

CERTIFICATION - DATA ANALYST - PRODUCT SALES



Validation et Nettoyage des Données

Pour garantir l'exactitude et la fiabilité de notre analyse, une étape cruciale de validation et de nettoyage des données a été effectuée sur le jeu de données `product_sales.csv`. Le DataFrame initial contenait 15 000 entrées et 8 colonnes.

Voici les étapes de validation et de nettoyage entreprises pour chaque colonne :

- **week (int64)** : Cette colonne représente la semaine de la vente depuis le lancement du produit. Après inspection, les valeurs varient de 1 à 6, ce qui est cohérent avec le lancement d'une nouvelle ligne de produits six semaines avant le début de l'analyse. Aucune valeur manquante ou anomalie n'a été détectée.
- **sales_method (object)** : Cette colonne catégorique indique la méthode de vente utilisée pour le client. Les valeurs attendues sont "Email", "Call" et "Email + Call". Aucune valeur manquante n'a été trouvée et les catégories sont conformes aux descriptions du projet.
- **customer_id (object)** : Cette colonne contient un identifiant unique pour chaque client. Aucune valeur manquante ou doublon n'a été détecté, assurant l'unicité de chaque entrée client.
- **nb_sold (int64)** : Représente le nombre de nouveaux produits vendus. Les valeurs sont des entiers non nuls, allant de 7 à 16. Aucune anomalie ni valeur manquante n'a été identifiée.
- **revenue (float64)** : Cette colonne contient le revenu généré par les ventes.
 - **Problème détecté** : Lors de l'inspection initiale, 1074 valeurs manquantes (NaN) ont été identifiées dans cette colonne.
 - **Action de nettoyage** : Étant donné que le revenu est une métrique essentielle pour cette analyse et que les valeurs manquantes représentaient environ 7.16% du jeu de données total, ces lignes ont été supprimées afin d'éviter d'introduire des biais par l'imputation et d'assurer l'intégrité des calculs de revenu.
 - **Résultat** : Le DataFrame final contient désormais 0 valeur manquante dans la colonne `revenue`.
- **years_as_customer (int64)** : Cette colonne indique le nombre d'années pendant lesquelles le client a été fidèle à l'entreprise.
 - **Problème détecté** : Bien que toutes les valeurs soient présentes (non nulles), l'inspection initiale a révélé une valeur maximale de 63 ans, ce qui est supérieur à l'âge de l'entreprise (fondée en 1984, soit environ 40 ans d'existence en date de ce projet). Des valeurs de 47 et 63 ans ont été identifiées comme des anomalies potentielles.
 - **Action de nettoyage** : Les deux lignes présentant une ancienneté supérieure à 40 ans ont été filtrées et exclues du jeu de données pour corriger ces erreurs de saisie manifestes.
 - **Résultat** : La valeur maximale de `years_as_customer` est maintenant de 39 ans, ce qui est cohérent avec la durée d'existence de l'entreprise.
- **nb_site_visits (int64)** : Cette colonne représente le nombre de visites du client sur le site web au cours des 6 derniers mois. Aucune valeur manquante ou anomalie n'a été détectée.
- **state (object)** : Cette colonne contient l'emplacement géographique du client (état). Aucune valeur manquante n'a été identifiée.

Après ces étapes de nettoyage, le DataFrame utilisé pour l'analyse (`df_final`) contient **13924 lignes complètes et validées**, assurant une base solide pour les analyses exploratoires à venir.

Analyse Exploratoire des Données (EDA)

L'analyse exploratoire des données vise à comprendre les caractéristiques principales du jeu de données et à répondre aux questions posées par le représentant des ventes.

Nombre de clients par méthode de vente

Pour commencer, nous avons analysé le nombre de clients uniques ayant participé à chaque méthode de vente testée pour la nouvelle ligne de produits. Initialement, des incohérences mineures dans les noms des méthodes de vente ('em + call' , 'email') ont été identifiées et corrigées en les standardisant respectivement en 'Email + Call' et 'Email' .

