

# 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



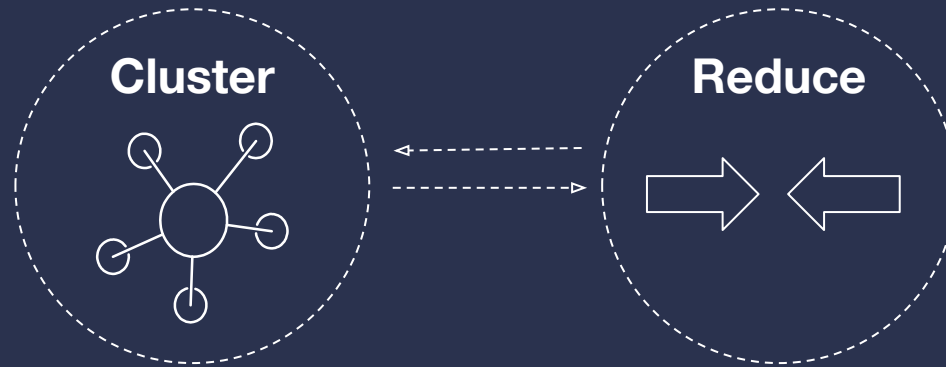
Future  
Performance

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

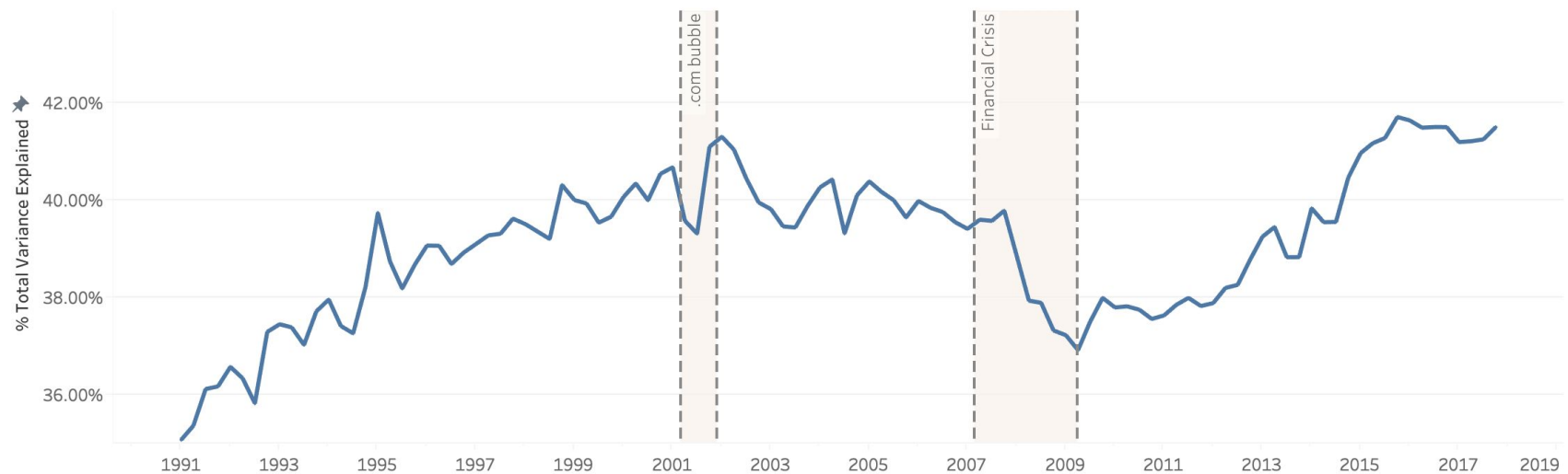
# Feature Engineering

1	Universe	<ul style="list-style-type: none"><li>• Compustat Unrestated Quarterly Financial Statements</li><li>• CRSP Monthly Pricing Data</li><li>• 1990 - 2018</li></ul>
2	Tradability Filters	<ul style="list-style-type: none"><li>• \$100M Market Cap (Inflation Adjusted)</li><li>• \$1M Dollar Volume (Inflation Adjusted)</li><li>• No Financial Firms</li></ul>
3	Data Cleaning	<ul style="list-style-type: none"><li>• NAs to Zeros</li><li>• Gap Filling</li><li>• Delisting and Relisting</li></ul>
4	Quarterly to TTM	<ul style="list-style-type: none"><li>• TTM Numbers for Income Statement and Cash Flow Items</li></ul>
5	Feature Creation	<ul style="list-style-type: none"><li>• 9 Feature Categories: Profitability, Asset Structure, Solvency, Utilization, Liquidity, Deployment, Sourcing, Growth, Acceleration</li></ul>
6	Feature Normalization	<ul style="list-style-type: none"><li>• Quantile Transformation</li></ul>

