A picture containing text

Description automatically generated

|  |  |
| --- | --- |
| Module code and name: | Business Intelligence Systems 5COSC016C-n |
| Assessment: | CW 2 |
| Student ID: | 00012938 |
| Submission deadline: | 26.03.2024 |
| Wordcount | 1160 |
| Results date and type of feedback |  |
| Result (Pass or Fail) |  |
| **Feedback** | |
|  | |

Table of Contents

[Introduction 2](#_Toc162310711)

[Problem Overview 3](#_Toc162310712)

[Dimensions and measures used for OLAP Cube creation 4](#_Toc162310713)

[Dimensions 4](#_Toc162310714)

[Data Cubes 5](#_Toc162310715)

[SalesOverTime 5](#_Toc162310716)

[SalesBySegment 6](#_Toc162310717)

[SalesByLocation 7](#_Toc162310718)

[SalesByCategory 8](#_Toc162310719)

[Limitations 9](#_Toc162310720)

[Data Mining Results 10](#_Toc162310721)

[Research Findings 11](#_Toc162310722)

[Sales Variability 11](#_Toc162310723)

[Sales among customer segments 12](#_Toc162310724)

[Sales By State 13](#_Toc162310725)

[Sales By Subcategory 14](#_Toc162310726)

[Recommendations 14](#_Toc162310727)

[Conclusion 14](#_Toc162310728)

# Introduction

The following report continues the exploration of one e-commerce business based on the available [dataset](https://www.kaggle.com/datasets/mmohaiminulislam/supershop/data). If the first coursework focused on creating a data warehouse with ETL, this coursework will prepare an analytical service for finding meaningful and useful information. Regarding the data warehouse:



# Problem Overview

Even though there are only 3 dimensions, they provide good indicators and a variety of data concerning different questions. Based on them, 4 main business questions were stated are interesting in terms of future business management and development:

1. Geographical Impact Concern

* Question: What specific regions/states/cities gave to business the biggest total sales? revenue
* Dimensions Effected: CustomerDim(Region, State, City) and SalesFact(total\_cost).
* Visualization Type[[1]](#footnote-1): A bar chart will be suitable because it clearly visualizes the total sales for each region/state/city and comparing its values with finding the top ones is easy for the human eye.

1. Effect of the category on profitability

* Question: Which product categories/products give the highest profit margins?
* Dimensions Effected: ProductDim(Category, ProductName), SalesFact(total\_cost, items\_number)
* Visualization Type: A Treemap is suitable because a large share will be clear to see.

1. Does the time matter

* Question: What is a general trend in business development regarding sales revenue?
* Dimensions Effected: DateTime(OrderYear, OrderMonth), SalesFact(total\_cost)
* Visualization Type: A line graph is ideal because the x-axis will be a time variable.

1. Impact of the customer segment

* Question: What are the top customer segments regarding sales revenue?
* Dimensions Effected: CustomerDim(Segment), SalesFact(total\_cost)
* Visualization Type: Stacked and clustered charts are the best variant for visualizing a distribution among segments(parts) taking the third constraint.

# Dimensions and measures used for OLAP Cube creation

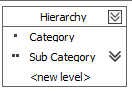
## Dimensions

CustomerSegmentDim: contains Segment Attribute

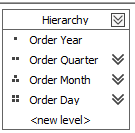
Geographical Dim: contains Country, Region, State, City attributes and respective hierarchy



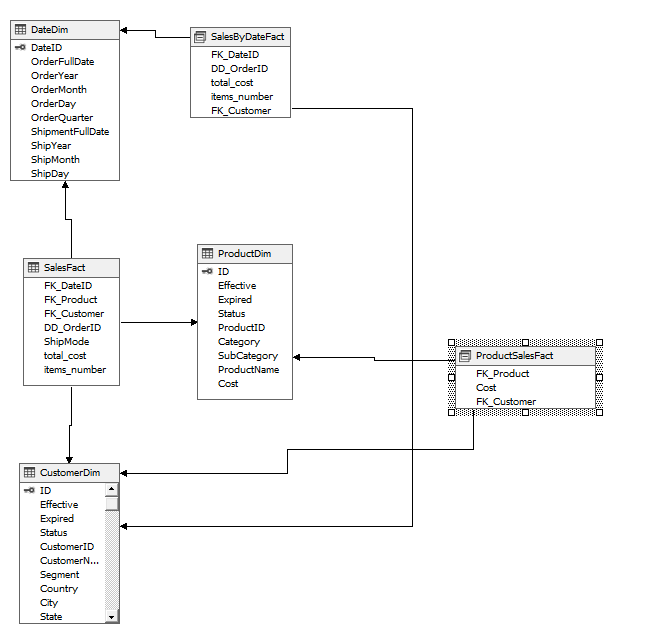
Category Dim: contains SubCategory, Category attributes and respective hierarchy



