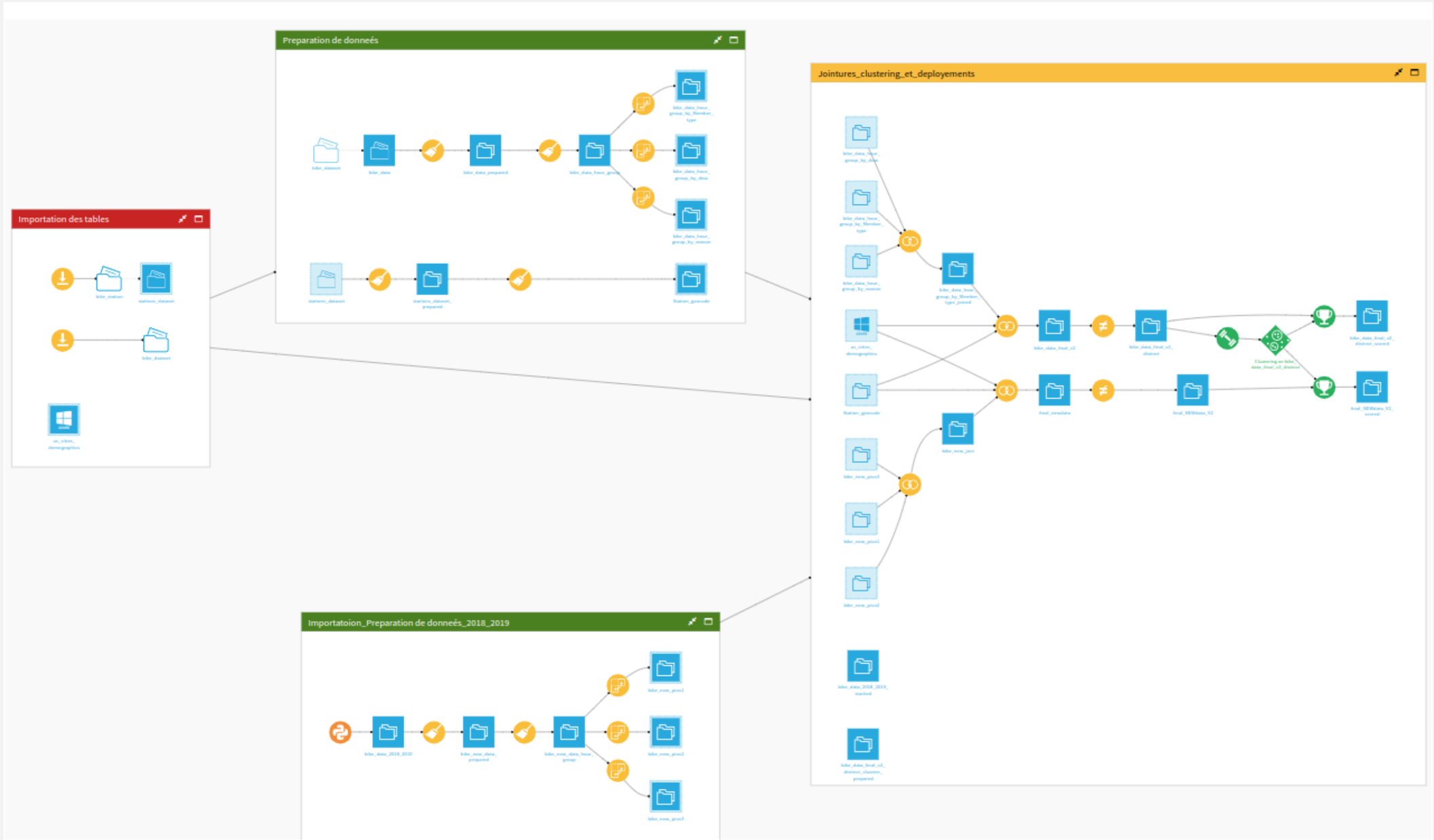
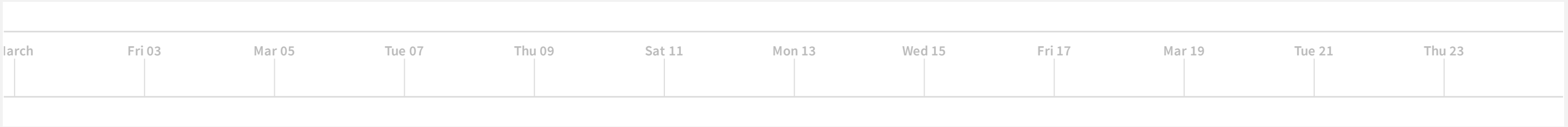
Table of Contents

Flow


Importation des tables
Preparation de donneés
Importatoion_Preparation de donneés_2018_2019
Jointures_clustering_et_deployements

Automatisation



Analyse du clustering sur les données originales

This slide is empty

Switch to edit mode to add tiles

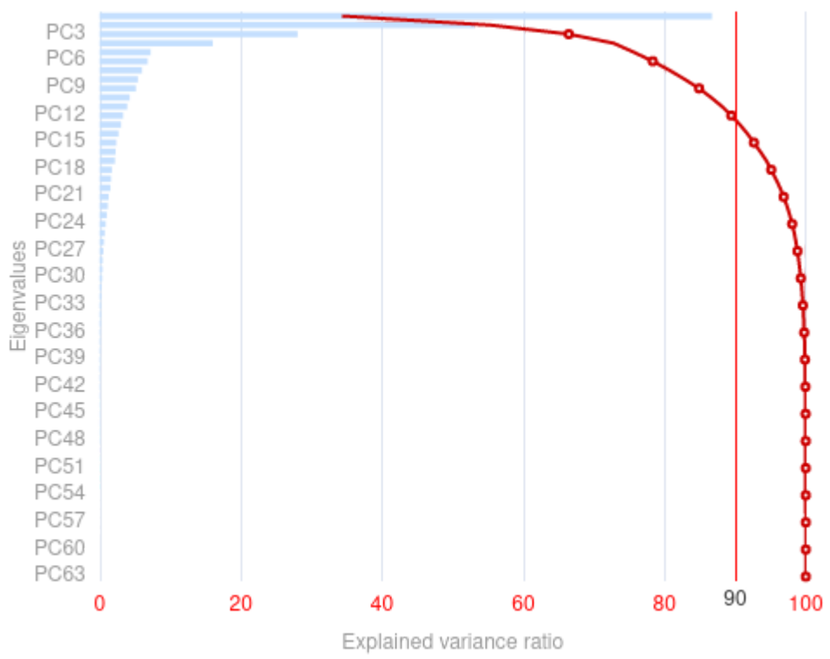
Dataset

Dataset

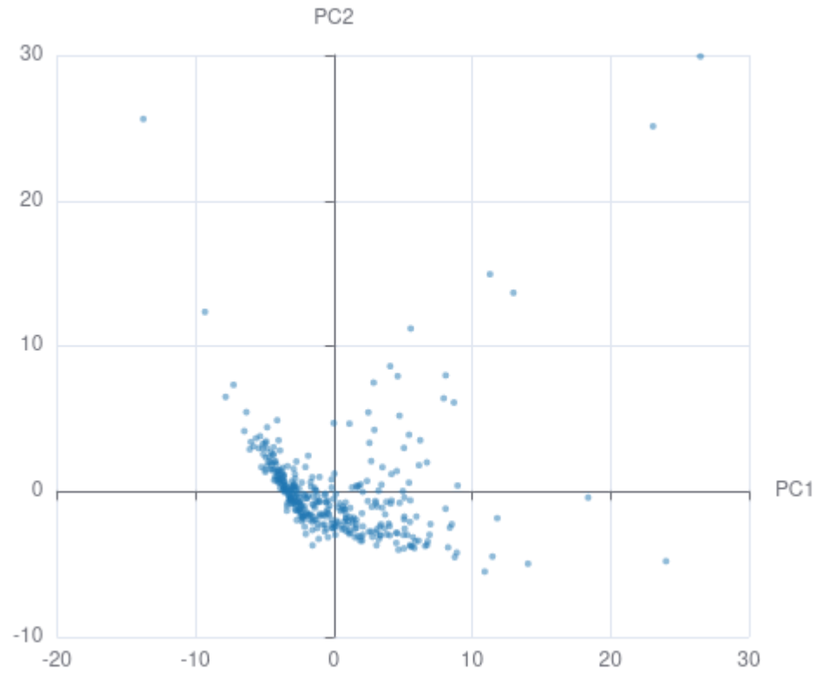
Start station	Casual_après-midi_Duration_avg	Casual_après-midi_count	Casual_matin_Duration_avg	Casual_matin_count	Casual_midi_Duration_avg	Casual_midi_count
10th & E St NW	2387.605623100304	9212	2653.5886458818054	2149	2751.8543077650024	6349
10th & Florida Ave NW	1498.7513134851138	571	1095.131188118812	404	1550.8791423001949	513
10th & G St NW	2477.932086213908	4918	2907.3828867761454	1157	2886.3701523545706	2888
10th & K St NW	2137.4146177916855	4474	2230.690534249612	4511	2460.890326209224	3556
10th & Monroe St NE	2416.953642384106	453	2027.936170212766	235	2540.1070038910507	514
10th & U St NW	1907.590481786134	1702	1570.849884526559	433	1844.0887096774193	868
10th St & Constitution Ave NW	2250.0402465918205	19952	3167.165839829908	1411	2518.3757809930944	12164
10th St & L'Enfant Plaza SW	2483.9716981132074	1590	4522.489847715736	394	2945.8633811603245	1603
11th & F St NW	2536.809474555234	4834	2994.7287581699347	1530	3109.731707317073	3239
11th & Girard St NW	1647.403314917127	181	1100.820512820513	117	1117.3482142857142	112
11th & H St NE	1595.548802946593	1086	1517.2250580046405	431	1856.7769028871392	762
11th & Kenyon St NW	1875.7108088761631	1397	1467.687341772152	790	1906.1576885406464	1021
11th & M St NW	1852.8367686170213	3008	1623.907358738502	1522	2097.886622675465	1667
11th & O St NW	1641.969696969697	1188	1821.1293302540416	866	1760.5726927939318	791
11th & S St NW	1378.3742331288342	978	1256.8648648648648	518	1993.2240837696336	955
12th & Army Navy Dr	2539.839035769829	1286	2667.95	780	2892.014720314033	1019
12th & Irving St NE	1907.5260869565218	230	1727.340425531915	141	3100.225641025641	195
12th & L St NW	1653.760975609756	1435	1877.1053497942387	1215	2154.310344827586	1102
12th & Newton St NE	3059.254054054054	185	1789.3362068965516	116	3401.193717277487	191
12th & U St NW	1876.3639774859287	2132	1237.5395973154361	745	1950.626099706745	1364
12th St & Pennsylvania Ave SE	2353.1146788990827	218	1481.811320754717	159	1775.3085106382978	188
13th & D St NE	1482.9666666666667	930	1403.7093333333332	1125	1951.3690078037905	897
13th & H St NE	1870.3808122424955	1699	1457.95295404814	914	1869.6612244897958	1225
13th St & Eastern Ave	2571.59509202454	163	2174.929411764706	85	3217.3867924528304	106

