



Problem Statement / Objective

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).

Index

Index			
SNo	Topic	Page No	Date
1	Cover Page and Objective	1	17/05/24
2	Index	2	17/05/24
3	List of Tables	2	17/05/24
4	List of Figures	2	17/05/24
5	Basic EDA	4 - 10	18/05/24
6	Question 1	10	19/05/24
7	Question 2,3	11	19/05/24
8	Question 4	12	19/05/24
9	Question 5, 6	13	20/05/24
10	Question 7	14	20/05/24
11	Question 8	15	20/05/24

List of Tables

- Top 5 rows of dataset
- Bottom 5 rows of dataset
- Number of pickups by hour
- Pickup trend due to weather

List of Figures

- Statistical data types
- Numerical data types

Data Dictionary

- 3-Uber_Data_New.csv

The top 5 rows

Observation

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

The last 5 rows

Observation

	pickup_dt	borough	pickups	spd	vsb	temp	dewp	slp	pcp01	pcp06	pcp24	sd	hday
29096	30-06-2015 23:00	EWB	0.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29097	30-06-2015 23:00	Manhattan	3828.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29098	30-06-2015 23:00	Queens	580.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29099	30-06-2015 23:00	Staten Island	0.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29100	30-06-2015 23:00	NaN	3.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N

Shape of dataset

Observation

- The dataset is having 29101 rows and 13 columns

Datatypes of each feature

Observation

- pickup_dt and borough are having object datatypes
- pickups, spd, vsb, temp, dewp, slp, pcp01, pcp06, pcp24, sd, hday contains numerical datatype

Null values

Observation

- Borough contains 3043 null values, needs to be checked
- pickups contains 2 null values, needs to be checked
- temp contains 359 null values, needs to be checked

Duplicate values

Observation

- Data set contains no duplicate values

Anomalies or wrong entry

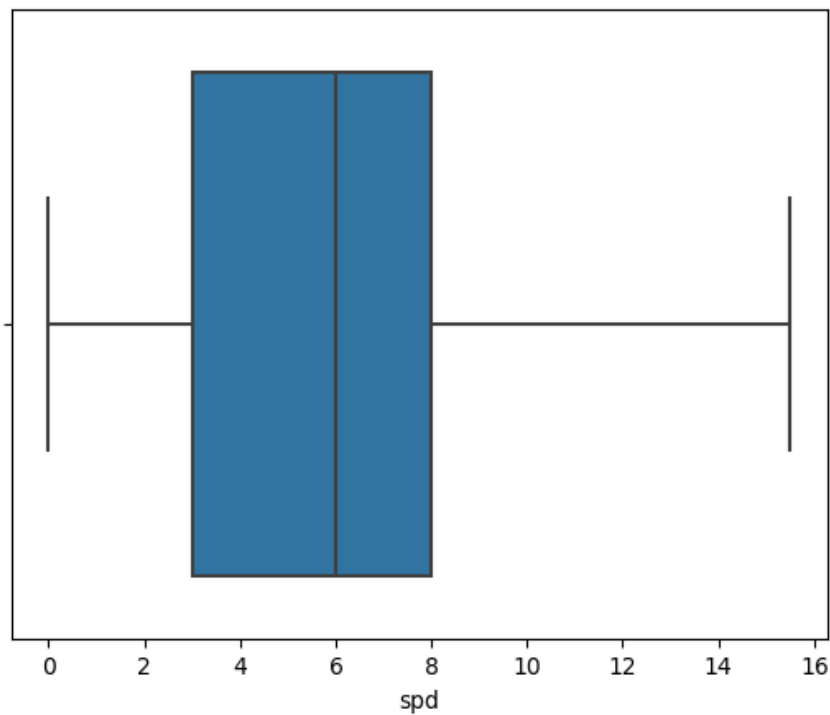
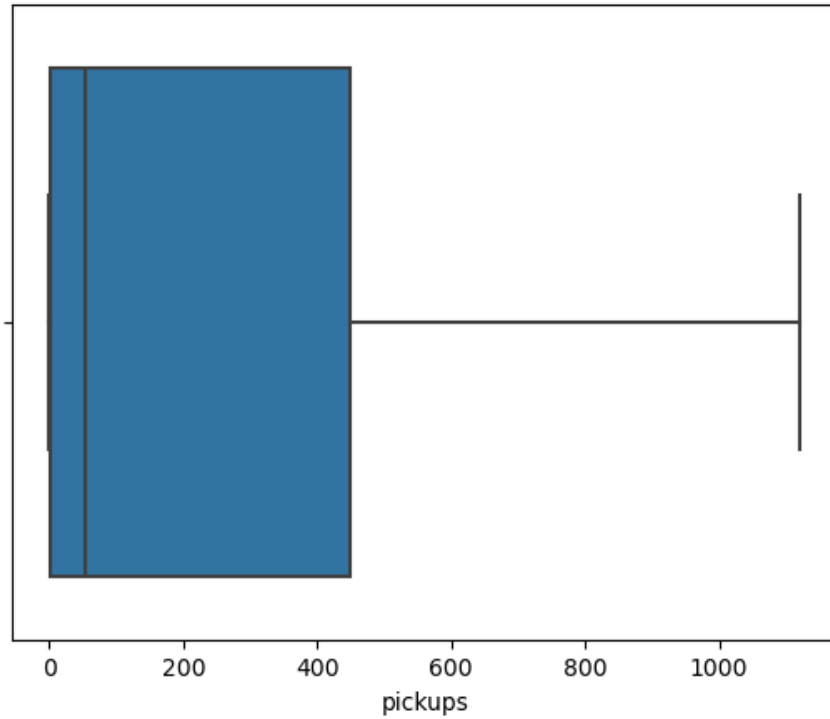
Observation

