PHASE 4 : DEVELOPMENT PART 2

WEBSITE TRAFFIC ANALYSIS

PRIYADARSHINI ENGINEERING COLLEGE

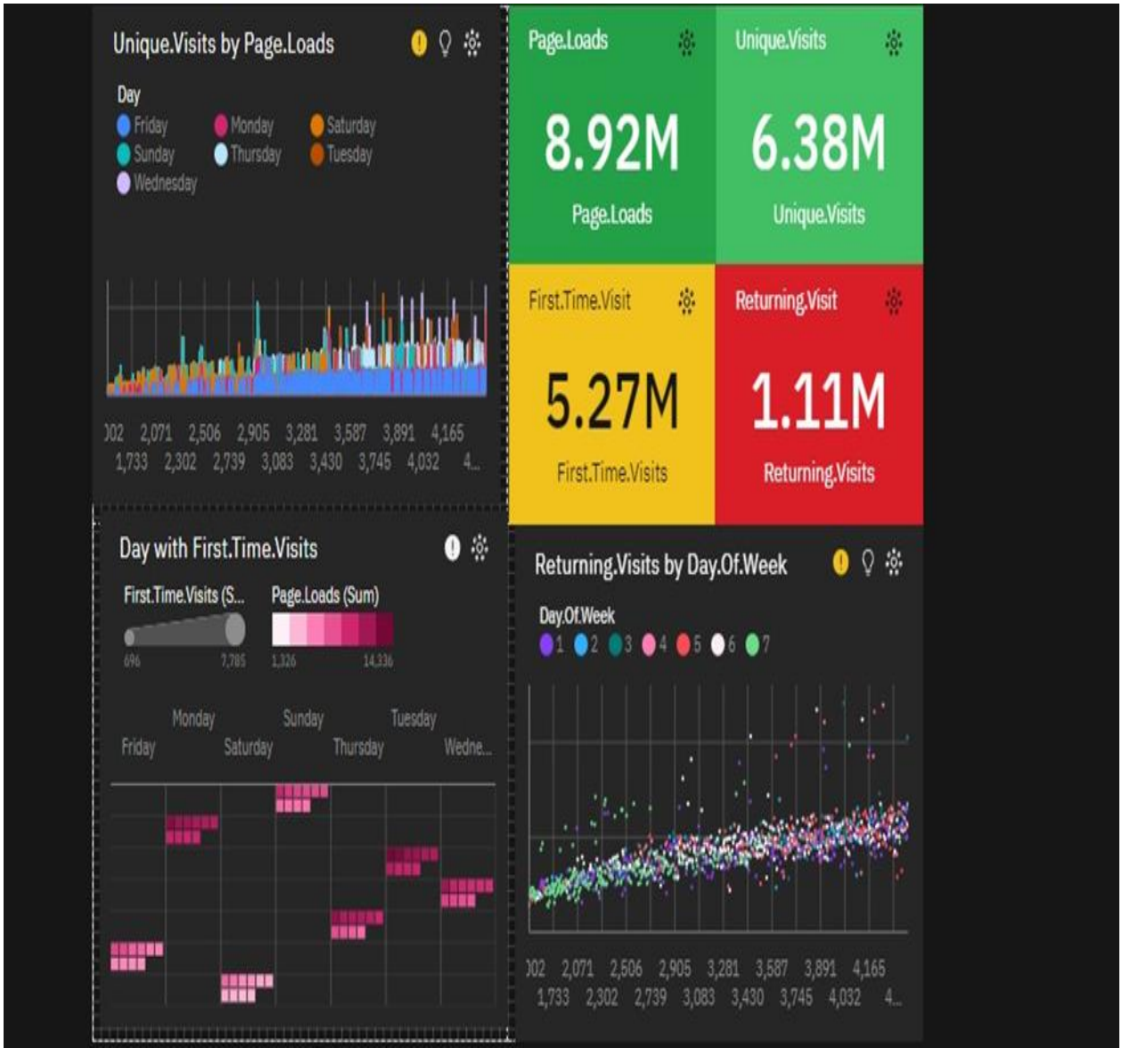
Ram Prasath/511921104055

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 tram c data for insights into user behavior, popular pages, and tram 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



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**.

