

WHAT ARE COMMON TRAITS OF
CAR INSURANCE FRAUD?

INTRODUCTION

- According to Insurance Information Institute¹: Auto insurers in the US lose at least \$29 billion a year to Insurance Fraud
- Multiple studies have researched the applicability of statistical methods to meaningfully predict frauds before they occur
- Using data analysis and visualizations, we try to analyze relationships between several claim-related factors and their ability to predict frauds based on large dataset



THE DATA

- Insurance Fraud incidents from accident data taken from Kaggle
- The data had 15,420 entries with 33 unique attributes
- The accident data took place from Jan 1994 – Dec 1996



FACTORS UNDER CONSIDERATION

All Available Factors

```
fraud_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15420 entries, 0 to 15419
Data columns (total 33 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Month                  15420 non-null  object
1   WeekOfMonth            15420 non-null  int64
2   DayOfWeek              15420 non-null  object
3   Make                   15420 non-null  object
4   AccidentArea           15420 non-null  object
5   DayOfWeekClaimed       15420 non-null  object
6   MonthClaimed           15420 non-null  object
7   WeekOfMonthClaimed     15420 non-null  int64
8   Sex                    15420 non-null  object
9   MaritalStatus          15420 non-null  object
10  Age                    15420 non-null  int64
11  Fault                  15420 non-null  object
12  PolicyType             15420 non-null  object
13  VehicleCategory        15420 non-null  object
14  VehiclePrice           15420 non-null  object
15  FraudFound_P           15420 non-null  int64
16  PolicyNumber           15420 non-null  int64
17  RepNumber              15420 non-null  int64
18  Deductible             15420 non-null  int64
19  DriverRating           15420 non-null  int64
20  Days_Policy_Accident   15365 non-null  object
21  Days_Policy_Claim      15419 non-null  object
22  PastNumberOfClaims     11068 non-null  object
23  AgeOfVehicle           15420 non-null  object
24  AgeOfPolicyHolder       15420 non-null  object
25  PoliceReportFiled      15420 non-null  object
26  WitnessPresent         15420 non-null  object
27  AgentType              15420 non-null  object
28  NumberOfSupplements    8373 non-null   object
29  AddressChange_Claim    15420 non-null  object
30  NumberOfCars           15420 non-null  object
31  Year                   15420 non-null  int64
32  BasePolicy             15420 non-null  object
dtypes: int64(9), object(24)
```



We filter down the available list of factors based on literature review. Factors which have been good predictors of fraud in past research are retained for the analysis

Factors used in the Analysis

```
fraud_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15420 entries, 0 to 15419
Data columns (total 28 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Month                  15420 non-null  object
1   WeekOfMonth            15420 non-null  int64
2   DayOfWeek              15420 non-null  object
3   Make                   15420 non-null  object
4   AccidentArea           15420 non-null  object
5   DayOfWeekClaimed       15420 non-null  object
6   MonthClaimed           15420 non-null  object
7   WeekOfMonthClaimed     15420 non-null  int64
8   Sex                    15420 non-null  object
9   MaritalStatus          15420 non-null  object
10  Age                    15420 non-null  int64
11  Fault                  15420 non-null  object
12  PolicyType             15420 non-null  object
13  VehicleCategory        15420 non-null  object
14  VehiclePrice           15420 non-null  object
15  FraudFound_P           15420 non-null  int64
16  Deductible             15420 non-null  int64
17  DriverRating           15420 non-null  int64
18  Days_Policy_Accident   15420 non-null  object
19  Days_Policy_Claim      15420 non-null  object
20  PastNumberOfClaims     15420 non-null  object
21  AgeOfVehicle           15420 non-null  object
22  AgeOfPolicyHolder       15420 non-null  object
23  PoliceReportFiled      15420 non-null  object
24  WitnessPresent         15420 non-null  object
25  NumberOfCars           15420 non-null  object
26  Year                   15420 non-null  int64
27  BasePolicy             15420 non-null  object
dtypes: int64(7), object(21)
```

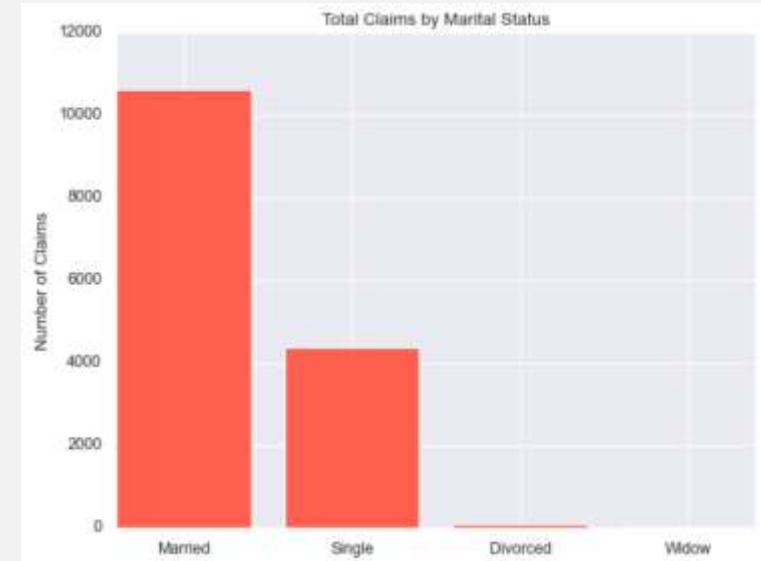
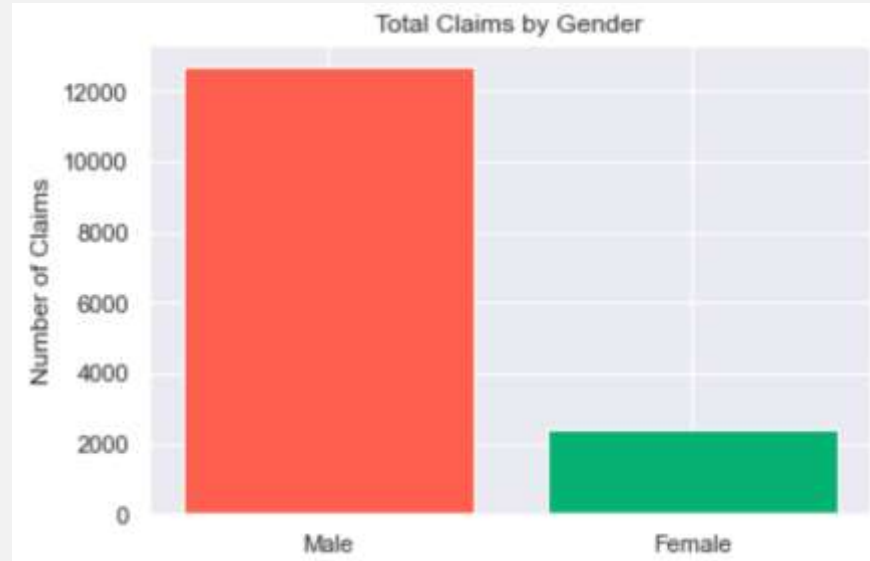
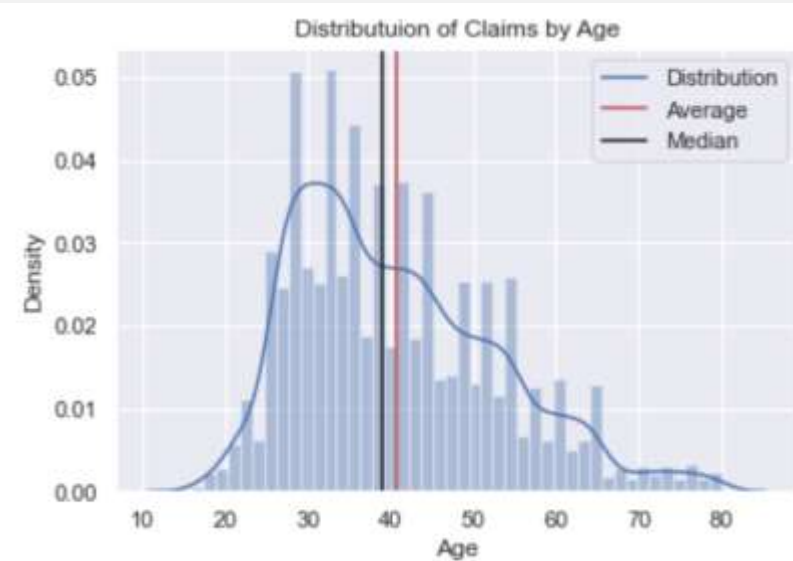
DATA CLEANING

