

Catégorie	Critères	RNPC	Détail	évaluation (entre 1-5)	Commentaire sur ce qui a été fait et possibilités d'amélioration	Ce qui a été fait par le groupe évalué que le groupe évaluateur n'a pas fait.		
Organisation	Planning	C16	Le planning est géré à l'aide d'un outil (trello...), il est créé en amont puis mis à jour régulièrement	0	voir ci-dessous	cleaning différent		
	Besoin Client	C1/C11	Le besoin client est synthétisé de manière claire (Cahier des charges - User Story)	5	voir ci-dessous			
Code	Lisibilité		Respect d'une convention syntaxique de nommage (PEP8) et nommage transparent et pertinent	5	voir ci-dessous			
	Structuration		Le code est organisé en modules et sous modules cohérents avec la logique du projet	5	voir ci-dessous			
	Commentaire		Les commentaires sont courts, utiles, et en anglais, présence de séparateurs entre les parties d'un même fichier	4	voir ci-dessous			
	factorisation		Le code ne se répète pas, les bons outils sont utilisés pour factoriser (classe, fonction, dictionnaire)	5	voir ci-dessous			
	Documentation		Le readme résume l'utilité de chaque fichier, toutes les fonctions possèdent un docstring	4	voir ci-dessous			
Conception de la base de données	Sourcing	C1	Recherche de sources de données et validation de ces sources	X	voir ci-dessous			
	MPD	C10	Un MPD est réalisé suivant les bonnes pratiques ou sont absence est justifiée	X	voir ci-dessous			
	Outil	C10/C2	Le choix de l'outil pour stocker la base de données est pertinent et justifié (SQL/no SQL ; Cloud/ moteur sql (mysql,sqlite...) / CSV)	X	voir ci-dessous			
	Qualité	C10	Les relations entre les tables sont respectées ainsi que les types des variables et les clefs primaires et secondaires	X	voir ci-dessous			
	Intégrité	C3	Le choix de la méthode d'import est justifié et l'intégralité des tables et des instances est respectée suite à l'import	X	voir ci-dessous			
	Sauvegarde	C10	Une procédure de sauvegarde a été pensée, justifiée et mise en place	X	voir ci-dessous			
	Qualification	C1	Par des représentations simples je donne les principales caractéristiques de ma base données	X	voir ci-dessous			
Data Cleaning	Sélection	C4	Toutes mais seules les données pertinentes ont été conservées dans la base de données intermédiaires	3	voir ci-dessous			
	Valeurs manquantes	C4	Les valeurs manquantes ont été repérées, traitées et justifiées	5	voir ci-dessous			
	Doublons	C4	Les doublons ont été repérés et traités	5	voir ci-dessous			
	Consistence	C4	La consistance des données a été évaluées, traitées et justifiées	4	voir ci-dessous			
	Encodage	C4	L'encode a été évalué et traité	4	voir ci-dessous			
	Valeur abérantes	C4	Les valeurs abérantes ont été repérées, traitées et justifiées	3	voir ci-dessous			
Data Viz	Outil	C1	Les outils choisis pour la réalisation des rendus et leur présentation sont pertinents (Librairie python (Seaborn/Bokeh); outil d'analyse (Tableau, Power BI...) outil de rendu (power point , Excel / Tableau/ Notebook...)	5	voir ci-dessous			
	Pertinence	C1	Les visualisations permettent d'apporter des éléments de réponse aux problèmes posés	5	voir ci-dessous			
	Qualité	C1	Les visualisations respectent les bonnes pratiques: claires, esthétiques, dotées de titre, de légende et de source	5	voir ci-dessous			
	Analyse	C1	Les visualisations sont analysées de manière succincte et pertinente dans un texte adossée à celles-ci	3	voir ci-dessous			
	Choix du modèle	C5	Le choix du ou des modèles retenus est cohérent avec le problème à traiter, les données et la capacité de calcul disponible. Ce choix est également clairement justifié	4	voir ci-dessous			

Machine learning	Préparation des données	C4	Le jeu de données a été divisé en différents sets. Les données sont en fonction du besoin, normalisées, discrétisées, encodées, échantillonnées	4	voir ci-dessous			
	évaluation du modèle	C8	Le ou les modèles sont évalués sur la base d'indicateurs pertinents. Le choix du modèle retenu et des ses hyperparamètres est justifié sur la base de cette évaluation	4	voir ci-dessous			
	Outils	C6	Les outils employés pour la modélisation sont pertinents (choix de la librairie python, utilisation d'un pipe)	4	voir ci-dessous			
points positifs								
excellente organisation du projet, code concis, commenté et factorisé, graphique coloré, lisible et pertinent (joli scatterplot), ajout de colonnes pertinentes pour le ML isolation des dates, hot encoder etc								
Commentaires				86				
nettoyage des données : il aurait fallu nettoyer les coordonnées pour cibler uniquement la zone de new york								
store_and_fwd_flag', 'passenger_count' pouvaient être utiles ou du moins analysés pour le ML								
le cleaning semble assez agressif on passe de 1 458 644 lignes à 609 803!								
data cleaning vous avez éliminé les vitesses en dessous de 10km/h, pourquoi ? vous avez également drop les distances inférieure à 1km, en centre ville ça semble juste, non ?								
vous n'avez pas répondu par des phrases, il faut lire le graphique pour trouver la réponse								
au niveau des tranches horaires, quand on lit 0-3h on comprend 3h exclus, il aurait peut être mieux fallu écrire 0-4h, 4-8h...								
c'est plus agréable de lire le temps en minutes quand tu dépasses 60s , ex dans le nb ml 306.0s =~ 5min								
sur vos graphiques de ml les points sont un peu gros du coup on ne peut pas voir certains outliers								