```
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
import numpy as np
import sklearn.metrics
from pylab import rcParams
%matplotlib inline
pd.set_option('display.max_columns', 500)
pd.set_option('display.max_rows', 500)
```

Business case

Claim related fraud is a huge problem in the insurance industry. It is quite complex and difficult to identify those unwanted claims. With Random Forest Non-Parametric Machine Learning Algorithm, I am trying to troubleshoot and help the General Insurance industry with this problem.

The data that I have is from Automobile Insurance. I will be creating a predictive model that predicts if an insurance claim is fraudulent or not. The answere between YES/NO, is a Binary Classification task. A comparison study has been performed to understand which ML algorithm suits best to the dataset.

```
#load & view raw data
df = pd.read_csv('insurance_claims.csv')
df.head(10)
```

```
   months_as_customer  age  policy_number policy_bind_date
policy_state  \
0                 328   48         521585       2014-10-17
OH
1                 228   42         342868       2006-06-27
IN
2                 134   29         687698       2000-09-06
OH
3                 256   41         227811       1990-05-25
IL
4                 228   44         367455       2014-06-06
IL
5                 256   39         104594       2006-10-12
OH
6                 137   34         413978       2000-06-04
IN
7                 165   37         429027       1990-02-03
IL
8                  27   33         485665       1997-02-05
IL
9                 212   42         636550       2011-07-25
IL
```

	policy_csl	policy_deductable	policy_annual_premium	umbrella_limit \
0	250/500	1000	1406.91	0
1	250/500	2000	1197.22	5000000
2	100/300	2000	1413.14	5000000
3	250/500	2000	1415.74	6000000
4	500/1000	1000	1583.91	6000000
5	250/500	1000	1351.10	0
6	250/500	1000	1333.35	0
7	100/300	1000	1137.03	0
8	100/300	500	1442.99	0
9	100/300	500	1315.68	0

	insured_zip	insured_sex	insured_education_level	insured_occupation \
0	466132	MALE	MD	craft-repair
1	468176	MALE	MD	machine-op-inspct
2	430632	FEMALE	PhD	sales
3	608117	FEMALE	PhD	armed-forces
4	610706	MALE	Associate	sales
5	478456	FEMALE	PhD	tech-support
6	441716	MALE	PhD	prof-specialty
7	603195	MALE	Associate	tech-support
8	601734	FEMALE	PhD	other-service
9	600983	MALE	PhD	priv-house-serv

	insured_hobbies	insured_relationship	capital-gains	capital-loss \
0	sleeping	husband	53300	0
1	reading	other-relative	0	0
2	board-games	own-child	35100	0

```
3     board-games            unmarried          48900        -62400
4     board-games            unmarried          66000        -46000
5  bungie-jumping            unmarried              0             0
6     board-games              husband              0        -77000
7    base-jumping            unmarried              0             0
8            golf            own-child              0             0
9         camping                 wife              0        -39300

  incident_date             incident_type   collision_type
incident_severity  \
0    2015-01-25  Single Vehicle Collision   Side Collision      Major
Damage
1    2015-01-21             Vehicle Theft                ?      Minor
Damage
2    2015-02-22   Multi-vehicle Collision   Rear Collision      Minor
Damage
3    2015-01-10  Single Vehicle Collision  Front Collision      Major
Damage
4    2015-02-17             Vehicle Theft                ?      Minor
Damage
5    2015-01-02   Multi-vehicle Collision   Rear Collision      Major
Damage
6    2015-01-13   Multi-vehicle Collision  Front Collision      Minor
Damage
7    2015-02-27   Multi-vehicle Collision  Front Collision
Total Loss
8    2015-01-30  Single Vehicle Collision  Front Collision
Total Loss
9    2015-01-05  Single Vehicle Collision   Rear Collision
Total Loss

  authorities_contacted incident_state incident_city
incident_location  \
0                Police             SC      Columbus      9935 4th
Drive
1                Police             VA     Riverwood        6608 MLK
Hwy
2                Police             NY      Columbus   7121 Francis
Lane
3                Police             OH     Arlington    6956 Maple
Drive
4                  None             NY     Arlington        3041 3rd
Ave
5                  Fire             SC     Arlington  8973 Washington
St
6                Police             NY   Springfield   5846 Weaver
Drive
7                Police             VA      Columbus        3525 3rd
Hwy
```

```
8                 Police             WV     Arlington     4872 Rock
Ridge
9                  Other             NC     Hillsdale    3066 Francis
Ave

   incident_hour_of_the_day  number_of_vehicles_involved
property_damage  \
0                         5                            1
YES
1                         8                            1
?
2                         7                            3
NO
3                         5                            1
?
4                        20                            1
NO
5                        19                            3
NO
6                         0                            3
?
7                        23                            3
?
8                        21                            1
NO
9                        14                            1
NO

   bodily_injuries  witnesses police_report_available
total_claim_amount  \
0                1          2                      YES
71610
1                0          0                        ?
5070
2                2          3                       NO
34650
3                1          2                       NO
63400
4                0          1                       NO
6500
5                0          2                       NO
64100
6                0          0                        ?
78650
7                2          2                      YES
51590
8                1          1                      YES
27700
9                2          1                        ?
```

```
42300

   injury_claim  property_claim  vehicle_claim  auto_make  auto_model
\
0          6510           13020          52080       Saab         92x

1           780             780           3510   Mercedes        E400

2          7700            3850          23100      Dodge         RAM

3          6340            6340          50720  Chevrolet       Tahoe

4          1300             650           4550     Accura         RSX

5          6410            6410          51280       Saab          95

6         21450            7150          50050     Nissan  Pathfinder

7          9380            9380          32830       Audi          A5

8          2770            2770          22160     Toyota       Camry

9          4700            4700          32900       Saab         92x


   auto_year fraud_reported  _c39
0       2004              Y   NaN
1       2007              Y   NaN
2       2007              N   NaN
3       2014              Y   NaN
4       2009              N   NaN
5       2003              Y   NaN
6       2012              N   NaN
7       2015              N   NaN
8       2012              N   NaN
9       1996              N   NaN

df.dtypes

months_as_customer               int64
age                              int64
policy_number                    int64
policy_bind_date                object
policy_state                    object
policy_csl                      object
policy_deductable                int64
policy_annual_premium          float64
umbrella_limit                   int64
insured_zip                      int64
insured_sex                     object
insured_education_level         object
```

```
insured_occupation              object
insured_hobbies                 object
insured_relationship            object
capital-gains                    int64
capital-loss                     int64
incident_date                   object
incident_type                   object
collision_type                  object
incident_severity               object
authorities_contacted           object
incident_state                  object
incident_city                   object
incident_location               object
incident_hour_of_the_day         int64
number_of_vehicles_involved      int64
property_damage                 object
bodily_injuries                  int64
witnesses                        int64
police_report_available         object
total_claim_amount               int64
injury_claim                     int64
property_claim                   int64
vehicle_claim                    int64
auto_make                       object
auto_model                      object
auto_year                        int64
fraud_reported                  object
_c39                           float64
dtype: object
```

```
df.columns
```

```
Index(['months_as_customer', 'age', 'policy_number',
'policy_bind_date',
       'policy_state', 'policy_csl', 'policy_deductable',
       'policy_annual_premium', 'umbrella_limit', 'insured_zip',
'insured_sex',
       'insured_education_level', 'insured_occupation',
'insured_hobbies',
       'insured_relationship', 'capital-gains', 'capital-loss',
       'incident_date', 'incident_type', 'collision_type',
'incident_severity',
       'authorities_contacted', 'incident_state', 'incident_city',
       'incident_location', 'incident_hour_of_the_day',
       'number_of_vehicles_involved', 'property_damage',
'bodily_injuries',
       'witnesses', 'police_report_available', 'total_claim_amount',
       'injury_claim', 'property_claim', 'vehicle_claim', 'auto_make',
       'auto_model', 'auto_year', 'fraud_reported', '_c39'],
      dtype='object')
```

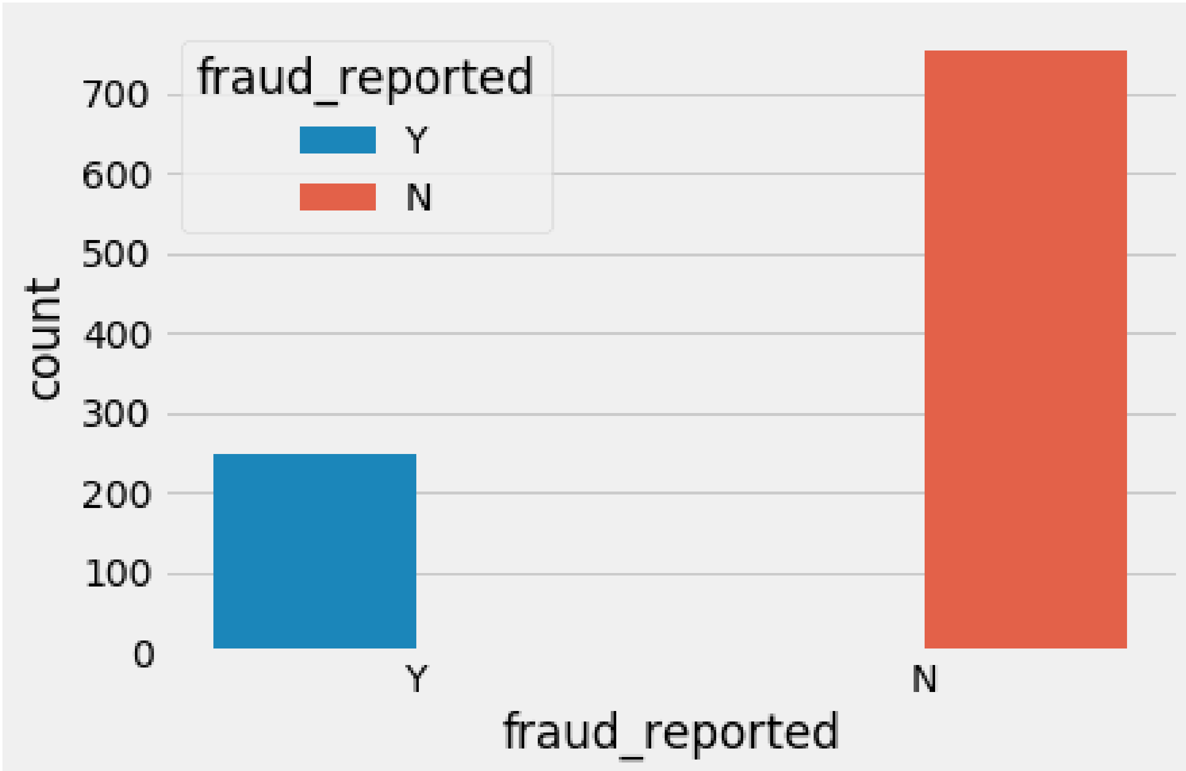
```
df.shape
```

```
(1000, 40)
```

```
df.nunique()
```

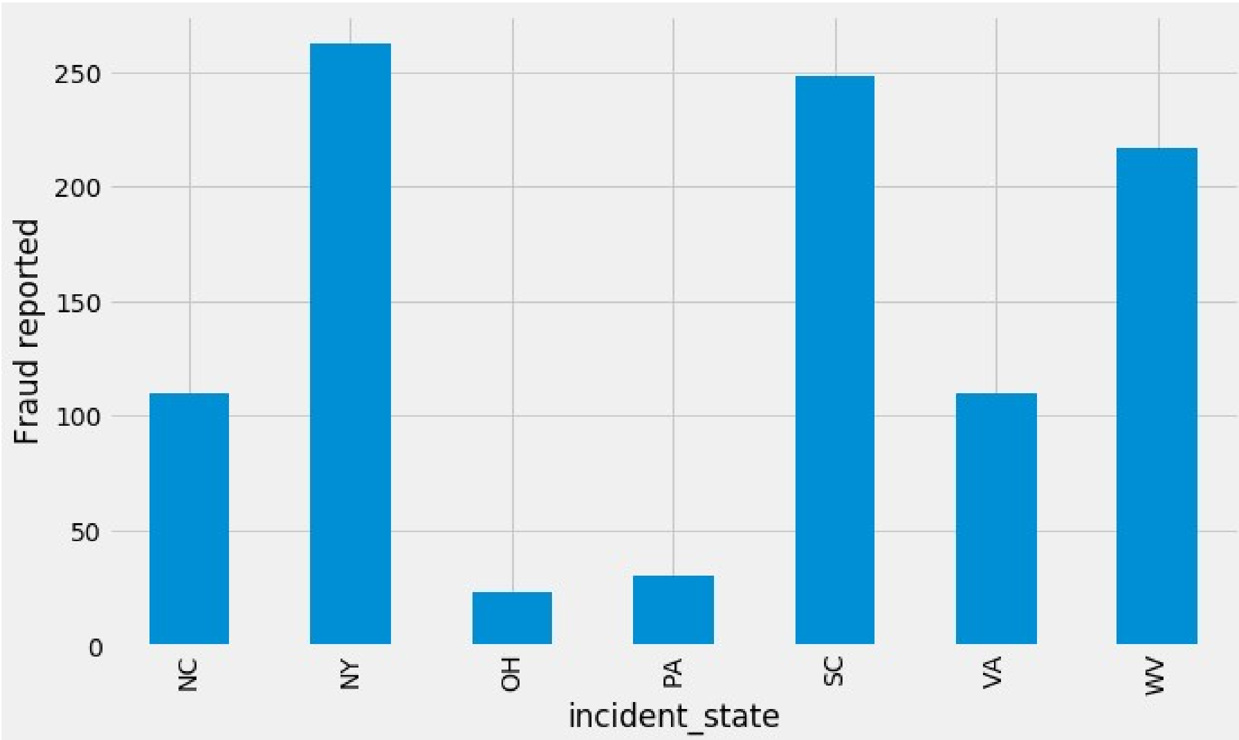
```
months_as_customer              391
age                              46
policy_number                  1000
policy_bind_date                951
policy_state                      3
policy_csl                        3
policy_deductable                 3
policy_annual_premium           991
umbrella_limit                   11
insured_zip                     995
insured_sex                       2
insured_education_level           7
insured_occupation               14
insured_hobbies                  20
insured_relationship              6
capital-gains                   338
capital-loss                    354
incident_date                    60
incident_type                     4
collision_type                    4
incident_severity                 4
authorities_contacted             5
incident_state                    7
incident_city                     7
incident_location              1000
incident_hour_of_the_day         24
number_of_vehicles_involved       4
property_damage                   3
bodily_injuries                   3
witnesses                         4
police_report_available           3
total_claim_amount              763
injury_claim                    638
property_claim                  626
vehicle_claim                   726
auto_make                        14
auto_model                       39
auto_year                        21
fraud_reported                    2
_c39                              0
dtype: int64
```

```
plt.style.use('fivethirtyeight')
ax = sns.countplot(x='fraud_reported', data=df, hue='fraud_reported')
```

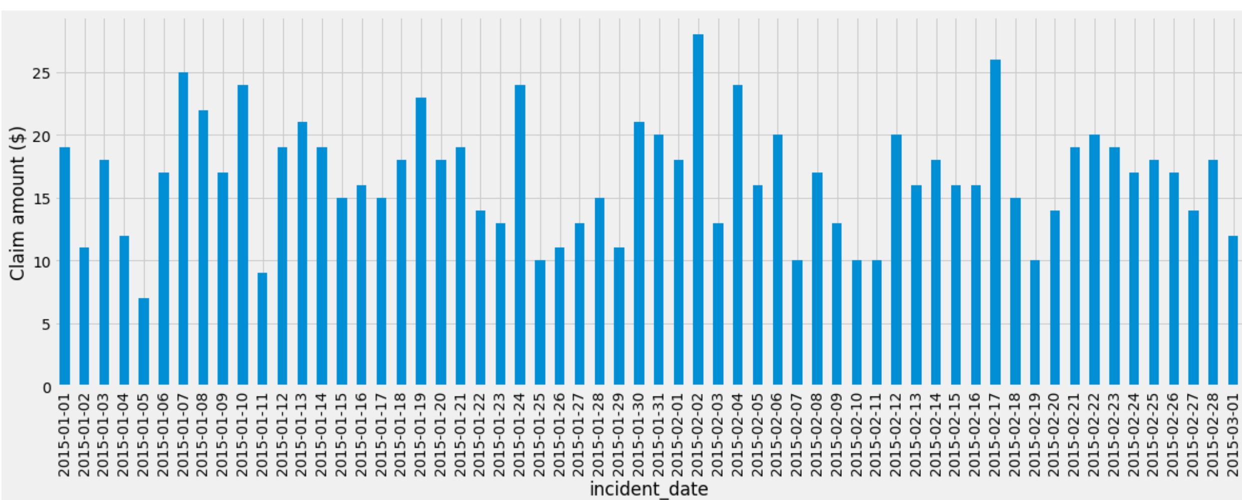


From abobe plot, like most fraud datasets, the label distribution is skewed.

```
df['fraud_reported'].value_counts() # Count number of frauds vs non-frauds

N    753
Y    247
Name: fraud_reported, dtype: int64

df['incident_state'].value_counts()

NY    262
SC    248
WV    217
VA    110
NC    110
PA     30
OH     23
Name: incident_state, dtype: int64

plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax = 
df.groupby('incident_state').fraud_reported.count().plot.bar(ylim=0)
ax.set_ylabel('Fraud reported')
plt.show()
```



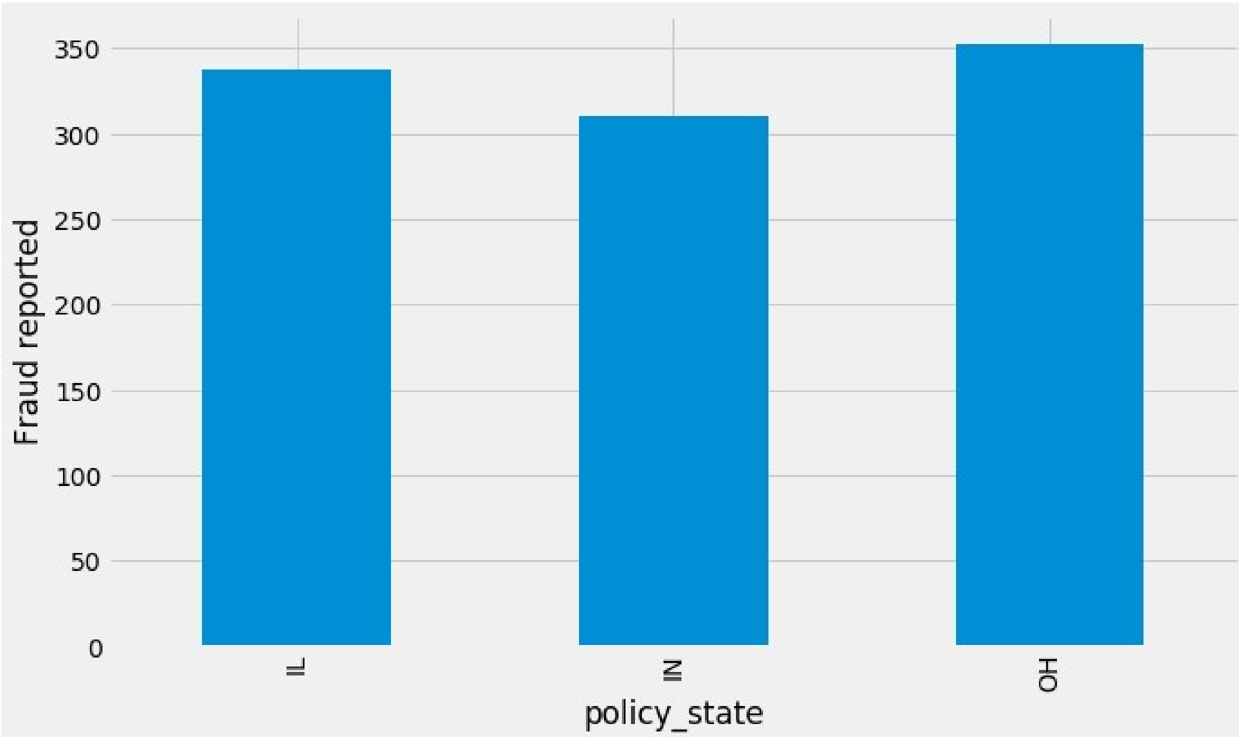

```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(18,6))
ax =
df.groupby('incident_date').total_claim_amount.count().plot.bar(ylim=0
)
ax.set_ylabel('Claim amount ($)')
plt.show()
```