PCA on 63 variables

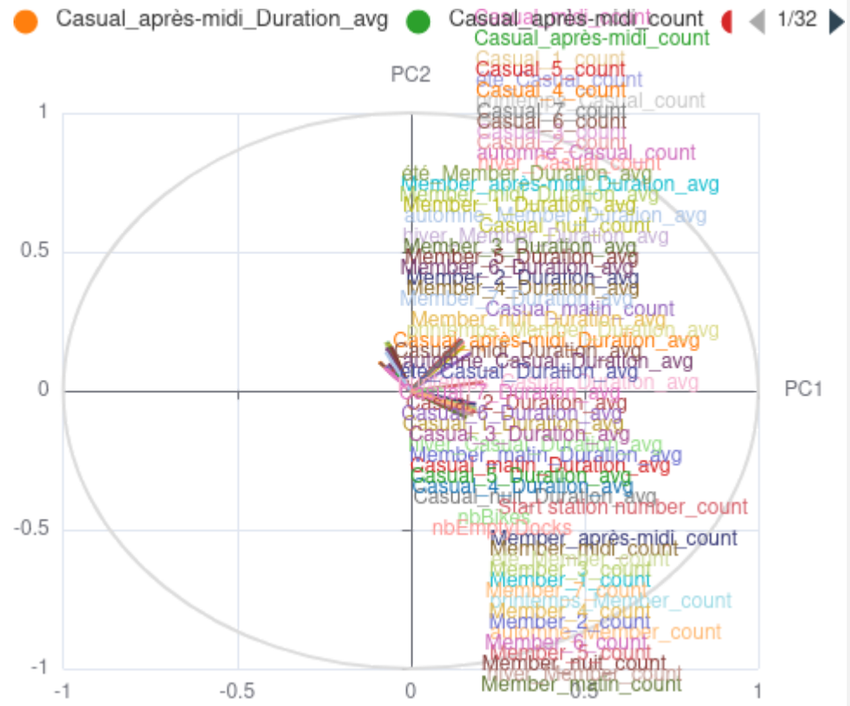
Scree plot



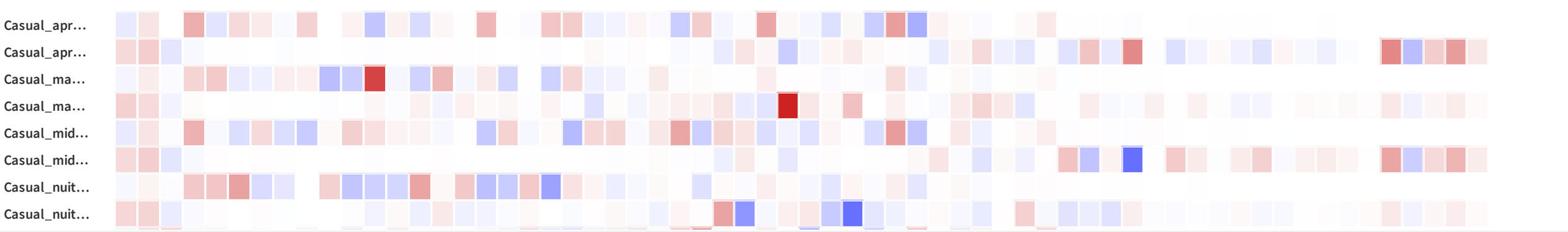
Scatter plot



Loading plot



Principal components



KMeans (k=4) (s3) - v1

KMeans (k=4)

Cluster outliers

Trained in 3 seconds on 488 records

Zone d'affaires		**81 (16.60%)**
Périphérie éloigné		103 (21.11%)
Centre et périphérie		284 (58.20%)
Zone Touristique		18 (3.69%)
Jeffeson&Linco…		2 (0.41%)

Zone d'affaires

Essentiellement utilisé par la membres pour aller au travail

Observations

- **hiver_Member_count** is in average 250% greater : mean of 9106 against 2599 globally
- **Member_midi_count** is in average 228% greater : mean of 9544 against 2906 globally
- **printemps_Member_count** is in average 241% greater : mean of 14015 against 4107 globally

Model

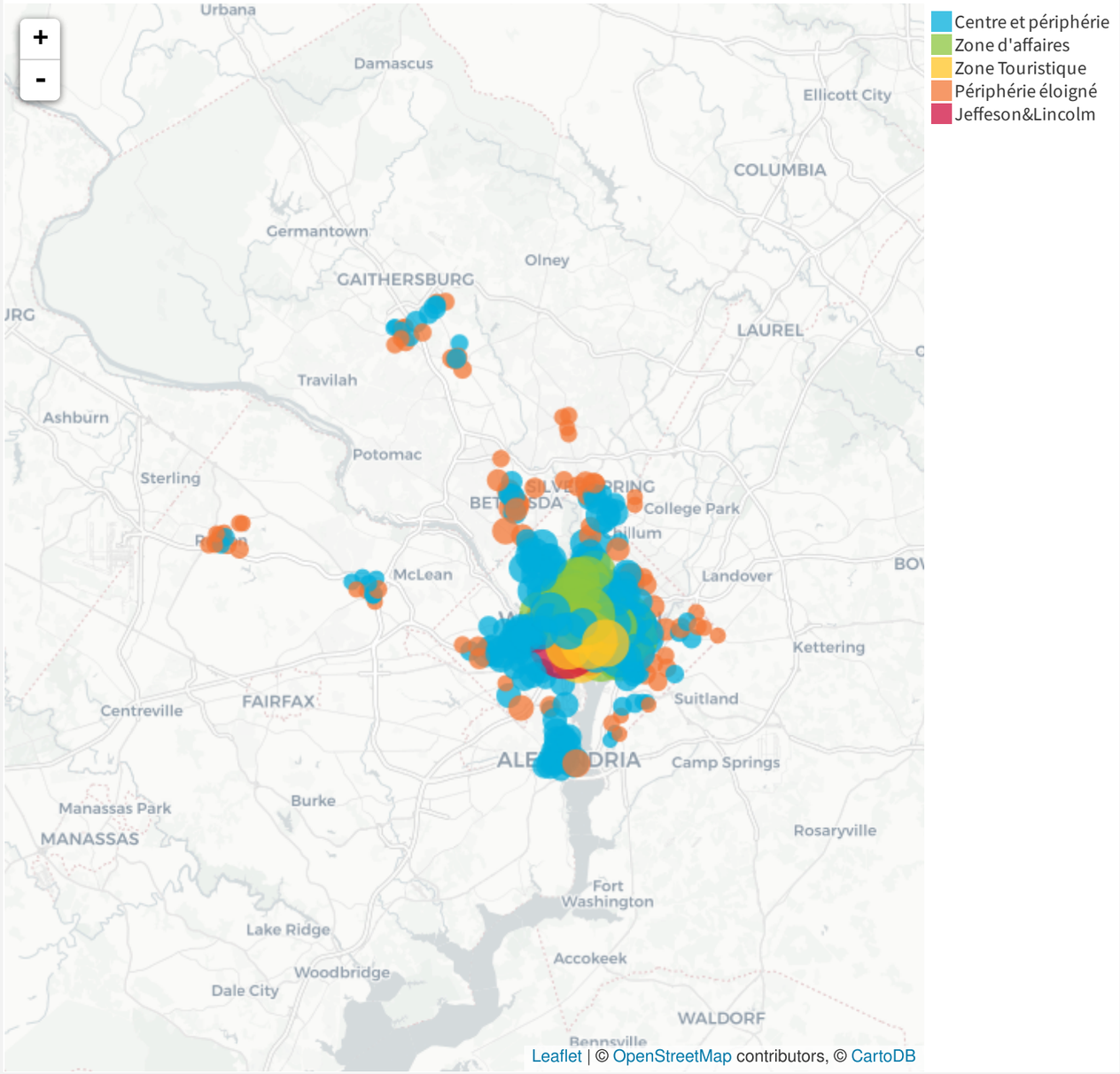
Model ID	**S-BIKESHARING-4mJSgeDX-1679494045815**
Model type	**Clustering**
Code Env	**DSS builtin env**
Python version	**3.6.8**

