

Data Modeling Assignments

You can share your work by uploading jupyter notebooks to a github repo and sharing a pull request for each assignment.

Grading Criteria:

1. Use of libraries
2. Execution
3. Selection of visualization plots
4. Comments
5. Insights and Inferences derived

Bonus Points For:

1. Going the extra mile to ask more questions related to the dataset
2. Going the extra mile to answer those questions using the existing dataset or bringing in another dataset
3. Following good software engineering practices

Assignment DM1- Scraping Websites

Scrap the wikipedia pages of your top 5 favourite movies, and scrape the following

1. The movie name
2. Director
3. Year of release
4. Overall gross earning
5. The cast

Assignment DM2 - Merging Dataset

We have provided you with a dataset that has a compiled list of car accidents.

<https://drive.google.com/drive/u/0/folders/1AISQxIXnAKMImYwpwVTs6FMMKz44mB4C>

Your task is to

1. Merge the datasets
2. Conduct exploratory data analysis to identify
 - a. Which city had the most amount of accidents
 - b. Identify if there was any time of the year or day that had more reported accidents? Try to infer why.
 - c. There are several other factors indicated in the dataset- weather condition, latitude-longitude, sunrise_sunset etc. Pick any one factor and predict what effect it might have had on the number of accidents. Confirm with a visualization. Were you right or wrong?

Assignment 3 - Finding Datasets

Pick your own dataset from one of the following domains (You can merge more than 1 data set for bonus points.)

1. Healthcare
2. Retail purchases
3. Sports

Come up with min 5 questions you want your data to answer and use visualizations to answer the same.