

# Healthcare Fraud Detection

## Exploratory Data Analysis Report

Dataset: Medicare Claims Data

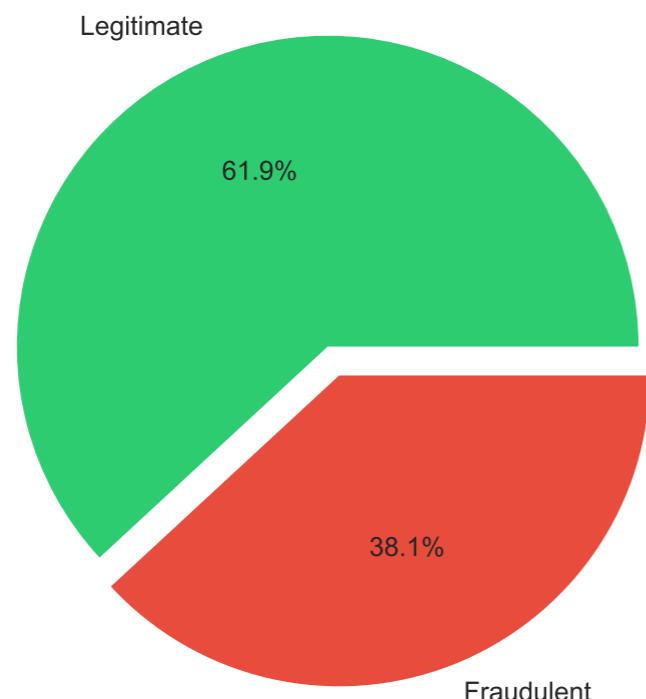
Total Records: 558,211

Unique Providers: 5,410

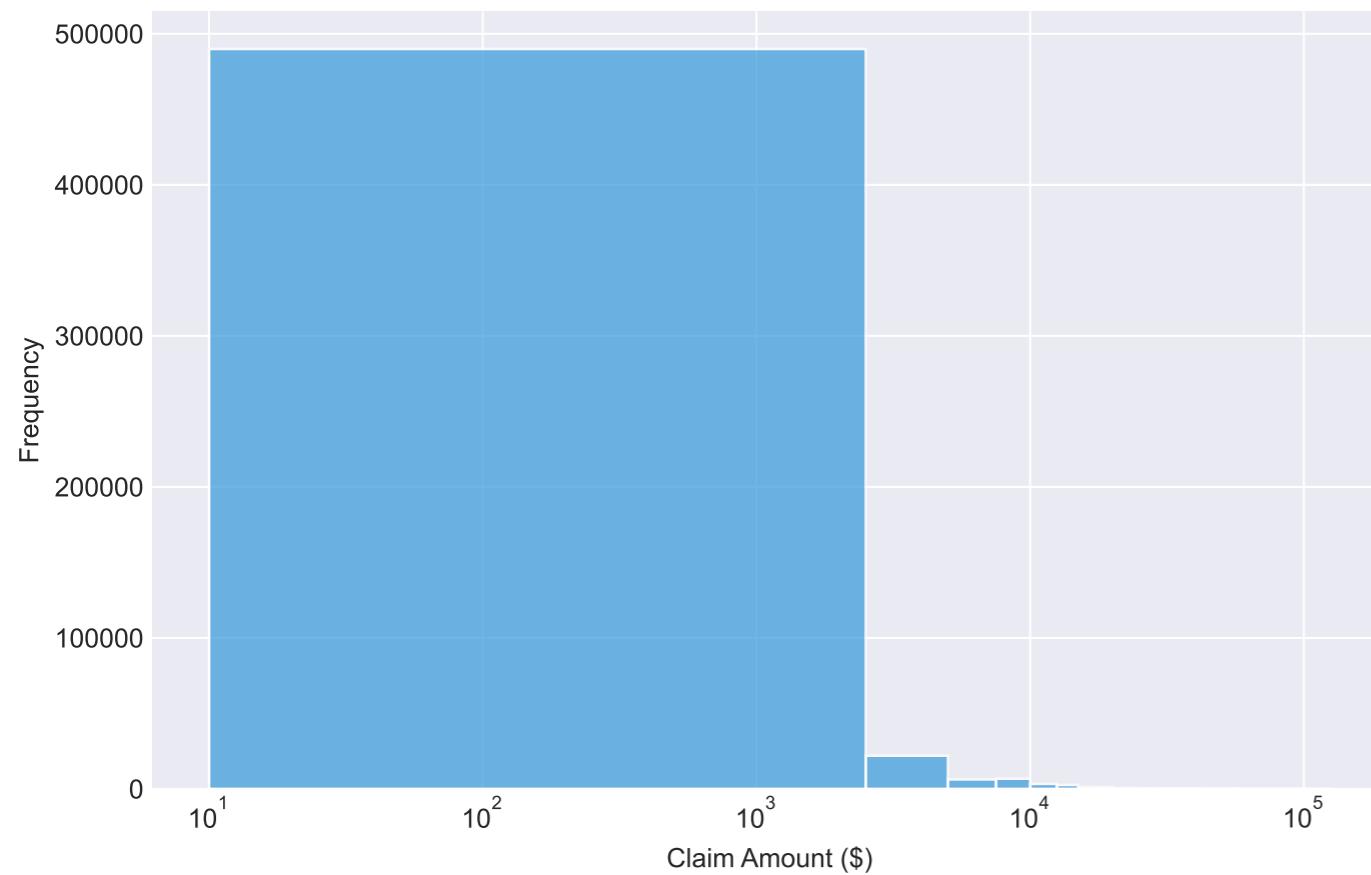
Fraud Rate: 38.12%

# Dataset Overview

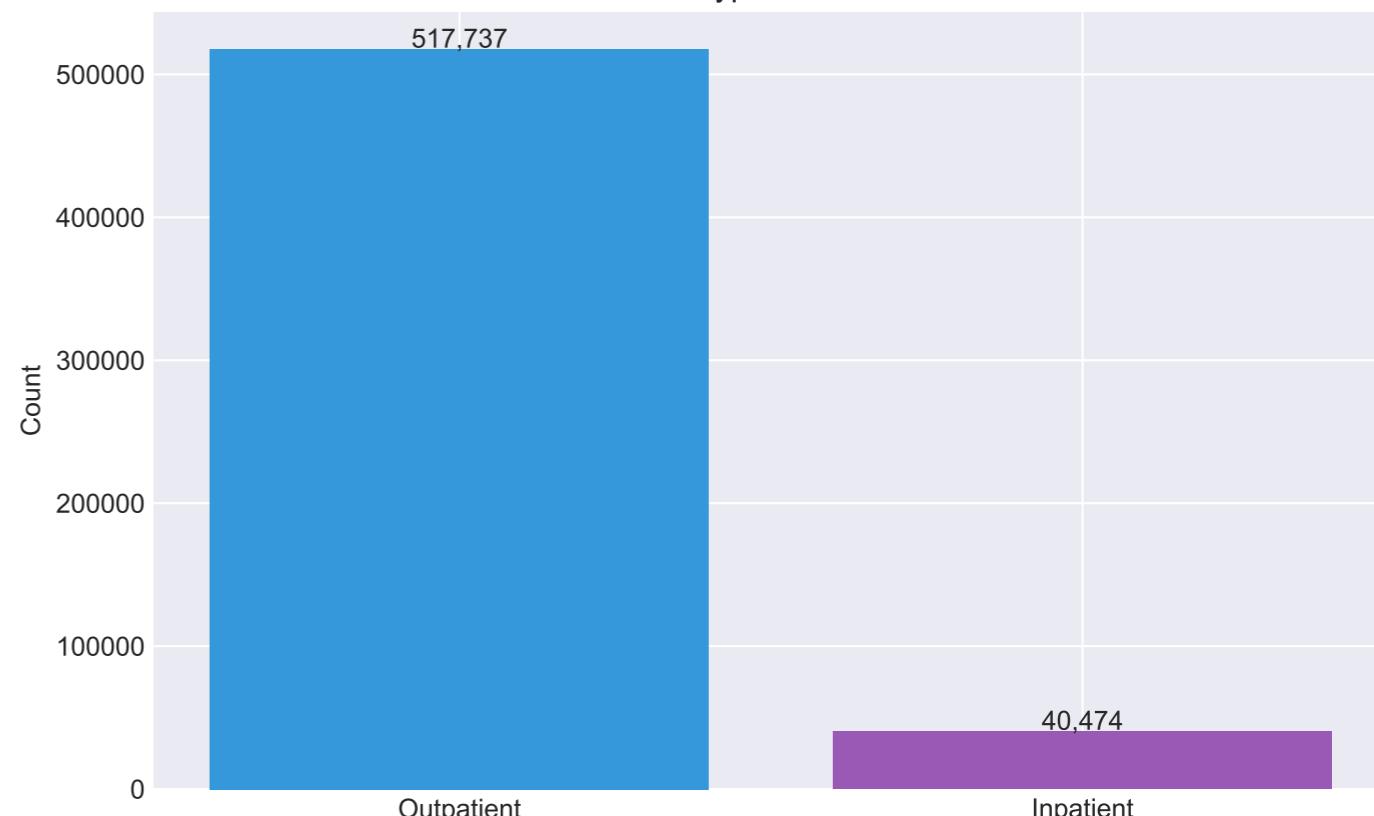
Fraud Distribution



Amount Distribution



Claim Type Distribution



## Dataset Statistics

=====

Total Claims: 558,211  
Unique Providers: 5,410  
Unique Patients: 138,556

## Fraud Statistics

=====

Fraudulent Claims: 212,796  
Legitimate Claims: 345,415  
Fraud Rate: 38.12%

## Financial Statistics

=====

Total Amount: \$556,543,140.00  
Mean Amount: \$997.01  
Median Amount: \$80.00  
Max Amount: \$125,000.00

