A

PROJECT REPORT

ON

**“EcoTrack”**

SUBMITTED BY

**Ankit Girdhar Bohare**

**Seat no. 10796**

UNDER THE GUIDANCE OF

**Mrs. Swati Bhat**

SAVITRIBAI PHULE PUNE UNIVERSITY (SPPU)

MASTER OF COMPUTER APPLICATIONS



DR. D. Y. PATIL UNITECH SOCIETY’S

DR. D. Y. PATIL INSTITUTE OF MANAGEMENT AND RESEARCH, PIMPRI, PUNE-18

2024-2025

**Dr. D. Y. Patil Unitech Society’s**

**Dr. D.Y. PATIL INSTITUTE OF MANAGEMENT & RESEARCH  
Sant Tukaram Nagar, Pimpri, Pune-411018**

**Recognized by the Savitribai Phule Pune University (SPPU-IMMP014220) AISHE CODE C-42109**

|  |  |  |
| --- | --- | --- |
| Recipient of the "Best College Award" of Savitribai Phule Pune University | Accredited by NAAC with "A++ Grade” (CGPA 3.52) | MBA & MCA Programme  Accredited by NBA |
|  |  |  |

**CERTIFICATE**

This is to certify that **Ankit Girdhar Bohare** has successfully completed the project on**‘‘EcoTrack’’**as a partial fulfillment of his **Master of Computer Applications (MCA-II Sem-III)** under the curriculum of **Savitribai Phule Pune University, Pune**  for the academic year 2024-25

Mrs. Swati Bhat Dr. Shikha Dubey Dr. Vishal Wadajkar

**Project Guide H.O.D. MCA Director**

Signature Signature

Name Name

**Internal Examiner**  **External Examiner**

Date : Date :

# Acknowledgement

The success and final outcome of the “**EcoTrack**” project were achieved thanks to the invaluable guidance and assistance from many individuals. We are deeply grateful for the support we received throughout the project’s completion. Our accomplishments are a direct result of this guidance, and we are truly thankful.

We sincerely thank to the Director **Dr. Vishal Wadajkar**, HOD **Dr. Shikha Dubey** and to my project guide **Mrs. Swati Bhat**, for providing us an opportunity to do the project work and give us all support and guidance which made us completes the project duty.

We are deeply indebted to our project guides for their continuous interest in our work and for steering us through the project by providing essential information necessary for developing an effective system.

We are also immensely grateful for the constant encouragement, support, and guidance from all the teaching staff, which played a significant role in the successful completion of our project. Furthermore, we extend our sincere appreciation to all the laboratory staff for their timely assistance.

Additionally, we would like to acknowledge the support of our families and friends, whose encouragement and understanding provided us with the motivation to persevere through the challenges of this project. Their unwavering belief in our capabilities helped us maintain focus and dedication, allowing us to achieve our goals.

# Table of Content

|  |  |  |
| --- | --- | --- |
| **Sr no.** | **Title** | **Page no.** |
| **1** | **Introduction to Proposed System**   1. Problem Definition 2. System Overview 3. Project Functionalities with Module Specification 4. Operating Environment (H/W & S/W Requirement Specification) |  |
| **2** | **Overview of the Proposed System**   1. Proposed System 2. Objectives of the System 3. Feasibility Study 4. User Requirement Specification |  |
| **3** | **System Analysis &Design**   1. Data Flow Diagram ( Context Level Diagram) 2. Class Diagram/ ERD 3. Activity Diagram 4. Data Dictionary with Table Specification 5. Used Case Diagram |  |
| **4** | **User Manual**   1. Operational Instructions 2. Input/output Screens 3. Reports |  |
| **5** | **System Limitation** |  |
| **6** | **Future Enhancement and Conclusion** |  |
| **7** | **Bibliography and Glossary** (Definitions, Acronyms and Abbreviations used in the Proposed System) |  |

**Introduction to Proposed System**

**1) Problem Definition**

The global economy is complex and interconnected, and one of the key metrics for understanding a country's economic health is its Gross Domestic Product (GDP). Understanding the factors affecting GDP can lead to better policy-making, improved economic forecasting, and more effective international cooperation. The problem that this project seeks to solve is providing a robust, analytical tool that will allow for the detailed exploration and prediction of GDP values across different nations. By utilizing available economic data, the project aims to build a predictive model that can identify relationships between various factors and GDP, with a focus on visualizing these patterns and helping policymakers, economists, and researchers make informed decisions.

