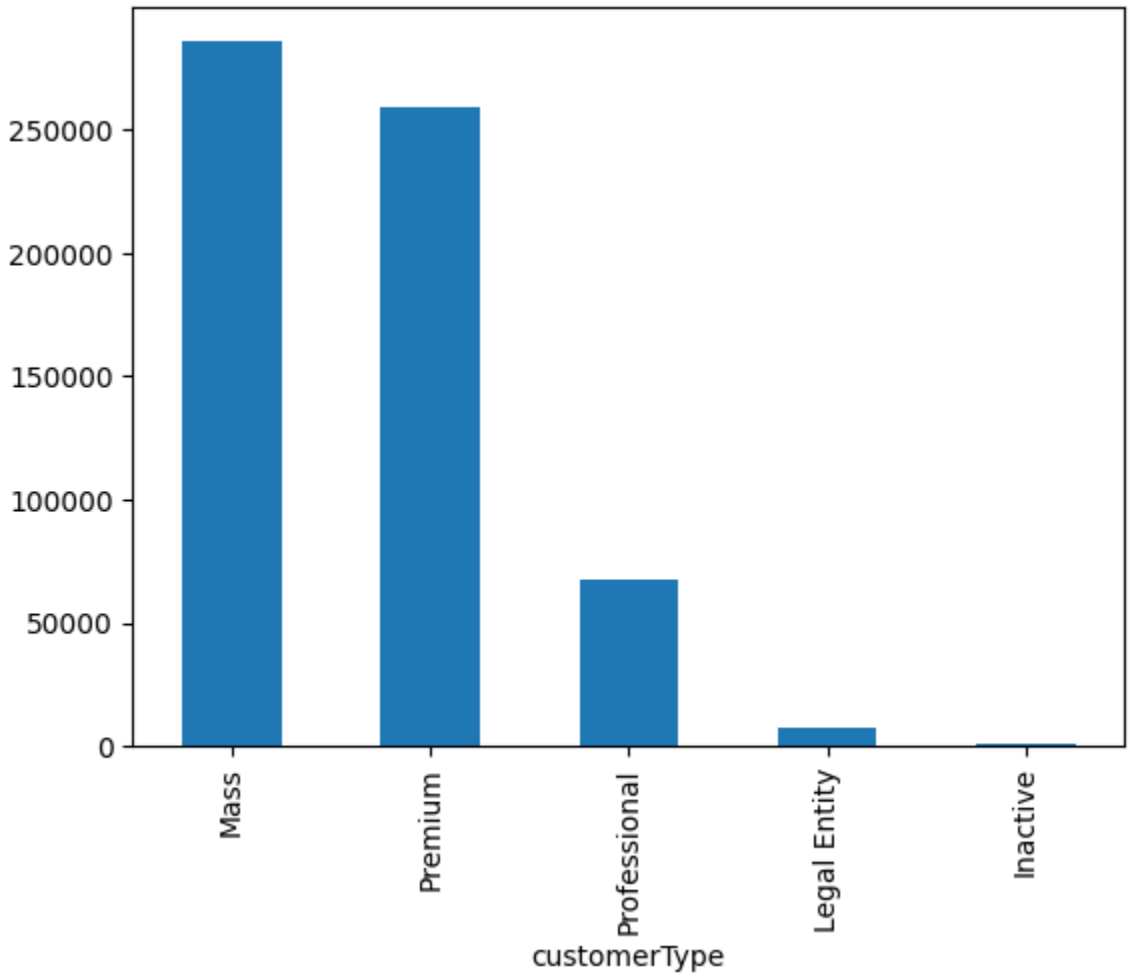
FAR-Trans Rough Draft

- Problem Introduction
    - For this project I will be analyzing a Financial Asset Dataset from which I will attempt to answer a few basic questions to gain a better understanding of investment behavior. This project will eventually help me gain a better understanding of Machine Learning and recommender systems for financial assets. I will attempt to answer the following questions:
        - What is the monthly distribution of transactions for each type of security (Bonds, Stocks, Mutual Funds).
        - Do the most volatile securities provide the most return on investment and who trades them (Individuals, Professionals, Institutions)?
        - Are the most active investors also the most profitable?
        - Are there any behavior changes in investments or customer risk tolerances before and after the COVID pandemic (2021)?
- Data Introduction
    - https://researchdata.gla.ac.uk/1658/
    - This data set was published by the University of Glasgow. It was intended to provide data for financial asset recommendation systems. The dataset is comprised of several documents which provide customer information, securities such as stocks, bonds, and mutual funds. It also includes a time series of customer transactions for securities between the years 2018 and 2022. The dataset contains customer profile information (excluding personal information) such as risk

tolerances, income range, customer type (individual, institution, etc.). Collectively, this dataset contains over 1-million records and over 20 different features from which to draw conclusions.

- Data Pre-processing
    - This dataset is relatively clean, however the document which contains the asset information has several null/missing values. However, this does not pose an issue because these features are not needed for my visualizations.
    - The original research paper from this dataset mentions that it has been pre-processed to address some of the common issues with financial data such as Sell transactions which do not have a Buy transaction preceding it at an earlier date. Assets with closing date gaps greater than 10 days have also been removed from the dataset to ensure consistency. Lastly, stock forward splits and reverse splits have been addressed by transforming the data preceding the split providing a consistent ratio across the time series.
    - I plan on filtering the data based on several criteria as well as performing several aggregate operations to group data according to my needs.

- Data Visualization



- Storytelling
    - The previous bar graph shows the number of transactions performed by customer type. This clearly shows Mass traders which are non-professional traders with less than 60K in investments execute most of the trades compared to Professional traders and Legal Entities.
    - While this visualization has not directly answered any of my questions, it will help me narrow down the most active investors to determine if they are more profitable than less active investors.

- Impact

- Sources
    - Javier Sanz-Cruzado, Nikolaos Droukas, Richard McCreadie. FAR-Trans: An Investment Dataset for Financial Asset Recommendation. IJCAI-2024 Workshop on Recommender Systems in Finance (Fin-RecSys). Jeju, South Korea, August 2024.