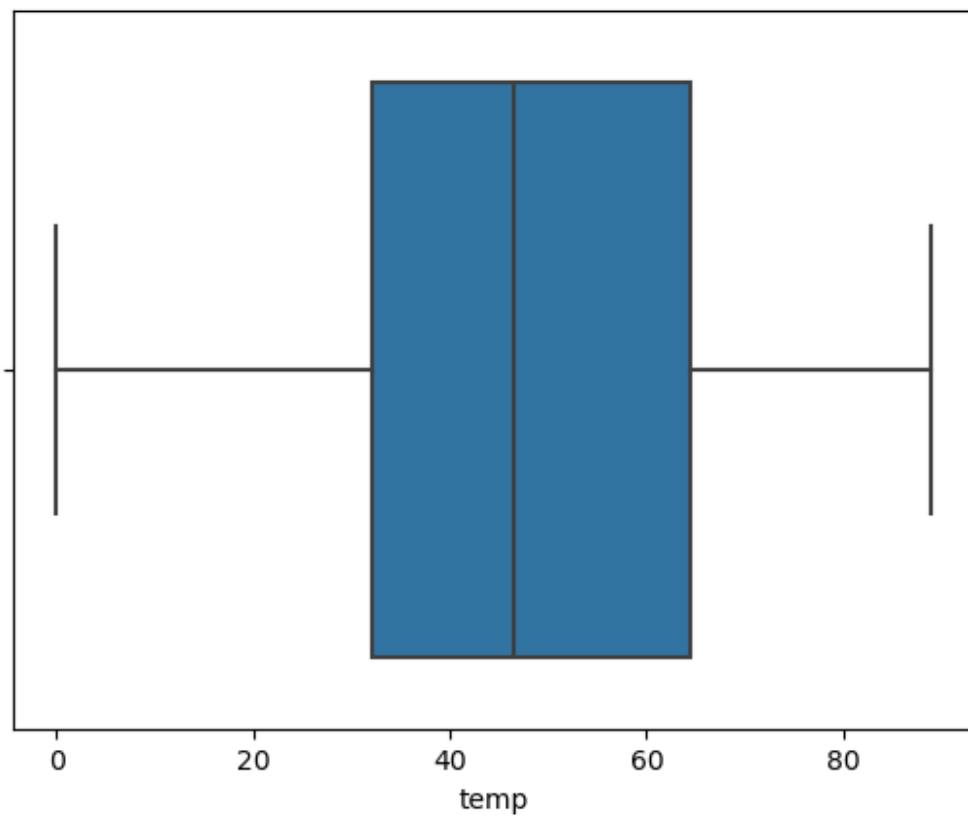
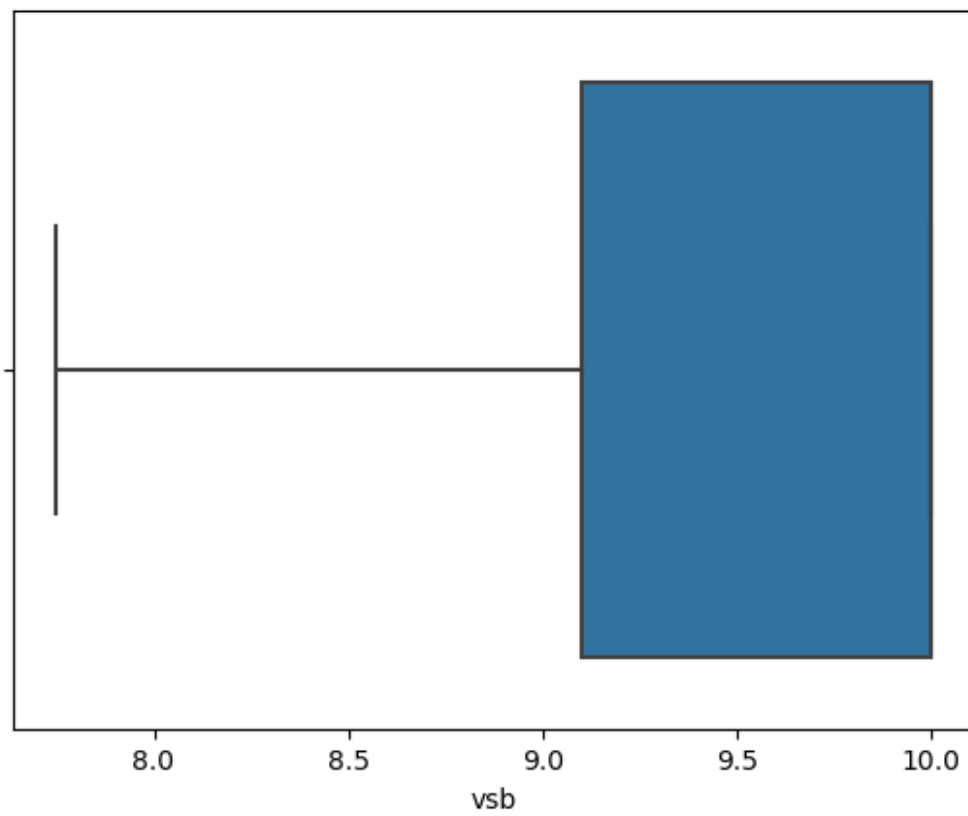
- hday contains contains '?'
- hday at row 1 and 123 needs further checking.

