

# 1. Problématique - Construction d'hypothèses

**Quelle est l'efficacité des différents canaux marketing sur les ventes de différents types de produits ?**

Hypothèse :

Les réseaux sociaux sont plus efficaces pour les produits ciblant les jeunes adultes, tandis que la télévision est plus adaptée pour les produits destinés à un public plus âgé.

Les campagnes marketing de longue durée sont plus efficaces pour générer des ventes que les campagnes courtes,

## 2. Préparation et Analyse Exploratoire des Données

- Nettoyage des Données :

Script python ok

- Statistiques Descriptives :

### Dataset Clients

- Âge : L'âge des clients varie de 18 à 69 ans, avec une moyenne d'environ 43.5 ans.
- Revenu Annuel : Le revenu annuel oscille entre 20,000.56 et 99,997.23, avec une moyenne d'environ 60,101.18.

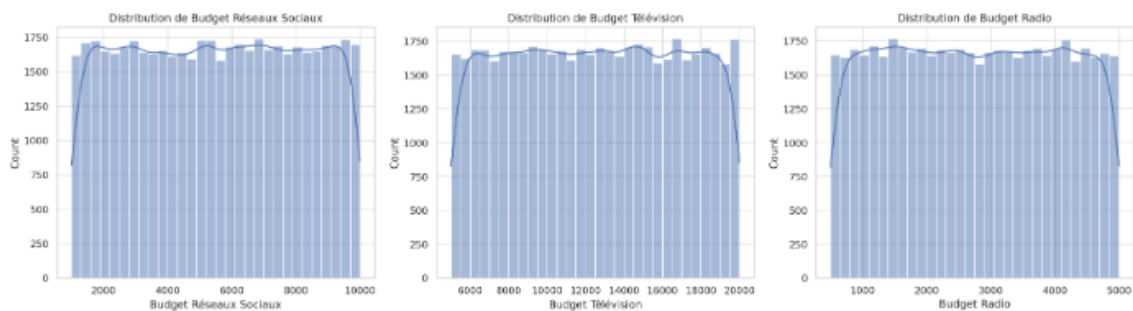
### Dataset Marketing

- Budget Réseaux Sociaux : Le budget pour les réseaux sociaux varie de 1,000.02 à 9,999.81, avec une moyenne d'environ 5,511.89.
- Budget Télévision : Le budget pour la télévision s'étend de 5,000.22 à 19,998.75, avec une moyenne d'environ 12,511.24.
- Budget Radio : Le budget pour la radio va de 500.01 à 4,999.98, avec une moyenne d'environ 2,747.40.
- Durée Campagne (Mois) : La durée des campagnes marketing varie de 1 à 11 mois, avec une moyenne d'environ 6 mois.
- Ventes : Les ventes générées par les campagnes marketing varient considérablement, de -1,302.24 (ce qui pourrait indiquer des retours ou

des remboursements) à 19,509.86, avec une moyenne d'environ 8,361.24.

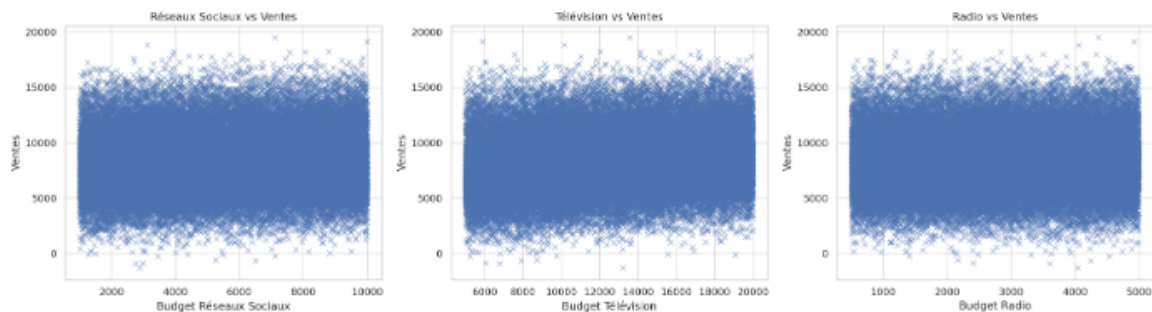
- Visualisation : Créez des histogrammes, des graphiques à barres, et des diagrammes de dispersion pour explorer visuellement les relations entre les variables.

Objectif : Visualiser la distribution des budgets alloués aux différents canaux marketing.



