Nato Jorjiashvili

Data Analyst, Ride-Hailing
Tbilisi, Georgia
Take Home Task

Main process of the project using Python:

1. First, I have imported the necessary libraries, such as Pandas, Matplotlib, Seaborn and NumPy.
2. I have loaded the two datasets ("Demand Data.xlsx" and "Supply Data.xlsx") into Pandas dataframes using the read_excel function.
3. I have cleaned and preprocessed the data as necessary. This included converting data types, removing duplicates, and filling in missing values.
4. Then I have merged the two dataframes based on the date field to create a single dataframe containing both demand and supply data.
5. After I have calculated the average demand and supply volume for each hour of the day, grouped by the day of the week.
6. I have created a line plot showing the demand and supply curves for each day of the week. This will give us a visual representation of the match/mismatch between supply and demand.
7. Then I have calculated the supply-demand gap for each hour and identified the undersupplied hours during the weekly period.
8. I have calculated the number of online hours required to ensure good Coverage Ratio during the peak hours identified in step 7.
9. Next, I have calculated the average earnings per hour for a driver during peak hours based on the average value of finished rides.
10. We can then set a guaranteed hourly rate for drivers that is higher than the average earnings per hour during peak hours.
11. Finally, I have created sophisticated plots to visualize the results of our analysis, such as heatmaps and interactive dashboards.

Question1: Please identify the time periods that are critical to us, i.e. when we are undersupplied, and explain your reasonings.

To identify the time periods when we are undersupplied, I have created a 24-hour curve of average demand and supply volume. This will help us see if there is any match or mismatch between the two. If the supply volume is consistently lower than the demand during certain hours, then those hours can be considered critical.

To identify undersupplied hours during a weekly period, we can calculate the difference between the supply volume and the demand for each hour of the day, for each day of the week. We can then rank the hours based on the difference and identify the top hours with the highest negative differences. These hours can be considered critical.

