PHASE 4 : DEVELOPMENT PART 2

WEBSITE TRAFFIC ANALYSIS

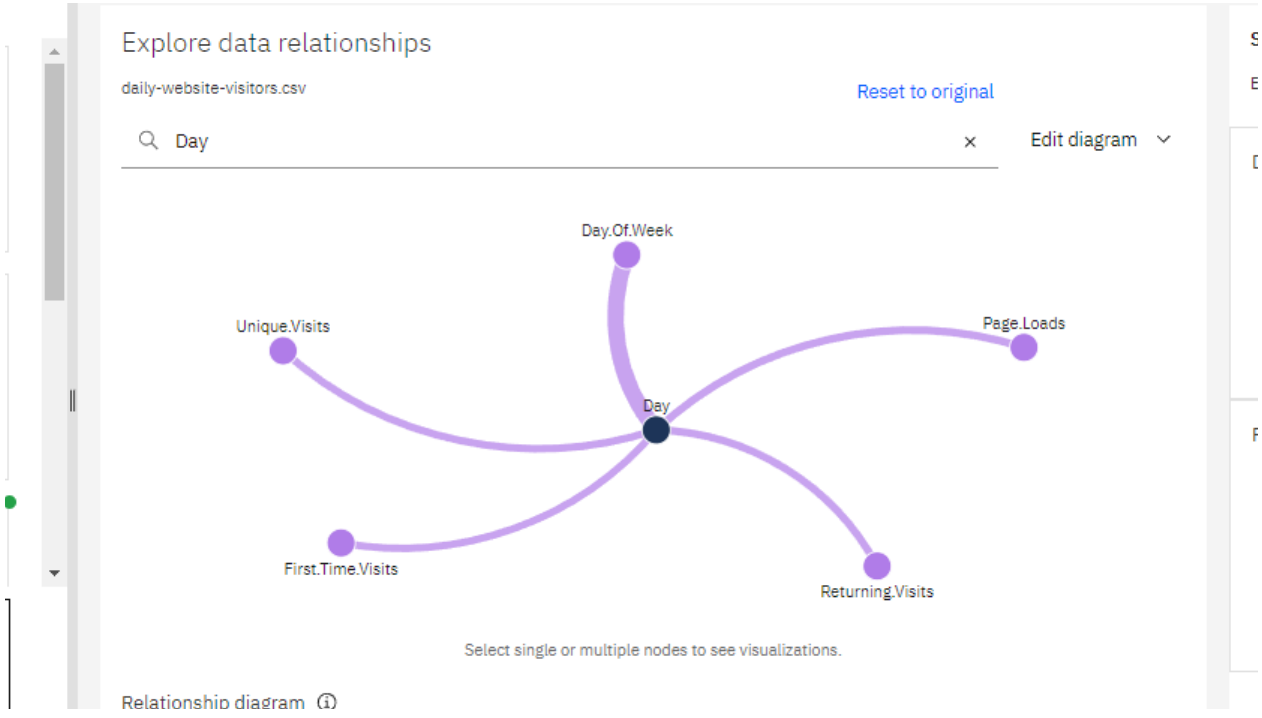
PRIYADARSHINI ENGINEERING COLLEGE

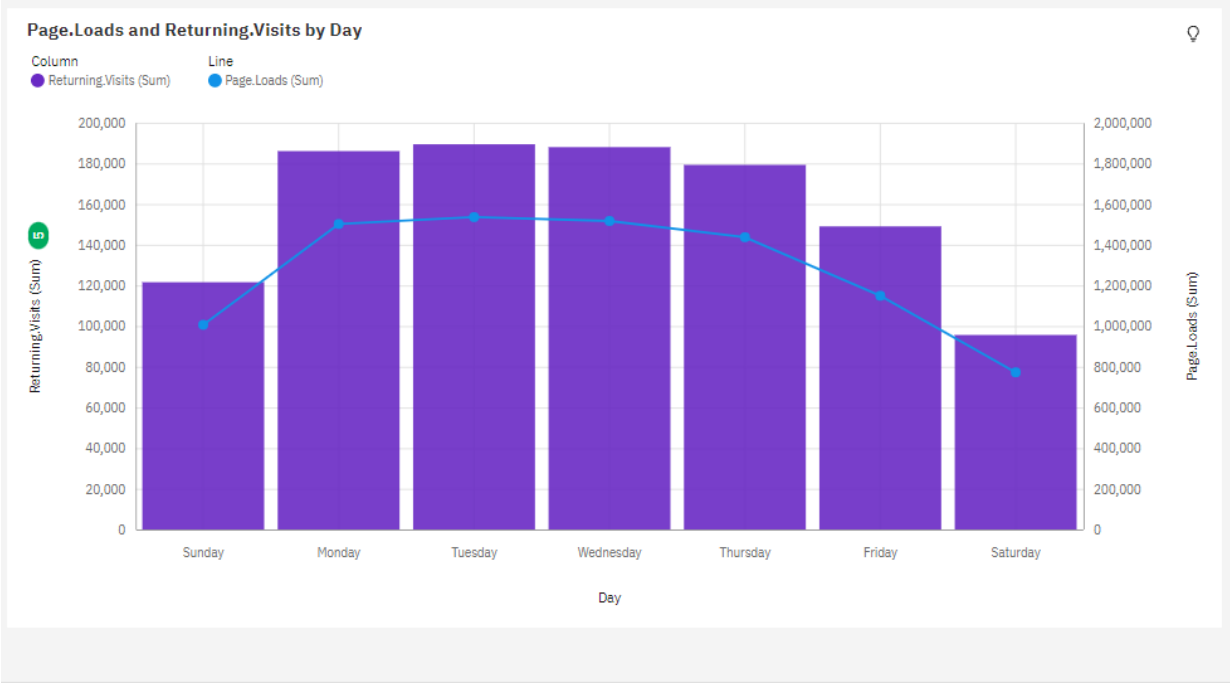
Website traffic analysis is the process of collecting, examining, and interpreting data related to the visitors and interactions on a website. Itprovides invaluable insights into user behavior, preferences, and trends, helping organizations make informed decisions, optimize their online presence, and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

Abstract:

