**I. Data Information**

* Process Step: Search, download, correct, structure and update data
* Attributes - Unit:
* ID : like item id, shop id, shop location,…
* Data : price, discount, stock, sold quantity,…
* Rating : rating star each comment, comment contents,…
* License: CC0: Public Domain

**II. User Story**

I made the user story table to analyze the purpose for each position. The best way to analyze, I realized that I put myself as the reader/ the audience. We analyze data to tell and make any person get our idea, not talk to our self.

The user guide I am using Agile methodology.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| USER STORY ID | PRIORITY | AS A [type of user] | I NEED TO [do some task] | SO THAT I CAN [get some result] | STATUS |
| 1 | Should have | As a director | know the total sale this category, the top 5 shop, the top comment, top brand | report all sales performance | Approved |
| 2 | Must have | As a Category Manager | know which the top 5 item, top 5 shop, top 5 brand, remain stock, store location | report category performance | Approved |
| 3 | Should have | As a supervisor | know total sales per shop/item, top % discount, which shop give highest discount, which item has highest discount | the dashboard report | Approved |
| 4 | Should have | As a Saleman | know stock of each item, which shop has stock | the storage board | Approved |

**III. Data Visualization**

I am using Power Bi to clean data and visualize data. The positive of Power Bi is ***required the outer should login to Power Bi service*** before read this. It is not a barrier because I can ***provide my individually account to you***, but I suggest you contact to me via my email or phone number. Futhermore, I willing to open with HR headhunt so please contact to me if you interested.

1. **Data Preparing**

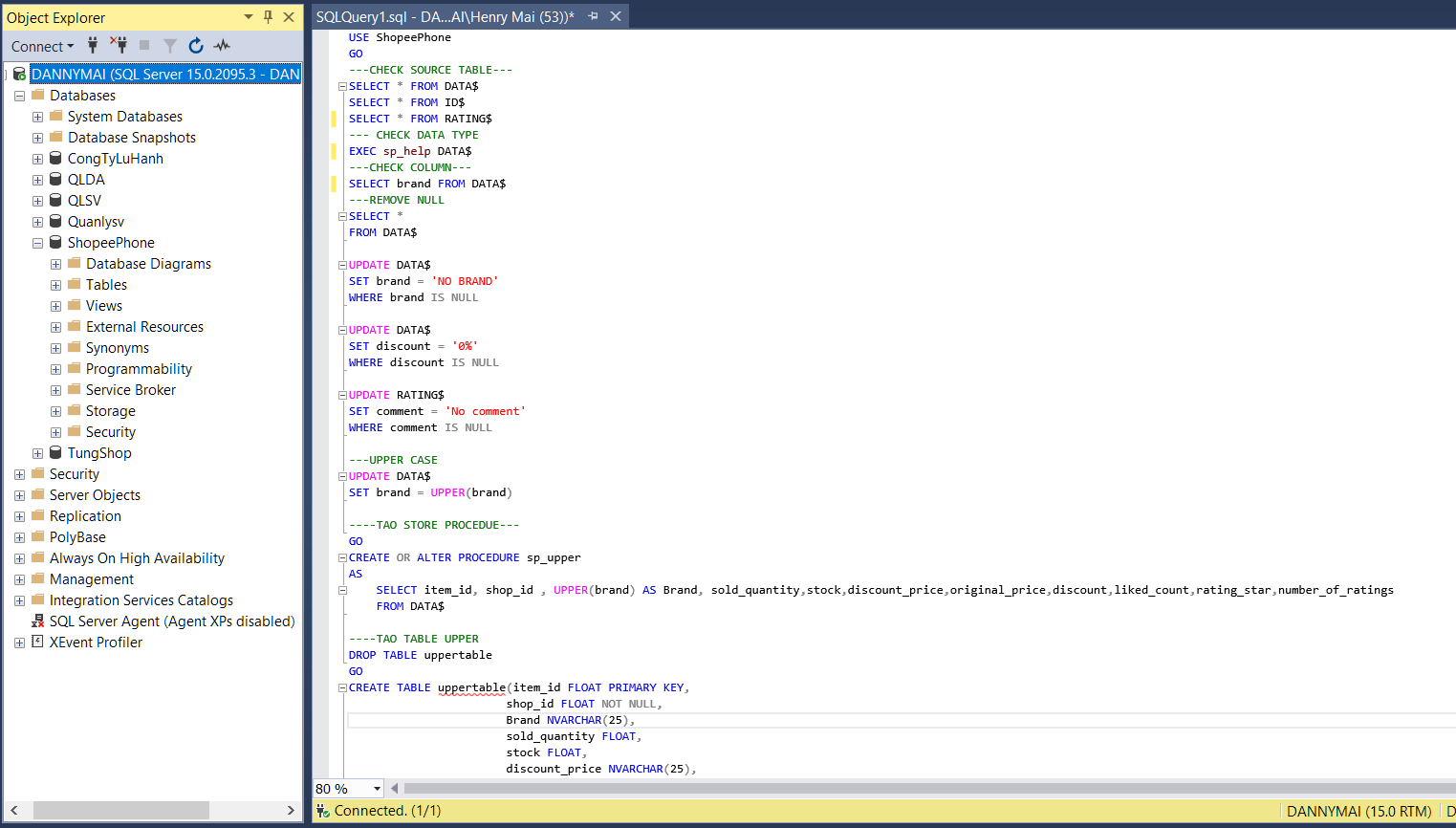
After I deeply check this data, I realized it need to make some tasks for preparing data before using this. First one I suggest using SQL to query in deep data then push it into Python to analyzing and remove error text in some field.

