College Park Restaurant Insights Project

Fordham Consulting Group - Lulu Yu | Uzair Ansar | Krystal Deng | Rock Lee

Table of Contents

- Introduction & Business Goal
- Data Overview & Methodology
- Analysis Breakdown
    - Cuisine Analysis
    - Sentiment Analysis
    - Pricing Analysis
    - Location Analysis
- Key Insights
- Recommendations
- Q&A

Introduction & Business Goal

1

Empowering Business Decisions with Data Analytics

Business Goal

Objective: As a consulting firm, our goal is to help Mr. Smith in making data-driven decisions for opening a successful restaurant in College Park. This includes identifying the most promising cuisine type, determining an optimal pricing strategy, and understanding key factors for maintaining a strong reputation. The ultimate goal is to help Mr. Smith establish a competitive restaurant that attracts customers and achieves long-term success.

Approach:

- Analyze Yelp review data for College Park restaurants.
- Uncover patterns, trends, and market opportunities.
- Utilizing tools such as Python and Tableau.
- Provide actionable recommendations.

Data Overview & Methodology

Empowering Business Decisions with Data Analytics

2

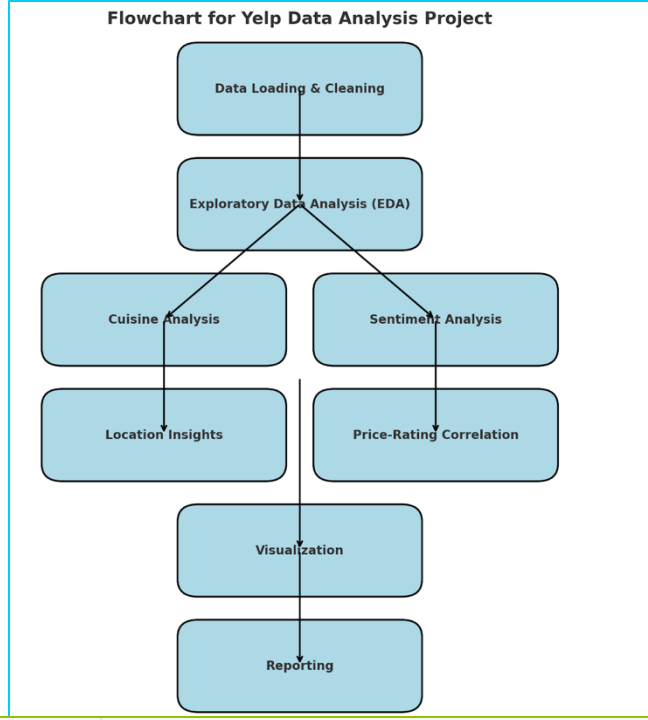
Data Overview

Datasets Used

- **restaurant.xlsx**: Includes restaurant details like rid, name, cuisine categories, address, price, and ratings.
- **review.xlsx**: Contains user reviews with ratings, dates, and textual feedback.

Preprocessing Steps

- Loaded datasets using Pandas.
- Clean missing values, standardized data types, and merged datasets on rid
- Used a dictionary to map similar cuisines together to reduce unique cuisine count.



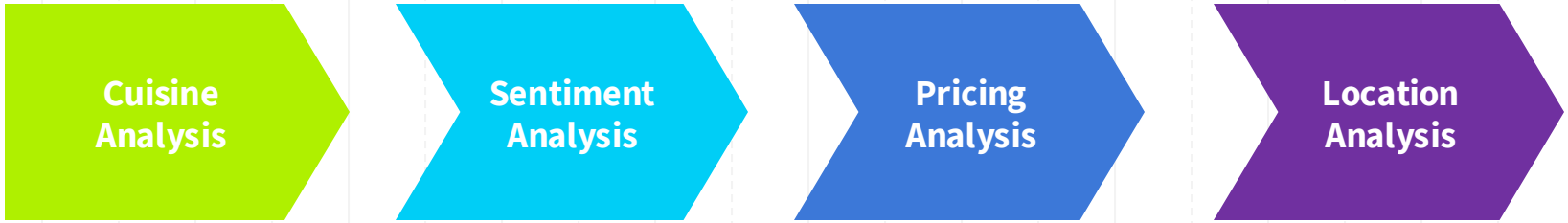

```python
# Define category mappings
category_mapping = {
    'American': 'American', 'American New': 'American', 'American Traditional': 'American', 'Southern': 'American',
    'Italian': 'Italian', 'Pizza': 'Italian', 'Pasta': 'Italian',
    'Chinese': 'Chinese', 'Dim Sum': 'Chinese', 'Cantonese': 'Chinese', 'Szechuan': 'Chinese',
    'Japanese': 'Japanese', 'Sushi Bars': 'Japanese', 'Ramen': 'Japanese', 'Izakaya': 'Japanese',
    'Indian': 'Indian', 'Pakistani': 'Indian',
    'Mexican': 'Mexican', 'Tex-Mex': 'Mexican',
    'French': 'French', 'Creperies': 'French', 'Bistro': 'French',
    'Mediterranean': 'Mediterranean', 'Middle Eastern': 'Mediterranean', 'Turkish': 'Mediterranean', 'Greek': 'Mediterranean',
    'Cafes': 'Cafes', 'Coffee & Tea': 'Cafes',
    'Fast Food': 'Fast Food', 'Burgers': 'Fast Food', 'Hot Dogs': 'Fast Food'
}
```

