NYC Mobility Trends

DATA 400

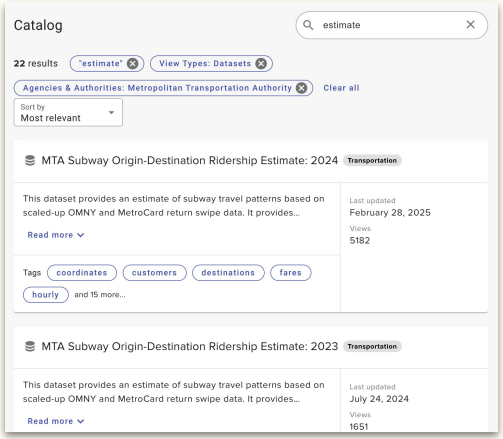
Ashley Kim, John Park

Mar. 25th, 2025

Tractable Data

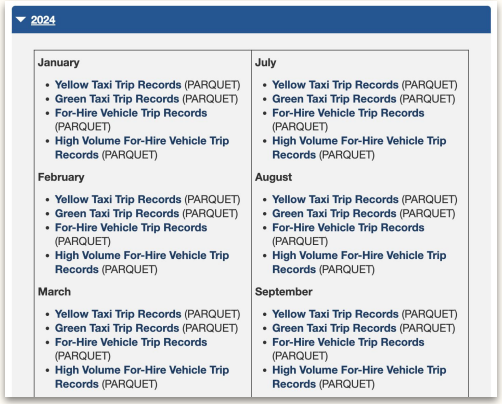
New York State Open Data

- open data portal, offering access to datasets from government agencies on topics like transportation, healthcare, education, and public safety



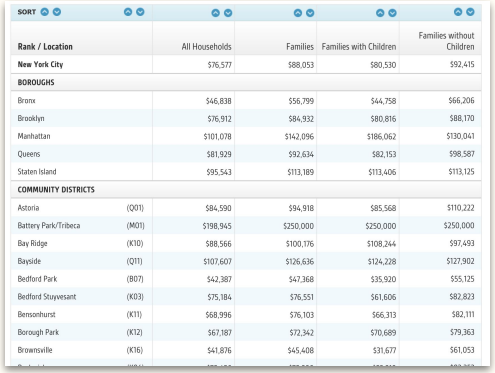
NYC Taxi & Limousine Commission

- regulates and oversees the city's yellow cabs, green cabs, and app-based ride services, providing data on ridership, licensing, and vehicle performance



Citizens' Committee for Children of NY

- nonprofit organization that provides research, data, and resources on child welfare, education, and family support



Rank / Location		All Households	Families	Families with Children	Families without Children
New York City		$76,577	$88,053	$80,530	$92,415
BOROUGHS					
Bronx		$46,838	$56,799	$44,758	$66,206
Brooklyn		$76,912	$84,932	$80,816	$88,170
Manhattan		$101,078	$142,096	$186,062	$130,041
Queens		$81,929	$92,634	$82,153	$98,587
Staten Island		$95,543	$113,189	$113,406	$113,125
COMMUNITY DISTRICTS					
Astoria	(Q01)	$84,590	$94,918	$85,568	$110,222
Battery Park/Tribeca	(M01)	$198,945	$250,000	$250,000	$250,000
Bay Ridge	(K10)	$88,566	$100,176	$108,244	$97,493
Bayside	(Q11)	$107,607	$126,636	$124,228	$127,902
Bedford Park	(B07)	$42,387	$47,368	$35,920	$55,125
Bedford Stuyvesant	(K03)	$75,184	$76,551	$61,606	$82,823
Bensonhurst	(K11)	$68,996	$76,103	$66,313	$82,111
Borough Park	(K12)	$67,187	$72,342	$70,689	$79,363
Brownsville	(K16)	$41,876	$45,408	$31,677	$61,053

Data Retrieval

- **Downloaded CSV files from NYC Open Data:**
  - MTA Subway Origin-Destination Ridership Estimate
    - Year: 2021 ~ 2024

- **Downloaded parquet files from NYC TLC :**
  - Yellow Taxi Trip Records (PARQUET)
  - Green Taxi Trip Records (PARQUET)
  - For-Hire Vehicle Trip Records (PARQUET)
  - High Volume For-Hire Vehicle Trip Records (PARQUET)
    - Year: 2021 ~ 2024 (all months)

