# Interborough Expresshorizontal line

**Group 2:  
Saachee Deshmukh**

**Tony Jose**

**Agaaz Oberoi**

**Jiaqi Guo**

**Parthkumar Kanani**

**05/14/2023**

Contents

**—**

1. **Overview**

*1.1. Abstract*

*1.2. Problem Statement*

*1.3. Proposed Solution*

1. **Technology and Tools**
2. **Dashboard**

*3.1. Description*

*3.2. Data analysis and processing methods*

*3.3. Component*

**4. Appendix**

## 1. OVERVIEW

## 1.1. Abstract

**—**

The New York City Metropolitan area is one of the largest cities in the world and relies heavily on its public transportation system to help millions of workers commute every day. This system has outgrown its initial design limitations, we have created visualizations to understand the commute demographics and data better to suggest solutions to the city’s most pressing transportation problems.

## 1.2. Scenario

**—**

New York City’s transportation network was built considering Manhattan at its core, allowing commuters to travel to Manhattan for work and commute back home after. For this reason, the network has been built to allow people to commute to, from, and in Manhattan most effectively. As the city has outgrown its urban core there is development in the outer Boroughs. Counties such as Kings County(Brooklyn) and Queens have grown at remarkable rates with other boroughs not far behind. As this explosive growth has taken place, more residents are commuting to work in Boroughs other than Manhattan. Without a public transportation network designed to cater to these commuters they are left with no choice but to travel through Manhattan in most cases to travel to other boroughs.

## 

## 1.3. Proposed Solution

**—**

To cater to the growing demand of interborough commuters trying to work in boroughs other than Manhattan we created visualizations to estimate commute patterns with the highest growing demand. After careful consideration and review of the data we arrived at a proposed solution for an Interborough Express and suggestions for the route alignment and high-volume stops to enable fast and accessible transportation for commuters allowing them to travel between the Boroughs, specifically Brooklyn, and Queens within 30 minutes compared to the existing travel time of 90 minutes.

## 2. Technology and Tools

**—**

Excel was used in the earliest stages of data cleaning because of its usability and user-friendly interface. The remaining data cleaning and processing activities in Python were made easier by this initial cleansing. Python was selected for thorough data cleaning and processing because of its strong libraries, adaptable handling of various data types, and capability of carrying out sophisticated calculations.

Specifically, we used Python for further data cleaning such as identifying and handling missing values, removing duplicates, and formatting data into pandas dataframe. Also, Python was used for data aggregation based on the coordinates for different boroughs. Besides, we used python for complex calculations such as the percentage of commuters traveling between boroughs. Python libraries such as folium, geopandas, matplotlib were used to create a Sankey map, pie chart, and geographic map of each borough.

## 3. Dashboard

**—**

**3.1. Description**

<https://colab.research.google.com/drive/1DirCLc3q9JVZEiEBtk_3s1LfmwUxGiJC?usp=sharing>

Made using Dash on python, the files and output are included in the attached Colab file. JupyterDash was used to render the dashboard components within the colab file itself, so it doesn’t need to be hosted online.

**Meeting the Needs of the Scenario:**

The dashboard helps the MTA understand the existing commuting patterns in the city. It provides detailed visualizations of commuting data, including pie charts showing the means of transportation used in different boroughs and a Sankey map showing commute flows within boroughs. The dashboard also includes a proposal for a new transit line connecting Eastern Queens to Eastern Brooklyn, which could reduce commute times by over half. This crucial information can inform decision-making, support transportation planning efforts, and drive initiatives to improve public transit accessibility and efficiency.

**3.2. Data Analysis and Processing Method**

**—**

**Overview of the Data Source, Audience, and Technology**

The data for this analysis was sourced from a shapefile containing geospatial data related to means of transportation to work in New York City. The intended audience for this dashboard includes urban planners, transportation authorities, policymakers, and researchers interested in urban transportation planning. The technology used includes Python for data cleaning, manipulation, and analysis, and libraries like GeoPandas, Matplotlib, and Folium for geospatial analysis and visualization.

