**WEBSITE TRAFFIC ANALYSIS**

**PROJECT OBJECTIVE:**

* The intention is to analyze website traffic data analysis to obtain insights into user behavior and experience, then utilize design thinking to improve the website's user experience.
* This will entail gathering, analyzing, and visualizing data to present website owners with actionable insights.

**DESIGN THINKING PROCESS:**

**EMPATHISE:**

* Connection is the first step towards understanding the user experience.
* This step of website traffic analysis includes gathering and analyzing data from multiple sources such as Analytics, heatmaps, user comments, and surveys.
* It's about putting yourself in the shoes of the user and knowing their interactions, pain spots, and preferences as they navigate the website.

**DESCRIBE:**

* This step entails identifying particular objectives and targets for the analysis.
* Website owners must decide the goal they intend to accomplish, whether it is to reduce bounce rates, increase engagement, or improve certain user experiences.
* The definition of key performance metrics (KPIs) aids to establish an obvious path for the analysis.

**IDEATE:**

* This stage entails coming up with methods and ways for analyzing website traffic data.
* It is about thinking creatively about how to evaluate and generate useful insights from the data that has been acquired.
* This might include exploring various visualization strategies, data segmentation, or identifying trends and patterns in user behavior.

**PROTOTYPE:**

* Creating a prototype entails establishing a methodology for analysis.
* It might entail using technologies like IBM Cognos to create dashboards, reports, and visual representations.
* This stage may also include the preliminary steps of creating Python code for data processing and analysis.

**TEST :**

* Checking and improving the chosen tactics is part of testing the analytical approach.
* It entails analyzing the data, assessing the visualizations, and verifying that the insights generated are consistent with the goals established in the Define phase.
* Testing also entails obtaining input and iteratively improving the analytical approach.

