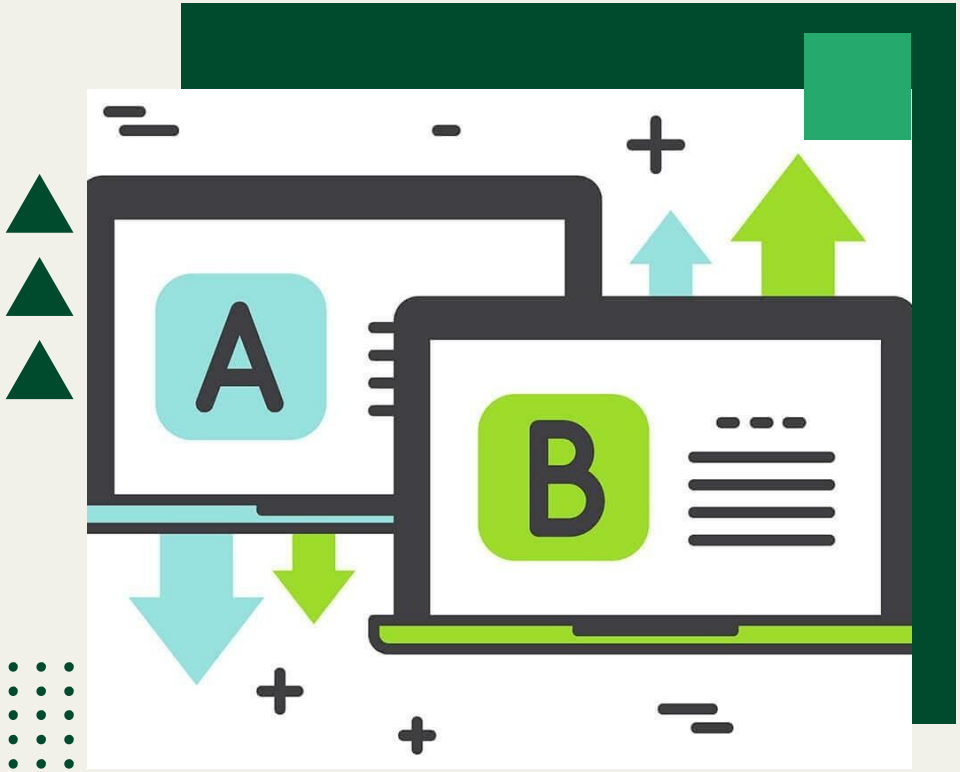
A/B Test
Mise en œuvre:

Nouveau Produit

Recherche de Stratégies

Marketing



Création : N’DOUBA MÔH.

Contenus

Business Understanding

CONTEXTE

Une chaîne de fast-food prévoit d'ajouter un nouvel élément à son menu.
Cependant, elle n'est toujours pas décidée entre trois campagnes de marketing possibles pour promouvoir le nouveau produit.
Afin de déterminer quelle promotion a le plus grand effet sur les ventes, le nouvel élément est introduit dans plusieurs marchés sélectionnés au hasard.
Une promotion différente est utilisée à chaque emplacement, et les ventes hebdomadaires du nouvel élément sont enregistrées pendant les quatre premières semaines.

- Type de promotion 1 : Remise sur vente flash
- Type de promotion 2 : Point de cashback spécial
- Type de promotion 3 : Offres groupées

QUESTION BUSINESS

Y a-t-il une différence dans la performance des ventes entre les trois stratégies marketing ?

OBJECTIFS

Y a-t-il une différence dans la performance des ventes entre les trois stratégies marketing ?

Aperçu des données

MarketID	MarketSize	LocationID	AgeOfStore	Promotion	week	SalesInThousands
1	Medium	1	4	3	1	33.73
1	Medium	1	4	3	2	35.67
1	Medium	1	4	3	3	29.03
1	Medium	1	4	3	4	39.25
1	Medium	2	5	2	1	27.81

Variable	Total Unique Value
MarketID	10
MarketSize	3
LocationID	548
AgeOfStore	25
Promotion	3
SalesInThousands	137
548 rows, 6 columns	

- MarketID : identifiant unique pour le marché
- MarketSize : taille de la zone de marché par les ventes (Large, Medium, Small)
- LocationID : identifiant unique pour l'emplacement du magasin
- AgeOfStore : âge du magasin en années
- Promotion : l'une des trois promotions qui ont été testées (1, 2, 3)
- Week : l'une des quatre semaines pendant lesquelles les promotions ont été réalisées (1, 2, 3, 4)
- SalesInThousands : montant des ventes pour un identifiant LocationID, Promotion et semaine spécifiques

Méthode & Workflow

Conception | Méthode d’analyse | Workflow

● Cette analyse a été réalisée dans plusieurs magasins à différents emplacements. Chaque emplacement mettra en œuvre un type de stratégie marketing différent. Les résultats de vente seront mesurés chaque semaine pendant quatre semaines consécutives dans chaque magasin.
● Les méthodes statistiques utilisées sont les tests pour différences de moyennes pour plus de deux groupes (paramétrique : test ANOVA ou non paramétrique : test de Kruskal-Wallis).

Step 1	Step 2	Step 3	Step 4	Step 5
Préparation des données: - Changer les noms de colonnes - Gestion des valeurs manquantes ou des valeurs invalides - Supprimer les données en double - Reformater le type de données (comme requis)	Compréhension des données & EDA	Test d'hypothèse : - Test de normalité - Test d'homogénéité Test supplémentaire: Test de bimodalité Coefficient & Test de Silverman’s	Test de différence : - Test ANOVA (si les hypothèses sont respectées) - Kruskal Wallis test (si les hypothèses ne sont pas respectées)	Test Post-hoc: - ANOVA: Tukey's HSD test - Kruskal Wallis: Test de Dunn avec correction de Bonferroni

Step 1: Préparation des données

Change column names:

- MarketID ➜ marché ID
- MarketSize ➜ taille_du_marche
- LocationID ➜ localisation_ID
- AgeOfStore ➜ age_du_magasin
- Week ➜ semaine
- SalesInThousands ➜ ventes_en_milliers
- Promotion ➜ promotion

	marche_ID	taille_du_marche	localisation_ID	age_du_magasin	promotion	semaine	ventes_en_milliers
0	1	Medium	1	4	3	1	33.73
1	1	Medium	1	4	3	2	35.67
2	1	Medium	1	4	3	3	29.03
3	1	Medium	1	4	3	4	39.25
4	1	Medium	2	5	2	1	27.81

Step 1: Préparation des données

- Dans ce cas, nous devons effectuer un nettoyage des données pour surmonter les incohérences entre les données et gérer les valeurs manquantes. Mais aucune donnée invalide ou valeur manquante n'a été identifiée.
- De plus, nous devons changer le type de données de plusieurs colonnes, y compris marche_ID, localisation_ID, promotion et semaine, de entier en chaîne car ces colonnes représentent des données catégorielles (même si elles sont sous forme numérique).

```
Total NaN value for each variable
marche_ID            0
taille_du_marche     0
localisation_ID      0
age_du_magasin       0
promotion            0
semaine              0
ventes_en_milliers   0
```

Il n’y a pas de valeurs doubles .

Colonne	Avant	Après
marche_ID	object	Object
taille_du_marche	int64	object
localisation ID	int64	object
age_du_magasin	int64	int64
promotion	int64	object
semaine	int64	object
ventes_en_milliers	float64	float64

Step 1: Préparation des données

- Avant de procéder à une analyse plus approfondie, il est nécessaire de créer un résumé des données qui montre les ventes totales sur 4 semaines (et non sous forme de montants de ventes hebdomadaires).
- De plus, la colonne ‘semaine' sera supprimée.

marche_ID	taille_du_marche	localisation_ID	age_du_magasin	promotion	semaine	ventes_en_milliers
1	Medium	1	4	3	1	33.73
1	Medium	1	4	3	2	35.67
1	Medium	1	4	3	3	29.03
1	Medium	1	4	3	4	39.25
1	Medium	2	5	2	1	27.81

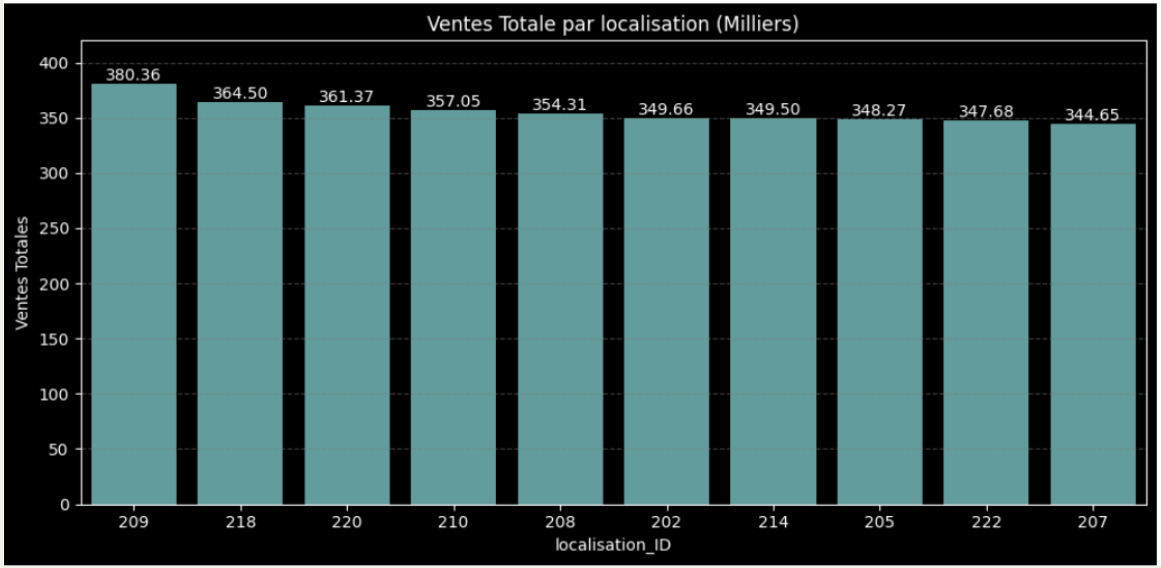
548 rows, 6 columns

marche_ID	taille_du_marche	localisation_ID	age_du_magasin	promotion	ventes_en_milliers
1	Medium	1	4	3	137.68
1	Medium	10	5	2	122.66
1	Medium	11	5	3	145.45
1	Medium	12	12	1	151.14
1	Medium	13	12	1	169.49

137 rows, 6 columns

Step 2: Compréhension des données & EDA

EDA QUESTION 1: Quelle est la vente moyenne au cours des 4 dernières semaines dans chaque emplacement ?



Le tableau dessiné à côté montre les 10 meilleurs emplacements avec les ventes totales de nouveaux produits les plus élevées sur la période de recherche de 4 semaines (lancement initial du produit).

L'emplacement avec l'ID 209 est classé au plus haut avec un chiffre d'affaires total de 380,36 milles. Ce chiffre présente un écart significatif avec le rang en dessous (comparé aux différences de rangs successifs).

Les rangs de deux à dix ont des ventes variant de 344 à 365 mille, les différences de ventes n'étant pas trop significatives.

Step 2: Compréhension des données & EDA

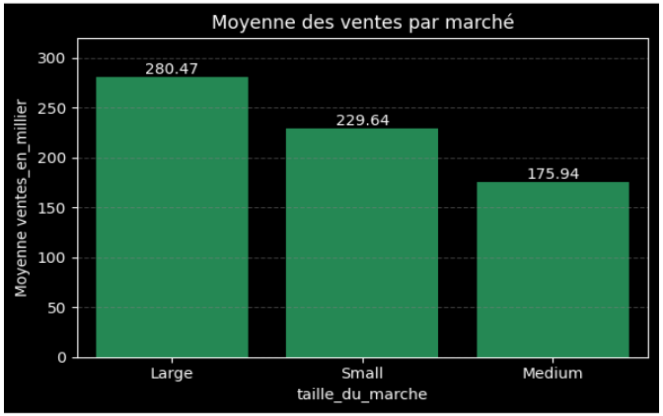
EDA QUESTION 2: Comment se présentent les performances des ventes pour chaque taille de marché ? La taille du marché est-elle positivement corrélée à la performance des ventes ?

