HKTV mall

☰ To-do

PROJECT PLANNING 🔍

🛒 ¹ PROGRESS

Workflow

WEB SCRAPING

DATA CLEANING & PREPROCESSING

POWERBI Dashboard

STREAMLIT Deployment

Potential Usage

HKTVmall

- Increase bargaining power by identifing the top/less performed partners and review the trading terms accordingly
- Make profit by selling the analyzed data to brands for their own review
- Distinguish the product popularity for virtual merchandising and website content update
- Identify the comment types and frequent comment wordings to leverage its comment session and service

Other e-commerce platforms

- Evaluate brand rating and frequent comments to identify new business partners
- Identify the market trend e.g. which category most common and inform the brands

Retailer/Wholesalers

- Determine consumer preference for inventory allocation
- Identify market trend for new product opportunities

Manufacturer

- Identify the best/worst rated products to reallocate the inventory
- New product development/discontinue low rated products
- Discover consumers' cared issue for quality control and improvement

Customer

- Identify similar and top performing products to avoid negative experience

To-do | DATA COLLECTION | PROGRESS

Web Scrapping

HKTVMall Website: Hot Product Page

- https://www.hktvmall.com/hktv/zh/main/search?page=**PageNumber**&q=**Category**
- **Selected number of Categories: 8**
- Values: ProductName, Brand, URL,Number of Reviews
- Libraries: BeautifulSoup, Selenium

HKTVMall Website: Product Page

- URL from scrapped data frame(50+ Reviews products)
- **Selected number of Products: 4K+**
- **Number of Reviews: 475K+**
- Values: ProductName, Brand, ShopName, Price, SaleNumber, ImageURL, Category , Reviews(Username, Rating, ReviewText,Date)
- Libraries: BeautifulSoup, Selenium

To-do [4]

PREPROCESSING 🔍

🛒 [4] PROGRESS

Data Cleaning

NULL values treatment

- Fill NA with '未有資料' or 'url:image Coming soon.jpg' for display

NLP

- Remove hktvmall cs reply
- Drop rows with pure special characters e.g. punctuations, emoji, digits
- Identify and remove English and Chinese stop words

Classification

- Manually label product subcategory and comment type
- Train and predict comment type by machine learning - limit by dataset size

≡ To-do

ANALYSIS 🔍

🛒 5 PROGRESS

PowerBI
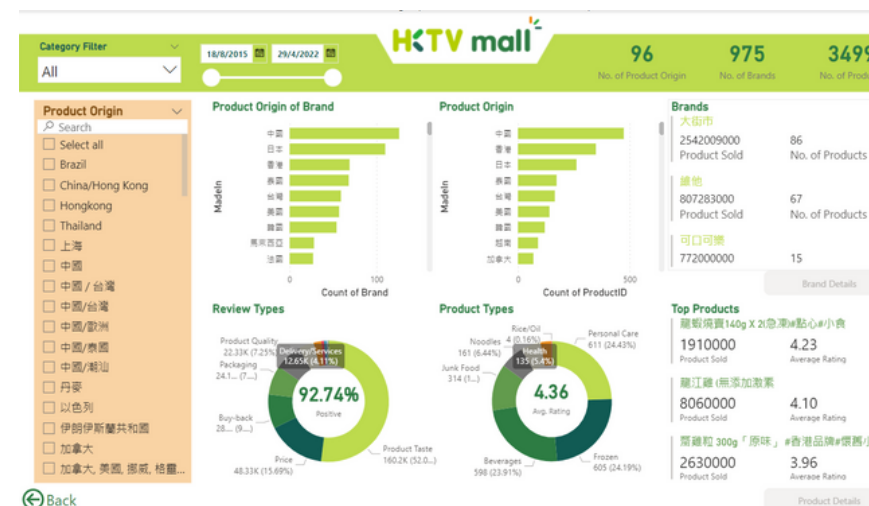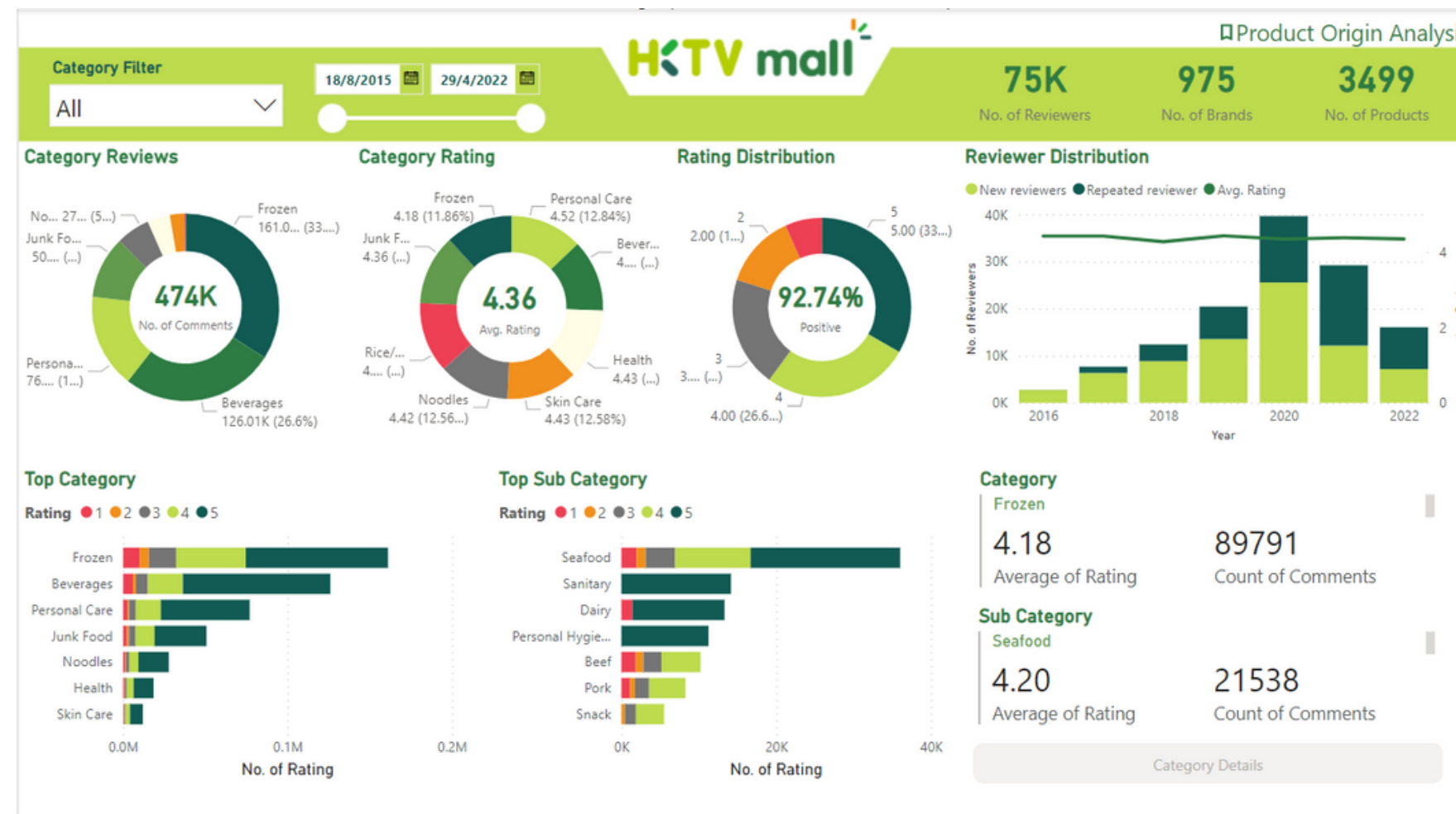
- **Video Demonstration:**
  - https://youtu.be/JobPgGj5-mc
- **File:**
  - https://github.com/naomitsang1016/Capstone_HKTVMALL_RecommendationSystem_BI_Dashboard

