



VEHICLE INSURANCE CLAIM FRAUD DETECTION

Hannah Lal
November 2023





**INSURANCE FRAUD IS A PREVALENT AND COSTLY
CHALLENGE FOR BOTH POLICYHOLDERS AND
INSURANCE COMPANIES**

FRAUDULENT ACTIVITIES MAY ENCOMPASS:

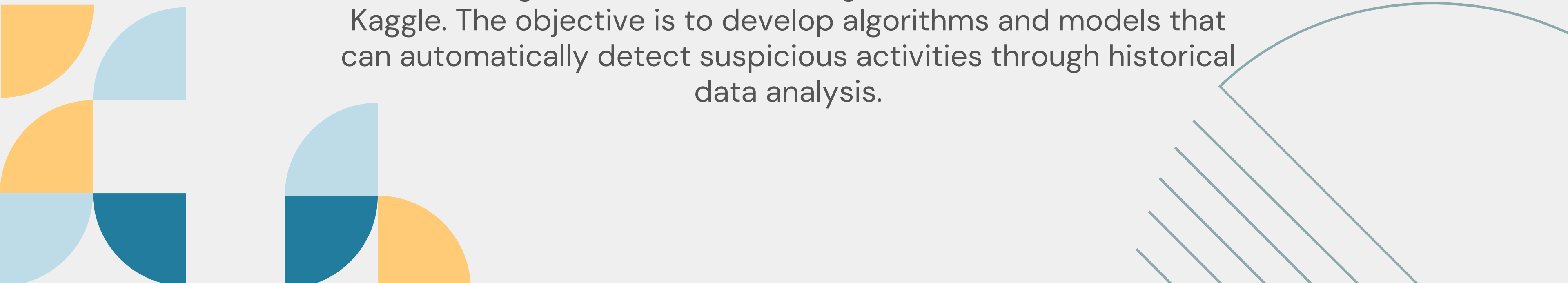
**FALSE CLAIMS
MISREPRESENTATION OF INFORMATION
ORGANIZED FRAUD SCHEMES**





PROJECT INTRODUCTION

Detecting and preventing these deceptive activities is crucial for minimizing financial losses for issuers as well as safeguarding policyholders. Given its significant impact, insurance fraud detection has emerged as a prominent research area in data science and machine learning. This capstone project aims to concentrate on identifying and preventing fraudulent or misleading insurance claims using the dataset available at Kaggle. The objective is to develop algorithms and models that can automatically detect suspicious activities through historical data analysis.



1 - DEFINE SUSPICIOUS

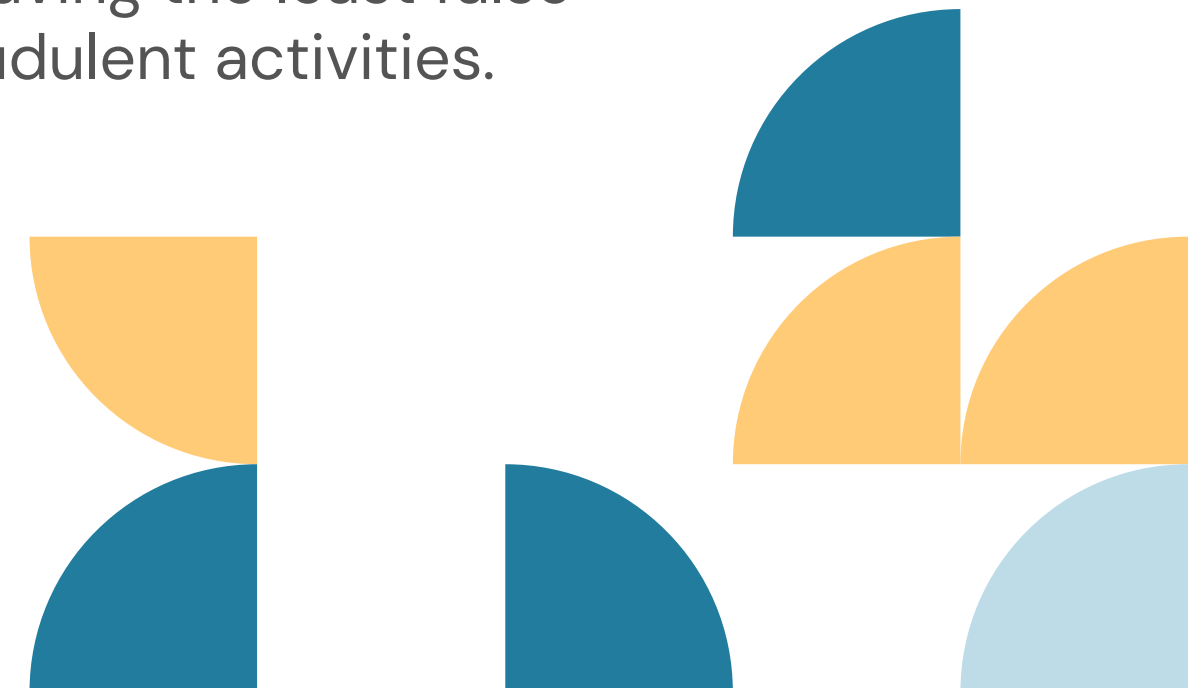
What is suspicious behavior in the world of insurance?

2 - IDENTIFY PATTERNS

Spot the patterns and anomalies in real world data that raises suspicion in insurance claims.

3 - PREVENT FRAUDS

Use machine learning methods, combined with some human monitored methods (mostly to prepare accurate data, having the least false positives) to prevent fraudulent activities.



BENEFITS OF FRAUD DETECTION

Stakeholders involved are

POLICYHOLDERS

POLICYHOLDERS STAND TO GAIN HEIGHTENED SECURITY AND PROTECTION AGAINST FRAUDULENT CLAIMS, FORTIFYING THEIR CONFIDENCE IN THE INSURANCE PROCESS.

