

Data Analytics Internship

by IBM - Edunet

Contents

Name: Niharsh Choudhary

Mail_id: niharsh.2024csit1184@kiet.edu

College_Name: KIET Group of Institutions

CollegeState: Uttar Pradesh

Internship_Domain: Data Analytics

Internship_startdate: 12-06-2023

Internship_enddate: 24-07-2023

Analysis of Superstore

- Introduction
- Working with Dataset
- Creating Loss Dataframe
- Visualizations
- Overall Analysis



Introduction

In this dataset, we have many features like ship mode, Segment, country, City, State, Postal code, Region, category, sub-category, sales, Quantity, and discount, and the Dependent variable is profit. The remaining are independent variables.

Working with DataSet

Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit
Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	261.9600	2	0.00	41.9136
Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	731.9400	3	0.00	219.5820
Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	14.6200	2	0.00	6.8714
Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	957.5775	5	0.45	-383.0310
Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	22.3680	2	0.20	2.5164

Creating loss dataframe

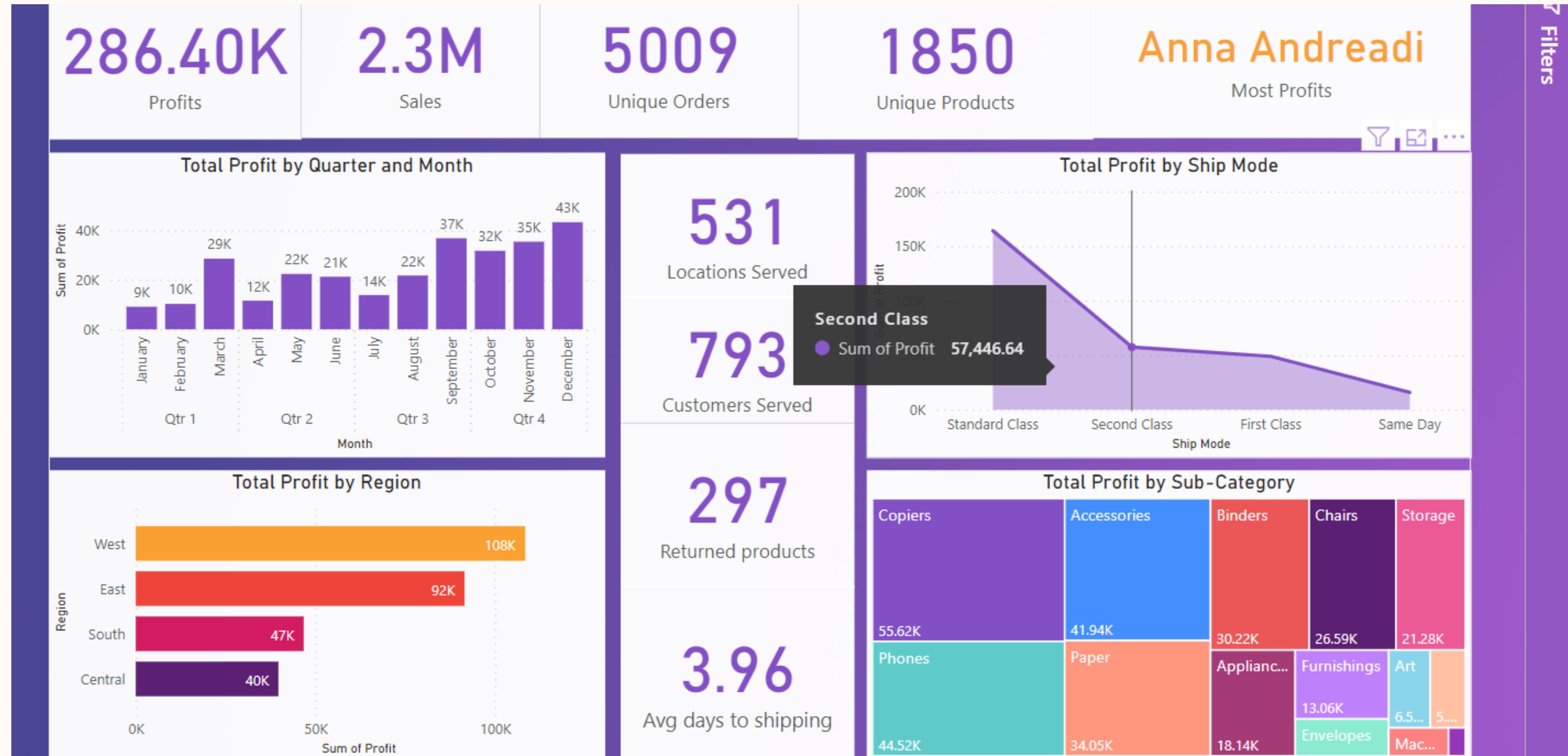
Now let's divide the overall data to draw some more accurate conclusions. Create a new dataframe, where profit is negative which means loss, and concentrate on these areas to improve.

loss_df=superstore_df[superstore_df['Profit'] < 0]

