NYC Real-Time Traffic Speed Analysis
Overspeed Rate and Jam Rate Analysis

Zhengda Liu, Eva Song, Jingwei Zhang, Shengyao Ye, Tianle Qiu

2024 Fall STAT605 Group 2

1 Introduction

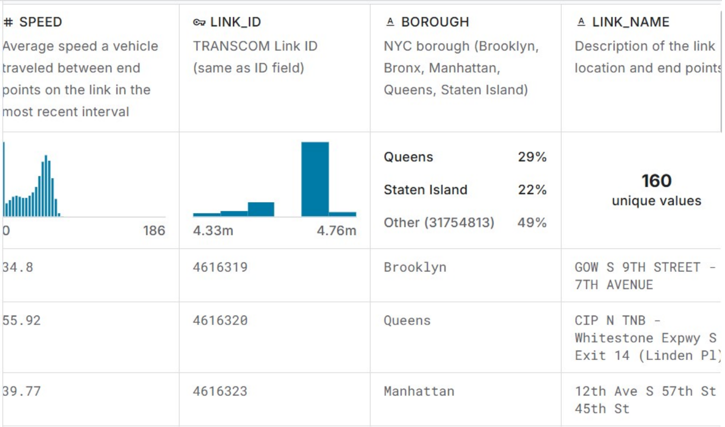
This report examines NYC real-time traffic data, focusing on overspeed and congestion rates. We calculated these rates and applied clustering to uncover traffic patterns. Persistent congestion on streets like 11th Avenue in Manhattan highlights shortcomings in traffic management. Results are presented through interactive visualizations, offering insights for urban traffic planning and policy improvements.

---

2 Body

2.1 Data Description

We used the NYC Real-Time Traffic Speed Data from Kaggle, a 28.36GB CSV file with 13 columns. Each record represents the average speed of vehicles traveling between endpoints on a link, defined as a sequence of latitude and longitude points.



# SPEED	LINK_ID	BOROUGH	LINK_NAME
Average speed a vehicle traveled between end points on the link in the most recent interval	TRANSCOM Link ID (same as ID field)	NYC borough (Brooklyn, Bronx, Manhattan, Queens, Staten Island)	Description of the link location and end points
0 186	4.33m 4.76m	Queens 29% Staten Island 22% Other (31754813) 49%	160 unique values
34.8	4616319	Brooklyn	GOW S 9TH STREET - 7TH AVENUE
55.92	4616320	Queens	CIP N TNB - Whitestone Expwy S Exit 14 (Linden Pl)
39.77	4616323	Manhattan	12th Ave S 57th St 45th St

Figure 1: Dataset Overview

Since the Kaggle dataset is not updated, we identified an updated version on NYC OpenData. While this version was not analyzed in this study, it is noted for future use.

2.2 Statistical Computation

All computations were performed using CHTC, as outlined below. The tasks include **A. computing summary statistics** (overspeed rate and jam rate) for each link, and **B. splitting data by months** for map visualization.

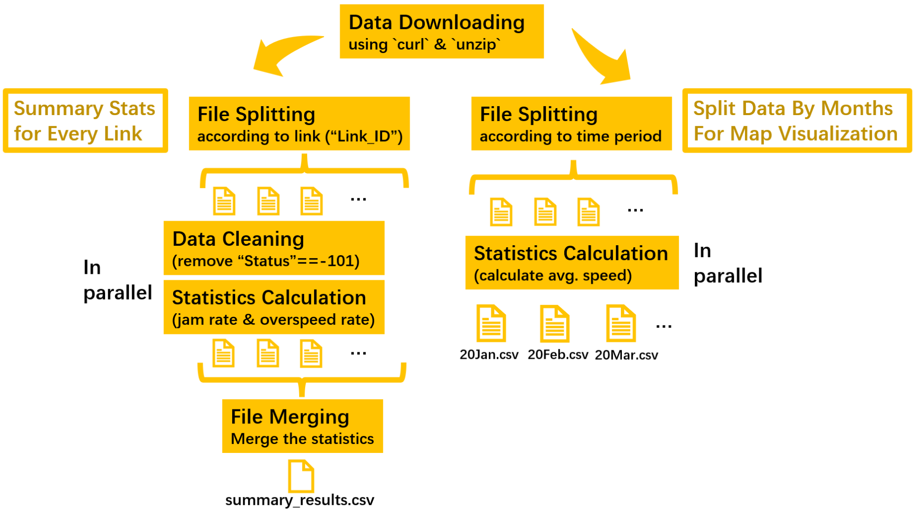


Figure 2: Computation Flowchart for Tasks A and B

2.2.1 Data Acquisition

Kaggle data was downloaded using `curl` and decompressed with `unzip`. After acquisition, two DAG files were executed to process tasks A and B.

2.2.2 Data Splitting and Cleaning

Data was split by key fields: `LinkID` for task A and month for task B. Records with `status` = -101, indicating zero speed, were removed as they distort summary statistics.