Metadata

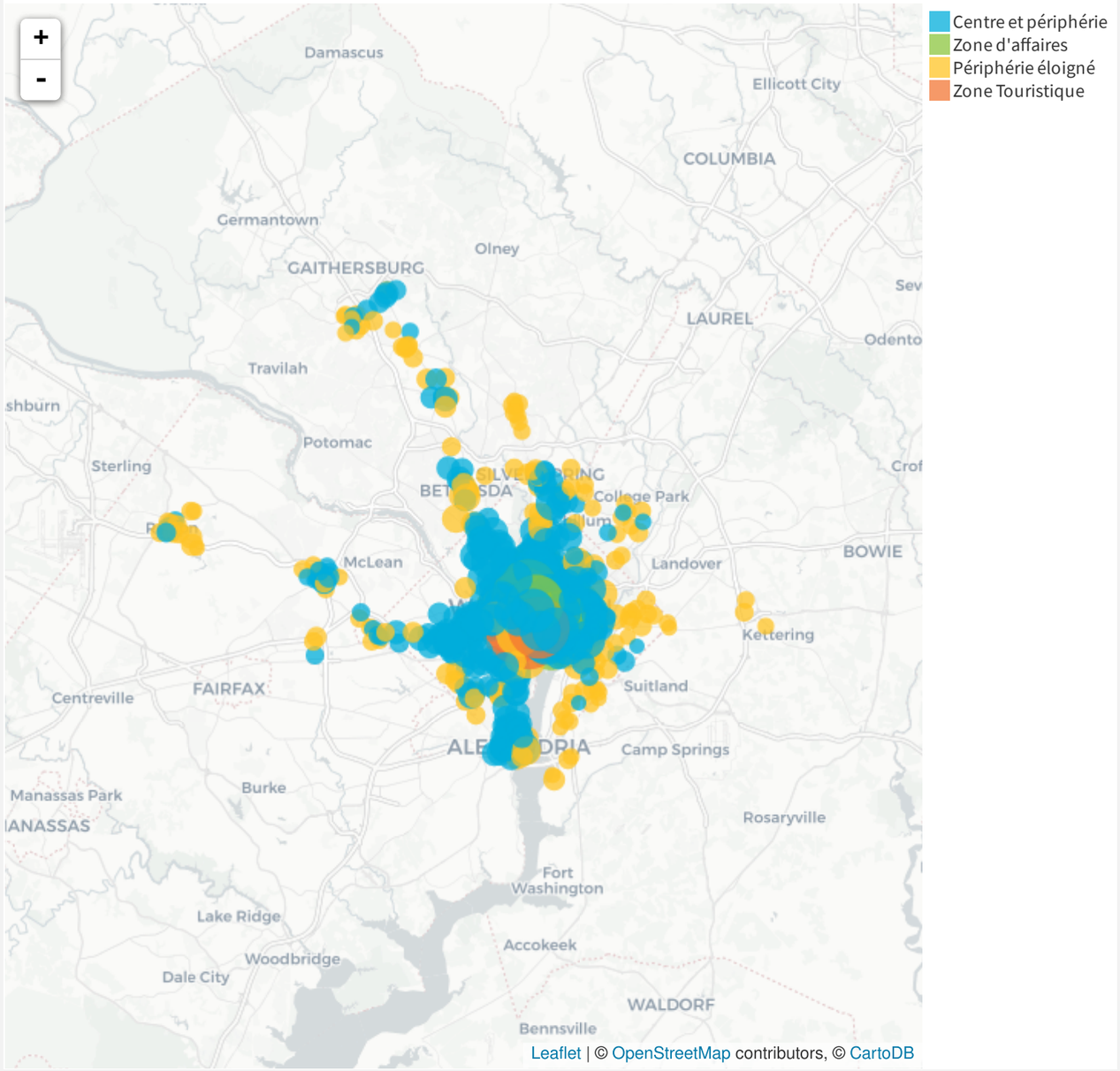
trainDataset:dataset-name	→	bike_data_final_v2_distinct
testDataset:dataset-name	→	bike_data_final_v2_distinct
evaluationDataset:dataset-name	→	bike_data_final_v2_distinct
model:algorithm	→	KMEANS
model:date	→	2023-03-22T08:47:20.263+0000
evaluation:date	→	2023-03-22T08:47:20.263+0000

Carte des clusters

Carte des clusters 2015-2017

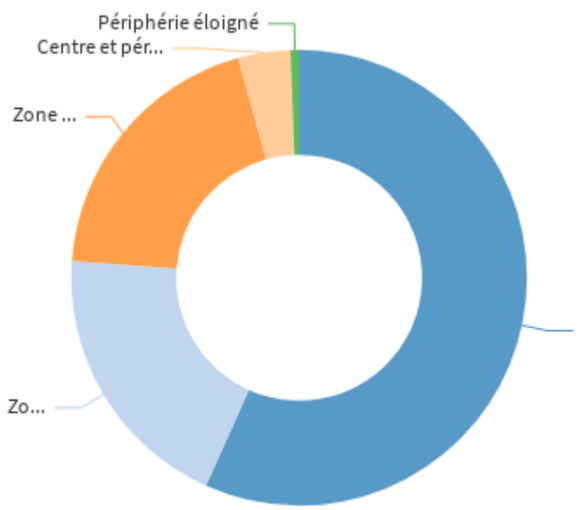


Carte des clusters 2018-2019

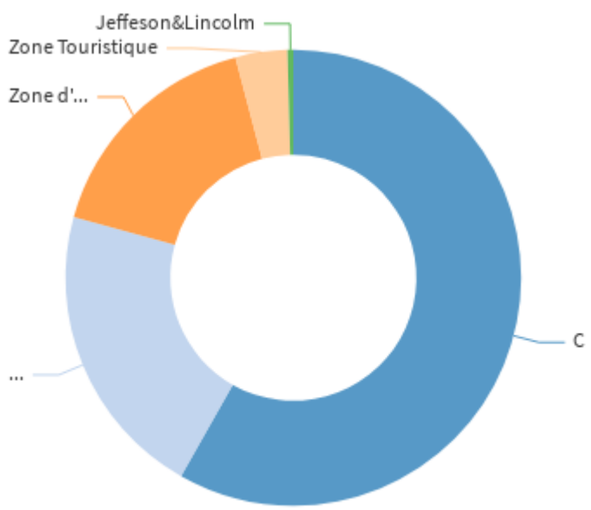


Données de bases sur clusters

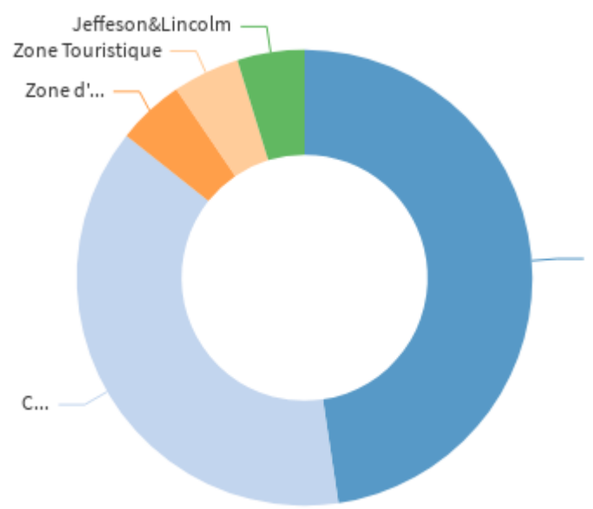
Nombre moyen de départ par cluster 2015-2017



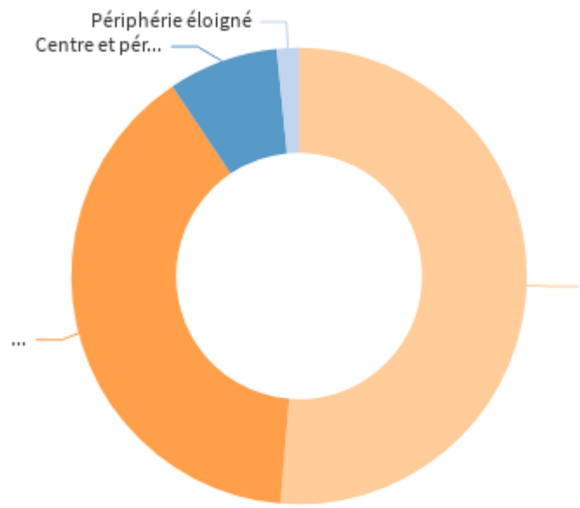
Nombre de stations par cluster 2015-2017



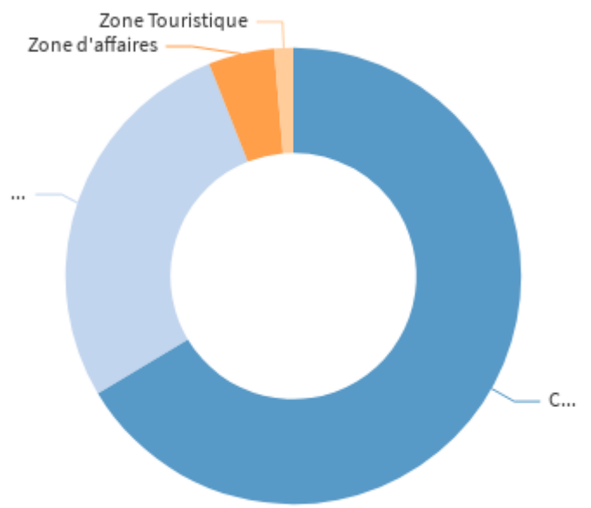
Nombre de villes par cluster 2015-2017



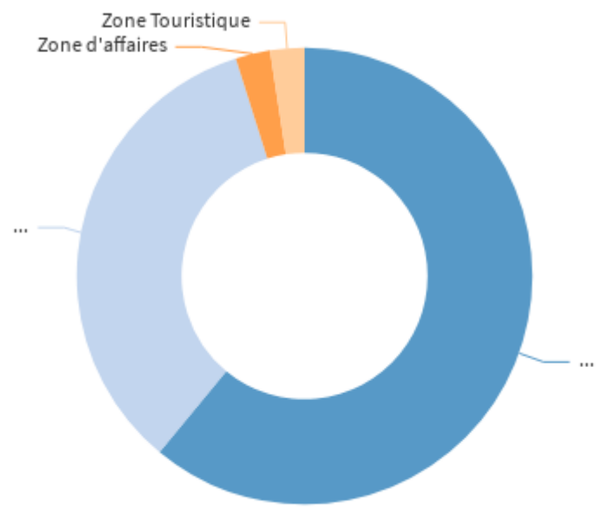
Nombre moyen de départ par cluster 2018-2019



