Website Traffic Analysis

Data Analytics with Cognos

Project Submission Part 5: Project Documentation & Submission

Sathishkumar / 511921104064

Priyadarshini Engineering College

Website Traffic Analysis – Data Analysis with Cognos:

Document Title: "Website Traffic Analysis"

Course : Data Analytics with Cognos (IBM).

Author / Editor: Sathish Kumar

Register Number: 511921104064

College Name: Priyadarshini Engineering College

Acknowledgments:

I would like to express my heartfelt gratitude to the individuals and organizations who have been instrumental in the completion of this project, "Website Traffic Analysis." Their unwavering support, guidance, and expertise have been invaluable, and I wish to extend my appreciation to each of them.

Priyadarshini Engineering College: I am deeply thankful to Priyadarshini Engineering College for providing the environment and resources that enabled me to pursue this endeavor.

Tutors from Skillup: I would also like to extend my thanks to the dedicated tutors from Skillup, whose guidance and expertise enriched my understanding of data analysis, making this project possible.

Table of Contents

Abstract

In an era driven by digital interactions and online experiences, understanding and optimizing website traffic have become paramount for website owners and businesses. This project, "Website Traffic Analysis on Daily Website Visitors," embarks on a journey to unveil the insights hidden within the data generated by daily visitors to a website.

The project follows a comprehensive approach to data analytics, integrating a design thinking methodology to identify, frame, and address critical issues related to website performance and user engagement. By employing a blend of data collection, analysis, and visualization techniques, this project sheds light on the daily patterns of website visitors, user behavior, and the factors that influence their interactions.

The key objectives of this project are to enhance our understanding of website traffic, identify opportunities for optimization, and provide actionable insights to website owners. This documentation not only outlines the methodology and development phases undertaken but also illuminates the significance of data-informed decision-making in a digital landscape.

Through this documentation, readers will gain valuable insights into the design thinking process, problem statement definition, and the overarching goals of the analysis, setting the stage for a deeper exploration of the data analytics journey that follows.

1. Introduction

1.1 Project Overview

The "Website Traffic Analysis" project revolves around the analysis of daily website visitor data. It offers an in-depth examination of website performance, user engagement, and the factors that influence daily visitor patterns. This project encompasses the entire data analytics process, from data collection to visualization and interpretation, with a specific focus on daily visitor trends.

1.2 Objectives of the Documentation

The objectives of this documentation are as follows:

- To provide a comprehensive project overview and establish the context for analyzing daily website visitor data.
- To outline the design thinking process and the various development phases employed in the project.

- To describe the specific analysis objectives and methodologies used for data collection and visualization, utilizing IBM Cognos and Python.

- To clarify how the insights derived from the analysis can empower website owners to optimize their online presence, improve user experiences, and make informed decisions based on daily visitor data.

These objectives guide the structure and content of the documentation, making it a valuable resource for website owners, analysts, and data enthusiasts seeking to leverage daily website visitor data for informed decision-making.

2.1 Problem Statement

The problem of "Website Traffic Analysis" lies in the need for organizations to effectively understand and leverage user behavior on their websites.

Necessity: In today's digital world, websites are central to business success, serving as primary touchpoints for customers, clients, and stakeholders.

Current Limitations: The current limitations in Website Traffic Analysis often revolve around the quality and depth of insights obtained.

Potential Benefits: Website Traffic Analysis offers numerous benefits.

Moreover, Website Traffic Analysis can uncover opportunities for innovation. By identifying patterns and trends in user behavior, organizations can develop new features or services that align with user needs. It can also help in personalizing user experiences, tailoring content and recommendations based on individual preferences.

2.2 Design Thinking

Empathize:
In the empathize phase, we are committed to truly understanding our users' perspectives and needs regarding website traffic analysis.

Define:
Building on our empathetic understanding, the define phase involves crystallizing the specific user needs and challenges related to website traffic analysis.

Ideate:
In the ideation phase, we foster a collaborative environment where cross-functional teams brainstorm creative solutions to address these user needs and challenges.

Prototype:
To transform these ideas into tangible solutions, we employ a prototype strategy.

Test:
Testing is a pivotal phase where we engage users to gather feedback on the prototypes.

Iterate:
Design thinking is inherently iterative, and we emphasize that we'll repeat the prototyping and testing stages as necessary.

2.3 Goals

The overarching goal of our website traffic analysis initiative is to elevate the effectiveness of our decision-making processes

Enhance Decision-Making : Our foremost objective is to empower our decision-makers with precise, user-centric information.

Analysis Objectives

Popular Pages: Identify the most visited pages on the website to understand user interests.

Traffic Trends: Analyze traffic patterns over time to detect seasonality, trends, or unusual spikes.

User Engagement Metrics:Evaluate user engagement through metrics like bounce rate, time on page, and conversion rates to assess website performance.

Traffic Sources: Determine the sources of traffic (organic, direct, referral, paid) and their contributions to user engagement and conversions.

3.Development Phases

3.1 Phase 2: Innovation

3.1.2 INNOVATION OBJECTIVES

Data Acquisition:
Download and import the "Daily Website Visitors" dataset from Kaggle into your analytics environment.

Exploratory Data Analysis (EDA):
Conduct EDA to understand the dataset's characteristics, patterns, and correlations. Visualize key metrics and trends in website traffic.

Analysing the Data:
Regularly analyze data collected from the real-time analytics tools and experiments. Adjust strategies based on insights derived from the dataset.

Predicting accurate values using LSTM network:
Analyzing data in a dataset is a crucial step in extracting meaningful insights, patterns, and knowledge from the raw information contained within it

3.1.3 INCLUSION OF MACHINE LEARNING MODEL

Regression Analysis: Linear regression or more advanced methods like polynomial regression can be used to model and predict traffic trends over time.

Time Series Analysis: Techniques like ARIMA (AutoRegressive Integrated Moving Average) can help in forecasting website traffic based on historical data.

Classification Algorithms: These can be used to categorize website visitors, such as decision trees, random forests, or support vector machines, to identify different user segments or traffic sources.

Recommendation Systems : Collaborative filtering or content-based recommendation algorithms can be be used to suggest content to users based on their behavior.

3.1.4 VISUALIZATION METHOD

1. Bar Charts: Bar charts are useful for displaying metrics like the number of page views, unique visitors, or bounce rates over a specific time period. You can create bar charts to compare different time intervals or website sections.

2. Line Charts: Line charts are effective for showing trends in website traffic data, such as changes in visitor numbers over time. They can help identify seasonal patterns and long-term trends.

3. Pie Charts: Pie charts can be used to represent the distribution of traffic sources, showing the percentage of traffic coming from direct visits, search engines, social media, etc.

4. Area Charts: Area charts are similar to line charts but can be used to display the cumulative effect of website traffic data over time. They are good for visualizing total page views or unique visitors.

5. Heat Maps: Heat maps can provide insights into user engagement by showing which parts of a webpage receive the most clicks or interaction. This helps in optimizing the website's layout.

Dashboards: Dashboards in IBM Cognos allow you to combine various visualizations and key metrics on a single screen, offering a comprehensive overview of website traffic.

3.2 Phase 3 Development Part 1

3.2.1 Data Visualization using IBM Cognos

3.2.1.1 Introduction to IBM Cognos :

IBM Cognos is a powerful and comprehensive business intelligence (BI) and performance management tool designed to help organizations make data-driven decisions and drive business success. It is part of the IBM Analytics portfolio and has gained widespread recognition for its ability to transform raw data into actionable insights, aiding in the management and optimization of business processes.

Cognos offers a wide range of features and capabilities, including data visualization, reporting, data exploration, and advanced analytics. It empowers users to access and analyze data from various sources, providing a holistic view of an organization's data landscape. This facilitates not only data exploration and reporting but also the creation of interactive dashboards and dynamic reports, enabling users to interact with their data in meaningful ways.

One of the key strengths of IBM Cognos is its user-friendly interface, which makes it accessible to both business analysts and IT professionals. With Cognos, non-technical users can perform sophisticated data analysis, create insightful reports, and share findings with stakeholders across the organization.

3.2.1.2 Creating Models on IBM Cognos

Creating a model in IBM Cognos typically involves defining the structure and relationships of the data you want to use for reporting and analysis. Here are the general steps for creating a model in IBM Cognos:

1. Access IBM Cognos:
   - Log in to your IBM Cognos environment with the appropriate credentials.

2. Data Source Connection:
   - Connect to your data source. Cognos can connect to a variety of data sources, including databases, flat files, and web services. Ensure that you have the necessary permissions to access the data source.

3. Create a New Project or Package:
   - In Cognos, you typically start by creating a project or package. A package is a container for organizing data and metadata. You can create a new package from scratch or use existing metadata.

4. Define Data Source Connections:
   - Specify the data source connections within your project or package. This involves configuring the database connection details, including server, port, credentials, and database schema.

5. Import Metadata:
   - Import metadata from the data source. Metadata describes the

structure and relationships of the data. Cognos can automatically discover and import metadata, or you can define it manually.

6. Define Data Items:
- Define the data items you want to include in your model. Data items are typically columns from tables in your data source. You can rename them and specify their data types.

7. Create Relationships:
- Define relationships between data items to establish the structure of your data model. You can specify primary and foreign key relationships to enable data drilling and analysis.

8. Create Calculations (Optional):
- You can create calculated data items in your model. These are derived from existing data items using mathematical operations, conditional statements, or other calculations.

9. Create Filters (Optional):
- Define filters to limit the data that users can access and analyze. Filters can be based on specific conditions or values.

10. Create Business Rules (Optional):
- Implement business rules or logic within the model to standardize data or provide default values.

11. Set Security and Permissions (Optional):
- Define user-level security and permissions for the model, specifying who can access, view, or modify it.

accurately reflects your data and provides the expected results.

12. Save and Publish:

- Save your model within the Cognos environment. Once it's ready, you can publish it for use in reports and dashboards.

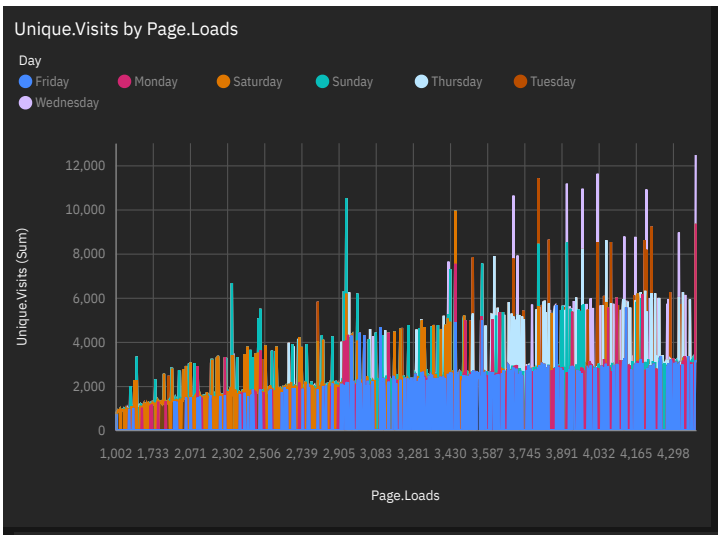
The steps may vary slightly depending on the specific version of IBM Cognos you are using. Additionally, it's a good practice to refer to the IBM Cognos documentation and guidelines for detailed, version-specific instructions on creating and managing models.

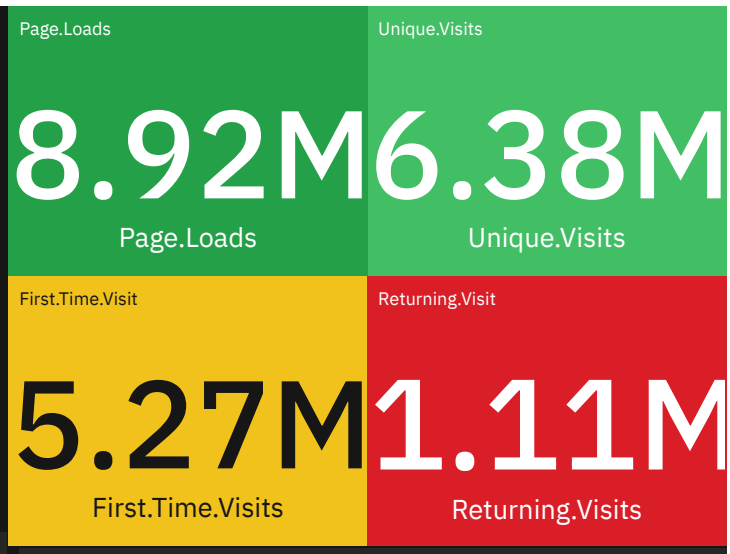
Here ,go with the Data Preprocessing, Exploration of Insights and analysis with Data set which have been provided by Kaggle , then segmented into day wise visits for the better understanding.

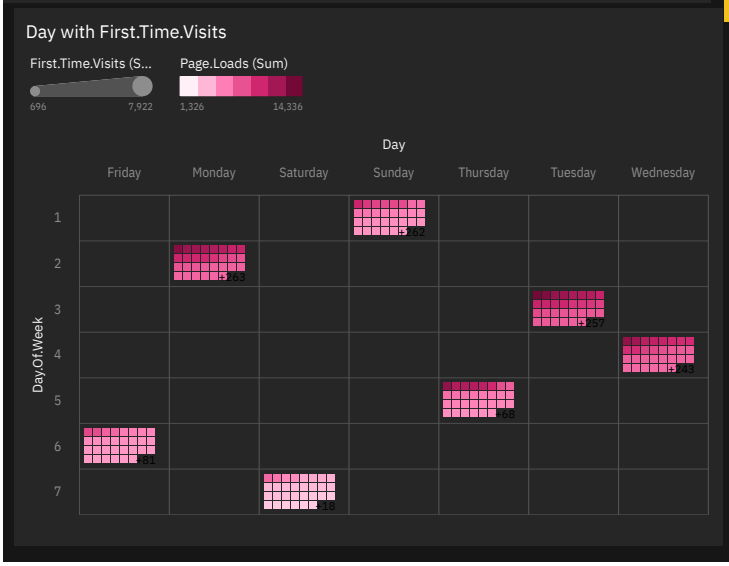
Under below the information provided is general all time overall visits and visits are segmented into day wise like

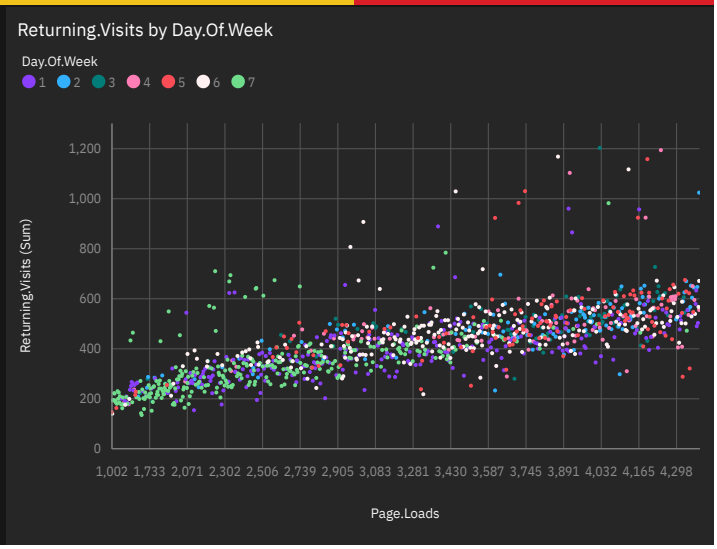
- Sunday
- Monday
- Tuesday
- Wednesday
- Thrusday
- Friday
- Saturday

Tab 1

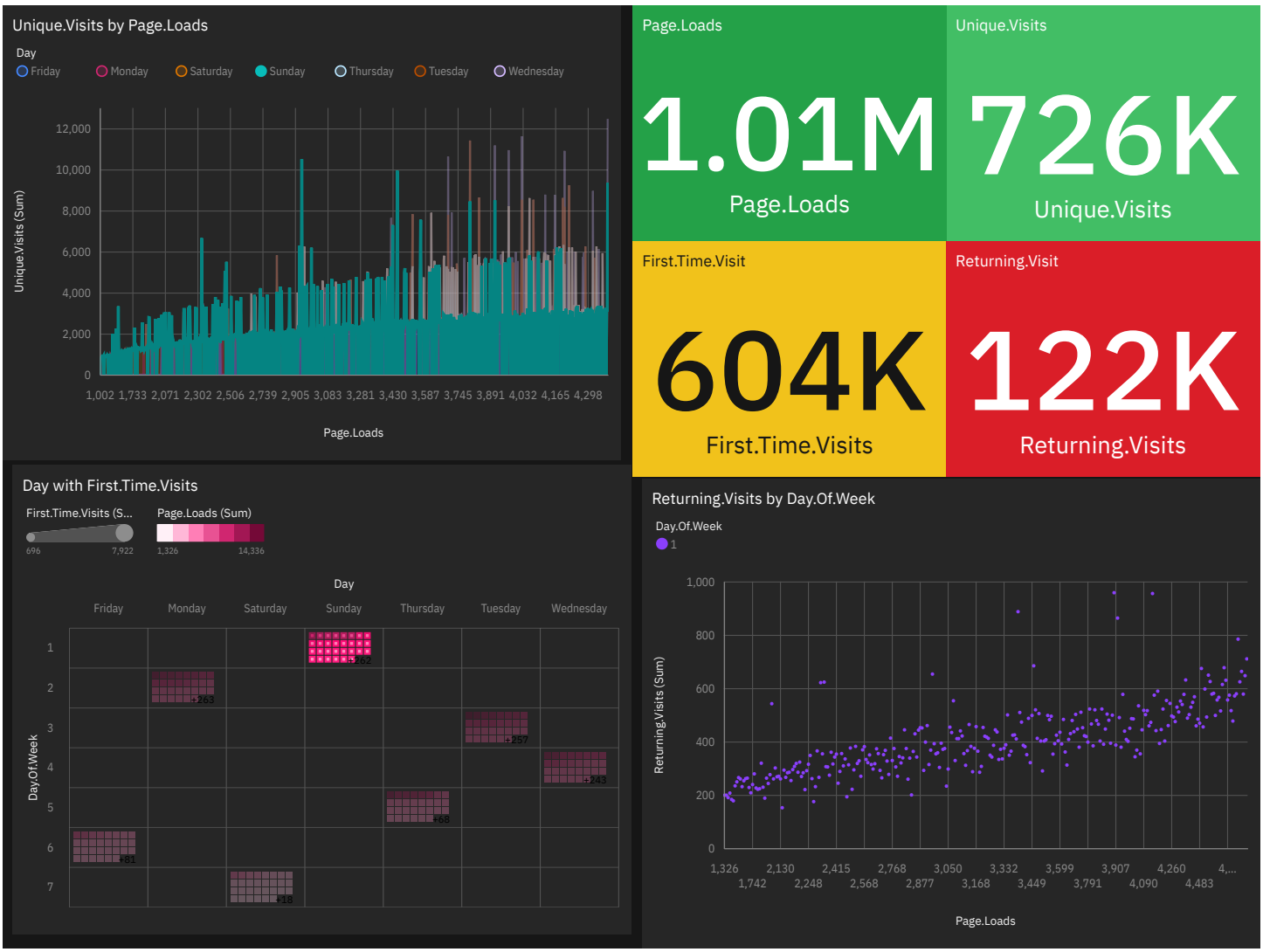

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
12,000
10,000
8,000
6,000
4,000
2,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads


Page.Loads
8.92M
Page.Loads
Unique.Visits
6.38M
Unique.Visits
First.Time.Visit
5.27M
First.Time.Visits
Returning.Visit
1.11M
Returning.Visits

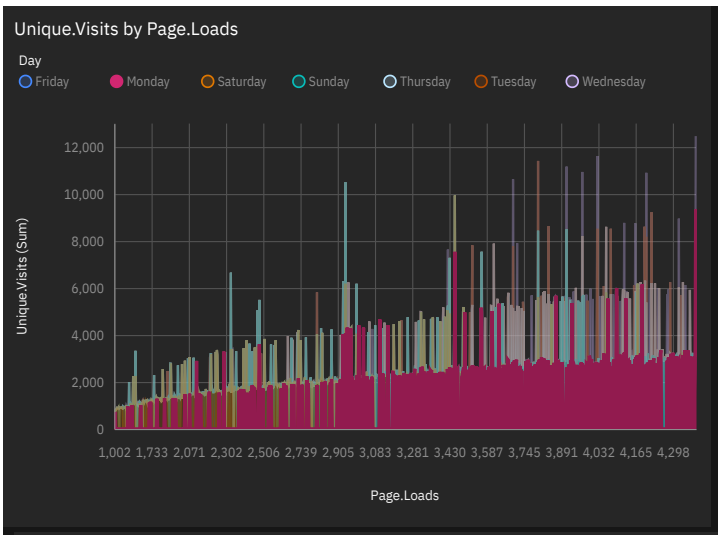

Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7

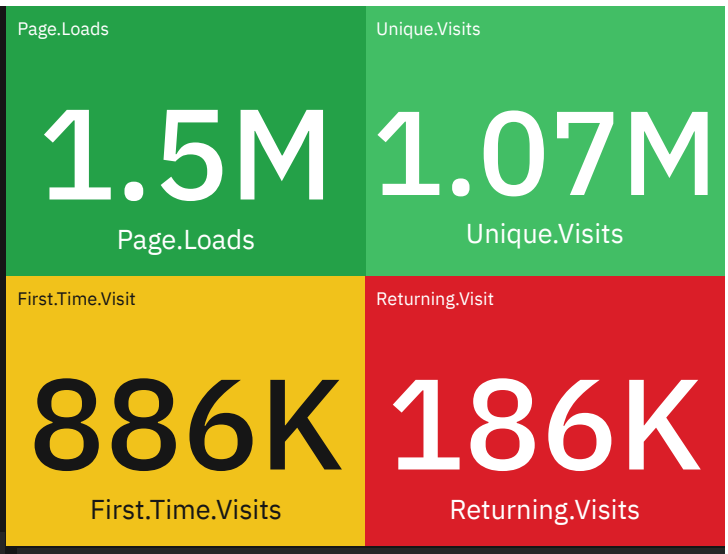

Returning.Visits by Day.Of.Week
Day.Of.Week
1
2
3
4
5
6
7
Returning.Visits (Sum)
1,200
1,000
800
600
400
200
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads

