

DATA ANALYTICS WITH COGNOS

WEBSITE TRAFFIC ANALYSIS

ABSTRACT:

Analyzing website traffic data is essential for optimizing user experiences and making informed decisions for website improvements. There is a great potential for such analyses to bring market value to the stakeholders by helping them improve web content quality and delivery.

PROBLEM STATEMENT:

To collect and analyze data on user traffic in websites by applying IBM Cognos analytics tool to obtain valuable data-driven insights on user preferences, understand user interaction and enhance user experience.

PROBLEM DEFINITION:

This 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.

DESIGN THINKING:

1. Analysis Objectives:

The objective of the project is to extract the following key insights based on the given website traffic dataset.

- **Bounce rate analysis:** To analyse the bounce rate for different pages. A high bounce rate may indicate that visitors are not finding what they're looking for or that there's a problem with the page's content or design.
- **Time-series analysis of traffic trends:** To examine traffic patterns over time, including daily, weekly, and seasonal fluctuations which helps in making business decisions on augmenting additional resources to the server to better handle user traffic.
- **Website traffic forecasting:** To predict future website visitor trends, volumes, and user behaviour based on historical data which helps in resource allocation and capacity planning.
- **Conversion Rate Analysis:** To analyse the behavior of users who return after an extended period can provide insights into their conversion rates compared to more active users. This information can inform conversion optimization strategies.
- **Popularity of the site:** To derive insights on the popularity of the site based on visitor behavior (inactive and returning visitors) to help improve content quality and overall user experience. This could also help tailor marketing and content strategies.

- **Identification of Technical Issues:** To identify and eliminate technical issues, such as slow page loading, which lead to prolonged inactivity thereby addressing performance issues.

2.Data Collection:

Generic website traffic analysis involves collection of user traffic data from various sources such as server logs, content management systems, content delivery networks, third party cookies, user surveys and feedback etc... The below dataset is the major source of data for the analysis phase. Given below is the overall description of the entities, attributes and tuples available in the dataset.

Dataset link: <https://www.kaggle.com/datasets/bobnau/daily-website-visitors>

This file contains 5 years of daily time series data for several measures of traffic on a statistical forecasting teaching notes website. There are 2167 rows of data spanning the date range from September 14, 2014, to August 19, 2020. A visit is defined as a stream of hits on one or more pages on the site on a given day by the same user, as identified by IP address. The following are the attributes in the above mentioned dataset.

- **Row (Type : Integer) :** Denotes the row number which is used to uniquely identify a particular data entry.
- **Day (Type : String) :** Consists of the day of the week that pertains to that row.
- **Day_of_week (Type : Integer) :** Denotes the day of the week in numerical form (numbered from 1 through 7 for each day of the week in that respective order).
- **Date (Format : dd/mm/yyyy) :** Denotes the date of entry (The entire data ranges over a period of 5 years).
- **Page Loads (Type : Integer) :** No. of page loads handled by the web server on that particular day on a daily basis.
- **Unique visits (Type : Integer) :** Daily number of visitors from whose IP addresses there haven't been hits on any page in over 6 hours. A visit is classified as "unique" if a hit from the same IP address has not come within the last 6 hours.
- **First time visits (Type : Integer) :** Number of unique visitors who do not have a cookie identifying them as a previous customer. Returning visitors are identified by cookies if those are accepted.
- **Returning visits (Type : Integer) :** Number of unique visitors minus first time visitors

3.Visualisation of Collected Data:

Creating meaningful dashboards and reports using IBM Cognos for visualizing insights from website traffic data requires a well-structured plan. Here's a step-by-step plan to guide you through the process:

- **Objectives of Data Visualization :** To visualise user demographics, website popularity and traffic magnitude assessment metrics.

- **Data Preprocessing :** To clean , structure and integrate data into a format that can be easily accessed by IBM Cognos.
- **Design of the data model :** To design a data model that represents the website traffic data. This may involve defining data sources, relationships, and calculations.
- **Create data queries :** Use IBM Cognos Query Studio or Report Studio to create data queries that extract the specific metrics and dimensions required for the visualizations.
- **Develop visualizations :** Use IBM Cognos Report Studio or Dashboard Studio to create a variety of visualizations, including bar charts, line graphs, pie charts, tables and funnel charts etc... to display trends in page views, traffic sources, distribution of traffic sources or demographics and user engagement over time and to track conversion rates and the user journey.
- **Define Filters and Parameters and create dashboards:** Implement interactive filters and parameters in the dashboards and reports. This allows users to customize the data they see, facilitating deeper exploration. Dashboards provide a consolidated view of key insights and allow users to explore the data dynamically.

By following this plan, IBM Cognos can be leveraged to create meaningful dashboards and reports that enable website owners to gain actionable insights from their traffic data, make informed decisions, and enhance the user experience.

