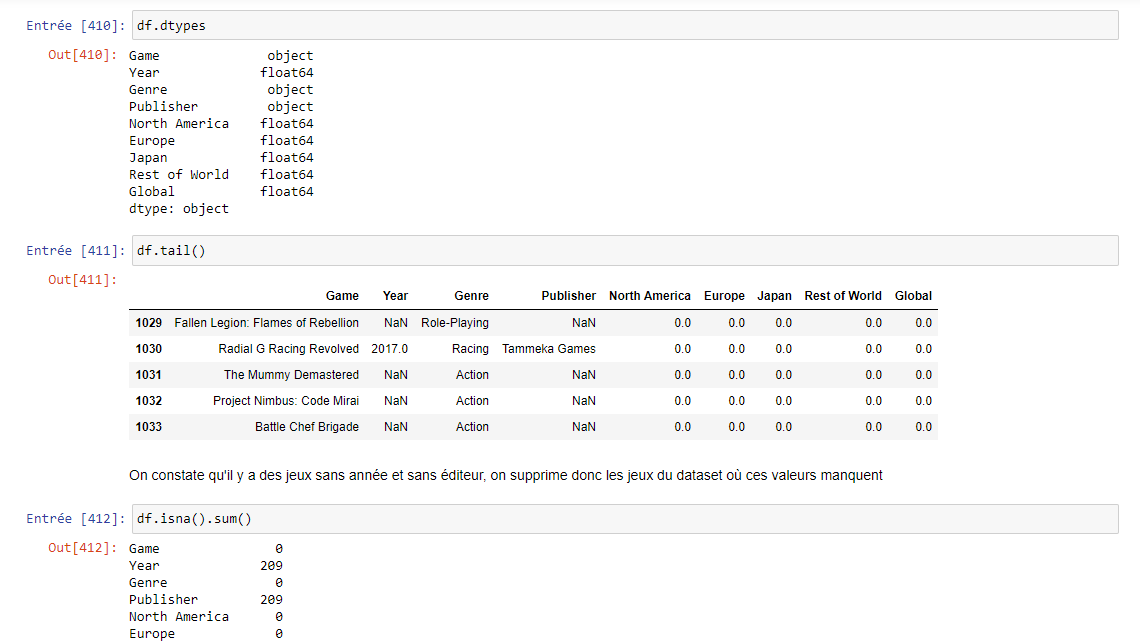
**Projet**

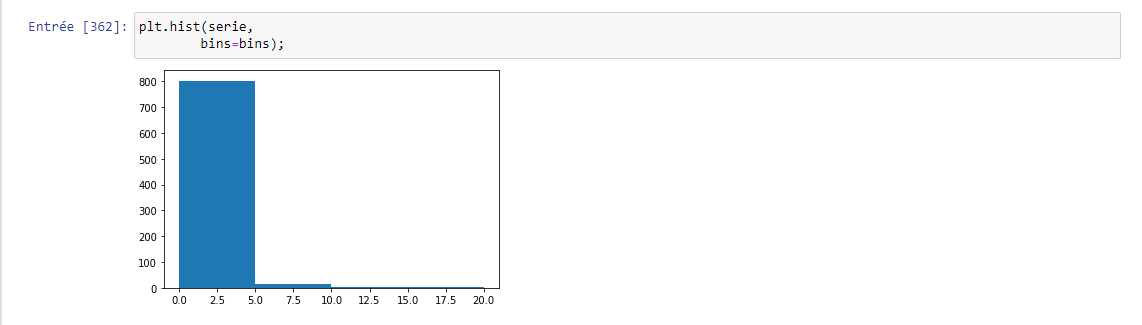
Pour ce projet on récupère un dataset regroupant les jeux vidéos vendus sur PS4. On commence ainsi par importer les librairies ainsi que le dataset :  
Une fois la librairie et le dataset importés, on commence le nettoyage du dataset :



