Website Traffic Analysis

Introduction:

In this Website Traffic Analysis Project, where we embark on the journey of transforming raw data into actionable insights. Our primary focus is on the critical phases of data cleaning and visualization. In this project, we commit to the meticulous process of data cleaning, ensuring that our insights are based on accurate and reliable information. To make sense of our website traffic data, we'll leverage the power of data visualization. Through charts, graphs, and interactive dashboards, we aim to present complex data in a visually engaging and understandable manner.

Data Collection and Preparation:

- Gather website traffic data from analytics tools and prepare it for analysis.
- Clean the data by handling missing values, duplicates, and inconsistencies.
- Convert data into a suitable format and ensure consistent date/time formatting.

Data Source:

- A good data source for website traffic analysis using machine learning should be Accurate, Complete, Accessible.

Dataset: (“ https://www.kaggle.com/datasets/bobnau/daily-website-visitors ”)

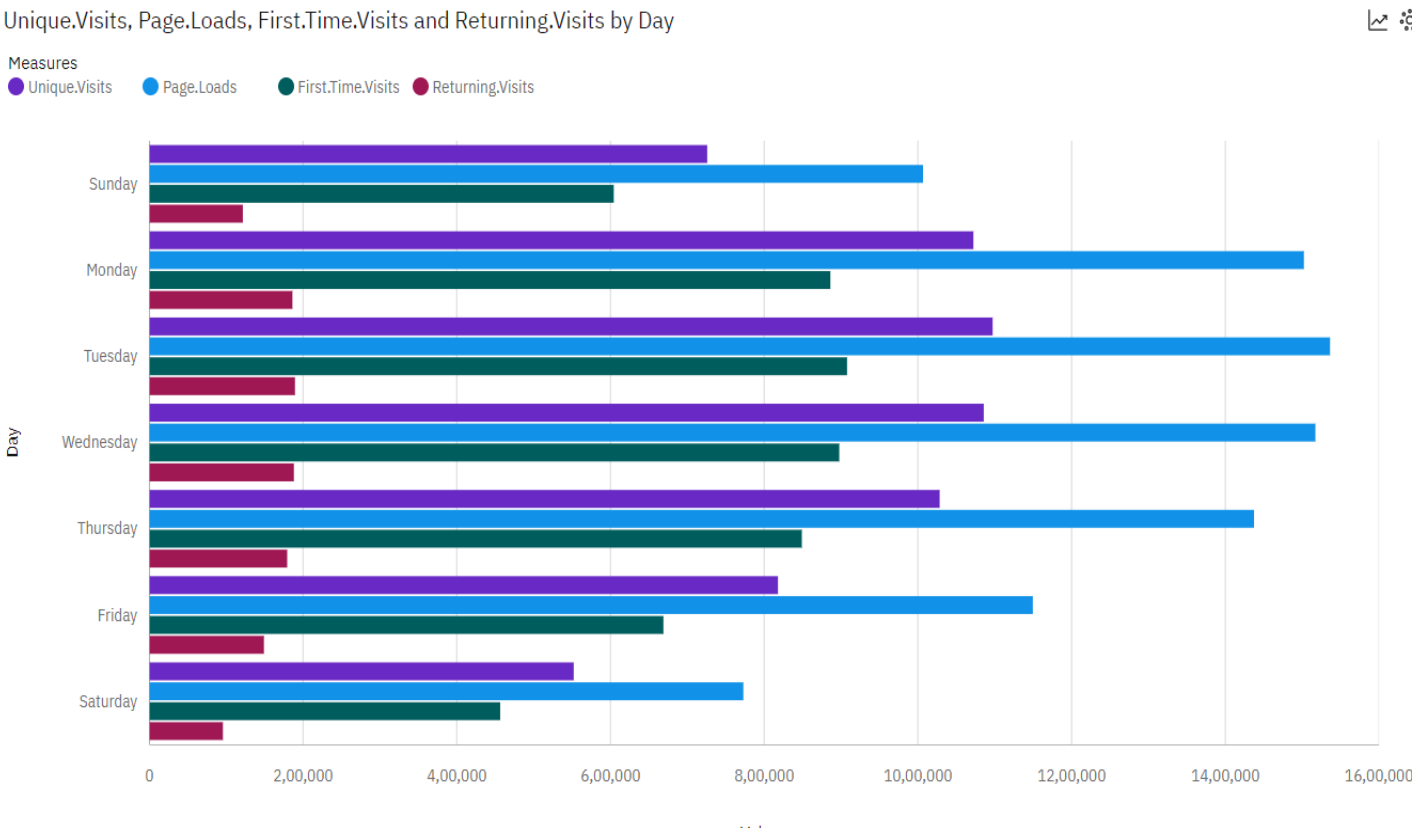
Date	Row	Day	Day.Of.Week	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
2014-09-14	1	Sunday	1	2146	1582	1430	152
2014-09-15	2	Monday	2	3621	2528	2297	231
2014-09-16	3	Tuesday	3	3698	2630	2352	278
2014-09-17	4	Wednesday	4	3667	2614	2327	287
2014-09-18	5	Thursday	5	3316	2366	2130	236
...	...	...	...	...	...	...	...
2020-08-15	2163	Saturday	7	2221	1696	1373	323
2020-08-16	2164	Sunday	1	2724	2037	1686	351
2020-08-17	2165	Monday	2	3456	2638	2181	457
2020-08-18	2166	Tuesday	3	3581	2683	2184	499
2020-08-19	2167	Wednesday	4	2064	1564	1297	267

