NYC Taxi and Limousine Commission (TLC) Analysis

Jessica Woods, Fabian, Nikhil Singh

Overview

- Background on the datasets
- Implementation of jobs in Flink
- Questions about the data - Analytics results
- Conclusions

Field Name	Description
VendorID	A code indicating the TPEP provider that provided the record. **1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.**
tpep_pickup_datetime	The date and time when the meter was engaged.
tpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
PULocationID	TLC Taxi Zone in which the taximeter was engaged
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged
RateCodeID	The final rate code in effect at the end of the trip. **1= Standard rate** **2=JFK** **3=Newark** **4=Nassau or Westchester** **5=Negotiated fare** **6=Group ride**
Store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server. **Y= store and forward trip** **N= not a store and forward trip**
Payment_type	A numeric code signifying how the passenger paid for the trip. **1= Credit card** **2= Cash** **3= No charge** **4= Dispute** **5= Unknown** **6= Voided trip**
Fare_amount	The time-and-distance fare calculated by the meter.
Extra	Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges.
MTA_tax	$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	$0.30 improvement surcharge assessed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.
Tolls_amount	Total amount of all tolls paid in trip.
Total_amount	The total amount charged to passengers. Does not include cash tips.

Yellow Taxi Data

- Traditionally hailed by signaling to a driver who is on duty (street hail)
  - May also be hailed using an e-hail app like Curb or Arro
- Yellow taxis are the only vehicles permitted to respond to a street hail from a passenger in all five boroughs
- Records collected and provided to the NYC TLC by tech. service providers

Trip_type	A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver. **1= Street-hail** **2= Dispatch**

Green Taxi Data

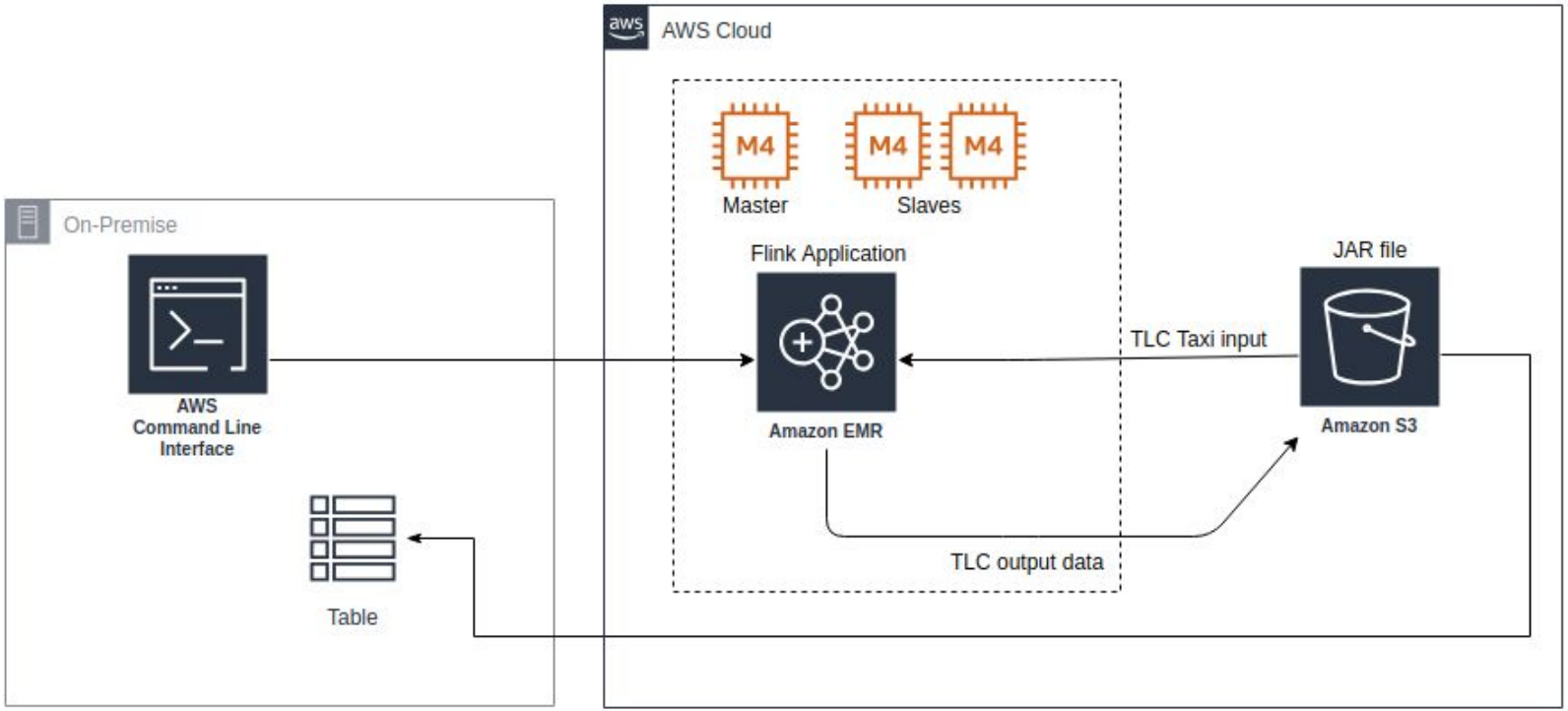
- Also known as boro taxis and street-hail liveries
- Introduced in August of 2013 to improve taxi service and availability in the boroughs
- Only operate in areas above W 110 St/E 96th St in Manhattan and in the boroughs

