

Analyse des données publiées par Uber Movement :

Le travail sur Uber Movement a été réalisé grâce aux données disponibles sur <https://movement.uber.com/>, celui-ci proposant, pour une liste de grandes villes, d'accéder à trois types de cartes (notons que chaque type de carte n'est pas disponible pour toutes les villes).

Ces cartes sont :

- Travel Speed : Représentant des mesures de temps de trajet des véhicules Uber entre deux zones.
- Speeds : Représentant la vitesse moyenne de ces mêmes véhicules en ville.
- Mobility Heatmap : Représentant le volume de fréquentation lié à ces véhicules en ville.

Ce sont les données de Travel Speed et de Mobility Heatmap qui ont été utilisées ici.

Travel Speed :

Les jeux de données utilisées pour travailler sur Travel Speed sont :

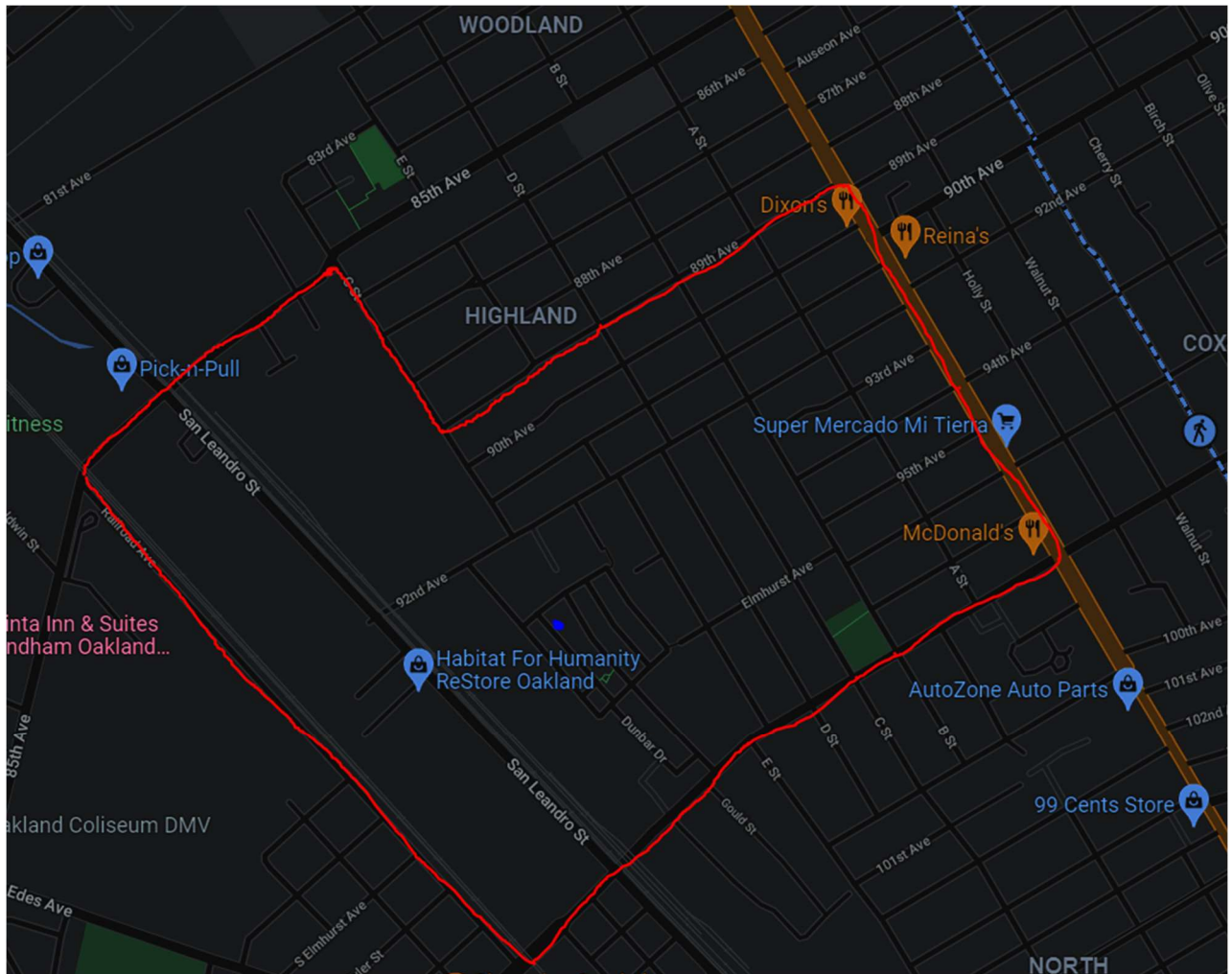
- Un fichier csv, traitant le premier semestre de 2020, contenant les attributs suivants :
 - o sourceid : Un identifiant correspondant à une position de départ.
 - o dstid : Un identifiant correspondant à une position d'arrivée.
 - o month : Correspondant au mois du trajet (01 pour janvier, 02 pour février, 03 pour mars).
 - o mean_travel_time : Le temps de trajet moyen (en secondes) entre sourceid et dstid.
 - o standard_deviation_travel_time : L'écart-type des temps de trajet entre sourceid et dstid.
 - o geometric_mean_travel_time : La moyenne géométrique des temps de trajet entre sourceid et dstid.
 - o geometric_standard_deviation_travel_time : L'écart-type géométrique des temps de trajet entre sourceid et dstid.
- Un fichier json permettant d'associer les identifiants du fichier csv à des polygones via un ensemble de positions GPS, ainsi qu'à un nom de rue et de ville.

