

Analyse des profils horaires de fréquentation et modélisation des passages sur les stations de comptage vélo à Bruxelles



OLIVIER LEBERT

Contexte

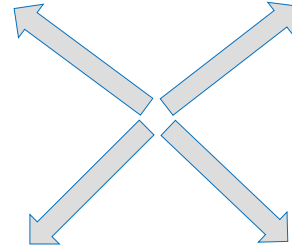
- Distinction des profils horaires des stations de comptage pour une meilleure compréhension des dynamiques de mobilité à Bruxelles
- 18 stations de comptage à Bruxelles
- Méthodes utilisées :
 - ✓ ACP
 - ✓ Clustering,
 - ✓ régressions linéaires
 - ✓ XGBoost



Principales découvertes

Les profils horaires
varient substantiellement
selon les jours de la semaine
et la météo

L'ACP et le clustering révèlent des
groupes de stations avec des
comportements distincts, en lien
avec le contexte urbain ou la
météo



La régression linéaire explique
une part non négligeable
de la variabilité
des passages.

XGBoost offre de bien
meilleures performances
prédictives

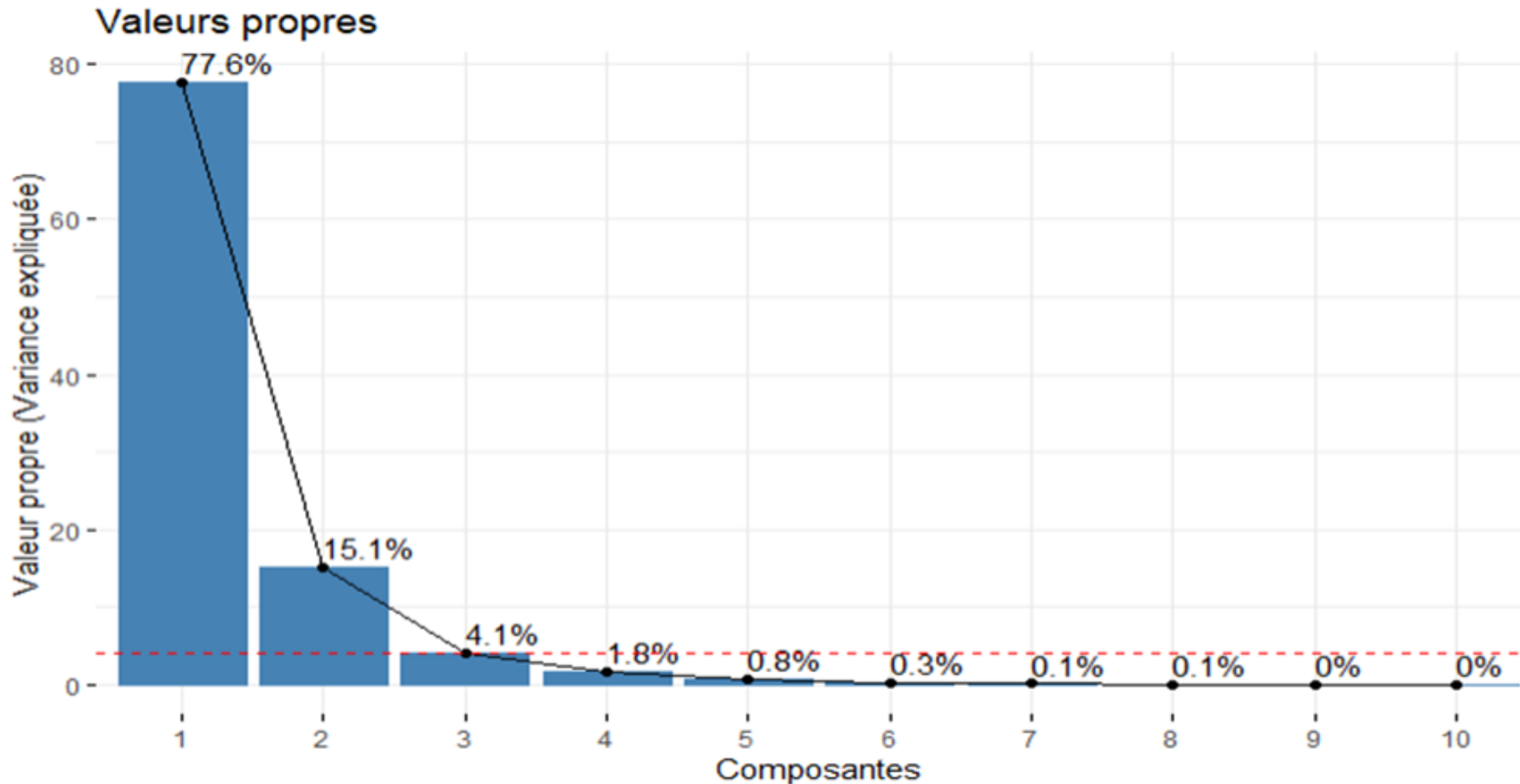
Traitement des données



Critère de Kaiser

Variance expliquée > $100\% / 24 = 4.17\%$

Valeur propre > 1

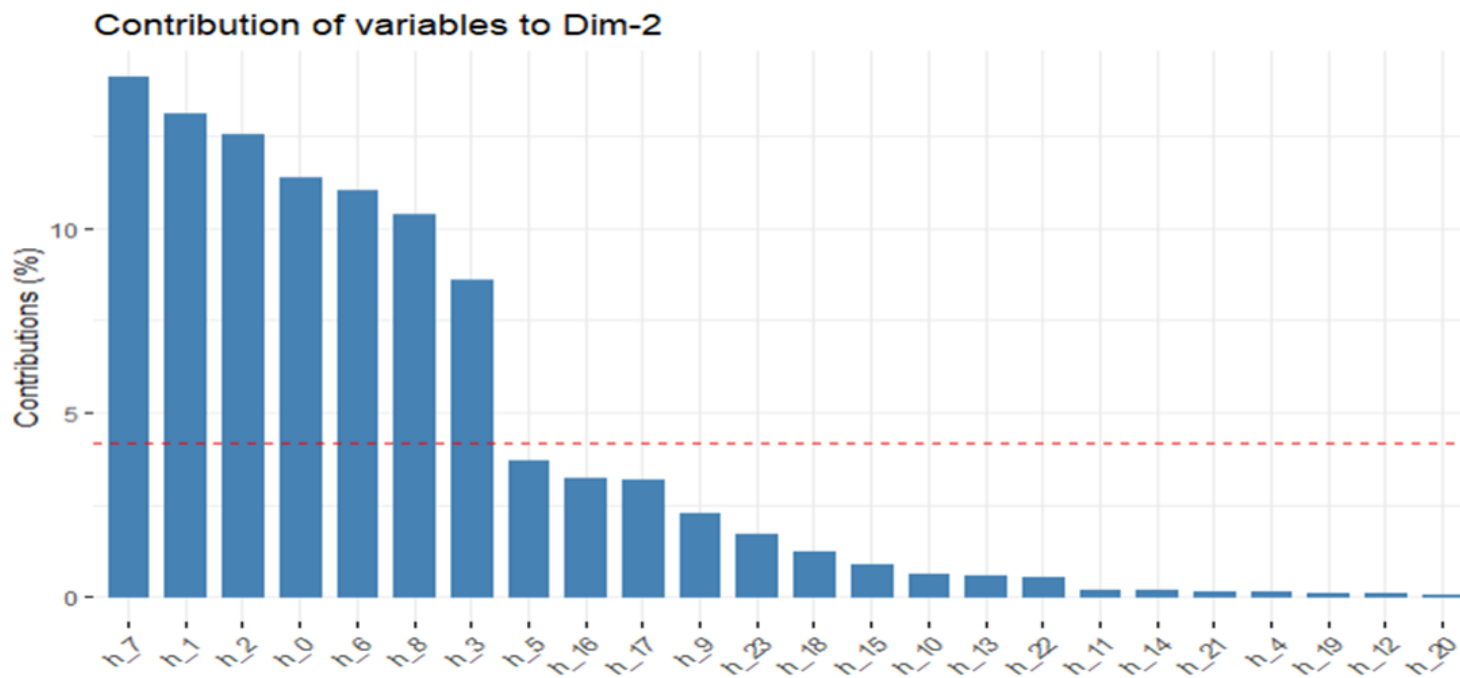
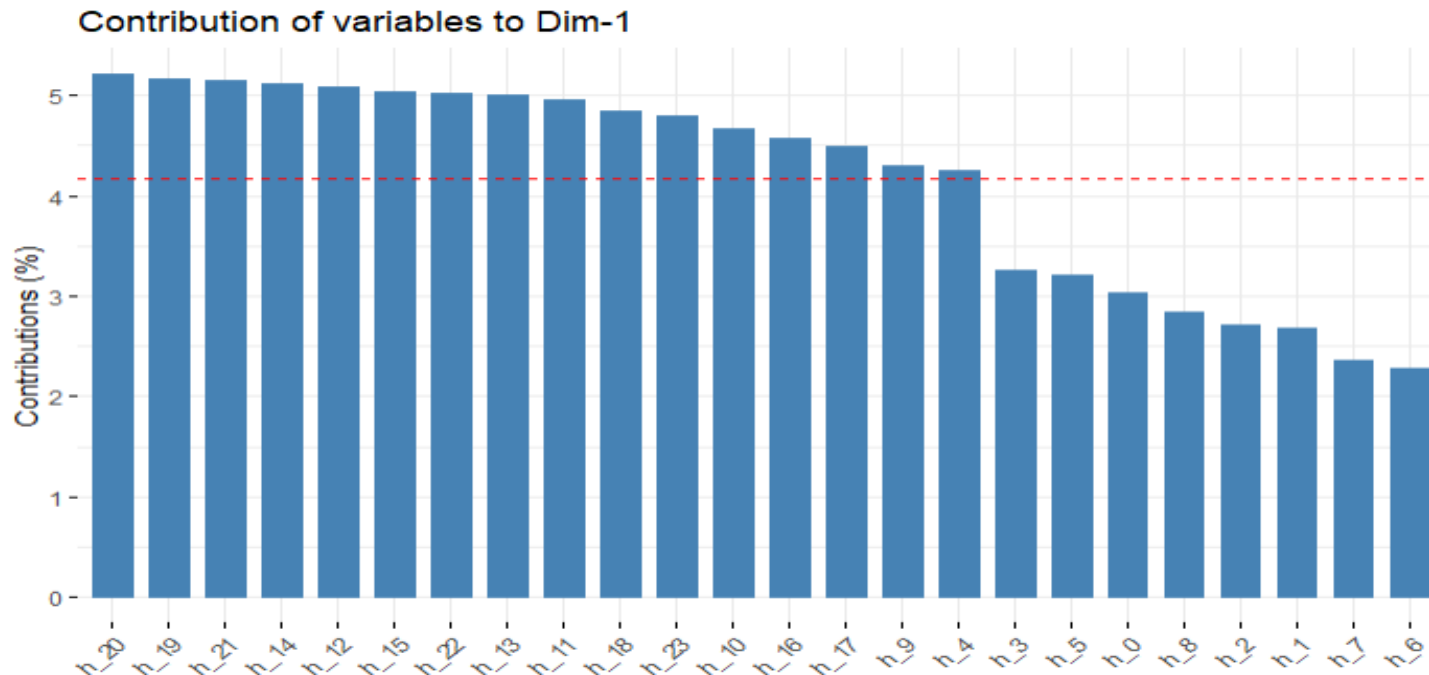


comp 1	comp 2
1.862628e+01	3.628900e+00
comp 12	comp 13
3.845143e-03	2.567018e-03
comp 23	comp 24
4.946759e-05	1.390039e-05

Les heures de la journée
(9h – 23h) sont celles
qui contribuent le plus à
la construction du
premier axe

