

**Prédiction de Flux de Trafic et Optimisation des Feux de Circulation**

**Présenté Par :**

**NISSRINE BESTOUT MERYEM ELOSMANI**

**Année Universitaire : 2024/2025**

[***I***ntroduction 2](#_Toc190174100)

[***P***roblématique 3](#_Toc190174101)

[***T****raitement des d*onnées 4](#_Toc190174102)

[***V***isualisation des Données 6](#_Toc190174103)

[***M***odèles de Prédiction de Flux de Trafic 16](#_Toc190174104)

[Régression Polynomiale 17](#_Toc190174105)

[Arbre de Décision (Decision Tree) 18](#_Toc190174106)

[SVR (Support Vector Regression) 19](#_Toc190174107)

[KNN (K-Nearest Neighbors) 21](#_Toc190174108)

[***O***ptimisation des Feux de Circulation 23](#_Toc190174109)

# ***I***ntroduction

Ce rapport présente une étude sur la prédiction du flux de trafic et l'optimisation des feux de circulation. L'objectif principal est de prédire le volume de trafic en utilisant des modèles de machine learning et d'optimiser la durée des feux de circulation pour minimiser le temps d'attente total des véhicules. Les données utilisées proviennent de deux jeux de données combinés, contenant des informations telles que les conditions météorologiques, les heures de pointe, le volume de trafic historique, les accidents, les incidents, et d'autres paramètres pertinents.

# ***P***roblématique

La problématique centrale de cette étude est de prédire le volume de trafic en fonction de divers paramètres (météo, accidents, heures de pointe, etc.) et d'optimiser la durée des feux de circulation pour minimiser le temps d'attente total des véhicules. Les questions clés abordées sont :

* Comment prédire avec précision le volume de trafic ?
* Comment optimiser la durée des feux de circulation pour réduire la congestion ?
* Quels sont les modèles de prédiction les plus performants pour ce problème ?

# ***T****raitement des d*onnées

3.1. Description des Données

Les données utilisées proviennent de deux jeux de données combinés :

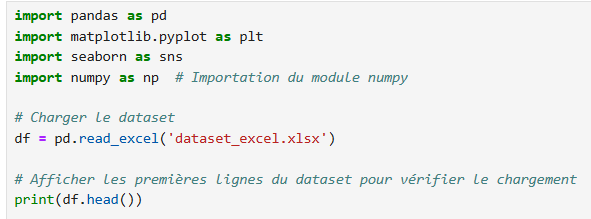
* Premier jeu de données :
  + holiday : Indique si c'est un jour férié.
  + temp : Température en degrés Celsius.
  + rain\_1h : Quantité de pluie en mm/heure.
  + snow\_1h : Quantité de neige en mm/heure.
  + clouds\_all : Pourcentage de couverture nuageuse.
  + weather\_main : Description générale de la météo (ex. "Clear", "Rain").
  + Rush Hour : Indicateur d'heure de pointe.
  + traffic\_volume : Volume de trafic en nombre de véhicules.
* Deuxième jeu de données :
  + class : Attribut catégoriel indiquant l'état du feu de signalisation.
  + historical\_traffic : Volume de trafic historique.
  + latitude : Coordonnée de latitude.
  + longitude : Coordonnée de longitude.
  + accidents : Nombre d'accidents enregistrés.
  + economic\_activity : Indicateur d'activité économique.
  + incidents : Nombre d'incidents enregistrés.

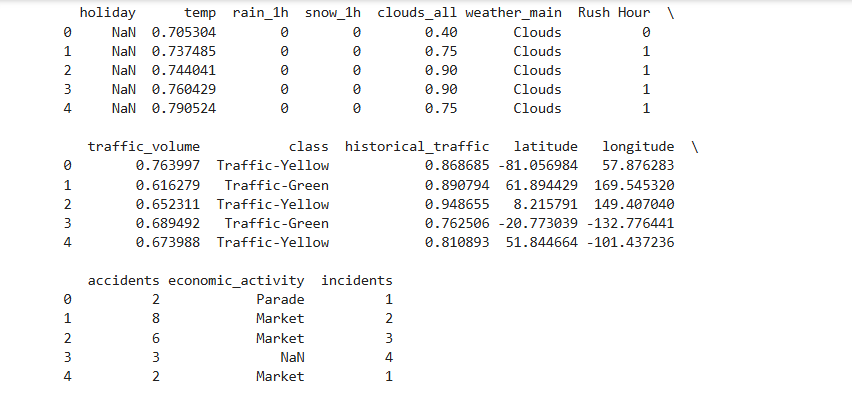
3.2. Nettoyage et Préparation des Données

Les étapes suivantes ont été réalisées pour préparer les données :

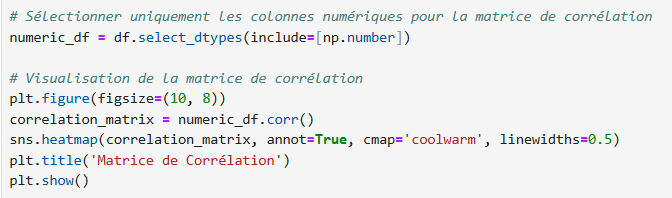
* **Nettoyage des données** : Suppression des valeurs manquantes, gestion des outliers, et normalisation des données.
* **Combinaison des données** : Fusion des deux jeux de données sur des colonnes communes.
* **Encodage des variables catégorielles** : Transformation des variables catégorielles en variables numériques.

