

# CAR INSURANCE PROJECT

Presented by  
Barkallah Ahmed  
Jebari Aziz  
Mbarki Mouhamed Amine  
Mlaouhi Yahya  
Mnif Noura  
Omri Marwen





# OUTLINE

- Introduction (Project Goals, Obstacles..)
  - Business and Data comprehension
  - Data Preparation
  - Data modeling and Evaluation
  - Data Visualization
- Conclusion and perspectives

# Introduction

- protection the right of policyholders
- the study of insurance operations companies
- give its opinion on any other question falling within its attributions Companies



# Buisness Understanding



Protection



Compensation



Justice

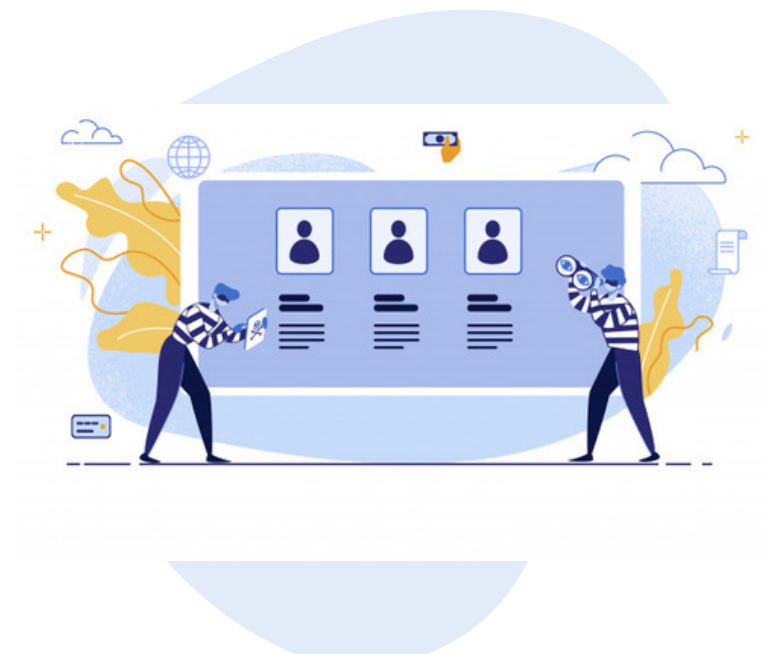
# Obstacles



Data growth



Human Limitation



Fraud

# Project Goals

Profit Maximization

Predict class Bonus Malus

Fraud Detection



# Data Comprehension



Internal data provided by  
CGA



assure



atlt
















police



Sinistre



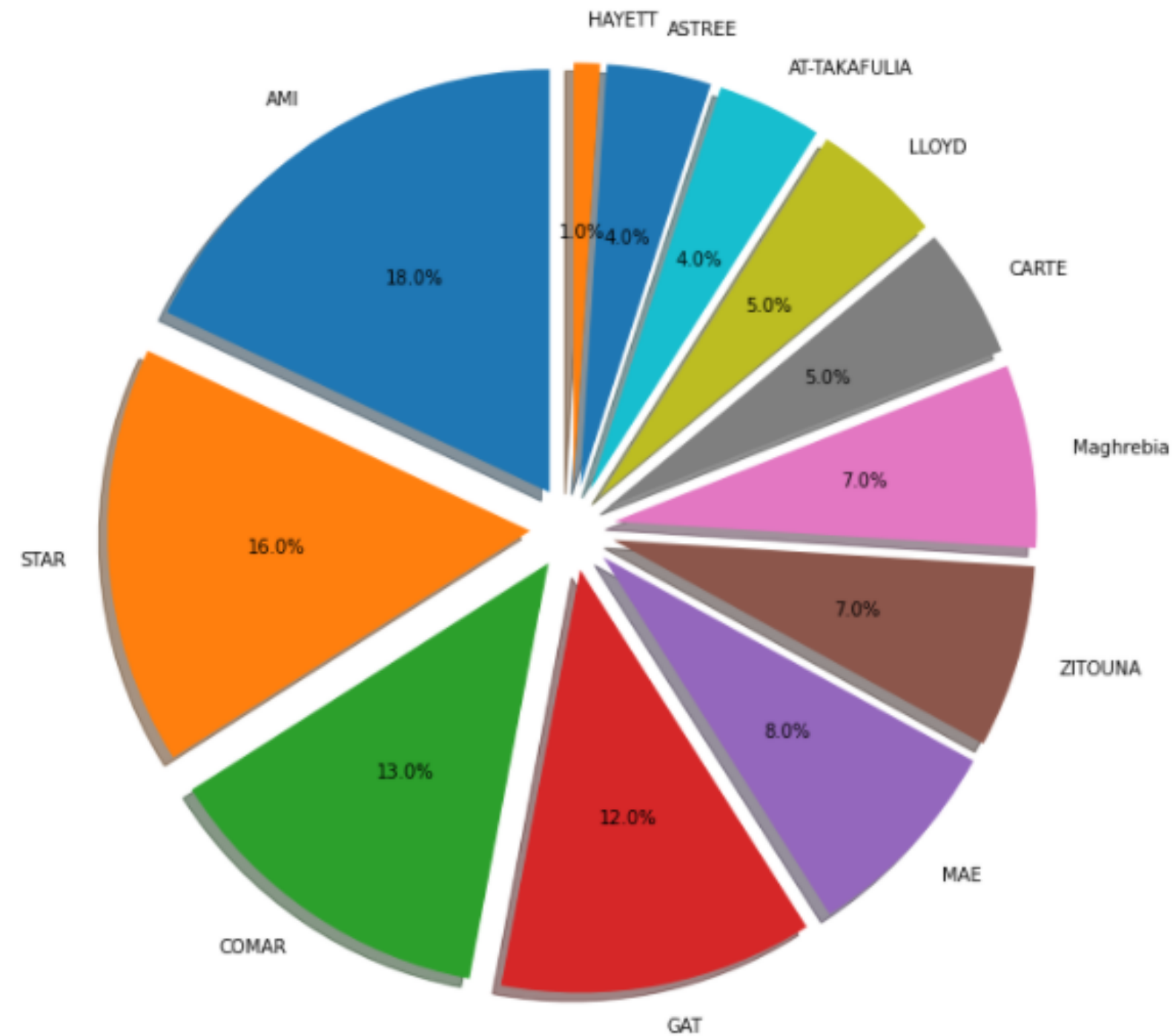
Vehicule

- +  dbo.Assure
- +  dbo.ClassesBonusMalus
- +  dbo.Compagnies
- +  dbo.FaitClassesBonusMalus
- +  dbo.FaitSinistre
- +  dbo.MarqueVehicule
- +  dbo.Police
- +  dbo.Sinistres
- +  dbo.Souscripteur
- +  dbo.Temps
- +  dbo.TypeInfo
- +  dbo.UsageVehicule
- +  dbo.Vehicule

# Data Comprehension



External data web  
scrapping using selenuin  
Customer satisfaction on  
insurance company





