
MTA EDA

COFFEE CART CATALYST

INTRODUCTION

- **Data Used:**
 - **MTA Turnstile Data**
 - **Google Maps Business Results**
- **Tools Utilized:**
 - **DB Browser for SQLite**
 - **Python Pandas, Seaborn, Datetime, Numpy, and Matplotlib libraries**

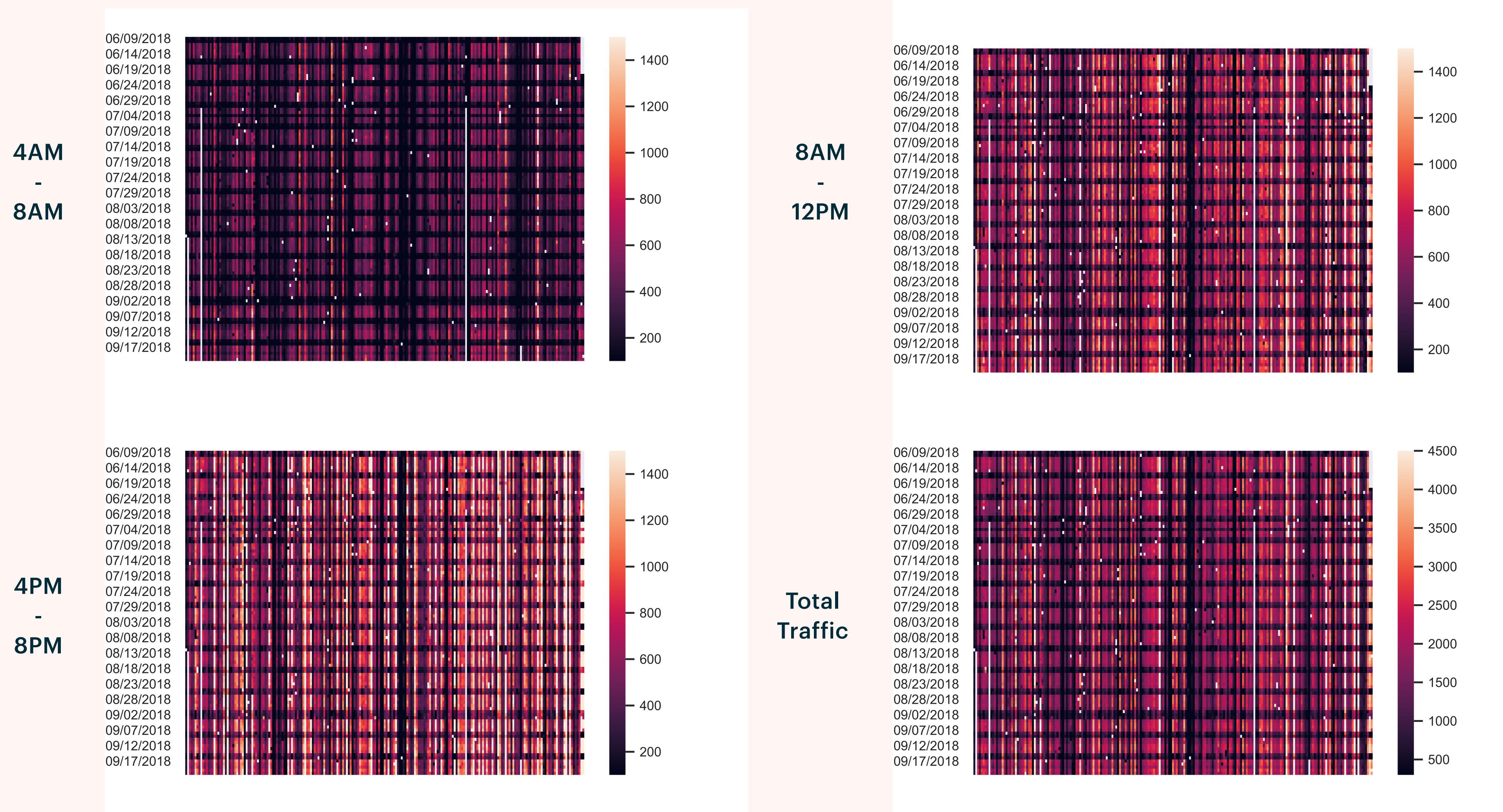


METHODOLOGY

- **Data Collection and Cleaning via SQL:**
 - Raw data scraped from MTA database and dropped into local SQL database
 - Data columns cleaned, and filter markers inserted via CASE statement
 - Data compiled back into python by date and time-slot using filter marker
 - **Data aggregation in python**
 - Time-slots separated into 4-8am, 8am-12pm, and 4pm-8pm to capture standard commuter rush hours
 - Station traffic sorted by varying time-slot to show highest traffic by station by time
-