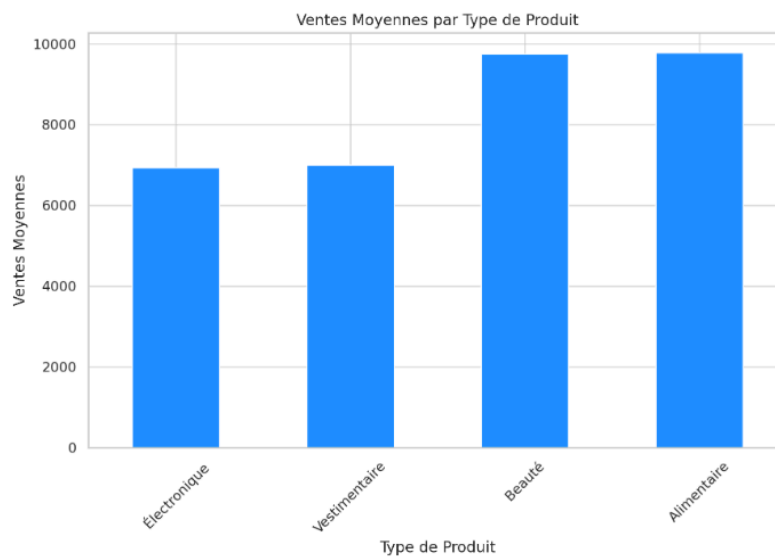
Les histogrammes offrent une vue d'ensemble sur la diversité des approches budgétaires adoptées dans le marketing. Ils révèlent que les entreprises varient significativement leurs investissements à travers les canaux de marketing, sans privilégier fortement un niveau de dépense spécifique.

Objectif : Examiner s'il existe une relation visible entre le budget alloué à chaque canal marketing et les ventes générées.



Les diagrammes de dispersion indiquent que, pour les trois canaux marketing analysés, il n'y a pas de corrélation linéaire forte et évidente entre le montant du budget alloué et les ventes générées. Cela suggère que d'autres facteurs, non

représentés dans ces graphiques, pourraient influencer l'efficacité des dépenses marketing.



### 3. Analyse de Corrélation et Tests Statistiques

#### Analyse de Corrélation

L'analyse de corrélation mesure la force et la direction de la relation linéaire entre deux variables continues. Les valeurs de corrélation (coefficient de Pearson) varient entre -1 et 1, où :

- 1 indique une corrélation positive parfaite (lorsque l'une augmente, l'autre aussi systématiquement),
- -1 indique une corrélation négative parfaite (lorsque l'une augmente, l'autre diminue systématiquement),
- 0 indique aucune corrélation linéaire.

Résultats de Corrélation :

- Réseaux Sociaux (0.059) : Cette valeur suggère une corrélation positive très faible entre le budget des réseaux sociaux et les ventes. Bien que statistiquement significative, la force de cette relation est si faible qu'elle implique que d'autres facteurs non mesurés pourraient être responsables des variations des ventes.
- Télévision (0.168) : Une corrélation positive modérée avec les ventes indique que les augmentations du budget de télévision sont relativement plus susceptibles de se traduire par des augmentations des ventes par rapport aux

réseaux sociaux ou à la radio. Cela suggère une efficacité plus élevée de la publicité télévisuelle dans la génération de ventes.

- Radio (0.027) : Comme pour les réseaux sociaux, la corrélation très faible avec les ventes indique peu d'effet direct du budget radio sur les ventes, soulignant la possibilité que l'efficacité de la radio en tant que canal de marketing soit limitée ou influencée par d'autres variables non incluses dans cette analyse.

### **Test statistiques :**

Le test ANOVA est utilisé pour comparer les moyennes de trois groupes ou plus pour déterminer si au moins un des groupes diffère significativement des autres en termes de moyenne. Le test produit deux valeurs principales :

- Statistique F : Mesure la variance entre les moyennes des groupes par rapport à la variance à l'intérieur des groupes. Une valeur F élevée indique une différence plus significative entre les moyennes des groupes.
- Valeur p : Indique la probabilité que les différences observées dans les moyennes des groupes soient dues au hasard. Une valeur p inférieure à 0.05 est généralement considérée comme significative.

Résultat ANOVA :

- $F = 5910.29$ ,  $p < 0.05$  : Indique des différences très significatives dans les ventes moyennes entre les types de produits. Cela suggère que certains types de produits sont nettement plus performants en termes de ventes que d'autres, ce qui peut orienter les décisions de marketing et de développement de produits.

## **5. Construction et Évaluation de Modèles avec Scikit-learn**

- Choix du Modèle :

Nous avons choisi d'utiliser un modèle de régression linéaire multiple pour aborder notre problématique, qui consiste à évaluer l'efficacité des différents canaux marketing sur les ventes. Cette décision repose sur plusieurs considérations importantes :

- Adéquation avec la Problématique : Notre objectif est de comprendre comment les investissements dans différents canaux marketing (réseaux sociaux, télévision, radio) peuvent influencer les ventes, une variable continue. La régression linéaire multiple est spécifiquement

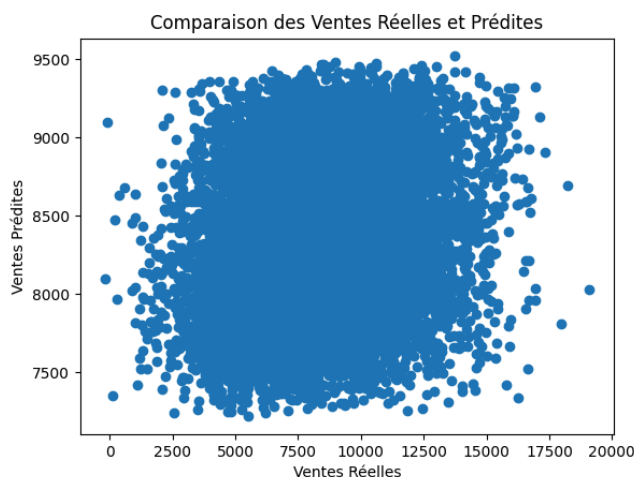
conçue pour examiner l'impact de plusieurs variables indépendantes sur une variable dépendante continue, ce qui correspond parfaitement à notre besoin d'analyse.

