
RECEIVE
SECURITY
INCREASE
PAGES
WEB
INTERNET
STATISTICS
DATA
SEO
MONITOR
TRAFFIC
VISITORS
SEND
OUTGOING
BANDWIDTH
WEBSITE
NUMBER
INCOMING
MEASURING
ANALYTICS
TREND
POPULARITY

website traffic analysis with IBM

Cognos Final Documentation

Team member

511921104063/Santhosh

1. INTRODUCTION:

Website analysis with IBM Cognos involves using Cognos analytics tools to gain insights into website performance, user behavior, and other key metrics. IBM Cognos is a business intelligence and performance management software suite that enables users to extract and analyze data from various sources, including websites, to make informed business decisions.

1.1 Data Collection:

Website analysis begins with collecting relevant data from your website. This data can include website traffic, user demographics, page views, bounce rates, and conversion rates. Tools like Google Analytics can collect this data and store it for further analysis.

1.2 Data Integration:

IBM Cognos allows you to integrate data from various sources, including your website analytics data. By integrating this data, you can create a comprehensive view of your website performance alongside other business data, such as sales figures or customer demographics.

1.3 Data Modeling:

After integrating the data, you can create data models using Cognos Framework Manager. Data modeling involves organizing and structuring the data in a way that makes it easier to analyze. You can define relationships, calculations, and business rules to prepare the data for analysis.

1.4 Reporting and Dashboard Creation:

Cognos offers a user-friendly interface for creating interactive reports and dashboards. You can use drag-and-drop features to visualize website analytics data. For instance, you can created.

reports that show website traffic trends over time, popular pages, user engagement, and conversion rates. Dashboards provide a real-time overview of your website's performance metrics in a visually appealing format.

1.5 Data Analysis and Exploration:

Cognos allows users to explore data in-depth. You can perform ad-hoc analysis, drill down into specific metrics, and identify patterns and trends. Advanced analytics tools in Cognos enable users to apply statistical methods and predictive analytics to gain valuable insights into user behavior and website performance.

1.6 Collaboration and Sharing:

Cognos enables collaboration among team members. You can share reports and dashboards with stakeholders, allowing them to interact with the data and gain insights. Scheduled reports and automated alerts can keep stakeholders informed about critical website metrics.

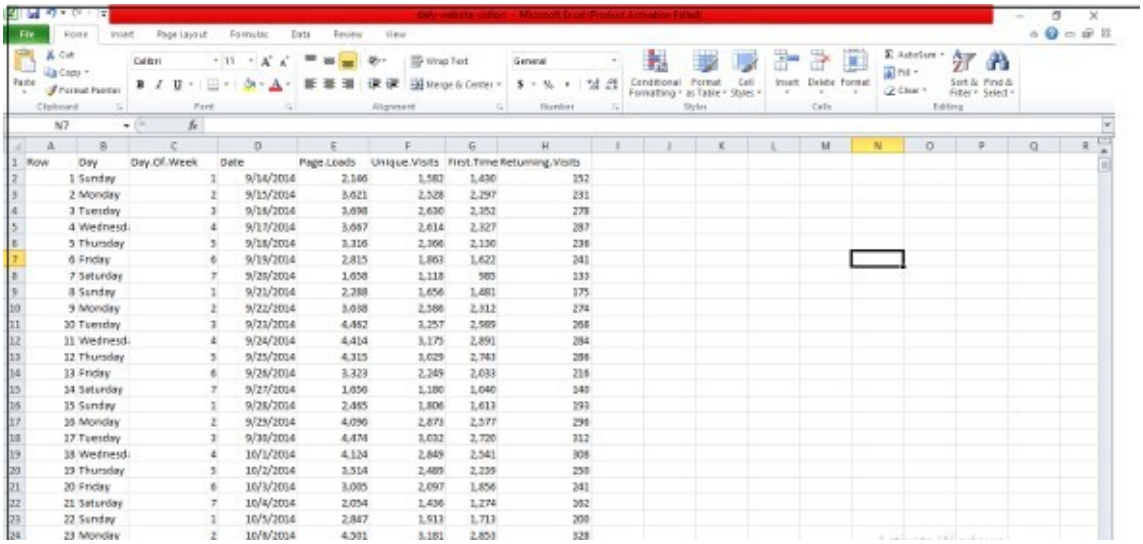
1.7 Performance Optimization:

By analyzing website data using Cognos, you can identify areas for improvement. For instance, if you observe a high bounce rate on certain pages, you can investigate further to understand why users are leaving those pages. Data-driven insights can guide website optimization efforts to enhance user experience and increase conversions.

2.0 DATASET:

The dataset used for this analysis is "Daily website visitors" document by kaggle.com website.

LINK: https://www.kaggle.com/datasets/bobnau/daily-website-visitors



Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
2	Monday	2	9/15/2014	3,621	2,528	2,297	231
3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
4	Wednesd	4	9/17/2014	3,667	2,614	2,327	287
5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
6	Friday	6	9/19/2014	2,815	1,863	1,622	241
7	Saturday	7	9/20/2014	1,658	1,118	985	133
8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
9	Monday	2	9/22/2014	3,698	2,586	2,312	274
10	Tuesday	3	9/23/2014	4,482	3,257	2,989	268
11	Wednesd	4	9/24/2014	4,414	3,175	2,891	284
12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
13	Friday	6	9/26/2014	3,323	2,249	2,033	216
14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
15	Sunday	1	9/28/2014	2,485	1,806	1,613	193
16	Monday	2	9/29/2014	4,096	2,873	2,577	296
17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
18	Wednesd	4	10/1/2014	4,124	2,849	2,541	308
19	Thursday	5	10/2/2014	3,514	2,489	2,239	250
20	Friday	6	10/3/2014	3,005	2,097	1,856	241
21	Saturday	7	10/4/2014	2,054	1,436	1,274	162
22	Sunday	1	10/5/2014	2,847	1,913	1,713	200
23	Monday	2	10/6/2014	4,501	3,181	2,853	328

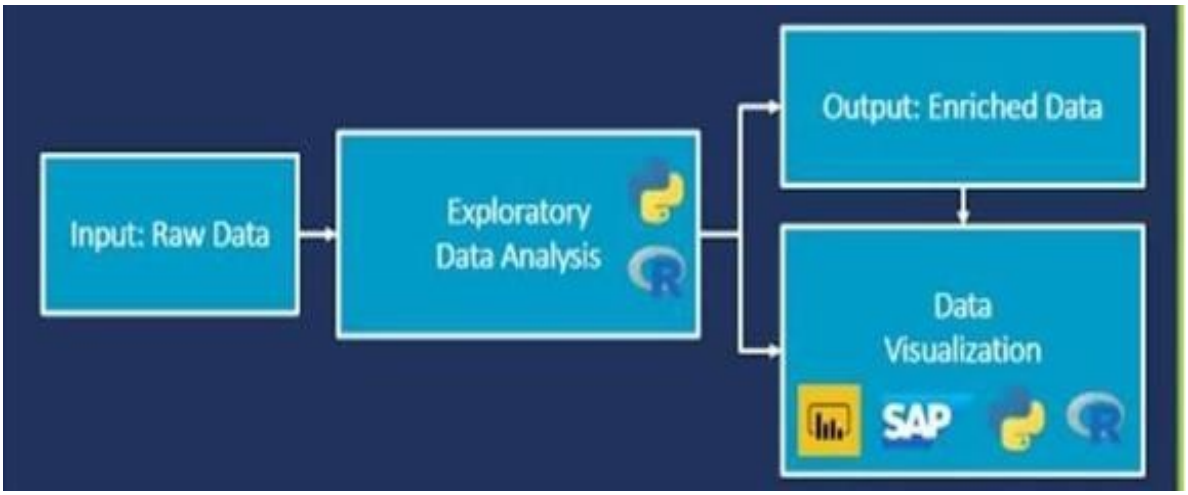
3.0 INNOVATION OBJECTIVES :

3.1 Data Acquisition:

Download and import the "Daily Website Visitors" dataset from Kaggle into your analytics environment.

3.2 Exploratory Data Analysis (EDA):

Conduct EDA to understand the dataset's characteristics, patterns, and correlations. Visualize key metrics and trends in website traffic.



3.3 Analysing the Data:

Regularly analyze data collected from the real-time analytics tools and experiments. Adjust strategies

based on insights derived from the dataset.

4.0 INCLUSION OF MACHINE LEARNING MODEL

4.1 Regression Analysis:

Linear regression or more advanced methods like polynomial regression can be used to model and predict traffic trends over time.

4.2 Time Series Analysis:

Techniques like ARIMA (AutoRegressive Integrated Moving Average) can help in forecasting website traffic based on historical data.

4.3 Classification Algorithms:

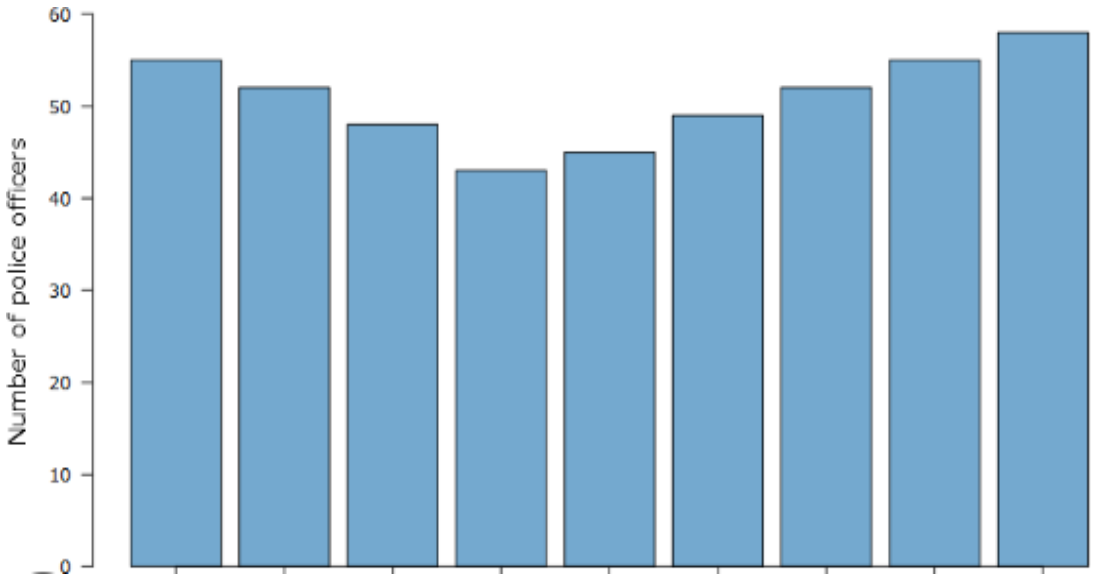
These can be used to categorize website visitors, such as decision trees, random forests, or support vector machines, to identify different user segments or traffic sources.

5.0 VISUALIZATIONs

5.1 Bar Charts:

Bar charts are useful for displaying metrics like the number of page views, unique visitors, or bounce rates over a specific time period.

Example of barchart:



5.2 Line Charts:

Line charts are effective for showing trends in website traffic data, such as changes in visitor numbers over time.

5.3 Pie Charts:

Pie charts can be used to represent the distribution of traffic sources, showing the percentage of traffic coming from direct visits, search engines, social media, etc.

5.4 Heat Maps:

Heat maps can provide insights into user engagement by showing which parts of a webpage receive the most clicks or interaction.This helps in optimizing the website's layout.

6.0 Problem Statement:

The problem of "Website Traffic Analysis" lies in the need for organizations to effectively

understand and leverage user behavior on their websites

The dataset given is "Daily web visitors" from kaggle.com

Link:https: //www.kaggle.com/datasets/bobnau/daily-website-visitors

6.1 Analytics using IBM cognos tool:

To perform website traffic analysis using IBM Cognos, start by collecting your website traffic data from various sources, such as web server logs or web analytics tools. Ensure that this data is well-structured and accessible, typically stored in a database or spreadsheet.

Next, create a data model in IBM Cognos Framework Manager to represent your website traffic data. This model will serve as the foundation for your analysis. Once your data model is in place, use IBM Cognos Report Studio to create reports and visualizations that will help you analyze and present the data effectively.

Define the key metrics and key performance indicators (KPIs) you want to focus on in your analysis. This could include metrics like page views, unique visitors, bounce rate, conversion rate, and more. With your key metrics in mind, set up filters and prompts that allow users to interact with the data. This can include selecting specific date ranges or traffic sources to narrow down the focus of the analysis.

Make sure to incorporate drill-down capabilities in your reports. This feature enables users to explore the data at different levels of detail. For example, users can start with an overview of website traffic and then drill down into specific pages or time periods for a more granular view.

Schedule your reports to run automatically at specific intervals and distribute them to relevant stakeholders via email or other preferred channels. Regular monitoring and analysis of the website traffic data will help identify trends, patterns, and insights that can inform decision- making and strategy.

Consider creating custom calculations and calculated fields within IBM Cognos to address specific analytical needs that are unique to your website traffic analysis.

6.2 IBM Cognos Chart Patterns:

- Column Chart: Column charts are used to display data in vertical bars. They are suitable for comparing values across categories or showing data changes over time.
- Bar Chart: Similar to column charts, bar charts display data using horizontal bars. They are useful when you want to make comparisons between categories or show ranking.
- Line Chart: Line charts represent data as a series of data points connected by lines. They are excellent for showing trends and changes in data over time.
- Area Chart: Area charts are like line charts but with the area under the lines filled, making them suitable for displaying data distributions over time.
- Pie Chart: Pie charts are used to display the parts of a whole. They are effective for showing the percentage distribution of categories within a dataset.
- Doughnut Chart: Doughnut charts are a variant of pie charts with a hole in the center. They show the same part-to-whole relationships but allow for additional data to be displayed in the center.
- Scatter Plot: Scatter plots display individual data points as dots on a two-dimensional graph. They are used to show the relationship between two variables.
- Bubble Chart: Bubble charts are like scatter plots but with bubbles of varying sizes to represent a third variable. They are useful for visualizing three dimensions of data.
- Stacked Chart: Stacked charts, both column and bar variants, allow you to display data in a way that shows both the total and the contribution of individual data points to that total.
- Combo Chart: Combo charts combine multiple chart types in a single chart. For example, you can have a column chart and a line chart in the same visualization.
- Gauge Chart: Gauge charts provide a way to display data as a dial or gauge, showing progress or performance against a predefined target or goal.
- Heat Map: Heat maps use color to represent values in a matrix or grid. They are helpful for visualizing data density or correlations.
- Waterfall Chart: Waterfall charts are used to show the incremental effect of positive and negative values in a process, such as financial data or project timelines.
- Map Chart: Map charts visualize data on geographical maps. They are useful for showing regional or location-based data.
- Radar Chart: Radar charts, also known as spider or star charts, are used to display multivariate data in a two-dimensional chart with multiple axes.