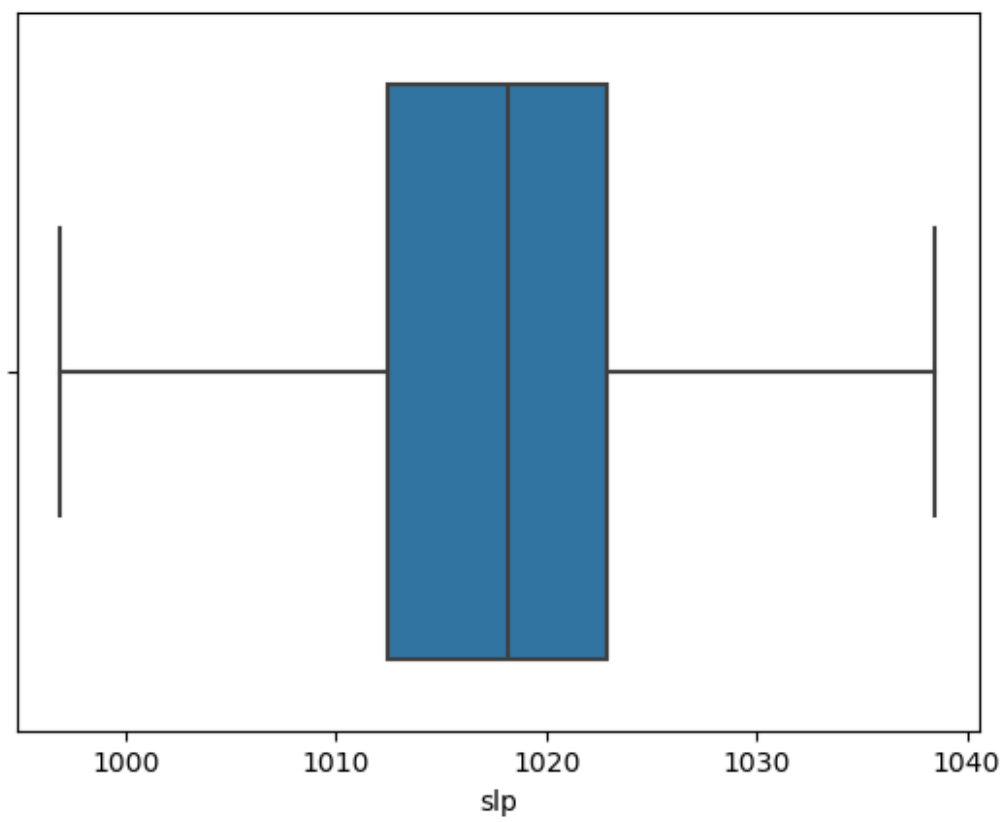
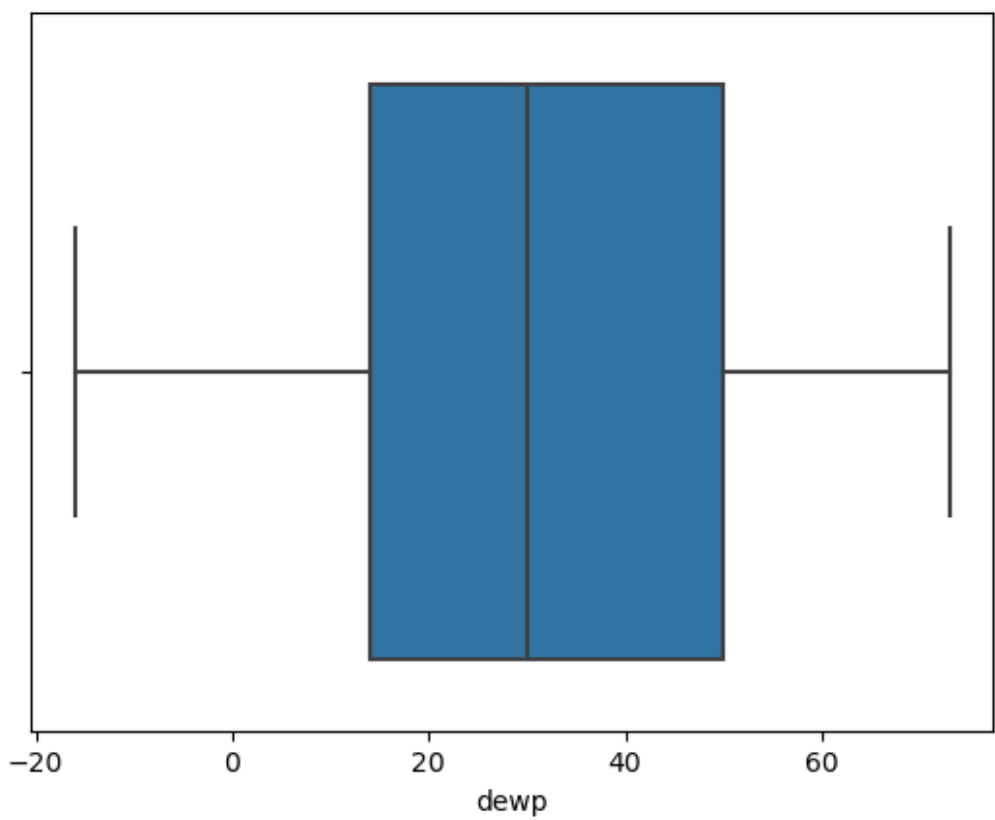
Outliers and their authenticity