- Within the dataset, a person's age appeared to be 0
- This was interpreted to be an N/A value and thus removed from the data

	# of Observations	# of Observations with Age = 0
Before Cleaning Age	15,420	320
Removing Obs. With Age = 0	15,100	0

- The column "NumberOfSupplements" had NaN values, but this attribute was not used in the analysis
- After cleaning, only 320 entries were removed (2.07% of the total data)
- There were no duplicate entries or any N/A entries

GENERAL CHARACTERISTICS OF THE DATA



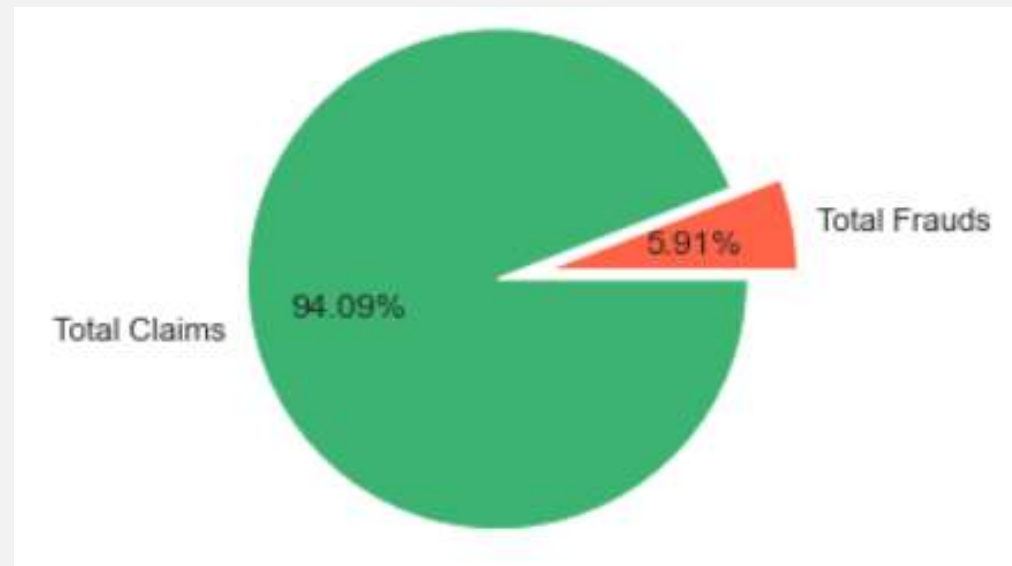
- Data shows that most claims on auto insurance are made by adults over the age of 30
- Average age of insurance claimers is ~ 40
- Historically, males and married sections of the demographic have made most of the claims

QUESTIONS TO CONSIDER

- Who is more likely to commit fraud, males or females?
- What make of vehicle had the highest amount of fraud cases?
- What age and age range has the highest proportion for fraud?
- What type of relationship status (Single, Married, Divorced, Widow) has a higher proportion for fraud?
- What policy type is most prevalent in fraud claims?
- Where do fraudulent incidents occur?
- How old are the vehicles involved in fraud claims?
- How expensive are the vehicles in fraud cases?
- What type of person commits the most fraud?

AMOUNT OF FRAUD PRESENT IN THE DATA

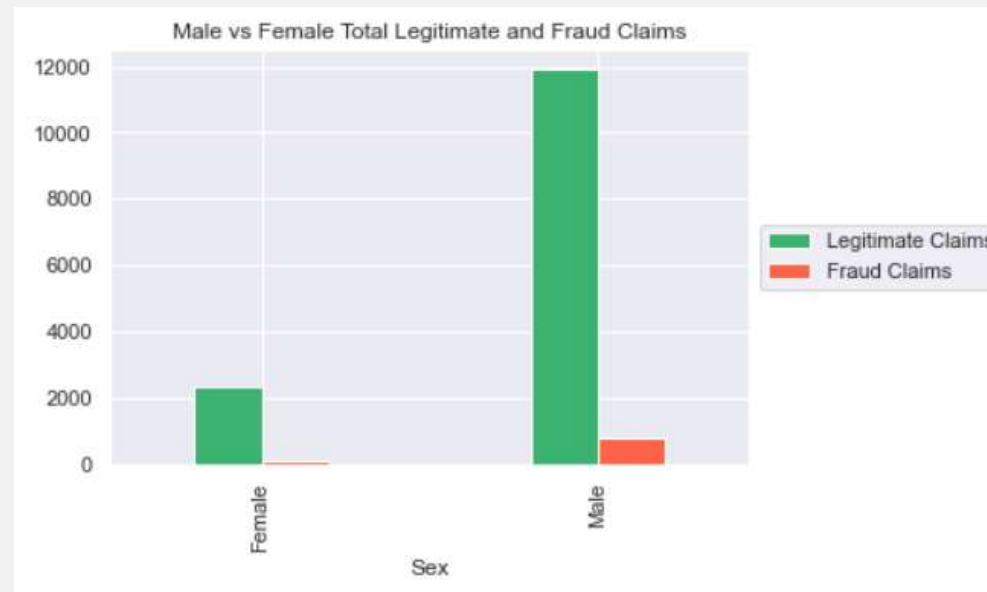
- Of the 15,100 accident claims, 892 were found to be fraudulent



MALE VS FEMALE FRAUD CASES

- The total claims for males: 12,680
- The total claims for females: 2,420
- The total fraudulent claims for males: 787
- The total fraudulent claims for females: 105