Field Name	Description
Dispatching_base_num	The TLC Base License Number of the base that dispatched the trip
Pickup_datetime	The date and time of the trip pick-up
DropOff_datetime	The date and time of the trip dropoff
PULocationID	TLC Taxi Zone in which the trip began
DOLocationID	TLC Taxi Zone in which the trip ended
SR_Flag	Indicates if the trip was a part of a shared ride chain offered by a High Volume FHV company (e.g. Uber Pool, Lyft Line). For shared trips, the value is **1**. For non-shared rides, this field is **null**. **NOTE:** For most High Volume FHV companies, only shared rides that were requested AND matched to another shared-ride request over the course of the journey are flagged. However, Lyft (base license numbers **B02510** + **B02844**) also flags rides for which a shared ride was requested but another passenger was not successfully matched to share the trip—therefore, trips records with SR_Flag=1 from those two bases could indicate EITHER a first trip in a shared trip chain OR a trip for which a shared ride was requested but never matched. Users should anticipate an overcount of successfully shared trips completed by Lyft.

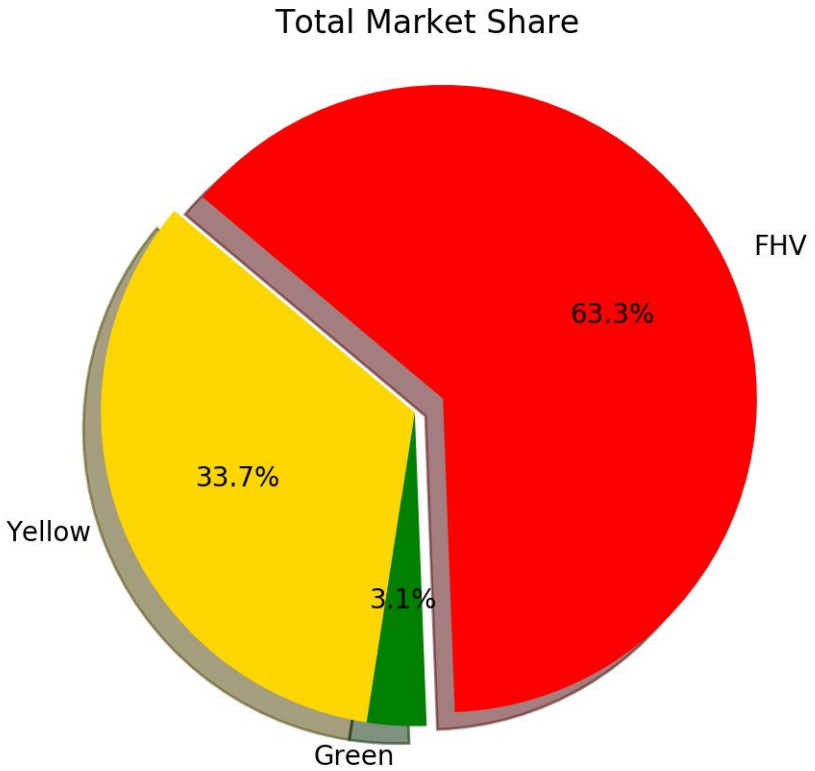
For-Hire Vehicles (FHV)

- FHV data includes trip data from high-volume for-hire vehicle bases
  - Uber
  - Lyft
- Records include fields capturing:
  - pick-up and drop-off dates/times
  - pick-up and drop-off locations
  - trip distances
- All data for February 2018