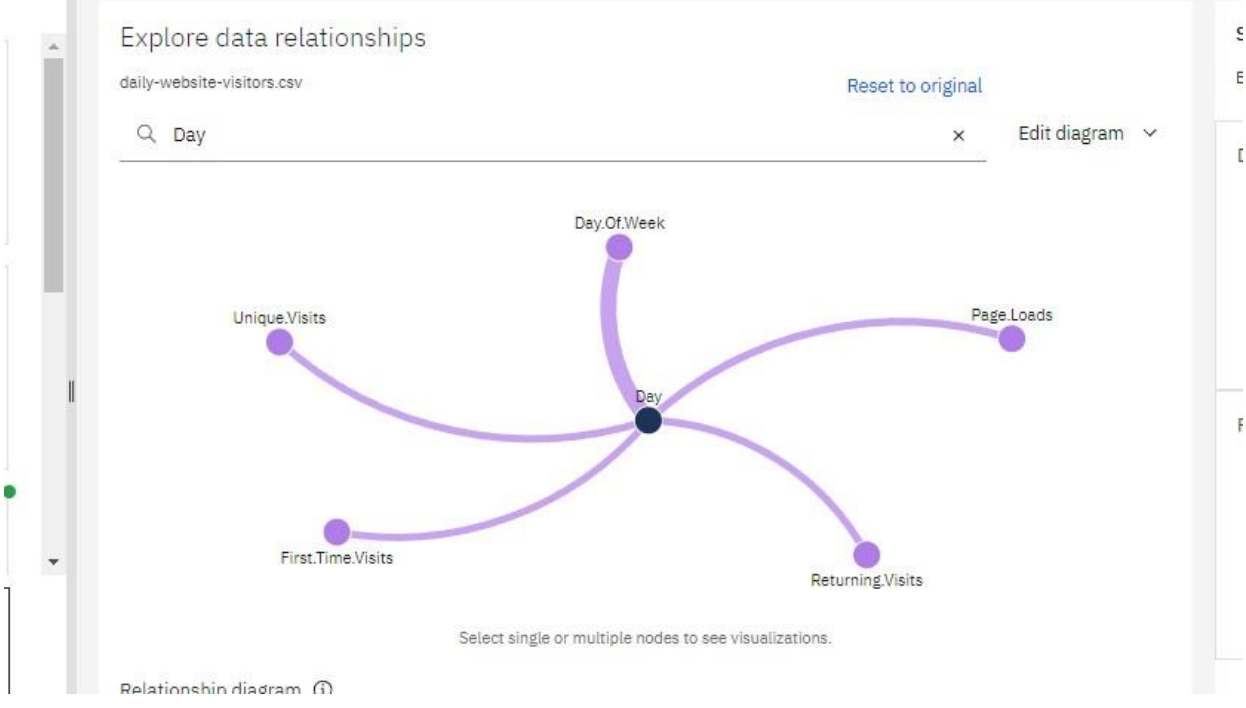
Over all **dates**, the average of **Returning.Visits** is **511.8**.

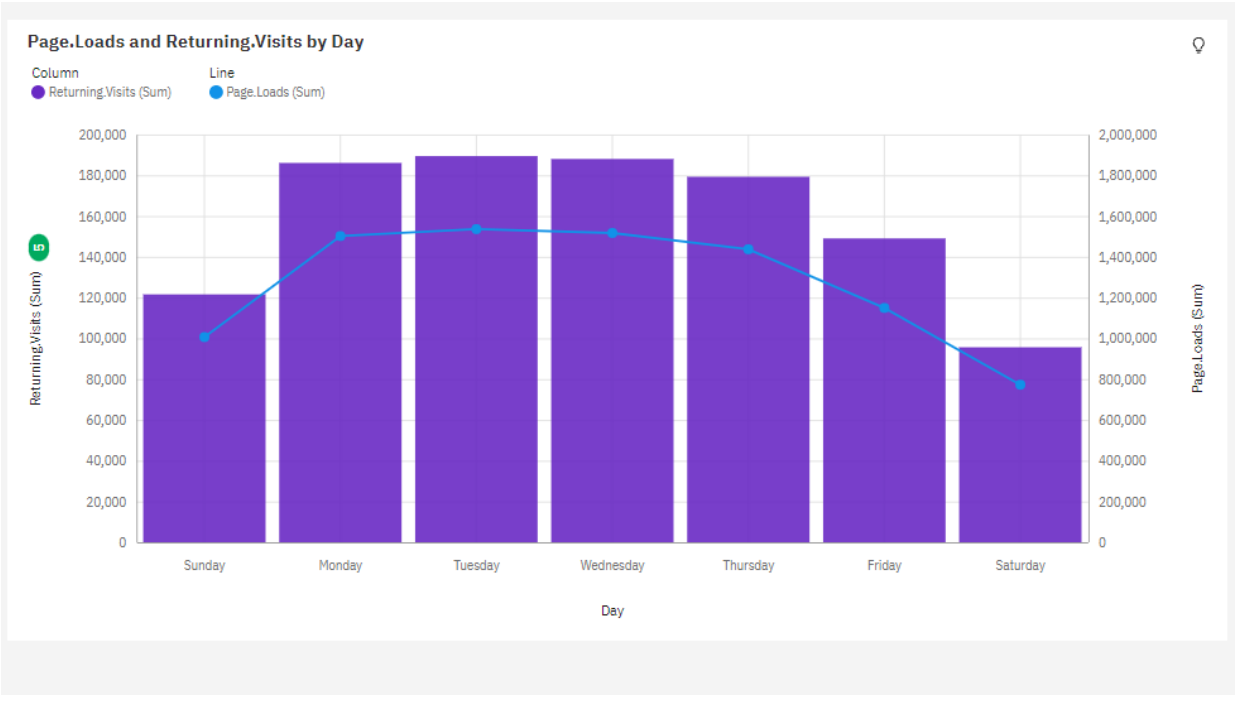
Over all **dates**, the average of **Unique.Visits** is **nearly three thousand**.

