DATA ANALYTICS WITH COGNOS

WEBSITE TRAFFIC ANALYSIS – PHASE 3

INTRODUCTION:

In Phase 3, the given dataset of the website traffic has been loaded to the Cognos Analytics tool. Then, the visualization of the data has been done by creating dashboard and data module using the Cognos tool.

DATASET LINK: https://www.kaggle.com/datasets/bobnau/daily-website-visitors

ABOUT DATASET:

The given data set has already been cleaned. Hence, no preprocessing is required.

➢ **Row:** An identifier or index for each row in the dataset used to identify each record uniquely.

➢ **Day:** Represents the day of the data record.

➢ **Day_of_week:** Indicates the day of the week represented as a numeric value (e.g., 1 for Sunday, 2

for Monday, etc.).

➢ **Date:** Denotes date of the entry where the entire data ranges over a period of 5 years.

➢ **Page Loads:** Number of page loads on the website for a given day. It's a count of how many times pages on the website were loaded or accessed.

➢ **Unique Visits:** Count of unique visitors to the website on a given day. It measures the number of distinct individuals whose IP addresses haven't been hits on any page in over 6 hours. A visit is classified as "unique" if a hit from same IP address has not come within the last 6 hours.

➢ **First-time Visits:** Indicates the number of first-time visitors to the website on a particular day. First-time visits typically represent new visitors who haven't visited the website before.

➢ **Returning Visits:** Count of returning visitors to the website on a specific day. Returning visits are from individuals who have visited the website before and are returning for subsequent visits.

ANALYSIS PURPOSE:

• **Analyze Traffic Patterns:** Allows to understand how website traffic varies on a daily basis, with insights into daily, weekly, and possibly academic calendar-related fluctuations.

• **Forecasting:** The data can be used to build forecasting models to predict future website traffic.

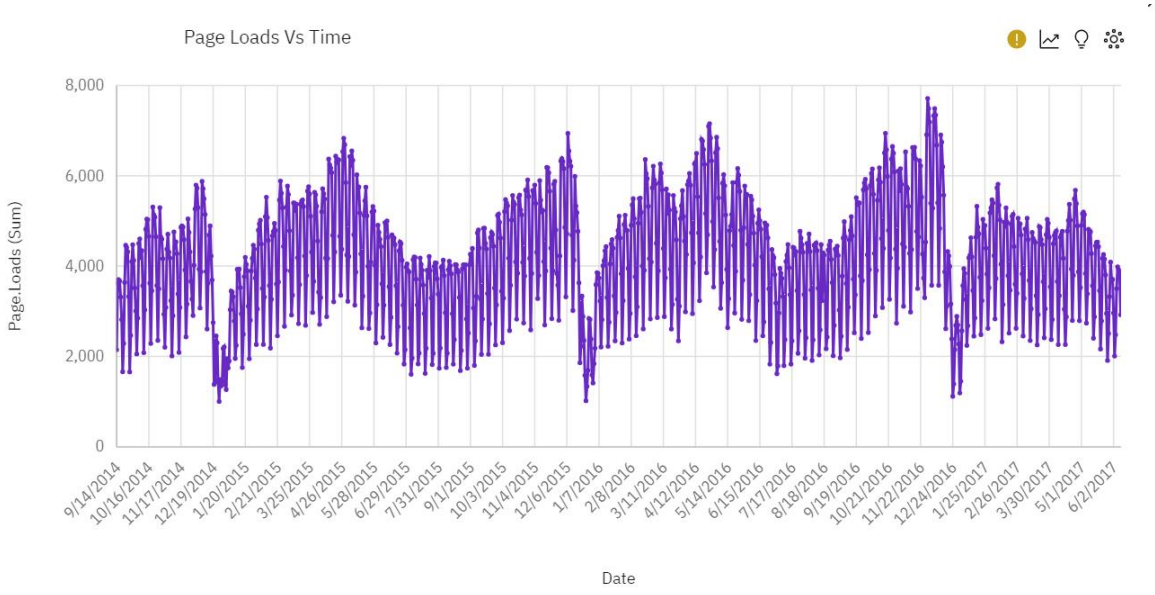
For example, you can develop models for 1-day-ahead forecasts, 7-day-ahead forecasts, and forecasts for the entire next week.

• **Visitor Segmentation:** By distinguishing between unique, first-time, and returning visitors, you can gain insights into the behavior of different types of website users.

• **Identify Trends and Seasonal Effects:** These data sets help in identifying trends and seasonality in website traffic, which can inform content strategies, marketing efforts, and resource allocation.

