PROJECT DOCUMENTATION AND SUBMISSION
WEBSITE TRAFFIC ANALYSIS

Date	1.11.2023
Team ID	718
Project Name	Website Traffic Analysis`

Introduction

Website traffic analysis is the process of monitoring and evaluating the visitors who come to your website. It's essential for understanding how users interact with your site and can provide valuable insights for improving its performance and achieving your goals

To perform website traffic analysis, webmasters typically use tools like Google Analytics, which provide detailed data and reports to help make informed decisions about site improvements and marketing strategies. Regularly reviewing and interpreting this data can enhance user experience, boost .conversions, and ultimately lead to a more successful online presence

This document delves into the comprehensive strategies employed in the Predicting Future Traffic Trends and User Behaviour Patterns" project. " Predicting future traffic trends and understanding user behaviour are pivotal in optimizing traffic management and digital platform performance. Through innovative methodologies, this project seeks to provide actionable insights for .stakeholders

Problem Statement

The problem at hand is to effectively analyze and interpret the traffic data of a website in order to address several key challenges and objective

The website needs to implement a robust traffic analysis solution, potentially utilizing tools like Google Analytics or similar analytics platforms. Additionally, creating a regular reporting and optimization cycle is essential to continuously improve the website's performance and achieve .its goals

:Objectives

The primary objective is to assess the overall performance of a website by tracking metrics such as traffic volume, user engagement, conversion rates, and bounce rates. This helps in understanding how effectively the website .is meeting its goals and where improvements may be needed

Another crucial objective is to use the insights gained from traffic analysis to make data-driven decisions. This includes optimizing content, marketing strategies, and user experience to enhance the website's .effectiveness, drive growth, and achieve specific business objectives

:Data Pre-proccessing

Handle Missing Values: You first checked for missing values in your dataset using the isnull() method and sum() function. This allowed you to .see how many missing values were present in each column

Data cleaning and preprocessing 3.1

Import Dependencies

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
dataset = pd.read_csv('/content/drive/MyDrive/daily-website-visitors.csv')
```

Dataset Exploration

```
dataset
```

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	3,69	2,63	2,352	27
3 4	Wednesday	4	9-17-2014	8	0	2,327	8
4 5	Thursday	5	9-18-2014	3,66	2,614	2,130	28
... ...	...	...	...	7	2,36	...	7
2162 2163	Saturday	7	8-15-	3,316	6 ...	1,373	23
2163 2164	Sunday	1	2020	...	1,696	1,686	6 ...
2164 2165	Monday	2	8-16-	2,221	2,03	2,181	323
2165 2166	Tuesday	3	2020	2,72	7	2,184	351
2166 2167	Wednesday	4	8-17-	4	2,63	1,297	45
2167 rows × 8 columns			2020	3,45	8		7
			8-18-	6	2,68		49
			2020 8-	3,581	3		9
			19-2020	2,06	1,564		26
				4			7

```
dataset.head()
```

	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	3,69	2,63	2,352	27
3	4	Wednesday	4	9-17-2014	8	0	2,327	8
4	5	Thursday	5	9-18-2014	3,66	2,614	2,130	28
					7	2,36		7

```
dataset.shape
```

```
(2167, 8)
```

					3,316	6		23
								6

```
dataset.columns
```

```
Index(['Row', 'Day', 'Day.Of.Week', 'Date', 'Page.Loads', 'Unique.Visits',
'First.Time.Visits', 'Returning.Visits'],
dtype='object')
```

```
dataset.dtypes
```

```
Row int64
Day object
Day.Of.Week int64
Date object
Page.Loads object
```

Z

```
Unique.Visits        object
First.Time.Visits    object
Returning.Visits object
dtype: object
```

Data Preprocessing

```
dataset.isnull()
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	False	False	Fals	Fals	Fals	Fals	Fals	Fals
1	False	False	e	e	e	e	e	e
2	False	False	Fals	Fals	Fals	Fals	Fals	Fals
3	False	False	e	e	e	e	e	e
4	False	False	Fals	Fals	Fals	Fals	Fals	Fals
…	…	…	e	e	e	e	e	e
2162	False	False	Fals	Fals	Fals	Fals	Fals	Fals
2163	False	False	e	e	e	e	e	e
2164	False	False	Fals	Fals	Fals	Fals	Fals	Fals
2165	False	False	e …	e …	e …	e …	e …	e …
2166	False	False	Fals	Fals	Fals	Fals	Fals	Fals

2167 rows × 8 columns

		Fals	Fals	Fals	Fals	Fals	Fals
		e	e	e	e	e	e
		Fals	Fals	Fals	Fals	Fals	Fals
		e	e	e	e	e	e
		Fals	Fals	Fals	Fals	Fals	Fals
		e	e	e	e	e	e
		Fals	Fals	Fals	Fals	Fals	Fals
		e	e	e	e	e	e
		Fals	Fals	Fals	Fals	Fals	Fals
		e	e	e	e	e	e

```
dataset.isnull().sum()
```

```
Row 0
Day 0
Day.Of.Week 0
Date 0
Page.Loads 0
Unique.Visits 0
First.Time.Visits 0
Returning.Visits 0
dtype: int64
```

```
dataset.isnull().sum().sum()
```

```
0
```

```
dataset.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