Visualization:

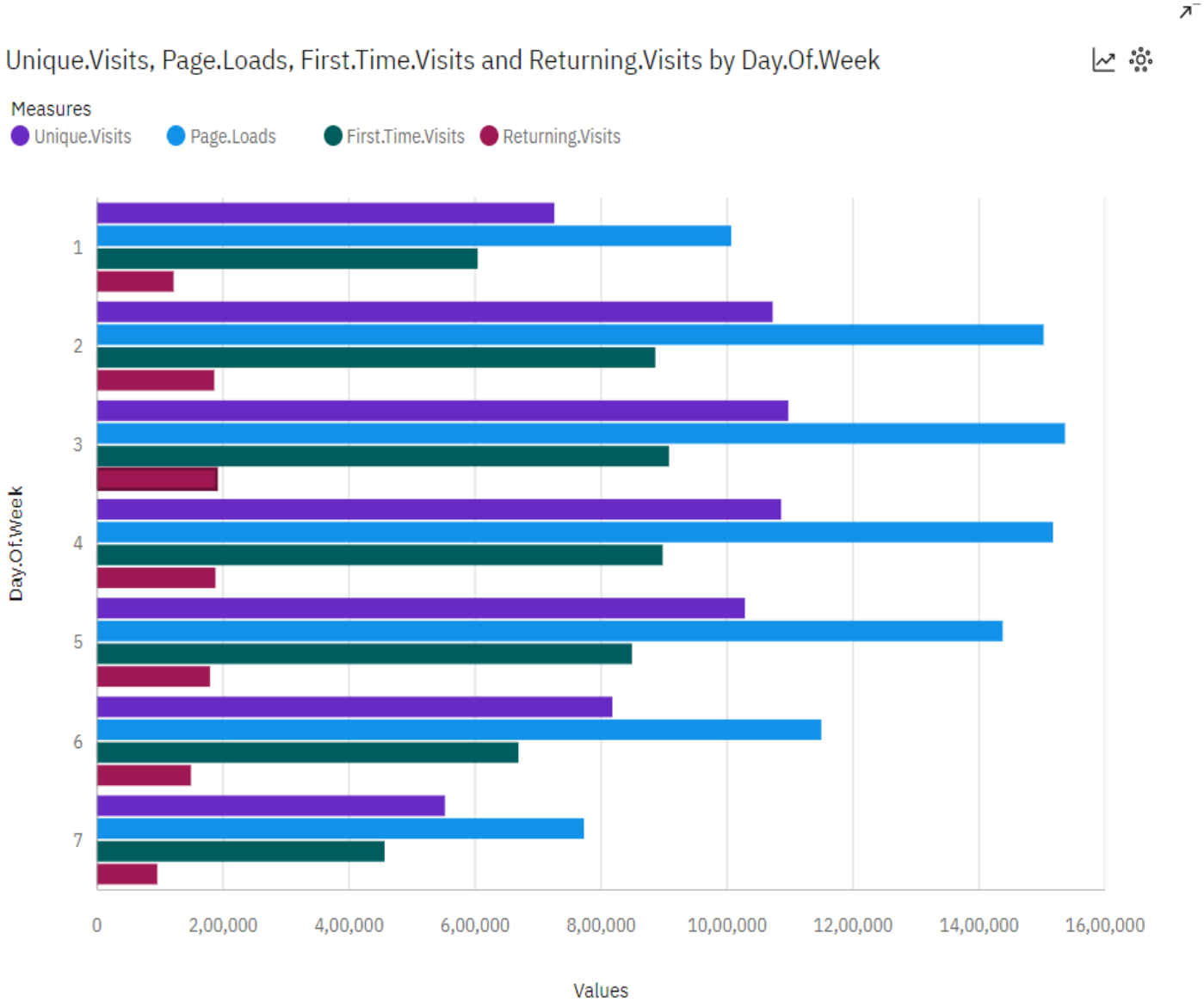
As a trusted and established leader in business intelligence and analytics, IBM Cognos equips us with a suite of advanced tools for data management and visualization. We harness the full potential of this platform to ensure accuracy and efficiency.

