### **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:** Primary Success Metric Proposal:

* **Metric**: First Ride Completion Rate
* **Definition**: The percentage of new sign-ups who successfully complete their first ride within a defined period (e.g., 30 days of signup). This is directly derived from the first\_completed\_date column in the dataset and by doing feature engineering has\_completed\_first\_ride.
* **Justification**: This metric is crucial because it directly reflects whether the redesigned app effectively enables new drivers to become active. An increase indicates that the improved features (earnings information, ratings, communication) are making the onboarding and initial driving experience more seamless and compelling, leading to quicker activation and value realization for Uber. It's a critical early activation point that serves as a strong predictor of long-term driver engagement.

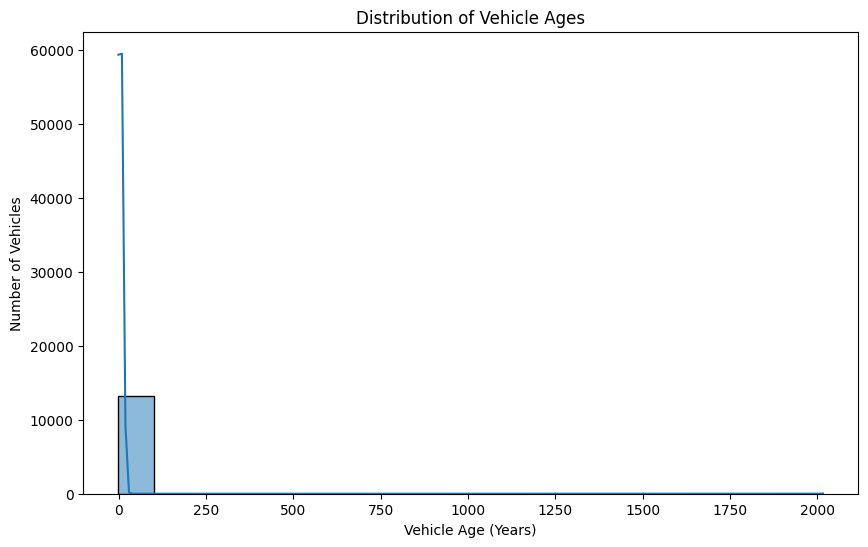
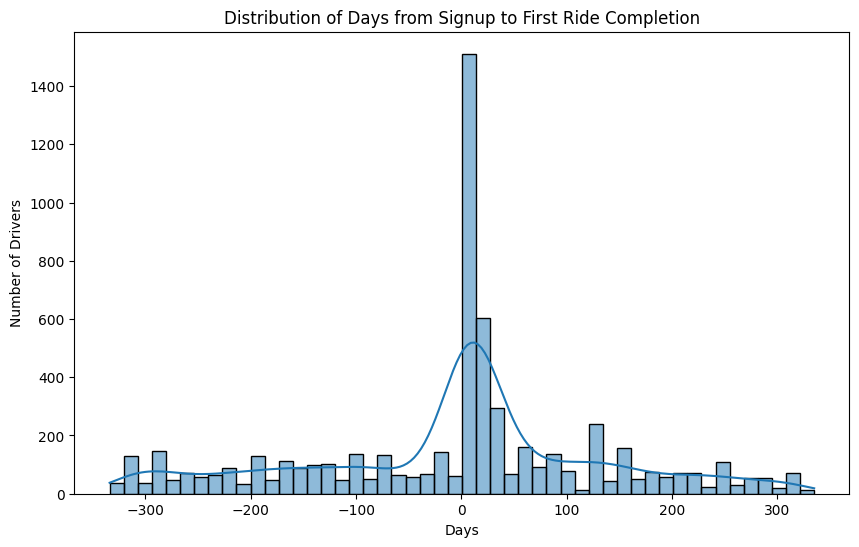
1. **Conduct necessary data cleaning, exploratory analysis, and/or visualizations using the provided dataset (brief descriptions or plots illustrating your approach are encouraged).**

**Ans:**

1. **Data Cleaning and Preprocessing Approach**

The following steps were taken to clean and prepare the dataset for analysis:

* **Date Conversion:** Columns containing date information (signup\_date, bgc\_date, vehicle\_added\_date, first\_completed\_date) were converted to datetime objects. This allows for proper chronological analysis and calculation of time differences. Errors during conversion were coerced to NaT (Not a Time).
* Handling Missing Values and Inconsistent Strings: 'NA' strings and empty strings ('') in categorical columns (signup\_os, signup\_channel, city\_name) were systematically replaced with NaN (Not a Number) to ensure consistent handling of missing data by pandas.
* **Feature Engineering:** Several new features were created to facilitate the analysis:
* has\_completed\_first\_ride: A binary indicator (1 or 0) derived from first\_completed\_date. If a date is present, it means the first ride was completed (1); otherwise, it's 0. This directly supports the primary success metric.
* Time Differences: Calculated the duration in days for key onboarding stages:
* signup\_to\_bgc\_days: Time taken from signup to background check completion.
* bgc\_to\_vehicle\_add\_days: Time taken from background check completion to vehicle addition.
* signup\_to\_first\_ride\_days: Total time from signup to the first completed ride.
* **vehicle\_age**: Estimated by subtracting the vehicle\_year from a reference year (assumed to be 2016, given the data context). Missing

vehicle\_year values resulted in NaN for vehicle\_age.

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:** In conjunction with the primary metric of First Ride Completion Rate, the following secondary metrics are proposed to provide a more comprehensive picture of the app’s performance and the effectiveness of the redesign.

1. **Time to First Ride Completion**

* **Definition:** The average or median number of days it takes a driver to complete their first ride, starting from their signup date. This metric is only calculated for drivers who successfully complete a first ride.
* **Justification:** A shorter time to first ride completion suggests a more efficient and intuitive onboarding process. The new app's features (improved information, better communication) are intended to streamline this process. A significant reduction in this time would provide strong evidence that the redesign is successfully reducing friction and enabling faster driver activation.

1. **Driver Conversion Rate at Key Onboarding Stages**

* **Definition:** This metric measures the proportion of drivers who successfully progress through critical steps in the onboarding funnel. Specifically, it can be broken down into:
* Signup to Background Check Completion Rate
* Background Check Completion to Vehicle Addition Rate
* **Justification:** The redesigned app aims to provide more clarity and support. Improvements in these specific conversion rates would indicate that the new app is effectively guiding drivers through administrative hurdles, reducing drop-off rates at these crucial stages and leading to a healthier overall driver pipeline.

1. **App Engagement Metrics (Conceptual, as not in dataset)**

* **Definition:** Metrics such as average time spent in the app, frequency of logging in, and usage of new features (e.g., viewing earnings reports, using the communication platform).
* **Justification:** These metrics would provide insight into whether the redesigned app is more engaging and useful to drivers, beyond just getting them to complete their first ride. An increase in engagement would suggest that the new features are valuable and could lead to better long-term driver retention and satisfaction, even if the primary metric change is modest.