```
dataset.describe(include='all')
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
count	2167.000000	2167	2167.000000	2167	2167	2167	2167	2167
unique	NaN	7	NaN	2167	1756	1658	1587	663
top	NaN	Sunday	NaN	9-14-2014	2,948	2,780	3,146	552
freq	NaN	310	1084.000000	NaN	1	5	5	5	12
mean	NaN	625.703338	NaN	3.997231	NaN	NaN	NaN	NaN	NaN
std	1.000000	NaN	2.000229	NaN	NaN	NaN	NaN	NaN	NaN
min	542.000000	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN
25%	1084.000000	NaN	2.000000	NaN	NaN	NaN	NaN	NaN	NaN
50%	1625.500000	NaN	4.000000	NaN	NaN	NaN	NaN	NaN	NaN
75%	2167.000000	NaN	6.000000	NaN	NaN	NaN	NaN	NaN	NaN
max	2709.000000	NaN	7.000000	NaN	NaN	NaN	NaN	NaN	NaN

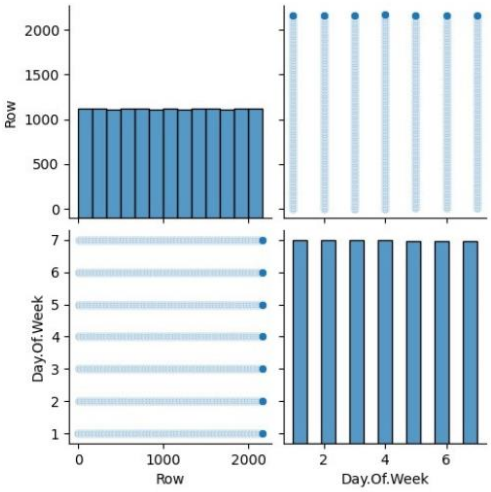
```
dataset.info()
```

```
<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
```

Data Visualization

```
plt.figure(figsize=(20,20))
sns.pairplot(dataset)
```

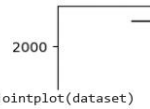
```
<seaborn.axisgrid.PairGrid at 0x7d9b3aee9060>
<Figure size 2000x2000 with 0 Axes>
```




```
sns.boxplot(dataset)
```

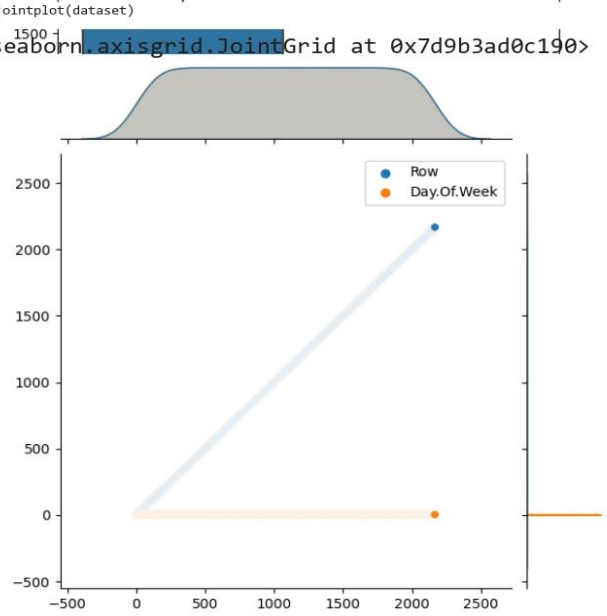
Z

<Axes: >




```
sns.jointplot(dataset)
```

<seaborn.axisgrid.JointGrid at 0x7d9b3ad0c190>



Correlation Visualization

```
dataset.corr()
```

```
<ipython-input-18-c187c74d1e71>:1: FutureWarning: The default value of numeric_only i
 dataset.corr()
```

