

WEBSITE TRAFFIC ANALYSIS USING IBM COGNOS

Introduction:

Our project embarks on the crucial task of loading and preprocessing the dataset to commence the website traffic analysis utilizing IBM Cognos for visualization.

This phase is foundational for extracting meaningful insights from the website traffic data.

Objectives:

1. Dataset Loading:

We've obtained the website traffic dataset from the designated source, encompassing various metrics vital for analysis.

2. Data Preprocessing:

Our primary goal is to ensure the quality and reliability of the dataset. Key preprocessing steps include:

- **Handling Missing Values:** Any gaps in the dataset have been carefully addressed to prevent distortions in the analysis.
- **Data Cleaning:** Identification and treatment of outliers and anomalies to maintain data integrity.
- **Data Transformation:** Necessary transformations have been applied to align the data with the requirements of the IBM Cognos visualization tool.

3. Dataset Loading:

The dataset, sourced from a specified location, incorporates metrics such as page views, unique visitors, and session durations—essential components for a comprehensive website traffic analysis.

4. Data Preprocessing:

Handling Missing Values: Missing data points were addressed through imputation methods or removal, depending on the nature of the data.

5. Data Cleaning:

Outliers were identified using statistical techniques and addressed to prevent skewing the analysis. Anomalies, if any, were meticulously examined and resolved.

6. Data Transformation:

To ensure compatibility with IBM Cognos, certain transformations such as data normalization or encoding were applied.

7. Data Visualization:

IBM Cognos serves as our visualization powerhouse, allowing for dynamic and insightful representations of website traffic patterns.

The tool's capabilities empower us to explore trends, user behaviors, and other critical aspects.

PROGRAM

```
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import mode
df = pd.read_csv('daily-website-visitors.csv')
df
df.isnull().sum
```

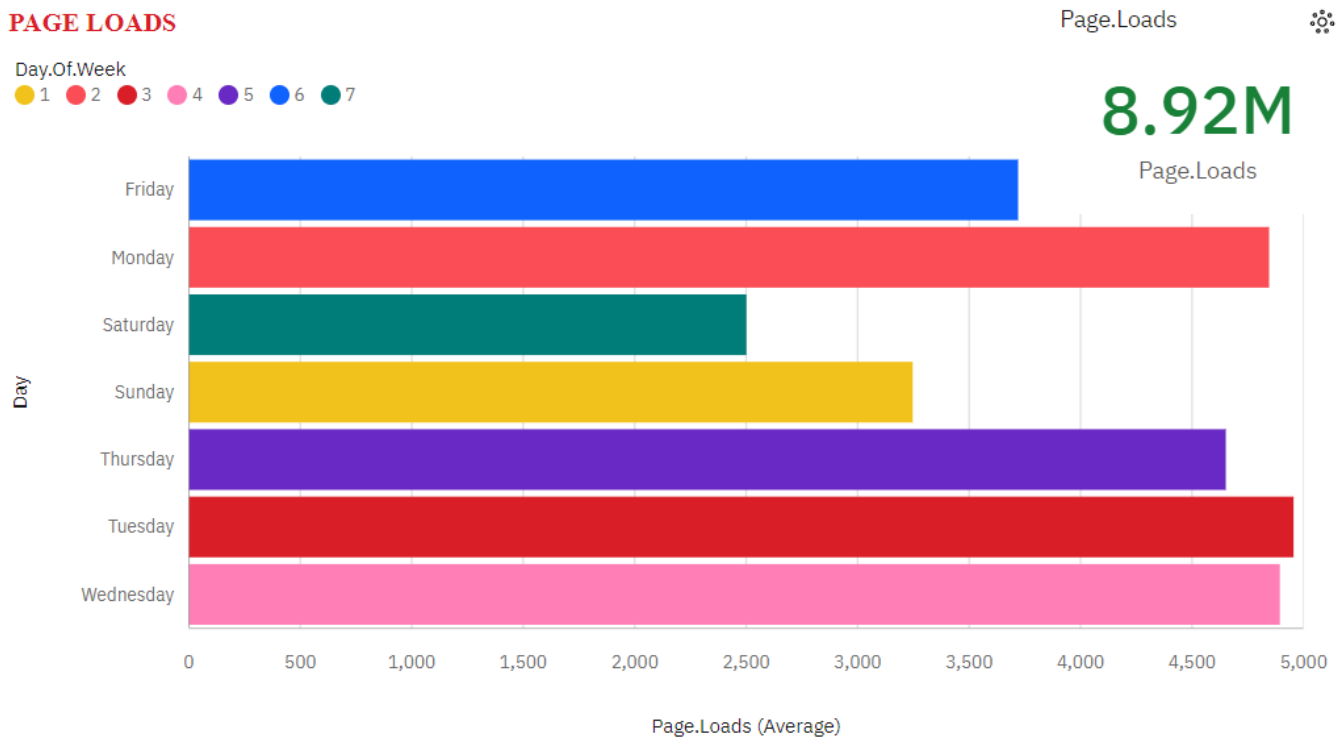
OUTPUT

```
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  object  
 5   Unique.Visits         2167 non-null  object  
 6   First.Time.Visits     2167 non-null  object  
 7   Returning.Visits      2167 non-null  object  
dtypes: int64(2), object(6)
memory usage: 135.6+ KB
```