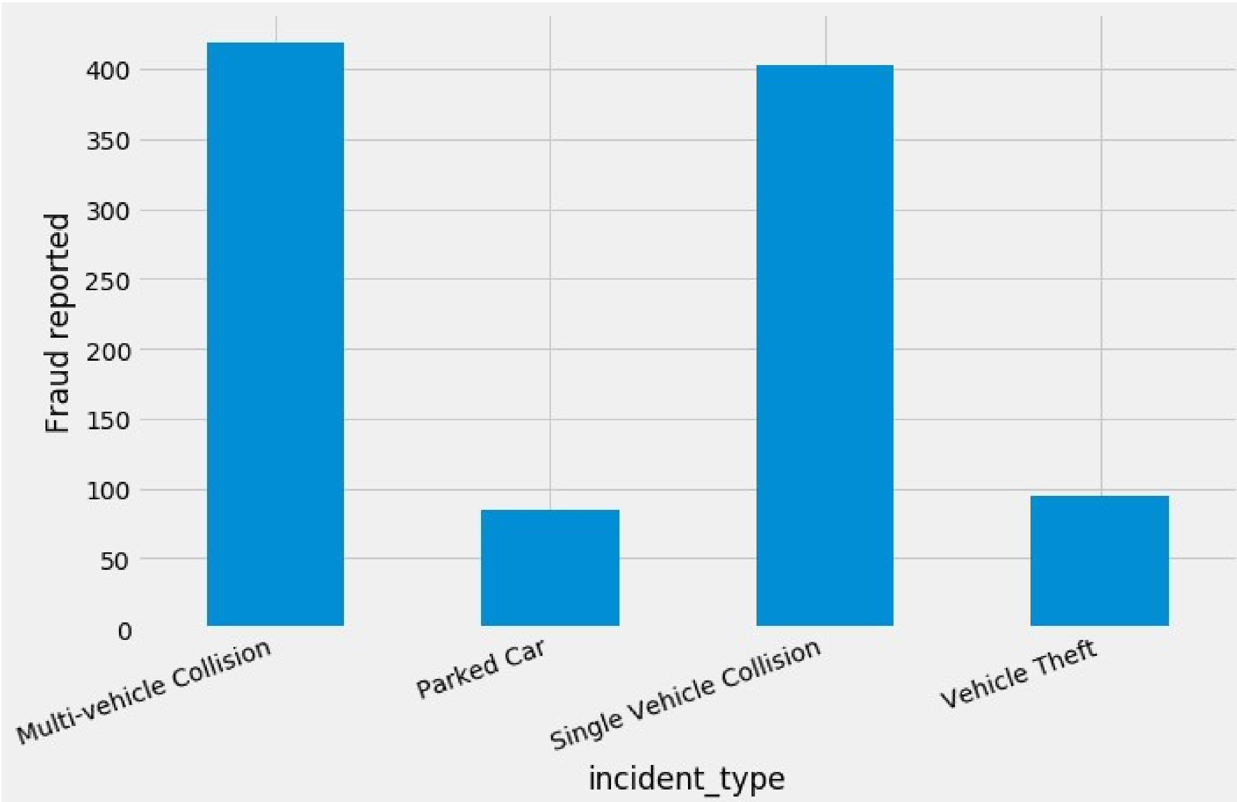
We see that, all the cases in above plot are for the months of January and February 2015

```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
```

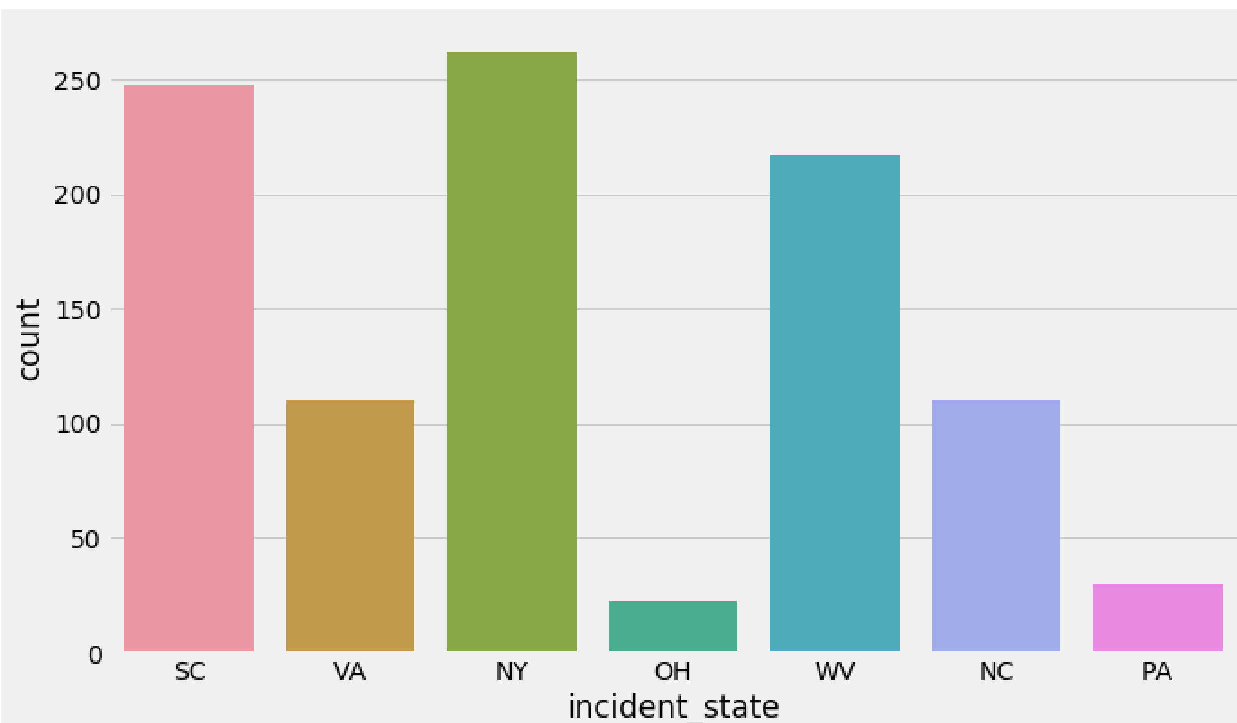
```
ax =
df.groupby('policy_state').fraud_reported.count().plot.bar(ylim=0)
ax.set_ylabel('Fraud reported')
plt.show()
```



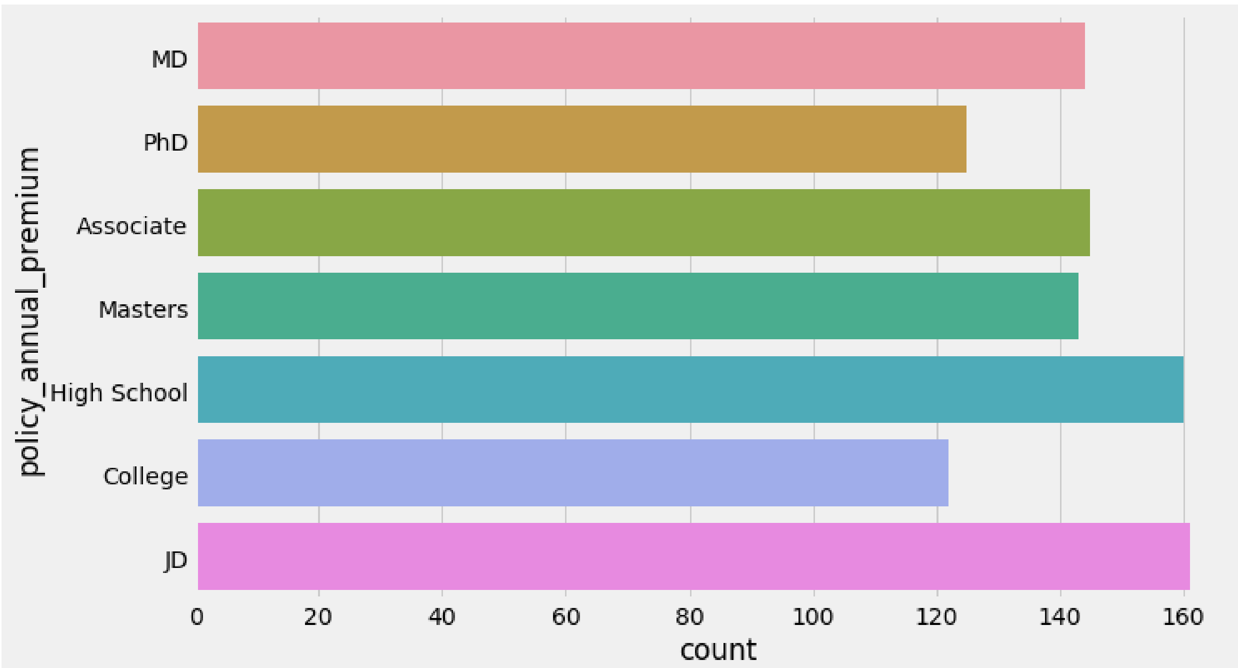

```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax =
df.groupby('incident_type').fraud_reported.count().plot.bar(ylim=0)
ax.set_xticklabels(ax.get_xticklabels(), rotation=20, ha="right")
ax.set_ylabel('Fraud reported')
plt.show()
```



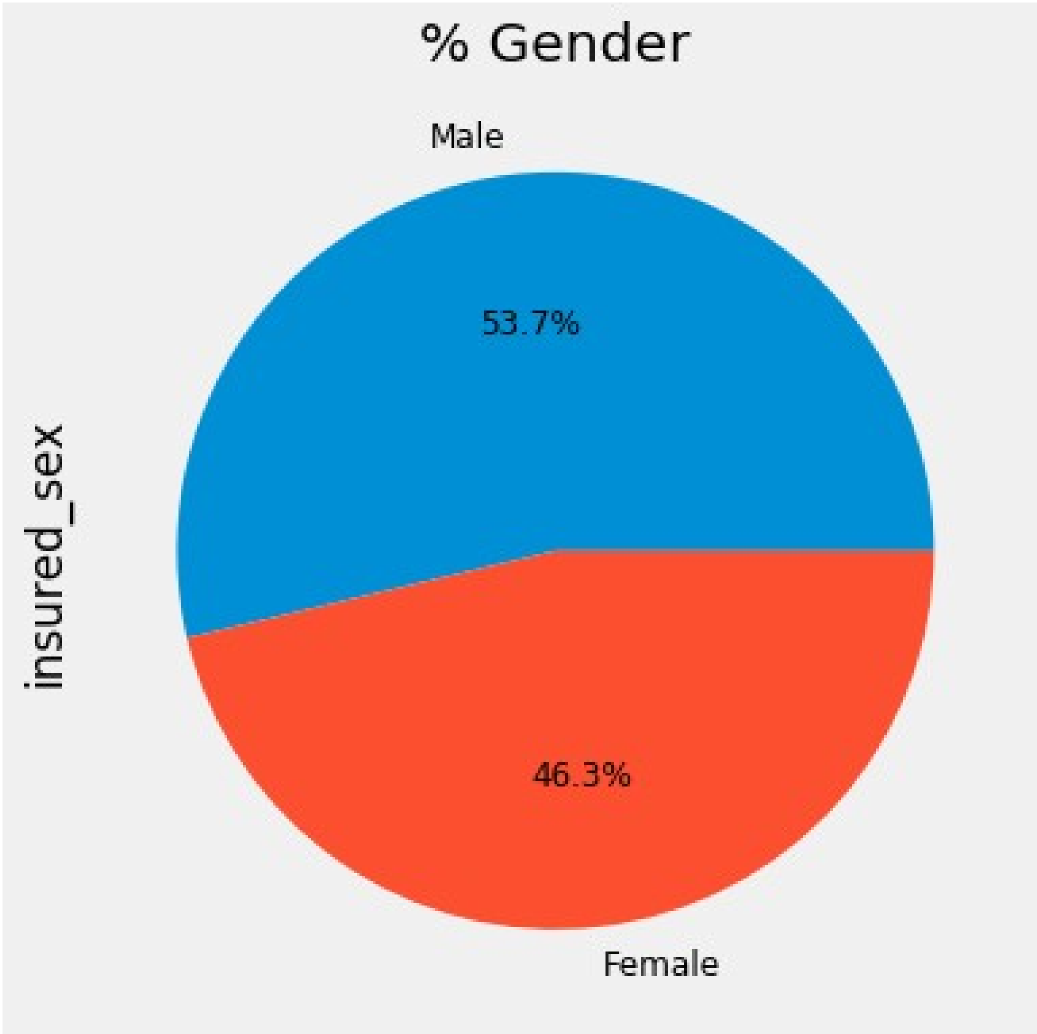

```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax = sns.countplot(x='incident_state', data=df)
```



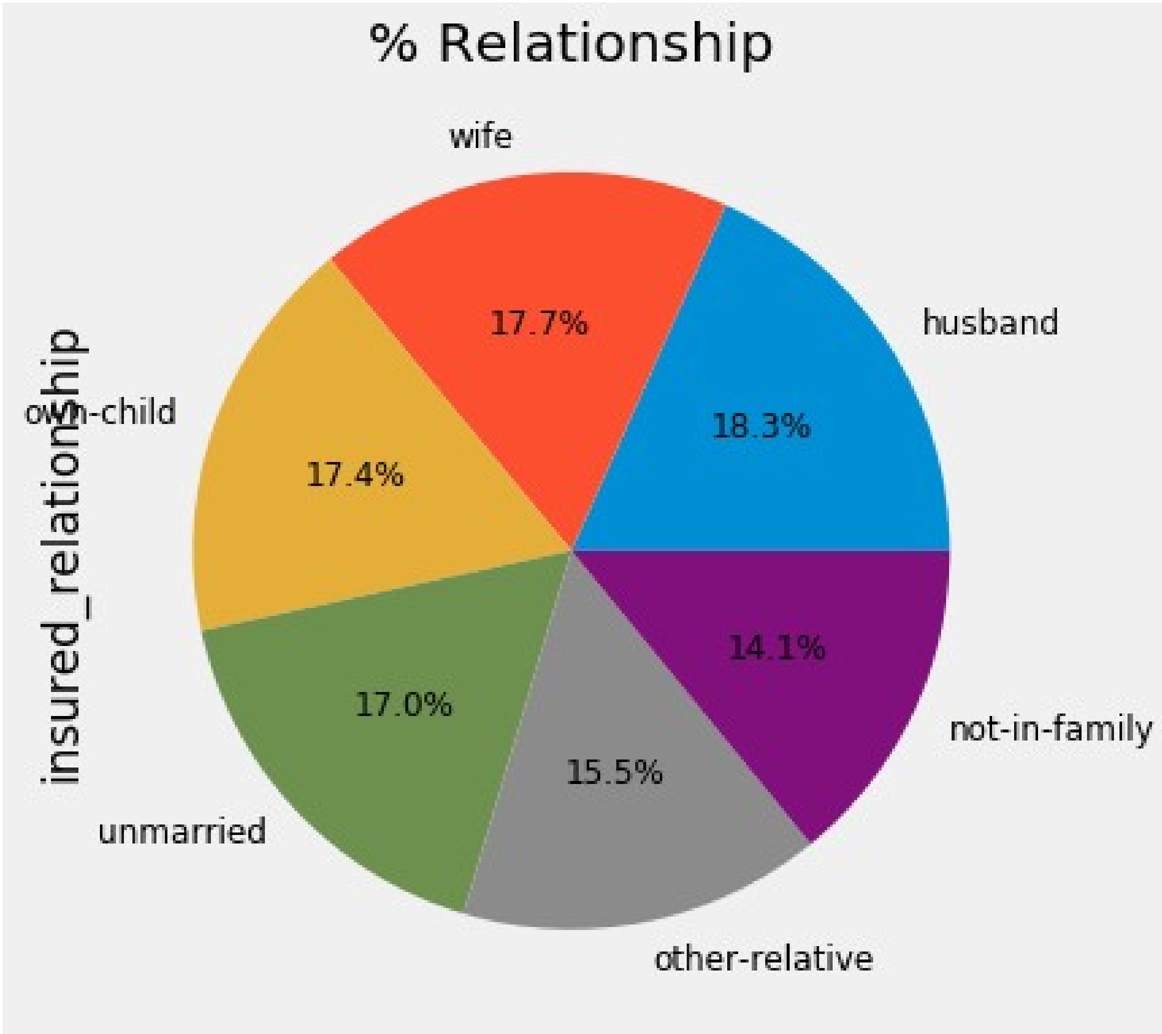
```
fig = plt.figure(figsize=(10,6))
ax = sns.countplot(y = 'insured_education_level', data=df)
ax.set_ylabel('policy_annual_premium')
plt.show()

# # Breakdown of Average Vehicle claim by insured's education level,
grouped by fraud reported
```



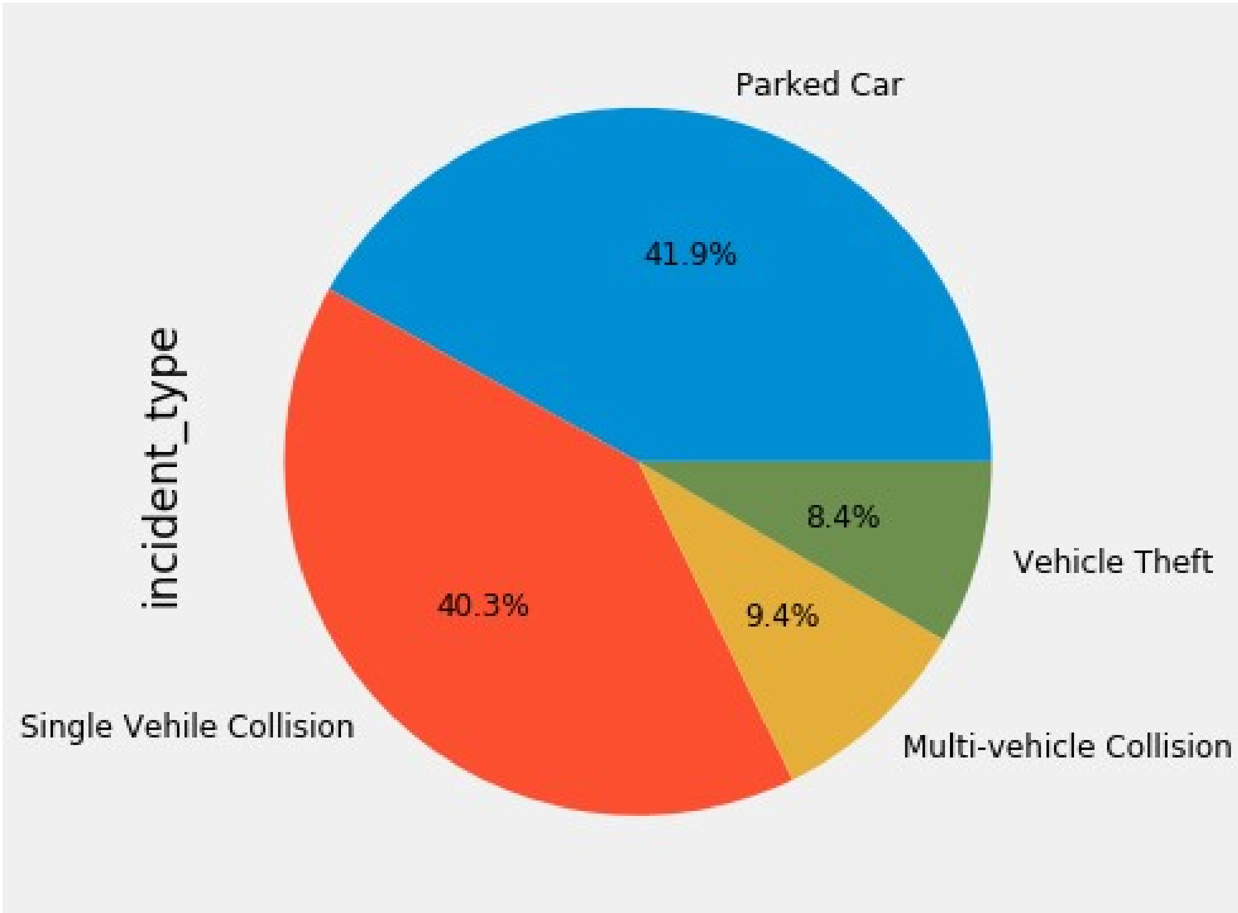

```
fig = plt.figure(figsize=(10,6))
ax = (df['insured_sex'].value_counts()*100.0 /len(df))\
.plot.pie(autopct='%.1f%%', labels = ['Male', 'Female'], fontsize=12)

ax.set_title('% Gender')
plt.show()
```



