

Rapport d'Exploration et d'Analyse Incident from 2018

Étapes de l'Analyse et du Nettoyage des Données :

1. Importation des Bibliothèques :
 - o Importation des bibliothèques pandas, matplotlib.pyplot, et seaborn pour la manipulation des données et la visualisation.
2. Chargement des Données :
 - o La taille de notre excel étant volumineuse, les données ont été chargées depuis un fichier .xlsb stocké sur Google Drive avec pd.read_excel() utilisant le moteur pyxlsb.
3. Aperçu des Données :
 - o Les premières lignes des données ont été affichées pour avoir une vue d'ensemble du dataset.

```

Incident Data:
  IncidentNumber DateOfCall CalYear TimeOfCall HourOfCall \
0 000008-01012018 43101 2018 0.003067 0
1 000009-01012018 43101 2018 0.003125 0
2 000010-01012018 43101 2018 0.003171 0
3 000011-01012018 43101 2018 0.003449 0
4 000014-01012018 43101 2018 0.005405 0

  IncidentGroup StopCodeDescription SpecialServiceType PropertyCategory \
0 False Alarm AFA NaN Non Residential
1 False Alarm AFA NaN Non Residential
2 Fire Secondary Fire NaN Outdoor Structure
3 Special Service Special Service RTC Road Vehicle
4 Fire Primary Fire NaN Road Vehicle

  PropertyType ... FirstPumpArriving_AttendanceTime \
0 Mosque ... 348.0
1 Pub/wine bar/bar ... 144.0
2 Common external bin storage area ... 232.0
3 Multiple Vehicles ... 22.0
4 Car ... 241.0

  FirstPumpArriving_DeployedFromStation SecondPumpArriving_AttendanceTime \
0 Finchley NaN
1 Beckenham NaN
2 Southgate NaN
3 Enfield NaN
4 Stratford NaN

  SecondPumpArriving_DeployedFromStation NumStationsWithPumpsAttending \
0 NaN 1.0
1 NaN 1.0
2 NaN 1.0
3 NaN 1.0
4 NaN 1.0

  NumPumpsAttending PumpCount PumpMinutesRounded Notional Cost (£) NumCalls
0 1.0 1 60 328 1.0
1 1.0 1 60 328 1.0
2 1.0 1 60 328 1.0
3 1.0 1 60 328 1.0
4 1.0 1 60 328 6.0

[5 rows x 39 columns]

```

- o Les types de variables ont été listés, notamment IncidentNumber, DateOfCall, CalYear, HourOfCall, IncidentGroup, et plusieurs autres colonnes.

Incident Data Variables:

	Variable	Type
0	IncidentNumber	object
1	DateOfCall	int64
2	CalYear	int64
3	TimeOfCall	float64
4	HourOfCall	int64
5	IncidentGroup	object
6	StopCodeDescription	object
7	SpecialServiceType	object
8	PropertyCategory	object
9	PropertyType	object
10	AddressQualifier	object
11	Postcode_full	object
12	Postcode_district	object
13	UPRN	int64
14	USRN	int64
15	IncGeo_BoroughCode	object
16	IncGeo_BoroughName	object
17	ProperCase	object
18	IncGeo_WardCode	object
19	IncGeo_WardName	object
20	IncGeo_WardNameNew	object
21	Easting_m	float64
22	Northing_m	float64
23	Easting_rounded	int64
24	Northing_rounded	int64
25	Latitude	float64
26	Longitude	float64
27	FRS	object
28	IncidentStationGround	object
29	FirstPumpArriving_AttendanceTime	float64
30	FirstPumpArriving_DeployedFromStation	object
31	SecondPumpArriving_AttendanceTime	float64
32	SecondPumpArriving_DeployedFromStation	object
33	NumStationsWithPumpsAttending	float64
34	NumPumpsAttending	float64
35	PumpCount	int64
36	PumpMinutesRounded	int64
37	Notional Cost (£)	int64
38	NumCalls	float64