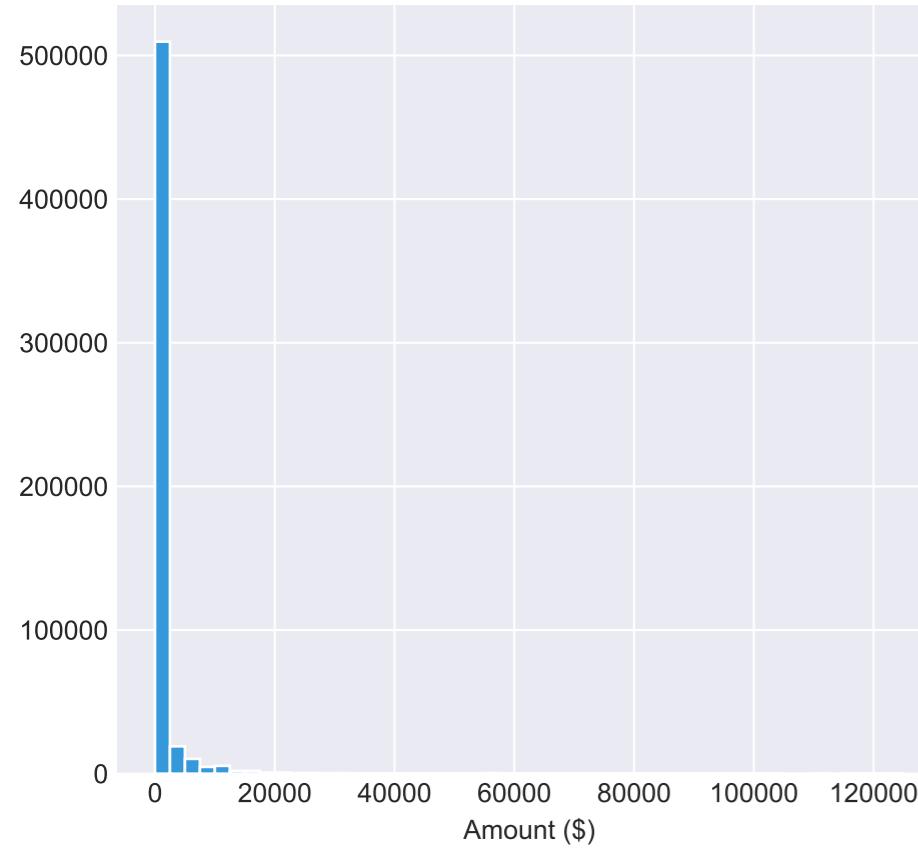
## Claim Types

=====

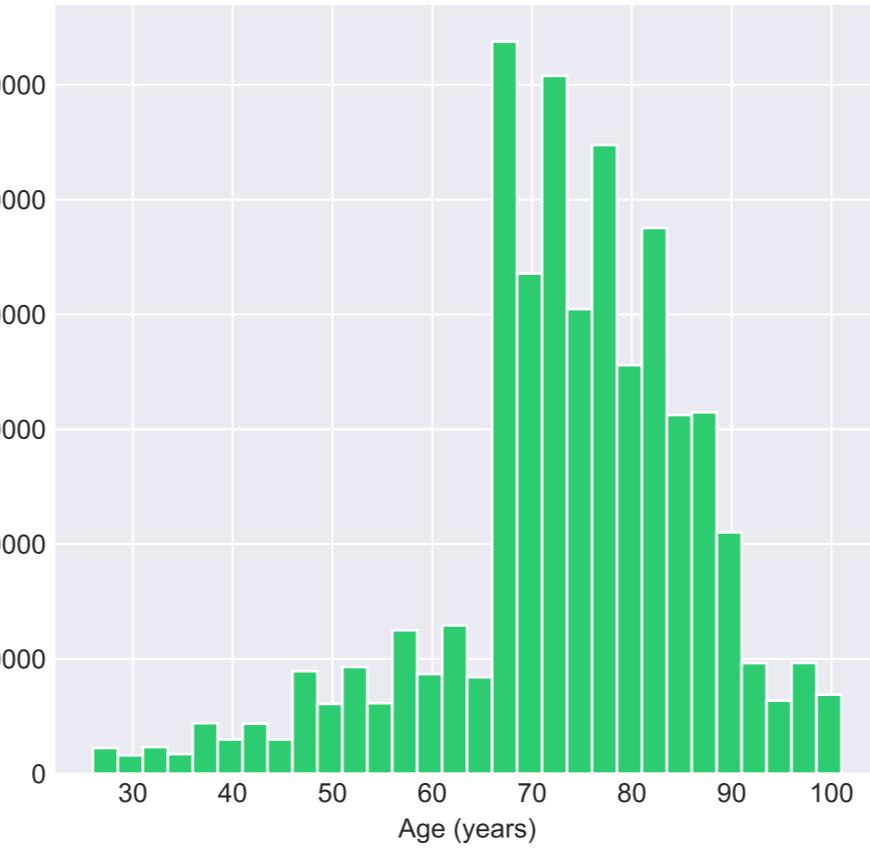
Inpatient: 40,474  
Outpatient: 517,737

# Numerical Feature Distributions

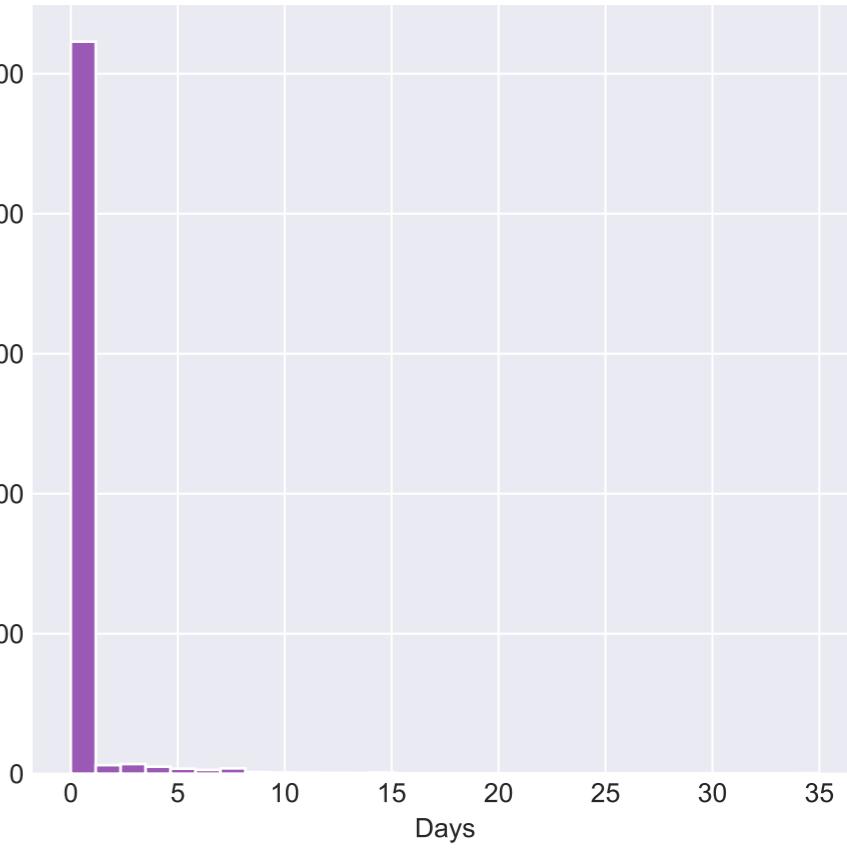
Claim Amount



