Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	739943
Project Title	Frappe Activity:mobile Phone Activity classification
Maximum Marks	6 Marks

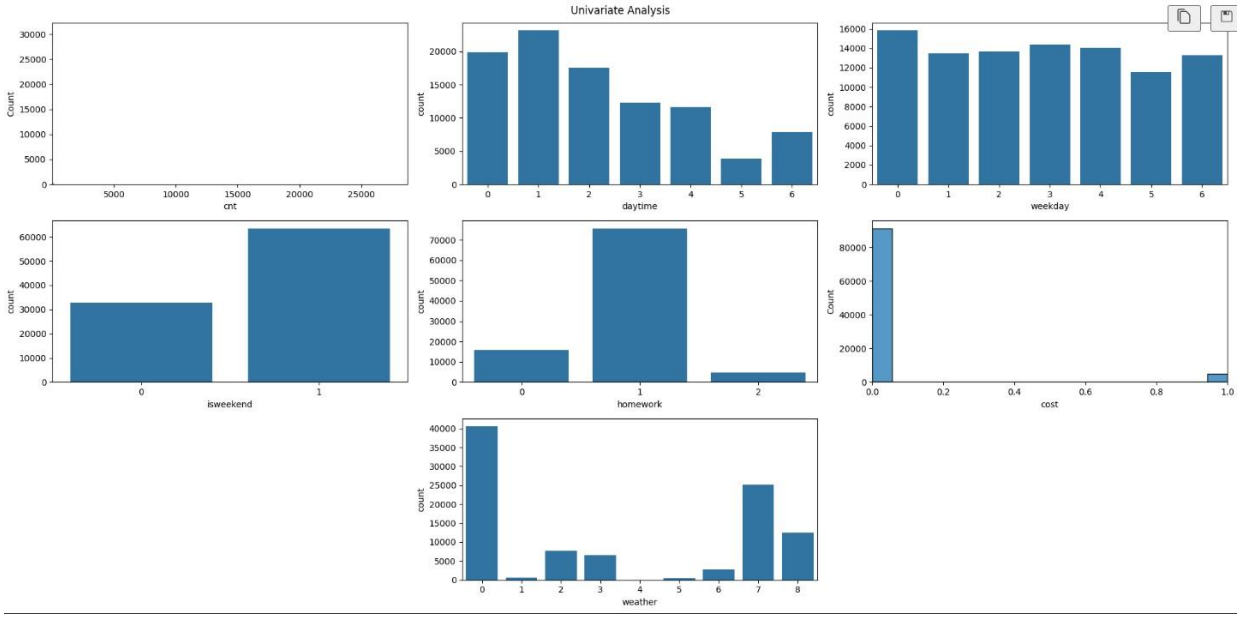
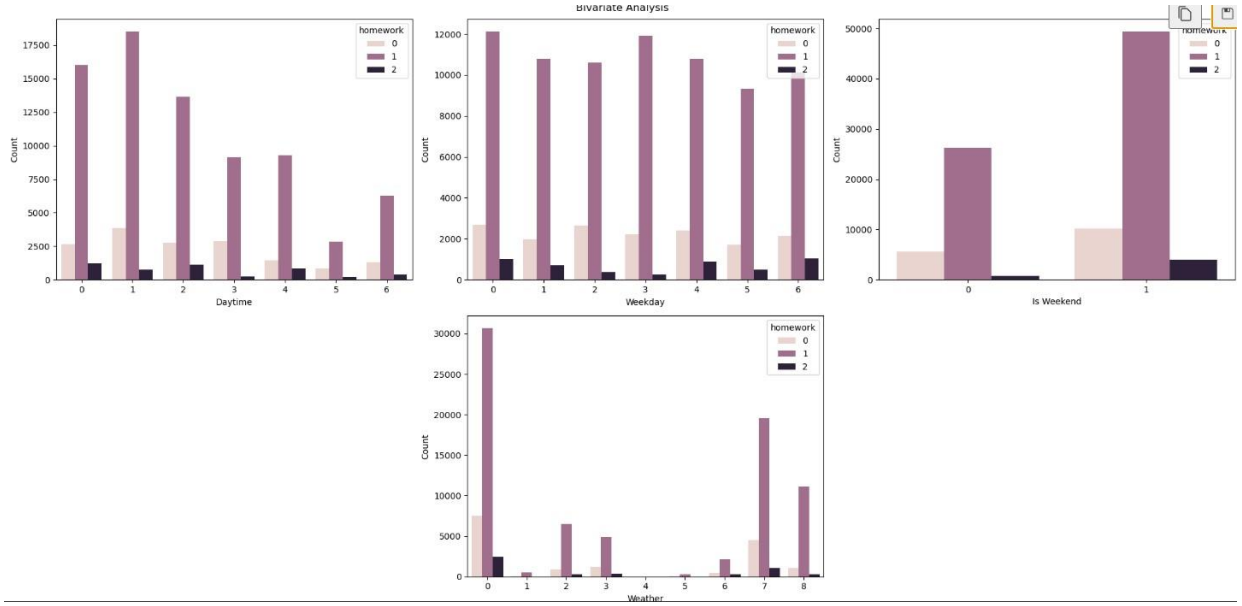
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will

Section	Description
Data Overview	Dimension: 923 rows × 49 columns Descriptive statistics: 
Univariate Analysis	

	user	item	cnt	daytime	weekday	isweekend	homework	cost	weather	city	name
count	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000	96203.000000
mean	189.646944	386.754207	81.330894	2.165317	2.885627	0.659792	0.885565	0.050872	3.432835	333.966176	1800.068990
std	156.458370	683.779627	395.414980	1.825046	2.009485	0.473781	0.447594	0.219737	3.380435	359.977130	1041.000073
min	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	74.000000	21.000000	2.000000	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1037.000000
50%	153.000000	77.000000	8.000000	2.000000	3.000000	1.000000	1.000000	0.000000	2.000000	327.000000	1489.000000
75%	271.000000	397.000000	36.000000	3.000000	5.000000	1.000000	1.000000	0.000000	7.000000	649.000000	2960.000000
max	956.000000	4081.000000	28752.000000	6.000000	6.000000	1.000000	2.000000	1.000000	8.000000	1087.000000	3508.000000

address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

	
Bivariate Analysis	

Outliers and Anomalies	-

Data Preprocessing Code Screenshots

Loading Data

```
df=df.merge(meta_app,on= 'item')

df.head()
```

	user	item	cnt	daytime	weekday	isweekend	homework	cost	weather	country	city	name
0	0	0	1	morning	sunday	weekend	unknown	free	sunny	United States	0	Any.DO To-do & Tasks List
1	1	1	7	afternoon	saturday	weekend	unknown	free	cloudy	Spain	0	Yahoo!
2	2	2	6	evening	monday	workday	unknown	free	cloudy	Spain	369	Compass PRO
3	3	3	1	sunset	thursday	workday	unknown	free	unknown	United States	1028	Instagram
4	4	4	428	night	thursday	workday	home	free	sunny	Switzerland	147	Shopping List

Handling Missing Data

```
df.isna().sum()
```

```
user          0
item          0
cnt           0
daytime       0
weekday       0
isweekend     0
homework      0
cost          0
weather       0
country       0
city          0
name          0
dtype: int64
```

Data Transformation

```
print(df['name'])
```

```
0         260
1        3217
2         698
3        1489
4        2523
         ...
96198     260
96199    1037
96200    1994
96201    1307
96202    2429
Name: name, Length: 96203, dtype: int32
```

Feature Engineering	Attached the codes in final submission.
Save Processed Data	-