

# **WEBSITE TRAFFIC ANALYSIS**

## **PHASE 4 : DEVELOPMENT PART 2**

**PRIYADARSHINI ENGINEERING COLLEGE**

**Sathish Kumar / 511921104064**

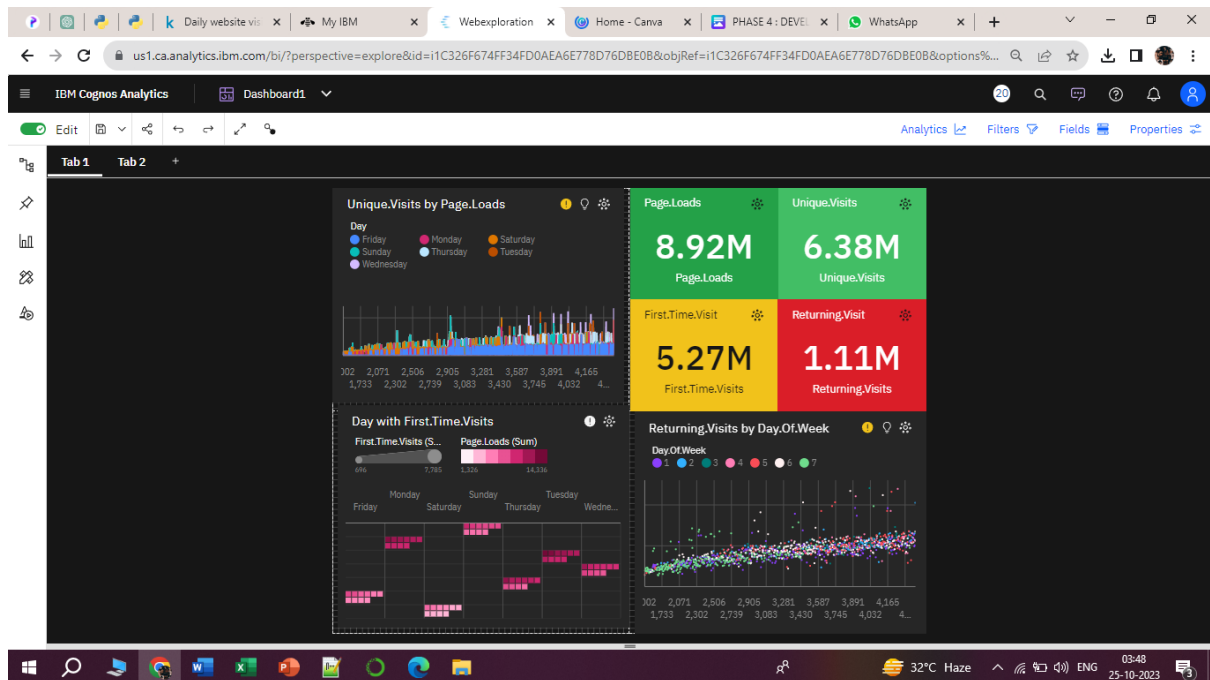
### **Introduction**

Website traffic analysis is the process of collecting, examining, and interpreting data related to the visitors and interactions on a website. It provides invaluable insights into user behavior, preferences, and trends, helping organizations make informed decisions, optimize their online presence, and enhance user experiences.

### **Abstract**

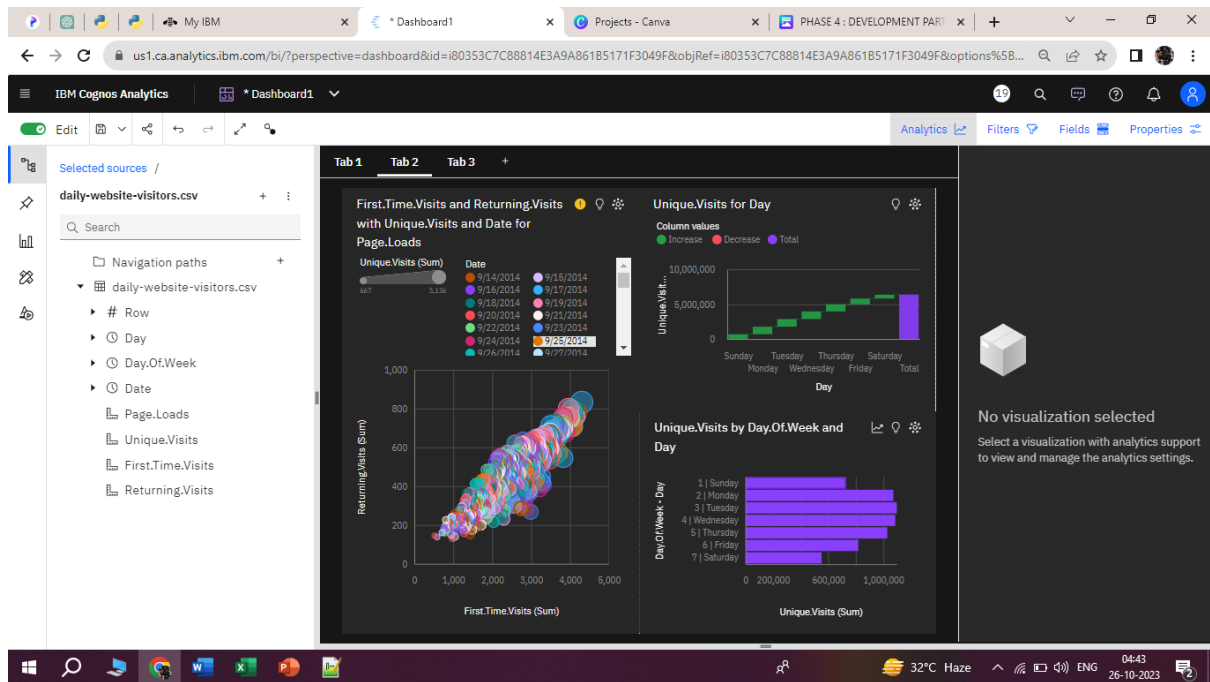
This project aims to analyze website traffic data for insights into user behavior, popular pages, and traffic sources. It involves data collection, visualization using IBM Cognos, and Python for advanced analysis. The goal is to optimize user experiences and enhance website performance.

# Data Exploration



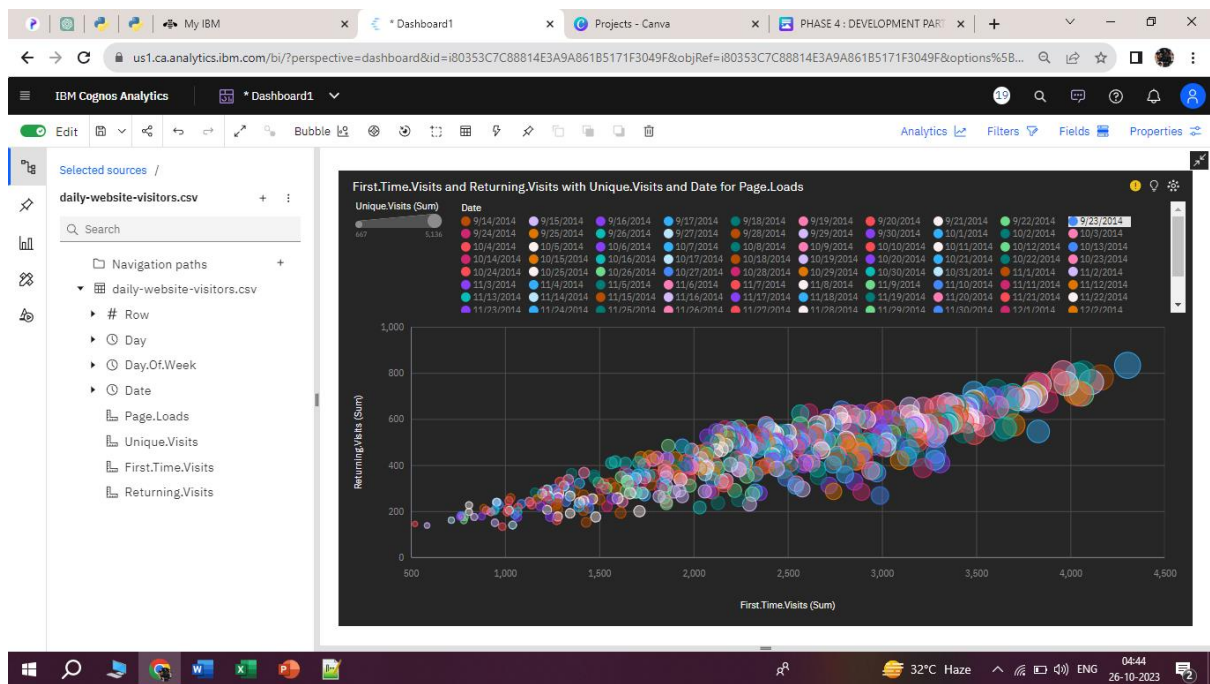
On Phase 2 Development part 1, made only the Data set 'Analysis visualization' of overall Page loads, Unique visits and more.

