

# TP : Big Data

## Objectifs :

Nous souhaitons recueillir par le biais de ce projet des données permettant d'identifier le jeux vidéo de l'année 2018 (sortie sur PS4).

## Outils :

Afin de mener à bien ce projet il nous avons utilisé différents outils.

Tout d'abord, nous devons récupérer des données sur notre sujet, pour cela nous avons utilisé l'API de Twitter. Elle nous permet de réaliser des recherches sur des Tweets triés par mots clés ou encore utilisateur.

Ensuite, nous avons besoin d'analyser le texte des Tweets afin d'en tirer le sentiment qui en émane. Pour cela nous avons utilisé TextBlob (License MIT), qui est une librairie qui permet l'analyse de langage naturel, de la traduction, ou encore de l'extraction de mots clefs dans une phrase.

Puis, nous avons utilisé Tweepy (License MIT) pour faire la communication avec l'API Twitter, cette librairie nous fournis un éventail de méthodes en Python, nous permettant de simplifier l'échange avec l'API.

Enfin, afin de confirmer nos résultats nous avons souhaité se servir de Google Trends, grâce à une librairie OpenSource pyTrends (License Apache2). Cette librairie nous renvoie les résultats sous forme de tableaux Pandas.DataFrame. Pandas est une librairie qui permet la manipulation de données avec des tableaux de type Excel (Ligne/Colonne).

Pour terminé nous avons utilisé ces librairies : matplotlib nous permettant de générer les graphes, et Tableau Software pour l'affichage des données sur une carte.

## Solutions :

Nous avons dans l'optique d'analyser cinq jeux, et pour chacun d'entre eux de récupérer les mêmes données afin de les comparer.

Les jeux choisis son :

- Red Dead Redemption 2
- Monster Hunter World
- Fallout 76
- Marvel Spider-Man
- God Of War

Pour ce faire voici nos deux propositions :

### 1 - À partir de la date de sortie

De base nous souhaitons réaliser nos statistiques à partir de la date de sortie du jeux afin de connaître l'engouement autour de ce dernier, en faisant une recherche par mot clefs et grâce aussi :

- Aux nombres de tweets global depuis sa sortie
- À l'analyse des sentiments sur les tweets
- Aux tendances
  - La période où il y était
  - Nombre de fois où il y était
- Nombre de Tweets un mois après la sortie
- Nombre de Tweets deux mois après la sortie

```
{
  "red_dead_redemption_2": {
    "release_date": "26/10/2018",
    "nb_tweet_release": 0,
    "keyword": "Red Dead Redemption 2",
    "sentiment": {},
    "trends": {
      "trending": false,
      "nb_times": 0,
      "date_start": {},
      "date_end": {}
    },
    "one_month_later": {
      "nb_tweet": 0
    },
    "two_month_later": {
      "nb_tweet": 0
    }
  }
}
```

Figure 1 : Structure des données pour un jeu au format JSON.

J'ai donc créé un fichier JSON (Figure 1) permettant le stockage de toutes ces données de façon structuré afin de faciliter la génération des différents graphes.

Malheureusement lors de nos recherches pour réaliser cette étude, nous avons compris, que l'API de Twitter ne permet pas de faire des recherches sur des Tweets datant de plus de sept jours (version gratuite de l'API). Cette contrainte réduit considérablement le champ des possibilités sur les statistiques que nous souhaitons faire.

Ce qui nous a amené sur la deuxième solution.

## 2 - Analyse journalière

Comme vu ci-dessus nous avons revu nos espérances sur les données que nous pouvions tirer de Twitter, et décider de baser notre analyse sur un jour en particulier. Cela risque potentiellement de fausser les résultats mais c'est notre dernière solution.

Pour se faire nous avons revu la structure de notre JSON, afin de se rapprocher des résultats que l'on souhaite.

```
{
  "red_dead_redemption_2": {
    "release_date": "26-10-2018",
    "nb_tweets": 0,
    "keyword": "Red Dead Redemption 2",
    "sentiment": {},
    "trends": {
      "start_at": "25-10-2018",
      "end_at": "28-10-2018",
      "cities": {},
      "evolutions": {}
    }
  }
}
```

Figure 2 : Nouvelle Structure de notre JSON.

De plus nous voulions utiliser la localisation des Tweets afin de réaliser une étude sur une ville en particulier, mais malheureusement trop peu d'utilisateurs tweet en utilisant la géolocalisation. Nous avons donc choisi de se baser uniquement sur la langue utilisée dans le tweet, ici l'anglais.