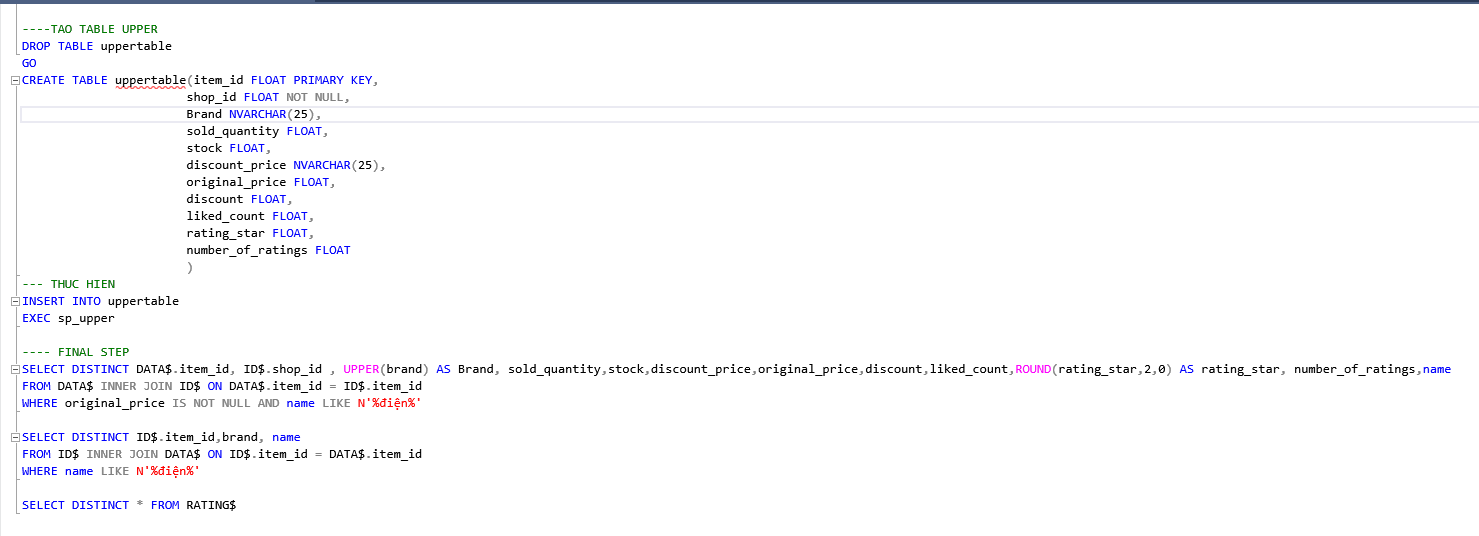
The Second steps I will use Azure Machine Learning to find out outliner number solution and resolve in data if data have.

Finally, we have the cleaned data to use and analyst by Power Bi as the result I show off above.

1. **SQL programing**

I am using SQL to extract and analyzing data. The data is not too big, so I can do some step driven data.

****

****

**B. Python**

We can see that the field name in Table ID$ is phone model but it definitely blinds text because of fatal input. I recommend using Python in order create the list official phone model to compare with field name in table ID$. Example We create two list, the one is all official phone model and the second list appended text in field name in table ID then we use pop function of Python.

