- ***Arash Khajooei***
- ***[arash.khajooei@gmail.com]***

# Report on Pricing Data Analyst Role Interview Task for SnappBox!

I am delighted to share this comprehensive report, showcasing my successful completion of the interview assignment for the Pricing Data Analyst role at Snapp! This document meticulously outlines the methodologies employed, analyses conducted, and invaluable insights drawn from the provided dataset and challenges. Its purpose is to spotlight my proficiency as a skilled and meticulous data analyst, focused on delivering precise results. Through this report, my aim is to present a robust and well-organized document that not only highlights the outcomes of the analysis but also provides transparent insights into the rationale behind each strategic step. I am committed to showcasing my ability to decipher intricate data, extract meaningful conclusions, and contribute substantively to data-centric decision-making within the framework of the Snapp! Company.

The opportunity to engage with this interview task has been highly rewarding, affording me an in-depth exploration of SnappBox's operational landscape and corporate ethos. I am confident that the competencies and expertise showcased in this report impeccably align with the demands of the Pricing Data Analyst position and Snapp's overarching objectives. I am sincerely appreciative of the chance to undertake this task, as it has enabled me to immerse myself in the intricacies of Snapp's operations and ethos. I would like to express my sincere appreciation for considering me as a potential member of your esteemed organization. I am eagerly looking forward to discussing this report and my contributions in further detail during the upcoming interview.

- ***This document does not include the source code. The main source code will send to the interviewers alongside this report and also uploaded on the GitHub repository.***

As a data analyst, data cleaning and preprocessing are essential steps in the data analysis process. Before delving into any analysis or generating insights, it is crucial to ensure that the data is accurate, consistent, and suitable for analysis. Data cleaning and preprocessing involve a series of techniques and procedures aimed at detecting and rectifying errors, inconsistencies, and missing values in the dataset.

- ***Firs Step: Data Cleaning and Data Preprocessing***

- **Data Cleaning:** Data cleaning is the initial step in the data analysis workflow. It involves identifying and correcting errors in the dataset to improve data quality. As a data analyst, my role would include:
    1. **Handling Missing Data:** Identifying missing data points and deciding the appropriate method to handle them, such as imputation or deletion. Missing data can significantly impact the analysis, and the chosen approach should be based on the nature of the data and the implications of imputing missing values.

2. **Removing Duplicates:** Identifying and removing duplicate records in the dataset. Duplicates can lead to skewed analysis results and must be eliminated to maintain data integrity.
3. **Handling Outliers:** Identifying and assessing outliers in the data. Outliers can distort analysis outcomes, and it is essential to decide whether to treat them or keep them in the dataset based on the context and objectives of the analysis.
4. **Data Standardization and Validation:** Ensuring that the data adheres to a consistent format, unit, and structure. Validation checks are performed to verify the accuracy and reliability of the data.

- **Data Preprocessing:** Data preprocessing involves transforming and preparing the data for analysis. This step ensures that the data is in a suitable format and aligns with the requirements of the analysis techniques to be applied. As a data analyst, my responsibilities would include:
    1. **Data Transformation:** This involves converting data into a standardized format, such as changing date formats, converting categorical variables into numerical values, or normalizing numerical data.
    2. **Feature Engineering:** Creating new features or variables based on existing data to improve model performance or enhance analysis insights. This could include deriving new metrics, aggregating data, or creating dummy variables.
    3. **Data Integration:** Merging data from various sources and combining datasets to create a unified and comprehensive dataset for analysis. This step is crucial when dealing with data from different systems or databases.
    4. **Data Reduction:** Reducing the dimensionality of the data to simplify analysis or improve computational efficiency. Techniques like Principal Component Analysis (PCA) or feature selection are commonly used for this purpose.

### Data Cleaning and Preprocessing: My Approach as a Data Analyst

As a data analyst, I understand the significance of data cleaning and preprocessing in ensuring the accuracy and reliability of the data I work with. Here's a breakdown of the steps I took to prepare the data for analysis:

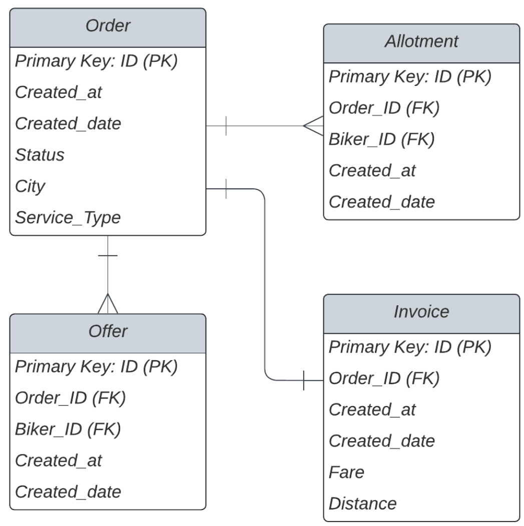
- ***ALL OF MY STEPS HAS BEEN DONE IN "GOOGLE COLLAB" PLATFORM SO SOME PART SHOULD CHANGE IN ORDER TO RUN THIS CODE ON YOUR OWN LOCAL SYSTEM.***

As the first step I'm going through following procedure to preparing our data for further exploration and analysis:

1. Handle missing data: Identify missing values and decide how to deal with them (e.g., impute, drop....).
2. Remove duplicate records, if any, to avoid skewing analysis results.
3. Address data inconsistencies, such as spelling errors, capitalization, or data entry mistakes.
4. Handle outliers if they exist and could influence the analysis.

