WEBSITE TRAFFIC ANALYSIS

PROJECT

PROJECT OBJECTIVE:

The objective of this website traffic analysis project is to provide website owners with valuable insights into user behaviour and engagement on their website.

By analysing website traffic data, the project aims to help website owners make data-driven decisions to improve user experience, increase engagement, and achieve their business goals.

DESIGN THINKING PROCESS:

1. Empathize: Understand the website owner's goals, target audience, and current challenges. Identify the key questions they want to answer through traffic analysis.

2. Define: Clearly define the objectives of the analysis, such as understanding user behaviour, identifying popular content, and optimizing conversion paths.

3. Ideate: Brainstorm data collection methods, tools, and visualization techniques. Consider potential insights that could be derived from the data.

4. Prototype: Create a plan for data collection, visualization, and analysis. Decide on the tools and technologies to be used, including IBM Cognos for data visualization and Python for data processing.

5. Test: Implement the data collection and analysis plan on a sample dataset to ensure it provides meaningful insights.

6. Implement: Execute the plan on the full dataset and continuously monitor the results.

DEVELOPMENT PHASES:

1. Data Collection:

- Utilize web analytics tools (e.g., Google Analytics) to gather data on website traffic, including page views, bounce rates, user demographics, and source of traffic.
- Collect user interaction data, such as click-through rates, scroll depth, and session duration, using event tracking.

2. Data Preprocessing:

- Clean and prepare the data, handling missing values, outliers, and data transformations if necessary.

3. Data Visualization with IBM Cognos:

- Create interactive dashboards in IBM Cognos to visualize website traffic data.
- Generate visualizations like line charts, bar graphs, and heatmaps to represent user behavior, traffic sources, and content popularity.

4. Python Code Integration:

- Develop Python scripts to automate data extraction and preprocessing from web analytics tools.
- Integrate Python scripts with IBM Cognos for seamless data updates and analysis.

5. Analysis Objectives:

- Identify high-traffic pages and user entry points.
- Analyse user flow and discover drop-off points in conversion paths.
- Determine the effectiveness of marketing campaigns and referral sources.
- Understand user demographics and tailor content accordingly.
- Monitor the impact of website changes on user engagement.

Improving User Experience:

Insights from the analysis can help website owners:

- Optimize content placement and layout to keep users engaged.
- Improve the user journey by addressing drop-off points.
- Tailor content to target audiences based on demographic data.
- Refine marketing strategies to focus on effective channels.

- Make data-driven decisions for website improvements.

INSTRUCTION TO LOADING AND PROCESSING THE DATASET IN IBM COGNOS ANALYTICS:

- Building a website traffic analysis in Python and visualization using IBM Cognos involves several steps, from data collection and preprocessing to model selection and evaluation.
- In this example, we'll create a simple website traffic analysis to determine traffic between return visitors and first-time visitors and page loads.
- Replicating website traffic analysis and generating visualizations using IBM Cognos and Python involves several steps.

Using IBM Cognos:

Step 1: Data Collection:

- Set up data sources for website traffic data in IBM Cognos.
- This may involve connecting to databases, data warehouses, or importing data files.

Step 2: Data Modelling:

- Create data models or data modules in IBM Cognos to define the structure of your website traffic data.
- Define dimensions, measures, and hierarchies for analysis.

Step 3: Analysis and Reports:

- Use IBM Cognos to build reports and analyses. Create queries and visualizations to explore the data.
- You can use Cognos' drag-and-drop interface to design your reports.

Step 4: Custom Calculations:

- If needed, perform custom calculations within IBM Cognos to derive metrics or KPIs specific to your website traffic analysis.

Step 5: Dashboard Creation:

- Build interactive dashboards to visualize website traffic trends and insights. Customize the layout and design of the dashboards to suit your needs.

Step 6: Automation:

- Schedule reports and analyses to run automatically at regular intervals to keep your analysis up to date.

Step 7: Export Data:

- Export the data or reports from IBM Cognos for use in Python. Common formats include CSV or Excel files.

USING PYTHON:

Step 1: Data Loading:

- Use Python, specifically libraries like Pandas, to load the data exported from Cognos.

Step 2: Data Preprocessing:

- Preprocess the data if needed.
- This might involve data cleaning, handling missing values, or transforming the data to suit your Python analysis.

Step 3: Data Analysis:

- Conduct in-depth analysis using Python libraries.
- You can perform statistical analysis, generate insights, and calculate additional metrics.

Step 4: Data Visualization:

- Create visualizations in Python using libraries like Matplotlib, Seaborn, or Plotly.
- Plot website traffic trends, user behaviour, and other insights.

