesis. We assumed that visitors would stay longer on the weekends due to a reduced work schedule. However, the duration is shorter on the weekends. Using this information, MOA needs to encourage stores to issue promotions during the middle of the week. The longer customers are in the store, the higher the probability of customers purchasing goods from the stores.

Q4: What is the average duration by floor level?

To analyze on what floor levels customers spend the most time, we created a Tableau visualization. Our visualization can be found in *Figure 12.2* of the Appendix. Floor levels 1, 0, and 3, highlighted in dark green to emphasize above-average duration values, are where customers spend most of their time. This is an important consideration for store locations, as it can highlight the most profitable stores, or it can indicate a need to relocate popular stores to these floor levels.

# **Usability of Research**

The research questions answered above are meant to benefit top management of Mall of America. In addition, the research can be used to benefit individual store owners located within MOA. Our research highlights the importance of capturing and using available Wi-Fi data to learn more about the shopping habits of customers. By learning more about customer tendencies, MOA can adjust their operations to satisfy loyal (recurring) customers. By improving customer convenience, MOA can more effectively compete with online shopping retailers.

# **Conclusions**

The conclusion section of this report will be broken down into the three main research categories: multiple visits, shopping habits, and duration:

*Multiple visits*

* Average time spent at food courts and restaurants is minimal but higher for recurrent visitors
* Recurrent visitors are less likely to visit on the weekends
* Duration is longer during the weekdays

*Shopping habits:*

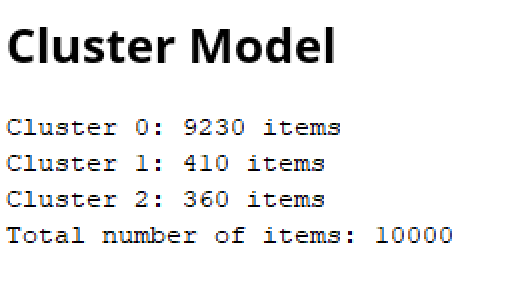
* Winter months are the most popular months
* Friday and Saturday are the most popular days
* South restaurant is the busiest food location
* Saturday- likely to get late-afternoon visitors

*Duration:*

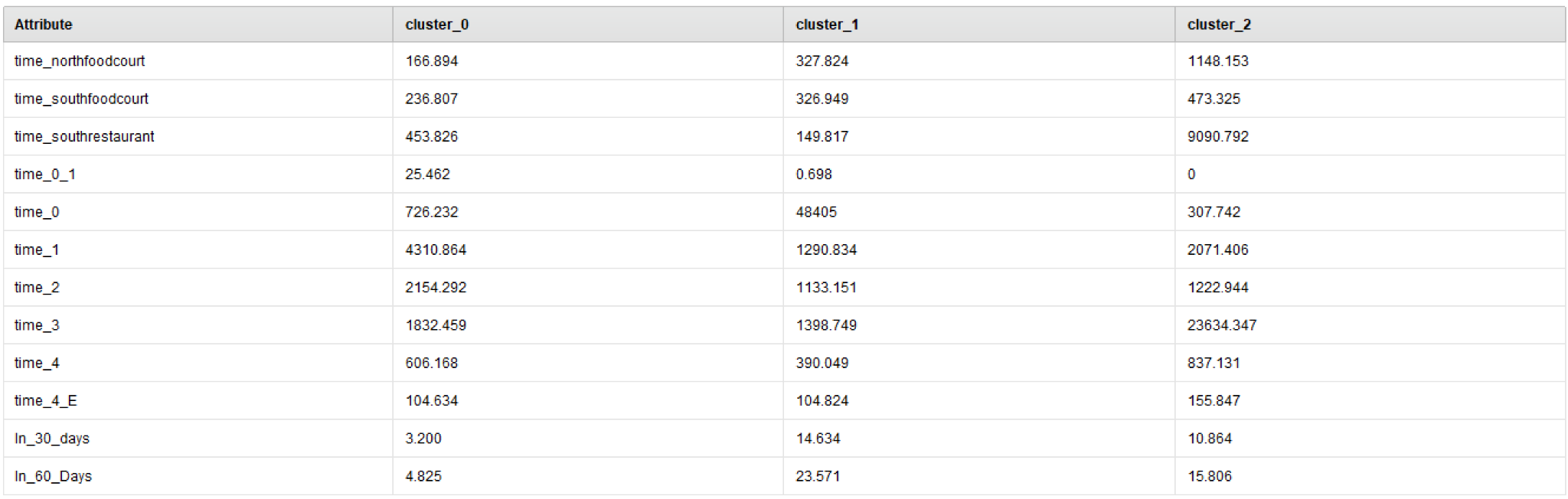
* South section is the busiest
* Apple users have higher duration values
* Tuesday and Wednesday have higher customer duration values
* Customers spend most time on floors 0, 1, and 3