IBM COGNOS ANALYTIC VIRTUALIZATION

1) PAGE LOAD VIRTUALIZATION

PAGE LOADS



INSIGHT

- **Day.Of.Week 3** has the highest values of both **Page.Loads** and **Returning.Visits**.
- **Day Tuesday** has the highest values of both **Page.Loads** and **Returning.Visits**.
- **Page.Loads** is unusually low when **Day** is **Saturday**.
- **Day** moderately affects **Page.Loads** (44%).
- **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).
- **Monday (14.3 %), Sunday (14.3 %), Wednesday (14.3 %), and Tuesday (14.3 %)** are the most frequently occurring categories of **Day** with a combined count of **1240** items with **Page.Loads** values (57.2 % of the total).
- Across all values of **Day** and **Day.Of.Week**, the average of **Page.Loads** is **over four thousand**.
- The average values of **Page.Loads** range from **over 2500** to **nearly five thousand**.
- **Day Saturday** has the lowest total **Page.Loads** at **nearly 773 thousand**, followed by **Sunday** at **over 1.0 million**.
- **Day Tuesday** has the highest total **Page.Loads** at **1536154.0**, followed by **Wednesday** at **1517114.0**.
- The overall number of results for **Page.Loads** is **over two thousand**.

2) UNIQUE VISITS VIRTUALIZATION

UNIQUE VISITS

Unique.Visits

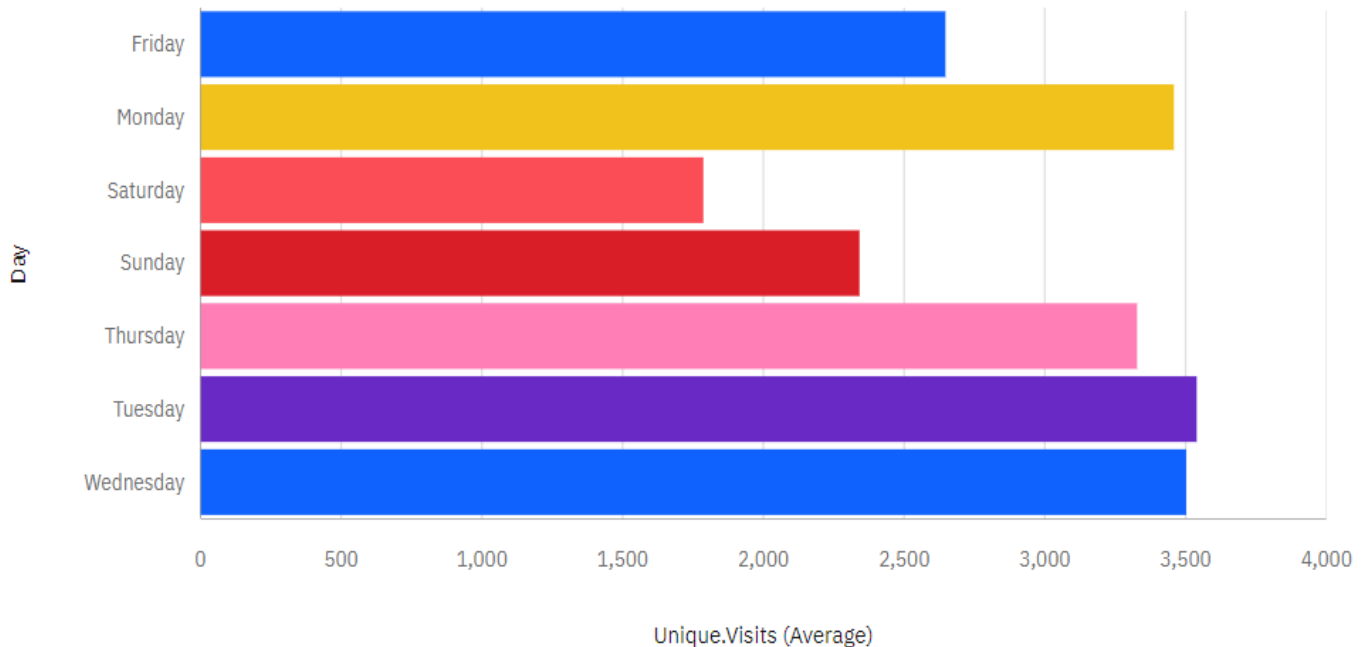


Day

Friday Monday Saturday Sunday Thursday Tuesday Wednesday

6.38M

Unique.Visits



INSIGHT

- **Day Saturday** has the lowest total **Unique.Visits** at **over 552 thousand**, followed by **Sunday** at **almost 726 thousand**.
- **Day Tuesday** has the highest total **Unique.Visits** at **1097181.0**, followed by **Wednesday** at **1085624.0**.
- Based on the current forecasting, **Unique.Visits** may reach **over 1500** by **Day Monday+1**.
- **Day** moderately affects **Unique.Visits (43%)**.
- **Monday (14.3 %)**, **Sunday (14.3 %)**, **Wednesday (14.3 %)**, and **Tuesday (14.3 %)** are the most frequently occurring categories of **Day** with a combined count of **1240** items with **Unique.Visits** values (**57.2 %** of the total).
- Over all **days** and **days**, the average of **Unique.Visits** is **nearly three thousand**.
- The average values of **Unique.Visits** range from **nearly two thousand** to **over 3500**.
- The overall number of results for **Unique.Visits** is **over two thousand**.