Exemple avec la première ligne du fichier csv concernant la ville de San Francisco :

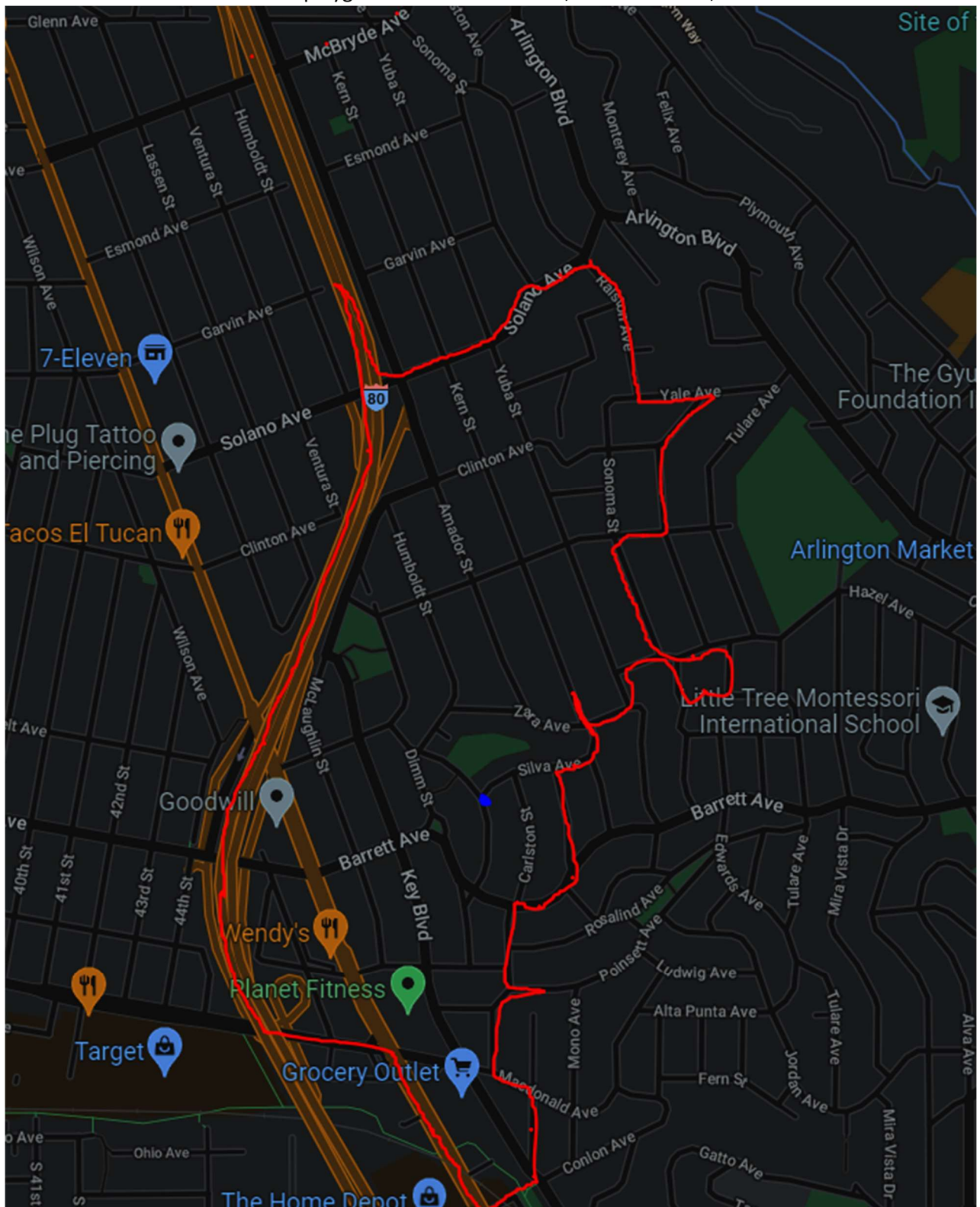
2220,102,02,1728.41,410.0,1683.83,1.25

Pour la sourceid 2220, le fichier json nous renvoie à ce polygone :

9400 Dunbar Dr, Oakland, CA 94603, États-Unis :

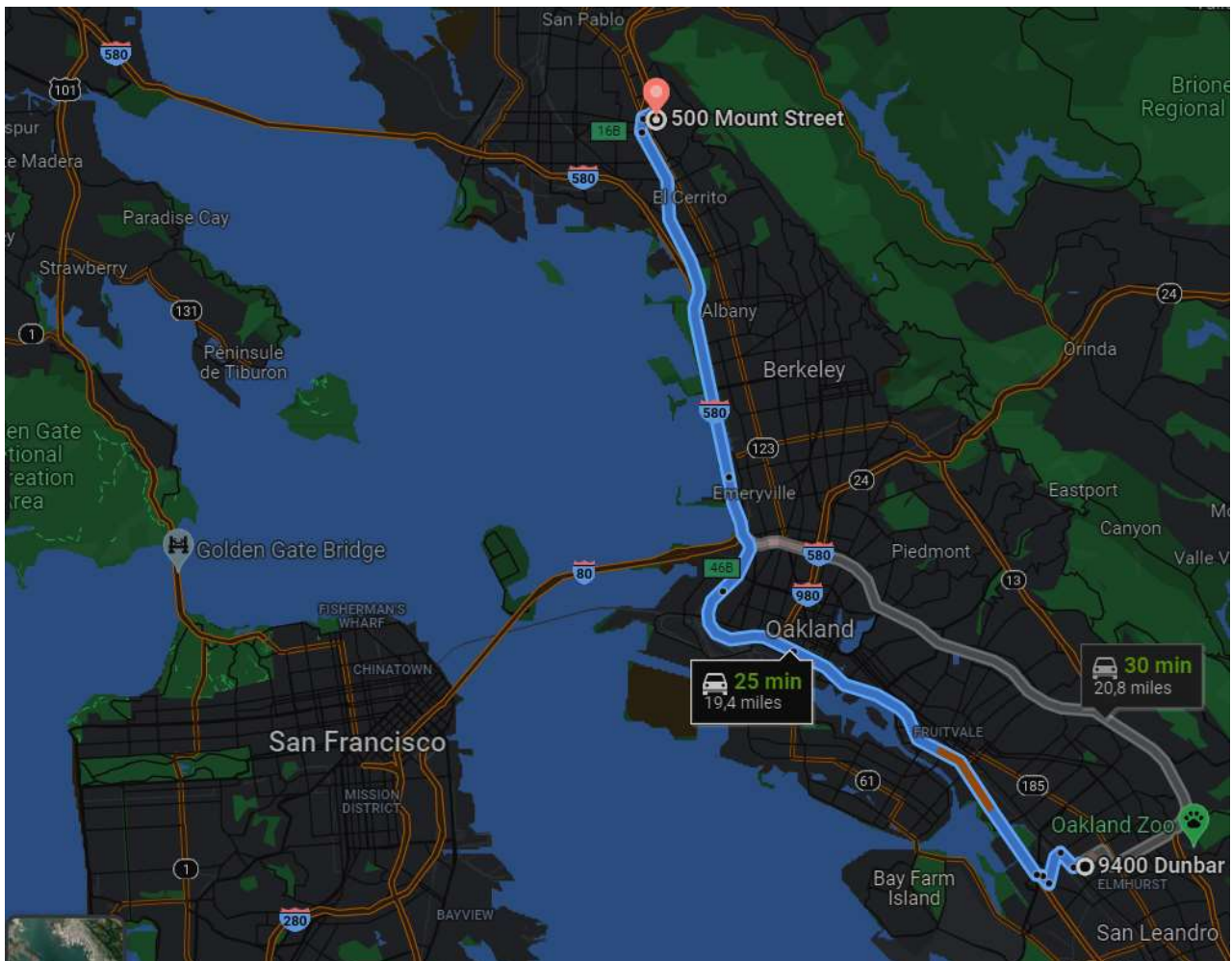


Et la dstid 102 nous renvoie à ce polygone : 500 Mount Street, East Richmond, Richmond



La moyenne de temps pour se déplacer entre ces deux points en février 2020 était donc de 1728,41s (environ 28,81 minutes).

