

# Rapport Test Stage Orange

Patricio José SILVA

Janvier 2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Exploration Statistique des Variables . . . . .	2
<b>2</b>	<b>Effets Locaux et <i>Insights Business</i></b>	<b>5</b>
2.1	Paris . . . . .	5
2.2	Lyon . . . . .	6
2.3	Marseille . . . . .	6
2.4	Lille . . . . .	7
2.5	Bordeaux . . . . .	8
2.6	Strasbourg . . . . .	8
2.7	Rennes . . . . .	9
2.8	Fort de France . . . . .	9
<b>3</b>	<b>Prévisions et Interprétation</b>	<b>11</b>
3.1	Méthodologie de Modélisation . . . . .	11
3.2	Prévisions pour Paris . . . . .	11
3.3	Prévisions pour Lyon . . . . .	12
3.4	Recommandations Opérationnelles . . . . .	13

# Chapitre 1

## Introduction

Le présent rapport est l'accompagnement demandé au code du test technical pour le stage de data scientist chez orange. Ce rapport inclus :

- Les principaux insights business
- La méthodologie de modélisation
- Les prévisions et leur interprétation
- Des recommandations opérationnelles

ainsi que les explications correspondantes.

### 1.1 Exploration Statistique des Variables

Les graphiques des variables dans la data peut nous montrer comment les revenus sont changés selon chaque variable. Dans la figure 1.1 les relations des revenus par rapport aux différentes variables peuvent être observées. Pour constater et mesurer la signification de l'effet des variables sur les revenus, régressions linéaires sont construites et les *p-values* sont observées pour mesurer l'importance de chaque variable sur les revenus. Pour montrer cela, comme exemple, la ville de Paris et la moyenne des revenus des modèles sont choisis et, les variables catégoriques, comme ils sont hiérarchiques, sont données valeurs numériques. Les graphiques montrés démontrent que chaque variable a des effets sur les revenus, où les variables continues montrent une corrélation dans les *scatterplots* et les revenus, par rapport au temps, ont la structure d'une série temporelle. Les variables discrètes ont d'un effet évident en observant les graphiques.

Pour chaque variable, une régression multiple linéaire a été faite entre les variables et le revenu de chaque modèle, d'où les paramètres et les significances (*p-values*) sont obtenus. Le tableau 1.1 montre les valeurs moyennes de coefficients (pentes) des différentes variables sur les revenus et le *p-value* maximale de chaque variable sur les modèles. De cette forme, l'effet moyen et la minimale signification de chaque variable est montrée et, de cette information, insight business peuvent être obtenus pour chaque ville, ainsi qu'une compréhension des effets des variables de manière locale. Un *p-value* moins que 5% est géné-

