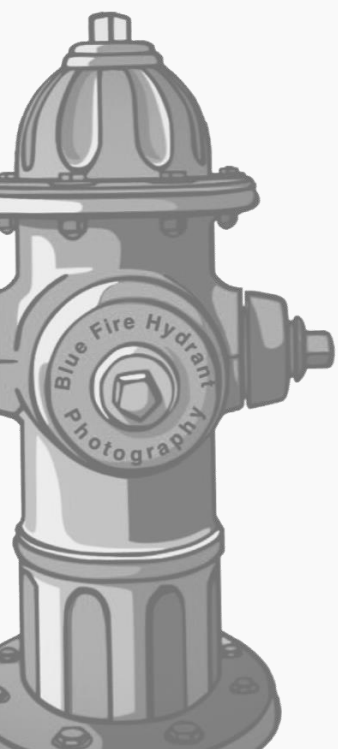


PROJET

DATA ANALYST

**“ Prédiction du temps de réponse d'un véhicule de la
Brigade des Pompiers de Londres ”**



Auteurs

BARRERA CEBRIAN, Ivan
COULANDREAU, Franck
HAEGEMAN, Cédric
HUBERT, Arnaud

Chef de cohorte

KASSEL, Raphael



DataScientest • com

Promotion Février 2021 “ Bootcamp ”

INDEX

I. ABSTRACT 3

- Introduction et résumé de l'objectif de notre étude
- Key words
- Niveau d'expertise autour du sujet à étudier

II. EXPLORATION ET SÉLECTION DES DONNÉES 4

Vision de l'ensemble des données :

- Brève description du jeu des données
- Description des variables sélectionnées pour l'analyse

III. DATAVIZ' 6

Représentations graphiques des données :

- Analyse
- Interprétation
- Commentaires

IV. MACHINE LEARNING 14

Application de Machine Learning aux datasets :

- Heatmap
- Modèle : Linear Regression
- Modèle : SGD Regressor
- Modèle : Gradient Boosting Regressor

V. CONCLUSIONS 20

Résumé de l'analyse du projet :

- Vision objective de l'ensemble des résultats
- Vision subjective
- Questionnement des résultats afin de proposer des nouvelles études sur le sujet

VI. DATA SOURCE 21

I. ABSTRACT

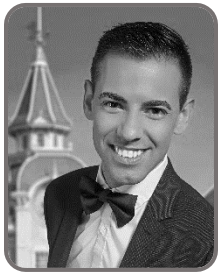
L'objectif de ce projet est d'analyser et/ou d'estimer les temps de réponse et de mobilisation de la Brigade des Pompiers de Londres (**London Fire Brigade, LFB**). La brigade des pompiers de Londres est le service d'incendie et de sauvetage le plus actif du Royaume-Uni et l'une des plus grandes organisations de lutte contre l'incendie et de sauvetage au monde.

Le *premier jeu de données* fourni contient les détails de chaque incident traité depuis janvier 2017 jusqu'à janvier 2021. Des informations sont fournies sur la date et le lieu de l'incident ainsi que sur le type d'incident traité.

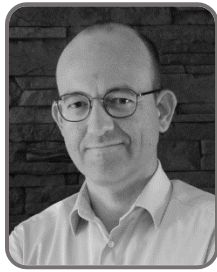
Le *second jeu de données* contient les détails de chaque camion de pompiers envoyé sur les lieux d'un incident depuis janvier 2017 jusqu'à janvier 2021. Des informations sont fournies sur l'appareil mobilisé, son lieu de déploiement et les heures d'arrivée sur les lieux de l'incident.

Key words : brigade, pompiers, Londres, mobilisation, incident, incendie, caserne, intervention, temps, LFB, London, fire.

Auteurs du projet



Ivan
BARRERA CEBRIAN



Franck
COULANDREAU



Cédric
HAEGEMAN



Arnaud
HUBERT

Niveau d'expertise autour du sujet à étudier

Ayant des profils professionnels très différents, la première approche de ce projet nous a permis d'avoir des points de vue complémentaires sur le sujet adressé. Bien que nous n'ayons pas d'expertise directe sur l'organisation d'une caserne de pompiers, ni sur la problématique de prévision d'un temps.

Afin d'approfondir nos connaissances, nous avons fait des recherches dans ce domaine en nous renseignant par exemple sur le nombre de casernes, la composition géographique de Londres ainsi que ses différents quartiers, le langage technique utilisé dans le métier, etc. L'ensemble de ces informations nous ont permis d'avoir une meilleure vision sur la problématique adressée.

En dehors de notre propre étude, nous n'avons pas pris contact avec des experts du métier de pompier, mais notre chef de cohorte, Raphael KASSEL, nous a bien guidé dans le côté technique afin d'appliquer nos connaissances dans l'analyse des données et proposer un modèle de prévision de prévision d'un temps.

II. EXPLORATION ET SÉLECTION DES DONNÉES

Comme il a été décrit précédemment dans l'Abstract, nous allons explorer et analyser en détail les 2 jeux de données correspondant aux « Incidents » et à la « Mobilisation » :

- **Premier jeu de données - Incidents** : *LFB_Incident_data_from_January_2017.xlsx*

Ce dataset est composé d'un total de 38 variables et 420 244 lignes :

1. IncidentNumber	14. UPRN	27. Longitude
2. DateOfCall	15. USRN	28. FRS
3. CalYear	16. IncGeo_BoroughCode	29. IncidentStationGround
4. TimeOfCall	17. IncGeo_BoroughName	30. FirstPumpArriving_AttendanceTime
5. HourOfCall	18. ProperCase	31. FirstPumpArriving_DeployedFromStation
6. IncidentGroup	19. IncGeo_WardCode	32. SecondPumpArriving_AttendanceTime
7. StopCodeDescription	20. IncGeo_WardName	33. SecondPumpArriving_DeployedFromStation
8. SpecialServiceType	21. IncGeo_WardNameNew	34. NumStationsWithPumpsAttending
9. PropertyCategory	22. Easting_m	35. NumPumpsAttending
10. PropertyType	23. Northing_m	36. PumpCount
11. AddressQualifier	24. Easting_rounded	37. PumpHoursRoundUp
12. Postcode_full	25. Northing_rounded	38. Notional Cost (£)
13. Postcode_district	26. Latitude	

- **Deuxième jeu de données - Mobilisation** : *LFB_Mobilisation_data_from_January_2017.xlsx*

Ce dataset est composé d'un total de 22 variables et 618 071 lignes :

1. IncidentNumber	12. TimeLeftTimezoneId
2. ResourceMobilisationId	13. DateAndTimeReturned
3. Resource_Code	14. TimeReturnedTimezoneId
4. PerformanceReporting	15. DeployedFromStation_Code
5. DateAndTimeMobilised	16. DeployedFromStation_Name
6. DateAndTimeMobile	17. DeployedFromLocation
7. TimeMobileTimezoneId	18. PumpOrder
8. DateAndTimeArrived	19. PlusCode_Code
9. TimeArrivedTimezoneId	20. PlusCode_Description
10. AttendanceTimeSeconds	21. DelayCodeId
11. DateAndTimeLeft	22. DelayCode_Description

Nous avons un total de 60 variables et 1 038 315 lignes, mais pour atteindre l'objectif de cette étude, l'estimation du temps de réponse et la mobilisation de la Brigade des Pompiers de Londres, nous avons fusionné les deux datasets grâce à la variable « *IncidentNumber* », qui nous servira d'index.

