Data Analytics

Name:-Manish Kumar

Start Date :-01-05-2025

End Date:-05-05-202

slide-2

Title :- "Insurance Claim Fraud Detection“

Domain:- "Finance“

Discription:- "This project explores insurance claim data to find signs of fraud. It uses simple analysis and charts to understand what makes a claim suspicious."

Skills:- "1 Python","2 Pandas","3 Statistics", "4 Data manipulation","5 Data Analysis", "6 Data transformation","7 Data cleaning", "8 Data visualization", "9 Project methodology"

Slide-3

- Introduction
- **Data Analytics** is the **process** of **collecting**, **cleaning**, **examining**, and **visualizing** data to find **useful information**, **patterns**, and **insights** for better **decision-making**.
- → To Pridicting Insurance Claim Fraud Detection.

Slide-4

- Objective.

- **Identify the most and least used appliances** in smart homes.
- **Measure energy consumption** across different devices
- (like Television, Oven, Refrigerator, etc.).

Slide-5

- **Data description**
- There are 1000 rows and 39columns.
- There are some count variable and some continuous.
- Gather raw data from databases, exel, websites, etc

Slide-6

- **Data Preprocessing**

- **Clean Data** Fix missing values, wrong formats, remove duplicates,Outliers
- **Visualize Data** Make graphs (bar charts, pie charts, heatmaps) to understand better
- **Get Insights** Find answers to important business or research questions

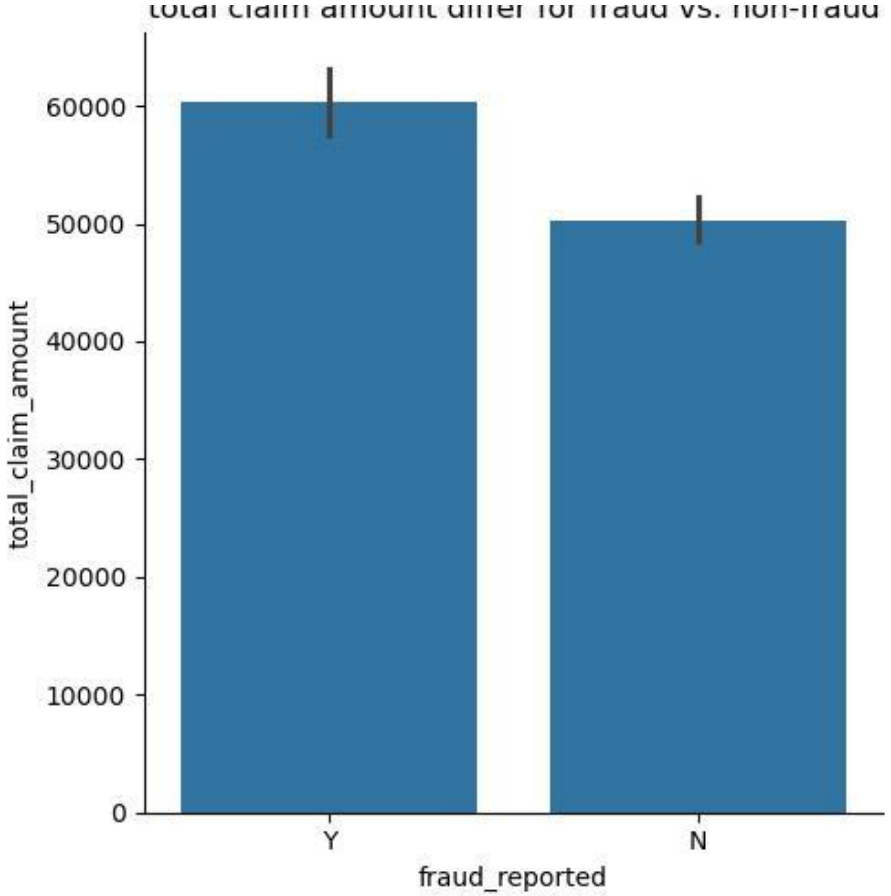
Slide-7

- Exploratory Data Analysis (EDA)
- visualizations: by using pie plot.
- distribution of fraud vs. non-fraud.
- Approx 75% real insurance.
- Only 25 % froud insurance

