IA Prédiction du flux routier

[wiki.lafabriquedesmobilites.fr/wiki/IA\_Prediction\_de\_flux\_routier](https://wiki.lafabriquedesmobilites.fr/wiki/IA_Prediction_de_flux_routier)

L'objet de ce commun est de produire de la connaissance pour identifier quelles sont les conditions minimales pour alimenter une IA capable de prédire les flux de circulation à un certain horizon de temps.

Table des matières

[I. Introduction 2](#_Toc47891052)

[II. Livrables 2](#_Toc47891053)

[III. Description des données à disposition en Open Data 3](#_Toc47891054)

[Données du trafic des autoroutes au Luxembourg 3](#_Toc47891055)

[Données du trafic des autoroutes 4](#_Toc47891056)

[Données des événements trafic 4](#_Toc47891057)

[Données météo 4](#_Toc47891058)

[Données du trafic place général de gaulle (mouans sartoux, France) 5](#_Toc47891059)

[IV. Analyse des données du trafic des autoroutes au Luxembourg 6](#_Toc47891060)

[Description des données 6](#_Toc47891061)

[Exploration graphique 6](#_Toc47891062)

[Prédiction d’indicateurs à l’aide d’un algorithme de classification 8](#_Toc47891063)

[V. Analyse des données de la caméra telraam 11](#_Toc47891064)

[Description des données 11](#_Toc47891065)

[Exploration graphique des données 12](#_Toc47891066)

[Gestion des valeurs manquantes 20](#_Toc47891067)

[Prédiction 21](#_Toc47891068)

[VI. Compte kaggle pour la FabMob 22](#_Toc47891069)

[VII. Vue globale 22](#_Toc47891070)

# Introduction

Les données à disposition en Open Data pour cette 1ère exploration sont

* Les données du trafic des autoroutes du Luxembourg à travers 186 caméras via [www.cita.lu](http://www.cita.lu)
* Les données de la caméra place général de gaulle (Mouans-Sartoux) via [www.telraam.net](http://www.telraam.net)

A partir de ces données, les questions à explorer sont

* Valider l’hypothèse qu'il est possible d'avoir une IA qui prédise les flux de circulation avec 15 minutes d’avance
* Identifier le faisceau de données nécessaires pour produire une IA pertinente
* Décrire les besoins pour mettre cet algorithme en production

Ce document résume les résultats de cette analyse.

D’un point de vue technique, le premier objectif est de développer une solution de captation de données, afin de valider la faisabilité et le faible cout de mise en œuvre d’une telle solution à l’aide des outils Open Source

# Livrables

Tous les sources sont disponibles ici : <https://github.com/fabmob/trafic>

#### Analyse

Pour la partie analyse, les outils utilisés sont les librairies python de « machine learning »  [scikit-learn.org/stable/](https://scikit-learn.org/stable/) [www.tensorflow.org/](https://www.tensorflow.org/) et [keras.io/](https://keras.io/)

#### La solution pour la captation des flux [cita.lu](http://cita.lu/) et telraam se compose de :

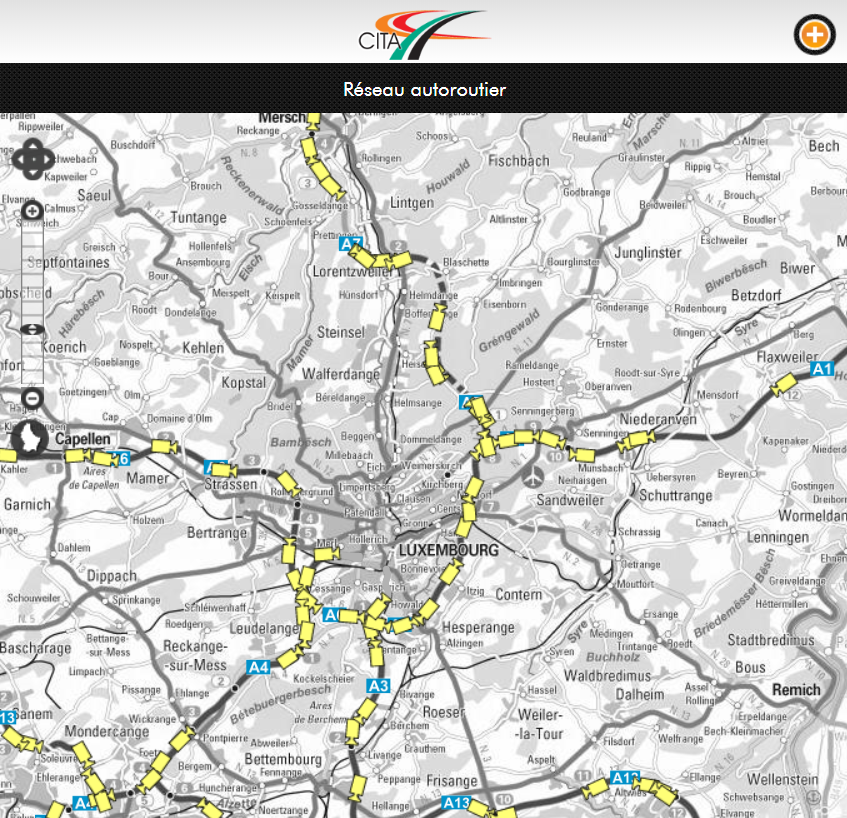
* Un batch d'alimentation : batch générique Open Source en java à l’aide du Framework Spring permettant de récolter les données des caméras ([spring.io/](https://spring.io/))
* Une base de donnée de type «Time Series » (timescaledb de PostgreSQL [www.timescale.com/](https://www.timescale.com/))
* Un outil de visualisation des données (Graphana [grafana.com/grafana/](https://grafana.com/grafana/))

Les livrables pour cette solution d’alimentation sont le code source (java + sql), la documentation de la mise en œuvre et des exemples de Dashboard de visualisation.

# Description des données à disposition en Open Data

## Données du trafic des autoroutes au Luxembourg

L’ administration des Ponts et Chaussées du Luxembourg met à disposition un ensemble de données ouvertes via la plate-forme de données luxembourgeoise [data.public.lu](http://data.public.lu/)  
( <https://data.public.lu/fr/organizations/administration-des-ponts-et-chaussees>  )



Les données suivantes peuvent être interrogées en temps réels au format DATEX II, format de données de trafic routier (norme européenne) (DATEX II : <http://trafic-routier.data.cerema.fr/la-norme-europeenne-datex-ii-a58.html> )

* Données du trafic des autoroutes
* Données des événements trafic
* Données météo

Données du trafic des autoroutes  
Pour chacune des 180 caméras sur les autoroutes (A1, A3, A4, A6, A7, A13, B40) les données suivantes sont exposées toutes les 5 minutes :  
Données :

* Identifiant de la caméra
* Timestamp de la mesure
* Sens de prise de vue
* Latitude et longitude de la caméra
* Distance de la caméra au début de l’autoroute (caméra au km x )
* Vitesse moyenne en km/h
* Concentration du trafic en pourcentage
* Flux du trafic en nombre de voitures par heure

Format : XML DATEX II  
URLs A1 : <https://www.cita.lu/info_trafic/datex/trafficstatus_a1>    
URL pour xx : <https://www.cita.lu/info_trafic/datex/trafficstatus_xx>   avec xx dans (A1, A3, A4, A6, A7, A13, B40)

Données des événements trafic  
L’API donne la liste des événements (situation) en cours.