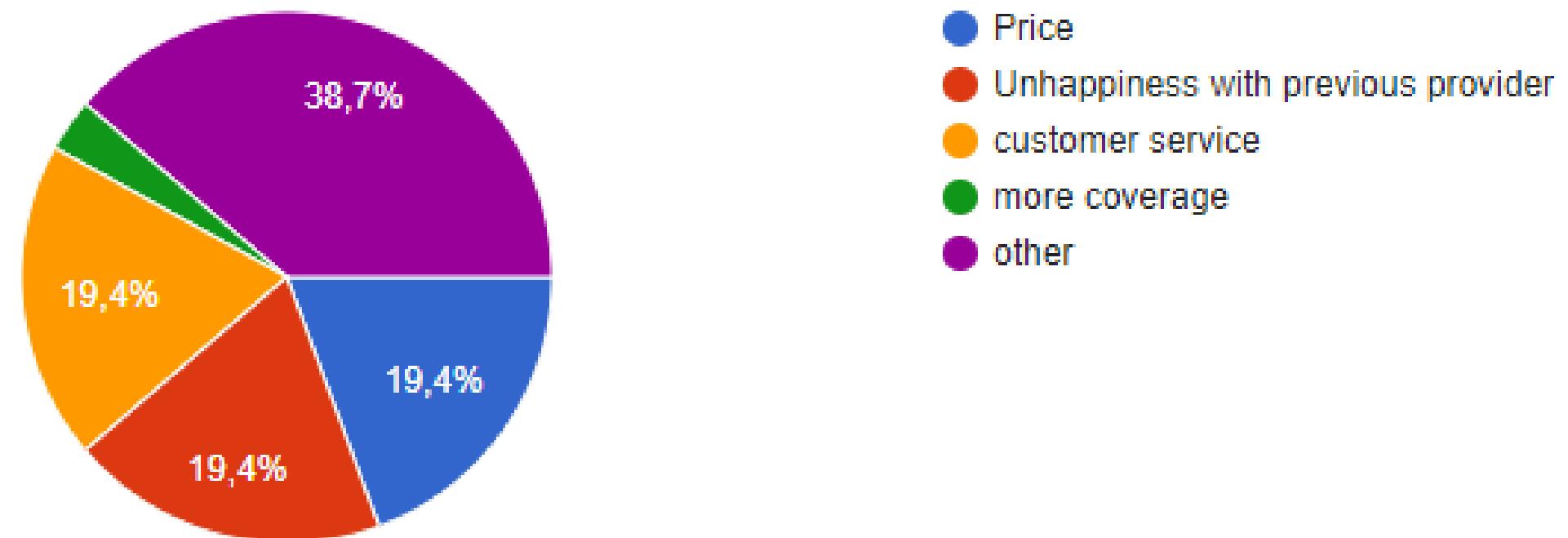
# Data Comprehension



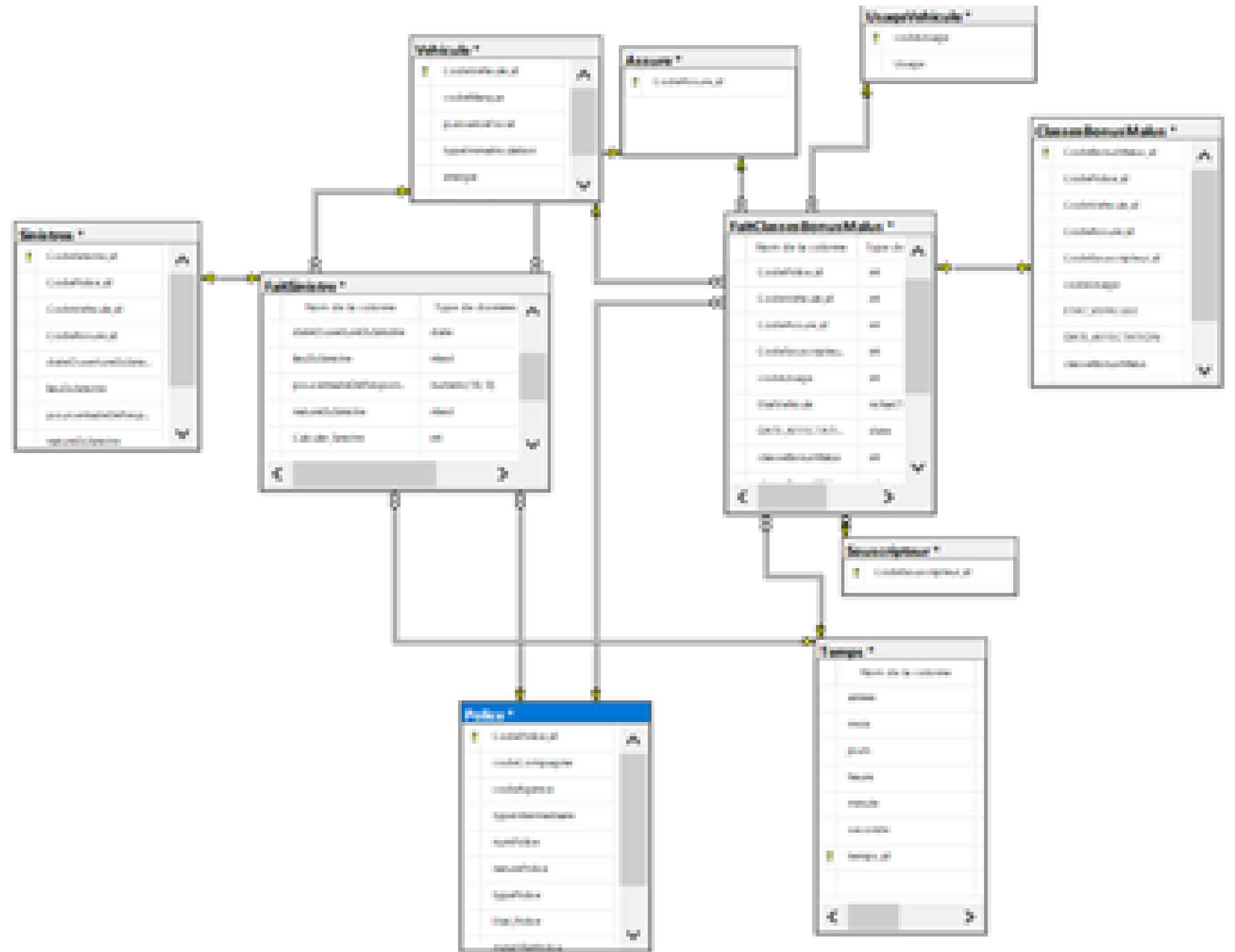
We launched a public form  
for more insights from  
customer

What was the main reason for switching your car insurance company?

31 réponses

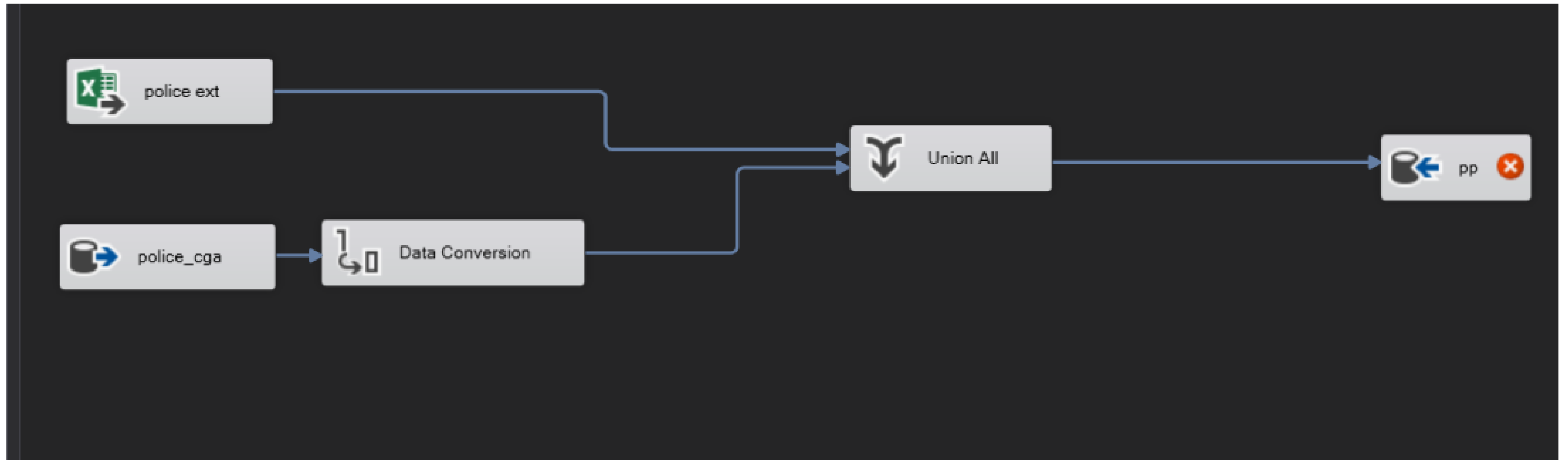


# Data warehouse



# Data preparation

Merging the data sources



# Data preparation

Cleaning the data by deleting the null values

```
#nombre de null  
for i in fait_Sinistre:  
    if fait_Sinistre[i].isnull().sum()!=0:  
        print(fait_Sinistre[i].name+' :'+str(fait_Sinistre[i].isnull().sum()))
```

```
heureSurveillanceSinistre :22353  
lieuDuSinistre :333491  
codeCompagnieAdverse :296753  
pourcentageCompagnieAdverse :22353  
numeroImmatriculationVehiculeAdverse :355306  
mouvementDuSinistre :22353  
typeImmatriculation :124458  
dateInsertion :124458  
etatVehicule :555747  
codeCompagnie :342759  
codeAgence :342759  
typeIntermediaire :342759  
naturePolice :342759
```