From the Continuation of Phase 3 development part 1 improved the analysis in a advanced analysis to improve the prediction and to show the visualization.

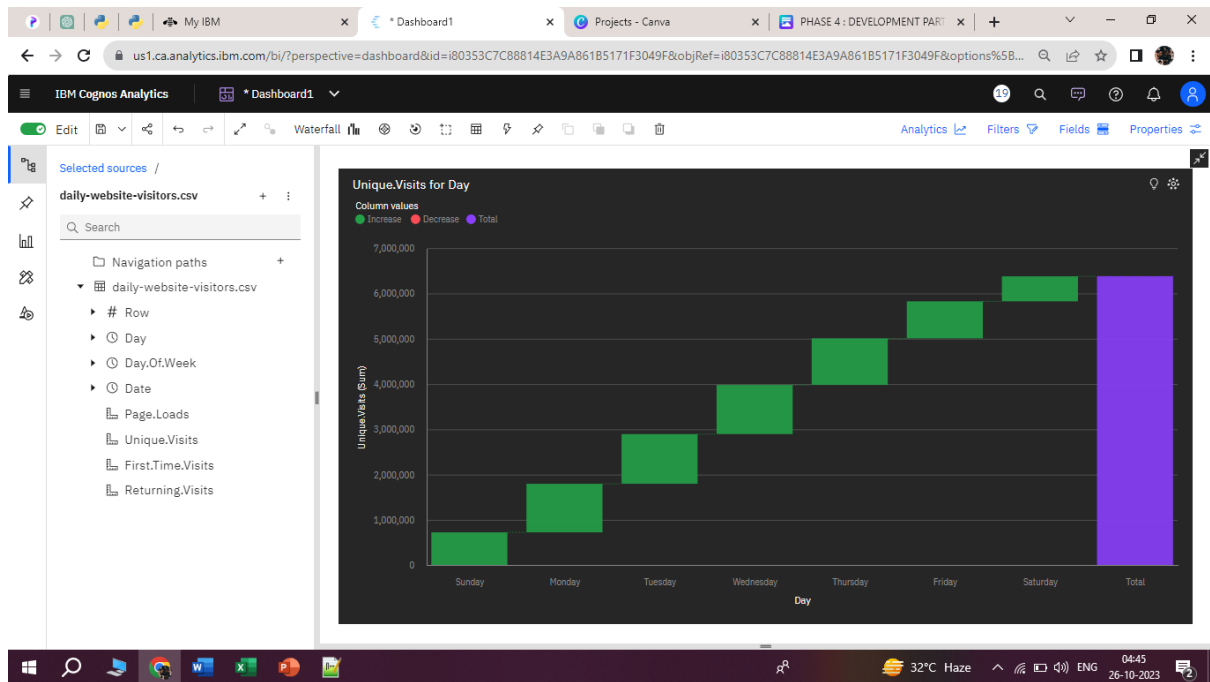


## Insights :

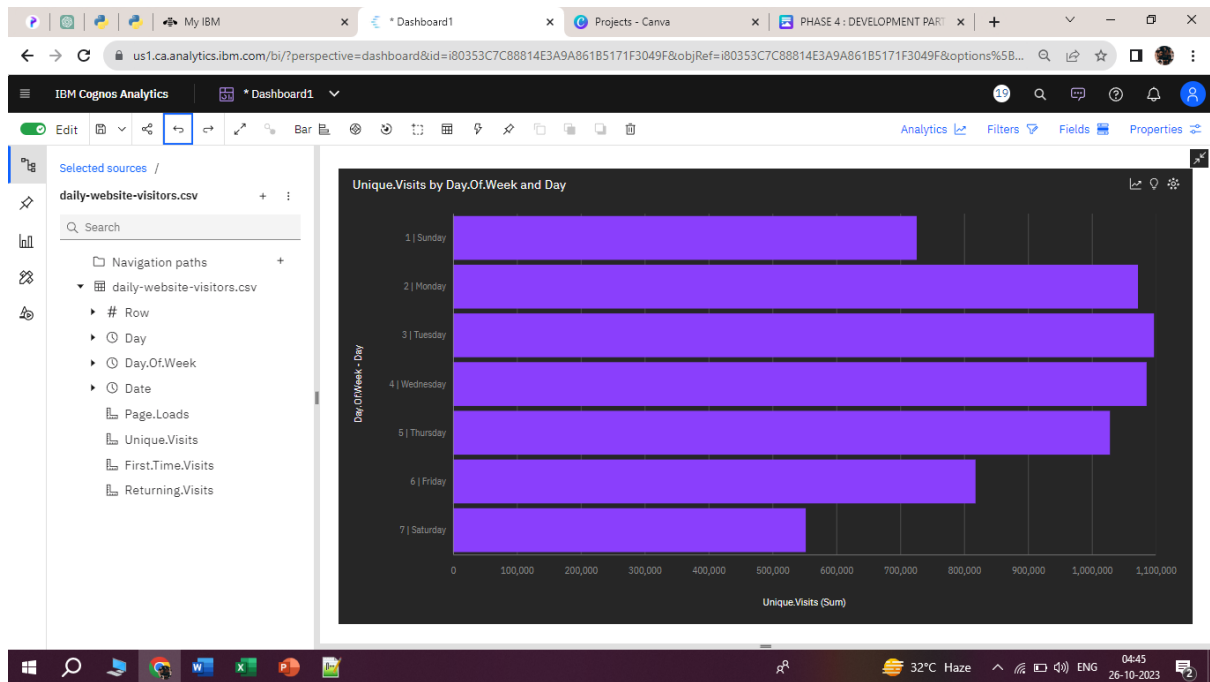
- Based on the current forecasting, **Page.Loads** may reach **nearly four thousand by Date 2021-10-27**
- **Page.Loads** has a strong weekly trend. The largest values typically occur on **Tuesday**, whereas the smallest values on **Saturday**.
- Over all **dates**, the average of **Unique.Visits** is **nearly three thousand**.
- Over all **dates**, the average of **First.Time.Visits** is **almost 2500**.



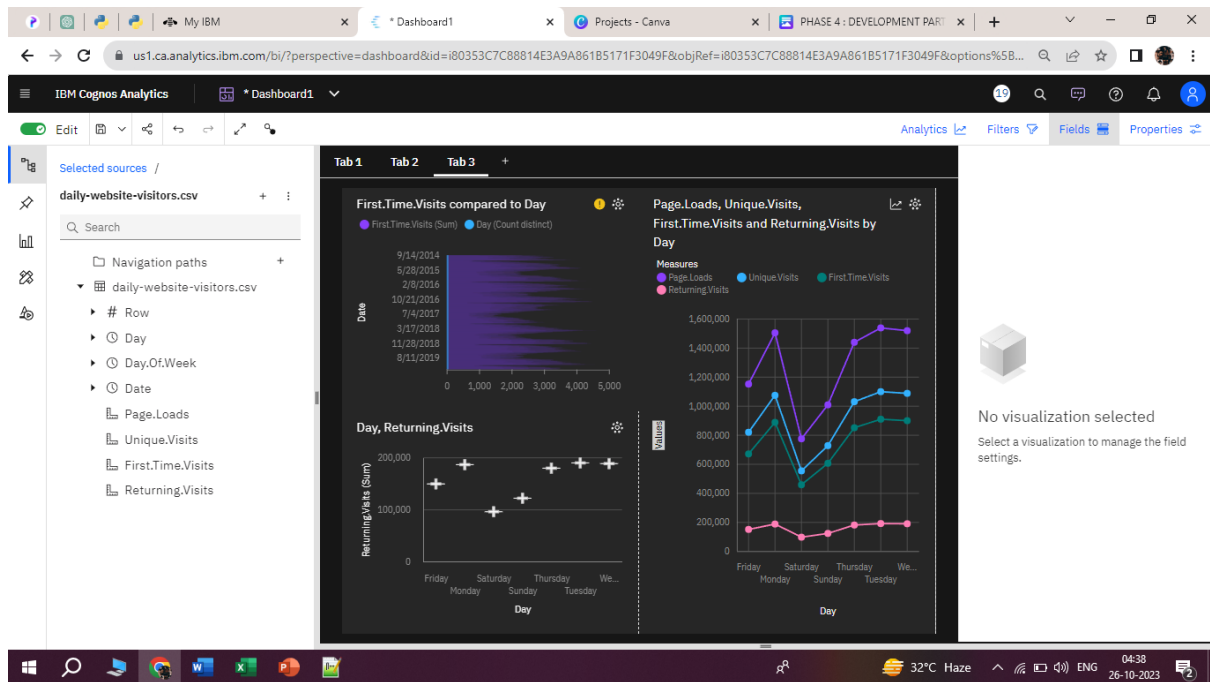
- ❖ First time visitors, Unique visitors, Returning Visitors are get Varying day-by-day on the daily basis based on the performance and expericence of the Website so, we can't able to judge the performance and visitors user exeperience.



