Phoon Huat Project

16 May 2022

Problem Statement

To increase profits from members by increasing revenue or reducing costs

Data Understanding

5 CSV files provided

(1) Orders - Historical

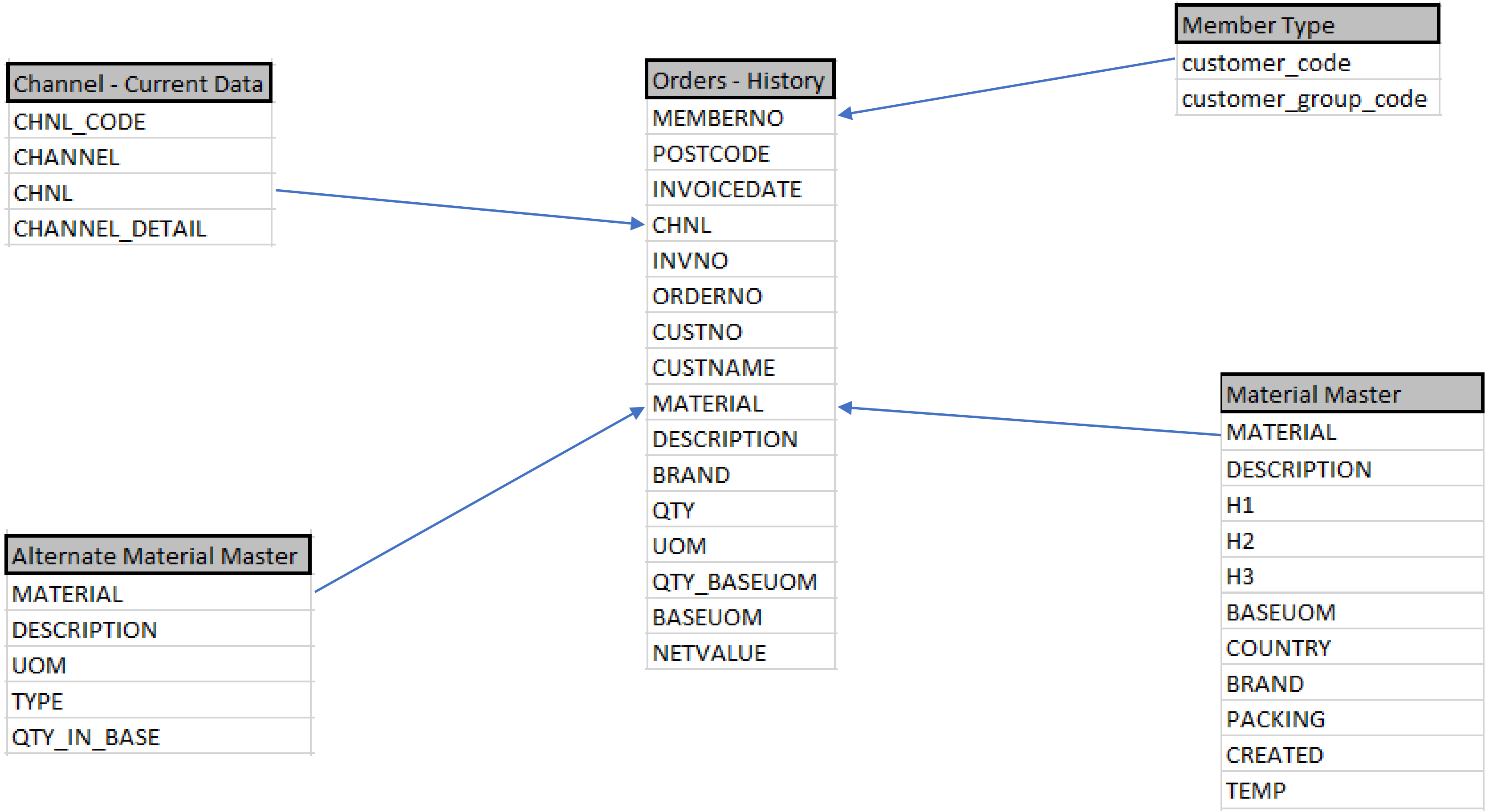
(2) Member type - Current data

(3) Material master - Current data

(4) Channel - Current data

(5) Alternate material master - Current data

Data Understanding – Star Schema



Channel - Current Data
CHNL_CODE
CHANNEL
CHNL
CHANNEL_DETAIL

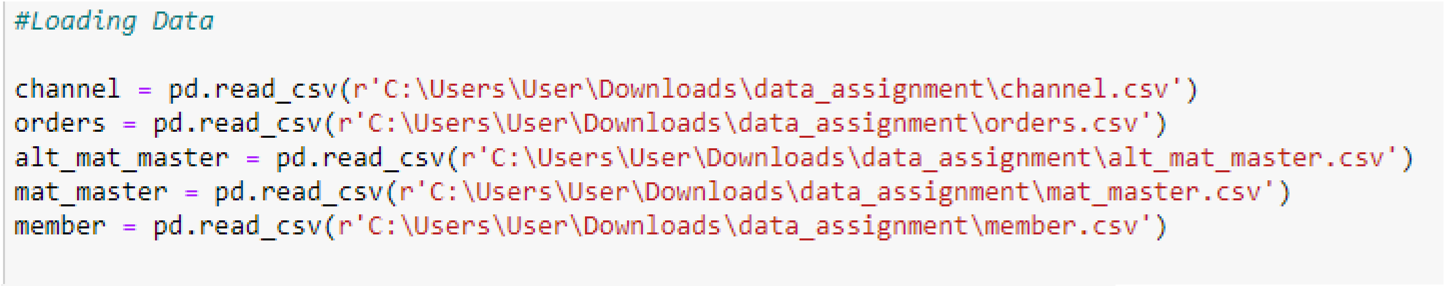
Orders - History
MEMBERNO
POSTCODE
INVOICEDATE
CHNL
INVNO
ORDERNO
CUSTNO
CUSTNAME
MATERIAL
DESCRIPTION
BRAND
QTY
UOM
QTY_BASEUOM
BASEUOM
NETVALUE