INSURANCE COMPANIES

INSURANCE COMPANIES AND FINANCIAL INSTITUTIONS REAP THE BENEFITS OF DIMINISHED FINANCIAL LOSSES RESULTING FROM FRAUDULENT ACTIVITIES, THEREBY BOLSTERING CUSTOMER TRUST AND ELEVATING THEIR STANDING AND REPUTATION WITHIN THE INDUSTRY

FINANCIAL INSTITUTIONS

DATASET OVERVIEW

KAGGLE

DATE OF ACCIDENT (BROKEN INTO YEAR, MONTH, WEEK OF MONTH, DAY OF WEEK)		
DATE OF CLAIM (SAME AS ACCIDENT DATE)		AGE OF VEHICLE OWENER
MAKE OF CAR	AGE OF CAR	VEHICLE PRICE
ACCIDENT AREA	SEX	MARITAL STATUS
FAULT	POLICY TYPE (COLLISION, LIABILITY, ALL PERILS)	
VEHICLE CATEGORY (SEDAN, SPORT, UTILITY)		DEDUCTIBLE
DRIVARE RATING	AGE OF POLICY HOLDER	POLICE REPORT FILED
WITNESS PRESENT	NUMBER OF PAST CLAIMS	AGENT TYPE
NUMBER OF CARS	NUMBER OF SUPPLIMENTS	BASE POLICY
ADDRESS CHANGE CLAIM	DAYS POLICY CLAIM	DAYS POLICY ACCIDENT