# Unsupervised Learning

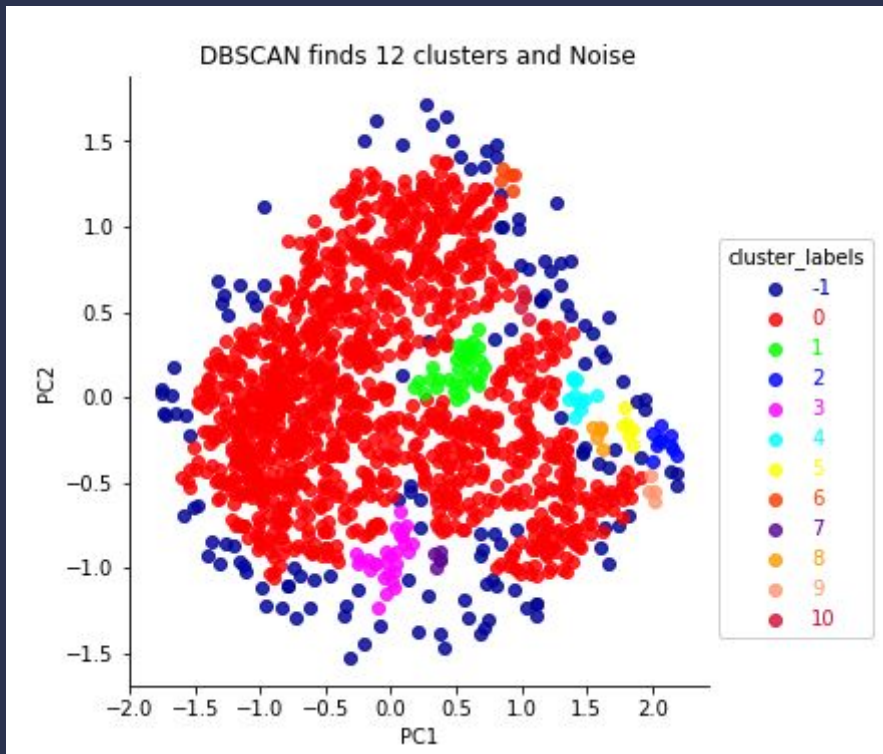


PCA: Total Variance Explained by the first two Principal Components

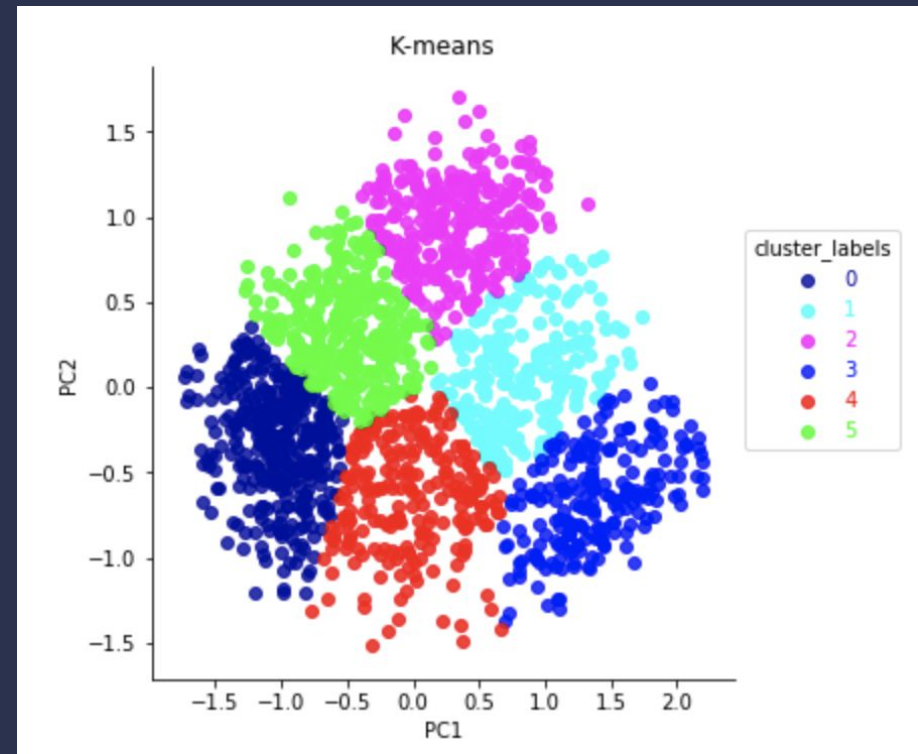


# Choosing the Clustering Algorithm: Visual Inspection

Bad Clustering



Good Clustering





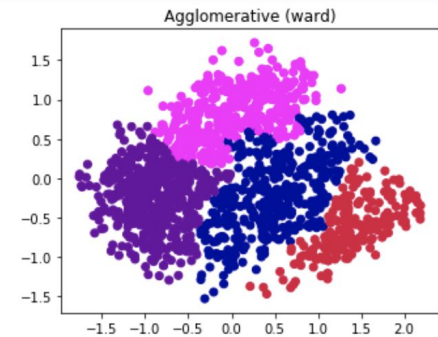
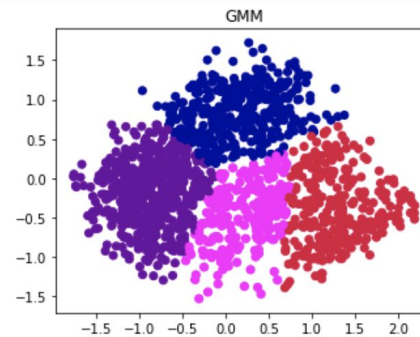
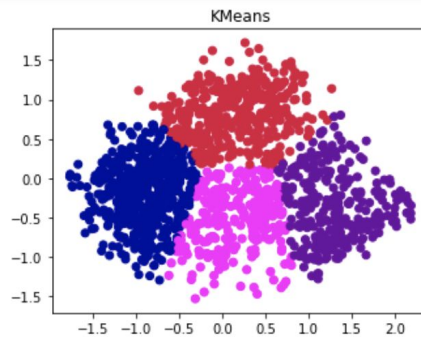
# Clustering Algorithms