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**.


Page.Loads, Unique.Visits, First.Time.Visits and Returning.Visits by Date
Measures
Page.Loads
Unique.Visits
First.Time.Visits
Returning.Visits
Values
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
9/14/...
1/20/2015
5/28/2015
10/3/2015
2/8/2016
6/15/2016
10/21/2016
2/...
Date
Day for Unique.Visits
Column values
Increase
Decrease
Total
Day (Count di...
1,000
500
0
667
1,444
1,806
2,102
2,366
2,624
2,833
3,045
Unique.Visits
Day by Unique.Visits
Unique.Visits
895
3,081
4,214
4,284
4,448
1,944
1,634
0
0.2
0.4
0.6
0.8
1
Day (Count distinct)


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



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.

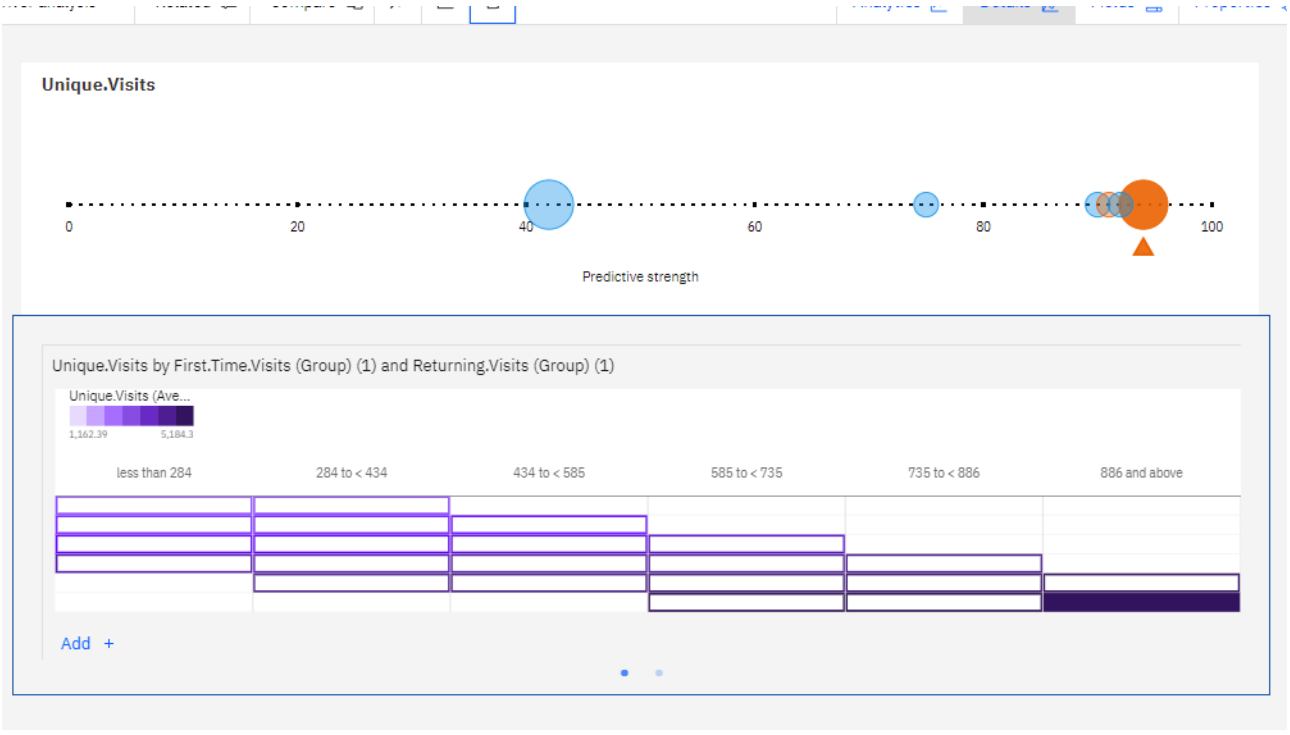
Across all values of **Day.Of.Week**, the average of **Page.Loads** is over four thousand.