This project aims to analyze website traffi c data for insights into user behavior, popular pages, and traffi c 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





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.

For **Page.Loads**, the most significant values of **Day** are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective **Page.Loads** values add up to over 7.1 million, or 80.1 % of the total.

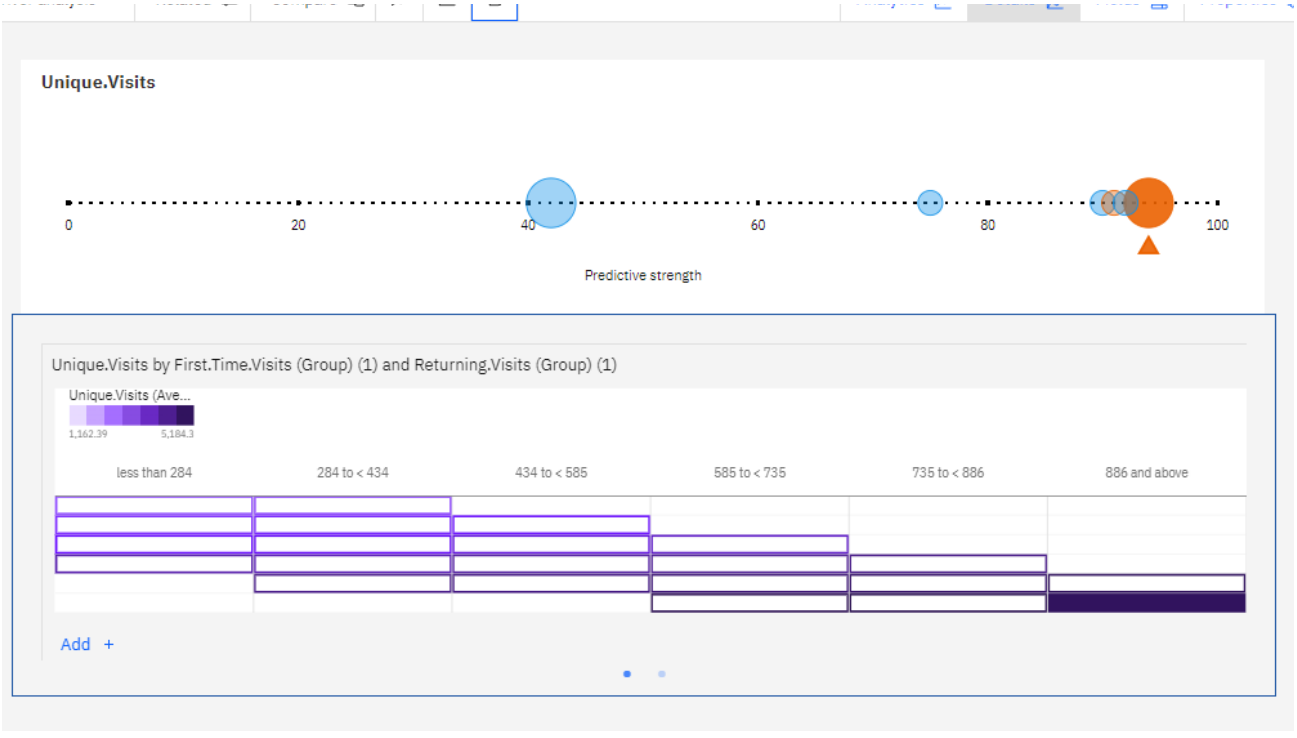
Across all values of **Day.Of.Week**, the average of **Page.Loads** is over four thousand.

