École Supérieure Privée de Management de Tunis
Esprit School of Business

Rapport de Stage d'Été

Réalisé par : Balkis Benslimane

« Analyse et Visualisation des Avis Clients »

Maître de stage : Abdelmonim KLAI

Année universitaire 2024/2025

Table des matières

Introduction

Dans un contexte économique en constante évolution, les entreprises cherchent à améliorer leur performance et leur compétitivité en plaçant la donnée au cœur de leurs décisions stratégiques. Les retours clients, en particulier, représentent une source d'information précieuse permettant de mieux comprendre les attentes, détecter les insatisfactions et orienter les actions d'amélioration.

C'est dans ce cadre que j'ai effectué mon stage d'été au sein de l'opérateur téléphonique **Ooredoo** Tunisie, l'un des leaders du secteur des télécommunications. Ce stage m'a permis de contribuer à un projet axé sur la collecte, le nettoyage, l'analyse et la visualisation des avis clients.

Mon objectif principal était de transformer une base de données brute en un outil d'aide à la décision à travers la création de tableaux de bord interactifs et d'une analyse sémantique par Word Cloud, en utilisant le langage Python.

Ce rapport vise à retracer les différentes étapes de cette expérience, depuis la découverte de l'environnement économique d'Ooredoo, jusqu'à la mise en œuvre technique des traitements de données, tout en mettant en lumière les compétences professionnelles et personnelles que j'ai pu consolider au cours de ce stage.

I. L'environnement économique du stage chez Ooredoo Tunisie

1. Le secteur : Télécommunications

A. Présentation du secteur

Le secteur des télécommunications joue un rôle clé dans le développement économique et social des pays. Il assure la transmission rapide de l'information, favorise la connectivité, soutient la croissance numérique et permet le développement des services en ligne dans des domaines aussi variés que l'éducation, la santé, le commerce ou encore l'administration publique.

Avec l'évolution des technologies de l'information et de la communication (TIC), le secteur des télécommunications ne cesse de se transformer, passant des simples appels vocaux aux services internet haut débit, à la 4G, à la fibre optique et aujourd'hui à l'anticipation de la 5G. Cette mutation impose aux opérateurs de constamment innover pour répondre à la demande croissante en connectivité et services numériques.

B. Le secteur économique en Tunisie

En Tunisie, le secteur des télécommunications est l'un des piliers de l'économie numérique. Il est encadré par l'Instance Nationale des Télécommunications (INT) et soutenu par des politiques publiques visant à moderniser l'infrastructure numérique du pays.

Trois grands opérateurs se partagent le marché tunisien : Ooredoo Tunisie, Tunisie Télécom et Orange Tunisie. La concurrence dans ce secteur est intense, poussant les acteurs à améliorer sans cesse la qualité de service, à proposer des offres innovantes et à investir dans les infrastructures de réseau.

Ce secteur contribue de manière significative au PIB national et représente un levier important pour l'employabilité des jeunes diplômés dans des domaines tels que l'ingénierie, la data science, le marketing digital, ou encore le service client.

2. L'entreprise par rapport au secteur

A. Historique de Ooredoo Tunisie

Ooredoo Tunisie, anciennement connue sous le nom de Tunisiana, a été fondée en 2002. Elle est devenue rapidement un acteur majeur des télécommunications en Tunisie. En 2014, l'entreprise a adopté la marque Ooredoo, rejoignant ainsi le groupe international Ooredoo Group, basé au Qatar et présent dans plusieurs pays du Moyen-Orient, d'Afrique du Nord et d'Asie.

Depuis sa création, Ooredoo Tunisie s'est distinguée par ses offres commerciales accessibles, ses innovations technologiques et sa stratégie orientée vers la satisfaction client. Elle a été pionnière dans plusieurs domaines, notamment dans le déploiement de la 4G et l'élargissement de l'accès à l'Internet mobile.