*Shapefile: U.S. Department of Transportation. (n.d.). Means of Transportation to Work. Retrieved May 11, 2023, from* [*https://geodata.bts.gov/datasets/usdot::means-of-transportation-to-work/explore?location=19.863468%2C0.314300%2C3.31&showTable=true*](https://geodata.bts.gov/datasets/usdot::means-of-transportation-to-work/explore?location=19.863468%2C0.314300%2C3.31&showTable=true)

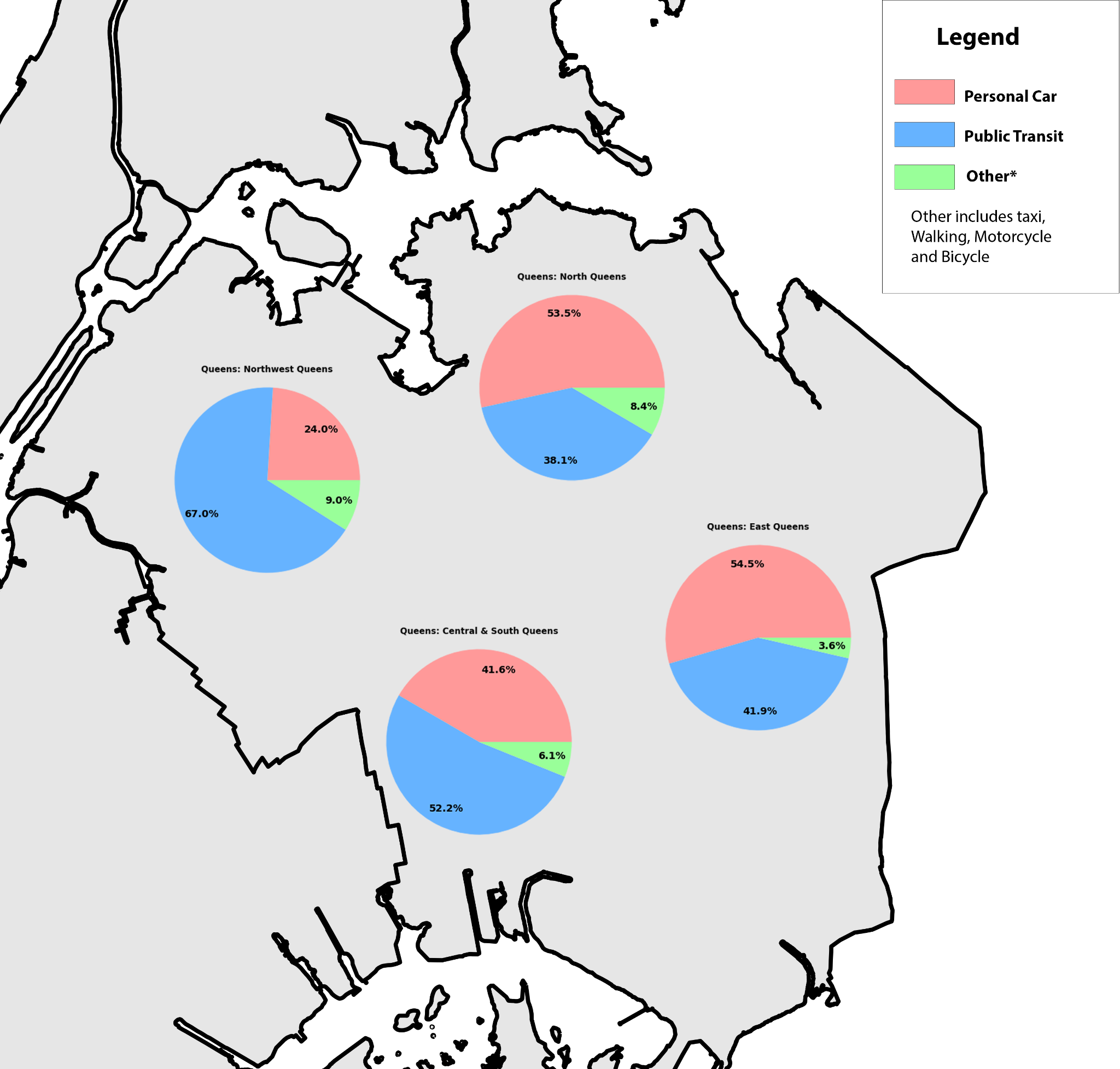
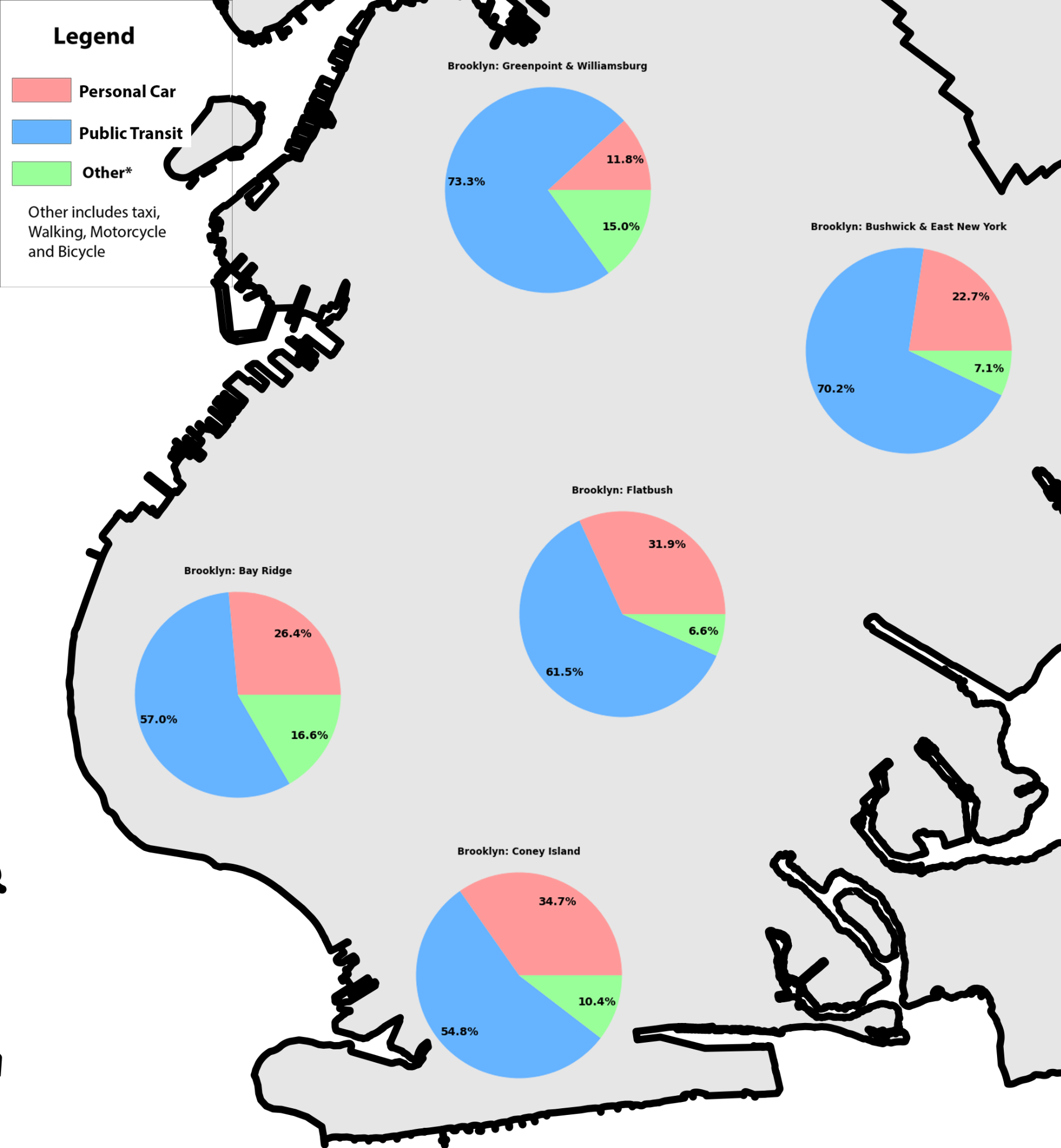
