Fast Food Restaurant Market Analysis

Group Project In Statistical Methods and Applications I

(Spring 2023)

By
Vaishnavi Asuri: 110544621
Meghana Vasanth Shettigar: 110571897
Debrup Basu: 110396574

TABLE OF CONTENTS

Fast Food Restaurants Market Analysis

In our project we want to gain insights on the data contributed by the establishment of Fast Food Chains all over the US. It is divided into 4 parts:

To begin, we would want to examine the prevalence of Fast Food Chains in various regions of the United States by categorizing the regions geographically. This will aid in determining which regions have a strong presence in fast food businesses and which do not.

Second, we want to look at which Fast Food businesses have a stronger presence than their competitors in various states around the United States. This will assist fast food restaurants that do not have a dominating market share in specific locations to plan their market share in order to increase sales. It would also assist the dominant Fast Food companies in analyzing the places where they may achieve a larger market share. It will also assist them in considering the establishment of branches in certain places based on their strengths and shortcomings.

Third, we want to identify which fast food restaurants have broken a given region's health policy in the United States. Improved Food Safety: Health agencies may determine which fast food companies need to improve their food safety policies by evaluating health violation data, ensuring that the public is protected from foodborne disease.
Better Consumer Awareness: The examination of health violation data can give customers useful information, allowing them to make educated decisions about where they eat.
Increased Competition: By making health violation data public, fast food restaurants are pushed to improve their food safety methods and maintain high standards in order to retain a favorable reputation and attract consumers.
Improved Regulation: Data analysis on health violations can assist health authorities in identifying areas where rules need to be improved or enforced, ensuring that all fast-food restaurants adhere to the same high standards.
Better Resource Allocation: By analyzing health violation data, health departments may more efficiently allocate resources to the areas where they are most needed.

Fourthly, we want to perform sentiment analysis of the yelp reviews for the most popular fast food chains all over the US. Improved Customer Satisfaction: By studying customer sentiment, restaurants may find areas for improvement in order to satisfy customer expectations and boost satisfaction. Increased Sales: Positive reviews can attract new customers and raise sales, whilst bad reviews can serve as a wake-up call for the restaurant to make improvements.
Competitive Advantage: By studying client emotion, restaurants may distinguish themselves from their competition and provide a one-of-a-kind customer experience.
Sentiment research may give insight into the most effective marketing methods, allowing restaurants to make informed decisions about how to contact potential consumers.
Better Reputation Management: Restaurants may handle negative comments and maintain a positive online reputation by monitoring and evaluating consumer opinion.

Proposed Data Source:
Chicago Food Inspections | Kaggle - Meghana
Fast Food Restaurants Across America | Kaggle- Vaishnavi
LA Restaurant & Market Health Data | Kaggle - Meghana
Yelp Dataset | Kaggle- Debrup

Potential real application of the project:
The insights provided by this project could be valuable to policymakers, health organizations, and consumers. Policymakers could use the findings to formulate regulations and policies aimed at improving the health and environmental impact of the fast-food industry. Health organizations could use the insights to educate consumers on the health impact of fast food and promote healthy eating habits. Consumers could use the insights to make informed decisions about their food choices and advocate for healthier and sustainable food options.

PART 1- ANALYSIS OF THE FAST FOOD RESTAURANT MARKET IN THE US

1.1 Context definition

1.1.a Motivation behind the project

Analyzing the fast food market in the US is important due to health concerns linked to unhealthy eating habits, economic contributions to the country, and understanding consumer behaviour. Insights from the analysis can help stakeholders make informed decisions and promote positive change. It is essential to ensure the analysis is free of plagiarism and uses professional language.

According to a report by the Centers for Disease Control and Prevention (CDC), on average, about 37% of American adults consume fast food on any given day. This three-part analysis project is an effort towards a healthier approach to such a widely consumed commodity. It contributes towards three main aspects:

- The project focuses on increasing economic revenue, improving service through consumer behaviour analysis, and promoting food safety through inspection data.
- The analysis is research-backed and aims to raise awareness of cleanliness in the fast food industry.
- The project is a significant step towards creating a healthier and safer fast food market in the US.

1.2 Research Questions/Problem Definition

The purpose of this project is to conduct an in-depth analysis of the fast food market in the US and answer the following questions:

1. To what extent is fast food consumption related to the level of development of each state?

2. Which fast food brands are prevalent in specific regions of the US, and what factors contribute to their presence?
3. Can we develop a predictive model to forecast the prominence of certain fast food brands on a state-by-state basis, and can this model be replicated for other brands and countries?

Through this project, we aim to gain insights into the fast food market's dynamics and identify trends and patterns that can help inform future research and policy decisions.

1.3 Methodology

Data description and exploration – how have you explained the source and variables and how they relate to your project?

The dataset used for the first part of the project, analyzing the fast food market and brand presence across the US, was obtained from Datafiniti, a well-known data import domain. The dataset contains 10 variables and 10,000 columns, providing extensive information on fast food brands across different US states. While the dataset's categorical nature presented some challenges, it was still deemed suitable for the project's objectives.

1.3.a Variable distributions:

1. The dataset used in the project consists of ten columns, including Address, Province, Name, Country, Latitude, Longitude, Postal Code, Website, and Key.
2. Postal Code, Website, and Key were deemed irrelevant and eliminated from the analysis.
3. The dataset is entirely categorical, requiring encoding at several stages of analysis.
4. Preliminary count analysis indicates the dominance of Mc. Donald's, Burger King, Taco Bell, Wendy's, and Arby's in that order. However, further analysis is needed to confirm these as the top brands.
5. Based on overall counts, the most prevalent states in the dataset are California, Texas, Ohio, Florida, and Indiana.
6. A new column called ‘Region’ was constructed which categorizes states into one of the four categories namely- Mid-west, South, West and North-East

1.3.b. Variable relationships

We performed various statistical tests to explore the relationship between states, regions, and brands. The Chi-square test and Pearson Chi-square test were conducted to examine the association between regions and brand names, and the result indicated a significant relationship between these variables. However, McNemar's test for region vs brands and brands vs province indicated no significant difference between the paired observations. Since these are all categorical, it is not possible to determine whether they have a positive or negative correlation with each other.

Finally, we conducted clustering analysis using an elbow plot and found that 3 or 4 clusters would be suitable for categorizing brands based on states and regions, but it was not as expected due to the categorical nature of the variables.

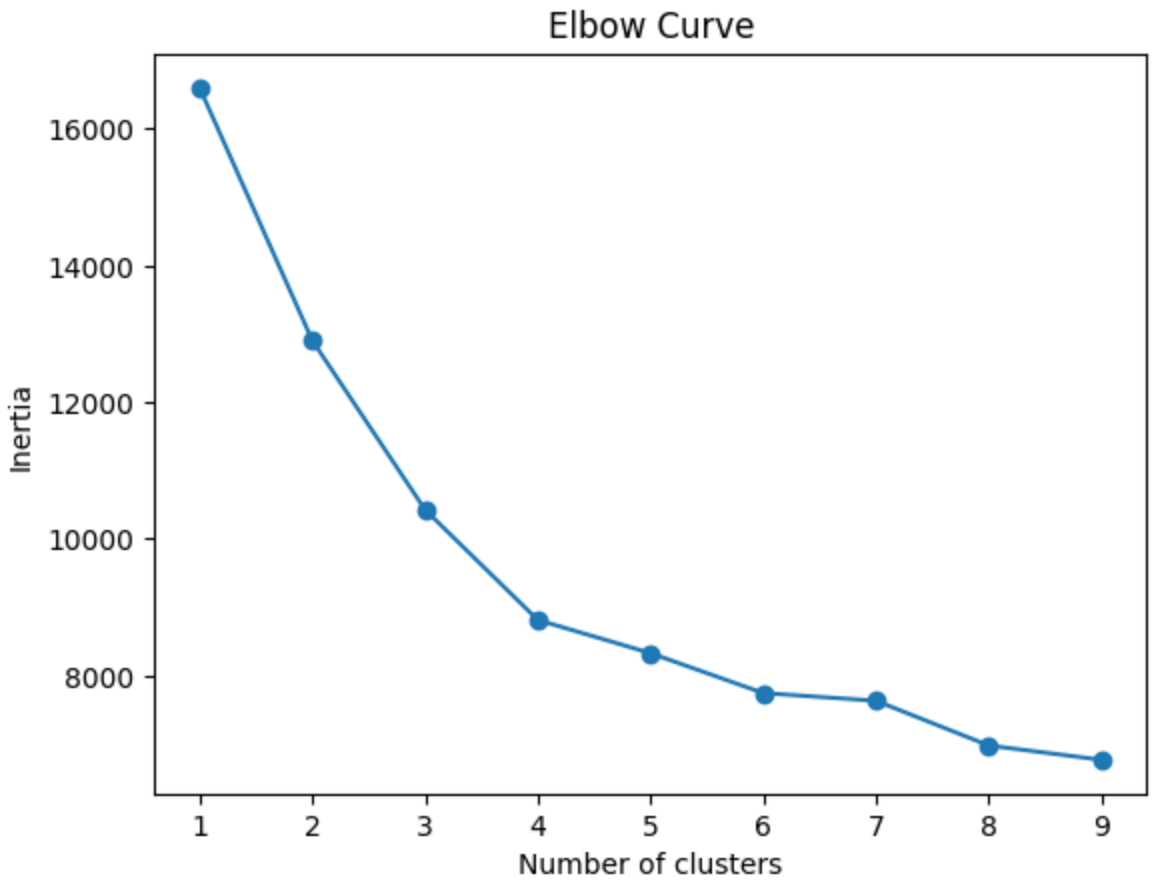


Fig 1: Elbow plot showing a possible number of clusters for the given dataset

1.3.c Model(s) Description

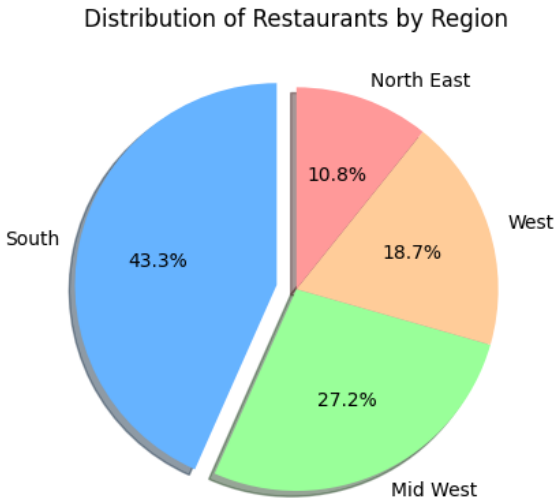
We employed five different models, namely Regression, Decision Trees, Naive Bayes, Random Forest, and Gradient Boost, to evaluate the forecasting capability of the dataset. However, even after tuning the hyperparameters, the maximum accuracy achieved was only 20% using the Decision Trees model. The limited accuracy may be attributed to the fact that the dataset consists purely of categorical data, which impacts the model's performance.

