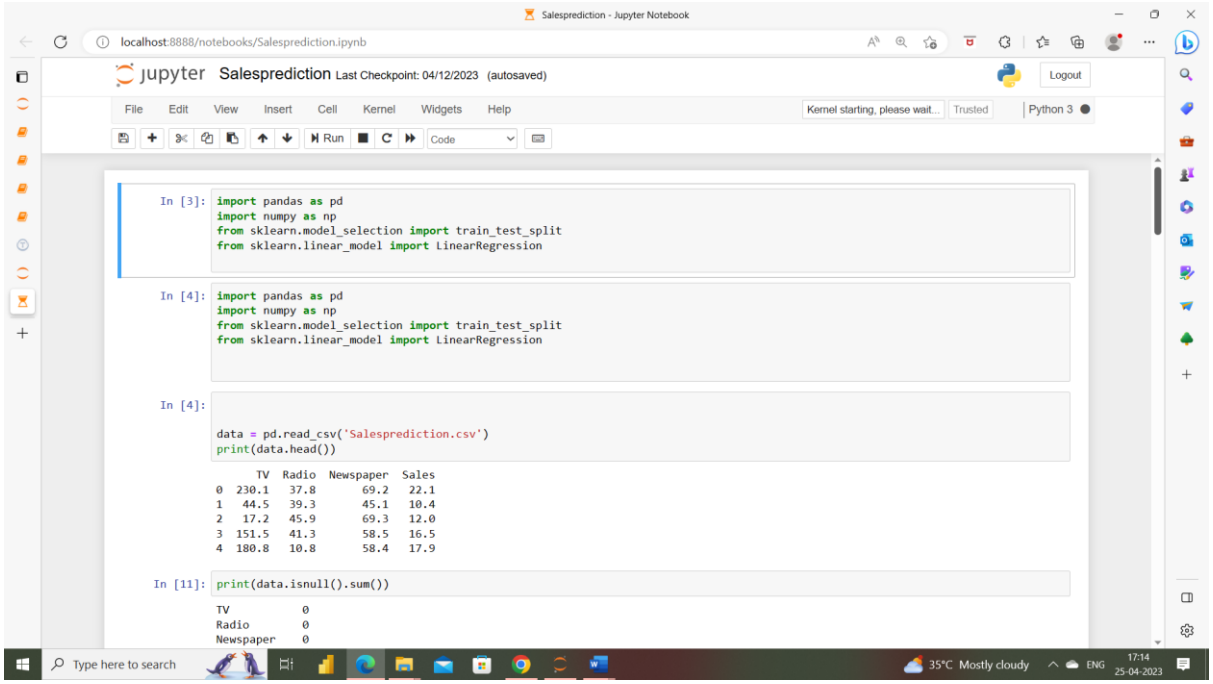
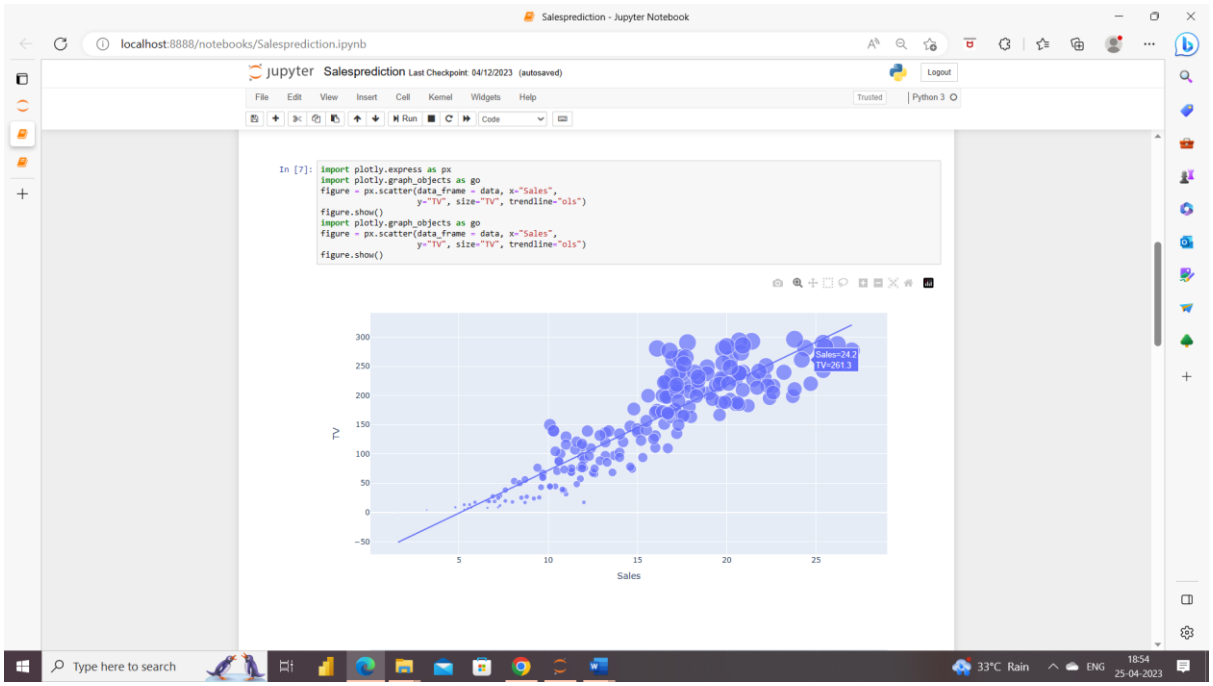
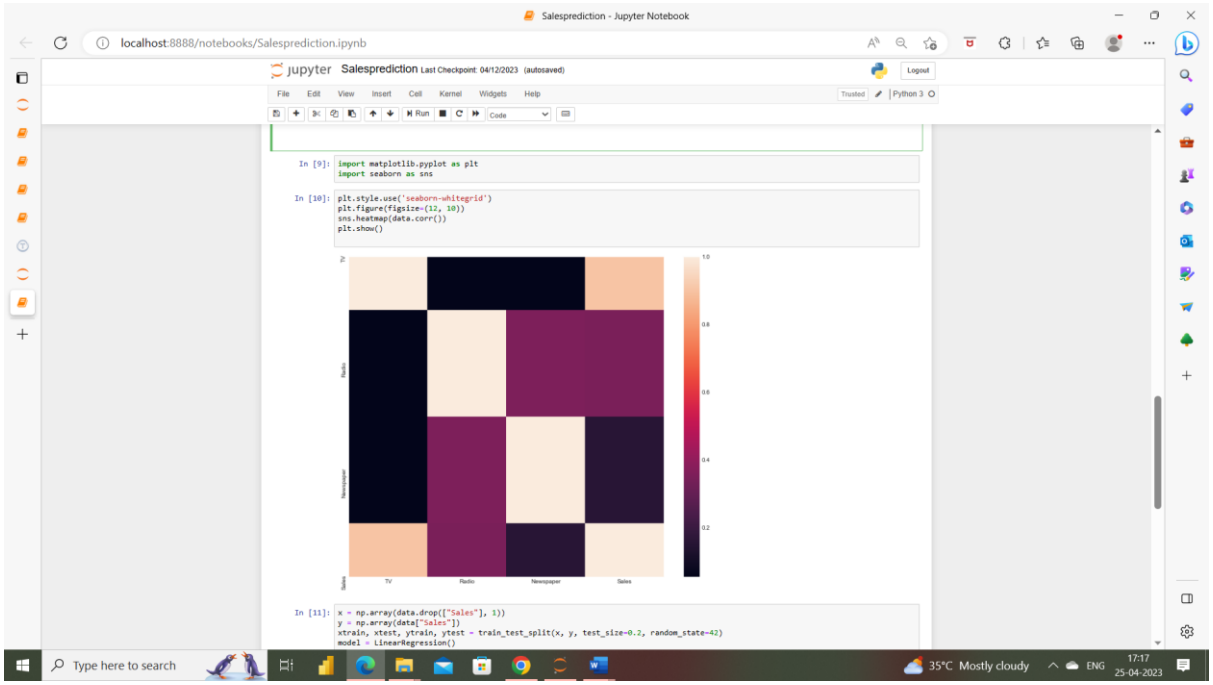
Python Projects Archive:Dr Ratika Datta