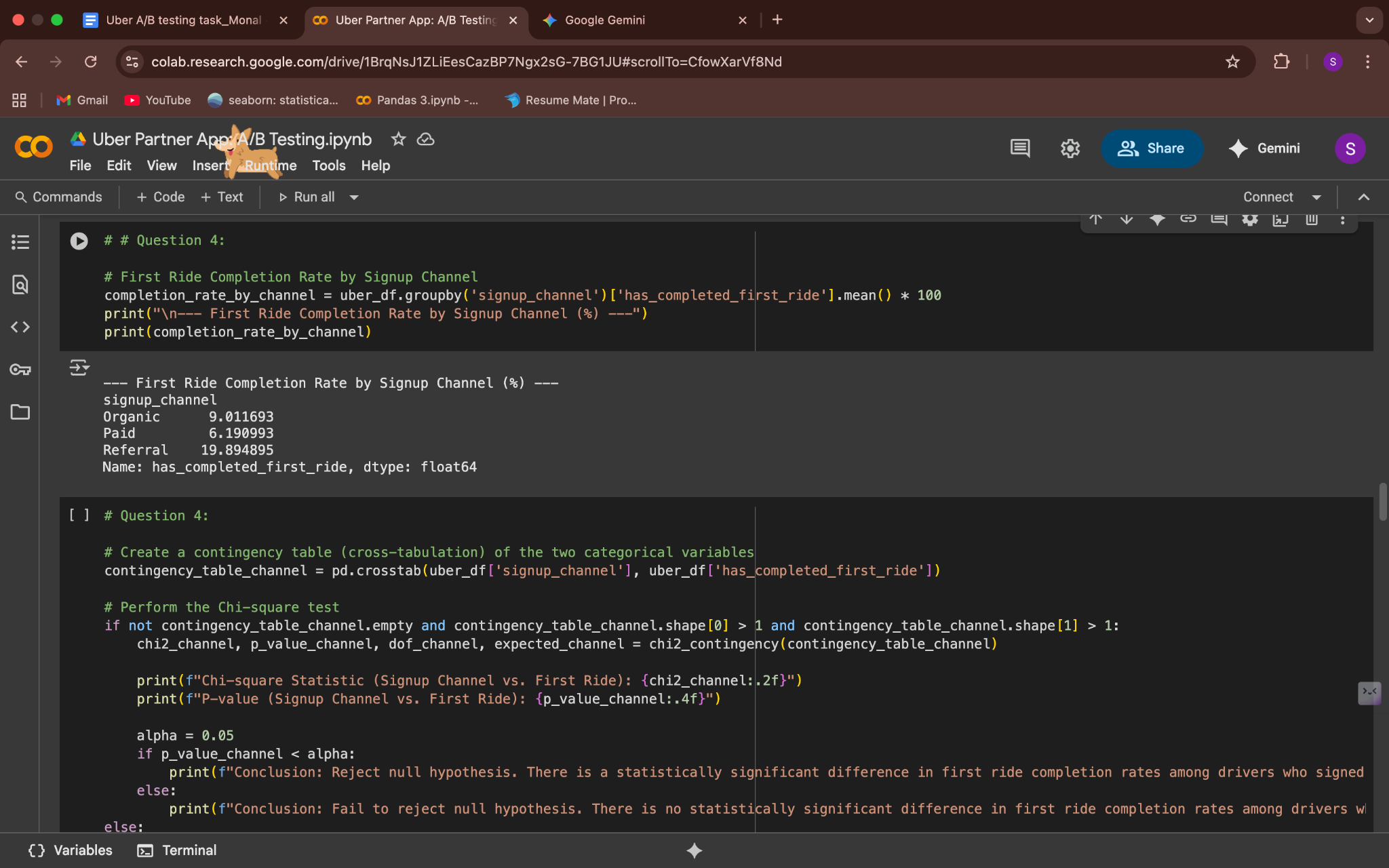
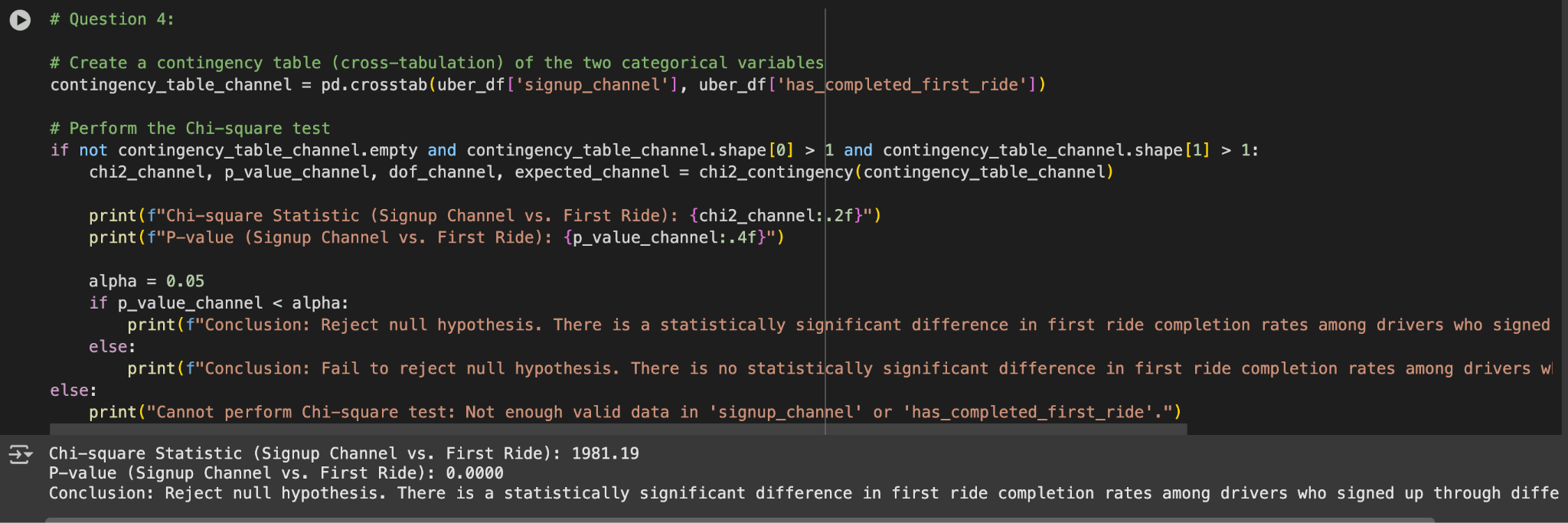
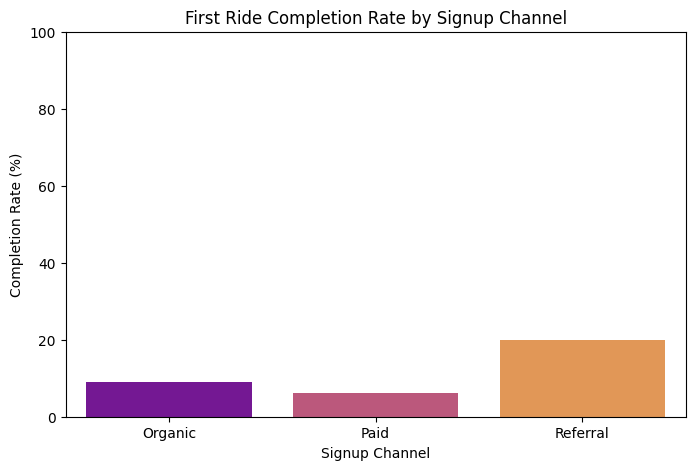
1. **Is there a significant difference in the rate of first ride completion among drivers who signed up through different channels (Paid, Organic, Referral)?**

