Berkeley Engineering | BerkeleyHaas

PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Office Hour #3 with Matilde D'Amelio

March 24, 2022 at 10 pm UTC

MODULE 2

Q&A Discussion Forum and Codio Activities

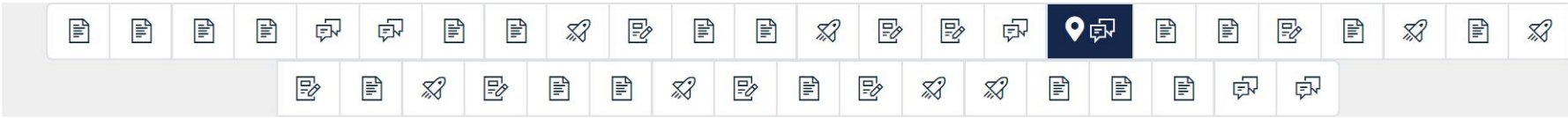
MODULE 3 OVERVIEW: Introduction to Data Analysis

- ➔ Data Science Lifecycle
- ➔ Visualisation Libraries (Pandas, Ploty and Seaborn)
- ➔ Data Sorting and Aggregation
- ➔ Data Indexing
- ➔ Data Filtering

TRY-IT ACTIVITY 3.1

Try-It Activity 3.1: Creating Data Visualizations - Section B

Mar 23 at 8:58am



Learning Outcomes Addressed:

3. Generate visualizations using pandas, Seaborn, and Plotly
4. Customize plots using externally sourced documentation

Building on the previous Codio activity, please select a dataset that is of interest to you. This could be a dataset provided in this program or one that you have sourced from the Internet. Utilizing the selected dataset, please create a visualization that best describes the data.

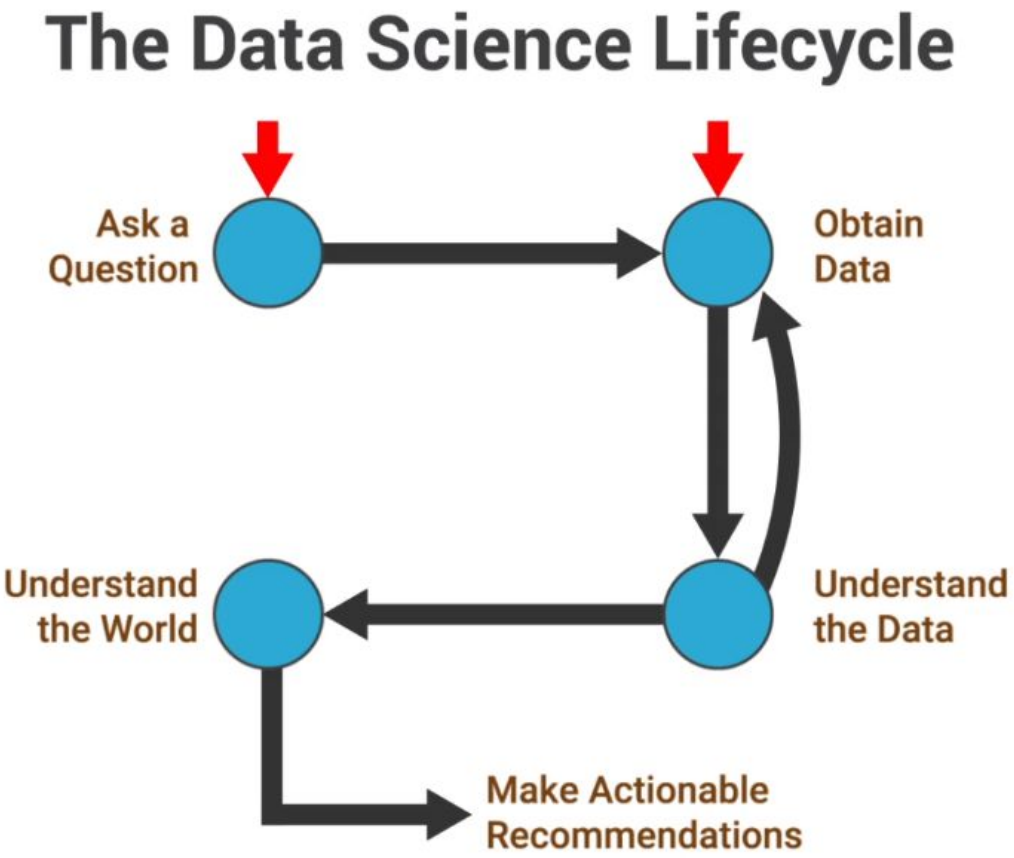
Post the following to the discussion board:

- Attach your visualization
- Explain the method you used to create the visualization
- Share what library you used
- Describe the results you found and what they tell you about the dataset you chose

Your visualization does not have to be something that you spend hours on making it look 'just right', as the rubric is purposefully vague when it comes to style, layout, etc. However, you may find the following resources helpful (and sometimes funny) in guiding your creation of your visualization

- 8 Ways to Turn Good Data into Great Visualizations: GoodData presents 8 principles of good data visualization
- Visualizations That Really Work: Harvard Business Review presents a framework for creating visualizations
- Why Is This Chart Bad? The Ultimate Guide to Data Visualization Evaluation using GoDVE: This resource both analyzes several visualizations, points out the flaws with each visualization, and presents a framework for developing visualizations

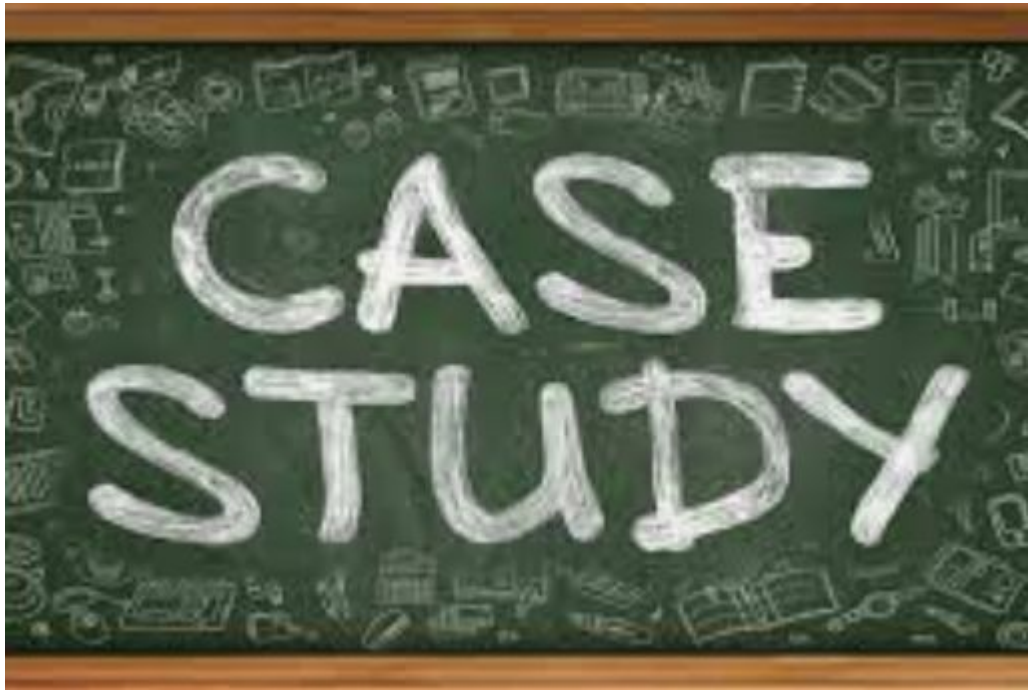
MODULE 3 OVERVIEW: Introduction to Data Analysis



CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

CRISP-DM CASE STUDY

Case Study: https://rpubs.com/Argaadya/crispr_dm



DISCUSSION

Did you work on any Data Science Lifecycle?

If yes….

- What was the **Business Problem** you were solving?
- Which **Problems/Challenges** did you face?

QUESTIONS?