3)FIRST TIME VISITS VIRTUALIZATION

FIRST TIME VISITS



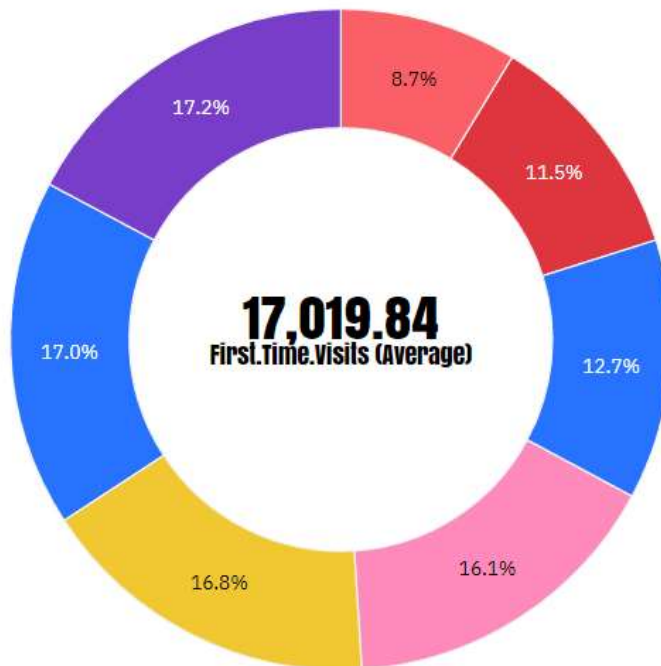
Day

- Saturday
- Sunday
- Friday
- Thursday
- Monday
- Wednesday
- Tuesday

First.Time.Visits

5.27M

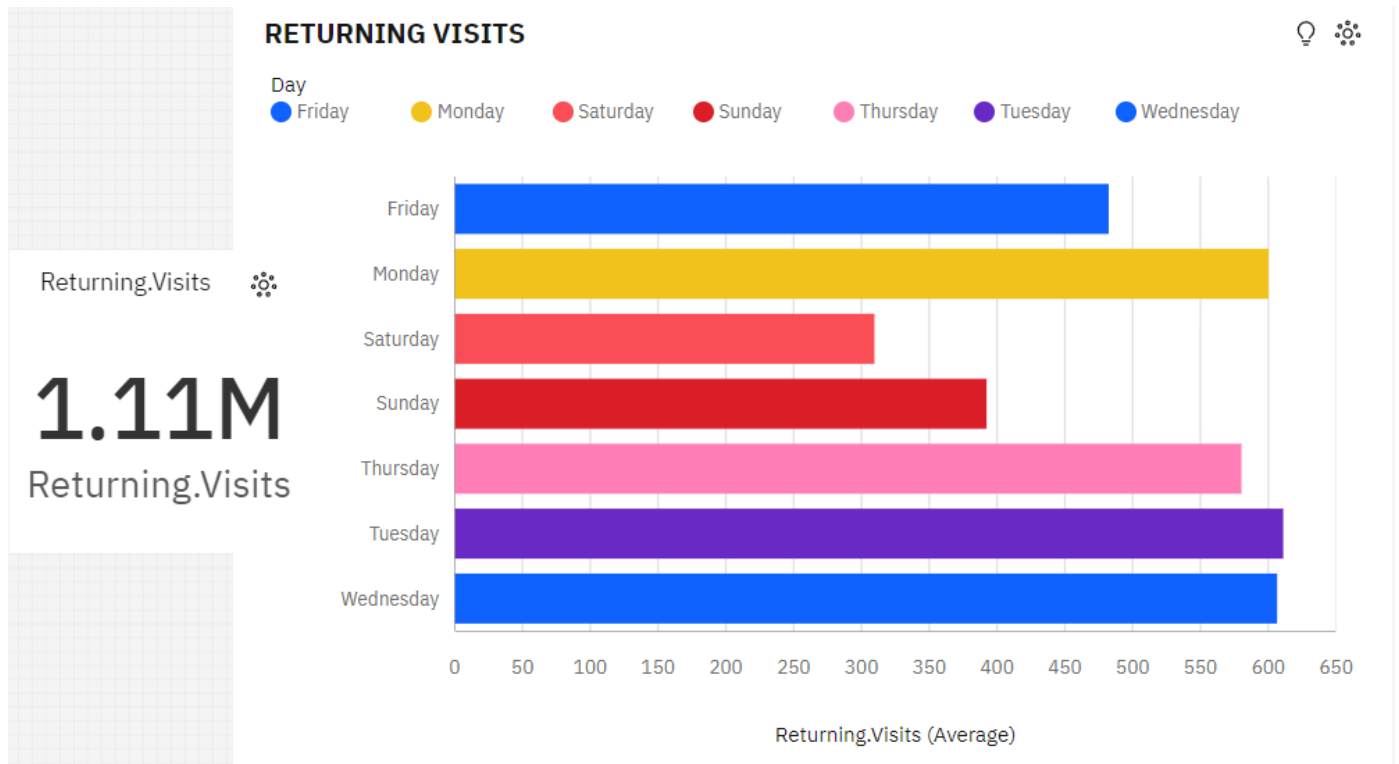
First.Time.Visits



INSIGHT

- **Day Saturday** has the lowest total **First.Time.Visits** at **over 456 thousand**, followed by **Sunday** at **over 604 thousand**.
- **Day Tuesday** has the highest total **First.Time.Visits** at **nearly 908 thousand**, followed by **Wednesday** at **almost 898 thousand**.
- Based on the current forecasting, **First.Time.Visits** may reach **over 395 thousand** by **Day Monday+1**.
- The overall number of results for **First.Time.Visits** is **over two thousand**.
- **Day** moderately affects **First.Time.Visits** (41%).
- **Monday (14.3 %)**, **Sunday (14.3 %)**, **Wednesday (14.3 %)**, and **Tuesday (14.3 %)** are the most frequently occurring categories of **Day** with a combined count of **1240** items with **First.Time.Visits** values (57.2 % of the total).
- Over all **days**, the average of **First.Time.Visits** is **almost 2500**.
- The average values of **First.Time.Visits** range from **almost 1500**, occurring when **Day** is **Saturday**, to **nearly three thousand**, when **Day** is **Tuesday**.