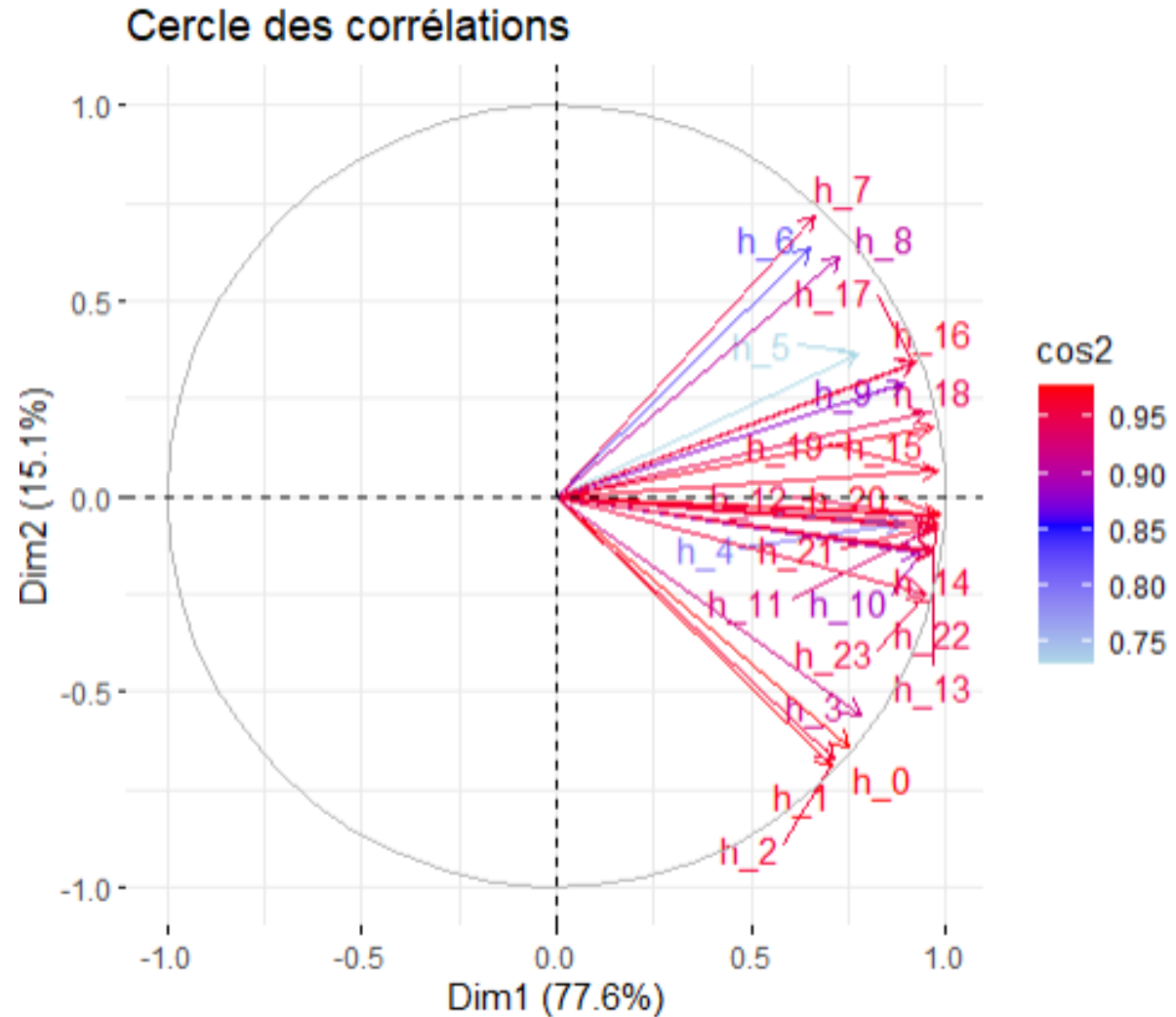
Construction des composantes principales

A l'inverse, les heures
nocturnes (0h – 8h) sont
mieux expliquées par le
second axe

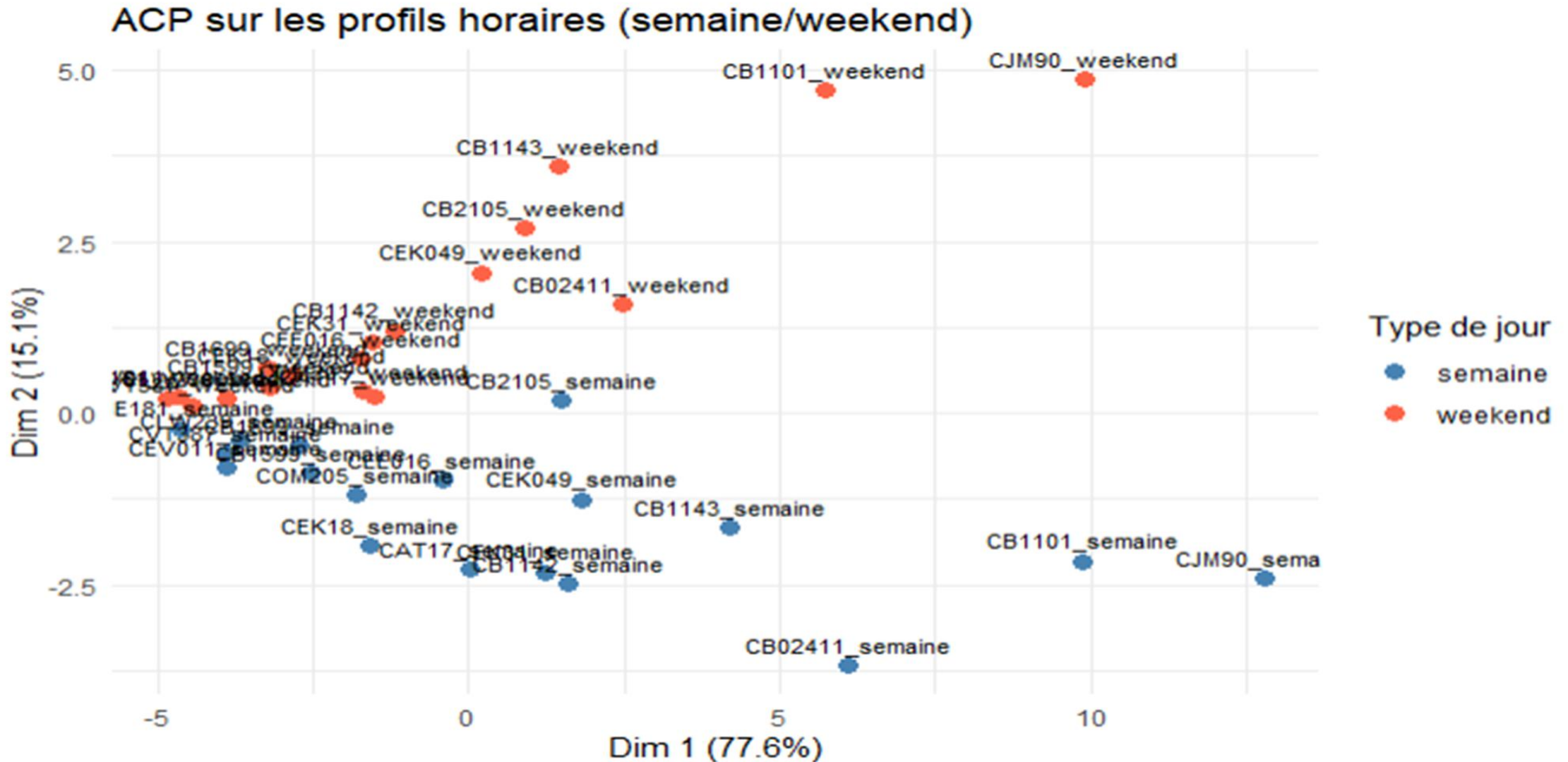


Qualité de représentation

- Toutes les heures sont très bien représentées sur les 2 premiers axes, qui expliquent près de 93% de la variance des données horaires à eux seuls.
- Les heures entre 4h et 6h sont les moins bien représentées.
- L'usage du vélo entre 0h et 3h est complètement indépendante de celle aux heures de pointe (angle de 90° entre les heures 6h-8h et 17h)