6.3 IBM Cognos Dashboard:

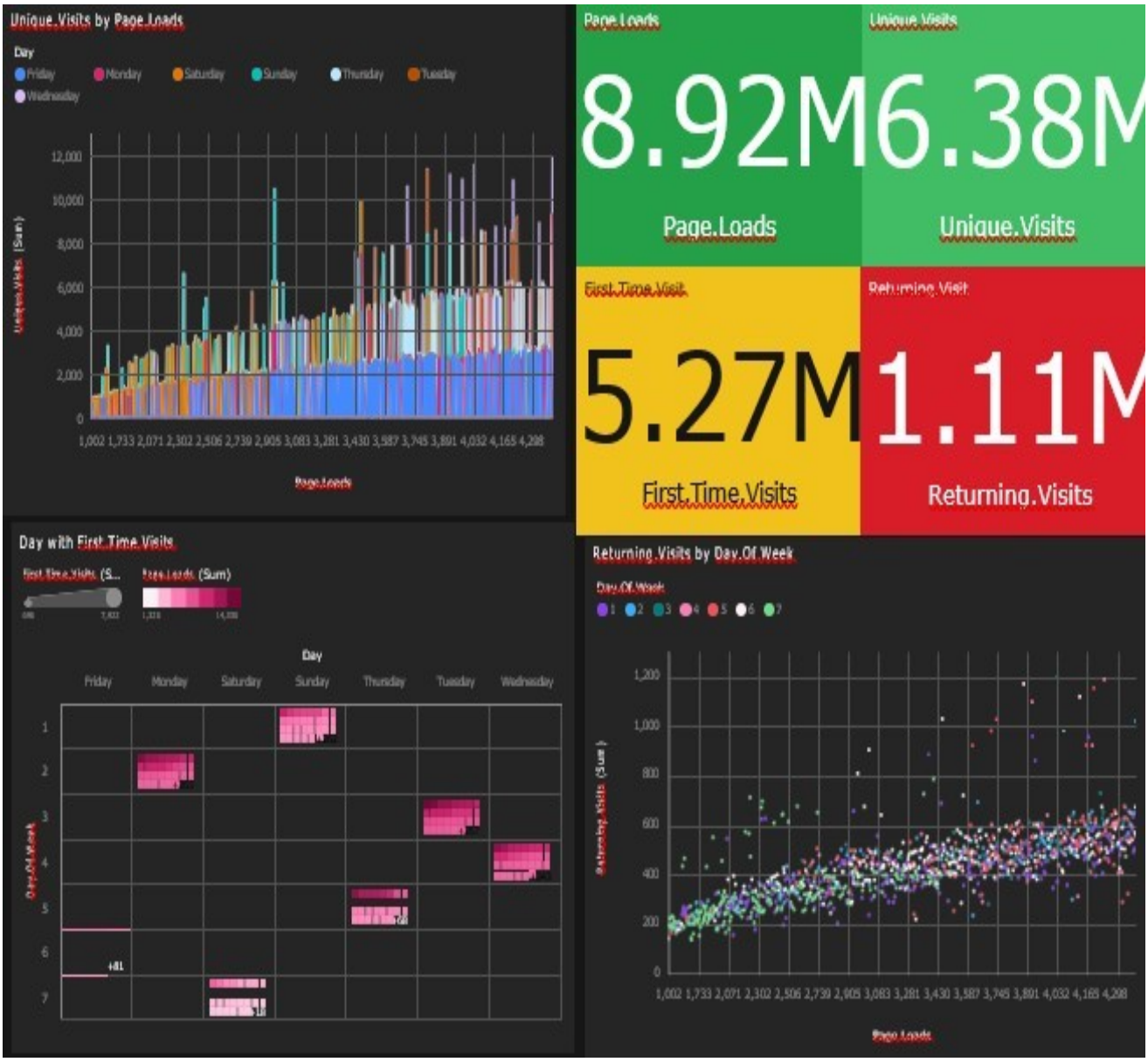
IBM Cognos Dashboard is a robust and user-friendly data visualization tool that empowers organizations to create interactive and insightful dashboards. These dashboards allow users to quickly access and analyze critical data, uncovering actionable insights. One of its key strengths is its ability to bring various data sources together, allowing for a comprehensive view of an organization's performance. With drag-and-drop functionality, users can effortlessly design dashboards by selecting and arranging widgets, charts, tables, and other visual elements to visually represent their data.

These dashboards are designed to be highly customizable, with users able to tailor the layout, colors, and fonts to match their organization's branding. Interactive elements like drill-through and filters allow users to delve deeper into the data and get answers to specific questions on the fly. Furthermore, the dashboards can be accessed on a wide range of devices, including desktop computers, tablets, and smartphones, ensuring that users can stay informed and make data-driven decisions from anywhere.

IBM Cognos Dashboard supports real-time data, which means that users can get up-to-the-minute information to keep their strategies aligned with current conditions. This tool plays a vital role in helping businesses stay competitive by providing them with the necessary data-driven insights to adapt and excel in a dynamic business environment. In conclusion, IBM Cognos Dashboard is an essential asset for organizations aiming to transform data into actionable knowledge through visually appealing and highly interactive dashboards.

6.4 Cognos Analytics using given Dataset:

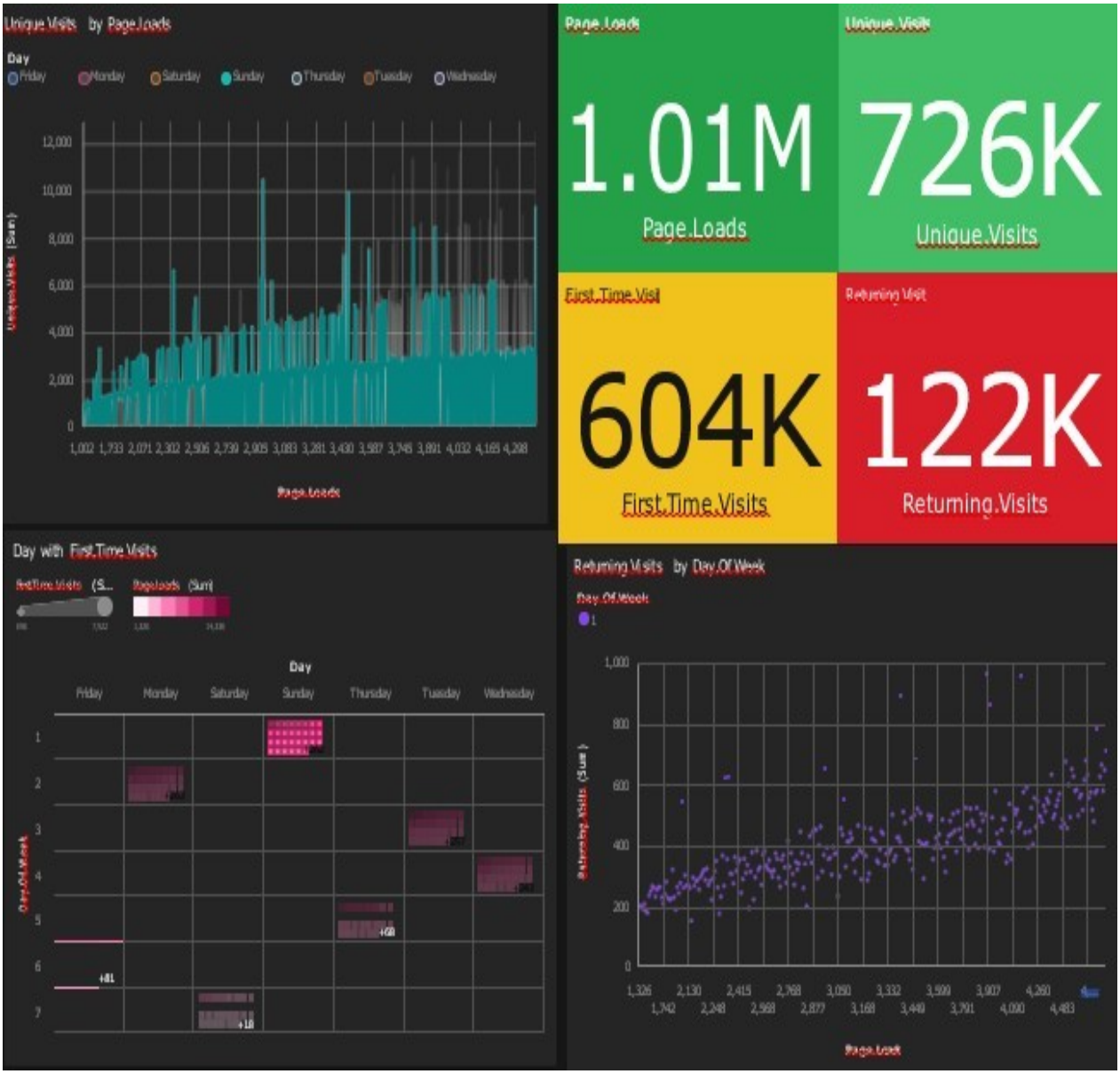
6.4.1 Overall



6.4.2 Sunday Visitors:

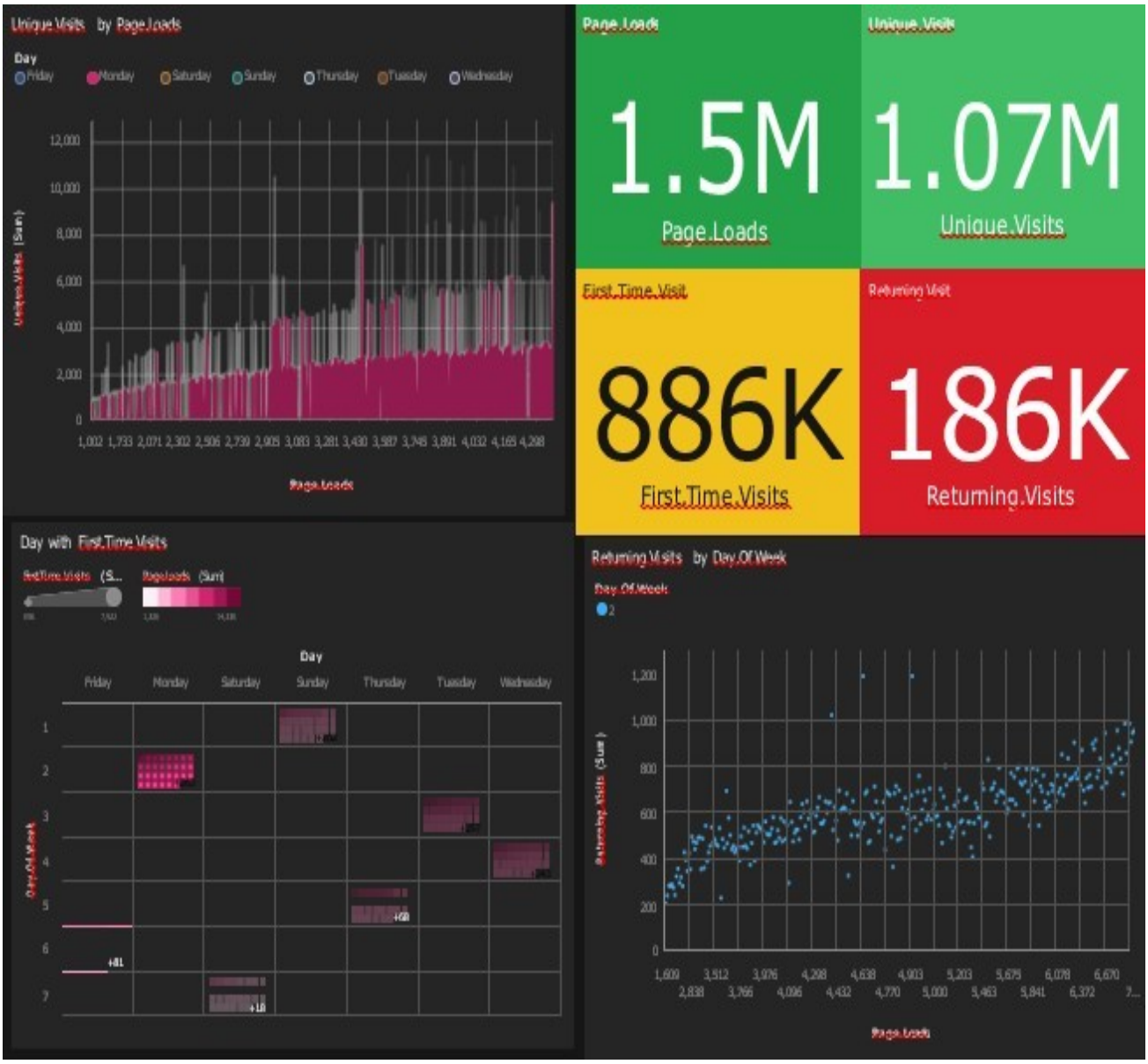
Analyzing Sunday visitors in website traffic analysis involves studying user behavior and engagement patterns on Sundays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week.

This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Sunday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



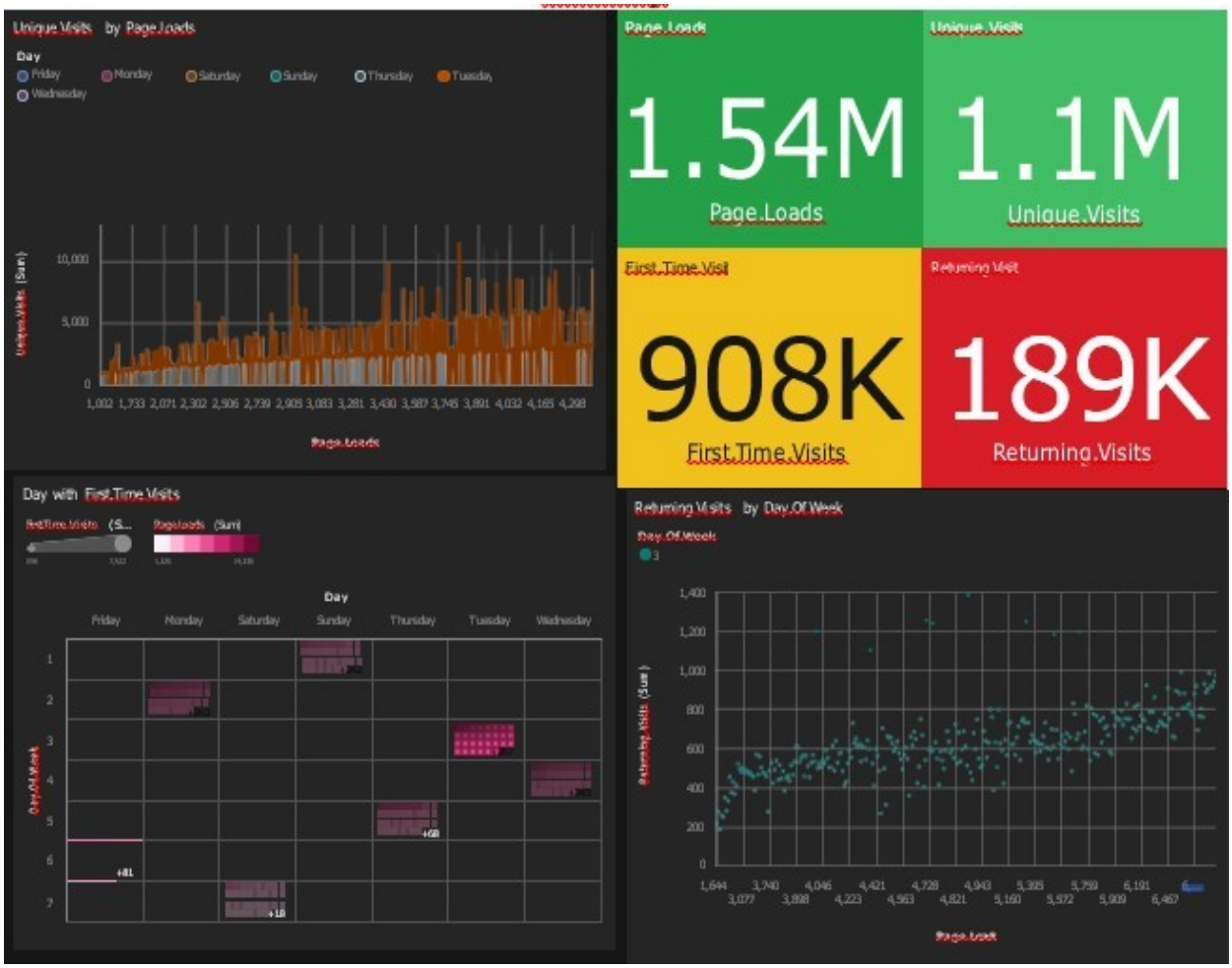
6.4.3Monday Visitors:

Analyzing Monday visitors in website traffic analysis involves studying user behavior and engagement patterns on Mondays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Monday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



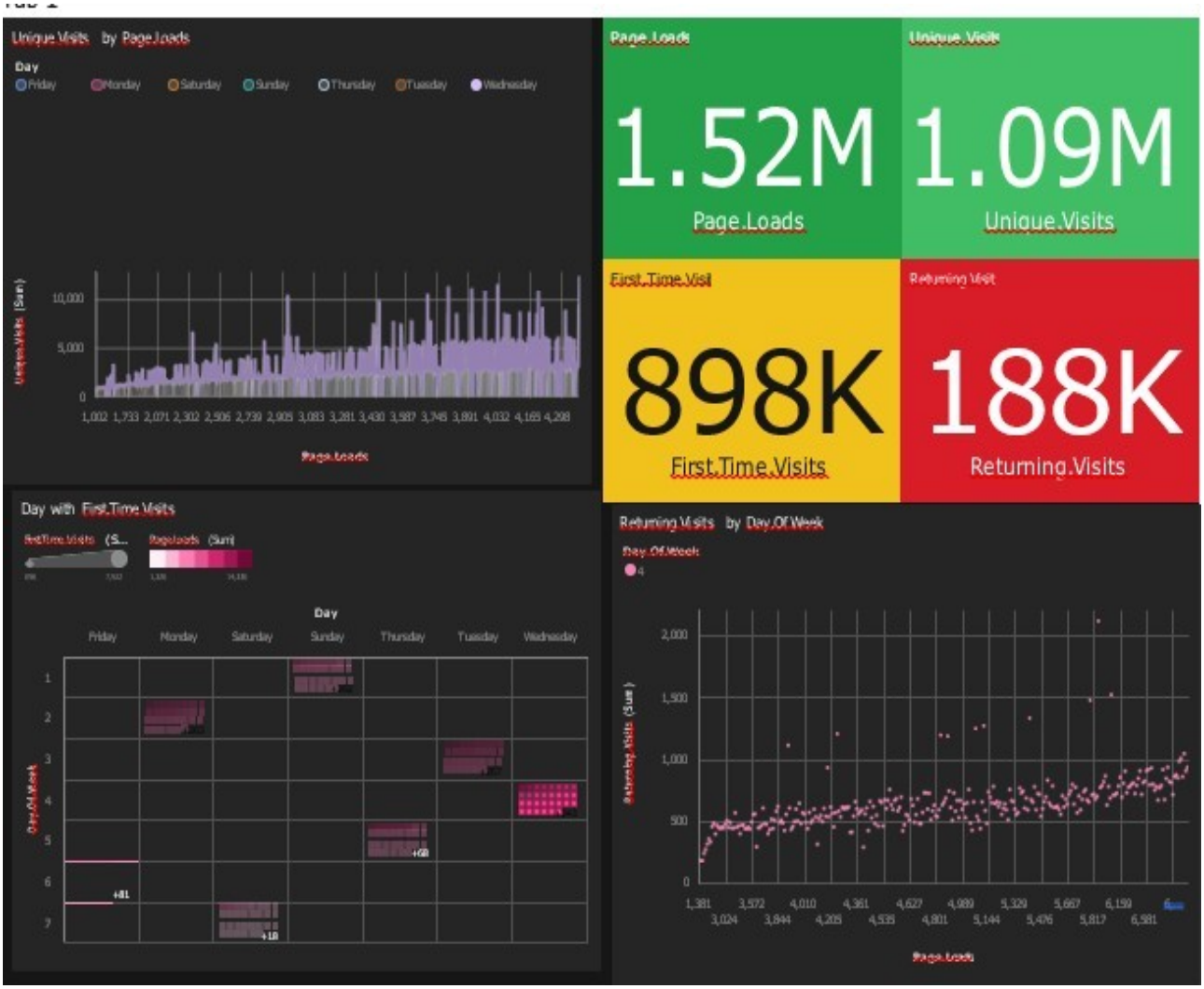
6.4.4Tuesday Visitors :

Analyzing Tuesday visitors in website traffic analysis involves studying user behavior and engagement patterns on Tuesdays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Tuesday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



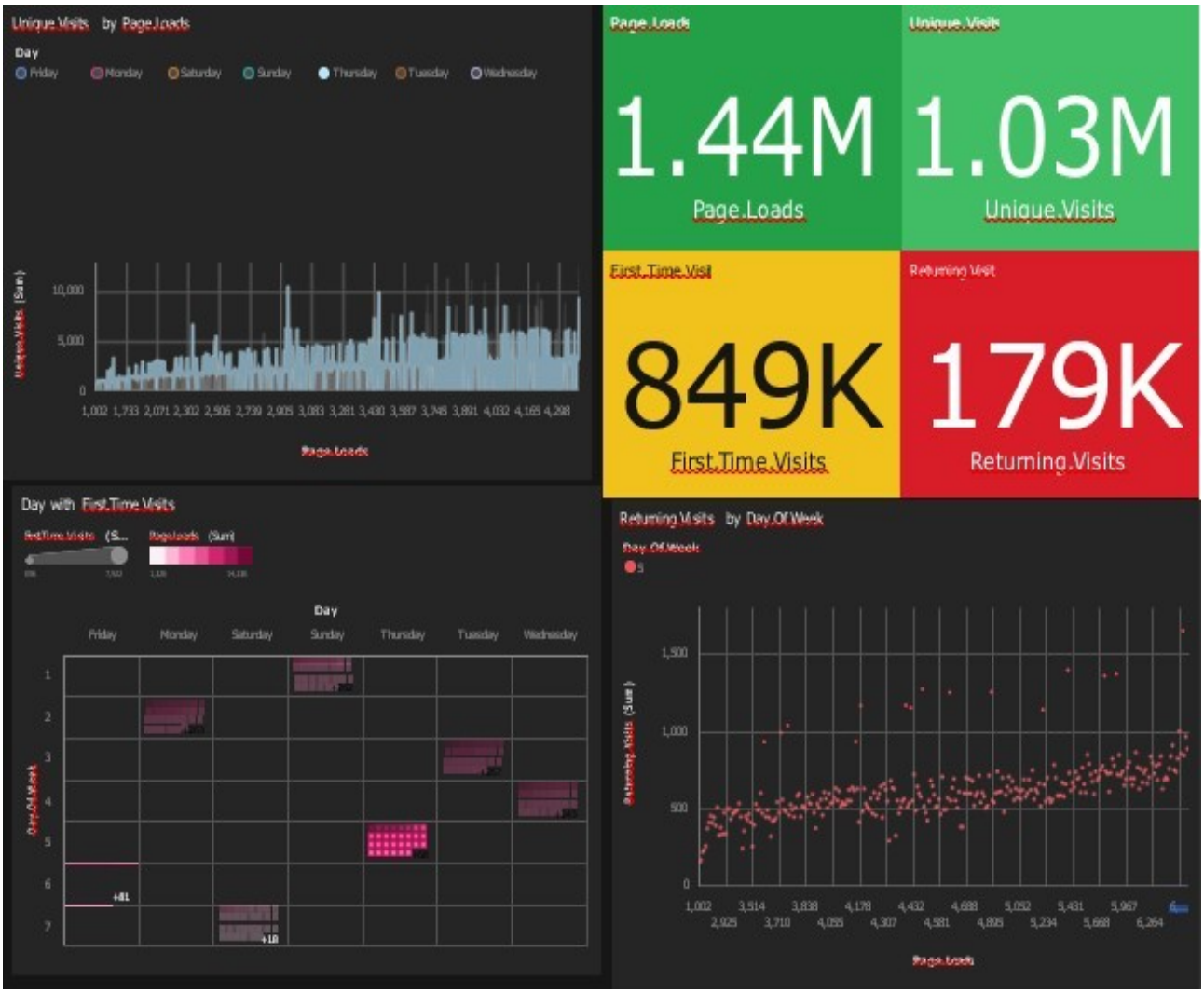
6.4.5 Wednesday visitors:

Analyzing Tuesday visitors in website traffic analysis involves studying user behavior and engagement patterns on Tuesdays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Tuesday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



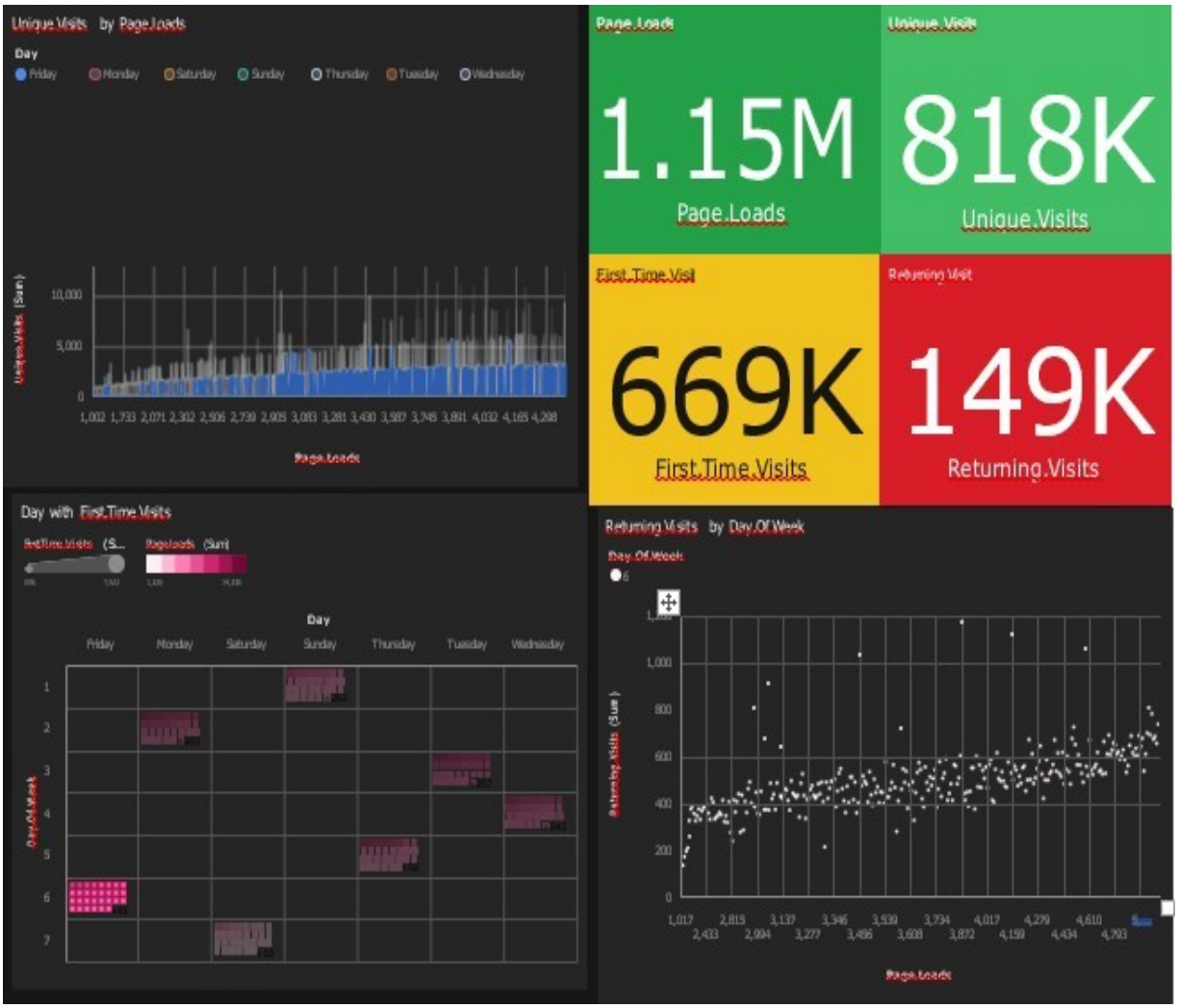
6.4.6Thursday:

Analyzing Tuesday visitors in website traffic analysis involves studying user behavior and engagement patterns on Tuesdays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Tuesday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



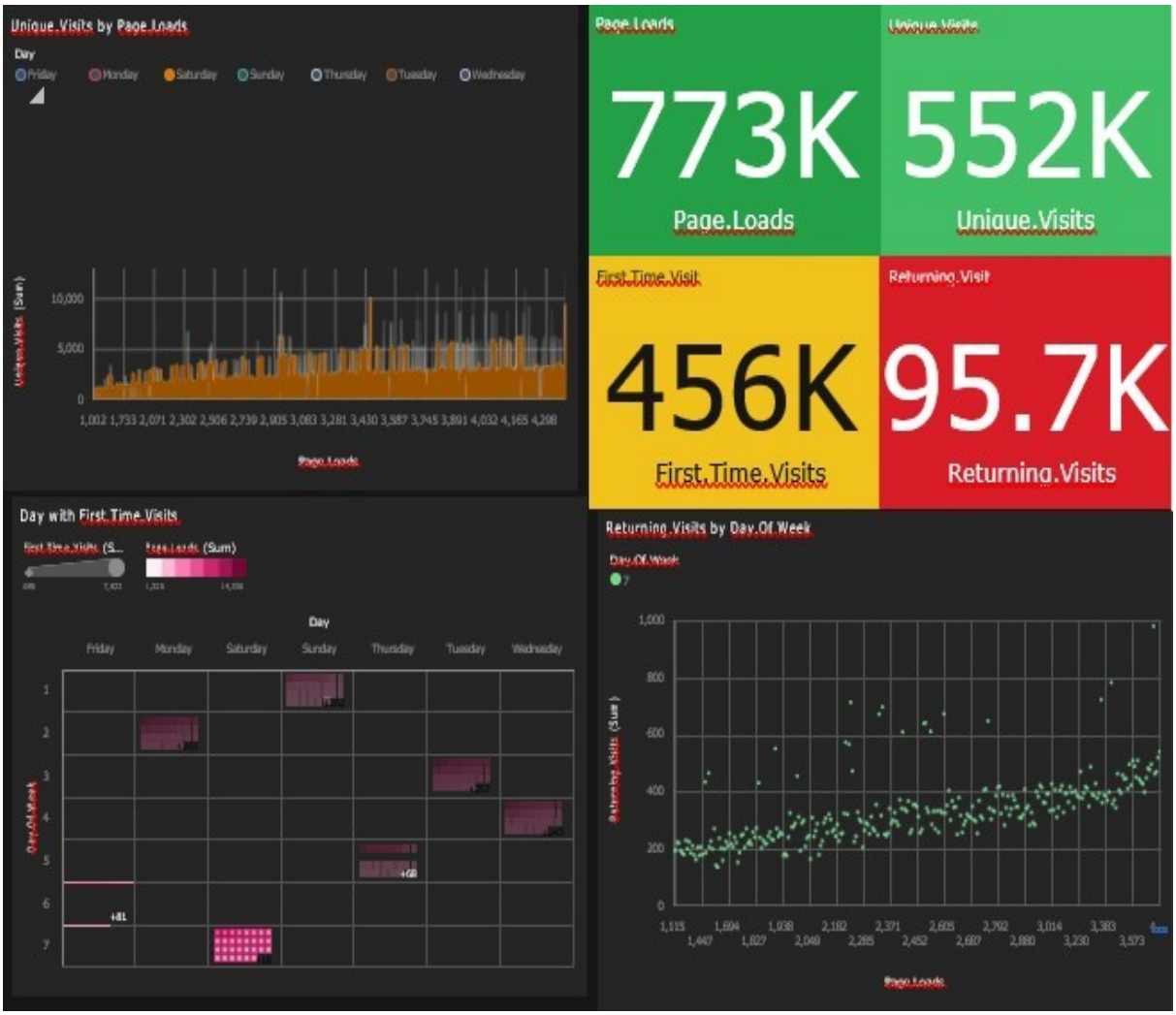
6.4.7 Friday:

Analyzing Tuesday visitors in website traffic analysis involves studying user behavior and engagement patterns on Tuesdays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Tuesday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



6.4.8 Saturday:

Analyzing Tuesday visitors in website traffic analysis involves studying user behavior and engagement patterns on Tuesdays, allowing you to uncover insights such as traffic trends, demographic preferences, content popularity, conversion rates, bounce rates, referral sources, peak activity times, device usage, and comparative performance with other days of the week. This information is invaluable for tailoring website content, marketing strategies, and user experiences to meet the specific needs and habits of Tuesday visitors, ultimately enhancing website performance and user satisfaction on this particular day.



