

# EXCEL PROJECT I: SHOPEE SALE STRATEGY

## THE FIRST STEP IN STRATEGY FORMULATION

**Chinh X. Mai**

*Last update: June 16, 2022, v.01*

### CONTENTS

1. Case description and Personas establishment .....	2
1.1. Case description .....	2
1.2. Personas .....	2
2. Storytelling design checklist .....	3
2.1. Background of the case .....	3
2.2. Audience personas .....	3
2.3. Checklist .....	3
3. Data extraction and cleaning .....	3
3.1. Data extraction .....	3
3.2. Data importing and cleaning .....	5
4. Analysis details .....	8
5. Design choices .....	12
6. Final visualization analysis .....	12
7. Related Information and Data .....	12
7.1. Dataset .....	12
7.2. Changelog .....	12
Appendix and References .....	12
References .....	12

## I. CASE DESCRIPTION AND PERSONAS ESTABLISHMENT

### I.1. CASE DESCRIPTION

Hal is the co-founder of a startup that is aiming to penetrate the Vietnamese market. His company is based in California USA and specializes in providing customized furniture and indoor appliances. The other founder, Huy, is a Vietnamese whose family has a history in assembling and retailing furniture. Huy is the strategy manager of the startup and the one who suggests this expansion. However, Hal is only convinced by statistics and so hires me to carry out the analysis using secondary data provided by a data vendor in Vietnam.

Hal does not know much about the Vietnamese retailing market, so he needs quick information about the potential sales and demand of the market. He possesses a basic understanding of statistics and likes to work with tables and charts. He needs a dashboard that presents important information about revenue, factors affecting revenue, trends in the market and product categories that have the highest revenue. This information is useful for his further discussion with Huy on their tactical decisions and forming their strategies to penetrate the retailing market in Vietnam.

I am given the access to the service of the data vendor in Vietnam. My tasks include extracting data from the vendor, cleaning the data, carrying out the analysis, and designing the dashboard to present my findings. Both Huy and Hal are young businessmen with good vision who focus more on the details and important information rather than the formality. However, it is still necessary to fulfil the aesthetic requirements of the dashboard so that important information is more efficiently delivered to the audience.

### I.2. PERSONAS

Persona 1: Hal



**Role:** Executive Officer

**Organization:** PKT international retailer

**Goals:**

- \* Get better understanding about the Vietnamese market
- \* Evaluate the market potential using sales statistics like revenue
- \* Know which factors impact the total revenue of a product in the long term

**“Money does not grow on trees; strategy does not come from thin air. Thoughtful consideration is where a great journey starts”**

Being a decision maker in a startup, Hal must make many decisions under high pressure. He is sensitive to changes and a careful thinker. Every decision he made was based on facts and statistics together with thorough discussions with his colleagues.

**Challenges and Needs:**

- \* Quick facts and figures for thought initiation and discussion
- \* Statistics is necessary for a good decision, too many statistics can lead to a bad decision
- \* He likes to focus on the most important points and only discuss the most meaningful findings

## 2. STORYTELLING DESIGN CHECKLIST

### 2.1. BACKGROUND OF THE CASE

#### Main tasks:

- Extract data from the data vendor
- Clean data
- Carry out data analysis to answer certain questions
- Design dashboard

#### Primary requirements:

- A dashboard for quick information and thought initiation
- Focus on top-level facts with strongly backed-up statistics
- Short enough for a page printout

### 2.2. AUDIENCE PERSONAS

#### Executive Officer Hal:

- Needs: quick but accurate information, clear presentation with focus on important points, detection of important factors
- Characteristics: careful thinker, sensitive to changes, prefer data-driven decision making

### 2.3. CHECKLIST

**WHO:** Hal is a stakeholder who is thoughtful and pay attention to details, while another audience is Huy who also joins the discussion with Hal with presented statistics. Both have a basic understanding of statistics and can interpret presented charts and tables.

**WHAT:** The data will be extracted from a data vendor in Vietnam, from the description of how the data is collected, it can be said that the source is quite reliable. Nothing can be said yet about the quality of this data. The data is up to date as the vendor provides constantly updated data on their platform.

**WHY:** Exploratory analysis on factors that affect sales revenue as well as indicators that can be used to estimate the potential of the market demand.

**HOW:** Constructing a dashboard to show quick information using charts and tables, which is short enough for a one-page printout

## 3. DATA EXTRACTION AND CLEANING

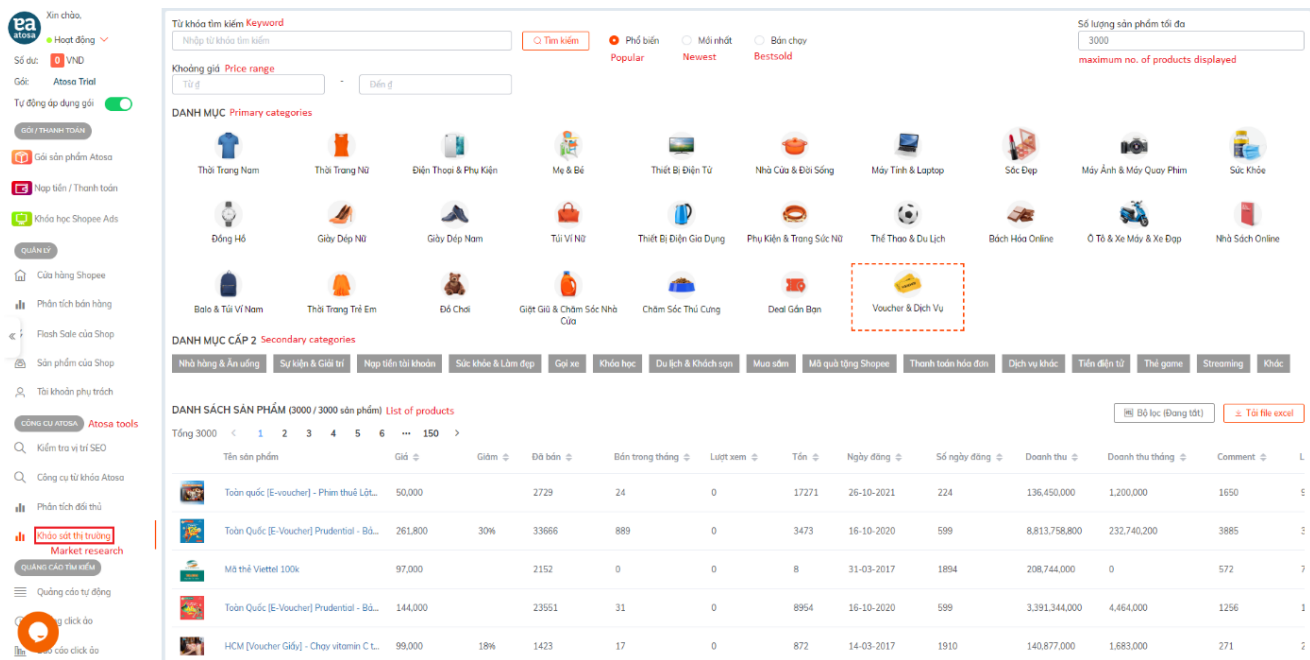
### 3.1. DATA EXTRACTION

#### Platform description:

Atosa (<https://app.atosa.asia/>) is a company that provides many marketing automation solutions for Shopee, an ecommerce platform in Vietnam. These solutions include Atosa Shopee Ads, Atosa Keyword Tools, and Atosa Agency. When having subscribed to these services, the user will be provided with many tools, one of which is the Market research tool, whose user interface is presented below

Since Shopee is an e-commerce platform in Vietnam, the user interface (UI) is in Vietnamese, so some texts have been added to the illustration to explain important elements of the UI. The UI is quite intuitive and there are many options for filtering and ordering data.