Enfin, nous avons quand même décidé de récupérer les tendances pour chaque jeu, par le biais de Google Trends. Pour ce faire nous avons choisi de récupérer des statistiques sur le

trafique des mots clefs, de 1 jour avant la sortie jusqu'à 3 jours après. Nous avons récupéré aussi la liste des pays où il y a eu le plus de trafics afin de générer une carte.

### 3 - Conclusion

Pour récapituler, nous voulions récupérer les tendances sur twitter, le jour de sortie de chaque jeu, mais nous avons vu que ce n'était pas possible.

Pour pallier à ce problème, nous avons opté pour recueillir les 500 tweets les plus récents (en anglais) pour chaque jeux, au même moment, nous permettant de savoir quel jeu est le moins détesté ce jours là (grâce à l'analyse des sentiments). Puis nous avons exploité les données de Google Trends, afin d'afficher sur une carte, pour chaque jeu, les pays et le résultat de son trafique.

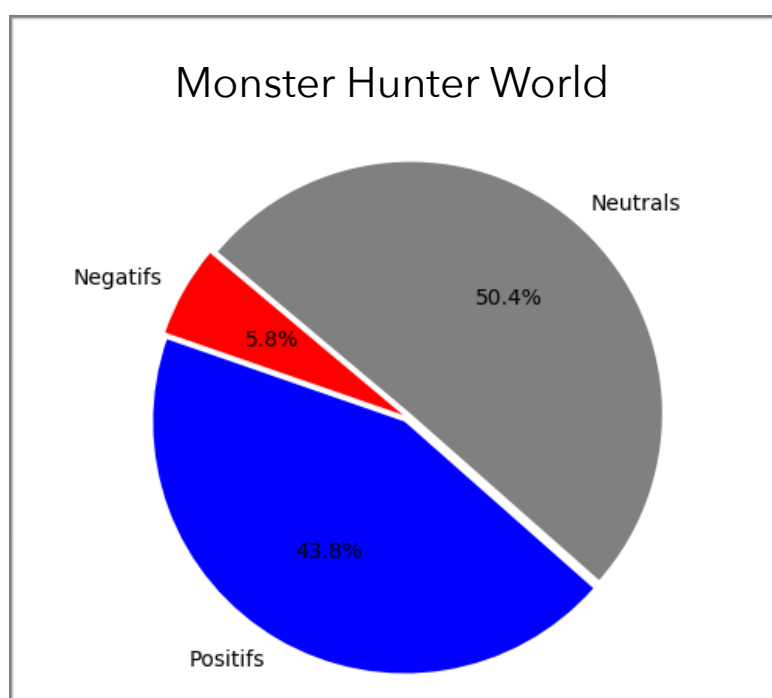
## Résultats :

### 1 - Sentiments sur les Tweets

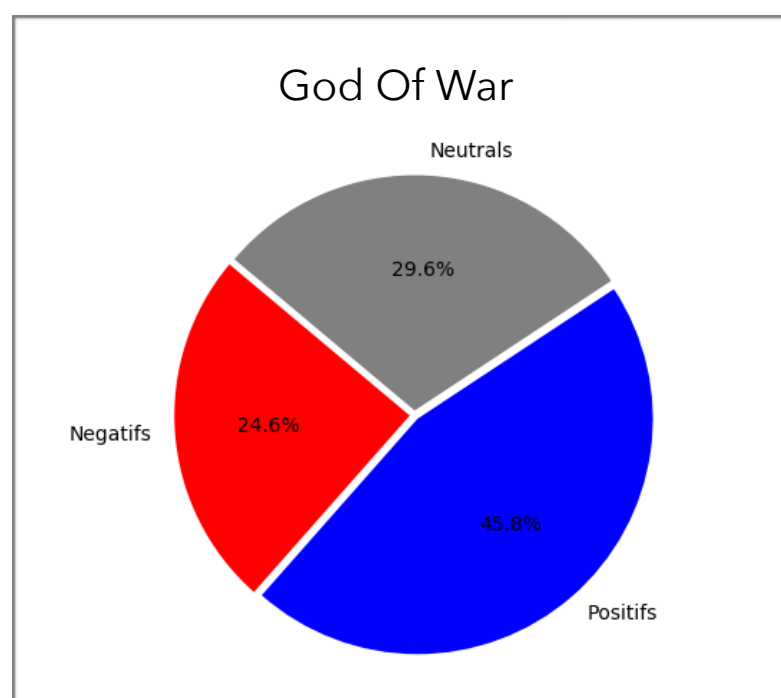
Comme vu précédemment nous avons fait notre analyse sur un échantillon de 500 tweets filtrés par le mots clefs associés au jeu.