1.4 Description of findings/insights

What are your main findings and how have you documented them in relation to your questions?

We made a pie chart as you can see in this figure(*fig 2*). Southern Region of the US has the highest number of Fast food restaurants. There are several reasons why the South has the highest number of fast-food restaurants. One reason is the population density in the region, as the South has the highest population of any region in the US. Additionally, the South has a culture that values convenience and

quick service, which aligns well with the fast food industry. The warm climate in the region also encourages people to eat out more often, leading to a higher demand for fast food



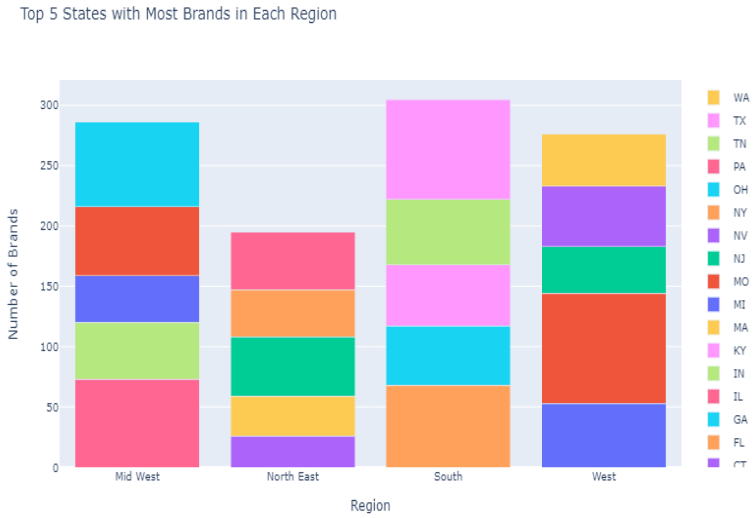


Fig 2(left): Region-wise prominence of Fast Food restaurants,
Fig 4(right): Stacked bar chart of 4 regions with wach bar consisting of the top 5 states with the highest store counts

As shown in *fig 3,* we constructed a tree map that shows all four regions which have each state with all top 3 fast food restaurant chains of that state. It seems like the highest number of chain restaurants in any given state are mostly owned by Mc. Donald's, Taco Bell, Arby's, or Burger King.

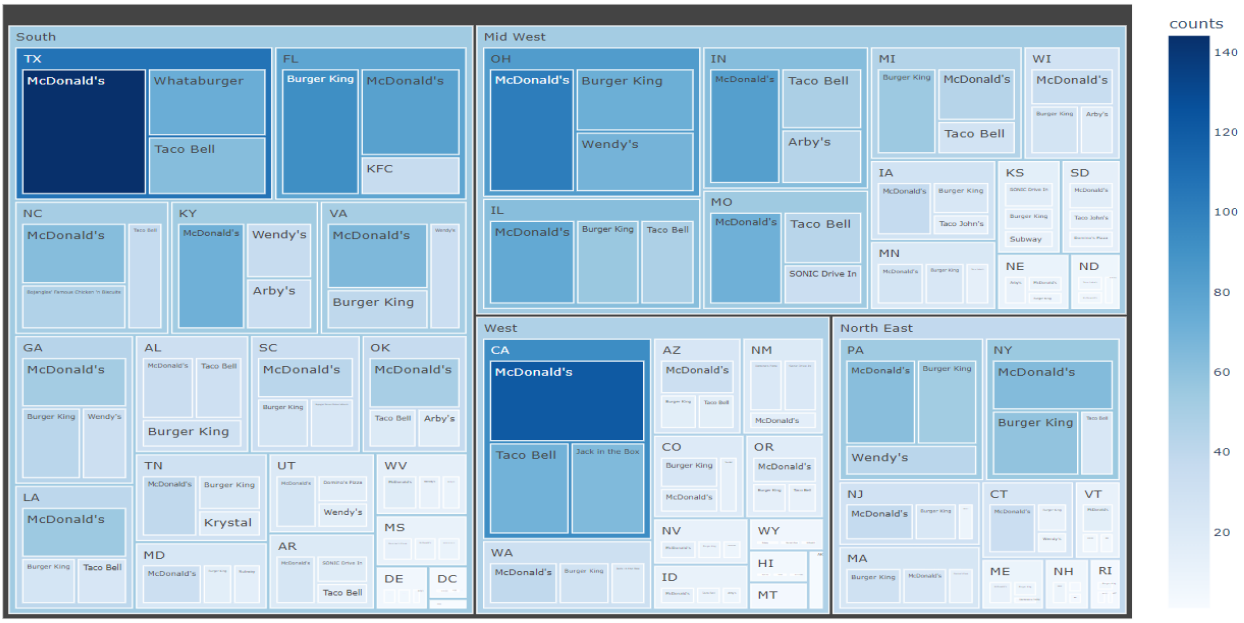


Fig 3: Treemap of region-wise store counts of the top three stores in each state of the region

The below picture depicts states in each region that have the highest number of restaurants. The highest number of stores are in the West with 91 stores in California alone. The second highest region is the South in which Texas is leading by 83 stores. In Midwest, Illinois has the next highest number of stores at 73. This is followed by Northeast which has 49 stores in New Jersey at the highest.

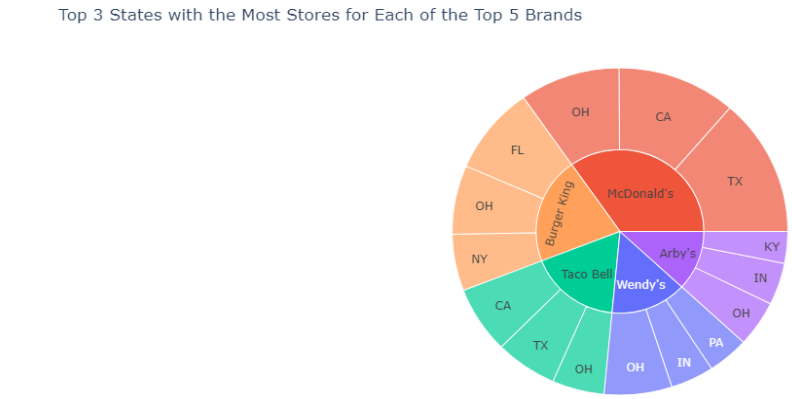


Fig 5: Donut chart with the top 5 brands and their top 5 states

As we can see in *fig. 5,* the top 5 brands turn out to be Mc. Donalds, Burger King, Taco Bell, Wendy's and Arby's. Among these, a common state that has a good number of all stores no matter what the brand is, turns out to be Ohio. So the state with the most number of popular brands can be concluded as Ohio.

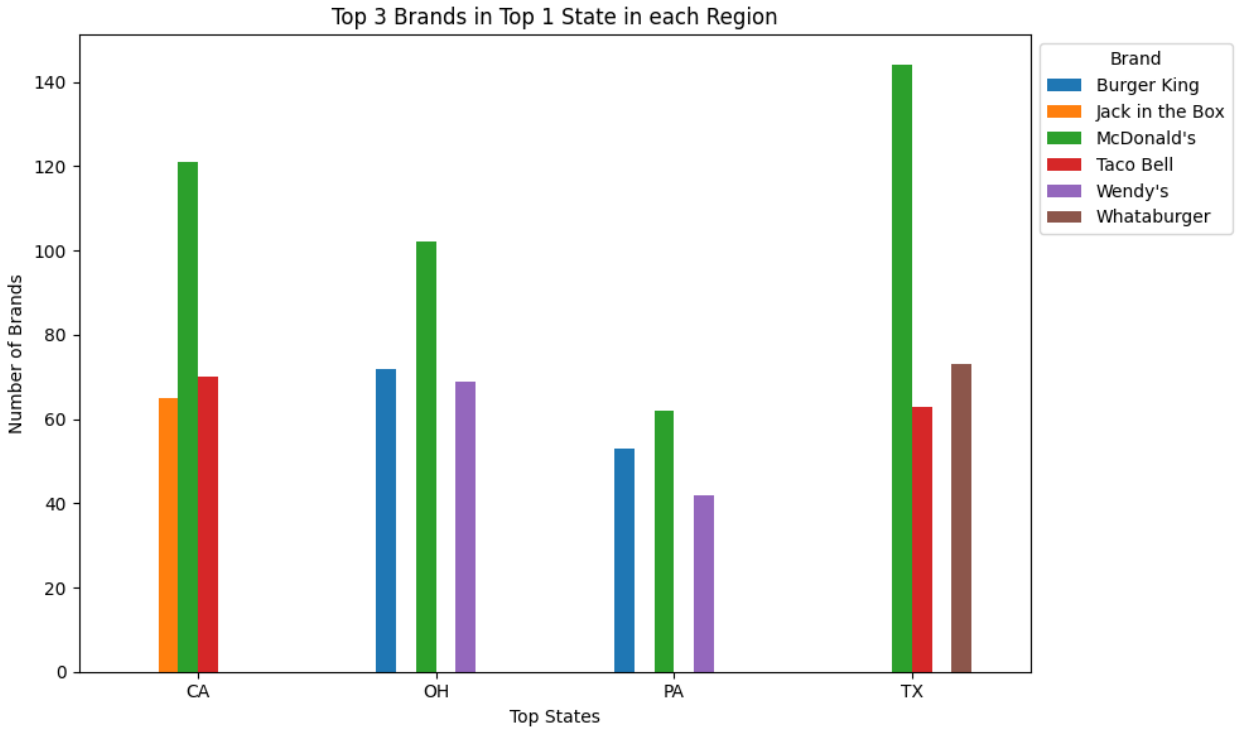


Fig 6: Top three brands in the Top most state of each Region

The above figure examines the top brands in each of the five states which have the most store in each region. One common trend to be observed here is that Mc. Donanld's repeatedly tops the charts whether it is region-wise, state-wise or else.