4. Vérification des Valeurs Manquantes :

o Colonne avec Valeurs Manquantes :

IncidentNumber	0
DateOfCall	0
CalYear	0
TimeOfCall	0
HourOfCall	0
IncidentGroup	6
StopCodeDescription	0
SpecialServiceType	484867
PropertyCategory	6
PropertyType	6
AddressQualifier	1
Postcode_full	412193
Postcode_district	0
UPRN	0
USRN	0
IncGeo_BoroughCode	0
IncGeo_BoroughName	0
ProperCase	0
IncGeo_WardCode	425
IncGeo_WardName	425
IncGeo_WardNameNew	425
Easting_m	412193
Northing_m	412193
Easting_rounded	0
Northing_rounded	0
Latitude	412193
Longitude	412193
FRS	0
IncidentStationGround	1
FirstPumpArriving_AttendanceTime	42383
FirstPumpArriving_DeployedFromStation	42389
SecondPumpArriving_AttendanceTime	473337
SecondPumpArriving_DeployedFromStation	473342
NumStationsWithPumpsAttending	7980
NumPumpsAttending	7980
PumpCount	0
PumpMinutesRounded	0
Notional Cost (£)	0
NumCalls	20
Year	0

- **IncidentGroup** : 6 valeurs manquantes.
- **PropertyCategory** : 6 valeurs manquantes.
- **PropertyType** : 6 valeurs manquantes.
- **AddressQualifier** : 1 valeur manquante.
- **IncidentStationGround** : 1 valeur manquante.
- **FirstPumpArriving_AttendanceTime** : 42,383 valeurs manquantes.
- **FirstPumpArriving_DeployedFromStation** : 42,389 valeurs manquantes.
- **SecondPumpArriving_AttendanceTime** : 473,337 valeurs manquantes.

- **SecondPumpArriving_DeployedFromStation** : 473,342 valeurs manquantes.
- **NumStationsWithPumpsAttending** : 7,980 valeurs manquantes.
- **NumPumpsAttending** : 7,980 valeurs manquantes.
- **Postcode_full**, **Easting_m**, **Northing_m**, **Latitude**, **Longitude** : plus de 412,193 valeurs manquantes.
- **SpecialServiceType** : 484,867 valeurs manquantes.

5. Imputation des Valeurs Manquantes :

- **Colonnes avec Peu de Valeurs Manquantes (< 1%) :**
 - **IncidentGroup** : Remplacement des valeurs manquantes par le mode (False Alarm).
 - **PropertyCategory** : Imputation des valeurs manquantes avec le mode (Non Residential).
 - **PropertyType** : Remplissage des valeurs manquantes par le mode (Car).
 - **NumCalls** : Remplacement des valeurs manquantes par la médiane (1.0).
 - **AddressQualifier** et **IncidentStationGround** : Remplacement par le mode (Unknown).
- **Colonnes avec Valeurs Manquantes Modérées (1% - 10%) :**
 - **FirstPumpArriving_AttendanceTime** : Imputation des valeurs manquantes avec la médiane (232.0 secondes).
 - **FirstPumpArriving_DeployedFromStation** : Utilisation du mode pour remplir (Unknown).
 - **NumStationsWithPumpsAttending** et **NumPumpsAttending** : Remplacement des valeurs manquantes par la médiane (1.0).
- **Colonnes avec Beaucoup de Valeurs Manquantes (> 60%) :**

- **SpecialServiceType** : Suppression de la colonne en raison de son manque de pertinence et du nombre élevé de valeurs manquantes.
- **Localisation (Postcode_full, Easting_m, Northing_m, Latitude, Longitude)** : Remplacement des valeurs manquantes par Unknown.
- **SecondPumpArriving_AttendanceTime** et **SecondPumpArriving_DeployedFromStation** : Suppression en raison de leur manque de pertinence.

6. Traitement des Lignes avec Beaucoup de Valeurs Manquantes :

- o Les lignes avec plus de 10 valeurs manquantes ont été supprimées pour améliorer la qualité des données.