- ***FIRST TASK: 1-SQL***

In order to extract the precise data and view presented in the "Distance Bucket" sheet, a SQL query needs to be composed using the following tables' structure. The functionality and relationships of these tables are explained in the subsequent section. The target of the query is to retrieve information for "City A" and "Service Type 1" on June 22nd, 2022. This query will allow us to access the specific data aligned with these criteria and recreate the intended view. First, I draw the Entity Relationship Diagram (ERD) to illustrate the relations between tables.




The ***.SQL*** code can be found in the repository. I going to explain each step of the query that I've written:

1. ***SELECT CLAUSE:***
    - The query starts with the SELECT clause, where I specify the columns that I want to retrieve in the result set.
    - Columns selected:
        - City, Service_Type, and Created_date from the Order table.
        - A CASE statement is used to create distance buckets based on the Distance value in the Invoice table.
        - Aggregate functions and calculated columns for analysis.
2. ***FROM CLAUSE:***
    - Specifies the source tables for the query.
    - Main table: Order (aliased as o).
    - Joins with other tables: Offer (aliased as ofr), Allotment (aliased as a), and Invoice (aliased as i).
3. ***LEFT JOINS:***
    - LEFT JOIN retrieves all rows from the left table (Order) and matching rows from the right tables (Offer, Allotment, Invoice).
    - Relationships are established using the relevant foreign keys (Order_ID) in the respective tables.
4. ***WHERE CLAUSE:***

- Filters the rows based on specified conditions:
    - City is 'A'
    - Service_Type is 1
    - Created_date is '6/22/2022'

5. ***GROUP BY CLAUSE:***
    - Groups the results based on specific columns:
        - City, Service_Type, Created_date, and the Distance Buckets (KM) calculated using the CASE statement.
6. ***AGGREGATE FUNCTIONS:***
    - COUNT (DISTINCT ...): Counts the number of distinct values for specific columns.
    - SUM(...): Calculates the sum of the specified column values.
    - ROUND(...): Rounds the calculated percentages and average fare to two decimal places.
7. ***ORDER BY CLAUSE:***
    - Orders the final result set based on the specified columns.

- ***FIRST TASK: 2- ANALYSIS***

What's your analysis of the pricing in this city? What are your suggestions to maximize the ride numbers?

***DATA LOADING:*** The first step was to load the data from the "Distance Bucket" sheet of the provided Excel file. This data contained information about different distance buckets, ride attributes, and pricing details.

***DATA CLEANING AND PREPROCESSING:*** The data was cleaned and preprocessed to handle missing values and ensure consistency. Columns with unnecessary symbols and spaces were renamed for better readability.

***DATA EXPLORATION AND VISUALIZATION:*** Initial insights were gained through exploratory data analysis. Scatter plots were created to visualize relationships between variables such as "Accepted-Offered %", "Fulfillment %", and "Average Ride Fare". These visualizations provided an overview of trends and potential outliers.

***OUTLIER DETECTION USING Z-SCORE METHOD:*** The Z-score method identifies a single outlier in the dataset. The identified outlier corresponds to the City "A," Service Type "1," and Created Date "2022-06-22" with the distance bucket "23-24."

***THIS OUTLIER HAS THE FOLLOWING CHARACTERISTICS:***

- Request: 14
- Offered Requests: 10
- Accepted Requests: 7
- Ride: 5
- Total Ride Fare: 1,655,000
- Offered-Created %: 71.43%
- Accepted-Offered %: 70.00%

- Fulfillment %: 35.71%
- Average Ride Fare: 331,000

***Accepted Requests:*** This point has a relatively low number of accepted requests (7) compared to the mean accepted requests (approximately 270). This suggests that on this specific day, a significantly lower number of requests were accepted, which could be due to various reasons, such as a shortage of available drivers or technical issues.

***Offered-Created %:*** The offered-created percentage is 0.714286, which means that only around 71% of the orders that were offered to bikers were created by customers. This value is relatively lower than the mean offered-created percentage (approximately 0.944085), indicating a lower conversion rate from offered orders to actual ride requests.

***Accepted-Offered %:*** The accepted-offered percentage is 0.7, which means that only around 70% of the offered orders were accepted by bikers. This value is relatively lower than the mean accepted-offered percentage (approximately 0.659208), indicating a lower acceptance rate of offered orders.

***Fulfillment %:*** The fulfillment percentage is 0.357143, which indicates that only around 36% of the accepted orders were successfully fulfilled as rides. This value is significantly lower than the mean fulfillment percentage (approximately 0.448408), suggesting that a larger proportion of accepted orders did not result in completed rides.