K-MEANS

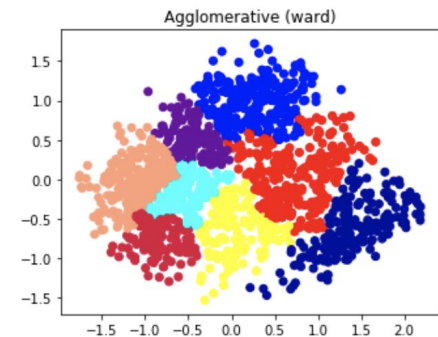
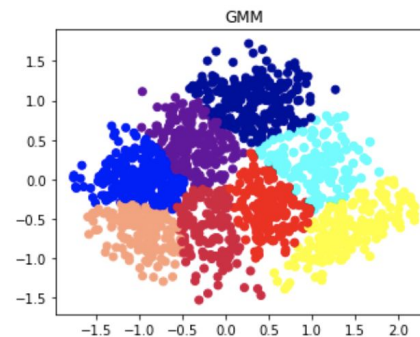
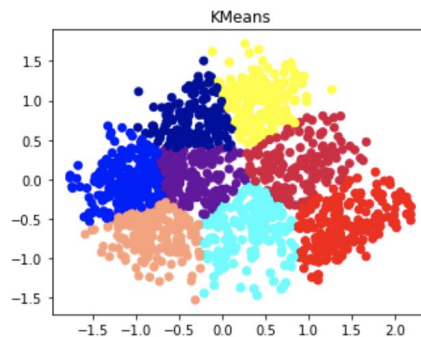
GAUSSIAN MIXTURE

AGGLOMERATIVE

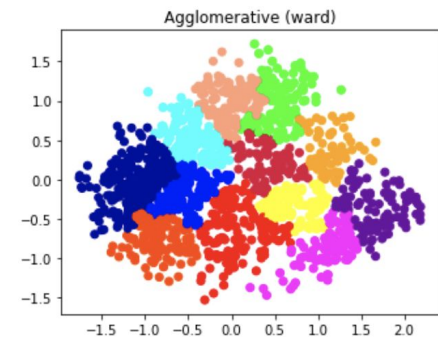
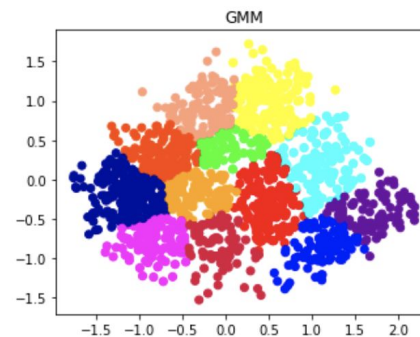
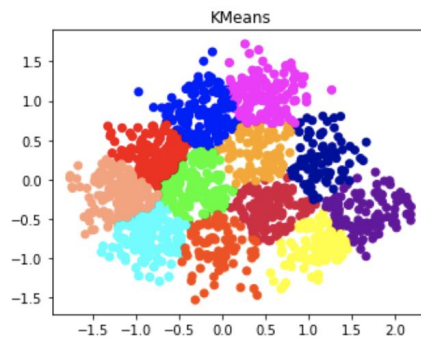
$k = 4$



