**Statement of Work: Plane Crash Database Analysis**

Saphyra Dulcio, Noah Growney, Shrihari Gunasekaran,

Madhan Saravanan, Janny Velazquez

**Background**

The objective of this project is to analyze historical plane crash data from 1920 to 2020. This analysis aims to identify patterns, trends, and contributing factors to aviation accidents over the past century.

**Scope**

This project may include, but is not be limited to:

* Collecting and cleaning plane crash data.
* Conducting exploratory data analysis (EDA) to identify trends over time.
* Analyzing contributing factors such as weather, mechanical failures, human error, and other variables.
* Developing predictive models to assess risks based on historical data.
* Delivering visualizations and a comprehensive report summarizing findings.

**Requirements**

* Cleaned and documented dataset.
* Interactive visualizations highlighting trends and patterns.
* Predictive models with documented methodology.
* Final report detailing:
  + Key findings.
  + Recommendations for aviation safety.
  + Limitations of the study.
  + Presentation for stakeholders summarizing the project outcomes.

**Tasks**

* Data Collection: Identify and acquire datasets on aviation incidents from 1920-2020.
* Data Cleaning: Handle missing values, inconsistencies, and duplicate entries.
* Exploratory Data Analysis (EDA): Identify trends in crash frequencies, locations, and contributing factors and analyze data by time periods (e.g., decades) and aircraft types.
* Predictive Modeling: Build machine learning models to predict potential risk factors.
* Visualization & Reporting:Create visualizations to communicate findings and draft a detailed report with recommendations for improving aviation safety.

**Deliverables**

* Cleaned and documented dataset.
* Interactive visualizations highlighting trends and patterns.
* Predictive models with documented methodology.
* Final report detailing with key findings
* Limitations of the study.

**Schedule**

* Problem Identification and Business Understanding (Week 1-2):Define the project objectives, identify the problem to solve, and establish its relevance to the business or study. Finalize project goals and ensure alignment with stakeholders.
* Data Collection and Exploration (Week 3-4): Gather relevant datasets and conduct an initial exploration to understand the data’s structure, features, and potential issues. Perform basic exploratory data analysis (EDA).
* Data Preparation and Cleaning (Week 5-6): Clean and preprocess the collected data to address missing values, inconsistencies, and outliers. Prepare the data for modeling, including feature engineering and transformation.
* Data Modeling and Analysis (Week 7-8): Build and refine predictive models based on the prepared data. Conduct in-depth analysis to draw insights and evaluate model performance.
* Model Evaluation and Interpretation of Results (Week 9-10): Assess model accuracy and reliability using appropriate metrics. Interpret the results in the context of the identified problem and derive actionable insights.
* Deployment and Communication of Findings (Week 11-12): Deploy the final solution or share the outcomes. Create visualizations, prepare a comprehensive report, and communicate findings effectively to stakeholders.