



Data Analytics Project: performing EDA on e- commerce platform sales data

NAZIM ATAKAN ERDOGAN | Data Analyst Intern

About me

- Currently pursuing BS in Computer Engineering at RWTH Aachen University, Germany; doing an internship in Data Analytics.
- Ex medical student at the University of Cologne
- Interested in scientific data analysis, astronomical data analysis, bioinformatics, clinical trial data analytics, biostatistics and healthcare data.
- As a newbie in data field, I am trying to combine my science background with data perspective.





AGENDA

1. Data Formatting and Cleaning
2. Preparing a table of Sales & Profit month-wise
3. Finding the average profit of each Product Category, region wise.
4. Finding the average sales of each Product, region wise.
5. Finding the the top 3 highest performing product and the bottom 3 lowest performing products.
6. Creating Chart of the month-wise table and region-wise table on the Profit.
7. Showing the statistical analysis of the Sales and Profit data.
8. Creating a sales dashboard



OVERVIEW OF THE STUDY

Business Scenario:

The objective is to design a comprehensive sales dashboard that integrates data from an e-commerce platform operating across various domains. The dashboard will focus on:

Sales Performance Analysis: Evaluate sales trends and metrics to track overall platform performance.

Category-Based Insights: Analyze sales distribution and performance across different product categories, enabling detailed comparisons and strategic decision-making.

Visual Representations: Incorporate charts, graphs, and other visual elements to provide actionable insights and facilitate data-driven strategies.

Data Overview: Order ID, Order Date, Ship Date, Aging, Ship Mode, Product Category, Products, Sales, Quantity, Discount, Profit, Shipping Cost, Order Priority, Customer ID, Customer Name, Segment, City, State, Country, Region and Months. More than 50000 rows and 20 columns of data

Project1

File

Edit

View

Insert

Format

Data

Tools

Extensions

Help

Project1

File

Edit

View

Insert

Format

Data

Tools

Extensions

Help

</



1. DATA CLEANING AND FORMATTING

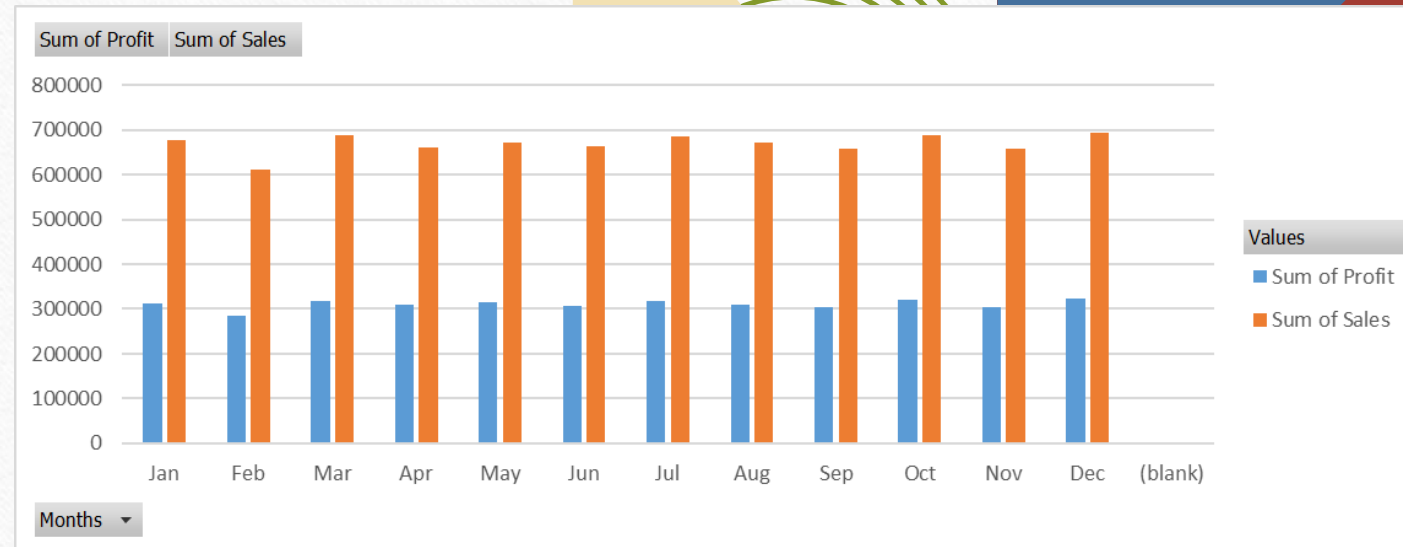
CUSTOM AND CONDITIONAL FORMATTING;
FINDING DUPLICATES AND MISSING VALUES

- No duplicates found, date formatted to different form, data cleaned and purified.

Product	Sales	Quantity	Column1	Profit	Shipping Cost	Order Priority
Media Players	\$ 140,0	1	0,01	\$ 58,6	\$ 5,9	High
Speakers	\$ 211,0	3	0,01	\$124,7	\$ 12,5	High
Body Covers	\$ 117,0	5	0,04	\$ 13,6	\$ 1,4	High
& Bike Care	\$ 118,0	5	0,02	\$ 26,2	\$ 2,6	Medium
e	\$ 250,0	3	0,02	\$155,0	\$ 15,5	High
e Tyres	\$ 72,0	5	0,05	\$ 14,4	\$ 1,4	Critical
Mat	\$ 54,0	1	0,02	\$ 54,0	\$ 5,4	High
Seat Covers	\$ 114,0	5	0,05	\$ 5,5	\$ 0,6	Critical
Pillow & Neck Rest	\$ 231,0	2	0,02	\$141,8	\$ 14,2	Critical
Media Players	\$ 140,0	1	0,03	\$ 55,8	\$ 5,6	High
Speakers	\$ 211,0	4	0,02	\$114,1	\$ 11,4	Critical
Body Covers	\$ 117,0	5	0,04	\$ 13,6	\$ 1,4	Critical
& Bike Care	\$ 118,0	5	0,04	\$ 14,4	\$ 1,4	Critical
e	\$ 250,0	4	0,05	\$120,0	\$ 12,0	Medium
e Tyres	\$ 72,0	5	0,05	\$ 14,4	\$ 1,4	Critical
Mat	\$ 54,0	5	0,03	\$ 10,8	\$ 1,1	Medium
Seat Covers	\$ 114,0	4	0,05	\$ 11,2	\$ 1,1	Critical
Pillow & Neck Rest	\$ 231,0	1	0,05	\$139,5	\$ 13,9	Critical
Media Players	\$ 140,0	5	0,01	\$ 52,0	\$ 5,2	Critical

