*Université Abdelmalek Essaadi Faculté des Sciences et techniques de Tanger Département Génie Informatique*



Rapport sur Atelier 1 :

« Régression »

Réalisé par :

***LASIRI Nouhaila***

***BENSLIMAN Hamza***

Encadré par :

***Pr . ELAACHAk LOTFI***

*Cycle Ingénieur: LSI S4*

*Machine Learning*

*Année : 2019/2020*

Objectif :

L’objectif principal de cet atelier est de pratiquer les deux concepts de la régression : la régression linière simple et la régression linière multiple, en traitant des données de plusieurs Data Sets

Outils :

Visual Studio Code, Anaconda, Python, Pandas, Sklearn, matplotlib.

Data Sets :

Expérience et Salaire : <https://www.kaggle.com/rohankayan/years-of-experience-and-salarydataset>

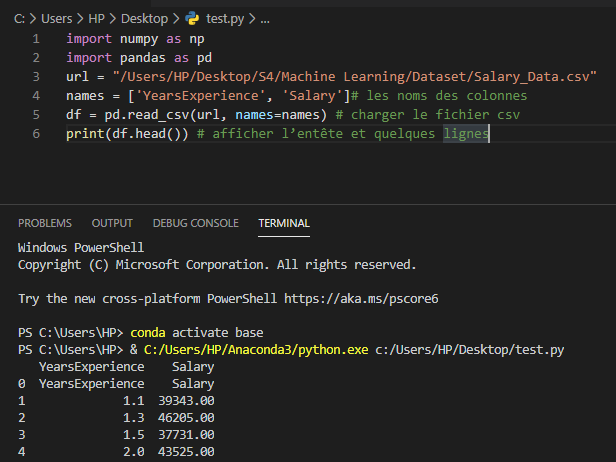
Assurance :

<https://www.kaggle.com/sinaasappel/tutorial-multiple-regression/data>

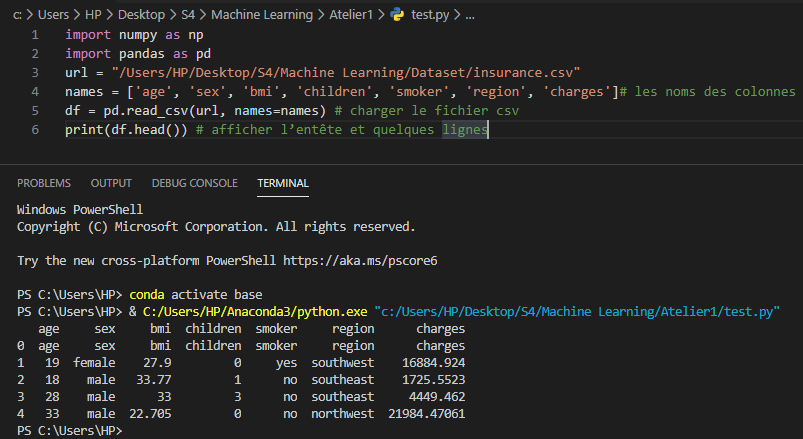
Partie 1 (Data Visualisation):