- ❖ Unique Visitors are the Daily and loyal visitors .who use the platform for they professional use with there brain storm addiciton for the website.

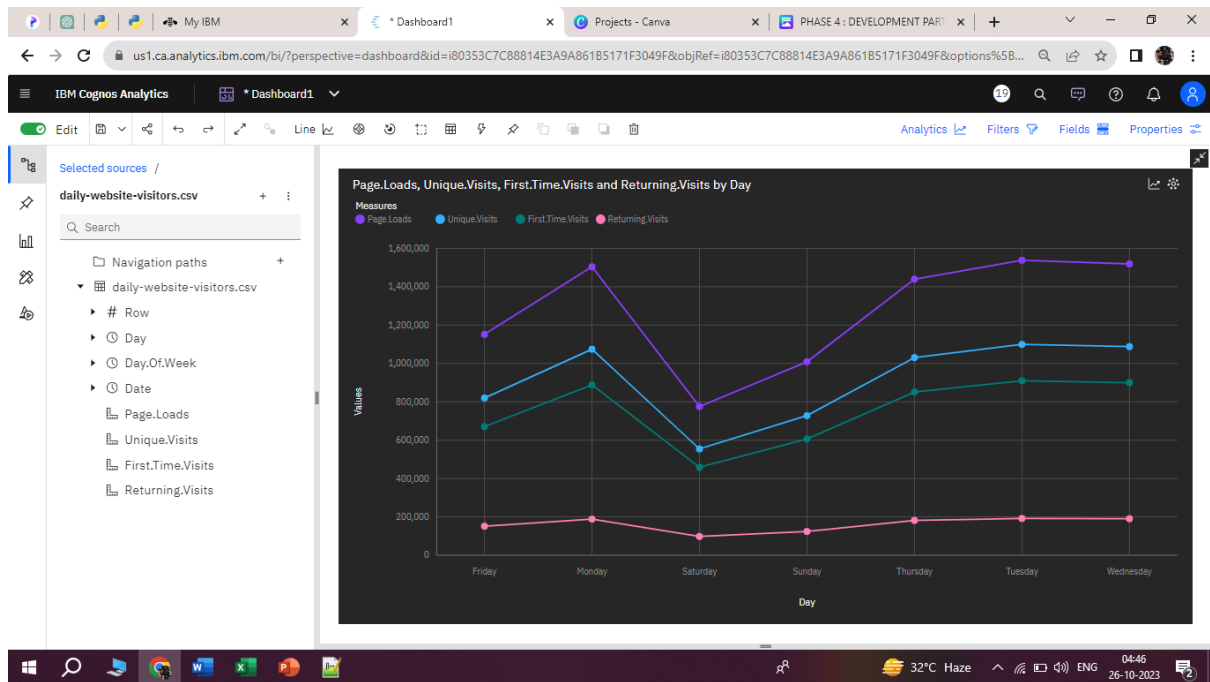


- ❖ There is major difference in week days and week ends visitors got varied in there need.
- ❖ When compared with the week ends the professionals are using the platform on there work basis on week days.



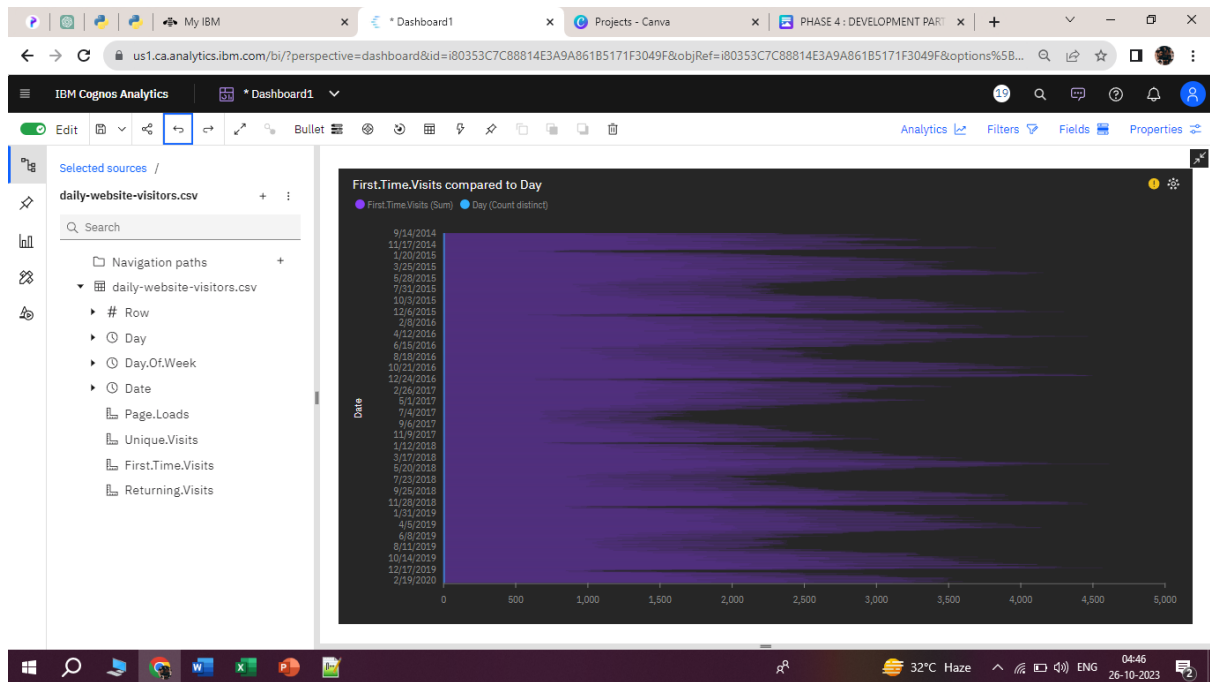
## Insights :

- Over all **dates**, the average of **Returning.Visits** is **511.8**.
- Across all **dates**, the average of **Page.Loads** is **over four thousand**.
- The total number of results for **FirstTime.Visits**, across all **dates**, is **over two thousand**.
- The total number of results for **Page.Loads**, across all **dates**, is **over two thousand**.

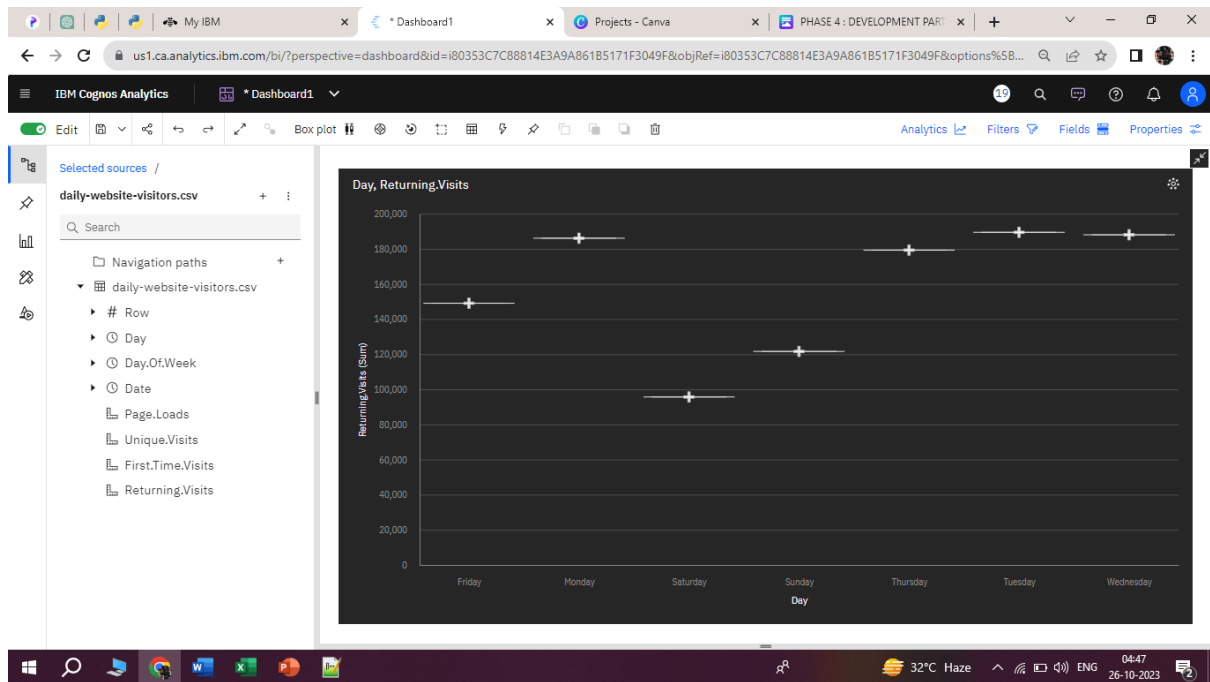


- ❖ Here You can go with Page loads got Analysed in comparison with the First time, Returning Visits, Unique visits .
- ❖ Loads are more and Traffic is more on starting of week days like Monday and end week days like Friday.





