

IUT Grand Ouest Normandie

Bachelor Universitaire de Technologie
Science des Données
Campus de Lisieux

Science des Données 1 - SAE 1.03
Préparation et synthèse d'un tableau de données

Thématique

Analyse des locations de vélos Citibike



Auteurs

GAMONDELE Maxime
DELAMARE Logan

Année universitaire 2023-2024

Table des matières

Introduction	3
1 Caractéristique des locations	4
1.1 Répartitions des locations selon le type de vélos	4
1.2 La variable duration	4
1.3 La répartition du temps de location des vélos	5
1.4 Distribution de la variable duration	5
1.5 Distribution de la durée de location selon le type de vélo	6
2 Analyse des fréquences et des trajets	7
2.1 Répartition des locations selon les jours de la semaine	7
2.2 Répartition des locations selon les heures de la journée	7
2.3 Top 10 des trajets les plus fréquentés	8
2.4 Top 10 des trajets avec station de départ et d'arrivée différentes	8
Conclusion	9
3 Annexe	10
Table des figures	11

Introduction

La compagnie CitiBike met à disposition les données d'utilisation de ses vélos dans la ville de New York.

Historiques des voyages Citibike

Ce jeu de données contient les informations suivantes (cf. Fig. 1) :

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID
- End station name
- End station ID
- Start latitude
- Start longitude
- End latitude
- End longitude
- Member or casual ride

ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat
1 B0A0F1DEFA872FC	electric_bike	2023-09-03 10:20:41	2023-09-03 10:24:16	E 1 St & Bowery	5636.13	E 10 St & 2 Ave	5746.02	40.72486	-73.99213	40.72971
2 2B26AB15647BF4EE	classic_bike	2023-09-27 15:44:23	2023-09-27 15:53:25	Pearl St & Hanover Square	4993.02	Allen St & Rivington St	5414.06	40.70465	-74.00913	40.72020
3 9D2B5971CA4E513F	classic_bike	2023-09-19 13:40:48	2023-09-19 13:48:11	E 1 St & Bowery	5636.13	E 10 St & 2 Ave	5746.02	40.72475	-73.99212	40.72971
4 17E6760596DC3ABE	classic_bike	2023-09-30 16:27:50	2023-09-30 16:56:35	Central Ave & Himrod St	4713.01	Mott St & Prince St	5561.04	40.69671	-73.92293	40.72318
5 97EFF376A7E2DC70	classic_bike	2023-09-21 16:59:53	2023-09-21 17:07:36	St Marks Pl & 2 Ave	5669.10	Mott St & Prince St	5561.04	40.72842	-73.98714	40.72318
6 D14BD5627029F9FEA	classic_bike	2023-09-11 14:50:19	2023-09-11 14:54:47	Meserole Ave & Manhattan Ave	5666.04	Franklin St & Dupont St	5944.01	40.72694	-73.95300	40.73564
7 D4AF7211BE4B6C54	classic_bike	2023-09-06 17:19:49	2023-09-06 17:23:40	Pearl St & Hanover Square	4993.02	Fulton St & William St	5137.11	40.70472	-74.00926	40.70960
8 775E8F8E42CF867	classic_bike	2023-09-26 17:48:10	2023-09-26 18:04:14	Adams St & Prospect St	4821.10	Washington Ave & Greene Ave	4419.03	40.70116	-73.98874	40.68650
9 2CC428B00C45A968	classic_bike	2023-09-15 13:14:38	2023-09-15 14:18:07	West Broadway & Watts St	5569.07	Allen St & Rivington St	5414.06	40.72323	-74.00314	40.72020
10 7FEB0C3990129D56	classic_bike	2023-09-06 03:29:02	2023-09-06 03:41:42	E 117 St & 1 Ave	7579.11	Tinton Ave & E 165 St	7991.01	40.79656	-73.93445	40.82480
11 B54C7B22E828FD7A	electric_bike	2023-09-13 08:27:35	2023-09-13 08:30:54	Pearl St & Hanover Square	4993.02	Fulton St & William St	5137.11	40.70472	-74.00926	40.70960
12 1BA80F46022B4613	classic_bike	2023-09-13 07:03:45	2023-09-13 07:08:40	St Marks Pl & 2 Ave	5669.10	Mott St & Prince St	5561.04	40.72852	-73.98724	40.72318
13 07FF3D443D7DFA14	classic_bike	2023-09-30 23:08:30	2023-10-01 01:12:04	Broad St & Bridge St	4962.08	Mott St & Prince St	5561.04	40.70365	-74.01168	40.72318
14 E3280089B7C5E57F	classic_bike	2023-09-01 18:42:41	2023-09-01 18:59:38	Meserole Ave & Manhattan Ave	5666.04	Allen St & Rivington St	5414.06	40.72708	-73.95310	40.72020
15 A918CFF9F19BF77B	classic_bike	2023-09-30 14:42:10	2023-09-30 14:48:17	E 1 St & Bowery	5636.13	E 10 St & 2 Ave	5746.02	40.72486	-73.99213	40.72971