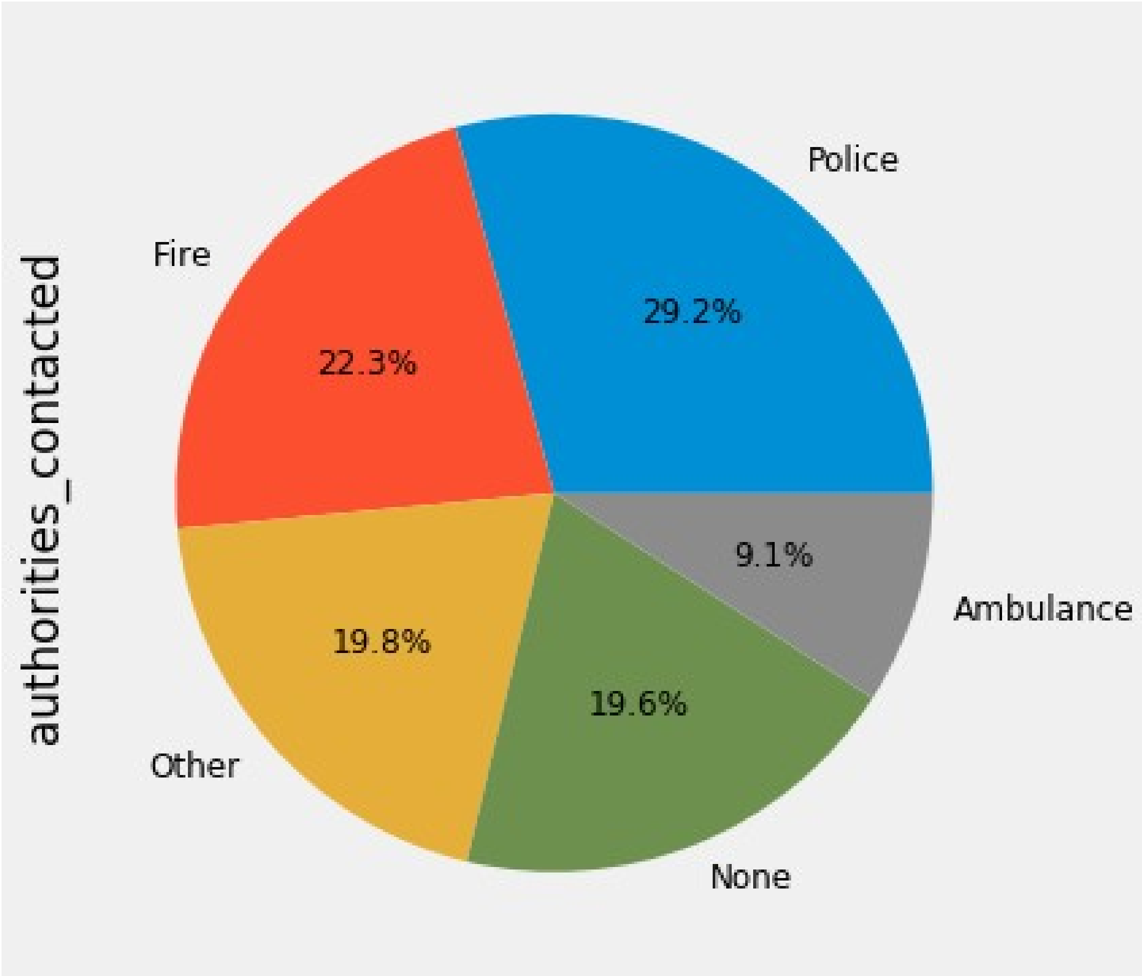

```
fig = plt.figure(figsize=(10,6))
ax = (df['insured_relationship'].value_counts()*100.0 /len(df))\
.plot.pie(autopct='%.1f%%', labels = ['husband', 'wife', 'own-child',
'unmarried', 'other-relative', 'not-in-family'],
         fontsize=12)

ax.set_title('% Relationship')
plt.show()
```



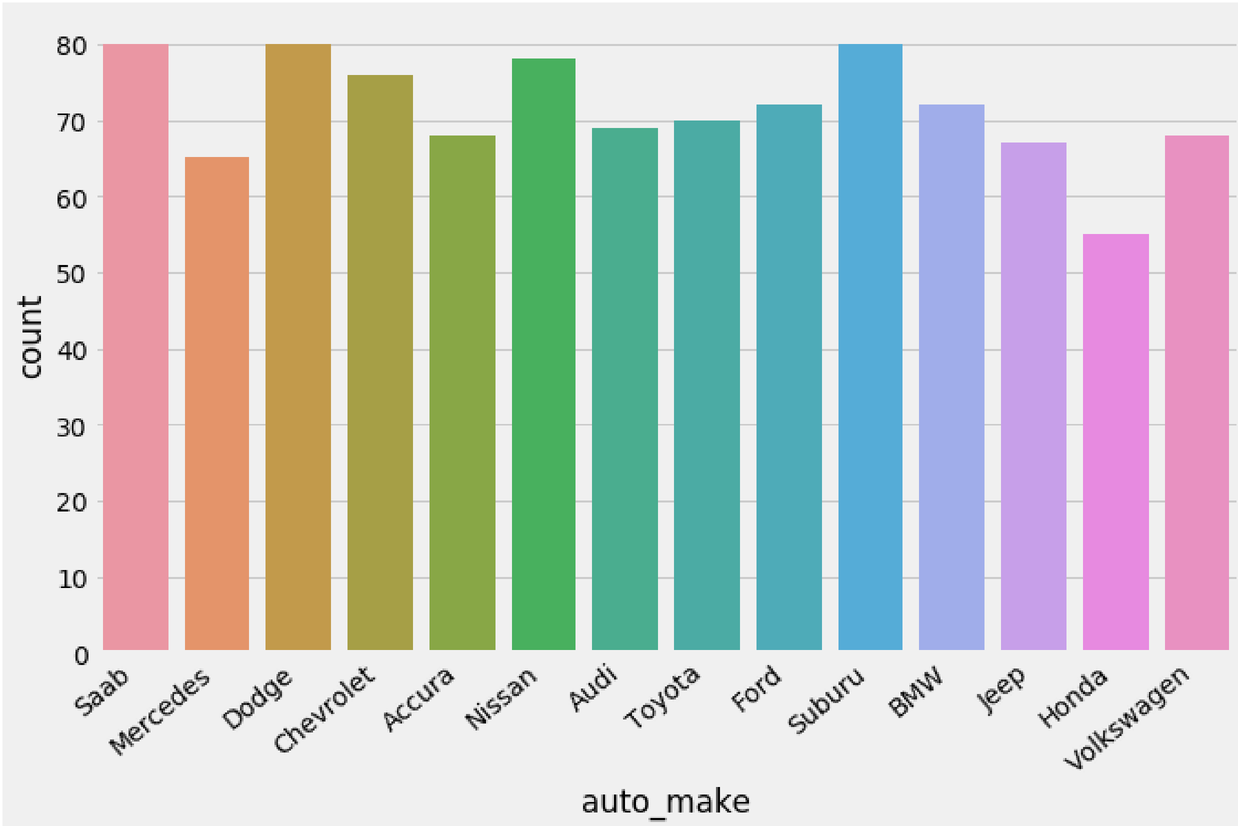

```
fig = plt.figure(figsize=(10,6))
ax = (df['incident_type'].value_counts()*100.0 /len(df))\
.plot.pie(autopct='%.1f%%', labels = ['Parked Car', 'Single Vehile
Collision', 'Multi-vehicle Collision', 'Vehicle Theft'],
         fontsize=12)
```



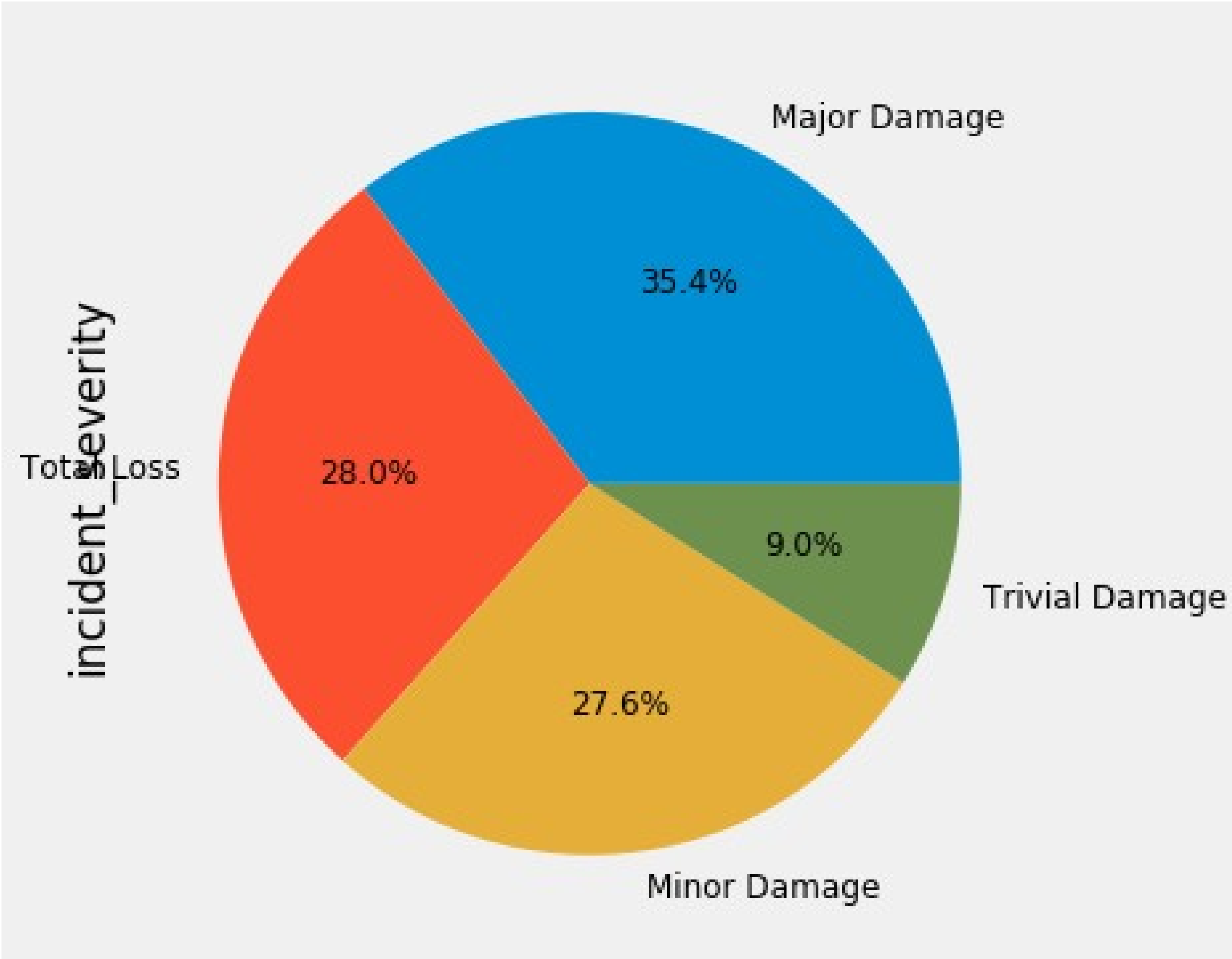

```
fig = plt.figure(figsize=(10,6))
ax = (df['authorities_contacted'].value_counts()*100.0 /len(df))\
.plot.pie(autopct='%.1f%%', labels = ['Police', 'Fire', 'Other',
'None', 'Ambulance'],
         fontsize=12)
```



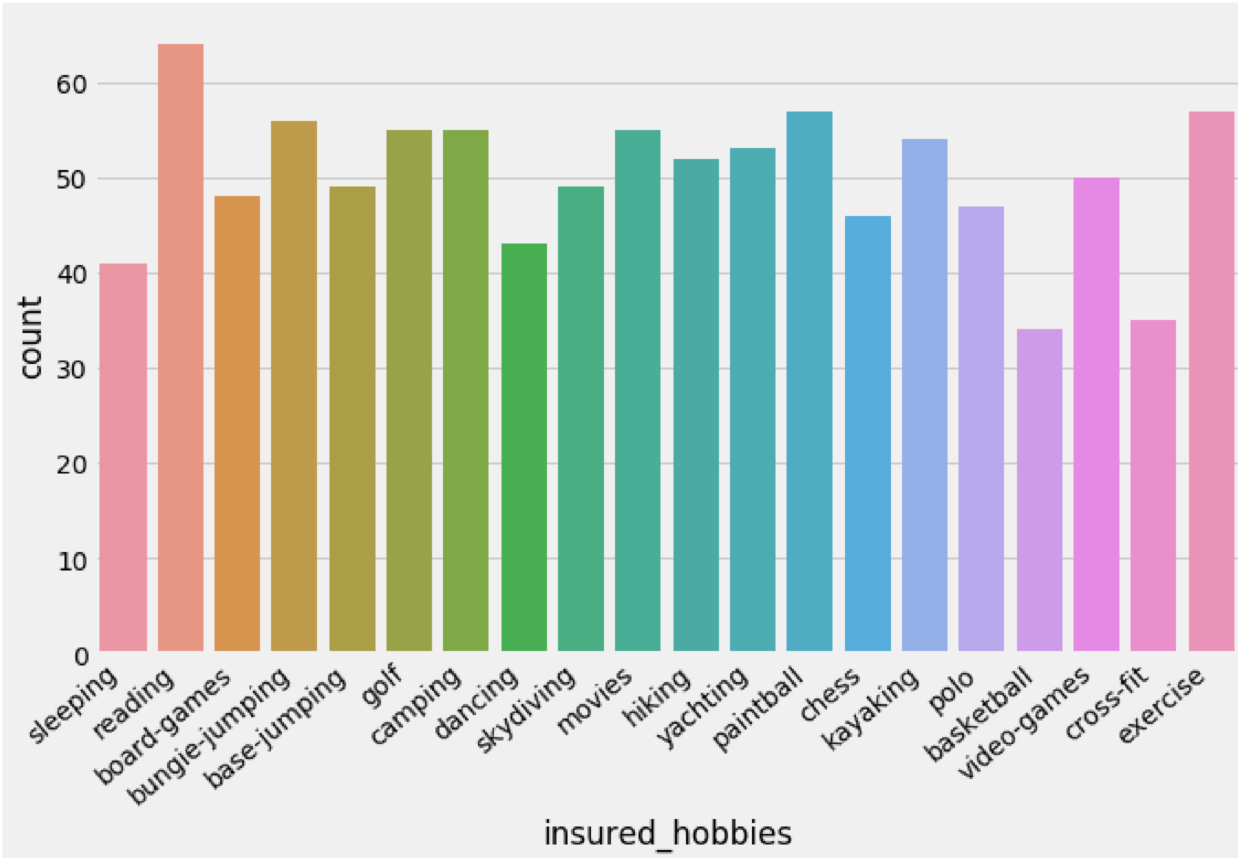

```
fig = plt.figure(figsize=(10,6))
ax = sns.countplot(x='auto_make', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.show()
```



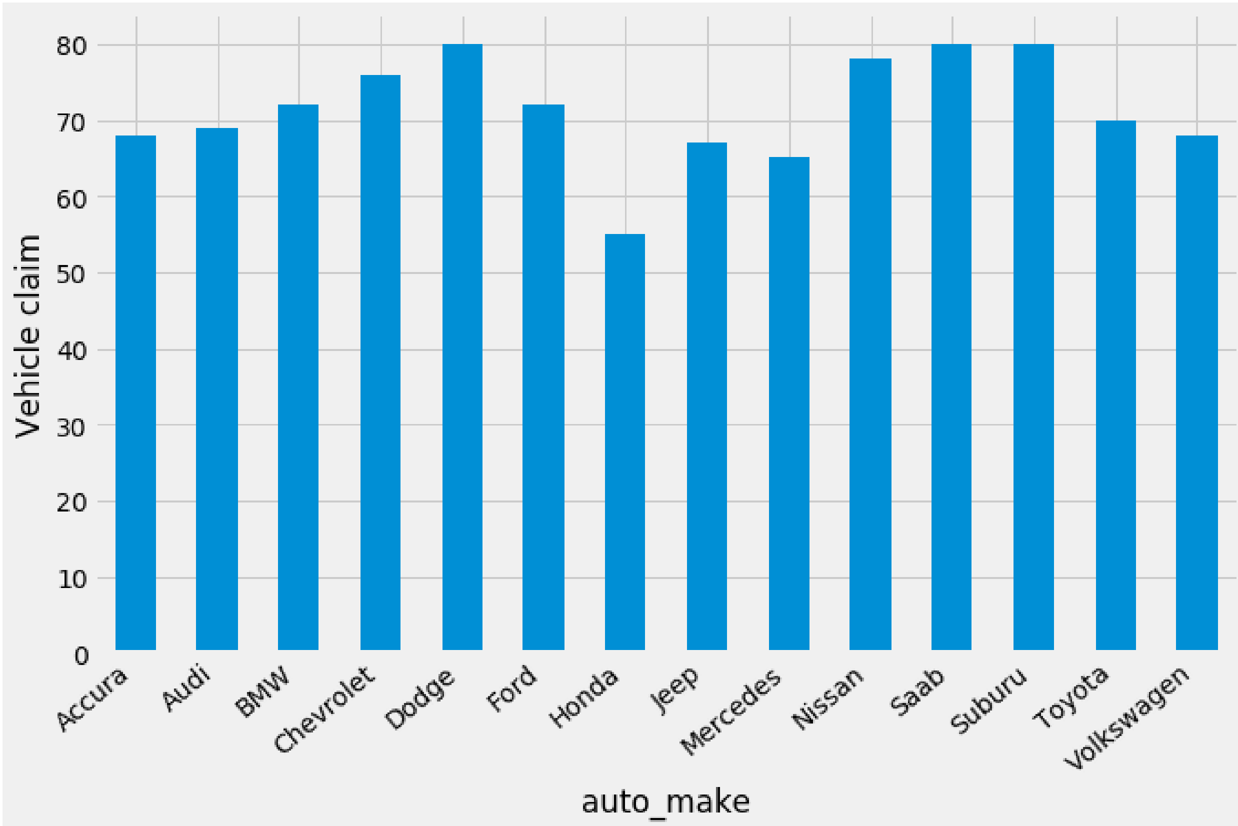

```
fig = plt.figure(figsize=(10,6))
ax = (df['incident_severity'].value_counts()*100.0 /len(df))\
.plot.pie(autopct='%.1f%%', labels = ['Major Damage', 'Total Loss',
'Minor Damage', 'Trivial Damage'],
         fontsize=12)
```



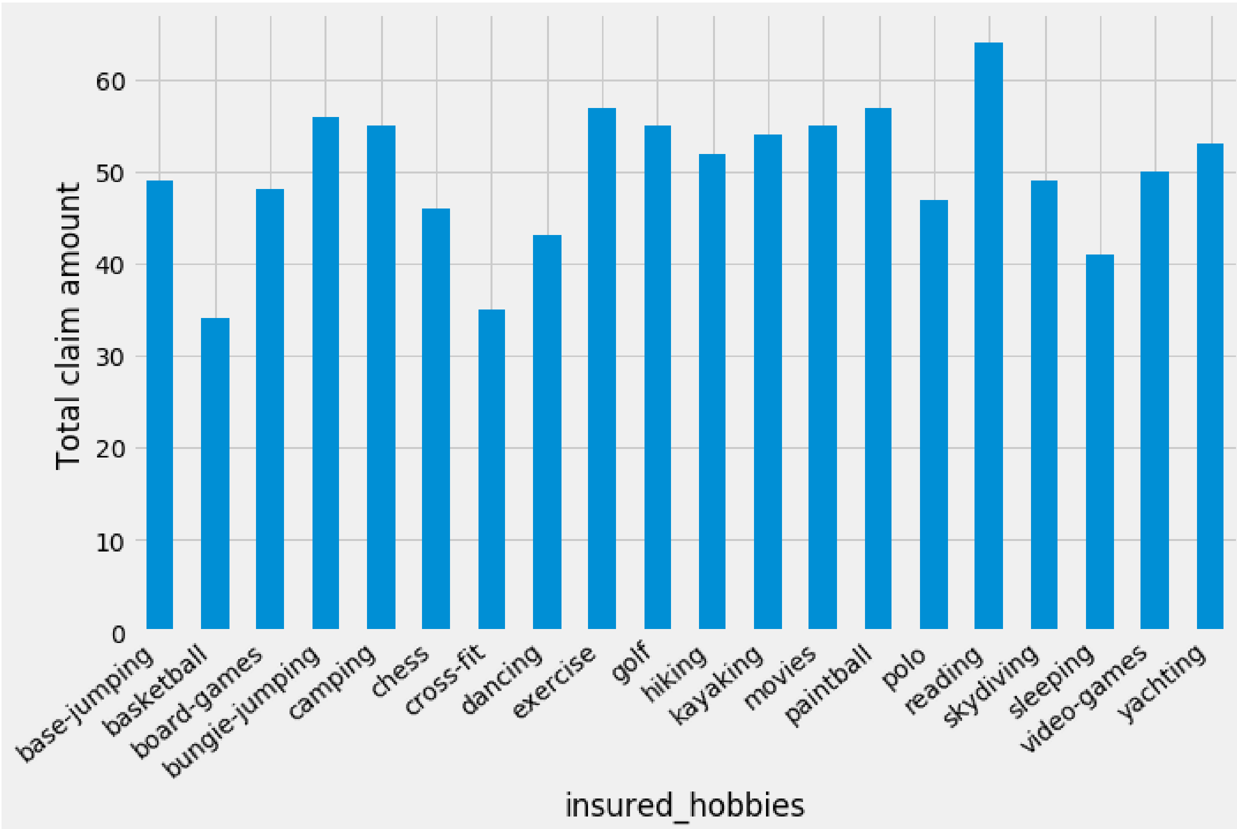

```
fig = plt.figure(figsize=(10,6))
ax = sns.countplot(x='insured_hobbies', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.show()
```




```
df["insured_occupation"].value_counts()

machine-op-inspct    93
prof-specialty       85
tech-support         78
exec-managerial      76
sales                76
craft-repair         74
transport-moving     72
priv-house-serv      71
other-service        71
armed-forces         69
adm-clerical         65
protective-serv      63
handlers-cleaners    54
farming-fishing      53
Name: insured_occupation, dtype: int64

plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax= df.groupby('auto_make').vehicle_claim.count().plot.bar(ylim=0)
ax.set_ylabel('Vehicle claim')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.show()
```




```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax=
df.groupby('insured_hobbies').total_claim_amount.count().plot.bar(ylim
=0)
ax.set_ylabel('Total claim amount')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.show()
```