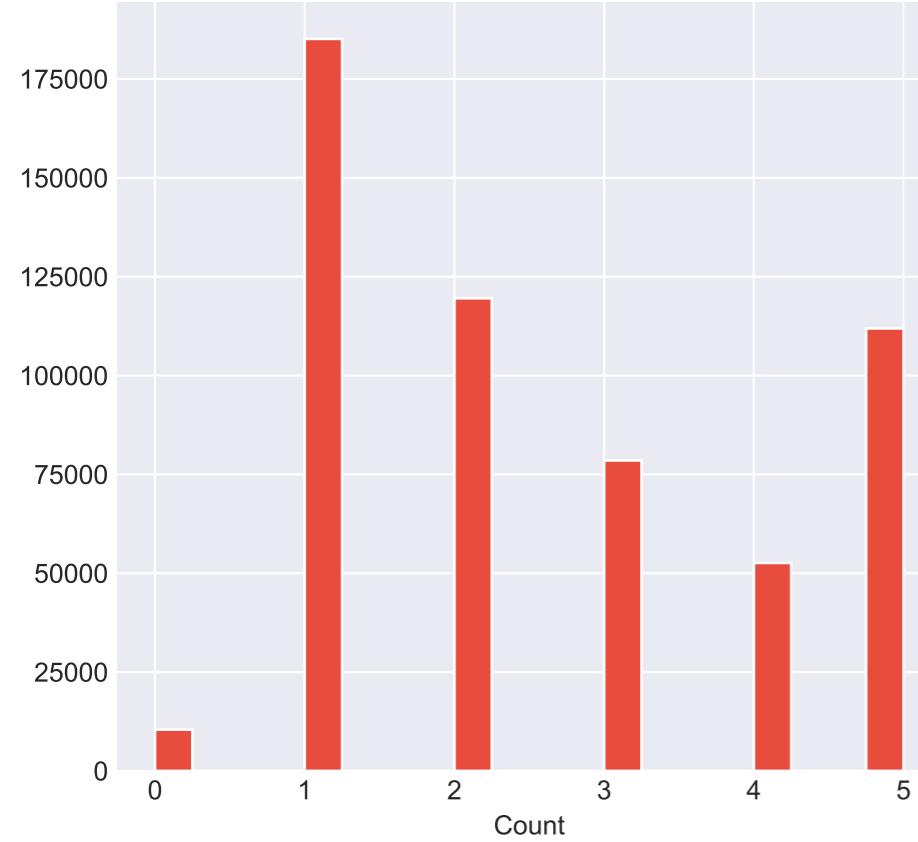
Patient Age



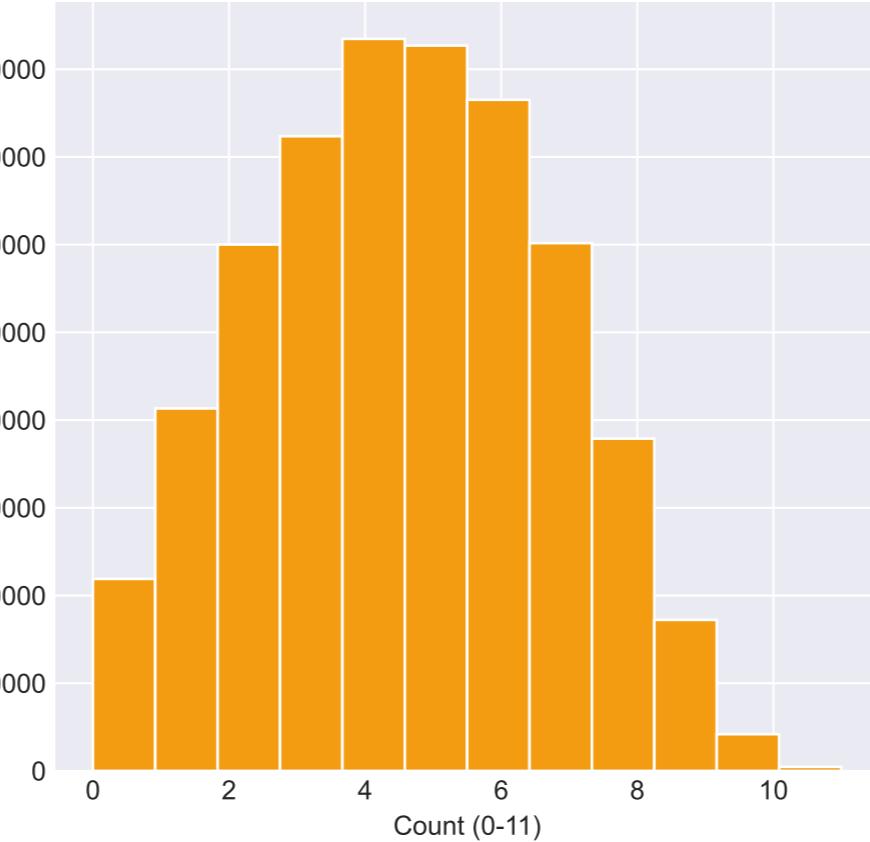
Length of Stay



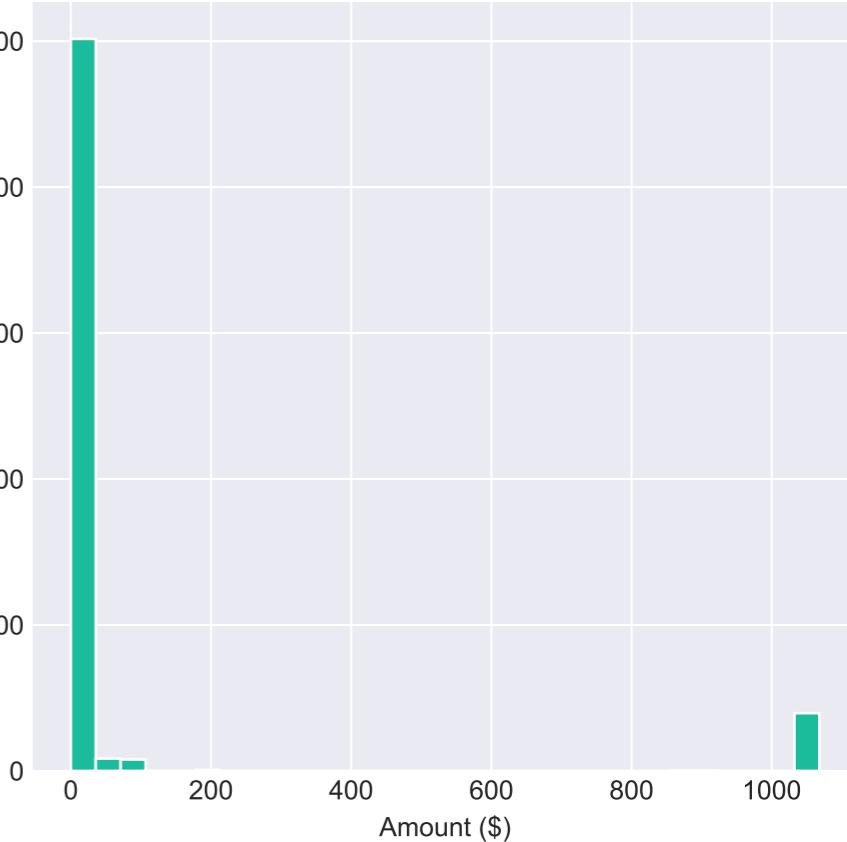
Number of Diagnoses



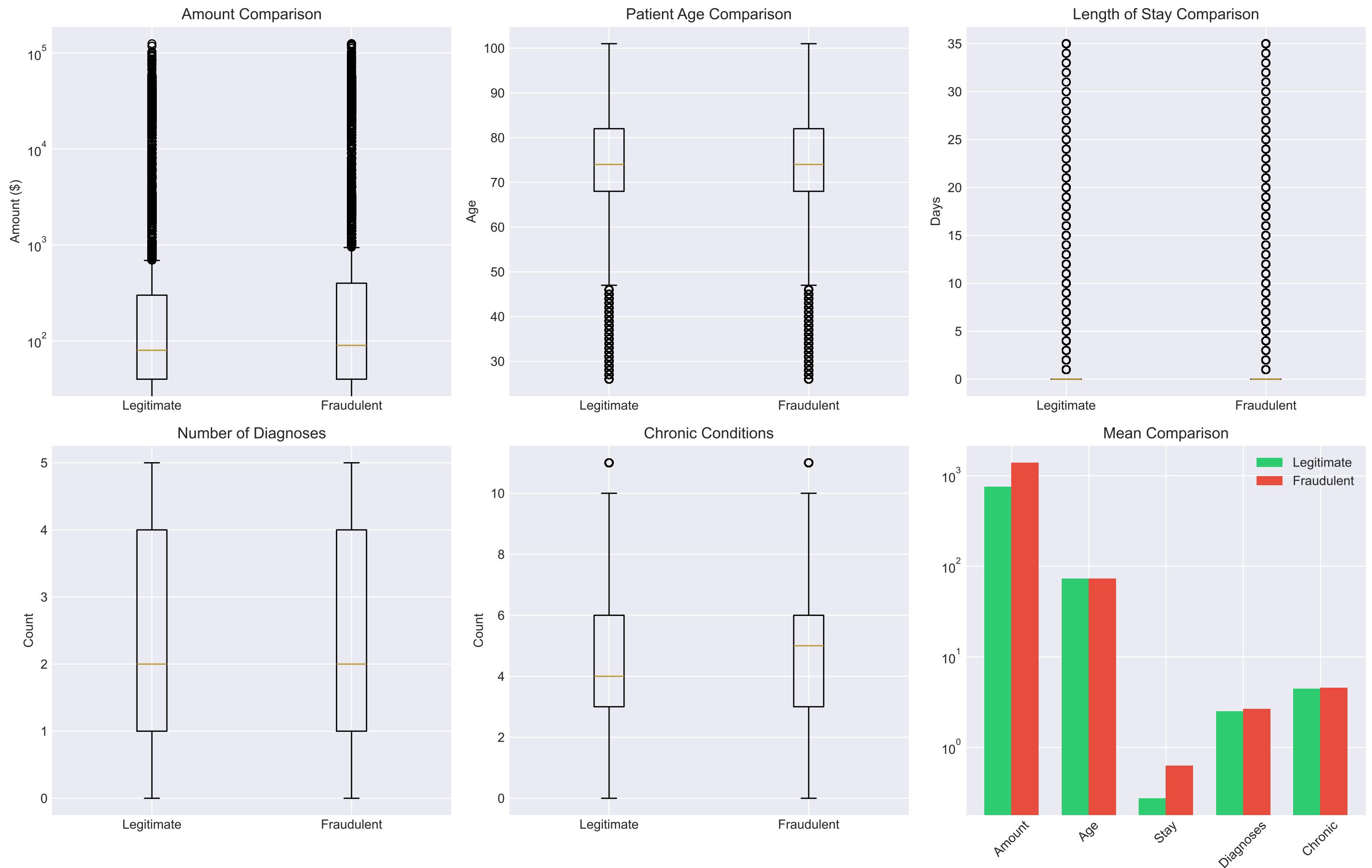
Chronic Conditions

