Daily Website Traffic Trends: A Time Series Analysis

Final Project-Technical Report

Group 9 Team Members:
1. Priya Kumari (kumarip@iu.edu)
2. Ayesha Tajammul Ahmed Mulla (amulla@iu.edu)
3. Aazin Asif Shaikh (aazshaik@iu.edu)

Application URL

We have hosted application on PythonAnywhere that allows users to write, run, and host Python applications in the cloud. It provides an integrated development environment (IDE) that supports various Python libraries and frameworks, making it easier for developers to build and deploy their applications without worrying about server management. The below link is valid for 3 months as the account is a Beginner account on PythonAnywhere with limited capabilities.

URL: http://amulla.pythonanywhere.com/

GitHub URL

We have uploaded the application source code with github.iu.edu at following URL.
URL: https://github.iu.edu/kumarip/TSA_amulla_kumarip_aazshaik

Project Purpose

The purpose of this project is to develop an application that can analyze website traffic data and provide valuable insights to website owners and managers. The application's main objective is to identify patterns, trends, and seasonality in the data to help stakeholders predict future website traffic, optimize website performance, plan marketing strategies, and forecast revenue accurately. The application will also feature an interactive dashboard, which will enable stakeholders to visualize the data and make informed decisions based on observed patterns. The target stakeholders for this project include marketing and advertising teams, IT and web development teams, as well as executives and senior-level stakeholders who require a high-level overview of website traffic trends to make strategic decisions.

Project Built

Our web-based time-series application is composed of multiple pages, each with a specific purpose. The landing page includes sign-up and login buttons for user authentication. The data visualization page displays various plots, including ACF/PCF plots, line plots, and a histogram that depicts weekly visit patterns. Users can access these plots without logging in or signing up.
The forecasting page provides users with a tabular view and a live forecast plot page, which offers detailed analysis. However, to view the forecast, users must log in or sign up.
Our application follows the Model-View-Controller (MVC) architecture, and we source our data from a CSV file obtained from Kaggle. The backend of our application is implemented using Flask, which facilitates all connections and deployment. The frontend of our application is designed with HTML, CSS, and JavaScript to create an intuitive user interface. We have integrated interactive features like Plotly for graph views, which enables users to explore and interact with the plots more efficiently. Finally, we have deployed our application on PythonAnywhere, ensuring easy access for users.

About Data

The provided dataset is in the form of a CSV file containing 5 years of daily time series information on various measures of traffic for a website. The data includes daily counts of page loads, unique visitors, first-time visitors, and returning visitors, and is categorized based on the day of the week and calendar. The dataset consists of 2167 rows of data spanning from September 14, 2014, to August 19, 2020. The definition of a visit is based on the IP address and is classified as "unique" if there hasn't been a hit from the same IP address in the last 6 hours. This dataset is suitable for time series analysis as it captures daily web traffic and can be further analyzed for specific periods.
The dataset is available at: https://www.kaggle.com/datasets/bobnau/daily-website-visitor.

Functionalities

1. Landing Page (Home Page): The landing page serves as the home page of our web application, where users can navigate to different pages. It consists of two buttons: sign-up and login. The sign-up button is for first-time visitors who need to register and create an account for personalized console and secure access. The login button

is for returning users who can log in with their credentials to access the application. It also consists of visualization and forecasting buttons.
2. Data Exploration and Visualization Page: Clicking the "Visualization" button on the landing page directs the user to the Data Exploration and Visualization page. It displays various plots, including ACF/PCF plots, interactive line plots, and a histogram that depicts weekly visit patterns.
3. Regression Analysis and Forecasting Page: The Regression Analysis and Forecasting page is where users can input the number of days for which they want to forecast the unique visits. Clicking on the "Forecast" button displays a tabular view of data, showing the unique visit values at the corresponding dates. If users do not want to go through every row in the table, there is another feature available.
4. Live Forecast Plot Page: The live forecast plot is an interactive visual plot of the predicted data that users can play with and perform detailed analysis. This is available as a separate page.

Team Evaluation

Name	Evaluation
Priya Kumari	During our project, we collaborated equally in various tasks. My contributions mainly focused on decomposing TS data, implementing LSTM and AR models, and working on the visualization page during integration. I was able to apply my experience effectively by selecting suitable models and thoroughly understanding the data, which helped us create accurate predictions and compelling visualizations. I found the visualization aspect particularly engaging, as it required a mix of creativity and precision to present complex data in a comprehensible way. In future projects, we should pay more attention to data preprocessing to ensure consistency and accuracy in our models and visualizations. Additionally, exploring more advanced models could further enhance the accuracy of our predictions. Overall, I enjoyed working on this project, and I found it to be an excellent opportunity to learn and grow.
Ayesha Mulla	Working on this project was an excellent learning experience for me. I spent a significant amount of time on learning and building interactive visualizations for better user experience. Looking ahead, we could also include more advanced TS Models that we have worked on for user to compare and analyze

	the predictions. In terms of my contribution to the teamwork, I performed data analysis, including checking for stationarity and white noise and conducting time series modeling using ARIMA. Additionally, I worked on the initial design of the web architecture layout using Figma, which helped to visualize the application's structure and functionality. Finally, I was responsible for the development of the User Interface, which involved creating the design and implementing it in code. I also worked on deploying the web application on the cloud server. Overall, I am extremely satisfied with the work we accomplished together as a team in delivering the project from vision to reality.
Aazin Shaikh	While developing this project, I think the concepts and techniques learned throughout the semester from lab sessions, coding practices and videos helped a lot. We were able to implement the TSA models for our data and provide visualizations which was a major goal for us. The teamwork was great and everyone made sure to contribute equally. My contribution towards this project was in taking the raw data and preprocessing it to make it suitable for modelling. I also helped in creating various time series visualizations. It was a bit tedious to research about the differen types of models and which one would be good for our data but I performed that well and also made contribution towards the write ups of each phase. Coming to what can be improved, I believe that a more user-friendly interface can be made to enhance user interaction. Additionally, more visualization plots can be added and more advanced models can be experimented with to get deeper insights into our data. All this can be implemented in the future scope of our project. In conclusion, I learned a lot while developing this project and I would say the team put great efforts to make this.