Date Dim: contains Order Day, Order Moth, Order Quarter, Order Year and respective hierarchy



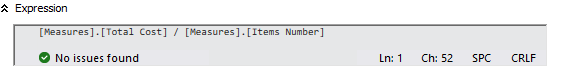
Note: SalesFact from the original data warehouse is not enough and may be not application for constructing all data cubes. Some of them needed aggregation. That’s why 2 additional ones were created. They are not adding a new value, but is used for aggregation of the main fact table:



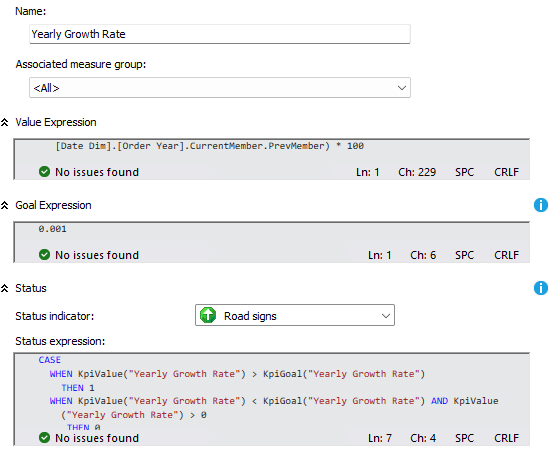
## Data Cubes

### SalesOverTime

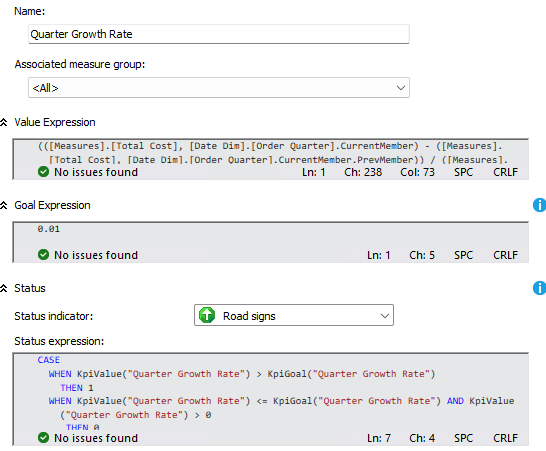
* Measures: Total Cost, Items Number, Sales Count
* Dimensions: Date Dim
* Calculations: AverageCostPerItem



