

Optimisation des Revenus des Smartphones : Analyse, Prévisions et Recommandations Stratégiques pour DataTel

Dans un contexte où la concurrence sur le marché des smartphones est de plus en plus intense, **DataTel**, opérateur télécom majeur en France, cherche à maximiser les performances de ses boutiques en optimisant la stratégie de vente de ses modèles phares de smartphones. Les revenus générés par ces modèles dépendent de multiples facteurs : les caractéristiques marketing, les indicateurs économiques, les conditions météorologiques, et même les spécificités locales propres à chaque ville.

Ce rapport vise à analyser les données historiques de ventes afin d'identifier les éléments déterminants qui influencent les performances commerciales et de développer un modèle prédictif robuste pour anticiper les revenus du premier trimestre 2025. Il présente une analyse approfondie des données, propose une modélisation des revenus et formule des recommandations stratégiques fondées sur les résultats obtenus. Ces démarches visent à renforcer la position concurrentielle de **DataTel** et à garantir un avantage décisif sur le marché des smartphones.

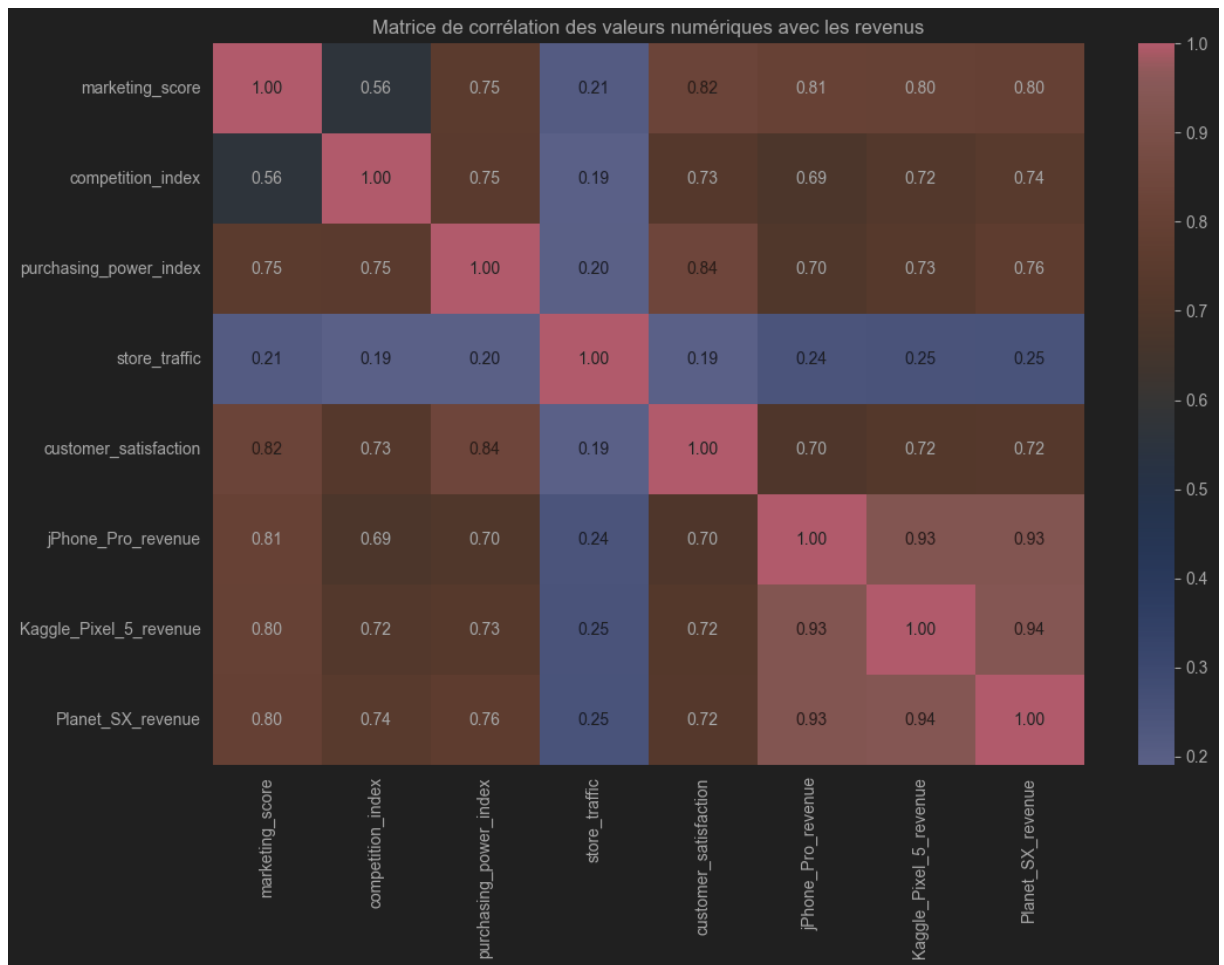
Les principaux insights business

L'analyse des données et des modèles prédictifs a révélé plusieurs points clés concernant les revenus des différents modèles de smartphones. Ces insights offrent une compréhension approfondie des facteurs ayant le plus grand impact sur les ventes et permettent de guider les décisions stratégiques :

1. Satisfaction client et marketing : des leviers majeurs

- La satisfaction client et le score marketing sont les deux facteurs les plus influents sur les revenus des trois modèles de smartphones (iPhone Pro, Kaggle Pixel 5, et Planet SX).
- Une forte satisfaction client se traduit directement par une augmentation des revenus, soulignant l'importance d'investir dans l'amélioration de l'expérience utilisateur.

- Les campagnes marketing bien conçues ont également montré un impact significatif sur les ventes.



Matrice de corrélation entre certaines variables et les revenus.

