TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS No.: M-131-MSTIM-2020-2023

A Study on usages of visualization tools to remove biases in decision making

by

Ishwor Khanal

A THESIS

SUBMITTED TO THE DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING IN PARTIAL FULFILLMENT OF REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN ENGINEERING IN TECHNOLOGY AND INNOVATION MANAGEMENT

DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING

LALITPUR, NEPAL

OCTOBER, 2023

COPYRIGHT

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis entitled "**A Study on usages of visualization tools to remove biases in decision making**" submitted by Ishwor Khanal in partial fulfilment of the requirements for the degree of Master of Science in Technology and Innovation Management.

Supervisor, Mr. Rajesh Kaji Kayastha
Associate Professor
Department of Mechanical and
Aerospace Engineering

External Examiner, Mr. Bibek Ropakheti
Associate Professor
Department Head ICT
Cosmos College

Committee Chairperson, Dr. Sudip Bhattarai
Head of Department
Department of Mechanical
and Aerospace Engineering

Date: October 13, 2023

ABSTRACT

Data visualization is important tool in modern-day decision-making process. It is used widely across industries to make effective and supporting proof to solve problems. The objective of this work is to study usage of Power BI tool to remove biasness in decision making. This is because Power BI tool is one of the easily available tools and widely accepted across the industry for reporting purpose. By use of Power BI tool, common type of biasness is highlighted. And respective solution and causes of manipulation are marked. Also, with the use of Q&A feature within the application, ease of creating bias free visualization is shown. This can help business user without any experience in visualization can generate desired chart. By excluding the cases of biasness highlighted and using tools effectively then decision making can be error free. And training of using this tool can add value to the managers or decision maker. Along with this, a python-based visualization model is created which helps to create instant charts without any knowledge of charts technicality.

Keywords: Visualization, Decision making, Manipulation, Power BI, Charts, Effective, Accuracy, Python, Matplotlib, Pandas

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and appreciation to my supervisor, Mr. Rajesh Kaji Kayastha for his genuine supervision throughout the preceding of the thesis. Likewise, I would like to express my gratitude to Dr. Sanjeev Maharjan, Program coordinator, MSc in Technology and Innovation Management for his encouragement before and during my thesis period. I am thankful to Head of Department and faculties of Mechanical and Aerospace Engineering, for providing me with opportunity to conduct this thesis.

Lastly, I would like to thank Er. Pravat Shrestha, Er. Sujan Kharel, from Strategy and Planning team at Daraz Nepal, Er. Rajan Sharma, Coordinator Mechatronics department at Balaju School of Engineering and Technology, my classmates of the MSTIM 076 batch and my family members for their endless support and encouragement.

Inadvertently if any inclusion for the gratitude is forgotten, I would like to express my sincere gratitude to all those who have directly or indirectly assisted me with my thesis work.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

CEO	Chief executive officer
GIS	Geographic information system
VR	Virtual Reality
AR	Augmented Reality
BI	Business Intelligence
MS	Microsoft
MOM	Month-over-month
MV	Market value
ML	Machine Learning
Q&A	Question and Answer
CSV	Comma Separated Value
GIS	Geographic Information System

CHAPTER ONE: INTRODUCTION

1.1 Background

Visualization is the process of representing data or information in a graphical or pictorial format. It is often used to help people understand complex data sets or to make decisions. There is a long history of using visualization for decision making. One of the earliest examples is the use of maps to help military leaders plan battles. In the 19th century, Charles Minard created a famous map that showed the progress of Napoleon's army during the Russian campaign. This map is still used today as an example of effective data visualization. In the 20th century, visualization became increasingly important as computers made it possible to create more complex and sophisticated graphics. Today, visualization is used in a wide range of fields, including business, science, and medicine. There are many benefits to using visualization for decision making. Visualization can help people to:

- See patterns and trends that would be difficult to see in a text-based report
- Identify relationships between different data sets
- Understand complex information more quickly and easily
- Make better decisions

However, there are also some challenges associated with using visualization for decision making. One challenge is that visualization can be difficult to create and interpret. Another challenge is that visualization can be biased, which can lead to poor decision making.