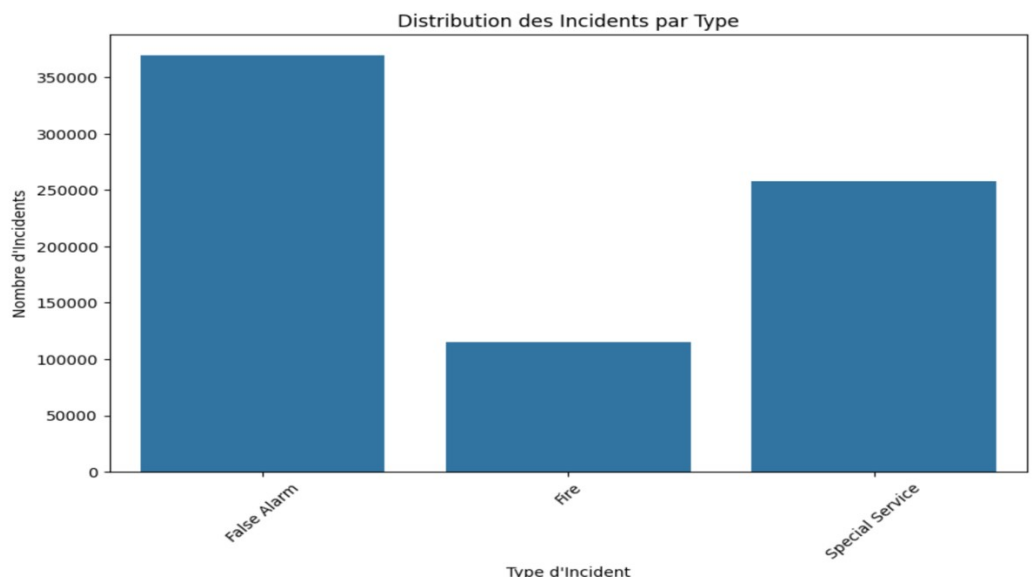
7. Vérification Finale des Valeurs Manquantes :

- o Une vérification finale a confirmé que toutes les valeurs manquantes restantes ont été correctement traitées.

```
IncidentNumber      0
DateOfCall          0
CalYear             0
TimeOfCall          0
HourOfCall          0
IncidentGroup       0
StopCodeDescription 0
PropertyCategory    0
PropertyType        0
AddressQualifier    0
Postcode_full       0
Postcode_district   0
UPRN                0
USRN                0
IncGeo_BoroughCode  0
IncGeo_BoroughName  0
ProperCase          0
IncGeo_WardCode     381
IncGeo_WardName     381
IncGeo_WardNameNew  381
Easting_m           0
Northing_m          0
Easting_rounded     0
Northing_rounded    0
Latitude            0
Longitude           0
FRS                 0
IncidentStationGround 0
FirstPumpArriving_AttendanceTime 0
FirstPumpArriving_DeployedFromStation 0
NumStationsWithPumpsAttending 0
NumPumpsAttending   0
PumpCount           0
PumpMinutesRounded  0
Notional Cost (£)   0
NumCalls            0
Year                0
dtype: int64
```

