Fondamentaux de Python

Solutions

# Explorer une base de données

Exerçons nous sur Pandas. Dans le drive, téléchargez le dataset suivant : “chipotle.csv”. Ouvrez une console en Python :

1. Importez les bonnes librairies

|  |
| --- |
| import pandas as pd |

1. Importez le dataset

|  |
| --- |
| dataset = pd.read\_csv("dataset.csv") |

1. Regardez les 10 premières lignes du dataset

|  |
| --- |
| dataset.head() |

1. Quelle est la taille de notre dataset ?

|  |
| --- |
| dataset.describe() |

1. Affichez toutes les colonnes de notre dataset

|  |
| --- |
| dataset.columns |

1. Quel est l’item le plus commandé
   1. Indice : On utilisera GROUP BY & Sort\_values()

|  |
| --- |
| dataset.groupby("item\_name").sum().sort\_values(["quantity"], ascending= False) |

1. Combien d’items ont été commandés au total chez Chipotle

|  |
| --- |
| dataset.quantity.sum() |

1. Combien de chiffre d’affaire à realisé Chipotle ?
   1. Convertissez item\_price en nombre décimal
      1. Regardez dataset[“item\_price”], que voyez vous ?
      2. Trouvons un moyen d’enlever ce $ qui nous gène. Pour cela, on peut utiliser l’objet *.str* (<https://bit.ly/2ClcdtN>)
      3. Convertissons maintenant la série en float au lieu de string
   2. Multipliez la quantité vendues par le prix de l’item
   3. Faites la somme du tout, combien trouvez vous?
   4. Arrondissez à deux chiffres après la virgule
      1. On pourra utiliser la fonction ([np.round()](https://docs.scipy.org/doc/numpy/reference/generated/numpy.ndarray.round.html#numpy.ndarray.round))

|  |
| --- |
| dataset.item\_price = dataset.item\_price.str[1:-1] dataset.item\_price = pd.Series(dataset.item\_price, dtype="float") (dataset.item\_price \* dataset.quantity).sum()  np.round((dataset.item\_price \* dataset.quantity).sum(), 2) |

1. Quel est le revenu moyen par commande ?

|  |
| --- |
| dataset["revenue"] = dataset["quantity"] \* dataset["item\_price"] dataset.groupby("order\_id").sum().mean() |