 **DHANALAKSHMI SRINIVASAN**

**COLLEGE OF ENGINEERING AND TECHNOLOGY**

**MAMALLAPURAM, CHENNAI - 603 104.**

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## DATA ANALYTICS WITH COGNOS

**WEBSITE TRAFFIC ANALYSIS**

**Group 1**

# A PROJECT REPORT

*Submitted by*

|  |  |
| --- | --- |
| **ARAVIND J** | 310521104012 |
| **ABINESH B** | 310521104004 |
| **KALAIYARASAN** | 310521104039 |
| **LIKITHESH NS** | - |

*in partial fulfillment for the award of the degree*

*of*

# BACHELOR OF ENGINEERING

*in*

# COMPUTER SCIENCE AND ENGINEERING

# ANNA UNIVERSITY :: CHENNAI - 600 025

**APRIL 2023-MAY 2023**



**ANNA UNIVERSITY :: CHENNAI – 600 025**

# BONAFIDE CERTIFICATE

The certificate that this project report **“WEBSITE TRAFFIC ANALYSIS**

**”** is the bonafide work of “**ARAVIND J (310521104012**), **ABINESH B (310521104004), KALAIYARASAN(310521104039**)**, LIKITHESH NS ”** who carried out the project work under my supervision.

## SIGNATURE SIGNATURE

**HEAD OF THE DEPARTMENT SUPERVISOR**

Department of Computer Science Department of Computer Science and and Engineering, and Engineering,

Dhanalakshmi Srinivasan College Dhanalakshmi Srinivasan College of of Engineering & Technology, Engineering & Technology,

Mahabalipuram. Mahabalipuram.

Submitted for the Practical Examination held on

## INTERNAL EXAMINER EXTERNAL EXAMINER

Website Traffic Analysis

**Project Definition:**

The project at hand involves analyzing website traffic data to gain insights into user behavior, popular pages, and traffic sources. The primary objective is to assist website owners in enhancing the user experience by understanding how visitors interact with the site. To accomplish this, we will undertake the following key steps:

**Analysis Objectives:**

Define the key insights we aim to extract from the website traffic data. This may include identifying popular pages, understanding traffic trends, and measuring user engagement metrics.

**Data Collection:**

Determine the data sources and methods for collecting website traffic data. This data may include metrics such as page views, unique visitors, referral sources, session duration, and more.

**Visualization:**

Plan how to visualize the insights effectively. We will leverage IBM Cognos to create meaningful dashboards and reports that provide a clear and actionable view of the website's performance.

**Python Integration:**

Consider incorporating Python code to perform advanced analysis on the website traffic data. This may include the application of machine learning models to predict future traffic trends or identify user behavior patterns.

**Design Thinking:**

**Analysis Objectives:**

**Objective 1: Identify Popular Pages**

Understanding which pages are most popular can help website owners optimize content, layout, and navigation.

**Objective 2: Analyze Traffic Trends**

Identifying traffic trends over time will provide insights into the seasonality of website traffic and help in planning for peak periods.

**Objective 3: Measure User Engagement Metrics**

Metrics such as bounce rate, time on page, and clickthrough rates can help in assessing the level of user engagement and identifying areas for improvement.

**Data Collection:**

**Data Sources:**

Website Analytics Tools: Utilize tools like Google Analytics, Adobe Analytics, or custom tracking scripts to gather data.

Server Logs: Extract data from server access logs to capture detailed information about each visitor.

**Data Collection Methods:**

Periodic Data Collection: Set up automated data collection processes to gather data regularly (e.g., daily or hourly).

Data Storage: Store collected data in a structured database or data warehouse for easy access and analysis.

**Visualization:**

**Dashboard Components:**

Utilize IBM Cognos to create visually appealing dashboards.

Key components may include line charts for traffic trends, bar charts for popular pages, and pie charts for traffic source distribution.

**Interactive Reports**:

Implement filters and interactive elements in the reports to allow users to explore data on their own.

Regular Updates:

Schedule automated report generation and distribution to relevant stakeholders.

**Python Integration:**

**Machine Learning Models:**

Explore the integration of machine learning models using Python libraries like scikitlearn or TensorFlow.

For example, predictive models can be trained to forecast future traffic trends based on historical data.

**Data Preprocessing:**

Ensure data is preprocessed, cleaned, and transformed as needed to feed into machine learning models.

**Model Evaluation:**

Implement model evaluation techniques to assess the accuracy and reliability of predictions.

**Feedback Loop:**

Continuously update models with new data to improve accuracy and adapt to changing user behavior

# INNOVATION:

Innovation in website traffic analysis is crucial for businesses and website owners to gain insights into their audience, improve user experiences, and optimize their digital marketing strategies. Here are some innovative approaches and technologies you can consider for website traffic analysis

## Machine Learning and AI:

Implement machine learning algorithms and artificial intelligence to analyze website traffic patterns, detect anomalies, and predict user behavior. AI can provide personalized recommendations and insights that can enhance user engagement.

