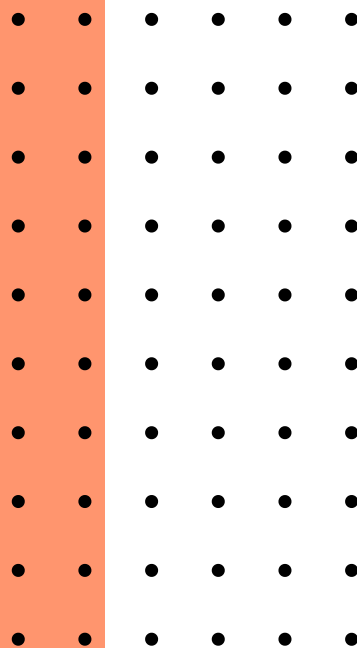




REJOIGNEZ NOTRE ÉQUIPE !



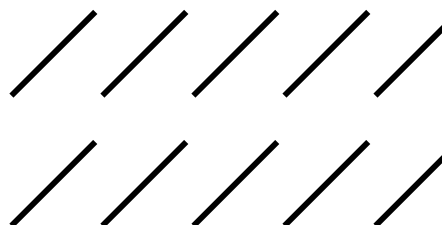
SPORTONE



ATTEIGNEZ LA PREMIÈRE MARCHE DU PODIUM

DE GUNST NOAH
DOIZON TOM
FARHAT ELIAS
LARRE VINCENT

ANNÉE 2023-2024



DONNEUR D'ORDRE : MESMOUDI AMIN

SPORTONE, L'INCONTOURNABLE RÉFÉRENCE POUR LES PRODUITS DE SPORT

Monsieur Mesmoudi, le directeur général de SportOne, une société spécialisée dans la vente d'articles de sports et de plein air, a sollicité notre équipe de développeurs pour la création d'un entrepôt de données visant à améliorer le traitement et l'analyse de données au sein de l'entreprise. L'objectif est de recueillir diverses informations liées aux ventes, clients, fournisseurs ou encore les produits et le temps afin de faciliter le fonctionnement de l'entreprise SportOne. La société nous a fourni la structure du modèle de l'entrepôt de données souhaité et nous avons donc ensuite nettoyée et liée les données fournies pour répondre aux exigences du modèle relationnel de l'entrepôt. De plus, l'entreprise imposait également que l'insertion de données dans cet entrepôt soit automatisée car l'entreprise a réalisé plus de 168 000 commandes au cours de l'année 2021. Il est donc évidemment impossible d'entrer chaque commande à la main. Finalement, un tableau de bord interactif est lié à l'entrepôt afin que SportOne puisse exploiter ces données et prendre des décisions qui pourront assurer une longueur d'avance sur leurs concurrents. Dans ce rapport, nous allons donc revenir sur chacune des étapes réalisées pour arriver à la conception finale de ce projet. À chaque étape, nous détaillerons les outils utilisés, la méthodologie employée, le besoin auquel cette étape a répondu, et les difficultés éventuellement rencontrées.

ETAPE 1 : COUP D'ENVOI : RÉUNION POUR MIEUX COMPRENDRE LE PROJET

Dans un premier temps, notre équipe s'est réunie pour échanger sur les contraintes et les exigences auquel l'entrepôt de données était soumis. Cette étape était relativement courte mais essentielle pour acquérir une bonne compréhension des attentes de Sportone et déterminer la meilleure approche pour atteindre les objectifs fixés. Nous avons également, dans cette étape, repéré les différents liens entre ce qu'on appelle la table de fait, la table qui stocke les commandes effectuées. Cette table était liée à des dimensions telles que le temps ou le client qui a effectué la commande. Nous avons discuté des possibles problèmes et incompatibilités qu'on pouvait rencontrer lié à des différences entre les données fournies et les contraintes de l'entrepôt de données. Pour illustrer cela, prenons la table des employés, chaque employé avait un numéro de référence contenant son numéro d'entreprise, département, section et groupe. Cependant, l'entrepôt de données ne demandait pas les numéros de référence mais les noms d'entreprise, département, section et groupe. Il fallait donc retrouver ces noms grâce aux fichiers fournis. Nous nous sommes alors rendu compte que les noms d'entreprise, section, groupe et départements étaient dans le même fichier que les employés eux-mêmes. Notre équipe s'est alors mis d'accord sur la manière dont nous allions traiter de tels problèmes.

