

Exploratory Data Analysis Report

Compiled by Zindazed Abdshahd Kasauli, Student of Masters Program in Computer Science at Makerere University as part of an Assignment in the Machine Learning Course offered in the masters program

Introduction:

The objective of this assignment was to perform an exploratory Data analysis on a dataset. I chose to use a dataset for one of our clients at work which contains transaction records for their business. The aim was to explore where their strengths are so that they can capitalize on them and decide how to increase their sales. Some key questions were drafted to help guide this exploration.

Exploratory Data Analysis Questions:

1. What products are most selling?
 - a. The Rationale for this question was to know what products to produce more of. In other words, once the business gets to know which of its products are mostly bought, then they can plan to ensure that such products are readily available.
2. What are the best performing regions?
 - a. The Rationale for this second question was to know what regions to send more products to. Through knowing the regions where a lot of sales happen, then you can work to ensure that such regions are well served given that they are where your key customers are.
3. What are the best performing Channels?
 - a. The Rationale: to know which channel to send more products through. Channels are how the sales get to happen, some sales happen through sales employees, which is the distribution channel. Others happen through contracts the business secures with institutions, which is the institution channel. So, knowing the channels that are best performing can also help the business know where to invest more.
4. What are the best-performing Sales Employees?
 - a. Rationale: To recognise the high performing people and reward them as a means of motivating the rest. It would be using a single stone to hit two birds. Since if the best performing sales employees are recognised and even rewarded for their efforts. They are motivated to make more sales, while this also motivates the under performing ones to work harder in order to get such treatment as their counter-parts.
5. What are the best performing BP?
 - a. Rationale: To do the best possible to keep the top Customers. Such customers could even be prioritized for example since they are the ones that buy most from the business. Strategies of how to retain them can be devised.

Data Wrangling:

The dataset used was gotten from our client at work. The dataset was found to not be clean. So, it had to first be cleaned.

First, some fields were found have datatypes different from the expected ones, for instance the Transaction Numbers fields were expected to be strings but they were integers. So, a conversion was done. This was because transaction numbers were not to be used in any arithmetic computations and

even if other characters were to be included, it wouldn't harm as their main purpose was to uniquely identify a transaction.

Next, Leading and trailing white spaces check was performed though none were found. Null and blank fields were replaced with 'Not specified' string for the string fields.

A total of 27 rows of duplicate values were found and removed. The fields to be used in explorations were checked to ensure that their values were feasible to be used in the analysis. Only '0' found in the product category was converted to Not specified since a category called zero just made no sense

A cleaned dataset was exported as an excel file so that it can serve as the basis of analysis for future investigations.

Exploratory Data Analysis:

The exploratory data analysis process was guided by the 5 formulated questions. Bar graphs and some tabular analysis were utilized to explore this data.

