

MARTA Optimize

Loren Klingman, Andre Hijaouy, Alora Wogsland, Kaushik Visvanathan, Sid Biswal, James Wang, and Osman Ghandour

MARTA Hackaton

Objectives

Using the MARTA APC data from the month of July, the goal of our project was to:

- Process APC data to calculate trip time
- Simulate current train schedule with rider data
- Optimize train schedule with machine learning
- Visualize our new train schedule

Introduction

Our group processed the APC data from July 2016 into a usable format to minimize trip time and delays. Combining the current train schedule, the distance between the stations and the processed rider data, a program was built to simulate the MARTA train system. Following that, we worked to build visualization for the simulation. The end goal from there was to optimize the simulation with machine learning to decrease overall time spent inside the MARTA system.

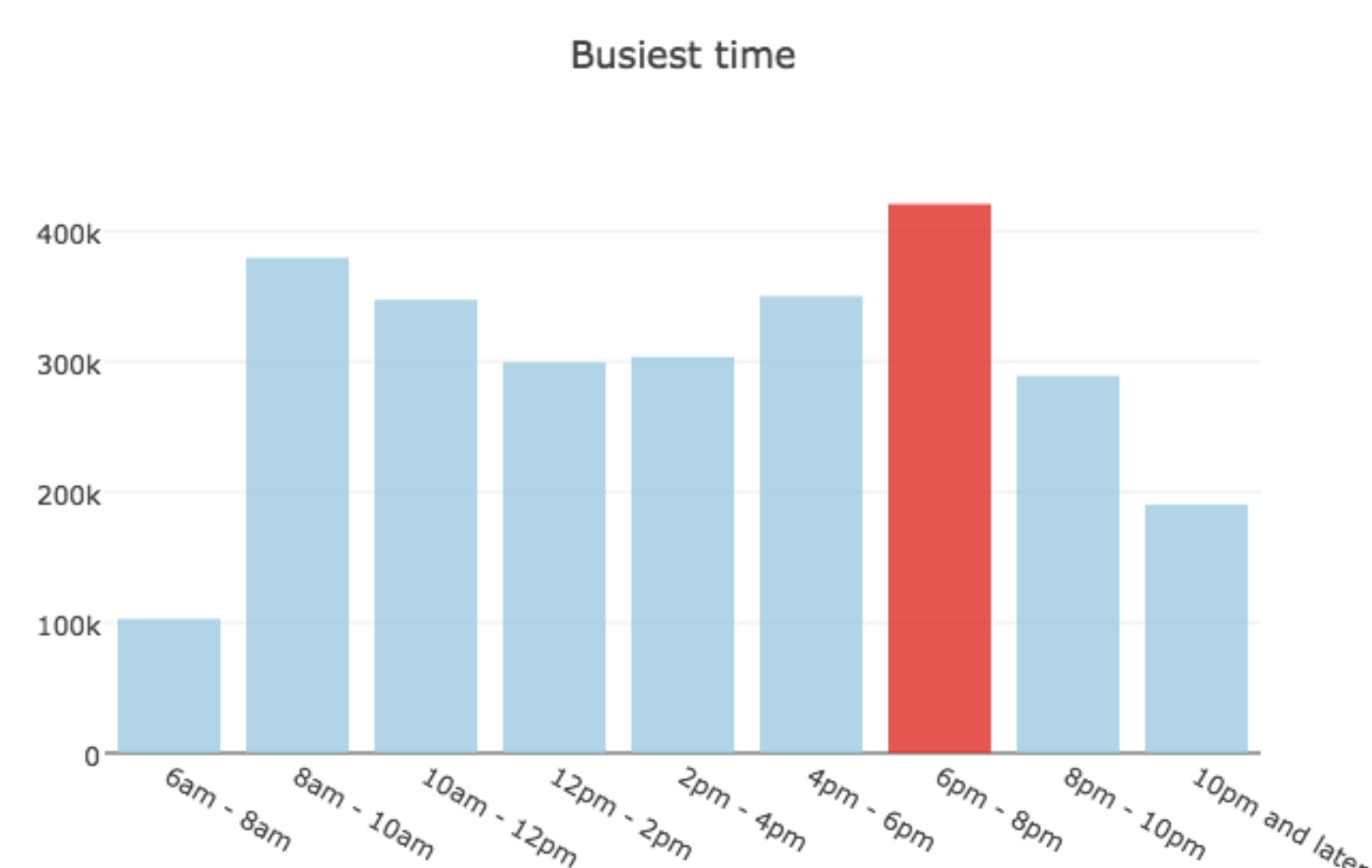


Figure 1: The busiest time of day on Marta is between 6 and 8pm

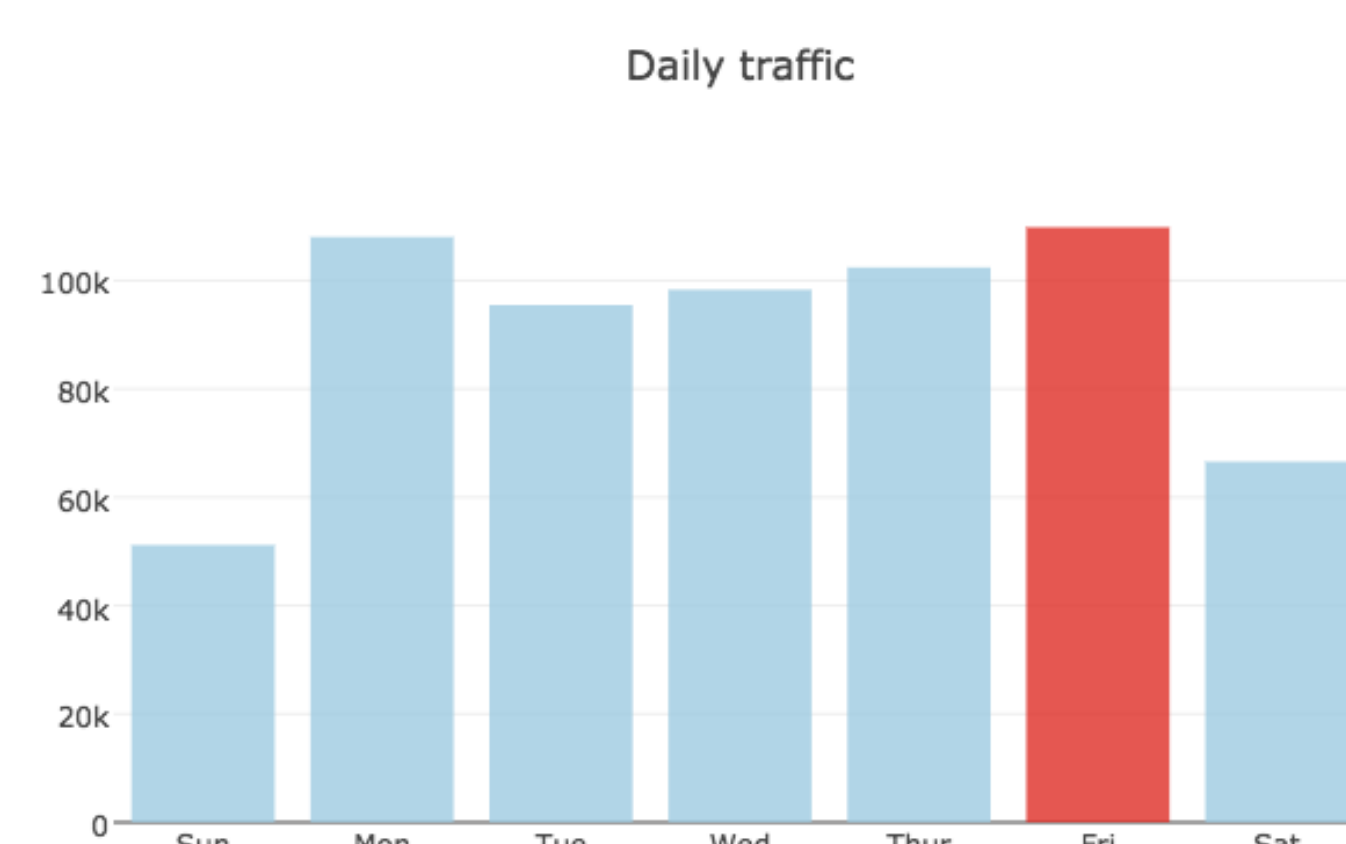


Figure 2: Friday was the busiest day of the week on MARTA in July 2016

APC Data Processing

To create the trip time data, we used the APC data collected during July 2016. The data was stored in a csv with entries and exits denoted by use_type. The data was cleaned. Then the data was separated into two tables one entry-only, one exit-only to be processed. The final processed data is represented in Table 1.

entry_station	entry_time	exit_station	exit_time	serial_num	trav_time
36	2016-07-01 00:07:00	47	2016-07-01 00:07:03	3E8CADDD4A774E4A	0:00:03
39	2016-07-01 00:07:00	41	2016-07-01 00:07:12	3E8CAE0248C14E4AE050	0:00:12

Table 1: Processed data

The entry and exit data was processed by iterating over both tables by serial number. Entry and exit times and stations were combined as one trip for the smallest difference between the entry and exit times for all instances of that serial number. This was necessary because each serial number could appear many more than twice for people who rode MARTA more than just once in the month of July. The final processed data was the exported to csv for use in the machine learning program.