ETAPE 2 : PYTHON MARQUE DES POINTS : UNE INSERTION RAPIDE ET EFFICACE

LES PRÉREQUIS :

Afin de pouvoir utiliser le code python fourni, il faut installer les librairies au préalable. Pour ce faire, il suffit tout simplement d'aller dans la boîte de commandes et d'écrire: "pip install " suivi du nom de la librairie :



Pandas : pour la lecture des fichiers csv, txt et excel ainsi que la manipulation des données



Pyodbc : pour la lecture des fichiers mdb



Datetime: pour gérer les dates



OS : pour vérifier l'existence des fichiers



SYS: pour arrêter le programme en cas d'erreur



Unicodedata : afin de décoder les caractères spéciaux



Re : la librairie regex de python qui permet de vérifier que du texte soit dans un format particulier

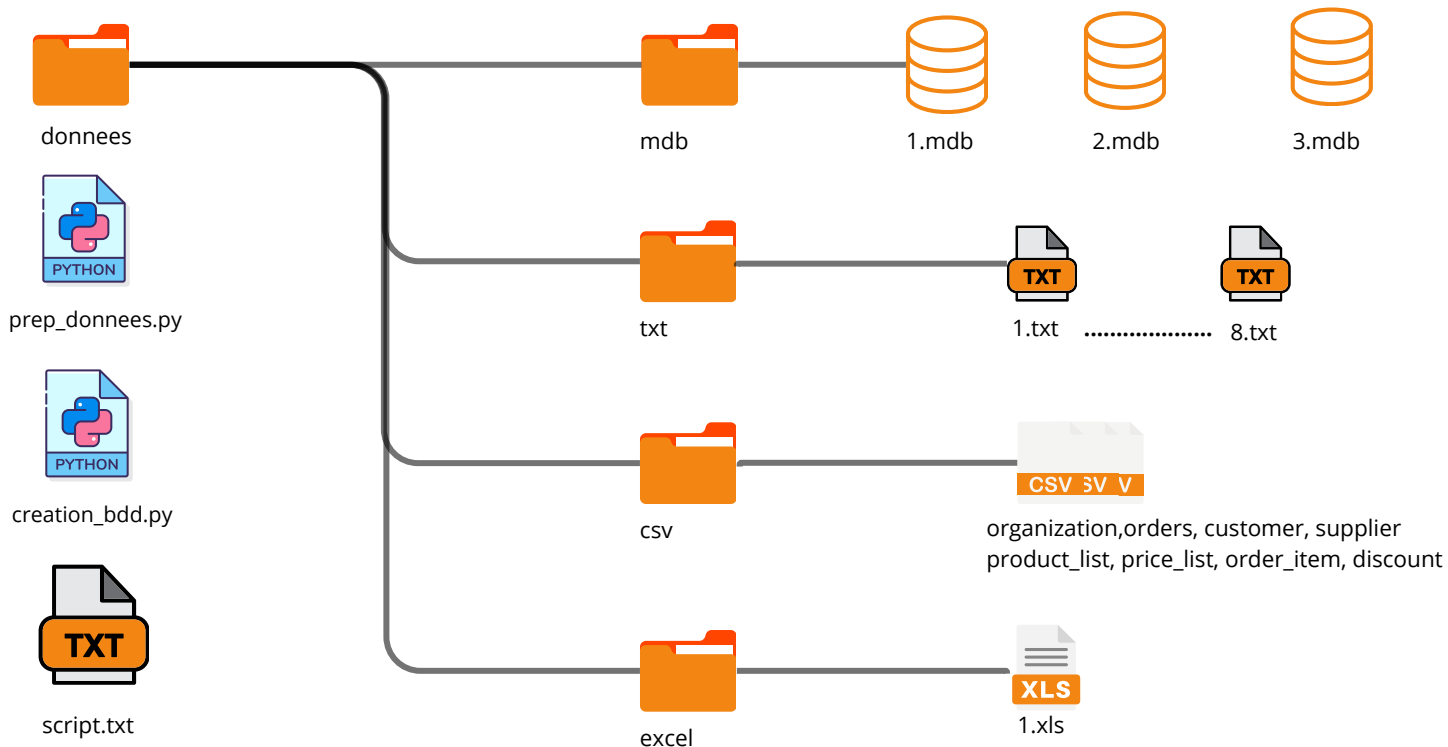


Cx_Oracle: permet d'insérer des données sous oracle avec python



COMMENT UTILISER LE PROGRAMME :

