Analyzation of Traffic Patterns in Relation to Delays in Boston Public Transit

Group Members

Cristian Mendivil
Luke Lyon
Theodore Freeman
Vamshi Arugonda

Questions Sought to Answer

Questions we aim to answer:

- Do delays in public transit times correlate with increased travel times for uber?
- What are the most influential weather patterns affecting bus travel time?
- Based on the delay from weather patterns, how much extra travel time should a Boston commuter account for?

Our group has been analyzing the difference between the frequency/ timeliness of Uber rides and the time reliability of public transit in the Boston area. We have developed an average travel time on Uber for various start and stop points throughout the city that correspond to the most utilized bus routes. We have also looked at which weather patterns affect transit times in order to provide the boston commuter with reliable data on when to expect delays

Data Preparation

Cleaning - Removed unuseful attributes and incomplete entries. Some entries in MTBA set were not useful and were removed.

Integration - For bus/weather analysis data joined together on date.

Reduction - Attributes in some data sets were redundant, and were removed.

List of Tools

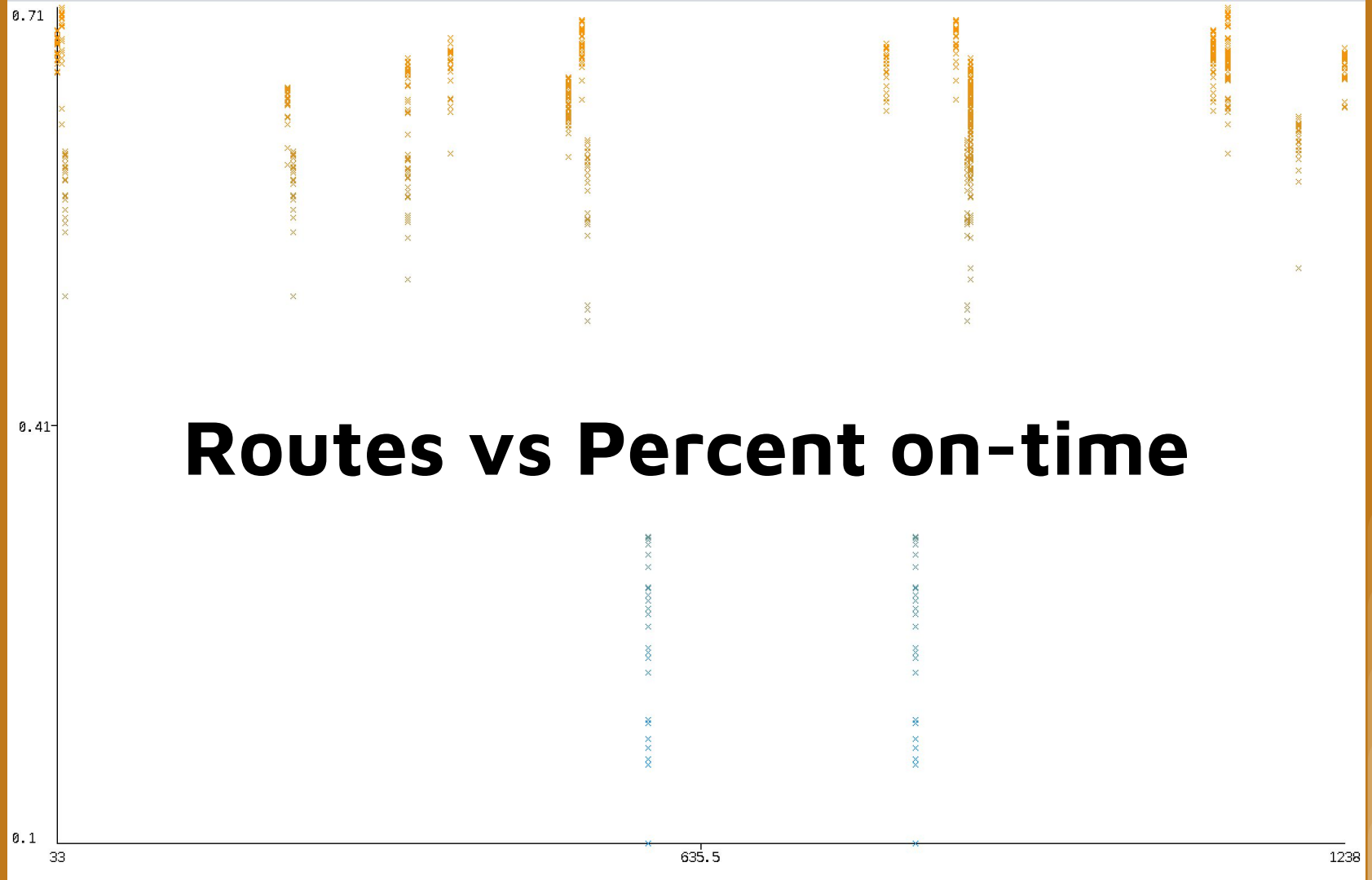
- Github
- Jupyter Notebooks
- Pandas
- Matplotlib
- NumPy
- SciKit Learn
- WEKA
- Google Drive