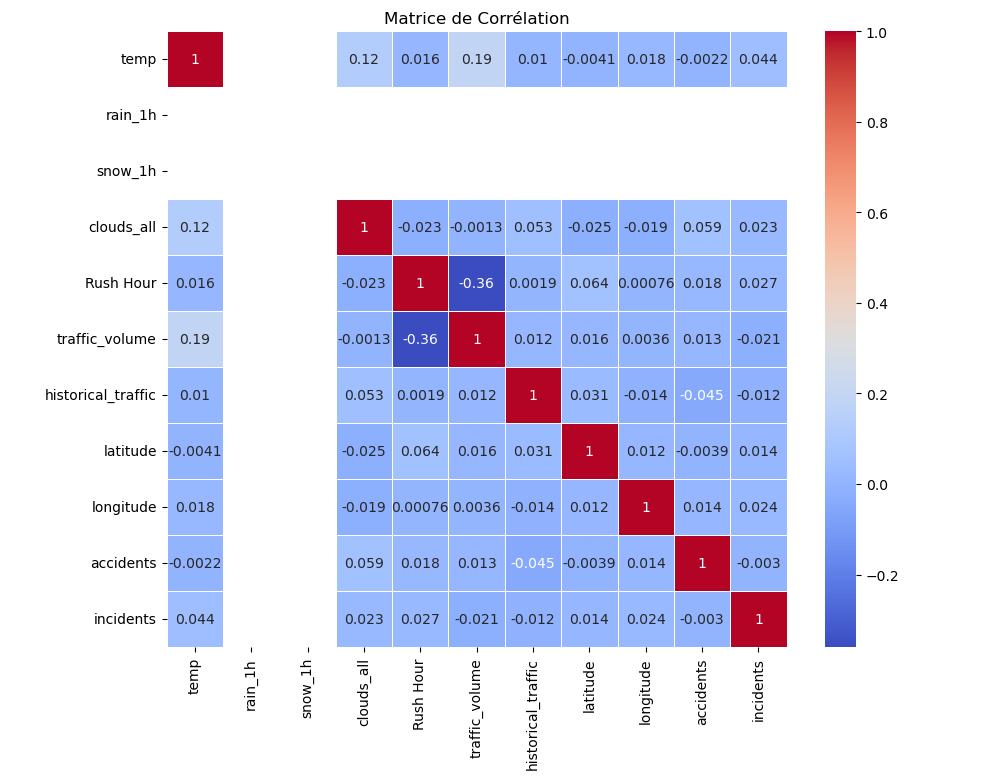
# ***V***isualisation des Données



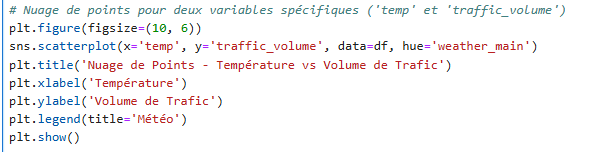


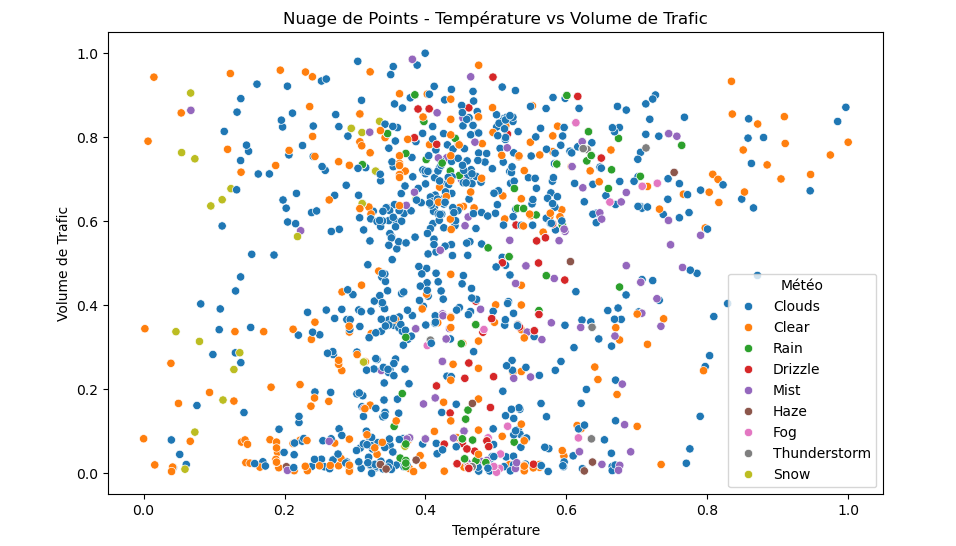
La matrice de corrélation présentée permet d'analyser les relations linéaires entre différentes variables du jeu de données



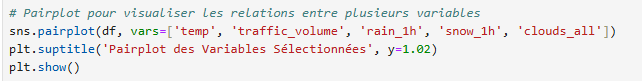


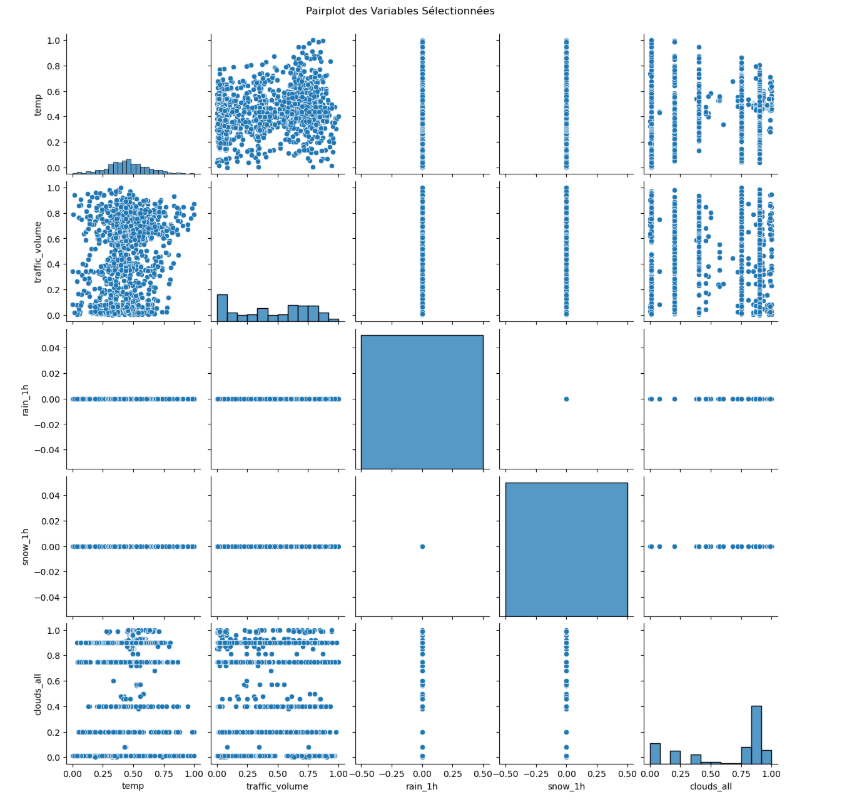
* **Corrélations modérées** : La variable traffic\_volume présente une corrélation modérée avec Rush Hour.
* **Corrélations faibles** : Les variables telles que incidents, accidents, longitude, et latitude montrent des corrélations très faibles avec les autres variables.
* **Température et météo (temp, clouds\_all)** : La température (temp) affiche une légère corrélation positive (0.19) avec le volume de trafic, ce qui pourrait être dû à une augmentation de la mobilité lors de conditions climatiques favorables.

