🔒 ga-students / **DSI_LDN_1_LESSON_NOTES** Private   👁 Watch ▾ 46   ★ Star 0   ⑂ Fork 26

<> Code   ⚠ Issues 0   ⑂ Pull requests 0   📋 Projects 0   📖 Wiki   ∿ Pulse   Ⅲ Graphs

Branch: master ▾   **DSI_LDN_1_LESSON_NOTES** / projects / capstone / **part-02** /   Create new file   Upload files   Find file   History

👤 **juanginzo** added project 7 and capstone 2   Latest commit b5243d1 23 hours ago

..

| 📄 part-02-rubric.md | added project 7 and capstone 2 | 23 hours ago |
| 📄 readme.md | added project 7 and capstone 2 | 23 hours ago |

📖 **readme.md**

GA Part 2: Problem Statement + EDA

Overview

This will be an overview of your approach with a well-articulated summary that includes your problem statement, outlines your proposed methods and models, defines any risks & assumptions, and includes any revisions from your initial goals & criteria, as needed.

In addition, you'll want to create a local database for your data, import and clean your data. You'll want to describe and create a data dictionary for your dataset(s), and perform your initial exploratory data analysis.

Goal: A summary notebook outlining your project's goals, methods, models, and EDA.

Requirements

Your work must:

- Articulate "Specific aim"
- Outline proposed methods and models
- Define risks & assumptions
- Revise initial goals & success criteria, as needed
- Create local database
- Describe data cleaning/munging techniques
- Create a data dictionary

- Perform & summarize EDA

- *Bonus:*

  ○ Explain how you intend to evaluate your results. What tuning metric and evaluation approaches do you intend to use?
  ○ Identify 1-2 additional datasets that may help you triangulate your findings. How might these relate to your data?
  ○ Create a blog post of at least 500 words (and 1-2 graphics!) that describes your assumptions and processes for EDA. Link to it in your Jupyter notebook.

Necessary Deliverables / Submission

- Materials must be submitted in a clearly labeled Jupyter notebook, including:

- Markdown writeups, code, and visualizations
- Materials must be submitted via a Github PR to the instructor's repo.
- Materials must be submitted by the end of Week 8.

Suggested Ways to Get Started

- Don't hesitate to write throwaway code to solve short term problems
- Read the docs for whatever technologies you use. Most of the time, there is a tutorial that you can follow, but not always, and learning to read documentation is crucial to your success!
- Write pseudocode before you write actual code. Thinking through the logic of something helps.
- Document **everything**.

Useful Resources

- Exploratory Data Analysis
- Best practices for data documentation

Project Feedback + Evaluation

Attached here is a complete rubric for this project.

Your instructors will score each of your technical requirements using the scale below:

Score	Expectations
0	*Incomplete.*
1	*Does not meet expectations.*
2	*Meets expectations, good job!*
3	*Exceeds expectations, you wonderful creature, you!*

This will serve as a helpful overall gauge of whether you met the project goals, but **the more important scores are the individual ones** above, which can help you identify where to focus your efforts for the next project!