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).



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.

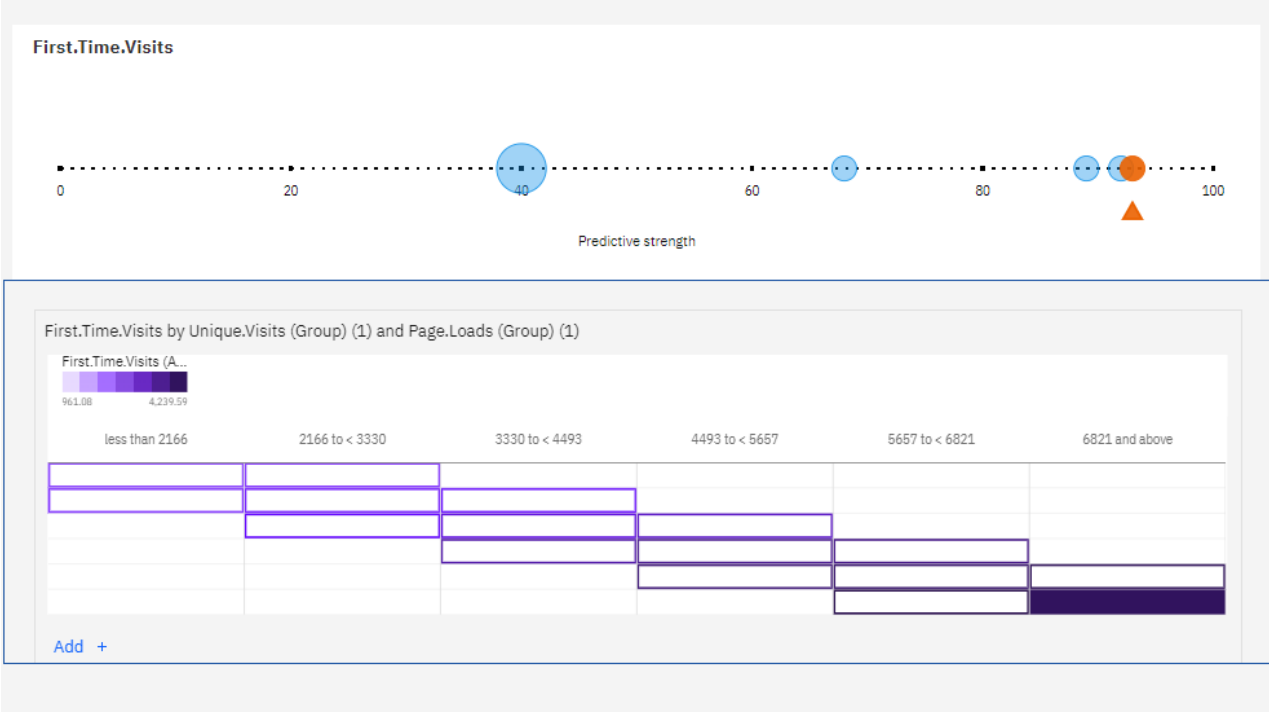
First.Time.Visits (Group) (3) and **Returning.Visits (Group) (2)** strongly affect **Unique.Visits** (96%).

