Yurika Tokoro

Text Analytics

February 14, 2020

Business Insights Report on Cosmetic Market in Japan

Introduction

Due to the growth in aging population around the world, it has certainly become a significant factor to boost desire among both women and men to maintain youthful appearances (Market Watch, 2019), and Japan's cosmetic market is one of the markets that have been affected most by this trend. Japan is one of the world's biggest markets for cosmetic products, and the domestic beauty industry reached a revenue of US$36 billion as of 2017 according to statista. Due to the fact, many global cosmetics companies have attempted to expand their shares to Japan's market, however, it was not as easy as pie. For instance, Sephora, one of the biggest personal care and beauty chain stores in the global cosmetic industry , entered Japan's market without doubt in success. Nevertheless, their branding strategy did not successfully penetrate throughout the market, thus they ended up leaving there after two years. Currently, LVMH, a parent luxury goods conglomerate of Sephora, has been challenging to grow awareness in its cosmetic brand MAKE UP FOREVER among Japanese consumers, and of course their strategies must be built based on their first failure with Sephora.

What did Sephora do wrong? What is the major factor of making it extremely difficult for new entries, including global gigantic companies like LVMH, to proceed their operations in Japan's market? This report emphasizes on unique insights about Japan's cosmetic market using text analysis conducted with several pdf files of market reports.

Insights from the Most Frequent Words

The most common words used over more than 20 times on the documents are those 17 words as shown in image 1, and mainly three interesting elements are found.

Firstly, you can see some companies dominating Japan's cosmetic market share like Shisendo, Kanebo, and Kao, and these are all traditional Japanese companies, and there is no single name of foreign cosmetics companies as a frequent token on the market report. It demonstrates that there is a high possibility that the cosmetic market in Japan still remains conservative, therefore, consumers' preferences incline towards domestic brands. With that in mind, adjustment for those foreign companies or brands to fit in the market may be inevitable to compete with the local traditional companies, as well as further thorough research on the local companies and brands will be needed.

Secondly, specific products such as "nail", "lip" and "eye" are mentioned repeatedly on the reports, which illustrates which category of products are generally targeted in the market. As famously known, Japanese nail art culture applies one of the best nail techniques in the world (tsunagu Japan, 2019), and

their techniques are admired due to its high quality and careful dedication to details. Taking this fact into consideration along with the result image 1, it is obvious that consumers' interest lies heavily in nail care segment.

The last element is that the word "premium" is surprisingly ranked as the third most frequent word on the market reports. Truth to be told, it is uncertain to exactly conclude that what context "premium" is being used in from this simple text analysis, but it is clear that being premium is a key to stay in Japan's market. Thus, it can be utilized as one of the marketing strategies when promoting a product or brand. To some extent, "premium" certainly adds outstanding value to either a product or a brand itself which leads consumers to self-fulfillment.

Insights from the Most Frequent Word-Pairs with Bigram

It is surprising how different a result of analysis with the most frequent word pair is compared to the previous frequent word analysis. Here, the most frequent pair shown in image 2 is mainly about "trade." Interestingly, it connects to the next most frequent word-pair pattern, that is, "vuitton sa" and "yves saint" which are foreign brands imported to Japan. This leaves a possibility that some elements regarding "trade" are happening in the market, therefore, having deep comprehension of any thwarting factors like different pharmaceutical regulations in terms of international cosmetics trading when reaching out Japan's market is significant

Insights from Sentiment Analysis

Relatively, cosmetics marketing reports of Japan indicates positive sentiment. As you can see from both Image 3 and 4, proportion of positive sentiment to negative sentiment is overall extremely large. Surprisingly, you see "loyalty" here again, which can connect to the possibility of Japan's market being conservative mentioned previously (Image 3). Either loyalty describes the market itself or consumer behaviors, it explains well enough that making consumers in Japan shift to a new product as an alternative will be hard because of their loyal attachment to specific products or brands.