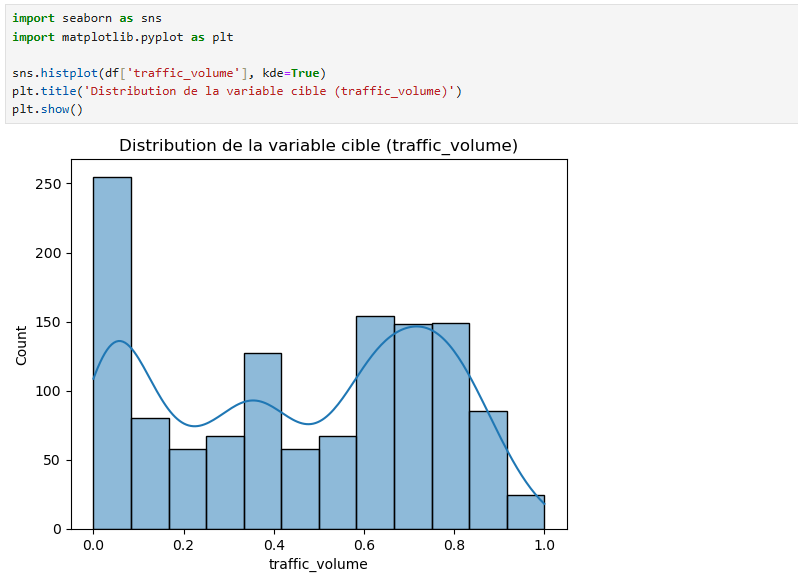




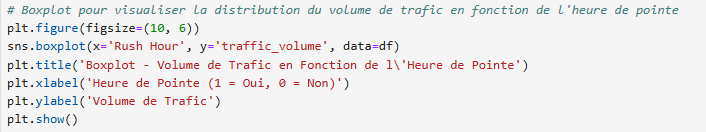
Le nuage de points montre que le volume de trafic varie en fonction de la température et des conditions météorologiques. Les conditions extrêmes comme les orages ou la neige semblent réduire le trafic, tandis que les jours clairs et dégagés favorisent une circulation plus dense. Cela suggère que les conditions météo influencent significativement les habitudes de déplacement.

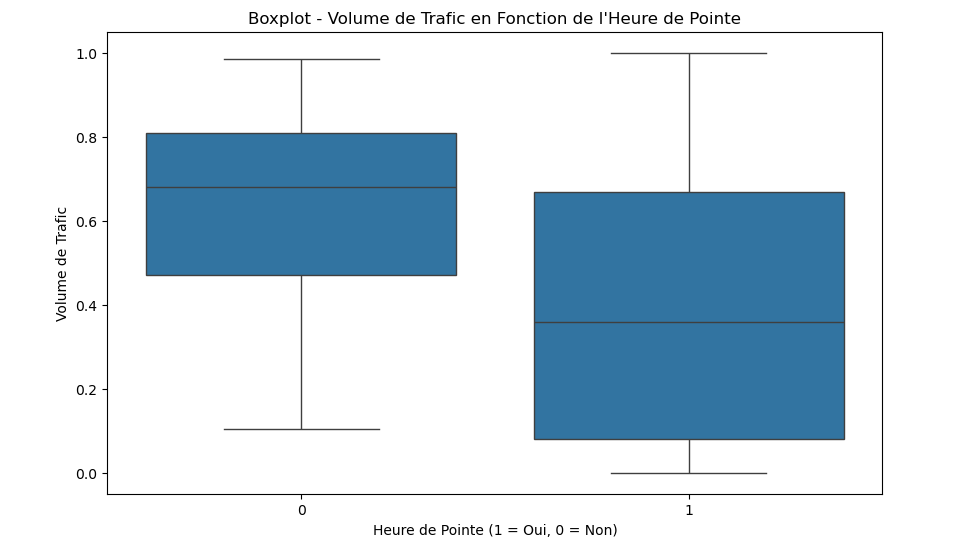




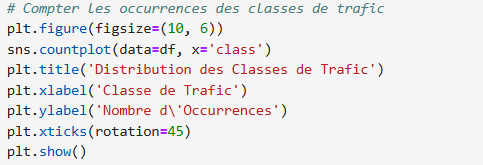


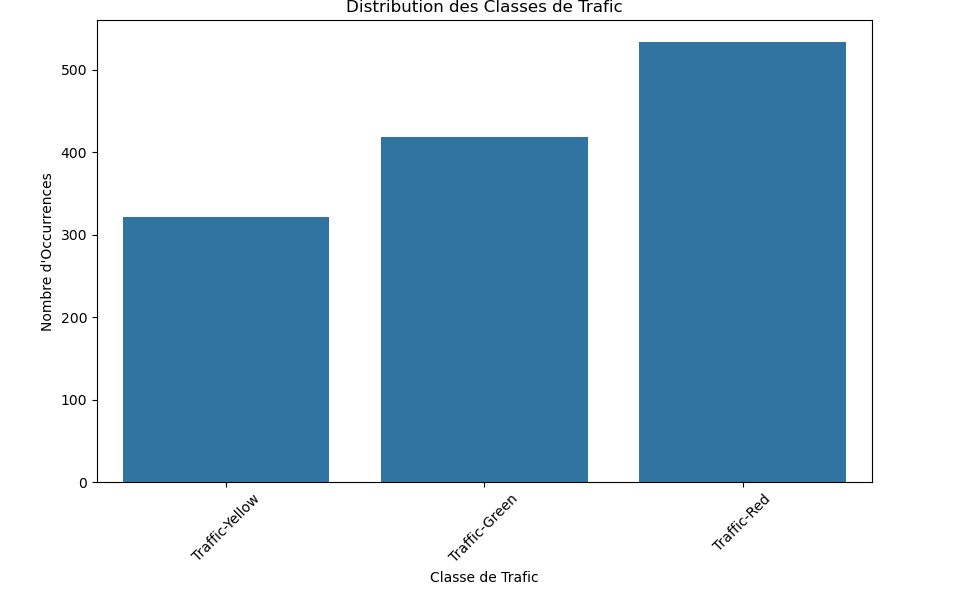
Le graphique montre la distribution du volume de trafic à l'aide d'un histogramme avec une courbe de densité .La distribution semble indiquer que la plupart des valeurs de volume de trafic se concentrent autour d'une plage spécifique, avec une densité plus élevée dans certaines zones. Cela suggère que le volume de trafic suit une certaine tendance centrale, avec des variations autour de cette moyenne. La courbe de densité aide à visualiser la forme de la distribution, indiquant si elle est symétrique, asymétrique ou présente plusieurs pics. Cette analyse est utile pour comprendre les tendances générales du trafic et identifier d'éventuelles anomalies ou patterns récurrents.