* KPIs:
  + Yearly Growth Rate

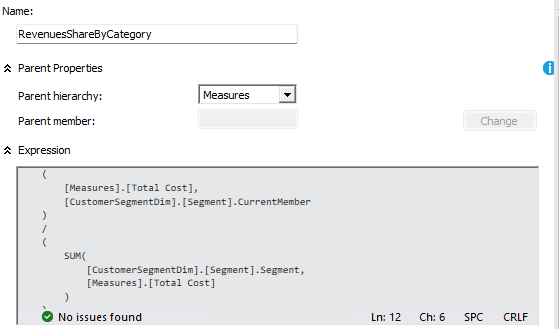


* + Quarter Growth Rate

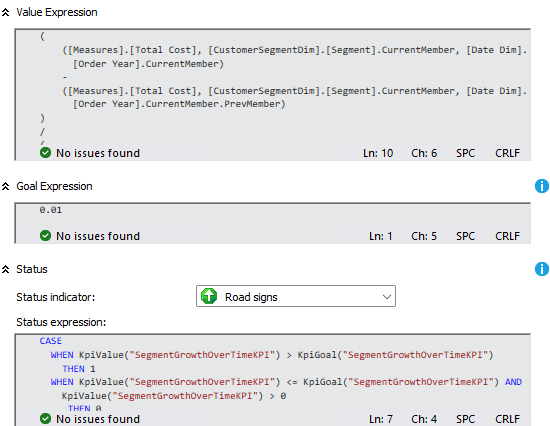


### SalesBySegment

* Measures: Total Cost, Items Number
* Dimensions: Date Dim, CustomerSegmentDim
* Calculations: RevenuesShareByCategory

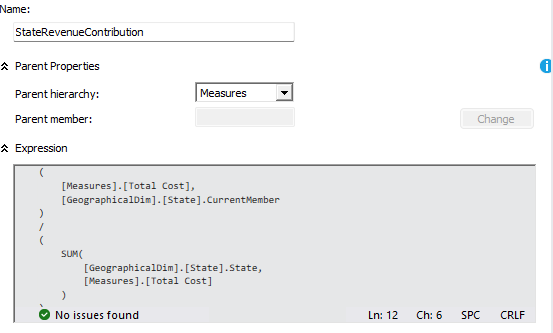


* KPI: SegmentGrowthOverTimeKPI

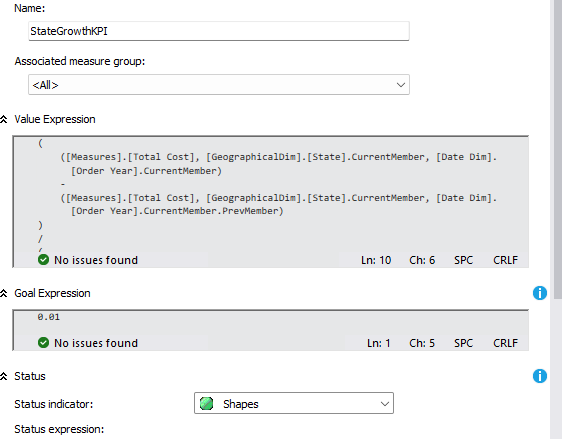


### SalesByLocation

* Measures: Total Cost, Items Number
* Dimensions: Date Dim, CustomerSegmentDim, GeographicalDim
* Calculations: StateRevenueContribution

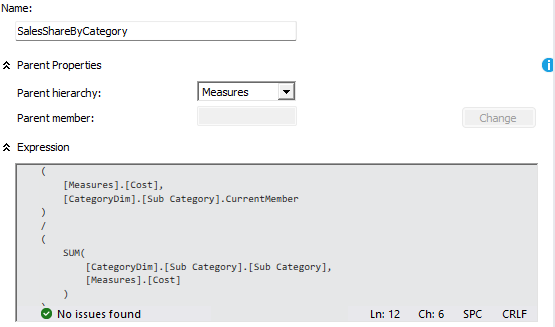


* KPI: StateGrowthKPI

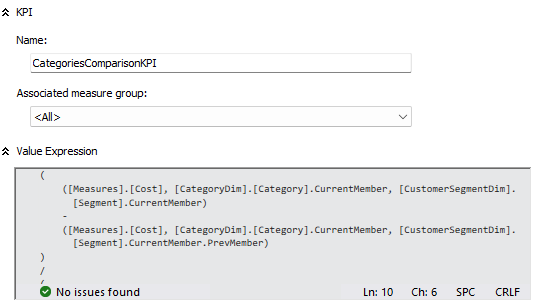


### SalesByCategory

* Measures: Cost, Product Sales Fact Count
* Dimensions: CategoryDim, CustomerSegmentDim
* Calculations: SalesShareByCategory