***Avg. Ride Fare:*** The average ride fare for this point is 331,000.0, which is relatively lower than the mean average ride fare (approximately 300,504). This could be due to the lower number of rides and the lower fulfillment rate, resulting in a lower overall revenue.

## Outlier Detection using IQR Method:

#### The IQR method identifies several outliers in the dataset.

The identified outliers share some common characteristics, such as being from City "A," Service Type "1," and Created Date "2022-06-22." Outliers include rows with distance buckets "23-24," "26-27," and "30-31." These outliers have varying characteristics, including Request, Offered Requests, Accepted Requests, Ride, Total Ride Fare, Offered-Created %, Accepted-Offered %, Fulfillment %, and Average Ride Fare.

## Descriptive Analysis for Numeric Columns:

- The descriptive analysis provides summary statistics for the numeric columns of the dataset.
- The dataset includes 34 records with Service Type "1."
- The Request column has a mean of approximately 421 and a standard deviation of about 471.
- The Offered Requests column has a mean of around 406 and a standard deviation of about 459.
- The Accepted Requests column has a mean of approximately 270 and a standard deviation of around 331.

- The Ride column has a mean of about 224 and a standard deviation of approximately 282.
- The Total Ride Fare column (Ride Fare) has a mean of around 38,285,590 and a standard deviation of approximately 42,131,990.
- The Offered-Created % column has a mean of about 0.944 and a standard deviation of around 0.073.
- The Accepted-Offered % column has a mean of approximately 0.659 and a standard deviation of about 0.172.
- The Fulfillment % column has a mean of around 0.448 and a standard deviation of approximately 0.130.
- The Average Ride Fare column has a mean of about 300,504 and a standard deviation of around 138,568.

### *Based on the outlier detection and descriptive analysis:*

1. The identified outlier from the Z-score method has relatively high values for Accepted Requests, Offered-Created %, and Accepted-Offered %, but a relatively low value for Fulfillment %. This could indicate a ***potentially abnormal behavior compared*** to the rest of the data points.
2. The identified outliers from the IQR method share similar characteristics, suggesting that these instances might represent specific patterns or unique situations that deviate from the majority of the data.
3. The points detected as ***outliers*** share common characteristics of lower-than-average accepted requests, lower-than-average fulfillment rates, and deviations from the mean in the offered-created and accepted-offered percentages. These patterns suggest that these specific instances experienced unique situations on the day of observation. These situations could involve operational challenges, external factors affecting demand, or issues in the service's ecosystem, resulting in deviations from the typical ride request and fulfillment patterns. Further investigation into these instances could provide insights into optimizing the service's performance and addressing potential issues that lead to such outlier behavior.

The summary statistics from the descriptive analysis provide insights into the central tendencies and variability of the numeric columns. For example, the relatively high standard deviations in some columns (e.g., Total Ride Fare) indicate significant variability in those measures.

### *Offered-Order / Created-Order ratio Analysis:*

***Observation:*** The "Offered-Order / Created-Order %" ratio indicates the percentage of offered orders compared to the total created orders. In the context of different distance buckets:

- As distance increases, the percentage of offered orders slightly decreases. This suggests that users might be more selective about using the service for longer distances.
- Users may prefer shorter rides for various reasons, such as convenience, cost-effectiveness, or time-saving.
- The decrease in this ratio could also reflect the impact of external factors like traffic congestion or travel time.

