Roman Dzadzic

09/30/2021

Business Presentation

Contents

We will cover the following topics

- Core business idea, objective, financial implications and approach
- Brief description of data
- Brief description of significant manipulations made to raw data
- Exploratory Data Analysis (EDA) with graphs and graphs' insights
- EDA: actionable insights – part 1, 2, and 3
- Business Insights and Recommendations: Actionable insights based on the analysis results
- Business Insights and Recommendations: Future Explorations & Recommendations

Business Problem Overview and Solution Approach

Core business idea

FoodHub is a food delivery app, which allows restaurants to receive direct online orders from customers. It has a pool of signed-in drivers who drive to a restaurant location to pick up the completed/prepared order. Once ready for pickup, delivery driver will confirm the pick-up in the app, and will travel to the customer's location to deliver the food. Driver confirms the drop-off in the app after completed delivery, and the customer can then rate the order in the app. FoodHub earns money by collecting a fixed margin rate on the total order delivery price from the restaurants.

Problem to tackle/objective

FoodHub wants to analyze the data to better understand the demand of different restaurants/cuisines which will help them to enhance the overall customer experience and achieve greater utilization of the app.

Business Problem Overview and Solution Approach

Financial implications and approach

- By enhancing the customer order and delivery experience though data analysis, such activity would likely lead to greater order satisfaction rates. This would in turn, hopefully, lead to greater utilization of the app (more orders placed and delivered via the app). Therefore, we could expect to see more orders resulting in more net revenue through larger number of marked up orders (margin rates).
- It’s worth mentioning that I took a slightly different approach to analyzing the data:
    - There are a lot of different restaurants, but not many restaurants are getting many orders/reviews, outside of maybe the top 5-10 restaurants.
    - Even if I could make a significant impact on a single restaurant or on a small group of top restaurants, I felt I wouldn’t make a big change in the revenue.
    - Instead, I observed restaurants by seeking patterns within their cuisine categories. I considered cuisine types to be ‘parent’ level categories of many restaurants. Therefore, if I could find good patterns in the cuisine level data, I could by impacting those cuisines also impact more of the restaurants belonging to them.

Data Overview

Actual data example below

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
0	1477814	62359	Pylos	Mediterranean	35.41	Weekday	4	21	29
1	1477665	231061	Han Dynasty	Chinese	34.19	Weekday	Not given	21	31
2	1477700	60039	Blue Ribbon Sushi	Japanese	33.37	Weekday	3	30	27

Brief description of data

- The dataset has 1,898 rows and 9 columns, where each row corresponds to an order placed by a customer.
- There are 3 dtypes: float64(1 column), int64(4 columns), object(4 columns)
- There are no missing data.
- The 'day_of_the_week' column has either a label named 'Weekend' or a label named 'Weekday’.
- The 'rating' column can have numeric values or a label 'Not given'.

Data Overview

Continued brief description of data

```
RangeIndex: 1898 entries, 0 to 1897
Data columns (total 9 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   order_id               1898 non-null   int64  
 1   customer_id            1898 non-null   int64  
 2   restaurant_name        1898 non-null   object 
 3   cuisine_type           1898 non-null   object 
 4   cost_of_the_order      1898 non-null   float64
 5   day_of_the_week        1898 non-null   object 
 6   rating                 1898 non-null   object 
 7   food_preparation_time  1898 non-null   int64  
 8   delivery_time          1898 non-null   int64  
dtypes: float64(1), int64(4), object(4)
memory usage: 133.6+ KB
```

Data Overview

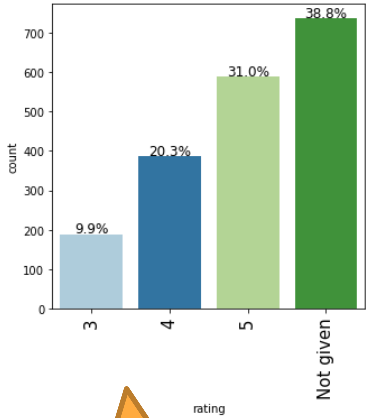
Brief description of significant manipulations made to raw data