On peut trouver les données suivantes à dispositions pour chaque situation, identifiant de la situation, timestamp de la création de la situation, position, description de l'événement.  
La description complète des données disponible est décrite ici : <https://diffusion-numerique.info-routiere.gouv.fr/IMG/pdf/interface-datexii-avectipi-2017-171124.pdf>

Le format est en XML DATEX II  
URL : <https://www.cita.lu/info_trafic/datex/situationrecord>

### Données météo

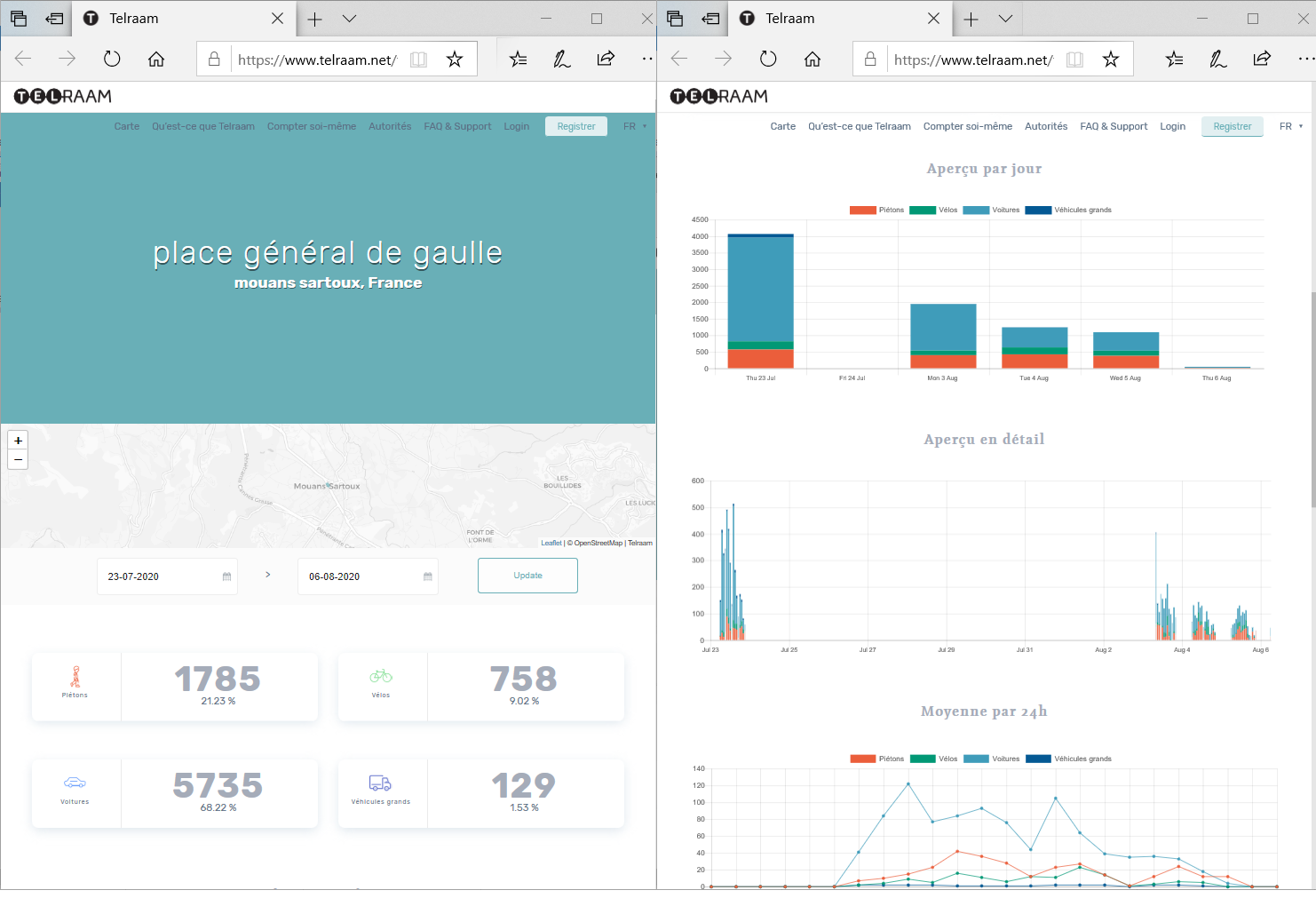
L’API donne des informations météos pour un ensemble de point de mesure.

* Identifiant du point de mesure
* Timestamp de la mesure
* Pourcentage d’humidité
* Précipitation en mm par heure
* Température de la surface de route
* Température de l'air
* Température de rosée
* Vitesse et direction du vent

Le format est en XML DATEX II  
URL : <https://www.cita.lu/info_trafic/datex/weather_dynamic.xml>

## Données du trafic place général de gaulle (mouans sartoux, France)

<https://www.telraam.net/fr/location/9000000411>



Cette caméra fait partie du dispositif Telraam qui met à disposition des citoyens intéressés des caméras de comptage. (<https://www.telraam.net/fr/what-is-telraam>)

Chaque heure la caméra relève les données suivantes :

* Pourcentage d'activité de la caméra
* Comptage des piétons, voitures, vélos, camions (total, gauche et droite de la rue)
* Histogramme des vitesses des voitures pour les intervalles [0-10[  [10-20[ [20- 30[     .....  [70 et plus [

Le format est en JSON  
URLs de récupération des données : <https://telraam-api.net/v0/reports/9000000411>

Documentation de l’API : <https://telraam.zendesk.com/hc/en-us/articles/360027325572-Want-more-data-Telraam-API>

Remarque : ce paragraphe décrit des données disponibles utilisables pour la prédiction du trafic. Dans la suite, uniquement les données issues des caméras sont utilisées pour cette exploration (Données cita du trafic des autoroutes Lux + Données telraam du trafic place général de gaulle)