B. Ooredoo aujourd’hui et perspectives

Aujourd’hui, Ooredoo Tunisie est l’un des leaders du marché national avec des millions d’abonnés mobiles et Internet. L’entreprise offre une large gamme de services allant de la téléphonie mobile à l’Internet haut débit, en passant par des services destinés aux entreprises (B2B) et aux particuliers (B2C).

Elle mise sur l’innovation, la transformation digitale et l’écoute active de sa clientèle pour renforcer sa position. Ooredoo investit également dans l’intelligence artificielle, le Big Data et les solutions cloud pour améliorer l’expérience utilisateur et optimiser ses processus internes.

À l’avenir, Ooredoo prévoit de développer la 5G, de renforcer son engagement envers le développement durable, et de soutenir l’inclusion numérique, en particulier dans les régions défavorisées du pays.

II. Présentation du travail effectué

1. Description du sujet

Durant ce stage, j'ai travaillé sur l'analyse des **avis clients collectés par Ooredoo Tunisie**, dans une démarche orientée **amélioration de l'expérience utilisateur et prévention du churn** (désabonnement). Le cœur du projet repose sur le traitement d'un jeu de données client, composé à la fois d'informations démographiques, contractuelles et textuelles, permettant une étude complète du comportement des abonnés.

L'objectif était de **nettoyer, enrichir et exploiter ces données** pour :

- Mieux comprendre les attentes des clients,
- Identifier les types de problèmes les plus fréquents,
- Évaluer le niveau de satisfaction à travers le **score de sentiment** et les **émotions détectées**,
- Anticiper les risques de désabonnement à travers une **probabilité de churn**,
- Fournir des visualisations explicites et des analyses textuelles (Word Cloud).

▪ Structure du dataset initial

Le dataset initial contenait **1 488 enregistrements** répartis sur **20 colonnes**, dont plusieurs présentaient un taux élevé de valeurs manquantes. Voici un aperçu de sa structure :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1488 entries, 0 to 1487
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Numéro Client         200 non-null    float64
 1   Age                   200 non-null    object
 2   Genre                 900 non-null    object
 3   Date de retour        749 non-null    datetime64[ns]
 4   Langue                797 non-null    object
 5   Avis Client           1486 non-null   object
 6   Note                  800 non-null    object
 7   Type de problème      251 non-null    object
 8   Localisation          201 non-null    object
 9   Score de sentiment    200 non-null    object
 10  Émotion détectée      200 non-null    object
 11  Probabilité de churn  200 non-null    float64
 12  Résumé de l'avis      198 non-null    object
 13  Connection Type       200 non-null    object
 14  Customer Type         200 non-null    object
 15  Customer Tenure       200 non-null    object
 16  Tarif nom (Forfait)   200 non-null    object
 17  Segment Client        200 non-null    object
 18  Unnamed: 18           0 non-null      float64
 19  Unnamed: 19           0 non-null      float64
dtypes: datetime64[ns](1), float64(4), object(15)
memory usage: 232.6+ KB
```

- Un aperçu du **nombre de valeurs manquantes** montre l'ampleur du travail de nettoyage nécessaire :

