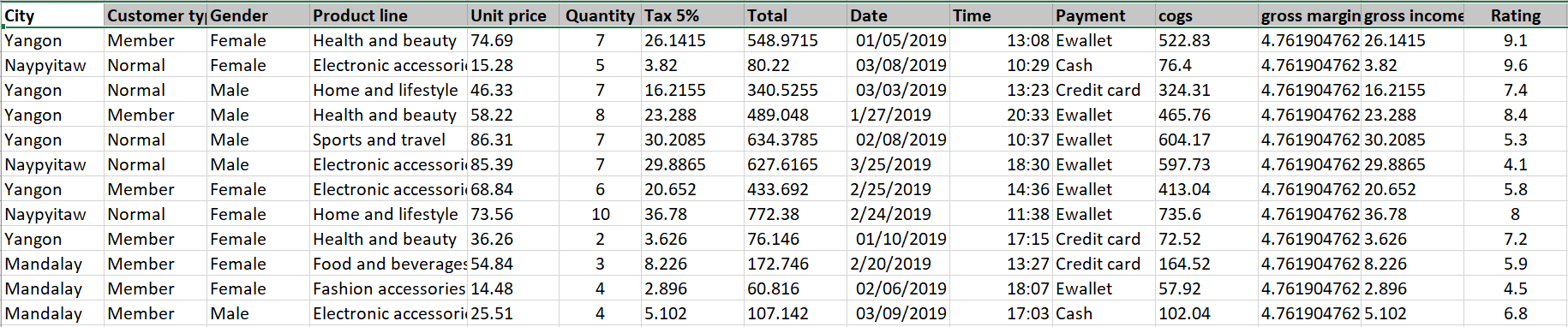
L’entreprise Sales4You sous-traite l’analyse de son jeu de données des transactions à la société *AnalyticsSARL*. Ce jeu de données comprend **un nombre important de lignes.** Chaque ligne représente une transaction et se compose des **caractéristiques** suivantes :

* **Invoice ID (string)** : Identifiant unique attribué à chaque facture.
* **Branch (string)** : Branche du supermarché où la vente a eu lieu (ex. : A, B, C).
* **City (string)** : Ville où la branche est située (ex. : Yangon, Mandalay, Naypyitaw).
* **Customer Type (string)** : Type de client (ex. : Member, Normal).
* **Gender (string)** : Genre du client (ex. : Male, Female).
* **Product Line (string)** : Catégorie de produit acheté (ex. : Health and beauty, Food and beverages).
* **Unit Price (float)** : Prix unitaire des articles vendus (en devise locale).
* **Quantity (int)** : Nombre d'articles achetés dans la transaction.
* **Tax 5% (float)** : Montant de la taxe appliquée (5% du total avant taxes).
* **Total (float)** : Montant total de la facture après taxe.
* **Date (date)** : Date de la transaction (ex. : 2019-01-01).
* **Time (string)** : Heure à laquelle la transaction a eu lieu (ex. : 13:45).
* **Payment (string)** : Mode de paiement utilisé (ex. : Cash, Credit card, Ewallet).
* **COGS (Cost of Goods Sold) (float)** : Coût des biens vendus avant taxes.
* **Gross Margin Percentage (float)** : Pourcentage de marge brute (fixé à 4.76%).
* **Gross Income (float)** : Revenu brut généré par la vente.
* **Rating (float)** : Évaluation donnée par le client pour l'expérience d'achat (échelle de 1 à 10).

Le tableau ci-dessous illustre un extrait du jeu de données contenu dans le fichier **supermarket\_sales.csv**:



**Objectifs :**

Etant un assistant data analyste dans la société *AnalyticsSARL*, vous étiez invité à concevoir des visualisations avec python qui permettront de mieux comprendre la composition des données, identifier les relations entre les variables, montrer la distribution et faciliter l'interprétation des données complexes afin d’aider les décideurs de la société à une bonne prise de décision.

