

PROJECT TITLE: The Website Traffic Analysis

Project Definition: The project involves analysing website traffic data to gain insights into user behaviour, popular pages, and traffic sources. The goal is to help website owners enhance the user experience by understanding how visitors interact with the site. This project encompasses defining the analysis objectives, collecting website traffic data, using IBM Cognos for data visualization, and integrating Python code for advanced analysis.

Project Objective:

Web traffic analysis is the process of collecting and analyzing data about the visitors and their behavior on a website. The objective of web traffic analysis is to understand how the website is performing, how to optimize it for better user experience, and how to increase conversions and revenue. Some of the common metrics used for web traffic analysis are:

Sessions: The number of times users interact with the website within a given time period.

- **Pageviews:** The number of times users view a page on the website.
- **Bounce rate:** The percentage of sessions that end after viewing only one page on the website.
- **Conversion rate:** The percentage of sessions that result in a desired action, such as making a purchase, signing up for a newsletter, or filling out a form.

Design Thinking:

1. **Analysis Objectives:** Define the key insights you want to extract from the website traffic data, such as identifying popular pages, traffic trends, and user engagement metrics.
2. **Data Collection:** Determine the data sources and methods for collecting website traffic data, including page views, unique visitors, referral sources, and more.
3. **Visualization:** Plan how to visualize the insights using IBM Cognos to create meaningful dashboards and reports.
4. **Python Integration:** Consider incorporating machine learning models to predict future traffic trends or user behaviour pattern.

Development Phase

Define Analysis Objectives:

Start by discussing the project objectives with the website owners or stakeholders.

Understand their specific goals and expectations from the analysis.

Clearly define what insights you want to extract from the website traffic data. For example, you may want to understand user demographics, behaviour patterns, or conversion rates.

Data Collection:

Identify the data sources for website traffic data. Common sources include Google Analytics, server logs, or other web analytics tools.

Set up data collection mechanisms to gather relevant data, ensuring that you capture the necessary metrics (e.g., page views, sessions, bounce rates, referral URLs, user agents).

Data Preparation:

Clean and preprocess the collected data to ensure it is accurate and consistent. This may involve handling missing values, removing duplicates, and transforming data as needed. Store the data securely in a structured format for analysis.

Data Visualization with IBM Cognos:

Use IBM Cognos or a similar data visualization tool to create visualizations and dashboards that provide insights into user behaviour, popular pages, and traffic sources. Generate charts, graphs, and reports that make it easy for stakeholders to interpret and derive insights from the data.

Advanced Analysis with Python:

Integrate Python into your analysis process to perform more advanced tasks, such as:

Clustering users based on behaviour.

Predicting user conversions.

Analysing traffic trends over time.

Write Python code to extract data from your data storage, perform analysis, and generate additional insights beyond what can be achieved with IBM Cognos alone.

Insights and Recommendations:

Analyse the results from both the data visualization tool and your Python-based analysis to derive actionable insights.

Based on these insights, provide recommendations to the website owners for improving the user experience. These recommendations should be data-driven and tailored to the specific findings.

Implementation and Monitoring:

Work with website owners to implement the recommended changes or enhancements to the website.

Continuously monitor the impact of these changes on user behavior and website performance.

Documentation and Reporting:

Document your analysis process, findings, and recommendations in a clear and concise manner.

