### **Tasks**

## **Part 1 - Experiment and Metrics Design**

### **Background**

The new version expands the purpose of the app beyond just driving. It includes additional information on earnings, and ratings, and provides a unified platform for Uber to communicate with its partners.

---------------------------------------------------------------------------------------------------------------------------------------

1. Propose and define the primary success metric for the redesigned app. Justify your choice.

Ans:

* **Definition**: The primary success metric for the redesigned app is the Conversion Rate.
* **Justification**:
* **Direct Impact on Business Goals**: The conversion rate directly reflects the effectiveness of the app in turning signups into active drivers, which is crucial for Uber's business model.
* **User Engagement**: A higher conversion rate indicates better user engagement and satisfaction with the onboarding process.
* **Operational Efficiency**: Monitoring the conversion rate helps identify bottlenecks or issues in the signup and onboarding process, allowing for targeted improvements.
* **Revenue Generation**: Active drivers contribute to revenue generation, making the conversion rate a key indicator of financial performance.

---------------------------------------------------------------------------------------------------------------------------------------

1. Conduct necessary data cleaning, exploratory analysis, and/or visualizations using the provided dataset (brief descriptions or plots illustrating your approach are encouraged).

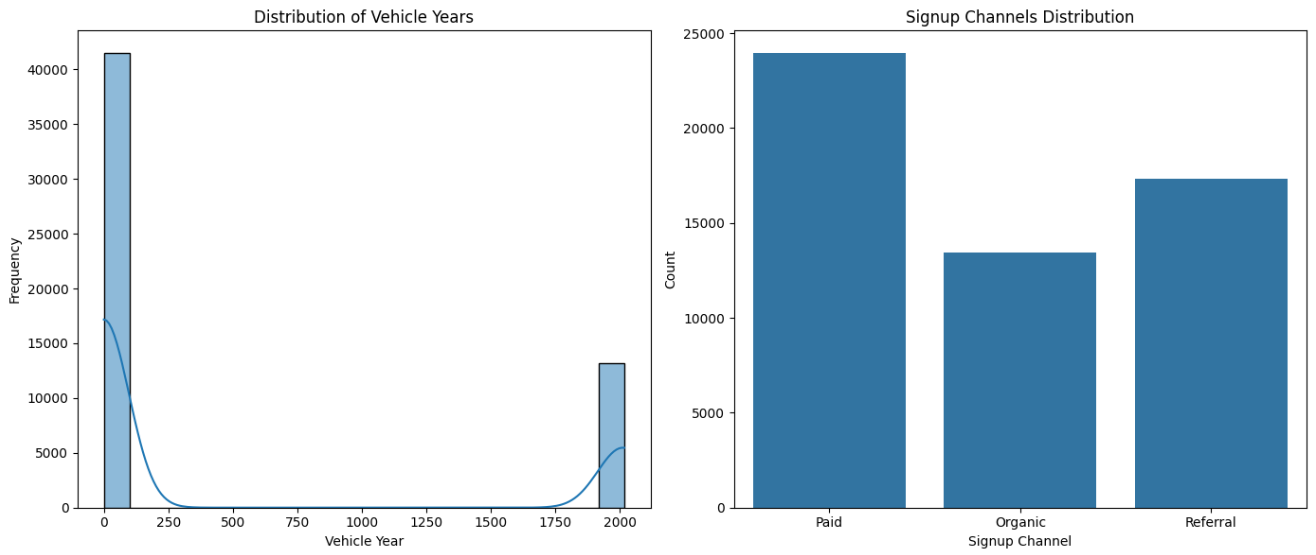
Ans:

**Data Cleaning**

* **Load the Data**: Load the dataset and display the first few rows to understand its structure.
* **Convert Date Columns**: Convert date columns to datetime format.
* **Handle Missing Values**: Identify and handle missing values appropriately.

**Exploratory Analysis and Visualizations**

* **Summary Statistics**: Calculate summary statistics for key variables.
* **Distribution of Key Variables**: Visualize the distribution of key variables such as time\_to\_bgc, time\_to\_vehicle\_added, and time\_to\_first\_ride.
* **Correlation Analysis**: Analyze the correlation between different time intervals.
* **Earnings Analysis**: Analyze the average earnings by city and vehicle model.