- **Downloaded CSV file from CCC of NY :**
  - Median Incomes

Data Cleaning & Preprocessing

Why Night-time?
- Reduced subway service

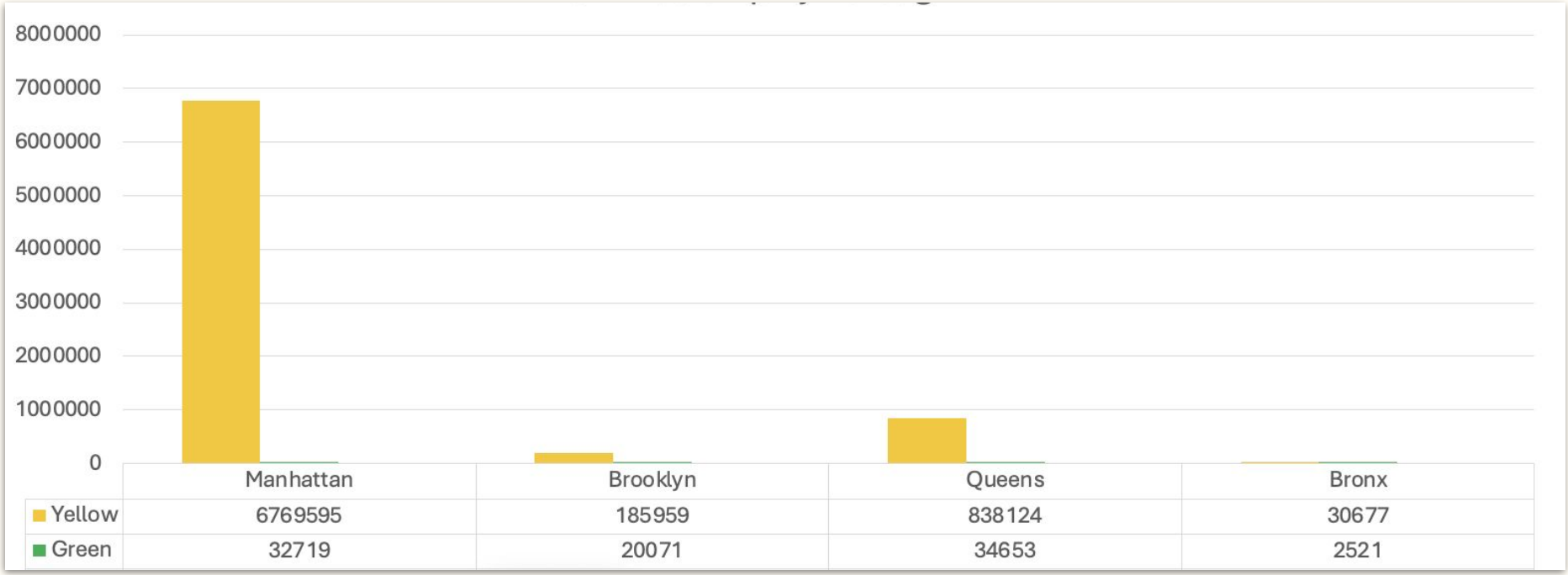
Taxi Datasets

- Key Identifier: LocationID
- Taxi_df = Merged yellow, taxi, fhvhv datasets vertically
  - Dropped day-time riderships
  - 10pm-6am
- Left joined taxi_df with taxizone_df on LocationID

```
['VendorID', 'pickup_datetime', 'dropoff_datetime', 'passenger_count',
 'trip_distance', 'RatecodeID', 'store_and_fwd_flag', 'PULocationID',
 'DOLocationID', 'payment_type', 'fare_amount', 'extra', 'mta_tax',
 'tip_amount', 'tolls_amount', 'improvement_surcharge', 'total_amount',
 'congestion_surcharge', 'Airport_fee', 'PU_Borough', 'PU_Zone',
 'PU_service_zone', 'DO_Borough', 'DO_Zone', 'DO_service_zone',
 'PU_Year', 'PU_Month', 'PU_Day', 'PU_Hour', 'PU_DayOfWeek', 'DO_Year',
 'DO_Month', 'DO_Day', 'DO_Hour', 'DO_DayOfWeek', 'Taxi_Tier'],
```