# Data preparation

## Fraud detection

- Individual policy type  
while the client has multiple cars
- personal vehicle  
while policy type flotte
- Flotte policy type  
while the client has less than 3 cars

## DATE\_AFFECTATION

0	2516-07-08 00:00:00
1	2115-03-02 00:00:00

- wrong assignment date
- rental vehicle  
while policy type individual

# Data Modelling

## Internal Data

labeled data

+

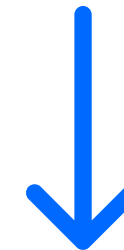
target class



supervised learning



Discrete data



Classification  
models



-Logistic Regression

-**Tree Decision**

-Multi Class Classification

-MultiLayer Perceptron

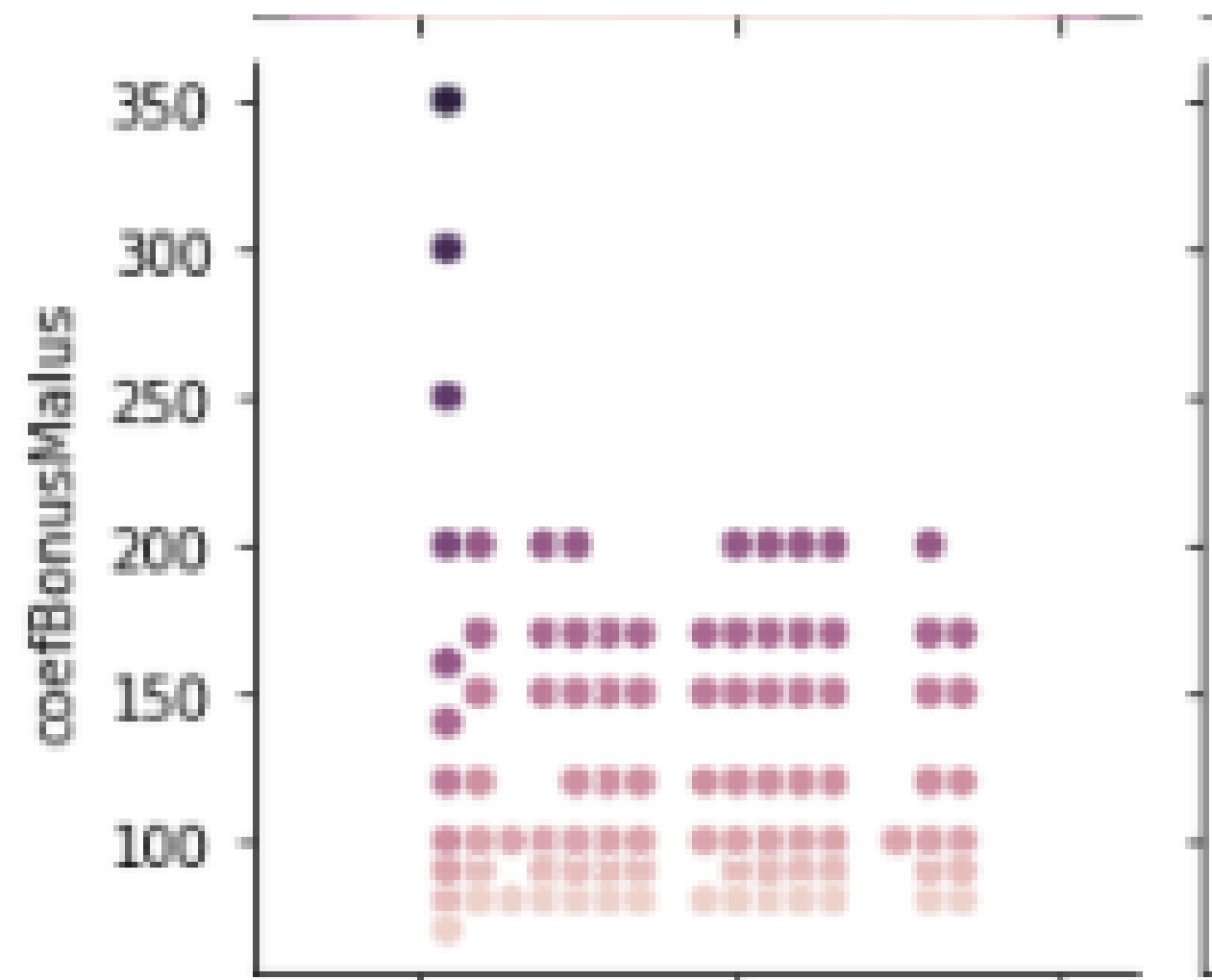
