

Synthèse de Mini-Projet de Data-science

Introduction

En tant que joueur de jeux-vidéos, j'ai pu observer dernièrement l'explosion des jeux battle royal et leur gratuité (pour la plupart), en faisant petit à petit une espèce de norme. Ces jeux étant extrêmement plebiscité, je me suis donc posé une question : le prix du jeu-vidéo a-t-il une influence significative et systématique sur les ventes de ce dernier ?

Pour y répondre, je vais extraire et utiliser les données d'un dataset Steam, regroupant près de 20000 jeux de tout prix et de toute population de joueurs.

Partie 1

La première chose à faire (après avoir chargé le csv et importé les librairies) est de voir si l'on y trouve des données aberrantes. Pour cela, j'ai utilisé la méthode `describe`, me donnant les données résumant le contenu du fichier. On peut remarquer que le prix maximum relevé est de 400\$, très loin de la norme du prix d'un jeu qui tourne le plus souvent autour de 60\$ maximum. Afin de savoir s'il existe beaucoup d'individus supérieurs notre norme, il faut effectuer et lire un diagramme à moustache. On peut noter que très peu de jeu se trouve avoir un prix supérieur à 60\$: il faut donc les supprimer.

Une fois cela fait, on remarque que dans la section «owners», qui désigne le nombre de joueur, les données sont un string contenant une fourchette de joueurs. Pour pouvoir avoir la moyenne, il faut créer deux nouvelles colonnes possédant chacune une donnée différente : la première, «min_o» sera le nombre minimal, et max_o sera le nombre maximal. L'étape suivante est le changement de type des colonnes, les transformant donc en int. Enfin, on affecte la soustraction du maximum au minimum afin d'avoir un chiffre médian qu'on affecte à la colonne «owners» en int.

Nous pouvons enfin utiliser notre dataset, et surtout notre colonne price, pour effectuer une représentation Gaussienne, afin de pouvoir vérifier que nos données soient exploitables. Le résultat étant parlant, nous pouvons continuer.

Afin d'avoir une indication sur la corrélation pour répondre à notre problématique, nous allons utiliser une Heatmap. Le résultat montre que la corrélation entre le prix et le nombre de joueurs n'est que de 4,3%. On peut se demander si cela change si l'on supprime les jeux gratuits dans le lot. On remarque une nette augmentation, avec 11%. Ainsi, la gratuité d'un jeu ne semble pas promettre un succès auprès des joueurs. Mais l'on remarque qu'il existe quand même une corrélation un peu plus importante entre le prix et le succès des jeux payants.