Avant d'exécuter le programme python, veuillez faire en sorte que votre dossier suive l'arborescence suivante et que vos noms de fichiers soit les mêmes :



Veuillez également penser à vous munir des informations nécessaires pour vous connecter à votre serveur oracle. On y retrouve l'adresse IP de votre serveur oracle, votre SID, votre numéro de port, votre nom d'utilisateur et votre mot de passe.

Lorsque vous démarrez le programme "prep_donnees.py", l'algorithme vous demandera de saisir ces informations. Attention, il est important de noter que si une de ces informations est au mauvais format, le programme vous demandera de les ressaisir à nouveau. S'il est impossible d'ouvrir une connexion avec les informations fournies, l'algorithme affichera un message d'erreur et vous devez redémarrer le programme. Si vous n'avez pas d'erreur alors le programme va ouvrir une connexion au serveur Oracle.

Après ouverture de la connexion, vous devez indiquer au programme si vous voulez créer ou recréer l'entrepôt de données. Dans le cas où vous voulez le recréer, le programme lancera le fichier "creation_bdd.py" qui exécute les instructions sql contenu dans le fichier "script.txt".

Suite à ça, le programme lira chacun des fichiers contenu dans le dossier "donnees" et utilisera ces fichiers pour créer des tables (dataframes) avec la librairie pandas puis il les manipulera de sorte à ce que les données soient de la même structure que les tables dans l'entrepôt de données et il les exportera en csv pour faire une sorte de dump de l'entrepôt.

Enfin, en cas de succès des étapes précédentes, le programme va insérer les données dans l'entrepôt en utilisant la librairie "Cx_Oracle" et vous affichera un message pour chaque table qui aura bien été alimentée.



Veuillez vous assurer que les fichiers que vous insérez ne sont pas déjà présents dans la base de données, sinon le programme indique une erreur de contrainte et l'insertion échoue. De plus, le temps d'exécution du programme est proportionnel à la quantité de données à insérer. Comme le fichier contenant les produits achetés contient plus de 900 000 lignes, le temps d'exécution peut monter jusqu'à 50 secondes.

LES DIFFICULTÉS LIÉS AU NETTOYAGE ET À L'INSERTION DES DONNÉES :

Malgré une bonne maîtrise du langage python dans notre équipe, un projet avec plus de 100 méga octets de données s'est avéré très compliqué. Pour nous, nous avons supposé que l'entreprise souhaite un entrepôt de données qui est alimenté quotidiennement ou alors très régulièrement avec des nouvelles données. Comme le temps d'exécution du programme n'est pas optimale, SportOne pourrait envisager d'insérer les données durant la nuit comme certains concurrents. Cela permettrait aux responsables de directement pouvoir consulter leur tableau de bord le lendemain matin tout en ayant accès aux données de la veille. Les responsables peuvent alors prendre des décisions stratégiques au sein de l'entreprise qui peuvent varier quotidiennement afin d'avoir un avantage face à la concurrence écrasante dans le secteur de la vente des produits sportifs. Malgré cet atout, le programme n'est pas très permissif, il faut une arborescence de fichiers bien défini et le temps d'exécution peut être long. De plus, le programme est assez long à cause des petites erreurs dans les données tels que des dates séparées par des doubles slash ou encore des caractères spéciaux dans les noms qui sont parfois mal encodés. Si vous voulez mieux comprendre le fonctionnement du code python, notre équipe vous invite à lire le code python qui est structuré et commenté.