As negative sentiment, some words like "hesitant", "odor" and "limited" appear. This needs further analysis to consolidate insights, since limited can be used as "limited edition" with a positive meaning. Moreover, "odor" can be an interesting category to look at since few local traditional cosmetics brands in Japan have their own perfumes or colognes, so it may become a segment that foreign companies may want to consider aiming for.

Insights from Bigram Graph

Based on the graphs on Image 5 and 6, two curious relationships between each word can be identified. The first one indicates specifically popular products of each category in cosmetics. For example, "lip"

connects to “gloss” instead of “stick”, and “eye liner” connects to “pencil” instead of “liquid”, which demonstrates what type of products consumers tend to like in the categories of eye and lip. Additionally, these two fragments follow “premium.” As mentioned above, premium can create a strong value to brands and products when it comes to marketing, therefore, it might become more successful if promoting eye and lip related products, or even a new category like fragrance with extra value like “premium”, “best”, “top”, etc.

The other interesting finding is “men’s cosmetics sets kits.” Apparently, men invest more dedication on their appearances and personal care nowadays compared to how men in general used to do in Japan. However, based on the words “sets” and “kits,” male consumers might not be as particular about what brand or product they use as female consumers are, but more likely, they incline towards something convenient, like purchasing every skin care essential all at a time as a “set” at one place. Thus, targeting male segment and providing more convenient personal care solution could be a chance for new entries instead of jumping into highly competitive female sector.

Conclusion

Based on the text analyses conducted based on Japan’s cosmetic market reports, several fascinating insights are revealed. The analysis with the most frequent word identified the market’s conservativeness towards foreign companies, consumers’ interests in cosmetics as well as a possible effective marketing strategy. With the bigram analysis, the possibility that some matters regarding trading in Japan’s cosmetics industry was demonstrated. With the sentiment analysis, we captured the whole overview of positive sentiment from the reports, and it also showed further needs on analysis on negative sentiment words to extract more insightful information about the market. Lastly, the bigram graph visualized the market’s specific trends in eye and lip related products with words creating superiority, and then it showed a new chance of targeting male sector by providing easier and more convenient sets or kits of personal care essentials.

In conclusion, the cosmetic market clearly demonstrates Japanese traits and culture. No matter how big the size of a company is, it is obvious that new entries must analyze and learn how Japan’s competitive market works and carefully aiming for a specific niche or segment of product is the major key to success.