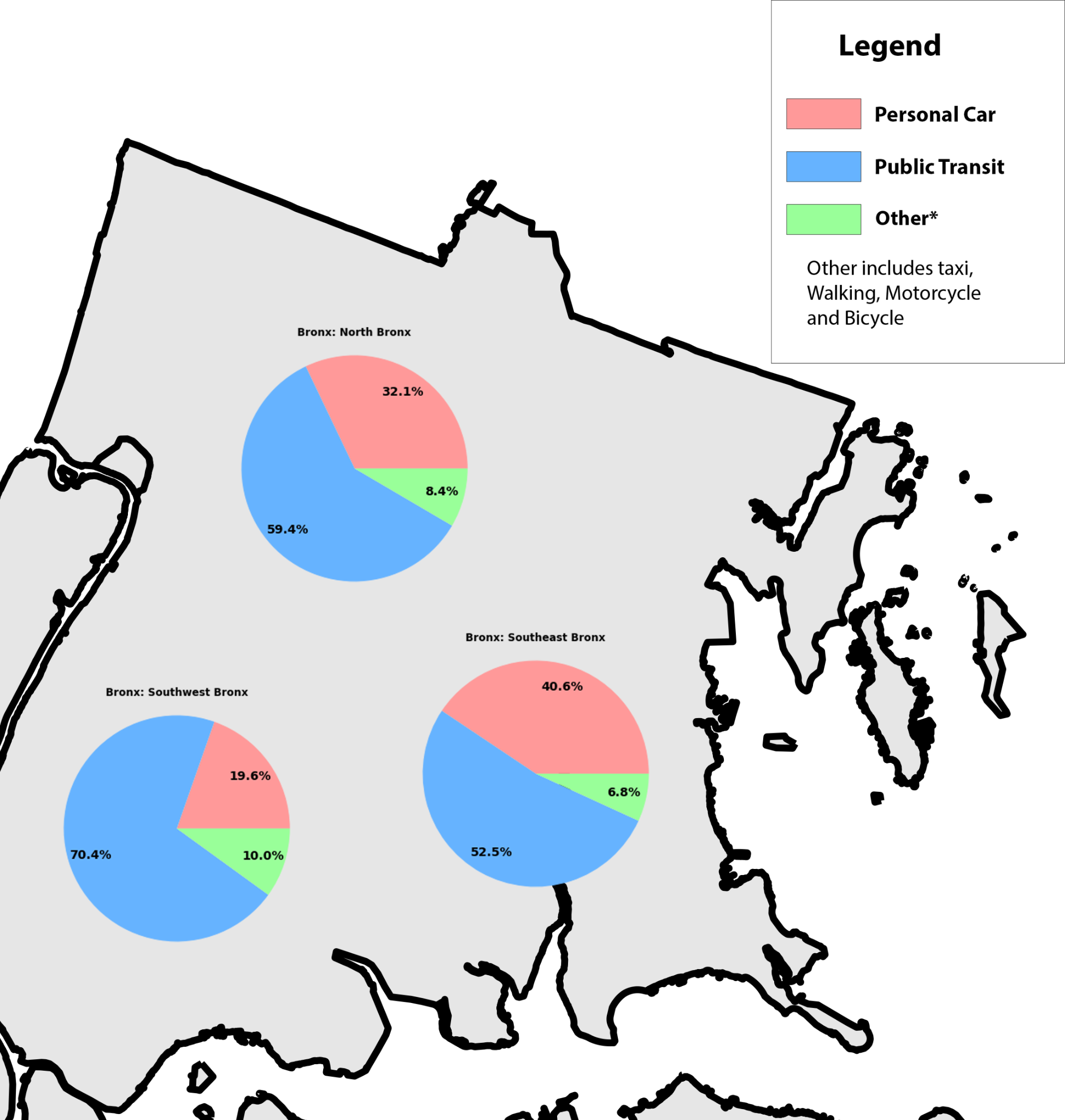
*Commute Flows: U.S. Census Bureau. (2015). Commuting flows. Retrieved May 11, 2023, from https://www.census.gov/data/tables/2015/demo/metro-micro/commuting-flows-2015.html*

