Business Intelligence
7. Outil de recommandation

Bado Benjamin

2020-2021

Dans le monde du jeux vidéo, quelle est la news qui vous a le plus fait rire cette semaine ?

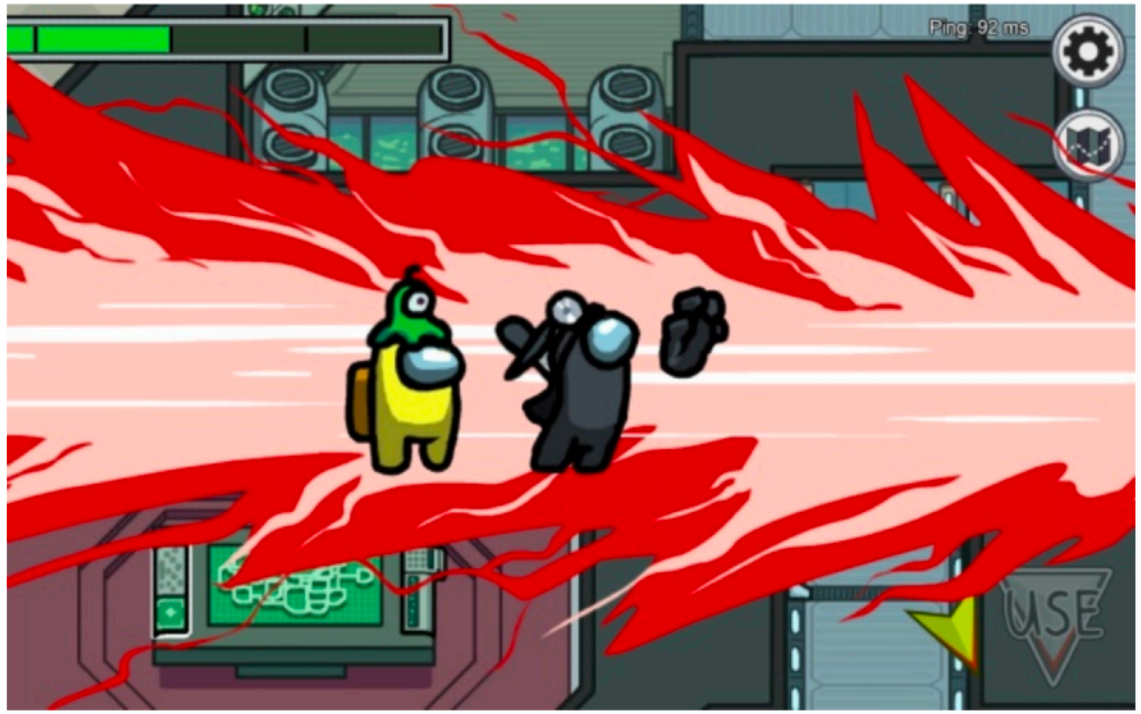
A massive spam attack is ruining public 'Among Us' games

As many as 5 million players may be affected by the hack.

