**PHASE 5 PROJECT SUBMISSION**

**PROJECT 1 - WEBSITE TRAFFIC ANALYSIS**

**TEAM MEMBERS:**

1. Nethaji K – 2021504531
2. Rohith M U – 2021504538
3. Sripushkalai S – 2021504547
4. Tharunsarwin U G – 2021504554

**PROBLEM DEFINITION:**

The project involves analyzing website traffic data to gain insights into user behavior, 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.

**PHASE OBJECTIVE:**

Outline the project's objective, design thinking process, and development phases. Describe the analysis objectives, data collection process, data visualization using IBM Cognos, and Python code integration. Explain how the insights from the analysis can help website owners improve user experience.

**DATASET LINK:**

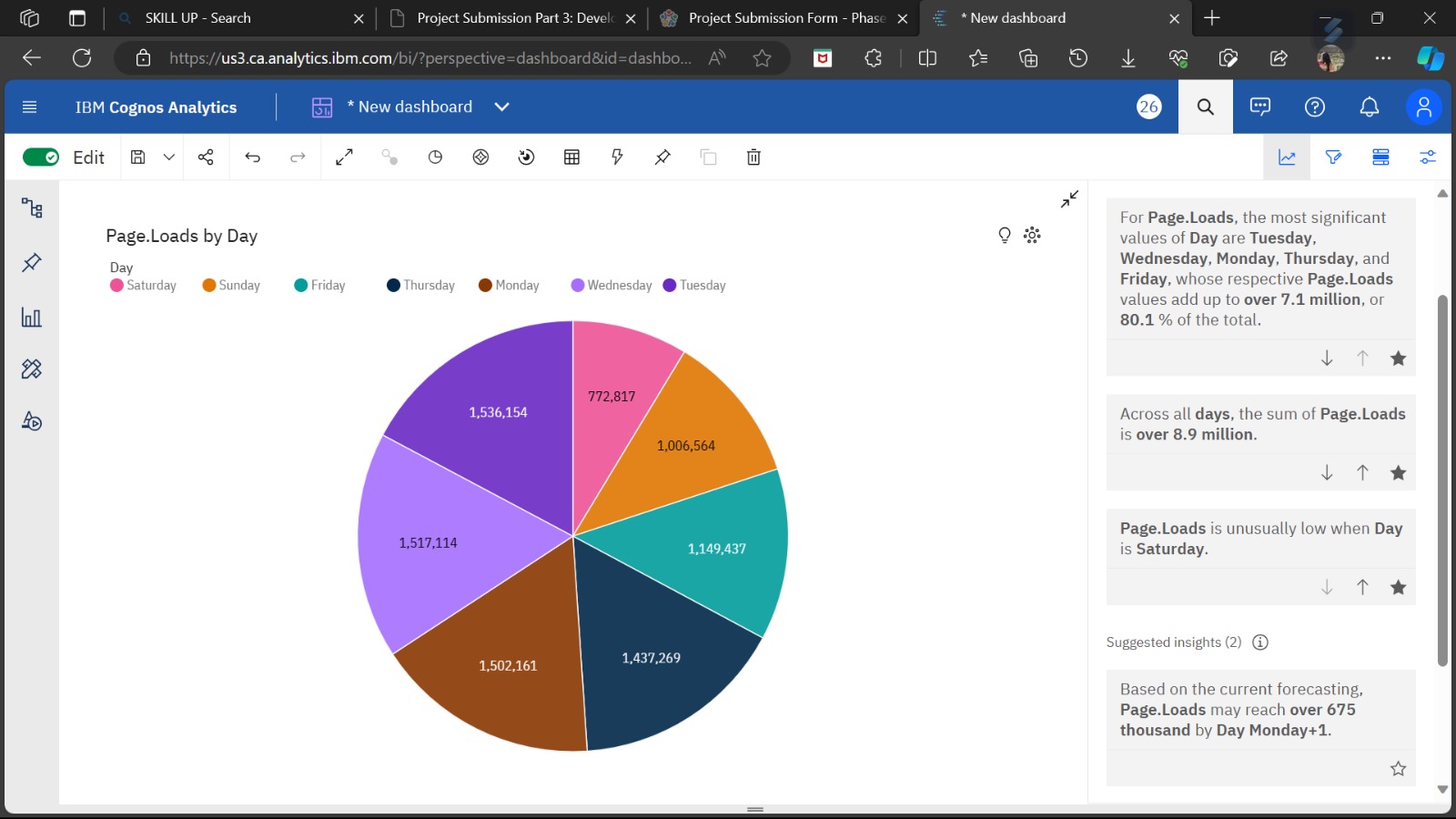
[**https://www.kaggle.com/datasets/bobnau/daily-website-visitors**](https://www.kaggle.com/datasets/bobnau/daily-website-visitors)

