PHASE 4 DEVELOPMENT PART 2:WEBSITE TRAFFIC ANALYSIS

Team member
R.Shreevathshan:511921104065

Naan mudhalvan id: au511921104065

1.0 Introduction:

Websitetrafficanalysisistheprocessofcollecting,examining, andinterpretingdatarelatedtothevisitorsand interactionsona website. Itprovides invaluable insights into user behavior, preferences,andtrends, helpingorganizationsmakeinformed decisions, optimize their online presence, and enhance user experiences.

1.1Data 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.

1.3Data 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:

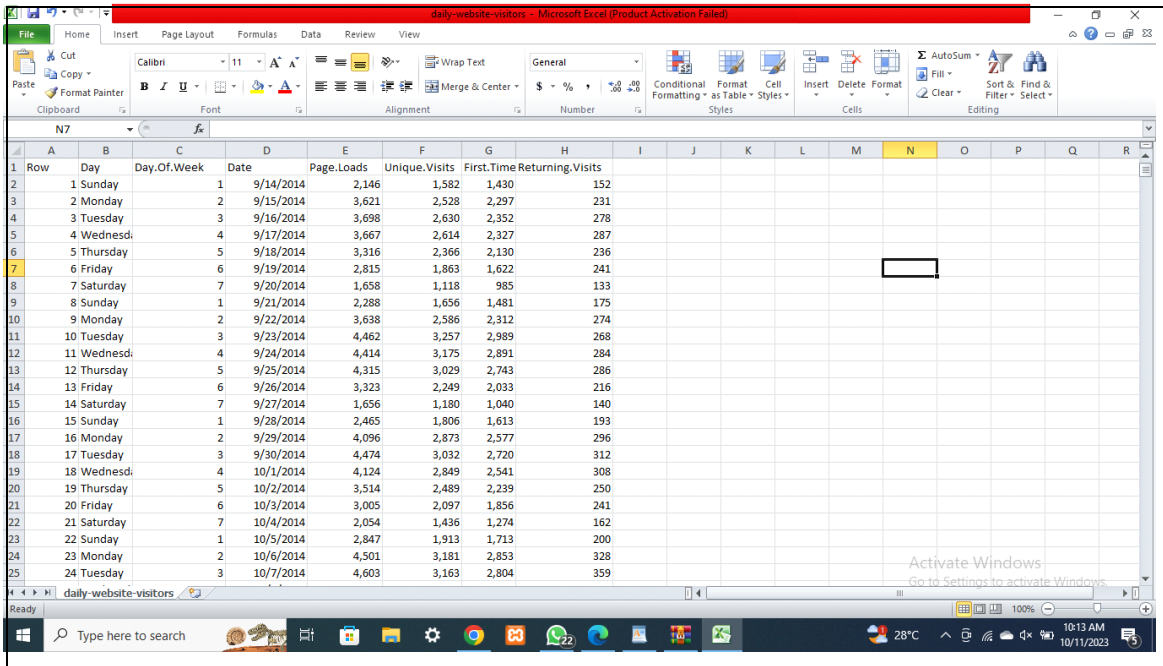
Cognos offers a user-friendly interface for creating interactive reports and dashboards. You can use drag-and-drop features to visualize website analytics data. For instance, you can create reports that show website traffic trends over time, popular pages, user engagement, and conversion rates. Dashboards provide a real-time overview of your website's performance metrics in a visually appealing format.

2.0 Abstract:

Thisprojectaimstoanalyzewebsitetraficdataforinsights intouserbehavior,popularpages,andtraficsources.It involvesdatacollection,visualizationusingIBMCognos,and Pythonforadvancedanalysis.Thegoalistooptimize user experiencesandenhancewebsiteperformance.

3.0 DATASET:

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



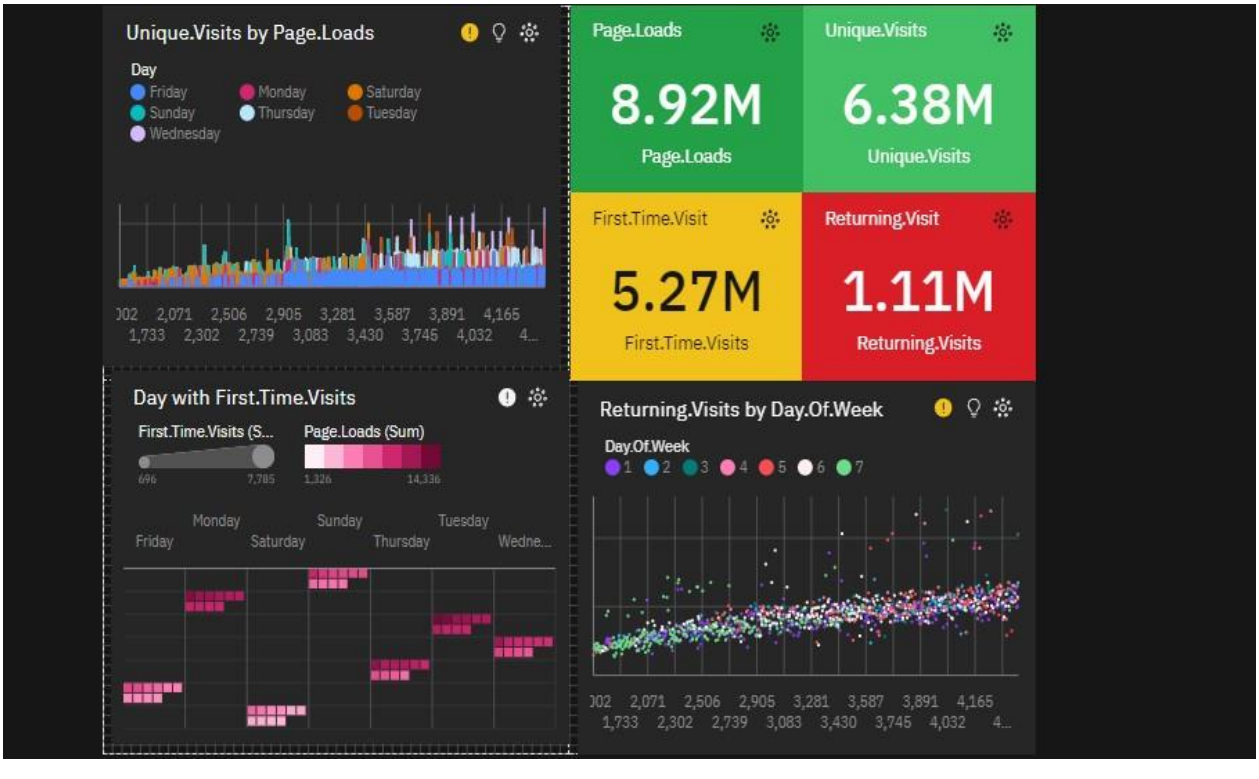
daily-website-visitors - Microsoft Excel (Product Activation Failed)

Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time	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	3,638	2,586	2,312	274
10	Tuesday	3	9/23/2014	4,462	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	2,465	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
24	Tuesday	3	10/7/2014	4,603	3,163	2,804	359

4.0 Data Exploration

Launch IBM Cognos and connect to data sources. Cognos supports a variety of data sources including databases, spreadsheets, and data warehouses.

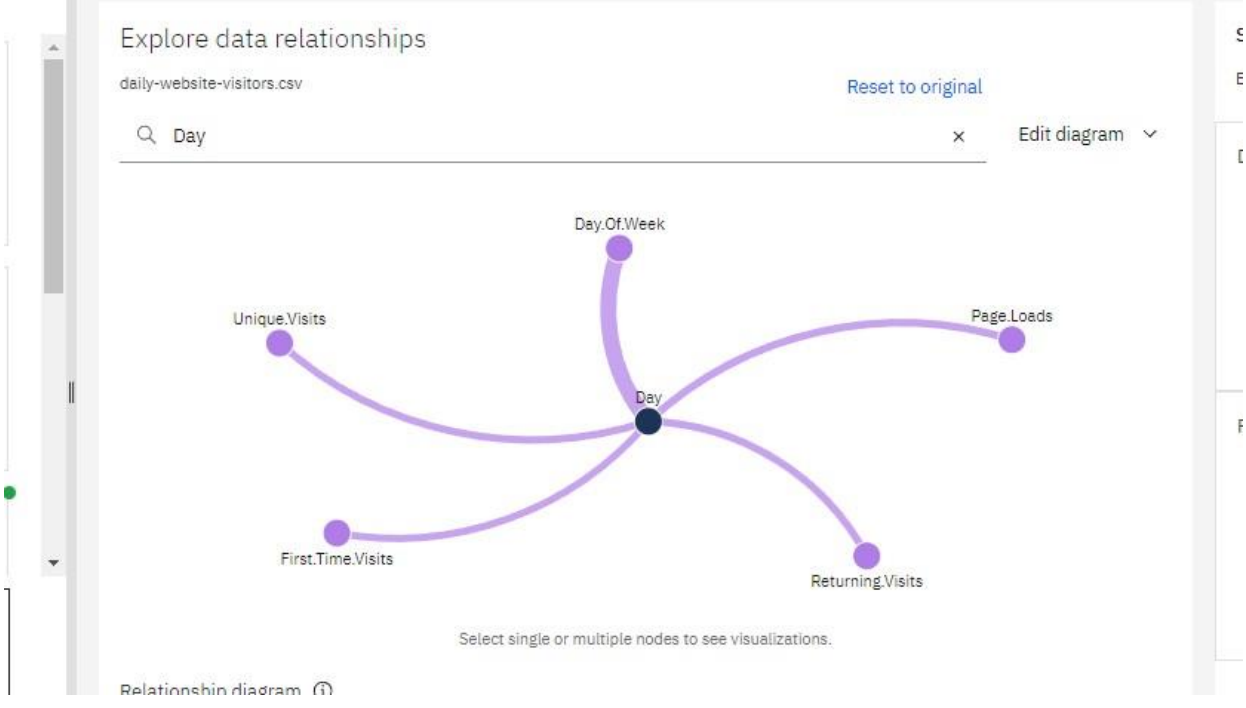
4.1 exploration 1



Analysis

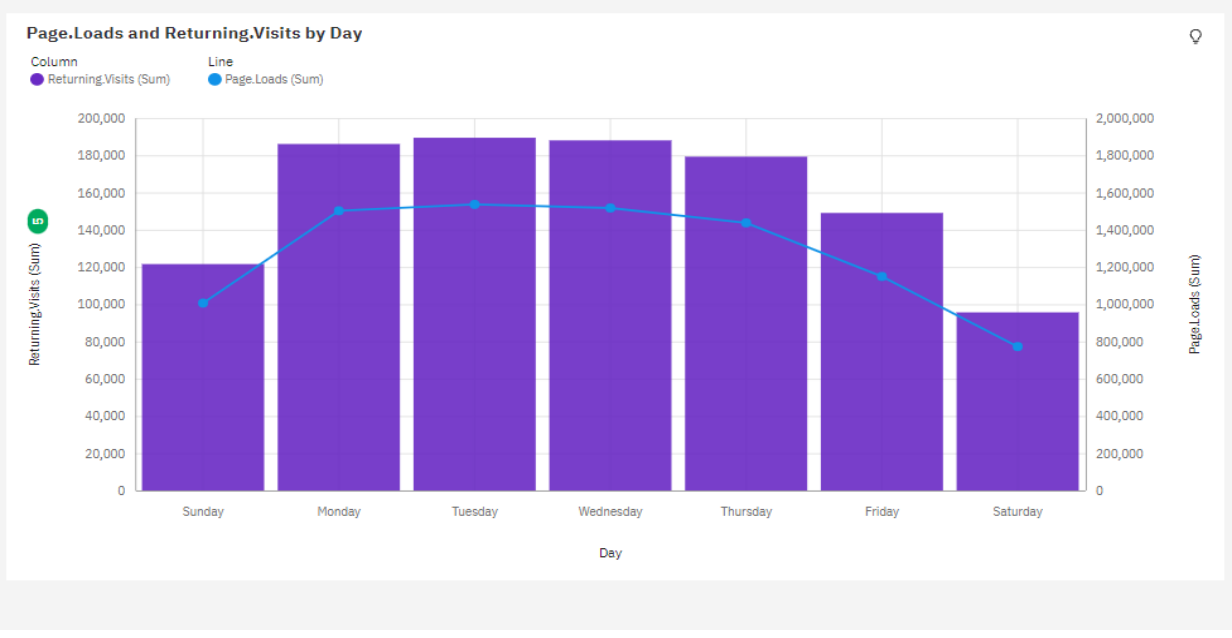