1.5 Limitation of Project and Future Work:

- Limited dataset: The dataset used in this analysis is limited to 10,000 rows and ten columns. A larger dataset can provide more detailed insights into the fast food market. Deeper analysis of regional variations: The current analysis provides a high-level overview of the fast food market in the US. A deeper analysis of regional variations can be carried out, looking into differences in consumer behaviour and preferences, marketing strategies, and competition.

- Categorical nature of data: The dataset consists purely of categorical data, which limits the models' accuracy and forecasting capabilities.

- Analysis of restaurant chain ownership: Analyzing the ownership of fast food restaurant chains and identifying common patterns and trends in the market can help inform future policy decisions.

1.6 Conclusion

The analysis conducted shows that the Southern region of the US has the highest number of fast-food restaurants. This could be attributed to population density, a culture that values convenience, and a warm climate that encourages eating out. The top three fast-food restaurant chains in each state of the US have been visualized using a tree map, and it appears that Mc. Donald's, Taco Bell, Arby's, and Burger King own the highest number of chain restaurants. The highest number of fast-food stores in the US is found in California, followed by Texas, Illinois, and New Jersey. The top five brands identified in this analysis are Mc. Donalds, Burger King, Taco Bell, Wendy's, and Arby's, with Ohio having the most number of popular brands. A common trend observed is that Mc. Donald's consistently tops the charts, whether it is region-wise, state-wise, or otherwise.

PART 2 -INSPECTION OF HEALTH VIOLATIONS BY BRAND

2.1 Context Definition:

This project aims to analyze health code violation data for fast food restaurants in Los Angeles and Chicago to identify any brands that may be struggling to run their operations in compliance with the standards. The insights from this analysis will help a private equity firm to decide if and where to invest and provide support to struggling brands.

2.2 Research Questions/Problem Definition:

The research questions that this project aims to answer are:

1. Which fast food brands have the highest number of health code violations in Los Angeles and Chicago?
2. How does the number of h violations compare to the number of stores for each brand?
3. Are there any specific brands that are consistently struggling to meet the health code standards, and therefore may need support from a private equity firm?

2.3 Methodology:

The methodology used in this study involves analysing the health inspection data for fast food restaurants in Los Angeles and Chicago. The data is analyzed to determine the total number of violations for each brand, the number of violations per store, and the overall performance of each brand. The analysis also includes a comparison of the performance of each brand with their respective number of stores. This analysis provides insights into which brands are facing challenges in running their operations due to health code violations.

The methodology for this study involves the following steps:

1. Data collection: The health inspection data for fast food restaurants in Los Angeles and Chicago is obtained from the County of LA Public Health. The data is in the form of a CSV file.
2. Data cleaning: The data is cleaned and pre-processed using R programming language. The missing values are removed, and the data is transformed into a suitable format for analysis.
3. Data exploration: The data is explored using R packages such as dplyr, tidyr, and ggplot2 to gain insights into the distribution of violations across brands and the number of violations per store.
4. Data analysis: The data analysis is conducted using R packages such as tidyverse, ggplot2, and reshape2. The total number of violations for each brand, the number of violations per store, and the overall performance of each brand are calculated and visualized using plots and charts.
5. Comparison of performance: The performance of each brand is compared with their respective number of stores to identify if the number of violations is proportional to the number of stores.

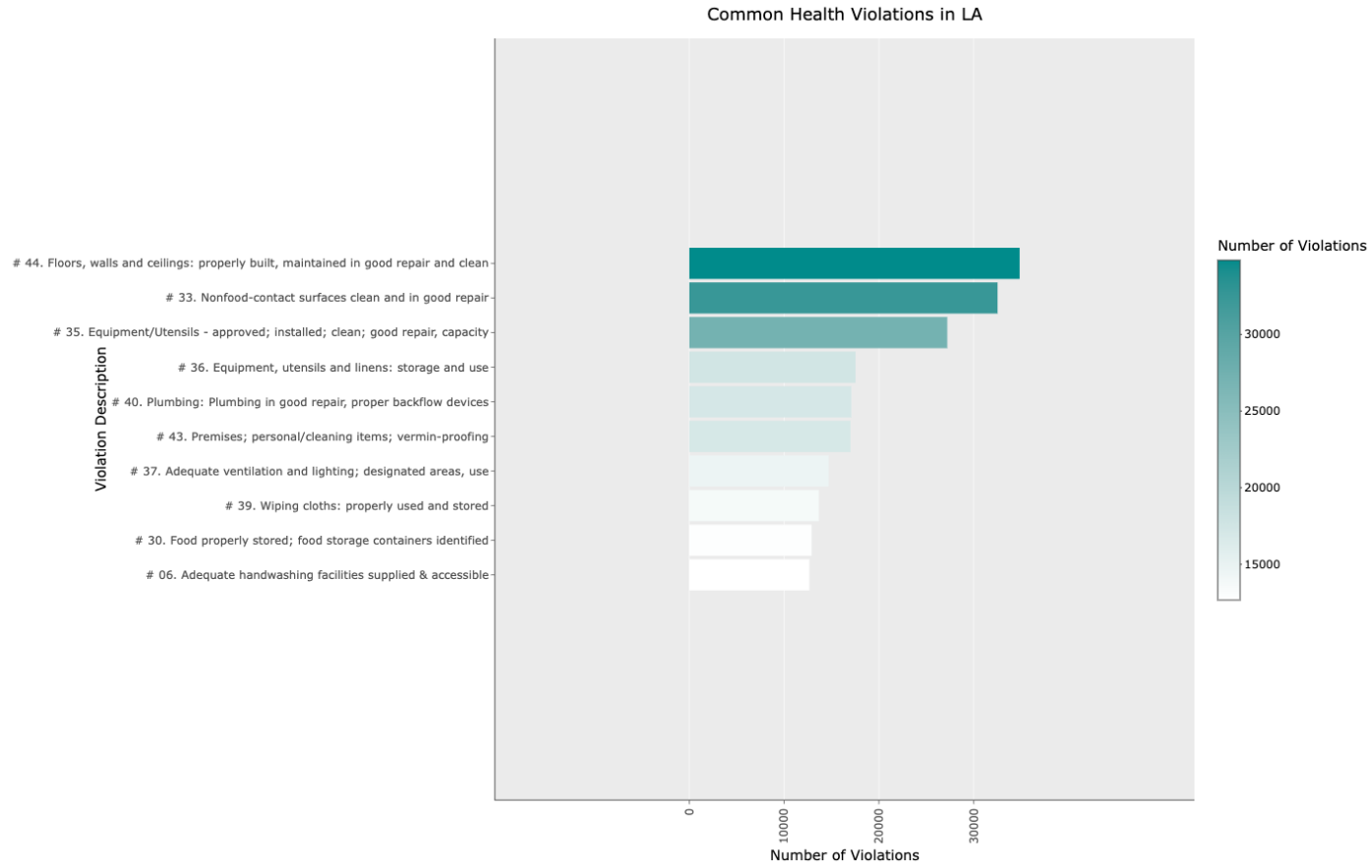
R studio packages/libraries used for this study:

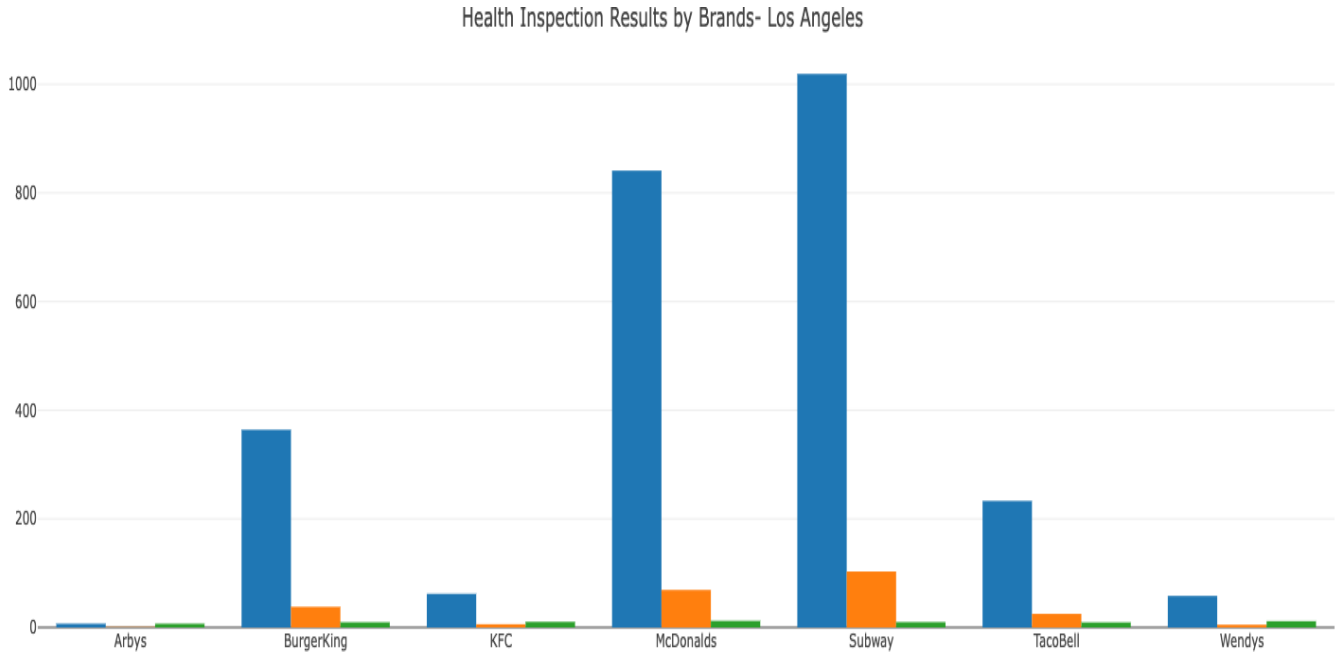
1. dplyr: Used for data manipulation, filtering, and summarization
2. tidyr: Used for data cleaning and transformation
3. ggplot2: Used for data visualization and creating plots
4. reshape2: Used for data manipulation and transformation