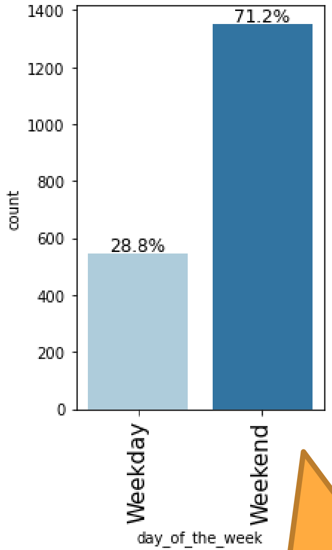
- I converted "objects" data to "category" to reduce the data space required to store the dataframe. To do this, I used the astype() function to convert 'restaurant_name', 'cuisine_type', and 'day_of_the_week' objects into categorical data.
- Upon the conversion, there were 5 numeric columns (four int64 and one float64), 3 categorical columns, and 1 object column.
- During the later part of the research (question 13) I removed observations (order records) from the raw data that had ratings marked with 'Not given'. Because 'Not given' ratings represent a lack of ratings, this action – I feel – has allowed me to have a better understanding of average ratings for various restaurants.
- Lastly, for the question 15, I have isolated two data sets from the original data, one where I had just restaurants and their mean ratings, and one where I had just restaurants and their total counts of ratings. Upon merging the two, this has allowed me to manipulate the data and understand which restaurants may be eligible for a promotion.

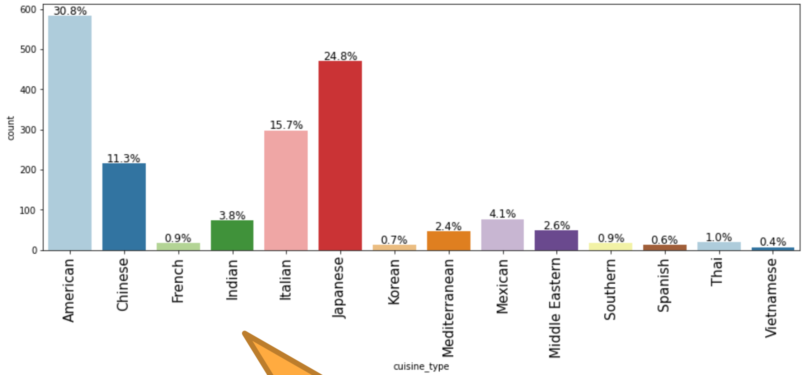
Exploratory Data Analysis (EDA)

- Graphs showing the factors most heavily impacting the target attribute
- Insights from the graphs showing the factors most heavily impacting the target attribute

Exploratory Data Analysis (EDA)