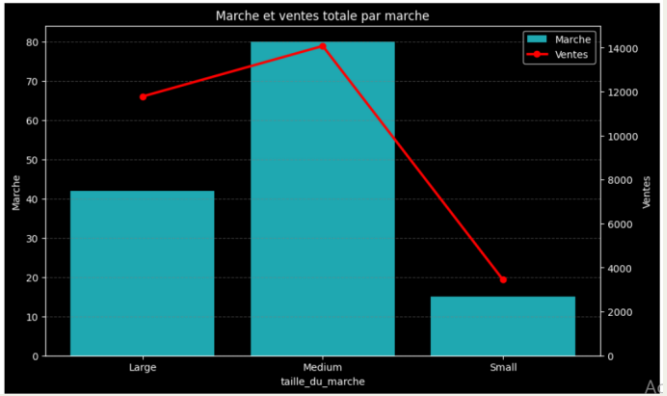
Le Grand Marché est le type de marché avec les ventes moyennes les plus élevées par rapport aux autres types de marchés, s'élevant à 280,47 mille.

L'anomalie ici est que le Petit Marché a en fait des ventes moyennes supérieures à celles du Marché Moyen, avec une différence significative où le Petit Marché a une vente moyenne de 229,64 mille, tandis que le Marché Moyen a une vente moyenne de 175,94 mille.

Le graphique montre que la taille du marché ne correspond pas toujours à la performance des ventes.

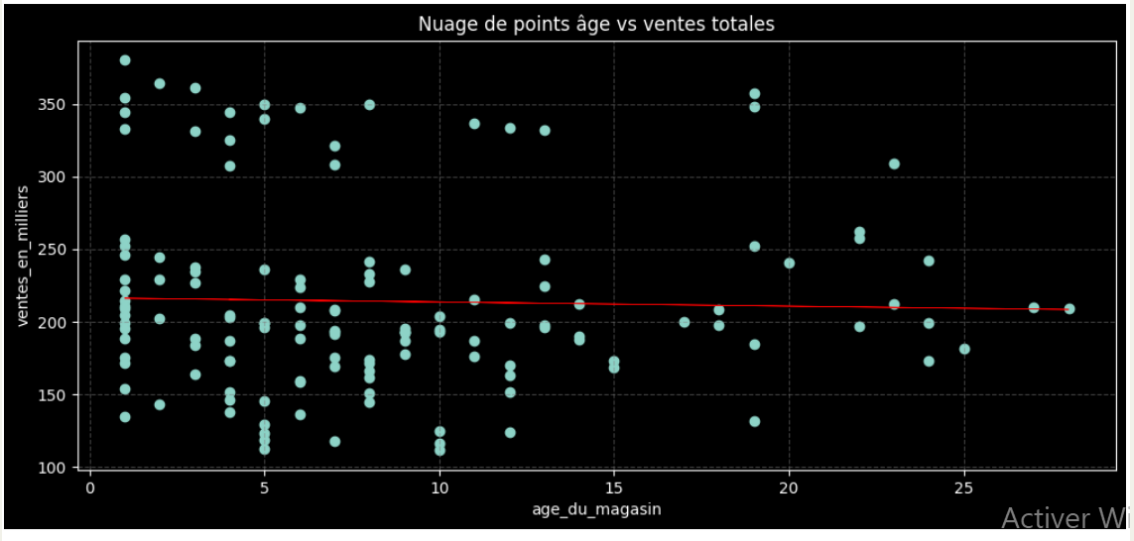
Après une enquête plus approfondie, bien que le marché moyen ait les ventes totales les plus élevées, ce chiffre est également soutenu par le grand nombre de magasins dans cette catégorie. Cela indique qu'il pourrait y avoir des magasins avec de faibles réalisations de ventes.





Step 2: Compréhension des données & EDA

EDA QUESTION 3: Quelle est la relation entre l'âge du magasin et la performance des ventes ?
L'âge du magasin est-il corrélé positivement avec la performance des ventes ?



Les ventes totales et l'âge d'un magasin n'ont pas de relation spécifique.

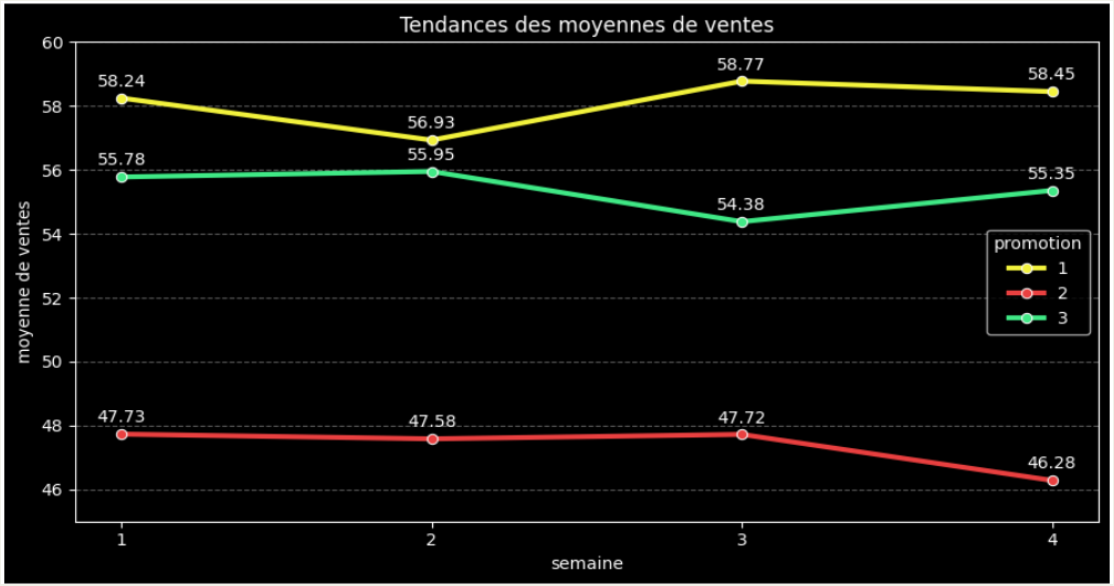
Si une ligne de tendance est tracée, il apparaît généralement que l'augmentation de l'âge d'un magasin n'est pas accompagnée d'une augmentation des ventes totales.

