

Introduction



This paper researches if there is a better way to cluster stocks into sectors then by company descriptions. By performing KMeans clustering algorithm on the QQQ exchange-traded fund, it measures the correlations between the predicted sectors and their stocks vs. the original sectors and their stocks using visuals created by Python's Seaborn library. If successful, this can improve the returns of popular funds and help better classify companies who are invested in multiple industries.