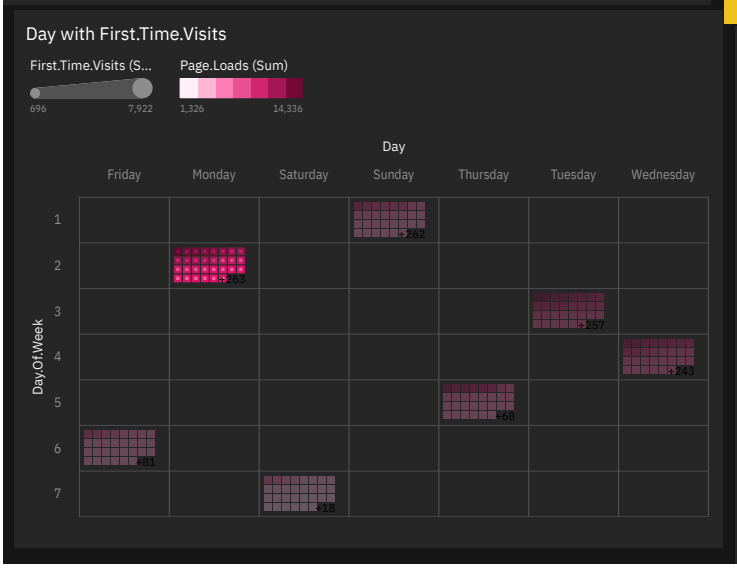
Tab 1

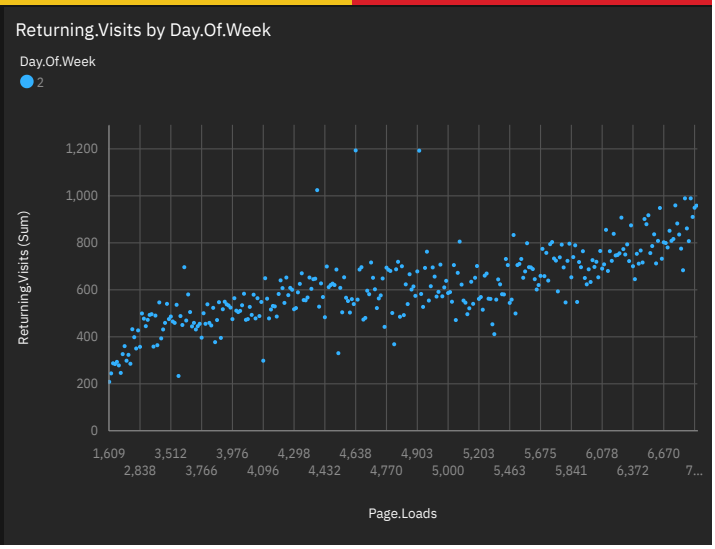

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
12,000
10,000
8,000
6,000
4,000
2,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.01M
Page.Loads
Unique.Visits
726K
Unique.Visits
First.Time.Visit
604K
First.Time.Visits
Returning.Visit
122K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
+262
+263
+257
+243
+68
+81
+48
Returning.Visits by Day.Of.Week
Day.Of.Week
1
Returning.Visits (Sum)
1,000
800
600
400
200
0
1,326 2,130 2,415 2,768 3,050 3,332 3,599 3,907 4,260 4,...
1,742 2,248 2,568 2,877 3,168 3,449 3,791 4,090 4,483
Page.Loads

Tab 1

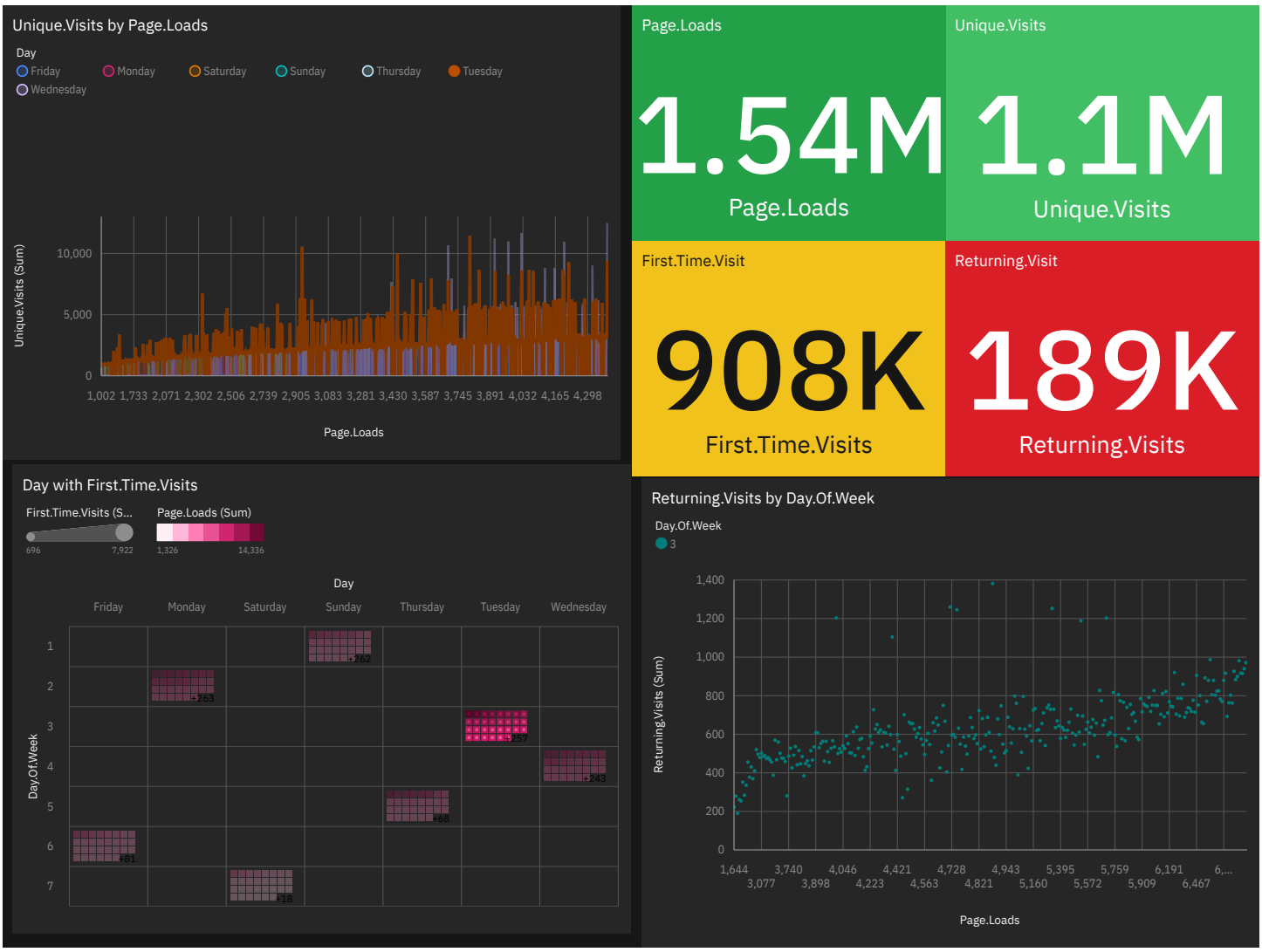

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
12,000
10,000
8,000
6,000
4,000
2,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads


Page.Loads
1.5M
Page.Loads
Unique.Visits
1.07M
Unique.Visits
First.Time.Visit
886K
First.Time.Visits
Returning.Visit
186K
Returning.Visits

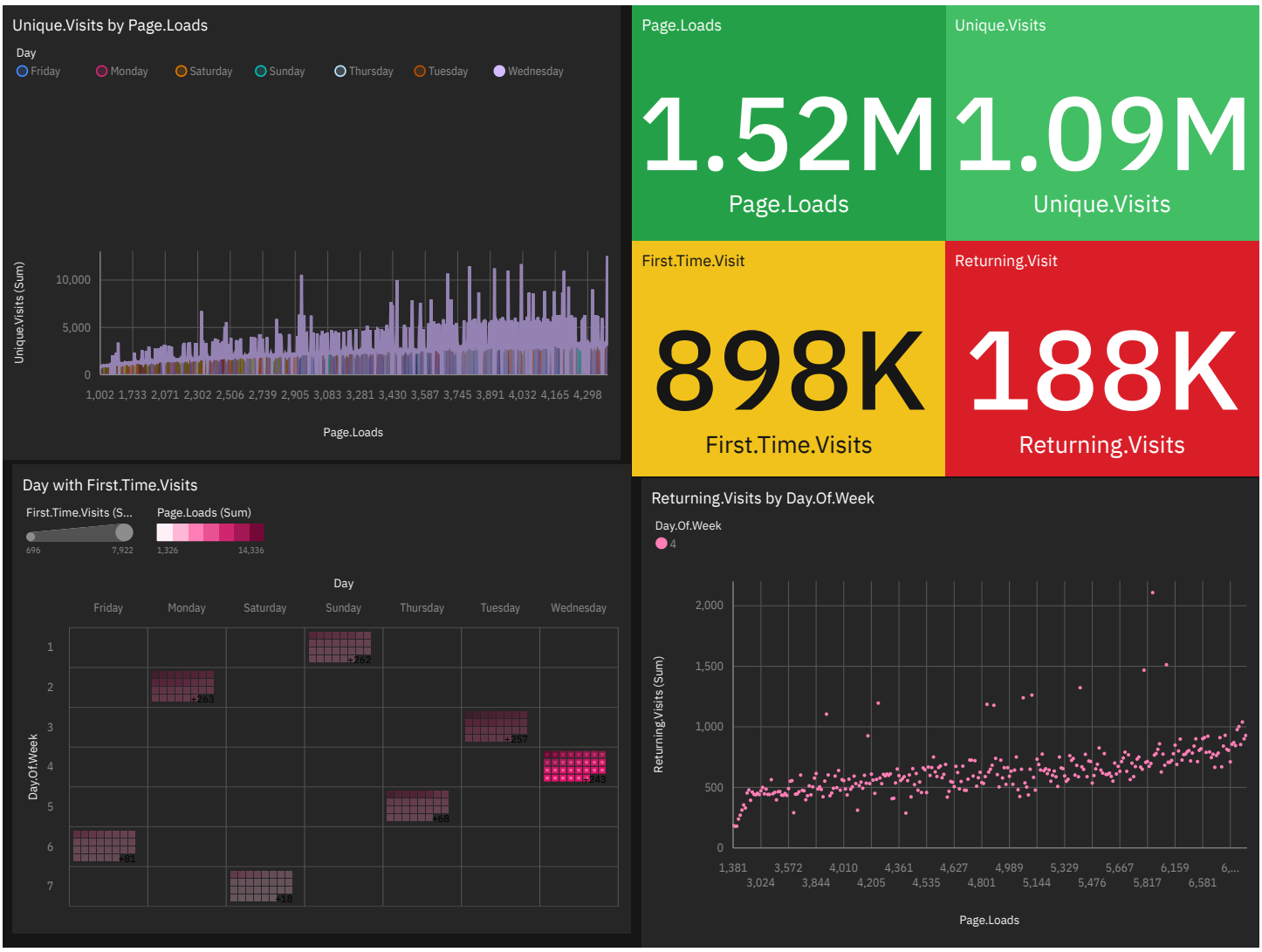

Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
+262
+263
+257
+243
+68
+81
+48


Returning.Visits by Day.Of.Week
Day.Of.Week
2
Returning.Visits (Sum)
1,200
1,000
800
600
400
200
0
1,609 3,512 3,976 4,298 4,638 4,903 5,203 5,675 6,078 6,670
2,838 3,766 4,096 4,432 4,770 5,000 5,463 5,841 6,372 7...
Page.Loads

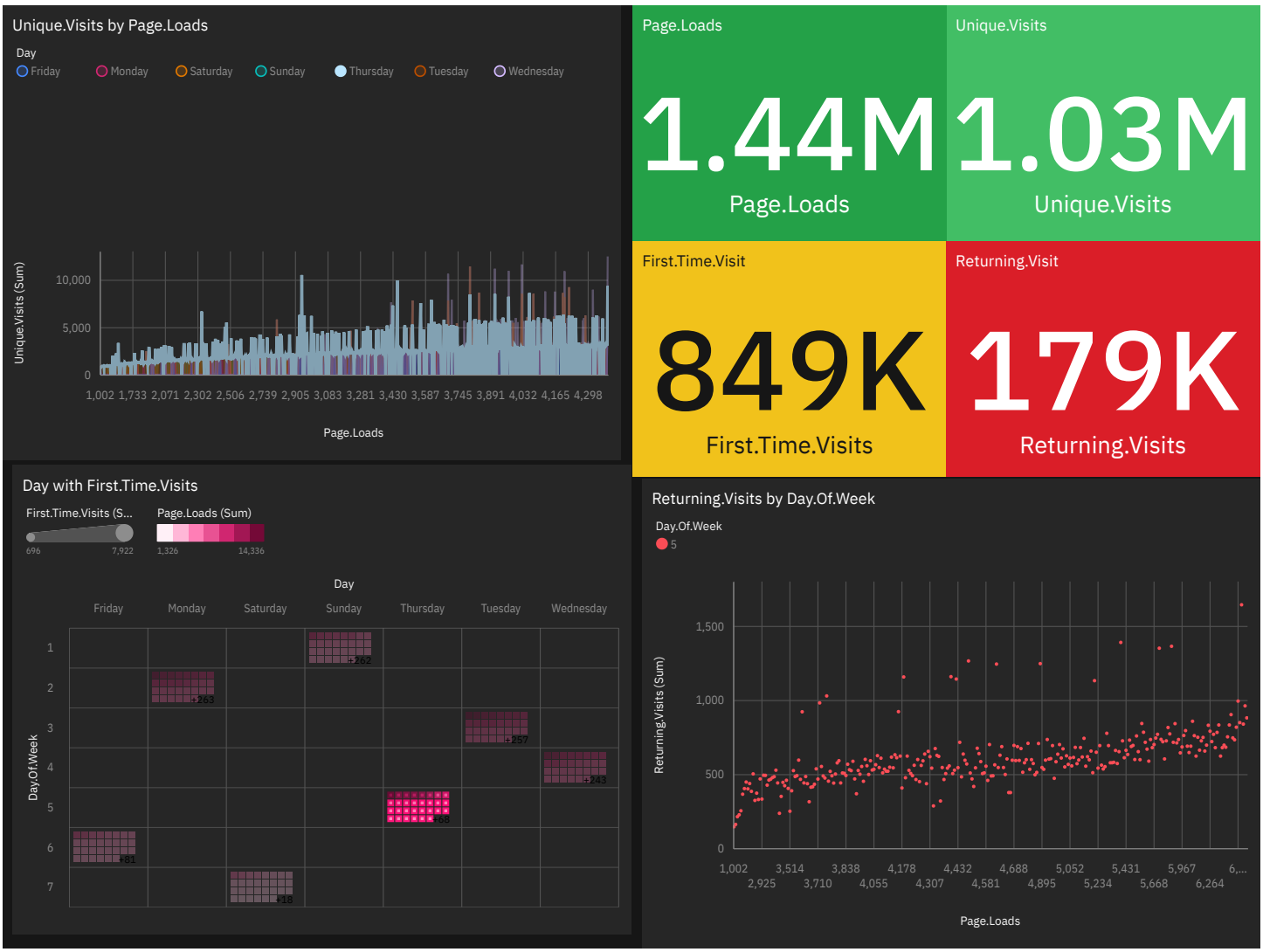
Tab 1


Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.54M
Page.Loads
Unique.Visits
1.1M
Unique.Visits
First.Time.Visit
908K
First.Time.Visits
Returning.Visit
189K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
Returning.Visits by Day.Of.Week
Day.Of.Week
3
Returning.Visits (Sum)
1,400
1,200
1,000
800
600
400
200
0
1,644 3,740 4,046 4,421 4,728 4,943 5,395 5,759 6,191 6,...
3,077 3,898 4,223 4,563 4,821 5,160 5,572 5,909 6,467
Page.Loads

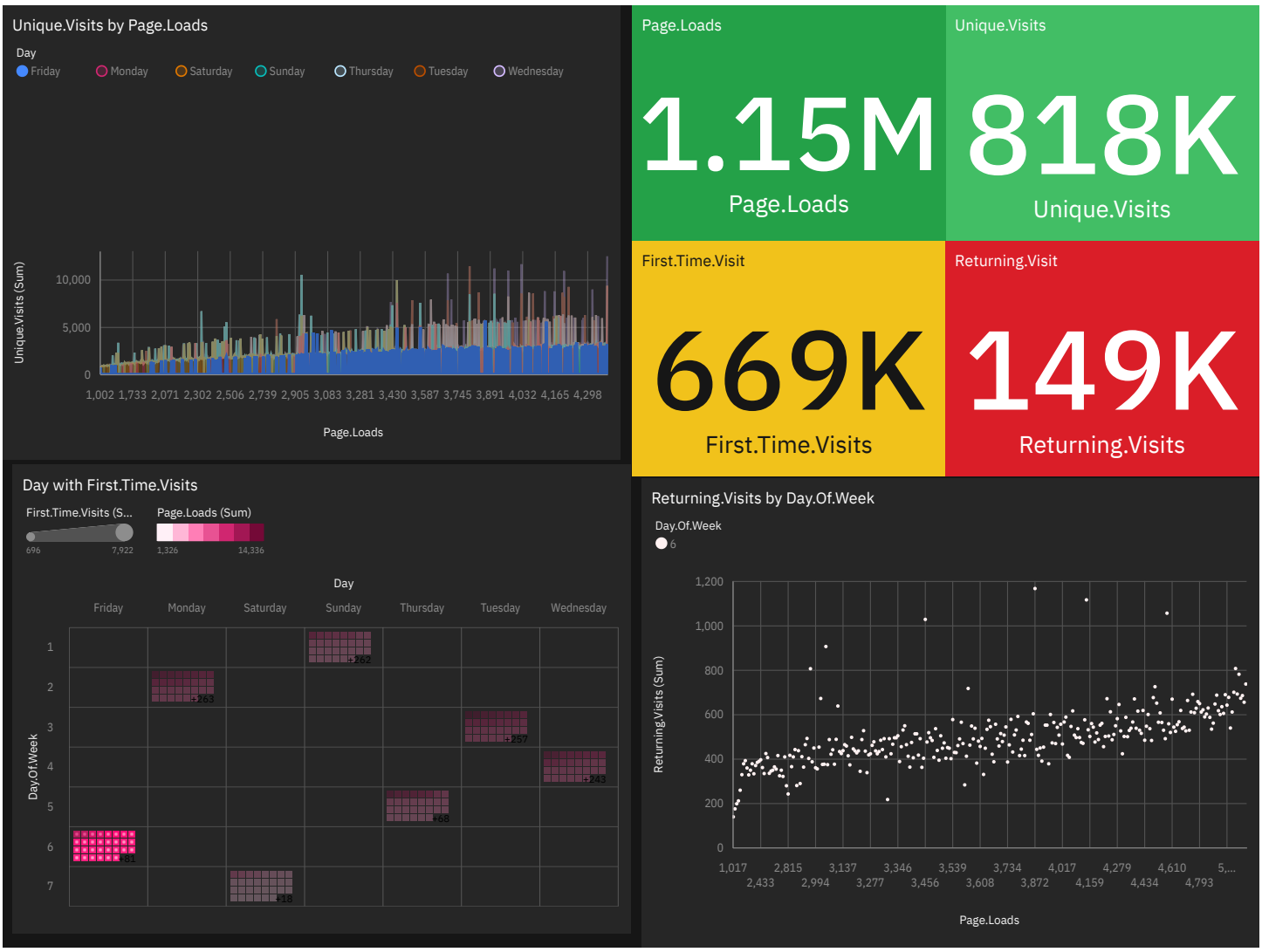
Tab 1


Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.52M
Page.Loads
Unique.Visits
1.09M
Unique.Visits
First.Time.Visit
898K
First.Time.Visits
Returning.Visit
188K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
+262
+263
+257
+243
+68
+81
+48
Returning.Visits by Day.Of.Week
Day.Of.Week
4
Returning.Visits (Sum)
2,000
1,500
1,000
500
0
1,381 3,572 4,010 4,361 4,627 4,989 5,329 5,667 6,159 6,...
3,024 3,844 4,205 4,535 4,801 5,144 5,476 5,817 6,581
Page.Loads

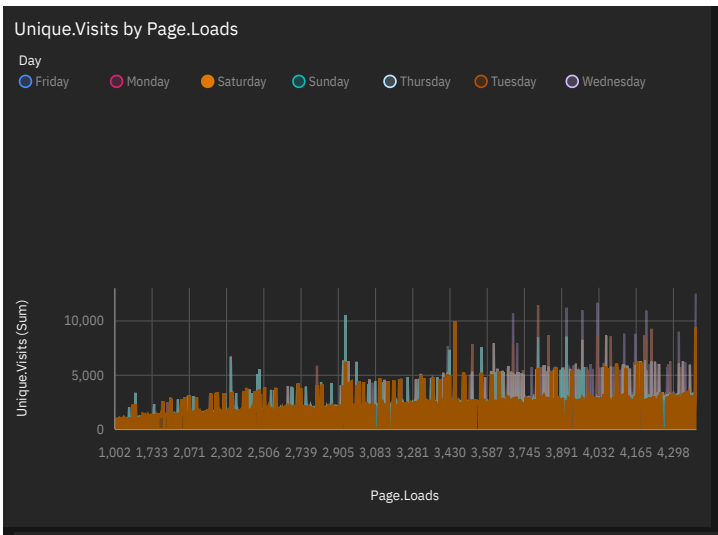
Tab 1

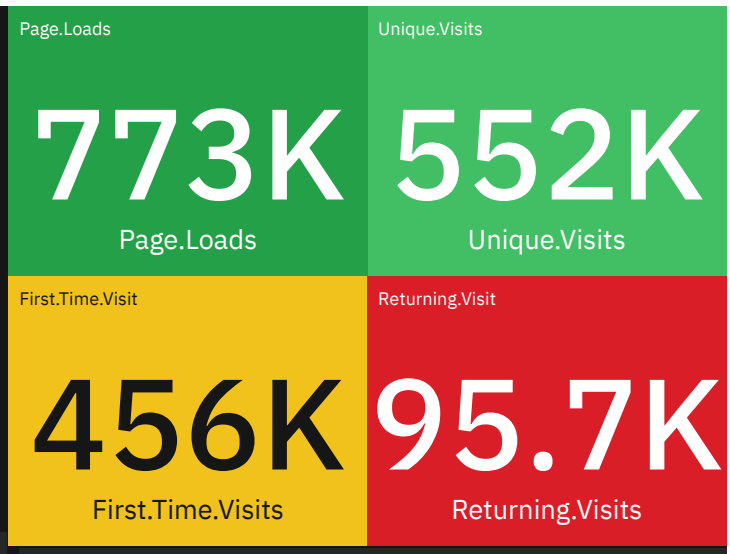

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.44M
Page.Loads
Unique.Visits
1.03M
Unique.Visits
First.Time.Visit
849K
First.Time.Visits
Returning.Visit
179K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
Returning.Visits by Day.Of.Week
Day.Of.Week
5
Returning.Visits (Sum)
1,500
1,000
500
0
1,002 2,925 3,514 3,710 3,838 4,055 4,178 4,307 4,432 4,581 4,688 4,895 5,052 5,234 5,431 5,668 5,967 6,264
Page.Loads