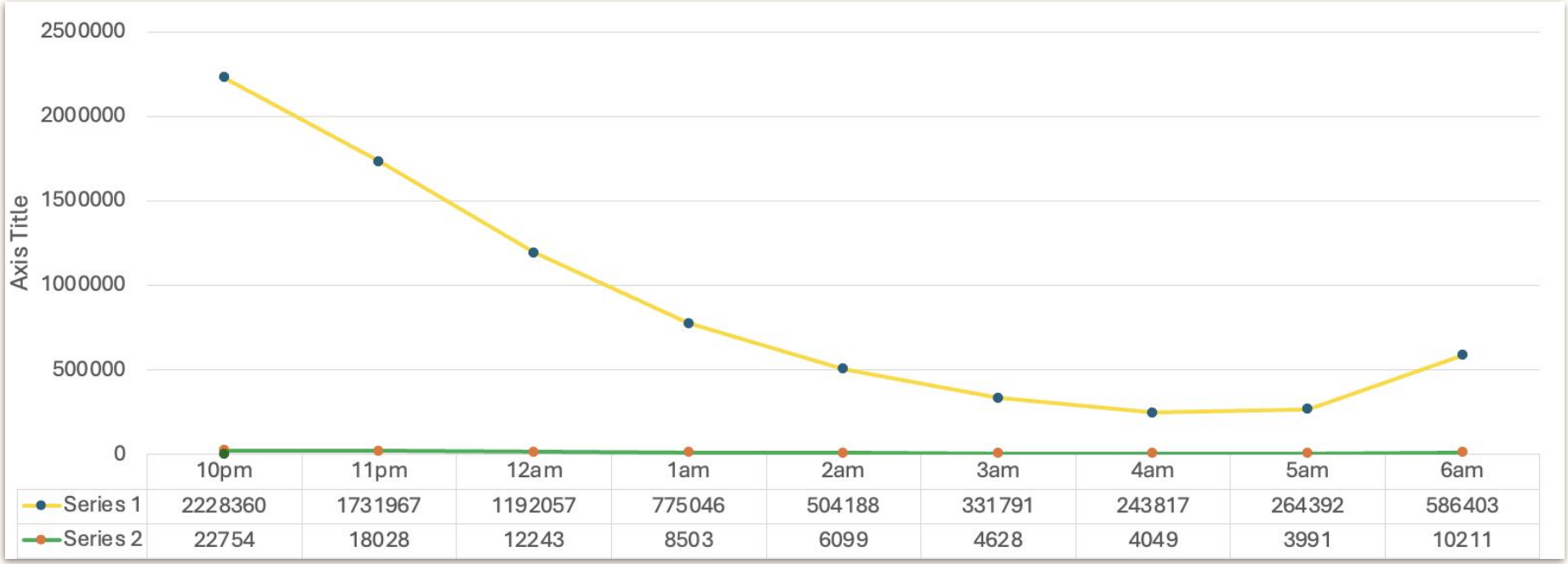
2024 Yellow/Green Taxi
Row: 7,948,527
Columns: 36

Taxi Riderships by Borough



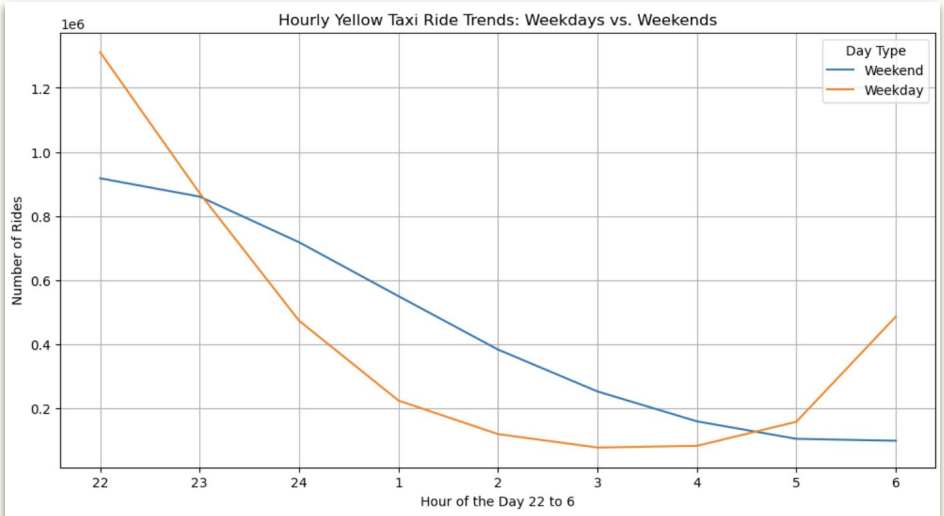
	Manhattan	Brooklyn	Queens	Bronx
Yellow	6769595	185959	838124	30677
Green	32719	20071	34653	2521