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.

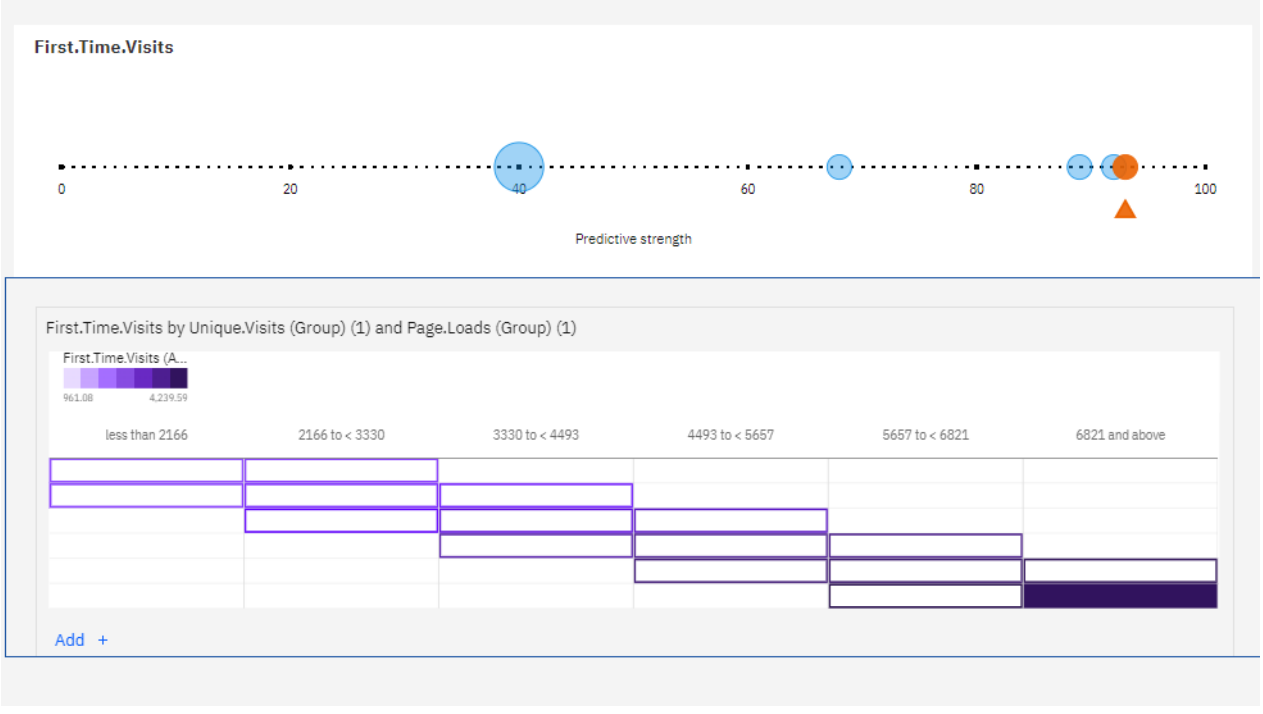
First.Time.Visits (Group) (3) and **Returning.Visits (Group) (2)** strongly affect **Unique.Visits** (96%).

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**.