2.3.a LA:

```
> print("data summary")
[1] "data summary"
> summary(dh)
 facility_name      violation_code     violation_description violation_status       points
 Length:313675      Length:313675      Length:313675         Length:313675      Min.   : 0.000
 Class :character   Class :character   Class :character      Class :character   1st Qu.: 1.000
 Mode  :character   Mode  :character   Mode  :character      Mode  :character   Median : 1.000
                                                                               Mean   : 1.328
                                                                               3rd Qu.: 1.000
                                                                               Max.   :11.000
    grade           facility_address   facility_city      facility_id        facility_state
 Length:313675      Length:313675      Length:313675      Length:313675      Length:313675
 Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character


 facility_zip           score          service_code     service_description    row_id
 Length:313675      Min.   : 64.00   Min.   :  1.000   Length:313675       Length:313675
 Class :character   1st Qu.: 90.00   1st Qu.:  1.000   Class :character    Class :character
 Mode  :character   Median : 92.00   Median :  1.000   Mode  :character    Mode  :character
                    Mean   : 91.53   Mean   :  7.802
                    3rd Qu.: 94.00   3rd Qu.:  1.000
                    Max.   :100.00   Max.   :401.000
> head(dh)
# A tibble: 6 × 15
  facility_name violation_code violation_description       violation_status points grade facility_address
  <chr>         <chr>          <chr>                      <chr>              <dbl> <chr> <chr>
1 kruangtedd    F030           # 30. Food properly stored… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
2 kruangtedd    F027           # 27. Food separated and p… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
3 kruangtedd    F035           # 35. Equipment/Utensils -… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
4 kruangtedd    F033           # 33. Nonfood-contact surf… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
5 kruangtedd    F029           # 29. Toxic substances pro… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
6 kruangtedd    F044           # 44. Floors, walls and ce… OUT OF COMPLIAN…      1 A     5151 HOLLYWOOD …
# ℹ 8 more variables: facility_city <chr>, facility_id <chr>, facility_state <chr>, facility_zip <chr>,
#   score <dbl>, service_code <dbl>, service_description <chr>, row_id <chr>
>
```

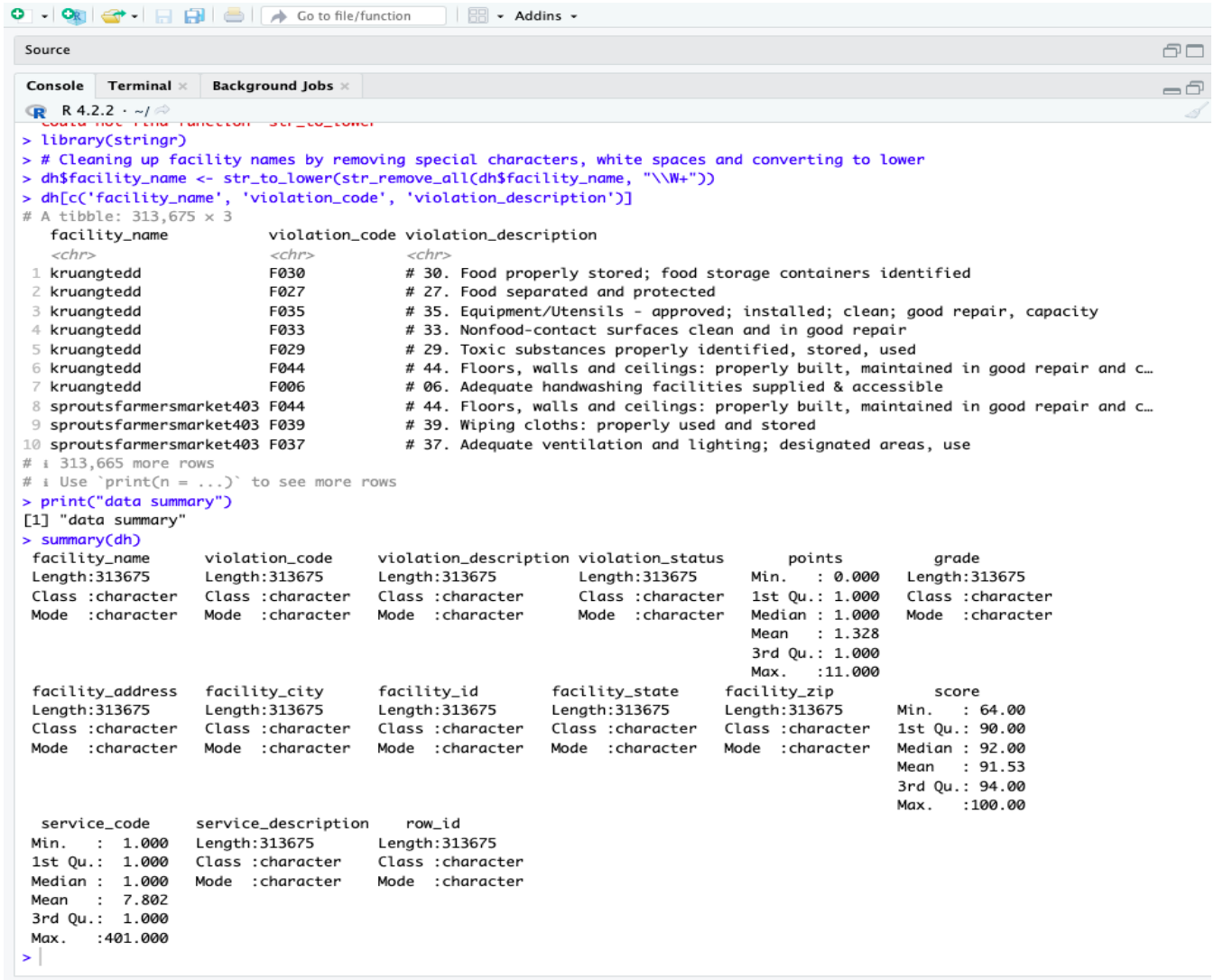




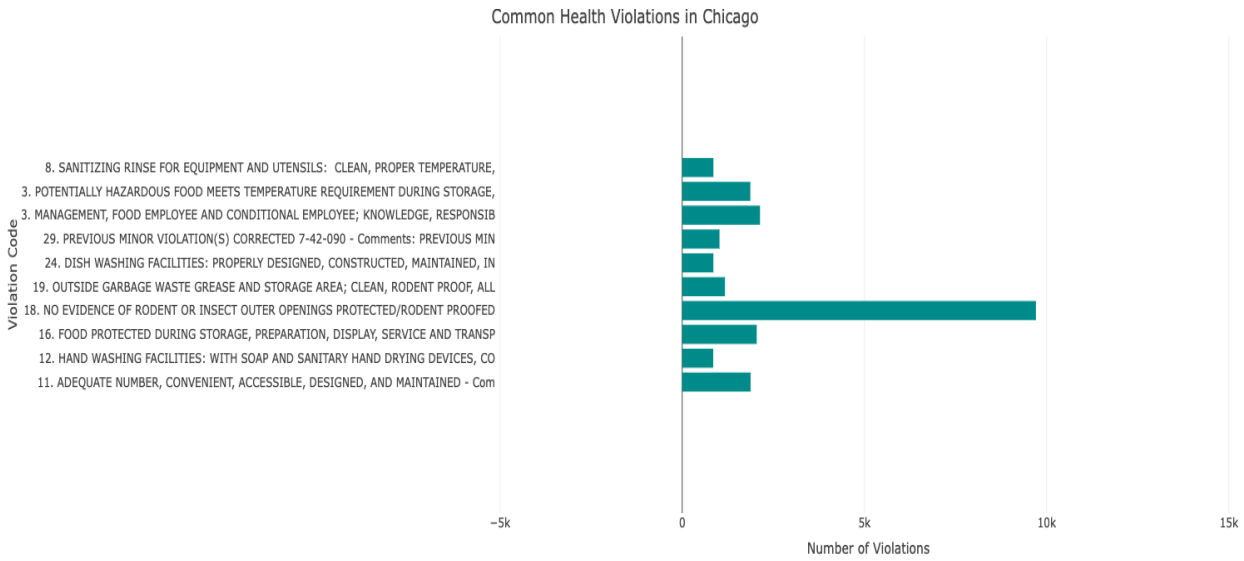
2.4 Summary:

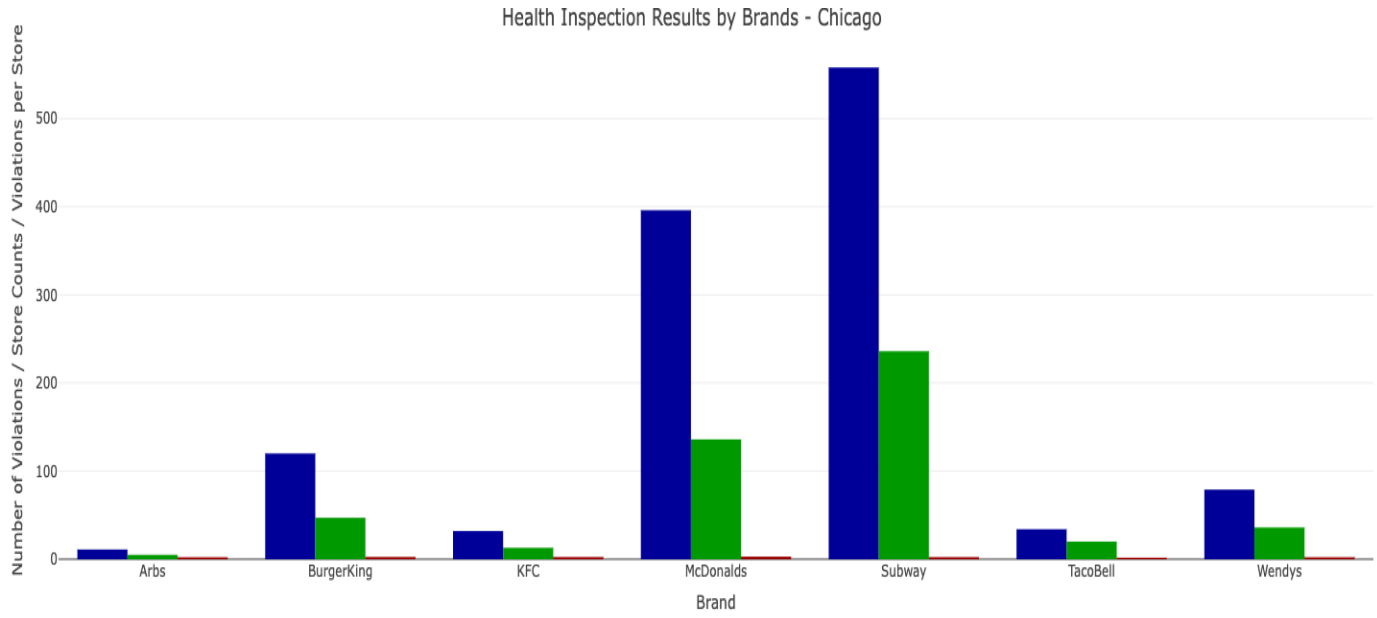
The analysis of health inspection data in Los Angeles revealed that Subway and McDonald's had the highest number of violations, likely due to their larger number of stores. KFC had the highest violations per store ratio, indicating potential difficulties in maintaining health code standards. Burger King and Taco Bell had the lowest number of violations, while KFC was found to be facing significant challenges in running their operations.

2.4.a Chicago:



```
> library(stringr)
> # Cleaning up facility names by removing special characters, white spaces and converting to lower
> dh$facility_name <- str_to_lower(str_remove_all(dh$facility_name, "\\W+"))
> dh[c('facility_name', 'violation_code', 'violation_description')]
# A tibble: 313,675 × 3
   facility_name           violation_code violation_description
   <chr>                   <chr>          <chr>
 1 kruangtedd              F030           # 30. Food properly stored; food storage containers identified
 2 kruangtedd              F027           # 27. Food separated and protected
 3 kruangtedd              F035           # 35. Equipment/Utensils - approved; installed; clean; good repair, capacity
 4 kruangtedd              F033           # 33. Nonfood-contact surfaces clean and in good repair
 5 kruangtedd              F029           # 29. Toxic substances properly identified, stored, used
 6 kruangtedd              F044           # 44. Floors, walls and ceilings: properly built, maintained in good repair and c…
 7 kruangtedd              F006           # 06. Adequate handwashing facilities supplied & accessible
 8 sproutsfarmersmarket403 F044           # 44. Floors, walls and ceilings: properly built, maintained in good repair and c…
 9 sproutsfarmersmarket403 F039           # 39. Wiping cloths: properly used and stored
10 sproutsfarmersmarket403 F037           # 37. Adequate ventilation and lighting; designated areas, use
# ℹ 313,665 more rows
# ℹ Use `print(n = ...)` to see more rows
> print("data summary")
[1] "data summary"
> summary(dh)
 facility_name      violation_code     violation_description violation_status       points          grade
 Length:313675      Length:313675      Length:313675         Length:313675      Min.   : 0.000   Length:313675
 Class :character   Class :character   Class :character      Class :character   1st Qu.: 1.000   Class :character
 Mode  :character   Mode  :character   Mode  :character      Mode  :character   Median : 1.000   Mode  :character
                                                                                Mean   : 1.328
                                                                                3rd Qu.: 1.000
                                                                                Max.   :11.000
 facility_address   facility_city      facility_id        facility_state     facility_zip           score
 Length:313675      Length:313675      Length:313675      Length:313675      Length:313675      Min.   : 64.00
 Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.: 90.00
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median : 92.00
                                                                             Mean   : 91.53
                                                                             3rd Qu.: 94.00
                                                                             Max.   :100.00
  service_code     service_description    row_id
 Min.   :  1.000   Length:313675       Length:313675
 1st Qu.:  1.000   Class :character    Class :character
 Median :  1.000   Mode  :character    Mode  :character
 Mean   :  7.802
 3rd Qu.:  1.000
 Max.   :401.000
>
```





2.5 Summary:

KFC and McDonald's are among the top 3 brands facing challenges in running their operations in both Chicago and Los Angeles, with repeated violations potentially leading to license loss. McDonald's larger presence may be contributing to their troubles, while KFC struggles despite having fewer stores.

2.6 Description of Findings/Insights:

The analysis revealed that Subway and Mcdonald's had the highest number of health code violations in Los Angeles and Chicago. However, this can be partially attributed to the larger number of stores in these cities. KFC had the highest violations per store ratio, indicating that they may be struggling to maintain

the health code standards despite having fewer stores. Taco Bell and Burger King had the lowest number of violations.

In Chicago, KFC and Mcdonald's were also among the top brands facing challenges in running their operations, with KFC having significantly fewer stores than Mcdonald's. These findings suggest that KFC may be in real trouble and may need support from a private equity firm.

Furthermore, the analysis also showed that the number of violations is not always proportional to the number of stores. For instance, Taco Bell and Burger King had more stores than Wendy's and Arby's, yet they registered fewer violations. This indicates that the size of the brand and the number of stores do not necessarily correlate with their compliance with health code standards.

Overall, the findings suggest that there is a need for increased attention to health code standards in the fast food industry, particularly for brands with higher violations per store ratios. The insights provided by this analysis can be used by private equity firms to identify potential investment opportunities in brands that may be struggling with health code compliance.

2.7 Conclusion:

Based on the analysis, it can be concluded that health code violations are a significant issue for fast food restaurants in both Los Angeles and Chicago. While some brands have higher violation rates due to their larger store counts, others are struggling to maintain the necessary standards despite having fewer stores. This analysis can help a private equity firm to decide which brands may need support in improving their operations and meeting the health code standards.

In addition, this analysis highlights the importance of compliance with health code standards in the fast food industry, as violations can lead to stiff penalties and even loss of licenses. It is essential for fast-food restaurants to prioritize the health and safety of their customers and maintain the necessary standards to avoid violations.

Overall, the insights provided by this analysis can assist private equity firms in identifying potential investment opportunities in the fast-food industry and supporting brands in improving their operations. Additionally, it can help fast-food restaurants to prioritize health code compliance and improve the overall safety and quality of their operations.

2.8 Limitations of Project and Future Work:

- The analysis only considers health code violations reported by the County of LA Public Health for Los Angeles and Chicago, which may not represent the overall health code compliance of the fast food industry in these cities. We can expand the analysis to other cities or regions to gain a more comprehensive understanding of the fast food industry's compliance with health code standards.

- The analysis only considers data up until the date of data cutoff and does not take into account any changes or improvements made by the brands after that date.
- The analysis does not take into account any qualitative factors that may affect the health code compliance of the brands, such as management practices or employee training. We could incorporate qualitative factors such as management practices and employee training to gain a more nuanced understanding of the factors that contribute to health code compliance.
- The analysis does not include data on any fines or penalties imposed on the brands for their health code violations, which may affect their financial performance.

Part 3- YELP Review Analysis on Fast Food Chains

3.1 Context Definition:

Yelp is a platform that allows customers to rate and review businesses, including fast food chains. This project aims to analyze Yelp reviews for fast food chains to gain insights into customer preferences and opinions.

3.1.a Objective:

The objective of this project is to perform data visualization and analysis on Yelp review data for fast food chains to identify popular menu items, positive and negative aspects of customer experiences, and overall trends in customer sentiment.

3.1.b Approach:

The dataset will be imported using the Pandas library in Python with a chunk size of 1000 to handle large file sizes. The data will be cleaned and preprocessed to remove irrelevant data and perform feature engineering. Exploratory data analysis is performed using data visualization techniques, including word clouds, to identify popular menu items and customer sentiments. Statistical analyses are performed to identify correlations between variables and develop predictive models.

3.1.c Research Questions/Problem Definition:

The questions asked in this project are:

- What are the top fast food brands in the US?
- What is the distribution of ratings for those brands?
- What is the number of fast food chains per brand?
- What is the average star ratings per brand?
- What are the brands with the most reviews?
- Is there a relationship between the average rating and the number of reviews for a brand?
- If yes, how does that relation vary for each brand?
- What are the top states with the most fast food brands?
- What is the distribution of average review ratings of food chains per state?