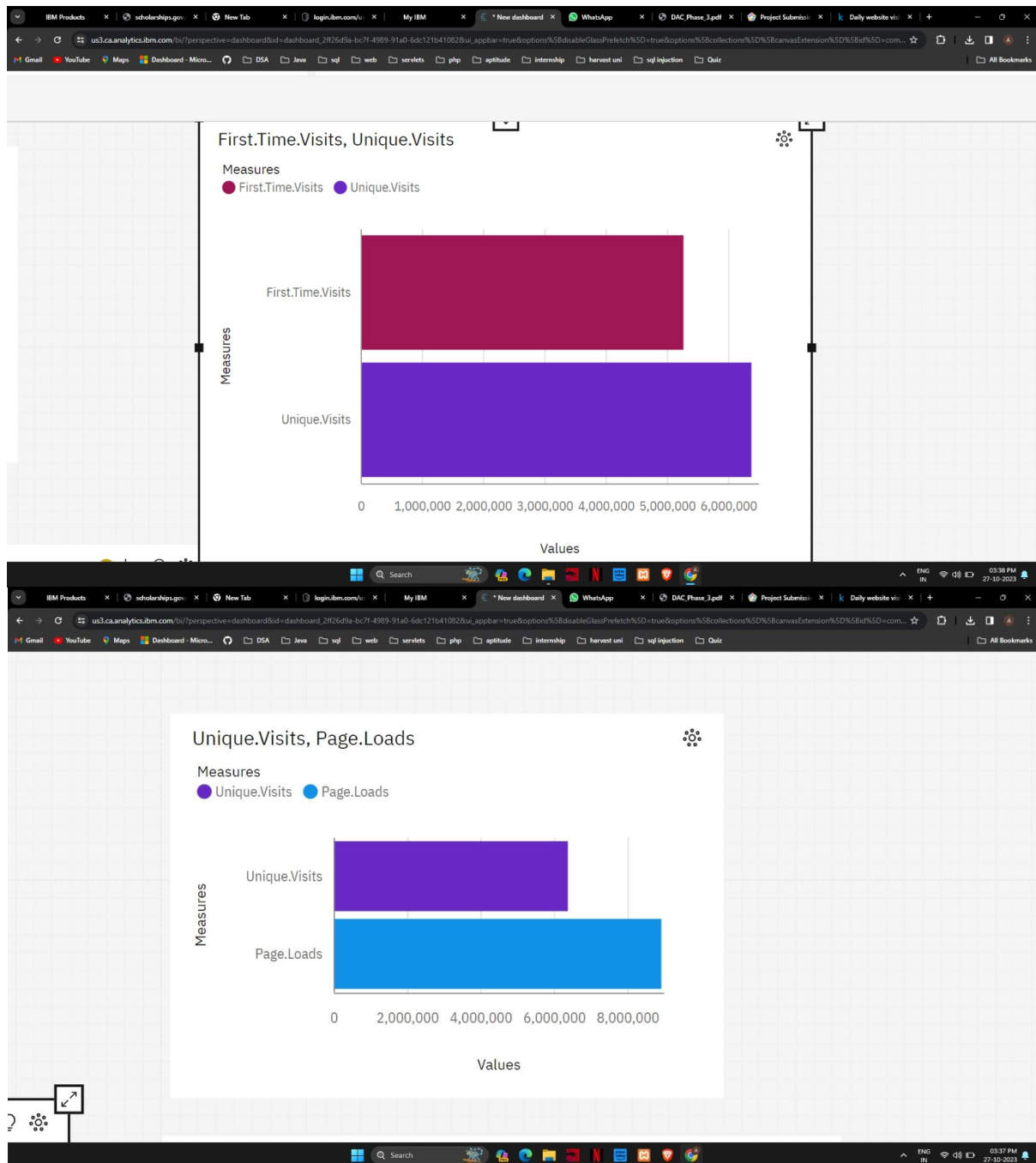
Create reports and presentations that can be shared with stakeholders.

