

# DataScience Project Details

There will be 13 mini projects and 4 major projects and for each concept we will be having a Real time Use Case scenario for better understanding of the Concept

## Mini Project 1

### **Title: Analysis of Capitals and Universities Data**

**Description:** This project aims to analyze and correlate data from two CSV files: capitals.csv and universities.csv.

capitals.csv contains detailed information about countries, including their names, capitals, geographical coordinates (latitude and longitude), country codes, and continents. This data provides insights into the geographical context of each country.

universities.csv lists various educational institutions, detailing their codes, names, national ranks, degrees offered, average costs, and scores. This information allows for a comparison of educational quality and accessibility across different countries.

The goal of the project is to explore relationships between countries and their educational institutions, examining factors such as national ranking, costs, and degree offerings in relation to geographical and demographic data. This analysis will provide valuable insights for students, educators, and policymakers interested in global education trends and the impact of geography on educational access and quality. Write the code using python functions.

## Mini Project 2

### **Title: Online Shopping System**

**Description:** This project implements a simple Online Shopping System that allows users to browse and purchase a selection of default products, including electronics, clothing, and fruits. Users can add items to their cart, update quantities, search products, remove products, or clear the cart entirely.

The system features a straightforward interface that enables users to view available products and their prices. Once users finalize their selections, they can place an order and receive an invoice detailing each item, its quantity, the

subtotal, a calculated 10% GST on each product price, and the total amount included in the invoice.

This project demonstrates key concepts in Python, such as object-oriented programming, data management, and basic arithmetic operations, providing a practical framework for understanding how online shopping systems operate.

## Mini Project 3

### **Title: Analytic American Association**

**Description:** The Analytic American Association is a robust application designed to optimize the functions of law firms by effectively managing clients, Employees (Lawyer), tasks, billing, and time sheets. This application enhances communication between clients and legal professionals while ensuring precise tracking of billable hours and generating invoices as well as time sheets for services rendered.

Key features of the project include:

- **Comprehensive Client Management:** Maintain detailed client profiles, including personal information and case descriptions. Generate tasks based on client needs and case specifics.
- **Efficient Employee (Lawyer) Tracking:** Track lawyer profiles, including their specialties, hourly rates, and availability. Allow Employees to manage assigned tasks and submit accurate timesheets for work completed.
- **Dynamic Task and Time Management:** Classify tasks as billable or non-billable, with clear tracking of their status and deadlines. Enable Employees to log hours worked, ensuring adherence to the 8-hour daily limit while facilitating approval processes.
- **Robust Billing and Invoicing System:** Generate detailed invoices for clients based on billable hours and specific lawyer rates. Create comprehensive billing records for employees, outlining their hours worked on various client tasks.
- The Analytic American Association will enhance productivity within the law firm by automating administrative tasks, ensuring accurate billing, and

improving overall client satisfaction. By centralizing all operations, the application will allow for better resource management and a more efficient workflow.

## Data Libraries

### Mini Project 4

#### **Title: Sales Data Insights**

#### **Description:**

This project provides a comprehensive analysis tool for processing and analyzing sales data from a CSV file named `Sales_Data_Analysis.csv`. The dataset includes various columns such as Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country, which collectively offer insights into sales transactions.

Key features of the project include:

- **Data Loading:** Efficiently read and load sales data from the CSV file for analysis.
  - **Sales Summary:** Generate a summary of total sales, including total revenue, number of transactions, and average order value.
  - **Product Analysis:** Identify the most and least sold products, along with their respective quantities and sales revenue.
  - **Customer Insights:** Analyze customer purchasing behavior based on Customer ID, including repeat purchases and average spending per customer.
  - **Country-Based Analysis:** Examine sales distribution by country, allowing for insights into geographic sales performance.
  - **Time-Based Analysis:** Analyze sales trends over time, identifying peak sales periods and seasonal variations.
- This project serves as a practical application of data analysis techniques using Python, Numpy, Pandas enabling users to extract valuable insights from sales data to inform business decisions and strategies. It showcases key data manipulation, making it a useful tool for both data analysts and business owners.

## Mini Project 5

### **Title: Employee Data Analysis**

#### **Description:**

This project implements an Employee Data Analysis Tool designed to process and analyze information from a CSV file named Employee\_Data.csv. The dataset contains comprehensive details about employees, including Emp ID, FirstName, LastName, StartDate, Exit Date, Title, Supervisor, and various other attributes that provide insights into the workforce.

