"Maximizing Profitability and Reseller Efficiency: A Data Analysis Approach for AdventureWorks Cycles"

Final Submission for the BDM Capstone Project

Submitted by

Name: Somesh

Roll number: 21f1001598

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Executive Summary and Title:

The Adventure Works Cycles is a global manufacturer and retailer specializing in bicycles and cycling-related equipment like components, accessories and clothing. The business operates through two channels: resellers and online sales. The project focuses on increasing profitability and optimizing relationships with resellers. Data was downloaded for the official Microsoft account consisting of 7 different tables with 49 features and more than 1 lakh rows capturing real-life challenges and trends making it ideal for analysis purposes. Python language and power BI were majorly used to find trends, insights and possible solutions to the problems and for that initial descriptive statistics were measured and exploratory data analysis was done using Python giving many insights about the data which later became decisive for strategies and processes used in the analysis process.

Six different analyses were performed, Channel analysis to identify which channel has high profitability, category-subcategory analysis to identify profitable product subcategories, geographic analysis to expand our understanding of profitability to the city level taking previous analysis into account. After this reseller performance analysis was performed to filter high and low-performing resellers using profit margin and gross profit matrix, followed by discount and pricing analysis to address inefficient discount patterns found in EDA to make strategies for discounts and offers tailored to the right customer/reseller, tailored to the right time of the year and tailored to the right city. The last analysis was performed to understand customer purchasing behaviour to formulate the right discount and cross-sell technique. Interactive visualization and tables helped find hidden patterns which were then carefully considered to recommend to business.

Proof of Originality

The Microsoft AdventureWorks dataset is a comprehensive, reliable and authentic database that truly illustrates real-world business scenarios. It was collected from the official microsoft github account which can be accessed in two formats given below.

↗ Dataset

↗ Excel Workbook

- BDM project files - Somesh (https://drive.google.com/drive/folders/172jSfom_meu6kVH1Chk4ndqV_DokOg_k?usp=sharing) All the files attached to this report can be accessed through this file folder.
- “AdventureWorks cycles” has been addressed as “business” throughout the report.

Metadata

Customer [Feature, Description]

City	The feature contains the name of the city where the customer either resides or made the purchase
Country-Region	A Categorical feature with values [Australia, United States, Canada, France, Germany, United Kingdom], and nearby countries are being put under one name.
Customer	Name of the customer
Customer ID	Unique identification number (UIN) for each customer [ex. AW0011000]
CustomerKey	The primary key of this table is a unique number for each row to help create data models (interconnection between tables)
Postal Code	An alphanumeric number assigned to a geographical location
State-Province	Name of the state (Higher administration after city)

Date

DateKey	The primary key of this table is a unique number for each row to help create data models (interconnection between tables)
Date	Date ranging from 01-07-2017 to 30-06-2020
Fiscal Year	Categorical feature consisting of values of the financial year 2018, 19 and 20 [FY2018, FY2019, FY2020]
Fiscal Quarter	Each financial year is grouped by 4 months making 4 quarters each year
Month	name of the month
Full Date	date in this format [2020 June, 01]
MonthKey	Unique identification number (UIN) for each month

Product

Category	Categorical feature, 4 categories [Bikes, Accessories, Clothing, Components]
Colour	Colour of the product being sold
List Price	MRP of a product
Model	A number of products have different models due to varying sizes and colour
Product	Name of the product
ProductKey	UIN number for products
SKU	Alphanumeric number for inventory management
Standard Cost	How much a product cost to the company
Subcategory	Each category is divided into subcategories for better classification of products.

Reseller

Business Type	What type of business does my reseller in, 3 categories [Specialty Bike Shop, Warehouse, Value-added Reseller]
City	Reseller's city name
Country-Region	A Categorical feature with values [Australia, United States, Canada, France, Germany, United Kingdom], and nearby countries are being put under one name.
Postal Code	An alphanumeric number assigned to a geographical location
Reseller	Name of the Reseller.
Reseller ID	UIN number for Resellers.
ResellerKey	Primary key of the table
State-Province	Name of the state (Higher administration after city)

SalesOrder **Sales Territory**

Channel	Business use two channels for selling, [Reseller, Online]	Sales TerritoryKey	UIN for grouped geographical location based on sales.
SalesOrderlineKey	UIN for the order line [Ex. 43659001]	Region	A combination of country-continent designed by business for better sales understanding.
Sales Order	UIN for placed order [Ex. SO43659]	Country	Name of the country
Sales Order Line	Order number with line number [Ex. SO43659 - 1]	Group	Think of them as continents

Sales

CustomerKey	UIN number for each online customer, -1 for all resellers
DueDateKey	Due date for delivery
Extended Amount	Order quantity X Unit price
Order Quantity	Total Quantity of a product
OrderDateKey	Date of ordering the products
Product Standard Cost	Cost to the company
ProductKey	UIN for products
ResellerKey	UIN for resellers
Sales Amount	The total amount paid by the customer.
SalesOrderLIneKey	Each order(bill) contains a separate line for what products were bought by the customer.
SalesTerritoryKey	UIN for grouped geographical location based on sales.
ShipDateKey	The date order was shipped.
TotalProductCost	Total cost the company bears for each product

Unit Price	Price for each product
Unit Price Discount %	% Discount business give on each product

Descriptive Statistics

The dataset contains 49 features(variables) spreading through 7 different tables and interconnected with primary and secondary keys with over 1,19,140 rows. We have 3 types of datatypes, Numeric, categorical and DateTime. The tools being used are PowerBi and Excel for graphical analysis and Python for exploratory data analysis. Data didn't require much cleaning except few columns with missing values which were filled after careful consideration. Here is a descriptive summary of the data,

For Categorical features:

Table. 1

index	SKU	Product	Color	Model	Subcategory	Category	Customer ID	Customer	Customer City	Customer State-Province	Customer Country-Region
count	119140	119140	119140	119140	119140	119140	119140	119140	119140	119140	119140
unique	266	266	9	107	35	4	18227	18144	269	53	7
top	WB-H098	Water Bottle - 30 oz.	unknown	Sport-100	Road Bikes	Accessories	[Not Applicable]	[Not Applicable]	[Not Applicable]	[Not Applicable]	[Not Applicable]
freq	4600	4600	33321	9033	20835	40298	59762	59762	59762	59762	59762

Customer Postal Code	Fiscal Year	Fiscal Quarter	Month	Full Date	Reseller ID	Business Type	Reseller	Reseller City
119140	119140	119140	119140	119140	119140	119140	119140	119140
323	3	12	36	1074	633	4	631	415
[Not Applicable]	FY2020	FY2020 Q2	2020 May	2020 Jun, 13	[Not Applicable]	[Not Applicable]	[Not Applicable]	[Not Applicable]
59762	82848	21904	8232	498	59378	59378	59378	59378

For Numerical features:

Table. 2

index	SalesOrderLineKey	ResellerKey	CustomerKey	ProductKey	OrderDateKey	DueDateKey	ShipDateKey	SalesTerritoryKey	Order Quantity	Unit Price
count	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0
mean	57550961.72118516	170.64384757428235	9390.18683061944	424.1220161154944	20191418.717827767	20191670.448615074	20191594.14194225	5.412842034581165	2.269649152257848	469.5606671084439
std	8844523.040871264	224.81330031714393	10172.12724589418	116.92232686450502	7960.338927276504	8005.2399465846665	7990.125079035944	2.8320688852381912	2.499843037800303	756.1232770622224
min	43659001.0	-1.0	-1.0	212.0	20170701.0	20170711.0	20170708.0	1.0	1.0	1.3282
25%	49849001.75	-1.0	-1.0	338.0	20190405.0	20190415.0	20190412.0	4.0	1.0	21.49
50%	56687001.5	3.0	-1.0	469.0	20191005.0	20191015.0	20191012.0	6.0	1.0	52.647
75%	65215016.25	327.0	18121.25	528.0	20200213.0	20200223.0	20200220.0	8.0	3.0	602.346
max	74689001.0	701.0	29483.0	606.0	20200608.0	20200618.0	20200615.0	10.0	44.0	3578.27

Extended Amount	Unit Price Discount Pct	Product Standard Cost	Total Product Cost	Sales Amount	Standard Cost	List Price	DateKey	MonthKey
119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0	119140.0
915.3287109467853	0.0	365.3182311658553	805.7492517231829	910.9300050008394	365.3182311658553	623.8977293511833	20191594.14194225	201915.78443008225
1708.4800150443461	0.0	543.517053159191	1663.0940001375836	1696.473850508397	543.517053159191	916.8768281136663	7990.125079035944	79.90434904300972
1.374	0.0	0.8565	0.8565	1.374	0.8565	2.29	20170708.0	201707.0
24.99	0.0	9.1593	10.8423	24.99	9.1593	24.49	20190412.0	201904.0
136.764	0.0	38.4923	104.7052	136.764	38.4923	63.5	20191012.0	201910.0
1120.49	0.0	486.7066	1030.9488	1120.49	486.7066	782.99	20200220.0	202002.0
30992.91	0.0	2171.2942	38530.3854	27893.619	2171.2942	3578.27	20200615.0	202006.0

The received data was in 7 different tables and each of the tools used require data to be in a certain format. For power BI, all the tables were connected to make the ‘data model’ using table keys, Multiple measures(calculations) were created, and new columns and tables were created to smooth out the analysis process. Measures include AOV(average order value), COGS(cost of goods sold), gross profit, profit margin, Total revenue, Total orders and more.

Similarly, data was transformed to be used with Python programming language. Firstly all the data tables were merged together(file 1) followed by data cleaning, and EDA(file 2) to ease out future data analysis process further.

Insights:

The goal of this project is to solve two major problems, What should business do to get more profit and how to improve reseller efficiency?

Keeping the goal in mind this is what we understood from the initial descriptive analysis,

The business sells 266 products of at least 9 different colours. Because of varying sizes and colors, we have 107 total models of products. These products are divided into 4 major categories and 30+ subcategories. Business is selling through two modes, Resellers and Online. It is important to segment the right customers and resellers however we are seeing "[Not applicable]" as a top value which needs further performance analysis.

We also have demographic data for both resellers and customers, covering city names, state-province, and country-region. Broadly 7 regions divide their sales territory which further drops down to state and city. Interestingly, the count of unique reseller IDs is less than customer IDs but the reseller city count beats the customer city. Geographic sales analysis will help to find the pattern behind it.

Among many Numerical features, some caught my attention more than others which were decisive on which analysis processes/methods to choose from.

"Order Quantity" - 50% of orders have only 1 item means online customers play a big role but with avg of 2.7 and std of 2.5 data is normalized by bulk resellers.

"Unit Price" - available products have listed prices from 1.3 dollars to 3500+ dollars with an avg of 469 dollars and a high standard deviation which justifies max value however our business has products for customers of all economic backgrounds.

"Discount" - The business is not offering any noticeable discounts.

"Standard cost" - Huge gap in avg cost and avg list price, margin analysis will help to understand how to leverage this for profitability.

“Sales Amount” - The median sales amount is $136, meaning half of the orders are below this value. However, the average sales amount is $910 with a standard deviation of $1696, indicating a high positive skewness in the distribution.

Explanation of Analysis process/methods

(Answering questions of what the method is and why it was chosen)

Every analysis is connected and previous analysis helps the next one to find more insights and solutions to business problems.

Channel Analysis;

Channels are the core of the business's sales model. Understanding their performance helps to identify where to focus efforts for profitability. To understand which channel is contributing more to sales, revenue and profitability a simple comparison was done. Measures like profit margin and revenue generated were compared across the channels.

Category-subcategory Analysis;

Comparison of different subcategories across channels using total revenue and gross profit generated to understand profitability patterns for different product categories/subcategories. It will help understand top-performing and underperforming subcategories.

Geographic Analysis;

The performance of each channel based on city was done using total revenue and gross profit. This analysis helps to understand channel performance at the regional level for better profitability strategies. Sales and profits were mapped by region to tailor marketing, inventory and profitability strategies to regional preferences and demand.

Reseller Performance Analysis;

Evaluation of associated resellers based on revenue generated and margin earned by the business. To improve the efficiency and performance of resellers a categorization of low and high-performing resellers was important for profitability and future expansion goals.

Discount and Pricing Analysis;

Based on product subcategories and channels what discount offers to give, whom to give and at what time, different scatter plots were drawn. The business is not giving noticeable discounts to customers and resellers and this analysis will help to identify what to offer, whom to offer and what time to offer.

Product affinity and cross-selling Analysis;

Using Python's mlxtend.frequent_patterns module which provides tools(functions) for performing market basket analysis, was used to identify relationships between items in transactional datasets. Two functions (algorithms) named apriori and association_rules were used. This analysis helps understand which products are being bought together, and how products are related to each other. It will help understand customer buying patterns and help form discount strategies.

Results and Findings

Channel Analysis

Through this analysis, we will try to understand which channel (reseller and internet) for sales is good for our business. Through this pie chart (fig.1) it is clear, that total sales through the reseller channel account for ¾ th of total sales. However, this is not enough we also need to look into profit margin which simply means how much of the total revenue business is left with after deducting costs. To calculate profit margin we had to calculate gross profit and COGS (cost of goods sold).

```
Profit Margin = DIVIDE([Gross Profit], [Total Revenue], 0) * 100

Gross Profit = [Total Revenue]-[COGS]

COGS = sum(Sales[Product Standard Cost])
```

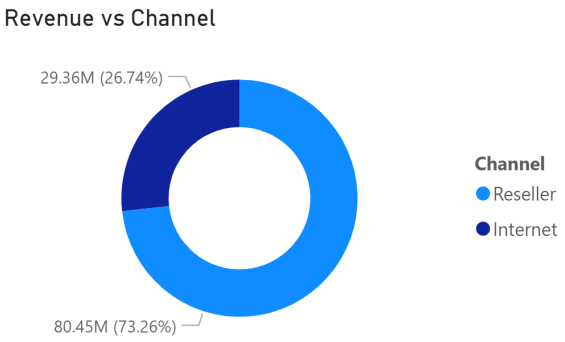


Fig. 1

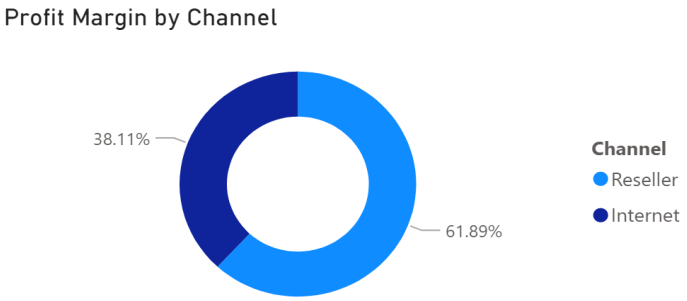


Fig. 2

Speaking through the language of data, the reseller channel is more profitable than the Internet. But which product, category, subcategory, and geographic regions give the business high profits and where it lacks? The following 2 analyses will help to understand the patterns.

