

Final Presentation

# Has Bankruptcy Become a Crisis in Poland?

*Can We Predict if a Company Will File For Bankruptcy*

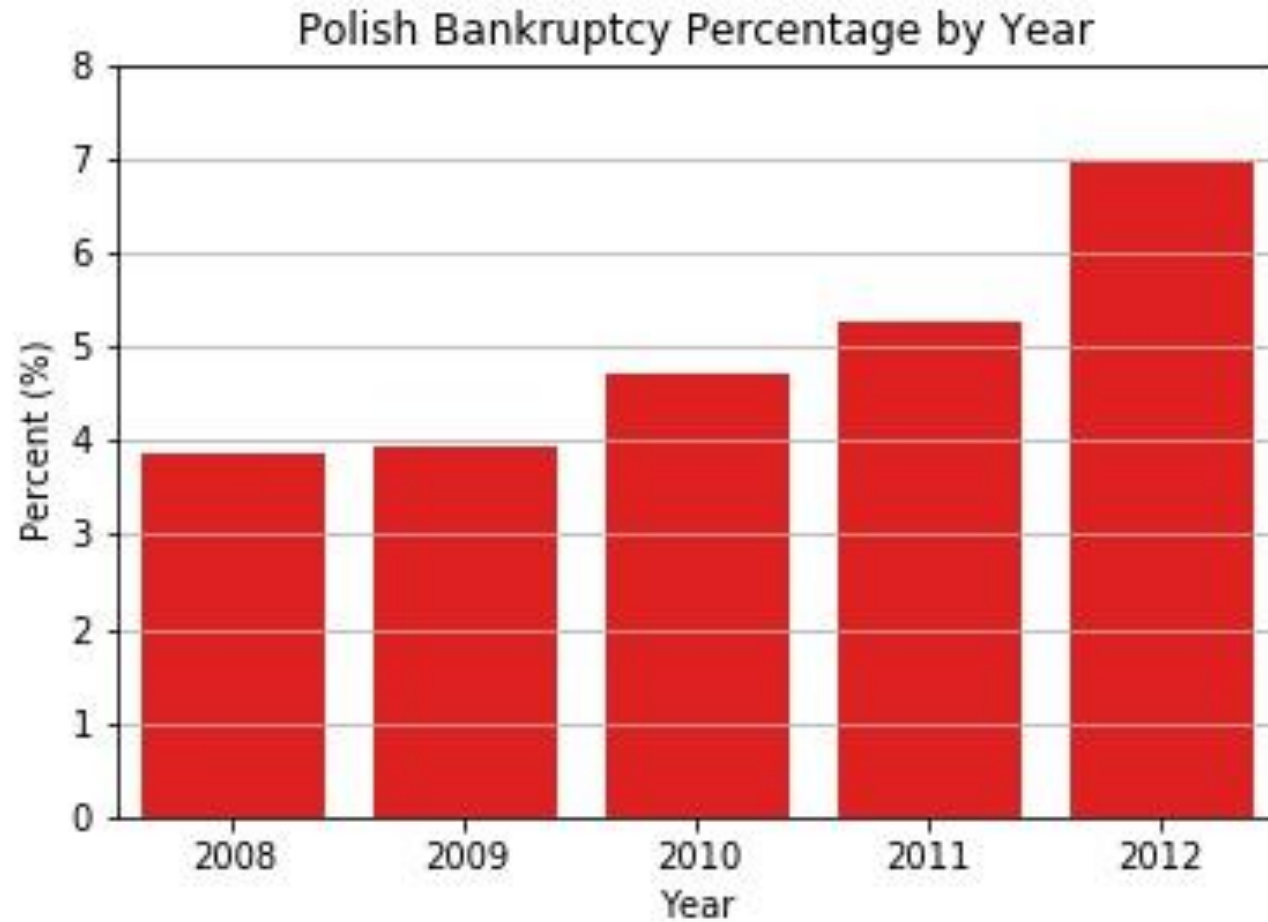
By

Gerald DeJean

# Questions to Consider

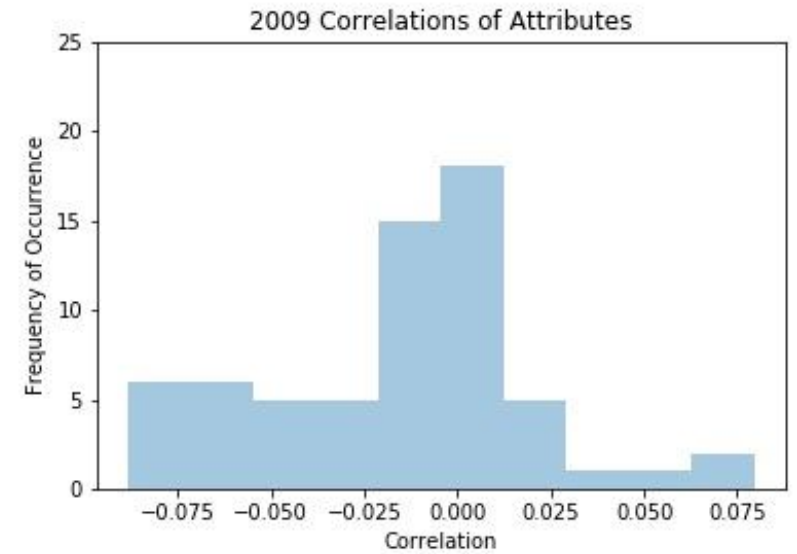
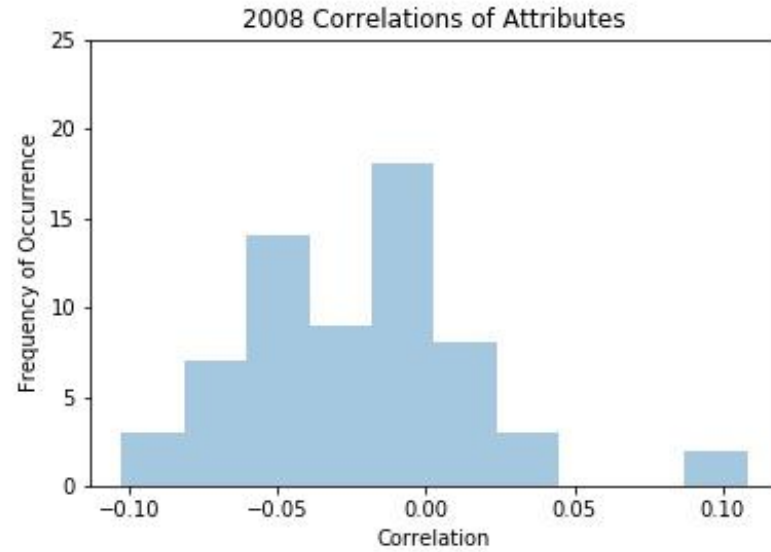
- Is it true that the bankruptcy rate in Poland has increased over the five-year period from 2008-2012?
- If so, what factors may be causing the increase in the bankruptcy rate in Poland?