ETAPE 3 : POWER BI : LE LOGICIEL PHARE POUR INTERPRÉTER LES DONNÉES

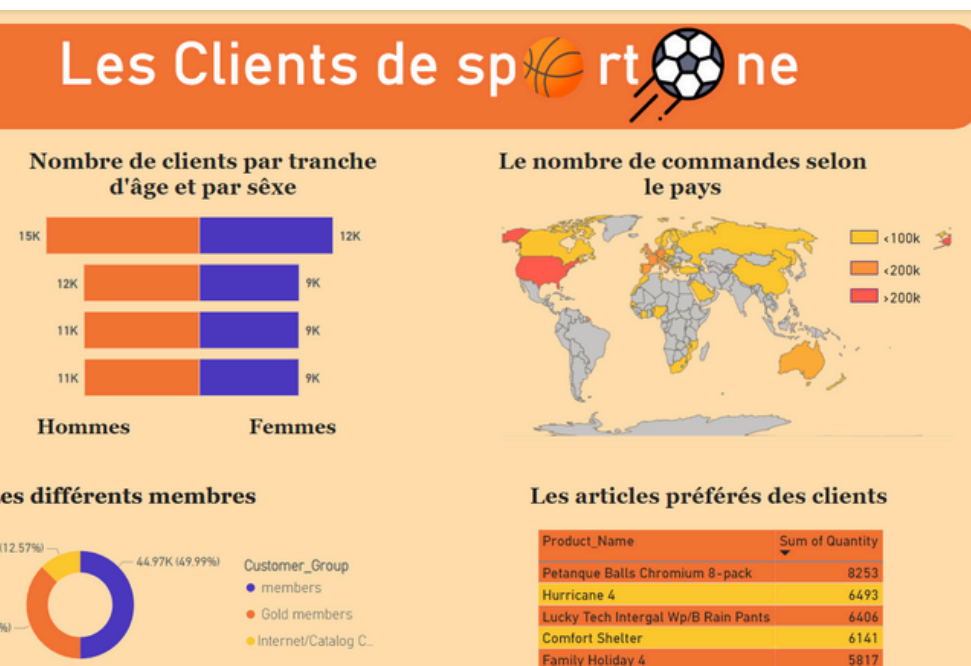
Afin de valoriser la donnée dans l'entrepôt, notre équipe devait concevoir un tableau de bord sous Power BI. Notre équipe a alors essayé de se mettre à la place des responsables de SportOne pour mieux comprendre les besoins auquel ce tableau de bord pouvait répondre tout en étant conscient du nombre conséquent des données qui sont à notre disposition. Notre vision était claire, cet outil devait permettre à SportOne d'avoir un aperçu rapide et concis de l'activité de l'entreprise mais avec assez de détails pour permettre à la structure de mettre en vigueur une stratégie qui pourrait permettre d'augmenter le bénéfice ou l'activité de l'entreprise et cela aussi bien dans leurs magasins physiques installés partout dans le monde que sur leur site d'e-commerce. De plus, les graphiques doivent être simples à comprendre et percutants. Au final, nous avons donc construit des problématiques réalistes autour de ces données que vous trouverez ci-dessous :

J'AIMERAIS AVOIR UN APERÇU RAPIDE ET CHIFFRÉ DE NOTRE ACTIVITÉ :



