Olympics History (1896-2016)

Data Analysis Project

Submitted To:-

Dr. Anita Budhiraja

National Institute of Electronics

and Information Technology

(NIELIT),Mohali

Submitted By:-

Bhavuk Thakur(07)

Gagan Sharma(03)

Munish Kumar(06)

Vishal Negi(09)

INDEX

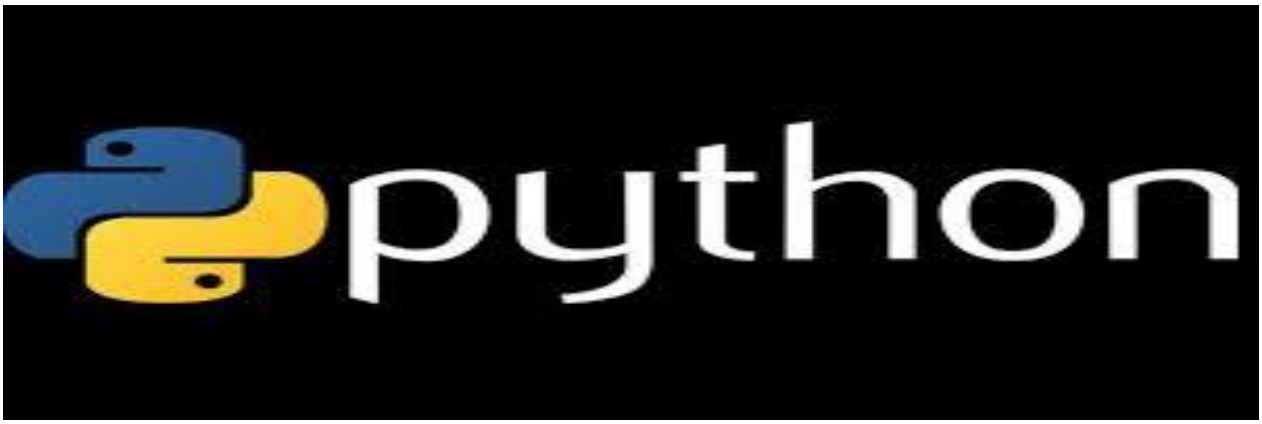
acknowledgement

We would like to express our sincere gratitude and special thanks to Dr. Anita Budhiraja who gave us a golden opportunity to learn under her vision and guidance. It is with her help that we are able to learn and successfully make a project based on Data Analysis - Olympics History (1896-2016).

Past few decades, we have seen a huge revolution in the Technical World. From room sized computer to pocket sized gadgets, from writing codes in low level languages to writing codes in high level languages and from waiting for hours for a single command to execute to executing the whole code in matter of a short while, everything has beautifully evolved.

With the advancement in Technology, one of the field that has bought a revolution to the Technical world is Artificial Intelligence and Data Science. The robots and gadgets are getting more and more intelligent and therefore making human life easier. Artificial Intelligence is a huge world in it itself. It's like a big umbrella and Machine Learning, Data Science are a part or subset of it.

The time to come will experience a huge revolution due to the further evolution of Artificial Intelligence. The lives of humans will be easier and will lead to evolution of many things advantageous to human lives.



Python is an interpreted, object-oriented programming language like PERL, which has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in several operating systems. Python was created by Guido van Rossum, a former resident of the Netherlands, whose favorite comedy group at the time was Monty Python's Flying Circus. Python uses dynamic typing, and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python supports modules and packages, which encourages program modularity and code reuse.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science. When Harvard Business Review called it "The Sexiest Job of the 21st Century", the term "data science" became a buzzword, and is now often applied to business analytics, business intelligence, predictive modeling, or any arbitrary use of data, or used as a glamorized term for statistics. In many cases, earlier approaches and solutions are now simply rebranded as "data science" to be more attractive, which can cause the term to become "dilute beyond usefulness."

Python is often the choice for developers who need to apply statistical techniques or data analysis in their work, or for data scientists whose tasks need to be integrated with web apps or production environments. Python really shines in the field of machine learning. Its combination of machine learning libraries and flexibility makes Python uniquely well-suited to developing sophisticated models and prediction engines that plug directly into production systems.