```
Numéro Client          1288
Age                    1288
Genre                   588
Date de retour          739
Langue                  691
Avis Client               2
Note                    688
Type de problème       1237
Localisation           1287
Score de sentiment     1288
Émotion détectée       1288
Probabilité de churn   1288
Résumé de l'avis       1290
Connection Type        1288
Customer Type          1288
Customer Tenure        1288
Tarif nom (Forfait)    1288
Segment Client         1288
Unnamed: 18            1488
Unnamed: 19            1488
dtype: int64
```

2. Nettoyage et préparation du dataset

Mon travail s'est articulé autour de plusieurs tâches successives de traitement de données :

a. Nettoyage des colonnes superflues

Les colonnes Unnamed: 18 et Unnamed: 19 étaient entièrement vides. Elles ont donc été supprimées pour alléger le dataset et faciliter l'analyse.

b. Création d'une colonne identifiant unique

Pour remplacer la colonne Numéro Client trop incomplète, j'ai généré une nouvelle colonne id_client avec une valeur auto-incrémentée, commençant à 216000. Cela a permis d'assurer l'unicité des enregistrements.

c. Ajustement des colonnes d'identification client

Pour garantir un identifiant unique et exploitable pour chaque client, j'ai procédé à une réorganisation des colonnes liées aux numéros clients.

- Premièrement, une nouvelle colonne a été générée avec un numéro aléatoire commençant par le chiffre 2 suivi de 7 chiffres aléatoires, afin d'assurer une base numérique suffisamment large et unique.
- Ensuite, une colonne « numéro final » a été créée en concaténant le préfixe fixe 216 (code régional ou spécifique à l'entreprise) avec ce numéro aléatoire.

- Enfin, une colonne définitive id_clt_final a été construite : elle conserve le numéro existant dans la colonne « Numéro Client » si celui-ci est présent, sinon elle prend la valeur du numéro final généré.

d. Traitement de la colonne Date de retour

Les valeurs manquantes dans la colonne Date de retour ont été remplacées par la date actuelle au moment du traitement, afin de conserver une structure cohérente pour l'analyse temporelle.

e. Enrichissement de la colonne Type de problème

La colonne Type de problème présentait une forte proportion de valeurs manquantes (1237/1488). J'ai donc appliqué un remplissage aléatoire contrôlé à partir d'une liste prédéfinie de types de problèmes. Voici l'évolution avant/après du contenu de cette colonne :

Avant traitement :

```
Type de problème
nan                     1237
Service client            86
Stabilité                 52
Couverture                38
Débit                     36
Autre                     30
Problème Application       6
Pas de problème            3
Name: count, dtype: int64
```

Après traitement :

```
Type de problème
Service client          506
Stabilité               312
Couverture              223
Débit                   212
Autre                   177
Problème Application     38
Pas de problème          20
Name: count, dtype: int64
```

Cette étape a permis de transformer une colonne inexploitée en une **variable catégorielle riche en information**, prête pour les visualisations et l'analyse croisée.

f. Remplissage de la colonne Score de sentiment

La colonne Score de sentiment, initialement très incomplète, contenait une majorité de valeurs manquantes. Voici la répartition avant traitement :

```
📊 Valeurs les plus fréquentes dans la colonne : Score de sentiment
Score de sentiment
NaN        1288
Négatif     127
Positif      41
Neutre       32
Name: count, dtype: int64
```

Pour rendre cette variable exploitable dans le cadre d'une analyse de satisfaction client, j'ai procédé à un **remplissage aléatoire contrôlé**, en respectant une répartition cohérente avec les types de retours clients observés. La répartition **après traitement** est la suivante :

