DATA ANALYTICS AND ORGANISATIONAL DECISION MAKING

University of Southampton – Masters of Science – Mathematics – Data and Decision Analytics

28965736 Vyom Khanna
vk1n23@soton.ac.uk

Table of Contents

LIST OF TABLES

LIST OF FIGURES

PART A

1. INTRODUCTION

In data science, the data is classified as structured, a predefined schema, and unstructured data which is also called as Big Data, a huge amount of data that does not have a predefined schema and cannot be processed either analyzed using conventional tools(Batko and Ślęzak, n.d.).

Examples of big data are traffic controlling system, financial and banking, health care units, logistics and supply chain, media and entertainment, government resources etc. Due to its large volume, analytical tools are used to extract information from the BD and it is very important to use these tools as BD lends valuable insights on organizational decision making and for increasing competitive advantages(Su et al., 2022). Judicious use of BD analytics in financial sector yields following benefits:

- *Streamlining their innovation process* to transform customer's information for new revenue generation like introduction of financial products, managing risks and enhancing customer service quality(Hajiheydari et al., n.d.).
- Helps financial analysts to *make better investment decisions*. Furthermore, it helps to understand the pattern of customer spending ability to draw some business models(Ali et al., 2021).
- Improves *knowledge on financial markets* and enhance the quality of service, thus refining the overall banking experience(Jouti, n.d.)
- Drives the financial institution to *leverage customer's data*, promotes algorithmic trading, creates transparency and transforms the culture(Hasan et al., 2020).

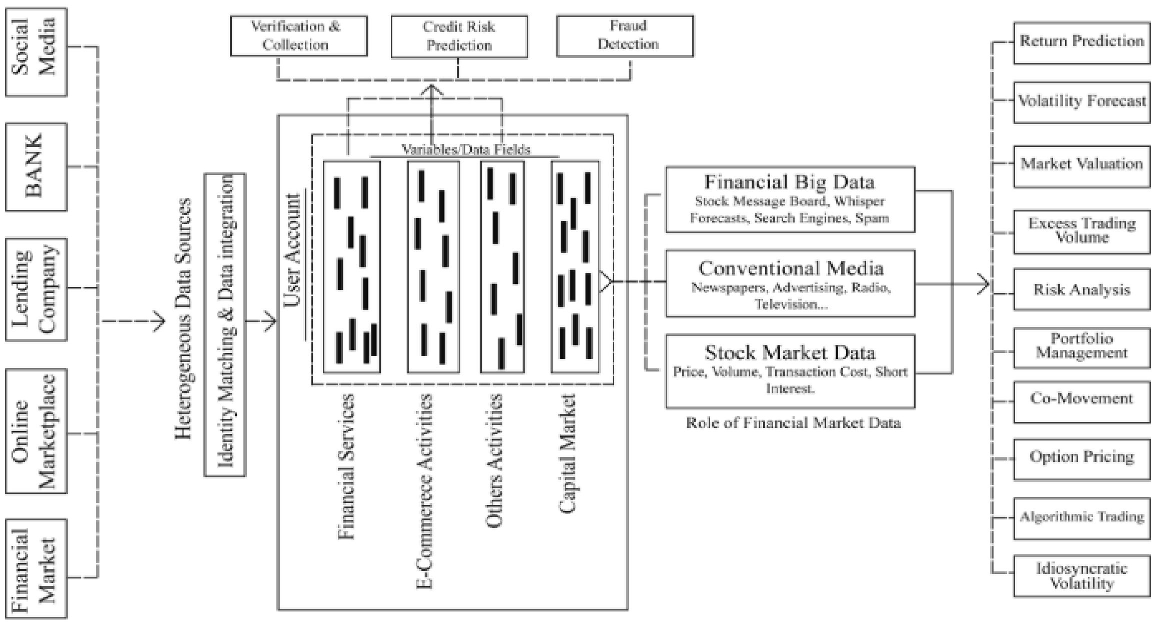


Figure 1: Theoretical framework of BDA in financial sector (Hasan et al., 2020)

1.1. Anticipated Issues Encountered During Implementation of Data Analytics in Organizations

Besides the recognition of many benefits of using DA in financial industry, many theories and models details some issues that could be faced during enactment of DA into the banking system.

- According to ***critical skepticism theory***, a financial institution might *face hindrance to tackle misreporting of information thereby compromising the objective of the organisation* if they introduce DA into their system. So, there would be call for tampering prevention and detection mechanism to mitigate the breach of data or proper auditing of the report(Alonso and Câmara, 2023)since the data flows between the legacy systems. As the theory focusses on the designer-agent-principal game, the principal would be *accounted for the effective auditing of information which is a risk* if there is a lack of proficiency since the financial banks comprises sensitive data.
- Inadequacy of knowledge on DA could *lead to cognitive biases*, however, *time constraint also plays a crucial role for the adoption of technology* as evaluation of available options consumes more time. With the support of ***rational decision theory***, the outcome of each option is evaluated and the biases could be reduced by portraying the benefits of digital strategy implementation to the workforce(Eastman et al., 2023)besides requiring more time for adoption to the technology.
- The application of digital based technologies like analytics in their decision-making system could lead to face challenges which might *arise due to uncertainties*. In accordance to ***dual theory of decision making***, fluctuations in market trends, critical thinking and formal decision support system, *sparse knowledge on adoption of these tools* could override the advantage of intuitive thinking that tacit expertise and practical wisdom would bring an explicit outcome and exhibit intrinsic insight by their practice(Svenson et al., 2024).
- ***The variance theory of deferred action*** alleges *requirement for an effective designing of DA* architecture as it is crucial to fledge a business process or to introduce any change in the financial decision-making operations. Imperfection in designing, effects the flexibility of processes thereupon requiring a perpetual investment on the development of the architecture, which impacts the financial budget allocation of the institution(Nandish V. Patel, 2011).
- Some technologies require trial and error method for evaluating the performance of them so that they could executed on large scale operations. ***The status quo bias theory*** explains that the bank might resist to get adopted to digital strategical management as it *requires potential costs for new set of procedures* if at all implemented(Almatrodi et al., 2023) like merging of departments calling for change in organisational structure.
- The *reliance of decision-making information* provided by supportive organizations would emphasis to maintain a strong relationship and autonomy if the bank does not have proper digital infrastructure and trained professionals to use the DA tools judiciously. This also consumes additional cost with

challenges arisen due to disputes on dominance perspective, and contractual agreements. So, train the trainers' model was used to train the personnel and look for opportunities to reduce the dependencies on third parties as stated by ***resource dependency theory***(Birken et al., 2017).

- The ***institutional and SIGMA theory*** elaborates on *implication of pressure on the resources* to comply with rules and procedures that mandates them not to think out of the box. This machinery model of working limits them to explore more options and even to lose many resources too within the bank(Birken et al., 2017)(Dietz and Mulder, n.d.).
- The perception of implementing DA into the system by the bank employees directly *influence their mental state of accepting the change*, mainly carrying the feel of job insecurity and *directly reflecting on their performance* to show off their neglection to DA adoption as explained by the ***theory of affordance***.(Ben-Zeev, 1981)
- There would be *consideration of architectural construction layer of an institution* if there is a need for implementing DA into their system. The ***Alpha theory's ontological model*** explains that the process would take long time to get implemented since the approval procedure would be time consuming especially in banking sector where there is a huge hierarchical structure(Dietz and Mulder, n.d.).

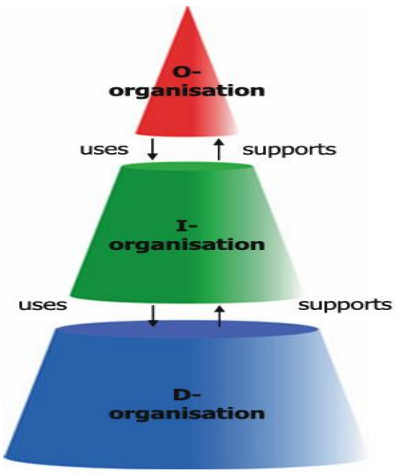


Figure 2:The ontology model from Alpha theory (Dietz and Mulder, n.d.)

D- Documentational

O- Original

I- Informational

The figure highlights how the information is being generated from the base level and flows upwards to higher level for requiring support

1.2. IMPACT OF DA ON CURRENT DECISION-MAKING PRACTICES AT BANK

DA carries both positive and negative impact on the current decision-making practices when adopted in the bank X.

Positive impact

- Based on ***PSI theory***, there would be *accountability of the artefacts* at each level of the bank X where the error in generation of information would be kept at minimal level thereby increasing the robustness(Dietz and Mulder, n.d.).

- Adoption of IS into the system would help to *coordinate the work flow and information* system between various users within the bank. Furthermore, they act as a *middleware between separate offices and legacy systems*(Stohr et al., 2001).
- Although there are issues with the quality of information being produced by using IS, they provide opportunity to *address the issues along with safety and efficiency through streamlining techniques*(Zayas-Cabán et al., 2023).
- Data Analytics permits more customization that helps to *provide personalized segmentation services based on multiple aspects* such as demographics, geographics, behaviour and selecting patterns, etc. This segmentation will lead to more effective use of resources and high productivity with highly targeted marketing and product offerings(Marshall et al., 2015)within the bank.
- Their capabilities allow for *better risk assessment and management*, reducing losses from defaults and fraud through more accurate prediction models where flexibility helps in issuing effective solutions to emerging risks(Safitri and Geraldina, 2023).
- Analytics ensures the organization's operations meet the ethical standards and social responsibility goals through internal and external factors in which internal *focuses on stability and economic* performance whereas the *external focuses on social contribution and environmental responsibility*. (Tan and Tsionas, 2022).

Negative Impact

- There would be *chances for continuous failure of enrolment, strategy deployment and occurrence for competition over period of time* among the employees and also the superior officials fail to negotiate and persuade people to use the technology. Apparently, this dynamic social reflection competes against the use of technology supporting participation.(Rivera and Cox, 2014)
- Due to the legacy system being used in the bank, there are two problematic ways for adoption: technical and social. Technical adoption issues include *deployment efficiency of information* as it takes long time to build, test and produce results that would prolong to take decisions. Social adoption includes *contradiction of data produced by various departments* leading to late informed decision making(Laukkanen et al., 2017).

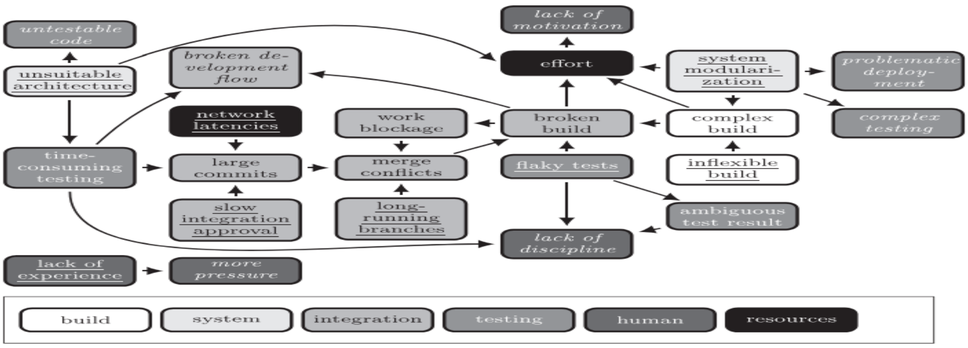


Figure 3: Effect of technical and social adoption issues on current decision-making practice at bank X (Laukkanen et al., 2017)

- An *appropriate usage of analytics software* could improve strategic decision making, failing to do so, *would not experience optimal decision making*(Liu et al., n.d.).
- More chance for *conflicting goals* between motivational experimentation and discouraging misreporting. This could decrease the productivity of knowledgeful resources and finally the goal of implementation is crossed(Alonso and Câmara, 2024).

1.3. MEASURES TO ADDRESS THE CHALLENGES

MEASURES	APPLICATION
Prioritization	• Investing in IT infrastructure and **selecting the right tools** to provide the necessary technical support(Farooq Aziz, 2023). • Bank should follow a **structured order** of priorities and promoting a data-driven culture and addressing talent needs(Kumar, 2018).
Establishing Data Center/ Hub	Develop and run a data center and data-driven culture within the organization for **educating and motivating** all employees from top management to operational staff about the benefits of data-driven decision-making and encouraging them to learn about data insights for data privacy and protection. (Zhang et al., 2018).
Data quality improvement	• Bank must prioritize a **robust data governance framework**. This includes developing comprehensive data standards and policies and regularly auditing data. • **Invest** in data cleansing and integration tools. • Tools to **automate the processing and merging** of data from different sources, thereby reducing missing or inconsistent data and ensuring reliable analytics(Kwon et al., 2014).
Data protection measures	• Ensure compliance with relevant **data protection regulations**. • Develop a **data privacy policy** and conduct regular compliance reviews. • Adopt **advanced security techniques**, including data encryption, access control and continuous monitoring systems(Dorr et al., 2015).
Organizational Culture Development	• Engage employees in the transformation process by **demonstrating the value** of data analytics through success stories to give internal staff confidence in the transformation. • **Comprehensive training** to enhance employees' skills and make them realize the importance of data analytics. • **Focus on culture and innovation**, the IT department and other business departments, such as HR, In close cooperation with finance, and manufacturing, features in (Cato et al., 2015). • **Collaboration and communication** with data actors that are external to the organization has been considered (Hawley, 2016).
Talent development	Develop **competitive compensation packages** and clear career paths to attract top talent. In addition, establishing a comprehensive **in-house training program** would help upgrading the skills of existing employees and ensure that data analytics tasks could be performed effectively(Popovič et al., 2018).
Technology Investments	**Selecting the right data analytics platforms** and tools that are easy to use and could integrate with existing systems. Continuously evaluate and innovate these platforms to keep up with technological advances and provide the necessary support for complex data analytics operations(Zhu et al., 2016). This enables more informed, flexible, and transparent decision-making, and ultimately gain a long-term competitive advantage

Table 1: Measures to prioritize the challenges

2. Importance of data quality in data analytics

In the era of big data, the importance of data quality is worth discussing as most of a company's operational and strategic decisions rely heavily on data (Batini et al., 2009).