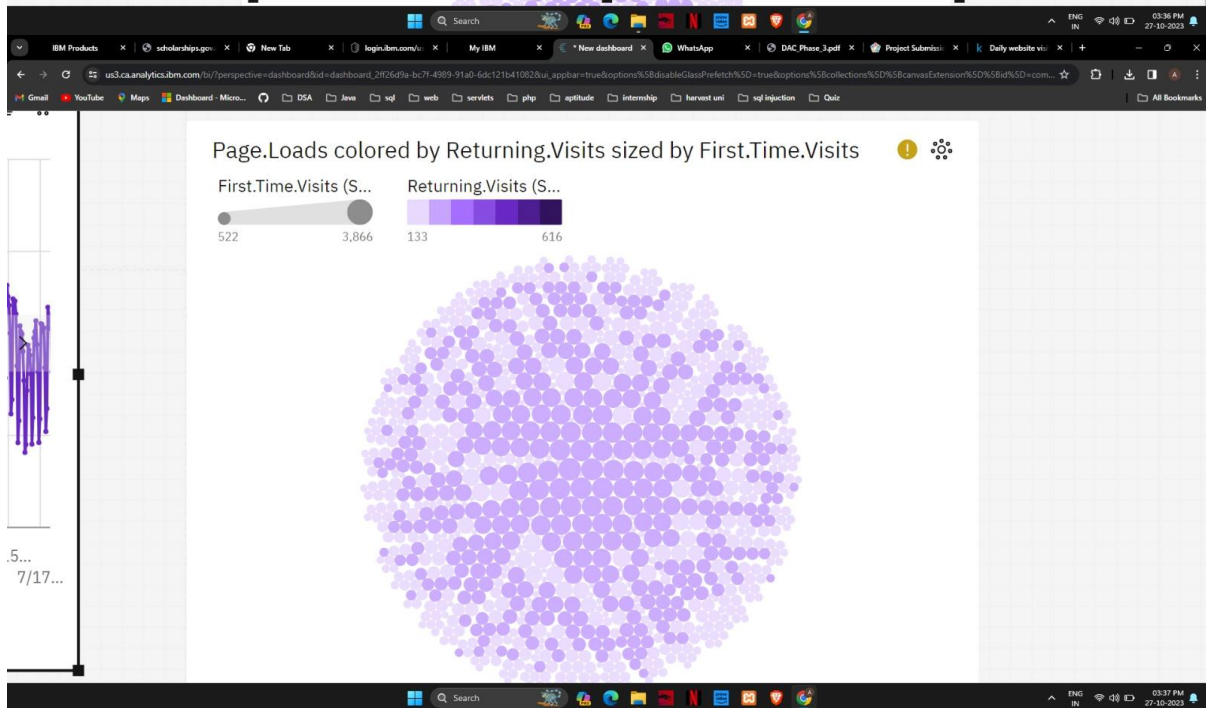
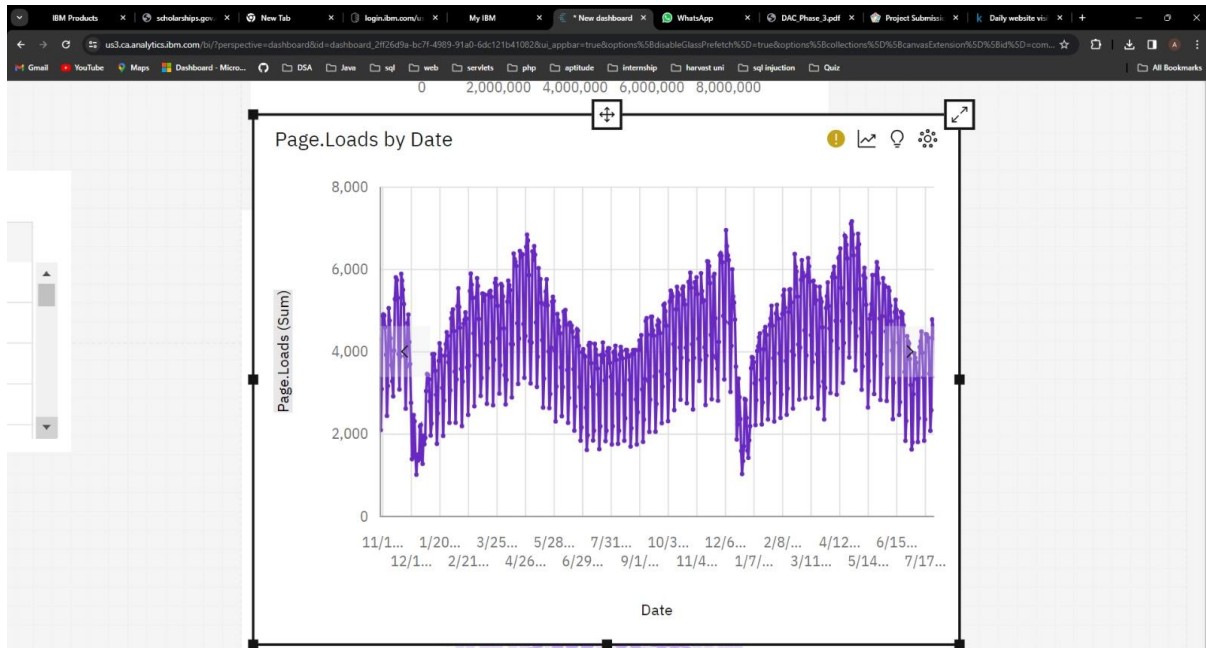
Iterate and Improve:

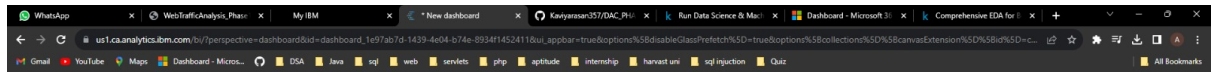
Website analytics is an ongoing process. Regularly update your analysis as new data becomes available and refine your recommendations based on evolving user behaviour and trends.

Throughout the project, it's essential to maintain data privacy and security, adhere to relevant regulations, and ensure that sensitive user information is handled responsibly. Additionally, effective communication with stakeholders is critical to ensure that the analysis results in meaningful improvements to the website's user experience.

VISUALIZATION USING IBM COGNOS ANALYTICS BEFORE PREPROCESSING



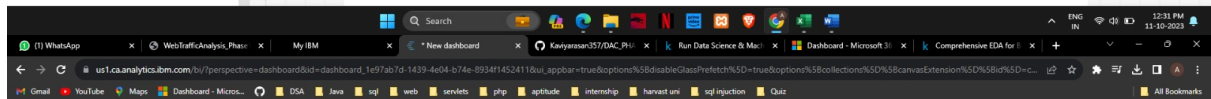
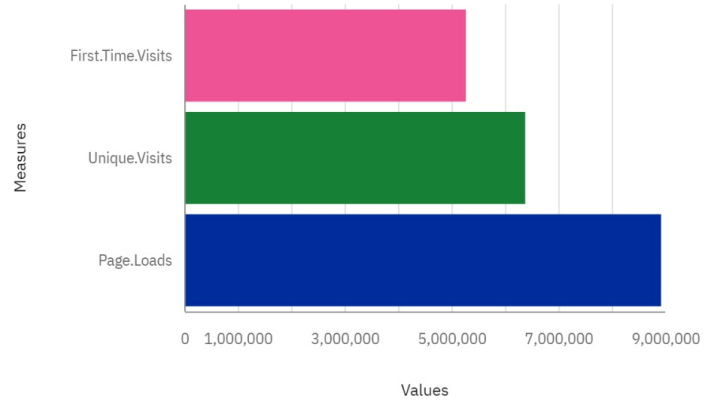