Page.Loads has a strong weekly trend. The largest values typically occur on Tuesday, whereas the smalest values on Saturday.
Based on the current forecasting, Page.Loads may reach nearly four thousand by Date 2021-10-27.
Over all dates, the average of First.Time.Visits is almost 2500.
Across all dates,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

4.2 exploration 2



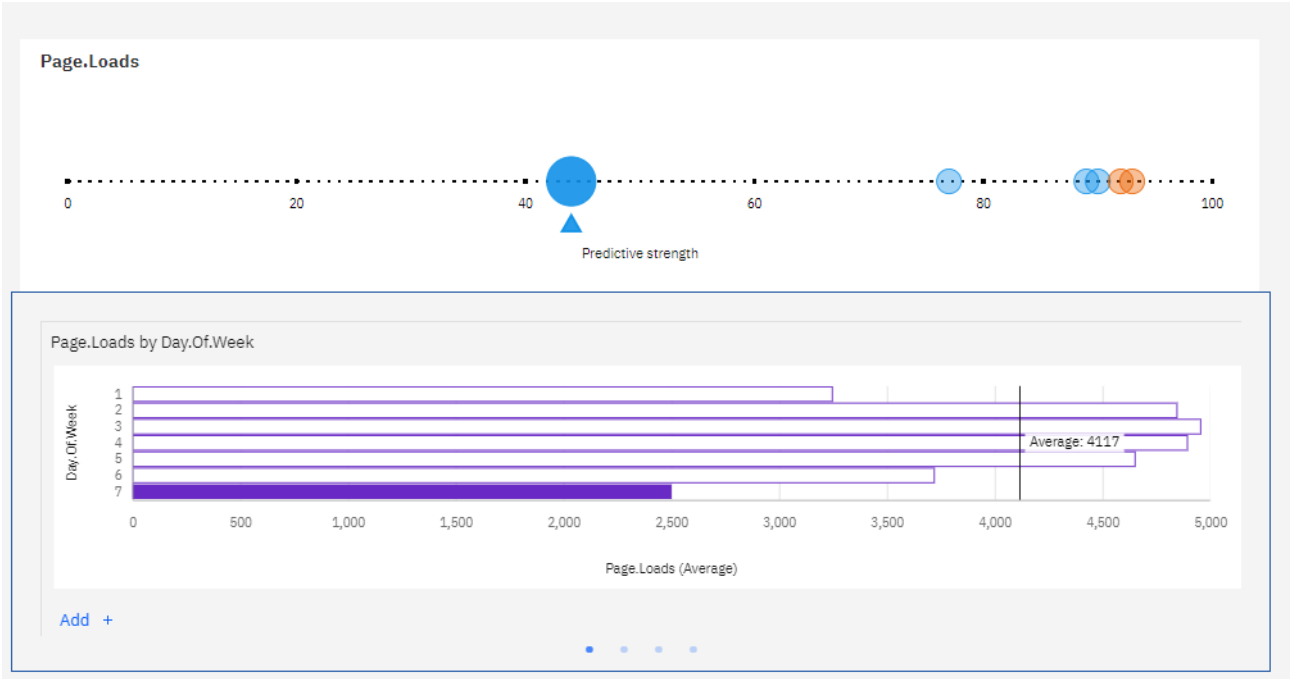
5.0 vizualization:

Utilize various visualization options such as bar charts, line charts, pie charts, and heat maps to represent data visually. Visualizations make it easier to identify patterns and outliers.