3.2 Methodology

3.2.a Data description

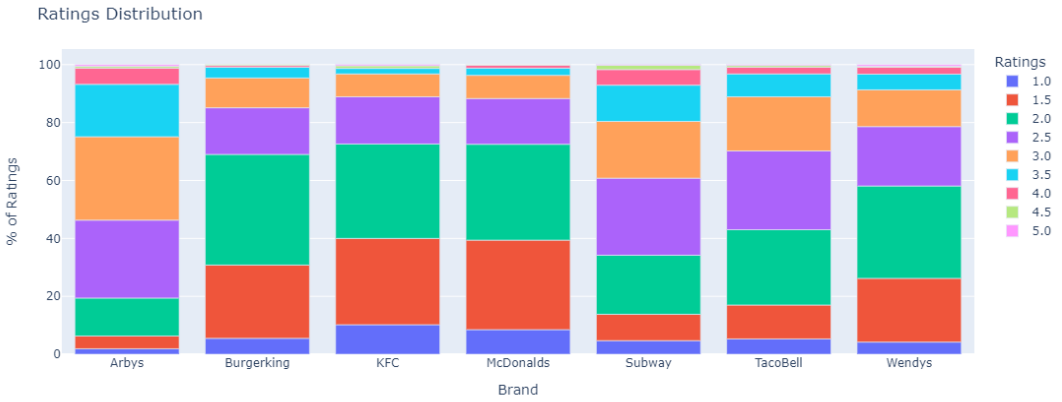
The Yelp dataset contains information on local businesses, reviews, and user data. Specifically, the dataset is divided into several JSON files:

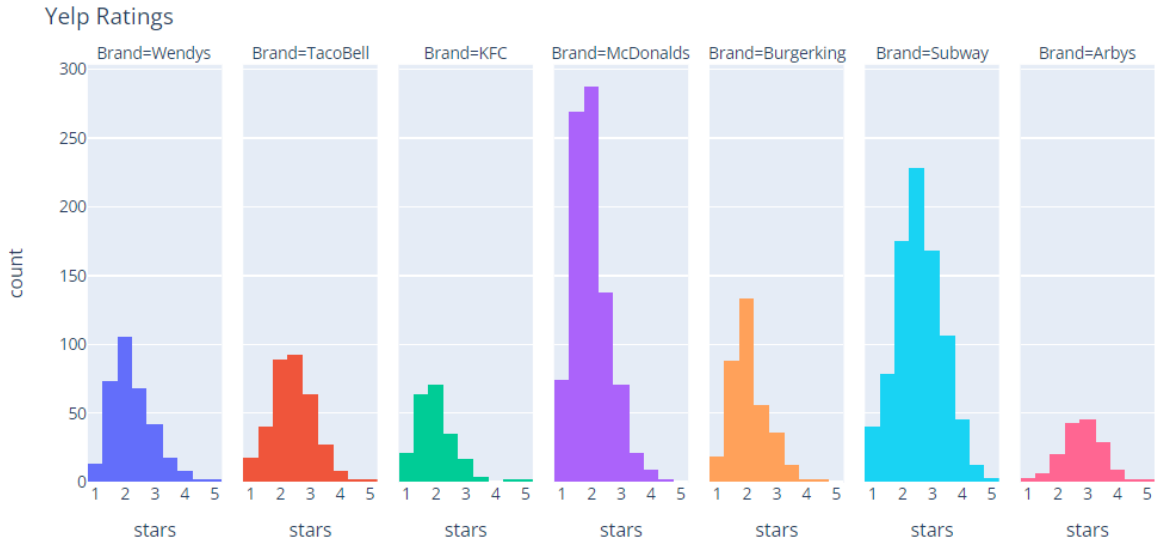
- yelp_academic_dataset_business.json - contains information on local businesses, including location data, attributes, and categories
- yelp_academic_dataset_review.json - contains review text, star ratings, and metadata about the reviews
- yelp_academic_dataset_user.json - contains user profile information
- yelp_academic_dataset_checkin.json - contains check-in activity data
- yelp_academic_dataset_tip.json - contains tips written by users on businesses
- yelp_academic_dataset_photo.json - contains photos uploaded by users to accompany their reviews

Each file contains various attributes.

- business_id: a unique identifier for the business
- name: the name of the business
- address, city, state, postal_code, latitude, longitude: location data for the business
- stars: the average rating of the business, from 1 to 5 stars
- review_count: the number of reviews for the business
- is_open: whether the business is open or closed
- categories: a list of categories that the business belongs to
- user_id: a unique identifier for the user
- review_id: a unique identifier for the review
- text: the text of the review
- date: the date the review was written
- useful, funny, cool: how many users marked the review as useful, funny, or cool
- compliment_count: how many compliments the review received
- fans: how many fans the user has
- tip: the text of the tip
- date: the date the tip was written
- likes: how many users liked the tip
- photo_id: a unique identifier for the photo
- caption: the caption for the photo
- label: whether the photo is a food photo or not

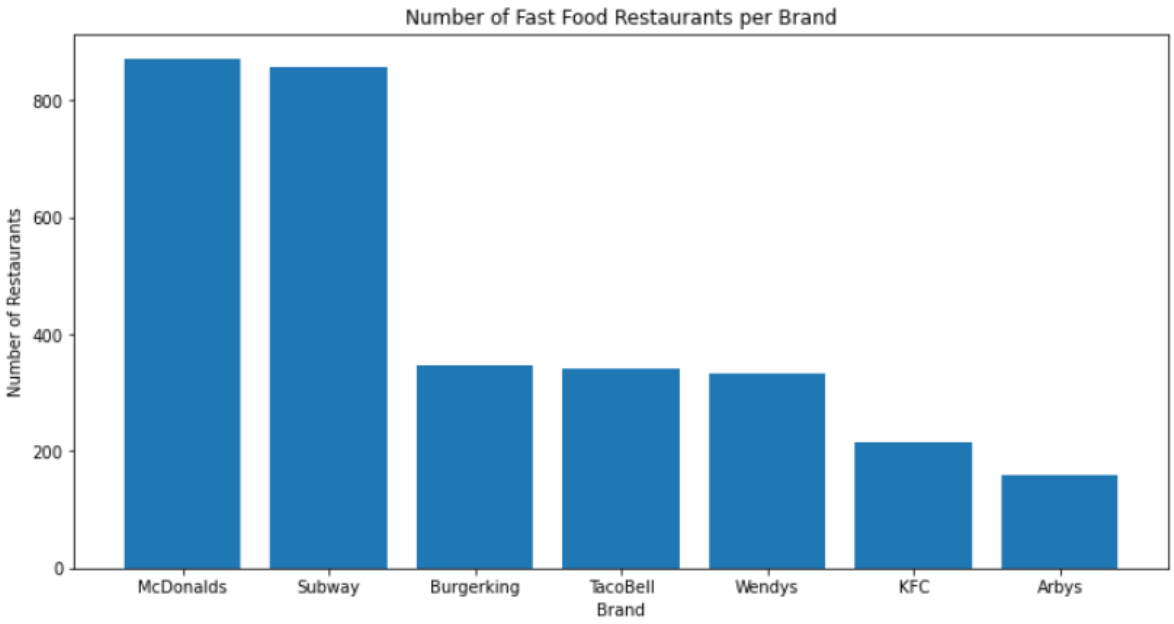
3.2.b Description of Findings, Variable Distributions, and Relationships:





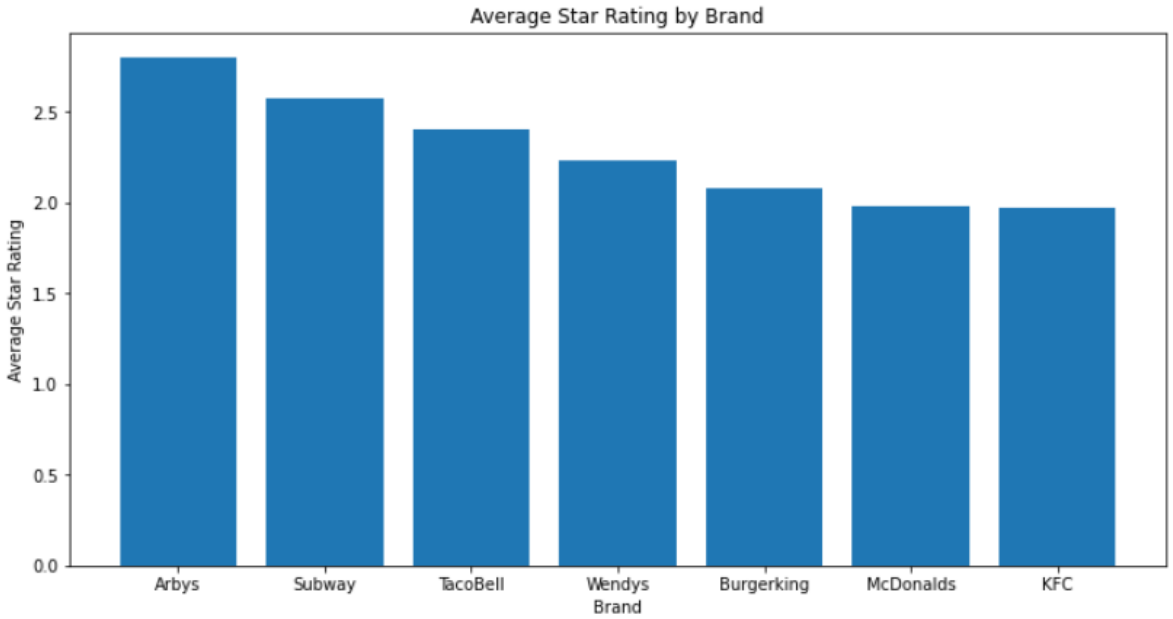
Mcdonalds seems to have the highest count of ratings compared to other brands like Wendy's, Tacobell, KFC, Burger King, Subway and Arby's. The average ratings for most of the fast food joints range from 2-2.5 out of 5.

So is there a relation between the number of reviews received by a fast food brand and the number of chains that a fast food brand has?



The inference drawn is that there is a positive relationship between the two variables, which means that as the number of branches increases, the name of the fast food chain also tends to increase. However, it's noted that the relationship is not completely linear, which could mean that the relationship is more complex than a simple linear trend, or that there are some outliers or anomalies that affect the relationship.

Overall, this image suggests that Mcdonald's has the greatest number of branches followed by Subway and that the relationship between a number of branches and the fast food chain is generally positive but not perfectly linear.