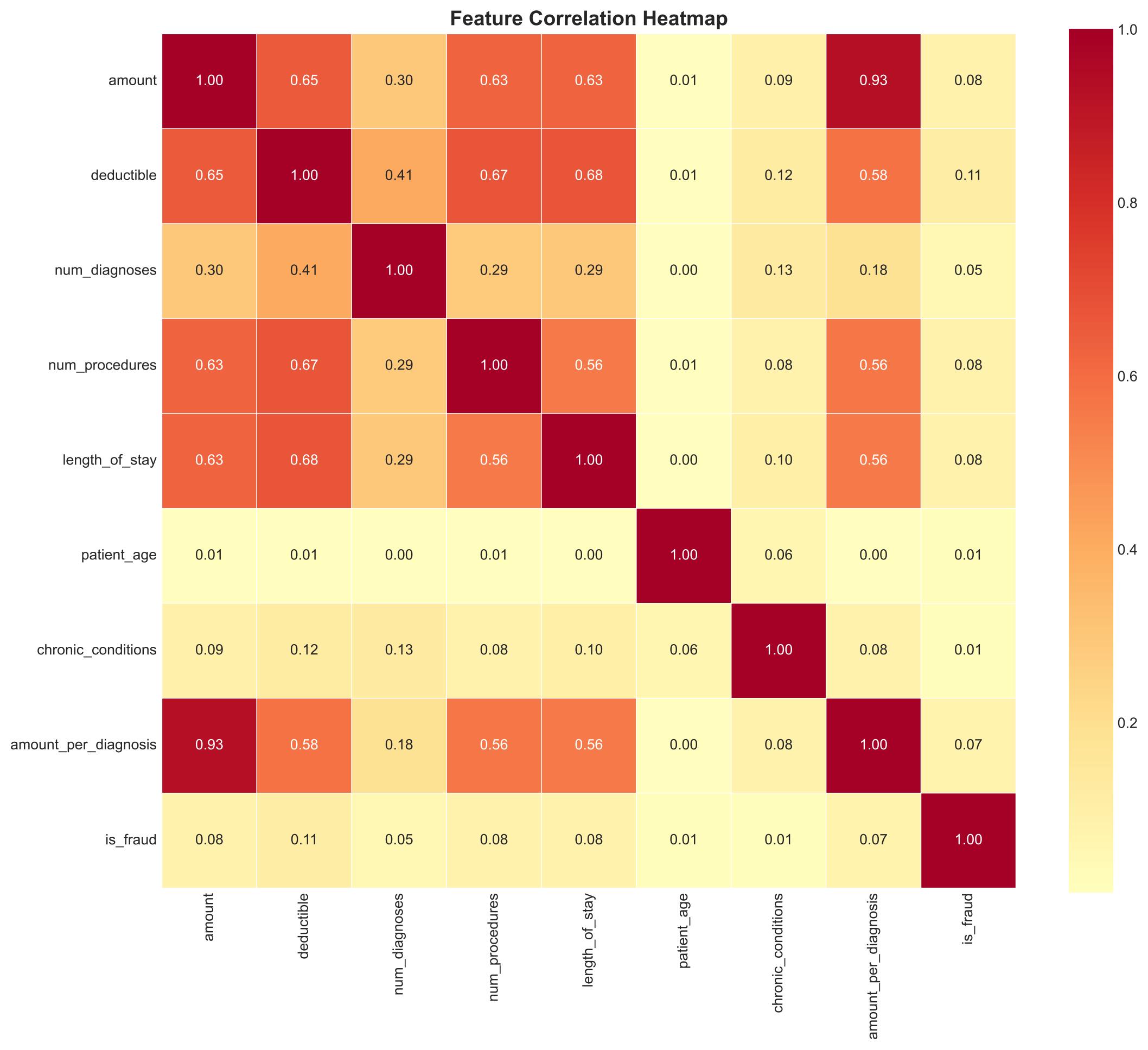


Deductible Amount



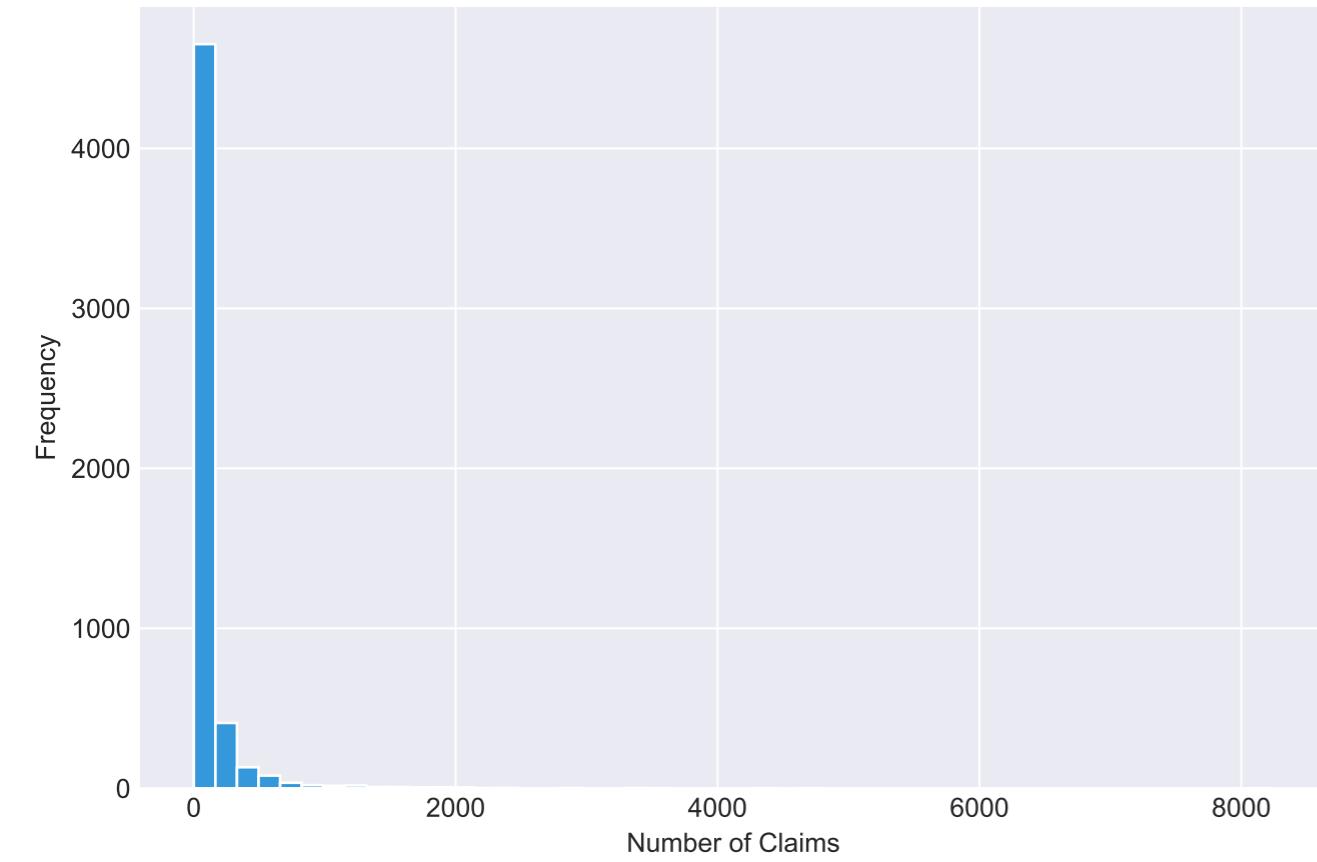
# Fraud vs Non-Fraud Comparison



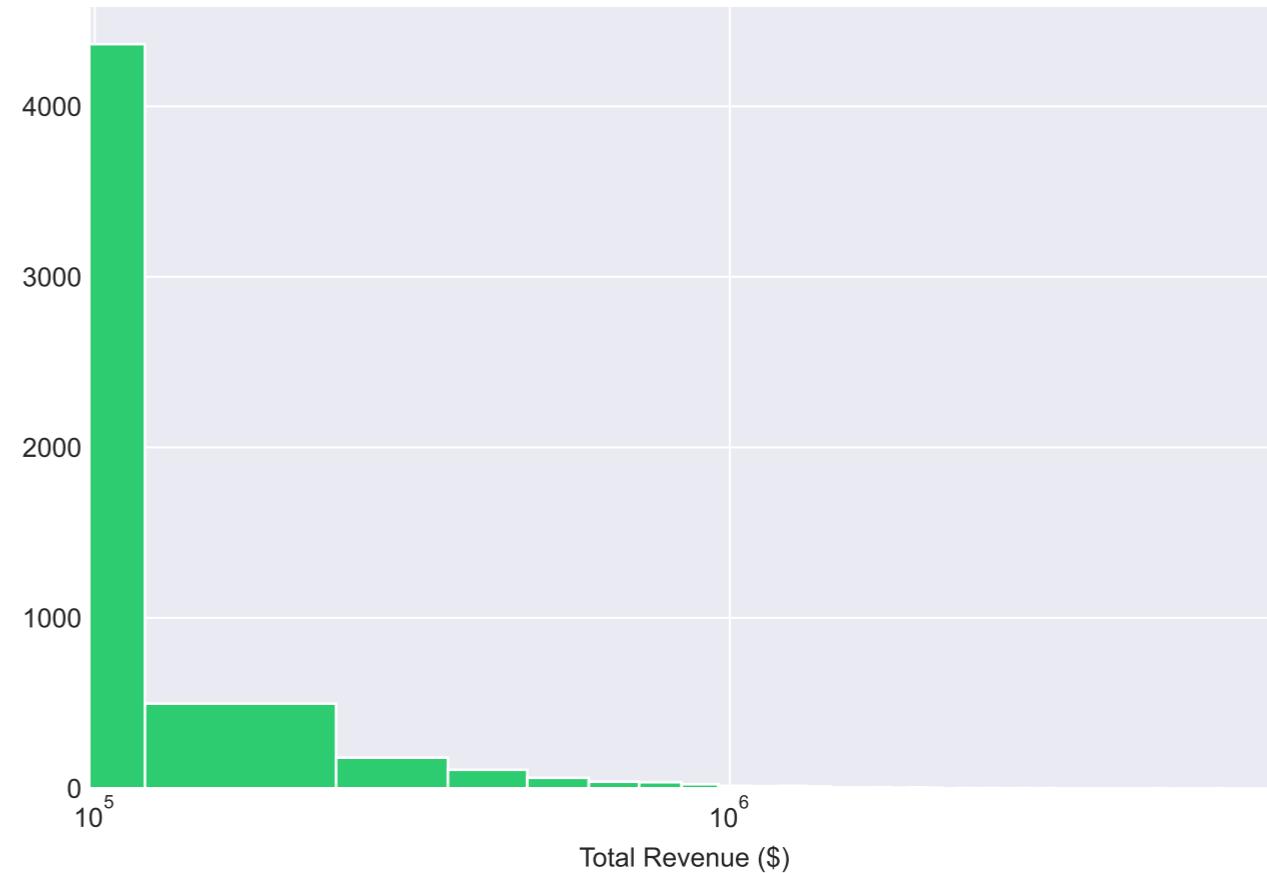


# Provider-Level Analysis

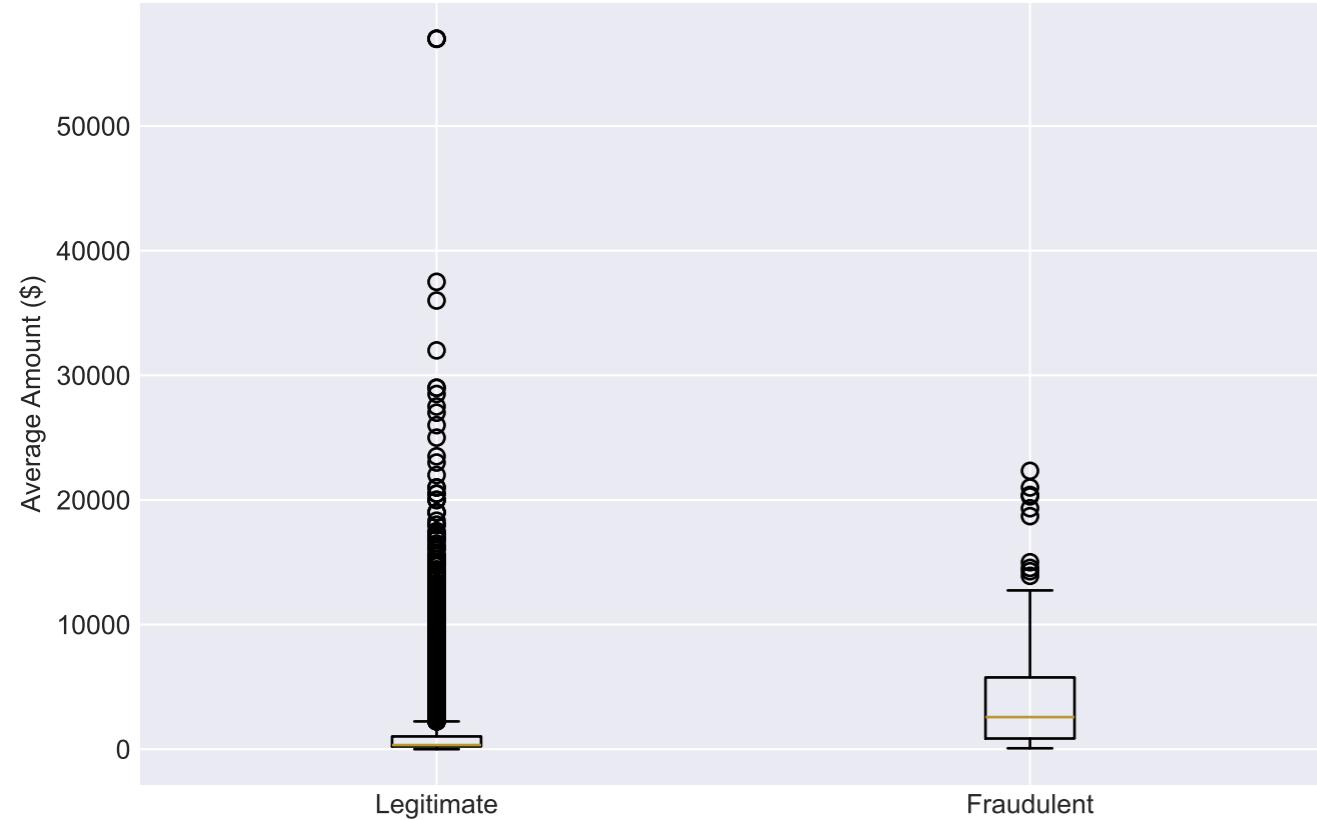