The basic function of this tool is to provide a list of products following given preferences of its user. The user can filter the products by choosing a set of categories (primary, secondary, and even tertiary categories) and order the results by popularity, novelty, or bestseller. At maximum, up to 3,000 products can be loaded from the database and displayed in the results list. However, this might not be all the available products following a chosen set of conditions for fetching.



Picture 1: The user interface of the platform with key elements translated for a more comprehensive presentation

## Data description

The data has many features whose names are displayed in Vietnamese. These features are (after translated if necessary):

- Product name (Tên sản phẩm)
- Price (Giá)
- Discount (Giảm)
- Sold quantity (Đã bán)
- Last month sold quantity (Bán trong tháng)
- Currently in stock (Tồn)
- Posted date (Ngày đăng)
- Posted duration in days (Số ngày đăng)
- Total Revenue (Doanh Thu)
- Last month revenue (Doanh thu tháng)
- Like
- 5 Star
- Rating
- 5 Star
- 4 Star
- 3 Star
- 2 Star
- 1 Star




Fetches data can be saved as an Excel file. However, the downloaded data does not include the last 5 features (5 Star, 4 Star, 3 Star, 2 Star, and 1 Star), which are necessary for understanding the rating distribution. This data is updated constantly so it might change depending on the collected date.

The given data here is collected on June 7, 2022, following a set of certain fetching and ordering conditions. These conditions are as follows

- Ordering by popularity
- Maximum 3,000 products are displayed
- Data is fetched using only primary categories (so there are 27 categories in total)

In total there are maximally 81,000 observations in the final consolidated data before cleaning.

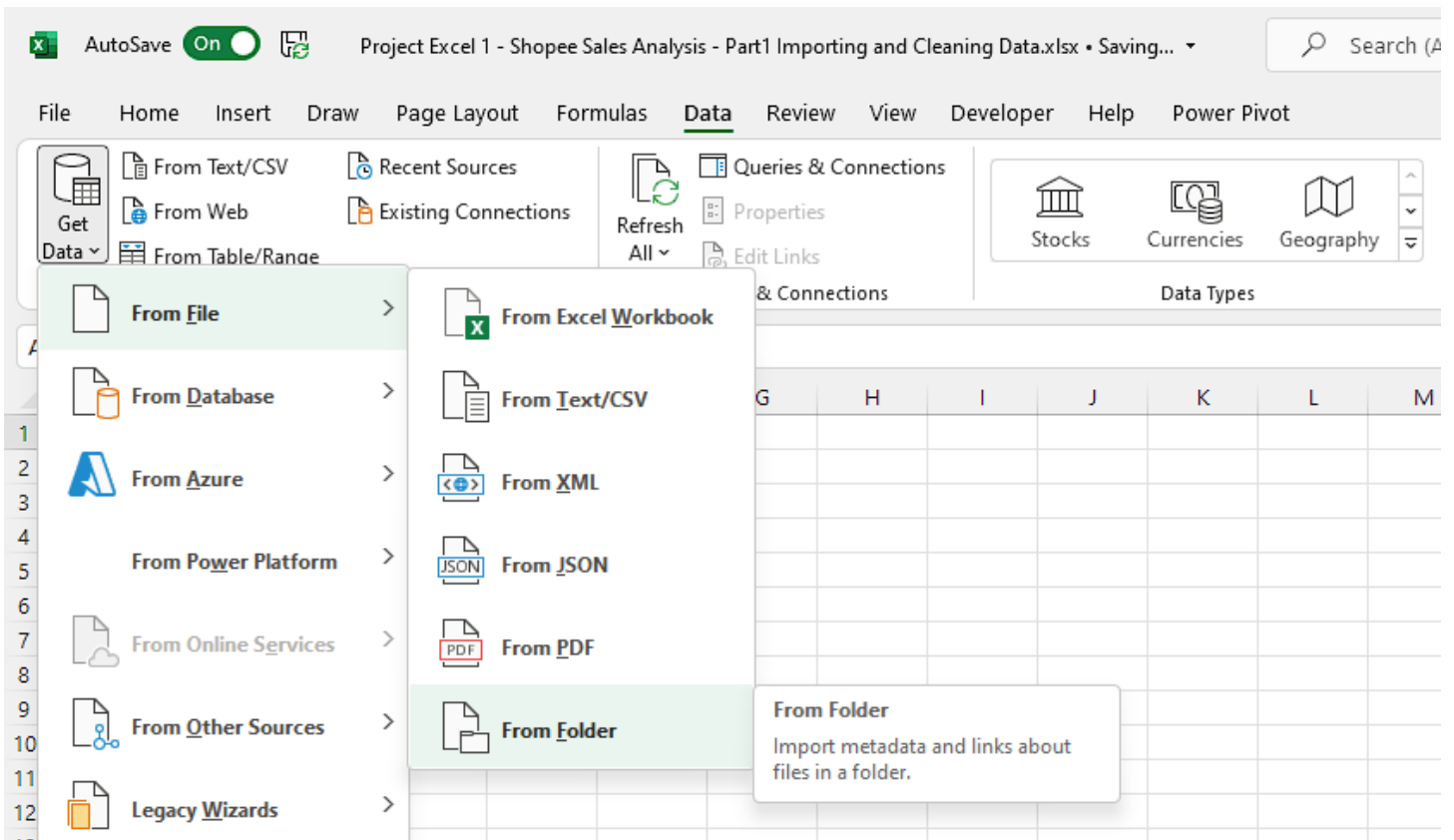
Only 3,000 products of each category are extracted and saved in an Excel file whose name follows the structure “[Data report] – [Category name] – [Date information].xlsx”. Since the data does not contain information about its category, this structure makes it more convenient to generate information about data categories later. A picture of the raw data is shown below