Step 5: Integration:

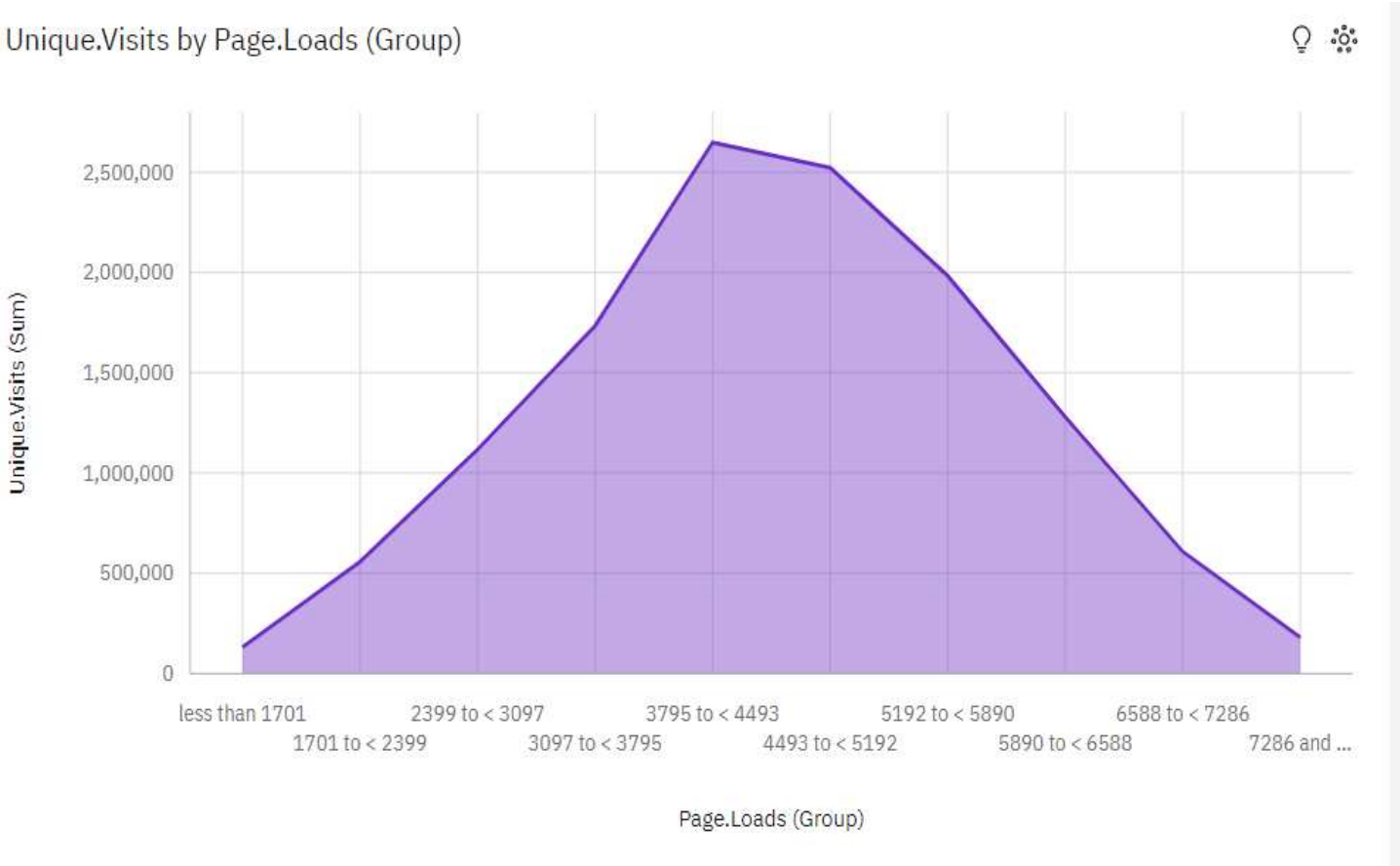
- Integrate your Python analysis with your IBM Cognos reports by sharing the results, insights, or visualizations.

Step 6: Documentation and Reporting:

- Document your analysis process and share your findings with stakeholders.
- You can use Google Collaboratory Notebooks or other reporting tools to create comprehensive reports.

