Milestone – 2 – Data Analytics Project

Power BI Project – Data Analysis and Visualization

Project Overview:

In this project, you will demonstrate your ability to clean, transform, analyze, and visualize data using **Power BI**. You will select one dataset from the provided list, prepare it using **Power Query Editor**, apply **DAX functions** to create meaningful calculations, and build a dynamic **dashboard** that highlights key insights. Finally, you will document your analysis, provide actionable recommendations, and present your findings in a professional report format.

Learning Objectives:

By the end of this project, you will be able to:

1. Select and import datasets into Power BI for analysis.
2. Clean and transform raw data using **Power Query Editor**.
3. Apply at least 10 **DAX functions** to derive key metrics and aggregations.
4. Build interactive dashboards with slicers, filters, and multiple visualization types.
5. Summarize insights through a written report highlighting key trends, patterns, and recommendations.

Selection of Dataset:

You need to choose one dataset from the list provided below. The datasets come from Kaggle and cover a variety of domains, such as sales, healthcare, sports, and more.

Available Datasets from Kaggle (Select One):

1. **Sales Data – Online shop**
   - **Link:** https://www.kaggle.com/datasets/gabrielramos87/an-online-shop-business
   - **Description:** This is a sales transaction data set of UK-based e-commerce (online retail) for one year. This London-based shop has been selling gifts and homewares for adults and children through the website since 2007.
2. **Heart Disease Dataset**
   - **Link:** https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

- **Description:** Data related to patients' medical information, which could be used to predict the likelihood of heart disease.

3. **Global Health Data Analysis**
   - **Link:** https://www.kaggle.com/datasets/kamaumunyori/global-health-data-analysis-1990-2019
   - **Description:** The "Global Health Data Analysis 1990-2019" dataset provides comprehensive health-related statistics across various countries, covering indicators such as life expectancy, disease prevalence, and healthcare expenditure over nearly three decades.
4. **Car Price Prediction**
   - **Link:** https://www.kaggle.com/datasets/zafarali27/car-price-prediction
   - **Description:** Contains car attributes such as make, model, year, mileage, etc., to predict car prices.
5. **Sales Forecasting Data**
   - **Link:** https://www.kaggle.com/datasets/tevecsystems/retail-sales-forecasting
   - **Description:** Sales forecasting dataset to predict future sales performance for retail businesses.
6. **Employee Attrition Dataset**
   - **Link:** https://www.kaggle.com/datasets/colearninglounge/employee-attrition
   - **Description:** Employee attrition dataset including employee details, job satisfaction, attrition status, and performance data.
7. **Movie Ratings Dataset**
   - **Link:** https://www.kaggle.com/datasets/asaniczka/tmdb-movies-dataset-2023-930k-movies
   - **Description:** A dataset containing movie ratings, genres, cast, crew, and other related data, suitable for sentiment analysis or recommendation systems.
8. **COVID-19 Dataset**

- **Link:** https://www.kaggle.com/datasets/imdevskp/corona-virus-report
- **Description:** A dataset containing COVID-19 case details in India state wise.

9. **Retail Store Transactions**
   - **Link:** https://www.kaggle.com/datasets/marian447/retail-store-sales-transactions
   - **Description:** Dataset with retail store transactions data, including product purchases, customer details, and sales amounts.
10. **World Happiness Report**
    - **Link:** https://www.kaggle.com/datasets/utkarshsen/happiness-dataset
    - **Description:** Data on happiness rankings across countries, with factors such as GDP per capita, life expectancy, social support, and more.

Project Tasks:

Project documentation must contain the following sections:

1. **Data Cleaning & Transformation (Power Query Editor)**
   - Handle missing values.
   - Remove duplicates and irrelevant data.
   - Standardize data types and formats.
   - Perform transformations such as splitting, merging, or reshaping data.
2. **DAX Functions**
   - Apply at least **10 DAX functions** to calculate measures and KPIs.
   - Examples: SUM(), AVERAGE(), COUNT(), DISTINCTCOUNT(), IF(), CALCULATE(), DATEADD(), RELATED(), ALL().
3. **Dashboard Creation**
   - Design a Power BI dashboard with:
     - Multiple visualization types (bar, line, pie, map, etc.).

- Filters and slicers for interactivity.
- A clean and logical layout to highlight insights.

4. **Insights & Analysis Report**
    - Provide a written summary that includes:
        - Brief dataset description and project objective.
        - Key observations, patterns, and anomalies.
        - Trends and relationships identified in the data.
        - Actionable recommendations based on findings.

Deliverables

1. **Power BI File (.pbix)** with cleaned dataset, DAX measures, and dashboard.
2. **Insights & Analysis Report (PDF/Word)** summarizing findings and recommendations.

Evaluation Rubric:

Evaluation Criteria	Marks
Data Cleaning & Transformation (Power Query Editor)	25
DAX Functions (Application of at least 10 functions)	25
Dashboard Design (Variety, Aesthetics, Interactivity)	25
Insights & Analysis Report (Depth & Quality of Insights)	25
Total	**100**