

Voici le **rapport final** intégrant toutes les analyses effectuées : segmentation K-means initiale (basée sur `Annual Income` et `Spending Score`), corrélation avec l'âge, corrélation avec le genre, et les recommandations marketing associées. Ce rapport est structuré pour être clair, concis et prêt à être utilisé par une équipe marketing. Les sections incluent les graphiques et tableaux mentionnés, que vous pouvez insérer en fonction des fichiers générés (par ex., `kmeans_clusters_final.png`, `age_by_cluster_boxplot.png`, `gender_by_cluster_barplot.png`, etc.).

Rapport Final : Segmentation des clients pour la personnalisation des offres

1. Contexte et objectif

Une entreprise souhaite regrouper ses clients en segments homogènes pour proposer des offres personnalisées, optimiser ses campagnes marketing et augmenter l'engagement client. À l'aide du dataset **Mall Customer Segmentation**, nous avons appliqué l'algorithme de clustering K-means pour identifier des groupes de clients basés sur leur **revenu annuel** (`Annual Income (k$)`) et leur **score de dépense** (`Spending Score`). Cette analyse a été enrichie par l'exploration des corrélations avec l'âge et le genre. Ce rapport présente les résultats de la segmentation, les profils des clients identifiés, et des recommandations pour des stratégies marketing adaptées.

2. Méthodologie

- **Données utilisées :** Dataset contenant les colonnes Annual Income (k\$), Spending Score (1-100), Age, et Gender.
- **Préparation des données :**
 - Normalisation des données avec StandardScaler pour garantir une échelle comparable.
 - Utilisation de l'algorithme K-means avec 5 clusters, déterminé comme optimal via la méthode du coude et un score de silhouette initial de 0.5547.
- **Outils :** Python avec les bibliothèques pandas, scikit-learn, matplotlib, et seaborn.
- **Analyses complémentaires :** Corrélation avec l'âge et le genre pour affiner les profils.

3. Résultats de la segmentation

L'algorithme K-means a identifié 5 clusters avec les caractéristiques suivantes (moyennes par cluster) :

Cluster	Annual Income (k\$)	Spending Score	Profil
0 (Clients moyens)	55.3	49.5	Revenu et dépenses équilibrés.
1 (Clients premium)	86.5	82.1	Haut revenu, dépenses élevées.
2 (Dépensiers à faible revenu)	25.7	79.4	Revenu modeste, dépenses élevées.
3 (Économies à haut revenu)	88.2	17.1	Haut revenu, dépenses faibles.
4 (Économies à faible revenu)	26.3	20.9	Revenu et dépenses faibles.

Visualisation des clusters

La figure ci-dessous montre la répartition des clients dans les 5 clusters, basée sur leur revenu annuel et leur score de dépense :

(Insérez ici la visualisation `kmeans_clusters_final.png`)

4. Analyse de la corrélation avec l'âge

Pour approfondir la segmentation, nous avons analysé la relation entre l'âge des clients et les clusters formés.

- **Distribution de l'âge par cluster :**

La figure ci-dessous montre un boxplot de l'âge par cluster, révélant des différences significatives :

(Insérez ici le graphique `age_by_cluster_boxplot.png`)

- **Cluster 0 (Clients moyens)** : Âge médian de 45 ans, avec une large dispersion (25-65 ans), indiquant une clientèle variée en âge.
- **Cluster 1 (Clients premium)** : Âge médian de 35 ans (25-40 ans), majoritairement jeunes.
- **Cluster 2 (Dépensiers à faible revenu)** : Âge médian de 30 ans (20-35 ans), très jeunes.
- **Cluster 3 (Économies à haut revenu)** : Âge médian de 45 ans (40-50 ans), d'âge moyen.
- **Cluster 4 (Économies à faible revenu)** : Âge médian de 50 ans (30-70 ans), incluant des clients plus âgés.

L'histogramme suivant confirme ces tendances, avec des distributions d'âge distinctes pour chaque cluster :

(Insérez ici le graphique `age_by_cluster_histogram.png`)

- **Corrélation avec l'âge :**

Les jeunes clients (20-35 ans) dominent dans les clusters à forte dépense (Clusters 1 et 2), tandis que les clients d'âge moyen (40-50 ans) et plus âgés (50-70 ans) sont plus présents dans les clusters économies (Clusters 3 et 4). Le Cluster 0 est hétérogène en âge.

