Zomato Restaurant Analysis

Executive summary

- **Western restaurant brands (like Domino's, KFC, and Subway) are leading in individual sales amounts (over ₹9M)**
- **May be of interest to focus on encouraging growth in Chinese cuisine and expanding Western offerings.**
- **By keeping average costs between ₹265 and ₹339, customers will be most likely to rate their purchases over 3.5.**

Introduction

As a junior analyst for Zomato, I was tasked to analyze the business performance of restaurants registered in the service. My project scope and objective for the Zomato dataset was specifically focused on the restaurant analysis. The purpose of analyzing the restaurants was to explore the data and find out what restaurants are the most popular, which ones have the highest revenue, finding any correlations between cuisine and revenue, and other key questions outlined further in the report.

Methodology and documentation

The data source I used was the Zomato_data.zip file that was provided to me for this analysis. Included in the archive were several tables, however my project scope and objectives led me to focus in on two tables in particular: 'restaurant.csv' and 'orders.csv'. After reviewing both tables, I combined both sheets into one Excel file, 'restaurant_and_orders_cleaned.xlsx', and began cleaning the data before importing the file into Tableau Public. The following data cleaning steps were applied to the raw data:

- For 'restaurant' sheet
    - Removed the first untitled column
        - Purpose: First column was just a numerical count of columns, which was not useful for further analysis.
    - Clean 'city' column to include only the primary city value, with no additional comma delineated values.
        - Purpose: Locations were aggregated at the primary city level and not at the neighborhood level.
    - 'Cost' column was cleaned to show numerical values only.
        - Purpose: Currency symbol was not displayed correctly. Proper currency value was attributed in Tableau Public
    - 'Cuisine' column was split into 'cuisine' and 'cuisine_secondary' by comma delineated values
        - Purpose: One assumption made was that each restaurant could have a primary and secondary cuisine.

        - Removed non-cuisine values, such as “Popular Brand Store”, blanks, unrelated text (operating hours, promo codes, etc.)
            - Purpose: Further data analysis required records with valid cuisine types. Any records that had an invalid ‘cuisine’ value but valid ‘cuisine_secondary’ value was edited to only include the valid cuisine in the primary slot.
        - Deleted unnecessary columns for final analysis: ‘lic_no’, ‘link’, ‘menu’.
            - Purpose: The three following columns were not used in the final analysis, either for joining purposes or for reference, and allowed for smaller file sizes when importing into Tableau.
- For ‘orders’ sheet
    - No data cleanup was necessary

The ‘restaurant_and_orders_cleaned.xlsx’ file was then imported into Tableau Public and the ‘orders’ table was left joined to the ‘restaurant’ table by the restaurant id (restaurant.id = orders.R_Id).

Data was further filtered by “Rating Count” to remove restaurants that were labeled as “Too Few Ratings”. One assumption made was that these restaurants may be too new to properly analyze. These rows were only included for the map, showing the overall distribution of restaurants, but otherwise left out of the remaining analysis.

The questions I used for research and analysis were the following:

- What is the geographic distribution of restaurants?
- What restaurants make the highest revenue?
- What restaurants are the most popular?
- What cuisines are the most popular?
- Is there any correlation between certain cuisines and revenue?
    - Hypothesis: The highest grossing restaurants will not be similar in cuisine type.
- What cities have the highest rated restaurants?
    - Hypothesis: The most populous cities will have the highest rated restaurants.
- What restaurants have the highest rated food vs the price?
- Do the least popular restaurants have something in common (high price, type of cuisine, etc.)
    - Hypothesis: The least popular restaurants will not be in the most populous cities.

The Visualizations I used in Tableau Public were the following:

- Bar charts (popular/least popular restaurants, highest revenue)
- Pie charts (share of cuisines)
- Scatterplots (cuisines vs revenue, popularity vs price, cuisine vs ratings)
- Treemaps (popular cuisines by sales)
- Maps (location of restaurants)

Data presentation

See Tableau Public presentation included with the project files.