Incorrect conclusions drawn from poor-quality data could have a detrimental effect on client satisfaction, operational efficiency, and business decisions. Ensuring accuracy, consistency, completeness, reliability, and timeliness are critical components of high-quality data and are necessary for making well-informed judgments. Data quality is crucial in big data analytics for the following reasons:

- **Accurate Insights and Decision-Making:** To find patterns, insights, and trends that guide decision-making processes, big data analytics analyzes enormous volumes of data. If the data itself is of low quality, containing missing values, inconsistencies, or errors, the insights generated from the analysis will not be reliable, *leading to incorrect conclusions and possible financial losses* (Hazen et al., 2014; Janssen et al., 2017).
- **Regulatory Compliance:** Organizations are required to comply with *strict reporting requirements and data quality standards*, and specifically the banking sector is highly regulated. Heavy fines and potential legal repercussions could arise from non-compliance due to poor data quality (Gartner, 2019).
- **Customer Experience:** To deliver personalized assistance, specific advertising campaigns, and customized offerings for products, firms use consumer data. *Inaccurate client information* could result in overlooked opportunities, inefficient marketing campaigns, and unsatisfactory customer experiences, which can ultimately impact customer happiness and retention (Kwon et al., 2007; Ofek et al., 2016).
- **Operational Efficiency:** In the banking sector, for instance, numerous banking procedures, including accounting, fraud detection, and credit risk assessment, rely heavily on data. The *precision and efficiency* of these procedures could be hampered by poor data quality, leading to raised costs and causing operational inefficiencies (Hazen et al., 2014; Janssen et al., 2017).
- **Risk of data analysis:** Chance for more *data errors could lead to conflicts*. Further, it is important to establish the connection between analytics and organizational risk management. This approach helps to ensure long-term sustainability by mitigating risks that produce insightful data-driven strategies.(Edwards and Taborda, 2016).
- **Efficiency of business processes:** The project should be completed within the stipulated timeline and prohibit delays. Accurate business analysis and forecasting help to *optimize business processes and improve effective productivity* with higher returns of profit which could be attained with good quality of data processing and analysis(Kubina et al., 2015).
- **Powerful architecture**: Data quality is very crucial for transparency and acceptance in BI systems. Faulty master data with inconsistency could affect the BI architecture(Karin, 2016).

- **Successful testing**: A successful end to end testing is important in multi-component and distributed systems. So, *proper metrics evaluates test data quality* which allow for the comparison of different data sets, enabling testers to arrive better decisions and ensures effectiveness of the process(Yury, 2016).

3. Strategies to Improve Data Quality at Bank X

In order to address the challenges with data quality, BankX should put into action a comprehensive list of plans comprising both technical and organizational elements.

a) Data Governance Framework

Establishing a strong data governance framework builds data stewardship, responsibility, and constant monitoring of data quality throughout the firm (Otto, 2011).

Implementation:

- **Establish a data quality management process:** Process-driven strategy consists of two main techniques: process control and process redesign(Batini et al., 2009). Design and implement an all-encompassing data quality management process that covers data collection, storage, processing, and analysis.
- **Data Stewardship:** Appoint data stewards who can take the responsibility to oversee data quality in different departments. By establishing explicit data governance procedures and communication channels, data stewards help departments collaborate and communicate more effectively, ensuring that all departments have the same data practices.
- **Policies and Standards:** Create and implement standards and policies for data quality, such as protocols for data entry, metadata management, and validation rules.
- **Data Quality Metrics:** Define and track critical metrics of data quality like accuracy, consistency, completeness, reliability, and timeliness (Khatri & Brown, 2010).

b) Data Cleaning and Transformation Tools

Data-driven is strategy for improving the quality of data by modifying the data value directly (Kwon et al., 2014). Bank X develops data-driven technologies such as data cleansing and pre-processing processes to scrutinize critical data such as customer information and transaction records. Effective data cleansing and transformation solutions may automate the process of discovering and correcting data quality concerns, which makes the data more reliable for analysis (Rahm & Do, 2000).

Implementation:

- **ETL Processes:** Clean and standardize data from external sources and legacy systems by using sophisticated Extract, Transform, Load (ETL) tools
- **Machine Learning Algorithms:** To identify and correct duplicates, anomalies, and inconsistencies in the data, employ machine learning algorithms. The adaptive learning capability of these algorithms enables data quality management procedures to evolve with the change in business needs and data patterns (Zhang et al., 2017). For example, ensure the accuracy of entered information when customers open accounts and conduct regular data consistency checks during data storage and analysis.

c) Centralized Data Repository

All the data in a centralized data repository is stored in one uniform format, which makes managing and analysing data easier. (Inmon, 2005).

Implementation:

- **Data Warehousing:** By implementing a data warehouse, data from a number of legacy systems can be consolidated into one unified repository.
- **Data Integration:** Data integration tools ensure that data from various sources is consistent and compatible by streamlining the process for consolidation (Vassiliadis et al., 2002).

d) Data Quality Training and Awareness

Sustained development requires that all the employees fully understand the importance of high-quality data and their role in retaining it (English, 1999). Provide targeted data quality training for account managers, data analysts, and IT staff.

Implementation:

- **Training Programs:** Provide employees with frequent training on best practices for data quality. Regular data management seminars are organized to share best practices and success stories.
- **Awareness Campaigns:** Awareness campaigns to showcase the impact of data quality on decision-making and business outcomes should be launched (Redman, 1998). Enhance staff's attention and skills on data quality.

e) Regular Data Audits and Quality Assessments

Regulatory bodies across various industries require organizations to maintain high data quality standards (Pansara, 2023). Regular quality assessments and audits support the early detection of issues in data quality and help tracking the success of data quality measures (Kim et al., 2003). Bank X need to develop and strictly enforce uniform data standards and specifications, such as customer information entry standards and transaction record formats.

Implementation:

- **Regular Audits:** Perform frequent data quality audits to assess adherence to data governance policies while identifying opportunities for improvement. Holding regular data quality assessment meetings to discuss and improve the data quality management process to adapt to business needs and environmental changes(Zhang et al., 2015).
- **Quality evaluations:** To evaluate the present level of data quality and track improvements over time, conduct periodical evaluations using data quality tools. This helps in identifying the gaps while ensuring that data quality measures of the firm are in line with or exceed industry expectations (Kim et al., 2003). Ensure that each department follows the same specifications in data processing to ensure data consistency and integrity.
- **Continuous Monitoring and Improvement:** Establish a continuous data quality monitoring mechanism, such as setting up a data quality dashboard to monitor data quality indicators in real-time.

f) Socio-Technical Aspects

Technical Aspects

- **Technology Stack:** Invest in cutting-edge tools for analytics and data management that can offer robust features including scalability and flexibility for handling various data types and increasing data volumes (Watson, 2012).
- **Integration Capabilities:** Make sure that all new systems offer smooth data integration and are compatible with legacy systems already in place. This minimizes the associated risks in data migration and transition of the systems preventing data being lost or corrupted, besides reducing the system downtime (Buhl et al., 2013).

Social/Organizational Aspects

- **Change Management:** To handle resistance and guarantee the smooth adoption of new procedures for data quality, develop and improve a change management strategy (Kotter, 1996).
- **Stakeholder Engagement:** Involving key stakeholders from all departments right from creation to execution of data quality programs guarantees cooperation and acceptance (Freeman, 2010).
- **Cultural Shift:** Promote a culture where data is viewed as a crucial asset and the maintenance of excellent data quality is emphasized (Schein, 2010).

g) Future Ambitions and Organizational Changes

BankX aspires to use big data analytics to stay ahead of the market competition, make better decisions, and provide better customer service. This needs considerable organizational adjustments to meet these objectives, including:

a) **Organizational restructuring:** To facilitate data-driven decision-making, rearrange organizational structures such as establishing specialized teams for the quality of data and analytics (Davenport & Harris, 2007).

b) **Process Redesign:** Revisit the design of business processes to include validation and perform data quality checks at every phase (Hammer & Champy, 1993).

c) **Technology Upgrades:** To meet the increasing demands of big data analytics, continuously improve your technology infrastructure (LaValle et al., 2011).

h) Challenges and Considerations

a) **Resistance to Change:** Employees might resist changes to their responsibilities and workflows (Dent & Goldberg, 1999).

b) **Resource Constraints:** Implementing extensive data quality initiatives requires significant investment in technology and training (Chaffey & White, 2010).

c) **Data complexity:** It is fundamentally difficult to manage and integrate data from many sources and formats (Gandomi & Haider, 2015).

To address these issues and ensure alignment and support during the implementation process, BankX has to take an organised approach, give priority to areas that will have the biggest impact, and keep channels of communication open with all the stakeholders.

PART B

1. Visualizations for question 1

Charts in the Power BI helps to comprehend complex information into understandable way by providing intuitive and instantaneous representations. Many charts were used based on the requirement of the problem statement(Aspin, 2022). Dataset has been processed as it included the returned products which would impact in terms of value of sales and profit.

a. Visualization of number of products under each category of Alpha sales

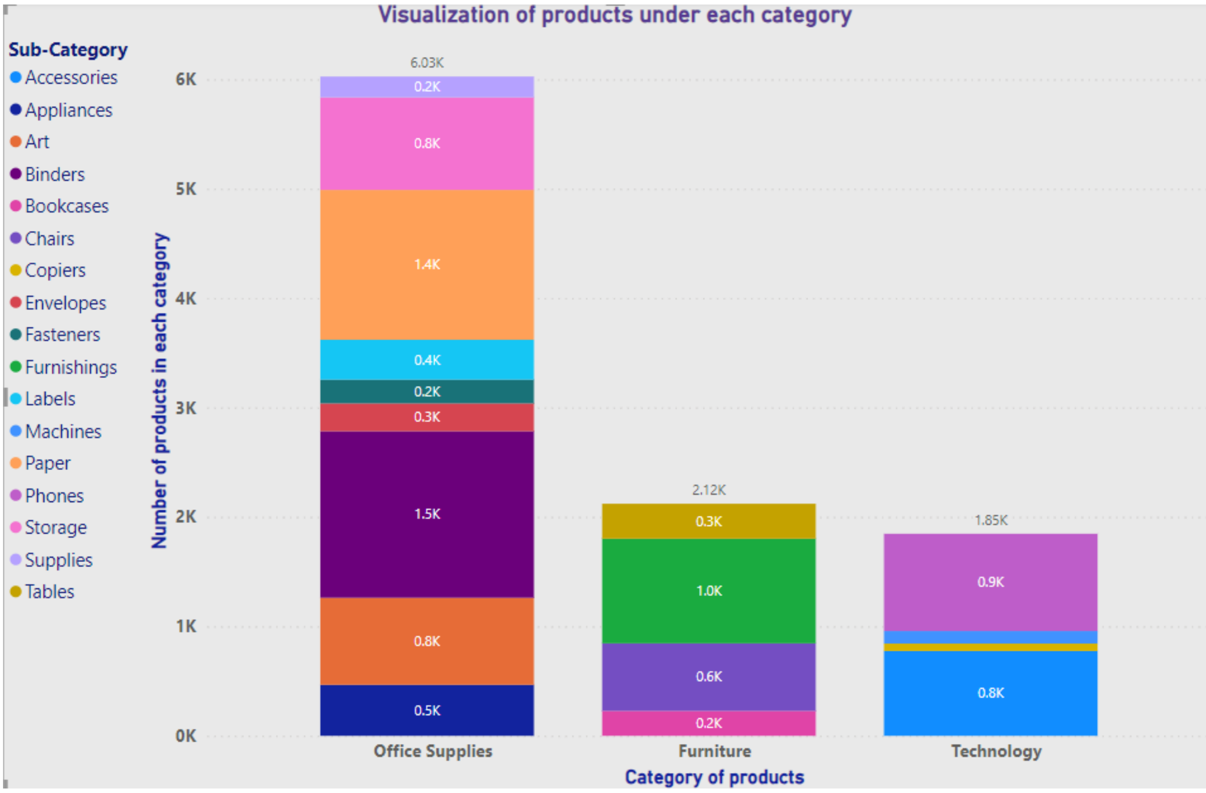


Figure 4: Stacked column chart for determining the number of products in each category

Stacked column charts are used to visualize the breakdown of sales by category wise and also aid in predicting the trend and comparisons within the dataset of sales. In this analysis, it could be observed that office supplies category has more products when compared to technical stuffs and more variety of binder product are available. It is suggested that more technical products could be sold as more operations are manually replaced by machines day by day, on the other hand this increases the sales volume also(Singh & Jadhav, 2022).

b. Visualization of top and bottom 10 products with respect to total sales volume

The charts like funnel, stacked bar, pie, clusters would aid in extracting the insight into sales performance for business Intelligence(Jolly, 2023).

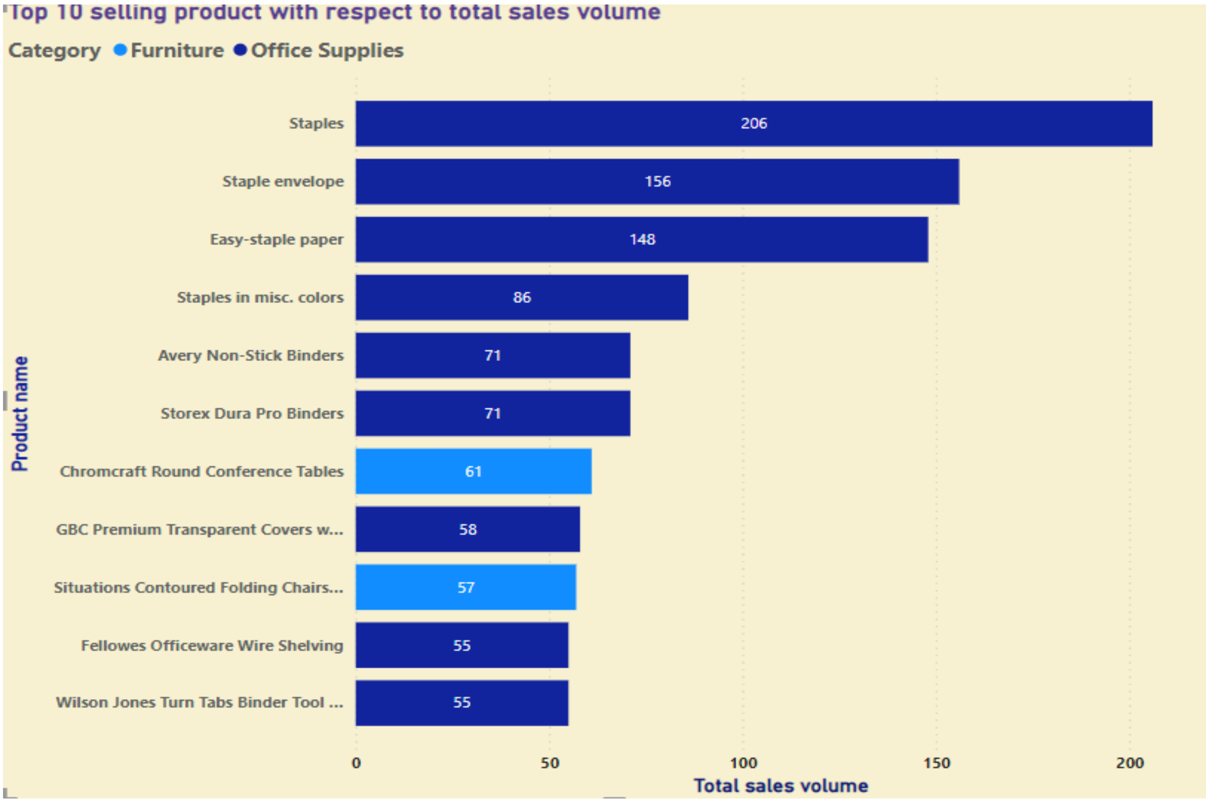


Figure 5: Stacked bar chart for visualizing top 10 products

The analysis presents that the more number staples was the most likely sold item and it could also be observed that office supplies are the most sold quantity that constitutes sales.