Image 1: Most frequent word used

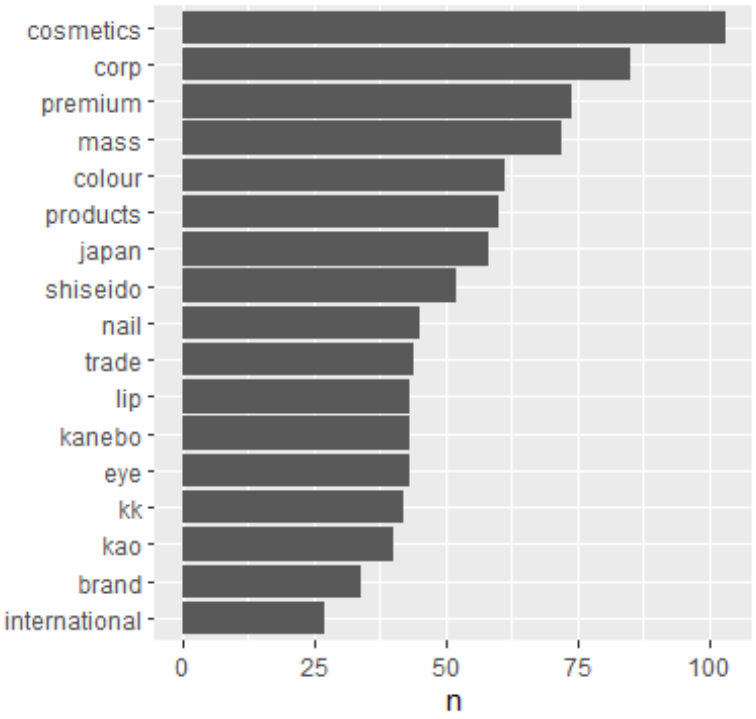


Image 2: Most frequent word-pair used

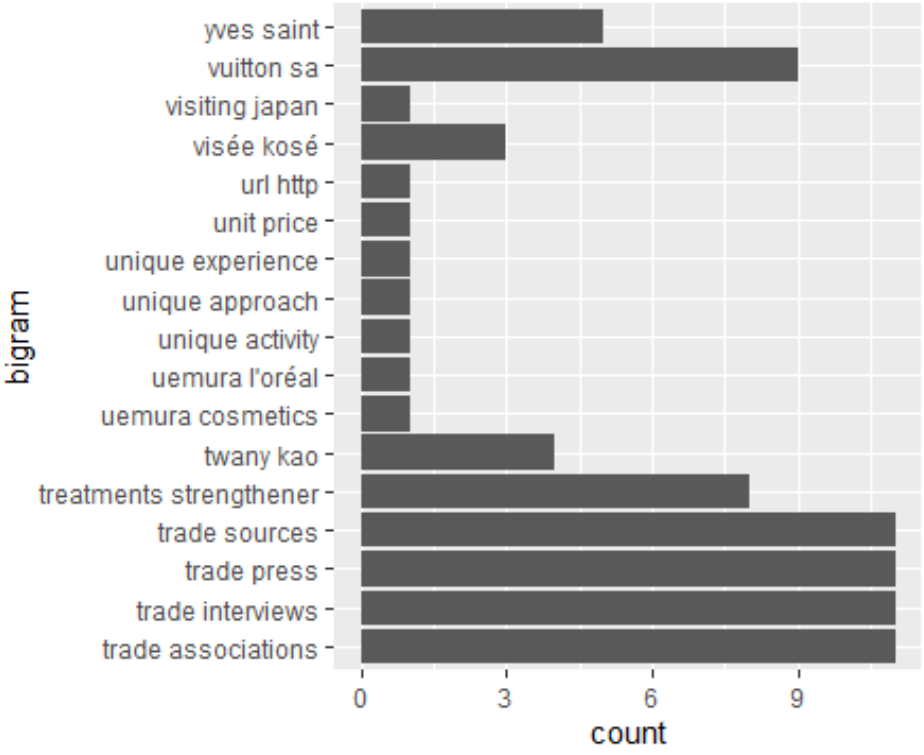


Image 3: Word cloud with sentiment (positive vs. negative)