2.2.3 Computation

For task A, overspeed rate and jam rate were computed as:

$$OverspeedRate = \frac{\sum_i I(s_i > 60)}{m}, \quad JamRate = \frac{\sum_i I(s_i < 10)}{m},$$

where s_i is the speed of the i-th record, and m is the total number of records. The results were merged into a single CSV for visualization.

For task B, average speeds for each link per minute were computed. The outputs included multiple files, each representing link speeds for a specific month.

2.3 Result

All the codes and results can be seen in our repository. Below is the command to clone it:

Listing 1: Git Clone Command

```
git clone https://github.com/JustinDs0205/STAT605_Final.git
```

2.3.1 Graphical summaries

We developed interactive tools to visualize traffic patterns:

Overspeed and Jam Rate Visualization

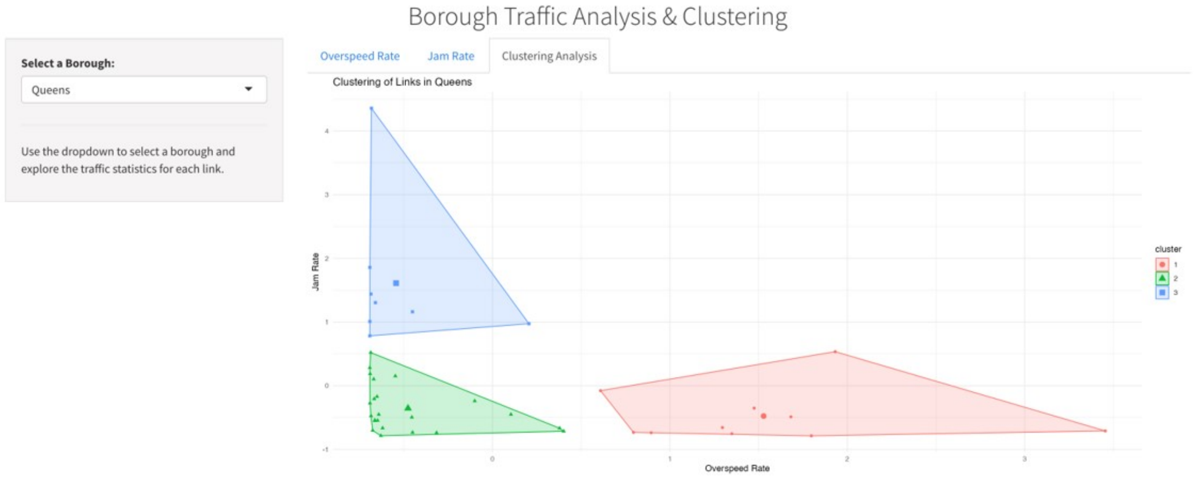
View the visualization here.

Interactive Traffic Map

Users can select streets and date ranges to visualize traffic speeds using an interactive map powered by Leaflet. Explore the map here.

Cluster Analysis

Overspeed rates and jam rates for clustering road sections with similar traffic patterns. This helps authorities identify and manage roads with comparable issues more effectively.



Borough Traffic Analysis & Clustering

Select a Borough: Queens

Use the dropdown to select a borough and explore the traffic statistics for each link.

Figure 3: Cluster Analysis Result

Table 1: Cluster Meansfor Overspeed Rate and Jam Rate

Cluster	Overspeed Rate	Jam Rate
1	0.099275982	0.01440353
2	0.009634136	0.01927453
3	0.006628807	0.09819441

Cluster 1 consists of road segments with high overspeed rates and low congestion, requiring enhanced traffic monitoring to address speeding. Cluster 2 includes stable road segments with minimal speeding or congestion issues. Cluster 3 comprises road segments with high congestion and low overspeed rates, suggesting the need for road optimization to reduce congestion.

2.3.2 Statistical computation

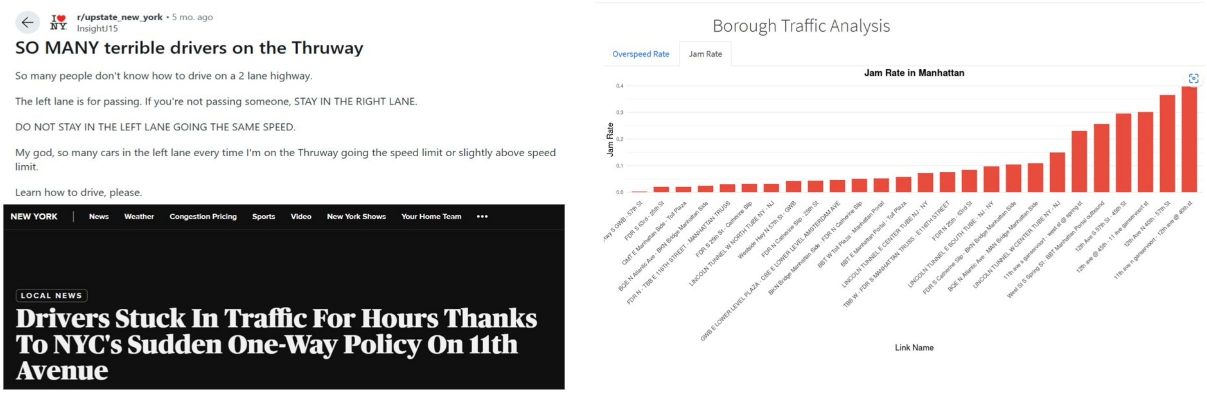
The number of jobs, the typical job time, memory, and disk space required are described as follows.