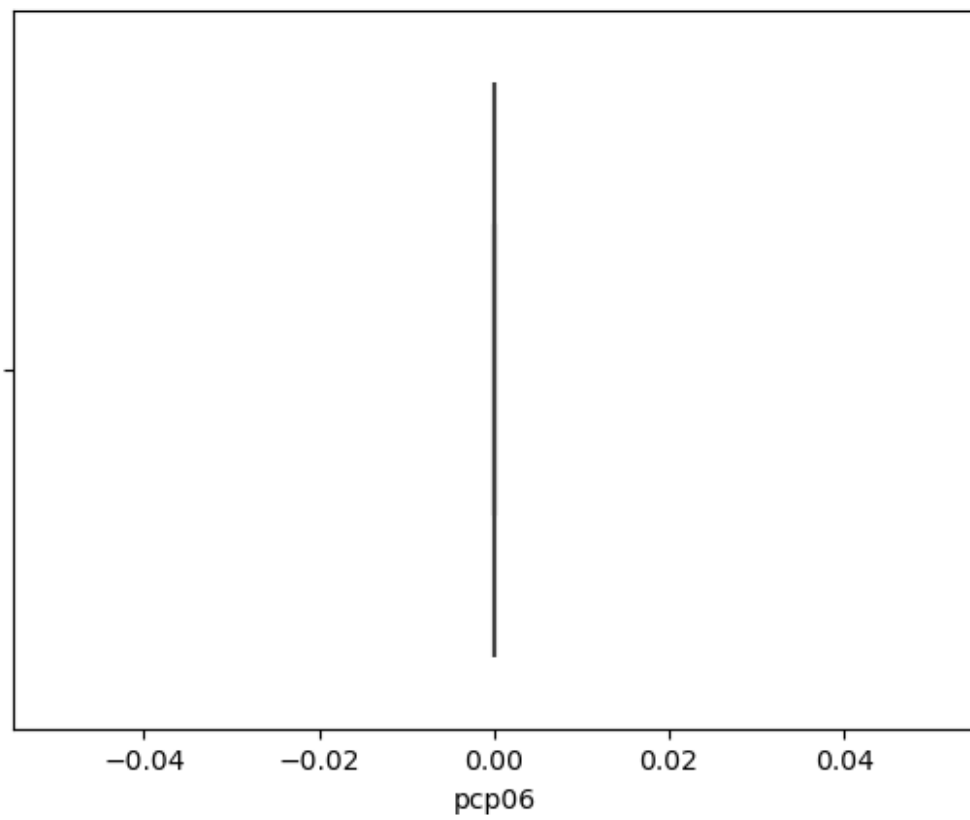
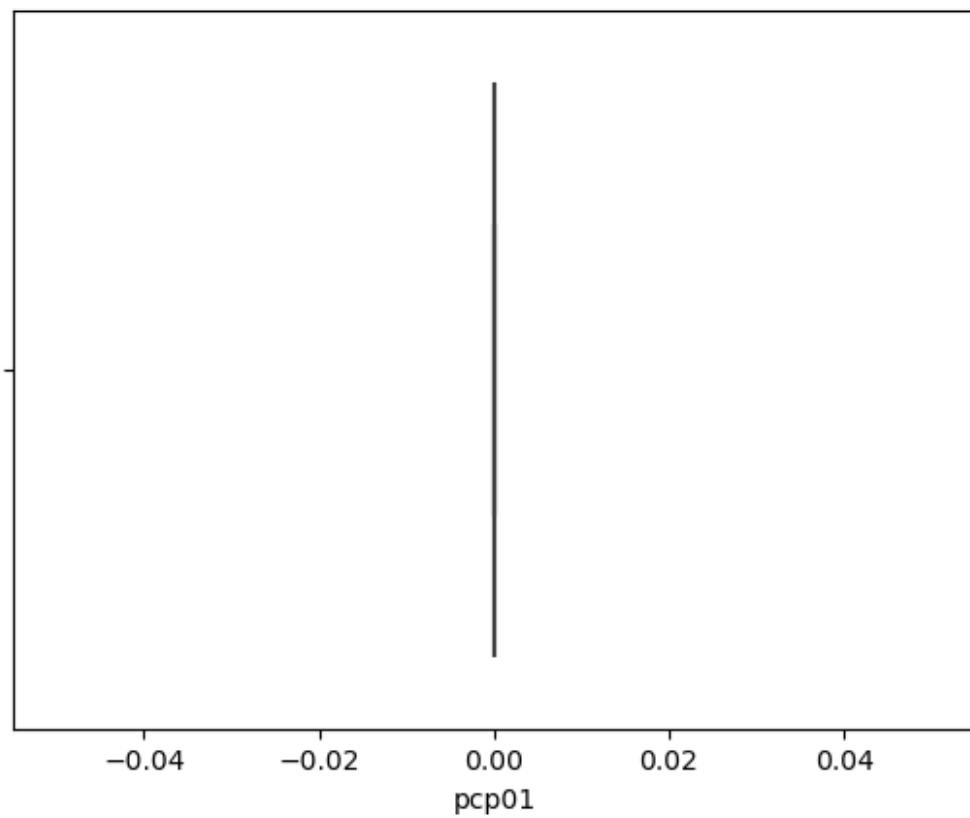
Observation