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	957.5775	5	0.45	-383.0310
14	Standard Class	Home Office	United States	Fort Worth	Texas	76106	Central	Office Supplies	Appliances	68.8100	5	0.80	-123.8580
15	Standard Class	Home Office	United States	Fort Worth	Texas	76106	Central	Office Supplies	Binders	2.5440	3	0.80	-3.8160
23	Second Class	Consumer	United States	Philadelphia	Pennsylvania	19140	East	Furniture	Chairs	71.3720	2	0.30	-1.0196
27	Standard Class	Consumer	United States	Philadelphia	Pennsylvania	19140	East	Furniture	Bookcases	3083.4300	7	0.50	-1665.0522
...
9920	Standard Class	Corporate	United States	Bryan	Texas	77803	Central	Office Supplies	Binders	22.3860	7	0.80	-35.8176
9921	Standard Class	Home Office	United States	Akron	Ohio	44312	East	Office Supplies	Binders	5.7420	3	0.70	-4.5936
9931	Standard Class	Consumer	United States	San Bernardino	California	92404	West	Furniture	Bookcases	683.3320	4	0.15	-40.1960
9937	Second Class	Corporate	United States	Los Angeles	California	90049	West	Furniture	Tables	71.0680	2	0.20	-1.7772
9962	First Class	Home Office	United States	Houston	Texas	77041	Central	Furniture	Bookcases	383.4656	4	0.32	-67.6704

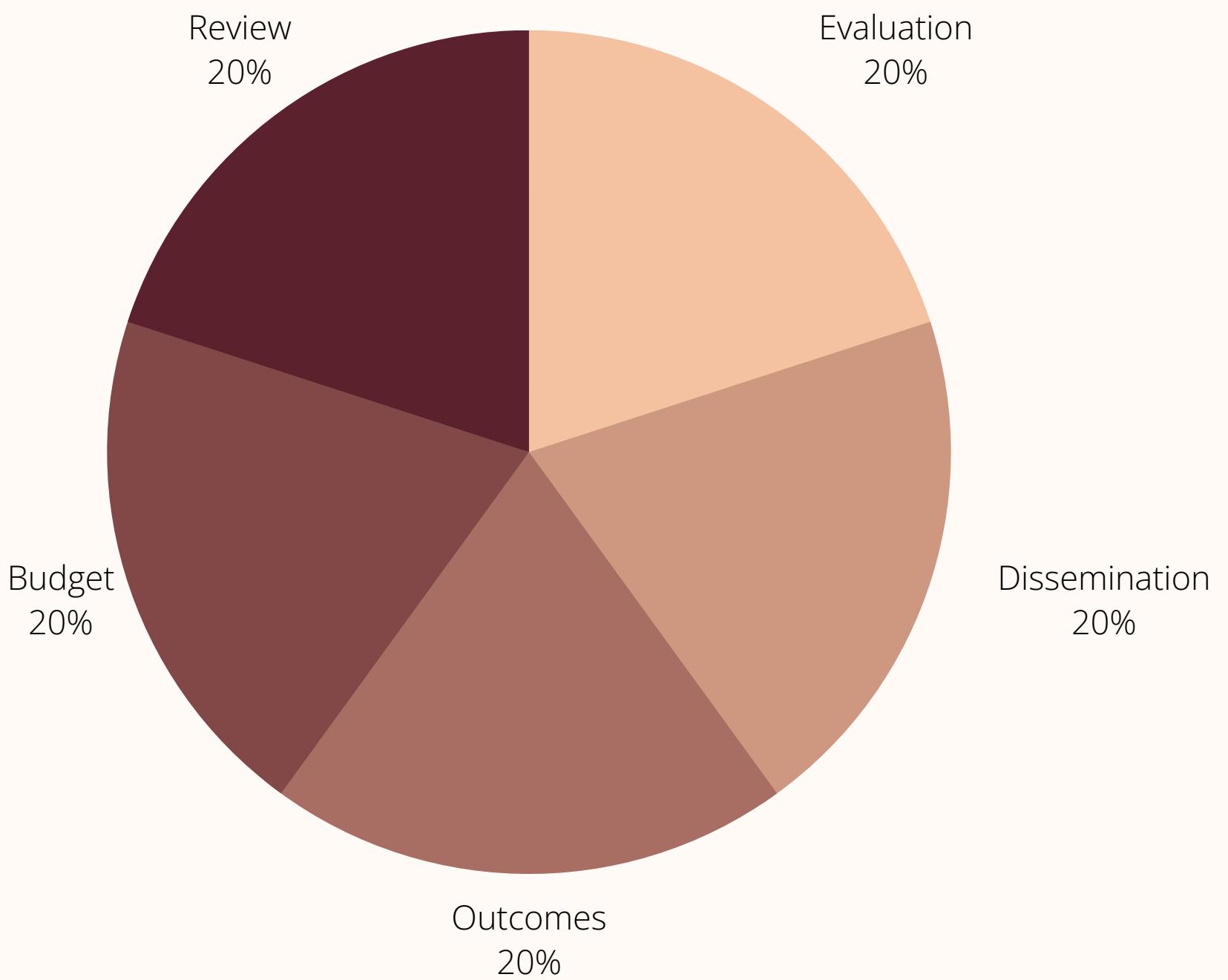
1871 rows × 13 columns

Visualization



Overall Analysis

- The main reason which leads to loss is Discount as if some areas lead to loss due to more discounts, and some areas lead to fewer sales due to fewer discounts, hence it needs to be improved.
- It is better to give more discounts during festival seasons, additionally, that will result in more sales.
- The Home office segment needs better improvement.
- Some cities have fewer sales, and lack of awareness can be the reason for this, hence advertising in those cities might help in more sales.



Reference Links

Project link:

https://github.com/Niharsh/Mini_project



thank
you