HERE THE SOME EXAMPLES FOR IBM COGNOS ANALYSIS WE USE:

PAGE LOADS BY UNIQUE VISITORS IN AREA CHART

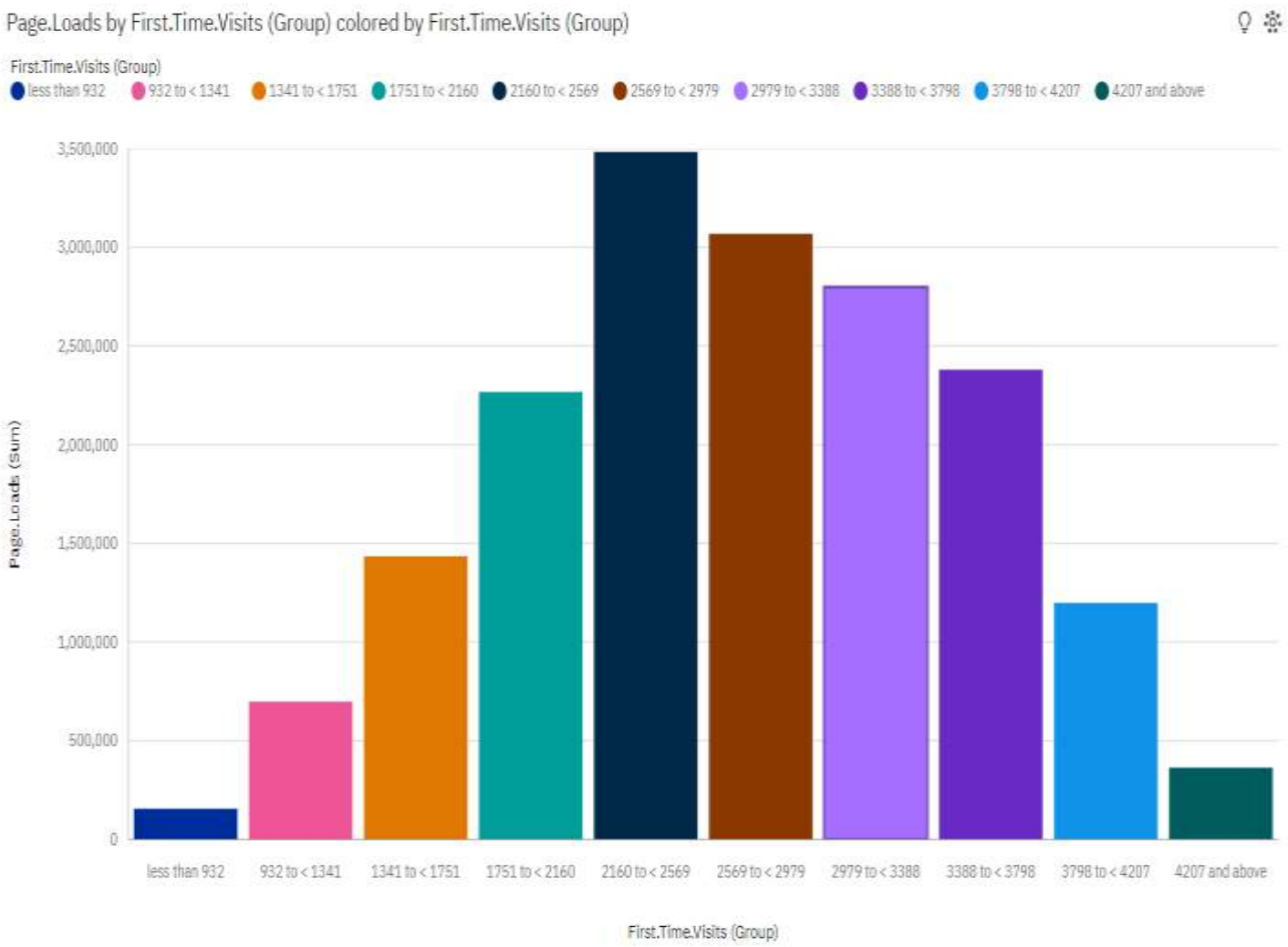


INSIGHTS:

- From **2020-05-04** to **2020-05-05**, **5890 to < 6588**'s **Unique.Visits** increased by **115%**.

- Over all values of **Page.Loads (Group)**, the sum of **Unique.Visits** is **almost thirteen million**.
- **Unique.Visits** ranges from **almost 132 thousand**, when **Page.Loads (Group)** is **less than 1701**, to **over 2.6 million**, when **Page.Loads (Group)** is **3795 to < 4493**
- For **Unique.Visits**, the most significant values of **Page.Loads (Group)** are **3795 to < 4493** and **4493 to < 5192**, whose respective **Unique.Visits** values add up to **nearly 5.2 million**, or **40.5** % of the total.