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**.

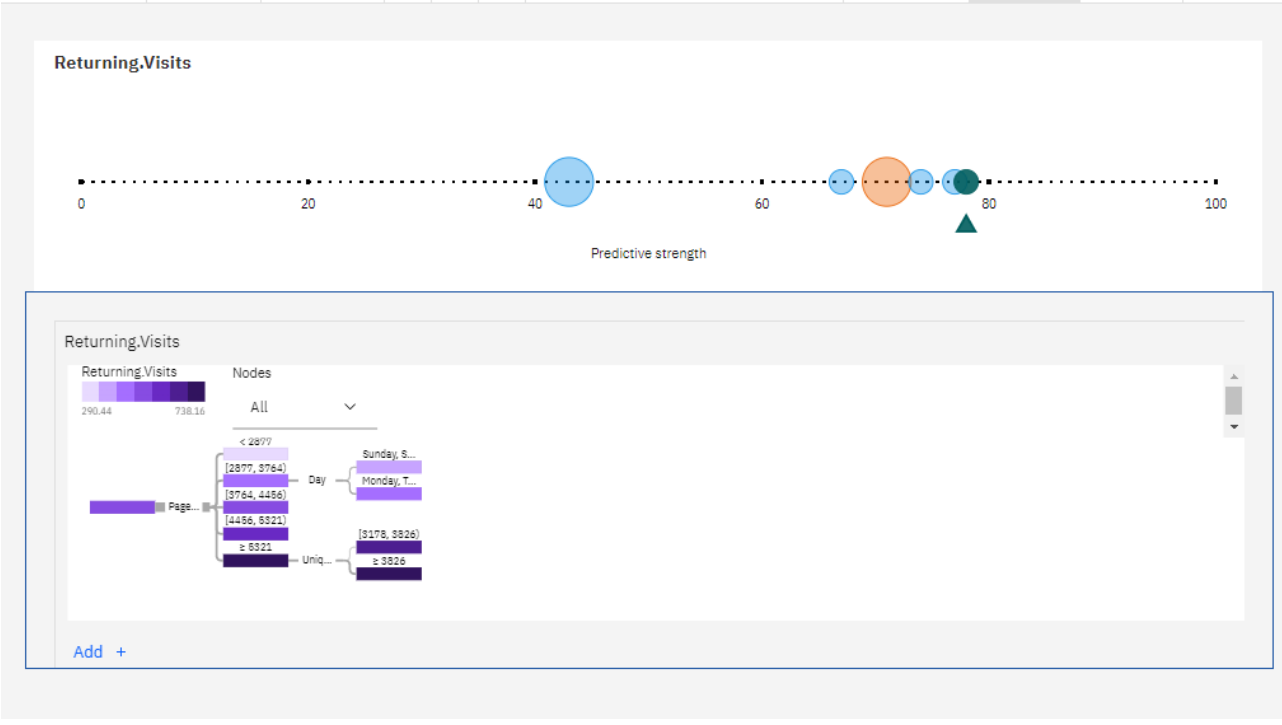


Unique.Visits is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.

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).