Segmentation améliorée avec l'âge

Nous avons effectué une segmentation K-means en incluant l'âge comme variable :

- **Nouveau score de silhouette** : 0.5547, identique au score initial (0.5547), indiquant que l'ajout de l'âge n'améliore pas la qualité de la segmentation.
- **Visualisation** :

La figure ci-dessous montre les clusters avec l'âge représenté par la taille des points :

(Insérez ici le scatterplot `kmeans_clusters_with_age.png`)

Les clusters restent cohérents avec ceux initiaux, confirmant que l'âge est mieux utilisé comme variable descriptive.

5. Analyse de la corrélation avec le genre

Nous avons analysé la répartition du genre (Male et Female) dans chaque cluster K-means.

- **Répartition du genre par cluster** :

La figure ci-dessous montre un barplot de la répartition absolue du genre par cluster :

(Insérez ici le graphique `Répartition du genre par cluster K-means`)

Les camemberts suivants détaillent les proportions par cluster :

(Insérez les graphiques `Proportion du genre - Cluster 0` à `Proportion du genre - Cluster 4`)

- Résultats :

Cluster	Female	Male
0 (Clients moyens)	59.3%	40.7%
1 (Clients premium)	53.8%	46.2%
2 (Dépensiers à faible revenu)	59.1%	40.9%
3 (Économies à haut revenu)	45.7%	54.3%
4 (Économies à faible revenu)	60.9%	39.1%

- **Cluster 0 (Clients moyens)** : Légère dominance féminine (59.3%), avec une clientèle variée en âge.
 - **Cluster 1 (Clients premium)** : Légère dominance féminine (53.8%), presque équilibré entre genres.
 - **Cluster 2 (Dépensiers à faible revenu)** : Dominance féminine marquée (59.1%), surtout parmi les jeunes.
 - **Cluster 3 (Économies à haut revenu)** : Légère dominance masculine (54.3%), parmi les clients d'âge moyen.
 - **Cluster 4 (Économies à faible revenu)** : Forte dominance féminine (60.9%), incluant des clients plus âgés.
- Corrélation avec le genre :

Les femmes sont majoritaires dans les clusters à faible revenu (Clusters 2 et 4) et dans le cluster moyen (Cluster 0), suggérant une prédominance féminine dans les comportements d'achat variés. Les hommes sont légèrement plus présents dans le cluster des économies à haut revenu (Cluster 3), indiquant une prudence accrue. Le cluster premium (Cluster 1) est presque équilibré.

6. Recommandations marketing

Basées sur les profils identifiés et les corrélations avec l'âge et le genre, voici des stratégies personnalisées pour chaque cluster :

- Cluster 0 : Clients moyens (55.3 k\$, 49.5 Spending Score, 59.3% Female, âge médian 45 ans)
 - **Profil** : Clients avec un comportement équilibré, majoritairement des femmes d'âge moyen à varié.
 - **Stratégie** : Offres généralistes avec une touche féminine : promotions saisonnières, bundles de produits populaires, programmes de fidélité.
 - **Exemple** : Réduction de 10 % sur les achats réguliers ou points de fidélité doublés.
- Cluster 1 : Clients premium (86.5 k\$, 82.1 Spending Score, 53.8% Female, âge médian 35 ans)
 - **Profil** : Clients aisés avec une forte propension à dépenser, légèrement dominés par les femmes jeunes.
 - **Stratégie** : Produits de luxe ou exclusifs : collections premium, services VIP, avec une légère emphasis sur des offres féminines (ex. : accessoires de luxe).
 - **Exemple** : Abonnement premium avec avantages exclusifs (livraison gratuite, accès anticipé).
- Cluster 2 : Dépensiers à faible revenu (25.7 k\$, 79.4 Spending Score, 59.1% Female, âge médian 30 ans)
 - **Profil** : Clients à revenu modeste mais dépensant beaucoup, majoritairement des femmes jeunes.
 - **Stratégie** : Offres promotionnelles sur des produits attractifs pour les femmes jeunes (ex. : cosmétiques, vêtements abordables).
 - **Exemple** : Coupon de réduction de 20 % sur les produits populaires.

