What is the title of your proposed class project? This should be a short sentence or phrase that generally describes your project topic and/or variable of interest.

Your Answer:

What Factors Contribute to A Top Tier Red Wine? An Analysis

Why did you choose this topic/dataset? Are you going to confirm/refute/replicate/extend a previous study? If so, cite (in APA format) the previous study (academic articles, news reports, etc.) and briefly describe how your study will relate to it. If not, what other justification do you have for wanting to research this topic?

Your Answer:

I chose this topic/dataset because I like red wine. I never fully under why different wine traits contributed to greater quality, thus sparking my interest in this dataset. This project will not be a comparison or extension of any sort to any other study done with this dataset. Further justification for wanting to research this topic would be my interest in fine dining and treats of those sorts such as charcuterie boards as well as my interest in classy affairs.

What is the source of your dataset? Be specific. Also, briefly describe how the data was collected.

Your Answer:

https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009 (Links to an external site.)

The poster of the dataset on Kaggle was able to collect the data from the UCI machine learning repository, and the data is based from a scientific research paper by Cortez et al., 2009.

Link to the paper: https://www.scitepress.org/Papers/2015/55519/55519.pdf (Links to an external site.)

The input data was collected by running different experiments with the data to collect measurements such as pH and alcohol concentration. The output data was collected using wine testers as sensory data.

Describe the population of interest (i.e. what you will be making an inference about with your future analysis). Can your dataset be considered a random sample from this population? Explain your reasoning and be specific.

Your Answer:

The population of interest would be all the red wine products that are produced around the world. The dataset cannot be considered a random sample from this population because they only come from one region of the world: Northwest Portugal. Thus there could be bias in my results due to the dataset not being representative of the red wine population.

good description, but maybe you should narrow your population of interest (remember, this is something you can decide!) so that your sample is more representative.

Describe your outcome variable. How was it measured? What are the units (if appropriate)?

Your Answer:

The outcome variable to be study would be quality of the red wine. It was measured by 3 taste testers and the median value of the 3 is the quality score. There are no appropriate units, but the range of scores for the quality of the wine are 1-10.

What predictor variables are you going to investigate in your analysis? Briefly describe how each one was measured, including units (if appropriate) for numeric variables and possible categories for categorical variables.

Your Answer:

The predictor variables i will investigate are: pH (no units), alcohol concentration ( %vol), fixed acidity (g(tartaric acid)/dm3), and residual sugar (g/dm3). The scientific article relating to this study did not explicitly state how these variable were collected, but it was stated that the values were collected using "physicochemical tests".

State the research questions you are going to attempt to answer with your analysis. Not all predictor variables need their own research question, but include at least two research questions and label them RQ1, RQ2, etc.

Your Answer:

Does an increase in alcohol concentration lead to an increase in the red wine quality, while controlling for pH, fixed acidity, and residual sugar?

Are pH, alcohol concentration, fixed acidity, and residual sugar significantly linearly related to the quality of red wine?