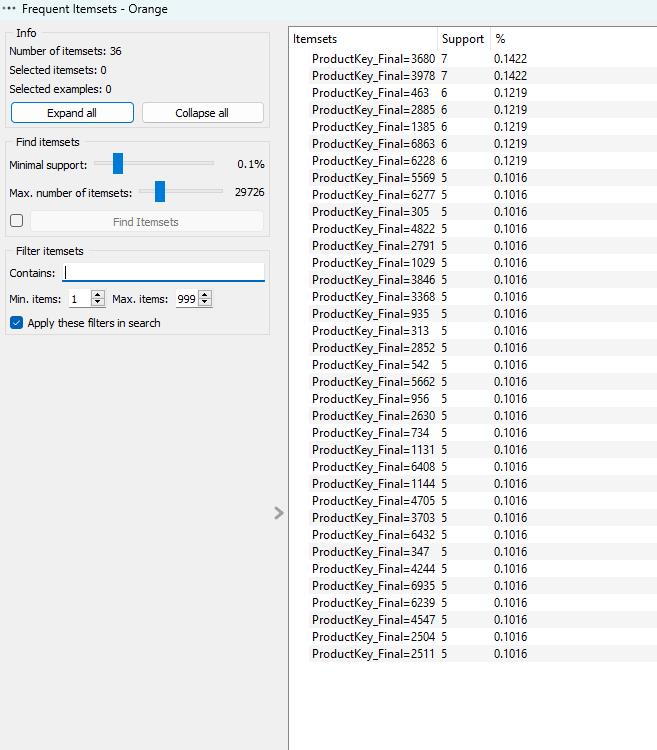
* KPI: CategoriesComparisonKPI



### Limitations

* Product Details: the dataset gives a limited description of the specifications. This information could be valuable for deep product analysis, especially when focusing on revenue constraint
* Customer Details: if there was a demographic data, it would be possible to identify some purchasing patterns. Currently, this is limited to the segment of the customer
* Purchase Details: data including if the discount was applied or what payment method used are not available, but could be quite useful for sales analysis

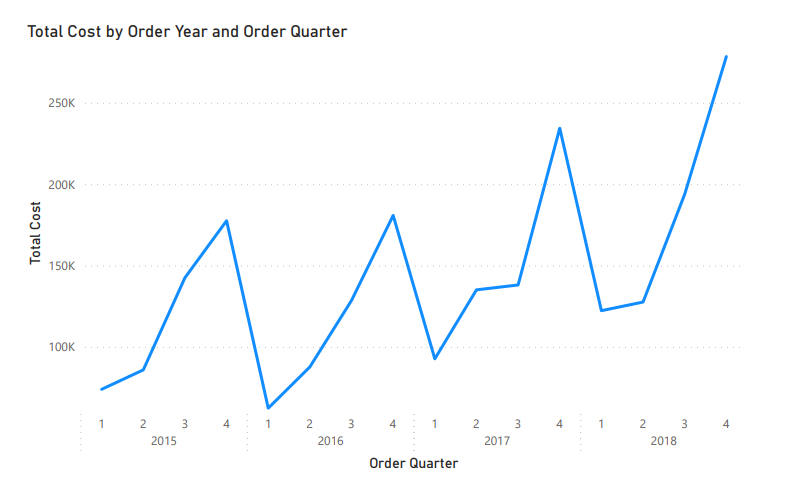
# Data Mining Results



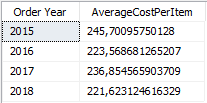
* Out of over item sets consisting of over 1200 products, 36 ones were identified that fulfilled the minimum requirement
* A minimum support threshold was set to 0.1% that is a relatively low
* Product 3680 (GBC DocuBind 300 Electric Binding Machine) and product 3978 (Avery Binding System Hidden Tab Executive Style Index Sets) are supposed to be purchased together more frequently with a support of 0.14%