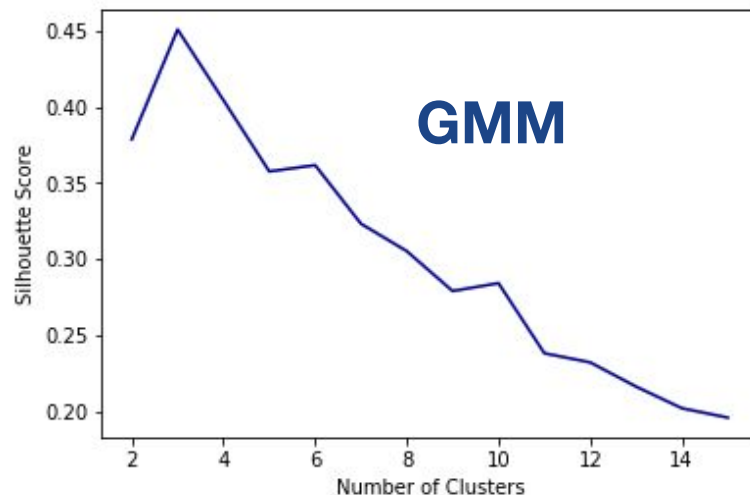
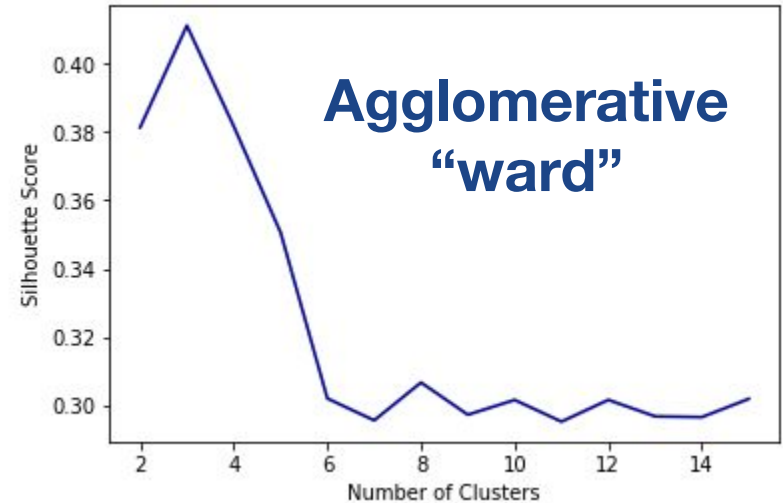
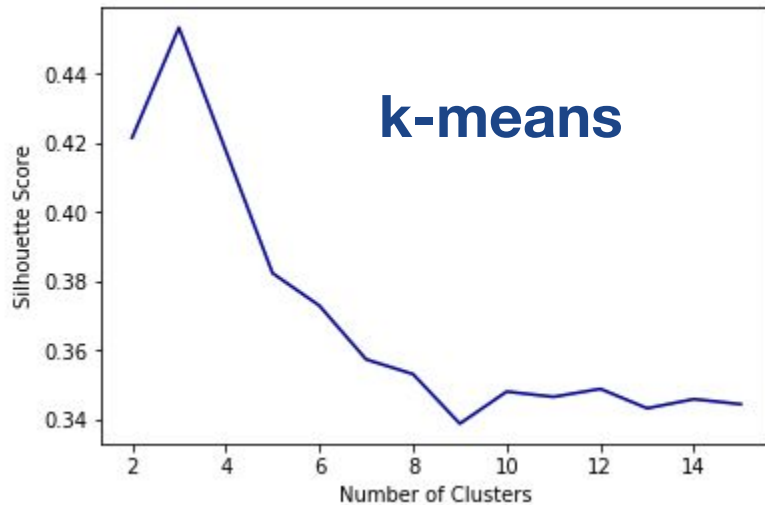
$k = 8$



$k = 12$



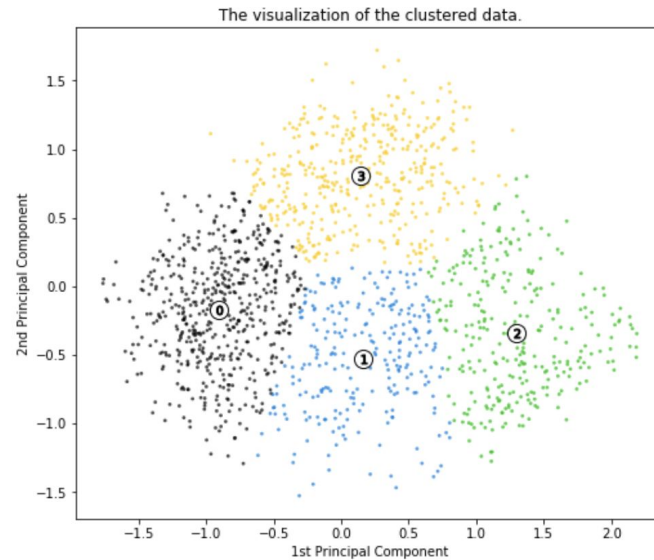
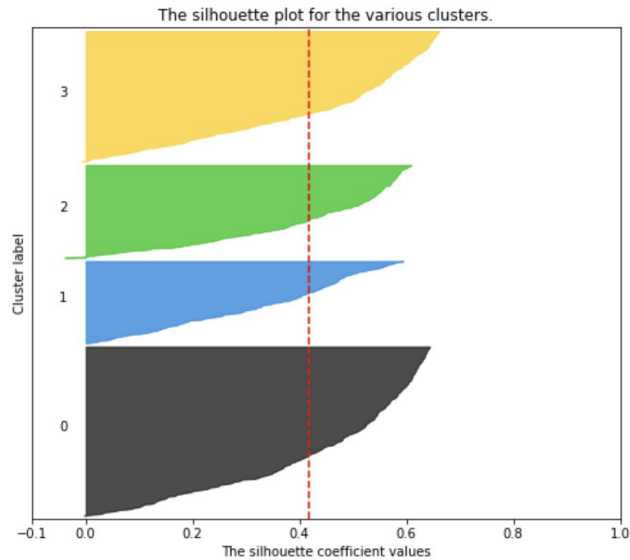
# Choosing the Clustering Algorithm: Silhouette Score & Optimal K