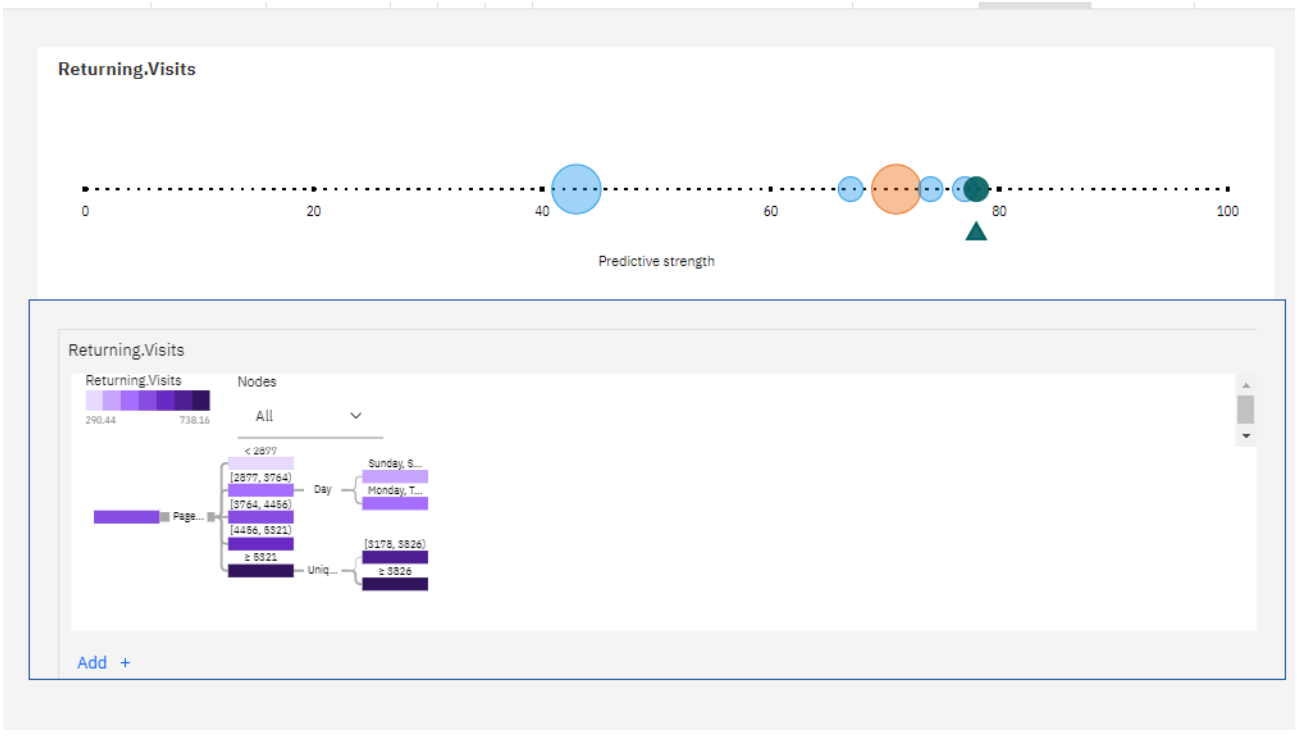


Unique.Visits is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.

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).

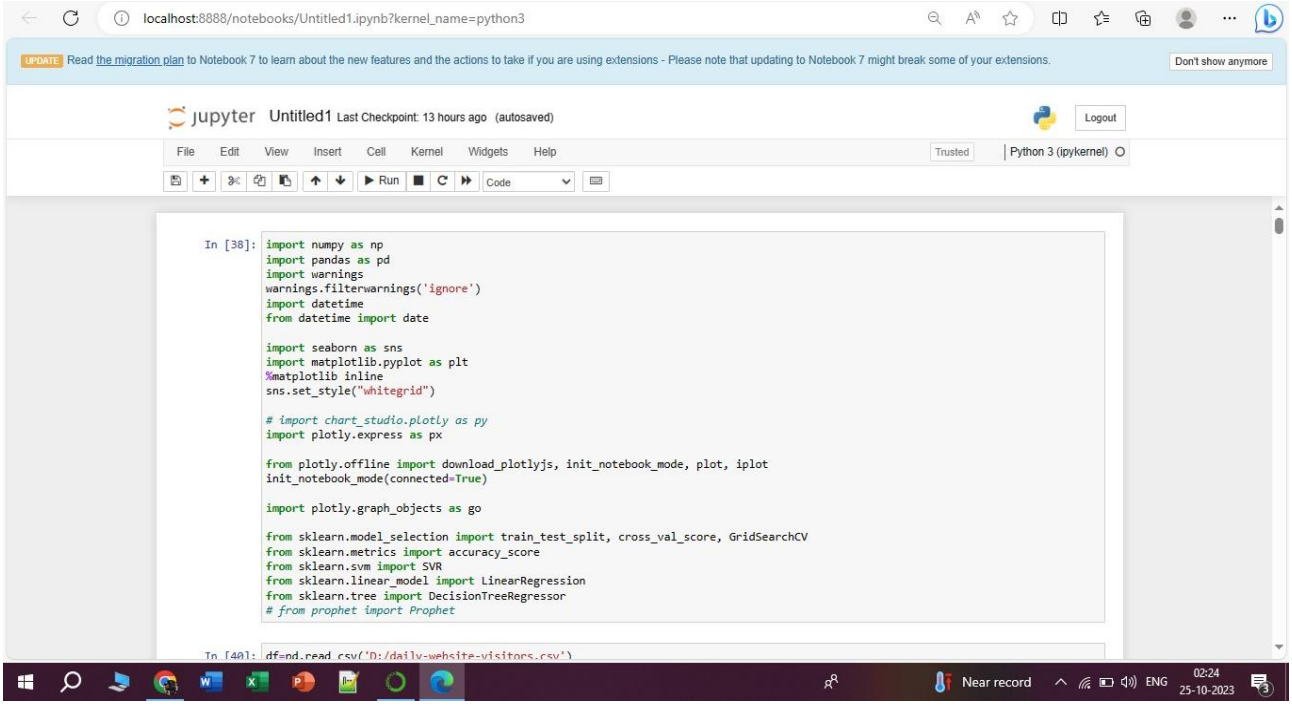
434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).