Name	Status	Date modified	Type	Size
 Báo cáo dữ liệu - Bách hoá online - ngày 07_06_2022 07_42.xlsx	 	6/8/2022 8:54 AM	Microsoft Excel Worksheet	446 KB
 Báo cáo dữ liệu - Balo và Túi Ví Nam - ngày 07_06_2022 07_55.xlsx	 	6/7/2022 7:55 AM	Microsoft Excel Worksheet	1,791 KB
 Báo cáo dữ liệu - Chăm sóc thú cưng - ngày 07_06_2022 08_16.xlsx	 	6/7/2022 8:16 AM	Microsoft Excel Worksheet	1,791 KB
 Báo cáo dữ liệu - Deal gần bạn - ngày 07_06_2022 08_16.xlsx	 	6/7/2022 8:17 AM	Microsoft Excel Worksheet	458 KB
 Báo cáo dữ liệu - Điện thoại và phụ kiện - ngày 07_06_2022 06_11.xlsx	 	6/7/2022 6:11 AM	Microsoft Excel Worksheet	1,879 KB
 Báo cáo dữ liệu - Đồ chơi - ngày 07_06_2022 08_04.xlsx	 	6/7/2022 8:04 AM	Microsoft Excel Worksheet	1,796 KB
 Báo cáo dữ liệu - Đồng hồ - ngày 07_06_2022 06_49.xlsx	 	6/7/2022 6:49 AM	Microsoft Excel Worksheet	1,802 KB
 Báo cáo dữ liệu - Giặt giũ và chăm sóc nhà cửa - ngày 07_06_2022 08_11.xlsx	 	6/7/2022 8:11 AM	Microsoft Excel Worksheet	1,775 KB
 Báo cáo dữ liệu - Giày dép nam - ngày 07_06_2022 07_03.xlsx	 	6/7/2022 7:03 AM	Microsoft Excel Worksheet	1,809 KB
 Báo cáo dữ liệu - Giày dép nữ - ngày 07_06_2022 06_53.xlsx	 	6/7/2022 6:53 AM	Microsoft Excel Worksheet	1,809 KB
 Báo cáo dữ liệu - Máy ảnh và máy quay phim - ngày 07_06_2022 06_36.xlsx	 	6/7/2022 6:37 AM	Microsoft Excel Worksheet	1,772 KB
 Báo cáo dữ liệu - Máy tính và laptop - ngày 07_06_2022 06_28.xlsx	 	6/7/2022 6:28 AM	Microsoft Excel Worksheet	1,790 KB
 Báo cáo dữ liệu - Mẹ và bé - ngày 07_06_2022 06_15.xlsx	 	6/7/2022 6:15 AM	Microsoft Excel Worksheet	1,790 KB
 Báo cáo dữ liệu - Nhà cửa và đời sống - ngày 07_06_2022 06_23.xlsx	 	6/7/2022 6:23 AM	Microsoft Excel Worksheet	1,789 KB
 Báo cáo dữ liệu - Nhà sách online - ngày 07_06_2022 07_50.xlsx	 	6/7/2022 7:50 AM	Microsoft Excel Worksheet	1,755 KB
 Báo cáo dữ liệu - Ô tô xe máy xe đạp - ngày 07_06_2022 07_47.xlsx	 	6/7/2022 7:47 AM	Microsoft Excel Worksheet	1,786 KB
 Báo cáo dữ liệu - Phụ kiện và trang sức nữ - ngày 07_06_2022 07_29.xlsx	 	6/7/2022 7:29 AM	Microsoft Excel Worksheet	1,798 KB
 Báo cáo dữ liệu - Sắc đẹp - ngày 07_06_2022 06_31.xlsx	 	6/7/2022 6:32 AM	Microsoft Excel Worksheet	1,792 KB
 Báo cáo dữ liệu - Sức khỏe - ngày 07_06_2022 06_45.xlsx	 	6/7/2022 6:45 AM	Microsoft Excel Worksheet	1,804 KB
 Báo cáo dữ liệu - Thể thao và du lịch - ngày 07_06_2022 07_37.xlsx	 	6/7/2022 7:37 AM	Microsoft Excel Worksheet	1,807 KB
 Báo cáo dữ liệu - Thiết bị điện gia dụng - ngày 07_06_2022 07_25.xlsx	 	6/7/2022 7:25 AM	Microsoft Excel Worksheet	1,799 KB
 Báo cáo dữ liệu - Thiết bị điện tử - ngày 07_06_2022 06_19.xlsx	 	6/7/2022 6:19 AM	Microsoft Excel Worksheet	1,796 KB
 Báo cáo dữ liệu - Thời trang nam - ngày 07_06_2022 05_47.xlsx	 	6/7/2022 5:47 AM	Microsoft Excel Worksheet	1,844 KB
 Báo cáo dữ liệu - Thời trang nữ - ngày 07_06_2022 05_52.xlsx	 	6/7/2022 5:52 AM	Microsoft Excel Worksheet	1,804 KB
 Báo cáo dữ liệu - Thời trang trẻ em - ngày 07_06_2022 07_59.xlsx	 	6/7/2022 7:59 AM	Microsoft Excel Worksheet	1,801 KB
 Báo cáo dữ liệu - Túi ví nữ - ngày 07_06_2022 07_12.xlsx	 	6/7/2022 7:12 AM	Microsoft Excel Worksheet	1,808 KB
 Báo cáo dữ liệu - Voucher và dịch vụ - ngày 07_06_2022 08_19.xlsx	 	6/7/2022 8:19 AM	Microsoft Excel Worksheet	1,743 KB

Picture 2: All the raw data files before importing and cleaning

### 3.2. DATA IMPORTING AND CLEANING

The importing and cleaning process involves in utilizing both the Power Query and Excel functions. Firstly, raw data is imported and put together into a table using Power Query. From the tab Data on the tools ribbon, select Get Data > From File > From Folder to import all the raw data files from their containing folder



This function allows for a mass import of all the data files in one folder, the result is then presented in the Power Query Editor window as follows

Using this tool, many cleaning steps have been taken. They are summarized on the right side of this windows in the APPLIED STEPS field. These steps include

- Rearranging column order for a more convenient analysis later

- Decomposing the column Posted Date into Posted Day, Posted Month, and Posted Year to formulate different time clusters which can be used to investigate how the product posting activity changes over time
- Deleting unnecessary columns such as the index of each raw data file, No. View (no view data has been recorded), and the column containing the Shopee link to the respective product
- Removing duplicates if existed

The data is then loaded to an Excel sheet where further cleaning steps are carried out:

Tên sản phẩm	Doanh mục sơ số	Giá	Giảm	Bã bán	Trong tháng	Tôn	Ngày đăng	Tháng đăng	Năm đăng	Số ngày đăng	Comment	Like	Rating	Doanh thu tháng	Doanh thu năm
Túi Tote Canvas Nhiều Màu Lựa Chọn Size 33x38	Balo và Túi Ví Nam	62000	0	5614	284	6281	10	11	2018	1304	684	720	5	17680000	348068
Túi Đeo Chéo Nam Nữ Thời Trang Leonardo - Zumba Bag - Vải Sợi Tổng Hợp Trượt Nước	Balo và Túi Ví Nam	160000	0.43	2888	142	10	26	1	2021	497	1197	549	5	22720000	462080
Ví Cầm Tay Clutch Nam Cao Cấp Da Nappa Mềm Đẹp Nhiều Ngăn Đựng Ipad Điện Thoại Tiền	Balo và Túi Ví Nam	100000	0.41	3347	107	6207	5	1	2021	517	711	2556	5	10700000	334700
Túi đeo chéo hông bụng chạy bộ thể thao nam nữ mini chống thấm nước 3 ngăn nhỏ gọn cực	Balo và Túi Ví Nam	45000	0.5	2402	0	32	12	8	2020	664	521	146	5	0	108090
Túi Laptop Dell 15.6 Inch	Balo và Túi Ví Nam	45000	0.36	3263	36	246	6	10	2018	1340	508	234	5	1620000	146835
Ví đựng tiền ví vải lanh cotton cổ điển	Balo và Túi Ví Nam	8459	0.42	963	35	326	16	1	2020	872	415	112	5	296065	8146
Túi đeo chéo vải canvas giá rẻ đi học đựng vừa A4 phong cách hàn quốc TV100	Balo và Túi Ví Nam	55000	0.31	4470	129	415	9	11	2021	209	2115	392	5	7095000	245850
[Hàng Loại 1] Túi Đeo Chéo Bụng Da Bò 4U Cao Cấp Cá Tính DB286	Balo và Túi Ví Nam	862290	0.33	5644	0	21356	5	11	2021	213	7	91	5	0	4866764
Áo mưa balo túi trùm balo Hiee dành cho balo thể tích dưới 30L - chống nước chống bụi - có q	Balo và Túi Ví Nam	45000	0.5	89	7	451	26	2	2021	466	28	6484	5	315000	4005
BALO ULZZANG TEEN THỜI TRANG THẾ THẢO - Túi CẤP XÁCH ĐI HỌC UNISEX CHO NAM - NỮ	Balo và Túi Ví Nam	199000	0.33	3362	1	20048	20	7	2021	322	600	6295	5	199000	669038
Túi Đeo Chéo nam nữ Unisex Hàn Quốc BEE 02	Balo và Túi Ví Nam	69000	0.31	4717	80	11036	28	2	2019	1195	1971	9208	5	5520000	329473
[3 màu] Túi CHỐNG SỐC LAPTOP / MACBOOK CAO CẤP 13.5 ICNH, 15.6 INCH ☞ FREE SHIP	Balo và Túi Ví Nam	168000	0.5	1051	3	139	22	5	2020	745	401	7709	5	504000	176566
Ví Nam, Bóp Nam Cao Cấp Đựng Tiền [FullBox] 6 màu, siêu nhiều ngăn để tiền để tiện du	Balo và Túi Ví Nam	136000	0.15	1803	67	210	12	10	2020	602	481	4268	5	9112000	245208
Ví đựng thẻ bằng vải cotton và vải lanh phong cách Hàn Quốc cổ điển IELGY	Balo và Túi Ví Nam	40000	0.22	636	0	3981	27	3	2020	801	300	141	5	0	25440
Ví Cầm Tay Da Bò Cao Cấp Hàng Hiệu POKO MEISDO Da Bò Thật 100%	Balo và Túi Ví Nam	1800000	0	303	0	4710	8	10	2021	241	51	4340	5	0	545400
Túi Đựng Card Visit, Atm, Cmnd Trong Suốt	Balo và Túi Ví Nam	1300	0.35	4164	435	7817	24	9	2020	621	290	37	5	565500	5413
Ví nam Beallerry Cao Cấp Cam Kết Bảo Hành 1 đổi 1, bóp ví da nam may viền sống lưng Phong	Balo và Túi Ví Nam	160000	0.38	5609	4	319	12	8	2021	298	923	69	5	640000	897440
Balo da nam nữ unisex Lohi thời trang ulzzang cao cấp hàn quốc đựng laptop phù hợp cho s	Balo và Túi Ví Nam	385000	0.26	1246	5	13	13	3	2021	450	629	682	5	1925000	479710
40x30cm Cặp da năng[Loại vải vừa] 8 ngăn dành cho học sinh giáo viên công sở, Cặp Laptop L	Balo và Túi Ví Nam	148000	0.44	379	14	16	5	2	2020	852	165	105	5	2072000	56092
Cặp Kẹp Cá Sấu 45mm (Đỏ + Đen)	Balo và Túi Ví Nam	2500	0	1204	32	899	3	1	2020	885	174	15	5	80000	3010
Ví nam nữ bóp dài cầm tay nam chính hãng Beallerry da pu cao cấp	Balo và Túi Ví Nam	95000	0	1207	34	20	31	10	2019	950	481	153	5	3230000	114665
Cặp 12 ngăn A4 Deli 5556	Balo và Túi Ví Nam	80000	0	779	30	1905	2	4	2020	796	239	114	5	2400000	62320
[XÁ KHO] Balo nữ chống nước đựng laptop, sách vở YYD BAG MYFAVORITE hot 2020	Balo và Túi Ví Nam	199000	0.01	6566	15	2410	11	6	2020	725	1800	5863	5	2985000	1306634
[CÓ KÈM HỘP] Túi Đeo Chéo Ví Cầm Tay Chữ X	Balo và Túi Ví Nam	80000	0.23	1919	149	1027	6	4	2021	427	839	372	5	11920000	153520
Ví nam da mềm dáng đứng, ngang cầm tay fullbox	Balo và Túi Ví Nam	119000	0	2728	29	126	1	6	2020	735	1112	577	5	3451000	324632
Túi chống sốc Laptop Macbook 2 ngăn to, quai xách quai đeo 2020	Balo và Túi Ví Nam	285000	0	520	34	87	14	9	2020	630	183	1132	5	9690000	148200
Balo thời trang phong cách hàn quốc HARRAS HR238	Balo và Túi Ví Nam	139000	0.47	6421	24	1055	18	2	2019	1205	2895	1247	5	3816000	1020935
Ví Đựng Thời Trang HARRAS HR264	Balo và Túi Ví Nam	19000	0.34	2908	74	3862	18	9	2020	626	1346	287	5	1406000	55252
Bóp Ví Nam Đựng Thẻ Thời Trang	Balo và Túi Ví Nam	180000	0	1747	0	2949	20	3	2020	809	59	24	5	0	314460
[DA THẬT] Ví da nam, ví dáng đứng nhỏ gọn, da Saffiano cao cấp Cosmos - thời trang OLUX - sả	Balo và Túi Ví Nam	249000	0.17	1985	157	1283	26	2	2021	465	239	336	5	39093000	494265
[DA THẬT] Ví nam, bóp nam nhỏ gọn cao cấp Cosmos da nhập khẩu - thời trang OLUX - sản p	Balo và Túi Ví Nam	239000	0.2	3882	290	40	26	2	2021	465	445	530	5	69310000	927798
Túi đeo chéo nam 3 ngăn, túi nam đeo chéo nam có lỗ tai nghe tiện lợi cách Hàn Quốc	Balo và Túi Ví Nam	125000	0.17	1096	11	116	13	11	2021	206	210	152	5	1375000	137000
Cặp đựng Laptop Dành Cho Máy 14 Inch - 15.6 Inch	Balo và Túi Ví Nam	70000	0	259	0	14	23	2	2019	1199	88	4725	5	0	18130
Túi đeo chéo nam vải bố thời trang	Balo và Túi Ví Nam	89000	0.35	8671	13	10	19	11	2019	930	3437	909	5	1157000	771715
Cặp sách đeo chéo vải chống thấm chống bám thời trang BAG U FS125	Balo và Túi Ví Nam	119000	0.29	537	28	44	11	5	2021	391	271	637	5	3332000	63903
Balo nam Đi Học Hàn Quốc, Balo Thời Trang Nam Công Nghệ Đựng lap top 16inch	Balo và Túi Ví Nam	250000	0	13144	2	6485	8	11	2019	942	736	1582	5	500000	3286000

These steps include

- Translating column names to English
- Translating the column `Primary Category` into English using the Sort & Filter > Advanced function in the Data tab to create the list of unique values then translate this list accordingly, after that using the XLOOKUP() function to look up the corresponding English translation of the listed categories in Vietnamese
- Adding the column Index
- Adding the column posted date using the DATE() function

The data is then properly formatted and put into a table for further manipulations. Then we have the final data, which is presented in the picture below. The final data has eighteen columns, including

- Index
- Product Description VIE
- Primary Category
- Posted Date
- Posted Day
- Posted Month
- Posted Year
- Price
- Discount
- Sales
- Sales Last Month
- Inventory



- Posted Duration
- No. Comment
- No. Like
- Rating
- Revenue Last Month
- Total Revenue

AutoSave

On

Project Excel 1 - Part 3 Sales Analysis Dashboard.xlsm • Saved ▾

Search (Alt+Q)