8. Dataviz pertinentes

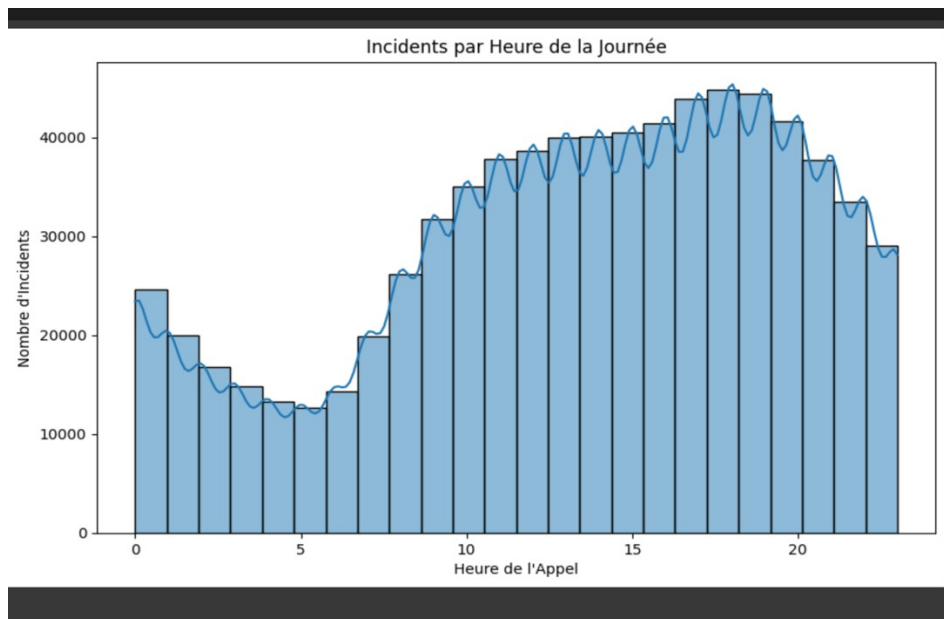
Distribution des Incidents par Type



- **Interprétation :**

- Ce graphique montre que la majorité des incidents sont des **False Alarms**, suivis par les **Special Service** et enfin par les incidents de **Fire**.
- La prépondérance des **False Alarms** peut indiquer un problème avec les systèmes d'alarme ou des protocoles d'activation inappropriés. Cela peut nécessiter des mesures pour réduire les fausses alertes, comme des ajustements dans la sensibilité des systèmes d'alarme.
- Les **Special Service** regroupent divers incidents comme les inondations, les accidents de la route, ou d'autres services non liés au feu, indiquant une part importante des activités des services d'urgence au-delà des simples interventions contre le feu.