# **Appendix**

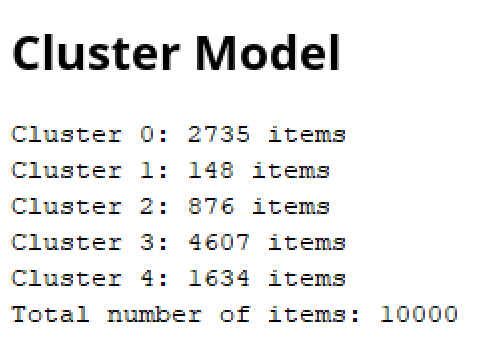
*Figure 7.1: Multiple Visits Q1 Cluster Sizes*



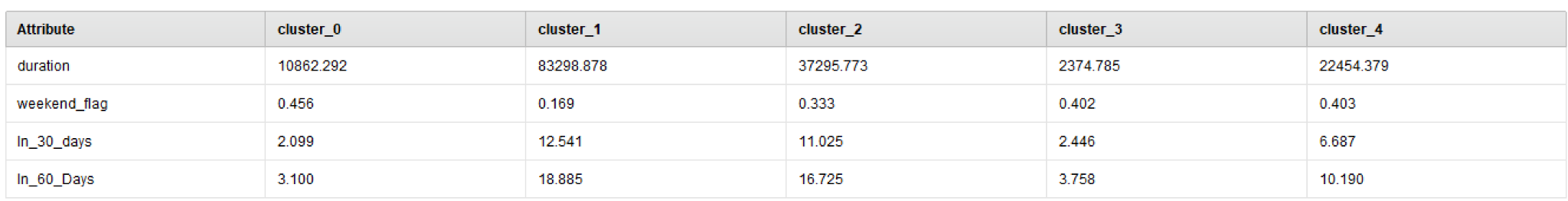
*Figure 7.2: Multiple Visits Q1 Centroid Table*



