|  |  |  |
| --- | --- | --- |
| PROJECT PROPOSAL | | |
| Project Title | | Ride Sharing in Chicago |
| Group Team Members | | |  |  | | --- | --- | | Naga Viraja Chunduru | [**naga.viraja@gmail.com**](mailto:naga.viraja@gmail.com) | | Hetal Modi | [**hetumodi@yahoo.com**](mailto:hetumodi@yahoo.com)**>** | | Chris Figy | [**ChrisFigy@gmail.com**](file:///C:\Users\U358343\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.Outlook\FLUEDOYF\ChrisFigy@gmail.com) | | Lorraine Kriho (Lynne) | [**LynneKriho@gmail.com**](mailto:LynneKriho@gmail.com) | | Rauf Nugmanov | [**nugmanov.rauf@gmail.com**](file:///C:\Users\U358343\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.Outlook\FLUEDOYF\nugmanov.rauf@gmail.com) | |
| Dataset to be used (including the source) | | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Description of Data | Source Link | **File Type** | **Row** | **Col** | **Sample Data** | | Ridesharing Trips - Uber, Lyft, etc for City of Chicago being November 2018 required by City Ordinance. | <https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p> | CSV | 129M | 21 |  | | Chicago Weather - This data will need to be extracted by Month and determined if hourly or daily will be need. | <https://www.meteoblue.com/en/weather/archive/export/chicago-o%27hare-international-airport_united-states-of-america_4887479> | CSV | Varies Daily Hourly by Month | 9 |  |  |  |  | | --- | --- | | Data Profiling of Subset of Ridesharing Trips |  | | | |
| Problem Description and Scope | | |
| Our analysis will try to look at ridesharing and how the weather impacts usage in the city of Chicago. In additional, we will look at trip’s lengths, busiest time periods and cost of fare. We will analyze neighborhoods and does location impact ride cost. | | |  | |
|  | | |  |  | |
| Proposed Approach and Methodology | | |  |  | |
| Biggest Challenge | The dataset is extremely large and contains multiple trips by time and date.  To handle the data on our laptops, we plan to create a random sample using an API to scrape data and randomly selecting records to create a sample dataset. | |
| Statistical methods to use to gather insight | We will look at distributions of the data and statistical information to understand outliers and evaluate the missing data and determine how to handle for each situation. | |
| Data preparation and analysis | Subset and filter out data using R since the dataset is so large. We will aggregate data to create visual representations of the data to tell our story using Tableau and R. We will combine Weather and Ridesharing Data. We will plan to look at Seasons, days of week, times of day, longitude & latitude, and census tract, and fares for our visualizations. We will clean our data set by running statistical analysis and analyzing for missing data. We will profile all the data to understand frequency distributions, correlations, and distinct values so we can have further insight. Create numerous graphs of data during exploration. | |