**2) System Overview**

"EcoTrack" is a data analysis and prediction system focused on GDP analysis. The system uses economic data from a variety of countries, including factors such as population, area, literacy rates, industry structure, and more, to predict GDP per capita. The system includes data preprocessing, exploratory data analysis, and machine learning models for predicting GDP. Users can visualize the relationships between different variables and explore trends across regions and countries. The system also includes model evaluation and comparison with various machine learning algorithms.

**3) Project Functionalities with Module Specification**

The system consists of several core functionalities:

* Data Import and Preprocessing: This module loads and cleans the dataset, handling missing values and transforming data into a usable format for analysis.
* Exploratory Data Analysis (EDA): Includes visualizations like correlation heatmaps, scatter plots, and bar charts to analyze the relationships between variables and GDP.
* Model Training and Prediction: Implements machine learning models (Linear Regression, Random Forest Regressor) to predict GDP values based on various economic factors.
* Model Evaluation: This module evaluates model performance using metrics such as RMSE (Root Mean Squared Error) and MSLE (Mean Squared Logarithmic Error).
* Visualization: Displays results of GDP predictions and comparisons between actual and predicted GDP values.
* Reporting: Generates reports on key insights, model performance, and country-wise GDP data.

**4) Operating Environment (H/W & S/W Requirement Specification)**

* Hardware Requirements:
  + Processor: Minimum Intel Core i5 or equivalent
  + RAM: 8 GB (recommended 16 GB)
  + Storage: 1 GB of free space for dataset and model storage
  + Display: 1080p resolution or higher
* Software Requirements:
  + Operating System: Windows 10/Ubuntu 18.04 or higher
  + Python 3.7 or higher
  + Required Libraries: Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn
  + Jupyter Notebook (for development and testing)
  + IDE: Visual Studio Code / PyCharm / Jupyter Notebook

**Overview of the Proposed System**

**1) Proposed System**

The "EcoTrack" system is designed to enable users to understand the relationship between economic variables and GDP by providing both data exploration and predictive modeling capabilities. The system integrates data processing, statistical analysis, and machine learning to create a comprehensive tool for GDP prediction and analysis. The system allows for the visualization of trends and comparisons of GDP across different countries and regions.

**2) Objectives of the System**

* To develop a comprehensive tool for analyzing GDP-related data from multiple countries.
* To build predictive models to estimate GDP per capita based on multiple economic indicators.
* To visualize and interpret the factors that influence GDP using correlation analysis and regression models.
* To provide stakeholders with insights into the economic performance of different regions and countries.

**3) Feasibility Study**

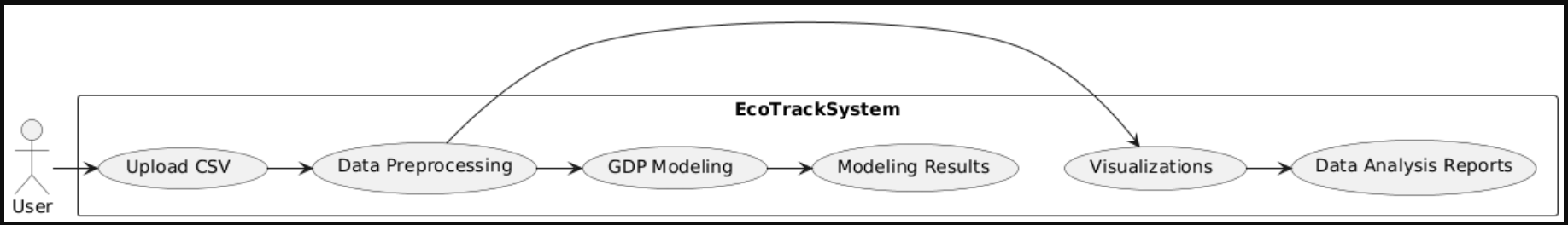
* Technical Feasibility: The project relies on widely used machine learning algorithms and standard libraries in Python, which are well-supported and have the capacity to handle the volume of data in the dataset.
* Operational Feasibility: The system is expected to be used by policymakers, researchers, and economists who are familiar with data analysis tools. The user interface is designed to be user-friendly, and reports can be easily generated for further use.
* Economic Feasibility: The project utilizes publicly available data, making it cost-effective to implement. The project provides valuable insights into economic trends, which can lead to better resource allocation and decision-making.

