



Analyzing relationship: Twitter tweet Frequency with the Stock prices of Telecom Companies

Amrita Shelar, Ching-yu Huang

Kean University, Union, USA {shelara, chuang}@kean.edu

Introduction

Twitter is a widely used online social media. In this paper, we investigate whether the daily number of tweets that mention any of the telecom companies i.e. **Verizon**, **T-Mobile**, AT&T and Sprint vis-à-vis stock prices. The basic premise of the study was to look for a correlation between the above for Telecom giants, using Statistical Methods: Z-score and Chi-Square - Test of Independence with data visualization. Our preliminary results also demonstrate the relation of frequency of tweets with stock prices of each day. Furthermore, it appears that Twitter tweets and stock prices are independent.

Research Steps:

- ⇒ Download dataset for each company using CRON:
 - ⇒ 14 days for tweets (Python Twitter API – Tweepy)
 - ⇒ 9 days for stock prices (Yahoo! stock market)
- ⇒ Schema design
- ⇒ ETL (Extract, Transform, and Load)
- ⇒ Application of Statistical Methods/Analysis
- ⇒ Interpretation and Data Visualization

Methods

Three Tier Architecture

- Presentation tier: Web Browser** ⇒ UI (HTML, CSS, JS)
- Application tier: Web Server** ⇒ PHP (JPGraph), Python (Tweepy)
- Data tier: Database Server** ⇒ DB (MySQL), JSON (Tweets)

Correlation:

Pearson correlation coefficient is used to measure the relationship between twitter tweets and stock prices.

Z-score:

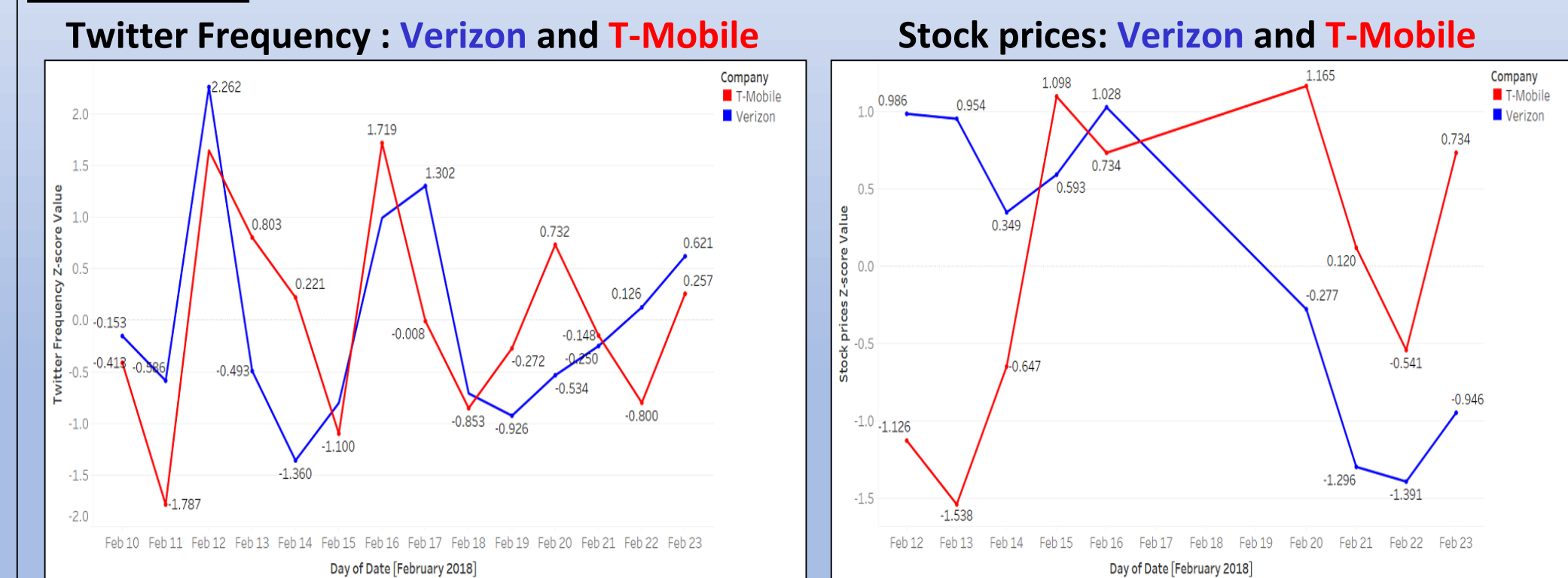
Z-score is used to calculate how many standard deviations from the mean a tweet count and stock price is.