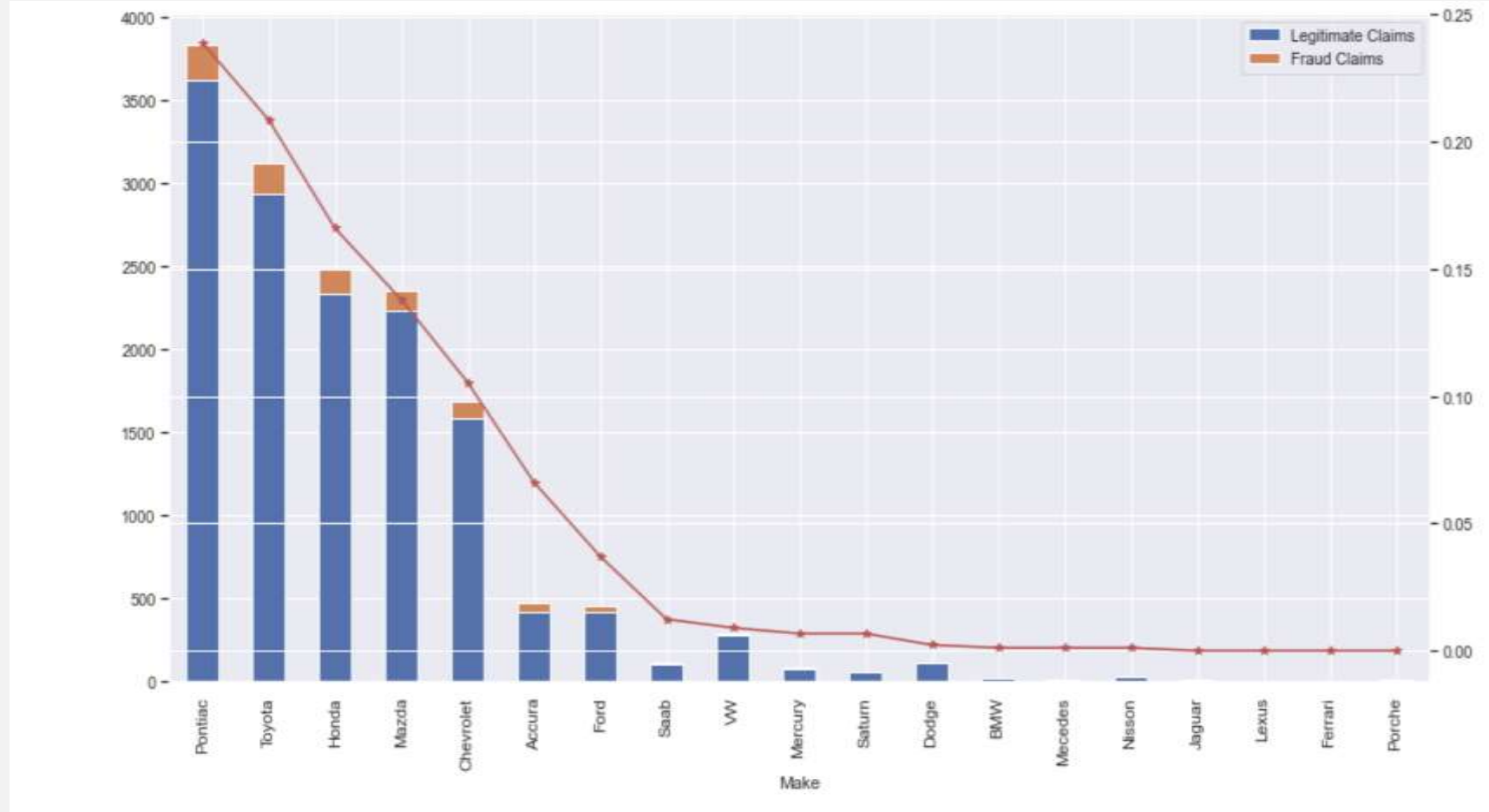
	Sex	FraudFound_P
0	Female	4.3388%
1	Male	6.2066%



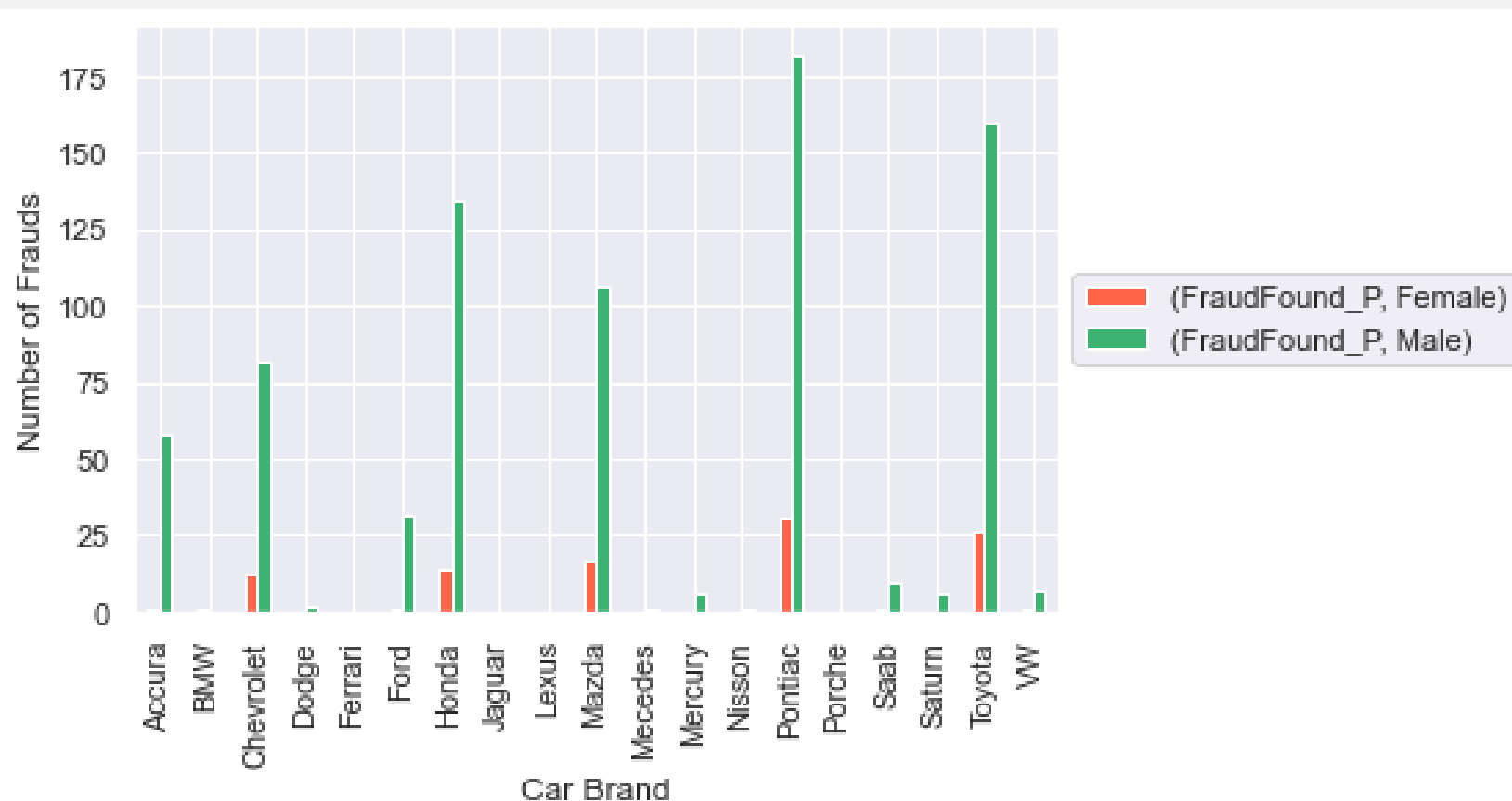
TOTAL CLAIMS BASED ON VEHICLE MAKE

Make	Total
Pontiac	3837
Toyota	3121
Honda	2482
Mazda	2354
Chevrolet	1681
Accura	472
Ford	450
VW	283
Dodge	108
Saab	108
Mercury	83
Saturn	58
Nissan	30
BMW	15
Jaguar	6
Porche	5
Mecedes	4
Ferrari	2
Lexus	1