Key Insights & Analysis

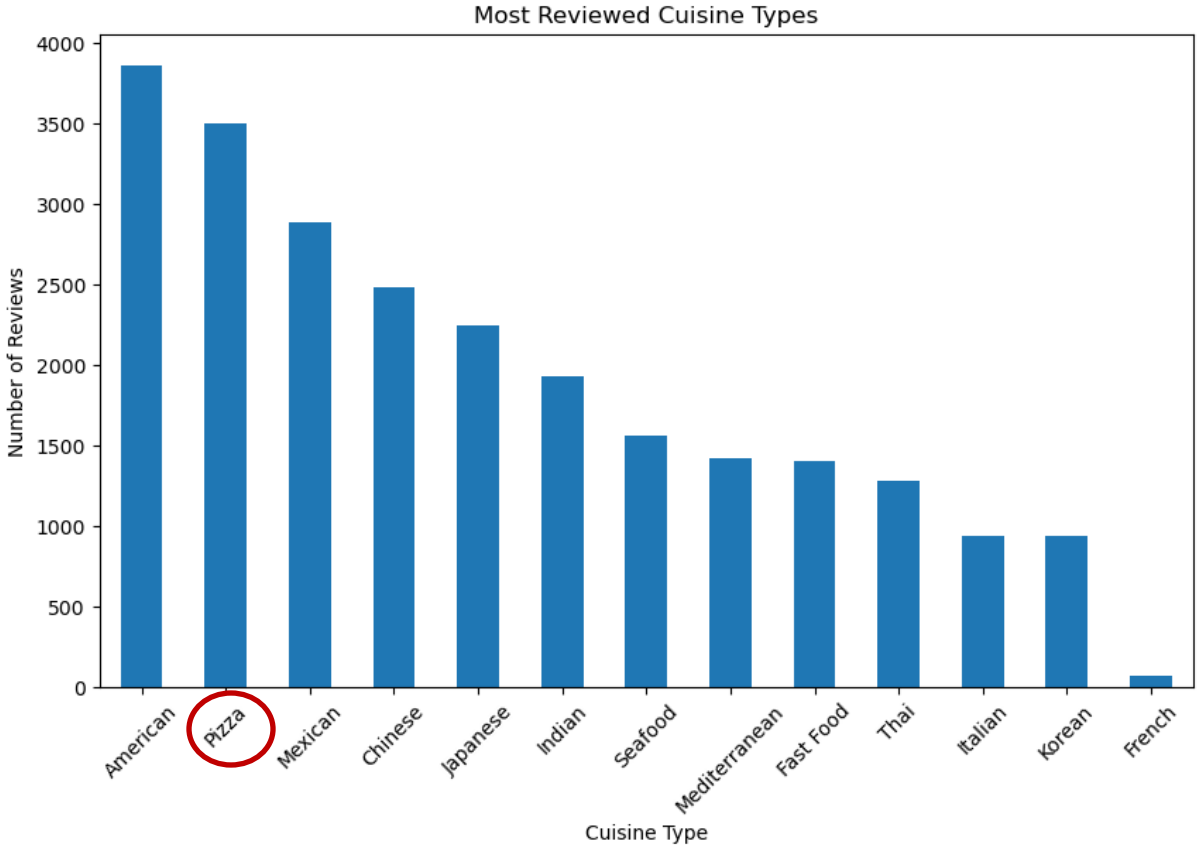
Empowering Business Decisions with Data Analytics