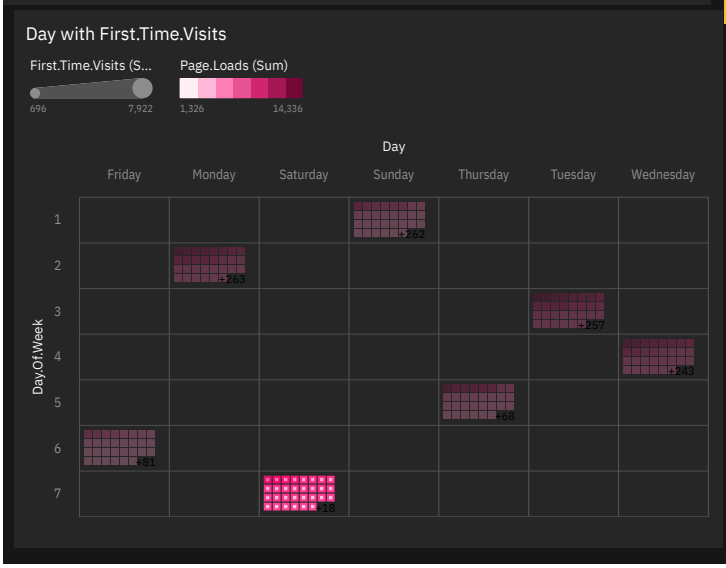
Tab 1

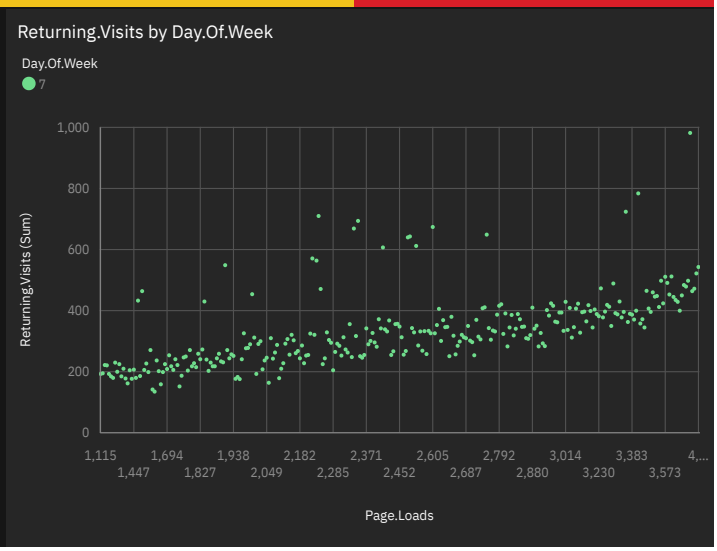

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.15M
Page.Loads
Unique.Visits
818K
Unique.Visits
First.Time.Visit
669K
First.Time.Visits
Returning.Visit
149K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
+262
+263
+257
+243
+68
+48
Returning.Visits by Day.Of.Week
Day.Of.Week
6
Returning.Visits (Sum)
1,200
1,000
800
600
400
200
0
1,017 2,815 3,137 3,346 3,539 3,734 4,017 4,279 4,610 5,...
2,433 2,994 3,277 3,456 3,608 3,872 4,159 4,434 4,793
Page.Loads

Tab 1

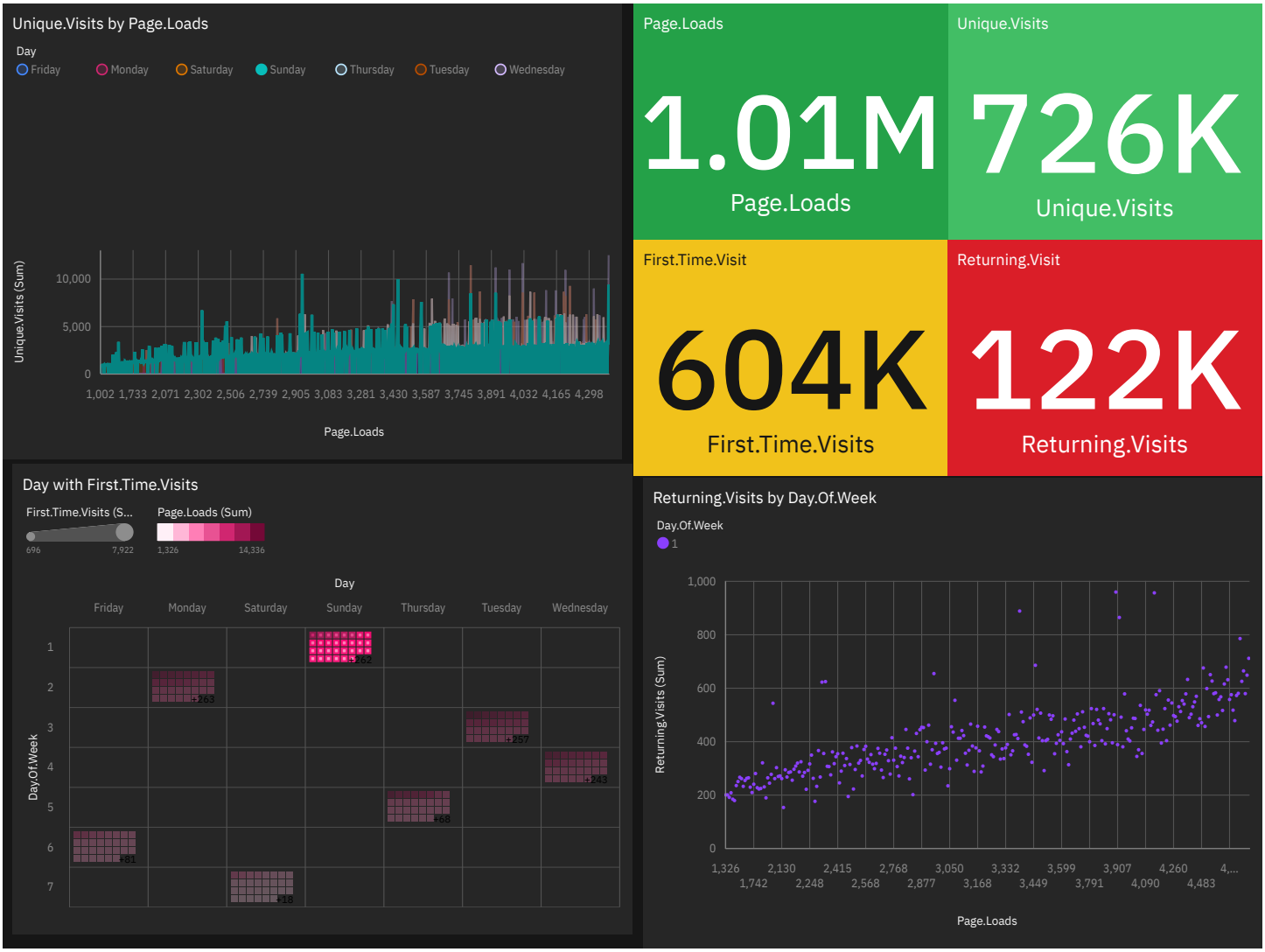

Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads


Page.Loads
773K
Page.Loads
Unique.Visits
552K
Unique.Visits
First.Time.Visit
456K
First.Time.Visits
Returning.Visit
95.7K
Returning.Visits


Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
+262
+263
+257
+243
+68
+81


Returning.Visits by Day.Of.Week
Day.Of.Week
7
Returning.Visits (Sum)
1,000
800
600
400
200
0
1,115 1,694 1,938 2,182 2,371 2,605 2,792 3,014 3,383
1,447 1,827 2,049 2,285 2,452 2,687 2,880 3,230 3,573
Page.Loads

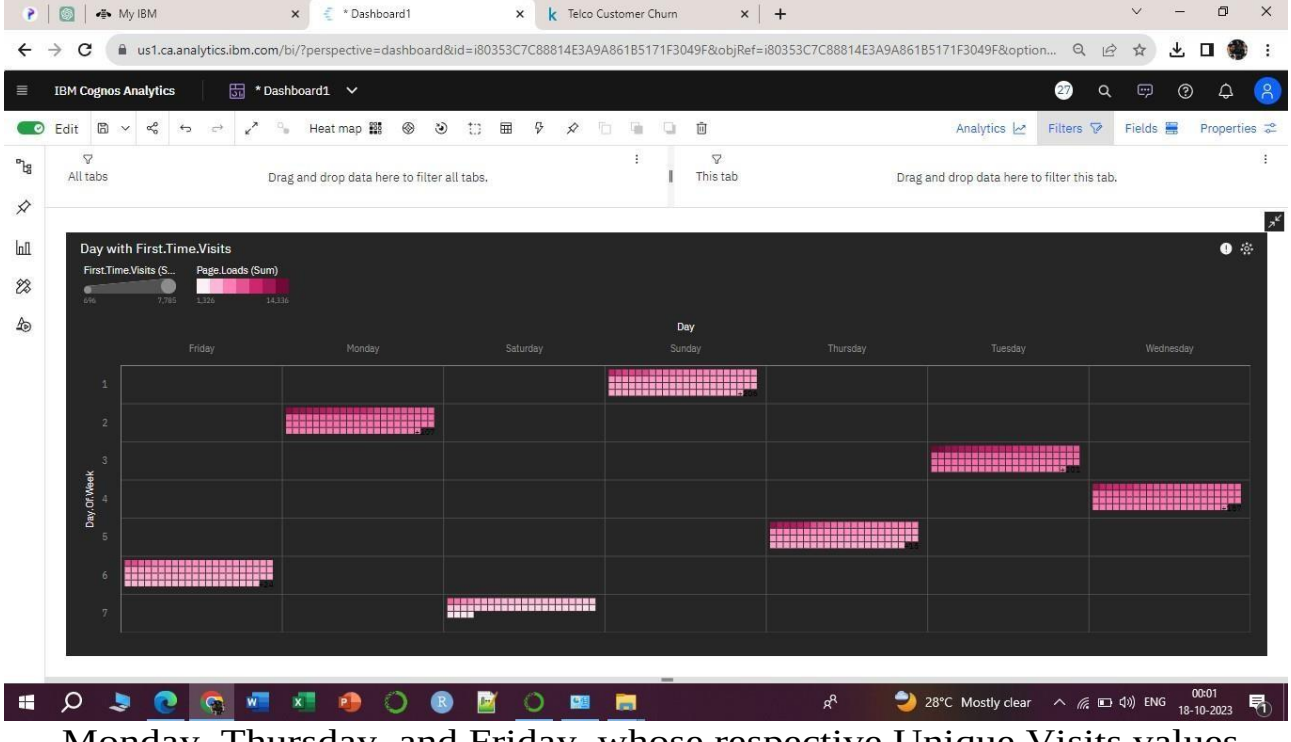
Tab 1


Unique.Visits by Page.Loads
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Unique.Visits (Sum)
10,000
5,000
0
1,002 1,733 2,071 2,302 2,506 2,739 2,905 3,083 3,281 3,430 3,587 3,745 3,891 4,032 4,165 4,298
Page.Loads
Page.Loads
1.01M
Page.Loads
Unique.Visits
726K
Unique.Visits
First.Time.Visit
604K
First.Time.Visits
Returning.Visit
122K
Returning.Visits
Day with First.Time.Visits
First.Time.Visits (S...
696
7,922
Page.Loads (Sum)
1,326
14,336
Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day.Of.Week
1
2
3
4
5
6
7
Returning.Visits by Day.Of.Week
Day.Of.Week
1
Returning.Visits (Sum)
1,000
800
600
400
200
0
1,326 2,130 2,415 2,768 3,050 3,332 3,599 3,907 4,260 4,...
1,742 2,248 2,568 2,877 3,168 3,449 3,791 4,090 4,483
Page.Loads

3.2.1.3 Insightsforabove**"IBM Cognos Analytics"**

Unique Visits :

- Unique.Visits is unusually low when Day is Saturday.
- Based on the current forecasting, Unique.Visits may reach almost 481 thousand by Day Monday+1.
- It is projected that by Monday+1, 4205 will exceed 3973 in Unique.Visits by almost 1500.
- Page.Loads 4376 has the highest Total Returning.Visits but is ranked #5 in Total Unique.Visits.
- Page.Loads 4638 has the highest Total Unique.Visits but is ranked #3 in Total Returning.Visits.
- Over all values of Page.Loads and Day, the sum of Unique.Visits is almost 6.4 million.
- The summed values of Unique.Visits range from 667 to nearly 13
- thousand. For Unique.Visits, the most significant values of Day are Tuesday, Wednesday,

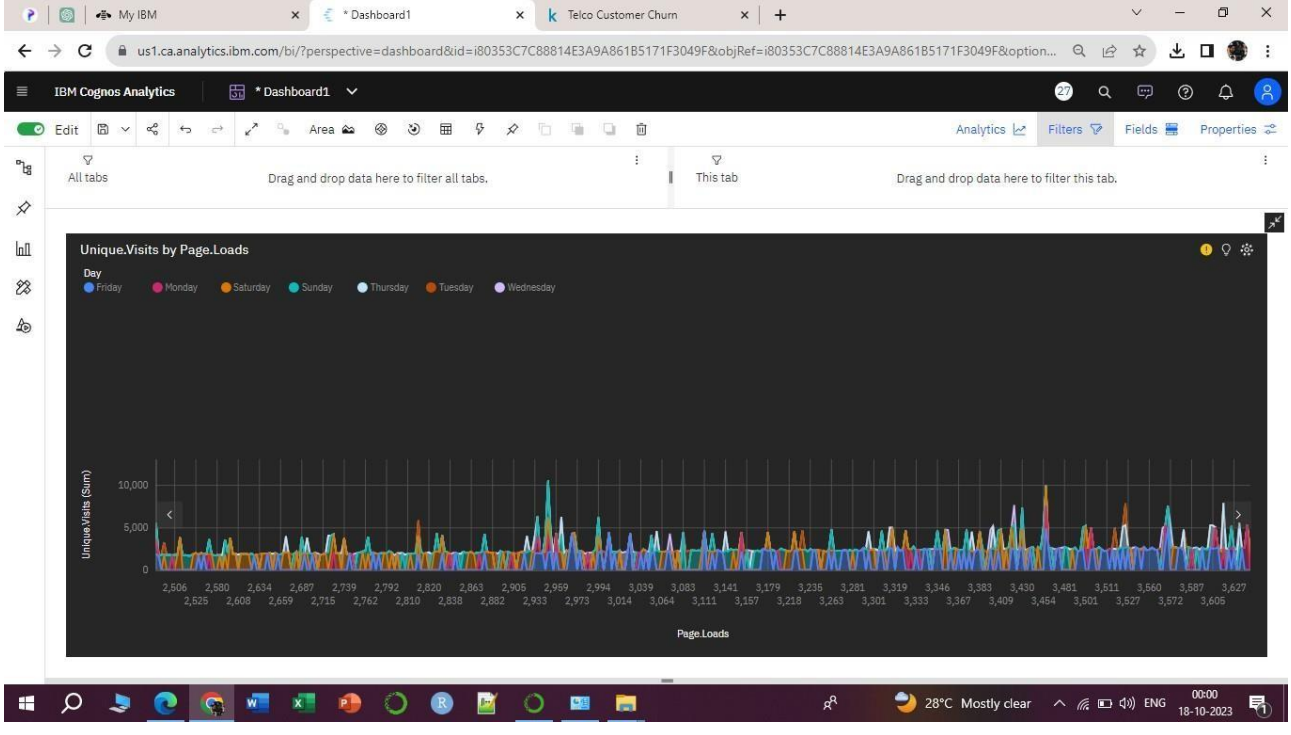


Monday, Thursday, and Friday, whose respective Unique.Visits values add up to over 5.1 million, or 80 % of the total.

First Time Visitors :

- Day.Of.Week 7 has the highest Unaggregated First.Time.Visits but is ranked #7 in Total Returning.Visits.
- Day Saturday has the highest Unaggregated First.Time.Visits but is ranked #7 in Total Returning.Visits.
- Day.Of.Week 3 has the highest Total Returning.Visits but is ranked #5 in Unaggregated First.Time.Visits.
- Day Tuesday has the highest Total Returning.Visits but is ranked #5 in Unaggregated First.Time.Visits.
- First.Time.Visits 3133 has the highest Total Returning.Visits but is ranked #2 in Total Page.Loads.
- First.Time.Visits 3146 has the highest Total Page.Loads but is ranked #2 in Total Returning.Visits.
- 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 First.Time.Visits values (57.2 % of the total).
- 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).
- Across all values of Day.Of.Week, the sum of Page.Loads is over 8.9
- million. The summed values of First.Time.Visits range from 0 to over 1500.
- The summed values of Page.Loads range from over a thousand to over fourteen thousand.
- For First.Time.Visits, the most significant values of Day.Of.Week are 3, 4, 2, 5, and 6, whose respective First.Time.Visits values add up to over 4.2 million, or
  79.9 % of the total.s
- 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.

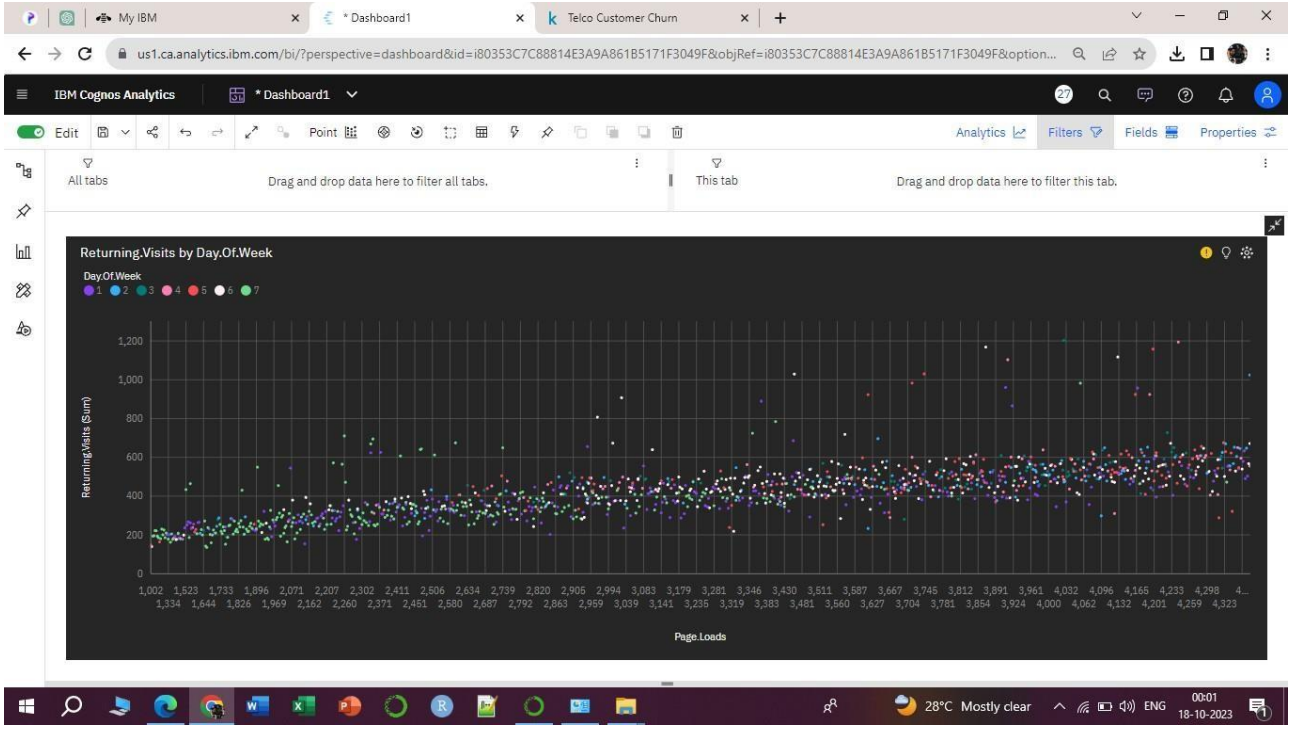
- For First.Time.Visits, the most significant values of Day are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective First.Time.Visits values add up to over 4.2 million, or 79.9 % of the total.
- For Page.Loads, the most significant values of Day are Tuesday, Wednesday, Monday, Thursday, and Friday, whose respective Page.Loads values add up to over 7.1 million, or 80.1 % of the total.



Returning Visit :

- Returning.Visits is unusually low when Day.Of.Week is 7.
- Based on the current forecasting, Returning.Visits may reach over 87 thousand by Day.Of.Week 9.
- It is projected that by 9, 4205 will exceed 3973 in Returning.Visits by 227.
- Across all values of Page.Loads and Day.Of.Week, the sum of Returning.Visits is over 1.1 million.
- The summed values of Returning.Visits range from 133 to over two thousand.

- For Returning.Visits, the most significant values of Day.Of.Week are 3, 4, 2, 5, and 6, whose respective



The Analysis using the Dashboard with the basic understanding gives the better understanding of visits that how to control the traffic and control the traffic on website with the help of scatter plot,heatmap and more.

3.2.2 **'Python Integration'** for Website Traffic Analysis

3.2.2.1 Role of Python in Data Analysis

Python plays a significant role in data analysis for a variety of reasons. It has become a popular and versatile programming language in the field of data analysis for the following key roles:

1. Data Manipulation: Python offers powerful libraries like Pandas, which provide data structures and functions for efficient data manipulation. Analysts can easily load, clean, transform, and filter data, making it suitable for analysis.

2. Data Visualization: Libraries like Matplotlib, Seaborn, and Plotly allow data analysts to create a wide range of charts and graphs to visually represent data. Python's visualization capabilities make it easier to communicate insights effectively.

3. Statistical Analysis: Python has libraries such as SciPy and StatsModels that provide tools for statistical analysis. Analysts can perform hypothesis testing, regression analysis, and other statistical procedures to draw meaningful conclusions from data.

4. Machine Learning: Python is a leading language for machine learning and artificial intelligence. Libraries like scikit-learn and TensorFlow enable data analysts to build and apply machine learning models to make predictions, classifications, and recommendations based on data.