4)RETURNING VISITS VIRTUALIZATION

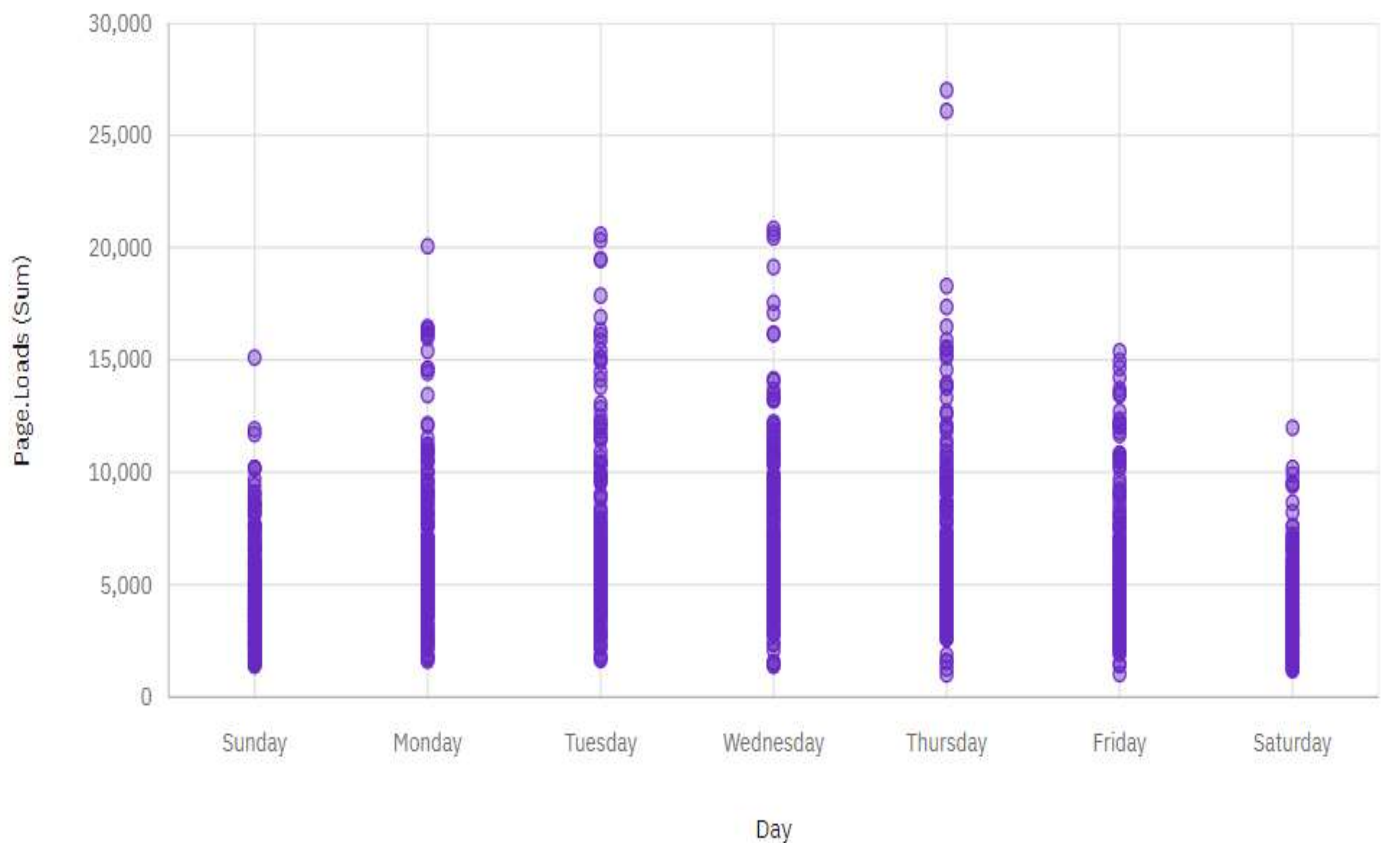


INSIGHT

- **Day Saturday** has the lowest total **Returning.Visits** at almost 96 thousand, followed by **Sunday** at almost 122 thousand.
- **Day Tuesday** has the highest total **Returning.Visits** at over 189 thousand, followed by **Wednesday** at over 188 thousand.
- Based on the current forecasting, **Returning.Visits** may reach over 87 thousand by **Day Monday+1**.
- The overall number of results for **Returning.Visits** is over two thousand.
- **Returning.Visits** is most unusual when the combinations of **Day** and **Day** are **Saturday and Saturday, Sunday and Sunday, Tuesday and Tuesday, Wednesday and Wednesday, Monday and Monday** and more.
- Based on the current forecasting, **Returning.Visits** may reach 282.6 by **Day Monday+1**.
- **Day** moderately affects **Returning.Visits** (44%).
- **Monday (14.3 %), Sunday (14.3 %), Wednesday (14.3 %), and Tuesday (14.3 %)** are the most frequently occurring categories of **Day** with a combined count of 1240 items with **Returning.Visits** values (57.2 % of the total).
- Over all **days** and **days**, the average of **Returning.Visits** is 511.8.
- The average values of **Returning.Visits** range from 309.6 to 611.1.