```
Score de sentiment
Négatif    961
Positif    291
Neutre     236
Name: count, dtype: int64
```

Ce remplissage a permis de simuler une situation réaliste pour la suite de l'analyse émotionnelle et comportementale.

g. Génération de la colonne *Émotion détectée* en lien avec le sentiment

Après le remplissage de la colonne Score de sentiment, j'ai créé une nouvelle variable appelée Émotion détectée. Celle-ci a été générée **automatiquement en fonction du score de sentiment** selon une logique simple :

Score de sentiment	Émotion détectée
Positif	Joie
Négatif	Colère
Neutre	Légère inquiétude

Ce traitement permet de **qualifier le retour client avec une émotion humaine**, ce qui est très utile dans le cadre d'une analyse qualitative des avis. Cette étape s'inscrit dans une démarche de **traitement du langage naturel (NLP)** orientée marketing relationnel.

h. Remplissage de la colonne Connection Type, Genre, Langue par la valeur la plus fréquente (mode)

1. **La colonne Connection Type** : contient des informations sur le mode de connexion du client (**'Prepaid ' , 'Paid' , 'Hybrid'**). Pour les valeurs manquantes, j'ai appliqué une **méthode simple et robuste** : le remplissage par le **mode**, c'est-à-dire la catégorie la plus fréquente dans la colonne.

2. **La colonne Genre (Sexe)** : La colonne **Genre** (Masculin / Féminin) a été complétée par la **valeur la plus fréquente (mode)**. Ce choix permet de préserver la répartition générale observée tout en évitant les biais introduits par une distribution aléatoire.

3. **La colonne Langue** : Les valeurs manquantes dans la colonne **Langue** ont été imputées également par la **langue la plus fréquente** dans le dataset, souvent **l'arabe** ou **le français**, selon les retours clients majoritaires.

i. Remplissage de la colonne *Customer Type* par la valeur dominante

Pour la colonne Customer Type, qui distingue les clients **B2B (professionnels)** des **B2C (particuliers)**, j'ai également utilisé la méthode du mode pour traiter les valeurs manquantes. L'objectif était aussi de respecter la contrainte métier : **les clients B2C doivent être plus nombreux que les B2B**, conformément à la réalité du marché.

j. Remplissage de la colonne Localisation avec les 24 gouvernorats

Pour traiter les valeurs manquantes dans la colonne **Localisation**, j'ai utilisé une approche de remplissage aléatoire à partir d'une liste complète des **24 gouvernorats tunisiens**.
Chaque valeur manquante a été remplacée par un gouvernorat choisi de façon aléatoire, garantissant ainsi une répartition réaliste sur le territoire national. De plus, toutes les localisations ont été converties en majuscules pour **uniformiser les données**.

Après avoir rempli la colonne *Localisation*, j'ai enrichi le jeu de données avec deux nouvelles colonnes : **Latitude** et **Longitude**, en associant à chaque gouvernorat ses coordonnées géographiques correspondantes. Cela permet par la suite de **cartographier les données.**