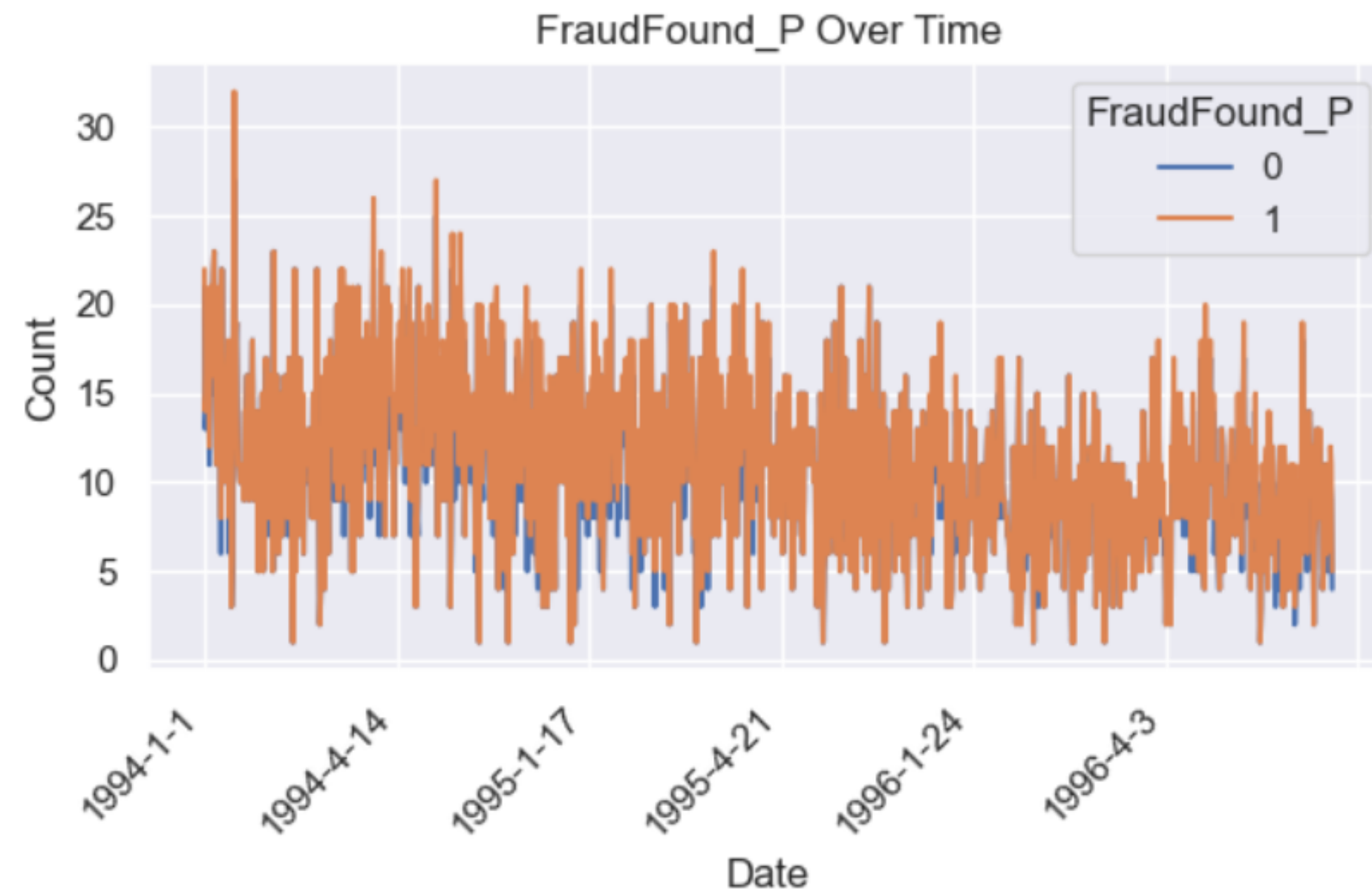
**FRAUDFOUND_P
(FAULT)**

FEATURE SET

TARGET VERIABLE

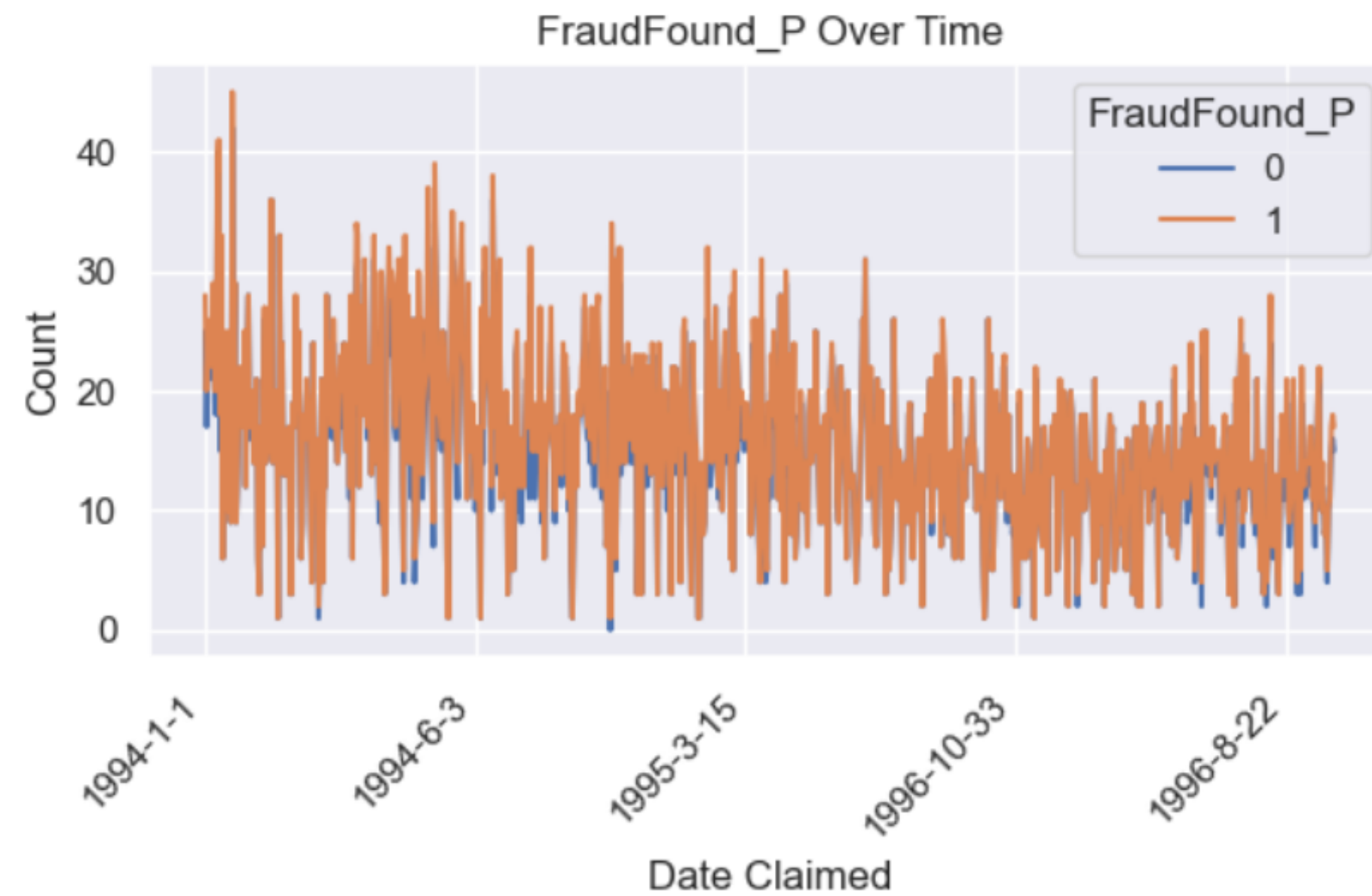
EDA

Progress of FraudFound_P over accident time



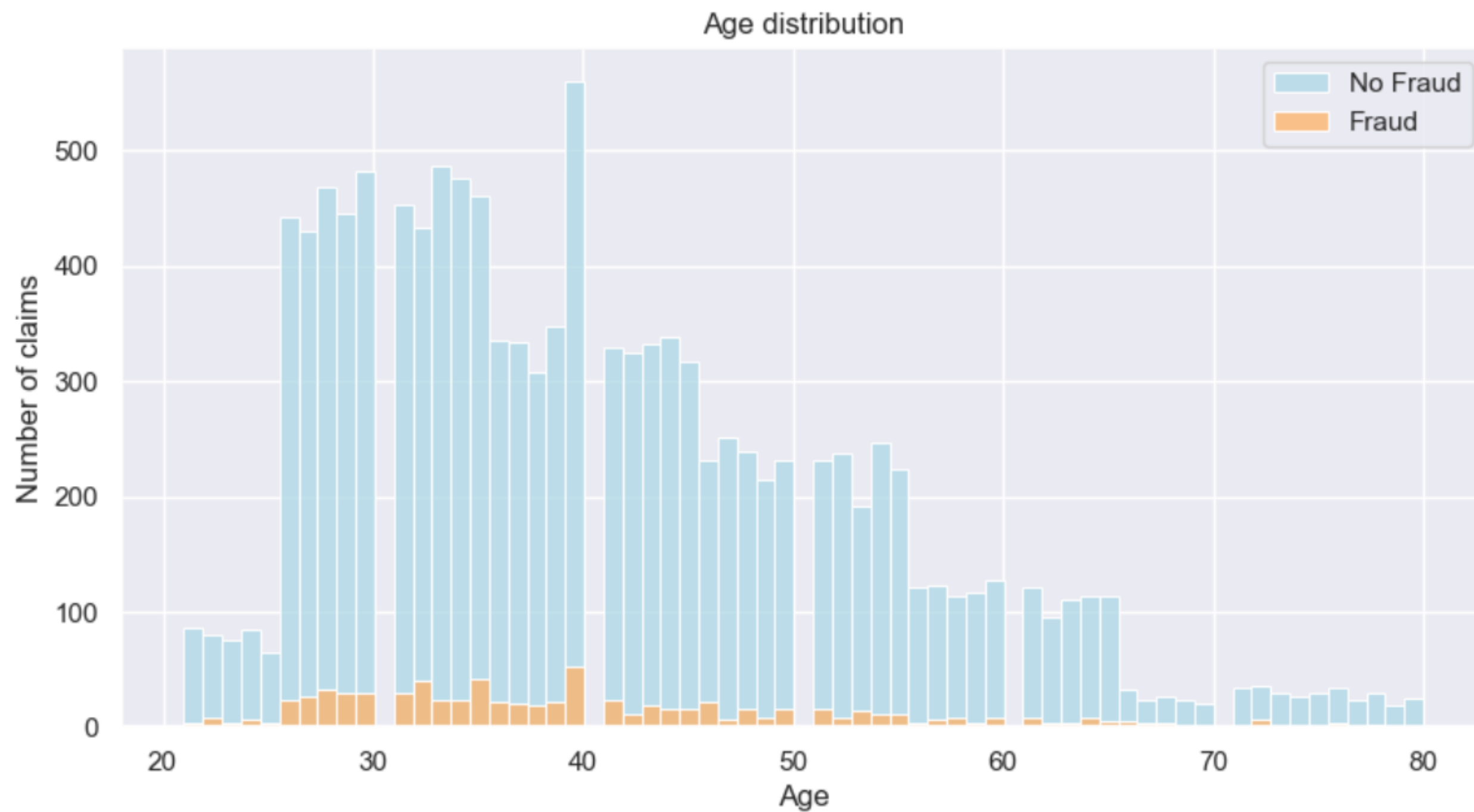
EDA

Progress of FraudFound_P over insurance claim time