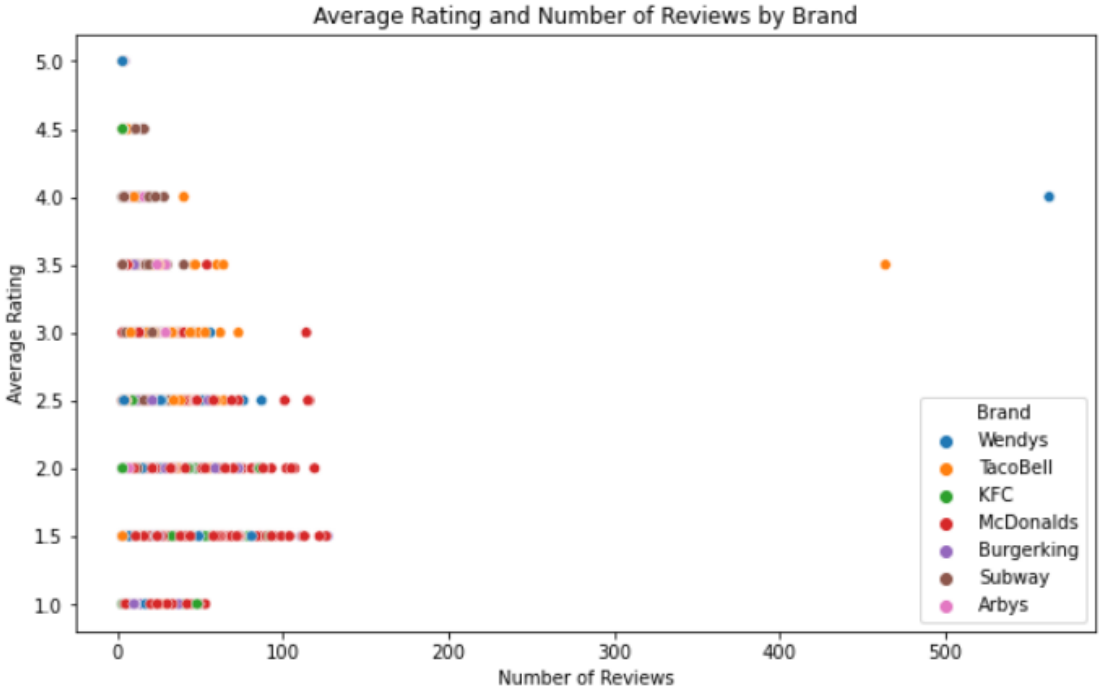


We see that the average ratings of Arby's are the greatest, followed by Subway, Tacobell, Wendy's, Burger King, Mcdonalds and KFC. Is there a relationship between the average ratings of a brand and the number of branches the fast food chain has?

The image might suggest a negative correlation between the average ratings and the number of branches as Arby's, which has the highest average rating, has the least number of branches. However, this assumption cannot be confirmed as Subway, which has the second-highest average rating, also has the second-highest number of branches in the US.

Therefore, the inference is that there is no direct significant relationship between the number of branches and the highest average rating of a fast food brand. Other factors like quality of food, service, and customer experience might be playing a role in determining the average rating of the brand.

We see that the Western Region of the US has the most number of fast food brands followed by South, Mid-West and North-East.

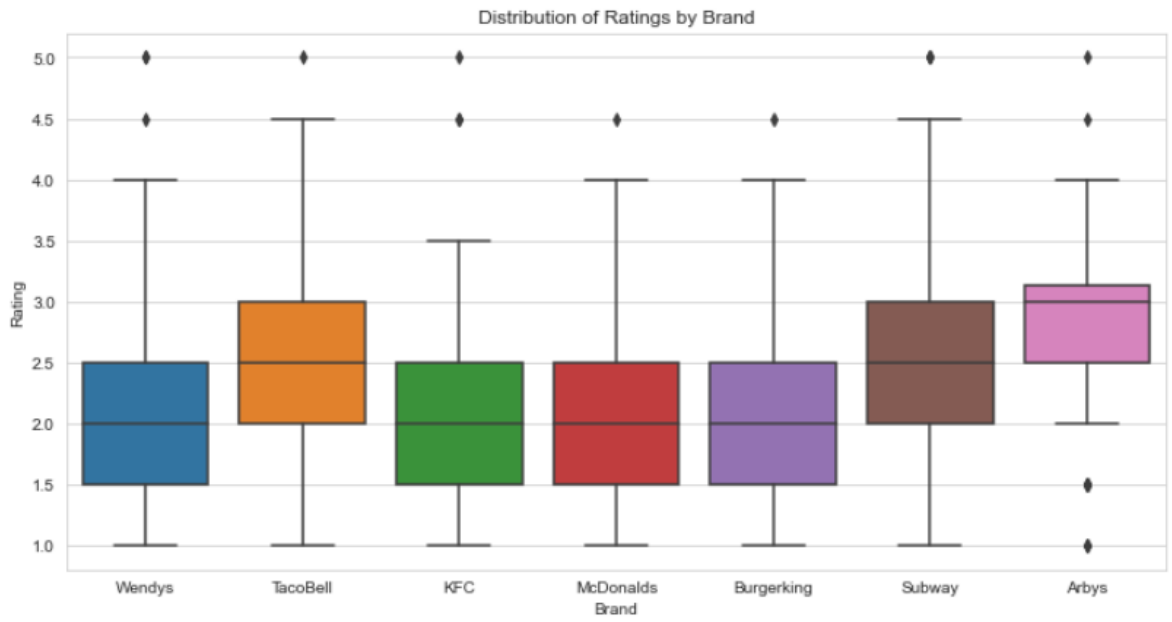


Here is an illustration of the Average Ratings and Number of Reviews by Brand.

The image describes the number of ratings and average ratings of three fast food chains: Wendy's, McDonald's, and Taco Bell. According to the analysis, Wendy's has the highest number of ratings, and

their average rating is 5. McDonald's has the most ratings in the range of 1.5-2, and Taco Bell has the most ratings in the range of 2.5-3.

From this analysis, we can infer that Wendy's has a higher overall customer satisfaction compared to McDonald's and Taco Bell, as their average rating is the highest among the three. On the other hand, McDonald's has a large number of low ratings, which might indicate a problem with their food or service. Similarly, Taco Bell has a large number of ratings in the range of 2.5-3, which might indicate a less satisfied customer base compared to Wendy's. Overall, this analysis can provide insights into the relative performance of these fast-food chains in terms of customer satisfaction.

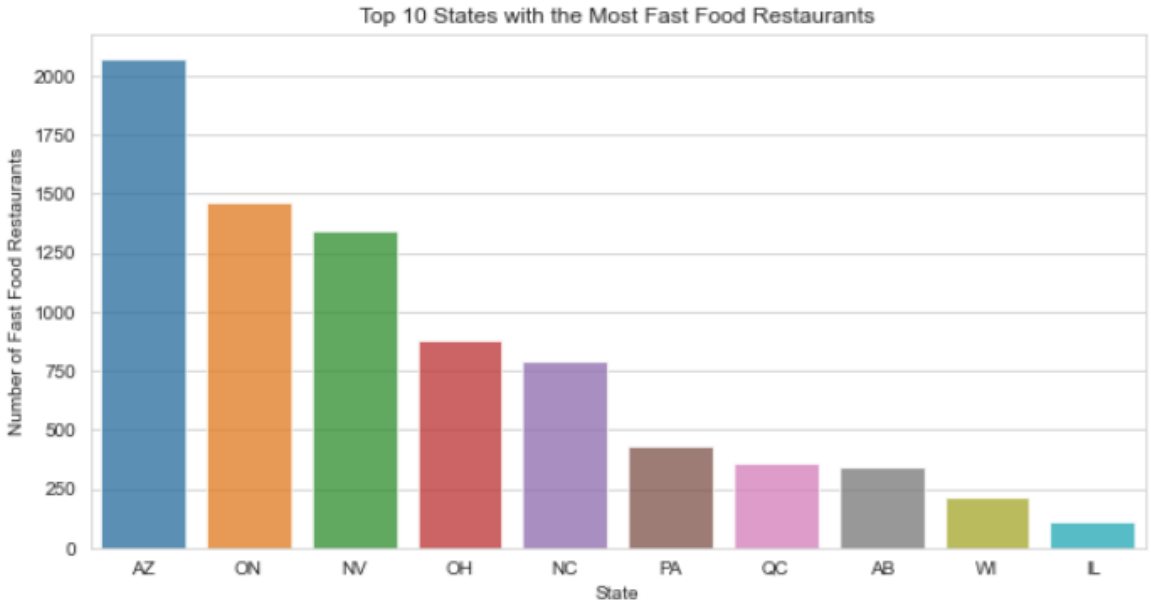


The given image talks about the distribution of median average ratings for different fast-food brands. A box plot is used to show the distribution of ratings for popular fast-food brands. The interpretation of the box plot reveals that:

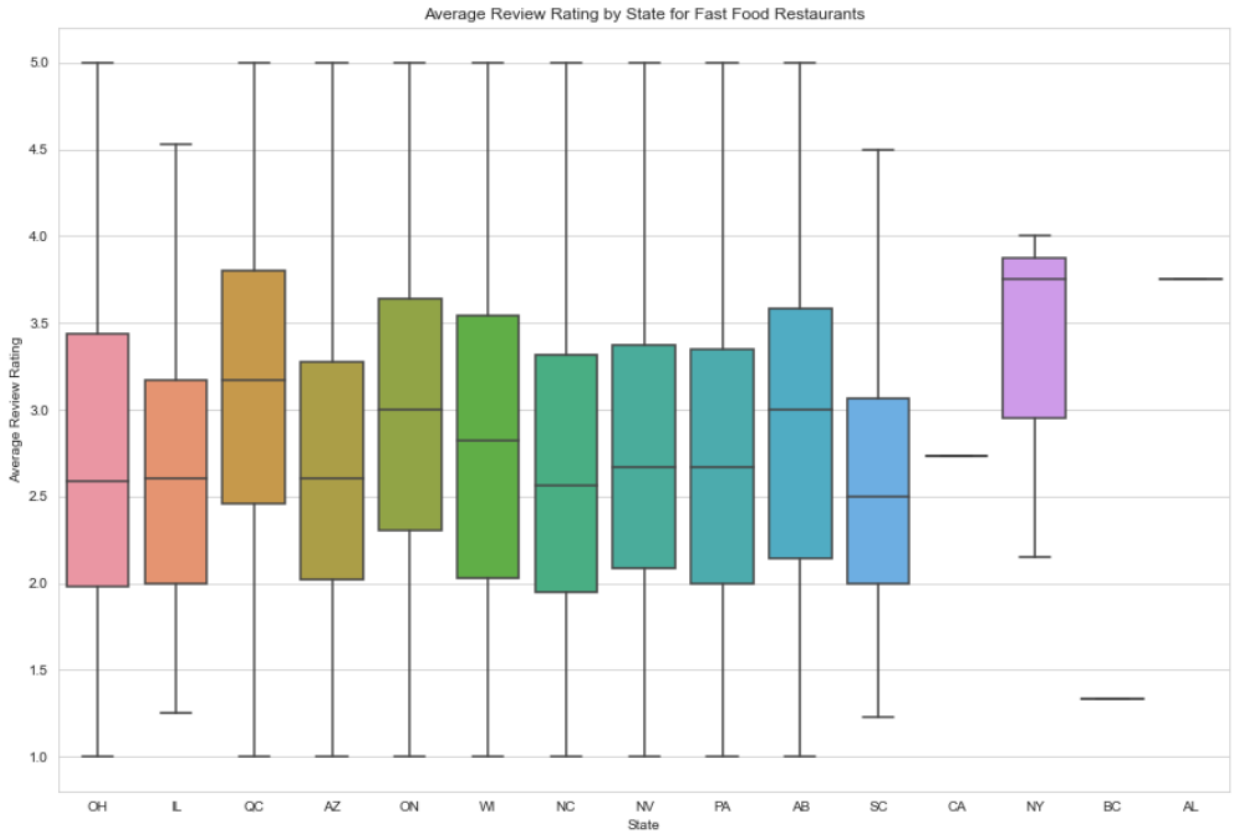
- Wendy's has the majority of ratings between 1.5 to 2.5, indicating that customers have mostly given them lower ratings.
- TacoBell has the majority of ratings between 2.5 to 3, indicating that customers have given them relatively higher ratings.
- KFC, McDonald's, and Burger King have the majority of ratings between 1.5 to 2.5, indicating that customers have mostly given them lower ratings.
- Subway has the majority of ratings between 2-3, indicating that customers have given them an average rating.

- Arby's has the majority of ratings between 2.5-3.2, indicating that customers have given them relatively higher ratings.

These inferences could suggest the popularity and overall satisfaction of customers with the different fast food brands, with some being more well-liked than others. However, it is important to note that the interpretation of the box plot and ratings may be subjective and dependent on individual experiences and preferences.

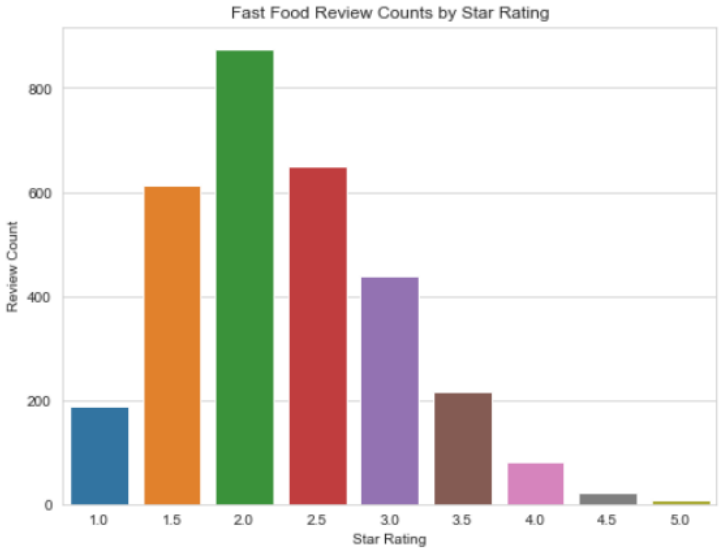


Arizona seems to have the highest number of fast food chains, and Illinois has the least number of fast food chains.



The given distribution shows the average review ratings of fast food chains across different states in the US. The statement suggests that people in New York have given higher ratings to fast food chains

compared to people in other states, indicating that the quality of food and service in fast food chains in New York might be better.



The given image suggests that among the different star ratings (ranging from 1 to 5), the 2-star rating has the highest number of reviews for all the fast-food chains included in the analysis. This is followed by 2.5-star and 1.5-star ratings. On the other hand, a 5-star rating has the least number of reviews, followed by 4-star and 4.5-star ratings.

However, the statement also suggests that there is not a clear correlation between the number of reviews and the star ratings for fast food chains. In other words, it cannot be said that fast-food chains with higher ratings necessarily have more reviews or vice versa. This could mean that factors other than the quality of food or service may be influencing customers to leave reviews, or that customers may not always rate a fast-food chain based on their overall experience. It is important to consider other factors such as location, price, or personal preferences when analyzing reviews for fast food chains.

3.4 Conclusion:

In conclusion, this project analyzed Yelp reviews for fast food chains to gain insights into customer preferences and opinions. The analysis included identifying popular menu items, positive and negative aspects of customer experiences, and overall trends in customer sentiment. The project used the pandas library in Python to import and preprocess the dataset, and data visualization techniques were used for exploratory data analysis. The findings revealed that McDonald's has the highest count of ratings compared to other brands, while Arby's has the highest average ratings. There is a positive relationship between the number of reviews received by a fast food brand and the number of chains that a fast food brand has, and there is no direct significant relationship between the number of branches and the highest

average rating of a fast food brand. The Western Region of the US has the most number of fast food brands followed by the South, Mid-West, and North-East. Overall, this project provides valuable insights for fast-food chains to improve their offerings and customer experience.

3.5 Limitations of the project and Future Work:

The analysis is limited to the Yelp dataset only, which might not represent the entire population of fast food chains in the US. We could collect data from other sources such as government databases, financial reports, and social media to provide more comprehensive insights into fast food chains in the US.

The dataset is limited to only certain attributes, and some important variables like sales, revenue, and profitability are not considered. We can perform sentiment analysis on the review text and could provide more granular insights into customer opinions and preferences.

The analysis is limited to only descriptive statistics and does not include predictive models or machine learning algorithms. We could develop predictive models using machine learning algorithms that could help in identifying factors that affect customer satisfaction and could help in improving business operations.

PRESENTATION OF WORK

We used R programming language and Python to analyze the Brand presence of various Fast food chains across America, Health inspection data provided by the County of LA Public Health for Los Angeles and Chicago and Performed a Sentiment analysis through a Yelp ratings Dataset. We used several R packages and libraries for data cleaning, wrangling, and visualization, including dplyr, tidyr, ggplot2, and reshape2.

The presentation of our work includes a detailed report on our methodology, findings, and insights. We have also provided visualizations in the form of charts and graphs to support our analysis and highlight the key findings. The report is organized in a clear and concise manner, with headings and subheadings to improve readability and comprehension. We have also included a list of references and citations to acknowledge our sources of information and data.

OVERALL CONCLUSIONS

This project analyzed Yelp reviews for fast food chains to gain insights into customer preferences and opinions. The analysis identified popular menu items, positive and negative aspects of customer experiences, and overall trends in customer sentiment. The findings revealed that McDonald's has the highest count of ratings compared to other brands, while Arby's has the highest average ratings. The Western Region of the US has the most fast-food brands followed by the South, Mid-West, and North-East. The project provides valuable insights for fast-food chains to improve their offerings and customer experience.

However, the analysis has limitations, such as being limited to the Yelp dataset, considering only certain attributes, having a small dataset, and being limited to descriptive statistics. The analysis did not include predictive models or machine learning algorithms, a deeper analysis of regional variations, oan r analysis of restaurant chain ownership.

Based on the analysis, it can be concluded that health code violations are a significant issue for fast-food restaurants in both Los Angeles and Chicago. This analysis can help private equity firms to decide which brands may need support in improving their operations and meeting the health code standards.

The insights provided by this analysis can assist private equity firms in identifying potential investment opportunities in the fast-food industry and supporting brands in improving their operations. Additionally, it can help fast-food restaurants to prioritize health code compliance and improve the overall safety and quality of their operations.

However, the analysis only considers health code violations reported by the County of LA Public Health for Los Angeles and Chicago, which may not represent the overall health code compliance of the fast food industry in these cities. The analysis does not take into account any qualitative factors that may affect the health code compliance of the brands, such as management practices or employee training. The analysis also does not include data on any fines or penalties imposed on the brands for their health code violations, which may affect their financial performance.

Links and References

- tidyverse: https://www.tidyverse.org/
- caret: https://topepo.github.io/caret/index.html
- stats: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
- psych: https://personality-project.org/r/psych/
- data.table: https://rdatatable.gitlab.io/data.table/
- Pandas documentation - https://pandas.pydata.org/docs/
- https://jakevdp.github.io/PythonDataScienceHandbook/
- Seaborn documentation - https://seaborn.pydata.org/
- Scikit-learn documentation - https://scikit-learn.org/stable/documentation.html
- Natural Language Toolkit (NLTK) documentation - https://www.nltk.org/
- Yelp dataset challenge - https://www.yelp.com/dataset/challenge
- Fast Food Industry Analysis 2021 - https://www.ibisworld.com/united-states/fast-food-restaurants-industry
- Centers for Disease Control and Prevention (CDC) - Food Safety - https://www.cdc.gov/foodsafety/index.html
- Food Safety and Inspection Service (FSIS) - https://www.fsis.usda.gov/
- Health Department websites for different cities/states (e.g. LA County Public Health, Chicago Department of Public Health)