Some of Python Libraries for Data Science and Data Analytics are:

1. NumPy is the foundational library for scientific computing in Python, and many of the libraries on this list use NumPy arrays as their basic inputs and outputs. In short, NumPy introduces objects for multidimensional arrays and matrices, as well as routines that allow developers to perform advanced mathematical and statistical functions on those arrays with as little code as possible.

2. SciPy builds on NumPy by adding a collection of algorithms and high-level commands for manipulating and visualizing data. This package includes functions for computing integrals numerically, solving differential equations, optimization, and more.

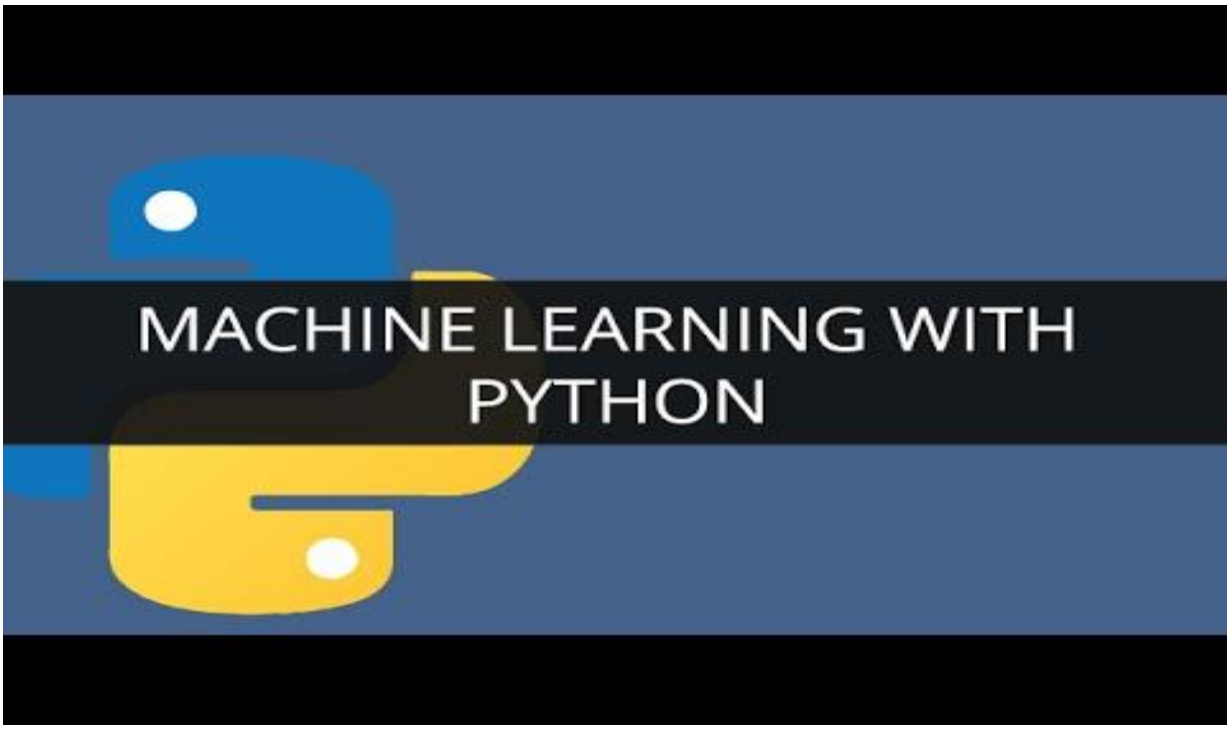
3. Pandas adds data structures and tools that are designed for practical data analysis in finance, statistics, social sciences, and engineering. Pandas works well with incomplete, messy, and unlabeled data (i.e., the kind of data you're likely to encounter in the real world), and provides tools for shaping, merging, reshaping, and slicing datasets.

4. IPython extends the functionality of Python's interactive interpreter with a souped-up interactive shell that adds introspection, rich media, shell syntax, tab completion, and command history retrieval. It also acts as an embeddable interpreter for your programs that can be really useful for debugging.

