**PROJECT PROPOSAL**

* *What dataset or datasets do you plan to use? What are the features, rows and data types of each?*

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset?select=IMDb+movies.csv>

I am planning to use IMDb movies.csv and IMDb ratings.csv files in the dataset. Here is the information for the files.

Table

Description automatically generated Table

Description automatically generated

* *What research or business questions do you want to answer?*

Does the director of the movies impact ratings?

* *What are your hypotheses going in?*

There is no difference between Woody Allen movies ratings and average ratings for all other directors’ movies.

* *How will you use your data to test your hypotheses?*

First, I will filter my dataset to have samples with common features. Then I will use A/A test for my samples. After concluding that confounding variables are controlled for, I will continue with t-test between two groups.

* *Who will find your findings valuable, and how will they use them?*

I will deliver data driven insights for streaming services. They can use this information to decide whether it worth buying these movies.

Busra Kurt