That are my ideas but unfortunately, I focus on analyst using PBI, SQL and Azure ML in this post so I skip this step. I will re-do this in near future.

**C. Azure Machine Learning**

In stead of using Python – Jupyter to analyzing outliner in data, I use a tool created by Microsoft, Azure Machine Learning.

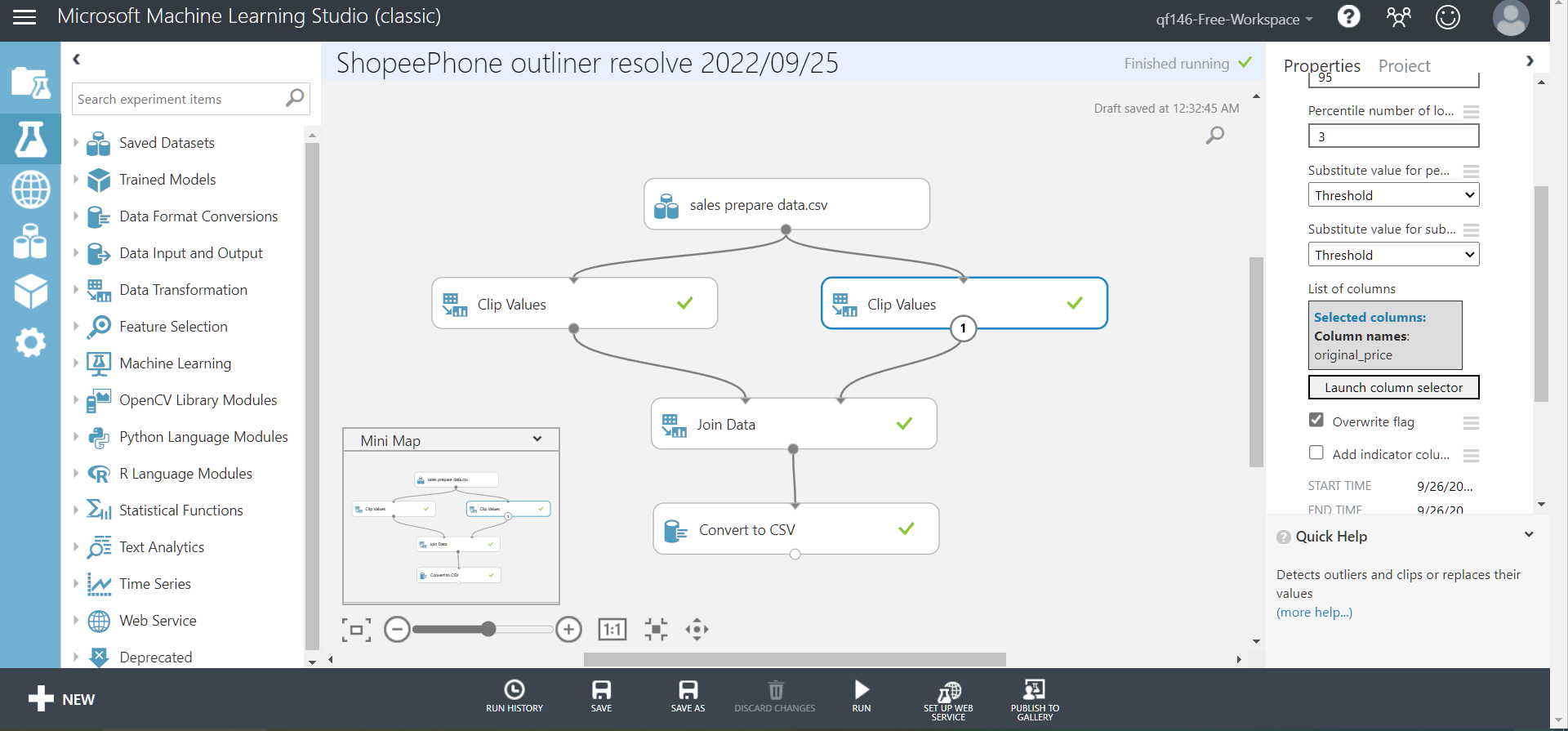
First step, I put dataset into model platform. I visualize this data and find out column original\_price and discount\_price on Data sheet has much outliner. My target is resolve this.

