

# Analyse exploratoire

Projet : Critiques de films



# Contexte

Le jeux de données contient 147 films depuis 2015 et possède des reviews de critiques et d'utilisateurs et propose une moyenne des notes des critiques et des utilisateurs.

Sur le site Fandango les notes des films semble sensiblement mieux noté que sur les autres sites. De plus Fandango a la particularité de vendre ces propres ticket de cinémas.

- Les notes sont-elles sur/sous évalué par Fandango pour augmenter ces vente ?

# Diagrammes en barres et graphiques à nuage de points

- Lire 'fandango\_scores.csv' dans un dataframe reviews.
- Sélectionner les colonnes suivantes et assigner les au dataframe norm\_reviews:
  - 'FILM'
  - 'RT\_user\_norm'
  - 'Metacritic\_user\_nom'
  - 'IMDB\_norm'
  - 'Fandango\_Ratingvalue'
  - 'Fandango\_Stars'
- Afficher les premières lignes de norm\_reviews.

# Diagrammes en barre

- Créer un sous-graphique avec `plt.subplot()` et assigner l'objet Figure à la variable `fig` et l'objet Axes à la variable `ax`.
- Générer un diagramme en barre avec:
  - Les coordonnées x de toutes les barres définie par `bar_positions` qui doit prendre les valeurs 1,2,3,4 et 5
  - La hauteur de chaque barres définie par `bar_heights` qui doit prendre les 5 colonnes qui nous intéressent pour le film 'Avengers: Age of Ultron (2015)'.
  - Avec l'argument `width` vous pouvez gérer l'épaisseur des barres. Fixé-là à 0.5
- Afficher le diagramme en barre.

# Aligner l'intitulé et le trait sur l'axe x

- Créer un sous-graphique et assigner l'objet Figure à la variable fig et l'objet Axes à la variable ax.
- Générer un diagramme en barre avec:
  - Les coordonnées x de toutes les barres définie par bar\_positions qui doit prendre les valeurs 1,2,3,4 et 5
  - La hauteur de chaque barres définie par bar\_heights qui doit prendre les 5 colonnes qui nous intéressent pour le film 'Avengers: Age of Ultron (2015)'
  - Avec l'argument width vous pouvez gérer l'épaisseur des barres. Fixé-là à 0.5
- Placer les 5 traits correspondant aux barres.
- Placer les intitulés de num\_cols et faire une rotation de 90 degrés.
- Nommer l'axe x par 'Sources de notation'.
- Nommer l'axe y par 'Note moyenne'.
- Définir le titre par 'Moyenne des notes utilisateurs pour le film Avengers: Age of Ultron (2015)'.
- Afficher le diagramme.

# Diagramme en barres horizontales

- Créer un sous-graphique et assigner l'objet Figure à la variable fig et l'objet Axes à la variable ax.
- Générer un diagramme en barres avec:
  - bottom pour les positions bar\_positions
  - width pour la largeur des barres bar\_widths
  - height fixé à 0.5
- Placer les 5 traits correspondants aux barres sur l'axe y.
- Placer les intitulés de num\_cols.
- Définir l'axe y par 'Sources de notes'.
- Définir l'axe x par 'Note moyenne'.
- définir le titre par 'Note moyenne utilisateurs pour le film Avengers: Age of Ultron (2015)'.
- Afficher le diagramme en barre.

# Graphique à nuage de points

- Créer un sous-graphique et assigner l'objet Figure à la variable fig et l'objet Axes à la variable ax.
- Générer un graphique à nuage de points avec la colonne 'Fandango\_Ratingvalue' sur l'axe x et la colonne 'RT\_user\_norm' pour l'axe y.
- Définir pour l'axe x 'Fandango' et 'Rotten Tomatoes' pour l'axe y.
- Afficher le graphique résultant.