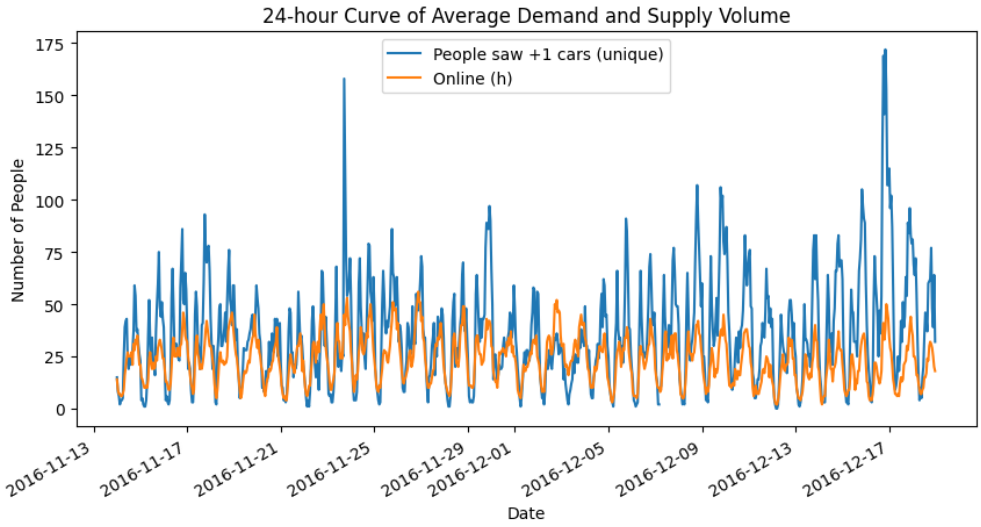

Fig.1. 24-hour Curve of Average Demand and Supply Volume

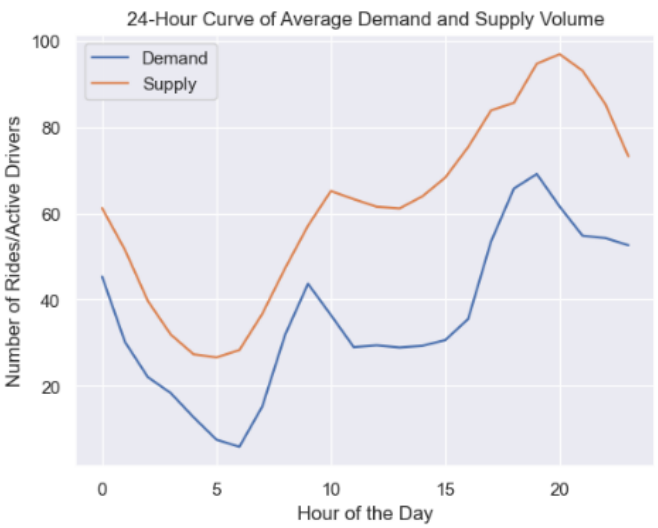

Fig.2. 24-hour Curve of Average Demand and Supply Volume

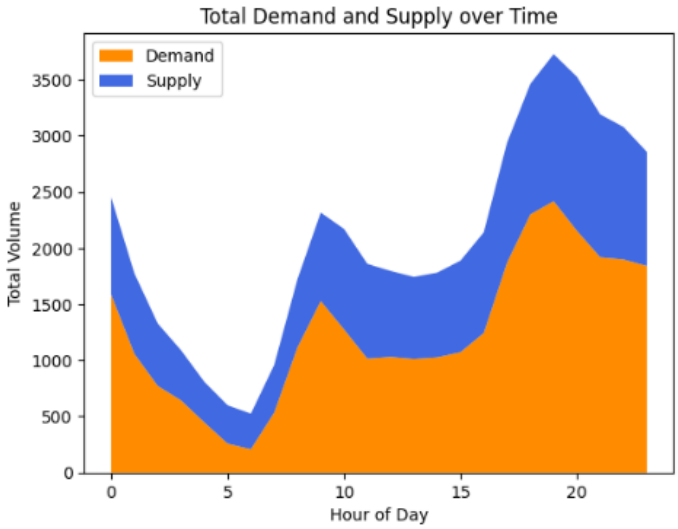

Fig.3. 24-hour total demand and suppy over time

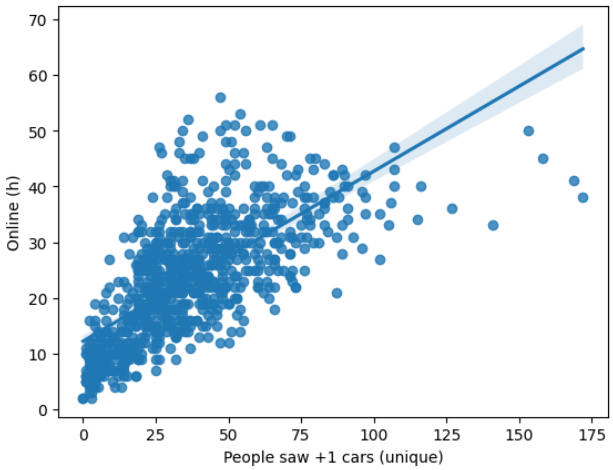

Fig.4. Online hours against People saw +1 car

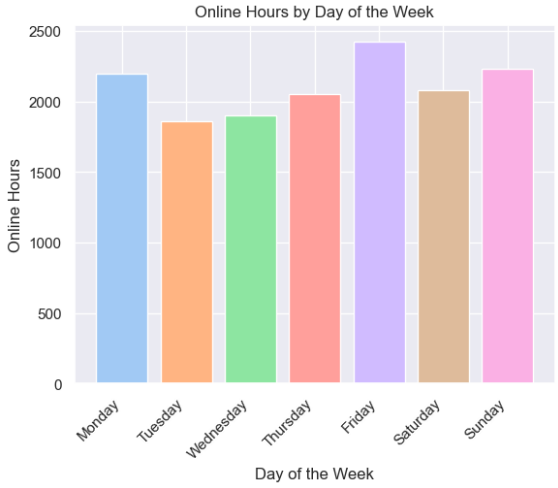

Fig.5. Online hours by Day of the Week

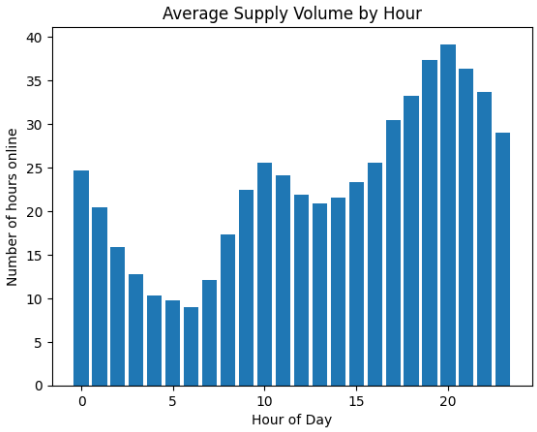

Fig.6. Average Demand Volume by Hour

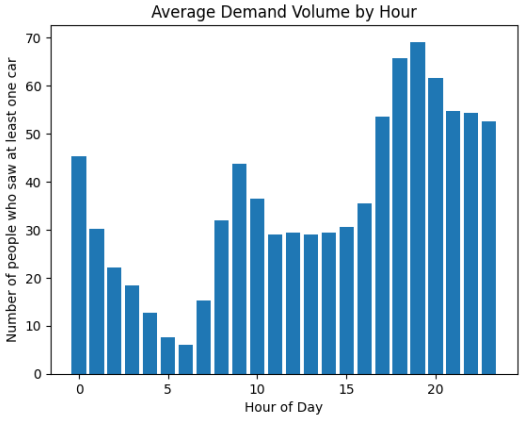