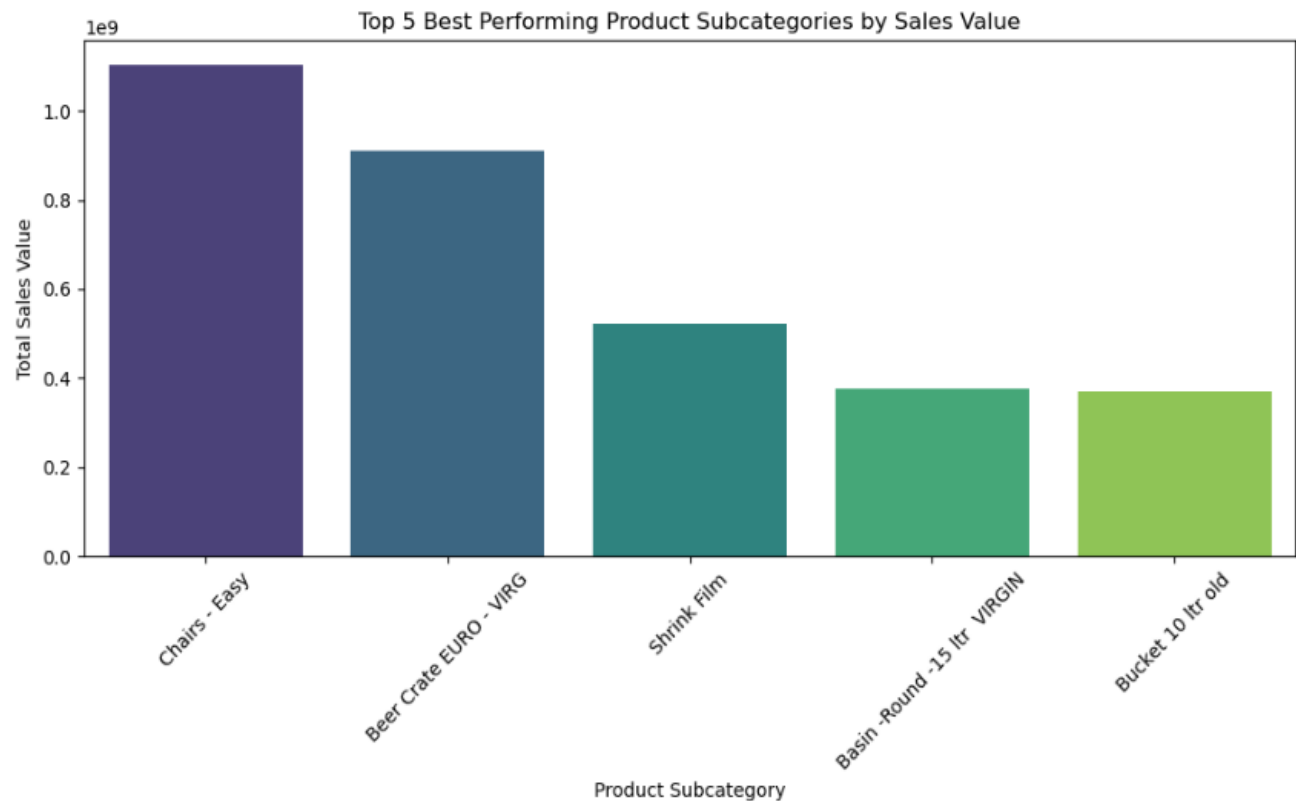
1. What products are most selling?

Item Description	Item Subcategory	Product Category	Sales Value
Beer Crate EURO - VIRG	Beer Crate EURO - VIRG	Packaging Crates	9.116641e+08
Packaging film -Shrink Film -340MM (Kgs)	Shrink Film	Packaging -Film	5.217466e+08
Chairs - Easy - White	Chairs - Easy	Furniture	5.116637e+08
Tooth brush-Adult (Classic Cardboard) - 24pkts...	Classic Toothbrushes	Dental	3.004925e+08
Chairs - Easy - Clip cap blue	Chairs - Easy	Furniture	2.938464e+08
Crate - 300ml VIRGIN - Red	Crate - 300ml VIRGIN - Coke	Packaging Crates	2.801845e+08
UHT Milk 500ML TETRA FINO	UHT Milk	Trading Items	2.131967e+08
Jerry Cans-20 ltr 750gms - Yellow	Jerry Cans-20 ltr 75gms	Packaging J/Cans	1.591797e+08
Packaging film - milk - 500ml	Milk Film	Not specified	1.356318e+08
Bucket 10 ltr - Green	Bucket 10 ltr old	Buckets	1.152324e+08
Beer Crate POLISH - VIRGIN	Beer Crate POLISH - VIRGIN	Packaging Crates	1.079345e+08
Watering Can	Watering Can	Agriculture	9.685043e+07
Basin -Round -15 ltr VIRGIN - Blue	Basin -Round -15 ltr VIRGIN	Basins	9.440018e+07
Chairs - Easy - Gold	Chairs - Easy	Furniture	9.036160e+07
Chairs - Legend - Clipcap Blue	Chairs - Legend	Furniture	8.273184e+07
Basin -Round -15 ltr VIRGIN - Green	Basin -Round -15 ltr VIRGIN	Basins	8.184436e+07
Bucket 20 ltr -B/White - IML - Weather guard (...)	Bucket 20 ltr	Packaging Buckets	7.912321e+07
Bucket 10 ltr - Blue	Bucket 10 ltr old	Buckets	6.764264e+07
Chairs - Easy - Grey	Chairs - Easy	Furniture	6.722711e+07
Packaging film - milk - 500ml	Milk Film	Packaging -Film	6.438808e+07