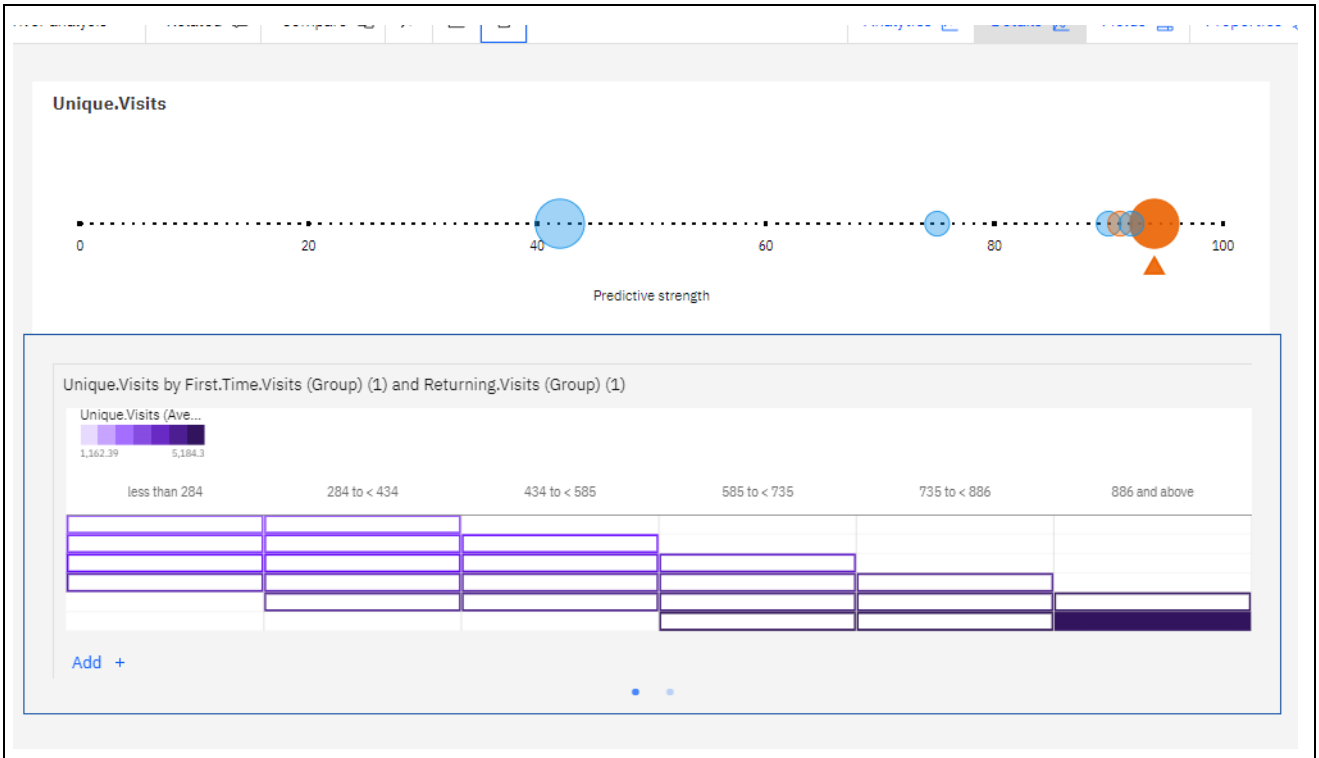
Analysis:

- 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.



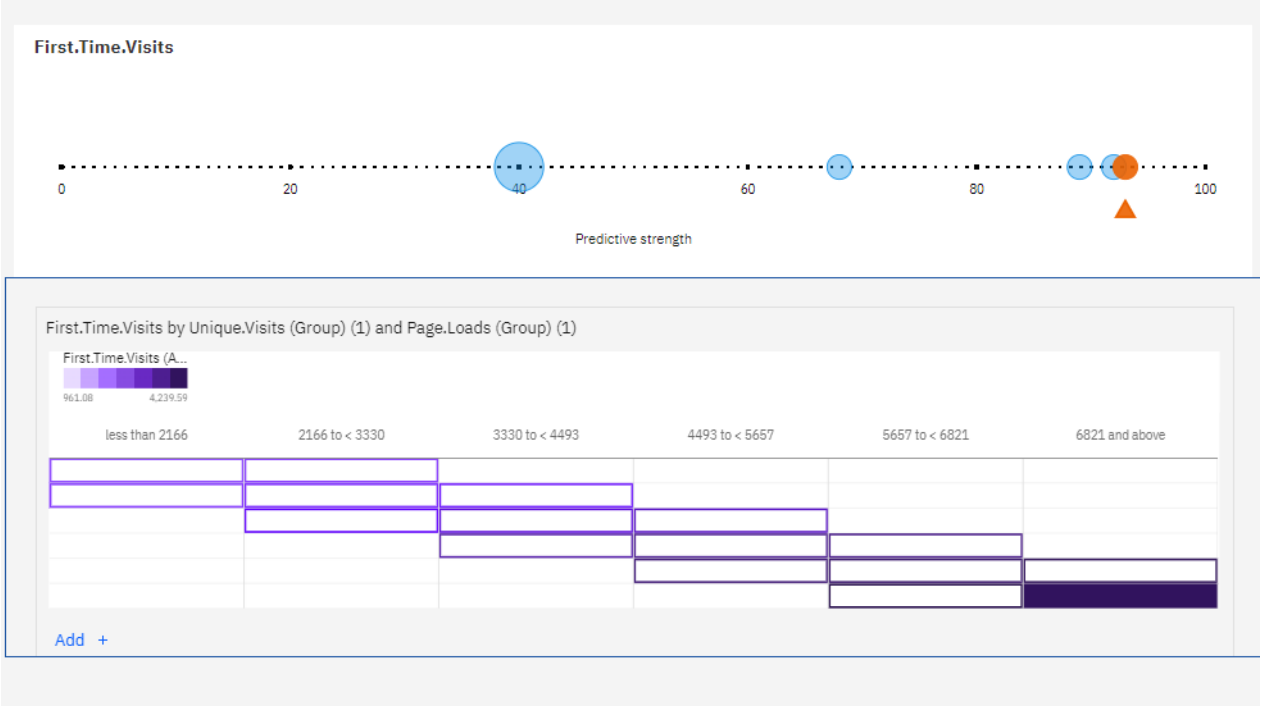
Analysis:

- 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 aGects **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).



Analysis:

- **First.Time.Visits (Group) (3)** strongly aGects **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 aGects **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 aGect
- **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)**
- 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**.