5) VIRTUALIZATION FOR PAGE LOAD BY DAY

Day by Page.Loads with points for Returning.Visits

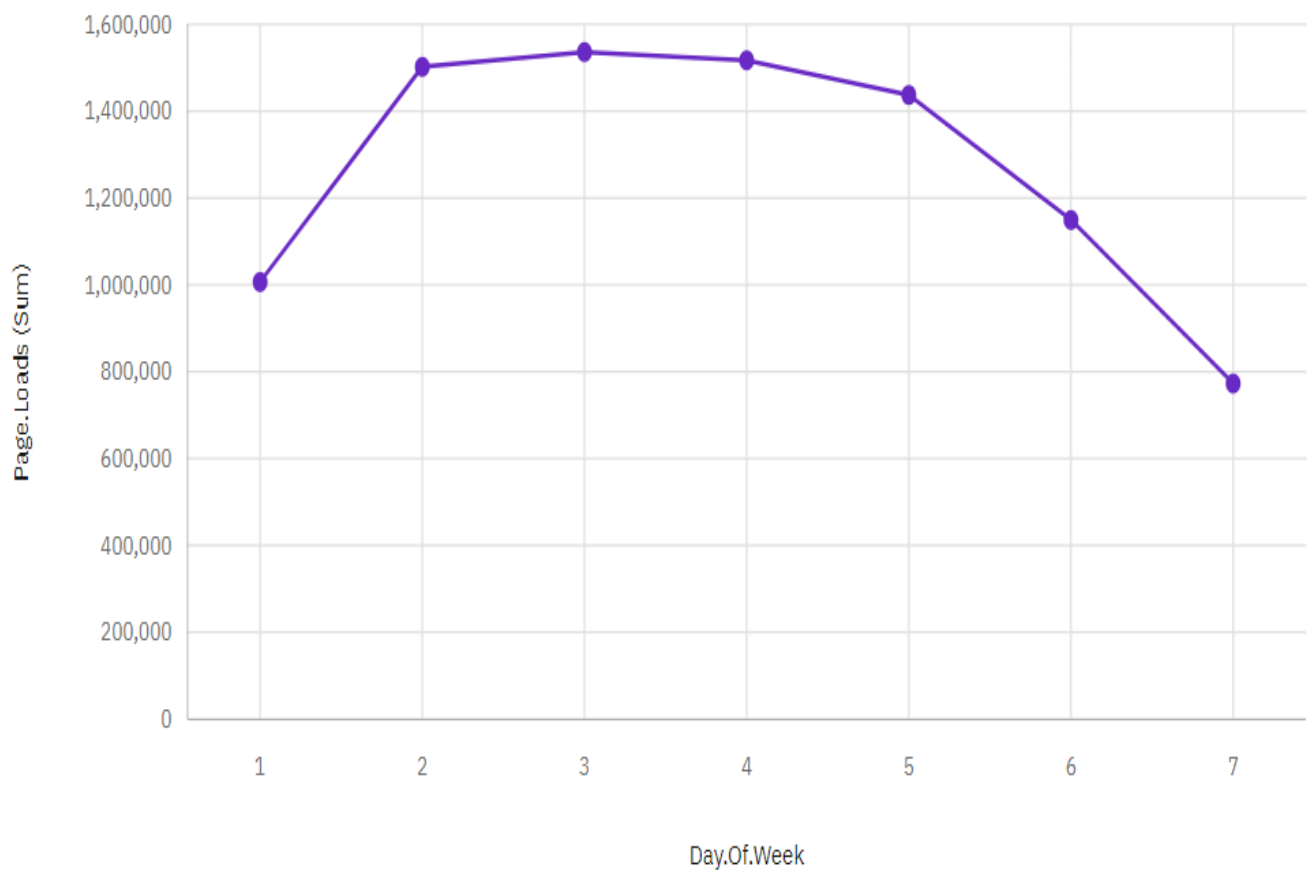


INSIGHT

- It is projected that by **Monday+1, 710** will exceed **552** in **Page.Loads** by almost **1500**.
- Based on the current forecasting, **Page.Loads** may reach **over 675 thousand** by **Day Monday+1**
- **Day Saturday** has the lowest total **Returning.Visits** at almost **96 thousand**, followed by **Sunday** at almost **122 thousand**.
- **Day Tuesday** has the highest total **Returning.Visits** at over **189 thousand**, followed by **Wednesday** at over **188 thousand**.