Category - subcategory analysis

266 SKUs are divided into 4 categories and 37 subcategories, however, the business is selling all 4 categories through the reseller channel but only 3 through the online channel by leaving “components” (Table. 4)

Table. 3

Accessories	Bikes	Clothing	Components
Bike Racks	Mountain Bikes	Caps	Brakes
Bike Stands	Road Bikes	Bib-Shorts	Chains
Bottles and Cages	Touring Bikes	Gloves	Cranksets
Cleaners		Jerseys	Derailleurs
Fenders		Shorts	Forks
Helmets		Socks	Handlebars
Hydration Packs		Tights	Headsets
Lights		Vests	Bottom Brackets
Locks			Mountain Frames
Panniers			Pedals
Pumps			Road Frames
Tires and Tubes			Saddles
			Touring Frames
			Wheels

Channel	Internet		Reseller		**Total**	
Category	Total Revenue	Gross Profit	Total Revenue	Gross Profit	**Total Revenue**	**Gross Profit**
Bikes	2,83,18,144.65	1,15,05,796.50	6,63,02,381.56	4,42,63,319.17	**9,46,20,526.21**	**5,57,69,115.67**
Components			1,17,99,076.66	75,09,942.63	**1,17,99,076.66**	**75,09,942.63**
Clothing	3,39,772.61	1,36,412.58	17,77,840.84	14,83,329.57	**21,17,613.45**	**16,19,742.15**
Accessories	7,00,759.96	4,38,674.57	5,71,297.93	5,00,175.04	**12,72,057.89**	**9,38,849.60**

Table. 4

In terms of revenue and gross profit (overall), Table 4 ranks categories in this order: “Bikes” is followed by “Components,” “Clothing,” and “Accessories.”

This trend remains the same for the Reseller channel, but in the internet channel, “Accessories” surpasses the “Clothing” category in both revenue and gross profit.

Top 10 subcategories (Internet)

Channel	Internet	
Subcategory	Total Revenue	Gross Profit
Road Bikes	1,45,20,584.04	55,37,299.70
Mountain Bikes	99,52,759.56	45,13,624.11
Touring Bikes	38,44,801.05	14,54,872.70
Tires and Tubes	2,45,529.32	1,53,700.75
Helmets	2,25,335.60	1,41,059.83
Jerseys	1,72,950.68	39,778.66
Shorts	71,319.81	44,646.16
Bottles and Cages	56,798.19	35,555.35
Fenders	46,619.58	29,183.90
Hydration Packs	40,307.67	25,232.57

Table. 5

Top 10 subcategories (Reseller)

Channel	Reseller	
Subcategory	Total Revenue	Gross Profit
Road Bikes	2,93,58,206.96	1,90,27,331.61
Mountain Bikes	2,64,92,684.38	1,88,49,520.59
Touring Bikes	1,04,51,490.22	63,86,466.97
Mountain Frames	47,13,672.15	29,83,724.91
Road Frames	38,49,853.34	23,25,360.59
Touring Frames	16,42,327.69	10,35,023.84
Wheels	6,79,070.07	4,99,803.60
Jerseys	5,79,308.71	4,40,791.28
Shorts	3,42,202.72	3,01,402.02
Helmets	2,58,712.93	2,22,693.64

Table. 6

Top 15 subcategories (overall)

Subcategory	Total Revenue	Gross Profit
Road Bikes	4,38,78,791.00	2,45,64,631.31
Mountain Bikes	3,64,45,443.94	2,33,63,144.70
Touring Bikes	1,42,96,291.27	78,41,339.67
Mountain Frames	47,13,672.15	29,83,724.91
Road Frames	38,49,853.34	23,25,360.59
Touring Frames	16,42,327.69	10,35,023.84
Jerseys	7,52,259.39	4,80,569.94
Wheels	6,79,070.07	4,99,803.60
Helmets	4,84,048.53	3,63,753.47
Shorts	4,13,522.53	3,46,048.18
Vests	2,59,488.37	2,25,337.31
Tires and Tubes	2,46,454.53	1,54,486.35
Gloves	2,42,795.87	2,03,400.05
Bike Racks	2,37,096.16	2,01,371.68
Cranksets	2,03,942.62	1,47,660.82

Table. 7

It is important to understand which products are doing better than others at both channels. Table 5 and Table 6 rank the top 10 subcategories of products in both channels and Table 7 ranks the top 15 subcategories overall. Because of variations in colour and size, different products come under a single subcategory but at a broader level subcategory trends are more important to understand which will help in further analysis.

Now we have understood the profitability trend of different subcategories and categories let us understand the profitability trend with geographical regions.

Geographic analysis

Although the business has divided its operations into territories, regions, groups, state-province and at the lowest level lies the 'city' feature, in which there are some cities where business is selling and serving either through resellers or online however there are some cities where both channels are present. We will divide them into 3 separate categories,

Type 1, Business through online channel

Type 2, Business through reseller channel

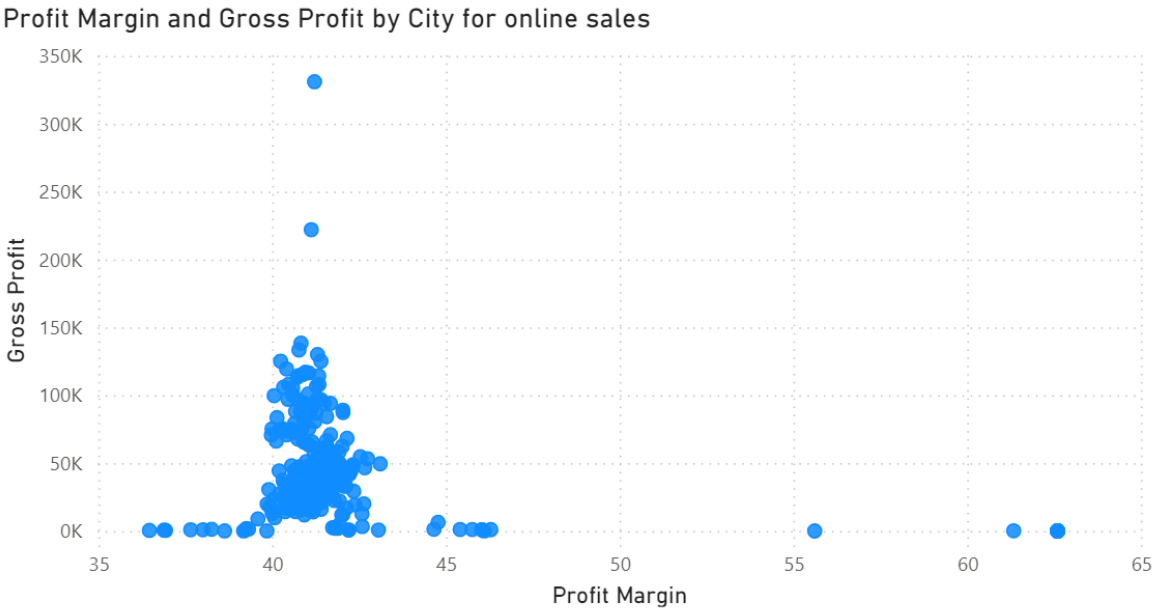
Type 3, Both channels are available

Table. 8

Customer_City	Gross Profit	Total Revenue
London	3,30,909.18	8,02,810.30
Paris	2,21,960.11	5,39,725.80
Wollongong	1,38,376.45	3,38,913.47
Warrnambool	1,33,333.74	3,27,036.37
Bendigo	1,29,933.41	3,14,568.72
Goulburn	1,25,104.21	3,10,875.90
Bellflower	1,25,152.09	3,02,278.81
Brisbane	1,19,361.48	2,95,353.58
Townsville	1,16,931.74	2,85,486.91
Geelong	1,16,499.02	2,83,802.17