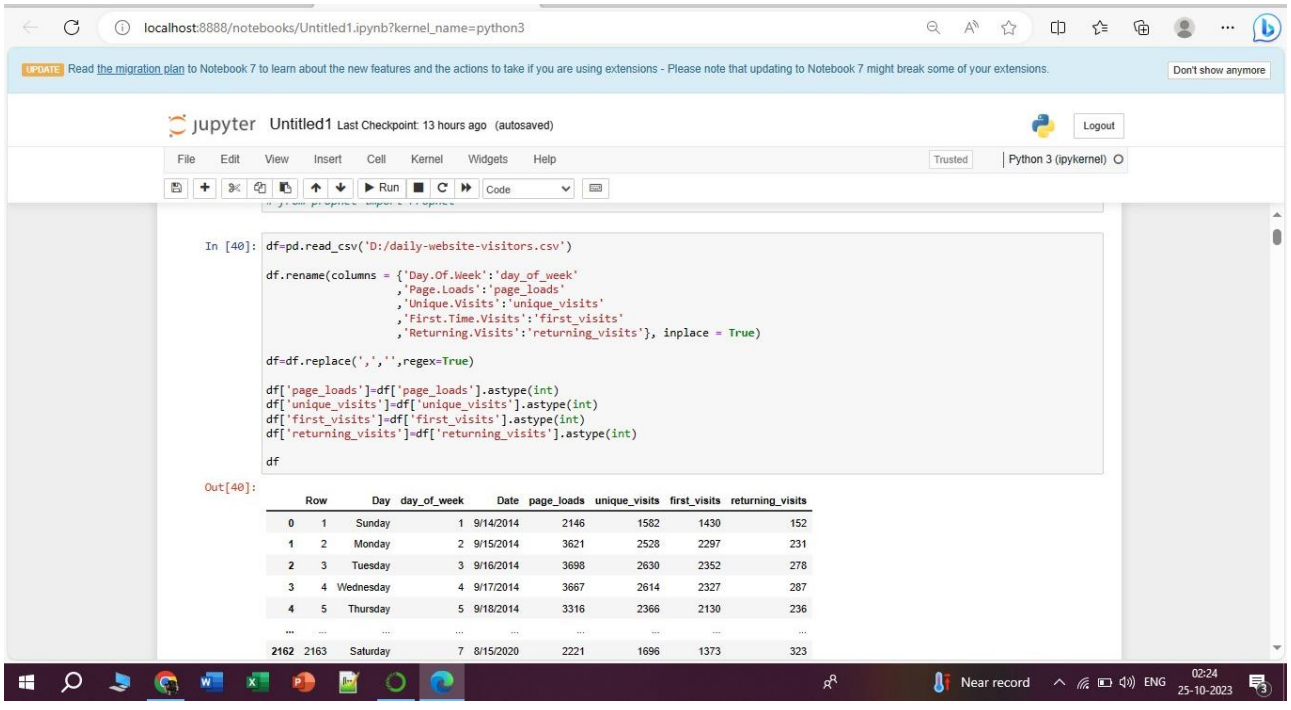
There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.