**Ans:** Yes, there is a statistically significant difference in the first ride completion rates among drivers who signed up through different channels.

**Explanation**:

The **Chi-square test of independence** yielded a p-value of 0.0000, which is significantly less than the standard alpha level of 0.05.

* **P-value < 0.05**: This result allows us to **reject the null hypothesis** (which states that there is no difference in completion rates across the channels).
* **Statistical Significance**: The extremely low p-value confirms that the observed differences in completion rates (as seen in your descriptive analysis, where Referral had a much higher rate) are highly unlikely to be due to random chance.

This finding suggests that the acquisition channel a driver comes from is a strong predictor of their likelihood to complete their first ride. The company should investigate why certain channels, such as Referral, are so much more effective in bringing in activated drivers.

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:**

**Conclusion**:

* It is not possible to perform this analysis with the provided dataset.

**Explanation**:

* The question asks for an analysis of how the new app affected driver earnings and whether this effect varied by city size or market type.
* Upon a careful review of the dataset (Uber A\_B testing.xlsx), we can see that it does not contain any information about driver earnings. There are no columns that track revenue, trip fares, or any other financial metric related to a driver's performance.
* Similarly, the dataset lacks any variables that categorize cities by their size or market type. The city\_name column only provides the city's name but not a classification of its market characteristics.

Therefore, because the necessary data points are missing, a meaningful quantitative analysis of this question cannot be conducted.

