PROJECT TITLE : The Website Traffic Analysis

Project Definition: The project involves analyzing website traffic data to gain insights into user behavior, popular pages, and traffic sources. The goal is to help website owners enhance the user experience by understanding how visitors interact with the site. This project encompasses defining the analysis objectives, collecting website traffic data, using IBM Cognos for data visualization, and integrating Python code for advanced analysis.

PREPROCESSING OF DATA

Data Cleaning

```
In [13]: import pandas as pd

         df = pd.read_csv('daily-website-visitors.csv')


         df.fillna(130, inplace = True)

         print(df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
5	6	Friday	6	9/19/2014	2,815	1,863	1,622	241
6	7	Saturday	7	9/20/2014	1,658	1,118	985	133
7	8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
8	9	Monday	2	9/22/2014	3,638	2,586	2,312	274
9	10	Tuesday	3	9/23/2014	4,462	3,257	2,989	268
10	11	Wednesday	4	9/24/2014	4,414	3,175	2,891	284
11	12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
12	13	Friday	6	9/26/2014	3,323	2,249	2,033	216
13	14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
14	15	Sunday	1	9/28/2014	2,465	1,806	1,613	193
15	16	Monday	2	9/29/2014	4,096	2,873	2,577	296
16	17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
17	18	Wednesday	4	10/1/2014	4,124	2,849	2,541	308

```
In [4]: import pandas as pd

        df = pd.read_csv('daily-website-visitors.csv')

        new_df = df.dropna()

        print(new_df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
5	6	Friday	6	9/19/2014	2,815	1,863	1,622	241
6	7	Saturday	7	9/20/2014	1,658	1,118	985	133
7	8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
8	9	Monday	2	9/22/2014	3,638	2,586	2,312	274
9	10	Tuesday	3	9/23/2014	4,462	3,257	2,989	268
10	11	Wednesday	4	9/24/2014	4,414	3,175	2,891	284
11	12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
12	13	Friday	6	9/26/2014	3,323	2,249	2,033	216
13	14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
14	15	Sunday	1	9/28/2014	2,465	1,806	1,613	193
15	16	Monday	2	9/29/2014	4,096	2,873	2,577	296
16	17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
17	18	Wednesday	4	10/1/2014	4,124	2,849	2,541	308

In [5]:
```
import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df['Date'] = pd.to_datetime(df['Date'])

print(df.to_string())
```