Igor Bonifacic, @igorbonifacic
October 23, 2020

12 Comments

295 Shares

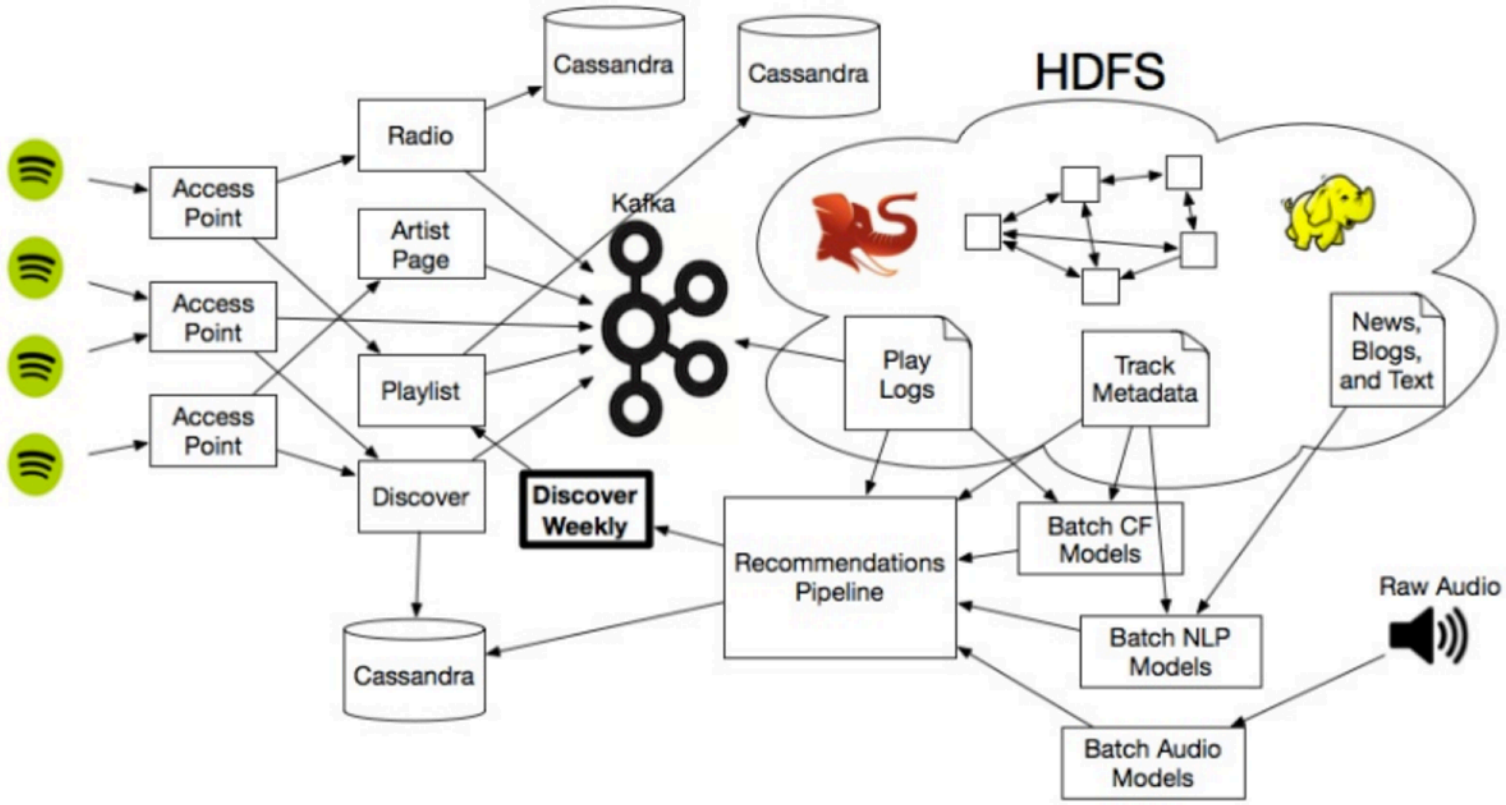


InnerSloth

Quel est le mécanisme utilisé par Spotify pour recommander des musiques ?

Architecture de Spotify

3 types de recommandation



Filtres Collaboratifs

Natural Language Processing

Modèles Audio

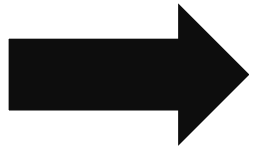
Src: https://blog.galvanize.com/spotify-discover-weekly-data-science/

Collaborative Filtering

Le filtrage Collaboratif est une méthodologie de prédiction automatique (Filtering) des centres d'intéret d'un utilisateur se basant sur la collection des préférences de nombreux utilisateurs (Collaborative)

Comment savoir à l’avance si un film risquerait de vous plaire ?