- Could accurate reporting of financial records provide significant information to help some businesses from not declaring bankruptcy?
- Can we predict if a company will file for bankruptcy

# 1. Is There an Increase In Bankruptcy?

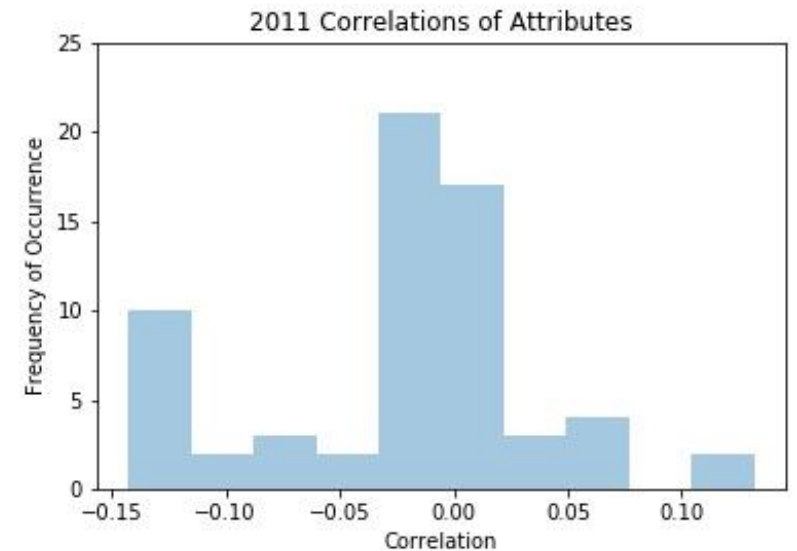
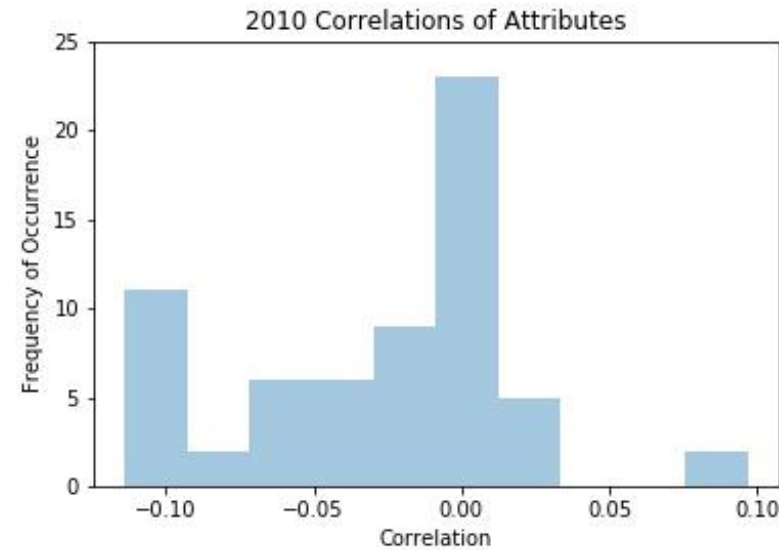


Early signs may point to an increase in bankruptcy of Polish companies from 2008-2012.