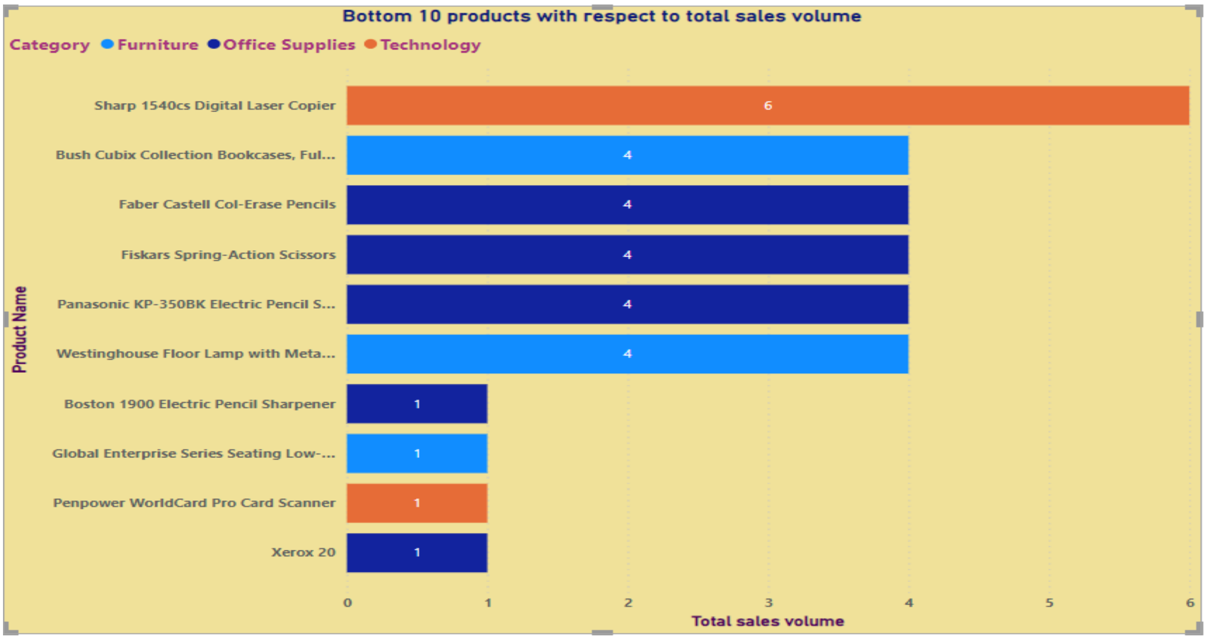


Figure 6: Stacked bar chart for visualizing bottom 10 products

As there were too items that constitutes same number of volumes sold, *average count of each sold item* is considered and plotted with respect to products. The visualization pictures that again office supplies sub categorical items were sold in least quantity. The strategies for improving the poor sold quantity of products could be devised and implemented like price adjustments, many brand for specific brand(Liu and Chen, 2022).

c. Performance of sales in various regions

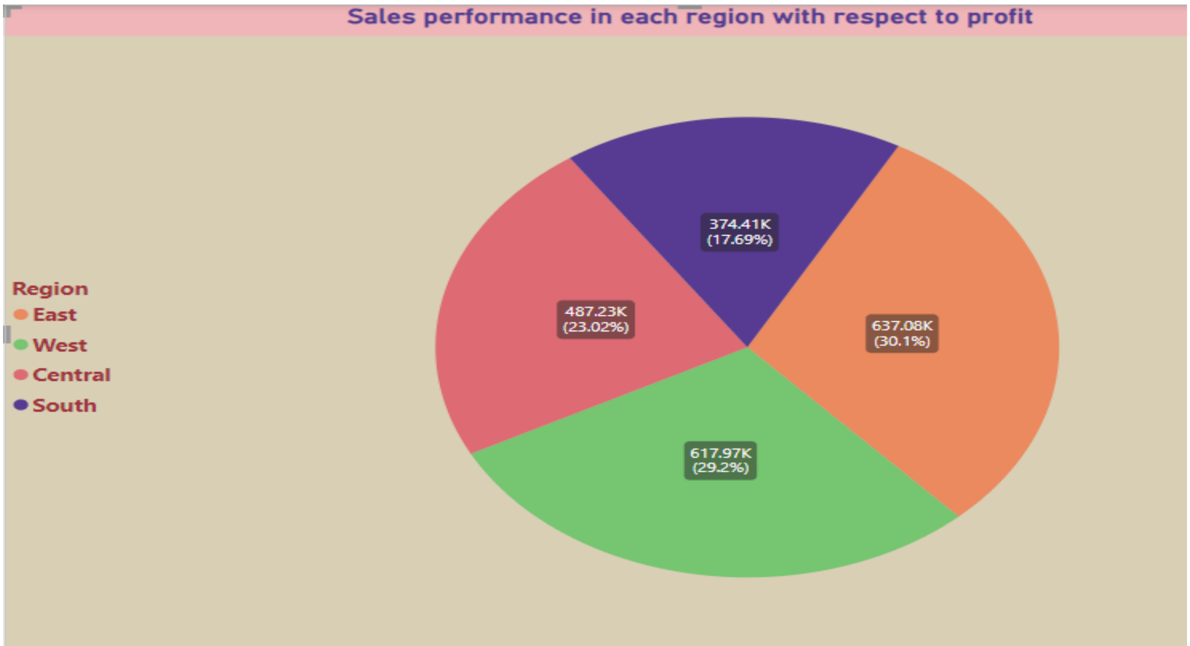


Figure 7: Pie chart for visualizing sales performance in all regions

Pie charts are most preferred for visualizing the sales performance as it aligns with cognitive learning theory to inflate cognitive load. When implicated into analysis, it is observed that east and west regions contribute high sales performance, on the other hand, south has poor performance which might be due to competitive market in that area or ineffective marketing of products or lack of selective products that are likely to be sold with respect to people's choice (Jie et al., 2018).

d. Top and bottom 10 products in each region

d.1. East region

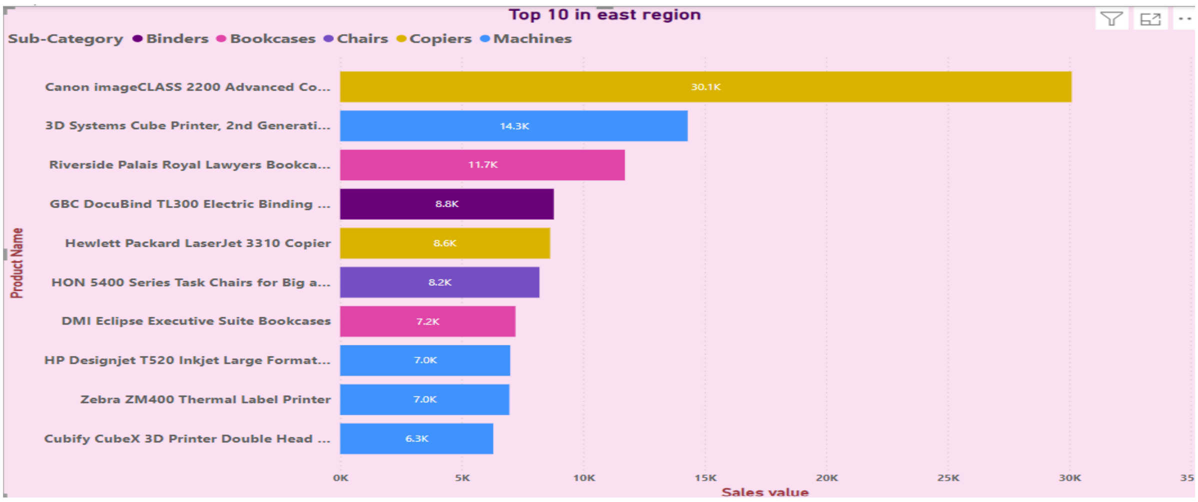


Figure 8: Top 10 products in East region

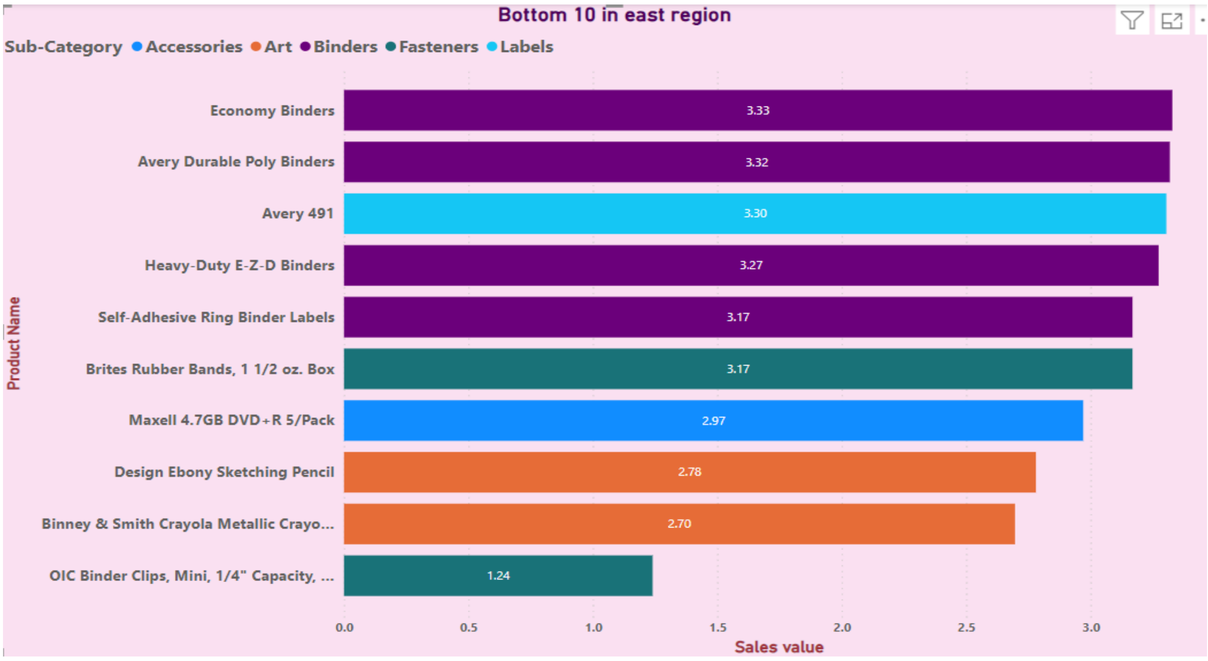


Figure 9: Bottom 10 products in east region

In contrast to central region, copiers were observed to be contributing more to the sales on the other hand, binders are the least number in the east region.

d.2. Central region

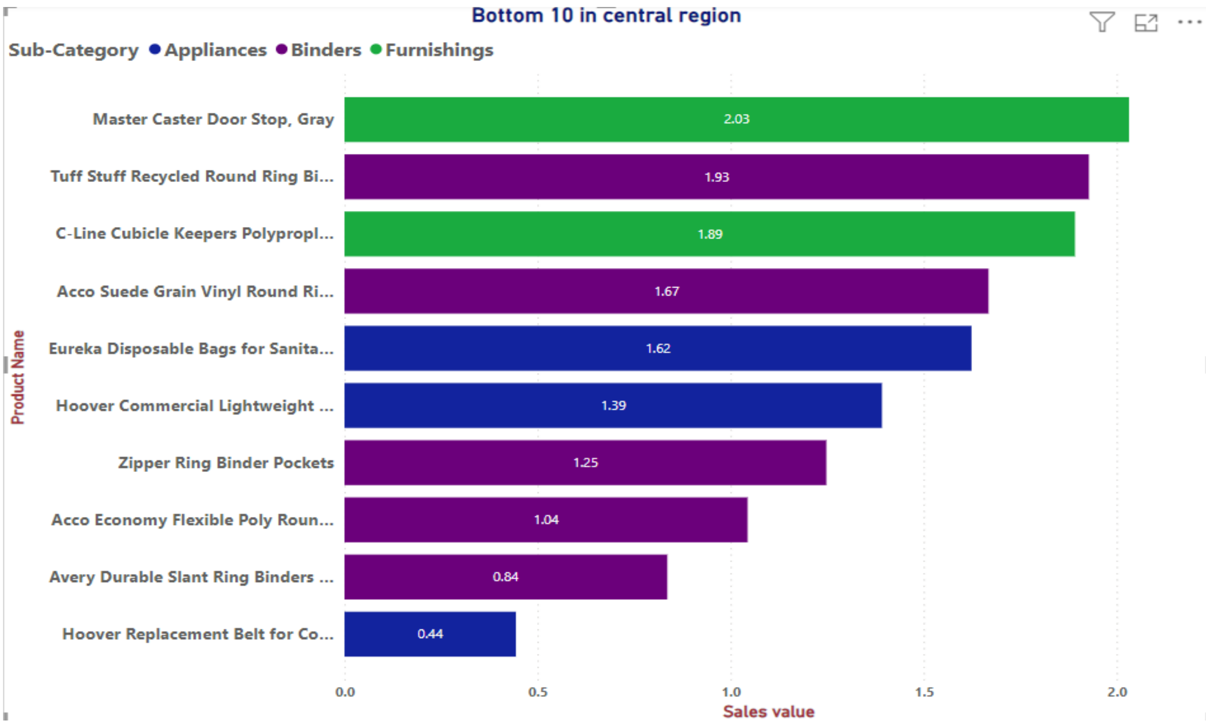


Figure 10: Bottom 10 products in central region

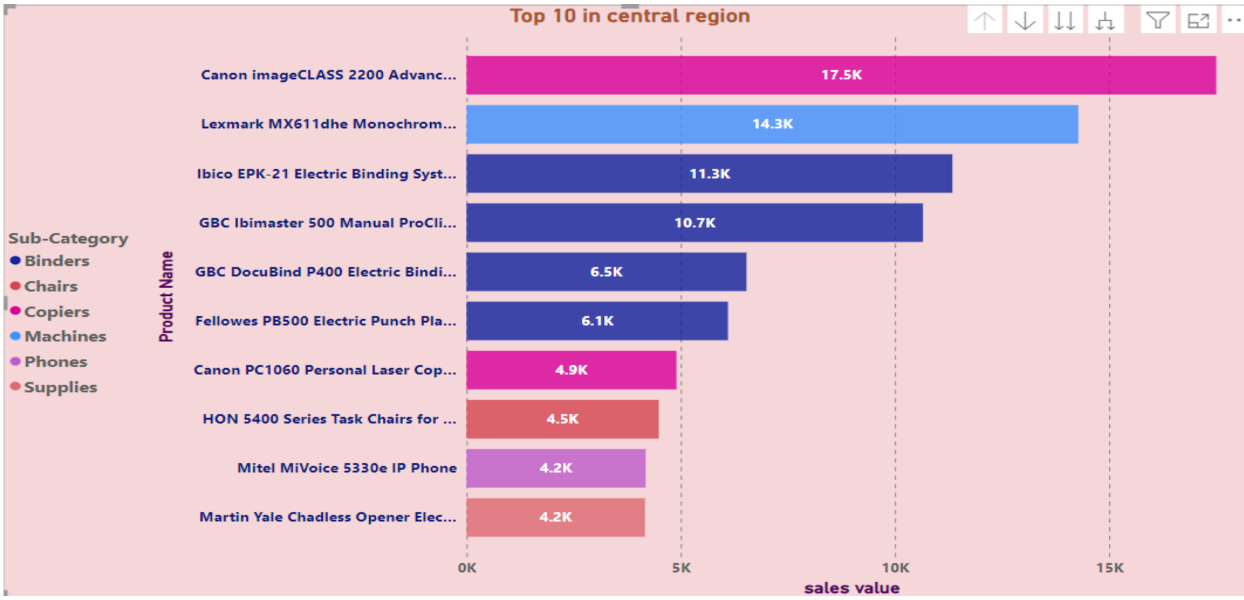


Figure 11: Top 10 products in central region

Binder products were found to be the least sold item in the central region whereas binders along with furniture items constitute to the top selling items. Copier categorical products are least sold which implies that could be more competitors who sell these items or people prefer to purchase from online platforms.

