

Exploratory Data Analysis

MTA Turnstile Data Analysis for a Food Advertising company

Apoorva Grampurohit

Introduction



- Advertisements in mass transit stations may be impactful
- The Problem : Targeted marketing by Food Ad agencies (Minorities, low-income)

Business Need



- Optimize the placement of “EatHealthy” Ads (both digital and poster)

Objective:

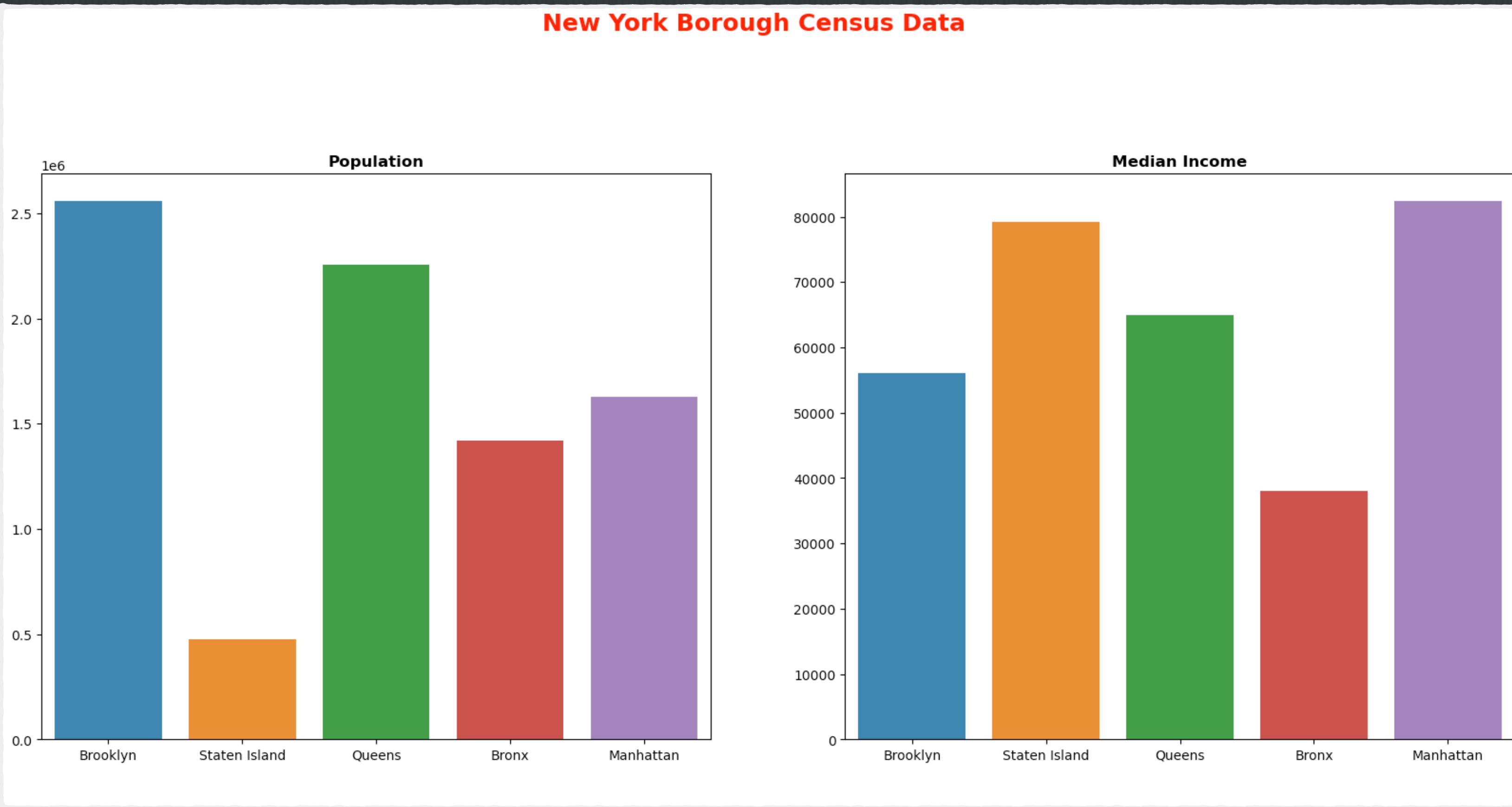
- Analyze New York City Census data to find the boroughs with Low median income**
- Use MTA Turnstile Data to find the top 10 Busiest stations in the selected boroughs.**
- Finding the busiest days for digital advertisement.**