Nombre de stations par cluster 2018-2019

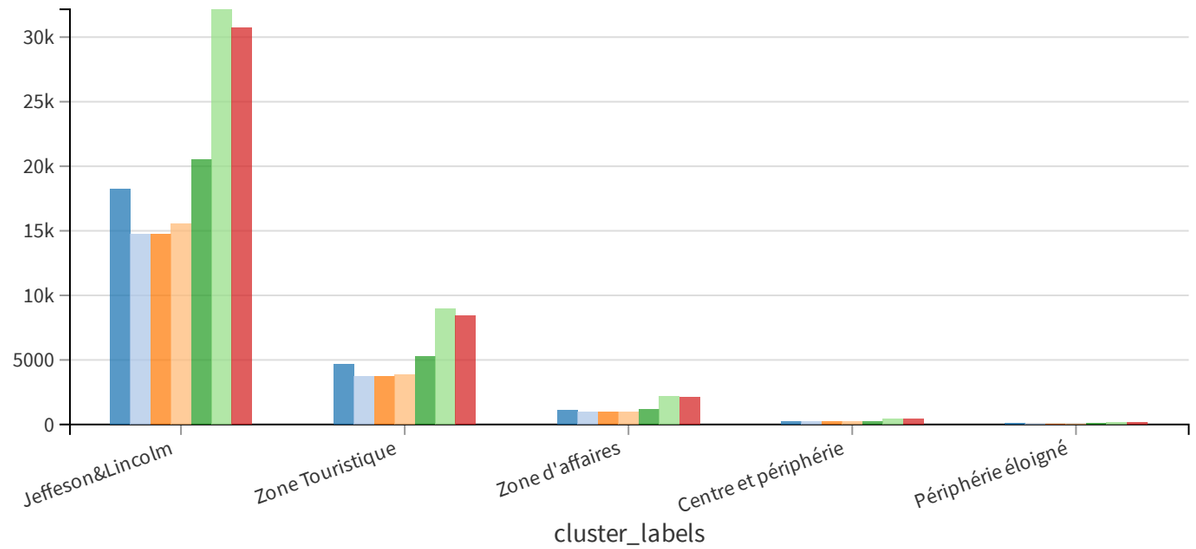


Nombre de villes par cluster 2018-2019

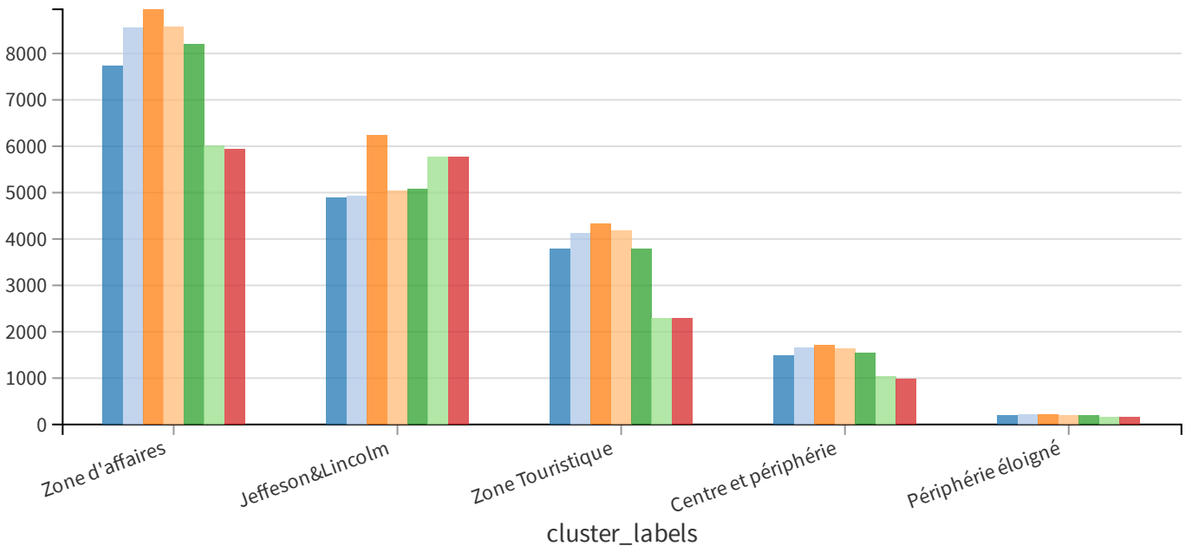


Répartition au cours de la semaine

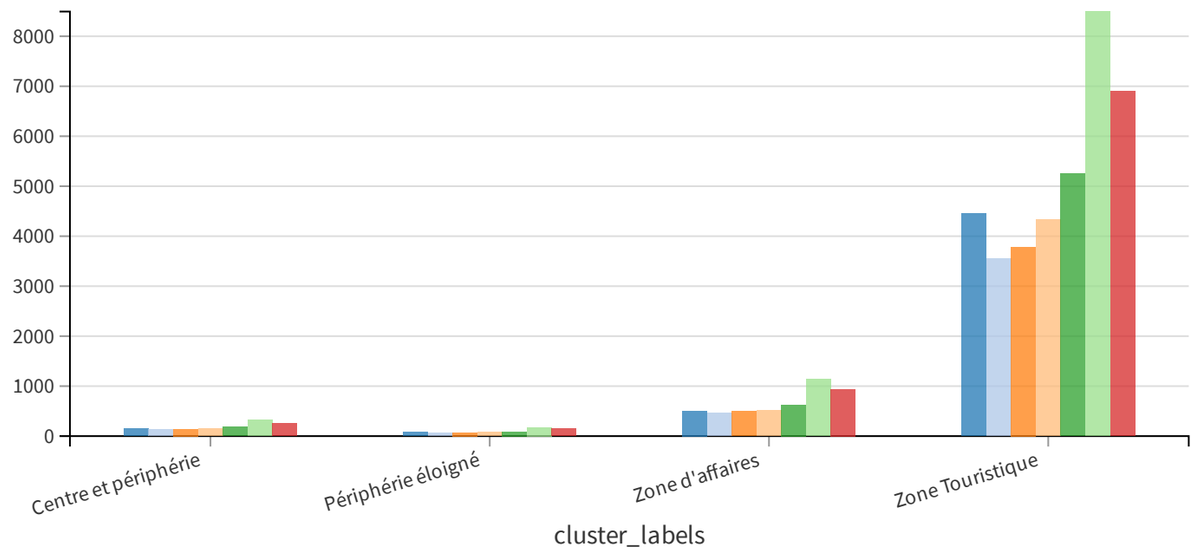
Nombre moyen de départs au cours de la semaine pour les utilisateurs occasionnels 2015-2017



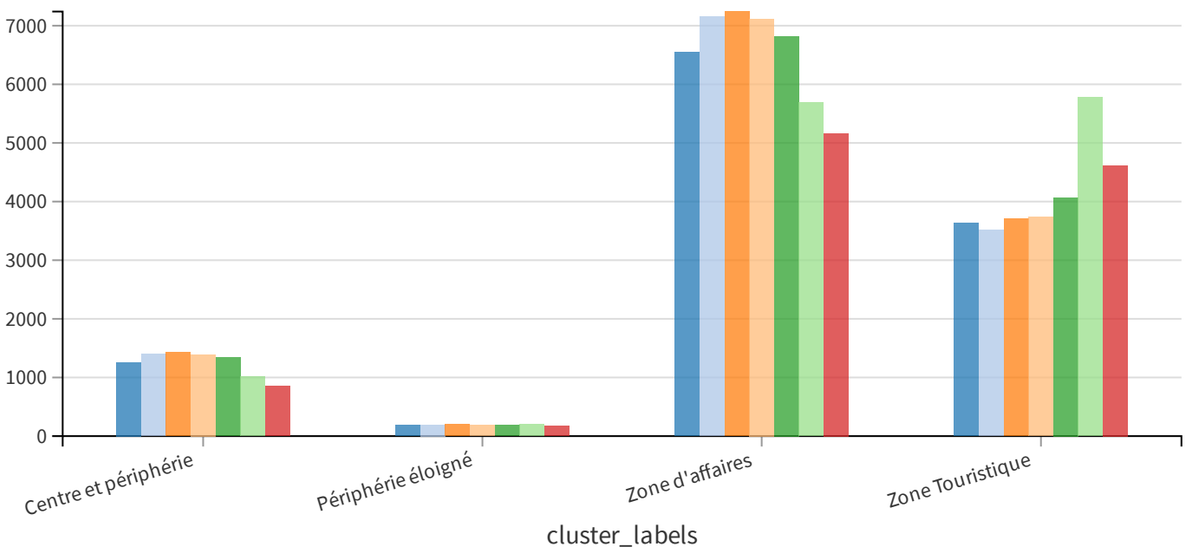
Nombre de voyage au cours de la semaine pour les membres 2015-2017



Nombre moyen de départs au cours de la semaine pour les utilisateurs occasionnels 2018-2019