Après standardisation, la répartition des clients par méthode est la suivante :

Méthode de Vente	Nombre de Clients Uniques
Email	6921
Call	4780
Email + Call	2223

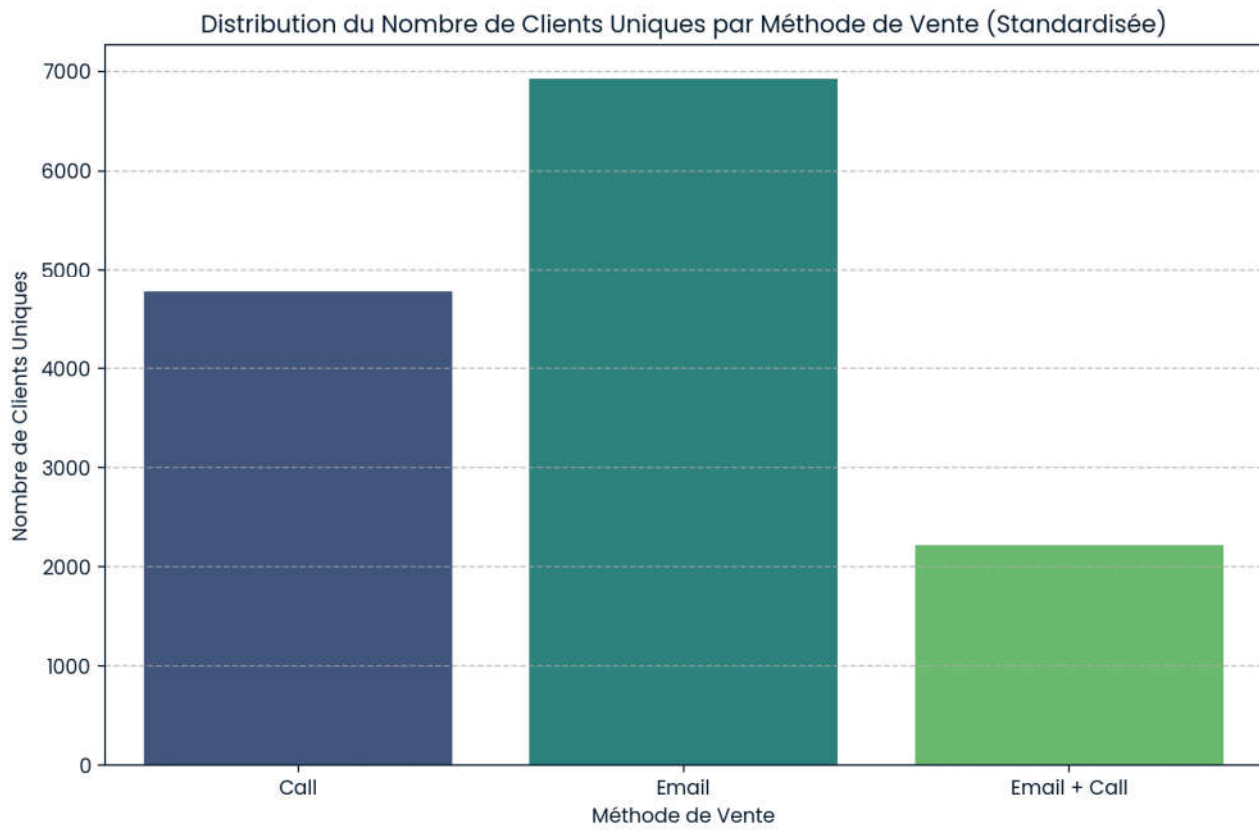
Observations :

- La méthode '**Email**' a attiré le plus grand nombre de clients uniques, avec 6921 participants.
- La méthode '**Call**' arrive en deuxième position avec 4780 clients.
- La méthode combinée '**Email + Call**' a le plus petit nombre de clients, avec 2223 participants.

Ces chiffres suggèrent que les approches 'Email' et 'Call' sont celles qui touchent le plus grand volume de clients, avec l'email étant la méthode la plus répandue.

Visualisation :

Le graphique à barres ci-dessous illustre visuellement cette distribution :



Répartition du Revenu Global et par Méthode de Vente

Nous avons ensuite examiné la distribution des revenus (`revenue`) pour comprendre la performance financière des ventes de la nouvelle ligne de produits.