2. SALES & PROFITS PER MONTH

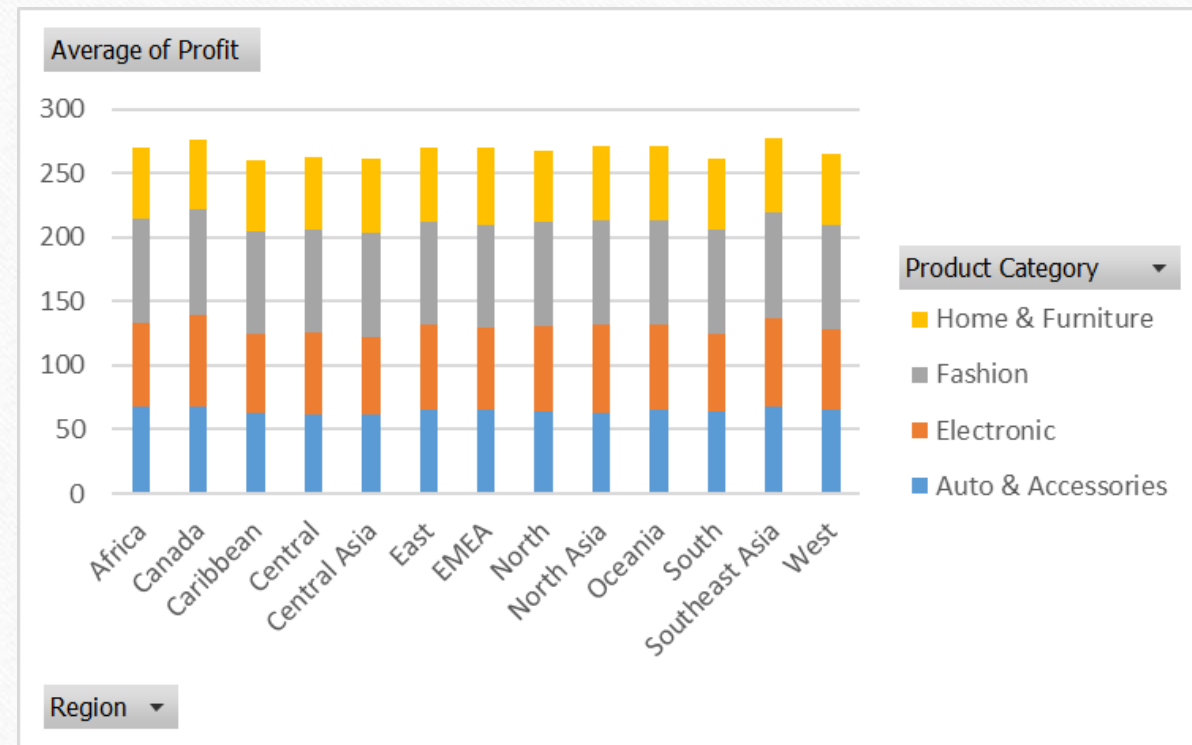
Row Labels	Sum of Profit	Sum of Sales
Jan	313566,3467	676313
Feb	286102,6233	610240
Mar	317186,0067	686681
Apr	308364,5133	659404
May	313751,2467	672547
Jun	307585,0233	664560
Jul	318703,2	685152
Aug	310442,8433	670788
Sep	305334,4567	658844
Oct	320748,67	689116
Nov	304716,1	656663
Dec	323401,92	693073
(blank)		
Grand Total	3729902,95	8023381



3. THE AVERAGE PROFIT OF EACH PRODUCT CATEGORY, REGION WISE.

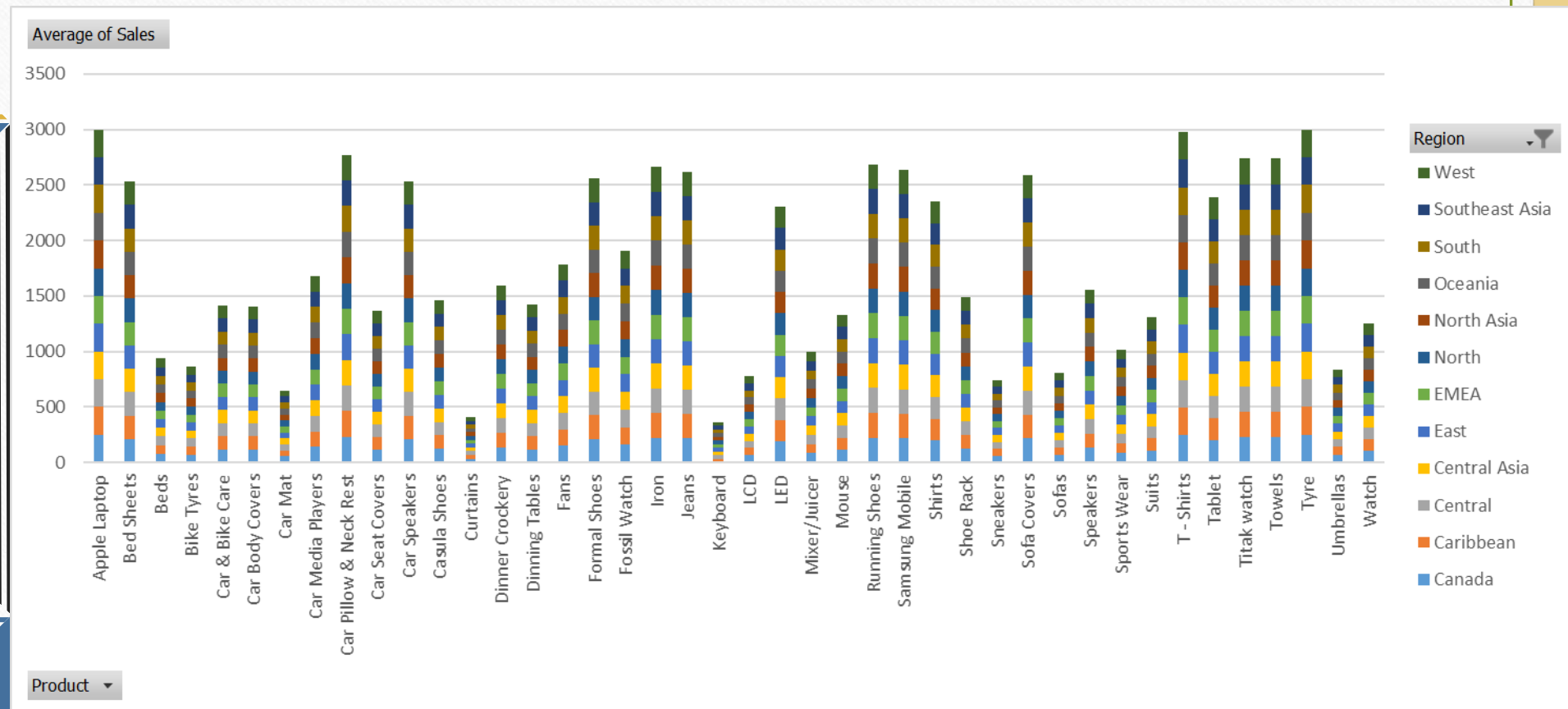
The company operated in a balanced way in all commercial areas, we can see a near equal distribution of average profit of each individual product.