Cela signifie que les nouveaux magasins n'ont pas toujours des ventes constamment faibles/élevées, et vice versa.

Il semble que la majorité des magasins avec des ventes élevées soient relativement nouveaux (par exemple, au-dessus de 300 mille).

Step 2: Compréhension des données & EDA

EDA QUESTION 4: Quelles sont les tendances de vente pour chaque type de promotion ?



Dans l'ensemble, le premier type de promotion domine les ventes hebdomadaires.

Les trois types de promotions ont connu des tendances de ventes fluctuantes au cours de la période de 4 semaines.

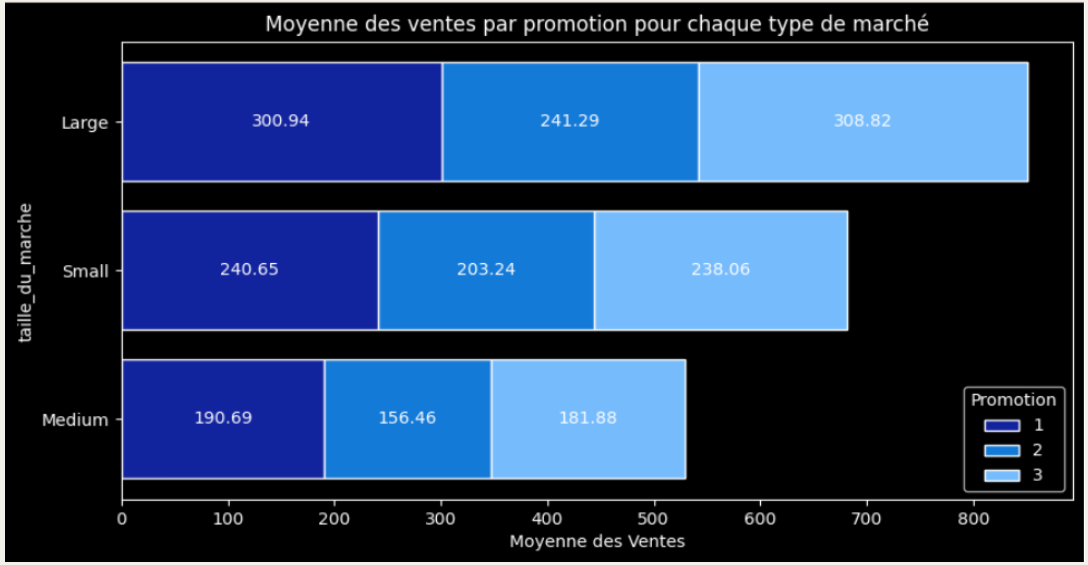
Chaque type de promotion a atteint ses ventes les plus élevées à différentes semaines.

- Le premier type de promotion a réalisé ses ventes les plus élevées au cours de la 3ème semaine avec un total de 58,77 milliers,
- le deuxième type dans la première semaine avec un total de 47,73 mille, et
- le troisième type dans la 2ème semaine avec un total de 55,95 mille.

Step 2: Compréhension des données & EDA

EDA QUESTION 5: Quel type de promotion a réalisé les chiffres de vente les plus élevés ?

Comparez également dans chaque taille de marché.



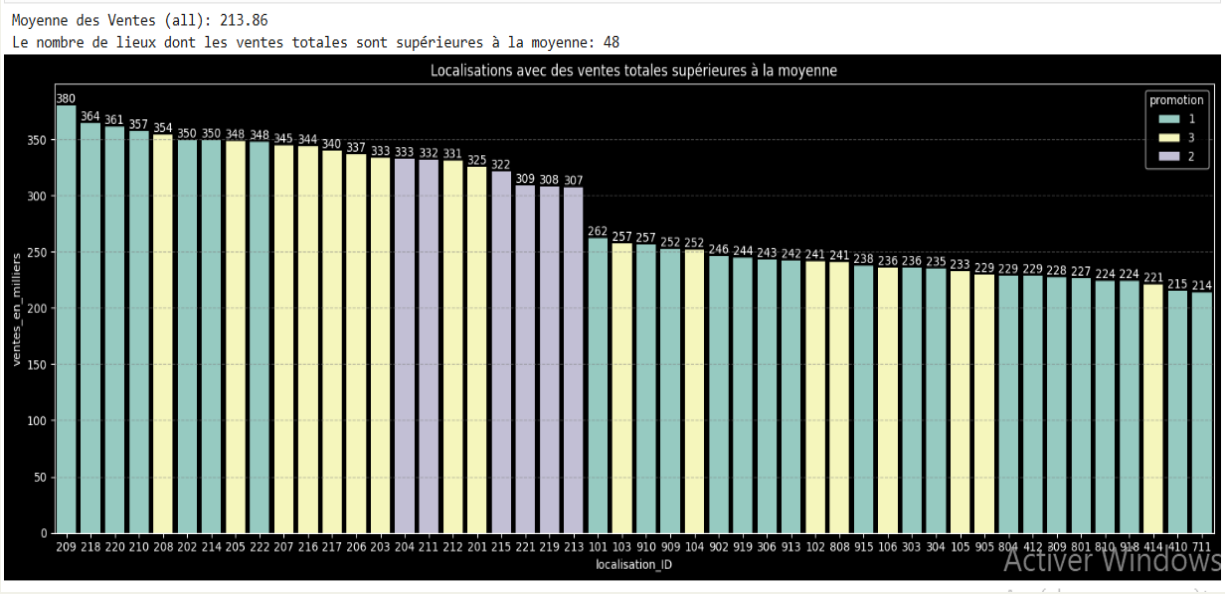
Comme expliqué dans la question 4 de l'EDA, le premier type de promotion domine globalement (même sur une base hebdomadaire).