**Partie 1 :**

1. Importer les bibliothèques de visualisation et de chargement des données (seaborn, pandas et pyplot de matplotlib) **(0.5pt)**
2. Lire le fichier *supermarket\_sales.csv* et affecte son contenu à un dataframe pandas **(0.5pt)**
3. Afficher l’entête du dataframe ? **(1pt)**
4. Initier une figure avec une figsize de (10,6) **(1pt)**
5. Utiliser Seaborn pour tracer un graphique en barres de nombres de transactions par genre. Genre sur l'axe des X et nombre de transactions sur l'axe des Y, et choisir une palette de couleurs ' viridis’ **(1pt)**
   1. Utiliser groupby premièrement pour produire un nouveau dataframe d’aggrégation
   2. Puis barplot
6. Définir le titre de graphique, l’axe des x, l’axe des y et afficher le graphique **(2pts)**
7. Copier coller le diagramme résultat puis Commentez le en détail ? **(1pt)**

**Partie 2 :**

On vous confie cette fois la tâche de conception de deux graphiques : 1) le premier pour analyser la relation entre la variable Total et la variable ‘Tax 5%’. 2) le deuxième pour visualiser la distribution des salaires mensuels.

1. Afficher les premières lignes du jeu **(0.5 pts)**
2. Initier une figure avec une figsize de (10,4) **(0.5 pts)**
3. Créer le graphique de nuage de point en spécifiant le jeu de données, la variable Total comme y et la variable ‘Tax 5%’ comme x. Spécifier le titre, l’axe des x, l’axe des y, la grille et afficher le graphique **(3pt)**
4. Copier coller le nuage de points résultant :
   1. Il y a une corrélation ?
   2. décrire bien le type et la forme de la corrélation ?
   3. interpréter le résultat
5. Créer un histogramme pour voir la distribution de la variable Total en définissant un titre, des labels pour les axes x et y, une couleur des points une grille, afficher le graphique.
   1. Interpréter le résultat de l’histogramme.
   2. Les factures sont-elles souvent de faibles montants ou bien réparties ?

**Partie 3 :**

Les deux visualisations ci-dessous sont conçues à partir du même jeu de données cité ci-avant. Durant un meeting, vous étiez invité à expliquer/interpréter les résultats illustrés afin de tirer des informations utiles pour le processus de prise de décision.

|  |  |
| --- | --- |
|  |  |

1. D’après le box plot, Quelle est la ville qui dispose des totaux des ventes les plus élevés pour les produits de type ‘Food and beverages’ ? justifier et détailler la réponse **(3pt)**
2. D’après les 2 graphiques, pourquoi les villes ont des totaux des ventes différents ? qu’est ce qui peut clarifier cela ? **(3pt)**

**Partie 4 :**

1. Utiliser Seaborn pour tracer un graphique en barres pour le/la :
   1. Nombre de transactions vs city
   2. Nombre de transactions vs branch
   3. Nombre de transactions vs product line
   4. Nombre de transactions vs customer type
   5. Total de transactions vs city
   6. Total de transactions vs branch
   7. Total de transactions vs product line
   8. Total de transactions vs customer type
2. **Remarque :** 
   1. Utiliser groupby premièrement pour produire un nouveau dataframe d’aggrégation
   2. Puis barplot
3. Copier coller le diagramme résultat puis Commentez le en détail ? **(1pt)**

Pour résumer cette partie, remplir ce tableau récapitulatif :

|  |  |
| --- | --- |
| Graphique en barres | Le graphique répond à quelles questions |
| * + Nombre de transactions vs city   + Nombre de transactions vs branch   + Nombre de transactions vs product line   + Nombre de transactions vs customer type   + Total de transactions vs city   + Total de transactions vs branch   + Total de transactions vs product line   + Total de transactions vs customer type |  |

**Partie 5 :**

1. Créer le graphique de nuage de point entre les variables :
   1. Total vs gross income
   2. Total vs Quantity
   3. Total vs Rating
2. Copier coller le nuage de points résultant :
   1. Il y a une corrélation ?
   2. décrire bien le type et la forme de la corrélation ?
   3. interpréter le résultat

Pour résumer cette partie, remplir ce tableau récapitulatif :

|  |  |
| --- | --- |
| Le nuage de points | Le nuage de points répond à quelles questions |
| * Total vs gross income * Total vs Quantity * Total vs Rating |  |

**Partie 6 : histogramme**

1. Créer un histogramme pour voir la distribution pour chacune des variables suivantes :
   1. Unit price
   2. Quantity
   3. Gross income
   4. Tax 5%
   5. Rating
2. Pour chaque histogramme:
   1. définir un titre, des labels pour les axes x et y, une couleur des points une grille, afficher le graphique.
   2. Interpréter le résultat de l’histogramme.

Compléter le tableau suivant

|  |  |
| --- | --- |
| L’histogamme | Ecrire les questions auxquelles répond chaque histogramme |
| * Unit price * Quantity * Gross income * Tax 5% * Rating |  |