ETFs or Exchange Traded Funds have become popular tradable securities since their inception in the early 1990s. ETFs are comprised of an assortment of assets, of various classes, and usually track various indices or offer investors exposure to certain industries.

We want to create a tool that can help investors, keen on investing in a certain industry, create a portfolio of industry-relevant ETFs to maximize their return and minimize their risk. The data set that we will be utilizing is the “Huge Stock Market Dataset” from Kaggle. Our starting point will be identifying the ETFs to use in the portfolio, and assigning weights to them. The weights assigned to each of these ETFs will be allocated based on two factors:

1. Historical Performance of ETFs
2. Identifying ETFs whose movement is strongly correlated with relevant market information.
3. Risk profile of the investor

This problem is important because of the rapid growth in ETFs and to help investors maximize their returns and minimize the risk. The data set is an important avenue for identifying the best performing ETFs which should give us a good starting point.