

LAPILUS Dylan

Rapport d’analyse et de prédiction

Table des matières

[Introduction 2](#_Toc54914369)

[Analyse des données 2](#_Toc54914370)

[Résultat 5](#_Toc54914371)

[Conclusion 5](#_Toc54914372)

# Introduction

Boston Police Department (BPD) reçoit chaque jour des rapports de crime auquels les officiers doivent répondre. Ces données sont maintenant stockées est formalisé depuis 2015. Analyser ces données pourra permettre aux officiers de mieux comprendre et visualiser les quartiers dit à risque. Ce document est un rapport des résultats obtenues dans le cadre de la résolution de la problématique suivante : **Est-il possible de prévoir où et quand un crime va être commis ?**

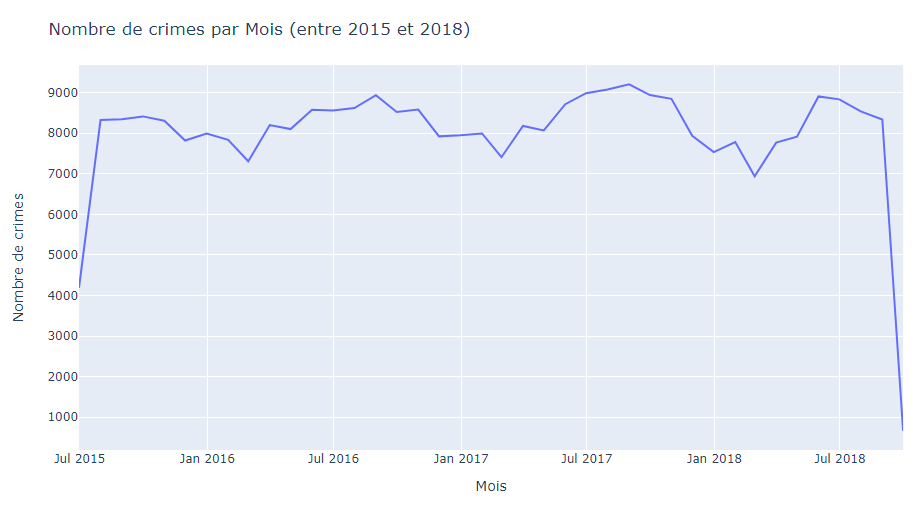
Pour cela, nous allons nous basée sur les données fournis, analyser les données afin de les comprendre et ainsi mettre en place un algorithme qui nous permettra de prédire les futurs crimes.

# Analyse des données

Les données transmis pour ce travail sont les suivantes :

|  |  |  |
| --- | --- | --- |
| **Données** | **Type** | **Exemple** |
| INCIDENT\_NUMBER | Texte | I182070945 |
| OFFENSE\_CODE | Nombre | 619 |
| OFFENSE\_CODE\_GROUP | Texte | Larceny |
| OFFENSE\_DESCRIPTION | Texte | LARCENY ALL OTHERS |
| DISTRICT | Texte | D14 |
| REPORTING\_AREA | Texte | 808 |
| SHOOTING | Texte (Boolean) | Y |
| OCCURRED\_ON\_DATE | Date | 2018-09-02 13:00:00 |
| YEAR | Nombre | 2018 |
| MONTH | Nombre | 9 |
| DAY\_OF\_WEEK | Texte | Sunday |
| HOUR | Nombre | 13 |
| UCR\_PART | Texte | Part One |
| STREET | Texte | LINCOLN ST |
| Lat | Nombre | 42.357791 |
| Long | Nombre | -71.139371 |
| Location | Tableau de nombre | (42.35779134, -71.13937053) |
| Category | Texte | Larceny |