Le nouveau DataFrame issue de la fusion, appelé *LFB_analysis_py.csv*, contient la sélection des 31 variables suivantes et 840 488 lignes :

Table 1. *Sélection des variables pertinentes pour l'analyse*

N°	NOM DES VARIABLES	DESCRIPTION DES VARIABLES
1	IncidentNumber	Numéro de l'incident
2	DateOfCall	Date de l'appel JJ/MM/YYYY
3	TimeOfCall	Temps de l'appel HH/MM/SS
4	IncidentGroup	Catégorie d'incident
5	StopCodeDescription	Sous-catégorie d'incident
6	SpecialServiceType	Type de service spécial
7	PropertyCategory	Type de propriété
8	PropertyType	Sous-catégorie du type de propriété
9	AddressQualifier	Qualité du renseignement sur la localisation de l'incident
10	Postcode_district	Code district
11	IncGeo_BoroughCode	Code arrondissement
12	ProperCase	Nom arrondissement en minuscule
13	IncGeo_WardCode	Code du quartier
14	IncGeo_WardNameNew	Nom du quartier
15	Latitude	Coordonnées GPS Latitude
16	Longitude	Coordonnées GPS Longitude
17	FirstPumpArriving_AttendanceTime	Délai d'arrivée de la 1ère brigade en secondes
18	FirstPumpArriving_DeployedFromStation	Nom de la brigade d'origine 1ère brigade
19	SecondPumpArriving_AttendanceTime	Délai d'arrivée de la 2ème brigade en secondes
20	SecondPumpArriving_DeployedFromStation	Nom de la brigade d'origine 2ème brigade
21	NumStationsWithPumpsAttending	Nombre de la brigade engagée pour 1 même incident
22	NumPumpsAttending	Nombre de camions déployés de la brigade pour 1 même incident
23	PumpCount	Nombre total de camions déployés toutes brigades confondues pour 1 même incident
24	PumpHoursRoundUp	Nombre d'heures d'intervention cumulées
25	Notional Cost (£)	Coût théorique de l'intervention
26	DateAndTimeMobile	Date et heure à laquelle les pompiers partent de la caserne
27	DateAndTimeArrived	Date et heure à laquelle les pompiers sont arrivés sur le lieu de l'incident
28	DateAndTimeLeft	Heure de départ des pompiers du lieu de l'incident
29	DeployedFromStation_Code	Code identifiant de la caserne
30	DeployedFromStation_Name	Nom de la brigade (nom du quartier)
31	DelayCodeId	Code identifiant de la raison du délai

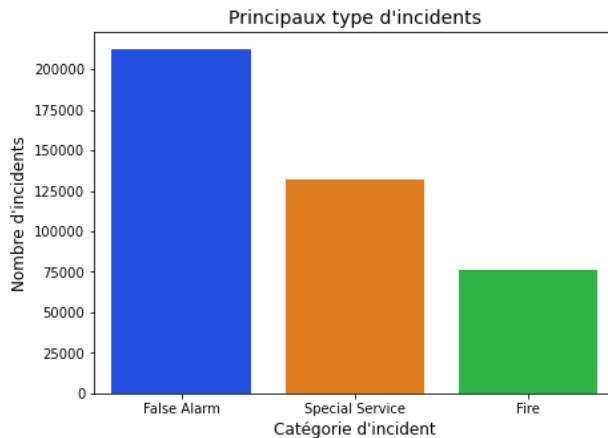
Notre **variable cible** pour créer notre modèle de prédiction du temps de réponse et de mobilisation de la Brigade des Pompiers de Londres sera « **FirstPumpArriving_AttendanceTime** », qui nous proportionne le délai de mobilisation et d'arrivée de la première brigade. Les valeurs de cette variable sont données en secondes. (*Se référer à notre tableau des 'Stats' Excel pour la distribution des valeurs*)

Dans le chapitre suivant, nous allons croiser différentes variables afin d'en tirer des conclusions pour pouvoir justifier la sélection des nos 31 variables parmi les 60 que nous avons au départ.

III. DATAVIZ'

Après notre analyse de l'ensemble des variables nous vous présentons une **analyse générale des données** :

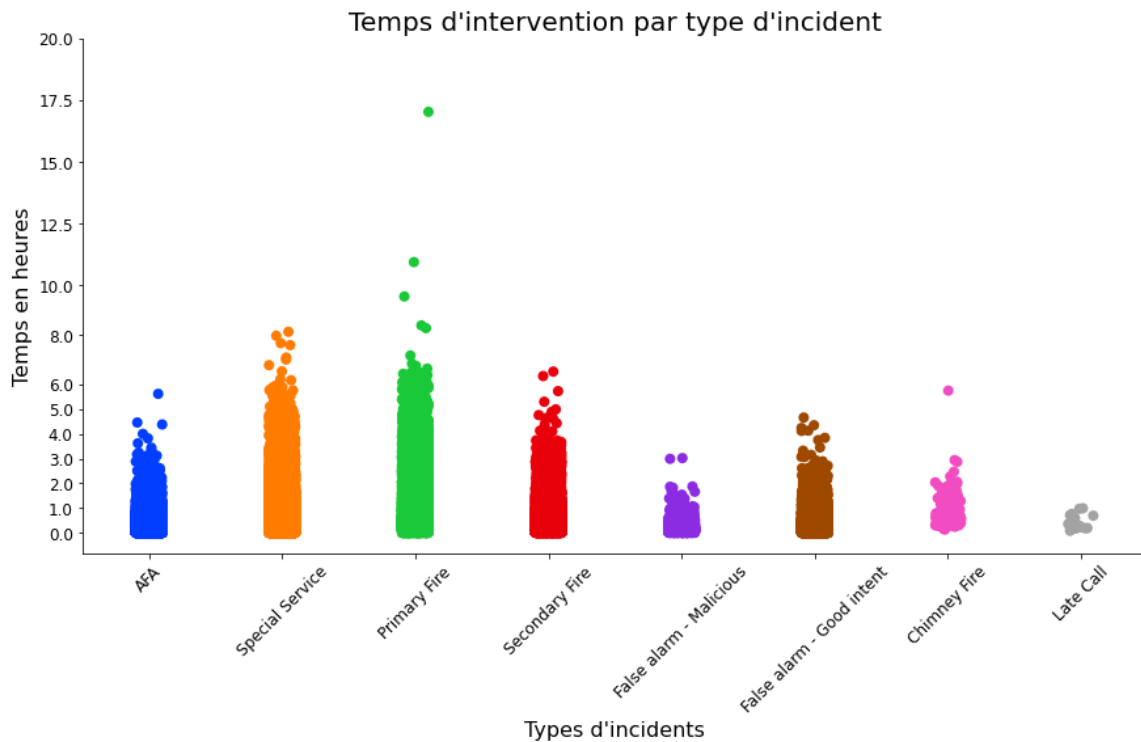
Graphique 1. *Principaux types d'incidents en nombre absolu*



Nous pouvons constater que la majorité des incidents signalés sont des fausses alarmes (automatique ou bien manuelle.)

Viennent ensuite les 'Special Service', puis en dernière position les incendies.

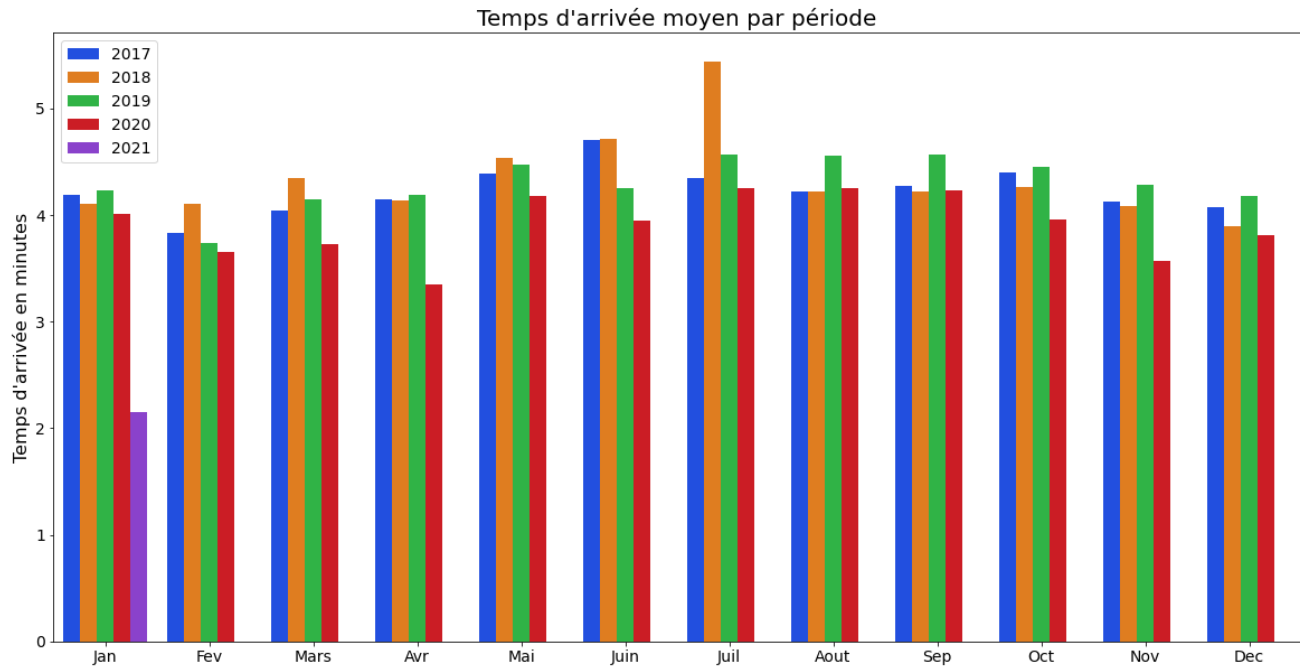
Graphique 2. *Temps d'intervention par type d'incident*



Nous constatons que les temps d'interventions par type d'incident sont généralement de 2 heures, avec des temps plus élevés comme pour les 'Incendies' et 'Special Services' (inondation, suicide, ouverture d'ascenseurs...)