Voici les résultats de nos recherches :

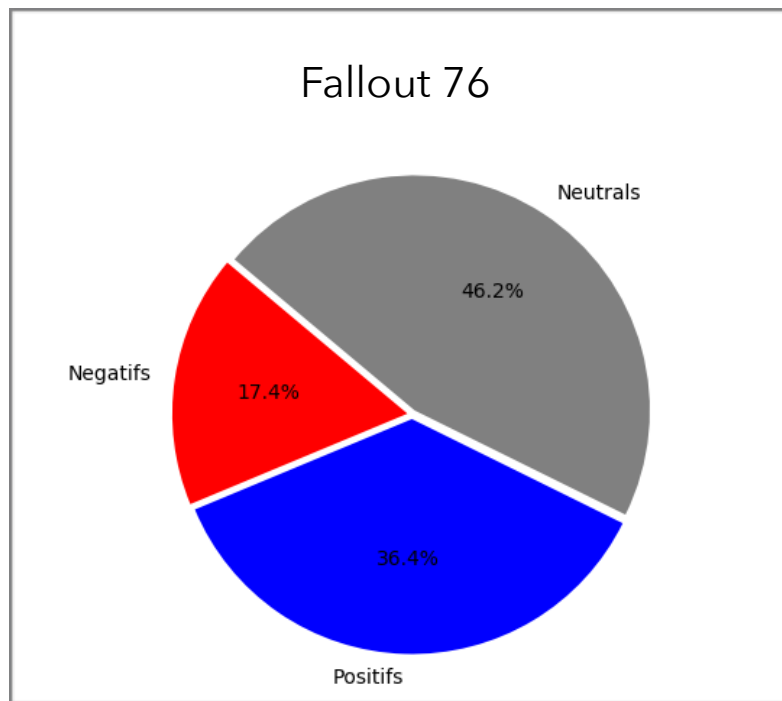
**#1**



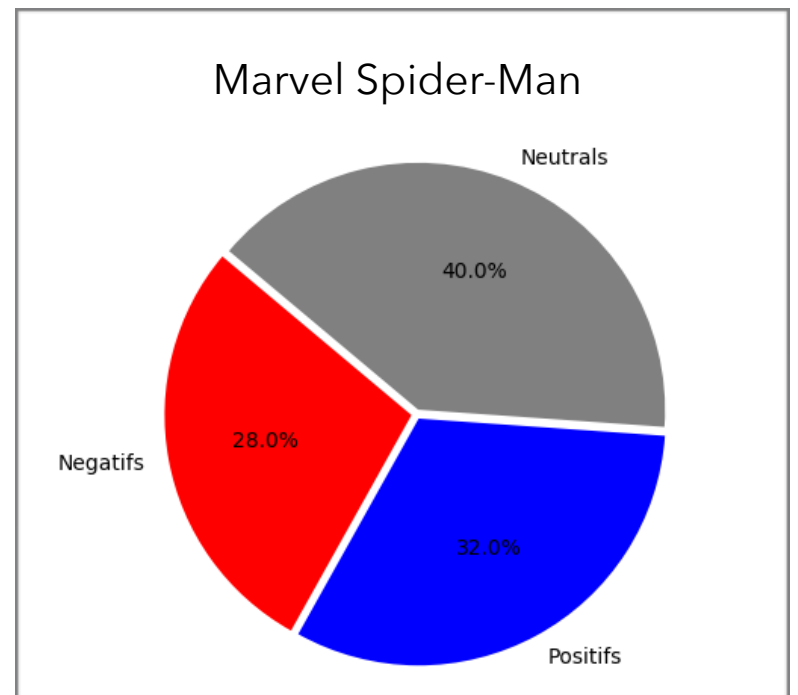
**#2**



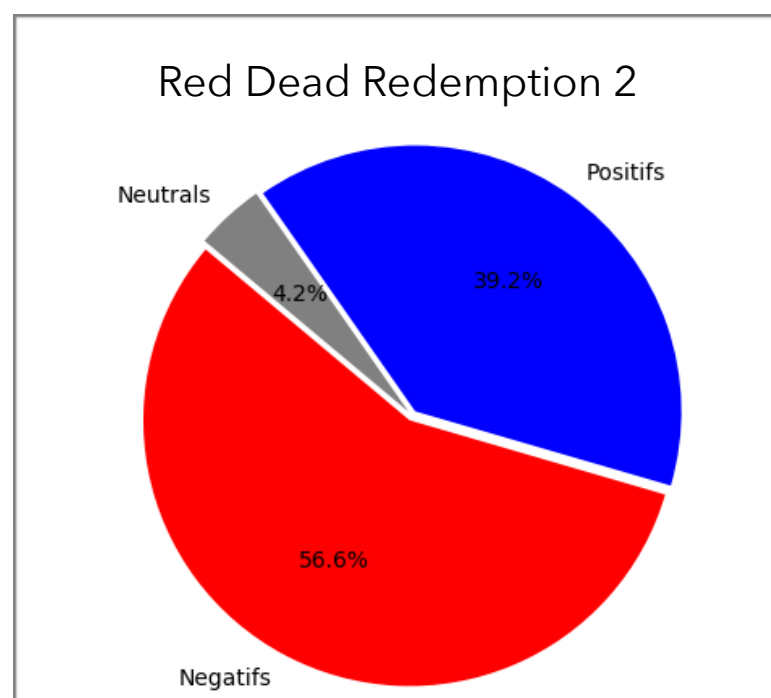
**#3**



**#4**



**#5**



Ces résultats nous donnent une première idée de ce que la communauté Twitter pense de ces cinq jeux.

## 1 - Tendances des recherches Google

Pour confirmer notre hypothèse nous avons utilisé Google Trends qui permet de connaître l'engouement autour d'une recherche. Ainsi nous avons choisi une période que l'on souhaite analyser (1 jour avant la sortie jusqu'à 3 jours après), afin de trouver une corrélation entre les deux résultats.

