



SUPERMARKET SALES ANALYSIS

Dataset Exploration Part 2

Appropriate Data Set Description

Data set description has already listed in previous data set exploration file and there is no need to change it.

Univariate Descriptive Statistics

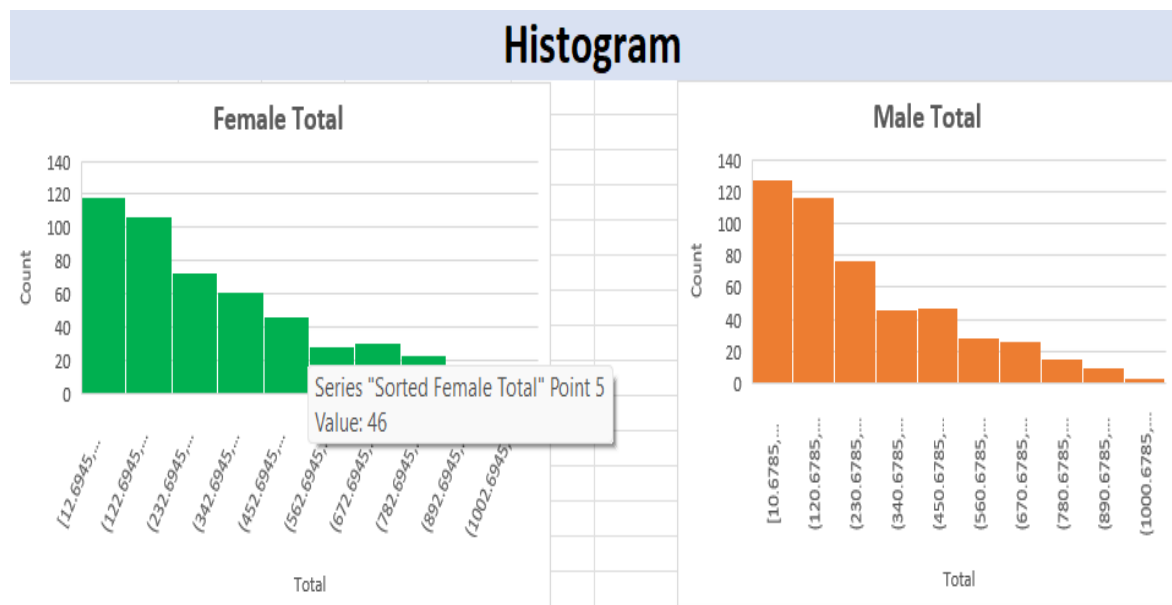
a) Total & Gender

- Central Tendency Metrics and Outliers

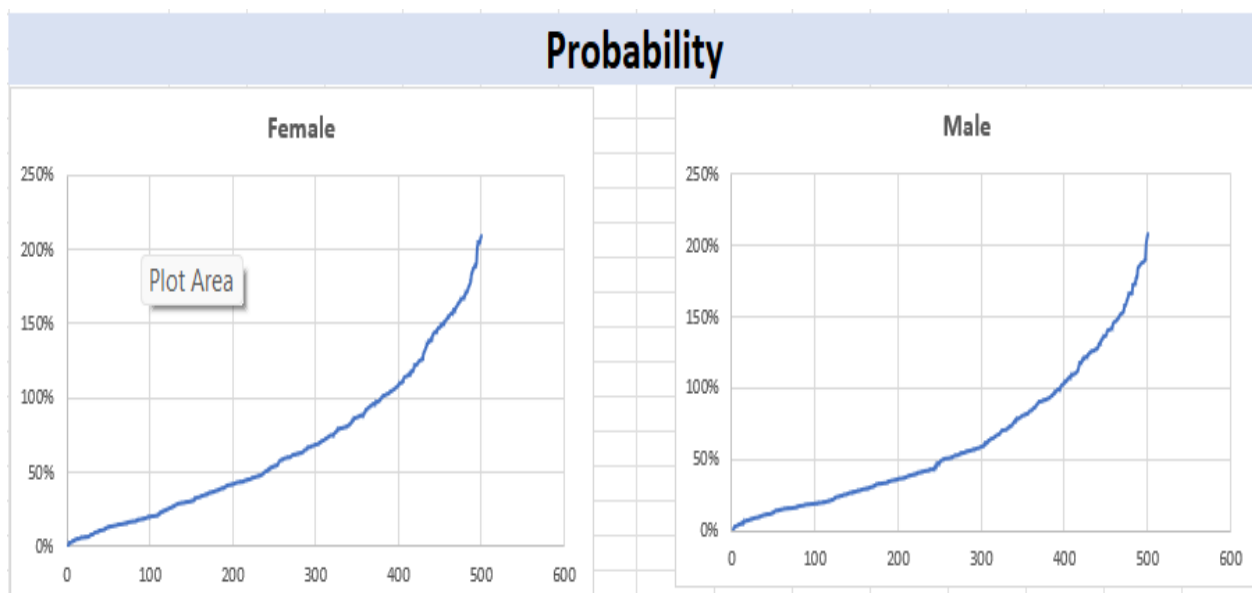
Central Tendency metrics		
Metrics	Female	Male
Mean	\$334.82	\$310.79
Median	✓ \$270.59	✓ \$242.60
Mode	✓ \$217.63	✓ \$175.92
Q3	487.7145	456.834
Q1	132.552	117.9465
IQR	355.1625	338.8875
Upper Outlier Range	1020.458	965.1653
Lower Outlier Range	-400.192	-390.385

- Total Frequency by Gender

Frequency of Total and gender is shown below. The distribution shows the frequency based on the total price of items and, genders are separated to perform statistics.



The first box and whisker chart is depicting total sales by female and male separately and second chart is representing then combinedly.