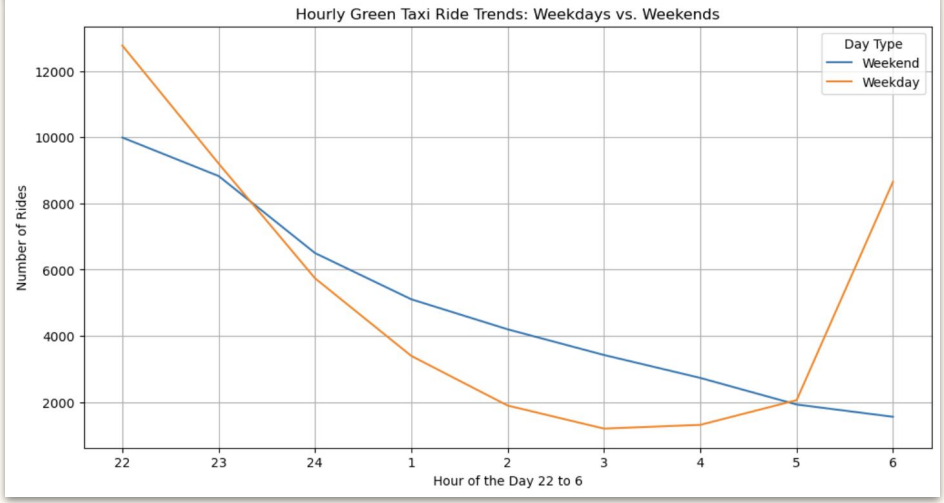
Taxi Riderships by Hour



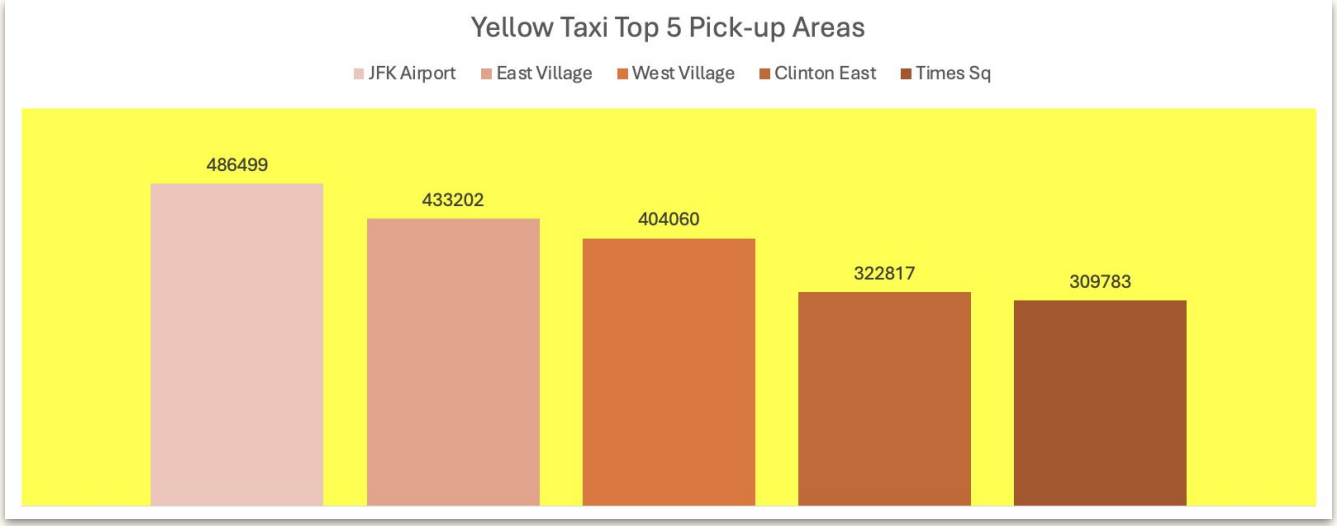
	10pm	11pm	12am	1am	2am	3am	4am	5am	6am
Series 1	2228360	1731967	1192057	775046	504188	331791	243817	264392	586403
Series 2	22754	18028	12243	8503	6099	4628	4049	3991	10211