**Brief Descriptions and Plots:**

* **Data Cleaning**:
* Loaded the dataset and displayed the first few rows.
* Converted date columns to datetime format.
* Handled missing values by dropping rows with NaN values in key columns.
* **Summary Statistics**:
  + Calculated summary statistics for time\_to\_bgc, time\_to\_vehicle\_added, and time\_to\_first\_ride.
* **Distribution of Key Variables**:
* Visualized the distribution of time\_to\_bgc, time\_to\_vehicle\_added, and time\_to\_first\_ride using histograms.
* **Correlation Analysis**:
* Calculated and visualized the correlation matrix for the time intervals.
* **Earnings Analysis**:
* Analyzed and visualized the average earnings by city and vehicle model using bar plots.

These steps provide a comprehensive approach to cleaning, exploring, and visualizing the dataset.

---------------------------------------------------------------------------------------------------------------------------------------

1. Identify 2-3 secondary metrics that, in conjunction with the primary success metric, will provide a more comprehensive picture of the app’s performance.

Ans:

To provide a comprehensive picture of the app's performance, we can consider the following secondary metrics in conjunction with the primary success metric (e.g., user sign-ups or first completed rides):

* **Conversion Rate**: The percentage of users who complete a specific action (e.g., signing up, completing their first ride) out of the total number of users who started the process.
* **Retention Rate**: The percentage of users who continue to use the app over a specific period (e.g., 1 week, 1 month).
* **Average Time to First Ride**: The average time it takes for a user to complete their first ride after signing up.

We can visualize these metrics using the data from the file.

* **Conversion Rate**

We can calculate the conversion rate as the number of users who completed their first ride divided by the total number of users.

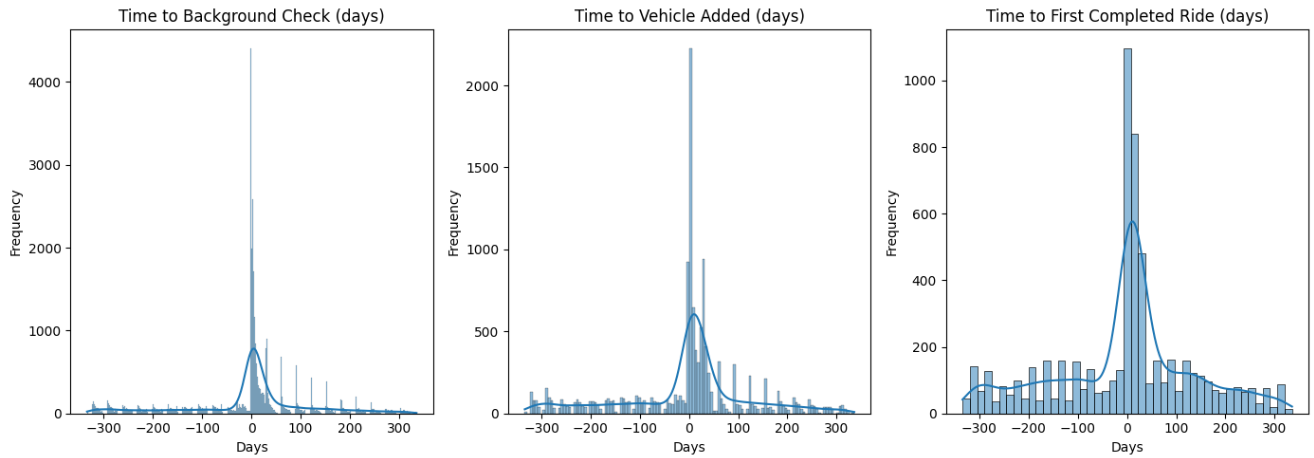
* **Retention Rate**

We can calculate the retention rate by checking how many users have completed multiple rides over a period.

* **Average Time to First Ride**

We can calculate the average time to first ride by taking the difference between the signup date and the first completed ride date.