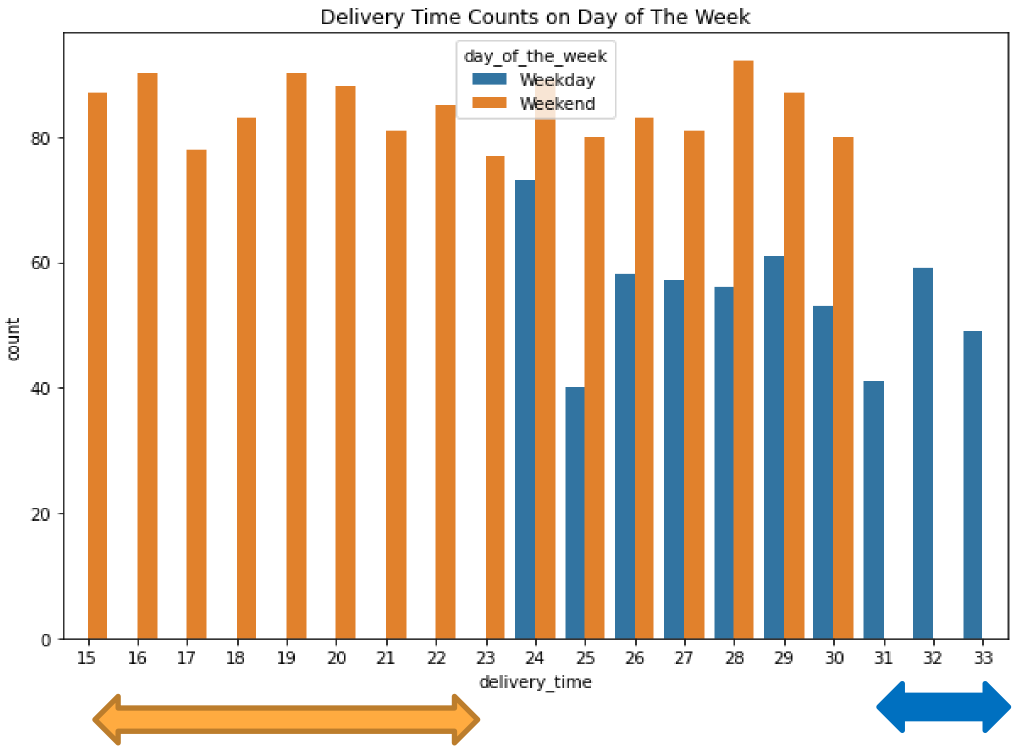


Nearly 40% of all orders are not rated. Orders rated 5 & 4 combined represent 51.3% of all orders.

Just above 71% of all orders happen on weekends.

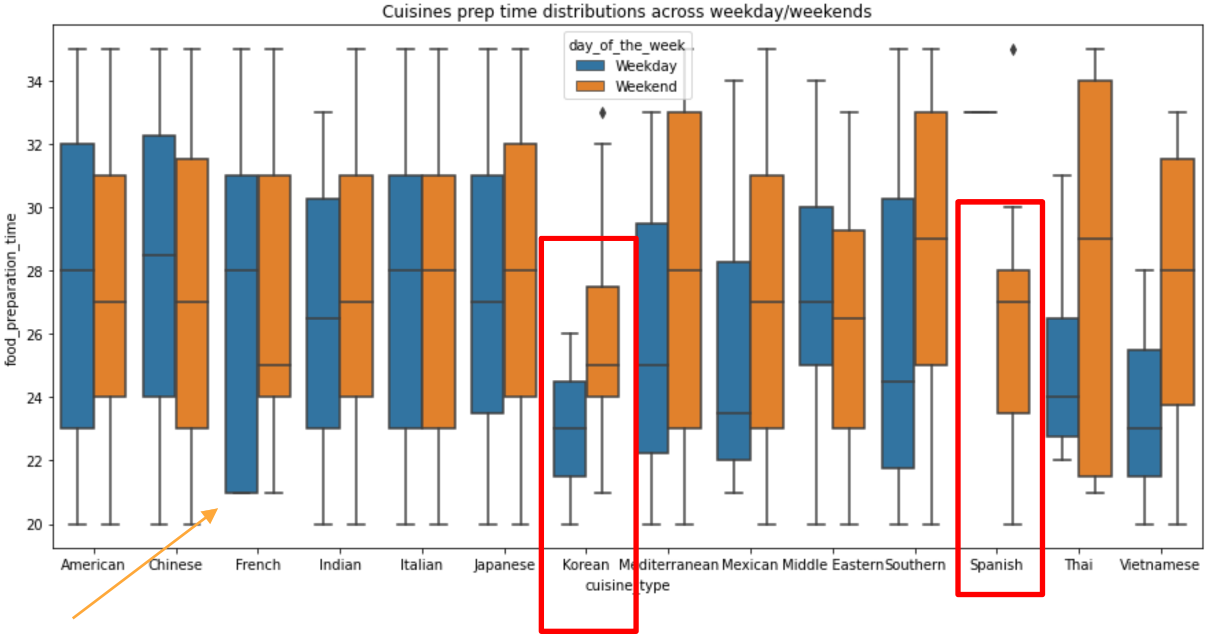
American, Japanese, Italian, and Chinese are the most popular cuisines.