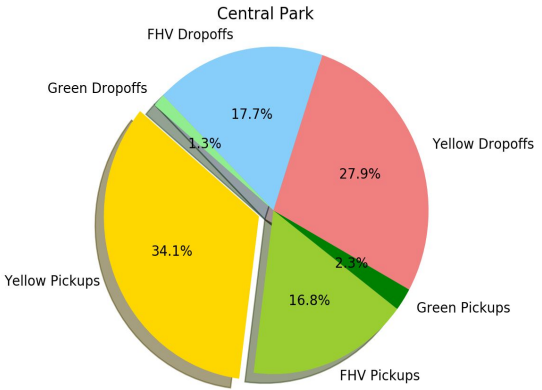
Flink Implementation

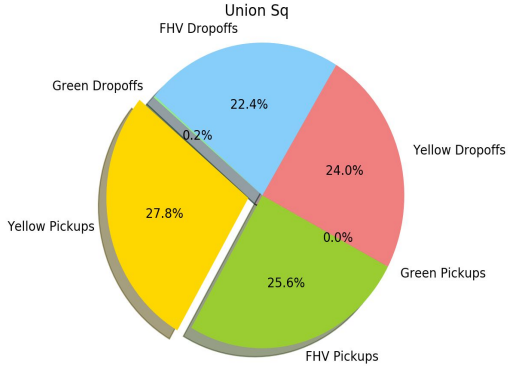


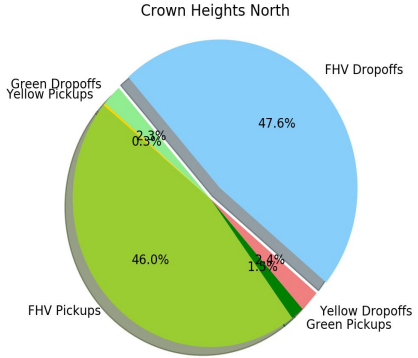
Total Market Share

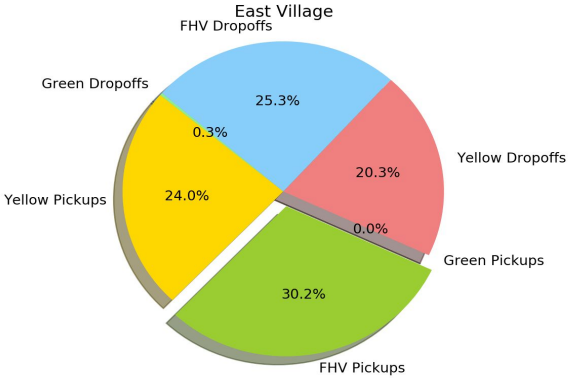


