PDO-1

Name: Ahmed Khawar
Reg No#: L1F20BSSE0558
Subject : Big Data Analytics

Project Planning:

Purpose: The purpose of this project is to analyze historical airplane crash data to enhance understanding of factors influencing air safety. By conducting a comprehensive examination of past incidents, the project aims to provide insights that can contribute to improved safety measures in the aviation industry.

Problem Description: The aviation industry faces the ongoing challenge of ensuring the safety of air travel. This project addresses the need to identify patterns and contributing factors to airplane crashes, enabling the industry to proactively implement safety enhancements.

Objectives:

a.Identify common factors contributing to airplane crashes, such as weather conditions, technical issues, or human factors.

b. Analyze trends in air safety over time, including changes in the frequency and severity of incidents.

c. Develop predictive models to forecast potential high-risk scenarios or regions.

Desired Outcomes:

a. Identification of key risk factors contributing to airplane crashes.

b. Establishment of actionable recommendations for enhancing aviation safety.

c. Development of a predictive model with a minimum accuracy of 80% in forecasting high-risk scenarios.

Consideration: Ensure that the dataset includes comprehensive information on airplane crashes, covering factors such as weather conditions, technical details, and human-related aspects. Additionally, verify that the dataset spans a sufficiently long time frame to capture historical trends and is sourced from reliable aviation databases and records.

Dataset Name: Airplane Crashes and Fatalities

Data Dictionary:

1. **Date:** Date of the airplane crash.
2. **Time:** Time of the airplane crash.
3. **Location:** Location where the crash occurred.
4. **Operator:** Airline or company operating the airplane.

5. **Flight:** Flight number.
6. **Route:** Flight route.
7. **Type:** Type of airplane.
8. **Registration:** Aircraft registration number.
9. **cn/In:** Construction or serial number of the airplane.
10. **Aboard:** Total number of people on board.
11. **Fatalities:** Number of fatalities in the crash.
12. **Ground:** Number of ground fatalities.
13. **Summary:** Brief summary of the crash.

Data Collection Plan:

- Utilize aviation databases, accident reports, and official records.
- Collect data spanning several decades to capture historical trends.

Raw Data: A sample of the raw data, with 100 entries, can be obtained from the dataset source.

Data Description:

- Data Type: Primarily numerical and categorical.
- Features: Date, Time, Location, Operator, Flight, Route, Type, Registration, cn/In, Aboard, Fatalities, Ground, Summary.
- Dataset Size: Varies based on the dataset source; typically includes information on hundreds to thousands of airplane crashes.

Objectives:

Identify Contributing Factors:
Goal: Identify and quantify the key factors contributing to airplane crashes.
Measurement: Develop a comprehensive list of variables, and conduct statistical analyses to determine the significance of each factor.
Temporal Analysis:
Goal: Analyze trends in air safety over time to understand any temporal patterns or changes.
Measurement: Create time-series visualizations and statistical models to identify trends, spikes, or decreasing patterns in airplane crashes.
Predictive Modeling:
Goal: Develop predictive models to forecast potential high-risk scenarios or regions.
Measurement: Build machine learning models with a focus on accuracy, precision, and recall. Achieve a minimum accuracy of 80% in forecasting high-risk scenarios.
Demographic Analysis:
Goal: Explore the impact of different variables, such as aircraft type, operator, or route, on crash outcomes.
Measurement: Conduct demographic analyses to understand the distribution of crashes across various categories and identify high-risk profiles.

Desired Outcomes:

Key Risk Factors Identification:

Insight: Identify and rank the most significant factors contributing to airplane crashes, such as weather conditions, technical issues, or human factors.

Temporal Trends Understanding:

Insight: Uncover trends in air safety over time, enabling the identification of periods with increased or decreased safety incidents.

Predictive Model Implementation:

Insight: Implement a predictive model capable of forecasting potential high-risk scenarios or regions, providing a proactive approach to safety management.

Demographic Risk Profiling:

Insight:Understand how different variables impact crash outcomes, allowing for the development of targeted safety measures for specific profiles.