1. Cluster using K-means (and maybe other clustering techniques)

Can try different

2. Build the various portfolios

k=3, k=5, k=7, k=9, k=11 for PCA, t-SNE on aggregate data

Pick the cluster with the highest average return

Balanced Portfolio

Control portfolio - S&P 500

# t-SNE cluster on best year

# t-SNE on worst year

3. Compare portfolios

Overall returns on 2019 data

Differences between stocks

Compare to S&P 500

4. Run analysis on t-SNE structure as a whole

How well does k-means capture the individual sectors from the t-SNE graph

5. How to split money in a portfolio