**DATA COLLECTION AND VISUALIZATION:**

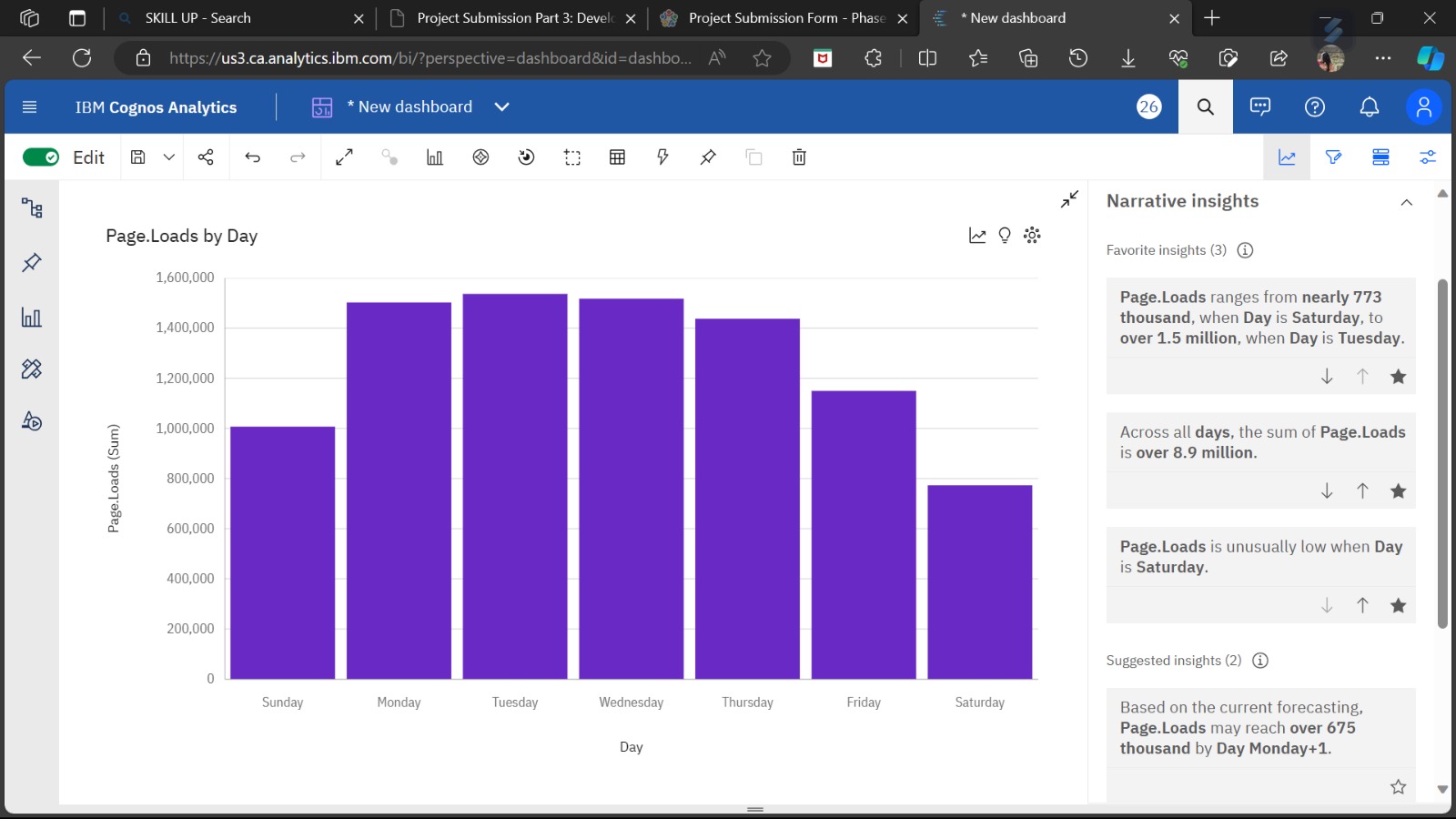
The input data file is obtained. Then the input data is pre-processed. Preprocessing process involves tasks like handling missing values, removing duplicates, and standardizing formats. This process ensures data quality and improves the accuracy of analytical results.