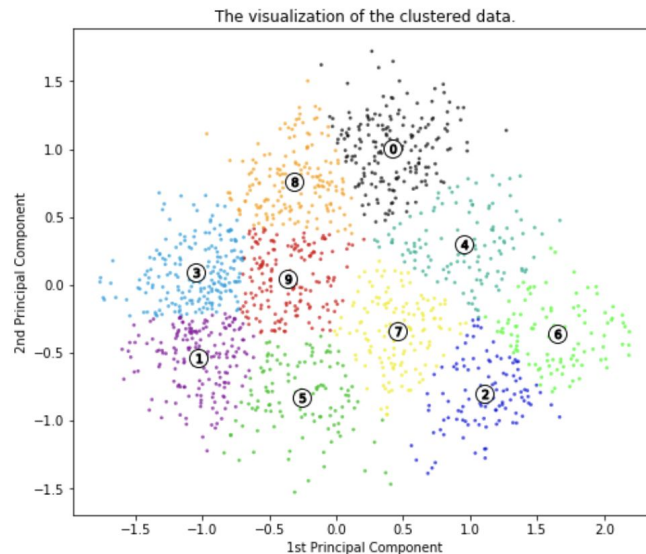
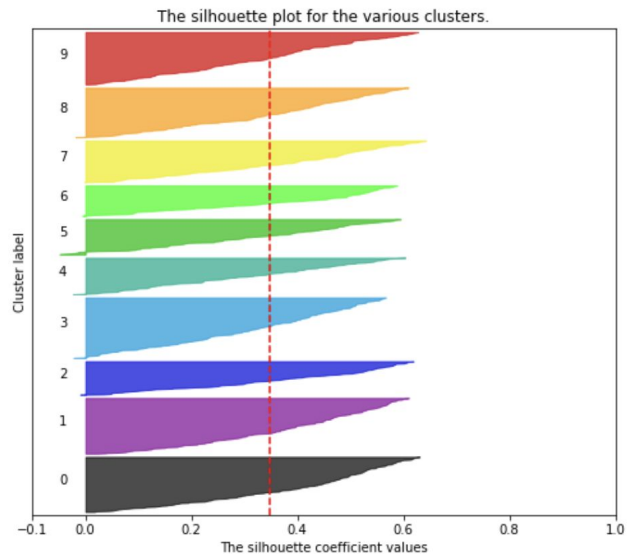
**Affinity Propagation:  
optimal k = 11**

# Evaluation: Silhouette Score

**Silhouette analysis for KMeans clustering on sample data with  $n\_clusters = 4$**



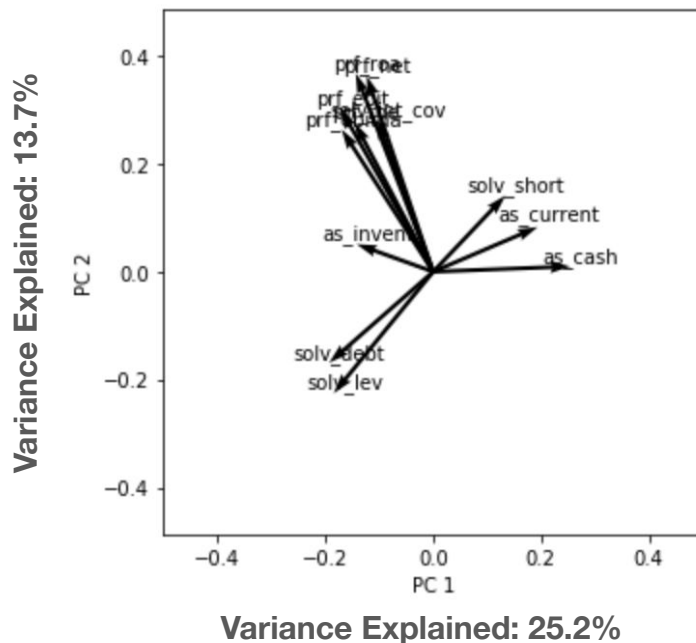
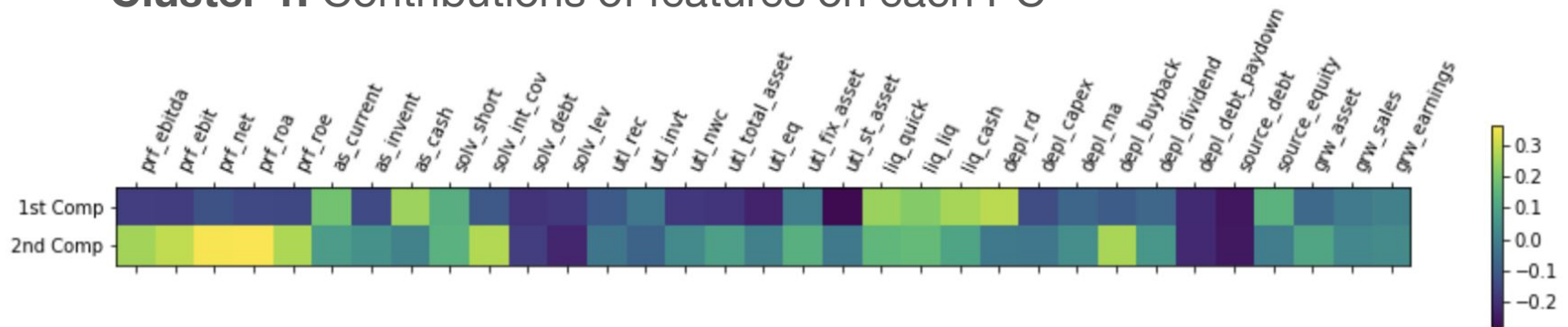
**Silhouette analysis for KMeans clustering on sample data with  $n\_clusters = 10$**