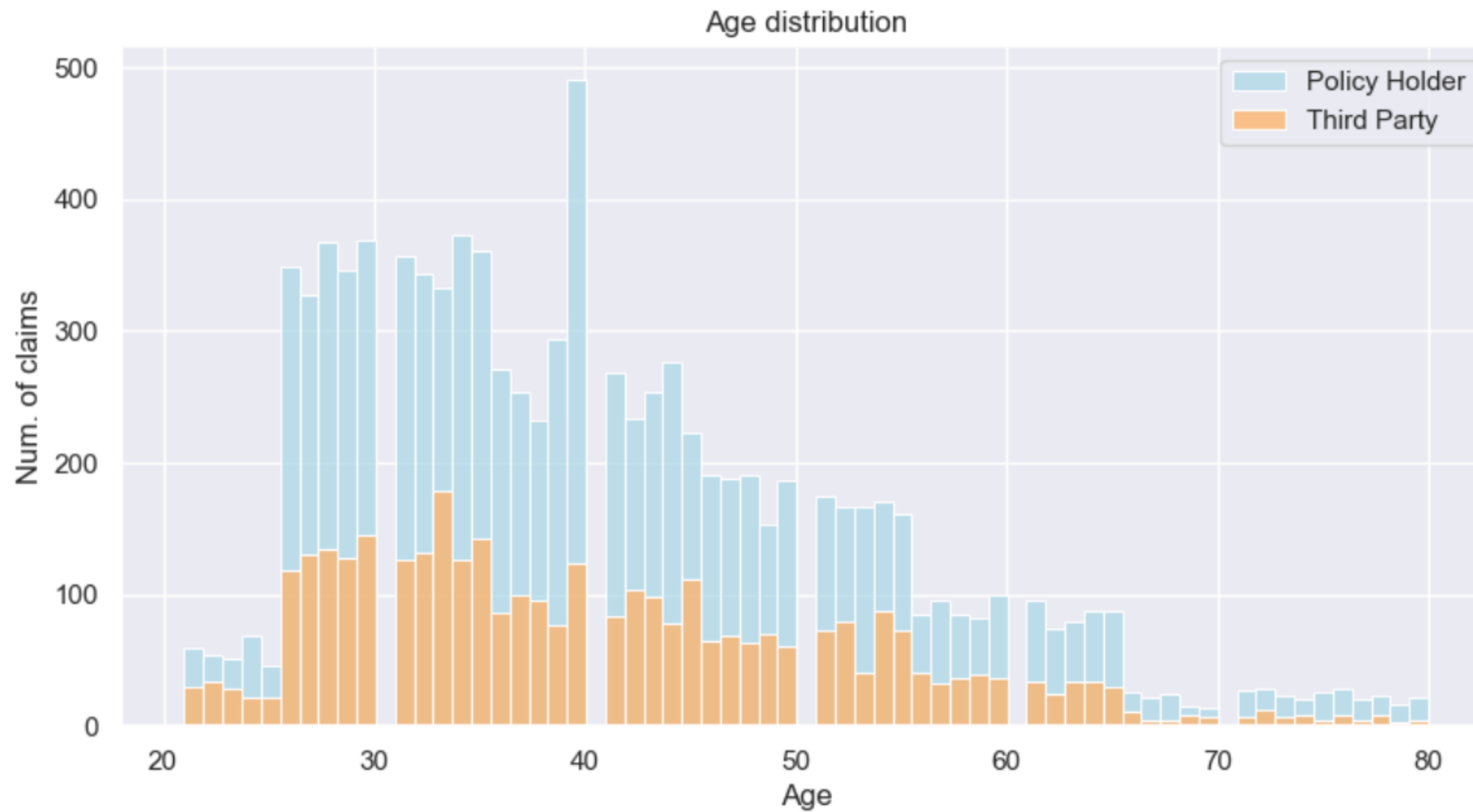
EDA

Relationship between Age and number of claims made for each category in our target variable (Fraud Found, and Fraud not found)



EDA

Relationship between Age and number of claims made for each category in 'Fault' variable (Policy holder, and third party)