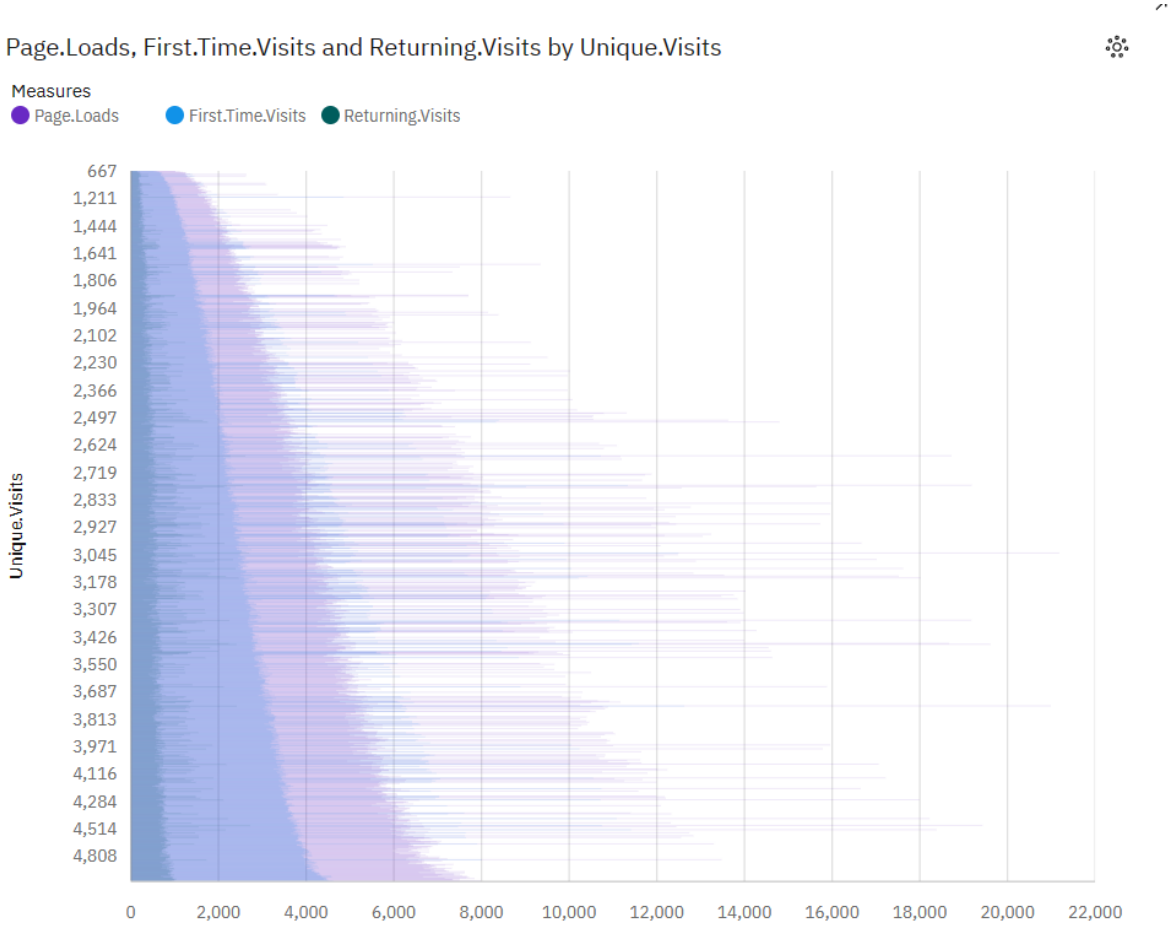
```
      Row        Day  Day.Of.Week       Date Page.Loads Unique.Visits First.Time.Visits Returning.Visits
0       1     Sunday            1 2014-09-14      2,146        1,582             1,430               152
1       2     Monday            2 2014-09-15      3,621        2,528             2,297               231
2       3    Tuesday            3 2014-09-16      3,698        2,630             2,352               278
3       4  Wednesday            4 2014-09-17      3,667        2,614             2,327               287
4       5   Thursday            5 2014-09-18      3,316        2,366             2,130               236
5       6     Friday            6 2014-09-19      2,815        1,863             1,622               241
6       7   Saturday            7 2014-09-20      1,658        1,118               985               133
7       8     Sunday            1 2014-09-21      2,288        1,656             1,481               175
8       9     Monday            2 2014-09-22      3,638        2,586             2,312               274
9      10    Tuesday            3 2014-09-23      4,462        3,257             2,989               268
10     11  Wednesday            4 2014-09-24      4,414        3,175             2,891               284
11     12   Thursday            5 2014-09-25      4,315        3,029             2,743               286
12     13     Friday            6 2014-09-26      3,323        2,249             2,033               216
13     14   Saturday            7 2014-09-27      1,656        1,180             1,040               140
14     15     Sunday            1 2014-09-28      2,465        1,806             1,613               193
15     16     Monday            2 2014-09-29      4,096        2,873             2,577               296
16     17    Tuesday            3 2014-09-30      4,474        3,032             2,720               312
17     18  Wednesday            4 2014-10-01      4,124        2,849             2,541               308
```