Mai Xuân Chính

File

Home

Insert

Draw

Page Layout

Formulas

Data

Review

View

Developer

Help

Power Pivot

Table Design

Query

Item ID

Product Description V6

Privacy Category

Posted Date

Posted Day

Posted Month

Posted Year

Price

Discount

Sales

Sales Last Month

Inventory

Posted Duration

No. Comment

No. Like

Rating

Revenue Last Month

Total Revenue

7	Túi đeo chéo vải canvas giá rẻ đi h Men Bag		11/9/2021	9	11	2021	VND	55,000.00	31%	4,470	129	415	209	2,115	392	5	VND	7,095,000.00	VND	245,850,000.00
8	Hàng Loại 1) Túi Đeo Chéo Đựng D Men Bag		11/5/2021	5	11	2021	VND	862,290.00	33%	5,644	-	21,356	213	7	91	5	VND	-	VND	4,866,764,760.00
9	Áo mưa bảo túi trùm bảo hieo da Men Bag		2/26/2021	26	2	2021	VND	45,000.00	50%	89	7	451	466	28	6,484	5	VND	315,000.00	VND	4,005,000.00
10	BALO UZZANG TIỀN THỜI TRANG TI Men Bag		7/20/2021	20	7	2021	VND	199,000.00	33%	3,362	1	20,048	322	600	6,295	5	VND	199,000.00	VND	699,036,000.00
11	Túi Đeo Chéo nam nữ Unisex Hàn Men Bag		2/28/2019	28	2	2019	VND	69,000.00	31%	4,717	80	11,036	1,195	1,971	9,208	5	VND	5,520,000.00	VND	325,473,000.00
12	5 màu) Túi CHỐNG SỐC LAPTOP / M Men Bag		5/22/2020	22	5	2020	VND	168,000.00	50%	1,051	3	139	745	401	7,709	5	VND	504,000.00	VND	176,568,000.00
13	Ví Nam, Bóp Nam Cao Cấp Đựng Ti Men Bag		10/12/2020	12	10	2020	VND	136,000.00	15%	1,803	67	210	602	481	4,268	5	VND	9,112,000.00	VND	245,208,000.00
14	Ví đựng thẻ bằng vải cotton và ví Men Bag		3/27/2020	27	3	2020	VND	40,000.00	22%	636	-	3,981	801	300	141	5	VND	-	VND	25,440,000.00
15	Ví cầm tay da Bò Cao Cấp Hàng Hi Men Bag		10/8/2021	8	10	2021	VND	1,800,000.00	0%	303	-	4,710	241	51	4,340	5	VND	-	VND	545,490,000.00
16	Túi Đựng Card Visit, Atm, Cmnid Tro Men Bag		9/24/2020	24	9	2020	VND	1,300.00	35%	4,164	435	7,817	621	290	37	5	VND	565,500.00	VND	5,413,200.00
17	Ví nam Beallery Cao Cấp Cam KẾT Men Bag		8/12/2021	12	8	2021	VND	160,000.00	38%	5,609	4	319	298	923	69	5	VND	640,000.00	VND	897,440,000.00
18	Balo da nam nữ unisex Loho thời t Men Bag		3/13/2021	13	3	2021	VND	385,000.00	26%	1,246	5	13	450	629	682	5	VND	1,925,000.00	VND	479,710,000.00
19	40x40cm Cặp da nân(Loại vải vớ) Men Bag		2/5/2020	5	2	2020	VND	148,000.00	44%	379	14	16	852	165	105	5	VND	2,072,000.00	VND	56,092,000.00
20	Cáp Xếp Cá Sấu 5mm (Bỏ + Dey) Men Bag		1/5/2020	5	1	2020	VND	2,500.00	0%	1,204	32	899	885	174	15	5	VND	80,000.00	VND	3,010,000.00
21	Ví nam nữ bóp dài cầm tay nam chi Men Bag		10/31/2019	31	10	2019	VND	95,000.00	0%	1,207	34	20	950	481	153	5	VND	3,230,000.00	VND	114,665,000.00
22	Cặp 12 ngăn A4 Deli 5556 Men Bag		4/2/2020	2	4	2020	VND	80,000.00	0%	779	30	1,905	796	239	114	5	VND	2,400,000.00	VND	62,320,000.00
23	(XÁ KHO) Balo nữ chống nước đựng Men Bag		6/11/2020	11	6	2020	VND	199,000.00	1%	6,566	15	2,410	725	1,800	5,863	5	VND	2,985,000.00	VND	1,306,634,000.00
24	(CỎ KEM HỒI) Túi Đeo Chéo Vĩ Cầm Men Bag		4/6/2021	6	4	2021	VND	80,000.00	23%	1,919	149	1,027	427	839	372	5	VND	11,920,000.00	VND	153,520,000.00
25	Ví nam da miễn đồng đựng, ngang Men Bag		6/1/2020	1	6	2020	VND	119,000.00	0%	2,738	29	126	735	1,112	377	5	VND	3,651,000.00	VND	334,632,000.00
26	Túi chống sốc Laptop Macbook 2 ng Men Bag		9/14/2020	14	9	2020	VND	285,000.00	0%	520	34	87	630	183	1,132	5	VND	9,690,000.00	VND	148,200,000.00
27	Balo thời trang phong cách hàn qu Men Bag		2/18/2019	18	2	2019	VND	159,000.00	47%	6,421	24	1,055	1,205	2,895	1,247	5	VND	3,816,000.00	VND	1,020,939,000.00
28	Ví Đựng Thoi Trang HARAS HRS26A Men Bag		9/18/2020	18	9	2020	VND	19,000.00	34%	2,908	74	3,862	626	1,346	287	5	VND	1,406,000.00	VND	55,252,000.00
29	Bóp Vĩ Nam Đựng Thê Thời Trang Men Bag		3/20/2020	20	3	2020	VND	180,000.00	0%	1,747	-	2,949	809	59	24	5	VND	-	VND	314,460,000.00
30	DA THẬT Vĩ da nam, Vĩ đồng đựng Men Bag		2/26/2021	26	2	2021	VND	249,000.00	17%	1,985	157	1,283	465	239	336	5	VND	39,093,000.00	VND	494,255,000.00
31	DA THẬT Vĩ nam, bóp nam nhỏ nh Men Bag		2/26/2021	26	2	2021	VND	239,000.00	20%	3,882	290	40	465	445	530	5	VND	69,310,000.00	VND	927,798,000.00
32	Túi đeo chéo nam 3 ngăn, túi nam Men Bag		11/13/2021	13	11	2021	VND	125,000.00	17%	1,096	11	116	206	210	152	5	VND	1,375,000.00	VND	137,000,000.00
33	Cáp đựng Laptop Dành Cho Máy 14 Men Bag		2/23/2019	23	2	2019	VND	70,000.00	0%	259	-	14	1,199	88	4,725	5	VND	-	VND	18,130,000.00
34	Túi đeo chéo nam và nữ thời trang Men Bag		11/19/2019	19	11	2019	VND	89,000.00	35%	8,671	13	10	930	3,437	909	5	VND	1,157,000.00	VND	771,719,000.00
35	Cáp sạc đeo chéo và chống thie Men Bag		5/11/2021	11	5	2021	VND	119,000.00	29%	937	28	44	391	271	637	5	VND	3,332,000.00	VND	68,901,000.00
36	Balo nam Đi Học Hàn Quốc, Balo TI Men Bag		11/8/2019	8	11	2019	VND	250,000.00	0%	13,144	2	6,485	942	736	1,582	5	VND	500,000.00	VND	3,286,000,000.00
37	Balo laptop nam 15 6 inch chổn Men Bag		10/5/2021	5	10	2021	VND	115,000.00	43%	1,391	8	80	245	498	179	5	VND	920,000.00	VND	159,965,000.00
38	✓ Túi Đeo Hồng Nhiều Ngăn Nhỏ Men Bag		6/2/2019	2	6	2019	VND	80,000.00	39%	3,193	2	4,329	1,101	811	350	5	VND	160,000.00	VND	235,440,000.00
39	Balo Da Nân Laptop, Thời Trang, C Men Bag		11/28/2018	28	11	2018	VND	179,000.00	28%	1,516	-	796	1,287	780	21,651	5	VND	-	VND	271,364,000.00
40	(HÀ FAMILI) Túi đựng 15m đơn 150x Men Bag		8/23/2017	23	8	2017	VND	119,000.00	40%	3,079	4	1,497	1,749	1,211	3,648	5	VND	476,000.00	VND	366,401,000.00
41	Túi chống sốc laptop cao cấp chính Men Bag		4/28/2021	28	4	2021	VND	337,500.00	33%	399	21	4,040	405	172	3,548	5	VND	7,087,500.00	VND	134,662,500.00
42	Vĩ Nam Da Bò Thật 100% Cao Cấp Sĩ Men Bag		12/18/2019	18	12	2019	VND	215,000.00	7%	25,869	2	13,076	901	278	1,545	5	VND	430,000.00	VND	5,561,835,000.00
43	Vĩ Nam Thời Trang Đẹp / Bóp Da K Men Bag		9/7/2020	7	9	2020	VND	39,000.00	0%	25,179	8	977	637	220	80	5	VND	312,000.00	VND	981,981,000.00
44	(da thật) vĩ namcard mini da thật Men Bag		10/13/2020	13	10	2020	VND	179,000.00	38%	394	1	2,561	601	78	4,509	5	VND	179,000.00	VND	70,526,000.00
45	Túi Đeo Hồng, Balo Đeo Ngang Đur Men Bag		10/1/2020	1	10	2020	VND	39,000.00	0%	2,860	22	62	613	191	22,009	5	VND	858,000.00	VND	111,540,000.00
46	Túi ĐUNG ATM, CMND,CARD VISIT, Men Bag		7/2/2021	2	7	2021	VND	2,000.00	50%	5,592	45	4,306	369	124	413	5	VND	90,000.00	VND	11,184,000.00
47	Túi HAI QUAI đeo càm Men Bag		5/18/2019	18	5	2019	VND	75,000.00	0%	2,164	-	20	1,115	659	938	5	VND	-	VND	162,300,000.00
48	Túi đeo chéo da nam nữ unisex chi Men Bag		6/9/2020	9	6	2020	VND	59,000.00	50%	307	7	722	727	136	27	5	VND	413,000.00	VND	18,113,000.00
49	Vĩ Nam Đựng Da Bò Đẳng Cấp Tầng Men Bag		2/27/2021	27	2	2021	VND	345,000.00	3%	34,219	1	7,397	464	6	10	5	VND	345,000.00	VND	11,808,533,000.00
50	Túi đeo hông chuy bộ Men Bag		3/18/2020	19	3	2020	VND	51,000.00	0%	331	-	249	820	73	1,090	5	VND	-	VND	16,481,000.00
51	Bỏ vẽ sinh Laptop, Máy sinh TIKA-S Men Bag		5/6/2021	6	5	2021	VND	40,000.00	0%	338	5	156	273	108	11	5	VND	100,000.00	VND	6,760,000.00
52	Túi đeo chéo nam thời trang CHEN Men Bag		8/29/2018	29	8	2018	VND	49,000.00	35%	5,158	24	464	1,377	2,511	15,043	5	VND	1,176,000.00	VND	252,742,000.00
53	Vĩ Nam Hàn Quốc Chất Da Bò Size 9 Men Bag		8/24/2020	24	8	2020	VND	104,500.00	50%	21,147	8	285,137	651	269	3,489	5	VND	836,000.00	VND	2,209,861,500.00
54	Túi chống sốc Laptop máy tính Mac Men Bag		1/16/2019	16	1	2019	VND	289,000.00	0%	2,583	76	236	1,237	894	4,703	5	VND	21,964,000.00	VND	746,487,000.00
55	(FRESH) Balo laptop Nam nữ Th Men Bag		5/28/2019	28	5	2019	VND	129,000.00	39%	5,374	1	1,166	1,570	30,451	5	VND	129,000.00	VND	435,246,000.00	
56	Vĩ Nam Đựng Da Bò Thật Đẳng Cấp Men Bag		11/16/2021	16	11	2021	VND	250,000.00	49%	20,333	-	2,222	202	4	9	5	VND	-	VND	5,083,250,000.00
57	Vĩ Xách Du Lịch Vải Dũ Cao Cấp (r Men Bag		4/27/2019	27	4	2019	VND	109,000.00	34%	6,041	4	28	1,136	1,920	1,132	5	VND	436,000.00	VND	658,469,000.00
58	Vĩ Da Nam Chính Hãng Beallery Sĩ Men Bag		3/24/2021	24	3	2021	VND	265,000.00	27%	29,419	-	2,022	440	8	11	5	VND	-	VND	7,796,035,000.00

