

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**:

City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin	gross income	Rating
Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	01/05/2019	13:08	Ewallet	522.83	4.761904762	26.1415	9.1
Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	03/08/2019	10:29	Cash	76.4	4.761904762	3.82	9.6
Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	03/03/2019	13:23	Credit card	324.31	4.761904762	16.2155	7.4
Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761904762	23.288	8.4
Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	02/08/2019	10:37	Ewallet	604.17	4.761904762	30.2085	5.3
Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761904762	29.8865	4.1
Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761904762	20.652	5.8
Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761904762	36.78	8
Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	01/10/2019	17:15	Credit card	72.52	4.761904762	3.626	7.2
Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2/20/2019	13:27	Credit card	164.52	4.761904762	8.226	5.9
Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	02/06/2019	18:07	Ewallet	57.92	4.761904762	2.896	4.5
Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	03/09/2019	17:03	Cash	102.04	4.761904762	5.102	6.8

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)**
 - a. Utiliser groupby premièrement pour produire un nouveau dataframe d'aggrégation
 - b. 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)**

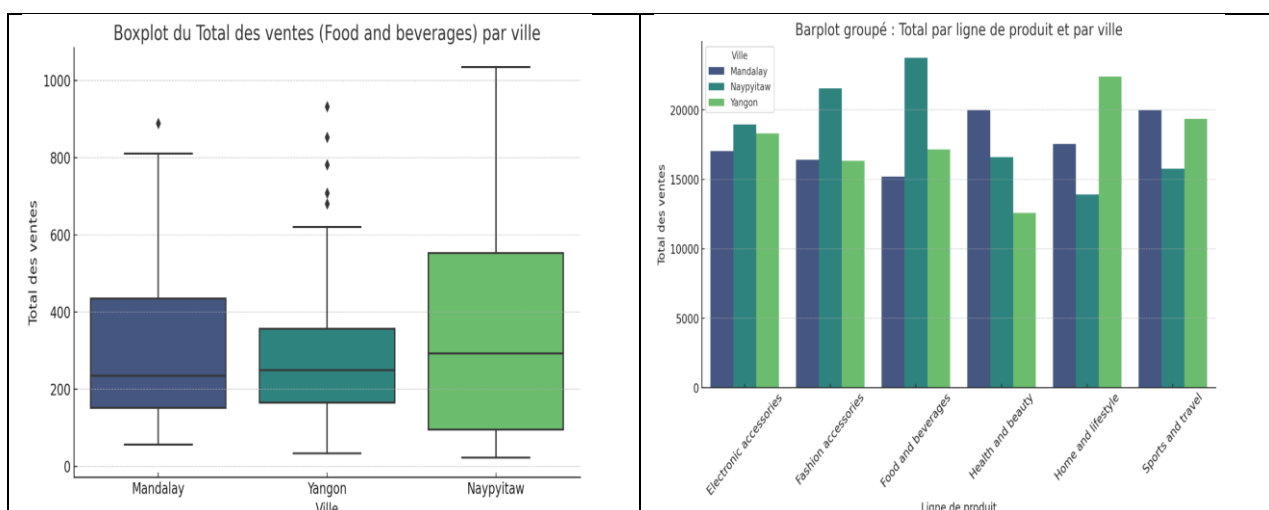
Partie 2 :

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 :
 - a. Il y a une corrélation ?
 - b. décrire bien le type et la forme de la corrélation ?
 - c. 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.
 - a. Interpréter le résultat de l'histogramme.
 - b. 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 :
 - a. Nombre de transactions vs city
 - b. Nombre de transactions vs branch
 - c. Nombre de transactions vs product line
 - d. Nombre de transactions vs customer type
 - e. Total de transactions vs city
 - f. Total de transactions vs branch
 - g. Total de transactions vs product line
 - h. Total de transactions vs customer type
2. **Remarque :**
 - a. Utiliser groupby premièrement pour produire un nouveau dataframe d'aggrégation
 - b. 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
<ul style="list-style-type: none"> ➤ 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 :
 - a. Total vs gross income
 - b. Total vs Quantity
 - c. Total vs Rating
2. Copier coller le nuage de points résultant :
 - a. Il y a une corrélation ?
 - b. décrire bien le type et la forme de la corrélation ?
 - c. interpréter le résultat

Pour résumer cette partie, remplir ce tableau récapitulatif :

Le nuage de points	Le nuage de points répond à quelles questions
<ul style="list-style-type: none"> ➤ Total vs gross income ➤ Total vs Quantity ➤ Total vs Rating 	

Partie 6 : histogramme