Exploratory Data Analysis (EDA)



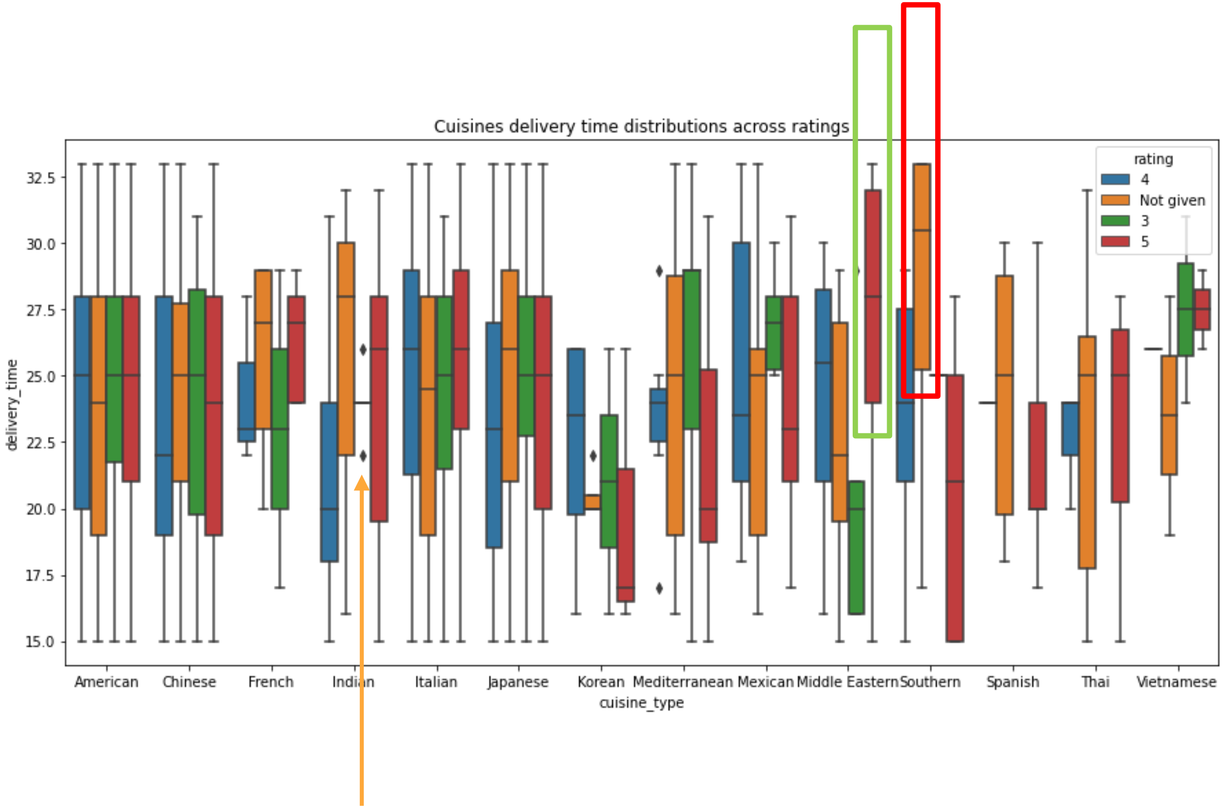
- From this countplot, we can see that orders between 15 and 23 minutes happen only on weekends. This is also likely indicating that the traffic on those days is more favorable to our delivery times.
- Similarly, we can see that a count of very high delivery times (between 31 and 33 minutes) happens only on weekdays.

Exploratory Data Analysis (EDA)