Member Type
customer_code
customer_group_code

Alternate Material Master
MATERIAL
DESCRIPTION
UOM
TYPE
QTY_IN_BASE

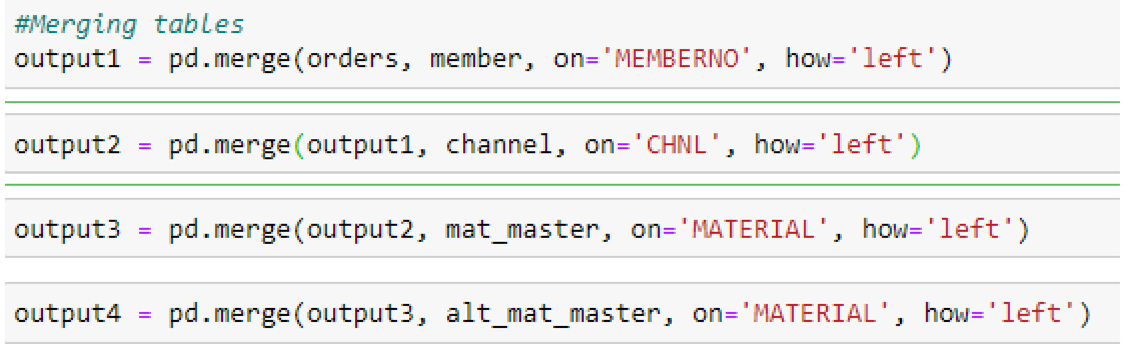
Material Master
MATERIAL
DESCRIPTION
H1
H2
H3
BASEUOM
COUNTRY
BRAND
PACKING
CREATED
TEMP

Data Preparation



```
#Loading Data

channel = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\channel.csv')
orders = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\orders.csv')
alt_mat_master = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\alt_mat_master.csv')
mat_master = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\mat_master.csv')
member = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\member.csv')
```

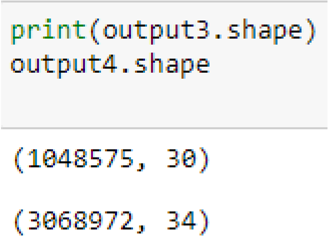


```
#Merging tables
output1 = pd.merge(orders, member, on='MEMBERNO', how='left')
```

```
output2 = pd.merge(output1, channel, on='CHNL', how='left')
```

```
output3 = pd.merge(output2, mat_master, on='MATERIAL', how='left')
```

```
output4 = pd.merge(output3, alt_mat_master, on='MATERIAL', how='left')
```



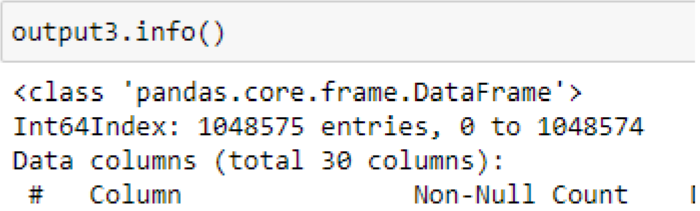
```
print(output3.shape)
output4.shape
```

```
(1048575, 30)

(3068972, 34)
```

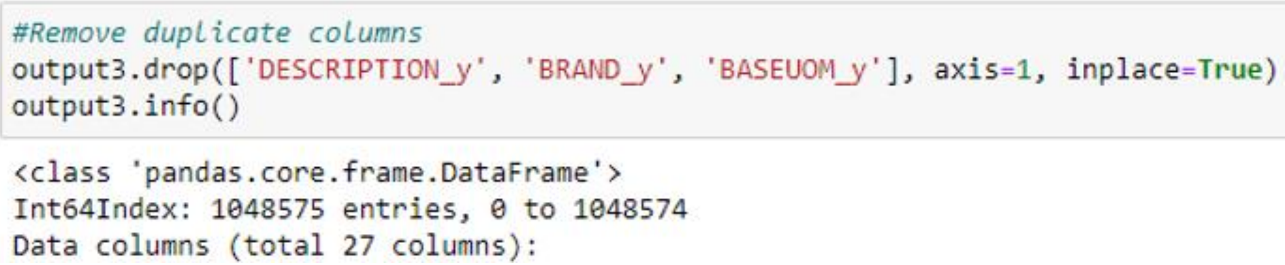
Will use output3 as the data as it has been preserved vs additional rows created in output4. Additional columns in alt_mat_master also do not affect the insights generated adversely.