3. Visits

First.Time.Visits, Unique.Visits, Page.Loads

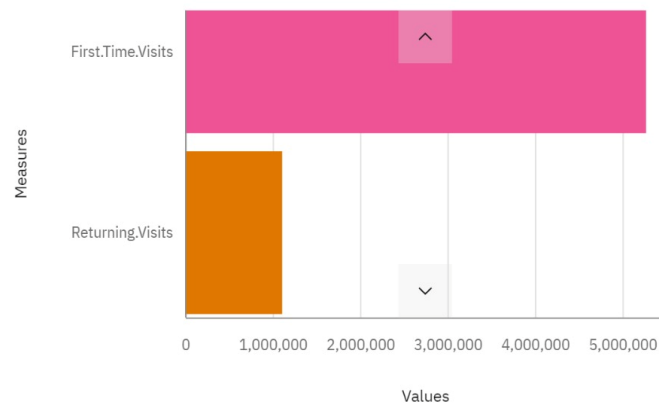
Measures
● First.Time.Visits ● Unique.Visits ● Page.Loads



4.

First.Time.Visits, Returning.Visits

Measures
● First.Time.Visits ● Returning.Visits



DATA PREPROCESSING USING PYTHON

```
In [13]: import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df.fillna(130, inplace = True)

print(df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
5	6	Friday	6	9/19/2014	2,815	1,863	1,622	241
6	7	Saturday	7	9/20/2014	1,658	1,118	985	133
7	8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
8	9	Monday	2	9/22/2014	3,638	2,586	2,312	274
9	10	Tuesday	3	9/23/2014	4,462	3,257	2,989	268
10	11	Wednesday	4	9/24/2014	4,414	3,175	2,891	284
11	12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
12	13	Friday	6	9/26/2014	3,323	2,249	2,033	216
13	14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
14	15	Sunday	1	9/28/2014	2,465	1,806	1,613	193
15	16	Monday	2	9/29/2014	4,096	2,873	2,577	296
16	17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
17	18	Wednesday	4	10/1/2014	4,124	2,849	2,541	308

```
In [4]: import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

new_df = df.dropna()

print(new_df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
5	6	Friday	6	9/19/2014	2,815	1,863	1,622	241
6	7	Saturday	7	9/20/2014	1,658	1,118	985	133
7	8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
8	9	Monday	2	9/22/2014	3,638	2,586	2,312	274
9	10	Tuesday	3	9/23/2014	4,462	3,257	2,989	268
10	11	Wednesday	4	9/24/2014	4,414	3,175	2,891	284
11	12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
12	13	Friday	6	9/26/2014	3,323	2,249	2,033	216
13	14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
14	15	Sunday	1	9/28/2014	2,465	1,806	1,613	193
15	16	Monday	2	9/29/2014	4,096	2,873	2,577	296
16	17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
17	18	Wednesday	4	10/1/2014	4,124	2,849	2,541	308

```
In [5]: import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df['Date'] = pd.to_datetime(df['Date'])

print(df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	2014-09-14	2,146	1,582	1,430	152
1	2	Monday	2	2014-09-15	3,621	2,528	2,297	231
2	3	Tuesday	3	2014-09-16	3,698	2,630	2,352	278
3	4	Wednesday	4	2014-09-17	3,667	2,614	2,327	287
4	5	Thursday	5	2014-09-18	3,316	2,366	2,130	236
5	6	Friday	6	2014-09-19	2,815	1,863	1,622	241
6	7	Saturday	7	2014-09-20	1,658	1,118	985	133
7	8	Sunday	1	2014-09-21	2,288	1,656	1,481	175
8	9	Monday	2	2014-09-22	3,638	2,586	2,312	274
9	10	Tuesday	3	2014-09-23	4,462	3,257	2,989	268
10	11	Wednesday	4	2014-09-24	4,414	3,175	2,891	284
11	12	Thursday	5	2014-09-25	4,315	3,029	2,743	286
12	13	Friday	6	2014-09-26	3,323	2,249	2,033	216
13	14	Saturday	7	2014-09-27	1,656	1,180	1,040	140
14	15	Sunday	1	2014-09-28	2,465	1,806	1,613	193
15	16	Monday	2	2014-09-29	4,096	2,873	2,577	296
16	17	Tuesday	3	2014-09-30	4,474	3,032	2,720	312
17	18	Wednesday	4	2014-10-01	4,124	2,849	2,541	308