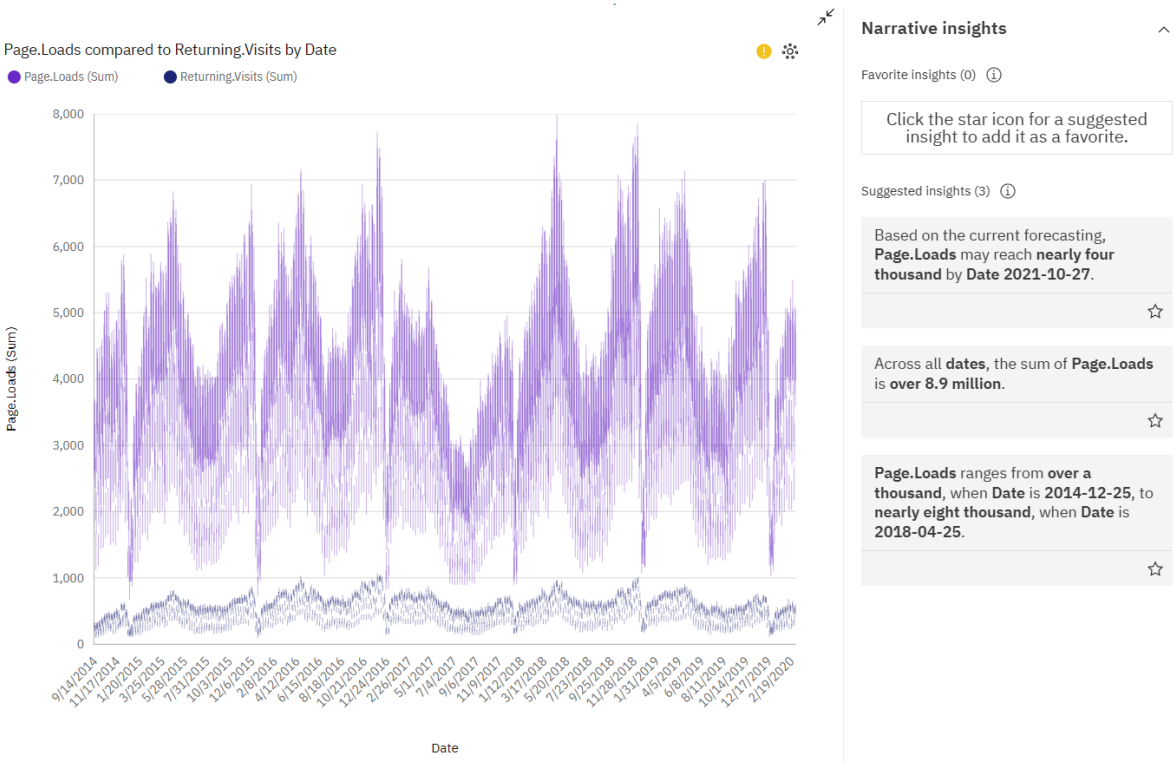
In [9]:
```
import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df.drop_duplicates(inplace = True)

print(df.to_string())
```

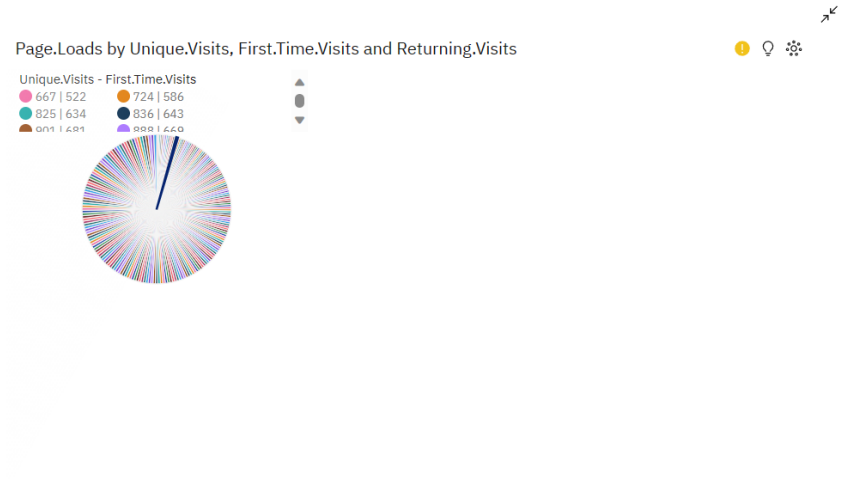
```
      Row        Day  Day.Of.Week       Date Page.Loads Unique.Visits First.Time.Visits Returning.Visits
0       1     Sunday            1  9/14/2014      2,146        1,582             1,430               152
1       2     Monday            2  9/15/2014      3,621        2,528             2,297               231
2       3    Tuesday            3  9/16/2014      3,698        2,630             2,352               278
3       4  Wednesday            4  9/17/2014      3,667        2,614             2,327               287
4       5   Thursday            5  9/18/2014      3,316        2,366             2,130               236
5       6     Friday            6  9/19/2014      2,815        1,863             1,622               241
6       7   Saturday            7  9/20/2014      1,658        1,118               985               133
7       8     Sunday            1  9/21/2014      2,288        1,656             1,481               175
8       9     Monday            2  9/22/2014      3,638        2,586             2,312               274
9      10    Tuesday            3  9/23/2014      4,462        3,257             2,989               268
10     11  Wednesday            4  9/24/2014      4,414        3,175             2,891               284
11     12   Thursday            5  9/25/2014      4,315        3,029             2,743               286
12     13     Friday            6  9/26/2014      3,323        2,249             2,033               216
13     14   Saturday            7  9/27/2014      1,656        1,180             1,040               140
14     15     Sunday            1  9/28/2014      2,465        1,806             1,613               193
15     16     Monday            2  9/29/2014      4,096        2,873             2,577               296
16     17    Tuesday            3  9/30/2014      4,474        3,032             2,720               312
17     18  Wednesday            4  10/1/2014      4,124        2,849             2,541               308
```