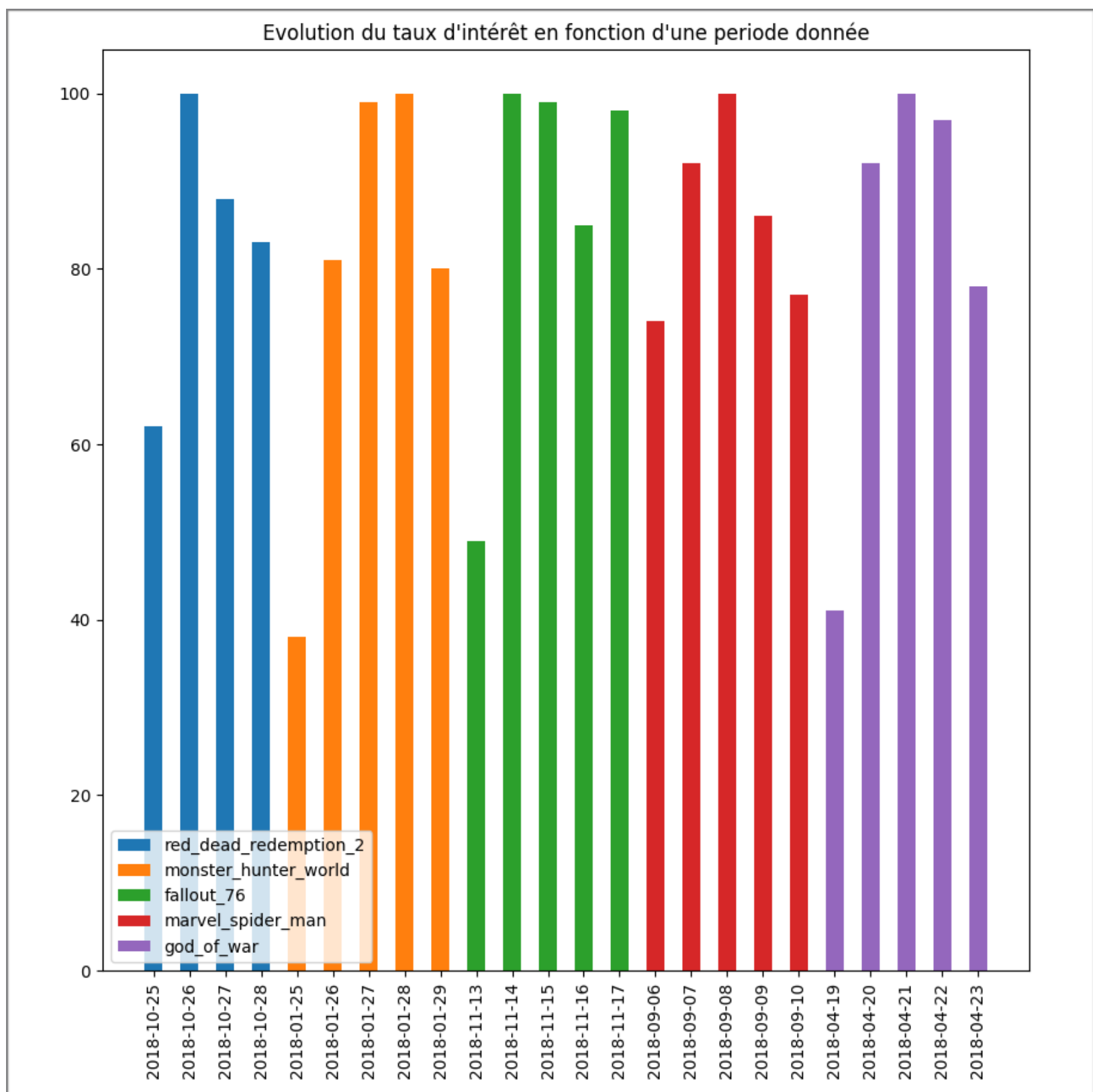


Figure 3 : Intérêt par période

Pour comprendre ce graphe il faut expliquer à quoi corresponde ces données. Google traque les différentes requêtes de ses utilisateurs anonymement, et les fournit à disposition de tous via un site internet et une API. Nous avons donc pour chaque jeu une évaluation de l'engouement des utilisateurs grâce a leurs recherches, sur une périodes données.

Ainsi on peut extraire que le jeu Fall Out 76 sort du lot avec un gros taux d'intérêt le jour de sa sortie puis plus ou moins le même, dans les trois jours qui suivent, ce qui montrent

une constance sur l'engouement autour de ce jeu. Donc nous pouvons réaliser un nouveau classement des jeux :

- #1** God Of War
- #2** Fall Out 76
- #3** Monster Hunter World
- #4** Marvel Spider-Man
- #5** Red Dead Redemption 2

D'après notre étude voici la prédiction du classement du jeux de l'année 2018.

### 3 - Les pays

Nous avons aussi pu grâce aux données de Google connaître pour chaque jeu le pays où il a été le plus recherché. Ainsi nous pouvons savoir où le jeu en question a fait le plus de bruits, nous pourrions même associer ce résultat à une étude sur la vente de ce jeu, il devrait y avoir corrélation. Nous allons donc regarder pour le gagnant, les données que nous avons pu extraire.