### ***Accepted-Order / Offered-Order Ratio Analysis:***

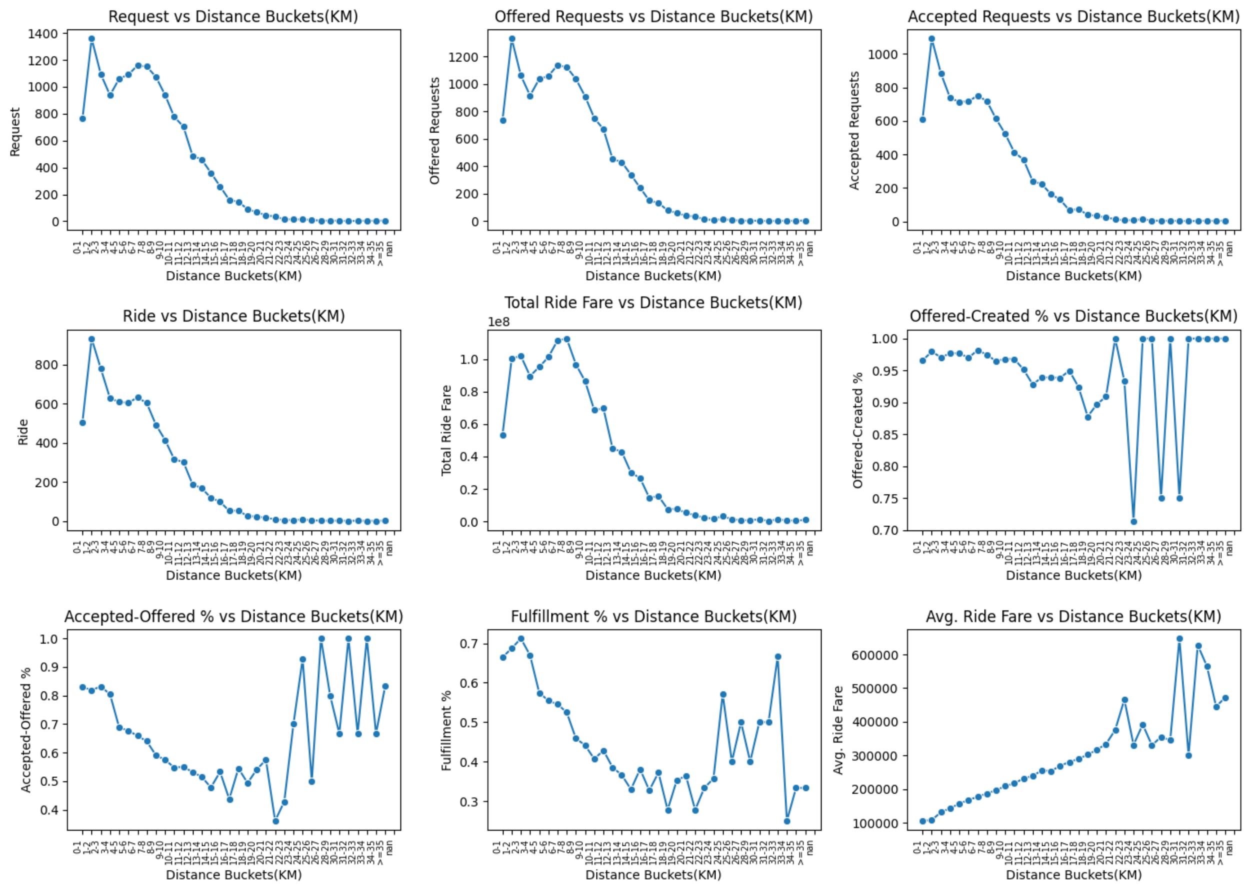
***Observation:*** The "Accepted-Order / Offered-Order%" ratio represents the percentage of accepted orders out of the total offered orders. Examining this ratio across distance buckets reveals:

- For shorter distances, the ratio is relatively higher, indicating a higher likelihood of bikers accepting rides for these distances.
- As distance increases, the ratio gradually declines, suggesting that bikers might be more hesitant to accept rides for longer distances.
- This pattern could be attributed to factors such as users' willingness to spend more time on longer trips or their preferences for more convenient transportation options for shorter distances.

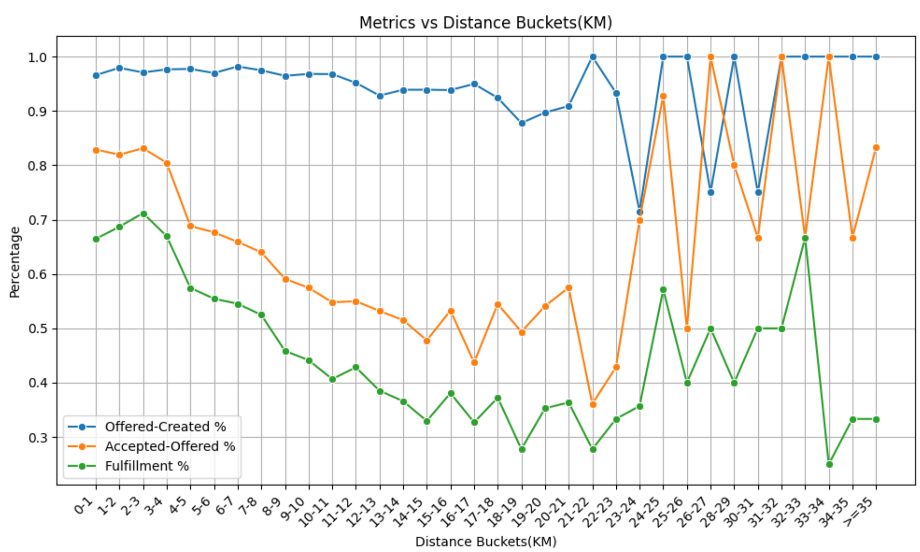
### ***Possible Reasons for the Observations:***

1. **User Behavior and Convenience:** Users might prioritize convenience for shorter distances. They might be more likely to accept rides that offer a quick and hassle-free solution for their immediate transportation needs.
2. **Perceived Value:** Users may perceive shorter rides as offering better value for money, leading to a higher acceptance rate. Longer rides might require more time and resources, influencing users' decisions to accept or decline.
3. **Travel Time and Commitment:** Longer rides are likely to take more time, which might discourage users from accepting them. Users could be hesitant to commit to longer trips due to potential time constraints or other engagements.
4. **Cost Considerations:** Users might be more conscious of costs for longer rides, leading to a lower acceptance rate. Longer distances often result in higher fares, which could impact users' decisions to accept offers.
5. **Alternative Transportation:** For longer distances, users might consider alternative transportation options like public transit, personal vehicles, or ridesharing services. This could influence their decision to accept or decline rides.
6. **Incentives:** Implementing incentives, such as reduced fares for longer distances or bonus rewards for drivers on long rides, could potentially encourage users to accept rides for these distances.