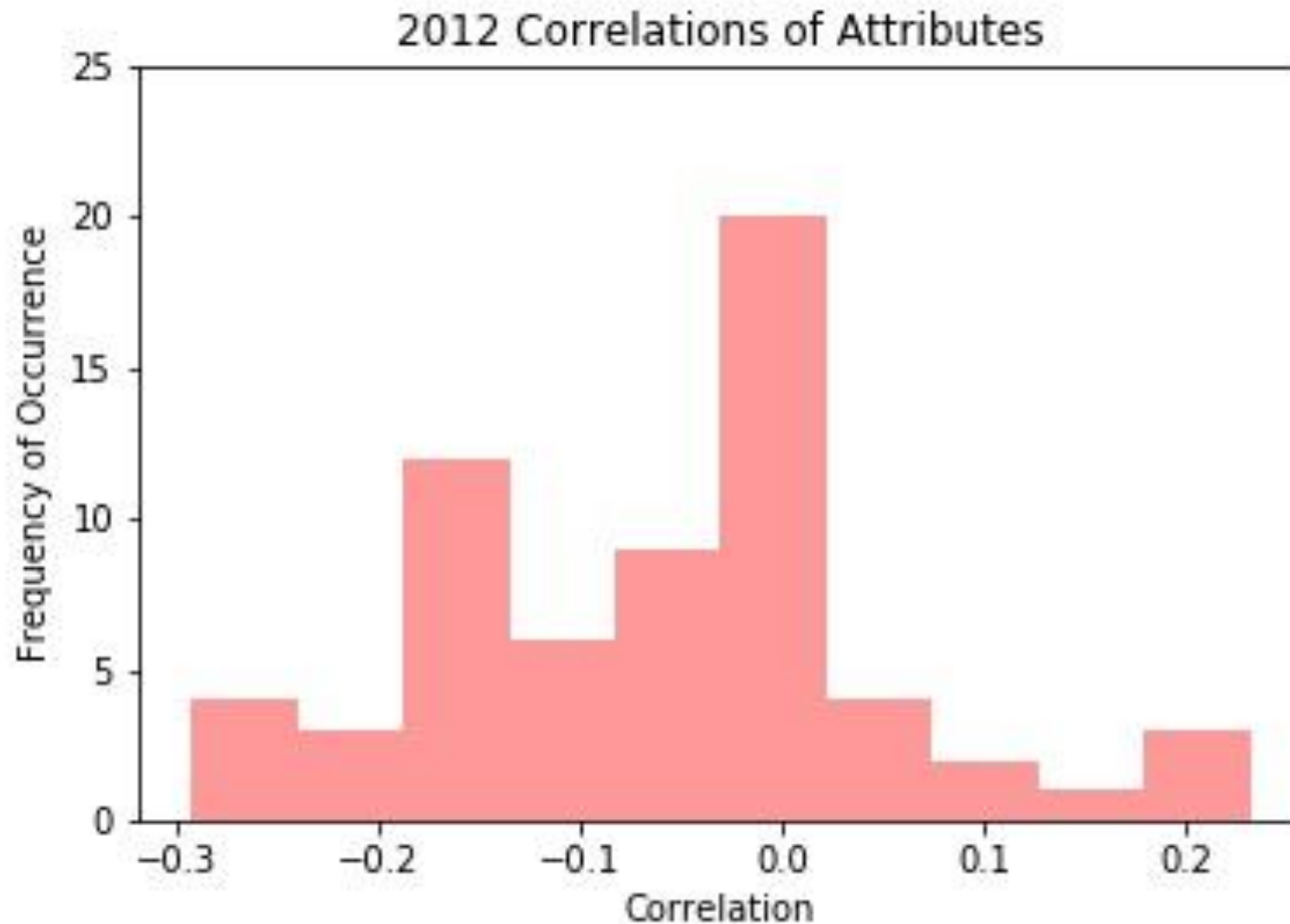
- Correlation Distribution of how a company's bankruptcy filing decision is correlated with the reported financial attributes



