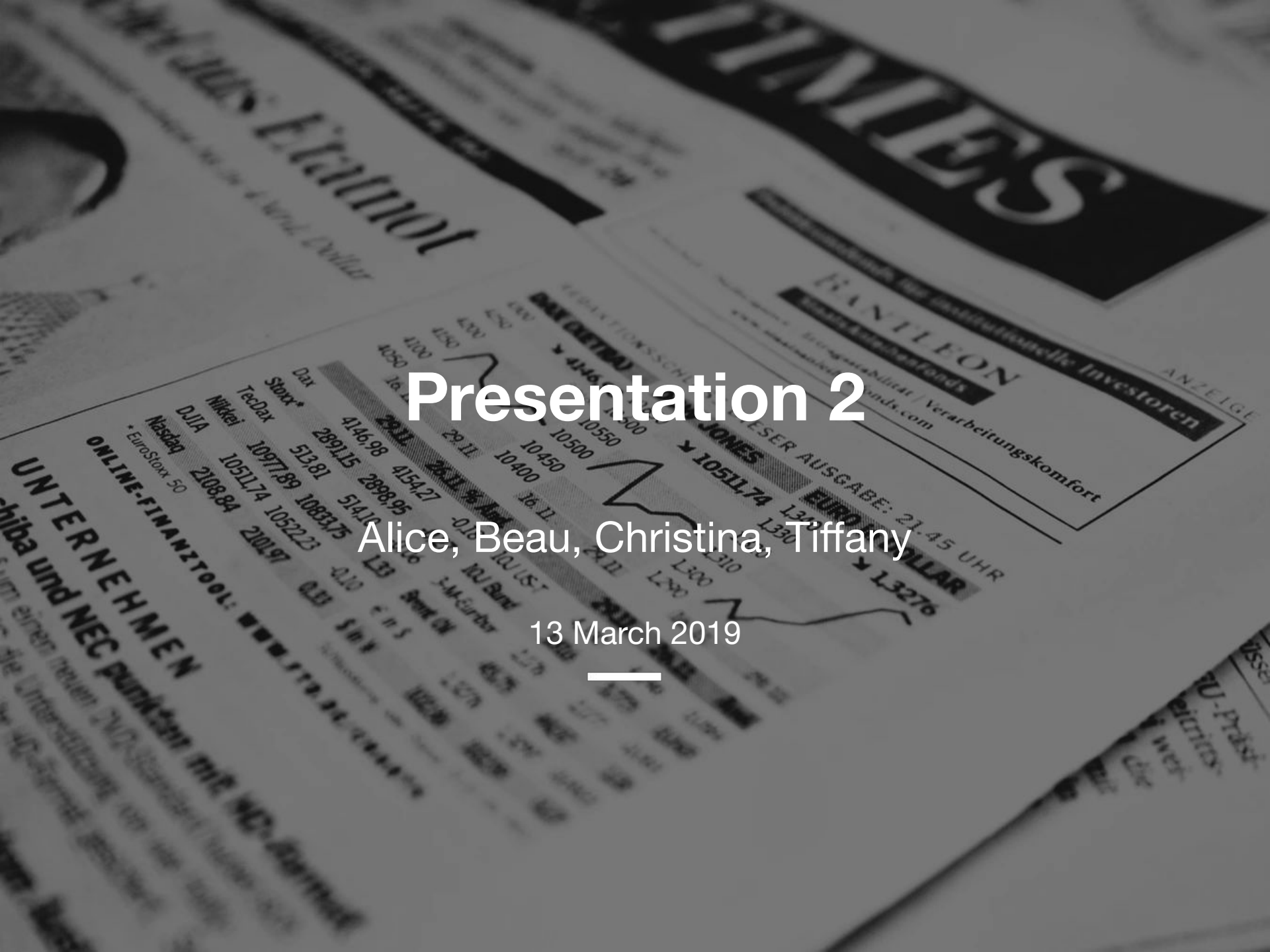


Presentation 2

Alice, Beau, Christina, Tiffany

13 March 2019

The Story Behind Our Product

"She knew there were **opportunities**

...if she could just **see** them."

OUR STORY

How can a manager have a **differentiated view** of markets **that is accurate**?