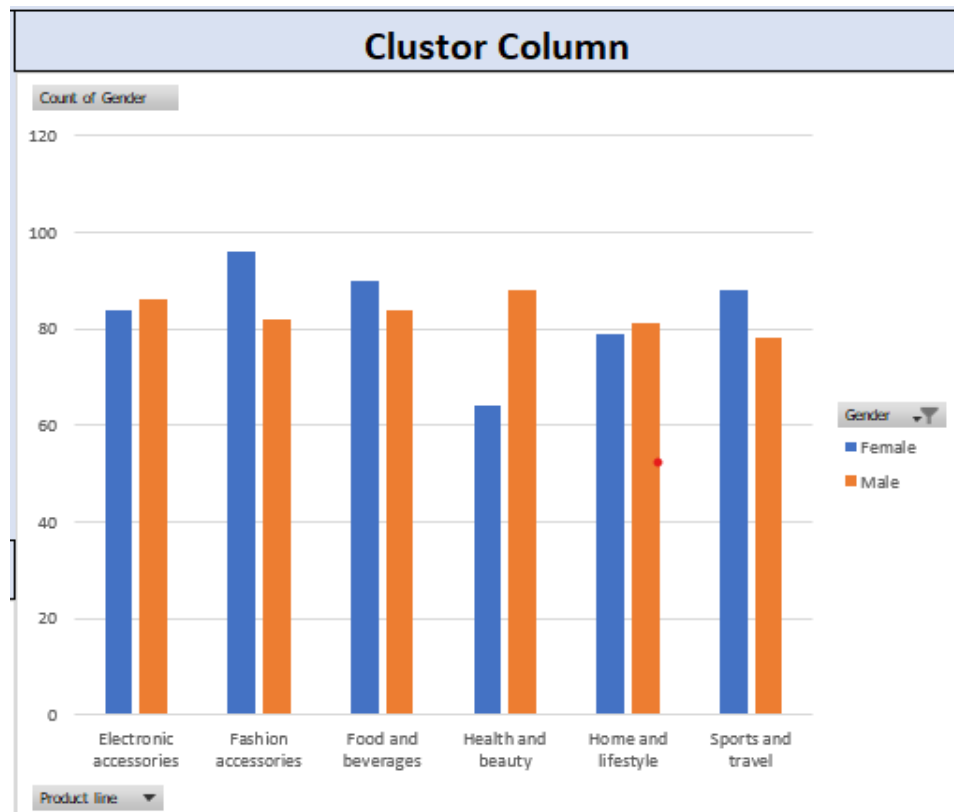
b) Product Line

- **Pivot table**

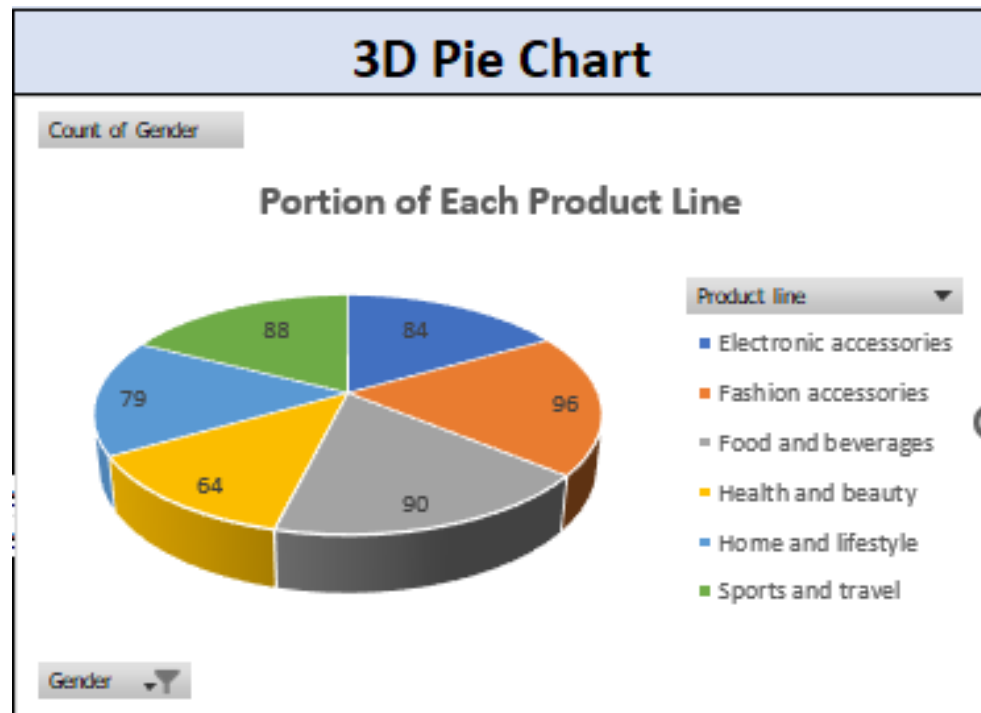
Pivot table			
Count of Gender	Column Labels ▼		
Row Labels ▼	Female	Male	Grand Total
Electronic accessories	84	86	170
Fashion accessories	96	82	178
Food and beverages	90	84	174
Health and beauty	64	88	152
Home and lifestyle	79	81	160
Sports and travel	88	78	166
Grand Total	501	499	1000

- **Product line sales by gender**

Frequency of sales of diff product line is shown based on gender. In which ratio of both male and female is almost equal in purchase of electronic accessories and home & lifestyle products. Women leads in purchase rate of fashion and accessories, while men lead contributing more sales in health and beauty products.