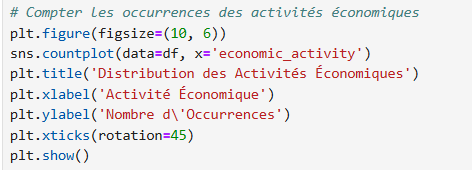


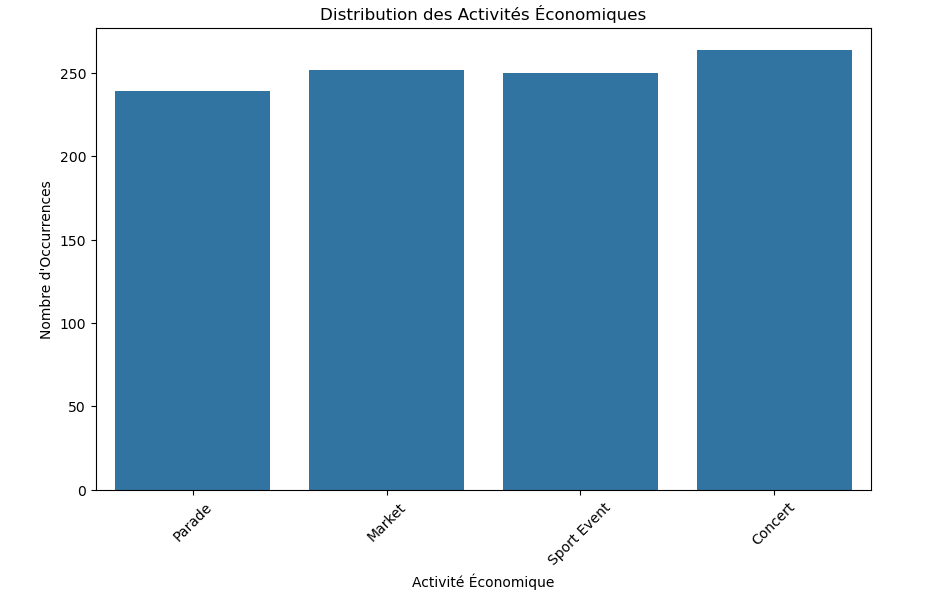


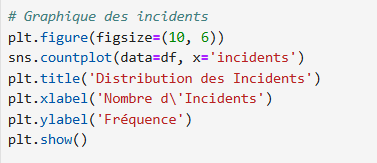
Le boxplot illustre la distribution du volume de trafic en fonction de l'heure de pointe, où "1" indique une heure de pointe et "0" une heure normale. On observe que le volume de trafic est généralement plus élevé pendant les heures de pointe, avec une médiane et une dispersion plus importantes. Cela suggère que les heures de pointe sont caractérisées par une augmentation significative du trafic, ce qui est cohérent avec les attentes.