It can be seen from the chart that the field “Fashion” plays a huge role in retail profits of the company,



4. THE AVERAGE SALES OF EACH PRODUCT, REGION WISE.

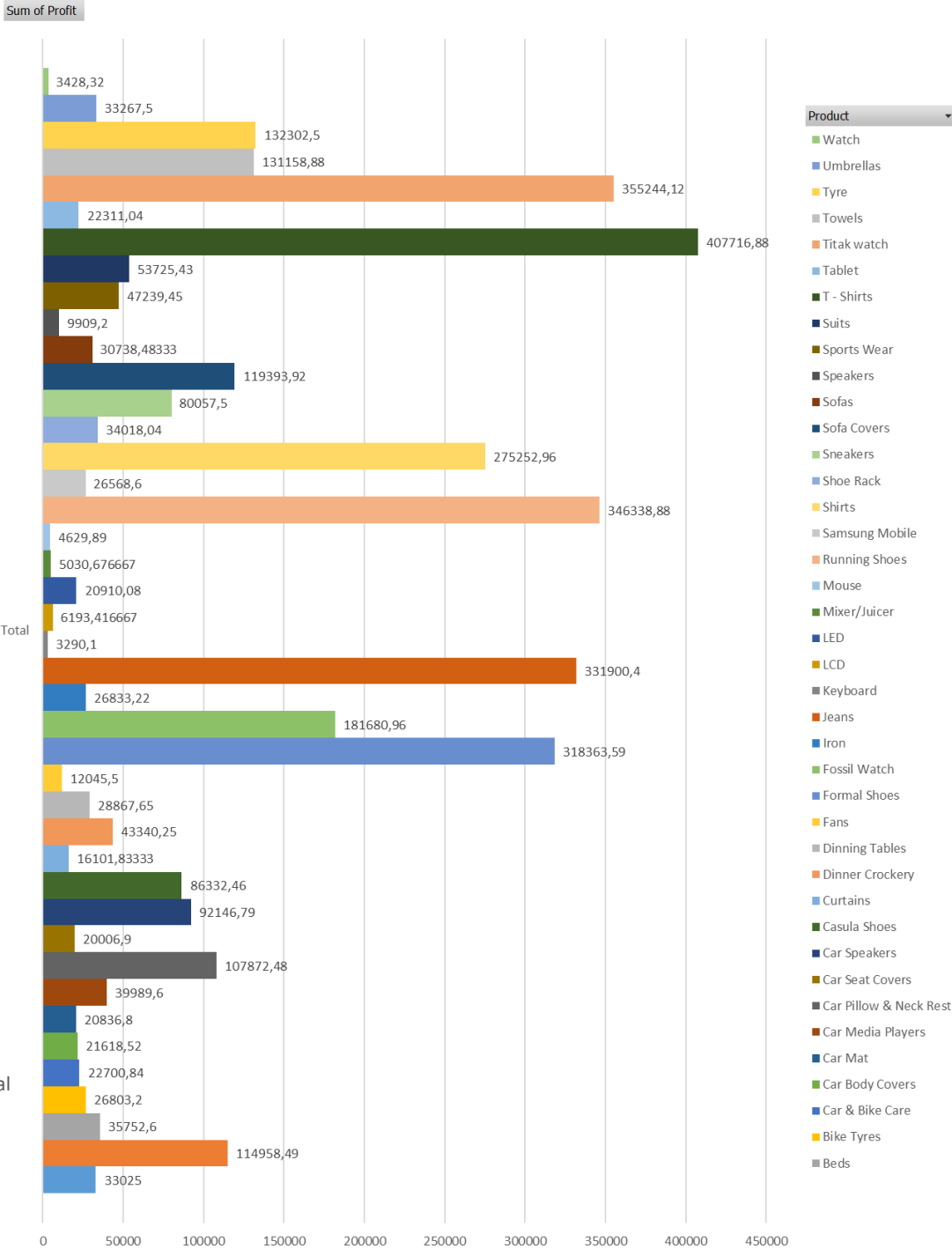
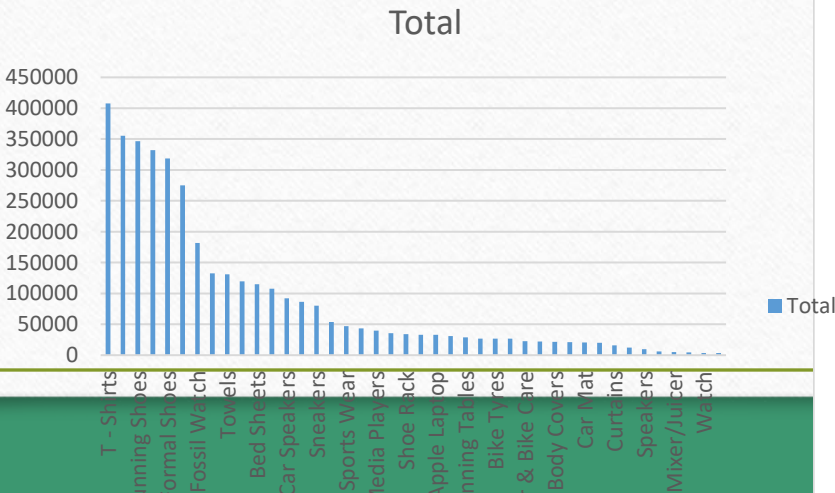
Here we can see the average sales of each products according to different regions.