	Localisation	Latitude	Longitude
0	SOUSSE	35.8256	10.6084
1	TUNIS	36.8065	10.1815
2	TUNIS	36.8065	10.1815
3	SOUSSE	35.8256	10.6084
4	NABEUL	36.4519	10.7363

k. Traitement de la colonne Âge

Dans le cadre du nettoyage des données, la colonne « Âge » présentait une grande hétérogénéité au niveau des tranches d'âge, avec des formats variés tels que "16-20", "25-35", ou encore "50+". Pour homogénéiser ces valeurs, une fonction a été mise en place afin de regrouper les âges similaires dans des catégories standardisées : "15-25", "25-35", "35-45", "45-55" et "55+". Cette étape a permis de faciliter l'analyse statistique ultérieure.

Ensuite, les valeurs manquantes ont été comblées en utilisant la modalité la plus fréquente, assurant ainsi une cohérence dans la distribution des âges.

Toutefois, on a observé une surreprésentation de la catégorie "55+", ce qui pouvait biaiser l'analyse. Pour remédier à cela, 60 % des individus appartenant à cette tranche ont été sélectionnés aléatoirement, puis redistribués vers des tranches plus jeunes selon des proportions ciblées : 40 % vers "15-25", 30 % vers "25-35", 20 % vers "35-45" et 10 % vers "45-55". Ce rééquilibrage a permis d'obtenir une répartition plus réaliste et représentative des différentes tranches d'âge dans le jeu de données final.

l. Remplissage de la colonne « Probabilité de churn »

Afin de compléter les valeurs manquantes dans la colonne *Probabilité de churn*, nous avons exploité l'information issue du *Score de sentiment* associé à chaque avis client. Tout d'abord, cette dernière colonne a été nettoyée en supprimant les espaces superflus et en harmonisant la casse (mise en minuscules). Ensuite, une fonction a été définie pour générer une probabilité de churn basée sur la tonalité du sentiment :

- **Sentiment négatif** → probabilité élevée de churn (entre 70 % et 100 %),
- **Sentiment neutre** → probabilité moyenne (entre 40 % et 60 %),
- **Sentiment positif** → faible probabilité (entre 0 % et 30 %).

Ce remplissage a été appliqué uniquement aux lignes où la valeur de la colonne *Probabilité de churn* était manquante, afin de conserver les données déjà existantes. Cette approche probabiliste permet d'enrichir le jeu de données de manière cohérente avec la logique métier : plus un avis est négatif, plus le risque de départ du client est élevé.

3. Tableau de bord

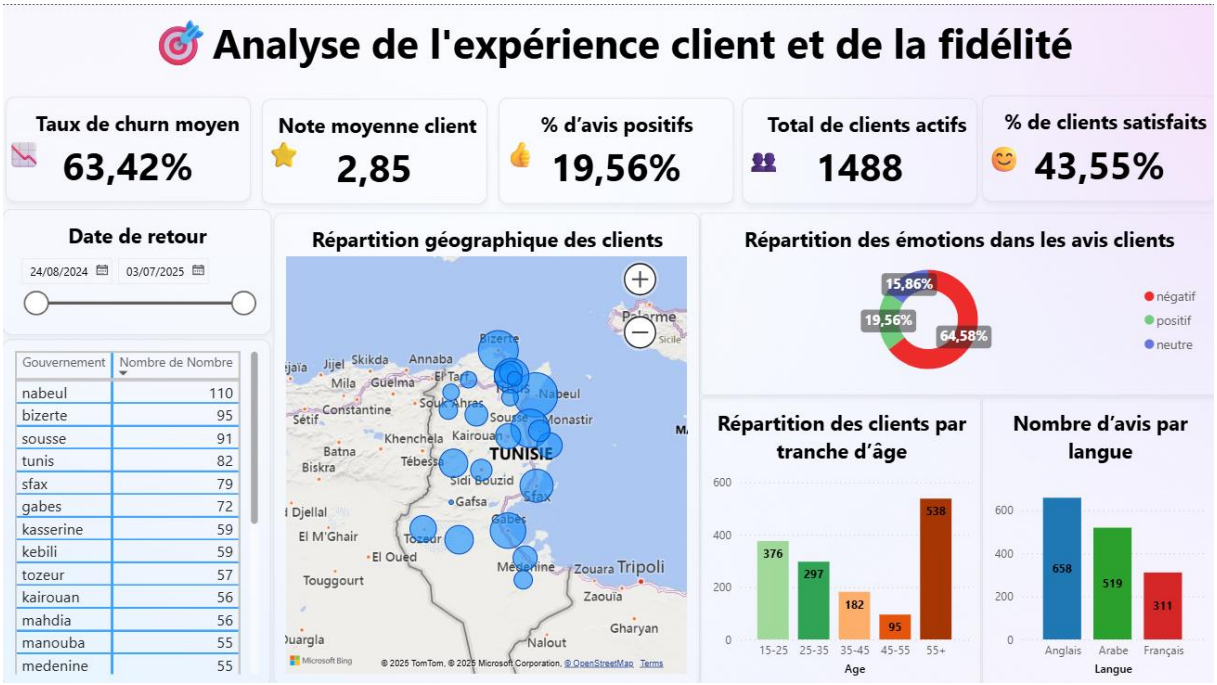
Afin de suivre efficacement l'activité des clients et d'évaluer leur comportement, plusieurs indicateurs clés ont été développés dans le tableau de bord. Voici les principaux :

KPI	Définition	Objectif	Formule / Méthode de calcul
Taux de churn moyen	Pourcentage moyen des clients ayant quitté la base	Évaluer la fidélité client et anticiper la perte	Moyenne(Taux_Churn)