1. Exploration des données :

*Expérience et Salaire :*

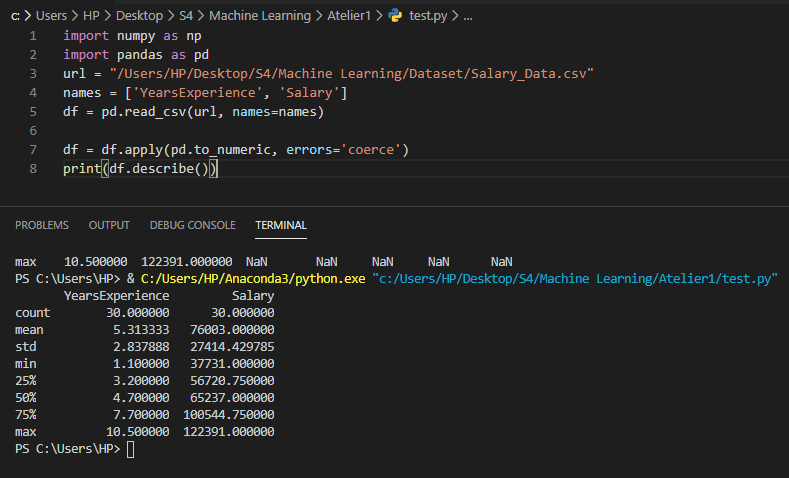


*Assurance :*



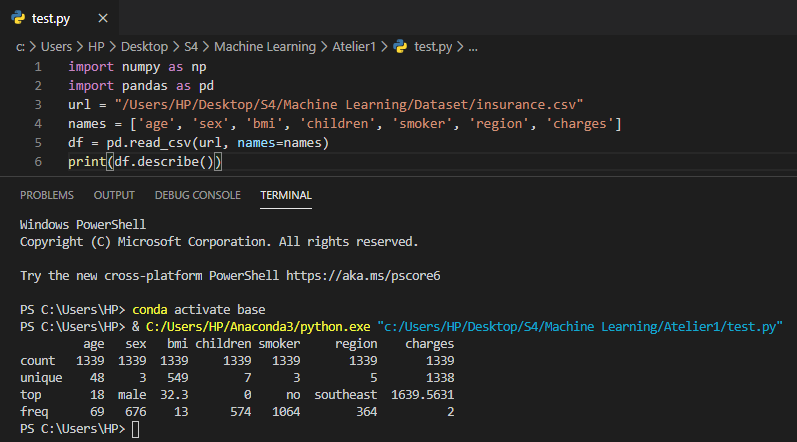
1. Le résumé statistique :

*Expérience et Salaire :* (df = df.apply(pd.to\_numeric, errors='coerce')) pour convertir tous les données aux données numériques, pour qu’il s’affiche tous les paramètres du résumé statistique.