En comparant cette donnée avec le temps estimé actuellement sur google maps :



On voit que cette mesure paraît plausible.

N'ayant pas accès aux données individuelles de chacun des trajets ayant servi pour réaliser ces moyennes, il est difficilement possible d'attaquer plus en détail ces données.

Mobility Heatmap :

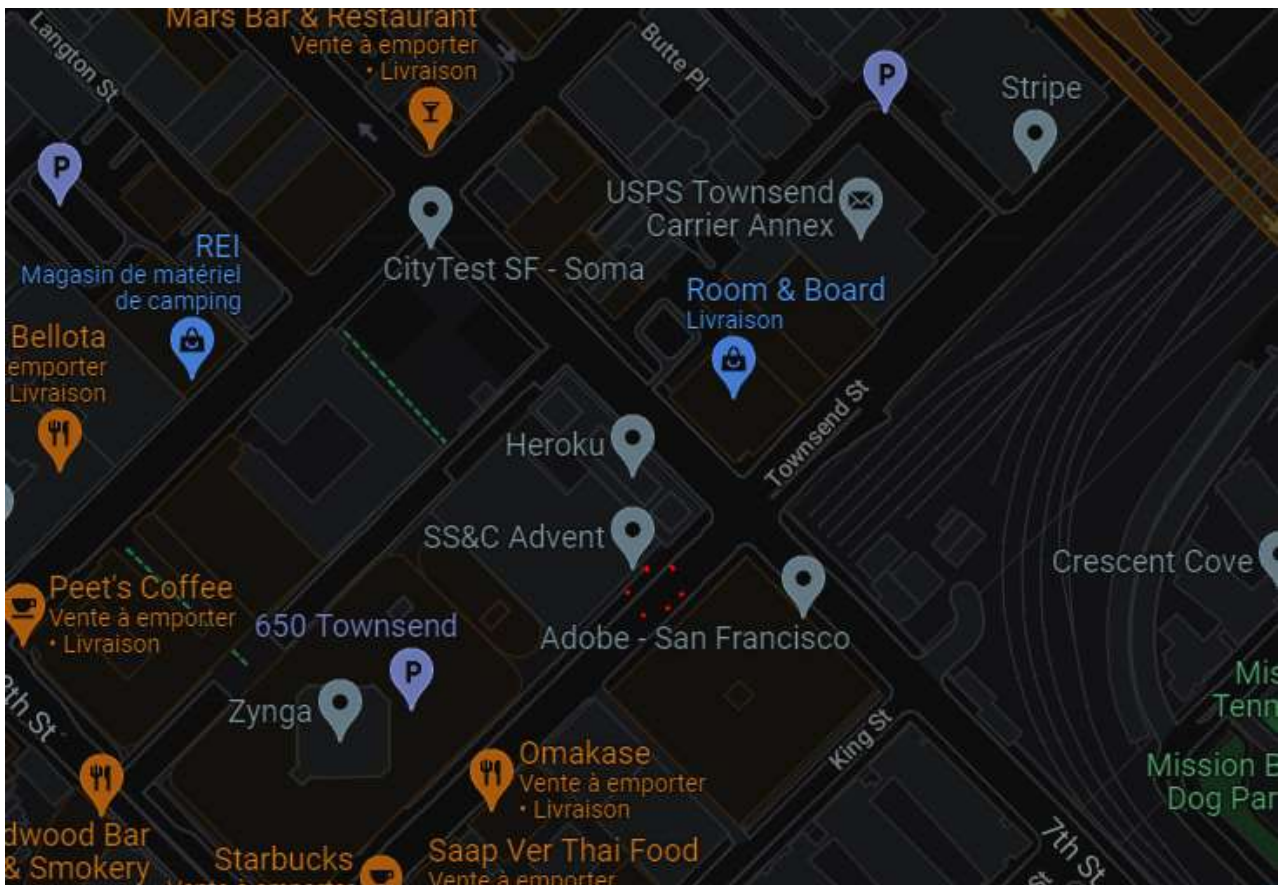
Le jeu de données disponible pour Mobility Heatmap est un fichier csv, traitant le premier semestre de 2020, contenant les attributs suivants :

- hexid : Un identifiant correspondant à une position.
- dayType : weekday ou weekend.
- traversals : Le nombre de traversées dans la zone sur le jour concerné.
- wktGeometry : 7 coordonnées GPS formant un hexagone (la première coordonnée étant la même que la dernière).