- Cluster 3 : Économies à haut revenu (88.2 k\$, 17.1 Spending Score, 54.3% Male, âge médian 45 ans)
 - **Profil** : Clients aisés mais prudents, légèrement dominés par les hommes d'âge moyen.
 - **Stratégie** : Offres incitatives mettant en avant la valeur pour les hommes (ex. : produits durables ou techniques).
 - **Exemple** : Offre d'essai gratuit pour un service premium ou remise sur le premier achat.
- Cluster 4 : Économies à faible revenu (26.3 k\$, 20.9 Spending Score, 60.9% Female, âge médian 50 ans)
 - **Profil** : Clients à faible revenu et dépenses limitées, majoritairement des femmes plus âgées.
 - **Stratégie** : Produits pratiques et économiques adaptés aux femmes seniors (ex. : articles ménagers, santé).
 - **Exemple** : Produits d'entrée de gamme à prix réduit ou offres combinées.

7. Prochaines étapes

1 Validation avec l'équipe marketing :

- Présentez ce rapport et les visualisations pour confirmer que les profils correspondent aux attentes.
- Discutez de la faisabilité des stratégies proposées.

2 Test des offres :

- Lancez des campagnes pilotes sur un échantillon de clients de chaque cluster.
- Mesurez les indicateurs clés : taux de conversion, chiffre d'affaires par segment, satisfaction client.

3 Automatisation :

- Intégrez le modèle K-means dans un système CRM pour attribuer automatiquement des clusters aux nouveaux clients :

```
python
```



```
def segment_new_data(new_data, scaler, kmeans_model):  
    new_X = new_data[['Annual Income (k$)', 'Spending Score (1-100)']].val  
    new_X_scaled = scaler.transform(new_X)  
    new_labels = kmeans_model.predict(new_X_scaled)  
    return new_labels
```

- Enrichissez avec des données sur l'âge et le genre pour personnaliser davantage.

4 Améliorations potentielles :

- Ajoutez d'autres variables (ex. : localisation, fréquence d'achat) si disponibles pour affiner la segmentation.
- Testez d'autres algorithmes (ex. : CAH) si nécessaire.

8. Conclusion

La segmentation K-means a permis d'identifier 5 groupes de clients distincts, offrant une base solide pour personnaliser les offres. Les analyses des corrélations avec l'âge et le genre ont révélé des tendances clés : les jeunes dominent les clusters à forte dépense, les femmes sont majoritaires dans les clusters à faible revenu, et les hommes sont légèrement plus présents dans les économies à haut revenu. Ces insights, combinés aux recommandations proposées, visent à maximiser l'engagement et les ventes. Une validation avec l'équipe marketing et des tests pilotes permettront de confirmer l'efficacité de cette approche.

Instructions pour utiliser le rapport

1 Insertion des graphiques :

- Remplacez les placeholders (*Insérez ici...*) par les fichiers générés :
 - `kmeans_clusters_final.png` pour la visualisation initiale.
 - `age_by_cluster_boxplot.png` et `age_by_cluster_histogram.png` pour l'analyse de l'âge.
 - Répartition du genre par cluster K-means et Proportion du genre - Cluster 0 à Proportion du genre - Cluster 4 pour l'analyse du genre.
- Assurez-vous que les images sont bien alignées dans le document final.

2 Personnalisation :

- Si vous avez des données supplémentaires (ex. : localisation, fréquence d'achat), mentionnez-les dans la section "Améliorations potentielles".
- Ajustez les recommandations si l'équipe marketing identifie des préférences spécifiques.

3 Exportation :

- Convertissez ce rapport en PDF ou PowerPoint pour une présentation professionnelle. Utilisez un outil comme Microsoft Word, Google Docs, ou Jupyter Notebook avec des extensions d'exportation.
- Ajoutez un en-tête avec le logo de l'entreprise et la date (10:54 AM CEST, mercredi 04 juin 2025) pour un rendu officiel.

4 Suivi :

- Après la mise en œuvre des campagnes, ajoutez une section au rapport pour documenter les résultats (ex. : impact sur les ventes) et ajuster les stratégies si nécessaire.

Résumé

- Le rapport final regroupe la segmentation initiale (5 clusters), les analyses de l'âge et du genre, et des recommandations marketing personnalisées.
- Les graphiques et tableaux fournis permettent une communication claire des résultats.
- Suivez les instructions pour intégrer les visualisations et exporter le rapport, puis passez à la validation et aux tests pilotes.

Si vous avez besoin d'aide pour intégrer les graphiques, exporter le rapport, ou explorer d'autres analyses (par ex., avec des données supplémentaires), faites-le-moi savoir !