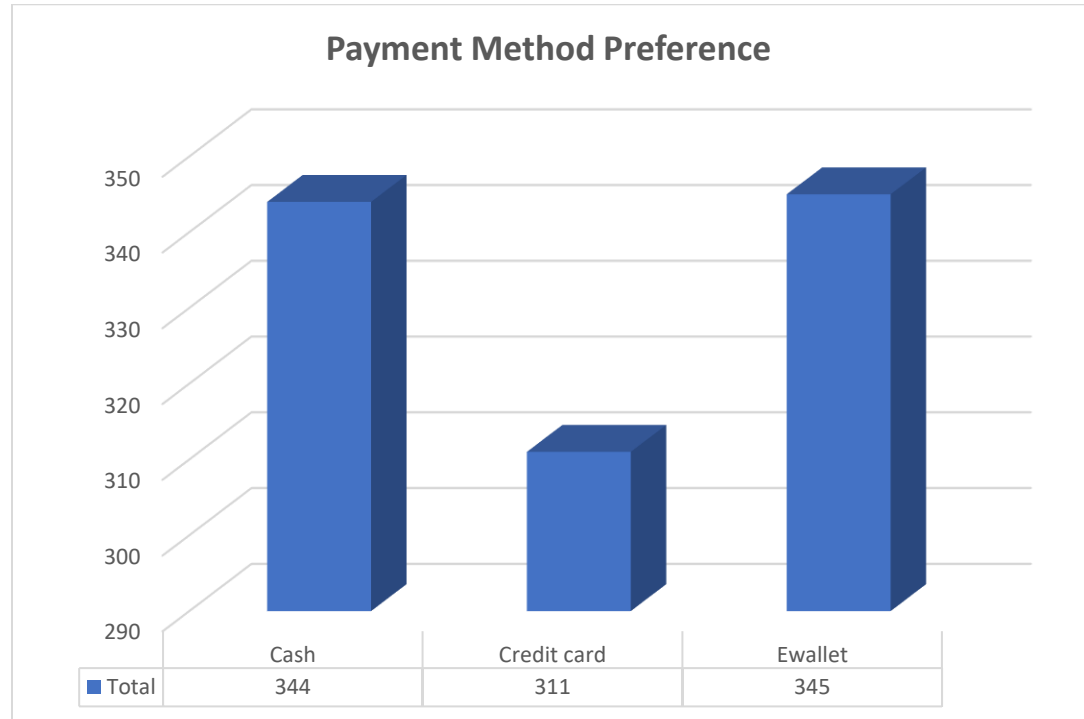
This Pie chart is representing count of each product type. Overall, sales of fashion accessories have been the highest and health and beauty products marked the lowest.



c) Payment Methods and Invoice id

- **Statistics of payment methods based on count of invoice id**

Below is the 3D column chart representing payment preference by customers. There are total three of them, 1. Cash, 2. Credit card and 3. E-wallet. These three are sorted and separated to perform these statistics. Stats depicts that customers prefer credit card the least and cash & e-wallet preference is almost equal.



- **Pivot table**

Pivot table lists the total invoices generated by cash, credit card and e-wallet.

Pivot table	
Row Labels	Count of Invoice ID
Cash	344
Credit card	311
Ewallet	345
Grand Total	1000

d) Tax & City

- Central Tendency Metrics

Central Tendency Metrics			
Metrics	Barrie	Orillia	Toronto
Mean	\$15.23	\$16.05	\$14.87
Median	\$12.04	\$12.92	\$11.29
Mode	\$12.57	\$39.48	\$10.33
Q3	22.17075	23.607375	21.873375
Q1	5.7005	5.779	6.51675
IQR	16.47025	17.828375	15.356625
Upper Outlier Range	46.876125	50.3499375	44.9083125
Lower Outlier Range	-19.004875	-20.9635625	-16.518188