**DEVELOPMENT PHASES:**

**DATA ANALYSIS OBJECTIVES:**

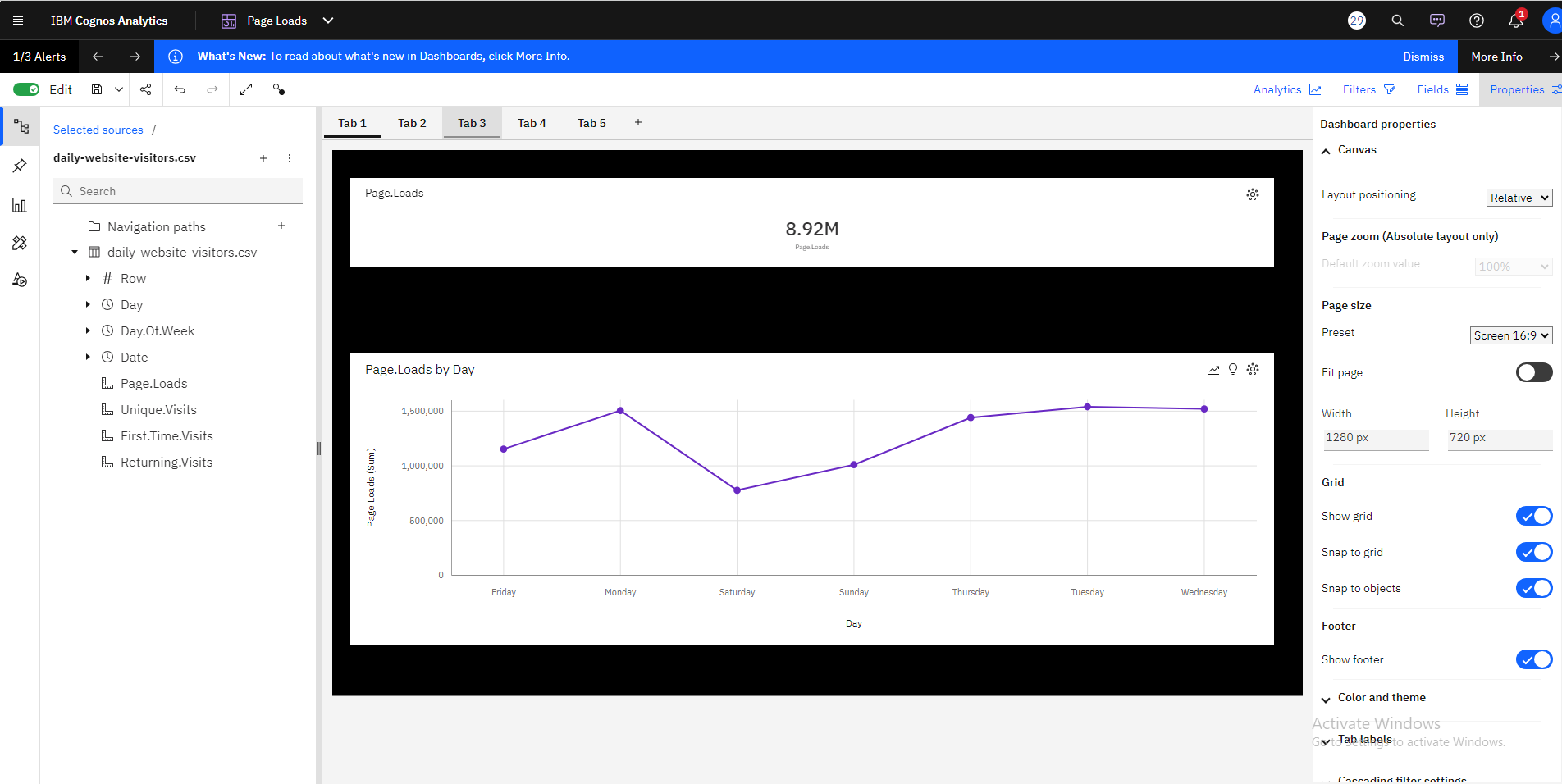
* Calculate the metrics related to user interaction (page views, time on site, bounce rate, etc.).
* To create a linear regression model based on different visitor-related features in order to forecast the quantity of page loads.
* To evaluate the model's performance with data visualization and assessment metrics.
* To see how certain features and page loads relate to one another.
* To ascertain whether the model is a good predictor of website page loads and how well it fits the data.

**DATA COLLECTION PROCESS:**

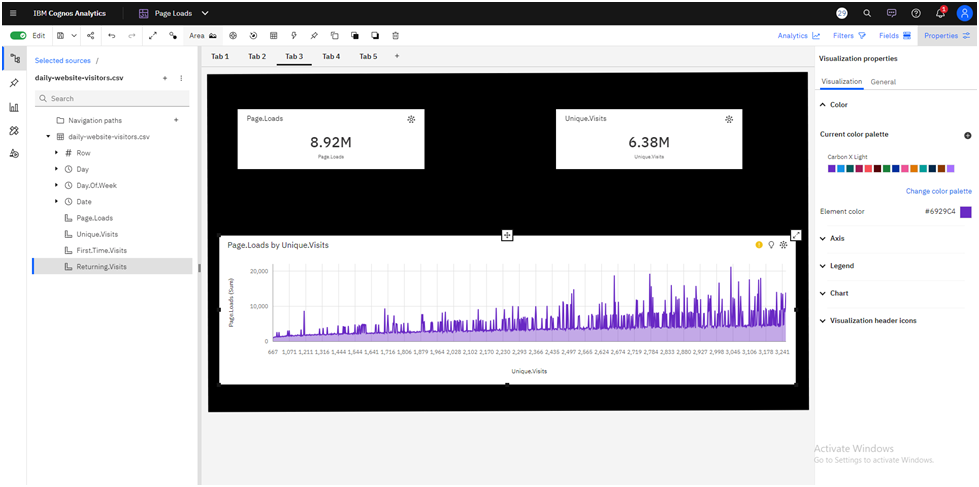
* Setting objectives and selecting data sources is the first step in the process of gathering website traffic data.
* Event tracking is configured, KPIs are defined, and tracking code or tools are put into place. Data that has been gathered is cleaned, kept safe, and updated often.
* Website optimization is driven by data analysis, and privacy compliance is essential.