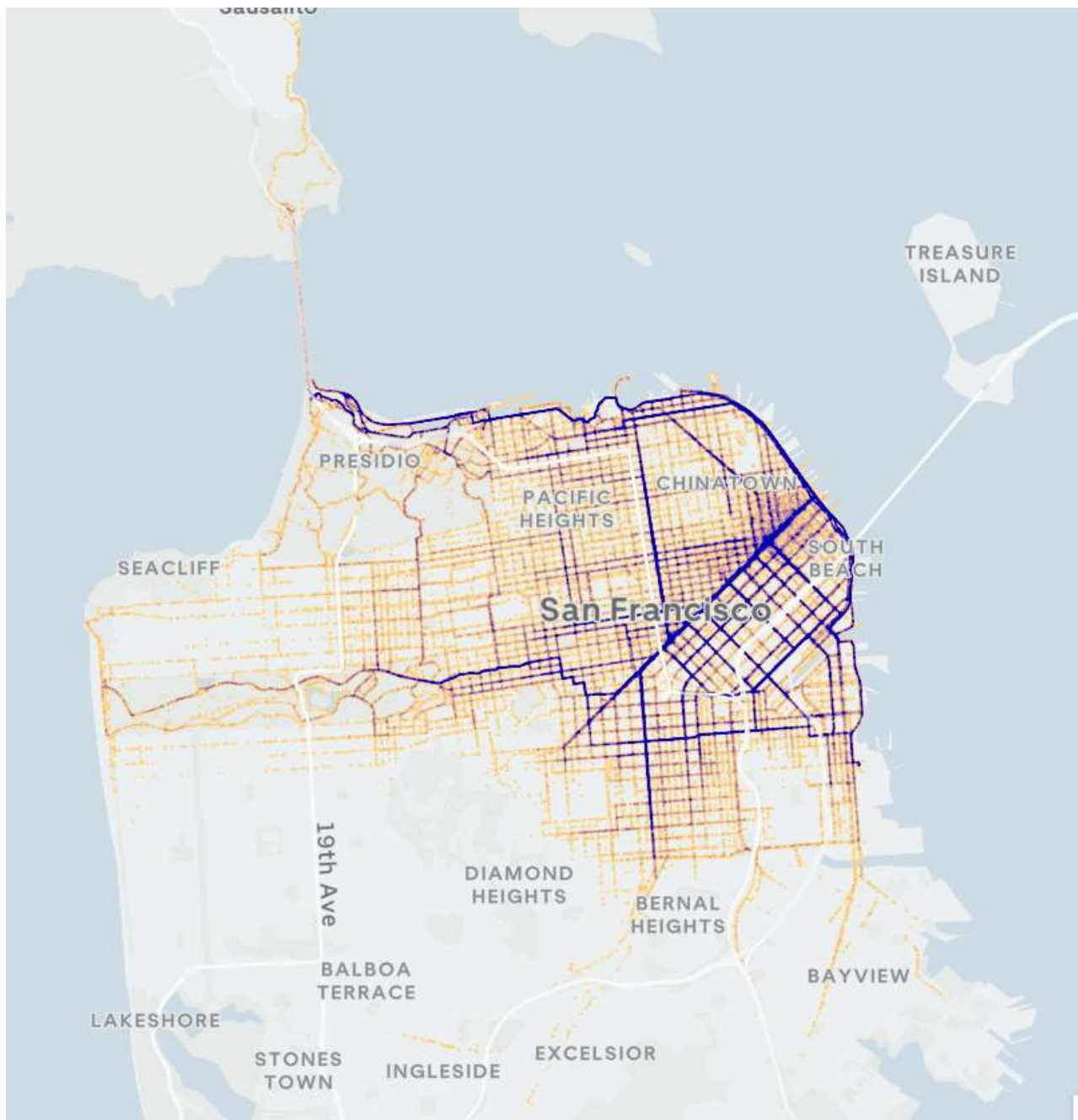
Exemple :

```
8c283082e2561ff,weekday,1032,"POLYGON ((-122.40206953427416 37.77145205659484, -  
122.40197821984188 37.77152436450601, -122.40200775554032 37.771620128151376, -  
122.40212860581218 37.771643583808135, -122.4022199201558 37.77157127584822, -  
122.4021903843163 37.771475512280276, -122.40206953427416 37.77145205659484)))"
```

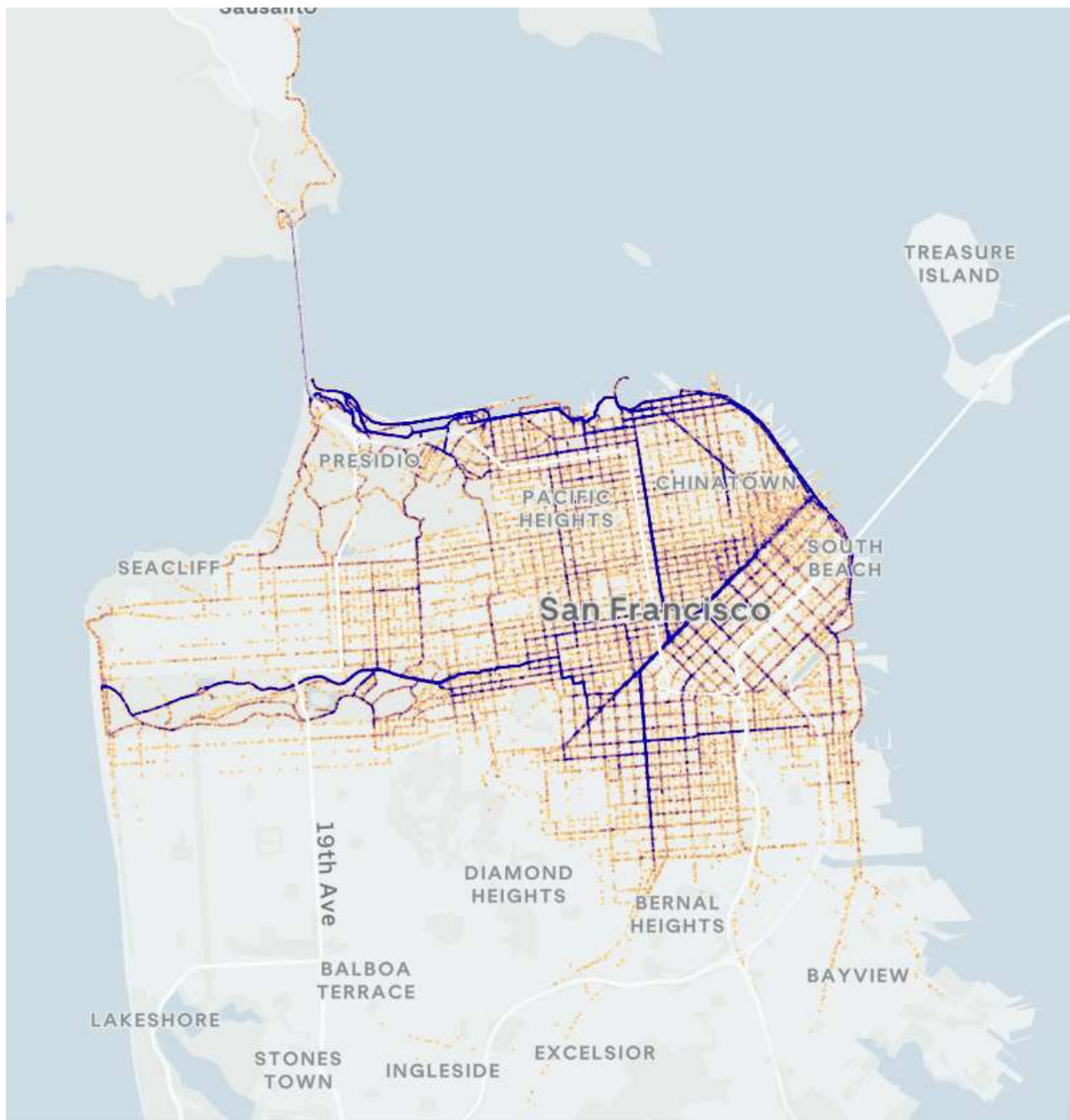
Dont le polygone correspond à ces points retranscrit sur une carte :