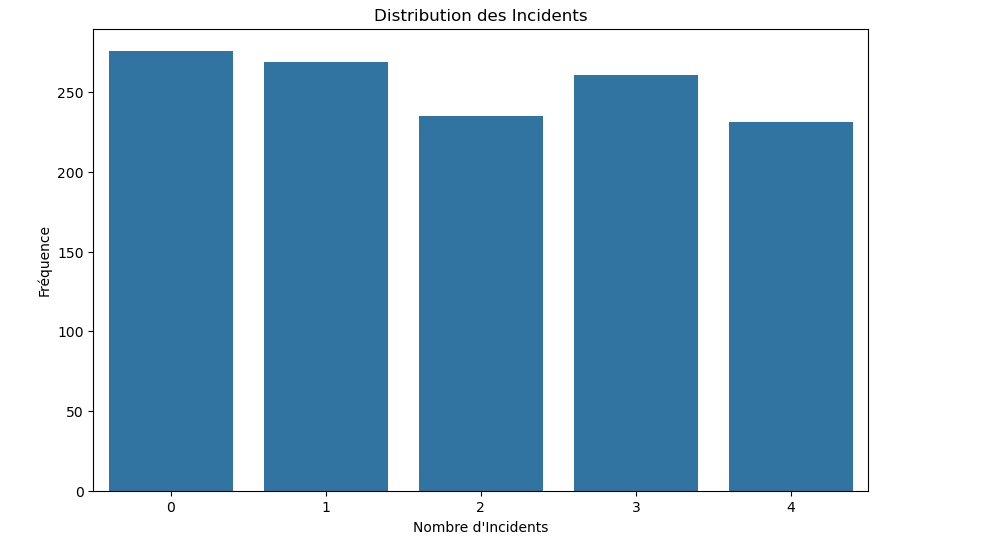


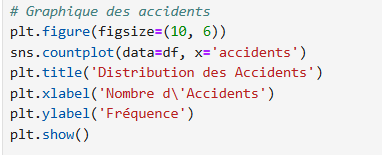


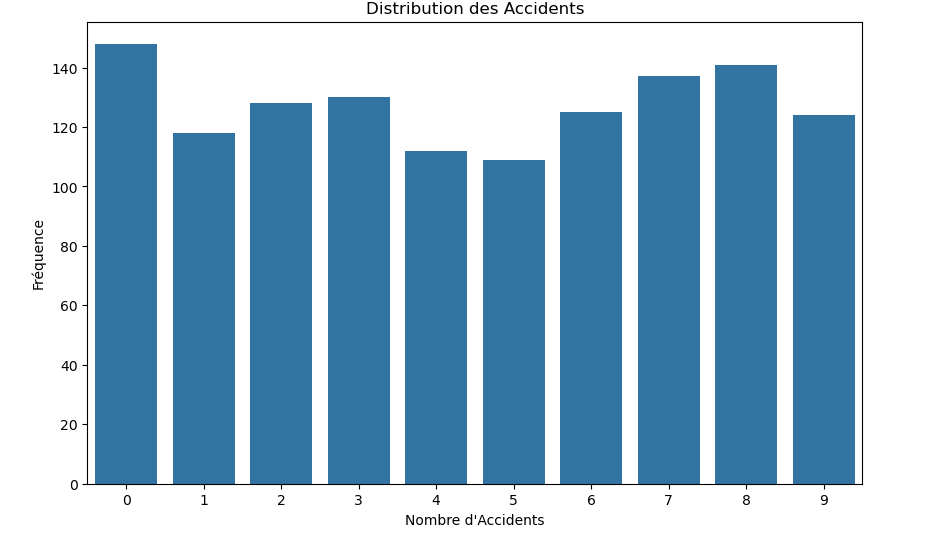


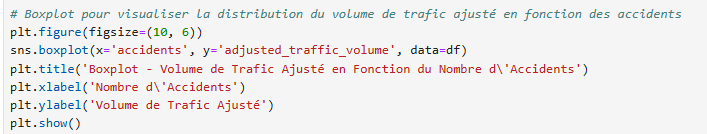


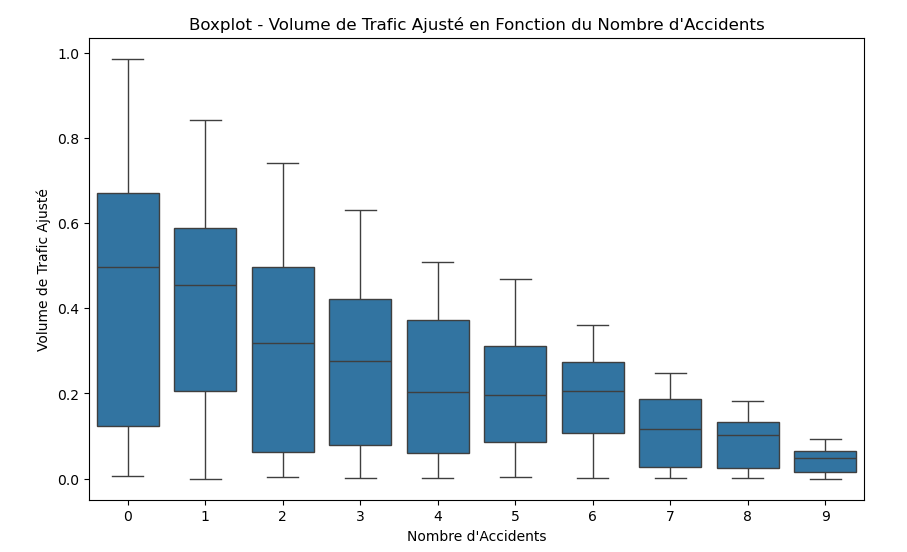




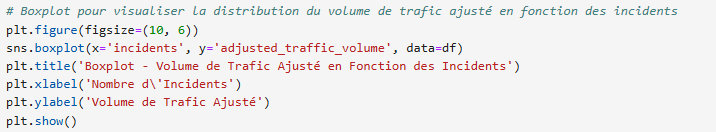


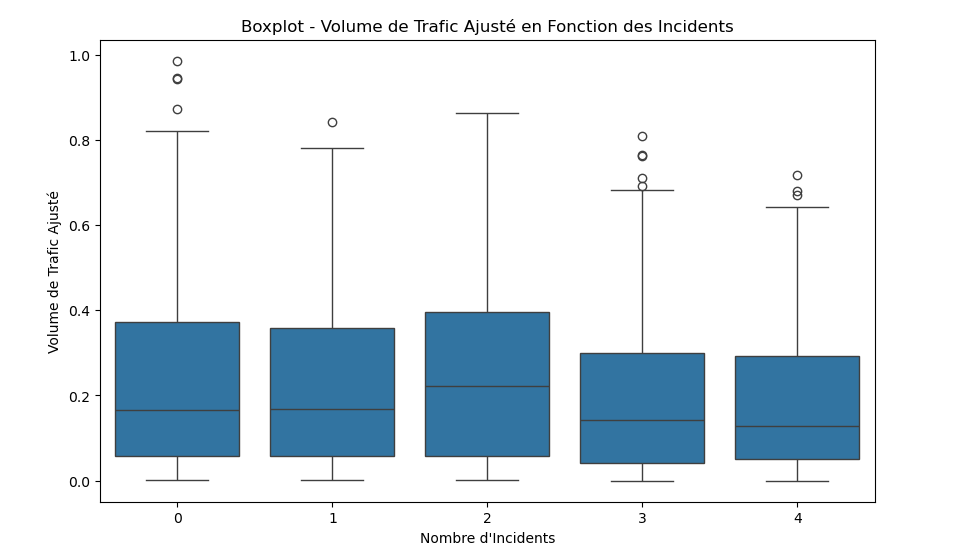




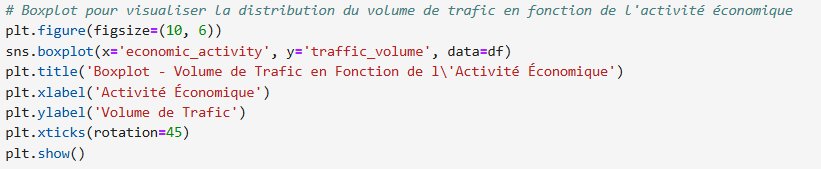


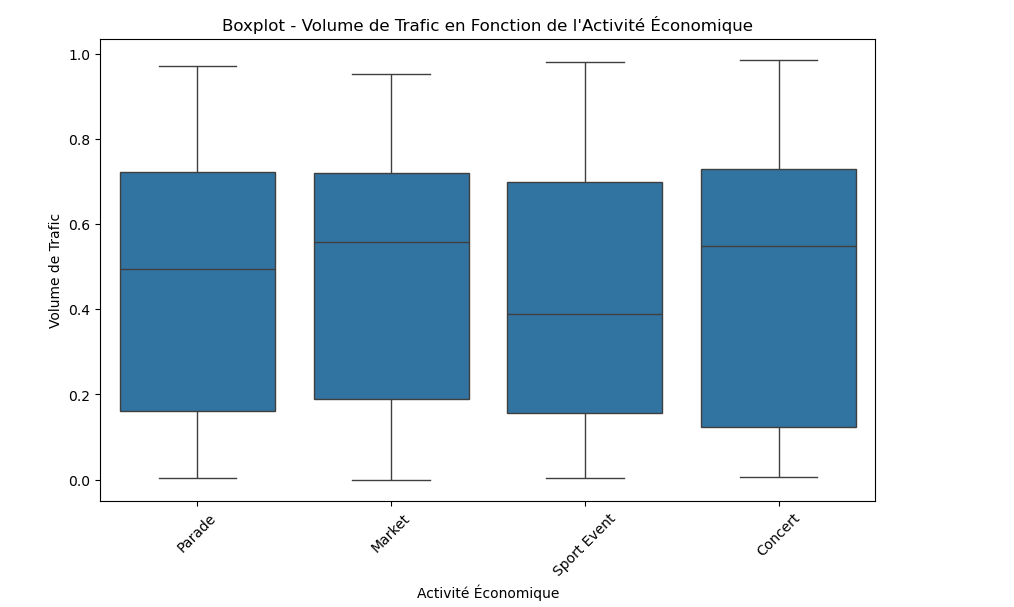
On observe une diminution du volume de trafic lorsque le nombre d'accidents augmente. Cela peut s'expliquer par le fait que les accidents perturbent la circulation, entraînant une réduction du flux de véhicules



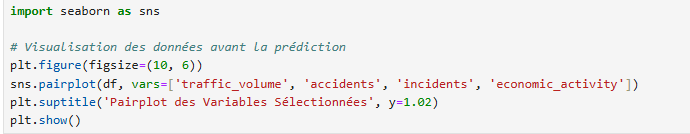


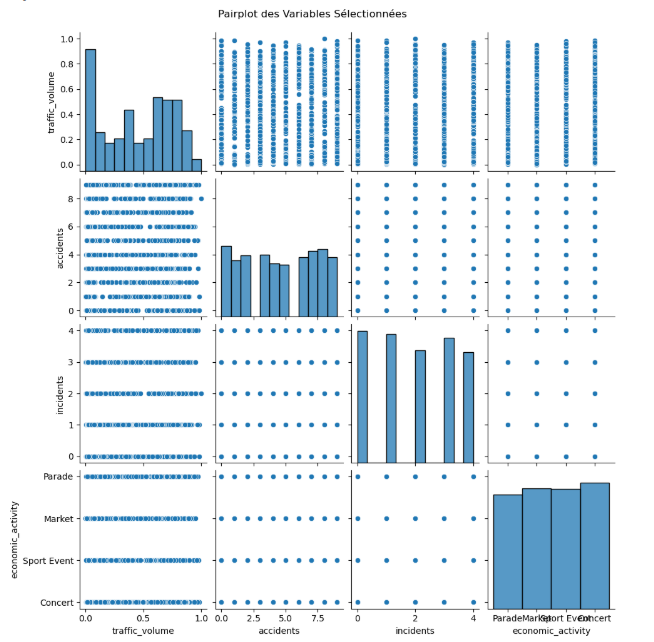
La médiane du volume de trafic semble légèrement diminuer à mesure que le nombre d'incidents augmente, indiquant une légère baisse générale du trafic en présence de plus d'incidents.