5. Data Cleaning and Preprocessing: Python is useful for cleaning and preprocessing data. Analysts can handle missing data, remove duplicates, and transform data into a suitable format for analysis.

6. Integration with Data Sources: Python can connect to various data sources, including databases, web APIs, and flat files. This capability allows analysts to access and analyze data from different origins.

to perform specific data analyses and calculations tailored to their unique requirements.

7. Data Exploration: Python, along with Jupyter Notebooks, provides an interactive environment for data exploration. Analysts can document their analysis steps and collaborate with others effectively.

8. Big Data Analysis: Python can be used in conjunction with big data processing frameworks like Apache Spark to analyze large datasets efficiently.

9. Open Source Ecosystem: Python's open-source nature means that a vast community of developers and data scientists contribute to a rich ecosystem of libraries and tools for data analysis, making it a versatile and continually evolving platform.

10. Data Reporting: Analysts can use Python to generate reports, automate report generation, and create dynamic dashboards with libraries like Plotly and Bokeh.

11. Accessibility and Ease of Learning: Python's clear and concise syntax makes it accessible to individuals with varying levels of programming experience. It is often considered a user-friendly language for data analysis.

The combination of these factors makes Python a preferred choice for data analysis across various domains, from business intelligence and finance to healthcare and scientific research. Its extensive ecosystem, coupled with its ease of use and flexibility, makes it a powerful tool for extracting insights and value from data.

Installing packages and modules in Python is a common and straightforward process. If you're working on a data analysis project and need specific packages or modules, here's how you can install and manage them:

1. Using pip:

[Pip](https://pypi.org/project/pip/):

- To install a package, open your command-line terminal and run:
  ```
  pip install package_name
  ```
  Replace `package_name` with the name of the package you want to install.

- To install a specific version of a package, specify the version number after the package name:
  ```
  pip install package_name==version_number
  ```

2. Managing Dependencies:

When you install a package, pip will automatically handle dependencies, which are other packages that the package you're installing relies on. Pip will download and install these dependencies for you.

3. Virtual Environments:

It's a good practice to create a virtual environment for your project. Virtual environments isolate your project's dependencies from the system-wide Python installation. Here's how you can create and activate a virtual environment:

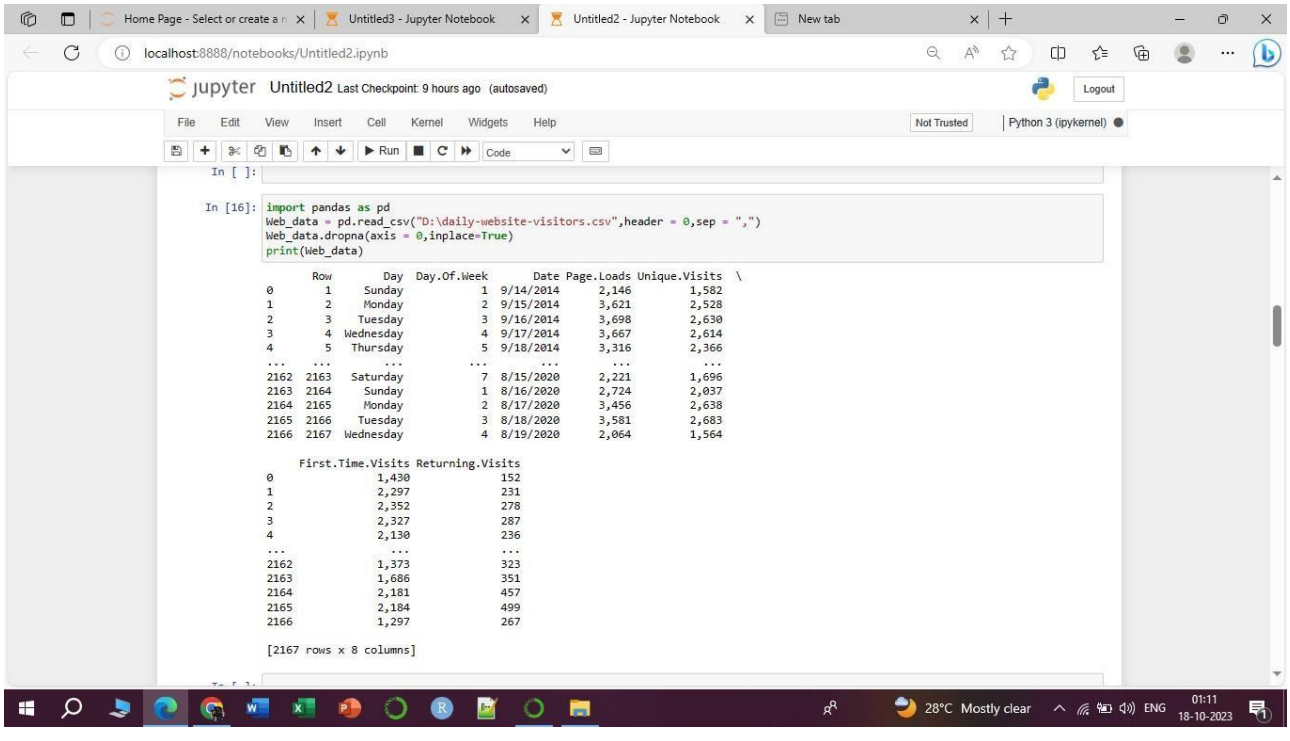
- Create a virtual environment:
  ```
  python -m venv myenv
  ```
  Replace `myenv` with the name you want to give your virtual environment.

- Activate the virtual environment:
  - On Windows:
    ```
    myenv\Scripts\activate
    ```
  - On macOS and Linux:
    ```
    ```

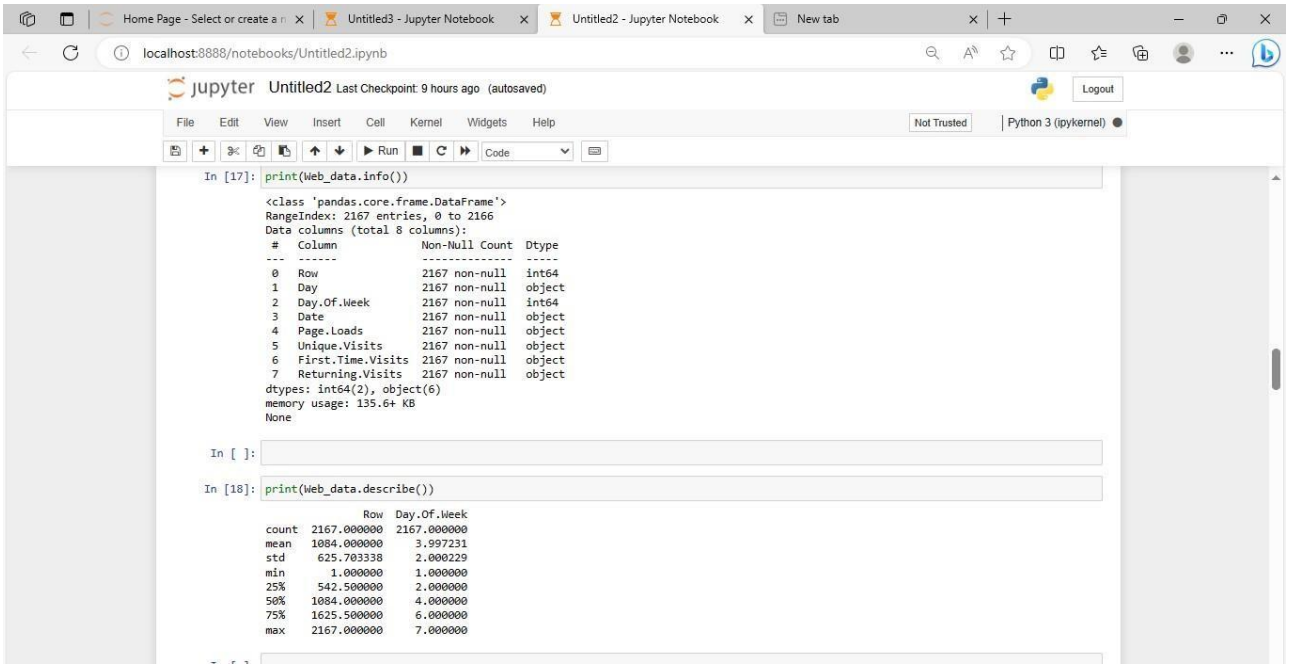
```

- Once activated, you can use pip to install packages, and they will be isolated within your virtual environment.

4. Managing Project Dependencies:

To manage project dependencies, it's a good practice to create a `requirements.txt` file that lists all the packages your project depends on. You can generate this file with the following command:

```
pip freeze > requirements.txt
```

To install the dependencies listed in the `requirements.txt` file on a new system or environment, you can use the following command:

```
pip install -r requirements.txt
```

5. Common Data Analysis Packages:

For a data analysis project, you might need packages like Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, Jupyter, and others. Install them with pip as needed:

```
pip install pandas numpy matplotlib seaborn scikit-learn jupyter
```

Make sure you have a clear understanding of the specific packages you need for your data analysis project, and install them as required. The `pip` package manager makes it easy to bring in the necessary libraries and modules to work on your project.
```

3.2.2.2 Analyzing Data with Python

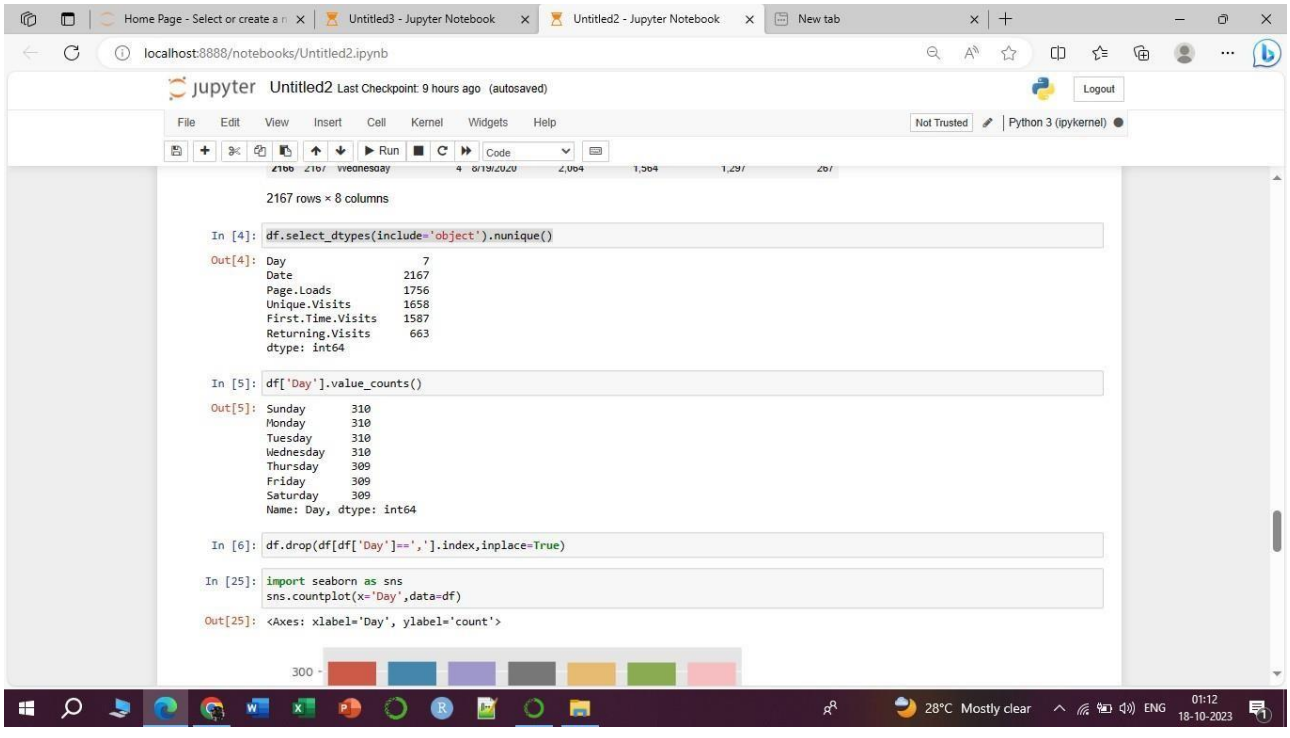
Extraction the Data file from the directory to the python text editor to excecute the data set .



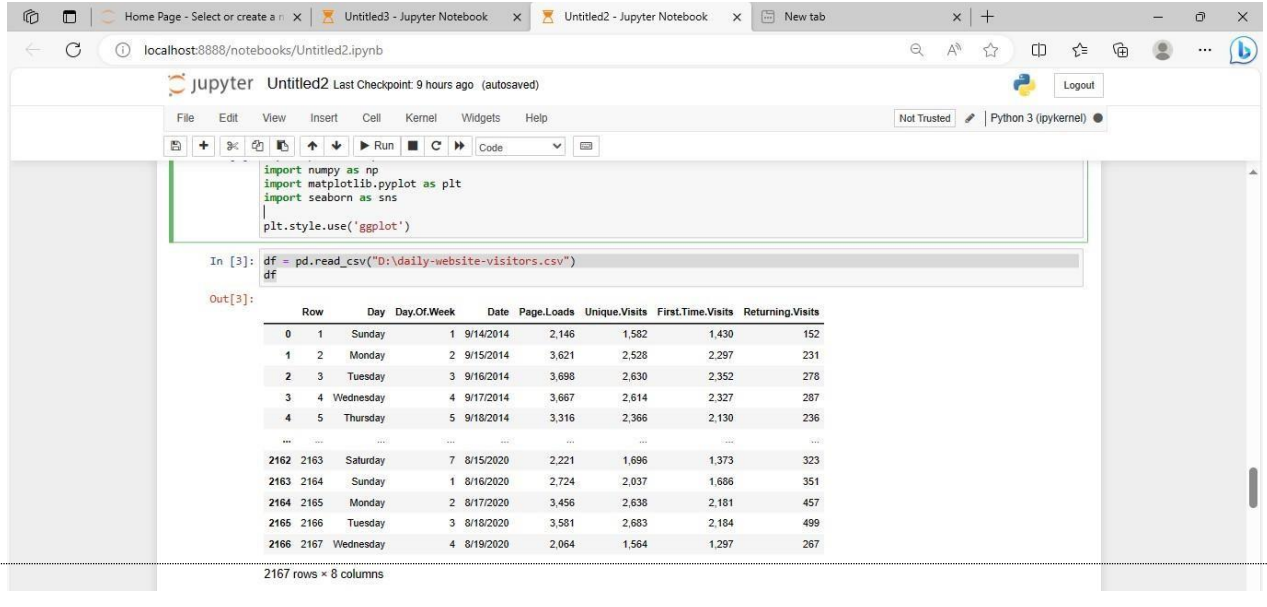
Getting of data set information using info function.


Home Page - Select or create a n
Untitled3 - Jupyter Notebook
Untitled2 - Jupyter Notebook
New tab
localhost:8888/notebooks/Untitled2.ipynb
jupyter Untitled2 Last Checkpoint: 9 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [17]: print(Web_data.info())
<class 'pandas.core.frame.DataFrame'>
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
None
In [ ]:
In [18]: print(Web_data.describe())
Row Day.Of.Week
count 2167.000000 2167.000000
mean 1084.000000 3.997231
std 625.703338 2.000229
min 1.000000 1.000000
25% 542.500000 2.000000
50% 1084.000000 4.000000
75% 1625.500000 6.000000
max 2167.000000 7.000000

Value Counts of each Insights of the data set content and Object Integration of the data set

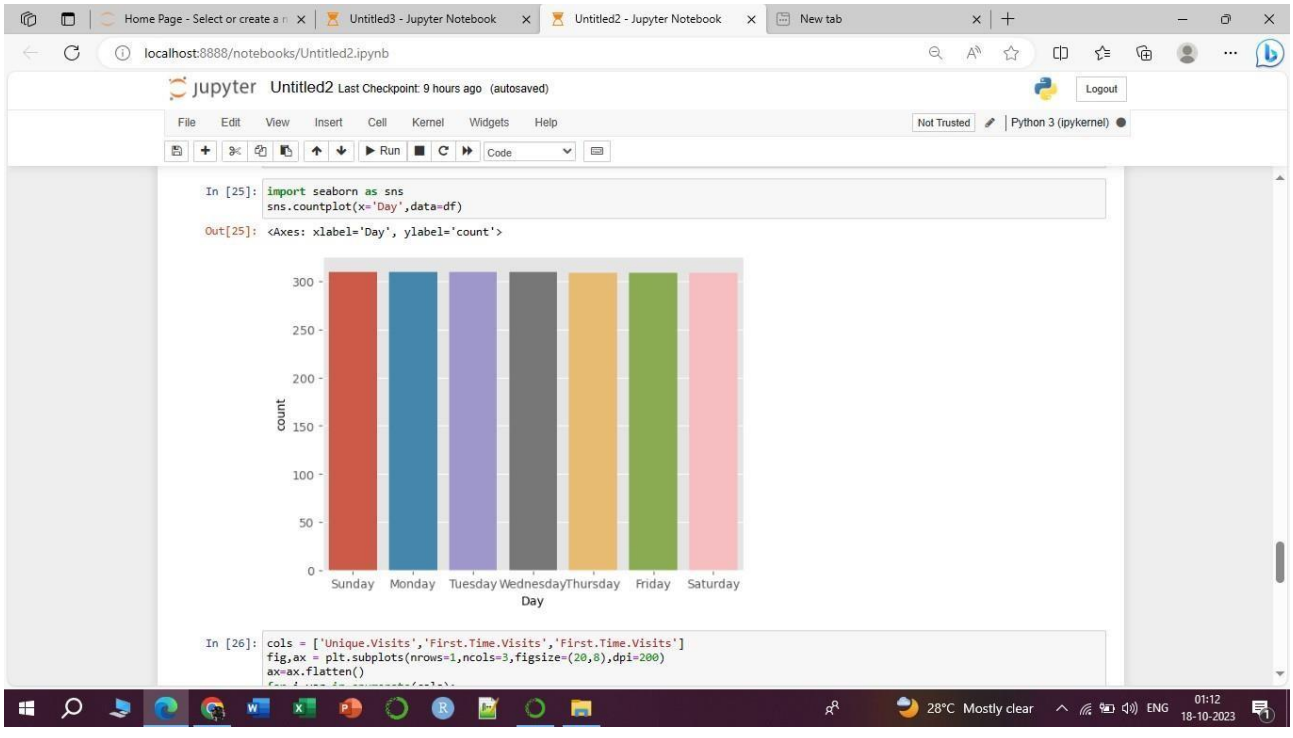


Again pointing out the data set with the help of Pandas Library.

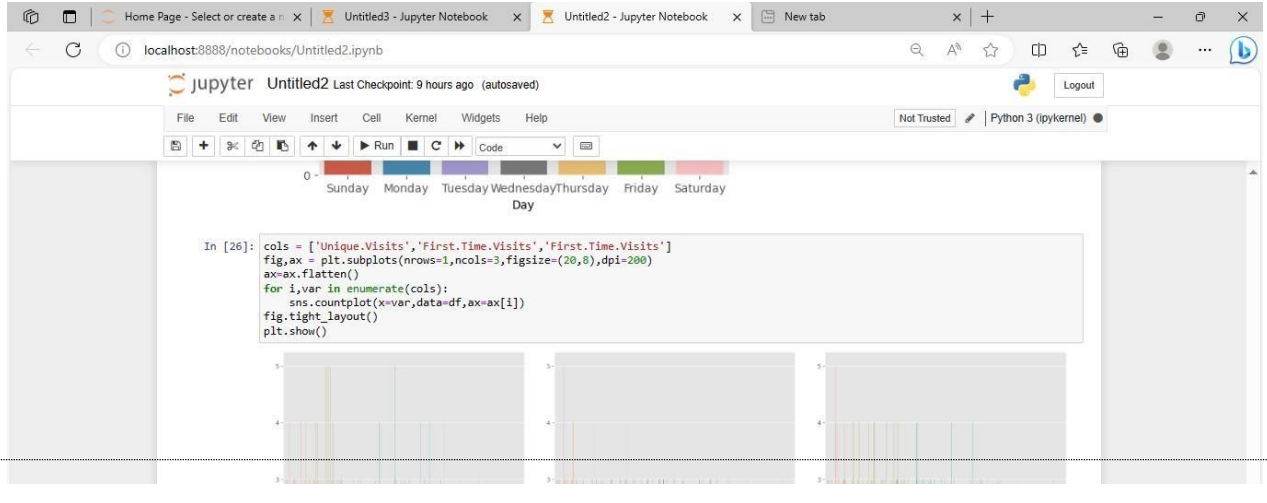


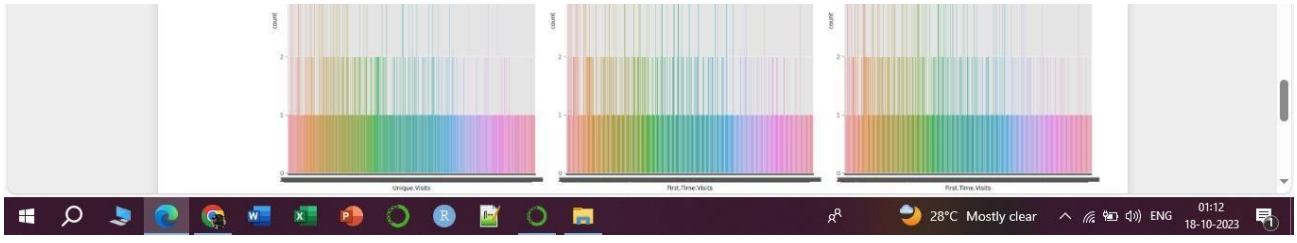
```
In [4]: df.select_dtypes(include='object').nunique()
Out[4]: Day              7
        Date          2167
        Page.Loads    1756
```

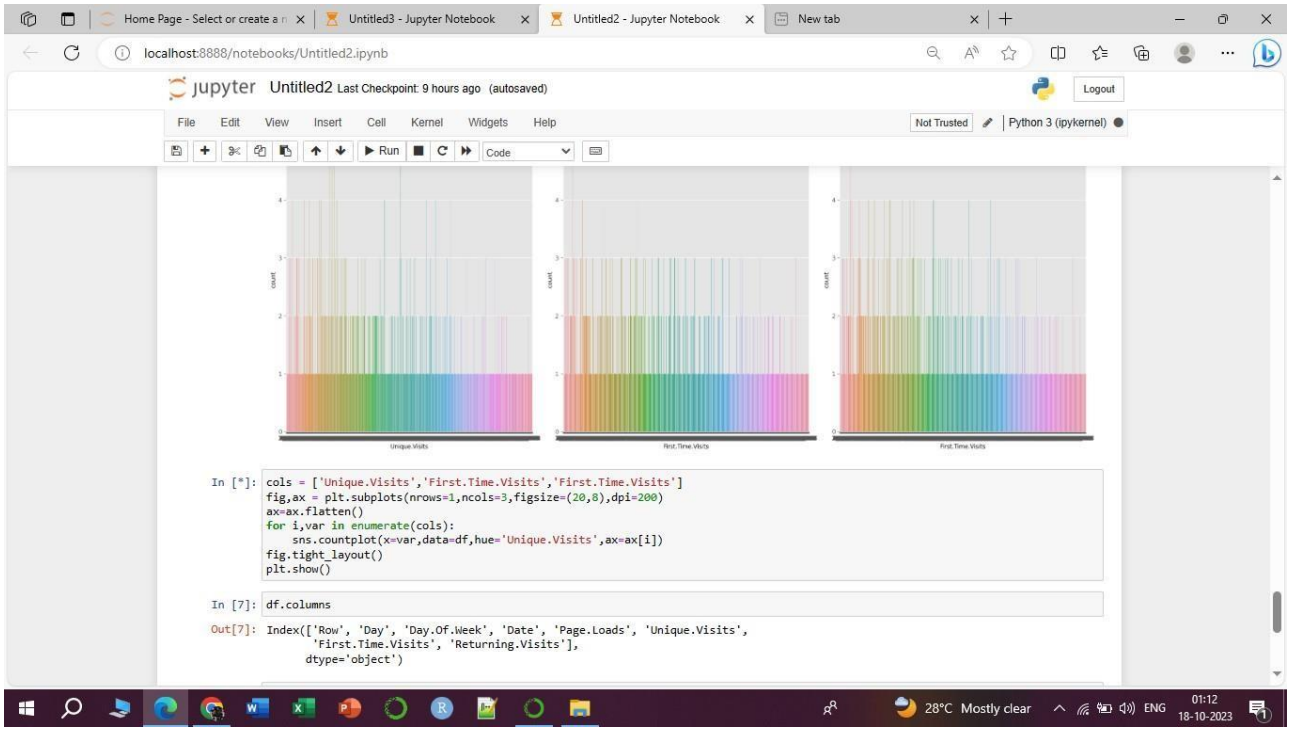
Listing the Number of Days that had been observed by the data set using Seaborn Library.