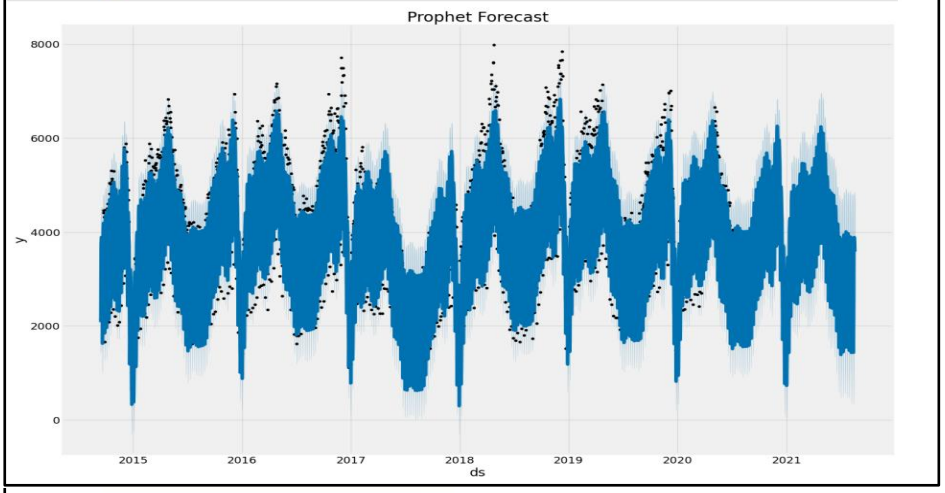
	Row	Day.Of.Week
Row	1.0000	0.0008
Day.Of.Week	0.0008	1.0000

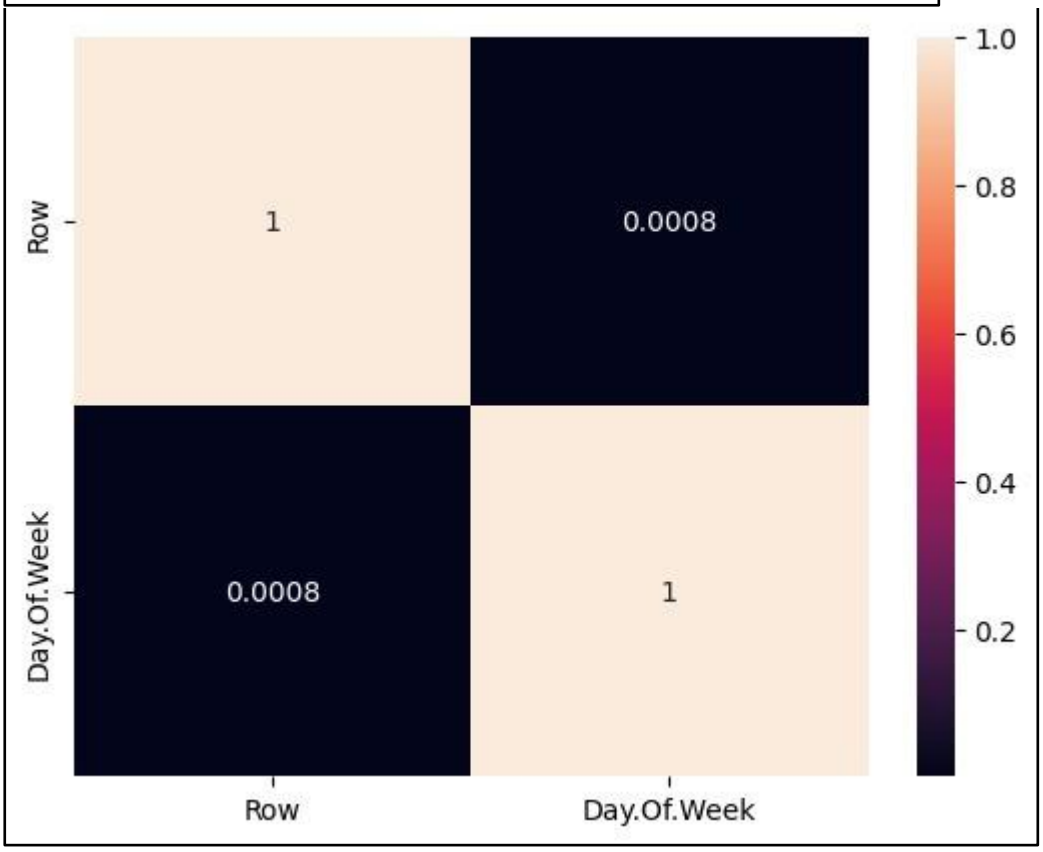
```
sns.heatmap(dataset.corr(),annot=True)
```

```
ipython-input-19-9d3fd451b567>:1: FutureWarning: The default value of numeric_only i >
 sns.heatmap(dataset.corr(),annot=True)
< :Axes>
```