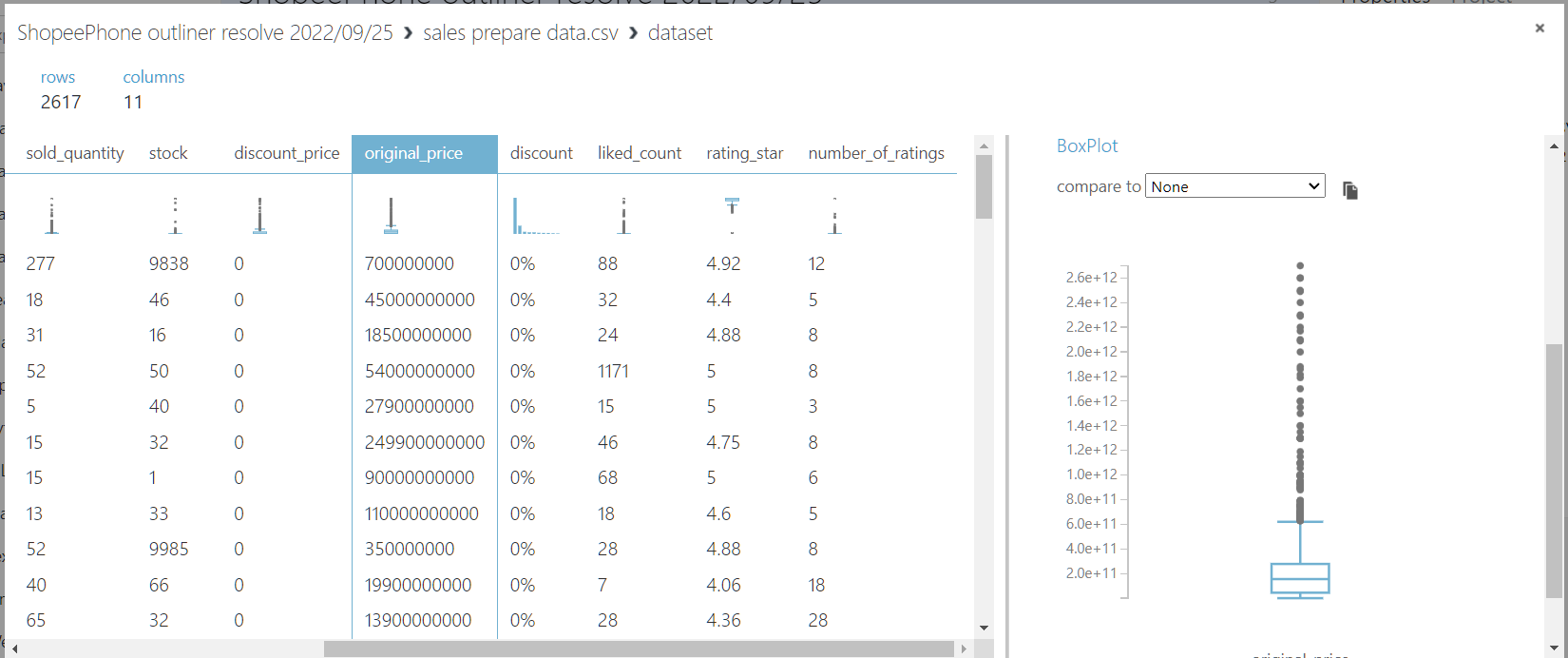
Second, I am using Clip Value tool to resolve outliner issuse. The method is cut out dirty data as the outliner data in peaks and subpeaks. I divived dataset into 2 parts because original\_price and discount\_price should be cut as two types.

