Shopee

Done By: Attenborough Sara Jo, Lim Ye Ju, Marcus Tay, Tiong Nai Shi

SMU SINGAPORE MANAGEMENT UNIVERSITY

ABOUT THE CLIENT

Shopee is a fast-growing Singaporean ecommerce platform for consumer products that caters to users in Southeast Asia and Taiwan. Due to the **mobile and social elements** prevalent in its business concept, Shopee was described as one of the "5 disruptive ecommerce start-ups we saw in 2015" by Tech In Asia.

BUSINESS PROBLEMS

In the midst of its rapid development, Shopee had been prioritising its business intelligence more on the operations flow than on consumers. Currently, it **lacks in awareness** of three main aspects:
1) **Upcoming product trends**, 2) **General customer sentiments**, and
3) **Competitor strategies** which could serve as critical data points.

PROPOSED SOLUTIONS

For the scope of this project, Shopee has identified that it would like to focus on *consumer technology products*.

A. Identification of Product Trends
To better position itself as a leading e-commerce site that provides products relevant to consumer demand, product trends will be identified at critical points in time, such as prior to the launch of a major sales campaign.

Data Retrieved: Latest tweet content of top Twitter tech accounts to be stored in a text file.

Data Retrieved: Crawled consumer tech-related hashtags (#) to identify trending hastags based on post frequency. Then, top/latest post content image for each # were retrieved and stored in a folder.

B. Sentiment Analysis of Shopee Reviews
To ensure customer retention, Shopee would need to understand the needs of its existing customers and address any dissatisfaction, hence customer reviews will be looked into and analysed.

Data retrieved: Customer reviews of 450 identified consumer tech products, including product category, timestamp, review content and ratings.

C. Competitor Analysis on Other E-commerce Platforms
To acquire more customers, Shopee has to ensure that it not only matches up to but outdoes its competitors by staying on top of business strategies and offering more desirable product deals. This will be based off Lazada's current offers and customer reviews.

Data Retrieved: Scraped customer reviews of 600 identified products, incl. product category, timestamp, review content and ratings. Individual products with original and discounted prices and seller IDs were also captured.
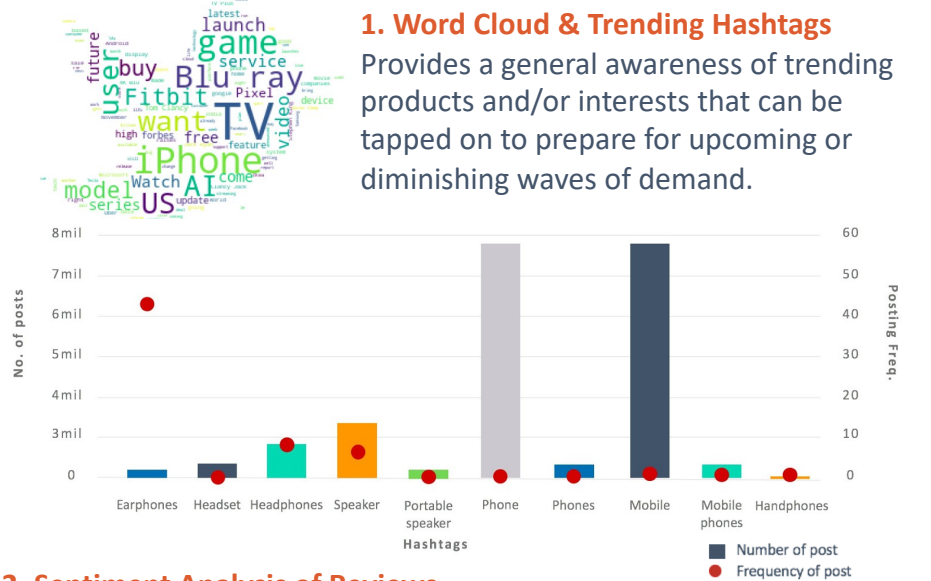
TECHNOLOGIES AND TOOLS

We used **Python** as our primary language for data mining, data preprocessing and data analysis. **Pandas** and **Numpy** libraries were utilized throughout our project to manipulate and store our data.
In order to web crawl and scrape from the various platforms, we used **urrlib.request** as well as **BeautifulSoup** to get hashtags, posts and images. Also, to bypass certain captchas, we had to import **webdriver** from **Selenium** to emulate a web browser to get data from the platforms.
We also used **Vader** to conduct our sentiment analysis as well as **PIL and WordCloud libraries** to generate our word cloud. We had to use **Tkinter** as well to generate a slideshow to showcase our images that we scraped from Instagram.

RESULTS AND INSIGHTS

1. Word Cloud & Trending Hashtags
Provides a general awareness of trending products and/or interests that can be tapped on to prepare for upcoming or diminishing waves of demand.


2. Sentiment Analysis of Reviews
Apart from getting a general sense of customer sentiments, reviews that are ranked higher based on their assigned sentiment scores will be more visible, allowing negative feedback to be addressed in time.

	Reviews	pos		Reviews	neg
834	GOOOD GOOD GOOD GOOD GOOD	0.921	509	Poor product quality Poor overall service	0.608
4935	Good bargain. I love this. Super cute. Nice pr...	0.862	3761	No comment as I was forced to return to seller...	0.608
7974	Good and nice. Receive a lovely gift.	0.861	4966	Received item defective not working.	0.529
3086	Thanks! Will enjoy good music.	0.823	5564	dont recommend due to bad print quality	0.529
1445	Fast delivery, great deal and awesome gift!!	0.745	6808	item received is faulty.	0.509
452	Awesome purchase & great delivery speed!	0.739	1055	The earpiece wire is a little long but no prob...	0.398

3. Competitor Price Comparisons
Dynamic fluctuations in competitor product prices can be identified and compared directly with that of Shopee's to ensure that the most price competitive deals are made available at all times.
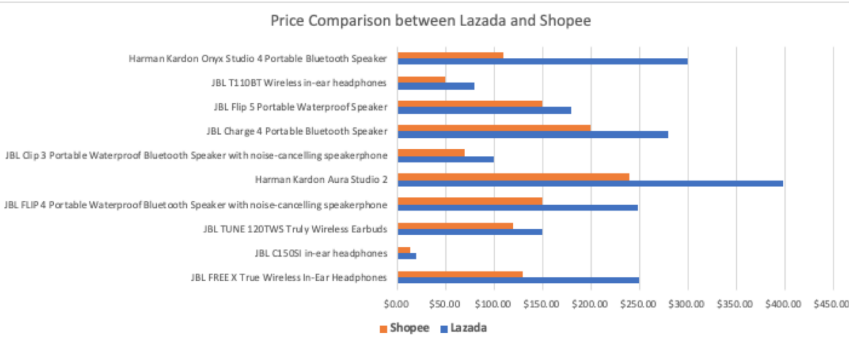
WHAT WE LEARNT

1. Impact of social analytics on businesses
With the proliferation of social media usage, contextualisation of such data can provide businesses with a competitive advantage by staying atop trends and being keenly aware of changing consumer demands.

2. Web scraping as valuable tool for data retrieval
Consumer sentiments and feedback can be tapped on to assess strengths and weaknesses, which when properly addressed will be synergistic to business growth and customer satisfaction.

3. Importance of data transformation
The filtering and cleaning of data to be of relevance to the business objective is just as important as the gathering of it.