Fig. 1. Extrait des données concernant les locations de vélos

L'analyse de ces données se fera sur la ville de New York uniquement et sur la période de septembre 2023. Le jeu de données étudié pour cette analyse comporte 3 575 162 observations. Le rapport est divisé en deux parties : une première partie sera consacrée aux caractéristiques des locations tandis que la seconde partie sera consacrée à l'analyse des fréquences et des trajets.

1 Caractéristique des locations

Cette première partie abordera l'analyse des données concernant les caractéristiques des locations. Nous nous pencherons sur les différents type de vélos loués, la durée moyenne de location, ainsi que la corrélation entre la durée de location et le type des vélos.

1.1 Répartitions des locations selon le type de vélos

En septembre 2023, nous pouvons observer (cf. Fig. 2) que 92,88 % des vélos utilisés sont des vélos classiques contre 7,12 % pour les vélos électriques. Nous en concluons que les vélos les plus utilisés sont les vélos de type classiques. Le lot de vélos électriques disponible sur New York paraît moins important que le nombre de vélos classique, cela expliquerait l'écart important des locations entre les deux types de vélo. Il pourrait être intéressant d'effectuer une analyse sur l'usage des vélos électriques et traditionnels en prenant en compte le nombre respectif de chaque type de vélo mis à disposition. On pourrait alors comparer la densité de location des types de vélo.

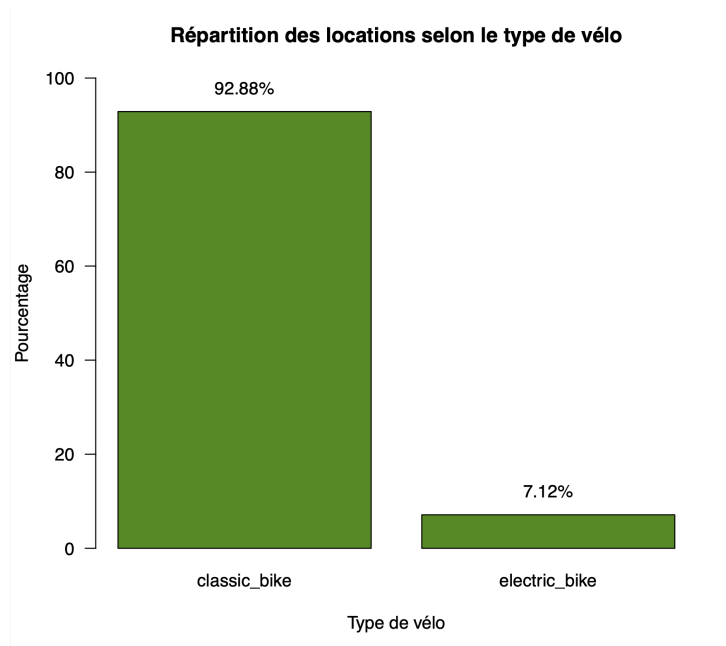


Fig. 2. Répartitions des types de vélos

1.2 La variable duration

Le résumé (cf. Fig. 3) nous permet de constater que le jeu de données contient des valeurs aberrantes. La valeur minimale est -549,25 cependant, il est impossible de louer un vélo pendant une durée négative. De ce fait, nous avons filtré le jeu de donnée pour avoir que des valeurs strictement positives, soit pour avoir des valeurs avec mouvement d'un vélo.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-549.25	5.22	9.42	15.92	16.72	1500.52

Fig. 3. Résumé de la variable duration

Ci dessous (cf. Fig. 4) la table nous montre la proportion de valeurs aberrantes qui est de 0,04%, c'est-à-dire 1469 valeurs aberrantes. Elles sont toutes des valeurs inférieur ou égale à zéro car on considère les valeurs négatives et égales à 0 inacceptable dans le cadre d'une étude sur les locations.

