

TABLE OF CONTENTS



- 01
- Présentaion du projet

Présentation des données et de l'objectif

04

Conclusion

- 02
- Analyse

Nettoyage et analyse des données

- 03
- Modélisation

Split des données et entrainement du modèle







Brief du projet



La cuisine centrale de Nantes Métropole produit par jour entre 15 000 et 16 000 repas qui sont livrés dans 87 établissements. Les familles ont la possibilité d'inscrire ou non les enfants le jour même.

Ce système entraine une grande variabilité dans la fréquentation. Celle-ci est effectivement impactée par le contexte sanitaire, la composition des menus, le jour de la semaine, la période de l'année...

Cette variabilité contraint les cantines, très en amont (S-3), à estimer systématiquement le nombre de convives. Ces estimations servent à évaluer les quantités à commander aux fournisseurs. Un constat : le gaspillage alimentaire sur l'ensemble de la métropole est passée de 5% en moyenne entre 2011 et 2016 à 8% sur 2017-2019.

En tant que responsable des finances de la cuisine, vous voulez réduire ce gaspillage de surproduction pour faire des économies. Un repas coûte en matière première 3 euros.

Vous allez donc analyser les données afin de trouver les facteurs qui influencent la fréquentation des cantines et essayer de proposer une solution pour réduire le gaspillage (un modèle statistique par exemple).





Les Missions



Analyser le jeu de données afin de repérer des variables pertinentes pour la mission.

Tout au long de l'analyse, produire des visualisations afin de mieux comprendre les données.

Nettoyer le jeu de données : identifier les colonnes inutiles (variables non pertinentes) et vérifier si toutes les lignes sont exploitables

Effectuer une analyse univariée pour chaque variable intéressante, afin de synthétiser son comportement. Variez les graphiques (boxplots, histogrammes, diagrammes circulaires, nuages de points...) pour illustrer au mieux votre propos.

Confirmer ou infirmer les hypothèses à l'aide d'une analyse multivariée (plusieurs variables). Le plus important sera de déterminer les facteurs qui jouent le plus sur la fréquentation.

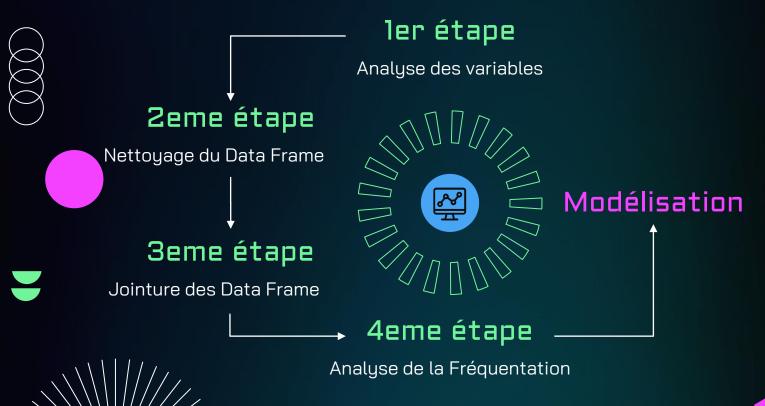
Identifier des arguments justifiant la faisabilité d'une fiabilisation des commandes en utilisant un modèle statistique. Est-ce possible d'après vous ?







Etapes de l'analyse

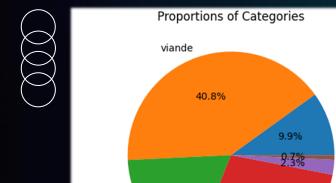


Proportion des catégories (menu)

porc

other

frites



18.1%

poisson

28.1%



Dans cette analyse concernant le menu, nous avons une représentions en forme de tarte qui nous permet de voir de façon global la constitution du menu.

Cette étape nous a permis de nous comprendre le jeu de données et les variables qui ont un impact sur la fréquentation.







Dans l'analyse du jeu de données de la fréquentation, certaines variables n'étaient pas pertinentes dans le sens où elles n'avaient pas d'impactes significatif sur la fréquentation.

Grèves, maladies : imprévisibles.

Vacances, Mercredis : Pas de fréquentation des cantines.