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

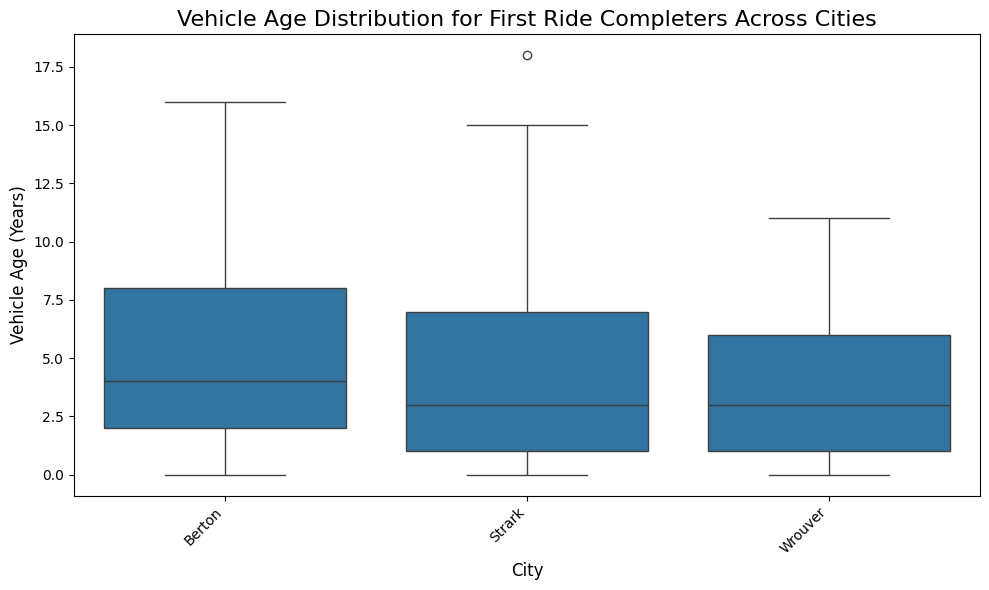
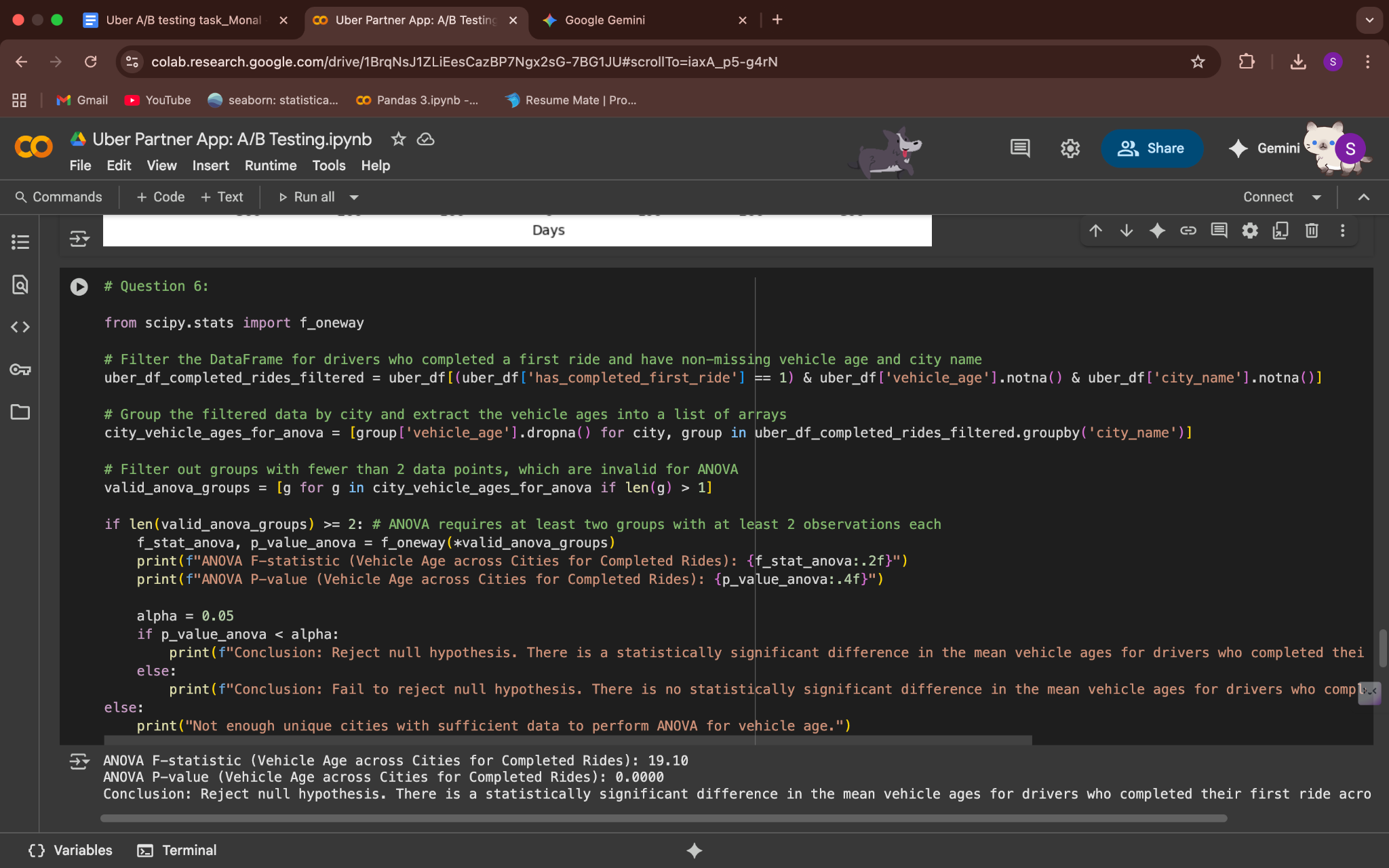
**Ans:**

**Conclusion:**

* Yes, there is a statistically significant difference in the mean vehicle ages for drivers who completed their first ride across different cities.

**Explanation:**

* The one-way ANOVA test yielded a p-value of 0.0000, which is a result of the F-statistic of 19.10. Since the p-value is extremely small and well below the standard significance level of 0.05, we can reject the null hypothesis.
* This means that the differences in the average vehicle age that you observe among the various cities are highly unlikely to be due to random chance. This finding suggests that factors specific to each city, such as local regulations, market preferences, or driver demographics, likely influence the age of vehicles used by active drivers.



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

**Ans:**

**Analysis:**

* It is not possible to perform this analysis with the provided dataset.

**Explanation:**

* The question asks for a correlation between demographic factors (such as city type and vehicle model) and higher earnings rates.
* As identified in the analysis for Question 5, the dataset does not contain any information on "earnings rates". Without a variable representing a driver's earnings, it is impossible to calculate a correlation with any other factor, whether demographic or not.
* Therefore, because the necessary data is missing, this analysis cannot be performed.