Overall Architecture

Model Output → Postgres DB → Website

Django + React

Key Features

to address the user's problem

Current Clusters

Shared Characteristics

Historical Comparisons

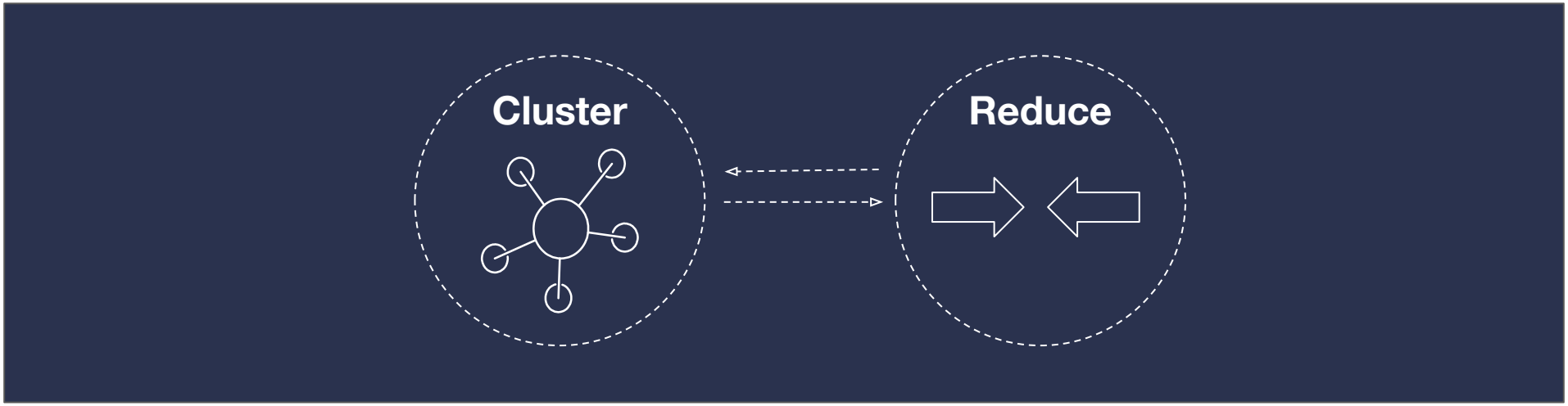
Future Performance

http://tiffapedia-pyxis.herokuapp.com/analytics/

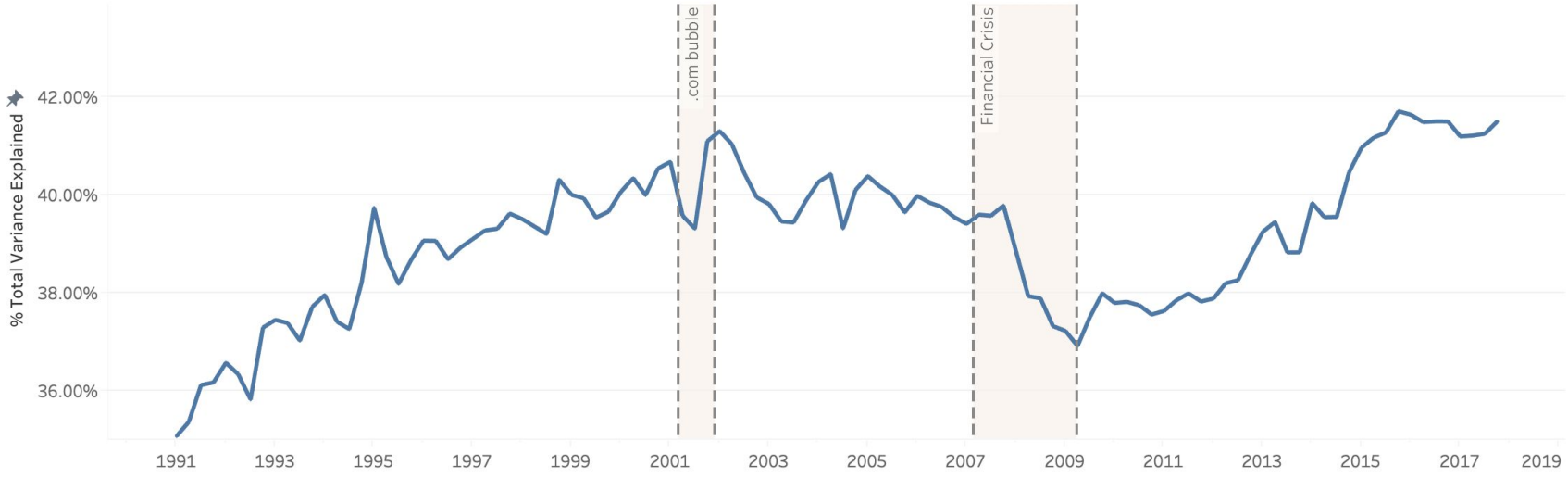
Feature Engineering

1	Universe	• Compustat Unrestated Quarterly Financial Statements • CRSP Monthly Pricing Data • 1990 - 2018
2	Tradability Filters	• $100M Market Cap (Inflation Adjusted) • $1M Dollar Volume (Inflation Adjusted) • No Financial Firms
3	Data Cleaning	• NAs to Zeros • Gap Filling • Delisting and Relisting
4	Quarterly to TTM	• TTM Numbers for Income Statement and Cash Flow Items
5	Feature Creation	• 9 Feature Categories: Profitability, Asset Structure, Solvency, Utilization, Liquidity, Deployment, Sourcing, Growth, Acceleration
6	Feature Normalization	• Quantile Transformation