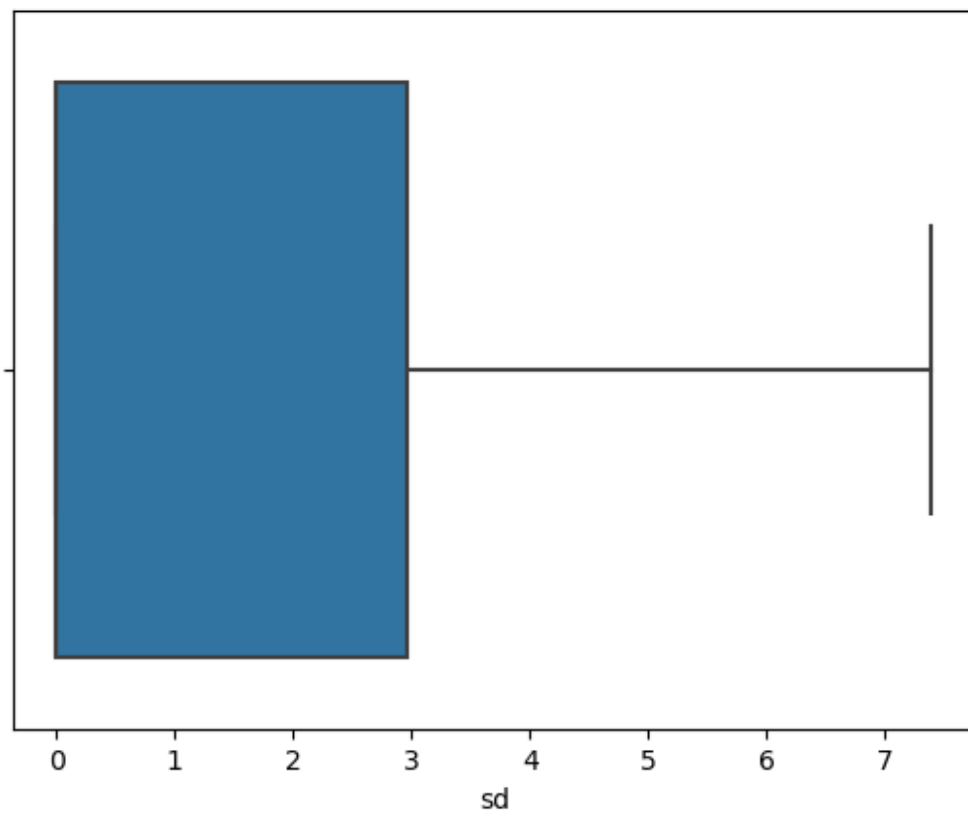
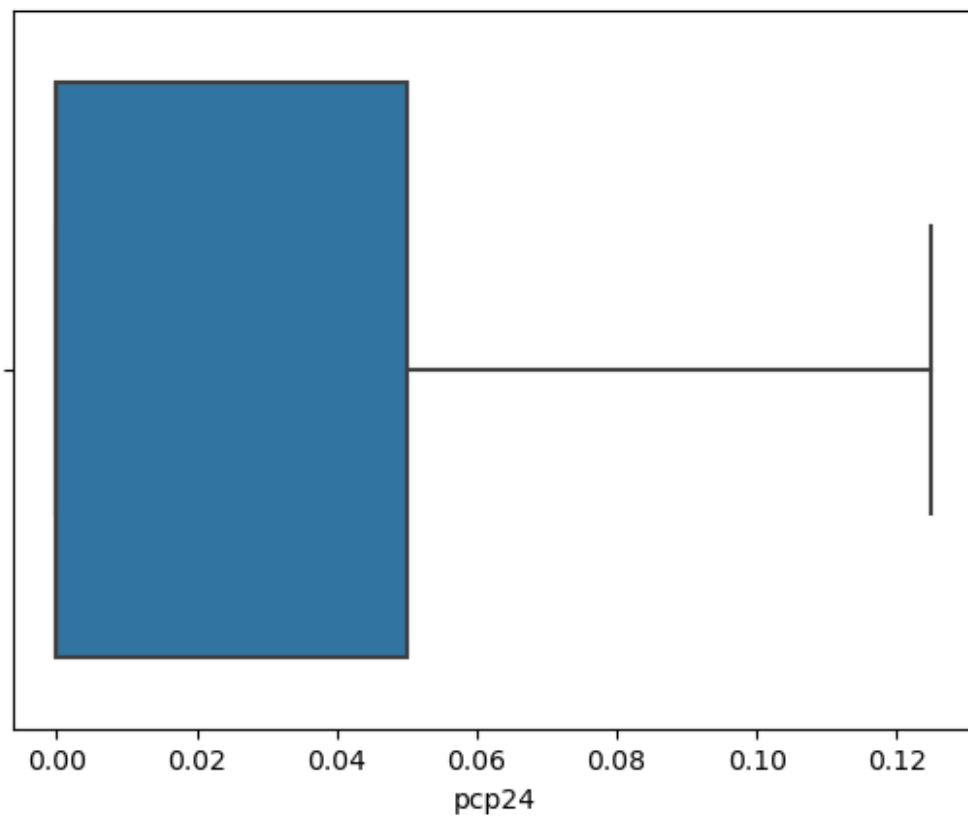
- After removing anomalies and converting datatypes of each column with wrong entries, the outliers representation through box plot are