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

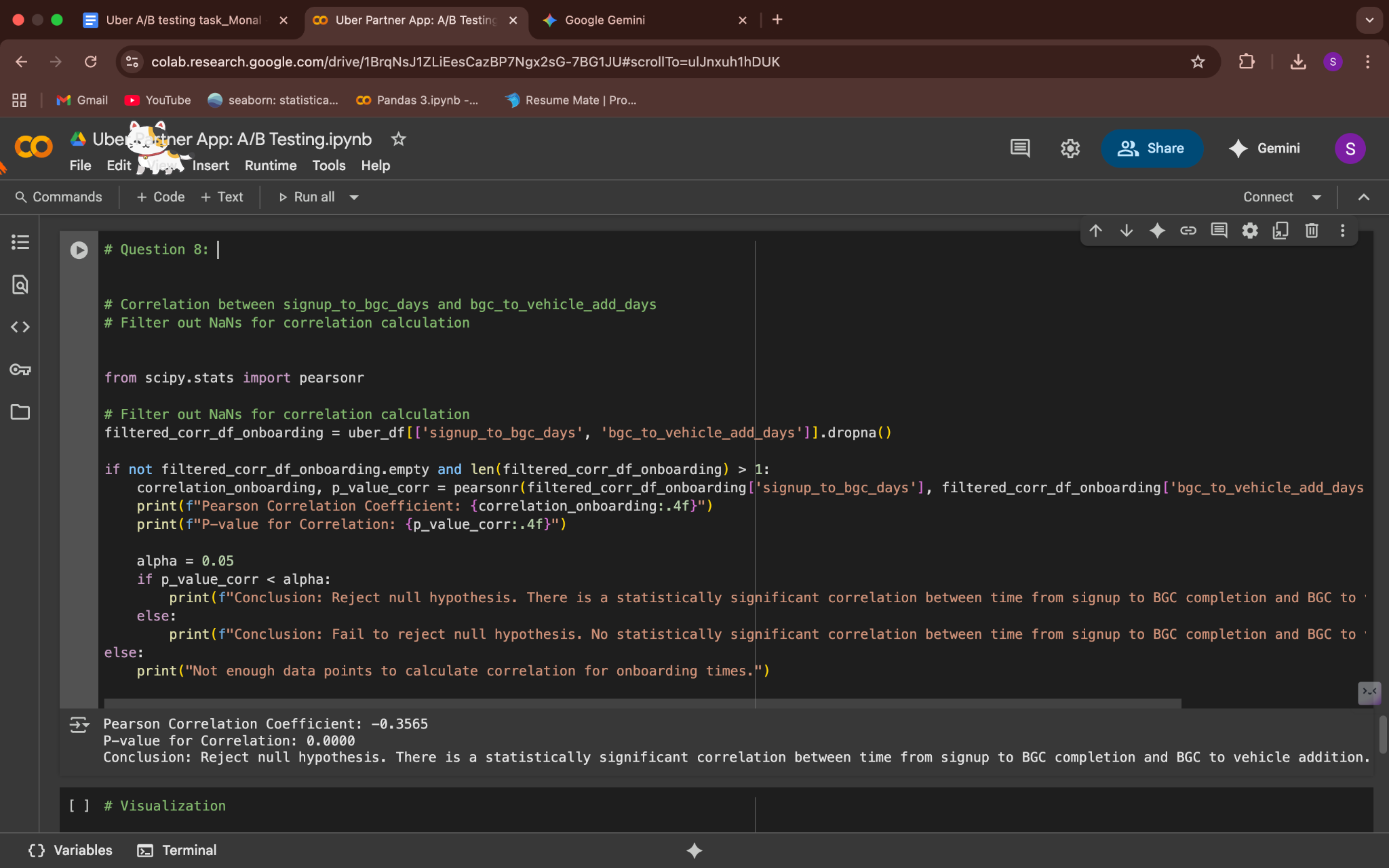
**Ans:**

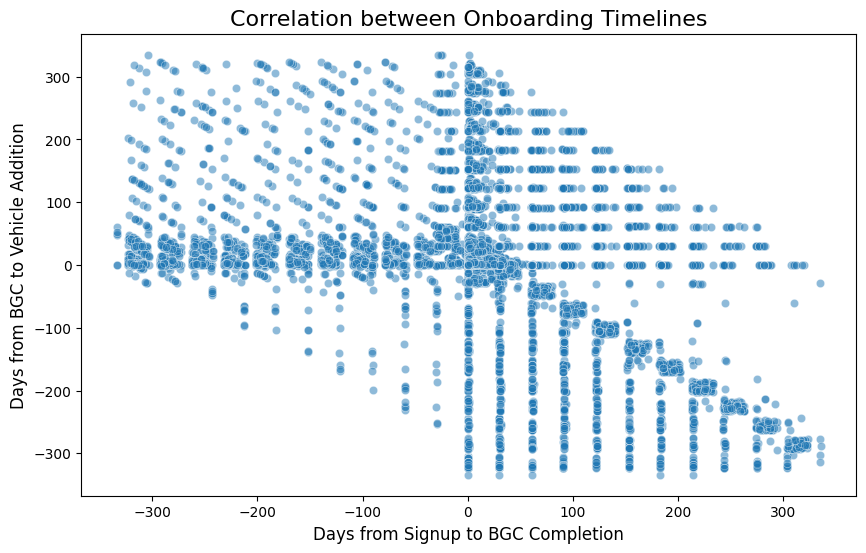
**Conclusion:**

* Yes, there is a statistically significant negative correlation between the time it takes from signup to background check completion and the time from background check to vehicle addition.

**Explanation:**

The Pearson correlation test yielded a correlation coefficient of -0.3565 and p-value of 0.0000, which is significantly less than the standard alpha level of 0.05.

* **P-value < 0.05:** This result allows us to reject the null hypothesis, confirming that a significant correlation exists between these two timeframes.
* **Correlation Coefficient (r=−0.3565):** The negative value indicates a negative correlation. This suggests that as the time from signup to background check completion increases, the time from background check to vehicle addition tends to decrease, and vice versa. While the correlation is statistically significant, it is a moderate relationship, not a perfect inverse one.

****

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:**

1. **A Robust Evaluation Plan**

A robust evaluation plan for this A/B test should encompass three main phases:

**Phase 1: Pre-Experiment Planning**

* **Clear Hypotheses:** Define the null and alternative hypotheses for the primary metric.
* **Null Hypothesis (H₀):** There is no difference in the First Ride Completion Rate between the control (old app) and treatment (new app) groups.
* **Alternative Hypothesis (Hₐ):** The First Ride Completion Rate for the treatment group is significantly higher than the control group.