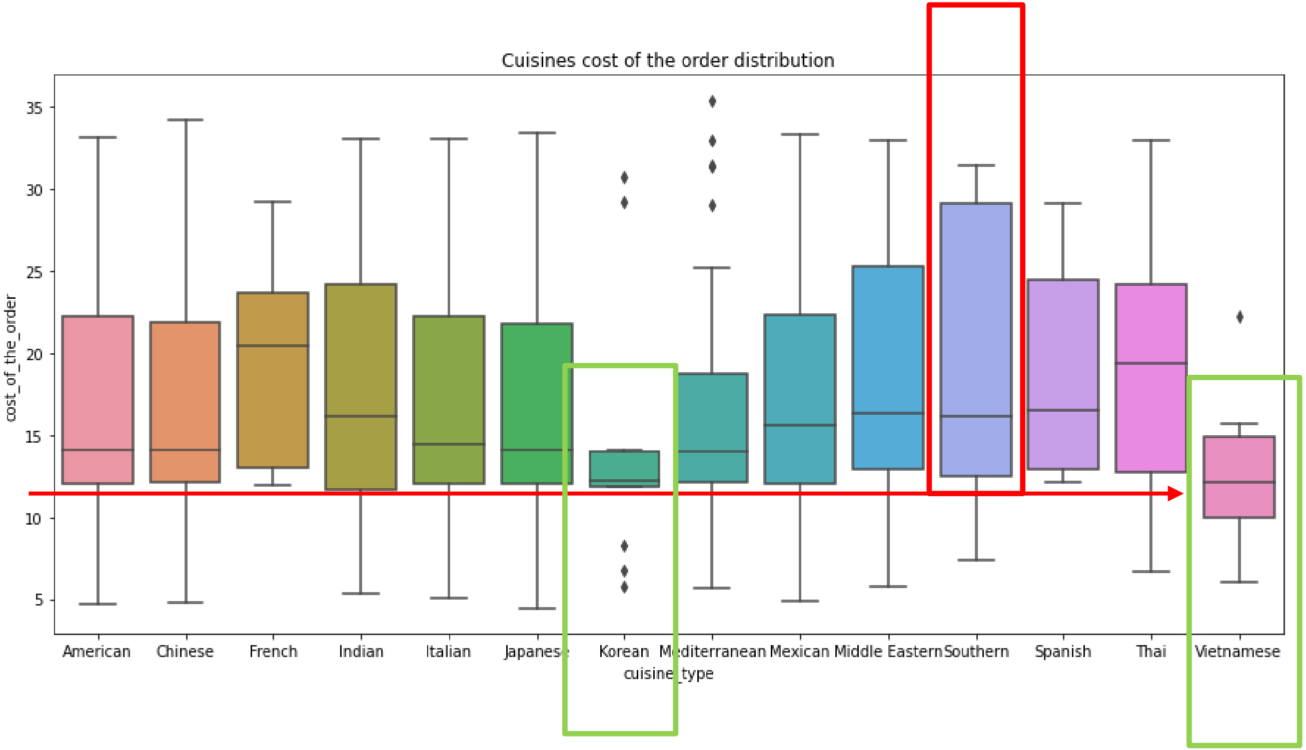
- While most cuisine types have similar food preparation times, Korean seems to lead in terms of record food preparation speeds on both weekdays and weekends.
- Thai cuisine has the biggest IQR range on weekends, from 22 to 34, while weekdays are one of the lowest ranges of all cuisines, except for Vietnamese and Korean.
- Spanish cuisine seems almost completely gone from weekdays, while on weekends they have decent food preparation times.
- And last but not least, French cuisine has the lowest 25th percentile of all cuisines at 21 min on weekdays.

Exploratory Data Analysis (EDA)



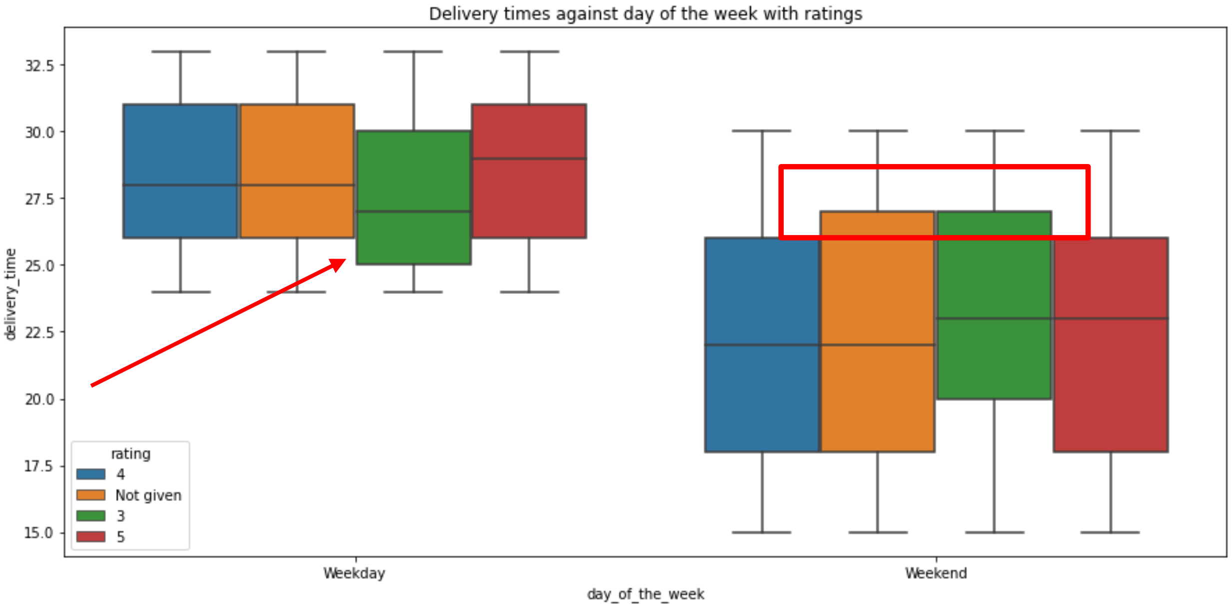
- Interestingly, for the top 4 cuisines (American, Japanese, Italian, and Chinese), they have consistent delivery time distributions across all 4 ratings.
- Indian cuisine receives very few ratings of 3, just like Southern, while Spanish, and Thai get no 3s.
- Middle Eastern cuisine receives 5s at highest IQR delivery times distribution while lower delivery time distributions have mostly rating of 3 (seems very counterintuitive), but perhaps long times spent on waiting for food are associated with the quality of food?!?
- Meanwhile, Southern cuisine 'Not given' IQR seems to be associated almost exclusively at 26 minutes and higher.