6.5 Insights for above “IBM Cognos Analytics”

unique visits:

Unique.Visits is unusually low when Day is Saturday.

Based on the current forecasting, Unique.Visits may reach almost 481 thousand by Day Monday+1.

It is projected that by Monday+1, 4205 will exceed 3973 in Unique.Visits by almost 1500.

Page.Loads 4376 has the highest Total Returning.Visits but is ranked #5 in Total Unique.Visits.

Page.Loads 4638 has the highest Total Unique.Visits but is ranked #3 in Total Returning.Visits.

Over all values of Page.Loads and Day, the sum of Unique.Visits is almost 6.4

million. The summed values of Unique.Visits range from 667 to nearly 13

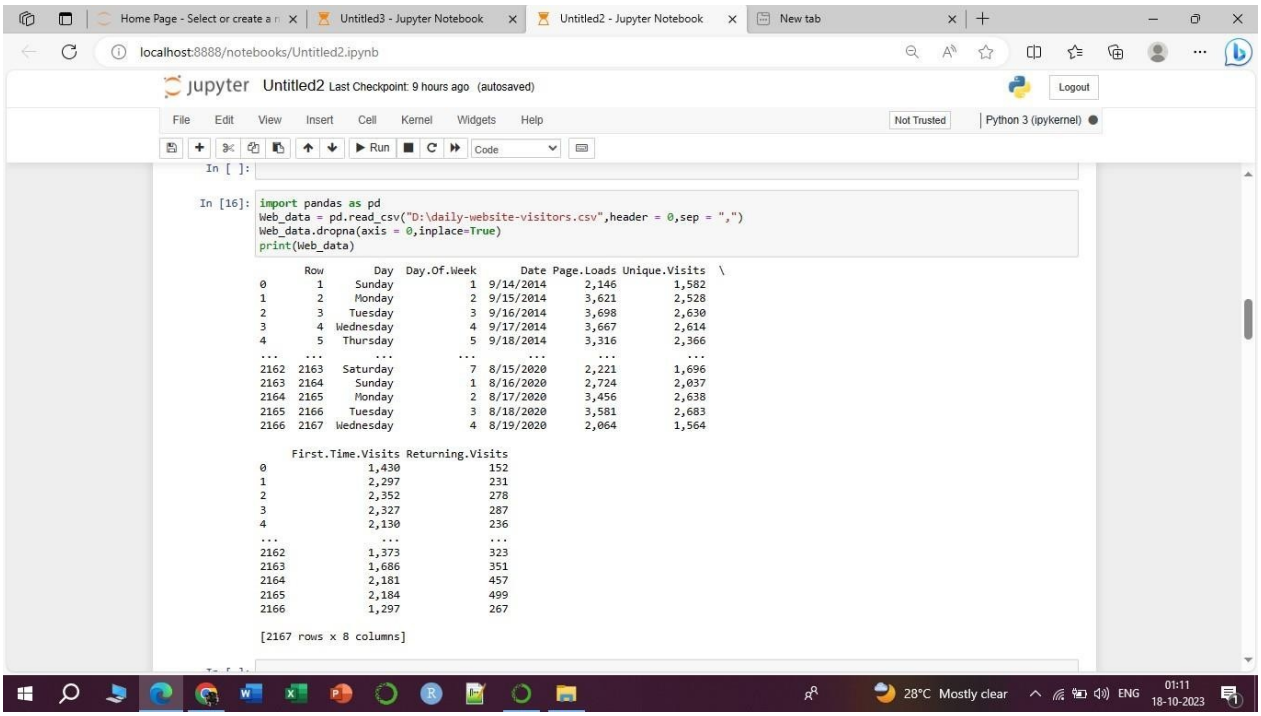
thousand.

For Unique.Visits, the most significant values of Day are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective Unique.Visits values add up to over 5.1 million, or 80 % of the total.

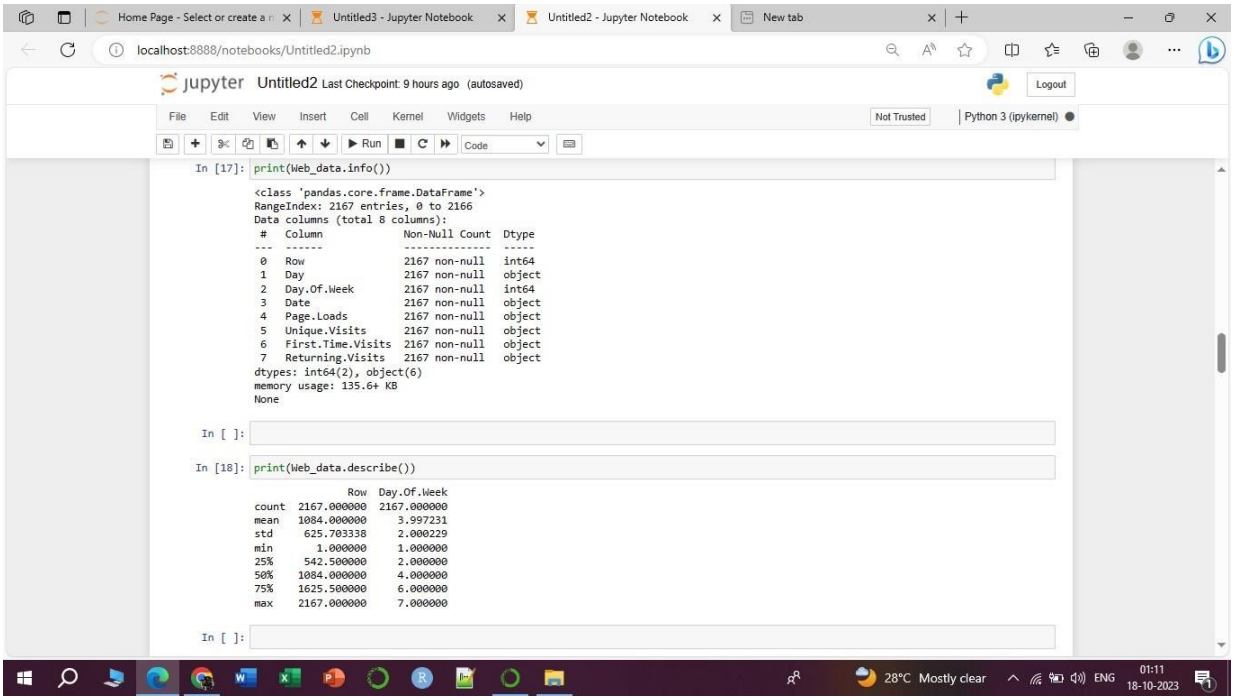
For Unique.Visits, the most significant value of Page.Loads is 4638, whose respective Unique.Visits values add up to over thirteen thousand, or 0.2 % of the total.

7.0 ‘Python Integration’ for Website Traffic Analysis

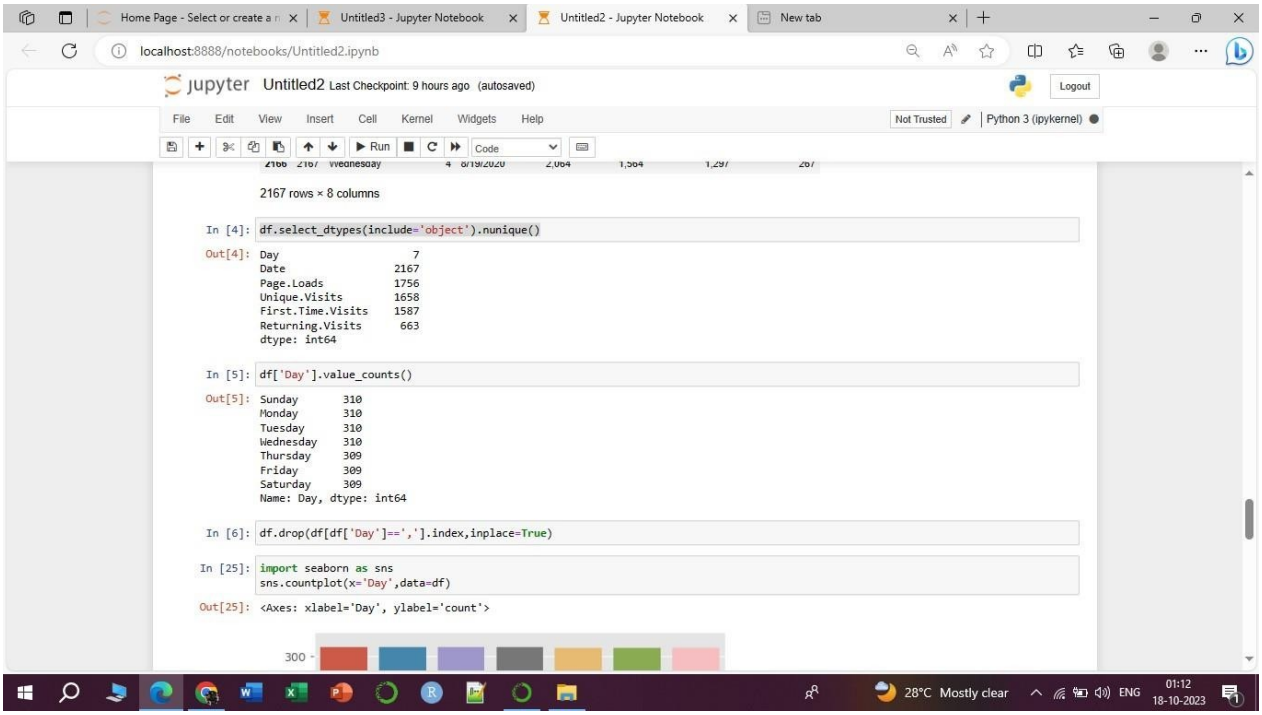
Extraction the Data file from the directory to the python text editor to excecute the data set.



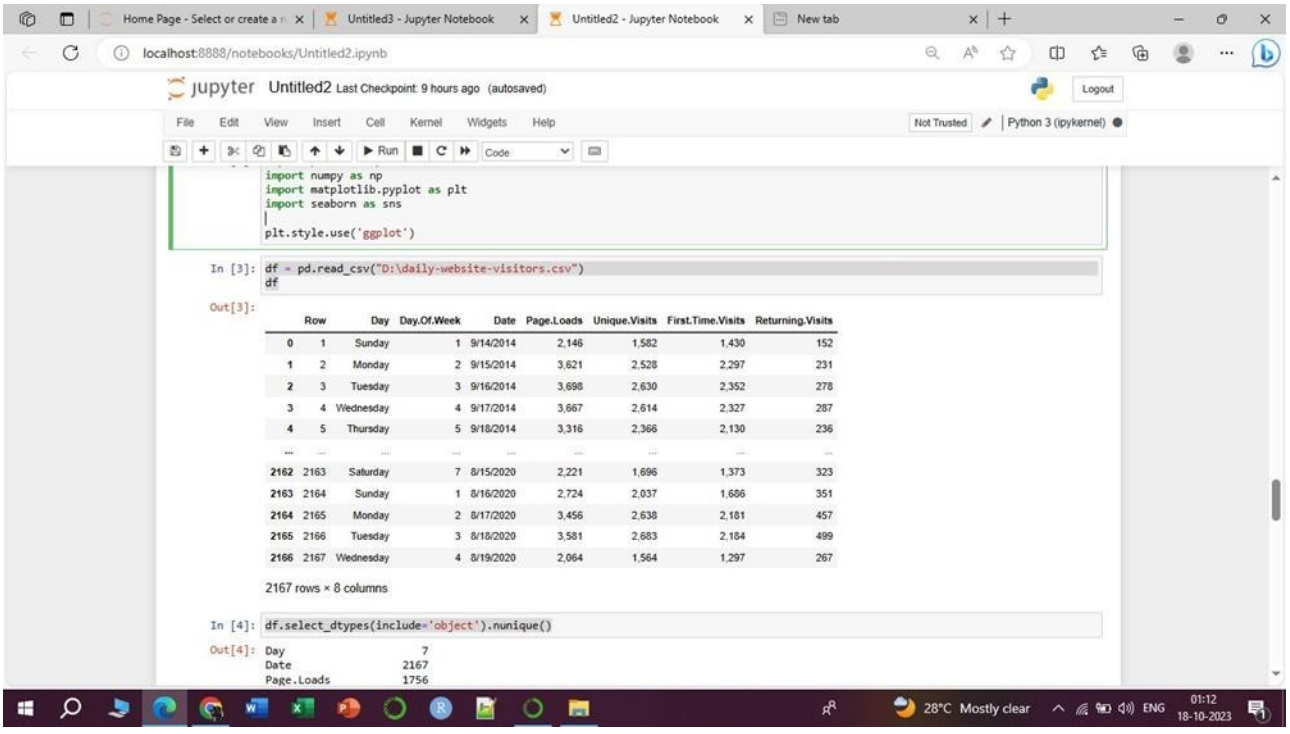
Getting of data set information using info function.