Unsupervised Learning

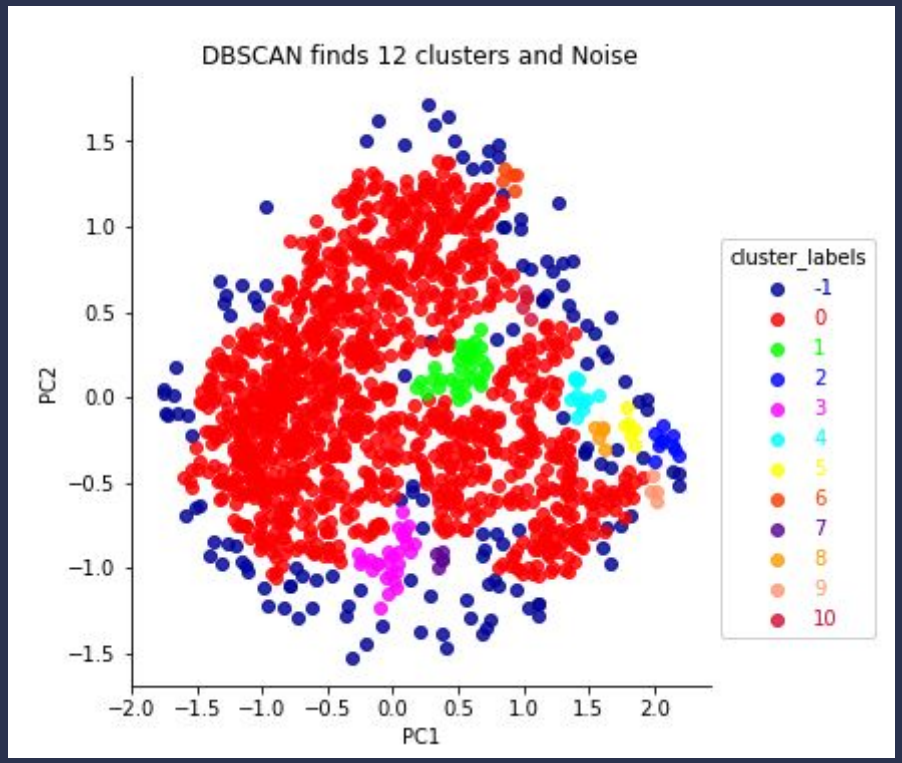


PCA: Total Variance Explained by the first two Principal Components

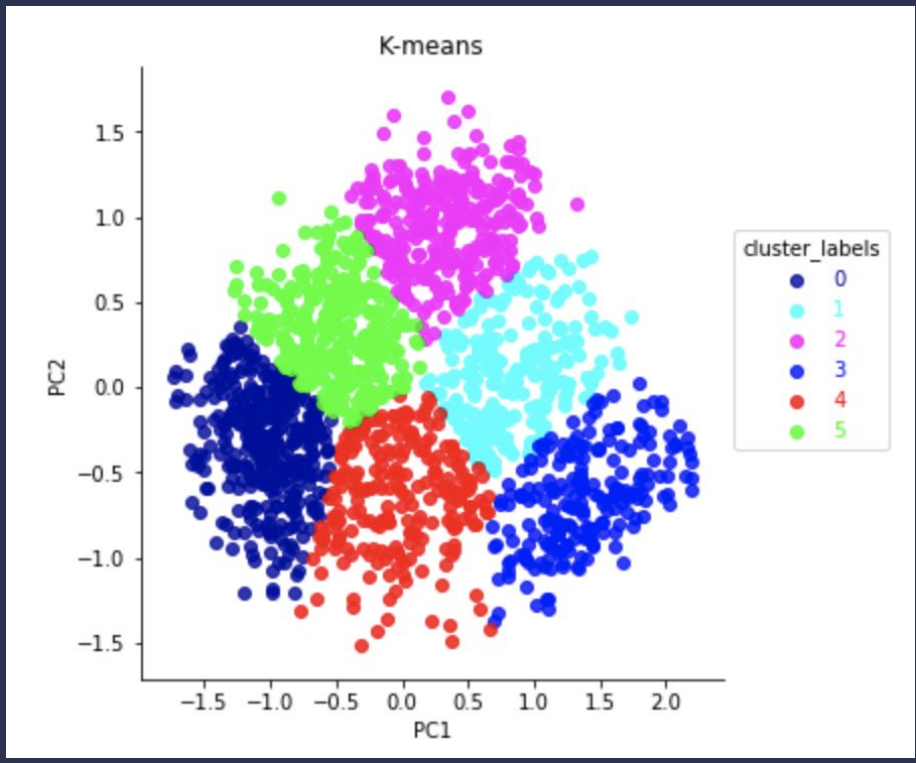


Choosing the Clustering Algorithm:
Visual Inspection

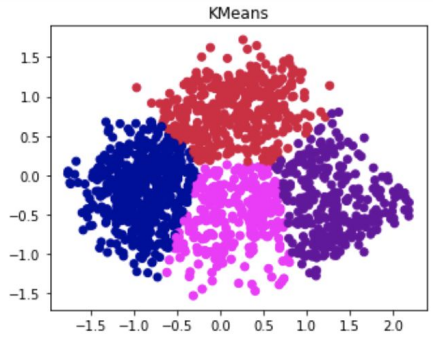
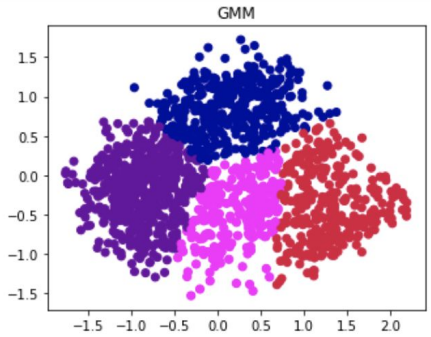
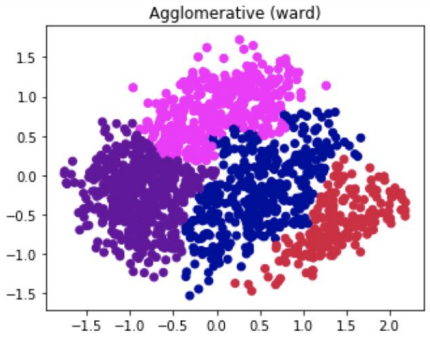
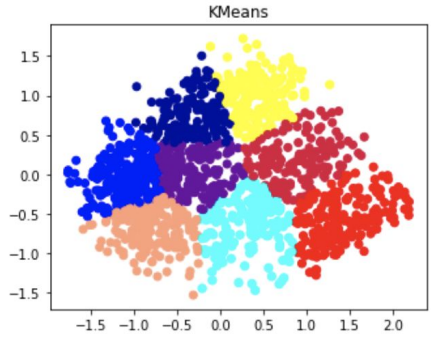
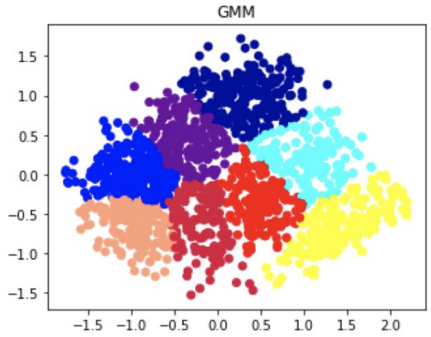
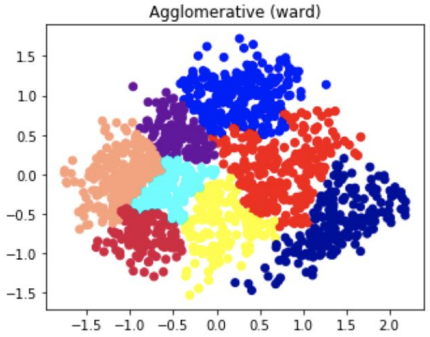
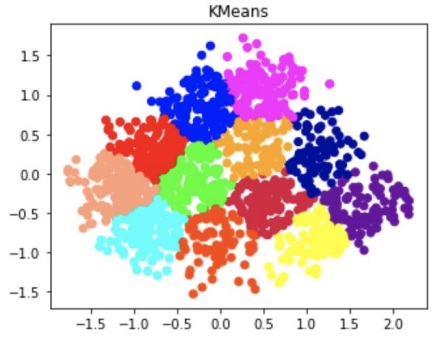
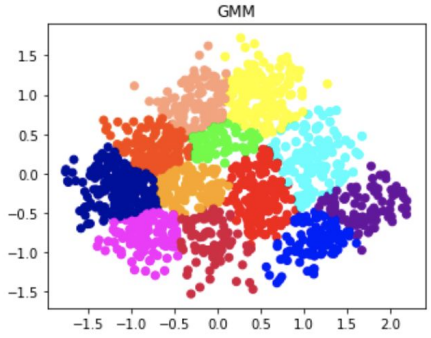
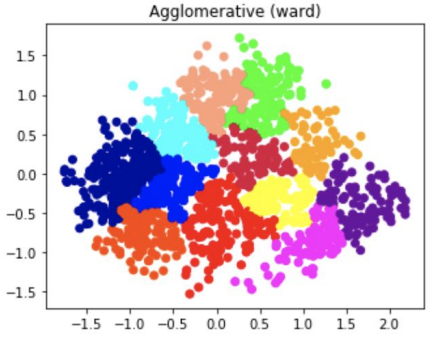
Bad Clustering 👎