Despite these challenges, visualization is a powerful tool that can be used to improve decision making. When used effectively, visualization can help people to see the big picture, identify problems, and make better decisions. Besides this, visualization is sometimes unable to show a picture of what is intended. Much research is carried out to highlight the importance of visualization in effective decision making. However, there is still no research available for how this can be used by mitigating risk of framing bias and improving risk understanding in strategic decision making.

Thus, this research paper act to highlight cases of biasness occurring in the industry and tries to provide guidance in removing such biasness. With the use of Power BI tool, the solution for removing biasness is summarized. Also, a python-based model is created which can help to resolve cases of potential biasness in the visual reporting.

1.2 Statement of Problem

Visualization is a powerful tool that can be used to communicate information and make decisions. However, it can also be used to introduce bias into decision making. This can happen when the way that data is visualized leads people to make decisions that are not in their best interests.

It is important to be aware of the potential for bias in visualization and to take steps to mitigate it. This can be done by carefully designing visualizations and by making sure that they are accurate and unbiased. These issues can be mitigated by opting out scenarios and considering ways to minimize biases in the visualization. This can be done by considering potential bias already while presenting a data report. And getting instant feedback from a variety of people and making changes into visualization as needed will help to avoid misleading and deceptive techniques.

With understanding of manipulations and distortion used by report designer, potential biases can be reduced. And individual or company involved in decision making can make proper necessary judgement.

Along with the research for highlighting impact of the visualization, a python-based model is also prepared which can help in visualization of different data type across the industry.

1.3 Research Objective

1.3.1 Main Objective

The main objective is to study usages of visualization tools to remove biases in decision making using Power BI tool.

1.3.2 Specific Objectives

1. To examine the biases issue faced while making decisions at a strategic level.
2. To study current trend use of visualization tools for decision making.
3. To analyze potential factors for removing biases and tendency to use while making decisions.
4. To develop python based model for visualization with use of Matplotlib.

1.4 Scope and Limitation of study

The use of visualization tools is increasing at a rapid pace. Not all professionals use every tool available in the market for visualization. Thus, this study output cannot hold true for

all of the use cases. Also, for this research Power BI tool is used for visual analysis. Since the Power BI desktop tool is free compared to Tableau.

Some of the limitation within this study are as follows:

- Study is based on the literature review and visualization manipulations with Power BI tool.
- 2-Dimensional analysis is only considered, as 3-Dimensional data for analysis was not needed.
- Study of changes in actual practice with result of the research.

1.5 Research Gap

Several researches are made to validate the point that visualization is an important parameter for decision making.

One of the latest research papers published on 25 August, 2021 by Dr.Karin Eberhard, entitled "The effects of visualization on judgment and decision-making: a systematic literature review". This study brings together information from different areas of social and information sciences to provide valuable insights. It subtly suggests how these insights can be applied to improve decision-making in management. She has pointed positive, negative as well as neutral conclusions from analysis of more than 170 research papers (Eberhard, 2023). Key outcome of this research can be illustrated as:

Information visualization improves decision accuracy and quality, Information visualization steers attention towards uncertainty, Information Visualization Speeds Up Cognitive Processing etc. Also, this research has highlighted negative effect as, it can always not be helpful as it can have risk of misguides. Also, some time visualization can impact in decision making by overconfidence of decision maker.

Thus, how to answer the research gap of knowing how can visualization help to mitigate framing bias needed to be resolved. Also, there was concern how visual aids can influence overconfidence in managerial decision-making. This was to be resolved by highlighting cases of framing in visual so that when removing these criteria visual output can be free from bias.

Research made to analyze risk of generative AI based report also suggest similar information's. The visualization created automatically of images and text raises questions about the role of human agency in the production of meaning in a graphical context, creating tensions in professional, legal and artistic contexts (Schetinger, et al., 2023).

With more human understanding towards the output context and by knowing importance of chart and information, the artistic context is preserved as well as the biases can also be controlled.

Here our research is supposed to provide valuable input and output manipulation types. With the output decision maker can already know about the types of manipulation there can be in the charts. Also, they can guide subordinate about the importance of changing parameters in changing elements within the visual reports.