# Research Findings

## Sales Variability

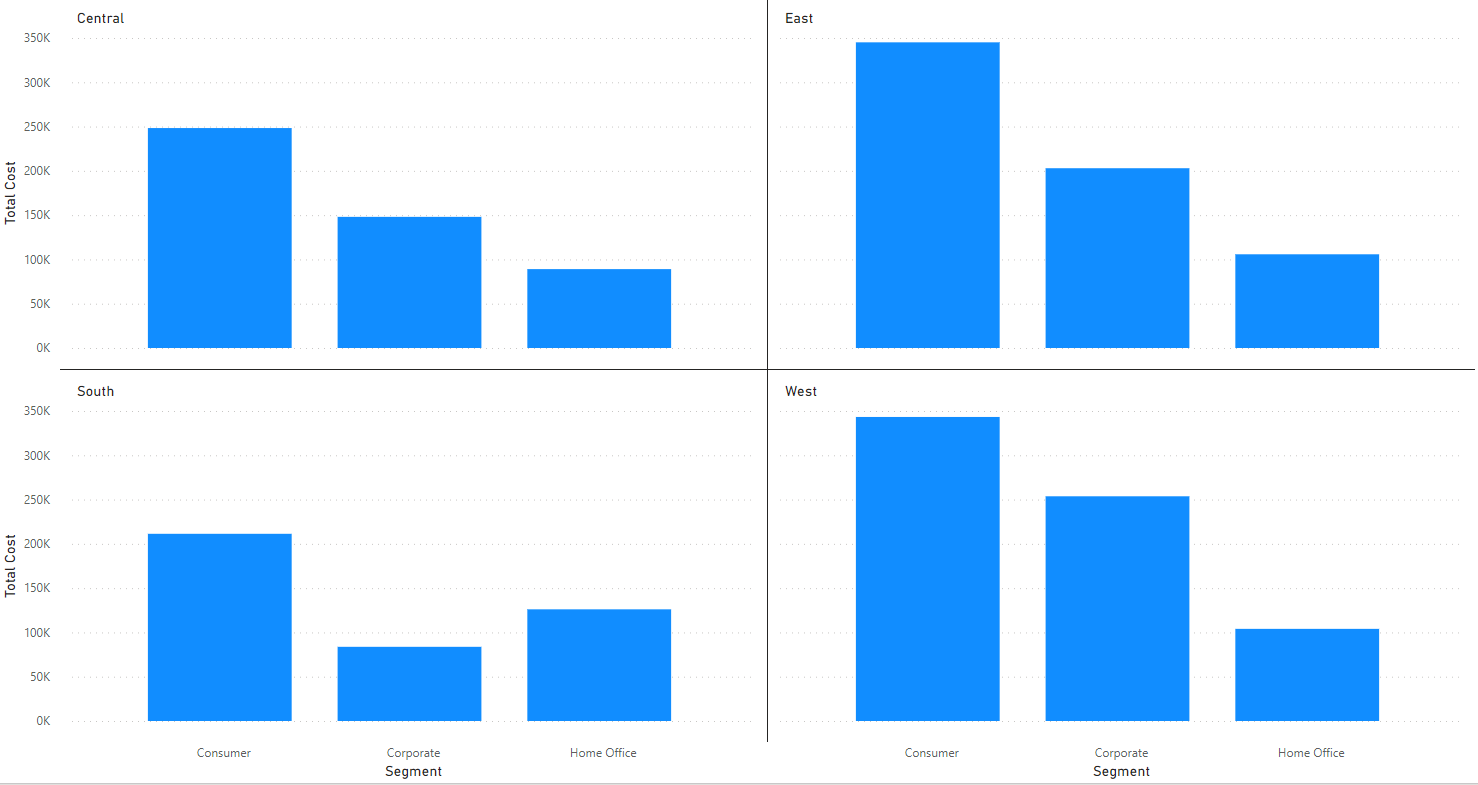


There is a certain pattern in the sales per quarter that repeats systematically. Each year, the sales reach its peak in the fourth quarter(Q4). First quarter is always the worst in terms of sales obtained. One of the possible reasons may be the popularity of purchasing presents online for holidays and quarter 4 includes one of the main holidays which is a Christmas. The first Quarter, on the contrary, contains a few big holidays. Regarding the yearly sales trend, there was a decline in 2016 followed by sharp growth in the next 2 years resulting in a multiplication of revenues.