Mining Techniques

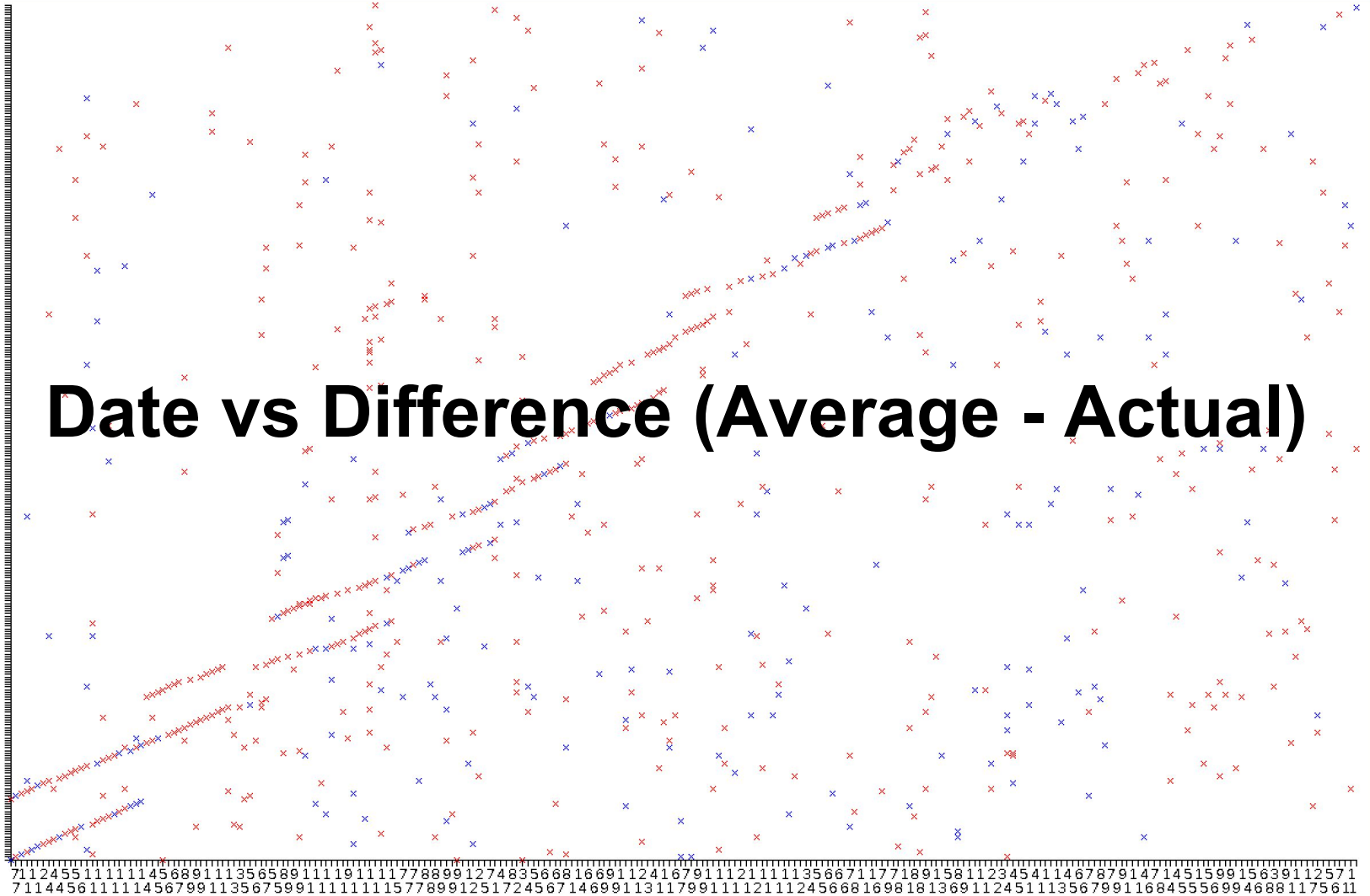
Correlation Analysis

