

Analyse de données - bibliothèque pandas - introduction

Code de partage avec Capytale : 54e6-1233372

1 Préparation

Dans un premier temps, nous allons importer un jeu de données.

Pour cela, avec Pyzo, il faut copier le fichiers films.csv et le placer dans le même dossier que le fichier Python avec lequel vous travaillerez.

Puis dans le fichier Python (ou de même dans l'activité sur Capytale), on écrira et on exécutera :

```
import pandas
films=pandas.read_csv('films.csv', delimiter=';', encoding='utf8')
```

2 Exercice

1. Avec ce jeu de données, on pourra tester les commandes suivantes et donner leur fonction :

<code>films.head()</code>	
<code>films.sample(7)</code>	
<code>films.columns</code>	
<code>films.dtypes</code>	
<code>films.describe()</code>	
<code>films.shape</code>	
<code>films.nlargest(..., 'revenue')</code>	
<code>films['release_year']</code>	
<code>films[['title', 'release_year']]</code>	
<code>films[(films['release_year'] ==2000) & (films['revenue'] >100000)]</code>	

2. Trouver des commandes qui permettent de répondre aux questions suivantes
 - De quelle année date le film le plus ancien et quelle était sa langue originale ?
 - Quel est le film de l'année 2011 le plus long ?
 - Quel a été le revenu cumulé des films des années 1990 à 1999 (incluses) ? de 2000 à 2009 (incluses) ?
 - Combien de langues sont représentées ?