# EECS 731 - Introduction to Data Science

## Topic: To be, or not to be

- Repository: https://github.com/Priyanka1527/Plays_Classification
- Project structure setup as suggested
- Shakespeare Plays Dataset was download and Pandas dataframe were used to read the data: Related Notebook - "Shakespeare_Plays.ipynb" under **/notebooks**
- Using Feature Engineering we can analyze the distribution number of players corresponding to a particular Play: Relevant histogram report for sample play "Henry IV" - "Player in Play.png" under **/reports**
- Classification Model I used - Decision Tree. I've committed all the process and details in the Jupyter Notebook "Classification Model_DecisionTrees.ipynb". However, the Accuracy score is very low.

As part of this assignment, I learned and got more familiar with the process of Data Preparation, Data Exploring, Feature Extraction and finally code a classifier model. However, i couldn't reach upto an optimal result with high accuracy. It could be due to my feature selection or classifier model.