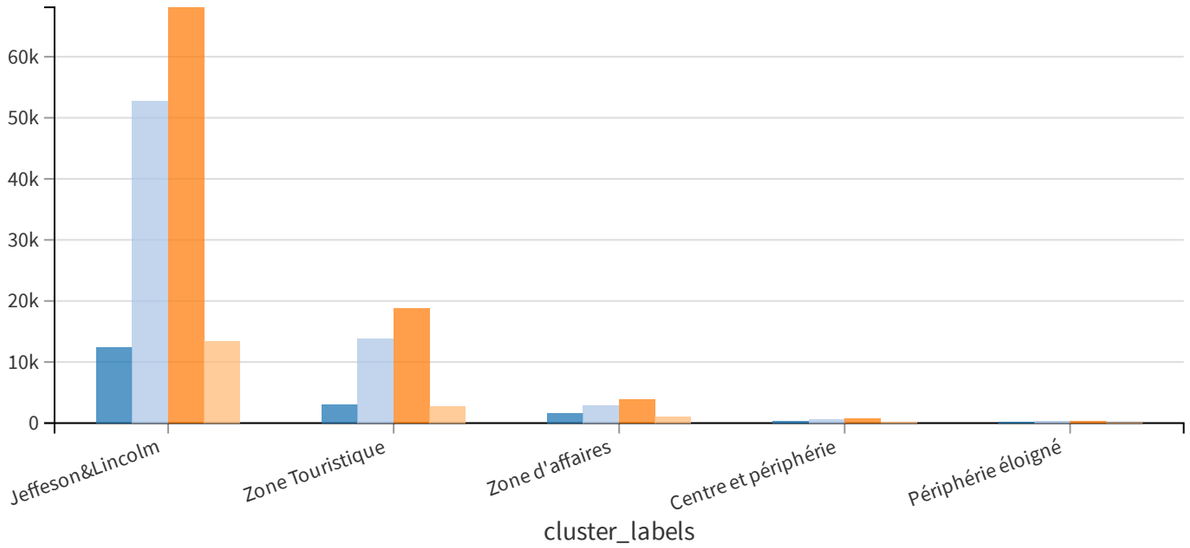


Nombre de voyage au cours de la semaine pour les membres 2018-2019

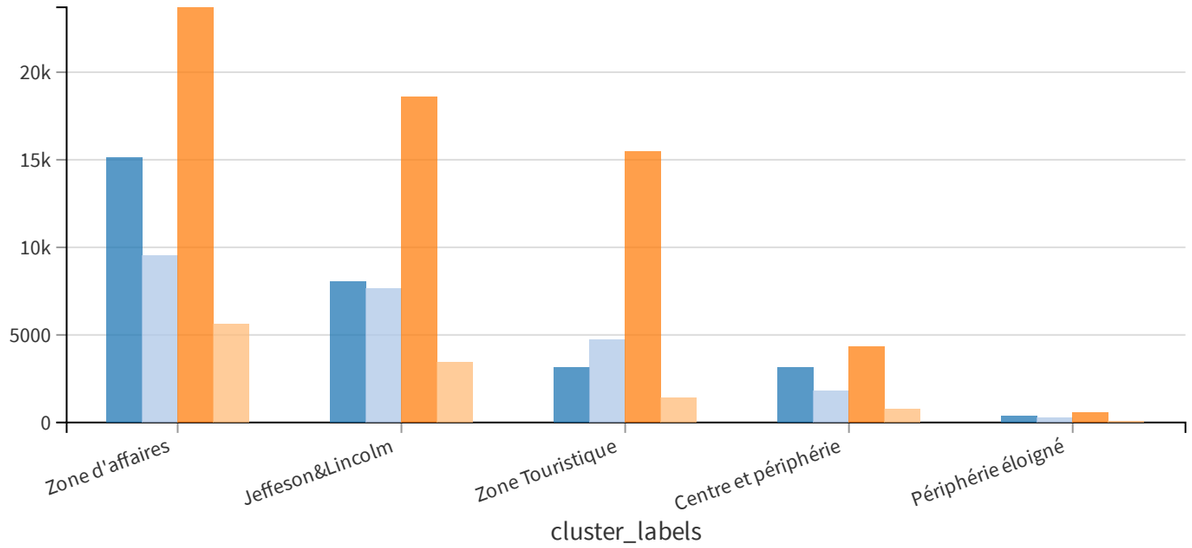


Répartition en fonction de la plage horaire

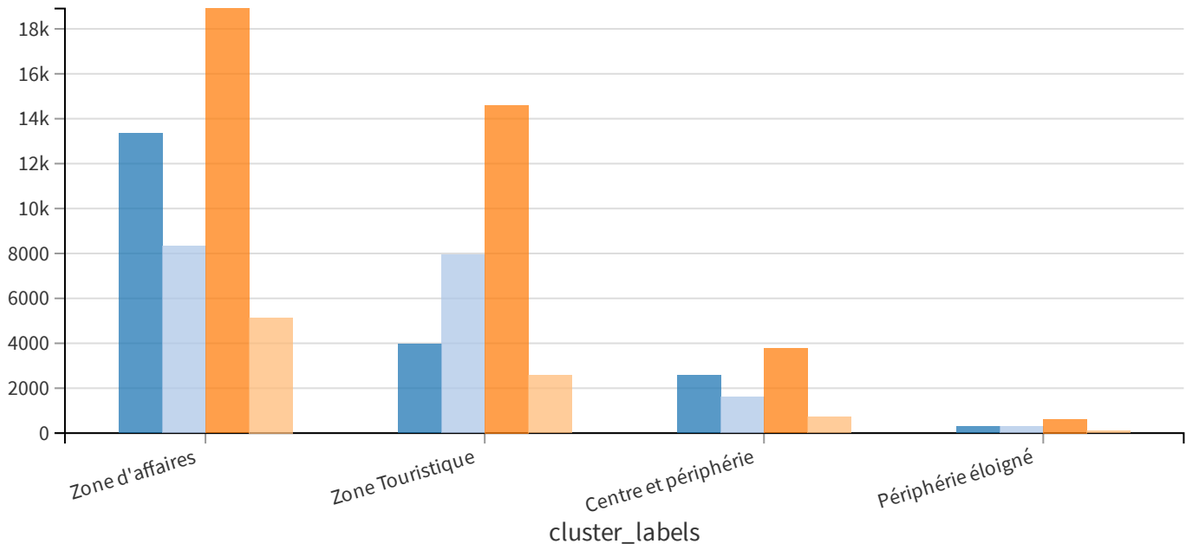
Nombre de voyage au cours de la journée pour les utilisateurs occasionnels 2015-2017



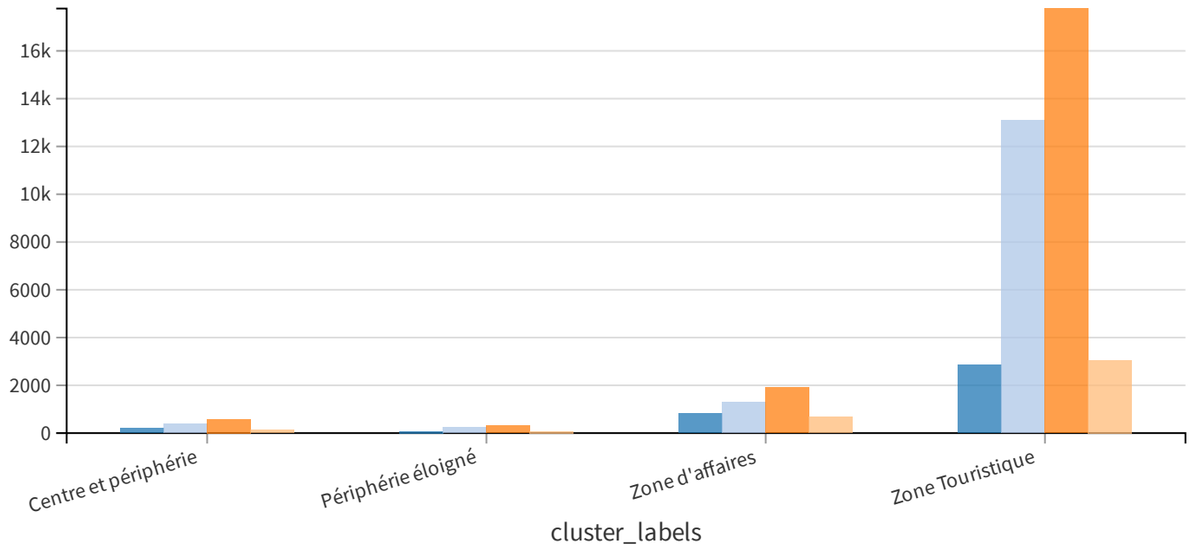
Nombre de voyage au cours la journée pour les membres 2015-2017



Nombre de voyage au cours de la journée pour les utilisateurs occasionnels 2018-2019

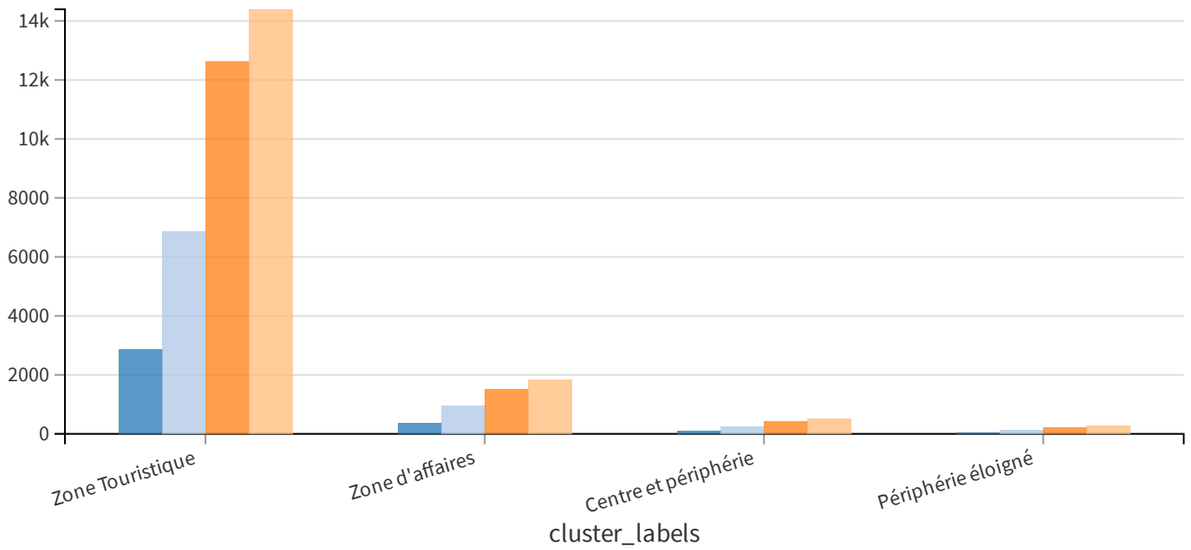


Nombre de voyage au cours la journée pour les membres 2018-2019

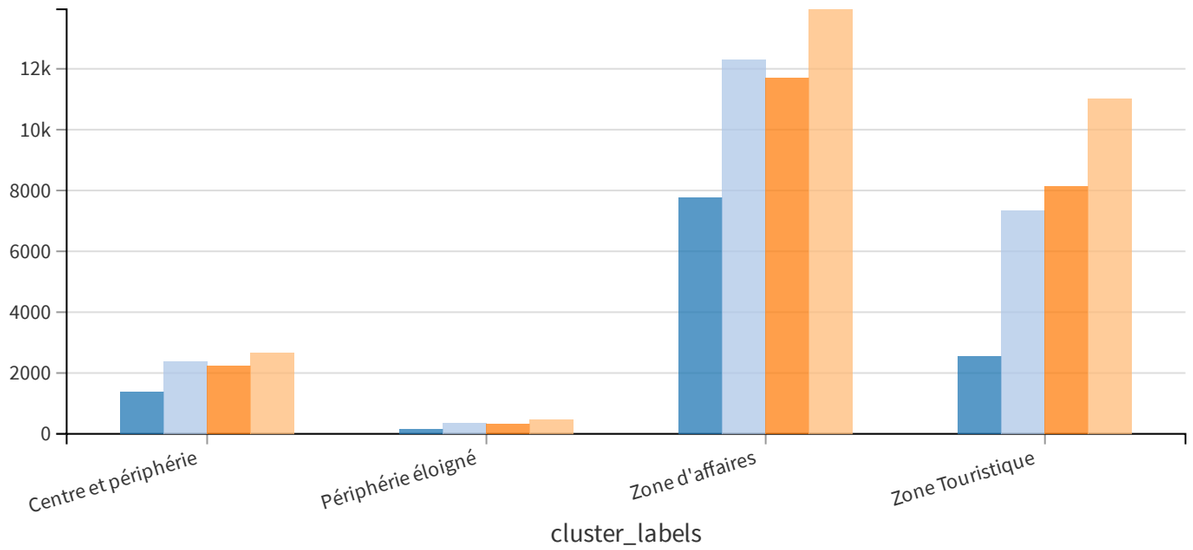


Répartition en fonction de la saison

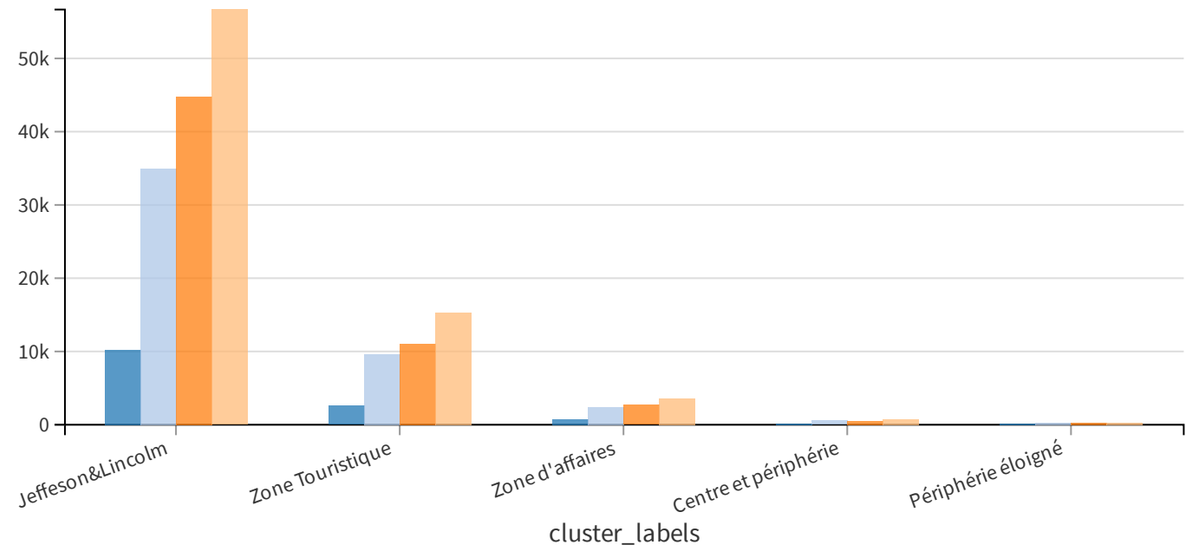
Nombre de voyages selon la saison pour les utilisateurs occasionnels 2018-2019