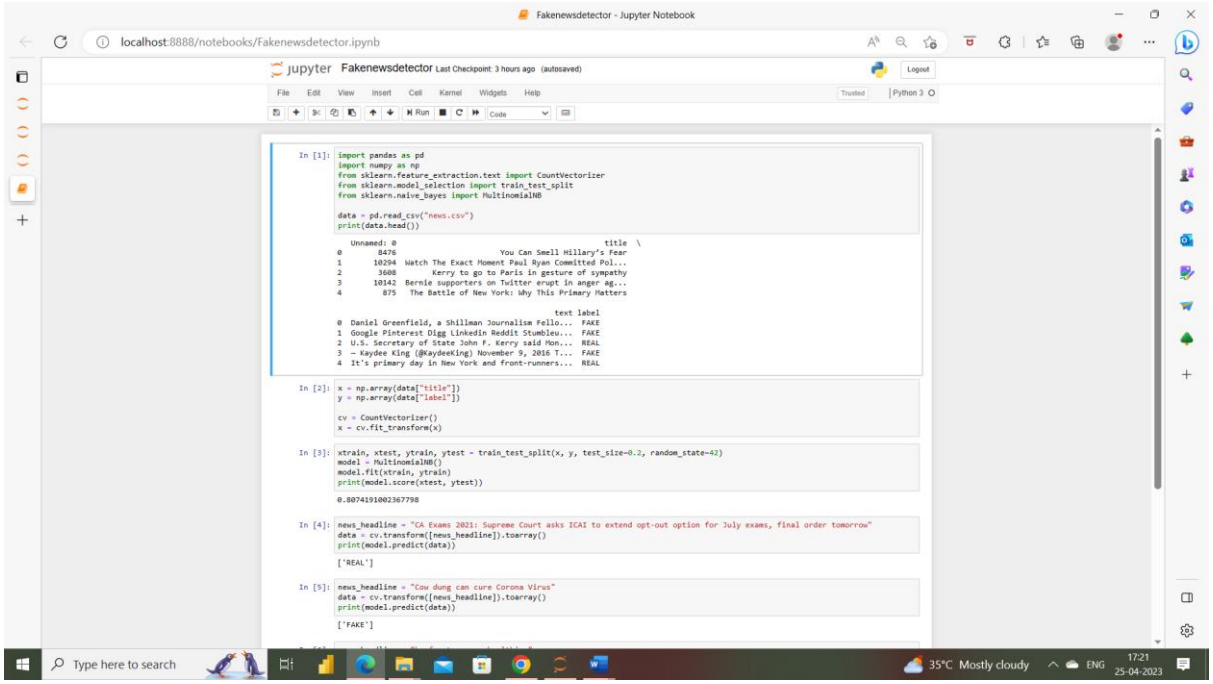
1.SalesPrediction Analysis



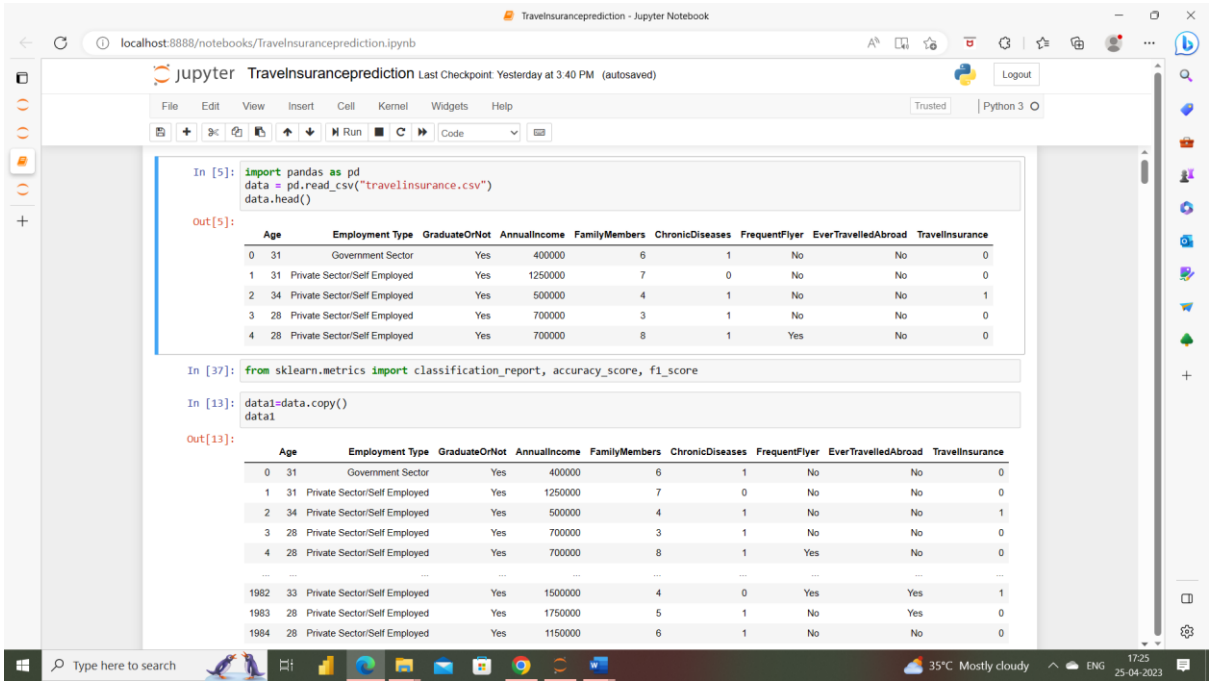


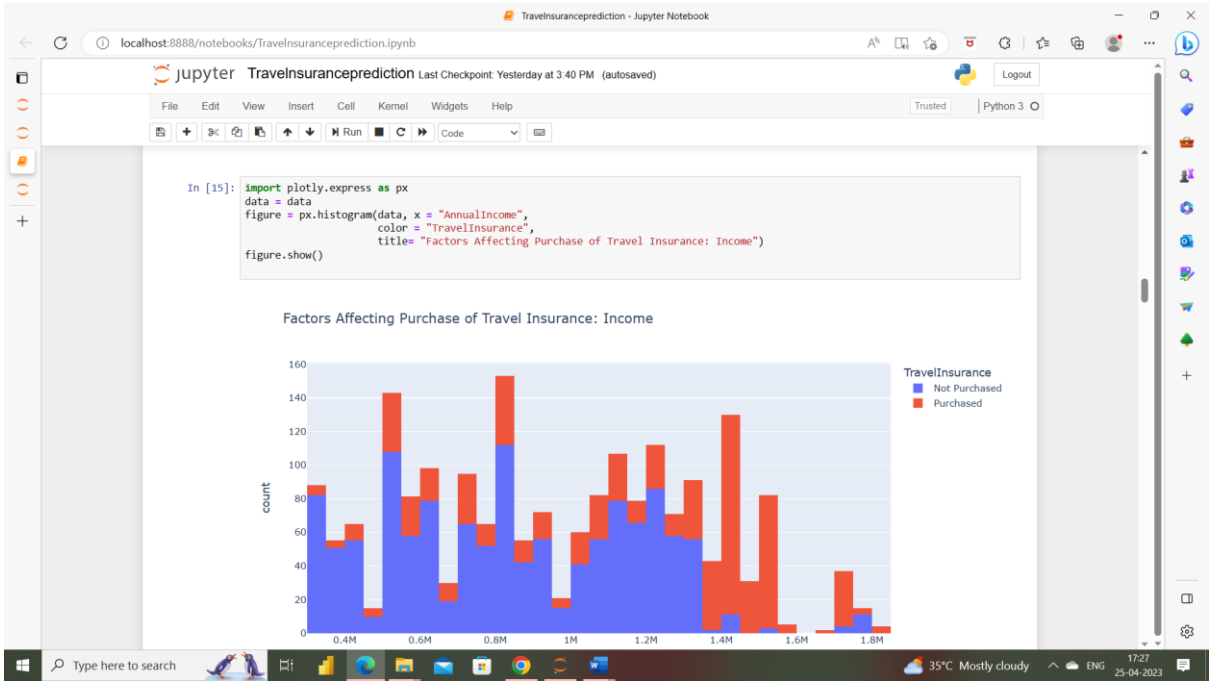


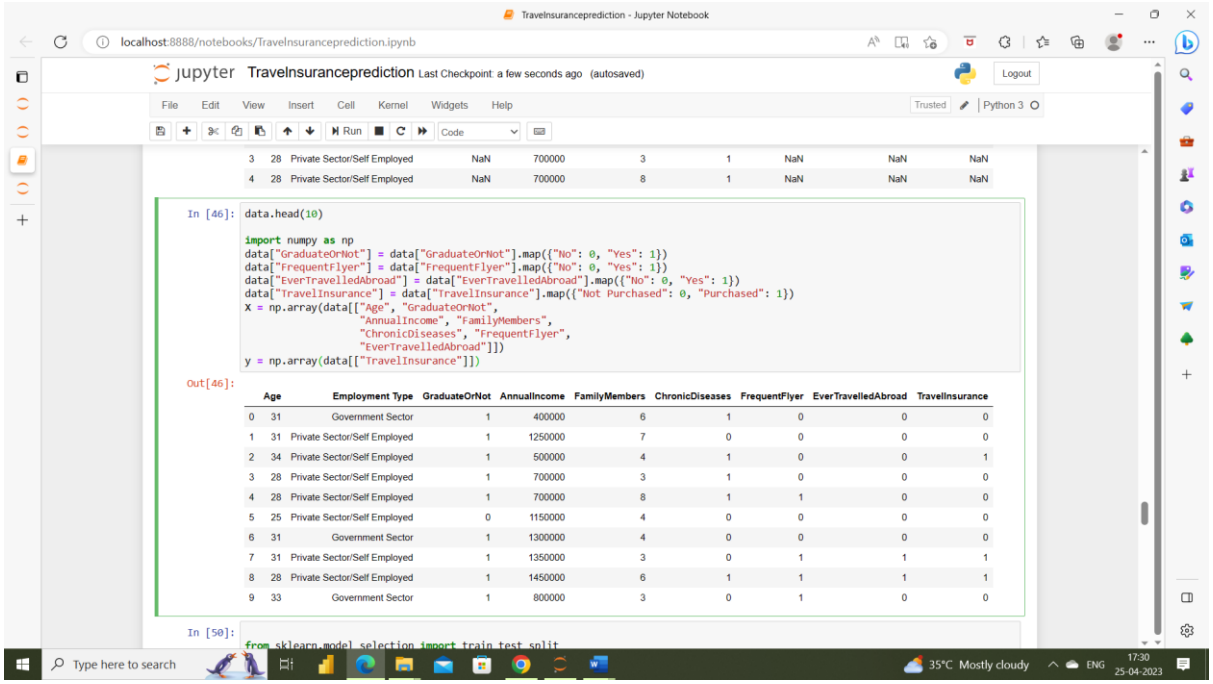
2. **Fake News Predictor**



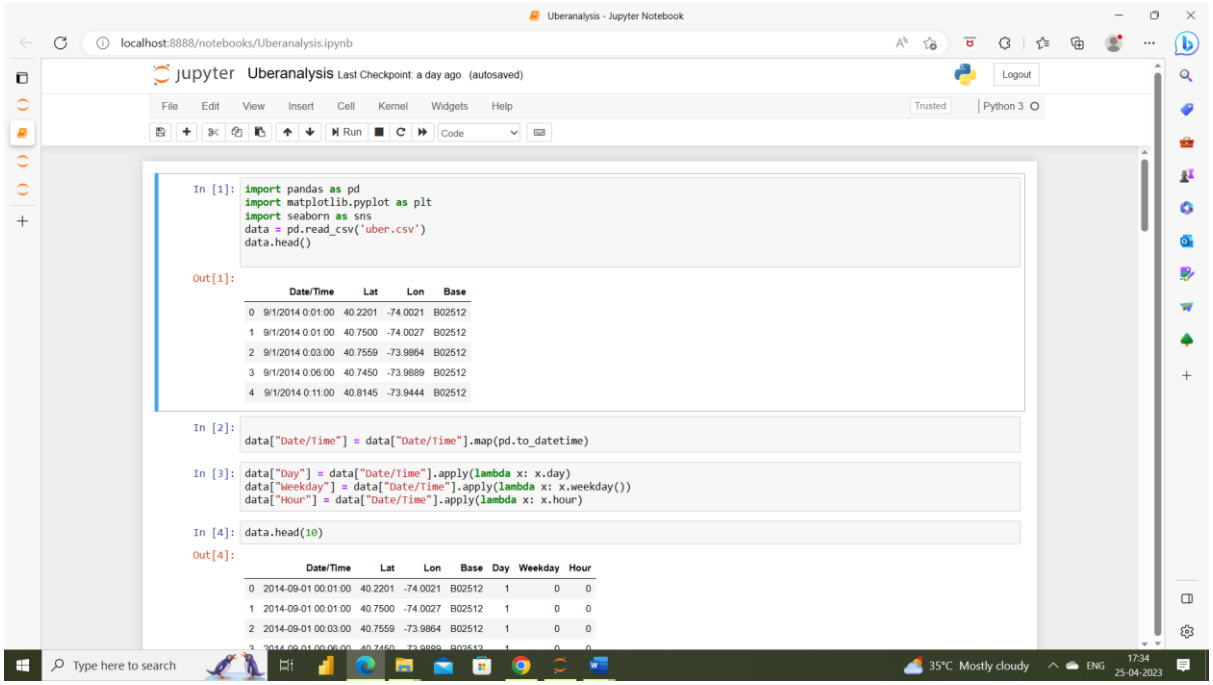
3. **Travel Insurance Prediction:**

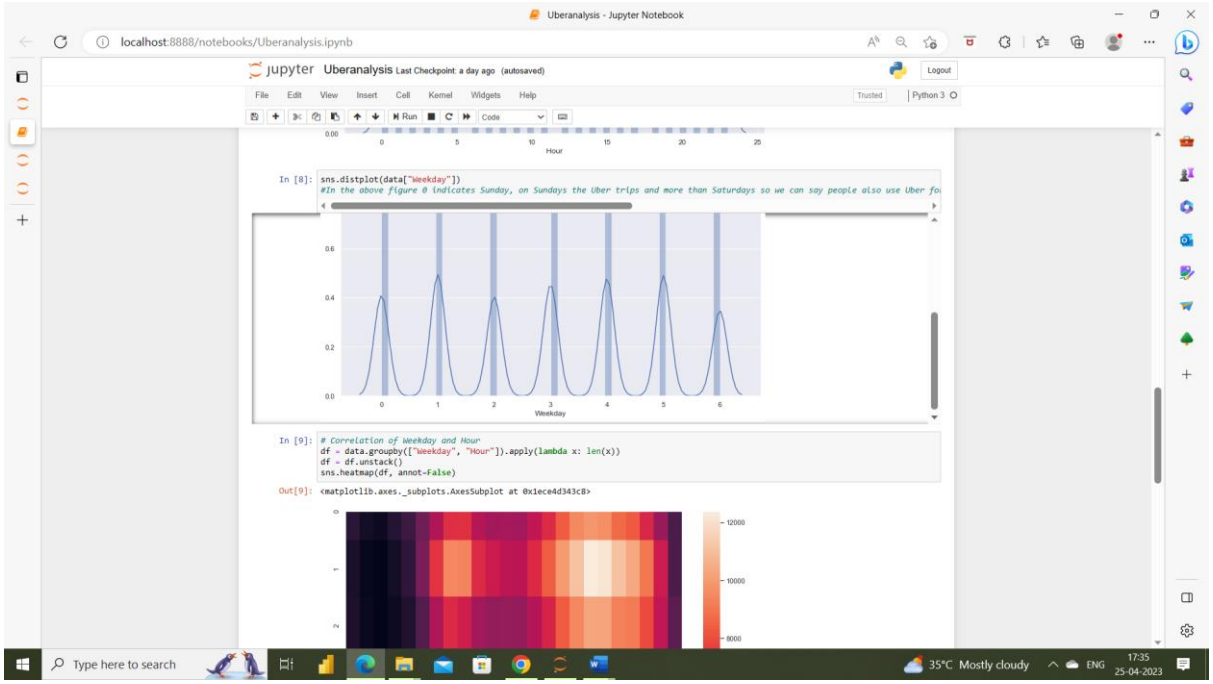


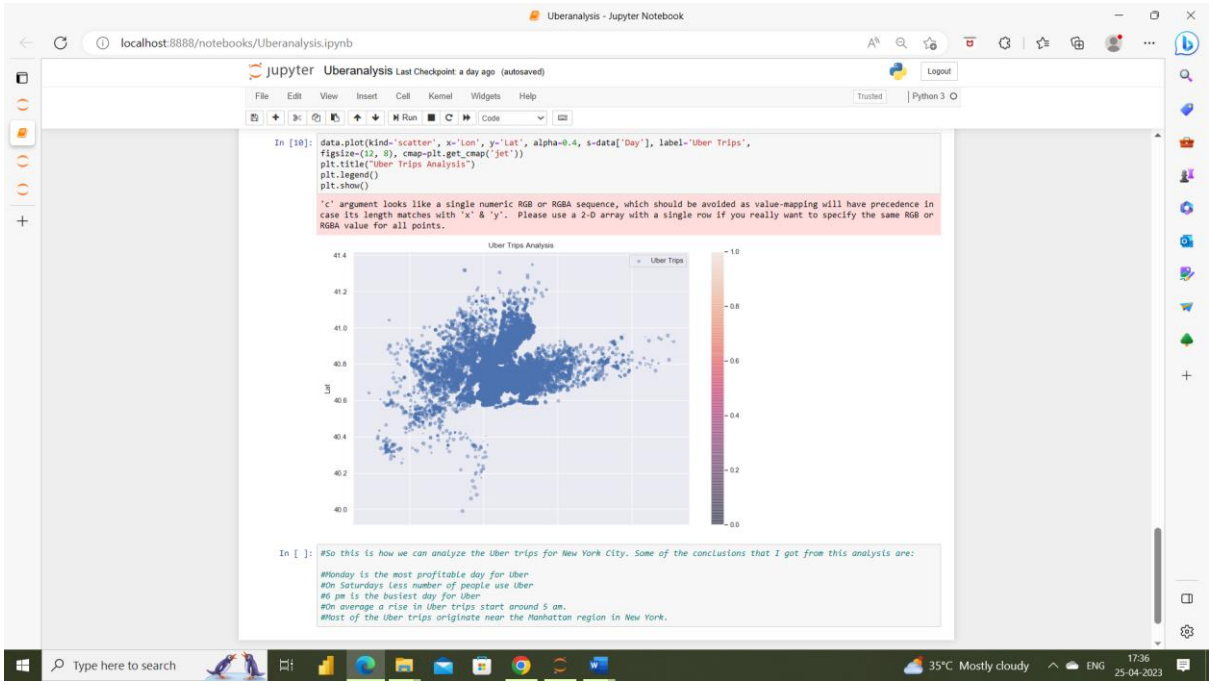




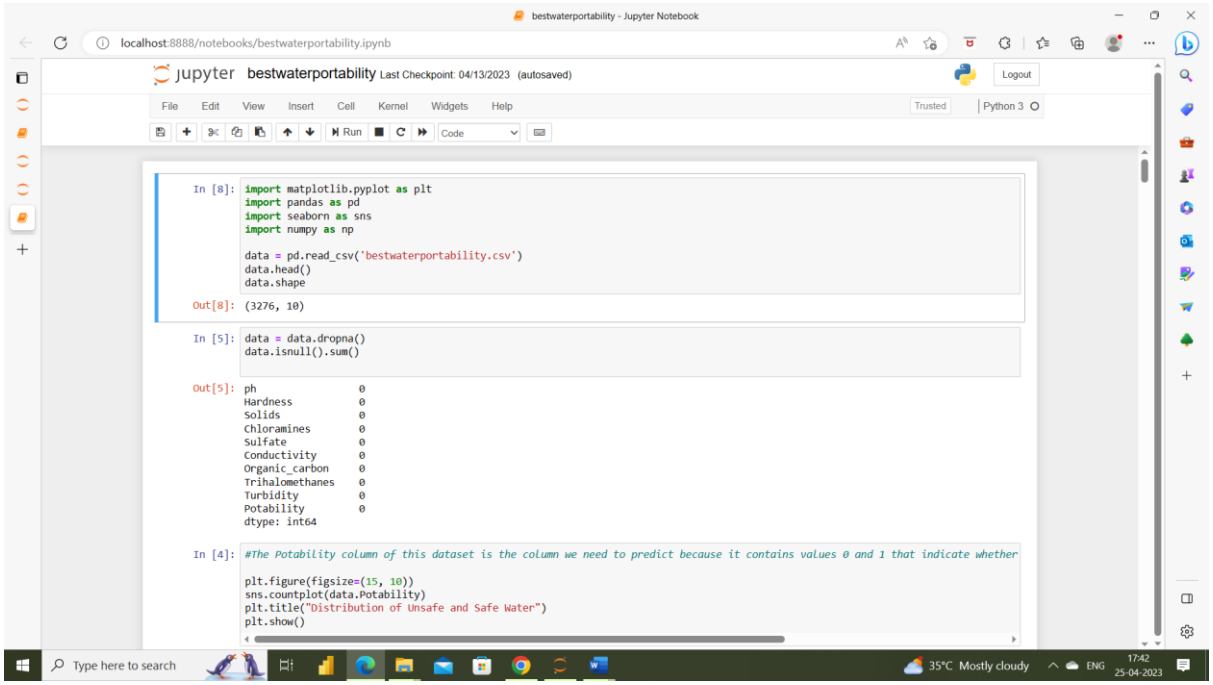
4. **Uber Trend Analysis for New York City Sales:**

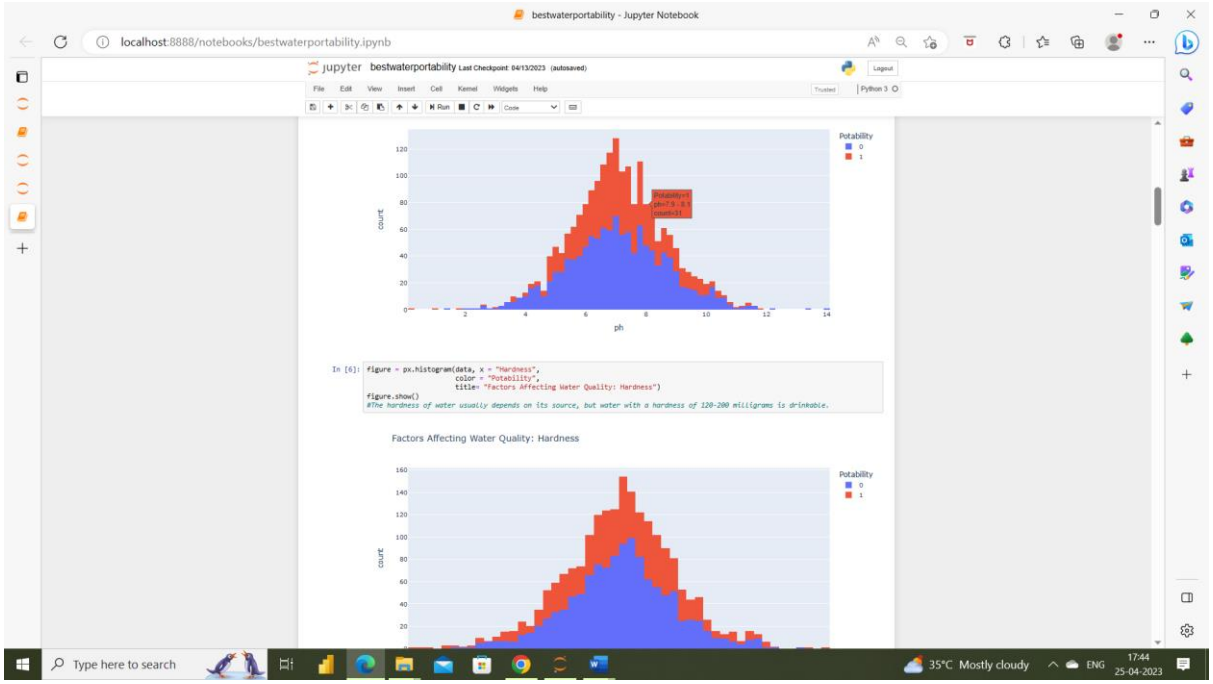



Uberanalysis - Jupyter Notebook
localhost:8888/notebooks/Uberanalysis.ipynb
jupyter Uberanalysis Last Checkpoint: a day ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
Hour
In [8]:
sns.distplot(data["Weekday"])
#In the above figure 0 indicates Sunday, on Sundays the Uber trips and more than Saturdays so we can say people also use Uber fo
Weekday
In [9]:
# Correlation of Weekday and Hour
df = data.groupby(["Weekday", "Hour"]).apply(lambda x: len(x))
df = df.unstack()
sns.heatmap(df, annot=False)
Out[9]:
<matplotlib.axes._subplots.AxesSubplot at 0x1ece4d343c8>
Type here to search
35°C Mostly cloudy
ENG
17:35
25-04-2023