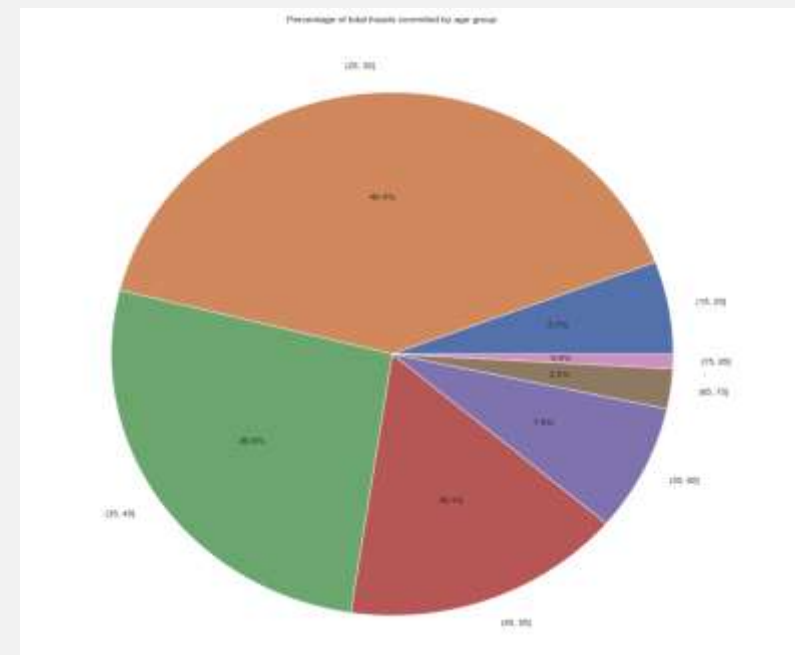
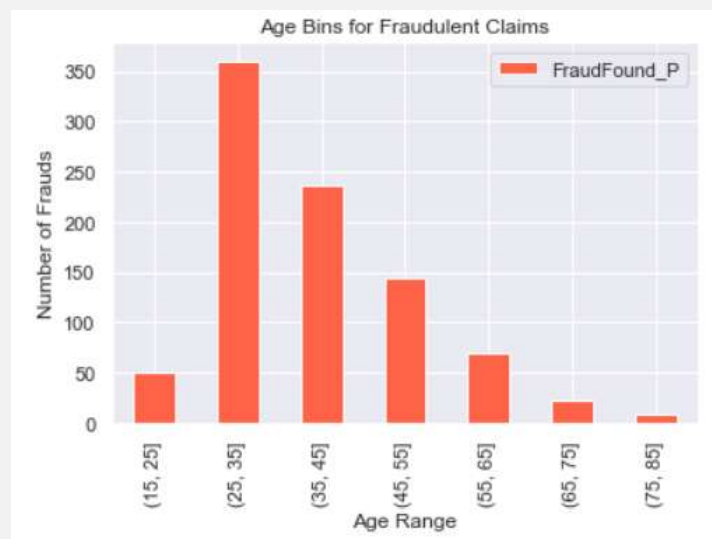
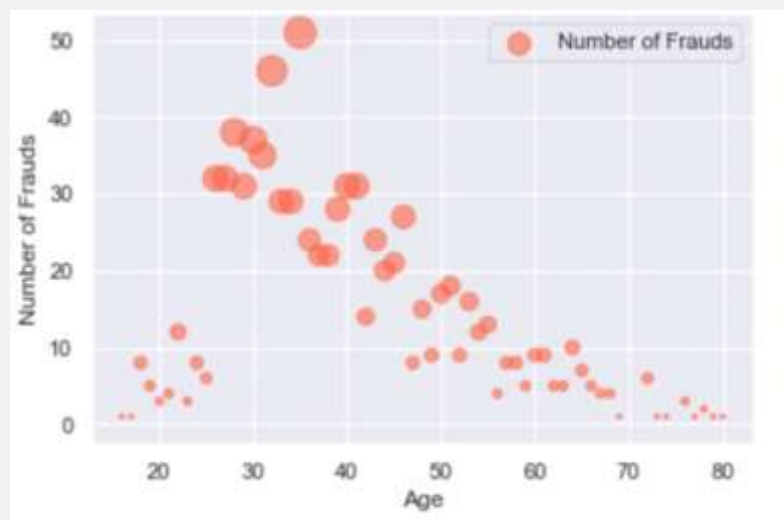
Make	Count	% of Total
Pontiac	213	23.88%
Toyota	186	20.85%
Honda	148	16.59%
Mazda	123	13.79%
Chevrolet	94	10.54%
Accura	59	6.61%
Ford	33	3.70%
Saab	11	1.23%
VW	8	0.90%
Saturn	6	0.67%
Mercury	6	0.67%
Dodge	2	0.22%
Nissan	1	0.11%
Mecedes	1	0.11%
BMW	1	0.11%