Cleaning Data

Observation

- Updated '?' to null value in columns- hday
- Null values of columns borough replaced by mode
- Null values of columns temp replaced by median
- Outliers of every column has been removed

Question 1

Pickup Analysis

- What is the total number of Uber pickups across all boroughs?
- Which borough has the highest average number of hourly pickups?
- How do the number of pickups vary across different hours of the day?
- Which day of the week has the highest number of pickups?

Observation

- Total number of Uber pickups: 8207691.0
- Borough with the highest average number of hourly pickups: Manhattan
- Number of pickups by hour:

pickup_dt	
1/1/2015 10:00	654.0
1/1/2015 11:00	903.0
1/1/2015 12:00	1147.0
1/1/2015 13:00	1361.0
1/1/2015 14:00	1578.0
1/1/2015 15:00	1738.0
1/1/2015 16:00	1807.0
1/1/2015 17:00	1845.0
1/1/2015 18:00	1862.0
1/1/2015 19:00	1850.0
- Day of the week with the highest number of pickups: N

Questions 2

Weather Impact

- What is the correlation between temperature and the number of pickups?
- How does visibility impact the number of pickups?
- Is there a relationship between wind speed and the number of pickups?
- How does precipitation (1-hour, 6-hour, 24-hour) affect the number of pickups?

Observation

- A correlation coefficient of 0.06 suggests a very weak positive correlation between temperature and the number of pickups
- A correlation coefficient of -0.002 indicates a very weak and almost negligible negative correlation between visibility (vsb) and the number of pickups (pickups).
- A correlation coefficient of -0.005 indicates a very weak and almost negligible negative correlation between wind speed (spd) and the number of pickups (pickups).
- pcp01 and pcp06: NaN values mean no data available, so no correlation calculated.
- pcp24: Has a weak negative correlation with pickups, suggesting a minimal impact of 24-hour precipitation on the number of Uber pickups.

Questions 3

Seasonal Trends

- What is the average number of pickups during holidays compared to non-holidays?
- How does snow depth influence the number of pickups?

Observation

- The average number of pickups during holidays compared to non-holidays is:
hday
N 282.180652
Y 279.061773
- A correlation of -0.025 indicates a very weak negative relationship between snow depth (sd) and the number of pickups, suggesting that snow depth has little to no influence on the number of pickups in this dataset.