6) VIRTUALIZATION FOR PAGE LOAD BY DAY OF WEEKS

Page.Loads by Day.Of.Week

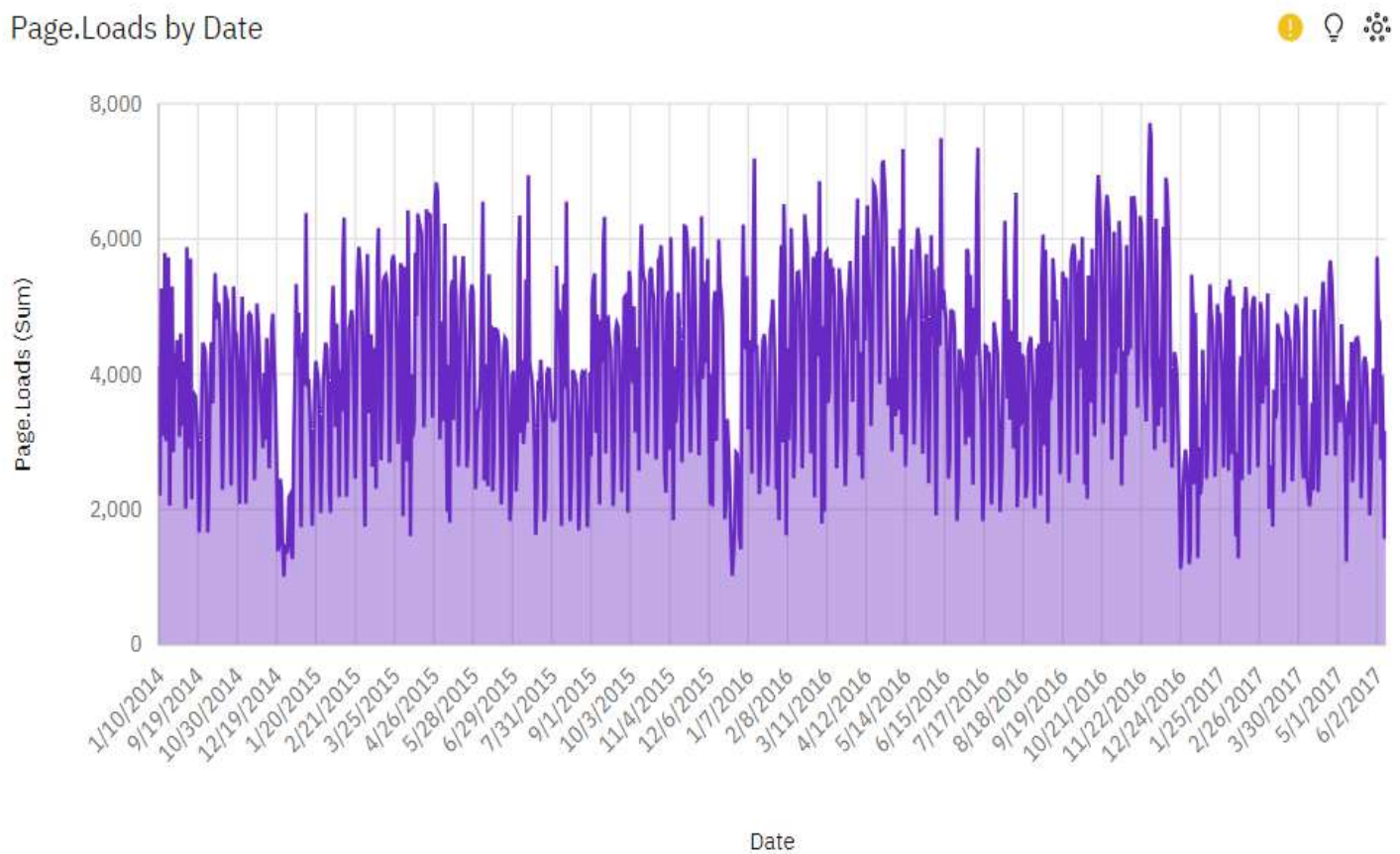


INSIGHT

- **Page.Loads** is unusually low when **Day.Of.Week** is 7.
- Based on the current forecasting, **Page.Loads** may reach **over 675 thousand** by **Day.Of.Week 9**.
- Across all values of **Day.Of.Week**, the sum of **Page.Loads** is **over 8.9 million**.
- **Page.Loads** ranges from **nearly 773 thousand**, when **Day.Of.Week** is 7, to **over 1.5 million**, when **Day.Of.Week** is 3.
- For **Page.Loads**, the most significant values of **Day.Of.Week** are 3, 4, 2, 5, and 6, whose respective **Page.Loads** values add up to **over 7.1 million**, or **80.1 %** of the total

7) VIRTUALIZATION FOR PAGE LOAD BY DATES

Page.Loads by Date



INSIGHT

- **Page.Loads** has a weak weekly trend. The smallest values typically occur on **Saturday**.
- Across all **dates**, the sum of **Page.Loads** is **over 8.9 million**.
- **Page.Loads** ranges from **over a thousand**, when **Date** is **2014-12-25**, to **nearly eight thousand**, when **Date** is **2018-04-25**.
- **Page.Loads** has most unusual values at **38** time points, the most notable of which are **2016-01-02**, **2016-03-07**, **2015-07-04**, **2019-08-06**, and **2018-10-12**.

Report Summary:

This initial phase has laid a robust foundation for our website traffic analysis project. The dataset is securely loaded and has undergone thorough preprocessing to ensure the accuracy and consistency of subsequent analyses. With IBM Cognos at our disposal, we are well-equipped to unveil intricate patterns and trends within the website traffic data.