5. matplotlib is the standard Python library for creating 2D plots and graphs. It's pretty low-level, meaning it requires more commands to generate nice-looking graphs and figures than with some more advanced libraries. However, the flip side of that is flexibility. With enough commands, you can make just about any kind of graph you want with matplotlib.

Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

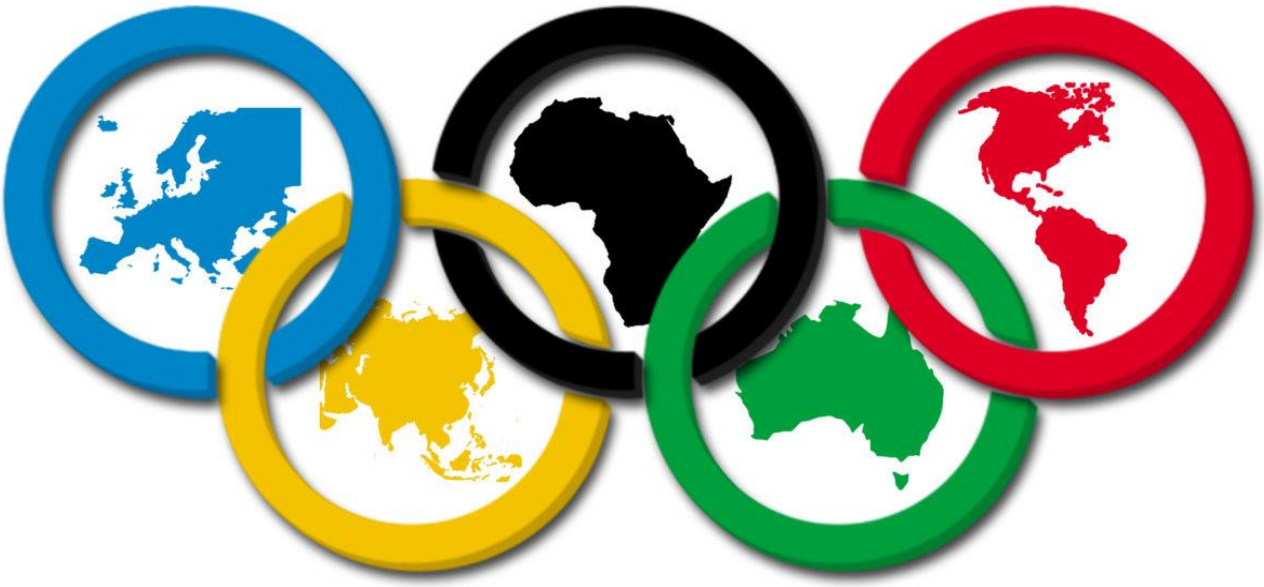


Machine learning sits at the intersection of Artificial Intelligence and statistical analysis. By training computers with sets of real-world data, we're able to create algorithms that make more accurate and sophisticated predictions, whether we're talking about getting better driving directions or building computers that can identify landmarks just from looking at pictures. The following libraries give Python the ability to tackle several machine learning tasks, from performing basic regressions to training complex neural networks:-

1. Scikit-learn builds on NumPy and SciPy by adding a set of algorithms for common machine learning and data mining tasks, including clustering, regression, and classification. As a library, scikit-learn has a lot going for it. Its tools are well-documented, and its contributors include many machine learning experts.

2. TensorFlow is another high-profile entrant into machine learning, developed by Google as an open-source successor to DistBelief, their previous framework for training neural networks. TensorFlow uses a system of multi-layered nodes that allow you to quickly set up, train, and deploy artificial neural networks with large datasets. It's what allows Google to identify objects in photos or understand spoken words in its voice-recognition app.

3. NLTK is a set of libraries designed for Natural Language Processing (NLP). NLTK's basic functions allow you to tag text, identify named entities, and display parse trees, which are like sentence diagrams that reveal parts of speech and dependencies. We can do sentiment analysis with it.



