Evon Troy Alexander
CAP4770
Professor Samira Zad
February 12, 2024

Assignment 1 Report

Introduction: This assignment provided unique challenges for me. While I am familiar with Python and actively use it to study Data Structures & Algorithms I have never used the Matplotlib library. Prior to commencing the assignment I viewed a Matplotlib Crash Course on the freeCodeCamp YouTube channel. This video provided me with a great overview of how to utilize this library for the assignment, additionally documentation was used as a reference.

Reflection: Steps 1-4 were pretty straight forward, the assignment itself provided a lot of guidance on syntax and common conventions to use. For example:

- Given a *DataFrame df*, to get a certain column, use the column label enclosed in brackets [] similar to when you want to retrieve data from a python dictionary: *df['fare']*
- Given a *DataFrame* column, you can find the mean, std, min, and max by calling *DataFrame.mean*, *DataFrame.std*, *DataFrame.min*, *DataFrame.max* respectively.
- Given a *DataFrame df*, to filter out observations with a specific condition, e.g. 'fare'>100, you can use the following syntax *filtered_df = df.loc[~(df['fare']>100)]* or *filtered_df = df.loc[df['fare']<=100]*
- Given a *DataFrame df*, to draw histogram of column 'fare' with 30 bins, you can use the following functions (after importing *matplotlib.pyplot as plt*):
  - *matplotlib.pyplot.hist(df['fare'], density=True, bins=30)*
  - *plt.ylabel('Probability')*
  - *plt.xlabel('Fare');*

I found this incredibly helpful.

One area of issue was step 5. I was not that confident in my calculation for ascertaining the IQR of fare and age and would appreciate any feedback.

Overall I really enjoyed this assignment, many of the things learned in this assignment have real world application and I appreciate assignments like that.