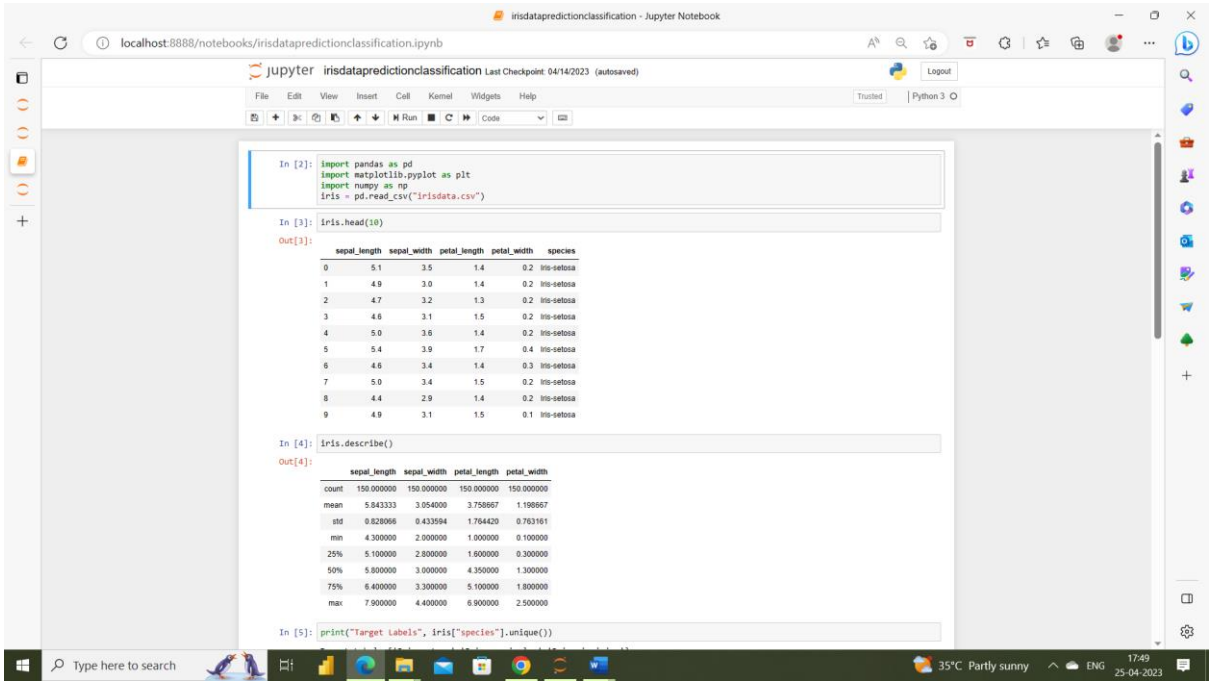

Uberanalysis - Jupyter Notebook
localhost:8888/notebooks/Uberanalysis.ipynb
jupyter Uberanalysis Last Checkpoint: a day ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [10]:
data.plot(kind='scatter', x='Lon', y='Lat', alpha=0.4, s=data['Day'], label='Uber Trips',
figsize=(12, 8), cmap=plt.get_cmap('jet'))
plt.title("Uber Trips Analysis")
plt.legend()
plt.show()
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
Uber Trips Analysis
Uber Trips
Lat
In [ ]:
#So this is how we can analyze the Uber trips for New York City. Some of the conclusions that I got from this analysis are:
#Monday is the most profitable day for Uber
#On Saturdays less number of people use Uber
#6 pm is the busiest day for Uber
#On average a rise in Uber trips start around 5 am.
#Most of the Uber trips originate near the Manhattan region in New York.
Type here to search
35°C Mostly cloudy
ENG
17:36
25-04-2023

5.Water Potability Analysis:





6.IRIS Classification Analysis



irisdatapredictionclassification - Jupyter Notebook

localhost:8888/notebooks/irisdatapredictionclassification.ipynb

jupyter irisdatapredictionclassification Last Checkpoint: 04/14/2023 (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

```
In [2]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np
        iris = pd.read_csv("irisdata.csv")
```

```
In [3]: iris.head(10)
```

Out[3]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

```
In [4]: iris.describe()
```

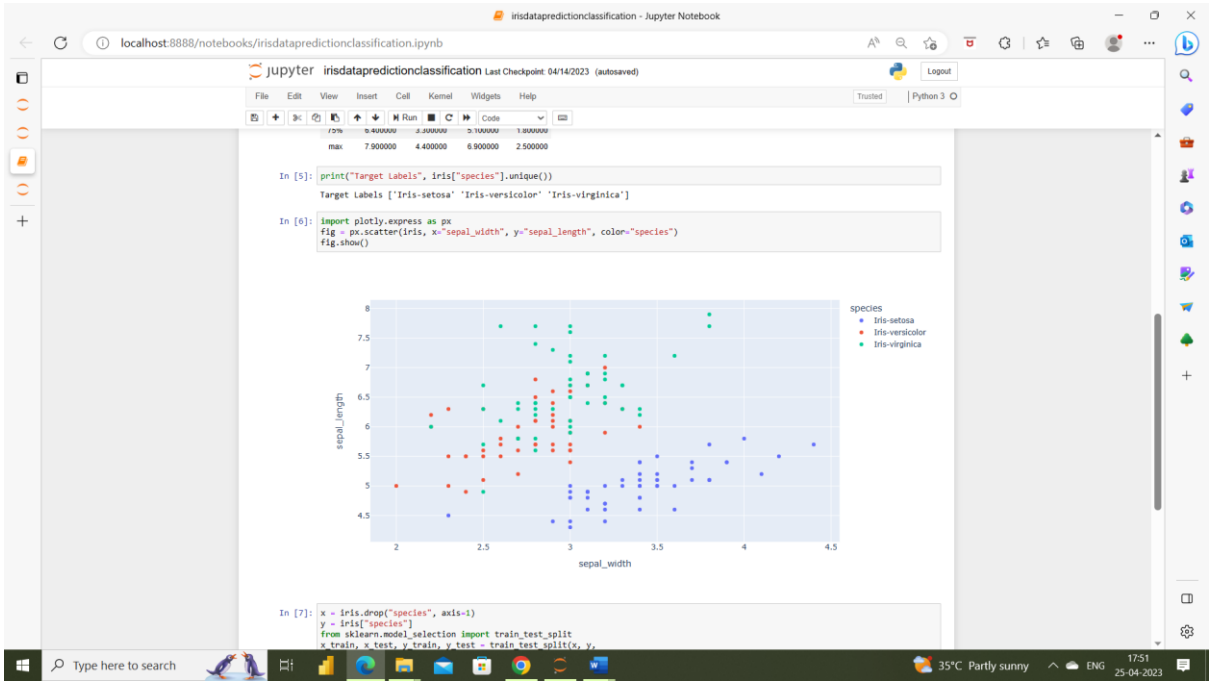
Out[4]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
In [5]: print("Target Labels", iris["species"].unique())
```

Type here to search

35°C Partly sunny ENG 17:49 25-04-2023



irisdatapredictionclassification - Jupyter Notebook

localhost:8888/notebooks/irisdatapredictionclassification.ipynb

jupyter irisdatapredictionclassification Last Checkpoint: 04/14/2023 (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

max	7.900000	4.400000	6.900000	2.500000