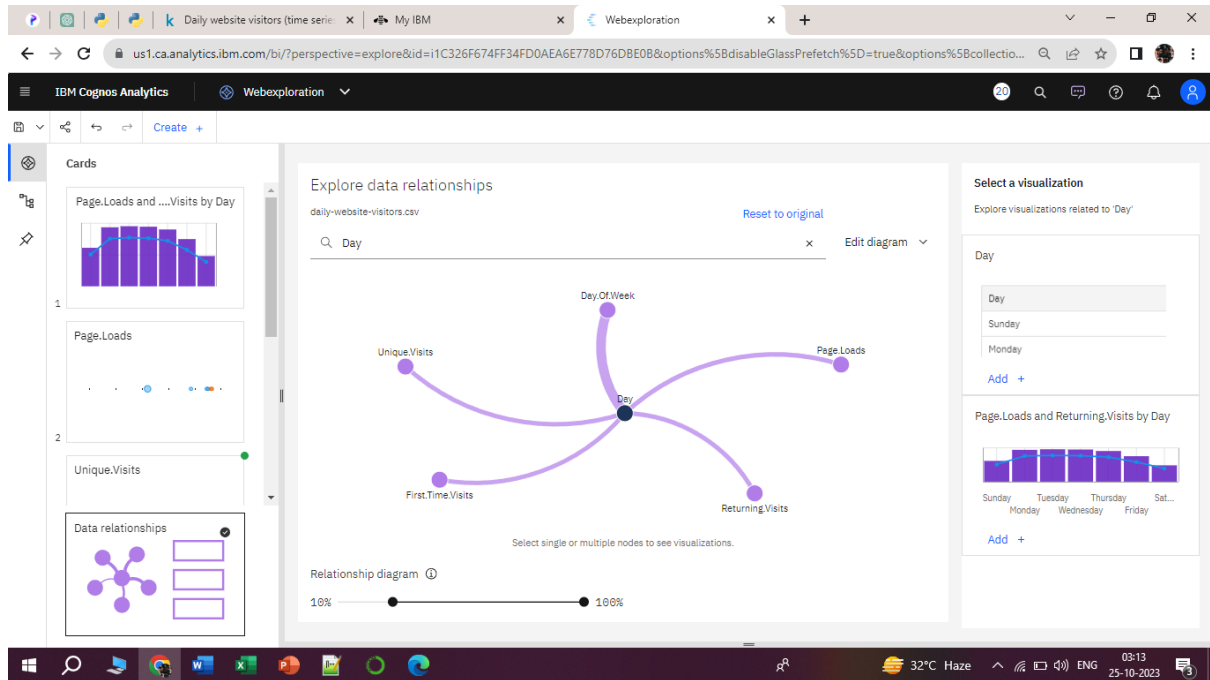
- ❖ First visitors are the new visitors for there necessity ,Product improvement and more.
- ❖ First visitors have the seasonal visit traffic on the above graph with the respective months and days.



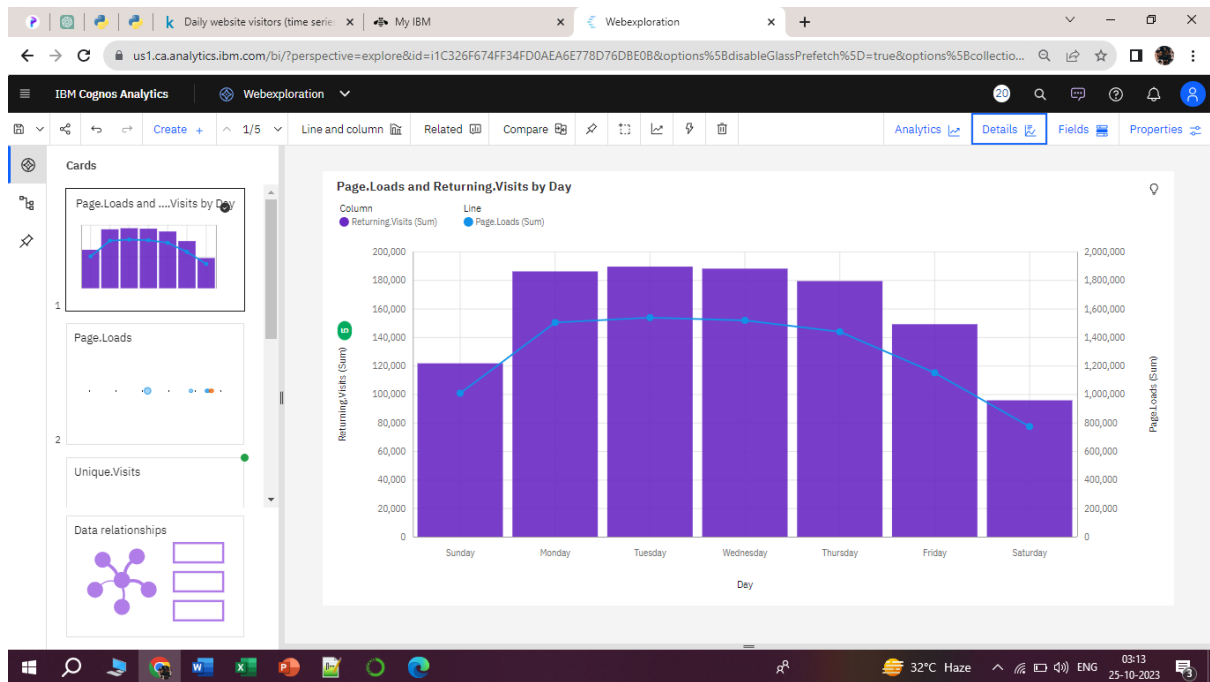
- ❖ First visitors are the new visitors for there necessity ,Product improvement and more.
- ❖ First visitors have the seasonal visit traffic on the above graph with the respective months and days.

# Report

## Segmentation

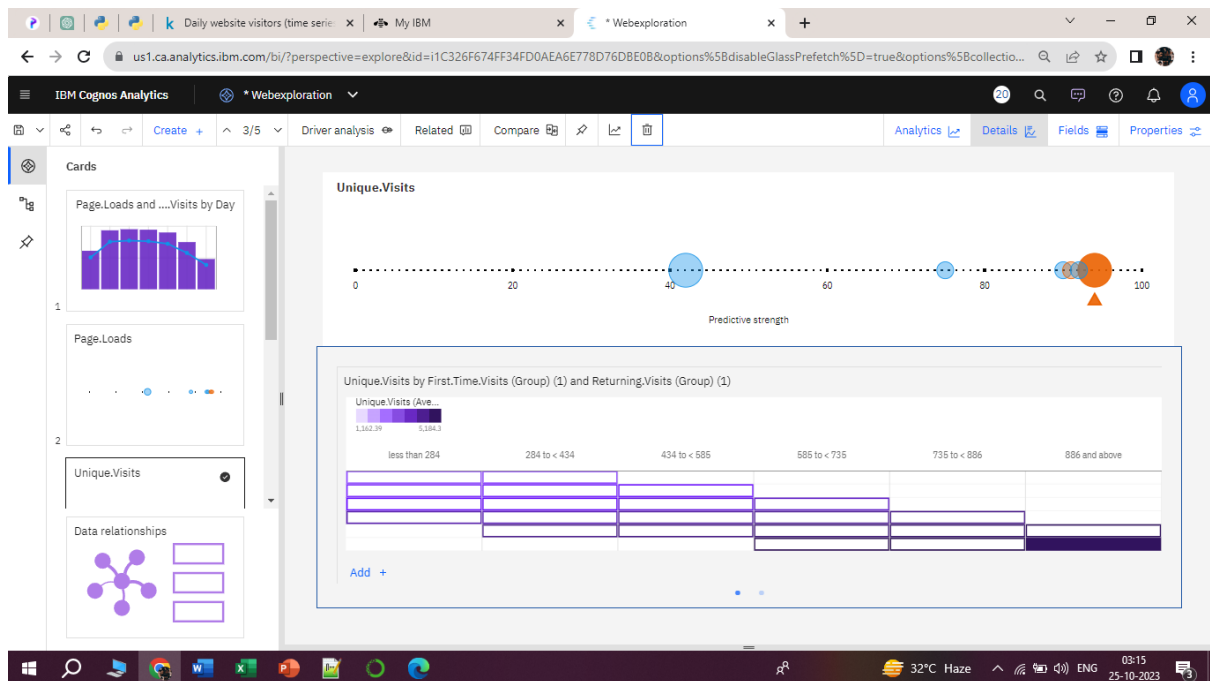


- Every insights are interlinked with the Day ,Here the Day is the major source which interlinks all the corresponding respectives.
- without any collapse and congestion every insights are linked with Day to show the Website Traffic.

