Car Insurance Claim Prediction

Submitted by: Gokul Priya

GUVI Data Science Course Project

Project Objective

This project aims to predict whether a customer will make a car insurance claim based on their demographic and policy details. The solution integrates data cleaning, exploratory data analysis (EDA), SQL analytics, machine learning modeling, and Power BI visualization to derive actionable insights.

Tools & Technologies Used

- Python
- SQL
- Pandas, NumPy, Matplotlib, Seaborn
- Scikit-learn
- Power BI

Dataset Details

- Source: Public or GUVI provided dataset
- Features: Customer demographics, policy details, claim history
- Size: 5000+ records
- Preprocessing: Missing value handling, encoding categorical features, scaling numeric features

Exploratory Data Analysis (EDA)

EDA involves visualizing data distributions, correlations, and feature importance. The insights help identify patterns and relationships between customer attributes and claim behavior.

[Mockup Charts & Graphs Here]

Model Building & Evaluation

Machine learning models such as Logistic Regression, Decision Trees, and Random Forests were used to predict insurance claims. Model evaluation metrics include Accuracy, Precision, Recall, and F1-Score.

[Mockup Tables Here]

Power BI Dashboard Visualizations

Power BI dashboards provide interactive visualizations of claim statistics, customer demographics, and policy trends.

[Mockup Dashboard Images Here]

Conclusion & Future Work

The project demonstrates the capability to predict car insurance claims using data-driven insights. Future work can include deploying the model as an API, integrating real-time claim data, and improving visualization interactivity in Power BI.