# Feature Importance in each Cluster

## Factor Maps

### Cluster 1: Contributions of features on each PC



$(\text{Explained Variance PC1}) \times (\text{Contribution to PC1})$



$(\text{Explained Variance PC2}) \times (\text{Contribution to PC2})$



**Score for Feature 1**

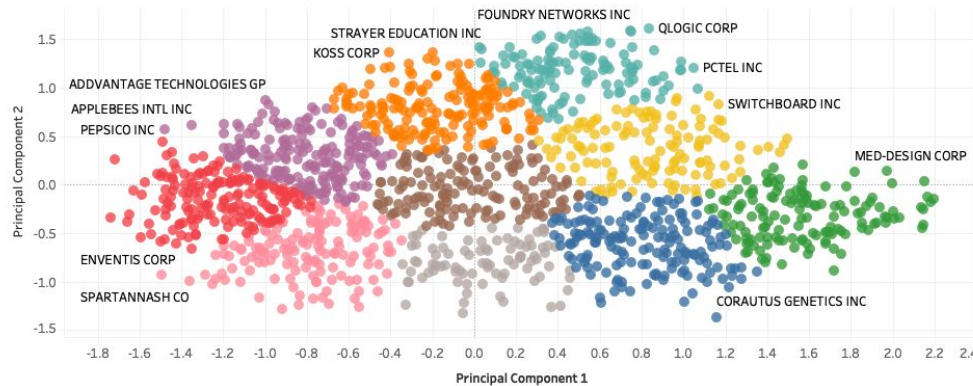


**Top Features in Cluster 1:**

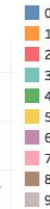
Source Debt  
Return on Assets (ROA)  
Liquidity Ratio  
Debt Paydown  
EBIT

## Stock Clustering

Note: Stocks are Clustered on the first two Principal Components



Cluster Labels



Choose Cluster Date:

12/31/2003

Find your Company!

Highlight Comm X

### Industry Distribution among Clusters

0.60% 66.67%

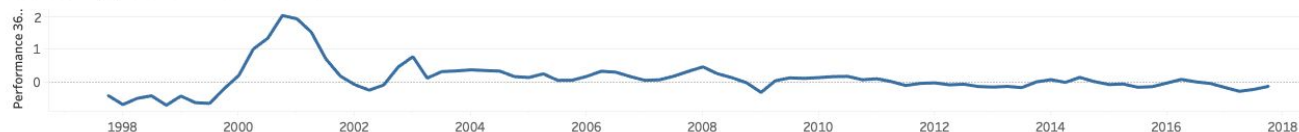
Cluster Labels	Consumer D..	Consumer S..	Energy	Health Care	Industrials	Information..	Materials	Telecommu..	Utilities	Grand Total
0	1.79%	0.60%	2.38%	41.07%	1.79%	51.19%	1.71%	0.60%	0.60%	100.00%
1	13.71%	4.57%	2.29%	30.86%	12.57%	32.00%	1.71%	2.29%		100.00%
2	26.74%	9.88%	5.23%	11.05%	27.33%	6.40%	4.07%	3.49%	5.81%	100.00%
3	4.65%	1.55%	0.78%	20.93%	5.43%	63.57%		3.10%		100.00%
4	0.72%		0.72%	66.19%	2.88%	27.34%	1.44%	0.72%		100.00%
5	0.79%		3.17%	24.60%	3.97%	66.67%		0.79%		100.00%
6	21.69%	7.94%	4.23%	16.40%	19.05%	15.34%	2.12%	8.99%	4.23%	100.00%
7	19.46%	4.03%	6.71%	10.74%	28.19%	14.77%	8.05%	7.38%	0.67%	100.00%
8	14.48%	2.07%	2.07%	15.86%	10.34%	51.03%	1.38%	2.76%		100.00%
9	10.47%		1.16%	15.12%	3.49%	63.95%	1.16%	4.65%		100.00%

### Performance Distribution

30.22% 69.78%

Cluster Labels	Losers	Winners
0	60.71%	39.29%
1	49.71%	50.29%
2	38.95%	61.05%
3	58.14%	41.86%
4	69.78%	30.22%
5	62.70%	37.30%
6	39.68%	60.32%
7	46.98%	53.02%
8	45.52%	54.48%
9	56.98%	43.02%

