

Assignment 1

DSSA 5001 – Introduction to Data Science | Due 12/11

Please submit the 2 questions with 2 separate files (.py, .ipynb, etc):

1. Download the spotify_charts ZIP file from Blackboard and unzip. Submit your Python code via Jupyter Notebook (.ipynb) and upload to Blackboard.

Import the spotify_charts.csv file into your Jupyter Notebook environment.

- Filter the original dataset to only include "Ed Sheeran" and data after December 31st, 2020. Save the result as a new Data Frame.
- Determine which Ed Sheeran song experienced the most amount of "MOVE_UP" trends
- Filter the Ed Sheeran dataset further by sub-setting by solely "Shape of you". Create a histogram of number of daily streams with a bin size of 10. What are the range of values that occur most frequently?

2. Find and download a dataset of your choosing, preferably a dataset that piques your interest.

Complete exploratory data analysis on this dataset in a Jupyter notebook. As a reminder, there is no set-by-set guide that is applicable to each dataset, as every dataset is different. However, your goals in general are to:

- Understand your variables
- Clean the dataset
- Analyze the relationship between variables

In addition, please meticulously annotate your notebook with your thoughts as you progress through the EDA process, writing down new questions that arise as you observe and analyze your data as comments. Not only is this best practice, but you'll be able to reference your analysis in detail in the future. The code is secondary to your thought process and insights about your findings.