

AUTOMOBILE SALES FORECAST USING SOCIAL LISTENING AND CONDITIONS TO FORECAST SALES

Hisaki Goto¹ and Yukiko Goto²

¹*Nagoya Gakuin University, Japan*

²*Enconnect Inc., Japan*

ABSTRACT

The amount of information created by end-users has become massive; it is often referred to as an ‘information explosion’, and social media is a typical tool for gathering that information. We gathered and analyzed tweets related to automobiles in Japan from June 2018 to May 2020, attempting to create a formula to forecast future automobile sales through content posted on Twitter. We found that forecasting the sale of automobiles is possible when the number of tweets for a model is large, which indicates that social media can be beneficial as alternative data to forecast product sales. Although we could successfully make a sales forecast on certain models with a large number of tweets, all the models were not possible to forecast. However, we propose a hypothesis that forecasting sales using Twitter is possible only after a product enters the so-called ‘cash cow’ phase under the Product Portfolio Matrix developed by Boston Consulting Group, which creates business strategy using market growth rates and relative markets.

KEYWORDS

Twitter, Alternative Data, Social Listening, Automobile Sales Forecasts

1. INTRODUCTION

In our previous study we had analyzed national elections in Japan through social listening by gathering and utilizing text data posted on Twitter (Goto, 2018, Goto, 2019). 2,009,520 tweets were gathered for the election analyses (Goto, 2019), making the study uniquely valuable because other previous studies had only used several hundred thousand tweets. We found that social media is a useful tool to make predictions about elections under the proportional representation system and that it can replace conventional prediction methods such as the RDD (Random Digit Dialing) method at low cost with frequency.

As Figure 1 shows, Japan’s Twitter population is ranked as the second following the U.S. Considering the U.S. population is 2.6 times large as that of Japan (329,500,000 in the U.S. and 125,800,000 in Japan in 2020), it is easy to understand how large the proportion of Twitter population is in Japan. TPS (Tweets Per Second) numbers also show the Japanese love of Twitter: Akimoto (2011) mentions that New Year’s tweets in Japanese in 2011 made the world record of 6,939 TPS, and 143,199 TPS during the broadcast of the animation movie ‘Castle in the Sky’ on August 2, 2013 renewed the world record. Also, the majority of top 10 TPS come from Japan. In this study, we attempted to forecast automobile sales in Japan by taking advantage of the Japanese love of Twitter.

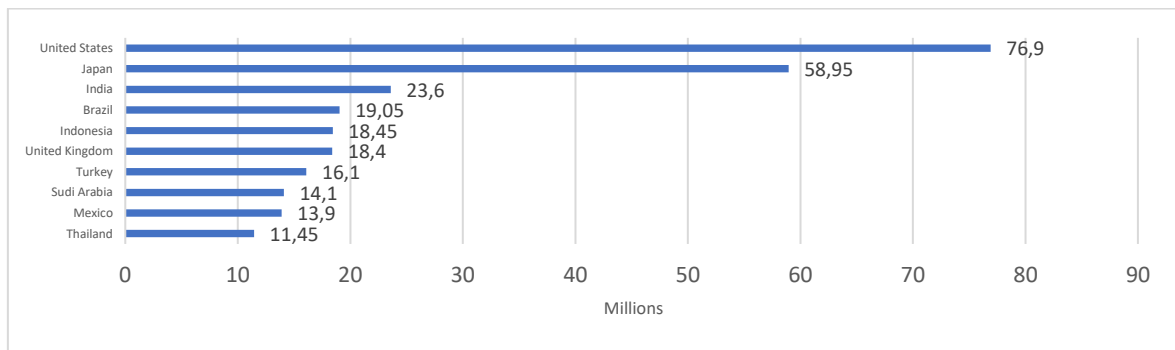


Figure 1. Leading countries based on number of Twitter users as of January 2022

Source : Statista (2022), <<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>>Retrieved on April 29, 2021