### 365 Return (%) of SILGAN HOLDINGS INC



# Demo Time!

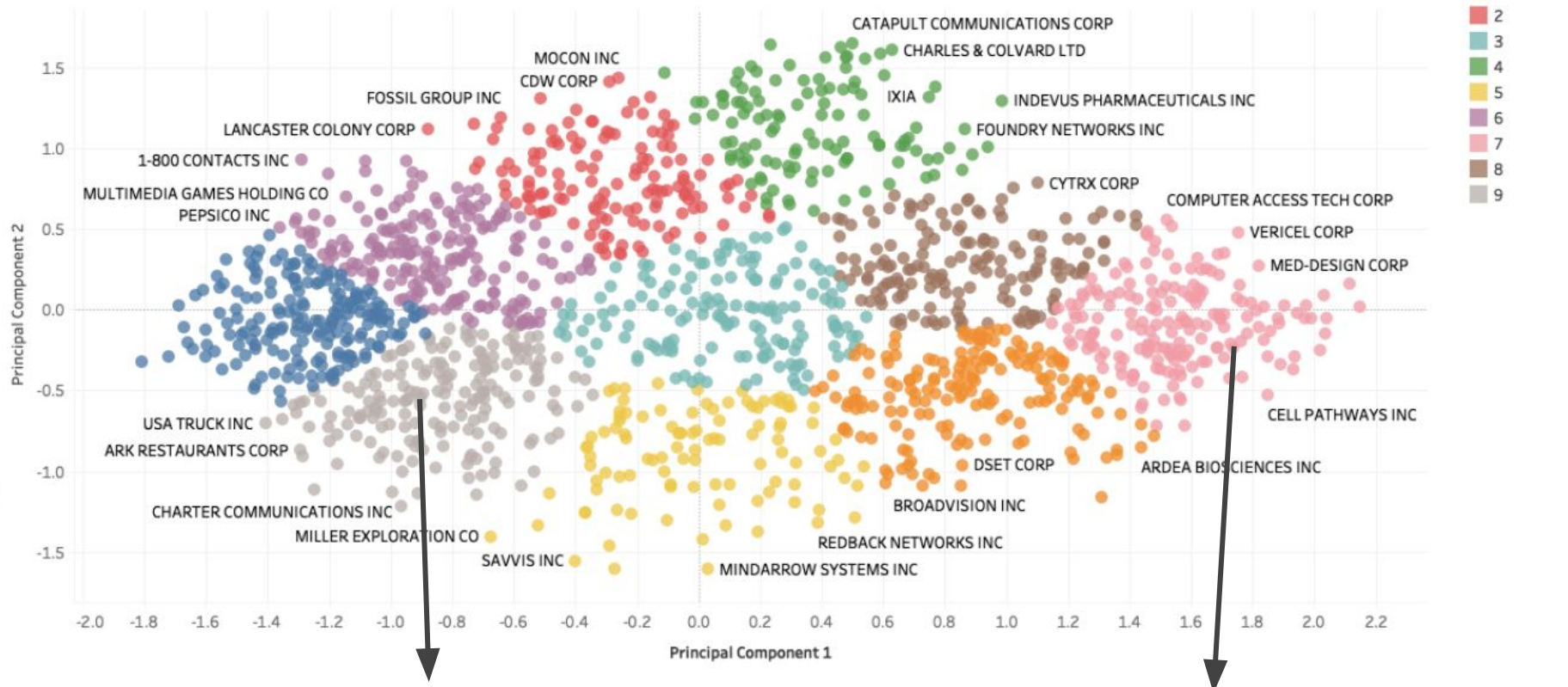
[https://public.tableau.com/shared/HFFHGBXKW?:display\\_count=yes](https://public.tableau.com/shared/HFFHGBXKW?:display_count=yes)



