



PRÉSENTATION DE PROJET

# NEW\_CAR

LA PLATEFORME



NEW\_CAR

# INTRODUCTION



Xxx €



## Contexte

Pouvoir prédire le prix d'une voiture en fonction de caractéristiques données

## Enjeux

- Charger les données
- Analyser les données
- Identifier des tendances
- Réaliser une régression linéaire
- Analyser les résultats



NEW\_CAR


# RÉCUPÉRATION DES DONNÉES

## Affichage

Affichage des données de façon basique

## Analyse sommaire

Première analyse des données ,  
définition des variables, de leur type  
ainsi que de la taille du dataset

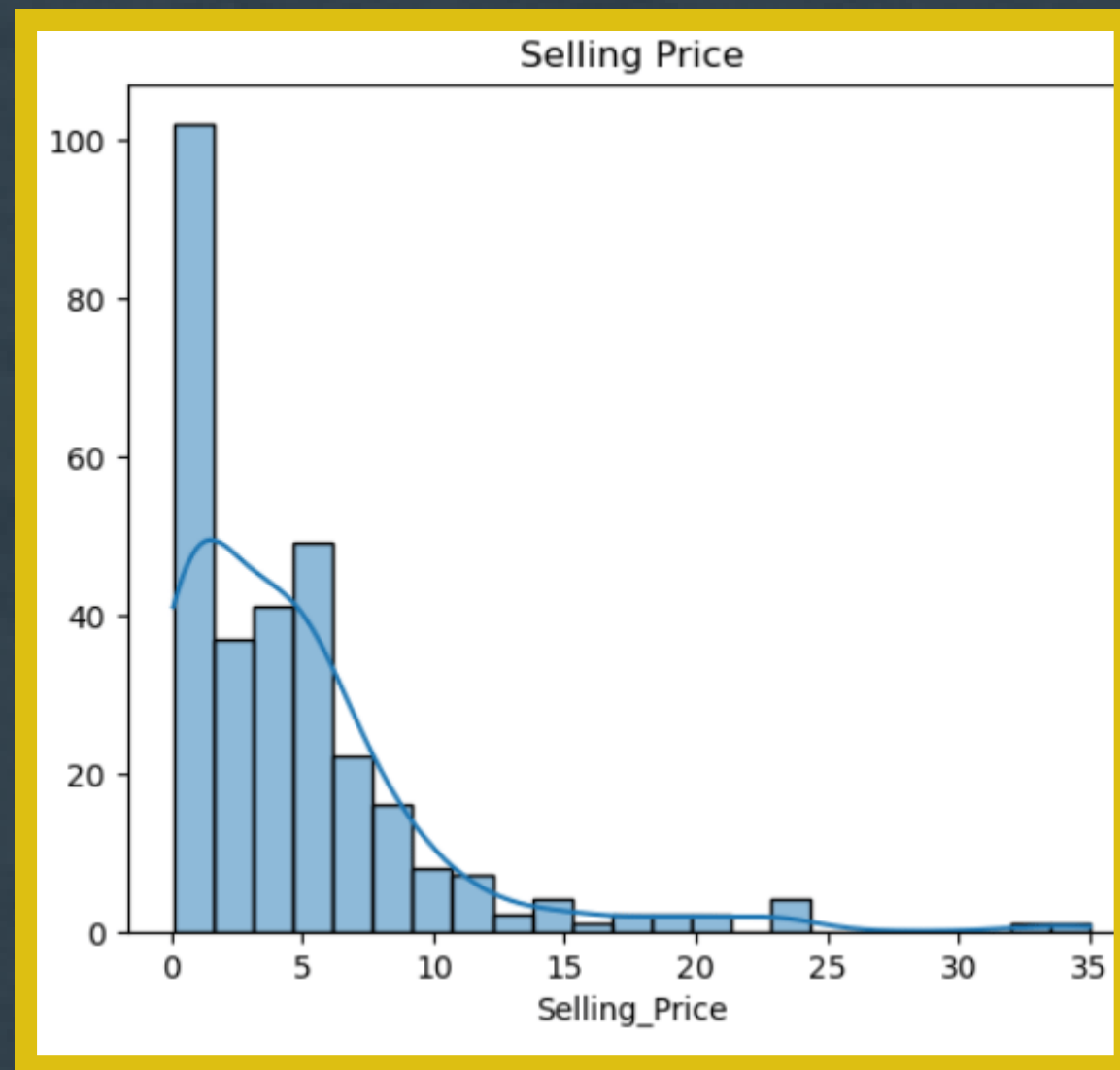


|      | C             | D             | E          | F         |             |
|------|---------------|---------------|------------|-----------|-------------|
|      | Selling_Price | Present_Price | Kms_Driven | Fuel_Type | Seller_Type |
| 2014 | 3.35          | 5.59          | 27000      | Petrol    | Dealer      |
| 2013 | 4.75          | 9.54          | 43000      | Diesel    | Dealer      |
| 2017 | 7.25          | 9.85          | 6900       | Petrol    | Dealer      |
| 2011 | 2.85          | 4.15          | 5200       | Petrol    | Dealer      |
| 2014 | 4.6           | 6.87          | 42450      | Diesel    | Dealer      |
| 2018 | 9.25          | 9.83          | 2071       | Diesel    | Dealer      |
| 2015 | 6.75          | 8.12          | 18796      | Petrol    | Dealer      |
| 2015 | 6.5           | 8.61          | 33429      | Diesel    | Dealer      |