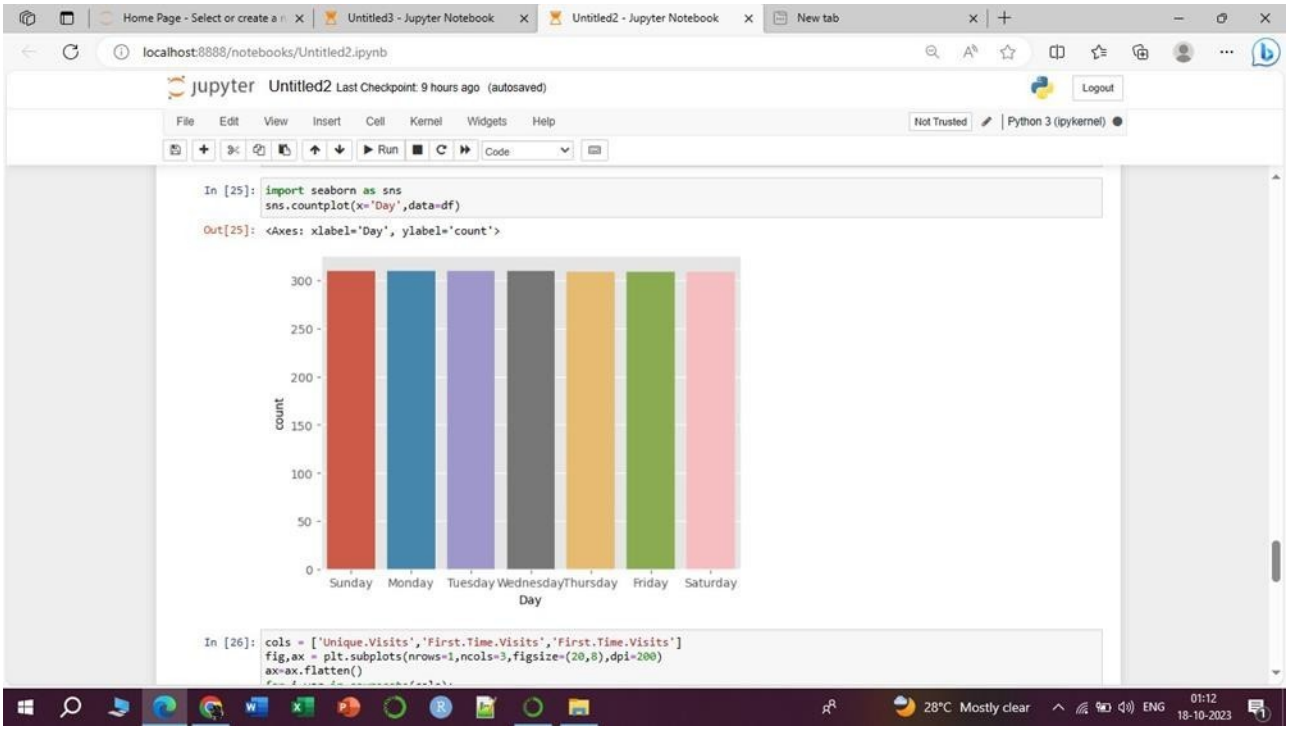
Value Counts of each Insights of the data set content and Object Integration of the data set



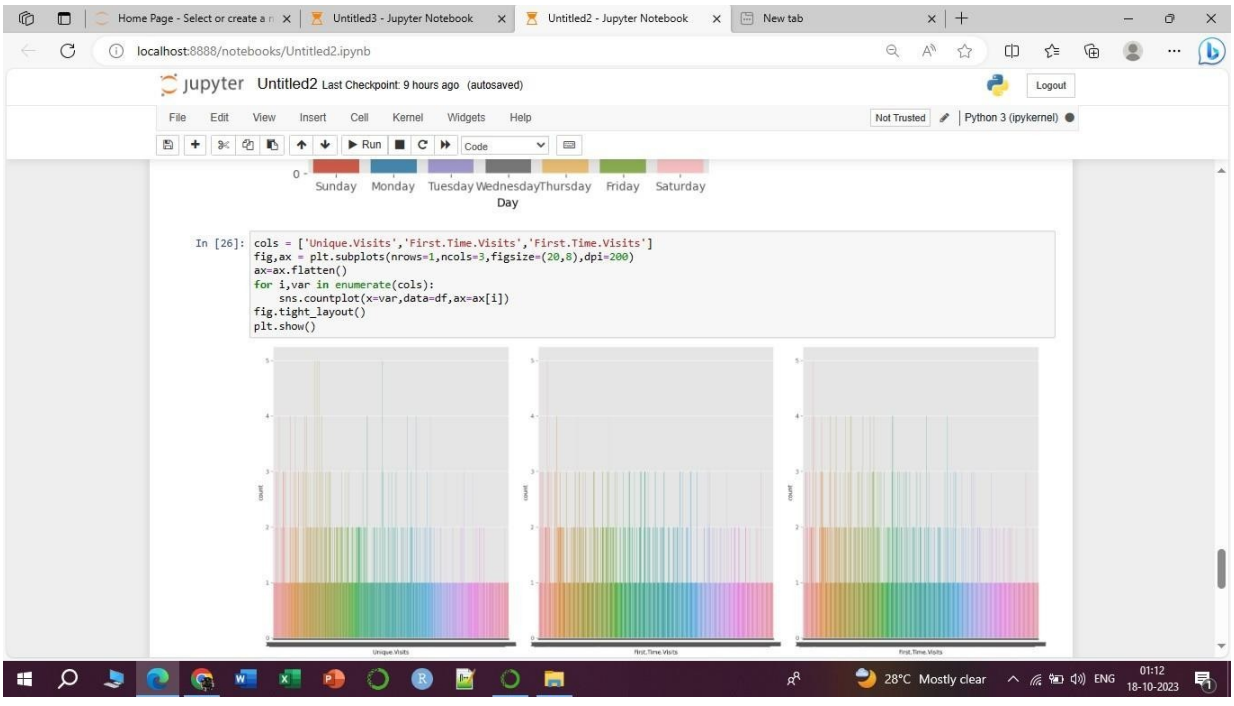
Again pointing out the data set with the help of Pandas Library.



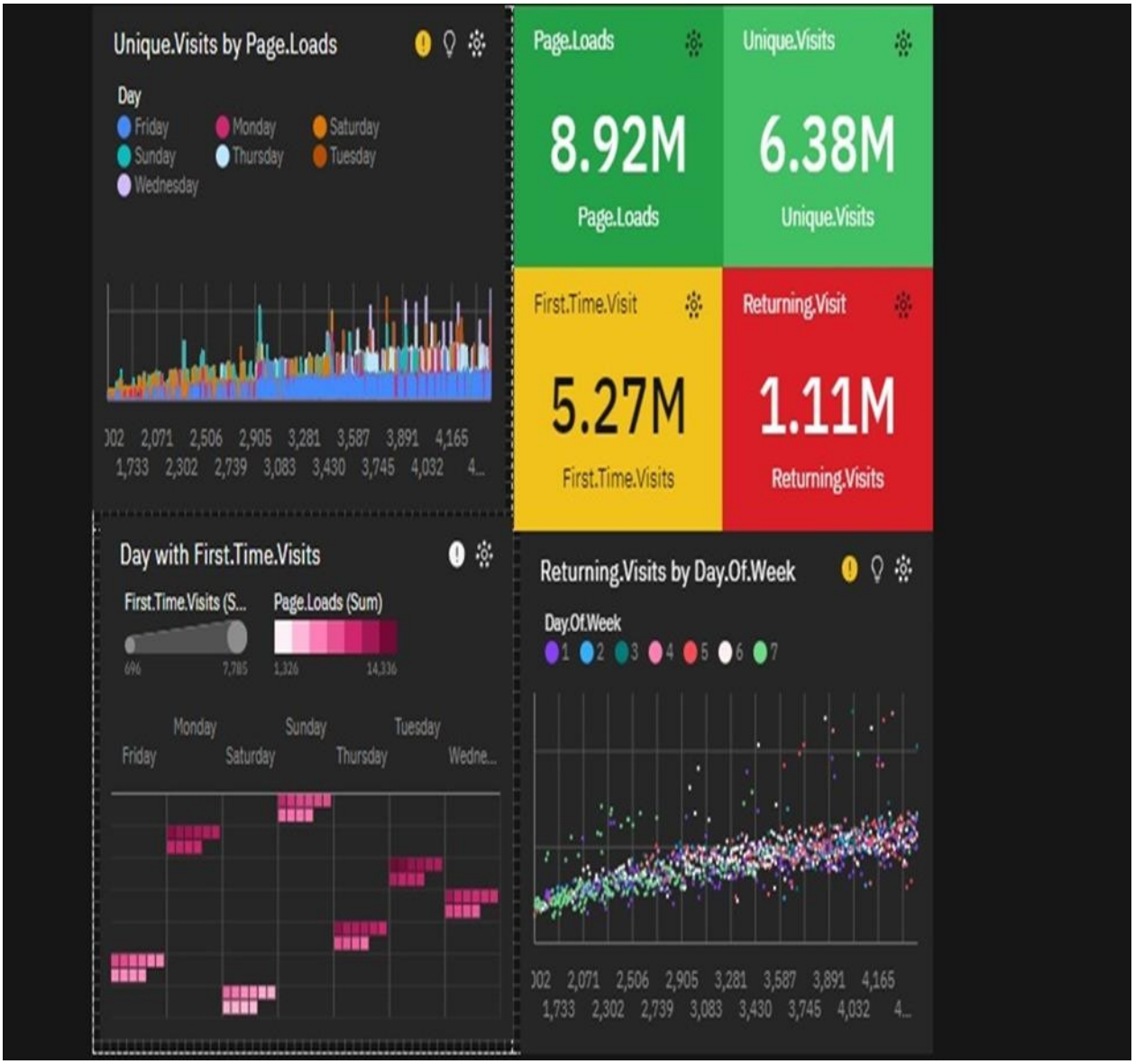
Listing the Number of Days that had been observed by the data set using Seaborn Library.



Plotting of Three important Stuffs like Unique visitor,First Time Visitor,Returning Visitor



8.0 Data Exploration:



- Page.Loads has a strong weekly trend. The largest values typically occur on Tuesday, whereas the smalest values on Saturday.

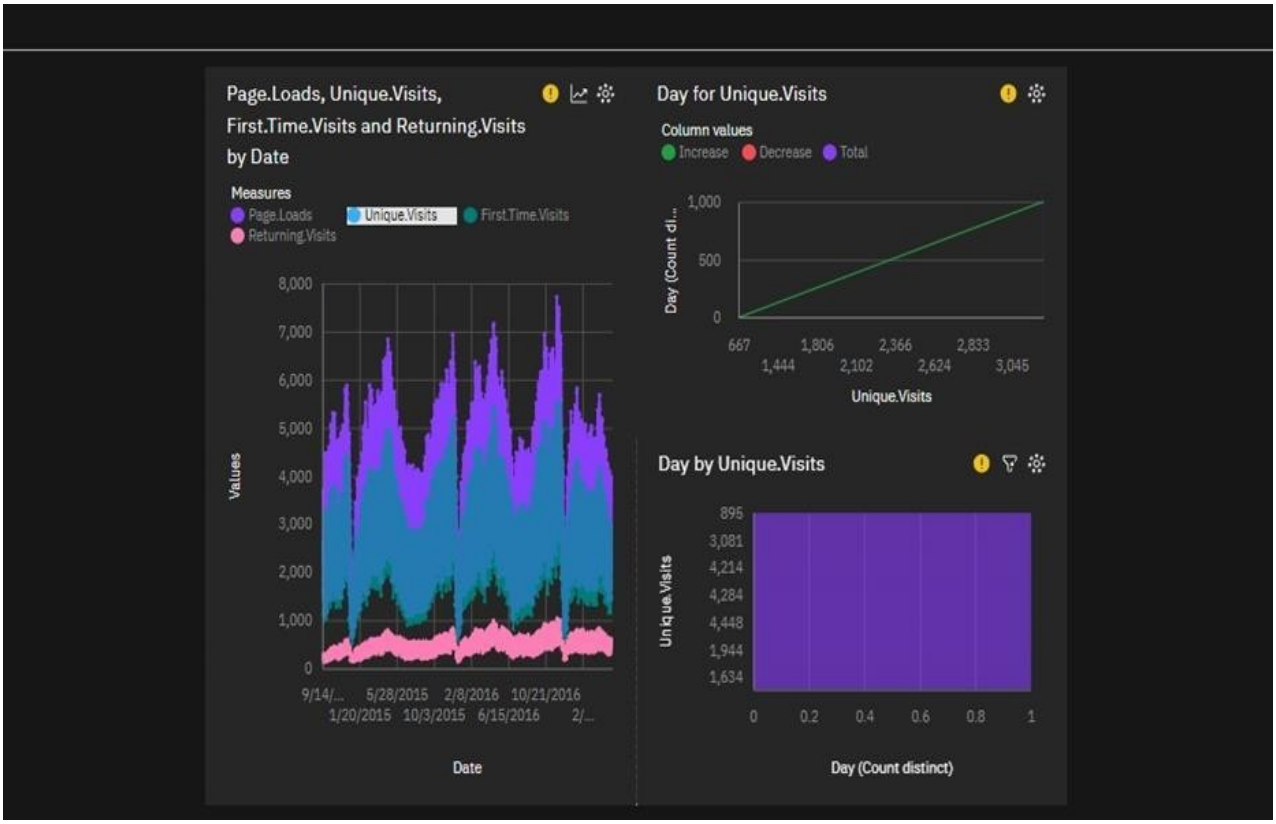
- Based on the current forecasting, Page.Loads may reach nearly four thousand by Date 2021-10-27

- Over all dates, the average of First.Time.Visits is almost 2500.

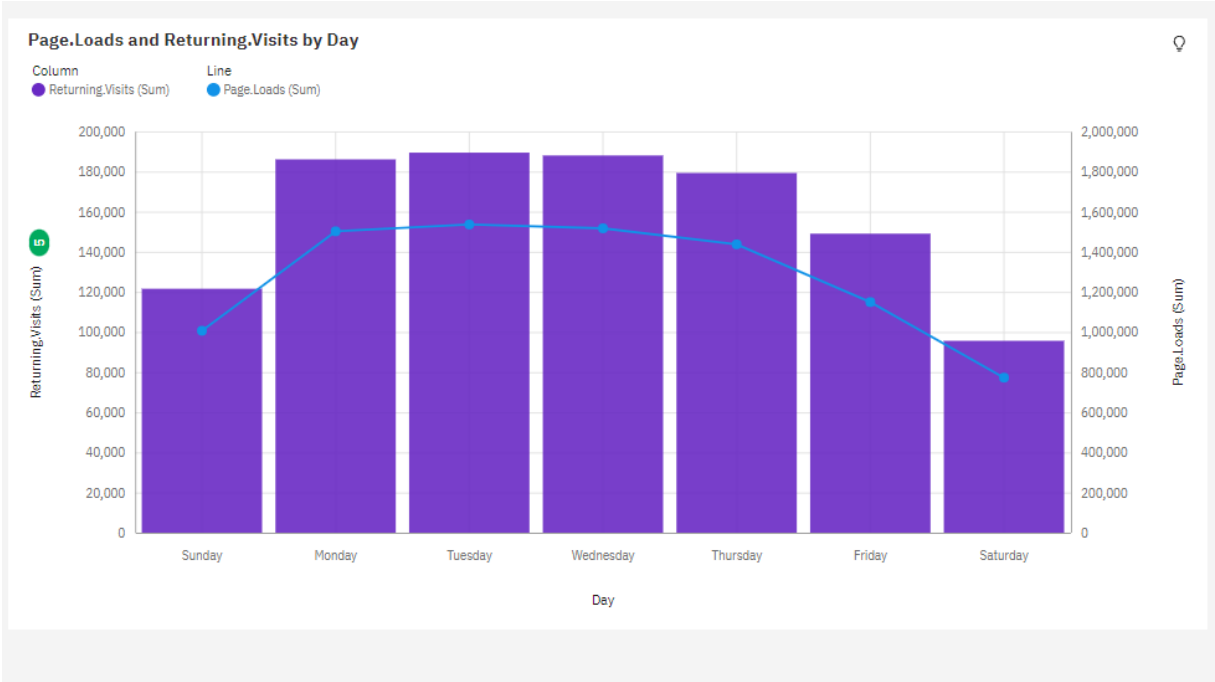
- Across all dates, the average of Page.Loads is over four thousand.
- Over all dates, the average of Returning.Visits is 511.8.
- Over all dates, the average of Unique.Visits is nearly three thousand.