Table. 8 ranks the top 10 profitable cities based on online channel. The business sells and serves in 269 cities with only online channel and it needs to filter the cities based on high profit margin and gross profit. The graph (fig. 3) gives a better understanding of which cities are doing better than others. Based on it business can strategies their marketing and expansion plans.

Fig. 3



This same pattern can be seen for Type 2 sales. Table 9 and Fig. 4 explain it explicitly.

Table 9

Customer_City	Gross Profit	Total Revenue
London	3,30,909.18	8,02,810.30
Paris	2,21,960.11	5,39,725.80
Wollongong	1,38,376.45	3,38,913.47
Warrnambool	1,33,333.74	3,27,036.37
Bendigo	1,29,933.41	3,14,568.72
Goulburn	1,25,104.21	3,10,875.90
Bellflower	1,25,152.09	3,02,278.81
Brisbane	1,19,361.48	2,95,353.58
Townsville	1,16,931.74	2,85,486.91
Geelong	1,16,499.02	2,83,802.17

Fig. 4

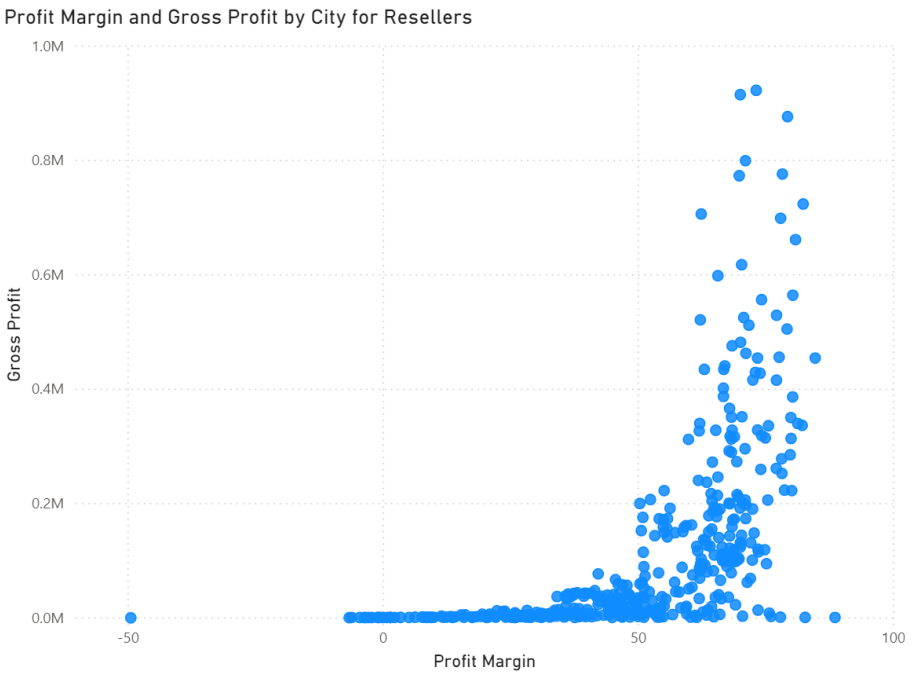


Fig. 5 shows the locations of cities where our resellers and online customers order. However, most online orders come from the close proximity of cities where resellers serve but some don't, like in Africa and Australia.

Resellers

Fig. 5

Online customers

For cities of Type 3, where both channels are used, business can compare the performance of both channels for which multiple factors can be considered like cost, profitability, expansion plans etc. Table 10 gives a glimpse of revenue and profit trends for cities.

Channel	Internet			Reseller		
City	Both Channels	Total Revenue	Gross Profit	Both Channels	Total Revenue	Gross Profit
Barstow	1	3,578.27	1,406.98	1	28,783.54	16,827.52
Basingstoke Hants	1	3,271.68	1,251.69	1	22,606.10	7,001.50
Baytown	1	25.48	15.95	1	14,042.08	7,468.36
Beaverton	1	1,61,959.43	68,272.17	1	1,98,487.88	1,26,252.09
Bell Gardens	1	5,920.24	2,475.11	1	6,221.05	1,747.80
Bellevue	1	2,049.10	943.29	1	4,43,316.11	3,35,228.51
Bellingham	1	2,07,613.25	87,279.32	1	7,48,202.72	5,55,694.02
Berks	1	46,935.11	20,011.97	1	1,41,834.27	98,238.78
Berkshire	1	58,119.86	23,628.91	1	3,02,043.77	2,10,540.78
Berlin	1	2,60,930.63	1,07,890.67	1	2,55,129.64	1,50,163.79
Billings	1	92.08	57.64	1	9,530.02	-89.29

Table 10

Reseller performance analysis

From earlier analysis we have a clear understanding of which channel is best for the business, and which product categories and subcategories give high profits, similar analyses were conducted with cities however reseller efficiency was the key player. It is essential to select the right kind of resellers that give high profit, and high margins at low COGS.