**After executing the code for data preprocessing, the following are the insights obtained from IBM Cognos:**

* A pie chart was plotted for number of pages loaded daily:



* It is observed that the number of pages loaded was high on Monday, Tuesday and Wednesday.
* Least number of pages were loaded on Saturday.
* A bar graph is plotted for number of webpages loaded daily:



**ANALYSIS OF PREDICTIVE MODELS:**

1. **Linear Regression:**

**Strengths:**

* Simple and easy to interpret.
* Assumes a linear relationship, which can be appropriate for certain types of data.
* Computationally efficient and quick to train.

**Weaknesses:**

* Assumes a linear relationship, which may not capture complex, non-linear patterns often seen in website traffic data.
* Sensitive to outliers.

1. **Support Vector Regression (SVR):**

**Strengths:**

* Capable of capturing non-linear relationships through the use of kernel functions.
* Effective in high-dimensional spaces and can handle complex relationships.
* Robust to outliers.

**Weaknesses:**

* Requires careful tuning of hyperparameters, which can be time-consuming.
* May be computationally more expensive compared to linear regression.

1. **Decision Tree Regression:**

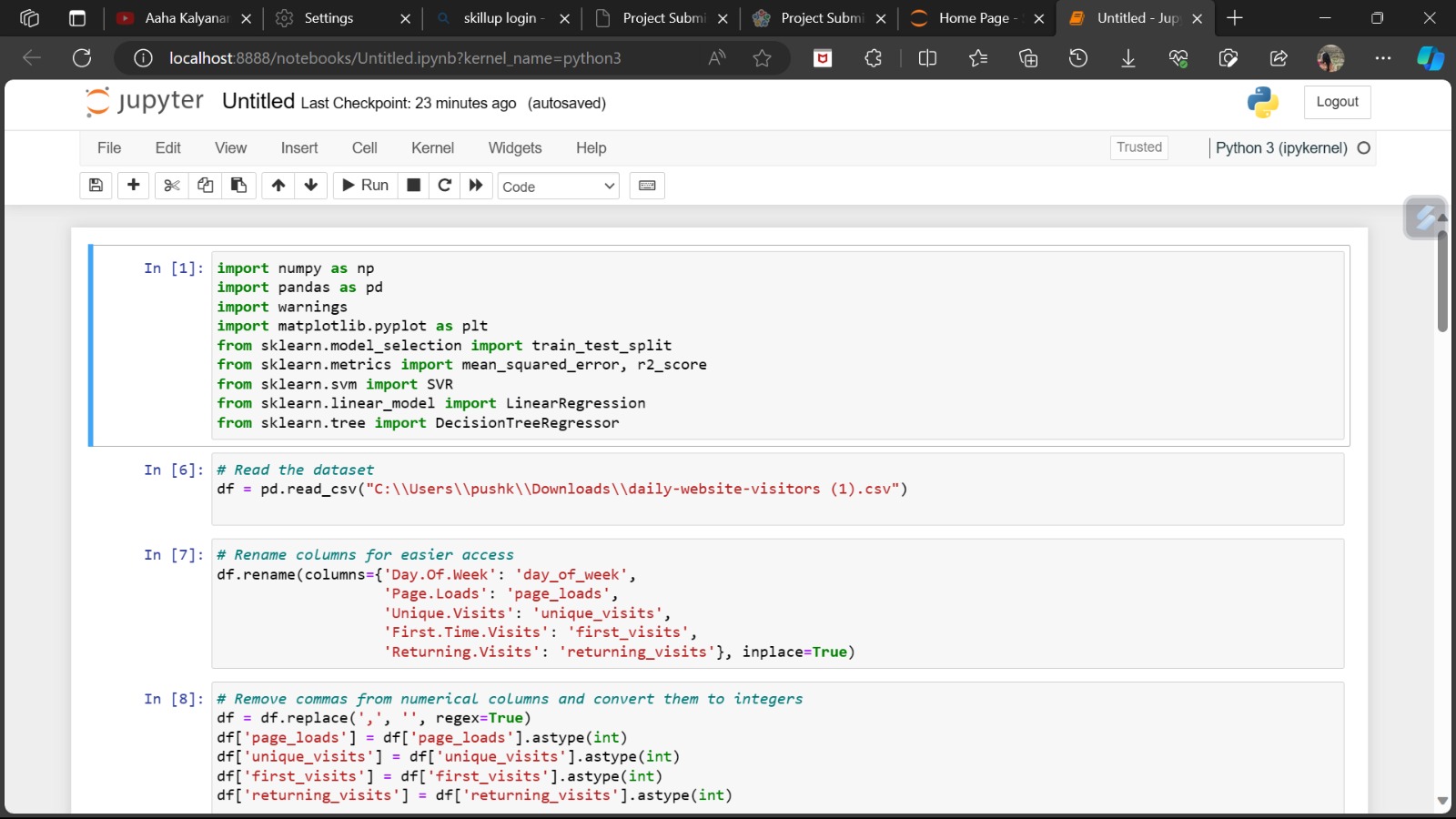
**Strengths:**

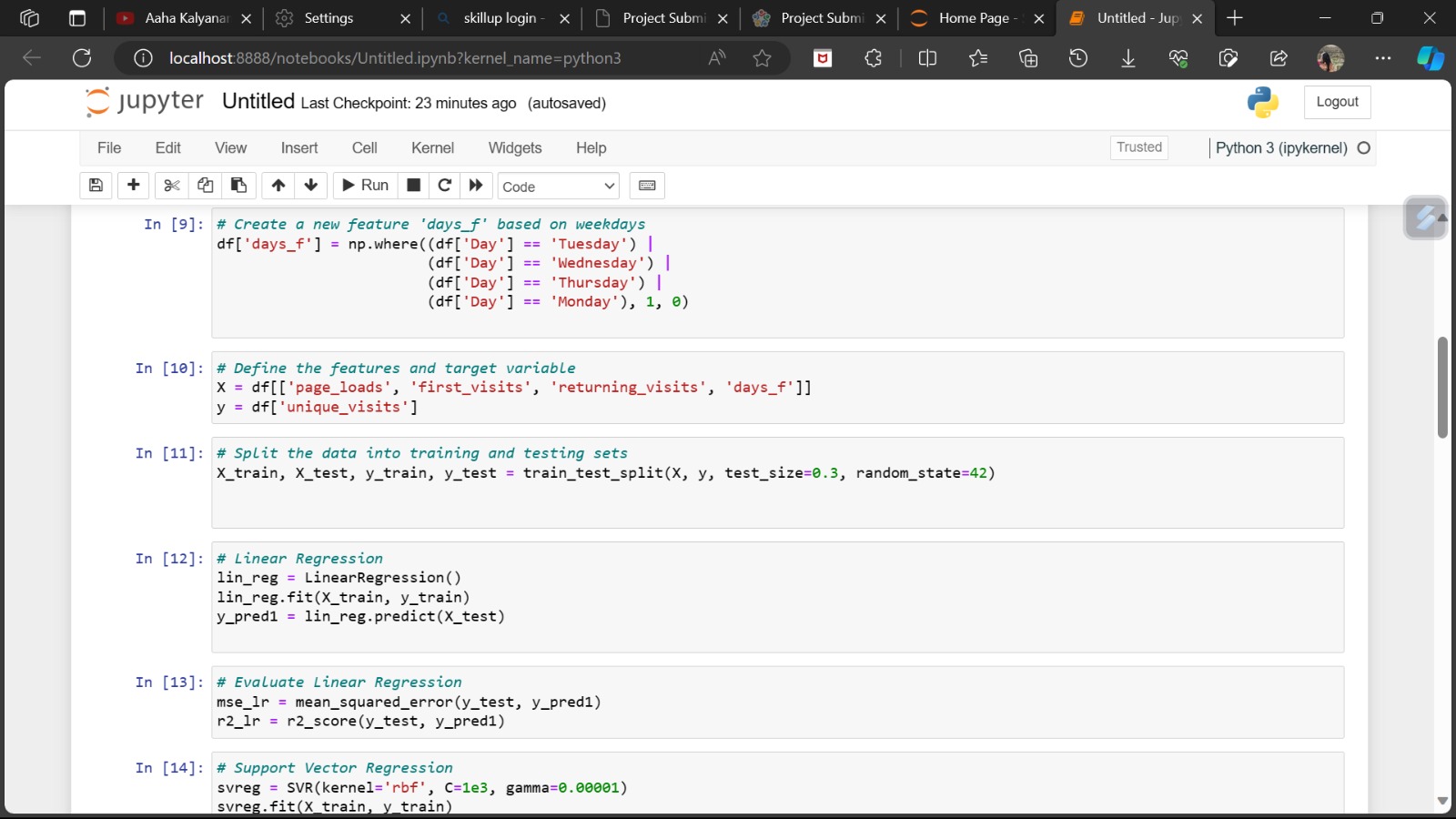
* Can model complex, non-linear relationships effectively.
* Provides interpretable results, making it easy to understand which variables are influential.
* Can handle both categorical and numerical variables.
* Robust to outliers.