## Predictive Analytics:

Use predictive analytics to forecast future website traffic based on historical data. This can help in planning marketing campaigns and allocating resources effectively.

## Realtime Analytics:

Realtime traffic analysis tools allow you to monitor user activities as they happen. This is crucial for ecommerce sites and online services to make immediate adjustments to improve user experience and address issues.

## User Behavior Analytics:

Dive deeper into user behavior on your website by tracking mouse movements, clicks, and scrolls. Heatmaps and session recordings can provide valuable insights into how users interact with your site.

## Voice and Conversational Analytics:

As voice search and chatbots become more prevalent,

analyzing user interactions through voice and chat interfaces can help you understand user intent and improve conversational experiences.

## Customer Journey Mapping:

Create detailed customer journey maps to visualize how users navigate through your website. Identify pain points and areas for improvement to enhance the overall user experience.

## Data Integration:

Integrate data from various sources, such as social media, CRM systems, and marketing platforms, to gain a holistic view of your audience and their interactions with your brand.

# INNOVATIVE APPROACHES TO WEBSITE TRAFFIC ANALYSIS

Analyzing website traffic is crucial for understanding user behavior, optimizing content, and improving the overall performance of your website. Innovative approaches to website traffic analysis can provide deeper insights and help you make datadriven decisions. Here are some innovative approaches you can

consider:

## User Behavior Analysis with Machine Learning:

Implement machine learning algorithms to analyze user behavior patterns, such as clustering users based on their interactions or predicting user preferences and engagement.

## Natural Language Processing (NLP):

Use NLP techniques to analyze usergenerated content, comments, and reviews to gain insights into sentiment analysis, customer feedback, and trending topics.

## Deep Learning for Image and Video Analysis:

If your website includes images or videos, employ deep learning models for image and video analysis to understand which visual content performs best and why.

## RealTime Analytics:

Implement realtime analytics to monitor website traffic as it happens. This can help you respond to emerging trends or issues quickly.

# INNOVATION IDEAS FOR WEBSITE TRAFFIC ANALYSIS

## Geographic and Demographic Insights:

Analyze the geographic locations and demographics of your website visitors. This data can help you tailor your content and marketing efforts to specific regions and target audiences.

## Voice Search Optimization:

With the increasing use of voiceactivated devices, analyze how voice search queries affect your website traffic. Optimize your content for voice search to capture this growing audience.

## Competitor Traffic Analysis:

Use competitive analysis tools to gain insights into your competitors' website traffic. Identify gaps in their strategies that you can exploit to attract more visitors to your own site.

## Customer Journey Mapping:

Create detailed customer journey maps to understand the various touchpoints and interactions users have with your website before converting. This can help you identify areas for improvement and optimization

# WHY IS INNOVATION IS NECESSARY?

**reasons:**

**Innovation is necessary in website traffic analysis for several**

1. **Stay Competitive**: The online landscape is highly

competitive. To stand out and attract and retain visitors, website owners need to

constantly adapt and improve their strategies. Innovation in traffic analysis allows them to stay ahead of the competition.

## User Experience Enhancement:

Understanding user behavior through innovative analytics tools helps improve the user experience. By identifying pain points and preferences, website owners can make necessary adjustments to make their site more userfriendly, leading to increased traffic and engagement.

## Data Security and Privacy:

As data security and privacy concerns become more prominent, innovation in traffic analysis should also focus on ensuring compliance with regulations like GDPR. This not only protects user data but also builds trust

among visitors.

# 

1. Innovation:

This documentation explains the code for visualizing website traffic data using Python. The code utilizes packages like pandas, numpy, and plotly.express to load, analyze, and visualize website traffic data.

2. Code Explanation:

2.1. **Importing the Packages:**

In this section, the necessary Python libraries are imported. These include:

`pandas` for data manipulation and analysis.

`numpy` for numerical operations.

`plotly.express` for creating interactive data visualizations.

2.2. **Loading and Inspecting the Data:**

```python

data = pd.read\_csv(r"C:\Users\DSCET\Desktop\dac\anali.csv")

```

Here, the code loads a CSV file named "anali.csv" from the specified file path into a Pandas DataFrame called `data`. This CSV file presumably contains website traffic data.

```python

data.head()

data.tail()

data.size

data.describe()

data.info()

data.isnull().sum()

```

These commands provide an initial inspection of the data:

`data.head()` displays the first few rows of the DataFrame.

`data.tail()` displays the last few rows of the DataFrame.

`data.size` returns the total number of elements in the DataFrame.

`data.describe()` provides summary statistics for the numerical columns.

`data.info()` gives information about the DataFrame's data types and nonnull values.

`data.isnull().sum()` counts and displays the number of missing values in each column.

**2.3. Data Visualization:**

```python

fig = px.scatter(data, x="Day.Of.Week", y="Unique.Visits", size='Day.Of.Week')

fig.show()

```

In this section, data visualization is performed using Plotly Express. It creates a scatter plot where "Day.Of.Week" is plotted on the xaxis and "Unique.Visits" on the yaxis. The `size` parameter is set to "Day.Of.Week," meaning the size of the data points will vary based on the values in the "Day.Of.Week" column. Finally, `fig.show()` displays the interactive scatter plot.