FALSE	TRUE
99.96	0.04

Fig. 4. Proportion des valeurs aberrantes

1.3 La répartition du temps de location des vélos

On observe, comme illustré dans la Figure ci-dessous (cf. Fig. 5) que la majorité des utilisateurs de vélos Citibike optent pour des trajets d'une durée allant de 0 à 20 minutes, avec une concentration particulière entre 5 et 10 minutes. Cependant, on constate également des utilisations s'étendant de 20 à 60 minutes. Au-delà de 60 minutes, le nombre de locations de vélos devient très faible.

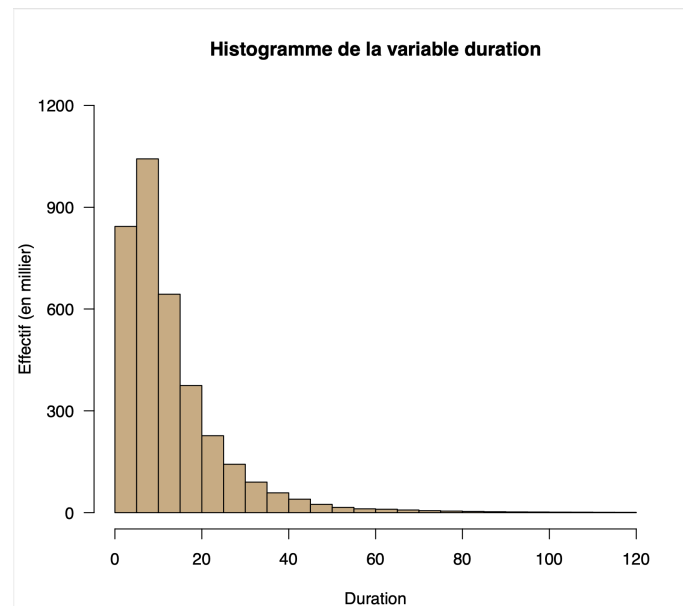


Fig. 5. Histogramme de la variable duration

La table ci-dessous (cf. Fig. 6) nous montre la proportion de valeurs supérieures à 60 minutes. Celle-ci nous permet de confirmer que seulement 1,69% des locations ont une valeur supérieur à 60 min. Par conséquent, nous considérerons les valeurs de location supérieures strictement à 60 minutes comme des valeurs extrêmes. Dorénavant, notre étude sera portée sur les valeurs dont la durée est comprise dans cet intervalle : $]0;60]$.

FALSE	TRUE
98.31	1.69

Fig. 6. Proportion des valeurs strictement supérieures à 60 minutes de locations

1.4 Distribution de la variable duration

La figure ci-dessous (cf. Fig. 6) nous confirme que le nombre de locations le plus important se situe entre 0 et 10 minutes. La seconde période où l'on retrouve le plus de locations est de 10 à 20 minutes, puis 20 à 30 minutes et ainsi de suite jusqu'à 50 à 60 minutes, soit de manière décroissante. Ce dernier intervalle représente le plus petit effectif de locations de vélo sur les valeurs comprises entre 0 et 60 minutes. Cette analyse suggère que les vélos sont principalement utilisés pour des trajets courts.

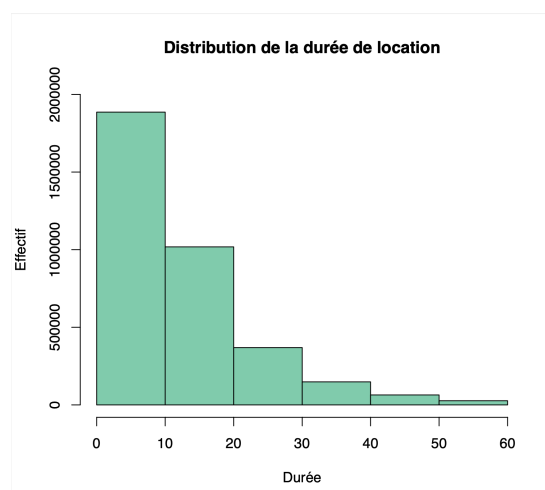


Fig. 7. Distribution de la durée de location

1.5 Distribution de la durée de location selon le type de vélo

Les deux résumés (cf. Fig. 8), (cf. Fig. 9) semblent similaires. On aperçoit une légère différence sur la valeur des nombres qui semblent être plus faibles pour le vélo électrique. Pour comparaison, prenons comme exemple la moyenne. Cette dernière est de 12,32 minutes pour la durée de locations des vélos classiques et de 10,73 minutes pour les vélos électriques. Pour mieux comprendre cette différence, nous allons étudier une représentation graphique de la variable durée pour ces deux paramètres (vélos classiques et vélo électriques).