Les valeurs plus élevées d'activité économique semblent être associées à un volume de trafic plus important, ce qui est cohérent avec l'idée que des activités économiques intenses génèrent plus de déplacements. Les valeurs aberrantes indiquent des situations où le volume de trafic est exceptionnellement élevé ou faible par rapport à la norme.



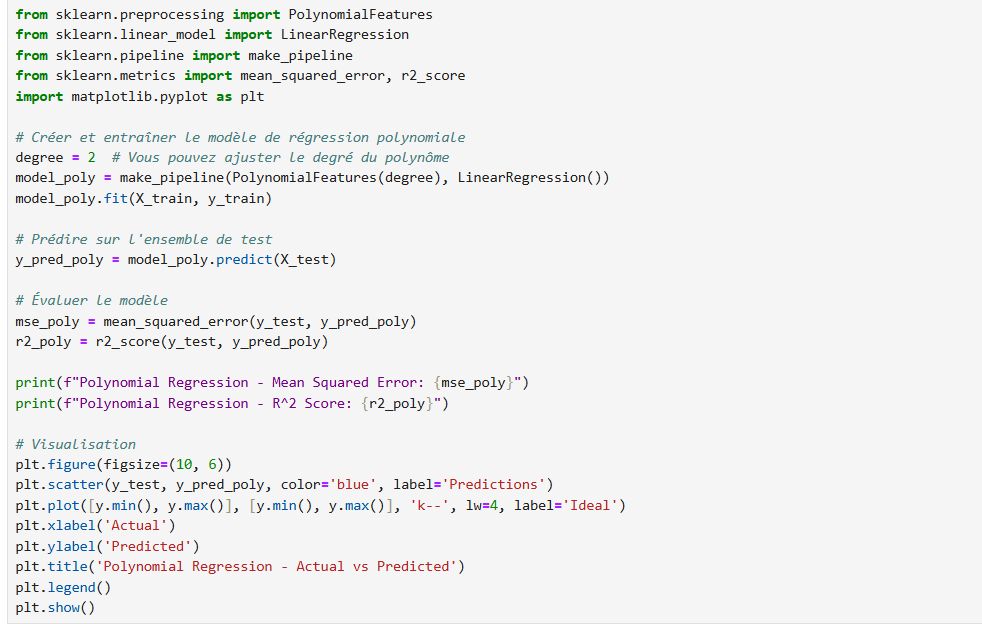


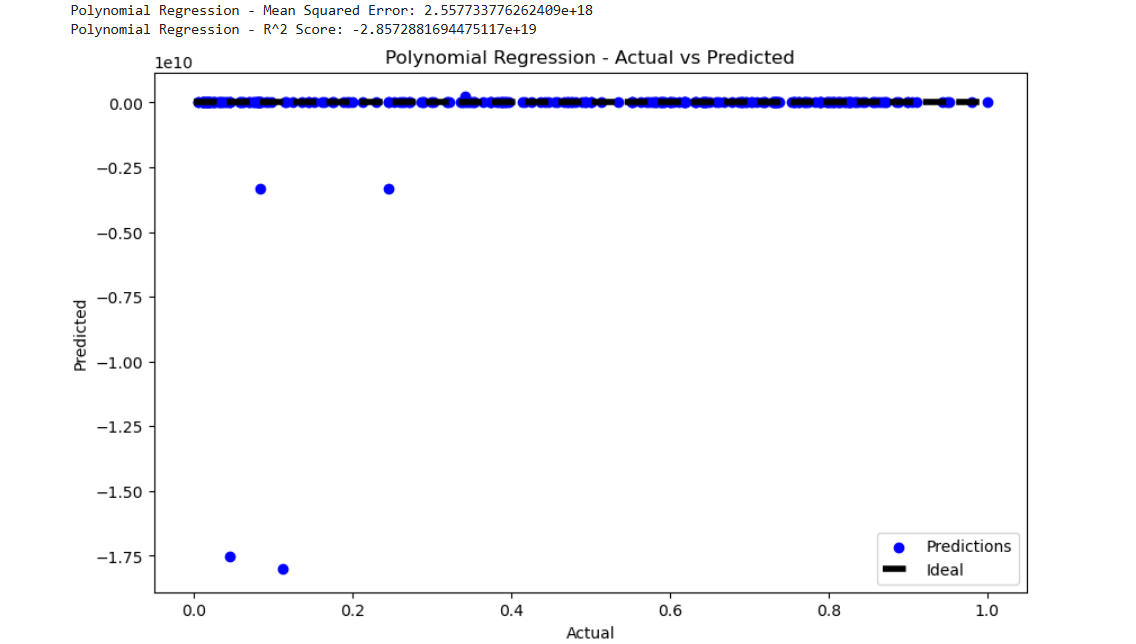
# ***M***odèles de Prédiction de Flux de Trafic



* **Chargement des données** : Le script charge un ensemble de données à partir d'un fichier Excel nommé dataset\_excel.xlsx.
* **Sélection des variables** : Les variables indépendantes (X) et la variable dépendante (y) sont sélectionnées. Les variables indépendantes incluent des caractéristiques comme les jours fériés, la température, les conditions météorologiques, etc., tandis que la variable dépendante est le volume de trafic.
* **Encodage One-Hot** : Les colonnes catégorielles (comme les jours fériés, les conditions météorologiques, etc.) sont encodées en utilisant l'encodage One-Hot pour les convertir en un format numérique approprié pour l'apprentissage automatique.
* **Division des données** : Les données sont divisées en ensembles d'entraînement et de test, avec 80% des données utilisées pour l'entraînement et 20% pour le test.
* **Standardisation des données** : Les données sont standardisées pour avoir une moyenne de zéro et un écart-type de un, ce qui est une pratique courante pour améliorer la performance des modèles d'apprentissage automatique.