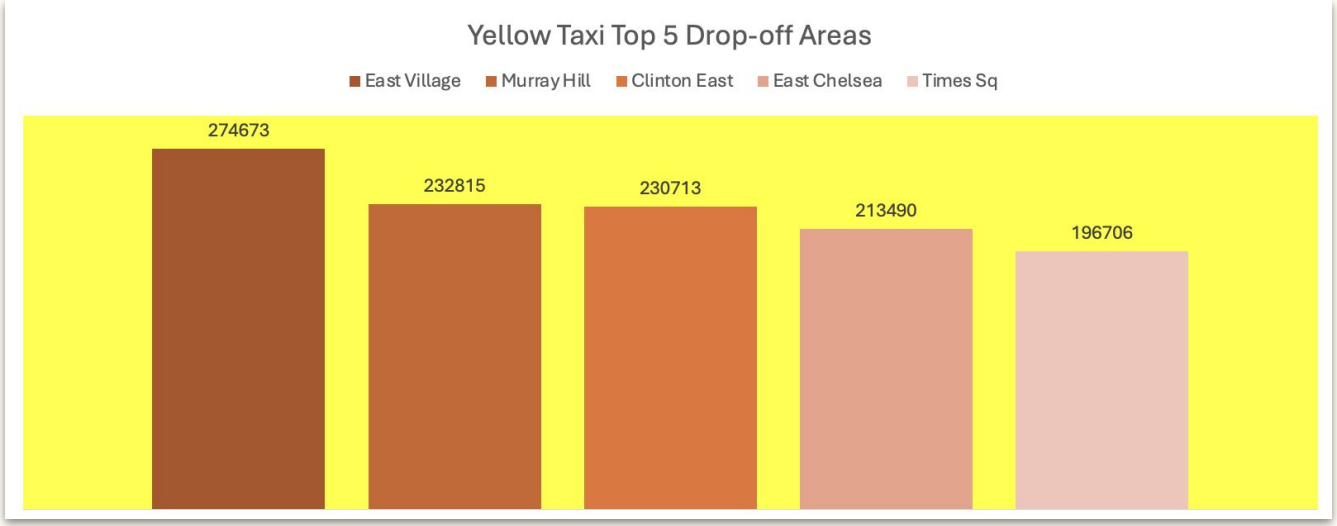
Hourly Taxi Ride Trends Weekdays vs. Weekdays