Claims per Provider



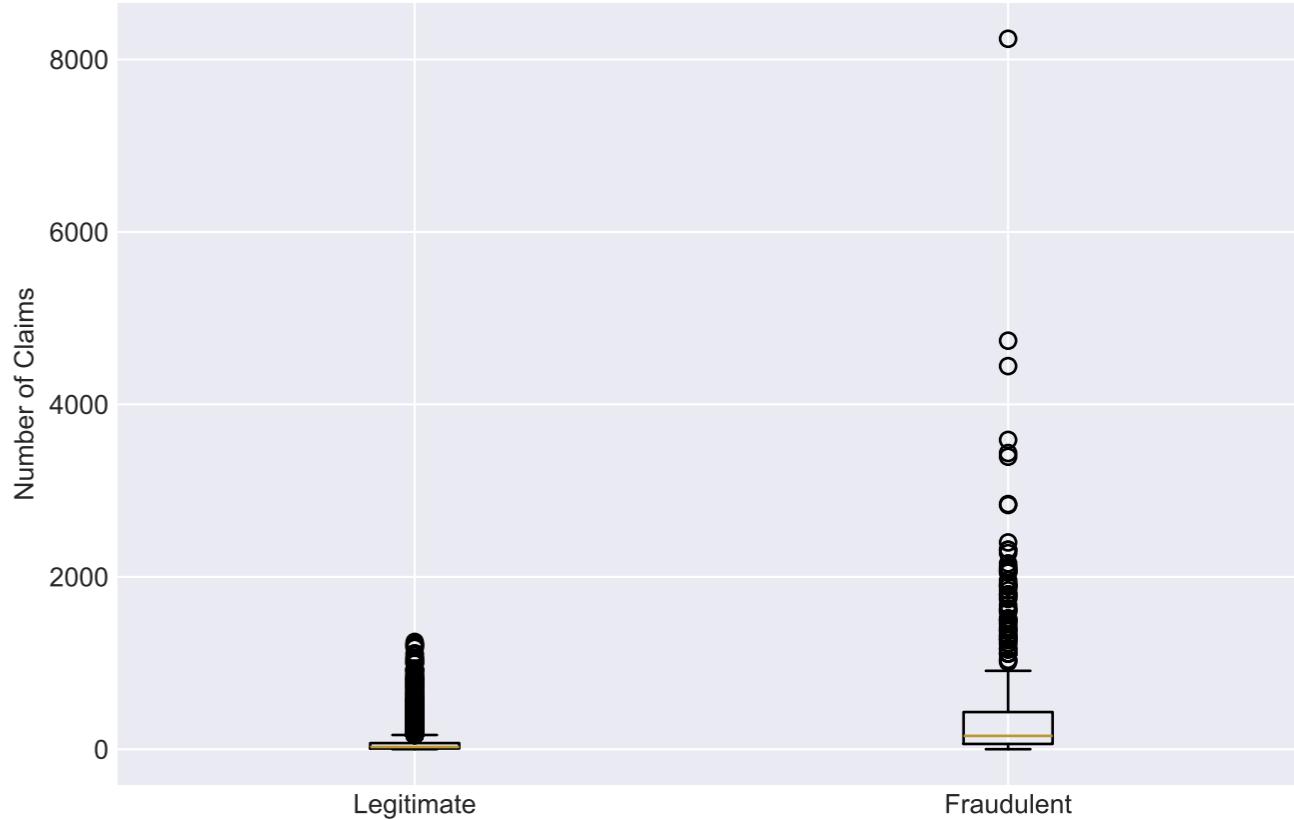
Revenue per Provider



Avg Claim Amount by Provider Type

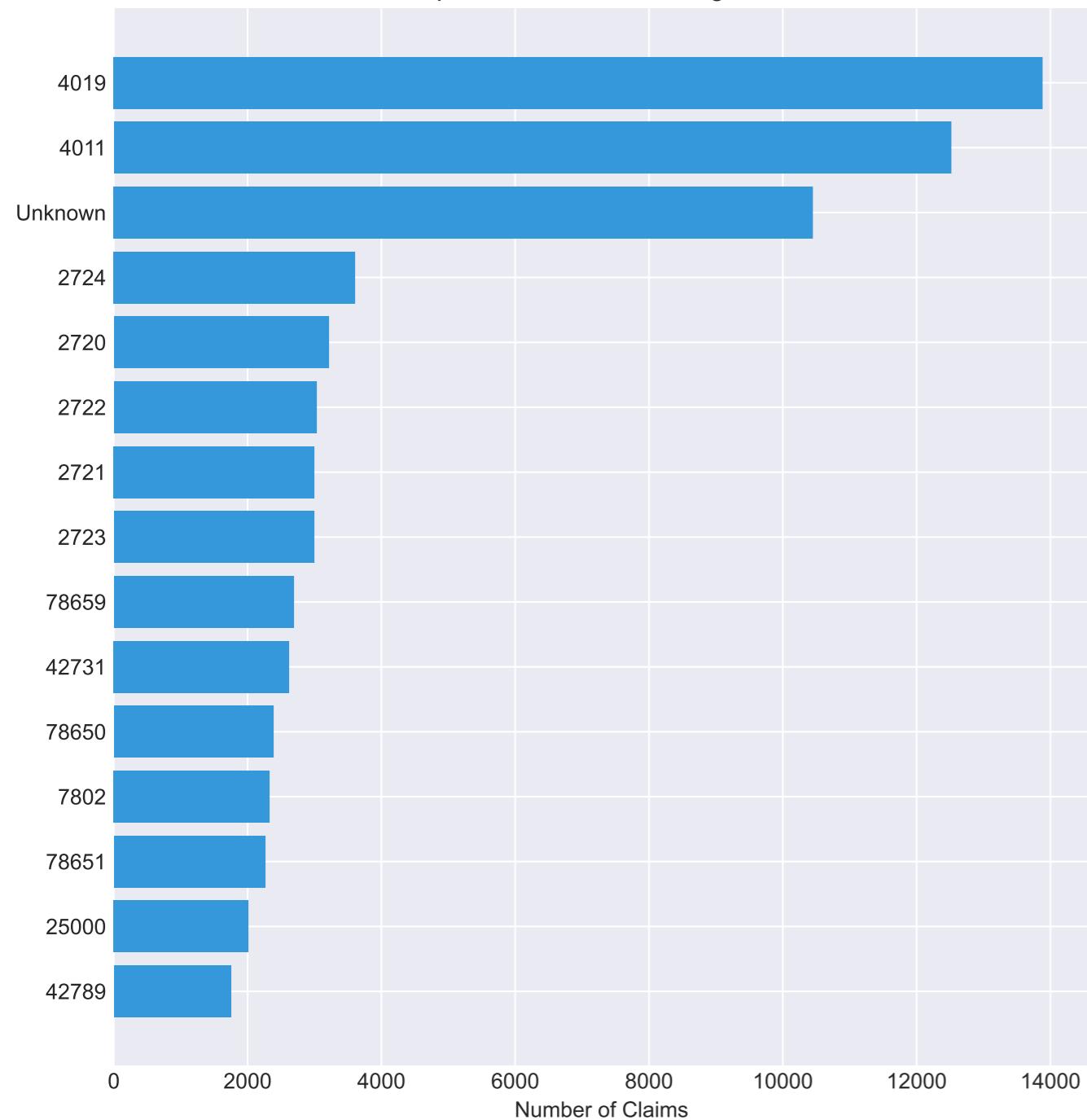


Claim Count by Provider Type

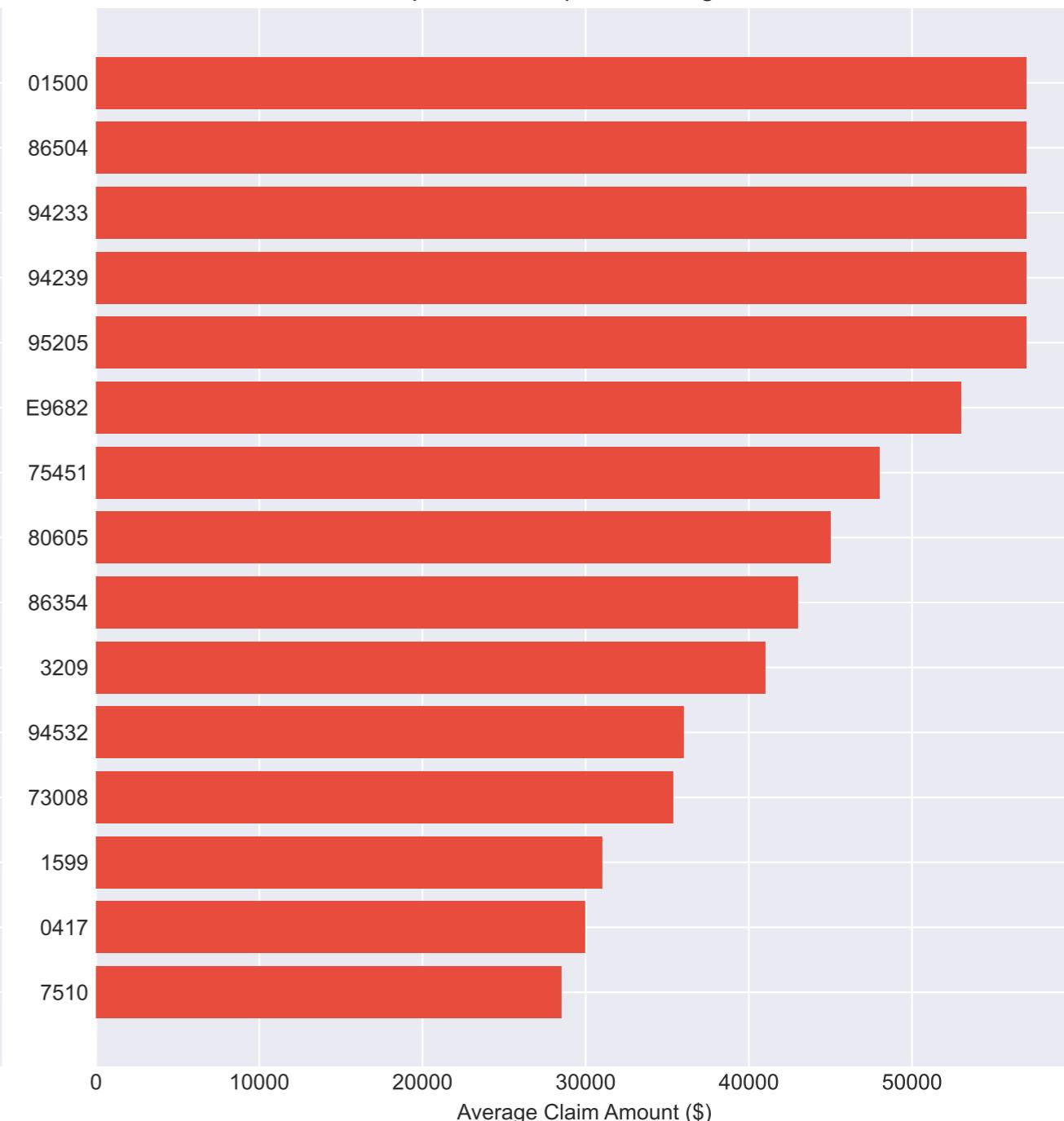