Data Processing

Cleaning up the data and prepare it for machine learning model.

```
df['fraud_reported'].replace(to_replace='Y', value=1, inplace=True)
df['fraud_reported'].replace(to_replace='N',  value=0, inplace=True)

df.head()

   months_as_customer  age  policy_number policy_bind_date
policy_state  \
0                 328   48         521585       2014-10-17
OH
1                 228   42         342868       2006-06-27
IN
2                 134   29         687698       2000-09-06
OH
3                 256   41         227811       1990-05-25
IL
4                 228   44         367455       2014-06-06
IL

  policy_csl  policy_deductable  policy_annual_premium  umbrella_limit
\
0    250/500               1000                1406.91               0
```

```
1    250/500               2000                    1197.22        5000000

2    100/300               2000                    1413.14        5000000

3    250/500               2000                    1415.74        6000000

4   500/1000               1000                    1583.91        6000000

   insured_zip insured_sex insured_education_level insured_occupation  \
0       466132        MALE                       MD       craft-repair

1       468176        MALE                       MD  machine-op-inspct

2       430632      FEMALE                      PhD              sales

3       608117      FEMALE                      PhD       armed-forces

4       610706        MALE                Associate              sales

  insured_hobbies insured_relationship  capital-gains  capital-loss  \
0        sleeping              husband          53300             0
1         reading       other-relative              0             0
2     board-games            own-child          35100             0
3     board-games            unmarried          48900        -62400
4     board-games            unmarried          66000        -46000

  incident_date             incident_type   collision_type
incident_severity  \
0    2015-01-25  Single Vehicle Collision   Side Collision      Major
Damage
1    2015-01-21             Vehicle Theft                ?      Minor
Damage
2    2015-02-22   Multi-vehicle Collision   Rear Collision      Minor
Damage
3    2015-01-10  Single Vehicle Collision  Front Collision      Major
Damage
4    2015-02-17             Vehicle Theft                ?      Minor
Damage

  authorities_contacted incident_state incident_city
incident_location  \
0                Police             SC      Columbus     9935 4th
Drive
1                Police             VA     Riverwood       6608 MLK
Hwy
2                Police             NY      Columbus  7121 Francis
Lane
```

```
3                 Police             OH     Arlington   6956 Maple
Drive
4                   None             NY     Arlington       3041 3rd
Ave

   incident_hour_of_the_day  number_of_vehicles_involved
property_damage  \
0                         5                            1
YES
1                         8                            1
?
2                         7                            3
NO
3                         5                            1
?
4                        20                            1
NO

   bodily_injuries  witnesses police_report_available
total_claim_amount  \
0                1          2                     YES
71610
1                0          0                       ?
5070
2                2          3                      NO
34650
3                1          2                      NO
63400
4                0          1                      NO
6500

   injury_claim  property_claim  vehicle_claim  auto_make
auto_model  \
0          6510           13020          52080       Saab        92x

1           780             780           3510   Mercedes       E400

2          7700            3850          23100      Dodge        RAM

3          6340            6340          50720  Chevrolet      Tahoe

4          1300             650           4550     Accura        RSX


   auto_year  fraud_reported  _c39
0       2004               1   NaN
1       2007               1   NaN
2       2007               0   NaN
3       2014               1   NaN
4       2009               0   NaN
```

```
df[['insured_zip']] = df[['insured_zip']].astype(object)

df.describe()

       months_as_customer          age  policy_number
policy_deductable  \
count         1000.000000  1000.000000    1000.000000
1000.000000
mean           203.954000    38.948000  546238.648000
1136.000000
std            115.113174     9.140287  257063.005276
611.864673
min              0.000000    19.000000  100804.000000
500.000000
25%            115.750000    32.000000  335980.250000
500.000000
50%            199.500000    38.000000  533135.000000
1000.000000
75%            276.250000    44.000000  759099.750000
2000.000000
max            479.000000    64.000000  999435.000000
2000.000000

       policy_annual_premium  umbrella_limit  capital-gains   capital-
loss  \
count            1000.000000    1.000000e+03    1000.000000
1000.000000
mean             1256.406150    1.101000e+06   25126.100000  -
26793.700000
std               244.167395    2.297407e+06   27872.187708
28104.096686
min               433.330000   -1.000000e+06       0.000000 -
111100.000000
25%              1089.607500    0.000000e+00       0.000000  -
51500.000000
50%              1257.200000    0.000000e+00       0.000000  -
23250.000000
75%              1415.695000    0.000000e+00   51025.000000
0.000000
max              2047.590000    1.000000e+07  100500.000000
0.000000

       incident_hour_of_the_day  number_of_vehicles_involved
bodily_injuries  \
count               1000.000000                   1000.00000
1000.000000
mean                  11.644000                      1.83900
0.992000
std                    6.951373                      1.01888
0.820127
```

```
min                       0.000000                      1.00000
0.000000
25%                       6.000000                      1.00000
0.000000
50%                      12.000000                      1.00000
1.000000
75%                      17.000000                      3.00000
2.000000
max                      23.000000                      4.00000
2.000000

         witnesses  total_claim_amount  injury_claim
property_claim  \
count  1000.000000          1000.00000   1000.000000     1000.000000

mean      1.487000         52761.94000   7433.420000     7399.570000

std       1.111335         26401.53319   4880.951853     4824.726179

min       0.000000           100.00000      0.000000        0.000000

25%       1.000000         41812.50000   4295.000000     4445.000000

50%       1.000000         58055.00000   6775.000000     6750.000000

75%       2.000000         70592.50000  11305.000000    10885.000000

max       3.000000        114920.00000  21450.000000    23670.000000


       vehicle_claim    auto_year  fraud_reported  _c39
count    1000.000000  1000.000000     1000.000000   0.0
mean    37928.950000  2005.103000        0.247000   NaN
std     18886.252893     6.015861        0.431483   NaN
min        70.000000  1995.000000        0.000000   NaN
25%     30292.500000  2000.000000        0.000000   NaN
50%     42100.000000  2005.000000        0.000000   NaN
75%     50822.500000  2010.000000        0.000000   NaN
max     79560.000000  2015.000000        1.000000   NaN
```

Some variables such as 'policy_bind_date', 'incident_date', 'incident_location' and 'insured_zip' contain very high number of level. We will remove these columns for our purposes.

```
df.auto_year.value_counts()  # check the spread of years to decide on further action.

1995    56
1999    55
2005    54
2011    53
2006    53
```

```
2007    52
2003    51
2010    50
2009    50
2013    49
2002    49
2015    47
1997    46
2012    46
2008    45
2014    44
2001    42
2000    42
1998    40
2004    39
1996    37
Name: auto_year, dtype: int64
```

auto_year has 21 levels, and the number of records for each of the levels are quite significant considering datasize is not so large. We will do some feature engineering using this variable considering, the year of manufacturing of automobile indicates the age of the vehicle and may contain valuable information for insurance premium or fraud is concerned.

```
df['vehicle_age'] = 2018 - df['auto_year'] # Deriving the age of the
vehicle based on the year value
df['vehicle_age'].head(10)

0    14
1    11
2    11
3     4
4     9
5    15
6     6
7     3
8     6
9    22
Name: vehicle_age, dtype: int64

bins = [-1, 3, 6, 9, 12, 17, 20, 24]  # Factorize according to the
time period of the day.
names = ["past_midnight", "early_morning", "morning", 'fore-noon',
'afternoon', 'evening', 'night']
df['incident_period_of_day'] = pd.cut(df.incident_hour_of_the_day,
bins, labels=names).astype(object)
df[['incident_hour_of_the_day', 'incident_period_of_day']].head(20)

   incident_hour_of_the_day incident_period_of_day
0                         5          early_morning
1                         8                morning
2                         7                morning
```

```
3                          5       early_morning
4                         20             evening
5                         19             evening
6                          0       past_midnight
7                         23               night
8                         21               night
9                         14           afternoon
10                        22               night
11                        21               night
12                         9             morning
13                         5       early_morning
14                        12           fore-noon
15                        12           fore-noon
16                         0       past_midnight
17                         9             morning
18                        19             evening
19                         8             morning
```

```
# Check on categorical variables:
df.select_dtypes(include=['object']).columns  # checking categorcial
columns
```

```
Index(['policy_bind_date', 'policy_state', 'policy_csl',
'insured_zip',
       'insured_sex', 'insured_education_level', 'insured_occupation',
       'insured_hobbies', 'insured_relationship', 'incident_date',
       'incident_type', 'collision_type', 'incident_severity',
       'authorities_contacted', 'incident_state', 'incident_city',
       'incident_location', 'property_damage',
'police_report_available',
       'auto_make', 'auto_model', 'incident_period_of_day'],
      dtype='object')
```

```
# dropping unimportant columns

df = df.drop(columns = [
    'policy_number',
    'insured_zip',
    'policy_bind_date',
    'incident_date',
    'incident_location',
    '_c39',
    'auto_year',
    'incident_hour_of_the_day'])

df.head(2)
```

```
   months_as_customer  age policy_state policy_csl  policy_deductable
\
0                 328   48           OH    250/500               1000
```

```
1                 228   42           IN    250/500              2000


   policy_annual_premium  umbrella_limit insured_sex
insured_education_level  \
0                1406.91               0        MALE
MD
1                1197.22         5000000        MALE
MD

  insured_occupation insured_hobbies insured_relationship  capital-
gains  \
0       craft-repair        sleeping              husband
53300
1  machine-op-inspct         reading       other-relative
0

   capital-loss             incident_type  collision_type
incident_severity  \
0             0  Single Vehicle Collision  Side Collision      Major
Damage
1             0             Vehicle Theft               ?      Minor
Damage

  authorities_contacted incident_state incident_city  \
0                Police             SC      Columbus
1                Police             VA     Riverwood

   number_of_vehicles_involved property_damage  bodily_injuries
witnesses  \
0                            1             YES                1
2
1                            1               ?                0
0

  police_report_available  total_claim_amount  injury_claim
property_claim  \
0                     YES               71610          6510
13020
1                       ?                5070           780
780

   vehicle_claim auto_make auto_model  fraud_reported  vehicle_age  \
0          52080      Saab        92x               1           14
1           3510  Mercedes       E400               1           11

  incident_period_of_day
0          early_morning
1                morning
```

```
# identify variables with '?' values
unknowns = {}
for i in list(df.columns):
    if (df[i]).dtype == object:
        j = np.sum(df[i] == "?")
        unknowns[i] = j
unknowns = pd.DataFrame.from_dict(unknowns, orient = 'index')
print(unknowns)
```

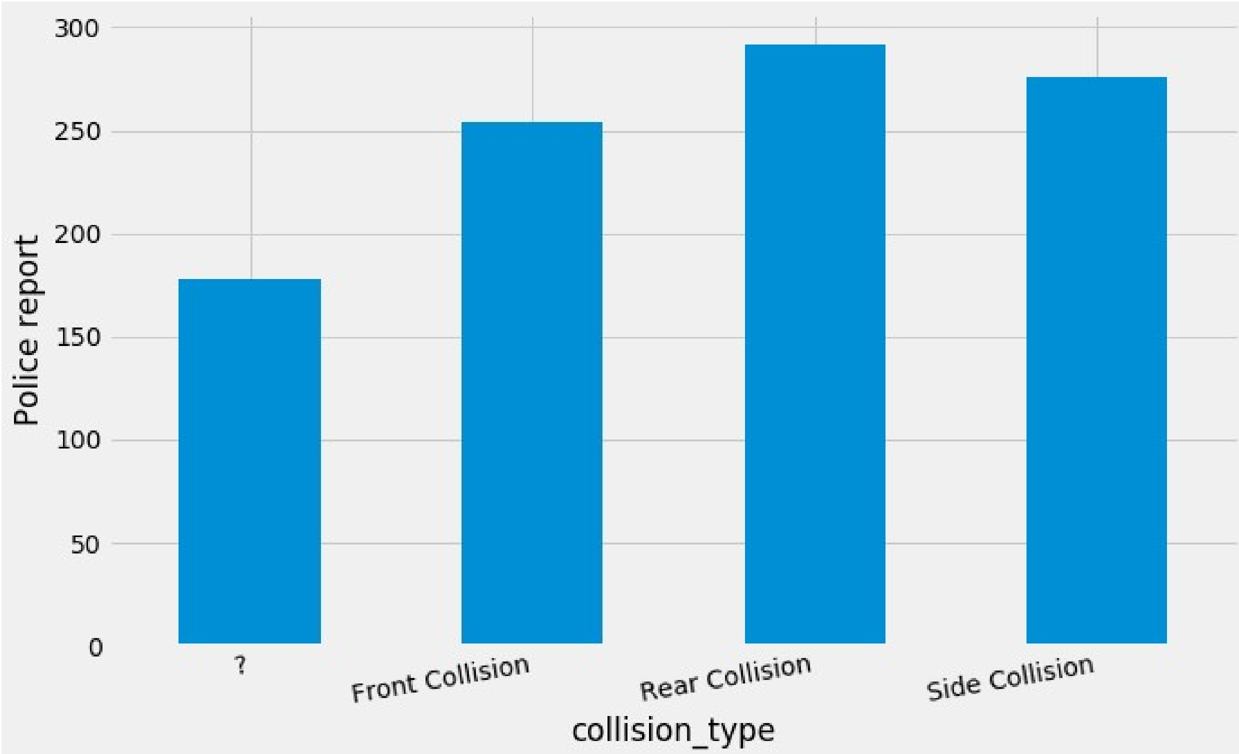
```
                               0
policy_state                   0
policy_csl                     0
insured_sex                    0
insured_education_level        0
insured_occupation             0
insured_hobbies                0
insured_relationship           0
incident_type                  0
collision_type               178
incident_severity              0
authorities_contacted          0
incident_state                 0
incident_city                  0
property_damage              360
police_report_available      343
auto_make                      0
auto_model                     0
incident_period_of_day         0
```

collision_type, property_damage, police_report_available contain many missing values. So, first isolate these variables, inspect these individually for spread of category values.

```
df.collision_type.value_counts()
```

```
Rear Collision     292
Side Collision     276
Front Collision    254
?                  178
Name: collision_type, dtype: int64
```

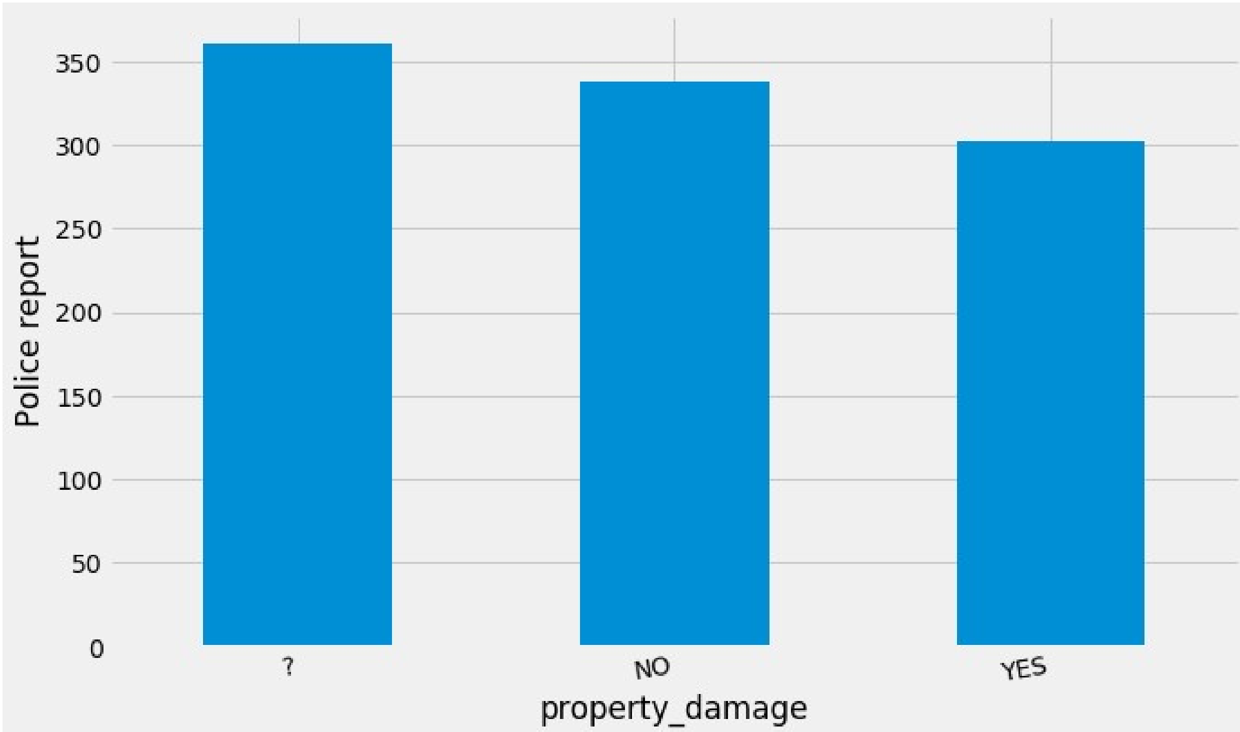
```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax=
df.groupby('collision_type').police_report_available.count().plot.bar(
ylim=0)
ax.set_ylabel('Police report')
ax.set_xticklabels(ax.get_xticklabels(), rotation=10, ha="right")
plt.show()
```




```
df.property_damage.value_counts()
```

```
?      360
NO     338
YES    302
Name: property_damage, dtype: int64
```

```
plt.style.use('fivethirtyeight')
fig = plt.figure(figsize=(10,6))
ax=
df.groupby('property_damage').police_report_available.count().plot.bar
(ylim=0)
ax.set_ylabel('Police report')
ax.set_xticklabels(ax.get_xticklabels(), rotation=10, ha="right")
plt.show()
```




```
df.police_report_available.value_counts()

?      343
NO     343
YES    314
Name: police_report_available, dtype: int64

df.columns

Index(['months_as_customer', 'age', 'policy_state', 'policy_csl',
       'policy_deductable', 'policy_annual_premium', 'umbrella_limit',
       'insured_sex', 'insured_education_level', 'insured_occupation',
       'insured_hobbies', 'insured_relationship', 'capital-gains',
       'capital-loss', 'incident_type', 'collision_type',
'incident_severity',
       'authorities_contacted', 'incident_state', 'incident_city',
       'number_of_vehicles_involved', 'property_damage',
'bodily_injuries',
       'witnesses', 'police_report_available', 'total_claim_amount',
       'injury_claim', 'property_claim', 'vehicle_claim', 'auto_make',
       'auto_model', 'fraud_reported', 'vehicle_age',
       'incident_period_of_day'],
      dtype='object')

df._get_numeric_data().head()  # Checking numeric columns

   months_as_customer  age  policy_deductable
policy_annual_premium  \
```

```
0                  328   48               1000                    1406.91

1                  228   42               2000                    1197.22

2                  134   29               2000                    1413.14

3                  256   41               2000                    1415.74

4                  228   44               1000                    1583.91


   umbrella_limit  capital-gains  capital-loss
number_of_vehicles_involved  \
0               0          53300             0
1
1         5000000              0             0
1
2         5000000          35100             0
3
3         6000000          48900        -62400
1
4         6000000          66000        -46000
1

   bodily_injuries  witnesses  total_claim_amount  injury_claim  \
0                1          2               71610          6510
1                0          0                5070           780
2                2          3               34650          7700
3                1          2               63400          6340
4                0          1                6500          1300

   property_claim  vehicle_claim  fraud_reported  vehicle_age
0           13020          52080               1           14
1             780           3510               1           11
2            3850          23100               0           11
3            6340          50720               1            4
4             650           4550               0            9
```

```
df._get_numeric_data().columns
```

```
Index(['months_as_customer', 'age', 'policy_deductable',
       'policy_annual_premium', 'umbrella_limit', 'capital-gains',
       'capital-loss', 'number_of_vehicles_involved',
'bodily_injuries',
       'witnesses', 'total_claim_amount', 'injury_claim',
'property_claim',
       'vehicle_claim', 'fraud_reported', 'vehicle_age'],
      dtype='object')
```

```
df.select_dtypes(include=['object']).columns  # checking categorcial
columns
```

```
Index(['policy_state', 'policy_csl', 'insured_sex',
'insured_education_level',
       'insured_occupation', 'insured_hobbies',
'insured_relationship',
       'incident_type', 'collision_type', 'incident_severity',
       'authorities_contacted', 'incident_state', 'incident_city',
       'property_damage', 'police_report_available', 'auto_make',
'auto_model',
       'incident_period_of_day'],
      dtype='object')
```