**4) User Requirement Specification**

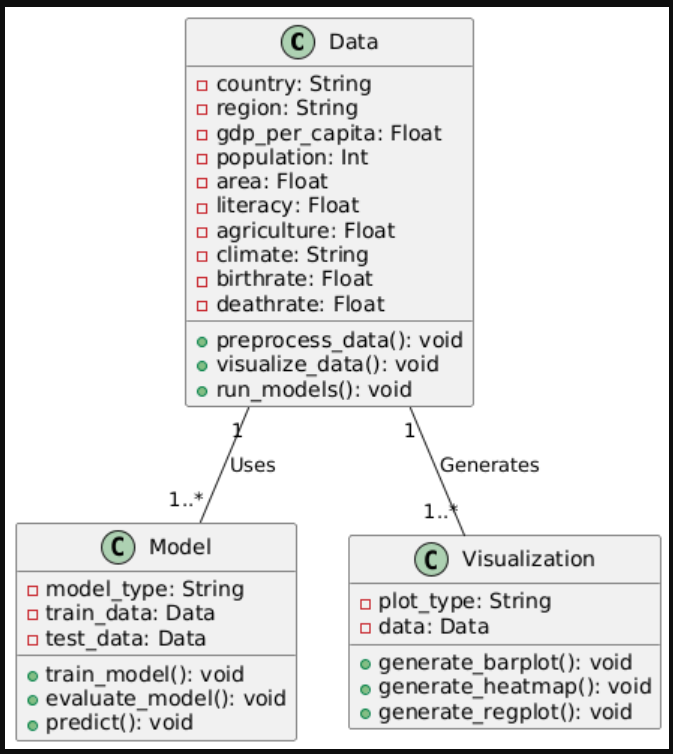
* End Users: Economists, policymakers, researchers, and anyone interested in understanding economic trends and GDP analysis.
* User Requirements:
  + The system should allow users to import and preprocess GDP-related data.
  + Users should be able to visualize correlations and trends between variables.
  + The system should allow for predictive modeling of GDP based on different economic factors.
  + Reports summarizing findings and predictions should be easily accessible.

**System Analysis & Design**

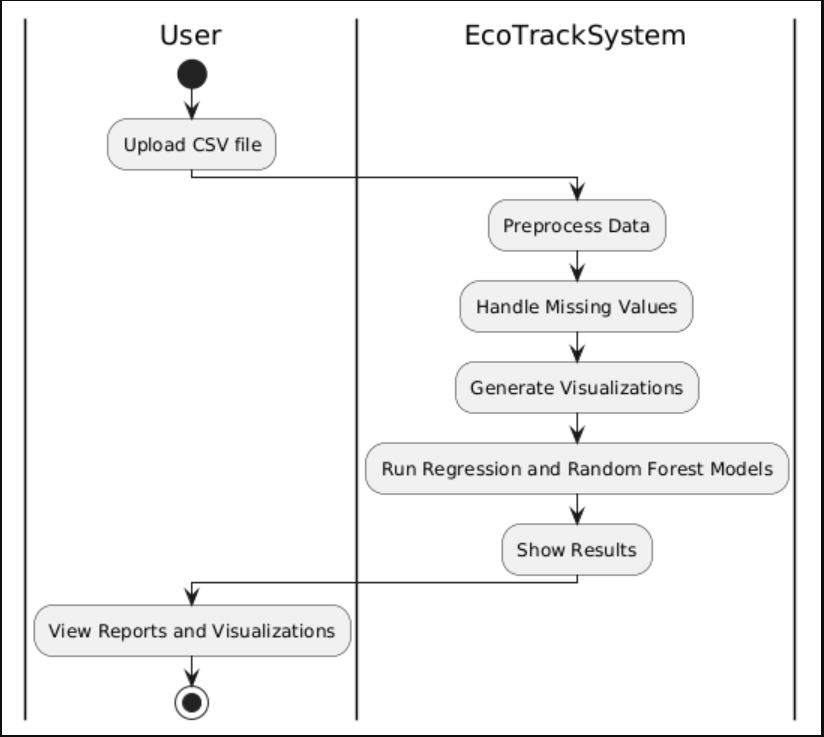
**1) Data Flow Diagram (Context Level Diagram)**