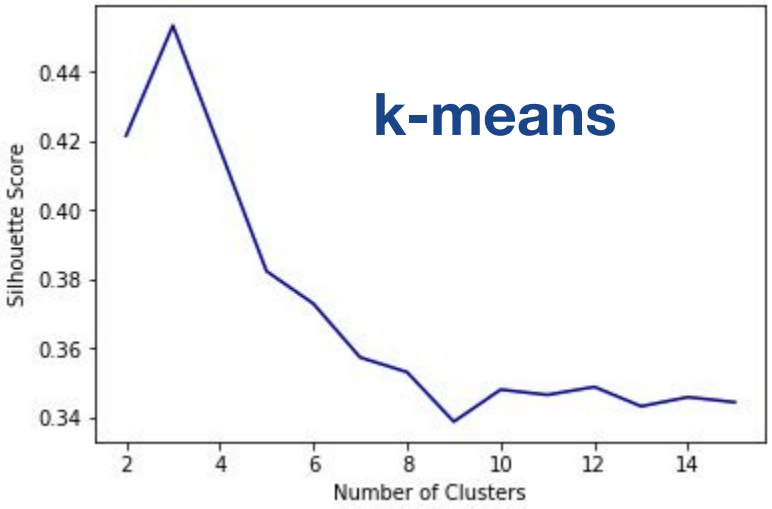
Good Clustering 👍

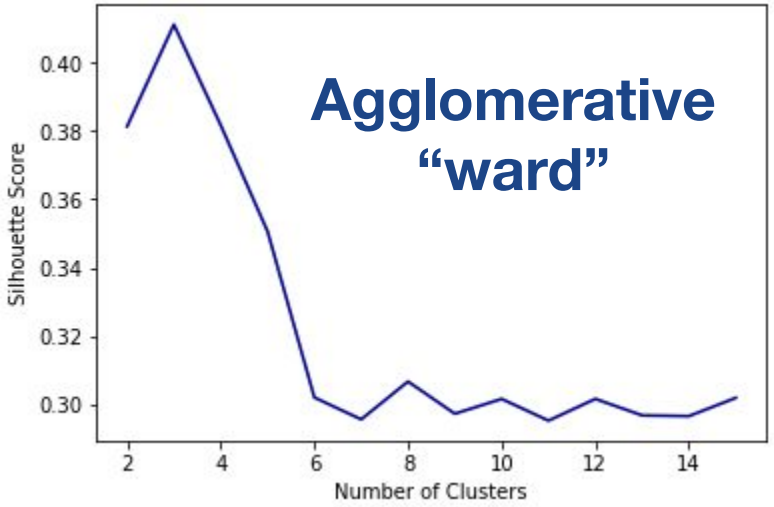


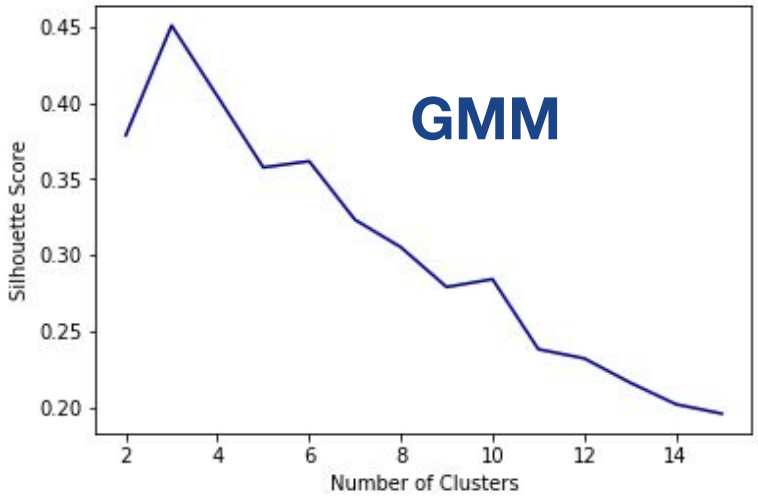
Clustering Algorithms

	K-means	Gaussian Mixture	Agglomerative
k = 4			
k = 8			
k = 12			

Choosing the Clustering Algorithm:
Silhouette Score & Optimal K

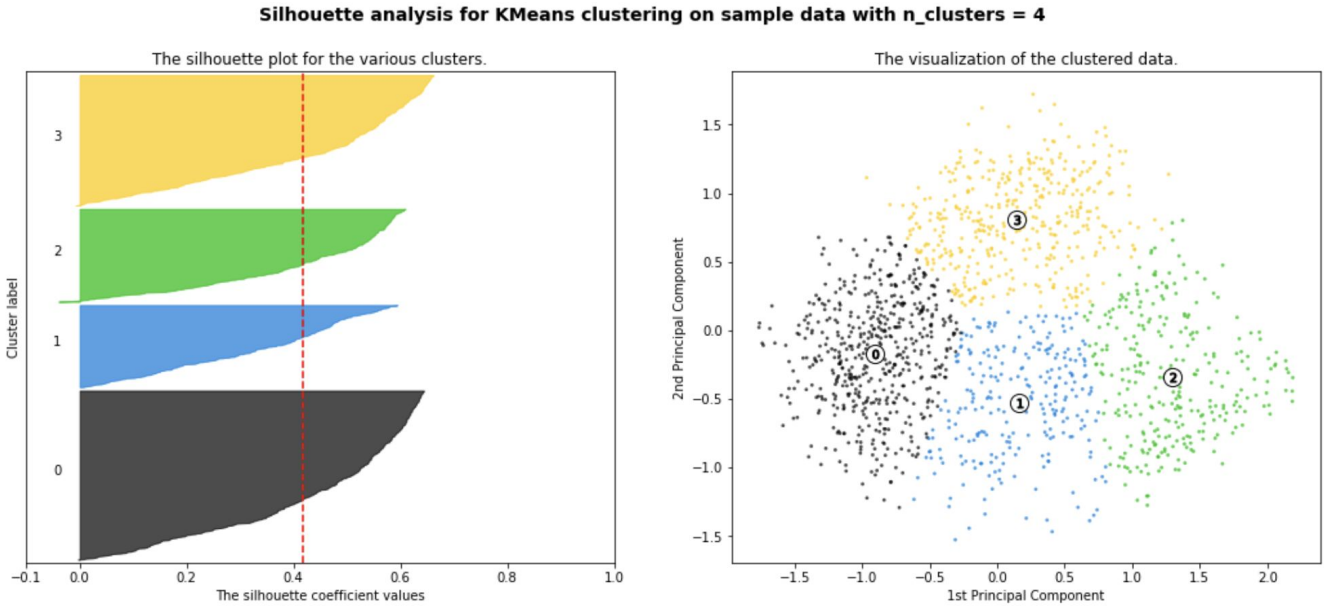