As we can see in the line charts of different parameters the amount of offered request and successful rides are clearly decreased in higher distance buckets.




We can also compare the percentage of offered request, accepted request and rides in a single line chart to observe this reduction better.



- ***Second Task: Ride Loss Calculation***

On June 8th, there was a technical issue, and bikers couldn't accept the orders after 10 p.m., and the ride number dropped. In this question, you should calculate the number of rides we lost because of this issue.

### 1. *Total Rides Lost Due to Technical Issue:*

The total number of rides lost due to the technical issue on June 8th, between 10:00 PM and 11:00 PM, is calculated to be **3734**. This represents the sum of rides that were not completed during the specified time period due to the technical issue.

### 2. *Average Lost Rides:*

The code that I've implemented also calculates the average number of rides lost on two reference days, June 6th and June 7th. The average lost rides are calculated to be **1293.5**. This value represents the average difference between the actual number of rides that occurred on the reference days and the number of rides that would have occurred if the technical issue hadn't happened.

### 3. *Comparison to Reference Days:*

The code uses the reference days (June 6th and June 7th) to estimate the average number of rides that would typically occur during the same time frame. It then compares this average with the actual number of rides that were completed on June 8th between 10:00 PM and 11:00 PM.

### 4. *Interpretation:*

The total number of rides lost (**3734**) indicates the direct impact of the technical issue on June 8th, with a significant number of rides not being completed. On average, the service lost around 1293.5 rides during the specified time frame, based on the comparison with the reference days. This suggests that the technical issue had a notable effect on ride completion during the affected hours.

### 5. *Future Considerations:*

To mitigate such issues in the future, it's important to address the technical problem promptly and efficiently to minimize the impact on service delivery and customer satisfaction. Additionally, continuous monitoring and swift response to technical issues can help prevent such disruptions.

- ***Third Task: Price Monitoring***

tell us what price action is needed to improve the business performance at 17:30.

In this problem I tried to extract the data of present moment in order to have a price action for the 17:30. If we consider 17 pm as present and we have to perform and strategy or price action to increase the business performance first we should observe our data in Realtime. Let's break down the analysis based on the provided data frames for different time slots and areas.

- **Request Data Frame:**

  The "Request" data frame provides the number of requests in different areas of the city for various time slots, including 17:00. This data indicates the level of demand in each area at 17:00. For instance, in Area Code 110, there were 8 requests at 17:00.

- **Accepted-Order / Offered-Order% Data Frame:**

  The "Accepted-Order / Offered-Order%" data frame represents the ratio of accepted orders to offered orders (requests) in different areas for different time slots, including 17:00. This ratio reflects the efficiency of order acceptance. For instance, in Area Code 110 at 17:00, the ratio was 0.875, indicating that 87.5% of offered orders were accepted.

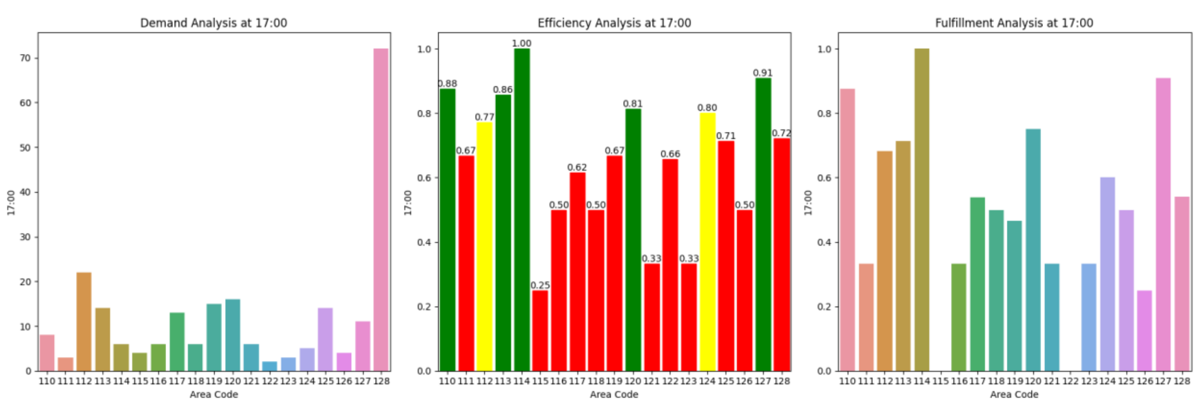
- **Fulfillment Rate% Data Frame:**

  The "Fulfillment Rate%" data frame shows the ratio of the number of successfully completed rides (orders) to the total number of requests in different areas for different time slots, including 17:00. This ratio reflects the business's ability to fulfill orders. For instance, in Area Code 110 at 17:00, the fulfillment rate was 0.875, indicating that 87.5% of orders were successfully completed.