```
In [5]: print("Target Labels", iris["species"].unique())
```

```
Target Labels ['Iris-setosa' 'Iris-versicolor' 'Iris-virginica']
```

```
In [6]: import plotly.express as px
        fig = px.scatter(iris, x="sepal_width", y="sepal_length", color="species")
        fig.show()
```

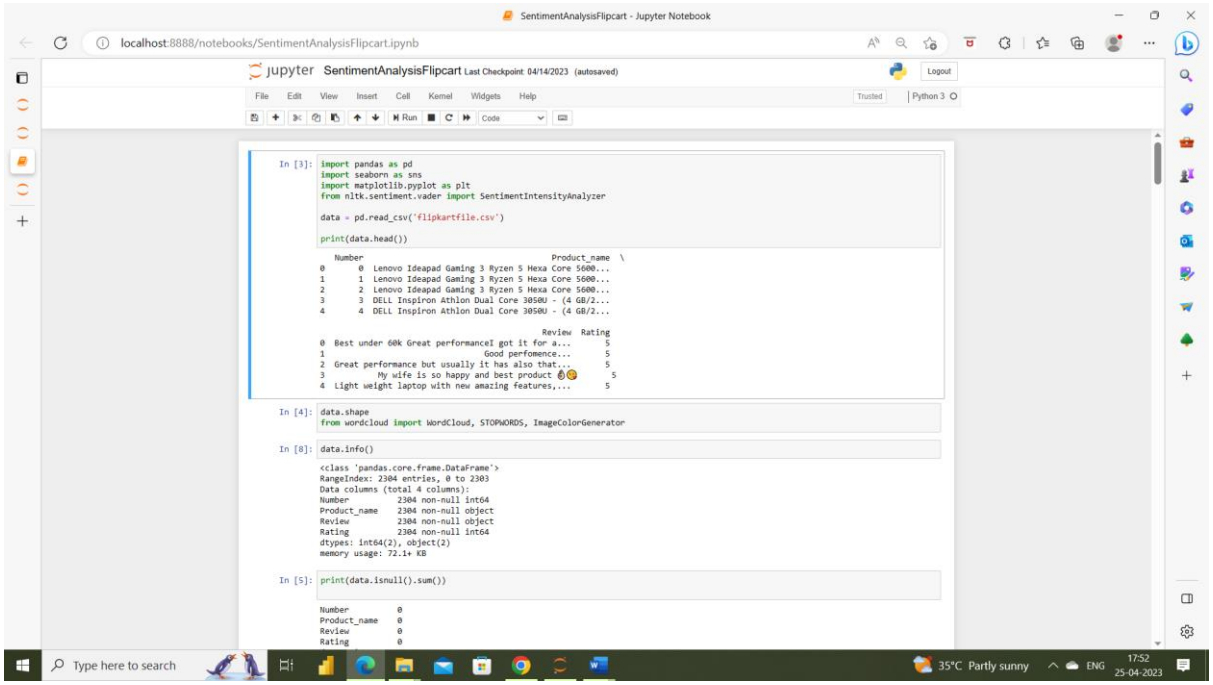


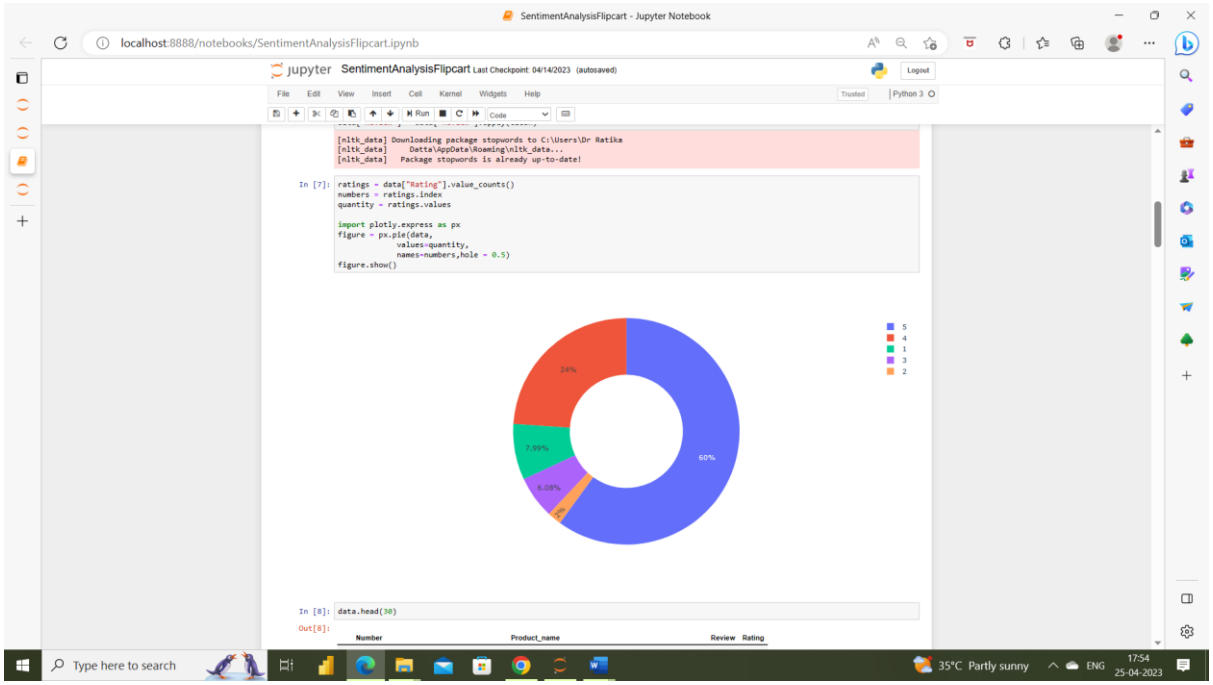
```
In [7]: x = iris.drop("species", axis=1)
        y = iris["species"]
        from sklearn.model_selection import train_test_split
        x_train, x_test, y_train, y_test = train_test_split(x, y,
```

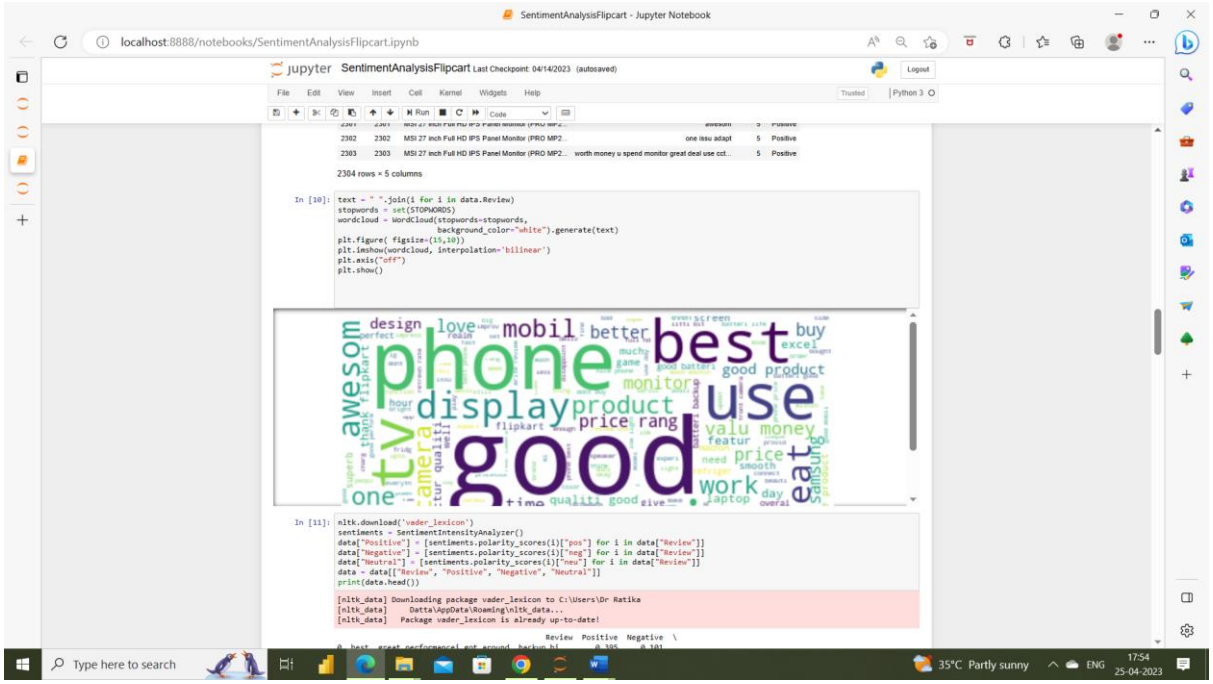
Type here to search

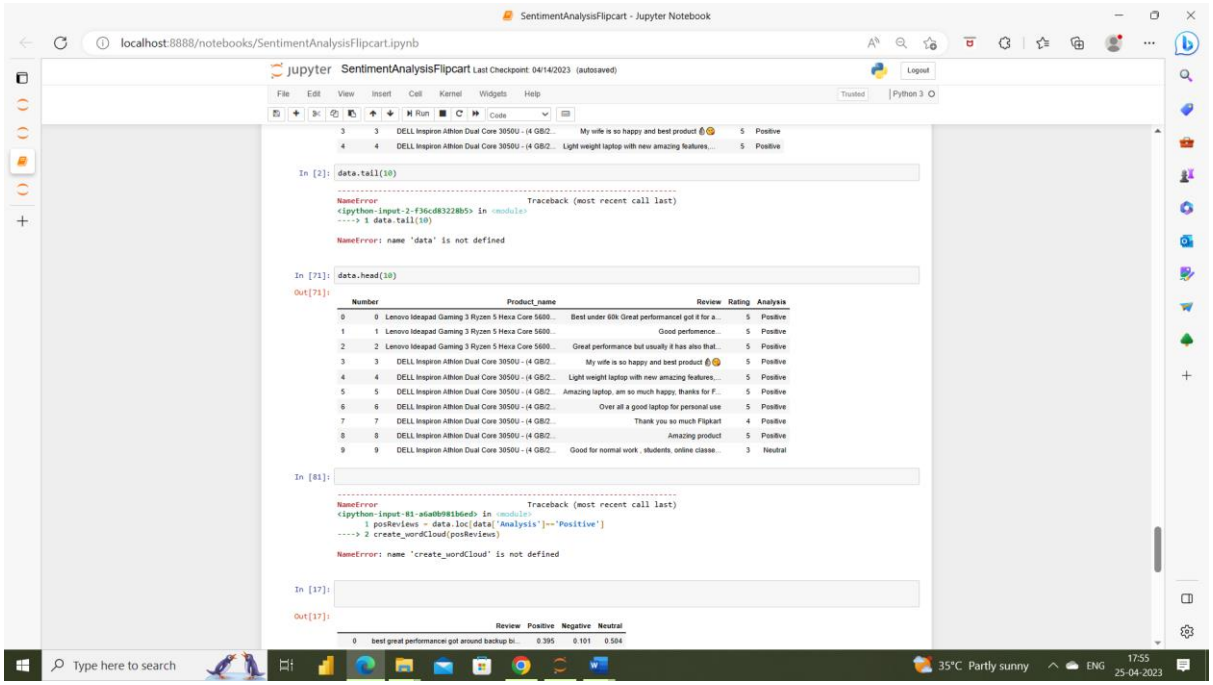
35°C Partly sunny ENG 17:51 25-04-2023

7.Sentiment Analysis eg. Flipcart

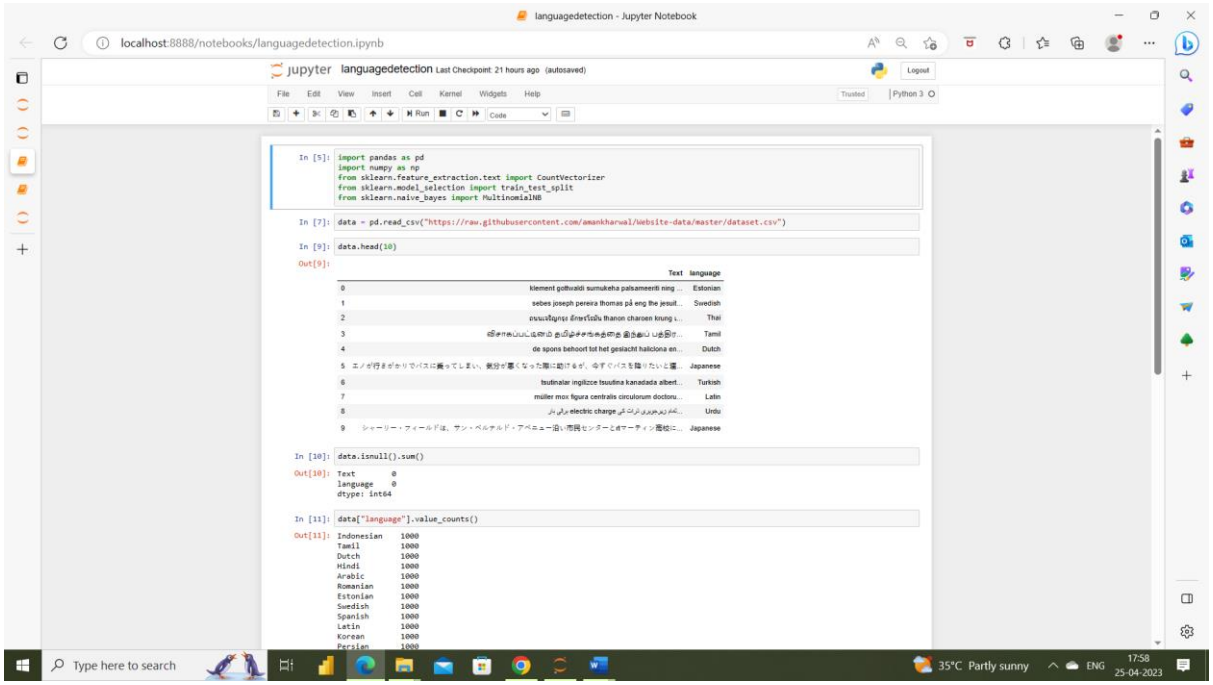


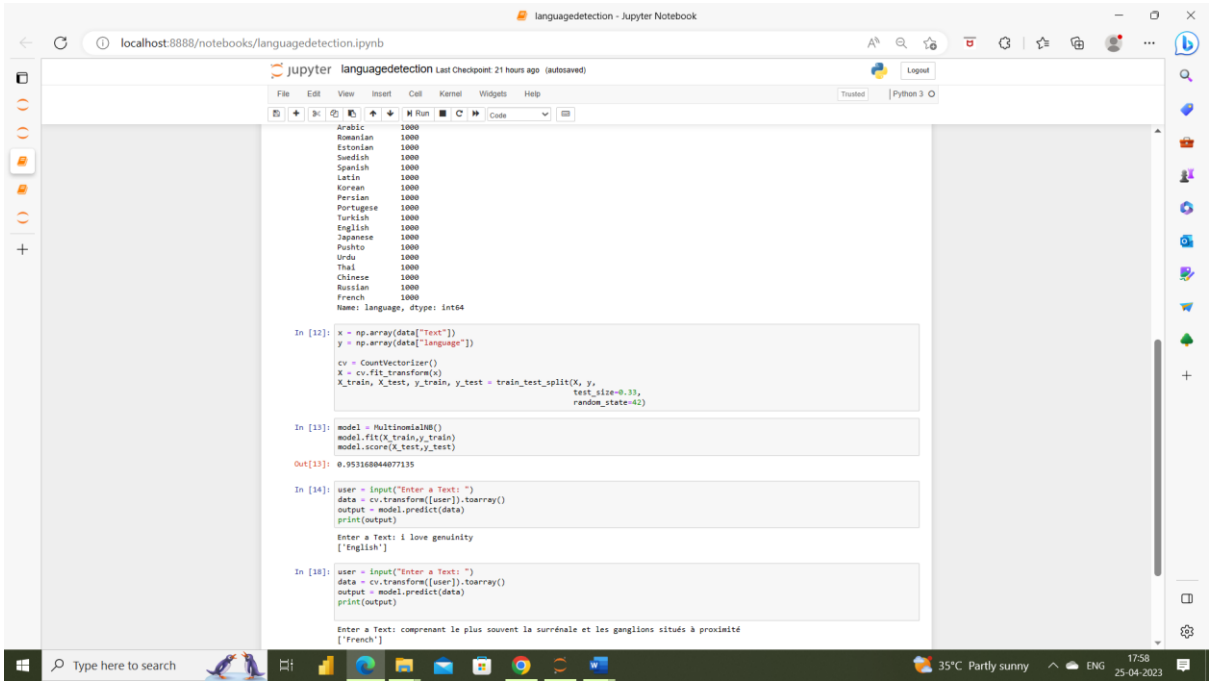



SentimentAnalysisFlipcart - Jupyter Notebook
localhost:8888/notebooks/SentimentAnalysisFlipcart.ipynb
jupyter SentimentAnalysisFlipcart Last Checkpoint: 04/14/2023 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
2302 2302 MSI 27 inch Full HD IPS Panel Monitor (PRO MP2... one issu adapt 5 Positive
2303 2303 MSI 27 inch Full HD IPS Panel Monitor (PRO MP2... worth money u spend monitor great deal use cct... 5 Positive
2304 rows × 5 columns
In [10]:
text = " ".join(i for i in data.Review)
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
background_color="white").generate(text)
plt.figure(figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
phone
good
use
best
display
tv
awesom
mobil
better
buy
product
monitor
price
work
eat
In [11]:
nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["Review"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["Review"]]
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["Review"]]
data = data[["Review", "Positive", "Negative", "Neutral"]]
print(data.head())
[nltk_data] Downloading package vader_lexicon to C:\Users\Dr Ratika
[nltk_data] Datta\AppData\Roaming\nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
Review Positive Negative \
Type here to search
35°C Partly sunny
ENG
17:54
25-04-2023

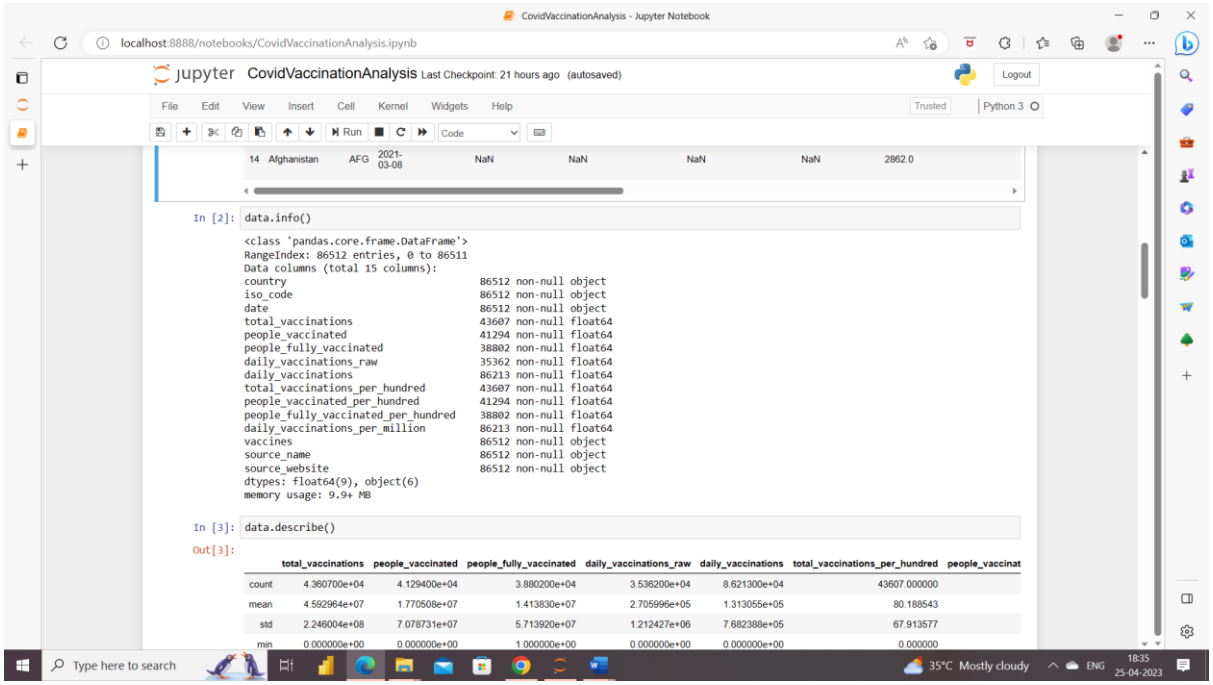

SentimentAnalysisFlipcart - Jupyter Notebook
localhost:8888/notebooks/SentimentAnalysisFlipcart.ipynb
jupyter SentimentAnalysisFlipcart Last Checkpoint: 04/14/2023 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
3 3 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... My wife is so happy and best product 5 Positive
