**Presentation Requirements**

* Questions that you found interesting and what motivated you to answer them
* Where and how you found the data you used to answer these questions
* The data exploration and cleanup process (accompanied by your Jupyter notebook)
* The analysis process (accompanied by your Jupyter notebook)
* Your Conclusion, including a numerical summary and visualization of the summary
* The implications of your findings: what do your findings mean?

**Development Requirements**

* Use Pandas to clean and format your dataset or datasets.
* Create a Jupyter notebook describing the data exploration and cleanup process.
* Create a Jupyter notebook illustrating the final data analysis.
* Use Matplotlib to create 6 to 8 visualizations of your data
* (ideally, at least 2 visualizations per “question” that you ask your data).
* Save PNG images of your visualizations to distribute to the class and instructional team—and for
* inclusion in your presentation.
* Create a write-up summarizing your major findings. This should include a heading for each “question”
* that you asked your data as well as a short description of your findings and any relevant plots.
* Bonus Use at least one API—if you can find one with data pertinent to your primary research questions.

**Suggestions for Data Sources**

* They are sufficiently large.
* They have a consistent format.
* They, ideally, contain more data than needed.
* They are well documented.

**How to approach it**

**Start with Questions:**

* Clearly define the questions you find interesting, such as:
* What factors are correlated with suicide rates?
* How has mental health awareness evolved over time?
* Are there demographic patterns in mental health conditions?
* Provide a numerical summary of key findings (e.g., correlation coefficients, means, etc.).
* Create visualizations (e.g., bar charts, line graphs) to summarize important trends.
* Offer insights into patterns or relationships discovered during analysis.
* Discuss the meaning and significance of your findings.
* Consider the practical implications for mental health policy, intervention, or awareness.
* Highlight any limitations of your analysis and suggest areas for future research.

**Define Your Research Question or Hypothesis:** Clearly articulate the research question or hypothesis you want to explore. For example, you might want to investigate the factors contributing to mental health issues or identify patterns leading to suicide.

**Data Collection:** Gather relevant data from reliable sources. This can include surveys, medical records, social media data, government reports, or other relevant datasets. Ensure that the data is representative of the population you are studying.

**Ethical Considerations:** Ensure that your research adheres to ethical guidelines, especially when dealing with sensitive topics such as mental health and suicide. Anonymize and secure the data to protect the privacy of individuals.

**Data Cleaning:** Clean the data to handle missing values, outliers, and inconsistencies. Ensure the data is in a format suitable for analysis. This may involve standardizing variable names, converting data types, and addressing any data quality issues.

**Data Exploration:** Conduct exploratory data analysis (EDA) to understand the characteristics of the data. Use descriptive statistics, visualizations, and summary measures to gain insights into the distribution of variables and potential patterns.

**Hypothesis Testing:** If you have specific hypotheses, use statistical tests to assess their significance. This could involve t-tests, chi-square tests, or more advanced methods depending on the nature of your data.

**Correlation Analysis:** Explore relationships between variables. Use correlation analysis to identify potential associations between mental health factors and other variables of interest.

**Predictive Modeling:** If applicable, consider building predictive models. Machine learning techniques can help predict mental health outcomes or identify risk factors for suicide. Ensure proper validation and evaluation of the models.

**Data Visualization:** Present your findings visually through charts, graphs, and other visualization tools. This helps make your results more accessible to a wider audience.

**Interpretation and Reporting:** Interpret your results in the context of your research question. Discuss the implications of your findings and their relevance to mental health and suicide prevention. Clearly communicate your results in a comprehensive report.

**Peer Review:** If possible, seek peer review from experts in the field to ensure the validity and reliability of your analysis.

Where to start

**CDC WONDER (Wide-ranging Online Data for Epidemiologic Research):** The CDC provides access to a wide range of public health data, including suicide rates, mental health indicators, and related statistics. You can use the CDC WONDER API to retrieve data directly into your Python environment.

**Kaggle Datasets:** Kaggle is a platform that hosts a variety of datasets. You can find datasets related to mental health, suicide rates, and associated factors. The "Suicide Rates Overview 1985 to 2016" dataset is one example that you can explore. <https://www.kaggle.com>

**World Health Organization (WHO) Data:** The WHO provides global health data, including information on mental health and suicide. You can explore their datasets or use the Global Health Observatory API to retrieve specific data. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>

CDC Wonder: <https://wonder.cdc.gov>

**Social Media Data:** Analyzing social media data can provide insights into mental health discussions. Platforms like Twitter and Reddit offer APIs that you can use to collect relevant data. Keep in mind the ethical considerations and privacy issues associated with analyzing social media data.

**National Institute of Mental Health (NIMH) Data:** The NIMH may offer datasets related to mental health research. Check their official website or contact them directly for access to relevant data.

<https://www.nimh.nih.gov/research/priority-research-areas/suicide-research>

<https://www.nimh.nih.gov/search-nimh?query=NIMH+database>

National Library of Medicine

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8699163/>

**Behavioral Risk Factor Surveillance System (BRFSS):** The BRFSS is a survey conducted by the Centers for Disease Control and Prevention (CDC) that collects data on health-related risk behaviors. It includes information on mental health and could be a valuable resource for your analysis. <https://www.cdc.gov/brfss/data_tools.htm>

<https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016>

<https://www.kaggle.com/code/lmorgan95/r-suicide-rates-in-depth-stats-insights>

**Thoughts**

**should we venture into how education can be done using the data to watch for signs of suicide and may be help with education and prevention**

**which group is more at risk**