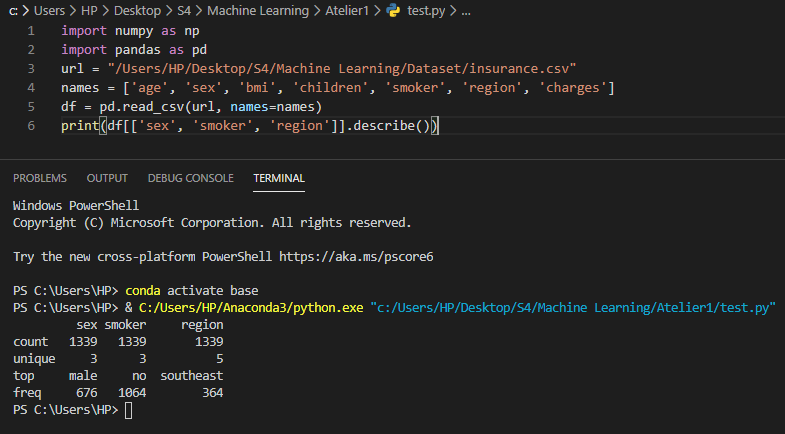


*Assurance :*

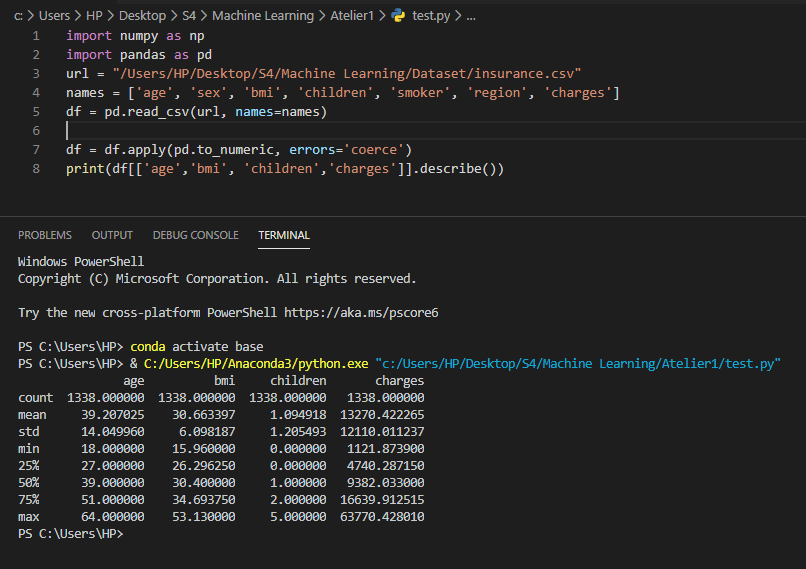
* Les données sont du format string et numérique donc seulement « count, unique, top, freq » s’affichent.



* Extraire les données du format string.

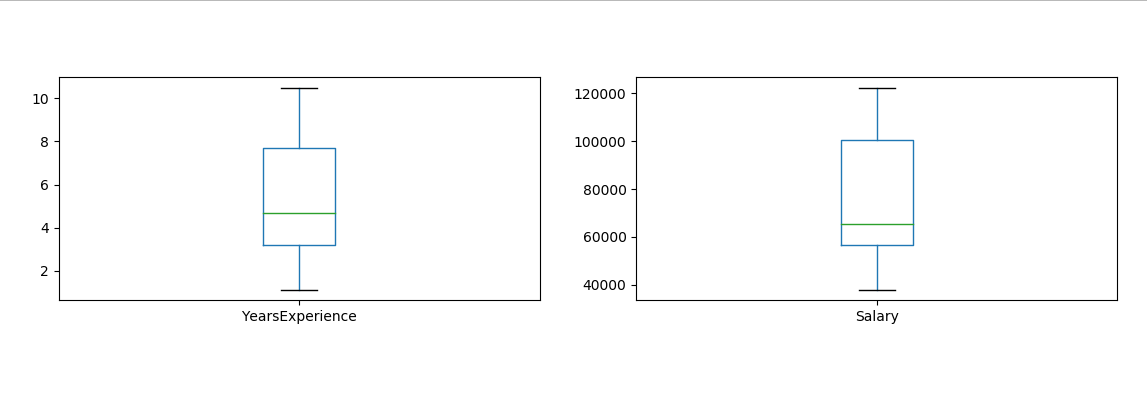


* Extraire les données du format numérique.