Data Preparation



```
output3.info()
```

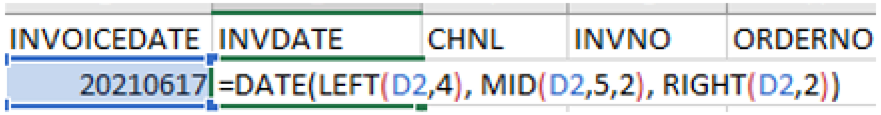
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1048575 entries, 0 to 1048574
Data columns (total 30 columns):
 #   Column                 Non-Null Count    I
```



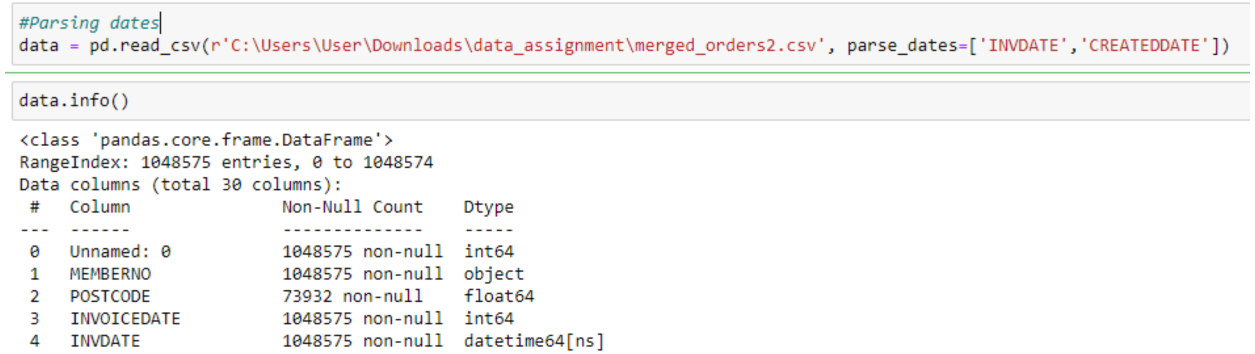
```
#Remove duplicate columns
output3.drop(['DESCRIPTION_y', 'BRAND_y', 'BASEUOM_y'], axis=1, inplace=True)
output3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1048575 entries, 0 to 1048574
Data columns (total 27 columns):
```

#Format number to date



INVOICEDATE	INVDATE	CHNL	INVNO	ORDERNO
20210617	=DATE(LEFT(D2,4), MID(D2,5,2), RIGHT(D2,2))			



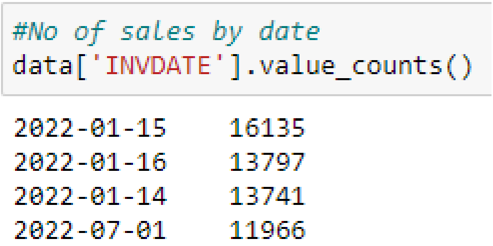
```
#Parsing dates
data = pd.read_csv(r'C:\Users\User\Downloads\data_assignment\merged_orders2.csv', parse_dates=['INVDATE','CREATEDDATE'])
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 30 columns):
 #   Column                Non-Null Count    Dtype
---  ------                --------------    -----
 0   Unnamed: 0            1048575 non-null  int64
 1   MEMBERNO              1048575 non-null  object
 2   POSTCODE              73932 non-null    float64
 3   INVOICEDATE           1048575 non-null  int64
 4   INVDATE               1048575 non-null  datetime64[ns]
```

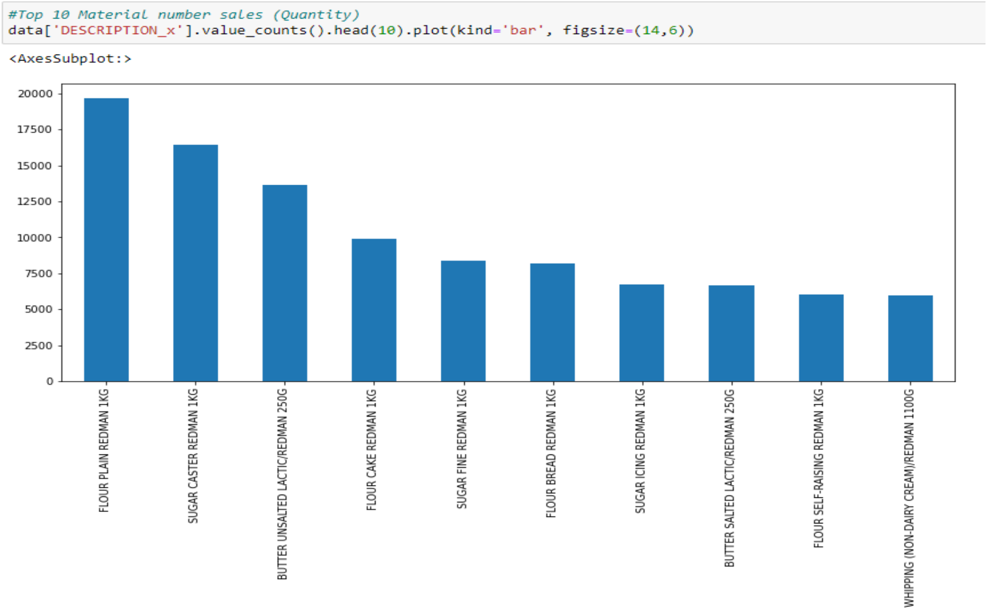
Data Insights - Python

Overview of the data



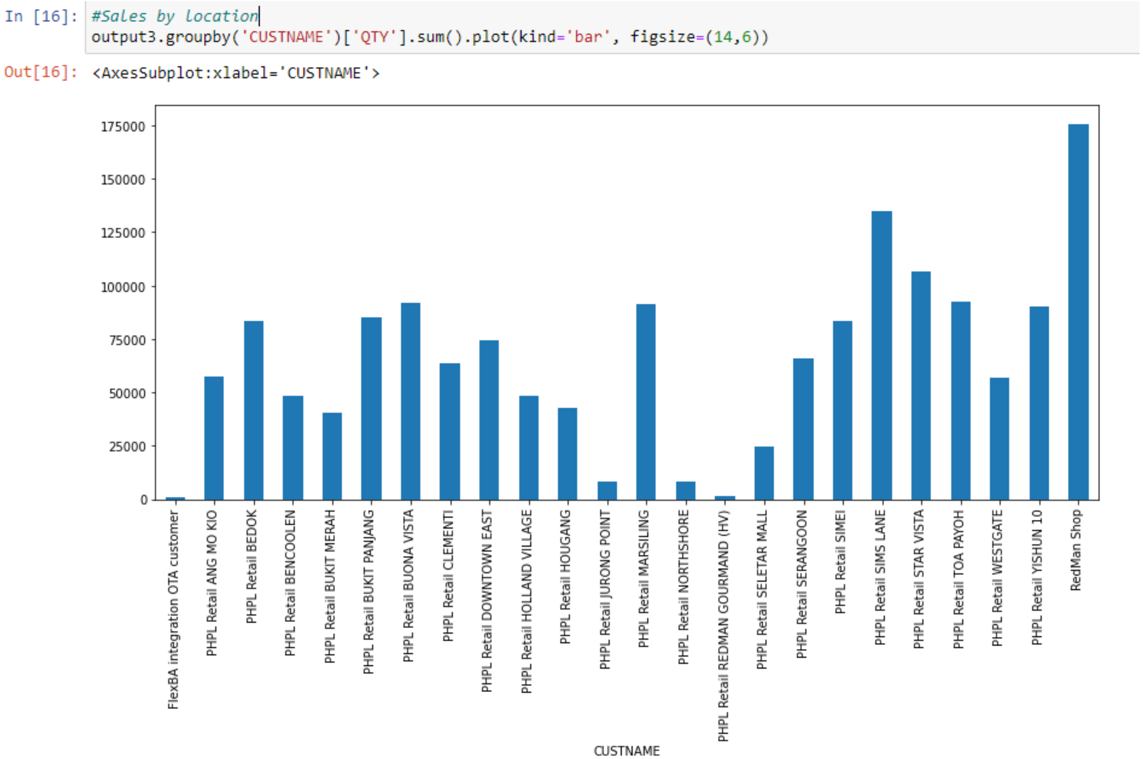
```
#No of sales by date
data['INVDATE'].value_counts()
```

```
2022-01-15    16135
2022-01-16    13797
2022-01-14    13741
2022-07-01    11966
```



```
#Top 10 Material number sales (Quantity)
data['DESCRIPTION_x'].value_counts().head(10).plot(kind='bar', figsize=(14,6))
```

```
<AxesSubplot:>
```



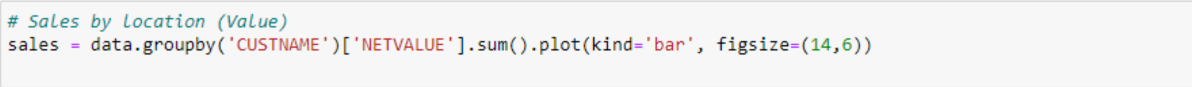
```
In [16]: #Sales by location
         output3.groupby('CUSTNAME')['QTY'].sum().plot(kind='bar', figsize=(14,6))

