

## Challenge 2021 : Prediction Part

L'objectif était de prédire le nombre de vélos qui passeront le 2 Avril entre minuit et 9h au totem de la place Albert Premier. Pour ce faire, j'ai récupéré les données fourni dans l'énoncé et y ait ajouté des données pluviométrique. Car les précipitations ne paraissent être un facteur déterminant dans l'utilisation du vélo chez les cyclistes. J'ai rencontré plusieurs difficultés pour obtenir ces données météorologiques. L'incendie chez l'hébergeur OVH Strasbourg a touché le site quand j'ai commencé mes travaux. J'ai donc opté pour le site [historique-meteo.net](https://historique-meteo.net) qui fourni au format csv les données météorologique depuis 2009. Cependant le site bloquait l'accès aux données via le script python. Pour contourner ce problème, une solution bancale mais efficace était d'importer les données dans une Google docs puis d'importer le Google doc dans un second temps.

Les données cyclistes décrivent chaque jour de manière très différente. Pour essayer de palier cette différence, j'ai choisi de ne garder pour chaque jour uniquement le dernier relevé de manière à limiter les disparités de relevé.

L'observation des données cyclistes journalières en fonction de la date permet de supposer qu'une régression linéaire ne sera pas très efficace au vue de la dispersion à partir de Novembre 2020. Le 2 Avril 2021 tombe un vendredi, on va observer uniquement les données des vendredis pour voir si on a une linéarité un peu plus criante qu'en prenant tout les jours de la semaine. On observe depuis Novembre 2020 une linéarité plus marquée qu'avant. Une valeur semble « loin » de cette supposé linéarité, en observant ce jour en particulier, le 19/03/2021, on remarque que la seule observation ce jour là, et donc la dernière, a été effectué à 8h26. Un relevé aussi tôt dans la journée ne peut évidemment pas refléter le trafic total de la journée.

En faisant tourner la méthode OLS sur les données postérieur à Octobre 2020, on trouve que la pluie est un facteur augmentant le nombre de vélo, ce qui est contre intuitif. On va réduire notre étude aux données depuis le premier janvier 2021. Dans ce nouvel échantillon, on retrouve aussi que la pluie est un facteur augmentant le nombre de cyclistes. Face à ce constat, on observe le nombre de jour de pluie dans les données depuis Novembre et depuis Janvier. On retrouve seulement 33 % de vendredi de pluie depuis novembre et 42 % depuis Janvier. Les valeurs des précipitations sont faibles,

depuis Mars 2020 seuls deux vendredi ont connus des précipitations de plus de 1cm. Ce qui nous amène à penser que les écarts de précipitations ne sont pas assez grands pour qu'on puissent les utiliser pour prédire le nombre de cyclistes.

On se rabat donc sur une méthode d'interpolation sans variable explicative mais toujours en ne prenant en compte que les vendredis. On récupère nos données depuis Novembre et Janvier et on effectue une interpolation d'ordre 2 sur le jeu de données démarrant en Janvier. L'interpolation sera d'ordre 3 sur les données datant de Novembre, cette différence d'ordre est due à une différence visuelle entre les allures des deux nuages de points. Ce n'est pas un problème d'avoir un ordre différent car cela permet d'avoir indépendamment des résultats plus précis sur chaque nuage. Les estimations par cette méthode ont environ 300 cyclistes d'écart. La méthode choisie pour récupérer une unique valeur journalière pour le vendredi 2 Avril 2021 est de faire une moyenne simple de ces deux valeurs.

Reste cependant que la question initiale porte sur la tranche horaire 0h-9h. J'ai arbitrairement accordé un point à chaque tranche horaire de la journée en fonction du passage pour pouvoir faire un ratio et déterminer le coefficient par lequel multiplié ma valeur journalière. Ainsi les horaires de bureaux ainsi que de la pause méridienne représente plus de cyclistes que les créneaux de début et fin de journée. La période 22h-5h est considérés comme n'ayant aucun poids. Le coefficient multiplicateur de la valeur journalière est 0.21.

Coef	1/4	2	2	1	1	2	2	1	1	1	1	2	2	1	1/2	1/2	0
Heure	6h	7h	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h-5h

La valeur finale de ma prédiction est 292 cyclistes entre minuit et 9h le 2 Avril 2021.

Vous trouverez le code dans mon dépôt Git : [accéder au dépôt Git](#)