CHAPTER TWO: LITERATURE REVIEW

The main objective of this literature review is to ascertain visualization history, how it is being used in the modern world, what benefits it has brought in the decision-making aspect and finally answer the question of removing biases while making decisions. During the study, the sources of information are Journals, Articles, Books, Thesis and dissertations, and Internet publications. Literature review is based on any studies related to the data visualization, its implementation, challenges, usage and effectiveness to decision making.

2.1 Evolution of data visualization

The early history of visualization can be traced to the 17th century. All pre historic visualization are recorded Pre-17th Century. The period is quoted as Early Maps and Diagrams era (Friendly, 2008),. During this period early maps and diagrams were used to represent data, but they were not very sophisticated. Visualization was only in the form of maps and diagrams. In period of 1600–1699, with the grow of interest in were seen. That has led to the development of Measurement and Theory era. During this time period, more sophisticated data visualization was introduced. Period of 1700–1799 called as New Graphic Forms, in this period, new graphic forms were developed, such as the line graph and the bar chart.

During early period of 19th century,1800–1850, Beginnings of Modern Graphics era occurred. During this period, the beginnings of modern graphics were seen, with the development of techniques such as the pie chart and the scatter plot. With growth of Statistical techniques which are still used today are introduced. Thus, period of 1850– called as The Golden Age of Statistical Graphics. During the period of 1900–1950, 20th century was a time of little innovation in the field of data visualization. This period is sometimes called the modern dark ages of data visualization. Thus period was defined as Modern Dark Age. In the mid-20th century, there was a rebirth of interest in data visualization. This led to the development of new techniques such as the choropleth map and the bubble chart. Thus the period 1950-1975 is known as Rebirth of Data visualization.

In the present day, there is a growing interest in high-dimensional, interactive, and dynamic data visualization. These techniques allow for more sophisticated and engaging visualizations of data. Period post 1975 till present is referred as High-D, Interactive and Dynamic Data Visualization.

The field of data visualization is both exciting and challenging. On the one hand, we are constantly developing new ways to collect, aggregate, analyze, and visualize data. This has led to a wealth of new insights into the world around us. On the other hand, these new technologies also raise important societal challenges, such as the potential for privacy violations and the misuse of data. As we enter the information age, it is both exciting and terrifying to imagine what the future holds for us. As Randy Bachman said, "We ain't seen nothin' yet!" (Insight Software, 2019). Data visualization, which was initially used to display land markers, cities, roads, and resources on maps, evolved into a more sophisticated tool in the 17th century to meet the growing demand for more accurate mapping and physical measurement. Dashboards and data discovery tools, scorecard applications, analytics suites, and an assortment of other software tools enable businesses, researchers, and individuals to explore their data in new and increasingly imaginative ways.

2.2 Types of visualization tools

Visualization tools are a diverse range of techniques that are designed to make complex data more accessible and meaningful. They use visual elements to enhance understanding, reveal patterns, and facilitate decision-making. Some fundamental types of visualization tools which are highlighted by Stephen (Few, 2012) and Munzner (Munzner, 2014) are as :

- Charts and graphs: These are the most common visualization tools, and they use bars, lines, pies, and other graphical elements to display data relationships. Bar charts compare quantities, line graphs show trends over time, and pie charts depict proportions.

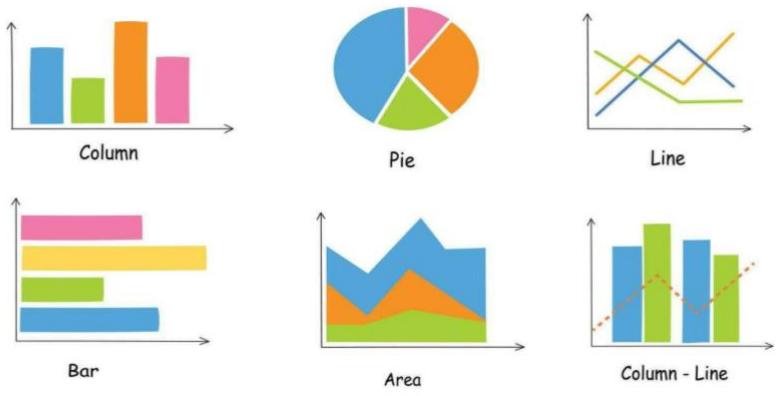

Figure 1 Charts and graphs

Source (Tran, 2023)

- Heatmaps: Heatmaps use color gradients to visualize data values on a grid. They are effective for displaying large datasets and identifying patterns or clusters.

