Ai Labs DS intern evaluation

Use this link:

https://www.dropbox.com/s/ge8dsb56vv9ofd4/movie_data.csv

to download a dataset called `movie_dataset.csv` that contains summaries of about 42,000 movies scraped from Wikipedia, along with some metadata about each movie. Use this dataset to answer the following questions:

- What are the 5 most popular genres?
- What words are characteristic of the movie summaries in those genres?
- An empirical observation known as Zipf's law is often used to describe the distribution of word frequencies in text corpora. Do you see evidence of Zipf's law in the summaries?

These prompts are intentionally open-ended. We're testing for your ability to:

- write well-documented, reproducible code.
- make reasonable assumptions about a dataset and understand how those assumptions might affect your results.
- effectively communicate your methods and results.

While working, please use git for version control. When you're finished, write up a short summary of your methods and results accessible to an audience of data scientists. Make sure to clearly explain any numbers or figures you provide. Submit the code necessary to reproduce your results (including the `.git` repository), a readme with dependency and environment info, and a PDF version of your summary in a zipped directory to the email address this prompt came from.