3

Analysis Breakdown

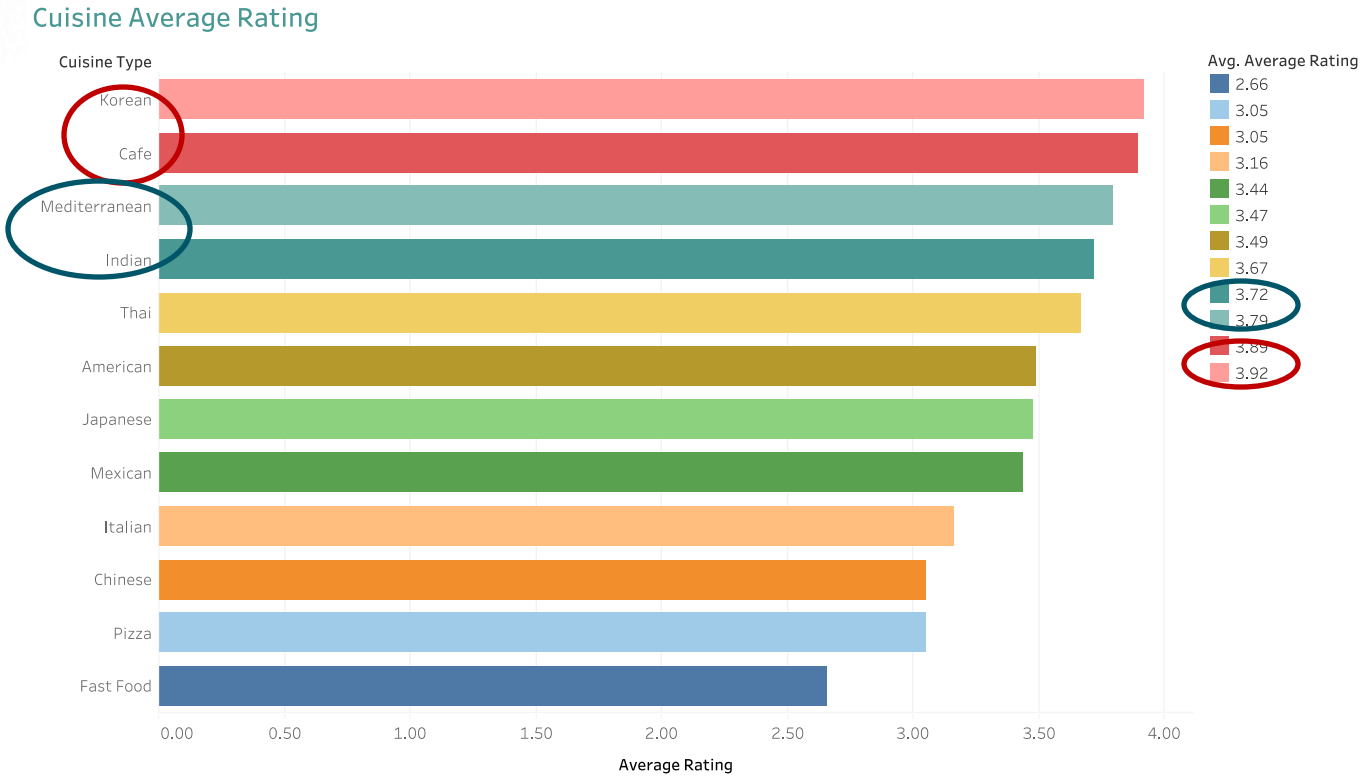


Cuisine Analysis



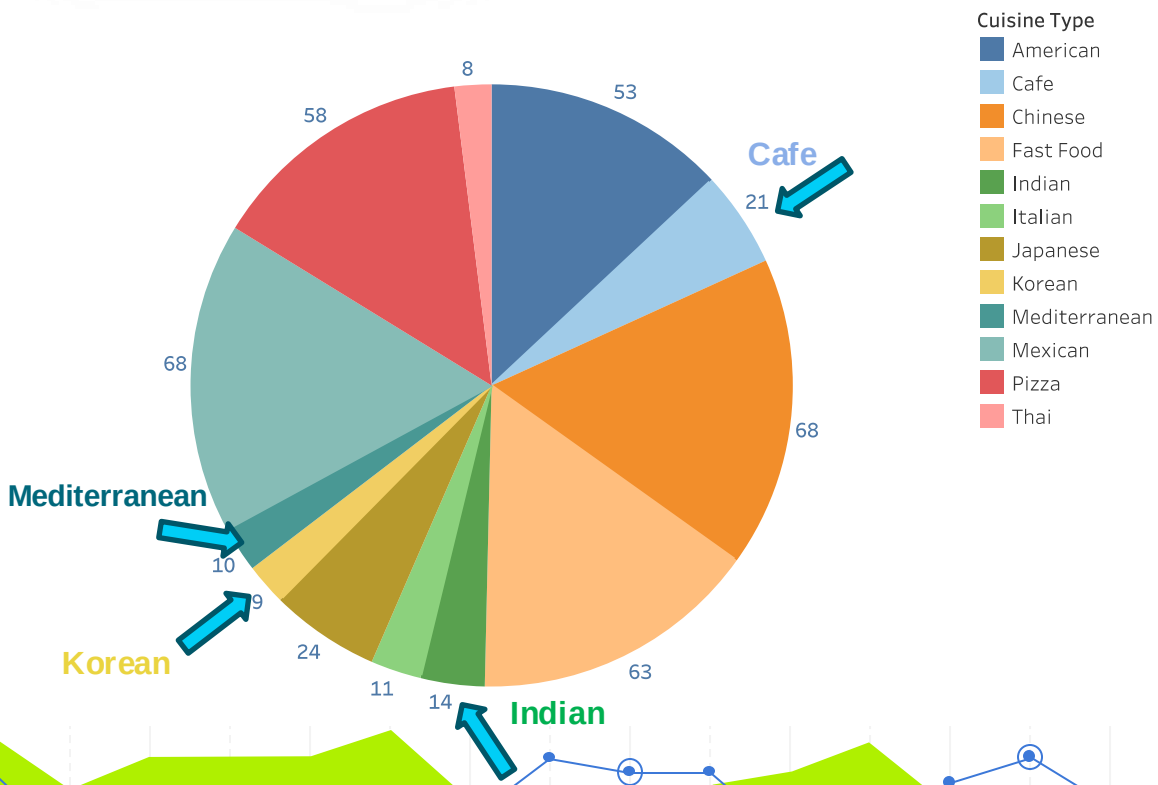
Cuisine Analysis

- Grouped data by cuisine type.
- Analyzed average ratings for each cuisine.
- Identified underrepresented but highly rated cuisines.