Methodology:

- Data Collection : MTA turnstile data and NYC Census data (Pandas and SQLAlchemy)**
- Data Cleaning**
- Merged the two Datasets to analyze the two selected boroughs**
- Data Visualization using Matplotlib and Seaborn**

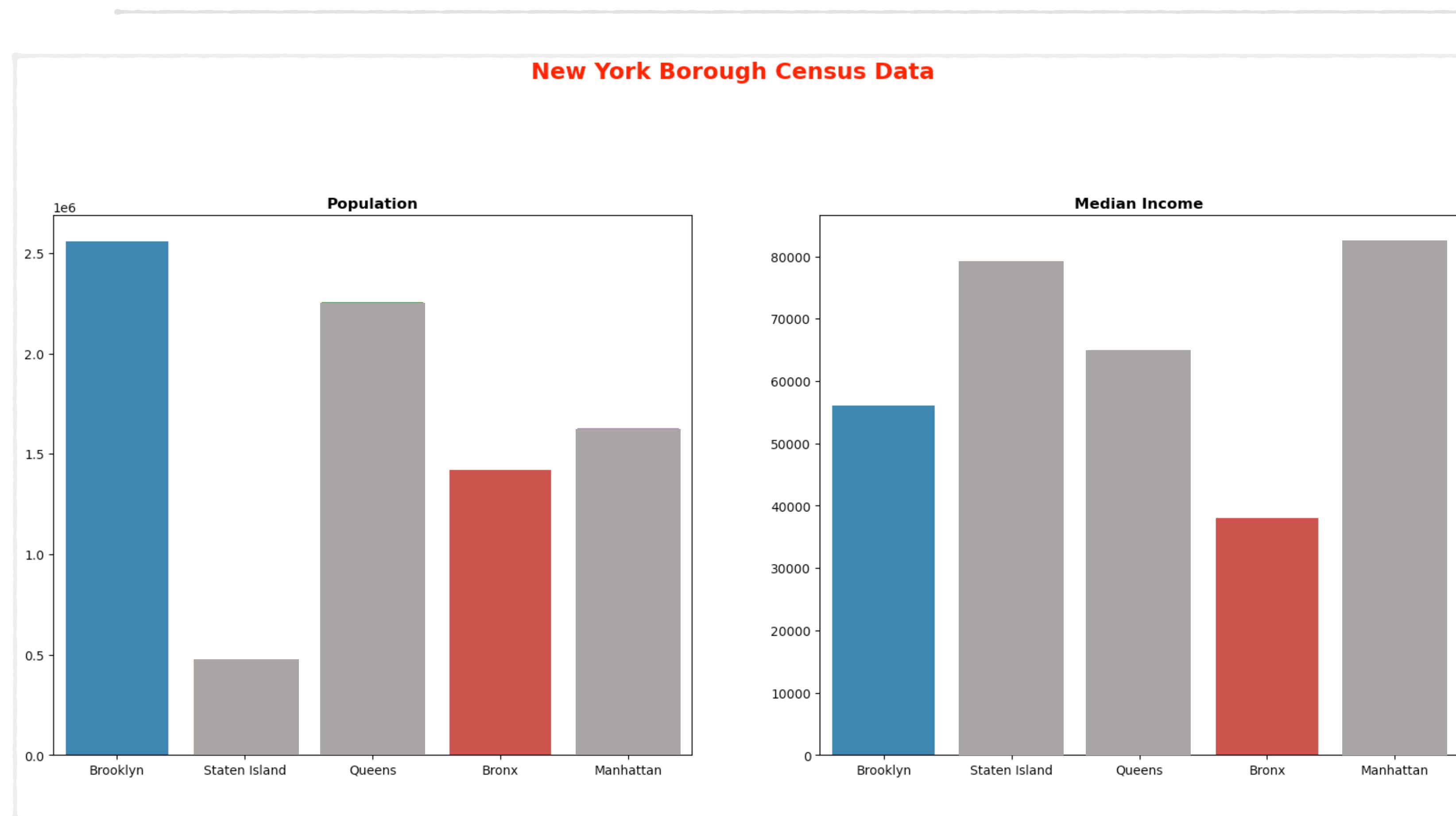
Census Data Findings:

New York Borough Census Data

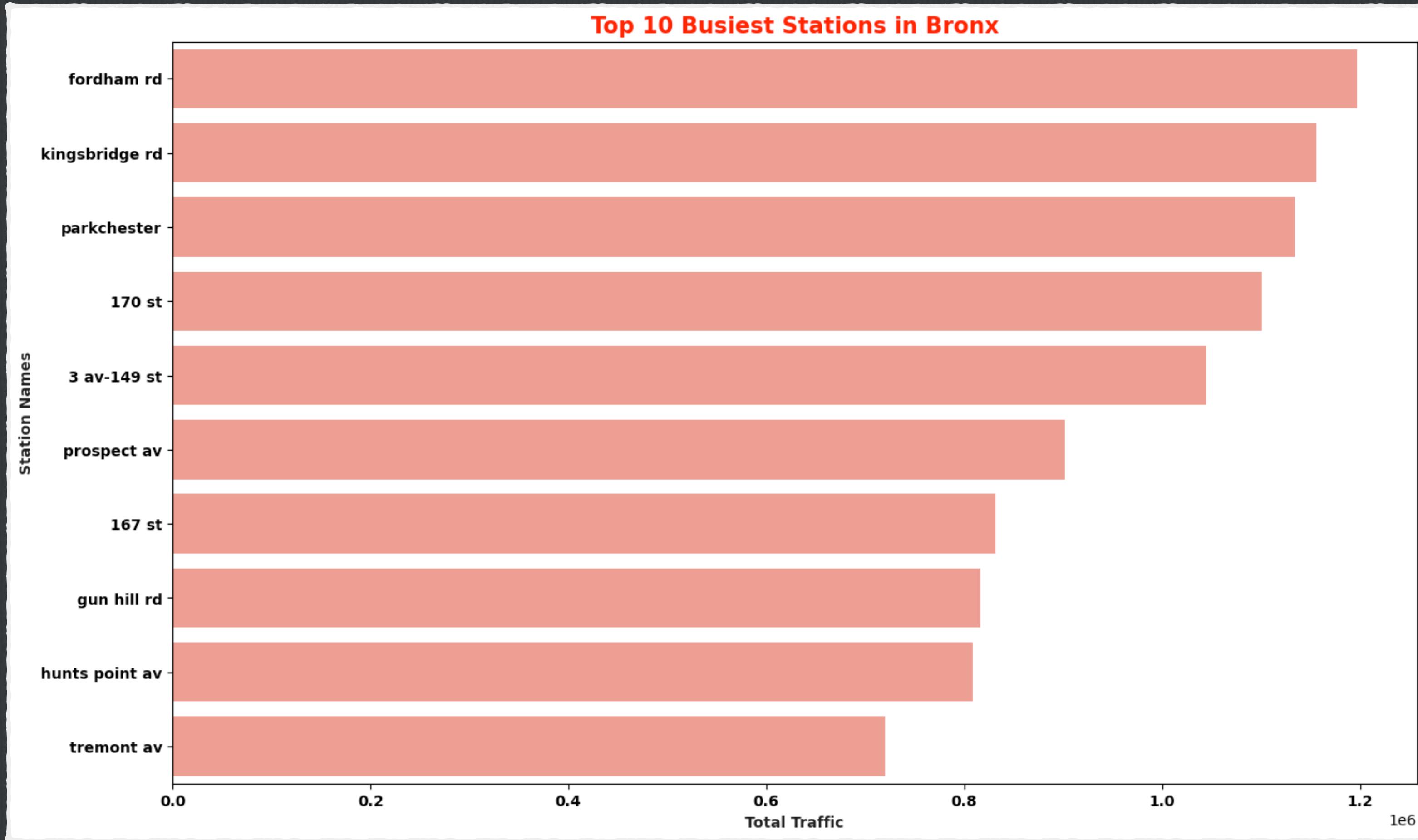


- Bronx and Brooklyn