Guidelines For Lab Reports

- Treat me as a client who is not a data analysis professional, but eagerly interested in your conclusions and how you got to same.
    - Do not try to snow me with a technical discussion which is not to the point of the question and your conclusions!
- Spend time reviewing the data features, what they mean and how they relate to one another
- Think about the sample and what are its limitations compared to perhaps a target population
- Clearly argue the connection between the filtered sub-sample you select, the features you use (and the related ones you did not select) and the question being asked.
- Operationalize in Intro. Describe features use and ones donot use and why. How will these feature answer the question being asked. Summarize what is the hypothesis testing circumstance. Treat it as an executive summary up to the conclusion
- Explore the data being used for summarization and shortcomings
    - Missing values
    - Obvious incorrect values
    - Data types (nominal, ordinal metric)
    - Relevant summaries and descriptive statistics for the data type (e.g. do not estimate means for ordinal data)
    - Create "Accounting Table" number samples started with and a row for the count of remove samples and reasons, followed by total
- Use appropriate charts and tables to make points and suggest inquiry directions, pushing the logic of argument forword realizing that these artifacts may not be diagnostic
    - Make all graphs complete and readable with axes, labels, titles, legends and same for all tables with title, column and row labels and aligned rows and columns elements
    - Give the reasons and purpose behind a particular graphic (figure or table)
    - Provide an interpretation of the graphics
    - Do not throw meaningless or illogical graphics into the report (e.g. donot run Q-Q Norm plots on ordinal data)
- Do not give me data or code dumps.
    - If your code computes a quantity that is not explained by the output, explain and label it
- Take a position and discuss with respect to data meeting assumptions
    - Normality
    - Independence
    - Sample size
    - Data type
- Argue the analysis circumstances of one group vs two groups, paired versus unpaired, one sided versus two sided, data scale before picking and performing test
- Use your responses to assumptions, analysis circumstances and decision tree (provided in separate file) to decide type of analysis and test prior to performing the test
    - Write out logic for this decision
- State clearly and in a crisp statistical form (required) , null and alternative hypotheses
    - For example H_0: $\mu=a$, versus select one H_A: $\mu\neq a$, H_A: $\mu<a$, H_A: $\mu>a$
    - Remember null and alternative hypothesis tests are stated on parameters not estimates

    - You should also state null and alternative in terms of the problem domain but this is not to be an alternative for crisp statistical notation
    - Distinguish between μ, σ and θ, do not use μ when testing the median θ
- Perform test(s) and provide interpretation of test results & your decision regarding null and alternative hypotheses
    - Hint NEVER accept null hypothesis, only reject it or fail to reject it
    - Never discuss CLT when variables are ordinal
    - Also interpret test results and effective significance in the language of the domain/problem
    - Run the test and compute the test statistic and P-value ⎫
        - Never change the null hypothesis and run more than one test
    - Draw conclusions ⎫
    - Use practical significance to draw a conclusion about the size and importance of your result, especially when the sample size is large ⎫
    - Use Cohen's d and correlation (r), when there is no more an understandable measure such as the difference between the means ⎫
    - Use proper correlation for ordinal variables
- Be direct and succinct, there is a word count limit, so do not give me a word salad
    - Show your client courtesy, check spelling and apply a modicum of grammar