# Analyse des données du trafic des autoroutes au Luxembourg

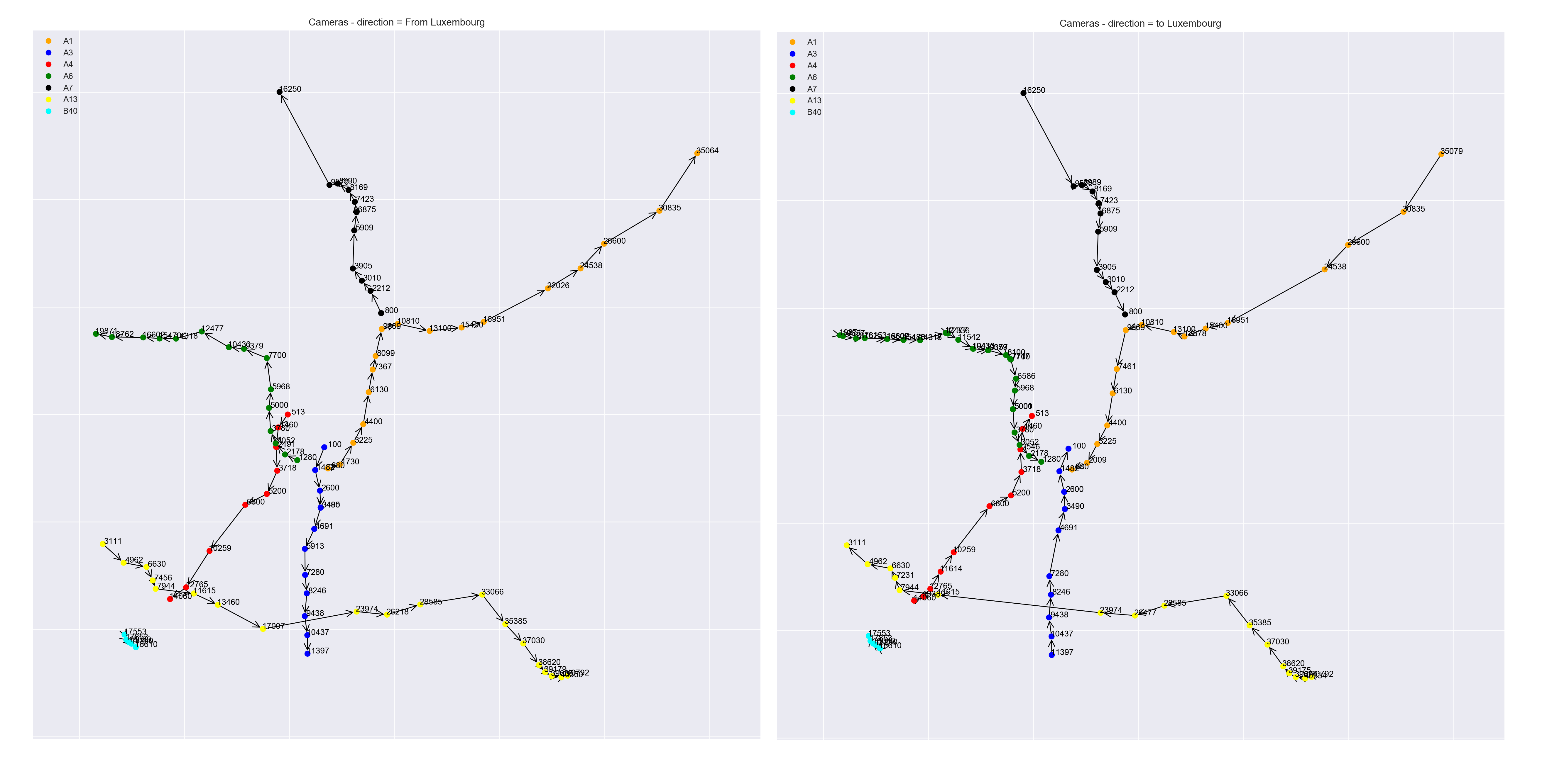
## Description des données

Les données analysées ont été captées sur les 186 caméras mises à disposition en Open Data par cita.lu entre le 19/11/2019 et le 26/12/2019.

Les données de chaque caméra sont mises à jour toutes les 5 minutes.

Pour cette période, cela représente un jeu de donnée de 1 887 379 enregistrements.

Si on extrapole à 1 année, cela représente environ 18 130 000 enregistrements.



*Figure : Les 186 caméras sur les autoroutes (A1, A3, A4, A6, A7, A13, B40)*

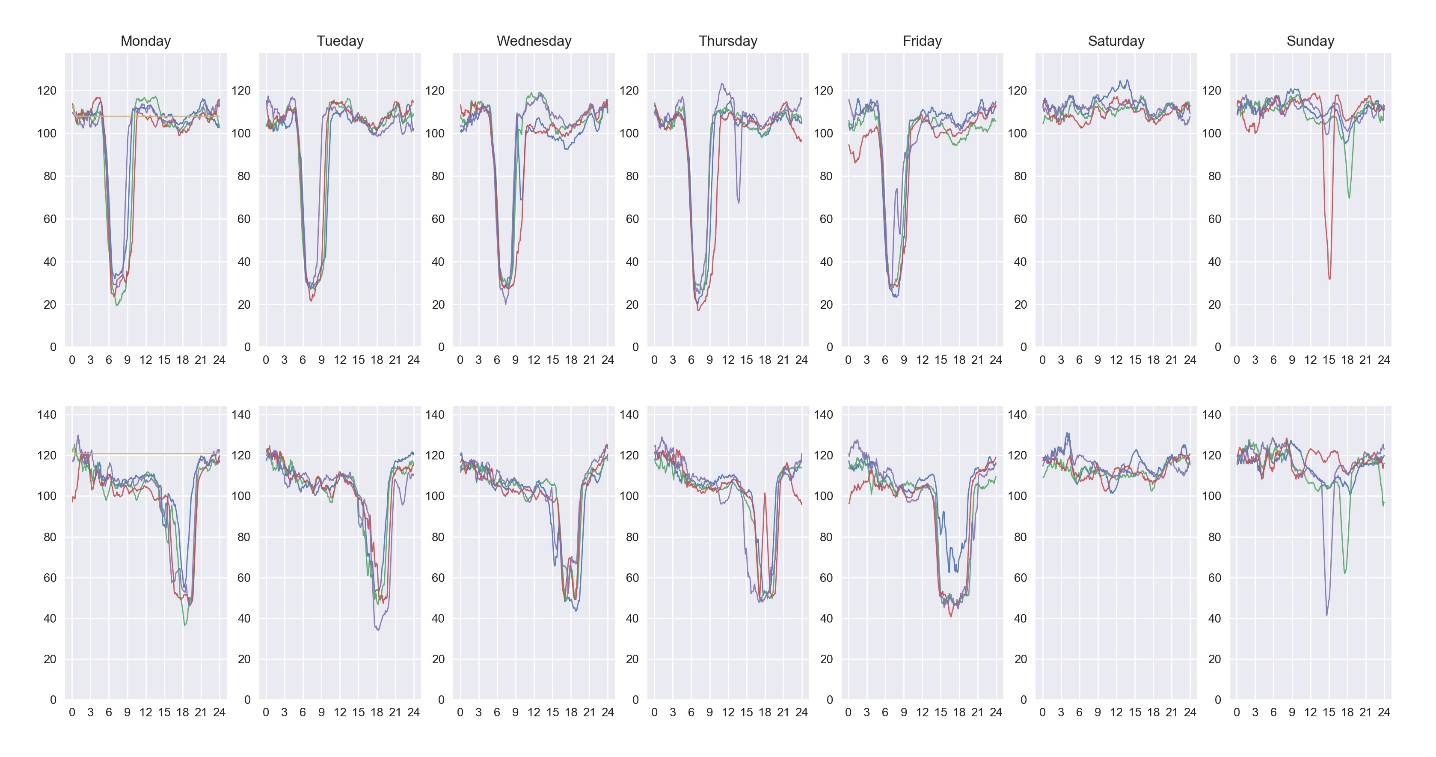
Les 3 mesures disponibles pour chaque caméra toutes les 5 minutes sont :