Chi-Square – Test of Independence:

Chi-square test for independence is implemented in a 2 x 2 contingency table to see if the two parameters are related.

Results

Z-Score



- ⇒ All Z-scores for **T-Mobile** are higher than **Verizon** on 02/13 & 02/20 due to it's 'T-Mobile Tuesday' campaigns
- ⇒ For **T-Mobile**, a sudden spike(from 54 to 249) is observed on 02/12 due to it's 'Solar energy initiative' campaign
- ⇒ On 02/12/2018, **Verizon** tweets were 2.262 standard deviations above the mean(only case)
- ⇒ Z-scores for **Verizon** and **T-Mobile** are between -2 and +2, hence none of them are very far from the mean and are consistent across the data set
- ⇒ For **T-Mobile**, since Z-score is positive for 5 days out of 9; **Verizon** tweets were more than average for 55.5% of the days

“T-Mobile tweets are more variable than Verizon and depend on campaigns to raise popularity”

Chi Square Test

Hypothesis: Assuming that there is no association between twitter frequency and stock price

Verizon: Twitter Frequency vs. Stock Price

Verizon	High Stock Value	Lower Stock Value	Marginal Row Totals
High Frequency of Tweets	2 (2.78)	3 (2.22)	5
High Frequency of Tweets	3 (2.22)	1 (1.78)	4
Marginal Column Totals	5	4	9 (Grand Total)

The chi-square statistic is **1.1025**
The p-value is .293718, not significant at p < .05

T-Mobile: Twitter Frequency vs. Stock Price

T-Mobile	High Stock Value	Lower Stock Value	Marginal Row Totals
High Frequency of Tweets	3 (3.33)	3 (2.67)	6
High Frequency of Tweets	2 (1.67)	1 (1.33)	3
Marginal Column Totals	5	4	9 (Grand Total)

The chi-square statistic is **0.025**
The p-value is .635256, not significant at p < .05

“The p-value is larger than the significance level, hence we conclude that the twitter frequency and stock price are not associated.”