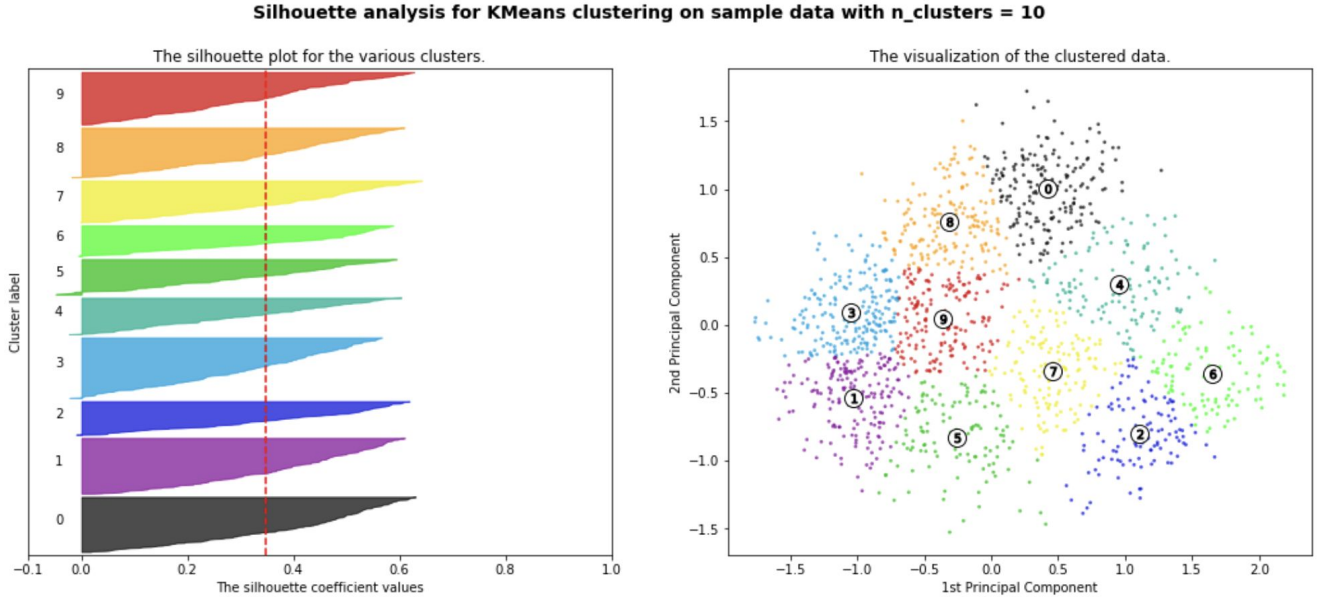




Affinity Propagation:
optimal k = 11

Evaluation: Silhouette Score

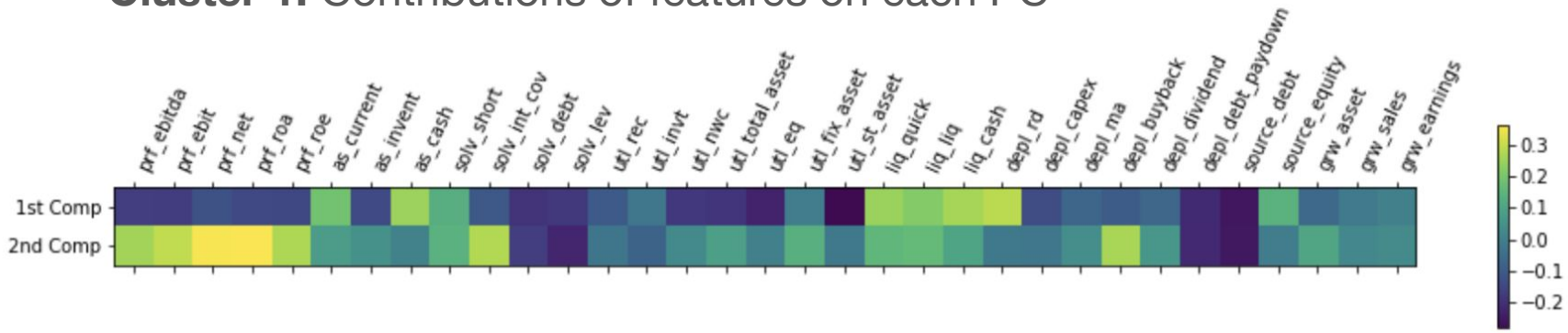


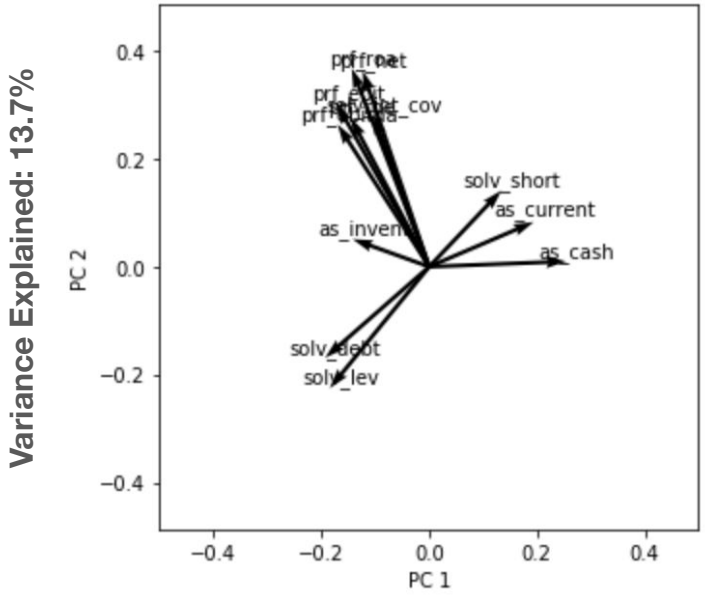


Feature Importance in each Cluster

Factor Maps

Cluster 1: Contributions of features on each PC





Variance Explained: 25.2%

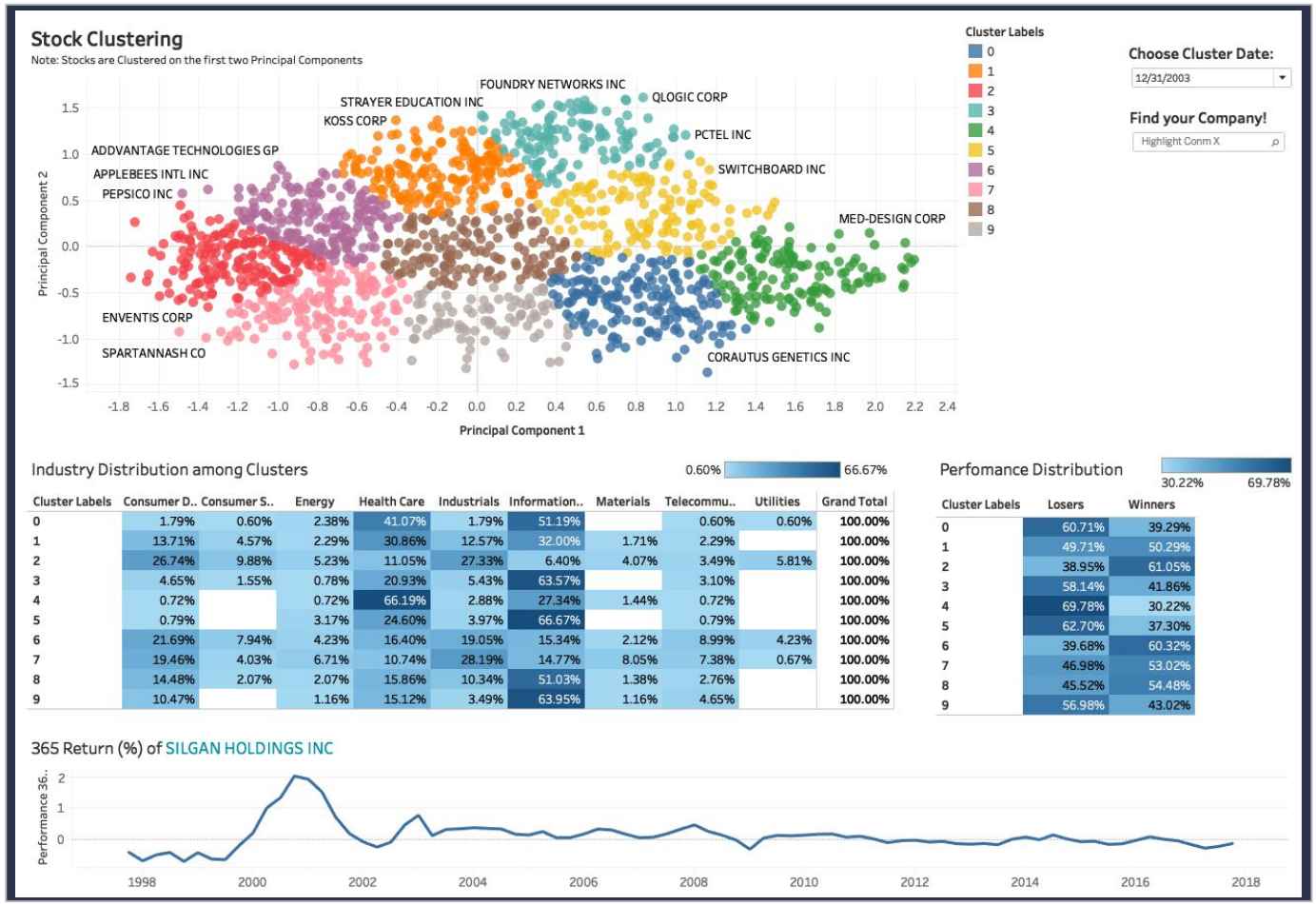