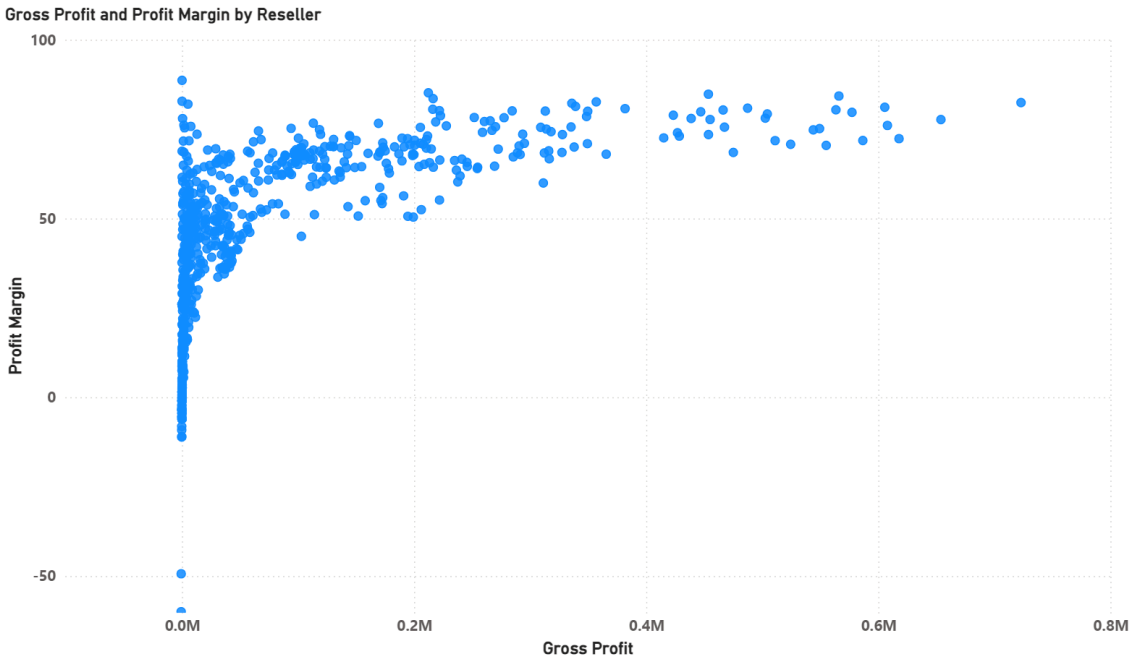


Fig. 6

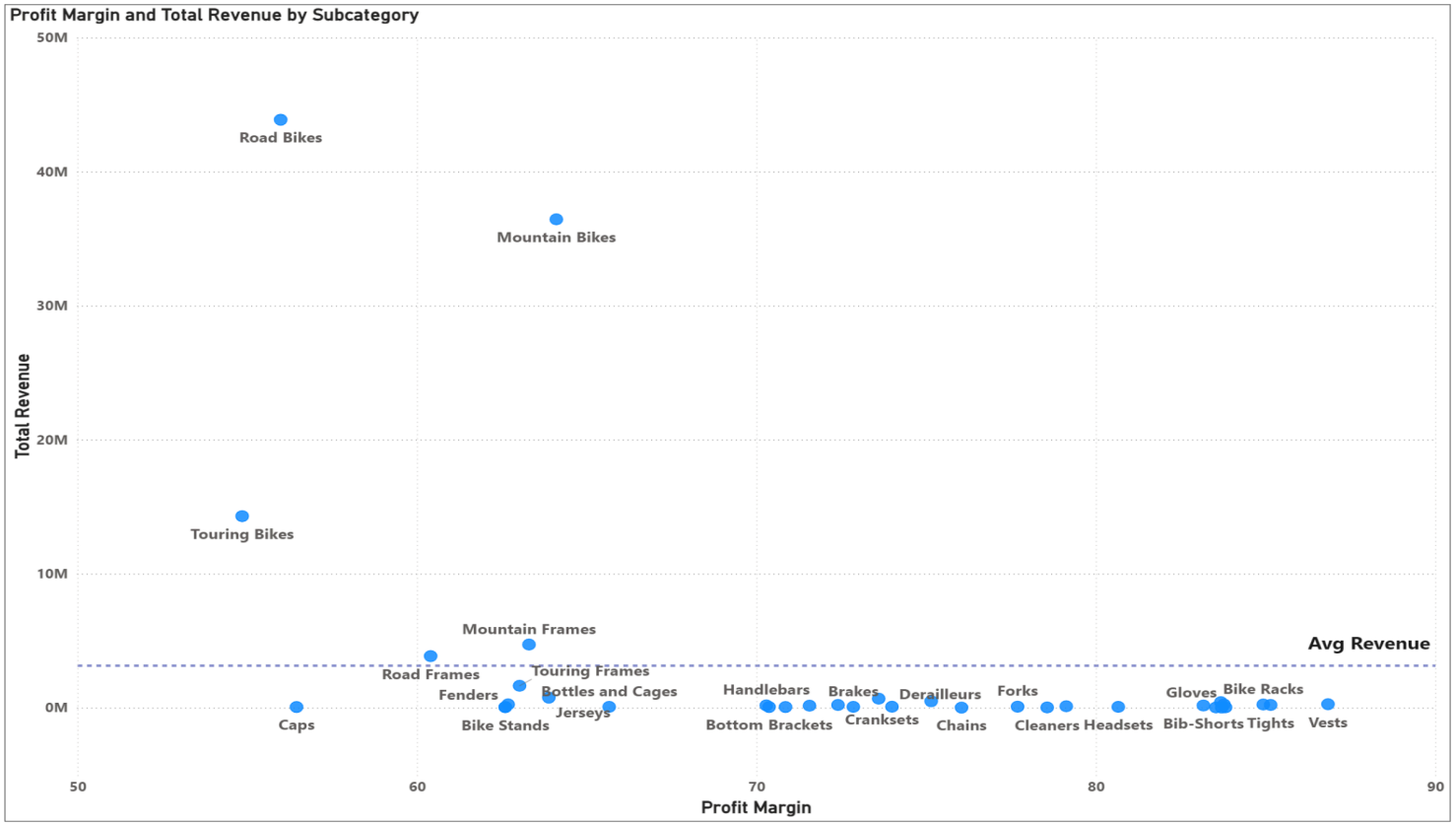
This graph (Fig.6) depicts all the resellers, as we can see number of resellers brings zero and even negative gross profits for the business. Ideal resellers are ones with higher gross profits and higher margins.

Discount and Pricing Analysis:

As we know business is offering unnoticeable discounts to online customers or resellers, and even to those who come under the metrics of loyal customers i.e. recency, frequency, and monetary value. We already know who are our loyal customers on both channels and how to filter them however there are more factors to consider. This analysis will help when to offer discounts, on which products and to whom.

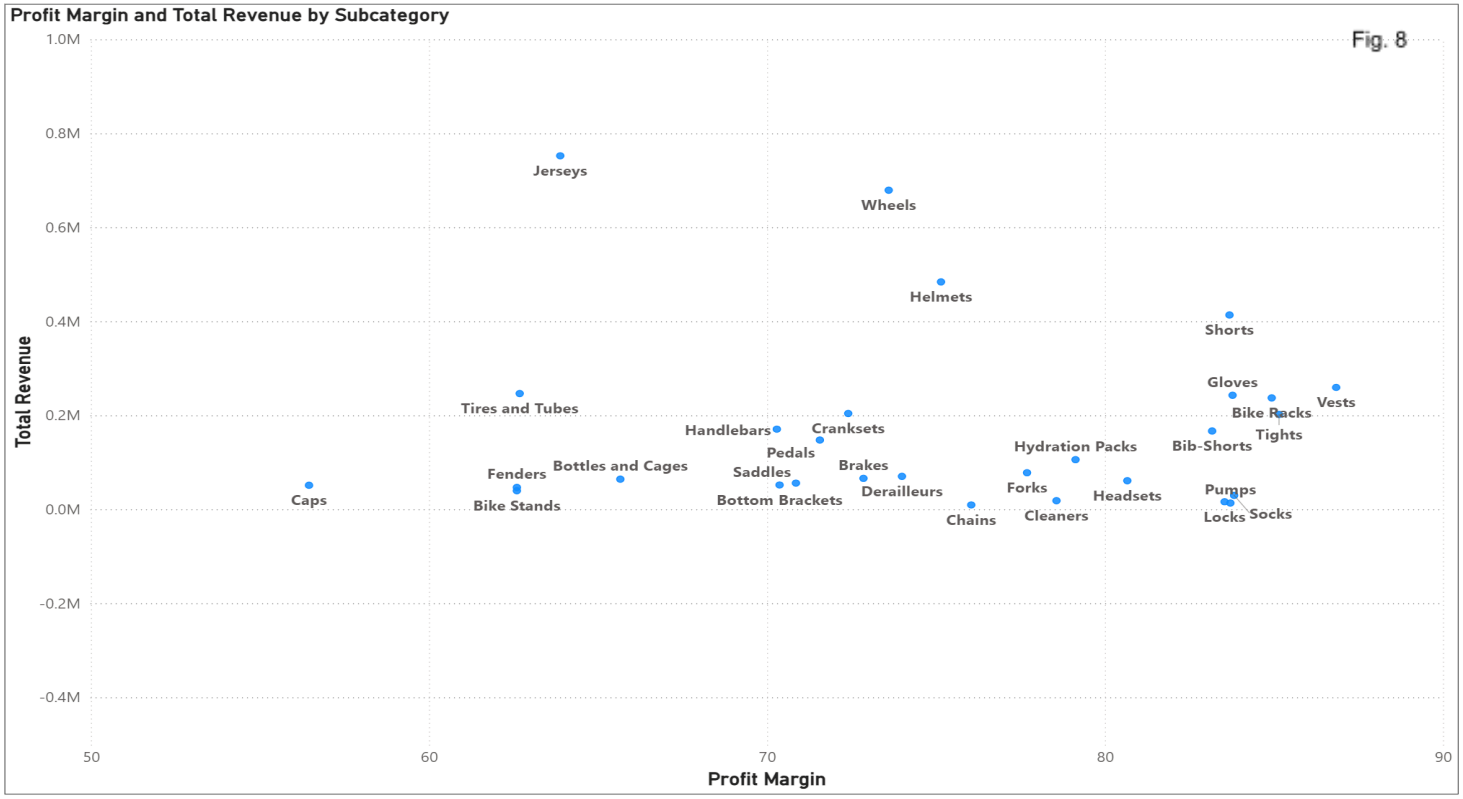
This graph (fig.7) gives us patterns of subcategories grouped with how much profit and revenue they bring to the business.