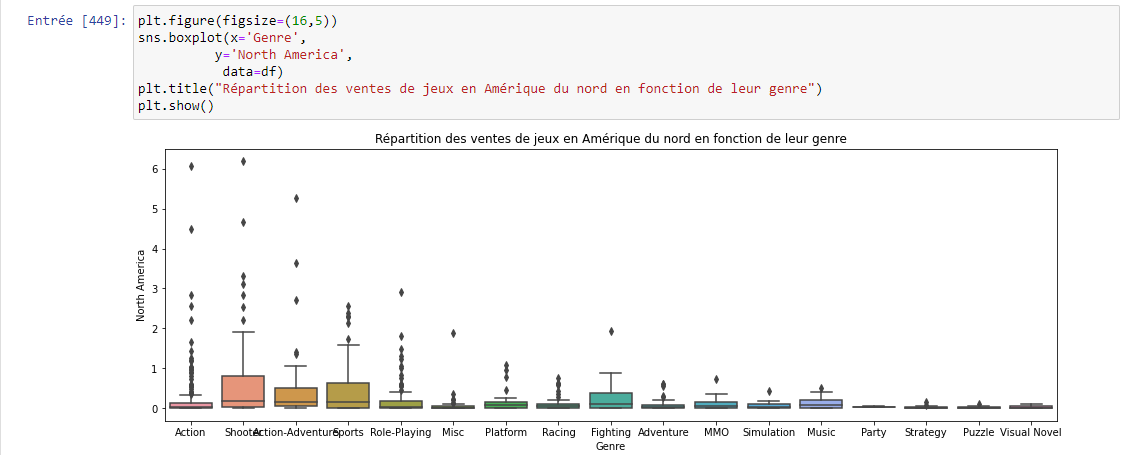
En recoupant les différents éléments et en ayant clean le dataset on peut en découler cette **problématique** : Existe-t-il une région plus propice à la vente d’un jeu ?

Pour répondre à cette problématique on réalise différents graphiques pour observer le comportement de nos colonnes :

On constate via ce graphique que la majorité des jeux ont peu de ventes totales. Nous allons donc chercher à trouver et nous rapprocher de la catégorie de jeu qui se vendent le plus :

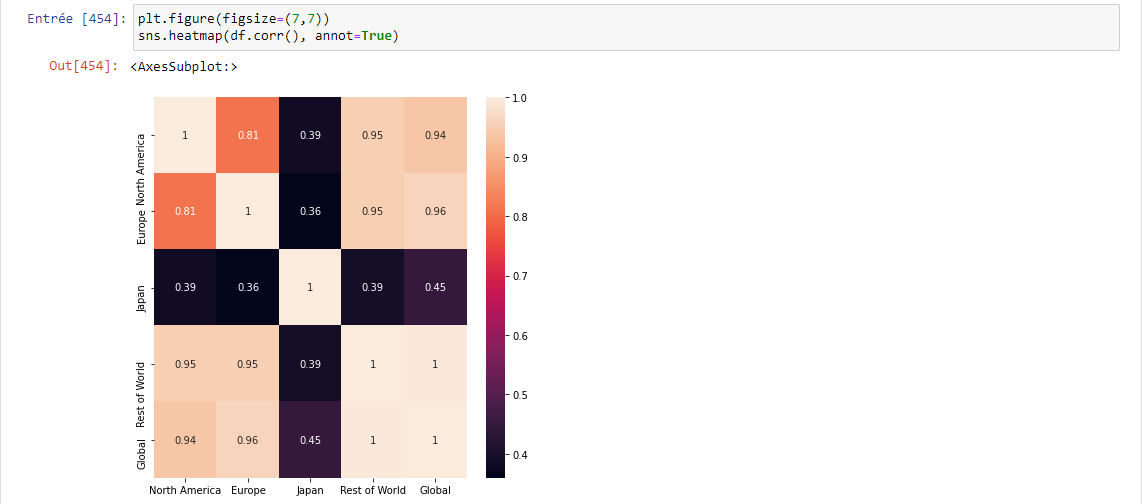


On va donc chercher à trouver quels sont les caractéristiques des jeux qui fonctionnent dans les différents pays. Pour cela nous utilisons les boîtes à moustaches :



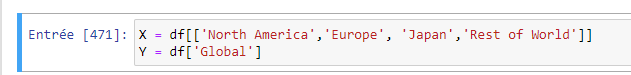
Cela nous permet de constater que certains genre de jeux ont plus de popularité selon certains pays.

Une fois que nous avons déterminé le genre nous pouvons analyser une heatmap pour voir quel corrélation existerait entre les ventes totales d’un jeu et les différentes région de distribution :



Cette Heatmap nous permet d’émettre l’hypothèse qu’il existe une corrélation forte entre les ventes totales d’un jeu et le nombre de ventes en Europe et Amérique du Nord.

Pour déterminer si cette hypothèse est vrai nous allons pouvoir tester nos variables dans un algorithme qui nous servira à faire des prédictions et démontrer des corrélations.



Nous allons pouvoir prendre commes critères nos régions, qui nous serviront à savoir si celles ci ont une corrélation forte avec les ventes totales.

Comme attendu, d’après nos hypothèses, il existe une corrélation forte entre les ventes en Amérique du Nord et en Europe par rapport aux ventes totales, et celle-ci est plus faible pour le Japon et le reste du monde :