d.3. South region

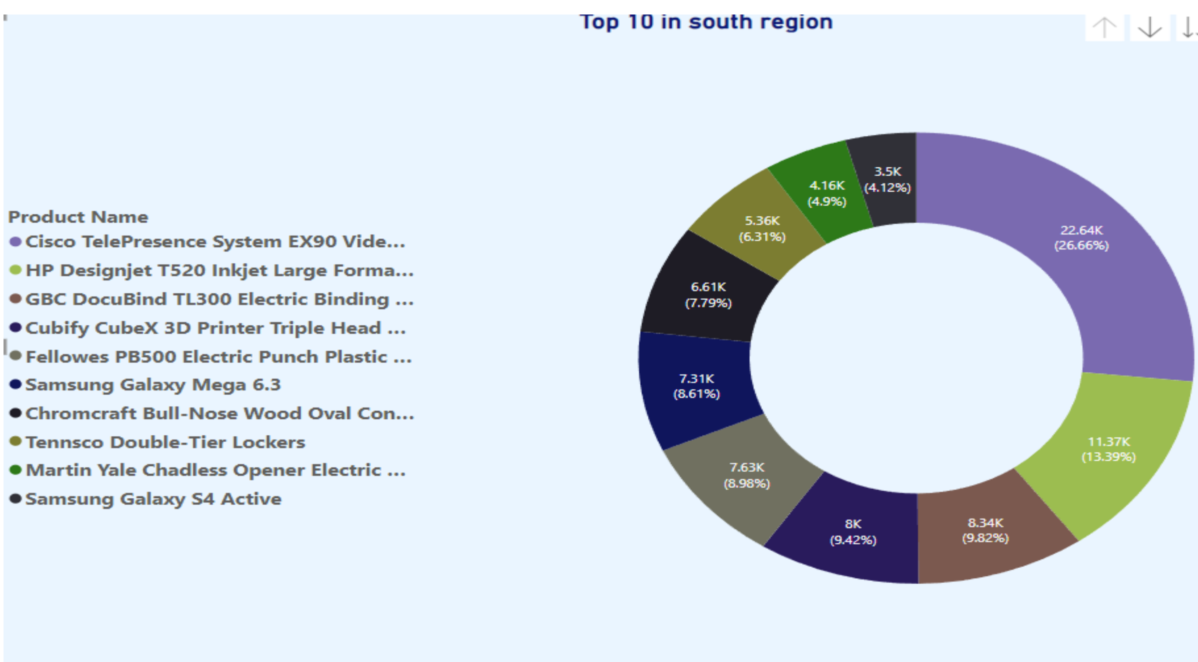


Figure 12: Doughnut chart for top 10 products in south region

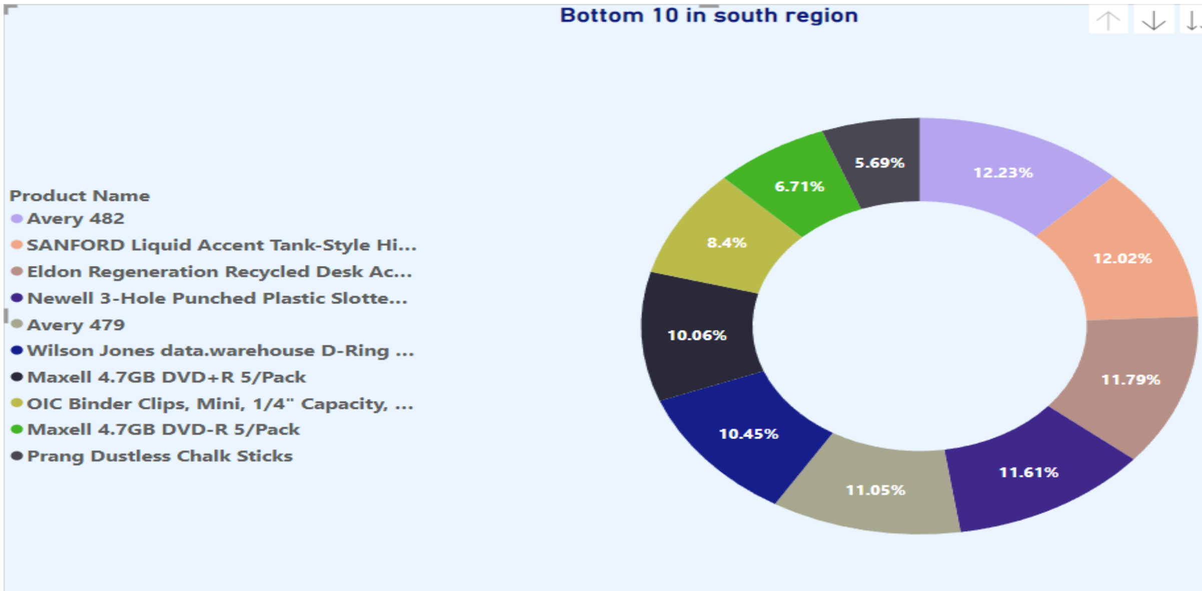


Figure 13: Doughnut chart for bottom 10 products in south region

The above donut figures the most and the least sold item in the south region. However, this chart clearly gives the percentage constitution of each product with clear visualization.

d.4. West region

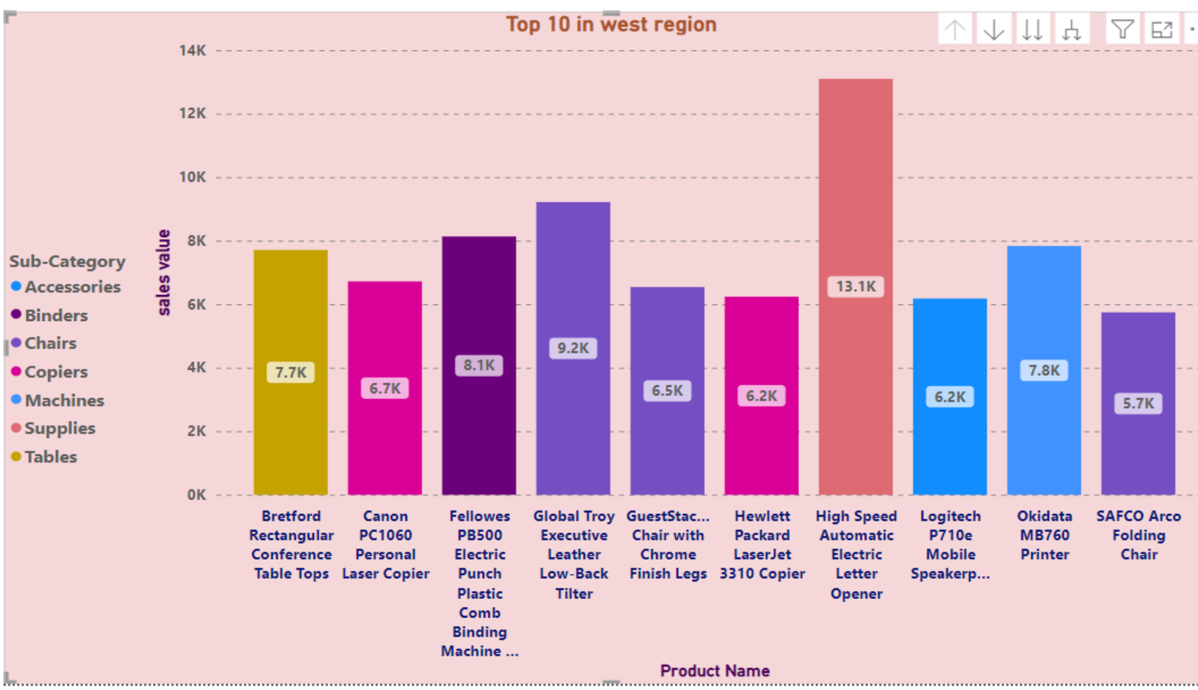


Figure 14: Top 10 products in west region

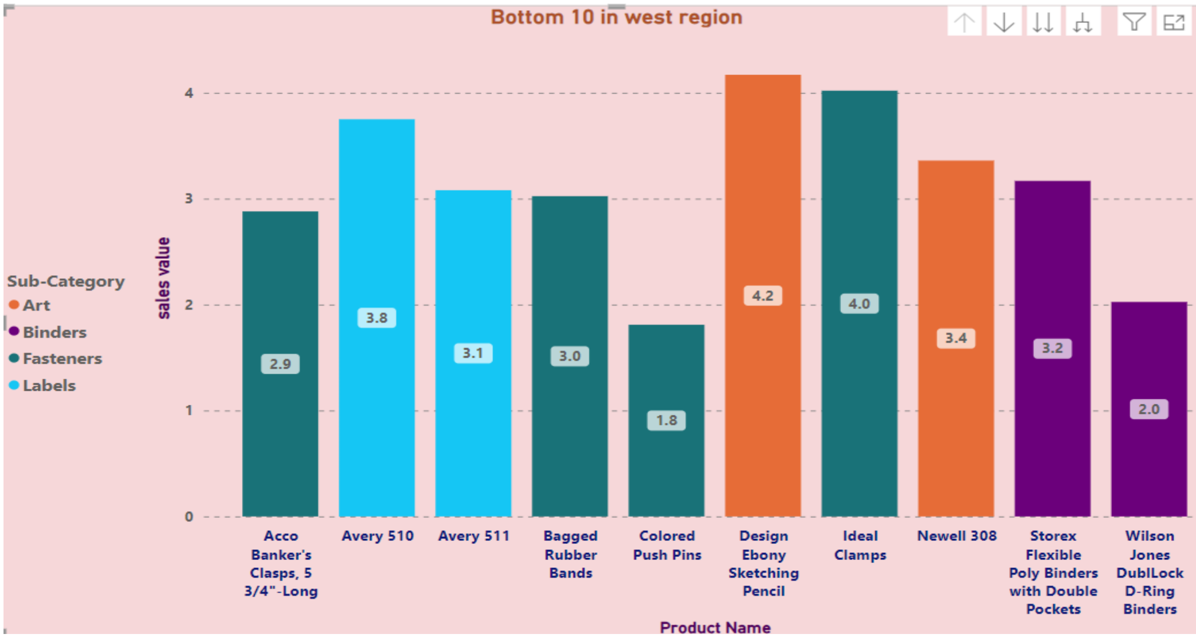


Figure 15:Bottom 10 products in west region

Inference

Different types of charts were used to visualize the top and bottom 10 products in each region as charts contribute for better understanding of data and to narrate data stories despite making analysis easy.

Using these tools, it would be useful to concentrate the sale of each product under respective category which in turn effects with the overall sales performance that accounts for profitability by identification of pattern in the purchase behavior of the customers(Jie et al., 2018; Jolly, 2023).

e. Visualization of total sales and profit for each year

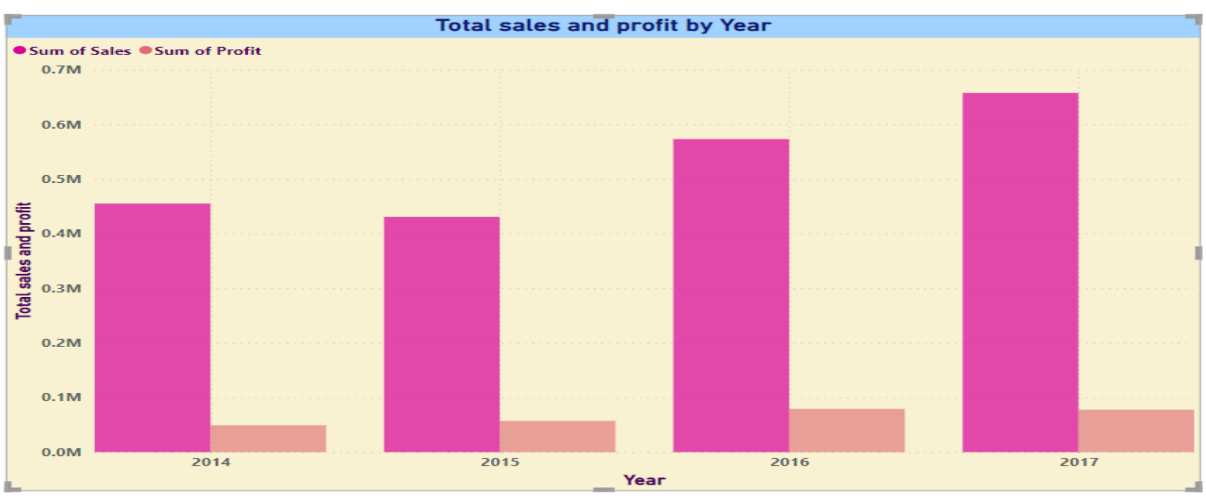


Figure 16: Clustered column chart for profit obtained with total sales each year

With the help of the above clustered column chart, it could be interpreted that the sales had been gradually increasing besides, the profit level was not obtained at an appreciated level. It could also be forecasted that the profit may retain the same level without improvement if sufficient actions were not taken to improve on the profitability of the store or else at a point of time, the increase in sales would dominate the profit level which would be diminished in the future(Harsoor and Patil, n.d.).