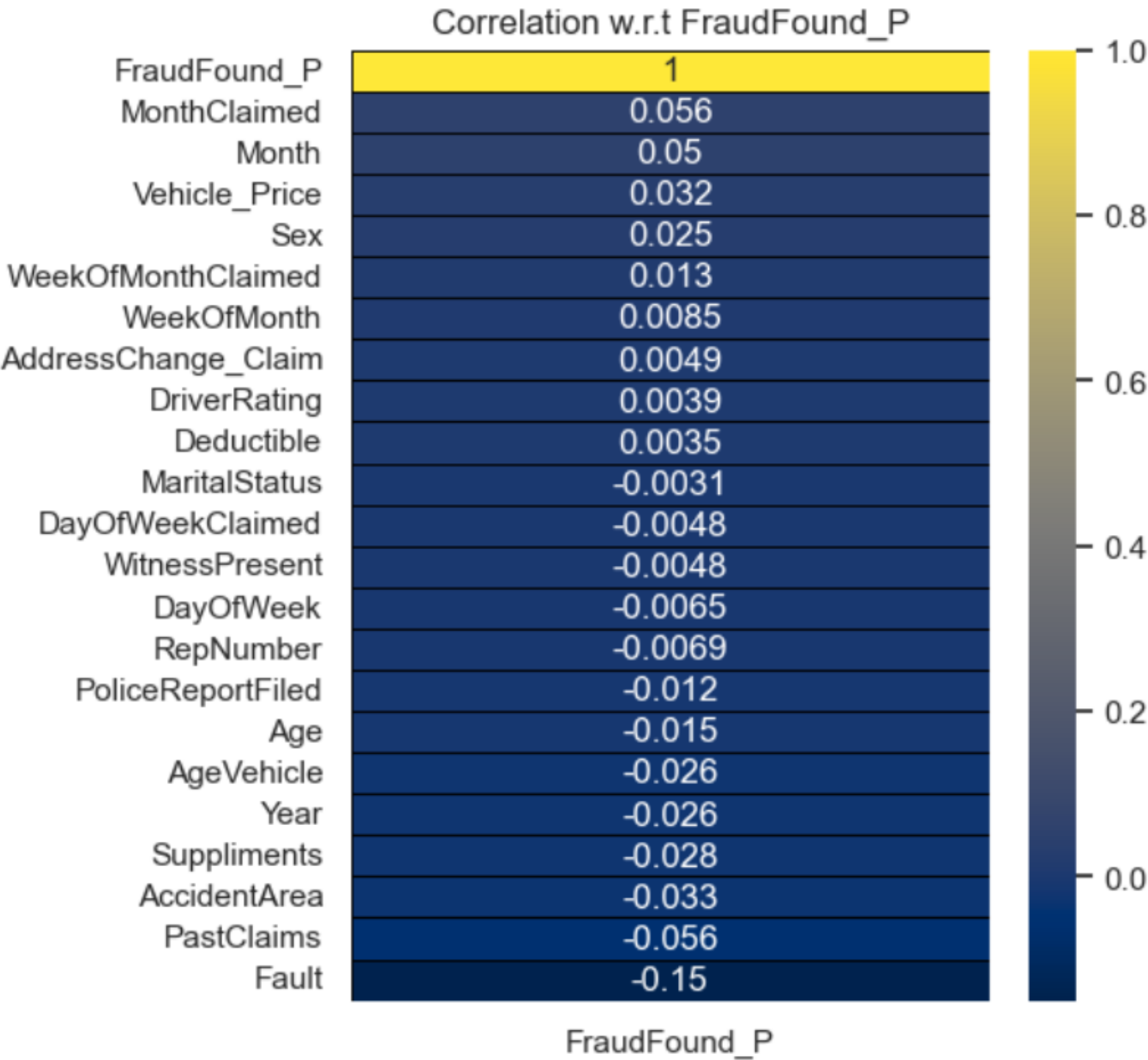
EDA

Correlation of feature set:

Month	1	-0.011	-0.00098	-0.011	-0.0041	0.76	0.015	0.014	0.01	0.0055	0.014	0.05	-0.0082	0.0071	-0.0093	-0.046	6.5e-05	-0.0035	-0.038	0.036	-0.02	-0.021
WeekOfMonth	-0.011	1	-0.012	-0.0043	0.011	-0.0015	0.22	0.0018	-0.014	0.0016	-0.019	0.0085	-0.0098	0.00042	0.0098	-0.0026	-0.0014	-0.0019	0.0065	-0.0049	0.004	-0.0017
DayOfWeek	-0.00098	-0.012	1	0.022	0.048	-0.0044	0.0038	0.014	-0.0044	-0.01	-0.014	-0.0065	0.0058	-0.012	-0.0023	-0.003	-0.0019	0.0055	-0.0049	0.033	-0.0036	0.0092
AccidentArea	-0.011	-0.0043	0.022	1	-0.0055	-0.01	0.011	-0.031	0.0014	0.0028	0.012	-0.033	0.0065	0.0018	-0.0025	-0.0025	-0.026	0.016	-0.0022	-0.0048	0.015	-0.011
WeekClaimed	-0.0041	0.011	0.048	-0.0055	1	0.016	-0.0086	0.0035	0.0083	-0.0043	-0.0083	-0.0048	-0.0092	-0.0042	-0.00011	0.011	0.01	-0.0029	-0.0064	-0.01	-0.0081	0.0032
MonthClaimed	0.76	-0.0015	-0.0044	-0.01	0.016	1	0.0064	0.015	0.0072	0.0027	0.0098	0.056	-0.002	0.011	-0.004	-0.052	-0.0019	-0.0068	-0.037	0.04	-0.027	-0.021
WeekClaimed	0.015	0.22	0.0038	0.011	-0.0086	0.0064	1	-0.0022	-0.0024	0.0079	0.0031	0.013	-0.017	-0.0054	-0.0033	-0.028	-0.0089	-0.0074	-0.0058	-0.0042	-0.0022	-0.022
Sex	0.014	0.0018	0.014	-0.031	0.0035	0.015	-0.0022	1	-0.12	0.11	0.0042	0.025	0.0032	0.019	-0.0006	-0.0067	0.0021	0.00098	0.0031	0.12	0.14	0.014
MaritalStatus	0.01	-0.014	-0.0044	0.0014	0.0083	0.0072	-0.0024	-0.12	1	-0.38	-0.014	-0.0031	-0.0061	-0.032	-0.017	0.012	0.009	-0.005	-0.016	0.03	-0.44	-0.034
Age	0.0055	0.0016	-0.01	0.0028	-0.0043	0.0027	0.0079	0.11	-0.38	1	-0.0085	-0.015	-0.0053	0.058	0.0052	-0.011	-0.0083	0.011	0.015	0.048	0.58	0.04
Fault	0.014	-0.019	-0.014	0.012	-0.0083	0.0098	0.0031	0.0042	-0.014	-0.0085	1	-0.15	-0.0053	-0.0051	-0.0061	0.026	0.058	0.0073	0.016	-0.039	0.019	0.01
FraudFound_P	0.05	0.0085	-0.0065	-0.033	-0.0048	0.056	0.013	0.025	-0.0031	-0.015	-0.15	1	-0.0069	0.0035	0.0039	-0.012	-0.0048	0.0049	-0.026	0.032	-0.026	-0.028
RepNumber	-0.0082	-0.0098	0.0058	0.0065	-0.0092	-0.002	-0.017	0.0032	-0.0061	-0.0053	-0.0053	-0.0069	1	-0.00083	0.015	0.011	0.0081	0.0058	0.0079	0.011	-0.004	-7.4e-05
Deductible	0.0071	0.00042	-0.012	0.0018	-0.0042	0.011	-0.0054	0.019	-0.032	0.058	-0.0051	0.0035	-0.00083	1	-0.0017	0.015	-0.0011	0.0028	-0.003	0.0038	0.095	-0.0011
DriverRating	-0.0093	0.0098	-0.0023	-0.0025	-0.00011	-0.004	-0.0033	-0.0006	-0.017	0.0052	-0.0061	0.0039	0.015	-0.0017	1	0.016	0.0094	-0.0067	-0.013	-0.0049	0.0002	0.003
ReportFiled	-0.046	-0.0026	-0.003	-0.0025	0.011	-0.052	-0.028	-0.0067	0.012	-0.011	0.026	-0.012	0.011	0.015	0.016	1	0.19	0.0021	0.015	0.014	-0.0035	0.013
WitnessPresent	6.5e-05	-0.0014	-0.0019	-0.026	0.01	-0.0019	-0.0089	0.0021	0.009	-0.0083	0.058	-0.0048	0.0081	-0.0011	0.0094	0.19	1	0.00059	-0.019	0.015	-0.0084	-0.012
AddressChange_Claim	-0.0035	-0.0019	0.0055	0.016	-0.0029	-0.0068	-0.0074	0.00098	-0.005	0.011	0.0073	0.0049	0.0058	0.0028	-0.0067	0.0021	0.00059	1	-0.0026	0.00031	0.0075	0.015
Year	-0.038	0.0065	-0.0049	-0.0022	-0.0064	-0.037	-0.0058	0.0031	-0.016	0.015	0.016	-0.026	0.0079	-0.003	-0.013	0.015	-0.019	-0.0026	1	-0.03	0.02	0.017
Vehicle_Price	0.036	-0.0049	0.033	-0.0048	-0.01	0.04	-0.0042	0.12	0.03	0.048	-0.039	0.032	0.011	0.0038	-0.0049	0.014	0.015	0.00031	-0.03	1	-0.075	0.0079
AgeVehicle	-0.02	0.004	-0.0036	0.015	-0.0081	-0.027	-0.0022	0.14	-0.44	0.58	0.019	-0.026	-0.004	0.095	0.0002	-0.0035	-0.0084	0.0075	0.02	-0.075	1	0.16
Supplements	-0.021	-0.0017	0.0092	-0.011	0.0032	-0.021	-0.022	0.014	-0.034	0.04	0.01	-0.028	-7.4e-05	-0.0011	0.003	0.013	-0.012	0.015	0.017	0.0079	0.16	1
PastClaims	0.02	0.006	0.034	0.062	-0.0025	0.0099	-0.014	-0.0053	0.022	-0.047	-0.13	-0.056	-0.00027	-0.0028	-0.0046	0.0018	-0.0094	0.018	0.014	0.061	-0.018	0.1
	Month	WeekOfMonth	DayOfWeek	AccidentArea	DayOfWeekClaimed	MonthClaimed	WeekOfMonthClaimed	Sex	MaritalStatus	Age	Fault	FraudFound_P	RepNumber	Deductible	DriverRating	PoliceReportFiled	WitnessPresent	AddressChange_Claim	Year	Vehicle_Price	AgeVehicle	Supplements