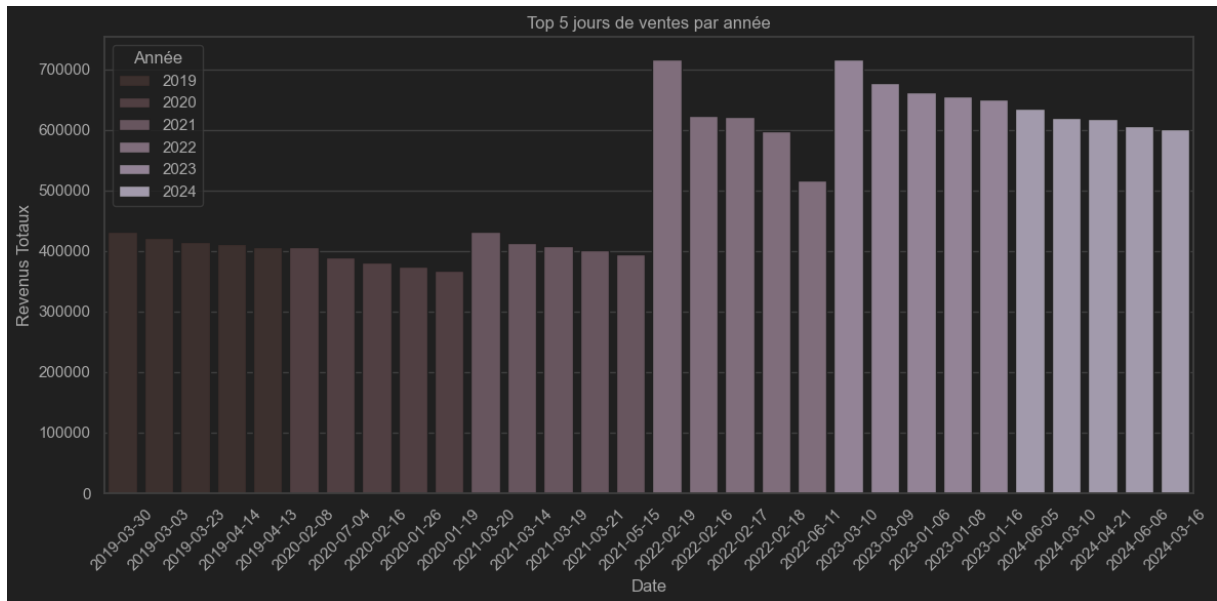
La matrice de corrélation nous montre bien :

- D'une part l'importance de certaines variables comme le pouvoir d'achat (environ 70%) et la satisfaction du client (environ 70%) montrant une forte influence sur les revenus.
- D'autre part une faible influence de certaines variables sur les revenus comme store traffic (24%)

2. L'importance des caractéristiques temporelles

- **Jour de la semaine et périodes spécifiques :**
 - Les jours de la semaine (notamment les week-ends) influencent les ventes de manière notable.

- Les revenus présentent des variations saisonnières, avec des pics observés durant certaines périodes de l'année (ex. : périodes festives ou lancements de nouveaux produits).

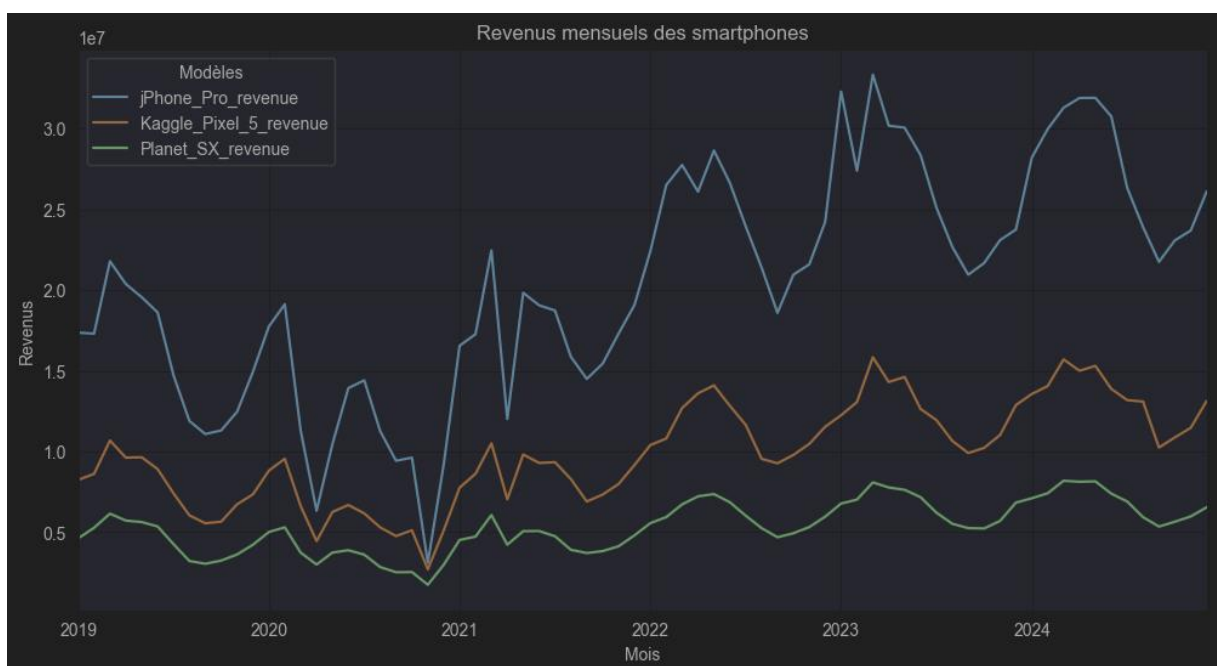


Graphique du Top 5 des jours de ventes par année entre 2019 et 2024

Ce graphique permet de visualiser clairement l'évolution des meilleurs jours par année. On remarque premièrement que les meilleures ventes s'améliorent au fil des années. Deuxièmement, Les meilleures ventes se situent très souvent dans le début de l'année.

- **Année et mois :**

- Les avancées technologiques, comme la transition vers des phases 5G plus matures, ont contribué à des hausses de revenus au fil du temps.