Findings and interpretation

Through an analysis of the data, I was able to find the following answers to my research questions:

- **Geographic distribution of restaurants**
  - The use of Zomato across India is widespread, with the majority of restaurants centered around the major cities such as Delhi, Bangalore, and Pune. This was a likely outcome as major metropolitan areas often have a multitude of restaurant options as compared to more rural areas.
- **Highest revenue generating restaurants**
  - By far the highest grossing restaurant on Zomato was Domino's Pizza, with ₹4,381,164 in total sales. This was over two times the second highest grossing restaurant, Sweet Tooth – Cake and Dessert (at ₹1,927,900) and third highest grossing restaurant, Pizza Hut (at ₹1,777,748).
- **Most popular restaurants?**
  - Restaurant popularity can be measured by either the quantity of sales or by the average rating of each restaurant. When looking at the total quantity of sales, once again Domino's Pizza comes out on top with a substantial 7,011 total sales. Trailing behind them are Pizza Hut (3,645 total sales), Subway (3,233 total sales), KFC (3,154 total sales), and Sweet Tooth – Cake and Dessert (2,851 total sales).
  - However, if we look at the top restaurants by average rating there are several restaurants with a 5.0 average:
    - Zaika Darbarr Junction in Noida
    - Vittles in Hyderabad
    - Veg Khana Hindustani in Delhi
    - Veer G Punjabi Swag in Delhi
    - Twenty Four Seven (247) in Gurgaon
  - This data may be skewed, as restaurants with a few 5.0 rankings would weigh more when looking at the average. To avoid this, filtering "Rating Count" to restaurants with over 100+ ratings yield a more diverse list:
    - Tony Bakery in Jaipur at 5.0
    - Shri Krishna Food Zone in Katni at 5.0
    - Sunshine Shake in Rohtak at 4.9
    - Ram Fast Food in Barnala at 4.9
    - Mahi Bakery in Jaipur at 4.9
- **Most popular cuisines?**
  - Cuisine popularity can easily be measured by measuring the total number of sales for each primary cuisine type. Overall, the top cuisines are North Indian (₹73.6M), Chinese (₹41.8M), Indian (₹39.5M), South Indian (₹34.6M), and Biryani (₹32.5M).
- **Relationship between certain cuisines and revenue?**

    - I initially hypothesized that there would not be any correlation between certain cuisines and revenue. After analyzing the data, there is a relationship between cuisine type and the total amount of revenue, as Indian food (including restaurants under North Indian, Indian, South Indian, and Biryani) makes up 42% of all total revenue (₹180.3M out of ₹425.8M total). Individually however, 4 out of 10 highest grossing restaurants (Domino's, Pizza Hut, KFC, and Subway) are Western restaurants.
- **Location of highest rated restaurants?**
    - I initially hypothesized that the most populous cities would have the highest rated restaurants. When mapping the top-rated restaurants, my hypothesis was proven correct as 7 out of 10 restaurants were located in the most populous areas (Delhi, Hyderabad, Pune, and Bangalore)
- **Relationship between the highest rated food vs the price?**
    - One of the more interesting data points in my analysis was charting the average restaurant rating against the average cost. Interestingly enough, there was a slight positive relationship between rating and price. Food that was, on average, rated 3.5 or higher had an average cost between ₹265 and ₹339. There were some interesting outliers as well, with one data point showing a low 1.2 average rating with an average cost of ₹338.
- **Least popular restaurants commonality?**
    - My final hypothesis for this data was that the least popular restaurants would not be in the most populous cities. When analyzing the data, I found that this hypothesis was mostly true, as 5 out of 10 of the lowest rated restaurants were in comparatively small cities (such as Chandigarh, Varanasi, Allahbad, Balasore, and Gurgaon).

Recommendations

- For increased revenue potential, look to expand the offerings of Western restaurants in the most populous cities.
- Outside of Indian cuisines, Chinese cuisine is the second most popular with potential to grow. Look into possible ways to boost restaurants offering this type of food.
- Encourage users to rate their purchases with incentives or require ratings before another purchase through Zomato. This will help future analysis of upcoming restaurants and refine the offerings currently on Zomato.
- Provide feedback to restaurants that the sweet spot for average cost vs rating is between ₹265 and ₹339. Promote meals from restaurants that fall within this zone, if possible.

Conclusion

In conclusion, Zomato's restaurant offerings are quite diverse and have the potential to increase revenue if the recommendations in this report are taken into account. Overall, individually Western restaurant brands are leading in individual sales amounts but there is a large number of Indian restaurants that make up the bulk of sales by cuisine. Cities are, by their nature, the most diverse in restaurant count and offerings, but there is potential in expanding to smaller secondary cities. It may be of specific interest to focus on encouraging growth in Chinese cuisine and

expanding Western offerings. Finally, by keeping average costs between ₹265 and ₹339, customers will be most likely to rate their purchases much more highly than meals outside of that target zone.

Appendices

Additional items can be found in the accompanying zip file (original data, cleaned data, Tableau file)