PROJECT ON UBER DATASET

DATA DESCRIPTION

The data contains the details for the Uber rides across various boroughs (subdivisions) of New York City at an hourly level and attributes associated with weather conditions at that time.

- pickup_dt: Date and time of the pick-up.
- borough: NYC's borough.
- pickups: Number of pickups for the period (hourly).
- spd: Wind speed in miles/hour.
- vsb: Visibility in miles to the nearest tenth.
- temp: Temperature in Fahrenheit.
- dewp: Dew point in Fahrenheit.
- slp: Sea level pressure.
- pcp01: 1-hour liquid precipitation.
- pcp06: 6-hour liquid precipitation.
- pcp24: 24-hour liquid precipitation.
- sd: Snow depth in inches.
- hday: Being a holiday (Y) or not (N).

Importing the libraries required for datasets.

- pandas as pd
- numpy as np
- seaborn as sns
- matplotlib.pyplot as plt

- calendar

Reading the given dataset.

Display the top 5 rows.

	0	1	2	3	4
pickup_dt	1/1/2015 1:00	1/1/2015 1:00	1/1/2015 1:00	1/1/2015 1:00	1/1/2015 1:00
borough	Bronx	Brooklyn	EWR	Manhattan	Queens
pickups	152.0	1519.0	0.0	5258.0	405.0
spd	5.0	5.0	5.0	5.0	5.0
vsb	10.0	10.0	10.0	10.0	10.0
temp	30.0	NaN	30.0	30.0	30.0
dewp	7.0	7.0	7.0	7.0	7.0
slp	1023.5	1023.5	1023.5	1023.5	1023.5
pcp01	0.0	0.0	0.0	0.0	0.0
pcp06	0.0	0.0	0.0	0.0	0.0
pcp24	0.0	0.0	0.0	0.0	0.0
sd	0.0	0.0	0.0	0.0	0.0
hday	Y	?	Y	Y	Y

- Temperature and holiday contains NaN values
- Holiday contains invalid entries

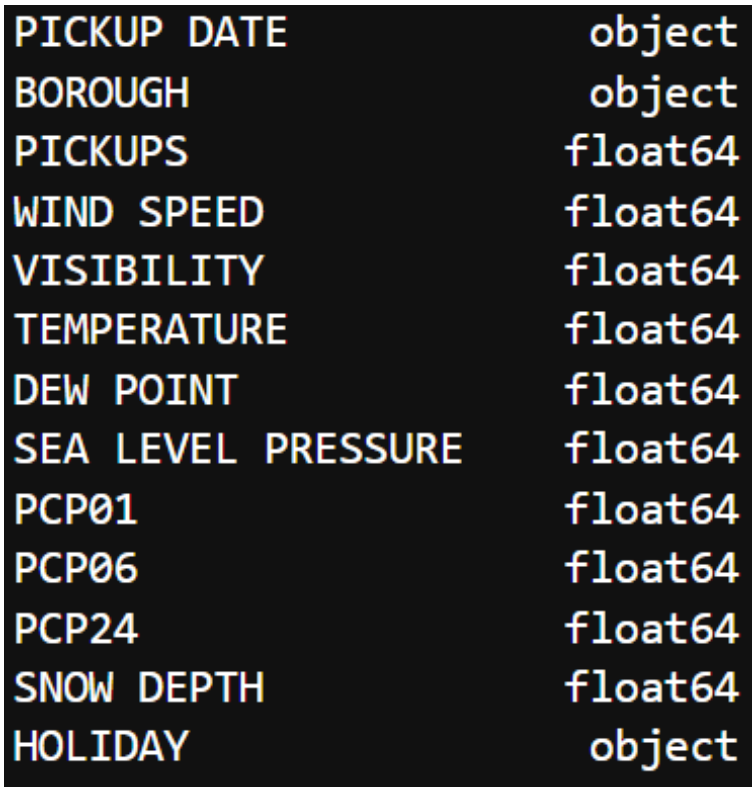
Renaming the columns

	0	1	2	3	4
PICKUP DATE	2015-01-01 01:00:00	2015-01-01 01:00:00	2015-01-01 01:00:00	2015-01-01 01:00:00	2015-01-01 01:00:00
BOROUGH	Bronx	Brooklyn	EWR	Manhattan	Queens
PICKUPS	152.0	1519.0	0.0	5258.0	405.0
WIND SPEED	5.0	5.0	5.0	5.0	5.0
VISIBILITY	10.0	10.0	10.0	10.0	10.0
TEMPERATURE	30.0	NaN	30.0	30.0	30.0
DEW POINT	7.0	7.0	7.0	7.0	7.0
SEA LEVEL PRESSURE	1023.5	1023.5	1023.5	1023.5	1023.5
PCP01	0.0	0.0	0.0	0.0	0.0
PCP06	0.0	0.0	0.0	0.0	0.0
PCP24	0.0	0.0	0.0	0.0	0.0
SNOW DEPTH	0.0	0.0	0.0	0.0	0.0
HOLIDAY	Y	NaN	Y	Y	Y

Checking the shape of dataset

- Data contains 29101 rows and 13 columns

Checking the datatpe of each feature



- Numeric columns are PICKUPS, WINDSPEED, VISIBILITY, TEMPERATURE, DEW POINT ,SEA LEVEL PRESSURE , PCP01 , PCP06 , PCP24 , SNOW DEPTH
- Categorical columns are BOROUGH , HOLIDAY

Checking the statistical summary.