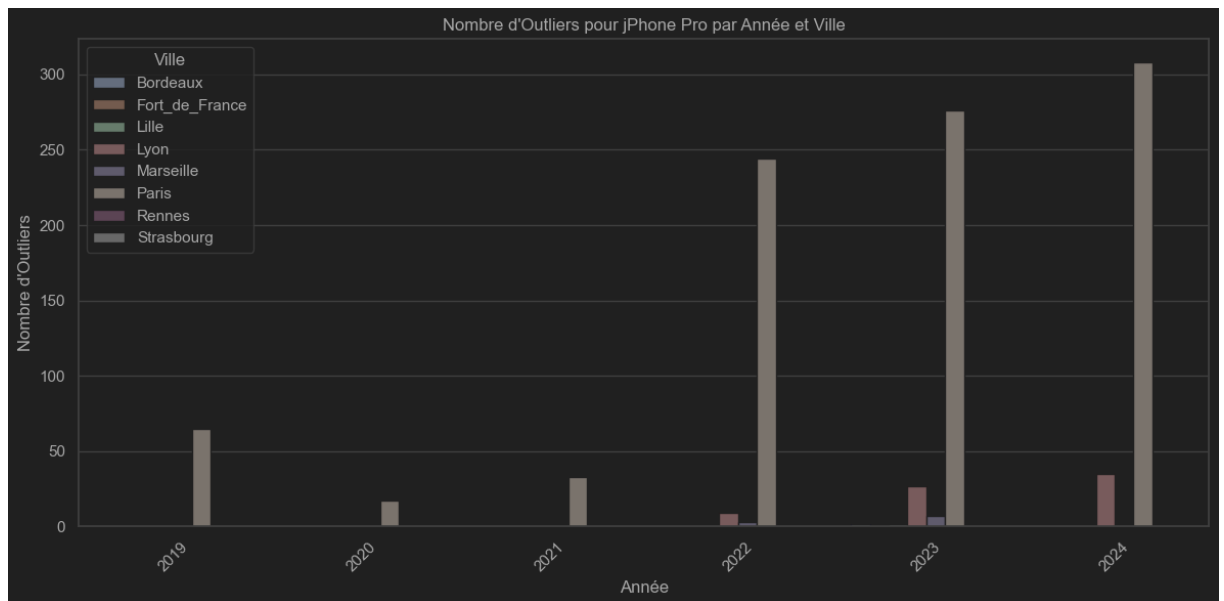


Revenus mensuels des smartphones entre 2019 et 2024

Ce graphique nous permet de voir les ventes mensuelles des trois modèles de téléphone. On remarque qu'au fil du temps, les ventes du iPhone_Pro s'améliorent, là où les performances des deux autres modèles semblent être plus constants. Il est également important de souligner un creux survenu en fin d'année 2020, qu'on associera à la période du covid-19.

3. Concentration géographique des valeurs aberrantes

- Une analyse des données a révélé une forte concentration de valeurs aberrantes à Paris.



Nombre d'outliers pour iPhone Pro par année et ville

Ce graphique met en évidence une concentration significative des valeurs aberrantes à Paris, avec une nette augmentation depuis 2022. Cette tendance peut être attribuée à des facteurs tels que la densité de population élevée, des événements spécifiques, ou encore une adoption rapide des nouvelles technologies, reflétant le pouvoir d'achat supérieur des consommateurs dans cette région.

4. Rôle des phases 5G

- Les phases de déploiement de la 5G jouent un rôle clé dans les ventes des smartphones :
 - La phase *5g_Mature-5G* est associée à des revenus plus élevés, indiquant une adoption accrue par les consommateurs lorsque cette technologie est dominante.
 - En revanche, les phases plus anciennes (*5g_Pre-5G* et *5g_Early-5G*) montrent un impact limité sur les ventes.

5. Effets des conditions météorologiques

- Les conditions météorologiques ont un effet modéré sur les ventes :
 - Les jours de "mauvais temps" (*weather_condition_Bad*) semblent légèrement décourager les achats en magasin.
 - Les jours de "bon temps" (*weather_condition_Good*) sont associés à une légère augmentation des ventes, particulièrement pour les modèles iPhone Pro et Kaggle Pixel 5.

Résumé des insights

Ces conclusions mettent en lumière l'importance de combiner des stratégies centrées sur le client (satisfaction et marketing), des optimisations géographiques (Paris comme zone stratégique) et une prise en compte des avancées technologiques et des dynamiques temporelles pour maximiser les revenus. Ces insights constituent une base solide pour affiner les prévisions et orienter les recommandations opérationnelles.

Méthodologie de Modélisation

Pour répondre à la problématique d'analyse et de prévision des revenus des smartphones, nous avons adopté une méthodologie rigoureuse et itérative incluant le prétraitement des données, la sélection des variables pertinentes, et l'entraînement de modèles prédictifs avancés.

1. Prétraitement des Données

- **Séparation par Ville** : Pour donner suite à l'identification d'une forte concentration de valeurs aberrantes à Paris, les données ont été séparées par ville. L'utilisateur peut choisir une ville spécifique pour une analyse personnalisée.
- **Traitement des Valeurs Manquantes** :
 - **Données numériques** : Les colonnes numériques manquantes ont été complétées par interpolation linéaire.
 - **Données catégoriques** : Les valeurs manquantes des colonnes catégoriques, comme les phases 5G ou les conditions météorologiques, ont été remplies avec les valeurs les plus fréquentes (*mode*) calculées pour chaque période mensuelle.
- **Ajout de Données Futures** : Pour anticiper les performances au premier trimestre 2025, des lignes ont été ajoutées pour chaque jour de cette période. Les données manquantes ont été interpolées ou remplies à partir des tendances observées dans les données historiques.

2. Sélection des Variables

Les variables explicatives ont été sélectionnées en fonction de leur pertinence statistique et de leur corrélation avec les revenus cibles. La liste des variables utilisées inclut des facteurs comme :

- **Variables numériques :**
 - marketing_score, competition_index, purchasing_power_index, store_traffic, customer_satisfaction.
- **Variables temporelles :**
 - year, month, day, day_of_week, week_of_year, is_weekend.
- **Variables catégoriques :**
 - weather_condition, 5g_phase, public_transport.

Ces variables capturent des dynamiques variées, telles que les tendances marketing, les comportements des clients, et les facteurs externes comme les conditions météorologiques.

3. Modélisation avec XGBoost

Pour la modélisation, nous avons voulu créer un modèle seulement pour la ville de Marseille, par manque de temps.

- **Modèle choisi :** Nous avons utilisé **XGBoost Regressor**, un algorithme puissant basé sur les arbres de décision, reconnu pour ses performances sur des données hétérogènes et complexes.
- **Optimisation des hyperparamètres :**
 - Les paramètres du modèle ont été optimisés à l'aide d'une validation croisée temporelle (TimeSeriesSplit) et de la méthode **GridSearchCV**.
 - Les hyperparamètres explorés incluent :
 - learning_rate: [0.01, 0.1, 0.2]
 - max_depth: [3, 5, 7]
 - n_estimators: [100, 200, 300]
 - subsample: [0.8, 1.0]

4. Interprétabilité avec SHAP

Pour interpréter les prédictions des modèles, nous avons utilisé **SHAP (SHapley Additive exPlanations)**. Cet outil fournit des explications détaillées sur l'impact de chaque variable sur les prédictions :

- **Graphiques récapitulatifs SHAP** : Ils montrent la distribution des impacts des variables sur les prédictions.
- **Graphiques en barres SHAP** : Ils illustrent l'importance moyenne des variables dans les modèles.