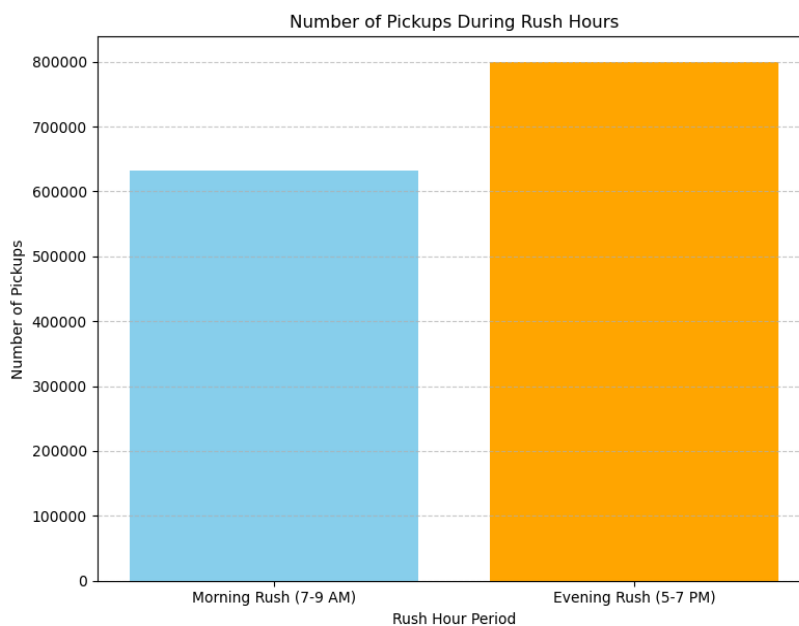
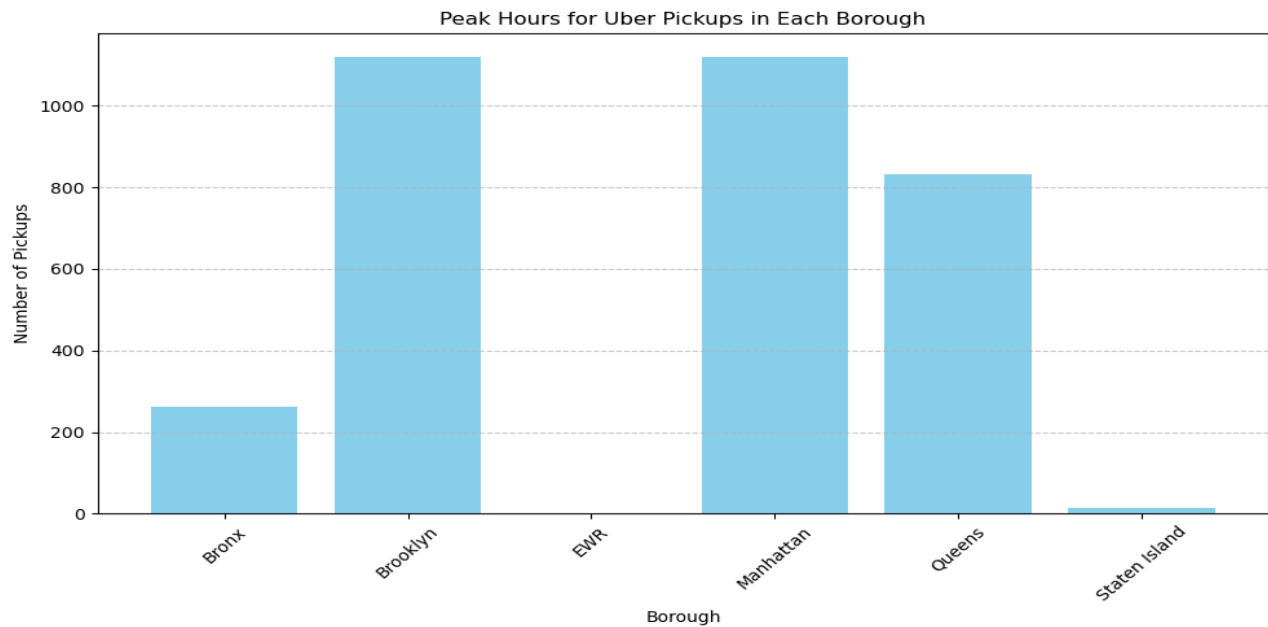
Question 4

Hourly Trends

- What are the peak hours for Uber pickups in each borough?
- How do the number of pickups change during rush hours (e.g., 7-9 AM, 5-7 PM)?
- What is the average number of pickups during late-night hours (e.g., 12 AM - 4 AM)?

Observation

- Average number of pickups during late-night hours (12 AM - 4 AM): 243.3
1617958492166



Question 5

Borough Comparison

- How do pickup trends differ between boroughs during different weather conditions?
- Which borough shows the highest increase in pickups during holidays?

Observation

- Pickup trends of pickups due to weather is:

	borough	hday	temp	dewp	pcp01	pcp06	pcp24	sd	pickups
0	Bronx	N	2.0	-16.0	0.0	0.0	0.0900	7.395833	27.200000
1	Bronx	N	3.0	-16.0	0.0	0.0	0.0900	7.395833	24.000000
2	Bronx	N	3.0	-15.0	0.0	0.0	0.0900	7.395833	36.000000
3	Bronx	N	4.0	-15.0	0.0	0.0	0.0900	7.395833	42.333333
4	Bronx	N	4.0	-11.0	0.0	0.0	0.0525	7.395833	38.000000
5	Bronx	N	5.0	-16.0	0.0	0.0	0.0900	7.395833	31.500000
6	Bronx	N	5.0	-11.0	0.0	0.0	0.0000	7.395833	24.000000
7	Bronx	N	5.0	-10.0	0.0	0.0	0.0075	7.395833	27.500000
8	Bronx	N	6.0	-15.0	0.0	0.0	0.0900	7.395833	19.500000
9	Bronx	N	6.0	-13.0	0.0	0.0	0.0900	7.395833	58.000000

- borough shows the highest increase in pickups during holidays: 'Queens'

Question 6

Weather Extremes

- How do extreme weather conditions (e.g., very high or very low temperatures, heavy rainfall, snowstorms) affect the number of pickups?
- What is the impact of visibility less than 1 mile on the number of pickups?