VISUALIZATION USING COGNOS:

1. **Page Loads Vs Time:**

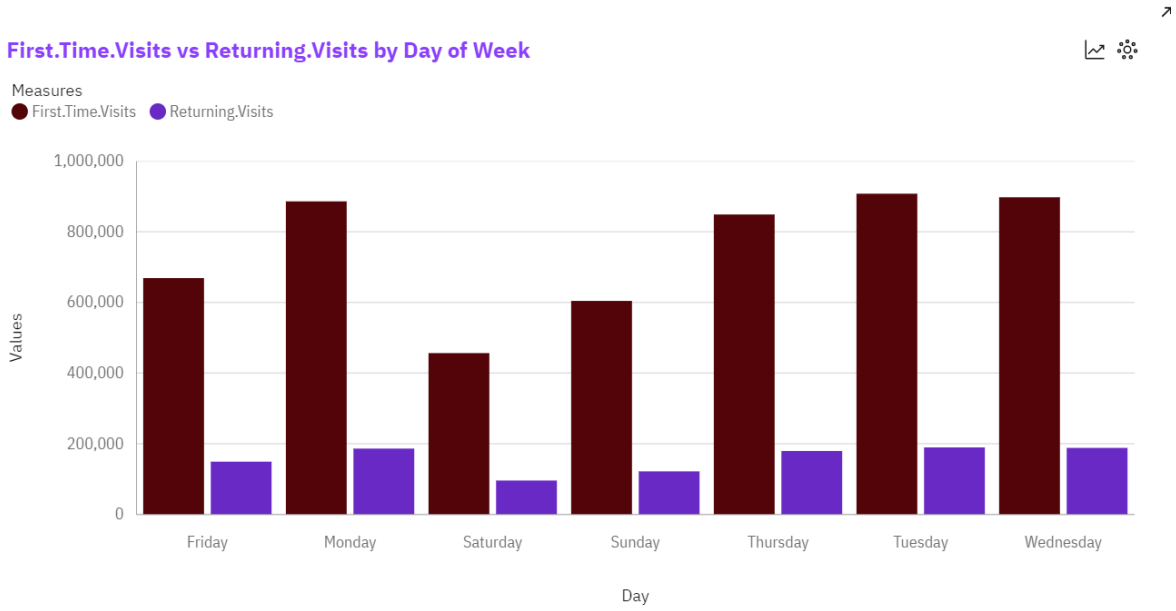


Insights:

- 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.

- The value of Page.Loads at the last observed time point 2020-08-19 is unusual. This may indicate incomplete data or a recent event that might require investigation.
- Page.Loads has most unusual values at 24 time points, the most notable of which are 2020-03-16, 2019-01-03, 2020-08-19, 2016-11-24, and 2019-01-02.
- From 2016-04-17 to 2016-04-18, Page.Loads increased by 62%.
- 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.

2. FIRST.TIME.VISITS VS RETURNING VISITS BY DAY OF WEEK

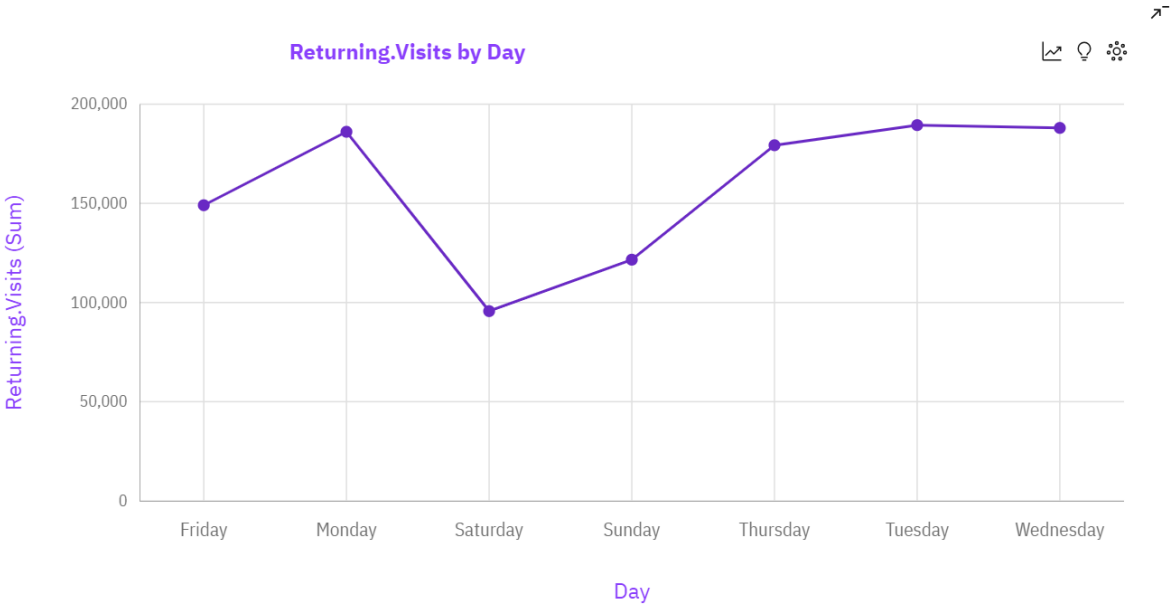


INSIGHTS:

- First.Time.Visits is unusually low when Day is Saturday.
- Based on the current forecasting, First.Time.Visits may reach over 395 thousand by Day Monday+1.
- 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).
- 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, the average of First.Time.Visits is almost 2500.
- Over all days, the average of Returning.Visits is 511.8.
- The total number of results for First.Time.Visits, across all days, is over two thousand.
- The total number of results for Returning.Visits, across all days, is over two thousand.
- First.Time.Visits ranges from over 456 thousand, when Day is Saturday, to nearly 908 thousand, when Day is Tuesday.

- Returning.Visits ranges from almost 96 thousand, when Day is Saturday, to over 189 thousand, when Day is Tuesday.

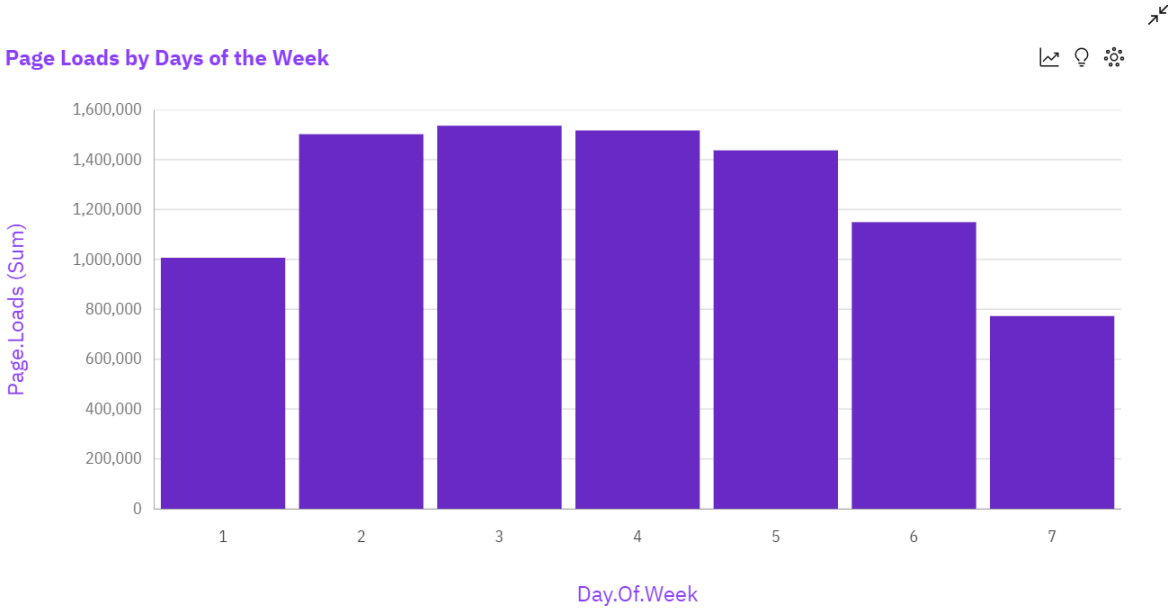
3. RETURNING.VISITS BY DAY



INSIGHTS:

- Returning.Visits is unusually low when Day is Saturday.
- Based on the current forecasting, Returning.Visits may reach over 87 thousand by Day Monday+1.
- 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.
- 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.