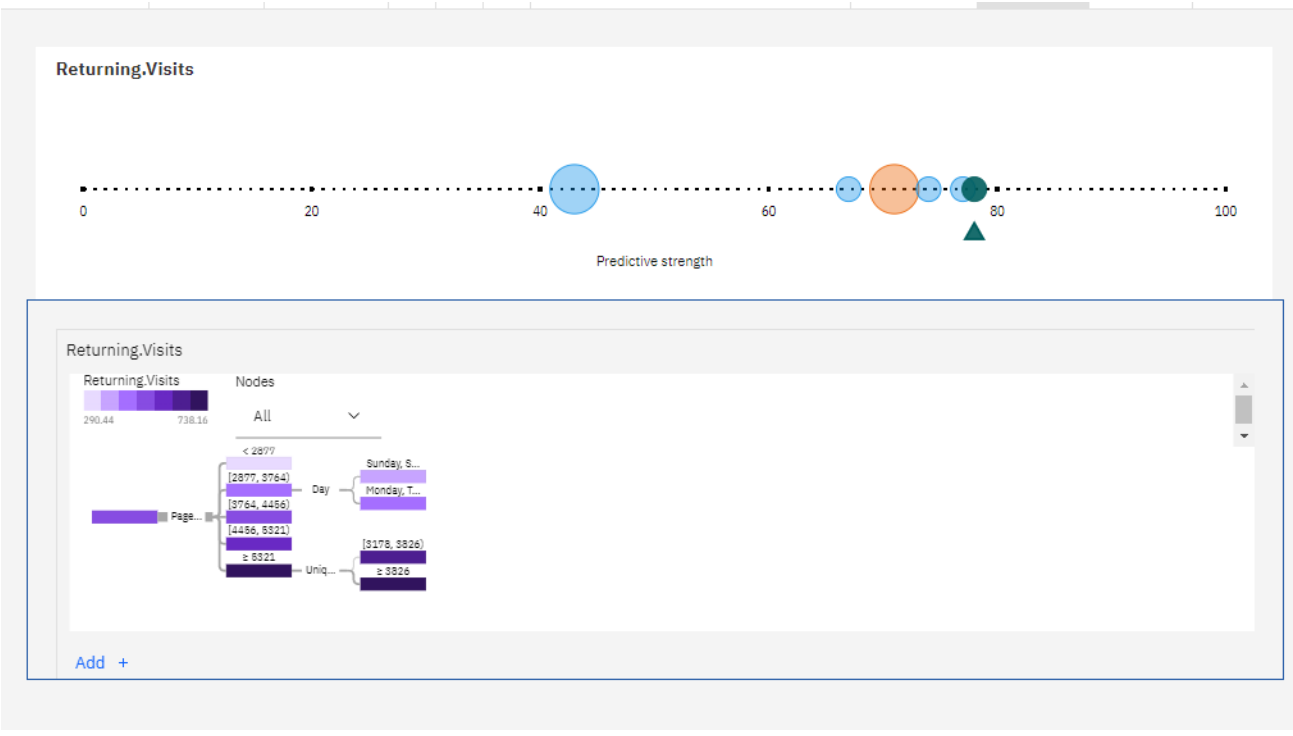
Analysis:

- **Unique.Visits** is unusually high when the combination of **First.Time.Visits (Group)**
- 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**.



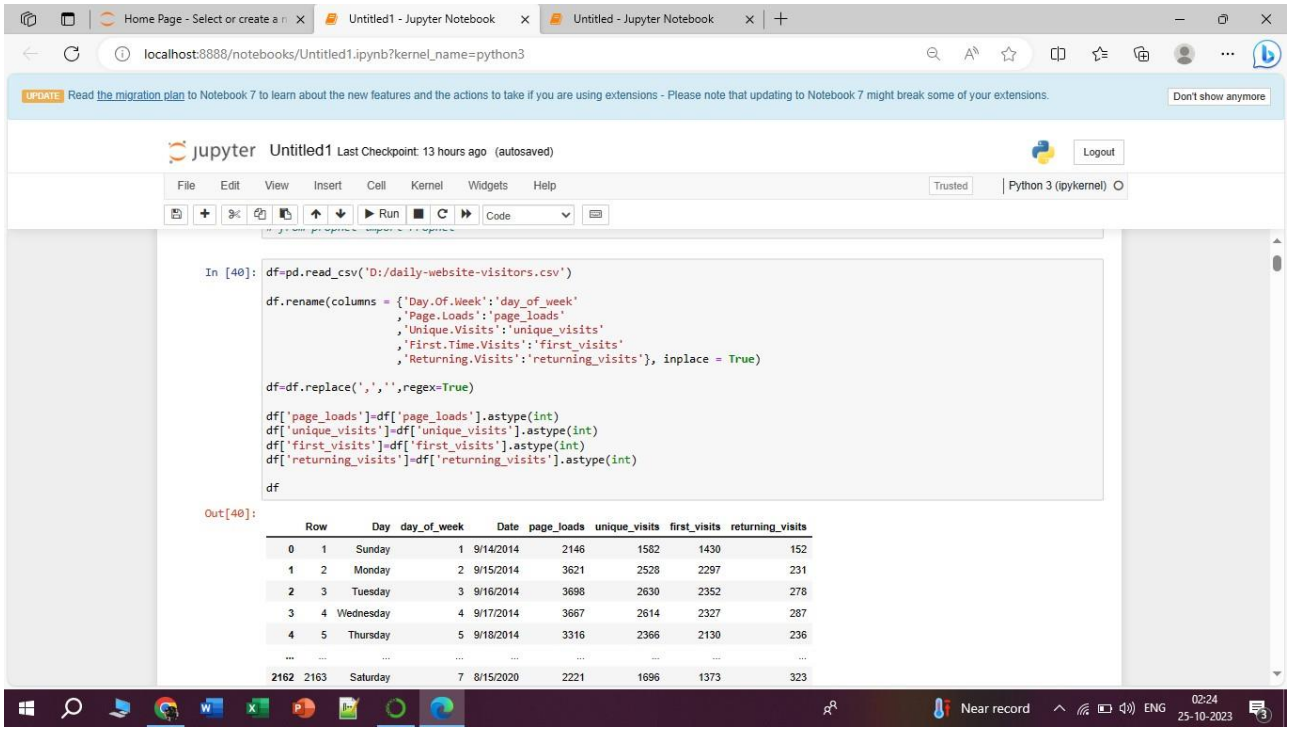
Analysis:

- **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.