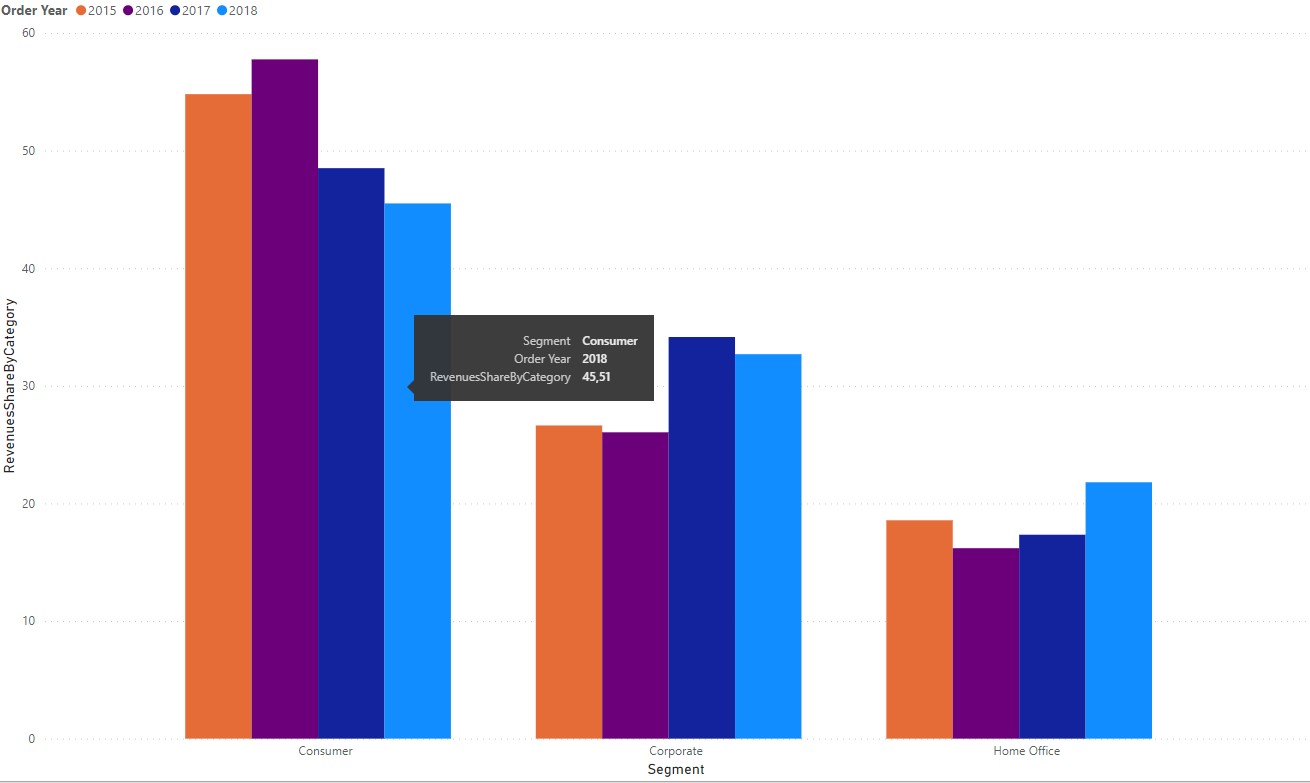


Moreover, there is an interesting result when focusing on the average item cost. It is seen that the average cost per item slowly declines over time. Usually, it is common that the average item cost in the first years of business operation is relatively low and increases over time. This may be a signal of a shift towards selling lower-cost items or a reduction of margin cost including taxes/transportation.