In the world of data, visualization is the key to unlocking meaningful insights. With IBM Cognos, we'll craft compelling data visualizations, interactive dashboards, and reports, enabling us to communicate complex website traffic patterns and trends effectively.

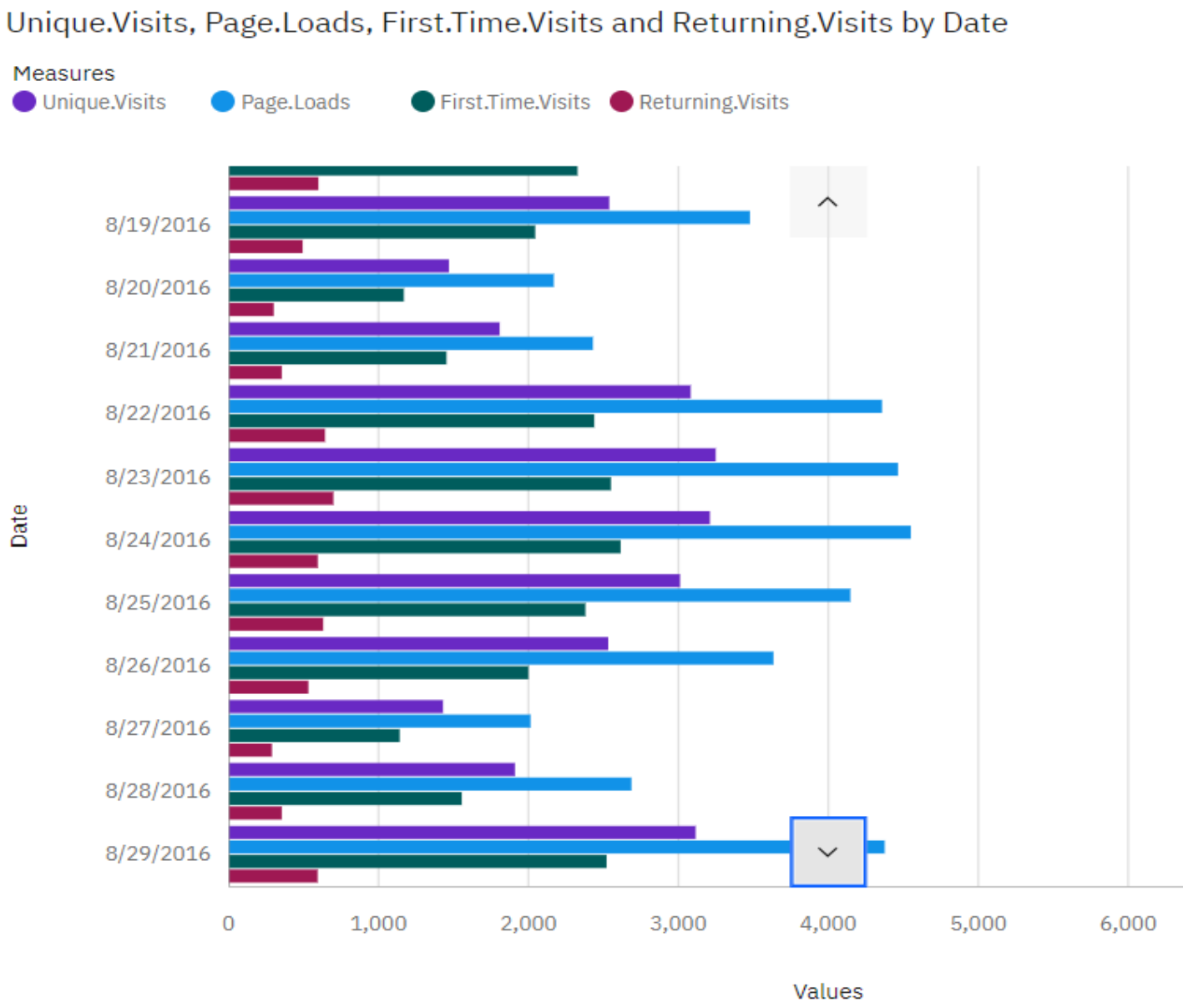
Visualization for Unique Visits, Page Loads, First time visit, Returning visit by Day:



Visualization for Unique Visits, Page Loads, First time visit, Returning visit by Day of Week



Visualization for Unique Visits, Page Loads, First time visit, Returning visit by Date:



Objectives:

- The primary objective is to ensure the integrity and accuracy of our website traffic data.
- We aim to clean, transform, and prepare the data for analysis within IBM Cognos, eliminating inconsistencies and errors to establish a reliable foundation for decision-making.
- We seek to employ the powerful visualization capabilities of IBM Cognos to create intuitive and informative data representations.
- The objective is to present website traffic trends, user behaviours, and performance metrics in a visually compelling and accessible manner.

Cleaning data:

The data cleaning phase of the Website Traffic Analysis Project focused on enhancing the quality and reliability of our raw web traffic data. This process involved:

- Identifying and rectifying missing values, duplicates, and inconsistencies.
- Standardizing data formats, such as date and URL formats.
- Ensuring that the dataset is now prepared for in-depth analysis and visualization.

Program:

```
import pandas as pd

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

df.drop_duplicates()

#in this section it deletes the duplicate data in the dataset
```

	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
...	...	...	...	...	...	...	...	...
2162	2163	Saturday	7	8-15-2020	2,221	1,696	1,373	323
2163	2164	Sunday	1	8-16-2020	2,724	2,037	1,686	351
2164	2165	Monday	2	8-17-2020	3,456	2,638	2,181	457
2165	2166	Tuesday	3	8-18-2020	3,581	2,683	2,184	499
2166	2167	Wednesday	4	8-19-2020	2,064	1,564	1,297	267

df.isnull().sum()

```
Row                 0
Day                 0
Day.Of.Week         0
Date                0
Page.Loads          0
Unique.Visits       0
First.Time.Visits   0
Returning.Visits    0
dtype: int64
```

#in this section it replaces the ‘/’ as ‘-‘ and makes standard form of all date in #the dataset

df['Date'] = df['Date'].str.replace('/', '-')

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

df['Date'] = df['Date'].dt.strftime('%m-%d-%Y')

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	09-14-2014	2,146	1,582	1,430	152
1	2	Monday	2	09-15-2014	3,621	2,528	2,297	231
2	3	Tuesday	3	09-16-2014	3,698	2,630	2,352	278
3	4	Wednesday	4	09-17-2014	3,667	2,614	2,327	287
4	5	Thursday	5	09-18-2014	3,316	2,366	2,130	236
...	...	...	...	...	...	...	...	...
2162	2163	Saturday	7	08-15-2020	2,221	1,696	1,373	323
2163	2164	Sunday	1	08-16-2020	2,724	2,037	1,686	351
2164	2165	Monday	2	08-17-2020	3,456	2,638	2,181	457
2165	2166	Tuesday	3	08-18-2020	3,581	2,683	2,184	499
2166	2167	Wednesday	4	08-19-2020	2,064	1,564	1,297	267

2167 rows × 8 columns

Conclusion:

The meticulous data cleaning phase, combined with the capabilities of IBM Cognos, has guaranteed the reliability and consistency of our web traffic dataset. This foundation is essential for generating accurate insights. The use of IBM Cognos for data visualization has enabled us to craft meaningful, visually engaging representations of website traffic trends and user behaviours.