```
In [9]: import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df.drop_duplicates(inplace = True)

print(df.to_string())
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236
5	6	Friday	6	9/19/2014	2,815	1,863	1,622	241
6	7	Saturday	7	9/20/2014	1,658	1,118	985	133
7	8	Sunday	1	9/21/2014	2,288	1,656	1,481	175
8	9	Monday	2	9/22/2014	3,638	2,586	2,312	274
9	10	Tuesday	3	9/23/2014	4,462	3,257	2,989	268
10	11	Wednesday	4	9/24/2014	4,414	3,175	2,891	284
11	12	Thursday	5	9/25/2014	4,315	3,029	2,743	286
12	13	Friday	6	9/26/2014	3,323	2,249	2,033	216
13	14	Saturday	7	9/27/2014	1,656	1,180	1,040	140
14	15	Sunday	1	9/28/2014	2,465	1,806	1,613	193
15	16	Monday	2	9/29/2014	4,096	2,873	2,577	296
16	17	Tuesday	3	9/30/2014	4,474	3,032	2,720	312
17	18	Wednesday	4	10/1/2014	4,124	2,849	2,541	308

```
In [6]: import pandas as pd

df = pd.read_csv('daily-website-visitors.csv')

df.drop_duplicates(inplace = True)

print(df.to_string())

x=df.to_csv('C:\\Users\\yoges\\Documents\\daily-website-visitors.csv')
```

Unnamed: 0.4	Unnamed: 0.3	Unnamed: 0.2	Unnamed: 0.1	Unnamed: 0	Row	Day	Day.Of.Week	Date	Page.Loads
Unique.Visits	First.Time.Visits	Returning.Visits							
0	0	0	0	0	1	Sunday	1	2014-09-14	2,146
1,582	1,430	152							
1	1	1	1	1	2	Monday	2	2014-09-15	3,621
2,528	2,297	231							
2	2	2	2	2	3	Tuesday	3	2014-09-16	3,698
2,630	2,352	278							
3	3	3	3	3	4	Wednesday	4	2014-09-17	3,667
2,614	2,327	287							
4	4	4	4	4	5	Thursday	5	2014-09-18	3,316
2,366	2,130	236							
5	5	5	5	5	6	Friday	6	2014-09-19	2,815
1,863	1,622	241							
6	6	6	6	6	7	Saturday	7	2014-09-20	1,658
1,118	985	133							
7	7	7	7	7	8	Sunday	1	2014-09-21	2,288
1,656	1,481	175							
8	8	8	8	8	9	Monday	2	2014-09-22	3,638

VISUALATION USING IBM COGNOS ANALYTICS AFTER DATA PREPROCESSING

show insights

@Cognos ⓘ

Please let me know which columns you are interested in.

For **dally-website-visitors.csv**, here is a list of relevant fields:

Page.Loads

Returning.Visits

Unique.Visits

First.Time.Visits

Row

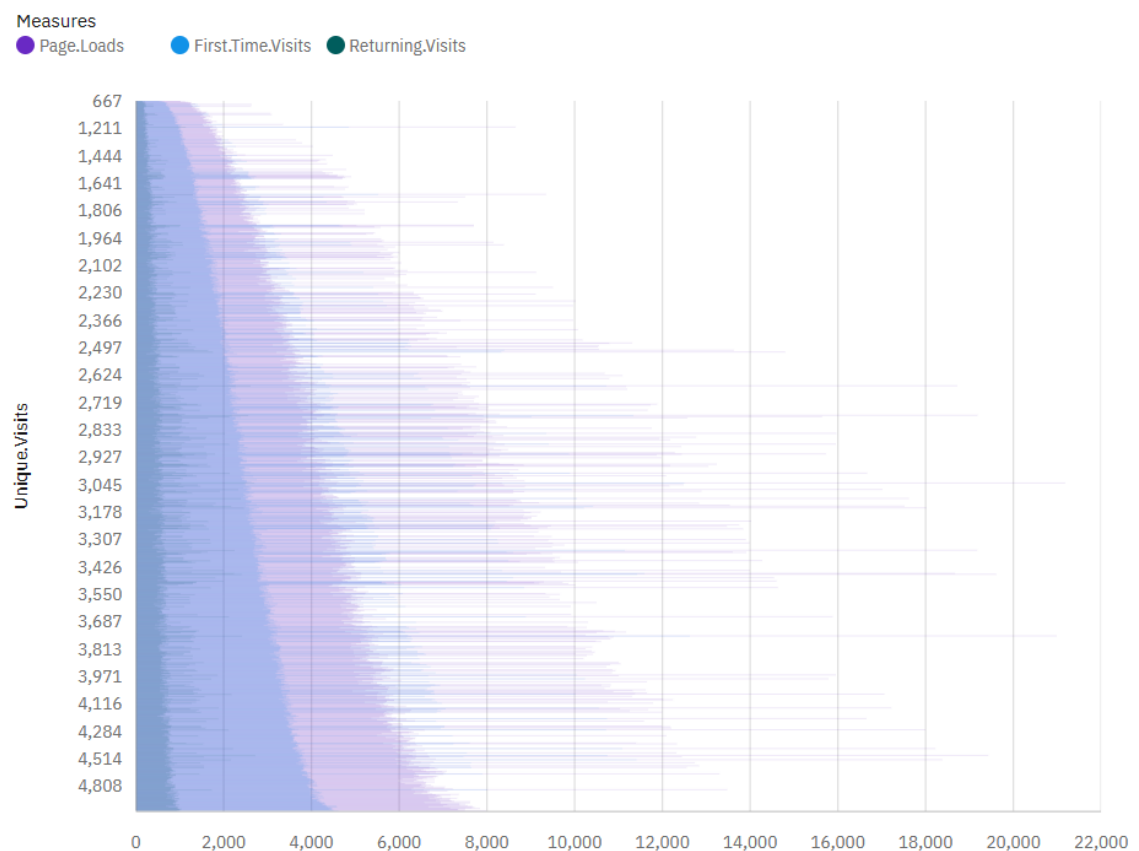
Unnamed: 0

Unnamed: 0.1

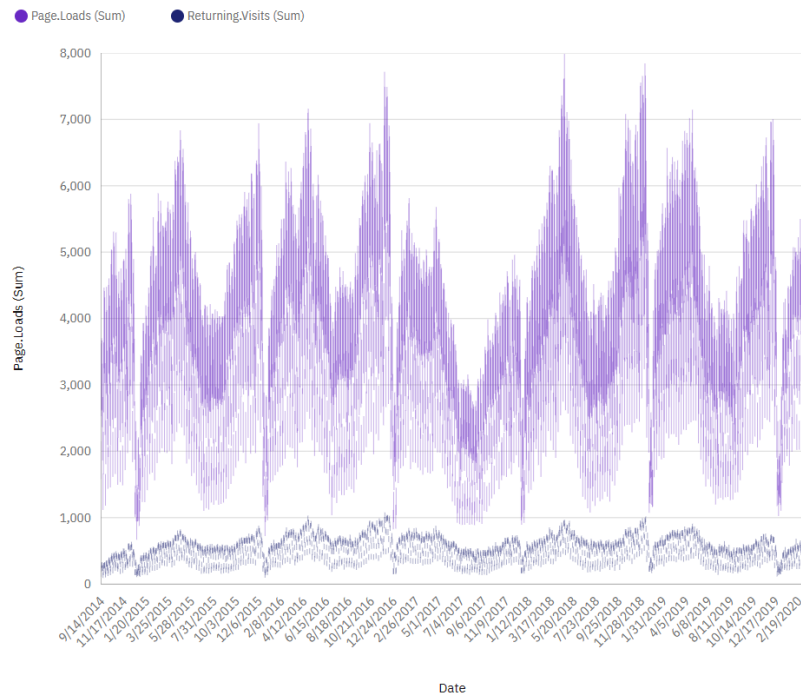
Unnamed: 0.2

Unnamed: 0.3

Page.Loads, First.Time.Visits and Returning.Visits by Unique.Visits



Page.Loads compared to Returning.Visits by Date



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (3) ⓘ

Based on the current forecasting, **Page.Loads** may reach **nearly four thousand** by **Date 2021-10-27**.



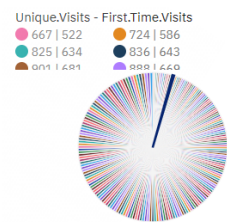
Across all **dates**, the sum of **Page.Loads** is **over 8.9 million**.



Page.Loads ranges from **over a thousand**, when **Date** is **2014-12-25**, to **nearly eight thousand**, when **Date** is **2018-04-25**.



Page.Loads by Unique.Visits, First.Time.Visits and Returning.Visits



Insights



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (8) ⓘ

Returning.Visits 552 has the highest total **Page.Loads** due to **First.Time.Visits 3635**.



It is projected that by **Monday+1, 2310** will exceed **2203** in **Page.Loads** by **366**.



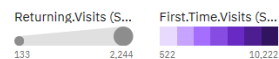
It is projected that by **Monday+1, 710** will exceed **552** in **Page.Loads** by **almost 1500**.



It is projected that by **Monday+1, 2844** will exceed **3039** in **Page.Loads** by **410**.