AMOUNT OF FRAUD CASES WITH RESPECT TO VEHICLE MAKE AND GENDER

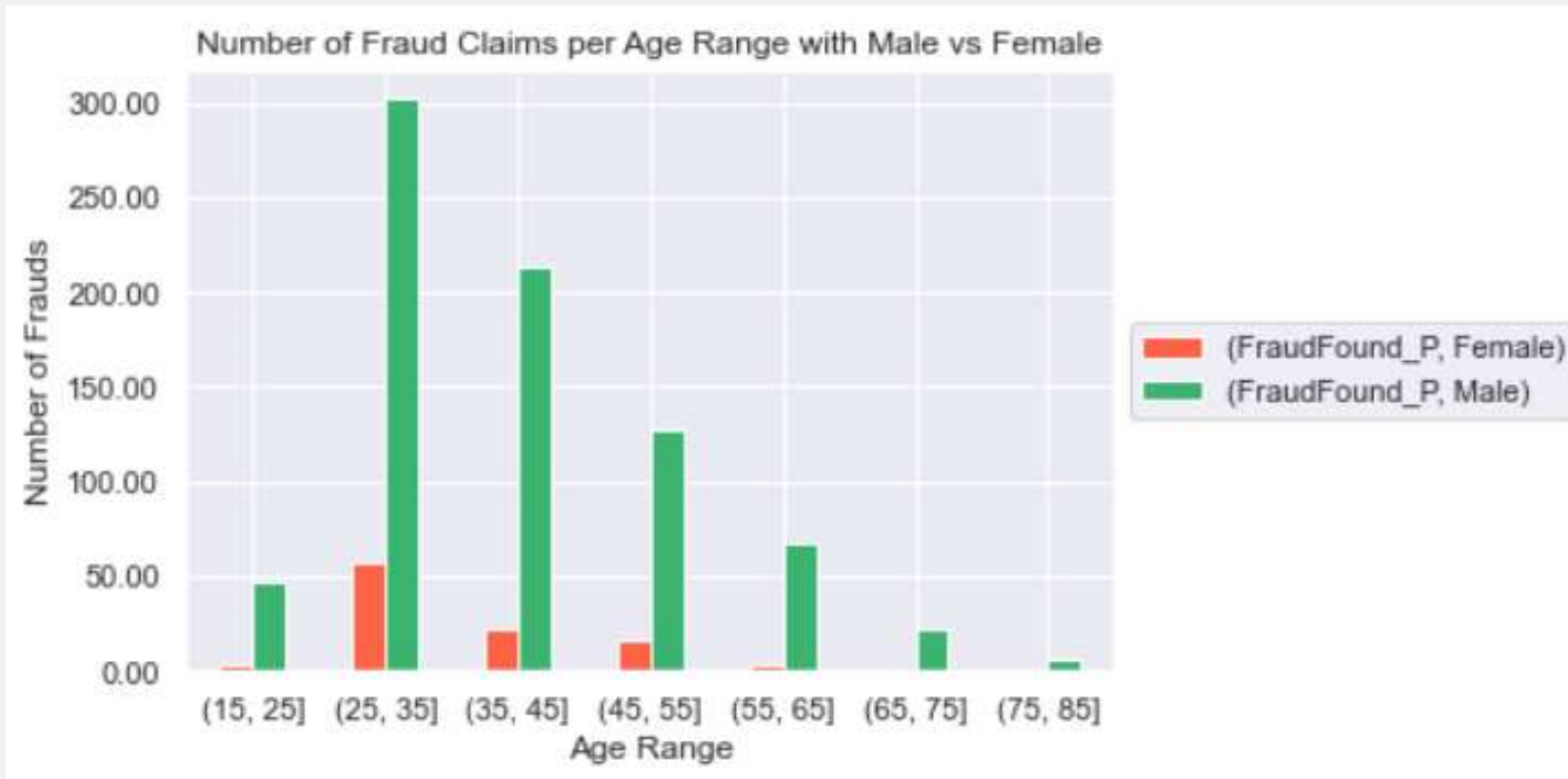


AGE WITH THE HIGHEST FRAUD CLAIMS



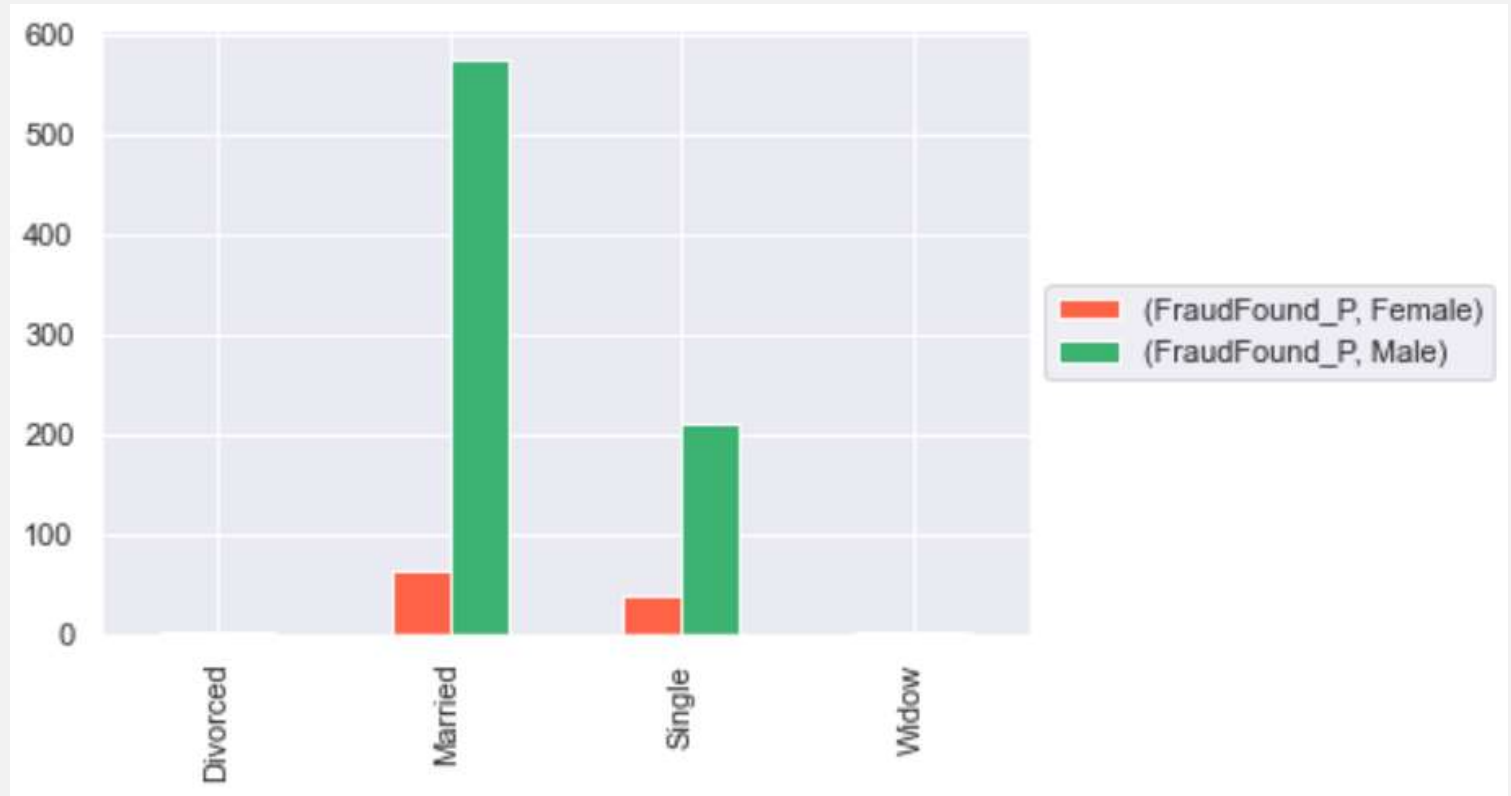
Fraud spikes after 25 years old, then gradually goes down as age increases
People who are 30 have the highest number of claims filed