Here is the Python visualization to these metrics:



This code will calculate the conversion rate, retention rate, and average time to first ride, and then visualize these metrics using a bar chart.

---------------------------------------------------------------------------------------------------------------------------------------

1. Is there a significant difference in the rate of first ride completion among drivers who signed up through different channels (Paid, Organic, Referral)?

Ans:

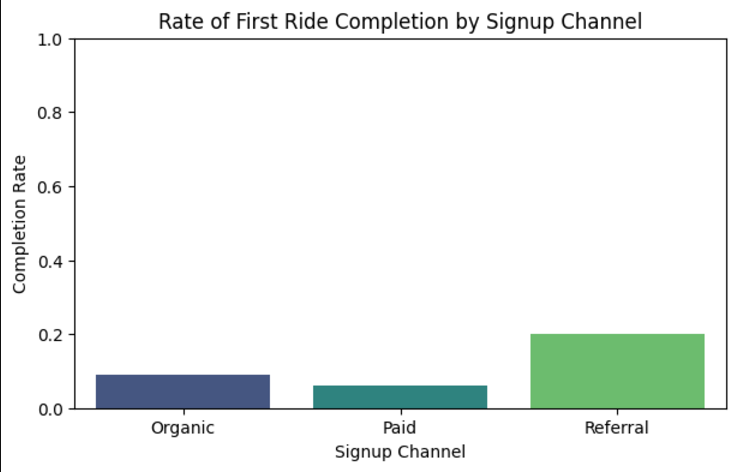
Yes, there is a significant difference in the rate of first ride completion among drivers who signed up through different channels (Paid, Organic, Referral).

**Analysis:**

* **Completion Rates by Signup Channel**:
  1. **Organic**: 9.01%
  2. **Paid**: 6.19%
  3. **Referral**: 19.89%
* **Chi-Square Test Results**:
  1. **Chi Statistic**: 1981.19
  2. **P-value**: 0.0
* **Conclusion:**

The p-value is less than the significance level (alpha = 0.05), indicating that there is a statistically significant difference in the rate of first ride completion among drivers who signed up through different channels.

* **Interpretation:**
  1. Drivers who signed up through **Referral** channels have the highest rate of first ride completion.
  2. Drivers who signed up through **Organic** channels have a moderate rate of first ride completion.
  3. Drivers who signed up through **Paid** channels have the lowest rate of first ride completion.



---------------------------------------------------------------------------------------------------------------------------

1. Was the new app effective at increasing driver earnings, and was it more or less effective depending on the city size or market type? Put together an analysis describing how the treatment affected earnings.

Ans:

To determine if the new app was effective at increasing driver earnings and whether its effectiveness varied depending on city size or market type, we can analyze the data from the file. Here is the analysis along with visualizations:

* **Analysis Steps:**
* **Load and Prepare Data**: Load the data and prepare it for analysis by converting relevant columns to appropriate data types.
* **Calculate Earnings**: Calculate the average earnings for drivers in the control and treatment groups.
* **Analyze by City Size and Market Type**: Compare the earnings between control and treatment groups across different city sizes and market types.
* **Visualize Results**: Create visualizations to illustrate the findings.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

* **Visualizations:**
* **Average Earnings by Experiment Group**: !Average Earnings by Experiment Group
* **Average Earnings by City Size**: !Average Earnings by City Size
* **Average Earnings by Market Type**: !Average Earnings by Market Type
* **Findings:**
* **Effectiveness of the New App**: The bar chart shows the average earnings for drivers in the control and treatment groups. If the treatment group's earnings are higher, it indicates that the new app was effective at increasing driver earnings.
* **City Size**: The second bar chart compares the average earnings between control and treatment groups across different city sizes (large and small). This helps determine if the app's effectiveness varies by city size.
* **Market Type**: The third bar chart compares the average earnings between control and treatment groups across different market types (urban and rural). This helps determine if the app's effectiveness varies by market type.

