Evaluation and Capstone Project

1/10/2022

Data Science Academy: Season II Week 8
Instructors: Oren Livne, Jiangang Hao, & Qiwei He

Evaluation and Certification

Data science is a highly hands-on and domain dependent discipline. As such, our evaluation will center around whether you can do the actual work you are likely to encounter in your future work at ETS. Once you pass the evaluation (the baseline is everyone should pass), a certificate from ETS will be granted. Specifically, the evaluation includes the following:

- *Attendance and completion of homework/projects.*
- *A data science knowledge test (multiple choice/constructed response, ~20 items, open book)*
- *A capstone project that requires you to parse some complex data files, extract designated features, create visualizations and dashboards. We will provide you the raw data from some real programs after de-identification and other security-related adjustments.*
- *A 10 min presentation of your capstone project.*

We believe that people who accomplish the above 1 – 4 should meet the learning goals of this course. Completing these learning goals will position you well to start addressing actual data science challenges you will encounter at your ETS day-to-day work and continue to improve your skills with your practical work.

https://etsorg1.sharepoint.com/sites/MSGrp_TrainingInitiativeFoundationsinAIDataScienceandLearnin/SitePages/Data-Science-Academy.aspx

Homework and Test

- If you (or your manager) ask for the certificate, you must submit the homework to us (e.g., you have organized them on the github per the week 7 homework).
- The data science knowledge test (open book) is a MS survey form. You have the whole week to complete it. The due date is 1/17/2022.

Capstone Project

- Develop a streamlit dashboard based on the data we provided.
- The codes should be hosted on GitHub
- The dashboard will be deployed on the streamlit cloud server
  - How do to this?
  - https://docs.streamlit.io/streamlit-cloud/get-started/deploy-an-app
- Requirements:
  - Include your name and group/center/department/division on the dashboard
  - You can work in groups, but everyone needs to create one dashboard
  - Please use your imagination to show your insights from the data and tell a cogent story to your audience
- Help
  - We will hold office hours 1/24/2022 (next next Monday) this time. Note that we won't meet next Monday due to the MLK holiday.
- Final presentation will be at our meeting on 1/31/2022. Before that, please send the link to your dashboard to us.
- Here is an oversimplified example, you should do better than this

  https://share.streamlit.io/jianganghao/capstone_dsa2021/main/capstone_das_hao.py

Technical Docs

- https://docs.streamlit.io/library/api-reference
- https://www.markdownguide.org/cheat-sheet/
- https://plotly.com/python/plotly-express/#highlevel-features

Data for Capstone Project

- data_capstone_das2021_2022.csv
- The data is a table that records response time, scores to questions, and some demographic info.
- Each row is a participant
- Column meaning

  - rt_gs_i: the response time to ith item in seconds. Note that there is no response time for the first item.
  - rt_total: total reponse time by the participant
  - gs_i: score of the ith item.
  - sum_score: total score of the participant
  - gender
  - home_computer: whether the participant has computer at home
  - state: the state the participant is living
  - age

Have Fun!