Principe algorithmique des Filtres Collaboratifs



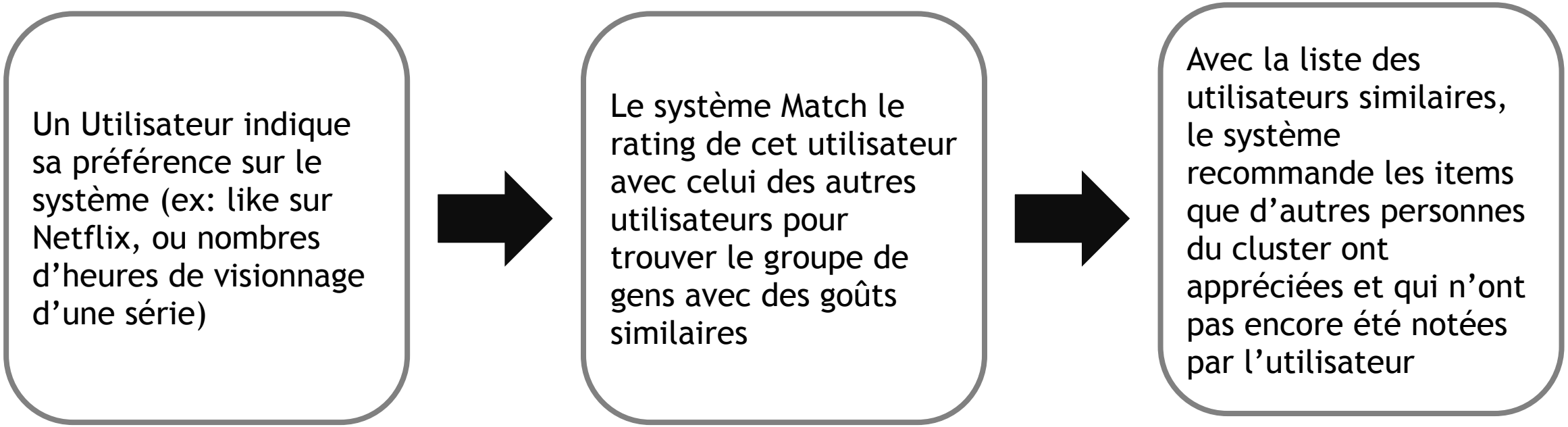
Ecouter l'avis d'un ami qui a des goûts similaires

Principe algorithmique des Filtres Collaboratifs

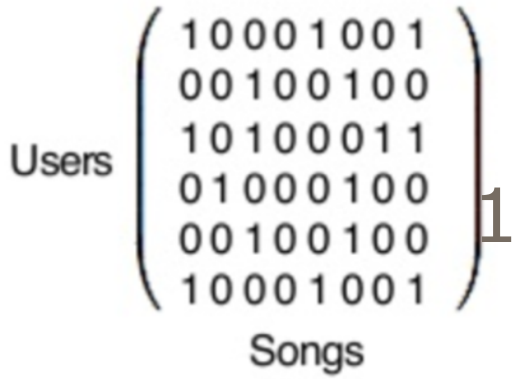
Un algorithme de Filtres Collaboratifs fonctionne en général en cherchant un sous-groupe de personnes ayant des goûts similaires.

Les Filtres Collaboratifs ont souvent besoin des éléments suivants:

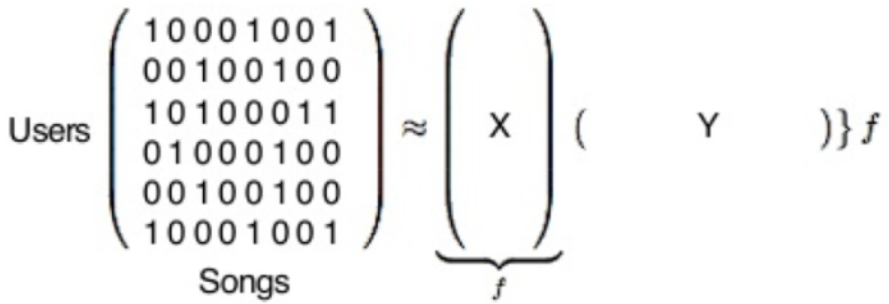
- Une participation active des utilisateurs (like, don't like, etc …)
- Une manière de représenter l'intérêt d'un utilisateur
- Un algorithme pour grouper les personnes avec des centres d'intérêts similaires



Représentation des données



Matrice gigantesque
150 millions d'utilisateurs * 30 millions de chansc



$$\min_{x,y} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i - \beta_u - \beta_i)^2 + \lambda(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