(Explained Variance PC1)*(Contribution to PC1)

(Explained Variance PC2)*(Contribution to PC2)

Score for Feature 1

Top Features in Cluster 1:

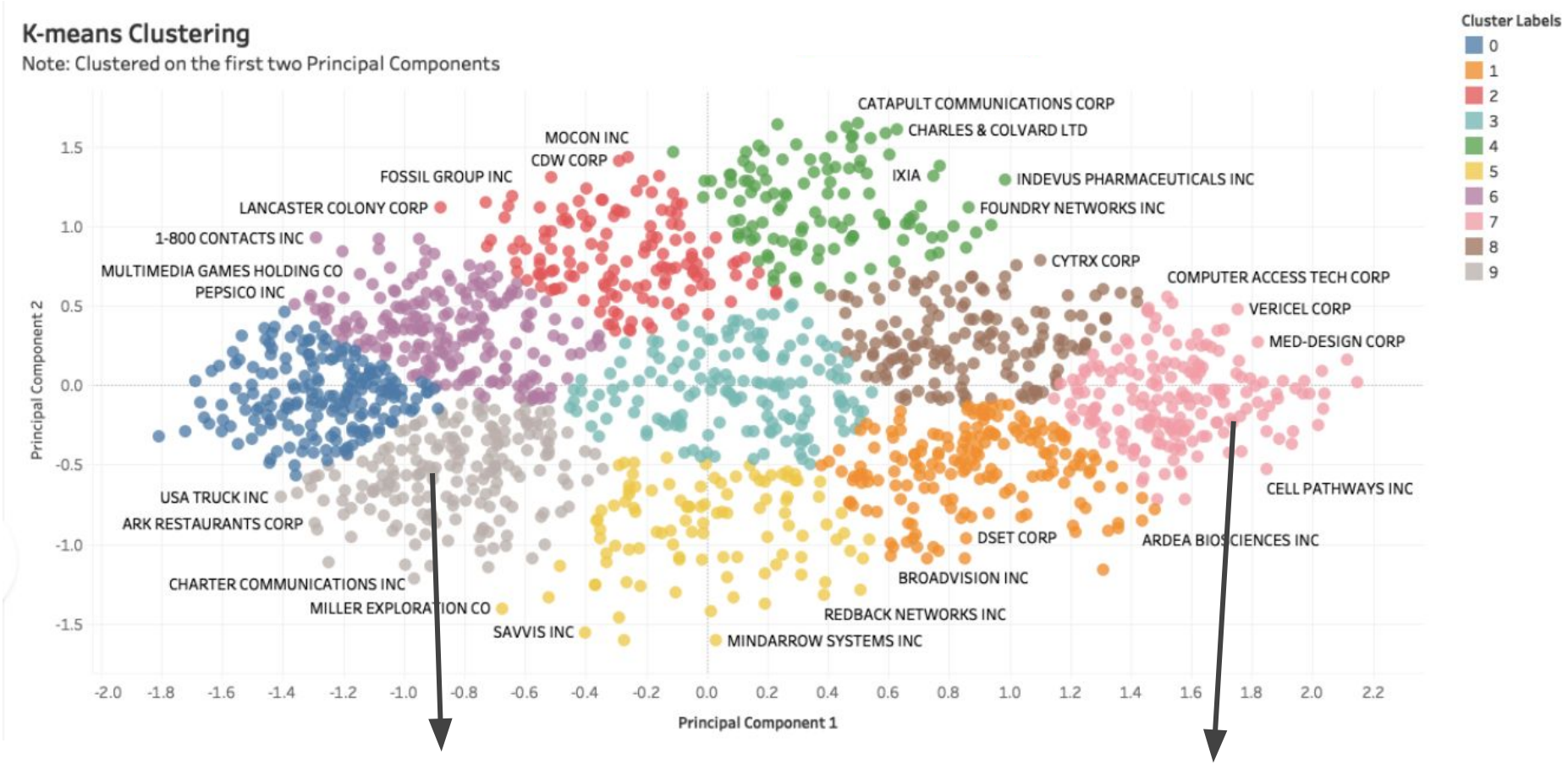
Source Debt
Return on Assets (ROA)
Liquidity Ratio
Debt Paydown
EBIT



Demo Time!

https://public.tableau.com/shared/HFFHGBXKW?:display_count=yes

What can we learn from these labels?