## Sales among customer segments

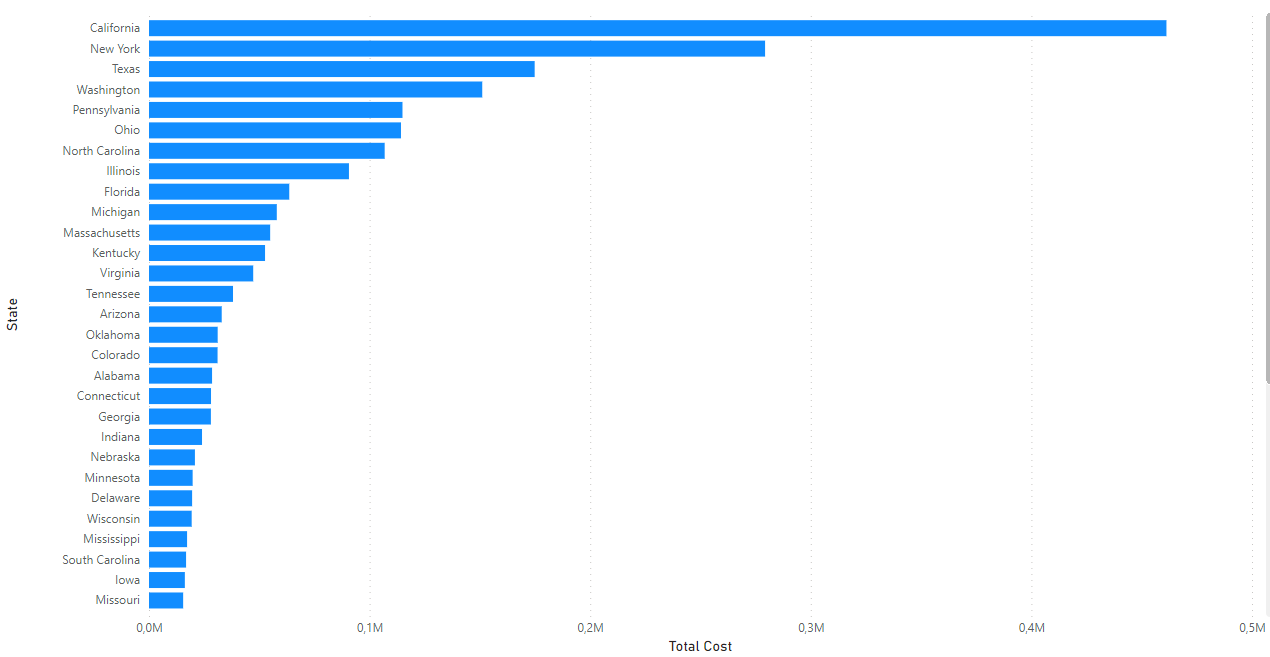


Looking at the sales gained form different types of customers and taking into account the difference in regions is beneficial for decision making. The results show that a consumer segment gives the highest revenue in all region. Then corporate and home office segments come. Only in the South, does the home office segment give a higher revenue than the corporate one.



However, looking at the segment sales over time prisma, it will be clear that the revenue from Consumers is decreasing, while others are increasing. Especially when looking at kpi of home office, it shows the greater growth.

## Sales By State



As seen, the distribution of sales among states is not uniform. The major sales are concentrated in California, New York, and Texas. The low sales in other states indicated the lost opportunities and potential bias when developing a marketing strategy.

## Sales By Subcategory



Chairs, Phones and Tables takes the largest revenue share. Smaller categories as Art, envelopes and labels suggest that a market should be explored for diversification.

# Recommendations

* Develop strategies to increase an average ‘basket’ of consumer without the sacrifice of overall sales volume. The main constraint can be increasing the diversity of higher-margin items
* Increase the marketing campaigns toward attracting more ‘Home Office’ segments, given the positive growth trends
* Develop sales programs and off-season promotions to stabilize the revenue over all quarters
* Make additional research to identify the unpopularity of the business’s service in specific popularity and develop an appropriate marketing campaign for reaching new customers
* Diversify the product category with different products. Try firstly to make a diversification in both high-demanded subcategories and low-demanded. If the low-demanded subcategory does not result in higher profits, then focus on promoting the high-demanded ones

# Conclusion

All things considered, the analysis reveals some patterns and trends related to business development and its prosperity. Specifically, it was found that the revenue highly depends on the season and is not stable, especially in Q4. Also, the declining average cost per item indicates that there are changes in product type (more cheaper products appeared in the market). Regarding the customer segment, the consumer segment gives the highest revenue among all regions, yet the ‘Home Office’ segment is rising over time. Moreover, when looking at the subcategory popularity, chairs, phones and tables dominate sales. It was recommended to focus on marketing campaigns in those states, that were not reached successfully enough and diversify the products available in the online shop. Finally, stabilizing the revenue by seasons by integrating promotions is suggested.

1. In order to give a clear answer to the question, visualization will be provided. The type of visualization type will be selected, so it could clearly reveal the answer. [↑](#footnote-ref-1)