Applying one-hot encoding to convert all categorical variables except out target variables

'collision_type', 'property_damage', 'police_report_available', 'fraud_reported'

```
dummies = pd.get_dummies(df[[
    'policy_state',
    'policy_csl',
    'insured_sex',
    'insured_education_level',
    'insured_occupation',
    'insured_hobbies',
    'insured_relationship',
    'incident_type',
    'incident_severity',
    'authorities_contacted',
    'incident_state',
    'incident_city',
    'auto_make',
    'auto_model',
    'incident_period_of_day']])

dummies = dummies.join(df[[
    'collision_type',
    'property_damage',
    'police_report_available',
    "fraud_reported"]])

dummies.head()
```

```
   policy_state_IL  policy_state_IN  policy_state_OH
policy_csl_100/300  \
0                0                0                1
0
1                0                1                0
0
2                0                0                1
1
```

```
3                1                0                0
0
4                1                0                0
0

   policy_csl_250/500  policy_csl_500/1000  insured_sex_FEMALE  \
0                   1                    0                   0
1                   1                    0                   0
2                   0                    0                   1
3                   1                    0                   1
4                   0                    1                   0

   insured_sex_MALE  insured_education_level_Associate  \
0                 1                                  0
1                 1                                  0
2                 0                                  0
3                 0                                  0
4                 1                                  1

   insured_education_level_College  insured_education_level_High
School  \
0                                0
0
1                                0
0
2                                0
0
3                                0
0
4                                0
0

   insured_education_level_JD  insured_education_level_MD  \
0                           0                           1
1                           0                           1
2                           0                           0
3                           0                           0
4                           0                           0

   insured_education_level_Masters  insured_education_level_PhD  \
0                                0                            0
1                                0                            0
2                                0                            1
3                                0                            1
4                                0                            0

   insured_occupation_adm-clerical  insured_occupation_armed-forces  \
0                                0                                0
1                                0                                0
2                                0                                0
```

```
3                                0                                1
4                                0                                0

   insured_occupation_craft-repair  insured_occupation_exec-managerial
\
0                                1                                   0

1                                0                                   0

2                                0                                   0

3                                0                                   0

4                                0                                   0


   insured_occupation_farming-fishing  insured_occupation_handlers-
cleaners  \
0                                   0
0
1                                   0
0
2                                   0
0
3                                   0
0
4                                   0
0

   insured_occupation_machine-op-inspct  insured_occupation_other-
service  \
0                                     0
0
1                                     1
0
2                                     0
0
3                                     0
0
4                                     0
0

   insured_occupation_priv-house-serv  insured_occupation_prof-
specialty  \
0                                   0
0
1                                   0
0
2                                   0
0
```

```
3                                    0
0
4                                    0
0

   insured_occupation_protective-serv  insured_occupation_sales  \
0                                    0                         0
1                                    0                         0
2                                    0                         1
3                                    0                         0
4                                    0                         1

   insured_occupation_tech-support  insured_occupation_transport-
moving  \
0                                 0
0
1                                 0
0
2                                 0
0
3                                 0
0
4                                 0
0

   insured_hobbies_base-jumping  insured_hobbies_basketball  \
0                             0                           0
1                             0                           0
2                             0                           0
3                             0                           0
4                             0                           0

   insured_hobbies_board-games  insured_hobbies_bungie-jumping  \
0                            0                               0
1                            0                               0
2                            1                               0
3                            1                               0
4                            1                               0

   insured_hobbies_camping  insured_hobbies_chess
insured_hobbies_cross-fit  \
0                        0                      0
0
1                        0                      0
0
2                        0                      0
0
3                        0                      0
0
4                        0                      0
```

```
0

   insured_hobbies_dancing  insured_hobbies_exercise
insured_hobbies_golf  \
0                        0                         0
0
1                        0                         0
0
2                        0                         0
0
3                        0                         0
0
4                        0                         0
0

   insured_hobbies_hiking  insured_hobbies_kayaking
insured_hobbies_movies  \
0                       0                         0
0
1                       0                         0
0
2                       0                         0
0
3                       0                         0
0
4                       0                         0
0

   insured_hobbies_paintball  insured_hobbies_polo
insured_hobbies_reading  \
0                          0                     0
0
1                          0                     0
1
2                          0                     0
0
3                          0                     0
0
4                          0                     0
0

   insured_hobbies_skydiving  insured_hobbies_sleeping  \
0                          0                         1
1                          0                         0
2                          0                         0
3                          0                         0
4                          0                         0

   insured_hobbies_video-games  insured_hobbies_yachting  \
0                            0                         0
```

```
1                            0                          0
2                            0                          0
3                            0                          0
4                            0                          0

   insured_relationship_husband  insured_relationship_not-in-family  \
0                             1                                   0
1                             0                                   0
2                             0                                   0
3                             0                                   0
4                             0                                   0

   insured_relationship_other-relative  insured_relationship_own-child
\
0                                    0                               0

1                                    1                               0

2                                    0                               1

3                                    0                               0

4                                    0                               0


   insured_relationship_unmarried  insured_relationship_wife  \
0                               0                          0
1                               0                          0
2                               0                          0
3                               1                          0
4                               1                          0

   incident_type_Multi-vehicle Collision  incident_type_Parked Car  \
0                                      0                         0
1                                      0                         0
2                                      1                         0
3                                      0                         0
4                                      0                         0

   incident_type_Single Vehicle Collision  incident_type_Vehicle Theft
\
0                                       1                            0

1                                       0                            1

2                                       0                            0

3                                       1                            0

4                                       0                            1
```

```
   incident_severity_Major Damage  incident_severity_Minor Damage  \
0                               1                               0
1                               0                               1
2                               0                               1
3                               1                               0
4                               0                               1

   incident_severity_Total Loss  incident_severity_Trivial Damage  \
0                             0                                 0
1                             0                                 0
2                             0                                 0
3                             0                                 0
4                             0                                 0

   authorities_contacted_Ambulance  authorities_contacted_Fire  \
0                                0                           0
1                                0                           0
2                                0                           0
3                                0                           0
4                                0                           0

   authorities_contacted_None  authorities_contacted_Other  \
0                           0                            0
1                           0                            0
2                           0                            0
3                           0                            0
4                           1                            0

   authorities_contacted_Police  incident_state_NC  incident_state_NY
\
0                             1                  0                  0

1                             1                  0                  0

2                             1                  0                  1

3                             1                  0                  0

4                             0                  0                  1


   incident_state_OH  incident_state_PA  incident_state_SC
incident_state_VA  \
0                  0                  0                  1
0
1                  0                  0                  0
1
2                  0                  0                  0
0
```

```
3                  1                  0                  0
0
4                  0                  0                  0
0

   incident_state_WV  incident_city_Arlington  incident_city_Columbus
\
0                  0                        0                       1

1                  0                        0                       0

2                  0                        0                       1

3                  0                        1                       0

4                  0                        1                       0


   incident_city_Hillsdale  incident_city_Northbend
incident_city_Northbrook  \
0                        0                        0
0
1                        0                        0
0
2                        0                        0
0
3                        0                        0
0
4                        0                        0
0

   incident_city_Riverwood  incident_city_Springfield
auto_make_Accura  \
0                        0                          0
0
1                        1                          0
0
2                        0                          0
0
3                        0                          0
0
4                        0                          0
1

   auto_make_Audi  auto_make_BMW  auto_make_Chevrolet  auto_make_Dodge
\
0               0              0                    0                0

1               0              0                    0                0
```

```
2              0             0                   0               1

3              0             0                   1               0

4              0             0                   0               0


   auto_make_Ford  auto_make_Honda  auto_make_Jeep  auto_make_Mercedes
\
0              0                0               0                   0

1              0                0               0                   1

2              0                0               0                   0

3              0                0               0                   0

4              0                0               0                   0


   auto_make_Nissan  auto_make_Saab  auto_make_Suburu
auto_make_Toyota  \
0                0               1                 0
0
1                0               0                 0
0
2                0               0                 0
0
3                0               0                 0
0
4                0               0                 0
0

   auto_make_Volkswagen  auto_model_3 Series  auto_model_92x
auto_model_93  \
0                     0                    0               1
0
1                     0                    0               0
0
2                     0                    0               0
0
3                     0                    0               0
0
4                     0                    0               0
0

   auto_model_95  auto_model_A3  auto_model_A5  auto_model_Accord  \
0              0              0              0                  0
1              0              0              0                  0
2              0              0              0                  0
```

```
3               0               0               0                   0
4               0               0               0                   0

   auto_model_C300  auto_model_CRV  auto_model_Camry  auto_model_Civic
\
0                0               0                 0                 0

1                0               0                 0                 0

2                0               0                 0                 0

3                0               0                 0                 0

4                0               0                 0                 0


   auto_model_Corolla  auto_model_E400  auto_model_Escape
auto_model_F150  \
0                   0                0                  0
0
1                   0                1                  0
0
2                   0                0                  0
0
3                   0                0                  0
0
4                   0                0                  0
0

   auto_model_Forrestor  auto_model_Fusion  auto_model_Grand Cherokee
\
0                     0                  0                          0

1                     0                  0                          0

2                     0                  0                          0

3                     0                  0                          0

4                     0                  0                          0


   auto_model_Highlander  auto_model_Impreza  auto_model_Jetta  \
0                      0                   0                 0
1                      0                   0                 0
2                      0                   0                 0
3                      0                   0                 0
4                      0                   0                 0

   auto_model_Legacy  auto_model_M5  auto_model_MDX  auto_model_ML350
\
```

```
0                  0              0               0                 0

1                  0              0               0                 0

2                  0              0               0                 0

3                  0              0               0                 0

4                  0              0               0                 0


   auto_model_Malibu  auto_model_Maxima  auto_model_Neon
auto_model_Passat  \
0                  0                  0                0
0
1                  0                  0                0
0
2                  0                  0                0
0
3                  0                  0                0
0
4                  0                  0                0
0

   auto_model_Pathfinder  auto_model_RAM  auto_model_RSX  \
0                      0               0               0
1                      0               0               0
2                      0               1               0
3                      0               0               0
4                      0               0               1

   auto_model_Silverado  auto_model_TL  auto_model_Tahoe
auto_model_Ultima  \
0                     0              0                 0
0
1                     0              0                 0
0
2                     0              0                 0
0
3                     0              0                 1
0
4                     0              0                 0
0

   auto_model_Wrangler  auto_model_X5  auto_model_X6  \
0                    0              0              0
1                    0              0              0
2                    0              0              0
3                    0              0              0
4                    0              0              0
```

```
   incident_period_of_day_afternoon
incident_period_of_day_early_morning  \
0                                 0
1
1                                 0
0
2                                 0
0
3                                 0
1
4                                 0
0

   incident_period_of_day_evening  incident_period_of_day_fore-noon  \
0                               0                                 0
1                               0                                 0
2                               0                                 0
3                               0                                 0
4                               1                                 0

   incident_period_of_day_morning  incident_period_of_day_night  \
0                               0                             0
1                               1                             0
2                               1                             0
3                               0                             0
4                               0                             0

   incident_period_of_day_past_midnight   collision_type
property_damage  \
0                                     0   Side Collision
YES
1                                     0                ?
?
2                                     0   Rear Collision
NO
3                                     0  Front Collision
?
4                                     0                ?
NO

  police_report_available  fraud_reported
0                     YES               1
1                       ?               1
2                      NO               0
3                      NO               1
4                      NO               0
```

```
X = dummies.iloc[:, 0:-1]
y = dummies.iloc[:, -1]
```

```
len(X.columns)
```

```
145
```

```
X.head(2)
```

```
   policy_state_IL  policy_state_IN  policy_state_OH
policy_csl_100/300  \
0                0                0                1
0
1                0                1                0
0

   policy_csl_250/500  policy_csl_500/1000  insured_sex_FEMALE  \
0                   1                    0                   0
1                   1                    0                   0

   insured_sex_MALE  insured_education_level_Associate  \
0                 1                                  0
1                 1                                  0

   insured_education_level_College  insured_education_level_High
School  \
0                                0
0
1                                0
0

   insured_education_level_JD  insured_education_level_MD  \
0                           0                           1
1                           0                           1

   insured_education_level_Masters  insured_education_level_PhD  \
0                                0                            0
1                                0                            0

   insured_occupation_adm-clerical  insured_occupation_armed-forces  \
0                                0                                0
1                                0                                0

   insured_occupation_craft-repair  insured_occupation_exec-managerial
\
0                                1                                   0

1                                0                                   0


   insured_occupation_farming-fishing  insured_occupation_handlers-
cleaners  \
0                                   0
0
```

```
1                                    0
0

   insured_occupation_machine-op-inspct  insured_occupation_other-
service  \
0                                      0
0
1                                      1
0

   insured_occupation_priv-house-serv  insured_occupation_prof-
specialty  \
0                                    0
0
1                                    0
0

   insured_occupation_protective-serv  insured_occupation_sales  \
0                                    0                         0
1                                    0                         0

   insured_occupation_tech-support  insured_occupation_transport-
moving  \
0                                 0
0
1                                 0
0

   insured_hobbies_base-jumping  insured_hobbies_basketball  \
0                              0                           0
1                              0                           0

   insured_hobbies_board-games  insured_hobbies_bungie-jumping  \
0                             0                               0
1                             0                               0

   insured_hobbies_camping  insured_hobbies_chess
insured_hobbies_cross-fit  \
0                         0                      0
0
1                         0                      0
0

   insured_hobbies_dancing  insured_hobbies_exercise
insured_hobbies_golf  \
0                         0                         0
0
1                         0                         0
0
```

```
   insured_hobbies_hiking  insured_hobbies_kayaking
insured_hobbies_movies  \
0                       0                          0
0
1                       0                          0
0

   insured_hobbies_paintball  insured_hobbies_polo
insured_hobbies_reading  \
0                          0                      0
0
1                          0                      0
1

   insured_hobbies_skydiving  insured_hobbies_sleeping  \
0                          0                         1
1                          0                         0

   insured_hobbies_video-games  insured_hobbies_yachting  \
0                            0                         0
1                            0                         0

   insured_relationship_husband  insured_relationship_not-in-family  \
0                             1                                   0
1                             0                                   0

   insured_relationship_other-relative  insured_relationship_own-child
\
0                                    0                               0

1                                    1                               0


   insured_relationship_unmarried  insured_relationship_wife  \
0                               0                          0
1                               0                          0

   incident_type_Multi-vehicle Collision  incident_type_Parked Car  \
0                                      0                         0
1                                      0                         0

   incident_type_Single Vehicle Collision  incident_type_Vehicle Theft
\
0                                       1                            0

1                                       0                            1


   incident_severity_Major Damage  incident_severity_Minor Damage  \
0                               1                               0
```

```
1                              0                               1

   incident_severity_Total Loss  incident_severity_Trivial Damage  \
0                             0                                 0
1                             0                                 0

   authorities_contacted_Ambulance  authorities_contacted_Fire  \
0                                0                           0
1                                0                           0

   authorities_contacted_None  authorities_contacted_Other  \
0                           0                            0
1                           0                            0

   authorities_contacted_Police  incident_state_NC  incident_state_NY
\
0                             1                  0                  0

1                             1                  0                  0


   incident_state_OH  incident_state_PA  incident_state_SC
incident_state_VA  \
0                  0                  0                  1
0
1                  0                  0                  0
1

   incident_state_WV  incident_city_Arlington  incident_city_Columbus
\
0                  0                        0                       1

1                  0                        0                       0


   incident_city_Hillsdale  incident_city_Northbend
incident_city_Northbrook  \
0                        0                        0
0
1                        0                        0
0

   incident_city_Riverwood  incident_city_Springfield
auto_make_Accura  \
0                        0                          0
0
1                        1                          0
0

   auto_make_Audi  auto_make_BMW  auto_make_Chevrolet  auto_make_Dodge
```

```
\
0                0              0                    0                0

1                0              0                    0                0


   auto_make_Ford  auto_make_Honda  auto_make_Jeep  auto_make_Mercedes
\
0                0                0               0                   0

1                0                0               0                   1


   auto_make_Nissan  auto_make_Saab  auto_make_Suburu
auto_make_Toyota  \
0                 0               1                 0
0
1                 0               0                 0
0

   auto_make_Volkswagen  auto_model_3 Series  auto_model_92x
auto_model_93  \
0                     0                    0               1
0
1                     0                    0               0
0

   auto_model_95  auto_model_A3  auto_model_A5  auto_model_Accord  \
0              0              0              0                  0
1              0              0              0                  0

   auto_model_C300  auto_model_CRV  auto_model_Camry  auto_model_Civic
\
0                0               0                 0                 0

1                0               0                 0                 0


   auto_model_Corolla  auto_model_E400  auto_model_Escape
auto_model_F150  \
0                   0                0                  0
0
1                   0                1                  0
0

   auto_model_Forrestor  auto_model_Fusion  auto_model_Grand Cherokee
\
0                     0                  0                          0

1                     0                  0                          0
```

```
   auto_model_Highlander  auto_model_Impreza  auto_model_Jetta  \
0                      0                   0                 0
1                      0                   0                 0

   auto_model_Legacy  auto_model_M5  auto_model_MDX  auto_model_ML350
\
0                  0              0               0                 0

1                  0              0               0                 0


   auto_model_Malibu  auto_model_Maxima  auto_model_Neon
auto_model_Passat  \
0                  0                  0                0
0
1                  0                  0                0
0

   auto_model_Pathfinder  auto_model_RAM  auto_model_RSX  \
0                      0               0               0
1                      0               0               0

   auto_model_Silverado  auto_model_TL  auto_model_Tahoe
auto_model_Ultima  \
0                     0              0                 0
0
1                     0              0                 0
0

   auto_model_Wrangler  auto_model_X5  auto_model_X6  \
0                    0              0              0
1                    0              0              0

   incident_period_of_day_afternoon
incident_period_of_day_early_morning  \
0                                 0
1
1                                 0
0

   incident_period_of_day_evening  incident_period_of_day_fore-noon  \
0                               0                                 0
1                               0                                 0

   incident_period_of_day_morning  incident_period_of_day_night  \
0                               0                             0
1                               1                             0

   incident_period_of_day_past_midnight  collision_type
```

```
property_damage  \
0                                     0  Side Collision
YES
1
0                ?               ?

  police_report_available
0                      YES
1                        ?
```

```
y.head()
```

```
0    1
1    1
2    0
3    1
4    0
Name: fraud_reported, dtype: int64
```

Label encoding

```
from sklearn.preprocessing import LabelEncoder
X['collision_en'] =
LabelEncoder().fit_transform(dummies['collision_type'])
X[['collision_type', 'collision_en']]
```

```
      collision_type  collision_en
0     Side Collision             3
1                  ?             0
2     Rear Collision             2
3    Front Collision             1
4                  ?             0
5     Rear Collision             2
6    Front Collision             1
7    Front Collision             1
8    Front Collision             1
9     Rear Collision             2
10   Front Collision             1
11   Front Collision             1
12    Rear Collision             2
13                 ?             0
14    Rear Collision             2
15    Side Collision             3
16    Rear Collision             2
17    Side Collision             3
18    Side Collision             3
19    Side Collision             3
20    Rear Collision             2
21    Side Collision             3
22    Rear Collision             2
```

```
23   Front Collision              1
24    Rear Collision              2
25    Rear Collision              2
26                 ?              0
27                 ?              0
28    Side Collision              3
29    Rear Collision              2
30    Side Collision              3
31    Side Collision              3
32   Front Collision              1
33   Front Collision              1
34    Side Collision              3
35   Front Collision              1
36    Rear Collision              2
37                 ?              0
38    Rear Collision              2
39   Front Collision              1
40    Rear Collision              2
41    Side Collision              3
42    Side Collision              3
43    Rear Collision              2
44   Front Collision              1
45    Rear Collision              2
46    Rear Collision              2
47   Front Collision              1
48                 ?              0
49    Rear Collision              2
50   Front Collision              1
51                 ?              0
52                 ?              0
53    Side Collision              3
54                 ?              0
55    Rear Collision              2
56   Front Collision              1
57                 ?              0
58   Front Collision              1
59    Side Collision              3
60    Rear Collision              2
61    Side Collision              3
62    Side Collision              3
63   Front Collision              1
64    Rear Collision              2
65   Front Collision              1
66    Side Collision              3
67    Side Collision              3
68   Front Collision              1
69                 ?              0
70    Side Collision              3
71   Front Collision              1
```

