Portfolio assignment 102

30 min: Look back at all the portfolio assignments you have done. Create a short report in which you use the portfolio assignments as examples to explain the tasks and process of a Data Scientist.

What are the tasks of data scientist?

In the given assignments we learn a bit of its activities. As a data scientist you are going to learn how to visualize data in a better way, and that is what in Assignments 3 to 6 explain us. If someone asks: What are the Locations where the life expectancy is at its highest or at its lowest, you need to make a visualization where you can get that information, use algorithms.

In Assignments 7 we discover how to calculate the frequency of data, how the distribution works, now we can realize that many rows have common attributes and can be classified in a certain group. That's important for the part of analyzing the data, getting enough information is a primary task of a data-scientist.

With those assignments we now can move from a visualization in a simple boring table to visualizing data in plots.

In Assignments 8 we can now calculate the confidence interval of data, we can now calculate what the interval in a certain attributes, and defining the confidence level, we can say with how much certainty that interval is. Knowing how trustworthy your data is is extremely important.

Assignment 9 & 10 teach us how to calculate the correlation between attributes. We can now see in a scatter plot how attributes are related with each other and how strong they are correlated. This remains in the part of analyzing the data, finding connections between attributes.

With Assignments 11 - 14 we now can calculate two attributes Categorical or Numerical, how they interact with each other. A example: What type of pokemon has the most points in Attack, or Defense etc. We can also calculate how strongly correlated they are, and the ratios between the categories, what also an important is for a data scientist.

Ans finally in Assignments 15 - 19 we learn one of the most interesting aspects from Data Science: Predicting data. We learn how to make a decision tree and regressive decision tree to analyze the attribute that might have influence in a dataset and also predict the values of the dataset. Also getting back at the aspect of how trustworthy the prediction is, we also learn to calculate that in those assignments.