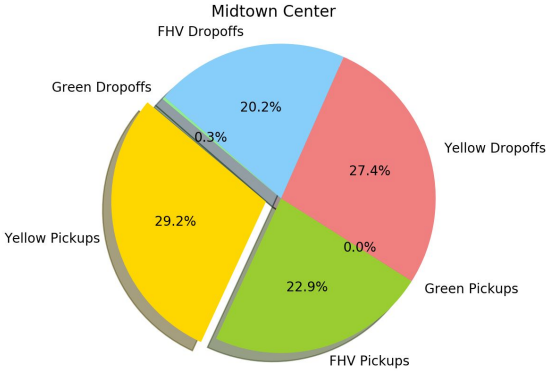
Borough Trends - Pickups/Drop offs



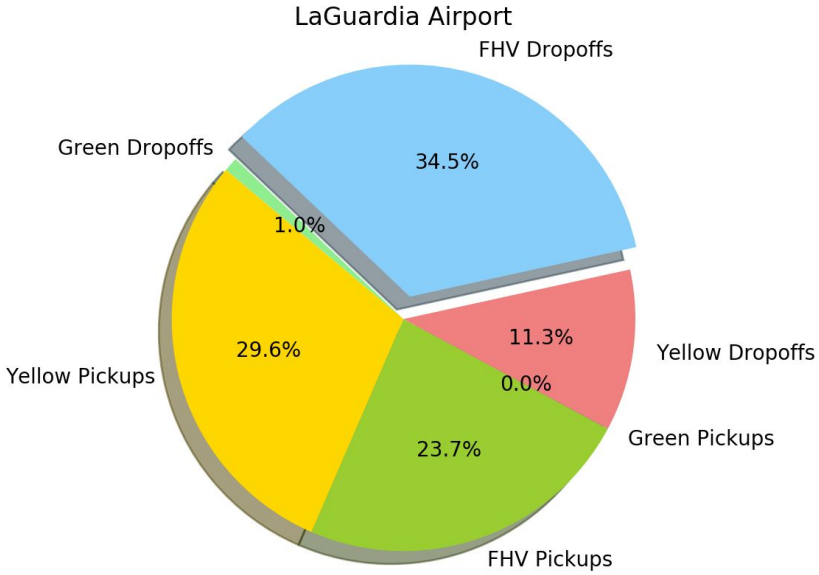


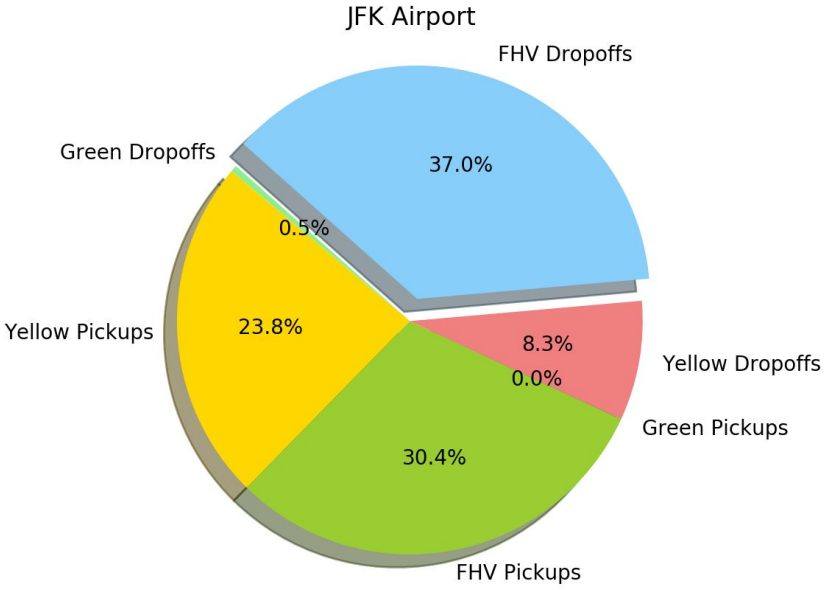




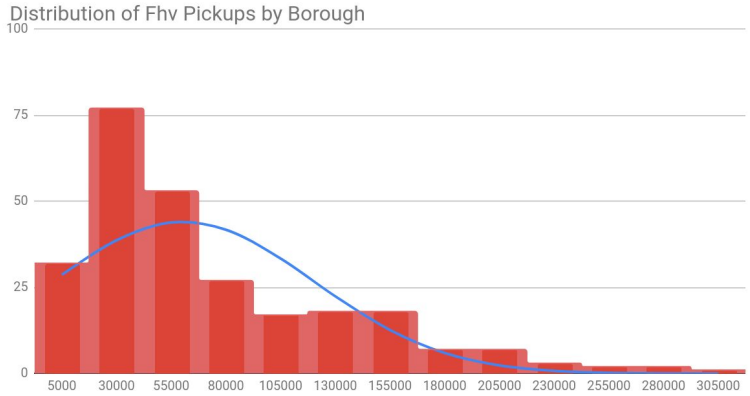


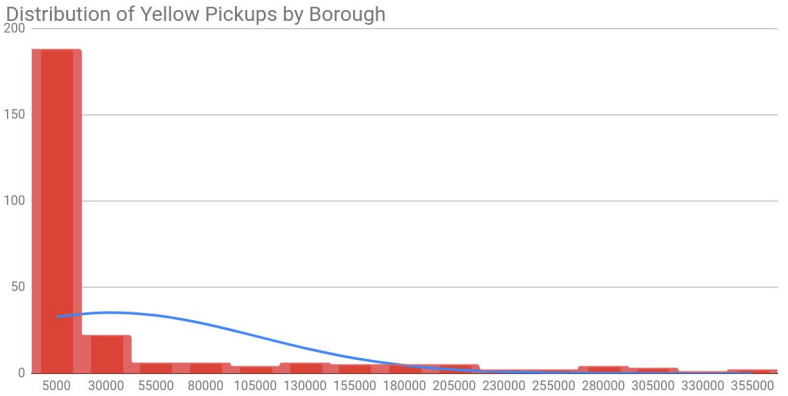