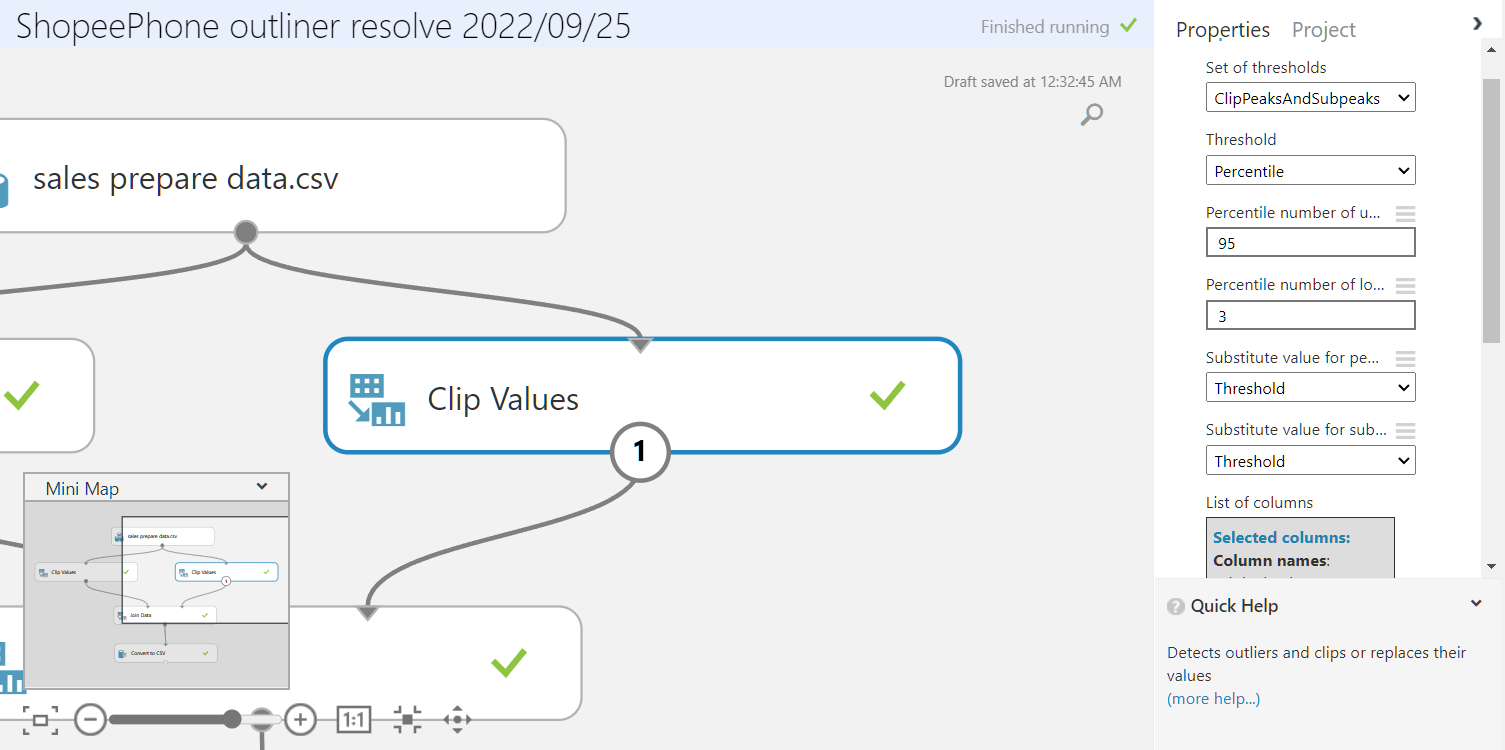
* Original\_price column should clip the peaks and subpeaks data by percentile number. I put it 95% and 3%.
* Discount\_price column should clip the peaks data, it has not dirtied data at below. I put it 95%.