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02	5.23	9.35	12.32	16.32	60.00

Fig. 8. Résumé de la variable durée pour les vélos classiques

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02	4.13	8.07	10.73	14.33	60.00

Fig. 9. Résumé de la variable durée pour les vélos électriques

À partir du graphique ci-dessous (cf. Fig. 10), on constate une similarité dans la durée d'utilisation des vélos, quel que soit leur type. Cependant, on observe un léger décalage vers la gauche de la courbe turquoise (vélo électrique). Cela explique les valeurs plus faibles du résumé des durées pour les vélos électriques vu précédemment. Cette disparité peut probablement être expliquée au fait que les utilisateurs se déplacent plus rapidement lorsqu'ils utilisent un vélo électrique.

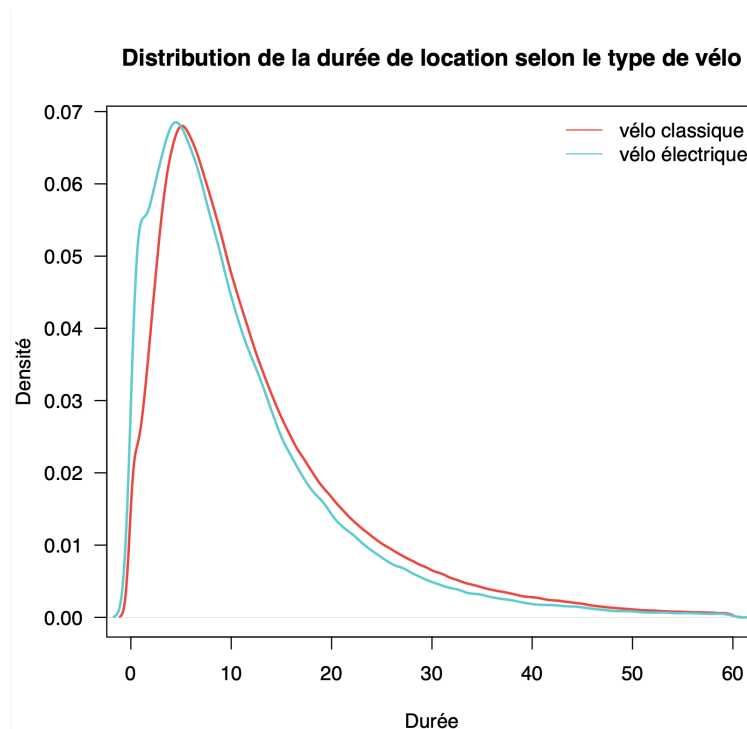


Fig. 10. Temps de locations en fonction du type de vélo

2 Analyse des fréquences et des trajets

Cette seconde section est consacrée à l'analyse des fréquences et des trajets des locations de vélos. Elle complétera la première en apportant des éclairages sur l'utilisation des vélos. Notre analyse portera sur les jours et les heures les plus propices à la location des vélos, ainsi que sur les trajets les plus fréquemment empruntés.

2.1 Répartition des locations selon les jours de la semaine

Le graphique ci-dessous (cf. Fig. 11), nous renvoie la répartition des locations de vélos en fonctions des jours. Grâce à ce dernier, on constate une utilisation moins importante le lundi et le dimanche avec environ 35 000 utilisations sur le mois de septembre respectivement pour les 2 jours. Pour les 5 autres jours, la fourchette d'utilisation des vélos sur le mois est comprise entre 541 454 et 609 411 utilisations. Le vendredi a été le jour avec le plus de locations de vélos sur le mois de septembre 2023 avec 609 411 locations au total.