Image 4: Words contribution to each type of sentiment

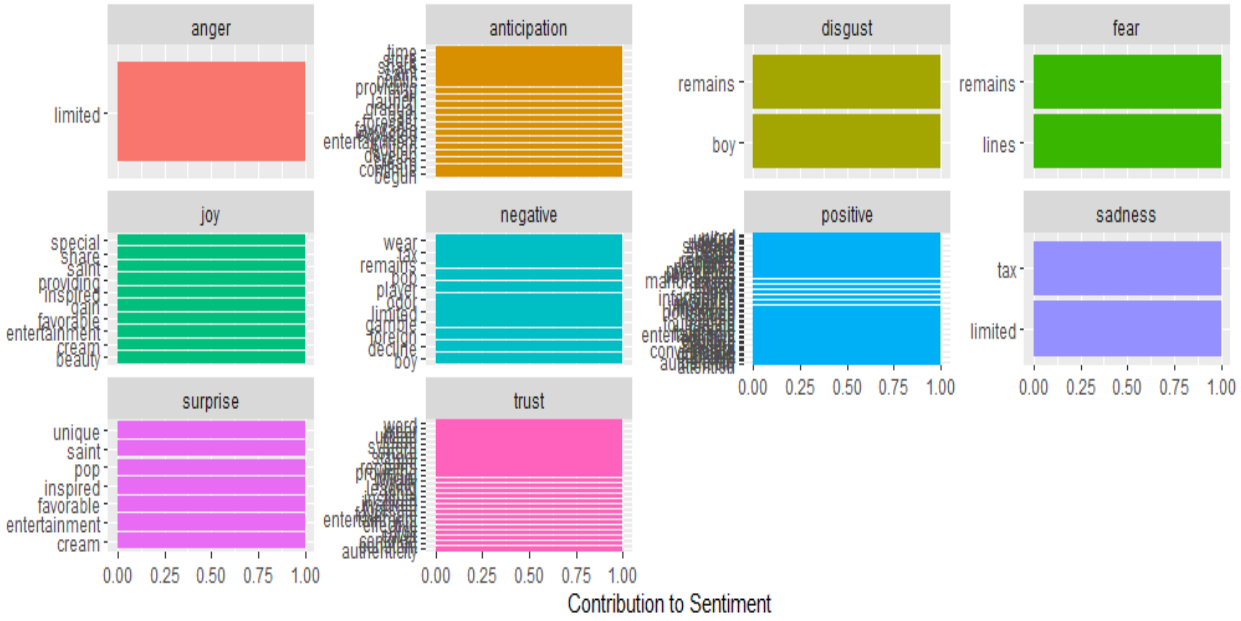


Image 5: Bigram Graph

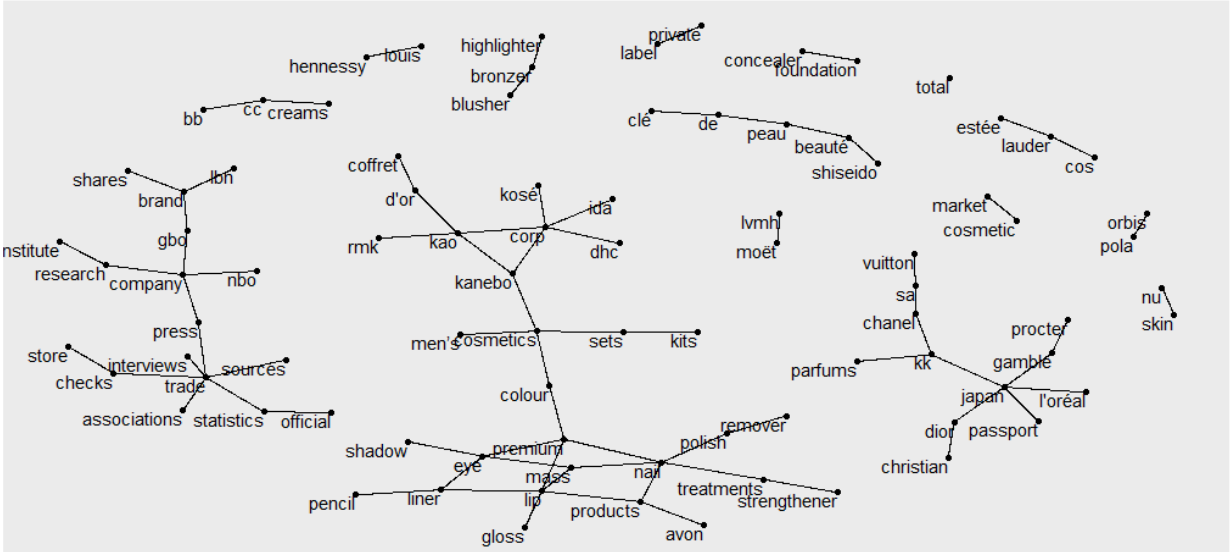
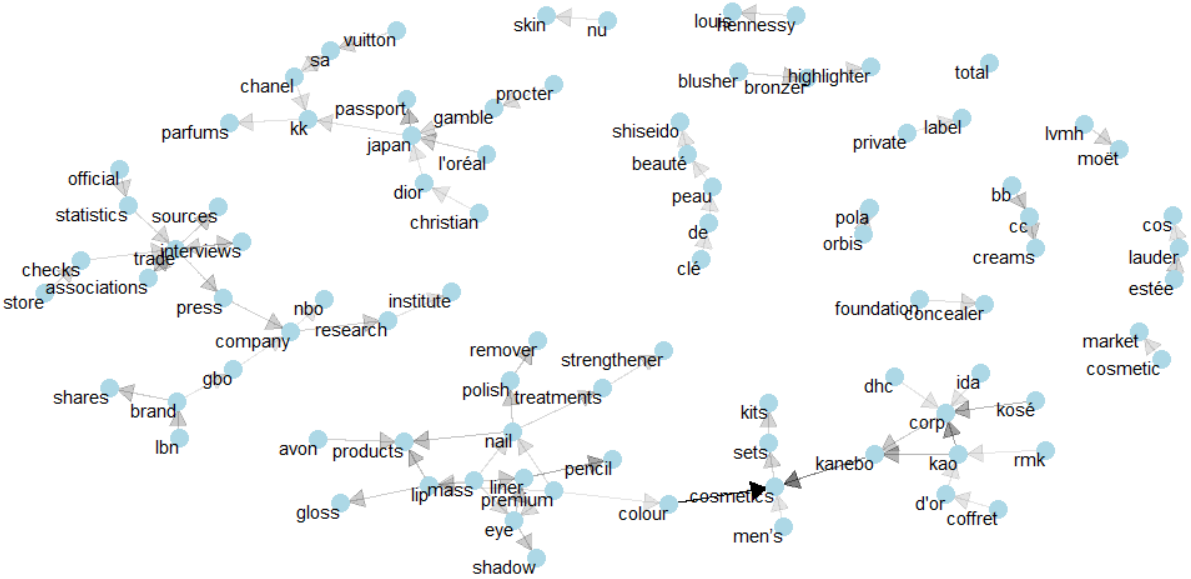