5. Évaluation des Modèles

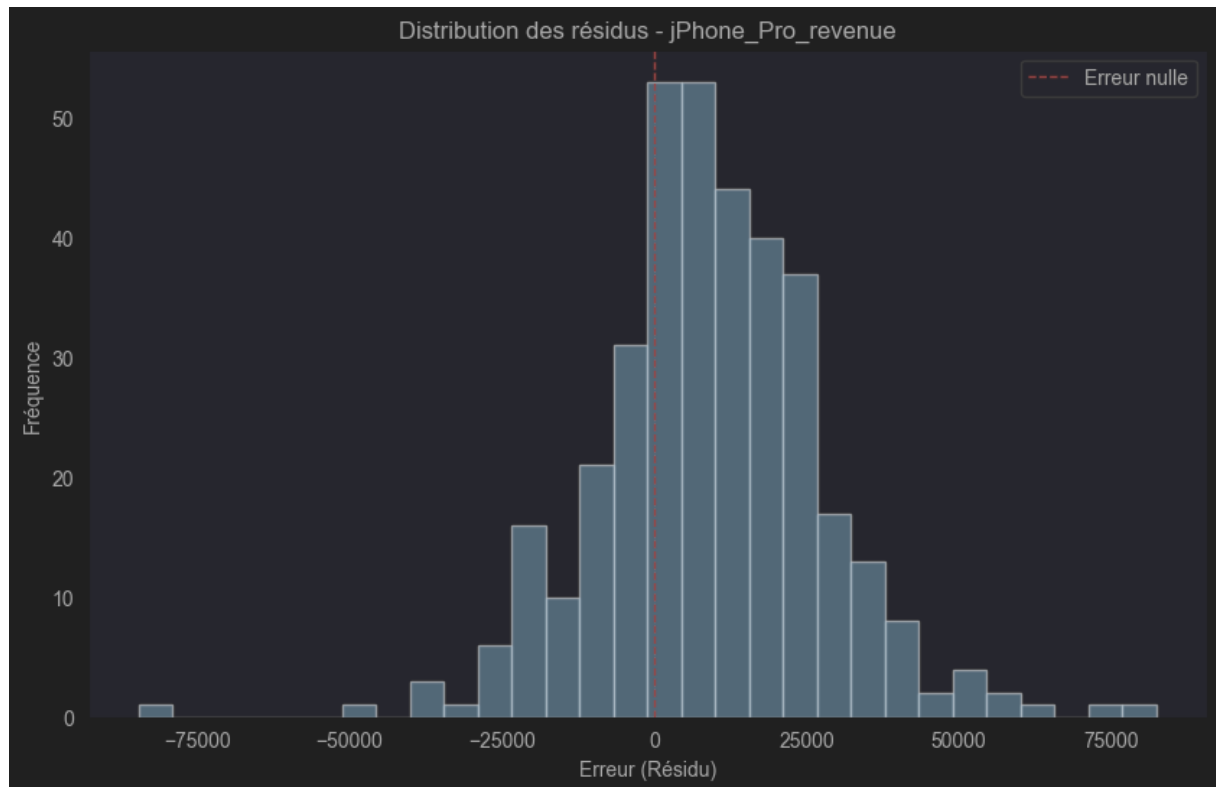
- Les modèles ont été évalués sur un ensemble de validation correspondant à l'année 2024. Les métriques utilisées sont :
 - **MAE** (Mean Absolute Error) : Mesure l'écart moyen entre les prédictions et les valeurs réelles.
 - **MSE** (Mean Squared Error) : Quantifie les erreurs moyennes au carré pour pénaliser les grandes erreurs.
 - **RMSE** (Root Mean Squared Error) : Représente la racine carrée de la MSE pour exprimer les erreurs dans les mêmes unités que les revenus.
 - **R² (Coefficient de détermination)** : Évalue la proportion de la variance expliquée par le modèle.

Évaluation des Performances des Modèles

1. iPhone Pro

- **Meilleurs paramètres du modèle :**
 - Taux d'apprentissage : 0.1
 - Profondeur maximale : 3
 - Nombre d'estimateurs : 100
 - Sous-échantillonnage : 1.0
- **Évaluation sur le jeu de validation :**
 - MAE : 15,784.83
 - MSE : 422,925,238.57
 - RMSE : 20,565.15
 - R² : 0.1
- **Distribution des résidus :**

- La distribution des résidus est symétrique, mais légèrement dispersée, avec des erreurs significatives atteignant $\pm 75\ 000$.
- Cela montre que le modèle capture globalement bien les tendances, mais qu'il existe encore des erreurs importantes, notamment pour certaines observations spécifiques.
- Une légère concentration autour de zéro indique que le modèle performe mieux pour la majorité des prédictions.



2. Kaggle Pixel 5

- **Meilleurs paramètres du modèle :**

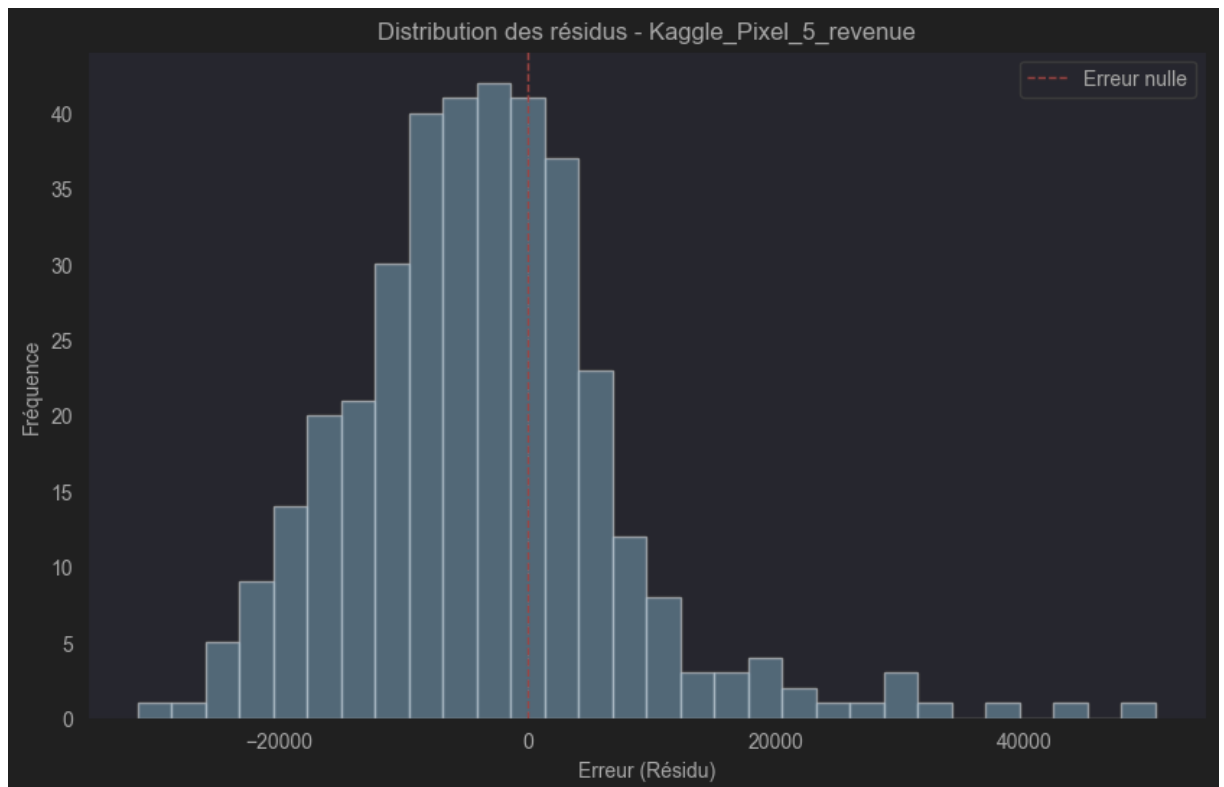
- Taux d'apprentissage : 0.1
- Profondeur maximale : 5
- Nombre d'estimateurs : 100
- Sous-échantillonnage : 1.0