-KNN

-Naive Bayes

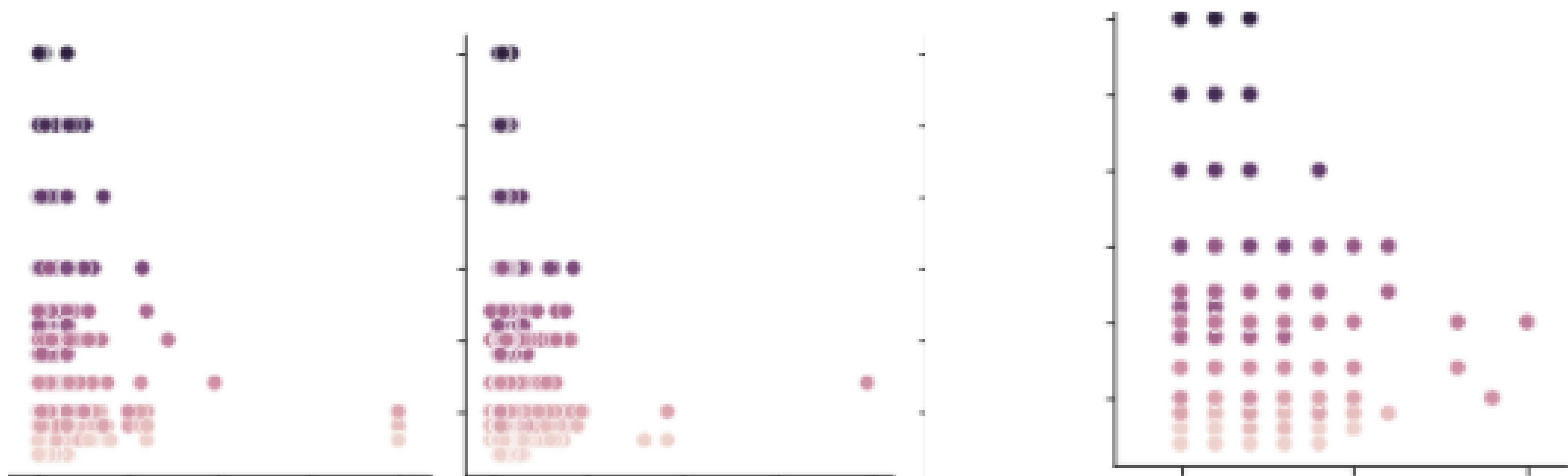
-SVM

-Random Forest

-Catboost



```
sns.pairplot(BM, hue="classeBonusMalus")
```

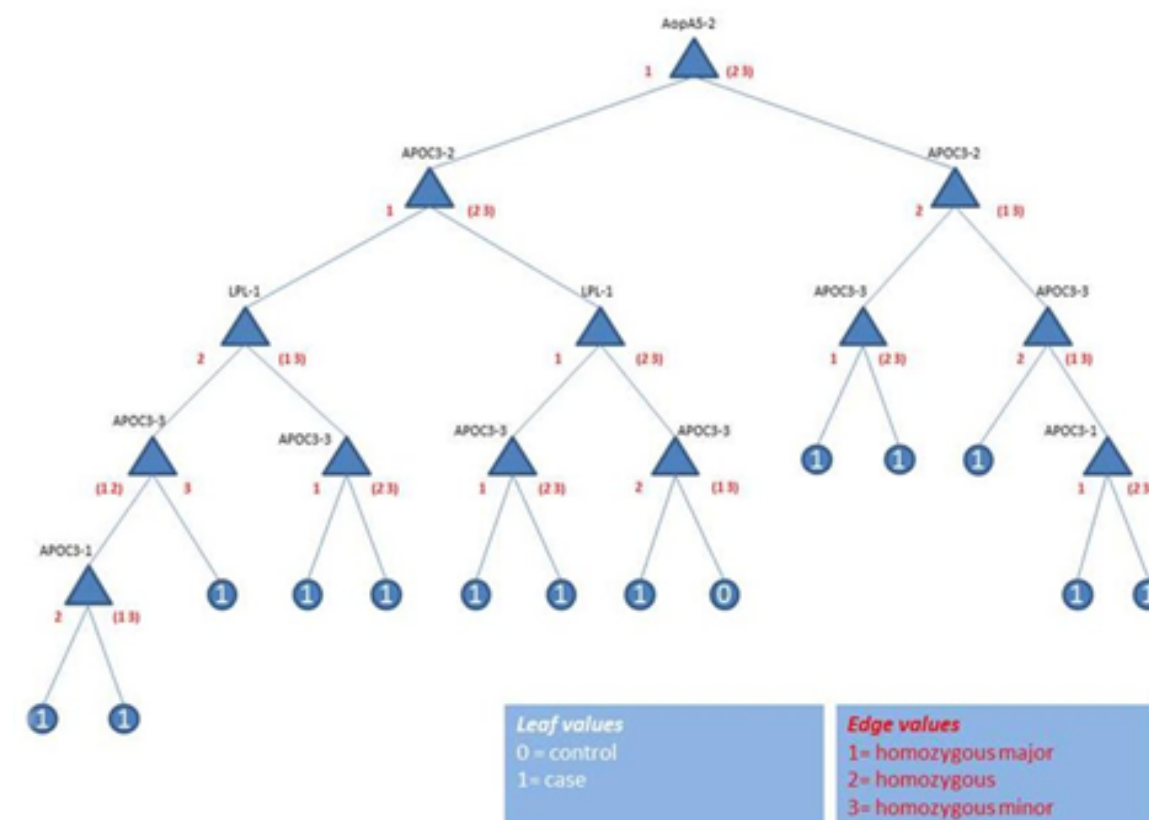


# Decision Tree

## Modelling

In finance, forecasting future outcomes and assigning probabilities to those outcomes

General business decision-making





# Decision Tree

## Evaluation

```
Entrée [585]: from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, random_state=3)
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier(random_state=0)
```

```
Entrée [586]: dt.fit(X_train, y_train)
```

```
Out[586]: DecisionTreeClassifier(random_state=0)
```