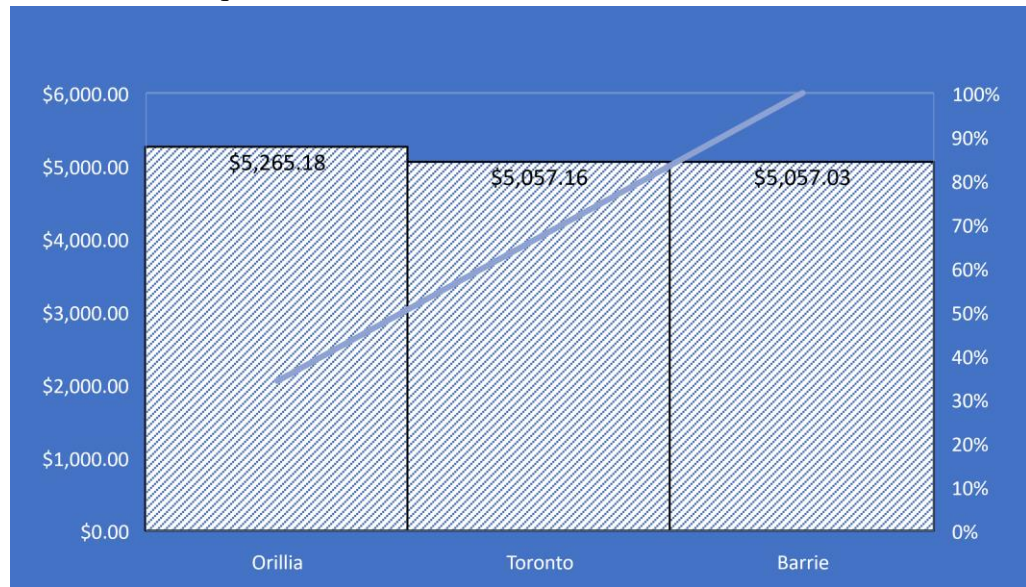
- Pivot table

Pivot table shows the no. of sales per city and its sum of tax

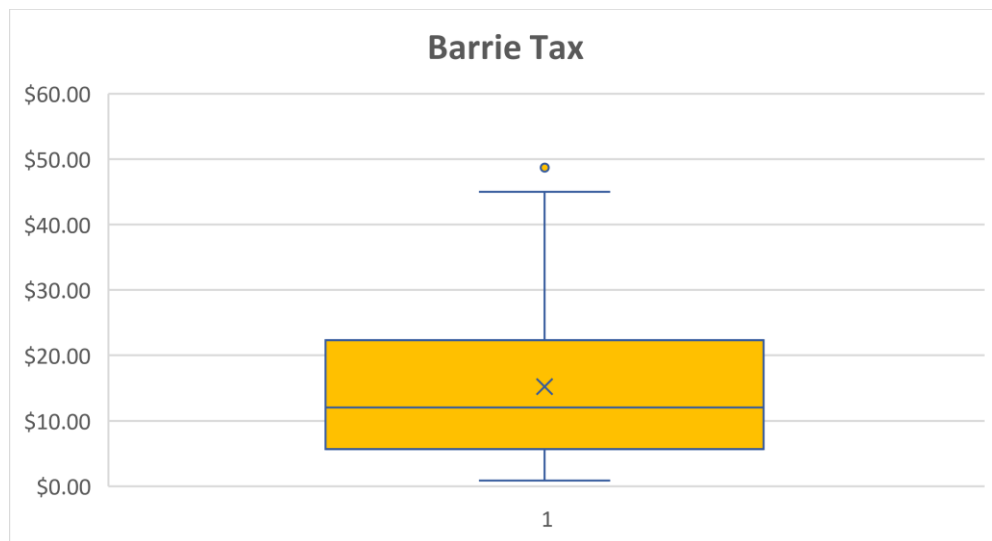
Row Labels	Count of City	Sum of Tax 5%
Barrie	332	5057.032
Orillia	328	5265.1765
Toronto	340	5057.1605
Grand Total	1000	15379.369