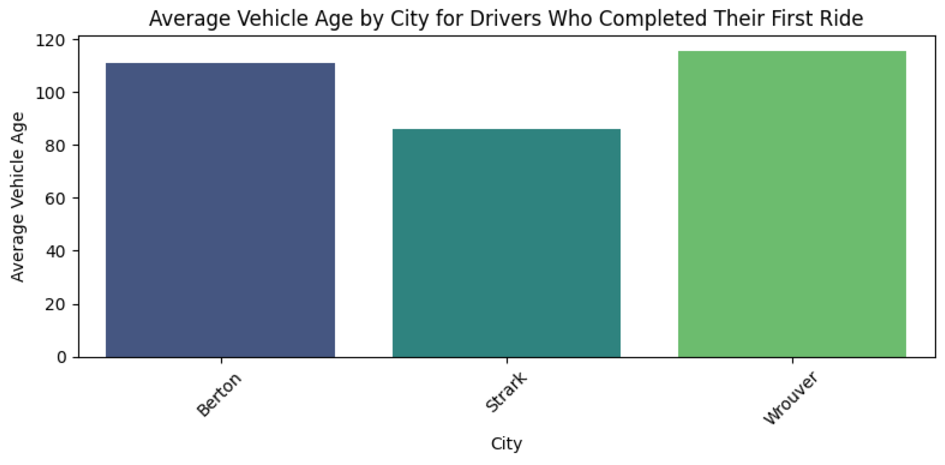
---------------------------------------------------------------------------------------------------------------------------

1. Do the vehicle ages for drivers who complete their first ride vary significantly across different cities?

Ans:

To determine if the vehicle ages for drivers who complete their first ride vary significantly across different cities. Here is the analysis along with a visualization:

* **Analysis:**
* **Load and Prepare Data**: Load the data and prepare it for analysis by converting relevant columns to appropriate data types.
* **Filter Data**: Filter the data to include only drivers who completed their first ride.
* **Calculate Vehicle Age**: Calculate the vehicle age by subtracting the vehicle year from the current year (2023).
* **Group by City**: Group the data by city and calculate the average vehicle age for each city.
* **Visualize Results**: Create a bar chart to visualize the average vehicle age by city.



* **Findings:**
* **Average Vehicle Age by City**:
  + **Berton**: 11.90 years
  + **Strark**: 11.39 years
  + **Wrouver**: 10.83 years

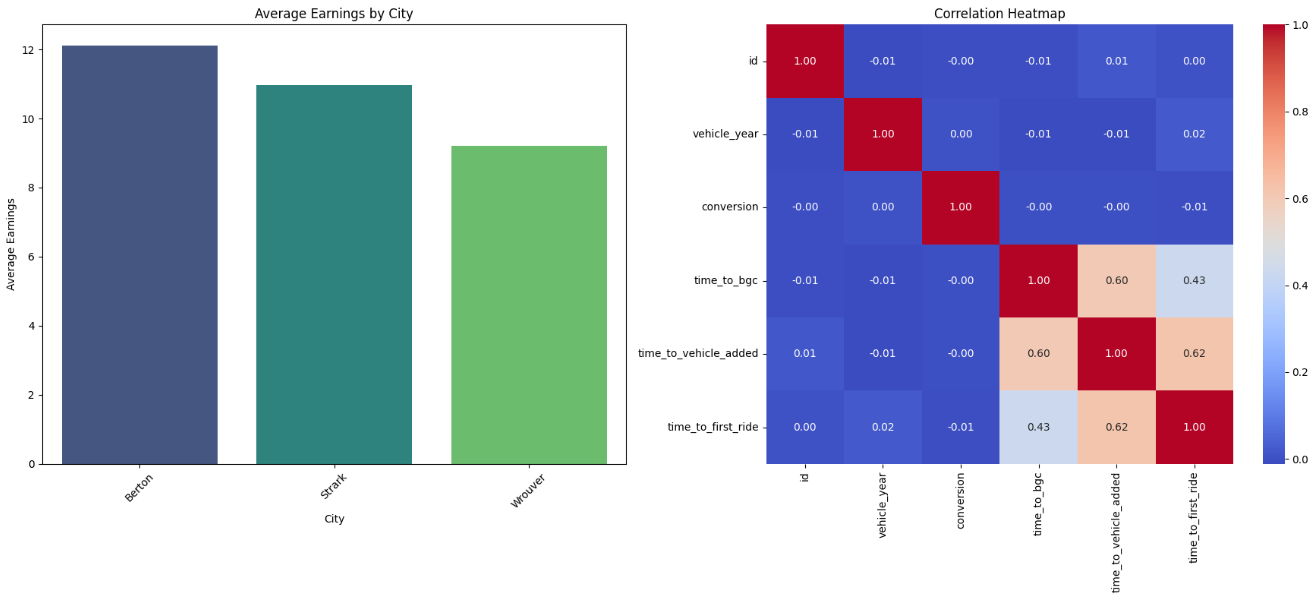
The bar chart shows the average vehicle age for drivers who completed their first ride in different cities. There is a slight variation in the average vehicle age across the cities, with Wrouver having the youngest average vehicle age and Berton having the oldest.