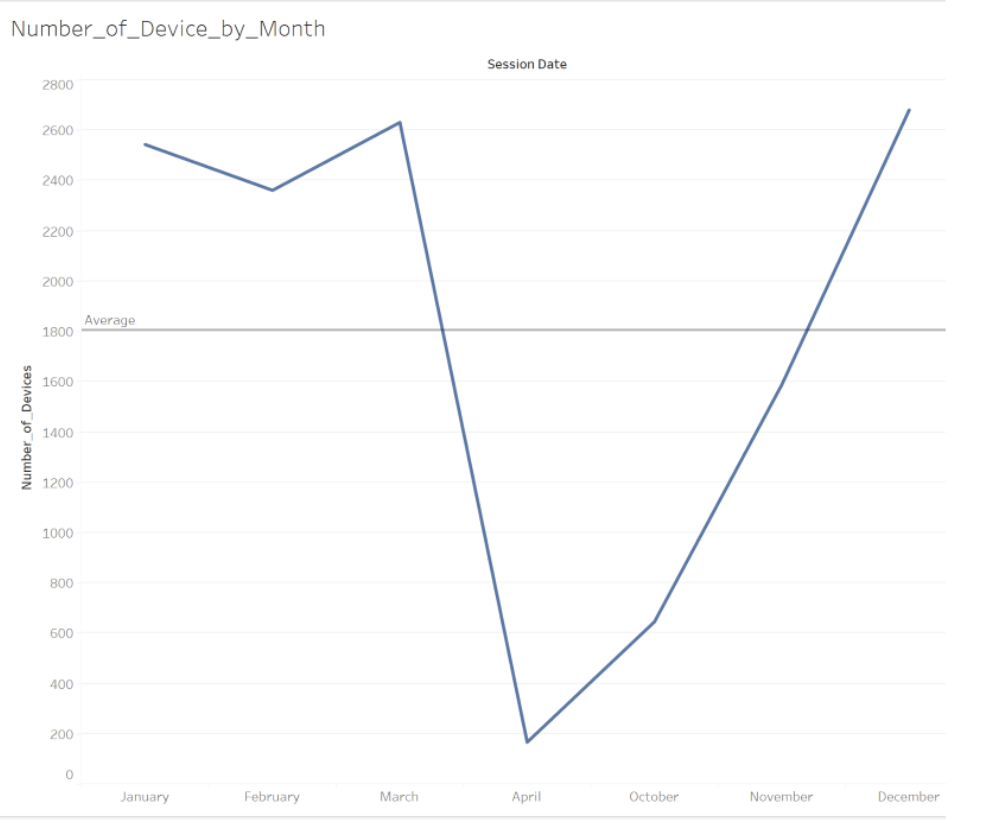
*Figure 8.1 Multiple Visits Q2 Cluster Sizes*