Cuisine Analysis

of Restaurant per Cuisine Type

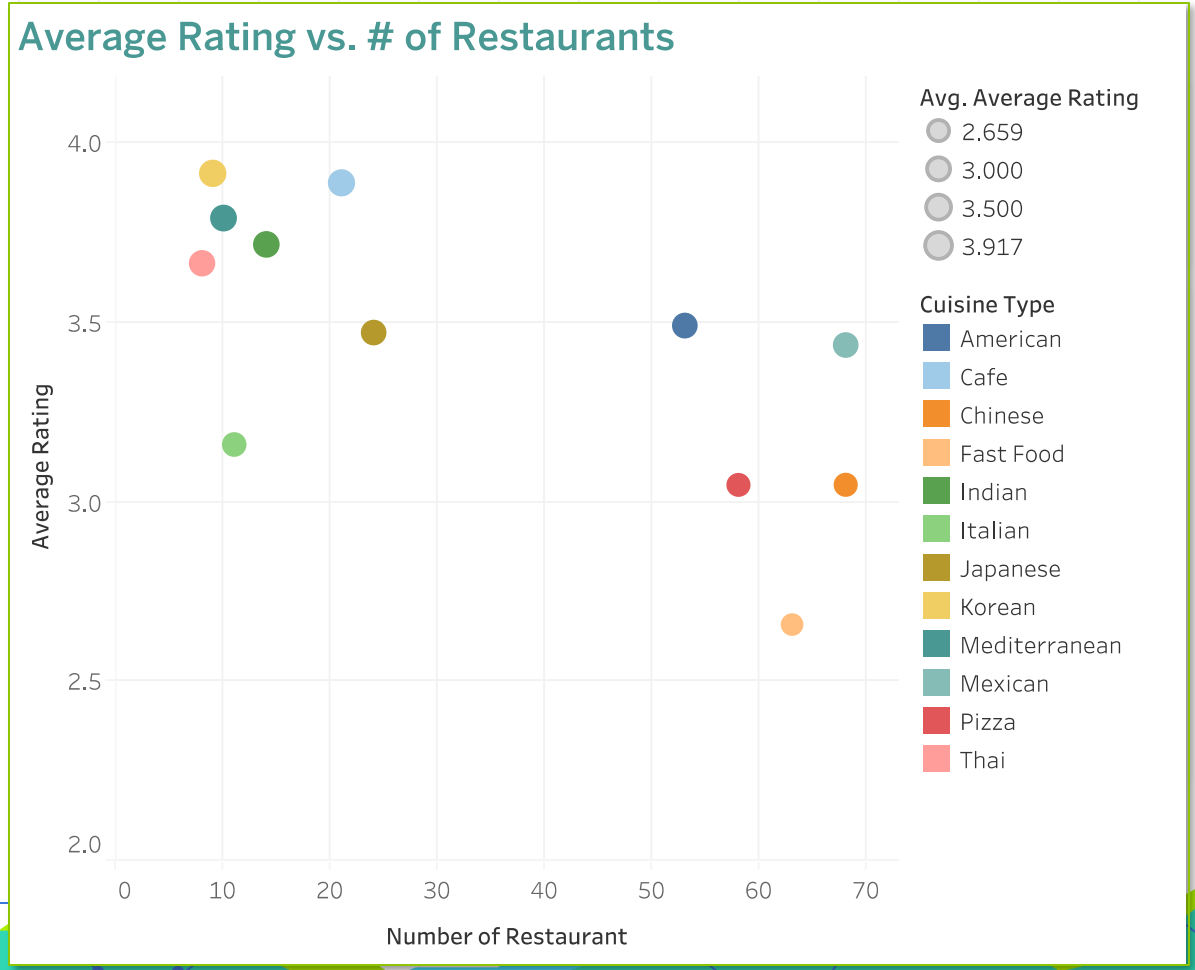


- Analyzed the number of restaurants by cuisine type to understand market saturation.
- Combined with the average ratings from the previous bar chart, **Cafe** and **Korean** cuisines stand out as highly rated but underrepresented, indicating a market gap.

Cuisine Analysis

- Analyzed the relationship between average ratings and the number of restaurants for each cuisine type.

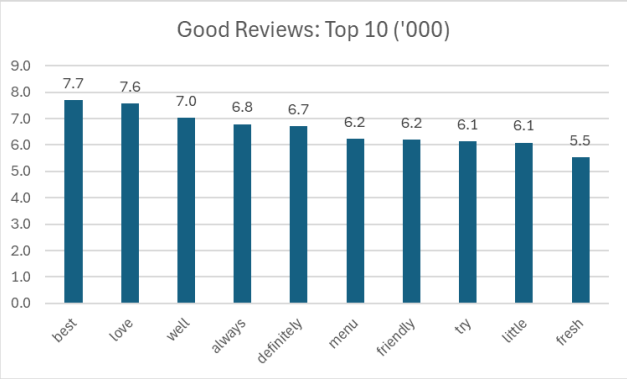
- **Cafe** and **Korean** cuisines have the **highest ratings** (around 4.0) but relatively **few restaurants**, indicating an underserved market opportunity.

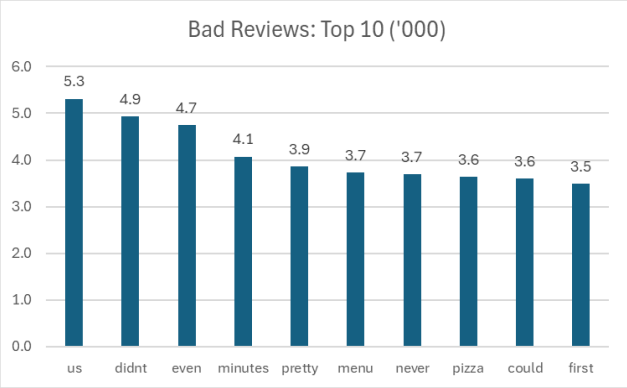


Sentiment Analysis

Sentiment Analysis

Sentiment Analysis





- Segmented reviews by good and bad reviews
  - Good Review: 4-5 star
  - Bad Review: 1-3 stars
- Removed stop words and few selected words
  - Ex.{'food', 'place', 'restaurant', 'good', 'bad','table', 'everything'}
- # of frequency of words under Good and Bad Reviews
- List of overlapping words
- Total frequency of overlapping words
- Ratio between good and bad reviews

Sentiment Analysis

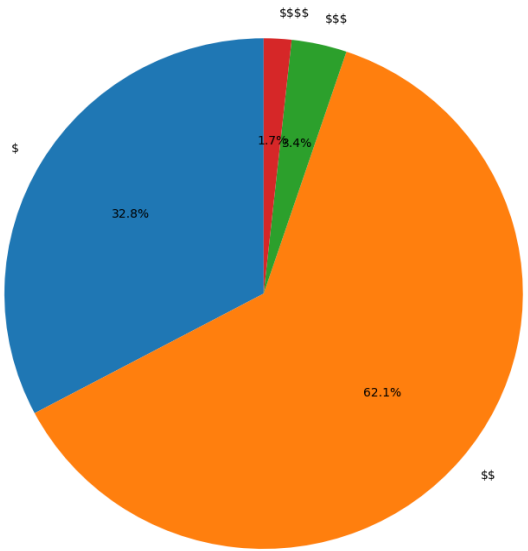
Analysis

- Having a good menu is important
- The "sauce" has to be good
- Area is important
- Friendly staff is important
- Fresh food is important
- Bar and drinks is important
- Lunch > Dinner
- Minutes – timing is important
- Experience is important
- Food – pizza, sauce, cheese, salad, fried, rice, meat, fries, drinks
- IMPORTANT: minutes (time), experience, drinks, fries, bar, meat

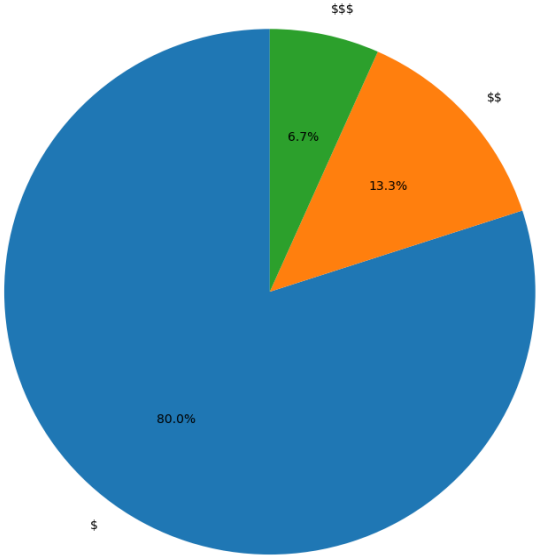
Bad % of Total	
minutes	72%
experience	47%
drinks	45%
fries	42%
bar	41%
meat	41%
pizza	40%
sauce	39%
rice	38%
menu	37%
salad	37%
cheese	37%
dinner	36%
lunch	35%
fried	33%
staff	33%
area	31%
spicy	26%
fresh	23%
friendly	21%