**DATA VISUALIZATION USING IBM COGNOS:**

**LINE CHART:**

****

* Page Loads Show How Frequently Visitors To Your Website View Or Load Specific Pages.
* The Independent variable and Dependent variable used here are Day and Page Loads(sum) respectively.
* Page.Loads are unusually low when Day is Saturday.
* 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.

****

* It is expected that by Monday+1, 2844 will have surpassed 3039 in Page.Loads

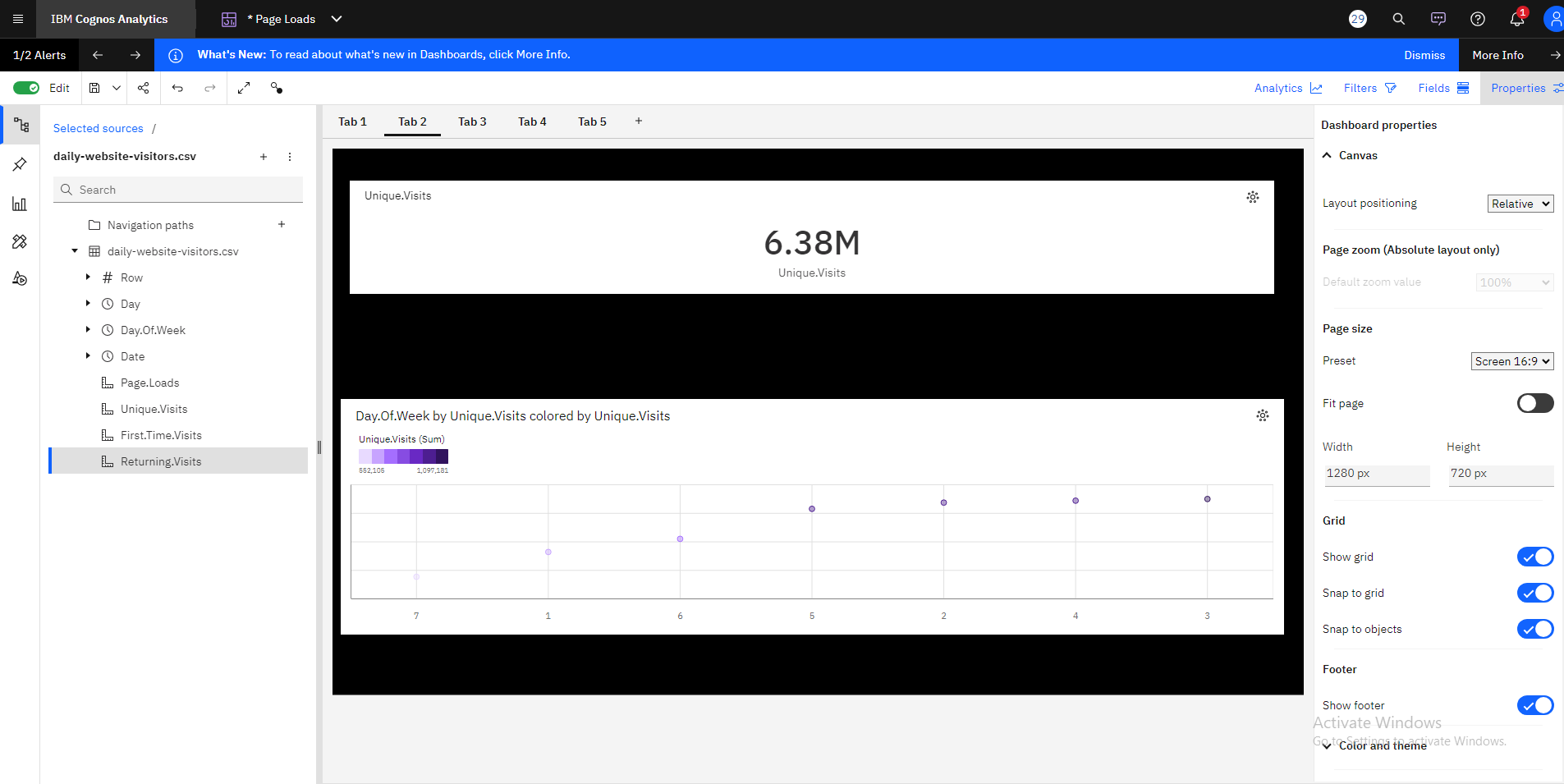
410 times

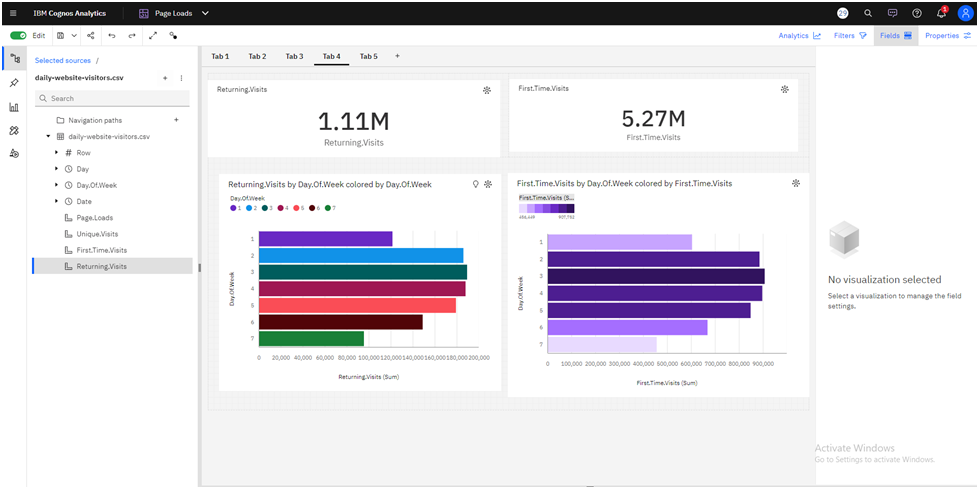
* The Independent variable and Dependent variable used here are Unique visit and Page Loads(sum) respectively
* Page.Loads can reach above a thousand when Unique.When Unique, the

number of visitors increases from 667 to over 21 thousand.The number of visits

is 3039.

**SCATTER PLOT:**



* You may see the link between the days of the week and the number of unique visits to your website by making a scatter plot using "Days of the Week" on the x-axis and "Unique Visits" on the y-axis.
* Based on the current forecasting, Unique.Visits may reach almost 481 thousand by Day.Of.Week 9.
* For Unique.Visits, the most significant values of Day.Of.Week are 3, 4, 2, 5, and 6.

● When the day of the week is 7, the number of returning visitors is exceptionally

Low.

● The total of Returning over all values of Day.Of.Week and Day.Of.Week.Over 1.1

million people have visited.Returning.Visits may exceed 87 thousand by

Day.Of.Week 9 based on current projections.