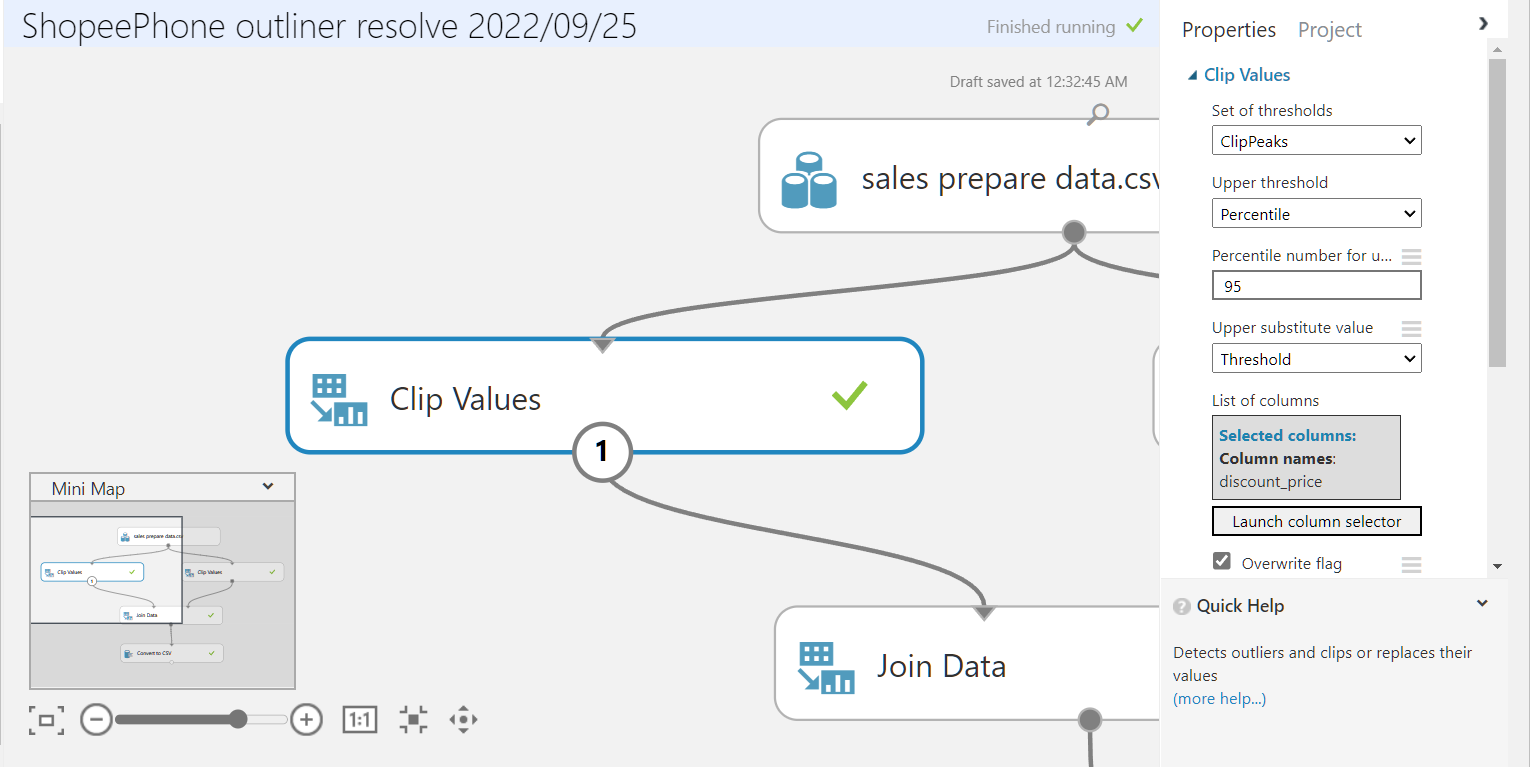
Third, file was joined from two-part data and prepare to take out.

Finally, I am used Conver to CSV tool in this platform. I need CSV file to put it into PBI.









**2.Data Visuazation**

**a. Data prepare**

I transform all data in PowerBi (PBI). I take a regular task to prepare analysing it. Creating Data source folder and Data modeling folder are the first step when I import data to PBI.

I do some tasks to resolve other small which still has after data prepare step form others tool. Such as I replaced null value by others value, remove duplicate row, etc.

This visualiazation dashboard has a dashboard for supervisor have country map. So, I create one more table which store location number. But when add to map visualize, we need create function import to table to store location data such as Longitude and Latitude.

**Graphical user interface, application, table, Excel

Description automatically generated**

**Text

Description automatically generated**

**Table

Description automatically generated**

**b. Data modeling**

The data is cleaned. I design a model to analyst data in PBI. These table are connected by relation one-many. Now we let start to visualize and analysis data.

Graphical user interface, application

Description automatically generated

**3. Data Analyst**

This section I present by using Vietnamese. Actually, I am Vietnamese people so that I proud of speaking Vietnamese and easy for me to depth analysis. Kindly understand for me.

Trong phần này, mình sẽ trình bày bằng ngôn ngữ tiếng Việt. Thực sự, mình là người Việt Nam nên mình tự hào trình bày bằng tiếng Việt và cũng dễ dàng hơn cho mình khi phân tích sâu dữ liệu này.