- The total number of results for First.Time.Visits, across all dates, is over two thousand.

- The total number of results for Page.Loads, across all dates, is over two thousand.




Explore data relationships
daily-website-visitors.csv
Reset to original
Day
Edit diagram
Day.Of.Week
Unique.Visits
Page.Loads
Day
First.Time.Visits
Returning.Visits
Select single or multiple nodes to see visualizations.
Relationship diagram


Page.Loads and Returning.Visits by Day
Column
Returning.Visits (Sum)
Line
Page.Loads (Sum)
200,000
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
Returning.Visits (Sum)
2,000,000
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
Page.Loads (Sum)
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day

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.

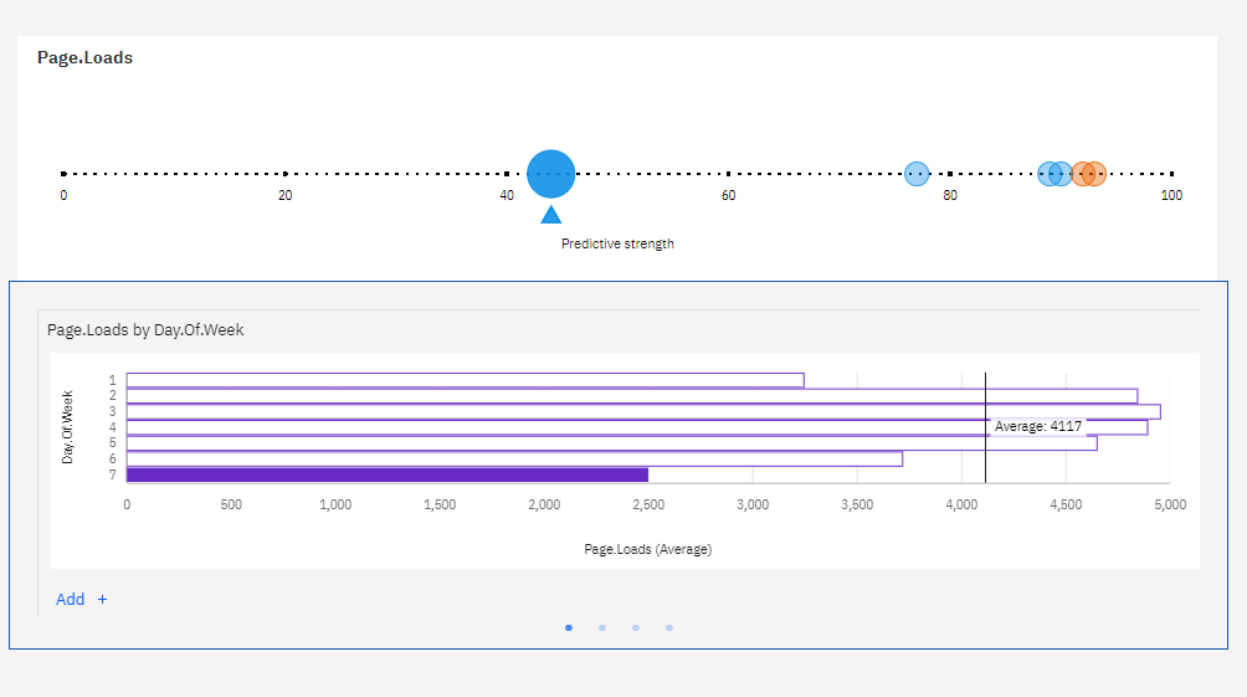
For **Returning.Visits**, the most significant values of **Day** are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective **Returning.Visits** values add up to almost 892 thousand, or 80.4 % of the total.

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 1.5million, 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.

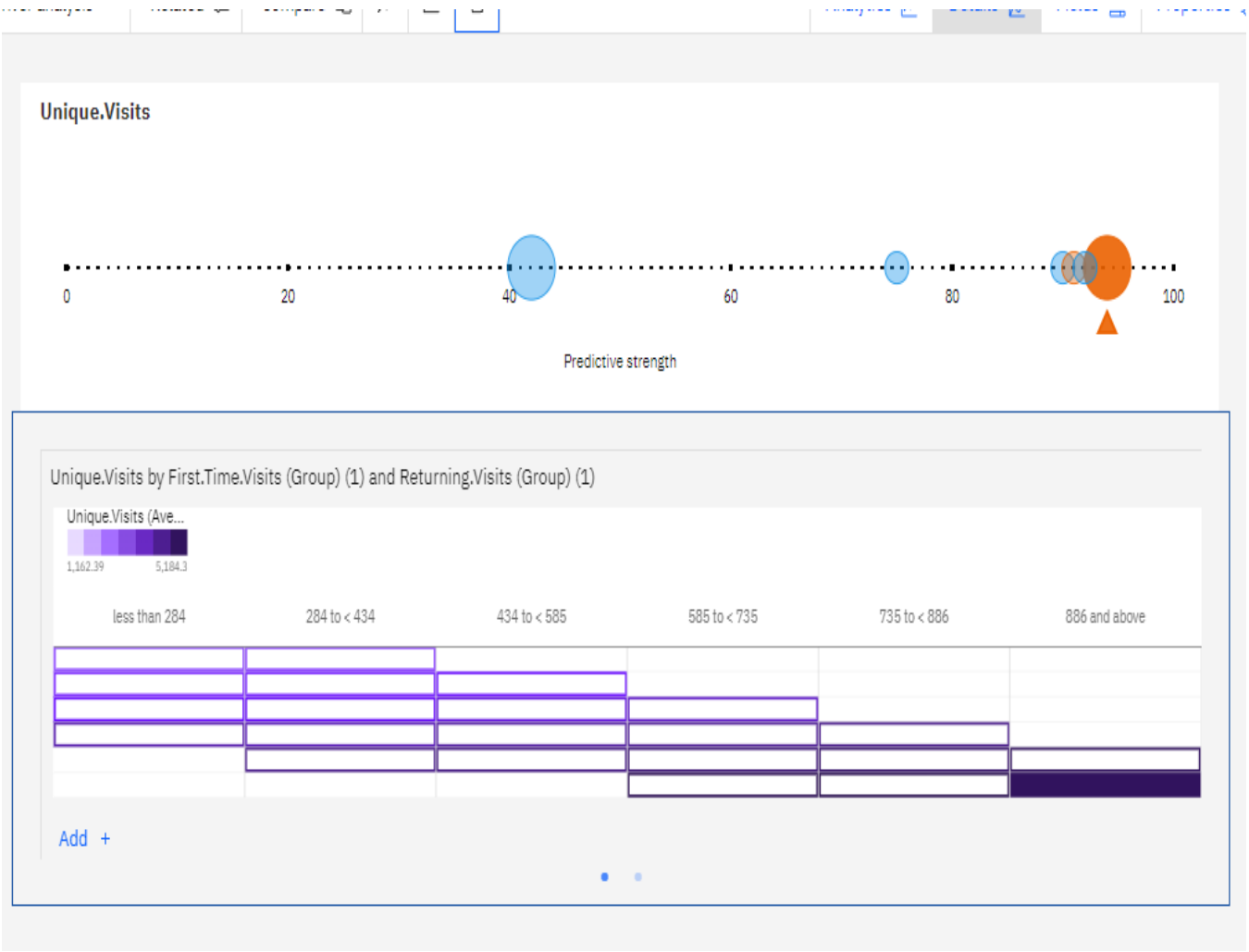
The average values of Page.Loads range from over 2500, occurring when

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)

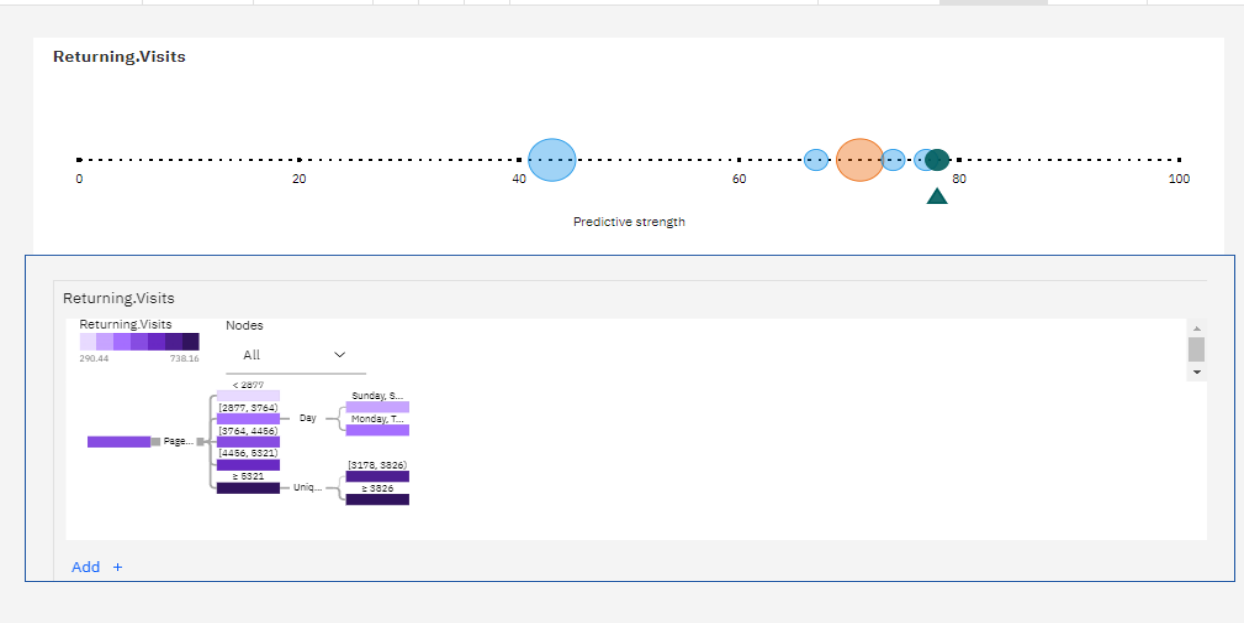
(3) and Returning.Visits (Group) (2) is 3934 and above and 886 and above.

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).

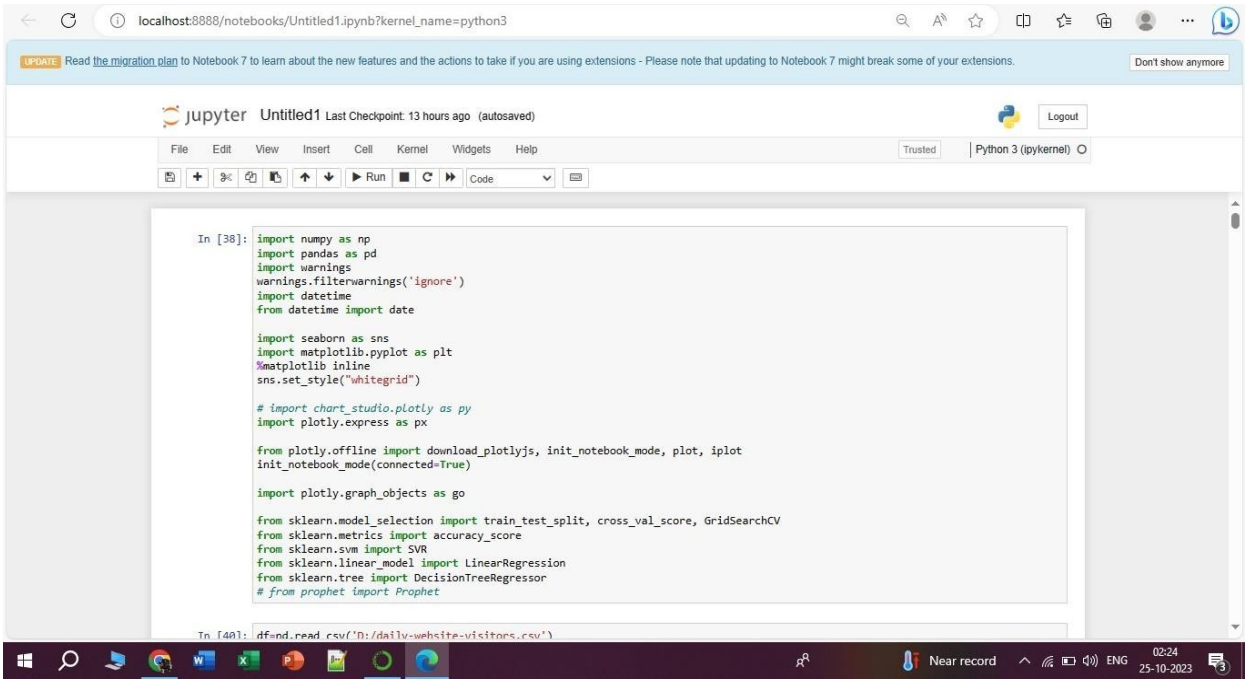
There is no significant impact of Returning.Visits (Group) (2) on the relationship between First.Time.Visits (Group) (3) and Unique.Visits.