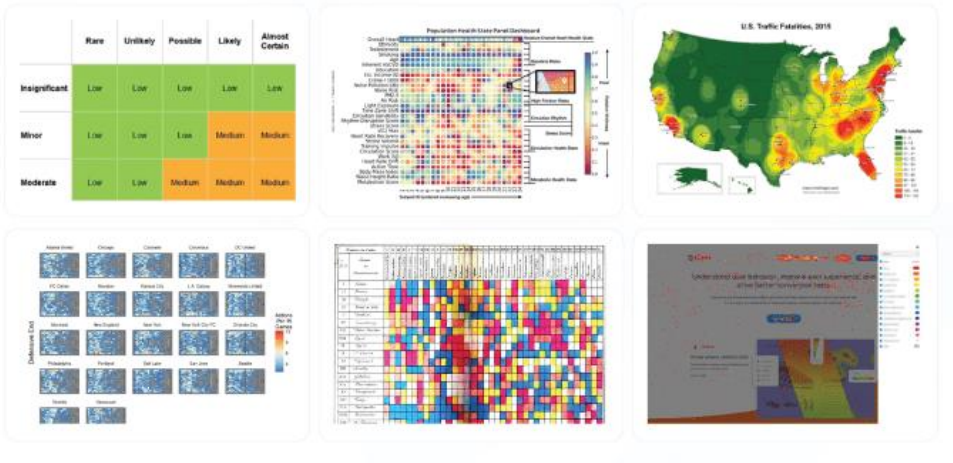

Figure 2 Heatmaps

Source (Capturly, 2023)

- Scatter plots: Scatter plots plot data points on a two-dimensional grid, with each point representing a unique observation. They reveal correlations and relationships between variables.

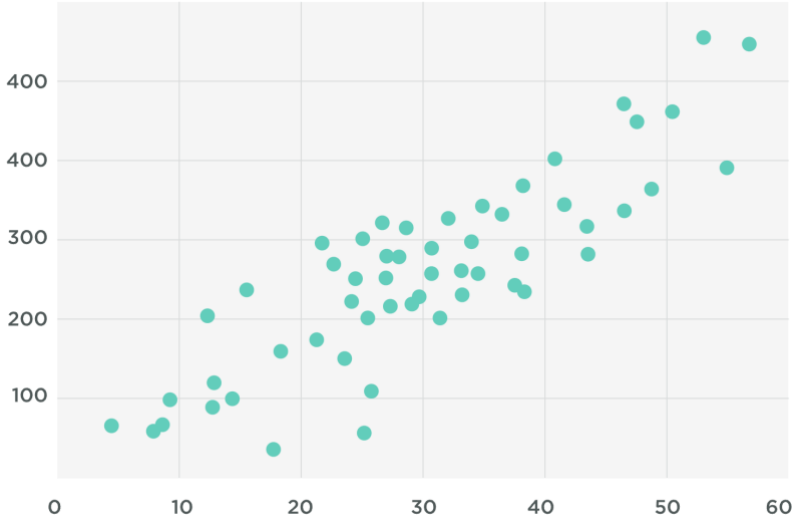

Figure 3 Scatterplot

Source (Spotfire, 2022)

- Network diagrams: Network diagrams illustrate relationships among interconnected entities. Nodes represent elements, while edges depict connections, aiding in the visualization of complex networks.