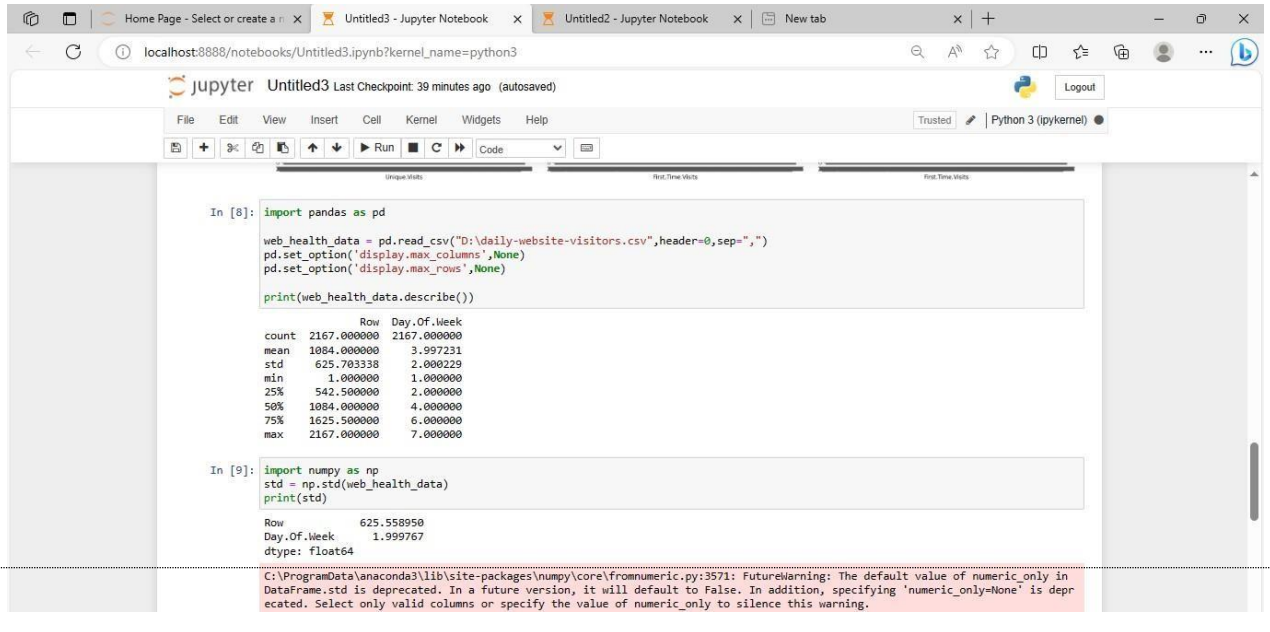
Plotting of Three important Stuffs like Unique visitor,First Time Visitor,Returning Visitor.

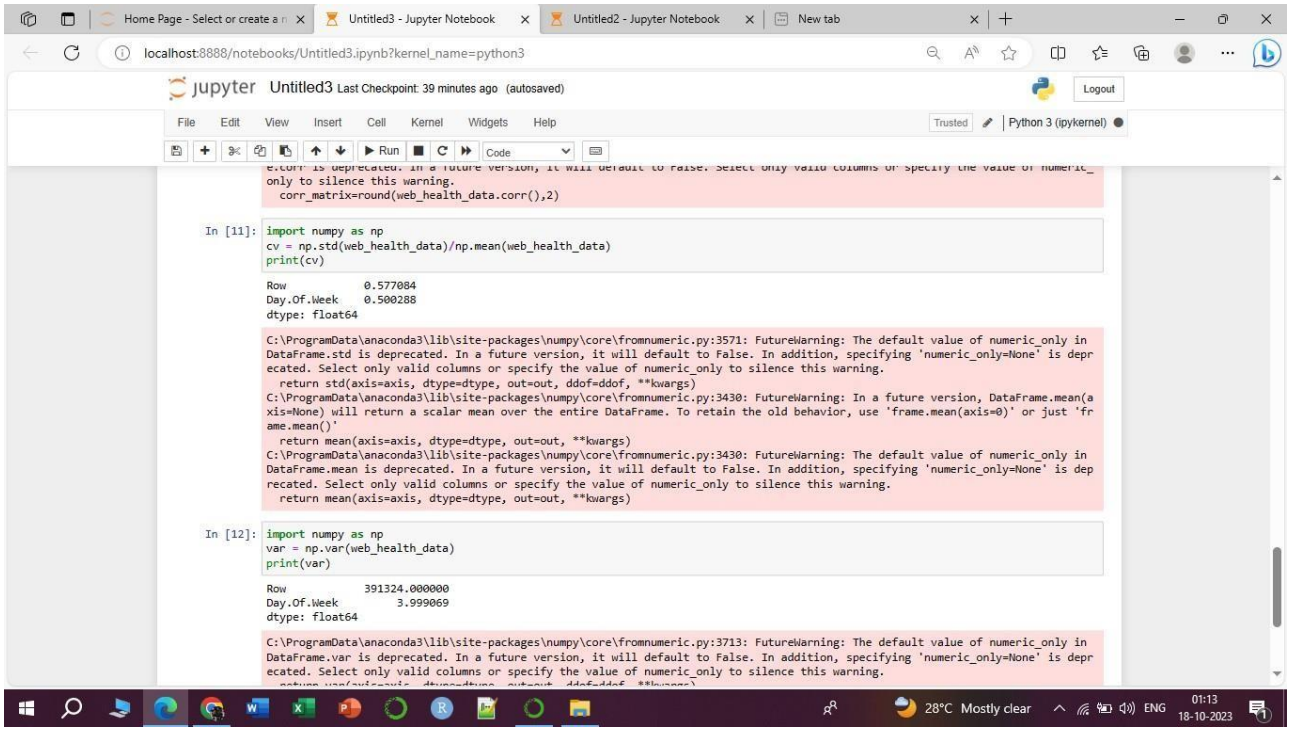




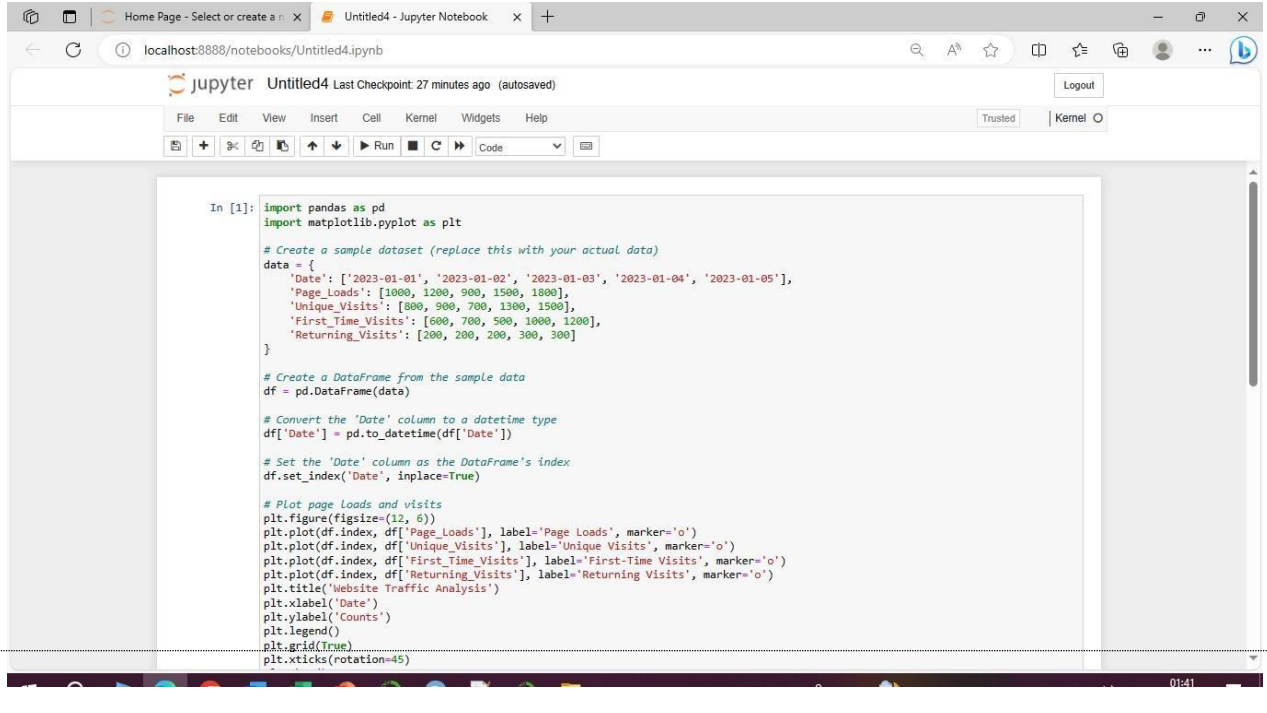


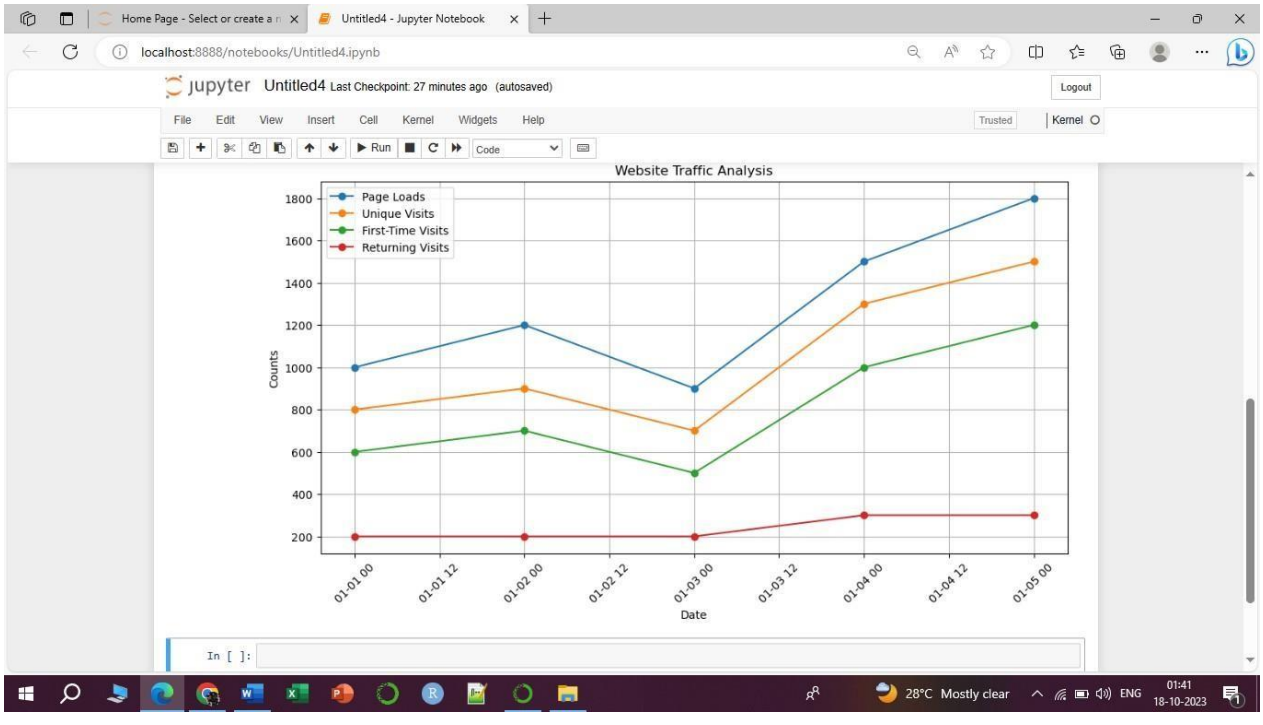
Mathematical Calculations of the data set with the help of Numpy Library.





An Basic Observation with the Limited Data points and to ploltting in the Line Graph.



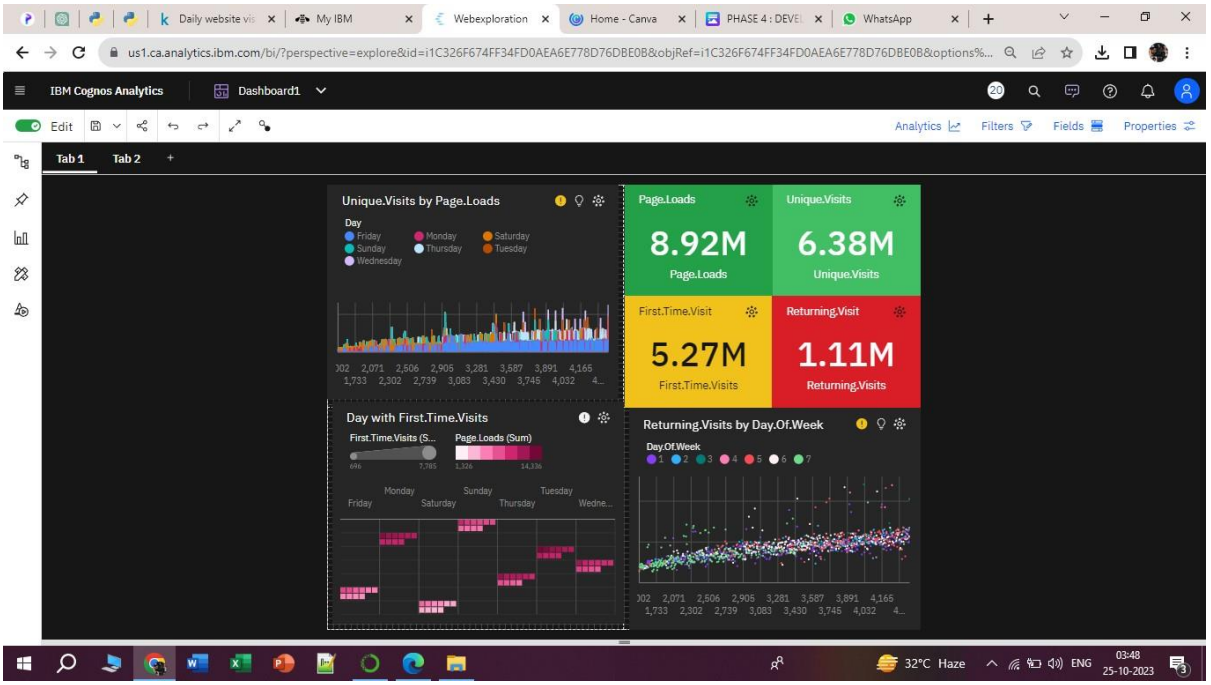

Home Page - Select or create a n
Untitled4 - Jupyter Notebook
localhost:8888/notebooks/Untitled4.ipynb
jupyter Untitled4 Last Checkpoint: 27 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Kernel
Run
Code
Website Traffic Analysis
Page Loads
Unique Visits
First-Time Visits
Returning Visits
Counts
Date
01-01 00
01-01 12
01-02 00
01-02 12
01-03 00
01-03 12
01-04 00
01-04 12
01-05 00
In [ ]:
28°C Mostly clear
ENG
01:41
18-10-2023

3.3 Phase 4: Development Part 2

IBM Cognos

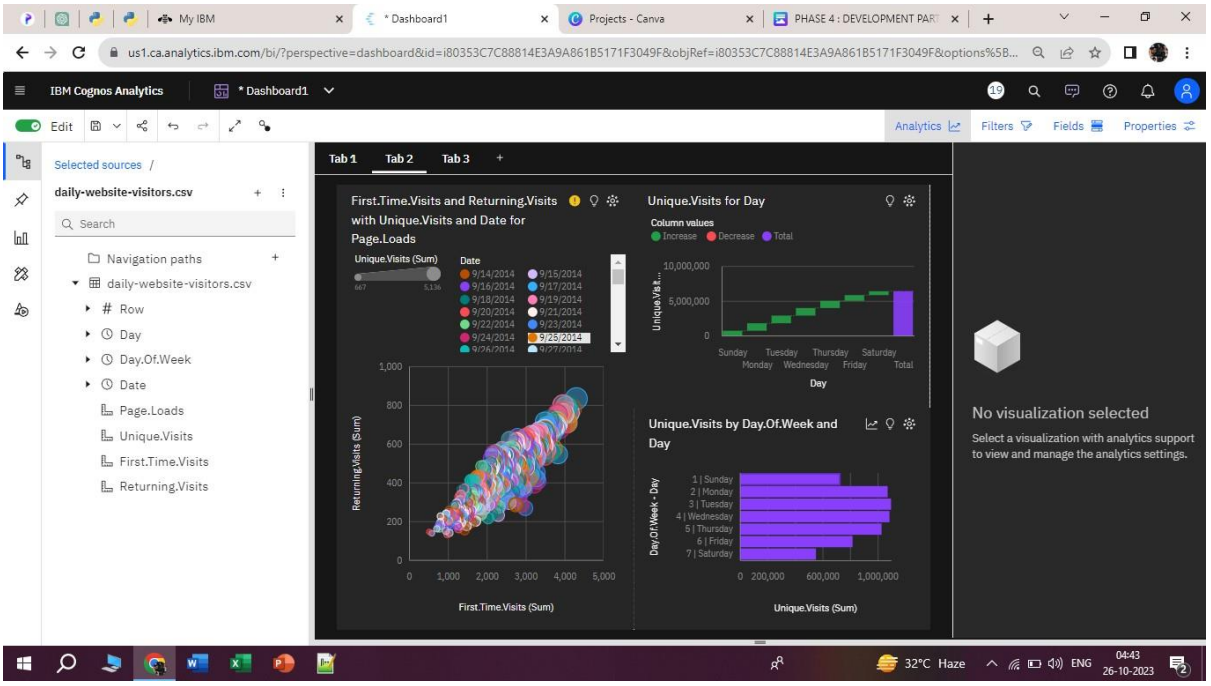
3.3.1 Building Dashboards and Reports

Data Exploration



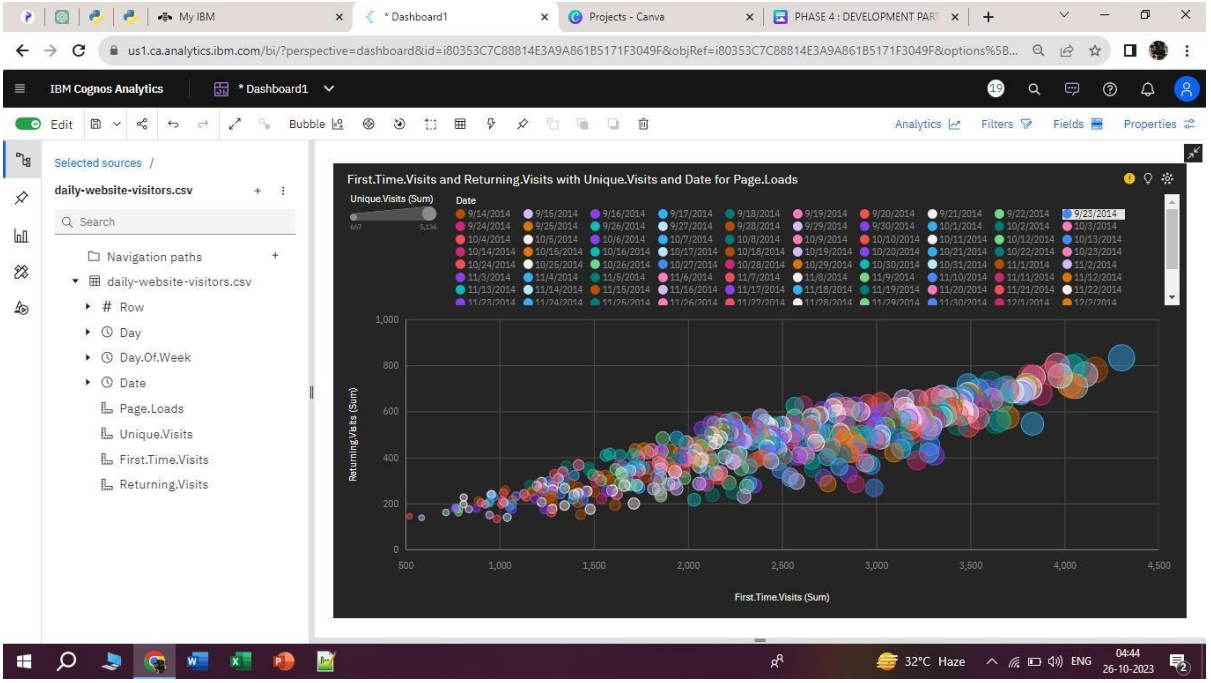
On Phase 2 Development part 1, made only the Data set 'Analysis visualization' of overall Page loads, Unique visits and more.