Page.Loads, First.Time.Visits, Returning.Visits



Page.Loads

1,225 1,296 1,380 1,474 1,546 1,606 1,664 1,709 1,742 1,795 1,831 1,857 1,922 1,954 1,998
1,002 1,265 1,334 1,430 1,523 1,578 1,644 1,685 1,733 1,780 1,826 1,848 1,896 1,946 1,969 2...



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (9) ⓘ

It is projected that by **Monday+1, 4205** will exceed **3973** in **First.Time.Visits** by a **thousand**.



2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **First.Time.Visits** values (**0.2** % of the total).



2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Unique.Visits** values (**0.2** % of the total).

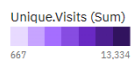


Over all values of **Page.Loads**, the average of **First.Time.Visits** is **almost 2500**.



Over all values of **Page.Loads**, the

First.Time.Visits by Returning.Visits with points for Page.Loads



Insights



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (4) ⓘ

Returning.Visits is most unusual when the values of **First.Time.Visits** are **2989, 8761, 10,517, 2891, 5599** and more.



It is projected that by **Monday+1, 4205** will exceed **3973** in **First.Time.Visits** by a **thousand**.

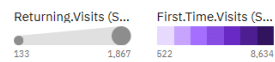


Returning.Visits and **First.Time.Visits** have a **strong** positive linear association, being **Returning.Visits = 39.86 + 0.1992 * First.Time.Visits**



The total of **Returning.Visits** is **over 1.1 million**.

Page.Loads colored by First.Time.Visits sized by Returning.Visits



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (9) ⓘ

It is projected that by **Monday+1**, **4205** will exceed **3973** in **First.Time.Visits** by a thousand.



2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Returning.Visits** values (**0.2** % of the total).



2948 is the most frequently occurring category of **Page.Loads** with a count of **5** items with **Unique.Visits** values (**0.2** % of the total).



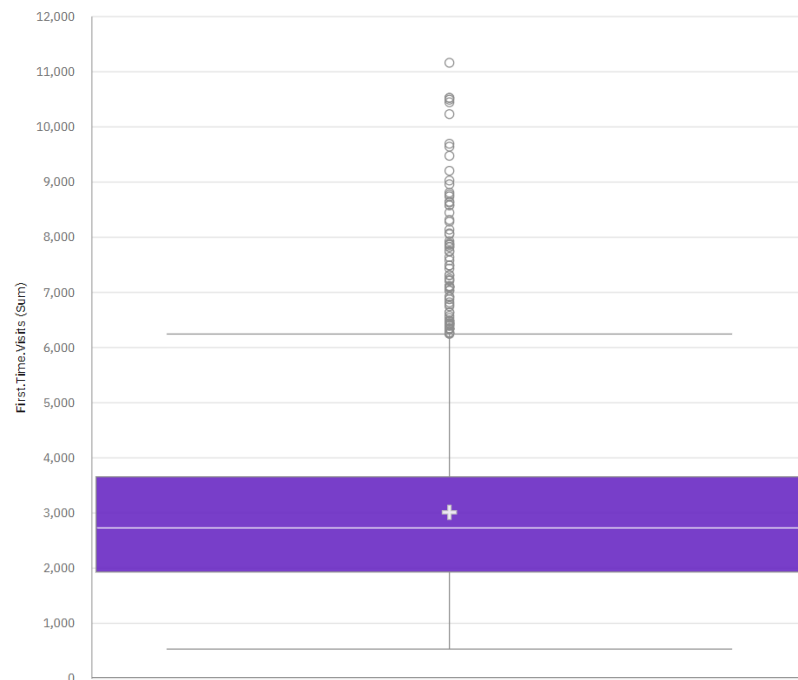
Over all values of **Page.Loads**, the average of **Returning.Visits** is **511.8**.



Over all values of **Page.Loads**, the average of **Unique.Visits** is **nearly**



First.Time.Visits by Page.Loads



Narrative insights

Favorite insights (0) ⓘ

Click the star icon for a suggested insight to add it as a favorite.

Suggested insights (4) ⓘ

It is projected that by **Monday+1**, **4205** will exceed **3973** in **First.Time.Visits** by a thousand.



Page.Loads 4376 has the highest **Total Returning.Visits** but is ranked **#6** in **Total First.Time.Visits**.

