

Projet traitement de l'information

Étude des campagnes Kickstarter et de leur réussite

Rodriguez Esteban, Coutan Thomas, Benoît Chénard

Table des matières

1	Introduction	2
2	Présentation des données	2
3	Présentation de la méthode	2
4	Analyse des résultats	3
4.1	Montant donné par donateur en fonction de la catégorie	3
4.2	Montant donné par donateur en fonction de la devise	4
4.3	Réussite des campagnes	5
5	Conclusion	5

1 Introduction

Dans ce rapport, nous étudierons les projets Kickstarters et essaierons de trouver des facteurs de réussite. Nous analyserons mais aussi quantitatives. Pour cela, nous utiliserons différentes méthodes pour traiter les données, notamment des régressions linéaires pour déterminer quelles catégories sont les plus prolifiques ou encore quel monnaie (et donc quel pays) est la plus donnée par donateurs.

Nous expliquerons plus précisément dans la section 2 la base de données que nous allons utiliser, dans la section 3, nous détaillerons la méthode avec laquelle nous avons traité ces données. Ensuite, dans la section 4, nous analyserons les différents résultats obtenus. Enfin, nous concluerons en section 5.

2 Présentation des données

Les données sur lesquelles nous allons travailler sont des données nous donnant diverses informations sur des projets Kickstarter. Kickstarter est une plateforme de financement participatif qui permet à de nombreux indépendants de récolter des fonds pour développer leur projet. Ainsi, des projets comme le jeu Star Citizen ou la montre Pebble Time ont pu voir le jour grâce au soutien des donateurs. Les données comportent des informations sur plus de 240 000 campagnes qui se sont déroulées entre 2010 et 2016, elles nous donnent comme information la catégorie, l'objectif initial, le nombre de donateurs et le montant levé au total. Nous avons aussi les dates de début et de fin de la campagne. Aussi, nous avons des informations sur les devises(monnaie) et pays d'origine de ces campagnes. Ainsi, ces données sont pour la plupart quantitatives et vont être relativement facile à analyser.

3 Présentation de la méthode

Nous avons décidé d'étudier plusieurs points: les catégories qui rapportent le plus d'argent par donateur, le montant donné par donateur en fonction de la devise et enfin de voir si les campagnes remplissent généralement leur objectif initial.

Pour les deux premiers points, nous utiliserons des régressions linéaires. Pour effectuer ces régressions linéaires, nous aurons besoin de calculer la moyenne des données selon x et y. On considère N valeurs d'un caractère x_1, x_2, \dots, x_N . La **moyenne**, notée \bar{x} , est:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

Et pour les y:

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_N}{N} = \frac{1}{N} \sum_{i=1}^N y_i$$

Il nous faut aussi la variance des x, et la covariance de x et y: La **variance d'une série** est la moyenne des carrés des écarts à la moyenne:

$$V(x) = \frac{n_1 x_1 - \bar{x}^2 + n_2 x_2 - \bar{x}^2 + \dots + n_N x_N - \bar{x}^2}{N} = \frac{1}{N} \sum_{i=1}^N n_i x_i - \bar{x}^2$$
 La **covariance** de x et y est

donnée par la formule :

$$Cov(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Le but d'une régression linéaire est de former une droite en fonction d'un nuage de points. Nous allons donc obtenir une droite de formule : $y=ax+b$ avec:

$$a = \frac{Cov(x, y)}{V(x)} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

Pour le troisième, nous représenteront les campagnes dans un espace en 3 dimensions avec comme axes: le montant récolté, l'objectif et le nombre de donateurs.

4 Analyse des résultats

4.1 Montant donné par donateur en fonction de la catégorie

Nous pouvons observer sur la figure 1 que certaines catégories attirent de plus gros donateurs. En effet, la catégorie "Arts" possède des donateurs bien plus généreux en moyenne. Des catégories comme "Musique", "Films" et autres ont aussi d'importantes moyennes de dons. En revanche, certaines catégories peinent à avoir d'importantes contributions en moyenne, notamment pour la catégorie "Comics".

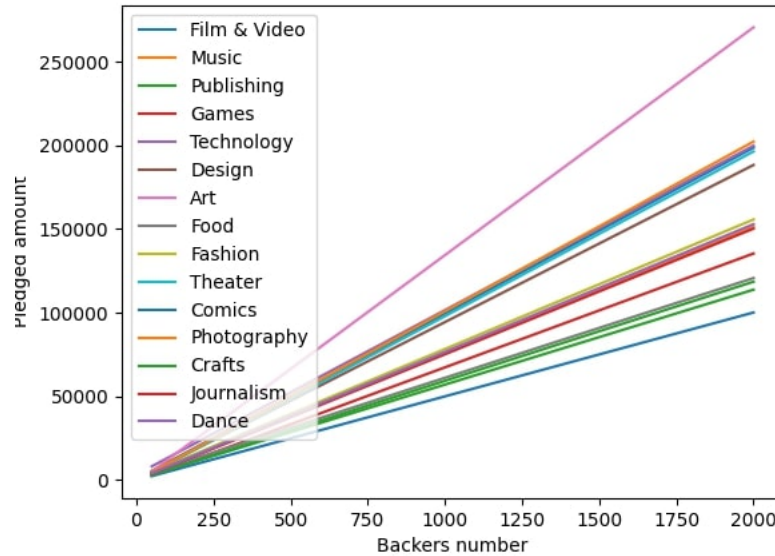


Figure 1: Montant donné en fonction du nombre de donateurs et de la devise

4.2 Montant donné par donateur en fonction de la devise

Nous pouvons observer sur la figure 2 un graphique représentant le montant total récolté (en équivalent US dollar) en fonction de la devise et du nombre de donateurs. Ainsi notre régression linéaire nous permet de déterminer la moyenne du montant des contributions par devise. En effet, cette moyenne est égale à la pente des droites. Ainsi, on voit se distinguer 3 devises dont le montant moyen des contributions est plus élevé: le Dollar de Honk Kong, la couronne danoise et la couronne norvégienne. Au contraire, la livre sterling est la monnaie dont les contributions sont en moyenne les plus faibles.