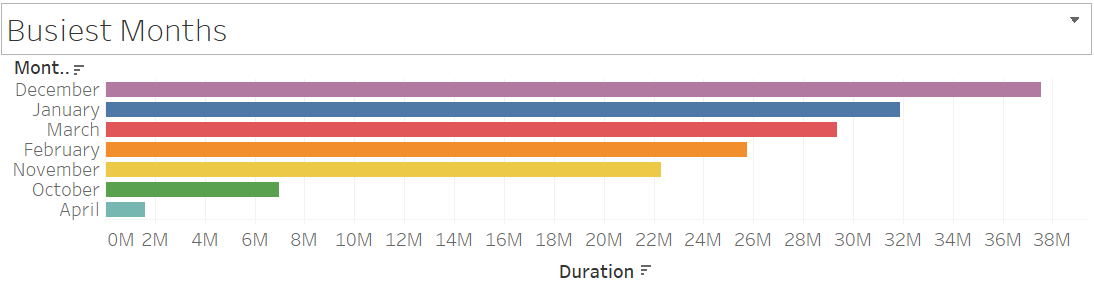
**

*Figure 8.2 Multiple Visits Q2 Centroid Table*

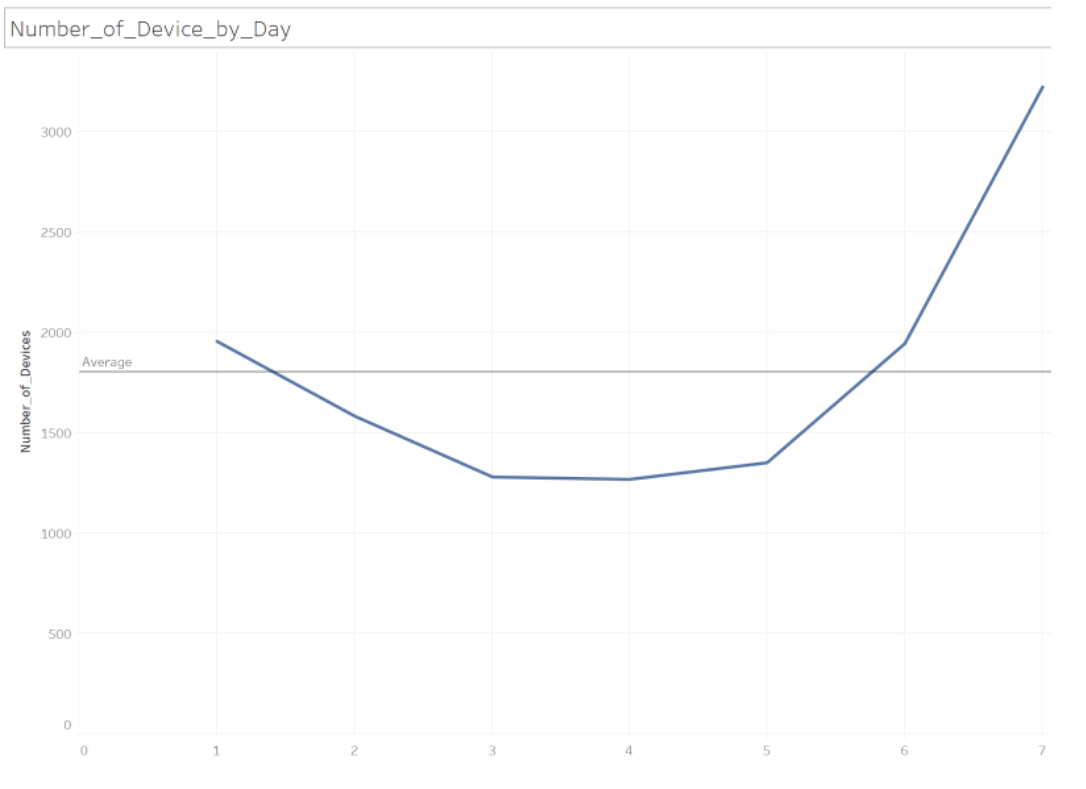
**

*Figure 9.1 Visits by Month*

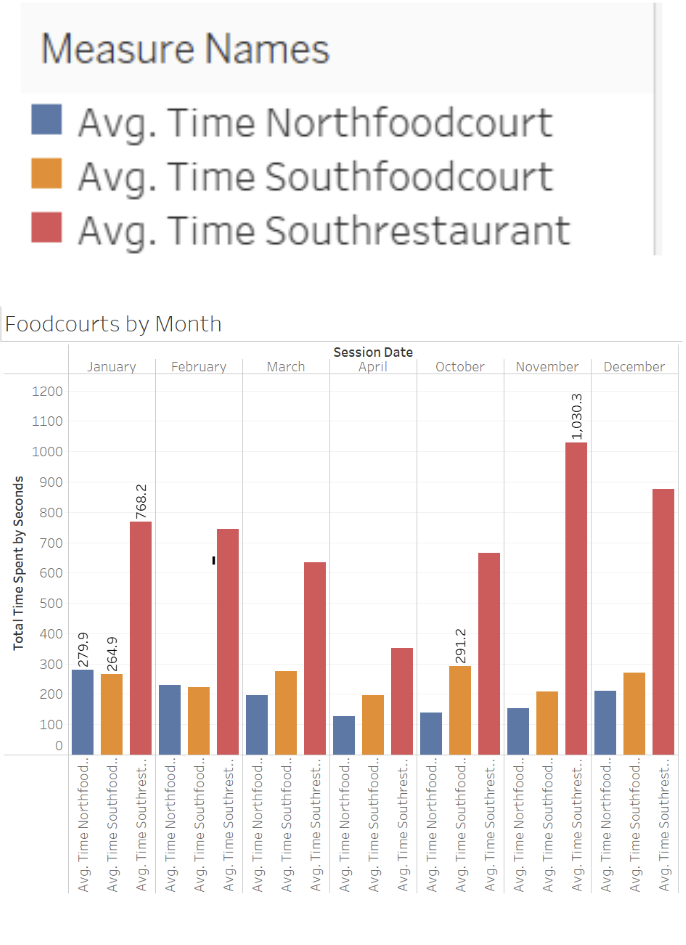