Cette étape nous a permis d'avoir un jeu de données propres, pour pouvoir passer à la jointure des DataFrames, « Menu » et « Fréquentions »

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1188 entries, 0 to 1187
Data columns (total 23 columns):
     Column
                      Non-Null Count Dtype
     Unnamed: 0
                      1188 non-null
                                      int64
                      1188 non-null
                                      object
     annee scolaire
                    1188 non-null
                                      object
                      1188 non-null
                                      object
     semaine
                      1188 non-null
                                      int64
                      1188 non-null
                                      object
     greves
                      1188 non-null
                                      int64
     ferie
                      1188 non-null
                                      int64
     veille ferie
                      1188 non-null
                                      int64
     retour ferie
                      1188 non-null
                                      int64
     vacances
                      1188 non-null
                                      int64
     retour vacances 1188 non-null
                                      int64
     veille vacances 1188 non-null
                                      int64
     fete musulmane 1188 non-null
                                      int64
     ramadan
                      1188 non-null
                                      int64
     fete chretienne 1188 non-null
                                      int64
     fete juive
                                      int64
                      1188 non-null
     inc grippe
                      1188 non-null
                                      int64
     inc gastro
                      1188 non-null
                                      int64
     inc varicelle
                      1188 non-null
                                      int64
     prevision
                      1188 non-null
                                      int64
    reel
                      1188 non-null
                                      int64
    effectif
                      1188 non-null
                                      int64
dtypes: int64(19), object(4)
memory usage: 213.6+ KB
```







Jointure des DataFrames

df_final.head()																				
																				Python
	date	annee_scolaire	jour	semaine	mois	greves	ferie	veille_ferie	retour_ferie	fete_musulmane	menu	porc	viande	poisson	bio	noel	frites	an_chinois	other	_merge
	2011- 01-03	2010-2011	Lundi	1.0	Janvier	0.0	0.0	0.0	0.0	0.0	['Pamplemousse', 'Hachis Parmentier', 'Salade									both
	2011- 01-04	2010-2011	Mardi	1.0	Janvier	0.0	0.0	0.0	0.0	0.0	['Crêpe au fromage', 'emincé de volaille à la									both
	2011- 01-06	2010-2011	Jeudi	1.0	Janvier	0.0	0.0	0.0	0.0	0.0	['thon', 'Mâche au thon', 'Croziflette', 'des									both
	2011- 01-07	2010-2011	Vendredi	1.0	Janvier	0.0	0.0	0.0	0.0	0.0	['Choux blancs mimolette', 'milanette en des',									both
4	2011- 01-10	2010-2011	Lundi	2.0	Janvier	0.0	0.0	0.0	0.0	0.0	['Endives / demi-\x9cuf', 'paella de la mer',									both