2. LITERATURE REVIEW

Previous studies that forecast product sales using social media include Dijkman, Ipeirotis & Aertsen (2015), Wijnhoven & Plant (2017) and Pai & Liu (2018). The study by Dijkman, Ipeirotis & Aertsen (2015) showed that commodities that receive less attention on social media did not show similar results as shown in books or movie tickets. This case study classified tweets into three type of tweet: the type of tweet, the type of user who posted the tweet, and the sentiment of the tweet in four countries. Their analysis suggested that positive tweets by persons may be useable to predict sales and that there was a relation between a peak in the number of positive tweets and an increase in sales in the following weeks. However, they pointed out that the relation between tweets and sales was not necessarily a causal relation but may be affected by a third factor for the sales. Wijnhoven & Plant (2017) explored the usefulness of sentiment analysis and Google Trend for automobile sales forecasts. They analyzed half a million postings related to the eleven automobile models on social media using linear regression models. Also, the study compared its results to Google Trends for predictability. The study concluded that sentiment analysis on social media showed little predictability on automobile sales, while Google Trends data and social mention volume showed meaningful results. Pai & Liu (2018) employed social media data, stock market values and time-series models to predict monthly sales of automobiles in the U.S. Sentiment analysis of tweets, stock market values and hybrid data were used for the analysis to forecast the total monthly sale of automobiles in the U.S. The study concluded that the values from social media and the stock market were essential for the monthly sales forecast and that the exclusion of seasonal variability improved the forecast of precise monthly sales. As such, most sales forecasts using Twitter have not seen success yet.

Our study gathered tweets related to 12 vehicle models sold in Japan for two years, from June 2018 to May 2020. The total number of tweets gathered was 3,247,796, quite large compared to other previous studies.

3. METHODS FOR THE STUDY

The study analyzed Twitter data of 12 automobile models and actual sales data published by Japan Automobile Dealers Association (JADA). In this chapter we discuss the methods for social listening and how the social listening data was analyzed.

3.1 Vehicle Models Subject for the Study

JADA publishes previous month's sales by models, and we selected the 25 top sold vehicle models published in May 2018. Those model names were searched by keywords on Real-Time search by Yahoo! Japan⁽¹⁾, and models that cannot be identified due to homonyms were excluded (for example, when searching for Toyota 'AQUA' ('Prius C' is the international name), non-related words such as 'aquarium' were excluded). Among 25 models, we selected twelve models after having removed model names with many homonyms (Table 1).

Table 1. Models subjected for study

Prius	VOXY	SIENTA	COROLLA
C-HR	Vitz	IMPREZA	VEZEL
ALPHARD	STEP WGN	X-TRAIL	DEMIO

Figure 2 is the sales record of the 12 models from June 2018 to May 2020 published by JADA. Toyota's models are always high on the list of the chart. Especially, Toyota Prius had been the most sold model since January 2019, but a Higashi-Ikebukuro runaway car accident⁽²⁾ on April 19, 2019 damaged the sales, and after August 2019 the sales dropped to 1/3 of its peak. On the other hand, Toyota Yaris has become the top seller since May 2020. Honda came second on the number of sales. Nissan had held the second ranking on the sales for a long time, but after the arrest of the former Nissan CEO Carlos Ghosn on charges of fraud related to the executive compensation, its sales dramatically plummeted, and although the Nissan X-TRAIL was subject to the research, the sales dropped by half during the data gathering.