In [34]:
```
import pandas as pd


x=df.to_csv('C:\\Users\yoges\Documents\daily-website-visitors.csv')
```

VISUALATION IS USING IBM COGNOS ANALYTICS







Insights

Narrative insights

Favorite insights (0)

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (8)

Returning.Visits 552 has the highest total **Page.Loads** due to **First.Time.Visits 3635**.

It is projected that by **Monday+1**, **2310** will exceed **2203** in **Page.Loads** by **366**.

It is projected that by **Monday+1**, **710** will exceed **552** in **Page.Loads** by **almost 1500**.

It is projected that by **Monday+1**, **2844** will exceed **3039** in **Page.Loads** by **410**.

Narrative insights

Favorite insights (0)

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (9)

It is projected that by **Monday+1**, **4205** will exceed **3973** in **First.Time.Visits** by **a thousand**.

2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **First.Time.Visits** values (**0.2** % of the total).

2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Unique.Visits** values (**0.2** % of the total).

Over all values of **Page.Loads**, the average of **First.Time.Visits** is **almost 2500**.

Over all values of **Page.Loads**, the

Insights

Narrative insights

Favorite insights (0)

Click the star icon for a suggested insight to add it as a favorite.

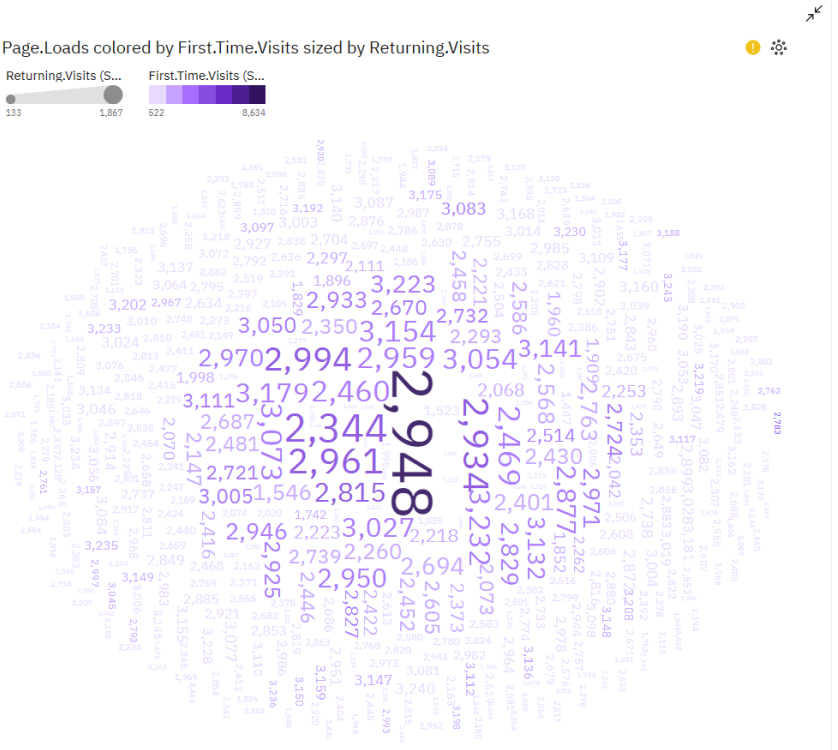
Suggested insights (4)

Returning.Visits is most unusual when the values of **First.Time.Visits** are **2989, 8761, 10,517, 2891, 5599** and more.

It is projected that by **Monday+1**, **4205** will exceed **3973** in **First.Time.Visits** by **a thousand**.

Returning.Visits and **First.Time.Visits** have a **strong** positive linear association, being **Returning.Visits = 39.86 + 0.1992 * First.Time.Visits**

The total of **Returning.Visits** is **over 1.1 million**.