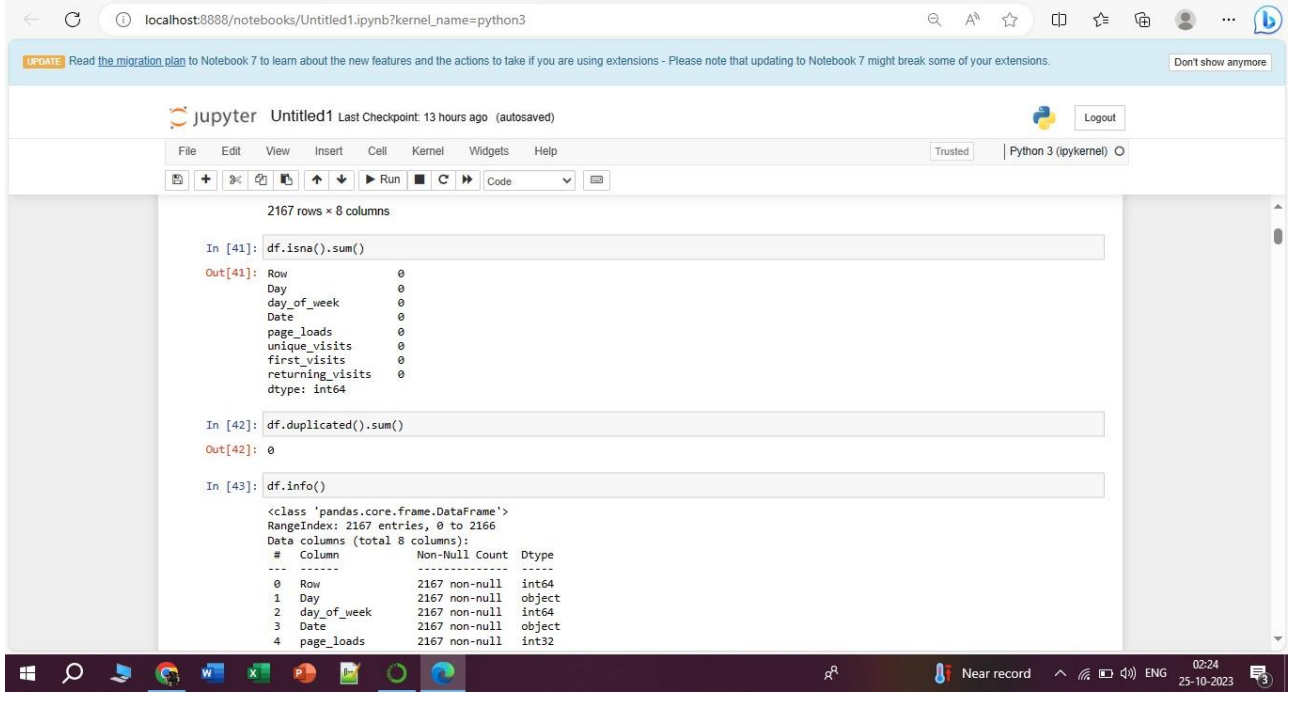


Page.Loads, **Unique.Visits**, and **Day** predict **Returning.Visits** with a strength of 78.1%.

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




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

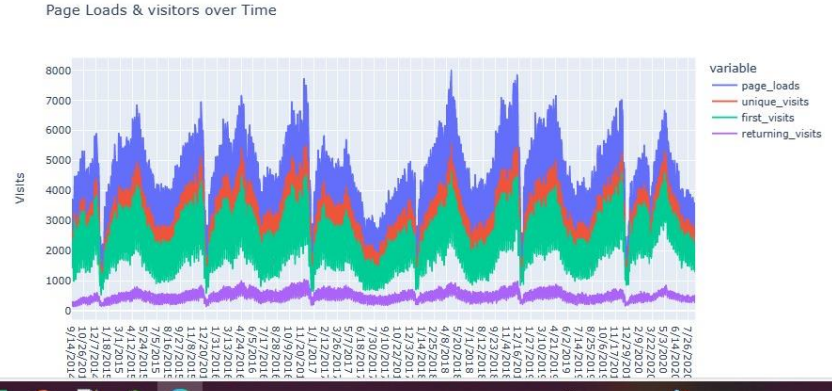
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