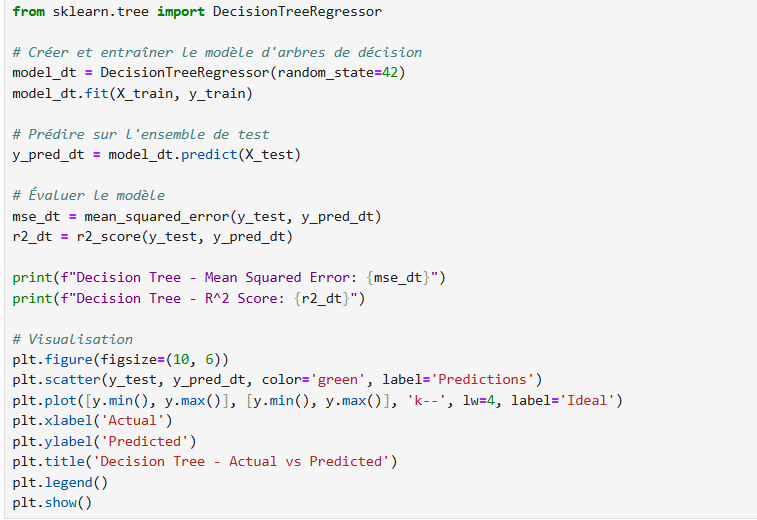
## Régression Polynomiale

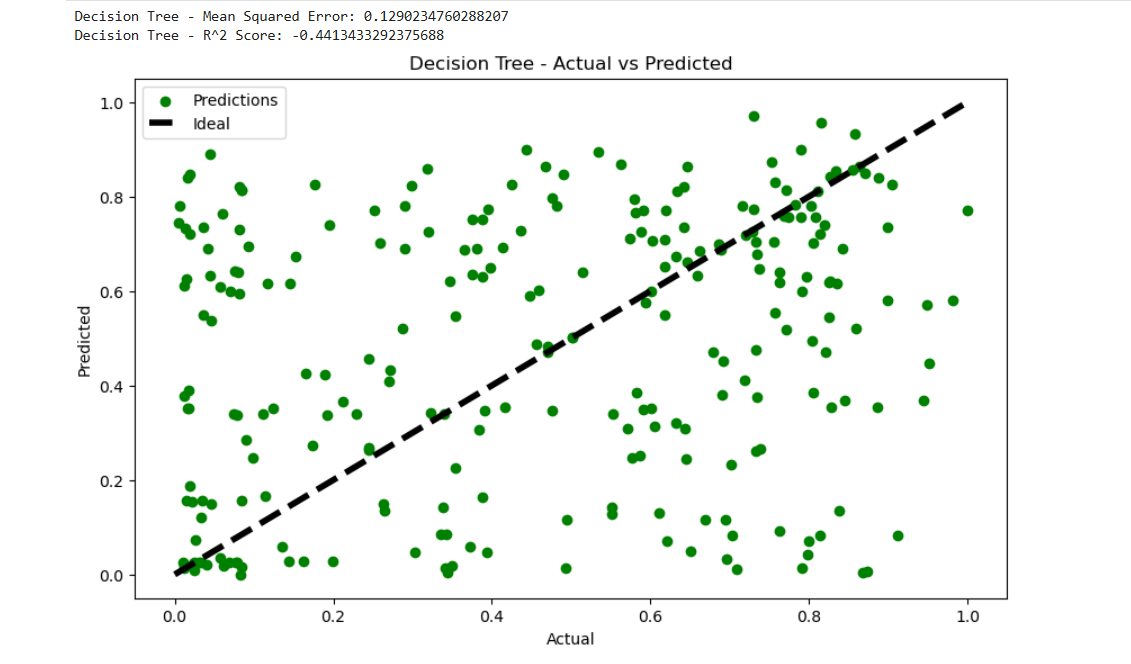




* **Mean Squared Error (MSE)** : Le MSE est extrêmement élevé. Cela signifie que les prédictions du modèle sont très éloignées des valeurs réelles, indiquant une grande erreur.
* **R^2 Score** : très inhabituel et extrêmement négatif.
* Le graphique montre une comparaison entre les valeurs réelles et prédites. Les valeurs prédites semblent être dans une plage très étroite, ce qui ne correspond pas bien aux valeurs réelles. Cela confirme que le modèle ne parvient pas à faire des prédictions précises

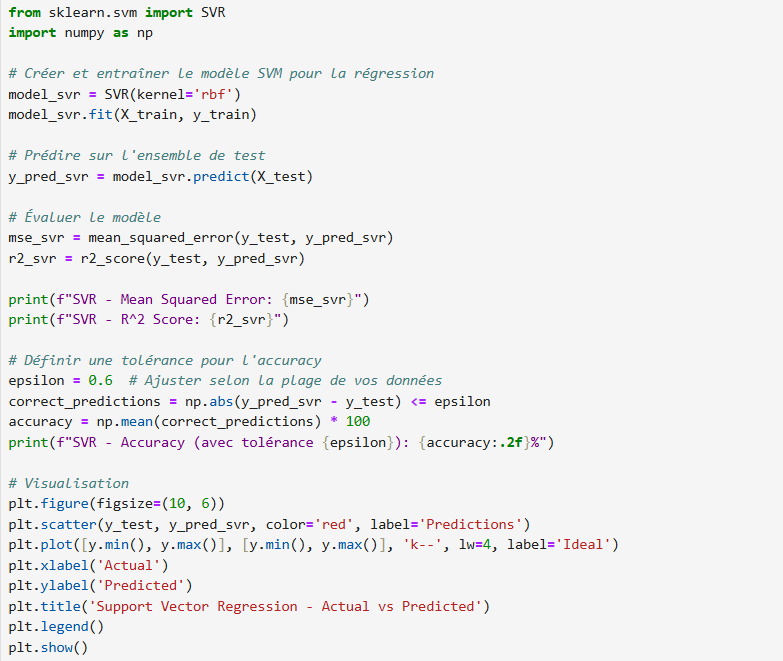
## Arbre de Décision (Decision Tree)