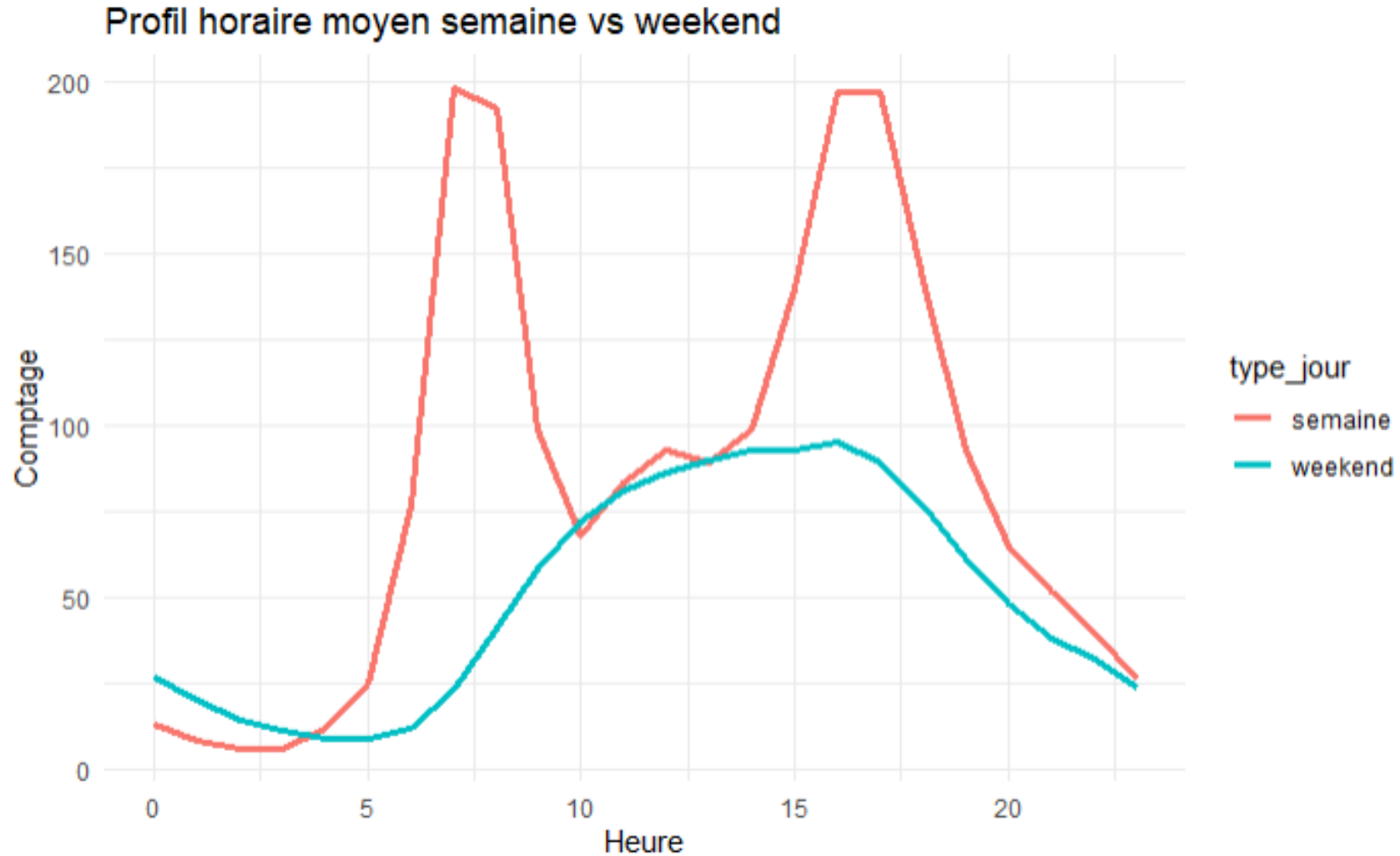


Représentation des individus (stations)