2. Visualization for question 2

2.1. Insight on relationship between each product category, sales and profit

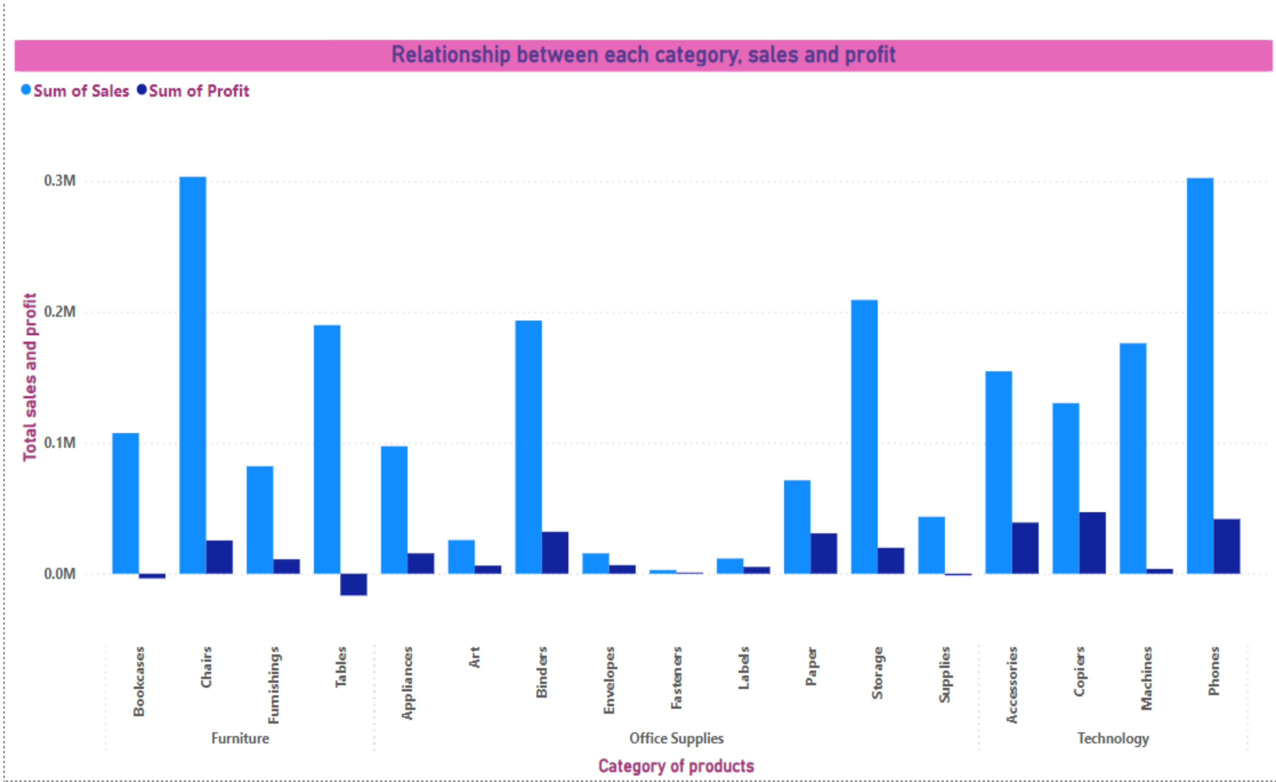


Figure 17: Relationship between each product category, sales and profit

The product from each category contributes to the total sales and profit. However, the products under office supplies are contributing less to the total profit. So, it could be inferred that quantity of each product sold is directly proportional to the total sales and profit(Al Rachmat et al., 2022).

2.2. Visualizing the trend of total sales in each region by using slicers

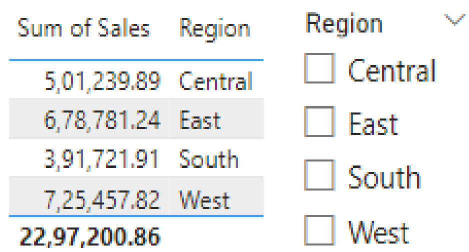
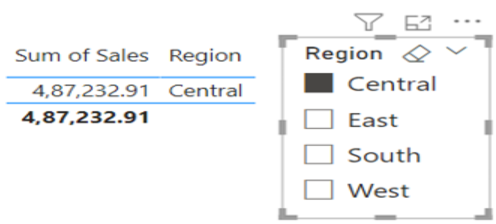
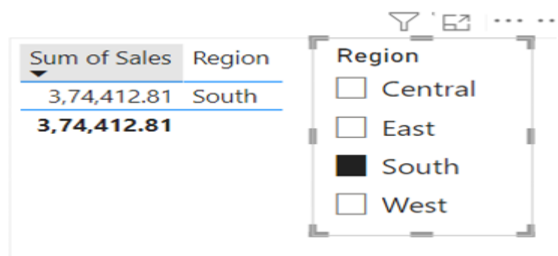


Figure 18: Slicer creation for sales in all regions

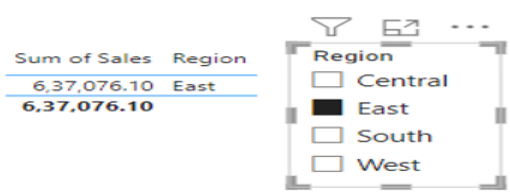
Central region:



South:



East:



West:

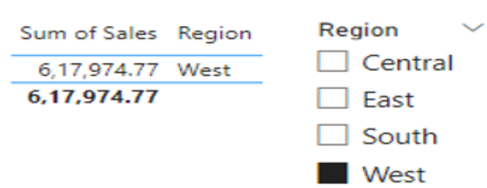


Figure 19: Separate slicer for all regions with sales

The east part of the country contributes for the maximum sales and the least being the south region. The segment of the customers might influence the sales and so choice of products to be sold in each region could be analysed before their introduction in the stores(Ighoteguonor, 2013) and this would increase the competitiveness also.

2.3. Visualize the total sales, profit, and profit rate for each product category and sub-categories

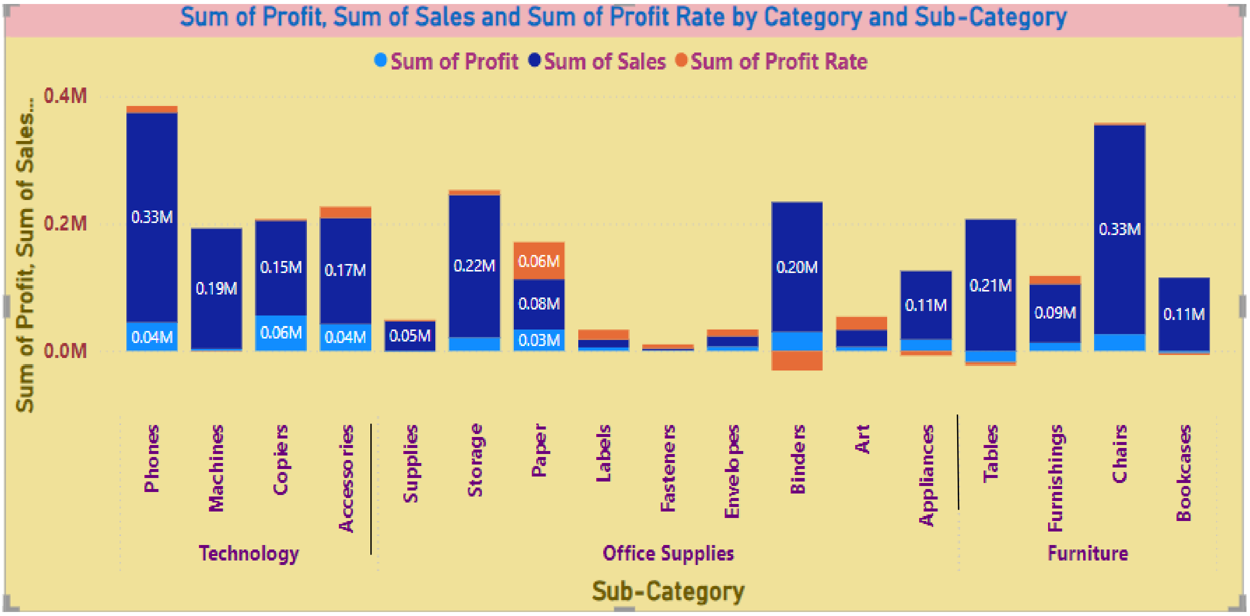


Figure 20: Visualization of total sales, profit, and profit rate

The chairs and phones contribute to the maximum profit rate on the other hand, the office supplies products gives either negative returns or meticulous profit. The sum of sales is inversely proportional to the profit obtained per product. Appropriate pricing of products has to be done to mitigate the future loss(Harsoor and Patil, n.d.).

2.4. Determination overall performance of each product with respect to customer segment

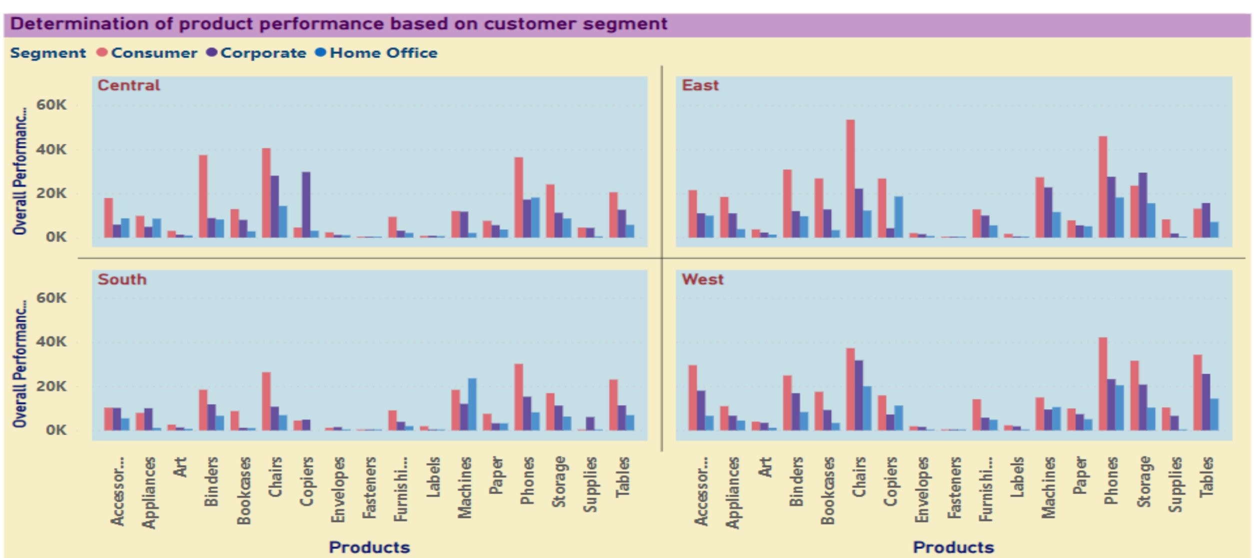


Figure 21: Performance of each product with respect to customer segment

Overall, the consumer segment of customers are major part of the sales and the products preferred by them are furniture and phones. Apparently, the East region has more consumers followed by corporate people. In the south, the population might also influence the sales of each product besides their choices. The products like envelopes, fasteners and labels are being sold in less quantity as people might not prefer to buy in any of the regions. So, they could be replaced by other products which might be sold in more numbers. A thorough analysis has to be demonstrated to find the attractive products that are likely to be preferred by all kinds of customers(Gautam and Kumar, 2022).

2.5. Contribution of each state towards the total sales in 2017

Contribution of each state towards sales in 2017

State	Sum of Sales
California	1,14,368.69
New York	84,239.26
Washington	47,614.09
Texas	42,432.35
Pennsylvania	40,195.12
Florida	26,430.38
Michigan	25,833.65
Illinois	24,188.12
North Carolina	23,456.83
Ohio	21,816.00
Georgia	19,157.24
Indiana	17,987.09
Kentucky	15,534.84
Delaware	13,317.05
Tennessee	11,905.61
Colorado	10,269.46
Arizona	10,095.74
New Jersey	9,483.55
Maryland	9,451.60
Missouri	9,350.80
Total	6,57,713.19

Figure 22: Creation of slicer for period

Central region:

Figure 23: Contribution of sales from central states in 2017

East:

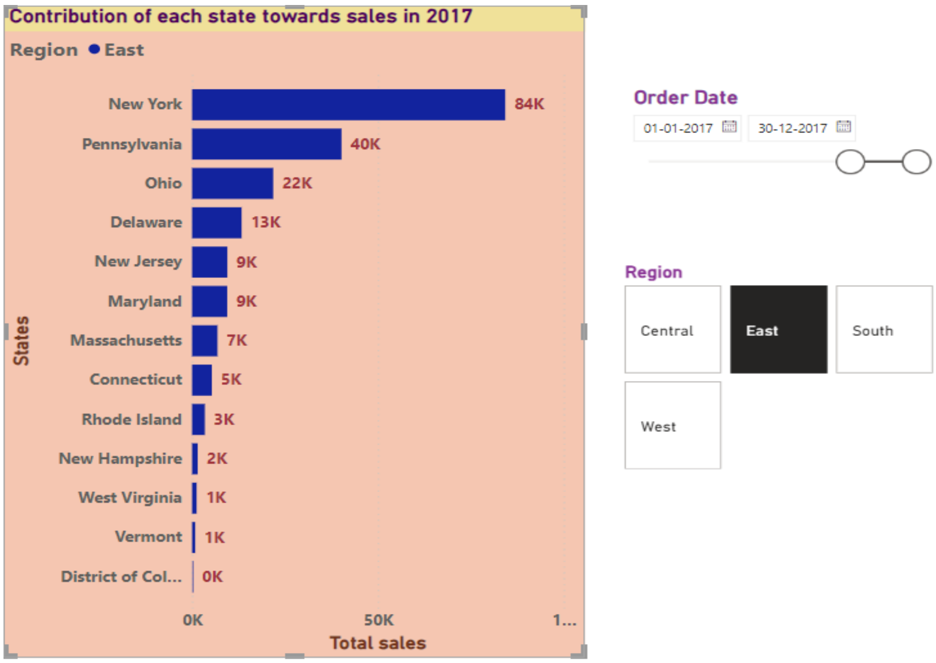


Figure 24: Contribution of sales from eastern states in 2017

West:

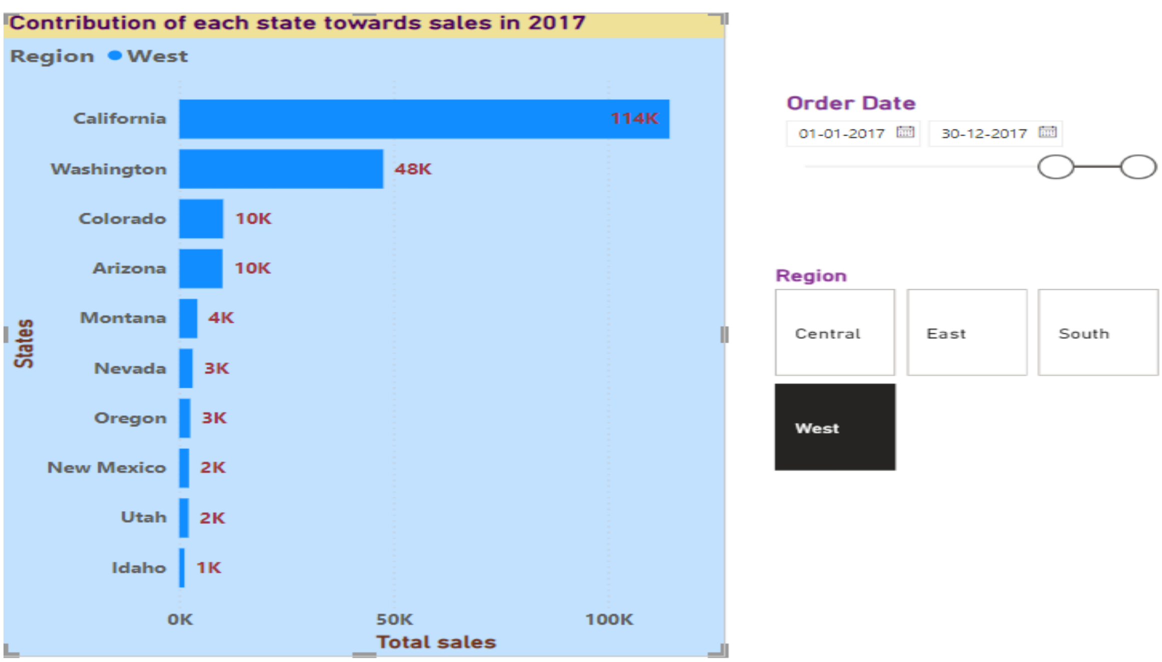


Figure 25:Contribution of sales from western states in 2017

South:

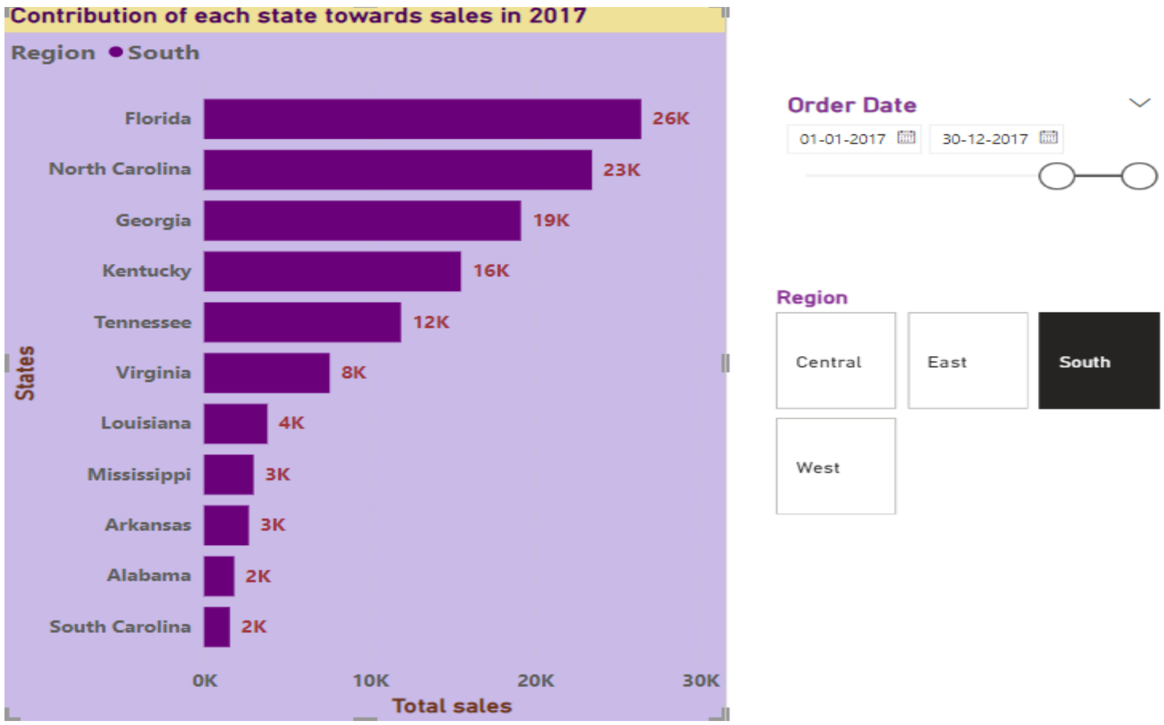


Figure 26:Contribution of sales from southern states in 2017

Region	State	Rank
Central	Texas	3
East	New York	2
West	California	1
South	Florida	4

Table 2:Ranking of regions with respect to sales in 2017

From the above table and figure, it could infer that central region contributes for the total sales. One of the reasons could be the density of population and also the segment of customers (from figure 21)who are professionals. The business development could be heavy concentrated in the central region(Hui, n.d.).

2.6. Compare the total sales of each city between 2014 to 2017

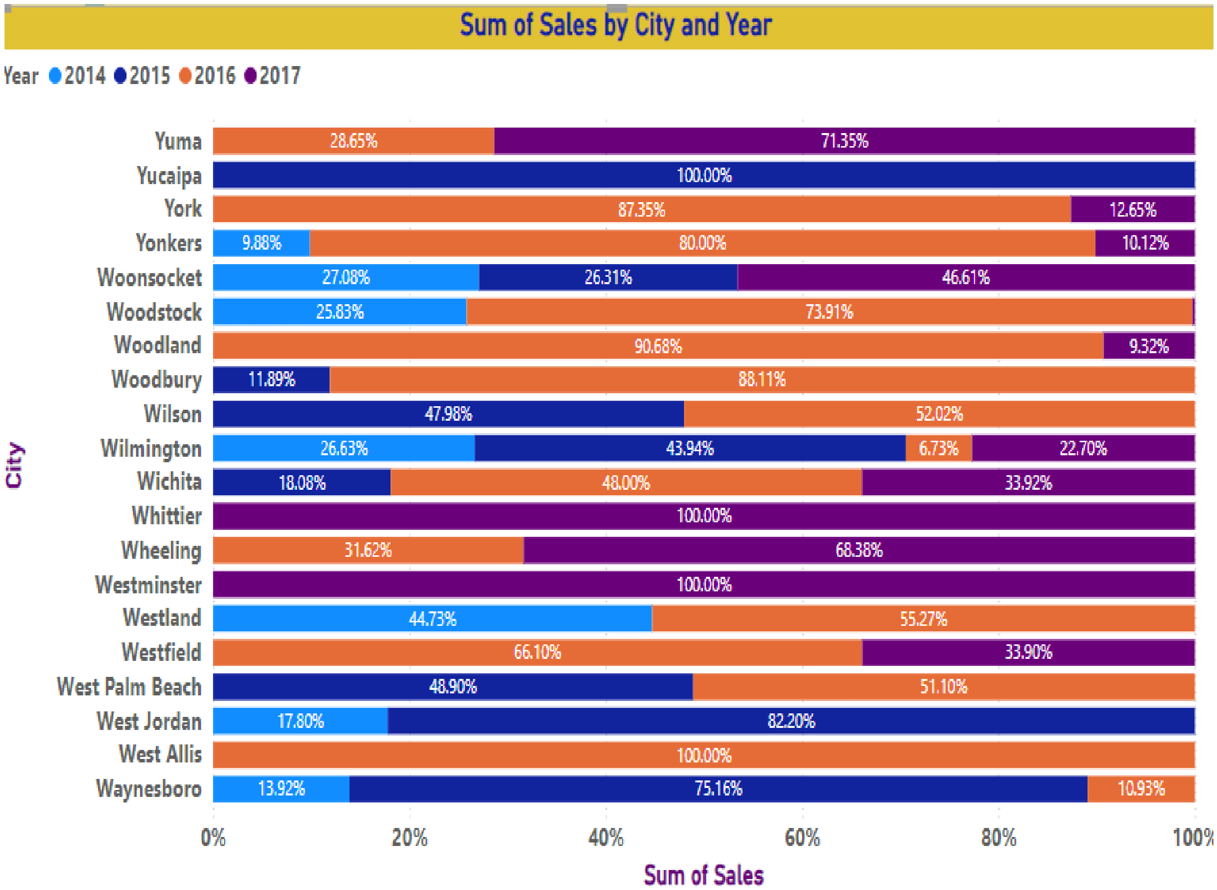


Figure 27:Sales in each city by year

There is no definite conclusion with respect to performance between cities but it is observed that many cities have their maximum sales in the period of 2016 and 2017. Certainly, it could be inferred that the performance of sales has been increasing annually with implementation of marketing strategies and upon complete analysis of existing market then and there(Jie et al., 2018).

2.7. Contribution of each state towards total sales of Alpha

Central:

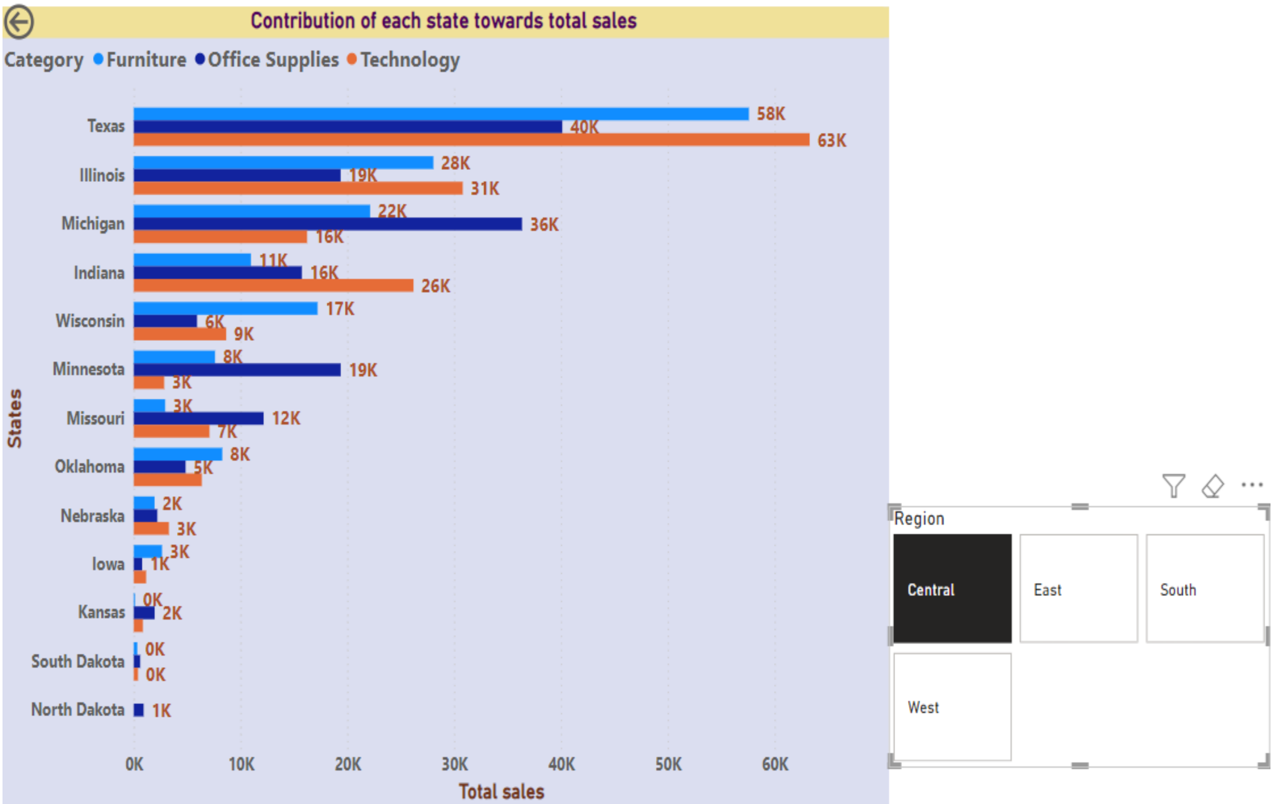


Figure 28: Contribution of central states towards total sales

East:

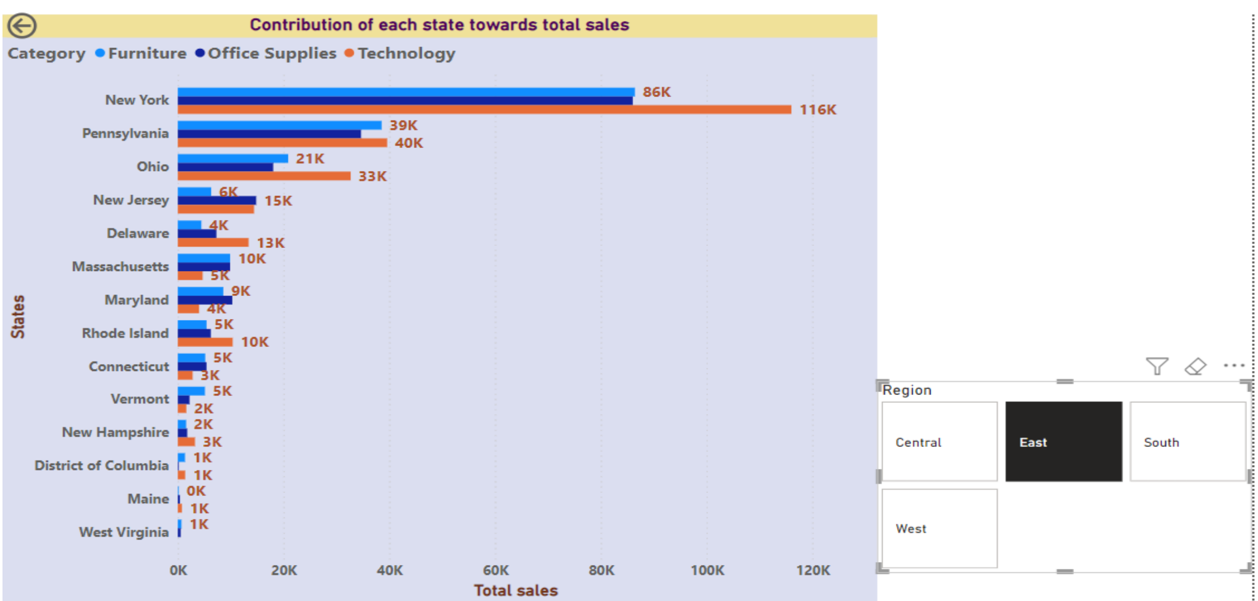


Figure 29:Contribution of eastern states towards total sales

West:

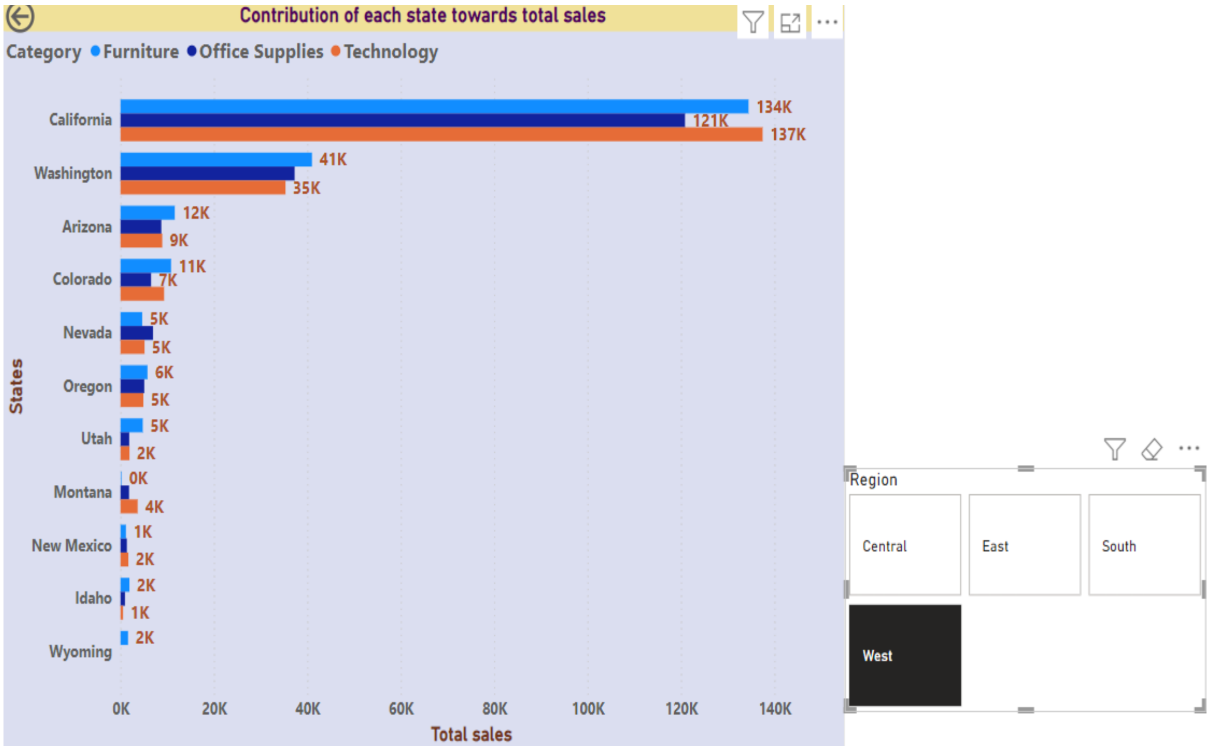


Figure 30:Contribution of western states towards total sales

South:

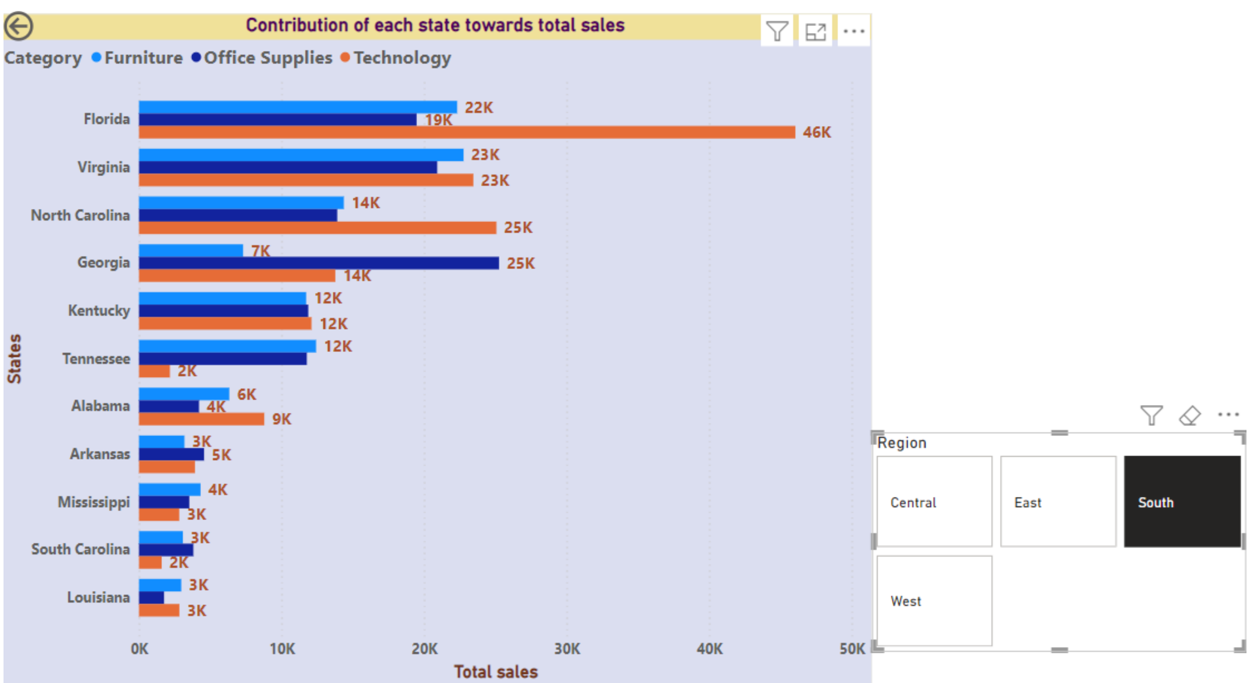


Figure 31:Contribution of southern states towards total sales

In all states, the technical products are found to be major contributor for Alpha sales. Office supply products are impacting the total sales of the store and so, replacement with another category of product would be much appreciated. This could increase customer base and also the competitiveness in the market(Wan et al., 2012).

2.8. Insights on relationship between each product and their quantity sold

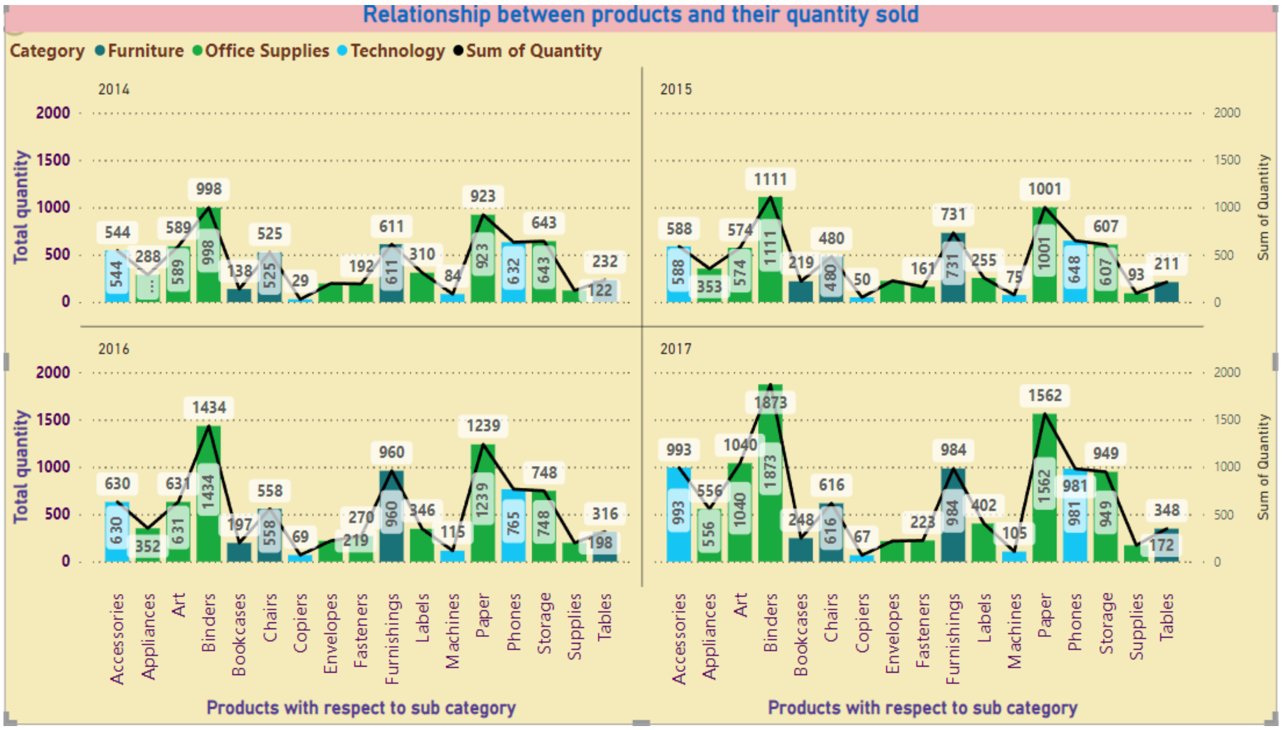


Figure 32:Relationship between each product and their quantity sold

The quantity of each product being sold by Alpha has gradually increased from 2014 to 2017. Still, some products are visualized as non performing ones in their sales volume, improvement could be made by introducing variety of office supplies. Hence, the sales could be increased by diversification of products in this category(Iskenderoglu, 2023).

2.9. Insights on relationship between customer segment and shipping mode

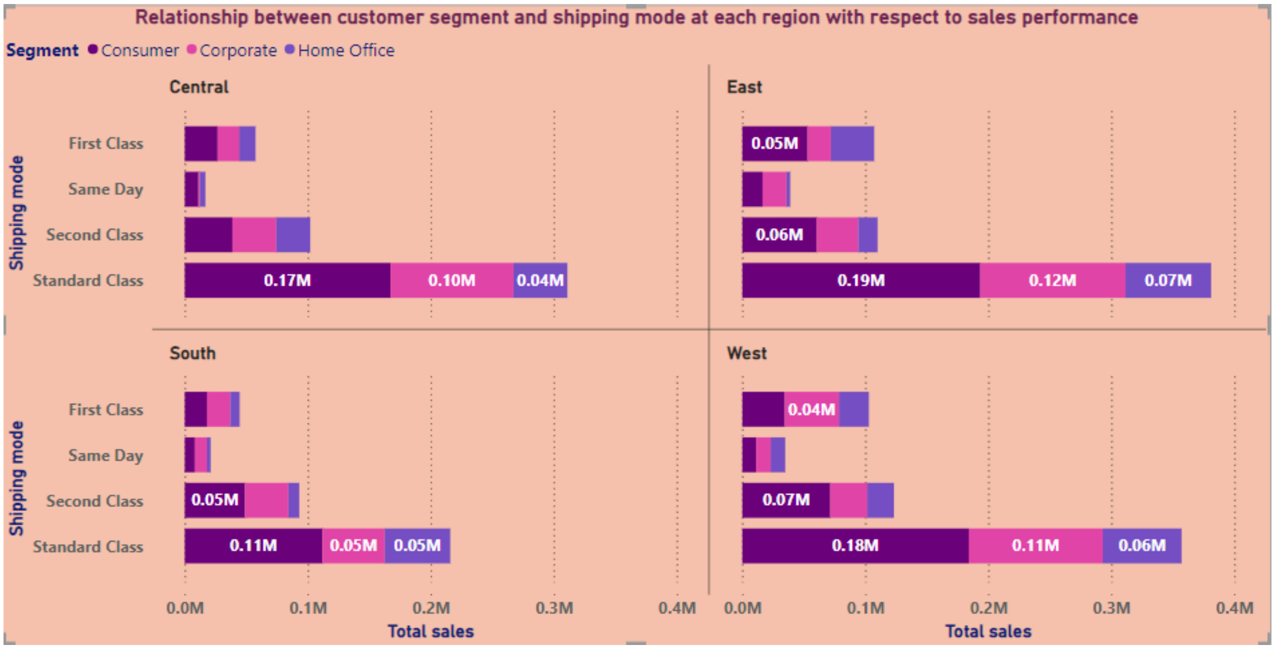


Figure 33:Relationship between customer segment and shipping mode

Most of the customers irrespective of the region, prefer to buy product through standard class which might take couple of days to be delivered. The consumer segment customers use this kind of shipping largely. Notably, the corporate people in West region opts for first class delivery. On an overall perspective, it could be inferred that the shipping charges could be heavily charged by the store for first class and same day delivery and this made the customers to choose standard mode(Wu et al., 2022).

2.10. Relationship between product category, sales and discount

Figure 34:Relationship between product category, sales and discount

Based on the earlier analysis and the above one, it could be justified that the store has given more discount on prices of office supplies to increase their sales. On the other hand, more discounts on technical products in order to attract customers. Attractive prices were not placed on furniture as the margin for each of them in this category might be less thereby affecting the total profit(Fasihah et al., 2020).

3. Dashboard for question 3

3.1. Alpha Store Sales Dashboard

Figure 35: Alpha Stores Sales Visualization

The above "Alpha Store Sales Visualization" dashboard provides a detailed sales performance from the year 2014-2017. Overall Total Sales of $2.30 million (2.30M) was during this period with total items sold being 5,009.

Through the ***regional sales bar graph*** from 2014-2017, a sales trend can be seen that the East and West regions have the highest sales figures, with notable peaks in 2017 (East: $0.21M; West: $0.16M). The

Central region shows a significant peak in 2016 at $0.15M, while the South region demonstrates a lower peak in 2015 at $0.10M.

- **Central:** $501,239.89
- **East:** $678,781.24
- **South:** $391,721.91
- **West:** $725,457.82

Through the ***yearly sales by category bar graph*** from 2014-2017depicts that the Technology has consistently outperformed the other categories with a peak of $0.23M in 2017. Furniture follows closely, with a peak of $0.20M in 2016, and Office Supplies maintains a steady sales performance, peaking at $0.18M in 2016.

Sales are further segmented into three distinct categories:

- **Consumer**
- **Corporate**
- **Home Office**

Consumer has the highest overall sales i.e. $1.16M followed by Corporate at $706.1K and Home Office being the lowest at 429K. Through the state wise bar graph, California from west region has provided the highest sales that being of ~$457K followed by New York (East) at ~$310K. The least performing state are North & South Dakota, Marine & West Virginia on regional strengths.

Key Insights

1. **Regional Performance:** With strong sales growth, especially in 2017, the East and West regions stand out as the best-performing locations. Despite their considerable contributions, the Central and South areas' aggregate sales are smaller than those of the East and West.
2. **Strengths of the Category:** Technology has the best sales performance, pointing to a robust product portfolio and significant market demand. Although they don't advance as quickly as technology, furniture and office supplies nevertheless make a substantial contribution, indicating room for improvement.
3. **State Contributions:** The significance of these markets is shown by the high sales contributions from states like Texas, New York, and California. Targeted marketing and sales tactics might further increase sales success in these areas.

Areas for Improvement

1. **Balanced Regional Growth:** Although the East and West areas do well, the Central and South regions require methods to increase sales. This could entail better distribution networks, locally tailored marketing campaigns, and product modifications in response to regional tastes.
2. **Segment Diversification:** At the moment, the consumer segment is the main focus. By include the Corporate and Home Office divisions in the research, more growth prospects may be found, and the sales portfolio may become more diversified.
3. **Improved Category Performance:** Furniture and office supplies show a lot of promise, even though technology is the dominating category. Sales in these categories could be improved by targeted marketing, innovative product development, and aggressive price strategies.

3.2.Alpha Store Return Analysis

Figure 36: Alpha Store Return Sales Analysis

Through the above dashboard, we can see that a total of ~$180k worth of items were returned with the count of orders being returned 296.