Row Labels	Sum of Profit
T - Shirts	407716,88
Titak watch	355244,12
Running Shoes	346338,88
Jeans	331900,4
Formal Shoes	318363,59
Shirts	275252,96
Fossil Watch	181680,96
Tyre	132302,5
Towels	131158,88
Sofa Covers	119393,92
Bed Sheets	114958,49
Car Pillow & Neck Rest	107872,48
Car Speakers	92146,79
Casula Shoes	86332,46
Sneakers	80057,5
Suits	53725,43
Sports Wear	47239,45
Dinner Crockery	43340,25
Car Media Players	39989,6
Beds	35752,6
Shoe Rack	34018,04
Umbrellas	33267,5
Apple Laptop	33025
Sofas	30738,48333
Dinning Tables	28867,65
Iron	26833,22
Bike Tyres	26803,2
Samsung Mobile	26568,6
Car & Bike Care	22700,84
Tablet	22311,04
Car Body Covers	21618,52
LED	20910,08
Car Mat	20836,8
Car Seat Covers	20006,9
Curtains	16101,83333
Fans	12045,5
Speakers	9909,2
LCD	6193,416667
Mixer/Juicer	5030,676667
Mouse	4629,89
Watch	3428,32
Keyboard	3290,1
Grand Total	3729902,95

5. MOST AND LEAST PERFORMING 3 PRODUCTS

It could be seen that most 3 performing products are T-shirts, Titak watches and running shoes; while least performing 3 products are Keyboard, Watch and Mouse with the values seen respectively.



TOP AND BOTTOM PERFORMING PRODUCTS

T - Shirts	407716,88
Titak watch	355244,12
Running Shoes	346338,88

Chart Title



■ T - Shirts ■ Titak watch ■ Running Shoes

Mouse	4629,89
Watch	3428,32
Keyboard	3290,1

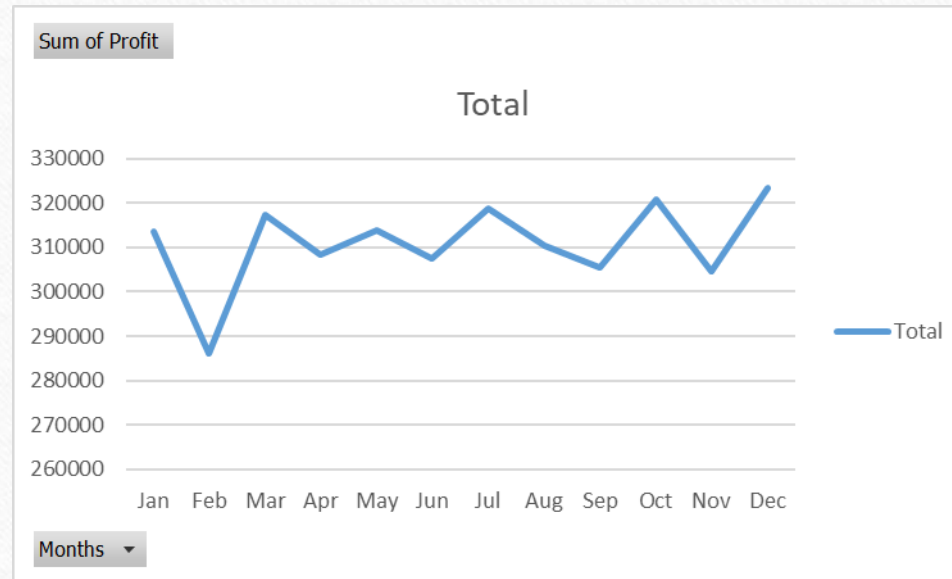
Chart Title



■ Mouse ■ Watch ■ Keyboard

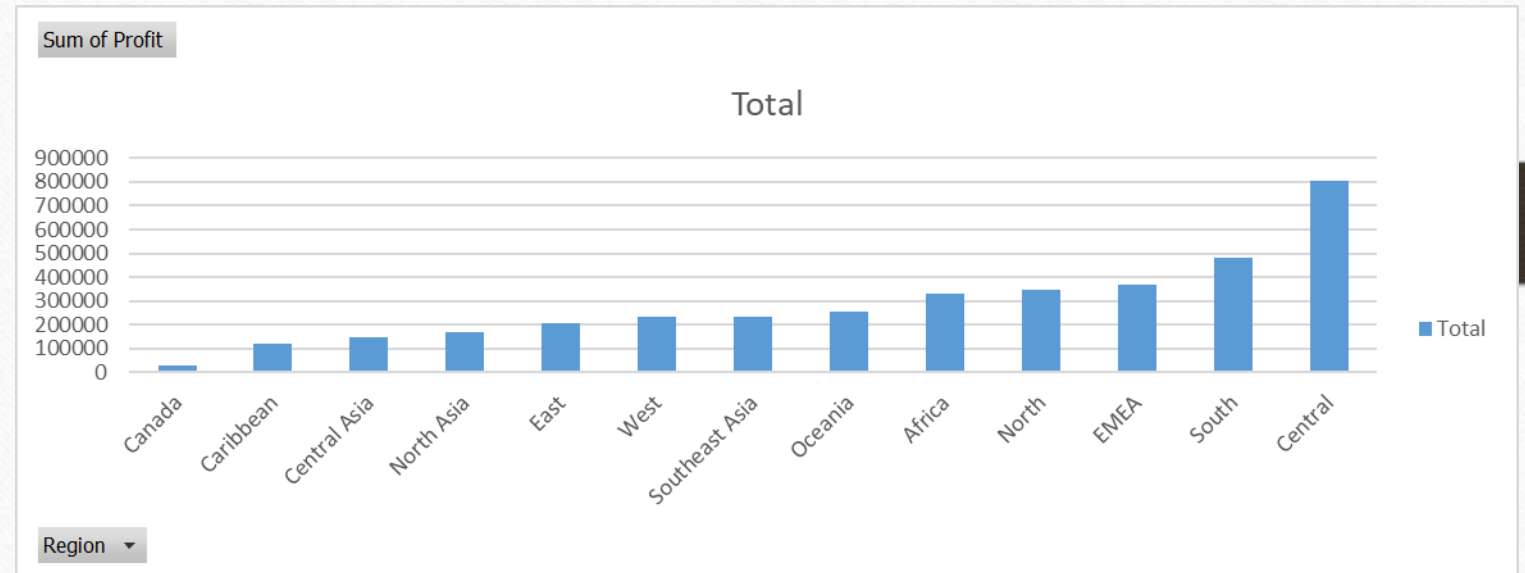
6.1 PROFIT BY MONTH

Row Labels	Sum of Profit
Feb	286102,6233
Nov	304716,1
Sep	305334,4567
Jun	307585,0233
Apr	308364,5133
Aug	310442,8433
Jan	313566,3467
May	313751,2467
Mar	317186,0067
Jul	318703,2
Oct	320748,67
Dec	323401,92
Grand Total	3729902,95