* La vitesse moyenne en km/h
* La concentration du trafic en pourcentage
* Le flux du trafic en nombre de voitures par heure

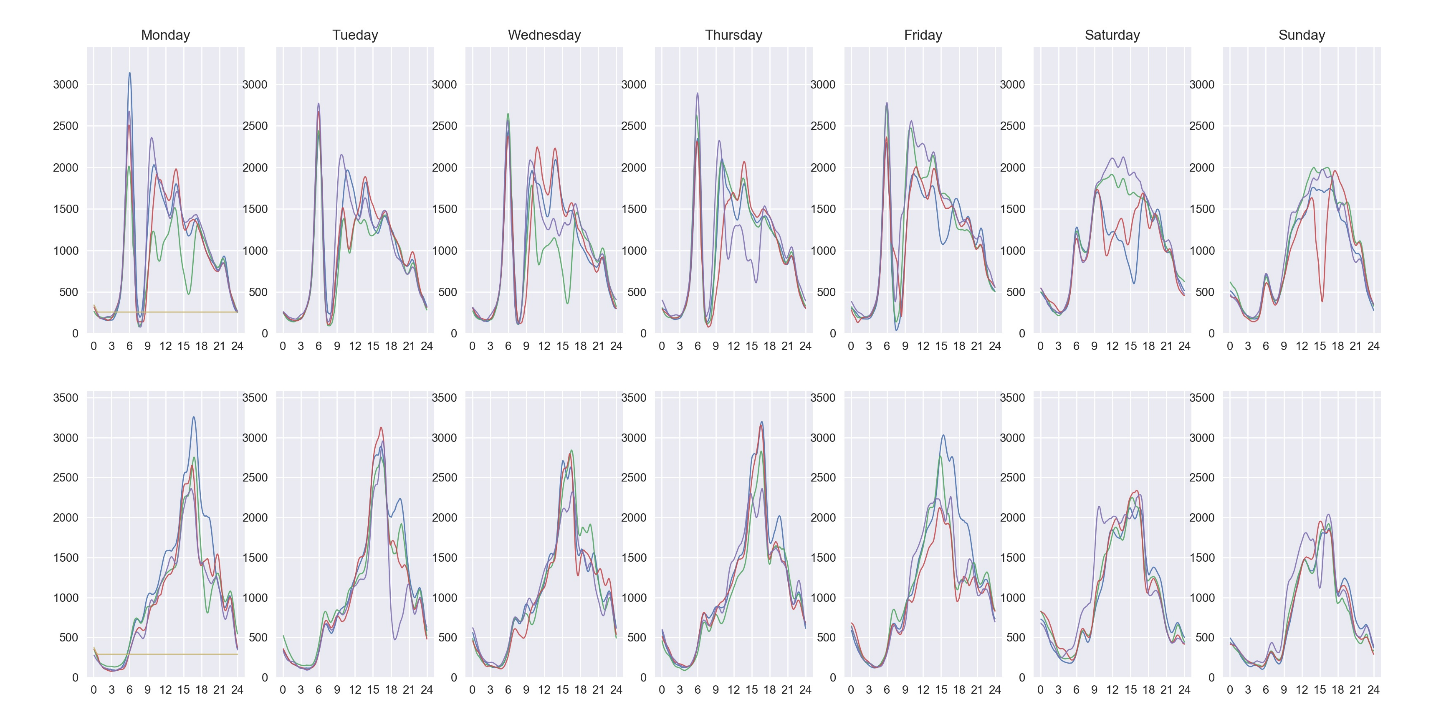
## Exploration graphique

Les graphiques suivants tracent ces 3 mesures pour la caméra au kilomètre 11 de l’A3, cette caméra se situe à la frontière de la France et du Luxembourg. Le trafic à la frontière connait des surcharges journalières le matin vers le Luxembourg et le soir vers la France.