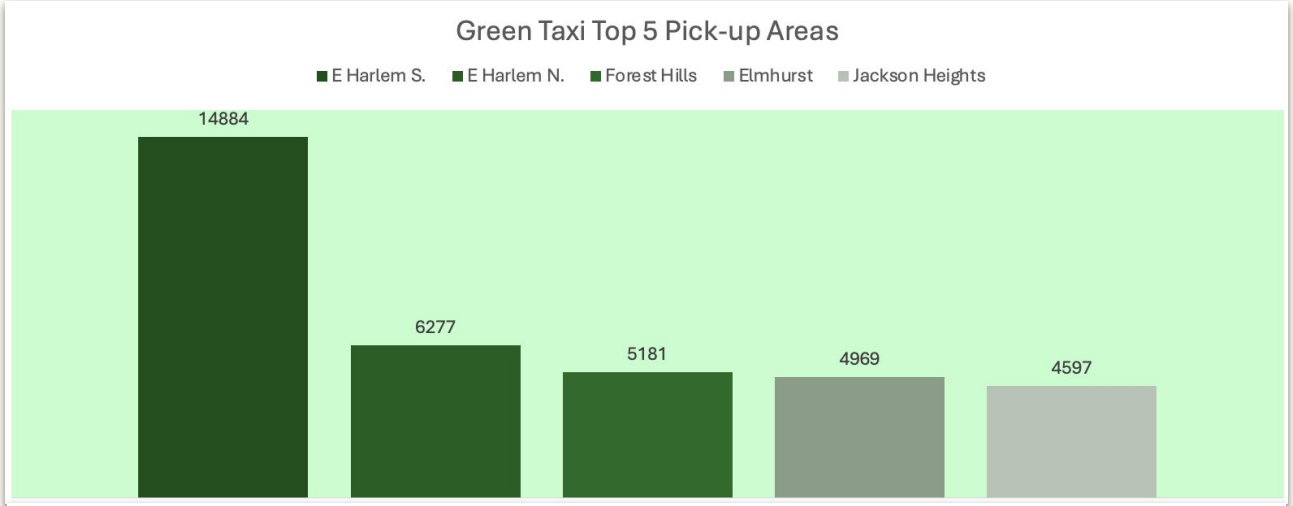


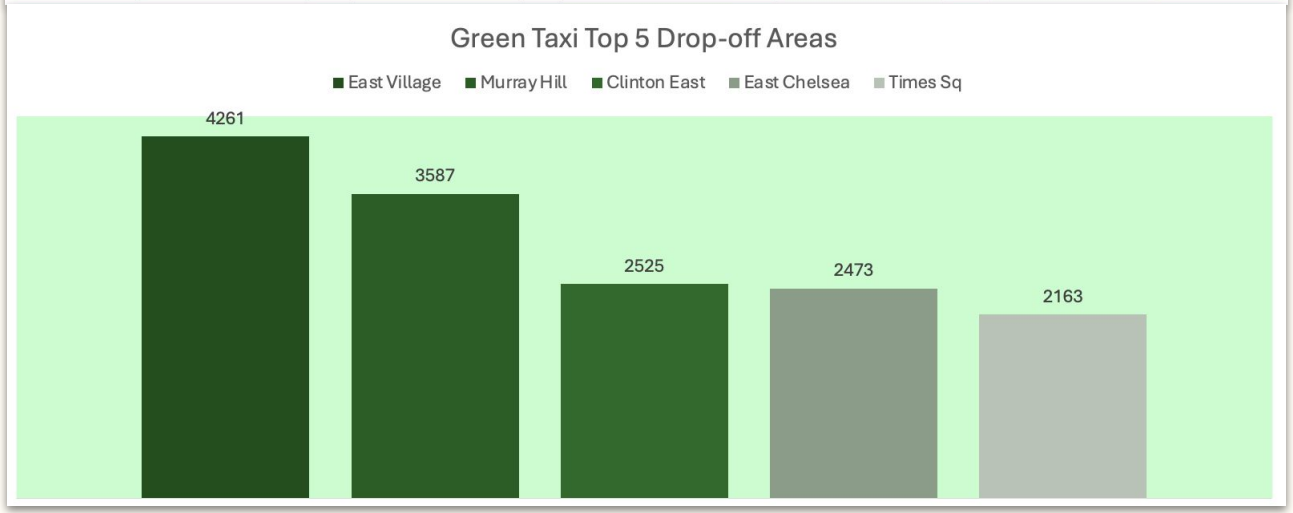
Yellow Taxi Top Pick-up & Drop-off Areas





Green Taxi Top Pick-up & Drop-off Areas





Research Questions...

1. How does subway and taxi ridership vary across neighborhoods with different median incomes?
2. Do higher-income neighborhoods have different peak ridership hours compared to lower-income areas?
3. How does late-night taxi and subway ridership compare across neighborhoods with different income levels?
4. Do neighborhoods with higher median incomes have a higher share of taxi rides compared to subway rides?

Machine Learning Model

Time Series Forecasting

- ARIMA to predict future Uber or subway ridership patterns based on historical data by time
- Use the hourly data of riderships

Tree Model

- To predict riderships (uber/yellow taxi/subway) by neighborhoods
- Classify neighborhoods based on demographic factors like household income, population density, etc.

Implications for Stakeholders

Government Agencies (MTA, TLC, NYC Planning):

- **Policy**: Use data to inform transportation planning and allocate resources equitably across neighborhoods.

Ride-Sharing Companies:

- **Service Expansion**: Use demographic insights to optimize service offerings in underserved neighborhoods.

Local Communities and Residents:

- **Accessibility**: Advocate for improved transportation options in areas with high reliance on public transit or ride-sharing.

Ethical, Legal, & Societal Implications

Data Privacy and Security:

- **Consent**: Uphold ethical standards by ensuring data collection practices are transparent and consent-based.

Equity and Fairness in Transportation Access:

- **Accessibility**: Promote equitable transportation solutions for underserved communities, especially low-income neighborhoods.

Accountability in Service Provision:

- **Responsibility:** Ensure transportation companies and government agencies are responsible for providing fair, reliable, and accessible services to all communities.

Thank You!