Z

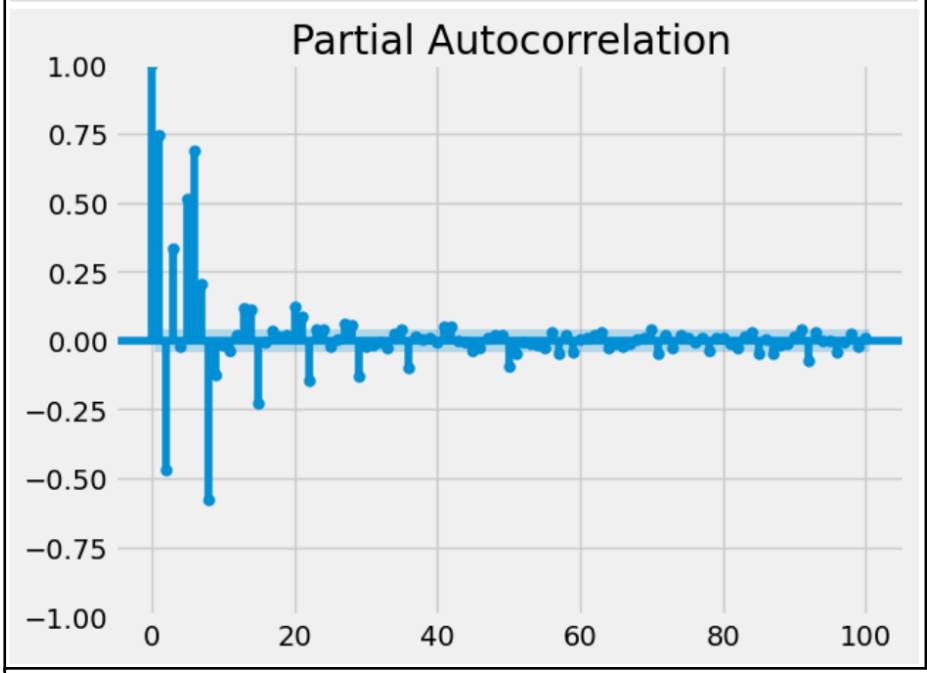
```
future = model.make_future_dataframe(periods=365)
forecast = model.predict(future)
fig = model.plot(forecast, figsize=(15, 10))
plt.title("Prophet Forecast")
plt.show()
```