# Diagnosis Code Analysis

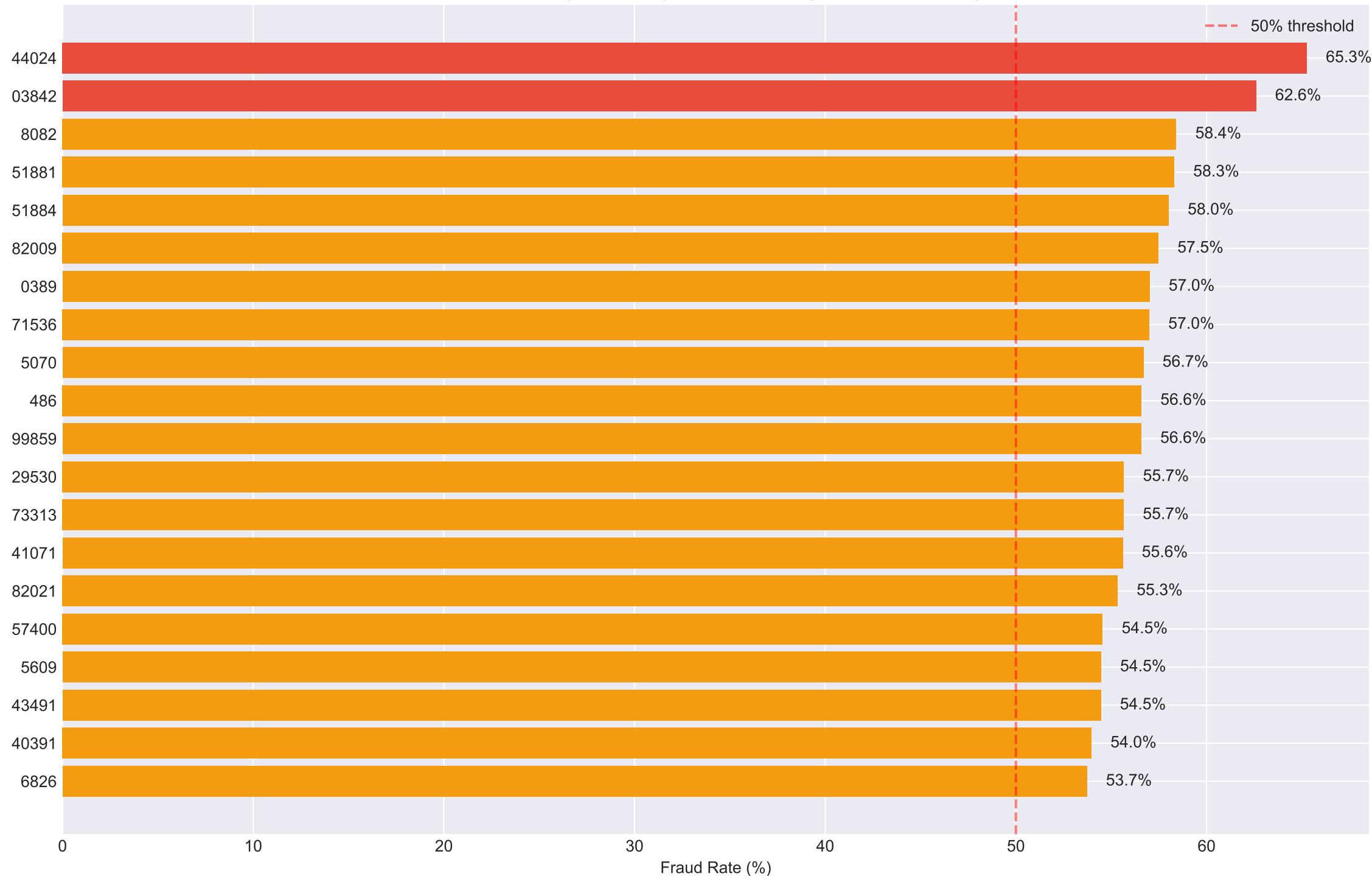
Top 15 Most Common Diagnoses



Top 15 Most Expensive Diagnoses

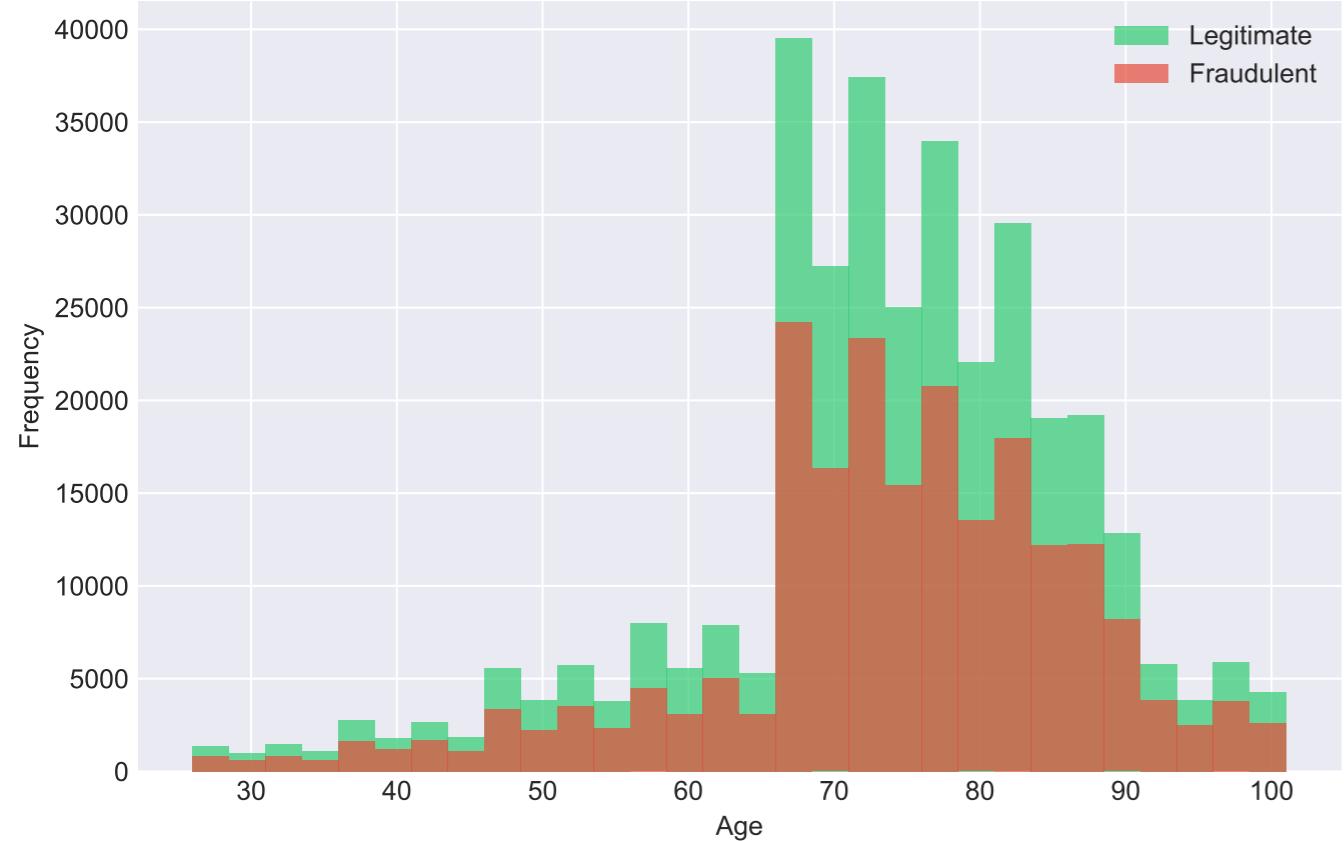


## Top 20 Diagnoses by Fraud Rate (min 100 claims)

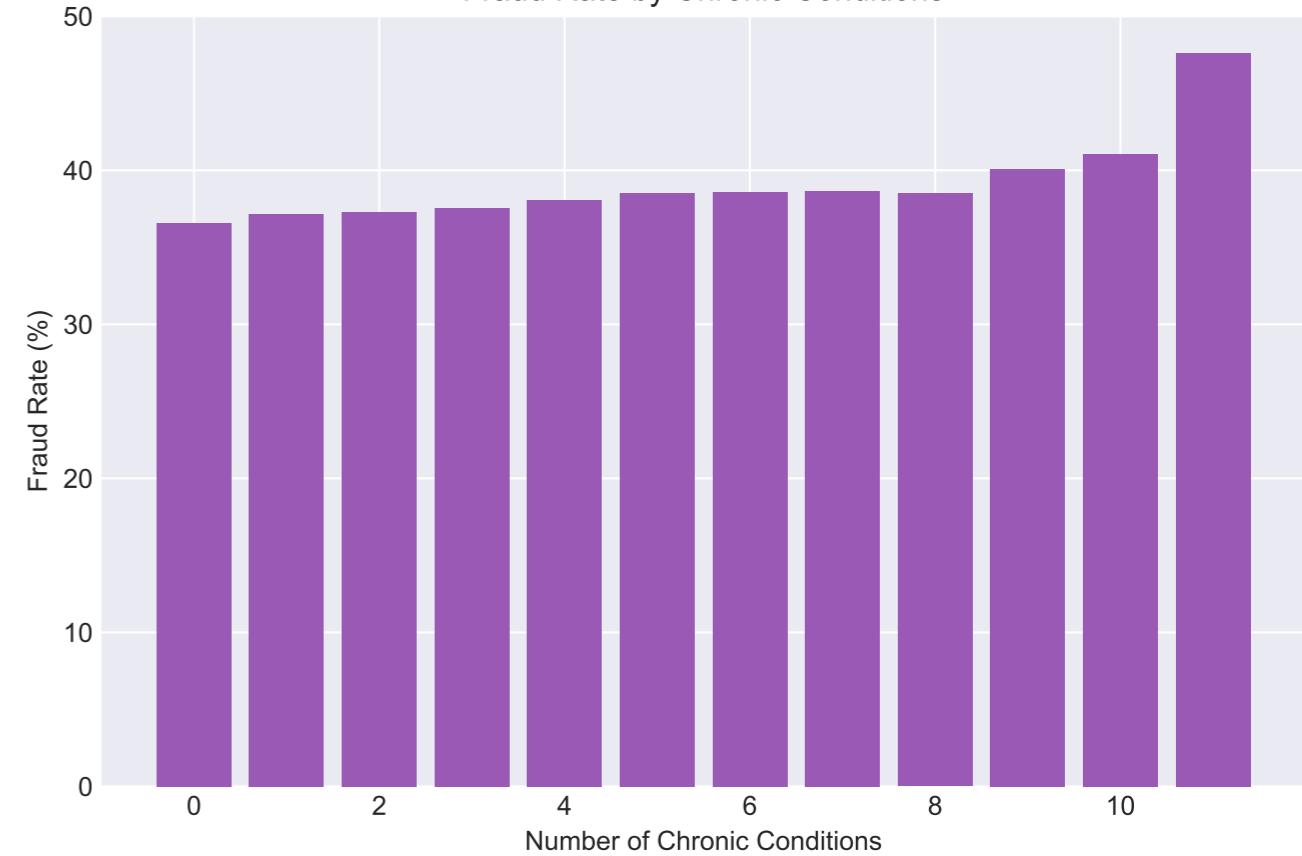