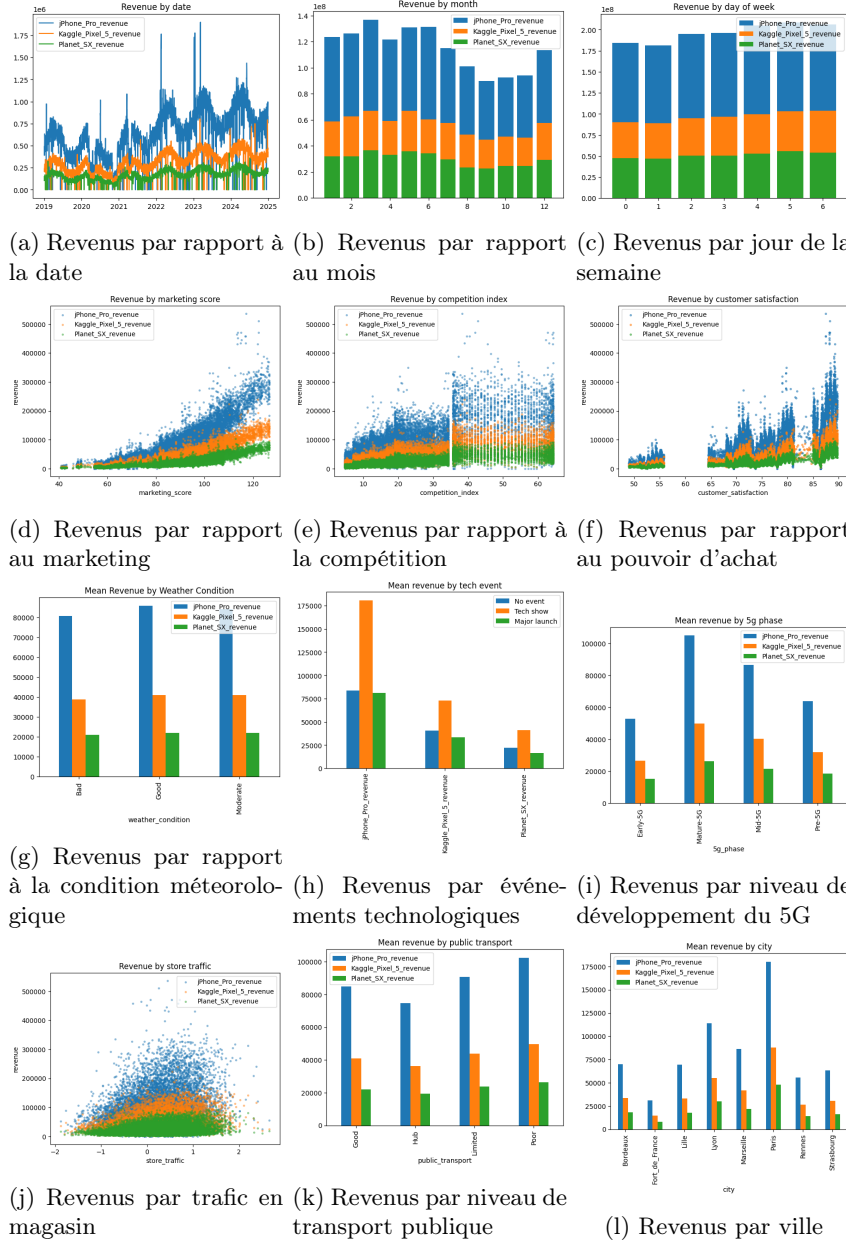


FIGURE 1.1 – Graphiques des revenus par rapport aux variables

	coeffs	pvalues
marketing score	1782.350672	3.283681e-83
purchasing power index	2251.918478	1.348907e-17
customer satisfaction	4185.630891	4.003269e-11
5g phase	6133.232151	6.894282e-10
weather condition	2863.831794	5.138940e-05
public transport	-1476.516145	3.535596e-03
store traffic	2304.670912	1.868250e-01
competition index	-115.177274	2.950163e-01
tech event	7849.596275	5.912788e-01

TABLE 1.1 – Coefficients et p-values des variables pour Paris

	coeffs	pvalues
marketing_score	843.106575	1.154761e-50
purchasing_power_index	1265.247867	1.348907e-17
customer_satisfaction	2286.793838	4.003269e-11
weather_condition	1498.407740	1.429182e-01
5g_phase	3056.744781	4.402143e-01
competition_index	-18.702265	8.789529e-01
tech_event	2601.821121	9.037570e-01
public_transport	-267.233903	9.538125e-01
store_traffic	817.334744	9.745636e-01

TABLE 1.2 – Coefficients et p-values pour la France

ralement considéré comme statistiquement significatif, où une valeur inférieure représente une signification plus grande. Une valeur plus grande que 5% peut signifier que l'effet de la variable peut être causé par du bruit. Le coefficient peut être lu comme une corrélation, c'est-à-dire, une valeur positive représente une relation positive entre la croissance de la variable et le revenu. Donc, ce tableau nous donne l'information sur l'importance de chaque variable (*p-value*) et l'effet moyen sur le revenu des téléphones. Est important que la valeur exacte des coefficients n'est pas prise en compte, car la magnitude est produite principalement par l'ordre de grandeur des données. Pour étudier l'effet des variables sur toutes les villes, le tableau 1.2 est fait où les coefficients sont les moyens sur les villes et les *p-values* sont le maximum sur les villes. De ce tableau, on peut observer que, globalement, seulement *marketing score*, *purchasing power index* et *customer satisfaction* sont statistiquement significatifs et ont d'un effet positif sur les revenus. Comme la comparaison avec le tableau 1.1 peut montrer, les effets varient entre les villes, d'où *insights business* peuvent être obtenus.