Lorsqu'on décompose par taille de marché, en termes de ventes moyennes, seul le Grand Marché (Large) n'est pas dominé par le premier type de promotion, mais plutôt par le troisième type.

Pour le petit marché (Small) et le marché moyen (Medum), les ventes les plus élevées sont réalisées par le premier type de promotion.

Step 2: Compréhension des données & EDA

EDA QUESTION 6: Quelles localités ont un chiffre d'affaires supérieur à la moyenne? Quel type de promotion a le plus d'impact pour augmenter les ventes dans ces lieux ?



Les ventes moyennes dans tous les lieux sont enregistrées à 213,86 milles.

Le nombre de lieux avec des ventes totales supérieures à la moyenne est enregistré à 48 lieux.

Les 4 premiers emplacements ayant les ventes les plus élevées utilisent le premier type de promotion. Dans l'ensemble, la majorité des emplacements avec des ventes supérieures à la moyenne utilisent également le premier type de promotion.

Seules quelques emplacements avec des ventes supérieures à la moyenne utilisent le deuxième type de promotion.

Step 3: Test d'hypothèse

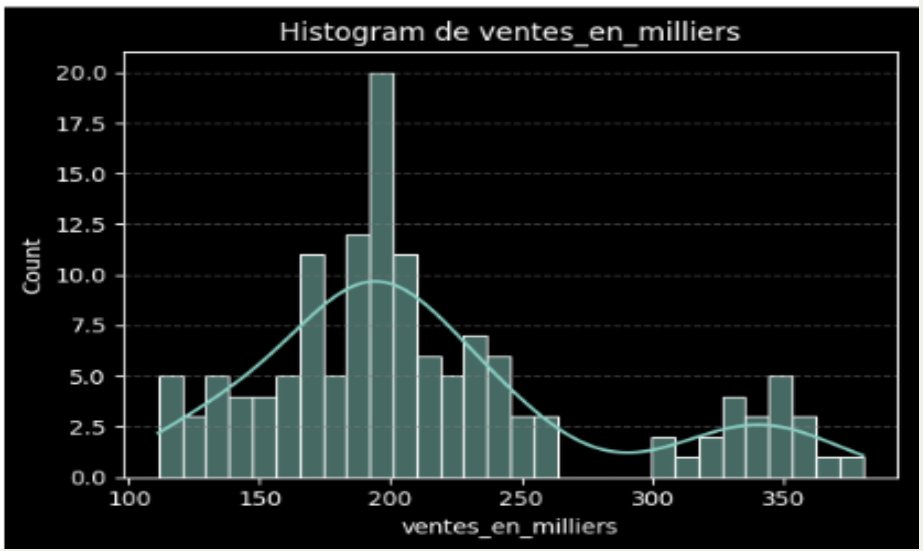
TEST DE NORMALITE

Hypothèses
H_0: Les données suivent une distribution normale.
H_1: Les données ne suivent pas une distribution normale..

Niveau Significatif level (alpha)
5% (0.05)

Shapiro-Wilk Test
Shapiro-wilk test group 1: p-value = 0.000157
Shapiro-wilk test group 2: p-value = 0.000024
Shapiro-wilk test group 3: p-value = 0.000135

Resultats
Rejetez l'hypothèse nulle parce que p-value < alpha (0.05).
Les données ne suivent pas une distribution normale.



Identification
Sur la base de l'histogramme présenté ('Histogramme des ventes (en milliers)'), il y a des soupçons que les données proviennent de deux groupes/populations différents (distribution bimodale), d'où la nécessité d'un test pour déterminer si les données sont identifiées comme ayant une distribution bimodale ou s'il s'agit simplement d'une collection de valeurs aberrantes (asymétrie positive).

Step 3: Test d’hypothèse

TEST DE DISTRIBUTION BIMODALE

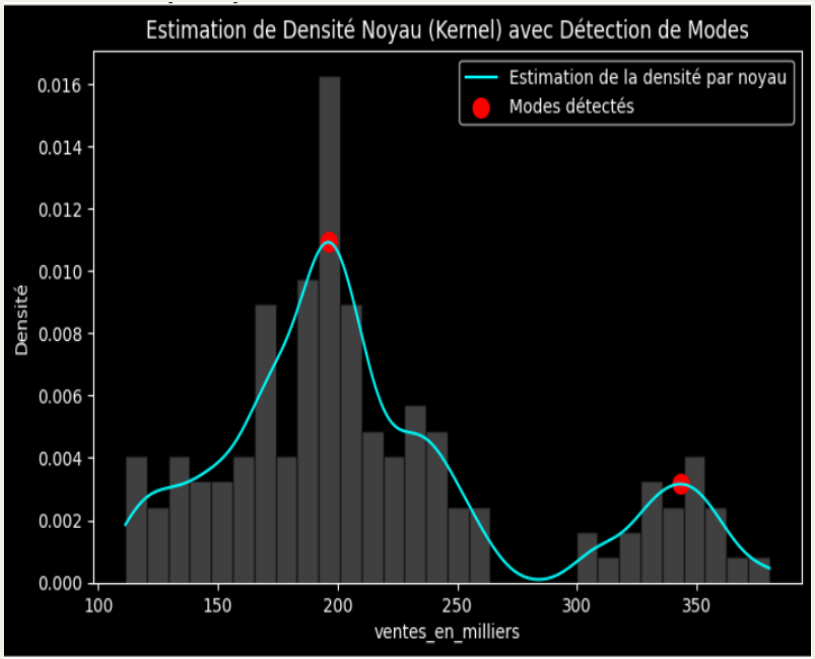
Test de l’utilisation du coefficient de bimodalité

- Si le coefficient de bimodalité est supérieur à 0,555 → distribution bimodale.
- Les résultats de l'analyse montrent que le coefficient de bimodalité (CB) : 0,571633. Ainsi, la distribution peut être classée comme bimodale.