The **Olympic Games** are an important international event featuring summer and winter sports. Summer Olympic Games and Winter Olympic Games are held every four years. Originally, the ancient Olympic Games were held in Ancient Greece at Olympia. The first games were in 776 BC. They were held every four years until the 5th century AD. The first "modern" Olympics happened in 1896 in Athens, Greece. Athletes participate in the Olympics Games to represent their country.

The celebration of the Games includes many rituals and symbols, such as Olympic flag and torch, as well

as the opening and closing ceremonies. The first, second, and third-place finishers in each event receive, respectively, gold, silver, and bronze medals.

It is coded in Python and its GUI is coded in Tkinter- a python library for designing effective and catchy GUI. The application is coded on Anaconda IDE.

The Data Set used for training was fetched from Kaggle

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

The file was a .csv file. The file contained approximately 210000 entries. The various entries in columns included Name, Team, Height, Weight, Age and Medals.

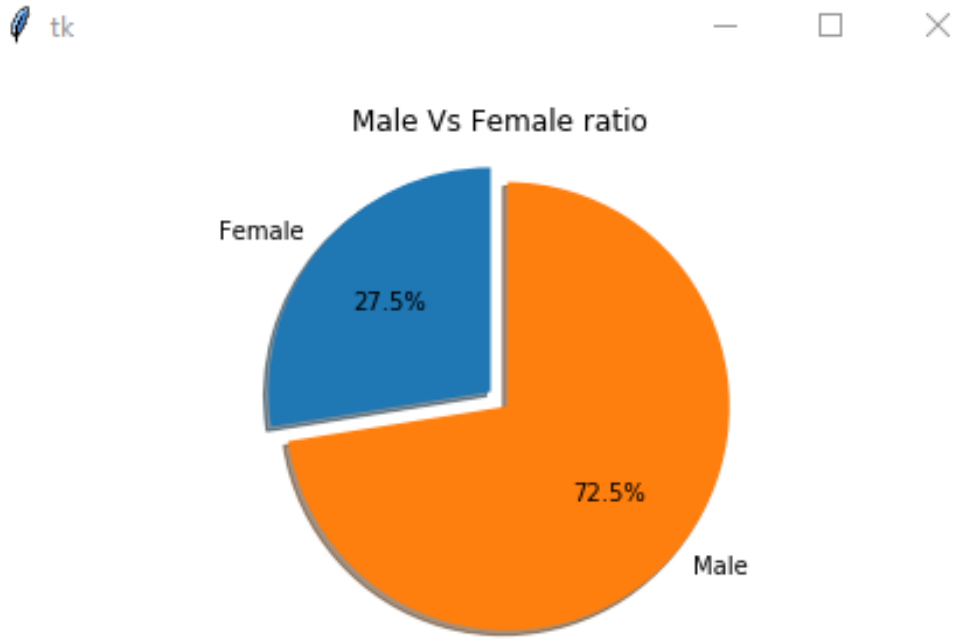
The reason for choosing Olympics as the project was that in the present time the old quote comes to action that Sports Unite the World.