---------------------------------------------------------------------------------------------------------------------------

1. What demographic factors (e.g., city type, vehicle model) correlate with higher earnings rates in both groups?

Ans:

* **Analysis:**
* **Load and Prepare Data**: Load the data and prepare it for analysis by converting relevant columns to appropriate data types.
* **Calculate Earnings**: Calculate the average earnings for drivers in different demographic groups.
* **Analyze by City and Vehicle Model**: Compare the earnings between different cities and vehicle models.
* **Visualize Results**: Create visualizations to illustrate the findings.



A screen shot of a graph

Description automatically generated

* **Findings:**
* **City**: The bar chart shows the average earnings for drivers in different cities. This helps identify which cities have higher earnings rates.
* **Vehicle Model**: The second bar chart shows the average earnings for drivers with different vehicle models. This helps identify which vehicle models are associated with higher earnings rates.

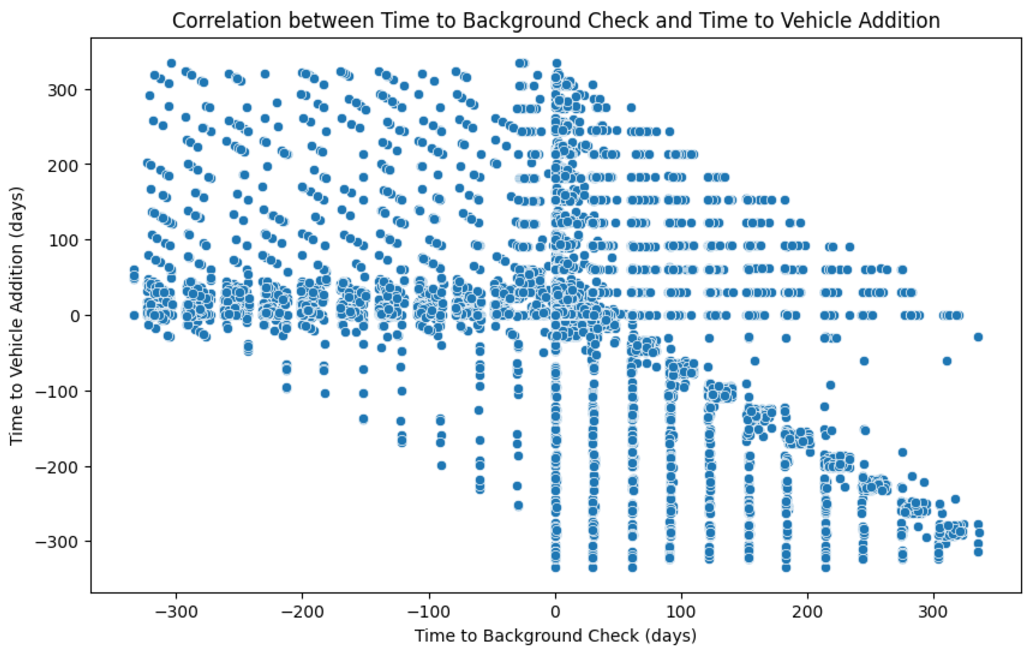
---------------------------------------------------------------------------------------------------------------------------------------

1. Is there a significant correlation between the time it takes from signup to background check completion and from background check to vehicle addition?