Key features of the project include:

- **Data Integrity:** Perform initial checks for missing values and duplicates, ensuring reliable analyses by dropping critical rows as necessary.
- **Performance Metrics:** Calculate average performance scores and ratings across various dimensions, such as departments and supervisors, to highlight areas of strength and improvement.
- **Employee Turnover Analysis:** Assess employee exits and termination reasons, providing insights into workforce stability and potential retention issues.
- **Demographic Insights:** Analyze gender and race distribution among employees, offering a clear view of diversity within the organization. This analysis aims to enhance understanding of employee performance and turnover while ensuring data quality. It serves as a valuable resource for HR professionals and organizational leaders to make informed decisions regarding employee management and engagement strategies.

## Machine Learning

## Mini Project 6

### **Title: Melbourne Housing Snapshot**

#### **Description:**

Melbourne real estate is BOOMING. Can you find the insight or predict the next big trend to become a real estate mogul... or even harder, to snap up a reasonably priced 2-bedroom unit?. This is a snapshot of a dataset created by Tony Pino. It was scraped from publicly available results posted every week from

Domain.com.au. He cleaned it well, and now it's up to you to make data analysis magic.

The dataset includes Address, Type of Real estate, Suburb, Method of Selling, Rooms, Price, Real Estate Agent, Date of Sale and distance from C.B.D.

Key features of the project include:

- **Data Loading:** Efficiently read and load sales data from the CSV file for analysis.
- **Sales Summary:** Generate a summary of total sales, including total revenue, number of transactions, and average order value.
- **Product Analysis:** Identify the most and least sold products, along with their respective quantities and sales revenue.
- **Customer Insights:** Analyze customer purchasing behavior based on Customer ID, including repeat purchases and average spending per customer.
- **Country-Based Analysis:** Examine sales distribution by country, allowing for insights into geographic sales performance.
- **Time-Based Analysis:** Analyze sales trends over time, identifying peak sales periods and seasonal variations. This project serves as a practical application of data analysis techniques using Python, Numpy, Pandas enabling users to extract valuable insights from sales data to inform business decisions and strategies. It showcases key data manipulation, making it a useful tool for both data analysts and business owners

## Remaining Mini Projects will be given as assignments

Captcha generator

Age Calculator

ToDoList

EmailSender

OtpGenerator

Alarm Clock

# Major Projects

## 1. Major Project 1

### **Title: Obesity Challenge**

#### **Description:**

Obesity is a serious and chronic disease with genetic and environmental interactions. It is defined as an excessive amount of fat tissue in the body that is harmful to health. The main risk factors for obesity include social, psychological, and eating habits. Obesity is a significant health problem for all age groups in the world. Currently, more than 2 billion people worldwide are obese or overweight. Research has shown that obesity can be prevented. In this study, artificial intelligence methods were used to identify individuals at risk of obesity. An online survey was conducted on 1610 individuals to create the obesity dataset. To analyze the survey data, four commonly used artificial intelligence methods in literature, namely Artificial Neural Network, K Nearest Neighbors, Random Forest and Support Vector Machine, were employed after pre-processing. As a result of this analysis, obesity classes were predicted correctly with success rates of 74.96%, 74.03%, 74.03% and 87.82%, respectively. Random Forest was the most successful artificial intelligence method for this dataset and accurately classified obesity with a success rate of 87.82%.

## 2. Major Project 2

### **Title: Financial Risk for Loan Approval**

#### **Description:**

This synthetic dataset comprises 20,000 records of personal and financial data, designed to facilitate the development of predictive models for risk assessment. It serves two primary purposes:

1. Risk Score Regression: To predict a continuous risk score associated with each individual's likelihood of loan default or financial instability.

2. Binary Classification: To determine the binary outcome of loan approval, indicating whether an applicant is likely to be approved or denied for a loan. The dataset includes diverse features such as demographic information, credit history, employment status, income levels, existing debt, and other relevant financial metrics, providing a comprehensive foundation for sophisticated data-driven analysis and decision-making.

### 3. Major Project 3

#### **FLM Bank Application**

##### **Description:**

In the evolving landscape of digital banking, there is a growing need for robust, secure, and user-friendly mobile banking applications that cater to the diverse financial needs of users. This project aims to develop a comprehensive banking application using Python and SQL, addressing the challenges of user authentication, secure transactions, and efficient data management. The application seeks to improve user experience in digital banking, ensuring seamless access to various banking services like account management, money transfers, and transaction history tracking, while maintaining high standards of security and data integrity.

### 4. Major Project 4

#### **Bank Statement Aggregator(BSA)**

##### **Description:**

In the evolving landscape of financial technology, there is a need for robust, secure, and user-friendly systems to aggregate and manage bank statements from multiple sources. This project aims to develop a comprehensive bank statement aggregation system, addressing secure data handling, efficient database management, and seamless cloud storage integration. Users will be able to log in, view their bank statement details, and store them in a database, enhancing their ability to manage financial data effectively.

The system will generate bank statements, upload them to a cloud storage service (such as AWS S3), and later allow users to download and convert these statements into objects for database storage.