Final Data

Questions

Dashboard

Ready

85%

These columns are categorized into separate groups featured by their colors for the presentation purpose. For this analysis, I want to focus on the Total Revenue, so it is highlighted by orange.

# 4. ANALYSIS DETAILS

Before start doing any analysis, it is important to define the questions the analysis needs to answer. For this analysis, the case description has stated the needs of Hal, but this information is still too general and needs to be articulated more clearly. The main points can be summarized shortly as follows:

- Which factors in Price, Discount, Number of Comments, Number of Likes, and Rating have the significant effect on the Total Revenue of products posted on the E-Commerce platform Shopee until June 7, 2022?
- How is Total Revenue distributed among Sales Categories of the products posted on the E-Commerce platform Shopee until June 7, 2022?
- How does the number of posted products on Shopee until June 7, 2022, changes throughout a year? And from year to year?
- For each Category of the products posted on Shopee until June 7, 2022, what are the Total Revenue, Last Month Revenue, and Distribution of Rating?

In this analysis, Sales statistics will be ignored as its information has already been conveyed by Total Revenue. It should be noted that the data set does not contain data for sales or revenue over time, but it only has the date when a product is posted. This data can be aggregated to illustrate the pattern of posting product on Shopee over time.

The first question can be answered by running a regression model of Total Revenue on Price, Discount, Number of Comments, Number of Likes, and Rating. This is a simple multiple regression model that has the following form

$$\text{Total Revenue} = \beta_0 + \beta_1 \times \text{Price} + \beta_2 \times \text{Discount} + \beta_3 \times \text{No. Comment} + \beta_4 \times \text{No. Like} + \beta_5 \times \text{Rating}$$

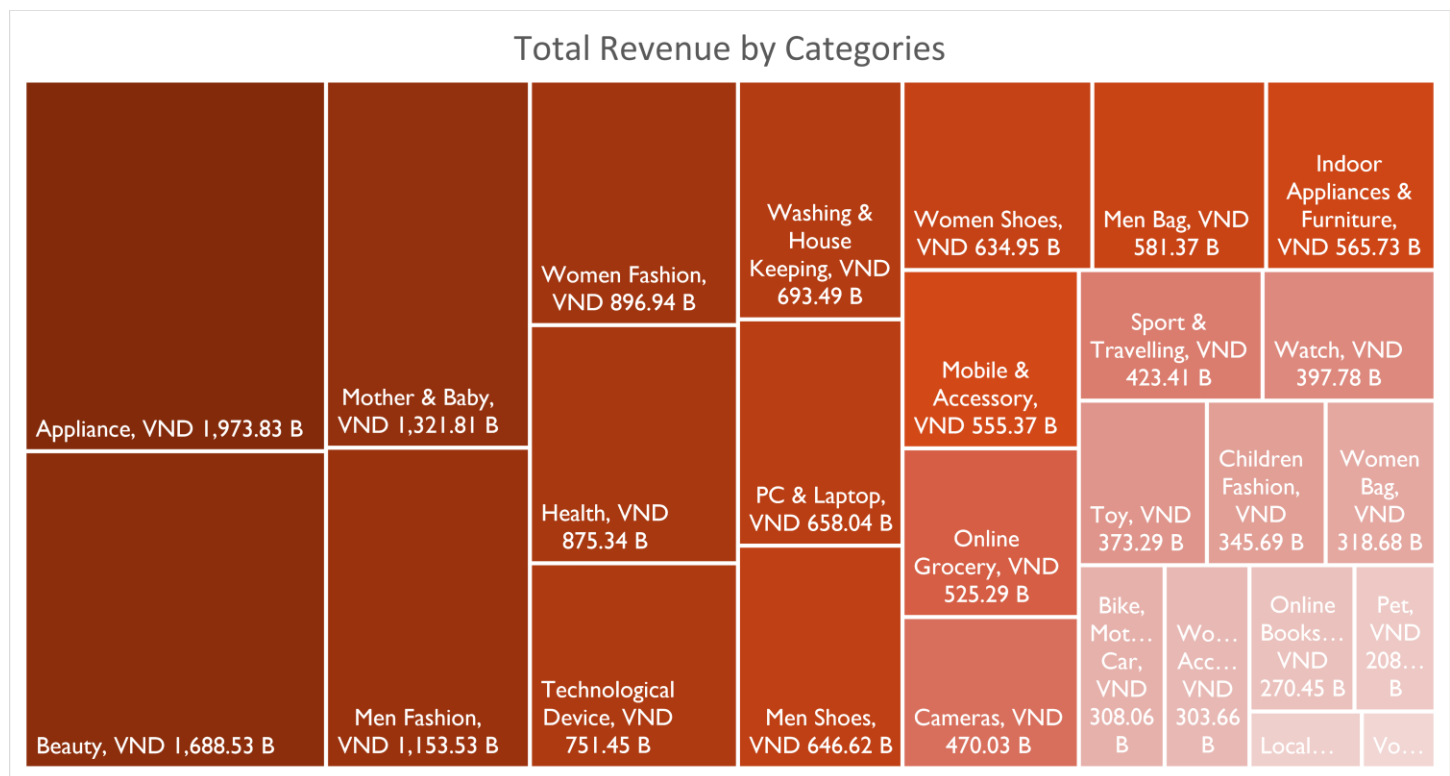


Where  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$  and  $\beta_5$  are respectively the intercept, effects of Price, Discount, No. Comment, No. Likes, and Rating. Since this multiple regression is still quite simple, it can be handled well by Excel. Using the Data Analysis function in the Data ribbon, the regression result is summarized in the following table