## What can we learn from these labels?

## K-means Clustering

Note: Clustered on the first two Principal Components

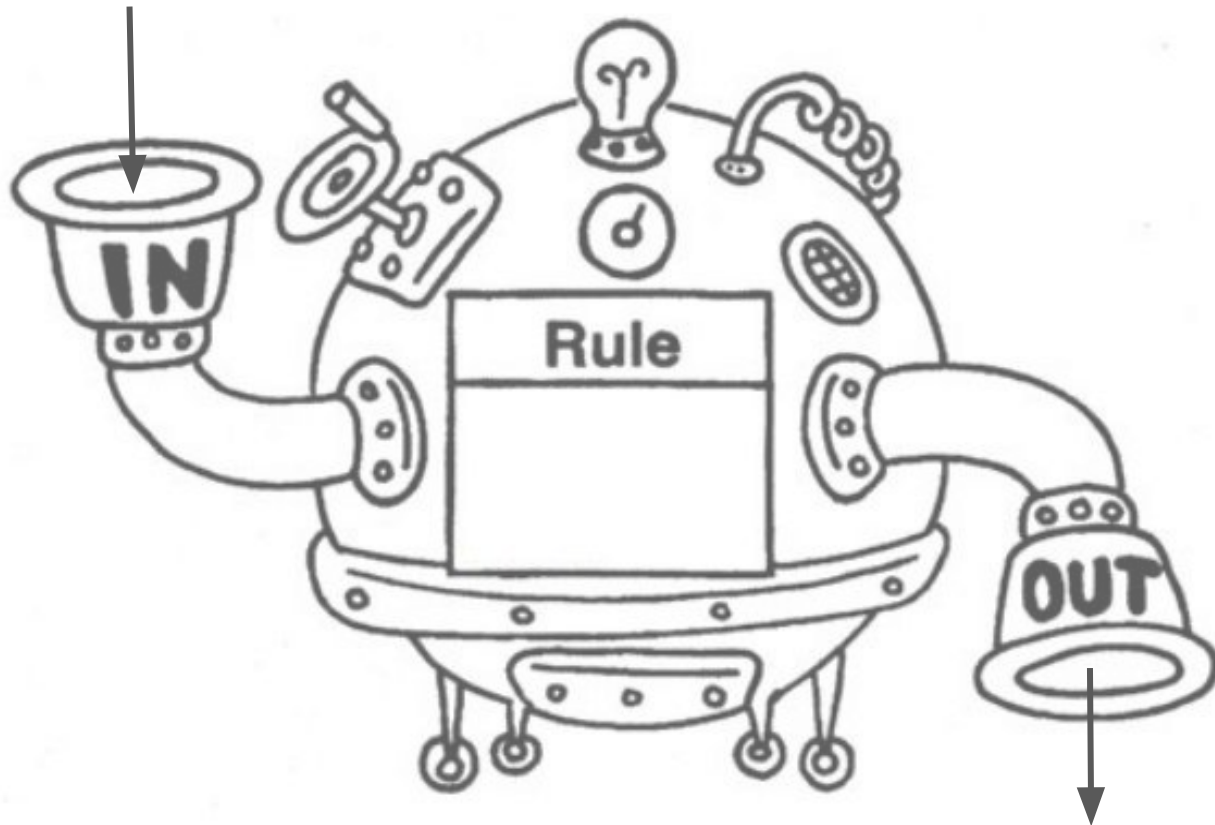


## Cluster to buy?

## Cluster to sell?

# Supervised Learning

**Cluster Labels**



**Stock Return Predictions**

Source: <https://www.kaggle.com/datasets/stock-market>

# Supervised Learning:

## Y labels and train-test scheme

Labels - Relative Performance vs S&P 500

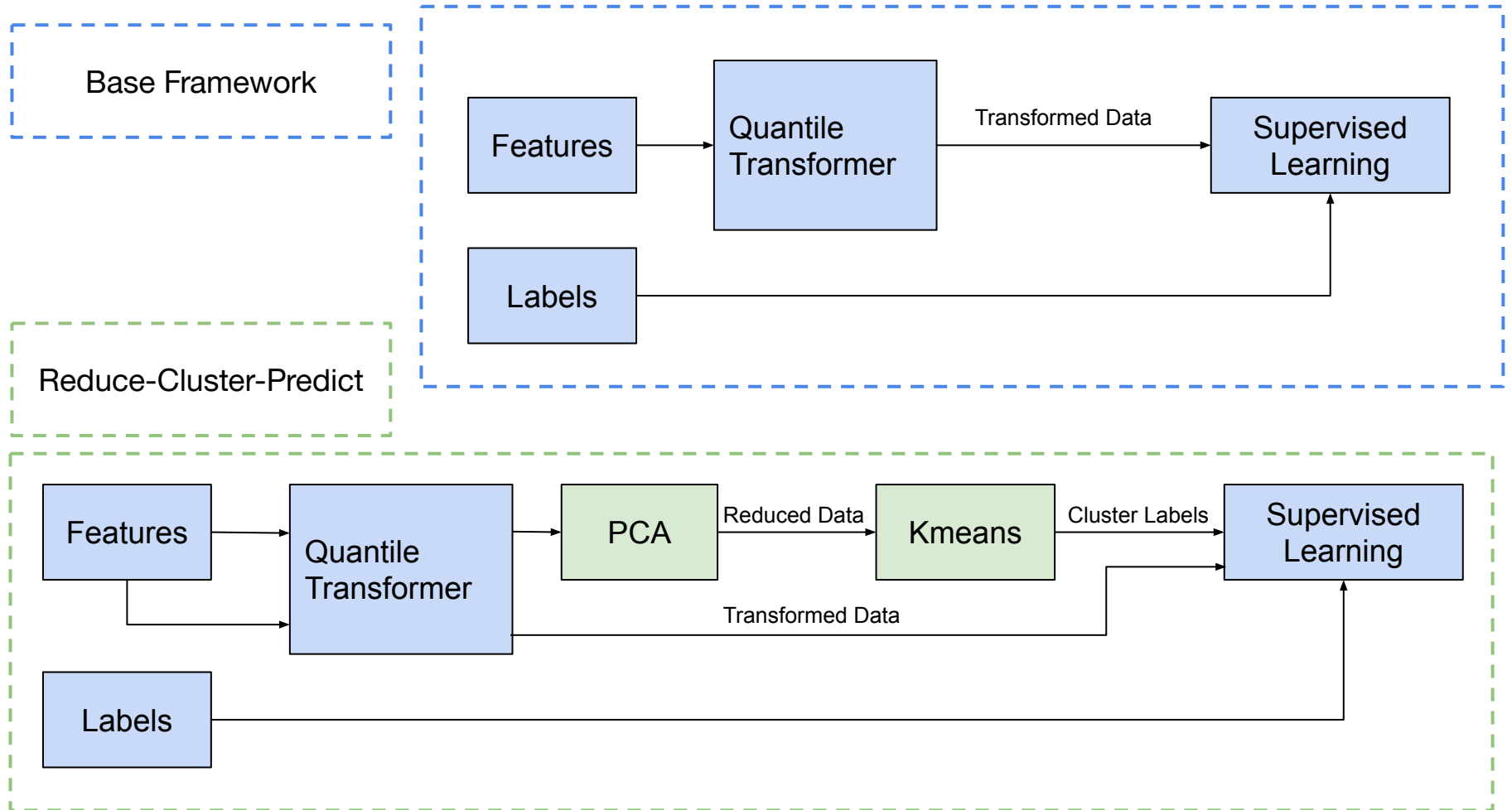
Relative Performance (%)	$[-100,-33)$	$[-33,-10)$	$[-10,10]$	$(10,33]$	$(33,\text{inf}]$
Y Label	-2	-1	0	1	2

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

A low-angle photograph of a tall, ornate building with multiple balconies and a flock of birds flying in the sky above it. The building features classical architectural details, including decorative moldings and wrought-iron railings on the balconies. The sky is a pale, overcast grey, and a large group of birds is captured in flight, scattered across the upper half of the frame. The text "In Closing" is centered over the image in a large, white, sans-serif font.

In Closing