● According to current projections, First.Time.The number of visitors may exceed

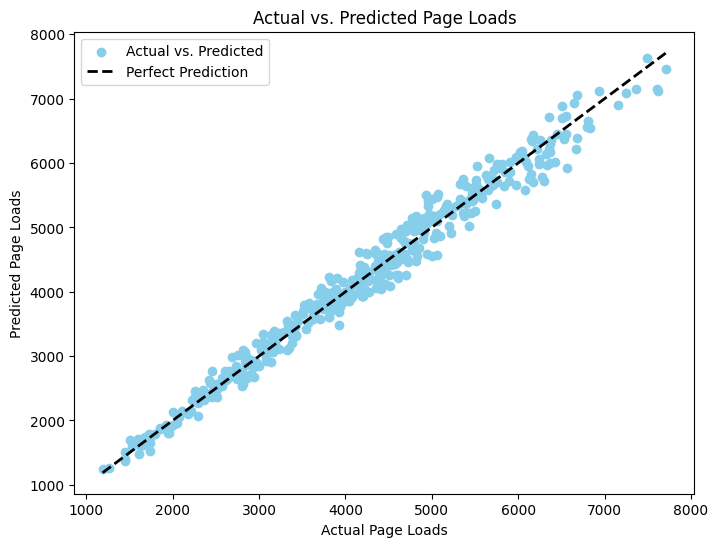
395 thousand every day.Week nine.

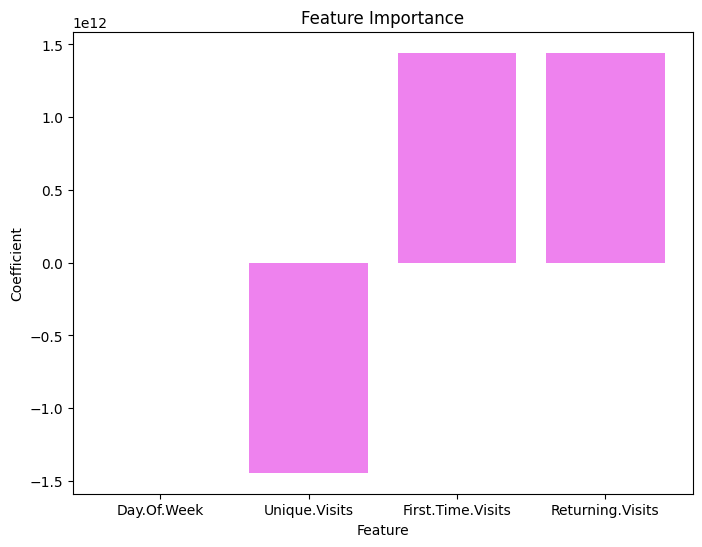
● The total of First.Time.Visits across all Day.Of.Week values is around 5.3 million.

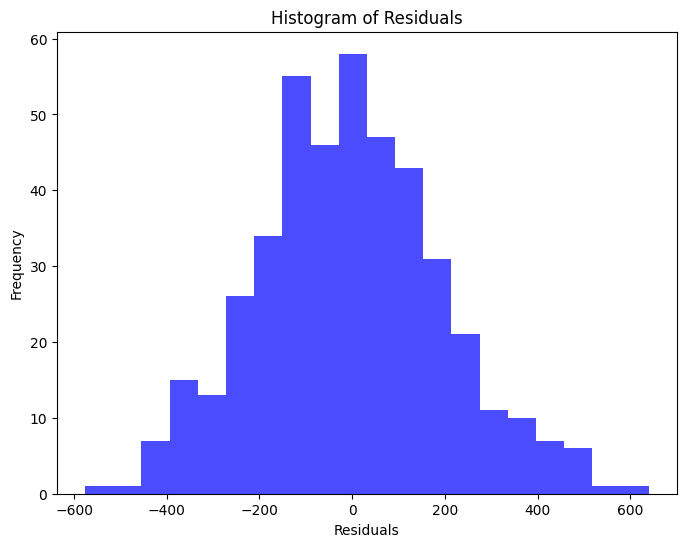
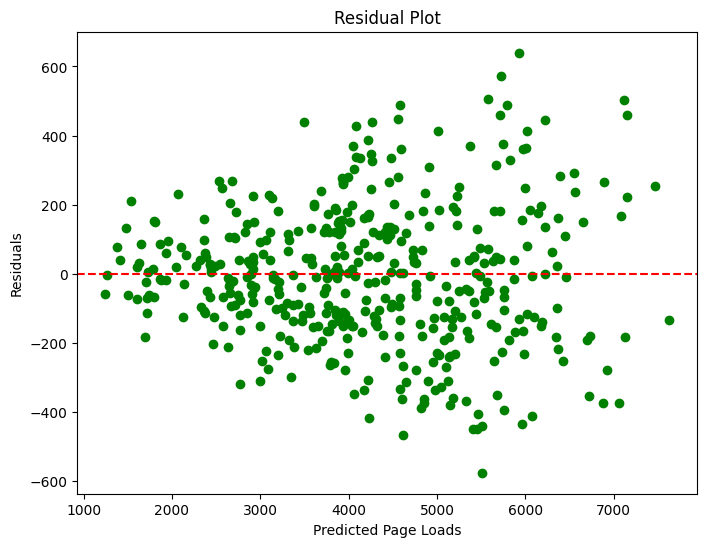
1. IBM Cognos has been critical in our analysis and visualization process.
2. We have used its powers to convert complicated website traffic data analysis into simple and understandable visual representations.
3. These insights have proven critical in understanding user behavior, popular site areas, and successful traffic sources, which have guided our efforts to improve user experiences and overall website performance.