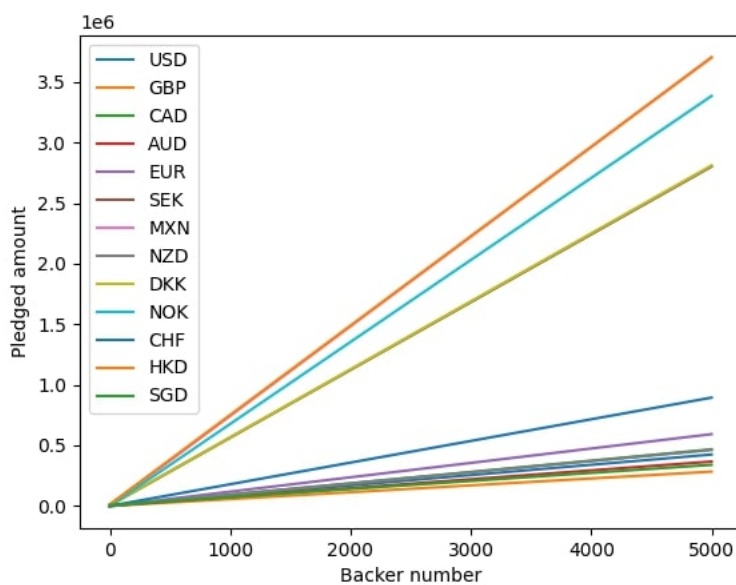


Figure 2: Montant donné en fonction du nombre de donateurs et de la catégorie

4.3 Réussite des campagnes

Nous pouvons observer sur la figure 3 qu'il y a énormément de campagnes qui se situent dans le coin en bas à gauche, c'est-à-dire que la plupart récolte pas énormément d'argent et mobilisent peu de donateurs. Cependant, une partie non négligeable des campagnes parviennent à dépasser leur objectif, il y a aussi des "anomalies", notamment une campagne qui n'a pas beaucoup de donateurs et un objectif initial peu élevé mais qui a tout de même récolté énormément d'argent.

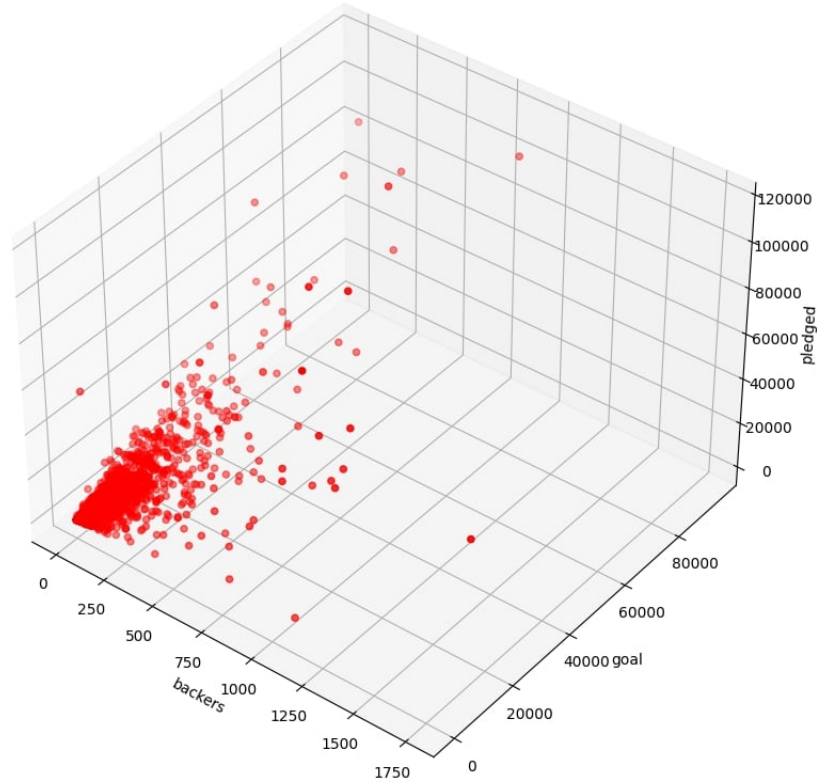


Figure 3: Représentation dans l'espace des campagnes Kickstarter

5 Conclusion

En somme, nous avons réussi déterminer les catégories dans lesquelles les donateurs donnent le plus mais aussi quel devise sont les plus généreusement donnée (par utilisateur). Ces deux points peuvent nous permettre, grâce aux régressions linéaires de prévoir le montant moyen que les donateurs vont donner en fonction de la catégorie et de la devise de la campagne. Nous pensons qu'une étude sur la durée des campagnes pourrait aussi être utile afin de trouver tous les paramètres les plus optimisés pour réussir sa campagne.