Pricing Analysis

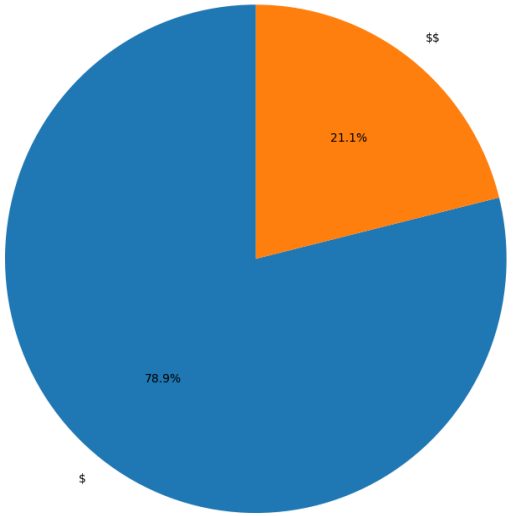
- $: Under $10
- $$: $11–$30
- $$$: $31–$60
- $$$$: Above $61



American Cuisine



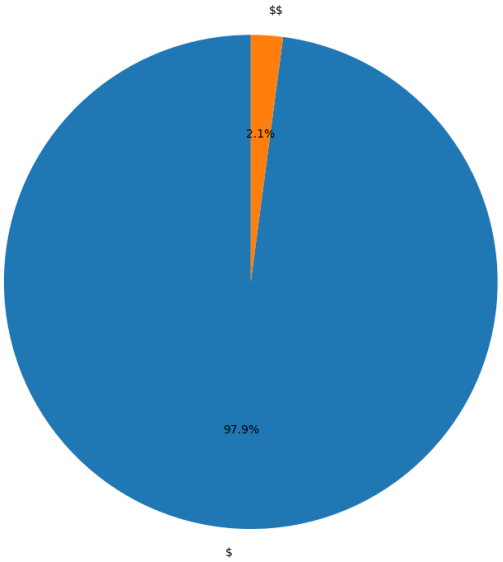
Café



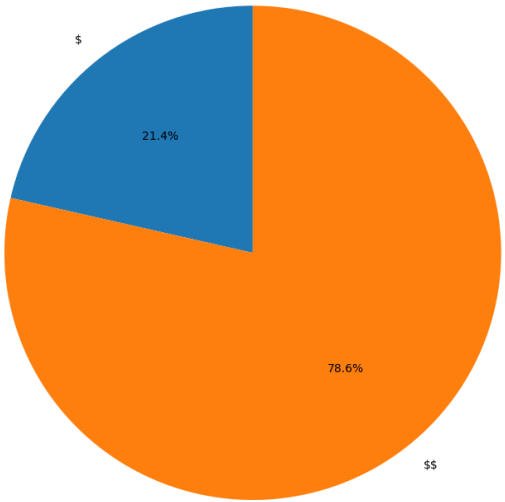
Chinese

Pricing Analysis

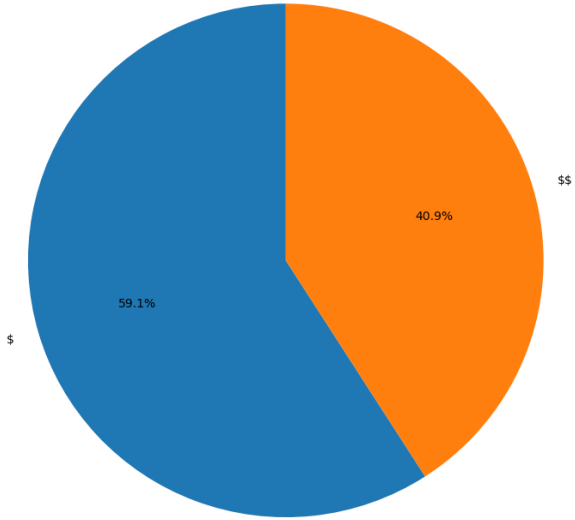
- $: Under $10
- $$: $11–$30



Fast Food



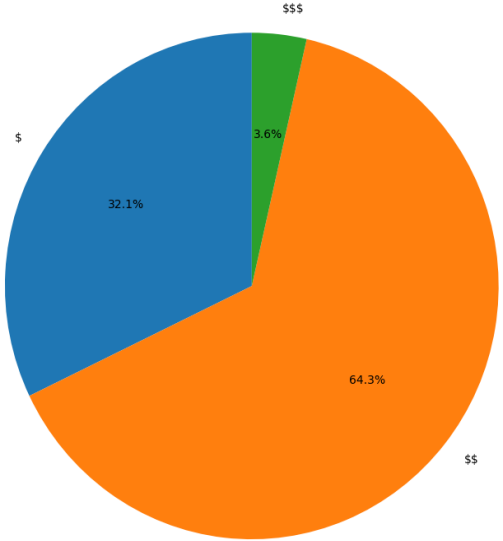
Indian



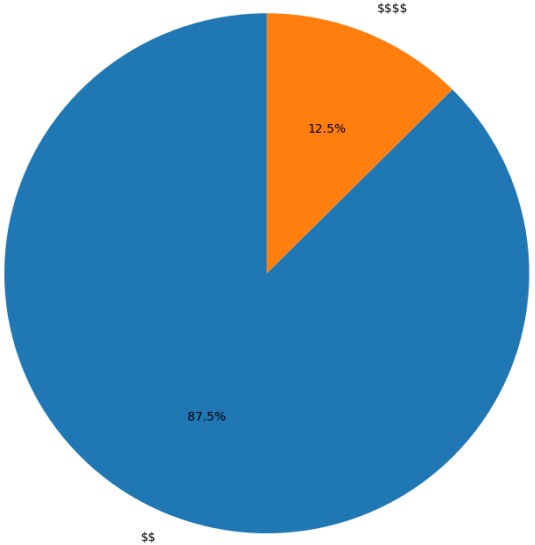
Italian

Pricing Analysis

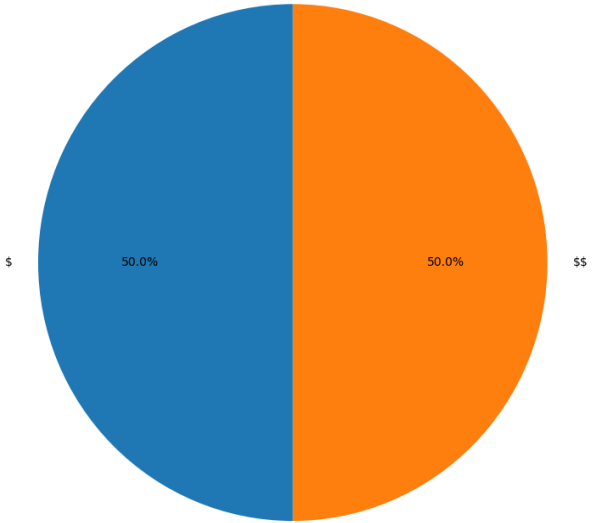
- $: Under $10
- $$: $11–$30
- $$$: $31–$60