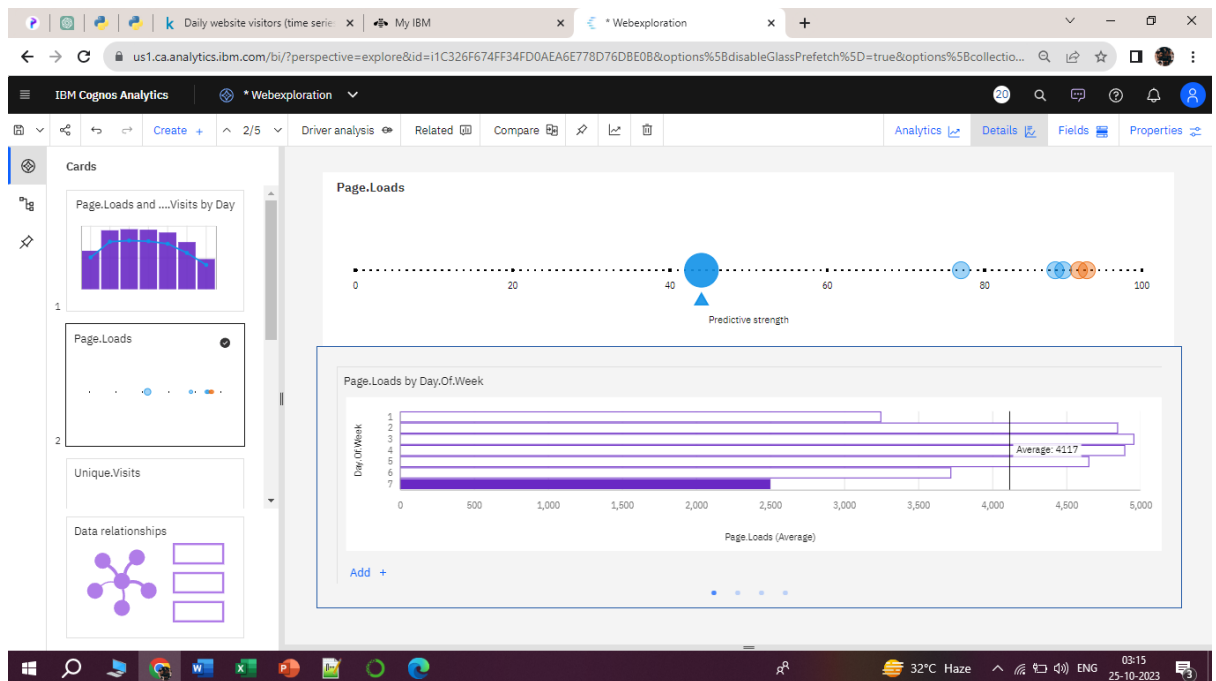


## Insights :

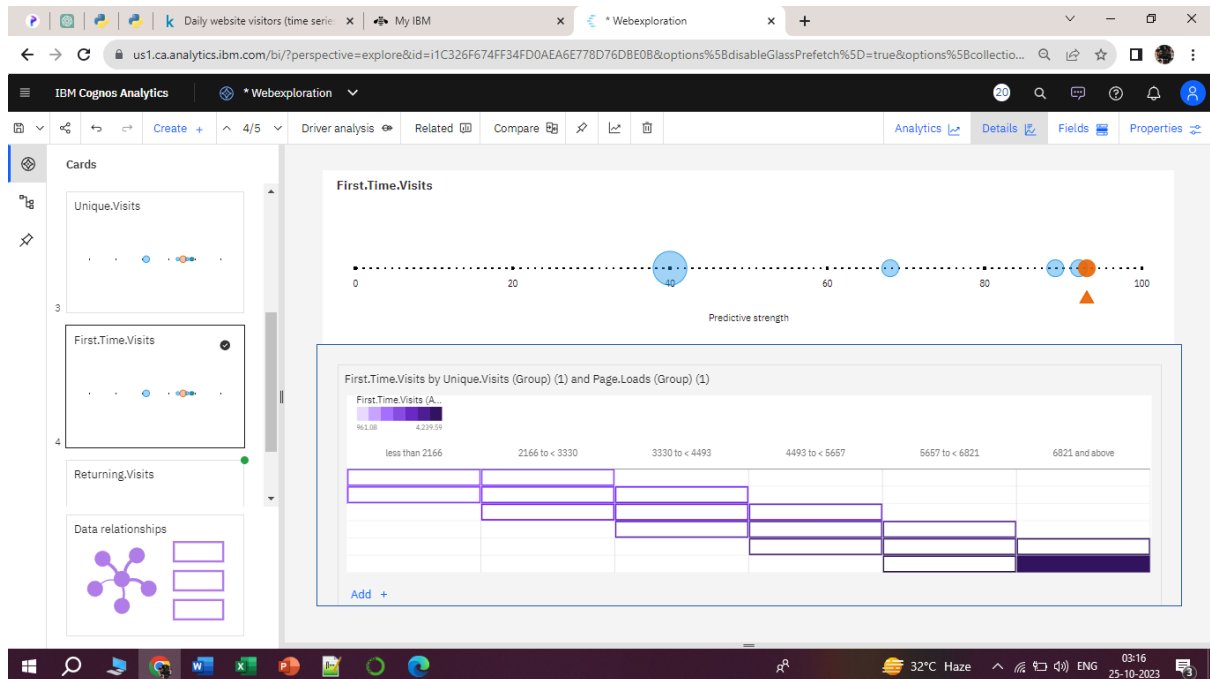
- Across all **days**, the sum of **Returning.Visits** is over 1.1 million.
- **Returning.Visits** ranges from almost 96 thousand, when **Day** is Saturday, to over 189 thousand, when **Day** is Tuesday.
- **Returning.Visits** is unusually low when **Day** is Saturday.
- For **Returning.Visits**, the most significant values of **Day** are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective **Returning.Visits** values add up to almost 892 thousand, or 80.4 % of the total.
- Across all **days**, the sum of **Page.Loads** is over 8.9 million.
- **Page.Loads** ranges from nearly 773 thousand, when **Day** is Saturday, to over 1.5 million, when **Day** is Tuesday.
- **Page.Loads** is unusually low when **Day** is Saturday.



- **First.Time.Visits (Group) (3)** strongly affects **Unique.Visits** (94%).
- **Unique.Visits** is most unusual when **First.Time.Visits (Group) (3)** is 3934 and above and less than 1205.
- **Returning.Visits (Group) (2)** strongly affects **Unique.Visits** (76%).
- **Unique.Visits** is unusually high when **Returning.Visits (Group) (2)** is 886 and above.
- Over all values of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)**, the average of **Unique.Visits** is nearly three thousand.
- The average values of **Unique.Visits** range from over a thousand to over five thousand.
- **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** strongly affect **Unique.Visits** (96%).
- **Unique.Visits** is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.
- 1887 to < 2569 is the most frequently occurring category of **First.Time.Visits (Group) (3)** with a count of 666 items with **Unique.Visits** values (30.7 % of the total).
- 434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).
- There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.



- Across all values of **Day.Of.Week**, the average of **Page.Loads** is over four thousand.
- The average values of **Page.Loads** range from over 2500, occurring when **Day.Of.Week** is 7, to nearly five thousand, when **Day.Of.Week** is 3.
- **Day.Of.Week** moderately affects **Page.Loads** (44%).
- **Page.Loads** is unusually low when **Day.Of.Week** is 7.
- 1 (14.3 %), 2 (14.3 %), 3 (14.3 %), and 4 (14.3 %) are the most frequently occurring categories of **Day.Of.Week** with a combined count of 1240 items with **Page.Loads** values (57.2 % of the total).