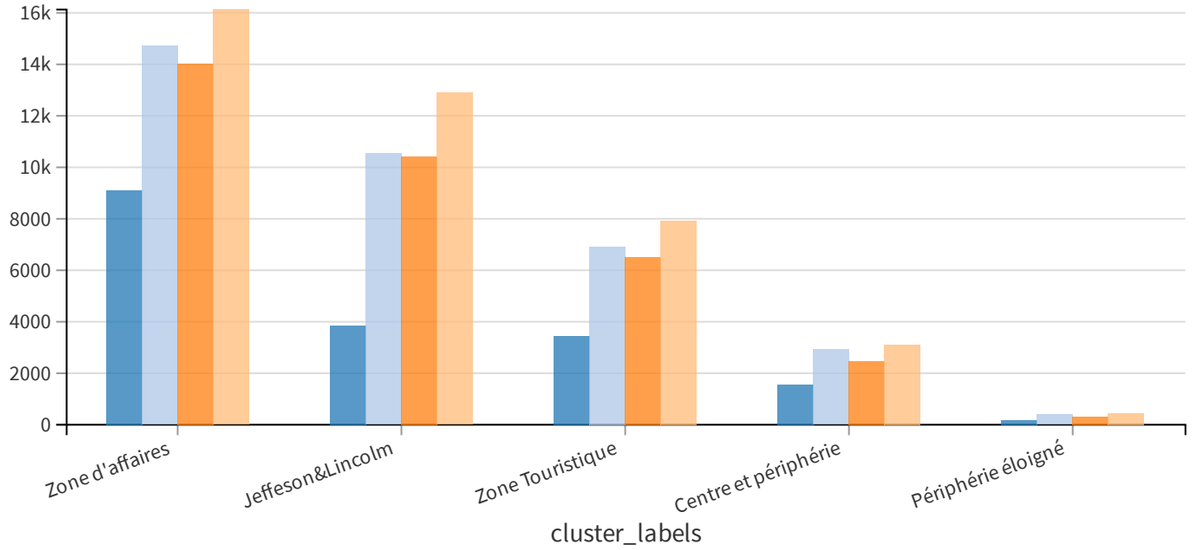
Nombre de voyages selon la saison pour les membres 2018-2019



Nombre de voyages selon la saison pour les utilisateurs occasionnels 2015-2017

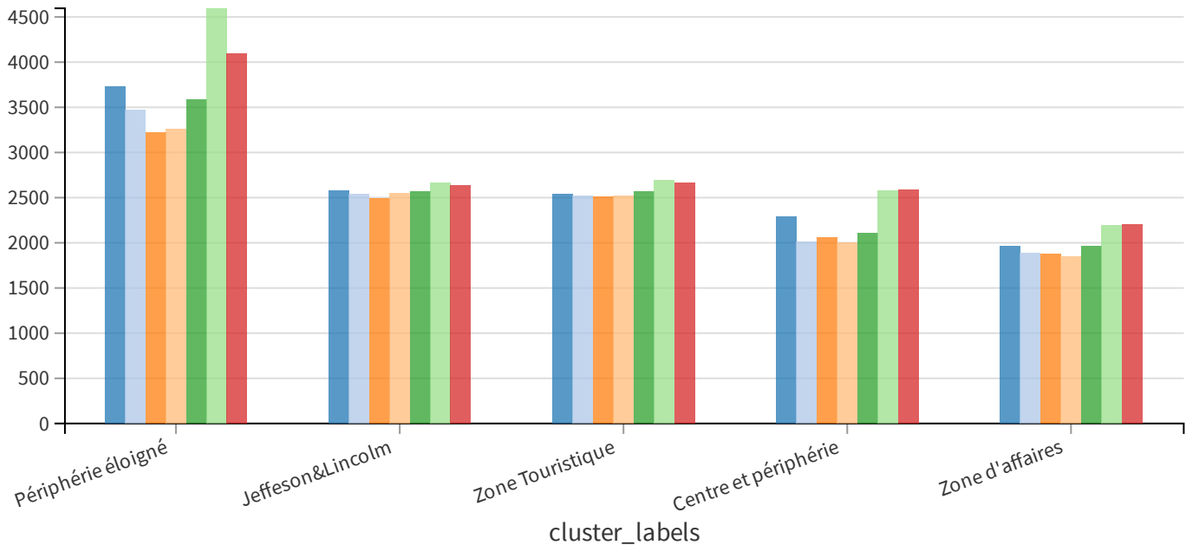


Nombre de voyages selon la saison pour les membres 2015-2017

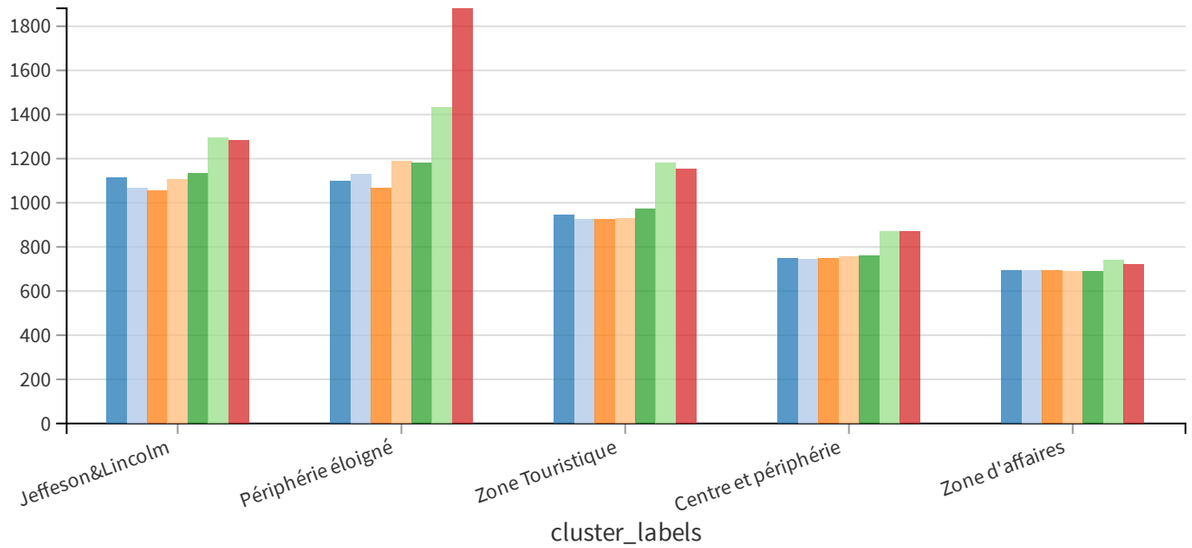


Répartition de la durée des voyages au cours de la semaine

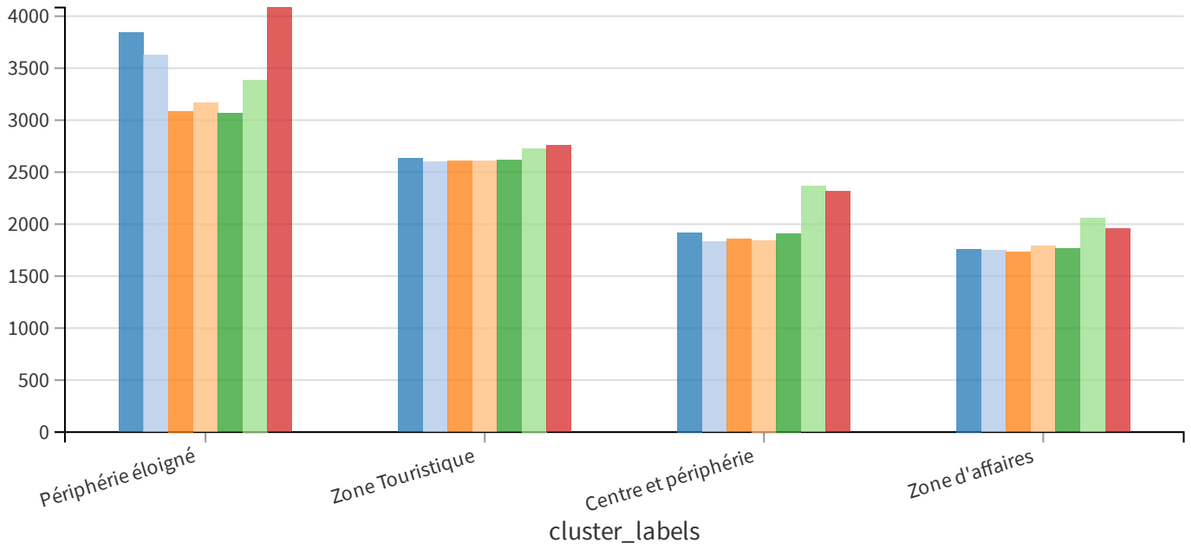
Durée moyenne des voyages au cours de la semaine pour les utilisateurs occasionnels 2015-2017



