Stock Data and Google Trends Correlation Assessment

TEXAS A&M UNIVERSITY SAN ANTONIO

Christopher Gutierrez
Faculty Sponsor: Dr. Izzat Alsmadi

Introduction

- Analyzing data from the stock exchange is not a relatively new idea, however the methods of by which that data is analyzed is something that is developing with the advent of big data and data mining companies collecting data from users.
- An Important tool for discovering changes on stock prices can come from searches for particular companies, and with this it would be an important piece of information to analyze a trend for a companies performance in the stock exchange.
- In this project we attempt to determine whether or not we can find significant correlation for a companies stock closing price and a stocks company as it is searched.
- The goal of this project is an extension of Selene Yue Xu's stock analysis for "Stock Price Forecasting Using Information from Yahoo Finance and Google Trend" However instead of collecting data for one stock data we collect data from a list that were part of the fortune 500 list.

Hypothesis

Online Google Search Results for a particular Stock name can Correlate with the stock prices closing price.

Method

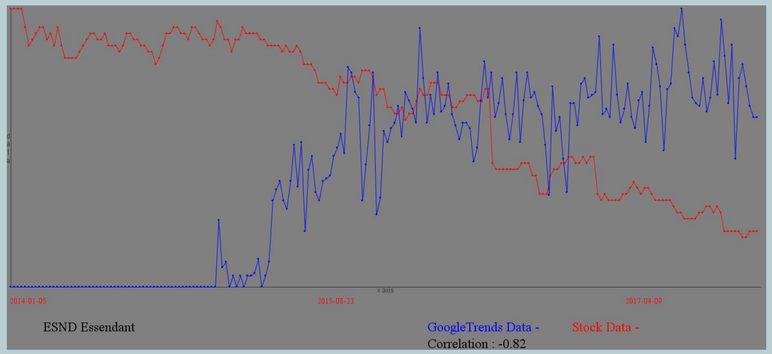
Tools:

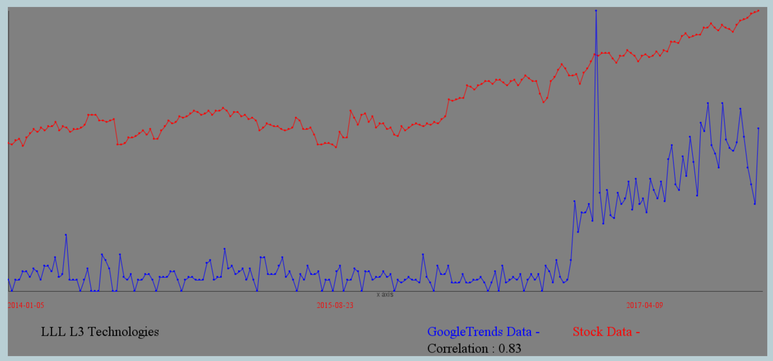
Project Utilizes the Google Trends API, and the Yahoo Finance API for collecting data. Programs and Languages for this project include R, java , and excel to collect and store data into CSV files.

Procedure:

Create a Program to pull data from the list of ticker symbols. Collect Stock Names, weekly historical data, and dates to verify the results. Use R to query search result information with the stock name query and store data. Write a Java program to graph data, and implement Pearson's r algorithm to find the correlation for stock and search result hits.

Results





Results, cont.

Stock nam	ticker	correlation
Essendant	ESND	-0.82
Comcast	CMCSA	-0.82
Facebook	FB	-0.81
Charter Cc	CHTR	-0.78
Assurant	AIZ	-0.77
eBay	EBAY	-0.76
JetBlue Air	JBLU	-0.73
Masco	MAS	-0.73
Time Warr	TWX	-0.73
Microsoft	MSFT	-0.71
NiSource	NI	-0.69
CDW	CDW	-0.66
Fisher Scie	TMO	-0.66
Harley-Da	HOG	-0.64
Southwest	LUV	-0.63
TravelCen	TA	-0.62
Lam	LRCX	-0.61
Dominion	D	0.59
XPO Logis	XPO	0.59
Kindred H	KND	0.6
McDonald	MCD	0.61
Sherwin-V	SHW	0.61
Tenet Heal	THC	0.64
Clorox	CLX	0.65
Genesis H	GEN	0.65
Costco Wh	COST	0.66
Republic S	RSG	0.66
Centene	CNC	0.68
Dollar Tre	DLTR	0.68
LKQ	LKQ	0.69
L3 Techno	LLL	0.83

Discussion

423 companies were studied in this project, and of those companies we found that 13.9% of the companies showed a correlation number greater than 0.5 and less than -0.5. Thus there is significance to a stocks name and how well it performs over time as the search appears over time.

This shows that for some companies the search results may play a role in a companies performance.

The data is limited for certain companies as the data was spanned over 3 years and in those 3 years, certain ticker symbols were recently created and stock prices for those are not complete.

An expansion on this project by Dr Alsmadi aims to target specific sectors for a list of words and find better correlation results for stocks within that word bank.

References

Xu, S. Y. (n.d.). *Stock Price Forecasting Using Information from Yahoo Finance and Google Trend.* Retrieved from https://www.econ.berkeley.edu/sites/default/files/Selene%20Yue%20Xu.pdf