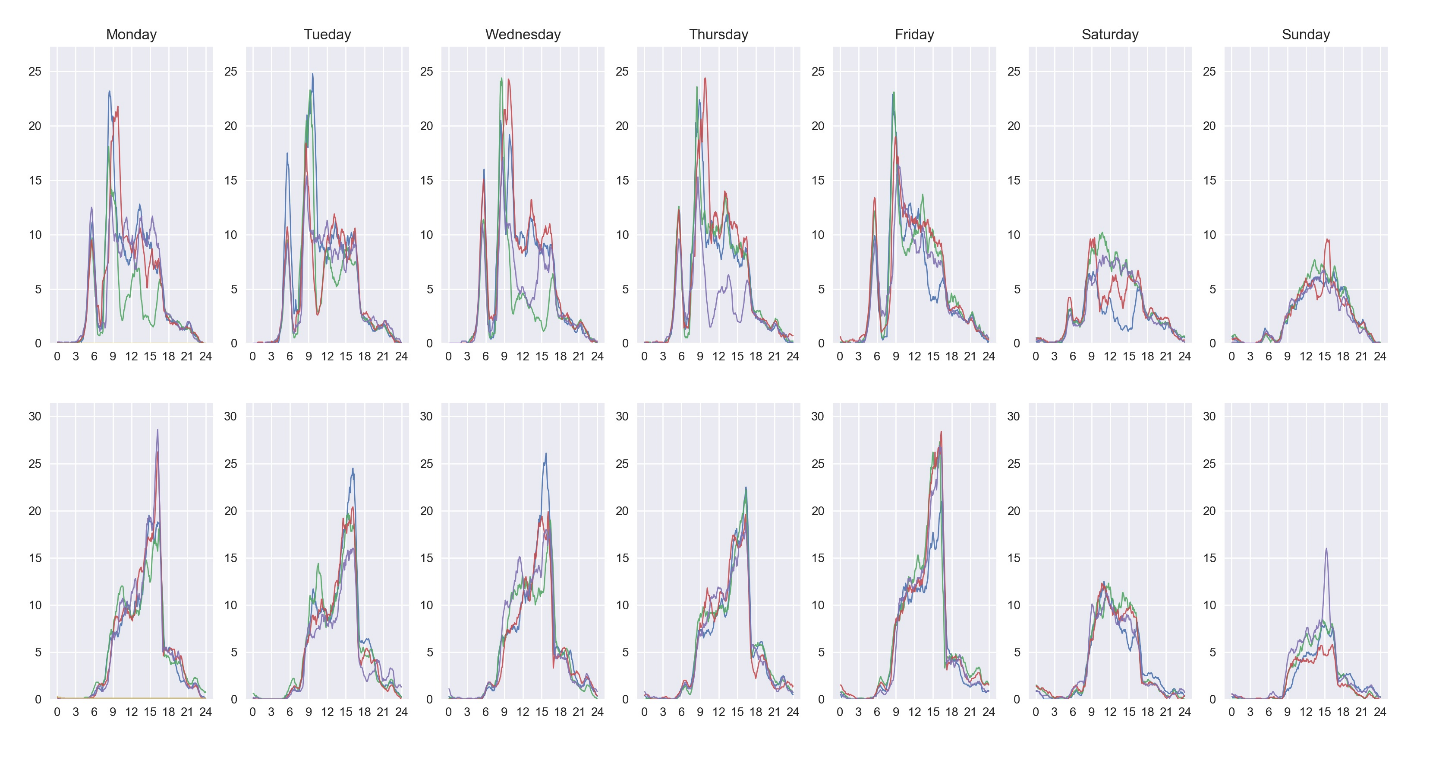
Vitesse moyenne en km/h sur 4 semaines (2019-11-25 au 2019-12-23) par jour, d’abord vers le Luxembourg, puis vers la France



Nombre de voitures par heure sur 4 semaines (2019-11-25 au 2019-12-23) par jour, d’abord vers le Luxembourg, puis vers la France



Concentration du trafic sur 4 semaines (2019-11-25 au 2019-12-23) par jour, d’abord vers le Luxembourg, puis vers la France



## Prédiction d’indicateurs à l’aide d’un algorithme de classification

Cette première analyse se concentre sur la prédiction d’indicateurs. Ces trois indicateurs simples (Oui/Non) sont les suivants :

* Baisse de la vitesse de 20% (par rapport à maintenant) dans 15 minutes.
* Baisse de la vitesse de 60% (par rapport à la vitesse moyenne globale) dans 15 minutes.
* Baisse de la vitesse et du flow de 60% (par rapport à la moyenne globale) dans 15 minutes.

Ces trois indicateurs tentent de qualifier respectivement un ralentissement, un potentiel bouchon, plus surement un bouchon.

Pour entrainer le modèle, les données suivantes ont été produites à partir des observations.

A un instant t, pour une caméra,

* La vitesse moyenne en km/h
* La concentration du trafic en pourcentage
* Le flux du trafic en nombre de voitures par heure
* La vitesse moyenne pour la caméra à 1km en amont
* La concentration du trafic pour la caméra à 1km en amont
* Le flux du trafic pour la caméra à 1km en amont
* Numéro de jour de la semaine
* Weekend : Oui/Non
* Heure (01, 02, …,14, 15, …, 24)
* Evolution de la vitesse en %, entre t et (t - 5 minutes), entre (t - 5 minutes) et (t - 10 minutes) et entre (t - 10 minutes) et (t - 15 minutes)
* Evolution du flux en %, entre t et (t - 5 minutes), entre (t - 5 minutes) et (t - 10 minutes) et entre (t - 10 minutes) et (t - 15 minutes)
* Evolution de la concentration en %, entre t et (t - 5 minutes), entre (t - 5 minutes) et (t - 10 minutes) et entre (t - 10 minutes) et (t - 15 minutes)
* Moyenne pondérée exponentielle de la vitesse sur les 15 dernières minutes
* Moyenne pondérée exponentielle de la concentration sur les 15 dernières minutes
* Moyenne pondérée exponentielle du flux sur les 15 dernières minutes

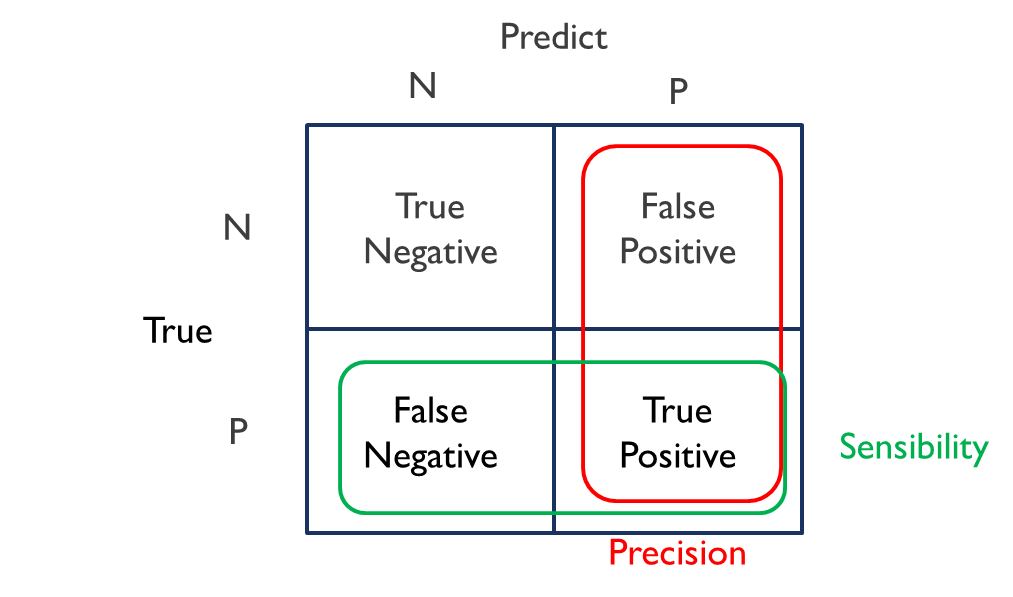
Les tests préliminaires de prédiction ont apporté les deux constats suivants :

* Entrainer 1 seul modèle pour toutes les caméras en même temps ne donnent pas de bon résultat. Les dynamiques de chaque point de l’autoroute sont différentes.
* Dans les variables d’une observation, si on ne prend que les données de la caméra, les prédictions à 15 minutes sont peu fiables. C’est pourquoi, les données de la caméra en amont ont été ajoutées.