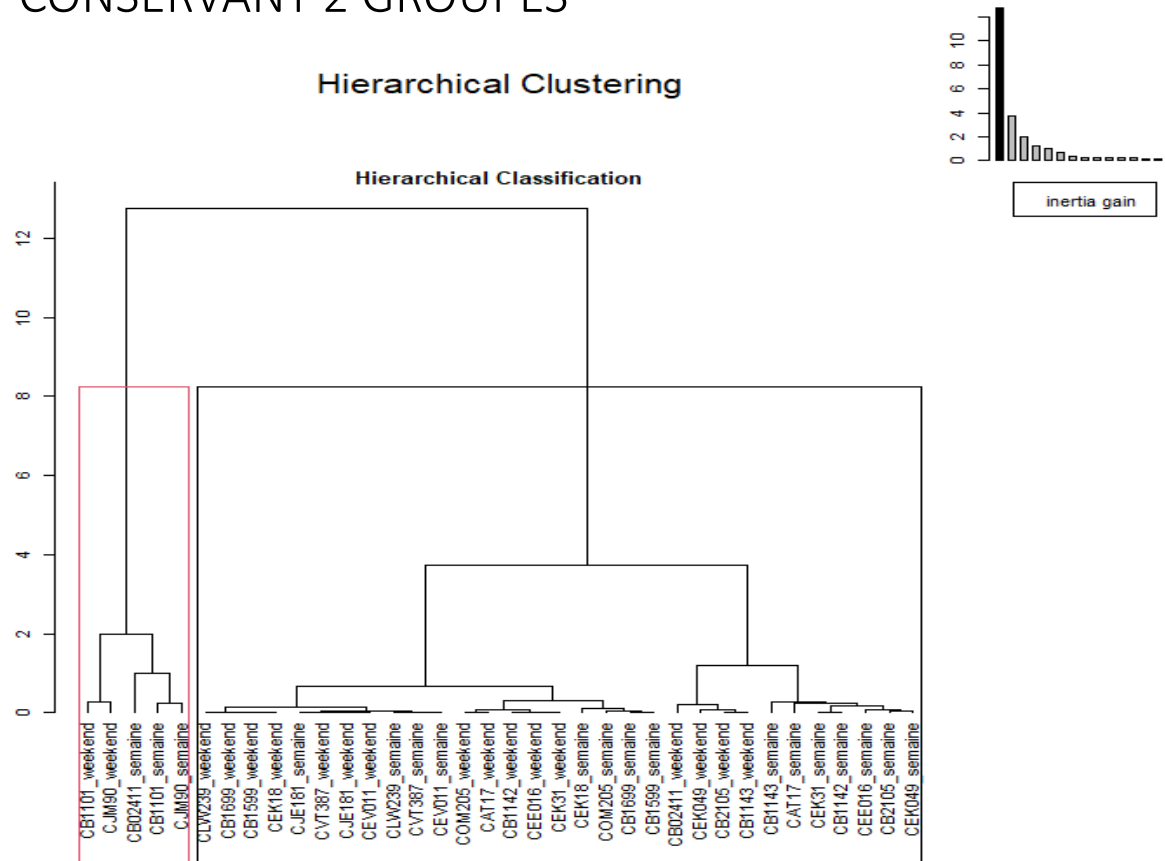
Habitudes de circulation

- Pics de circulation aux heures de pointe en semaine (rouge) et circulation plus lisse le week-end (bleu)
- Nombre de cyclistes plus élevé entre 0h et 3h le week-end par rapport à la semaine
- Comportement similaire entre 10h et 15h la semaine et le week-end

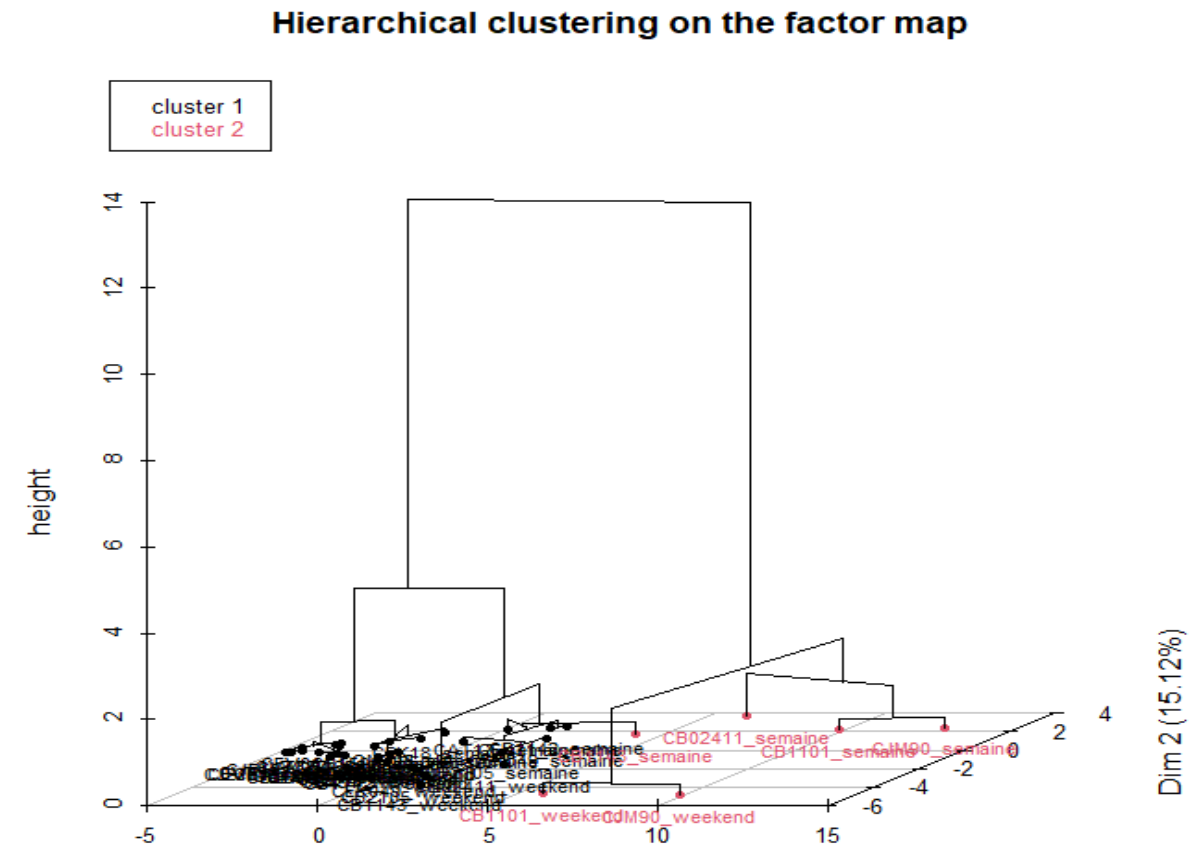


Dendrogrammes

ON MINIMISE LE GAIN D'INERTIE EN
CONSERVANT 2 GROUPES



ENCORE PLUS ÉVIDENT EN 3D (PLAN DES
FACTEURS EN X ET Y ET DISSIMILARITÉ EN Z)



Clustering (k = 2) sur les profils horaires



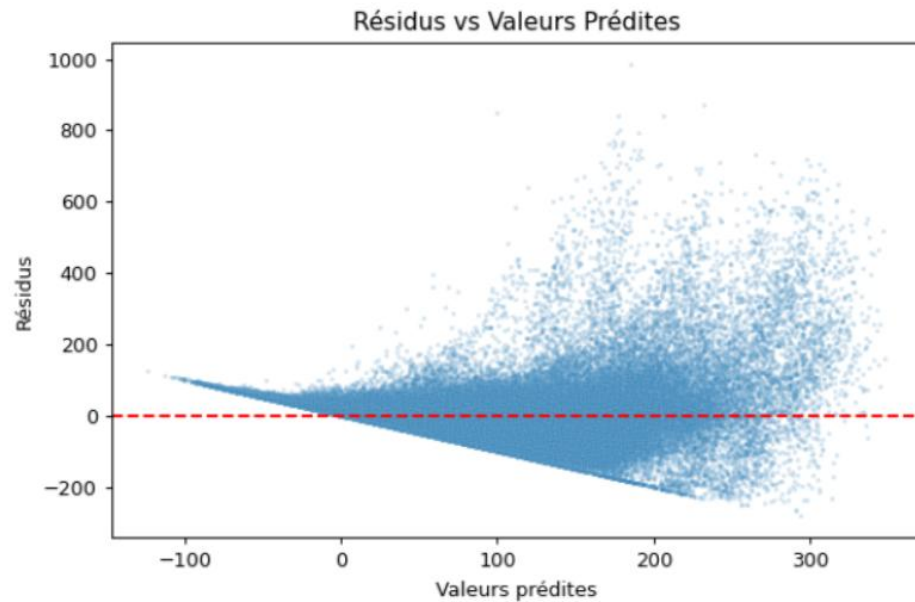
Seules les stations CB02411 et CB1143 font partie du cluster 1 pour leur composante weekend et du cluster 2 pour leur composante semaine