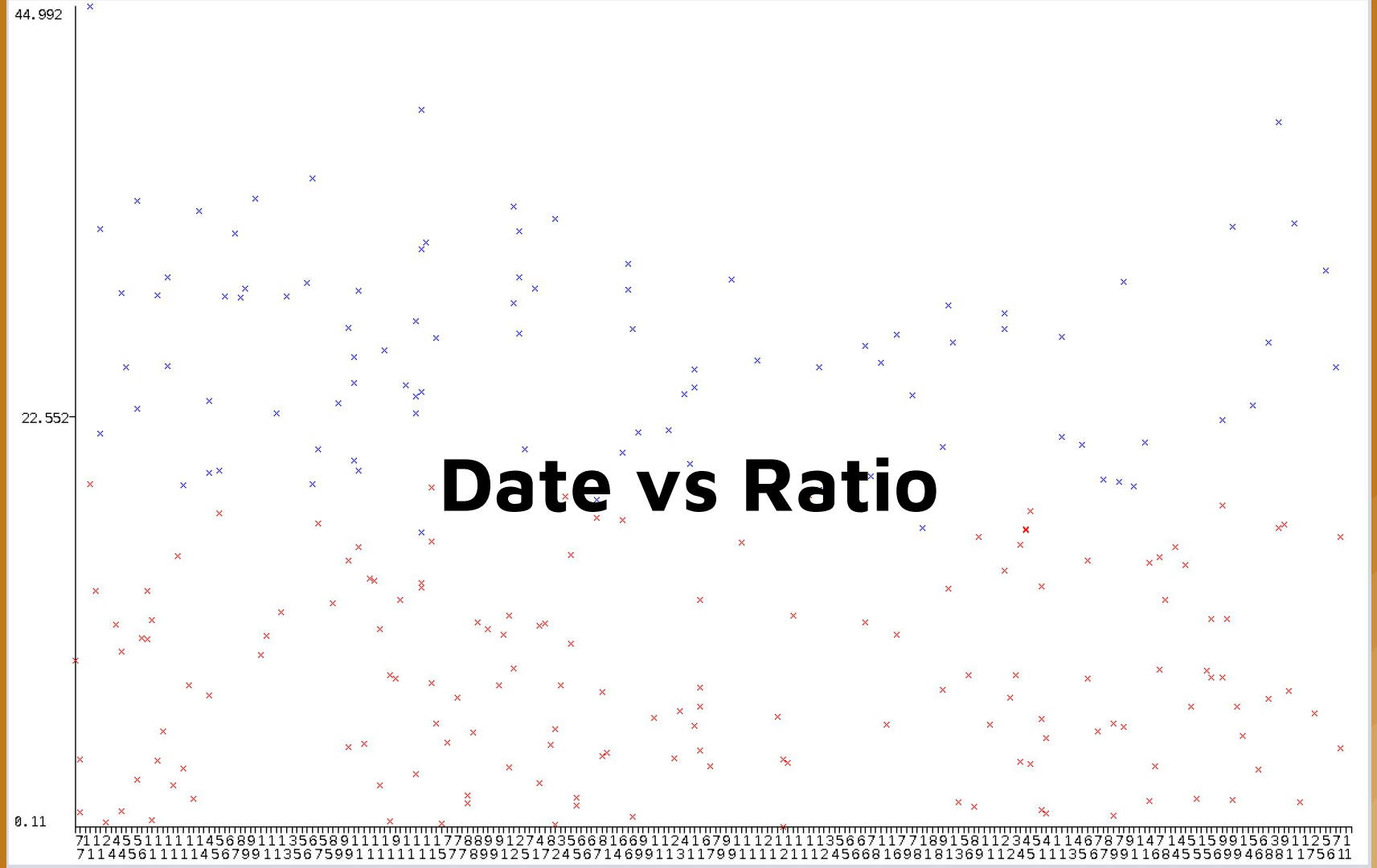
Regression models such as Logistic, Random Forest, Neural Network