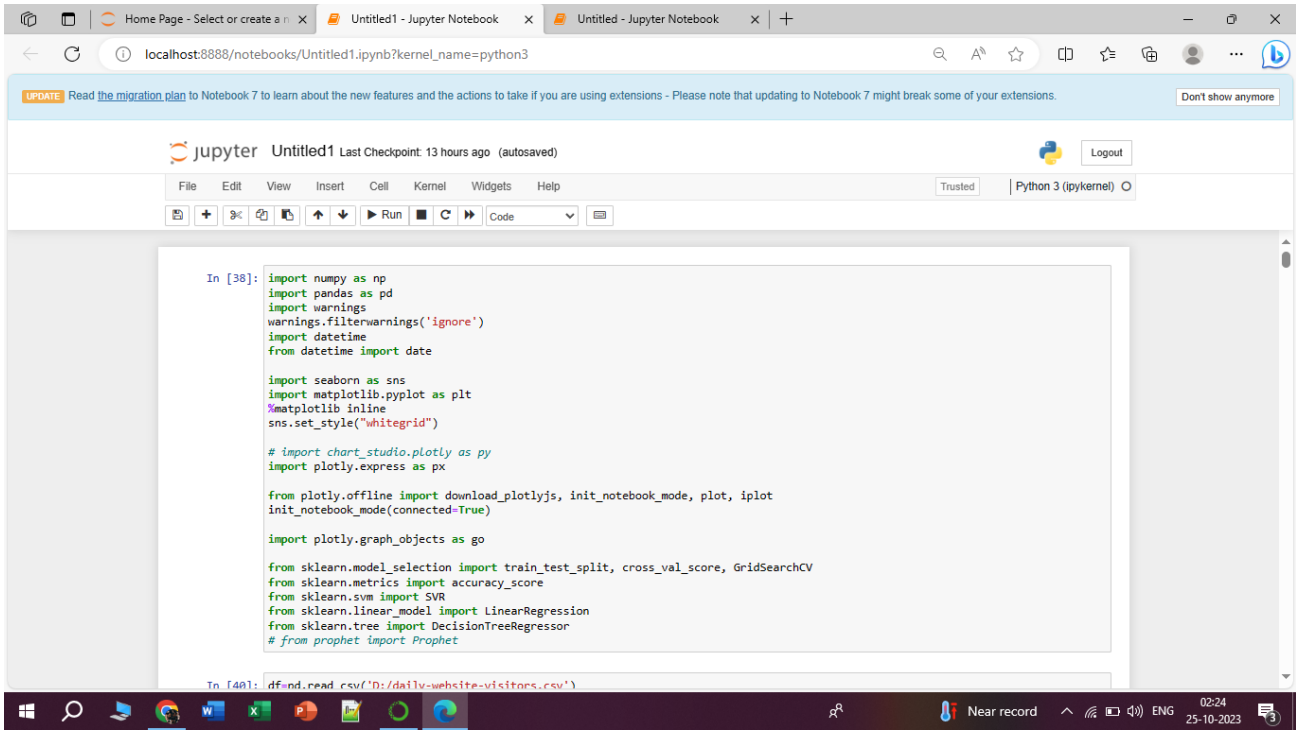
434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).

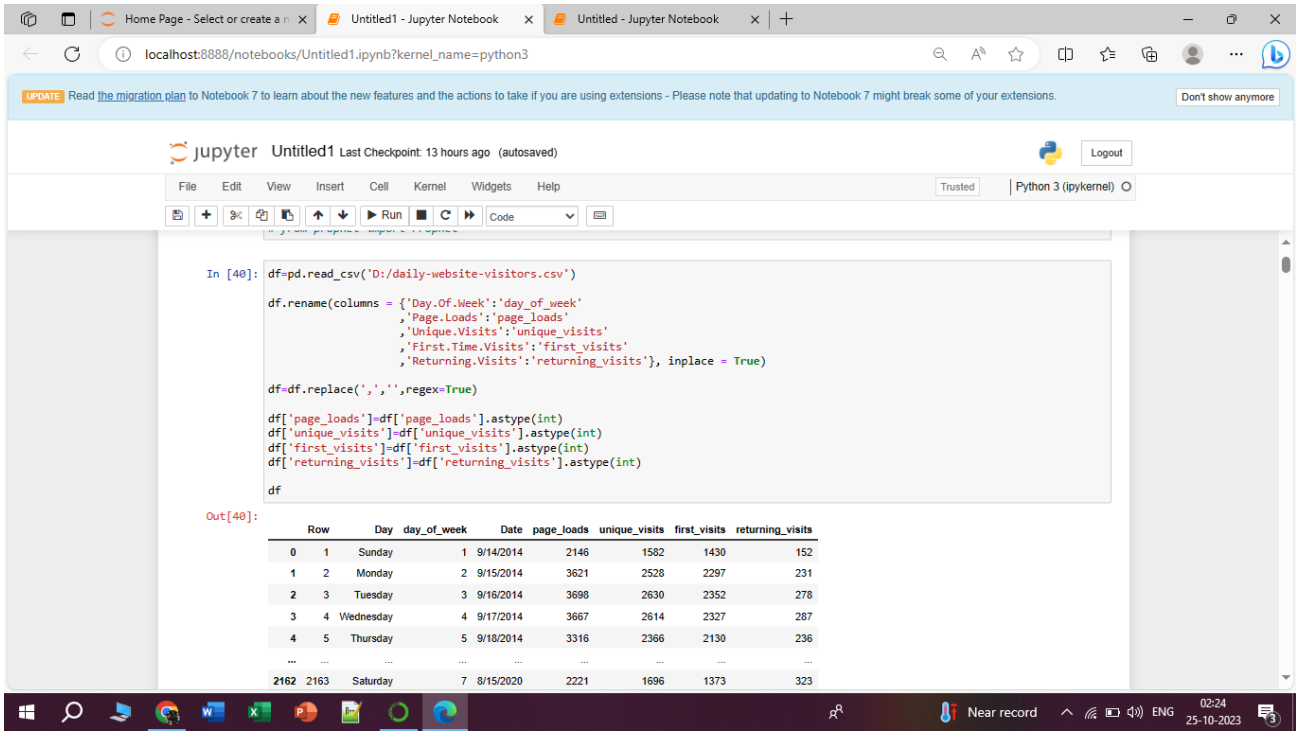
There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.