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

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

Run Code

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

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

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')
```

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

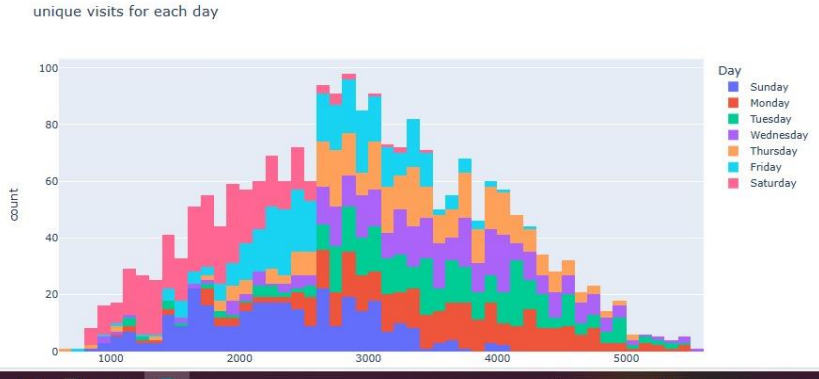
jupyter Untitled1 Last Checkpoint: 13 hours ago (autosaved) Logout

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

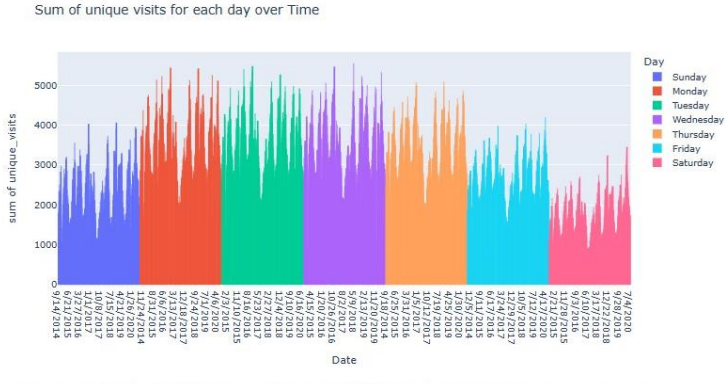
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

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

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

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

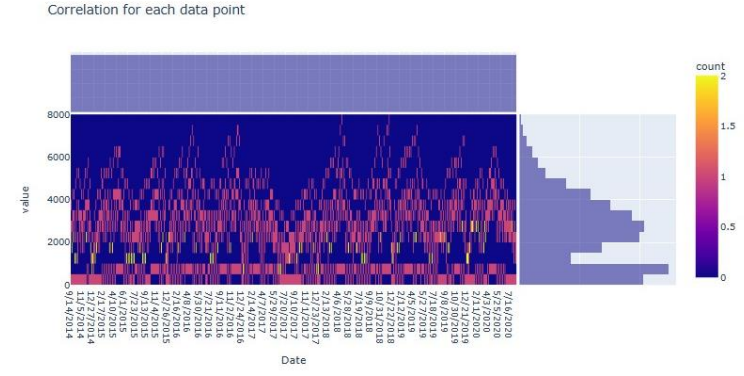
UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

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

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

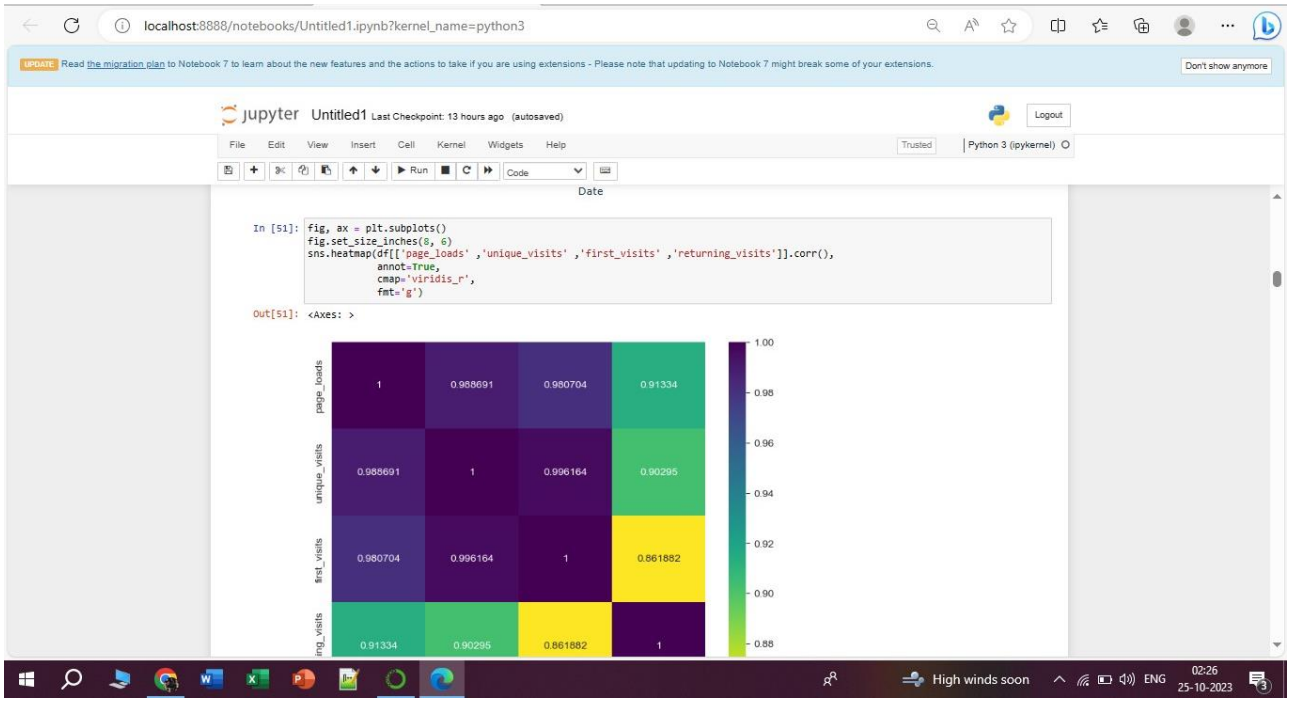
Run Code

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



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.