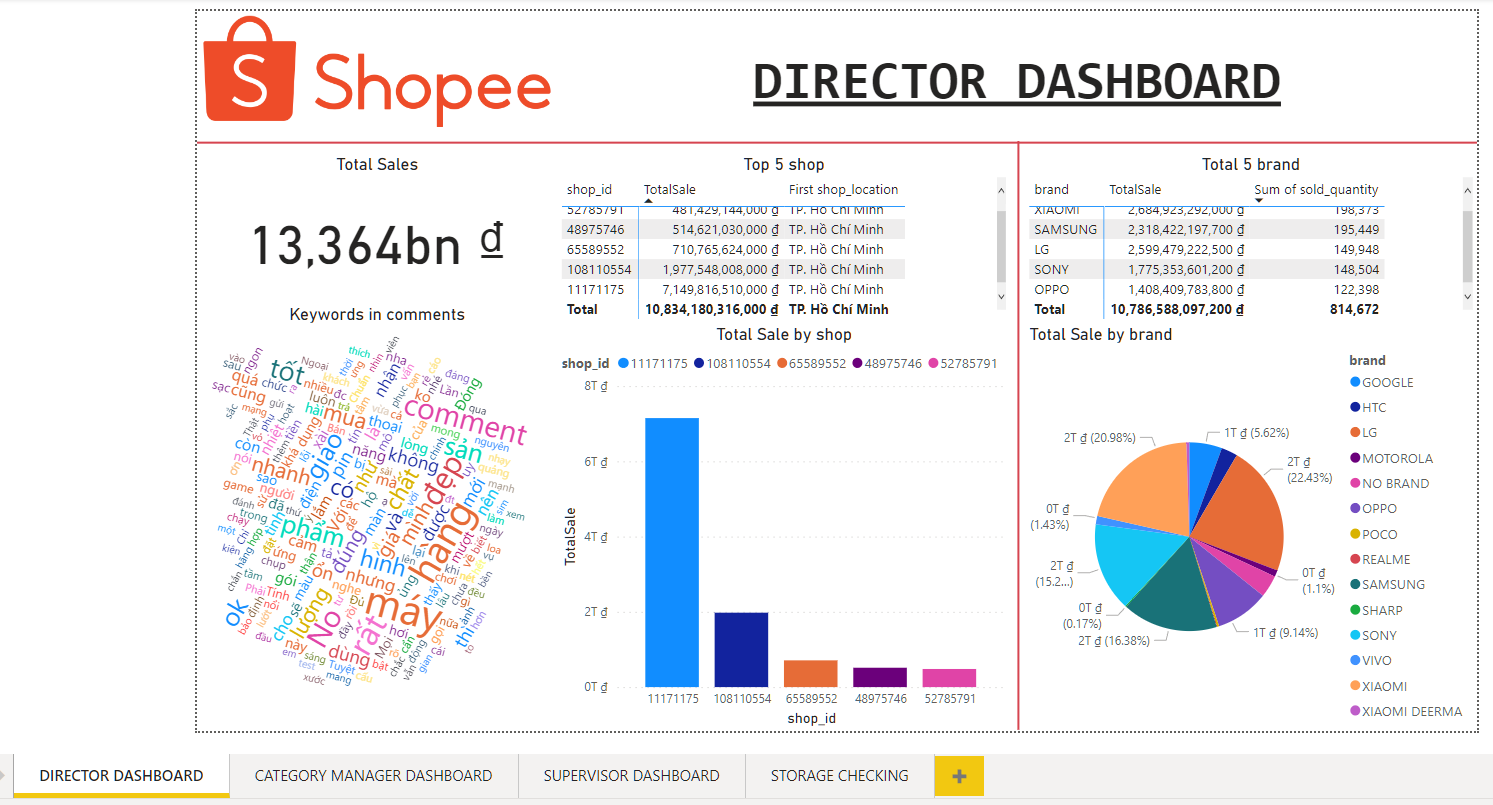
-------------------------------------------

Đầu tiên, chúng ta quan sát bảng dữ liệu DIRECTOR DASHBOARD, bảng này tập trung vào 4 phần: Tổng doanh thu tính đến nay, Top 5 cửa hàng có doanh thu lớn nhất, Top 5 thương hiệu bán chạy nhất và biểu đồ thị phần của các nhãn hiệu hiện nay.

Tổng doanh thu ghi nhận hiện nay của Shopee về mảng điện thoại di động là 13.364 tỷ đồng.

Như chúng ta có thể thấy toàn bộ 5 cửa hàng có doanh thu lớn nhất đều tập trung tại TP.HCM với tổng doanh thu là 10.834.180.316.000d trên tổng số 13 nghìn tỷ của cả nước, tức chiếm 80% doanh thu của tất cả các shop trên cả nước. Một con số rất khủng khiếp nhưng cũng không quá ngạc nhiên vì Sài Gòn luôn là nơi giao thương tấp nập nhất nước ta.