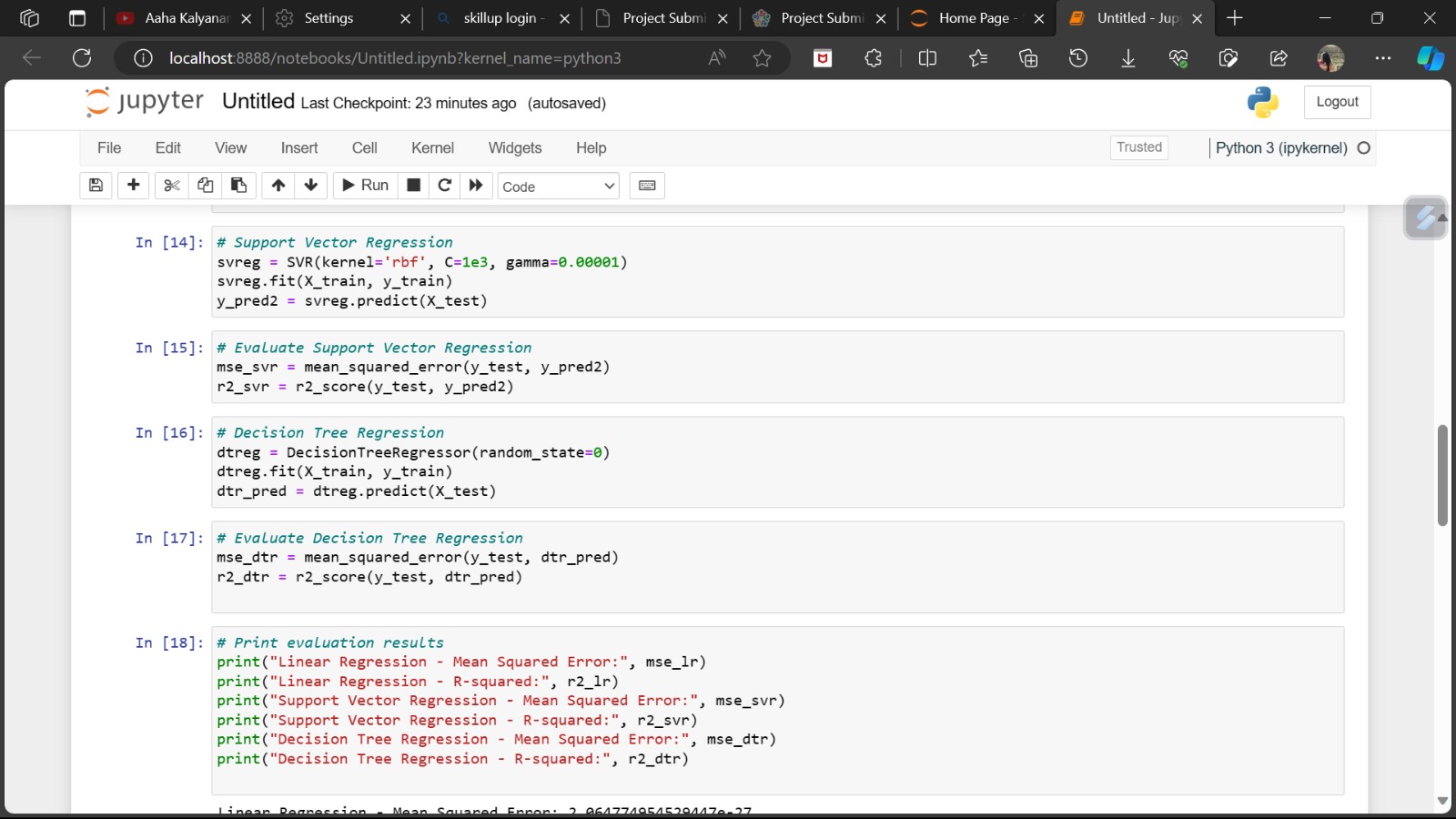
**Weakness:**

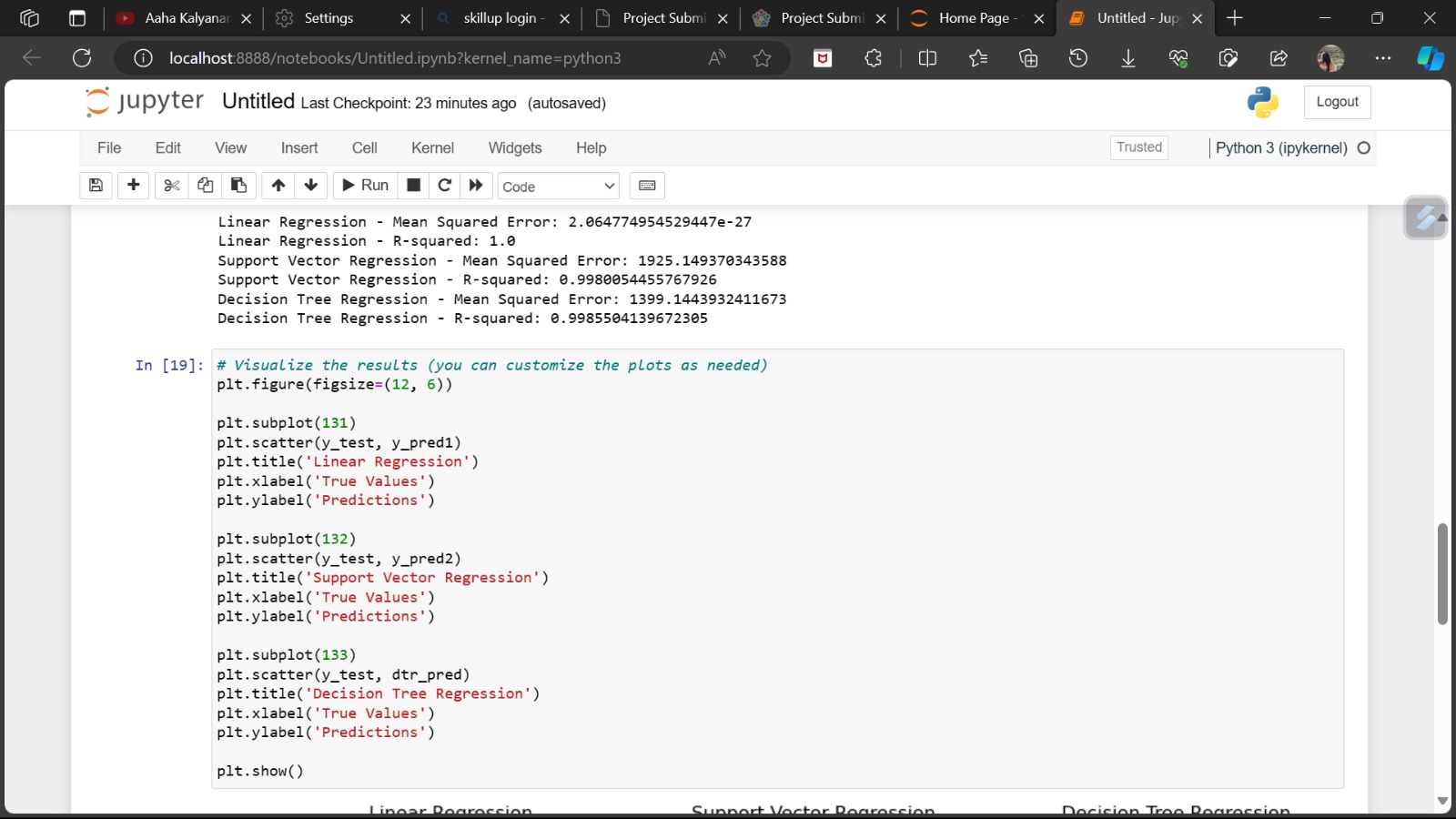
* Can be prone to overfitting, which requires careful tuning of parameters or the use of ensemble methods like Random Forests.