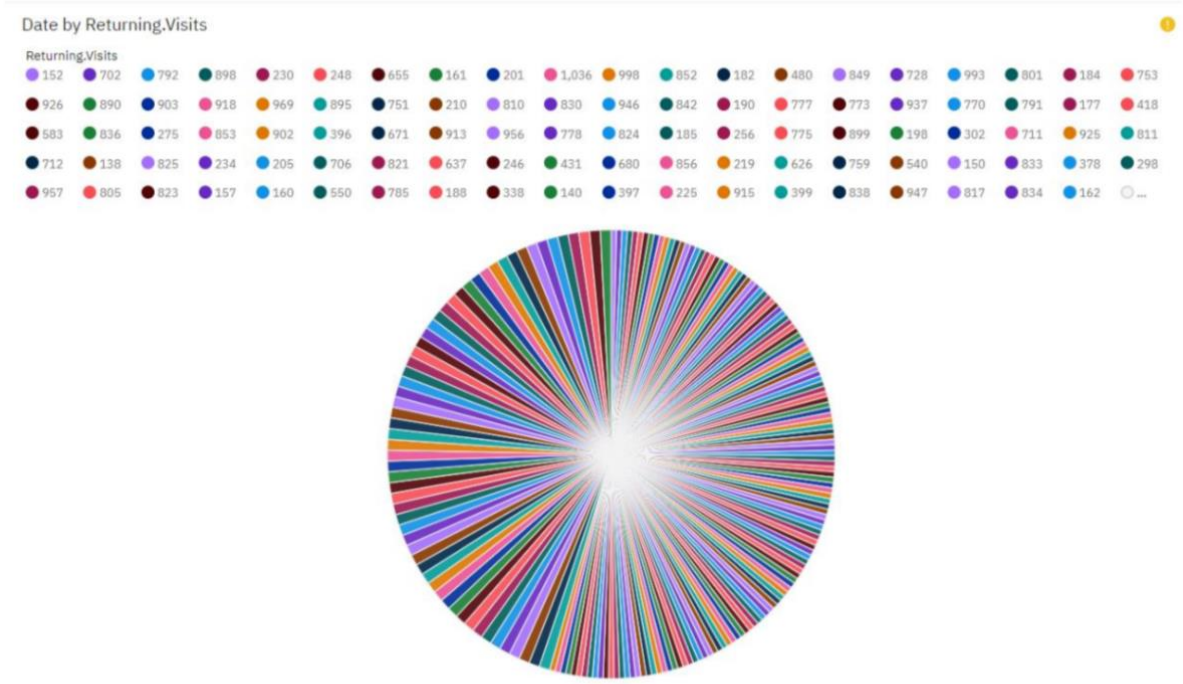
4. Page Loads by Days of the Week



INSIGHTS:

- 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.

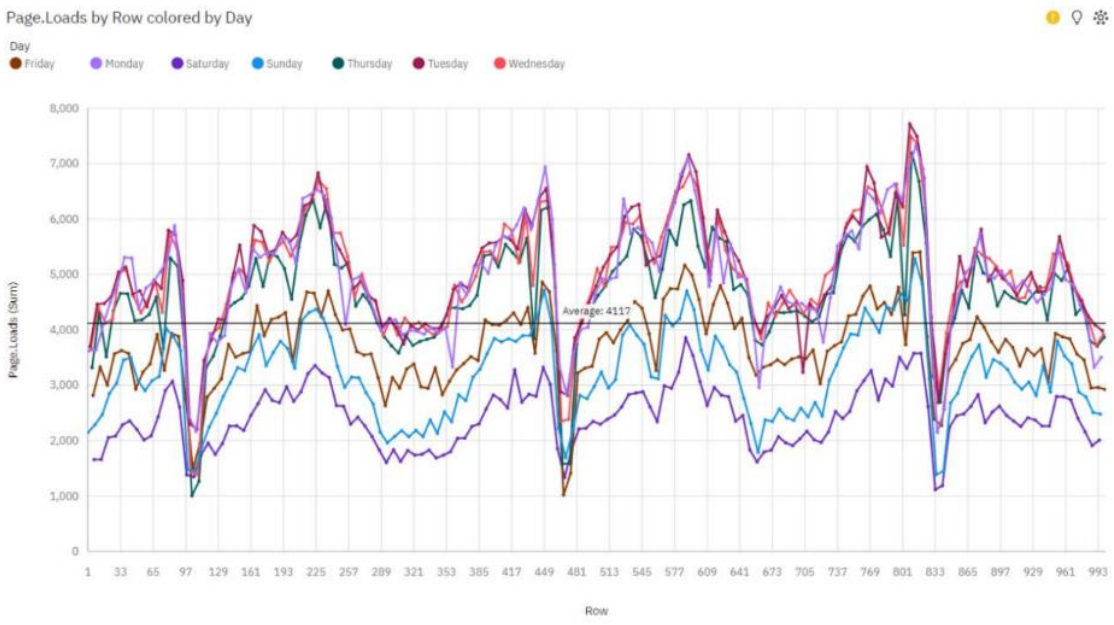
5. DATE BY RETURNING.VISITS



INSIGHTS:

• The total number of results for Date, across all Returning Visits, is over two thousand.

6. PAGE LOADS BY ROW :



INSIGHTS:

• Row 1320 has the highest total Page Loads due to Day Wednesday.

• Page Loads is unusually low when Day is Saturday.

• Based on the current forecasting, Page Loads may reach over 675 thousand by Day Monday+1.

• Day Tuesday has the highest Page Loads at over 1.5 million, out of which Row 808 contributed the most at over 7500.

• Across all rows and days, the sum of Page Loads is over 8.9 million.

• The summed values of Page Loads range from over a thousand to nearly eight thousand.

• 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.

• For Page Loads, the most significant value of Row is 1320, whose respective Page Loads values add up to nearly eight thousand, or 0.1 % of the total.

CONCLUSION:

Data visualisation for the time series analysis were obtained using IBM Cognos tools and various analytics insights were inferred. By collecting and analyzing data related to how users interact with a website, this would help stakeholders gain valuable insights and make informed decisions to improve visitors online presence.