From the Continuation of Phase 3 development part 1 improved the analysis in a advanced analysis to improve the prediction and to show the visualization.

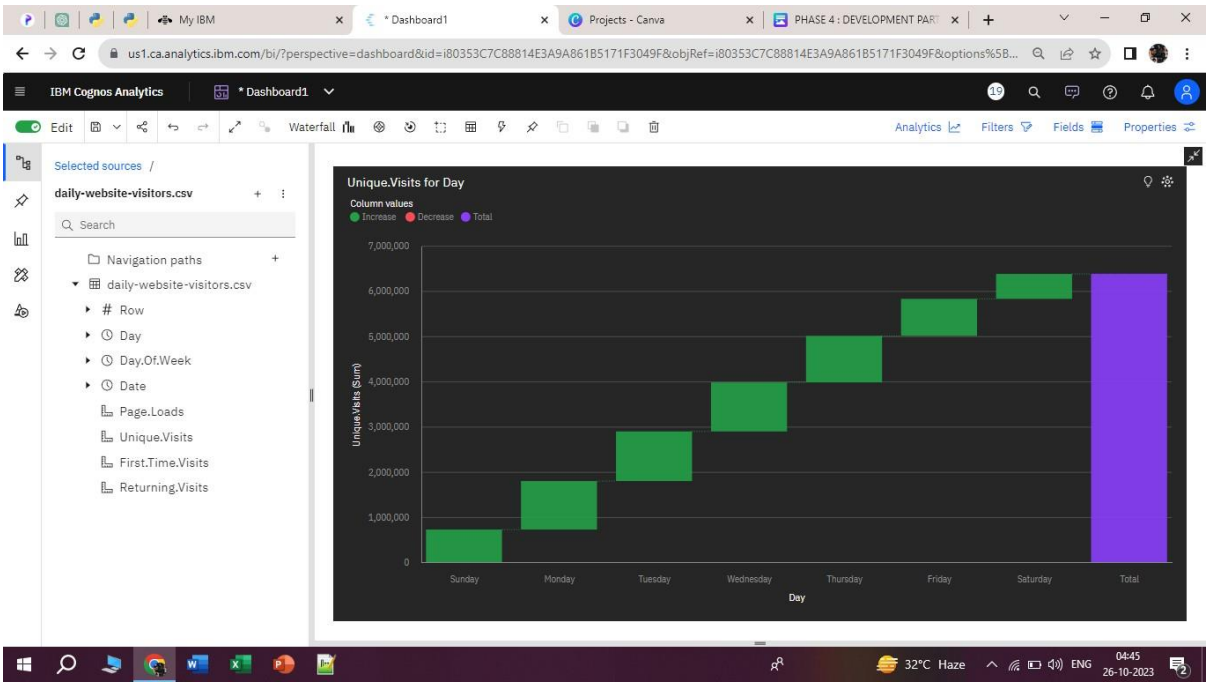


Insights :

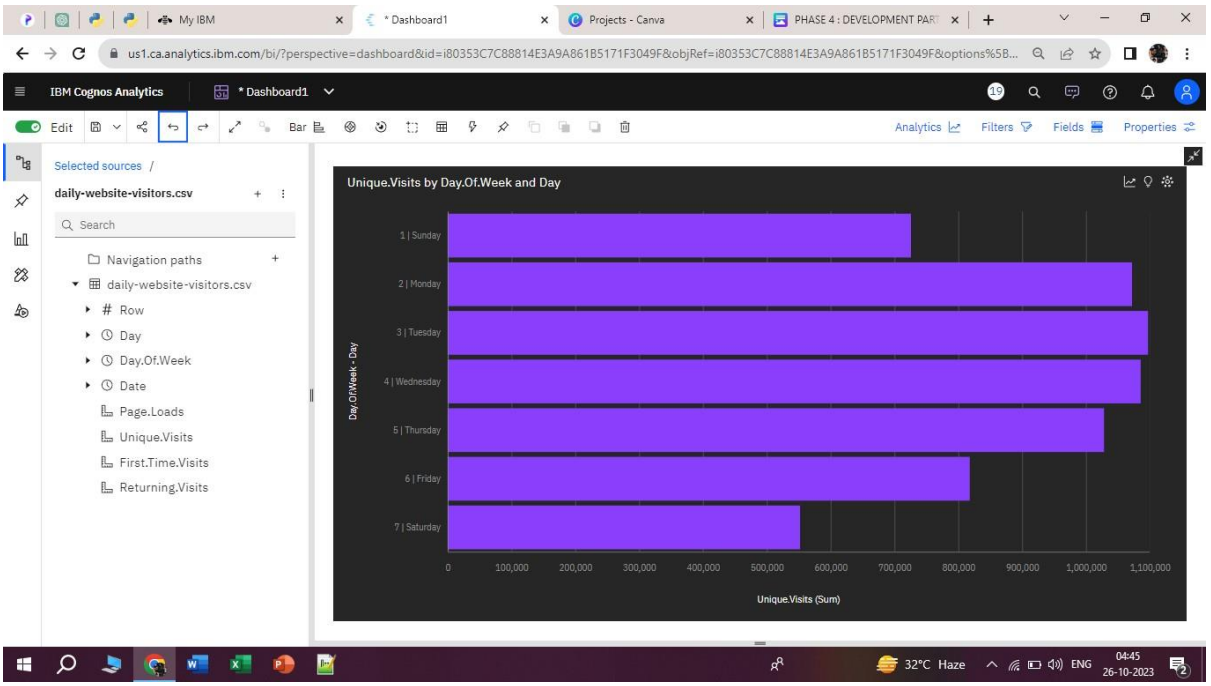
- Based on the current forecasting, **Page.Loads** may reach **nearly four thousand** by **Date 2021-10-27**
- **Page.Loads** has a strong weekly trend. The largest values typically occur on **Tuesday**, whereas the smalest values on **Saturday**.
- Over all **dates**, the average of **Unique.Visits** is **nearly three thousand**.
- Over all **dates**, the average of **First.Time.Visits** is **almost 2500**.



- First time visitors, Unique visitors, Returning Visitors are get Varying day-by-day on the daily basis based on the performance and expericence of the Website so, we can't able to judge the performance and visitors user exeperience.

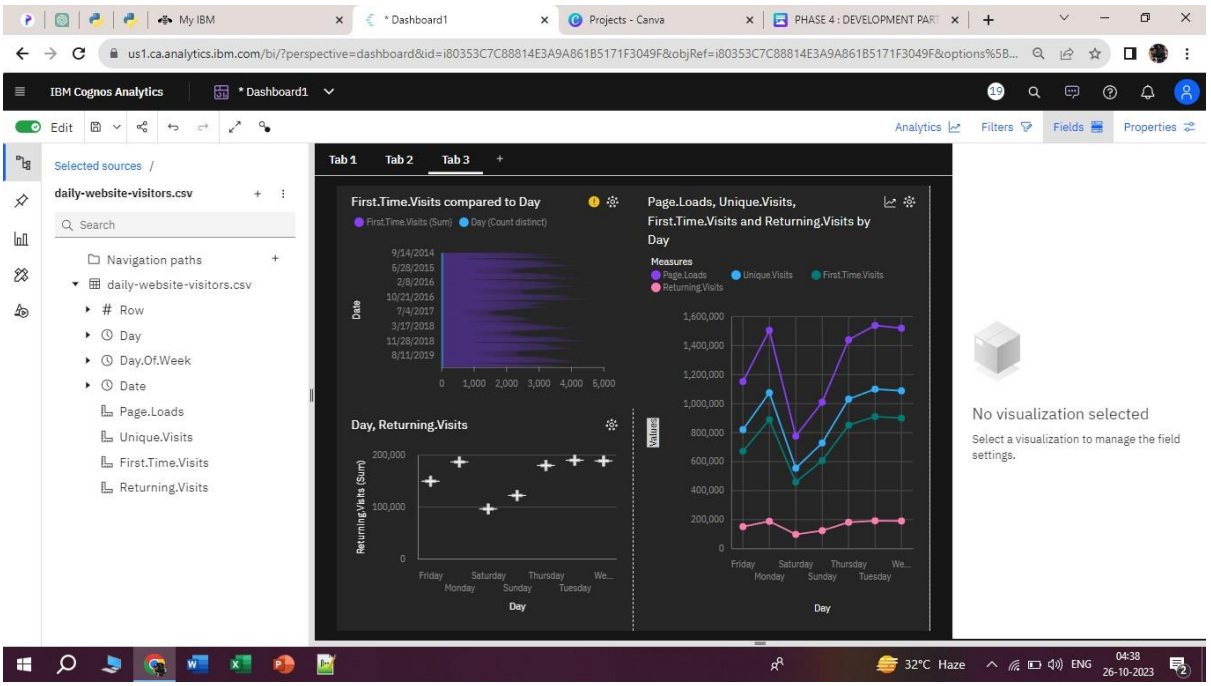


- Unique Visitors are the Daily and loyal visitors .who use the platform for they professional use with there brain storm addiciton for the website.



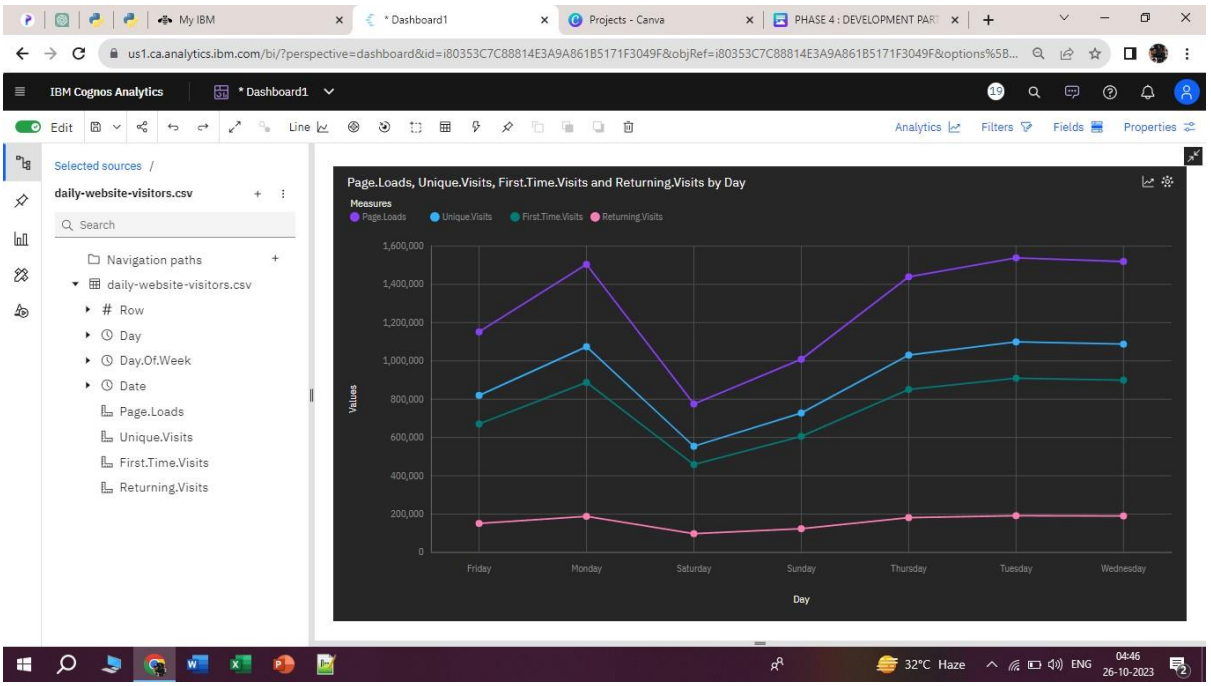
- There is major difference in week days and week ends visitors got varied in there need.

- When compared with the week ends the professionals are using the platform on there work basis on week days.



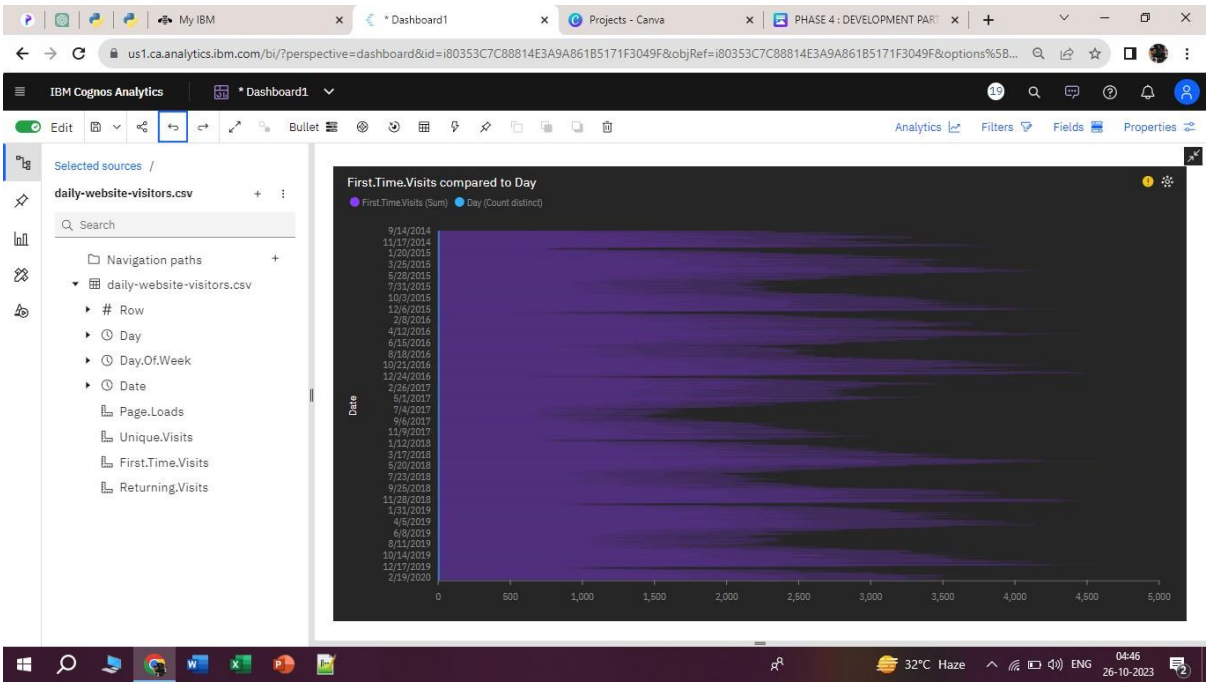
Insights :

- Over all **dates**, the average of **Returning.Visits** is **511.8**.
- Across all **dates**, the average of **Page.Loads** is **over four thousand**.
- The total number of results for **First.Time.Visits**, across all **dates**, is **over two thousand**.
- The total number of results for **Page.Loads**, across all **dates**, is **over two thousand**.



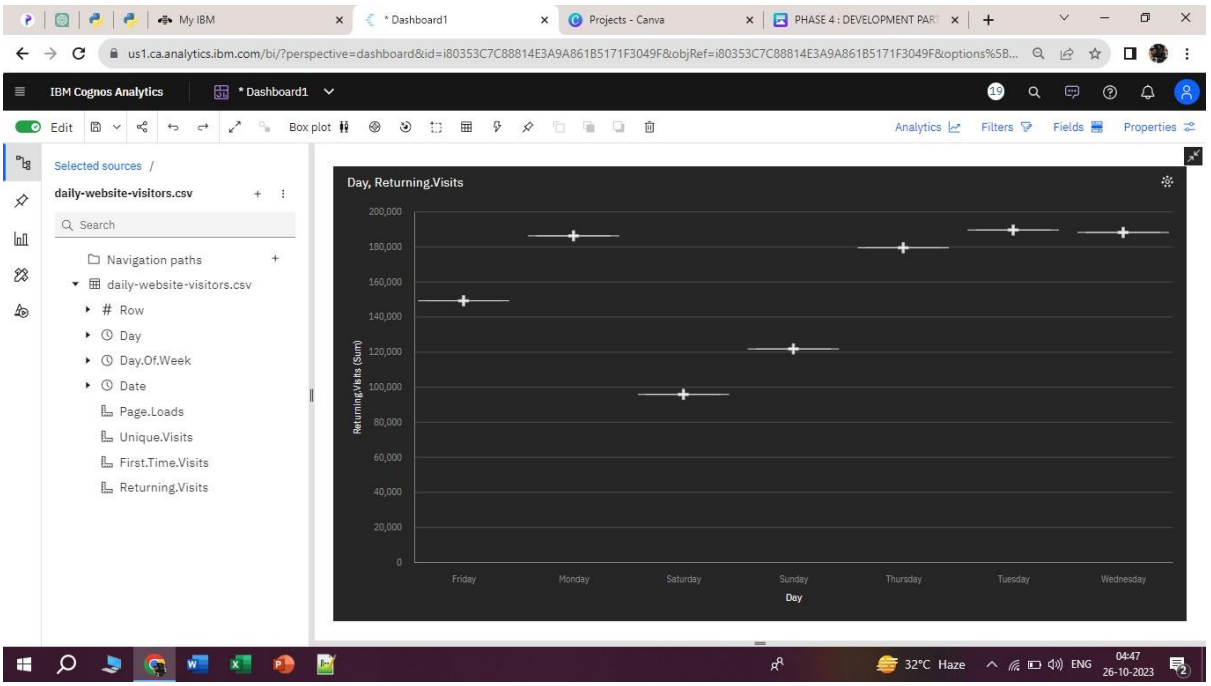
- Here You can go with Page loads got Analysed in comparison with the First time, Returning Visits, Unique visits .

- Loads are more and Traffic is more on starting of week days like Monday and end week days like Friday.



- First visitors are the new visitors for there necessity ,Product improvement and more.

- First visitors have the seasonal visit traffic on the above graph with the respective months and days.



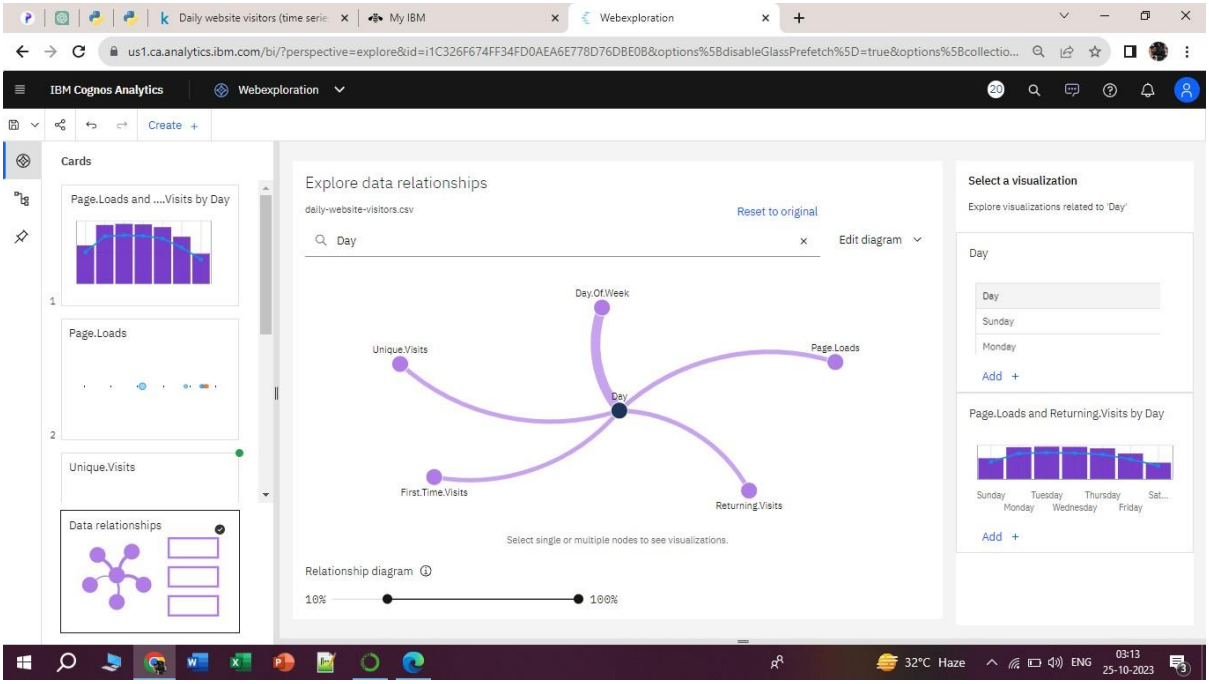
- First visitors are the new visitors for there necessity ,Product improvement and more.

- First visitors have the seasonal visit traffic on the above graph with the respective months and days.