AGE RANGE AND GENDER FRAUD CLAIMS



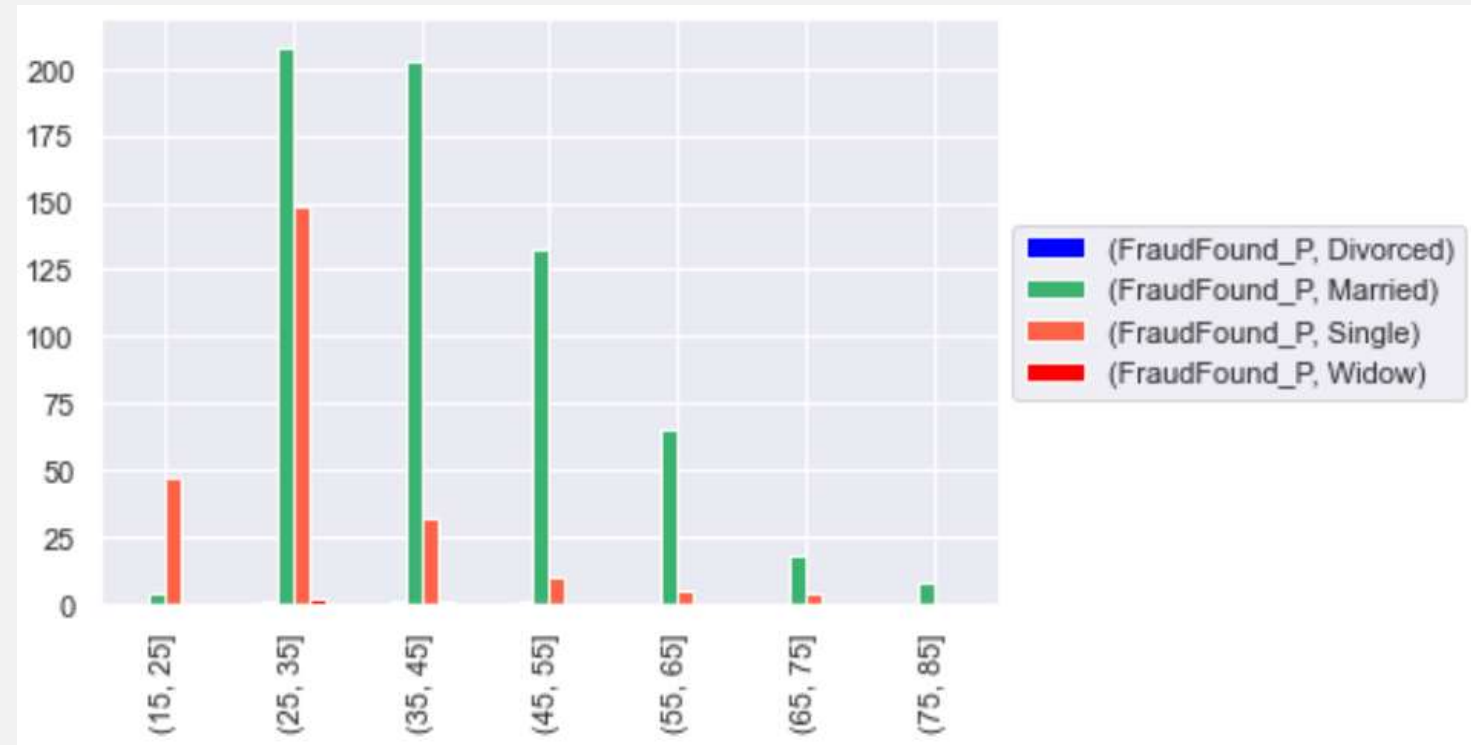
AMOUNT OF FRAUD CLAIMS BASED ON MARITAL STATUS AND GENDER

Sex	MaritalStatus	FraudFound_P	
		Female	Male
0	Divorced	1	2
1	Married	64	575
2	Single	38	209
3	Widow	2	1



AMOUNT OF FRAUD BASED ON MARITAL STATUS AND AGE

		FraudFound_P			
	Age_Range	Divorced	Married	Single	Widow
0	(15, 25]	0	4	47	0
1	(25, 35]	1	208	149	2
2	(35, 45]	1	203	32	1
3	(45, 55]	1	133	10	0
4	(55, 65]	0	65	5	0
5	(65, 75]	0	18	4	0
6	(75, 85]	0	8	0	0



Fraud is committed predominantly by married people between 25 and 45 years of age

WHERE DOES FRAUD OCCUR?

Total Claims

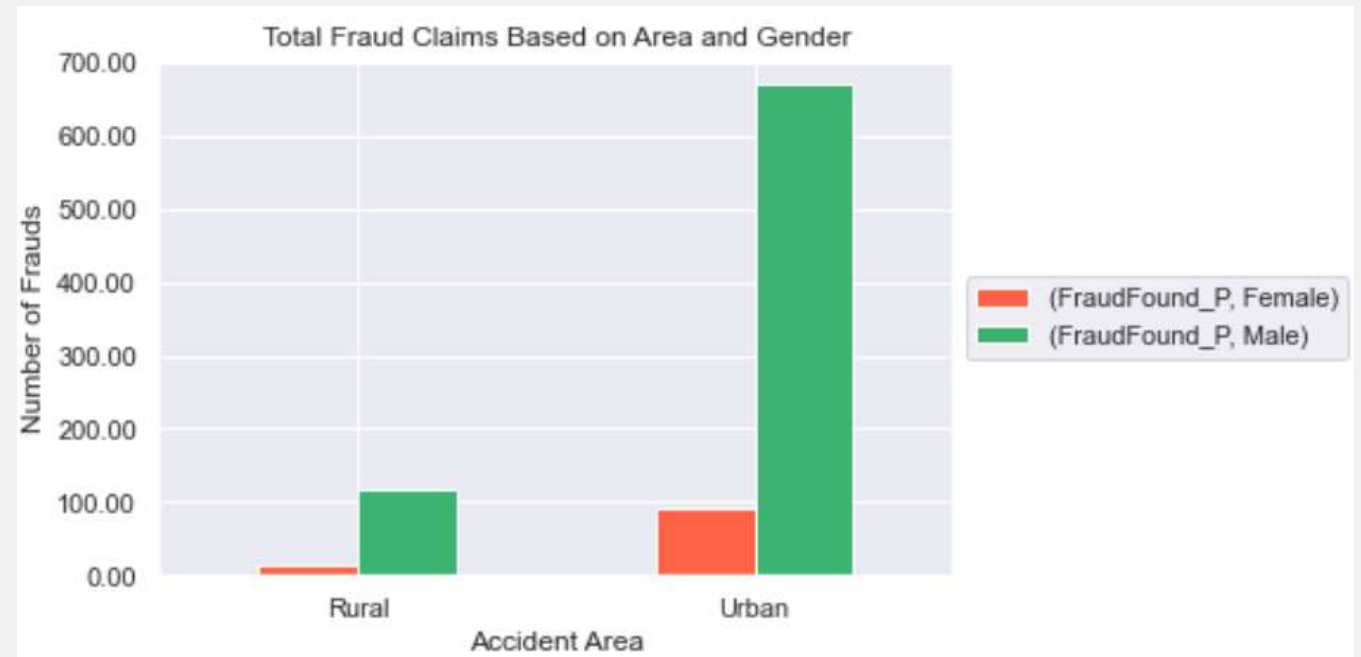
	AccidentArea
Urban	13554
Rural	1546

Total Fraud Claims

	AccidentArea
Urban	761
Rural	131

Total Fraud Claims with Gender

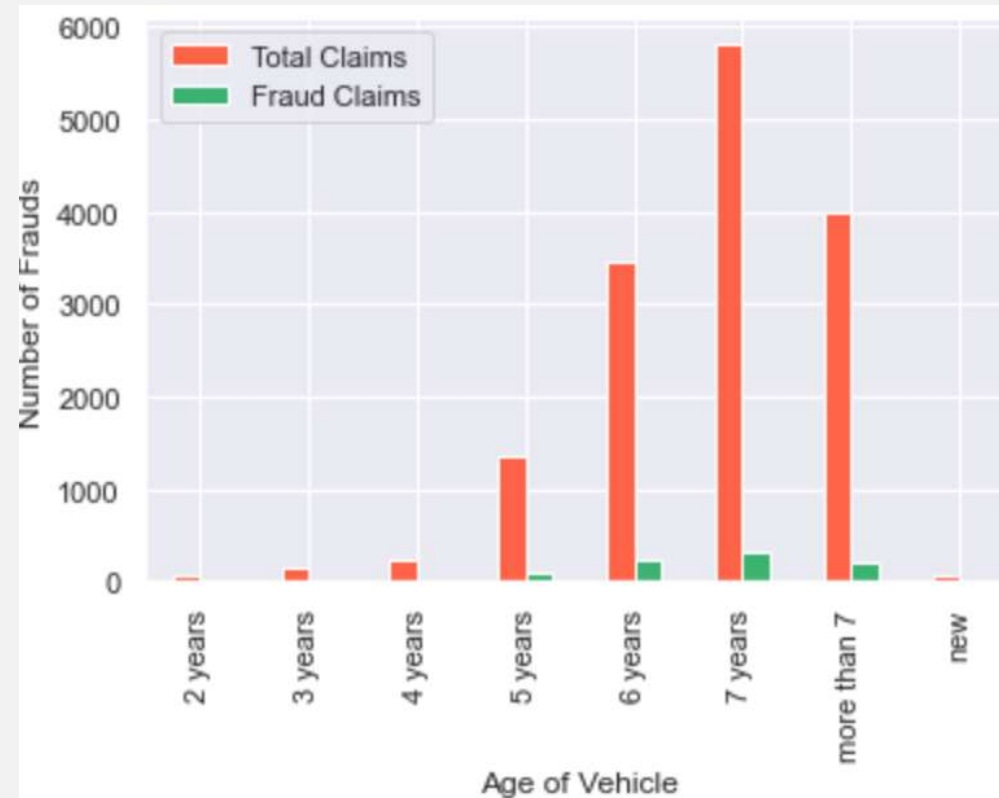
Sex	AccidentArea	FraudFound_P	
		Female	Male
0	Rural	14	117
1	Urban	91	670



Most frauds happen in urban areas where the population is higher, and traffic accidents are more severe.