Borough Trends - Pickups/Drop offs

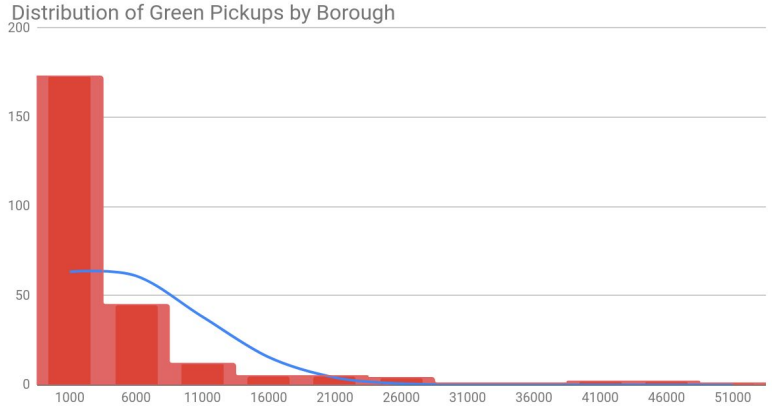




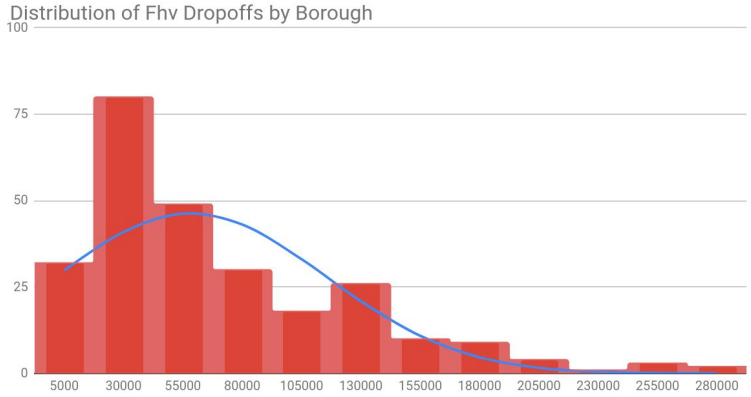
Distribution of Pickups by Borough

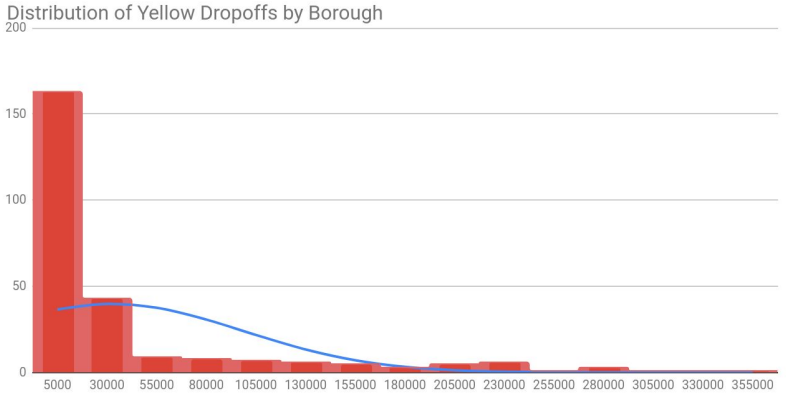