Report

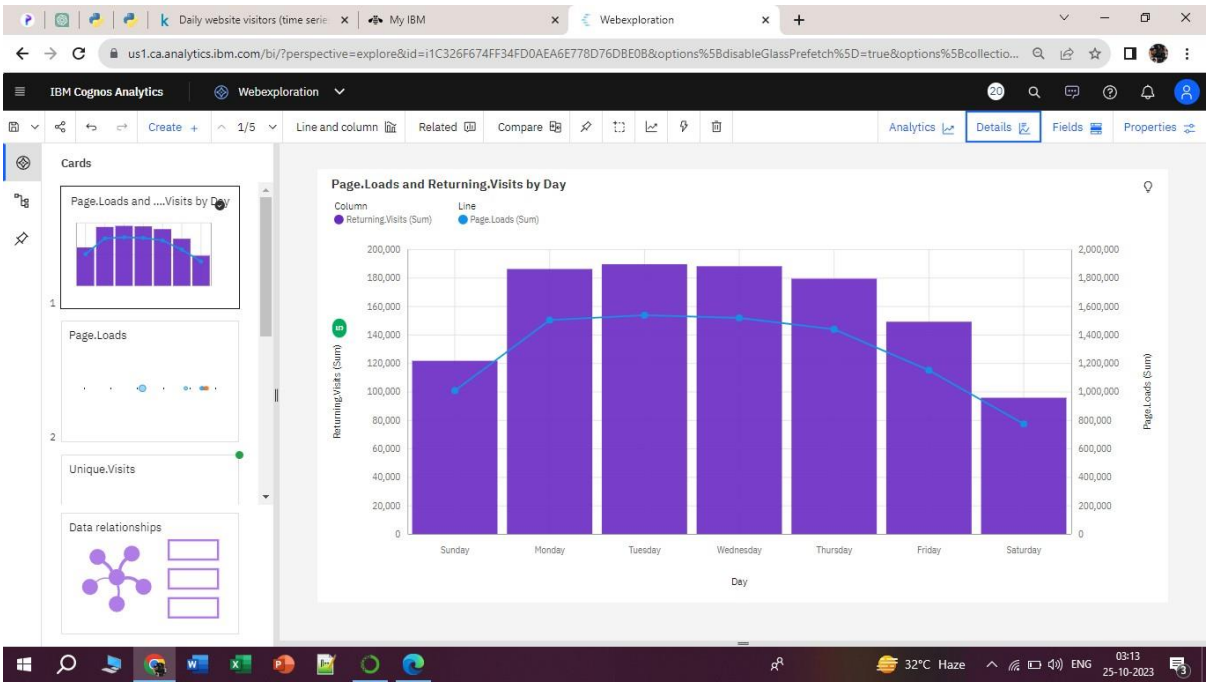
3.3.1.2 Visualization Techniques

Segmentation



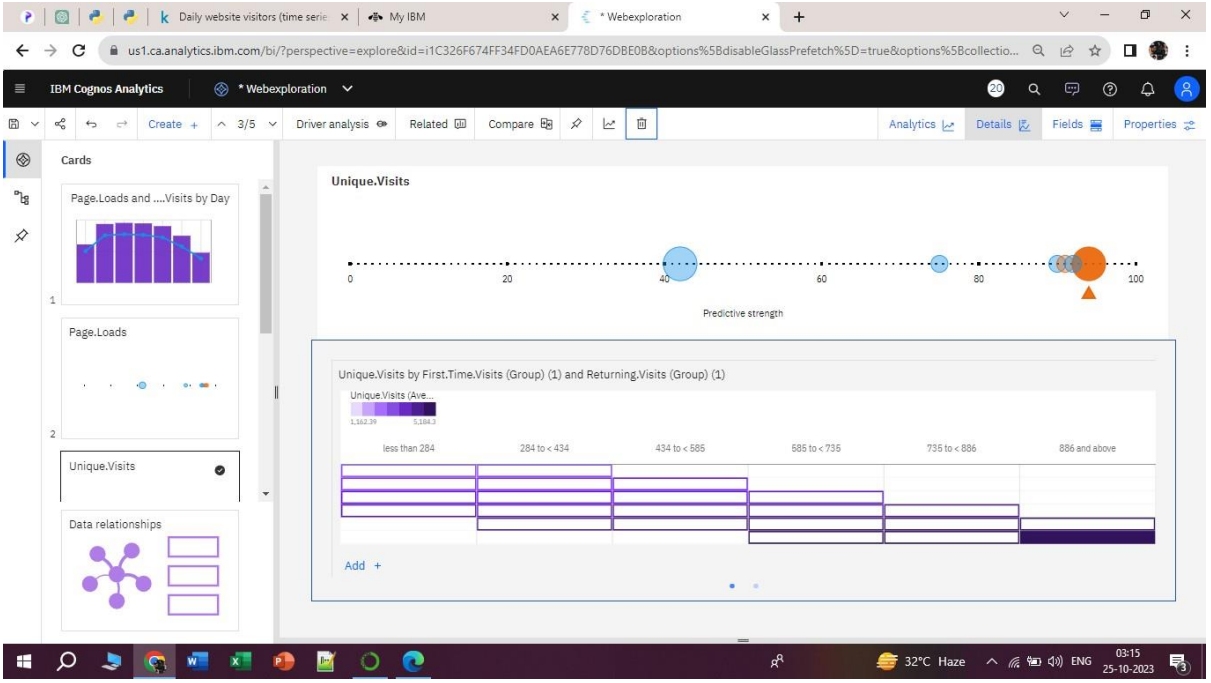
- Every insights are interlinked with the Day ,Here the Day is the major source which interlinks all the corresponding respectives.

- without any collapse and congestion every insights are linked with Day to show the Website Traffic.

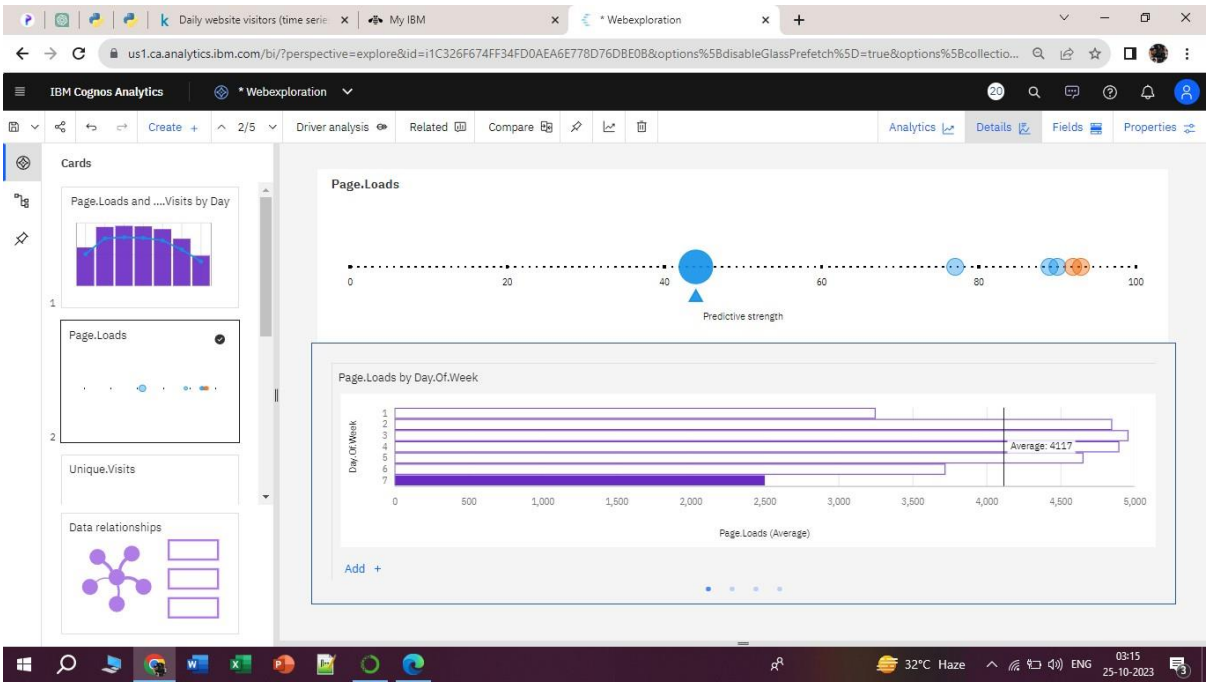


Insights :

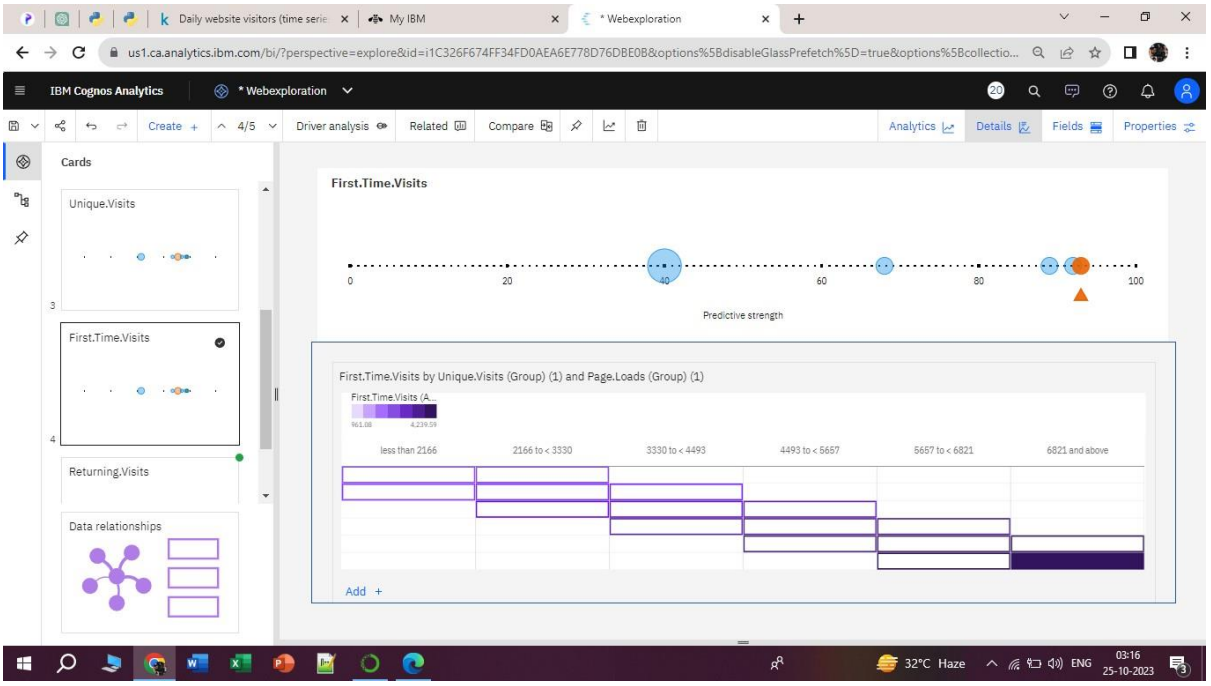
- Across all **days**, the sum of **Returning.Visits** is over 1.1 million.
- **Returning.Visits** ranges from almost 96 thousand, when **Day** is Saturday, to over 189 thousand, when **Day** is Tuesday.
- **Returning.Visits** is unusually low when **Day** is Saturday.
- For **Returning.Visits**, the 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.
- Across all **days**, the sum of **Page.Loads** is over 8.9 million.
- **Page.Loads** ranges from nearly 773 thousand, when **Day** is Saturday, to over 1.5 million, when **Day** is Tuesday.
- **Page.Loads** is unusually low when **Day** is Saturday.



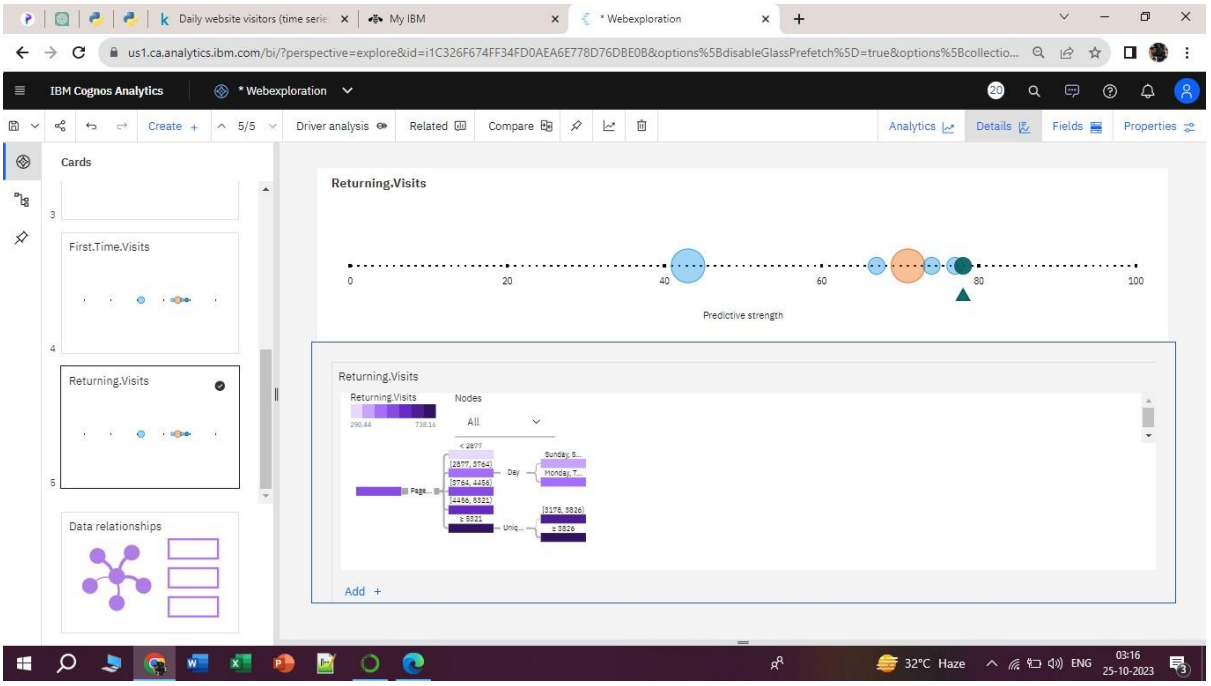
- **First.Time.Visits (Group) (3)** strongly affects **Unique.Visits** (94%).
- **Unique.Visits** is most unusual when **First.Time.Visits (Group) (3)** is 3934 and above and less than 1205.
- **Returning.Visits (Group) (2)** strongly affects **Unique.Visits** (76%).
- **Unique.Visits** is unusually high when **Returning.Visits (Group) (2)** is 886 and above.
- Over all values of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)**, the average of **Unique.Visits** is nearly three thousand.
- The average values of **Unique.Visits** range from over a thousand to over five thousand.
- **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** strongly affect **Unique.Visits** (96%).
- **Unique.Visits** is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.
- 1887 to < 2569 is the most frequently occurring category of **First.Time.Visits (Group) (3)** with a count of 666 items with **Unique.Visits** values (30.7 % of the total).
- 434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).
- There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.



- Across all values of **Day.Of.Week**, the average of **Page.Loads** is over four thousand.
- The average values of **Page.Loads** range from over 2500, occurring when **Day.Of.Week** is 7, to nearly five thousand, when **Day.Of.Week** is 3.
- **Day.Of.Week** moderately affects **Page.Loads** (44%).
- **Page.Loads** is unusually low when **Day.Of.Week** is 7.
- 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).



- **Unique.Visits** is unusually high when the combination of **First.Time.Visits (Group) (3)** and **Returning.Visits (Group) (2)** is 3934 and above and 886 and above.
- 1887 to < 2569 is the most frequently occurring category of **First.Time.Visits (Group) (3)** with a count of 666 items with **Unique.Visits** values (30.7 % of the total).
- 434 to < 585 is the most frequently occurring category of **Returning.Visits (Group) (2)** with a count of 734 items with **Unique.Visits** values (33.9 % of the total).
- There is no significant impact of **Returning.Visits (Group) (2)** on the relationship between **First.Time.Visits (Group) (3)** and **Unique.Visits**.



- **Page.Loads**, **Unique.Visits**, and **Day** predict **Returning.Visits** with a strength of 78.1%.

- **Page.Loads** is the most significant predictor of **Returning.Visits** being three times better than any other field.

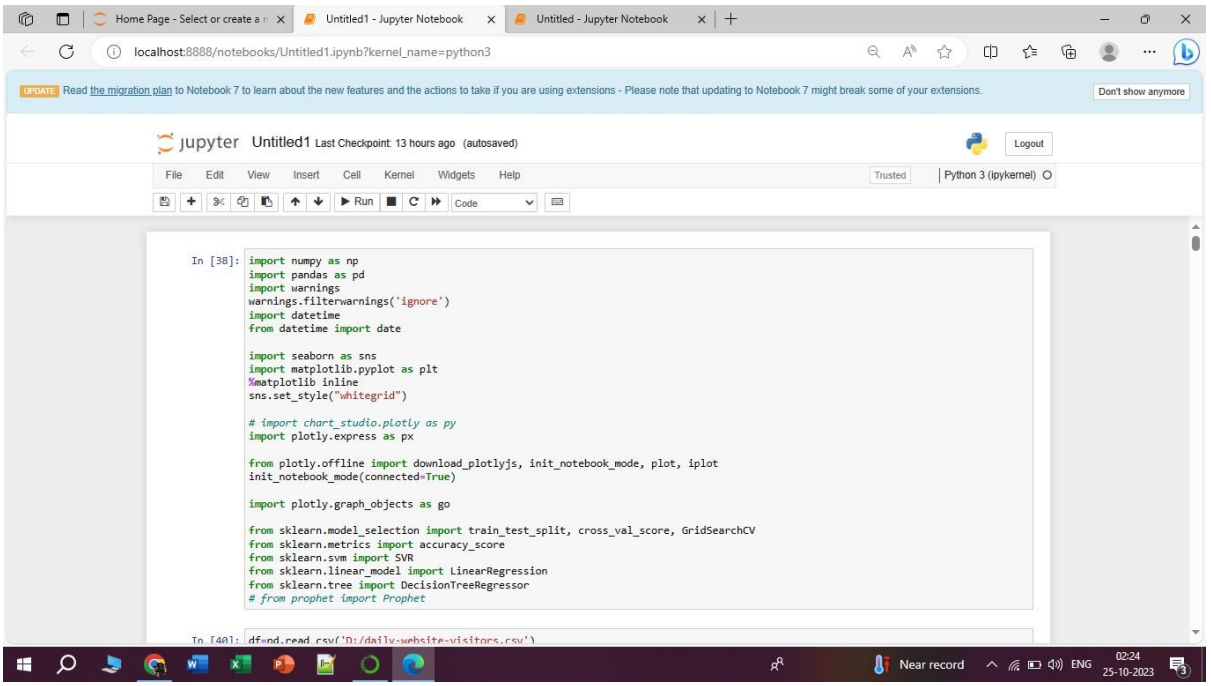
3.3.2 Python Integration

3.3.2.1 Advances Analyzing Data with Python

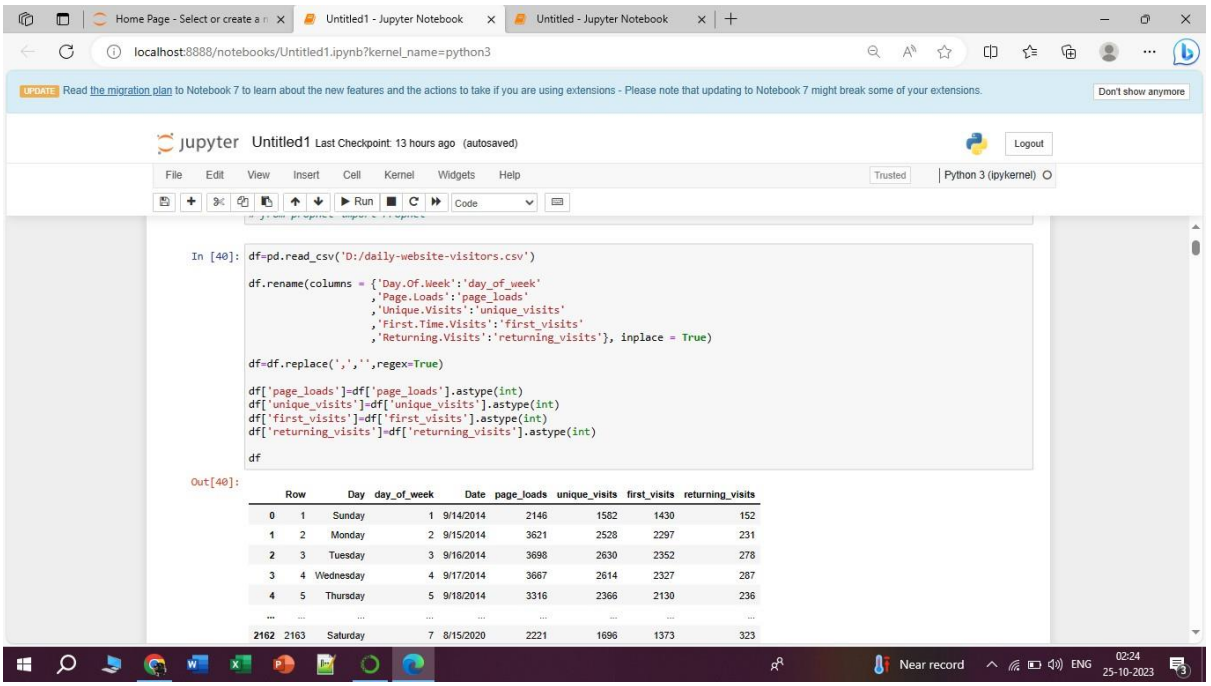
Under Python Integration Part is to make the analysis of traffic to make better and predictive with advanced techniques.

Using popular Library Modules

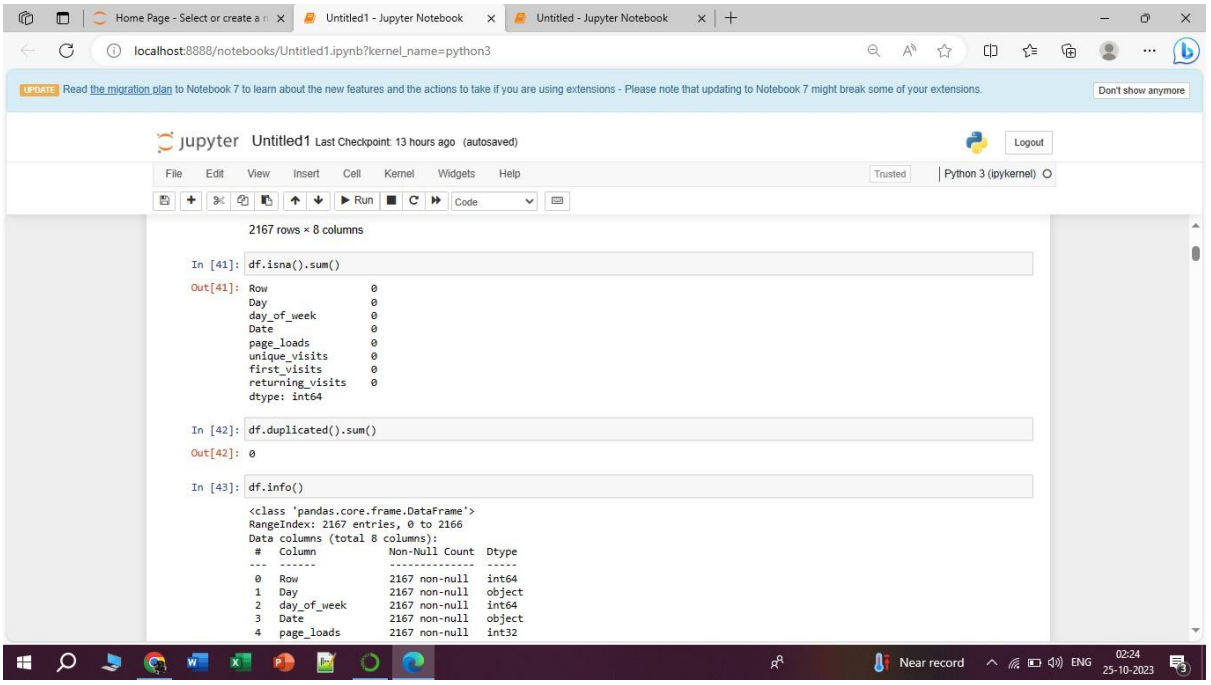
1. Numpy
2. Pandas
3. Matplot
4. Scipy
5. Seaborn and more for visualization ,analysis.