4. Python Integration for further analysis:

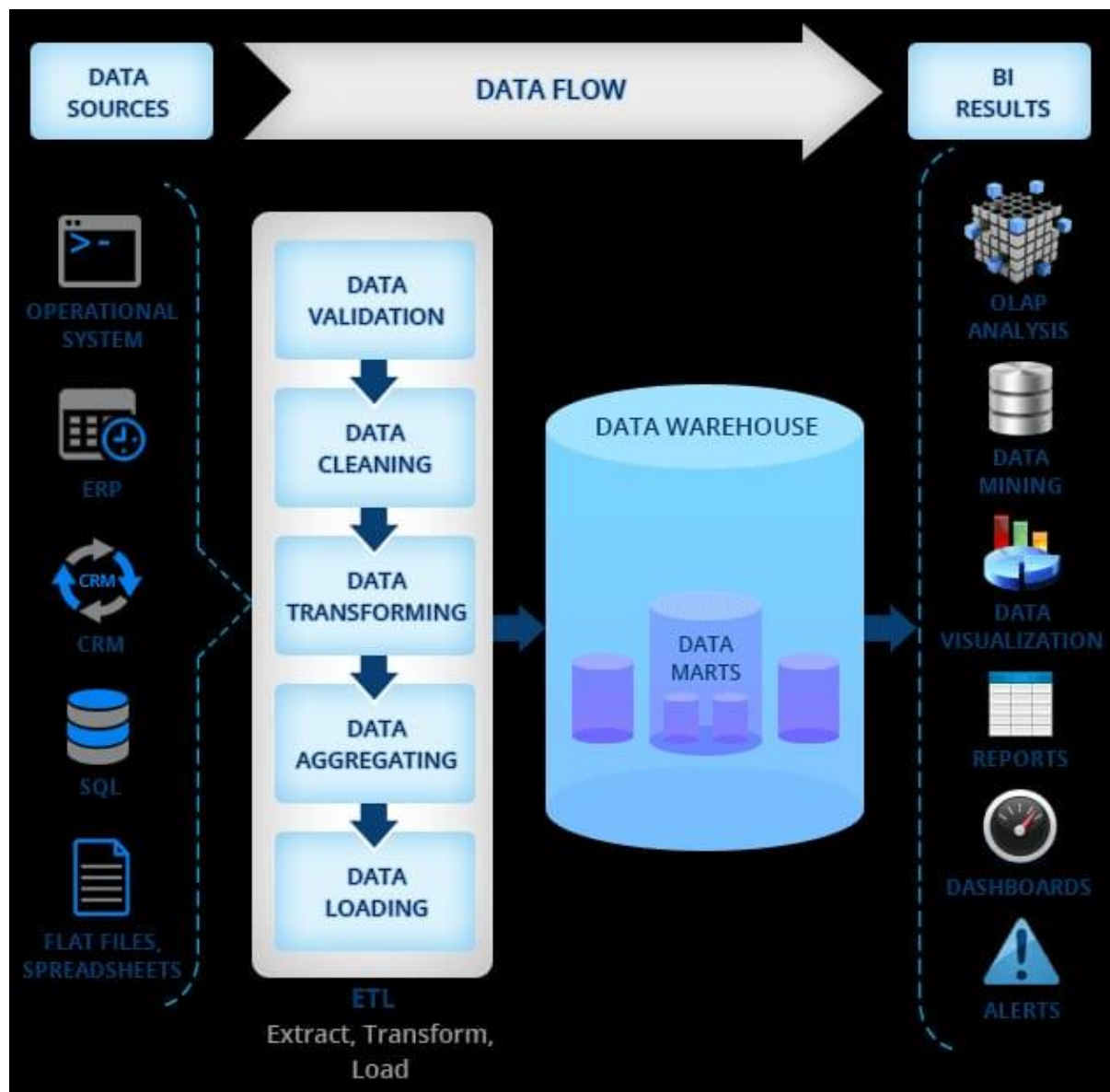
Incorporating machine learning models to predict future traffic trends or user behavior patterns can provide valuable insights for website optimization and business decision-making. Here are several possible ways to leverage machine learning in this context:

- **Time Series Forecasting :** Use time series forecasting models such as ARIMA, Exponential Smoothing, or Prophet to predict future website traffic trends based on historical data. This can help you anticipate traffic spikes, seasonality, and overall traffic patterns.
- **Regression Analysis :** Apply regression models to predict specific metrics like page views, bounce rates, or conversion rates based on various features such as marketing spend, advertising channels, or content quality. Linear regression, polynomial regression, and multiple regression are commonly used techniques.
- **Time Series Anomaly Detection :** Detect anomalies in website traffic data using machine learning models. Unusual spikes or drops in traffic can indicate technical issues, bot attacks, or significant user behavior changes.
- **User Churn Prediction :** Predict user churn or attrition by building models that identify patterns leading to users discontinuing their engagement with the website. This can inform retention strategies.
- **Online Learning :** Implement online learning models that adapt to changing user behavior patterns in real-time. These models can continuously update and retrain as new data becomes available.

- **Data Integration and Fusion :** Combine website traffic data with other data sources, such as CRM data, social media data, or external market data, to enhance the predictive power of your models.

Incorporating machine learning models for predicting future traffic trends or user behavior patterns can provide a competitive advantage by enabling data-driven decision-making and proactive strategies to enhance the user experience on your website.

PROCESS FLOW:



CONCLUSION:

This website traffic analysis would offer several advantages for website owners, marketers, and businesses. By collecting and analyzing data related to how users interact with a website, this would help stakeholders gain valuable insights and make informed decisions to improve visitors online presence.