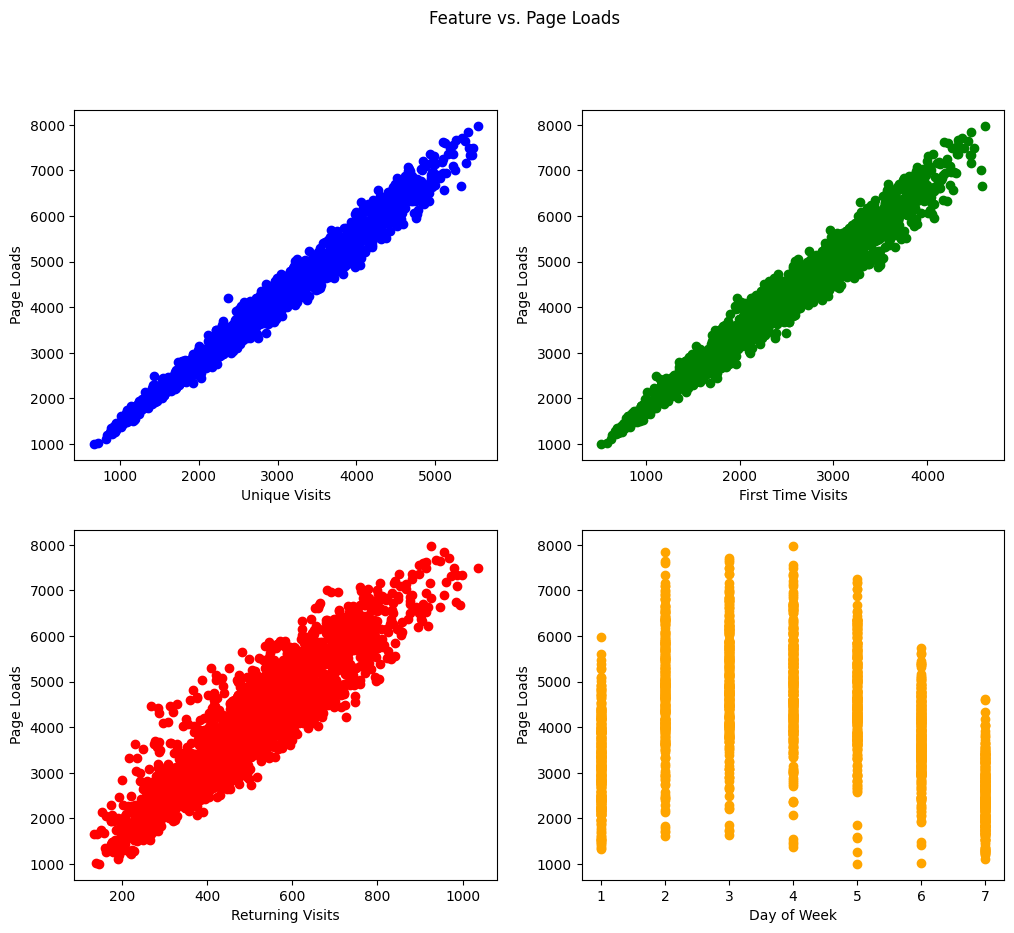
**PYTHON CODE INTEGRATION:**

* This Python script divides the data into training and testing sets after loading and preprocessing it.
* After that, it creates a linear regression model, forecasts data, and uses metrics like MAE, MSE, RMSE, and R2 to assess the model.
* A bar plot for feature relevance, residual plots, a histogram of residuals, and scatter plots for actual vs. predicted values are some examples of the visualizations.
* It also looks at the connection between specific features and page loads using four scatter plots.
* In terms of predicting website page loads, this code provides an extensive study of the prediction performance and feature relationships of the model.









**PERFORMANCE METRICS :**

A key performance metric in website traffic analysis is accuracy. It provides a simple approach to evaluate the overall accuracy of your model's predictions by tracking the percentage of instances in your dataset that were properly predicted. While lower accuracy reveals a need for prediction correctness improvement, higher accuracy indicates a better-performing model.

| **Metric** | **Value** |
| --- | --- |
| **Mean Absolute Error (MAE)** | 308.16 |
| **Mean Squared Error (MSE)** | 147,553.41 |
| **Root Mean Squared Error** | 384.02 |
| **R-squared (R2) Score** | 0.94 |

**Mean Absolute Error (MAE):**

* The average absolute difference between the expected and actual values is measured by the Mean Absolute Error.
* The average absolute difference in this instance between the actual and anticipated "Page Loads" is roughly 308.16.
* It displays the mean size of the model's prediction mistakes.

**Mean Squared Error (MSE):**

* The average of the squared discrepancies between the expected and actual values is called the mean squared error.
* This model's mean square error (MSE) is roughly 147,553.41, meaning that higher errors are given more weight.