6.2 PROFIT BY REGION

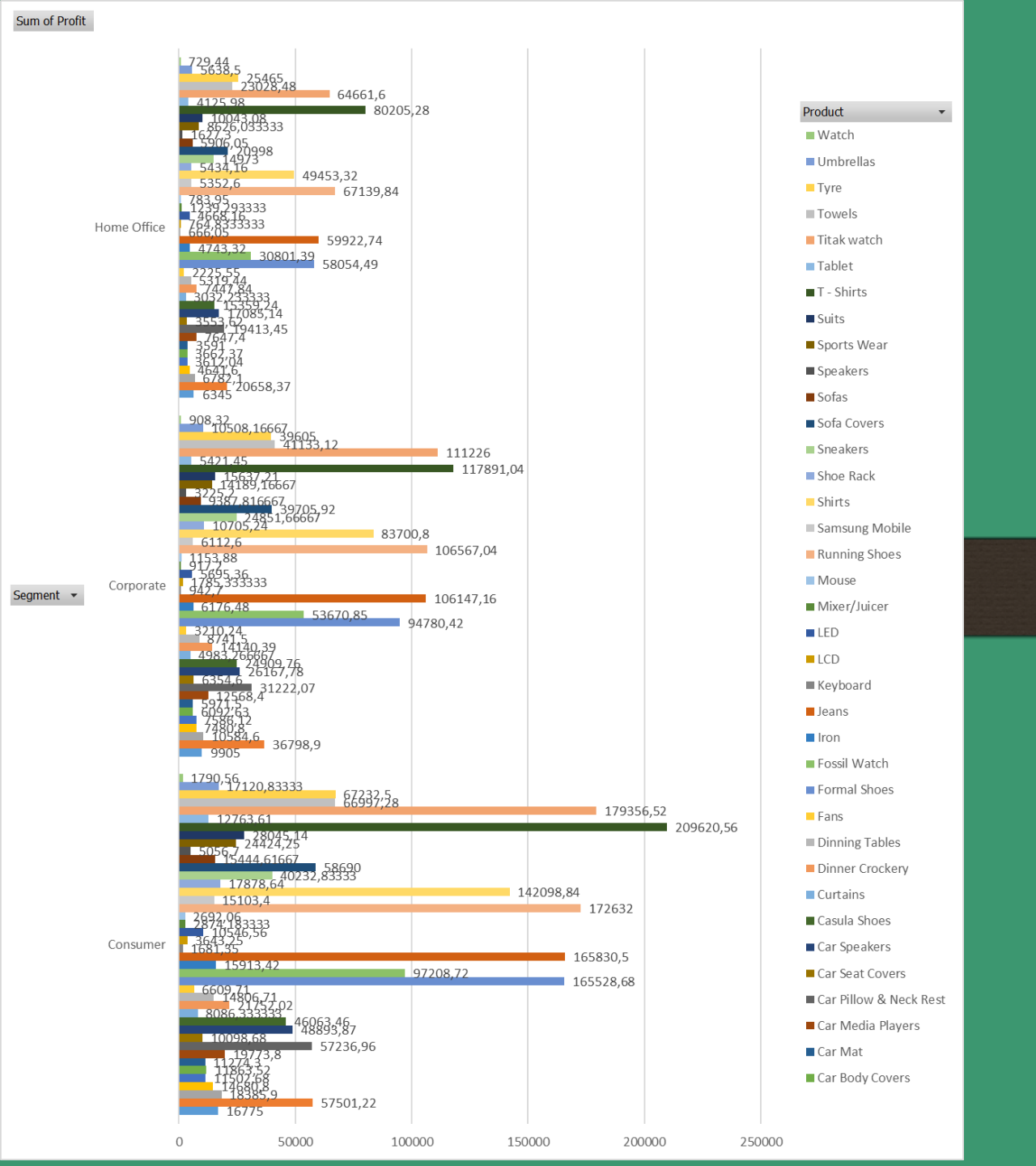
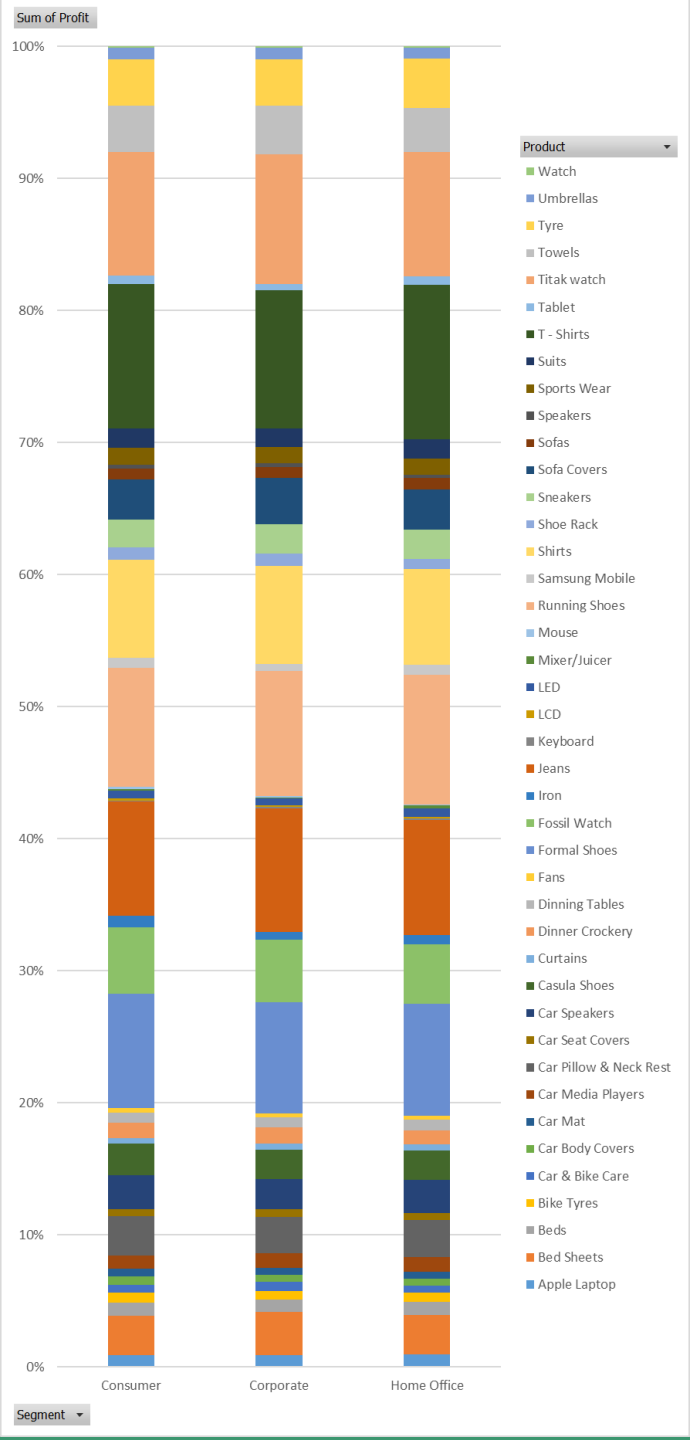
Row Labels	Sum of Profit
Canada	27782,30667
Caribbean	120602,6967
Central Asia	149526,0267
North Asia	170911,3567
East	206683,0267
West	231763,69
Southeast Asia	235587,8467
Oceania	253680,8133
Africa	332206,8267
North	349285,21
EMEA	368553,72
South	479146,82
Central	804172,61
Grand Total	3729902,95