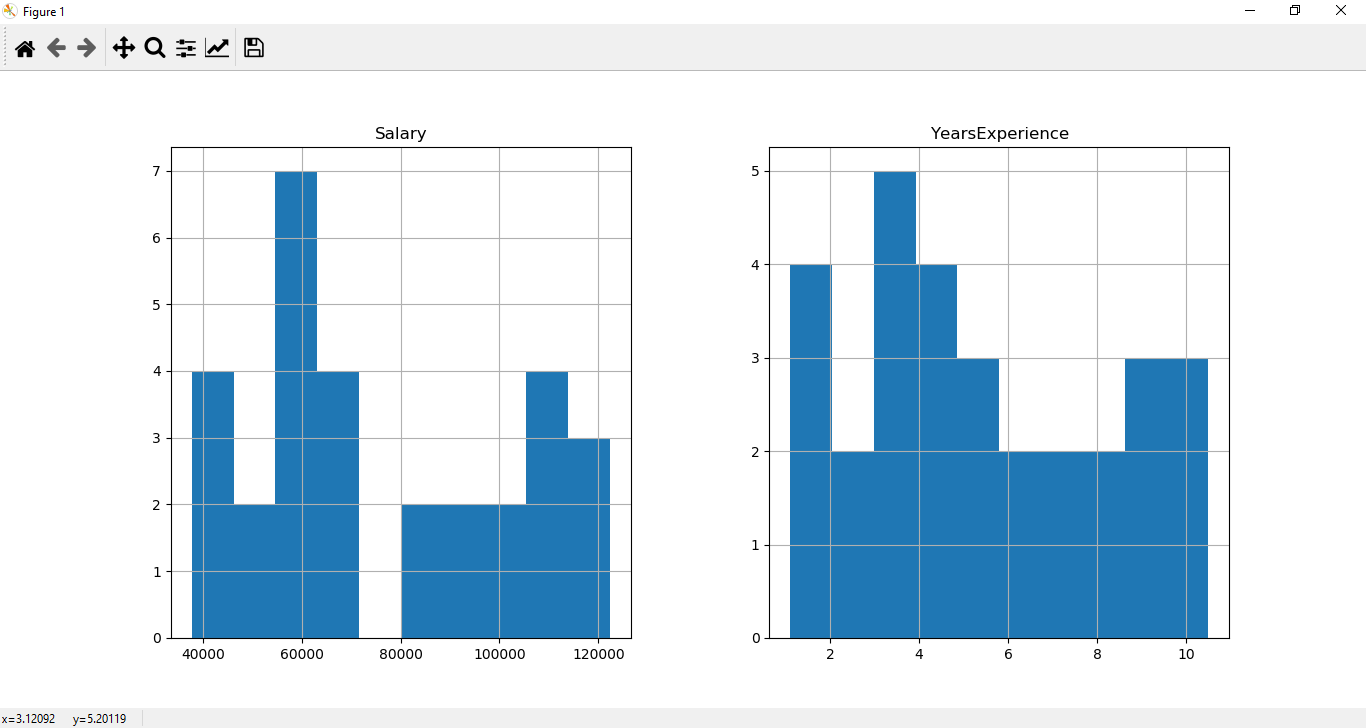


3- Les nuages des points du dataset « Expérience / Salaire » :

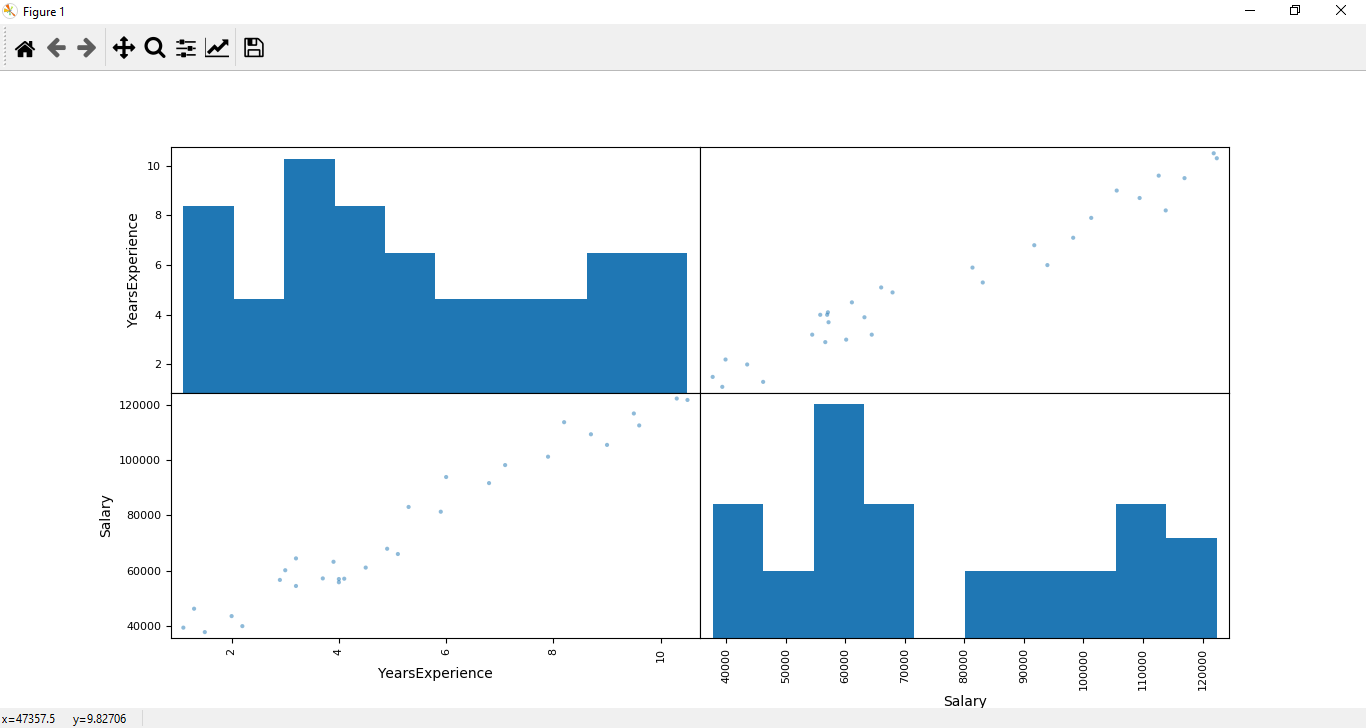
* Boite à moustache :