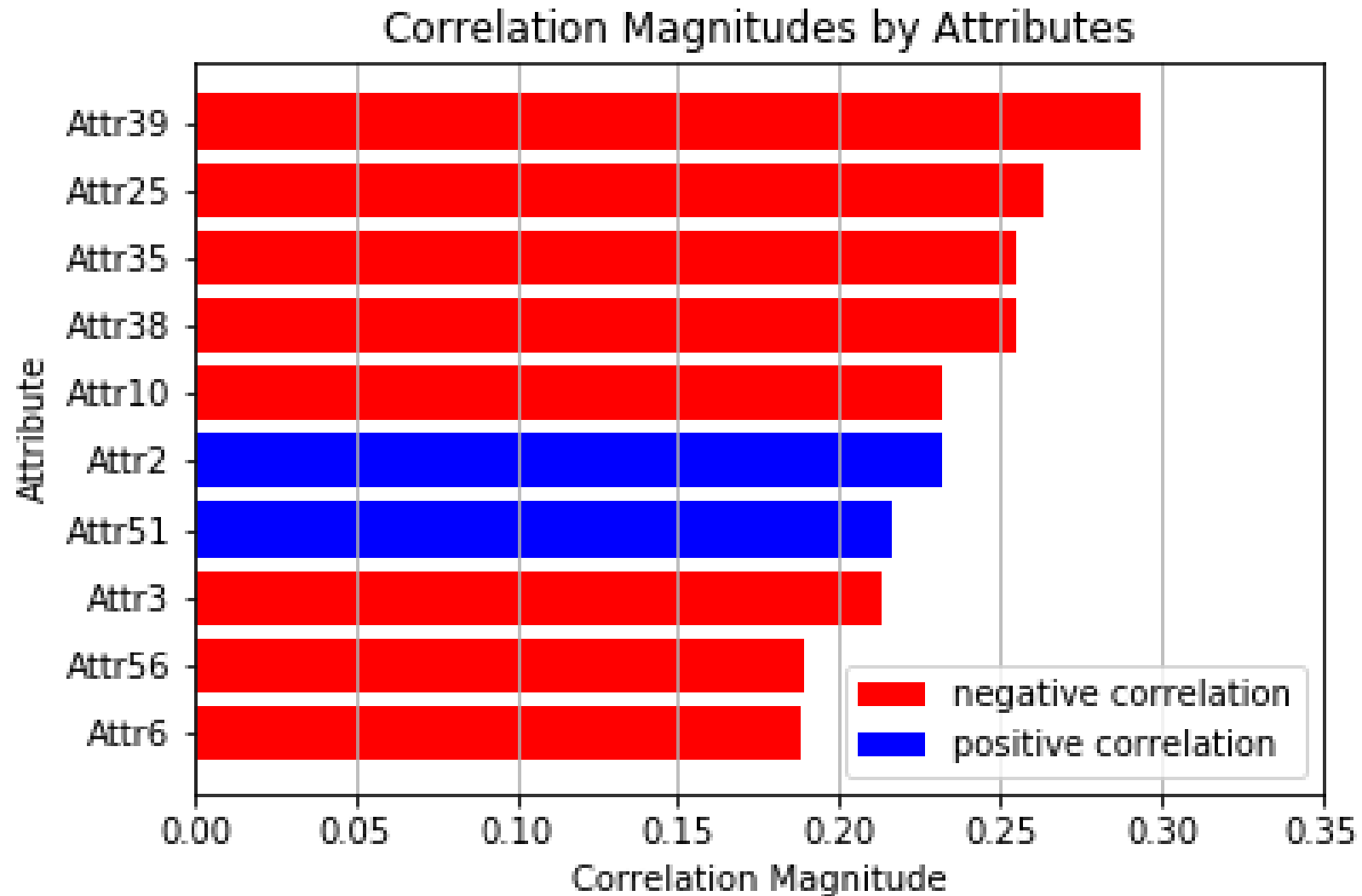
- Tails of distribution show relatively low correlations (2008-2011)



The larger magnitude of tails of the 2012 distribution shows that there may be some factor(s) contributing to a company's bankruptcy decision.