Fig. 7



If you look closely we have two areas of focus first one is those subcategories that are generating high revenue but have a low margin and the second one is those subcategories where the business has high margins of profit but a low number of sales.
If we take a closer look at the second type of subcategory by filtering total revenue below 1 million dollars we could observe which subcategories we need to target.

Now that have understood which subcategories/categories we need to experiment with discount and pricing strategies, let's explore two more aspects.
When to offer discounts or experiment pricing strategies and to whom it to offer.



The graph below(fig.9) traces patterns of months where we tend to have less revenue. If we look closely months of March, Oct we have had high dips since the last 3 years of collected data.

Fig. 9

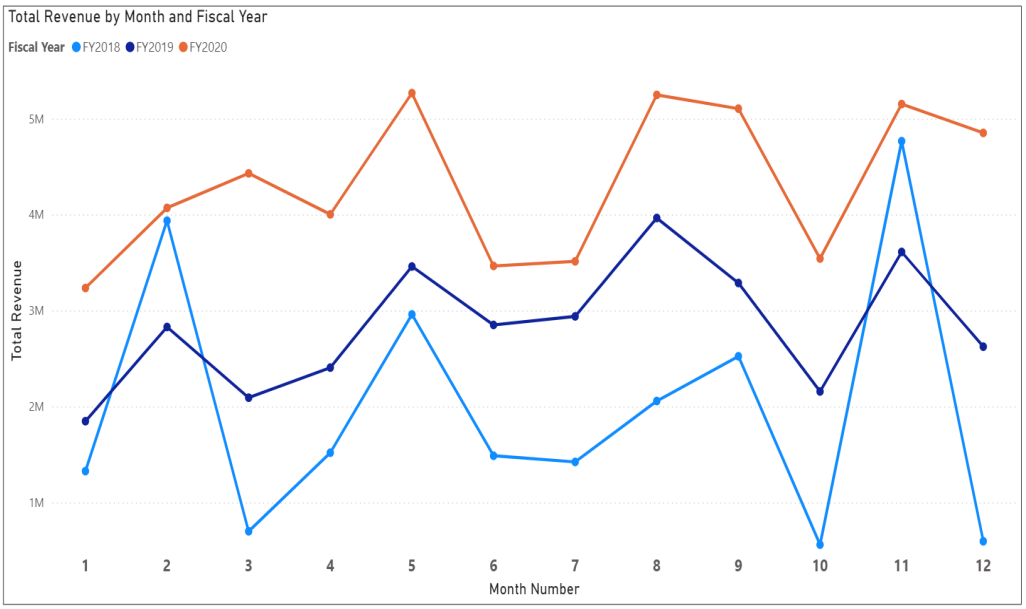


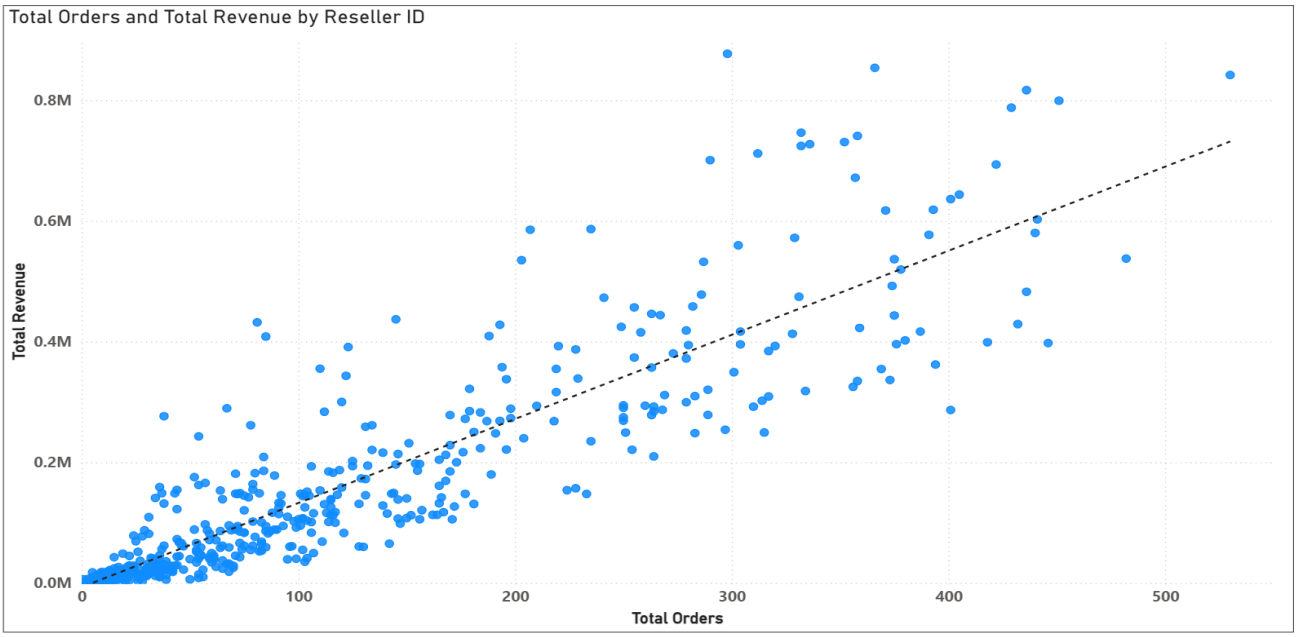
Table 11

best to worst month by revenue

Month Number	Month Name	Total Revenue
11	November	1,35,27,908.97
5	May	1,16,85,618.82
8	August	1,12,69,868.99
9	September	1,09,15,641.41
2	February	1,08,35,870.84
12	December	80,72,018.63
4	April	79,27,860.59
7	July	78,76,112.37
6	June	78,03,561.30
3	March	72,23,140.06
1	January	64,10,553.78
10	October	62,61,118.45

Through reseller performance analysis (fig. 6) we have already filtered out our best-performing resellers with whom discount offers/strategies can be implemented. The graph below (fig. 10) also reiterates this.