| 2016 | 8.75          | 8.89          | 20273      | Diesel    | Dealer      |
| 2015 | 7.45          | 8.92          | 42367      | Diesel    | Dealer      |
| 2017 | 2.85          | 3.6           | 2135       | Petrol    | Dealer      |
| 2015 | 6.85          | 10.38         | 51000      | Diesel    | Dealer      |
| 2015 | 7.5           | 9.94          | 15000      | Petrol    | Dealer      |
| 2015 | 6.1           | 7.71          | 26000      | Petrol    | Dealer      |
|      | 2.25          | 7.21          | 77427      | Petrol    | Dealer      |
|      | 7.75          | 10.79         | 43000      | Diesel    | Dealer      |
|      | 7.25          | 10.79         | 41678      | Diesel    | Dealer      |
|      | 7.5           | 10.79         | 43000      | Diesel    | Dealer      |
|      | 5.5           | 5.09          | 35500      | CNG       | Dealer      |
|      | 7.98          | 7.98          | 41442      | Petrol    | Dealer      |
|      | 3.95          | 3.95          | 25000      | Petrol    | Dealer      |
|      | 5.9           | 5.71          | 2400       | Petrol    | Dealer      |
|      | 4.4           | 8.01          | 50000      | Petrol    | Dealer      |
|      | 2.5           | 3.46          | 45280      | Petrol    | Dealer      |
|      | 4.41          | 4.41          | 56879      | Petrol    | Dealer      |
|      | 4.99          | 4.99          | 20000      | Petrol    | Dealer      |