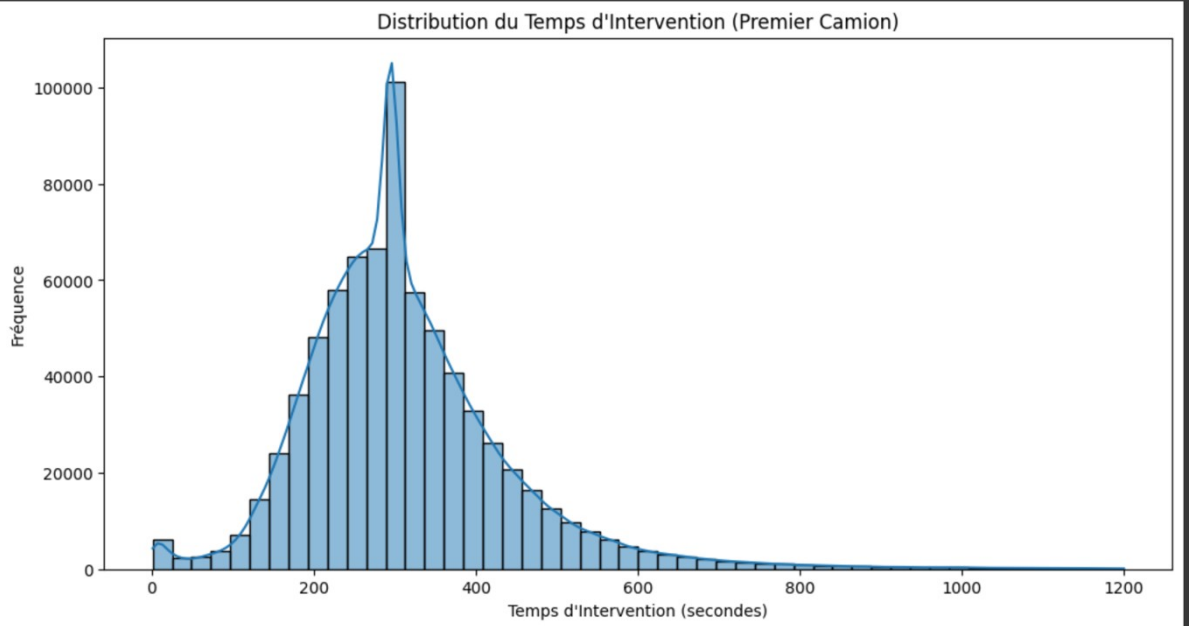
Incidents par Heure de la Journée



- **Interprétation :**

- Le nombre d'incidents augmente progressivement à partir de 6h du matin pour atteindre un pic entre 10h et 20h.
- Cette tendance pourrait s'expliquer par une activité humaine plus intense pendant la journée, avec des heures de travail, de trafic, et d'autres activités où des incidents sont plus susceptibles de se produire.
- Les heures creuses, tôt le matin et tard le soir, enregistrent moins d'incidents, ce qui correspond aux périodes de moindre activité publique.

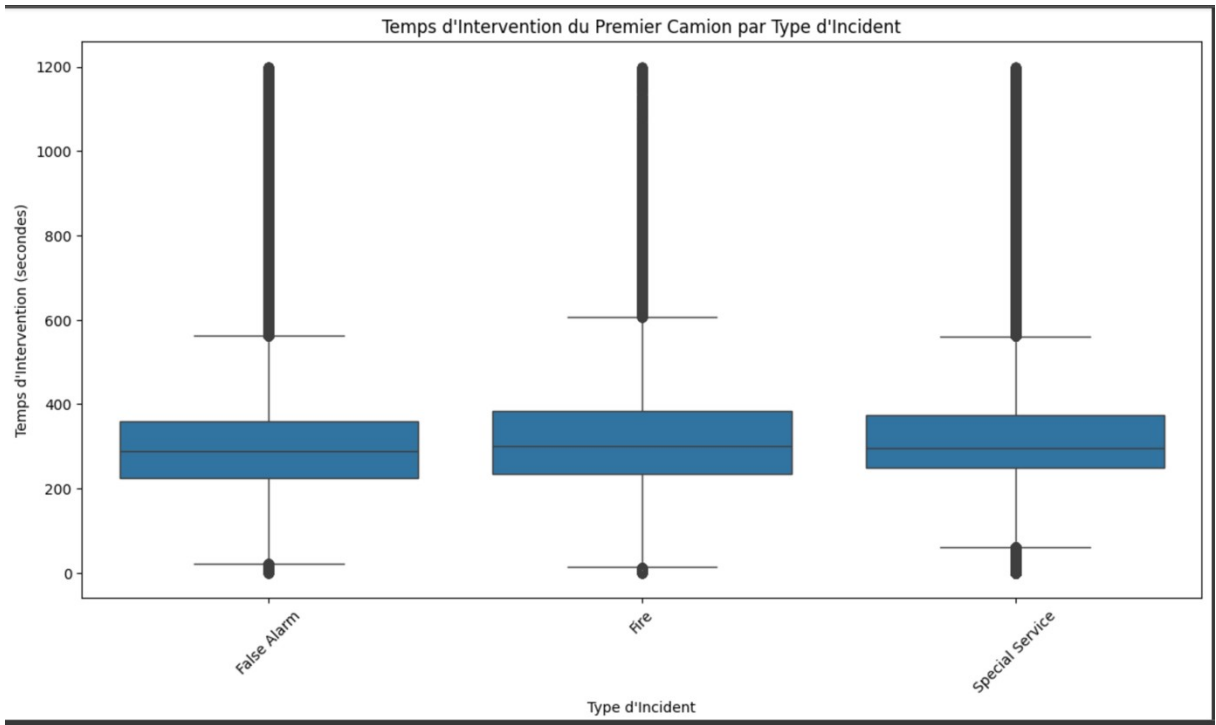
Distribution du Temps d'Intervention (Premier Camion)



- **Interprétation :**

- o La plupart des interventions du premier camion se situent entre 200 et 400 secondes (environ 3 à 7 minutes).
- o Ce temps reflète l'efficacité des services d'intervention pour atteindre rapidement les lieux d'incident, ce qui est crucial pour limiter les dommages et assurer la sécurité.
- o La distribution montre une asymétrie à droite, avec quelques interventions beaucoup plus longues, probablement dues à des incidents plus éloignés ou des obstacles particuliers rencontrés en route.