Exploratory Data Analysis (EDA)



- There is an interesting "floor" in the 25th quartile of most cuisine types except for the Vietnamese cuisine.
- Korean and Mediterranean cuisine restaurants have a few outliers.
- Korean and Vietnamese cuisine restaurants appear to have the smallest range of menu prices.
- Korean and Vietnamese cuisine restaurants seem to have a ceiling of upper quartile at around 15 dollars.
- Southern cuisine has the largest IQR.
- Only French cuisine has its mean closer to the 75the percentile.

Exploratory Data Analysis (EDA)

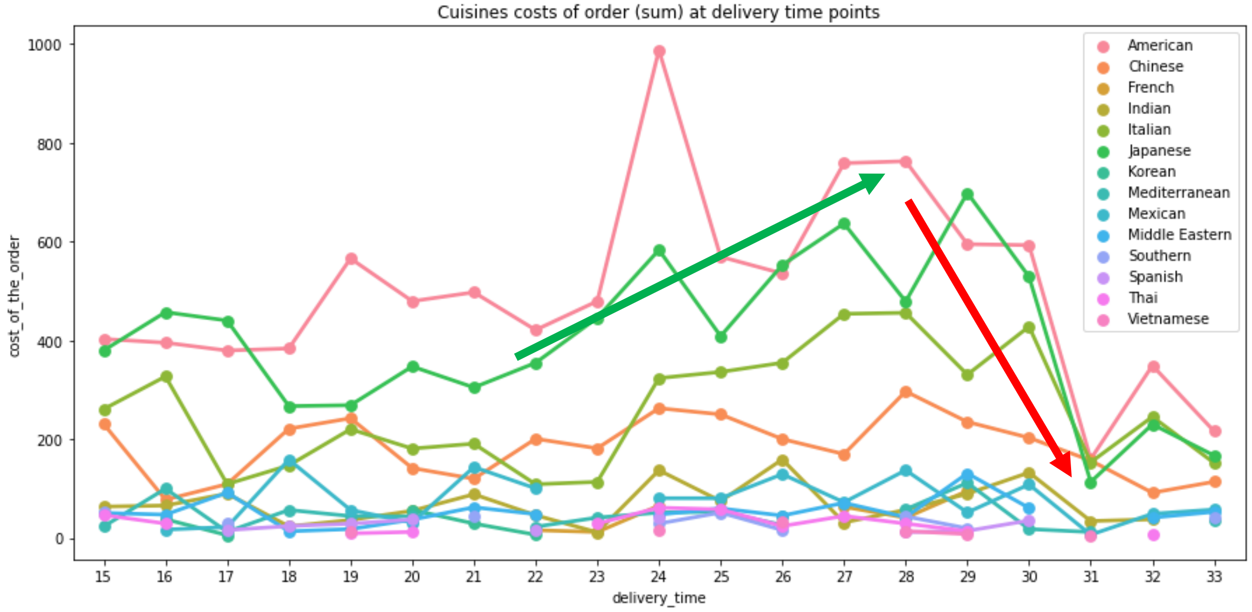


- On weekends, food delivery times with slightly higher 75th percentile distributions seem to represent lower ratings of 3 or 'Not given' ratings. We could intuitively assume that during weekends, slightly longer upper percentile delivery times are more prone to the lower rating marks -- below 5, or as a sign of dissatisfaction, orders are simply not rated.