Importing the modules

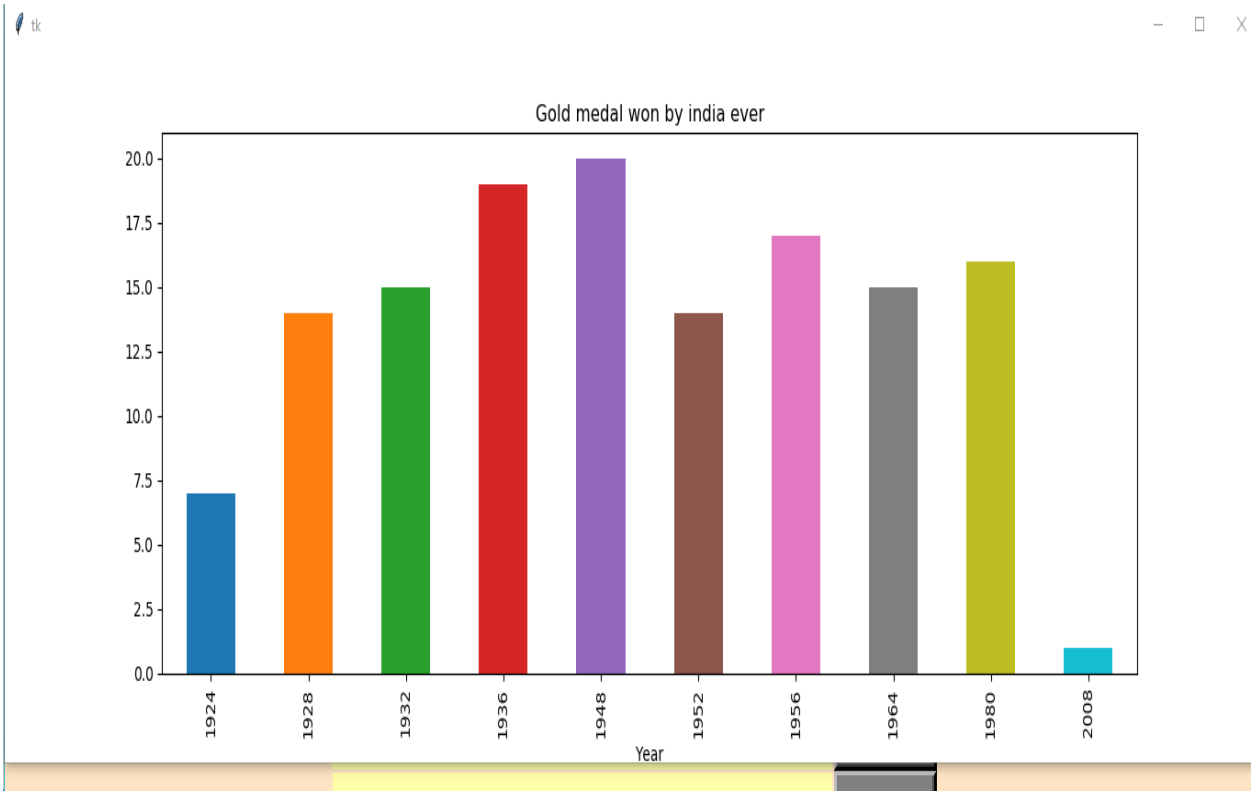
```
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt

import os
print(os.listdir("../input"))
```

Male Vs. Female Ratio



Gold Medals won by India



Data Science and Its Growing Importance – An interdisciplinary field, data science deals with processes and systems, that are used to extract knowledge or insights from large amounts of data.

So while retail is a tangible field where the effects of data science is clearly visible, data science can have far reaching implications in other fields as well. These include healthcare, energy and education.

Because these fields are constantly evolving, the importance of data science is also rapidly increasing.

In the field of healthcare, new drugs are being constantly discovered one hand and there is a need to create better care for patients on the other hand.

Data science with its use of methods and techniques can help the healthcare sector to find solutions that help take patient care and satisfaction to the next level.

The healthcare industry is constantly evolving and data science can help them create better care for patients at all stages. Another field that can truly benefit from data science is education.

With technology like smartphones and laptops becoming an integral part of the education system, data science can help create better opportunities to help students learn and enhance their knowledge in a constructive manner.

Another example of how data science can help society is through its application and use in energy. The energy sector is today on the cusp of radical change and transformation. From oil to gas to renewable energy, we need to find new and innovative ways to use energy.

Data science can help us meet the challenges of the increasing demand and sustainable future while ensuring the best solutions. This means that data scientists will have to come up with a wide range of solutions to meet challenges across all sectors.

This is not an easy task and that is why they need the resources and systems that will help them achieve this goal. Across sectors and economies, data scientists will have to become creative thinkers who use high end tools to create solutions that can be adopted across all verticals.

All in all, data scientists are the future of the world today. They will soon become an integral part of the organisation and help the world address major global challenges, Which in turn can have far reaching impacts across countries.

That is why the need of the hour is to develop the skill and creativity of data scientists across the world so that they can help people to experience life, products and services in a brand new manner.

- https://www.sas.com/en_in/insights/analytics/machine-learning.html
- https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results
- https://www.python.org
- https://www.upwork.com/hiring/data/15-pythonlibraries-data-science
- http://scikit-learn.org/stable/modules/tree.html


Thank you