HOW OLD ARE THE VEHICLES IN FRAUD CASES?

Age	AgeOfVehicle	Sex	Count
32	6 years	Male	23
35	6 years	Male	21
30	6 years	Male	20
41	7 years	Male	19
35	7 years	Male	19
36	7 years	Male	18
39	7 years	Male	18
27	6 years	Male	16
31	6 years	Male	14
26	6 years	Male	14

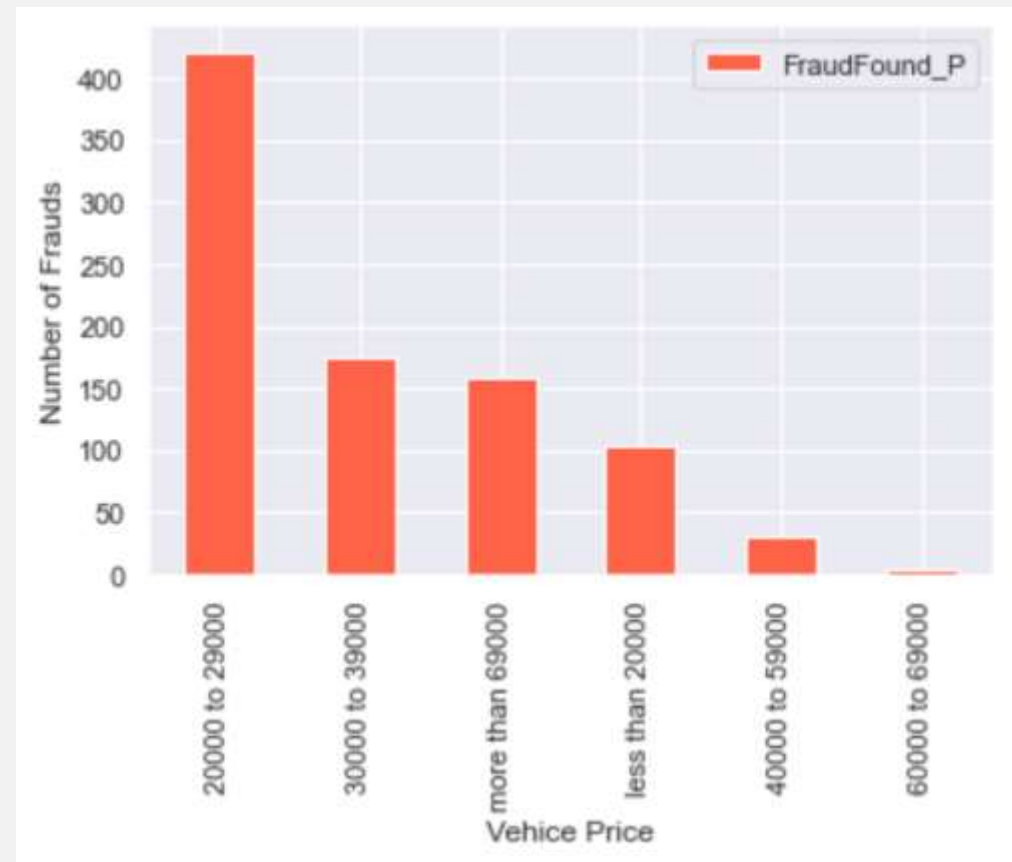


As the vehicle age increases, a car insurance fraud becomes more likely

PRICE OF VEHICLES IN FRAUD CLAIMS

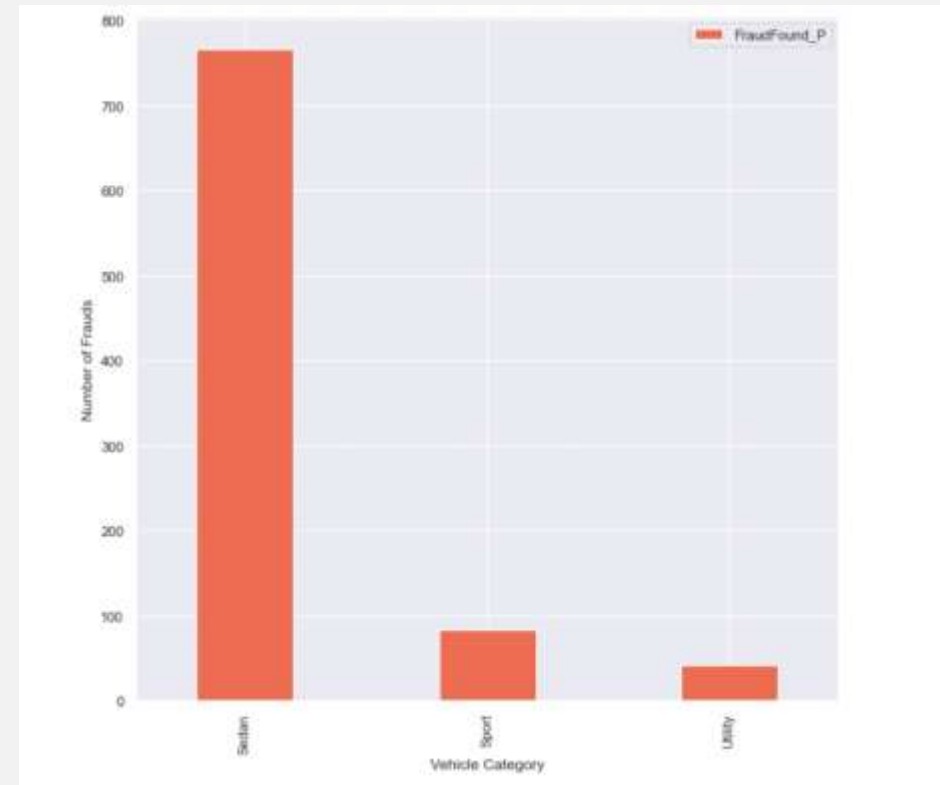
	VehiclePrice	FraudFound_P
0	20000 to 29000	421
1	30000 to 39000	175
5	more than 69000	158
4	less than 20000	103
2	40000 to 59000	31
3	60000 to 69000	4