- On weekdays, delivery times seem to have a type of a sensitivity to rating of 3 even with IQR from 26-30 being lower than the rest of the ratings.

Exploratory Data Analysis (EDA)

This chart shows the sums of all orders at a particular delivery time point. For example, we can see that the American cuisine has a sum of all orders equal around 1,000 dollars on delivery times that take around 24 minutes.



- American, Japanese, Italian, and Chinese cuisines have the highest cost of the order sums at almost all delivery time points with exception of the 31-minute delivery, where the top 3 cuisines drop below the Chinese cuisine.
- As delivery times increase, but only after the 30-minute deliveries, most cuisine order value sums drop to various extents.
- Interestingly, we also see that some cuisines are consistent, but at a lower cost of the order sum points and remain visibly unaffected by the delivery time.
- From the 15- to the 22-minute delivery time points, most cuisines have somewhat consistent cost of the order, but after the 22-minute delivery point, American, Japanese, and Italian cuisines' cost of the order sums are experiencing an increase up to a certain point (which tends to be between 29- to 30-minute deliveries) where order value sums tend to drop again.

EDA: Actionable Insights

Part I

- The company should ensure that enough delivery drivers are available in the peak traffic rush hours during weekdays, when delivery times on average take longer. It's important to add that cost of order sums for most cuisine types drops after 30 minutes marks suggesting various negative dynamics at play once delivery times pass the 30-minute mark, including less revenue.
- On weekends, delivery drivers' availability should be ensured once IQR of delivery times starts reaching between 21-26 min or extending beyond the 26 minutes mark.
- On weekdays, delivery drivers' availability should be ensured once IQR of delivery times start reaching between 25-30.
- With only 50% of all ratings on the positive side, the company should strive to improve its negative ratings in the areas such as: reduced delivery times for the restaurants in the cuisine types that are prone to lower ratings with increased delivery times (e.g. Vietnamese, Mexican, and Japanese to mention a few).
- Over 29% of all orders in the dataset are over 20 dollars -- almost 1/3. And, as a percentage of total revenue, such orders represent nearly half of all orders revenue, or 47.12% of revenue. The company should use this as a sign to push forward and experiment with small surcharges right away on larger orders, therefore contributing immediately to its bottom line revenue.

EDA: Actionable Insights

Part II

- There is an imaginary "floor" in the 25th quartile of most cuisine types at around 12.5 dollars, and yet the 75th percentile seems to vary a lot. This could suggest that a price elasticity exists around the 75th quartile, and that there is an openness of customer personas with deeper pockets to accept higher cost of orders for different cuisines. With exception of Korean and Vietnamese cuisines (restaurants), which seem to have smaller IQR ranges, and likely customers with smaller pockets, FoodHub could work with other cuisine restaurants to promote delivery specials. They could jointly push more specials out, and possible higher prices to particular persona types with higher orders costs by allowing such customers to venture out into the less explored but more unique cuisines with bigger tolerances for higher cost of orders (e.g. Southern, Middle Eastern, Indian and French).
- Since certain cuisine types (and therefore restaurants serving such foods) are more resistant to poor ratings (receive less ratings of 3), the company could push out various promotions to increase awareness and orders from such restaurants. The hope would be that even with increased demand, such cuisines would be able to continue to maintain higher order ratings - making FoodHub look good.