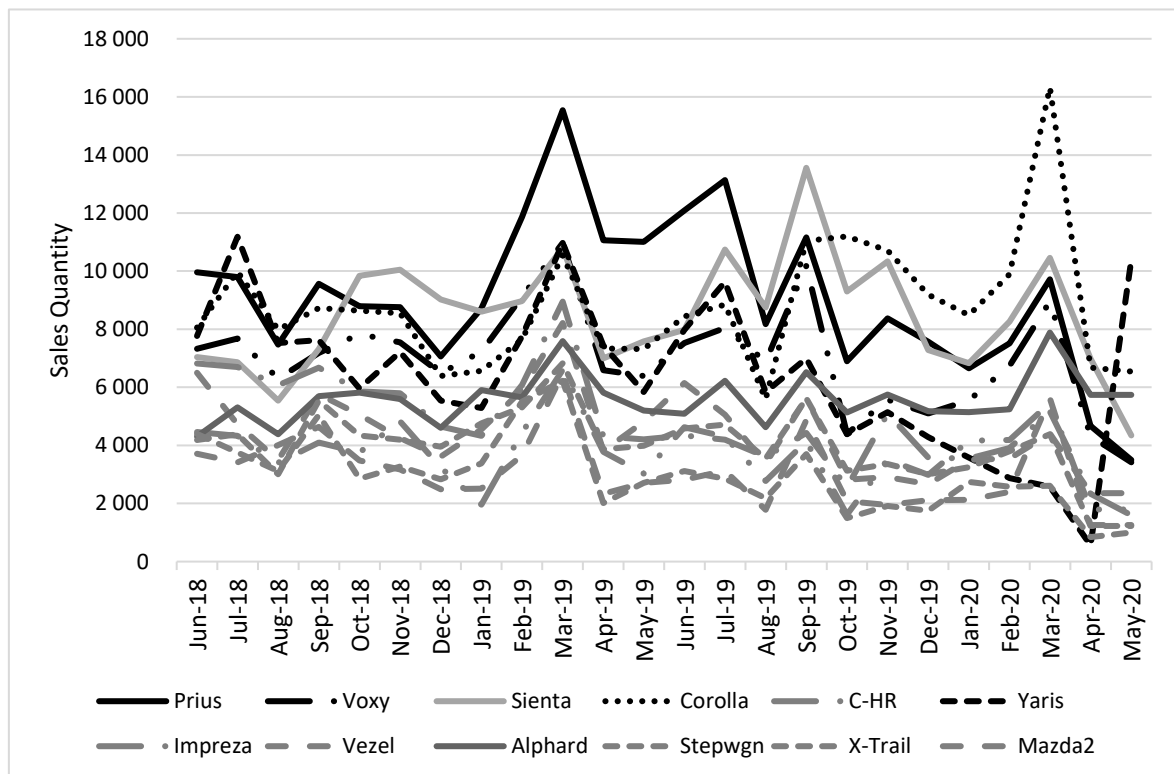


Figure 2. Sales by models

Source : JADA (2018-2020), <<http://www.jada.or.jp/data/month/m-r-hanbai/m-r-type>>
Retrieved on April 29, 2021. (Vits and Demio were renamed Yaris and Mazda 2, respectively.)

3.2 Methods for Social Listening

We obtained data on vehicle models for social listening through the real-time search provided by Yahoo! Japan. Through the real-time search, Japanese tweets that are written in Japanese and are not set in private can be searched. Therefore, tweets written overseas (for example, tweets from Honolulu, Hawaii, which has a 19 hour time-difference to Japan) are also included in the data. We believed that gathering Twitter information by key words makes the information volume bigger and more precise than gathering hash-tagged information.

Figure 3 is the number of tweets on each model. This shows only positive tweets identified through sentiment analysis, of which criteria of ‘negative’ and ‘positive’ are set by Yahoo. Also, since Japanese anime characters occasionally take after Japanese car model names, words related to anime characters had to be excluded when researching Yahoo! Japan. For example, when searching a model name ‘Sienta’, key words for search looks like this: ‘Sienta -Sientaris -Senisienta -Sienta Super Beast -Enamel Made Sienta’ (‘-’ means ‘to exclude’). The figure shows that tweets about Toyota Prius increase around every January. The reason for the increase is because Toyota announces minor changes to the model during that period. Also, as Figure 4 shows, the previously mentioned Higashi-Ikebukuro runaway car accident brought the fire on Twitter; thus, we decided to use only positive tweets for the analysis. The tweets on Toyota Prius gradually took a downhill slide since April 2019 shown in Figure 3, while tweets on Toyota Yaris showed rapid increase.