**Data Analysis and Processing**

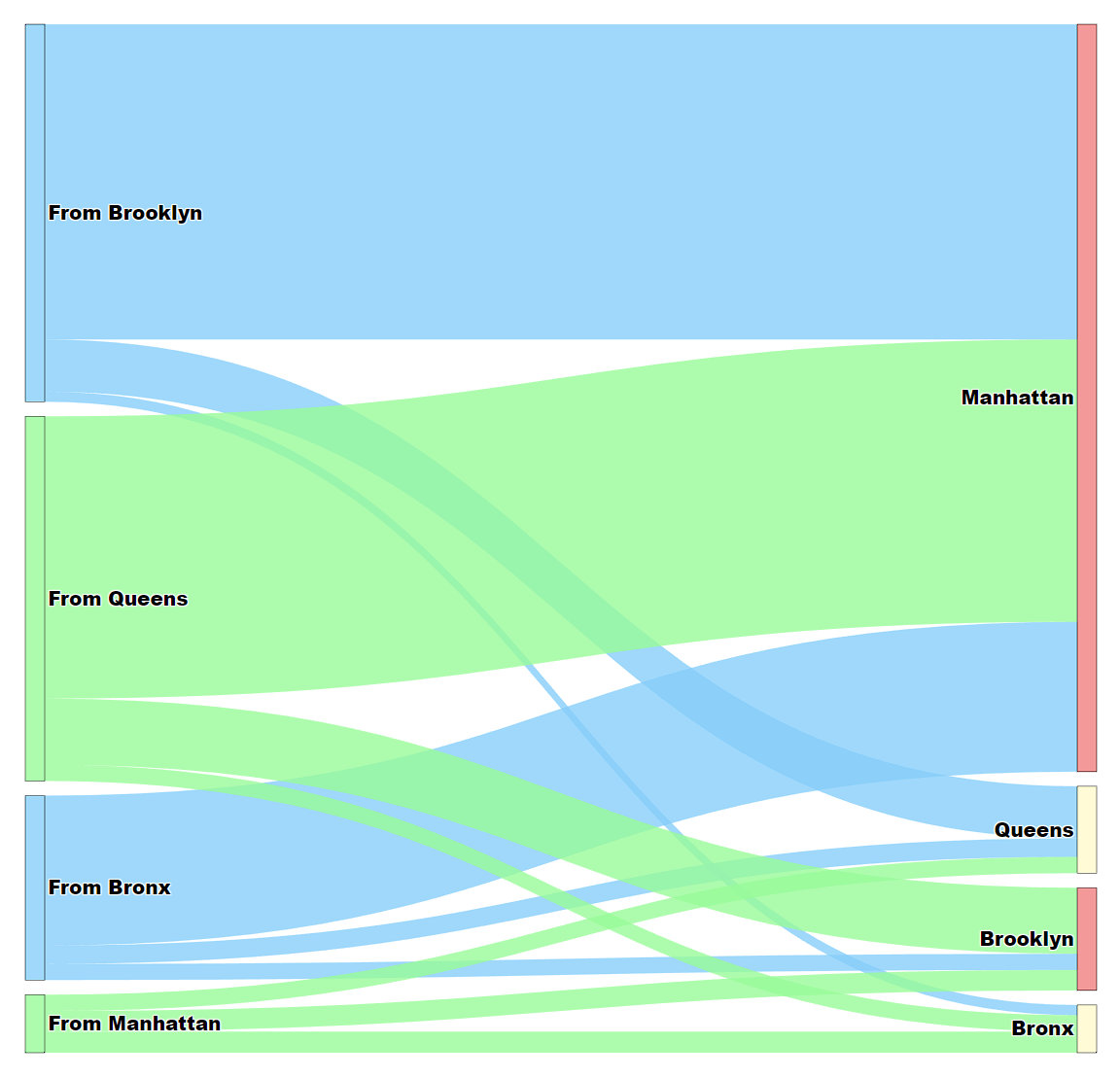
The data was extensively cleaned and processed to meet the needs of the analysis. This involved steps like filtering relevant data, renaming columns for clarity, combining data from several columns into a single column, and dropping unnecessary columns. The data was also divided into specific parts using KMeans clustering based on geographical coordinates. Further, commute patterns were visualized using pie charts and a Sankey map to represent the proportion of commuters using different means of transportation and the commute flows within boroughs. A proposal for a new transit line was also developed, showing the potential to significantly reduce commute times. These metrics, analyses, and visualizations were selected to provide a comprehensive understanding of commuting patterns in New York City and to propose actionable solutions to improve commute efficiency.