4 4 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Light weight laptop with new amazing features,... 5 Positive
In [2]:
data.tail(10)
NameError Traceback (most recent call last)
<ipython-input-2-f36cd83228b5> in <module>
----> 1 data.tail(10)
NameError: name 'data' is not defined
In [71]:
data.head(10)
Out[71]:
Number Product_name Review Rating Analysis
0 0 Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600... Best under 60k Great performancel got it for a... 5 Positive
1 1 Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600... Good perfomence... 5 Positive
2 2 Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600... Great performance but usually it has also that... 5 Positive
3 3 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... My wife is so happy and best product 5 Positive
4 4 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Light weight laptop with new amazing features... 5 Positive
5 5 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Amazing laptop, am so much happy, thanks for F... 5 Positive
6 6 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Over all a good laptop for personal use 5 Positive
7 7 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Thank you so much Flipkart 4 Positive
8 8 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Amazing product 5 Positive
9 9 DELL Inspiron Athlon Dual Core 3050U - (4 GB/2... Good for normal work , students, online classe... 3 Neutral
In [81]:
NameError Traceback (most recent call last)
<ipython-input-81-a6a0b981b6ed> in <module>
1 posReviews = data.loc[data['Analysis']=='Positive']
----> 2 create_wordCloud(posReviews)
NameError: name 'create_wordCloud' is not defined
In [17]:
Out[17]:
Review Positive Negative Neutral
0 best great performancel got around backup bi... 0.395 0.101 0.504
Type here to search
35°C Partly sunny
ENG
17:55
25-04-2023

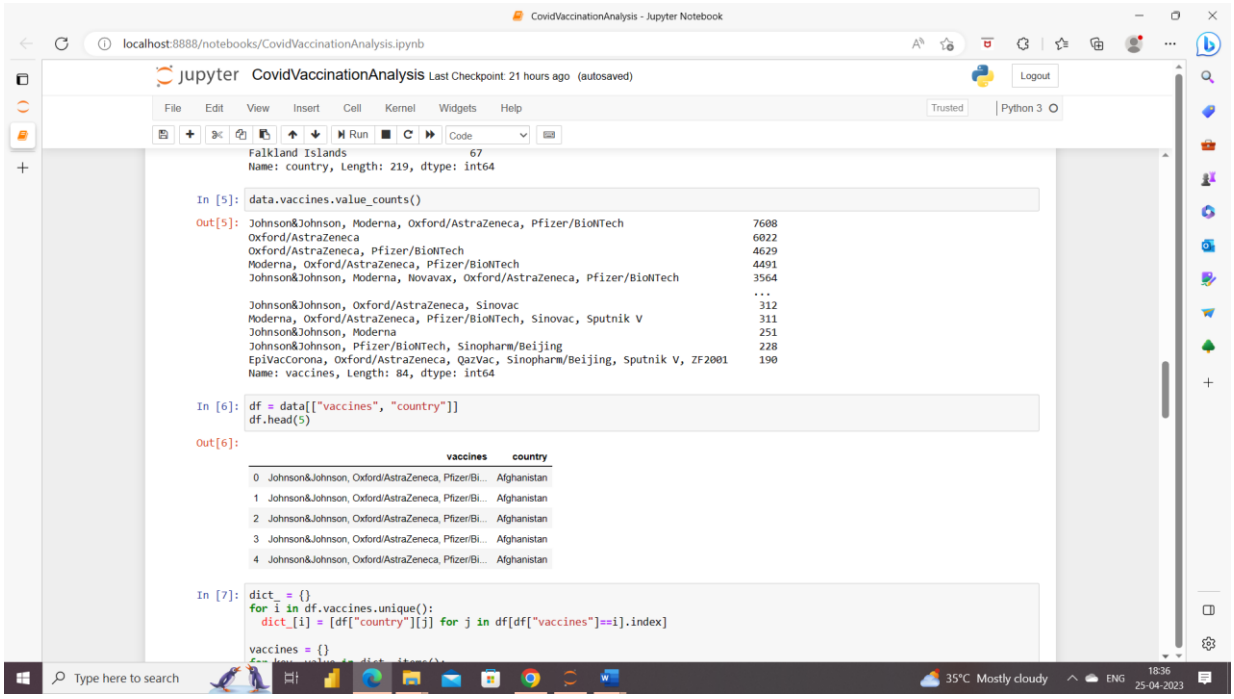
8.Language Prediction Analysis:





9. Covid19 Vaccine Analysis





CovidVaccinationAnalysis - Jupyter Notebook

localhost:8888/notebooks/CovidVaccinationAnalysis.ipynb

jupyter CovidVaccinationAnalysis Last Checkpoint: 21 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

3 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... Afghanistan

4 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... Afghanistan

In [7]:
```
dict_ = {}
for i in df.vaccines.unique():
  dict_[i] = [df["country"][j] for j in df[df["vaccines"]==i].index]