Out[16]: <AxesSubplot:xlabel='CUSTNAME'>
```

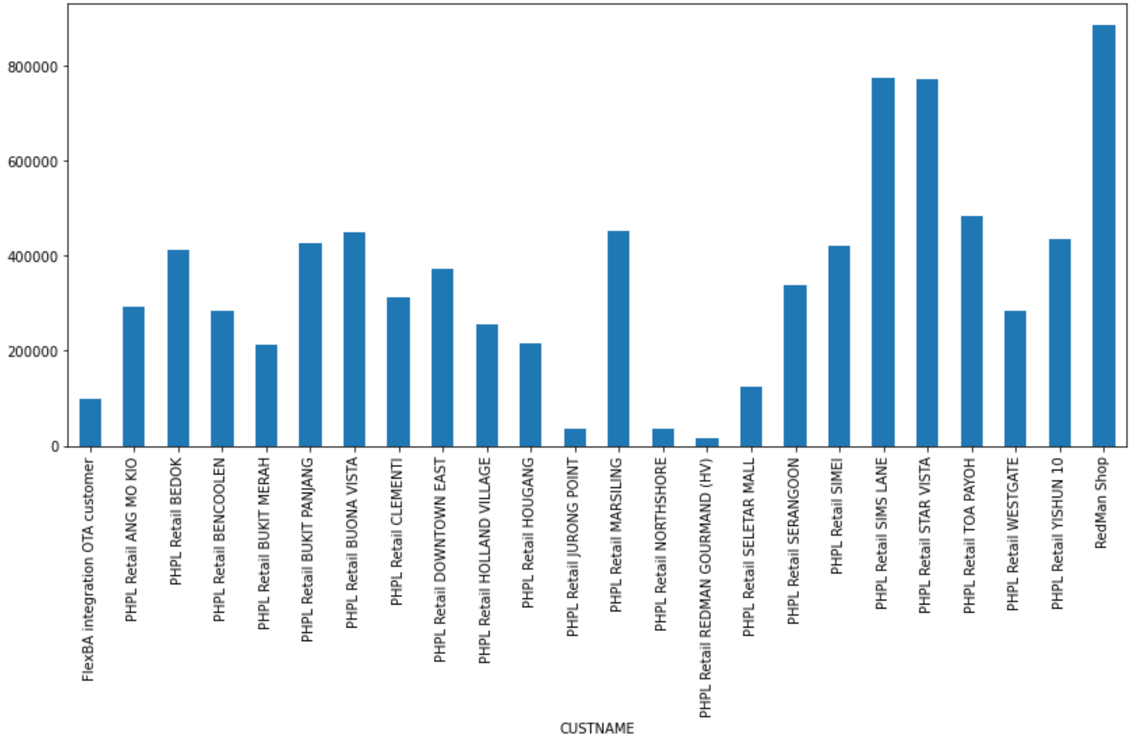
Just a general overview of the data. We can see that RedMan Shop pulls in highest number of sales by quantity. Note that this is the ecommerce channel while the retail is distributed among the outlets. Plain Flour 1kg is the most sold product. (Data here is incomplete but fixed in the Power BI slides)

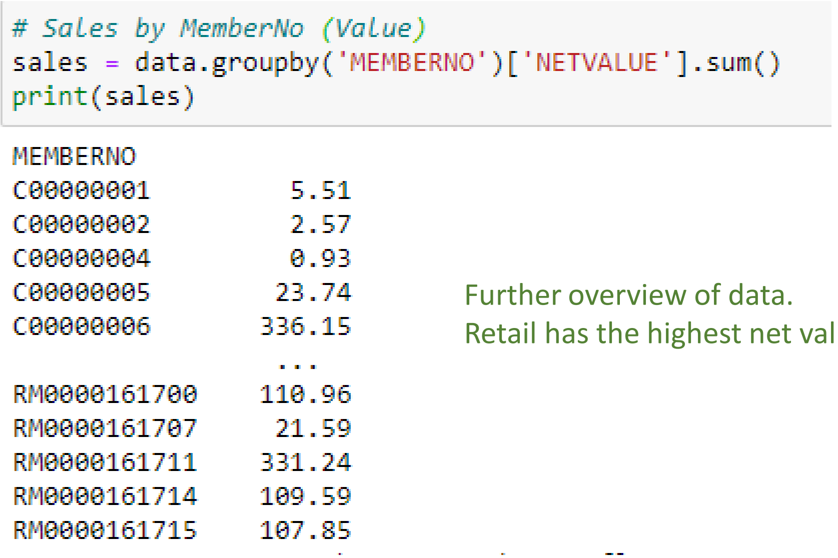
Data Insights - Python

Overview of the data



```
# Sales by location (Value)
sales = data.groupby('CUSTNAME')['NETVALUE'].sum().plot(kind='bar', figsize=(14,6))
```

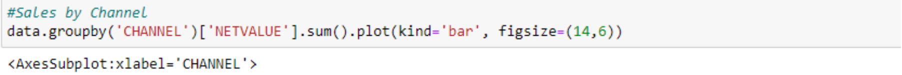




