Data Wrangling with Pandas

Using the dataset stored in **bit.ly/PGGM_dataset.**

1. Define a function that takes the dataset and creates a DataFrame which columns are: ‘Variable’, ‘Number of Nulls’, ‘Percentage of Nulls’.

2. Create a function that replaces all the occurrences of the string “Inc.” for the string “B.V.” in any given vector.

3. Include a new variable in the dataset that corresponds the quarter of the year.

4. Create a subset where the columns are the Quarter of the year, and Industry Group the Rows, agregating by ROA. Is there any pattern? Export the subset in csv.

5. Create a subset where the original dataset is sorted by Market Cap USD and determine which month has the highest value in the dataset.

6. Which of the overall Industry Group has the lowest 5 years sales growth in average

7. Which of the sectors has the highest 5 years sales growth in 2017

8. Create a subset of the portfolio which includes all the values where Universe Returns indexes higher than the 75 percentile.

9. Create a function that can label any given vector in a dicotomic way: above the median and below the median.

10. Create an excel file which contains 3 different sheets respectively for 2016, 2017, 2018. Each including the same combination of the portfolio. The portfolio should be equally represented by all sectors.

11. Create a story in a notebook that reports the Data Exploration of the dataset.