- Total count of pickup borough and temperature are lesser than the shape of dataset which means it contains null values
- Wind speed range is between 0 and 21
- Visibility range is between 0 and 10
- Temperature range is bewteen 0 and 89(Most probably an outlier according to new york)

Checking the null values

PICKUP DATE	0
BOROUGH	3043
PICKUPS	2
WIND SPEED	0
VISIBILITY	0
TEMPERATURE	359
DEW POINT	0
SEA LEVEL PRESSURE	0
PCP01	0
PCP06	0
PCP24	0
SNOW DEPTH	0
HOLIDAY	0

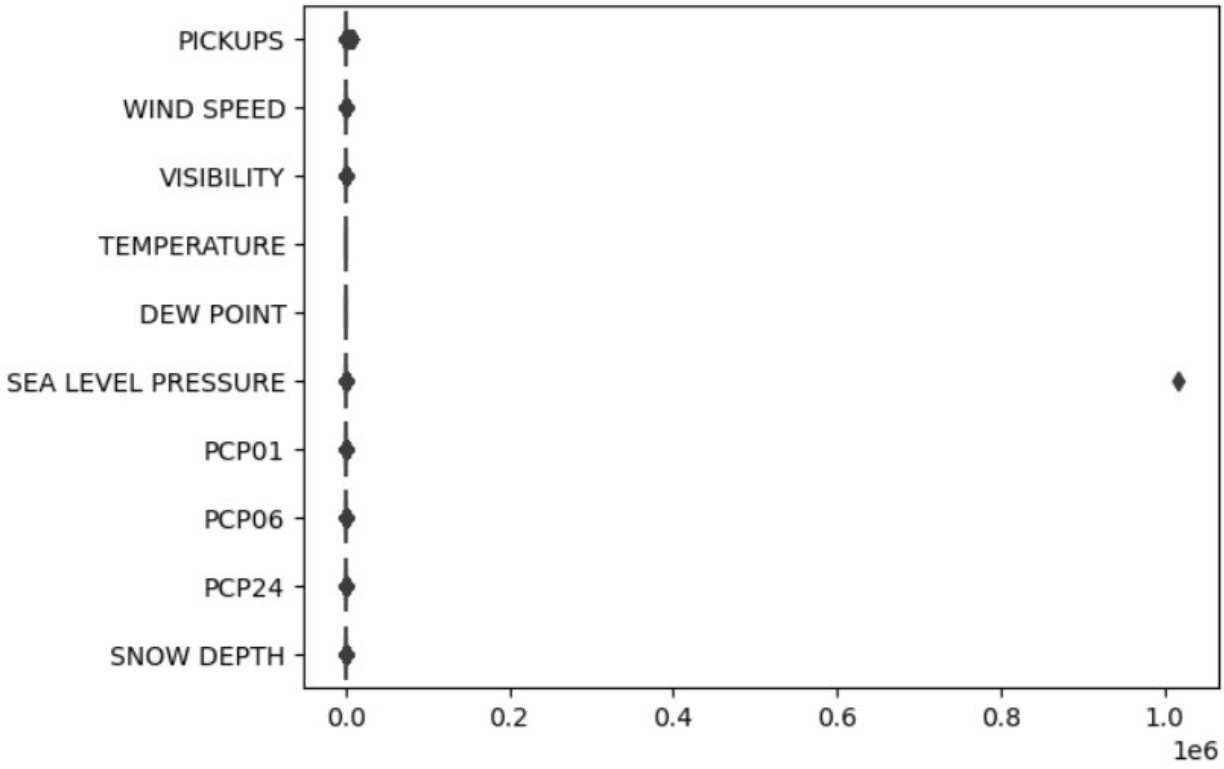
Checking the duplicate values

- There are 0 duplicate entries

Checking the anomalies or wrong entries

- Holiday column contains wrong entries

Plotting boxplot

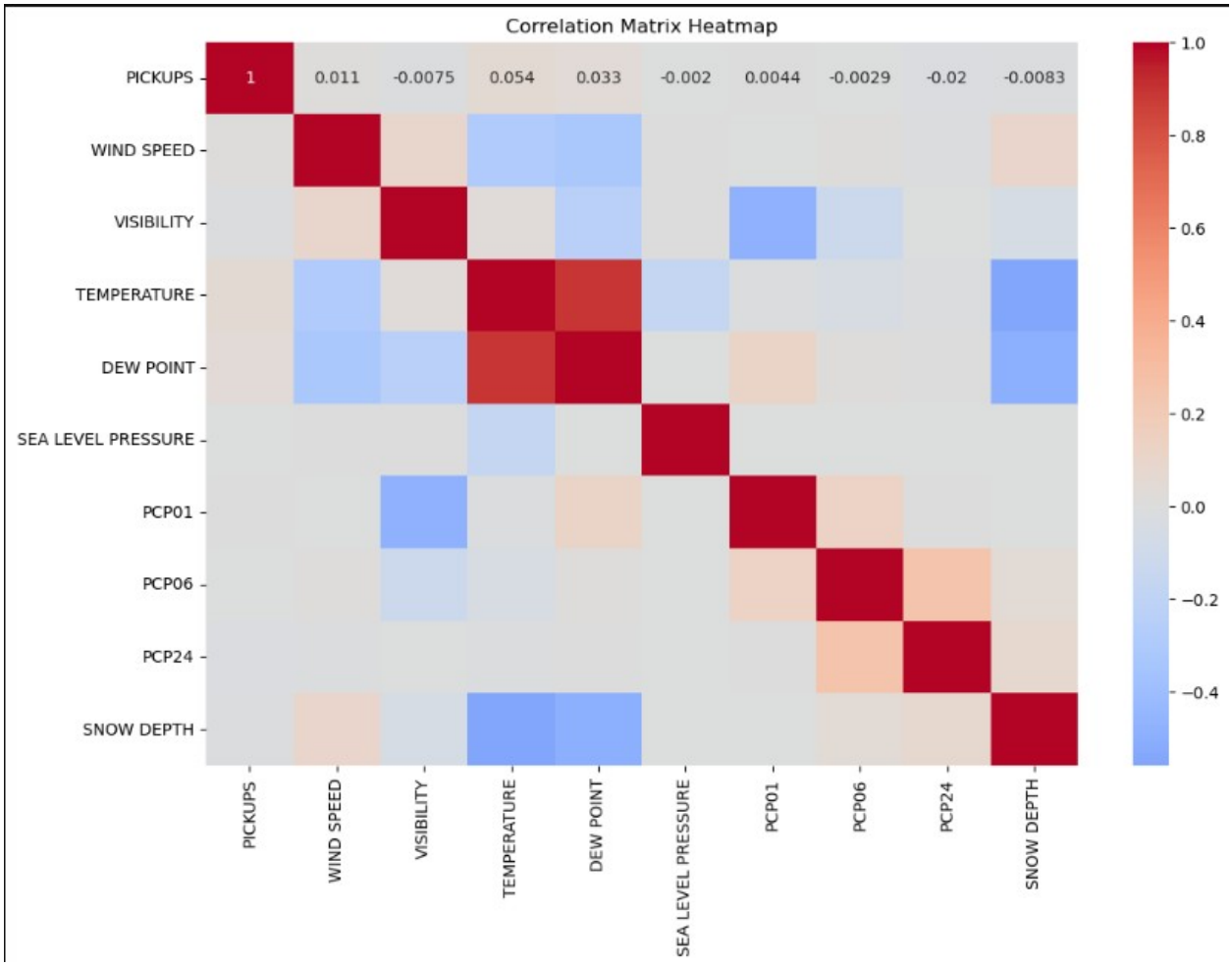


- Sea level pressure have a outlier so high that it disturbs all boxplots

Doing the necessary data cleaning steps

Converting pickup date to datetime format

Finding Correlation



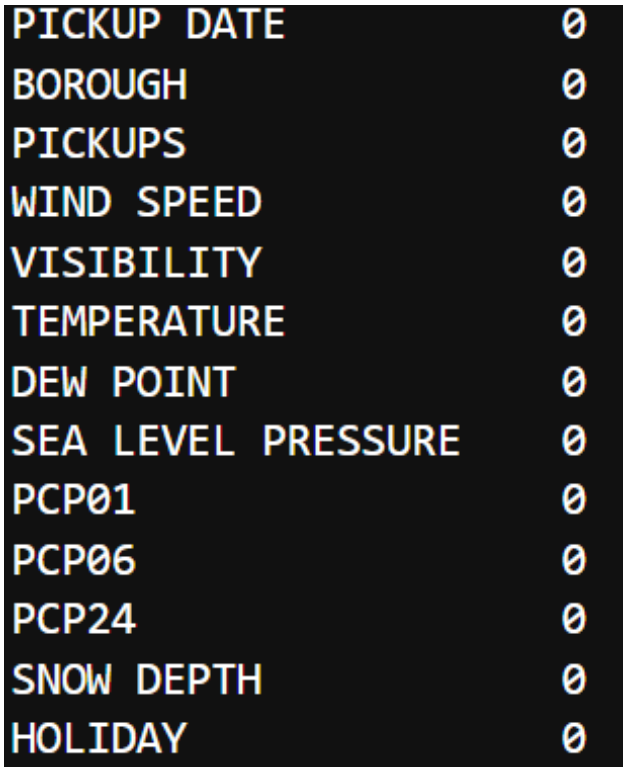
- High correlation between dew point and temperature otherwise no correlation between any column

Removing outliers using IQR

- Original DataFrame shape: (29101, 13)* DataFrame without outliers shape: (20292, 13)

Null value imputation

- Replaced null value of numeric columns with median
- Replaced null value of categorical columns with mode



