

S'entraîner avec pandas

Jouez les pandas-détectives

Tout comme l'exercice sur numpy, celui-ci a pour but de vous forger une expérience « pandas ». À travers plusieurs petites missions, vous utiliserez ce que l'on a vu en classe et vous découvrirez, par la même occasion, de nouvelles fonctionnalités.

Le plan de votre enquête sera divisé en plusieurs grandes parties¹ :

1. Commençons par le commencement.
2. Analysons de plus près les « ratings ».
3. Un peu de nettoyage n'a jamais fait de mal...
4. Recherchons les applications populaires.

Commençons par le commencement

1. Chargez votre fichier de données et nommez-le « google ».
2. Enquêtez sur sa structure. Regardez sa taille (shape), le nombre total d'informations (size) et le type des colonnes (info).
3. Sélectionnez les 10 premières lignes de votre fichier de données.
4. Sélectionnez les 10 dernières lignes de votre fichier de données.

Analysons de plus près les « ratings »

1. Sélectionnez la colonne « Rating » et enregistrez-en une copie (méthode copy) dans la variable « ratings ».
2. Observez ce que la méthode « describe » vous renvoie sur cette dernière.
3. Supprimez les valeurs aberrantes s'il y en a.
4. Réopérez la méthode « describe » afin de vérifier si tout est normal.
5. Essayons d'en savoir plus sur la composition de nos données avec l'aide de [value counts](#).
6. Affinez les résultats obtenus grâce au paramètre nommé « bins » et assignez-lui la valeur 5.
7. Vous avez gagné en netteté de l'information. Néanmoins, créez vos propres « bins » en utilisant « [interval range](#) ».
8. Grâce à ces dernières informations, réalisez un graphique en camembert. Pour ce faire, mettez vos données dans un « [pandas. Series](#) » et appelez dessus la méthode « [plot.pie \(\)](#) ».

¹ Il n'y a pas de suite logique entre les différentes étapes (la première est toutefois l'exception qui confirme la règle).

Un peu de nettoyage n'a jamais fait de mal...

1. Sauvegardez dans une variable « android_ver » une copie de la colonne « Android Ver ».
2. Observez les données avec la méthode « value_counts ».
3. Créez une méthode qui ne retient que le chiffre de la version du système d'exploitation. Par exemple, si la version vaut 4.0.3, alors vous retenez le chiffre 4.
4. Passez en revue toutes vos données avec la méthode « [apply](#) » et persistez les résultats dans une nouvelle colonne prénommée « Major Android Ver ».
Attention, veuillez ne garder que des valeurs entières !
5. Effectuez, cette fois-ci, un « value_counts » sur votre nouvelle colonne. De plus, calculez les pourcentages qui leur sont relatifs (un certain paramètre nommé de « value_counts » peut s'avérer assez utile).
6. Enfin, réalisez un histogramme afin de rendre tous ces chiffres un peu plus parlants.

Recherchons les applications populaires

1. Triez vos informations, par ordre **décroissant**, à travers le nombre d'installations et affichez les noms des 10 premiers résultats ainsi que l'information que vous avez utilisée pour le tri.
2. En y regardant de plus près, nous pouvons remarquer la présence de doublons. Remontez dans votre notebook et supprimez-les en employant « [drop_duplicates](#) ». N'oubliez pas de relancer toutes vos cellules.
3. Améliorons notre tri en passant cette fois-ci deux noms de colonnes : « Installs » et « Reviews ».