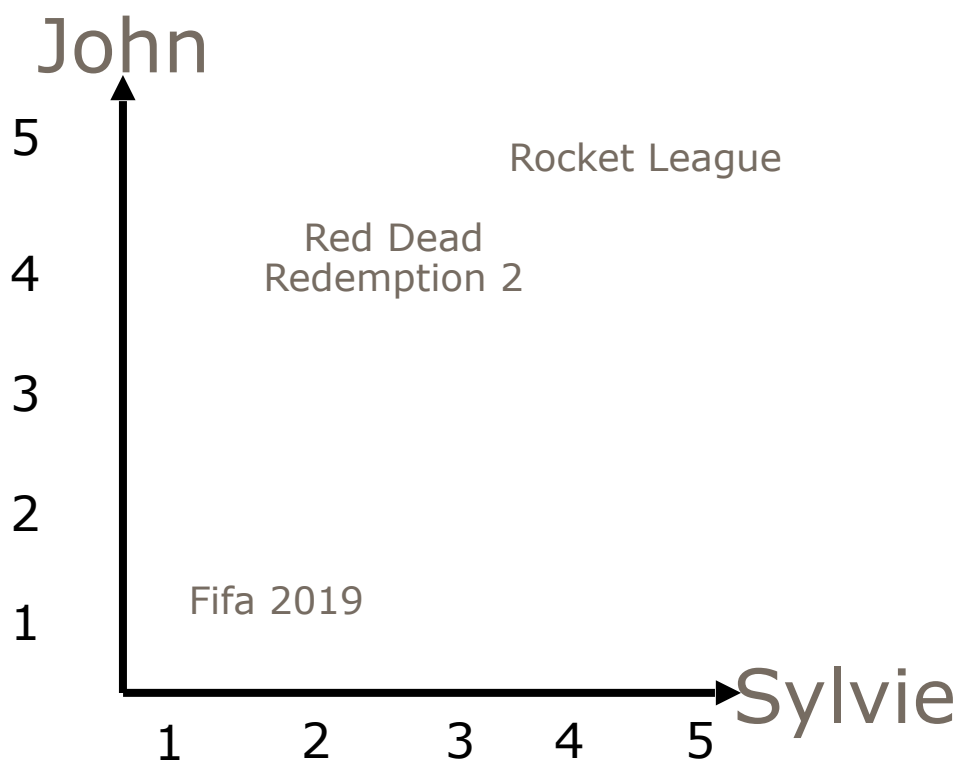
Simplifié grâce à une factorisation matricielle
Plus d'infos sur
https://www.slideshare.net/MrChrisJohnson/from-idea-to-execution-spotifys-discover-weekly/31-1_0_0_0_1

Algorithme de recommandation de jeux vidéo

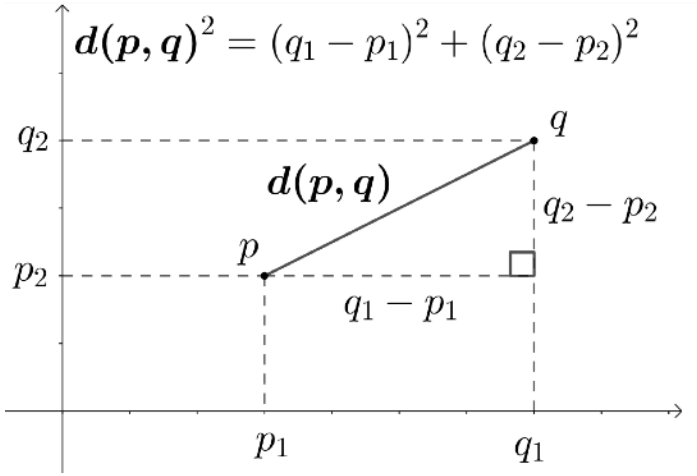
Trouver des utilisateurs similaires

Nous verrons ici 2 méthodes pour calculer un score de similarité entre 2 utilisateurs : la distance euclidienne et la corrélation de Pearson.

Distance euclidienne :
Une manière simple de calculer la similarité entre 2 personnes est de représenter les scores que les utilisateurs ont donnés en commun sur un graphe.