Fig.7. Average Supply Volume by Hour

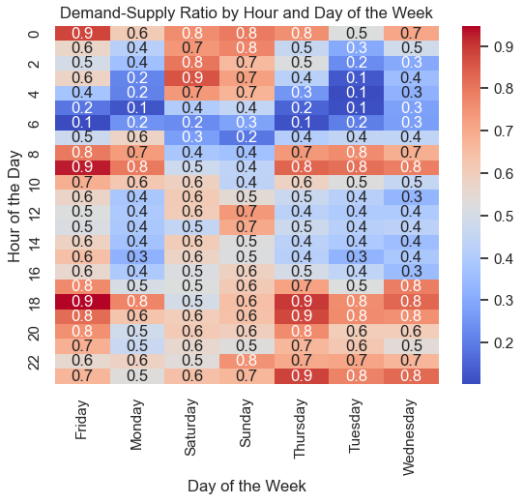

Fig.8. Heatmap of Demand-Supply Mismatch: Heatmap shows the hours of the day and days of the week where there is a significant mismatch between demand and supply. You can see critical days of the weeks and hours in blue colors where Demand-Supply ratio is less than 0.5

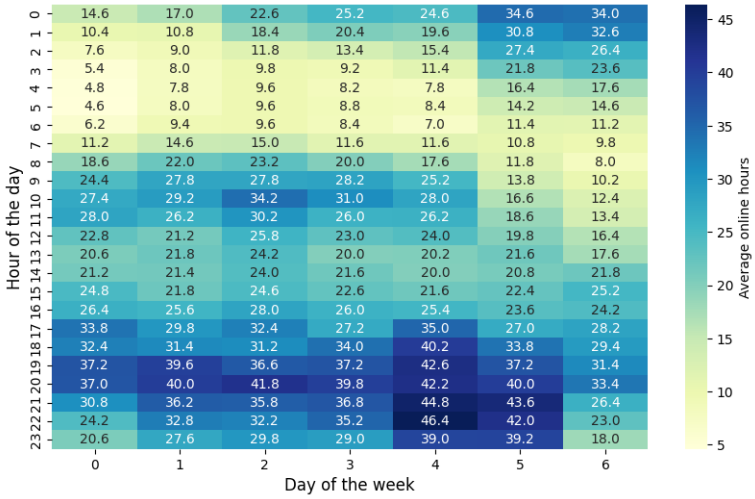

Fig.9. Heatmap of Demand-Supply Mismatch: Heatmap shows the hours of the day and days of the week where there is a significant mismatch between demand and supply. You can see critical days of the weeks and hours in blue colors

