

# Movie Comparison Case Study Rubric

A UVA Data Science Case Study by Maajid Husain, May 2024

**How will I know if I have succeeded?** You will meet the expectations on the Movie Comparison Case Study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"><li>• Repository – A github repository containing all of the materials<ul style="list-style-type: none"><li>○ Submit a link to the repository</li><li>○ Everything is contained in the repo or linked</li></ul></li><li>• Contents include<ul style="list-style-type: none"><li>○ README.md</li><li>○ DATA directory</li><li>○ SCRIPTS directory</li><li>○ OUTPUTS directory</li><li>○ LICENSE</li></ul></li></ul>
README.md	<ul style="list-style-type: none"><li>• GoalL Oreint the reader with the contents of the GitHub repository</li><li>• Explanation of how/where the data was queried from</li><li>• Give a tree of the contents of the repository and what is included in each folder</li><li>• SCRIPTS<ul style="list-style-type: none"><li>○ Explain how to run the code and include what dependencies are needed to execute code</li></ul></li><li>• DATA<ul style="list-style-type: none"><li>○ Explain what each of the data files contain as well as what the sources are</li><li>○ Explain how the data was cleaned</li></ul></li><li>• OUTPUTS<ul style="list-style-type: none"><li>○ Explain what figures/tables should be seen in this folder</li></ul></li></ul>

	and what script should produce which figure
LICENSE	<ul style="list-style-type: none"> <li>• Goal: This file should explain how someone can use the contents of the repository</li> <li>• Use the MIT license</li> </ul>
SCRIPTS	<ul style="list-style-type: none"> <li>• Goal: This folder should contain all the scripts that can be executed by someone viewing the repository</li> </ul>
DATA	<ul style="list-style-type: none"> <li>• Goal: Your uncleaned and cleaned datasets should be included</li> <li>• Include a data dictionary explaining what each field in each dataset represents</li> <li>• Links to where the data was queried from should be used</li> </ul>
OUTPUTS	<ul style="list-style-type: none"> <li>• Goal: Should contain figures you used in your analysis</li> <li>• *Hint: Some useful figures could show sentiment scores distributions for each of the movies selected</li> </ul>
Presentation	<ul style="list-style-type: none"> <li>• Goal: Demonstrate your thought process, findings, and conclusions</li> <li>• Should include the following: <ul style="list-style-type: none"> <li>○ Title</li> <li>○ Motivation</li> <li>○ Hypothesis and Approach</li> <li>○ Data Acquisition</li> <li>○ Analysis Plan</li> <li>○ Tricky Analysis Decision</li> <li>○ Results / Conclusion</li> <li>○ Next Steps</li> <li>○ References</li> </ul> </li> </ul>

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from [Streifer & Palmer \(2020\)](#).