vaccines = {}
for key, value in dict_.items():
  vaccines[key] = set(value)
for i, j in vaccines.items():
  print(f"{i}:>>{j}")
```

```
Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing:>>{'Belize', 'Namibia', 'Trinidad and Tobago', 'Afghanistan', 'Cameroon'}
Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V:>>{'Albania', 'Azerbaijan', 'Bosnia and Herzegovina', 'Oman'}
Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V:>>{'Zimbabwe', 'Algeria'}
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech:>>{'United Kingdom', 'Isle of Man', 'Japan', 'Sint Maarten (Dutch part)', 'Fiji', 'Guernsey', 'Jersey', 'Sweden', 'Andorra', 'Finland', 'Australia'}
Oxford/AstraZeneca:>>{'Nigeria', 'Montserrat', 'Mali', 'Pitcairn', 'Falkland Islands', 'Liberia', 'Sao Tome and Principe', 'Tuvalu', 'Kiribati', 'Angola', 'Samoa', 'Tonga', 'Solomon Islands', 'Nauru', 'Vanuatu', 'Saint Helena', 'Papua New Guinea', 'Democratic Republic of Congo', 'Togo', 'Saint Vincent and the Grenadines'}
Oxford/AstraZeneca, Pfizer/BioNTech:>>{'Bermuda', 'Kosovo', 'Saint Lucia', 'Gibraltar', 'Saint Kitts and Nevis', 'Cayman Islands', 'Saudi Arabia', 'Anguilla', 'New Zealand', 'Costa Rica', 'Panama'}
Oxford/AstraZeneca, Pfizer/BioNTech, Sputnik V:>>{'Antigua and Barbuda'}
CanSino, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V:>>{'Argentina'}
Moderna, Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V:>>{'Armenia'}
Pfizer/BioNTech:>>{'Cook Islands', 'Niue', 'Tokelau', 'Aruba', 'Turks and Caicos Islands', 'Monaco', 'New Caledonia'}
Johnson&Johnson, Moderna, Novavax, Oxford/AstraZeneca, Pfizer/BioNTech:>>{'Netherlands', 'South Korea', 'Italy', 'Czechia', 'Germany', 'Slovenia', 'Austria', 'Lithuania'}
Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech:>>{'Eswatini', 'Grenada', 'Bahamas'}
Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik Light, Sputnik V:>>{'Bahrain'}
Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac:>>{'Bangladesh'}
Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing:>>{'Barbados', 'Maldives', 'Suriname', 'Peru', 'Dominica'}
Sinopharm/Beijing, Sputnik V:>>{'Belarus', 'Kyrgyzstan'}
Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech:>>{'Romania', 'Croatia', 'Jamaica', 'Luxembourg', 'Portugal', 'Greece', 'Spain', 'Iceland', 'Cyprus', 'Canada', 'Belgium', 'Bulgaria', 'Estonia', 'Ireland', 'France', 'Malta', 'Poland'}
```

Type here to search

35°C Mostly cloudy ENG 18:37 25-04-2023

CovidVaccinationAnalysis - Jupyter Notebook

localhost:8888/notebooks/CovidVaccinationAnalysis.ipynb

jupyter CovidVaccinationAnalysis Last Checkpoint: 21 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