* It is a measurement of the squared average error magnitude.

**Root Mean Squared Error (RMSE):**

* The MSE's square root is known as the Root Mean Squared Error. The target variable and it are both stated in the same unit.
* The RMSE in this instance is roughly 384.02, which offers a more comprehensible indication of the inaccuracy.
* It helps to comprehend the average magnitude of prediction mistakes in the model.

**R-squared (R2) Score:**

* The percentage of variance in the target variable (Page Loads in this case) that can be predicted from the independent variables (Day of the Week, Unique Visits, First Time Visits, and Returning Visits) is measured by the R-squared score.
* With an R2 value of 0.940, the independent variables in the model can account for almost 94% of the variance in "Page Loads".
* This shows that, depending on the metrics supplied, the model can accurately forecast "Page Loads" and has a strong fit with the data.

**IMPROVING USER EXPERIENCE:**

* Gained insights can direct content initiatives, resulting in the creation of more pertinent and interesting content.
* Knowing user demographics helps you customize the website to the interests of particular users.
* User behavior analysis aids in UI/UX improvement and site navigation optimization.
* Adjustments that improve user experience can be validated and implemented through A/B testing based on the insights.

**CONCLUSION:**

* This website traffic analysis project seeks to provide website owners with critical insights into user behavior and key traffic sources. The careful data collecting assures depth and quality by focusing on important metrics like page views, unique visitors, and referrals.
* Using IBM Cognos for clear visualizations, this investigation strives to improve user experiences and overall website performance. By concentrating on user interaction, traffic sources, and navigation, practical insights for a more user-centric website building strategy were obtained.