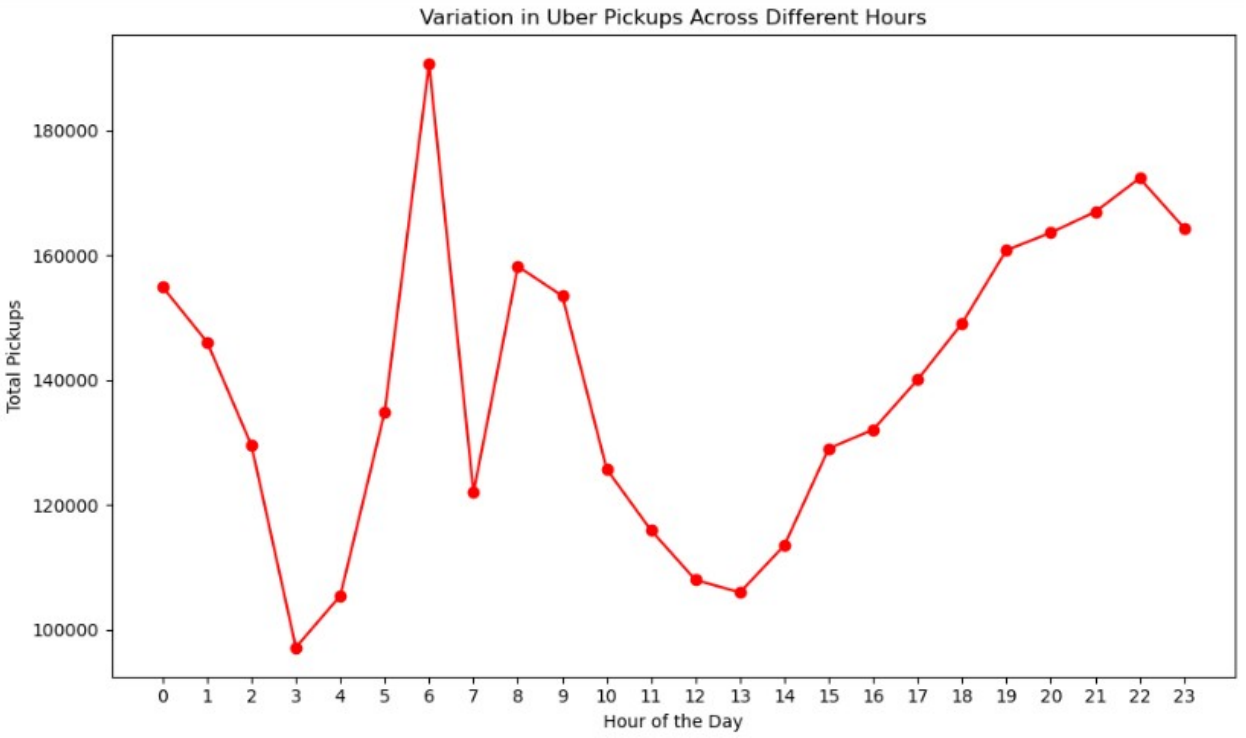
Total numbers of Uber pickups across all boroughs

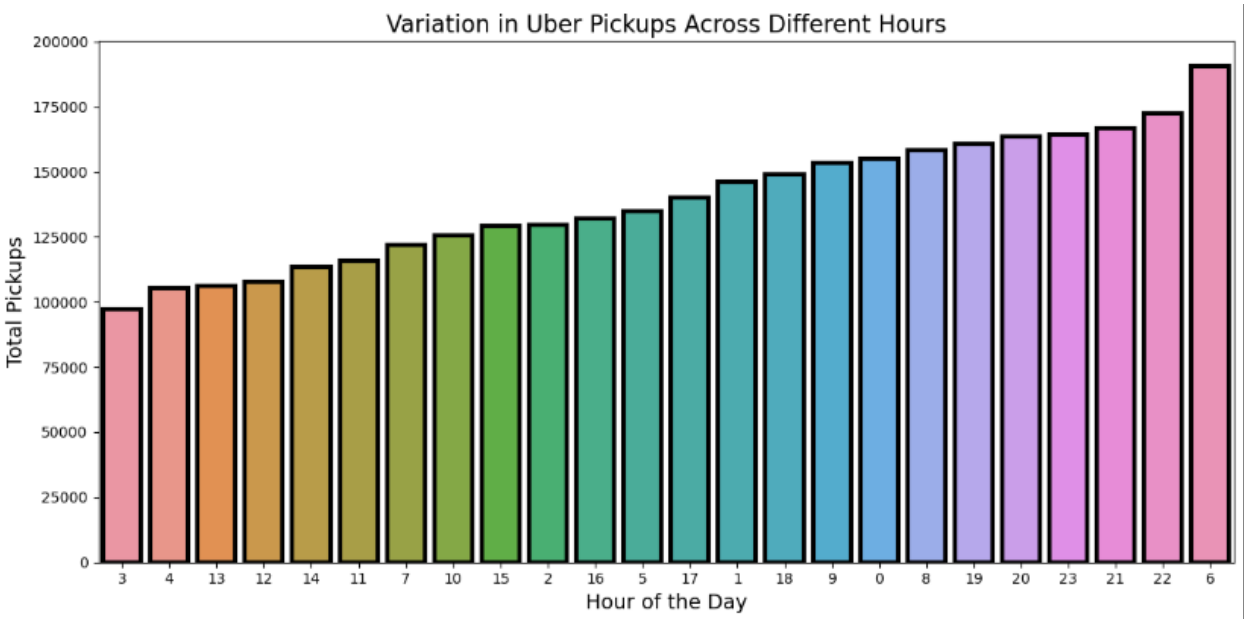
- Total number of Uber pickups across all boroughs: 3340485.0