**

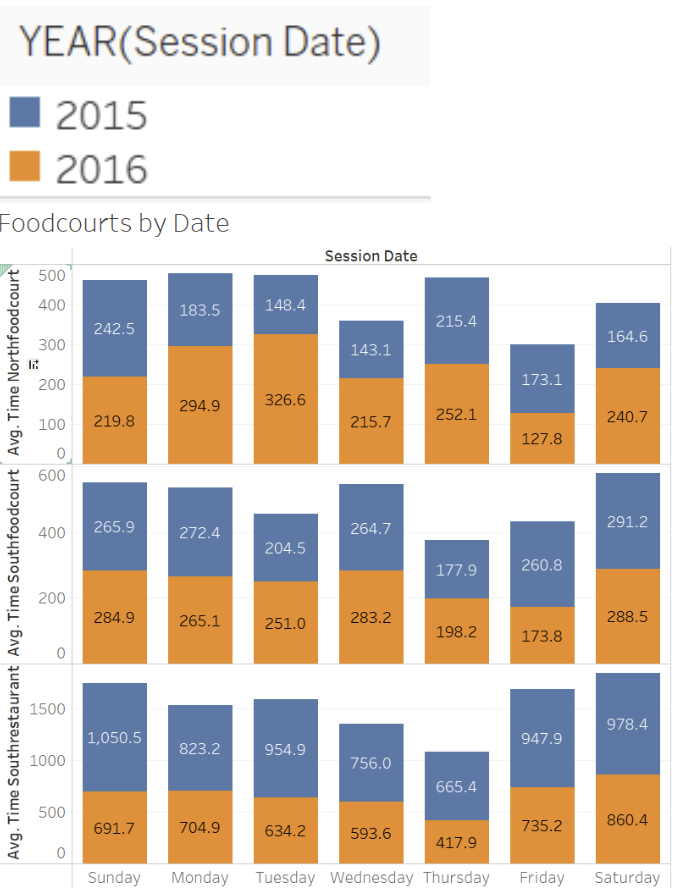
*Figure 9.2: Visits by Day*



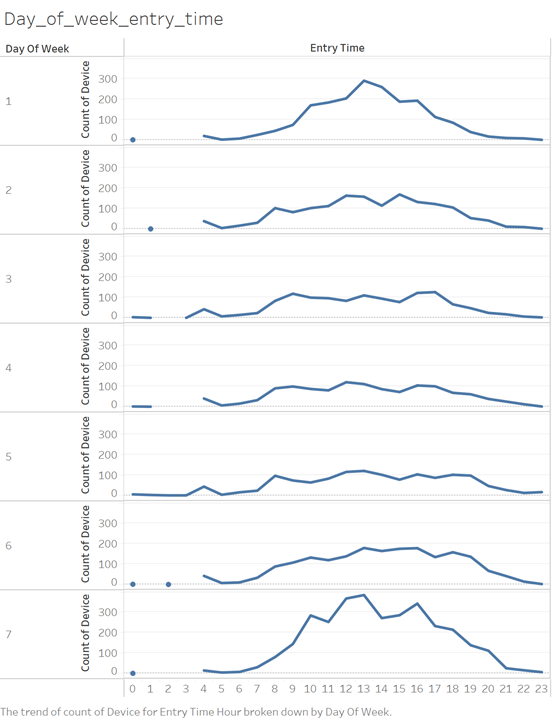
*Figure 10.1: Food Courts by month*



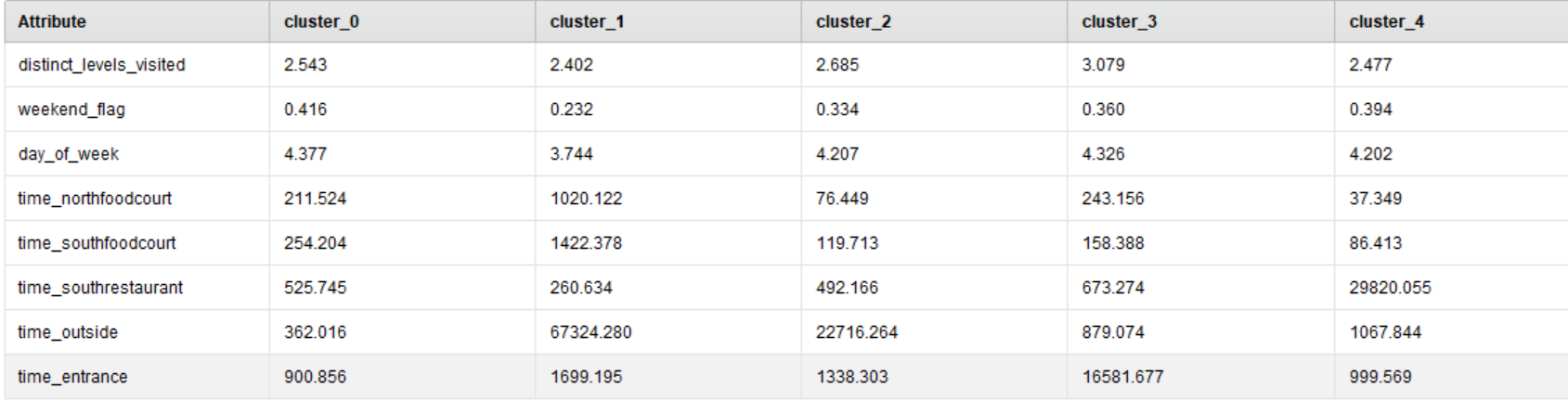
*Figure 10.2 Food Court by Year*