****

**2) Class Diagram/ERD**

****

**3) Activity Diagram**

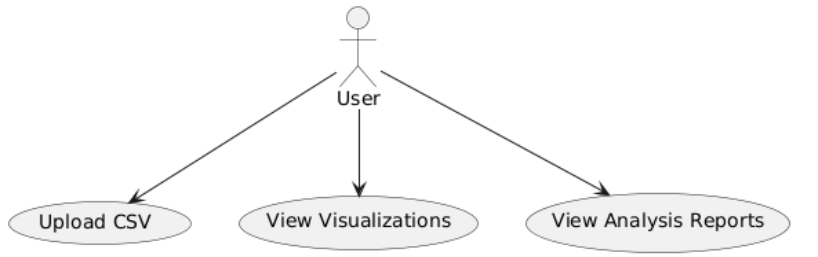
****

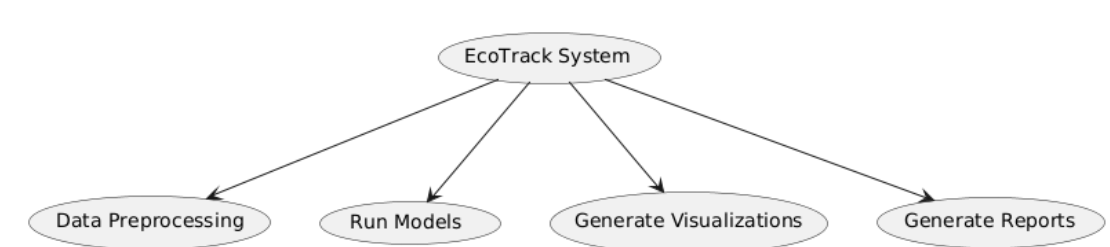
**4) Data Dictionary with Table Specification**

**Table:** world.csv (Main Dataset)

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Country | String | Name of the country |
| Region | String | Region to which the country belongs |
| GDP ($ per capita) | Float | GDP per capita for the country |
| Population | Integer | Population of the country |
| Area (sq. mi.) | Float | Total area of the country in square miles |
| Literacy | Float | Literacy rate of the country |
| Agriculture | Float | Contribution of agriculture to the economy |
| Climate | String | Climate type (e.g., Tropical, Dry, etc.) |
| Birthrate | Float | Birthrate (per 1000 people) |
| Deathrate | Float | Deathrate (per 1000 people) |
| Infant mortality | Float | Infant mortality rate per 1000 births |
| Net migration | Float | Net migration rate (per 1000 people) |
| Phones (per 1000) | Integer | Number of phones per 1000 people |
| Arable (%) | Float | Percentage of arable land |
| Crops (%) | Float | Percentage of land used for crops |
| Other (%) | Float | land usage percentage |
| Coastline (coast/area ratio) | Float | Coastline to area ratio |
| Region\_label | Integer | Encoded value for Region |
| Climate\_label | Integer | Encoded value for Climate |

