CONDENSED GUIDELINES

METAGOALS:

- Identify problems in a situation and goals in analyzing a related data set
- Create and explain reasoning for choosing specific approach to solve problems and acheive goals
- Apply selected analysis methods to analyze dataset
- Interpret results of analysis
- Give context and broader meaning to results
    - Provide scientific and social context
    - Acknowledge and address any issues related to privacy and ethics

FINAL PROJECT GUIDELINES:

- Final project notebook should include all code used in the project
    - Cleaning
    - Visualization
    - Analysis
- Submitted notebook should have code pre-evaluated and outputs (e.g. graphs) present
- Notebook should be self-contained so entire project can be evaluated from notebook alone

SECTIONS:

Overview:

- 3-4 sentences
- Summarize topic and project

Research Question:

- 1-2 sentences
- Precisely describe research question

Background and Prior Work:

- 2-3 paragraphs
- ≥2 references
- Describe currently known information about the topic
    - Provide references for similar related projects (publications, blogs, company sites etc.)
        - Explain their findings
- Answer the following questions
    - Why do each of you find the question you chose interesting?
    - What background info led to hypothesis?
    - Why is this topic important?
    - What work has already been done on topic?
        - What do we know already?

Hypothesis:

- Main hypothesis/predictions
- Explain why

Dataset(s):

- Describe data used to answer your question
    - Number of observations (n=__)
    - Content/features
    - Etc.
- Must use at least one dataset containing ≥1000 observations
- If using multiple datasets
    - Briefly explain how you will combine them together
- List sources of dataset(s)

Setup:

- Include analysis packages required to run data collection/analysis/visualization code

Data Cleaning:

- Describe all methods used to clean the data
- Describe steps taken to clean data before analyzing
- Answer the following questions:
    - How ‘clean’ is the data?

        - See guidelines from lecture
    - What steps were required to get data into useable format?
    - What pre-processing steps were required?
        - E.g. checking data distributions for normalcy or transformations

Data Analysis and Results:

- Include markdown text and code walking graders through the following:
    - Exploratory data analysis (EDA)
        - Describe variable distributions
        - Describe any outliers
        - Describe significant relationships between variables
    - Analysis
        - Describe analysis approaches
            - Justify why
        - Describe significant results of analysis
        - Discuss interpretations of results
    - Data visualization
        - ≥3 data visualizations throughout data analysis and results section
        - For each visualization:
            - Label all axes
            - Don't include unnecessary details that would crowd or add clutter
            - Provide an interpretation
            - Discuss what should be learned

Ethics and Privacy:

- 1-2 paragraphs
- Address any ethical/privacy concerns regarding question/dataset/results/analyses
    - See relevant lecture and/or Deon's Ethics Checklist for more details
    - Answer the following questions:
        - permission to use this data?
            - For this purpose?
        - Any privacy concerns regarding the data you used?
            - Any terms of use you needed to comply with?
        - Any potential biases that might make analysis/results unequitable?
            - In terms of ppl it samples
            - In terms of how it was collected
                - E.g. excluding certain populations, reflecting human biases
        - General/other data privacy and equitable impact issues
    - Discuss how you addressed/minimized/resolved the identified issues

Conclusion and Discussion:

- Recapitulate/summarize data characteristics and question
- Redescribe/summarize most important aspects of analysis
- Summarize results and conclusion
- Discuss challenges/limitations
- Discuss broader implications/connections/impact on society

GRADING:

- Final project worth 40% overall
    - 10% project proposal; 2% check in; 3% project survey; 25% final project notebook
- Address all rubric sections using cell markdown for textual descriptions
    - Overview, Question, & Background; **10%**
    - Data Description; **10%**
    - Data Cleaning/Processing; **10%**
    - Data Visualization; **15%**
    - Data Analysis & Results; **25%**
    - Ethics & Privacy; **15%**
    - Conclusion & Discussion; **15%**