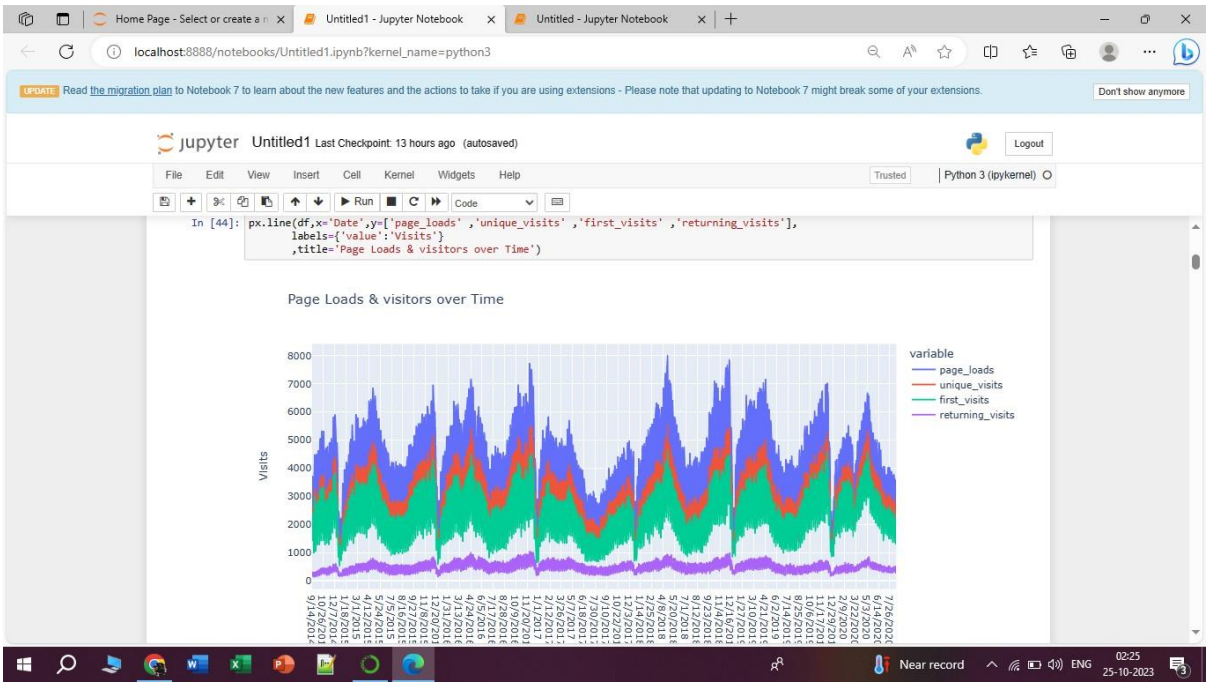
Importing the necessary modules to make the analysis better to know the work flow of traffic .



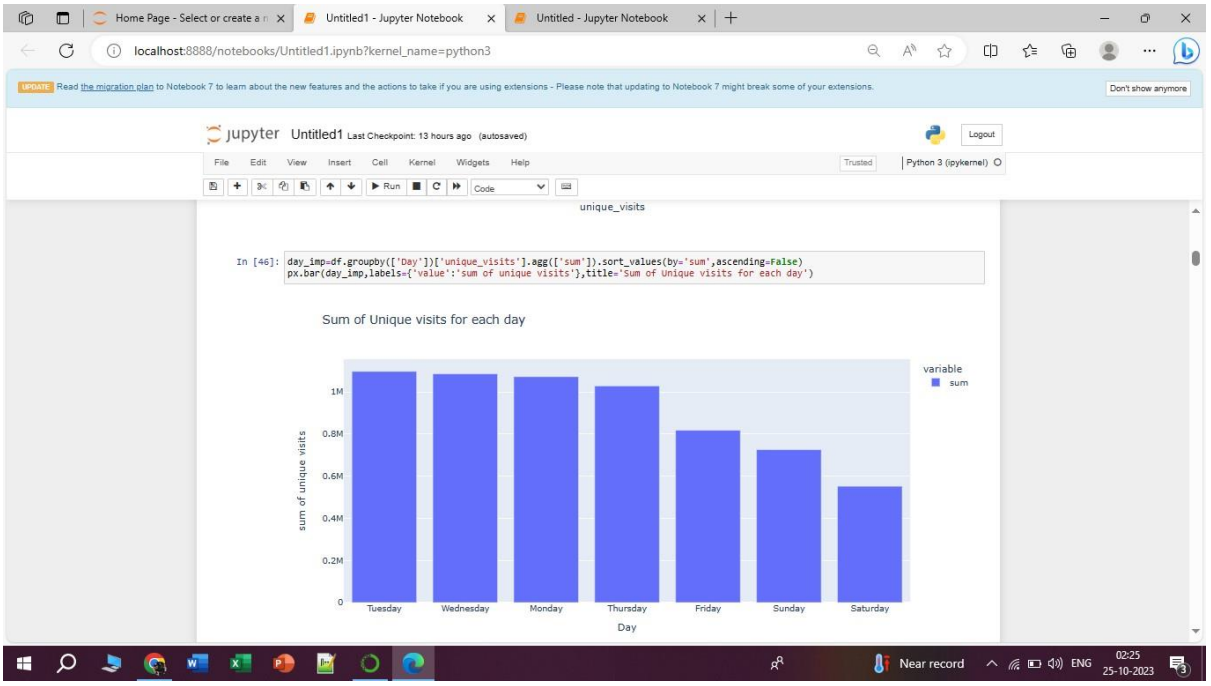
Printing the Insights like Rows, Columns, Datas inside the Data sheet which provided on the Kaggle Website for Website Traffic Analysis



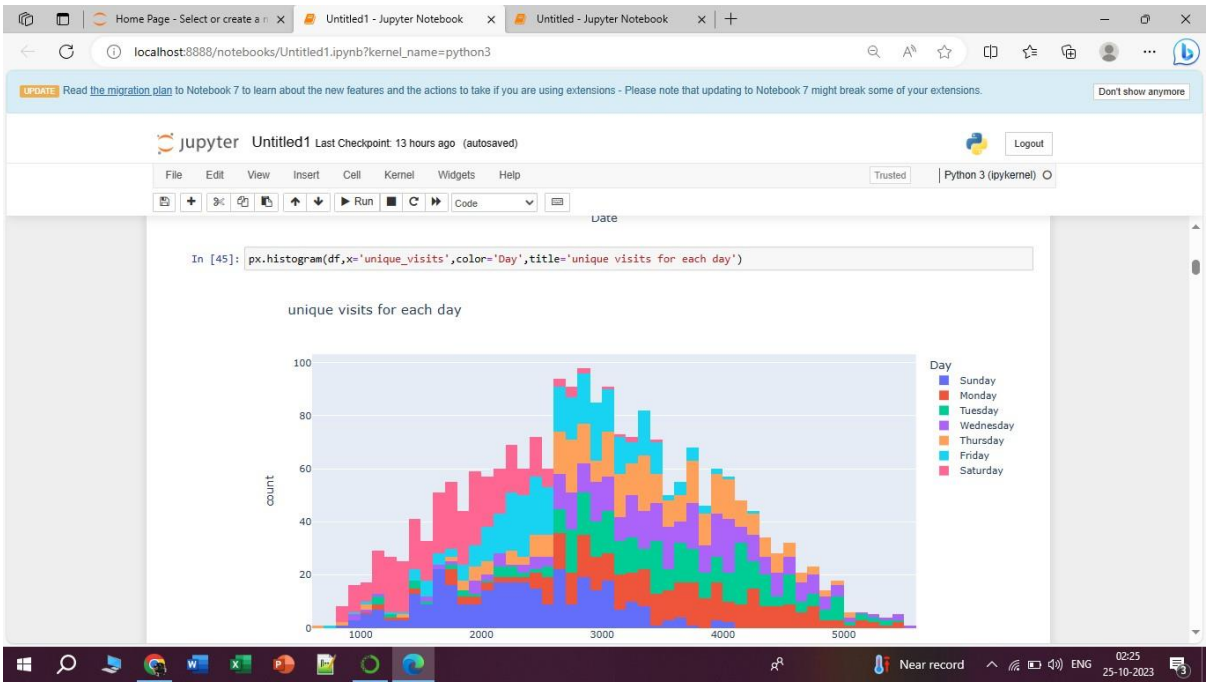
Showcasing the general information about the Data set for the better understanding for analysis.

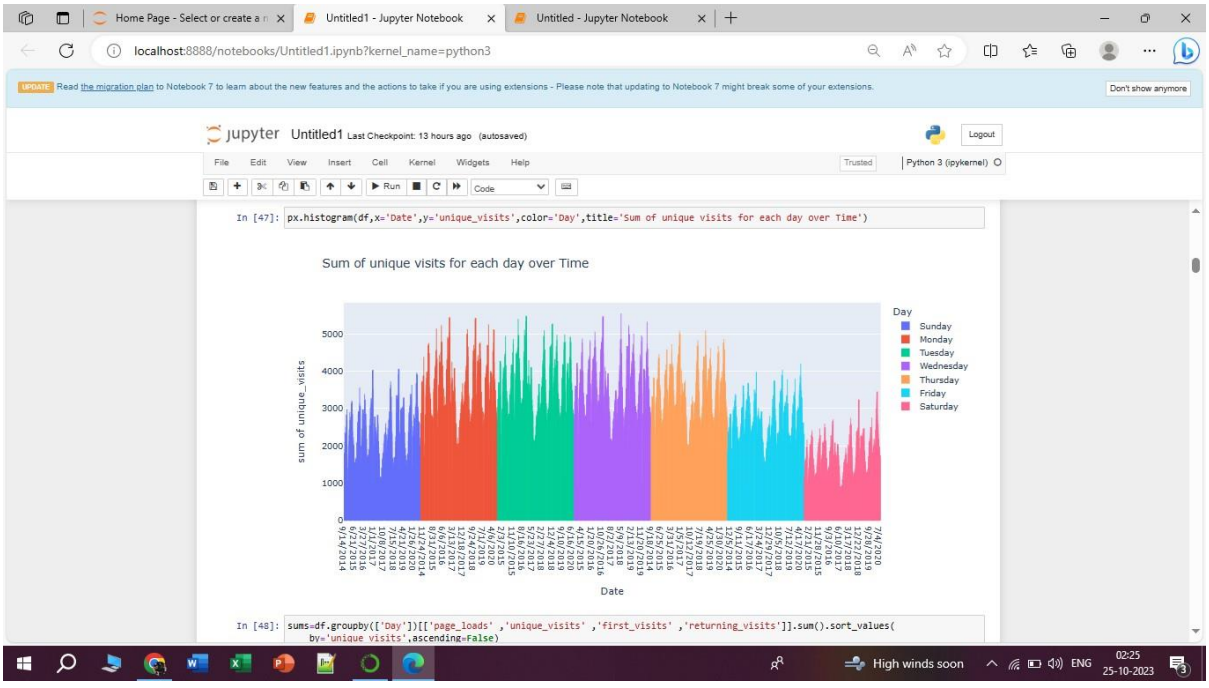


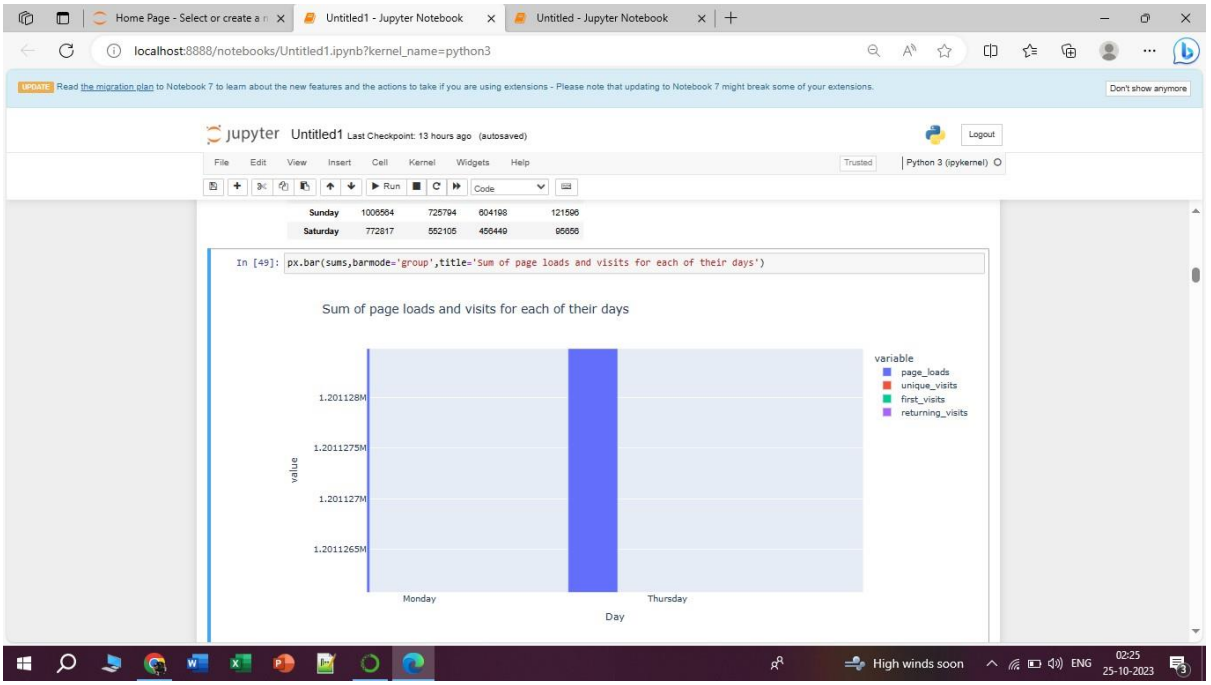
Presenting the **Time Series** for Page loads, Unique Visits, First time, Returning with the help of provided Data set to analyse the Traffic



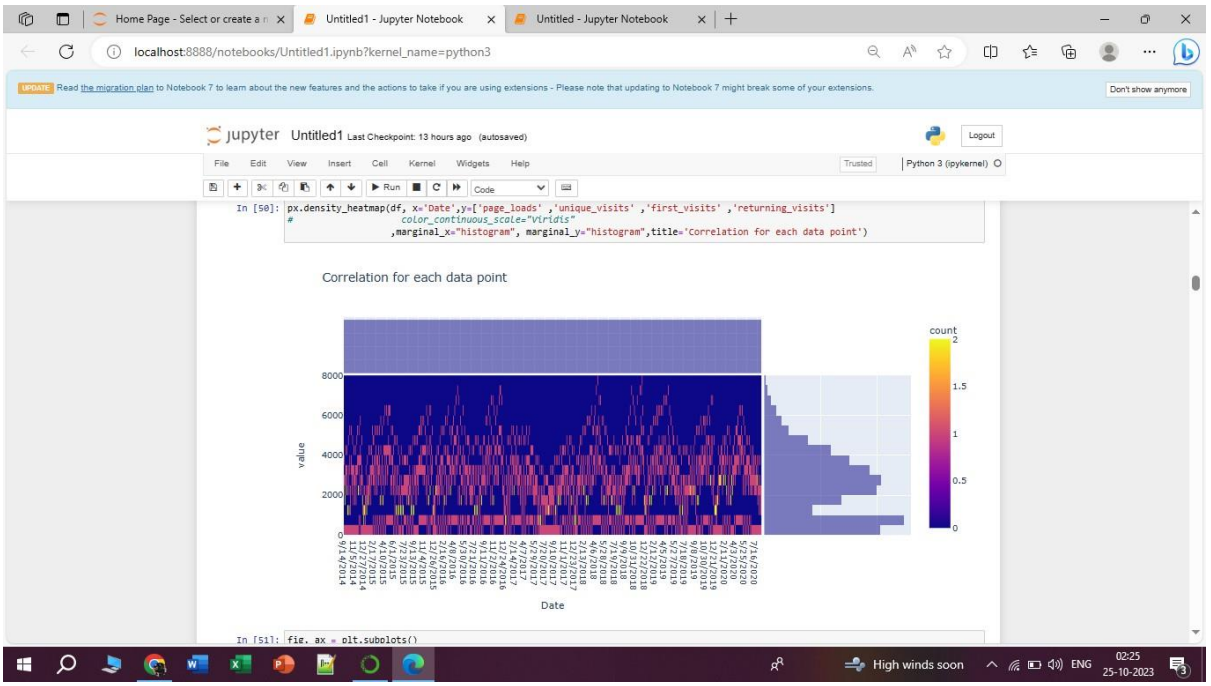
Calculating the Sum values for the Unique Visitors with the help of visualizaiton tool like Bar Graph and Histogram.



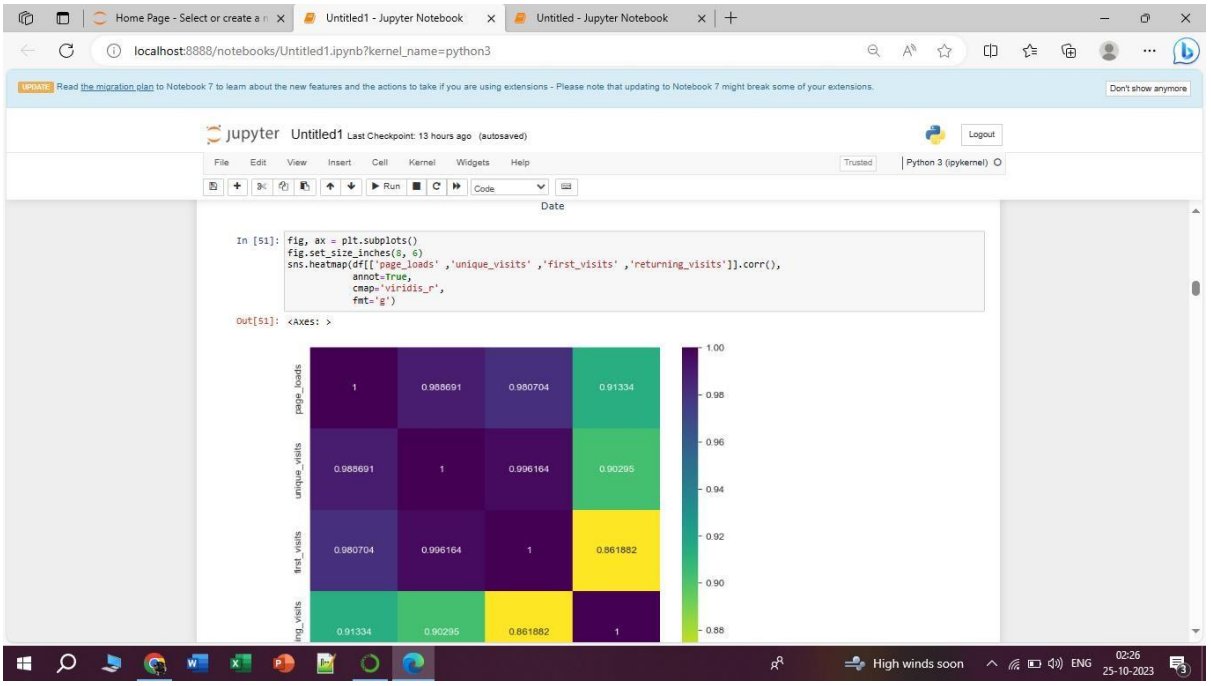




Visualizing the Sum of the Page loads in the bar graph to under stand the Traffic in the week.



Representing the Correlation of the Whole Data set for **EDA** Predcition to know the Stuffs that used by the visitors for there necessity calculation.



4.Conclusion

4.1 Summary of Project

Project Summary: Website Traffic Analysis

A successful online presence hinges on understanding the behavior of website visitors and leveraging data-driven insights to enhance user experiences and optimize performance. The project, "Website Traffic Analysis," embarks on a comprehensive journey to unravel the intricacies of data analytics and transform raw website traffic data into actionable insights.

This project encompasses a multi-faceted approach, leveraging a combination of data collection, data manipulation, visualization, and integration with Python for advanced analysis.

4.2 Achievement and challenges

That's a significant achievement in your "Website Traffic Analysis" project. Understanding that there are more visitors on weekdays compared to weekends provides valuable insights for website owners. This information can lead to several actionable strategies and decisions, such as:

1. Content Scheduling: Website owners can plan to release important or engaging content during weekdays to maximize its reach. This aligns content publication with high visitor traffic.

2. Advertising and Promotion: For marketing campaigns and promotions, targeting weekdays might be more effective, as more users are actively engaging with the website.

3. User Engagement: Knowing that weekdays are busier, website owners can focus on strategies to keep visitors engaged during this time, such as live chats, interactive features, or tailored content.

4. Resource Allocation: Allocating resources like customer support, server capacity, or content updates based on this insight can help optimize the user experience during peak times.

5. Data-Driven Decision-Making: Your project highlights the importance of data-driven decision-making. Website owners can use similar analyses and insights to make informed choices in the future.

4.3 Future Directions

1. Visitor Behavior Analysis: Delve deeper into visitor behavior patterns during weekdays. Are there specific times of day when traffic is highest? Do certain types of content or interactions attract more visitors on weekdays?

2. Weekend User Engagement: Investigate the reasons for lower weekend traffic. Are there particular content types or user engagement strategies that can be employed to boost weekend user numbers? This could involve running special weekend promotions or events.

3. Content Strategy: Develop a content strategy that takes into account the weekday and weekend traffic patterns. Consider aligning your content with these patterns to maximize engagement and reach.

4. A/B Testing: Implement A/B testing to evaluate the effectiveness of different strategies on weekdays versus weekends. This can help in fine-tuning your approach to optimize user experiences.

5. User Segmentation: Segment users based on their behavior during weekdays and weekends. This can help tailor content and features to different user groups.

6. Geographic Analysis: Explore whether there are geographic variations in the traffic patterns. Are certain regions more active on weekdays or weekends? This information can be used for targeted marketing efforts.

5.References

5.1 Citations and Sources used

When using a dataset from Kaggle or any other source, it's important to provide appropriate attribution and citations. Here's how you can format citations for a Kaggle dataset and references for other materials:

Kaggle Dataset and Codings:

Title: Daily Website Visitors Dataset

Author : **Bob Nau**

Source: Kaggle

URL: [Link to the Kaggle dataset](https://www.kaggle.com/datasets/bobnau/daily-website-visitors)