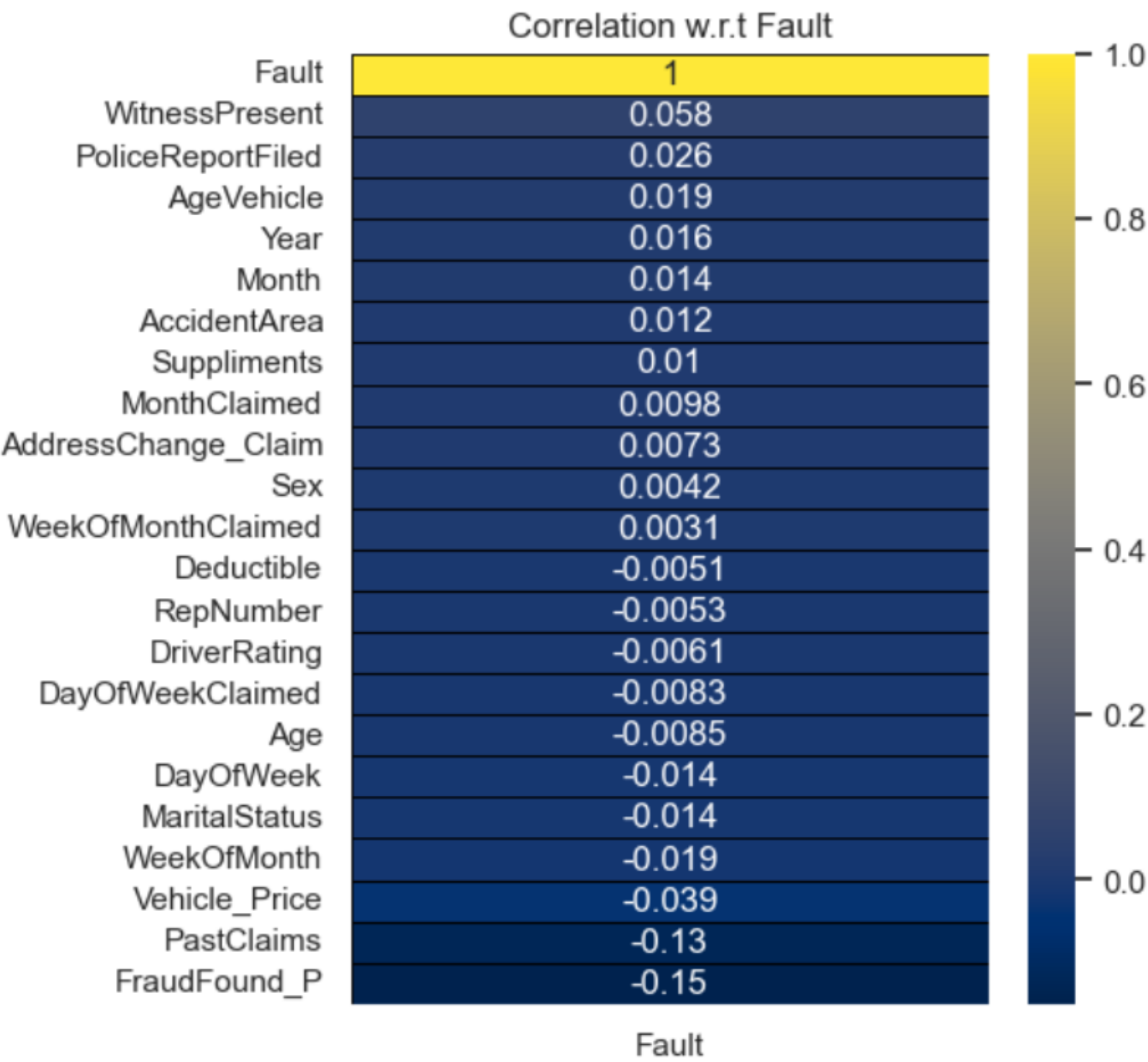
EDA

Correlation of feature set with respect to the target variable FraudFound_P:



EDA

Correlation of feature set with respect to Fault Variable:

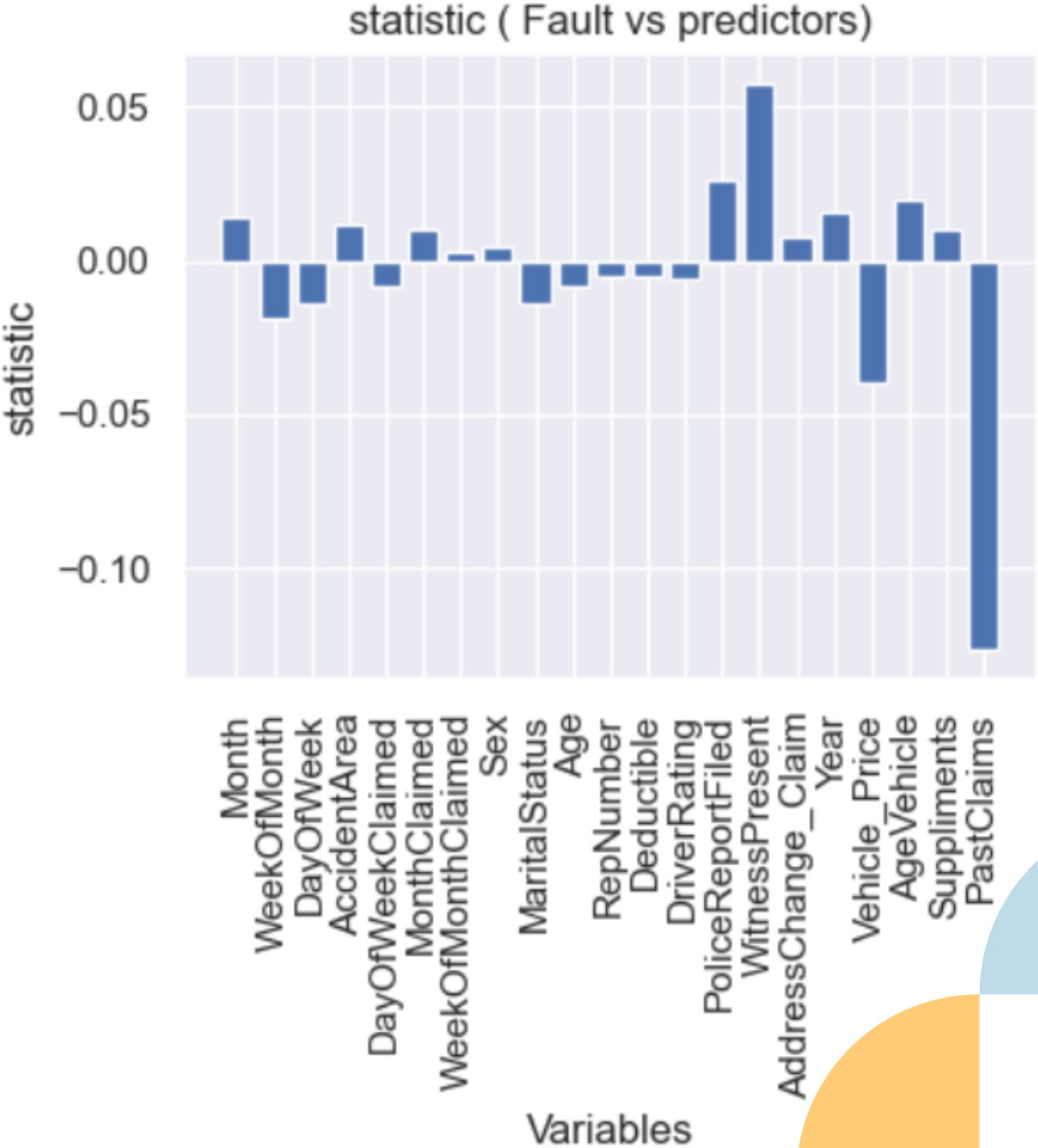
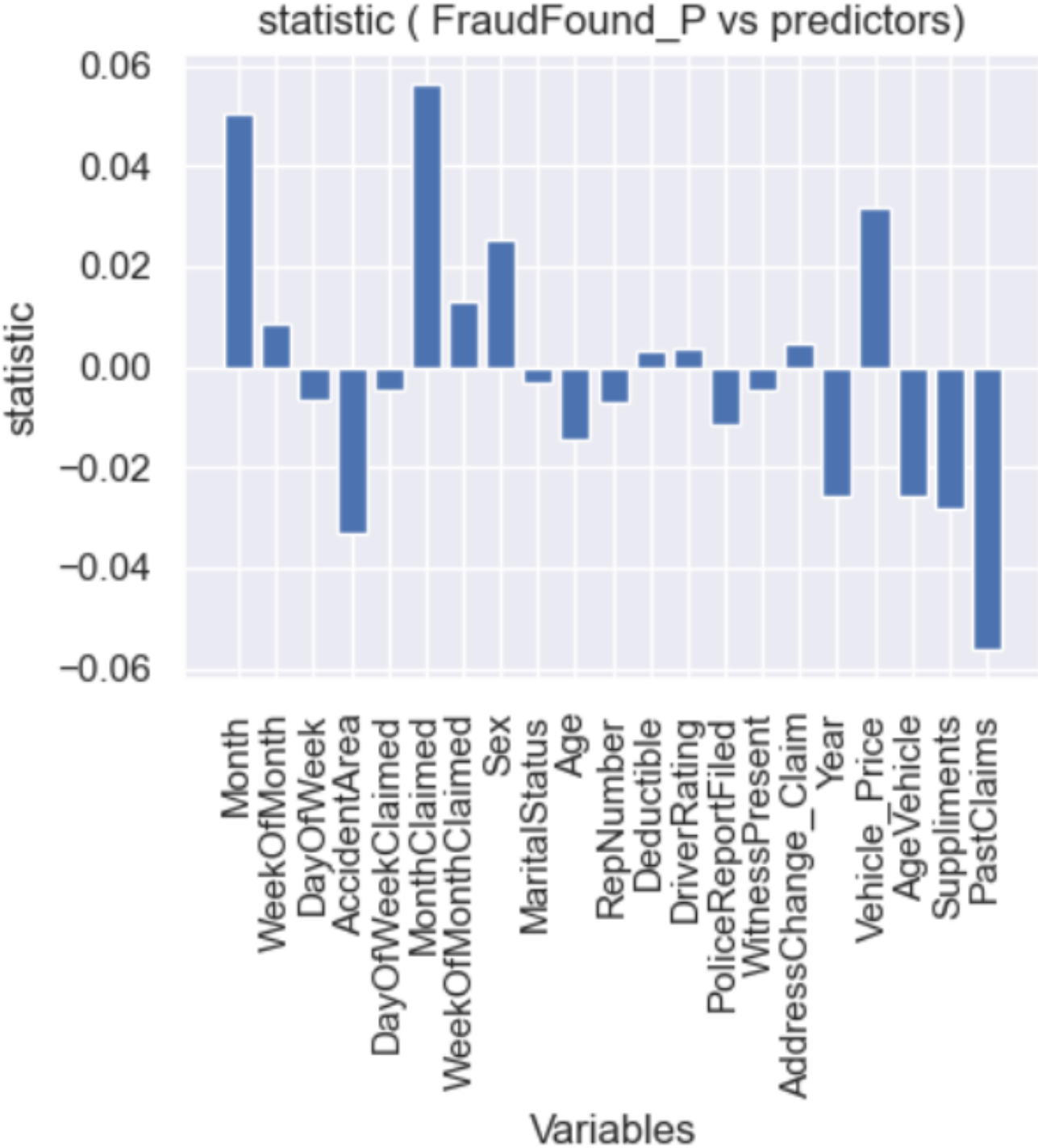