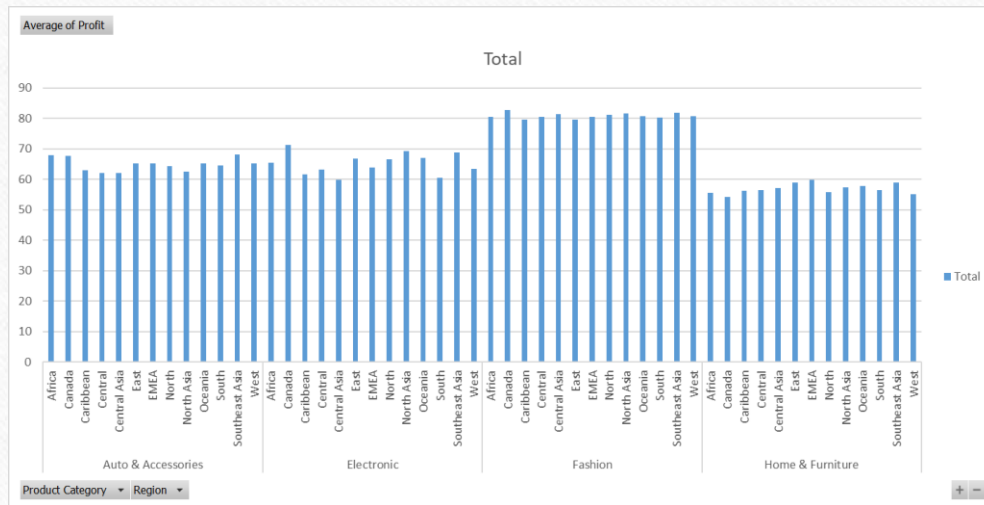
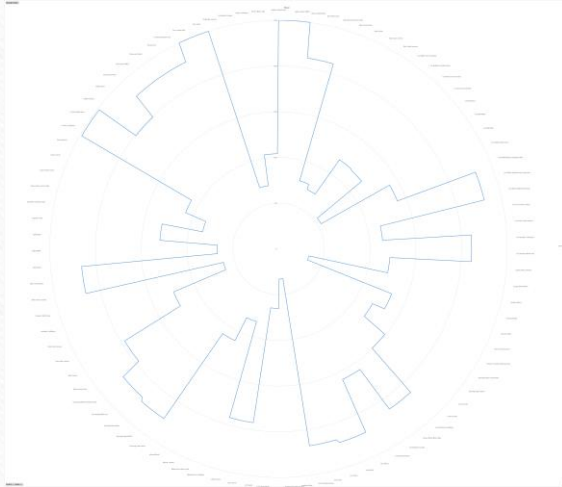
7. Statistical analysis

MODE		228		17
MEDIAN	\$	159,0	\$	67,0
SD		65,78197498		49,09966448
VARIANCE		4327,268233		2410,777052
MEAN	\$	156,4	\$	72,7
MIN	\$	33,0	\$	0,5
MAX	\$	250,0	\$	167,5