L’algorithme choisi est le « Random Forest », un algorithme de classification, basés sur l’agrégation d’arbre de décisions.

Les métriques utilisées pour mesurer la fiabilité des prédictions sont les suivantes :

* **Score global de prédiction**Accuracy = Probability (correct test) = (TP+TN)/(TP+FP+TN+FN)
* **Pourcentage de prédiction correcte parmi les prédiction positives**Precision= Probability (positive | positive test) = TP / (TP+FP)
* **Pourcentage de prédiction correcte parmi les observations négatives**Specificity = Probability (negative test | negative) = TN / (FP+TN)



Les données ont été divisées en deux jeux :

* 80% des données sont utilisées pour entrainer le modèle (train data)
* 20% des données sont utilisées pour mesurer la fiabilité des prédictions (test data)

Résultat la caméra A3.MV.10437 (A3 au kilomètre 10) en direction du Luxembourg.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Baisse relative de la vitesse de 20% | Baisse de la vitesse de 60% par rapport à la vitesse moyenne globale | Baisse de la vitesse et du flow de 60% par rapport à la vitesse moyenne globale |
| Les 5 variables les plus discriminantes | ['vehicleFlowRate', 'prev\_station\_vehicleFlowRate', 'avgVehicleSpeed', 'EWMtraffic', 'EWMavg'] | ['avgVehicleSpeed', 'EWMavg', 'prev\_station\_avgVehicleSpeed', 'EWMtraffic', 'prev\_station\_vehicleFlowRate'] | ['prev\_station\_vehicleFlowRate', 'vehicleFlowRate', 'EWMflow', 'EWMavg', 'EWMtraffic'] |
| Matrice de confusion | [[1501 **8**]  [ **35** 14]] | [[1394 13]  [ 12 138]] | [[1532 1]  [ **10**  14]] |
| Accuracy | 0.972 | 0.984 | 0.993 |
| Precision | 0.636 | 0.914 | 0.933 |
| Sensitivity | 0.286 | 0.92 | 0.583 |

# Analyse des données de la caméra telraam

La question est de trouver quelles connaissances peut-on extraire de ces données de comptage afin d’offrir le l’information pertinente à des usagers. Ces usagers peuvent être les piétons, les conducteurs mais aussi les acteurs des communautés de communes, la mairie, etc..

## Description des données

Les données analysées ont été captées entre le 16/06/2020 et le 07/08/2020.

Chaque heure la caméra relève les données suivantes :

* Pourcentage d'activité de la caméra
* Comptage des piétons, voitures, vélos, camions (total, gauche et droite de la rue)
* Histogramme des vitesses des voitures pour les intervalles [0-10[  [10-20[ [20- 30[     .....  [70 et plus [

Exemple :

"segment\_id": "9000000411",

"date": "2020-07-23T05:00:00.000Z",

"pct\_up": 0.584444444444444,

"timezone": "Europe/Paris",

"pedestrian": 27.3764258555133,

"bike": 10.2661596958175,

"car": 366.15969581749,

"lorry": 11.9771863117871,

"pedestrian\_lft": 15.3992395437262,

"bike\_lft": 0,

"car\_lft": 22.2433460076046,

"lorry\_lft": 0,

"pedestrian\_rgt": 11.9771863117871,

"bike\_rgt": 10.2661596958175,

"car\_rgt": 343.916349809885,

"lorry\_rgt": 11.9771863117871,

"car\_speed\_histogram": [10.2661596958175,27.3764258555133,107.794676806084,164.25855513308,34.2205323193916,11.9771863117871,5.13307984790875,5.13307984790875],

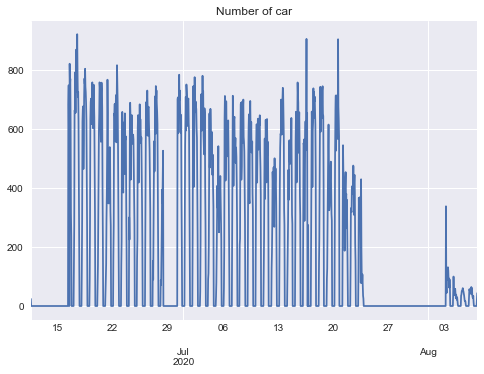
"car\_speed\_bucket": [0,1,2,3,4,5,6,7]

Pour une année, cela représente au plus 8760 enregistrements

Cette caméra a des contraintes, car pour réaliser les comptages, elle fait de la reconnaissance d’image :

* La caméra n’est pas active la nuit
* La caméra n’est pas active 100% du temps en plein jour c’est pourquoi, le pourcentage d’activité est indiqué. Lorsque la caméra est partiellement active, les comptages sont proratisés à l’aide de ce pourcentage pour estimer l’activité dans l’heure observée.
* La caméra peut être hors de service pendant une période

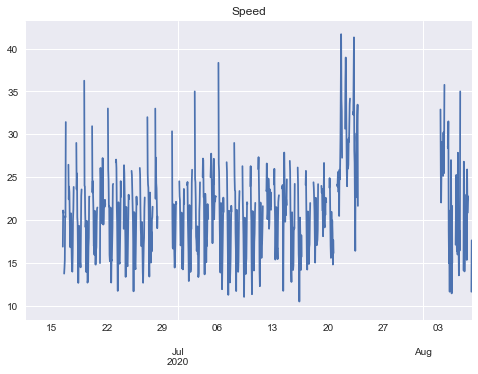
La capture des données n’est donc pas en continue comme on peut le voir sur le graphique de comptage des voitures sur cette période.



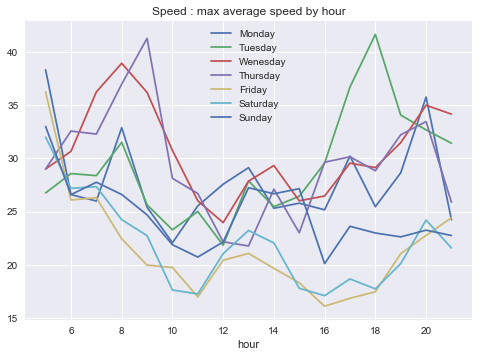
## Exploration graphique des données

A partir de l’histogramme des vitesses des voitures, on peut estimer une vitesse moyenne.