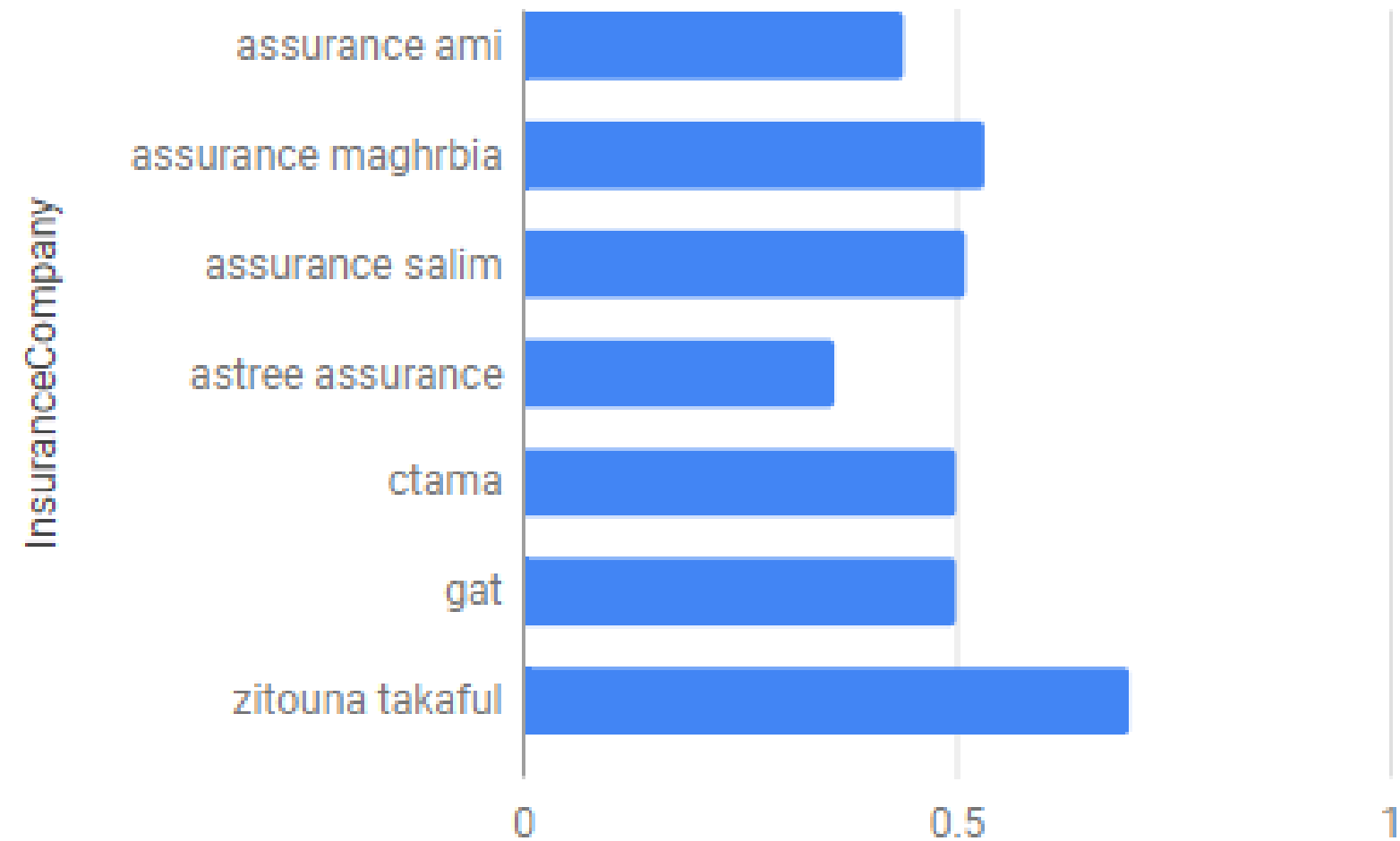
```
Entrée [588]: print('Le test score est :', dt.score(X_test, y_test))
```

```
Le test score est : 0.9495293567008165
```