PAGE LOAD BY FIRST TIME VISITORS IN COLUMN CHART

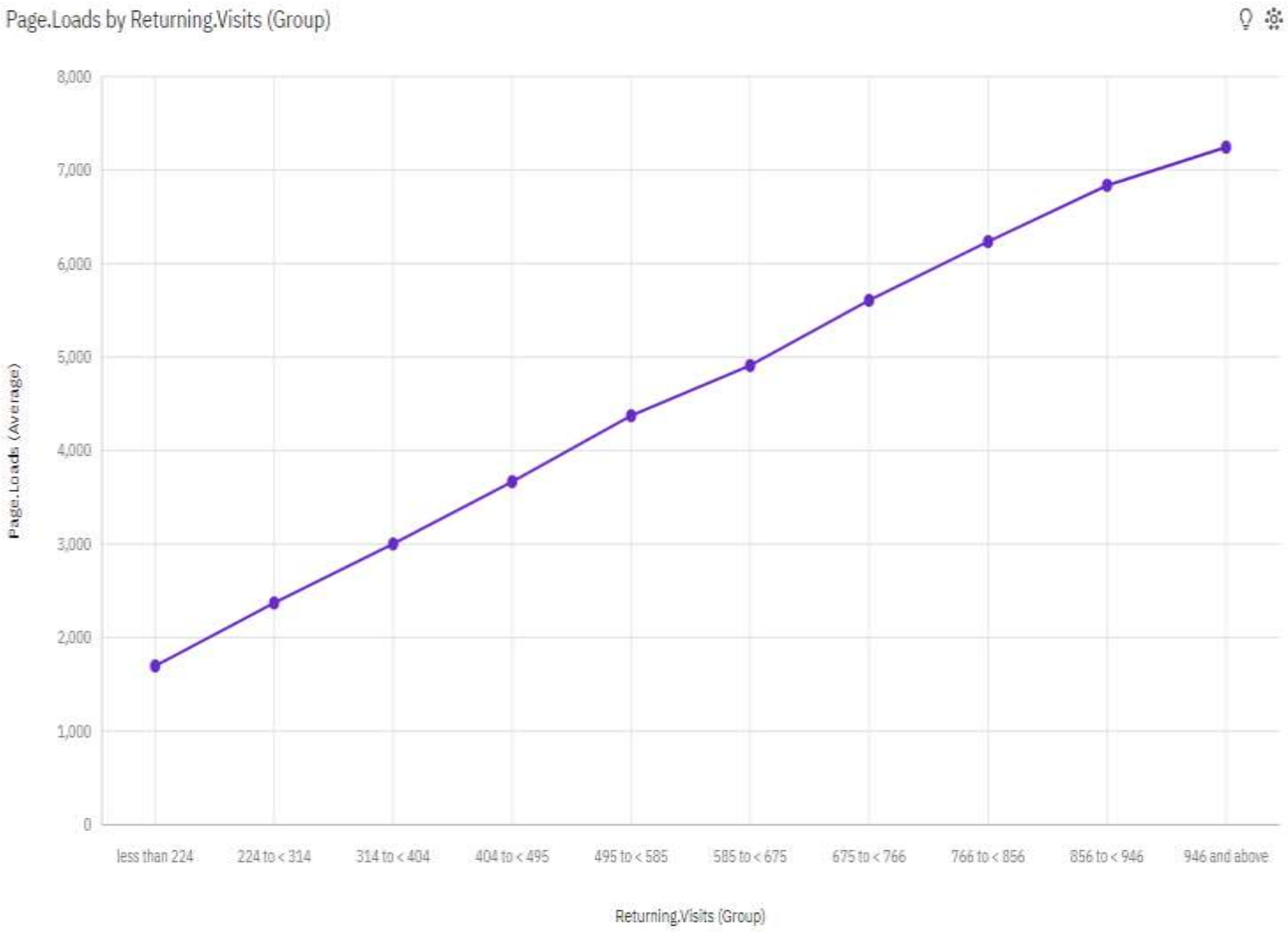


INSIGHTS:

- From **2020-05-04** to **2020-05-05**, **3798 to < 4207**'s **Page.Loads** increased by **114**%.
- Over all values of **First.Time.Visits (Group)** and **First.Time.Visits (Group)**, the sum of **Page.Loads** is **nearly eighteen million**.

- The summed values of **Page.Loads** range from **over 155 thousand** to **nearly 3.5 million**
- For **Page.Loads**, the most significant values of **First.Time.Visits (Group)** are **2160 to < 2569**, **2569 to < 2979**, **2979 to < 3388**, **3388 to < 3798**, and **1751 to < 2160**, whose respective **Page.Loads** values add up to **nearly fourteen million**, or **78.4** % of the total.

