I. Introduction

1.1 Overview

This project revolves around the use of data visualization in E-commerce businesses. Here we analyze obtained data from the business and perform data analytics using IBM Cloud and IBM Cognos Analytics and finally deliver a dynamic dashboard along with our insights from the visualizations.

1.2 Purpose

The decision makers of the company can simply work with the dashboard instead of looking at plain tabulated numbers. With this project we aim at aiding them in better understanding their customer base and making more effective and informed decisions.

II. Literature Survey

2.1 Existing problem

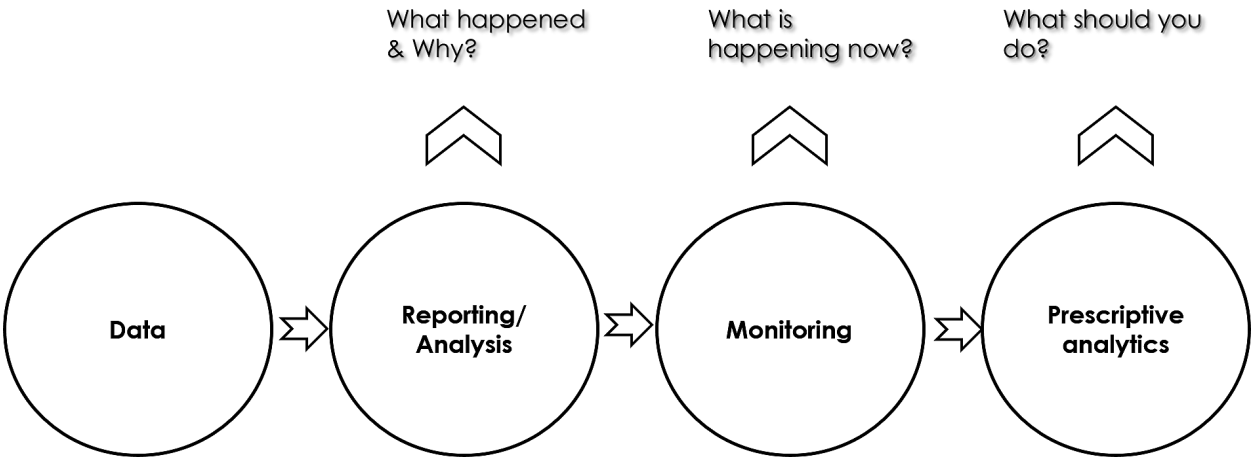
The current methods of deriving insights from large amounts of data relies on excel sheets and sql queries. But these approaches do not give the entire big picture.

2.2 Proposed solution

Dashboards providing a dynamic platform for analysis is the solution we propose for this problem. With the help of well crafted inter connected graphs, better insights can be drawn.