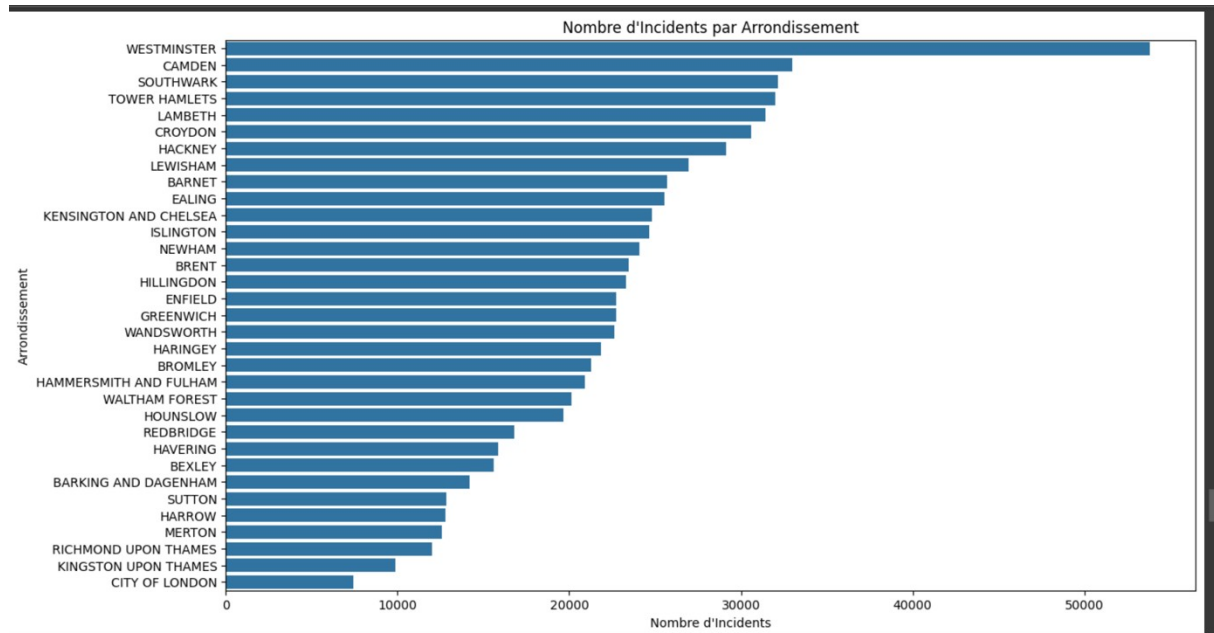
Box Plot du Temps d'Intervention par Type d'Incident



- **Interprétation :**

- o Les temps d'intervention sont similaires pour les **False Alarms**, **Fire**, et **Special Service**, avec une médiane autour de 350 secondes.
- o Les **False Alarms** présentent des valeurs aberrantes importantes, indiquant des cas où l'intervention a été retardée. Ceci pourrait être dû à des difficultés d'accès ou de localisation de l'incident.
- o Le temps d'intervention constant pour les différents types d'incidents suggère que les équipes sont déployées de manière similaire quel que soit
- o le type d'incident, optimisant les temps de réponse.