- Interprétabilité des Résultats : Nous apprécions la capacité de la régression linéaire multiple à fournir des insights clairs et interprétables sur la relation entre les variables. En particulier, les coefficients du modèle nous indiquent directement l'effet prévu sur les ventes d'une augmentation unitaire des budgets de chaque canal marketing, tout en tenant compte des autres canaux. Cette clarté est cruciale pour formuler des recommandations stratégiques basées sur les données.

Objectif: Explorer l'efficacité des différents canaux marketing (réseaux sociaux, télévision, radio) sur les ventes, en utilisant un modèle de régression linéaire multiple. La problématique consiste à déterminer comment les investissements dans ces canaux peuvent influencer les ventes d'un produit, permettant ainsi d'optimiser les stratégies marketing pour une efficacité maximale.

Script OK

Résultat :



RMSE: 2691.8549150172

$R^2$ : 0.031525358492048516

Distribution des Points :

- Les points sont dispersés sur une large gamme de valeurs de ventes réelles, mais les ventes prédites semblent être confinées à une plage plus étroite.
- Il n'y a pas de tendance claire ou de corrélation linéaire apparente indiquant que les prédictions correspondent étroitement aux valeurs réelles.

Écart entre les Ventes Réelles et Prédites :

- La concentration des points le long d'un espace horizontal suggère que le modèle n'est pas en mesure de capturer la variabilité des ventes réelles et tend à prédire des valeurs dans un intervalle restreint, quelle que soit la valeur réelle.
- Les prédictions ne semblent pas augmenter proportionnellement avec les ventes réelles, ce qui indiquerait un modèle bien ajusté.

Performances du Modèle :

- L'absence de corrélation visible et la dispersion des points suggèrent que le modèle de régression linéaire multiple n'est pas particulièrement efficace pour prédire les ventes.
- Le modèle pourrait ne pas être suffisamment complexe ou manquer de variables explicatives pertinentes pour capturer les dynamiques influençant les ventes.

## Conclusion

D'après le graphique, il est clair que notre modèle de régression linéaire multiple a des limites significatives dans sa capacité à prédire avec précision les ventes. Cela pourrait être dû à l'omission de variables influentes qui ne sont pas capturées par le budget marketing seul, comme les facteurs saisonniers, les tendances du marché, la qualité des campagnes, ou les préférences changeantes des consommateurs. Pour améliorer notre modèle, il serait judicieux d'explorer des modèles plus complexes qui peuvent intégrer des interactions entre variables, ou d'enrichir notre ensemble de données avec des informations supplémentaires susceptibles d'influencer les ventes.

Au terme de notre analyse approfondie, nous avons exploré l'efficacité des différents canaux marketing sur les ventes de divers types de produits. Nous avons initialement formulé deux hypothèses clés : premièrement, que les réseaux sociaux seraient plus efficaces pour les produits ciblant les jeunes adultes, et deuxièmement, que la télévision conviendrait mieux aux produits destinés à un public plus âgé. Nous avons également posé que les campagnes de marketing de plus longue durée seraient plus bénéfiques pour générer des ventes.

À travers notre préparation et notre analyse exploratoire des données, nous avons constaté une grande variété dans la distribution des budgets marketing et des ventes, sans tendance claire indiquant une efficacité particulière de l'un ou l'autre canal. Les diagrammes de dispersion et les analyses de corrélation ont souligné l'absence d'une corrélation linéaire forte entre les budgets marketing et les ventes, suggérant l'influence d'autres facteurs non examinés dans notre modèle initial.

Notre analyse de corrélation a révélé des liens très faibles entre les budgets des réseaux sociaux et des radios avec les ventes, tandis que la corrélation avec le budget télévision était légèrement plus marquée, mais toujours modeste. De plus, les tests ANOVA ont montré des différences significatives dans les performances des ventes entre les différents types de produits, ce qui implique que la nature du produit peut jouer un rôle crucial dans l'efficacité d'une campagne.

En construisant et en évaluant notre modèle de régression linéaire multiple avec Scikit-learn, nous avons constaté que, bien que nos prédictions se soient avérées peu précises, les résultats offraient une base sur laquelle construire. Le modèle a fourni des coefficients qui nous ont permis de saisir l'effet relatif de chaque canal marketing sur les ventes, mettant en évidence l'importance de considérer chaque canal dans le cadre d'une stratégie marketing intégrée.

Pour conclure, bien que nos analyses aient révélé des relations moins fortes que prévu entre les canaux marketing et les ventes, elles ont ouvert la voie à de nouvelles questions et à l'importance d'intégrer des données supplémentaires pour une compréhension plus complète.