**5) Use Case Diagram**

****

****

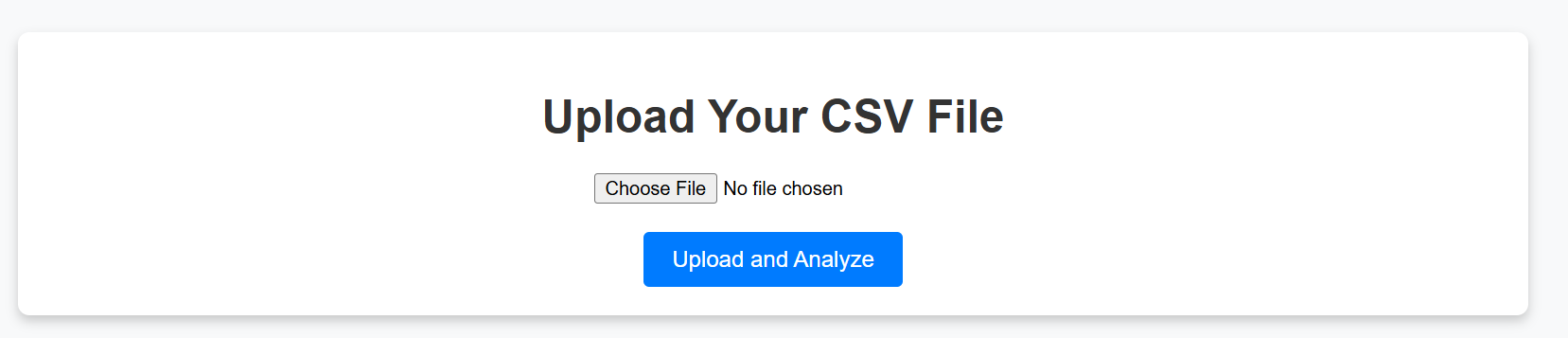
**User Manual**

**1) Operational Instructions**

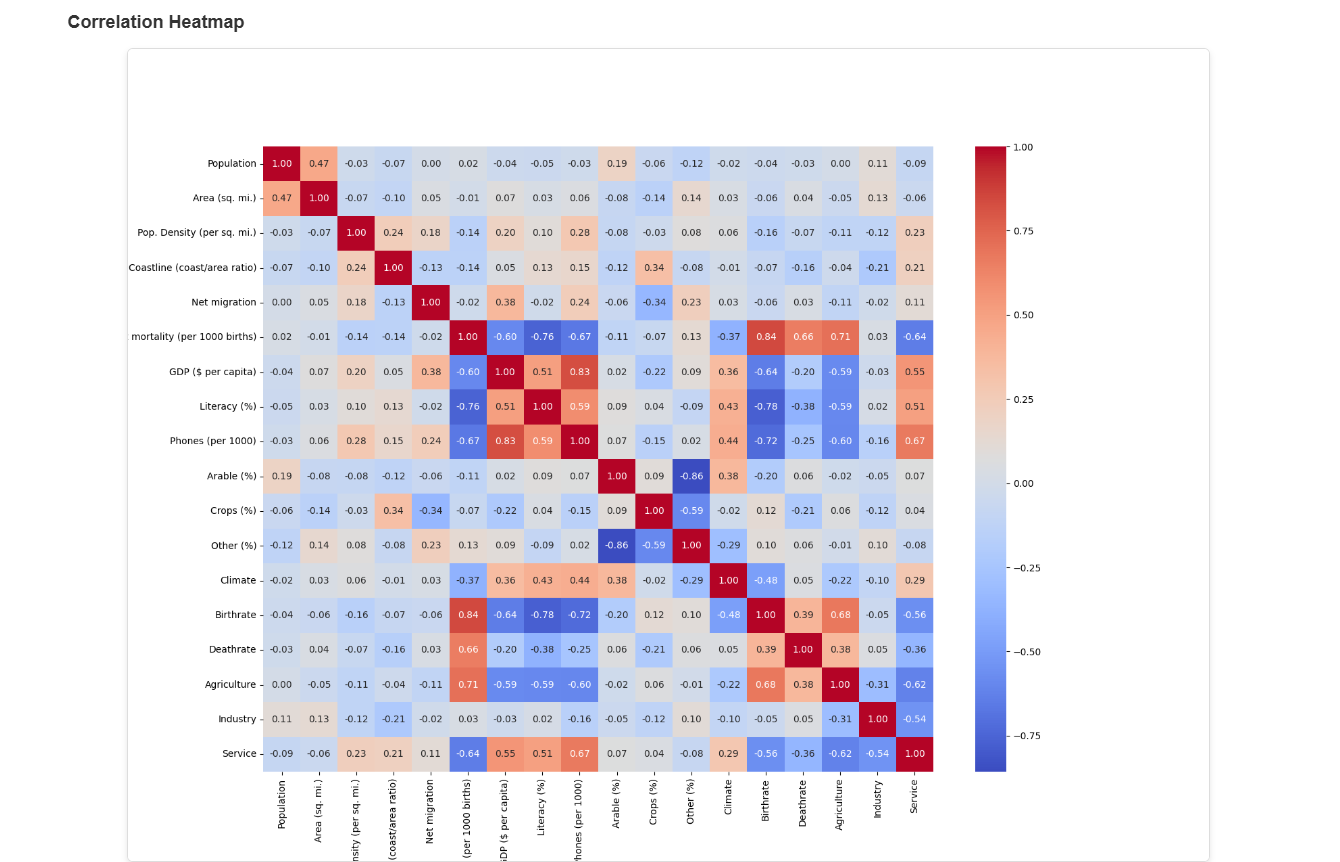
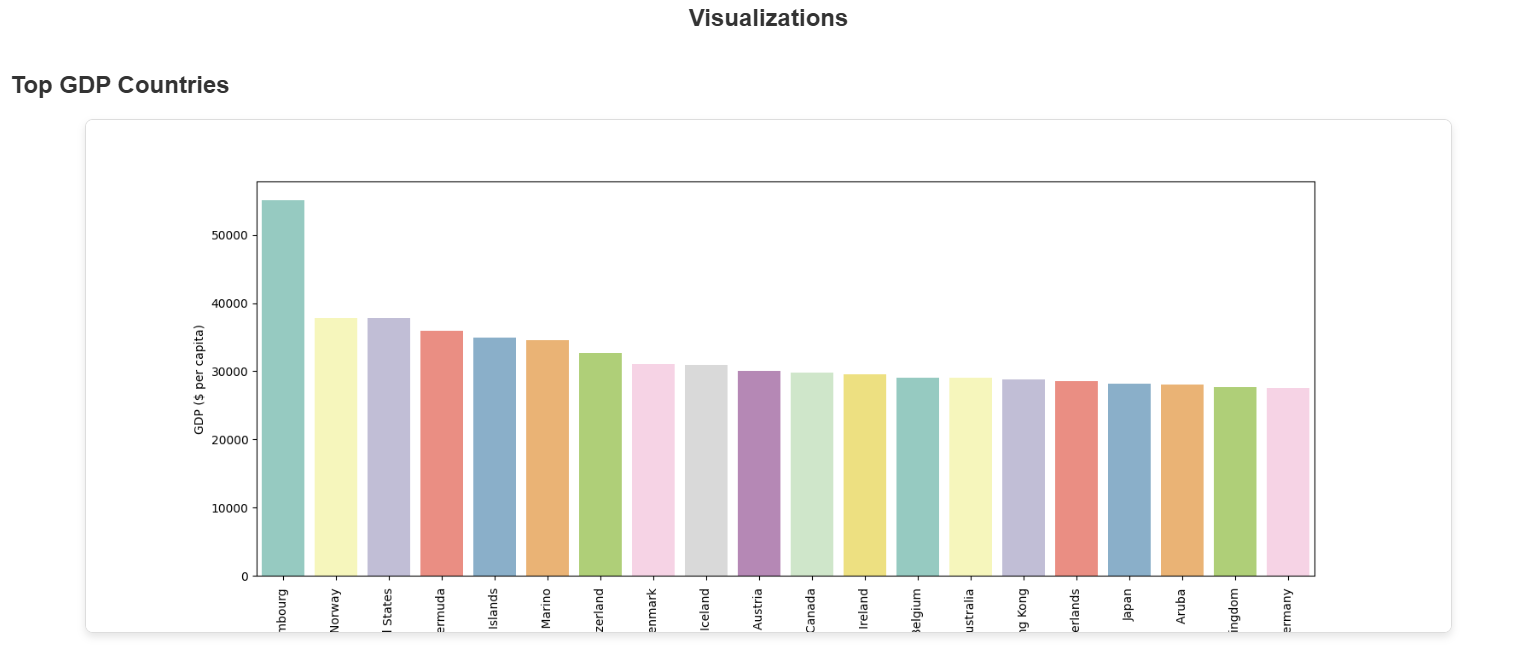
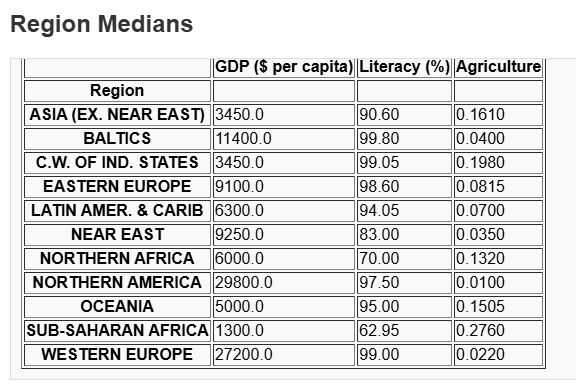
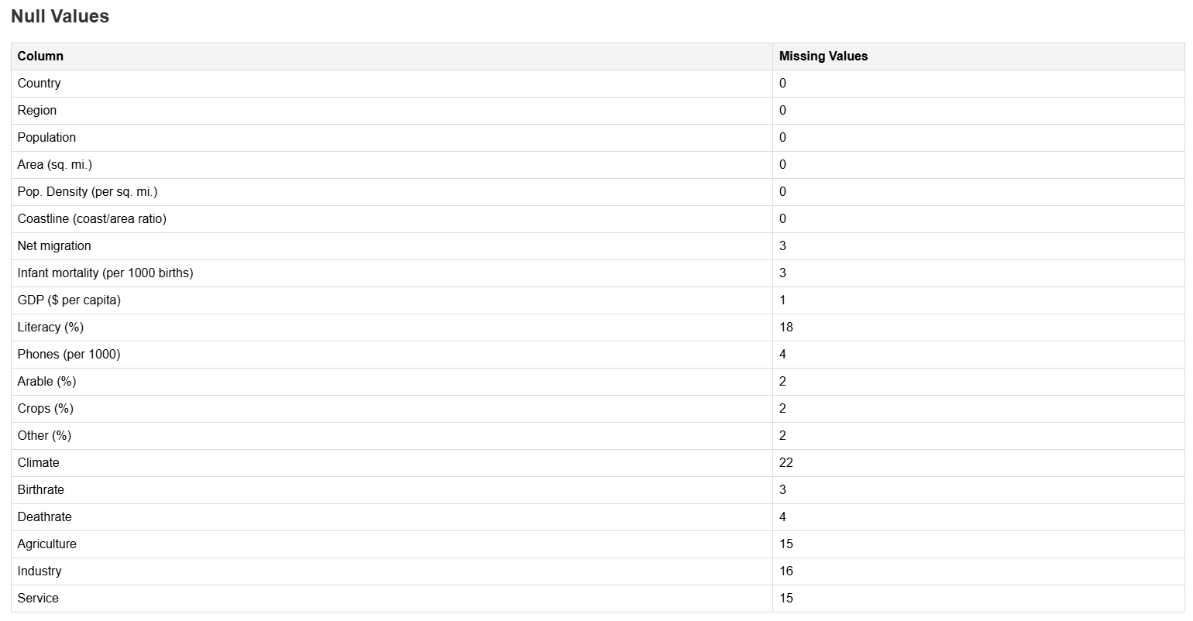