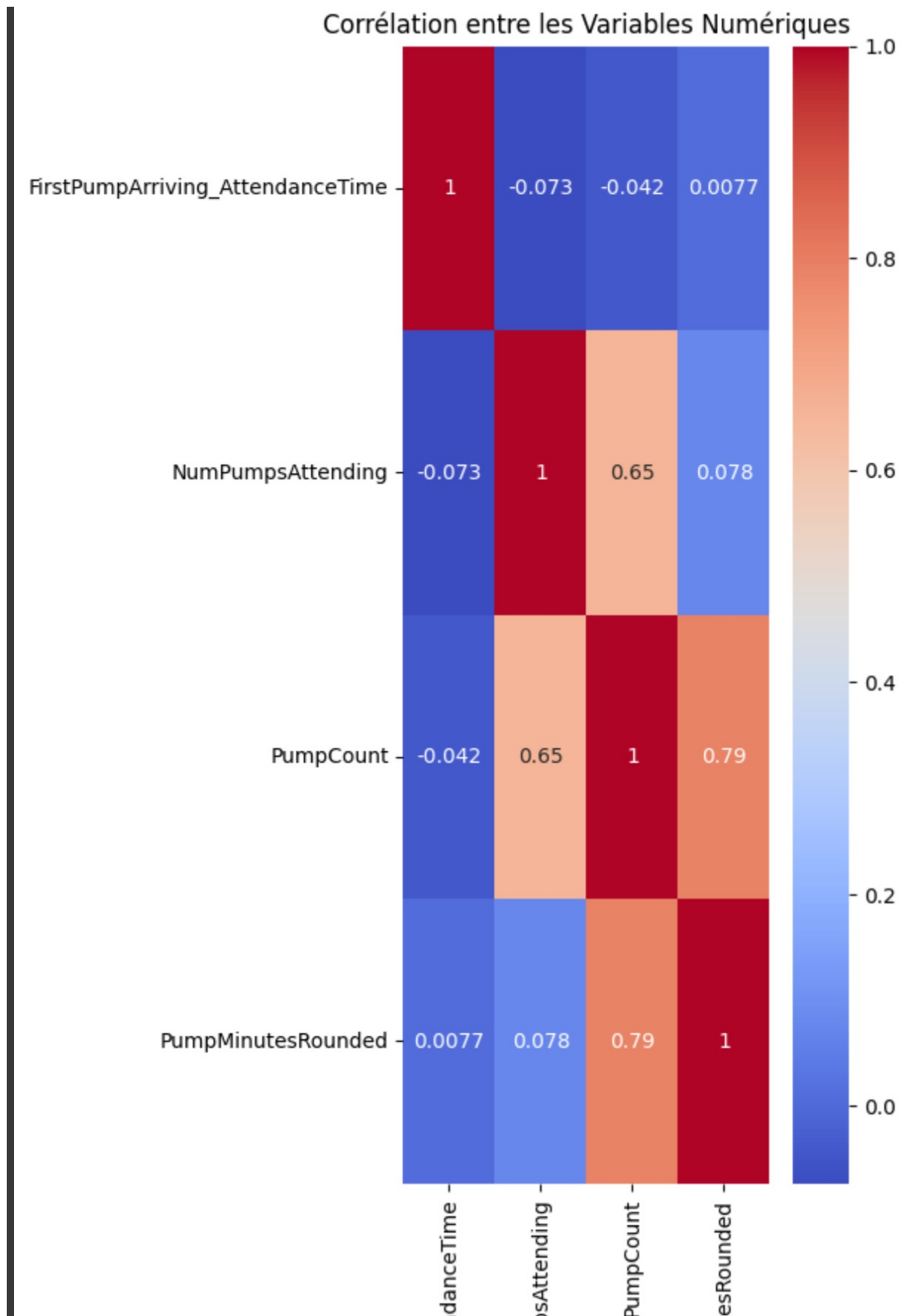
Nombre d'Incidents par Arrondissement



- **Interprétation :**

- **Westminster** et **Camden** sont les arrondissements avec le plus grand nombre d'incidents, ce qui peut s'expliquer par leur forte densité de population et d'activités commerciales.
- Les arrondissements périphériques comme **Kingston upon Thames** et **la City of London** enregistrent moins d'incidents, probablement en raison de leur plus faible densité et des activités moins intenses par rapport aux zones centrales.
- Ce graphique peut aider à identifier les zones nécessitant une allocation plus importante de ressources pour la prévention et l'intervention.