Pour la prochaine partie, chaque ville sera étudiée de manière individuelle et les effets locaux des variables seront étudiés pour obtenir les *insights* pour la ville.

## Chapitre 2

# Effets Locaux et *Insights Business*

Pour chaque ville, les coefficients statistiquement significatifs sont considérés pour faire des *insights business* (en ordre de signifiante) et les valeurs de ces coefficients pour déterminer si les variables doivent être augmentées ou diminuées.

### 2.1 Paris

Le tableau 2.1 montre les effets des variables sur les revenus à Paris, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	1782.350672	3.283681e-83
purchasing_power_index	2251.918478	1.348907e-17
customer_satisfaction	4185.630891	4.003269e-11
5g_phase	6133.232151	6.894282e-10
weather_condition	2863.831794	5.138940e-05
public_transport	-1476.516145	3.535596e-03
store_traffic	2304.670912	1.868250e-01
competition_index	-115.177274	2.950163e-01
tech_event	7849.596275	5.912788e-01

TABLE 2.1 – Effets des variables sur les revenus de Paris

- Augmenter la commercialisation
- Rechercher ou investir dans des domaines à plus haut pouvoir d'achat
- Accroître la satisfaction des clients
- Rechercher ou investir dans des domaines où le développement de la 5G est plus élevé

- Rechercher ou investir dans des régions où les conditions météorologiques sont meilleures
- Éviter les zones avec beaucoup de transports en commun

## 2.2 Lyon

Le tableau 2.2 montre les effets des variables sur les revenus à Lyon, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	1115.717126	3.658020e-74
purchasing_power_index	1421.302344	1.151616e-19
customer_satisfaction	3000.784067	8.143372e-14
5g_phase	4029.599637	1.400652e-09
weather_condition	2397.388395	7.801454e-09
store_traffic	1281.232131	2.551215e-01
tech_event	5879.735427	3.753990e-01
public_transport	-306.694764	5.810217e-01
competition_index	-84.891194	8.682022e-01

TABLE 2.2 – Coefficients et p-values pour Lyon

- Augmenter le marketing
- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter la satisfaction client
- Investir dans les zones avec un réseau 5G plus développé
- Investir dans les zones avec de meilleures conditions météorologiques

## 2.3 Marseille

Le tableau 2.3 montre les effets des variables sur les revenus à Marseille, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	984.135257	1.160484e-71
purchasing_power_index	1281.598036	1.080138e-25
customer_satisfaction	2577.195495	1.783411e-20
5g_phase	3769.340078	2.722837e-12
store_traffic	1317.937838	1.563454e-02
weather_condition	850.642759	1.429182e-01
tech_event	2466.620132	8.003446e-01
competition_index	-37.218573	8.756711e-01
public_transport	-56.105086	9.018851e-01

TABLE 2.3 – Coefficients et p-values pour Marseille

- Augmenter le marketing
- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter la satisfaction client
- Investir dans les zones avec un réseau 5G plus développé
- Investir dans les zones avec un trafic en magasin plus élevé

## 2.4 Lille

Le tableau 2.4 montre les effets des variables sur les revenus à Lille, d'où les suivants *insights business* peuvent être obtenus.

Variable	Coeffs	P-values
marketing_score	671.099290	9.835377e-55
purchasing_power_index	1172.236116	1.005141e-32
customer_satisfaction	2185.755696	2.885565e-16
weather_condition	1426.156218	3.857575e-10
5g_phase	2679.678974	6.547810e-10
store_traffic	485.686509	2.113656e-01
tech_event	446.260815	6.125121e-01
competition_index	-3.220110	7.814894e-01
public_transport	99.327255	8.510277e-01

TABLE 2.4 – Coefficients et p-values pour Lille.