72	Rear Collision	2
73	Rear Collision	2
74	Side Collision	3
75	Front Collision	1
76	Front Collision	1
77	Front Collision	1
78	?	0
79	Rear Collision	2
80	Side Collision	3
81	?	0
82	?	0
83	?	0
84	Side Collision	3
85	Front Collision	1
86	Front Collision	1
87	Side Collision	3
88	?	0
89	Side Collision	3
90	Front Collision	1
91	Side Collision	3
92	?	0
93	Front Collision	1
94	Rear Collision	2
95	?	0
96	Side Collision	3
97	Rear Collision	2
98	?	0
99	?	0
100	Rear Collision	2
101	Side Collision	3
102	Front Collision	1
103	?	0
104	Side Collision	3
105	?	0
106	Rear Collision	2
107	Front Collision	1
108	Front Collision	1
109	Rear Collision	2
110	Rear Collision	2
111	Front Collision	1
112	Rear Collision	2
113	Side Collision	3
114	?	0
115	Side Collision	3
116	Rear Collision	2
117	Side Collision	3
118	Rear Collision	2
119	Rear Collision	2
120	Side Collision	3

121	Front Collision	1
122	Front Collision	1
123	Front Collision	1
124	Rear Collision	2
125	Rear Collision	2
126	Side Collision	3
127	?	0
128	Front Collision	1
129	Front Collision	1
130	Front Collision	1
131	Front Collision	1
132	Front Collision	1
133	Side Collision	3
134	Side Collision	3
135	Rear Collision	2
136	?	0
137	Rear Collision	2
138	Rear Collision	2
139	Front Collision	1
140	Rear Collision	2
141	?	0
142	?	0
143	Rear Collision	2
144	Rear Collision	2
145	Front Collision	1
146	Side Collision	3
147	Front Collision	1
148	Rear Collision	2
149	Side Collision	3
150	Side Collision	3
151	Rear Collision	2
152	Rear Collision	2
153	Rear Collision	2
154	Rear Collision	2
155	Front Collision	1
156	Rear Collision	2
157	?	0
158	Front Collision	1
159	?	0
160	?	0
161	Front Collision	1
162	Side Collision	3
163	Front Collision	1
164	Rear Collision	2
165	Side Collision	3
166	Rear Collision	2
167	Front Collision	1
168	?	0
169	?	0

170	Rear Collision	2
171	Rear Collision	2
172	Rear Collision	2
173	Front Collision	1
174	?	0
175	Rear Collision	2
176	Front Collision	1
177	Rear Collision	2
178	Front Collision	1
179	?	0
180	Side Collision	3
181	Side Collision	3
182	Side Collision	3
183	Rear Collision	2
184	Front Collision	1
185	Front Collision	1
186	Front Collision	1
187	?	0
188	Rear Collision	2
189	Rear Collision	2
190	?	0
191	Side Collision	3
192	Front Collision	1
193	?	0
194	Rear Collision	2
195	Side Collision	3
196	?	0
197	?	0
198	Front Collision	1
199	?	0
200	?	0
201	Rear Collision	2
202	?	0
203	Rear Collision	2
204	Rear Collision	2
205	Rear Collision	2
206	Rear Collision	2
207	Side Collision	3
208	Side Collision	3
209	?	0
210	?	0
211	?	0
212	Side Collision	3
213	Side Collision	3
214	Side Collision	3
215	Side Collision	3
216	Front Collision	1
217	?	0
218	Rear Collision	2

```
219   Side Collision              3
220  Front Collision              1
221  Front Collision              1
222   Side Collision              3
223   Side Collision              3
224   Side Collision              3
225  Front Collision              1
226   Rear Collision              2
227  Front Collision              1
228  Front Collision              1
229   Rear Collision              2
230   Side Collision              3
231   Side Collision              3
232   Side Collision              3
233   Rear Collision              2
234  Front Collision              1
235   Side Collision              3
236  Front Collision              1
237   Rear Collision              2
238   Side Collision              3
239  Front Collision              1
240  Front Collision              1
241  Front Collision              1
242                ?              0
243   Rear Collision              2
244                ?              0
245   Rear Collision              2
246   Side Collision              3
247   Rear Collision              2
248                ?              0
249   Rear Collision              2
..               ...            ...
750                ?              0
751   Side Collision              3
752   Rear Collision              2
753   Rear Collision              2
754   Side Collision              3
755   Side Collision              3
756   Rear Collision              2
757   Rear Collision              2
758  Front Collision              1
759  Front Collision              1
760   Side Collision              3
761   Rear Collision              2
762   Rear Collision              2
763  Front Collision              1
764   Rear Collision              2
765   Rear Collision              2
766   Rear Collision              2
```

767	Front Collision	1
768	Rear Collision	2
769	Side Collision	3
770	Front Collision	1
771	Side Collision	3
772	Front Collision	1
773	Rear Collision	2
774	Side Collision	3
775	?	0
776	Front Collision	1
777	Rear Collision	2
778	Front Collision	1
779	Rear Collision	2
780	Front Collision	1
781	Front Collision	1
782	?	0
783	?	0
784	Front Collision	1
785	Front Collision	1
786	Rear Collision	2
787	Front Collision	1
788	Front Collision	1
789	Rear Collision	2
790	?	0
791	Rear Collision	2
792	Rear Collision	2
793	Side Collision	3
794	Front Collision	1
795	Side Collision	3
796	Rear Collision	2
797	Side Collision	3
798	Side Collision	3
799	?	0
800	Rear Collision	2
801	Rear Collision	2
802	Front Collision	1
803	Rear Collision	2
804	?	0
805	Rear Collision	2
806	Side Collision	3
807	Rear Collision	2
808	Side Collision	3
809	Rear Collision	2
810	Side Collision	3
811	?	0
812	Front Collision	1
813	?	0
814	Front Collision	1
815	Rear Collision	2

816	Side Collision	3
817	Rear Collision	2
818	?	0
819	Side Collision	3
820	?	0
821	Side Collision	3
822	Front Collision	1
823	Front Collision	1
824	Rear Collision	2
825	Front Collision	1
826	Front Collision	1
827	Side Collision	3
828	Rear Collision	2
829	Side Collision	3
830	Side Collision	3
831	Rear Collision	2
832	?	0
833	Rear Collision	2
834	?	0
835	?	0
836	Rear Collision	2
837	?	0
838	Side Collision	3
839	Rear Collision	2
840	?	0
841	Rear Collision	2
842	?	0
843	Front Collision	1
844	Side Collision	3
845	Side Collision	3
846	Side Collision	3
847	Side Collision	3
848	Side Collision	3
849	?	0
850	Side Collision	3
851	Side Collision	3
852	Side Collision	3
853	Front Collision	1
854	Side Collision	3
855	Rear Collision	2
856	Side Collision	3
857	Front Collision	1
858	Side Collision	3
859	Side Collision	3
860	Rear Collision	2
861	Side Collision	3
862	Rear Collision	2
863	Front Collision	1
864	Front Collision	1

865	Rear Collision	2
866	Front Collision	1
867	Rear Collision	2
868	Front Collision	1
869	Rear Collision	2
870	Front Collision	1
871	?	0
872	Side Collision	3
873	Rear Collision	2
874	Side Collision	3
875	Side Collision	3
876	?	0
877	Front Collision	1
878	Rear Collision	2
879	Front Collision	1
880	Front Collision	1
881	Rear Collision	2
882	Rear Collision	2
883	Side Collision	3
884	Front Collision	1
885	Rear Collision	2
886	Front Collision	1
887	?	0
888	Rear Collision	2
889	Side Collision	3
890	Side Collision	3
891	Rear Collision	2
892	?	0
893	?	0
894	?	0
895	Front Collision	1
896	?	0
897	Front Collision	1
898	Rear Collision	2
899	?	0
900	Side Collision	3
901	Side Collision	3
902	Side Collision	3
903	Side Collision	3
904	Side Collision	3
905	Side Collision	3
906	Rear Collision	2
907	Front Collision	1
908	?	0
909	Rear Collision	2
910	Side Collision	3
911	Front Collision	1
912	Rear Collision	2
913	Side Collision	3