Dans un autre côté, les incendies sont les types d'incident qui peuvent retenir les pompiers plusieurs dizaines d'heures. Nous pouvons visualiser aussi que les 3 types d'alarmes engrangent énormément de temps cumulé inutile.

Graphique 3. Temps d'arrivée moyen par période de l'année

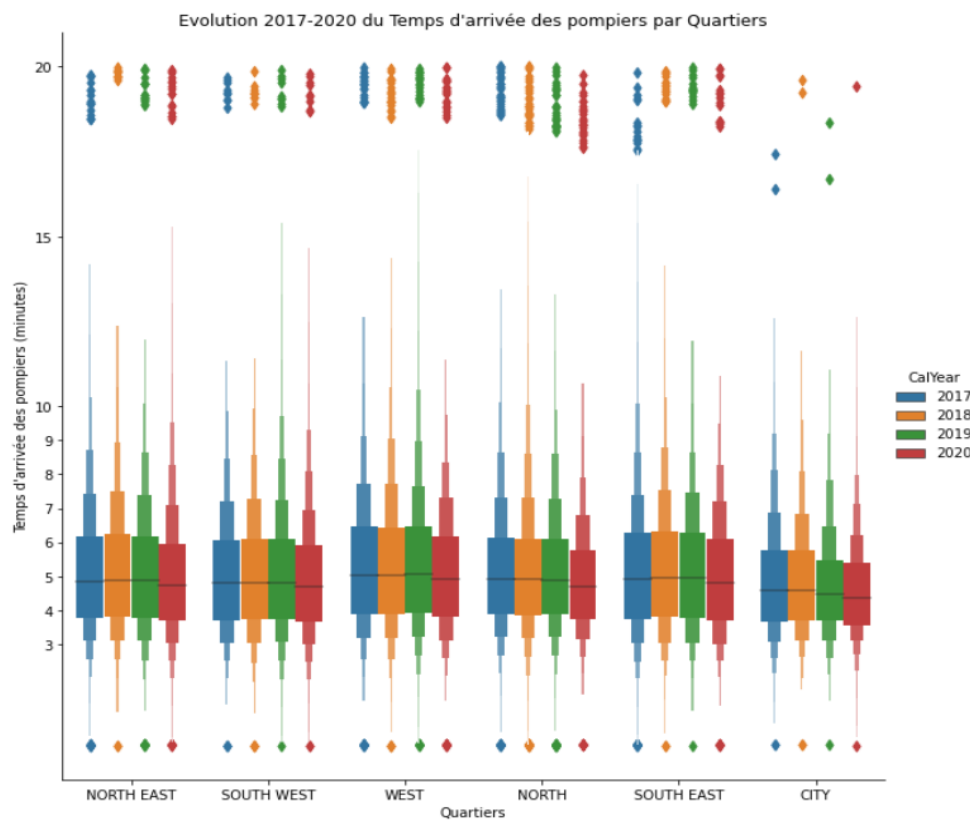


* Notre dataset contient des données jusqu'au 12/01/2021

Nous pouvons remarquer que le temps d'arrivée moyen fluctue selon la période de l'année, avec des temps élevés sur la période estivale, et en diminution sur la période hivernale.

Nous constatons également un pic sur le mois de juillet 2018.

Graphique 4. Évolution du temps moyen d'arrivée des pompiers par secteurs géographiques



Pour mieux comprendre le graphique, *Le Grand Londres* est composé de 33 quartiers qu'il est commun de regrouper en 6 secteurs géographiques :

- **SOUTH EAST** : Bexley, Bromley, Greenwich, Lewisham, Southwark.
- **NORTH EAST** : Havering, Redbridge, Newham, Barking & Dagenham, Waltham Forest, Tower Hamlets.
- **NORTH** : Enfield, Barnet, Haringey, Hackney, Islington, Camden, Westminster.
- **WEST** : Harrow, Hillingdon, Brent, Ealing, Hounslow, Hammersmith & Fulham, Kensington & Chelsea.
- **SOUTH WEST** : Richmond upon Thames, Kingston upon Thames, Wandsworth, Merton, Lambeth, Sutton, Croydon.
- **CITY** : City of London.

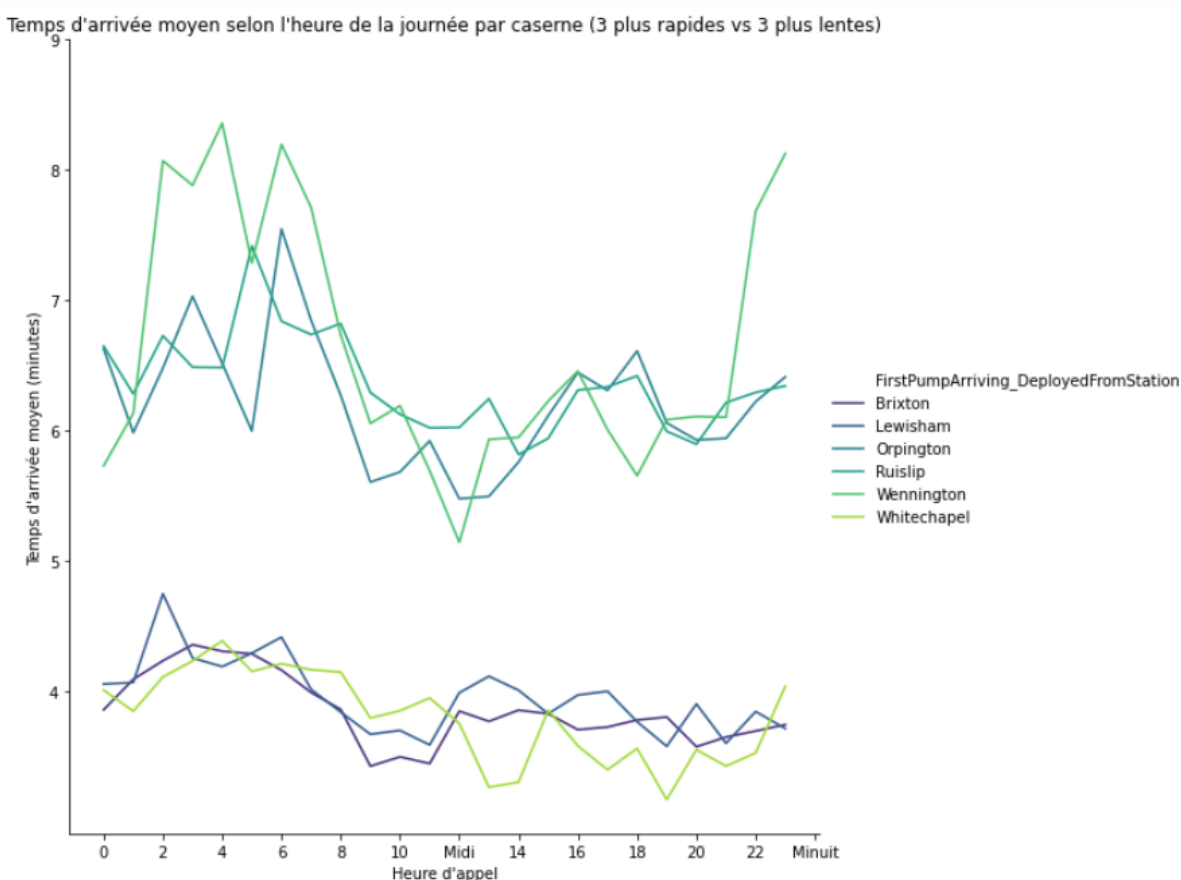


Cartographie : "Secteurs du Grand Londres"

Pour les 6 secteurs géographiques du Grand Londres on peut constater que la médiane des temps d'arrivée des pompiers se situe aux alentours de 5 minutes et que dans la majorité des cas les pompiers mettent entre 4 et 6 minutes entre une alerte et leur arrivée sur les lieux. On observe que les pompiers sur le secteur de la City mettent moins de temps à intervenir que dans les autres secteurs.