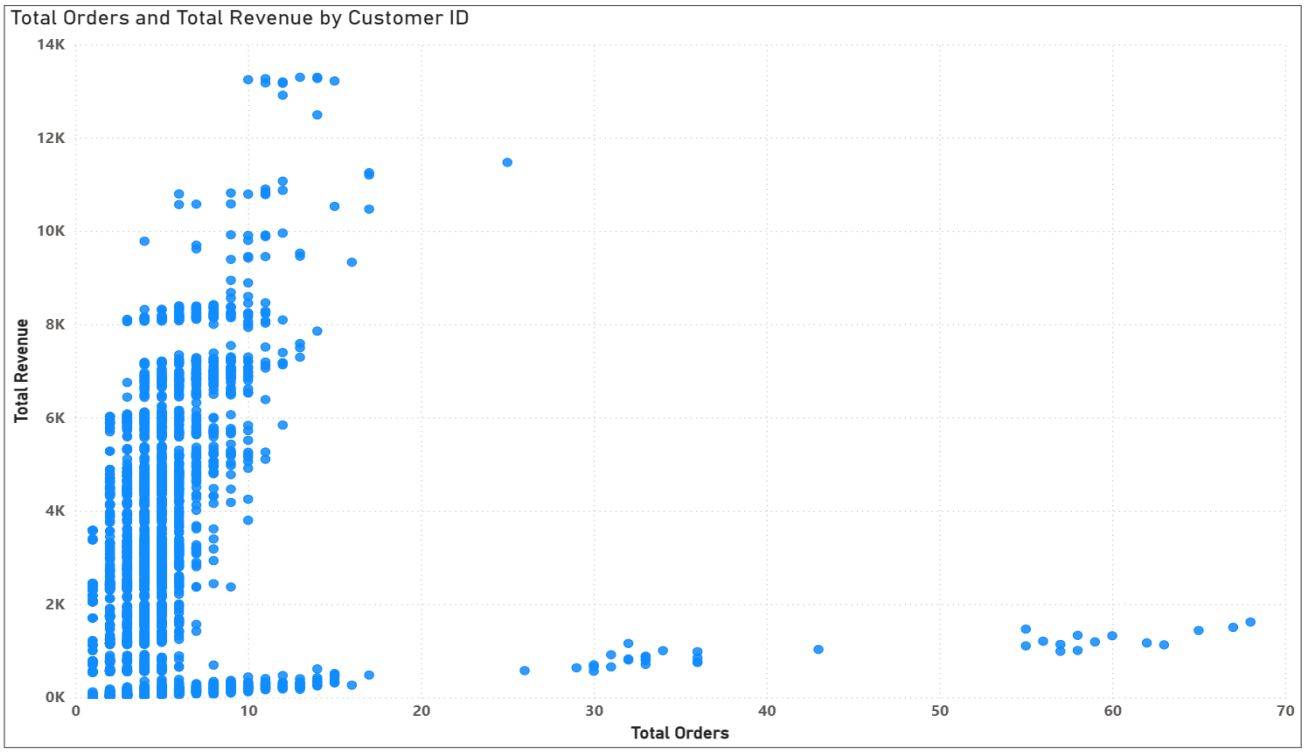
Fig. 10



However in the case of customers, this trend is different, here discount strategies should be based on two types of segmented customers, first; low volume - high revenue (left cluster), second; high volume, low revenue (right cluster).

The ideal quadrant is high volume - high revenue.

Fig. 11



Product Affinity and cross-sell analysis using Python

Now let's understand customer purchasing patterns, identify cross-selling opportunities and better product bundling combinations for discount offers. The results can also be used for product recommendations.

The tables have been successfully merged (refer to file 1), and subsequent data cleaning and exploratory data analysis have been completed (refer to file 2).

Steps involve,

1. Transforming the dataset into a transaction-product matrix (new data frame), by grouping data using the 'sales order' column, where the first column represents a transaction (sales order) and the second column consists of all the products under that sales order(sale order line).

2. For this analysis we used 'mlxtend.frequent_patterns' module which provides tools(functions) for performing 'market basket analysis', which is used to identify relationships between items in transactional datasets. Two functions (algorithms) named apriori and association_rules were used.

apriori - This function is used to identify frequent itemsets in transactional data. It applies the Apriori algorithm, which finds product combinations that appear together in transactions.

association_rules - This function generates association rules from the frequent itemsets discovered by apriori. Rules help understand the relationship between products, showing how the purchase of one item (antecedent) leads to the purchase of another (consequent).

In the result, we get two tables, the rules table, and the frequent_itemsets table. Table 1(fig. 12) only lists the first two columns of the rules table while Table 2(fig. 13) lists down the first 10 observations of frequent_items table. More detailed and in-depth explanations can be found in shared code files.

The frequent items table provides insights into which products or combinations of products are commonly purchased together by customers and the rule table which builds on the frequent table shows the relationship between products(purchasing pattern).

Fig. 12

```
[32]  1  rules.loc[:, ['antecedents', 'consequents']]
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_
  and should_run_async(code)
```

	antecedents	consequents
15765	(Classic Vest, S, AWC Logo Cap, Classic Vest, M)	(Short-Sleeve Classic Jersey, XL, Short-Sleeve...
16219	(Hitch Rack - 4-Bike, AWC Logo Cap, Classic Ve...	(Short-Sleeve Classic Jersey, XL, Short-Sleeve...
20240	(Hitch Rack - 4-Bike, AWC Logo Cap, Hydration ...	(Short-Sleeve Classic Jersey, XL, Short-Sleeve...
22845	(Classic Vest, S, Classic Vest, M, Bike Wash -...	(Short-Sleeve Classic Jersey, XL, Short-Sleeve...
25095	(Classic Vest, S, Hydration Pack - 70 oz., Bik...	(Short-Sleeve Classic Jersey, XL, Short-Sleeve...
...	...	...
159339	(Water Bottle - 30 oz.)	(Classic Vest, S, Short-Sleeve Classic Jersey,...
160358	(Water Bottle - 30 oz.)	(Short-Sleeve Classic Jersey, XL, Classic Vest...
165970	(Water Bottle - 30 oz.)	(Classic Vest, S, Short-Sleeve Classic Jersey,...
166480	(Water Bottle - 30 oz.)	(Classic Vest, S, Short-Sleeve Classic Jersey,...
167500	(Water Bottle - 30 oz.)	(Short-Sleeve Classic Jersey, XL, Classic Vest...

169034 rows × 2 columns

Fig. 13

```
1  frequent_itemsets.head(10)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel
  and should_run_async(code)
```

	support	itemsets
0	0.107519	(AWC Logo Cap)
1	0.042187	(Bike Wash - Dissolver)
2	0.017644	(Classic Vest, M)
3	0.021682	(Classic Vest, S)
4	0.067430	(Fender Set - Mountain)
5	0.011477	(Full-Finger Gloves, L)
6	0.011286	(Full-Finger Gloves, M)
7	0.015292	(HL Mountain Frame - Black, 42)
8	0.015610	(HL Mountain Frame - Silver, 38)
9	0.044381	(HL Mountain Tire)

Interpretation of Results and Recommendations

Interactive reports using Power BI and a recommendation system using Python were created during the implementation of the different analysis processes. Business can interact with different graphs and get a deeper understanding of patterns and insights.