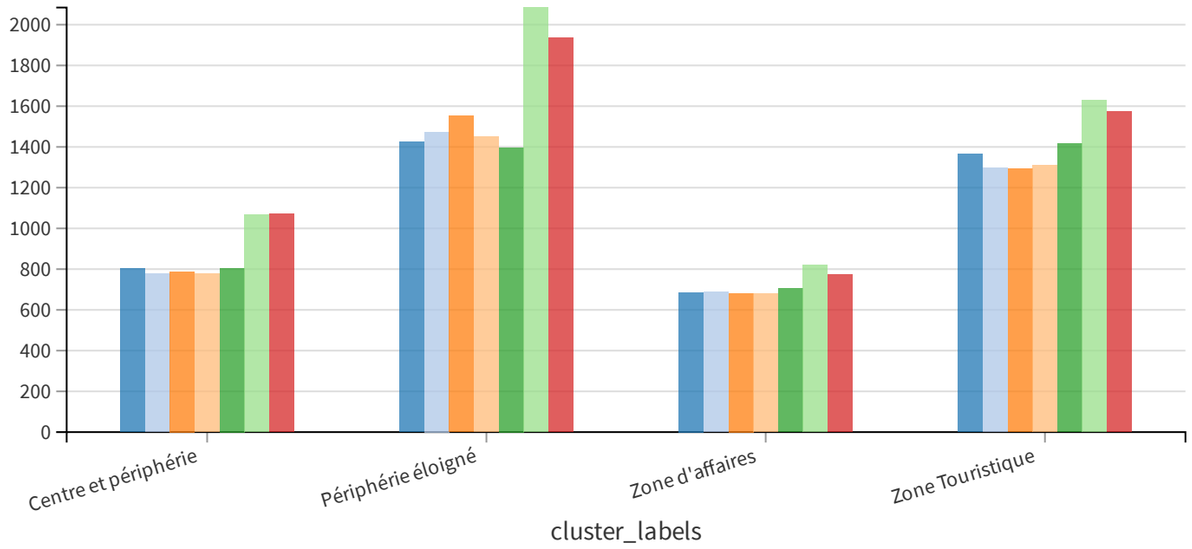
Durée moyenne des voyage au cours de la semaine pour les membres 2015-2017



Durée moyenne des voyages au cours de la semaine pour les utilisateurs occasionnels 2018-2019

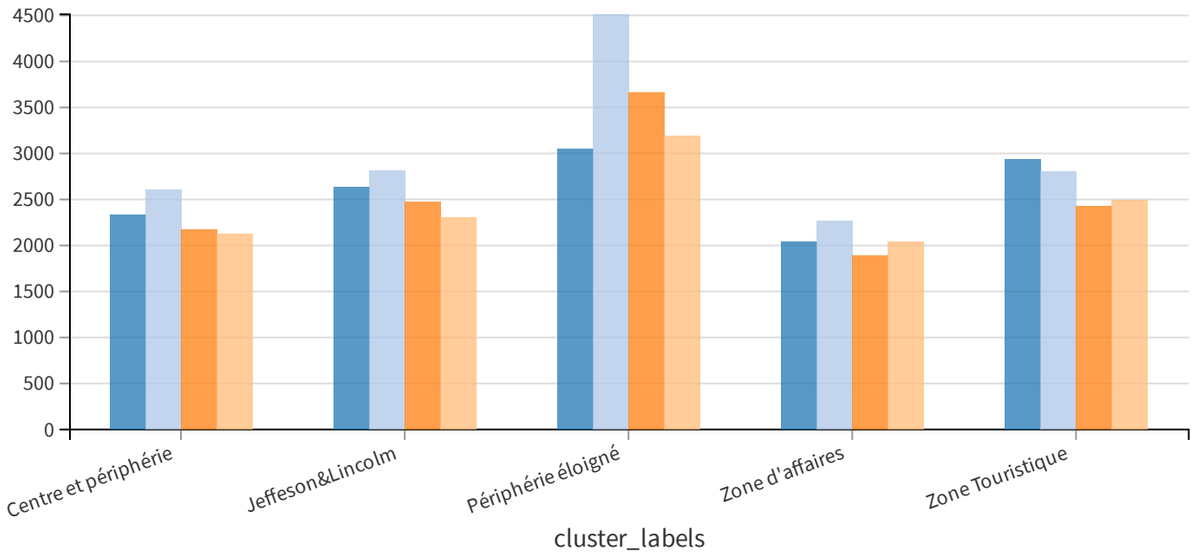


Durée moyenne des voyage au cours de la semaine pour les membres 2018-2019

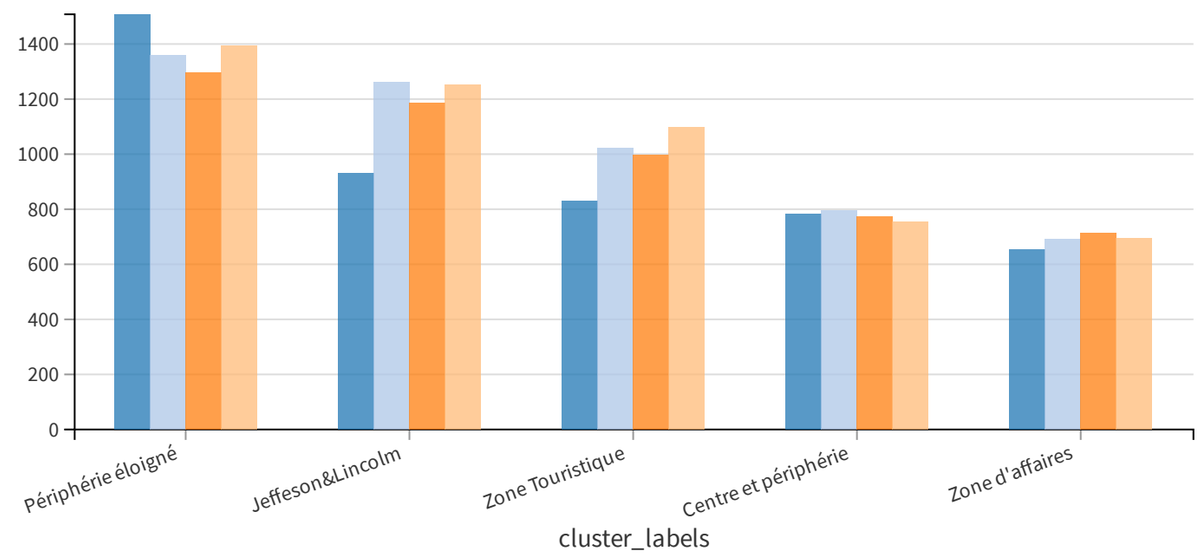


Répartition de la durée des voyages au cours la journée

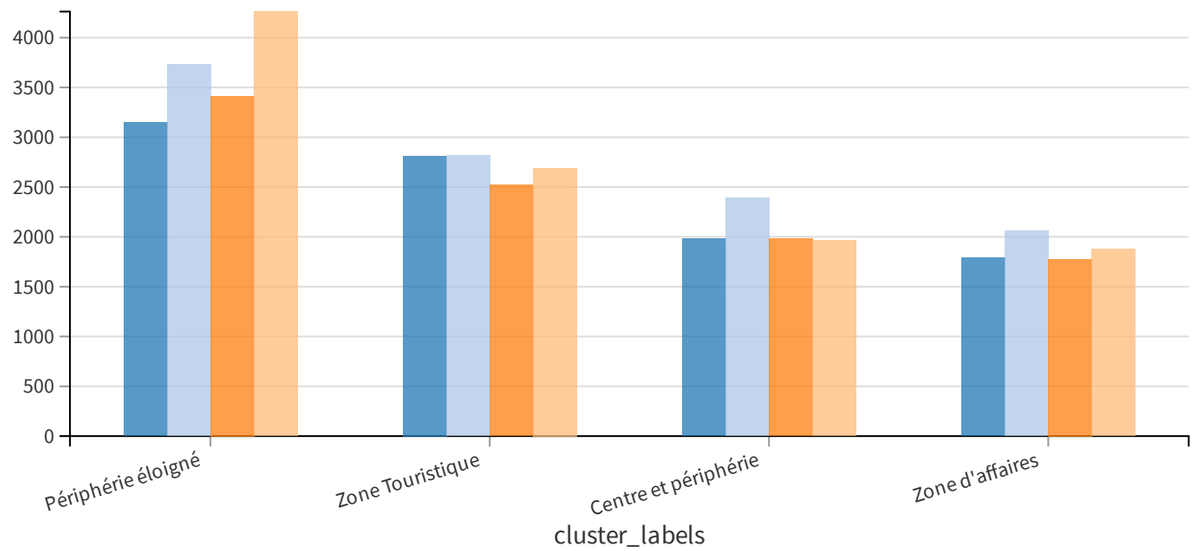
Durée moyenne des voyages au cours de la journée pour les utilisateurs occasionnels 2015-2017



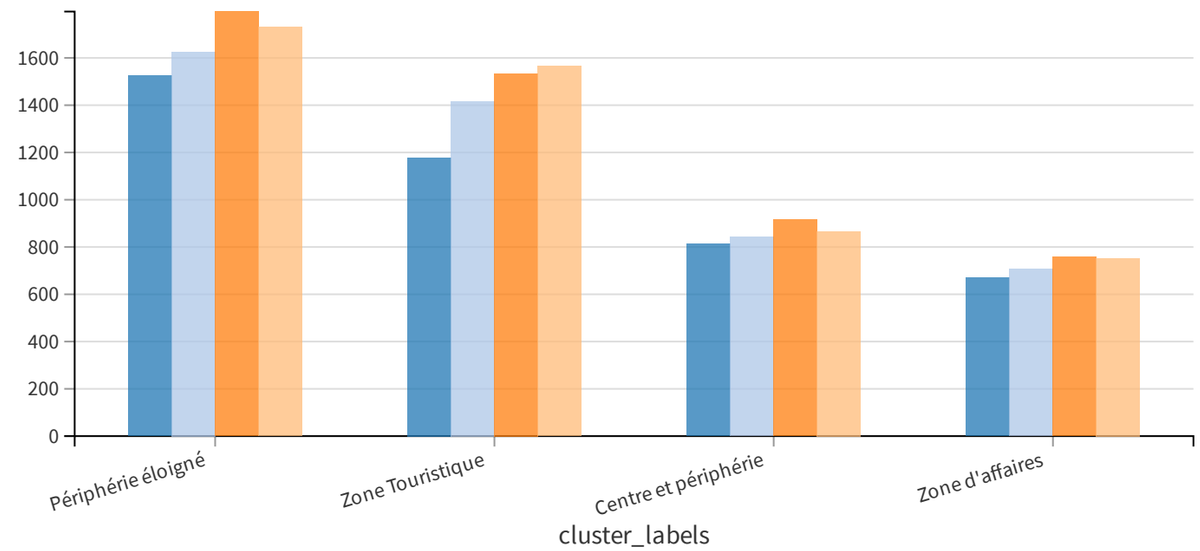
Durée moyenne des voyages au cours de la journée pour les membres 2015-2017