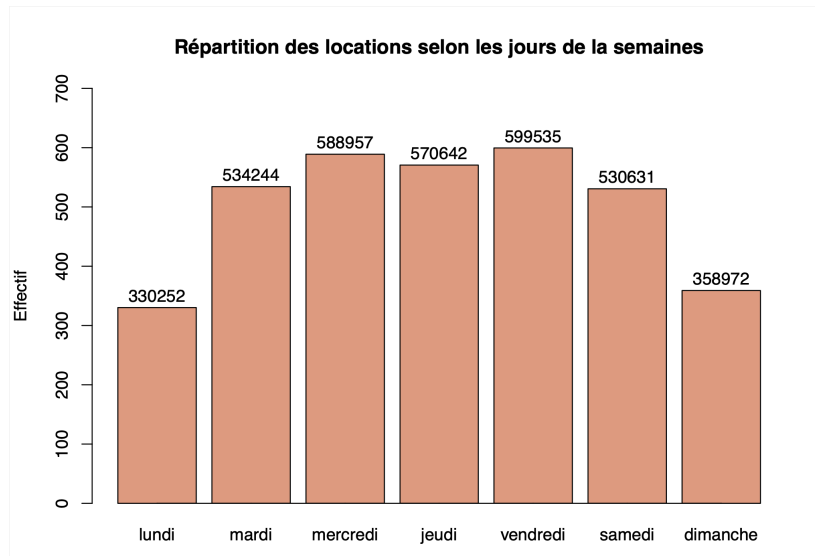


Fig. 11. Location en fonction des jours

2.2 Répartition des locations selon les heures de la journée

On peut observer que le graphique ci-dessous (cf. Fig. 12) illustre la répartition des locations de vélos en fonction des heures de la journée. Nous observons tout d'abord un pic d'utilisation aux alentours de 8 heures, ce qui pourrait correspondre à l'ouverture des bureaux dans la ville. Le second pic est situé aux alentours de 17-18 heures, soit les horaires de fermetures de bureau ou bien les activités de loisirs. Ce pic d'utilisations des vélos est plus important, le pourcentage cumulé entre 17 heures et 19 heures est de 17,5%.

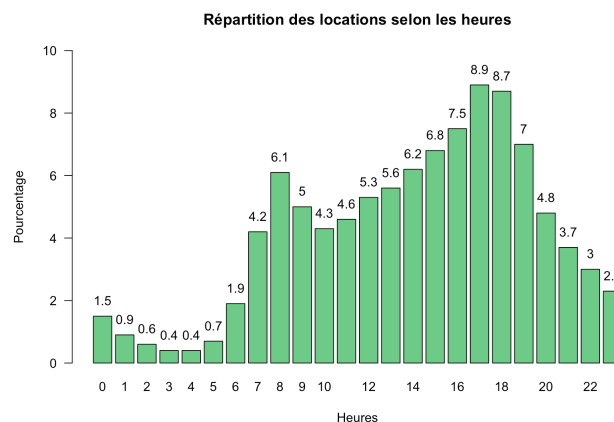


Fig. 12. Location en fonction des heures

2.3 Top 10 des trajets les plus fréquentés

Via la (cf. Fig. 13), on peut constater les 10 trajets les plus fréquents à vélo Citibike en novembre 2023 à New York. Sur ce classement, nous notons que 6 parmi les 10 trajets partent et arrivent à la même station, comme c'est le cas pour le trajet le plus fréquent en novembre 2023 : "Central Park S and 6 Ave" à "Central Park S and 6 Ave", qui a été effectué 1479 fois au cours du mois.

Central Park S & 6 Ave ; Central Park S & 6 Ave	Dock St & Front St ; Old Fulton St
1479	1095
Old Fulton St ; Dock St & Front St	West St & Chambers St ; West St & Chambers St
924	807
Broadway & W 58 St ; Broadway & W 58 St	11 Ave & W 41 St ; 11 Ave & W 41 St
706	701
7 Ave & Central Park South ; 7 Ave & Central Park South	North Moore St & Greenwich St ; Vesey St & Church St
692	664
E 10 St & Avenue A ; E 11 St & Avenue B	12 Ave & W 40 St ; 12 Ave & W 40 St
646	638

Fig. 13. Top 10 des trajets les plus effectués

2.4 Top 10 des trajets avec station de départ et d'arrivée différentes

La table ci-dessous (cf. Fig. 14) représente les proportions de trajets ayant les mêmes stations d'arrivée et de départ. On observe que un peu plus de 5% des trajets partagent la même station de départ et d'arrivée. De ce fait, nous allons réexaminer les 10 trajets les plus fréquemment empruntés en excluant les valeurs avec des stations identiques.