1. From channel analysis we found that resellers account for approximately 73.26% of total revenue generated making it the dominant source of sales. This trend is also found in how much profit margin business gets through the reseller channel which is approximately 62%. Business should prefer segmented resellers(not all) over online channel after considering which product subcategories are profitable for which region and which region(city) is best for which kind of channel because of varying costs.

2. Table 5,6,7 lists down which product categories business gets the most revenue and gross profit. Business can draw more resources and attention to these categories, rearrange inventory, clear stocks of low-performing subcategories using discount offers and focus future plans for filtered subcategories.

3. From geographic analysis we understood that there were some cities where business was selling and serving either through online channel or reseller channel and in some cities both channels were active. We filtered out the best city for each type.

We divided all cities into 3 types,

For the first type of city where business is only using online channel, Business should strengthen their online presence with targeted marketing and exclusive offers(using insight from product affinity and cross-sell analysis). Business should also plan for opening reseller partnerships to get higher margins.

For the second type of city where business is only using reseller channel, It should filter out reseller performance, incentivize high-performing resellers and rethink relationships with low/negative performing resellers either by complete withdrawal or by giving required support.

For the third type of city where business is using both channels, using Table 7 business can understand that at a given city which channel is performing well and plan accordingly. It should use a Hybrid strategy where one channel supports the other. Online channel support resellers by marketing incentives/offers for store visits and similarly stores should also use strategies to influence/entice store customers.

For this hybrid model business can use insights from product affinity and cross-sell analysis to understand customer behaviour and modify strategies accordingly.

4. From the Reseller performance analysis we created a scatterplot depicting each reseller on two parameters, first gross profit and second profit margin. Business needs to incentivize

resellers with high profit and high margin and rethink their relations with low-performing ones.

We found that there were resellers who were generating revenue either in negative or close to zero which the business needs to address effectively.

5. In discount and pricing analysis we tried to find answers 3 questions, If we are offering discounts then what product subcategories we should choose from, what time of the year we should experiment with discount and pricing stunts and which resellers/customers should we target? What we found,

We found that there was an imbalance of data when it came to product subcategories's revenue and profit margin trends. There were a number of subcategories whose revenue was below the average level. We found a cluster of subcategories with high margins but low revenue.

We also found a seasonal trend where in few months sales were really bad throughout the span of 3 financial years. Table 11 ranks all the months by total revenue generated.

we found that through the reseller channel bulk orders translate to high revenue but with online customers who tend to buy fewer items, different clusters were formed in our scatterplot for different customers and their buying patterns. There was a range of revenue value generated for the same level of sales order bcs some customers were buying premium products and others weren't. Some customers were ordering high but generating little revenue

Through our finding business could strategies its discount and pricing strategies like this,

- Promotion of high-margin low revenue products using bundling strategies, seasonal promotions and offers.
- For low-margin high revenue products business could renegotiate margin terms or reduce COGS
- Discount campaigns in low revenue months using customer-specific preference, organizing flash sales.
- Incentivizing high-performing resellers, rethinking relationships with low-performing resellers
- For low volume, high revenue customers offer discount offers for repeat purchases, For high volume, low revenue customers offer upselling or conditional discounts.
- Dynamic pricing for online customers

6. We found two tables as a result of apriori and association_rules algorithms. Frequent Items Table and Rules Table

—> Frequent items table

This table provides insights into which products or combinations of products are commonly purchased together by customers. It has two key columns:

Itemset: Lists the individual product(s) or combinations of products. Example: {Product A, Product B} means these two products are often bought together.

Support: Represents the percentage of transactions where this product or product combination appears.

Example: If the support for {Product A, Product B} is 0.10, it means 10% of all transactions include this combination.

Business Insight:

The frequent items table helps business understand which products are popular or frequently bought together, allowing us to create product bundles or prioritize their availability in stock.

For example, if customers often buy "bicycle tyres" and "repair kits" together, business can bundle them as a package to boost sales.

—> Rules Table

This table builds on the frequent items to show relationships between products. It tells us what happens when customers buy certain items. The key columns include:

Antecedents: The product(s) customers purchase first (e.g., {Product A}).

Consequents: The product(s) they tend to buy afterward (e.g., {Product B}).

Support: The percentage of transactions where both the antecedents and consequents appear together.

Confidence: The likelihood that customers who buy the antecedent will also buy the consequent. For instance, if confidence is 80%, 80% of people buying {Product A} also buy {Product B}.

Lift: How much more likely a customer is to buy the consequent when the antecedent is purchased, compared to random chance. A lift > 1 indicates a strong relationship.

Conviction, Zhang's Metric, and Other Metrics: Additional metrics for advanced analysis of rule strength and reliability.

Business Insight:

The rules table reveals actionable cross-selling opportunities. For example:

If the rule {Bicycle Frame} → {tyres} has high confidence (e.g., 75%) and lift (e.g., 2.5), it means customers who buy frames are highly likely to buy tyres.

Using this business can recommend tyres to customers purchasing frames, either as a bundle or with a small discount to encourage sales.

Recommendations

1. Channel Optimization

- Focus on reseller channels for profitability, contributing 73.26% of total revenue and 62% profit margin.
- Segment resellers by product subcategory and regional performance for better alignment with channel-specific costs.

2. Inventory and Product Strategy

- Prioritize high-revenue, high-gross-profit categories; reallocate resources accordingly.
- Clear low-performing subcategories using discount offers.

3. Geographic Strategy

- Online-only cities: Strengthen online presence through targeted marketing and exclusive offers; consider establishing reseller partnerships for higher margins.
- Reseller-only cities: Incentivize high-performing resellers; reevaluate or withdraw from low/negative-performing resellers.
- Hybrid cities: Leverage insights to determine the dominant channel; adopt a hybrid strategy where one channel supports the other using cross-channel incentives and promotions.

4. Reseller Management

- Incentivize high-performing resellers with high profit and margin.
- Address resellers generating low or negative revenue through performance review or support strategies.

5. Discount and Pricing Strategies

- Promote high-margin, low-revenue products through bundling, seasonal campaigns, and special offers.
- Reevaluate COGS or renegotiate terms for low-margin, high-revenue products.
- Target low-revenue months with customer-specific discounts and flash sales.
- Incentivize repeat purchases for high-revenue, low-volume customers. Offer upselling and conditional discounts for high-volume, low-revenue customers.
- Use dynamic pricing to maximize revenue from online customers.

6. Seasonality Insights

- Focus discount and promotional activities during months with consistently low sales.
- Use historical trends to tailor campaigns and optimize resource allocation during high-potential periods.

7. Product Bundling and Cross-Selling Opportunities

- Bundle Frequently Bought Together Items:
  Utilize insights from the Frequent Items Table to create attractive bundles of commonly purchased items, such as "bicycle tyres + repair kits," to boost sales and improve customer convenience.
- Cross-Sell Using Rules Table:
  Leverage high-confidence and high-lift rules to recommend related products to customers. For instance, if {Bicycle Frame} → {tyres} has a high lift, offer a targeted discount or recommendation for tyres to customers purchasing frames.