Note moyenne client	Moyenne des notes données par les clients sur leurs expériences	Mesurer la satisfaction générale	AVERAGE([Note_Client])
% d'avis positifs	Pourcentage d'avis exprimant une opinion favorable (ex. note ≥ 4)	Identifier la part de clients satisfaits	DIVIDE(COUNTROWS(Filtre_Pos), COUNTROWS(Total_Avis), 0)
Total de clients actifs	Nombre de clients ayant un statut actif	Suivre la taille actuelle de la base client	COUNTROWS((Client)
% de clients satisfaits	Pourcentage de clients ayant donné une note satisfaisante (ex. note ≥ 3)	Suivre le niveau global de satisfaction	DIVIDE(Nombre_Clients_Satisfaits, Nombre_Total_Clients, 0) par DAX

- **Filtre temporel (Slicer)** :
  Un **slicer** a été ajouté pour permettre la **filtration des données par date de retour**. Cela permet d'analyser les indicateurs selon des périodes spécifiques.
- **Graphique en anneau : Score de sentiment**
  - **Titre** : *Répartition des émotions dans les avis clients*
  - Visualise la distribution des sentiments (positif, négatif, neutre) extraits des avis.
  - Permet de mesurer rapidement l'humeur générale des clients à travers leurs retours.
- **Graphique à barres : Répartition par tranche d'âge**
  - Montre la distribution des clients selon différentes **tranches d'âge**.
  - Aide à comprendre quel segment d'âge est le plus représenté parmi les avis ou les retours clients.
- **Graphique à barres : Nombre d'avis par langue**
  - Visualise la **répartition linguistique** des avis clients.
  - Permet de détecter la diversité culturelle ou les préférences linguistiques des utilisateurs.
- **Carte géographique (Map)**

- Représente les **24 gouvernorats tunisiens**.
- Montre la **répartition des clients** ou des avis selon leur localisation.
- Offre une vue géospatiale claire sur la présence des clients à travers le pays.

- **Tableau croisé (Table)**
  - Détaille les **gouvernorats et le nombre de clients** par région.
  - Complète la carte pour offrir une lecture numérique plus précise.



4. Word Cloud

A. Méthode 1 : Génération avec Python

L'objectif de cette visualisation est d'**identifier les mots les plus fréquemment utilisés** par les clients dans leurs avis. Cela permet d'identifier rapidement les thématiques récurrentes (positives ou négatives) liées à leur expérience.

En commence par **Nettoyage et Prétraitement du Texte** tel que le texte brut des avis a été soumis à plusieurs étapes de traitement :

- Suppression de la ponctuation et des caractères spéciaux,
- Passage en minuscules,
- Standardisation des mots proches (ex. : *connect*, *connected*, *connection* ⟶ *connexion*),

- Remplacement des fautes d'orthographe ou des variantes (ex. : *ooreedoo*, *اوريدو* ⟶ *ooredoo*),
- Suppression des mots vides (stopwords) en arabe, français et anglais,
- Réorganisation des mots arabes pour respecter la direction droite-gauche (RTL) à l'aide des bibliothèques arabic_reshaper et bidi.

