Emre Kemerci / 311802017 / BDA 503 Fall 2018 Final

Part 1

Q1

I noted my insights for both languages below:

Python + I have not enough experience but according to what I heard commonly, Python is more capable in modelling & machine learning + I think syntax of Python is easier to read and write comparing the R. R has disadvantage especially due to paranthesis + As far as I know, Python is preferred more among data analists/scientists

R + Better for charming visualisation + Better reporting capabilities and customization opportunities with Markdown comparing Jupyter notebook + R Studio looks more friendly for new starters comparing PyCharm + My insights are that R have variety of libraries, on the other hand coming with consistency problems of the libraries

Opinions stated in the dataschool website commonly overlap with my thoughts but at the end of the day, both languages have advantages and disadavntages. So I think that a data scientist should have capabilities to an extend in both languages and should prefer the one better satisfing the need.

Q2

First, the objective - question to be identified - should be clarified. Then I determine what I need to answer the question and continue with understanding what I have on hand. I design the execution of analysis considering the outputs from understanding the data on hand and execute accordingly to reach a conclusion answering the initial objective.

Following the example in the question; I can measure the impact of each project through finding the correlations of overall wellfare with each project. Delta of overall welfare per unit of improvement project. Most probably marginal benefit to overall welfare will decrease and considering the marginal benefit curve can be used to allocate funds efficiently. If I found a clear picture showing which projects should be focused for each welfare levels, info becomes valuable for me.

I prefer to present what the data says but I must explain the results of the analysis in a meaningful way.

Q3