Dans cette étape nous avons procédé à la jointure des DataFrames« Menu » ,et « Fréquentation ». Cela nous a permis de faire une étude générale sur la fréquentation de la cantine.

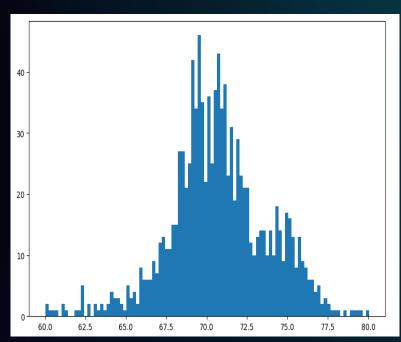






Analyse de la fréquentation





Nous avons ici un histogramme qui nous permet d'avoir une représentation du taux de fréquentation par rapport à l'effectif réel.

Ainsi le taux de fréquentation moyen est d'environ 67.85%.

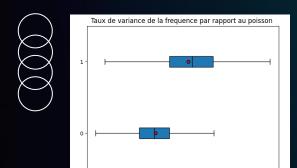
L'écart type est assez faible. Il est d'environ 12.48. Cela veut dire que les différentes valeurs de fréquentation sont relativement proches de la moyenne.

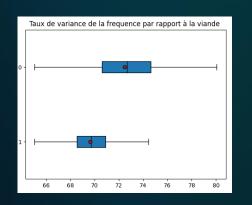


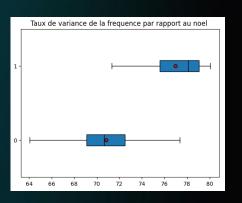




Analyse de la fréquentation et des menus







Grace ces box plot nous avons une vue générale sur les variables qui influence le plus la fréquentation de la cantine. Ces variables sont le poisson, la viande et le repas de noël.

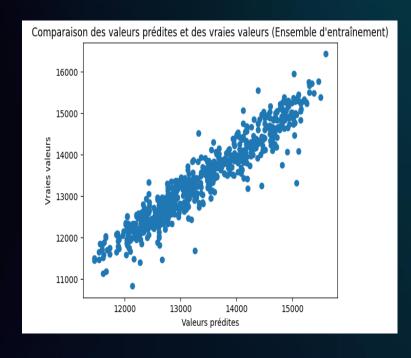
- Poisson : Il y a en moyenne plus de monde quand il y a du poisson.
- Viande : Le deuxième graphique montre qu'en général il y a plus de monde quand il n'y a pas de viande
- Repas de noel : Il y a beaucoup plus de personnes quand il y a le repas de noël.





Modélisation de l'entrainement





Dans cette dernière étape du projet qui concerne la modélisation. Nous avons utilisé la régression linéaire pour pouvoir entrainer notre modèle à prédire et à devenir plus précis.

lci nous avons la partie entrainement pour perfectionner le modèle, en vue d'avoir une mse faible.

r2

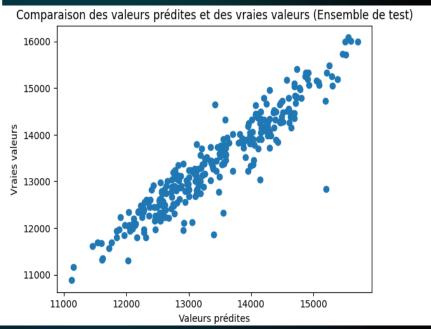
0.9010237473624512

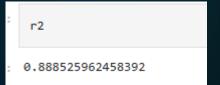


Modélisation du test

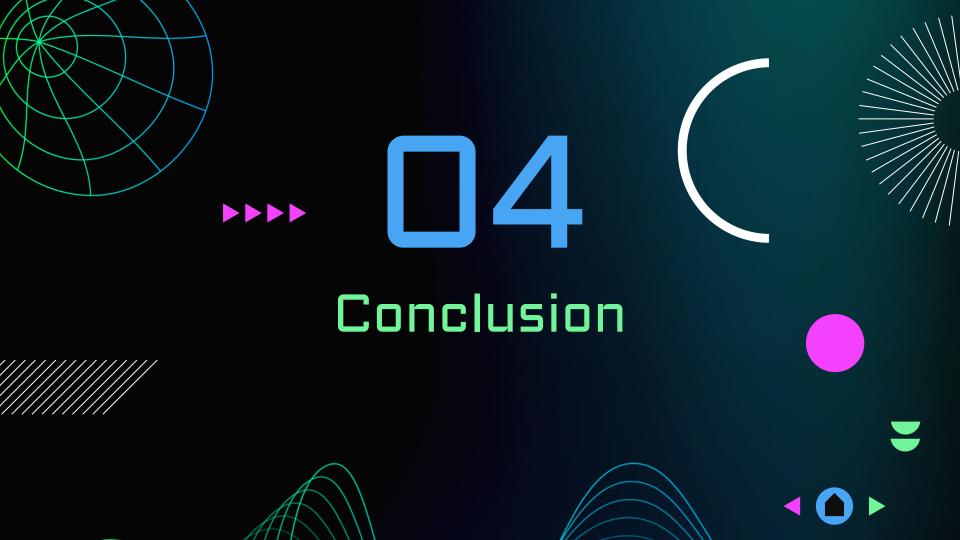
Dans cette étape de test de notre modèle nous avons un R square proche de 1, ce qui signifie que notre explique 88% de la variance des données. Cela est considéré comme une bonne performance.

Notre modèle a donc une bonne capacité à prédire la variable cible (réel) qui représente le nombre de personnes qui fréquente la cantine en fonction des variables indépendantes













En conclusion, le projet de gestion de la fréquentation de la cantine scolaire et des prévisions pour limiter le gaspillage alimentaire a été réalisé avec succès en utilisant la bibliothèque Pandas. Deux DataFrames ont été utilisés, l'un contenant les données de fréquentation et l'autre contenant les menus. Une jointure des DataFrames a été effectuée pour obtenir un jeu de données final cohérent, sur lequel des analyses ont été menées.

Ainsi nous avons un modèle de régression linéaire pour pouvoir prédire et limiter le gaspillage alimentaire dans les cantines.

Il est recommandé de considérer ces résultats comme une première approche et en collaboration avec des experts en data science nous pourrions avoir de meilleurs résultats et ainsi limiter davantage le gaspillage alimentaire.







Notre équipe













Massinissa BELHARET



Adrien FORMOSO





Rapport journalier >>>>



1er jour	2ème jour	3ème jour	4ème jour			
Compréhension du projet, et distribution des tâches	Nettoyage et jointure des DataFrames	Analyse et création des boxplots	Calcul de la corrélation			
5ème jour	6ème jour	7ème jour				
Modélisation	Création du Power Point	Revu d'ensemble du projet et organisation du notebook				









MERCI!