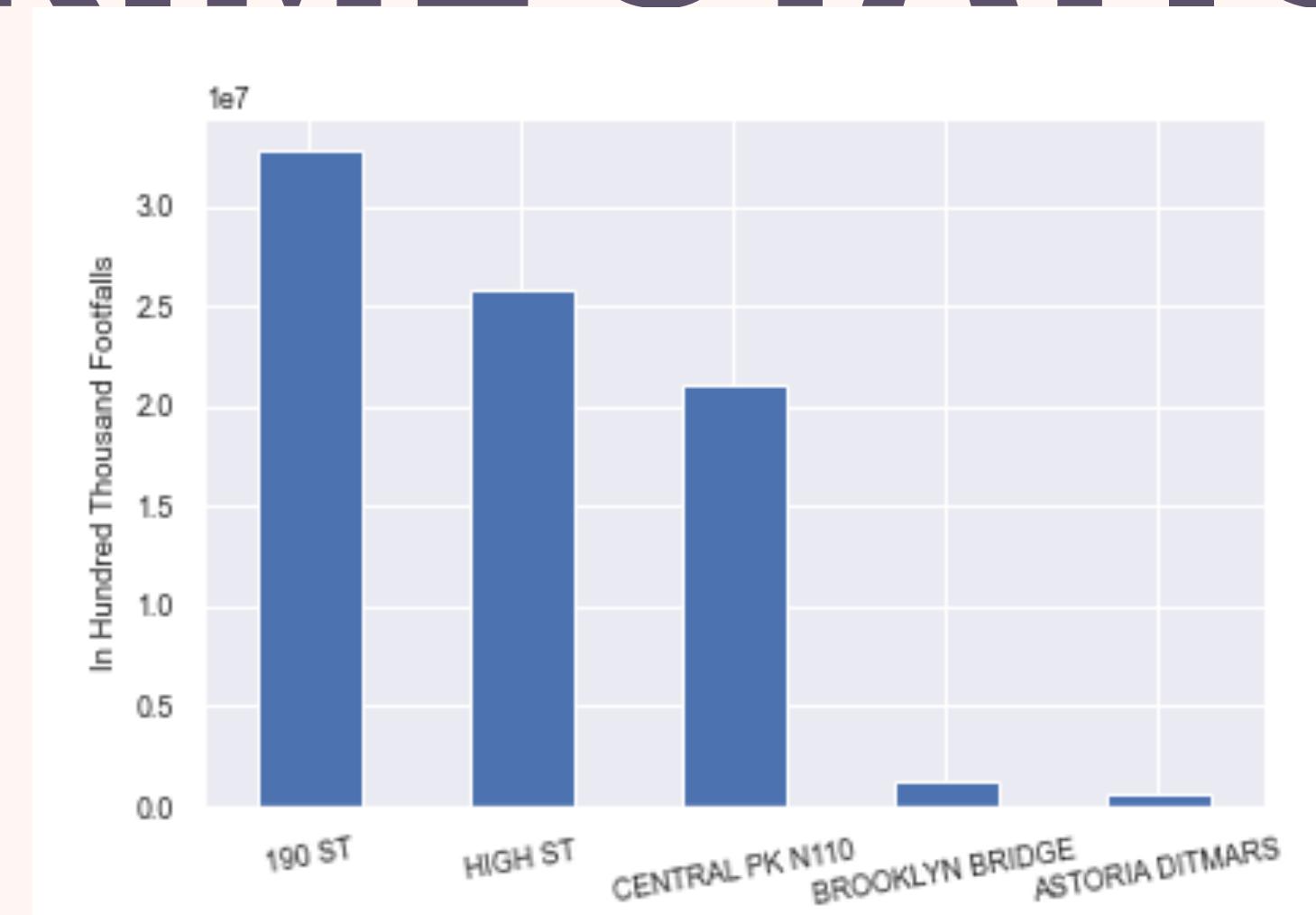
RESULTS

SUMMER TRAFFIC BY TIME-SLOT

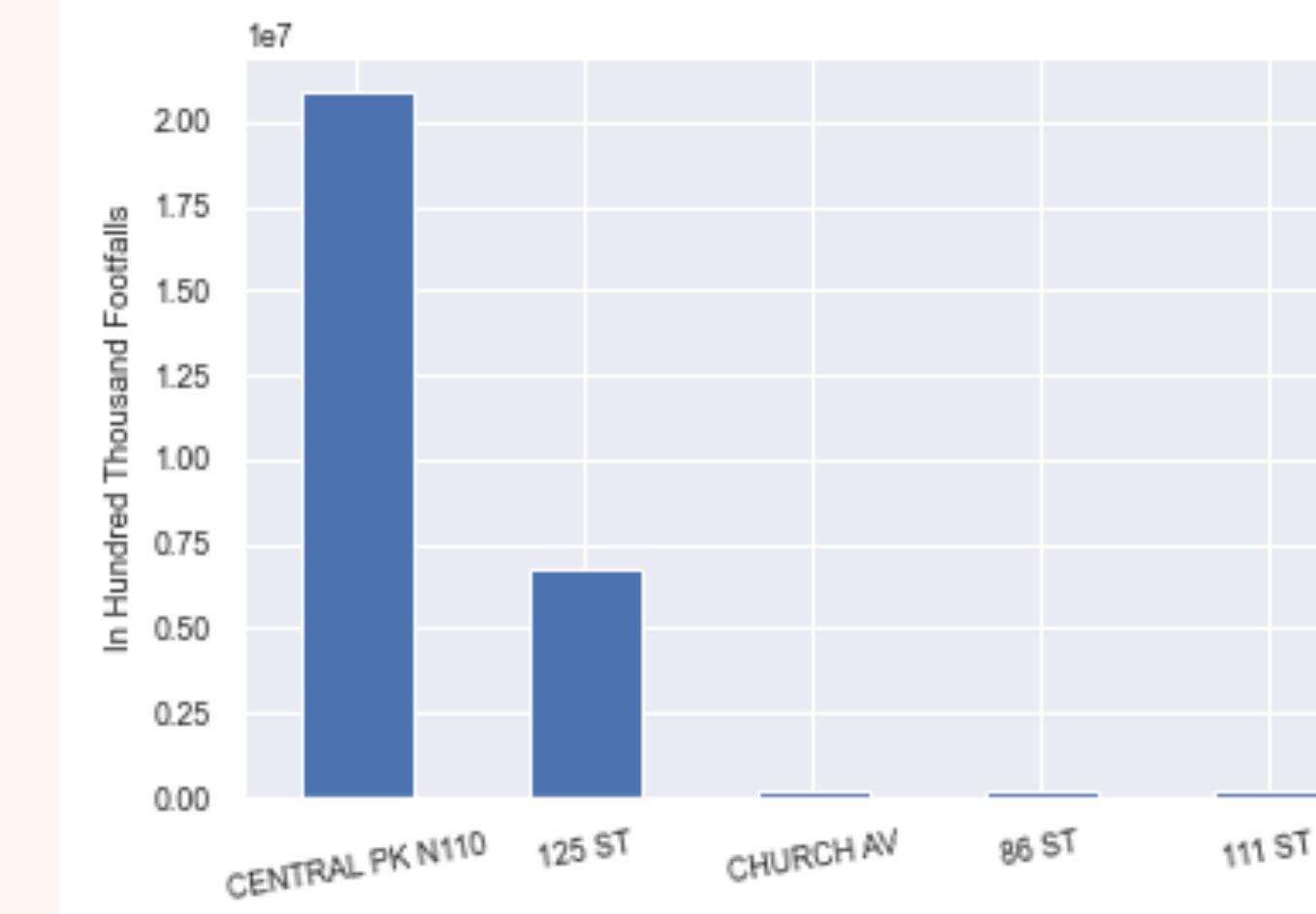


PRIME STATIONS BY TIME-SLOT