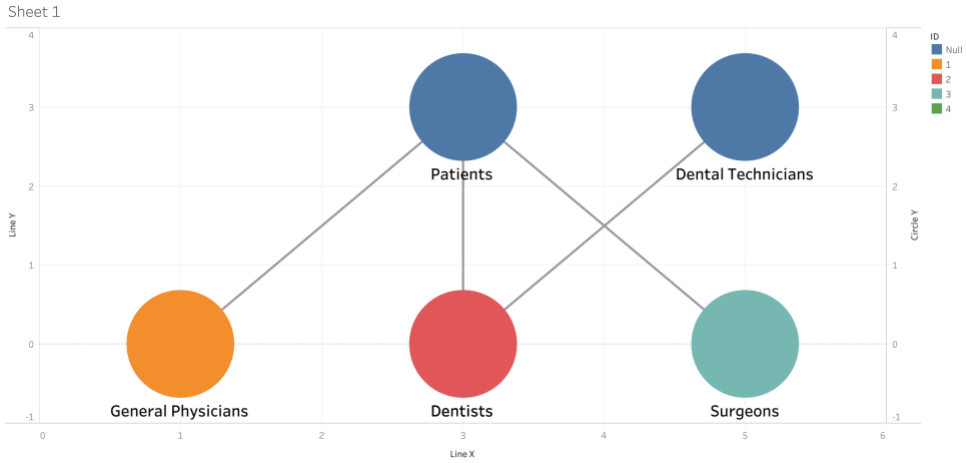

Figure 4 Network Diagram

Source (Chintakindi, 2022)

- Interactive dashboards: Interactive dashboards provide a customizable interface with multiple visualizations. Users can manipulate parameters, filter data, and explore insights dynamically.
- Geospatial visualizations: Geospatial tools map data onto geographical locations. Geographic Information Systems (GIS) display data on maps, helping visualize spatial patterns.

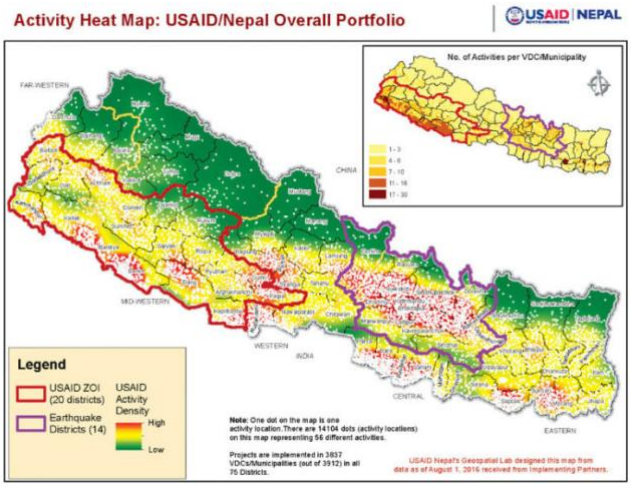

Figure 5 Heatmap

Source (USAID, 2022)

- Tree maps: Tree maps use nested rectangles to display hierarchical data structures, such as file systems or organizational hierarchies.