T VALUES AND P VALUES

PREDICTORS WITH
HIGHER
CORRELATION

MONTH
MONTH CLAIMED
YEAR
PAST CLAIMS
(FOR FRAUDFOUND_P)

WITNESS PRESENT
VEHICLE PRICE
PAST CLAIMS
(FOR FAULT)



SUMMARY REPORT

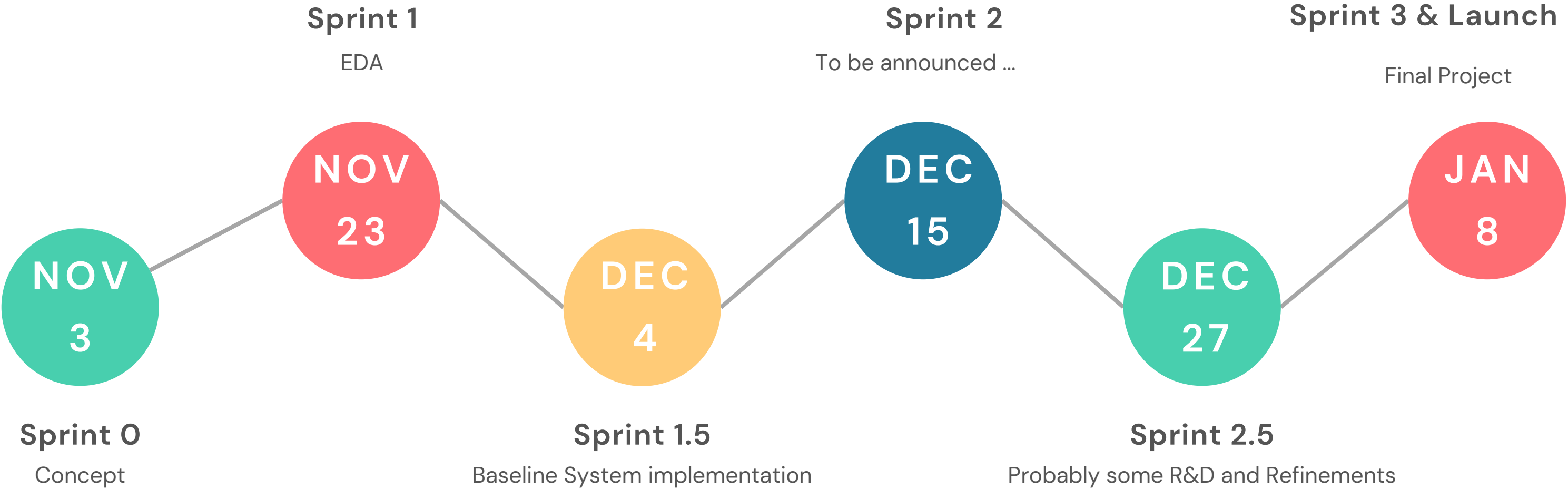
FraudFound_P:

- Month, WeekOfMonth, MonthClaimed, WeekOfMonthClaimed, Sex, Dedictible, DriverRating, AddressChange_Claim and Vehicle_Price are positively correlated with our target variable FraudFound_P.
- DayOfWeek, AccidentArea, DayOfWeekClaimed, MaritalStatud, Age, RepNumber, PoliceReportFiled, WitnessPresent, Year, Age of Vehicle and Number of suppliments are negatively correlated with our target variable FraudFound_P.]

Fault:

- Month, AccidentArea, MonthClaimed, WeekOfMonthClaimed, Sex, PoliceReportFiled, WitnessPresent, AddressChange_Claim, Year, Age of Vehicle and Number of suppliments are positively correlated with Fault variable.
- WeekOfMonth, DayOfWeek, DayOfWeekClaimed, MaritalStatud, Age, RepNumber, Dedictible, DriverRating, Vehicle_Price and Past Number of Claims are negatively correlated with number of t_registered rentals.

PROJECT TIMELINE





THANK YOU

Questions?