* **Metric Operationalization:** Precisely define how each metric will be calculated, including the timeframe (e.g., First Ride Completion Rate within 30 days of signup).
* **Power Analysis & Sample Size:** Conduct a power analysis to determine the required sample size for each group. This calculation depends on the desired statistical power (e.g., 80%), significance level (alpha = 0.05), and a minimum detectable effect (the smallest increase in completion rate that is considered a business success, e.g., a 1% absolute increase). This dictates the experiment's duration.
* **Randomization:** Implement a user-level random assignment to ensure that new sign-ups are randomly and evenly distributed between the control and treatment groups, preventing bias.
* **Pre-Analysis Plan:** Document the entire analysis methodology (data cleaning, statistical tests, interpretation rules) before the experiment begins to prevent analysis bias.

**Phase 2: Experiment Execution & Monitoring**

* **Phased Rollout (Optional but Recommended):** Start by releasing the new app to a small percentage of users (e.g., 5-10%) to catch major technical bugs or severe negative impacts early. Gradually increase the rollout to the full treatment group.
* **Data Integrity Checks:** Continuously monitor the traffic split between groups to ensure it remains close to 50/50. Check for any data logging errors or inconsistencies.
* **Guardrail Metrics:** Monitor crucial business metrics that are not expected to change, or should not drop, such as overall driver retention rate, number of support tickets, and app crash rate. If these metrics show a significant negative trend, it would be a signal to halt the experiment.