FALSE	TRUE
5.02	94.98

Fig. 14. Proportion des valeurs ayant une station de départ et arrivé différentes

Après exclusion des valeurs où les stations de départ et d'arrivé étaient identique, nous trouvons le classement suivant (cf. Fig. 15). Ainsi, le trajet le plus fréquent est : "Dock St. & Front" à "Old Fulton St." Ce trajet a été pris 1095 fois le mois de septembre 2023.

Dock St & Front St ; Old Fulton St	Old Fulton St ; Dock St & Front St
1095	924
North Moore St & Greenwich St ; Vesey St & Church St	E 10 St & Avenue A ; E 11 St & Avenue B
664	646
Broadway & W 51 St ; Broadway & W 53 St	St Marks Pl & 2 Ave ; St Marks Pl & 1 Ave
619	601
6 Ave & W 34 St ; Broadway & W 36 St	W 41 St & 8 Ave ; Broadway & W 41 St
566	562
West St & Chambers St ; Pier 40 - Hudson River Park	Central Park S & 6 Ave ; 5 Ave & E 87 St
551	536

Fig. 15. Top 10 des trajets les plus fréquentés avec station différentes

Conclusion

A travers ce rapport nous proposons une analyse des locations des vélos Citibike à New York au cours du mois de septembre 2023. Cette étude se divise en deux parties distinctes, la première mettant en lumière les caractéristiques des locations, tandis que la seconde offre une analyse détaillée des fréquences de locations et des trajets.

La première partie nous montre que sur le réseau de New York, les vélos les plus utilisés en septembre 2023 sont les vélos classiques à plus de 92%. Le temps de locations des vélos électrique et classique est similaire, on observe une moyenne de 12 minutes et 19 secondes pour les vélos classiques et de 10 minutes et 43 secondes pour les vélos électriques. On en conclut que la majorité des personnes louant les vélos optent pour des trajets courts, aux alentours de 10 minutes.

Tout d'abord en analysant les jours d'utilisations des vélos, le vendredi, est le jour ayant le plus de locations. Cependant, la différence avec le deuxième jour où il y a le plus de locations est légère, on note un écart de 12 915 locations. On remarque que le lundi et le dimanche sont les jours avec le moins d'utilisation. Secondement, les horaires d'utilisations. Deux points importants ont été relevés à 8 heures et aux alentours de 17 et 19 heures ce qui serait explicable par les horaires de bureau. La majorité des trajets sont des déplacements de station à station, en particulier depuis et vers la station de Central Park, suggérant des trajets récréatifs. De ce fait, une deuxième étude a été réalisée en examinant les déplacements entre des stations de départ et d'arrivée différentes. Les trajets les plus fréquemment effectués sont ceux de Dock St & Front à Old Fulton St.

3 Annexe

La compagnie Citibike propose toute une liste de jeux de données sur la location de leurs vélos, comme le montre l'image ci-dessous :

[How it Works](#)[Pricing](#)[Find a Bike](#)[Explore NYC](#)[Help](#)

System Data

Where do Citi Bikers ride? When do they ride? How far do they go? Which stations are most popular? What days of the week are most rides taken on? We've heard all of these questions and more from you, and we're happy to provide the data to help you discover the answers to these questions and more. We invite developers, engineers, statisticians, artists, academics and other interested members of the public to use the data we provide for analysis, development, visualization and whatever else moves you.

This data is provided according to the [Citi Bike Data Use Policy](#).

Citi Bike Trip Histories

We publish [downloadable files of Citi Bike trip data](#). The data includes:

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID

Ce rapport se base sur les données relatives aux locations de vélos dans la ville de New York. Toutes les données utilisées sont disponibles via le lien suivant :

[Historiques des voyages CitiBike](#)

Table des figures

1	Extrait des données concernant les locations de vélos	3
2	Répartitions des types de vélos	4
3	Résumé de la variable duration	4
4	Proportion des valeurs aberrantes	4
5	Histogramme de la variable duration	5
6	Proportion des valeurs strictement supérieures à 60 minutes de locations	5
7	Distribution de la durée de location	5
8	Résumé de la variable duration pour les vélos classiques	6
9	Résumé de la variable duration pour les vélos électriques	6
10	Temps de locations en fonction du type de vélo	6
11	Location en fonction des jours	7
12	Location en fonction des heures	7
13	Top 10 des trajets les plus effectués	8
14	Proportion des valeurs ayant une station de départ et arrivé différentes	8
15	Top 10 des trajets les plus fréquentés avec station différentes	8