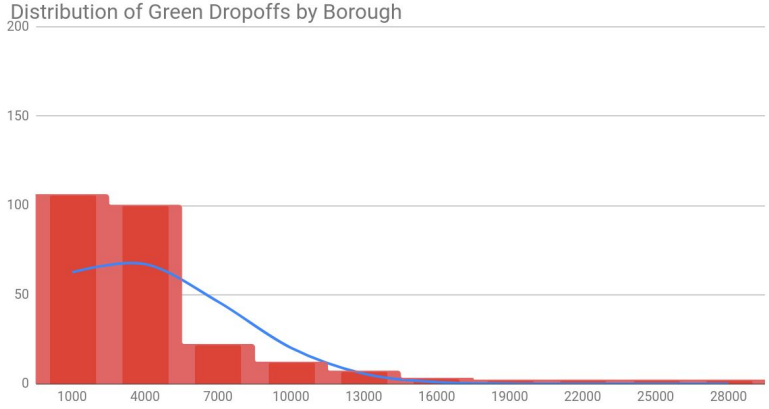




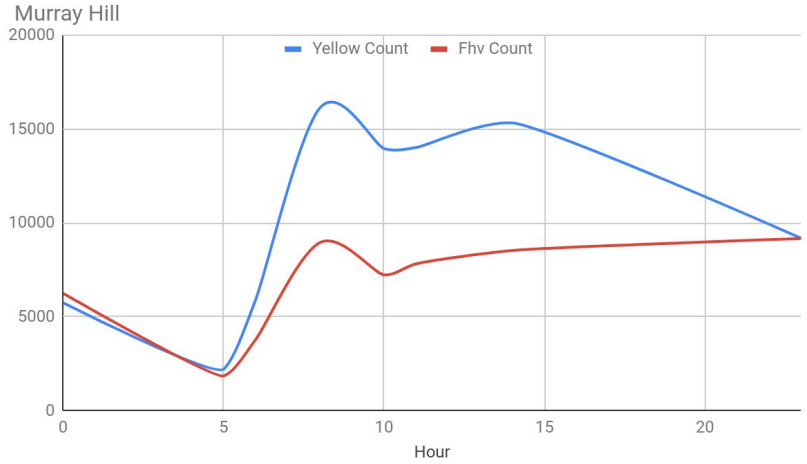
Distribution of Drop offs by Borough

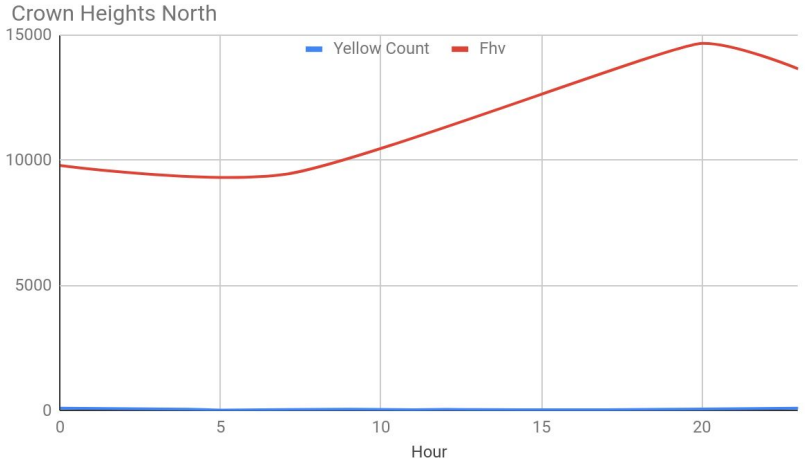


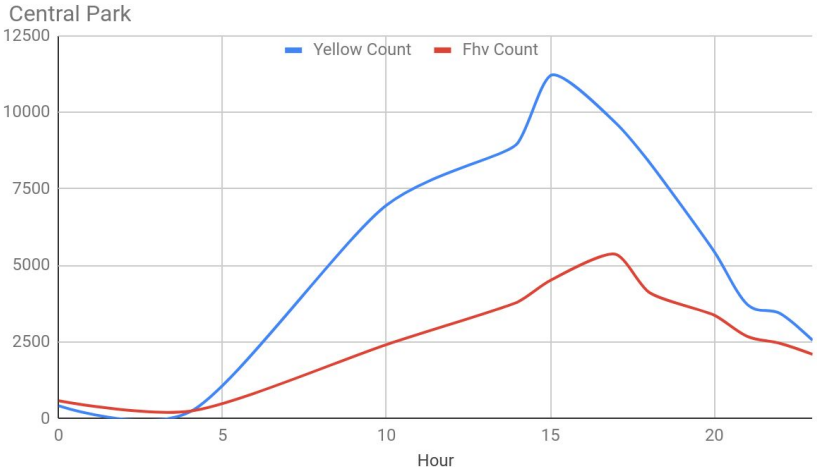


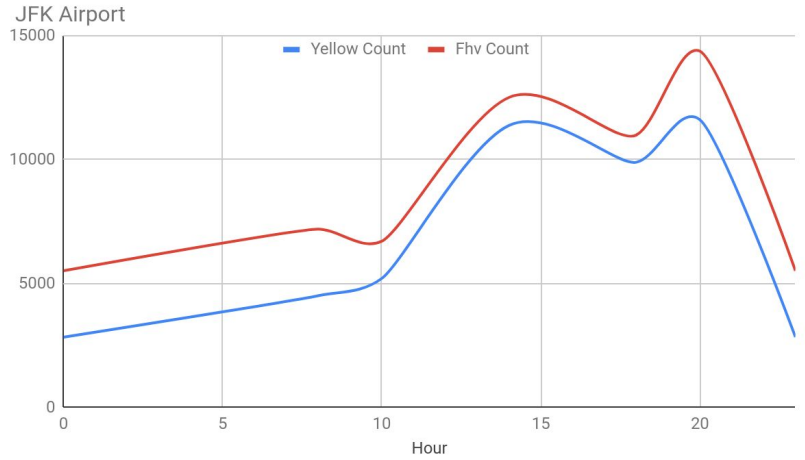