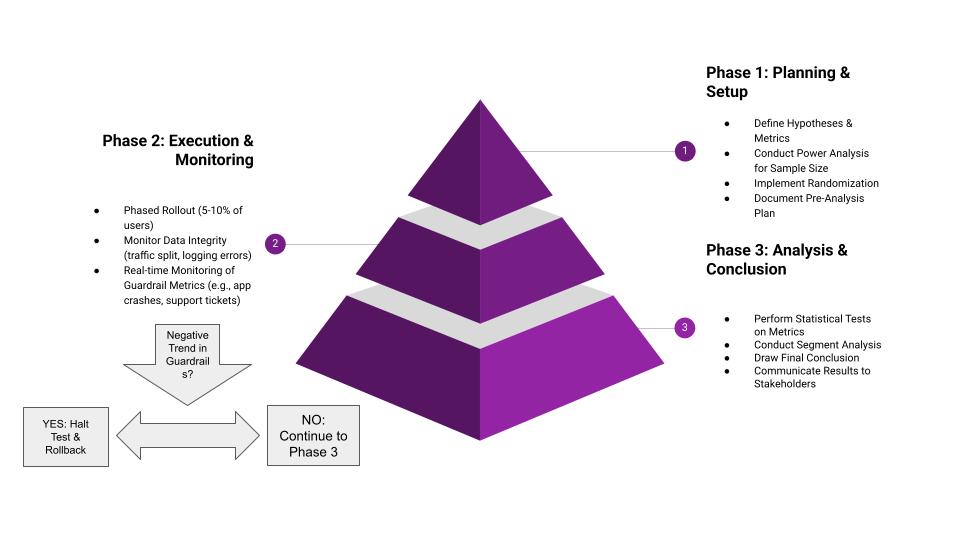
**Phase 3: Post-Experiment Analysis**

* **Statistical Tests:** Once the experiment reaches the predetermined sample size and duration, apply the appropriate statistical tests:
* **For the primary metric (First Ride Completion Rate):** Use a two-sample proportion test (e.g., Z-test) to determine if the difference between the two groups is statistically significant.
* **For secondary metrics (Time to First Ride Completion):** Use a t-test or a non-parametric equivalent to compare the means.
* **Segment Analysis:** Analyze the results across different segments (e.g., city\_name, signup\_channel, signup\_os) to see if the new app had a differential impact on specific driver groups.

1. **Reconciling Rapid Results, Statistical Validity, and Risk Monitoring**

Balancing these three aspects is a critical challenge in A/B testing:

* **Statistical Validity vs. Rapid Results:**
* **Reconciliation:** The desire to get results quickly must be balanced against the need for statistical validity. The predetermined sample size from the power analysis is the minimum duration the experiment should run to ensure the results are reliable.
* **Strategy:** Resist the urge to "peek" at results and stop early. Communicate to stakeholders that prematurely ending a test can lead to false positives (launching a bad feature) or false negatives (killing a good feature).
* **Monitoring Potential Risks:**
* **Reconciliation:** While aiming for a significant result on the primary metric, it is crucial to monitor for unintended negative consequences.
* **Strategy:** Implement real-time monitoring of guardrail metrics. If a guardrail metric (like driver retention or crash rate) shows a significant negative change, an immediate halt and rollback of the new app version may be necessary, regardless of the primary metric's performance. Qualitative data from app store reviews and driver feedback channels should also be monitored for early warning signs of dissatisfaction.



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 Results for an Informed Decision**

Making a final decision on whether to fully implement the new app design requires a comprehensive interpretation of the test results, balancing statistical significance with practical business impact. The process would follow a logical hierarchy:

1. **Analyze the Primary Metric First:**

* The primary metric, First Ride Completion Rate, is the most critical factor. The decision is largely based on its performance.
* **Scenario A:** Positive & Statistically Significant Result: If the First Ride Completion Rate for the treatment group is significantly higher than the control group (as confirmed by a p-value < 0.05), this is the strongest signal to proceed. The magnitude of the increase should also be considered; a small, statistically significant increase might not justify a full rollout if the costs are high.
* **Scenario B:** Negative & Statistically Significant Result: If the First Ride Completion Rate for the treatment group is significantly lower, the new design is detrimental to driver activation. The decision should be to revert to the previous version immediately.
* **Scenario C:** No Statistically Significant Result: If there is no significant difference between the groups (p-value ≥ 0.05), this means the new app did not have a measurable impact on the primary goal. In this case, the recommendation would be to not implement the new design and either return to the old version or pivot to a new experiment.

1. **Use Secondary Metrics as Supporting Evidence:**

* Once the primary metric has been evaluated, secondary metrics provide a more nuanced picture.
* **Reinforcement:** If the primary metric showed a positive trend, and secondary metrics like Time to First Ride Completion also improved (e.g., decreased), it strengthens the case for full implementation.
* **Identifying Trade-offs:** If the primary metric is positive but a critical secondary metric (e.g., driver satisfaction, support tickets) showed a negative impact, this signals a trade-off. A decision would need to be made on whether the gain from the increased completion rate is worth the cost of the negative impact. This could lead to a decision to iterate on the new design to address the issues before a full rollout.

1. **Consider the Holistic Business Context:**

* The final decision is not purely data-driven; it must also consider qualitative factors and business goals.
* Cost-Benefit Analysis: Does the cost of implementing and maintaining the new design outweigh the expected benefits from the increase in activated drivers?
* Strategic Alignment: Does the new app align with Uber's long-term strategy for partner relations and the driver experience?
* Qualitative Feedback: User feedback from surveys, app store reviews, and driver support interactions can provide valuable context that quantitative metrics might miss.

### **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