III. Theoretical Analysis

3.1 Block diagram

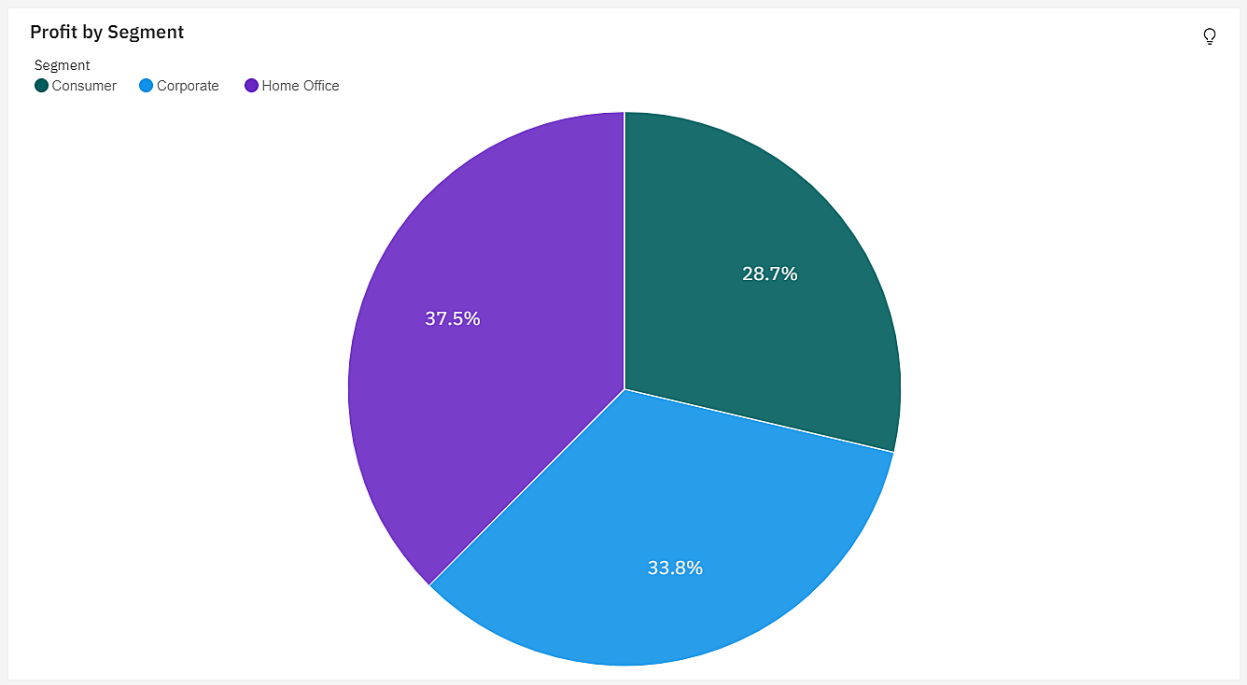


3.2 Software Designing

IBM tools have been used for this project. IBM Cloud and IBM Cognos Analytics.

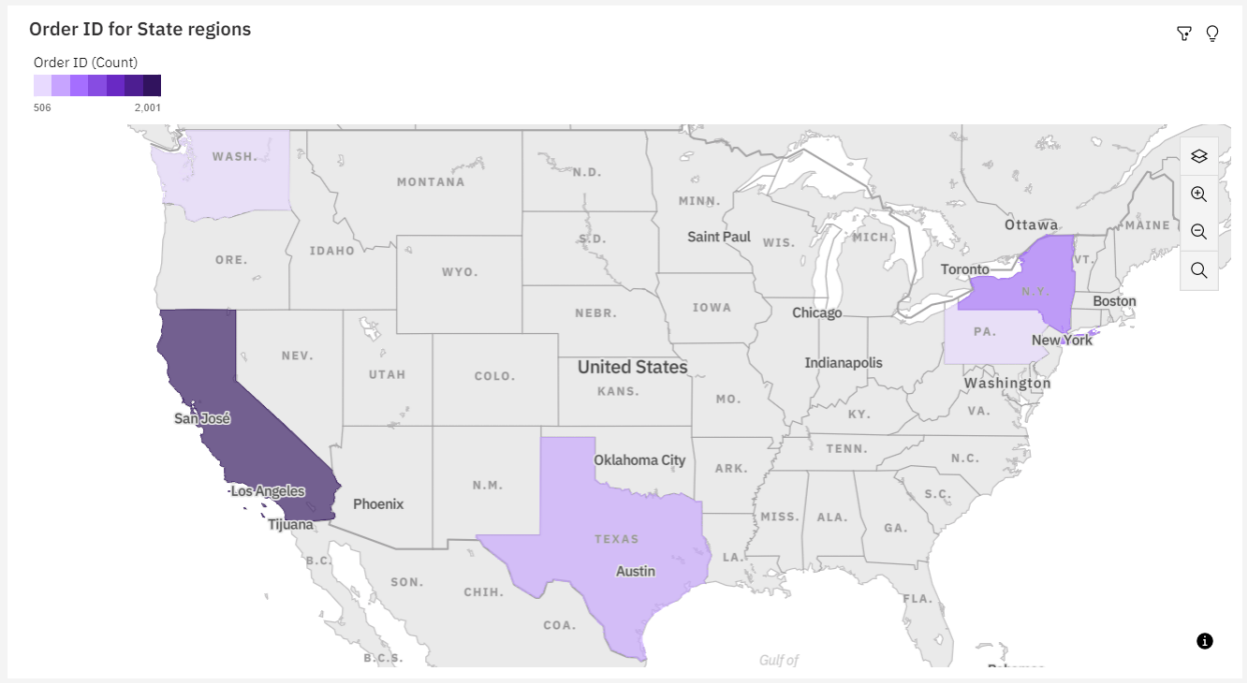
IV. Experimental Investigations

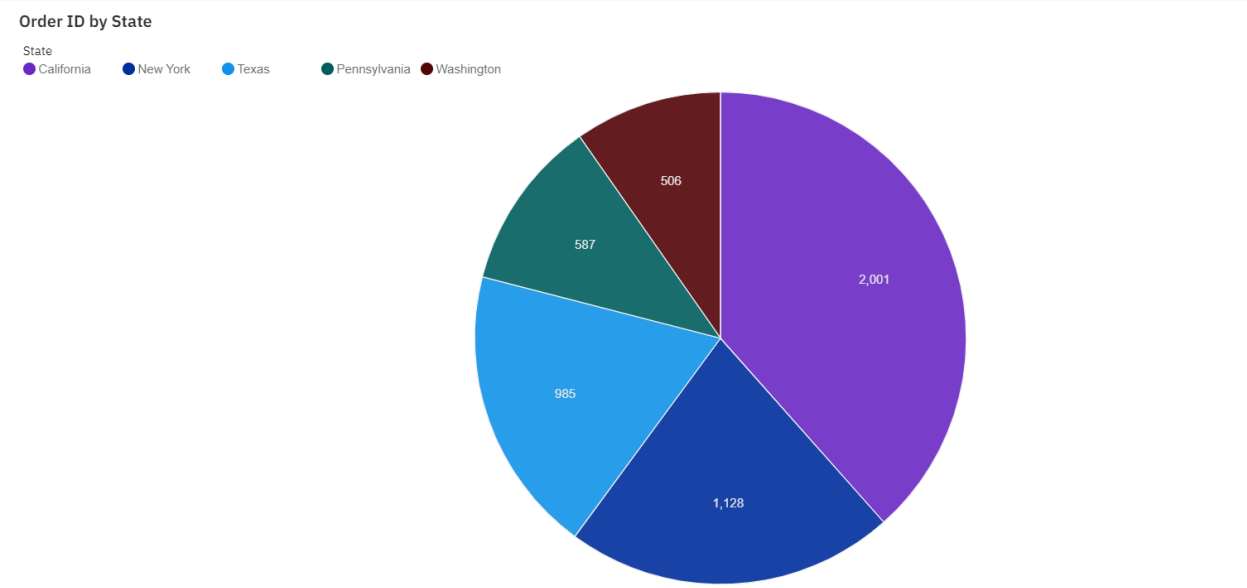
How are profits with respect to the segments?