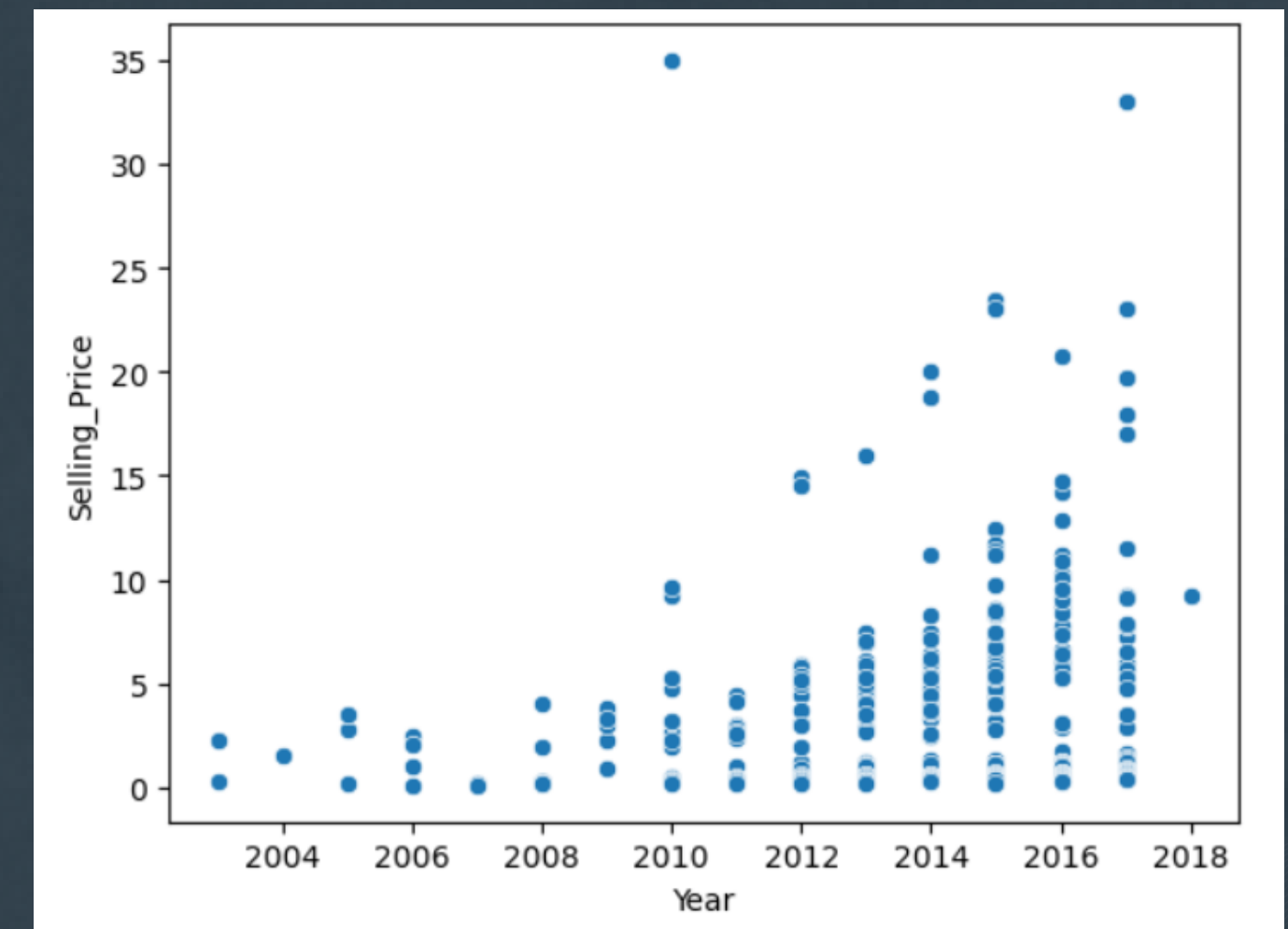
NEW\_CAR

# DATA VISUALISATION

Corrélation des  
variables



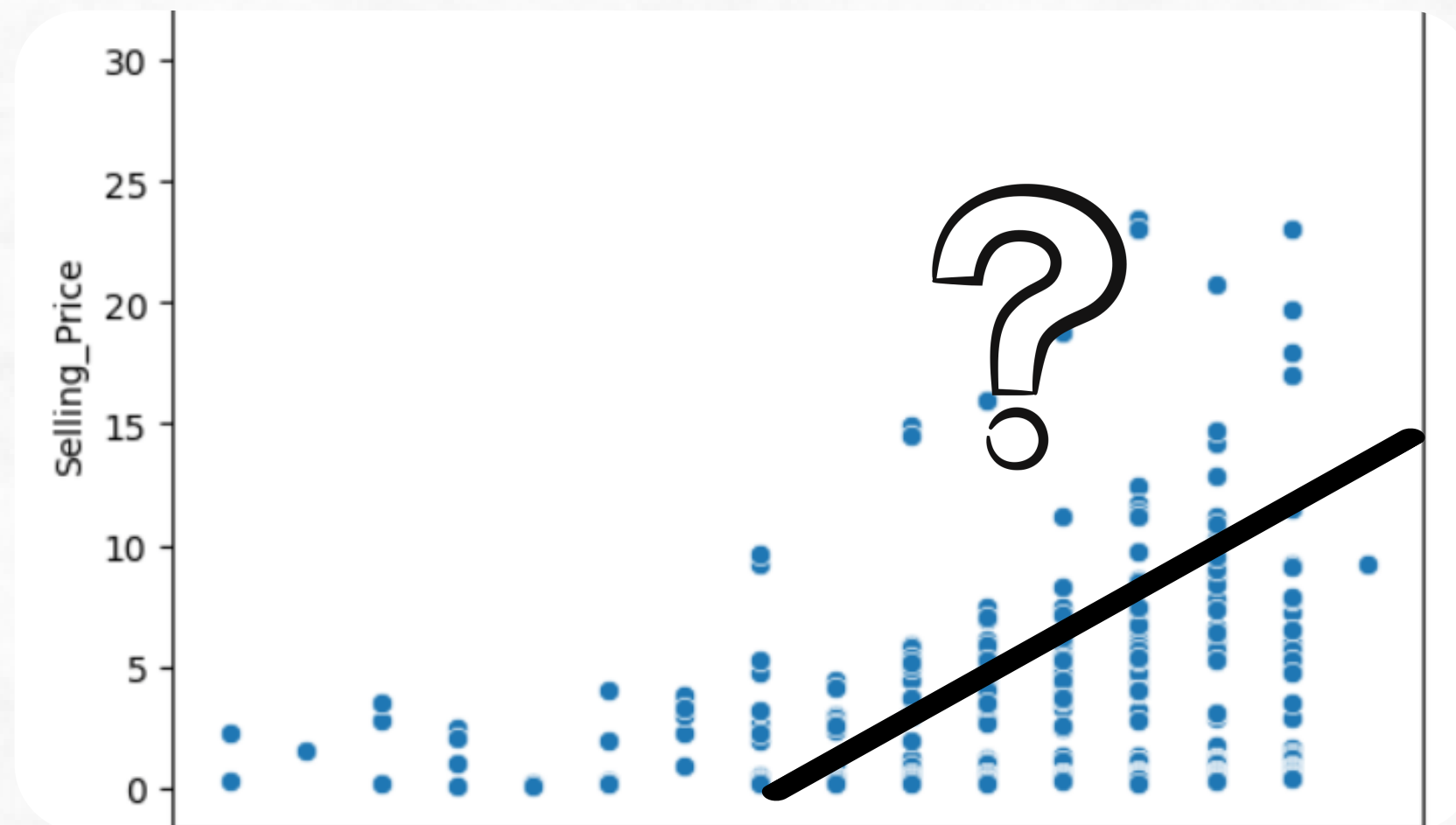
Analyse de  
répartition





NEW\_CAR