Observation

- Percentage of pickups during extreme weather conditions: 11.11%
- Total pickups during visibility less than 1 mile: 0.0

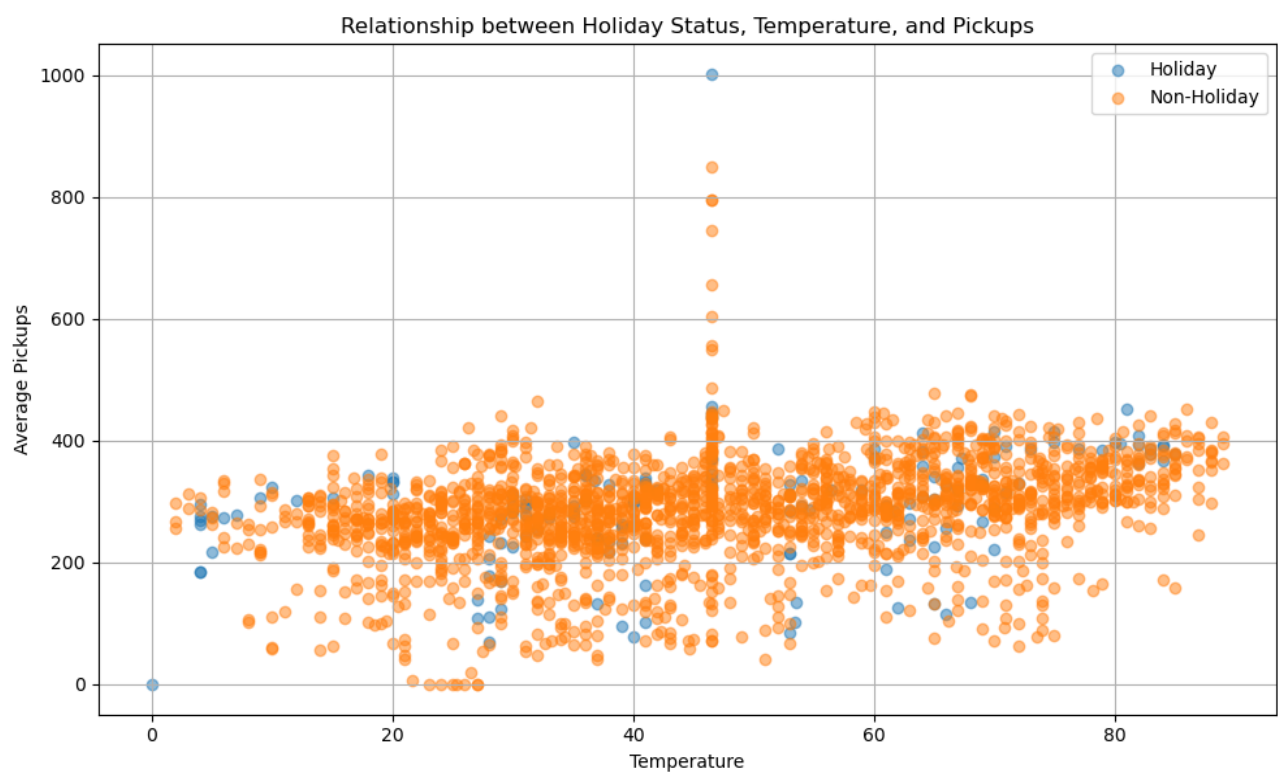
Question 7

Data Correlations

- Is there a correlation between sea level pressure and the number of pickups?
- How do different weather variables (temperature, dew point, wind speed, visibility) collectively impact the number of pickups?

Observation

- The correlation coefficient of -0.01 suggests a very weak negative correlation between sea level pressure (slp) and the number of pickups



Question 8

Growth Insights

- Which weather conditions are most favorable for Uber pickups, and how can this information be used to optimize driver availability?
- Based on the data, what recommendations can be made to Uber to increase pickups during low-demand periods?

Observation

- The temperature between 30 to 70 is the favorable temperature for pickup
- From this data we can say
 1. To increase pickups during low-demand periods, Uber can:
 2. Implement targeted promotions and discounts during late-night hours (12 AM - 4 AM).
 3. Consider surge pricing or incentives for drivers during extreme weather conditions to increase availability.
 4. Offer promotional discounts during temperatures outside the favorable range of 30 to 70 degrees Fahrenheit.
 5. Adjust driver incentives to increase pickups during mid-morning and late-night hours.
 6. Use borough-specific strategies to target areas with lower demand.