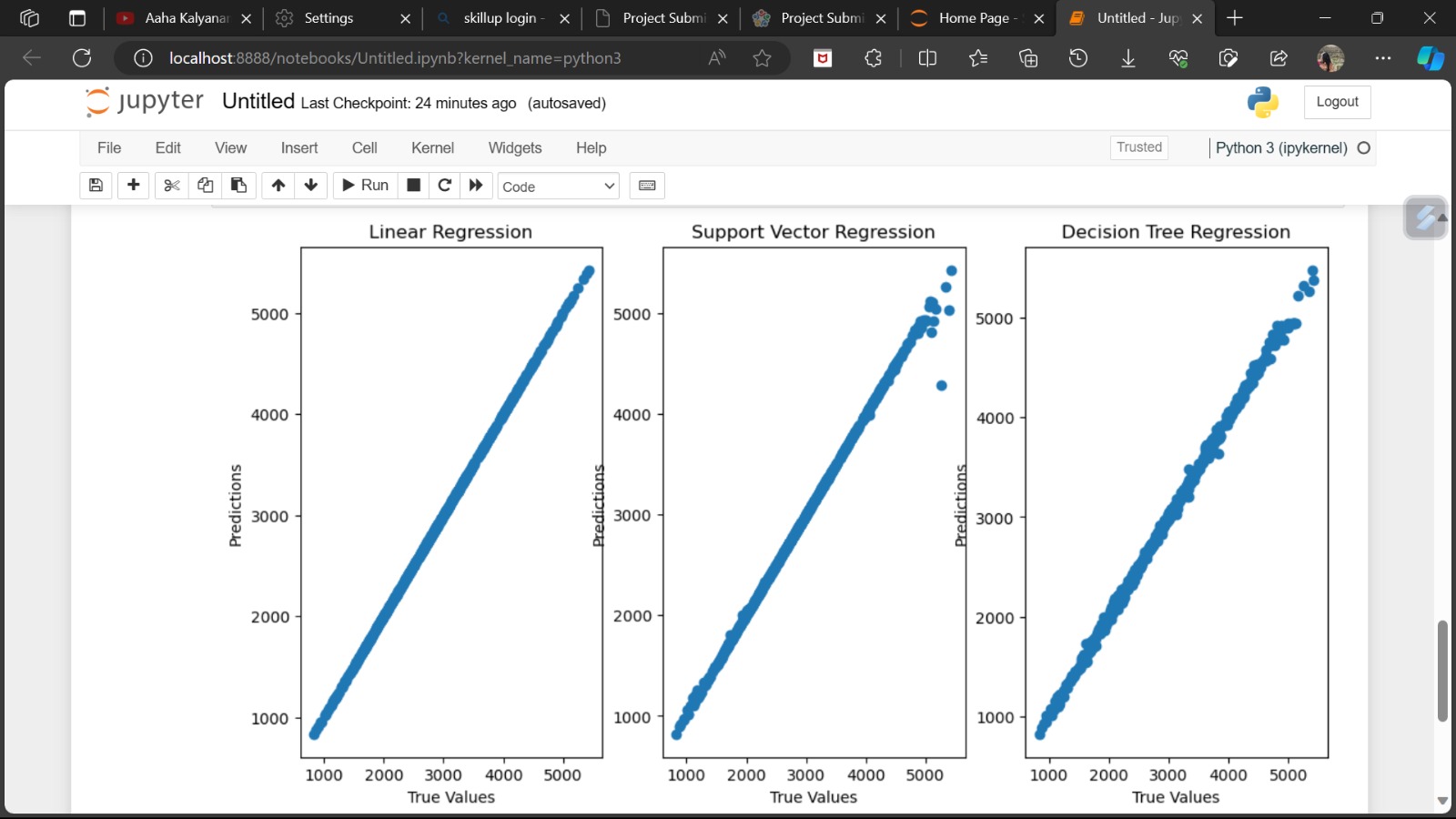
**PYTHON CODE:**

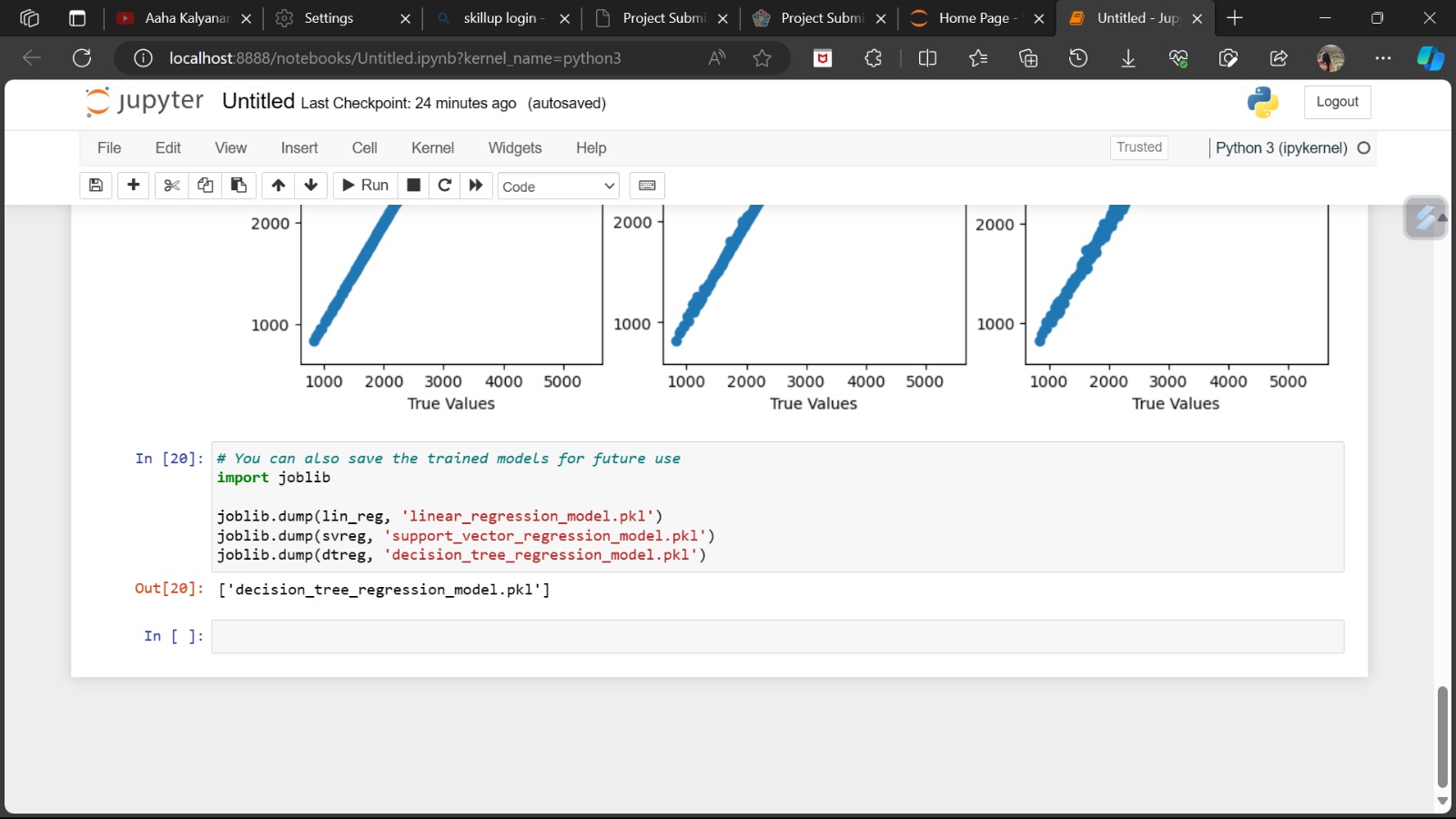












**USE OF THESE ANALYSIS:**

By examining user behaviour, traffic patterns, and engagement metrics, owners can identify areas for improvement. For instance, if analysis reveals high bounce rates on certain pages, it may indicate a need for more engaging content or a better layout. Understanding popular content can help in creating more of what resonates with visitors. Analysing traffic sources can guide marketing efforts, focusing on channels that bring in the most valuable traffic. Additionally, insights can highlight areas for technical optimization, ensuring faster load times and smoother navigation. Overall, data-driven insights empower website owners to make informed decisions, leading to a more user-friendly and satisfying experience for visitors.

**CONCLUSION:**

Thus we have built a predictive model in python using various machine learning algorithms which predicts website traffic based on the website load data of the users and different visualisations were carried out in IBM Cognos to analyse the important features.