```
# Sales by MemberNo (Value)
sales = data.groupby('MEMBERNO')['NETVALUE'].sum()
print(sales)
```

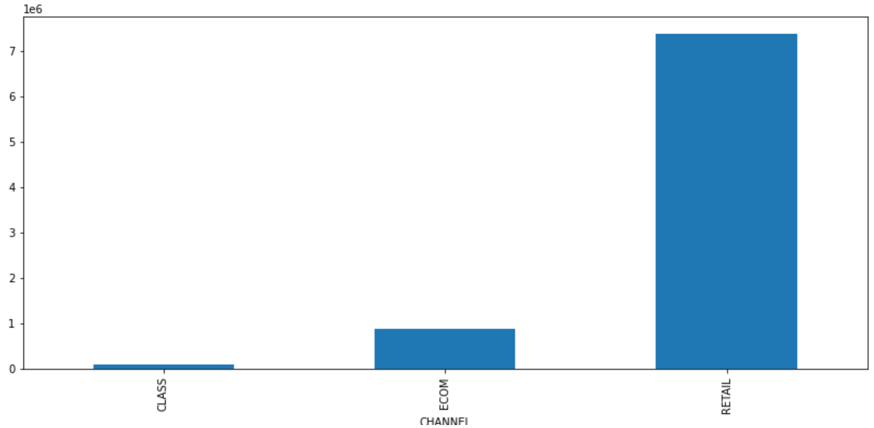
```
MEMBERNO
C00000001           5.51
C00000002           2.57
C00000004           0.93
C00000005          23.74
C00000006         336.15
                    ...
RM000161700       110.96
RM000161707        21.59
RM000161711       331.24
RM000161714       109.59
RM000161715       107.85
```

Further overview of data.
Retail has the highest net value

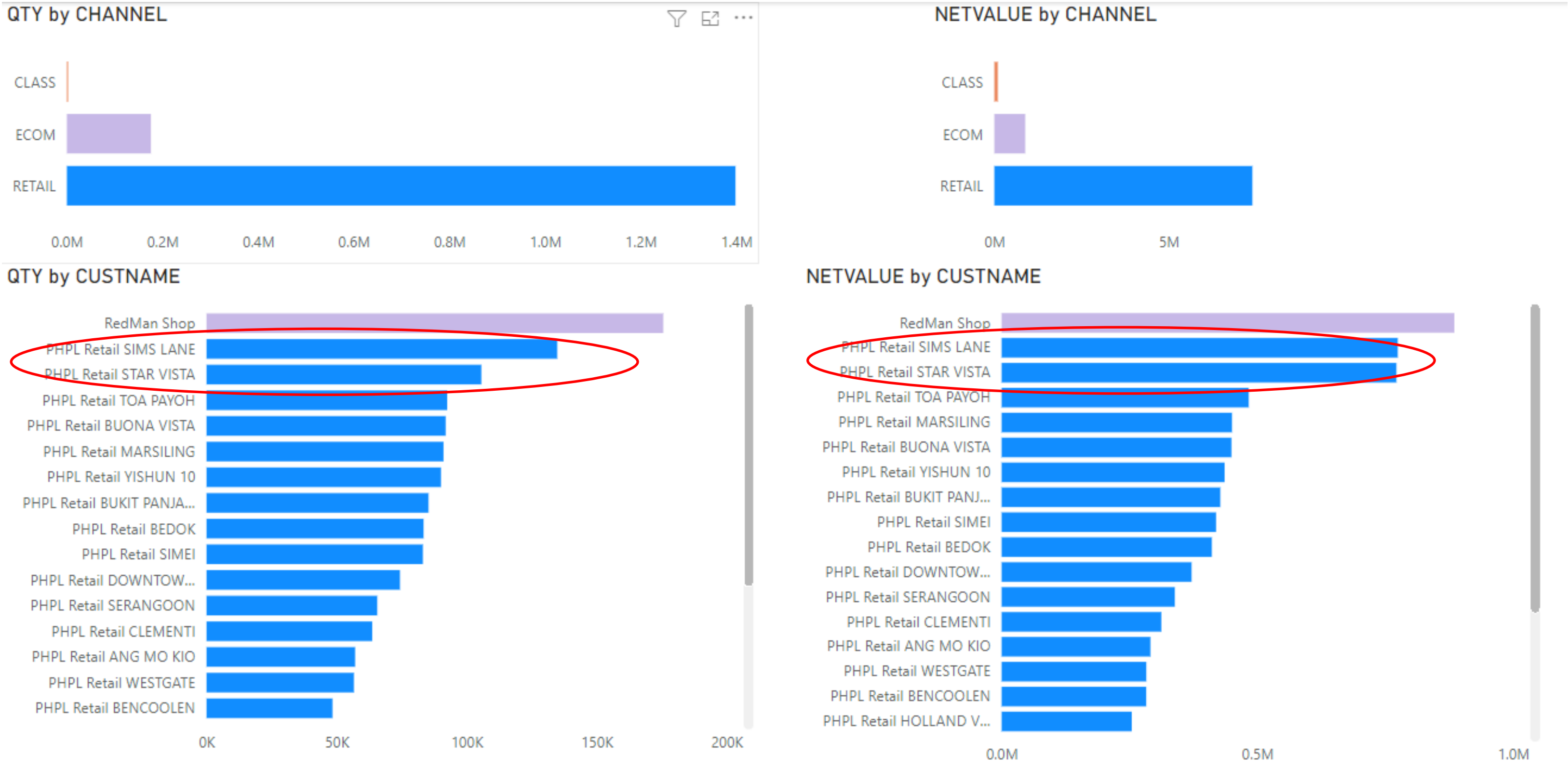


```
#Sales by Channel
data.groupby('CHANNEL')['NETVALUE'].sum().plot(kind='bar', figsize=(14,6))
```