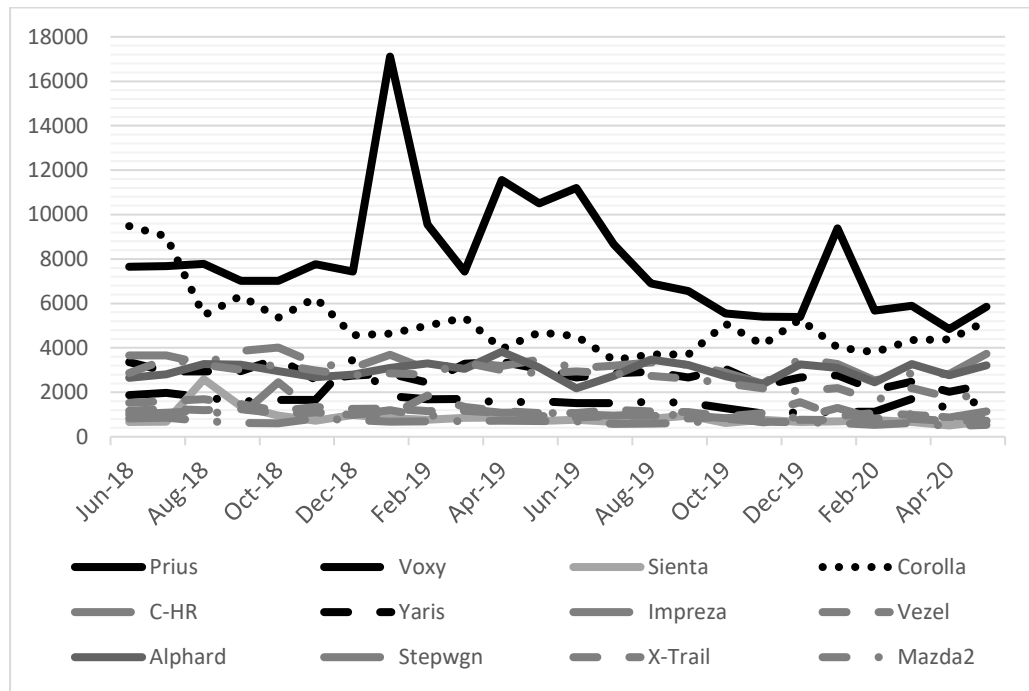


Figure 3. Positive tweets by month

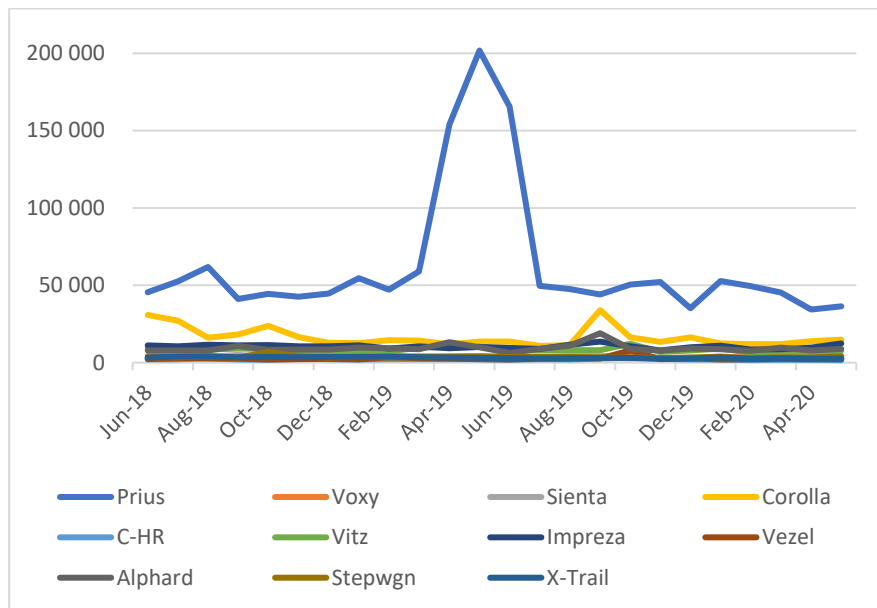


Figure 4. Tweets by month

3.3 Methods for Analysis

Because the data used for the study is time-series data, we first performed a cross correlation calculation for the models to measure time-lag between tweets and actual sales. Analysis was performed after the time-lag was adjusted.

