Final Project #6

Comprehensive Data Analytics Project

Project Description

This is the final project where you will be able to enjoy the fruits of your hard labour and utilise all the skills you have acquired and put them to work into creating a marvellous final delivery

Project Goal

- Deliver a full end to end data analysis project from idea to presentation where you autonomously leverage everything you have learnt so far

Final project main pointers!

Find a problem which intrigues you! A problem where you feel excited to be able to provide a solution or part(s) of a solution. You should **clearly** define your problem & make sure that the data set you choose can bring insights to the problem you want to solve. We can not put enough emphasis on the point of **CLEARLY!** defining your problem. Make sure that your scope is reasonable given the timeframe. Try to think one step ahead, what are some likely obstacles and what could a possible solution to the problem have for an impact?

Note, you can here choose to do something of your choice related to data if you want to deep dive into something specific/have an idea that you want to test out. In that case you need to check with the instructors that your idea is possible to execute. Otherwise we recommend you to follow the proposed workflow below and do a full Data Analysis/Machine Learning project since this covers all that we learnt during the BC and also looks fantastic when put up on GitHub.

Lastly, push yourselves to deliver results that you can take pride in. Do not allow yourself to get stuck for too long on details, continuously zoom out and look at the bigger picture to enable yourself to prioritise wisely. Use the teachers for discussions & feedback to stay on track.

Project Requirements and Instructions day to day

Below you will see a layout of a possible workflow for your final project. View this as a helpful overview rather than instructions that you must follow.

Also, remember that data analysis is iterative. So, from time to time you may need to step back to a previous phase and iterate.

Day 1 - Define, Plan, get & clean your data

- Idea generation and planning
- Clearly defining your problem
- Get your data
- Clean your data (if necessary)

For this project you will likely spend more time than usual searching for the data

Day 2 – Finish cleaning + Exploratory Data Analysis

- Finish up your data cleaning
- EDA
- Cluster analysis

You will explore, analyse, and visualize the data applying the variety of techniques you've learned throughout the bootcamp. Here it is time to really think about the building blocks of the story that you (and the data!) want to tell to communicate the insights from your analysis.

Day 3 – Finish EDA + Start building the your ML model

- Choice of metric
- Feature selection
- Evaluate and compare the different models

Focus on choice of metric. You might want to evaluate the model based on more than one metric. Either way, you should have a clear explanation for your choice of metric.

Day 4 – Machine Learning second iteration

- Choose the best model (you can keep more than one model until the end also)
- Hyperparameter tune the best model(s)
- Analyze the results of your predictions on test data
- Organise the ML steps into a full ML pipeline
- Start preparing for presentation

Day 5 – Structure you code + Final analysis & Presentation

- Organize your code into a neat and structured workflow/pipeline
  - Remove unnecessary code
  - Add comments to your code
- Connect the dots in your story
- Create your presentation
- Put your project on Github

Time is limited so prioritize wisely! How did the story unfold? From framing the problem to the EDA to the Analysis of the models predictions? And what happens after the model has made its predictions on new data? How & why is your analysis & code able to provide value? What are the most important insights and how should they best be communicated?

Presentation

The presentation should take **max 10 minutes**

- Title of the project + Student name
- Students' choice

Schedule

The presentations will take place on Friday after lunch

Good Luck!!