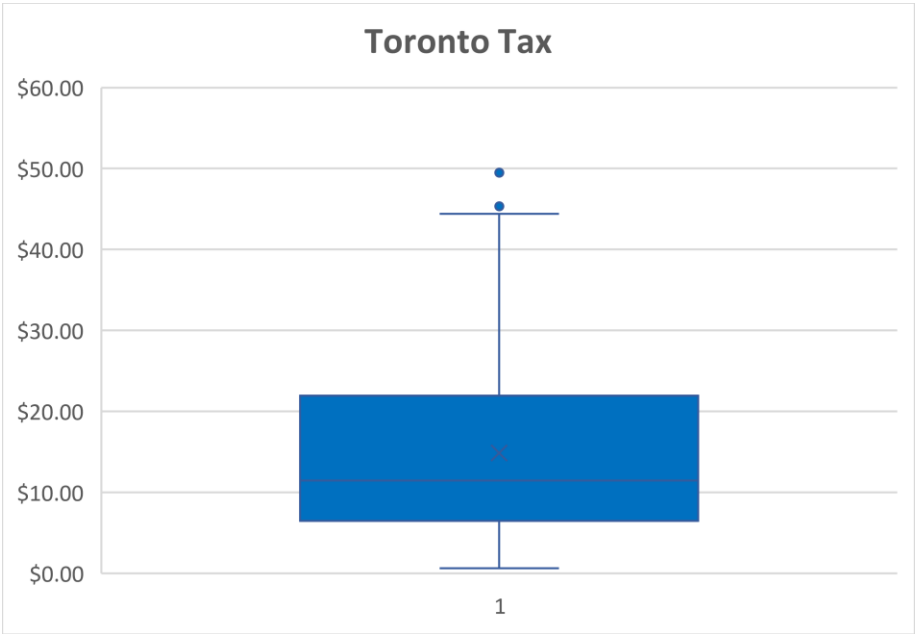
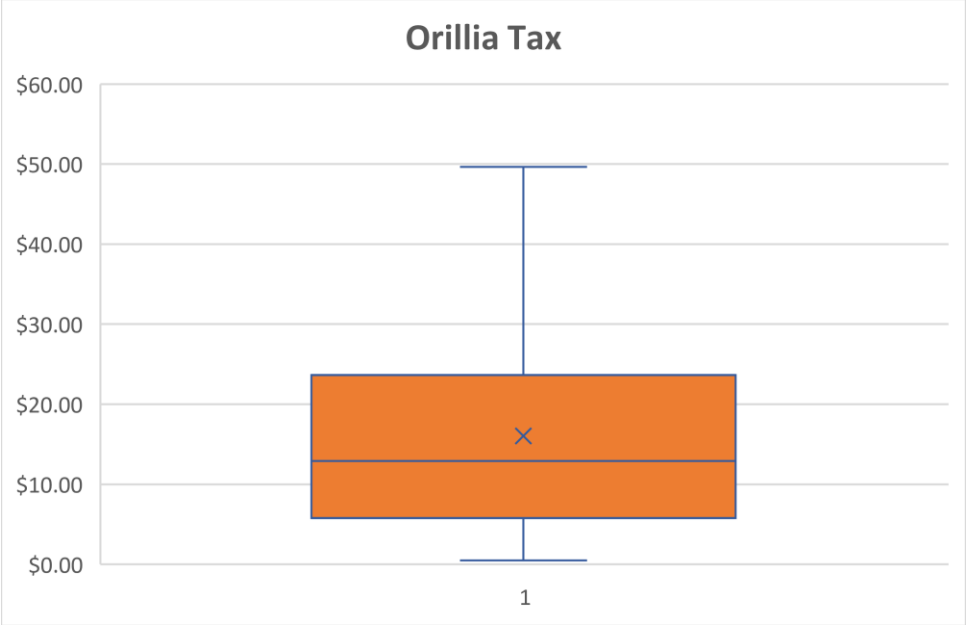
- **Tax based on city**

1. **Pareto chart:** This chart is plotted showing total sales tax collected from different cities, where we can see that Orillia generated maximum tax and Barrie and Toronto are at equal level.



