

Correlation and Predictive Analysis of Twitter data and Spending Activity in the UK

Ding Jin Wong¹, Aidan Slingsby¹ and Julian Flavin²

¹City, University of London, ²Barclays

Project Background

Whether it's for social networking or content sharing, the power and influence of social media have expanded and impacted every industry. However, utilising these data in businesses can be tricky but extremely rewarding. Correlation and sales prediction can help companies capitalise on social media for marketing strategies and influence future sales. This paper aims to find how Twitter can improve current states of analytic models to generate deeper market and customer insights. Five restaurants all offering a very similar type of cuisine were chosen to conduct a proof of concept.

Data and Methods

Two datasets were used: Twitter and transaction data. Twitter data was obtained through Brandwatch and Twitter API. Owned (company's tweets) and Earned tweets (every other tweet mentioning the company) were separated with Twitter API to study impact of owned tweets. Sentiment and mood variables were also generated as part of Brandwatch's data. Transactional datasets were queried in a controlled environment where factors such as branch openings or closings, were kept constant. This allowed a more reliable study to understand Twitter's impact on sales. Both datasets were grouped together on a weekly basis, ranging between June 2017 and June 2019. Two additional datasets were created:

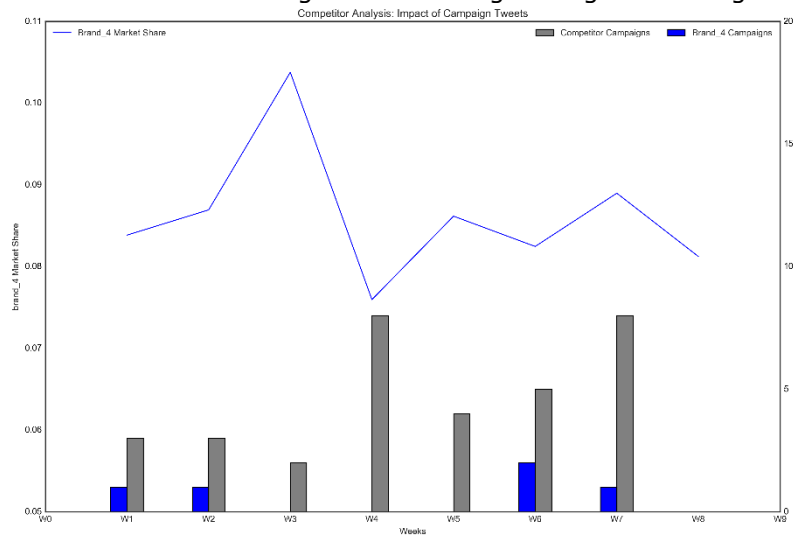
1. Transactions from customers of age 16-24
2. Sales aggregated on quarterly basis

Raw values were converted to market share for competitive analysis. Time series properties were considered to account for stationarity. To find correlation, Pearson, Spearman, Kendall, Distance and Information Criterion were utilised. Each correlation was put through Hypothesis test, thresholding and Granger-causality test to identify the variables of highest correlation. To predict sales, ARIMA and ARIMAX models were implemented and compared to determine if Twitter had any predictive power in each case. Impact of owned tweets was studied to identify how companies should manage their Twitter account to boost Twitter activities that drive sales. For a competitive analysis, a case study on one brand

against all its competitors measured the extent of Twitter-based marketing / promotions.

Key Findings

Correlation combined with statistical tests identified 13 pairs of Twitter and transaction variables that contained high correlation. Tweets that sparked 'joy' were found to be significant against market sales in two brands but no industry trends were found. Stationarity proved to be important in improving ARIMA where the largest negative change in RMSE was 11.92%



when Twitter variables were added. In general, 80% of models where stationarity was achieved on all variables, found a decrease in RMSE with Twitter variables. For owned tweets, retweet count and highly-favourited tweet count coincided with peaks of Twitter variables that were highly correlated with sales. Discounts

and competition-related tweets were commonly found in these tweets. In competitive analysis, brand 4's fluctuations and promotions were investigated against its competitors on a smaller scale. Promotions were found to impact market sales the week after its posting. It also identified the extent of each promotion's influence on market sales. The case study is visualised below with tweet notes removed.

Value of the Research

Both ARIMA and event-based case study contributed to the potential of Twitter on sales prediction. ARIMAX with Twitter improved long term predictions, allowing clients to obtain a better idea on future sales patterns. Furthermore, the case study demonstrated the short-term impact of tweets within the same week and especially the next. This allows companies to reassess their risk management techniques, Twitter marketing strategies and cash flow projections to prepare for the following week's rise or fall. Future studies should involve daily trends to understand how the impact of a promotion diminishes over time, which could be highly beneficial in a competitive market. Steps taken here can be replicated in other brands or industries.