```
# 📌 Stopwords en arabe
stopwords_arabe = set([
    "ماذا", "ما", "ذلك", "تلك", "ذلكم", "هذه", "هذا", "الذي", "التي", "الذين", "إلى", "الى", "عن", "مع", "على", "من", "في",
    "لماذا", "هل", "كل", "كان", "كانت", "يكون", "قد", "لقد", "لم", "لن", "إن", "أن", "إنه", "أنها", "أنه", "هو", "هي",
    "أي", "إي", "إنه", "إنها", "إذا", "حتى", "ثم", "كيف", "أين", "أنتم", "أنت", "انت", "انتم", "نحن", "أنا", "هن", "هم",
    "او", "أو", "بل", "و", "وهو", "وهي", "وهذا", "وهذه", "فيه", "فيها", "بها", "به", "بهم", "عند", "حول", "بعد", "قبل", "بين",
    "بدون", "غير", "معه", "معها",
    "باهي", "لي", "عندي", "الا", "انا", "كي", "ة", "ر", " ", "فى", "اليها", "اليه", "عليها", "عليه", "بسبب", "ذلك", "أثناء", "حين", "كما", "لكن"
])
```

```
# 📌 5. Stopwords personnalisés
stopwords_personnalises = set([
    "the", "i", "and", "to", "is", "a", "it", "you", "for", "w", "in", "but", "of", "have", "that", "el", "with", "my",
    "في", "on", "can", "not",  "they", "or", "de", "و", "its", "if", "this", "so",  "get", "are", "your", "as","from", "le",
    "be", "good", "لا", "just", "at", "la", "ة", "نحب", "what", "pas", "ken", "im", "u", "same", "ma", "0", "per" , "was" , "go" ,
    "all" ,"any" , "li" ,"l", "pou", "yes", "des" , "shit", "here","mais" , "ama" , "6",
    "je", "about", "there", "dont", "no", "do", "use", "think", "me", "will", "thats",
    "en", "et" , "50 " , " نيجه " , "لا" , "an" , "also" , "even"  , "then" , "than" ,"we",  "ne" ,
    "50", "20", "100", "1", "2", "3",  "ki" ,"mon"," now",  "f" ,"est",  "ppr", "etc" , "their" ,"less","nul","fama","10",
    "them","did","ads","ww","b","3al","jai","fi","un","vous","que","didnt","which","should","those","eli","ur","ask","live","had",
    "by", "am","30","one" , "nbadel", "I" , "wala" ,"how", "who","cap","buy", "3andi" ,"dans","got","am","pour" ,"more" ,"les", "fel",
    "une","cant" ,"8", "mta3", "ge","9" ,"24","two","yet","say","be","why","wanna","tho","youll","bech","famech","doesnt", "much","pay","paying","where"
])
```

Par la suite, **Création du Nuage de Mots (Word Cloud).** Après avoir regroupé tous les mots restants, une analyse de fréquence a été faite grâce à collections.Counter. Puis, les résultats ont été utilisés pour générer un **nuage de mots** :

- Taille des mots proportionnelle à leur fréquence,
- Mots arabes correctement affichés,
- Limitation à 200 mots maximum.

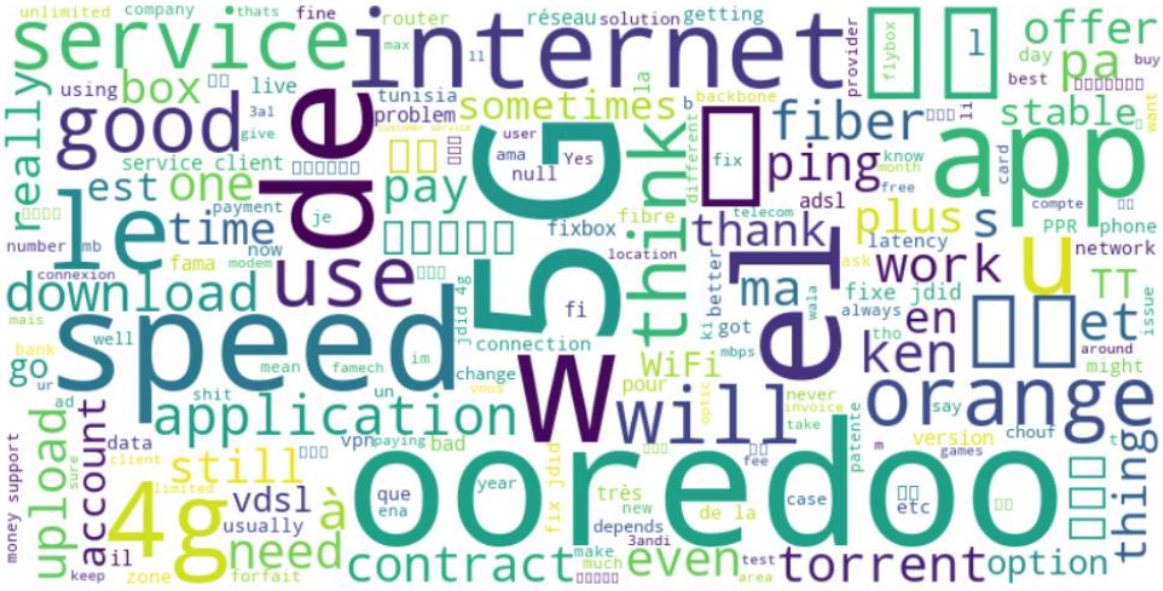
Les mots les plus fréquents sont les suivants :

Mot	Fréquence
5g	212
Ooredoo	150
Application	90
Service	87
4g	77
Speed	63

B. Méthode 2 : Génération du Word Cloud avec Power BI

La deuxième approche consiste à **intégrer dynamiquement le nuage de mots dans un tableau de bord Power BI**. Cette méthode permet de proposer une **visualisation interactive et exploitable** directement par les utilisateurs métiers.

1. Première essai



➔ On remarque qu'il des objets inconnus les carrées

2. Deuxième essai

J’ai essai de faire un Word cloud avec la langue arabe seulement à l’aide de script python en Power BI.

3. Troisième essai