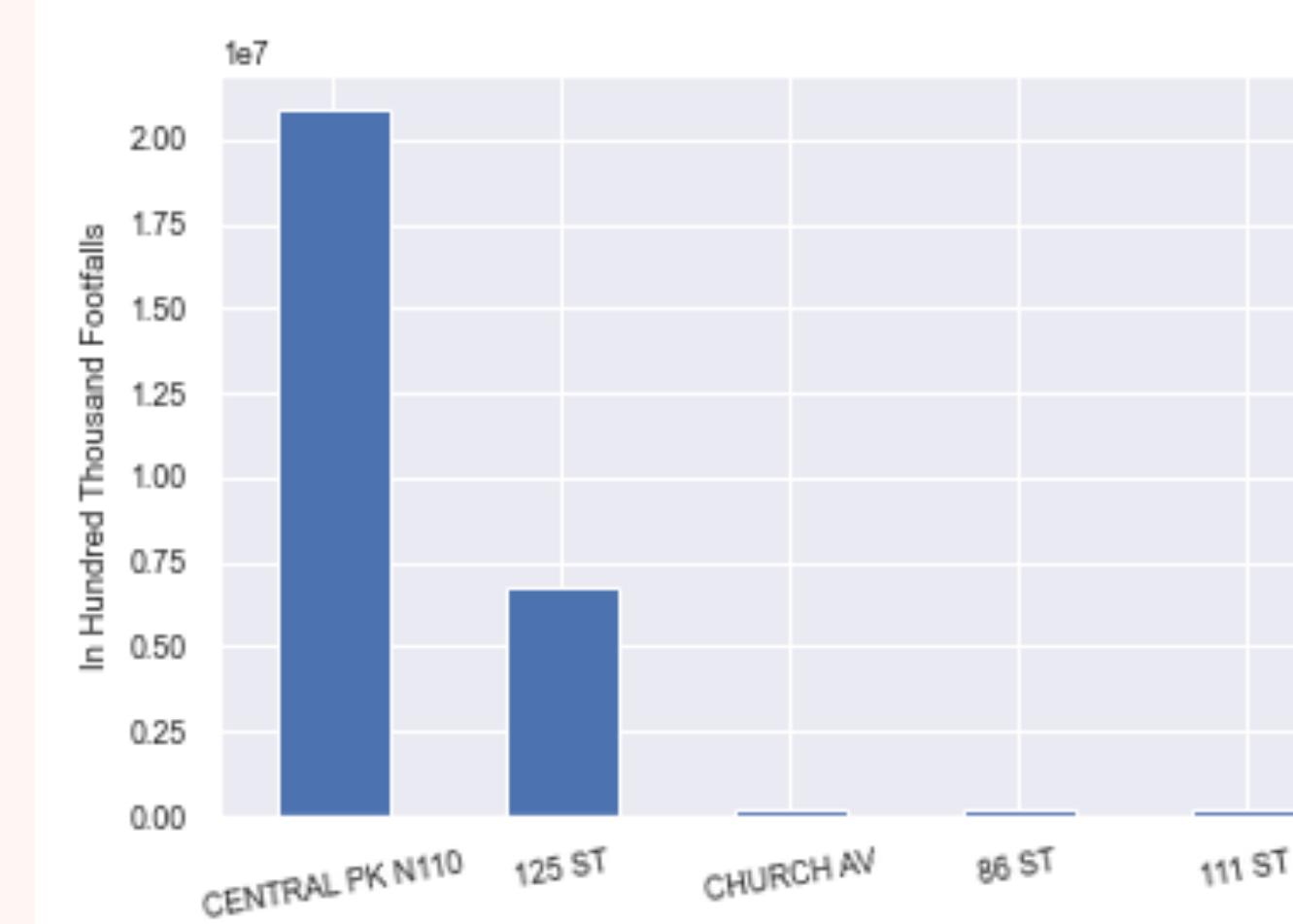
4AM
-
8AM



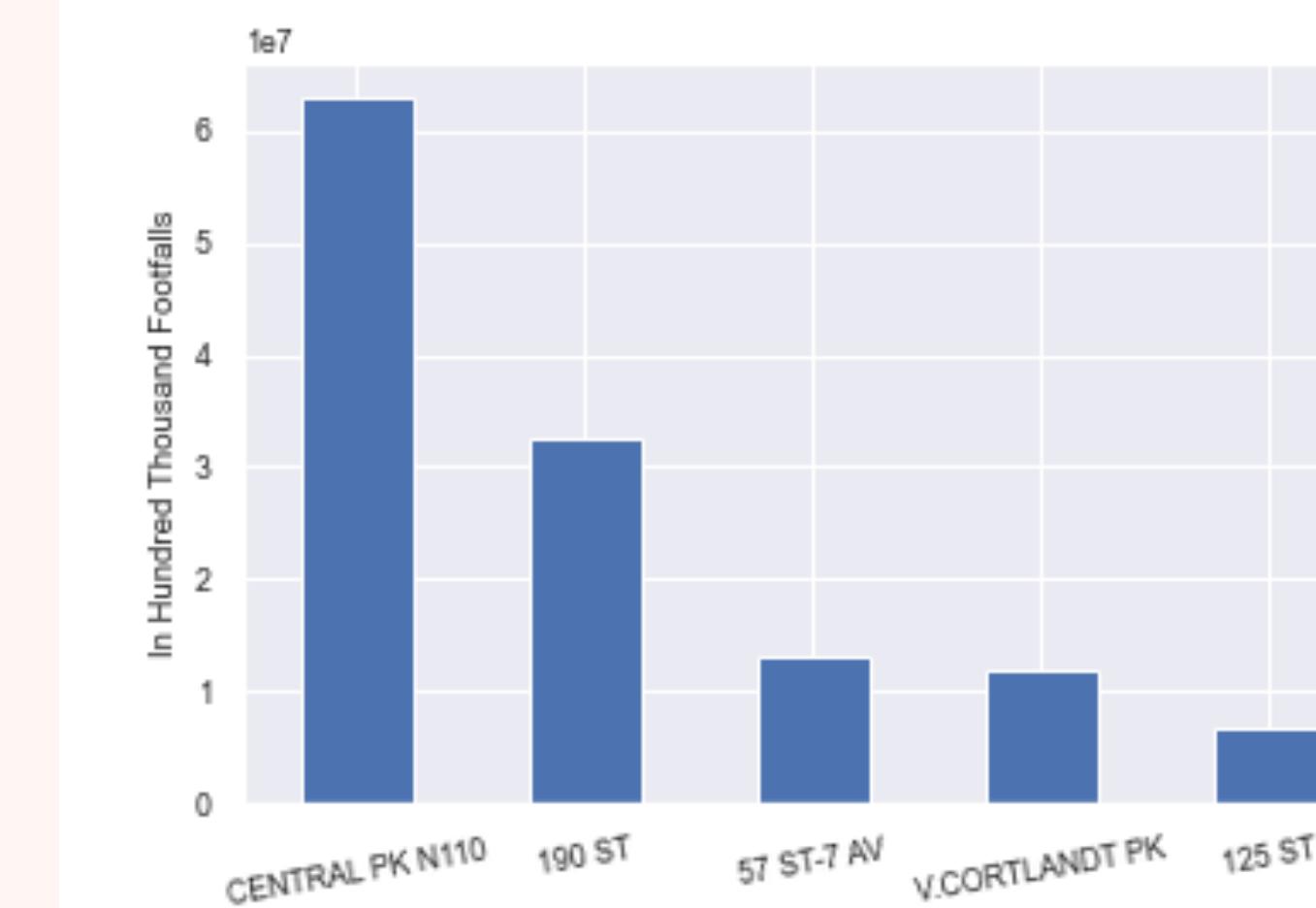
8AM
-
12PM



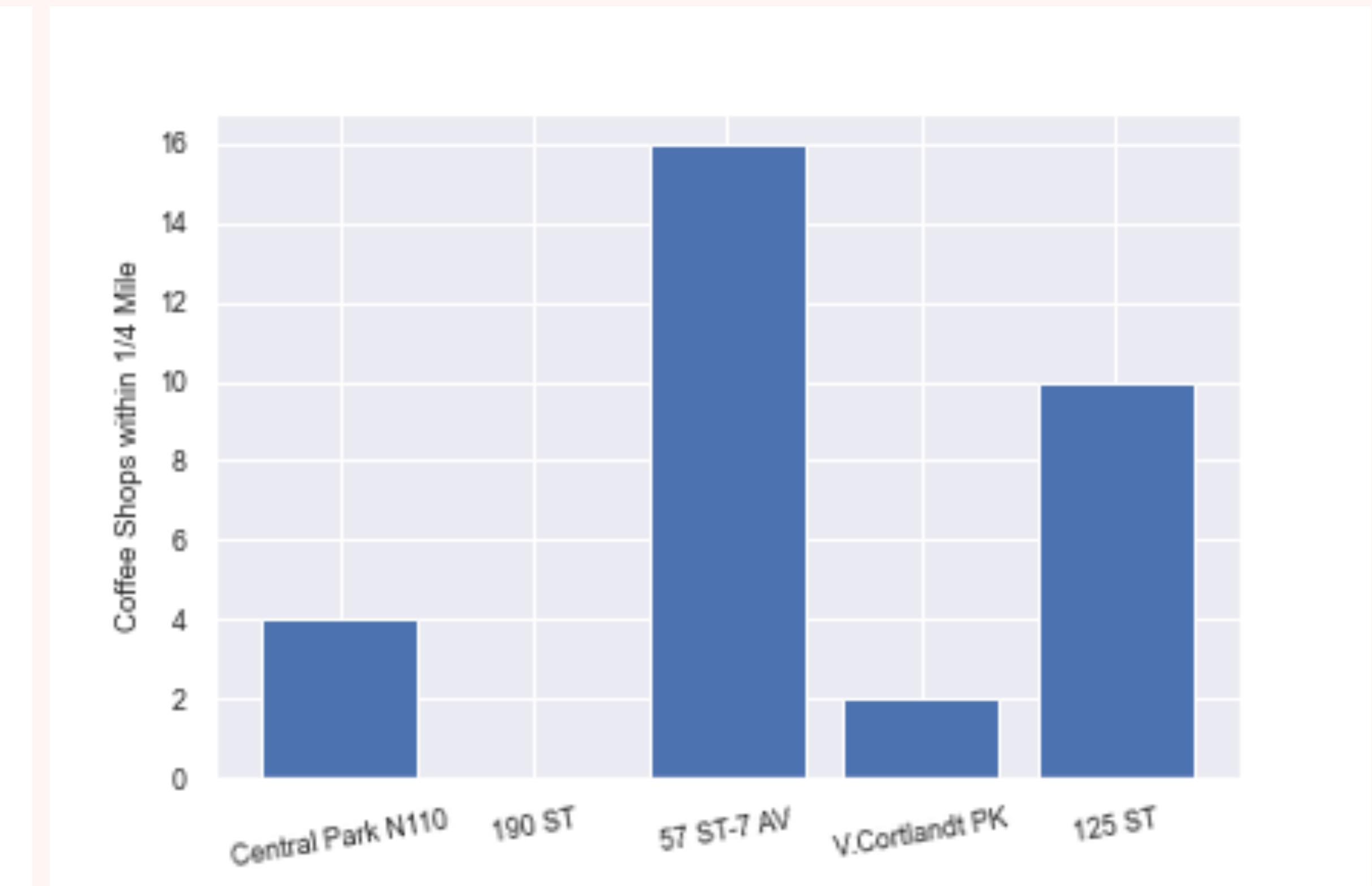