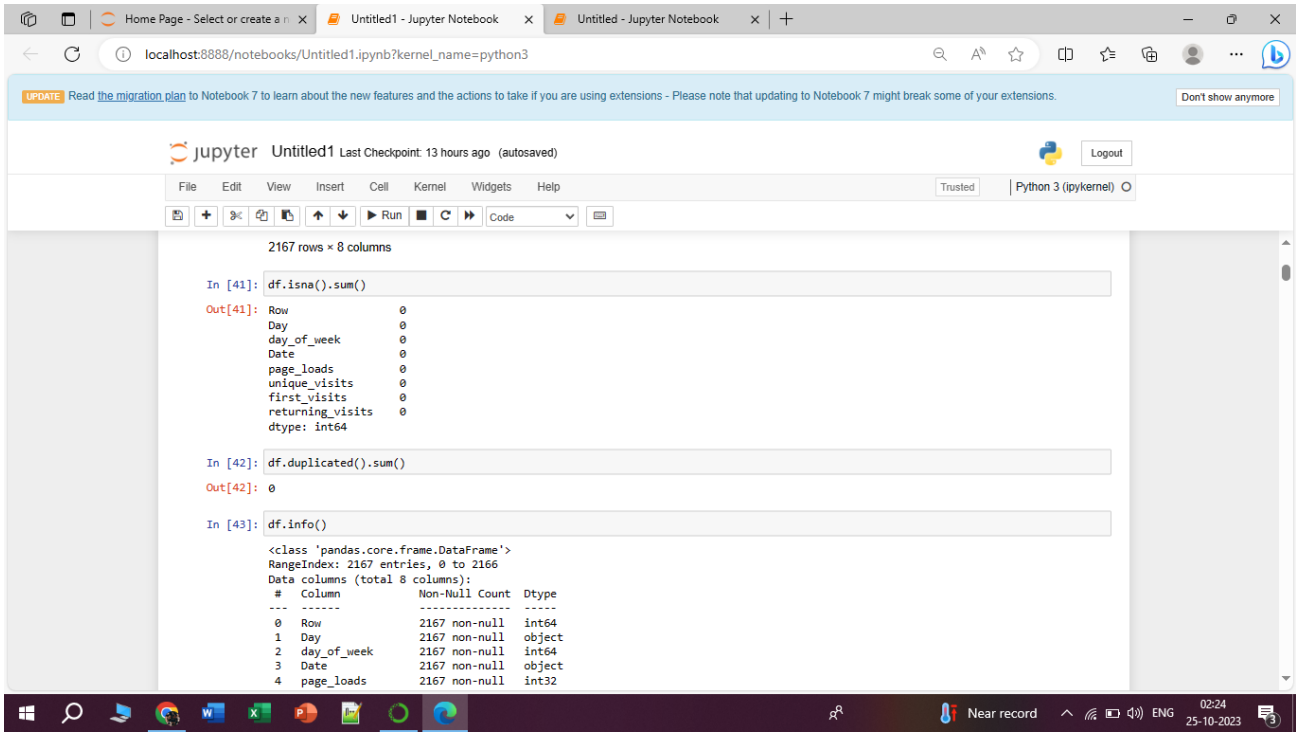


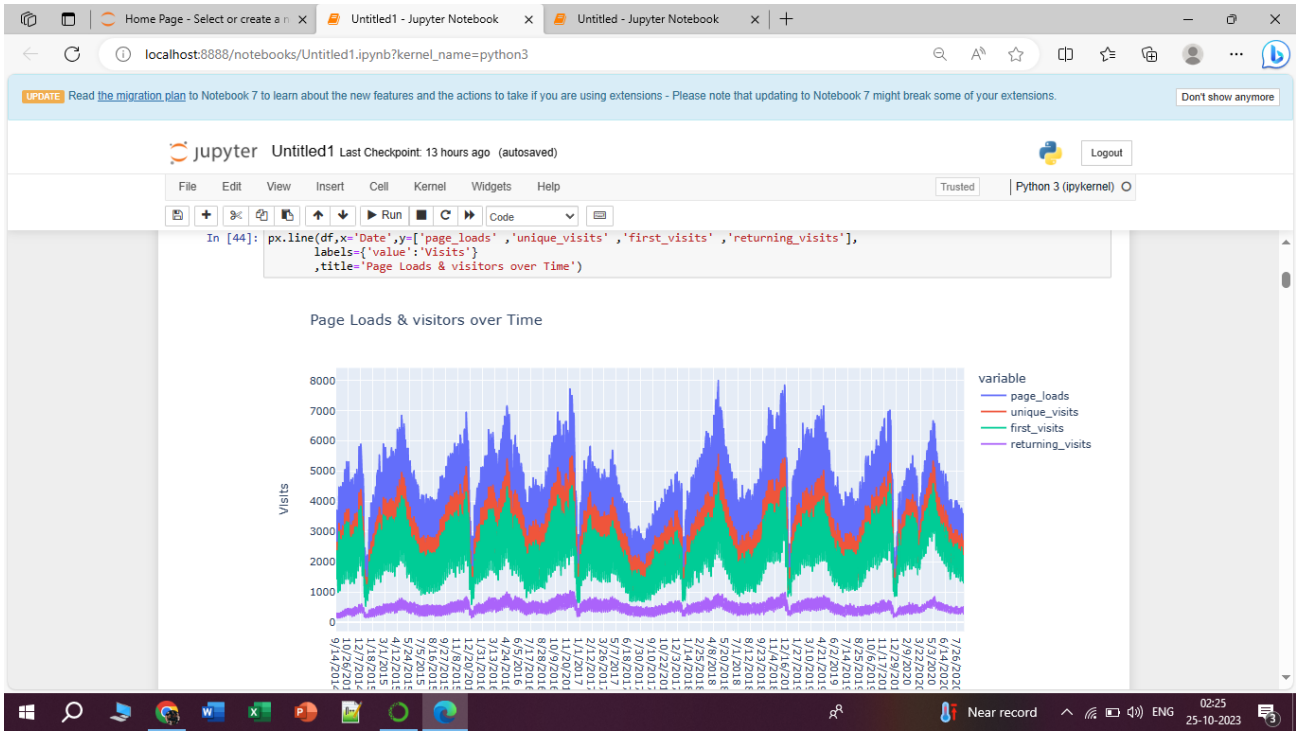
Page.Loads, **Unique.Visits**, and **Day** predict **Returning.Visits** with a strength of 78.1%.

