

Part I: BestRunCorp

<https://github.com/Emmanuel75/Data>

Packages

```
from datetime import datetime
import pandas as pd

import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt

import numpy as np
```

Data

1. Get BestRunCorp_Retail.csv from <https://github.com/Emmanuel75/Data> and read it in a pandas dataframe
df = pd.read_csv("BestRunCorp_Retail.csv")
2. Display the shape and the head of the dataframe
3. get the description (df.describe())

Data cleansing

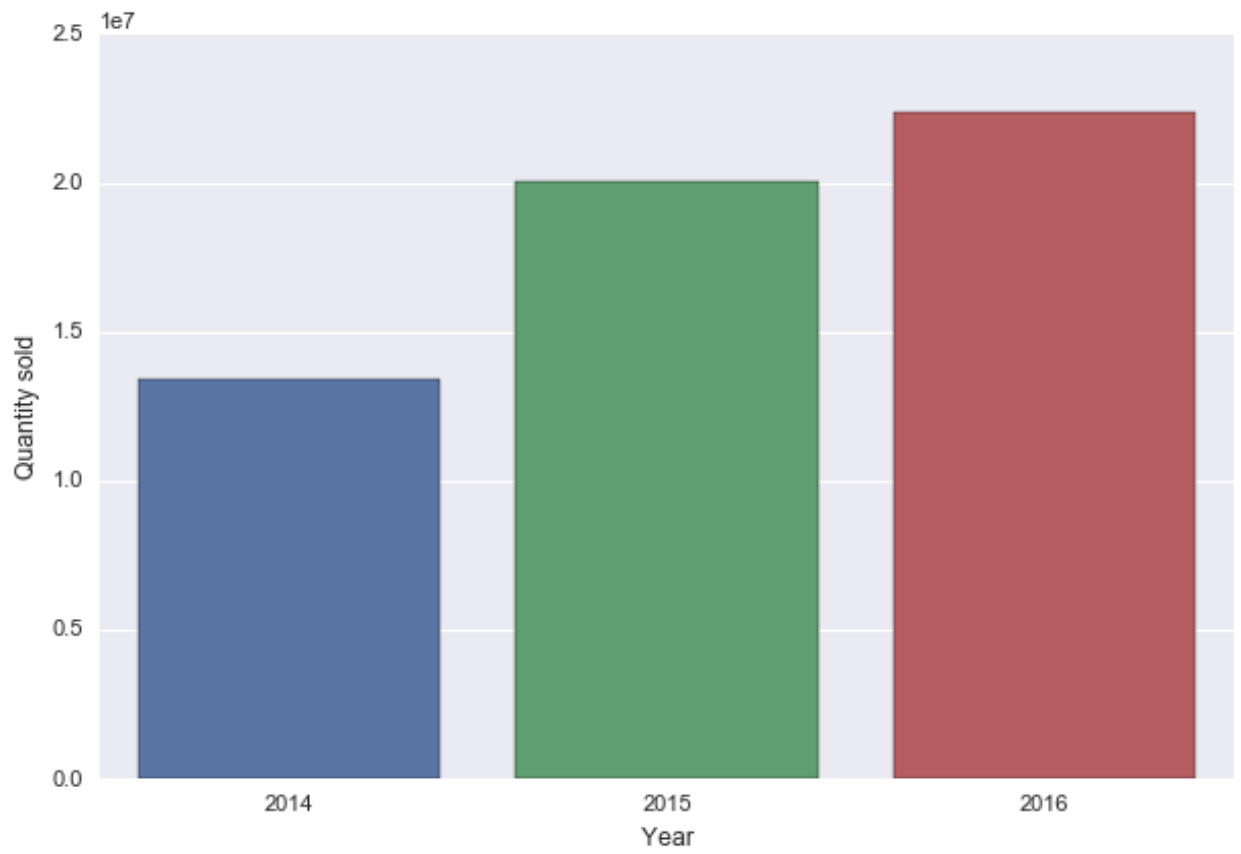
Instead of a column Date with datetime and a column Year-Quarter we want 4 columns