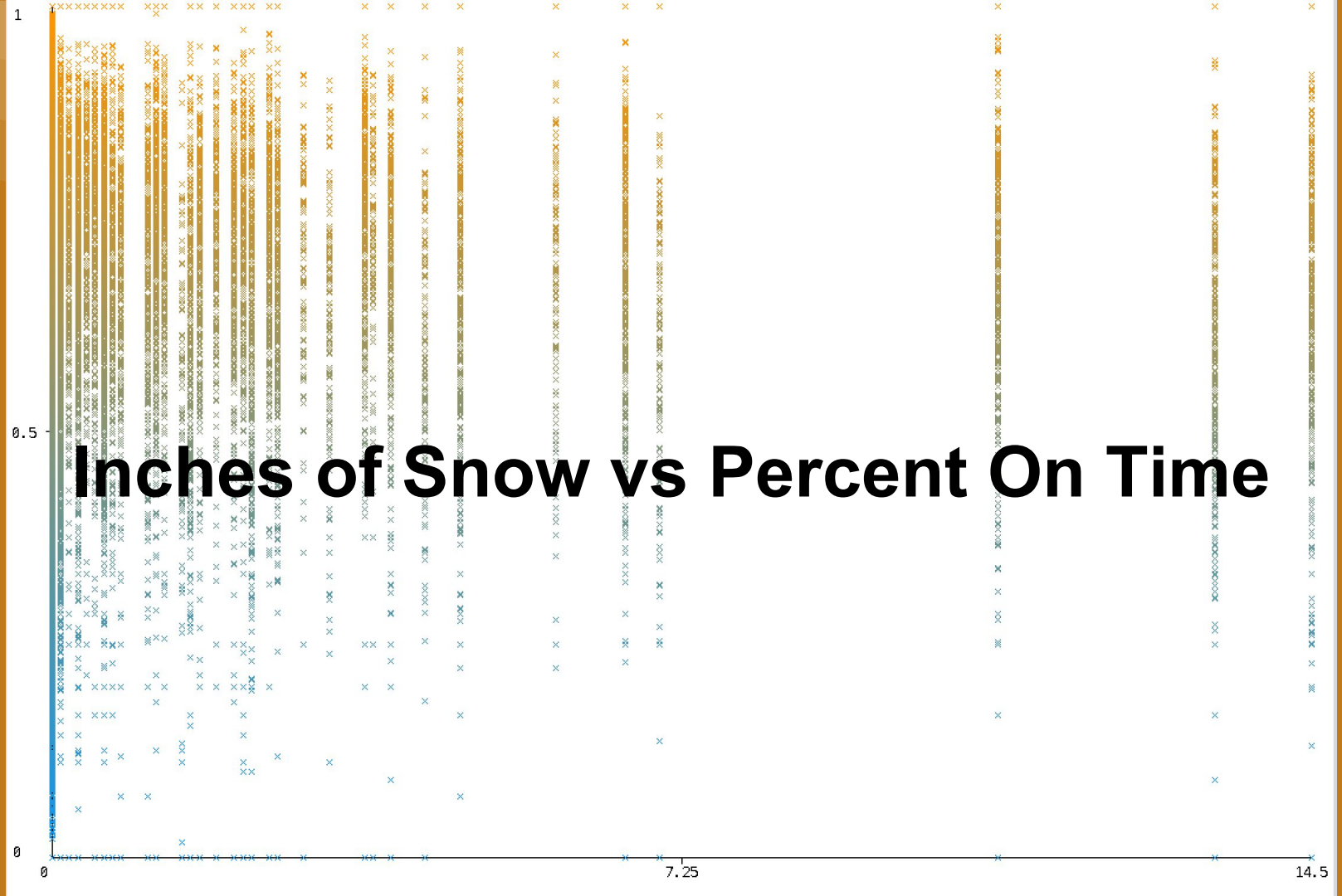
K-Means Clustering



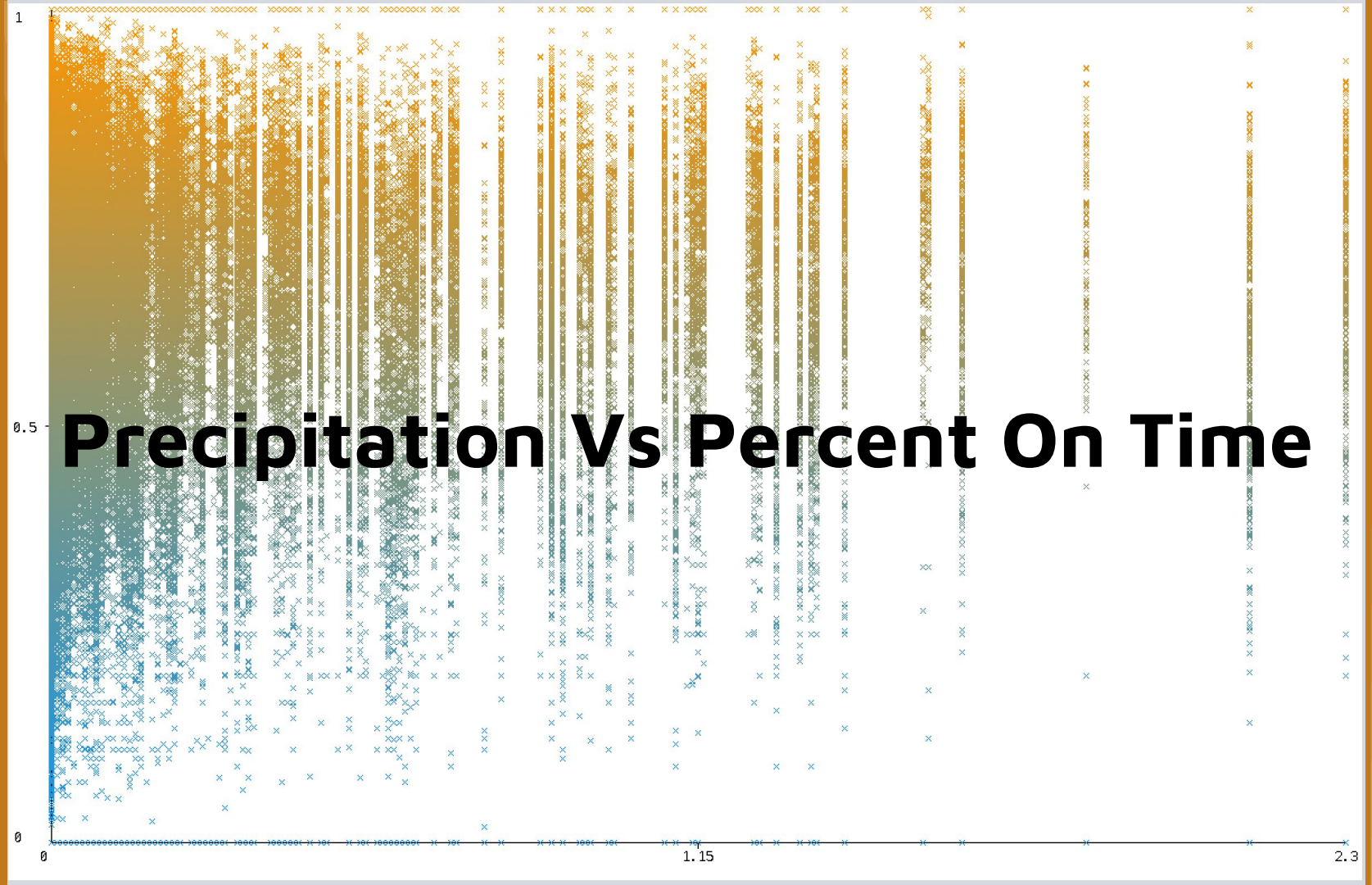
Routes vs Percent on-time


Date vs Difference (Average - Actual)


44.992
22.552
0.11
Date vs Ratio



Inches of Snow vs Percent On Time



Precipitation Vs Percent On Time

Knowledge Gained

Bus and Uber travel times generally supported positive correlation

The regression algorithms were unable to produce a model that accurately predicted percent of buses on time given weather conditions - likely there are more factors that need to be considered to predict this.

Applications

Uber drivers could use the MTBA real-time data streams to try to forecast what their delays might be and plan their routes accordingly

We could look for further data to incorporate into building a predictive model for percentage of buses on time