INVESTIGATION OF A DATASET - TMDB MOVIES DATASET

In this task, you will be performing advanced data analysis on a dataset using Python libraries such as NumPy, Pandas, and others. The dataset provided contains information about various movies, including their budget, revenue, genre, and more. Your task is to answer the following 20 questions by applying data analysis techniques and providing clear instructions for using the mentioned Python libraries.

1. Single Variable (1D) Questions:
    1.1. Which year had the highest release of movies?
    1.2. Which genre had the highest release of movies?
    1.3. Which 10 actors are casted the most?
    1.4. Identify the keywords that appear most frequently in movie descriptions.
    1.5. Calculate the average popularity for each genre category.
    1.6. Calculate the total revenue for each genre category.
    1.7. Calculate the average vote average for movies released before the year 2000.
    1.8. Perform a time series analysis of movie releases and visualize the trend over the years.
2. Multivariable (2D…) Questions:
    2.1. Which length (runtime) is suits viewers most based on movie popularity?
    2.2. What is the correlation between movies' budgets and their revenue?
    2.3. What is the correlation between average ratings and revenue generated?

Instructions:

1. Work in a Jupyter Notebook.
2. Clean the data extensively before analysis.
3. Use Python libraries (NumPy, Pandas …) mostly (and minimal Python code).
4. Include comments where necessary please to expound more on your approach.

Deliverables:

1. A final report of your analysis in the form of a PDF document (including the introduction with table of contents, wrangling section, cleaning section, exploration section, conclusion).
2. A link to a simple presentational slide of the summary of your report (< 10 pages).
3. A link to a GitHub repo with all your source code.

Note: You are free to explore and analyse the dataset beyond the mentioned questions, as long as you cover all the tasks specified. Make sure to address the business and data science aspects wherever applicable. Feel free to consult online documentation and resources for Python libraries to enhance your analysis.

Happy Data Analysis!