MARTA Schedule Simulation

We created a program to simulate riders in the MARTA system based on the current train schedule, the processed APC data, and the distance between the stations. In our experimentation, we found that when you set dwell time at Five Points for all trains which start after 8:30pm or before 6am, you see the total times in Figure 3. For this we tried the current schedule and 1 - 3 minutes of dwell time at five points for all trains so more people can make connections. As the figure shows, the increased delay decreased the total time spent on MARTA.

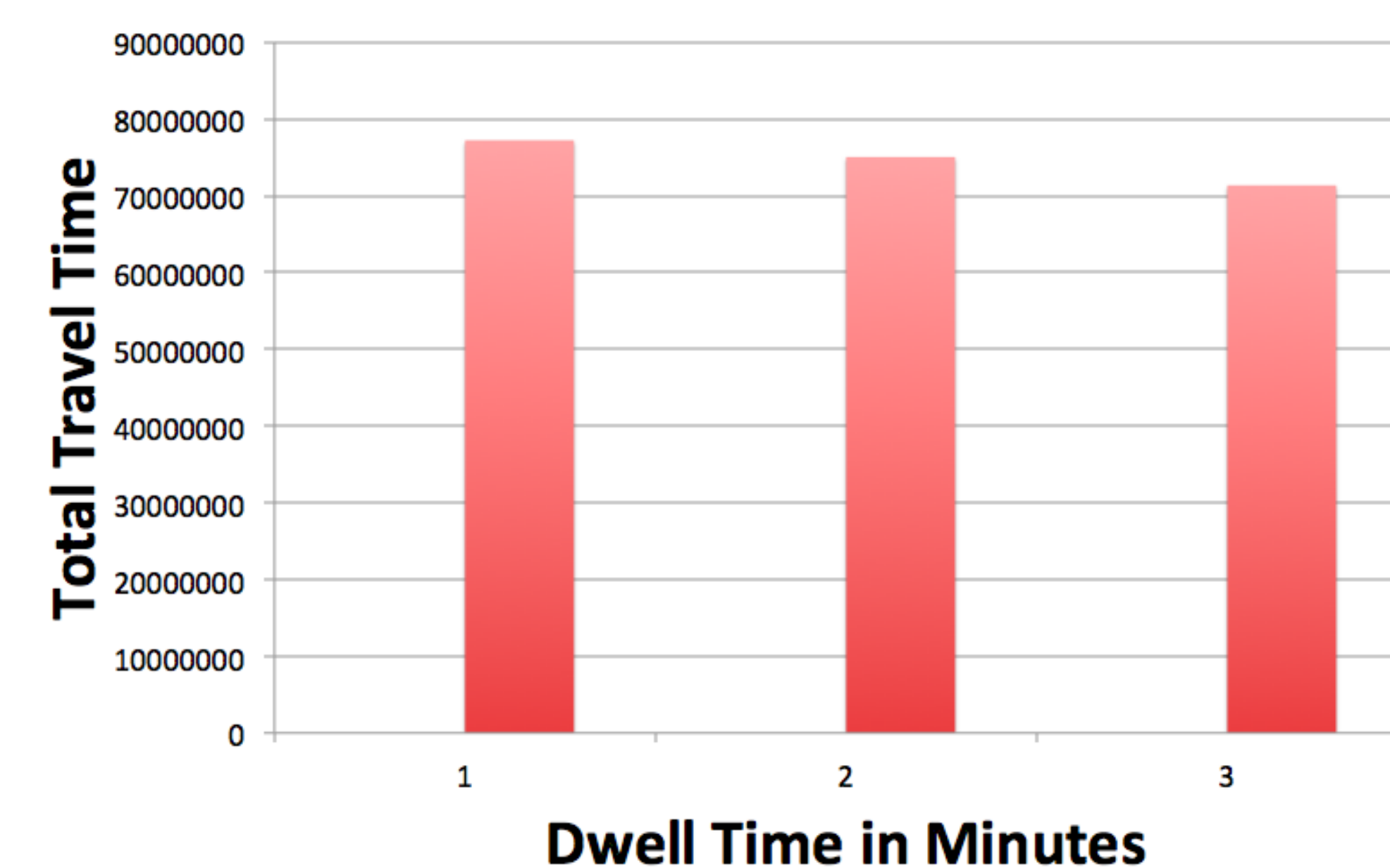


Figure 3: Dwell times at Five Points and their effect on total time on MARTA for the month.