```
<AxesSubplot:xlabel='CHANNEL'>
```



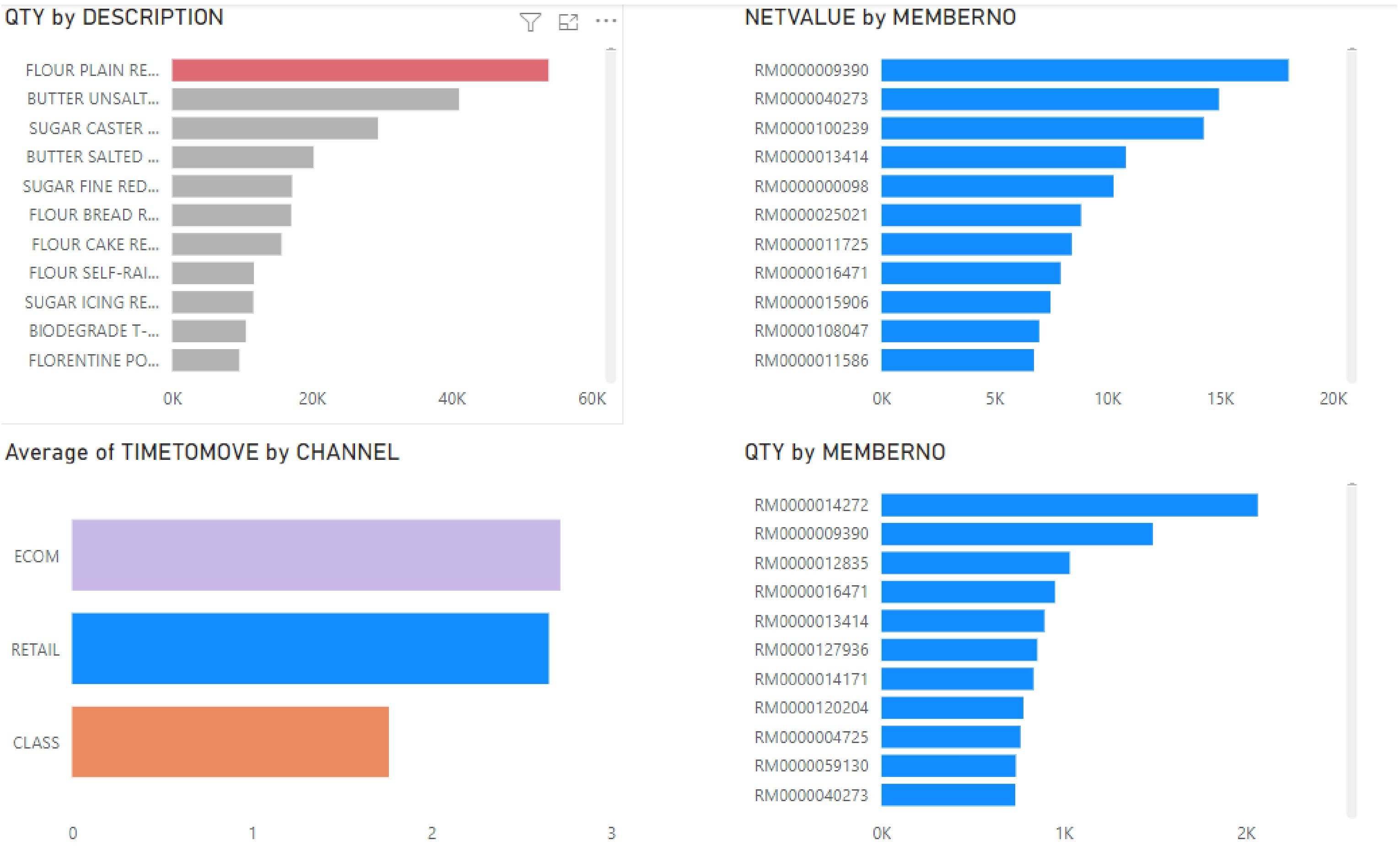
Data Insights – Power BI



Note that net value for Sims Lane and Star Vista is similar, but quantity is lower for the latter. This means net value per quantity is better for Star Vista. To check if we can improve Sims Lane net value.

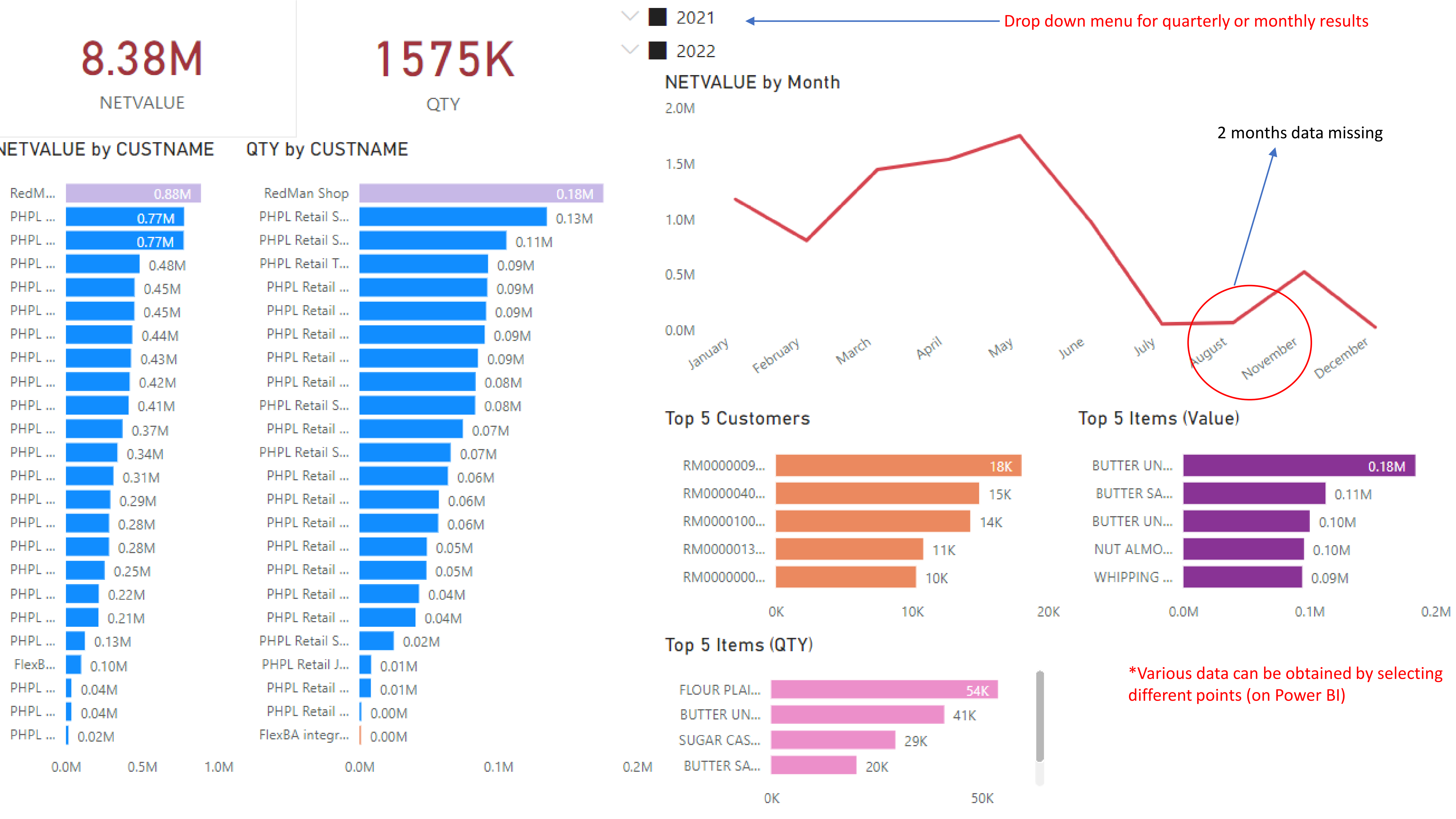
General overview using Power BI. We can see that this is consistent in the previous graphs where retail has the highest in terms of quantity and net value.

Data Insights – Power BI



The right tables show the top members sales. It may be of use to understand their demographic and the reasons why they tend to spend more than the other members. Targeted marketing to those demographics may help to boost sales.

There is not much time difference between created date and invoice date for ecommerce or retail. There may be an opportunity to manage the inventory turn-around time better instead of holding on to too much stock.

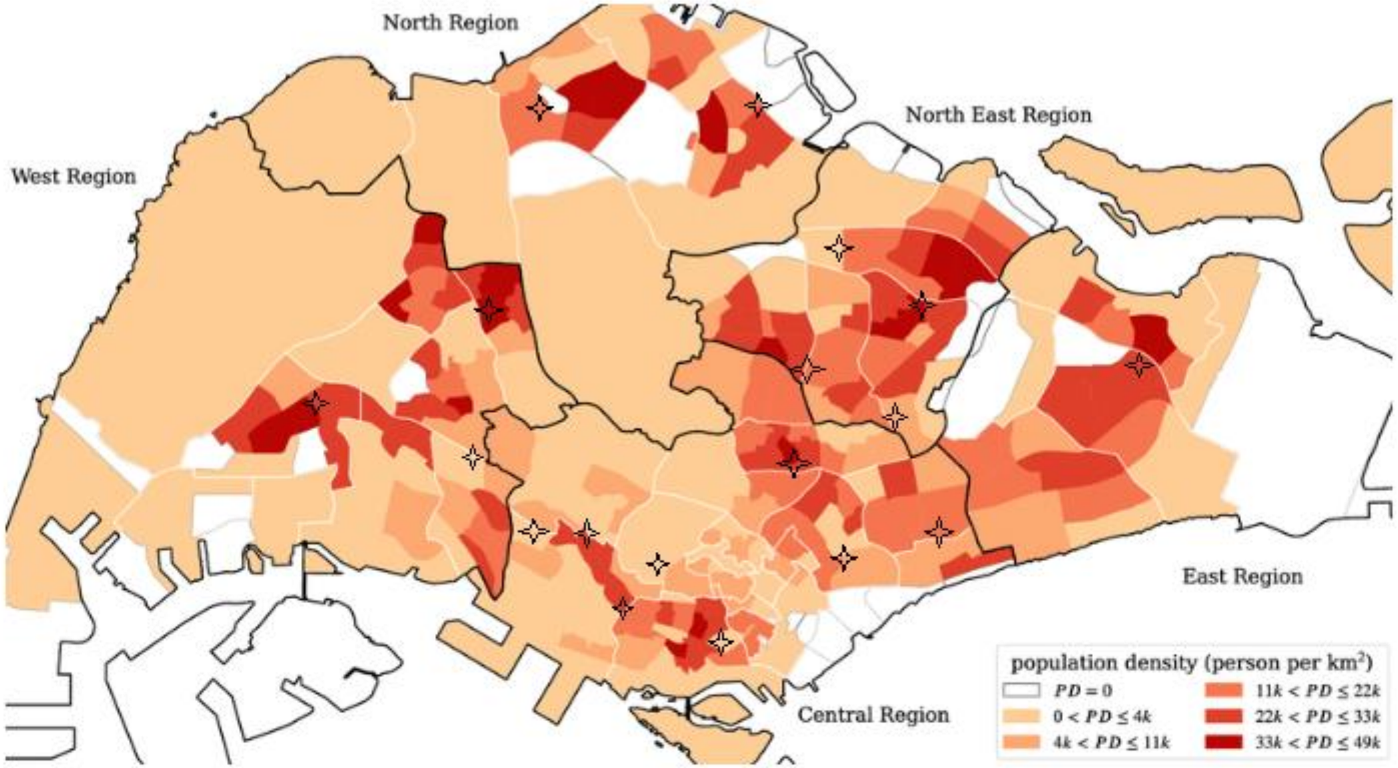


Drop down menu for quarterly or monthly results

2 months data missing

*Various data can be obtained by selecting different points (on Power BI)

Population Density vs Store Sales

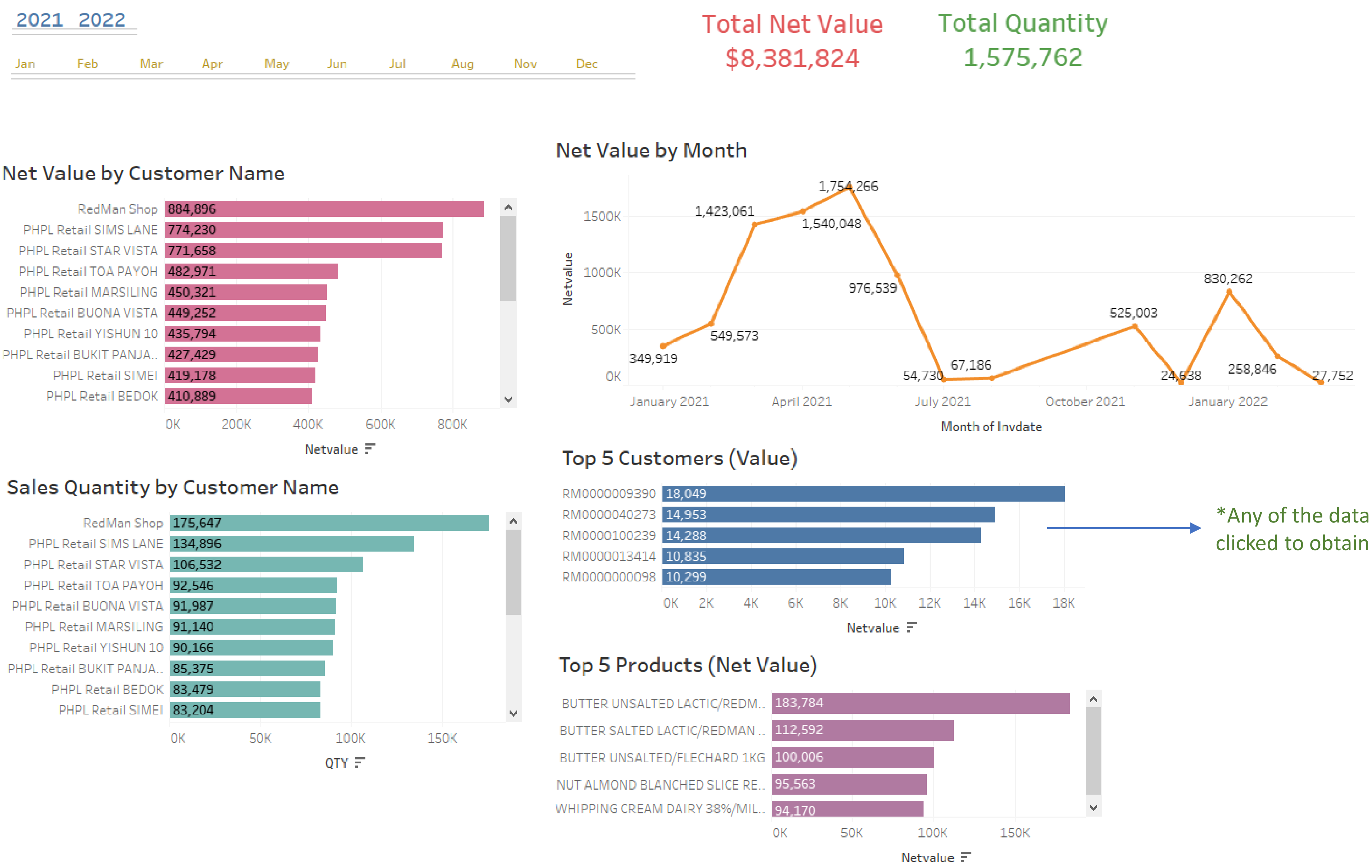


Top 3 Density Location
- Bukit Panjang
- Jurong Point
- Hougang

Top 3 Store Sales
- Sims Lane
- Star Vista
- Toa Payoh

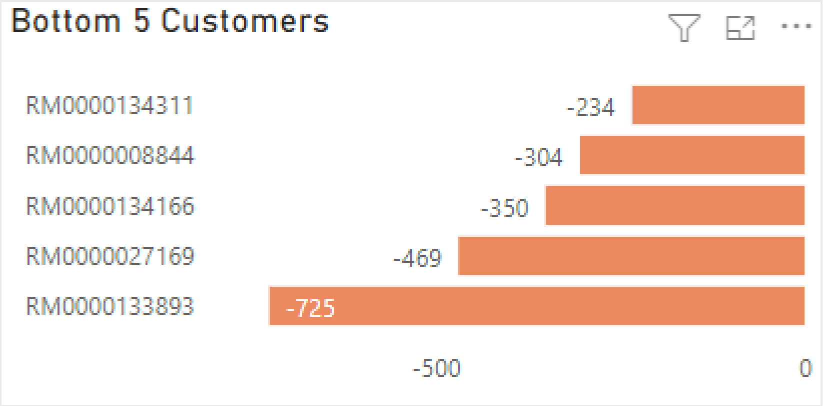
No strong correlation between density of a store and the number of sales it brings in. Hence, when deciding to open a new store, population density does not need to have high consideration.