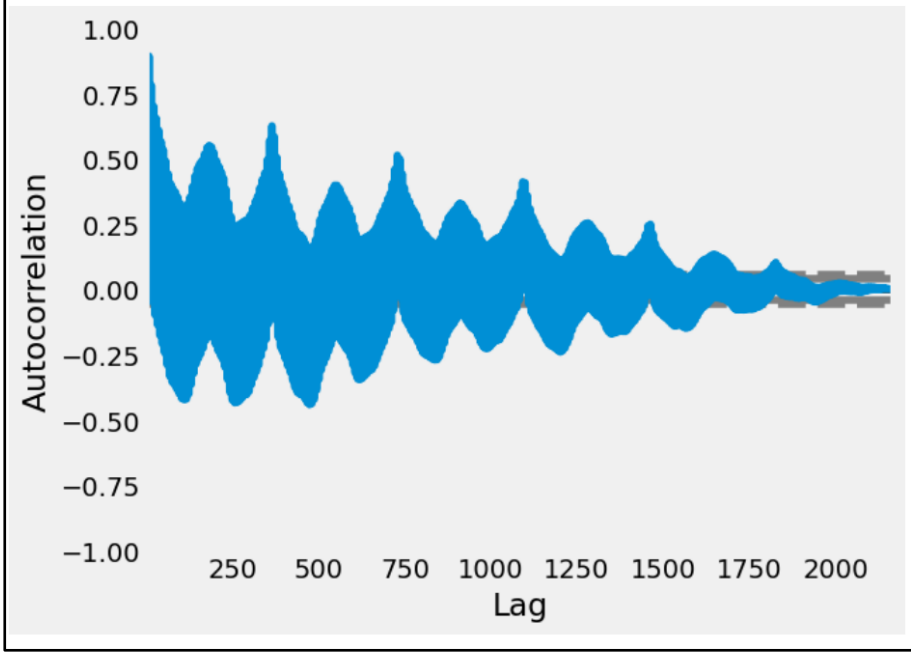


Z

```
plot_pacf(data["y"], lags = 100)
```

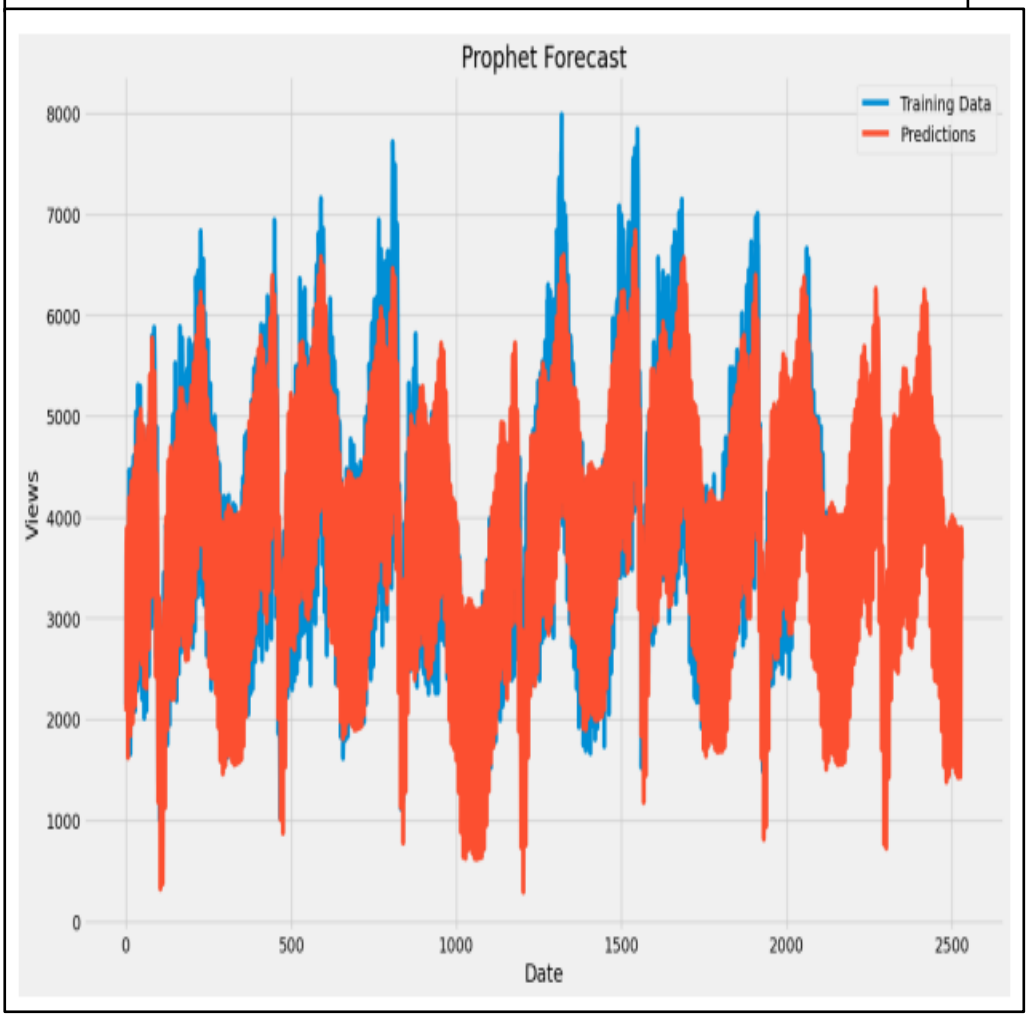



```
<Axes: xlabel='Lag', ylabel='Autocorrelation'>
```



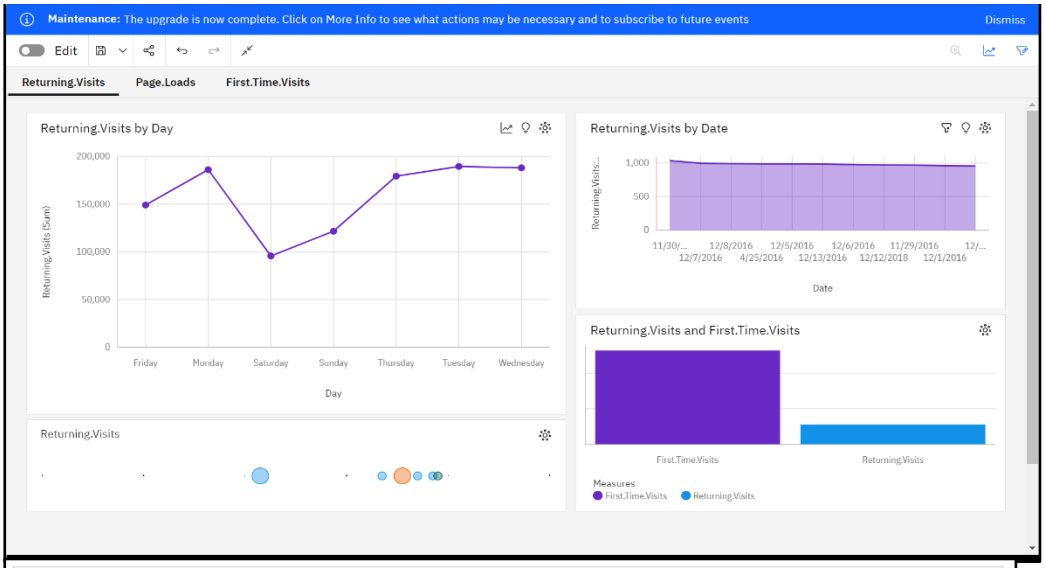
z

```
data["y"].plot(legend=True, label="Training Data", figsize=(15, 10))
forecast["yhat"].plot(legend=True, label="Predictions")
plt.title("Prophet Forecast")
plt.xlabel("Date")
plt.ylabel("Views")
plt.show()
```