# Age and Chronic Conditions Analysis

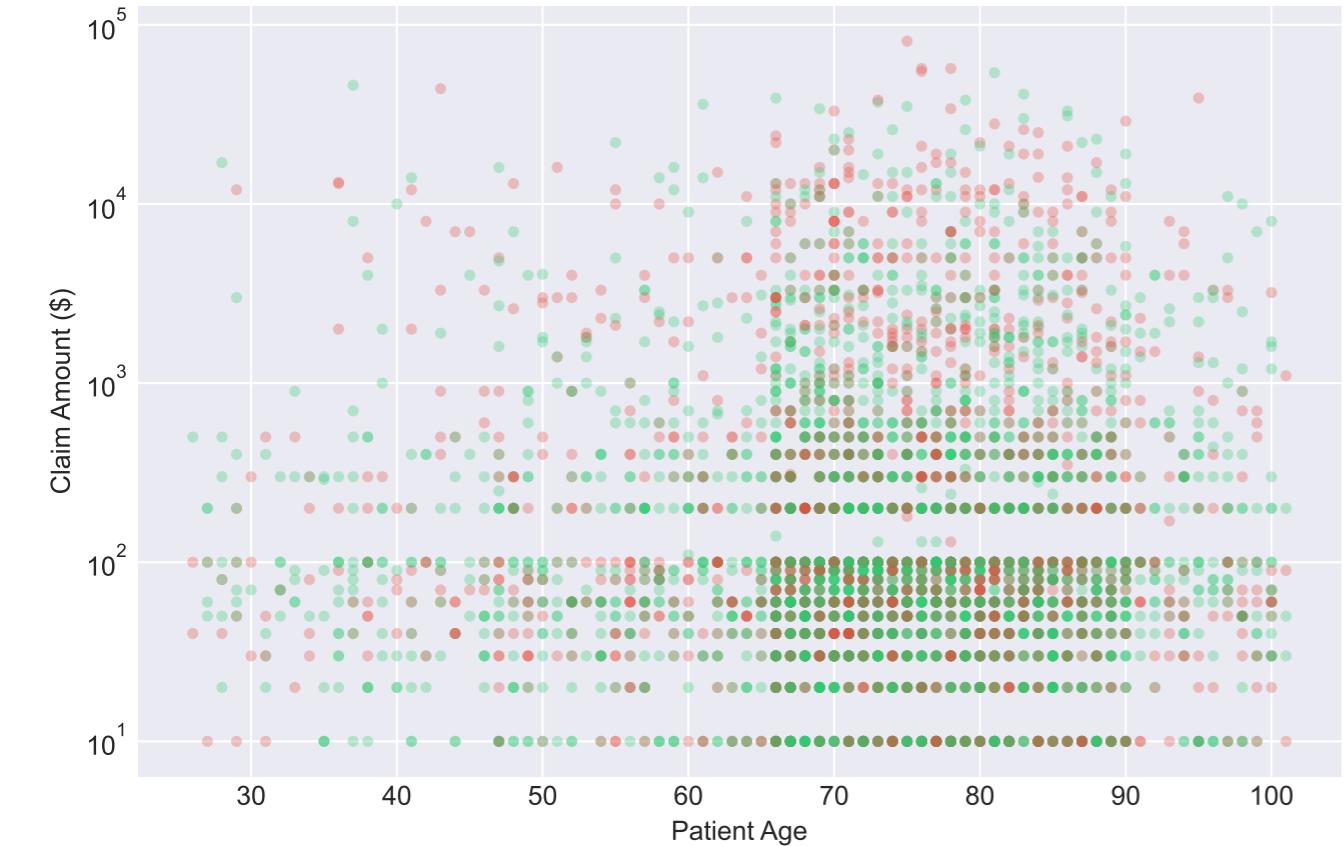
Age Distribution by Fraud Status



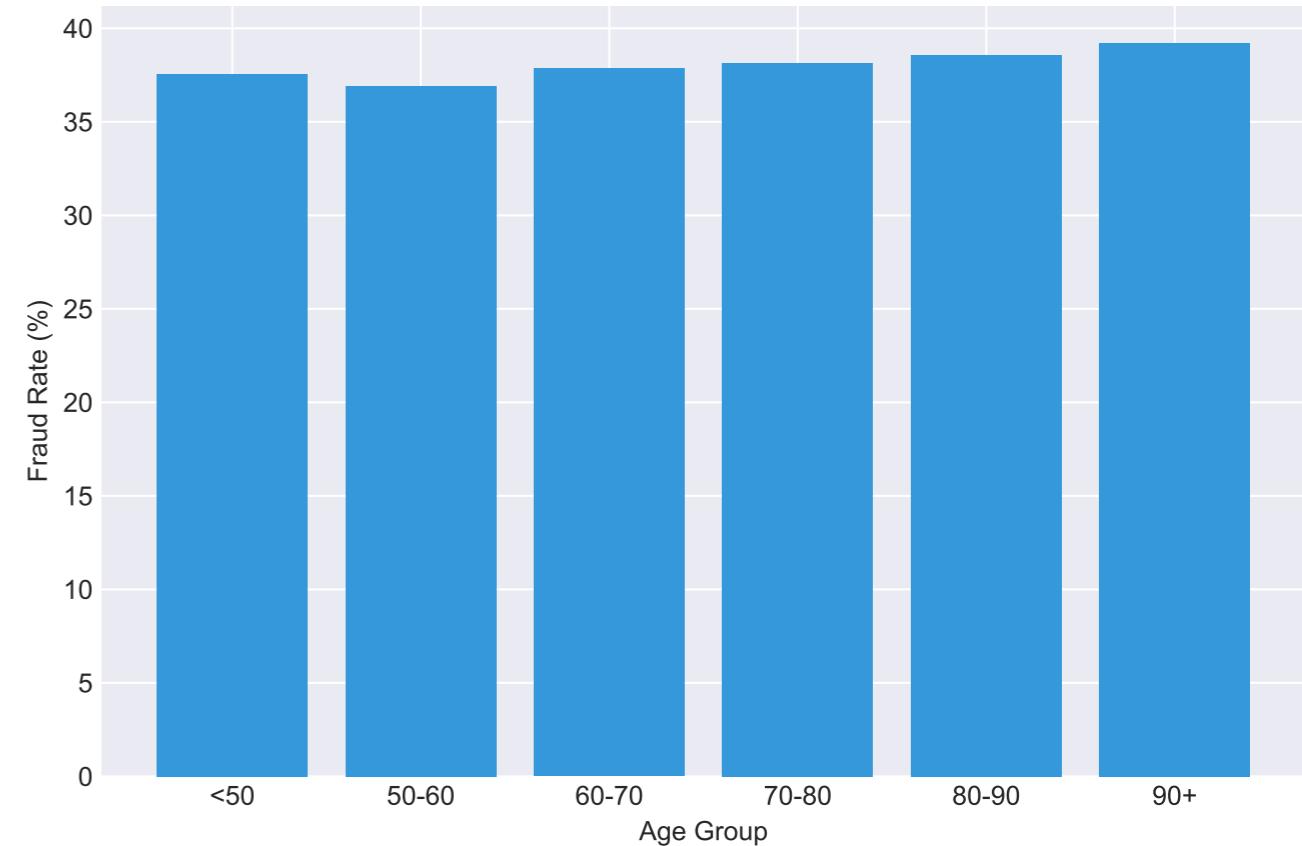
Fraud Rate by Chronic Conditions



Age vs Amount (Green=Legit, Red=Fraud)