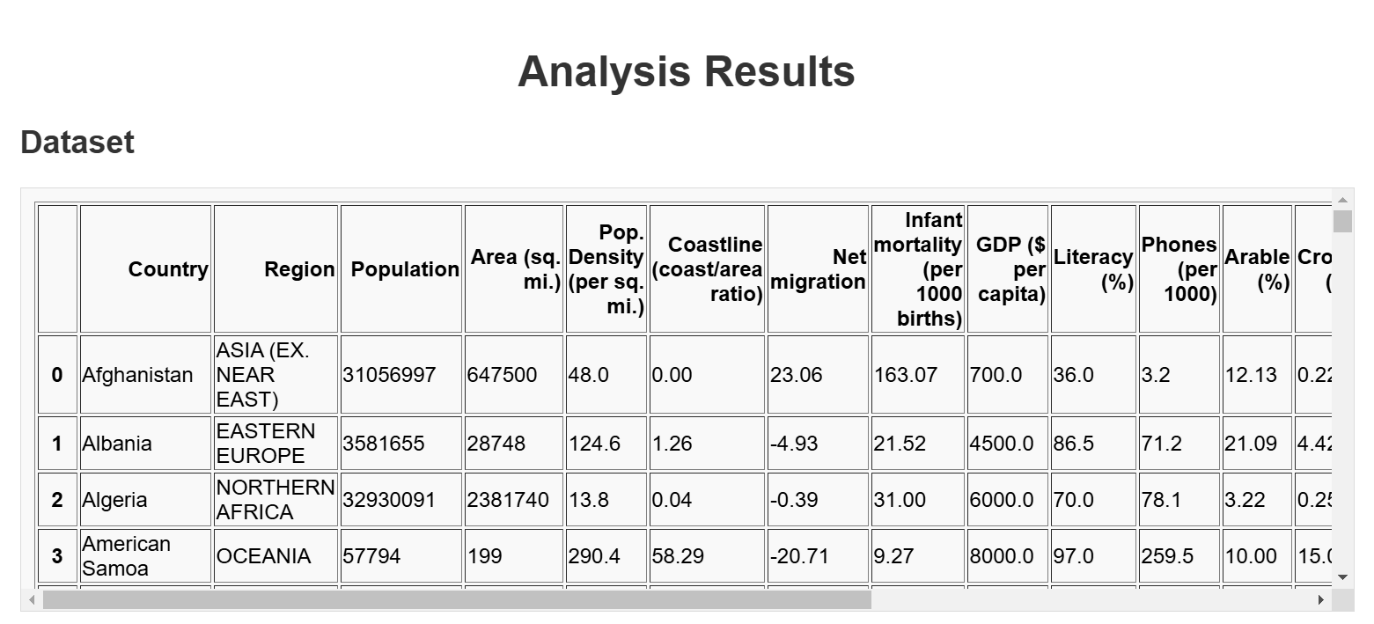
* Step 1: Load the GDP data by selecting the appropriate CSV file from the system’s menu.
* Step 2: The system will automatically preprocess the data, handling missing values and encoding categorical variables.
* Step 3: Select the desired visualizations (e.g., correlation heatmaps, GDP vs. other variables) to explore the data.
* Step 4: Choose a machine learning model (Linear Regression or Random Forest) to predict GDP.
* Step 5: Generate reports that summarize model performance and key findings.