Bên cạnh đó trong top 5 thương hiệu dẫn đầu là những cái tên quen thuộc như Xiaomi, Samsung, LG, Sony và Oppo . Trong đó 3 vị trí dẫn đầu là Xiaomi, Samsung, LG lại không kém cạnh nhau bao nhiêu khi doanh số rơi vào khoảng 2 nghìn tỷ đồng.



Biểu đồ tròn là biểu đồ tổng thị phần của các nhãn hiệu điện thoại (ngoại trừ Apple) đang kinh doanh hiện nay trên sàn thương mại Shopee. Trong đó có một vài tên quen thuộc của nước Mỹ nhưng thị phần kinh doanh hiện nay chỉ chiếm khoảng 1-6% như Google chiếm 5.5% thị phần còn Motorola chỉ chiếm 1.1% thị phần hiện nay tại sàn thương mại Shopee.

Một cái tên nhãn hiệu cũng khá lạ là Xiaomi Deerma cũng xuất hiện trong biểu đồ này với tỷ lệ thị phần tương đương với nhãn hiệu Motorola là 1.1%

Chart, pie chart

Description automatically generated

Tiếp đến chúng ta cùng đi sang CATEGORY MANAGER DASHBOARD để cùng phân tích một vài số liệu quan trọng liên quan đến các mặt hàng kinh doanh hiện nay.

Map

Description automatically generated with medium confidence

Ngoại trừ 2 thành phần mà chúng ta đã xem xét bên DIRECTOR DASHBOARD thì chúng ta có ở đây 3 biểu đồ dữ liệu quan trọng khác là Top 5 mẫu điện thoại được bán nhiều nhất, bản đồ phân bổ số lượng cửa hàng kinh doanh tại Việt Nam và bảng kiểm tra hàng tồn.

Trong 5 mẫu điện thoại có doanh số lớn nhất thì SAMSUNG có một mẫu sản phẩm, Oppo có một mẫu và LG đóng góp 3 mẫu. Đó là các mẫu điện thoại:

* Samsung Galaxy M2 ( doanh thu ~ 472 tỷ đồng )
* Oppo A16 (doanh thu ~ 172 tỷ đồng )
* LG V40 ( doanh thu ~ 123 tỷ đồng )
* LG V50 S ( doanh thu ~ 120 tỷ đồng )
* LG V50 ( doanh thu ~ 151 tỷ đồng )

Chúng ta chuyển sang biểu đồ phân bổ các cửa hàng kinh doanh tại Việt Nam.

Map

Description automatically generated

Như bản đồ trên chúng ta cũng có thể thấy rằng 2 khu vực tập trung nhiều cửa hàng điện thoại mở trên sàn thương mại Shopee nhất là khu vực miền Nam và miền Bắc. Trong đó số lượng cửa hàng tại Hà Nội chiếm cao nhất với tổng cộng 291 cửa hàng.

Tiếp đến là số lượng cửa hàng tại TP.HCM chiếm vị trí thứ 2 với 222 cửa hàng và vị trí thứ 3 thuộc về Nam Định với 28 cửa hàng. Thông tin thêm rằng Nam Định từ lâu đã được mệnh danh là thủ phủ ra đời của rất nhiều loại điện thoại chính ngạch và “like new”.

Chúng ta không có quá nhiều vấn đề để nói trong bảng SUPERVISOR DASHBOARD tiếp theo vì phần nhiều các số liệu ở đây đều là chức năng được xây dựng cho đặc thù công việc của các bạn này vị trí Suppervisor. Chẳng hạn như biểu đồ xem doanh thu các shop , biểu đồ xem chấm điểm các shop, bảng biểu xem tổng doanh thu khuyến mãi của từng cửa hàng, biểu đồ xem tổng doanh thu khuyến mãi của từng sản phẩm.

A picture containing table

Description automatically generated

Bên cạnh đó chúng ta cũng có ba con số đáng lưu tâm. Đó là hiện nay trên sàn thương mại Shopee đã có 2.617 mặt hàng điện thoại di động, với sự đóng góp của 699 cửa hàng trên toàn quốc đã nhận được 114 nghìn lượt đánh giá.

Cuối cùng là bảng STORAGE CHECKING BOARD. Nhìn chung đây là một bảng chức năng dành cho nhân viên của sàn thương mại Shopee dùng để kiểm tra tồn kho , giá cả của sản phẩm.

Table

Description automatically generated

Chúng ta đã vừa đi qua tất cả các báo cáo dữ liệu này. Hy vọng tất cả các dữ liệu và trình bày ở trên sẽ hữu ích với tất cả quý anh chị.