- Augmenter le marketing
- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter la satisfaction client
- Investir dans les zones avec de meilleures conditions météorologiques
- Investir dans les zones avec un développement plus avancé de la 5G



## 2.5 Bordeaux

Le tableau 2.5 montre les effets des variables sur les revenus à Bordeaux, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	727.401715	3.792782e-59
purchasing_power_index	1242.188896	5.179089e-38
customer_satisfaction	1920.807608	1.144918e-13
5g_phase	2644.897535	3.196548e-08
weather_condition	945.979534	6.509937e-03
store_traffic	893.948687	3.118325e-02
public_transport	-245.497027	5.473816e-01
tech_event	-1153.296746	7.310576e-01
competition_index	38.922322	8.789529e-01

TABLE 2.5 – Coefficients et p-values pour Bordeaux

- Augmenter le marketing
- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter la satisfaction client
- Investir dans les zones avec un meilleur développement de la 5G
- Investir dans les zones avec de meilleures conditions météorologiques
- Investir dans les zones avec un trafic en magasin plus élevé

## 2.6 Strasbourg

Le tableau 2.6 montre les effets des variables sur les revenus à Strasbourg, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	605.859557	2.745533e-64
purchasing_power_index	931.972487	1.988208e-29
weather_condition	1699.231441	4.898938e-14
customer_satisfaction	1676.213609	1.397459e-12
5g_phase	2330.877314	8.234879e-08
competition_index	26.152241	7.790042e-01
tech_event	1247.779938	9.037570e-01
public_transport	-9.914161	9.538125e-01
store_traffic	143.778297	9.745636e-01

TABLE 2.6 – Coefficients et p-values pour Strasbourg

## 2.7 Rennes

Le tableau 2.7 montre les effets des variables sur les revenus à Rennes, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
marketing_score	468.548890	7.852141e-51
purchasing_power_index	866.816125	7.323172e-38
customer_satisfaction	1380.345261	2.890575e-11
5g_phase	2275.978858	1.178155e-07
weather_condition	995.471078	4.131633e-07
competition_index	69.979270	4.707636e-01
store_traffic	200.599683	5.692335e-01
public_transport	-87.175839	6.903256e-01
tech_event	2824.939355	8.340452e-01

TABLE 2.7 – Coefficients et p-values pour Rennes

- Augmenter le marketing
- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter la satisfaction client
- Investir dans les zones avec un développement plus avancé de la 5G
- Investir dans les zones avec de meilleures conditions météorologiques

## 2.8 Fort de France

Le tableau 2.8 montre les effets des variables sur les revenus à Fort de France, d'où les suivants *insights business* peuvent être obtenus.

	coeffs	pvalues
purchasing_power_index	953.950453	3.409854e-101
marketing_score	389.740091	1.154761e-50
customer_satisfaction	1367.618081	1.738573e-23
weather_condition	808.560698	3.133945e-06
5g_phase	590.353701	4.402143e-01
competition_index	-44.164797	5.288692e-01
store_traffic	-89.176108	6.622493e-01
public_transport	-55.295457	7.771807e-01
tech_event	1252.933770	8.185720e-01

TABLE 2.8 – Coefficients et p-values pour Fort de France

- Investir dans les zones à plus fort pouvoir d'achat
- Augmenter le marketing
- Augmenter la satisfaction client

— Investir dans les zones avec de meilleures conditions météorologiques

## Chapitre 3

# Prévisions et Interprétation

### 3.1 Méthodologie de Modélisation

Les revenus, après d'une observation visuelle, ont été déterminés comme saisonnières, c'est-à-dire, qu'il existe une oscillation sur le temps qui, sommée à une croissance, a la structure d'une série temporelle. Pour cette raison, les prévisions ont été produites par cette modèle. Les variables utilisées pour prédire les revenus sont, pour la plupart, aussi saisonnières, alors elles ont été prédites avec séries temporelles aussi. Une validation avec un 5,6% des données d'entraînement a montré que le modèle de séries temporelles fonctionne de manière assez bonne pour ce problème, alors comme les séries temporelles sont le modèle d'élection pour les prédictions temporelles, j'ai choisi cette modèle. Il faut tenir en compte que les séries temporelles seulement prédit le comportement d'un phénomène stationnaire, donc pour transformer les données à données stationnaires, pour observer la variation et la prédire, une régression linéique par rapport au temps a été faite et après soustraite des observations. Après la prédiction avec les séries temporelles, la même régression linéique a été utilisée (sommée) pour de-normaliser les données. Pour montrer et interpréter les prévisions, les villes de Paris et Lyon sont choisies.