Borough Trends - taxi requests vs time of day



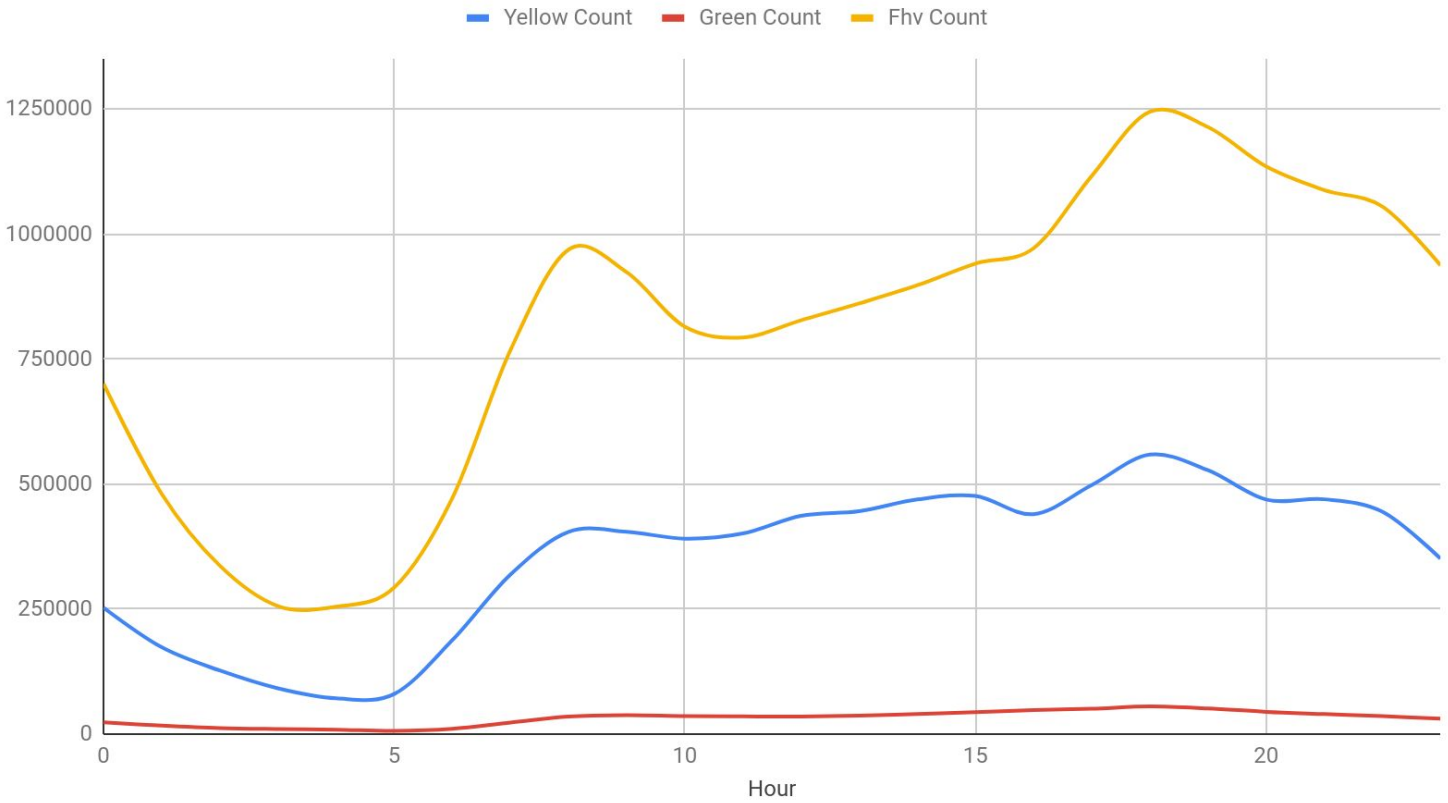






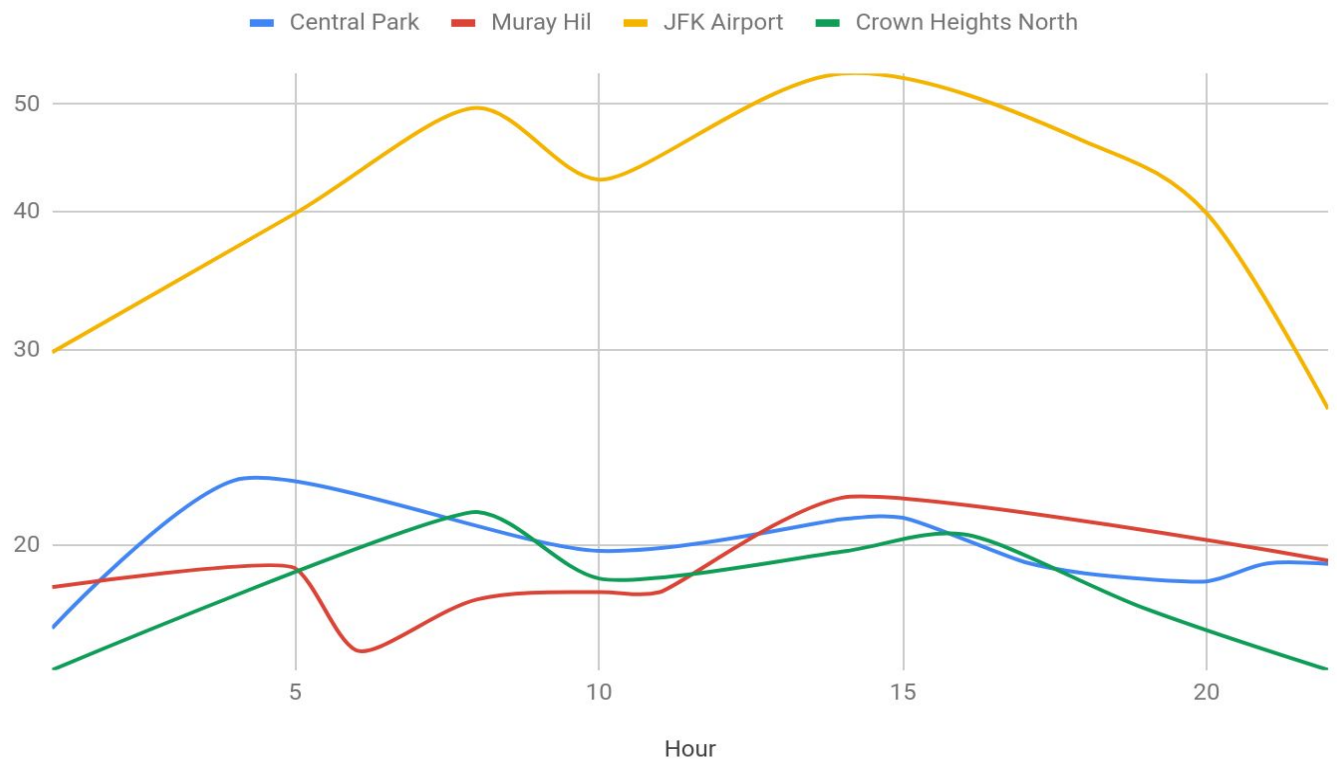
Popular times to travel

When do users usually request taxis?



