****

**BAHRIA UNIVERSITY, Karachi Campus**

*Department of Software Engineering*

REPORT

**Course Title:**  Introduction To Data Science **Course Instructor**: Dr.Hina Shakir

**Lab Instructor:** Engr. Ayesha Khan  **Class**: BSE- (5B)

PROJECT TITLE:

Data Analysis & Prediction On Olympics

GROUP MEMBERS LIST:

|  |  |  |
| --- | --- | --- |
| S.NO | Enrollment | Name |
| 01 | 02-13121212-027 | Kanwal Shehzadi |
| 02 | 02-13121212-028 | Aiman Zia Satti |

**Submission Date: 09-January-2024**

# **INTRODUCTION**

## **Introduction**

## The Olympics Data Analysis and Prediction project embark on a comprehensive exploration of historical Olympics datasets, employing a multifaceted approach that amalgamates data analysis, visualization, and machine learning techniques. Spanning multiple years, the project seeks to unravel trends, patterns, and influential factors shaping performance across various sports. The ultimate goal is to create predictive models capable of forecasting winners in specific Olympic events, leveraging selected input features.

## **Problem Statement**

The project addresses several challenges inherent in the analysis and prediction of Olympic Games outcomes. First and foremost is the complexity of the historical datasets, which require meticulous exploration and cleaning to ensure accurate and reliable insights. Handling missing values and inconsistencies is crucial to maintaining the integrity of the analysis. The deployment of the predictive model for real-time predictions on Streamlit introduces challenges related to system integration, scalability, and ensuring consistent performance in a dynamic environment.

## **Proposed Solution**

In response to the expansive scope of the project, the proposed solution entails a systematic and thorough analysis pipeline. Beginning with data exploration and cleaning, the project ensures the integrity of the datasets through in-depth examination and preprocessing to handle missing values and inconsistencies. Descriptive analytics form a pivotal aspect, involving the visualization of trends in Olympic participation and the analysis of medal distributions by country, sport, and athlete demographics.

To facilitate a nuanced understanding of athlete performance, the project enables users to compare historical data, providing insights into the consistency and improvement of athletes over multiple Olympics. A country-wise analysis offers a comprehensive view of each nation's performance in different sports, visualizing the distribution of medals over the years. The project also delves into seasonal trends, scrutinizing variations in Olympic Games participation and performance.

**Technologies Used**

* Streamlit
* PyCharm

**PROJECT FUNCTIONALITIES:**

**Data Exploration and Cleaning:**

* In-depth exploration of historical Olympics datasets.
* Cleaning and preprocessing of data to handle missing values and inconsistencies.

**Descriptive Analytics:**

* Visualizing trends in Olympic participation over the years.
* Analyzing the distribution of medals by country, sport, and athlete demographics.

**Athlete Performance Comparison:**

* Allow users to compare the historical performance of athletes in a specific sport.
* Provide insights into the consistency and improvement of athletes over multiple Olympics.

**Country-wise Analysis:**

* Explore the overall performance of countries in different sports.
* Visualize the distribution of medals over the years for each country.

**Seasonal Trends:**

* Analyze seasonal trends in Olympic Games and how participation and performance vary.

**Predictive Modeling:**

* Developing machine learning models for predicting winners in specific sports.
* Evaluating and selecting the most suitable model for the task.

**User Interface:**

* Building a user-friendly interface for users to input parameters and receive predictions.
* Integration with visualization tools for presenting analysis results.

**Deployment:**

* Deploying the predictive model and analysis on streamlit for real-time predictions.