- day
- month
- quarter
- year

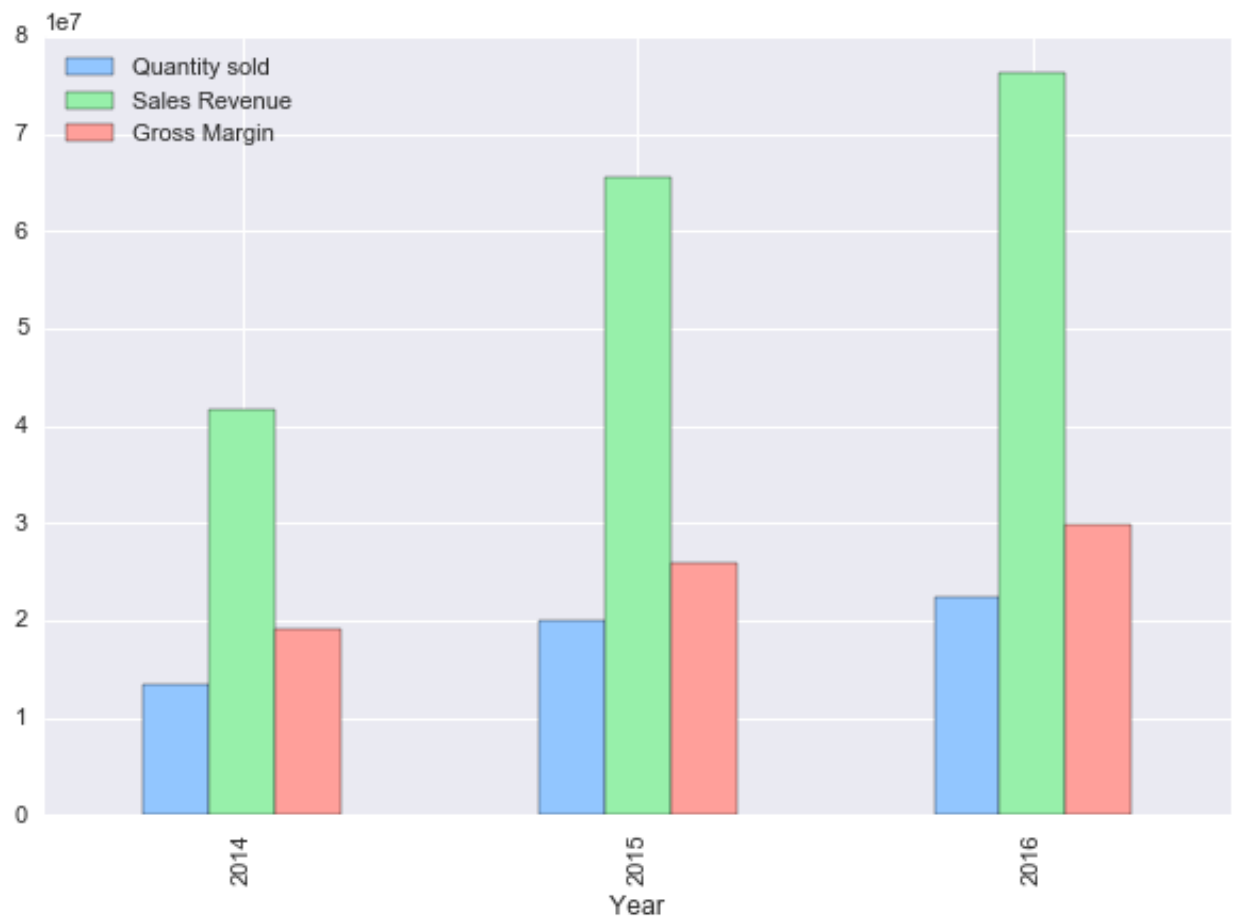
check columns data types and do what need to have have integer/float instead of object when relevant/if needed