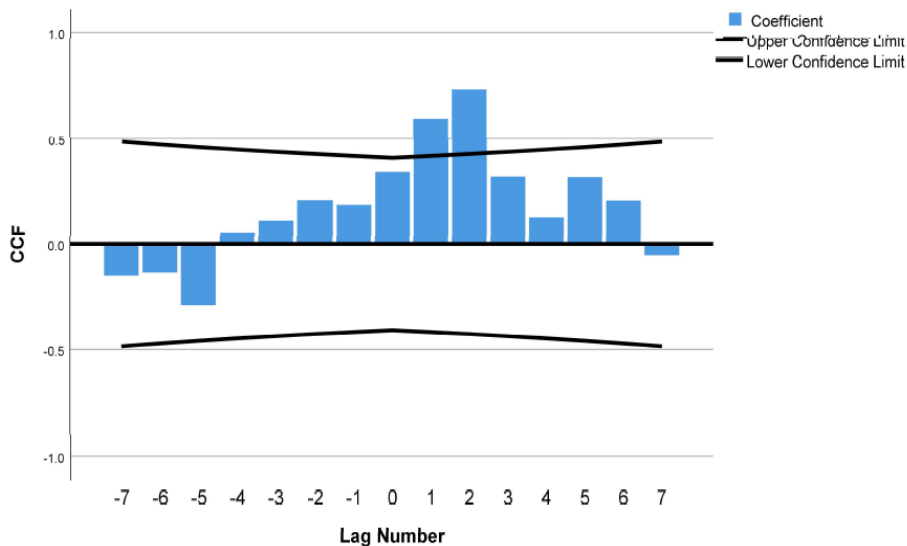


Figure 5. Prius Twitter with Prius Sales Quantity

Figure 5 is a cross correlation graph between the number of tweets on Prius and its actual sales. The cross-correlation coefficient is the highest at the lag value of 2. This means the leading index is two months ahead of actual sales, which means two months of time-lag exists between the tweets and actual sales. This time-lag can be justified, because in Japan brand-new cars are assembled after a sales contract, and actual delivery takes 2 to 3 months after the contract.

4. ANALYSIS

In this chapter we analyze the sales forecast for Prius, which was the most tweeted from June 2018 to May 2020 (1,516,016 tweets), with the methods discussed in the previous chapter. We set up a regression formula by creating a scatter chart using the total number of tweets of a month and actual sales two months after those tweets were posted. Sales forecast was made for August and September 2020, using the regression formula. Figure 6 is a scatter chart and a regression formula from June 2018 to May 2020. Figure 7 is a correlation coefficient between tweets on Prius and the number of sales during that period, which is the mark of 1% and is significant.

From the regression formula:

The number of sales = $0.87 \times$ number of tweets two months before + 1,830

Sales forecast of Prius in August 2020 is:

Number of tweets in June 2020: 5,309

Sales forecasts of Prius in August 2020: $0.87 \times 5,309 + 1,830 = 6,448.8$

Number of sales of Prius in August 2020: 6,429

Relative error: $(6,429 - 6,449) \div 6,449 = -0.3\%$

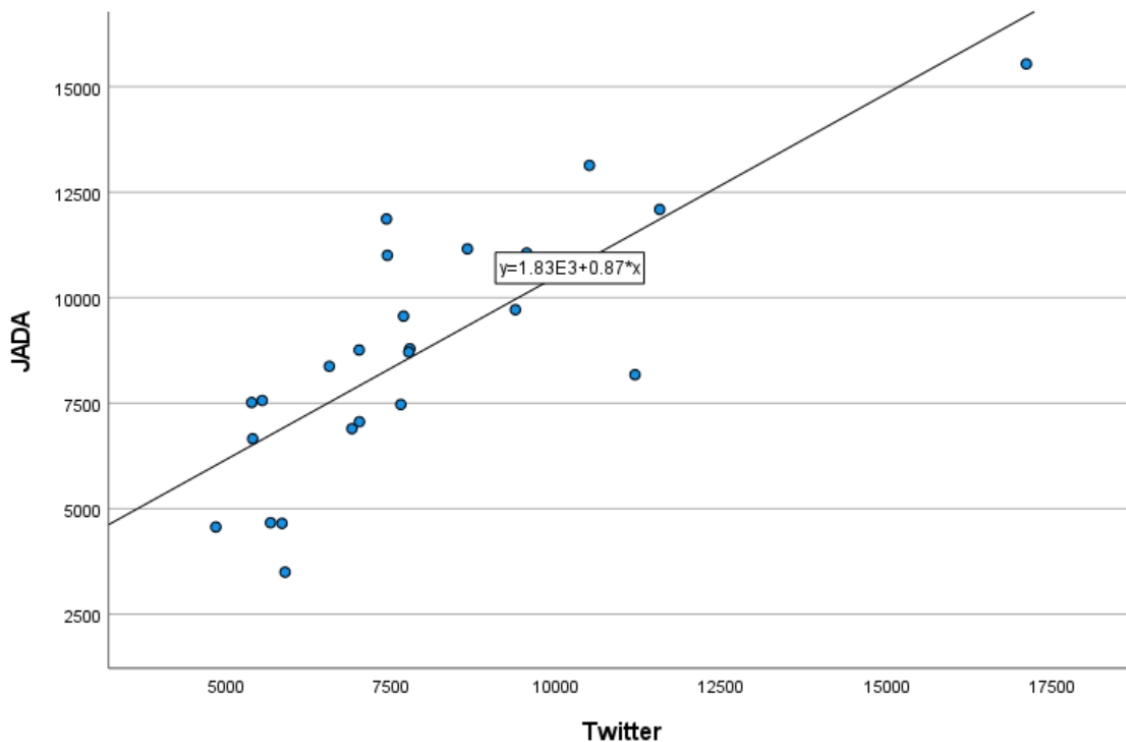


Figure 6. Scatter chart and regression formula for tweets on Prius

Correlations

		JADA	Twitter
JADA	Pearson Correlation	1	.796**
	Sig. (2-tailed)		<.001
	N	24	24
Twitter	Pearson Correlation	.796**	1
	Sig. (2-tailed)	<.001	
	N	24	24

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 7. Correlation coefficient of number of tweets and sales for Prius

Sales forecast of Prius in September 2020 is:

Number of tweets in July 2020: 4,943

Sales forecasts of Prius in September 2020: $0.87 \times 4,943 + 1,830 = 6,130.4$

Number of sales of Prius in September 2020: 5,818

Relative error: $(5,818 - 6,130) \div 6,130 = -5.1\%$

Next, we discuss Corolla, which had the second largest number of tweets during the 24 months of data gathering. According to Toyota, Corolla went on the market in 1966 in Japan. The model was sold 0.4 million units at its peak in the 70s, but recently sells less than 0.1 million (Toyota, 2022). However, outside Japan its annual sales is over one million since 2000, and repeating minor changes contribute to the accumulated total sales of over 50 million so far. The total number of tweets about Corolla was 393,523, which was 26% of Prius', and the sales forecast of the model was not possible.

The model with the third highest number of tweets was Vits (currently Yaris). Its sales volume has been rapidly increasing after renaming to Yaris in February 2020. The number of tweets about Vits has been increasing since the spring of 2020. The forecast was not successful with the current data, but we expect that the forecast will be possible with more Twitter data. Forecasts for other models were not successful either.

Table 2. Number of tweets on each model

Prius	Voxy	Sienta	Corolla	C-HR	Vitz
1,516,016	96,697	68,470	393,523	54,468	204,176
Impreza	Vezele	Alphard	Stepwgn	X-Trail	Mazda2
250,772	65,692	231,582	89,675	78,066	198,658

Now we discuss each model. According to the sales data published by JADA (Figure 2), Prius was in the transition period from a maturity stage to a decline stage, Corolla was in a decline stage in Japan, and Yaris was in a growth stage. When those stages were applied to the PPM (Product Portfolio Management) Model⁽³⁾ by BCG (Boston Consulting Group) (Henderson, 1979), Prius was in a transitional period from 'star' to 'cash cow' and then to 'dog', and Yaris was in a transitional period from 'star' to 'cash cow'. To confirm that, we created matrix diagrams for the models applying PPM. Figure 8 is a matrix by models in the first half of 2019. The vertical axis shows a growth rate (an increase or decrease of the total automobile sales) from the latter half of 2018 to the first half of 2019, and the horizontal axis is a share of each model in the first half of 2019. The size of the circle indicates the sales of each model. Figure 9, 10 and 11 are created in the same manner.