Image 6: Bigram Graph with Arrows



Appendix

```
## Setting all packages needed

library(textreadr)
library(dplyr)
library(stringr)
library(tidytext)
library(tidyr)
library(scales)
library(wordcloud)
library(reshape2)
library(textdata)
library(igraph)
library(ggraph)



## Importing pdf files

pdf1 <-
read_document(file="C:/Users/User/Desktop/DataScience_R/Individual/Colour_Cosmetics_in_Japan.pdf")
pdf2 <- read_document(file="C:/Users/User/Desktop/DataScience_R/Individual/1601.pdf")

## Combine the two pdf files into one
pdf_combo <- c(pdf1, pdf2)
```

```
## Convert pdf_combo into a data frame
a <- 836
b <- 1
pdf_df <- as.data.frame(matrix(nrow=a, ncol=b))

for(z in 1:b){
  for(i in 1:a){
    pdf_df[i,z]<- pdf_combo[i*b+z-b]
  }
}



## Stop words
data(stop_words)

## Creating my own stop words and assigning it to "bind"
num <- c(seq.int(0, 6500, 0.1))
num2 <- c("1.0", "2.0", "3.0","4.0", "100.0", "11.0", "5.0", "6.0", "9.0", "8.0", "12.0",
        "23.0", "26.0", "13.0", "21.0", "0.0", "103.0", "130.0", "19.0", "2410.0", "211.0",
        "24.0", "25.0", "28.0", "40.0", "410.0", "7.0", "8620", "110,000", "122.0", "16.0",
        "22.0", "358.0", "42.0", "426.0", "43.0", "45.0", "49.0", "5371", "58.0", "6912",
        "708.0", "73.0", "732.0", "yano.co.jp", "www.yanoresearch.com", "jpy680.9", "jpy920.7",
        "jpy780.9", "1,000.0", "714.0", "621.0", "376.0", "2136.0", "166.0", "473.0", "425.0",
        "338.0", "353.0", "2417.0", "2139.0", "3523.0", "298.0", "623.0", "171.0", "1131.0", "1005.0",
        "usd", "usd17.3", "usd21.6", "yano")

c <- data_frame(word = c("table", "euromonitor international", "copyright","euromonitor", num2),
lexicon = c("custom"))
n <- data_frame(word = num, lexicon = c("number"))
```

```
bind <- rbind(c, n)

## Binding my custom stop words and default stop words and assigning it to "junk"
junk <- bind_rows(bind, stop_words)



#################
## Tokenizing ##
#################
pdf_df1 <- pdf_df$V1
pdf_df2 <- data_frame(line = 1:a, text = pdf_df1)
tidy_df <- pdf_df2 %>%
  unnest_tokens(word, text) %>%
  anti_join(junk) %>%
  count(word, sort = TRUE) %>%
  ungroup()

## Plotting ranking of the most frequent tokens
tidy_df %>%
  filter(n > 25) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()
```