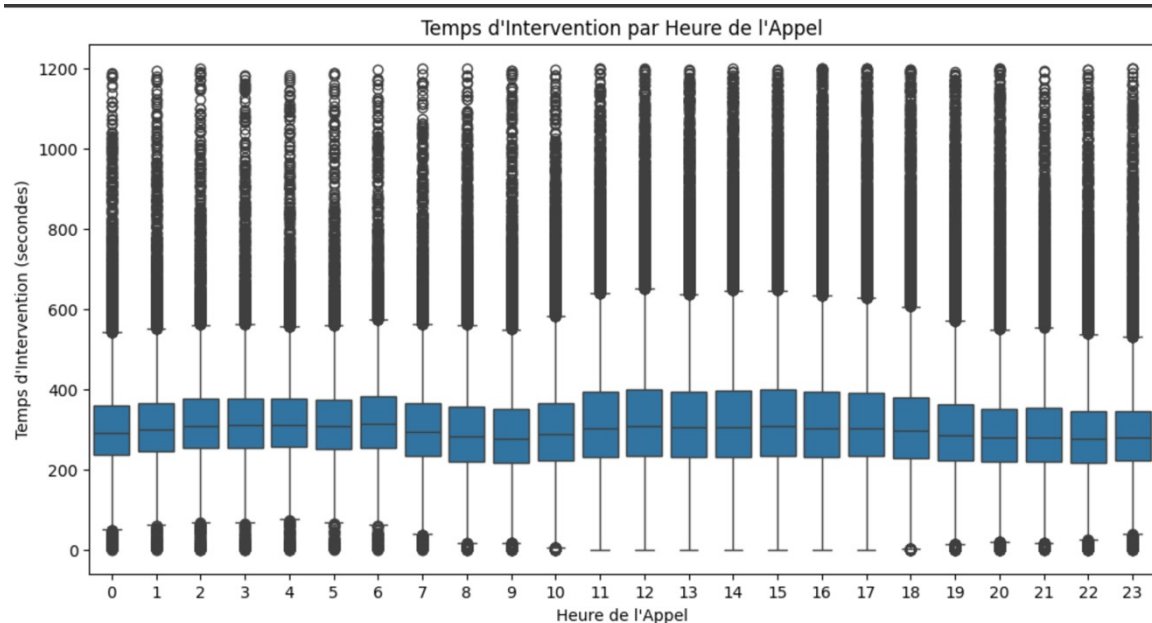
Heatmap de Corrélation entre les Variables Numériques



- **Interprétation :**

- o Le graphique montre une faible corrélation entre le **FirstPumpArriving_AttendanceTime** et les autres variables, indiquant que le temps d'intervention n'est pas fortement influencé par le nombre de pompes ou le nombre de stations avec des pompes en intervention.
- o Une corrélation modérée entre **NumPumpsAttending** et **PumpCount (0.65)** suggère qu'un plus grand nombre de pompes présentes sur les lieux est lié au nombre total de pompes déployées.
- o La faible corrélation générale suggère que d'autres facteurs externes (comme le trafic ou les conditions routières) pourraient avoir un impact plus significatif sur les temps d'intervention.

Temps d'Intervention par Heure de l'Appel



• Interprétation :

- o Le temps d'intervention médian reste relativement stable tout au long de la journée, avec une légère augmentation en fin de matinée et en début d'après-midi.
- o Les heures entre **10h et 16h** montrent des temps d'intervention légèrement plus élevés, potentiellement dus à un trafic plus dense ou une plus grande complexité des incidents pendant ces heures de forte activité.
- o Les temps d'intervention plus longs en période de pointe pourraient indiquer la nécessité d'améliorer les protocoles de

déploiement pendant ces périodes pour maintenir des réponses rapides.