On peut également noter que pour les 6 secteurs géographiques le temps d'arrivée des pompiers a diminué sur l'année 2020, probablement à cause de la circulation plus fluide observée pendant les périodes de confinement liées à la crise sanitaire du Covid19.

Graphique 5. Temps moyen d'arrivée des pompiers par caserne



Parmi les 105 casernes du Grand Londres, on constate que le temps d'arrivée moyen de la première équipe sur les lieux d'un incident peut varier significativement entre les casernes.

Une différence de l'ordre de 2 à 3 minutes est observée entre les casernes qui interviennent le plus rapidement et celles qui mettent le plus de temps (quel que soit l'heure de l'appel dans la journée). On peut imaginer facilement que sur un départ d'incendie ces 2 à 3 minutes peuvent avoir de lourdes conséquences.

Pour construire cette visualisation il a fallu ne tenir compte que des incidents pour lesquels c'est la caserne du secteur qui est intervenue. En effet, il arrive parfois qu'une caserne extérieure au secteur arrive en premier sur les lieux d'un incident avec un temps forcément un peu supérieur à ses temps habituels.

Table 2. Étude de corrélation entre le temps d'arrivée des pompiers et la caserne du lieu de l'incident (ANOVA)

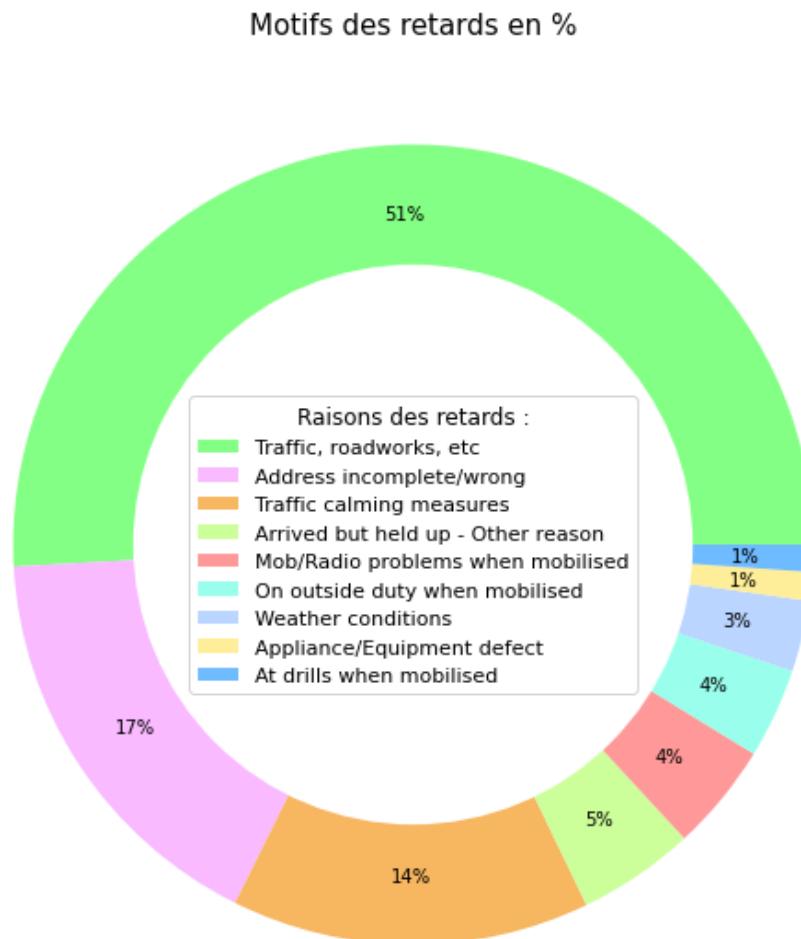
	df	sum_sq	mean_sq	F	PR(>F)
IncidentStationGround	101.0	3.744967e+08	3.707889e+06	224.890618	0.0
Residual	389079.0	6.414948e+09	1.648752e+04	NaN	NaN

La p-value (PR(>F)) à 0.0 est inférieure à 5 % donc on peut rejeter l'hypothèse selon laquelle *IncidentStationGround* (caserne du lieu d'un incident) n'influe pas sur *FirstPumpArriving_AttendanceTime* (temps d'arrivée des pompiers sur les lieux).

Par la suite et après une analyse approfondie de l'ensemble des variables, nous vous présentons les différentes représentations graphiques sur **l'impact du temps de réponse et mobilisations** des pompiers.

Cette deuxième partie de la représentation va servir comme introduction pour ensuite créer des modèles de prédiction à l'aide de Machine Learning.

Graphique 6. *Motifs des retards*

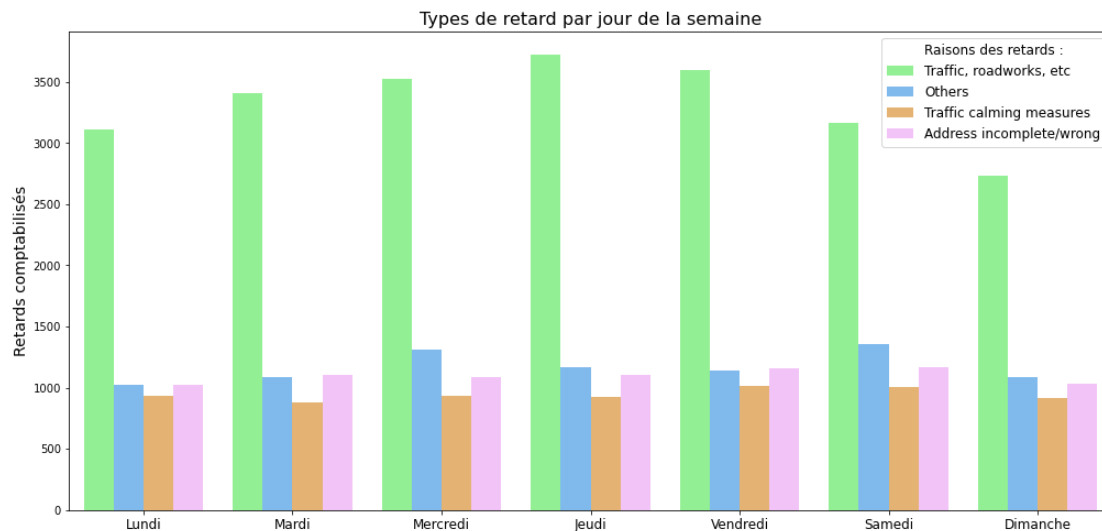


Nous constatons que les embouteillages représentent une grande majorité des raisons de retard avec 51%, suivi par 17% des anomalies d'adresses et, enfin, 14% des retards sont dus aux mesures de régulation de la circulation.

Le reste des motifs reste très minoritaire et donc nous allons les regrouper dans la variable 'Autres' pour le reste de cette analyse.

Nous allons maintenant étudier le type de retard par période de la semaine et de la journée et enfin calculer leurs temps moyens.