Which borough has the highest average number of hourly pickups

- Top borough with highest average hourly pickups: Manhatt with 775.304211

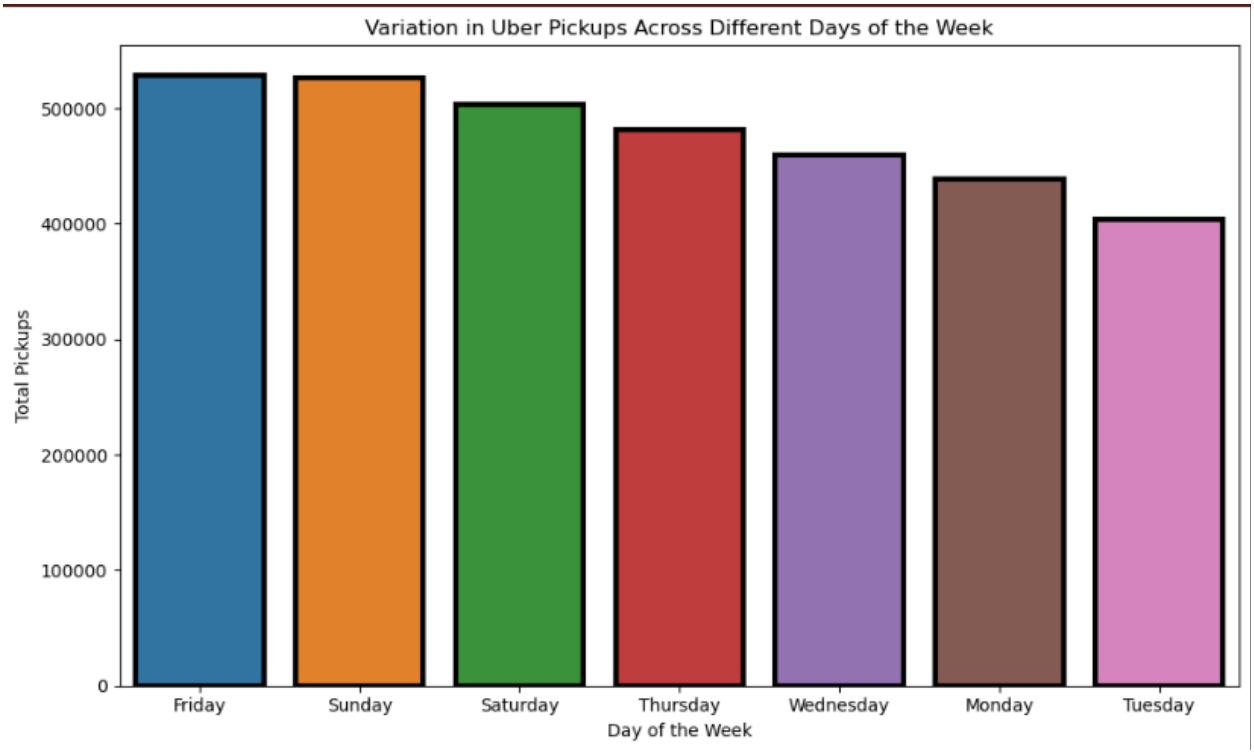
Number of pickups in different hours of the day