Exemple de visualisation de données sur 24h complètes sur la ville de San Francisco en semaine :



Puis en weekend :



Nous avons réalisé des statistiques sur ce jeu de données en utilisant la librairie pandas pour Python :

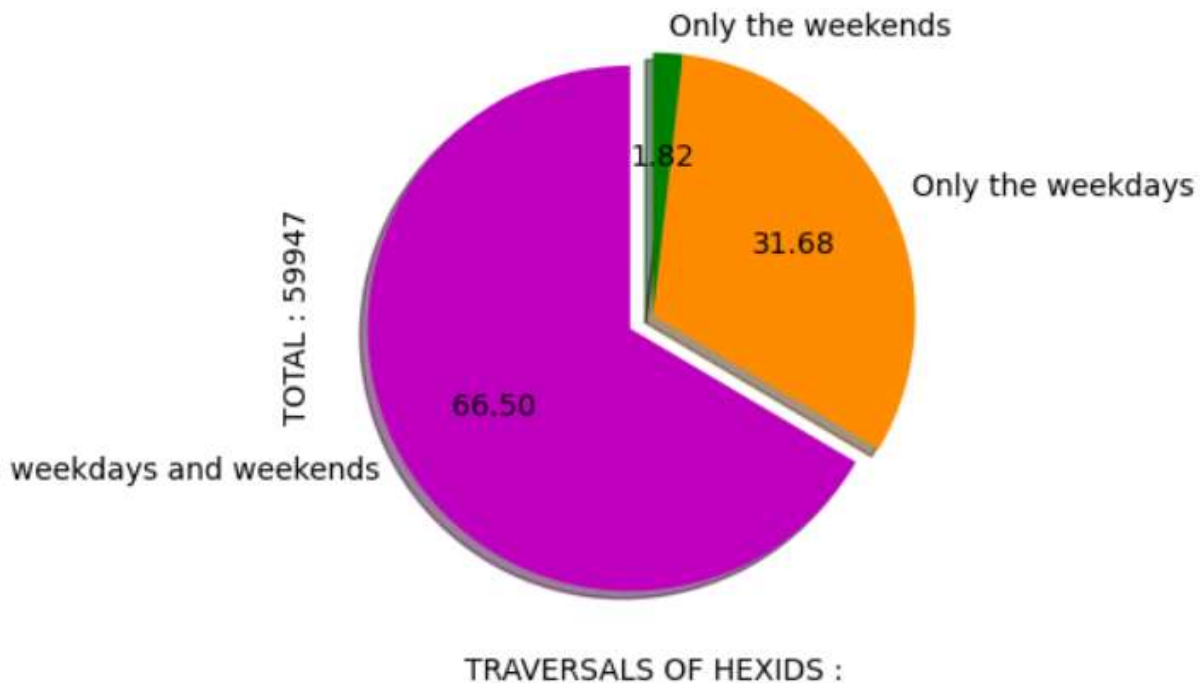
Différents ensembles de données issus des valeurs ont été traités :

- Un ensemble contenant toutes les valeurs : All
- Un ensemble ne contenant que les valeurs qui apparaissent en semaine : All-weekdays
- Un ensemble ne contenant que les valeurs qui apparaissent en semaine : All-weekends
- Un ensemble contenant les valeurs n'apparaissant qu'en semaine ou que le week-end : Xor-all
- Un ensemble contenant les valeurs n'apparaissant qu'en semaine et pas le week-end : Xor-weekdays
- Un ensemble contenant les valeurs n'apparaissant que le weekend et pas en semaine : Xor-weekends
- Un ensemble contenant les valeurs apparaissant en semaine ainsi qu'en weekend : And-all
- Un ensemble contenant les valeurs apparaissant en semaine parmi celles apparaissent à la fois en semaine et en weekend : And-weekdays
- Un ensemble contenant les valeurs apparaissant en weekend parmi celles apparaissent à la fois en semaine et en weekend : And-weekends

Des statistiques ont été calculées pour chacun de ces ensembles, par exemple pour All :

```
Values of traversals :  
traversals  
count    99813.000000  
mean      109.178614  
std       187.981887  
min       15.000000  
25%       26.000000  
50%       48.000000  
75%      106.000000  
max      3864.000000  
Median is 48.0
```


Et un graphe a été généré pour connaître la répartition des polygones selon les ensembles :