To-do   MODEL   🔍   6 PROGRESS

Recommendation System

- **Display basic product information**

- **Collaborative Filtering**

-Recommend products brought by users who have purchased the same item

-Recommend similar products (by Brand, Sub-Category)

- **If the ProductID is invalid:**

     Recommend 4 randomly chosen products

- **Demonstration: Streamlit**

    https://share.streamlit.io/naomitsang1016/capstone_hktvmall_recommendationsystem_bi_dashboard/main/RecommendationSystem/Streamlit_hktvmall.py

To-do   LIMITATIONS & FUTURE IMPROVEMENT   🔍   🛒 7   PROGRESS

E-Commerce Platform Related Limitations

- **No easy access to commercial data**
  - HKTVmall denied our request for accessing their product database
  - Web scraping will get your IP address being blocked temporary by HKTVmall
  - Manual labelling consumes huge amount of time

  Soluton:
  - Wait for a specific period of time between each visit to the page
  - Hire a team (e.g. data entry part-time workers) to perform labelling

- **The ever-changing nature of e-commerce platform**
  - A product is no longer available/the product link changes
  - The program stops working if the script cannot establish connection with the product page

  Soluton:
  - Expect the unexpected: make good use of Python's try and except statement

To-do

LIMITATIONS & FUTURE IMPROVEMENT

PROGRESS

Machine Learning Related Limitations

- **NLP issue**
  - We tried to train a model to categorises comments; it takes time to categorise the comments
  - Varieity of comments: Traditioanl vs Simplified Chinese, English mixing with Chinese
  - Typos in Chinese, e.g. mistakently type '貸品' instead of '貨品' can easily confuse the model
  - There is not way to identify the brand name in a comment
  - After removing the stop words, it is difficult, even impossible to tell what the reviewer is saying

  Soluton:
  - Request reviewers to indicate the nature of their review
  - Train a team specialise in Cantonese Lingustics and NLP

Software and Hardware Limitations

- **Hareware Issue**
  - When training the NLP model, the program shutted down due to the collosal amount of data
  - PowerBI takes time to load the data

  Soluton:
  - Use a high-end computer
  - Utilise Google CLoud Compute Engine

- **Software Issue**
  - PowerBI free version disables the colloborative function

  Soluton:
  - Subscribe the Pro/Premium version

To-do

CONCLUSION

PROGRESS

Findings

- Consumers usually comment on Product Taste, Price, and Product Quality, Delivery/Services
- Some reviewers complain that HKTVmall do not email back to them
- HKTVmall customer service team responds to negative comments in a timely manner
- Consumers usually provide generic comments e.g. nice, good, or copy/paste previous comment
- Most consumers give high rating (4-5 out of 5)

Suggestions

- Recommend HKTVmall to segment its comment and rating section by comment type for easier referral on the area the product is top/least performing
- HKTVmall may allow reviewer to either rate or comment on a proudct
- Suggest HKTVmall to promote "Quality Comment Campaigns" to better identify improvement areas (on Amazon, consumers can vote helpfulness of a review)

To-do ⁱⁱⁱ ¹⁰

Q&A 🔍

🛒 ¹⁰ PROGRESS

Thank You!
Q&A