DATA ANALYSIS

Computer

Statistics

Business

DATA

INFORMATION

KNOWLEDGE

DATA SCIENCE

ANALYTICS

BUSINESS TASK

Questioning the Stakeholder

DATA DRIVEN - DECISIONS MAKING

OPTIMIZING TIME AND COUST

INCRESE THE ROI

PROBLEM SOLVING

Questioning - Querying the Data

DATA

HOW

ANALYSE

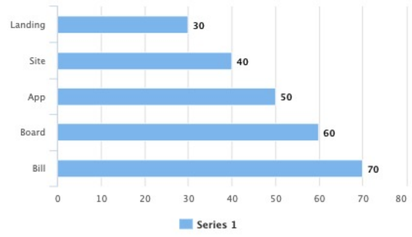
SQL/ PYTHON /R

TABLEAU/BI

PREPARE
Cleaning
Excluding outliers
Query for information
Vizualize

DASHBOARDS/ PRESENTATIONS

REPORT



Aurea Montero

DATA ANALYSIS PROJECT

WITH PYTHON

THE WEATHER DATASET

Este projeto foi desenvolvido utilizando o Weather Dataset
(Dataset sobre condições climáticas de determinados locais).
O Dataset foi disponibilizado pelo canal Data Science Lovers como um arquivo CSV (valores separados por vírgula) e analisado utilizando Python, através da Interface Jupyter Notebook e utilizando o DataFrame Pandas.

Aurea Montero

Questões:

Q. 1) Find all the unique 'Wind Speed' values in the data.
Q. 2) Find the number of times when the 'Weather is exactly Clear'.
Q. 3) Find the number of times when the 'Wind Speed was exactly 4 km/h'.
Q. 4) Find out all the Null Values in the data.
Q. 5) Rename the column name 'Weather' of the DataFrame to 'Weather Condition'.
Q. 6) What is the mean 'Visibility' ?
Q. 7) What is the Standard Deviation of 'Pressure' in this data?
Q. 8) What is the Variance of 'Relative Humidity' in this data ?
Q. 9) Find all instances when 'Snow' was recorded.
Q. 10) Find all instances when 'Wind Speed is above 24' and 'Visibility is 25'.

Comandos:

* head() - It shows the first N rows in the data (by default, N=5)
* shape - It shows the total no. of rows and no. of columns of the dataframe
* index - This attribute provides the index of the dataframe
* columns - It shows the name of each column
* dtypes - It shows the data-type of each column
* unique() - In a column, it shows all the unique values. It can be applied on a single column only, not on the whole dataframe.
* nunique() - It shows the total no. of unique values in each column. It can be applied on a single column as well as on the whole dataframe.
* count - It shows the total no. of non-null values in each column. It can be applied on a single column as well as on the whole dataframe.
* value_counts - In a column, it shows all the unique values with their count. It can be applied on a single column only.
* info() - Provides basic information about the dataframe.

Aurea Montero