**2) Input/Output Screens**

* Input Screen:



* Output Screen

****

**3) Reports**

* The system generates reports that include the model's accuracy, factors influencing GDP, and country-level GDP predictions, etc.

**Data Exploration and Preprocessing**:

* Missing values were handled using median (for continuous variables) and mode (for categorical variables) based on regions.
* Summary statistics revealed important insights into variables like GDP per capita, literacy rates, and agriculture's contribution.

**Key Visualizations and Insights:**

* Top GDP Countries: A bar plot of the top 20 countries by GDP per capita highlighted wealthier nations like Luxembourg and Norway.
* Correlation Analysis: A heatmap showed strong correlations between GDP and factors like literacy and industrialization.
* Factors Affecting GDP: Scatterplots revealed that higher literacy, industrialization, and infrastructure correlate with higher GDP per capita.
* Total GDP Distribution: A barplot and pie chart visualized the dominance of a few countries in the global GDP.

**Predictive Modeling:**

* Linear Regression Model: Predicted GDP per capita with moderate accuracy, yielding RMSE of 4,612.41 (train) and 4,831.25 (test).
* Random Forest Model: Outperformed the Linear Regression model, with lower RMSE and MSLE scores, indicating better handling of complex relationships.

**Model Evaluation:**

* Both models showed reasonable performance, but the Random Forest model was more accurate in predicting GDP, especially for extreme cases.

**System Limitation**

* Data Availability: The system relies on publicly available data, which may not cover all regions or be updated frequently.
* Model Generalization: The predictive models may not generalize well to all countries, particularly those with unique or non-standard economic structures.
* Complexity of Economic Systems: The models may not fully capture the complexity of global economics, as GDP is influenced by a multitude of unpredictable factors such as geopolitical events and global crises.

**Future Enhancement and Conclusion**

**Future Enhancement**

* More Features: Incorporating additional economic indicators such as unemployment rate, inflation, and trade data could improve the accuracy of GDP predictions.
* Real-time Data: Implementing a data scraping module to gather real-time economic data from online sources could make the system more dynamic and accurate.
* Advanced Models: Using advanced models such as neural networks and deep learning could further enhance predictive capabilities.

**Conclusion**

The "EcoTrack" system provides an effective platform for exploring, analyzing, and predicting GDP using machine learning techniques. It helps users understand the relationships between economic variables and GDP, and can serve as a valuable tool for policymakers, economists, and researchers. By combining data analysis, machine learning, and visualizations, the system offers an intuitive interface for both exploring data and generating predictions.

**Bibliography and Glossary**

**Bibliography:**

1. “An Introduction to Statistical Learning” - Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.
2. “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow” - Aurélien Géron.
3. OECD Economic Data - OECD, 2020.

**Glossary:**

* GDP: Gross Domestic Product - A measure of a country's economic output.
* RMSE: Root Mean Squared Error - A metric for evaluating the accuracy of predictions.
* MSLE: Mean Squared Logarithmic Error - A metric for measuring model performance that is sensitive to large deviations in predictions.

**Acronyms:**

* EDA: Exploratory Data Analysis
* ML: Machine Learning
* RF: Random Forest
* CSV: Comma-Separated Values