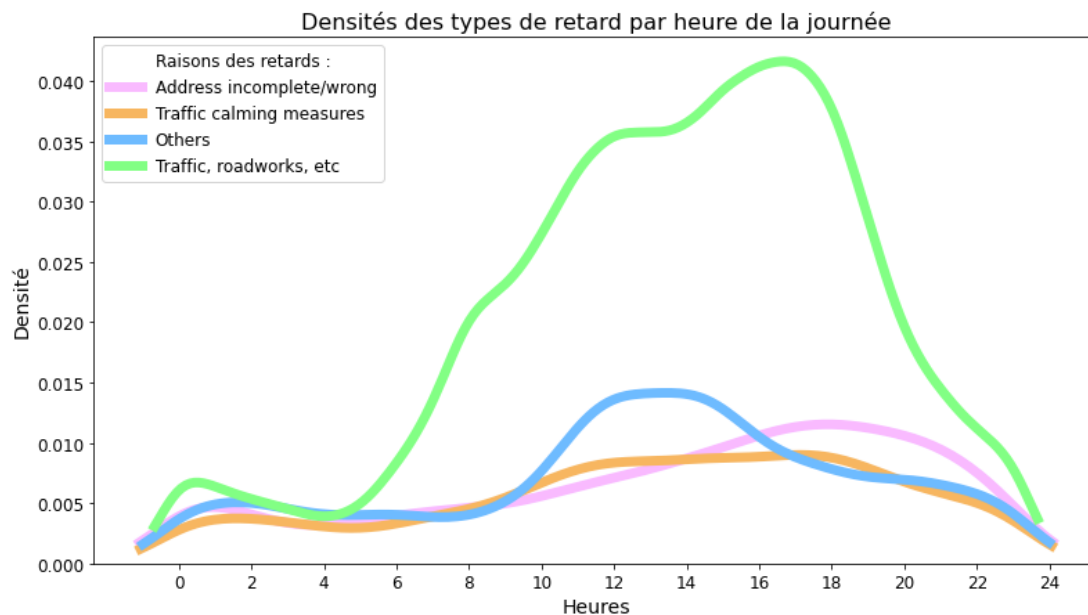
Graphique 7. Types de retard par jour de la semaine



Comme le démontrait le graphique précédent, les embouteillages sont fortement représentés, ceux-ci connaissent un pic en milieu de semaine et une régression le week-end.

Une fois regroupé, la variable « *Autres* » représente la 2eme cause des retards, avec les journées du mercredi et samedi en pic.

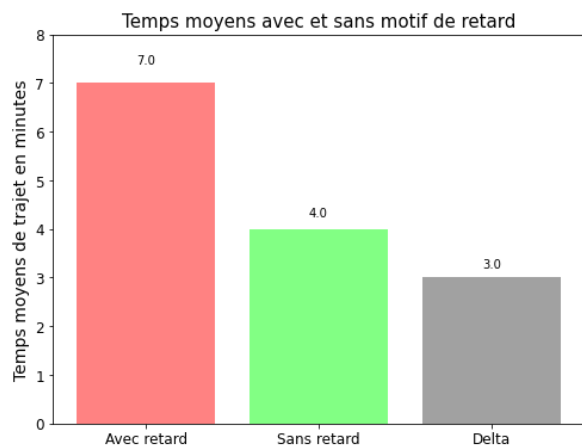
Graphique 8. Densités des types de retard par heure de la journée



Intéressons-nous à la variable la plus forte : « Embouteillage ». Sans surprise, celle-ci est très prononcée lors des heures de travail, entre 7h et 20h.

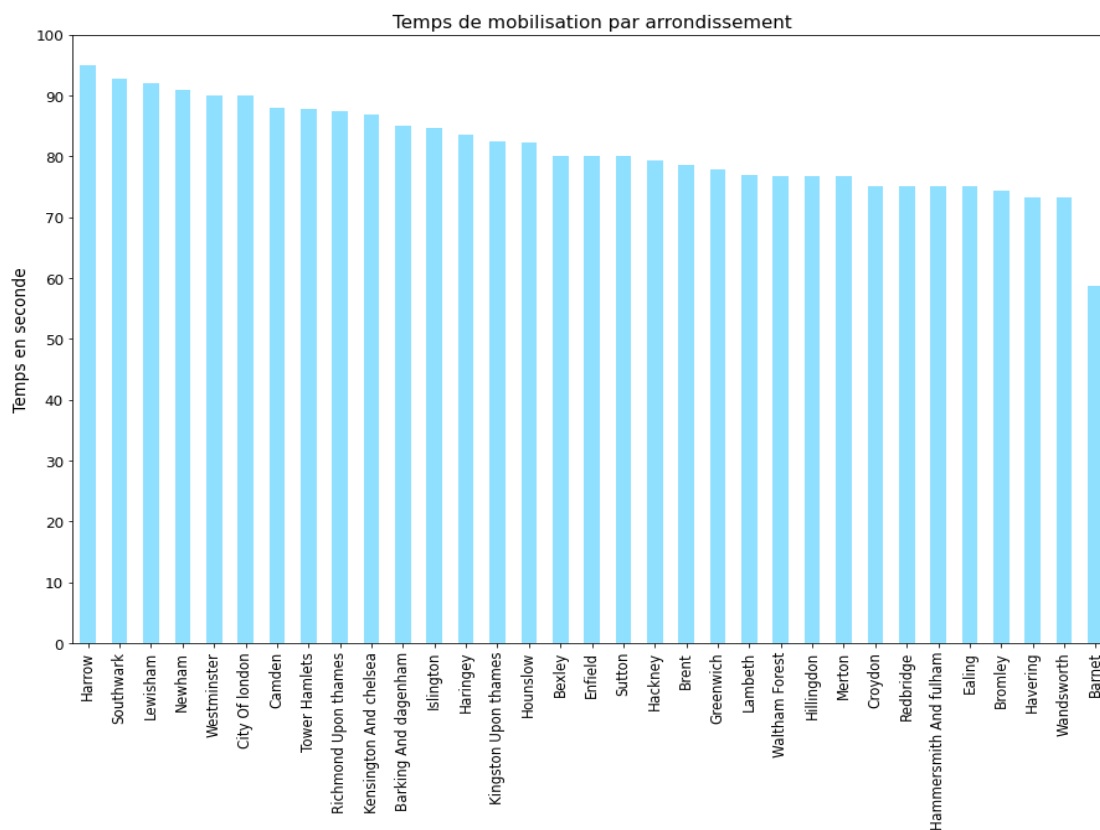
C'est ensuite la variable « *Autres* » qui prend le relais avec un pic entre 10h et 16h. Et enfin la variable « *Anomalies d'adresses* », qui se distingue en soirée.

Graphique 9. Temps moyens avec et sans motifs de retard



Une intervention avec retard mettra en moyenne 7 minutes, contre 4 minutes en temps normal, entre le départ et l'arrivée des pompiers sur les lieux des incidents. Il y a donc une moyenne de 3 minutes dès qu'un retard est enregistré.

Graphique 10. Temps de mobilisation par arrondissement

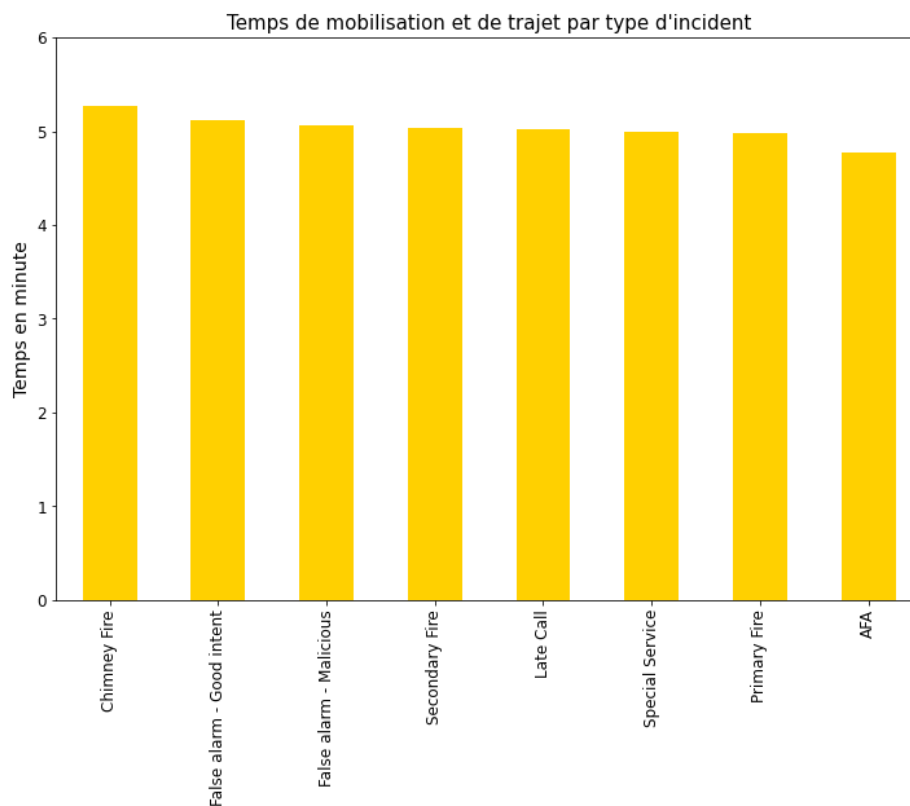


Ce graphique représente par arrondissement, le temps moyen écoulé entre l'heure de l'appel et l'heure de départ des pompiers.