EDA: Actionable Insights

Part III

- To reduce the demand on certain cuisines (and their restaurants), especially during delivery peak times, FoodHub could work with under-represented cuisines type restaurants during certain times of the week to push more customers their way. For example, with Spanish cuisine being almost completely underutilized during weekdays, which are busier delivery and food preparation times, FoodHub could send more customers towards Spanish cuisine restaurants and offload some of the pressure and demand on other popular cuisine restaurants. Similarly, Korean cuisine seems to lead in terms of record food preparation speeds on both weekdays and weekends. Incentivizing more customers to order from Korean restaurants could offload some of the pressure from dominant cuisines and compensate for the higher delivery times with lower food preparation times of the Korean restaurants.

Business Insights and Recommendations

Actionable insights based on the results of the analysis

- There are 555 orders in the dataset that are over $20. Observed as percentage, over 29% of all orders in the total set are over $20, or almost 1/3 of all orders. However, observed as a percentage of total revenue, orders over $20 represent almost 50% (or 47.12%). FoodHub should explore imposing surcharges for certain order types in order to increase its chance of positive financial outcomes.
- Top 5 restaurants are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, followed by Parm. There aren't that many restaurants that receive large number of reviews. The company should attempt to work with other underrepresented restaurants with positive ratings to bring them more traffic. This would do two things: 1) it would offload the traffic in heavy peak hours from some of the most popular restaurants, 2) while also increasing the chance of more positive ratings with underrepresented restaurants.
- The mean delivery times for all orders and orders over $20 is around 24 minutes. The company should strive to keep steady at that level as much as possible, as higher delivery times (past the 30-minute mark) have shown to have lower costs of orders, and with some cuisines lower order ratings.
- Restaurants with highest costs of order, such as Pylos Mediterranean at $35.41 (order_id: 1477814) should be rewarded through various incentives. High order costs benefit everyone: restaurants and FoodHub's markups.
- Delivery mean for all orders on weekends is approximately 6 minutes shorter than the delivery mean for all orders on weekdays. This might suggest that weekdays are generally busier, and therefore it might take drivers longer time to deliver meals. To maintain consistent stream of revenue during peak hours, FoodHub should ensure strong pool of drivers during peak delivery times.

Business Insights and Recommendations

Actionable insights based on the results of the analysis

- A successful marketing campaign was executed to provide several restaurants with promotional offer in advertisements. Conditions were over 50 ratings and mean rating over 4. Sadly, not many restaurants fit this mold. FoodHub should work to reduce the campaign requirements, therefore enabling more restaurants to advertise and therefore bring in more revenue (see table on the right).

- Overall, the company helps restaurants process a total of around $31,314.82 dollars (in a given time period) from almost 1900 orders. If the company was to apply a 25% surcharge for orders over $20 AND 15% surcharge for orders between $5-$20, it would generate a net revenue of around $6,166.30 dollars in the given time period. That's around 20% markup on all processed orders.

	restaurant_name	count_of_ratings	mean_of_ratings
0	Shake Shack	133	4.278195
1	The Meatball Shop	84	4.511905
2	Blue Ribbon Sushi	73	4.219178
3	Blue Ribbon Fried Chicken	64	4.328125

Business Insights and Recommendations

Future Explorations and Recommendations

- From the further analysis, it would be great to understand which specific days (if data is available) have the most observed orders in order to understand how to distribute drivers during those days.
- It would also be great to understand which days have the best order reviews, and even at what times of day.
- Attempt to understand why as many as 38.8% of all possible ratings were not provided ('Not given'). Having more reviews and ratings would help the company have a better perception of its value, utilization, and feature usability for future improvements.
- The company should find the way to incentivize their customers to rate more deliveries, and possibly provide reasons for rating, not just rating numbers.
- There doesn't seem to be a significant impact of order delivery or preparation times on cost of orders, with exception of delivery times exceeding 30-minute deliveries. There is potentially a sign that there is a level of price elasticity that exists on the market. The company could explore and experiment with higher delivery surcharges in order to extract more net revenue from end users (restaurant's customers), given enough drivers to maintain order delivery times below 30-minute mark.
- Understand the cost of orders that take longer to prepare and deliver and use that information to reshape profit plans and methodologies. For example, orders that take long to prepare and deliver, might be charged higher delivery surcharges to compensate for poorer ratings or dropped order/customer adjustments.
- It would be great to acquire the local areas demographic data to better understand which restaurants and cuisines to incentivize and promote.

greatlearning
Power Ahead

Happy Learning !