## 2a. What Factors May Be Contributing to This Increase?



## Definitions of the Top Ten Reported Financial Attributes

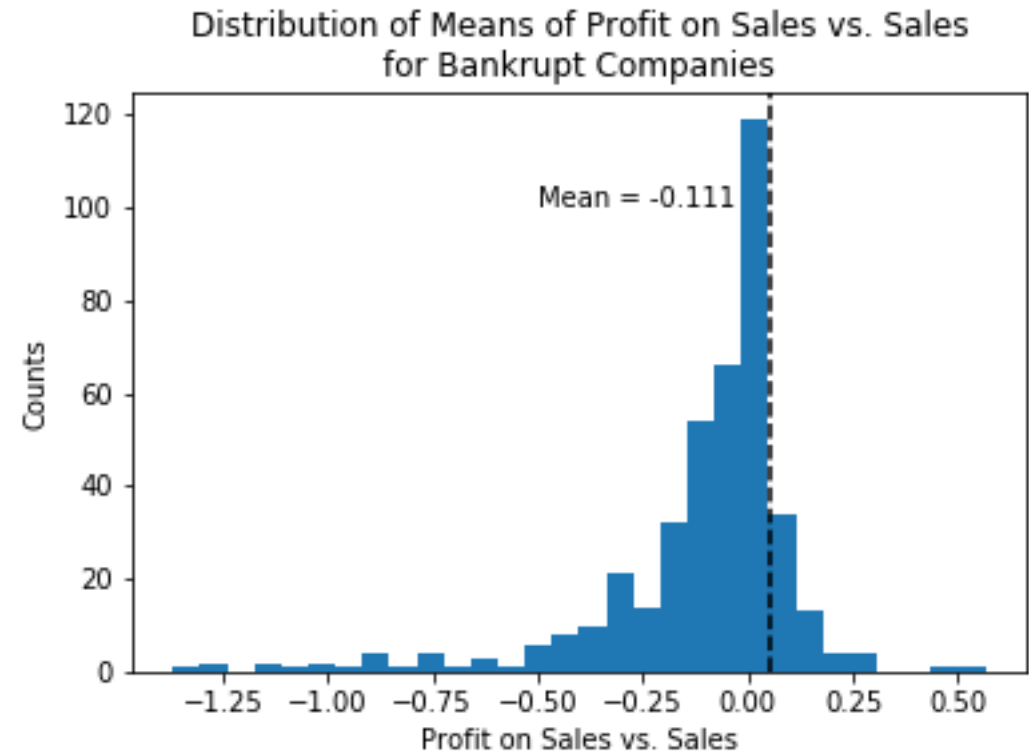
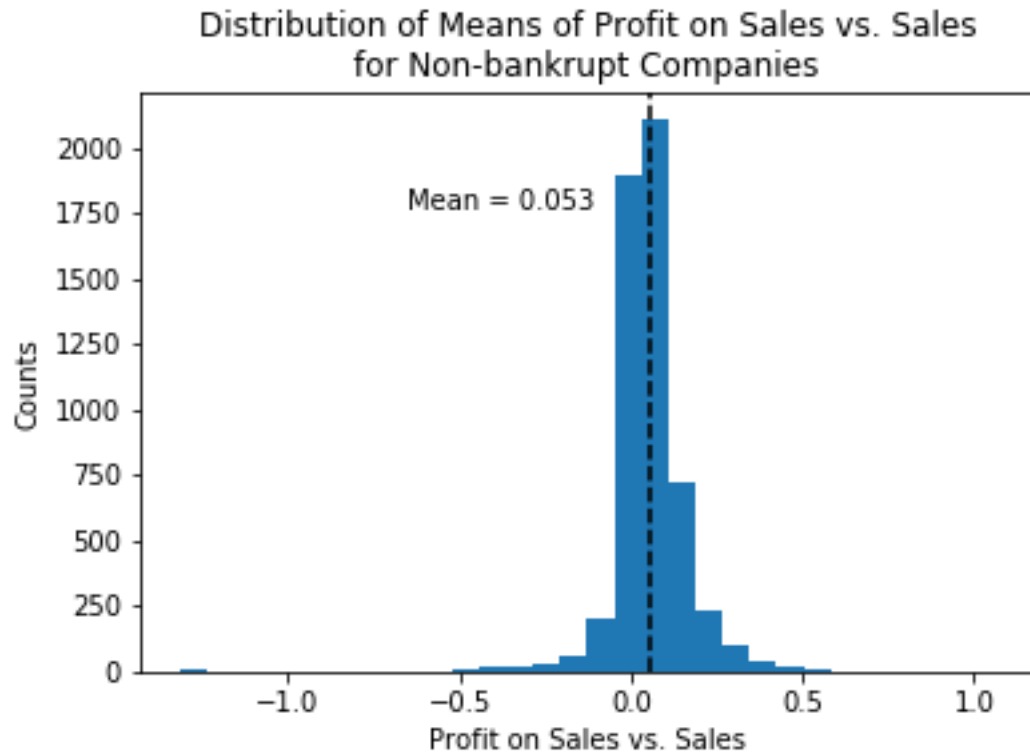
Index	Definition
Attr39	$\text{profit on sales} / \text{sales}$
Attr25	$(\text{equity} - \text{share capital}) / \text{total assets}$
Attr35	$\text{profit on sales} / \text{total assets}$
Attr38	$\text{constant capital} / \text{total assets}$
Attr10	$\text{equity} / \text{total assets}$
Attr2	$\text{total liabilities} / \text{total assets}$
Attr51	$\text{short-term liabilities} / \text{total assets}$
Attr3	$\text{working capital} / \text{total assets}$
Attr56	$(\text{sales} - \text{cost of products sold}) / \text{sales}$
Attr6	$\text{retained earnings} / \text{total assets}$

80% are inversely proportional to total assets, while the other 20% are inversely proportional to sales.

Thus, total assets and sales may be important factors to predicting bankruptcy decisions.

## 2b. Further Analysis of Attr39 (Profit on Sales / Sales)

An initial investigation of this attribute when split into two samples (bankruptcy and no bankruptcy) show a difference in means of 0.164.



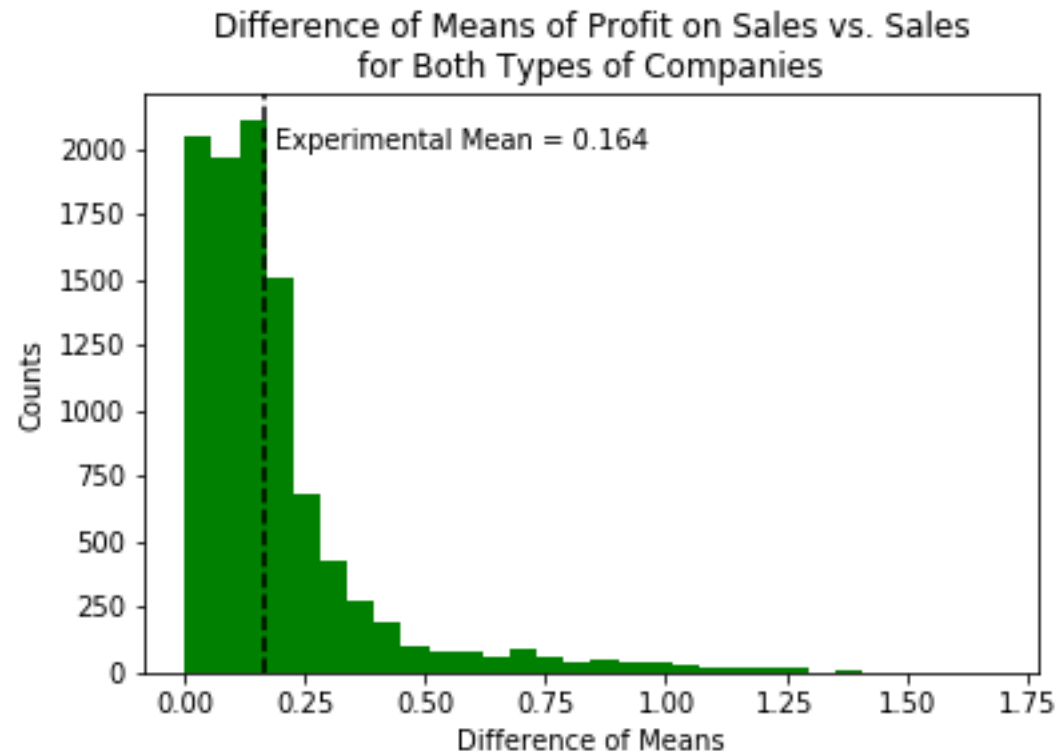
**BUT** are the true means really different?