It is more likely for owners to commit frauds if their vehicle value is low



TYPE OF VEHICLES IN FRAUD CLAIMS

	VehicleCategory	FraudFound_P
0	Sedan	767
1	Sport	83
2	Utility	42



It is more likely for owners of Sedans to commit frauds

BUILDING A PROFILE FOR FRAUD CLAIMS

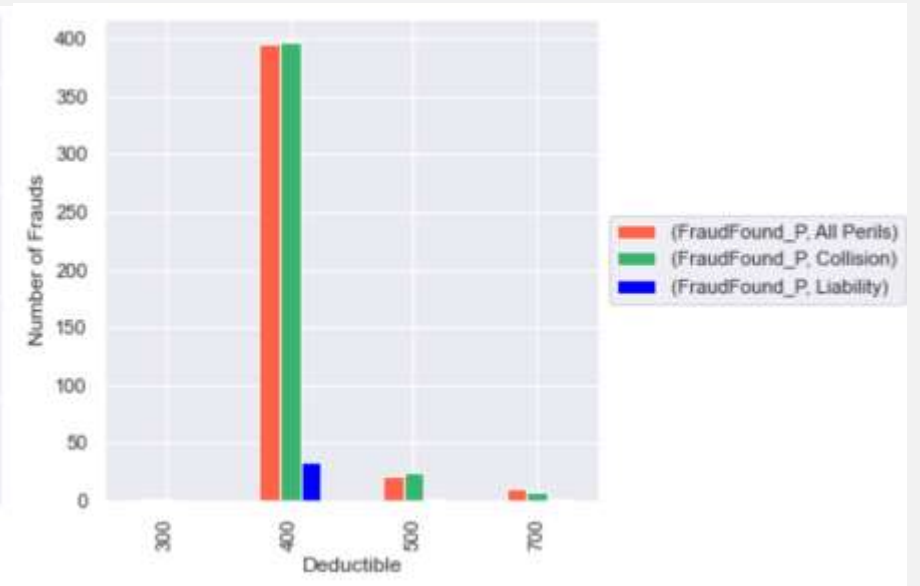
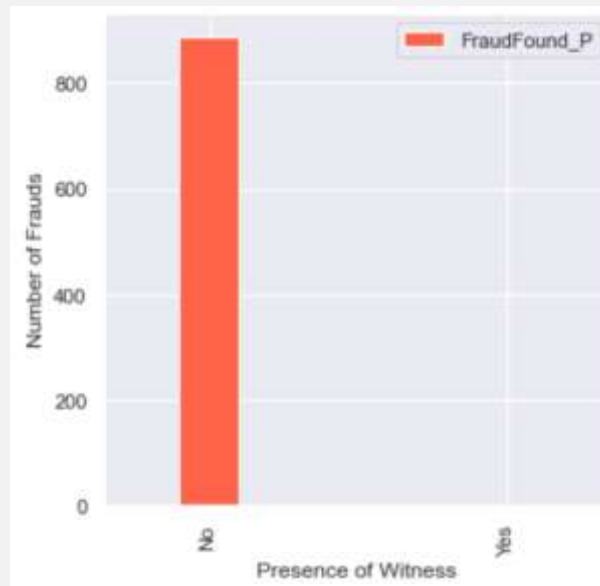
AgeRange	MaritalStatus	Sex	VehicleCategory	AgeOfVehicle	VehiclePrice	AccidentArea	Count
(35, 45]	Married	Male	Sedan	7 years	20000 to 29000	Urban	46
(25, 35]	Married	Male	Sedan	6 years	20000 to 29000	Urban	38
(25, 35]	Single	Male	Sedan	6 years	20000 to 29000	Urban	33
(25, 35]	Married	Male	Sedan	7 years	20000 to 29000	Urban	29
(45, 55]	Married	Male	Sedan	more than 7	20000 to 29000	Urban	21
(35, 45]	Married	Male	Sedan	7 years	30000 to 39000	Urban	18
(25, 35]	Married	Male	Sedan	7 years	30000 to 39000	Urban	17
(45, 55]	Married	Male	Sedan	more than 7	30000 to 39000	Urban	16
(35, 45]	Married	Male	Sedan	more than 7	30000 to 39000	Urban	14
(25, 35]	Single	Male	Sedan	6 years	30000 to 39000	Urban	13

BUSINESS INSIGHTS

- Insurance companies should use the profile we developed to hedge the risk associated with a given policy holder.
- Insurance companies can raise premiums on people fitting this profile to reduce financial risk and protect the bottom line.
- Insurance ads can be targeted towards certain demographics of the population where higher fraud rates are detected

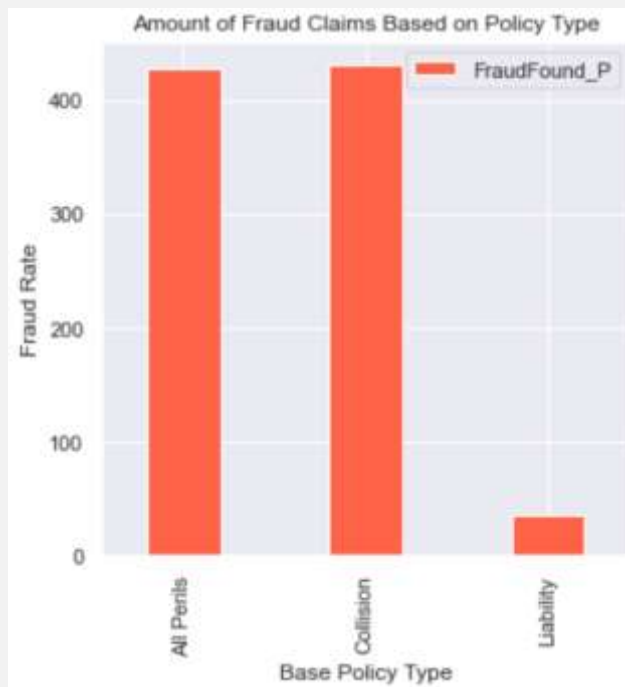
APPENDIX

OTHER DATA EXAMINED BUT DETERMINED TO BE NOT INSIGHTFUL

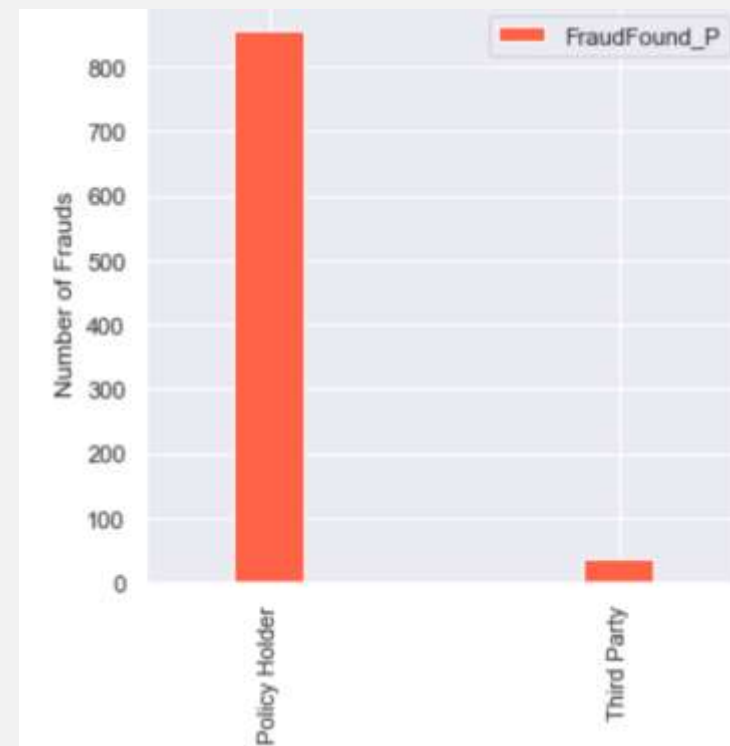


	WitnessPresent	FraudFound_P
0	No	889
1	Yes	3

OTHER DATA EXAMINED BUT DETERMINED TO BE NOT INSIGHTFUL (CONT.)



	BasePolicy	FraudFound_P
0	All Perils	427
1	Collision	430
2	Liability	35



	Fault	FraudFound_P
0	Policy Holder	855
1	Third Party	37