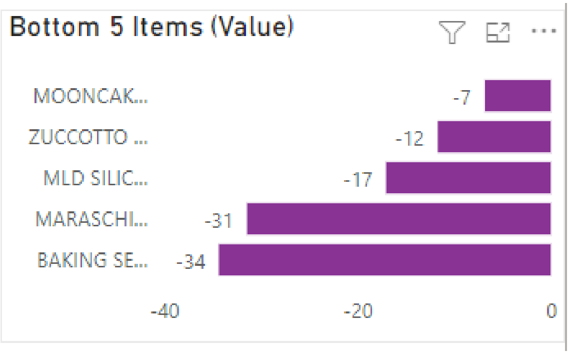
Data Insights - Tableau

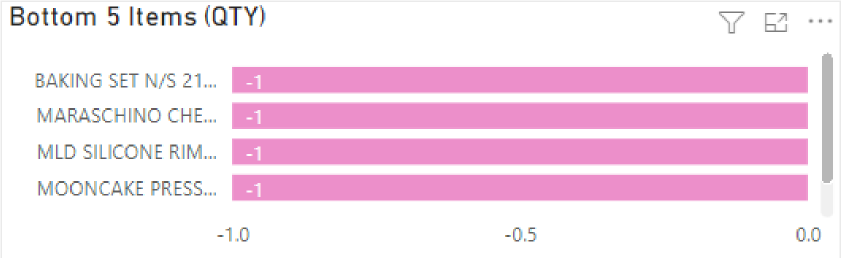

2021 2022
Jan Feb Mar Apr May Jun Jul Aug Nov Dec
Total Net Value
$8,381,824
Total Quantity
1,575,762
Net Value by Customer Name
RedMan Shop 884,896
PHPL Retail SIMS LANE 774,230
PHPL Retail STAR VISTA 771,658
PHPL Retail TOA PAYOH 482,971
PHPL Retail MARSILING 450,321
PHPL Retail BUONA VISTA 449,252
PHPL Retail YISHUN 10 435,794
PHPL Retail BUKIT PANJA.. 427,429
PHPL Retail SIMEI 419,178
PHPL Retail BEDOK 410,889
0K 200K 400K 600K 800K
Netvalue
Net Value by Month
Netvalue
1500K 1000K 500K 0K
349,919
549,573
1,423,061
1,540,048
1,754,266
976,539
54,730
67,186
525,003
24,638
830,262
258,846
27,752
January 2021 April 2021 July 2021 October 2021 January 2022
Month of Invdate
Sales Quantity by Customer Name
RedMan Shop 175,647
PHPL Retail SIMS LANE 134,896
PHPL Retail STAR VISTA 106,532
PHPL Retail TOA PAYOH 92,546
PHPL Retail BUONA VISTA 91,987
PHPL Retail MARSILING 91,140
PHPL Retail YISHUN 10 90,166
PHPL Retail BUKIT PANJA.. 85,375
PHPL Retail BEDOK 83,479
PHPL Retail SIMEI 83,204
0K 50K 100K 150K
QTY
Top 5 Customers (Value)
RM0000009390 18,049
RM0000040273 14,953
RM0000100239 14,288
RM0000013414 10,835
RM0000000098 10,299
0K 2K 4K 6K 8K 10K 12K 14K 16K 18K
Netvalue
Top 5 Products (Net Value)
BUTTER UNSALTED LACTIC/REDM.. 183,784
BUTTER SALTED LACTIC/REDMAN .. 112,592
BUTTER UNSALTED/FLECHARD 1KG 100,006
NUT ALMOND BLANCHED SLICE RE.. 95,563
WHIPPING CREAM DAIRY 38%/MIL.. 94,170
0K 50K 100K 150K
Netvalue

*Any of the data points can be clicked to obtain further information

Data Insights







Some members only have negative net values,
which could mean returns of products followed
by no more business from them. It may help
sales by turning them to returning customers.
By looking at the lowest performing products, it may
be beneficial to stop future purchases to save costs.

Upon digging into information by the bottom 5 customers, the negative values relate to online courses or the Wilton Course. I infer that these courses were given out free to them and the company reflected the costs in the data.

Increase Revenue or Decrease Costs Summary

- Reduce overhead costs at the stores by possibly negotiating for lower rent. Star Vista sales quantity is lower than Sims Lane but as almost equal net value sales. Other methods would be to reduce manpower required.
- Reduce purchase price of the popular items by buying in bulk. Popular items are proven to sell and costs can be reduced this way.
- A/B price testing for popular products. Does increasing the price result in quantity of sales to remain the same? Does decreasing price subsequently cause a larger quantity to be sold to offset the difference and create more profits?
- Sales trend through the year looks like a seasonal nature. Sales pick up in May. Is this due to the mid-year school break and families tend to want to cook/bake and hence purchase the supplies?
- Study the top few customers and identify demographics.
- Some net values are negative, could be related to returns. Reduce returns by ensuring quality of the products

Thank You

python™

Power BI

tableau®