8. Visualization using Cognos

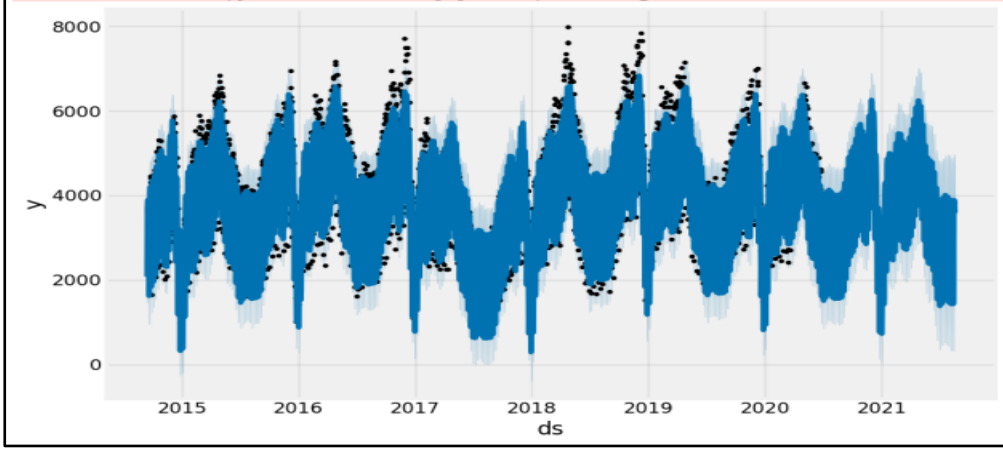
z



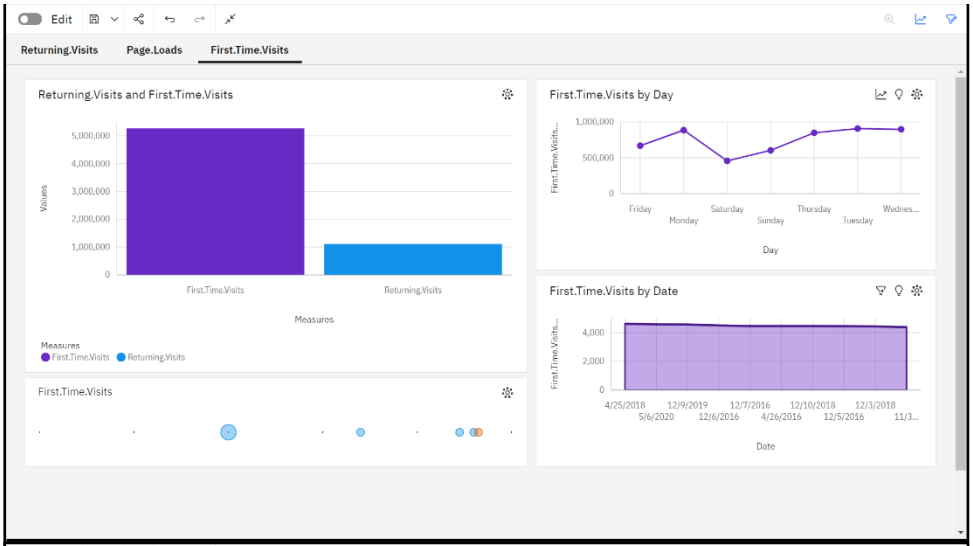

```
data = data.rename(columns={"Date": "ds", "Page.Loads": "y"})
data['ds'] = pd.to_datetime(data['ds'])
if data['y'].dtype == object:
    data['y'] = data['y'].str.replace(',', '').astype(float)
data = data.dropna(subset=['y'])
if len(data) < 2:
    print("")
else:
```

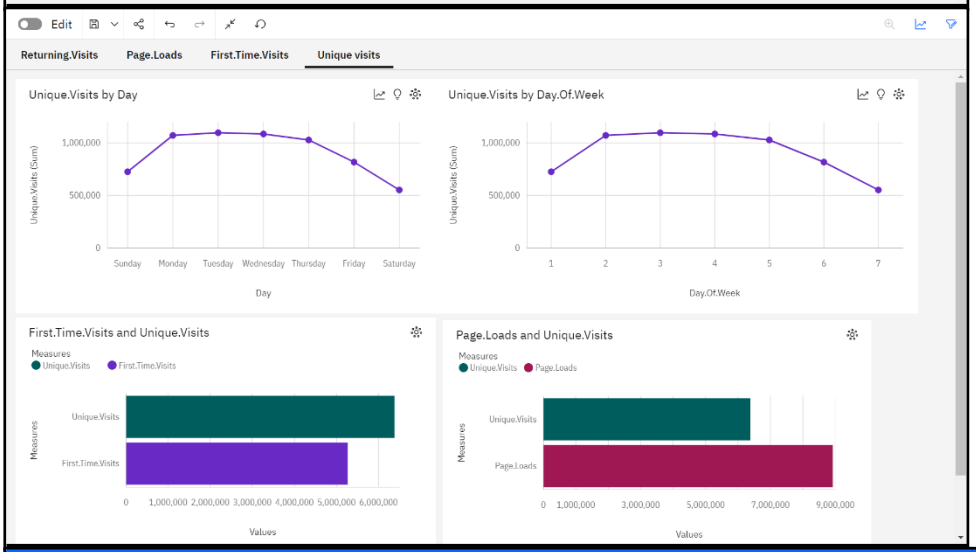
```
    model = Prophet()
    model.fit(data)
    future = model.make_future_dataframe(periods=365)
    forecast = model.predict(future)
    fig = model.plot(forecast)
```