- Highest number of overall pickups are highest at 6 Am and lowest are at 3 Am

Which day of the week has the highest number of pickups

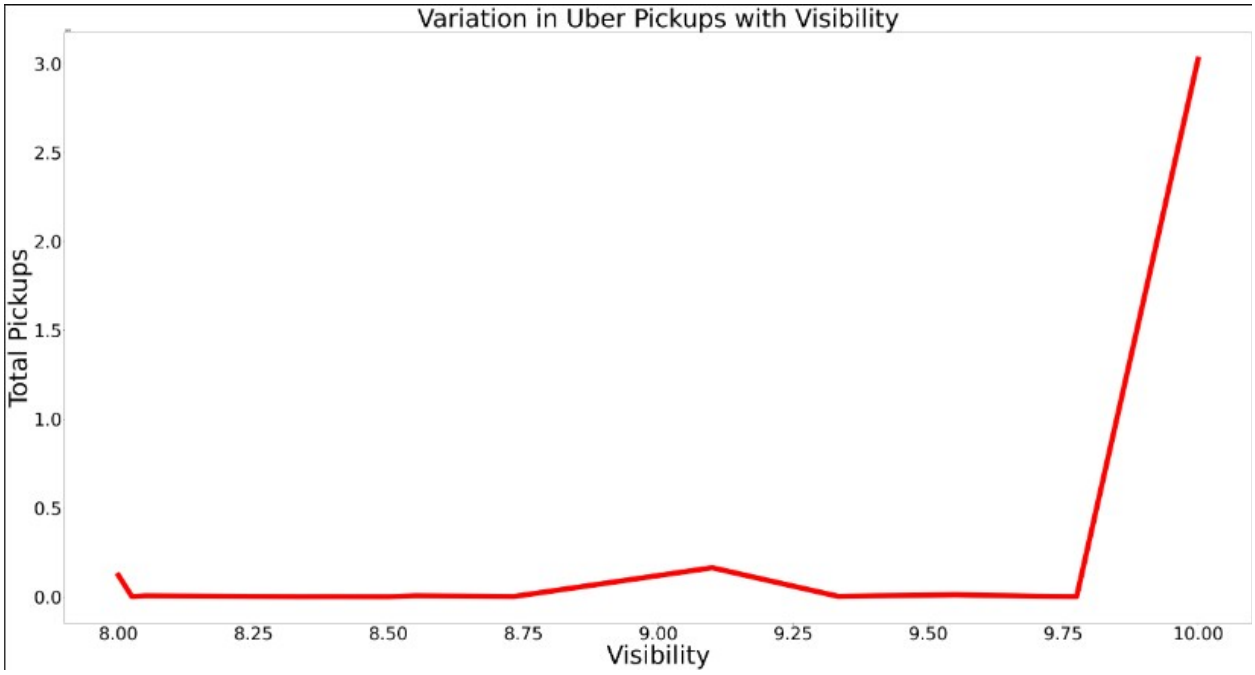


- Friday have highest number of pickups with around 520000 pickups
- Tuesday hav lowest number of pickups with arounf 420000 pickups

Correlation between temperature and pickups

- Correlation between temperature and number of pickups: 0.08936396921929778

Variation in pickup with visibility



- Highest number of pickups are when visibility is above 9.75 or around 10
- Correlation between temperature and number of pickups: -0.00027571779901785533

Relation between wind speed and pickups

- Correlation between windspeed and number of pickups: -0.019813343926501296

Relation between precipitation and number of pickups

- Correlation between PCP01 and number of pickups: -0.003957444334561385*
  Correlation between PCP06 and number of pickups: -0.006318169702729992
- Correlation between PCP24 and number of pickups: -0.016150554112423353