914	Front Collision	1
915	Rear Collision	2
916	?	0
917	Side Collision	3
918	Front Collision	1
919	Rear Collision	2
920	Rear Collision	2
921	Rear Collision	2
922	?	0
923	Rear Collision	2
924	Rear Collision	2
925	?	0
926	Front Collision	1
927	Front Collision	1
928	?	0
929	Side Collision	3
930	Rear Collision	2
931	Side Collision	3
932	Side Collision	3
933	Side Collision	3
934	Rear Collision	2
935	Side Collision	3
936	Front Collision	1
937	Front Collision	1
938	Front Collision	1
939	Rear Collision	2
940	?	0
941	Side Collision	3
942	?	0
943	Front Collision	1
944	Side Collision	3
945	Front Collision	1
946	Side Collision	3
947	Side Collision	3
948	Front Collision	1
949	Side Collision	3
950	?	0
951	Side Collision	3
952	Front Collision	1
953	?	0
954	Rear Collision	2
955	Side Collision	3
956	Side Collision	3
957	Rear Collision	2
958	Front Collision	1
959	?	0
960	Rear Collision	2
961	?	0
962	Rear Collision	2

```
963                 ?              0
964                 ?              0
965  Front Collision              1
966   Rear Collision              2
967   Rear Collision              2
968   Side Collision              3
969                 ?              0
970   Side Collision              3
971  Front Collision              1
972   Rear Collision              2
973   Rear Collision              2
974   Side Collision              3
975   Rear Collision              2
976   Side Collision              3
977   Side Collision              3
978  Front Collision              1
979   Rear Collision              2
980   Rear Collision              2
981  Front Collision              1
982  Front Collision              1
983                 ?              0
984   Side Collision              3
985   Side Collision              3
986   Rear Collision              2
987   Side Collision              3
988   Rear Collision              2
989   Rear Collision              2
990   Rear Collision              2
991   Rear Collision              2
992  Front Collision              1
993   Side Collision              3
994                 ?              0
995  Front Collision              1
996   Rear Collision              2
997   Side Collision              3
998   Rear Collision              2
999                 ?              0

[1000 rows x 2 columns]
```

```
X['property_damage'].replace(to_replace='YES', value=1, inplace=True)
X['property_damage'].replace(to_replace='NO', value=0, inplace=True)
X['property_damage'].replace(to_replace='?', value=0, inplace=True)
X['police_report_available'].replace(to_replace='YES', value=1,
inplace=True)
X['police_report_available'].replace(to_replace='NO', value=0,
inplace=True)
X['police_report_available'].replace(to_replace='?', value=0,
inplace=True)
```

```
X.head(10)
```

```
   policy_state_IL  policy_state_IN  policy_state_OH
policy_csl_100/300  \
0                0                0                1
0
1                0                1                0
0
2                0                0                1
1
3                1                0                0
0
4                1                0                0
0
5                0                0                1
0
6                0                1                0
0
7                1                0                0
1
8                1                0                0
1
9                1                0                0
1

   policy_csl_250/500  policy_csl_500/1000  insured_sex_FEMALE  \
0                   1                    0                   0
1                   1                    0                   0
2                   0                    0                   1
3                   1                    0                   1
4                   0                    1                   0
5                   1                    0                   1
6                   1                    0                   0
7                   0                    0                   0
8                   0                    0                   1
9                   0                    0                   0

   insured_sex_MALE  insured_education_level_Associate  \
0                 1                                  0
1                 1                                  0
2                 0                                  0
3                 0                                  0
4                 1                                  1
5                 0                                  0
6                 1                                  0
7                 1                                  1
8                 0                                  0
9                 1                                  0
```

```
   insured_education_level_College  insured_education_level_High
School  \
0                                0
0
1                                0
0
2                                0
0
3                                0
0
4                                0
0
5                                0
0
6                                0
0
7                                0
0
8                                0
0
9                                0
0

   insured_education_level_JD  insured_education_level_MD  \
0                           0                           1
1                           0                           1
2                           0                           0
3                           0                           0
4                           0                           0
5                           0                           0
6                           0                           0
7                           0                           0
8                           0                           0
9                           0                           0

   insured_education_level_Masters  insured_education_level_PhD  \
0                                0                            0
1                                0                            0
2                                0                            1
3                                0                            1
4                                0                            0
5                                0                            1
6                                0                            1
7                                0                            0
8                                0                            1
9                                0                            1

   insured_occupation_adm-clerical  insured_occupation_armed-forces  \
0                                0                                0
1                                0                                0
```

```
2                                          0                                          0
3                                          0                                          1
4                                          0                                          0
5                                          0                                          0
6                                          0                                          0
7                                          0                                          0
8                                          0                                          0
9                                          0                                          0

   insured_occupation_craft-repair  insured_occupation_exec-managerial
\
0                                          1                                              0

1                                          0                                              0

2                                          0                                              0

3                                          0                                              0

4                                          0                                              0

5                                          0                                              0

6                                          0                                              0

7                                          0                                              0

8                                          0                                              0

9                                          0                                              0


   insured_occupation_farming-fishing  insured_occupation_handlers-
cleaners  \
0                                      0
0
1                                      0
0
2                                      0
0
3                                      0
0
4                                      0
0
5                                      0
0
6                                      0
0
7                                      0
0
```

```
8                                    0
0
9                                    0
0

   insured_occupation_machine-op-inspct  insured_occupation_other-
service  \
0                                     0
0
1                                     1
0
2                                     0
0
3                                     0
0
4                                     0
0
5                                     0
0
6                                     0
0
7                                     0
0
8                                     0
1
9                                     0
0

   insured_occupation_priv-house-serv  insured_occupation_prof-
specialty  \
0                                   0
0
1                                   0
0
2                                   0
0
3                                   0
0
4                                   0
0
5                                   0
0
6                                   0
1
7                                   0
0
8                                   0
0
9                                   1
```

```
0

   insured_occupation_protective-serv  insured_occupation_sales  \
0                                   0                         0
1                                   0                         0
2                                   0                         1
3                                   0                         0
4                                   0                         1
5                                   0                         0
6                                   0                         0
7                                   0                         0
8                                   0                         0
9                                   0                         0

   insured_occupation_tech-support  insured_occupation_transport-
moving  \
0                                0
0
1                                0
0
2                                0
0
3                                0
0
4                                0
0
5                                1
0
6                                0
0
7                                1
0
8                                0
0
9                                0
0

   insured_hobbies_base-jumping  insured_hobbies_basketball  \
0                             0                           0
1                             0                           0
2                             0                           0
3                             0                           0
4                             0                           0
5                             0                           0
6                             0                           0
7                             1                           0
8                             0                           0
9                             0                           0

   insured_hobbies_board-games  insured_hobbies_bungie-jumping  \
```

```
0                              0                                 0
1                              0                                 0
2                              1                                 0
3                              1                                 0
4                              1                                 0
5                              0                                 1
6                              1                                 0
7                              0                                 0
8                              0                                 0
9                              0                                 0

   insured_hobbies_camping  insured_hobbies_chess  
insured_hobbies_cross-fit  \
0                        0                      0  
0  
1                        0                      0  
0  
2                        0                      0  
0  
3                        0                      0  
0  
4                        0                      0  
0  
5                        0                      0  
0  
6                        0                      0  
0  
7                        0                      0  
0  
8                        0                      0  
0  
9                        1                      0  
0  

   insured_hobbies_dancing  insured_hobbies_exercise  
insured_hobbies_golf  \
0                        0                         0  
0  
1                        0                         0  
0  
2                        0                         0  
0  
3                        0                         0  
0  
4                        0                         0  
0  
5                        0                         0  
0  
6                        0                         0  
```

```
0
7                        0                          0
0
8                        0                          0
1
9                        0                          0
0

   insured_hobbies_hiking  insured_hobbies_kayaking
insured_hobbies_movies  \
0                       0                          0
0
1                       0                          0
0
2                       0                          0
0
3                       0                          0
0
4                       0                          0
0
5                       0                          0
0
6                       0                          0
0
7                       0                          0
0
8                       0                          0
0
9                       0                          0
0

   insured_hobbies_paintball  insured_hobbies_polo
insured_hobbies_reading  \
0                          0                     0
0
1                          0                     0
1
2                          0                     0
0
3                          0                     0
0
4                          0                     0
0
5                          0                     0
0
6                          0                     0
0
7                          0                     0
0
```

```
8                              0                       0
0
9                              0                       0
0

   insured_hobbies_skydiving  insured_hobbies_sleeping  \
0                          0                         1
1                          0                         0
2                          0                         0
3                          0                         0
4                          0                         0
5                          0                         0
6                          0                         0
7                          0                         0
8                          0                         0
9                          0                         0

   insured_hobbies_video-games  insured_hobbies_yachting  \
0                            0                         0
1                            0                         0
2                            0                         0
3                            0                         0
4                            0                         0
5                            0                         0
6                            0                         0
7                            0                         0
8                            0                         0
9                            0                         0

   insured_relationship_husband  insured_relationship_not-in-family  \
0                             1                                   0
1                             0                                   0
2                             0                                   0
3                             0                                   0
4                             0                                   0
5                             0                                   0
6                             1                                   0
7                             0                                   0
8                             0                                   0
9                             0                                   0

   insured_relationship_other-relative  insured_relationship_own-child
\
0                                    0                               0

1                                    1                               0

2                                    0                               1

3                                    0                               0
```

```
4                                    0                              0
5                                    0                              0
6                                    0                              0
7                                    0                              0
8                                    0                              1
9                                    0                              0

   insured_relationship_unmarried  insured_relationship_wife  \
0                               0                          0
1                               0                          0
2                               0                          0
3                               1                          0
4                               1                          0
5                               1                          0
6                               0                          0
7                               1                          0
8                               0                          0
9                               0                          1

   incident_type_Multi-vehicle Collision  incident_type_Parked Car  \
0                                      0                         0
1                                      0                         0
2                                      1                         0
3                                      0                         0
4                                      0                         0
5                                      1                         0
6                                      1                         0
7                                      1                         0
8                                      0                         0
9                                      0                         0

   incident_type_Single Vehicle Collision  incident_type_Vehicle Theft
\
0                                       1                            0
1                                       0                            1
2                                       0                            0
3                                       1                            0
4                                       0                            1
5                                       0                            0
```

```
6                                       0                             0
7                                       0                             0
8                                       1                             0
9                                       1                             0

   incident_severity_Major Damage  incident_severity_Minor Damage  \
0                               1                               0
1                               0                               1
2                               0                               1
3                               1                               0
4                               0                               1
5                               1                               0
6                               0                               1
7                               0                               0
8                               0                               0
9                               0                               0

   incident_severity_Total Loss  incident_severity_Trivial Damage  \
0                             0                                 0
1                             0                                 0
2                             0                                 0
3                             0                                 0
4                             0                                 0
5                             0                                 0
6                             0                                 0
7                             1                                 0
8                             1                                 0
9                             1                                 0

   authorities_contacted_Ambulance  authorities_contacted_Fire  \
0                                0                           0
1                                0                           0
2                                0                           0
3                                0                           0
4                                0                           0
5                                0                           1
6                                0                           0
7                                0                           0
8                                0                           0
9                                0                           0

   authorities_contacted_None  authorities_contacted_Other  \
0                           0                            0
1                           0                            0
2                           0                            0
```

```
3                              0                              0
4                              1                              0
5                              0                              0
6                              0                              0
7                              0                              0
8                              0                              0
9                              0                              1

   authorities_contacted_Police  incident_state_NC  incident_state_NY
\
0                             1                  0                  0

1                             1                  0                  0

2                             1                  0                  1

3                             1                  0                  0

4                             0                  0                  1

5                             0                  0                  0

6                             1                  0                  1

7                             1                  0                  0

8                             1                  0                  0

9                             0                  1                  0


   incident_state_OH  incident_state_PA  incident_state_SC
incident_state_VA  \
0                  0                  0                  1
0
1                  0                  0                  0
1
2                  0                  0                  0
0
3                  1                  0                  0
0
4                  0                  0                  0
0
5                  0                  0                  1
0
6                  0                  0                  0
0
7                  0                  0                  0
1
8                  0                  0                  0
0
```

```
9                    0                    0                    0
0

   incident_state_WV  incident_city_Arlington  incident_city_Columbus
\
0                  0                        0                       1

1                  0                        0                       0

2                  0                        0                       1

3                  0                        1                       0

4                  0                        1                       0

5                  0                        1                       0

6                  0                        0                       0

7                  0                        0                       1

8                  1                        1                       0

9                  0                        0                       0


   incident_city_Hillsdale  incident_city_Northbend
incident_city_Northbrook  \
0                        0                        0
0
1                        0                        0
0
2                        0                        0
0
3                        0                        0
0
4                        0                        0
0
5                        0                        0
0
6                        0                        0
0
7                        0                        0
0
8                        0                        0
0
9                        1                        0
0

   incident_city_Riverwood  incident_city_Springfield
```

```
auto_make_Accura  \
0                        0                           0
0
1                        1                           0
0
2                        0                           0
0
3                        0                           0
0
4                        0                           0
1
5                        0                           0
0
6                        0                           1
0
7                        0                           0
0
8                        0                           0
0
9                        0                           0
0

   auto_make_Audi  auto_make_BMW  auto_make_Chevrolet  auto_make_Dodge
\
0                0              0                    0                0

1                0              0                    0                0

2                0              0                    0                1

3                0              0                    1                0

4                0              0                    0                0

5                0              0                    0                0

6                0              0                    0                0

7                1              0                    0                0

8                0              0                    0                0

9                0              0                    0                0


   auto_make_Ford  auto_make_Honda  auto_make_Jeep  auto_make_Mercedes
\
0                0                0               0                   0

1                0                0               0                   1
```

```
2                0                0                0                    0

3                0                0                0                    0

4                0                0                0                    0

5                0                0                0                    0

6                0                0                0                    0

7                0                0                0                    0

8                0                0                0                    0

9                0                0                0                    0


   auto_make_Nissan  auto_make_Saab  auto_make_Suburu
auto_make_Toyota  \
0                 0               1                 0
0
1                 0               0                 0
0
2                 0               0                 0
0
3                 0               0                 0
0
4                 0               0                 0
0
5                 0               1                 0
0
6                 1               0                 0
0
7                 0               0                 0
0
8                 0               0                 0
1
9                 0               1                 0
0

   auto_make_Volkswagen  auto_model_3 Series  auto_model_92x
auto_model_93  \
0                     0                    0               1
0
1                     0                    0               0
0
2                     0                    0               0
0
3                     0                    0               0
0
```

```
4                      0                     0                0
0
5                      0                     0                0
0
6                      0                     0                0
0
7                      0                     0                0
0
8                      0                     0                0
0
9                      0                     0                1
0

   auto_model_95  auto_model_A3  auto_model_A5  auto_model_Accord  \
0              0              0              0                  0
1              0              0              0                  0
2              0              0              0                  0
3              0              0              0                  0
4              0              0              0                  0
5              1              0              0                  0
6              0              0              0                  0
7              0              0              1                  0
8              0              0              0                  0
9              0              0              0                  0

   auto_model_C300  auto_model_CRV  auto_model_Camry  auto_model_Civic
\
0                0               0                 0                 0

1                0               0                 0                 0

2                0               0                 0                 0

3                0               0                 0                 0

4                0               0                 0                 0

5                0               0                 0                 0

6                0               0                 0                 0

7                0               0                 0                 0

8                0               0                 1                 0

9                0               0                 0                 0


   auto_model_Corolla  auto_model_E400  auto_model_Escape
auto_model_F150  \
```

```
0                    0                0                  0
0
1                    0                1                  0
0
2                    0                0                  0
0
3                    0                0                  0
0
4                    0                0                  0
0
5                    0                0                  0
0
6                    0                0                  0
0
7                    0                0                  0
0
8                    0                0                  0
0
9                    0                0                  0
0

   auto_model_Forrestor  auto_model_Fusion  auto_model_Grand Cherokee
\
0                     0                  0                          0

1                     0                  0                          0

2                     0                  0                          0

3                     0                  0                          0

4                     0                  0                          0

5                     0                  0                          0

6                     0                  0                          0

7                     0                  0                          0

8                     0                  0                          0

9                     0                  0                          0


   auto_model_Highlander  auto_model_Impreza  auto_model_Jetta  \
0                      0                   0                 0
1                      0                   0                 0
2                      0                   0                 0
3                      0                   0                 0
4                      0                   0                 0
```

```
5                      0                    0                  0
6                      0                    0                  0
7                      0                    0                  0
8                      0                    0                  0
9                      0                    0                  0

   auto_model_Legacy  auto_model_M5  auto_model_MDX  auto_model_ML350 
\
0                  0              0               0                 0

1                  0              0               0                 0

2                  0              0               0                 0

3                  0              0               0                 0

4                  0              0               0                 0

5                  0              0               0                 0

6                  0              0               0                 0

7                  0              0               0                 0

8                  0              0               0                 0

9                  0              0               0                 0


   auto_model_Malibu  auto_model_Maxima  auto_model_Neon 
auto_model_Passat  \
0                  0                  0                0
0
1                  0                  0                0
0
2                  0                  0                0
0
3                  0                  0                0
0
4                  0                  0                0
0
5                  0                  0                0
0
6                  0                  0                0
0
7                  0                  0                0
0
8                  0                  0                0
0
9                  0                  0                0
```

```
0

   auto_model_Pathfinder  auto_model_RAM  auto_model_RSX  \
0                      0               0               0
1                      0               0               0
2                      0               1               0
3                      0               0               0
4                      0               0               1
5                      0               0               0
6                      1               0               0
7                      0               0               0
8                      0               0               0
9                      0               0               0

   auto_model_Silverado  auto_model_TL  auto_model_Tahoe
auto_model_Ultima  \
0                     0              0                 0
0
1                     0              0                 0
0
2                     0              0                 0
0
3                     0              0                 1
0
4                     0              0                 0
0
5                     0              0                 0
0
6                     0              0                 0
0
7                     0              0                 0
0
8                     0              0                 0
0
9                     0              0                 0
0

   auto_model_Wrangler  auto_model_X5  auto_model_X6  \
0                    0              0              0
1                    0              0              0
2                    0              0              0
3                    0              0              0
4                    0              0              0
5                    0              0              0
6                    0              0              0
7                    0              0              0
8                    0              0              0
9                    0              0              0

   incident_period_of_day_afternoon
```

```
incident_period_of_day_early_morning  \
0                                    0
1
1                                    0
0
2                                    0
0
3                                    0
1
4                                    0
0
5                                    0
0
6                                    0
0
7                                    0
0
8                                    0
0
9                                    1
0

   incident_period_of_day_evening  incident_period_of_day_fore-noon  \
0                               0                                 0
1                               0                                 0
2                               0                                 0
3                               0                                 0
4                               1                                 0
5                               1                                 0
6                               0                                 0
7                               0                                 0
8                               0                                 0
9                               0                                 0

   incident_period_of_day_morning  incident_period_of_day_night  \
0                               0                             0
1                               1                             0
2                               1                             0
3                               0                             0
4                               0                             0
5                               0                             0
6                               0                             0
7                               0                             1
8                               0                             1
9                               0                             0

   incident_period_of_day_past_midnight   collision_type
property_damage  \
0                                     0   Side Collision
1
```

```
1                                      0                ?
0
2                                      0   Rear Collision
0
3                                      0  Front Collision
0
4                                      0                ?
0
5                                      0   Rear Collision
0
6                                      1  Front Collision
0
7                                      0  Front Collision
0
8                                      0  Front Collision
0
9                                      0   Rear Collision
0

   police_report_available  collision_en
0                        1             3
1                        0             0
2                        0             2
3                        0             1
4                        0             0
5                        0             2
6                        0             1
7                        1             1
8                        1             1
9                        0             2
```

```
X = X.drop(columns = ['collision_type'])
X.head(2)
```

```
   policy_state_IL  policy_state_IN  policy_state_OH
policy_csl_100/300  \
0                0                0                1
0
1                0                1                0
0

   policy_csl_250/500  policy_csl_500/1000  insured_sex_FEMALE  \
0                   1                    0                   0
1                   1                    0                   0

   insured_sex_MALE  insured_education_level_Associate  \
0                 1                                  0
1                 1                                  0

   insured_education_level_College  insured_education_level_High
```

```
School  \
0                                0
0
1                                0
0

   insured_education_level_JD  insured_education_level_MD  \
0                           0                           1
1                           0                           1

   insured_education_level_Masters  insured_education_level_PhD  \
0                                0                            0
1                                0                            0

   insured_occupation_adm-clerical  insured_occupation_armed-forces  \
0                                0                                0
1                                0                                0

   insured_occupation_craft-repair  insured_occupation_exec-managerial
\
0                                1                                   0

1                                0                                   0


   insured_occupation_farming-fishing  insured_occupation_handlers-
cleaners  \
0                                   0
0
1                                   0
0

   insured_occupation_machine-op-inspct  insured_occupation_other-
service  \
0                                     0
0
1                                     1
0

   insured_occupation_priv-house-serv  insured_occupation_prof-
specialty  \
0                                   0
0
1                                   0
0

   insured_occupation_protective-serv  insured_occupation_sales  \
0                                   0                         0
1                                   0                         0
```

```
   insured_occupation_tech-support  insured_occupation_transport-
moving  \
0                                0
0
1                                0
0

   insured_hobbies_base-jumping  insured_hobbies_basketball  \
0                             0                           0
1                             0                           0

   insured_hobbies_board-games  insured_hobbies_bungie-jumping  \
0                            0                               0
1                            0                               0

   insured_hobbies_camping  insured_hobbies_chess
insured_hobbies_cross-fit  \
0                        0                      0
0
1                        0                      0
0

   insured_hobbies_dancing  insured_hobbies_exercise
insured_hobbies_golf  \
0                        0                         0
0
1                        0                         0
0

   insured_hobbies_hiking  insured_hobbies_kayaking
insured_hobbies_movies  \
0                       0                         0
0
1                       0                         0
0

   insured_hobbies_paintball  insured_hobbies_polo
insured_hobbies_reading  \
0                          0                     0
0
1                          0                     0
1

   insured_hobbies_skydiving  insured_hobbies_sleeping  \
0                          0                         1
1                          0                         0

   insured_hobbies_video-games  insured_hobbies_yachting  \
0                            0                         0
1                            0                         0
```

```
   insured_relationship_husband  insured_relationship_not-in-family  \
0                             1                                   0
1                             0                                   0

   insured_relationship_other-relative  insured_relationship_own-child
\
0                                    0                               0

1                                    1                               0


   insured_relationship_unmarried  insured_relationship_wife  \
0                               0                          0
1                               0                          0

   incident_type_Multi-vehicle Collision  incident_type_Parked Car  \
0                                      0                         0
1                                      0                         0

   incident_type_Single Vehicle Collision  incident_type_Vehicle Theft
\
0                                       1                            0

1                                       0                            1


   incident_severity_Major Damage  incident_severity_Minor Damage  \
0                               1                               0
1                               0                               1

   incident_severity_Total Loss  incident_severity_Trivial Damage  \
0                             0                                 0
1                             0                                 0

   authorities_contacted_Ambulance  authorities_contacted_Fire  \
0                                0                           0
1                                0                           0

   authorities_contacted_None  authorities_contacted_Other  \
0                           0                            0
1                           0                            0

   authorities_contacted_Police  incident_state_NC  incident_state_NY
\
0                             1                  0                  0

1                             1                  0                  0


   incident_state_OH  incident_state_PA  incident_state_SC
```

```
incident_state_VA  \
0                  0                  0                  1
0
1                  0                  0                  0
1

   incident_state_WV  incident_city_Arlington  incident_city_Columbus
\
0                  0                        0                       1

1                  0                        0                       0


   incident_city_Hillsdale  incident_city_Northbend
incident_city_Northbrook  \
0                        0                        0
0
1                        0                        0
0

   incident_city_Riverwood  incident_city_Springfield
auto_make_Accura  \
0                        0                          0
0
1                        1                          0
0

   auto_make_Audi  auto_make_BMW  auto_make_Chevrolet  auto_make_Dodge
\
0               0              0                    0                0

1               0              0                    0                0


   auto_make_Ford  auto_make_Honda  auto_make_Jeep  auto_make_Mercedes
\
0               0                0               0                   0

1               0                0               0                   1


   auto_make_Nissan  auto_make_Saab  auto_make_Suburu
auto_make_Toyota  \
0                 0               1                 0
0
1                 0               0                 0
0

   auto_make_Volkswagen  auto_model_3 Series  auto_model_92x
auto_model_93  \
```

```
0                     0                     0                1
0
1                     0                     0                0
0

   auto_model_95  auto_model_A3  auto_model_A5  auto_model_Accord  \
0              0              0              0                  0
1              0              0              0                  0

   auto_model_C300  auto_model_CRV  auto_model_Camry  auto_model_Civic
\
0                0               0                 0                 0

1                0               0                 0                 0


   auto_model_Corolla  auto_model_E400  auto_model_Escape
auto_model_F150  \
0                   0                0                  0
0
1                   0                1                  0
0

   auto_model_Forrestor  auto_model_Fusion  auto_model_Grand Cherokee
\
0                     0                  0                          0

1                     0                  0                          0


   auto_model_Highlander  auto_model_Impreza  auto_model_Jetta  \
0                      0                   0                 0
1                      0                   0                 0

   auto_model_Legacy  auto_model_M5  auto_model_MDX  auto_model_ML350
\
0                  0              0               0                 0

1                  0              0               0                 0


   auto_model_Malibu  auto_model_Maxima  auto_model_Neon
auto_model_Passat  \
0                  0                  0                0
0
1                  0                  0                0
0

   auto_model_Pathfinder  auto_model_RAM  auto_model_RSX  \
0                      0               0               0
1                      0               0               0
```

```
   auto_model_Silverado  auto_model_TL  auto_model_Tahoe
auto_model_Ultima  \
0                     0              0                 0
0
1                     0              0                 0
0

   auto_model_Wrangler  auto_model_X5  auto_model_X6  \
0                    0              0              0
1                    0              0              0

   incident_period_of_day_afternoon
incident_period_of_day_early_morning  \
0                                 0
1
1                                 0
0

   incident_period_of_day_evening  incident_period_of_day_fore-noon  \
0                               0                                 0
1                               0                                 0

   incident_period_of_day_morning  incident_period_of_day_night  \
0                               0                             0
1                               1                             0

   incident_period_of_day_past_midnight  property_damage  \
0                                     0                1
1                                     0                0

   police_report_available  collision_en
0                        1             3
1                        0             0

X = pd.concat([X, df._get_numeric_data()], axis=1)  # joining numeric
columns
X.head(2)

   policy_state_IL  policy_state_IN  policy_state_OH
policy_csl_100/300  \
0                0                0                1
0
1                0                1                0
0

   policy_csl_250/500  policy_csl_500/1000  insured_sex_FEMALE  \
0                   1                    0                   0
1                   1                    0                   0

   insured_sex_MALE  insured_education_level_Associate  \
```

```
0                 1                                   0
1                 1                                   0

   insured_education_level_College  insured_education_level_High
School  \
0                                0
0
1                                0
0

   insured_education_level_JD  insured_education_level_MD  \
0                           0                           1
1                           0                           1

   insured_education_level_Masters  insured_education_level_PhD  \
0                                0                            0
1                                0                            0

   insured_occupation_adm-clerical  insured_occupation_armed-forces  \
0                                0                                0
1                                0                                0

   insured_occupation_craft-repair  insured_occupation_exec-managerial
\
0                                1                                   0

1                                0                                   0


   insured_occupation_farming-fishing  insured_occupation_handlers-
cleaners  \
0                                   0
0
1                                   0
0

   insured_occupation_machine-op-inspct  insured_occupation_other-
service  \
0                                     0
0
1                                     1
0

   insured_occupation_priv-house-serv  insured_occupation_prof-
specialty  \
0                                   0
0
1                                   0
0
```

```
   insured_occupation_protective-serv  insured_occupation_sales  \
0                                    0                          0
1                                    0                          0

   insured_occupation_tech-support  insured_occupation_transport-
moving  \
0                                 0
0
1                                 0
0

   insured_hobbies_base-jumping  insured_hobbies_basketball  \
0                              0                            0
1                              0                            0

   insured_hobbies_board-games  insured_hobbies_bungie-jumping  \
0                             0                                0
1                             0                                0

   insured_hobbies_camping  insured_hobbies_chess
insured_hobbies_cross-fit  \
0                        0                      0
0
1                        0                      0
0

   insured_hobbies_dancing  insured_hobbies_exercise
insured_hobbies_golf  \
0                        0                         0
0
1                        0                         0
0

   insured_hobbies_hiking  insured_hobbies_kayaking
insured_hobbies_movies  \
0                       0                         0
0
1                       0                         0
0

   insured_hobbies_paintball  insured_hobbies_polo
insured_hobbies_reading  \
0                          0                     0
0
1                          0                     0
1

   insured_hobbies_skydiving  insured_hobbies_sleeping  \
0                          0                         1
1                          0                         0
```

```
   insured_hobbies_video-games  insured_hobbies_yachting  \
0                            0                         0
1                            0                         0

   insured_relationship_husband  insured_relationship_not-in-family  \
0                             1                                   0
1                             0                                   0

   insured_relationship_other-relative  insured_relationship_own-child
\
0                                    0                               0

1                                    1                               0


   insured_relationship_unmarried  insured_relationship_wife  \
0                               0                          0
1                               0                          0

   incident_type_Multi-vehicle Collision  incident_type_Parked Car  \
0                                      0                         0
1                                      0                         0

   incident_type_Single Vehicle Collision  incident_type_Vehicle Theft
\
0                                       1                            0

1                                       0                            1


   incident_severity_Major Damage  incident_severity_Minor Damage  \
0                               1                               0
1                               0                               1

   incident_severity_Total Loss  incident_severity_Trivial Damage  \
0                             0                                 0
1                             0                                 0

   authorities_contacted_Ambulance  authorities_contacted_Fire  \
0                                0                           0
1                                0                           0

   authorities_contacted_None  authorities_contacted_Other  \
0                           0                            0
1                           0                            0

   authorities_contacted_Police  incident_state_NC  incident_state_NY
\
0                             1                  0                  0
```

```
1                              1                  0                  0   

   incident_state_OH  incident_state_PA  incident_state_SC  
incident_state_VA  \
0                  0                  0                  1  
0   
1                  0                  0                  0  
1   

   incident_state_WV  incident_city_Arlington  incident_city_Columbus  
\
0                  0                        0                       1  

1                  0                        0                       0  


   incident_city_Hillsdale  incident_city_Northbend  
incident_city_Northbrook  \
0                        0                        0  
0   
1                        0                        0  
0   

   incident_city_Riverwood  incident_city_Springfield  
auto_make_Accura  \
0                        0                          0  
0   
1                        1                          0  
0   

   auto_make_Audi  auto_make_BMW  auto_make_Chevrolet  auto_make_Dodge  
\
0               0              0                    0                0  

1               0              0                    0                0  


   auto_make_Ford  auto_make_Honda  auto_make_Jeep  auto_make_Mercedes  
\
0               0                0               0                   0  

1               0                0               0                   1  


   auto_make_Nissan  auto_make_Saab  auto_make_Suburu  
auto_make_Toyota  \
0                 0               1                 0  
0   
1                 0               0                 0  
0   
```

```
   auto_make_Volkswagen  auto_model_3 Series  auto_model_92x
auto_model_93  \
0                     0                    0               1
0
1                     0                    0               0
0

   auto_model_95  auto_model_A3  auto_model_A5  auto_model_Accord  \
0              0              0              0                  0
1              0              0              0                  0

   auto_model_C300  auto_model_CRV  auto_model_Camry  auto_model_Civic
\
0                0               0                 0                 0

1                0               0                 0                 0


   auto_model_Corolla  auto_model_E400  auto_model_Escape
auto_model_F150  \
0                   0                0                  0
0
1                   0                1                  0
0

   auto_model_Forrestor  auto_model_Fusion  auto_model_Grand Cherokee
\
0                     0                  0                          0

1                     0                  0                          0


   auto_model_Highlander  auto_model_Impreza  auto_model_Jetta  \
0                      0                   0                 0
1                      0                   0                 0

   auto_model_Legacy  auto_model_M5  auto_model_MDX  auto_model_ML350
\
0                  0              0               0                 0

1                  0              0               0                 0


   auto_model_Malibu  auto_model_Maxima  auto_model_Neon
auto_model_Passat  \
0                  0                  0                0
0
1                  0                  0                0
0
```

```
   auto_model_Pathfinder  auto_model_RAM  auto_model_RSX  \
0                      0               0               0
1                      0               0               0

   auto_model_Silverado  auto_model_TL  auto_model_Tahoe
auto_model_Ultima  \
0                     0              0                 0
0
1                     0              0                 0
0

   auto_model_Wrangler  auto_model_X5  auto_model_X6  \
0                    0              0              0
1                    0              0              0

   incident_period_of_day_afternoon
incident_period_of_day_early_morning  \
0                                 0
1
1                                 0
0

   incident_period_of_day_evening  incident_period_of_day_fore-noon  \
0                               0                                 0
1                               0                                 0

   incident_period_of_day_morning  incident_period_of_day_night  \
0                               0                             0
1                               1                             0

   incident_period_of_day_past_midnight  property_damage  \
0                                     0                1
1                                     0                0

   police_report_available  collision_en  months_as_customer  age  \
0                        1             3                 328   48
1                        0             0                 228   42

   policy_deductable  policy_annual_premium  umbrella_limit  capital-
gains  \
0               1000                1406.91               0
53300
1               2000                1197.22         5000000
0

   capital-loss  number_of_vehicles_involved  bodily_injuries
witnesses  \
0             0                            1                1
2
1             0                            1                0
```

```
0

   total_claim_amount  injury_claim  property_claim  vehicle_claim  \
0               71610          6510           13020          52080   
1                5070           780             780           3510   

   fraud_reported  vehicle_age  
0               1           14  
1               1           11  
```

```
X.columns
```

```
Index(['policy_state_IL', 'policy_state_IN', 'policy_state_OH',
       'policy_csl_100/300', 'policy_csl_250/500',
'policy_csl_500/1000',
       'insured_sex_FEMALE', 'insured_sex_MALE',
       'insured_education_level_Associate',
'insured_education_level_College',
       ...
       'capital-loss', 'number_of_vehicles_involved',
'bodily_injuries',
       'witnesses', 'total_claim_amount', 'injury_claim',
'property_claim',
       'vehicle_claim', 'fraud_reported', 'vehicle_age'],
      dtype='object', length=161)
```

```
X = X.drop(columns = ['fraud_reported'])
X.columns
```

```
Index(['policy_state_IL', 'policy_state_IN', 'policy_state_OH',
       'policy_csl_100/300', 'policy_csl_250/500',
'policy_csl_500/1000',
       'insured_sex_FEMALE', 'insured_sex_MALE',
       'insured_education_level_Associate',
'insured_education_level_College',
       ...
       'capital-gains', 'capital-loss', 'number_of_vehicles_involved',
       'bodily_injuries', 'witnesses', 'total_claim_amount',
'injury_claim',
       'property_claim', 'vehicle_claim', 'vehicle_age'],
      dtype='object', length=160)
```

We now have a dataset that we could use to evaluate an algorithm sensitive to missing values like LDA.

