## ☐ Zeth-Abney / Flatiron-projects (Public)

₽ main ▼ Flatiron-projects / phase-1-project /	•••
Zeth-Abney Create README.md	2 minutes ago 🖰 History
presentation	19 minutes ago
unzippedData	6 days ago
<u>Canvas</u> .canvas	6 days ago
	6 days ago
CONTRIBUTING.md	6 days ago
LICENSE.md	6 days ago
MS_studios_analysis.ipynb	19 minutes ago
PROJECTDESC.md	16 minutes ago
☐ README.md	2 minutes ago
first_round_eda.ipynb	4 days ago
second_round_eda.ipynb	16 hours ago
setup_hollywood.ipynb	3 days ago
setup_testBase.ipynb	4 days ago
🗅 sqlite3	6 days ago
unzipper.ipynb	4 days ago

## README.md

0

This exploratory data analysis serves as my first milestone project at Flatiron school as a data science student.

The data used is originally provided in a zipfolder full of csv/tsv files. It is currently in the folder 'unzipped'. The data files were originally unzipped in the notebook 'unzipper' The raw data is passed to into a sqlite database 'testBase' in the 'setup\_testbase' notebook. This raw data is then explored in 'first\_round\_eda', where the desired tables are determined. The notebook 'setup\_hollywood' sets up the hollywood.db sqlite database, queries the desired tables from testBase.db into pandas dataframes, performs some preliminary data cleaning, and passes those dataframes into hollywood.db as tables. The notebook 'second\_round\_eda' then explores further the data in hollywood.db, engineers new data features, and seeks to discover informative behaviors of the data via plotting with matplotlib and seaborn.

The final analysis is presented in the notebook MS\_studios\_analysis. A non-technical presentation of the analysis is available in a video file in the presentation folder.