# Intervertir les axes

- Assigner à la variable fig la figure retourné par plt.figure()
- Nous allons ici un graphiques avec deux sous-graphiques à l'aide de la méthode add\_subplot() répartie sur une ligne et deux colonnes:
  - Avec fig.add\_subplot(...,...,...) assigné à ax1 l'axe de la 1<sup>ère</sup> figure.
  - Avec fig.add\_subplot(...,...,...) assigné à ax1 l'axe de la 2<sup>ème</sup> figure.
- Pour le sous-graphique associé à la variable ax1.
  - Générer un graphique à nuage de points avec la colonne 'Fandango\_Ratingvalue' sur l'axe x et la colonne 'RT\_user\_norm' sur l'axe y
  - Définir pour l'axe x 'Fandango' et pour l'axe y 'Rotten Tomatoes'
- Pour le sous-graphique associé à la variable ax2.
  - Générer un graphique à nuage de points avec la colonne 'RT\_user\_norm' sur l'axe x et la colonne 'Fandango\_Ratingvalue' sur l'axe y
  - Définir pour l'axe x 'Rotten Tomatoes' et pour l'axe y 'fandango'
- Afficher le graphique résultant.



# Comparaison de corrélations

Tracer un graphiques avec 3 sous-graphiques:

- Pour le sous-graphique associé à la variable ax1:
  - Générer un graphique à nuage de points avec la colonne 'Fandango\_Ratingvalue' sur l'axe x et la colonne 'RT\_user\_norm' sur l'axe y
  - Définir pour l'axe x 'Fandango' et pour l'axe y 'Rotten Tomatoes'
  - Définir les limites des axes x et y en les fixant de 0 à 5 avec `.set_xlim(...)` et `set_ylim(...)`
- Pour le sous-graphique associé à la variable ax2:
  - Générer un graphique à nuage de points avec la colonne 'Fandango\_Ratingvalue' sur l'axe x et la colonne 'Metacritic\_user\_nom' sur l'axe y
  - Définir pour l'axe x 'Fandango' et pour l'axe y 'Metacritic'
  - Définir les limites des axes x et y en les fixant de 0 à 5
- Pour le sous-graphique associé à la variable ax3:
  - Générer un graphique à nuage de points avec la colonne 'Fandango\_Ratingvalue' sur l'axe x et la colonne 'IMDB\_norm' sur l'axe y
  - Définir pour l'axe x 'Fandango' et pour l'axe y 'IMDB'
  - Définir les limites des axes x et y en les fixant de 0 à 5
- Afficher le graphique.

# Distribution de fréquences

- Utiliser la méthode `value_counts()` pour retourner le décompte des fréquences de la colonne 'IMDB\_norm'. Trier l'objet Series résultant par son index et assigner le résultat à la variable `imdb_distribution`.
- Afficher le résultat.
- Créer un sous-graphique et assigner les objets Figure à la variable `fig` et Axes à la variable `ax`.
- Générer un histogramme des valeurs de la colonne 'Fandango\_Ratingvalue' en utilisant un range de 0 à 5.
- Utiliser `plt.show()` pour afficher le graphique.

# Comparaison d'histogrammes

Tracer un graphique avec 4 sous-graphiques:

- Pour le sous-graphique associé à la variable ax1:
  - Générer un histogramme des valeurs de la colonne 'Fandango\_Ratingvalue' en utilisant 20 classes (argument bins=20) et un intervalle de 0 à 5 (argument range=(0,5)).
- Pour le sous-graphique associé à la variable ax2:
  - Générer un histogramme des valeurs de la colonne 'RT\_user\_norm' en utilisant 20 classes (bins) et un intervalle de 0 à 5
- Pour le sous-graphique associé à la variable ax3:
  - Générer un histogramme des valeurs de la colonne 'Metacritic\_user\_nom' en utilisant 20 classes (bins) et un intervalle de 0 à 5
- Pour le sous-graphique associé à la variable ax4:
  - Générer un histogramme des valeurs de la colonne 'IMDB\_norm' en utilisant 20 classes (bins) et un intervalle de 0 à 5
- Pour tous les graphiques définir l'intervalle sur l'axe y de 0 à 50 en utilisant Axes.set\_ylim() et choisir un titre.
- Afficher les histogrammes.

# Plusieurs diagrammes à boîtes

- Créer un sous-graphique et assigner les objets Figure à la variable fig et Axes à la variable ax.
- Générer un diagramme à boîtes pour chaque colonne de num\_cols = ['RT\_user\_norm', 'Metacritic\_user\_nom', 'IMDB\_norm', 'Fandango\_Ratingvalue'].
- Définir pour chaque diagramme un intitulé correspondant au nom de la colonne num\_cols et faire une rotation à 90 degrés de cet intitulé.
- Définir l'intervalle de l'axe y de 0 à 5.
- Afficher le graphique.