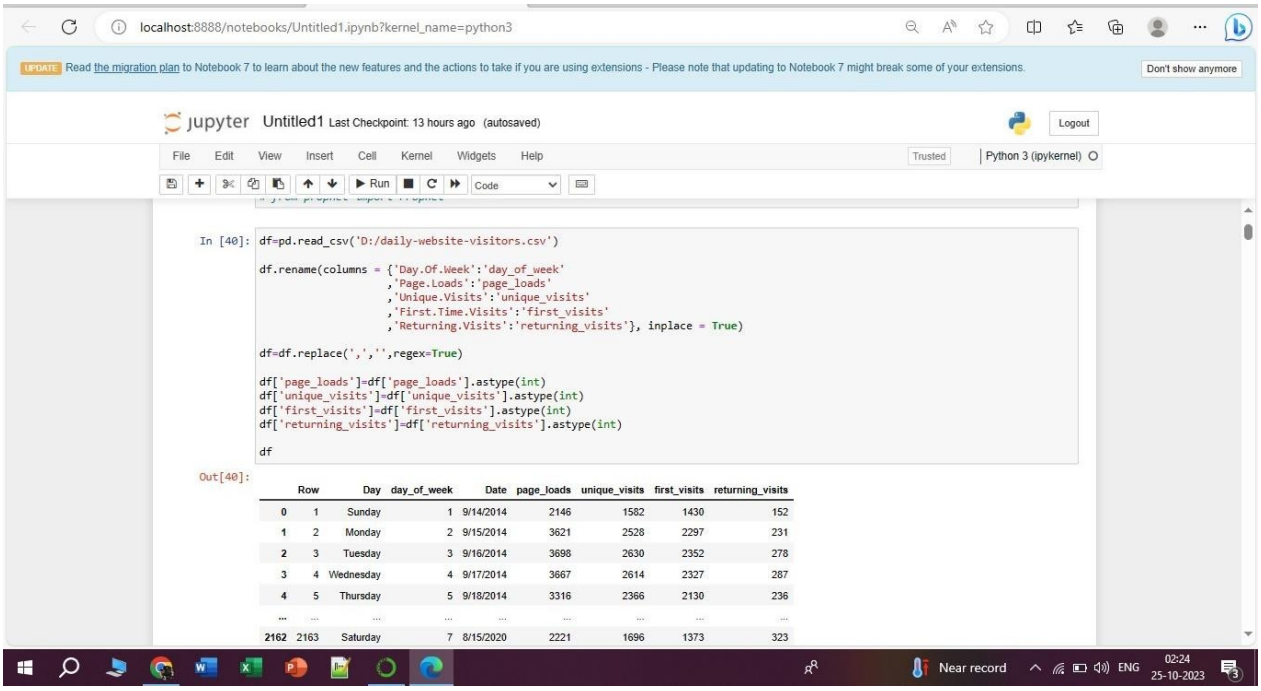


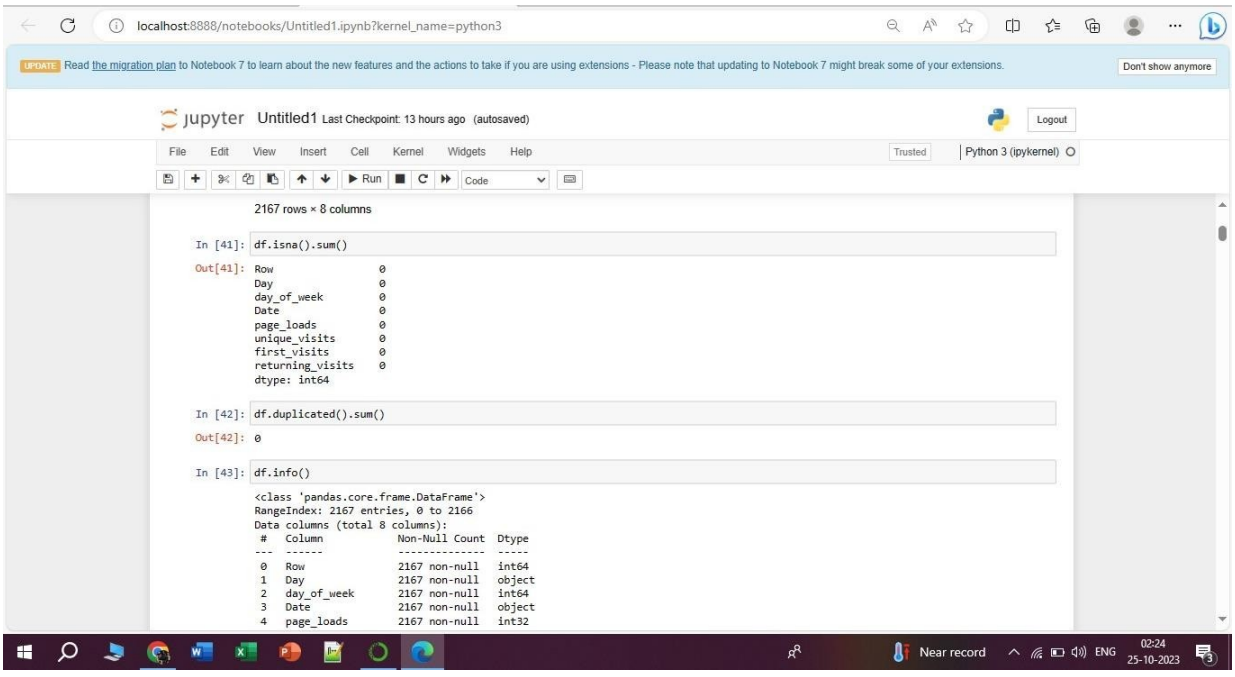
Page.Loads, **Unique.Visits**, and **Day** predict **Returning.Visits** with a strength of 78.1%.

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

9.6 Data Exploration using 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


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

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

localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3

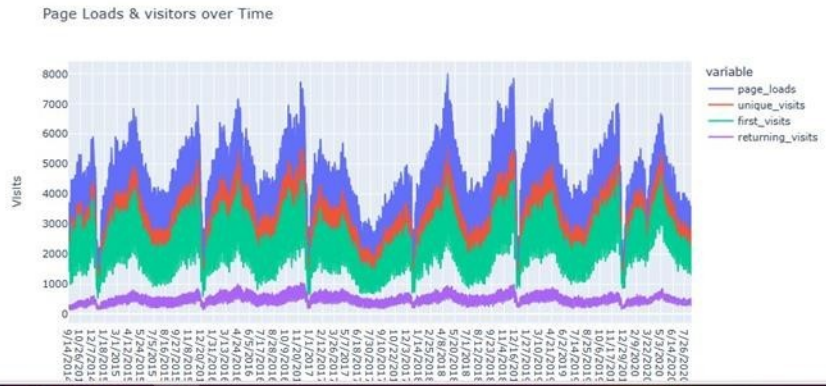
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved) Logout

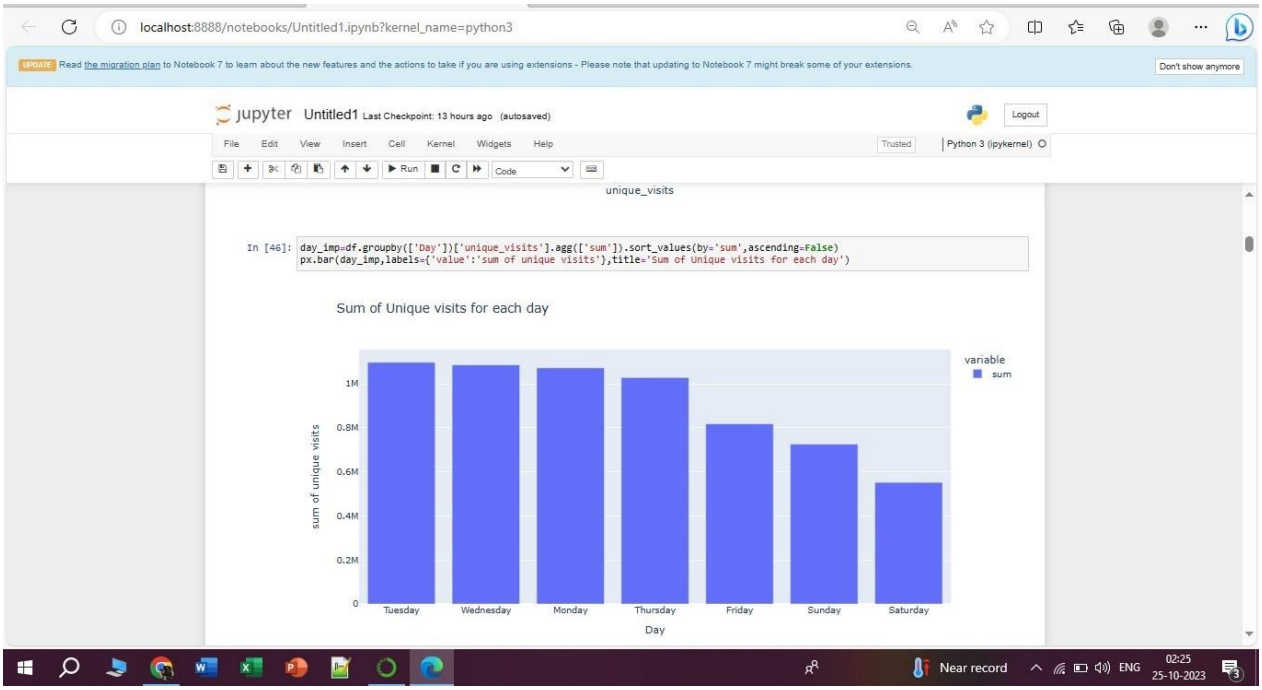
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

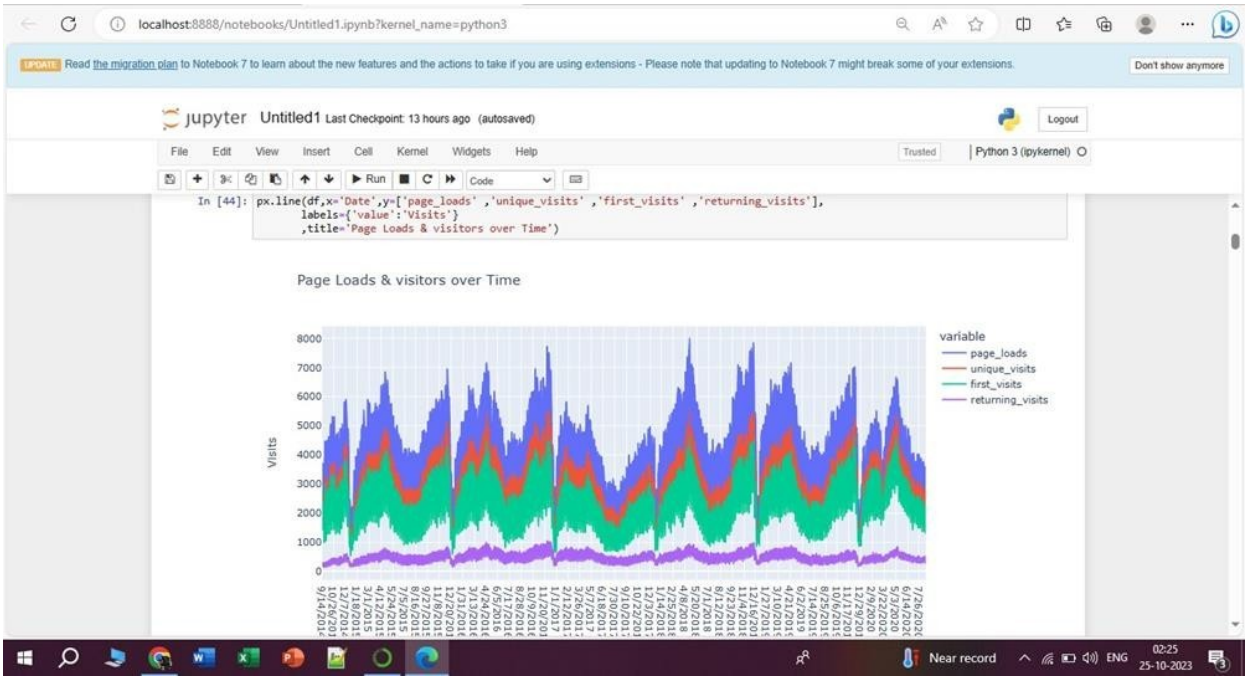
Run Code

```
In [44]: px.line(df,x='Date',y=['page_loads' ,'unique_visits' ,'first_visits' ,'returning_visits'],
                 labels={'value':'Visits'}
                 ,title='Page Loads & visitors over Time')
```



Near record ENG 02:25 25-10-2023


localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
unique_visits
In [46]:
day_imp=df.groupby(['Day'])['unique_visits'].agg(['sum']).sort_values(by='sum',ascending=False)
px.bar(day_imp,labels={'value':'sum of unique visits'},title='Sum of Unique visits for each day')
Sum of Unique visits for each day
variable
sum
sum of unique visits
1M
0.8M
0.6M
0.4M
0.2M
0
Tuesday
Wednesday
Monday
Thursday
Friday
Sunday
Saturday
Day
Near record
ENG
02:25
25-10-2023


localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [44]:
px.line(df,x='Date',y=['page_loads' ,'unique_visits' ,'first_visits' ,'returning_visits'],
labels={'value':'Visits'}
,title='Page Loads & visitors over Time')
Page Loads & visitors over Time
variable
page_loads
unique_visits
first_visits
returning_visits
Visits
8000
7000
6000
5000
4000
3000
2000
1000
0
Near record
ENG
02:25
25-10-2023

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

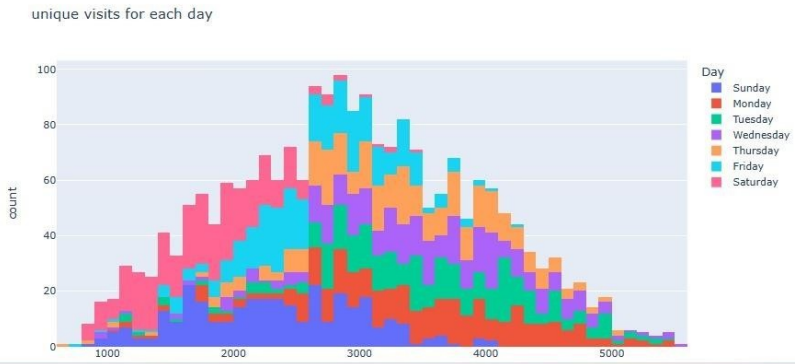
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

Date

In [45]:
```
px.histogram(df,x='unique_visits',color='Day',title='unique visits for each day')
```



Near record ENG 02:25 25-10-2023

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

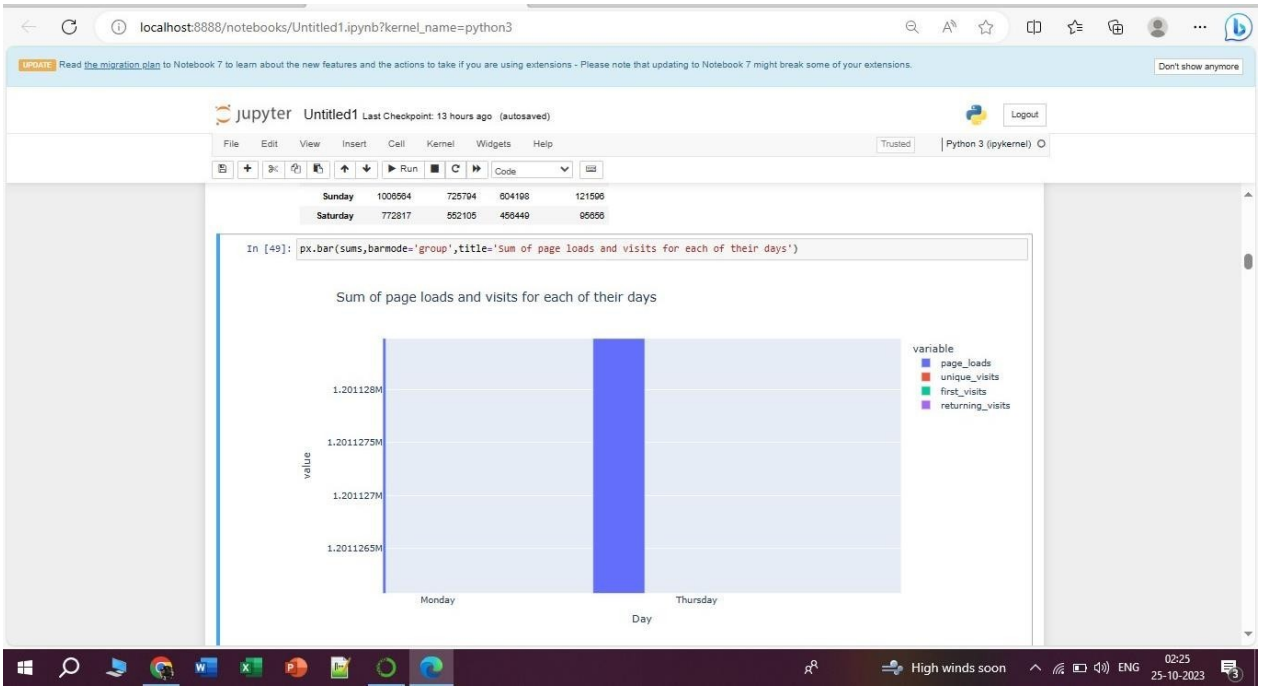
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