Pickups during holiday and non holiday

HOLIDAY

NO = 3209477

YES = 131008

Snow Depth relation with pickups

- Correlation between snowdepth and number of pickups: -0.02071555436587074

Peak hours for Uber pickups in each borough

Bronx: Hour=22 Pickups=10650

Brooklyn: Hour=21 Pickups=89406

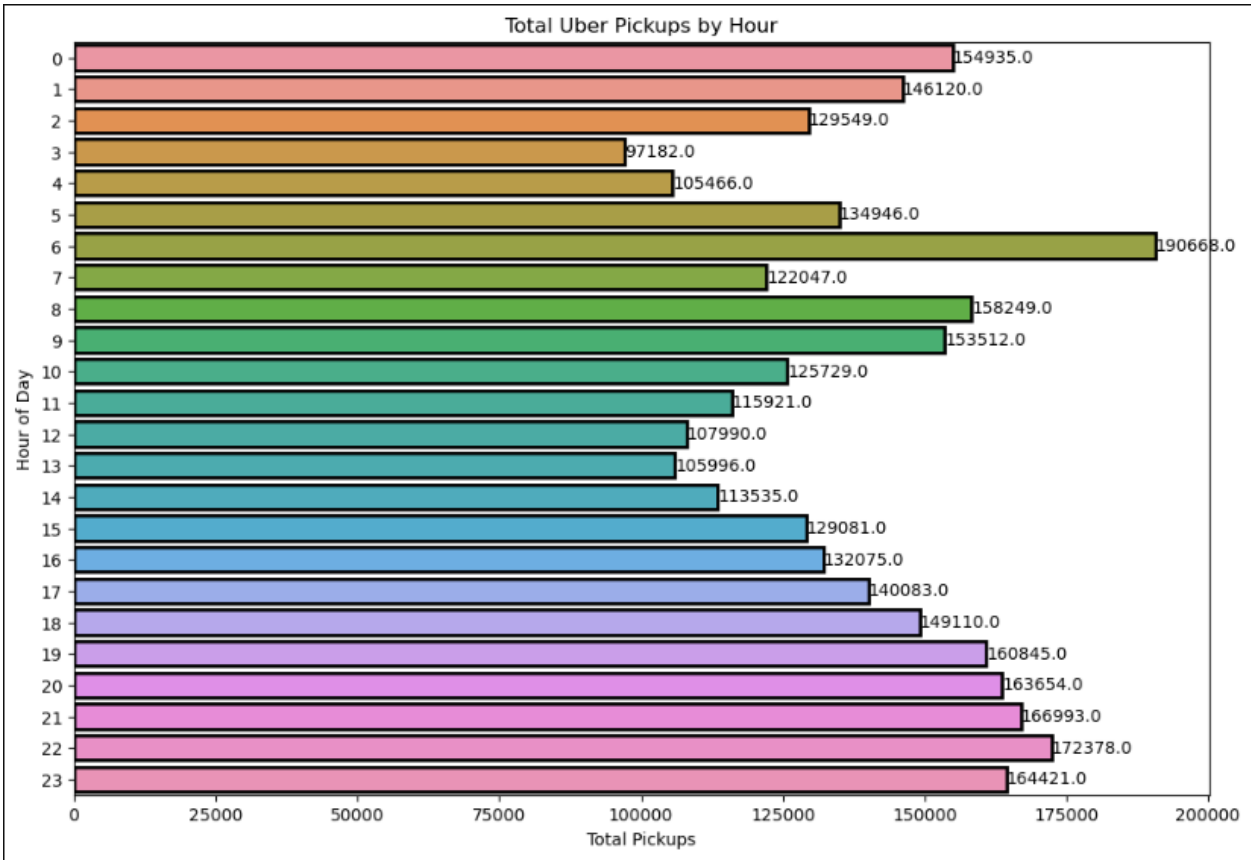
EWR: Hour=7 Pickups=331

Manhattan: Hour=6 Pickups=116196