Clearly the average profits gained are not segment dependent. Each segment has a equal contribution to the profit.

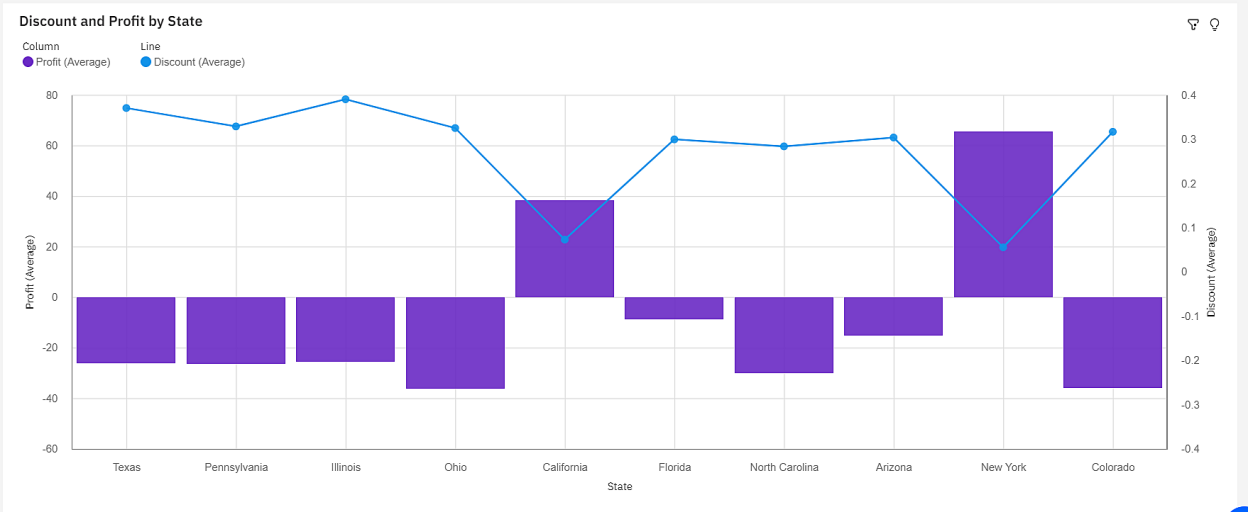
What regions account for a greater number of orders?





These are the top 5 regions with respect to total number of orders, which are a total of around 10k. California being the highest with around 2k orders, accounting for 20% of the total orders. Followed by New York and Texas with around 1k orders each, followed by Pennsylvania and Washington with 0.5k orders each.

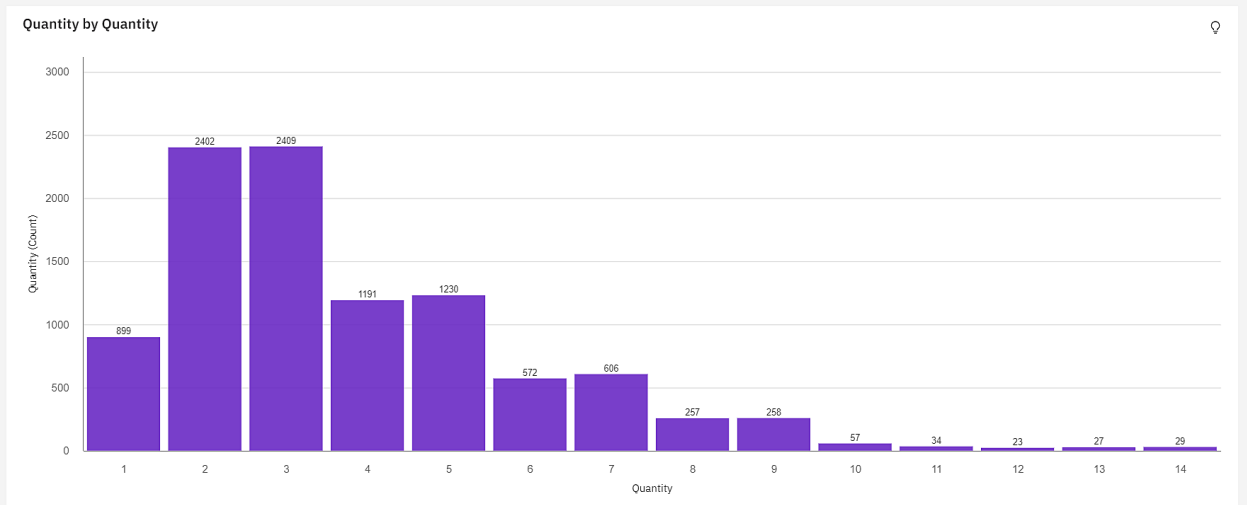
What is the trend between discount offered and profits obtained?