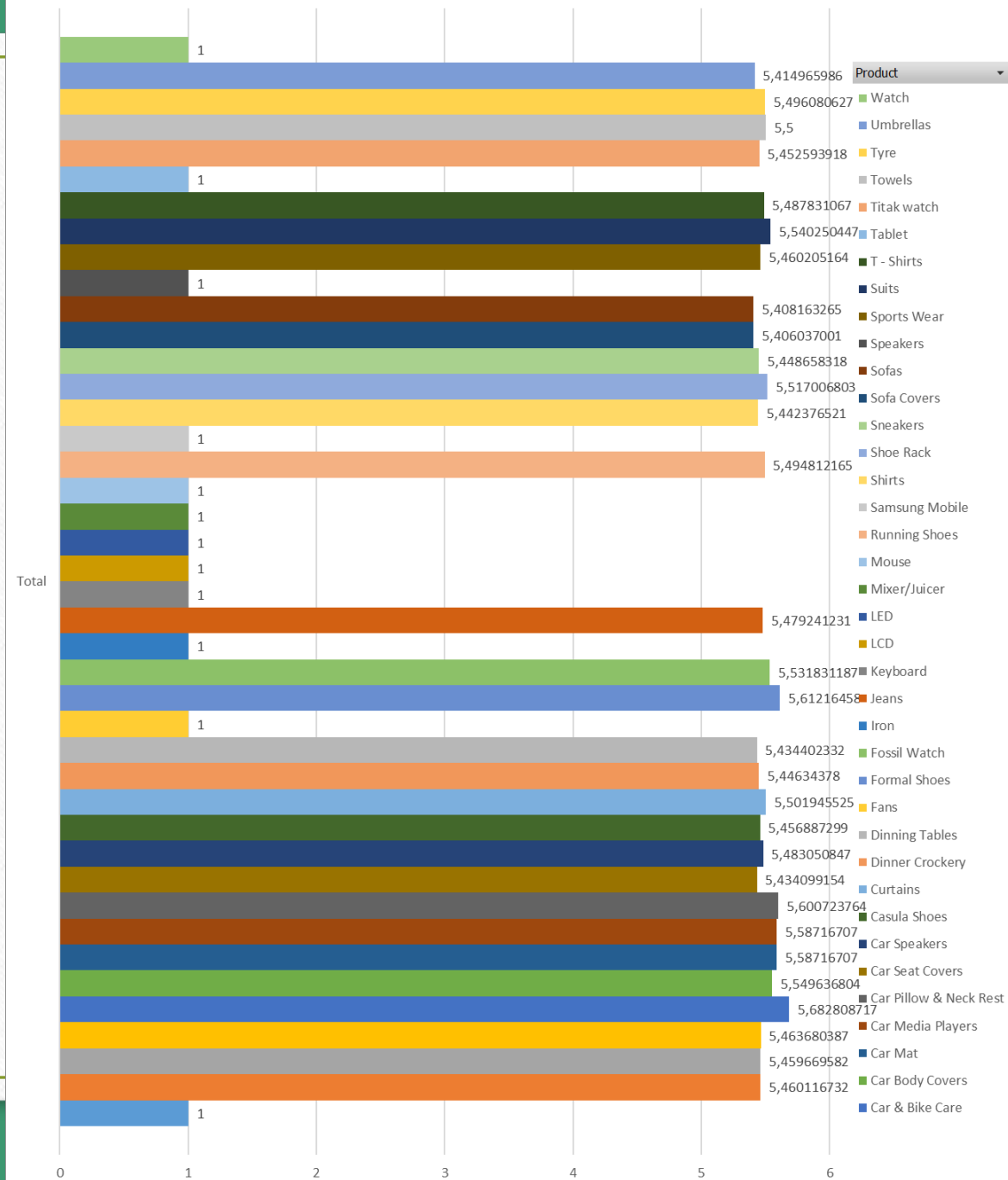
- Here we use the statistical tools like mode, median, standard deviation, variance, mean, min and max.
- The formulas used
 - = MODE()
 - = AVERAGE()
 - = STDEV.S()
 - = VAR.S()
 - = MEDIAN()
 - = MIN/MAX()