Nous constatons un delta moyen de 37 secondes entre les 2 arrondissements en valeurs extrêmes.

La moyenne générale est de 80 secondes.

Graphique 11. Temps de mobilisation et de trajet par type d'incident



Ce graphique représente par type d'incident, le temps moyen écoulé entre le moment de l'appel et l'arrivée sur les lieux. Afin d'avoir les valeurs les plus réelles, les données n'incluent aucun type de retard.

Au vu du graphique, nous pouvons en déduire que les temps de mobilisation et de trajet restent les mêmes pour n'importe quel type d'incident avec un delta de 29 secondes entre les 2 valeurs extrêmes.

Les équipes et le matériel sont donc déjà prêt pour tous types d'interventions.

- Résumé de l'analyse -

Par ces graphiques, nous avons pu obtenir une meilleure représentation visuelle de la répartition générale de notre dataset. Nous avons également pu étudier en détail, l'impact de certains facteurs tels que :

- Les motifs de retard.
- Les types d'interventions, selon l'hypothèse que certaines d'entre elles pourraient nécessiter une préparation spécifique et donc ralentir les délais d'intervention.
- Les temps de mobilisation par arrondissement, afin d'évaluer la performance par secteur géographique.

Ces DataViz' ont démontré que seule la variable « *retards* » influait sur le temps d'arrivée moyen et de manière modérée. Nous allons donc, au vu de ces premières informations visuelles, étudier des modèles de prédictions.

IV. MACHINE LEARNING

Pour la partie Machine Learning, nous avons essayé d'estimer le temps d'arrivée des pompiers à partir de l'heure d'appel à leurs services.

La variable cible, étant une variable continue, nous avons choisi d'entraîner des méthodes de Régression Linéaires : LinearRegression, RidgeCV et LassoCV.

La **variable cible** identifiée est :

- FirstPumpArriving_AttendanceTime

Les **variables numériques** de nos datasets que nous avons retenues sont :

- DateOfCall retravaillée en CalYear, CalMonth, CalDay, CalWeekDay
- HourOfCall
- Easting_rounded
- Northing_rounded

Les **variables catégorielles** de nos datasets que nous avons retenues sont :

- IncidentGroup
- IncidentStationGround
- StopCodeDescription
- FirstPumpArriving_DeployedFromStation
- SpecialServiceType
- DeployedFromLocation
- AddressQualifier
- DelayCodeId
- ProperCase

De plus nous avons enrichi nos données en **rajoutant des variables** calculées :

- Casernes_quartier : le nombre de casernes du quartier de l'incident (ex. dans le quartier de Westminster il y a 8 casernes)
- Casernes_sous_quartier : le nombre de casernes du sous quartier de l'incident (ex. dans le sous quartier de Barkingside il y a 2 casernes)

Pour améliorer les performances de nos modèles basés sur la Régression Linéaire nous sommes partis de la matrice de corrélation de nos variables numériques. Malgré l'ajout des variables sur le nombre de casernes par quartier et par sous-quartier il fallait construire de nouvelles variables susceptibles d'être retenues par nos modèles de régression.

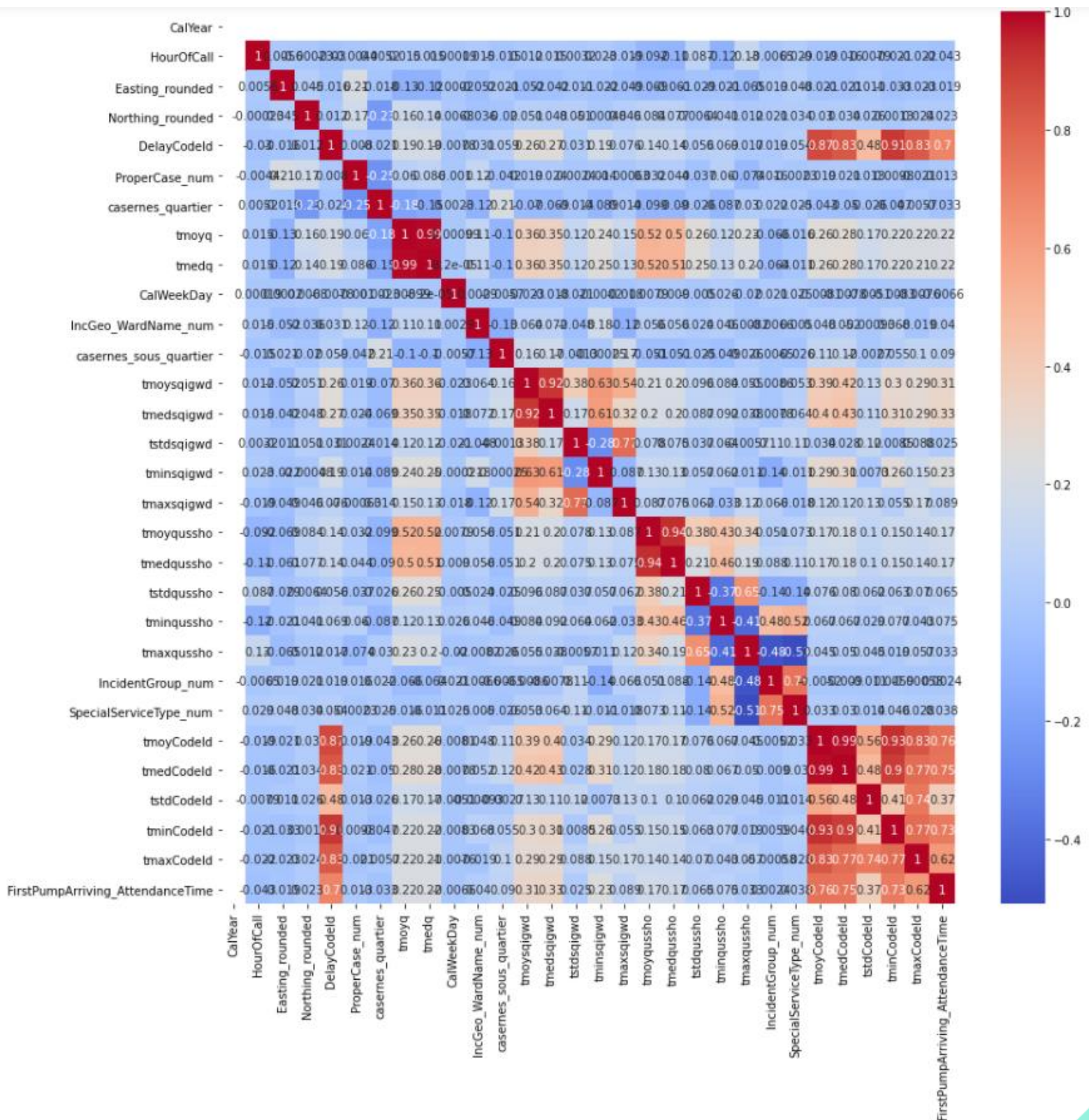
C'est pourquoi nous avons décidé d'ajouter des données statistiques de l'année antérieure à nos échantillons d'entraînement et de test.

Pour cela nous nous sommes référés aux variables identifiées lors de notre phase d'exploration des données (Exploration et Dataviz) et qui montraient des variations significatives de notre variable cible.

Nous avons donc retenu le quartier ('ProperCase'), le sous-quartier ('IncGeo_WardName'), la catégorie d'incident ('IncidentGroup'), le jour de la semaine calculé à partir de la date de l'appel ('DateOfCall'), le type de service spécial ('SpecialServiceType'), l'heure d'appel ('HourOfCall') et enfin le type de retard lorsque renseigné ('DelayCodeId'). A partir de ces variables catégorielles nous avons calculé les temps moyens observés sur l'année antérieure à nos échantillons d'entraînement et de test. Exemple : temps moyens observés en 2019 utilisés pour nos échantillons d'entraînement et de test sur l'année 2020.