An unusual trend noticed here is that regions being offered a high average discount are tending to average loss, whereas California and New York account for high profits while having a rather low average discount.

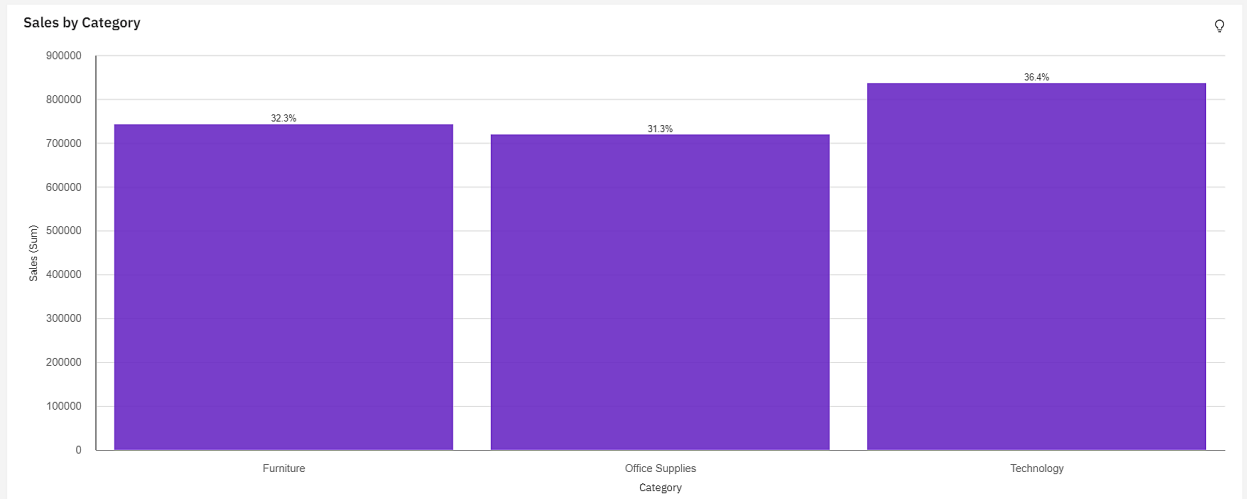
One reason that comes to my mind for this is that the customer base is well-established in these areas as they also account for the maximum number of orders. The marketing team and the delivery service or the product qualities of those regions are higher and well managed. The company can do a study on the teams performing in these areas to understand what is leading to success and apply those to other regions as well.

What is the frequency of quantity ordered?



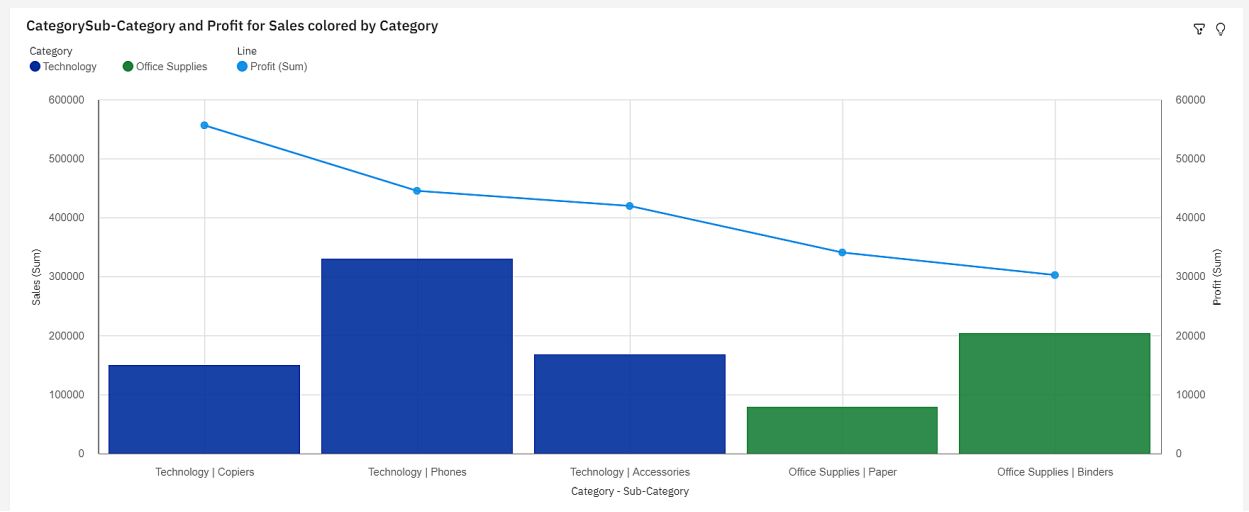
The distribution is right skewed with maximum orders containing 2 or 3 items. 48.1% of the total orders have 2-3 items in them. The customer base clearly is more inclined to buying 2-3 items. The company could work on their recommender system to shift the graph and eventually make it left skewed.