Queens: Hour=22 Piclups=70905

Staten Island: Hour=18 Pickups=347

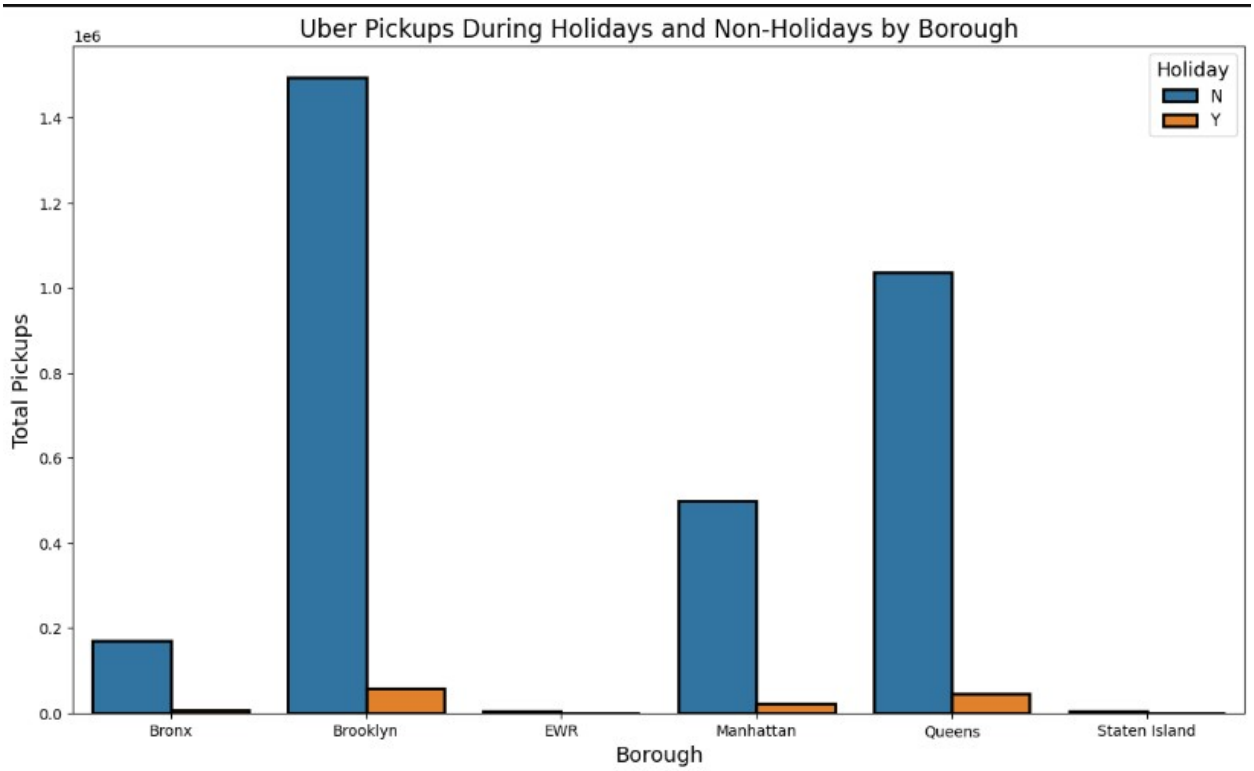
Number of pickups change during each hour



Average number of pickups during late-night hours(12Am-4Am)

HOURS	PICKUPS
0	185.773381
1	169.512761
2	146.880952
3	107.264901
4	113.282492

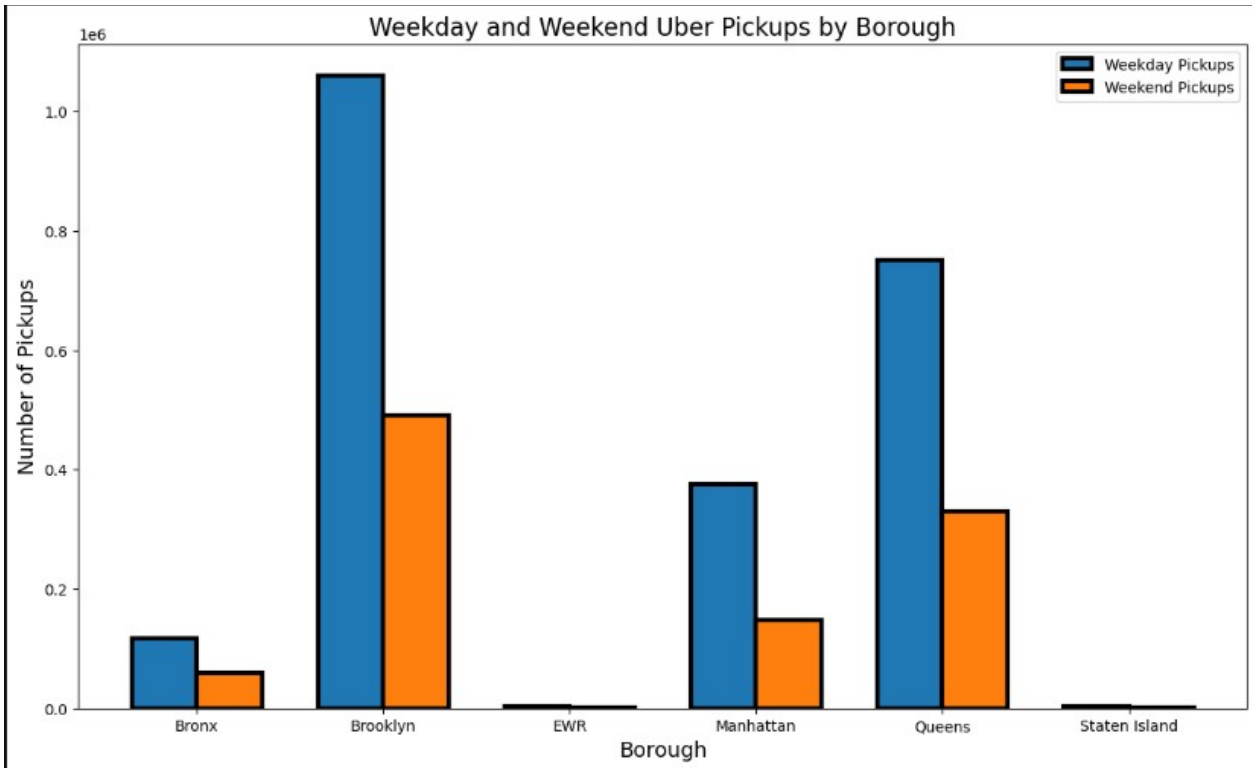
Increase in pickups of each borough during holidays



