Table des matières

[Projet 2 : Faites une analyse des ventes pour un e-commerce 1](#_Toc137809194)

[Projet 3 : Créez et utilisez une base de données immobilière avec SQL 1](#_Toc137809195)

[Projet 4 : Réalisez une étude de santé publique avec Python 1](#_Toc137809196)

[Projet 5 : Optimisez la gestion des données d'une boutique avec Python 2](#_Toc137809197)

[Projet 6 : Analysez les ventes d'une librairie avec Python 2](#_Toc137809198)

[Projet 7 : Analysez des indicateurs de l'égalité femme-homme avec Knime 3](#_Toc137809199)

[Projet 8 : Faites une étude sur l'eau potable avec Tableau 4](#_Toc137809200)

[Projet 9 : Produisez une étude de marché avec Python 4](#_Toc137809201)

# Projet 2 : Faites une analyse des ventes pour un e-commerce

* Générer des graphiques adaptés aux types de données (courbe, diagramme en barre, histogramme, boîte à moustache, nuage de points...)
* Synthétiser des résultats à destination d'un client
* Interpréter les informations provenant d'un dashboard
* Utiliser des fonctionnalités avancées d'Excel

# Projet 3 : Créez et utilisez une base de données immobilière avec SQL

* Extraire les données nécessaires à l’analyse.
* Définir des règles de gestion de nettoyage des bases de données (formatage, suppression des doublons, typage valeurs aberrantes…).
* Définir des règles de gestion de structuration des différentes bases de données entre elles.
* Création d'un schéma relationnel sur PostgreSQL en respectant la 3NF.
* Création des fichiers à intégrer dans les tables.
* Création de la base de données (tables, champs, clés primaires et étrangères, contraintes, index)
* Importation des données dans les tables.
* Requêtage des données (jointure, requêtes imbriquées, Windows fonctions, création de vues, champs calculés...).

# Projet 4 : Réalisez une étude de santé publique avec Python

Analyse des ressources alimentaires au niveau mondial

* Extraction des données du site de la FAO
* Nettoyage, préparation et exploration des données
* Analyse des données : pays, population, disponibilité alimentaire, aide alimentaire, sous-nutrition
* Utilisation des librairies spécialisées pour les traitements data (numpy, pandas)
* Utilisation des librairies spécialisées pour les visualisations (matplotlib, seaborn)
* Rédaction et présentation d'une méthodologie d'exploration et d'analyse des données
* Manipulation des DataFrames

# Projet 5 : Optimisez la gestion des données d'une boutique avec Python

* Nettoyage, préparation et exploration des données (gestion types, doublons, valeurs manquantes)
* Analyse univariée : décrire la répartition d'une variable
  + Mesures de tendance centrale (moyenne, mode, médiane, moyenne tronquée...),
  + Mesures de dispersion (variance, écart type, coefficient de variation),
  + Mesures de forme (asymétrie : skewness empirique, aplatissement : kurtosis)
* Gestion des valeurs aberrantes (méthode de l'IQR et du zscore)
* Visualisation de l'analyse univariée (histogramme, boîte à moustaches)
* Tests statistiques de normalité (Kolmogorov-Smirnov, Shapiro-Wilk) d'une variable
* Transformation au logarithme et à la racine
* Tests paramétriques et non paramétriques
* Création de fonctions de synthèse automatisant toute l'analyse univariée

# Projet 6 : Analysez les ventes d'une librairie avec Python

1. Analyse générale des données de ventes (Chiffre d'affaires, produits, clients...)
2. Analyse du comportement des clients en ligne en étudiant les liens en plusieurs variables
3. Analyse générale :

* Nettoyage, préparation et exploration des données (gestion types, doublons, valeurs manquantes, valeurs aberrantes)
* Analyse du chiffre d'affaires :
  + Année, mois, catégories de livres,
  + Moyennes mobiles, modèle ARIMA
  + Analyse de la saisonnalité
* Analyse du nombre de ventes
* Analyse des produits (classement, répartition, courbe de Lorenz et indice de Gini)
* Analyse des clients (classement, courbe de Lorenz et indice de Gini)

1. Analyse du comportement des clients :

* Analyse bivariée entre 2 variables qualitatives (sexe et catégories de livres)
  + Tableau de contingence, Chi2, calcul des résidus standardisés
* Analyse bivariée entre 2 variables quantitatives (âge clients et montants des achats)
  + Visualisation (nuage de points, coefficient de Pearson, Régression Linéaire)
* Analyse bivariée entre une variable quantitative et qualitative (âge clients et catégories de livres)
  + ANOVA
* Réalisation de tests statistiques (Shapiro Wilk, Spearman, Levene, Kruskal-Wallis)
* Utilisation de numpy, pandas, scipy, matplotlib, seaborn

# Projet 7 : Analysez des indicateurs de l'égalité femme-homme avec Knime

1. Mettre en place un processus ETL
2. Collecter des données en respectant le RGPD
3. Préparer des données pour l'analyse en respectant les normes internes à l’entreprise

Créer un workflow sur Knime qui :

* Importe les données et les transfère vers une zone de préparation
* Nettoie et prépare les données
* Créer des indicateurs, tests statistiques et visualisations sur l'égalité Homme/Femme de l'entreprise
* Anonymise les données en respectant le RGPD
* Transfert le fichier pour des analyses sur Tableau

# Projet 8 : Faites une étude sur l'eau potable avec Tableau

Aider un investisseur sur un projet d'infrastructure lié à la gestion de l'eau

1. Synthétiser des résultats à destination d'un client
2. Créer un tableau de bord répondant à des questions analytiques
3. Analyser un besoin client pour formuler des questions analytiques
4. Générer des graphiques adaptés aux types de données

* Créer un workflow de nettoyage et préparation des données sur Tableau Prep
* Export du fichier sur Tableau
* Création de champs calculés
* Choix des vues répondant aux problématiques
* Choix des types de visualisations, des indicateurs et des filtres
* Création d'une vue mondiale et continentale
* Création de 3 vues nationales permettant de filtrer les pays suivants différents critères

# Projet 9 : Produisez une étude de marché avec Python

Déterminer quel pays peut convenir pour exporter son entreprise

Création de son jeu de données :

* Analyse Pestel : choix des variables pertinentes à ajouter
* Import de fichier depuis plusieurs sources de données (données de base, FAO, banque mondiale)
* Exploration des données pour synthétiser des variables
* Nettoyage des données (gestion types, doublons, valeurs manquantes, valeurs aberrantes)
* Création d'un diagramme causale à partir des variables retenues

Analyse :

* Réalisation de l'ACP (Analyse en composantes principales) : kmo, éboulis des valeurs propres, cercle des corrélations, représentation des individus sur le plan factoriel
* Réalisation du clustering ascendant hiérarchique
* Réalisation du KMeans
* Visualisation des clusters sur 2 et 3 dimensions

Proposition de pays pour la nouvelle implantation de l'entreprise.

# Projet 10 : Détectez des faux billets avec Python

* Réaliser une analyse prédictive
* Réaliser une régression logistique
* Réaliser une régression linéaire
* Opérer des classifications automatiques pour partitionner les données