For further investigation the data I extract these data to interpret the proper action which needed for the 17:30 pm:

- Areas with efficiency more than 75%
- Areas with high demand but lower fulfillment rates,
- Areas with high demand and efficiency lower than 75%:

***YOU CAN SEE THESE DATA AND SELECTED AREA CODE FOR EACH CATEGORIES IN THE CODE.***




***THE RED BAR CHART SHOWS LOWER EFFICIENCY THAN 75%***

***THE GREEN BAR CHART SHOWS HIGHER EFFICIENCY MORE THAN 75%***

***THE YELLOW BAR CHART SHOWS EFFICIENCY IN RANGE OF 75% TO 80% WHICH IS PREFERRED***

- Areas with Efficiency More than 75%:

  These areas demonstrate a strong efficiency in terms of order acceptance and fulfillment. This indicates that a high percentage of offered orders are being accepted and fulfilled successfully. In these areas, the supply and demand seem to be well-balanced. Given the high efficiency, your strategy could be to maintain the current pricing strategy as it appears to be effective in achieving a balanced supply and demand.

- Areas with High Demand but Lower Fulfillment Rates:

  These areas experience high demand but have relatively lower fulfillment rates. This suggests a potential imbalance between the number of orders and the available supply. To address this, consider implementing a targeted pricing increase during peak demand hours. This dynamic pricing strategy could incentivize more drivers to accept orders in these high-demand areas, leading to improved fulfillment rates.

- Areas with High Demand and Efficiency Lower than 75%:

  These areas experience both high demand and lower efficiency in terms of order acceptance and fulfillment. This combination could indicate a need to optimize both pricing and operational strategies. Implement dynamic pricing to encourage higher order acceptance rates while also focusing on strategies to improve ride completion rates. We might consider offering temporary incentives, bonuses, or rewards to drivers who consistently maintain high efficiency levels.

- ***IMMEDIATE PRICE ACTION STRATEGY:***

  Based on the identified categories, here's a suggested immediate price action strategy:

  ***Maintain Pricing for Efficient Areas:*** For areas with efficiency levels exceeding 75%, consider maintaining the existing pricing strategy as it seems to be working well to balance supply and demand.

  ***Dynamic Pricing for High Demand, Low Fulfillment Areas:*** Implement dynamic pricing in areas with high demand and lower fulfillment rates. Increase prices during peak demand hours to attract more drivers and improve ride completion rates. We can Monitor the impact of this strategy on fulfillment rates and adjust pricing as needed.

  ***Optimize Pricing and Operations for High Demand, Low Efficiency Areas:*** In areas with both high demand and efficiency levels below 75%, consider a two-pronged approach:

  - ***Dynamic Pricing:*** Apply dynamic pricing to encourage higher order acceptance rates. Increase prices during peak demand times to attract drivers.
  - ***Driver Incentives:*** Offer incentives, bonuses, or rewards to drivers who consistently maintain high efficiency levels by accepting and completing orders. This could help improve both order acceptance and fulfillment rates.

  ***Continuous Monitoring and Adjustment:*** We should regularly monitor the impact of our price action strategies on the identified areas. Use real-time data to make dynamic adjustments to pricing and operational strategies. I think Regular communication with drivers and analyzing their feedback can provide valuable insights for refining our approach.

I found some different pricing strategy which can be used in similar situation for long term:

***Surge Pricing:*** Increase prices during periods and also recognized area with the potential of drop the efficiency which are high demand and low supply. Also, other factors can make situation even worse such as rush hours or bad weather. This strategy encourages more drivers to be on the road and helps manage supply-demand imbalances.

***Dynamic Pricing:*** Adjust prices in real-time based on various factors like demand, supply, weather, traffic conditions, and events. This ensures that prices are always aligned with current market conditions. We can also consider the pricing strategy of our competitors for our business in the same areas.

***Time-of-Day Pricing:*** Set different pricing tiers for different times of the day. For example, prices could be higher during peak hours and lower during off-peak hours to incentivize more rides during quieter times.

***Distance-based Pricing:*** Charge riders based on the distance traveled. Longer distances could be priced higher to compensate drivers for longer rides.

***Zone Pricing:*** Divide the service area into zones and assign different prices for each zone. Areas with higher demand or longer travel times could have higher prices.

***Flat Rate Pricing:*** Offer fixed prices for specific routes or distances. This provides predictability for riders and encourages usage for common routes.

***Membership or Subscription Pricing:*** Offer riders a subscription plan where they pay a fixed monthly fee for a certain number of rides. This can encourage riders to use the service more frequently.

***Promotional Pricing:*** Offer temporary discounts or promotions to attract more riders during specific times or events.

***Tiered Pricing:*** Create different pricing tiers based on the level of service provided. For example, offer premium services with higher prices and added benefits.

***Personalized Pricing:*** Use rider profiles and historical data to offer personalized pricing based on user behavior and preferences.

***Geofencing Pricing:*** Use geofencing technology to set specific prices for certain areas or events. For instance, you could offer lower prices near popular venues to attract more riders.

***Demand Forecast Pricing:*** Use predictive analytics to forecast future demand and adjust prices in advance to optimize driver availability and rider satisfaction.