PAGE LOAD BY RETURNING VISITORS IN LINE CHART

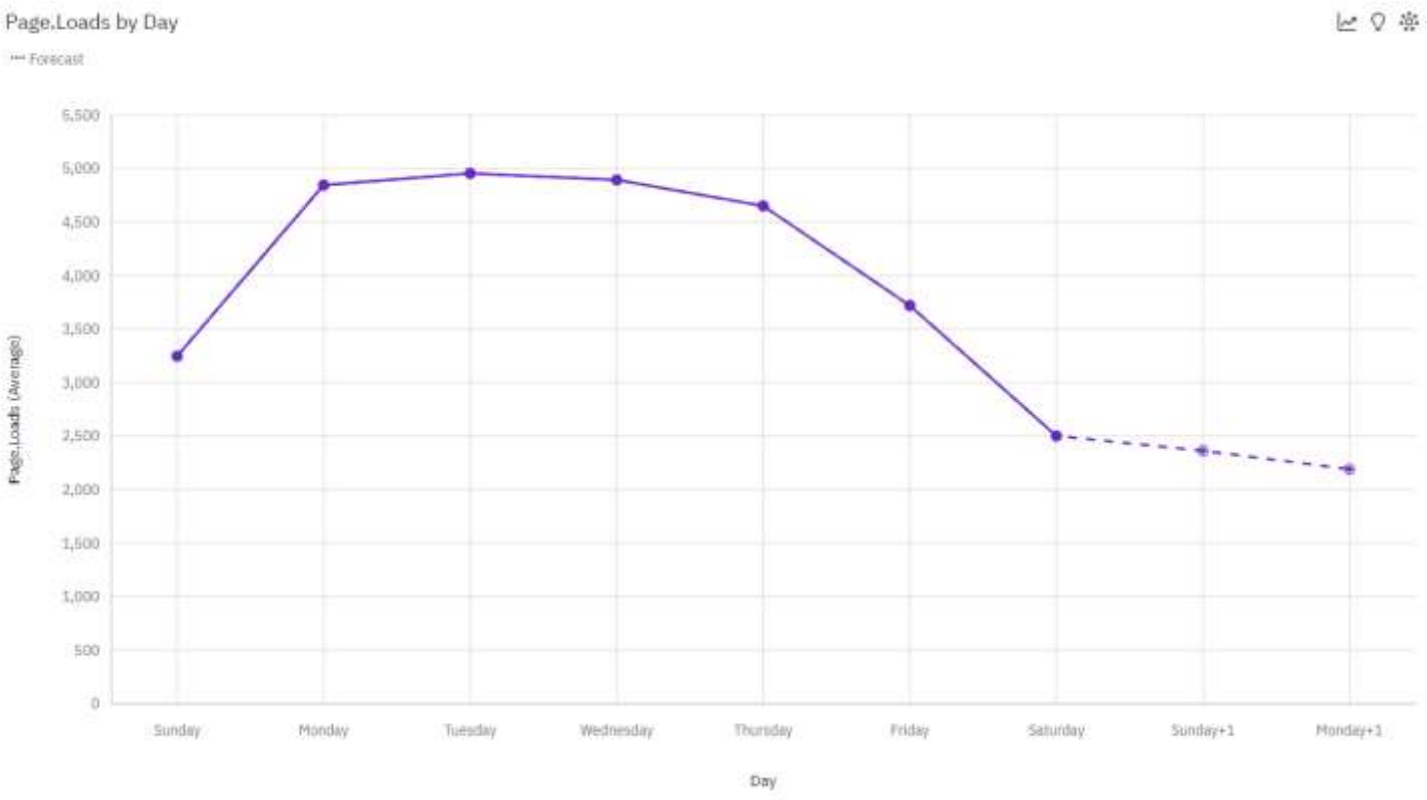


INSIGHTS:

- From **2018-10-07** to **2018-10-08**, **585 to < 675**'s **Page.Loads** increased by **48**%.
- **Returning.Visits (Group)** strongly affects **Page.Loads** (**81%**)

- **495 to < 585** (**19.9** %), **404 to < 495** (**19.1** %), **585 to < 675** (**15.9** %), **314 to < 404** (**13.5** %), and **675 to < 766** (**10.8** %) are the most frequently occurring categories of **Returning.Visits (Group)** with a combined count of **3432** items with **Page.Loads** values (**79.2** % of the total) .
- Across all values of **Returning.Visits (Group)**, the average of **Page.Loads** is **over four thousand**.
- The average values of **Page.Loads** range from **over 1500**, occurring when **Returning.Visits (Group)** is **less than 224**, to **over seven thousand**, when **Returning.Visits (Group)** is **946 and above**.