Test de l’utilisation Silverman’s (KDE + Mode Detection)

- Si le nombre de modes est supérieur à un, la distribution peut être considérée comme bimodale. Si elle n'en a qu'un, alors la distribution est unimodale.
- Les résultats de l'analyse montrent que le mode = 2, donc les données peuvent être classées comme une distribution bimodale.
- Selon le graphique, les valeurs de la modalité apparaissent à deux points (196 et 343).

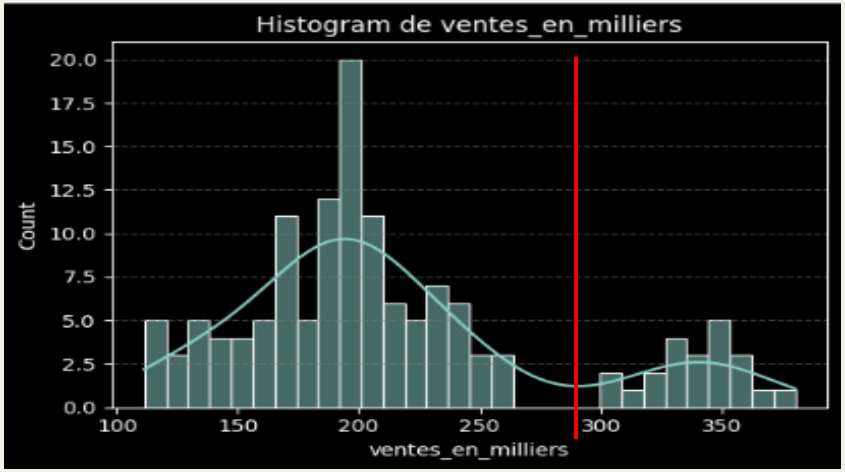
```
Coefficient de Bimodalité: 0.571633185838776
La distribution peut être classée comme bimodale
```



Step 3: Test d’hypothèse

SEPARATION DES DONNEES

- Nous utiliserons l'histogramme pour estimer le seuil entre deux groupes de données. Basé sur l'histogramme, le seuil des deux groupes est à 275 .
- Deux ensembles de données ont été formés (df_1 et df_2), chacun contenant respectivement 115 lignes et 22 lignes.



Data (df_1) shape: (115, 6)	
Promotion	Total
1	36
2	41
3	38

Data (df_2) shape: (22, 6)	
Promotion	Total
1	7
2	6
3	9

Step 3: Test d'hypothèse

NORMALITY TEST FOR EACH DATASET

- **Premier Dataset (df_1)**
  Shapiro-wilk test for df_1 **group 1: p-value = 0.049387**
  Shapiro-wilk test for df_1 **group 2: p-value = 0.000631**
  Shapiro-wilk test for df_1 **group 3: p-value =0.397674**
  Avec un niveau de signification (alpha) de 0,05, rejetez H_0 parce que le p-value est plus petit que alpha (pour le groupe 1 et le groupe 2), indiquant que les données ne suivent pas une distribution normale.

- **Second Dataset (df_2)**
  Shapiro-wilk test for df_2 **group 1: p-value = 0.233013**
  Shapiro-wilk test for df_2 **group 2: p-value = 0.075242**
  Shapiro-wilk test for df_2 **group 3: p-value =0.996028**
  Avec un niveau de signification (alpha) de 0,05, nous ne rejetons pas H_0 parce que le p-value est supérieur à alpha (pour tous les groupes), indiquant que les données suivent une distribution normale.

TEST D'HOMOGENEITE

Hypotheses
H_0: Les données sont homogènes.
H_1: Les données ne sont pas homogènes.

Niveau significatif (alpha)
5% (0.05)

Levene Test
Levene Test df_1 **(1er dataset): p-value =0.799583**
Levene Test df_2 **(2nd dataset): p-value =0.611496**

Resultats
Échec de rejet de H_0 car p-value est supérieure à alpha (pour l'ensemble des données), indiquant que les données sont homogènes.

Step 4: Test de Difference

PREMIER DATASET: KRUSKAL WALLIS TEST

Hypotheses
H_0: Il n'y a aucune différence dans la médiane/de la distribution des ventes pour chaque type de promotion.

H_1: Au moins un type de promotion a une médiane/de la distribution des ventes différente.

Niveau Significatif (alpha)
5% (0.05)

Kruskal-Wallis Test
Kruskal-Wallis Test: statistic = 20.838426, p-value = 0.0000299

Resultats

Rejeter l'hypothèse nulle (H_0) car la valeur p < alpha (0,05). Donc, au moins un type de promotion a une médiane/distribution des ventes différente.

SECOND DATASET: ANOVA TEST

Hypothesis
H_0: Il n'y a pas de différence dans les ventes moyennes pour chaque type de promotion.

H_1: Au moins un type de promotion a des ventes moyennes différentes.

Niveau Significatif (alpha)
5% (0.05)

Kruskal-Wallis Test
ANOVA Test: statistic = 22.788589, p-value = 0.000009

Result
Rejetez l'hypothèse nulle (H_0) car la p-value < alpha (0,05). Donc, au moins un type de promotion a des ventes moyennes différentes.

Step 5: Post-Hoc Test

PREMIER DATASET: DUNN TEST

	1	2	3
1	1	0.000021	0.297623
2	0.000021	1	0.013036
3	0.297623	0.013036	1

D'après le tableau de sortie ci-dessus, on peut conclure que :

- Les types de promotion 1 et 2 ont une différence significative résultats de vente avec un p-value de 0.000021.
- Les types de promotion 1 et 3 ont une différence significative résultats de vente avec un p-value de 0.297623.
- Les types de promotion 2 et 3 ont une différence significative résultats de vente avec un p-value de 0.013036.