Durée moyenne des voyages au cours de la journée pour les utilisateurs occasionnels 2018-2019

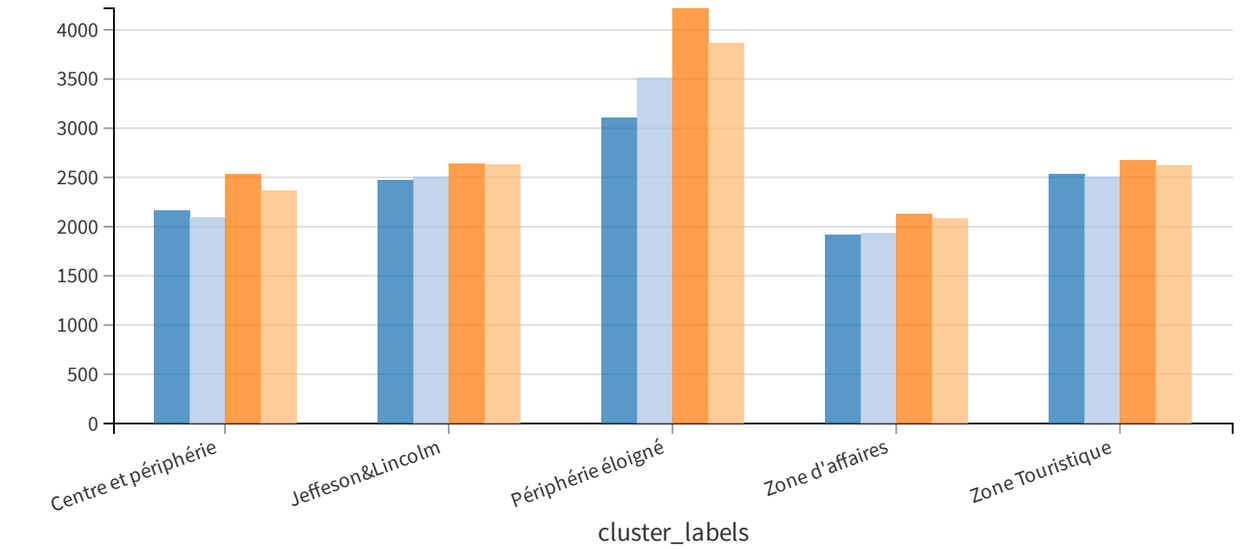


Durée moyenne des voyages au cours de la journée pour les membres 2018-2019

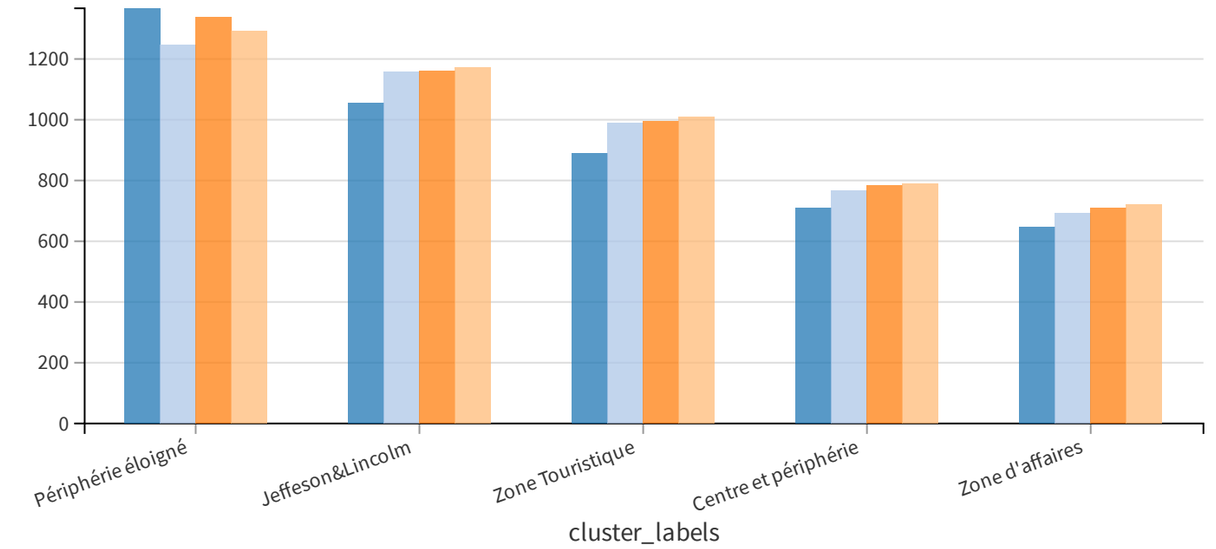


Répartition de la durée des voyages selon la saison

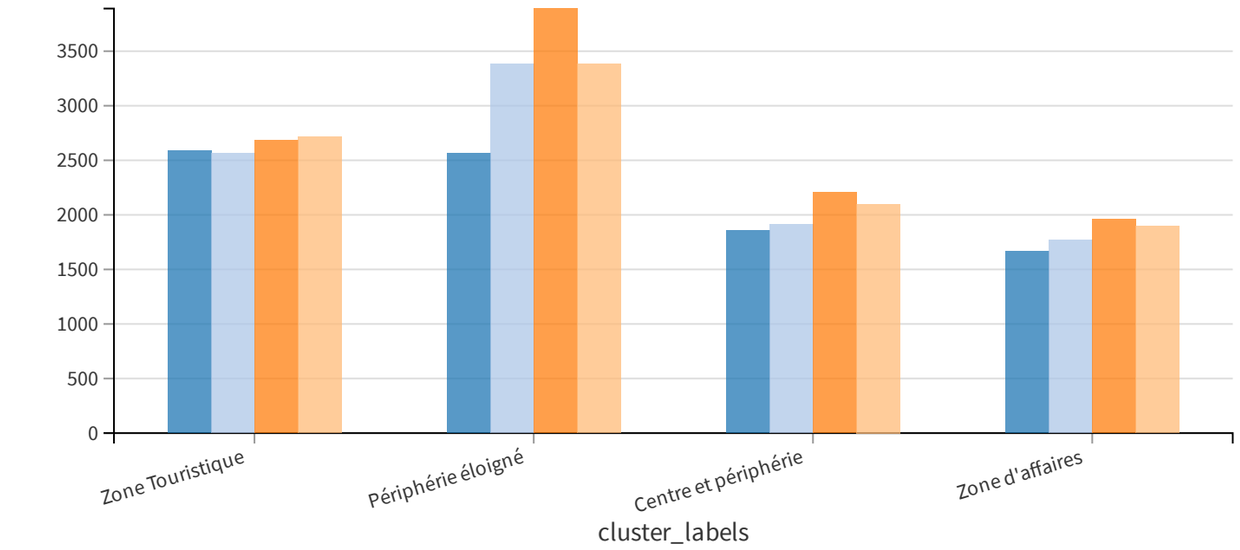
Durée moyenne des voyages selon la saison pour les utilisateurs occasionnels 2015-2017



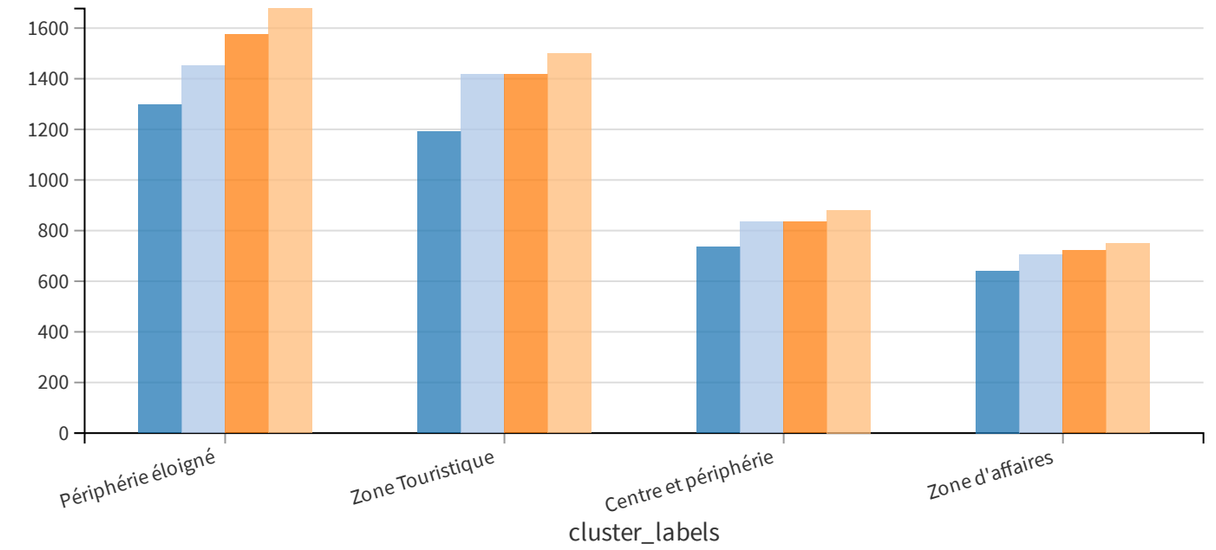
Durée moyenne des voyages selon la saison pour les membres 2015-2017



Durée moyenne des voyages selon la saison pour les utilisateurs occasionnels 2018-2019



Durée moyenne des voyages selon la saison pour les membres 2018-2019



Zone Touristique

- **fort trafic, en particulier le weekend (casual et membres)**
- **localisé à Washington, dans une région comprenant de nombreux monuments**

Centre et périphérie

- **Voyage plus court, semblable au cluster d'affaires**
- **faible trafic**

Périphérie éloignée

- **Voyages longs mais peu nombreux**
- **faible trafic**

Zone d'affaires

- **trajet court avec un pique d'influence le matin (aller) et l'après-midi (retour)**
- **fort trafic (membres)**

Recommandation

Disponibilité des vélos

- **S'assurer une quantité de vélos suffisante pour les stations ayant le plus fort trafic**
- **disponibilité des vélos électriques pour les stations où les trajets sont les plus long**

Entretien de l'infrastructure

- **Surveiller l'état des stations les plus utilisées**

Offre promotionnel et touristique

- **proposer des offres adaptées aux caractéristiques des clusters**

Avantages

- **Gain de productivité**
- **Concentration sur l'analyse des données et du clustering plutôt que le codage**
- **Centralisation des processus**
- **Facilité de prise en main**

Inconvénients

- **Manque de flexibilité**
- **Plugins capricieux**
- **Documentation obscure**
- **Performances**
- **Sécurité des données et dépendance au service**