- **Unique.Visits** is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.
- 1887 to < 2569 is the most frequently occurring category of **First.Time.Visits (Group) (3)** with a count of 666 items with **Unique.Visits** values (30.7 % of the total).
- 434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).
- There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.



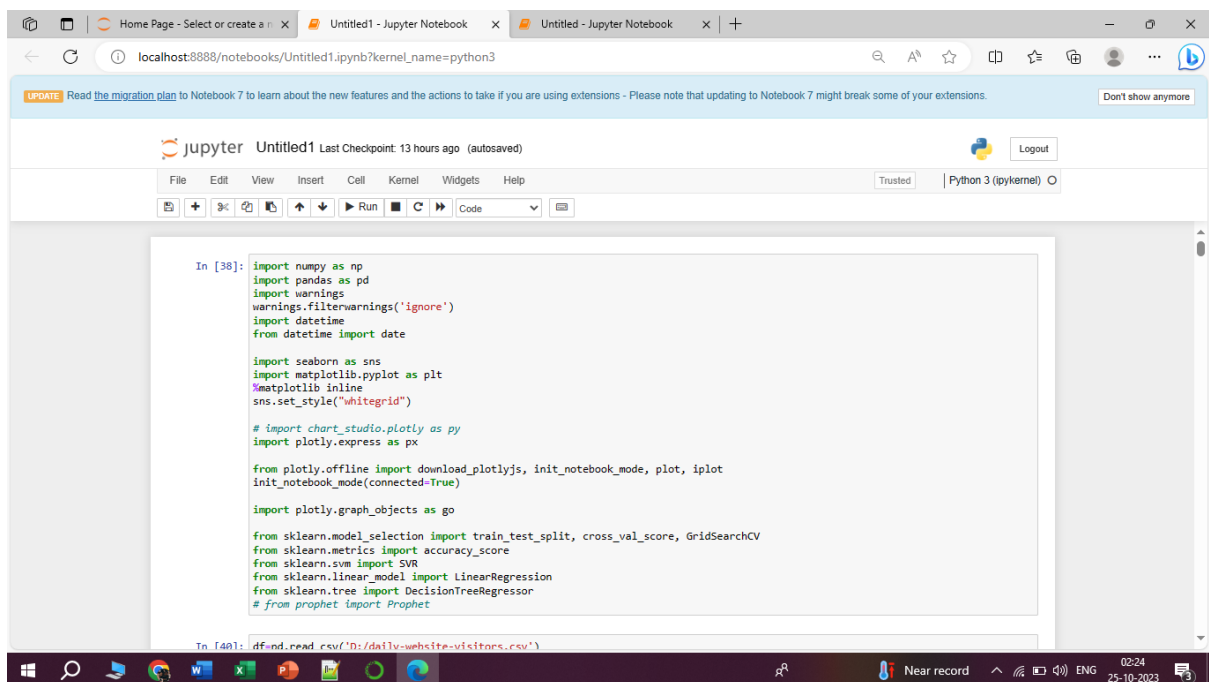


# Python Integration

Under Python Integration Part is to make the analysis of traffic to make better and predictive with advanced techniques.

Using popular Library Modules

1. Numpy
2. Pandas
3. Matplot
4. Scipy
5. Seaborn and more for visualization ,analysis.



The screenshot shows a Jupyter Notebook interface in a web browser. The browser address bar shows 'localhost:8888/notebooks/Untitled1.ipynb?kernel\_name=python3'. The Jupyter Notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running code, and other functions. The main area displays a code cell with the following Python code:

```
In [38]: import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
import datetime
from datetime import date

import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set_style("whitegrid")

# import chart_studio.plotly as py
import plotly.express as px

from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
init_notebook_mode(connected=True)

import plotly.graph_objects as go

from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
from sklearn.metrics import accuracy_score
from sklearn.svm import SVR
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
# from prophet import Prophet
```

Below the code cell, the prompt 'In [40]:' is visible, followed by the start of a new line of code: 'df = pd.read\_csv('D:/daily-website-visitors.csv')'.

Importing the necessary modules to make the analysis better to know the work flow of traffic .

The screenshot shows a Jupyter Notebook interface with a code cell containing the following Python code:

```
In [40]: df=pd.read_csv('D:/daily-website-visitors.csv')

df.rename(columns = {'Day.Of.Week':'day_of_week',
                    'Page.Loads':'page_loads',
                    'Unique.Visits':'unique_visits',
                    'First.Time.Visits':'first_visits',
                    'Returning.Visits':'returning_visits'}, inplace = True)

df=df.replace(',','',regex=True)

df['page_loads']=df['page_loads'].astype(int)
df['unique_visits']=df['unique_visits'].astype(int)
df['first_visits']=df['first_visits'].astype(int)
df['returning_visits']=df['returning_visits'].astype(int)

df
```

The output of the code is a preview of the DataFrame:

```
Out[40]:
```

Row	Day	day_of_week	Date	page_loads	unique_visits	first_visits	returning_visits
0	1	Sunday	1 9/14/2014	2146	1582	1430	152
1	2	Monday	2 9/15/2014	3621	2528	2297	231
2	3	Tuesday	3 9/16/2014	3698	2630	2352	278
3	4	Wednesday	4 9/17/2014	3667	2614	2327	287
4	5	Thursday	5 9/18/2014	3316	2366	2130	236
...	...	...	...	...	...	...	...
2162	2163	Saturday	7 8/15/2020	2221	1696	1373	323

Printing the Insights like Rows, Columns, Datas inside the Data sheet which provided on the Kaggle Website for Website Traffic Analysis

The screenshot shows a Jupyter Notebook interface with three code cells. The first cell contains:

```
In [41]: df.isna().sum()
```

The output is:

```
Out[41]:
```

Row	0
Day	0
day_of_week	0
Date	0
page_loads	0
unique_visits	0
first_visits	0
returning_visits	0
dtype:	int64

The second cell contains:

```
In [42]: df.duplicated().sum()
```

The output is:

```
Out[42]: 0
```

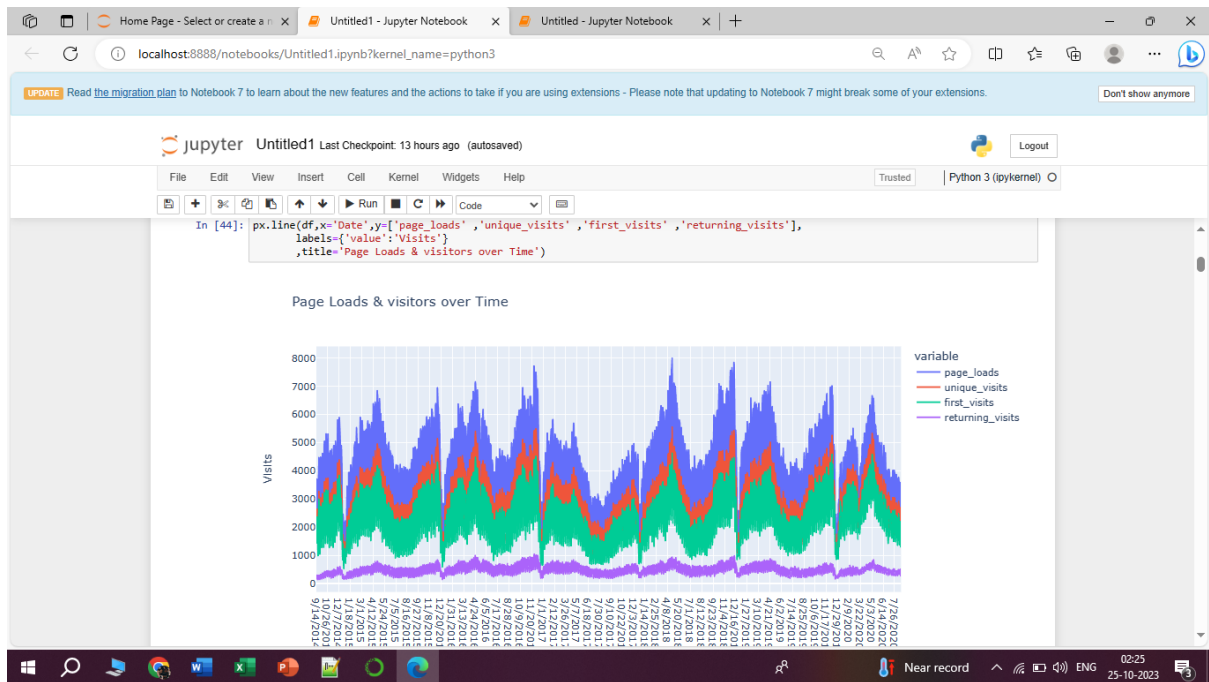
The third cell contains:

```
In [43]: df.info()
```