SECOND DATASET: TUKEY'S HSD TEST

Group 1	Group 2	Mean diff	p-value	Reject
1	2	-40.2202	0	TRUE
1	2	-18.9019	0.0064	TRUE
2	3	21.3183	0.0035	TRUE

D'après le tableau de sortie ci-dessus, on peut conclure que:

- Les types de promotion 1 et 2 ont une différence significative résultats de vente avec un p-value de 0.0.
- Les types de promotion 1 et 3 ont une différence significative résultats de vente avec un p-value de 0.0064.
- Les types de promotion 2 et 3 ont une différence significative résultats de vente avec un p-value de 0.0035.

Step 5: Post-Hoc Test

Comparaison entre les groupes

Plus le rang moyen, la médiane et la moyenne sont élevés, meilleure est la performance des ventes. Cela indique que la stratégie promotionnelle mise en œuvre a été couronnée de succès.

Les deux tableaux ci-dessous montrent des résultats cohérents, à savoir que **le type de promotion 1 a le rang moyen, la médiane et la moyenne** les plus élevés par rapport aux autres types de promotion, suivi du type de promotion 3 en deuxième position et du type de promotion 2 en troisième position.

PREMIER DATASET

promotion	Mean Rank	Median
1	74.416667	211.815
2	40.219512	183.89
3	61.631579	196.33

SECOND DATASET

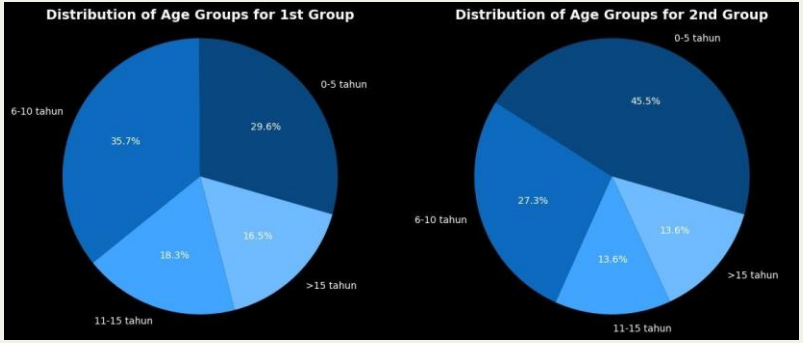
promotion	Average Sales
1	232.396047
2	189.31766
3	221.457872

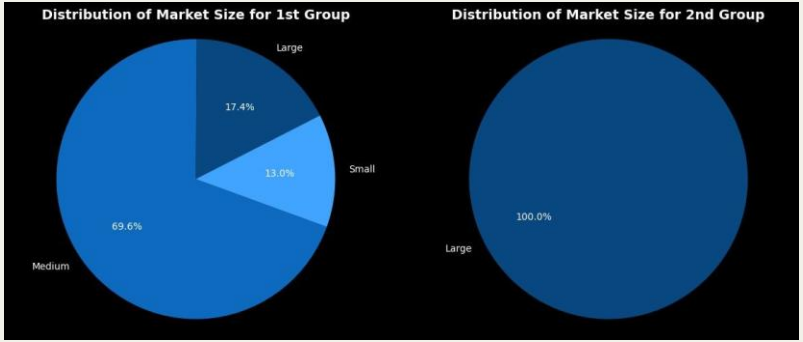
PERSPECTIVE : COMPARAISON ENTRE GROUPES DE DONNEES

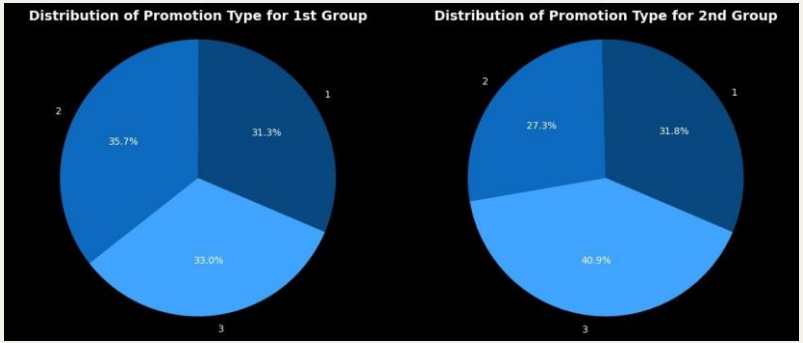
Les résultats de l'analyse ont identifié deux groupes de données. Le graphique à côté illustre les différences entre les deux groupes en fonction de la répartition de l'âge des magasins, de la taille des magasins et des promotions mises en œuvre.

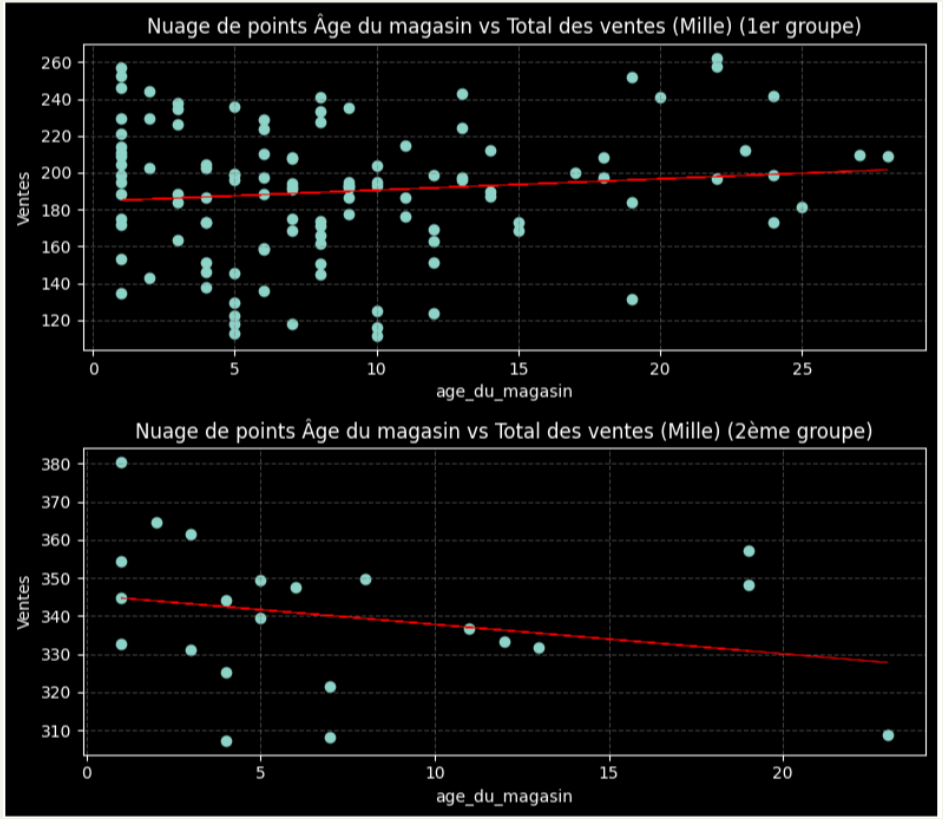
- En fonction de l'âge du magasin, le premier groupe se compose principalement de magasins âgés de 6 à 10 ans, tandis que le deuxième groupe est dominé par des magasins âgés de 0 à 5 ans.
- Sur la base de la taille du marché, le premier groupe est dominé par des marchés moyens, tandis que le second groupe est entièrement composé de grands marchés.
- En fonction du type de promotion, la mise en œuvre des stratégies promotionnelles dans le premier groupe est relativement répartie de manière uniforme entre les trois types. Cependant, dans le deuxième groupe, le type de promotion 3 est le plus dominant.