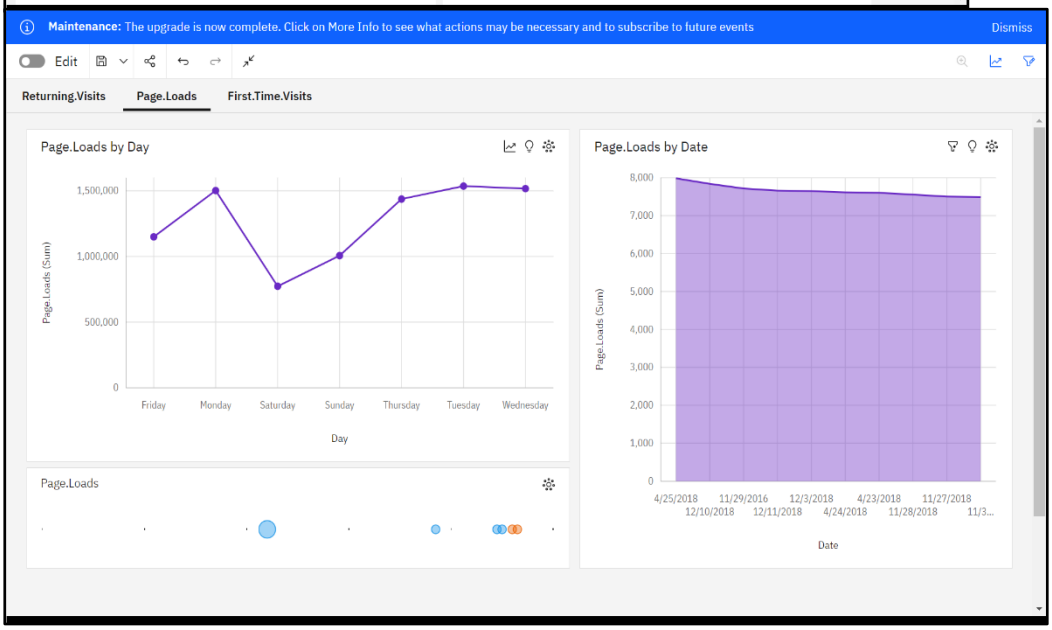
```
12:35:28 - cmdstanpy - INFO - Chain [1] start processing
12:35:28 - cmdstanpy - INFO - Chain [1] done processing
```



Z


Edit
Returning.Visits
Page.Loads
First.Time.Visits
Returning.Visits and First.Time.Visits
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
Values
First.Time.Visits
Returning.Visits
Measures
Measures
First.Time.Visits
Returning.Visits
First.Time.Visits by Day
1,000,000
500,000
0
First.Time.Visits...
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednes...
Day
First.Time.Visits
First.Time.Visits by Date
4,000
2,000
0
First.Time.Visits...
4/25/2018
5/6/2020
12/9/2019
12/6/2016
12/7/2016
4/26/2016
12/10/2018
12/5/2016
12/3/2018
11/3...
Date


Edit
Returning.Visits
Page.Loads
First.Time.Visits
Unique visits
Unique.Visits by Day
1,000,000
500,000
0
Unique.Visits (Sum)
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day
Unique.Visits by Day.Of.Week
1,000,000
500,000
0
Unique.Visits (Sum)
1
2
3
4
5
6
7
Day.Of.Week
First.Time.Visits and Unique.Visits
Measures
Unique.Visits
First.Time.Visits
Measures
Unique.Visits
First.Time.Visits
0
1,000,000
2,000,000
3,000,000
4,000,000
5,000,000
6,000,000
Values
Page.Loads and Unique.Visits
Measures
Unique.Visits
Page.Loads
Measures
Unique.Visits
Page.Loads
0
1,000,000
3,000,000
5,000,000
7,000,000
9,000,000
Values


Maintenance: The upgrade is now complete. Click on More Info to see what actions may be necessary and to subscribe to future events
Dismiss
Edit
Returning.Visits
Page.Loads
First.Time.Visits
Page.Loads by Day
1,500,000
1,000,000
500,000
0
Page.Loads (Sum)
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day
Page.Loads by Date
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
Page.Loads (Sum)
4/25/2018
12/10/2018
11/29/2016
12/11/2018
12/3/2018
4/24/2018
4/23/2018
11/28/2018
11/27/2018
11/3...
Date
Page.Loads

Z

Design Thinking Approach

:Empathize

Begin by empathizing with your website users. Conduct user interviews, surveys, and usability tests to gain insights into their needs, behaviors, and pain points. Understand their goals when visiting your site and the .challenges they face

:Actions

Data Collection: The primary action is collecting data on website visitors, .their activities, and interactions with the site using analytics tools