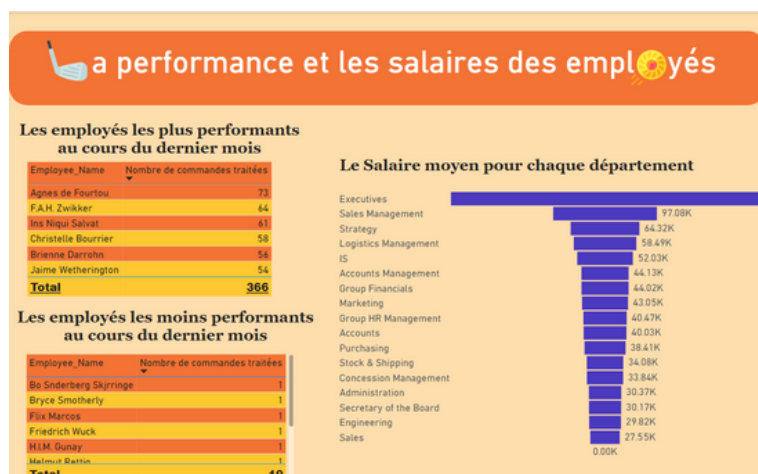
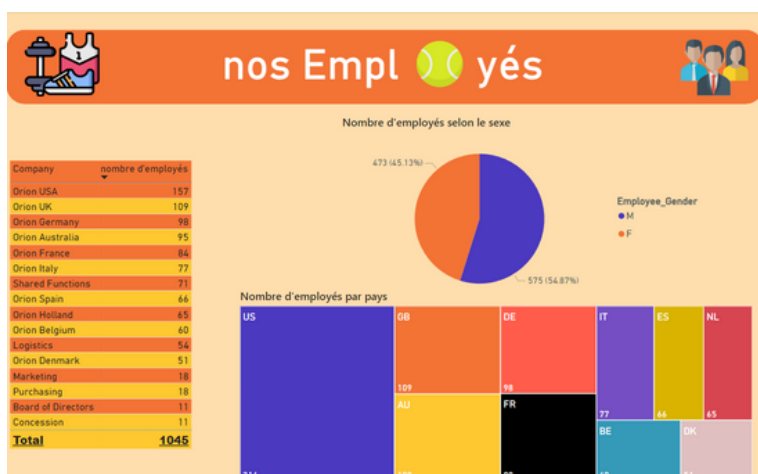
Lorsque vous ouvrez le tableau de bord, une page d'accueil s'ouvre à vous. A gauche, vous trouverez des chiffres sur l'activité globale de la société SportOne au cours du dernier mois. A droite, vous trouverez un navigateur simple qui permet de rapidement naviguer les pages du tableau de bord.

QUI SONT LES CLIENTS DE SPORTONE ?



La page "connaître les clients" permet de mieux se rendre compte de la grande diversité des clients, on voit alors que les clients ont majoritairement en dessous de 30 ans et sont notamment du sexe masculin. Etant donné que les moins de 30 ans ont une assez bonne connaissance de l'entreprise il serait préférable de développer de nouveaux axes de communication comme la télévision, la radio ou encore des affiches où les autres tranches d'âge sont souvent plus sensibles. Ces clients proviennent notamment des pays de l'est ainsi que de l'Australie. La moitié de ces clients sont des membres classiques et que 12,5% de leurs clients commandent leurs produits sur le site e-commerce. L'article le plus vendu sont les boules de pétanques avec plus de 8 000 exemplaires vendus avec une avance considérable sur le deuxième produit les "Hurricane 4".

ET QUI SONT LES EMPLOYÉS DE SPORTONE ? QUEL EST LEUR PRODUCTIVITÉ ?



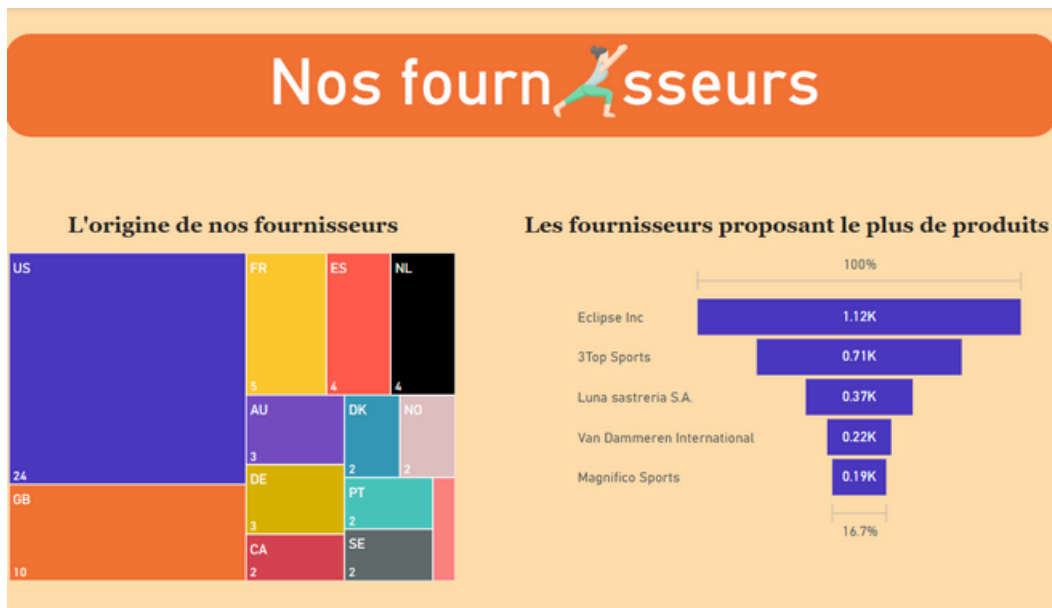
Ci-dessus vous trouverez un tableau de bord sur l'identité des employés (à gauche) et un sur leurs performances (à droite). La page que vous pouvez visualiser à gauche utilise des données plutôt qualitatives. Nous pouvons par exemple voir que près de 55% des employés sont des hommes et qu'une grande partie des employés travaille aux Etats-Unis ou en Europe (treemap en bas à droite du tableau de bord à gauche). A gauche de la page, vous trouverez également un tableau qui indique le nombre d'employés pour chaque entreprise affiliée avec SportOne. Sur l'autre page, vous pouvez voir, d'après les deux tableaux qui sont les employés les plus et les moins performants au cours du dernier mois. A noter que ces tables excluent l'employé numéro 99999 qui représente la vente sur internet. A droite de cette page, vous voyez le salaire moyen pour chaque département. On voit alors que les cadres gagnent près de 3 fois plus que les commerciaux. Ces deux pages permettent donc à SportOne d'avoir un premier aperçu de l'activité de la structure.