Conclusions

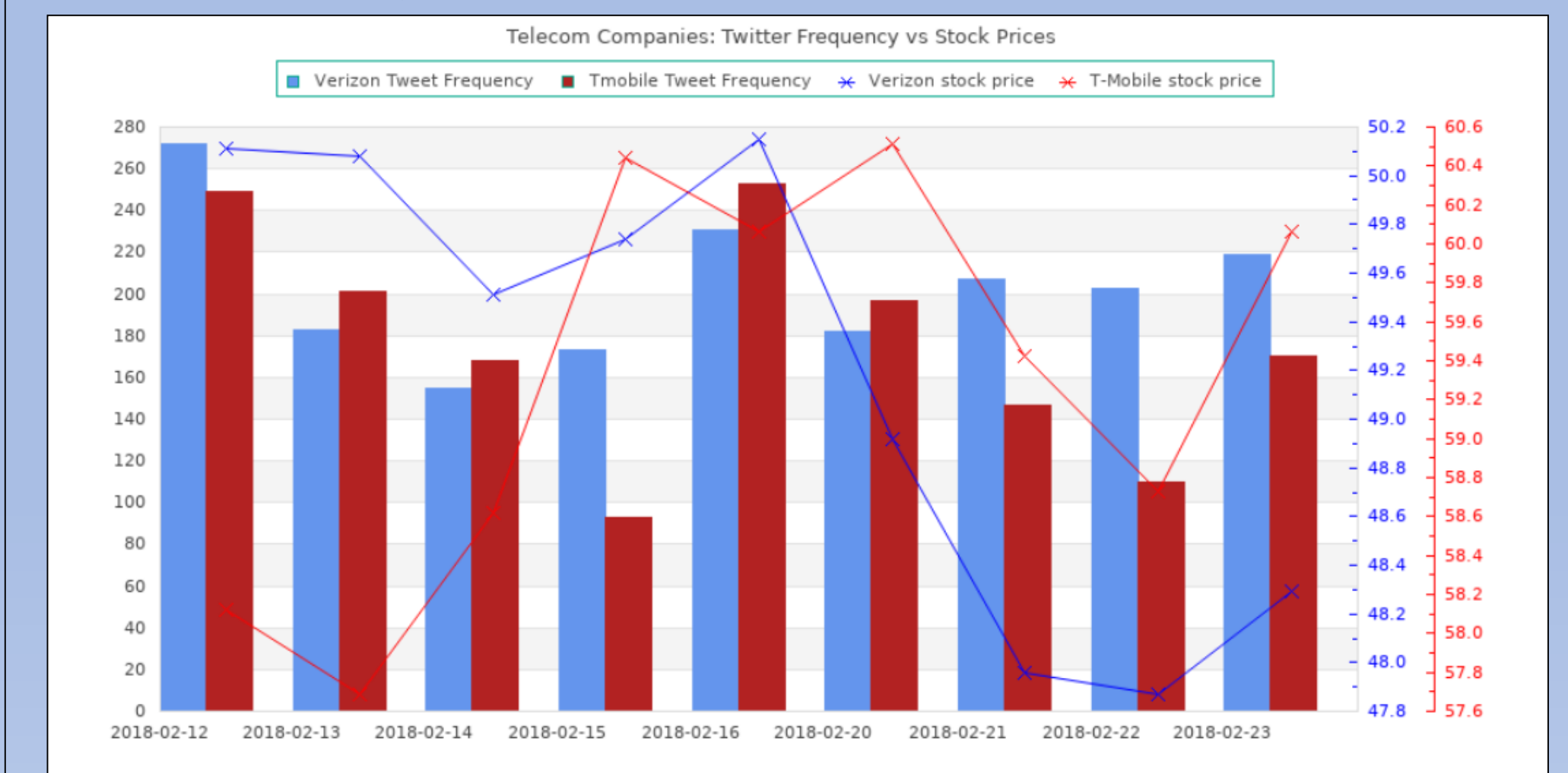
Telecom Companies: Twitter Frequency vs Stock prices

Our results demonstrate that daily number of tweets is not correlated with the stock prices for a period of 9 days.

Given the data sets analyzed for tweets and stock price:

- ⇒ **Verizon:** Correlation **-0.102** indicates **Low Positive Correlation**
- ⇒ **T-Mobile:** Correlation **-0.221** indicates **Low Negative Correlation**

Scale: X-axis-Date, Y-axis- Frequency of tweets for Verizon and T-Mobile (numbers), Y1-axis- Stock price for Verizon (billion dollars), Y2-axis- Stock price for T-Mobile (billion dollars)



“No Correlation between tweet mentions and stock prices”

Experience & Roadmap

- ⇒ Twitter allows 140 characters tweet – fewer keywords captured
- ⇒ Only two parameters included in research
- ⇒ Lots of dirty data observed while experiment
- ⇒ Roadmap:
 - ⇒ Conduct research for longer duration
 - ⇒ Include an additional dimension in the data sets
 - ⇒ Consider AT&T and Sprint data sets
 - ⇒ Performing sentimental analysis over tweets to confine tweets to telecom

References

- [1] Andrew Huang, "Data Mining Twitter to Model Stock Price Movement", *Proceedings of 2016 Multidisciplinary International Social Networking Conference*, Union, New Jersey, August 15-17, 2016.
- [2] Adil Moujahid. "An Introduction to Text Mining Using Twitter Streaming API and Python," *Adilmoujahid.com*, July 2014.
- [3] "Yahoo! Finance", *Yahoo*, <http://finance.yahoo.com/>
- [4] "Statistics Standard Score", *Laerd.com*, <https://statistics.laerd.com/statistical-guides/standard-score-2.php>
- [5] "Chi-Square Test of Independence", *Statistics Solutions*, <https://www.statisticssolutions.com/non-parametric-analysis-chi-square/>