# IDENTIFICATION DES TENDANCES

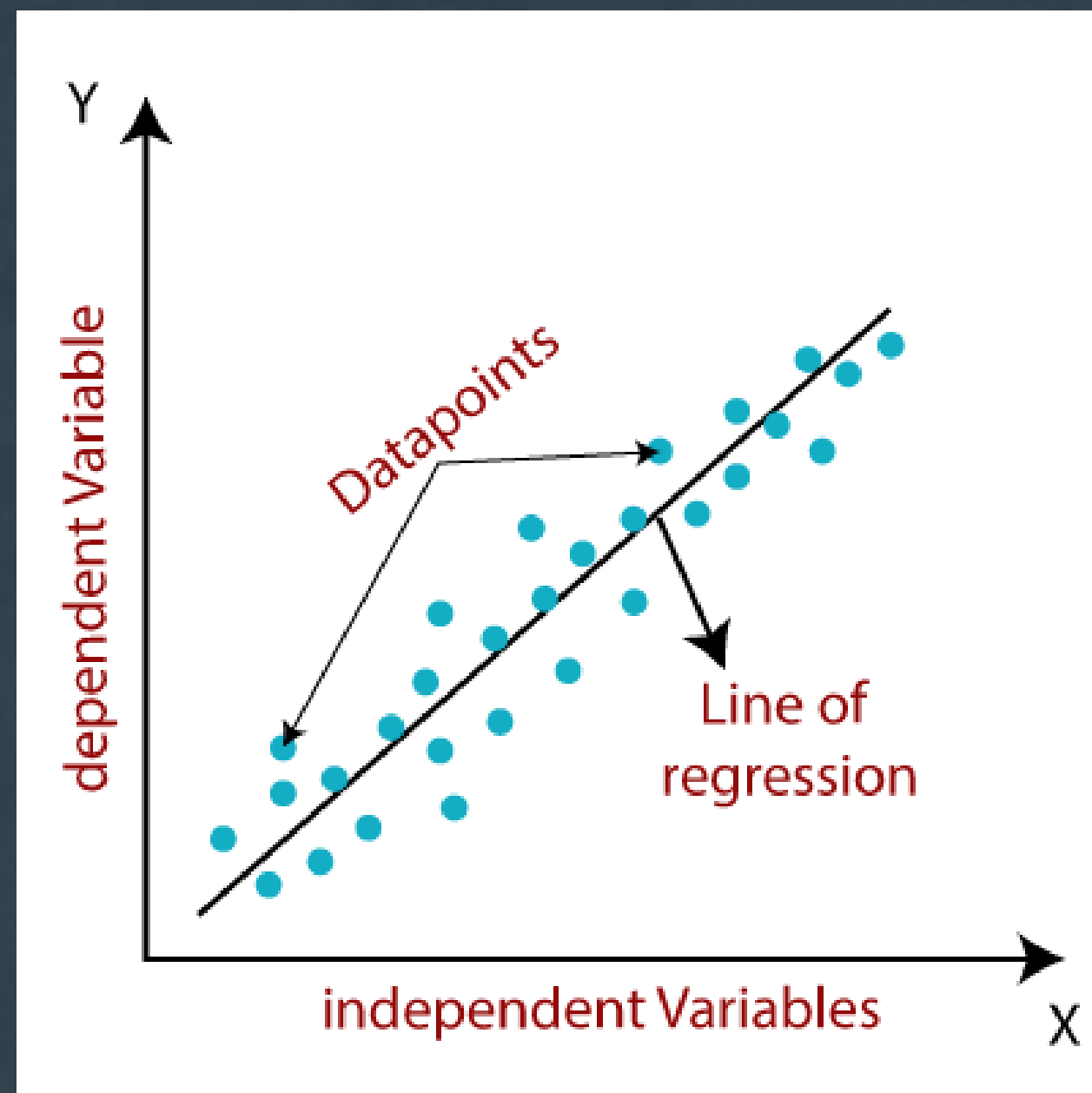


## Tendances linéaires ?

Une tendance linéaire peut indiquer qu'une régression linéaire peut être utilisée à des fins de prédiction

NEW\_CAR

# RÉGRESSION LINÉAIRE ?



## Qu'est-ce que c'est ?

C'est une droite dont l'équation est du format  $ax + b$  passant par le plus de datapoints possible