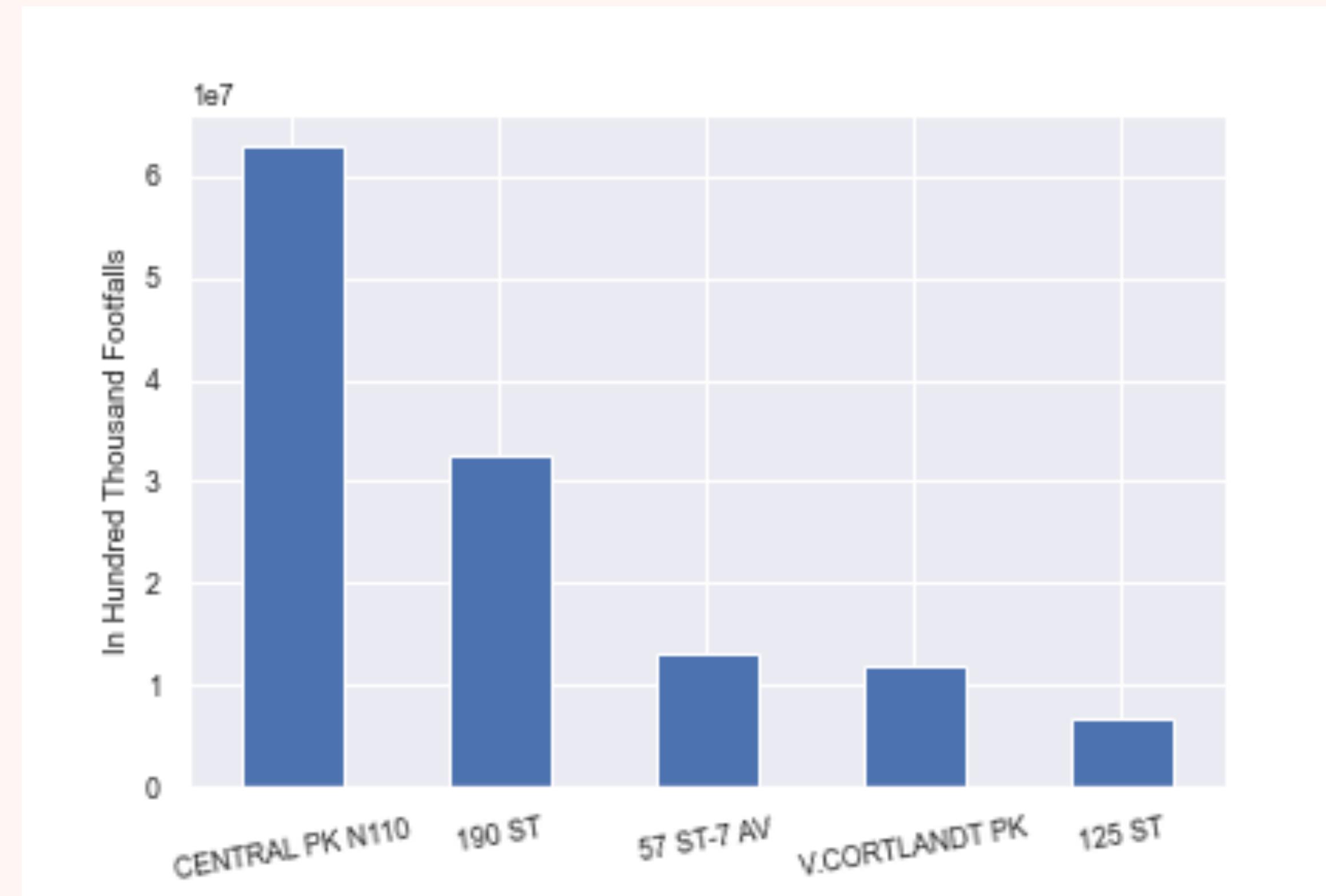
4PM
-
8PM



Total
Traffic



COFFEE SATURATION



RECOMMENDATION

BY NUMBER OF PROPOSED CARTS

- Given a single introductory cart
 - Central Park N110 ST is the only station appearing in the top 3 of all time slots, and the saturation of other coffee stations in the area is relatively low compared to other Prime stations.
 - Given multiple carts
 - Due to the heavy saturation of other coffee outlets within a quarter mile of the 3rd and 5th Prime candidates, and the complete lack of competition near the 2nd Prime, I would recommend 190 St Station first, Central Park N110 St second, and V Cortlandt Park third
-