Null Hypothesis (H0): the means are different

Alternative Hypothesis (H1): the means are the same

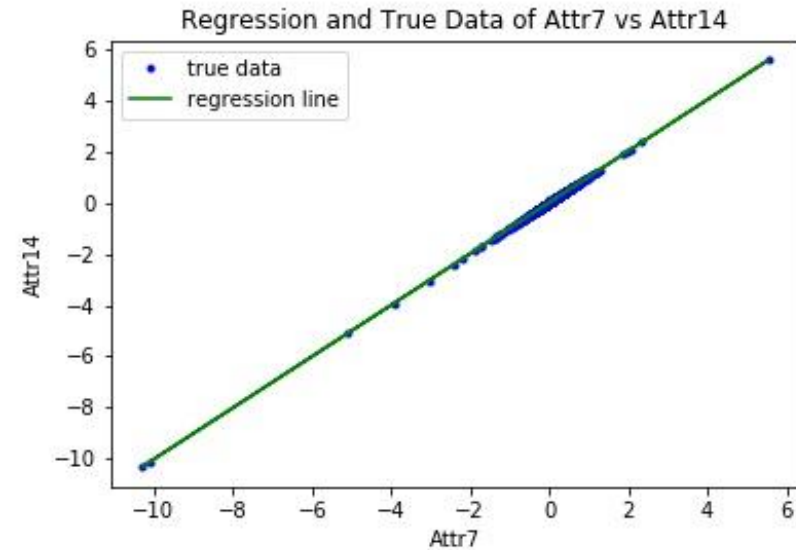
Bootstrap inference method was used to get 10,000 samples of the difference of means of bootstrapped Attr39 bankrupt and non-bankrupt samples.



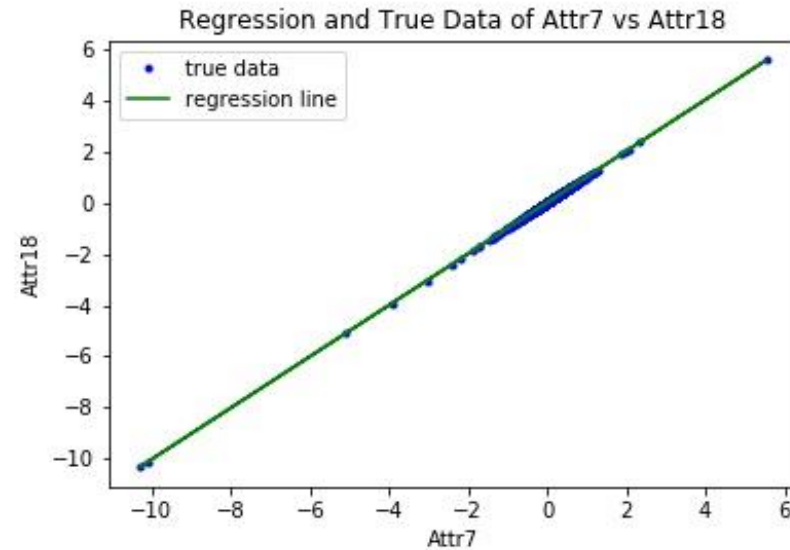
- **p-value:** 0.4
- **Implication:** 40% chance that we would see a mean as extreme as 0.164; thus, we must accept the null that the means are different.