Table 2: Job Resource Usage

Job	Time	Memory	Disk
getKdata.sub	1 hour 2 mins	1GB	40GB
A_pre.sub	1 hour 17 mins	1GB	40GB
A_para.sub	58 mins	1GB	40GB
B_para.sub	1 hour 35 mins	1GB	40GB

2.3.3 News Validation

Streets with high congestion rates in Manhattan, like 11th Avenue, reveal issues in policy implementation and enforcement. A Reddit post and a CBS News report highlight ongoing traffic challenges, with drivers ignoring lane discipline. Our visualization confirms that 11th Avenue has the highest congestion rate, underscoring the need for improved traffic management and stricter policy enforcement.



r/upstate_new_york • 5 mo. ago
Insight/15

SO MANY terrible drivers on the Thruway

So many people don't know how to drive on a 2 lane highway.

The left lane is for passing. If you're not passing someone, STAY IN THE RIGHT LANE.

DO NOT STAY IN THE LEFT LANE GOING THE SAME SPEED.

My god, so many cars in the left lane every time I'm on the Thruway going the speed limit or slightly above speed limit.

Learn how to drive, please.

NEW YORK | News Weather Congestion Pricing Sports Video New York Shows Your Home Team •••

LOCAL NEWS

Drivers Stuck In Traffic For Hours Thanks To NYC's Sudden One-Way Policy On 11th Avenue

Borough Traffic Analysis
Overspeed Rate
Jam Rate
Jam Rate in Manhattan
Jam Rate
0.0
0.1
0.2
0.3
0.4
Hwy S GWB - 57th St
FDR S 63rd - 25th St
QMT E Manhattan Side - Toll Plaza
BQE N Atlantic Ave - BKN Bridge Manhattan Side
FDR N - TBB E 116TH STREET - MANHATTAN TRUSS
FDR S 25th St - Catherine Slip
LINCOLN TUNNEL W NORTH TUBE NY - NJ
Westside Hwy N 57th St - GWB
FDR N Catherine Slip - 25th St
GWB E LOWER LEVEL PLAZA - CBE E LOWER LEVEL AMSTERDAM AVE
BKN Bridge Manhattan Side - FDR N Catherine Slip
BBT W Toll Plaza - Manhattan Portal
BBT E Manhattan Portal - Toll Plaza
LINCOLN TUNNEL E CENTER TUBE NJ - NY
TBB W - FDR S MANHATTAN TRUSS - E116TH STREET
FDR N 25th - 63rd St
LINCOLN TUNNEL E SOUTH TUBE - NJ - NY
FDR S Catherine Slip - BKN Bridge Manhattan Side
BQE N Atlantic Ave - MAN Bridge Manhattan Side
LINCOLN TUNNEL W CENTER TUBE NY - NJ
11th ave s ganservoort - west st @ spring st
West St S Spring St - BBT Manhattan Portal outbound
12th Ave S 57th St - 45th St
12th ave @ 45th - 11 ave ganservoort st
12th Ave N 40th - 57th St
11th ave n ganservoort - 12th ave @ 40th st
Link Name

Figure 4: Validation for news and plots

2.3.4 Weaknesses

During data cleaning, we did not remove data related to special occasions, such as farmers markets or festival parades. Users can only know the overall historical congestion level of a road, but cannot get real-time information.

3 Conclusion

This report analyzes NYC traffic speed data, focusing on overspeed and congestion rates across streets. High congestion on streets like 11th Avenue in Manhattan highlights inefficiencies in traffic management and policy enforcement.

Clustering analysis identified patterns among streets with similar traffic characteristics, providing valuable insights into traffic behavior. The findings suggest that targeted measures, such as enhanced lane management and stricter enforcement, could mitigate congestion and overspeeding.

Future research could integrate additional factors, such as weather, to achieve a more comprehensive analysis. Predictive models may also enable real-time traffic monitoring using NYC OpenData.

4 Contributions

Member	Proposal	Coding	Presentation	Report	PPT
Zhengda Liu	1	1	1	0.6	0.4
Eva Song	0.6	1	1	1	0.2
Jingwei Zhang	1	0	1	0	1
Shengyao Ye	0	0.2	1	0	1
Tianle Qiu	0	0	1	1	0

Table 3: Group Member Contribution Chart

- In the chart above, 1 = full contribution, 0.1–0.9 = partial contribution, 0 = no contribution.