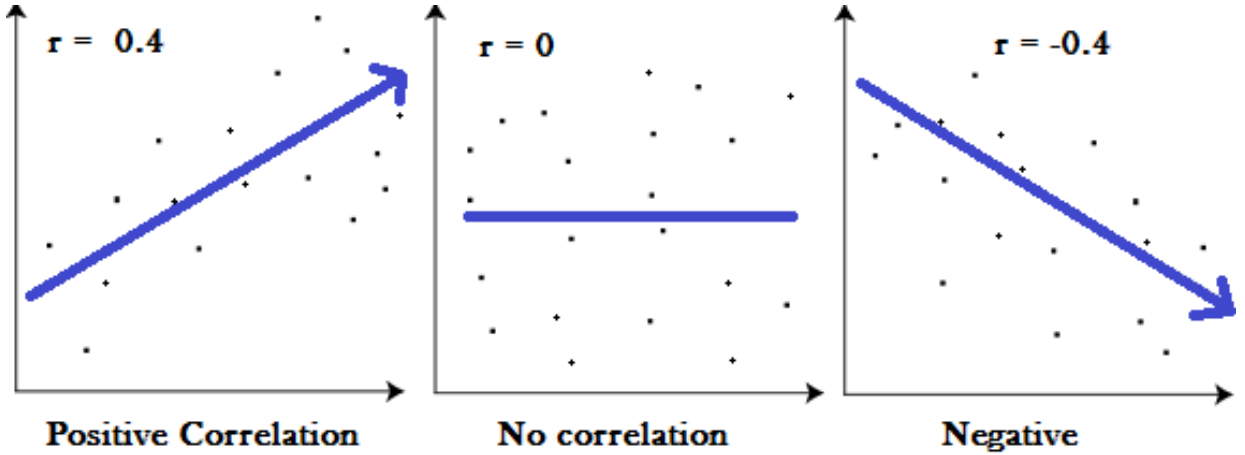


La distance euclidienne, se basant sur le théorème de Pythagore, est calculée de la manière suivante



Trouver des utilisateurs similaires – Pearson correlation score

Une corrélation de Pearson est un nombre entre -1 et 1 qui indique la corrélation linéaire entre 2 variable X et Y



L'avantage de la corrélation de Pearson sur la distance euclidienne, est que cette dernière corrige les erreurs d'inflations. Par exemple: Un utilisateur qui donne toujours un score légèrement plus grand qu'un autre, mais qui a les mêmes préférences.

Exercice

Résultat attendu

```
print(sim_distance(preferences, 'lol', 'Benj'))
```

```
0.31096533279812266
```

```
topMatches(preferences, 'Benj')
```

	users	similarity
152	Alfa Khrisna	1.000000
2	Mayné	1.000000
96	HUET	1.000000
38	Julien	1.000000
104	Alexandre Michiels	0.981981

```
getRecommendations(preferences, 'Benj')

# ca donne une idée du score que j'aurais d
```

```
C:\Anaconda3\lib\site-packages\ipykernel_la
```

Out[47]:

	game	score
10	Crusader Kings 3	3.079372
2	Pokemon: épée et bouclier	2.884676
4	FIFA 21	2.291513
0	Last of us 2	NaN
1	Super Smash Bros. Ultimate	NaN

Trouver des utilisateurs similaires

Créez une fonction sim_distance(preferences, personne1, personne2):

Cette fonction doit retourner la distance de Pearson entre la personne 1 et la personne 2

La distance de Pearson ne doit pas être calculée pour les jeux vidéo où les 2 personnes ont donné une note

La distance de Pearson est calculée en utilisant :
scipy.stats.pearsonr

Représentation des données

1) Prennez le fichier Recommendations.csv
2) Importez le dans Python en utilisant Pandas

```
import pandas
from math import sqrt

preferences = pandas.read_csv('Recommendations.csv', sep = ';')
```

3) Remplacez les nulls par des 0
   preferences.fillna(value = 0, inplace = True)

Trouver la liste des utilisateurs qui sont les plus similaires à une autre personne