6.0 vizualization with python using anaconda

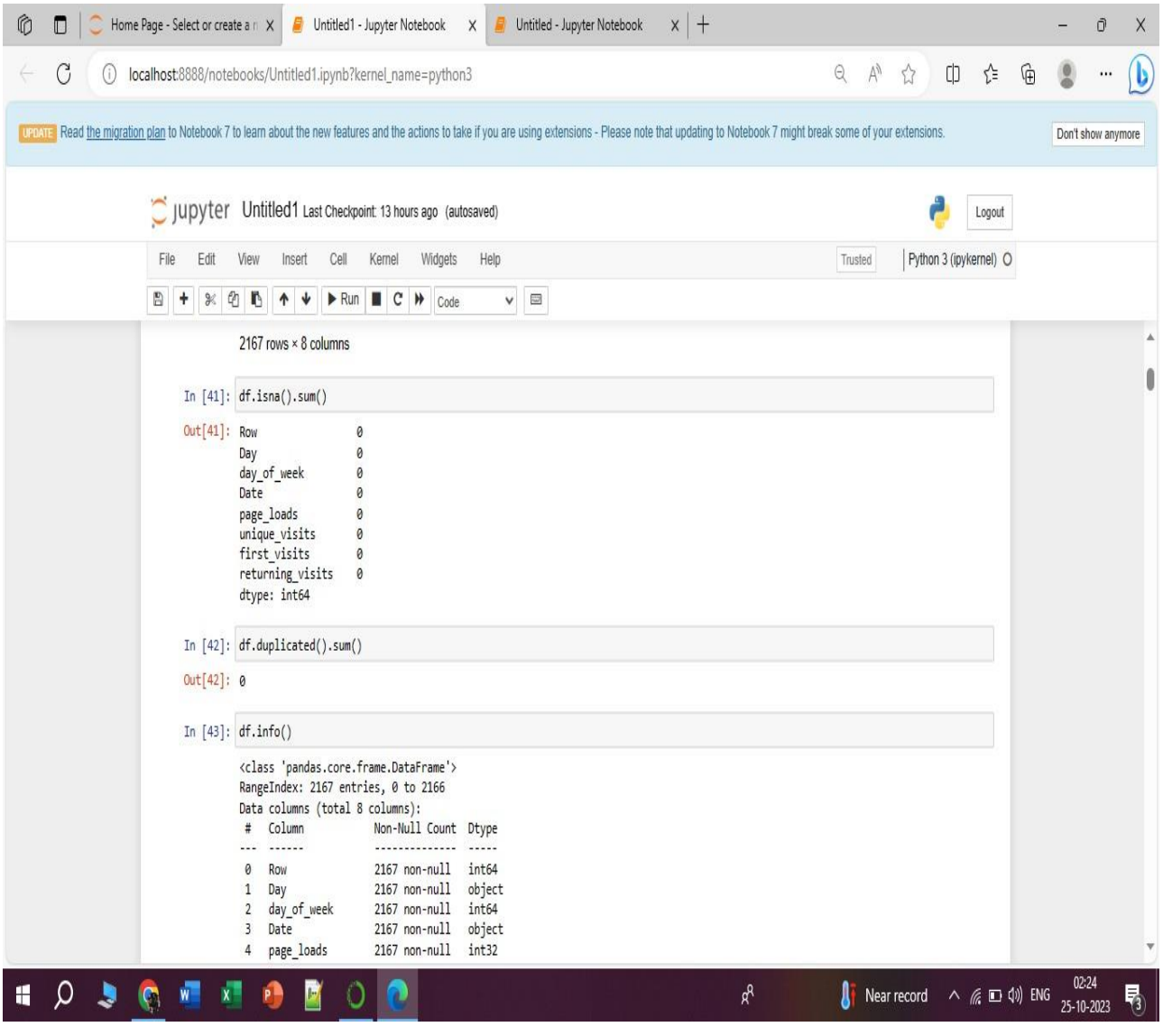
6.1 loading of dataset:



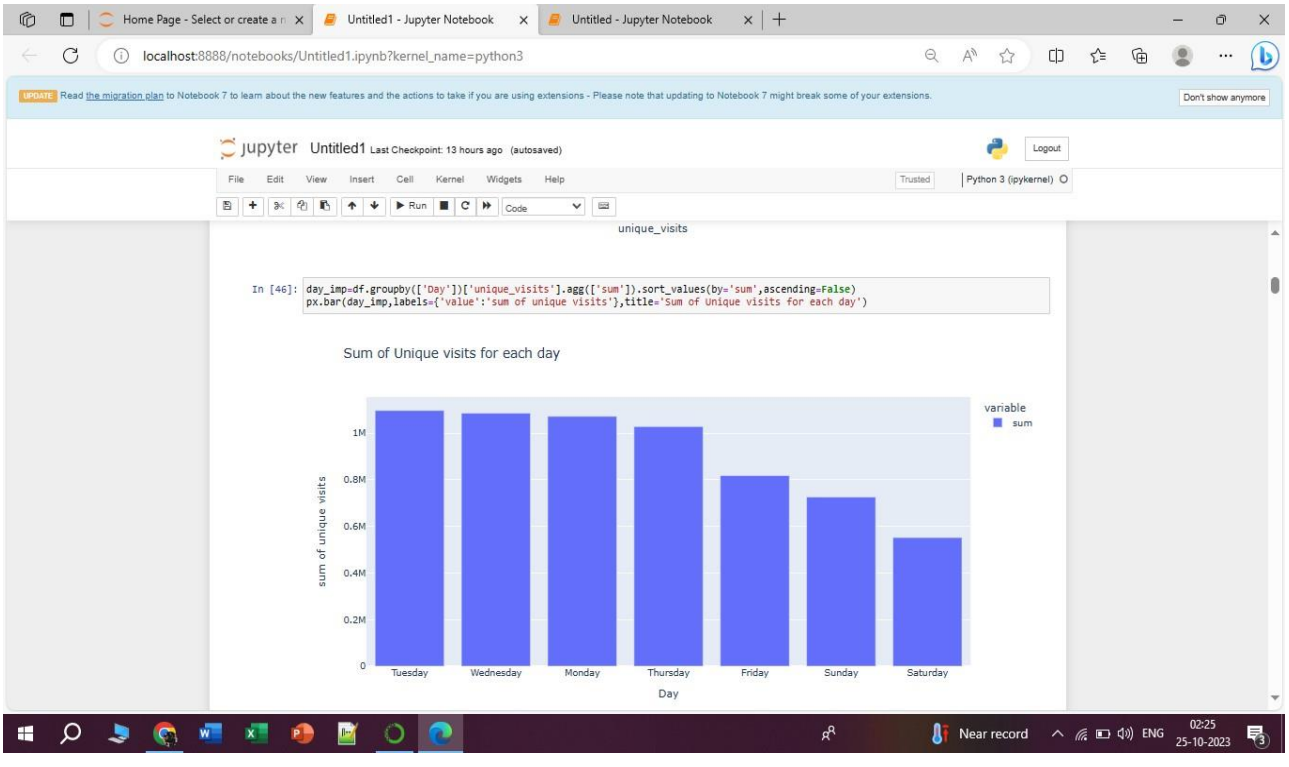
- Here we have loaded our dataset into the anaconda for further exploratory and visualization process.

6.2 data processing:

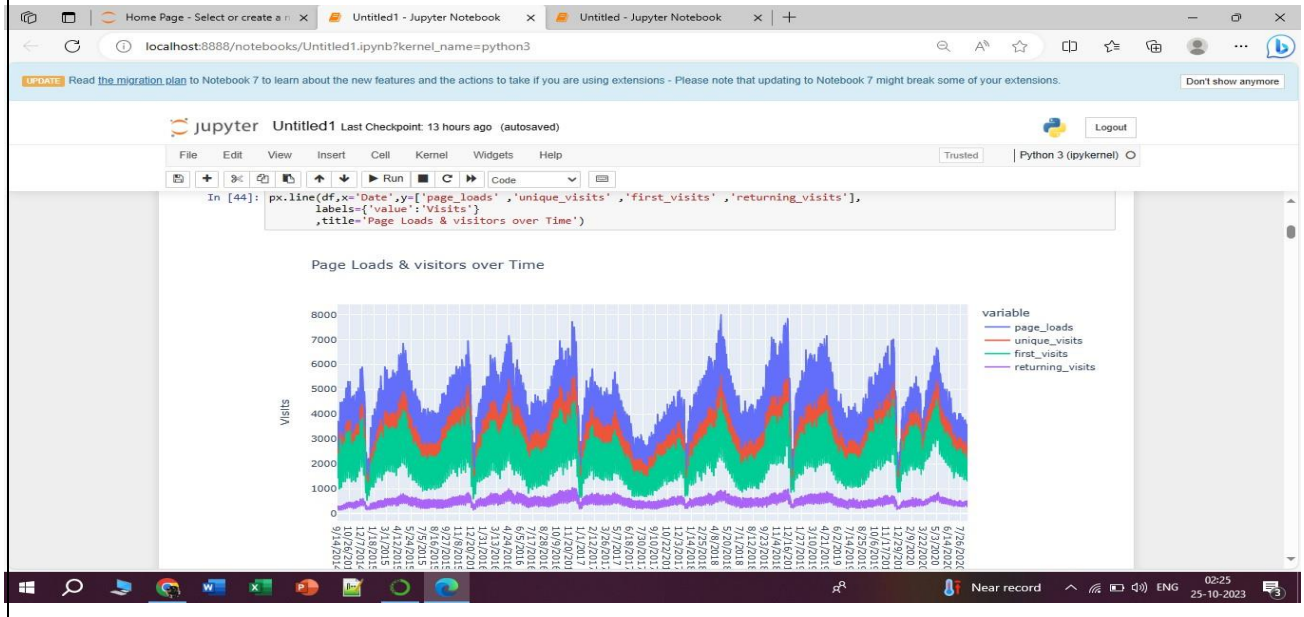
- Data processing refers to the conversion of raw data into meaningful information through a series of operations



6.3 data visualization:



- Creating a bar graph in Python using the Matplotlib library is straightforward



Peaks represents the highest reading in whole data

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

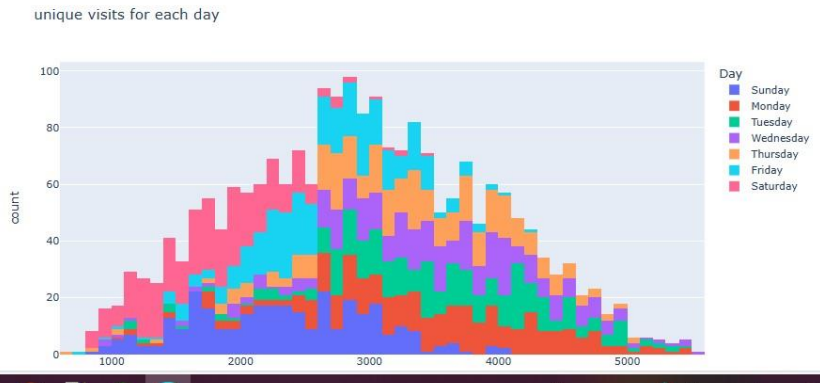
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved) Logout