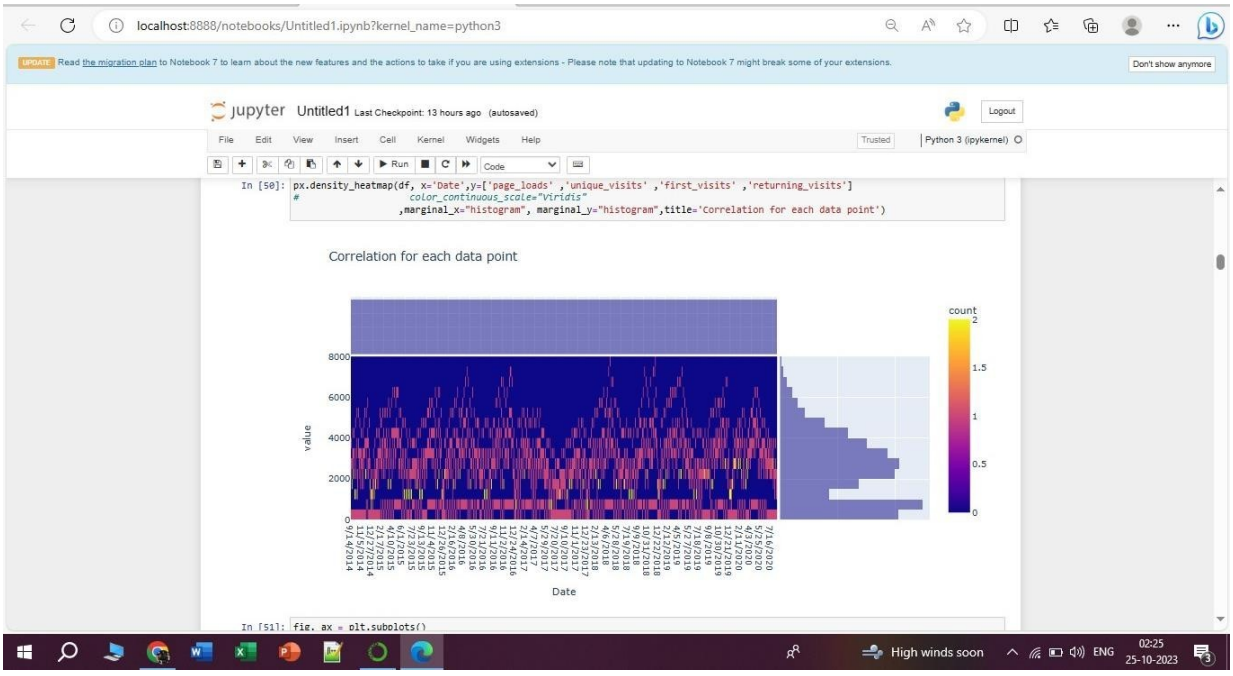
Run Code

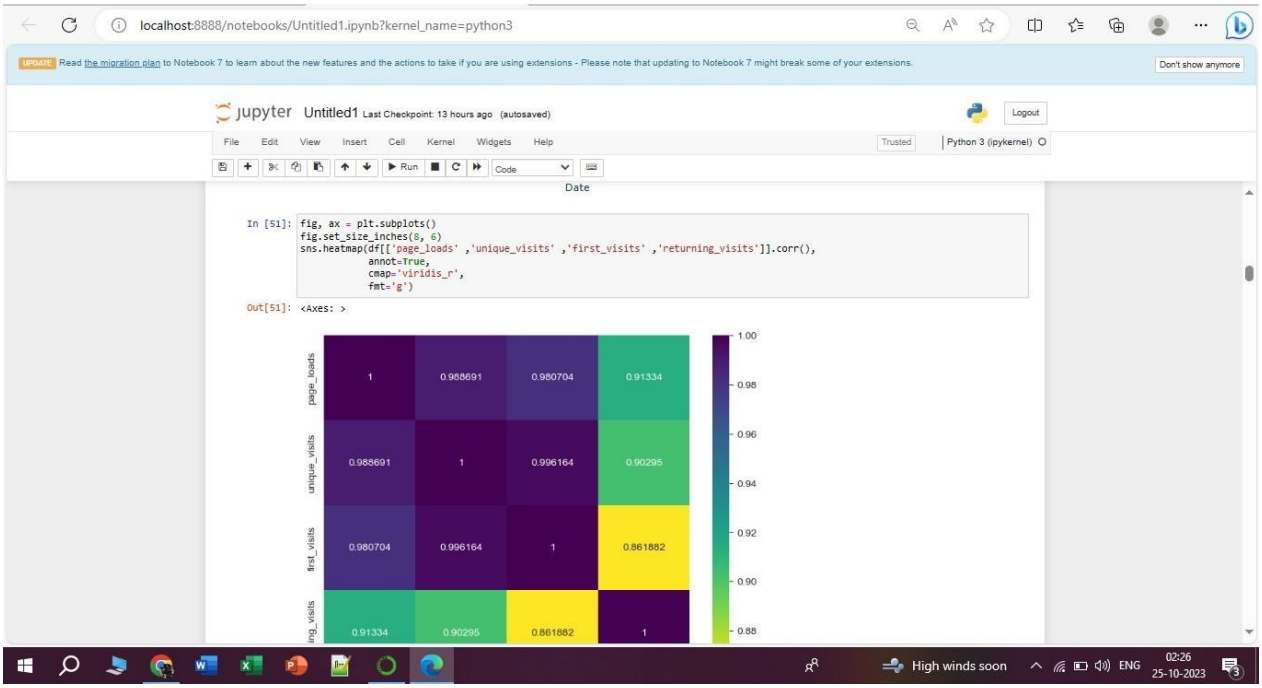
In [47]:
```
px.histogram(df,x='Date',y='unique_visits',color='Day',title='Sum of unique visits for each day over Time')
```

In [48]:
```
sums=df.groupby(['Day'])[['page_loads' ,'unique_visits' ,'first_visits' ,'returning_visits']].sum().sort_values(
    by='unique_visits',ascending=False)
```

High winds soon ENG 02:25 25-10-2023


localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Sunday 1006564 725794 604198 121596
Saturday 772817 552105 456449 95656
In [49]: px.bar(sums,barmode='group',title='Sum of page loads and visits for each of their days')
Sum of page loads and visits for each of their days
variable
page_loads
unique_visits
first_visits
returning_visits
1.201128M
1.2011275M
1.201127M
1.2011265M
value
Monday
Thursday
Day
High winds soon
ENG
02:25
25-10-2023


localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [50]: px.density_heatmap(df, x='Date',y=['page_loads' ,'unique_visits' ,'first_visits' ,'returning_visits']
color_continuous_scale="Viridis"
,marginal_x="histogram", marginal_y="histogram",title='Correlation for each data point')
Correlation for each data point
count
8000
6000
4000
2000
0
value
Date
In [51]: fig, ax = plt.subplots()
High winds soon
ENG
02:25
25-10-2023



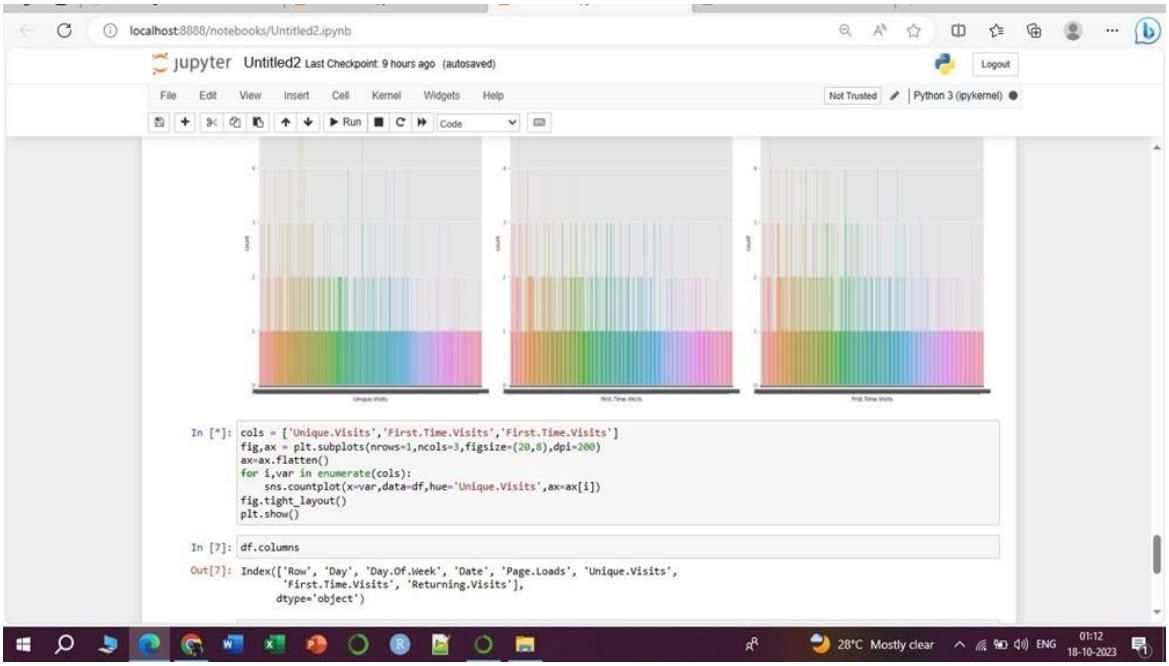
8.1 Data Expolaration using Jypyter Notebook:

Analyzing website traffic using Jupyter Notebook involves leveraging Python and relevant libraries like Pandas, Matplotlib, and Seaborn. The process begins with data collection, which can be sourced from web server logs or analytics tools like Google Analytics. Once you've set up your Jupyter Notebook and imported the necessary libraries, you load the data into a Pandas DataFrame. From there, you explore the data, perform data cleaning, and conduct in-depth analysis. This analysis may encompass examining traffic trends over time, counting page views, assessing geographic user distribution, identifying traffic sources, and studying user behavior. Data visualization plays a pivotal role in conveying insights effectively, and you can create various plots to present your findings. Lastly, you can document your insights and automate the process for ongoing analysis and reporting.

In conclusion, website traffic analysis in Jupyter Notebook is a comprehensive data-driven process that involves data collection, cleaning, exploration, analysis, and visualization. It empowers you to gain valuable insights into user behavior, website performance, and traffic

sources, and allows for effective reporting and automation for continuous monitoring and optimization of your website's performance.

Example:



9.0 Predictive Analytics and Machine Learning in Website Traffic Analysis:

9.1 Predictive Analytics:

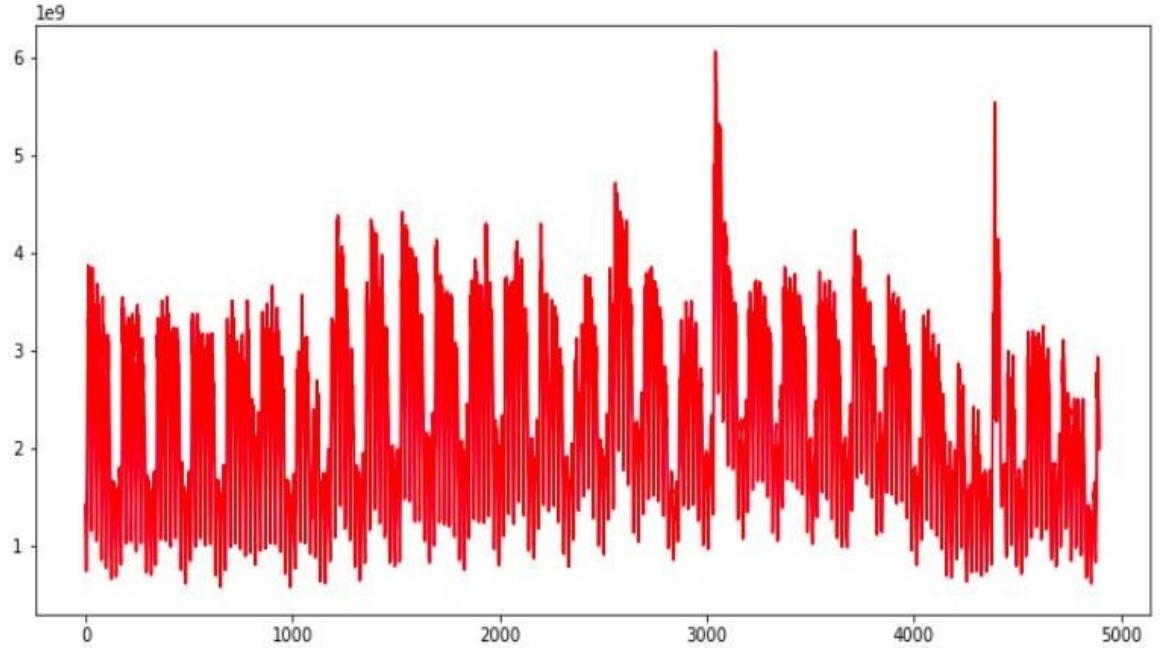
Predictive analytics is the practice of using historical data and statistical algorithms to predict future events or outcomes. It involves identifying patterns, relationships, and trends in data and using this information to forecast what might happen in the future. Predictive analytics often includes techniques like regression analysis, time series analysis, and classification. For example, in the context of website traffic analysis, predictive analytics could be used to forecast future traffic trends, user behavior, or conversion rates. It enables businesses to anticipate customer needs, optimize operations, and make proactive decisions.

Predictive Analysis Dashboard:

9.2 Machine Learning:

Machine learning is a subset of artificial intelligence that focuses on developing algorithms and models that allow computers to learn from data and improve their performance over time. Machine learning models can automatically identify patterns, classify data, and make predictions without being explicitly programmed. There are various types of machine learning, including supervised learning (where models are trained on labeled data), unsupervised learning (where models discover patterns without labeled data), and reinforcement learning (where models learn through trial and error). In website traffic analysis, machine learning can be used to build models that predict user behavior, segment users, detect anomalies, and provide personalized recommendations.

Example:



10.0 Conclusion:

In conclusion, the website traffic analysis project has provided a comprehensive understanding of the dynamics of our online platform and user interactions. By meticulously collecting, processing, and analyzing the data, we have gained valuable insights that can inform our strategies and enhance the user experience.

The analysis has revealed key trends in website traffic, user behavior, and performance over time. We have identified peak traffic periods, popular pages, and user demographics, which will guide content optimization and marketing efforts. Moreover, our predictive analytics and machine learning models have allowed us to forecast future trends and user behavior, empowering us to make proactive decisions to improve our website's performance.

This project's impact is significant, as it enables data-driven decision-making, resource allocation, and the implementation of targeted strategies for improving the website's efficiency and user engagement. As we move forward, the insights gained from this project will serve as a foundation for continuous optimization and enhancement of our online presence, ensuring that we meet and exceed our users' needs and expectations. The website traffic analysis project has proven to be a valuable tool in our ongoing efforts to provide an exceptional online experience.


THANK
YOU!