## **LIBRARIES USED:**

import streamlit as st

import pandas as pd

import preprocessor,helper

import plotly.express as px

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.figure\_factory as ff

import numpy as np

import time

import pickle

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

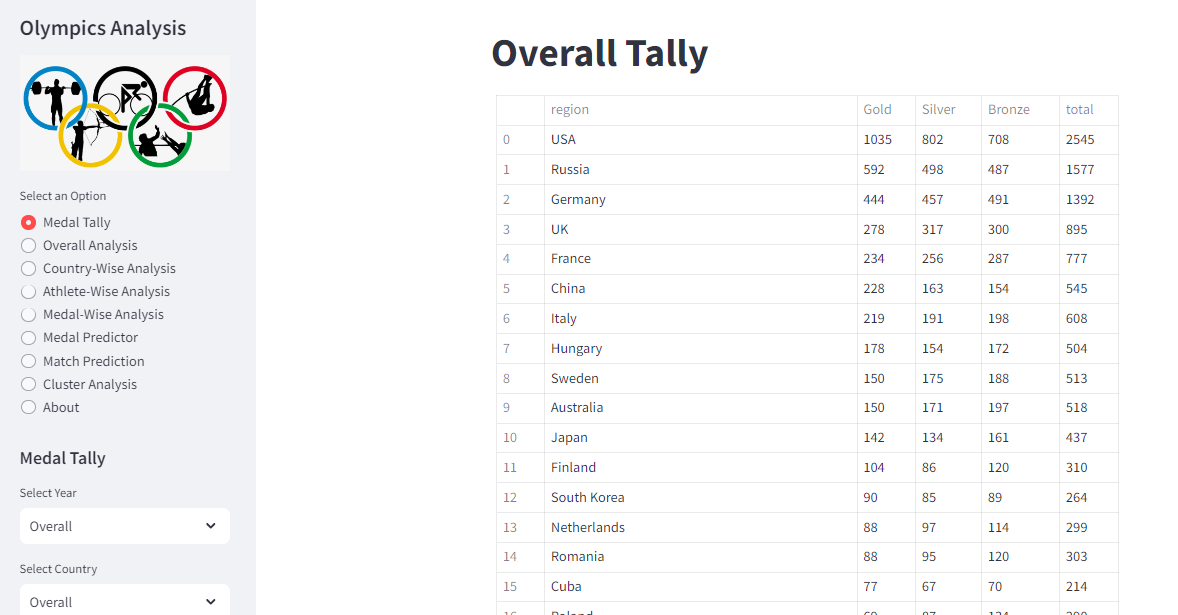
from sklearn.metrics import classification\_report, confusion\_matrix

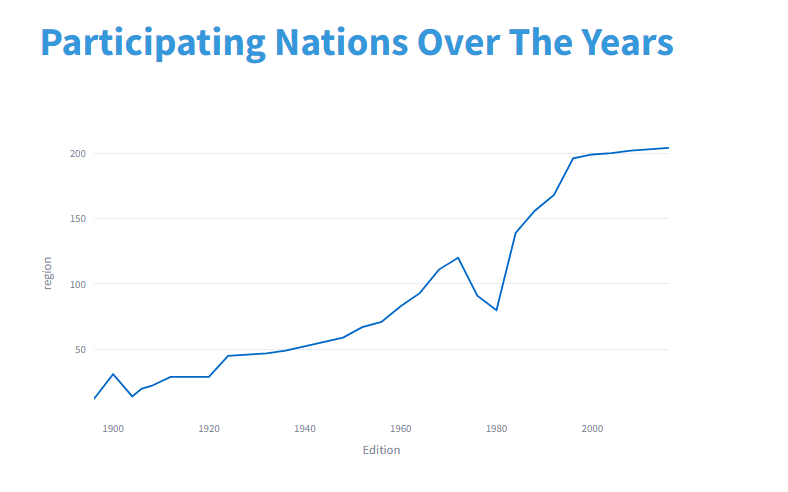
from sklearn.linear\_model import LogisticRegression

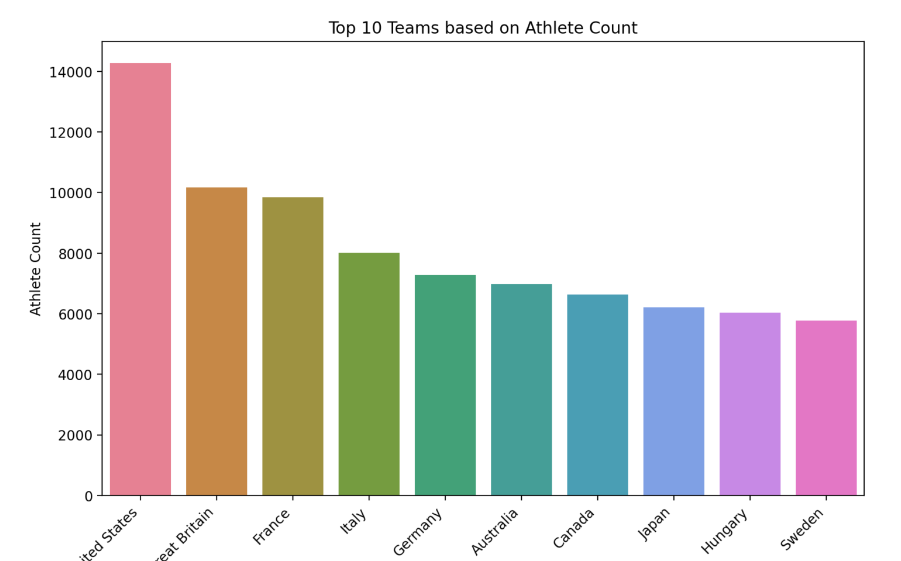
from sklearn.metrics import accuracy\_score