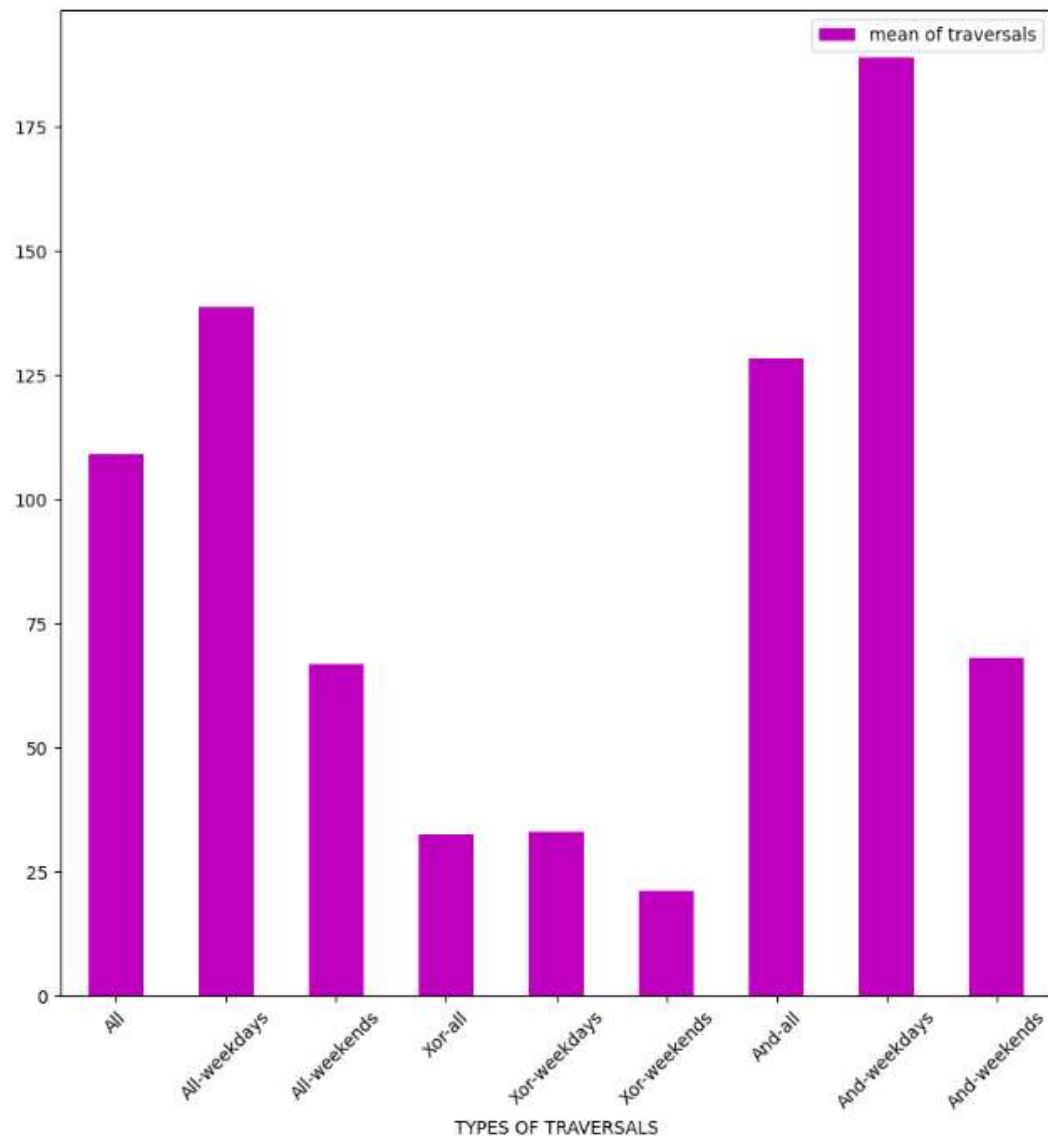


Le graphe entier représente donc All, la partie violette représente ici And-weekdays et And-weekends et représente donc également And-all dont la taille serait divisée par deux car chaque valeur de polygone est présente une seule fois et non deux pour le graphe.

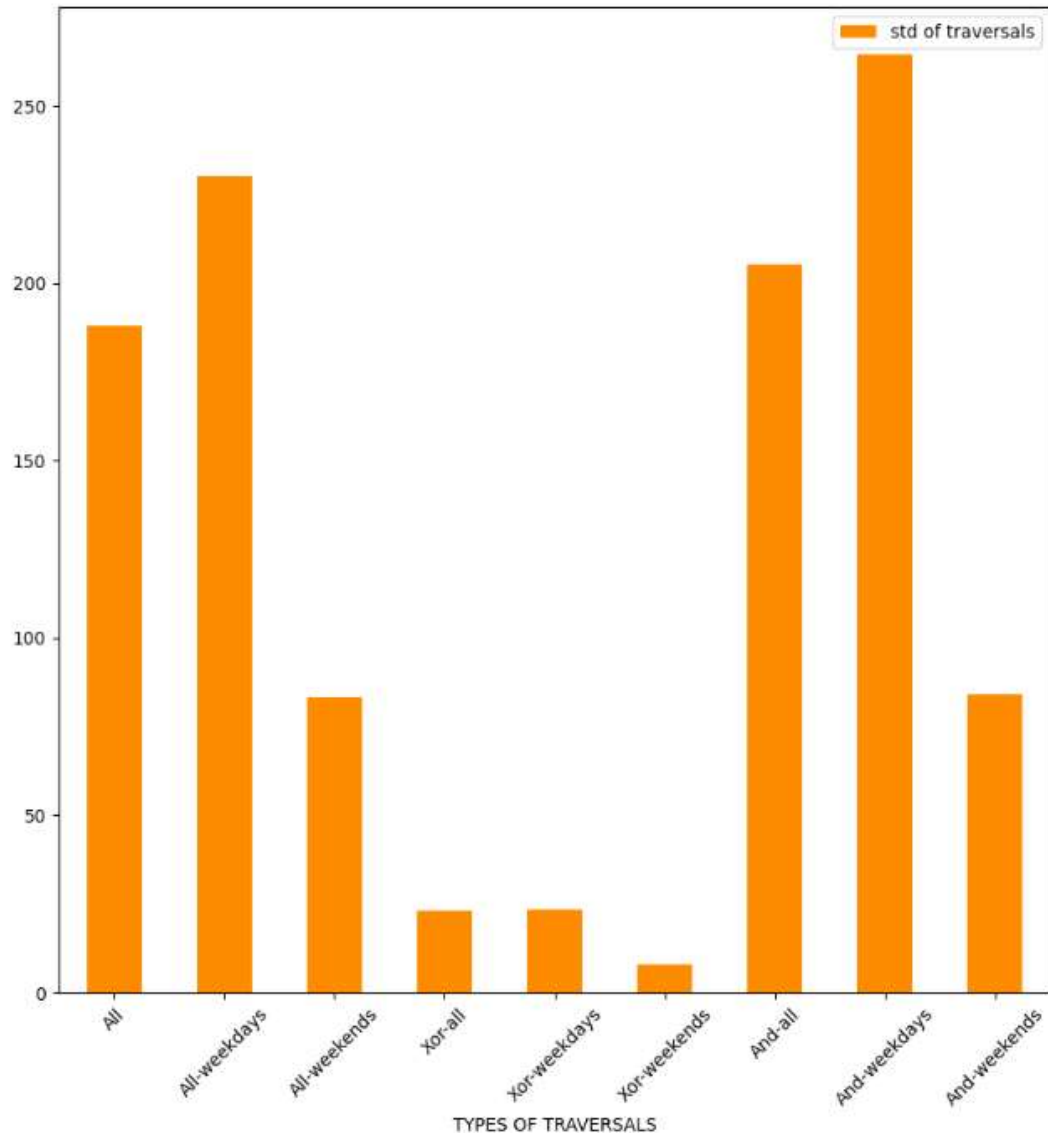
Xor-weekdays correspond à la partie orange tandis que Xor-weekends à la partie verte et Xor-all correspondant donc à la partie orange plus la partie verte.