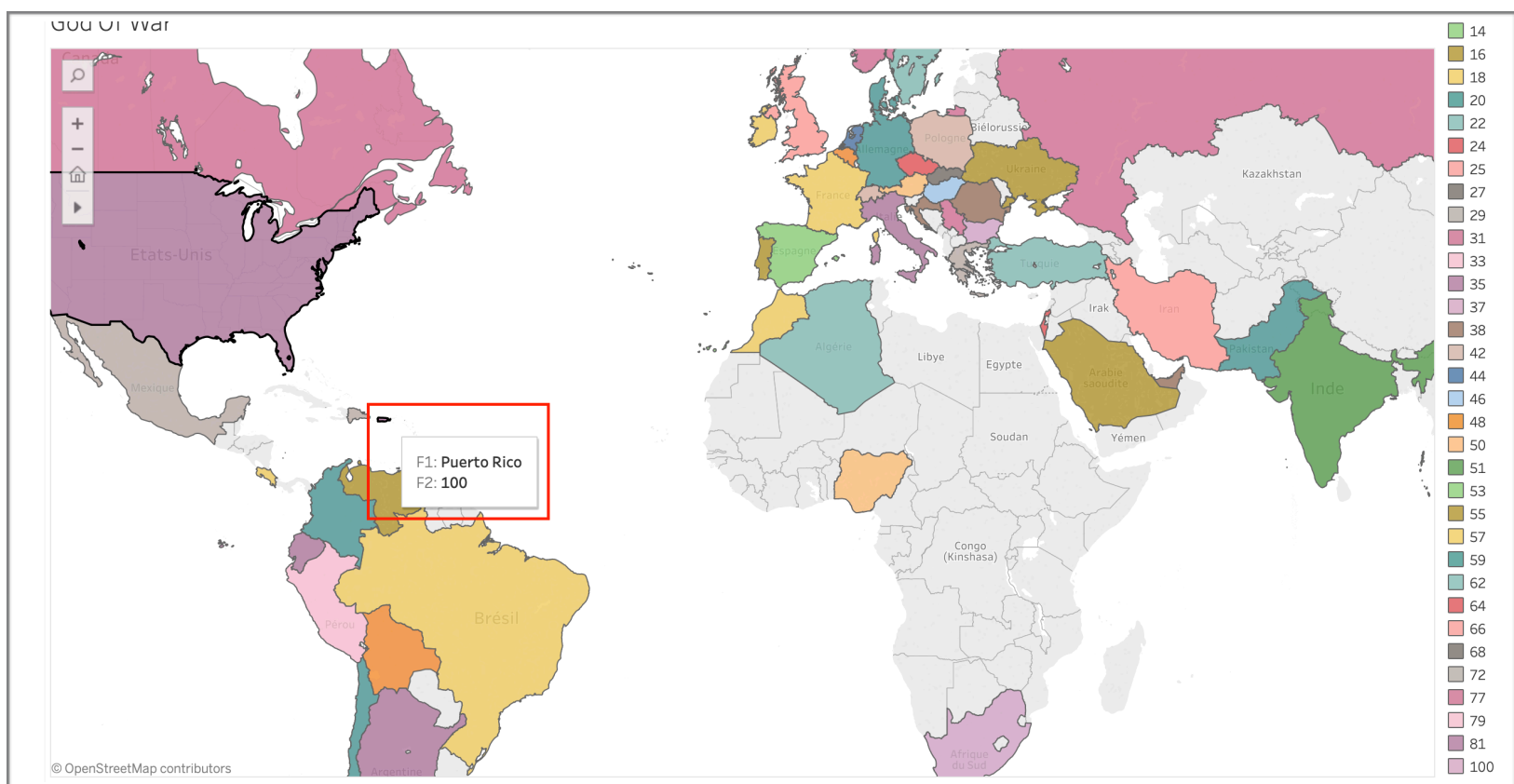


Figure 4 : Carte avec l'intérêt par Pays pour God Of War

Cette carte a été réalisée avec Tableau Software en exportant les données de Google au format CSV, puis Tableau nous permet d'avoir le rendu. Vous trouverez sur [mon profil](#) Tableau le résultat pour chacun des jeux.

## 4 - Conclusion

Pour conclure, notre prédiction pour le jeu de l'année 2018, sur PS4, est : God Of War. Cette prédiction est confirmée par un site web spécialisé en jeux vidéo : [jeuxvideo.com](#).

Filtrer parRéinitialiser tous les filtres

Plateforme

PlayStation 4 ×

Genre ▾


Mode ▾

—

Décennie ▾

Année ▾

328 résultats



### God of War

sur PS4

Dans ce nouvel épisode de God Of War, le héros évoluera dans un monde aux inspirations nordiques, très forestier et montagneux. Dans ce beat-them-all, un enfant accompagnera le héros principal, pouvant apprendre des actions du joueur.

**Sortie:** 20 avril 2018


20/20  
Test sur PS4

17.7/20  
Avis sur PS4

Forum

Vidéo test

Acheter



### Red Dead Redemption II

sur PS4

Suite du précédent volet multi récompensé, Red Dead Redemption II nous permettra de nous replonger sur PS4 dans une ambiance western synonyme de vastes espaces sauvages et de villes malfamées. L'histoire se déroule en 1899, avant le pr

**Sortie:** 26 octobre 2018

19/20  
Test sur PS4

17.2/20  
Avis sur PS4

Forum

Vidéo test

Acheter

Figure 5 : Résultat de la recherche du meilleur jeu de l'année 2018 sur PS4



Nous nous rendons quand même compte que nos résultats sont à prendre avec des pincettes. Car les données ont été prises le même jour et que certains jeux sont sortis il y a plus de dix mois. On risque donc de trouver moins d'avis négatifs sur un de ces jeux, car moins présent sur le devant de la scène et il restera uniquement les fans qui en discuteront.

## Code source

Le code source est disponible sur ce [repository](#) Github afin de faciliter l'échange. Vous y trouverez un petit « readme » si vous souhaitez plus de détails sur le code.