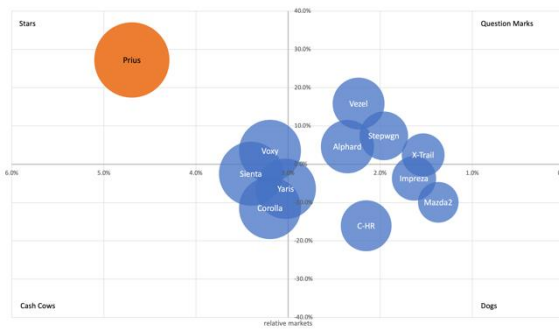


Figure 8. PPM in the 1st half of 2019 by models

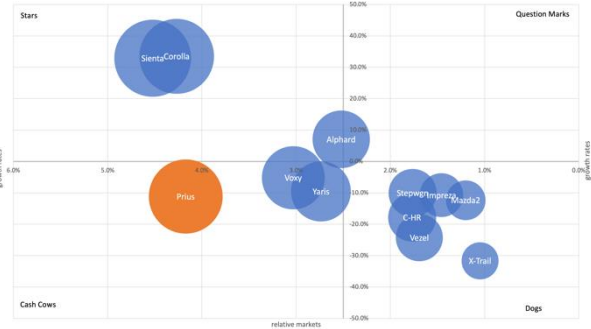


Figure 9. PPM in the 2nd half of 2019 by models

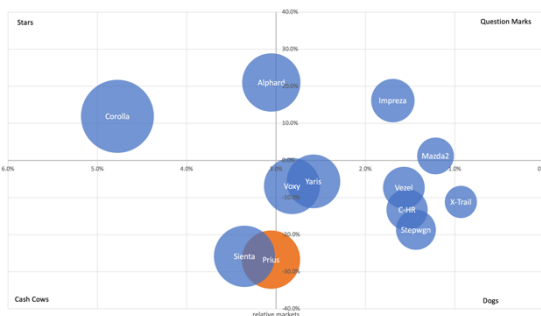


Figure 10. PPM in the 1st half of 2020 by models

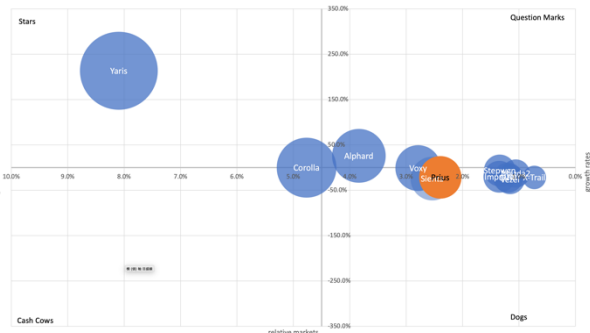


Figure 11. PPM in the 2nd half of 2020 by models

As these figures show, Prius was in the position of star in the first half of 2019 but shifted to cash cow in the latter half of the year. The model was in the transitional period to dog in the first half of 2020 and positioned in dog in the latter half of the year. Other models transition in random order.

From the above, we assumed that although a number of tweets are necessary to make sales forecast through social listening, it is important for products to undergo the four stages of life cycle, which consist of introduction, growth, maturity and decline, to be forecast. In other words, forecast through social listening is not possible on products which shift to decline stage straight after introduction stage.

It is likely that sales forecast is possible during the process of products shifting from cash cows to dogs after positioning as stars and that the products in stars and dogs phases are difficult to forecast sales. This is because people are less likely to write about products in question mark and dog phases. There is a possibility that both a large number of tweets and the product being shifted from cash cows to dogs after positioning as star contributed to the correct sales forecast of Prius.