What are the sales per category?



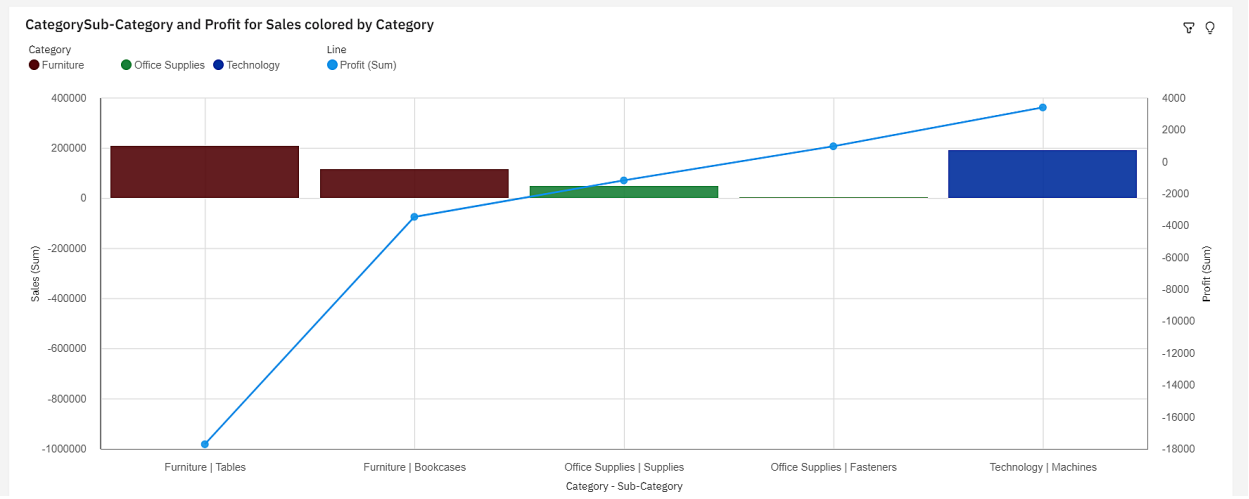
Sales are not much affected by the product categories since all the three categories are near the equal mark with technology having a slight edge, which is also of a insignificant difference.

Which are the most profitable product categories?



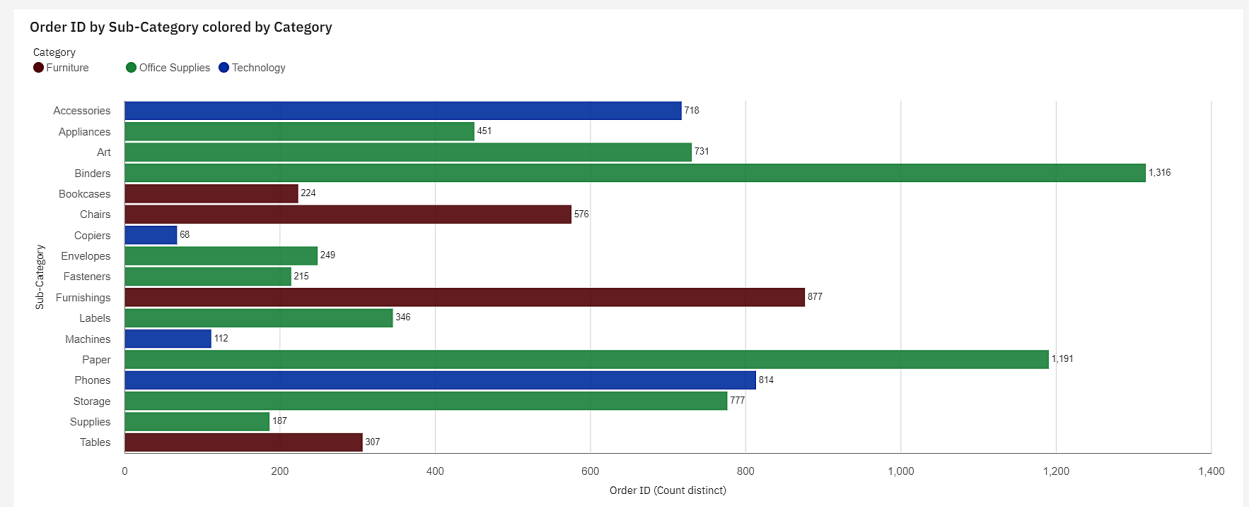
These are the top 5 profitable sub-categories of products, along with their sales. Copiers under the Technology category have low sales but gain the most profit. Therefore, promoting the copier can be profitable. Phones have high sales and profit. Under office supplies, Paper is seen to have a high profit too as compared to their sales and so good marketing can introduce high profit to this sub-category too.