Analysis and Insights: Analyzing the collected data to gain insights into .user behavior, traffic sources, and areas for improvement

Optimization: Implementing changes based on the analysis to improve the .website's performance, user experience, and conversion rates

:Define

Clearly define the problem areas or opportunities for improvement based on the insights gathered. This could involve identifying specific pages with .high bounce rates, conversion bottlenecks, or user journey issues

:Objectives

The primary objective is to assess the overall performance of a website by tracking metrics such as traffic volume, user engagement, conversion rates, and bounce rates. This helps in understanding how effectively the website .is meeting its goals and where improvements may be needed

Another crucial objective is to use the insights gained from traffic analysis to make data-driven decisions. This includes optimizing content, marketing strategies, and user experience to enhance the website's .effectiveness, drive growth, and achieve specific business objectives

:Ideate

Collaboratively brainstorm solutions to the defined problems. Encourage cross-functional teams to generate creative ideas. For website traffic analysis, this might involve brainstorming ways to enhance content, .improve navigation, or optimize landing pages

Prototype

Create prototypes or wireframes of potential website changes based on the ideation phase. These can be low-fidelity representations to test concepts .before full implementation

:Actions

- Determine where your website traffic is coming from, such as search .engines, social media, or direct visits
- .Analyze which sources are sending the most visitors to your site
- Use this information to focus your marketing efforts on the most .effective channels
- Identify which pages on your website are the most popular or have the .highest bounce rates
- Assess which pages lead to the most conversions or desired actions .(e.g., sign-ups or purchases)
- Optimize or improve underperforming pages to enhance user engagemen
- Keep an eye on the conversion rate for key actions or goals on your .website (e.g., form submissions or product sales)
- .Track changes in conversion rates over time

Test

Conduct A/B testing or usability testing with real users to validate your prototypes. Analyze the impact of changes on website traffic, user engagement, and conversion rates. Iteratively refine your designs based on .user feedback and data

Implement

Once you've identified effective changes through testing, implement them on your website. Ensure that tracking mechanisms are in place to monitor .the impact of these changes on traffic and user behavior

Iterate

The Design Thinking process is iterative. Continuously gather and analyze website traffic data to assess the impact of your changes. Make further .improvements as needed to align with evolving user needs and goals

Design and Innovation Strategies .3

Data Collection and Feature Engineering .3.1

Innovation: Comprehensive Data Gathering

Implement advanced data collection techniques, including web scraping, API integration, and data enrichment, to gather diverse datasets .encompassing traffic data and user interactions

Apply innovative feature engineering techniques to extract meaningful .insights from structured and unstructured data sources

Data Preprocessing .3.2

Innovation: Data Cleansing and Transformation

Implement data cleansing and transformation procedures to handle missing .values, outliers, and data quality issues

Utilize natural language processing (NLP) and text analytics for textual .data preprocessing, enabling sentiment analysis and topic modeling

Model Selection and Training .3.3

Innovation: Hybrid Models

Employ a combination of traditional machine learning models (e.g., regression, classification) and deep learning models (e.g., neural networks) .to predict traffic trends and user behavior

Develop hybrid models that leverage the strengths of both traditional and .deep learning approaches for enhanced predictive accuracy

Geographic Analysis .3.4

Innovation: Geospatial Insights

Integrate geospatial analysis to gain insights into the geographic patterns .of traffic trends and user behavior

Implement innovative spatial visualization techniques, such as heatmaps .and geospatial clustering, to identify spatial trends

User Behavior Modeling 3.5

Innovation: Sequence Analysis

Utilize sequence modeling techniques, including recurrent neural networks (RNNs) and hidden Markov models (HMMs), to model and predict user .behavior patterns

Analyze user journeys, session durations, and entry/exit points for .improved user experience

Predictive Analytics .3.6

Innovation: Forecasting and Anomaly Detection

.Develop predictive models for forecasting traffic trends and user behavior

Implement anomaly detection techniques to identify unusual patterns that .may require immediate attention

Continuous Improvement .3.7

Innovation: Realtime Monitoring and Feedback Loop

Create realtime monitoring dashboards that provide instant updates on .traffic trends and user behavior

Establish a feedback loop to continuously improve models based on new .data and changing user preferences

Innovation: Forecasting and Anomaly Detection
.Develop predictive models for forecasting traffic trends and user behavior
Implement anomaly detection techniques to identify unusual patterns that .may require immediate attention

Continuous Improvement .3.7
Innovation: Realtime Monitoring and Feedback Loop
Create realtime monitoring dashboards that provide instant updates on .traffic trends and user behavior
Establish a feedback loop to continuously improve models based on new .data and changing user preferences

Conclusion.10
Analyzing website traffic data can provide valuable insights. The conclusion will depend on the specific data and objectives, but you might summarize key points like traffic sources, popular content, user demographics, and conversion rates to make informed decisions for .improving the website's performance

z