Through the ***regional return bar graph*** from 2014-2017, a sales trend shows that the East and West regions have the highest return figures,

1. **West: $107,483.06**
2. **East: $41,705.14**
3. **South: $17,309.10**
4. **Central: $14,006.98**

In Category, Technology has seen the highest number of returns,

1. **Technology:** $72,708.17
2. **Furniture:** $59,219.17
3. **Office Supplies:** $48,576.93

Recommendations to reduce returns

1. **High Returns in Key Regions:** Despite high sales, regions like the West, particularly California, also likely experience a high volume of returns given their sales contribution. Through this, we can imply that the number of returns is directly proportional to the sales. Perform Customer surveys and analyze the feedback to improve the product quality.
2. **Improve Customer Support:** Providing customers with how-to-use guides and customer support can help in easy adoption.
3. **Changes in Return Policy:** Having a strict return policy can help reduce the redundancy of orders track the customers with high returns and issue a warning to reduce the unnecessary return numbers.
4. **Tailor made product:** Make more tailor made product after analyzing the segment , product sold and the state as every region might have a different geo graphical conditions which can lead to different customer needs

REFERENCES

Al Rachmat, R., Sartika Pratiwi, T., Ekonomi Manajemen Akuntansi Dan Keuangan, J., 2022. JANUARI 2021 page: 121-128| 121 Sales Volume, Operating Cost and its Effect on Profitability (Study on Listed Companies in The Indonesia Stock Exchange (IDX) for 2017-2020) Volume Penjualan, Biaya Operasional dan Pengaruhnya terhadap Profitabilitas (Studi pada Perusahaan terdaftar di Bursa Efek Indonesia tahun 2017-2020) Padriyansyah 3, 2017–2020. https://doi.org/10.53697/emak.v3i1

Ali, Q., Yaacob, H., Parveen, S., Zaini, · Zaki, 2021. Big data and predictive analytics to optimise social and environmental performance of Islamic banks 41, 616–632. https://doi.org/10.1007/s10669-021-09823-1

Almatrodi, I., Li, F., Alojail, M., 2023. Organizational Resistance to Automation Success: How Status Quo Bias Influences Organizational Resistance to an Automated Workflow System in a Public Organization. Systems 11. https://doi.org/10.3390/systems11040191

Alonso, R., Câmara, O., 2024. Organizing Data Analytics. Manage Sci 70, 3123–3143. https://doi.org/10.1287/mnsc.2023.00207

Alonso, R., Câmara, O., 2023. Organizing Data Analytics. https://doi.org/10.1287/mnsc.2023.00207. https://doi.org/10.1287/MNSC.2023.00207

Aspin, A., 2022. Charts in Power BI Desktop, in: Pro Power BI Dashboard Creation. Apress, Berkeley, CA, pp. 115–138. https://doi.org/10.1007/978-1-4842-8227-4_6

Batini, C., Cappiello, C., Francalanci, C., Maurino, A., 2009. Methodologies for data quality assessment and improvement. ACM Comput Surv 41. https://doi.org/10.1145/1541880.1541883

Batko, K., Ślęzak, A., n.d. The use of Big Data Analytics in healthcare. https://doi.org/10.1186/s40537-021-00553-4

Ben-Zeev, A., 1981. J.J. Gibson and the ecological approach to perception. Stud Hist Philos Sci 12, 107–139. https://doi.org/10.1016/0039-3681(81)90016-9

Birken, S.A., Bunger, A.C., Powell, B.J., Turner, K., Clary, A.S., Klaman, S.L., Yu, Y., Whitaker, D.J., Self, S.R., Rostad, W.L., Chatham, J.R.S., Kirk, M.A., Shea, C.M., Haines, E., Weiner, B.J., 2017. Organizational theory for dissemination and implementation research. Implementation Science 12. https://doi.org/10.1186/s13012-017-0592-x

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data. Business & Information Systems Engineering, 5(2), 65-69.

Cato, P., Gölzer, P. and Demmelhuber, W., 2015, November. An investigation into the implementation factors affecting the success of big data systems. In *2015 11th International Conference on Innovations in Information Technology (IIT)* (pp. 134-139). IEEE.

Chaffey, D., & White, G. (2010). Business information management: improving performance using information systems. Pearson Education.

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business Review Press.

Dent, E. B., & Goldberg, S. G. (1999). Challenging "resistance to change". The Journal of Applied Behavioral Science, 35(1), 25-41.

Dorr, B.J., Greenberg, C.S., Fontana, P.C., Przybocki, M.A., Le Bras, M., Ploehn, C.A., Aulov, O., Michel, M., Golden, E.J. and Chang, W.L., 2015, October. The nist iad data science research program. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics* (pp. 1-10).

Dietz, J.L.G., Mulder, H.B.F., n.d. Enterprise Ontology A Human-Centric Approach to Understanding the Essence of Organisation.

Eastman, N., Adshead, G., Fox, S., Latham, R., Whyte, S., 2023. Decision-making theories, in: Eastman, N., Adshead, G., Fox, S., Latham, R., Whyte, S. (Eds.), Oxford Casebook of Forensic Psychiatry. Oxford University PressOxford, pp. 3–34. https://doi.org/10.1093/med/9780198842057.003.0001

Edwards, J.S., Taborda, E.R., 2016. Using knowledge management to give context to analytics and big data and reduce strategic risk, in: Procedia Computer Science. Elsevier B.V., pp. 36–49. https://doi.org/10.1016/j.procs.2016.09.099

English, L. P. (1999). Improving data warehouse and business information quality: methods for reducing costs and increasing profits. John Wiley & Sons.

Farooq Aziz, 2023. Data analytics impacts in the field of accounting. World Journal of Advanced Research and Reviews 18, 946–951. https://doi.org/10.30574/wjarr.2023.18.2.0863

Freeman, R. E. (2010). Strategic management: A stakeholder approach. Cambridge University Press.

Gandomi, A., & Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.

Fasihah, N., Jamaluddin, B., Siti, &, Esa, A., 2020. Effect of Price on Sales Volume, Journal of

Undergraduate Social Science and Technology.

Gartner. (2019). The Chief Data Officer's Guide to Upholding Data Ethics. Retrieved from https://www.gartner.com/en/documents/3906483/the-chief-data-officer-s-guide-to-upholding-data-ethics

Hammer, M., & Champy, J. (1993). Reengineering the corporation: A manifesto for business revolution. HarperCollins.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. International Journal of Production Economics, 154, 72-80.

Hawley, D., 2016. Implementing business analytics within the supply chain: Success and fault factors. *Electronic Journal of Information Systems Evaluation*, *19*(2), pp.pp112-120.

Hajiheydari, N., Delgosha, M.S., Wang, Y., Olya, H., n.d. Exploring the paths to big data analytics implementation success in banking and financial service: an integrated approach. https://doi.org/10.1108/IMDS-04-2021-0209

Harsoor, A.S., Patil, A., n.d. FORECAST OF SALES OF WALMART STORE USING BIG DATA APPLICATIONS, IJRET: International Journal of Research in Engineering and Technology.

Hasan, M.M., Popp, J., Oláh, J., 2020. Current landscape and influence of big data on finance. J Big Data 7. https://doi.org/10.1186/s40537-020-00291-z

Hui, Z., n.d. Experiment With Multiple Regression Models for Sales Forecast and Location Analysis. Available from: 10.32920/23296004.v1. 2023.

Ighoteguonor, Godfrey., A.Ngozi., U.B., 2013. The Effect of Market Segmentation on Global Marketing: A Conceptual Overview.

Inmon, W. H. (2005). Building the data warehouse. John Wiley & Sons.

Iskenderoglu, C., 2023. Product market competition and the value of diversification. Financ Res Lett 58, 104049. https://doi.org/10.1016/j.frl.2023.104049

Jie, L.Y., Wong, D.H. Ten, Zain, Z.M., Sjarif, N.N.A., Ibrahim, R., Maarop, N., 2018. Metrics and benchmarks for empirical and comprehension focused visualization research in the sales domain. Indonesian Journal of Electrical Engineering and Computer Science 12, 1340–1348.

https://doi.org/10.11591/ijeecs.v12.i3.pp1340-1348

Jolly, J., 2023. Identify Patterns and Trends. pp. 275–304. https://doi.org/10.1007/978-1-4842-9013-2_11

Jouti, A.T., n.d. An integrated approach for building sustainable Islamic social finance ecosystems. https://doi.org/10.1108/IJIF-10-2018-0118

Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148-152.

Kim, W., Choi, B. J., Hong, E. K., Kim, S. K., & Lee, D. (2003). A taxonomy of dirty data. Data Mining and Knowledge Discovery, 7(1), 81-99.

Kotter, J. P. (1996). Leading change. Harvard Business Review Press.

Karin, Hartl., O.J., 2016. The Role of Data Quality in Business Intelligence - An empirical study in German medium-sized and large companies. 33-42. . 2016 33–42.

Kwon, O., Lee, N. and Shin, B., 2014. Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, *34*(3), pp.387-394.

Kubina, M., Koman, G., Kubinova, I., 2015. Possibility of Improving Efficiency within Business Intelligence Systems in Companies. Procedia Economics and Finance 26, 300–305. https://doi.org/10.1016/s2212-5671(15)00856-4

Kumar, P., 2018. Impact of Data Analytics in Changing the Future of Business and Challenges Facing It. International Journal of Scientific and Research Publications (IJSRP) 8. https://doi.org/10.29322/ijsrp.8.5.2018.p7711

Laukkanen, E., Itkonen, J., Lassenius, C., 2017. Problems, causes and solutions when adopting continuous delivery—A systematic literature review. Inf Softw Technol. https://doi.org/10.1016/j.infsof.2016.10.001

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52(2), 21.

Liu, Y., Chen, X., 2022. Application of Big Data Analysis Based on Power BI in Sales Forecasts, in: ACM International Conference Proceeding Series. Association for Computing Machinery, pp. 722–726. https://doi.org/10.1145/3569966.3571272

Liu, Y., John, S., Han, H., Debello, J.E., n.d. The Challenges of Business Analytics: Successes and Failures.

Marshall, A., Mueck, S. and Shockley, R., 2015. How leading organizations use big data and analytics to innovate. *Strategy & Leadership*, *43*(5), pp.32-39.

Nandish V. Patel, 2011. Deferred action: Theoretical model of process architecture design for emergent business processes. Operations Management: A Modern Approach.

Otto, B. (2011). A morphology of the organisation of data governance. ECIS 2011 Proceedings, 270.

Popovič, A., Hackney, R., Tassabehji, R. and Castelli, M., 2018. The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, *20*, pp.209-222.

Pansara, R., 2023. Cultivating Data Quality to Strategies, Challenges, and Impact on Decision-Making. *International Journal of Managment Education for Sustainable Development*, *6*(6), pp.24-33.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin, 23(4), 3-13.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. Communications of the ACM, 41(2), 79-82.

Redman, T. C. (2018). *Data Driven: Creating a Data Culture*. Harvard Business Press.

Rivera, G., Cox, A., 2014. An evaluation of the practice based approach to understanding the adoption and use of information systems. Journal of Documentation 70, 878–901. https://doi.org/10.1108/JD-06-2013-0080

Safitri, J., Geraldina, I., 2023. The Implementation Of Banking Risk Management In ASEAN Countries.

Schein, E. H. (2010). Organizational culture and leadership (Vol. 2). John Wiley & Sons.

Singh, M.S., Jadhav, L., 2022. Data Analysis and Visualization of Sales Dataset using Power BI 10. https://doi.org/10.22214/ijraset.2022.44132

Sidi, F., Panahy, P.H.S., Affendey, L.S., Jabar, M.A., Ibrahim, H. and Mustapha, A., 2012, March. Data quality: A survey of data quality dimensions. In *2012 International Conference on Information Retrieval & Knowledge Management* (pp. 300-304). IEEE.

Stohr, E.A., Howe, W.J., Zhao, J.L., 2001. Workflow Automation: Overview and Research Issues, Information Systems Frontiers. Kluwer Academic Publishers.

Surbakti, F.P.S., Wang, W., Indulska, M. and Sadiq, S., 2020. Factors influencing effective use of big data: A research framework. *Information & Management*, *57*(1), p.103146.

Su, X., Zeng, W., Zheng, M., Jiang, X., Lin, W., Xu, A., 2022. Big data analytics capabilities and organizational performance: the mediating effect of dual innovations. European Journal of Innovation Management 25, 1142–1160. https://doi.org/10.1108/EJIM-10-2020-0431

Svenson, F., Peuser, M., Çetin, F., Colecraft Aidoo, D., Launer, M.A., 2024. Decision-making styles and trust across farmers and bankers: Global survey results. Decision Analytics Journal 10, 100427. https://doi.org/10.1016/j.dajour.2024.100427

Tan, Y., Tsionas, M.G., 2022. Modelling sustainability efficiency in banking. International Journal of Finance and Economics 27, 3754–3772. https://doi.org/10.1002/ijfe.2349

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP (pp. 14-21).

Valera, I., Gomez-Rodriguez, M., 2018. Enhancing the Accuracy and Fairness of Human Decision Making.

Watson, H. J. (2012). The business case for analytics. BizEd, 11(3), 48-54.

Wan, X., Evers, P.T., Dresner, M.E., 2012. Too much of a good thing: The impact of product variety on operations and sales performance. Journal of Operations Management 30, 316–324. https://doi.org/10.1016/J.JOM.2011.12.002

Wu, Y., Lu, Y., Huang, S., 2022. Impacts of Delivery Charge on the Possibility of Consumers Using Online Food Delivery. Sustainability (Switzerland) 14. https://doi.org/10.3390/SU14031795

Yury, Chernov. (2016), 2016. Test-Data Quality as a Success Factor for End-to-End Testing,An Approach to Formalisation and Evaluation . Chernov, YIn Proceedings of the 5th International Conference on Data Management Technologies and Applications - DATA; ISBN 978-989-758-193-9; ISSN 2184-285X, SciTePress,. DOI: 10.5220/0005971700950101 95–101.

Zayas-Cabán, T., Okubo, T.H., Posnack, S., 2023. Priorities to accelerate workflow automation in health care. Journal of the American Medical Informatics Association. https://doi.org/10.1093/jamia/ocac197

Zhang, J., Zhang, Z., & Yang, Y. (2017). Machine learning in big data analytics. In Big Data Mining and Analytics (pp.

Zhang, J., Yu, F.R., Wang, S., Huang, T., Liu, Z., Liu, Y., 2018. Load balancing in data center networks: A survey. IEEE Communications Surveys and Tutorials 20, 2324–2325.

https://doi.org/10.1109/COMST.2018.2816042

Zhu, X., Song, B., Ni, Y., Ren, Y., Li, R., Zhu, X., Song, B., Ni, Y., Ren, Y. and Li, R., 2016. Big Data—From Raw Data to Big Data. *Business Trends in the Digital Era: Evolution of Theories and Applications*, pp.1-22.