La première dimension correspond à l'intensité de l'utilisation et des pics de circulation aux heures de pointe

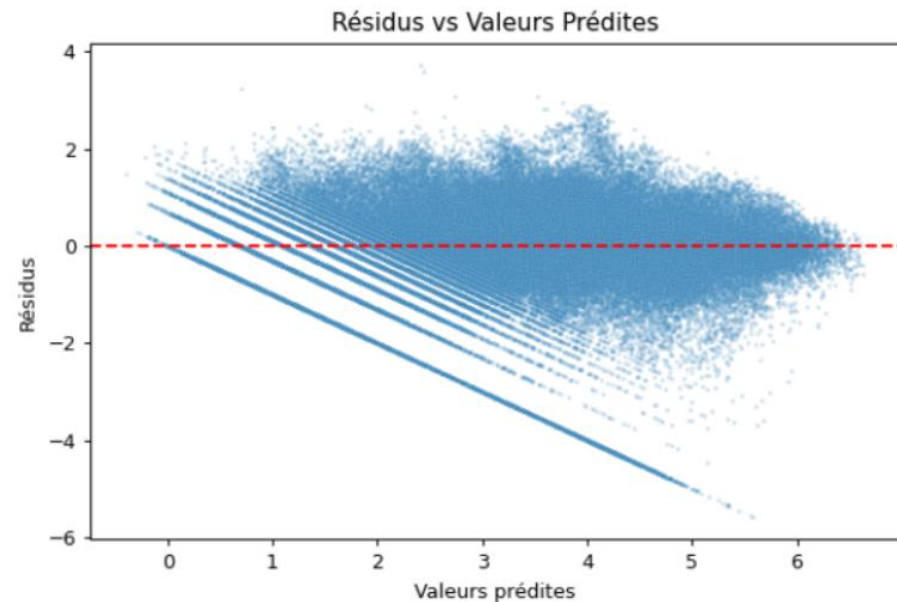
La seconde dimension correspond à l'opposition entre semaine et w-e mais aussi, entre nuit et jour

L'hypothèse d'homoscédasticité est rejetée pour les 2 modèles de régression linéaire

MODÈLE 1 (LIN-LIN)



MODÈLE 2 (LOG-LIN)