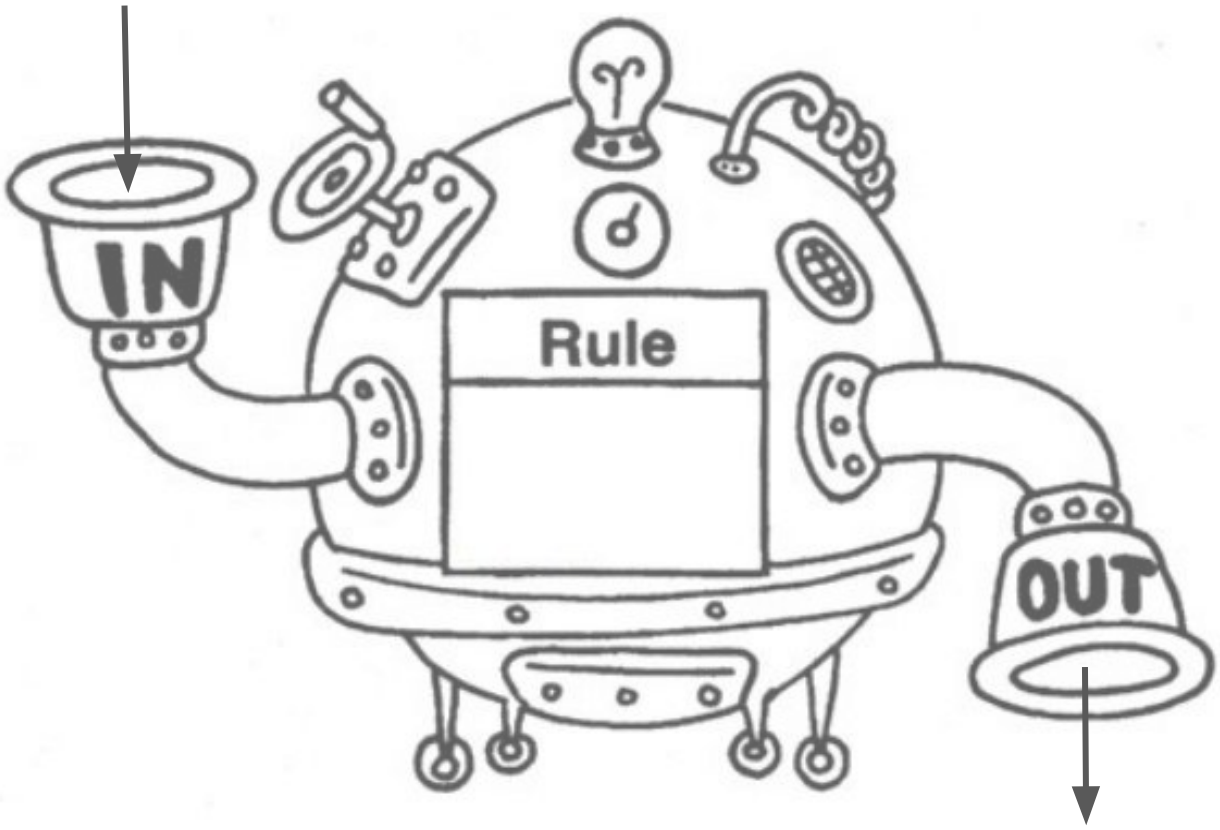


Cluster to buy?

Cluster to sell?

Supervised Learning

Cluster Labels



Stock Return Predictions

Supervised Learning:
Y labels and train-test scheme

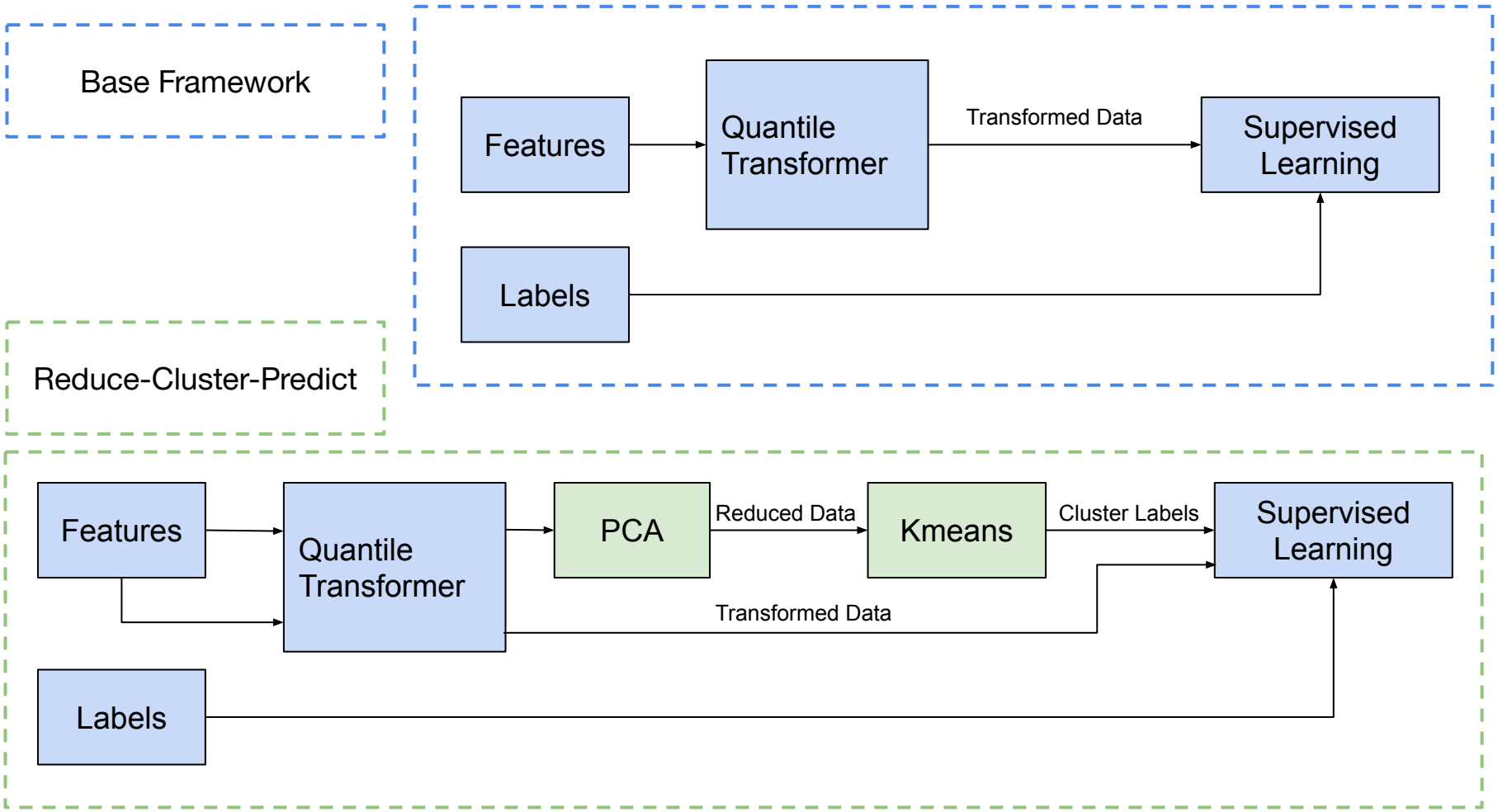
Labels - Relative Performance vs S&P 500