PAGE LOADS BY DAY AND IT IS FORECASTED IN LINE CHART



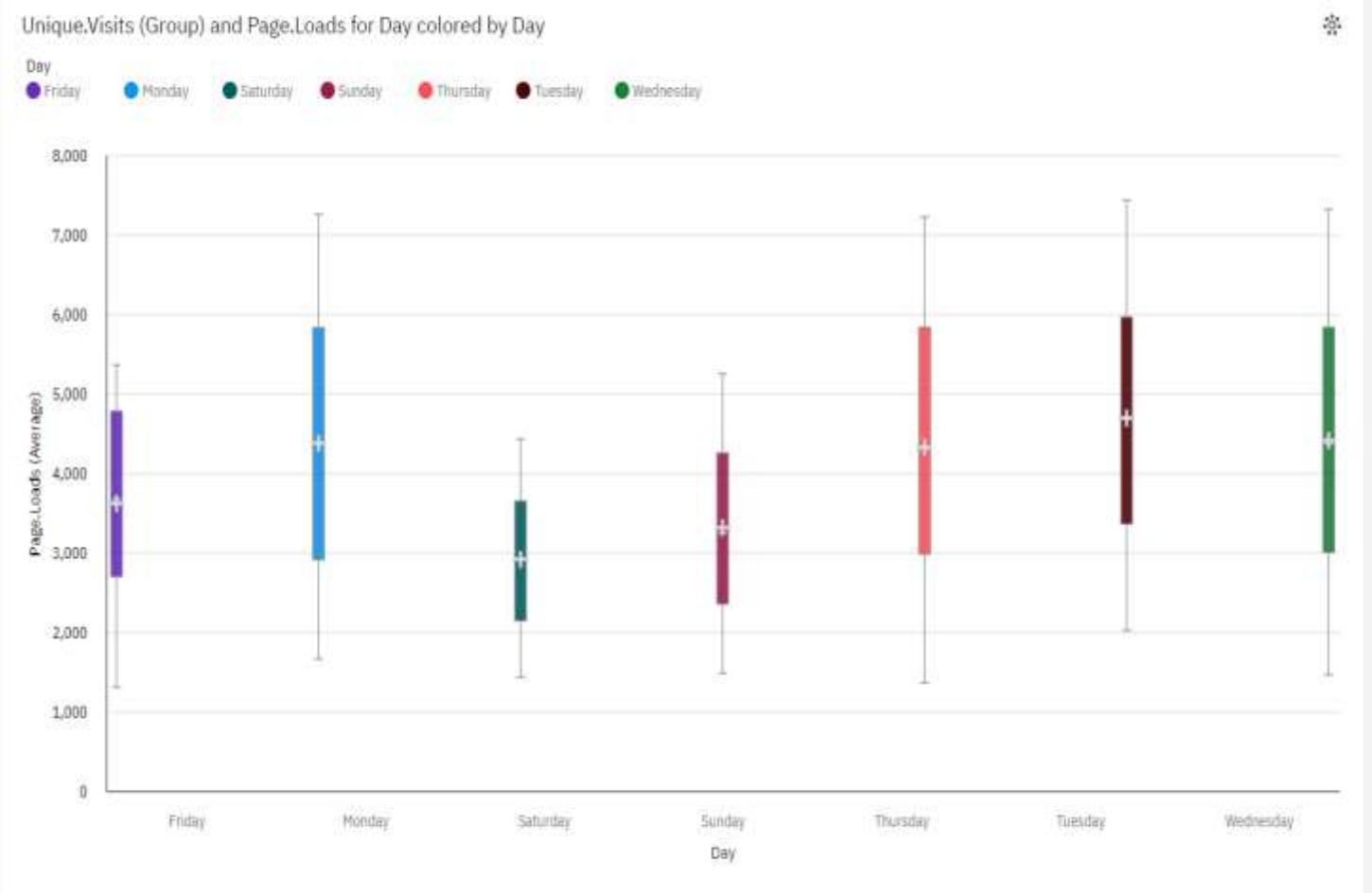
INSIGHTS:

- **Page.Loads** is unusually low when **Day** is **Saturday**.
- Based on the current forecasting, **Page.Loads** may reach **over two thousand** by **Day Monday+1**.
- **Day** moderately affects **Page.Loads** (**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 **2480** items with **Page.Loads** values (**57.2** % of the total).

- Across all **days**, the average of **Page.Loads** is **over four thousand**.
- The average values of **Page.Loads** range from **over 2500**, occurring when **Day** is **Saturday**, to **nearly five thousand**, when **Day** is **Tuesday**.