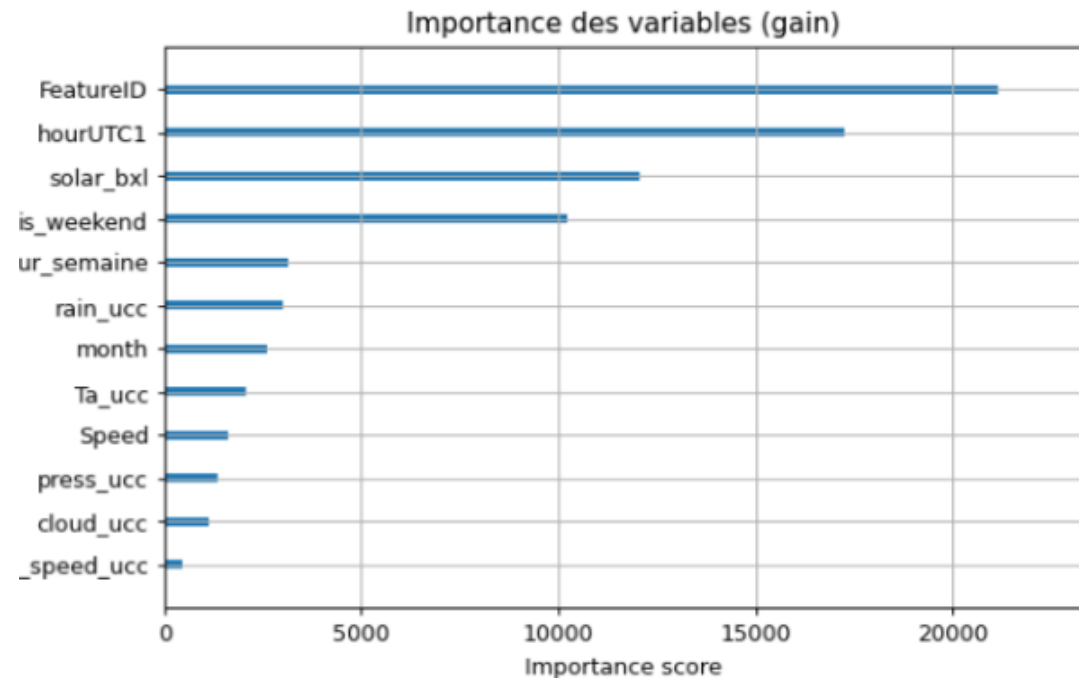


Régression linéaire

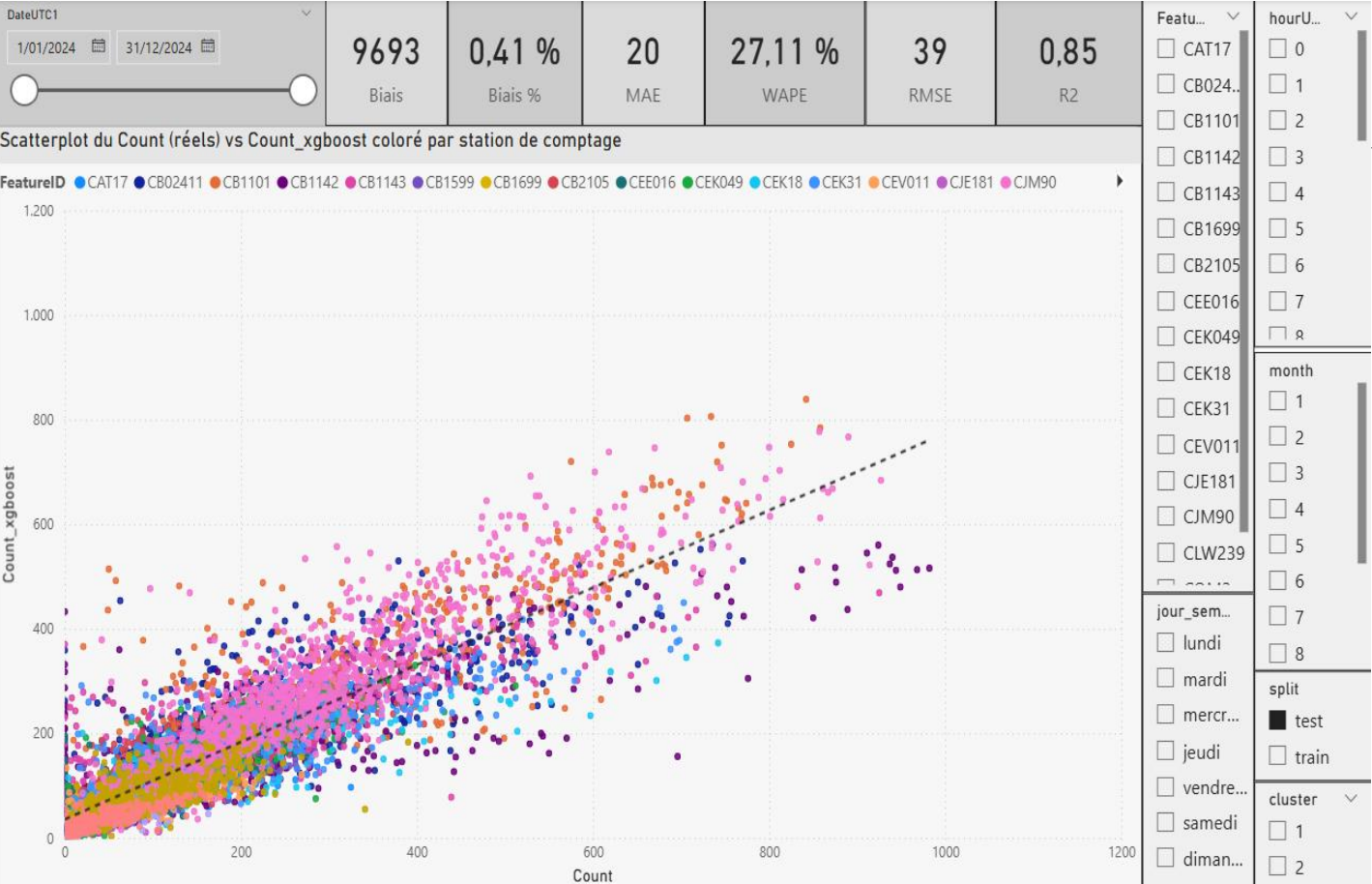
XGBoost

OLS Regression Results						
Dep. Variable:	np.loglp(Count)	R-squared:	0.715			
Model:	OLS	Adj. R-squared:	0.715			
Method:	Least Squares	F-statistic:	333.4			
Date:	Tun., 01 sept. 2025	Prob (F-statistic):	4.28e-18			
Time:	14:52:31	Log-Likelihood:	-1.8917e+05			
No. Observations:	158094	AIC:	3.785e+05			
DF Residuals:	158040	BIC:	3.790e+05			
DF Model:	53					
Covariance Type:	cluster					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.0578	0.812	-1.303	0.193	-2.649	0.534
C(hourUTC1)[T..1]	-0.3865	0.018	-21.291	0.000	-0.422	-0.351
C(hourUTC1)[T..2]	-0.6748	0.037	-18.288	0.000	-0.747	-0.602
C(hourUTC1)[T..3]	-0.7019	0.071	-9.951	0.000	-0.840	-0.564
C(hourUTC1)[T..4]	-0.3245	0.119	-2.718	0.007	-0.558	-0.090
C(hourUTC1)[T..5]	0.2007	0.154	1.300	0.194	-0.102	0.503
C(hourUTC1)[T..6]	1.0014	0.155	6.458	0.000	0.697	1.305
C(hourUTC1)[T..7]	1.8498	0.139	13.353	0.000	1.578	2.121
C(hourUTC1)[T..8]	2.0397	0.112	18.278	0.000	1.821	2.258
C(hourUTC1)[T..9]	1.6989	0.097	17.508	0.000	1.509	1.889
C(hourUTC1)[T..10]	1.5639	0.097	16.135	0.000	1.374	1.754
C(hourUTC1)[T..11]	1.7001	0.093	18.195	0.000	1.517	1.883
C(hourUTC1)[T..12]	1.8017	0.090	20.117	0.000	1.626	1.977
C(hourUTC1)[T..13]	1.7793	0.086	20.615	0.000	1.610	1.948
C(hourUTC1)[T..14]	1.8527	0.091	20.383	0.000	1.675	2.031
C(hourUTC1)[T..15]	2.0753	0.096	21.582	0.000	1.887	2.264
C(hourUTC1)[T..16]	2.2984	0.090	25.512	0.000	2.122	2.475
C(hourUTC1)[T..17]	2.2985	0.082	28.028	0.000	2.138	2.459
C(hourUTC1)[T..18]	2.0287	0.071	28.455	0.000	1.889	2.168
C(hourUTC1)[T..19]	1.6710	0.063	26.542	0.000	1.548	1.794
C(hourUTC1)[T..20]	1.3465	0.052	25.699	0.000	1.244	1.449
C(hourUTC1)[T..21]	1.1004	0.042	26.143	0.000	1.018	1.183
C(hourUTC1)[T..22]	0.8162	0.029	27.893	0.000	0.759	0.874
C(hourUTC1)[T..23]	0.4014	0.021	18.836	0.000	0.360	0.443
C(jour_semaine)[T..lundi]	0.2945	0.032	9.346	0.000	0.233	0.356