En général, le deuxième groupe se caractérise par de nouveaux grands marchés qui mettent fortement en œuvre le type de promotion 3. Cependant, le premier groupe montre plus de variété en termes d'âge des magasins et de type de promotion, les marchés moyens étant la majorité.









APERCU: COMPARAISON ENTRE GROUPES DE DONNEES

Lorsqu'on examine la relation entre l'âge du magasin et la performance des ventes, **la majorité des magasins dans les deux groupes se situent dans la tranche d'âge de 0 à 10 ans**, avec seulement de légères différences observées entre les deux groupes.

1. Le premier groupe montre une tendance à la hausse**, ce qui signifie que plus le magasin est ancien, meilleures sont les performances de vente** de ses nouveaux produits.

2. En revanche, le deuxième groupe présente une tendance à la baisse, indiquant que **les nouveaux magasins ont tendance à générer des ventes** plus élevées. Plus le magasin est jeune, meilleure est la performance de vente de ses nouveaux produits. Une constatation intéressante dans le deuxième groupe est que les ventes les plus élevées ont été générées par les nouveaux magasins.

Conclusions

- Sur la base de l'analyse EDA, il est évident que le **Grand Marché génère les ventes moyennes les plus élevées** par rapport aux autres tailles de marché.
- Concernant la performance promotionnelle :
  - **Le premier type de promotion a effectivement suscité l'intérêt des clients**, entraînant les ventes moyennes les plus élevées au cours des quatre dernières semaines. Cette performance n'est que légèrement différente de celle du troisième type de promotion.
  - En revanche, **le deuxième type de promotion affiche des chiffres de vente beaucoup plus bas** que les deux autres types de promotion.

- Les informations dérivées de la corrélation entre l'âge du magasin et la performance des ventes sont les suivantes :
  - Il existe un groupe de magasins relativement récents **(en particulier le nouveau Grand Marché) qui ont réussi à enregistrer des ventes élevées**. **Une analyse plus approfondie est nécessaire** pour déterminer les facteurs derrière cette réalisation, tels que :
    a) 'Des stratégies de marketing efficaces': Un type de promotion qui introduit et popularise avec succès le nouveau produit dans cet endroit spécifique.
    b) 'Innovation de produit réussie': Un nouveau produit qui s'aligne avec les intérêts des clients dans cette localité.
    c) 'Facteurs externes favorables': Facteurs externes qui causent des anomalies (comme des performances de vente exceptionnellement élevées).

    Les exemples incluent des emplacements de magasin stratégiques, l'impact des événements locaux, des moments spécifiques (longs congés), etc.

    Il y a plusieurs **anciens magasins qui contribuent encore de manière significative aux ventes** et qui devraient donc être maintenus.

Prochaines Etapes

Domaine de recherche

Réaliser des recherches **en prenant en compte des facteurs externes** qui peuvent potentiellement influencer les résultats, tels que les différences de localisation qui entraînent des variations dans les conditions géographiques et démographiques, des tailles d'échantillons différentes entre les groupes qui peuvent impacter la validité de l'analyse, et éviter d'autres facteurs externes tels que le temps (par exemple, des jours commémoratifs spécifiques, des événements locaux, etc).

Domaine d'activité

- **Mettre en œuvre des stratégies promotionnelles similaires** qui se sont révélées efficaces dans cette étude pour la promotion de nouveaux produits à l'avenir (tout en tenant compte des changements de tendances).
- **Réévaluer les ventes de nouveaux produits dans les magasins ayant une faible performance de vente** , car les produits pourraient ne pas être demandés. Réaliser des recherches sur le développement de nouveaux produits adaptés aux tendances de la région.
- Continuer à **surveiller les ventes de nouveaux produits dans les nouveaux magasins** ayant de bonnes performances de vente pour déterminer si le succès est dû à des facteurs externes ou si les produits sont vraiment populaires auprès des clients.
  a) 'Innovation de produit réussie': Introduire continuellement de nouveaux produits ou des produits uniques pour maintenir l'intérêt des clients et attirer une clientèle plus large.
  b) 'Facteurs externes favorables': Les exemples incluent des emplacements stratégiques de magasins, l'impact d'événements locaux ou d'autres circonstances avantageuses. Ceux-ci devraient être exploités pour amplifier la notoriété du produit, peut-être en collaborant avec des influenceurs pour produire du contenu engageant, ou en organisant des concours liés au produit pour améliorer la visibilité.

MERCI

Chaque octet de données renferme une histoire. Plongez profondément pour découvrir, analyser et créer avec confiance.

N’DOUBA MÔH-ADJONLIN J.C, Passionné de données.