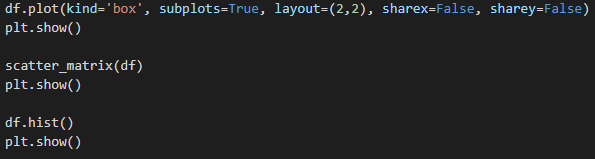
* Histogramme :



* Scatter plot matrix :



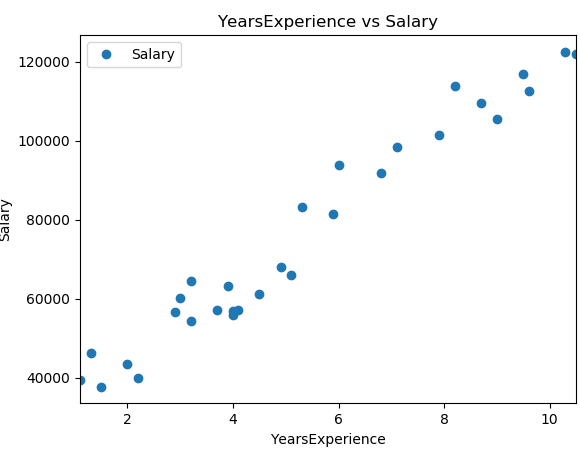
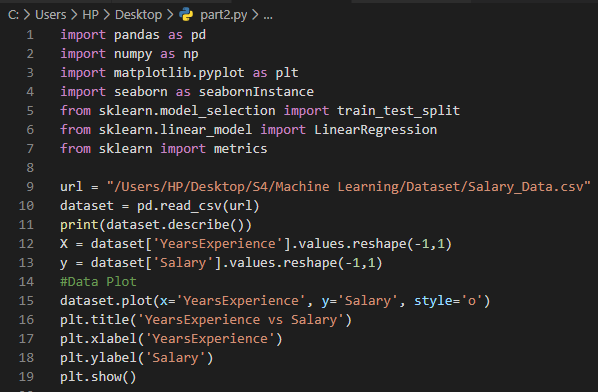
* Source Code :



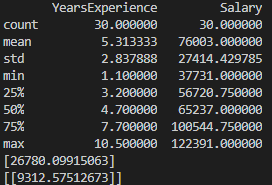
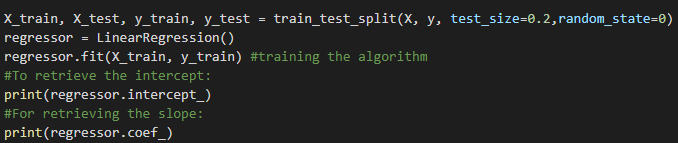
Partie 2 (Régression Simple cas Expérience Salaire):