- **Évaluation sur le jeu de validation :**

- MAE : 8,644.92
- MSE : 131,752,591.15
- RMSE : 11,478.35
- R^2 : 0.01

- **Distribution des résidus :**

- La distribution des résidus est relativement concentrée autour de zéro, mais moins symétrique que pour le iPhone Pro.
- Les erreurs sont principalement comprises entre -20 000 et +40 000, ce qui montre une certaine variabilité dans les prédictions.
- Une asymétrie notable suggère que le modèle tend à sous-estimer les revenus dans certains cas.



3. Planet SX

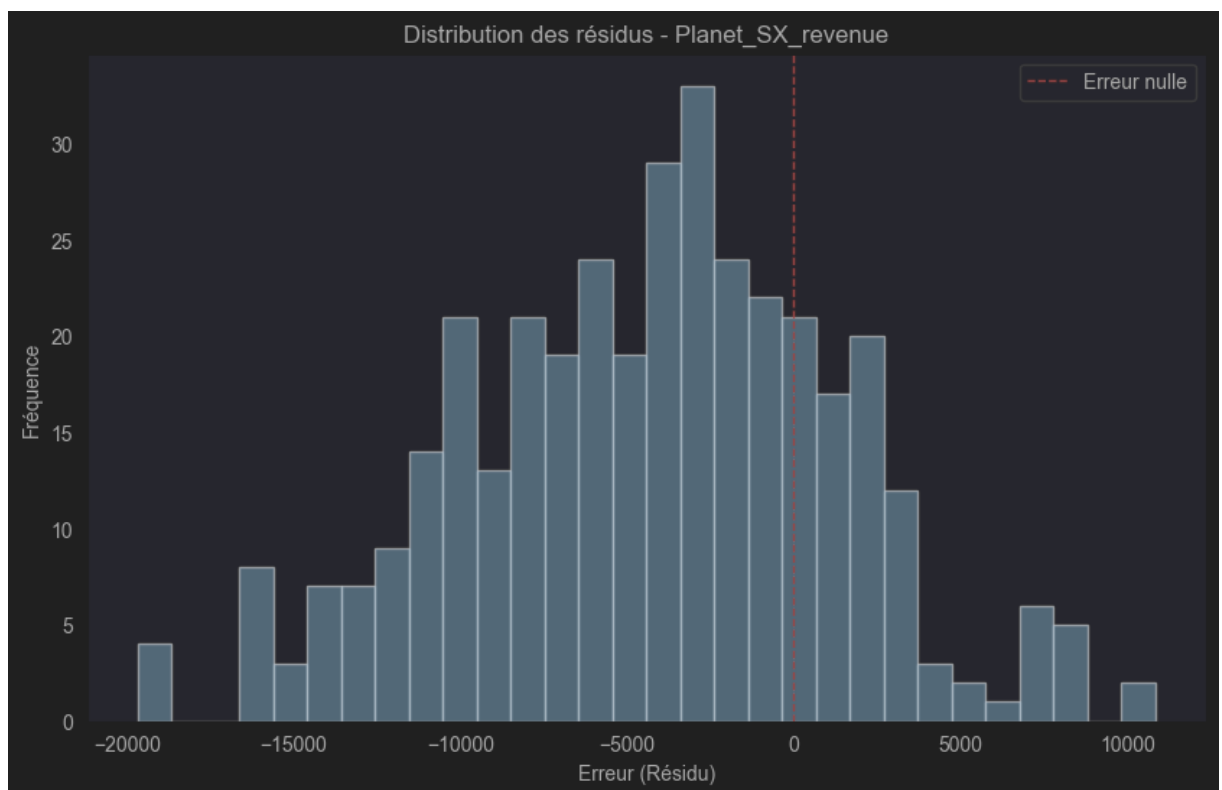
- **Meilleurs paramètres du modèle :**

- Taux d'apprentissage : 0.2
- Profondeur maximale : 3
- Nombre d'estimateurs : 200
- Sous-échantillonnage : 1.0

- **Évaluation sur le jeu de validation :**

- MAE : 5,691.13
- MSE : 50,698,980.43
- RMSE : 7,120.32

- R^2 : -0.39
- **Distribution des résidus :**
 - La distribution des résidus est moins centrée autour de zéro, avec une concentration plus élevée dans la plage de -10 000 à -5 000. Cela indique que le modèle a tendance à sous-estimer les revenus de manière systématique.
 - Les erreurs positives sont relativement rares, suggérant une faiblesse dans la prédiction des valeurs élevées.



Les résultats montrent que la performance globale des modèles est variable selon les cibles, avec des scores R^2 particulièrement faibles ou négatifs pour certains modèles (comme Planet SX).