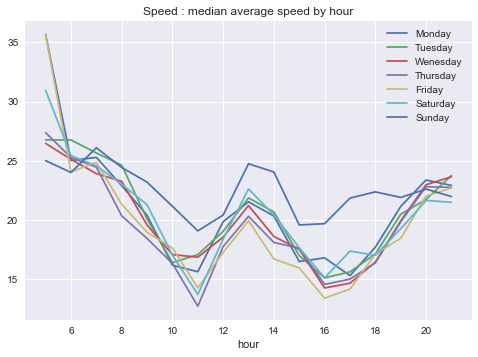
#### Graphique : évolution de la vitesse moyenne



#### Graphique : maximum de la vitesse moyenne journalière

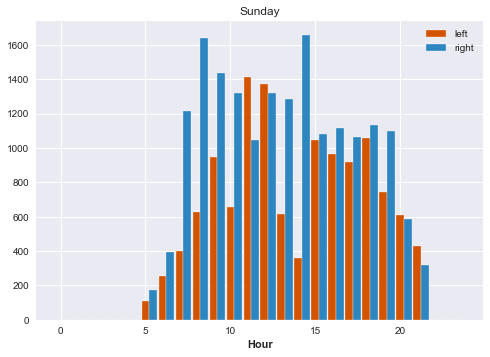
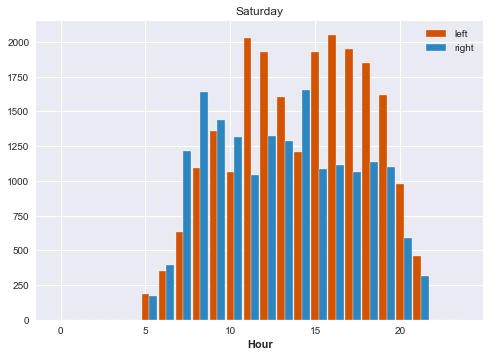
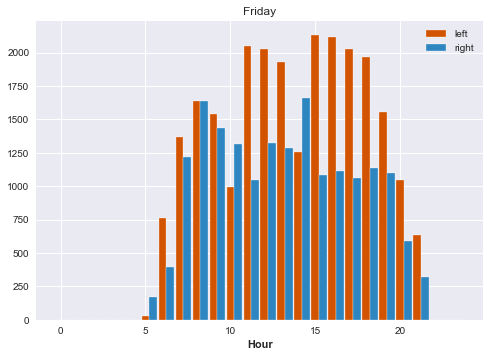
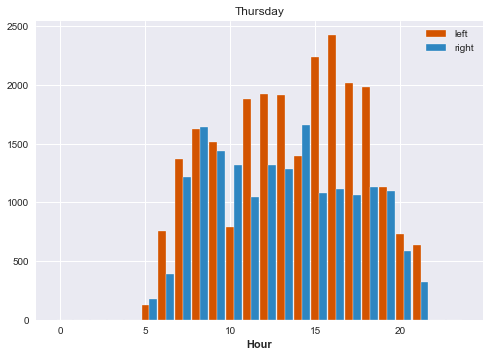
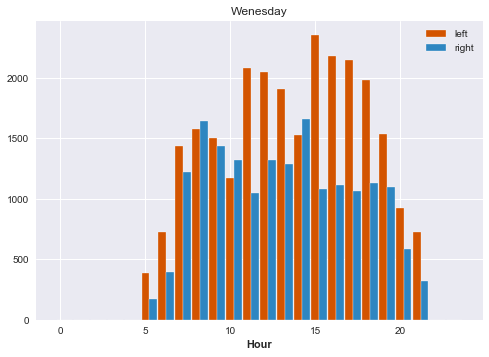
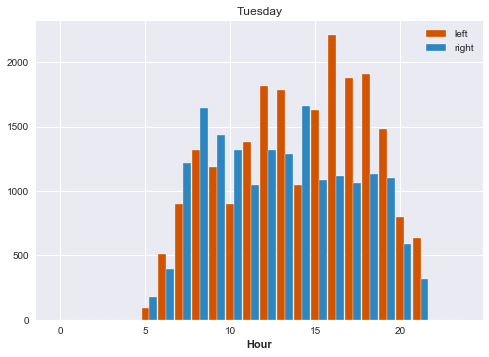
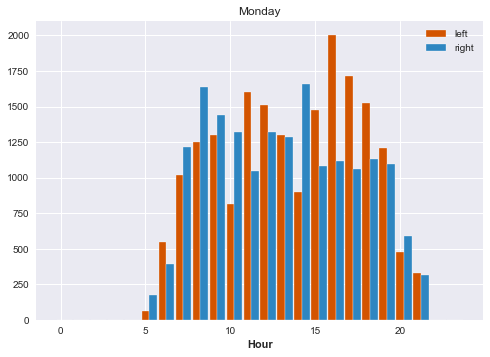


#### Graphique : médiane de la vitesse moyenne journalière



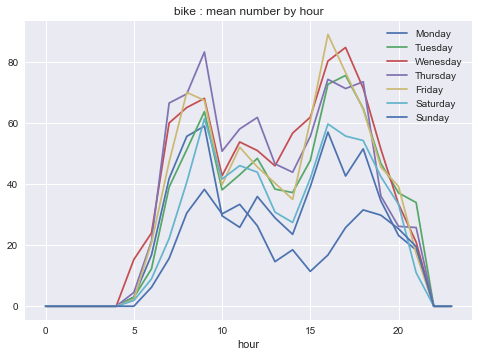
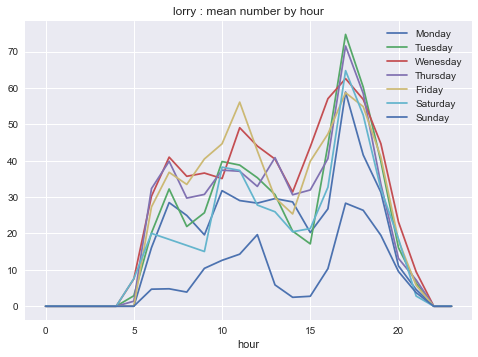
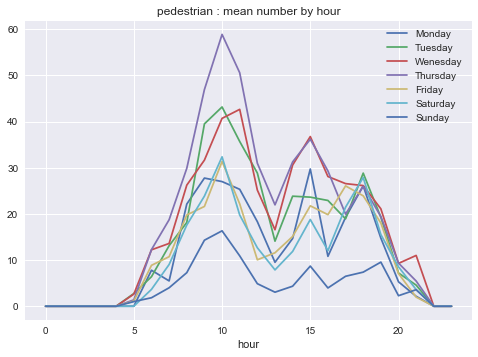
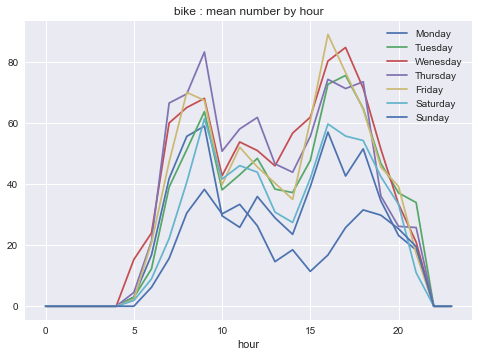
On peut aussi s’intéresser dans quel sens et à quel moment de la journée la rue est-elle plus passante, en comparant le comptage des voitures à droite et à gauche.