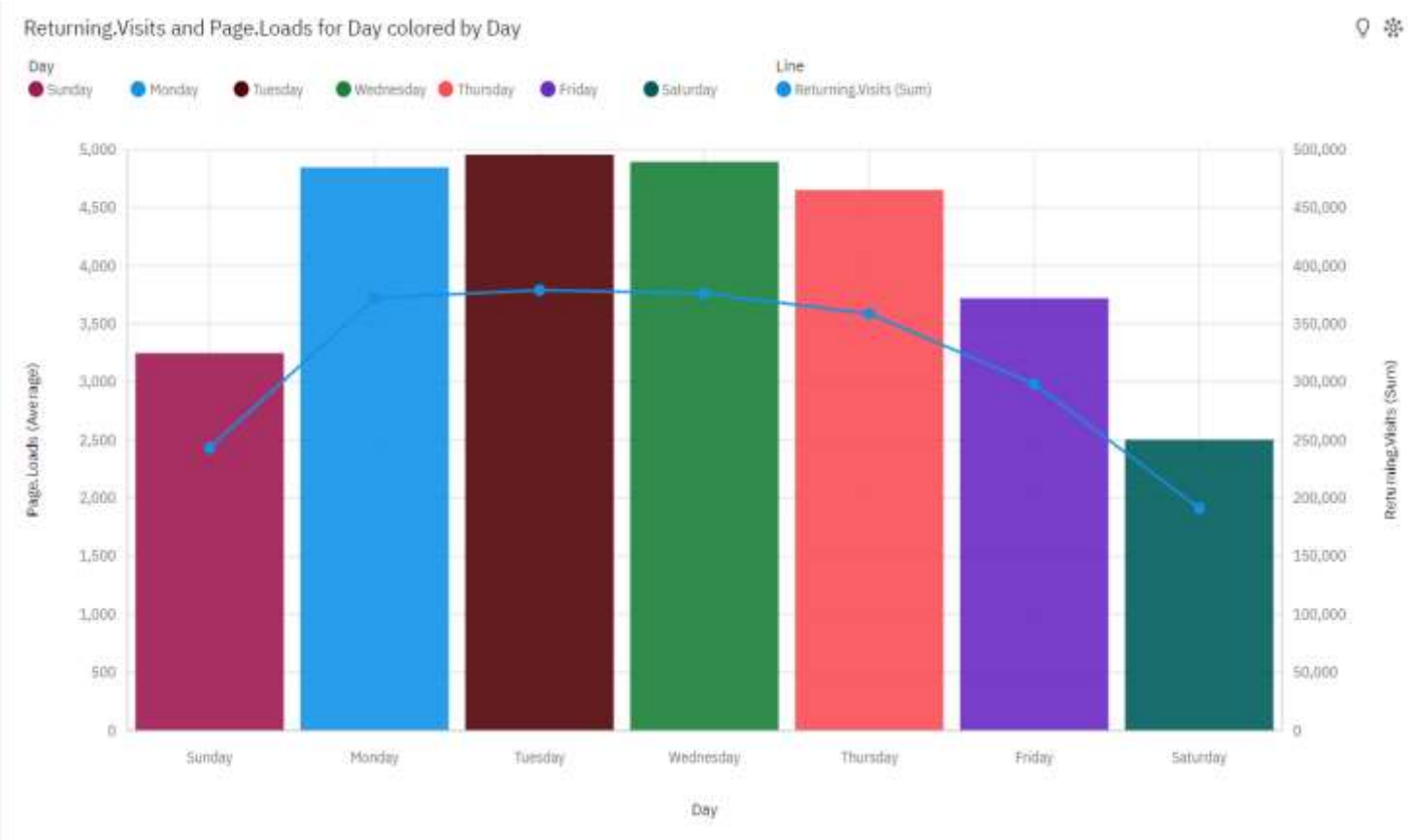
UNIQUE VISITOR AND PAGE LOADS BY DAY IN BOX PLOT



INSIGHTS:

- Based on the current forecasting, **Page.Loads** may reach **over two thousand** by **Day Monday+1**.
- Based on the current forecasting, **Page.Loads** may reach **over two thousand** by **Day Monday+1**.
- The overall number of results for **Page.Loads** is **nearly 4500**.