***Incentive-based Pricing:*** Offer rewards or discounts to riders who consistently use the service, creating loyalty and encouraging repeat business.

- ***THE LAST ONE: PRICE TEST***

  We increased the price of service type 1 in 6 of our cities for two days to test the effect of this change on business. You can see the business parameter changes caused by the price change. What's your analysis of the test result based on this data? Should we increase our prices in these cities? Explain your answer.

let's break down the analysis by considering the different aspects of the data. I first analyze these two days separately and then combine the result together.

***Test Day 1 Results for 6 Cities:***

On the initial day of testing, the following observations were made:

- Request: The average decrease in requests was significant, with a reduction of 5.00. This indicates that the price increase had a substantial impact on the number of ride requests, leading to reduced customer interest.
- Ride: A parallel decrease in rides, averaging at -2.50, mirrored the decrease in requests. This demonstrated a clear correlation between the pricing change and customer behavior.
- Price Conversion: The average decrease in price conversion, at -0.50, highlighted that the higher prices deterred potential customers from converting ride requests into actual rides. This suggests that the price increase affected the willingness of customers to proceed with their ride plans.
- Accepted-Order / Offered-Order%: The average increase in accepted-order percentage, at 1.17, indicated that drivers were more inclined to accept ride requests due to the lower demand. This is a positive impact from the business perspective, as drivers had more opportunities to fulfill orders.
- Fulfillment Rate%: A significant increase in the average fulfillment rate by 1.33 demonstrated that drivers were able to meet a higher proportion of ride requests due to the decreased demand. This could positively affect the overall customer experience.
- GMV: The fact that the average GMV remained steady at 0.00 despite the decrease in the number of rides implies that the increased pricing compensated for the decline in ride volume. This could be indicative of the potential to enhance revenue per ride.
- Average Fare Per KM: The average increase in fare per kilometer was substantial, at 4.17. This showcased the potential for revenue generation despite the decline in ride volume, as customers were paying more for each kilometer.

***Test Day 2 Results for 6 Cities:***

On the second day of testing, the results were as follows:

- Request: The trend of decreased ride requests persisted, with the average decrease standing at-4.33. This demonstrated that the impact of the pricing change remained consistent over the two testing days.
- Ride: The decrease in the average number of rides was slightly less severe than on the first day, at-2.00. While still negative, this indicates a potential stabilization in customer response to the new prices.

- **Price Conversion:** The average decrease in price conversion deepened to -1.83, signaling a continued challenge in convincing customers to proceed with their ride requests under the revised pricing structure.
- **Accepted-Order / Offered-Order%:** The consistent average increase of 1.17 in accepted-order percentage underscored the positive impact on driver behavior, wherein they were more willing to accept offered rides.
- **Fulfillment Rate%:** A further increase in the average fulfillment rate to 1.50 confirmed that drivers were readily fulfilling a higher proportion of ride requests due to decreased demand.
- **GMV:** The average GMV increase of 2.00 on the second day revealed that the pricing strategy was more effective in generating revenue despite the reduction in ride demand.
- **Average Fare Per KM:** The average fare per kilometer increase remained positive, though slightly reduced at 3.50. This still pointed toward the potential for increased revenue generation through higher fares.

***Combining Test Day 1 and Day 2 Results:***

Aggregating the results of both days, a pattern emerged:

- Ride demand experienced a significant **reduction** due to increased prices, leading to fewer requests and rides.
- Drivers exhibited improved willingness to accept and fulfill orders, translating to positive driver behavior and customer satisfaction.
- Revenue generation potential was maintained through increased fares, **despite the reduction in ride volume**.

***Recommendations:***

- **Balanced Approach:** I think we should Strive for a balanced approach that aligns with business objectives. While increased fares can lead to higher revenue per ride, they also result in decreased demand. A delicate balance is essential to maximize overall revenue.
- **Market Segmentation:** Consider segmenting the customer base to tailor pricing strategies to different groups. This can help manage the impact of price changes on customer loyalty and behavior.
- **Continuous Monitoring:** Regularly monitor key performance indicators to understand the evolving impact of pricing changes on the business. This will enable timely adjustments to pricing strategies.

I think we should consider that a comprehensive evaluation of revenue, rider experience, and driver satisfaction is crucial for a successful pricing strategy that strikes the right balance between maximizing revenue and maintaining customer demand. In analyzing the provided test data for two consecutive days, the effects of a price increase on various business parameters were examined, with a focus on short-term and long-term impacts. I will discuss about it in the following:

***Short-Term Impacts:***

The **immediate effects** of the price increase were observed over the two testing days:

- **Reduced Demand:** There was a consistent decrease in both ride requests and completed rides, indicating that the price change had an immediate negative impact on the user demand. This reduction in demand can be attributed to the increased cost, causing potential customers to explore alternative options or delay their trips.
- **Price Conversion:** The average price conversion showed a decline, indicating that the higher prices led to fewer riders converting their ride requests into completed rides. This suggests that users were hesitant to proceed with their requests due to the increased fares.
- **Driver Response:** On a positive note, the **accepted-order percentage** and **fulfillment rate** slightly increased. This suggests that drivers were more motivated to accept and fulfill ride requests due to the **lower demand**. This can enhance short-term driver satisfaction and overall service quality.
- **Revenue Fluctuations:** Despite the reduction in demand, the GMV remained stable on Day 1 and increased on Day 2. This suggests that the increased fares offset the decrease in the number of rides, yielding stable or improved revenue levels. However, relying solely on fare increases may not be sustainable in the long term.

