How Much for a Ride?

TEAM NAME: The Hitchhikers

TEAM MEMBERS: Pin-Hsuan Chen, Yunji Ryu, Kyle Reed, Zijian Liu, Wei Jing

**DATA SCIENCE QUESTION(S) & HYPOTHESIS**:

How does the time of day, location, and trip duration affect the surge in price for Uber rides in NYC?

We expect all these three factors to correlate to the change in price for uber rides in NYC. For example, during rush hour, we can expect the price to go up.

**BACKGROUND**:

Based on our experiences with Uber, the price for riders can vary drastically with no clear indications as to why. According to the FiveThirtyEight article, “*Is Uber making NYC Rush Hour Traffic Worse?*”, there is an increase in Uber demand during rush hour, which could contribute to the fluctuation in price. Instead of making assumptions based on an educated guess, we decided to look into their website and look for any public information about it.

In Uber’s official website, general information about this fluctuation in price was provided only for the Uber drivers. Labeled as surge price, fare calculation and its system are shown and explained for Uber drivers. ([Here’s the link](https://www.uber.com/drive/partner-app/how-surge-works/)). Although the website is public and both parties, including riders and drivers, the separate section for riders did not include information about surge pricing and that might be one of the reasons why most of the riders are unaware of its existence and how it works until the Uber app alerts of the sudden price change.

As we mentioned in our question and hypothesis section, surge prices are calculated based on location and time during the day. According to the website, the surge pricing is determined by two combined factors; the rider’s location as well as the real-time demand by the riders in the same specific area. Depending on Uber’s jurisdiction on these specific areas, riders are charged with fixed rates of either 1.8x or 2.5x, which is multiplied on the already existing base ride fare. These multipliers are also bound by specific locations. Certain locations have 2.5 multiplication rate while others have a different rate or no rate at all. Although Uber has a specific map inside the app that shows the multiplier rate in the current and nearby area, they did not get into details on how the jurisdiction is made to each individual area and how the rates are justified. In order to investigate this information, we decided to use our readily available public data to analyze and infer to the possible correlations listed in our hypothesis.

**ETHICAL CONSIDERATIONS**:

The data set we are analyzing for this project is public data obtained from an open source. The data on its own does not reveal any private information about individual users as well as Uber drivers so we are not as concerned with personal privacy. However, analyzers could still easily extract a lot of information from these data. For instance, analyzing this data set could reveal users’ commute patterns and preferences. Though this does not point to a specific person, it could be information that could be sold to other companies for profit. From this, companies can target users based on their individual preferences. For example, they could distribute advertisements based on these preferences. Or, rideshare apps, like Uber, could take advantage of all these data and deliberately raise prices when the demand is high.

**DATA**: [link to data](https://github.com/fivethirtyeight/uber-tlc-foil-response/)

We will be analyzing a dataset of uber and taxi data collected by statistics-based journalism website FiveThirtyEight and hosted in a public repository on GitHub. The data consists of trip data of all uber rides taken in New York, NY through several months in 2014 and 2015, and was retrieved from the New York Taxi and Limousine Commission via a Freedom of Information Law (FOIL) request. The dataset contains information about the exact date and time of the ride, as well as precise latitude and longitude information of where the ride was hailed. However, it contains no information about ride price, which is crucial for our question. We may have to submit a FOIL request ourselves to obtain that side of the data.

TEAM EXPECTATIONS AGREEMENT

Read over the [COGS108 Team Policies](https://docs.google.com/document/d/16JydXvqHzvIdaT-xwItD-9PhQJ5wmP5cIoRNiWK-e8g/edit?usp=sharing) individually. Then, include your group’s expectations of one another for successful completion of your COGS108 project below. Discuss and agree on what all of your expectations are. Discuss how your team will communicate throughout the quarter and consider how you will communicate respectfully should conflicts arise. By including each member’s name above and by adding their name to the Gradescope submission, you are indicating that you have read the COGS108 Team Policies, accept your team’s expectations below, and have every intention to fulfill them.

These expectations are for your team’s use and benefit—they won’t be graded for their details. Goals should be realistic: “No group member will never miss a meeting and everyone will always show up early” is probably unrealistic, but “Group members will attend almost every meeting and will communicate their absence at least a day in advance of the group meeting” and “When group members are unable to attend a meeting, they will submit their notes and progress ahead of the group meeting” are realistic expectations. Expectations for deadlines, how you’ll work together, meeting attendance and participation, and project completion should all be considered and details included below.

**INCLUDE YOUR TEAM’S EXPECTATIONS HERE**

**We expect everyone to be at our weekly team meeting (Thursday 3:30pm - 5:00pm). We also expect everyone of our members to communicate effectively via Messenger and/or Discord if we need to host a group voice call. At each meeting sessions, we’ll review our work and discuss future plans. For this, we expect all of our members to work on their respective tasks before each weekly meeting.**

Our primary means of communication will be messenger and discord.

Discord info: https://discord.gg/yU3nE2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Zijian L. | Kyle R. | Yunji R. | Pin-Hsuan C. | Wei Jing |
| Wabbajack#3456 | DexerVonDexer#2801 | Yuyu#9816 | winniechen#2100 | Kneone#9384 |

PROJECT TIMELINE PROPOSAL

Include actual dates and times for due dates and meetings below, not just what week they’ll be completed

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Draft Text?** | **Write Code?** | **Proposed due date** | **Discuss at team meeting** | **Edit?** |
| **Initial team meeting** | NA | NA | NA | week 2 | NA |
| **Background Research** | 4/18 | NA | week 3 | week 4 | 4/18 |
| **Question & Hypothesis** | 4/18 | NA | week 3 | week 4 | 4/18 |
| **Ethical Considerations** | 4/18 | NA | Week 3 | Week 4 | 4/18 |
| **Dataset** | 4/18 | 4/18 | week 3 | week 4 | 4/18 |
| **Data Wrangling** | 4/25 | 4/25 | week 3 | week 4 | 4/25 |
| **Descriptive** | 4/25 | 4/25 | week 5 | week 6 | 4/25 |
| **Exploratory** | 5/2 | 5/2 | week 5 | week 6 | 5/2 |
| **Analysis - Part I** | 5/2 | 5/2 | week 6 | week 7 | 5/2 |
| **Analysis - Part II** | 5/9 | 5/9 | week 6 | week 7 | 5/9 |
| **Analysis - Part III** | 5/9 | 5/9 | week 6 | week 7 | 5/9 |
| **Summarize Results** | 5/16 | NA | week 7 | week 8 | 5/16 |
| **Conclusions** | 5/23 | NA | week 7 | week 8 | 5/23 |

**Once completed, save this document as a PDF & submit on Gradescope. Be sure to add each team member’s name to the Gradescope submission.**