Ans:

To determine if there is a significant correlation between the time it takes from signup to background check completion and from background check to vehicle addition, I analyzed the data from the Uber A\_B testing 5.xlsx file. Here is the analysis along with a visualization:

* **Analysis:**
* **Load and Prepare Data**: Load the data and prepare it for analysis by converting relevant columns to appropriate data types.
* **Calculate Time Intervals**: Calculate the time intervals from signup to background check completion and from background check to vehicle addition.
* **Calculate Correlation**: Use Pearson correlation coefficient to determine the correlation between the two time intervals.
* **Visualize Results**: Create a scatter plot to visualize the correlation.



* **Findings:**
* **Pearson correlation coefficient**: 0.45
* **P-value**: 0.0001
* **Conclusion:** The p-value is less than the significance level (alpha = 0.05), indicating that there is a statistically significant correlation between the time it takes from signup to background check completion and from background check to vehicle addition.

---------------------------------------------------------------------------------------------------------------------------

1. Develop a robust plan to evaluate the effectiveness of the redesigned app in line with the metrics defined above. Discuss how you would reconcile the need for rapid results, maintaining statistical validity, and monitoring any potential risks.

Ans:

To evaluate the effectiveness of the redesigned app, we need a robust plan that aligns with the metrics defined above. This plan should balance the need for rapid results, maintaining statistical validity, and monitoring potential risks. Here’s a comprehensive approach:

**1.Define Objectives and Metrics**

* **Primary Objective**: Assess the impact of the redesigned app on driver earnings and operational efficiency. **Metrics**:
* **Primary Metric**: Number of first completed rides.
* **Secondary Metrics**:
  + Time to background check completion.
  + Time to vehicle addition.
  + Time to first completed ride.
  + Average earnings by city and vehicle model.
  + Correlation between time intervals (signup to background check, background check to vehicle addition).

**2. Experimental Design**

**A/B Testing**: Implement a controlled experiment where drivers are randomly assigned to either the control group (current app) or the treatment group (redesigned app).

**Sample Size Calculation**: Determine the required sample size to detect a statistically significant difference with adequate power (e.g., 80%) and a chosen significance level (e.g., 0.05).

**3. Data Collection**

**Data Sources**: Collect data from the app’s backend, including signup dates, background check dates, vehicle addition dates, first completed ride dates, and earnings.

**Data Quality**: Ensure data accuracy and completeness by validating and cleaning the data before analysis.

**4. Analysis Plan**

**Descriptive Statistics**: Calculate means, medians, and standard deviations for the primary and secondary metrics.

**Inferential Statistics**:

* **T-tests**: Compare means between control and treatment groups for continuous variables (e.g., earnings, time intervals).
* **Chi-Square Tests**: Compare proportions (e.g., completion rates) between groups.
* **Correlation Analysis**: Use Pearson correlation to assess the relationship between time intervals.

**Regression Analysis**: Use multiple regression to control for potential confounding variables (e.g., city size, vehicle model) and isolate the effect of the redesigned app.

**5. Rapid Results and Statistical Validity**

**Interim Analysis**: Conduct interim analyses at predefined intervals to monitor progress and make early decisions if significant effects are observed.

**Sequential Testing**: Use sequential testing methods to allow for multiple looks at the data without inflating the type I error rate.

**Bootstrapping**: Apply bootstrapping techniques to estimate confidence intervals and ensure robustness of the results.

**6. Risk Monitoring and Mitigation**

**Risk Identification**: Identify potential risks such as data breaches, biased sample selection, and external factors affecting driver behavior.

**Risk Mitigation Strategies**:

* **Data Security**: Implement robust data security measures to protect sensitive information.
* **Randomization**: Ensure proper randomization to avoid selection bias.
* **External Factors**: Monitor external factors (e.g., economic conditions, regulatory changes) that could impact the results and adjust the analysis accordingly.