Distribution Globale du Revenu :

Les statistiques descriptives du revenu global sont les suivantes :

- **Nombre d'observations** : 13924
- **Moyenne** : \$93.94
- **Écart-type** : \$47.44
- **Minimum** : \$32.54
- **25ème percentile** : \$52.47
- **Médiane (50ème percentile)** : \$89.50
- **75ème percentile** : \$107.33
- **Maximum** : \$238.32

L'histogramme de la distribution globale du revenu révèle une distribution multi-modale, avec des pics de fréquence distincts. Cela suggère qu'il existe différents niveaux de revenu générés par les ventes, potentiellement liés aux différentes méthodes de vente ou aux types de produits vendus.

Visualisation (graphique univarié) :

Répartition du Revenu par Méthode de Vente :

Pour approfondir, nous avons analysé les statistiques de revenu pour chaque méthode de vente :

Méthode de Vente	Nombre (count)	Moyenne du Revenu	Écart-type du Revenu	Médiane du Revenu	Max Revenu
Call	4780	\$47.60	\$8.61	\$49.05	\$71.36
Email	6921	\$97.13	\$11.21	\$95.58	\$148.97
Email + Call	2223	\$183.65	\$29.08	\$184.74	\$238.32

Observations clés du revenu par méthode :

- La méthode '**Email + Call**' génère le revenu moyen par vente le plus élevé (183.65). Elle présente également la plage de revenus la plus élevée (jusqu'à 238.32).
- La méthode '**Email**' se situe en deuxième position avec un revenu moyen de \$97.13, ce qui est près du double de la méthode '**Call**'.
- La méthode '**Call**' a le revenu moyen par vente le plus bas (\$47.60).

Ces différences sont illustrées de manière frappante par les box plots, où les boîtes sont clairement séparées et classées en fonction de leur niveau de revenu. La méthode '**Email + Call**' se distingue nettement par des revenus bien supérieurs.

Visualisation (graphique multivarié) :

image-1.png

Évolution du Revenu Moyen au fil du Temps par Méthode de Vente

Pour évaluer la performance durable de chaque approche, nous avons analysé l'évolution du revenu moyen généré par semaine pour chaque méthode de vente sur une période de six semaines.

Le tableau ci-dessous présente le revenu moyen par semaine pour chaque méthode :

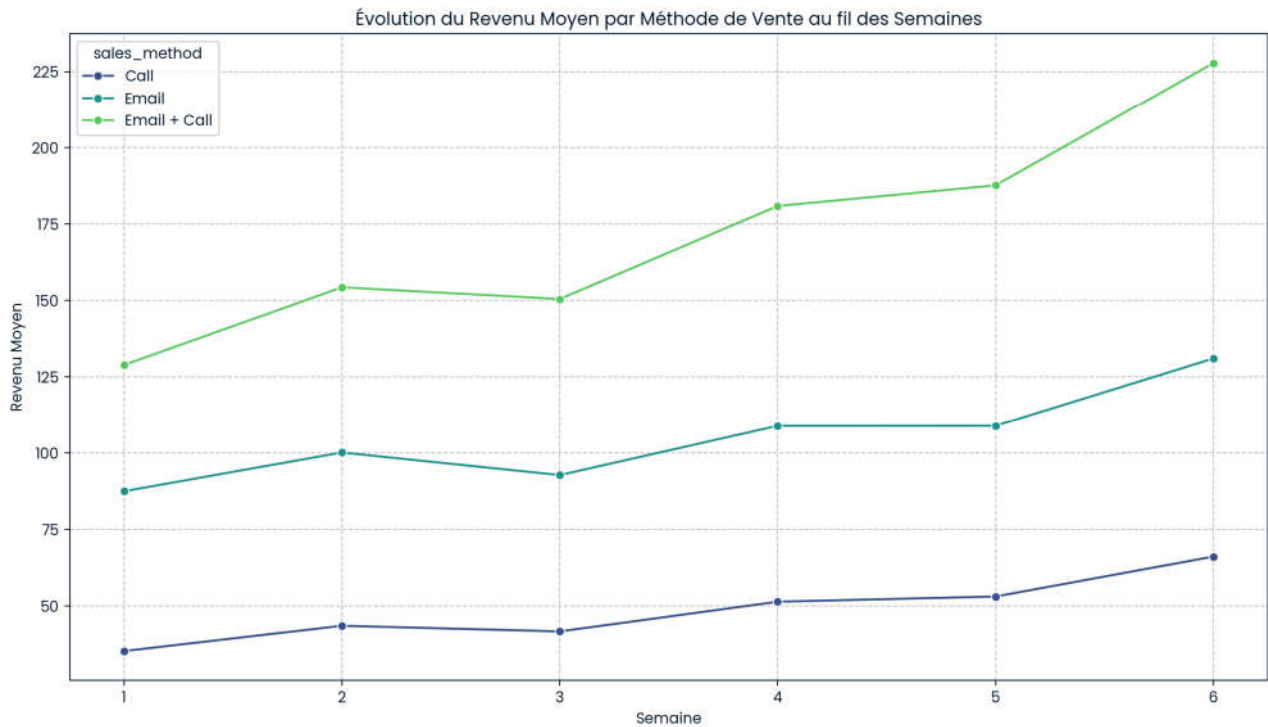
Semaine	Méthode de Vente	Revenu Moyen
1	Call	\$35.35
1	Email	\$87.50
1	Email + Call	\$128.90
2	Call	\$43.60
2	Email	\$100.14
2	Email + Call	\$154.25
3	Call	\$41.76
3	Email	\$92.76
3	Email + Call	\$150.42
4	Call	\$51.45
4	Email	\$108.88
4	Email + Call	\$180.82
5	Call	\$53.15
5	Email	\$108.89
5	Email + Call	\$187.59
6	Call	\$66.17
6	Email	\$130.98
6	Email + Call	\$227.77