Relative Performance (%)	[-100,-33)	[-33,-10)	[-10,10]	(10,33]	(33,inf]
Y Label	-2	-1	0	1	2

Train-Test Scheme

Q1 20X4	Q2 20X4	Q3 20X4	Q4 20X4	Q1 20X5	Q2 20X5	Q3 20X5	Q4 20X5	Q1 20X6
Train				Test				
	Train				Test			
		Train				Test		
			Train				Test	
				Train				Test

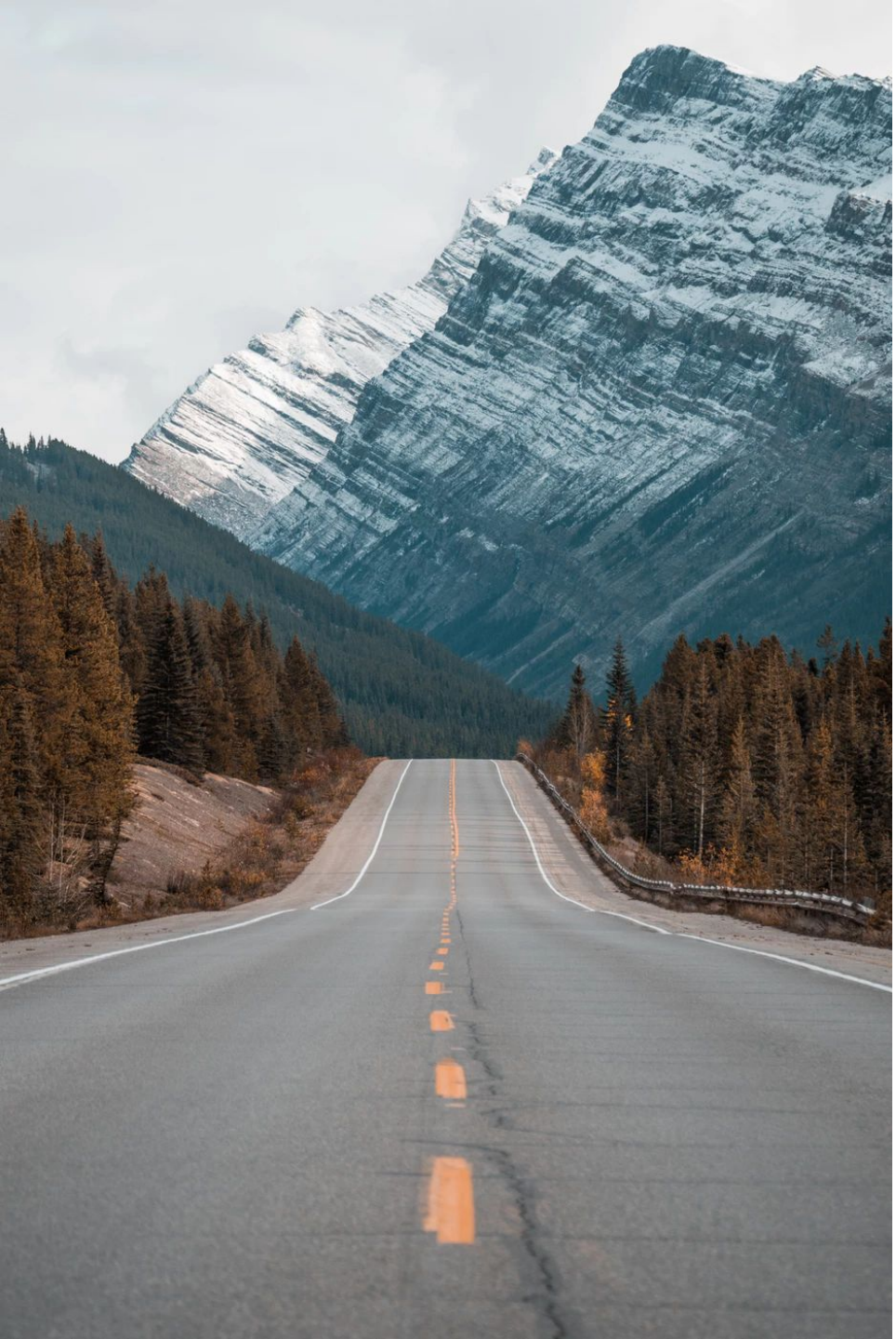
The Use of Cluster Labels in Supervised Learning



Preliminary (Untuned) Results

Using RandomForestClassifier

	Base	Reduce-Cluster-Predict
Weighted Average Accuracy	0.269	0.293
Top Features	depl_capex	**cluster_10**
	grw_earnings	**cluster_1**
	grw_sales	**cluster_7**
	source_equity	**cluster_9**
	solv_int_cov	depl_capex
	prf_roe	**cluster_8**
	grw_asset	utl_nwc



The Road Ahead

Steps remaining

- Tune models
- Get data into format for website
- Create the d3 visualization
- Adjust the frontend to make sure all the components are working properly

Help Needed

- Clustering in prediction frameworks

In Closing