Analyse SHAP : Importance des Caractéristiques

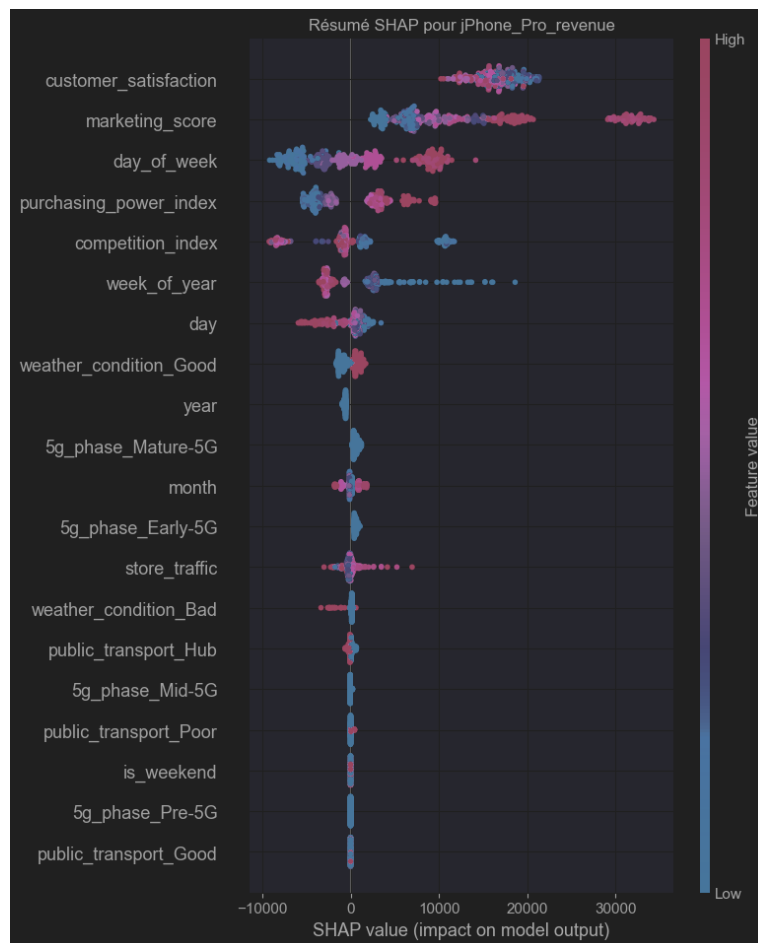
L'analyse SHAP a permis de comprendre les contributions individuelles des caractéristiques aux prédictions des revenus. Voici les résultats détaillés par modèle.

1. iPhone Pro

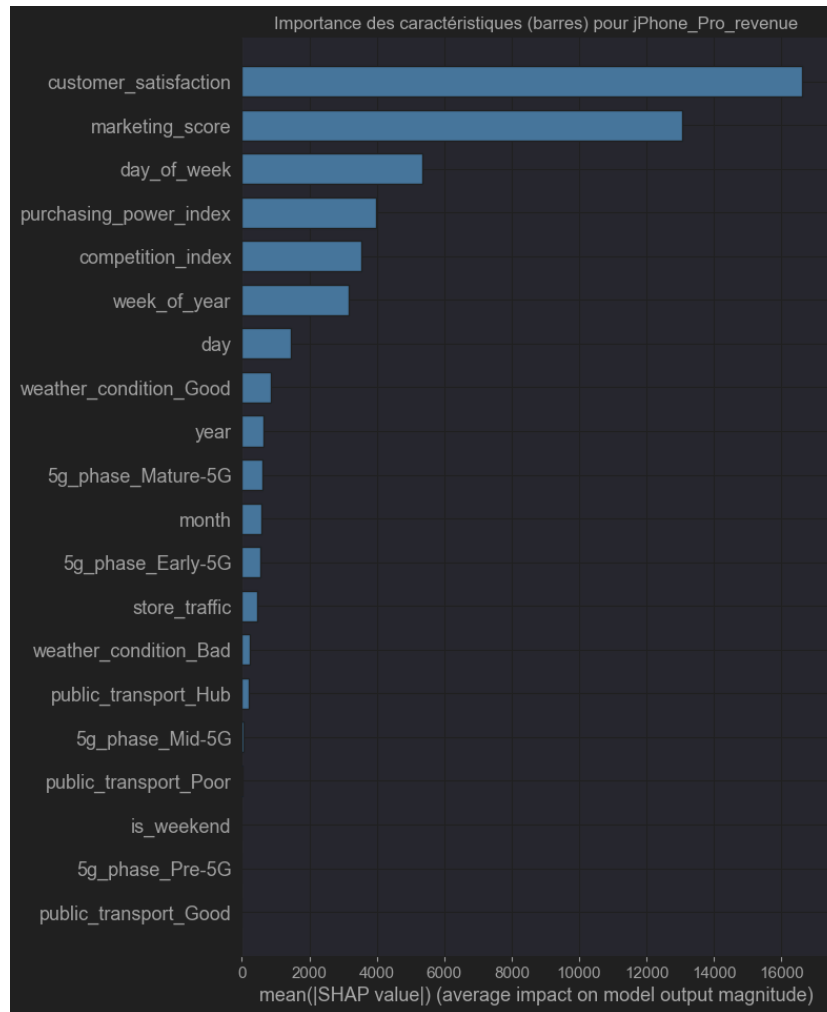
- Les caractéristiques les plus influentes :

Les variables les plus influentes sur les revenus de l'iPhone Pro incluent la **satisfaction client** et le **score marketing**, qui montrent une forte corrélation positive avec les revenus. Ces facteurs soulignent l'importance de maintenir des niveaux élevés de satisfaction client et de renforcer les campagnes marketing pour maximiser les performances. Les caractéristiques temporelles comme le **jour de la semaine** jouent également un rôle clé, avec des variations importantes selon les périodes.

- Visualisation SHAP :
 - Résumé des impacts des caractéristiques :



- Importance moyenne des caractéristiques :