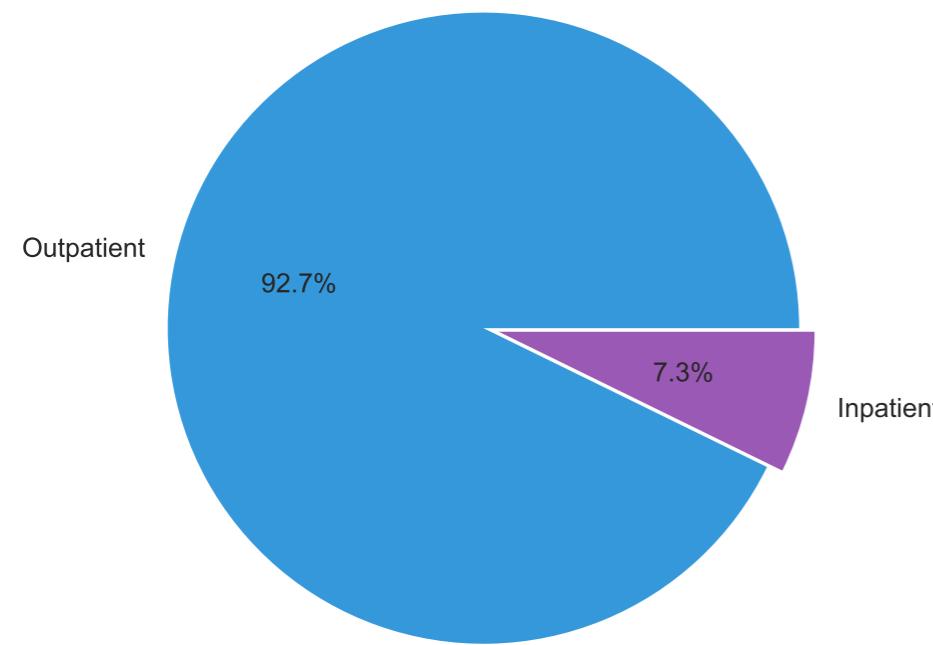


Fraud Rate by Age Group

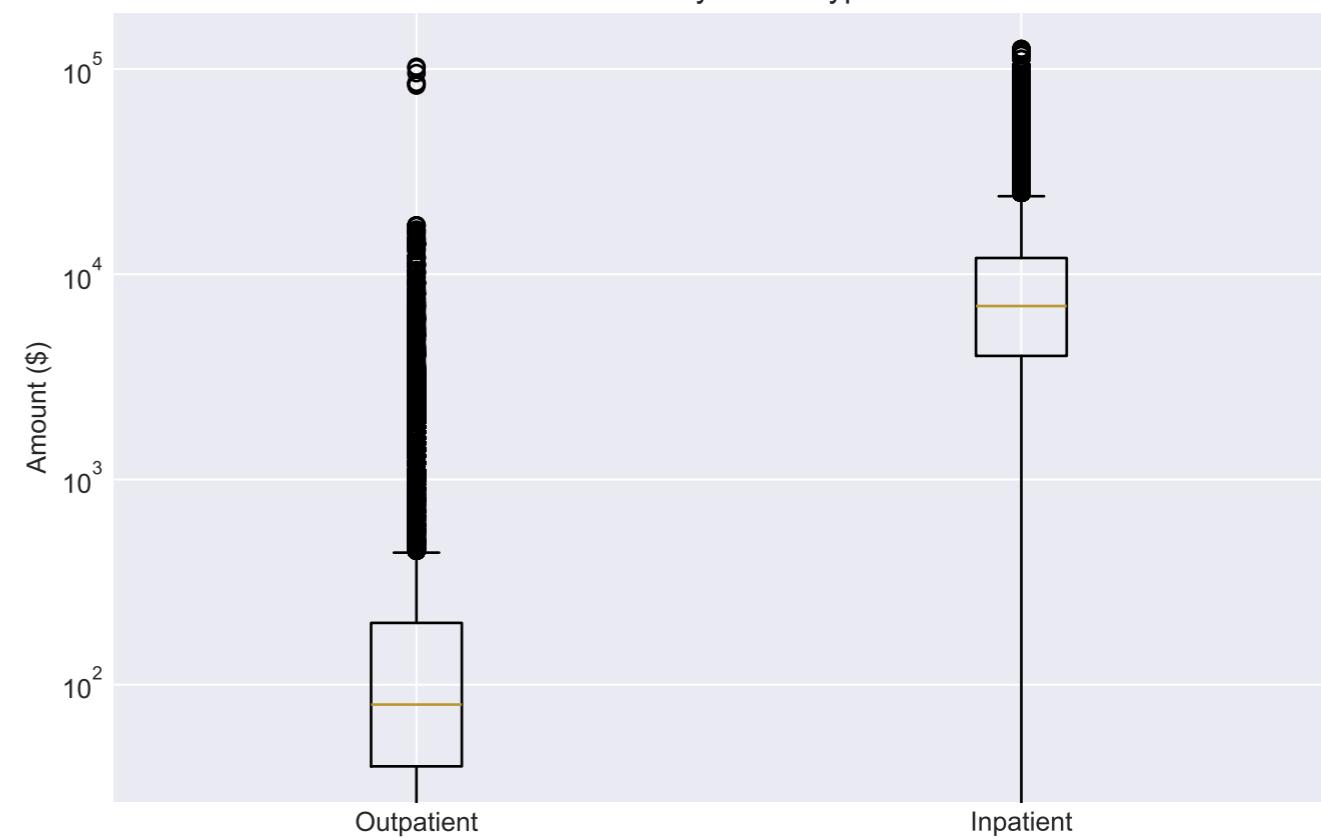


# Inpatient vs Outpatient Analysis

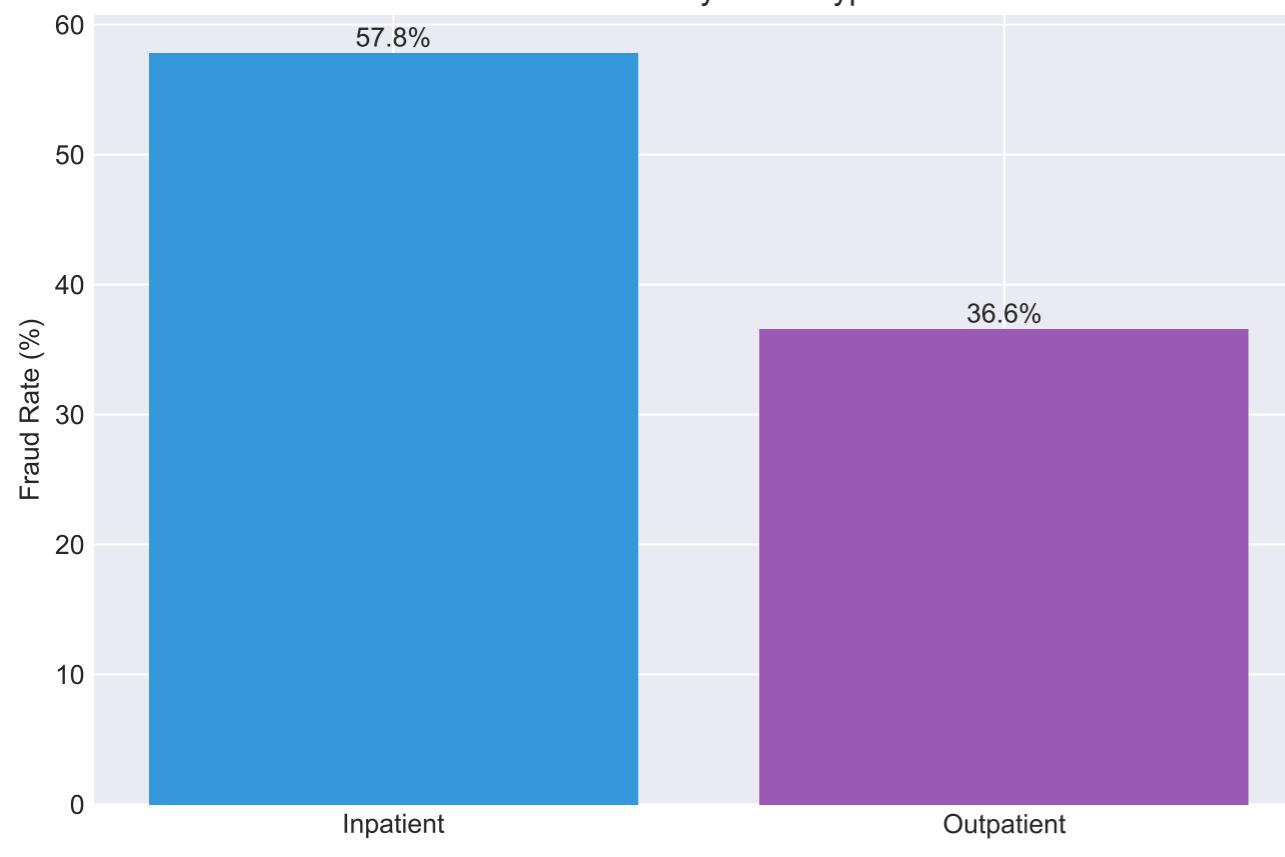
Claim Type Distribution



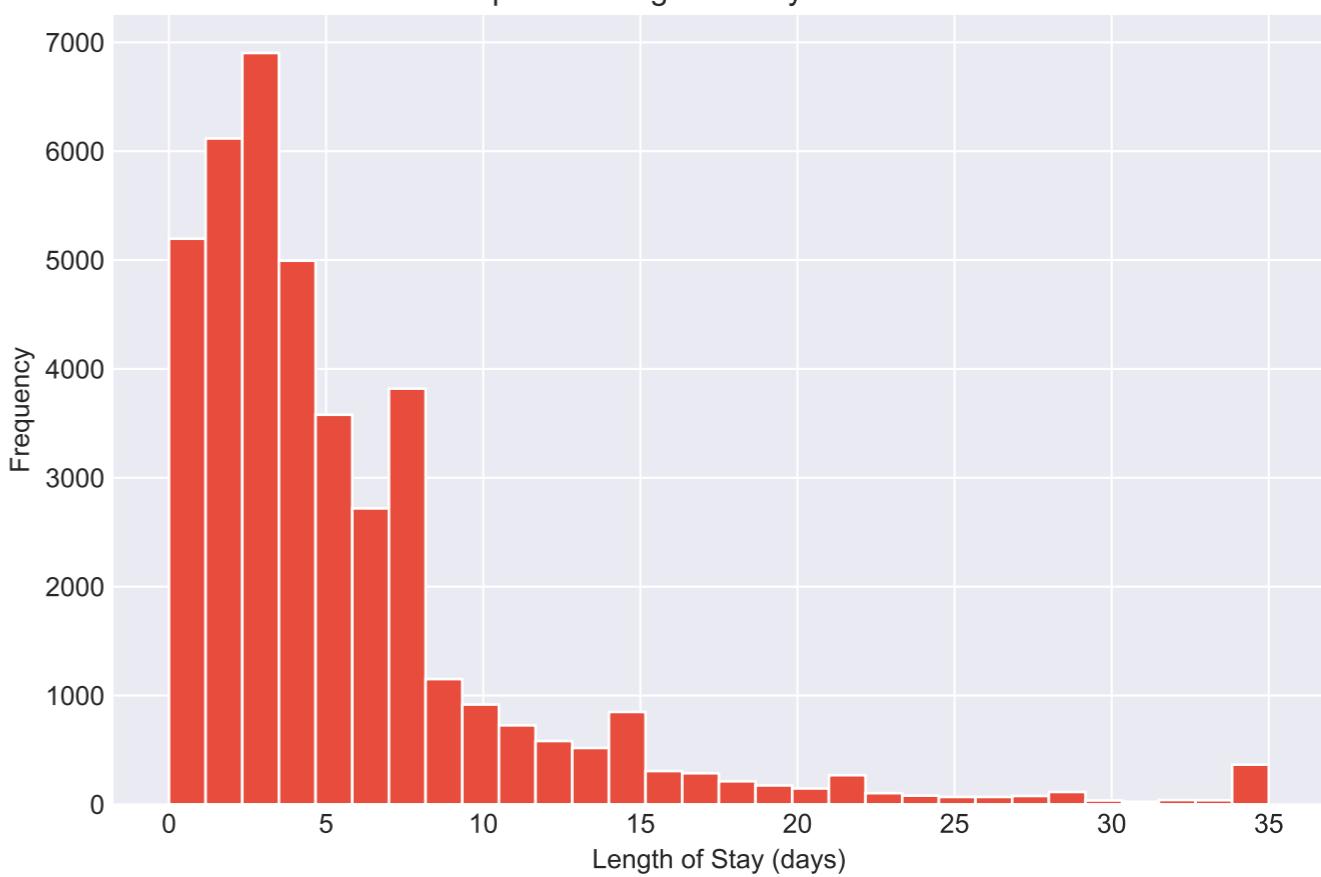
Amount by Claim Type



Fraud Rate by Claim Type



Inpatient Length of Stay Distribution



## KEY FINDINGS - HEALTHCARE FRAUD DETECTION EDA

---

### DATASET SUMMARY

---

- Total Claims Analyzed: 558,211
- Unique Providers: 5,410
- Unique Patients: 138,556
- Date Range: Medicare Claims Data

### FRAUD STATISTICS

---

- Overall Fraud Rate: 38.12%
- Fraudulent Claims: 212,796
- Legitimate Claims: 345,415

### FINANCIAL INSIGHTS

---

- Total Claims Value: \$556,543,140.00
- Average Claim Amount: \$997.01
- Median Claim Amount: \$80.00
- Highest Single Claim: \$125,000.00

### FRAUD PATTERNS IDENTIFIED

---

- Fraudulent claims have higher average amounts
- Certain diagnosis codes show 60%+ fraud rates
- Inpatient claims show different fraud patterns than outpatient
- Provider-level aggregation reveals clearer fraud signals

### TOP FRAUD INDICATORS

---

1. Diagnosis Code 44024 (Atherosclerosis w/ Gangrene): 65.2% fraud rate
2. Diagnosis Code 03842 (E. Coli Septicemia): 62.6% fraud rate
3. High claim amounts relative to provider type
4. Unusual patterns in claims per patient

### RECOMMENDATIONS

---

- Focus on provider-level analysis for fraud detection
- Flag high-risk diagnosis codes for manual review
- Implement amount thresholds by provider type
- Monitor claims per patient ratios