All-weekdays correspond à la partie violette plus la partie orange et All-weekends correspond à la partie violette plus la partie verte.

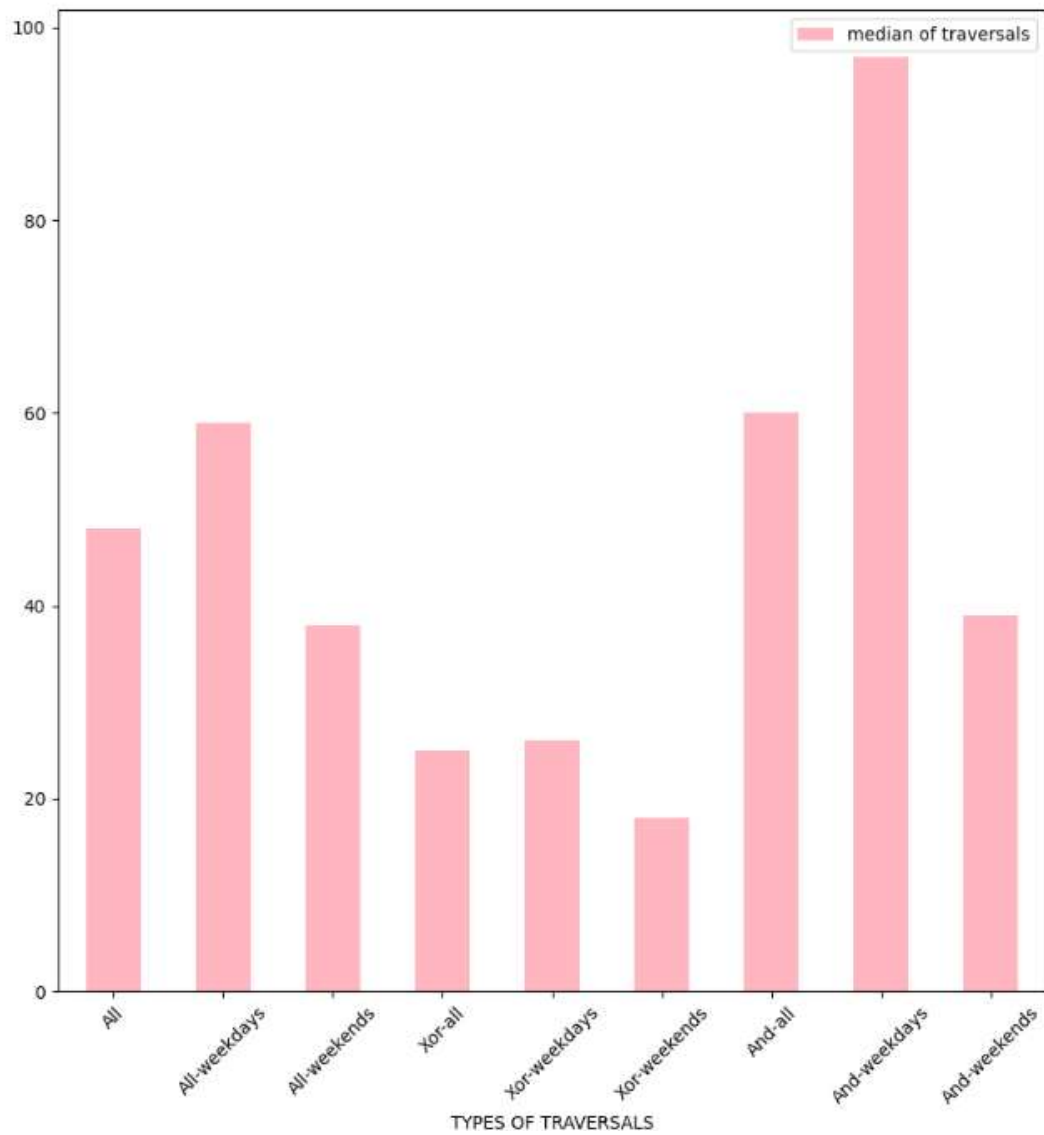
Différents autres graphes ont été réalisés sur le nombre de traversées des zones, notamment pour la moyenne :



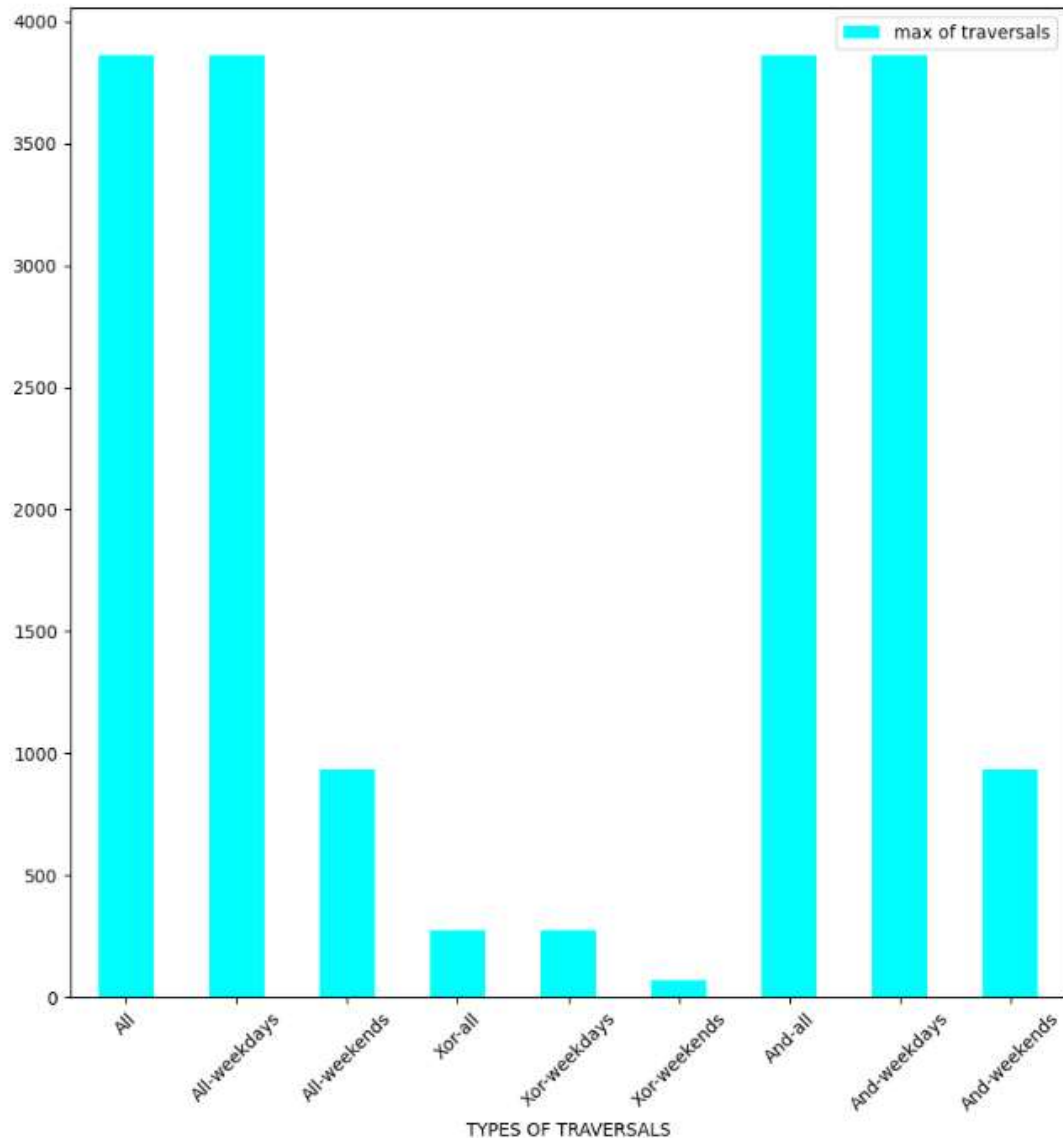
Pour l'écart-type :