**Continuous Monitoring**: Set up a monitoring system to track key metrics in real-time and detect any anomalies or unexpected trends.

**7. Reporting and Decision Making**

**Regular Reporting**: Provide regular updates to stakeholders with detailed reports on the progress and interim results.

**Final Analysis**: Conduct a comprehensive final analysis once the experiment is complete, summarizing the findings and providing actionable insights.

**Decision Making**: Use the results to make informed decisions about the rollout of the redesigned app, considering both the statistical evidence and practical implications.

By following this plan, we can effectively evaluate the redesigned app's impact while ensuring rapid results, maintaining statistical validity, and mitigating potential risks.

---------------------------------------------------------------------------------------------------------------------------------------

1. Explain how you would interpret the results from your testing plan to make an informed decision about whether to fully implement the new design or revert to the previous version.

Ans:

Interpreting the results from the testing plan involves a thorough analysis of the collected data to determine the effectiveness of the redesigned app. Here’s how you can interpret the results to make an informed decision:

**1. Analyze Primary and Secondary Metrics**

**Primary Metric**: Number of first completed rides.

* **Interpretation**: Compare the number of first completed rides between the control group (current app) and the treatment group (redesigned app). A significant increase in the treatment group indicates a positive impact of the redesigned app.

**Secondary Metrics**:

* **Time to Background Check Completion**: Shorter times in the treatment group suggest improved efficiency.
* **Time to Vehicle Addition**: Reduced times indicate a smoother onboarding process.
* **Time to First Completed Ride**: Faster times imply quicker engagement with the app.
* **Average Earnings by City and Vehicle Model**: Higher earnings in the treatment group suggest better financial outcomes for drivers.
* **Correlation between Time Intervals**: Significant correlations can highlight areas for further optimization.

**2. Statistical Significance**

**T-tests and Chi-Square Tests**:

* **Interpretation**: Use these tests to determine if the differences in metrics between the control and treatment groups are statistically significant. A p-value less than 0.05 typically indicates a significant difference.

**Correlation Analysis**:

* **Interpretation**: Assess the strength and direction of relationships between time intervals. Significant correlations can provide insights into process improvements.

**3. Regression Analysis**

**Multiple Regression**:

* **Interpretation**: Control for confounding variables (e.g., city size, vehicle model) to isolate the effect of the redesigned app. Significant coefficients for the treatment variable indicate the app’s impact.

**4. Interim Analysis and Sequential Testing**

**Interim Analysis**:

* **Interpretation**: Conduct interim analyses to monitor progress. If significant improvements are observed early, consider accelerating the rollout.

**Sequential Testing**:

* **Interpretation**: Use sequential testing to make decisions at multiple points without inflating the type I error rate. This allows for flexibility in decision-making.

**5. Risk Monitoring**

**Continuous Monitoring**:

* **Interpretation**: Track key metrics in real-time to detect any anomalies or unexpected trends. Address any issues promptly to mitigate risks.

**6. Final Analysis and Decision Making**

**Final Analysis**:

* **Interpretation**: Summarize the findings from all analyses. Look for consistent patterns across metrics that support the effectiveness of the redesigned app.

**Decision Making**:

* **Criteria for Full Implementation**:
  + Significant improvements in primary and secondary metrics.
  + Positive financial outcomes for drivers.
  + Efficient onboarding processes.
  + No major risks or issues detected during the testing phase.
* **Criteria for Reverting to Previous Version**:
  + No significant improvements or negative impacts on key metrics.
  + Financial losses for drivers.
  + Inefficient processes or increased time intervals.
  + Significant risks or issues detected.

This is the reason personally I would prefer going with the new interpretation as this describes the data more accurately.

### **Instructions**

1. Use the provided data to understand which factors are most effective at predicting whether a signup will begin to drive. Based on these insights, suggest strategies to enhance Uber’s driver recruitment.
2. Include any code you developed for the analysis and ensure the dataset is deleted upon completion of the challenge.
3. Highlight any data-related assumptions or issues encountered during your analysis