RETURNING VISITORS AND PAGE LOADS BY DAYS IN LINE AND COLUMN CHAT



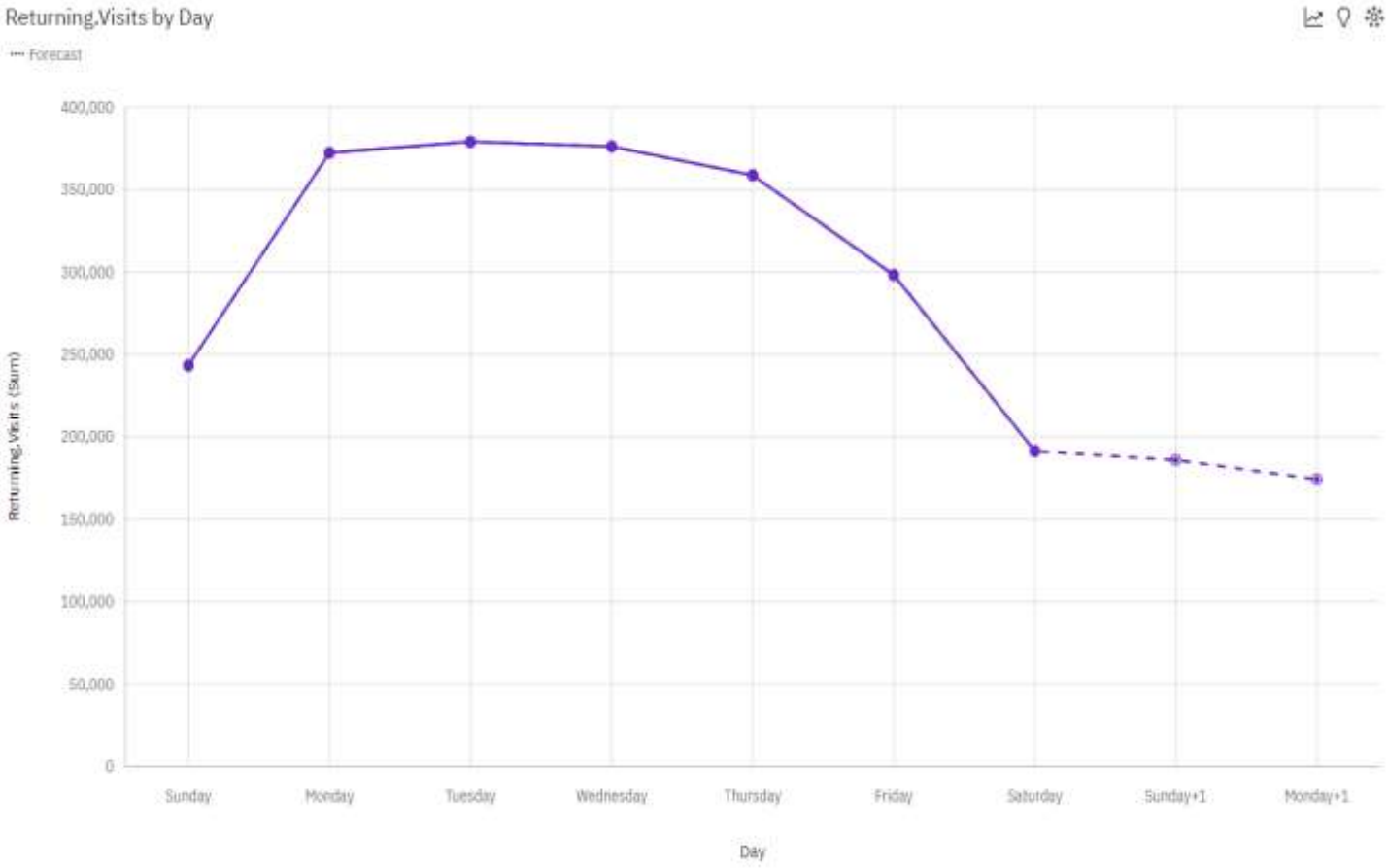
INSIGHTS:

- **Page.Loads** 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, **Page.Loads** may reach **over two thousand** by **Day Monday+1**.
- **Day** moderately affects **Page.Loads** (**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 **2480** items with **Page.Loads** values (**57.2** % of the total).
- Across all **days** and **days**, the average of **Page.Loads** is **over four thousand**.
- The average values of **Page.Loads** range from **over 2500** to **nearly five thousand**.
- Based on the current forecasting, **Returning.Visits** may reach **over**

174 Thousand by **Day Monday+1**.

- **Returning.Visits** ranges from **over 191 thousand**, when **Day** is **Saturday**, to **almost 379 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 1.8 million**, or **80.4** % of the total.

RETURNING VISITORS BY DAY AND IT IS FORECASTED IN LINE CHART:



INSIGHTS:

- **Returning.Visits** is unusually low when **Day** is **Saturday**.
- Across all **days**, the sum of **Returning.Visits** is **over 2.2 million**.
- **Returning.Visits** ranges from **over 191 thousand**, when **Day** is **Saturday**, to **almost 379 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 1.8 million**, or **80.4** % of the total.

CONCLUSION

In this project, we have outlined the objectives, design thinking process, and development phases for analysing user behaviour on a website. By collecting and visualizing data using tools like IBM Cognos and integrating Python for in-depth analysis, valuable insights have been generated. These insights enable website owners to make data-driven improvements that enhance the user experience, ultimately leading to increased user engagement and satisfaction. A user-centric approach to website optimization is crucial for achieving business objectives and ensuring long-term success.