This code provides an example of loading and visualizing website traffic data. It serves as an initial step in understanding the data and identifying potential patterns or insights.

## Code Explanation

### Importing Required Libraries

`pandas`: A library for data manipulation and analysis.

`matplotlib.pyplot`: A plotting library for creating visualizations.

`numpy`: A library for numerical operations.

`sklearn`: A library for machine learning and data mining.

`plotly.express`: A library for creating interactive plots.

### Loading the Dataset

`df = pd.read\_csv(r"C:\Users\nithy\OneDrive\Desktop\IBM project\dac.csv")`: Reads the CSV file and loads it into a pandas DataFrame named `df`. The file path provided should be updated to the correct location of the CSV file on your system.

### Exploring the Dataset

`print(df.columns)`: Prints the column names of the DataFrame.

`print(df.head())`: Prints the first few rows of the DataFrame.

`print(df.shape)`: Prints the dimensions (number of rows and columns) of the DataFrame.

`df.info()`: Prints information about the DataFrame, including the data types of columns and memory usage.

### Preprocessing the Data

`df['Day'] = pd.to\_datetime(df['Day'])`: Converts the 'Day' column to datetime format.

`df.set\_index('Day', inplace=True)`: Sets the 'Day' column as the index of the DataFrame.

### Visualizing Time Series Data

`df['Returning.Visits'].plot()`: Plots the 'Returning.Visits' column as a time series.

`plt.xlabel('Day')`: Sets the xaxis label of the plot.

`plt.ylabel('Returning.Visits')`: Sets the yaxis label of the plot.

`plt.title('Unique.Visits')`: Sets the title of the plot.

`plt.show()`: Displays the plot.

### Clustering with KMeans

`features = ['engagement\_metric\_1', 'engagement\_metric\_2', 'engagement\_metric\_3']`: Defines the features used for Kmeans clustering.

`X = df[features]`: Extracts the feature columns from the DataFrame.

`kmeans = KMeans(n\_clusters=3)`: Initializes a Kmeans clustering model with 3 clusters.

`kmeans.fit(X)`: Fits the Kmeans model to the data.

`df['cluster'] = kmeans.labels\_`: Adds a new column 'cluster' to the DataFrame with the cluster labels assigned by the Kmeans model.

### Linear Regression Modeling

`features = ['feature1', 'feature2', 'feature3']`: Defines the features used for linear regression.

`target = 'target\_variable'`: Defines the target variable for linear regression.

`X = df[features]`: Extracts the feature columns from the DataFrame.

`y = df[target]`: Extracts the target variable from the DataFrame.

`X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)`: Splits the data into training and testing sets.

`model = LinearRegression()`: Initializes a linear regression model.

`model.fit(X\_train, y\_train)`: Trains the linear regression model on the training data.

`y\_pred = model.predict(X\_test)`: Makes predictions using the trained linear regression model on the test data.

This code performs data loading, preprocessing, visualization, clustering with Kmeans, and linear regression modeling on the given dataset. It provides insights into the data and creates predictive models for further analysis.

**Conclusion**

In this documentation, we have explored a Python code example that focuses on analyzing website traffic data. The project's main objectives include gaining insights into user behavior, identifying popular pages, and understanding traffic trends and user engagement metrics. The code workflow encompasses data loading, exploration, preprocessing, time series analysis, K-means clustering, and linear regression modeling.

We began by importing essential libraries like pandas, numpy, and plotly.express to facilitate data manipulation, visualization, and analysis. The dataset, assumed to contain website traffic data, was loaded using pandas from a CSV file. The data was inspected to gain an initial understanding, checking for the structure of the data and any missing values.

A key part of the code involved data visualization, where we used Plotly Express to create an interactive scatter plot. This visualization showcased the relationship between the day of the week and the number of unique visits, potentially revealing patterns in website traffic.

For advanced analysis, the code introduced K-means clustering to segment the data into clusters based on engagement metrics. Additionally, it implemented linear regression to make predictions based on selected features, providing insights into relationships between variables.

The project further emphasized the importance of innovation in website traffic analysis, suggesting cutting-edge approaches and technologies. These included machine learning and AI for predictive analytics, real-time analytics for immediate adjustments, and user behavior analysis through mouse tracking and heatmaps. It also recommended the integration of data from various sources to gain a holistic view of user interactions.

In conclusion, the code and project outline demonstrated the essential steps in analyzing website traffic data, from data loading to visualization and advanced analysis. Innovations in data analysis and visualization techniques can provide valuable insights to improve user experiences and enhance digital marketing strategies. Analyzing website traffic is essential for staying competitive, enhancing the user experience, and ensuring data security and privacy compliance. Continuous innovation in this field can lead to better decision-making and more successful online ventures.