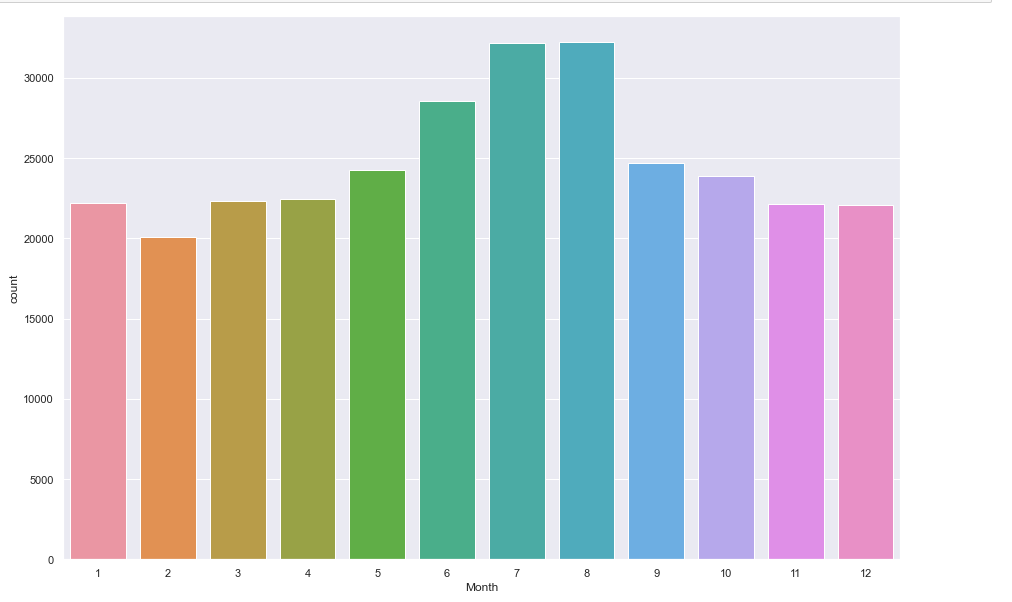
Dans cette liste de donnée, « Category » a été ajouté afin de pouvoir regrouper les crimes et mieux comprendre les types de crimes. Cette donnée étant moins éparpiller nous a permis de faire des analyses plus précises sur les différents crimes à boston, en regroupant par exemple les différents vols qu’ils soient à l’arraché ou sur propriété privé.



*Figure 1 : Graphique représentant le nombre de crimes entre 2015 et 2018*

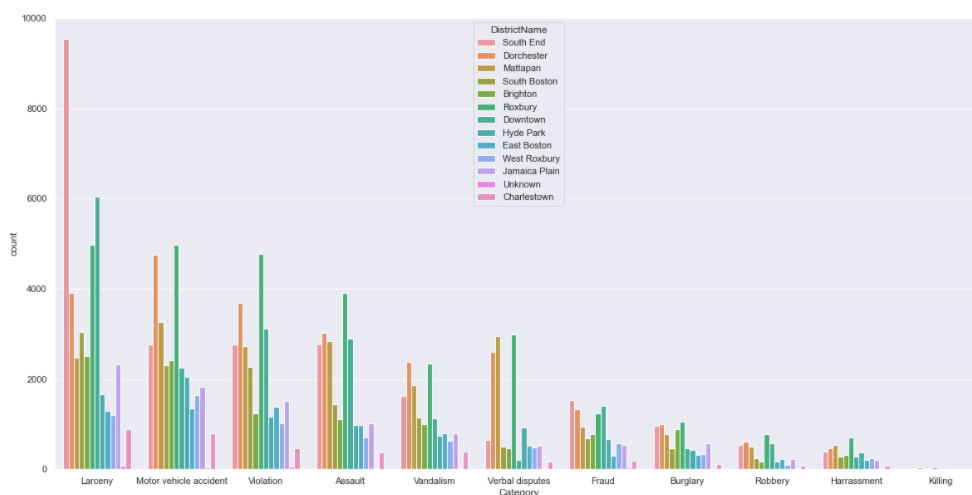
A partir de ce graphique, on a pu constater plusieurs points. Le premier est que les données fournis ne sont comprises qu’entre Juillet 2015 et Août 2018, ce qui a faussé les données sur le nombre de crimes par années.