#### Graphiques : somme totale des voitures par heure et par jour



On peut aussi établir des tendances journalières moyennes

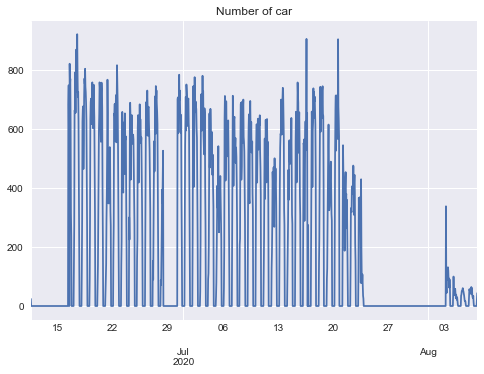
#### Graphique : comptages moyens par heure et par jour :



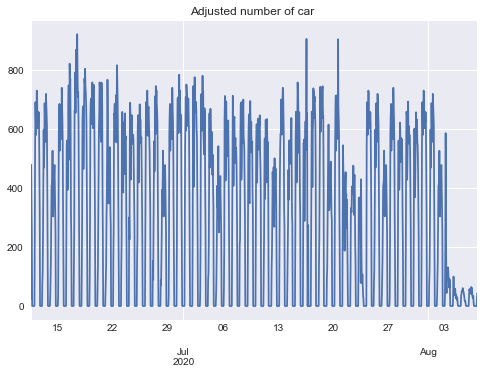
## Gestion des valeurs manquantes

Pour utiliser des modèles sur ces séries de comptage, les valeurs manquantes vont être remplacées par la valeur moyenne du jour de la semaine et de l’heure.

#### Graphique : série originale



#### Graphique : série ajustée



## Prédiction des flux futures

Aux vues des observations de comptage il existe un pattern des données journalières.

L’idée est de générer à partir d’une série de comptage, par exemple les voitures, deux journées en avance pour voir si on peut capter cette structure.

Sur le graphique ci-dessous on voit que les données sont hiératiques à partir du 23/07. Il n’y a pas assez de données pour savoir si c’est normal, l’activité au mois d’aout serait très faible, ou s’il y a une anomalie de captation. Pour trancher ce point, il faudrait les données sur au moins une année.

Ces comptages de trafic générés par une caméra équipée de reconnaissance d’image peuvent en effet générer des tendances journalières et des prédictions de flux.

Néanmoins les résultats devraient être beaucoup plus fiable si l’historique des données étaient suffisant pour observer toutes les saisonnalités dans ces séries, plus de 1 année.

L’algorithme LTSM qui prédit une ou plusieurs valeurs futures à partir d’une fenêtre dans le passé serait plus fiable.

Et à l’aide de SARIMA or EXP\_SMOO, plus de données dans le futur pourrait être générées pour capter les tendances.

Enfin avec plus de données, nous pourrions créer des sous séries qui ont des motifs journaliers très semblable pour faire des modèles prédictifs ciblées : les lundis ouvrés, les dimanches et jours fériés etc. …

The question is to find what knowledge can be extracted from this data in order to provide the relevant information to users.

# Compte kaggle pour la FabMob

User : [lafabriquedesmobilites@gmail.com](mailto:lafabriquedesmobilites@gmail.com)

Pwd :

# Description de l’infrastructure

Cameras

telraam / cita

Extraction

Stockage

NOTES

Exemple

"segment\_id": "9000000411",

 "date": "2020-07-01T22:00:00.000Z",

 "pct\_up": 0.448981481481482,

"timezone": "Europe/Paris",

"pedestrian": 604.391643145026,

"bike": 1553.10258573979,

"car": 9226.5770122573,

"lorry": 1063.81217958995,

"pedestrian\_lft": 284.86737748684,

 "bike\_lft": 894.621476307705,

"car\_lft": 5193.30653377209,

"lorry\_lft": 364.777851777449,

"pedestrian\_rgt": 319.524265658186,

 "bike\_rgt": 658.481109432086,

 "car\_rgt": 4033.27047848521,

"lorry\_rgt": 699.034327812504,

"car\_speed\_histogram": [2289.38747172499,3104.1790056431,2722.70916279142,725.617402833619,143.54595136182,65.1015948034983,58.1040594288343,117.932363670011],

            "car\_speed\_bucket": [0,1,2,3,4,5,6,7]

Livrables attendus ( repris de la fiche commun)  
Connecteur Spring : Connecteur batch générique Open Source en java à l'aide du framework Spring permettant de récolter les données des caméras

Base de données « Time Series » : Intégration intégration et alimentation d'une base de données « Time Series »

Outil de visualisation : oOutil de visualisation de données  
Test IA de prédiction à l’aide d’outil de « machine learning »  
Rapport d'analyse sur les résultats de l'algorithme prédictif et description de l’infrastructures  
nécessaires au bon fonctionnement de l'IA à son cycle de vie.

Yann :

Les technologies choisies répondent à ces critères :

+ Open Source

+ Très bien documentées

+ Très utilisées et la connaissance est très répandue chez les développeurs

**Solution pour la captation pour les flux**[**cita.lu**](http://cita.lu/)**et telraam**

Batch d'alimentation : java à l'aide du framework Spring (<https://spring.io/>)

DB « Time Series »  : timescale de Postgresql <https://www.timescale.com/>

Visualisation : Graphana <https://grafana.com/grafana/>

(timescaledb et graphana tournent dans un container Docker)

Livrables attendus

+ Code sources dans un repo Git  (java + sql)

+ Documentation de mise en oeuvre sur un poste local

+ Deux exemples de dashboard Graphana : 1 cita + 1 telraam

**Analyse / prédiction**

Outils : python à l'aide des librairies <https://scikit-learn.org/stable/> <https://www.tensorflow.org/> et <https://keras.io/>

Code sources dans un repo Git

Documentation de mise en oeuvre

Rapport d'analyse pour la mise en oeuvre d'un algorithme prédictif :

+ Quelles sont les données nécessaires ?

+ Quelles prédictions ( flow, vitesse ) , quel horizon de temps?

+ Décrire les besoins pour mettre cet algorithme en production