*Figure 10.3 Entry Time by Hour and Day*

**

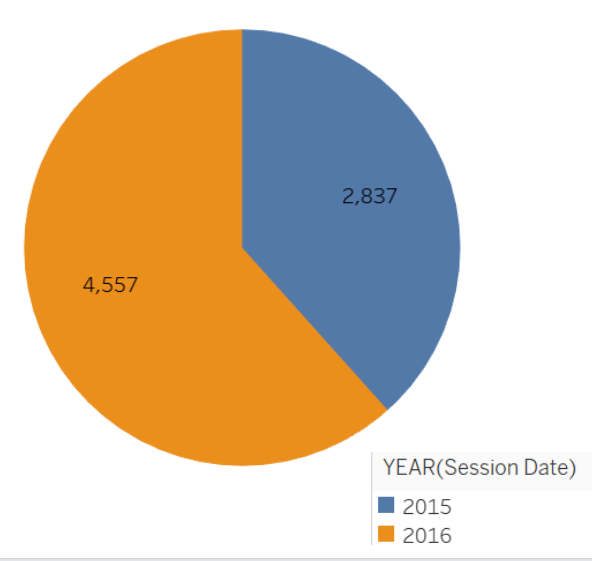
*Figure 10.4 Time entrance and Time Outside Cluster Analysis*

**

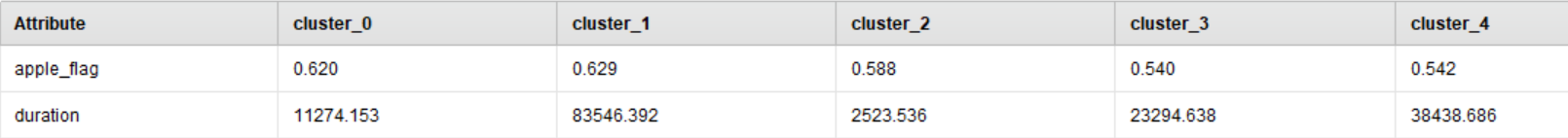
*Figure 11.1: Average Duration by Section and Floor*

