**Group Member Information:**

Please read the COGS 9 team policies to best understand how to approach group work and to understand what the expectations are of you in COGS 9.

|  |  |  |
| --- | --- | --- |
| **First Name** | **Last Name** | **PID** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Question**

Clearly state the specific data science question you're interested in answering. This question can be the same as what you submitted for your project proposal. Alternatively, you can edit your original question or change your topic completely. (2 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Hypothesis**

Write down your group’s hypothesis to your question. Provide justification how you came to this hypothesis. (What background information or instinct led you to that hypothesis?). You should incorporate the feedback you received on your proposal. (2 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Background Information**

Include a few paragraphs of background research and information on your topic. This should include at least 2 citations to work from others. Including hyperlinks to reputable sources are fine. (3 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Data**

Include a description of the perfect dataset you would need to answer this question. How many observations would you need? What variables would you collect? Explain the perfect dataset that you would want to answer this question.

Then, look online for available datasets. Find a dataset that could be used to answer this question. Describe how many observations are included and what variables have been collected. Discuss the dataset’s limitations and how it differs from your ideal dataset. If you collected your own data, explain what information you collected, from whom you collected it, and a link to the data. (2 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Ethical Considerations**

Read the data science ethics checklist from lecture. Then, discuss what ethical considerations must be made when answering your specific data science question. Brainstorm and explain how you would address these considerations for each of the following categories in your specific project: Team Bias, Sampling Bias, Data Bias, Consent, Data Privacy / Ownership, Algorithmic Bias / Discrimination, Transparency, Unintended Consequences, Continued Monitoring / Accountability. Feel free to write about additional ethical considerations you would make that aren't included on the checklist. Note that data privacy is NOT the only ethical consideration for a data science project. It is a piece, but there is a lot more that has to be considered. (3 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Analysis Proposal**

Here, you will propose how you would use and analyze data to answer your question(s) of interest. You are neither expected nor encouraged to carry out the analyses to answer your question(s). You will describe, in detail, what you would need to do to prepare your dataset for analysis (data wrangling) and what type of analysis you would do to answer your question(s). Explain how you would interpret the results from this analysis. We are looking for the correct conceptual understanding and application of ideas discussed in class, not specific and technical implementations. For example, if you are applying machine learning to some categorical data, it’s important to specify whether you will be performing regression or classification. If you are unsure about the details of anything above, ask on Piazza, come to office hours, and/or do further research on your own (Stack Exchange, Google, Wikipedia, etc.).

Specifically, you are required to incorporate *at least four different methods*, exploring ideas from a combination of:

* Data Collection (web scraping, APIs, etc.)
* Data Wrangling
* Descriptive & Exploratory Data Analysis (summary stats, correlation, etc.)
* Data Visualization
* Statistical Analysis (Inference, A/B testing, etc.)
* Predictive Analysis (machine learning, classification, regression, etc.)
* Text Analysis (Sentiment Analysis, TF-IDF, etc.)
* Geospatial Analysis (choropleth maps, geospatial statistics, etc.)

(15 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Discussion**

How would you interpret the results of your proposed analysis? What are the limitations, pitfalls, and potential confounds of your methods, or biases in your data sources (e.g., how does the selection of the sources of your crowds affect your outcomes?)? How would you set out to address them? In addition, outline how you would address any societal and/or ethical implications of your proposed project discussed in your Ethical Considerations section. (10 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE

**Group Participation**

Include one paragraph briefly outlining the contribution of each group member throughout the quarter while working on this project. Each of you must also fill out the survey (link provided toward the end of the quarter) about individual and group participation. (3 pts)

ADD YOUR GROUP’S TEXT HERE AND DELETE INSTRUCTIONS ABOVE