### 3. Is Data Being Reported Accurately?

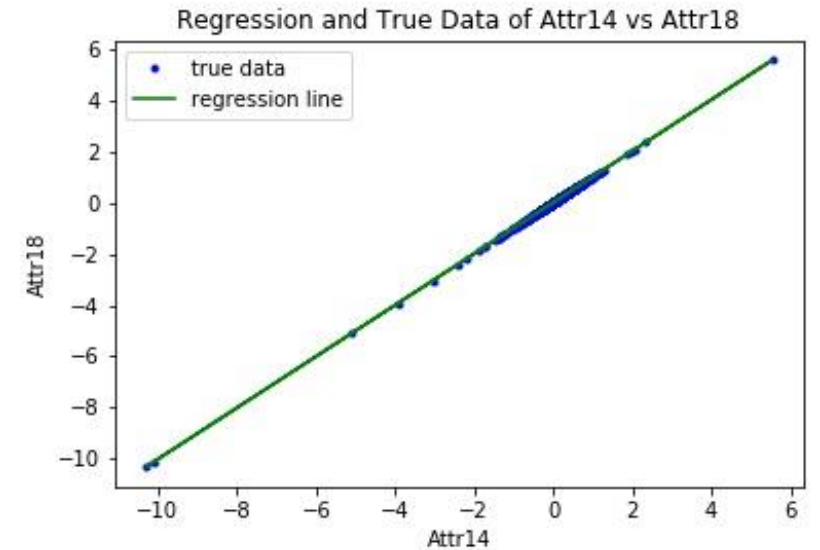
Attr7, Attr14, Attr18 seem to have the same data. 5901 of 5910 entries have the same data.



slope = 0.9999847768361899  
intercept = 2.623950924899314e-05



slope = 0.9999945803259087  
intercept = 9.341657898464208e-06



slope = 0.9999822669729542  
intercept = -1.525488991409506e-05

- Slopes  $\approx 1$ , Y-Intercepts  $\approx 0$  seem to provide support to above statement.
- What does this suggest?

Index	Definition
Attr7	EBIT / total assets
Attr14	(gross profit + interest) / total assets
Attr18	gross profit / total assets

**EBIT = Gross profit – Operating expenses – SG&A – Depreciation(Amortization)**

- If Attr7, Attr14, and Attr18 are the same for almost every instance, then it is possible that companies are not accurately reporting interest, operating expenses, SG&A, and depreciation (and/or amortization).
- This could significantly impact the decision of whether a company files for bankruptcy.

## 4. Can We Accurately Predict If a Company Will File For Bankruptcy?

- 3 machine learning algorithms are used to create models for classification
  - Standard Logistic Regression
  - XGBoost
  - LightGBM
- The 64 features were reduced to 31 based on attributes that have a VIF < 5.

### Confusion Matrices

Standard Logistic Regression

		Predicted	
		0	1
Actual	0	1631	14
	1	116	12

XGBoost

		Predicted	
		0	1
Actual	0	1626	19
	1	86	42

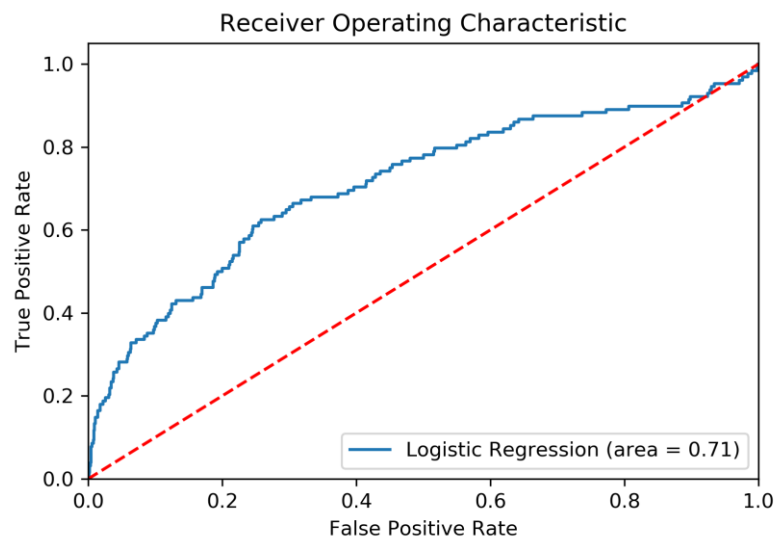
LightGBM

		Predicted	
		0	1
Actual	0	1627	18
	1	75	53

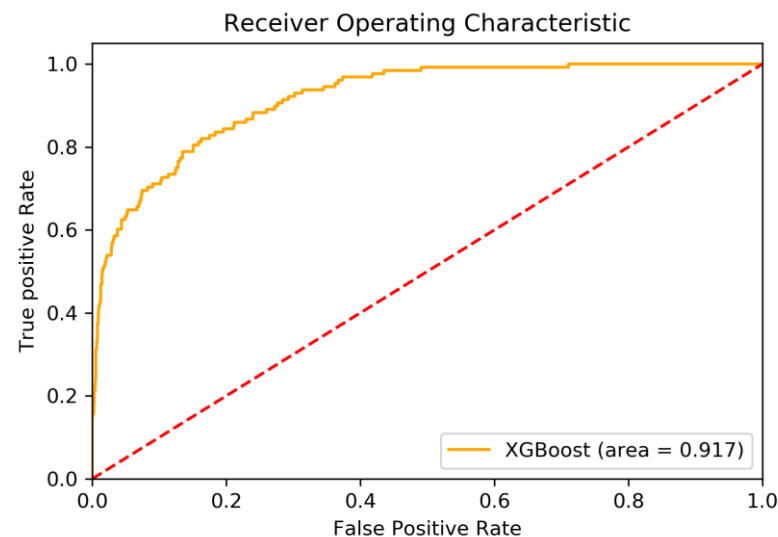
XGBoost and LightGBM perform much better at predicting companies that filed for bankruptcy.