**3.3. Components**

**—**

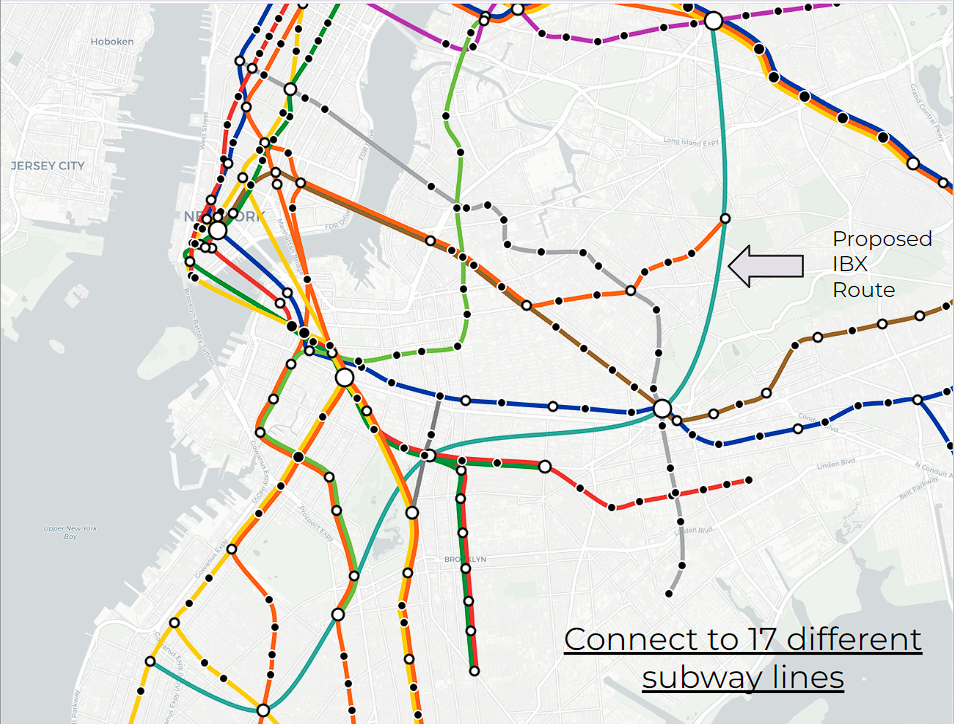
****

The above components show the different parts each borough was divided into using clustering (we used the shapefile data for this) and then further analyzed to aggregate the 3 different transit categories based on the groupings we found most appropriate. This was made by Saachee, Agaaz, Tony and Jiaqi as a collaborative effort. The python script was mostly taken care of by Agaaz and Jiaqi, while Tony and Saachee helped with the visualizations and data cleaning intricacies. The same effort was used for the Sankey map below, which used our Commuter Flows Dataset, showing the flow of workers between NYC boroughs on a daily basis:



The thickness of the flows from each node corresponds to the number of workers.

Finally, Tony, Sachee and Jiaqi used [this Github Page](https://jpwright.github.io/subway/) to create our suggested Interborough Express line for the MTA, based on our understanding of the data and its implications:



1. **Appendix**

| **Index** | **Date** | **Attendees** |
| --- | --- | --- |
| 1 | 4/20/23 at 2PM | Saachee, Jiaqi, Tony, Agaaz |
| 2 | 4/27/23 at 2PM | Saachee, Jiaqi, Tony, Agaaz, Parth |
| 3 | 5/4/23 at 4:30PM | Saachee, Jiaqi, Tony, Agaaz |
| 4 | 5/5/23 at 3PM | Saachee, Jiaqi, Tony, Agaaz |
| 5 | 5/7/23 at 6PM | Saachee, Jiaqi, Tony, Agaaz |
| 6 | 5/8/23 at 3PM | Saachee, Jiaqi, Tony, Agaaz |
| 7 | 5/9/23 at 5PM | Saachee, Jiaqi, Tony, Agaaz |
| 8 | 5/10/23 at 2PM | Saachee, Jiaqi, Tony, Agaaz |
| 9 | 5/14/23 at 1PM | Saachee, Jiaqi, Tony, Agaaz |