Which are the least profitable product categories?



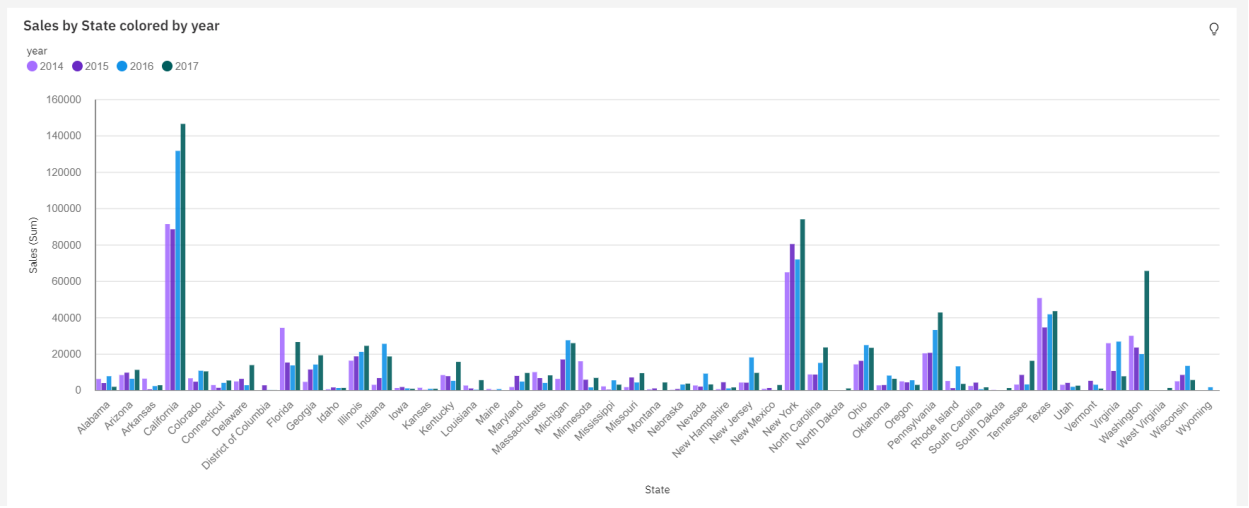
These are the bottom 5 sub-categories while taking profit into account. Even with sales higher than that of the copier, which is the most profitable product, tables are seen incurring a high loss to the company. The company can do a study here to see what is happening because the sales are doing fine. Slight loss is also seen in case of bookcases and these both belong to the furniture category.

How many times is a product type ordered?



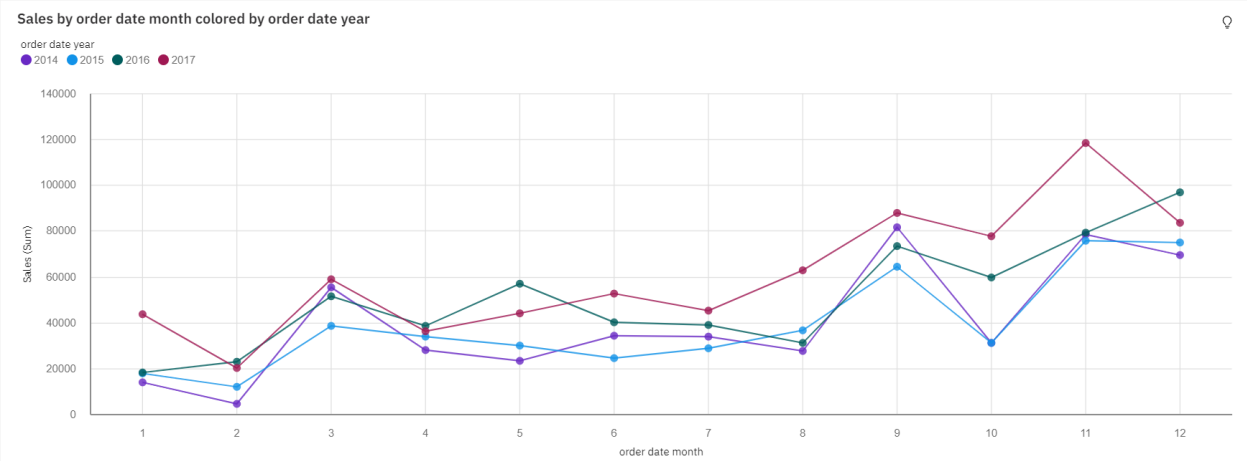
Binders (15.2%) and Paper (13.7%) are the most ordered categories with a combined count of 2893 items which is 28.9% of the total. It is also seen that Office Supplies is the most ordered category with 6026 items or 60.3 % of the total orders placed.