	Standard Regression	XGBoost	LightGBM
Average Accuracy of Prediction	0.93	0.94	0.95
Precision (Bankruptcy = 0 or False)	0.93	0.95	0.96
Precision (Bankruptcy = 1 or True)	0.46	0.69	0.75
Recall (Bankruptcy = 0 or False)	0.99	0.99	0.99
Recall (Bankruptcy = 1 or True)	0.09	0.33	0.41
Misclassification Rate	0.07	0.06	0.05
Area Under the Curve (AUC)	0.71	0.92	0.92

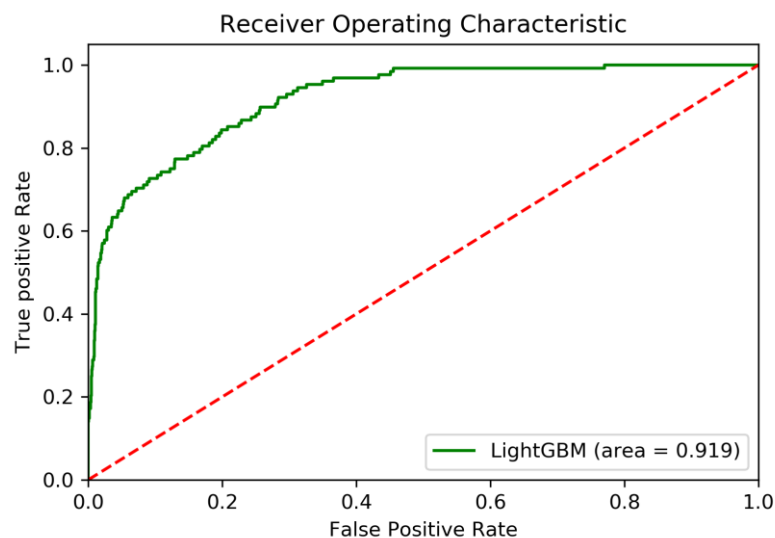
## Standard Logistic Regression



## XGBoost



## LightGBM



- The ROC curve shows that the accurate of prediction is better across different thresholds for XGBoost and LightGBM.
- Here, the area is the AUC, where the AUC is around 0.92 for XGBoost and LightGBM as opposed to 0.71 for the standard logistic regression.

- Scaling the features improved the standard logistic regression AUC significantly.

	Unscaled			Scaled		
	Standard Regression	XGBoost	LightGBM	Standard Regression	XGBoost	LightGBM
Average Accuracy of Prediction	0.93	0.94	0.95	0.92	0.94	0.95
Precision (Bankruptcy = 0 or False)	0.93	0.95	0.96	0.94	0.95	0.95
Precision (Bankruptcy = 1 or True)	0.46	0.69	0.75	0.44	0.75	0.77
Recall (Bankruptcy = 0 or False)	0.99	0.99	0.99	0.99	0.99	0.99
Recall (Bankruptcy = 1 or True)	0.09	0.33	0.41	0.14	0.34	0.39
Misclassification Rate	0.07	0.06	0.05	0.08	0.06	0.05
Area Under the Curve (AUC)	0.71	0.92	0.92	0.85	0.92	0.93

- The logistic regression statistics were taken for the 31 features.
- Using a 95% confidence interval, the 22 features that were statistically insignificant were removed, and new models were created.

	31-feature unscaled			9-feature unscaled		
	Standard Regression	XGBoost	LightGBM	Standard Regression	XGBoost	LightGBM
Average Accuracy of Prediction	0.93	0.94	0.95	0.92	0.94	0.94
Precision (Bankruptcy = 0 or False)	0.93	0.95	0.96	0.93	0.94	0.95
Precision (Bankruptcy = 1 or True)	0.46	0.69	0.75	0.38	0.69	0.67
Recall (Bankruptcy = 0 or False)	0.99	0.99	0.99	0.99	0.99	0.99
Recall (Bankruptcy = 1 or True)	0.09	0.33	0.41	0.09	0.24	0.27
Misclassification Rate	0.07	0.06	0.05	0.08	0.06	0.06
Area Under the Curve (AUC)	0.71	0.92	0.92	0.81	0.88	0.89



Index	expression
Attr22	profit on operating activities / total assets
Attr24	gross profit (in 3 years) / total assets
Attr25	(equity - share capital) / total assets
Attr3	working capital / total assets
Attr48	EBITDA (profit on operating activities - depreciation) / total assets
Attr49	EBITDA (profit on operating activities - depreciation) / sales
Attr55	working capital
Attr56	(sales - cost of products sold) / sales
Attr62	(short-term liabilities *365) / sales

- Looking at the 9 statistically significant features in the reduced model, all but working capital include an inversely proportional relationship between total assets and sales.
- This highlights an earlier conjecture about the importance of this relationship in predicting a company's bankruptcy decision.

# Conclusions

- There is an increase in the bankruptcy rate in Poland from 2008 to 2012.
- The inversely proportional relationship of total assets and sales may be important factors that influence the increase.
- Machine learning algorithms accurately predicted the bankruptcy decision of 92% of the companies in the test set.
- Accurate reporting of financial data could be significant in determining if a company files for bankruptcy.