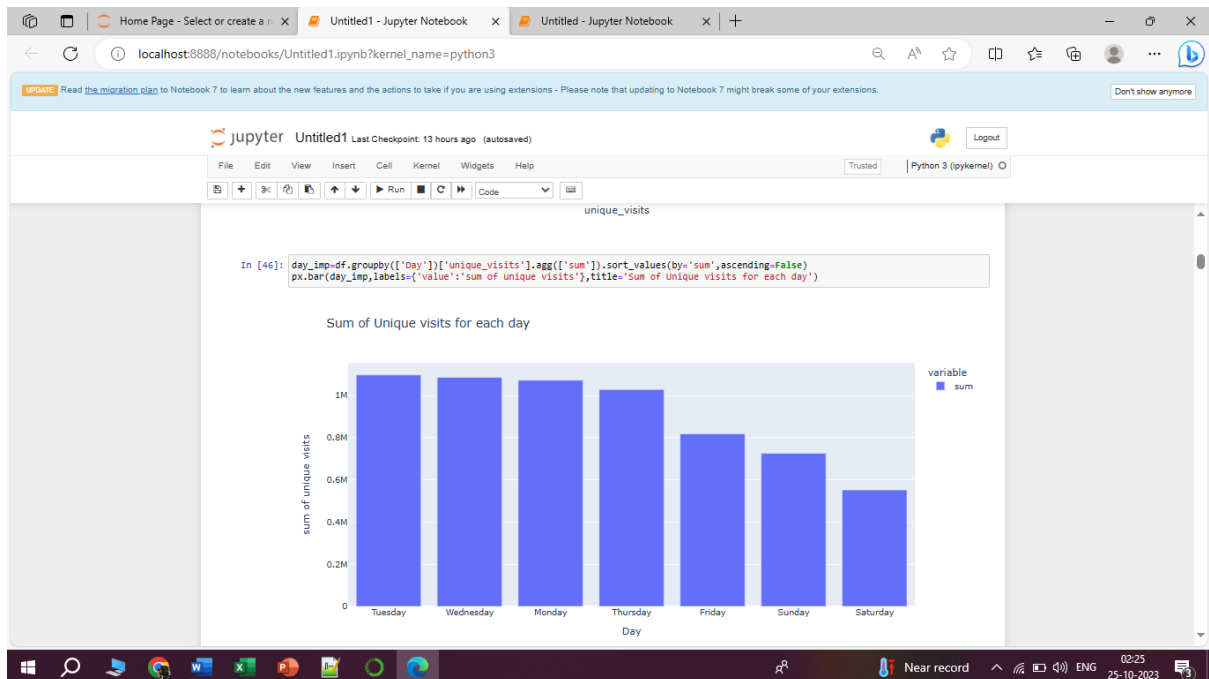
The output is:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2167 entries, 0 to 2166
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Row                   2167 non-null  int64
1   Day                   2167 non-null  object
2   day_of_week           2167 non-null  int64
3   Date                  2167 non-null  object
4   page_loads            2167 non-null  int32
```

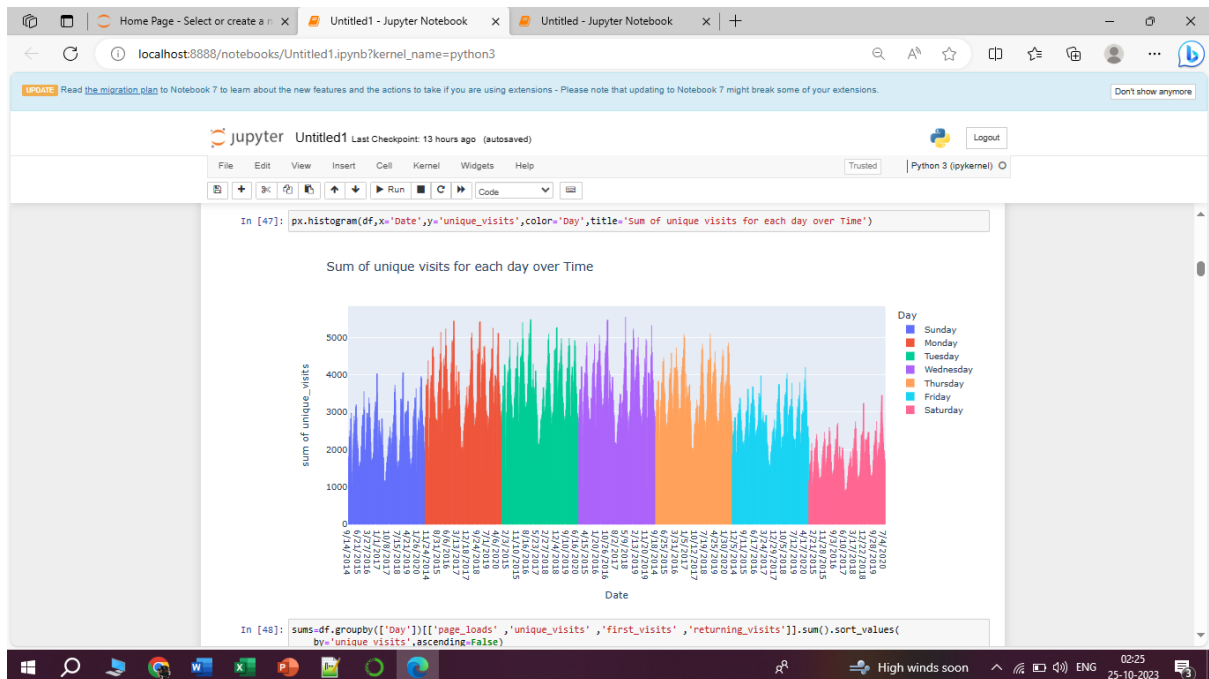
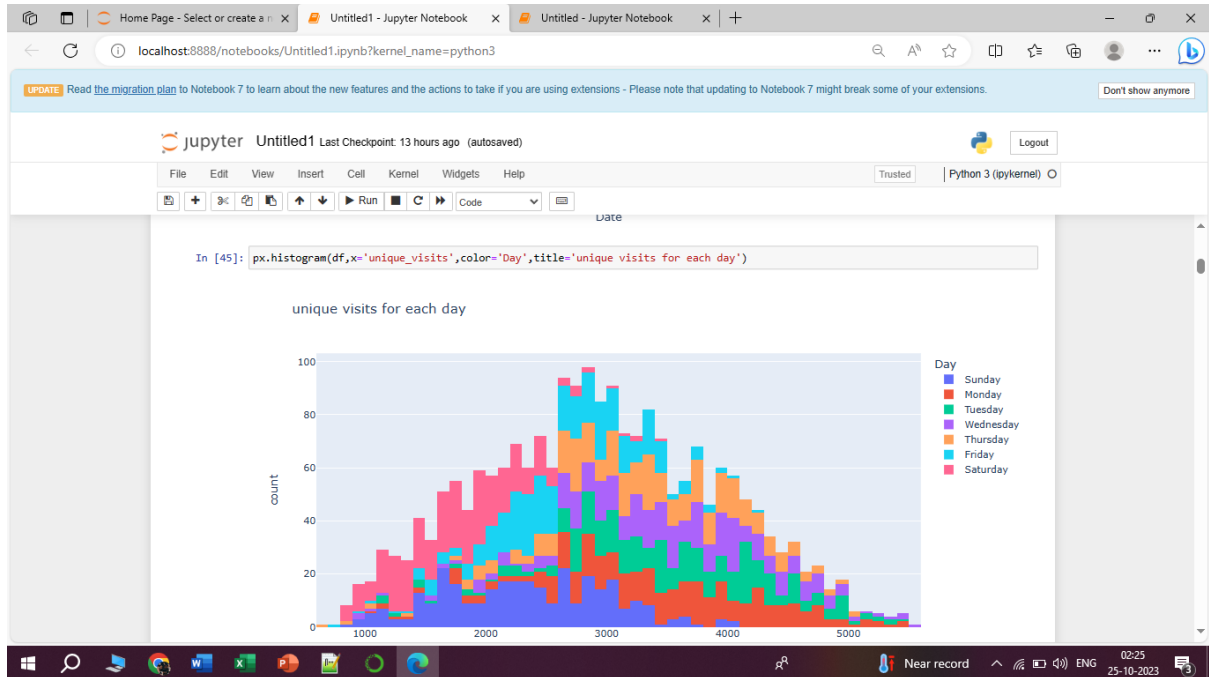
Showcasing the general information about the Data set for the better understanding for analysis.

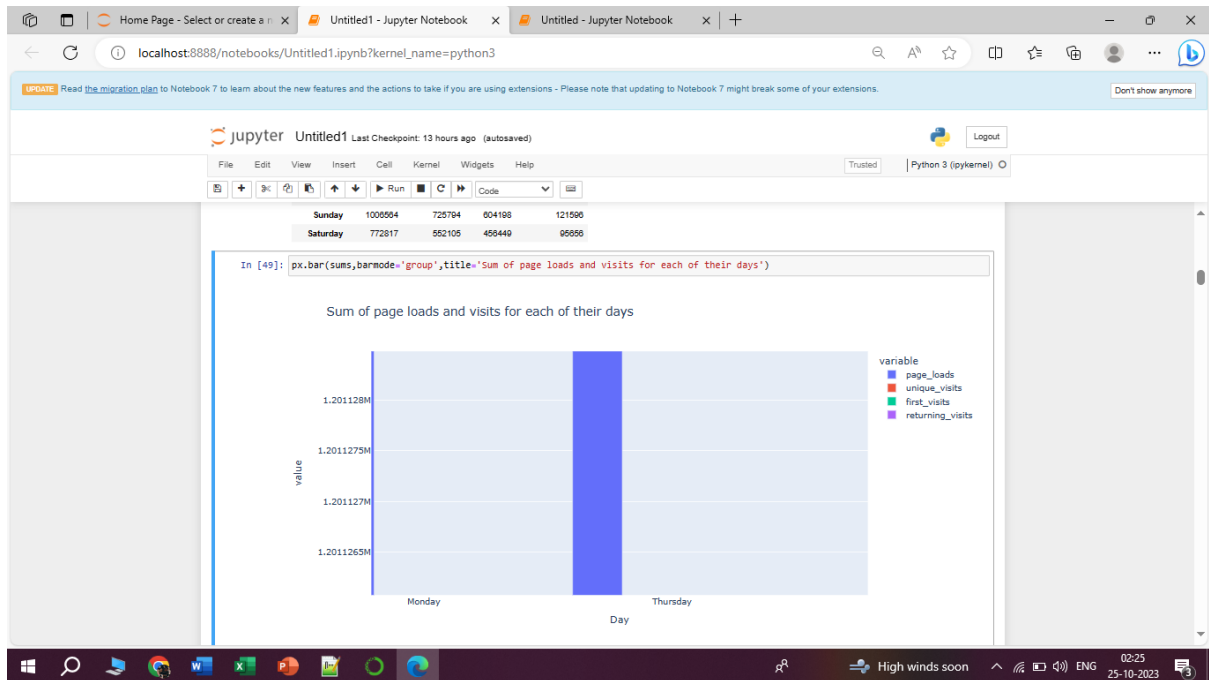


Presenting the **Time Series** for Page loads, Unique Visits, First time, Returning with the help of provided Data set to analyse the Traffic