How are the yearly sales for various states?



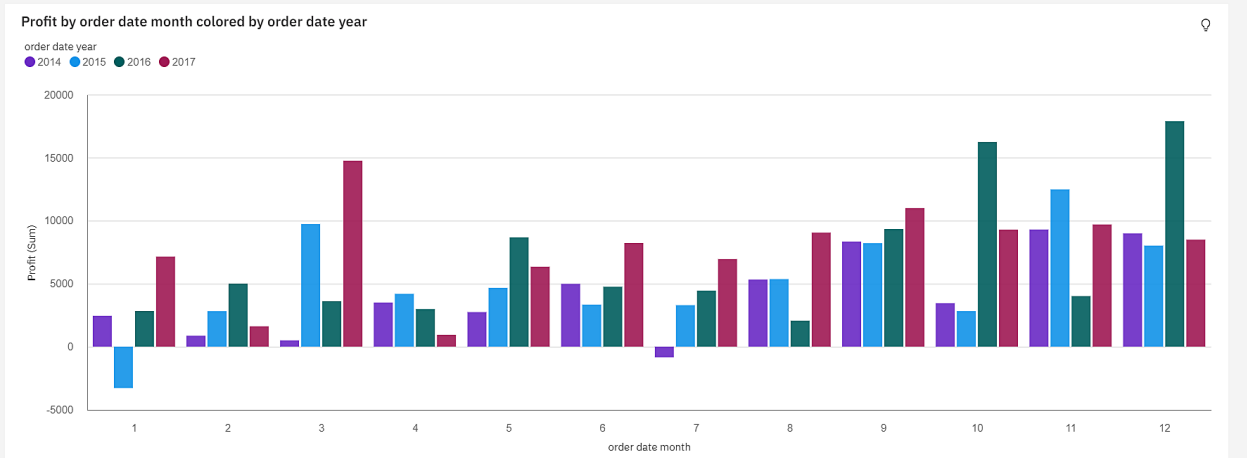
Clearly 2017 has been a good year in terms of sales. 2016 and 2017 combined contribute to 58.4% of the total sales. California started off moderately and shot high in 2016 and 2017. Georgia, Illinois, Michigan, North Carolina are some the states showing a gradual increase in sales without any decline. New Jersey, Rhode Island and Wisconsin had comparatively high sales in 2017, which can be investigated to bring a gradual increasing trend.

How are the trends in sales over time?



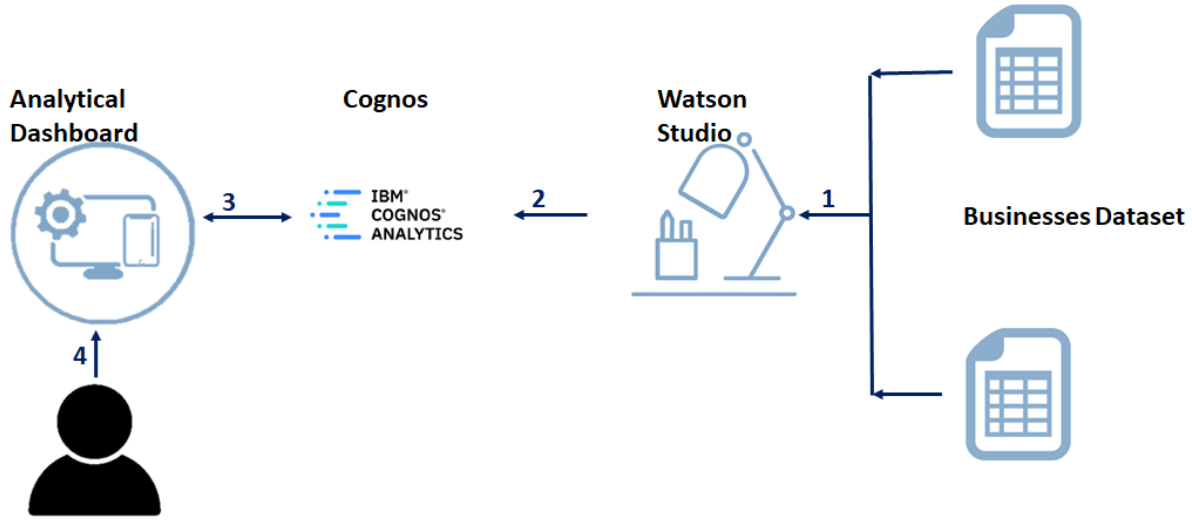
Sales are usually high in the months of September, November and December, in total accounting for 42.9% of the total sales. Years 2016 and 2017 saw the highest sales. In general the graph is increasing across the year with the highest sales in the year ends. The company could use this to their advantage and introduce discounts and marketing to attract more customers during this time of the year.