From the above tabular representation, it can be seen that Beer Crate Euro – Virg is the top most selling product. It belongs to the Beer Crate Euro – Virg sub category and the Packaging Crates Category.

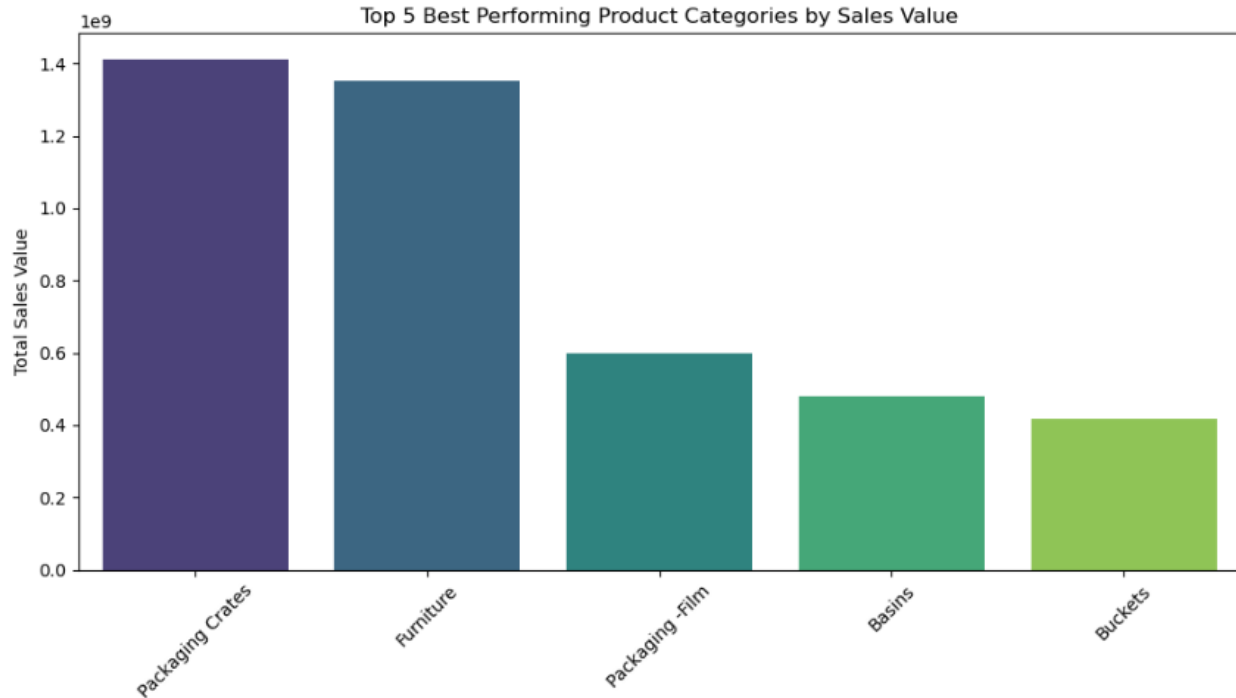
Item Description is the product itself.