OTHER ANALYSIS

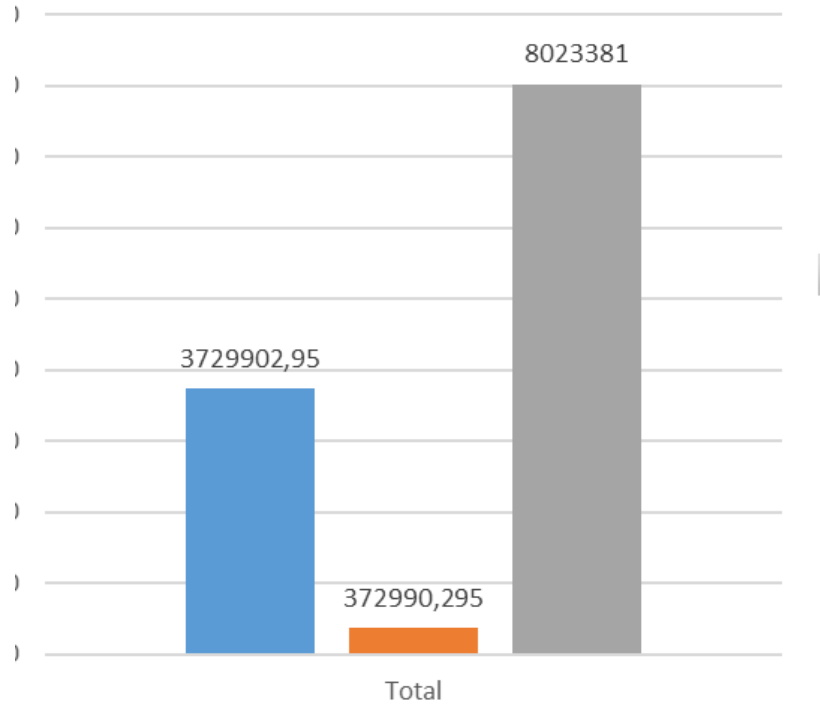


Average of Aging

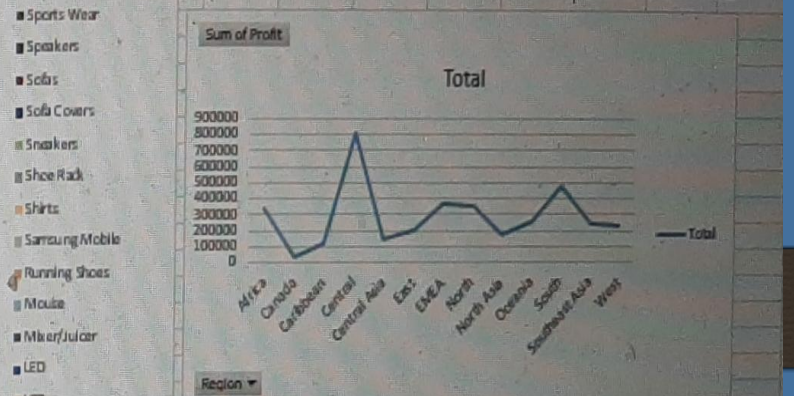
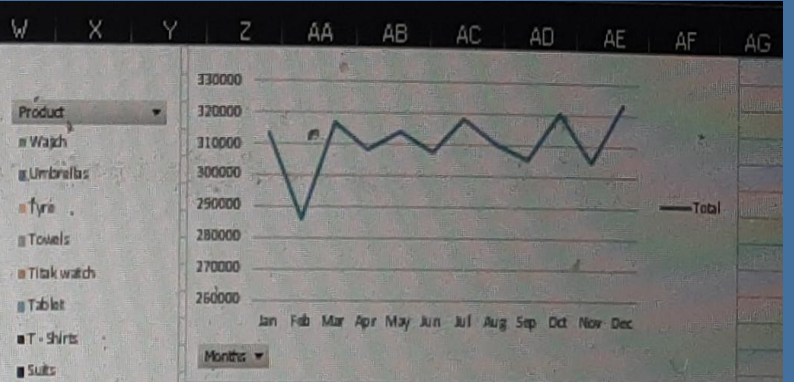
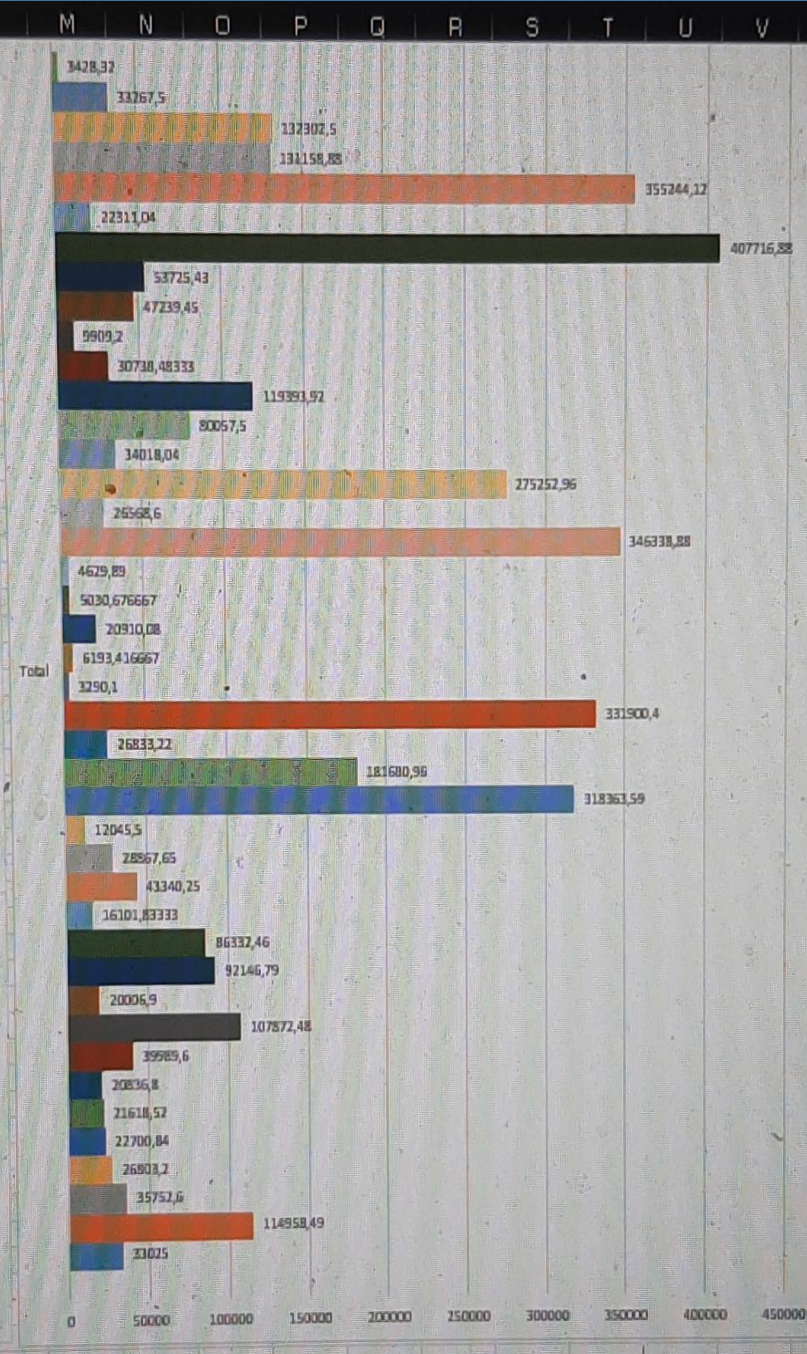
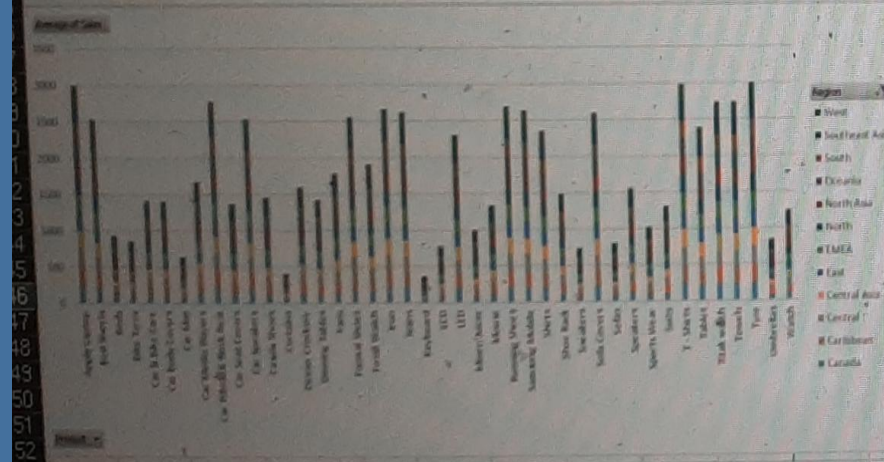
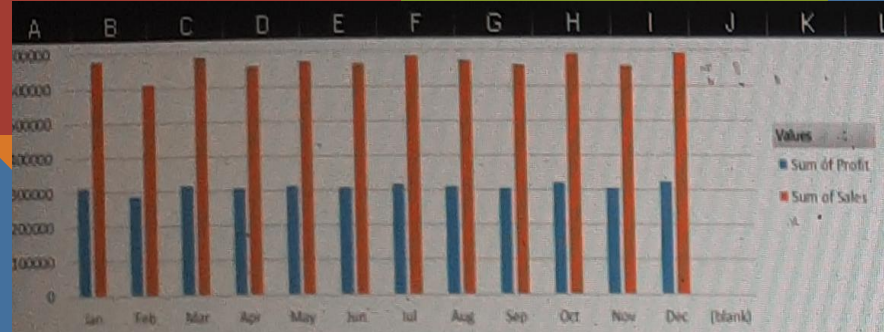


Recommendations and insights

Profit Sum of Shipping Cost Sum of Sales



- According to the given data set, it could be said that the performance of the company in North America, EMEA and other parts of the commercial areas of the world was nearly equally distributed.
- Inside the US, there is a profit peak in the Central, and there is a low position by February. So in order to make profits more balanced the company should lean on improving its performance also in west, north and south regions of the US and the also the regulating the performance in underperforming periods of the year, .
- Underperforming products should be analyzed and their profit rates should be improved,
- More seasonal, regional, product-wise analysis should be done.
- Logistic costs should be analysed
- Total sum of profits, shipping costs and sales could be seen on the graph, during the presentation we have seen the most and least performing products and regions too. So that we have done our seasonal, regional, product-wise analysis.



THANK YOU

Nazim Atakan Erdogan