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

Run Code

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

Home Page - Select or create a n | Untitled1 - Jupyter Notebook | Untitled - Jupyter Notebook

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

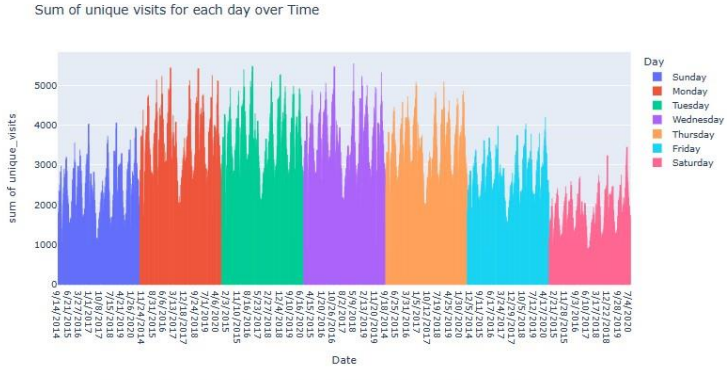
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

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

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

Run Code

In [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

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

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')
```

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

Home Page - Select or create a n | Untitled1 - Jupyter Notebook | Untitled - Jupyter Notebook

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

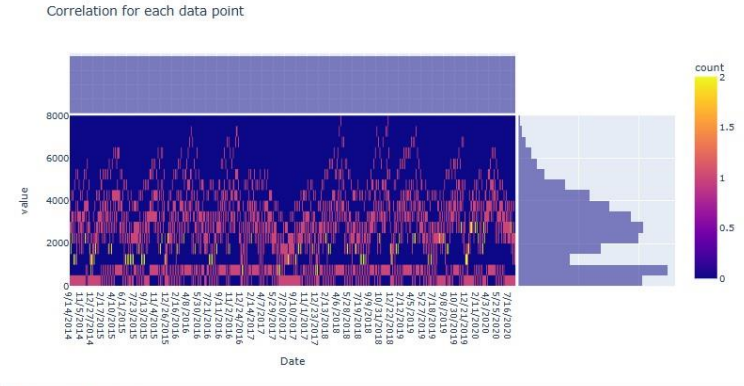
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

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

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

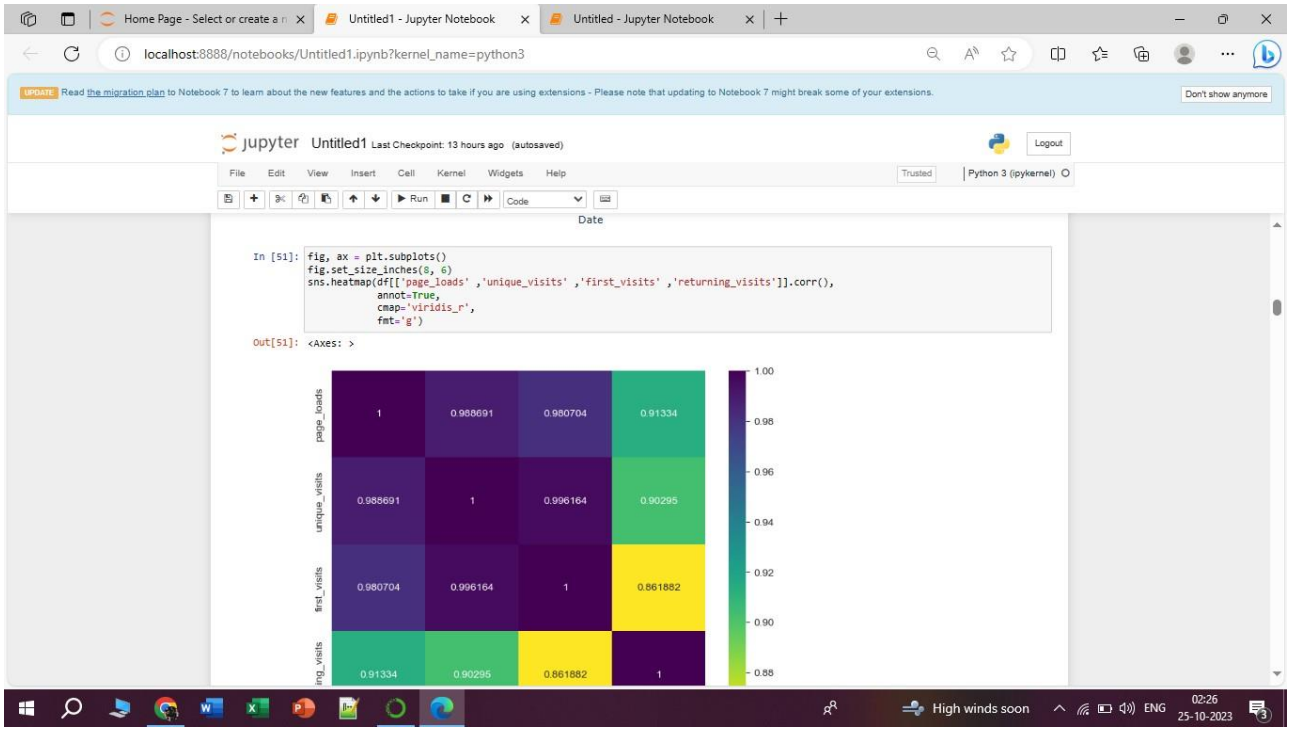
Run Code

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')
```



In [51]:
```
fig, ax = plt.subplots()
```

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



- Heatmap for the dataset is plotted

Conclusion

Website trafic 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 trafic data, enabling organizations to monitor key metrics, detect trends, and make informed decisions to drive business success. It oGers the ability to create interactive reports and dashboards, making it easier for teams to collaborate and act on the insights derived from website trafic data, ultimately leading to improved online performance and user engagement.