```
vaccine_map = px.choropleth(data, locations = 'iso_code', color = 'vaccines')
vaccine_map.update_layout(height=400, margin={"r":0,"t":0,"l":0,"b":0})
vaccine_map.show()
```

vaccines

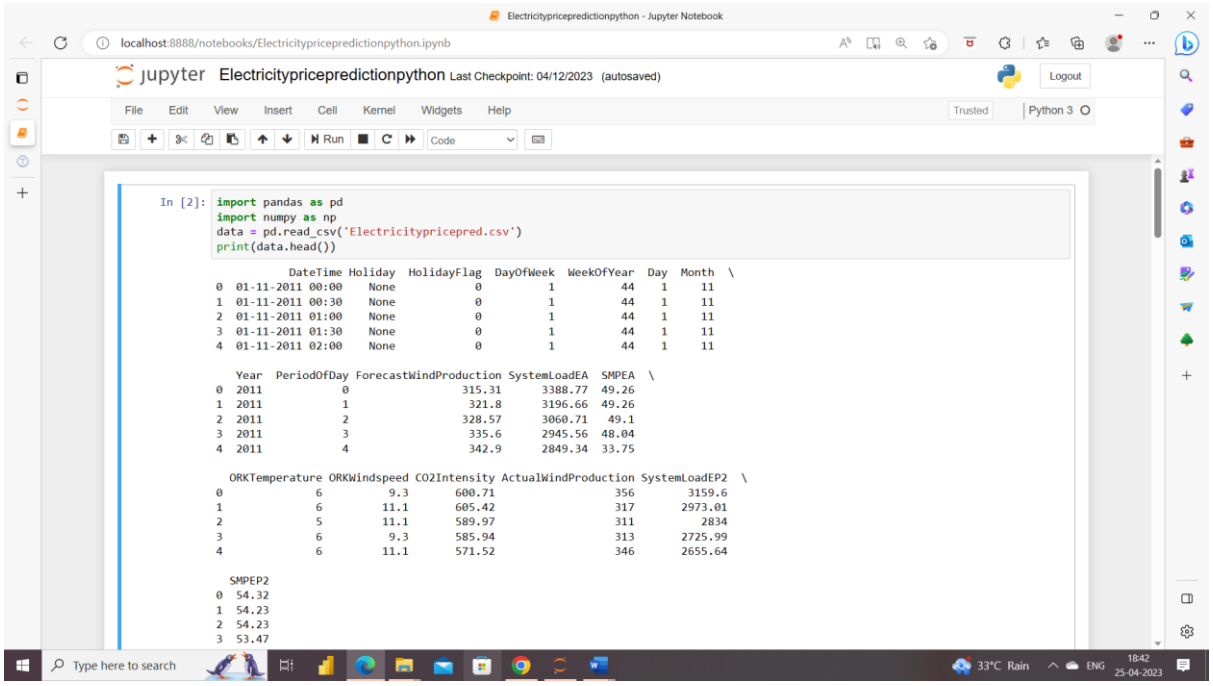
- Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
- Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V
- Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V
- Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
- Oxford/AstraZeneca
- Oxford/AstraZeneca, Pfizer/BioNTech
- Oxford/AstraZeneca, Pfizer/BioNTech, Sputnik V
- CanSino, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V
- Moderna, Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V
- Pfizer/BioNTech
- Johnson&Johnson, Moderna, Novavax, Oxford/AstraZeneca, Pfizer/BioNTech
- Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech
- Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik Light, Sputnik V
- Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac
- Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
- Sinopharm/Beijing, Sputnik V
- Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
- Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac
- Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
- Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V

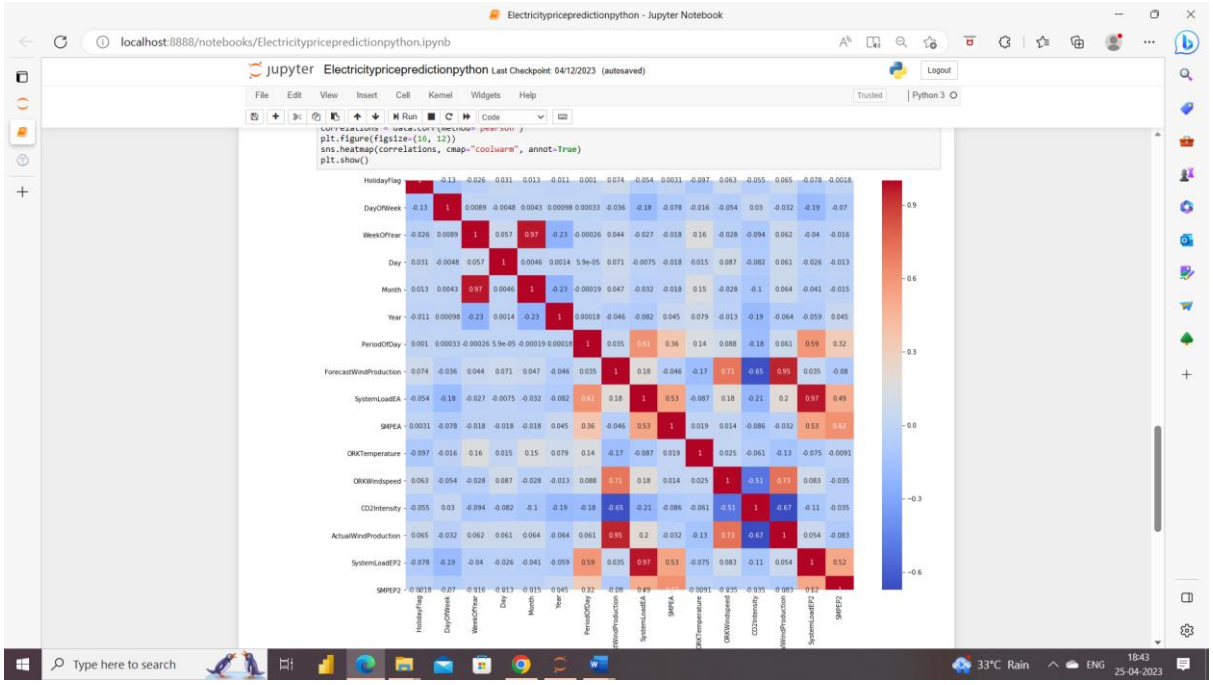
In [ ]:
```
#The The code starts by creating an empty dictionary called dict_. It then iterates through the unique values of the vaccines co
```

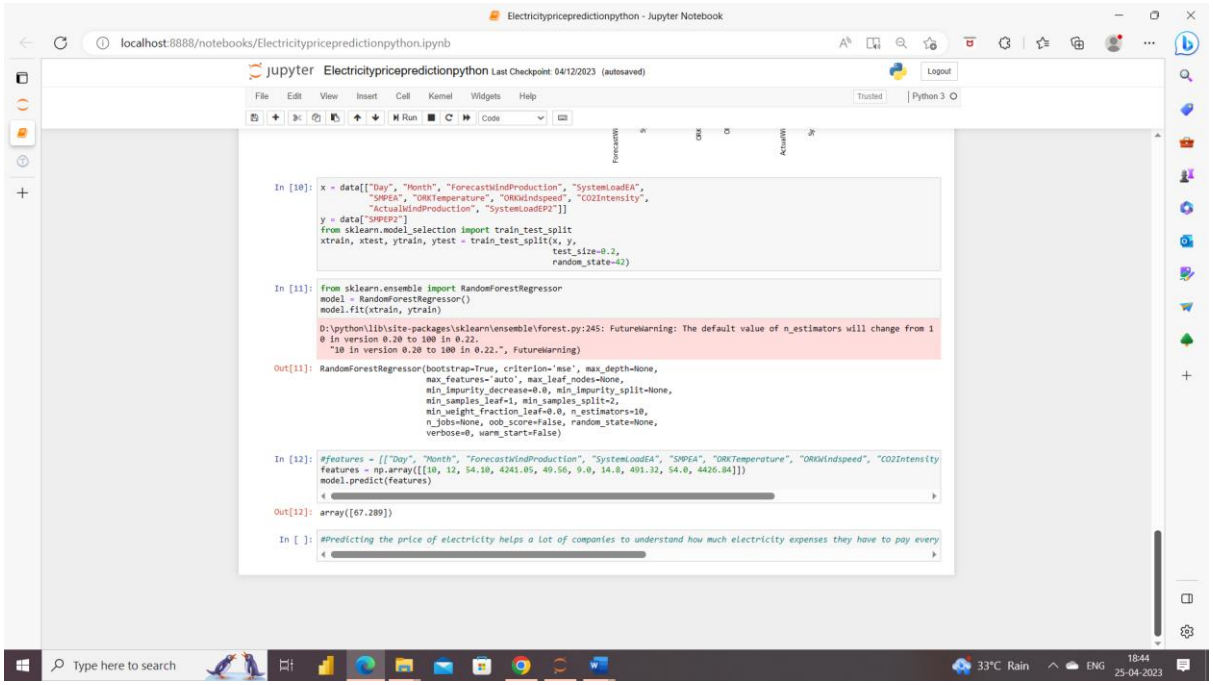
Type here to search

33°C Rain ENG 18:40 25-04-2023

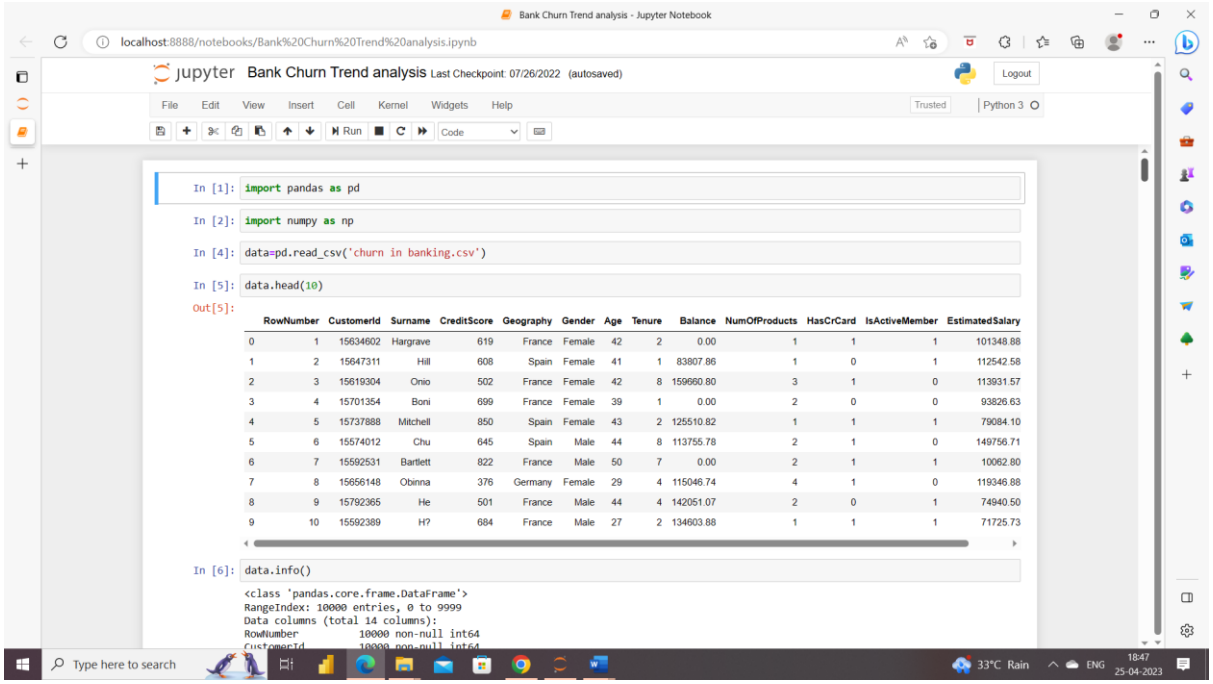
10. Electricity Price Prediction:

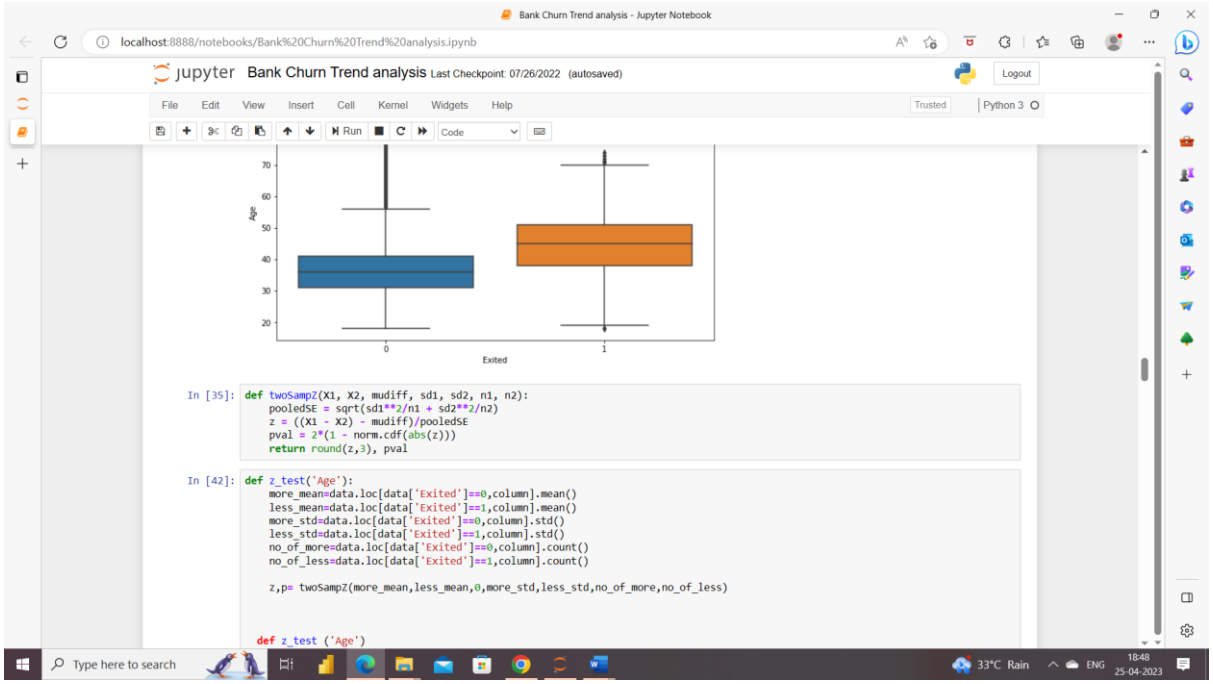


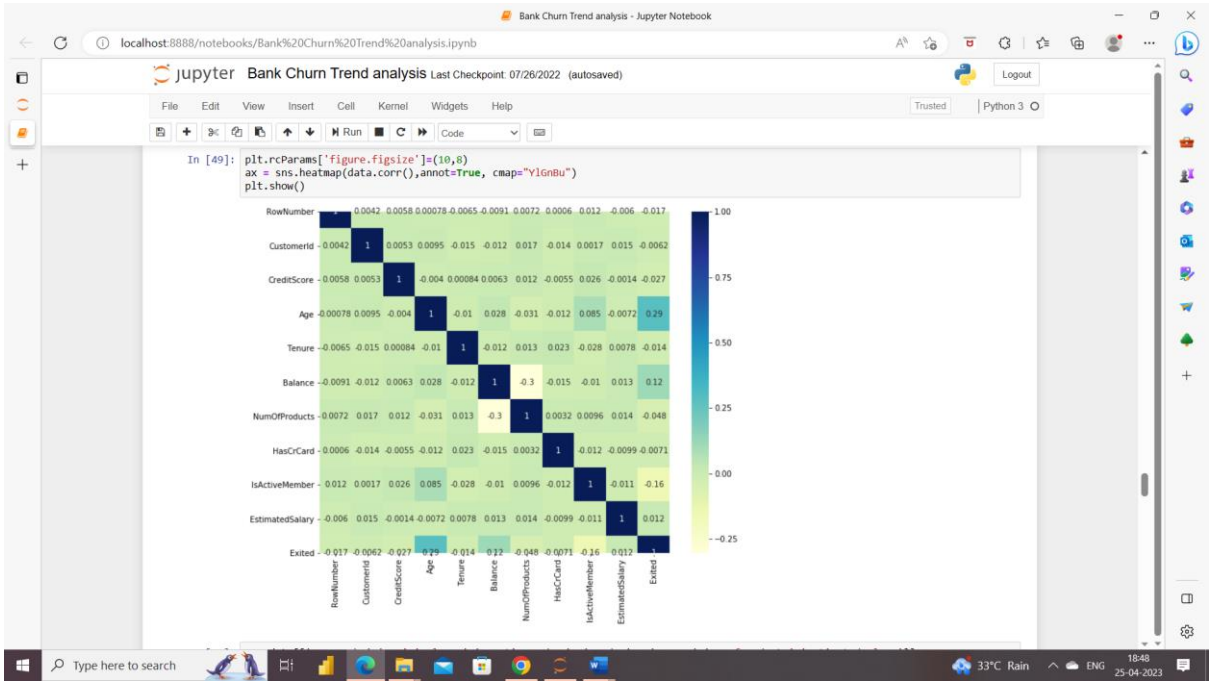




11.Bank Churn Analysis




Bank Churn Trend analysis - Jupyter Notebook
localhost:8888/notebooks/Bank%20Churn%20Trend%20analysis.ipynb
jupyter Bank Churn Trend analysis Last Checkpoint: 07/26/2022 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
Age
Exited
In [35]:
def twoSampZ(X1, X2, mudiff, sd1, sd2, n1, n2):
pooledSE = sqrt(sd1**2/n1 + sd2**2/n2)
z = ((X1 - X2) - mudiff)/pooledSE
pval = 2*(1 - norm.cdf(abs(z)))
return round(z,3), pval
In [42]:
def z_test('Age'):
more_mean=data.loc[data['Exited']==0,column].mean()
less_mean=data.loc[data['Exited']==1,column].mean()
more_std=data.loc[data['Exited']==0,column].std()
less_std=data.loc[data['Exited']==1,column].std()
no_of_more=data.loc[data['Exited']==0,column].count()
no_of_less=data.loc[data['Exited']==1,column].count()
z,p= twoSampZ(more_mean,less_mean,0,more_std,less_std,no_of_more,no_of_less)
def z_test ('Age')
Type here to search
33°C Rain
ENG
18:48
25-04-2023


Bank Churn Trend analysis - Jupyter Notebook
localhost:8888/notebooks/Bank%20Churn%20Trend%20analysis.ipynb
jupyter Bank Churn Trend analysis Last Checkpoint: 07/26/2022 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [49]:
plt.rcParams['figure.figsize']=(10,8)
ax = sns.heatmap(data.corr(),annot=True, cmap="YlGnBu")
plt.show()
RowNumber
CustomerId
CreditScore
Age
Tenure
Balance
NumOfProducts
HasCrCard
IsActiveMember
EstimatedSalary
Exited
Type here to search
33°C Rain
ENG
18:48
25-04-2023

12.Google Word Trend Analysis