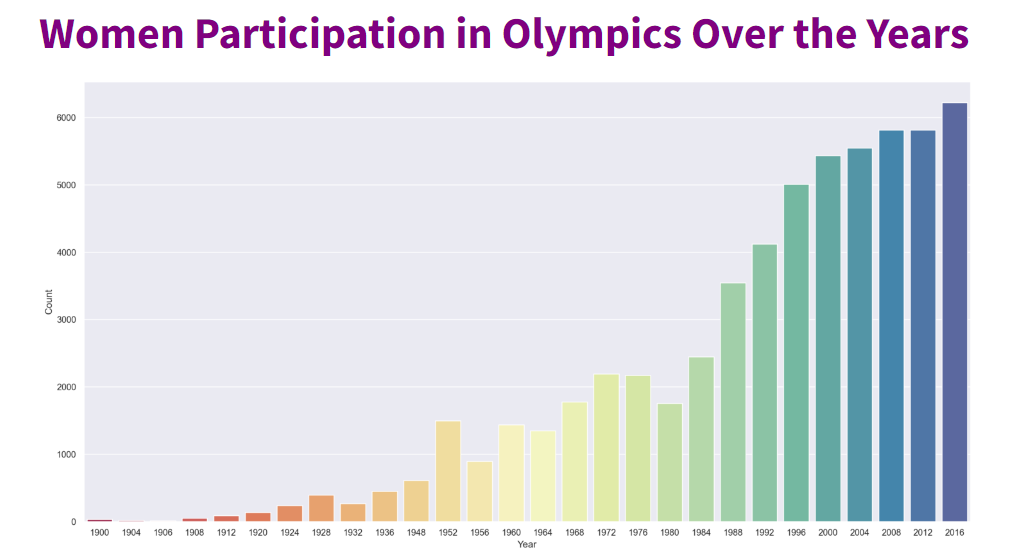
from sklearn.ensemble import RandomForestClassifier

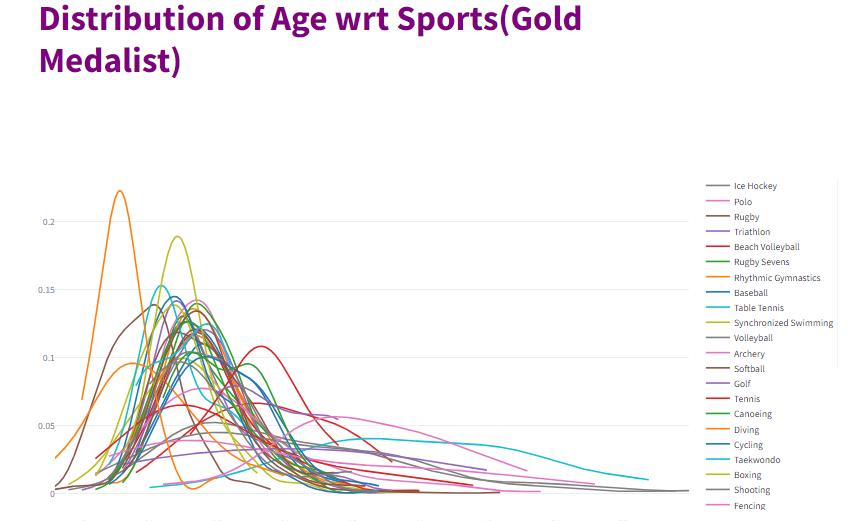
## **INTERFACES:**











## 

## 

## **CONCLUSION****:**

## The Olympics Data Analysis and Prediction project have traversed an extensive journey, uncovering valuable insights into historical Olympics data and providing a predictive model for forecasting outcomes of future Olympic events. The in-depth exploration and cleaning of datasets laid the foundation for robust analyses, overcoming challenges related to missing values and inconsistencies.By embracing ongoing data enrichment, model refinement, and user feedback, the project can remain at the forefront of sports analytics, contributing valuable insights to the ever-evolving world of Olympic Games.

The deployment on Streamlit ensures real-time access to predictions, aligning the project with the dynamic nature of the Olympic Games.