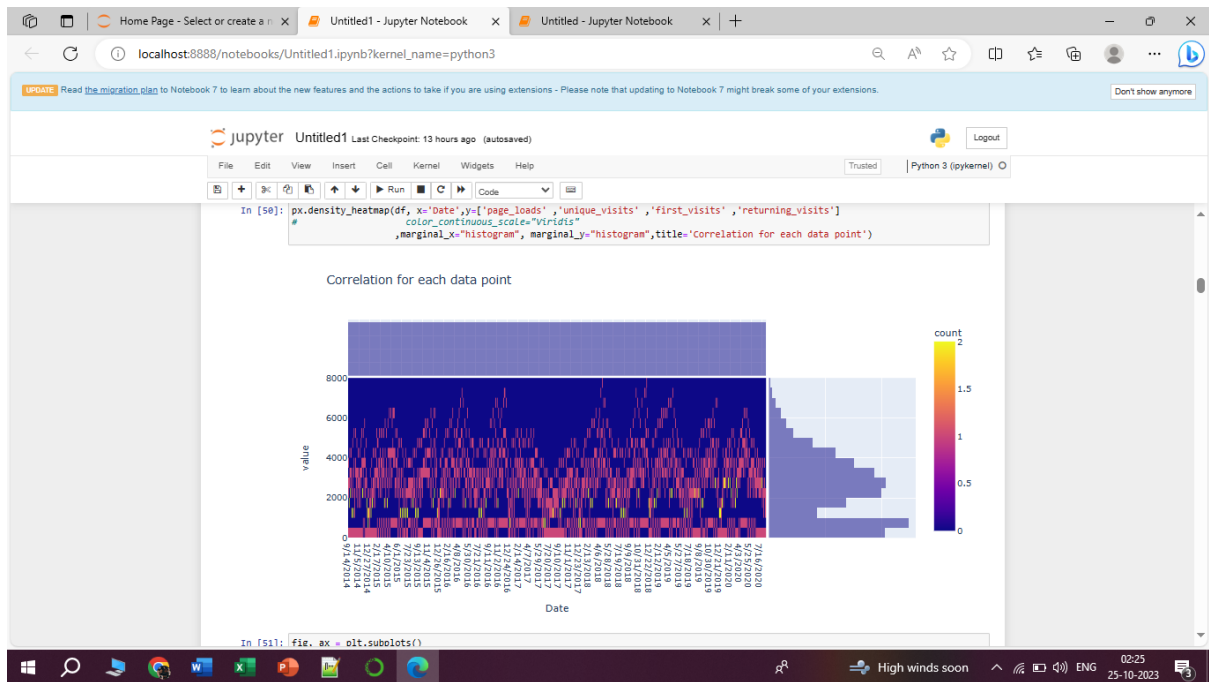


# Calculating the Sum values for the Unique Visitors with the help of visualizaition tool like Bar Graph and Histogram.

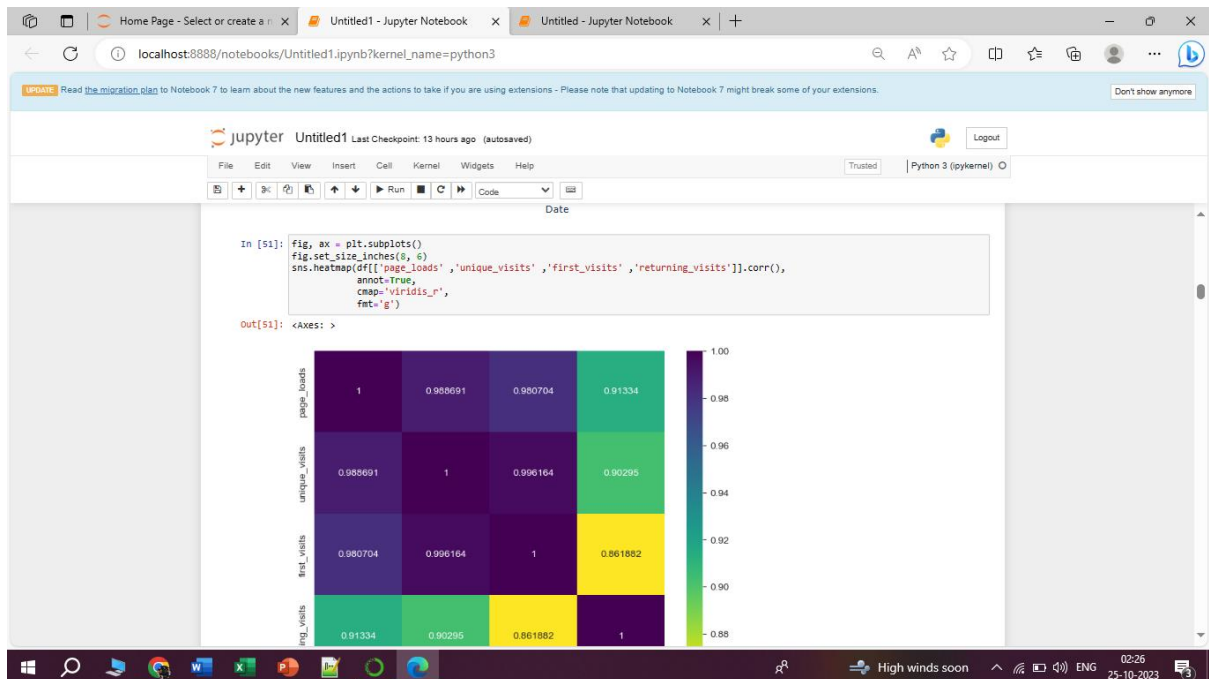




Visualizing the Sum of the Page loads in the bar graph to under stand the Traffic in the week.



Representing the Correlation of the Whole Data set for **EDA** Predcition to know the Stuffs that used by the visitors for there necessity calculation.



## **Conclusion**

Overall Analysis of Development Part 2 with help of IBM Cognos Analysis and Python Integration. We can able to understand that the visitors traffic varies on the basis of week-ends and week-days under there use and necessity for personal or professional use.