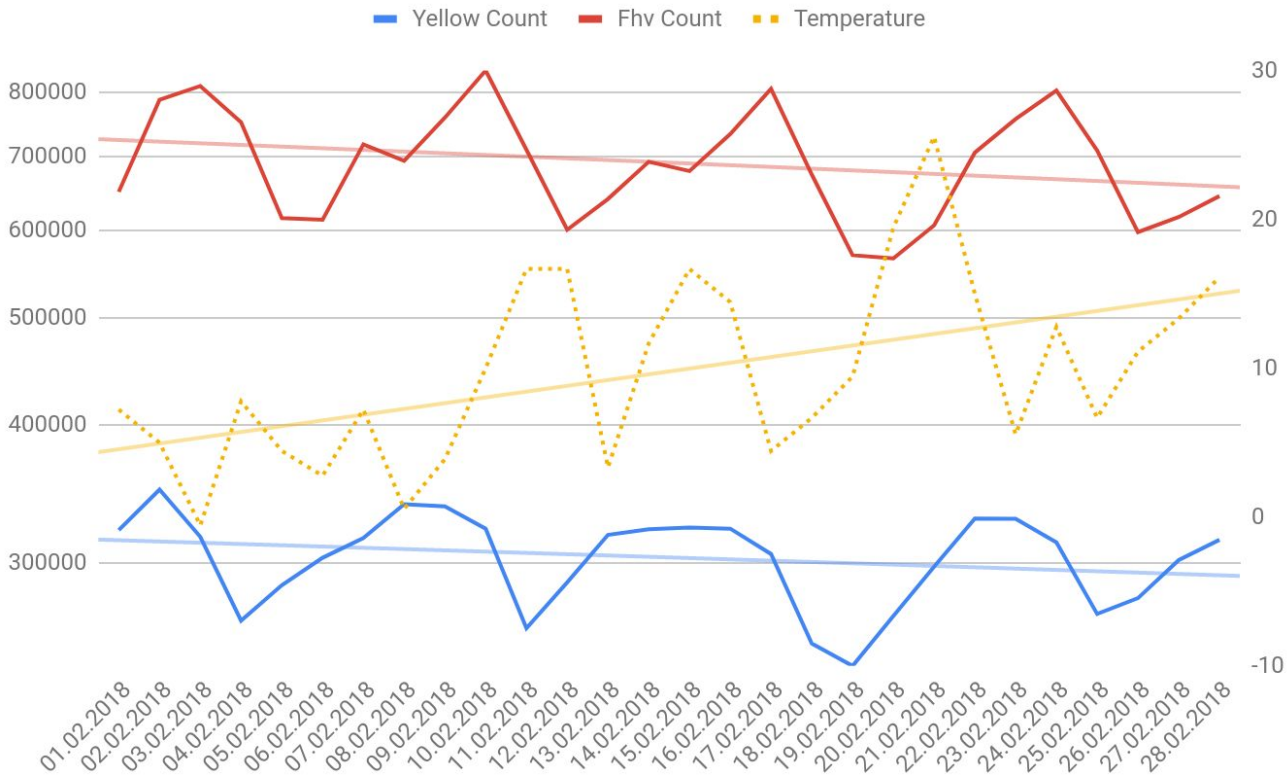
Average Travel Time

Average travel times per borough throughout the day



Weather and Taxi usage

How does weather affect taxi usage in a day



Conclusions

- **Overall, fhv are dominating the market**
- **Boroughs have taxi preferences**
  - Yellow taxis still dominate Manhattan
  - Brooklyn prefer fhv
  - Airports show a more equal preference
    - Fhv being more preferred for drop off
    - Yellow taxis more preferred for pickup
- **Distribution of taxi pickups**
  - Fhv have a more even distribution
  - Other taxis are skewed because they are restricted from certain areas
  - Distribution of taxi drop offs are more even
    - Introduction of green taxis helped to even the distribution for yellow taxis

Conclusions

- **Popular times to travel**
  - Times - before and after work
  - Green taxis fairly constant rate
  - Yellow taxi and fhv have similar peak times
  - Fhv have a larger number of requests in the evening
- **Traffic plays a crucial role in average commute**
  - Airport time having the greatest numbers.
- **There exists a correlation between weather and taxi usage**
  - Fhv peaks on Saturdays
  - Yellow taxi peaks on Thursdays
  - Yellow taxis most inactive on Sundays
  - Fhv are most inactive on Mondays

References

Gitlab: git@gitlab.tubit.tu-berlin.de:jhw0014/CITinitProject.git

Flink produced datasets:
https://docs.google.com/spreadsheets/d/1BxWVdVOO_9X5HjbBbSXjAJ1u5fBOADDEGt791iiZQdM/edit?usp=sharing

Original TLC data: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page