5. CONCLUSION

In this study we gathered tweets related to automobile models in Japan and used them as alternative data for social listening to forecast the sales of each model. A cross correlation between the number of tweets and the actual number of sales published by JADA showed two month of time-lag. In Japan brand-new cars are assembled after a sales contract, and actual delivery takes two to three months after the contract; thus, this time-lag is deemed accurate. However, models with a lower number of tweets were unforecastable and a versatile sale forecast model which can be applied to all models has not been built yet. For this study we gathered 3,247,796 tweets from June 2018 to May 2020, and Prius accounted for 1,516,016 tweets. This large number of tweets contributed to the correct sales forecast of Prius. In addition, after a PPM analysis, we came to think that the fact that Prius shifted from cash cow to dog after positioning in star may be also a factor for the forecast. Hence, we propose a hypothesis that sales forecast is possible for products whose life cycle undergoes the growth, maturity, and decline; in other words, sales can be forecast during the process of products shifting from cash cows to dogs after positioning in stars. Our future task is to find the number of

tweets necessary for forecast and to establish a versatile model which takes into account life cycle's stage changes in the PPM analysis model caused by various factors such as sales decline due to negative tweets and change of sales price.

ACKNOWLEDGEMENT

This work was supported by Nagoya Gakuin University.

NOTES

- (1). Real-time search is an internet search that reflects and adds postings on the real-time web service like Twitter on the second-by-second basis. Google initiated the services in October 2009, but halted the service in July 2011 due to the expiration of the access provision contract with Twitter. Yahoo! Japan started the real-time research services on June 14, 2011. Yahoo! in the U.S. does not provide the real-time search services.
- (2). The Higasi Ikebukuro runaway car accident is the fatal car accident which occurred in Tokyo on April 19, 2019 caused when a driver mistakenly stepped on the accelerator instead of the brake, resulting in two deaths and nine injuries. The driver, who was a retired elite public bureaucrat and was 87 years old at that time, was driving Prius and insisted that its brake pedal did not work. There were several factors why the number of tweets increased. There were criticism about Prius itself, but also his status as a retired elite public bureaucrat helped the number of tweets related to Prius increase. Toyota announced that its driving record showed no brake pedal was hit right before the accident, and he was sentenced to five years in jail. This accident caused many elders to surrender their driver's license.
- (3). PPM is also called the growth-share matrix, BCG-matrix or Boston matrix. Crainer (1998)

REFERENCES

- Akimoto, A. (2011). *Japan, the Twitter nation*. The Japan Times. May 18th. Retrieved June 3, 2021, from <http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/twitterinjapan-jt.pdf>
- Crainer, S., (1998). *Key Management Ideas: Thinkers that Changed the Management World*, Financial Times Pitman Pub.
- Dijkman, R., Ipeirotis, P., Aertsen, F., & van Helden, R. (2015). Using Twitter to Predict Sales: A Case Study. *arXiv preprint arXiv:1503.04599*.
- Goto, H & Goto, Y. (2018). Analysis of Elections Using Social Listening in Japan. *51st Hawaii International Conference on System Sciences*. pp.1742-1751
- Goto, H. & Goto, Y. (2019). Regression Analysis of National Elections in Japan Using Social Listening. *IADIS*. 2019.
- Henderson, B. D., (1979). *Henderson on Corporate Strategy*, Harpercollins College Div.
- Pai, P. F., & Liu, C. H. (2018). Predicting vehicle sales by sentiment analysis of Twitter data and stock market values. *IEEE Access*, 6, 57655-57662.
- Toyota (2022). ‘*カローラ累計販売台数* (English Translation: Accumulated sales volume of Corolla)’, https://global.toyota/jp/corolla50th/history/by_the_numbers/sales_results_01/, retrieved in June 3, 2021
- Wijnhoven, F., & Plant, O. (2017). Sentiment Analysis and Google Trends Data for Predicting Car Sales. *ICIS*.