### 3.2 Prévisions pour Paris

La graphique de la prévision peut être observée dans la figure 3.1.

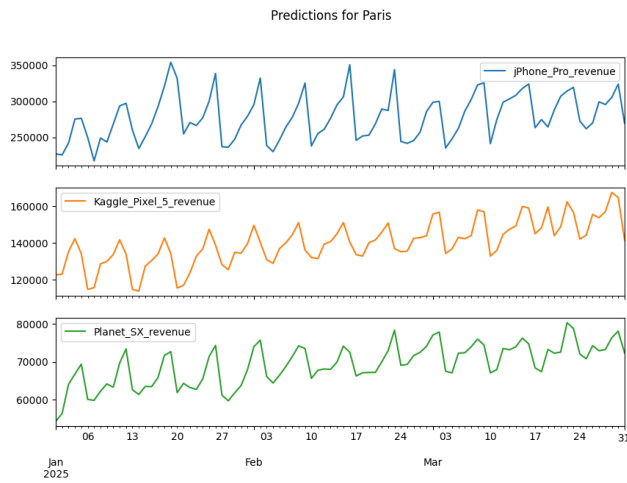


FIGURE 3.1 – Prévisions pour Paris

où les premières interprétations peuvent être la relative croissance et l'oscillation des revenus par modèle. Ça signifie que le modèle de prédiction espère que le marché des téléphones continue à grandir et que les revenus varient dans chaque mois et dans les jours de la semaine. Aussi, la croissance des modèles Kaggle Pixel et Planet SX semblent à grandir proportionnellement plus que le modèle leader, le iPhone Pro.

### 3.3 Prévisions pour Lyon

La graphique 3.2 montre la prévision pour Lyon.

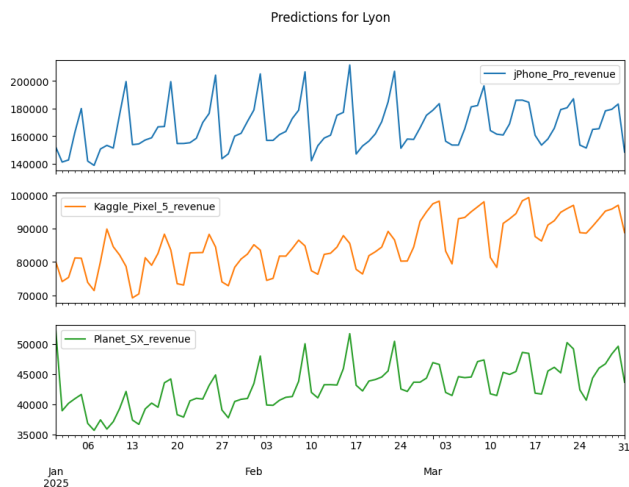


FIGURE 3.2 – Pr vision pour Lyon

Ici, le mod le leader, iPhone Pro ne montre pas une croissance marqu e, plut t, une stationnarit  pour les mois    tudier. Le mod le Planet SX a un comportement similaire. Le mod le Kaggle Pixel, en revanche, montre un bon grandissement. Peut- tre que le mod le n'esp re pas une croissance marqu e pour le march  de cette ville.

### 3.4 Recommandations Op rationnelles

En plus des recommandations op rationnelles donn es dans le premier chapitre de ce rapport, on peut recommander d'ajuster le *stock* et le budget selon les pr visions donn es dans la section pr c dente. En plus, les revenus peuvent  tre grandis gr ce   une inversion selon les *insight business* donn s dans l' tude de variables locaux.