Project Plan - Video Game Analysis

My data explorations project will be analysis of video game sales. More specifically how total game sales relate to other attributes such as critic or user ratings. My ultimate goal is to determine which factor is most correlated to total video game sales and then analyze why.

I've chosen to do an implementation project and I will be focusing on data visualization. I will have a dataset of games that I will be making various charts and graphs to help find trends in game sales. I will be using python with several packages to read in and create the visuals. There are a number of old datasets available, but they contain outdated information so I will be creating my own dataset using a web scraping tool. This tool scraps a video game database website and returns a csv of every game in the database. I will run this tool before I begin my analysis so that I have the most up to date data possible.

The database I will be getting my data from has tens of thousands of games each with varying degrees of data associated with them. Many games contain missing or inaccurate information. For example, many games don't have data on how many total units have been sold. After I have the full dataset I will go through and clean the data up a bit so that it will be more useful for my analysis. This cleaning will mainly involve deleting any games that don't have total sales information, but I also might add in some data from other sources to fill in some of the gaps.

Milestones

1. Get initial dataset from scraping tool.
2. Clean data to make it more useful for my analysis.
3. Write the code to create the visualizations.
4. Analyze the visualizations and determine which results I want to share.
5. Create a final presentation.