How are the trends in profits over time?



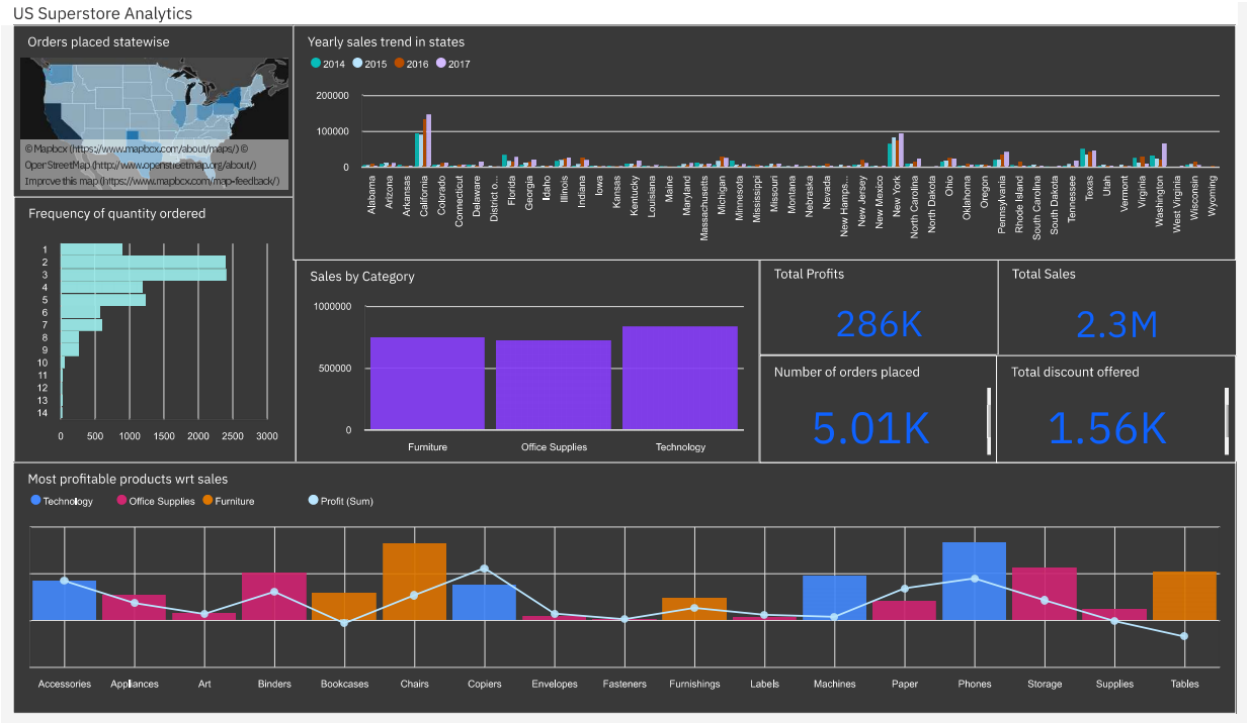
The profit in the month of December is the highest, accounting for 15.1% of the total. 61.2% of the profits in gained during the years 2016 and 2017. It seems that the company has bought in some changes during the last two years and they have fulfilled the objective of getting more profits.

V. Flowchart



VI. Result

The following is the final dashboard created. On clicking over any region or product category or product subcategory or year, changes the entire dashboard to reflect the data with that reference only. For example, selecting California and Technology and 2015 will show the parameters with those filters applied.



VII. Advantages and Disadvantages

Advantages

- Enhanced visibility
- Timesaving efficiency
- Better forecasting
- Key performance indicators
- Real-time customer analytics
- Better decision-making

Disadvantages

- No predetermined rules and hierarchies for how dashboard metrics are used
- Difficulty in attaching supporting data to a dashboard
- Failure of data to refresh automatically

VIII. Applications

The dashboard can help the company improve by giving them a starting point to work on the problem by giving a concise view of the entire dataset. The stakeholders can finally make sense of the data they hold and make informed decisions for the company and implement effective changes. This will also give them an edge over their competitors by helping them understand buying patters and their customer base better.

IX. Conclusion

We started with a huge dataset from an e-commerce buisness, which was first cleaned and then studied by using descriptive statistics techniques. IBM Cloud was then used to initialize the IBM Cognos Analytics service. A great amount of exploration was done using various visualizations, following which a dashboard was built to incorporate all the key findings.

X. Future Scope

Furthermore, many more comparisons and visualizations can be done to get insights on any unusual patterns. There are also suggestions of improvements from the analysis done in the previous section which can be inculcated into the company's process.

XI. Bibilography

https://www.ibm.com/docs/en/cognos-analytics/11.0.0