J’ai essai de le faire avec la langue français et anglais.

4. Quatrième essai

J'ai tenté de créer un Word Cloud multilingue en regroupant les avis clients rédigés dans les trois principales langues utilisées : l'arabe, le français et l'anglais.

C. Comparaison entre Python et Power BI

Critère	Python	Power BI
Flexibilité	Très élevé : possibilité de personnaliser la forme, les couleurs, les polices, les stopwords, etc.	Limité aux options disponibles dans le visual Word Cloud
Support du multilingue	Excellente prise en charge (notamment avec NLTK, regex, etc.)	Limité, difficulté à gérer l'arabe ou les caractères spéciaux
Prétraitement des données	Avancé : nettoyage, normalisation, suppression des stopwords personnalisés	Très limité sans passer par Power Query
Visualisation interactive	Moins interactif nativement (sauf si intégré dans un outil web)	Plus interactif dans le rapport Power BI
Intégration avec d'autres analyses	Facilement intégrable à des pipelines de data science ou de machine learning	Moins adapté à des traitements complexes

Critère	Python	Power BI
Courbe d'apprentissage	Plus technique, nécessite des compétences en programmation	Plus accessible aux utilisateurs métier

➔ Bien que Power BI soit pratique pour une visualisation rapide, Python reste beaucoup plus flexible pour un Word Cloud multilingue et personnalisé, surtout lorsqu'on travaille avec des textes en arabe, français et anglais. Il permet un contrôle total sur le traitement linguistique et l'apparence finale du nuage de mots.

Conclusion

Ce stage d'été au sein de l'entreprise Ooredoo Tunisie a constitué une expérience formatrice, tant sur le plan technique que professionnel. En intégrant une problématique réelle – l'analyse des avis clients – j'ai pu mettre en pratique mes connaissances en data science et renforcer mes compétences en nettoyage, transformation et visualisation de données.

Le travail réalisé m'a permis de parcourir toutes les étapes d'un projet d'analyse de données : depuis la préparation d'un jeu de données complexe et hétérogène, jusqu'à la création d'indicateurs de performance et de visualisations pertinentes facilitant la prise de décision. L'utilisation du langage Python m'a offert une grande flexibilité, notamment pour le traitement multilingue et l'extraction sémantique, tandis que Power BI a permis de proposer des représentations interactives accessibles aux utilisateurs métiers.

Ce projet m'a également sensibilisée à l'importance d'une donnée propre, structurée et contextualisée dans une démarche orientée client. Il a mis en évidence le rôle stratégique que peuvent jouer les analyses de sentiment et les visualisations avancées dans l'amélioration de la relation client et la prévention du churn.

Au-delà des compétences techniques acquises, ce stage a renforcé ma rigueur, ma capacité d'adaptation ainsi que mon esprit analytique. Il marque une étape importante dans mon parcours vers une carrière dans l'analyse de données, et me motive à poursuivre dans cette voie, avec l'ambition de contribuer à des projets à fort impact décisionnel.