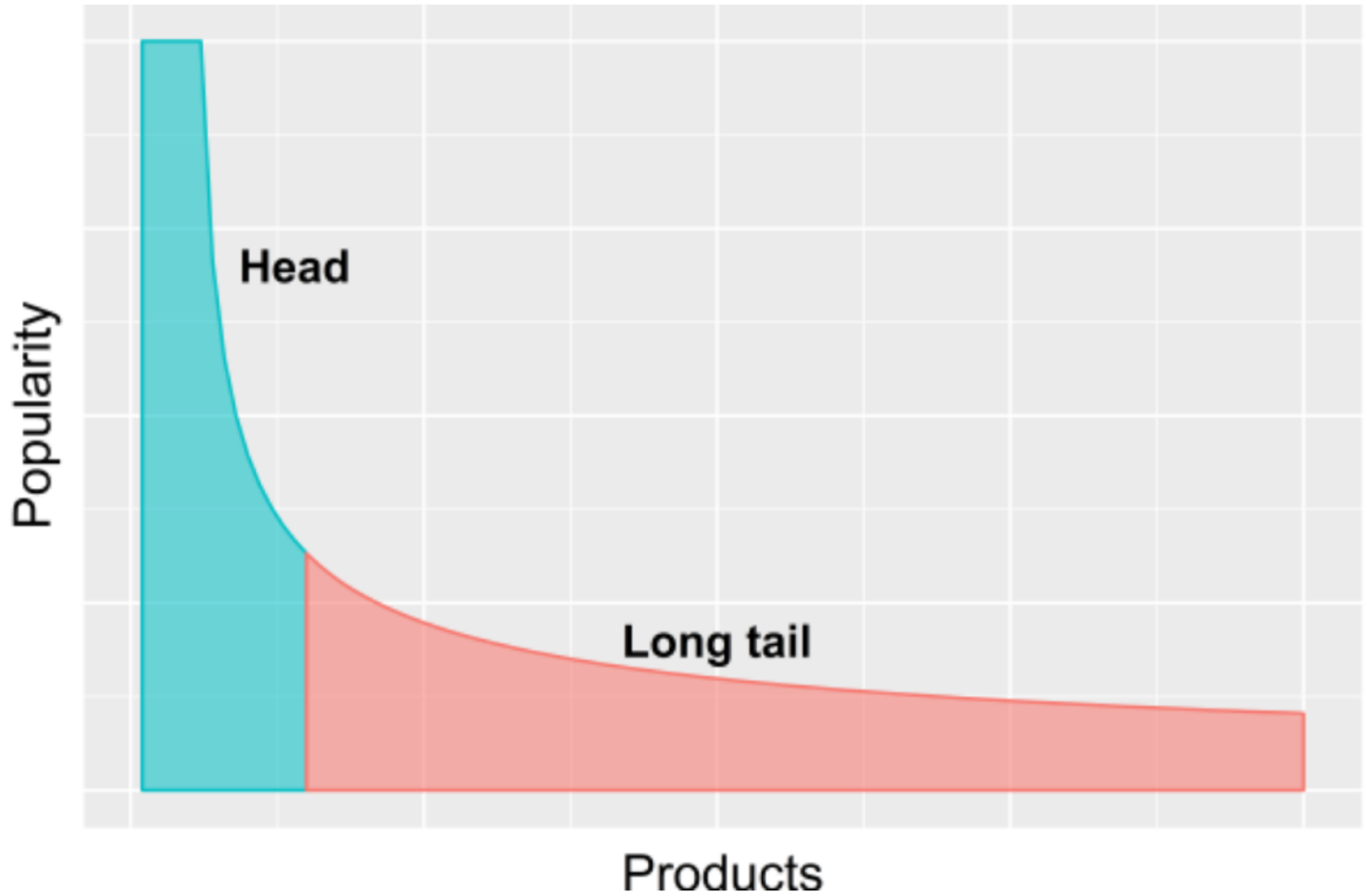
Créez une fonction topMatches(preferences, personne , nbre_de_preferences)

Cette fonction doit retourner la liste des nbre_de_preferences utilisateurs qui ont les goûts les plus proches de la personne passée en paramètre

Avoir des critiques similaires c'est bien, mais nous voulons un système de recommandation de jeux vidéo:

- On pourrait regarder la liste des personnes ayant les goûts similaires, mais ce serait trop permissif: on pourrait avoir des recommandations de personnes n'ayant pas joué aux jeux que j'aime bien

=> Il faut donner un poids pour les différents ratings de jeux


Popularity
Head
Long tail
Products

Créer la fonction pour recommander un jeu à un utilisateur

Personne	Score de similarité	Red Dead Redemption	Similarité Red Dead Redemption	Rocket League	Similarité Rocket League
John	0,99	3	2,97	5,0	4,97
Sylvie	0,38	3	1,14	3,0	1,14
Total			4,11		6,11
Somme des similarité			1,37		1,37
Total / Somme des similarité			3		4,45

Créez une fonction getRecommendations(preferences, personne):

- Cette fonction renverra la liste des jeux à recommander pour ‘personne’

Créer la fonction pour recommander un jeu à un utilisateur pseudo-code

```
# pour toutes les personnes dans préférences sauf la personne passée en paramètre
        # obtenir le score de similarité en utilisant sim_distance
        # pour toutes les préférences de l'utilisateurs
        # Pour tout les jeux auxquels je n'ai pas joué et avec les personnes ayant une similarité > 0
                # calculer la Total = similarité * score
                # calculer la valeur cumulée de similarité  SimSums

# calculer le ranking = Total / SimSums
# trier Total
# retourner Total
```

Et comment faire pour regarder les jeux similaires ?

Simple: on retourne la matrice utilisateurs jeux