Japanese



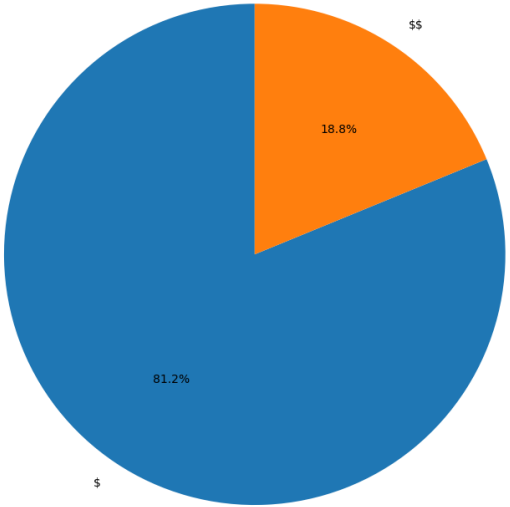
Korean



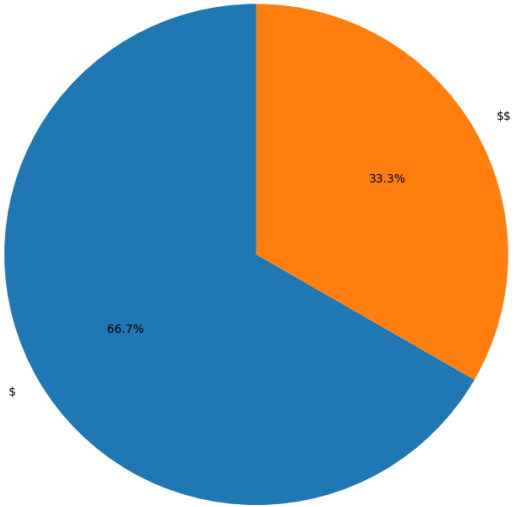
Mediterranean

Pricing Analysis

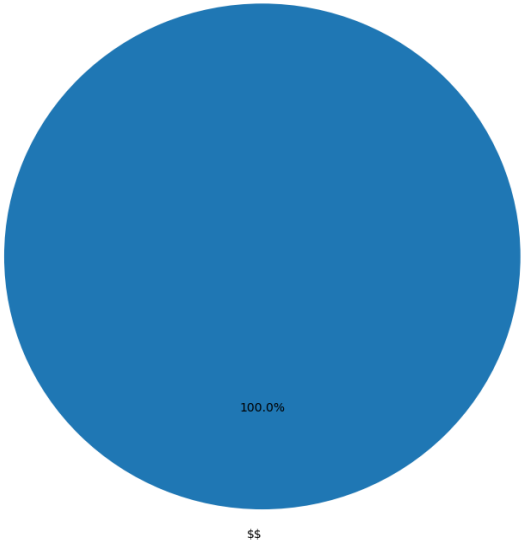
- $: Under $10
- $$: $11–$30



Mexican

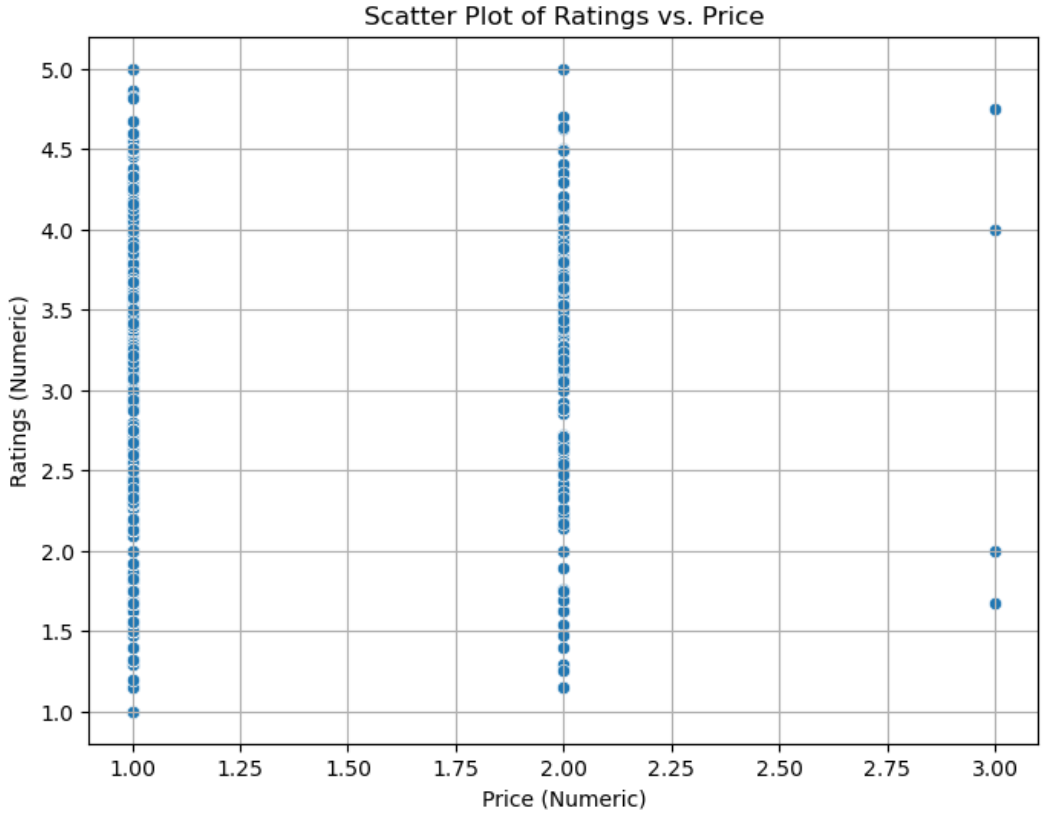


Pizza



Thai