**

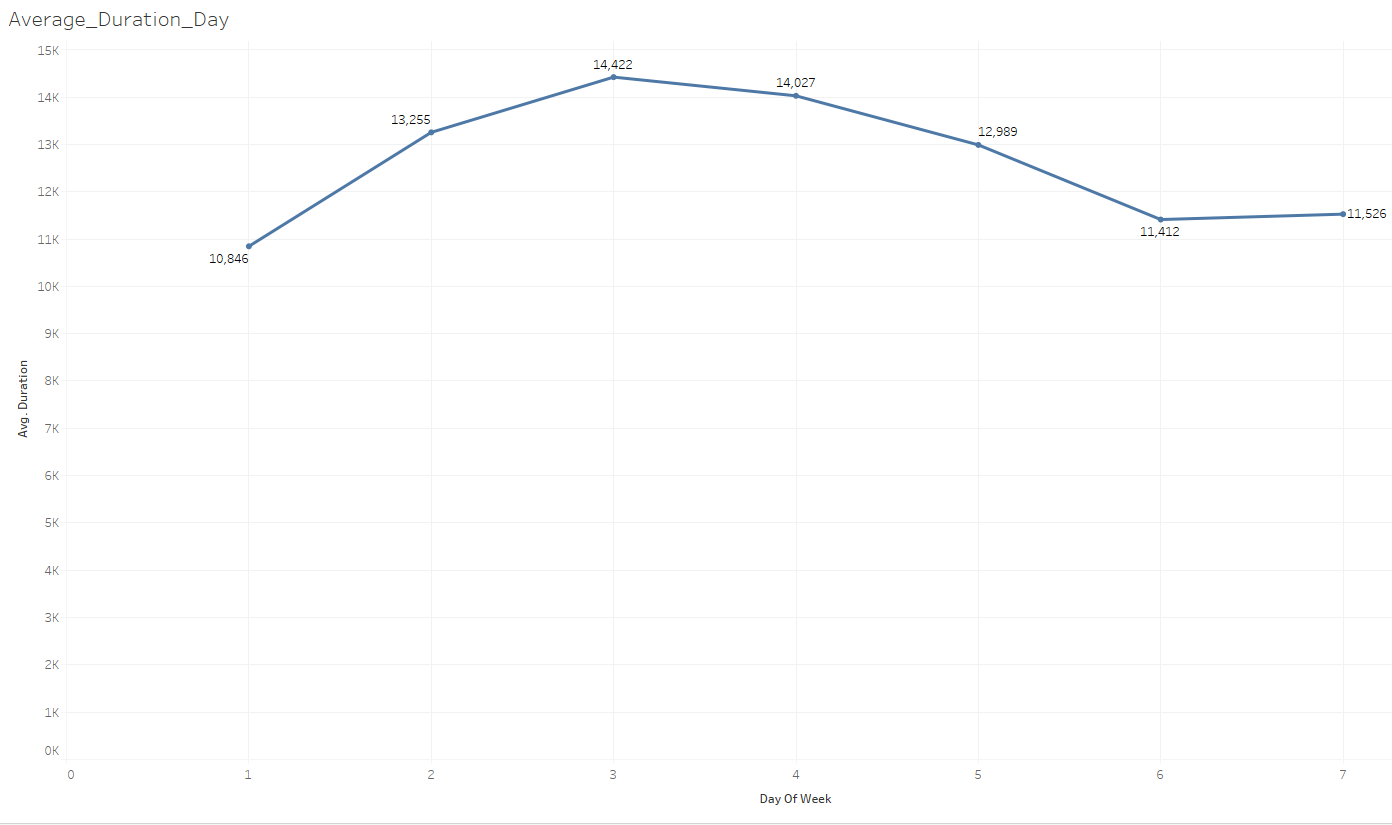
*Figure 11.2: Apple Users in 2015 vs 2016*

**

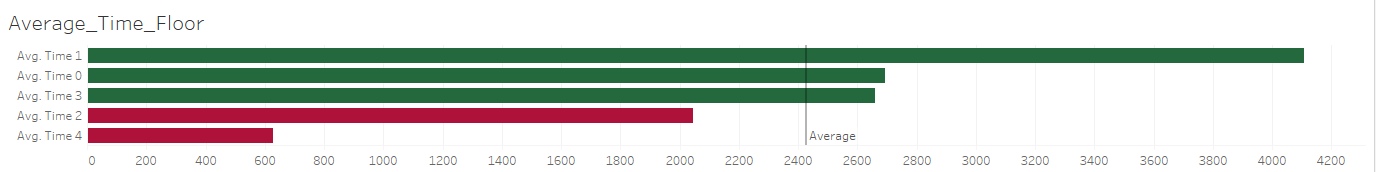
*Figure 11.3: Apple Flag versus Duration in each Clusters*

**

*Figure 12.1: Average Duration by Day of Week*



*Figure 12.2: Average duration by Floor*

**

# 