Narrative insights

Favorite insights (0)

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (9)

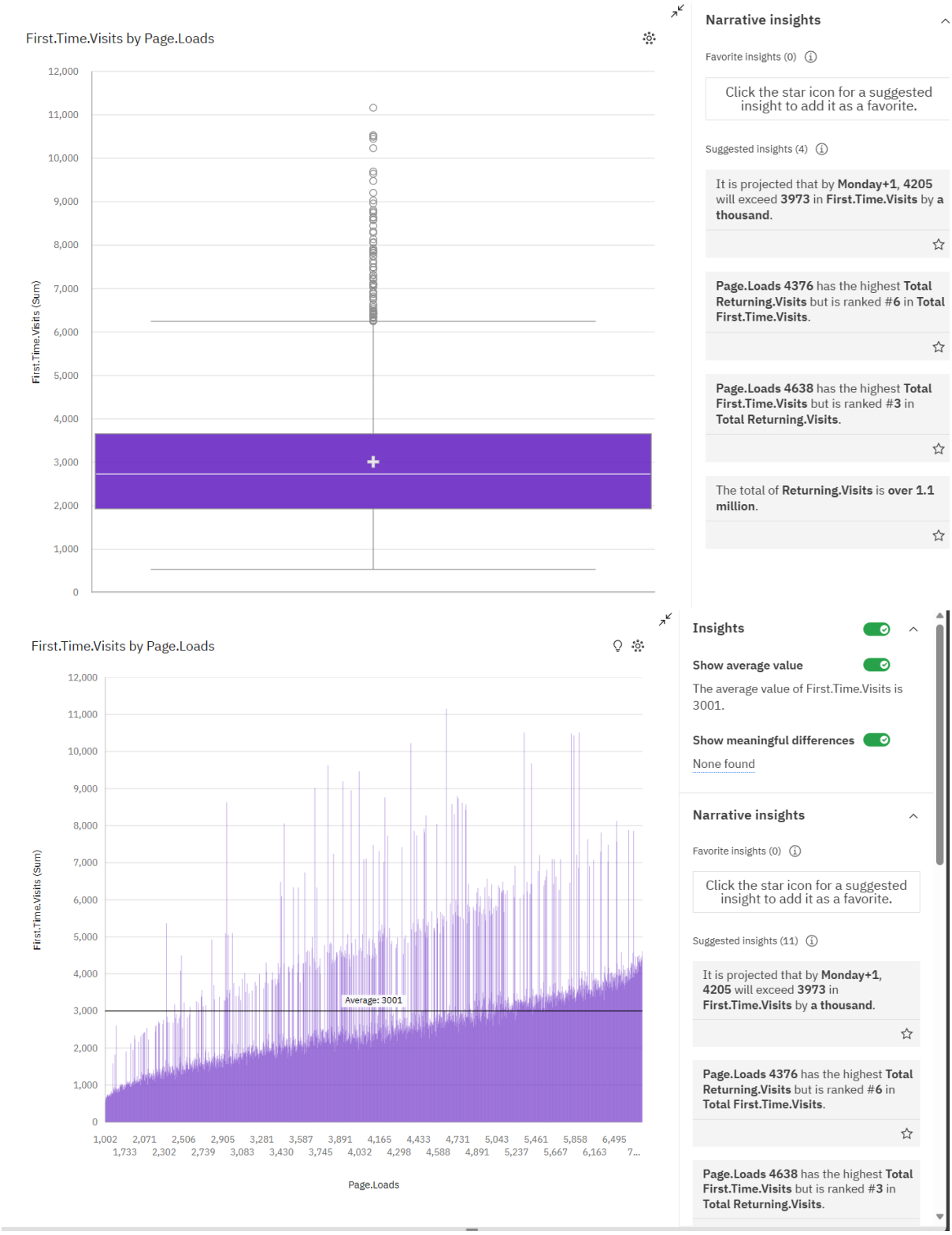
It is projected that by **Monday+1**, **4205** will exceed **3973** in **First.Time.Visits** by **a thousand**.

2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Returning.Visits** values (**0.2** % of the total).

2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Unique.Visits** values (**0.2** % of the total).

Over all values of **Page.Loads**, the average of **Returning.Visits** is **511.8**.

Over all values of **Page.Loads**, the average of **Unique.Visits** is **nearly**



CONCLUSION

In conclusion, leveraging IBM Cognos Analytics for website traffic analysis involves meticulous data preprocessing and insightful visualization techniques. By

ensuring data accuracy and employing compelling visualizations, businesses can extract meaningful insights, enabling informed decision-making and optimizing online performance.