OUTPUT Image 1

```
########################
## Sentiment Analysis ##
########################

## Plotting word cloud with bing (positive vs. negative)
tidy_df %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                   max.words = 100,
                   scale = c(1, 1),
                   fixed.asp = TRUE,
                   title.size = 1)
```

OUTPUT Image 3

```
## Plotting contribution to each sentiment with NRC
tidy_df_nrc <- tidy_df %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

tidy_df_nrc %>%
  group_by(sentiment) %>%
  top_n(5) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
```

```
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Contribution to Sentiment",       x = NULL) +
  coord_flip()
```

OUTPUT Image 4

```
#############
## Bigram ##
#############

## Tokenization
pdf_bigram <- pdf_df2 %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

## Separating each pair
pdf_bi_separated <- pdf_bigram %>%
  separate(bigram, c("word1", "word2"), sep = " ")

## Removing unnecessary words
pdf_bi_nojunk <- pdf_bi_separated %>%
  filter(!word1 %in% junk$word) %>%
  filter(!word2 %in% junk$word)

## Removing NA values
pdf_bi_filtered <- na.omit(pdf_bi_nojunk)
```

```
## Count each word in word1 per word2
pdf_bi_counts <- pdf_bi_filtered %>%
  count(word1, word2, sort = TRUE)

## Reuniting word1 and word2 as bigram
bigram_united <- pdf_bi_filtered %>%
  unite(bigram, word1, word2, sep=" ")

## Plotting ranking of the most frequent pairs
bigram_united %>%
  top_n(80, bigram) %>%
  ggplot(aes(bigram)) +
  geom_bar() +
  coord_flip()
```

OUTPUT Image 2

```
###################
## Bigram Graph ##
###################

pdf_bi_graph <- pdf_bi_counts %>%
  filter(n > 5) %>%
  graph_from_data_frame()

ggraph(pdf_bi_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
```

```
  geom_node_text(aes(label = name), vjust = 1, hjust = 1)
```

OUTPUT Image 5

```
## Bigram Graph with arrows
set.seed(2016)
arrow <- grid::arrow(type = "closed", length = unit(.15, "inches"))

ggraph(pdf_bi_graph, layout = "fr") +
  geom_edge_link(aes(edge_alpha = n),
                 show.legend = FALSE,
                 arrow = arrow, end_cap = circle(.07, 'inches')) +
  geom_node_point(color = "lightblue", size = 5) +
  geom_node_text(aes(label = name),
                 vjust = 1, hjust = 1) +
  theme_void()
```

OUTPUT Image 6

Reference

Market Watch, 07/11/2019, “Cosmetics Products Market 2019 Global Industry Trends, Shae, Size, Demand, Growth Opportunities, Industry Revenue, Future and Business Analysis by Forecast – 2023”, Retrieved 02/12/2020 from : https://www.marketwatch.com/press-release/cosmetics-products-market-2019-global-industry-trends-share-size-demand-growth-opportunities-industry-revenue-future-and-business-analysis-by-forecast-2023-2019-07-11

Statista, 10/17/2019, “Cosmetics Market in Japan – Statistics & Facts”, Retrieved 02/12/2020 from : https://www.statista.com/topics/4661/beauty-industry-in-japan/

Tsunagu Japan, 12/10/2018, “8 Things We Learned About Japanese Nail Art $ Trends at the Tokyo Nail Expo 2018”, Retrieved 02/12/2020 from: https://www.tsunagujapan.com/8-things-learned-japanese-art-trends-tokyo-nail-expo-2018/