COMMENT L'ACTIVITÉ DE SPORTONE ÉVOLUE-T-IL AU COURS DU TEMPS ?



Pour répondre à cette problématique, notre équipe a élaboré une page qui montre l'évolution des ventes au cours du temps, en haut à droite, vous pouvez voir que le nombre de commandes est le plus élevé en été. En effet entre 2017 et 2021, le nombre de commandes monte jusqu'à 65 000 sur les 4 ans pendant la période juillet-août, soit dix mille de plus qu'en novembre. Ces disparités se constatent davantage grâce au graphique de droite. Par exemple, en 2021, on peut voir que le montant total de ventes s'élève à seulement 25 millions contre près de 40 millions au trimestre 3 soit un écart de 15 millions entre ces deux trimestres. Dans ce cas, nous conseillons à l'entreprise de mettre en place des campagnes marketing comme des soldes pendant les périodes creuses pour attirer davantage de clients. De plus, la pancarte en bas à gauche montre le nombre de commandes effectués. Comme SportOne cherche à profiter pleinement de cette transition vers une structure "e-commerce", il est important d'en voir l'évolution. C'est pourquoi, notre équipe a mis un bouton qui permet de faire varier les graphiques en fonction de l'année. Chaque graphique s'adaptera alors en fonction de l'année sélectionnée. Il est également possible de mettre en pause sur une année ou d'avancer d'année en année.

QUI SONT LES FOURNISSEURS ?



La dernière page intitulée “Nos fournisseurs” permet de mieux connaître les fournisseurs des produits vendus par SportOne. On voit sur le treemap situé à gauche que les fournisseurs sont situés dans les mêmes pays que les employés, c’est à dire les Etats-Unis et l’Europe ce qui est plutôt logique comme le nombre de commandes est plus important. De plus, Le graphique à droite montre que les fournisseurs principaux sont Eclipse Inc et 3TopSports qui fournissent respectivement 1120 et 710 produits par mois. SportOne pourrait envisager des partenariats avec ces fournisseurs afin de diminuer le coût d’achat et donc augmenter les gains.

COUT DE SIFFLET FINAL

Pour conclure, ce projet a été un véritable défi pour notre équipe mais grâce à notre persévérance nous avons pu en arriver à terme notamment grâce à la maîtrise de python et power BI mais aussi grâce à la communication au sein du groupe. La mise en place de cet entrepôt de données va permettre à la fois à la société SportOne d’améliorer leur efficacité grâce à l’automatisation du traitement des données mais également de les aider dans leur prise de décisions pour améliorer leur Chiffre d’Affaires. Pour encore plus faciliter le traitement des données et prendre de l’avance sur les concurrents, SportOne pourrait dans le futur mettre en place des pipelines qui prépare automatiquement les fichiers puis qui lance le programme python réalisé lors de ce projet.