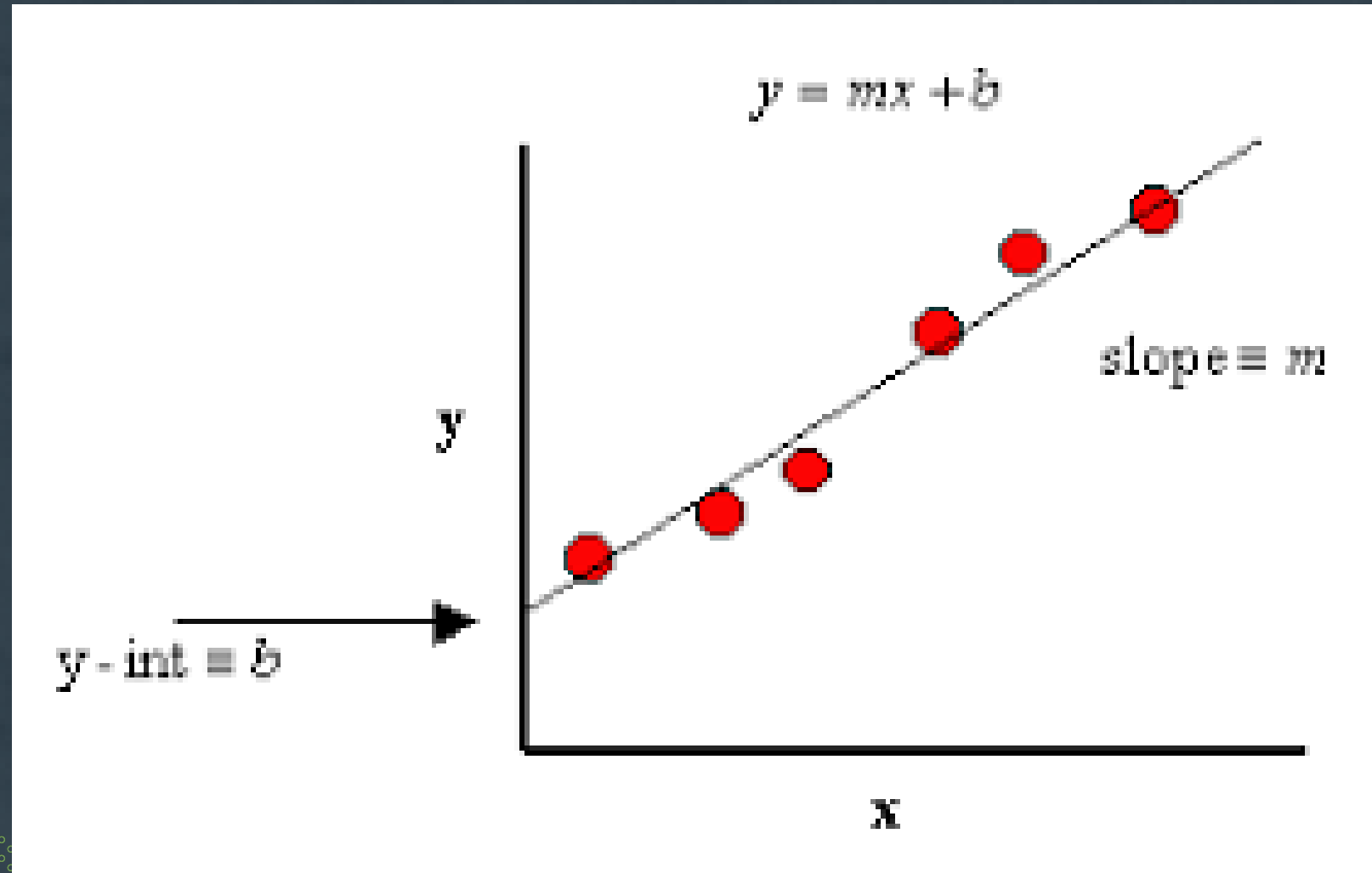
## A quoi ça sert ?

- Etablir une relation entre claire et quantifiable entre les variables du dataset
- Utiliser l'équation de la droite ainsi obtenue pour faire de la prédiction

NEW\_CAR

# REGRESSION LINÉAIRE ?

*m*



## Comment ça marche ?

- Tests de coefficients a et b
- Calculer l'erreur entre prédiction et les données d'entraînement
- Réajuster les coefficients a et b