***Long-Term Implications:***

When considering the long-term implications of the price increase:

- **Customer Retention:** The consistent reduction in ride demand over the two days indicates that the price change could lead to long-term customer retention challenges. Sustained high prices might alienate riders and prompt them to explore other transportation options, impacting brand loyalty.
- **Driver Behavior:** While drivers responded positively to the price change in the short term, long-term impacts on their willingness to accept rides should also be considered. If the overall rider base diminishes, drivers might experience periods of reduced demand, potentially affecting their income and job satisfaction.
- **Competitor Response:** Long-term high prices could prompt competitors to attract customers with more **competitive pricing**. Monitoring competitive pricing strategies is essential to maintain a balanced competitive position.

***Analysis of Ride Per Check Parameter and Its Significance***

The "Ride Per Check" parameter, which quantifies the ratio of successful rides to the number of times a price has been checked, is a valuable metric for understanding rider behavior and its impact on pricing strategies. Analyzing this parameter provides insights into rider sensitivity to price changes, conversion rates, and their decision-making process. Here's how to use this parameter for analysis and extract meaningful results:

1. ***Rider Price Sensitivity:***

The Ride Per Check metric allows us to gauge how sensitive riders are to pricing changes. A higher positive value indicates that a significant portion of riders continue to proceed with ride requests despite checking the price, suggesting that they are less price-sensitive. Conversely, a lower positive value or a negative value indicates that riders are more **price-sensitive**, leading to canceled or unfulfilled rides due to perceived high prices.

2. ***Conversion Rate Analysis:***

By correlating Ride Per Check with other metrics such as Price Conversion, you can assess the conversion rate of riders who check the price. A higher Ride Per Check coupled with a high Price Conversion indicates that riders are willing to the higher prices. On the other hand, a low-Price Conversion despite a high Ride Per Check could suggest that riders are checking prices but opting for alternative transportation methods due to the cost.

3. ***. Pricing Strategy Validation:***

The Ride Per Check parameter can help validate the effectiveness of pricing strategies. If Ride Per Check remains steady or increases even after a price increase, it might indicate that users perceive the value in the service and are willing to pay more. Conversely, a sharp drop in Ride Per Check after a price increase could signal that the pricing strategy has led to user attrition.

4. ***Demand Elasticity Assessment:***

The metric offers insights into demand elasticity – how much ride demand changes in response to price changes. If Ride Per Check remains relatively constant despite pricing changes, it could suggest that demand is inelastic, meaning riders continue to use the service regardless of price fluctuations. A decrease in Ride Per Check would imply higher demand elasticity, as riders are more reactive to price changes.

5. ***Decision-Making Patterns:***

Analyzing Ride Per Check can uncover patterns in rider decision-making. A consistent positive Ride Per Check indicates stable rider behavior, while fluctuations suggest changes in riders' perception of value. If Ride Per Check becomes negative, it implies that riders are more inclined to opt out of rides due to perceived high costs.

6. ***6. Long-Term Effects:***

Tracking Ride Per Check over an extended period can help assess the long-term impact of pricing changes. If Ride Per Check gradually decreases over time, it may indicate that riders are becoming more price-sensitive and exploring other alternatives. Conversely, a stable or increasing Ride Per Check could suggest resilience to price changes.

***Extracting Results:***

To extract results from the Ride Per Check parameter:

- We can compare Ride Per Check across different pricing scenarios or over time to identify trends and changes in rider behavior.
- We can correlate Ride Per Check with other relevant metrics, such as Price Conversion, GMV, and rider retention rates, to understand its impact on overall business performance.
- Utilize Ride Per Check to refine pricing strategies. For example, if Ride Per Check remains positive despite price increases, it suggests a strong value perception; if it decreases significantly, it might be time to reevaluate pricing.

Ride Per Check is a dynamic and insightful metric that can guide pricing decisions and provide valuable insights into rider behavior and demand elasticity. By understanding how riders react to price changes and correlating this parameter with other metrics, our businesses can develop more effective pricing strategies and ensure sustainable growth while maintaining rider satisfaction.

I will provide the final code and implementation in "***.IPYNB***" format. Its also uploaded to GitHub repository in order to ease of access you can open the source code in Google Collab and run each cell. If you open the code in collab platform keep in mind that you should provide both dataset file in the root path which is "***/CONTENT/***"

### CONCLUSION

Thank you for giving me this opportunity to join Snapp! Company which is my favorite workplace. I am excited about the possibility of joining your team and contributing my skills to your innovative projects. I am eager to be part of your talented workforce and make a positive impact. Looking forward to your decision.

Sincerely,

***ARASH KHAJOOEI***

***30 AUGUST 2023***