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score

# evaluate an LDA model on the dataset using k-fold cross validation
model = LinearDiscriminantAnalysis()
```

```
kfold = KFold(n_splits=5, random_state=7)
result = cross_val_score(model, X, y, cv=kfold, scoring='accuracy')
print(result.mean())
```

```
C:\Users\Sarit\Anaconda3\envs\python\lib\site-packages\sklearn\
discriminant_analysis.py:388: UserWarning: Variables are collinear.
  warnings.warn("Variables are collinear.")
C:\Users\Sarit\Anaconda3\envs\python\lib\site-packages\sklearn\
discriminant_analysis.py:388: UserWarning: Variables are collinear.
  warnings.warn("Variables are collinear.")
C:\Users\Sarit\Anaconda3\envs\python\lib\site-packages\sklearn\
discriminant_analysis.py:388: UserWarning: Variables are collinear.
  warnings.warn("Variables are collinear.")
C:\Users\Sarit\Anaconda3\envs\python\lib\site-packages\sklearn\
discriminant_analysis.py:388: UserWarning: Variables are collinear.
  warnings.warn("Variables are collinear.")

0.841

C:\Users\Sarit\Anaconda3\envs\python\lib\site-packages\sklearn\
discriminant_analysis.py:388: UserWarning: Variables are collinear.
  warnings.warn("Variables are collinear.")
```

84.1% accuracy without standardizing the data. This looks good to go for Random Forest Classification method. Random Forest is a tree-based model and hence does not require feature scaling. The convergence and numerical precision issues, which can sometimes trip up the algorithms used in logistic and linear regression, as well as neural networks, aren't so important in case of random forest.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
train_size=0.8, random_state=1234)
print('length of X_train and X_test: ', len(X_train), len(X_test))
print('length of y_train and y_test: ', len(y_train), len(y_test))
```

```
length of X_train and X_test:  800 200
length of y_train and y_test:  800 200
```

Random Forest Classification

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, recall_score,
classification_report, cohen_kappa_score
from sklearn import metrics

# Baseline Random forest based Model
rfc = RandomForestClassifier(criterion = 'gini', n_estimators=1000,
verbose=1, n_jobs = -1,
                             class_weight = 'balanced', max_features =
'auto')
rfcg = rfc.fit(X_train,y_train) # fit on training data
predictions = rfcg.predict(X_test)
```

```
print('Baseline: N_features: ', len(list(X.columns)))
print('Baseline: Accuracy: ', round(accuracy_score(y_test,
predictions)*100, 2))
print( 'Cohen Kappa: '+ str(np.round(cohen_kappa_score(y_test,
predictions),3)))
print('Baseline: Recall: ', round(recall_score(y_test,
predictions)*100, 2))
print('\n Classification Report:\n',
classification_report(y_test,predictions))
```

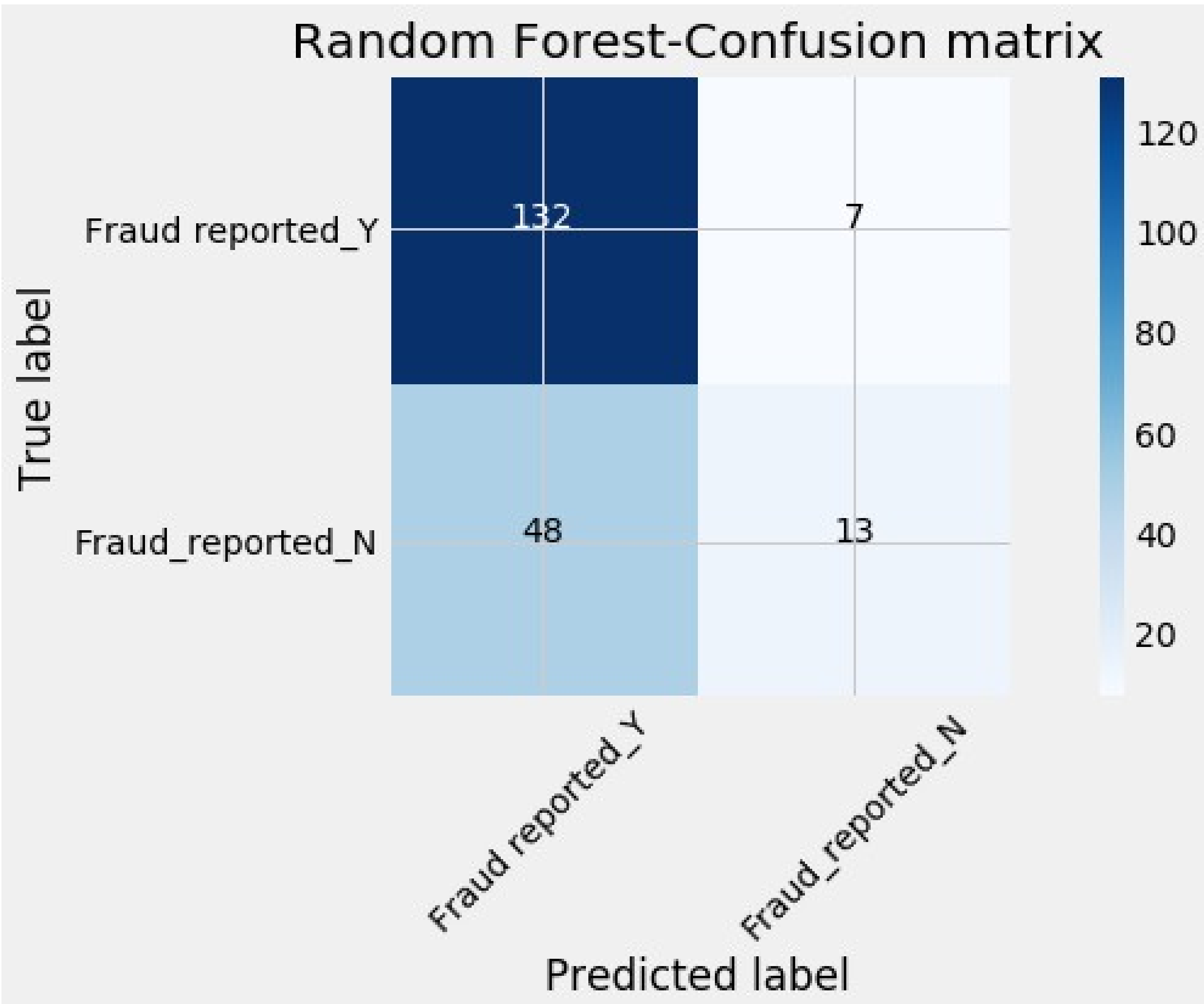
```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8
concurrent workers.
[Parallel(n_jobs=-1)]: Done  34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=-1)]: Done 184 tasks      | elapsed:    0.2s
[Parallel(n_jobs=-1)]: Done 434 tasks      | elapsed:    0.5s
[Parallel(n_jobs=-1)]: Done 784 tasks      | elapsed:    1.0s
[Parallel(n_jobs=-1)]: Done 1000 out of 1000 | elapsed:    1.3s
finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent
workers.
[Parallel(n_jobs=8)]: Done  34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 434 tasks      | elapsed:    0.0s

Baseline: N_features:  160
Baseline: Accuracy:  72.5
Cohen Kappa: 0.201
Baseline: Recall:  21.31

 Classification Report:
               precision    recall  f1-score   support

           0       0.73      0.95      0.83       139
           1       0.65      0.21      0.32        61

    accuracy                           0.73       200
   macro avg       0.69      0.58      0.57       200
weighted avg       0.71      0.72      0.67       200


[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 1000 out of 1000 | elapsed:    0.2s
finished
```

```
rfcg
```

```
RandomForestClassifier(bootstrap=True, class_weight='balanced',
                       criterion='gini', max_depth=None,
max_features='auto',
                       max_leaf_nodes=None, min_impurity_decrease=0.0,
```

```
                       min_impurity_split=None, min_samples_leaf=1,
                       min_samples_split=2,
min_weight_fraction_leaf=0.0,
                       n_estimators=1000, n_jobs=-1, oob_score=False,
                       random_state=None, verbose=1, warm_start=False)

from sklearn.metrics import confusion_matrix

import itertools

#Evaluation of Model - Confusion Matrix Plot
def plot_confusion_matrix(cm, classes, title ='Confusion matrix',
normalize=False, cmap = plt.cm.Blues):
    """
    This function prints and plots the confusion matrix.
    Normalization can be applied by setting `normalize=True`.
    """
    print('Confusion matrix')

    print(cm)

    fig = plt.figure(figsize=(10,6))
    plt.style.use('fivethirtyeight')
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    fmt = '.2f' if normalize else 'd'
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]),
range(cm.shape[1])):
        plt.text(j, i, format(cm[i, j], fmt),
                 horizontalalignment="center",
                 color="white" if cm[i, j] > thresh else "black")

    plt.ylabel('True label')
    plt.xlabel('Predicted label')
    plt.tight_layout()

# Compute confusion matrix
cnf_matrix = confusion_matrix(y_test, predictions)
np.set_printoptions(precision=2)


# Plot non-normalized confusion matrix
plt.figure()
plot_confusion_matrix(cnf_matrix, classes=['Fraud
```

```
reported_Y','Fraud_reported_N'],
                              title='Random Forest-Confusion matrix')

Confusion matrix
[[132   7]
 [ 48  13]]

<Figure size 432x288 with 0 Axes>
```



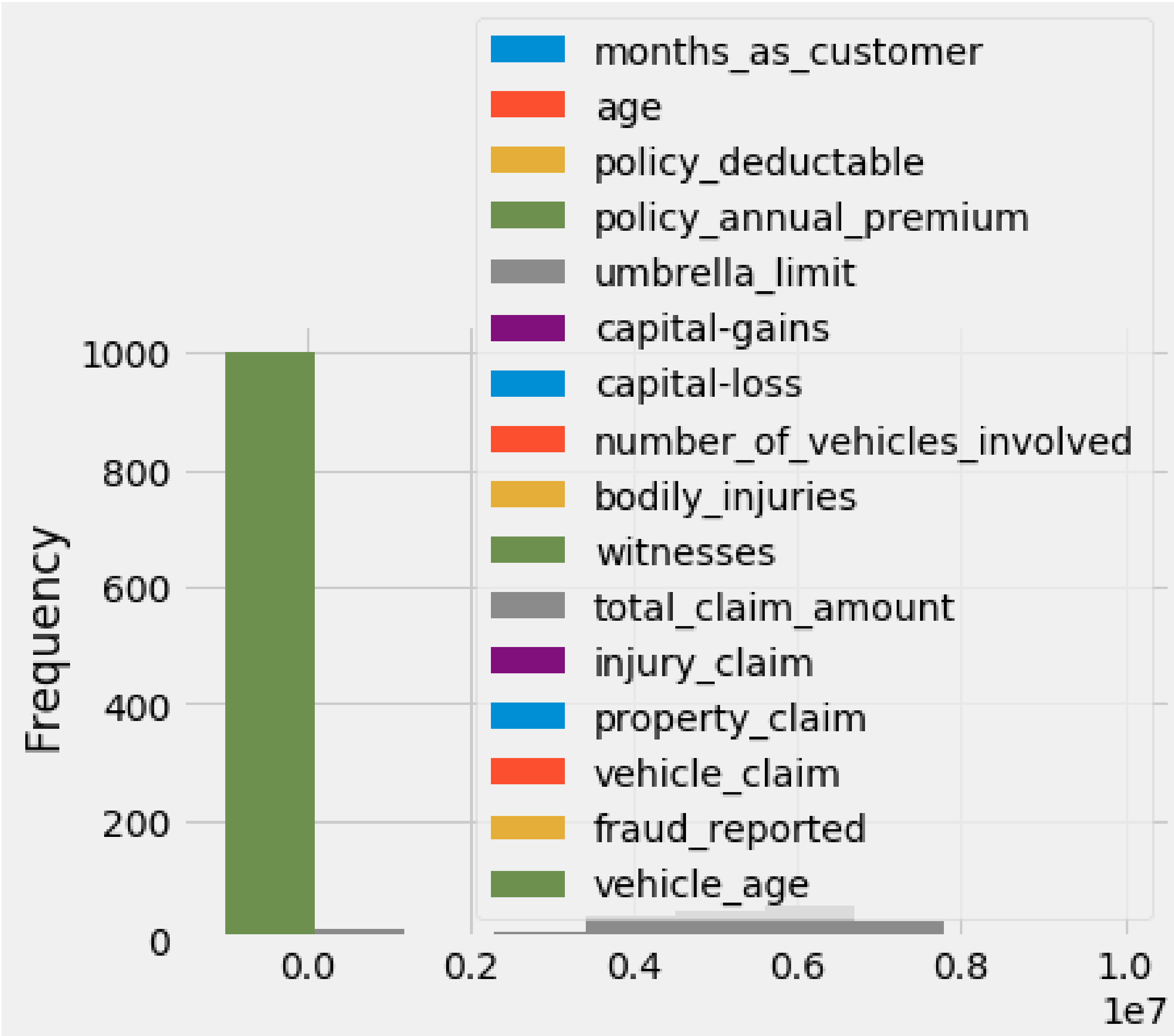
With 72.5% accuracy, we take a closer look at the confusion matrix:

- 132 transactions were classified as valid that were actually valid
- 7 transactions were classified as fraud that were actually valid (type 1 error)
- 48 transactions were classified as valid that were fraud (type 2 error)
- 13 transactions were classified as fraud.

Err = ((FP+FN)/ (TP+TN+FN+FP) = {(48+7) / (132+7+48+13)}*100 = 0.275

So, the algorithm misclassified 27.5% fraudulent transactions. We looked at other measures too like the Cohen Kappa, Recall, and F1 score. In each case, the scores are closer to 1.

```
# Generate a Histogram plot for anomaly detection
df.plot(kind='hist')
plt.show()
```




```
# Minimum and maximum premium
print('Minimum premimum ' + str(df['policy_annual_premium'].min()))
print('Maximum premium ' + str(df['policy_annual_premium'].max()))
```

```
Minimum premimum 433.33
Maximum premium 2047.59
```

```
# Minimum and maximum age of vehicle
print('Vehicle age-minimum ' + str(df['vehicle_age'].min()))
print('Vehicle Age-maximum ' + str(df['vehicle_age'].max()))
```

```
Vehicle age-minimum 3
Vehicle Age-maximum 23
```

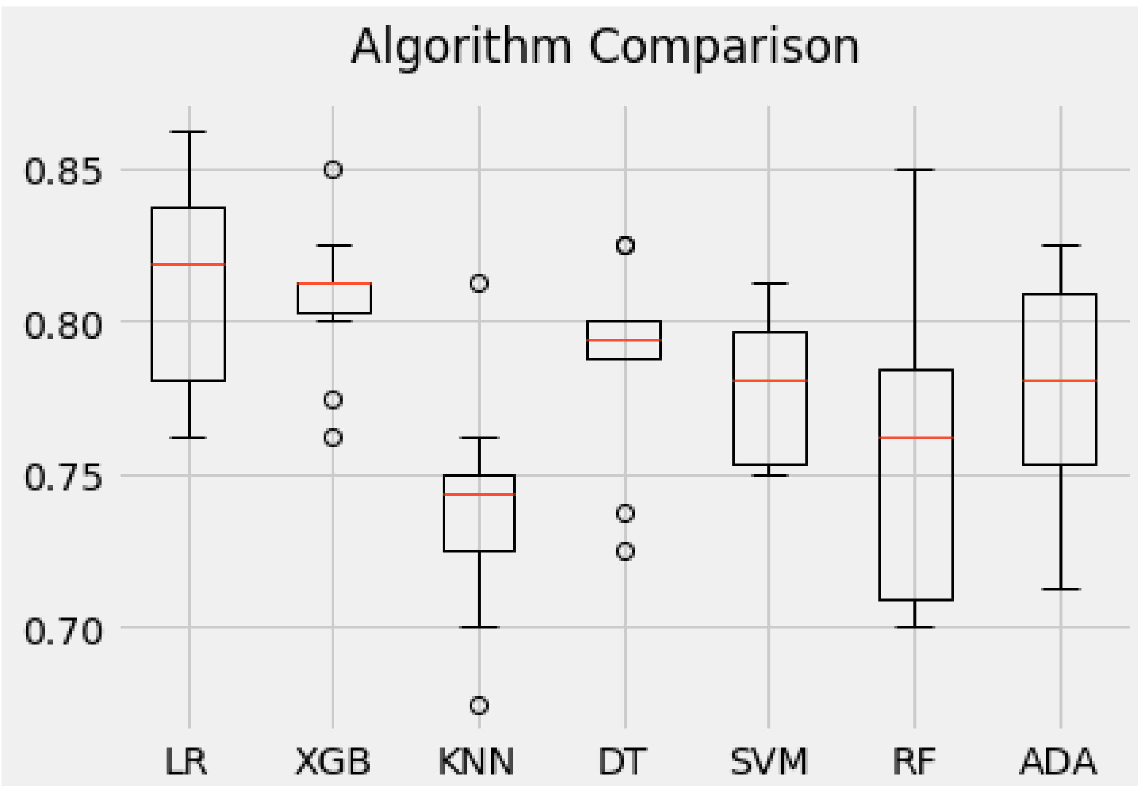
```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler(with_mean=False)
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

The 10-fold cross validation procedure is used to evaluate each algorithm, importantly configured with the same random seed to ensure that the same splits to the training data are performed and that each algorithms is evaluated in precisely the same way.

```
from xgboost import XGBClassifier
from sklearn import model_selection
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LogisticRegressionCV
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import AdaBoostClassifier

xgb = XGBClassifier()
logreg2= LogisticRegressionCV(solver='lbfgs', cv=10)
knn = KNeighborsClassifier(5)
svcl = SVC()
adb = AdaBoostClassifier()
dtclf = DecisionTreeClassifier(max_depth=5)
rfclf = RandomForestClassifier()

# prepare configuration for cross validation test harness
seed = 7
# prepare models
models = []
models.append(('LR', LogisticRegressionCV(solver='lbfgs',
max_iter=5000, cv=10)))
models.append(('XGB', XGBClassifier()))
models.append(('KNN', KNeighborsClassifier()))
models.append(('DT', DecisionTreeClassifier()))
models.append(('SVM', SVC(gamma='auto')))
models.append(('RF', RandomForestClassifier(n_estimators=100)))
models.append(('ADA', AdaBoostClassifier(n_estimators=100)))

# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'
for name, model in models:
    kfold = model_selection.KFold(n_splits=10, random_state=seed)
    cv_results = model_selection.cross_val_score(model,
X_train_scaled, y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
```

```
    print(msg)

# boxplot algorithm comparison
fig = plt.figure()
fig.suptitle('Algorithm Comparison')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```

```
LR: 0.812500 (0.034460)
XGB: 0.807500 (0.023184)
KNN: 0.738750 (0.035111)
DT: 0.787500 (0.031125)
SVM: 0.777500 (0.022220)
RF: 0.756250 (0.045843)
ADA: 0.778750 (0.034483)
```



Above a list of each algorithm, the mean accuracy and the standard deviation accuracy and a box & whisker plot showing the spread of the accuracy scores across each cross validation fold for each algorithm.

It is clear that the LR or LDA is good enough for both feature selection (81% and 84% accuracy with 100 features) as well as model selection.

I will analyse both both logistic regression and linear discriminate analysis further on this problem.