# Data Modelling

## External Data

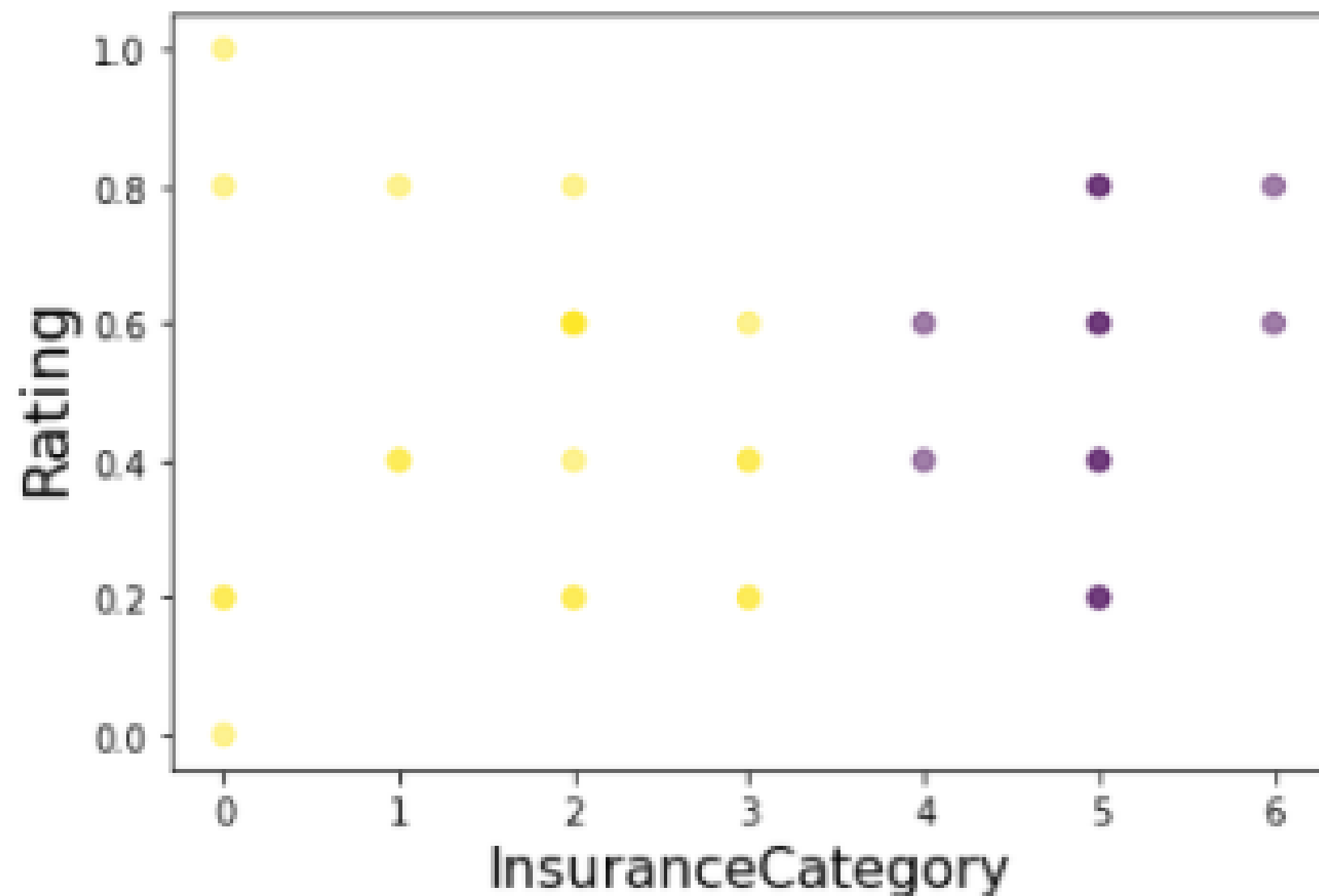
InsuranceVote



# Data Modelling

## Unsupervised Model

```
from sklearn.cluster import KMeans  
k_means = KMeans(init = "k-means++", n_clusters = 2, n_init
```



```
28] k_means.fit(X)
```

# Data Visualization



# Detect fraudulent client

## Companies Informations

Company:

Agence Code:

## Client Informations

Usage Code:

BonusMalus Class:

Number of accidents:

## Car Informations

Car Brand:

Horse Power:

Fuel:

## Policy Informations

Intermediate type:

Policy Nature:

Policy Type:

Policy State:

Predict

# Predict BM class

## Client Informations

Usage Code:

Number of accidents:

## Vehicle Informations

Car Brand:

Horse power:

Fuel:

## Policy Informations

Intermediate type:

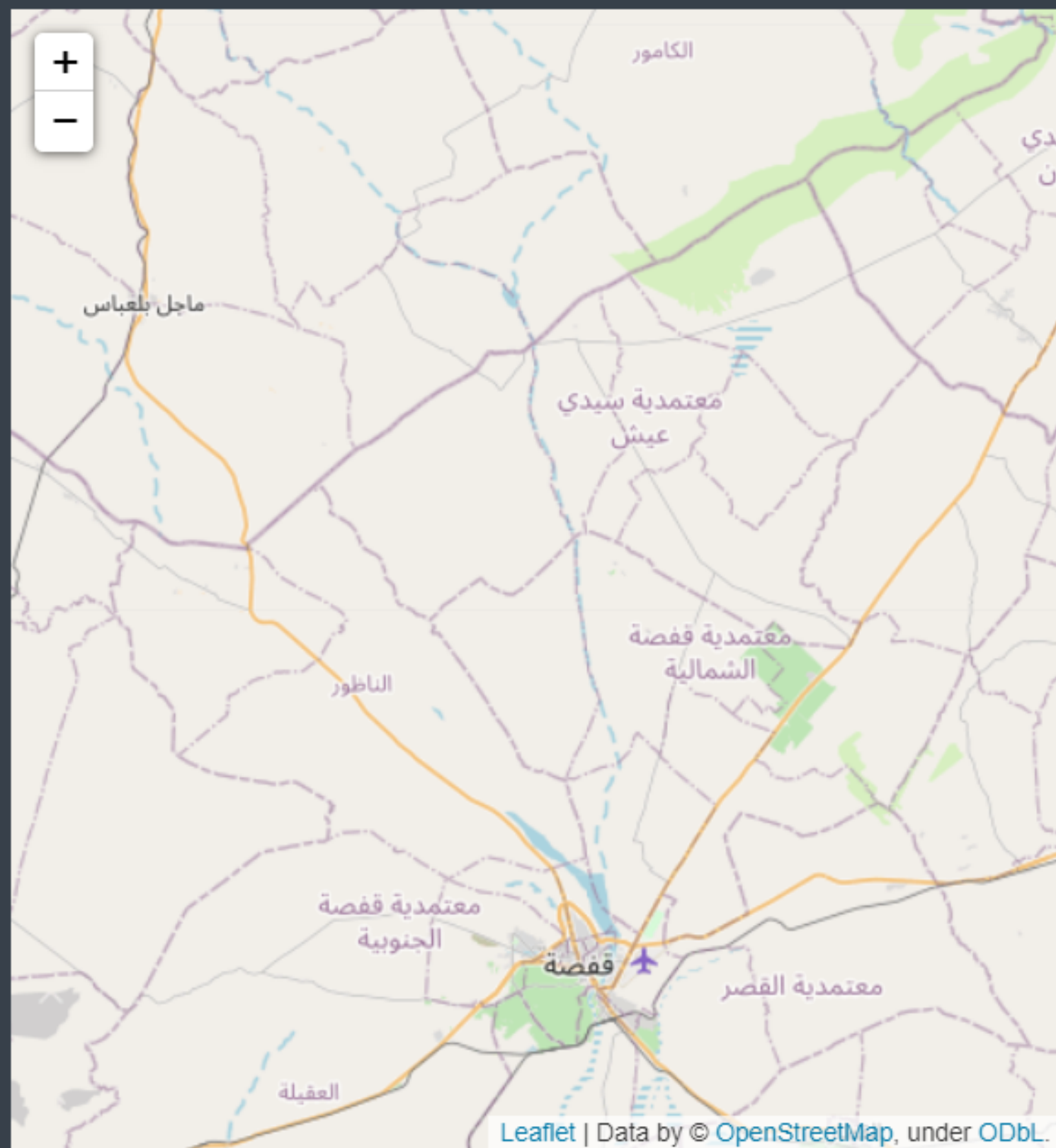
Policy Nature:

Policy Type:

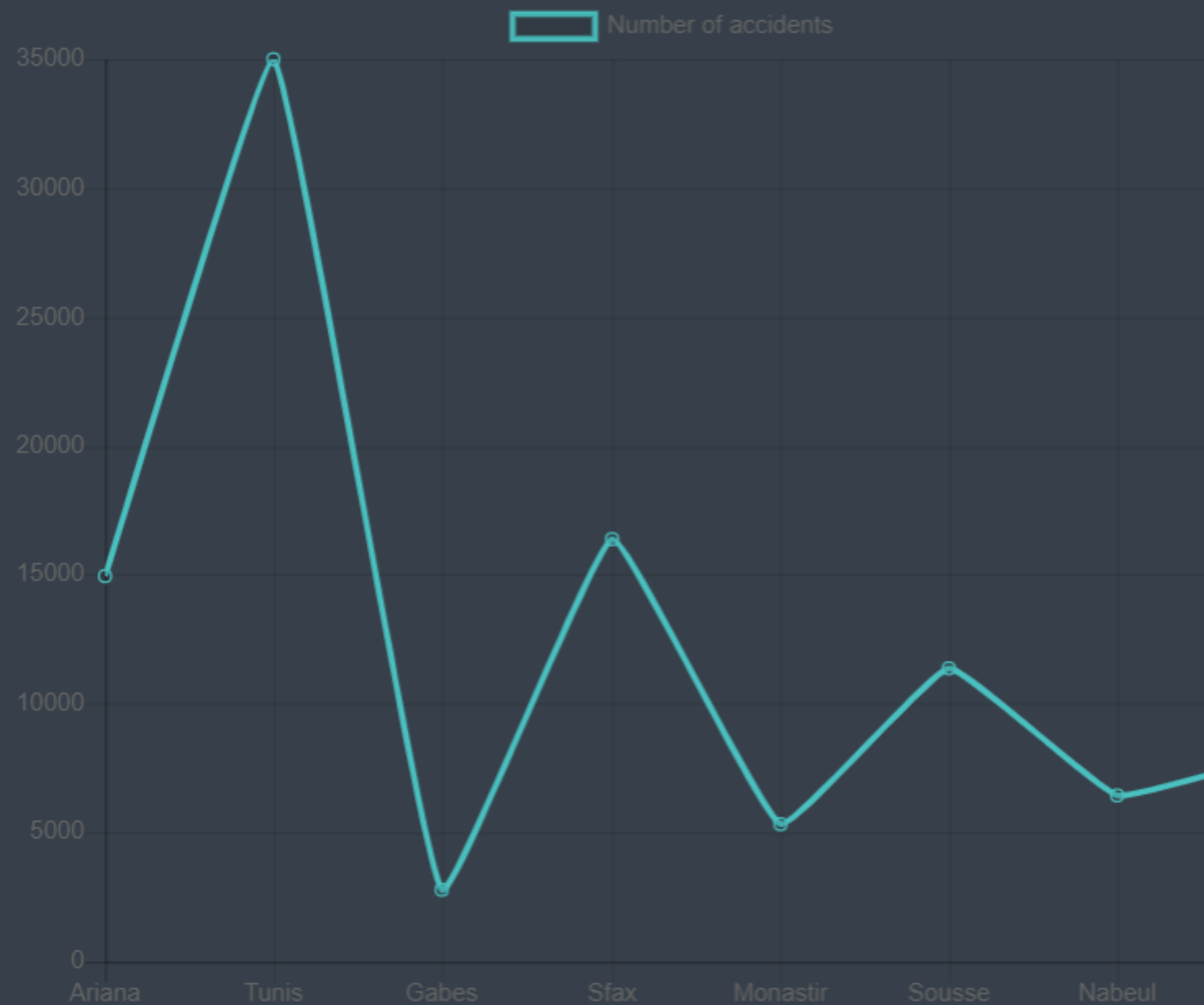
Policy State:

Predict

MAP OF ACCIDENTS BY STATE



NUMBER OF ACCIDENTS PER STATE





c o n c l u s i o n



THANK YOU FOR YOUR ATTENTION