Pricing Analysis

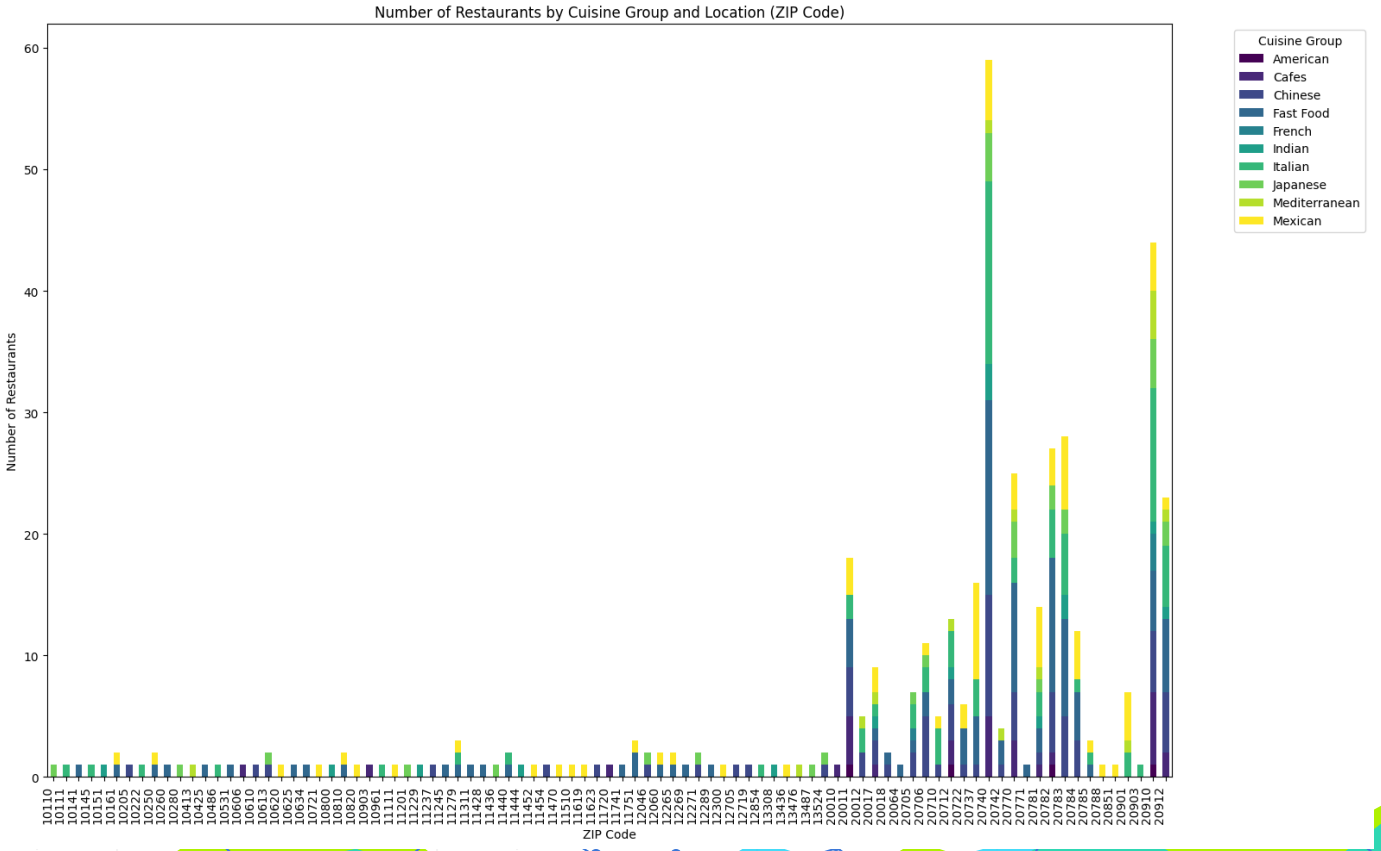


Rating and price have a weak correlation of 0.25

Location Analysis

- Plotted restaurant locations using address data
- Identified high-traffic areas with dense restaurant clusters.
- Zip Codes with high restaurant density:

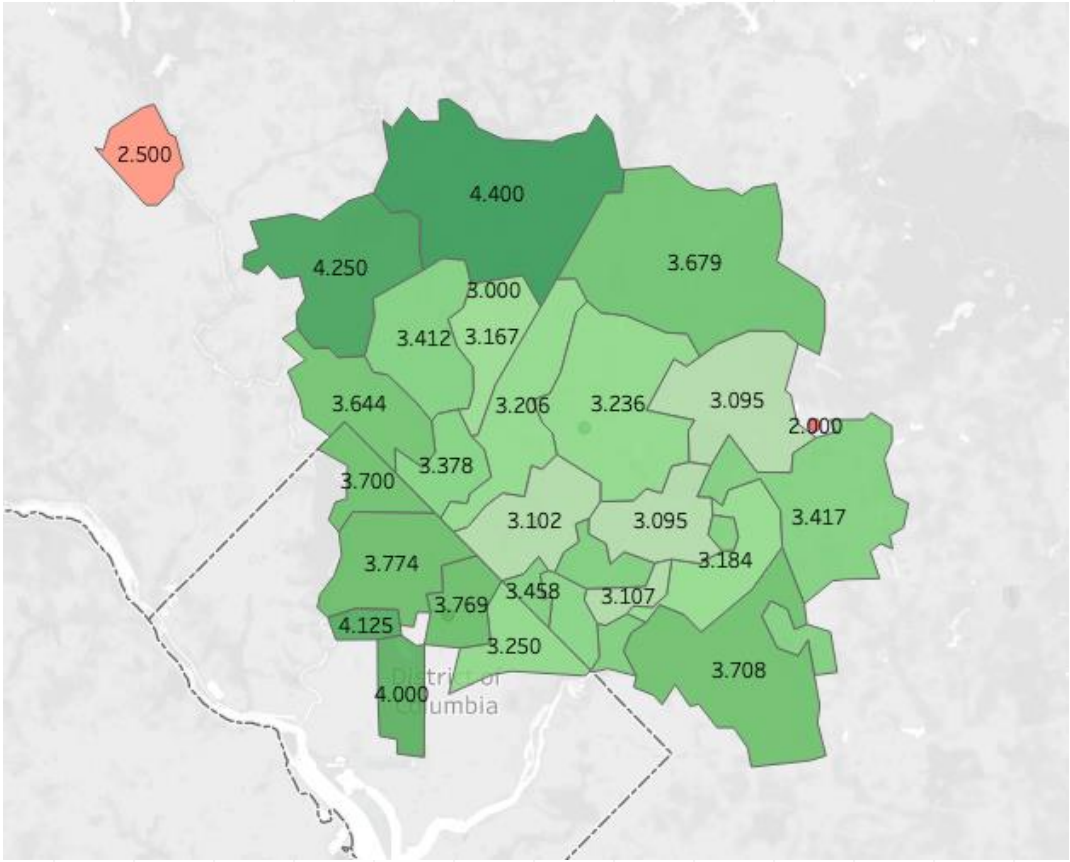
20740 | 20910 | 20783



Location Analysis

- Plotted average ratings by Zip Code on a map
- Identified high-rated areas.
- Zip Codes with highly rated restaurants:

20904 | 20902 | 20010 | 20001



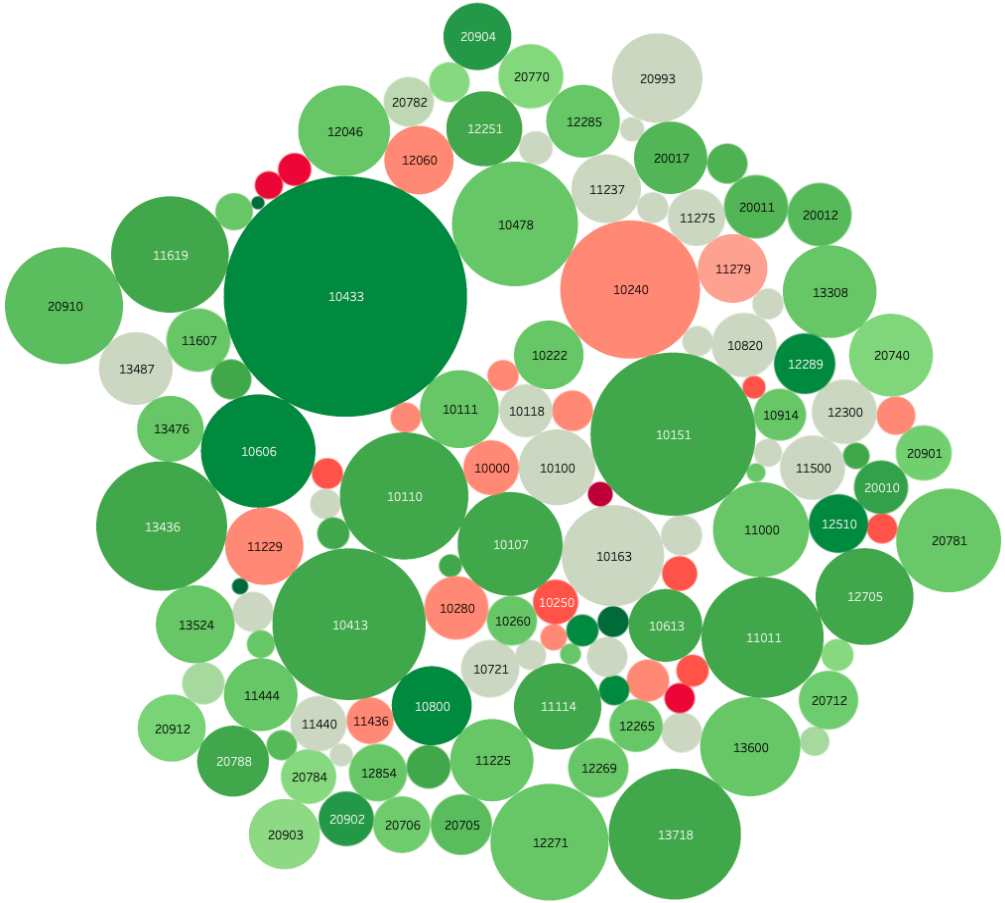
Location Analysis

- The **size** of the bubble represents the **number of reviews**

- The **Color** represents the **average rating** for that Zip Code

- Zip Codes with high reviews:

10433 | 10151 | 10413



Key Insights

4

Empowering Business Decisions with Data Analytics

Key Insights

Cuisine analysis highlights **Cafe** and **Korean** cuisines as high-potential opportunities due to their high average ratings and low market presence, indicating a clear market gap.

Sentiment analysis revealed that menu, area, staff, bar are critical. Service time, experience, drinks, fries, and meat were especially sensitive to bad ratings. Customers tend to have more opinions on lunch than dinner.

Price analysis has shown that majority of the restaurants in the area is with the price range under $10, and price has low correlation with the ratings.

Location analysis shows that certain ZipCodes perform better in terms of review count and rating:

20740 – highest # of restaurants

10433 – most # of positive reviews

20904 – highest rated restaurants

Recommendations

5

Empowering Business Decisions with Data Analytics

We recommend opening a Korean or Café restaurant

Cuisine

Korean or Cafe has low presence and high ratings.

Pricing

Focus on $10 pricing with lunch specials for college students.

Location Targeting

20740: High restaurant density for traffic.

10433: Most positive reviews show loyalty.

20904: Highest-rated reflects quality preference.

Experience

Menu | Service | Ambiance

Q&A 6

THANK YOU!

Your Restaurant’s Future with Data Starts Here!