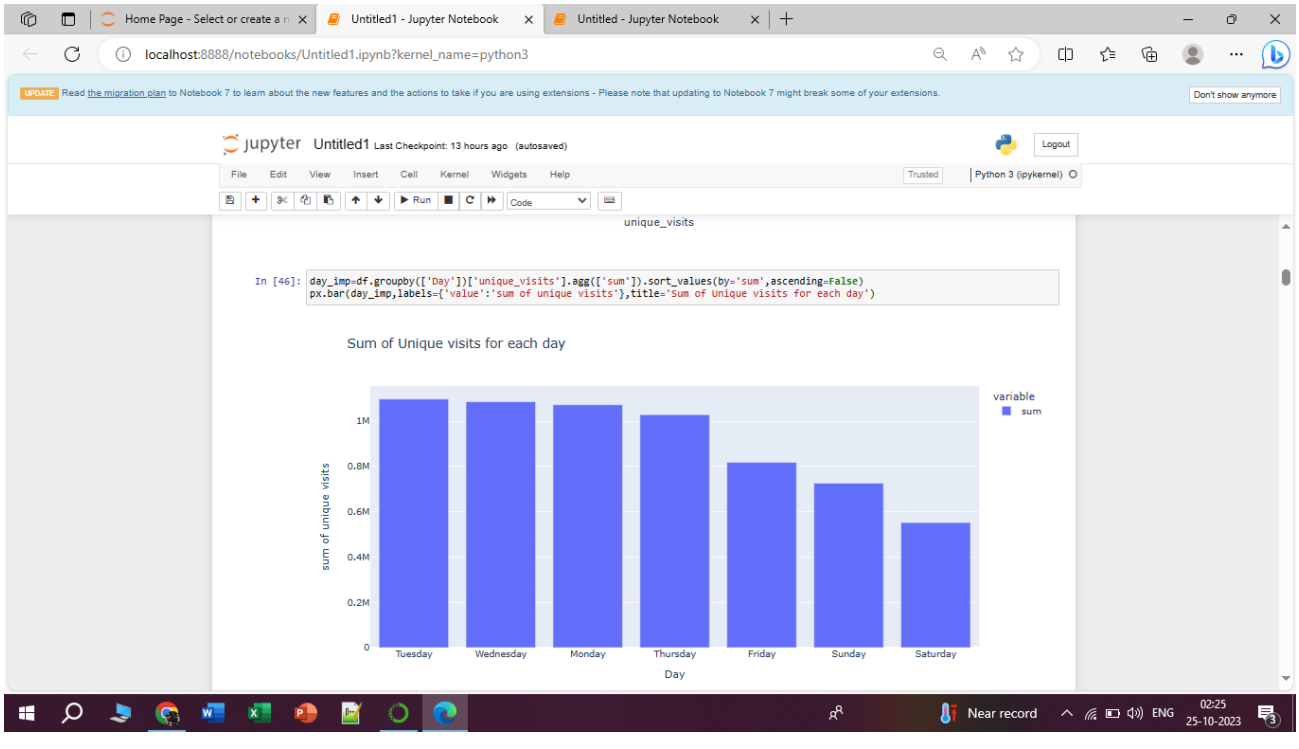
Page.Loads is the most significant predictor of **Returning.Visits** being three times better than any other field.




Home Page - Select or create a n
Untitled1 - Jupyter Notebook
Untitled - Jupyter Notebook
localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.
Don't show anymore
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
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
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
Near record
ENG
02:24
25-10-2023


Home Page - Select or create a n
Untitled1 - Jupyter Notebook
Untitled - Jupyter Notebook
localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.
Don't show anymore
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2167 rows × 8 columns
In [41]:
df.isna().sum()
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
In [42]:
df.duplicated().sum()
Out[42]:
0
In [43]:
df.info()
<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
Near record
ENG
02:24
25-10-2023





Conclusion

Website traffic analysis using IBM Cognos Analytics allows businesses to gain valuable insights into the performance of their websites. This analysis helps in making data-driven decisions to enhance user experience,

optimize content, and improve marketing strategies. IBM Cognos Analytics provides robust tools for tracking and visualizing web traffic data, enabling organizations to monitor key metrics, detect trends, and make informed decisions to drive business success. It offers the ability to create interactive reports and dashboards, making it easier for teams to collaborate and act on the insights derived from website traffic data, ultimately leading to improved online performance and user engagement.