La heatmap de corrélation obtenue à l'issue de cette phase d'enrichissement montre bien que les statistiques de l'année antérieure sont susceptibles d'aider nos modèles de prédiction.

Graphique 16. *Heatmap*



• Modèle LinearRegression

Le coefficient de détermination du modèle sur l'échantillon d'entraînement est de 0.5855081424710049

Le coefficient de détermination du modèle sur l'échantillon de test est de 0.5784596619364113

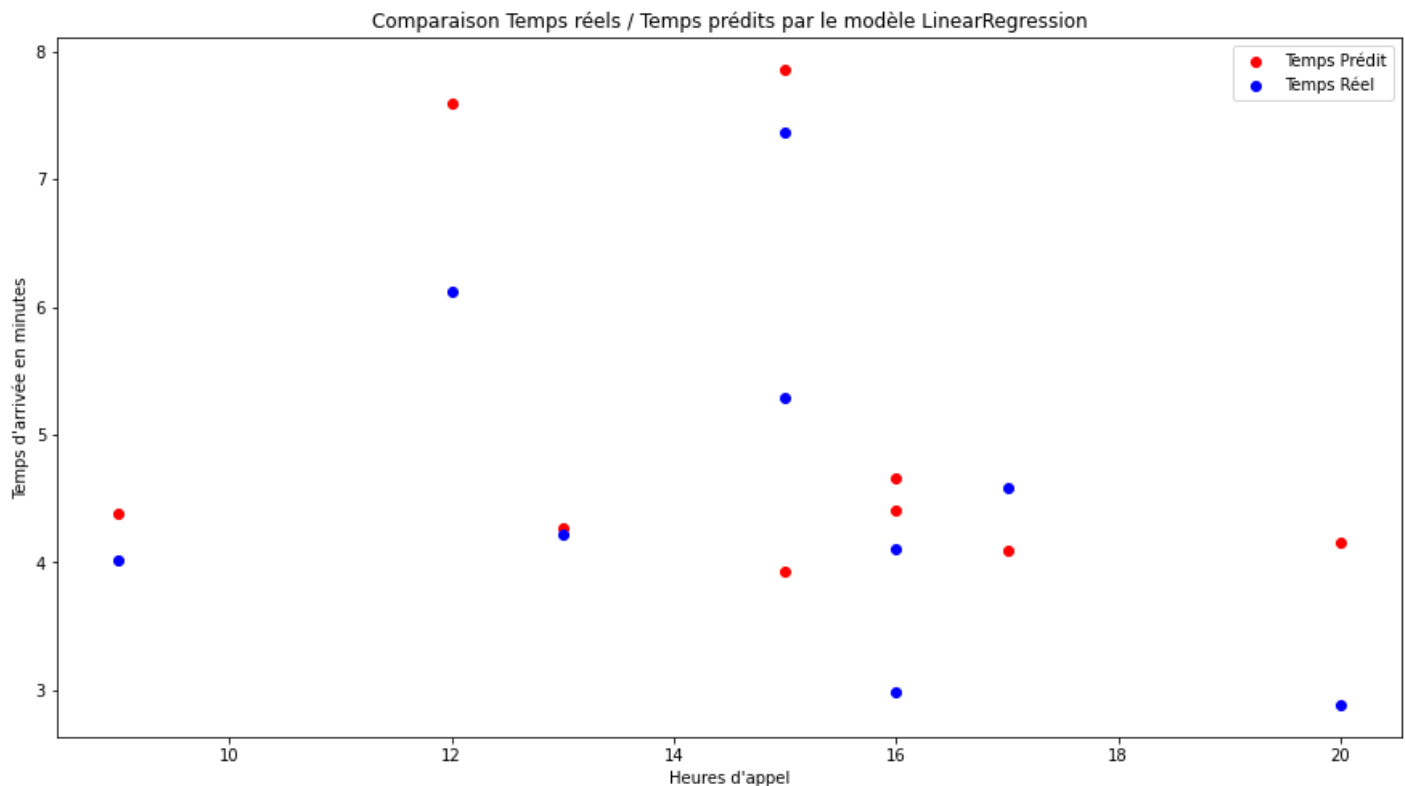
Le mean_squared_error sur l'échantillon d'entraînement est de 0.4164908522480654

Le mean_squared_error sur l'échantillon de test est de 0.41230732117478597

Les scores sont faibles mais relativement proches entre nos 2 échantillons.

Le graphique suivant nous montre sur le quartier de Merton pour un lundi et pour des services spéciaux de l'échantillon de test les écarts entre les temps réels d'arrivée des pompiers et les temps prédits par le modèle LinearRegression à différentes heures de la journée.

Graphique 17. *Visualisation Modèle LinearRegression*



• Modèle SGDRegressor

Le score du modèle sur l'échantillon d'entraînement est de 0.583734580320258

Le score du modèle sur l'échantillon de test est de 0.5775630142437822

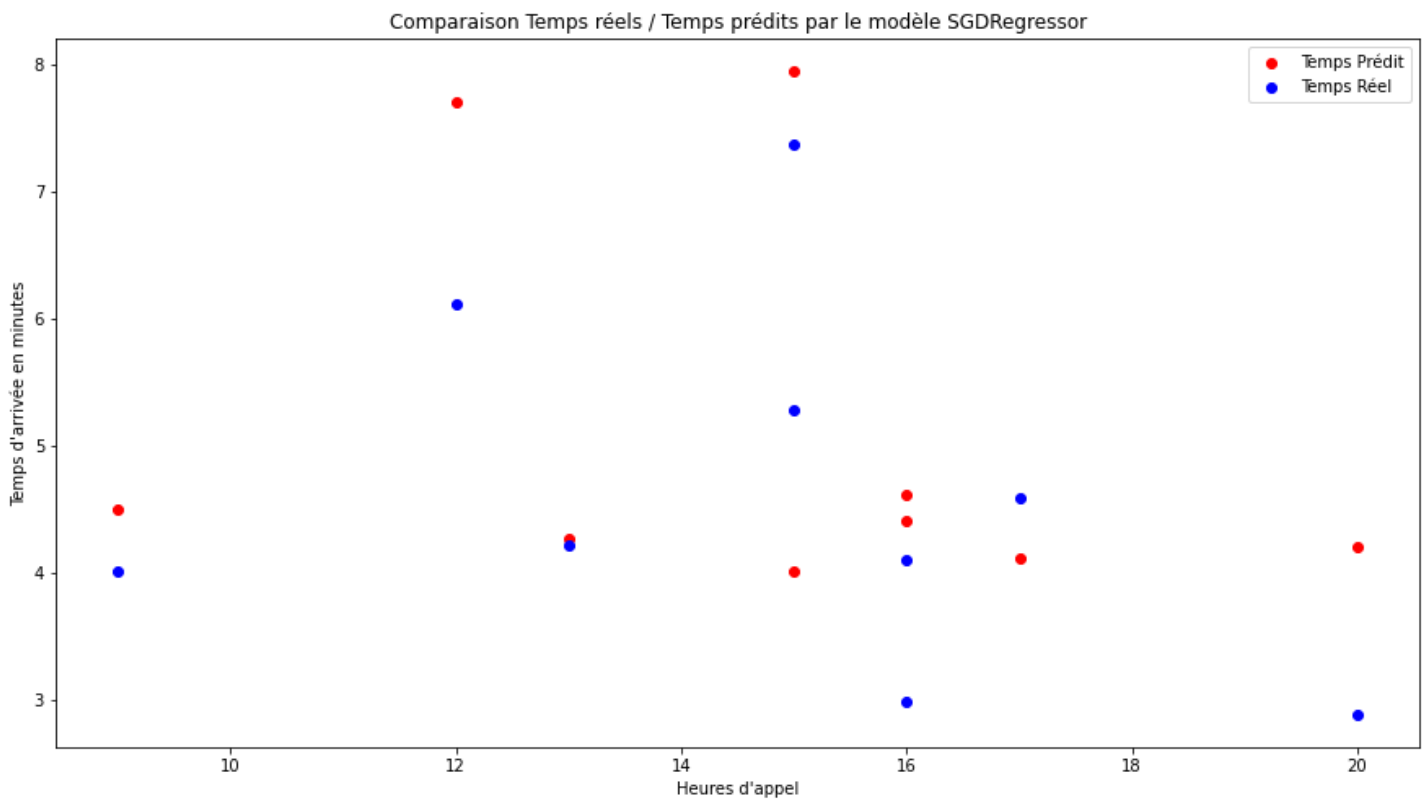
Le mean_squared_error sur l'échantillon d'entraînement est de 0.41548348577161687

Le mean_squared_error sur l'échantillon de test est de 0.4256085013915583

Les scores sont faibles mais relativement proches entre nos 2 échantillons.