Also subcategory performance was analysed and this is what was found:



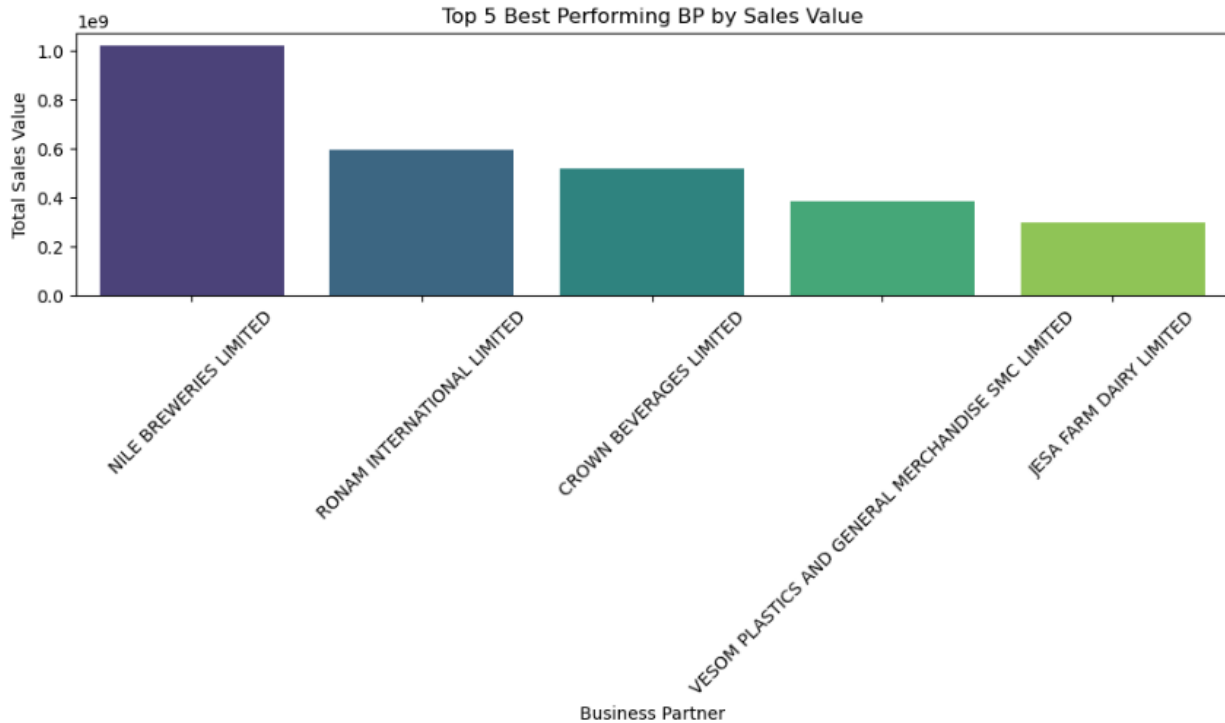
Chairs-Easy subcategory tops the list. Which is no surprise since in the top 20 products table, 4 of its products are well represented which makes it the most represented subcategory in the top 20 products table. Easy chair white is in the 3rd place, Easy chair clip cap blue in the 5th. Then Easy chair Gold in the 14th and finally chair easy grey in the 19th

Product category Analysis yielded these results:

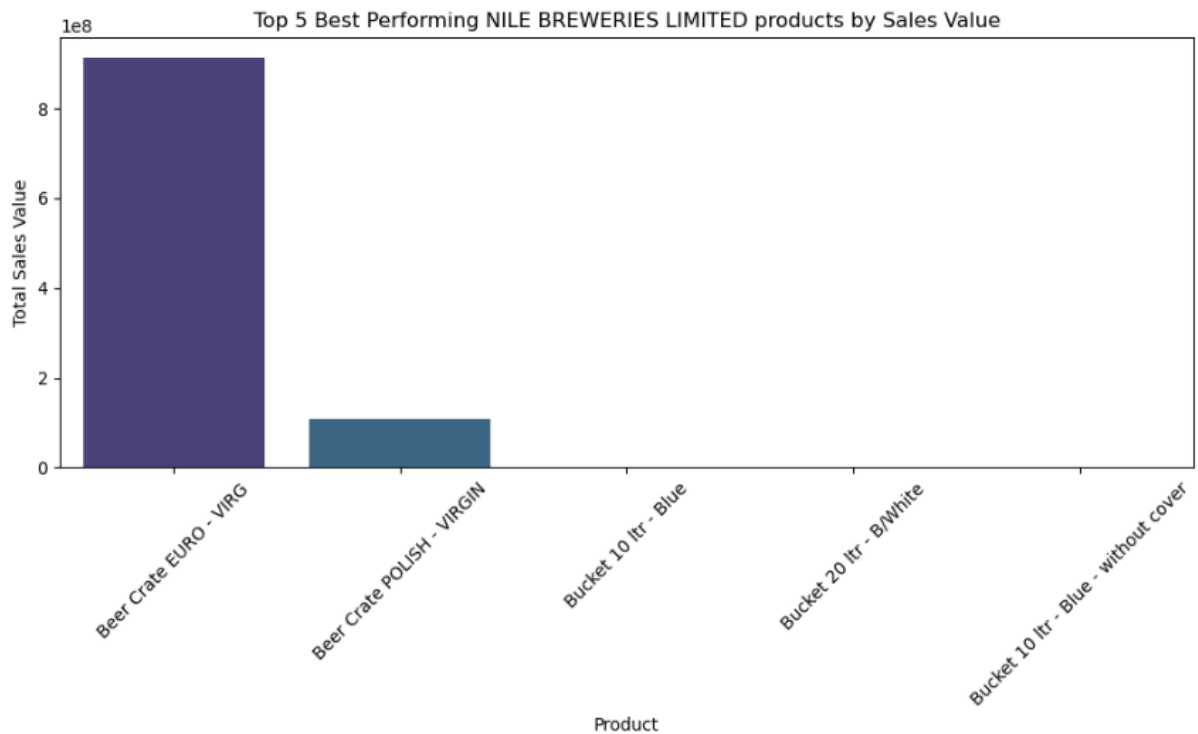


When it comes to Product Categories, the category of the top selling product tops the list. Explains a lot since as it can be seen from the top 20 table, the Beer Crate Euro – Virg leads by a mile and much as the other packaging crates in the top 20 products are from different subcategories, they are also selling well. Making Packaging crates to top the category list. We can suspect a Beverage Customer appear among the top buying Business partners.

For that reason, let's see what the Business Partners analysis results which will also answer the question; What are the best performing BP?:



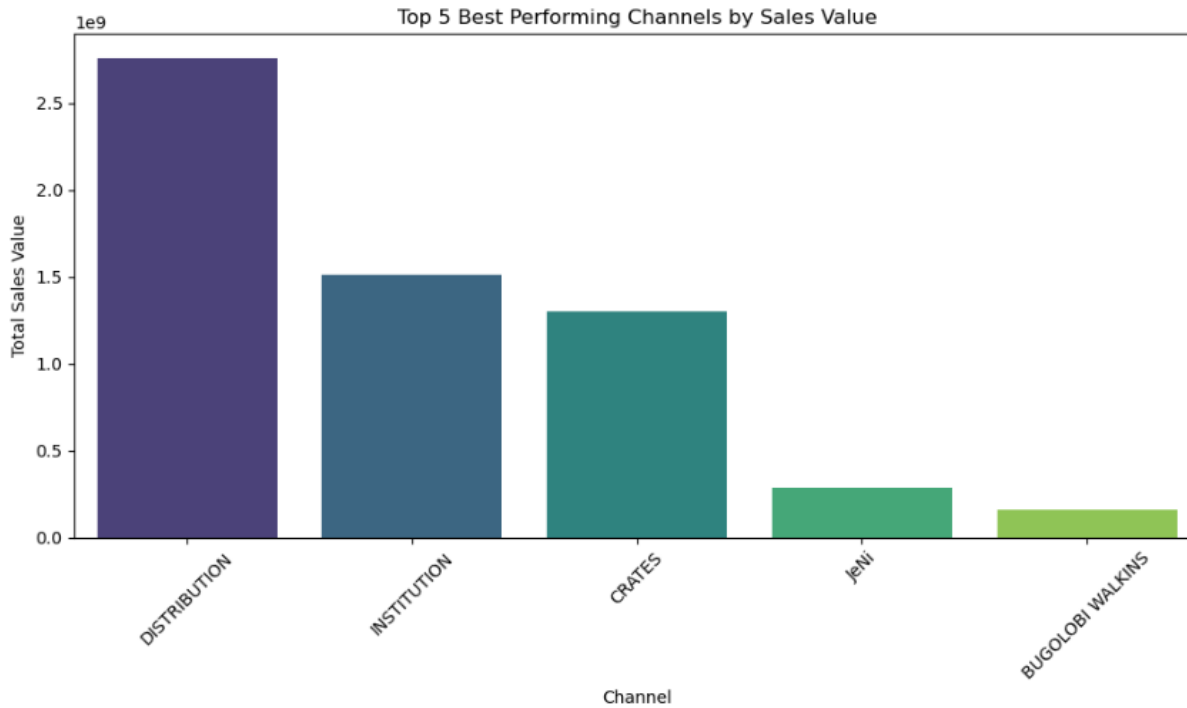
As expected, Nile Breweries limited, a Beverage company tops the list. And we also have Crown Beverages limited coming in at position 3. Upon close analysis, it become apparent the Nile Breweries isn't just the top buying BP but could also be the one buying most of the best-selling product. Making it a very crucial customer. This can be illustrated in the graph below:



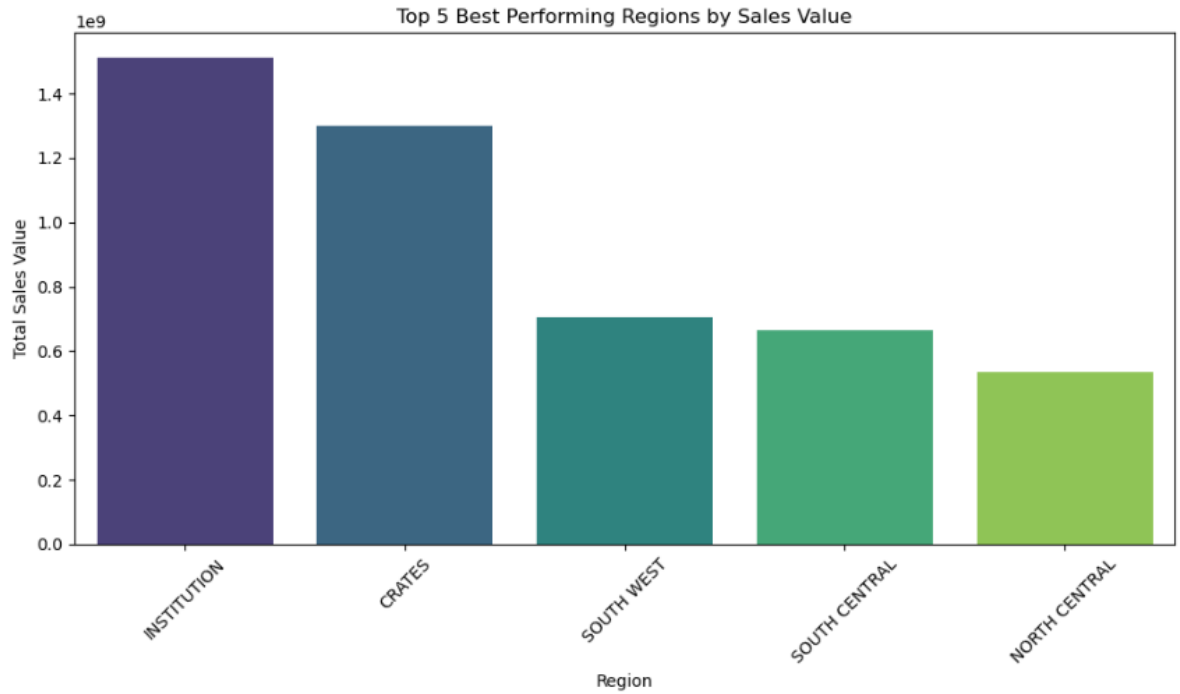
It also buys Beer crate Polish – virgin which is the 3rd most selling packaging crates products and also comes in the 11th place in the overall top selling products rankings as can be seen in the top 20 selling products table.