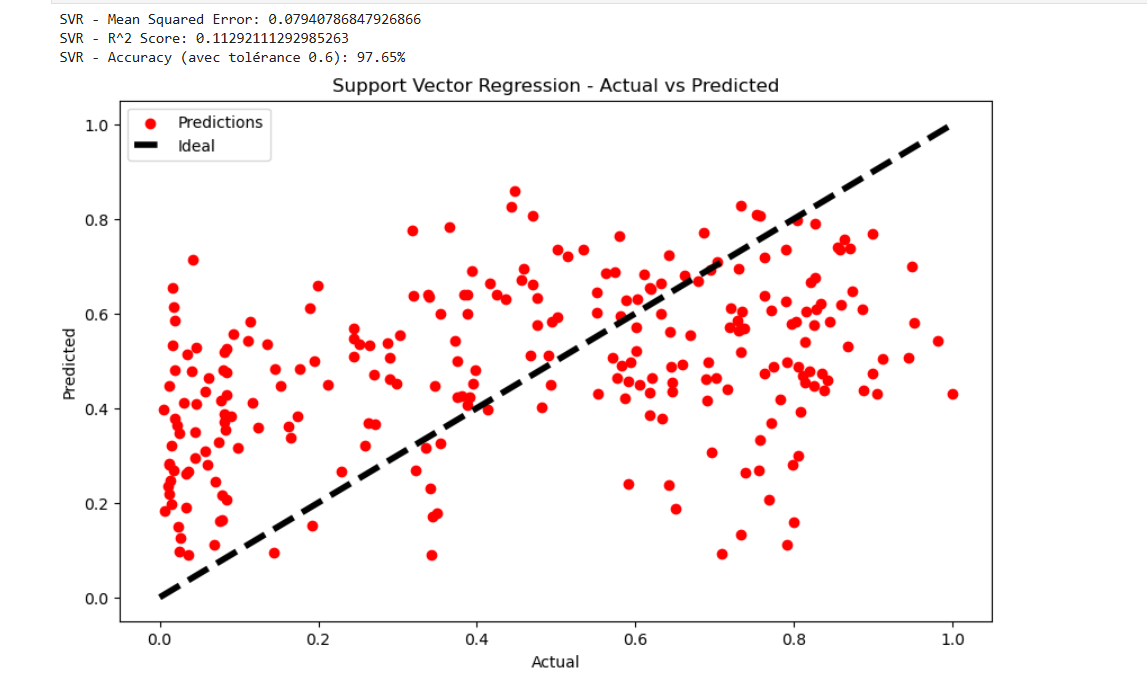




* **Mean Squared Error (MSE)** : Le MSE est de 0.129, ce qui est relativement faible par rapport au modèle de régression polynomiale. Cela indique que les prédictions sont plus proches des valeurs réelles, bien qu'il y ait encore une certaine erreur.
* **R^2 Score :**  est négatif. Cela indique que le modèle ne parvient pas à capturer correctement la variance des données.
* Le graphique montre une comparaison entre les valeurs réelles et prédites. Les valeurs prédites semblent être dans une plage de 0.0 à 1.0, mais elles ne correspondent pas bien aux valeurs réelles, ce qui confirme que le modèle ne fait pas des prédictions précises.

## SVR (Support Vector Regression)

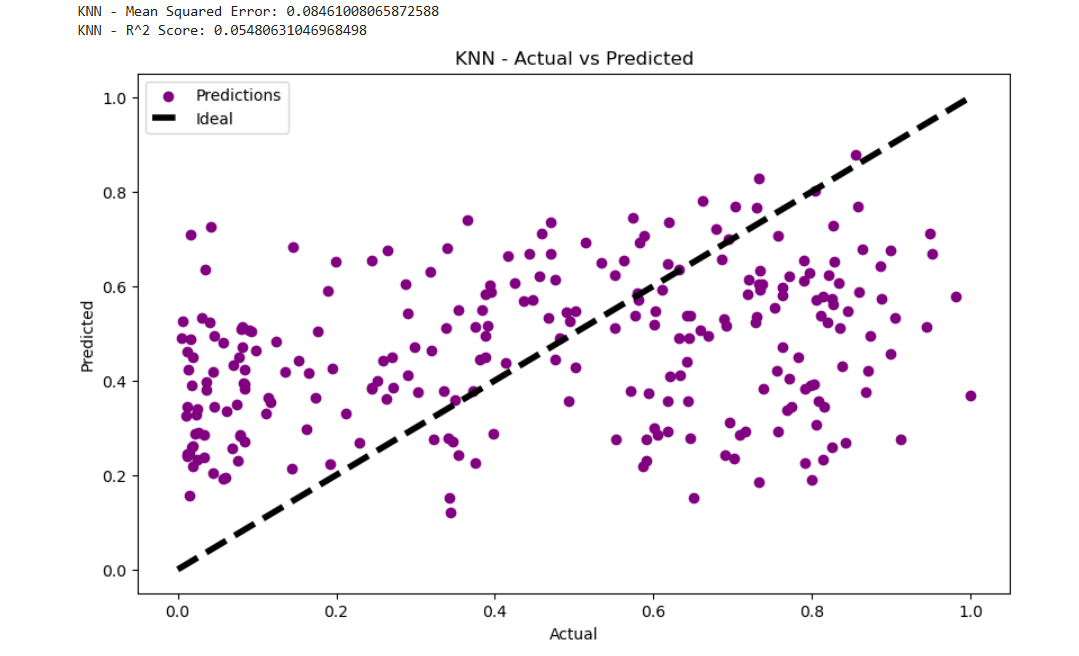




* **Mean Squared Error (MSE)** : indique que les prédictions du modèle sont assez proches des valeurs réelles, avec une erreur réduite.
* **R^2 Score** : est positif mais relativement faible. Un score R^2 proche de 1 indiquerait une excellente adéquation du modèle aux données.
* **Accuracy (avec tolérance 0.6)** : Le modèle atteint une précision de 97.65% avec une tolérance de 0.6. Cela signifie que dans 97.65% des cas, les prédictions du modèle sont à moins de 0.6 unité des valeurs réelles, ce qui est très bon.
* Le graphique montre une comparaison entre les valeurs réelles et prédites. Les valeurs prédites semblent bien correspondre aux valeurs réelles, bien qu'il y ait encore quelques écarts.

## KNN (K-Nearest Neighbors)





* **Mean Squared Error (MSE)** : est relativement faible. Cela indique que les prédictions du modèle sont assez proches des valeurs réelles, bien qu'il y ait encore une certaine erreur.
* **R^2 Score** : est positif mais faible, le score suggère que le modèle ne capture qu'une petite partie de la variance des données.
* Le graphique montre une comparaison entre les valeurs réelles et prédites. Les valeurs prédites semblent correspondre aux valeurs réelles dans une certaine mesure, mais il y a des écarts, en particulier pour les valeurs réelles plus élevées.

Discussion

***Meilleur Modèle :*** Le modèle Support Vector Regression (SVR) est le meilleur parmi ceux présentés. Il a le MSE le plus faible (0.079), un score R^2 positif (0.113), et une précision très élevée (97.65%) avec une tolérance de 0.6. Bien que le score R^2 ne soit pas très élevé, il est nettement meilleur que ceux des autres modèles.

# ***O***ptimisation des Feux de Circulation

 Formulation du Problème d'Optimisation

L'objectif de cette étape est de minimiser le temps d'attente total des véhicules en optimisant la durée des feux de circulation. Le problème a été formulé comme suit :

* **Variables de décision** :
  + ns\_duration : Durée des feux Nord-Sud (NS).
  + ew\_duration : Durée des feux Est-Ouest (EW).
* **Fonction objectif** : Minimiser le temps d'attente total.
* **Contraintes** :
  + La durée totale d'un cycle de feux ne doit pas dépasser 120 secondes.
  + Une direction peut être prioritaire si son volume de trafic est significativement plus élevé.



 Résolution du Problème

Le problème a été résolu en utilisant la bibliothèque pulp en Python. Les résultats obtenus sont les suivants :

* **Statut de la solution** : Optimal.
* **Durée des feux NS** : 15.0 secondes.
* **Durée des feux EW** : 22.5 secondes.
* **Temps total d'attente minimisé** : 17.029 unités de temps.
* **Interprétation des Résultats**
* **Durée des feux** : Les feux NS ont une durée plus courte que les feux EW, ce qui indique que la direction EW a un volume de trafic plus élevé.
* **Priorité** : La direction EW a été priorisée en raison de son volume de trafic plus important.