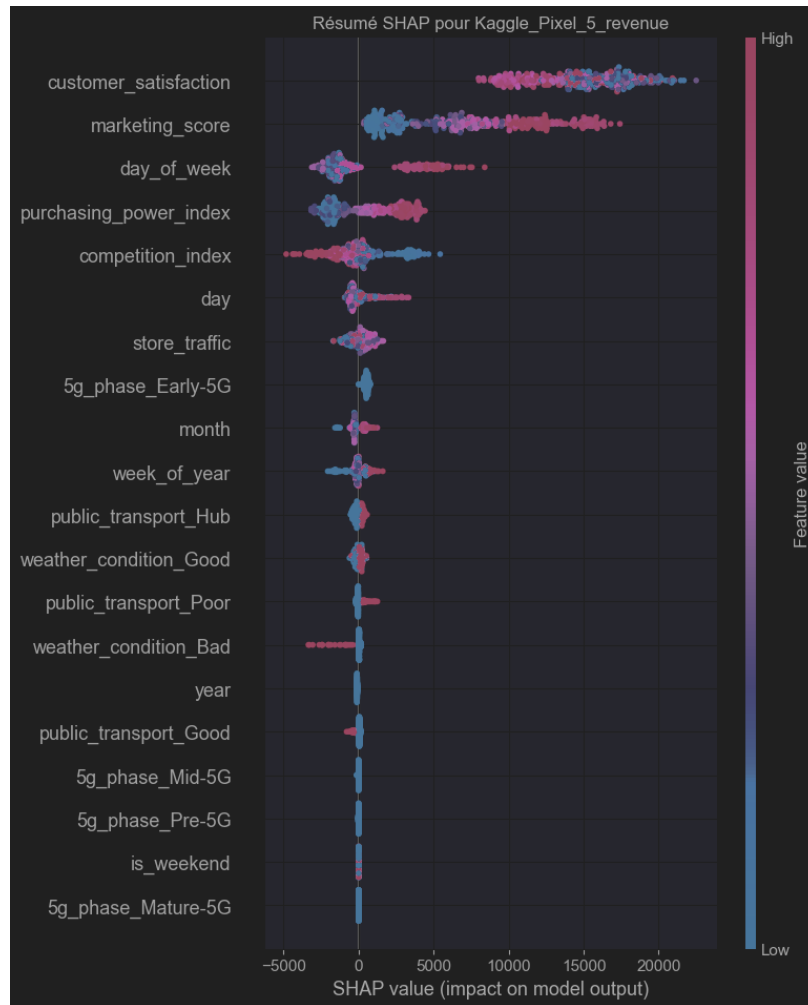
2. Kaggle Pixel 5

- Les caractéristiques les plus influentes :

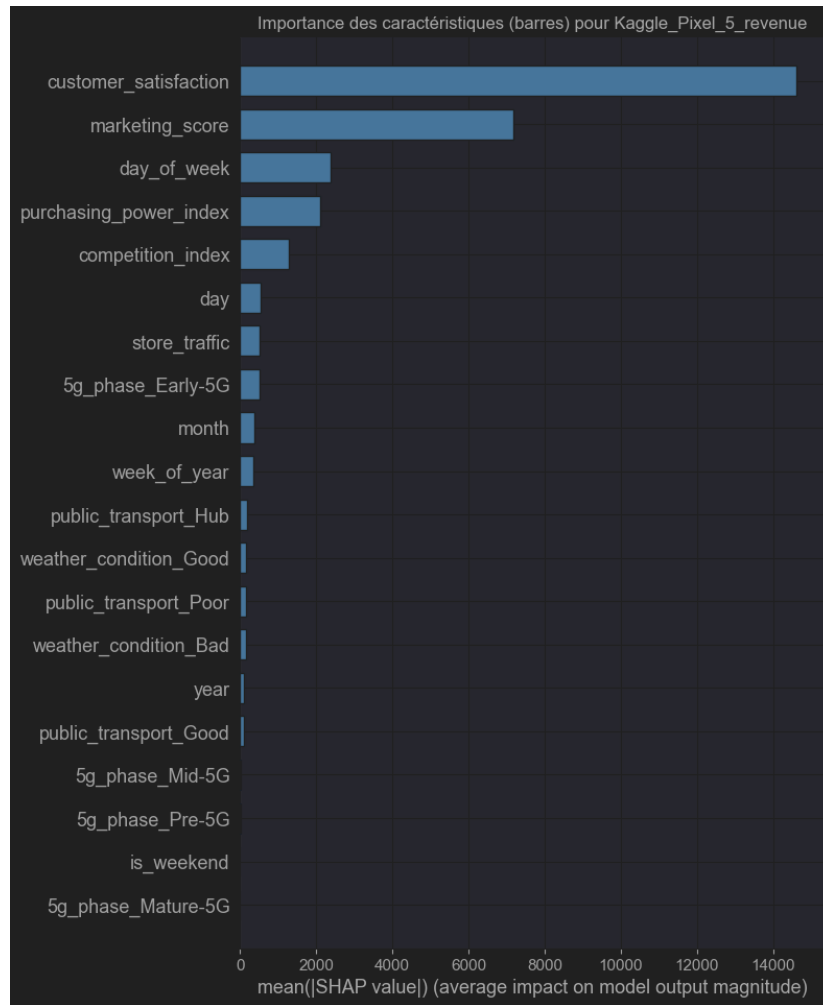
Pour le Kaggle Pixel 5, la **satisfaction client** domine en tant que facteur clé, suivi par le **score marketing**. Ces deux caractéristiques sont essentielles pour optimiser les ventes de ce modèle. Les variations selon le **jour de la semaine** et les **indices économiques**, comme le pouvoir d'achat et la concurrence, influencent également significativement les performances de ce modèle.

- Visualisation SHAP :

- Résumé des impacts des caractéristiques :



- Importance moyenne des caractéristiques :



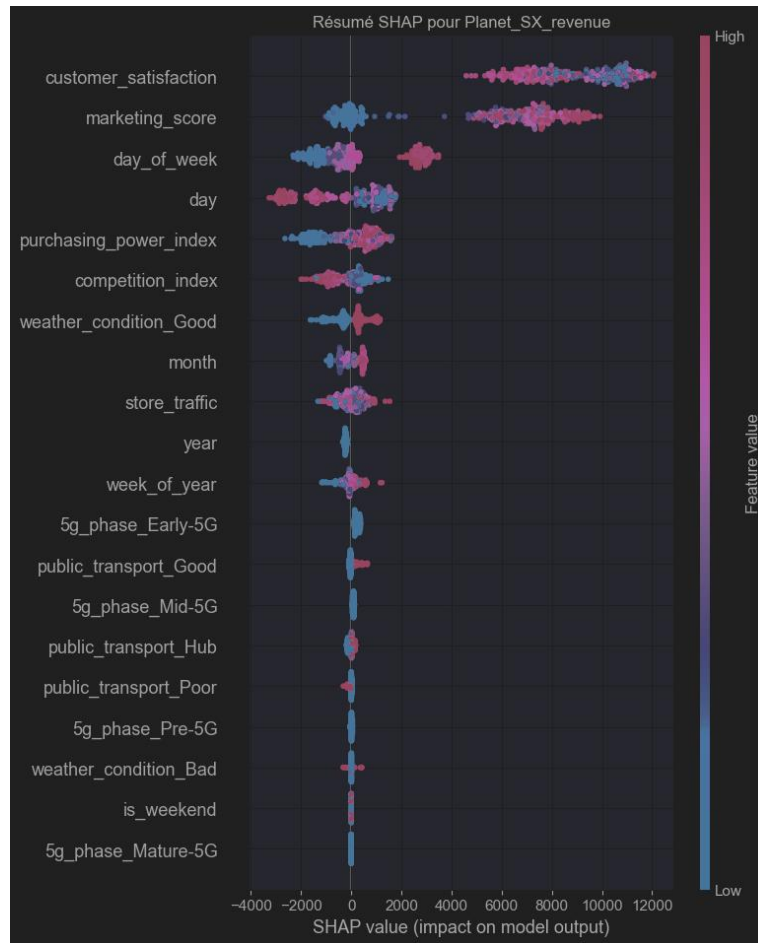
3. Planet SX

- Les caractéristiques les plus influentes :