Observations clés de l'évolution temporelle :

- La méthode '**Email + Call**' démontre une performance remarquable, non seulement en générant le revenu moyen par vente le plus élevé dès la première semaine, mais aussi en affichant une **croissance constante et la plus significative** sur les six semaines. Son revenu moyen passe de 128.90 en Semaine 1 à 227.77 en Semaine 6.
- La méthode '**Email**' montre également une tendance positive, avec une augmentation progressive du revenu moyen de 87.50 en Semaine 1 à 130.98 en Semaine 6.
- La méthode '**Call**', bien que générant les revenus moyens les plus bas, affiche aussi une amélioration au fil du temps, passant de 35.35 en Semaine 1 à 66.17 en Semaine 6.

En conclusion, toutes les méthodes ont montré une amélioration de leur revenu moyen au fil du temps, mais la méthode 'Email + Call' se démarque clairement comme la plus performante et la plus prometteuse en termes de croissance de revenu.

Visualisation (graphique multivarié) :



Recommandations et Définition de la Métrique Commerciale

Sur la base de l'analyse exploratoire des données, nous pouvons formuler des recommandations stratégiques pour l'équipe de vente et proposer une métrique clé pour le suivi de la performance.

Recommandation de Méthode de Vente

L'objectif est d'identifier la méthode qui non seulement génère des revenus, mais le fait de manière efficace, en tenant compte du temps investi par l'équipe.

- **Méthode 'Email + Call' :**
 - **Revenu Moyen :** Cette méthode génère de loin le revenu moyen par vente le plus élevé (\$183.65).
 - **Évolution Temporelle :** Elle montre une croissance constante et la plus significative du revenu moyen sur les six semaines.
 - **Temps Investi :** Bien qu'elle implique un appel, celui-ci est relativement court (environ 10 minutes par client).
- **Méthode 'Email' :**
 - **Revenu Moyen :** Elle génère un revenu moyen respectable (\$97.13).
 - **Temps Investi :** L'avantage majeur est le "très peu de travail pour l'équipe" requis, ce qui la rend extrêmement efficace en termes de ressources humaines.
 - **Volume :** Elle attire le plus grand nombre de clients (6921).
- **Méthode 'Call' :**
 - **Revenu Moyen :** Cette méthode affiche le revenu moyen par vente le plus bas (\$47.60).
 - **Temps Investi :** Elle est la plus coûteuse en temps (environ 30 minutes par client).