Machine Learning

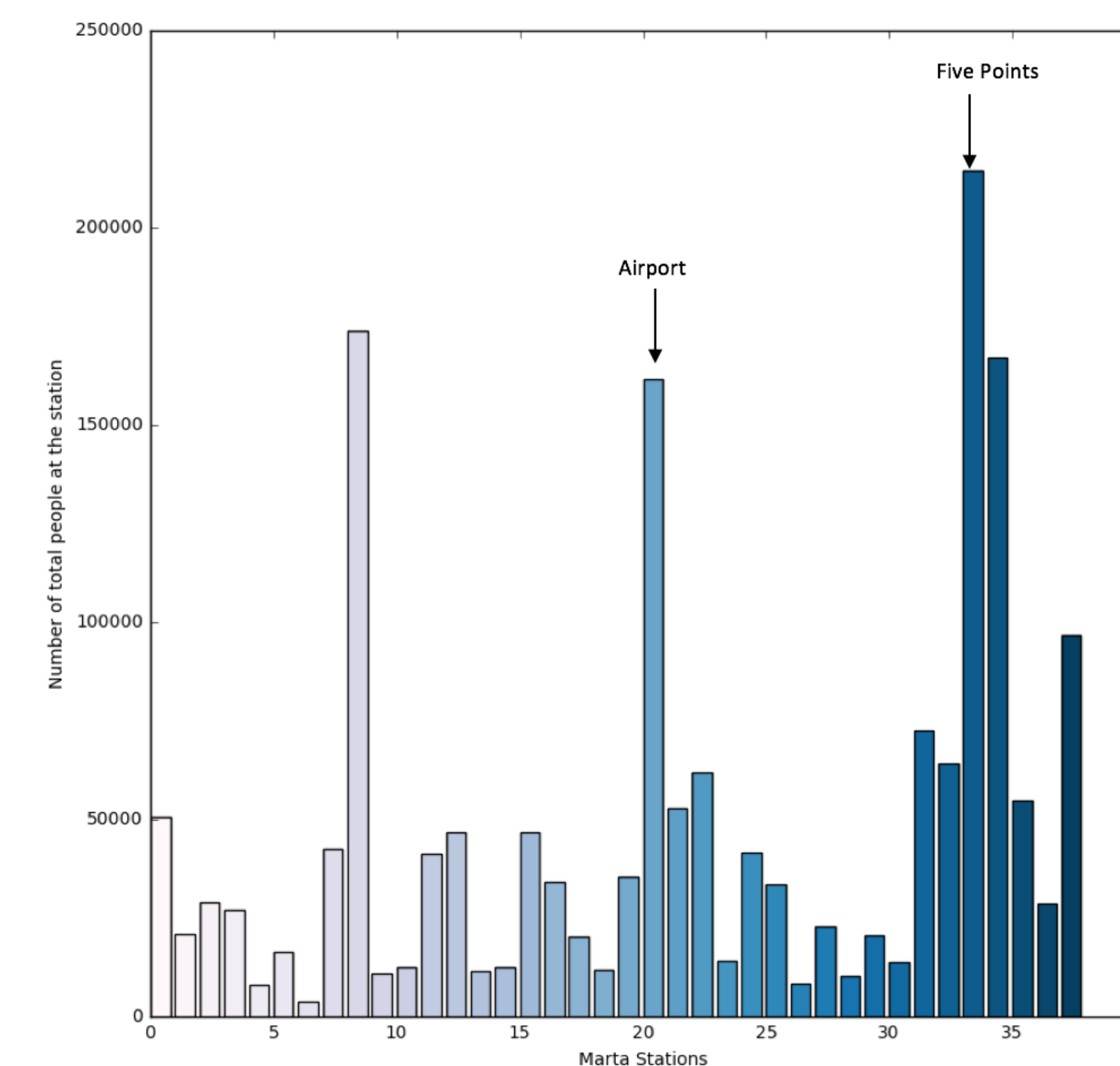


Figure 4: We found out which stations had higher volume of travelers and looked at the relationship between that number of people and amount of time spent at different stations. We believe that optimizing for the right amount of time at a station can reduce the total travel time for passengers. We also built a regression model to predict to find important factors related to travel time.

Conclusion and Next Steps

In July of 2016, commuters spent 72,121,074 minutes riding MARTA over the weekdays. The goal of our project was to optimize the MARTA train schedule in such a way that it would decrease the amount of time riders spent on MARTA, by minimizing delays, especially around Five Points Station. Over the course of this project, we have done a lot of work, but more time is necessary to reach our end goal. However, we have created a working MARTA simulation based on actual MARTA data. We found interesting things about the current system and, with further research, could implement machine learning to attempt to reach our goal.

Acknowledgements

Thanks to all the hosts and sponsors of this event who are the only reason this project came to be.