slide-8

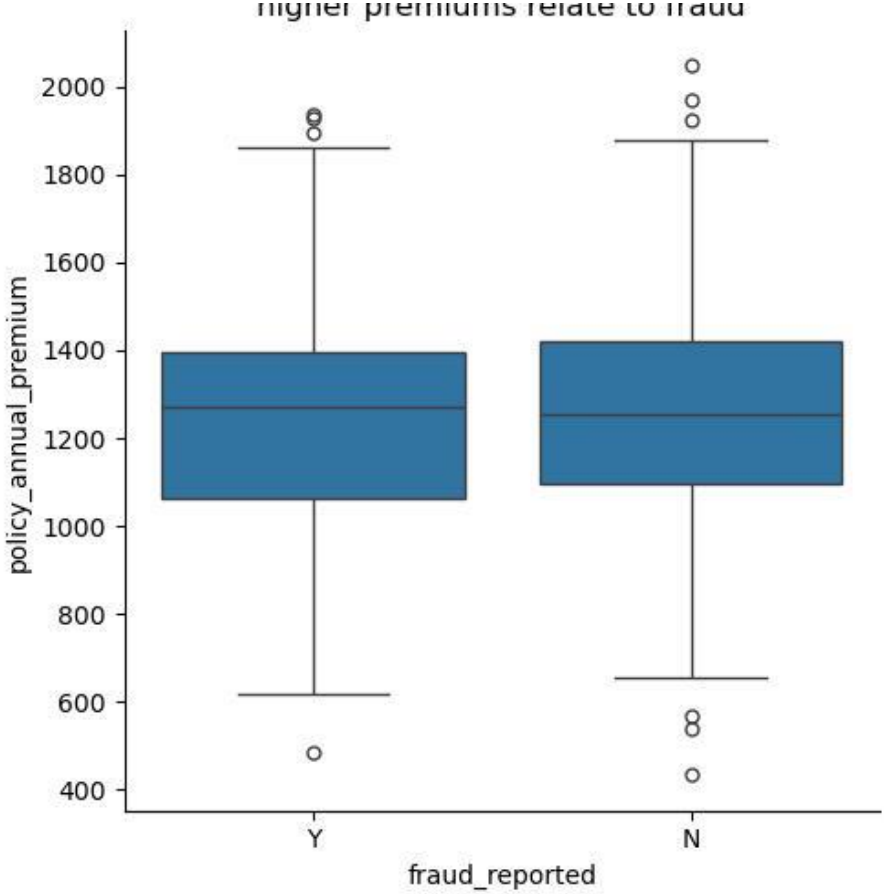
- Exploratory Data Analysis (EDA)
- visualizations: by using cat plot.
- distribution of fraud vs. non-fraud.
- total claim amount differ for fraud vs. non-fraud.

N= 50288.605578
Y =60302.105263



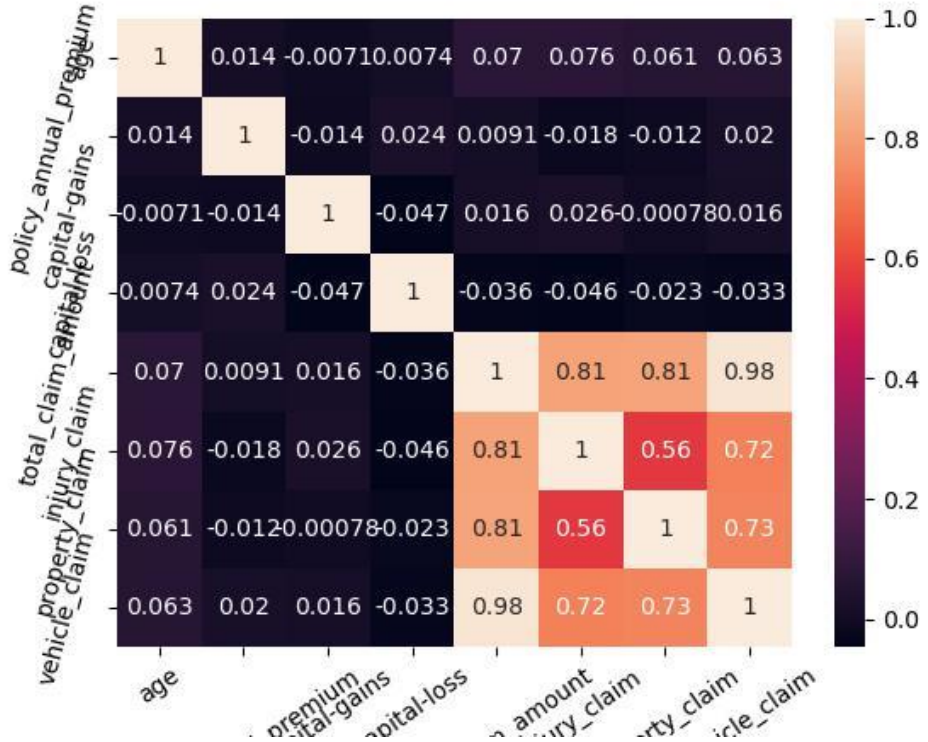
Slide-9

- Exploratory Data Analysis (EDA)
- visualizations: by using cat plot.
- the higher premiums relate to fraud.
- 1935.85 maximum



Slide-10

- Exploratory Data Analysis (EDA)
- visualizations: by using heat map.



Slide-10

Strong Positive Correlations:-

total_claim_amount has a very strong positive correlation with vehicle_claim (0.98), property_claim (0.81),and injury_claim (0.81).

Weak Correlations:-

policy_annual_premium, capital-gains, and capital-loss also generally show weak correlations with each other.
age shows very weak correlations with most other variables.

Near-Zero or Very Weak Negative Correlations:-

There are some very weak negative correlations, such as between age and capital-gains (-0.0071) #or age and capital-loss (-0.0074).