La deuxième constatation est les différentes chutes de la courbe durant le printemps, celles-ci remontent durant l’été. Afin de comprendre ce qu’il se passe durant cette période, nous nous sommes penchés plus précisément sur sujet.



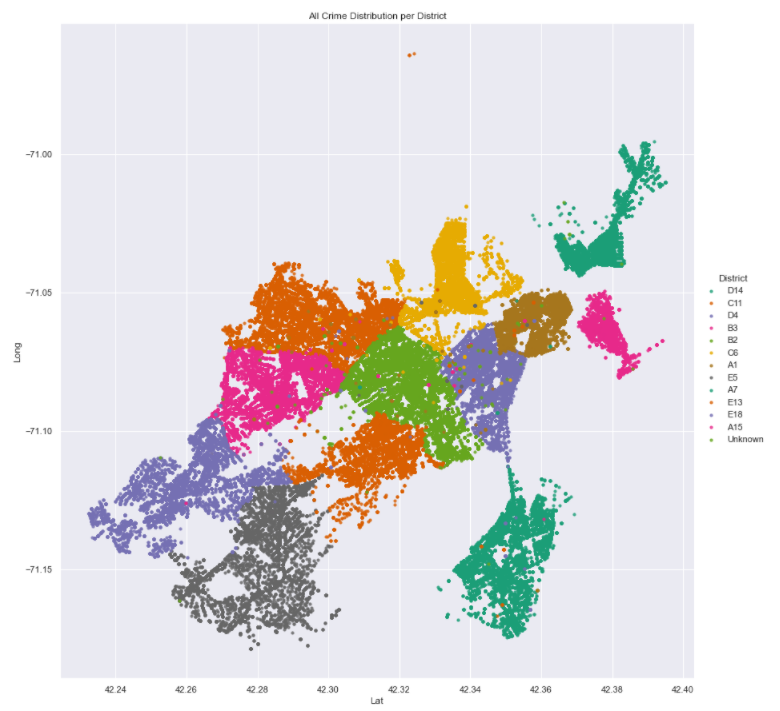
*Figure 2 : Graphique représentant le nombre de crimes par Mois*

Avec ce graphique, nous avons pu confirmer que durant l’été le nombre de crimes augmentent belle et bien. Voyons maintenant quels sont les crimes les plus présents selon les districts.



*Figure 3 : Graphique représentant la catégorie de crime par district*

Pour ce graphique, nous avons regroupé tous les crimes en catégorie afin de voir lequel domine le plus. Les crimes sont dans l’ordre décroissant et comme nous pouvons le voir nous avons une barre dominante « Larceny » dans le district de « South End ».



*Figure 4 : Nuage de point représentant les crimes à Boston*

Afin de mieux voir la répartition des crimes sur le territoire de Boston, nous avons réalisé un nuage de point les représentant par district. Nous pouvons observer certains points de couleurs qui se trouve dans en dehors de son district. Nous en avons conclu qu’ils avaient été signalés dans le district correspondant à la couleur, cela peut représenter, par exemple, un vol lors d’un trajet entre le domicile et le lieu de travail.

# Résultat

Après avoir étudier l’ensemble des données, nous avons entrainés plusieurs algorithmes pour prédire les crimes. Suite aux tests de ses algorithmes avec une partie des données écarté au préalable, nous avons pu voir et comparé leurs scores de prédictions :

|  |  |
| --- | --- |
| Algorithme | Score maximal |
| Random Forest (RFC) | 39% |
| K Nearest Neighbors (KNN) | 36% |
| Support Vector Classification (SVC) | Aucun résultat |

Nous avons pu remarquer que l’algorithme ayant le meilleur score de prédiction est Random Forest. Nous avons donc livré cet algorithme en production.

# Conclusion

L’objectif de cette analyse était de savoir si nous pouvions trouver le rapport entre différentes variables et en prédire les crimes. Celles-ci ont montré qu’il était difficile de les prévoir car notre meilleur algorithme, RFC, reçoit un score maximal de 39%. Ce qui est compréhensible car il est difficile de prévoir le comportement humain.