Conclusion des Recommandations :

1. **Prioriser 'Email + Call' :** Cette méthode est la plus performante en termes de revenu par transaction et de croissance. Elle devrait être la méthode privilégiée pour maximiser la valeur de chaque vente, même si elle demande un investissement en temps modéré. L'entreprise devrait envisager d'allouer plus de ressources à cette approche pour cibler les prospects à plus fort potentiel.
2. **Maintenir et Optimiser 'Email' :** En raison de sa faible exigence en temps et de sa capacité à toucher un large volume de clients avec un bon revenu moyen, la méthode 'Email' est idéale pour une couverture de marché étendue et pour la génération d'un grand nombre de leads à faible coût. Des efforts peuvent être concentrés sur l'optimisation du contenu des emails pour augmenter encore leur taux de conversion et le revenu par transaction.
3. **Réévaluer ou Cibler 'Call' :** La méthode 'Call' pure est la moins efficace en l'état actuel des données, en raison de son faible revenu moyen et de son coût élevé en temps. L'entreprise devrait soit la réévaluer (par exemple, en la réservant à des segments de clients très spécifiques ou à des produits à très haute marge), soit chercher des moyens de réduire drastiquement le temps d'appel ou d'augmenter le revenu moyen par appel.

Définition d'une Métrique pour l'Entreprise : Revenu par Heure de Travail Investi (RPTI)

Pour surveiller l'efficacité des méthodes de vente et guider les décisions futures, nous recommandons la métrique "Revenu par Heure de Travail Investi (RPTI)". Cette métrique permet d'évaluer la valeur générée pour chaque heure de travail consacrée par l'équipe de vente.

Comment la surveiller :

Le RPTI devrait être calculé et suivi sur une base régulière (par exemple, hebdomadaire ou mensuelle) pour chaque méthode de vente. L'entreprise peut :

- Fixer des objectifs de RPTI pour chaque méthode.
- Identifier les tendances : Une augmentation du RPTI indique une amélioration de l'efficacité.
- Comparer les performances : Le RPTI permet une comparaison directe de l'efficacité opérationnelle entre les différentes méthodes, indépendamment du volume de clients.
- Allouer des ressources : Un RPTI élevé pour une méthode justifie un investissement plus important en temps d'équipe.

Estimation des Valeurs Initiales pour le RPTI (sur la période de 6 semaines) :

Pour estimer le RPTI, nous avons besoin du revenu total généré par chaque méthode et du temps total approximatif investi.

- **Méthode 'Call' :**
 - Revenu Total : 4780 clients * 47.60/*client* = 227 608
 - Temps Total Estimé : 4780 clients * (30 minutes/client) = 4780 * 0.5 heure = 2390 heures
 - **RPTI (Call) :** $227608 / 2390 \text{ heures} = **95.23 / \text{heure}**$
- **Méthode 'Email + Call' :**
 - Revenu Total : 2223 clients * 183.65/*client* = 408 043
 - Temps Total Estimé : 2223 clients * (10 minutes/client) = 2223 * 0.1667 heure ≈ 370.5 heures
 - **RPTI (Email + Call) :** $408043 / 370.5 \text{ heures} = **1101.33 / \text{heure}**$
- **Méthode 'Email' :**
 - Le document spécifie que cette méthode "a nécessité très peu de travail pour l'équipe". Sans une estimation précise du temps par client, il est difficile de calculer un RPTI comparable. Cependant, il est clair que son efficacité en termes de temps investi est extrêmement élevée, faisant du Revenu Moyen par Client (\$97.13) et du volume de clients (6921) des indicateurs clés à surveiller.

Synthèse des RPTI Initiaux :

Méthode de Vente	Revenu Total (sur 6 semaines)	Temps Total Investi (heures)	RPTI (Revenu par Heure de Travail)
Call	\$227 608	2390	\$95.23 / heure
Email + Call	\$408 043	370.5	\$1101.33 / heure
Email	\$672 384	Très faible	Extrêmement élevé

Les valeurs initiales du RPTI confirment que la méthode '**Email + Call**' est de loin la plus rentable en termes de temps investi, générant plus de 10 fois plus de revenus par heure que la méthode

'Call'.