NEW\_CAR

# ANALYSE DES RESULTATS



UNIVARIEE

$$R^2 = 6,5\%$$

- Score de régression très faible
- Données trop étalées



MULTIVARIEE

$$R^2 = 26,1\%$$

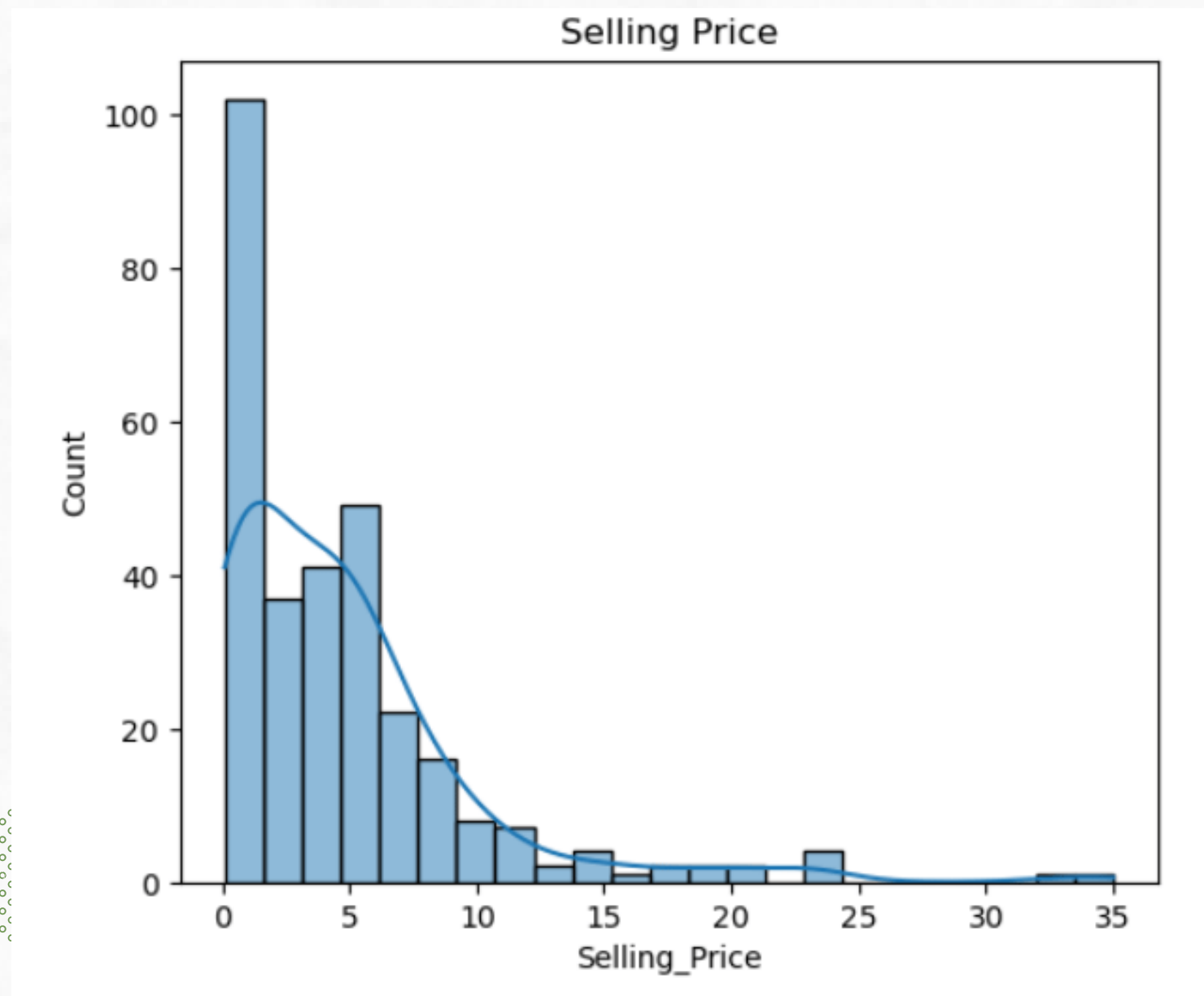
- Score de régression bas
- Résultats inexploitable pour faire des prédictions





NEW\_CAR

# METRICS DE QUALITÉ



RMSE

Erreur quadratique  
moyenne



MAE

Erreur absolue moyenne

NEW\_CAR

# CONCLUSIONS



## RESULTATS NON CONCLUANTS

Les métriques de qualité de régression indiquent que les résultats sont peu probants

Plusieurs facteurs peuvent en être la cause :

- Données trop générales
- Facteurs externes non pris en compte
- Jeu de donnée trop petit
- ...



NEW\_CAR

# MERCI

POUR VOTRE ATTENTION