First analytics

1. display the sales per year

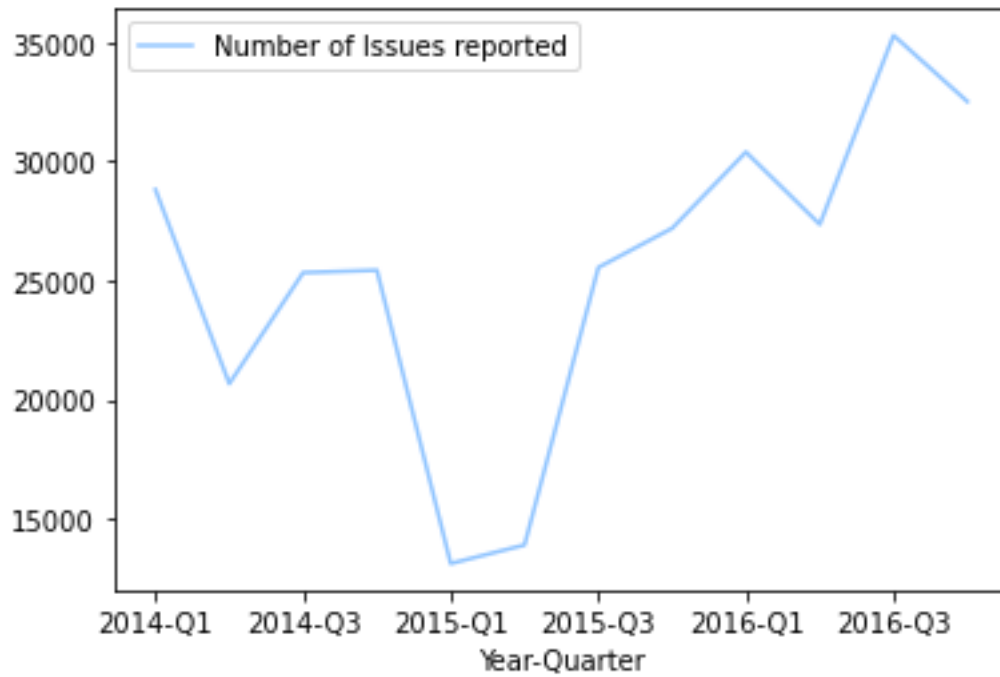


- display the sales per year and then the sales per quarter and year as a bar



- experiment with different matplotlib styles

4. Display the number of issues reported as a line chart



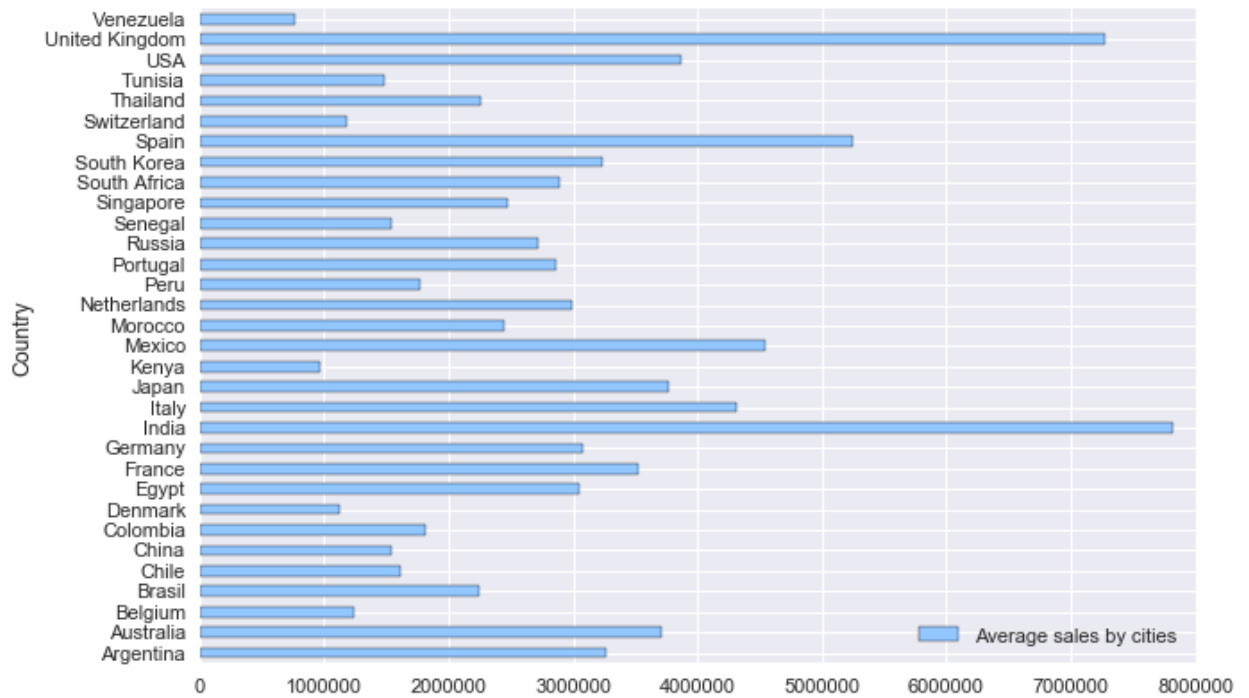
Does the sum of customer satisfaction mean something? ¶