Variable	Coefficients	Standard Error	t-Stat	P-value	Lower 95%	Upper 95%
Intercept	(36,688,510.80)	19,225,143.68	(1.91)	0.06	(74,369,679.06)	992,657.46
Price	46.98	2.26	20.81	0.00	42.56	51.41
Discount	26,464,359.98	17,665,273.79	1.50	0.13	(8,159,472.48)	61,088,192.44
No. Comment	132,607.95	4,482.53	29.58	0.00	123,822.21	141,393.68
No. Like	6,954.91	709.93	9.80	0.00	5,563.45	8,346.36
Rating	35,466,020.25	3,888,292.03	9.12	0.00	27,844,990.80	43,087,049.69

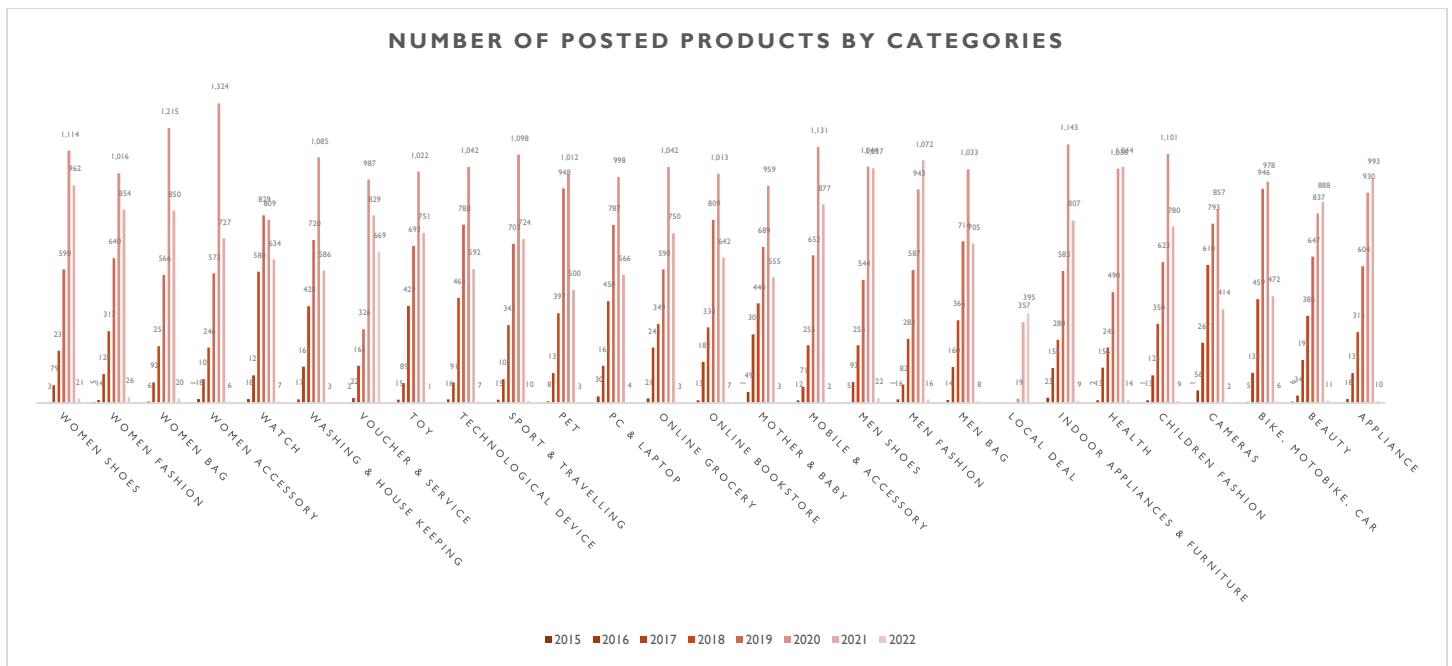
It can be seen from the table that the effect of Price, No. Comment, No. Like and Rating are all statistically significant since their P-value is almost zero. However, the effect of discount is not significant as its P-value is larger than 0.1.

The distribution of Total Revenue among Sales Categories of the posted products can be presented by a tree map. Using a pivot table to aggregate data and construct the map from this data (actually, the version of Excel being used in this analysis still does not allow for the direct construction of a tree map from a pivot table, so a workaround has been used), we have the following result

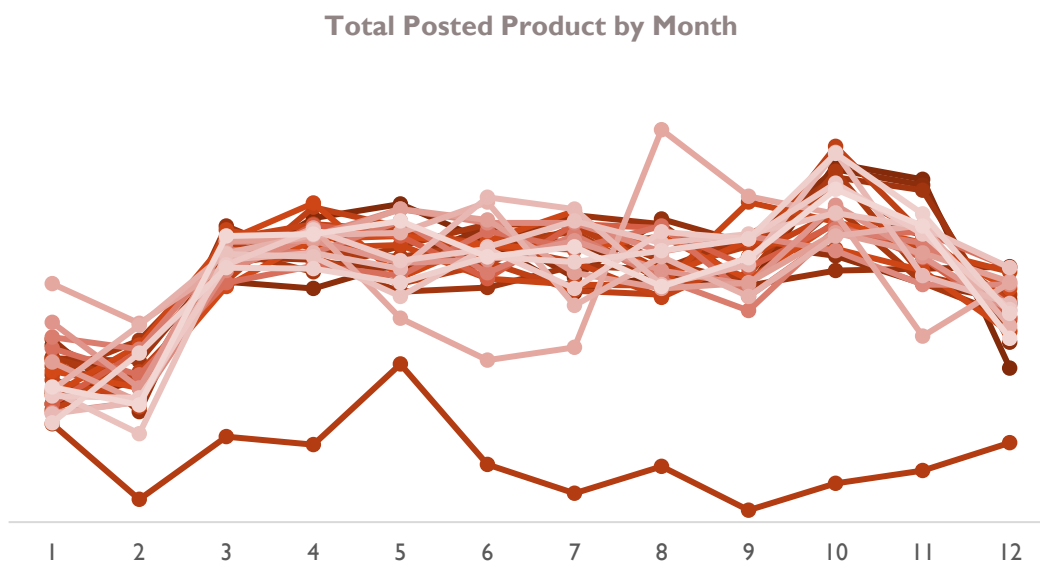


Tree map is a proper presentation for this type of information besides bar chart. In this case, I would prefer tree map to bar chart because it shows which category has the highest Total Revenue clearly (as Western People usually read from left to right, and top to bottom) and a sense of relative magnitude between categories, while still saving lots of space due to the more compact size of its illustration. Furthermore, change in color is also used to illustrate this difference in size. From this illustration, Appliance, Beauty, and Mother & Baby are the categories that have the highest Total Revenue.