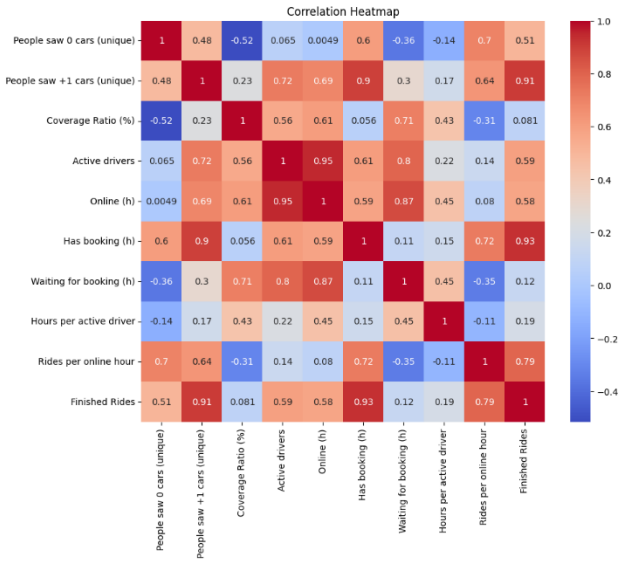

Fig.10. Correlation heatmap shows correlation between each column given in datasets

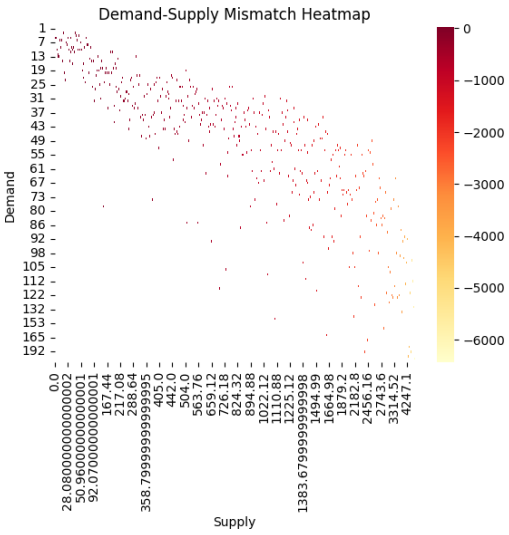

Fig.11. Demand supply mismatch heatmap.

The reason of undersupply might be sleeping hours early in the morning and traffic jams within the day time.

Question2: Please calculate the number of online hours required to ensure that we have a good Coverage Ratio during the peak hours you identified above.

Once we have identified the peak hours, we can calculate the number of online hours required to ensure a good coverage ratio. To do this, we can calculate the average number of rides per hour during the peak hours and divide it by the rides per online hour. This will give us the required online hours.

Number of online hours required are following:

1. First peak hours were calculated by selection of dataset where people who saw 0 cars are higher than people who saw +1 car;
2. Then required online hours were obtained by division of number of 'people who saw +1 car' from selected dataset in 1 to 'Rides per online hour'

Number of Online Hours Required for Good Coverage Ratio During Peak Hours: 2322.891

Question3: During peak hours, we can guarantee the drivers a certain amount of income. If the drivers make less than the guaranteed amount, we will pay them the difference. Please calculate how much earning we can guarantee so that we can attract more supply.

Guaranteed earning was obtained by the formula:

guaranteed_earning = (0.8 * peak_hours['People saw +1 cars (unique)'].sum() * 10) / peak_hours['Online (h)'].sum()

Earning We Can Guarantee to Attract More Supply During Peak Hours: €12.5