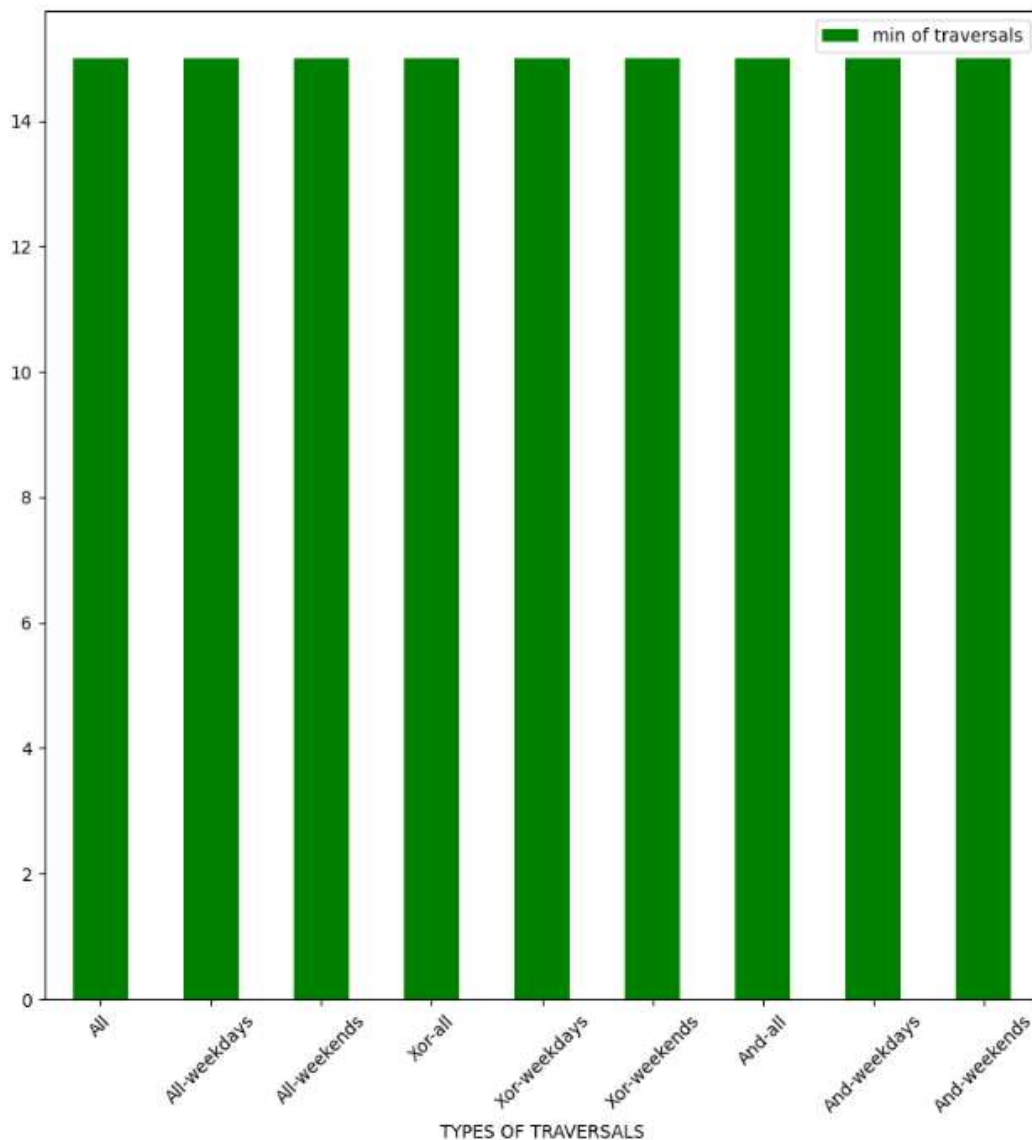
Pour la valeur médiane :



Pour la valeur maximale :



Ainsi que pour la valeur minimale :



Pour ce dernier graphe, nous remarquons que la valeur ne semble pas descendre en dessous de 15, on peut donc supposer que Uber Movement n'utilise pas les valeurs des zones avec moins de 15 traversées. Afin d'éviter toute coïncidence, le même traitement a été réalisé avec les données de la ville de Paris, et le même résultat est survenu, venant confirmer notre supposition.

Il était difficile d'attaquer plus en détail ces données car les traversées ne sont pas liées directement à des personnes dans les jeux de données et on ne peut pas suivre une de ses personnes pour connaître son itinéraire. Également, le jeu de données de nous permet pas d'avoir des données découpées plus finement qu'en jour ce qui limite également leur utilisation, bien que sur leur site nous puissions visualiser la carte sur une plage horaire choisie.