1. Aggregate dataframe with sum and mean
2. Display the mean of issue reported by month

Breaking down data frame by cities

1. Display the number of cities by countries

2. Displays the average sales by cities using pandas.plot



3. experiment with the different styles from plt.style.available:

NULL behavior

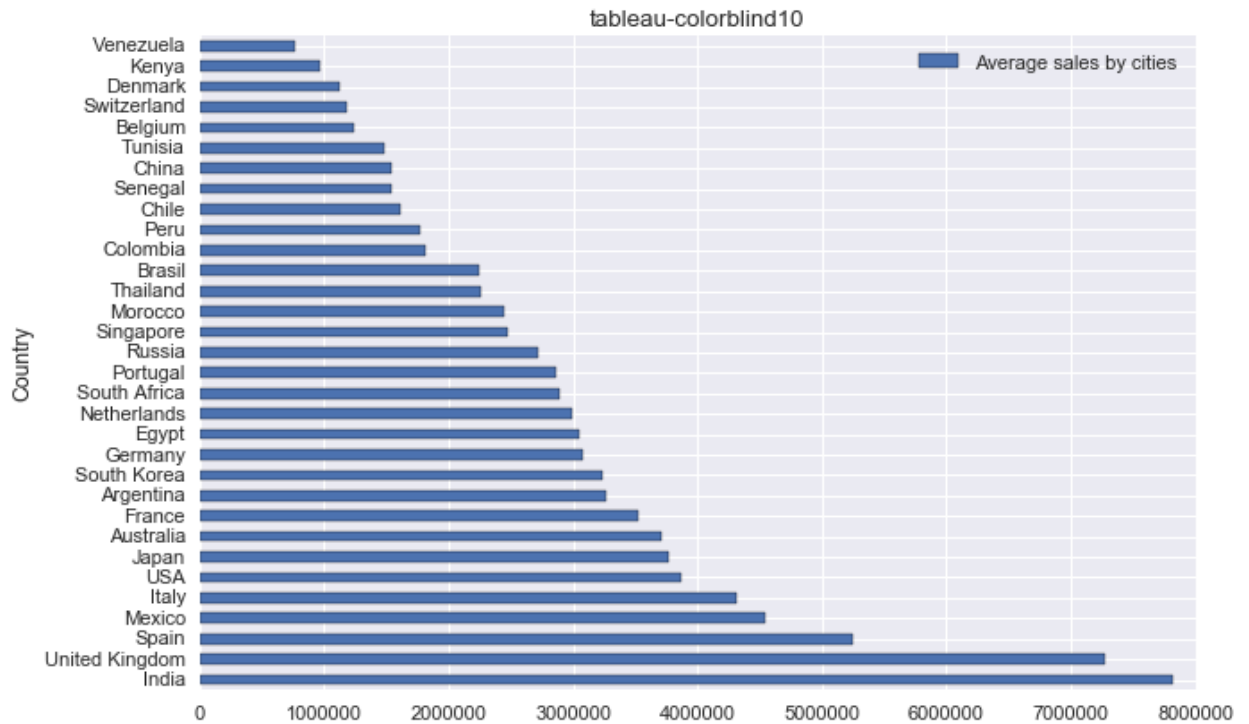
Create the following toy dataframe

```
df_null = pd.DataFrame({"Person":  
    ["John", "Myla", "Lewis", "John", "Myla"],  
    "Age": [24., np.nan, 21., 33, 26],  
    "Single": [False, True, True, True, False]})
```

1. How np.nan == np.nan evaluates?
2. How can we count the NaN ?

Orders and ranks

1. Display the average sales by cities in ascending order



2. Rank countries per number of cities. Experiment with the various ranking method {'average', 'min', 'max', 'first', 'dense'}
3. How does rank behave wrt Null ? experiment with previous df_null dataset
4. What are the top best countries?

Lookup

1. Get data about country population for example
from: <https://data.worldbank.org/indicator/SP.POP.TOTL> and add a column to previous dataset (df8) with country population
2. Check if it work for Brasil