



RAPPORT

Conseils d'investissements

BPI-France cherche à distinguer les performances de ses entreprises et à remarquer les principaux générateurs de profit parmi des critères tels que leur géolocalisation et les dépenses.

Nous avons donc conçu un modèle prédictif en prenant avantage de la distribution normale des données. Ce modèle se caractérise par régression linéaire multiple prenant comme paramètres : les dépenses en R&D, Marketing, Administration, ainsi que la ville où opère l'entreprise.

5 start-ups digne d'attention

	Profit	prediction	RATIO
45	64926.08	49268.116335	131.781129
36	90708.19	72827.009208	124.552952
38	81229.06	65644.277738	123.741265
43	69758.98	61621.330858	113.205896
16	126992.93	114854.040827	110.568970

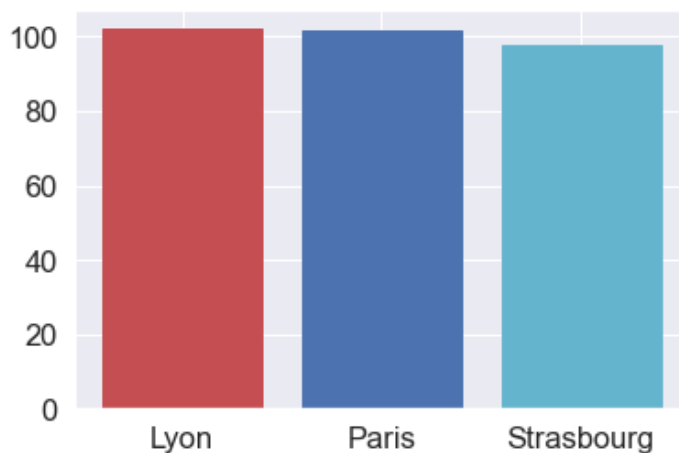
Après comparaison entre les prédictions de profit et le profit réel généré par ses entreprises, nous avons découvert 5 entreprises ayant des gains supérieurs au modèle de prédiction. Ces entreprises sont, par ID :

- n°45 avec 131,78% de la prédiction
- n°36 avec 124,55% de la prédiction
- n°38 avec 123,74% de la prédiction
- n°43 avec 113,20% de la prédiction
- n°16 avec 110,56% de la prédiction

Classements des villes les plus performantes

Si les entreprises basées à Lyon sont celles qui semblent, en moyenne, être les plus performantes, la différence est sensiblement minime.

	RATIO
villes	
Lyon	102.098000
Paris	101.573120
Strasbourg	97.911019



Tom MARCHAL

Antoine MEYER

Conseils d'investissements

Notre modèle de prédiction

$$y = 41011.48935223 + 0.828371991 X_R + 0.0503972419 X_M + 0.0221861926 X_A + 0.0062648 X_V$$

Où y représente le profit

X_R les dépenses en R&D

X_A les dépenses en Administratif

X_M les dépenses en Marketing

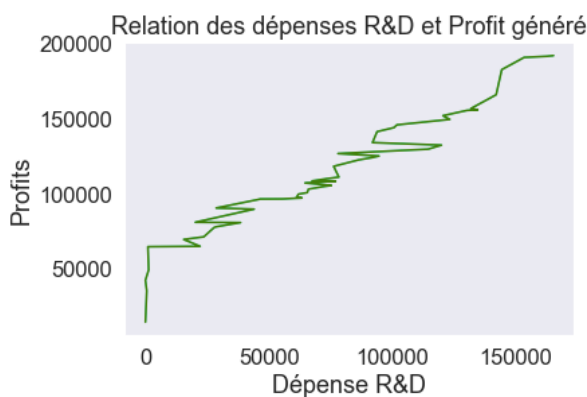
X_V la géolocalisation (encodé)

Ce modèle de régression linéaire multiple nous a été obtenu par Machine Learning avec un $R^2 = 0.9576$. Nous observons que le plus haut coefficient est celui de la R&D.

La Backward Elimination nous confirmant que la R&D est le seul paramètre véritablement significatif (voir notebook), nous avons pu simplifier notre modèle, à $R^2 = 0.944$:

$$y = 48416.29766139 + 0.8516228 X_R$$

R&D, meilleur générateur de profit



Il semble donc judicieux de focaliser les investissements strictement dans la R&D plus que tout autre paramètre, ce dernier démontrant statistiquement son influence sur les profits générés.

Conseils d'investissements

Conclusion

Pour nous convaincre de la pertinence de notre modèle, nous avons comparé nos résultats avec un deuxième outil d'analyse. Nous remarquons que les coefficients générés sont sensiblement les mêmes, que cela soit pour la régression linéaire multiple ou la régression linéaire simple.

Modèle 2: MCO, utilisant les observations 1-50
Variable dépendante: Profit

	coefficient	éc. type	t de Student	p. critique
const	50122,2	6572,35	7,626	1,06e-09
RD	0,805715	0,0451473	17,85	2,63e-22
Administration	-0,0268160	0,0510288	-0,5255	0,6018
Marketing	0,0272281	0,0164512	1,655	0,1047
Moyenne var. dép.	112012,6	Éc. type var. dép.	40306,18	
Somme carrés résidus	3,92e+09	Éc. type régression	9232,335	
R2	0,950746	R2 ajusté	0,947534	
F(3, 46)	295,9781	P. critique (F)	4,53e-30	
Log de vraisemblance	-525,3857	Critère d'Akaike	1058,771	
Critère de Schwarz	1066,420	Hannan-Quinn	1061,684	

Modèle 3: MCO, utilisant les observations 1-50
Variable dépendante: Profit

	coefficient	éc. type	t de Student	p. critiqu
const	49032,9	2537,90	19,32	2,78e-24
RD	0,854291	0,0293056	29,15	3,50e-32
Moyenne var. dép.	112012,6	Éc. type var. dép.	40306,18	
Somme carrés résidus	4,26e+09	Éc. type régression	9416,349	
R2	0,946535	R2 ajusté	0,945421	
F(1, 48)	849,7889	P. critique (F)	3,50e-32	
Log de vraisemblance	-527,4365	Critère d'Akaike	1058,873	
Critère de Schwarz	1062,697	Hannan-Quinn	1060,329	

Ceci confirme la pertinence de notre modèle d'analyse.

Nous vous conseillons donc d'investir :

Dans les entreprises performant mieux que le modèle.

Dans les entreprises ayant un grand potentiel de R&D.

Merci d'avoir choisi My Company.

Tom MARCHAL

Antoine MEYER