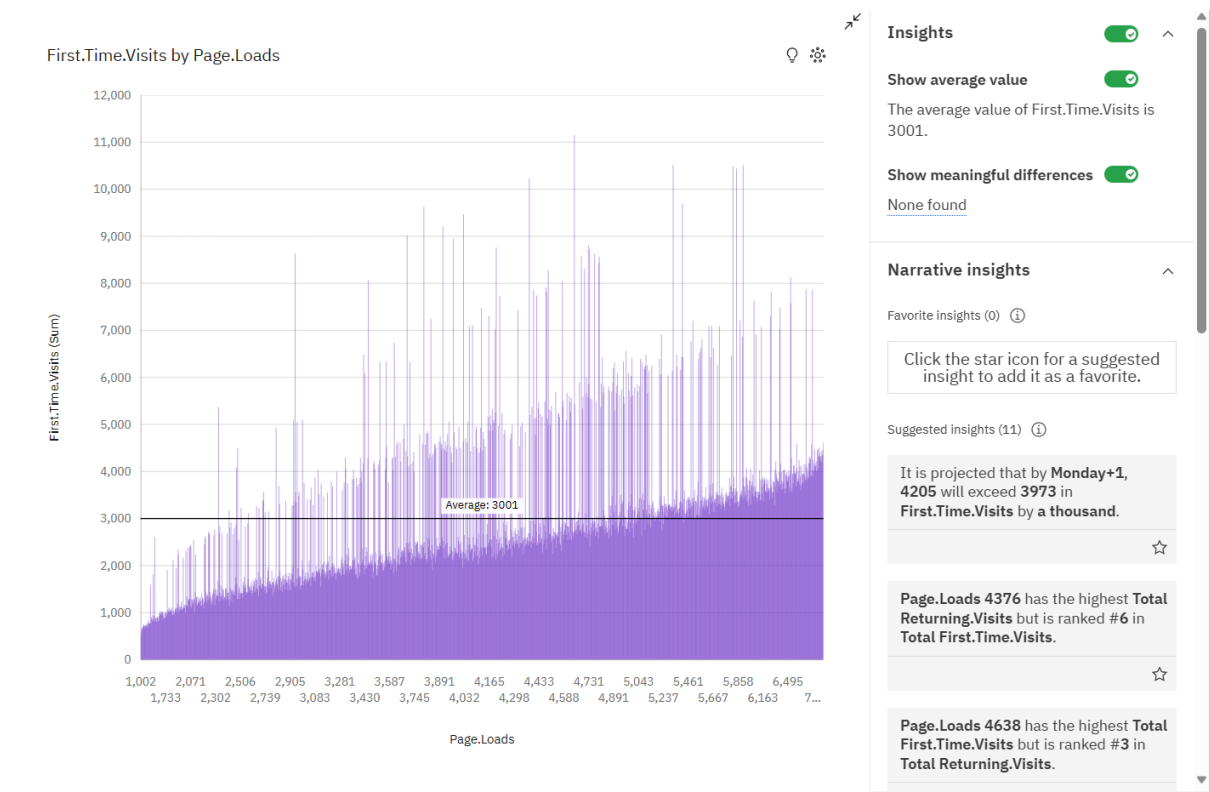


Page.Loads 4638 has the highest **Total First.Time.Visits** but is ranked **#3** in **Total Returning.Visits**.



The total of **Returning.Visits** is over **1.1** million.





Website traffic analysis provides valuable insights that can significantly benefit the website owner. Here's how these insights can help the owner improve their website and enhance the overall visitor experience:

1.Understanding Audience Behavior:

Insight: Analyzing user behavior helps the owner understand what visitors are looking for on their site.

Benefit: By knowing what users want, the owner can tailor content, products, or services to meet visitor needs effectively.

2.Identifying Popular Content:

Insight: Knowing which pages or posts are popular indicates what resonates with the audience.

Benefit: The owner can create more content similar to the popular ones, increasing user engagement and time spent on the website.

3.Improving User Experience:

Insight: Analyzing exit pages and bounce rates helps identify weak points in the user journey.

Benefit: By enhancing these pages or providing clearer calls to action, the owner can reduce bounce rates and improve conversion rates.

4. Optimizing Marketing Strategies:

Insight: Understanding traffic sources helps the owner identify which marketing channels are most effective.

Benefit: Focus marketing efforts and budget on channels that bring in high-quality traffic, leading to better ROI.

can rank higher in search engine results, increasing visibility and organic traffic.

5. Data-Driven Decision Making:

Insight: Regular reports and data trends provide a basis for decision-making.

Benefit: The owner can make informed decisions on content, marketing, and user experience, leading to a more effective website strategy.

6. Enhancing User Engagement:

Insight: Analyzing user interactions with multimedia and interactive content.

Benefit: By creating more engaging multimedia content, the owner can increase user interaction and time spent on the website.

7. Staying Competitive:

Insight: Benchmarking against competitors provides context for performance.

Benefit: Understanding how the website performs compared to competitors can inform strategies, helping the owner stay ahead in the market.

CONCLUSION

In summary, website traffic analysis empowers the owner to make data-driven decisions, improve content, enhance user experience, and ultimately increase the website's effectiveness in achieving its goals, whether it's sales, lead generation, or content engagement.