C(jour_semaine)[T..mardi]	0.3844	0.034	11.467	0.000	0.319	0.450
C(jour_semaine)[T..mercredi]	0.3828	0.032	11.850	0.000	0.319	0.446
C(jour_semaine)[T..jeudi]	0.4125	0.032	12.840	0.000	0.350	0.476
C(jour_semaine)[T..vendredi]	0.3827	0.029	13.216	0.000	0.326	0.439
C(jour_semaine)[T..samedi]	0.1155	0.015	7.735	0.000	0.086	0.145
C(featureID)[T..CB02411]	0.8093	0.009	92.584	0.000	0.792	0.826
C(featureID)[T..CB11101]	1.1481	0.014	83.965	0.000	1.121	1.175
C(featureID)[T..CB11142]	-0.3554	0.010	-36.217	0.000	-0.375	-0.336
C(featureID)[T..CB11143]	0.5601	0.002	253.522	0.000	0.556	0.564
C(featureID)[T..CB11591]	-0.5099	0.009	-54.722	0.000	-0.528	-0.492
C(featureID)[T..CB1699]	-0.4958	0.013	-37.999	0.000	-0.521	-0.470
C(featureID)[T..CB2105]	0.4558	0.003	134.451	0.000	0.449	0.462
C(featureID)[T..CB0416]	-0.0482	0.007	-6.619	0.000	-0.062	-0.034
C(featureID)[T..CBK049]	0.4165	0.005	79.517	0.000	0.406	0.427
C(featureID)[T..CEK18]	-0.2655	0.023	-11.565	0.000	-0.311	-0.221
C(featureID)[T..CEK31]	0.1842	0.010	18.550	0.000	0.165	0.204
C(featureID)[T..CBV011]	-1.0743	0.025	-42.831	0.000	-1.124	-1.025
C(featureID)[T..CJE181]	-1.4506	0.023	-63.995	0.000	-1.495	-1.406
C(featureID)[T..CJM90]	1.4001	0.000	5670.789	0.000	1.400	1.401
C(featureID)[T..CLW239]	-0.9404	0.003	-319.771	0.000	-0.946	-0.935
C(featureID)[T..COM205]	-0.2807	0.013	-22.209	0.000	-0.305	-0.256
C(featureID)[T..CVT387]	-1.1370	0.019	-61.280	0.000	-1.173	-1.101
Speed	0.0125	0.007	1.678	0.093	-0.002	0.027
Ta_ucc	0.0174	0.009	2.030	0.042	0.001	0.034
wind_speed_ucc	-0.0090	0.007	-1.370	0.171	-0.022	0.004
press_ucc	0.0028	0.001	4.070	0.000	0.001	0.004
cloud_ucc	-0.0128	0.003	-4.839	0.000	-0.018	-0.008
rain_ucc	-0.1500	0.018	-8.496	0.000	-0.185	-0.115
solar_bxl	0.0155	0.103	0.150	0.880	-0.187	0.218
Omnibus:	50273.945	Durbin-Watson:	0.402			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	307778.595			
Skew:	-1.394	Prob(JB):	0.00			
Kurtosis:	9.241	Cond. No.	1.11e-05			