The posting pattern of products in all categories can be presented by two charts. Due to the substantial number of categories, it is more convenient to separate a supposed-to-be-one illustration into multiple charts so that they are more comfortable to look at and interpret. In this analysis, I choose to visualize the number of posted products aggregated by category, by month, and by year. This offers an insight into the posting pattern throughout the year and from year to year. Moreover, two combinations of illustration are used to present the pros and cons of each means of illustration while allowing me to utilize their strengths as well as minimize their shortcomings. With a slicer, these visualizations can be simplified to compare the patterns between two or more categories or only show the pattern of one category.



The first chart shows the number of posted products of each category from year to year. Cluster bar chart is chosen to present this information as it shows all categories clearly as well as how each category changes over time. I want to show the magnitude of the number of posted products of each category in each year, so the bar is the best option for presenting this information. It might look a bit messy at first, but the slicer will make the chart look tidier later.



The second chart shows how the number of posted products of each category changes throughout the year. Line chart is different from bar chart in such a way that it focuses more on the changes rather than the magnitude within a category. When considering the number of posted products of a category within a year, changes in the number are more important because it allows the user to choose the right time to post their product so that he or she could have tactical advantaged depending on his or her sales strategy. This is different from the number of posted products of a category from year to year as it offers a rough estimate of the market size. These different charts serve different purposes and when combined with the right slicers, they provide very insightful information about the categories of choice.

The final question can be answered easily by a combination of pivot tables. Total Revenue, Last Month Revenue, and Distribution of Rating of each category can be aggregated easily using their corresponding pivot tables. The pivot tables do not need to show the categories because they can be linked to the slicers that are used to control the tables in the previous question. When the slicer is

used, the aggregated statistics will be updated accordingly in the pivot tables of Total Revenue, Last Month Revenue, and Distribution of Rating.

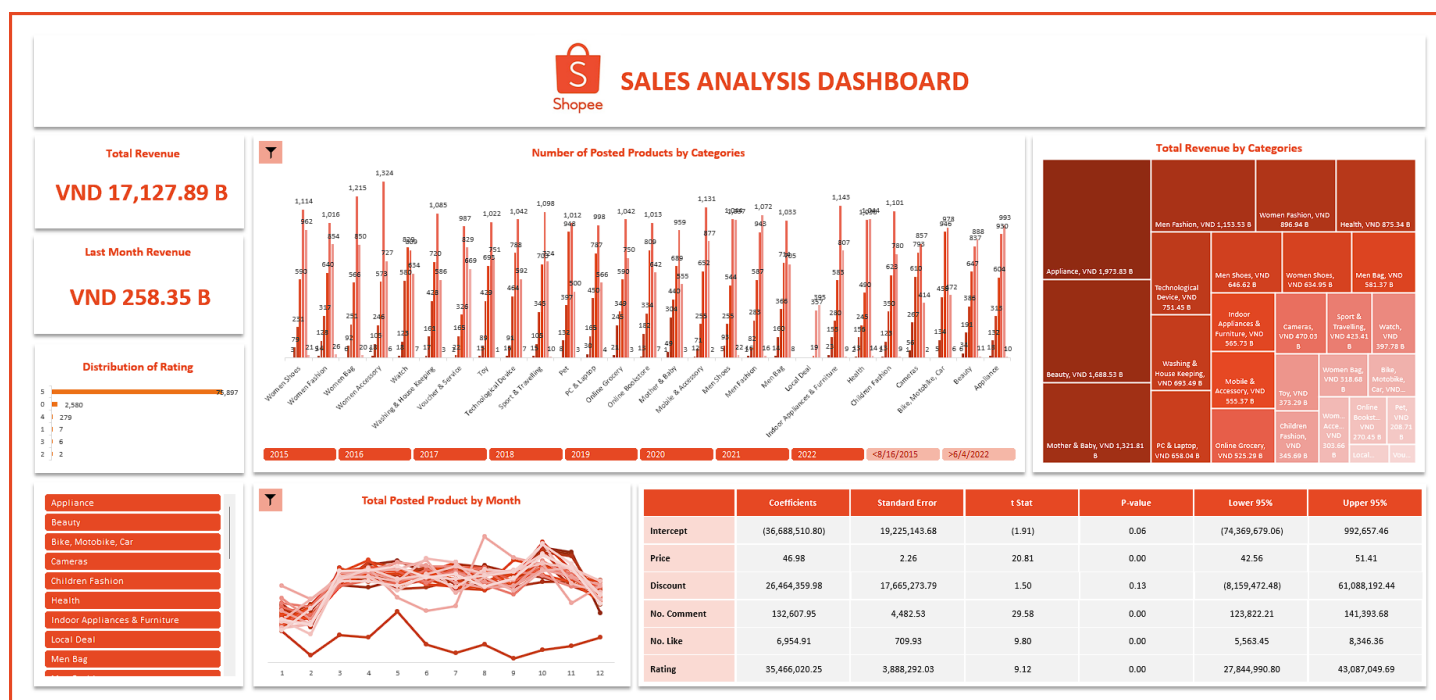
Sum of Total Revenue	
VND	17,127.89 B

Sum of Revenue Last Month	
VND	258.35 B

Row Labels	Distribution of Rating
2	2
3	6
1	7
4	279
0	2,580
5	75,897
<b>Grand Total</b>	<b>78,771</b>

The slicer is an important part of the Excel pivot table because it provides a way to interact with the information being presented, making the visualization become dynamic rather than just static figures.

These charts are then brought together to construct a dashboard to fulfill the needs of Hal, so that he can investigate further to formulate his strategy and tactical choices with the available data.



Overall, this dashboard provides information for the user to estimate the magnitude of market for each category, and important analytical information for forming strategy and making tactical decisions if he or she wants to penetrate the market. With slicers to filter categories and years, the dashboard allows the user to investigate the overall market, the market of each category as well as compare them. Regression table also provides information about the factors affecting Total Revenue, while line chart presents the patterns of how sellers in each category post their products on Shopee throughout the year. Details of the design choices are discussed in the next section of this documentation.

## 5. DESIGN CHOICES

The dashboard has many elements including information card, table, line chart, vertical bar chart, and tree map. The reasons for these choices are discussed in the previous section. In this dashboard, the alignment of objects as the first pre-attentive attribute. The dashboard has two rows, the upper is for information about the revenue, and the lower is for information about the products posting tactics. The slicers, which are used to interact with the dashboard, are placed on the left side of the dashboard. Since the Western people usually read from top to bottom, left to right, the dashboard user will encounter with important elements and information first following his or her natural eye movement.

The color choice is only to make the dashboard have the same color tone with the icon of Shopee. I did not intend to make it a pre-attentive attribute, so I used the provided color palette of Excel. The palette did a decent job on providing a consistent presentation of the color tone and it seems to give the impression that objects of the darken orange have higher values. However, color is not an effective pre-attentive attribute here because the dashboard does not need to show which values are more important than the others. Moreover, there is too much information to be convey and they seem to be equally important.

Some of the Gestalt principles are employed in the design choices to reduce mental clutter when processing the visualizations, they include:

- Continuity is used to show the changes in the data as the category change from one to another
- Enclosure is used to separate information presented by different charts

As the intended audiences are familiar with the subject and require arguments to be backed up by statistics, they will focus more on the quick, well-articulated overall trends rather than details. As they have no problem with viewing the visualization, these charts are well designed to focus their attention on important points while still providing enough information in case when they need to dive deeper into the statistics and correlation among the presented illustrations.

## 6. FINAL VISUALIZATION ANALYSIS

## 7. RELATED INFORMATION AND DATA

### 7.1. DATASET

The dataset is given together with the dashboard, the raw data is put in a separate folder.

### 7.2. CHANGELOG

**18/06/2022:** v0.1 first establishment of the dashboard

## APPENDIX AND REFERENCES

### REFERENCES