The other metrics that were also analysed together with their results are listed below:

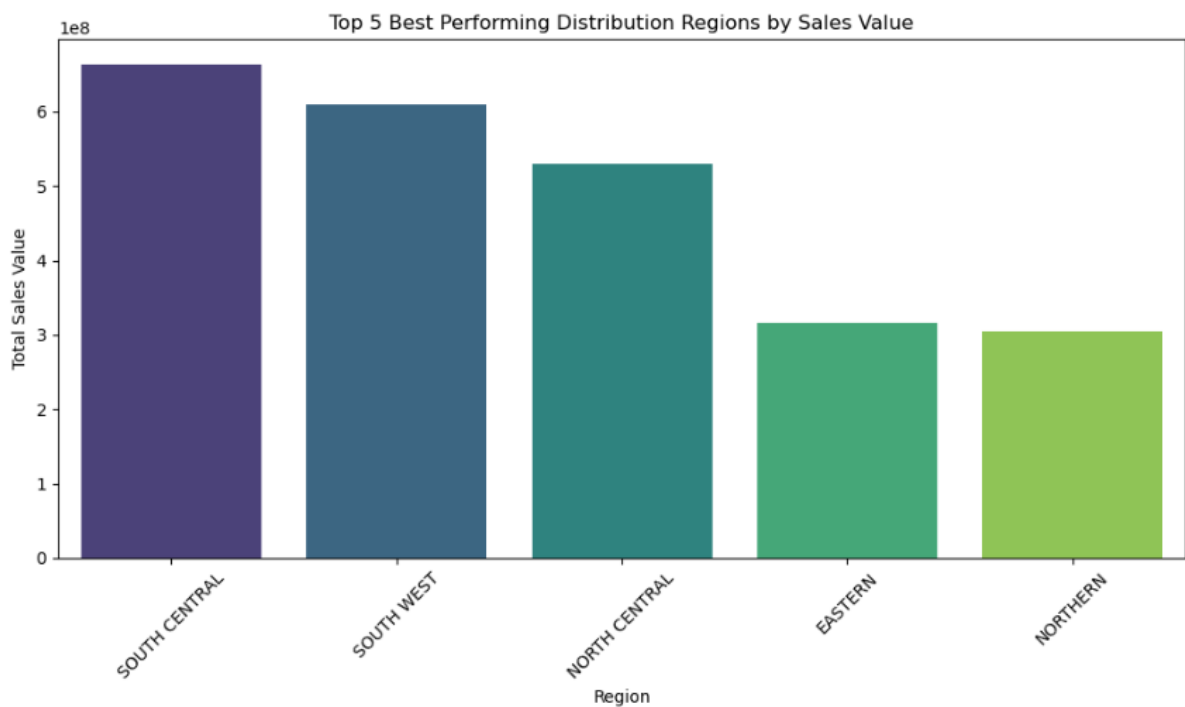
What are the best performing Channels?



1. What are the best performing regions?



We also analysed the best performing regions from the best performing channel. The results were as below:



Then lastly

What are the best-performing Sales Employees?



Given such information, Our client can clearly see where their strengths are and then can use these findings to capitalize on them so that they can enhance their profits as a consequence.