

Conseils pour le TP Prepa Feature



A n'utiliser que si vous êtes bloqué.

Vous apprendrez mieux si vous n'utilisez pas ces conseils, mais cherchez par vous même.

1.1

Utiliser la méthode pandas `read_csv`.

Utiliser la méthode pandas `sample` pour afficher des exemples.

1.2

idem 1.1

1.3

Utiliser la fonction `info()` sur chaque DataFrame

1.4

Utiliser la fonction `mean()` sur la colonne `Survived`

2.1

Regrouper par GroupBy et appliquer la fonction mean() à chaque groupe.

2.2

Idem 2.1

2.3

La taille de la famille est le nombre de frères/soeurs/conjoints (sibsp) + le nombre de parents/enfants (parch) + 1 (la personne).

On peut simplement ajouter les colonnes et utiliser la vectorisation pour ajouter 1.

2.4

Encore un `groupBy` auquel on applique les opérations `count()` et `mean()`

2.5

On peut contruire IsAlone en testant si la colonne FamilySize est > 1

2.6

Toujours un GroupBy

2.7

Vous pouvez utiliser l'exemple du fichier `cours1_exemples_data_collect_iris.ipynb`
Attention à bien définir la liste `discrete_features`

2.8

Reprendre l'exemple du fichier `cours1_exemples_data_collect_iris.ipynb`

3.1

Fonction fillna

3.2

Fonction GroupBy et mean

3.3

Utiliser les fonctions fillna et median

3.4

Utiliser cut ou qcut (voir exemple Cours1_IntroPython.ipynb)

3.5

Fonction GroupBy et mean

3.6

Pour compter le nombre de valeurs manquantes, on peut combiner les fonctions `isnull` (qui renvoie 1 si la valeur est nulle) et `sum`.

L'écart type peut être calculé avec la fonction `std()`

La fonction `np.random.randint()` permet de générer une liste aléatoire d'entiers entre 2 bornes.

3.7

Utiliser cut ou qcut (voir exemple Cours1_IntroPython.ipynb)

3.8

Fonction GroupBy

4.1

Les titres sont suivis d'un point. Faire une recherche d'expression régulière en utilisant la librairie python `re`.

4.2

Aller voir la fonction `crosstab` dans la documentation pandas.

4.3

Utiliser la fonction `pandas.replace` sur la colonne voulue.

5.1

Utiliser la fonction `map` comme dans l'exemple.

5.2

Utiliser la fonction `map` comme dans l'exemple.

5.3

Utiliser la fonction `fillna` comme dans l'exemple.

5.4

Utiliser quatre fois la fonction `loc` avec une condition sur la variable `Fare`.
Finissez par la fonction `astype()` pour convertir en entier.

5.5

Utiliser la fonction `map` comme dans l'exemple.

5.6

Utiliser la fonction `drop` en faisant attention aux différences entre les datasets Train et Test.

5.7

Utiliser à nouveau la fonction `info()`.

5.8

Utiliser la fonction `to_csv` avec le paramètre `index`.