2. **Box and whisker** on each city separately.





e) Gross Income & Branches

- Central Tendency Metrics

Central Tendency metrics			
Metrics	A	B	C
Mean	\$14.87	\$15.23	\$16.05
Median	\$11.47	\$12.04	\$12.92
Mode	\$10.33	\$12.57	\$39.48
Q3	21.8425	22.17075	23.60738
Q1	6.4545	5.7005	5.779
IQR	15.388	16.47025	17.82838
Upper Out	44.9245	46.87613	50.34994
Lower Out	-16.6275	-19.0049	-20.9636

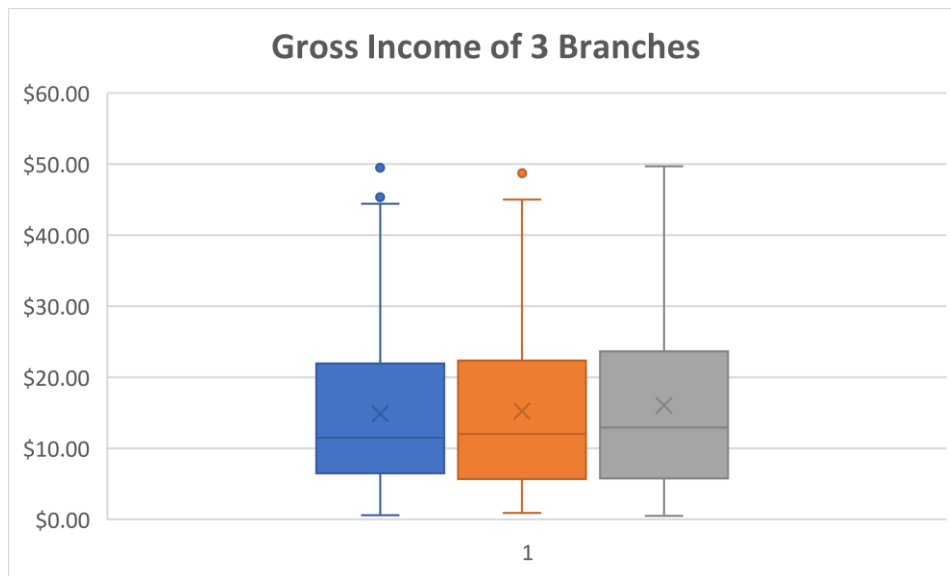
- Pivot table

Pivot table shows gross income per branch

Pivot Table	
Row Labels	Sum of Gross Income
⊕ A	5057.1605
⊕ B	5057.032
⊕ C	5265.1765
Grand Total	15379.369

- **Statistics of gross income based on 3 branches.**

This box and Whisker chart depicts the gross income on each item by 3 different branches.



Finer Research Question

1. Which city leads in sales? On that note, which location's branch should be chosen for expansion and which category of items it should focus on?
 - As we can see, there is data of 3 cities. So, the sales among them may be compared which could answer above question
2. What is the purchasing preference of men and women? Is there any difference in category they prefer more?
 - In this dataset, there are different categories of items which can be preferred by different genders
3. Is there any relationship between COGS and ratings? Using these, shall gross income be predicted?
 - Generally, ratings not only depend on customer service one gets, but also depend on the price of the goods as customer may compare the price from one store with another. And using the current trend future gross income may be calculated
4. Which is the day, the products are sold maximum? And which hour of the day is busiest?
 - Looking at the date and time the products bought, thought of above question
5. Product sales by product line and by city with month slicer so that data statistics can be seen for each month.
6. Same as above for each day and hour.
7. What is the rating distribution across the board?



Demonstrated Tracking

- To track total price by gender, first created 2 different columns. Followed by that, formed a central tendency metrics calculating mean, median, mode, Q3, Q1, IQR, upper outlier and lower outlier range. In which female had higher upper outlier range than male.
- Likewise for all variable suitable charts are being made and tables are generated.

References

Historical record of sales data in 3 different supermarkets. 2019. [Supermarket sales | Kaggle](#)