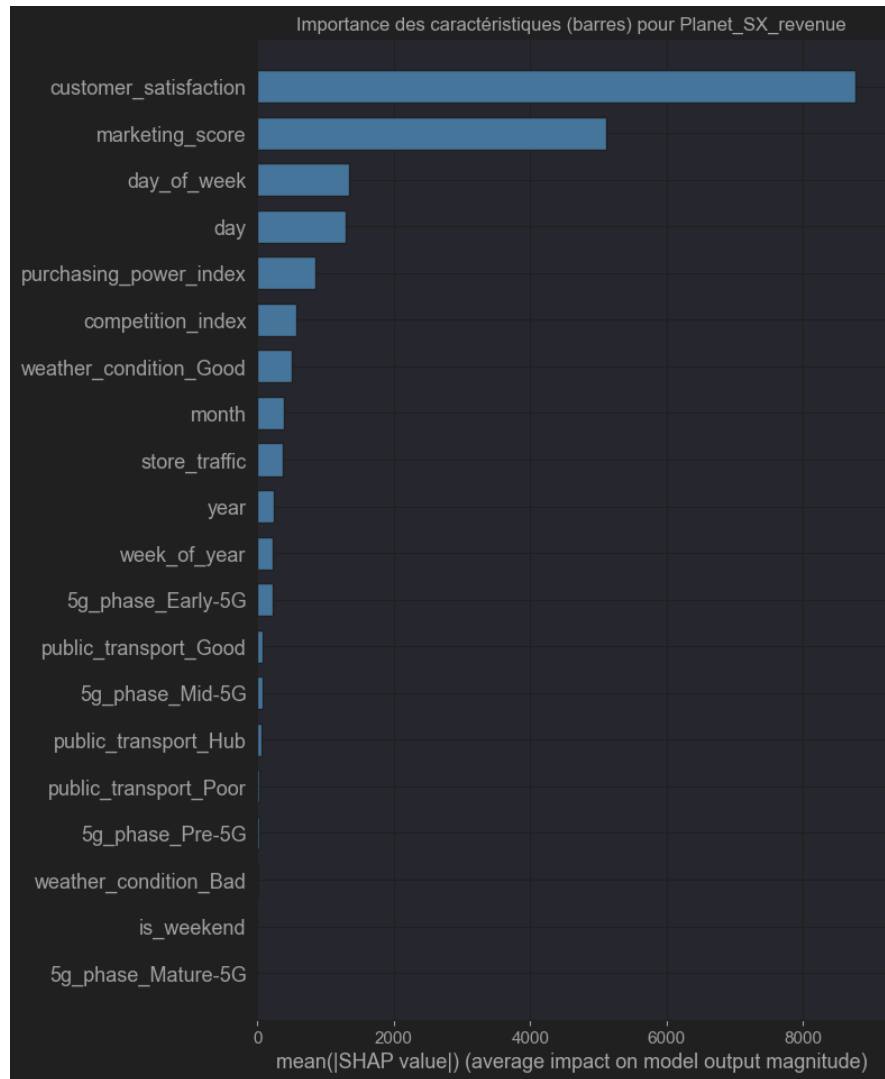
Le modèle Planet SX est fortement influencé par la **satisfaction client**, avec un impact significatif du **score marketing**. Les facteurs temporels comme le **jour** et le **mois** jouent également un rôle non négligeable, indiquant des variations liées à des périodes spécifiques de l'année. De plus, l'adoption des phases 5G (notamment la phase **Early-5G**) semble avoir un effet mesuré sur ses revenus.

- Visualisation SHAP :

- Résumé des impacts des caractéristiques :



- Importance moyenne des caractéristiques :



Dans l'ensemble, ces analyses montrent que, bien que certains facteurs comme la satisfaction client et le marketing soient constants dans leur impact, chaque modèle présente des sensibilités spécifiques à d'autres variables telles que le calendrier ou les indices économiques.

Recommandations opérationnelles

1. Amélioration de la Satisfaction Client

- La **satisfaction client** est le facteur le plus influent sur les revenus pour tous les modèles. Il est crucial d'investir dans des initiatives qui augmentent la satisfaction des clients, telles que :
 - Renforcer la qualité du service après-vente.
 - Offrir des programmes de fidélité pour encourager les achats récurrents.
 - Simplifier le processus d'achat en ligne et en magasin.

2. Optimisation des Stratégies Marketing

- Le **score marketing** a également un impact significatif sur les revenus. Les entreprises devraient :
 - Augmenter les investissements publicitaires pendant les jours où les ventes sont historiquement faibles.
 - Suivre les performances des campagnes en temps réel pour ajuster les messages et maximiser leur efficacité.

3. Gestion des Variations Temporelles

- Les variations dans les revenus selon les **jours de la semaine**, les **mois** et les **phases 5G** nécessitent des stratégies flexibles :
 - Planifier des promotions et des lancements de produits pendant les périodes à fort potentiel de ventes identifiées.
 - Ajuster les ressources (par exemple, le personnel et les stocks) en fonction des prévisions de demande pour chaque période.

4. Exploitation des Phases 5G

- Les données montrent que la phase **5g_Mature-5G** a un impact positif sur les revenus. Les recommandations incluent :
 - Concentrer les efforts de vente et de marketing dans les régions où cette phase est dominante.
 - Collaborer avec les opérateurs téléphoniques pour accélérer l'adoption de la 5G dans les zones encore en phase précoce.

5. Adaptation aux Conditions Météorologiques

- Les **conditions météorologiques** ont un impact léger mais notable, surtout pour certains modèles. Des actions spécifiques pourraient inclure :
 - Proposer des offres spéciales pendant les périodes de mauvais temps pour stimuler les ventes en ligne.
 - Adapter les campagnes marketing en fonction des prévisions météorologiques.

6. Focus sur les Caractéristiques Secondaires

- Bien que moins influents, certains facteurs comme le **trafic en magasin** et les **conditions de transport public** peuvent être exploités pour améliorer les performances :

- Améliorer l'expérience en magasin pour attirer plus de trafic.
- Collaborer avec les services de transport pour promouvoir l'accès facile aux magasins.

Conclusion

Dans le cadre de la mission confiée par DataTel, l'objectif principal était d'optimiser la stratégie de vente de smartphones dans les boutiques de l'opérateur en exploitant les données historiques et en développant un modèle de prévision robuste pour le premier trimestre 2025. Cependant, en raison de contraintes de temps, cette analyse a été réalisée exclusivement sur les données de Marseille. À travers cette étude ciblée, plusieurs étapes clés ont été menées, aboutissant à des résultats concrets et des recommandations stratégiques spécifiques à cette ville.