1. Entraînement du modèle :

* Affichage :

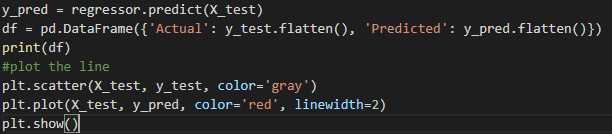


* Entraînement :

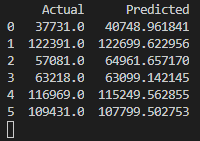


2- Prédiction des données :

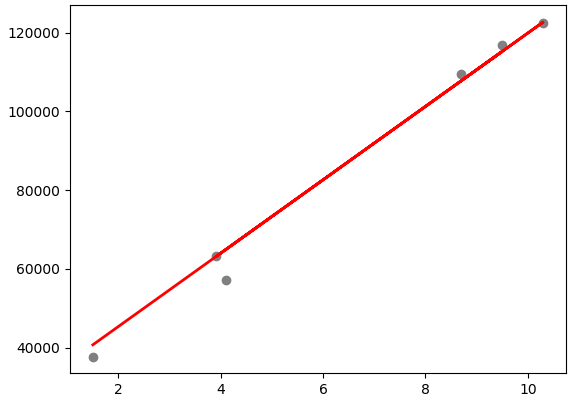
* Source Code :



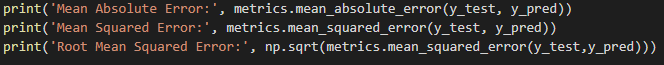
* Résultats :



3- Affichage des résultats sous forme d’un graphe :



4- Evaluation du modèle :



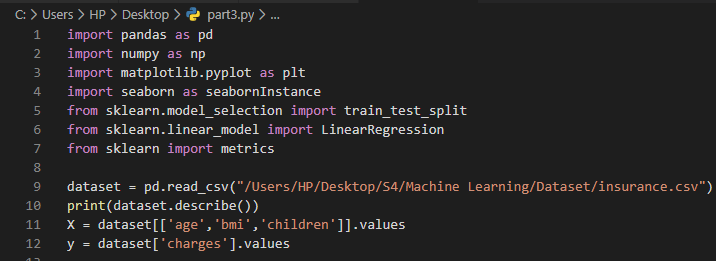
* Résultat :



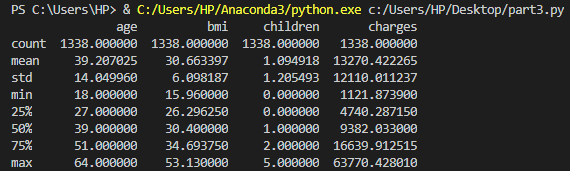
Partie 3 (Régression multiple cas d’assurance):

1. Entraînement du modèle :

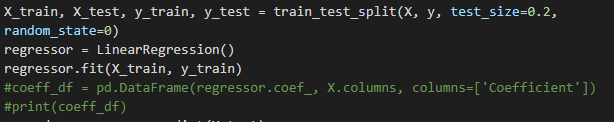
* Affichage :



* Resultat :

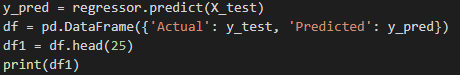


* Entrainement :

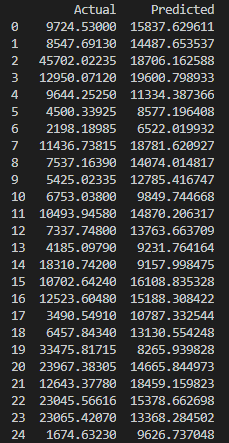


1. Prédiction des données :

* Source Code :

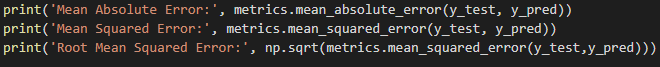


* Resultat :



4- Evaluation du modèle :

* Source Code :



* Resultat :