had the least Median income

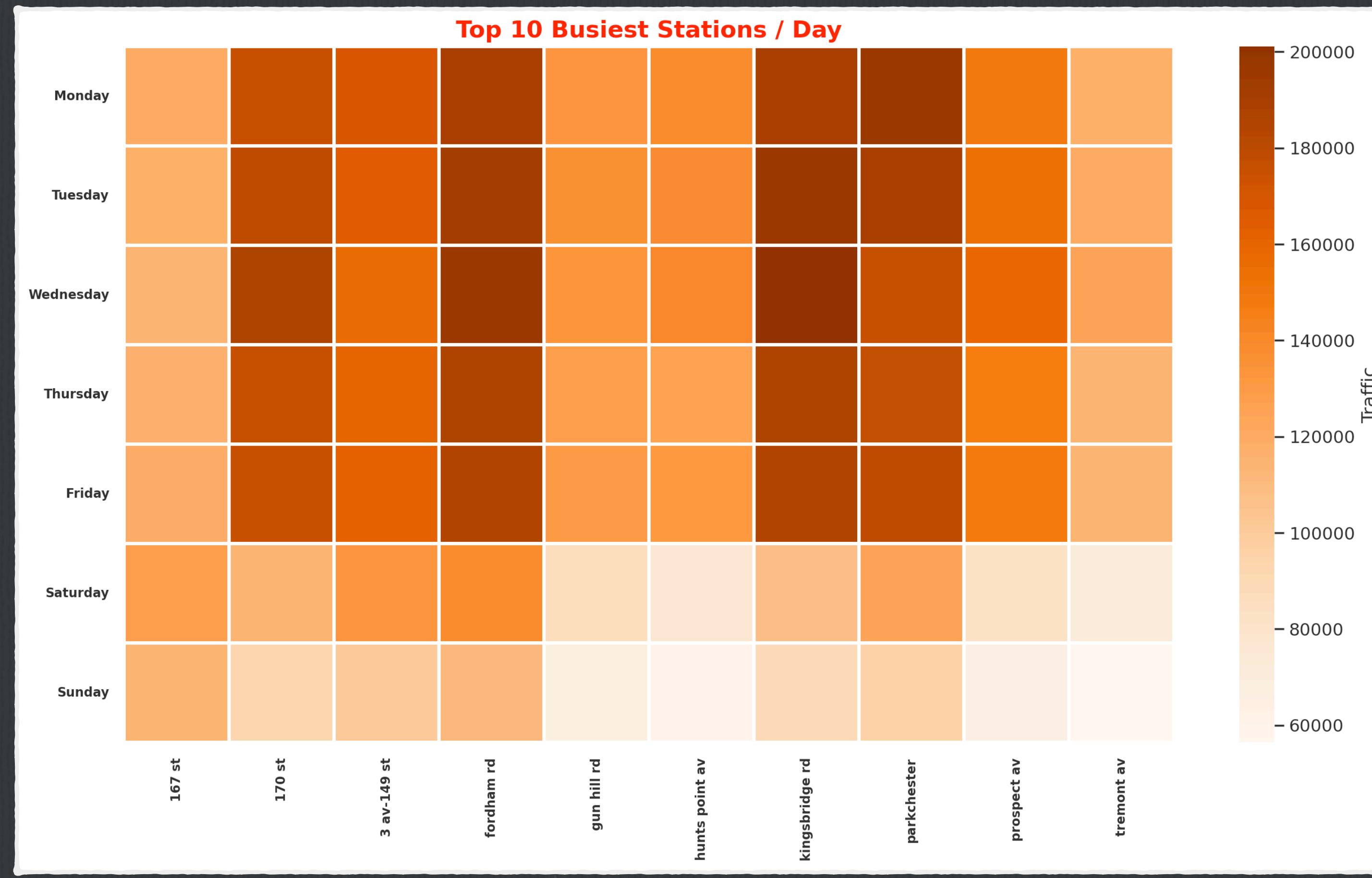
Census Data Findings:



MTA Turnstile Data Findings: Bronx

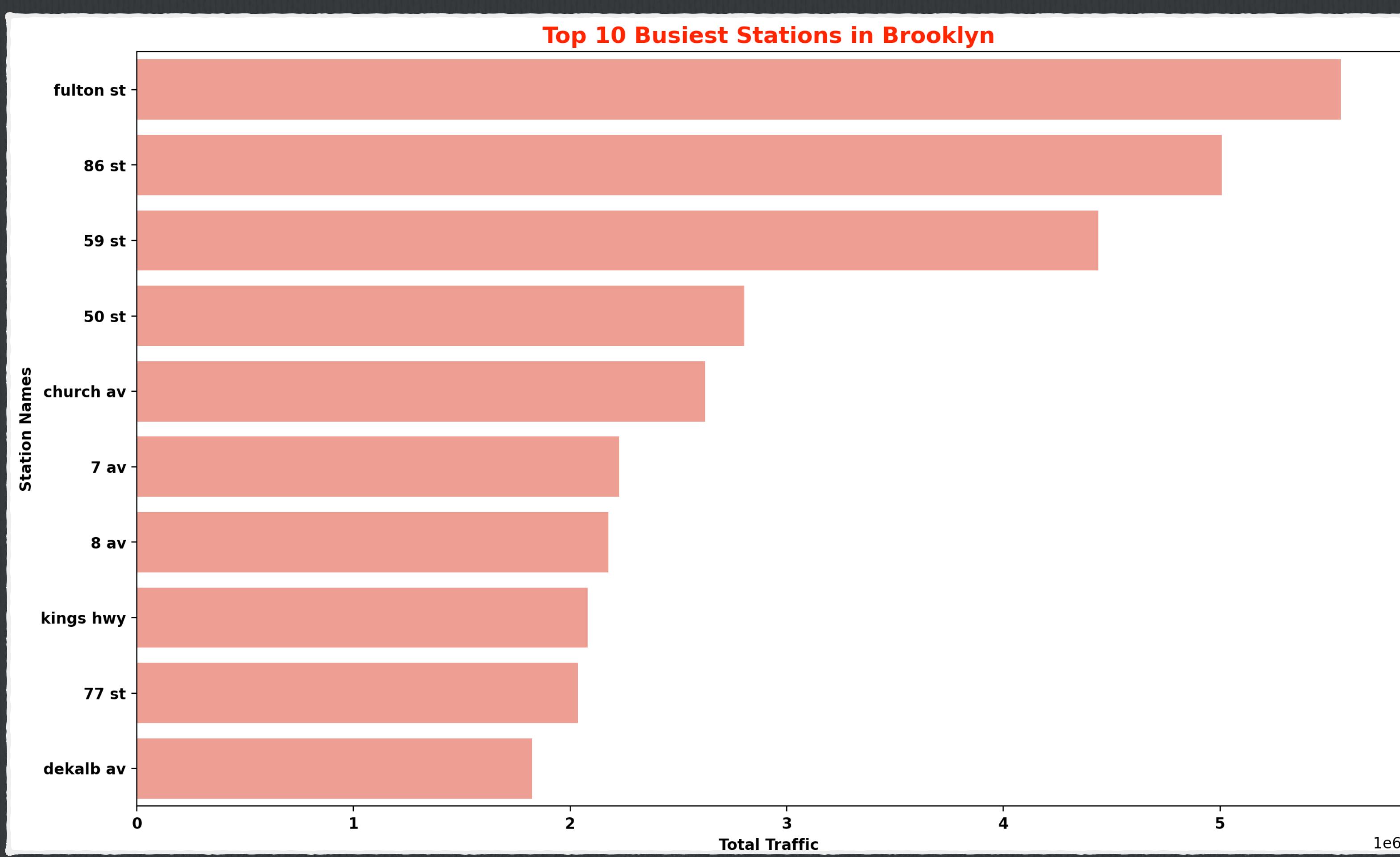


MTA Turnstile Data Findings: Bronx

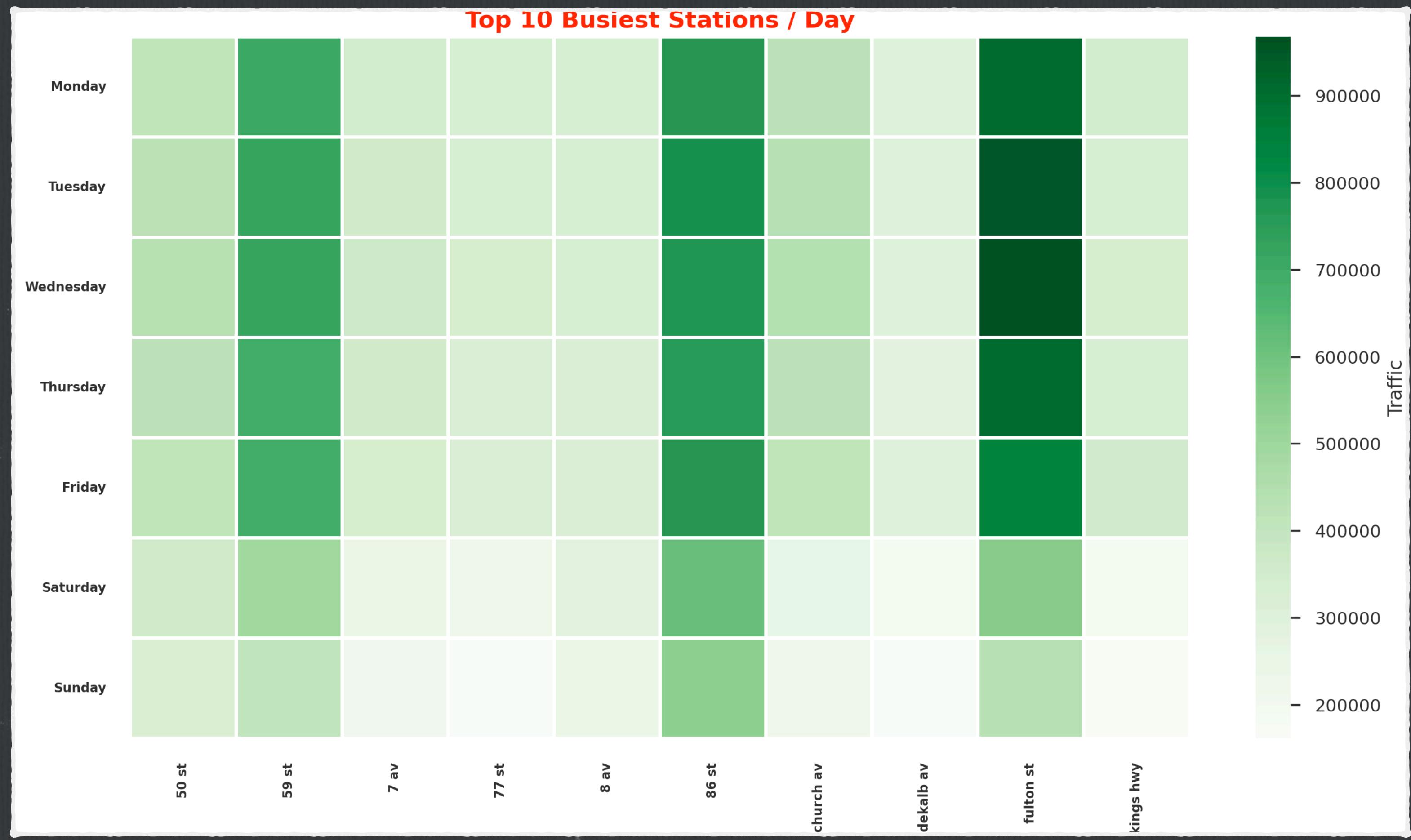


Heatmap of busiest days.

MTA Turnstile Data Findings: Brooklyn



MTA Turnstile Data Findings: Brooklyn



Heatmap of busiest days.

Conclusions and Further analysis:

- Bronx and Brooklyn boroughs can be chosen for targeted marketing.
- Digital advertising : Fordham rd (Bronx) and Fulton st (Brooklyn) - Monday-Friday
- Further analysis should include:
 - Analyzing Census data based on zip-code.