MSE : 1543.90
R2 : 0.85
MAE (counts) : 20.13
Poisson deviance (moy) : 11.08
Valeurs négatives : 0 sur 158094 (0.00%)

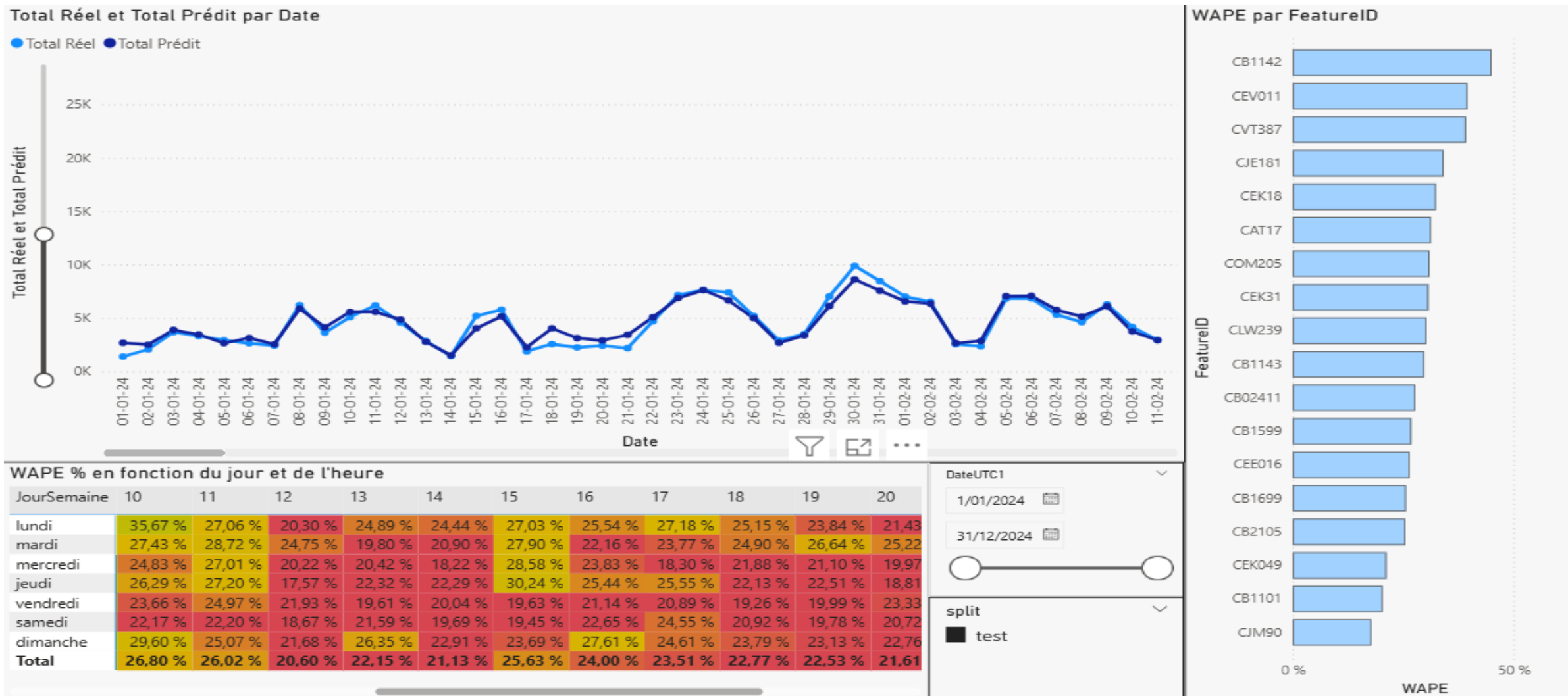


Dashboard PowerBI



Plus d'erreurs le samedi et
dimanche à 6-7h

Ensoleillement vs Nombre de cyclistes : corrélation



Meilleure capacité prédictive entre 9-23h (WAPE% plus faible)

Stations du cluster 2 mieux prédites par XGBoost

Merci !