- Brooklyn have the highest number of difference between pickups in holiday and on non holiday days

- On the other hand this difference is least in Staten Island

Number of pickups compare between weekdays and weekends for each borough



- Difference is highest in Brooklyn
- Difference is lowest in EWR and

Correlation between sea level pressure and pickups

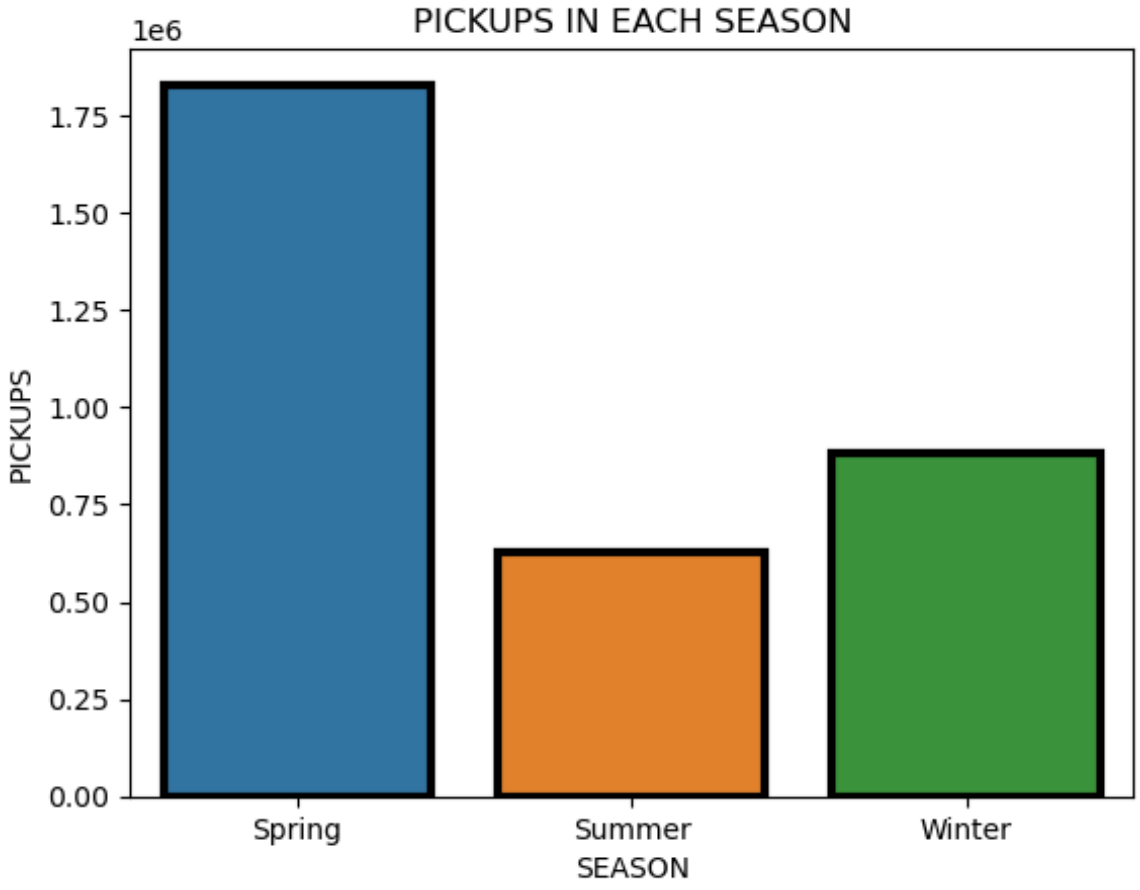
- Correlation between sea level pressure and number of pickups: -0.020560188998132122

Number of operations in different seasons

Spring: 1830144.0

Summer: 628888.0

Winter: 881453.0



THANKS YOU :) :) :)