Le graphique suivant nous montre sur le quartier de Merton pour un lundi et pour des services spéciaux de l'échantillon de test les écarts entre les temps réels d'arrivée des pompiers et les temps prédits par le modèle SGDRegressor à différentes heures de la journée.

Graphique 18. *Visualisation Modèle SGDRegressor*



● Modèle GradientBoostingRegressor

Le score du modèle sur l'échantillon d'entraînement est de 0.6156006822570951

Le score du modèle sur l'échantillon de test est de 0.6110778763441289

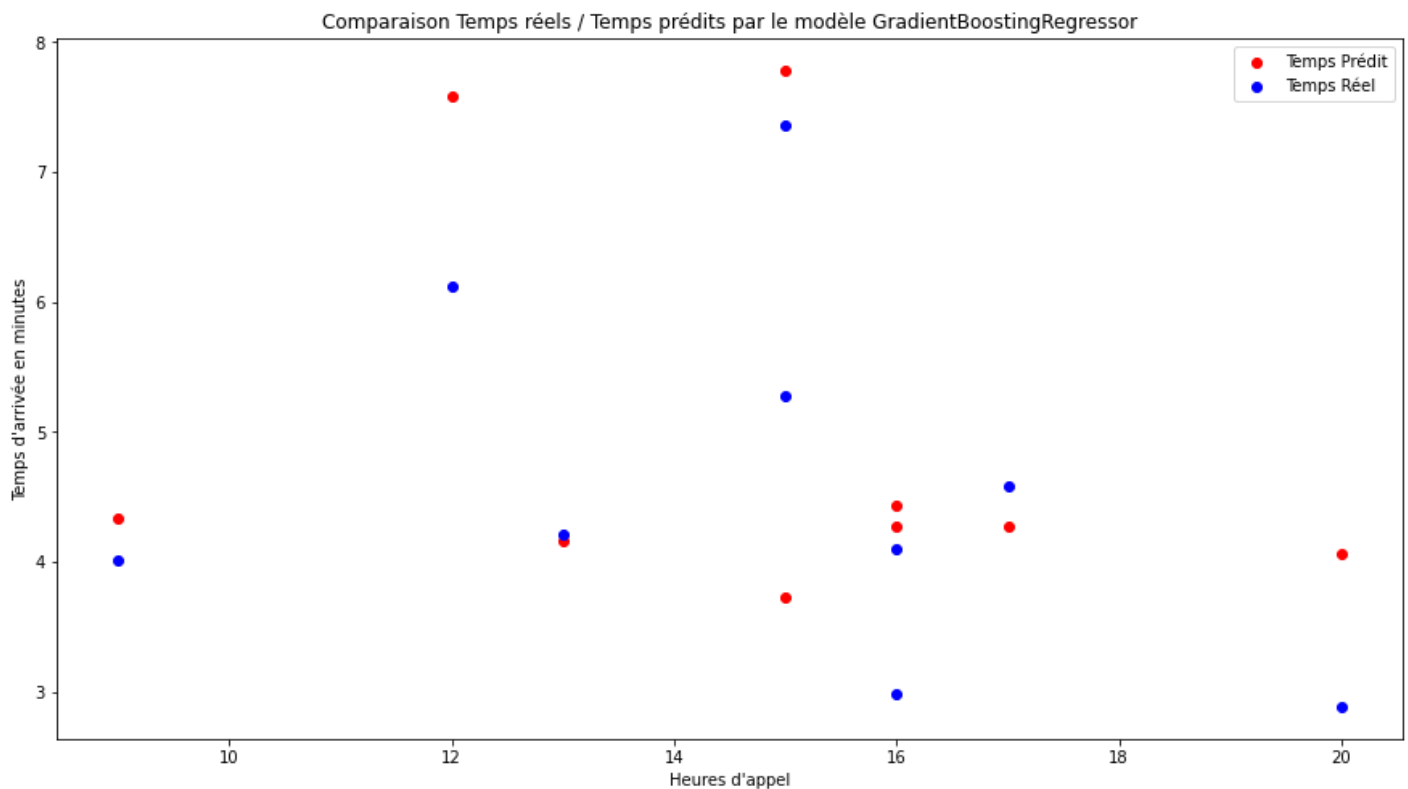
Le mean_squared_error sur l'échantillon d'entraînement est de 0.386885187097919

Le mean_squared_error sur l'échantillon de test est de 0.3788480879452526

Les scores et les erreurs quadratiques moyennes sont un peu meilleurs que les modèles précédents. De plus les scores et les erreurs quadratiques moyennes sont très proches entre les échantillons d'entraînement et de test ce qui montre une certaine stabilité du modèle.

Le graphique suivant nous montre sur le quartier de Merton pour un lundi et pour des services spéciaux de l'échantillon de test les écarts entre les temps réels d'arrivée des pompiers et les temps prédits par le modèle GradientBoostingRegressor à différentes heures de la journée.

Graphique 19. Visualisation Modèle GradientBoostingRegressor



- Bilan Machine Learning -

Le modèle avec les meilleures performances est le modèle *GradientBoostingRegressor*.

Les scores sont légèrement meilleurs que pour les autres modèles.

L'apport des nouvelles colonnes calculées à partir des variables catégorielles significatives a été bénéfique.

Il reste quelques grands écarts entre la réalité et les prédictions qui selon nous proviennent d'aléas externes à nos données.

On observe que de nombreuses prédictions sur 2020 donnent des temps d'arrivée des pompiers supérieures à la réalité probablement à cause des conditions de circulation meilleures en 2020 par rapport à 2019 du fait des effets de la crise sanitaire.

V. CONCLUSIONS

L'objectif principal du projet, qui est « La prédiction du temps d'intervention des Pompiers de Londres », a été atteint :

Le modèle de Machine Learning que nous avons retenu avec la meilleure performance est le :

- Gradient Boosting Regressor, avec un coefficient de détermination de 62 %.

Ce modèle se rapproche de la réalité, avec une marge de progression potentielle de 38% à ce jour.

Nous pensons que ce modèle pourrait encore être amélioré en confortant les calculs de prédiction avec l'étude des variables et facteurs suivants :

- Superficie / Densité de la population par quartier (règle sur les implantations de casernes de pompiers en fonction de la densité et / ou superficie d'intervention.)
- Informations sur le trafic (ratio de feux rouges par nombre d'habitants, signalétique routière, etc...)
- Kilométrage des routes

L'ensemble de ce projet pourrait être utile dans un futur proche, à la fois pour :

- Outil orienté utilisateur / App :

Aujourd'hui, seul le numéro des secours permet de rentrer en contact avec les pompiers.

Nous pouvons imaginer, demain, qu'une fois raccroché, une application mobile puisse prendre le relais afin de prédire le temps d'arrivée des pompiers, géolocaliser le véhicule et informer l'utilisateur, par exemple, des 1^{ers} gestes de secours à effectuer en attendant l'arrivée sur les lieux et donc de sauver des vies.

Cela pourrait également servir de support de communication à la communauté sur les actions et l'actualité des pompiers de leurs quartiers.

- Outil orienté pompiers :

- Améliorer l'organisation entre casernes (classement de l'efficacité des temps d'arrivée).
- Faciliter et automatiser une partie du travail de la personne chargée de prendre les appels en pouvant prédire le temps d'arrivée aux victimes.
- Proposer notre projet de prédiction à la brigade de pompiers de Paris.
- Appliquer notre modèle à d'autres services comme les ambulances, la police, etc..., où le temps d'intervention joue un rôle crucial.

VI. DATA SOURCE

Le jeu de données utilisé pour cette étude provient du site officiel du Gouvernement du Royaume-Uni, London Datastore, London Fire Brigade (LFB).

Ces données sont de libre accès comme indiqué ci-dessous :

Le premier jeu de données ("Incidents") :

<https://data.london.gov.uk/dataset/london-fire-brigade-incident-records>

Le second jeu de données ("Mobilisation") :

<https://data.london.gov.uk/dataset/london-fire-brigade-mobilisation-records>