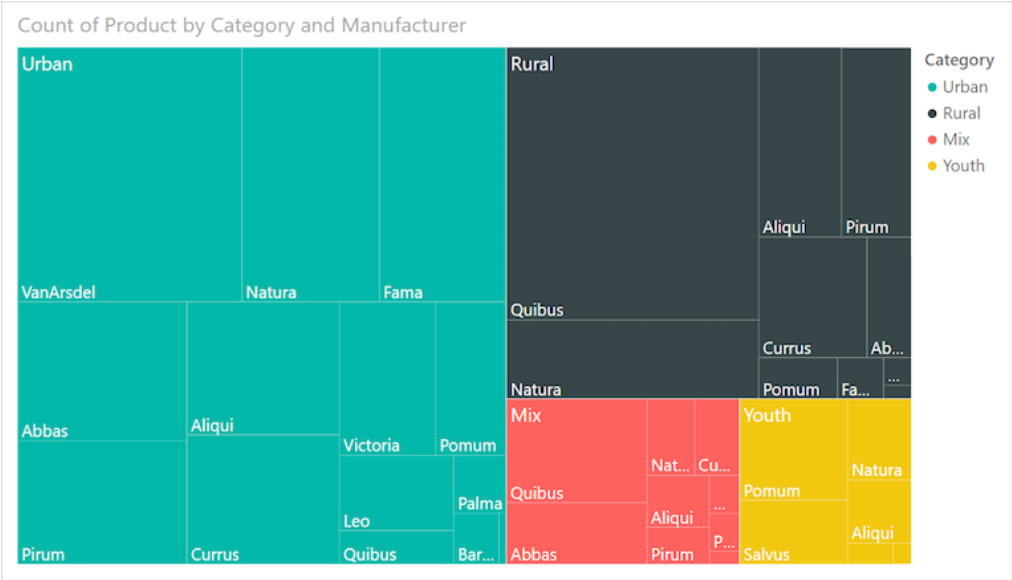

Figure 6 Tree maps

Source (Microsoft, 2023)

- Word clouds: Word clouds visually represent word frequency in a text, with larger words indicating higher frequency.

Figure 7 Word clouds

Source (Outlook India, 2023)

- Flowcharts and process diagrams: Flowcharts and process diagrams visualize workflows, processes, or decision paths, aiding in understanding complex systems.

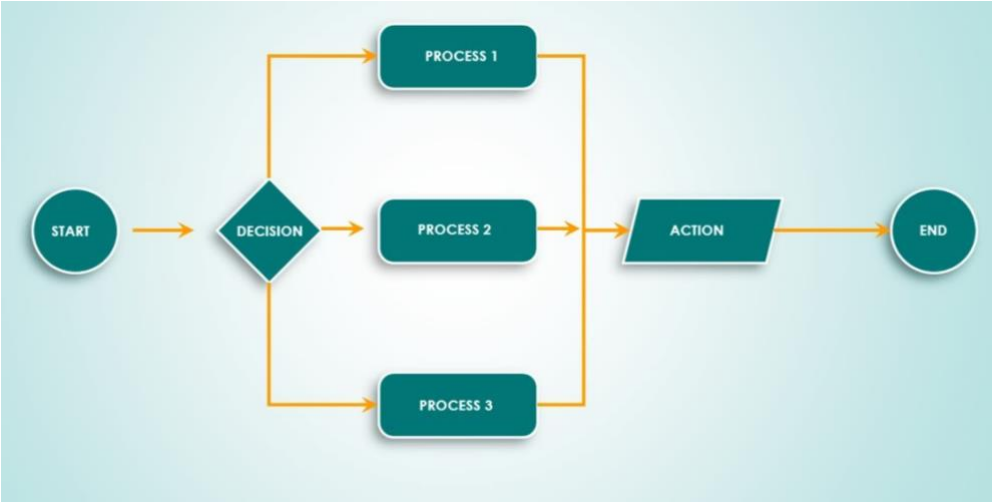

Figure 8 Flow diagram

Source (Ihi org, 2018)

- Virtual Reality (VR) visualizations: VR tools immerse users in a virtual environment where they can interact with data, providing an immersive and engaging visualization experience.
- Augmented Reality (AR) visualizations: AR tools overlay digital information onto the real world, allowing users to visualize data within their physical environment.

These tools can be used in a variety of fields, including business, science, and education. They can help to make data more understandable and accessible, and they can be used to identify patterns and trends that would not be visible in traditional data representations.

2.3 Decision Biases and Visualization Tools

Human decision making is often susceptible to cognitive biases, leading to deviations from rational and optimal choices. These biases, rooted in mental shortcuts and heuristics, can significantly impact the quality and accuracy of decisions. Cognitive biases, such as confirmation bias (Nickerson, 1998), anchoring bias (Tversky & Kahneman, 1974), and availability heuristic (Tversky & Kahneman, 1973), influence how information is processed and can lead decision makers to systematically deviate from rational judgment. These biases may result in suboptimal outcomes, distorted risk assessments, and hindered problem-solving. Recognizing and addressing these biases is crucial for improving decision-making processes across various domains.

Visualization tools offer a promising avenue for mitigating decision biases. Visualization leverages the human brain's visual processing capabilities to represent complex data in intuitive and comprehensible forms. Visualization tools encompass a range of techniques,

such as graphs, charts, heatmaps, and interactive dashboards, designed to present information in ways that facilitate clearer understanding, pattern recognition, and informed decision making. By externalizing information and fostering a holistic view, visualization tools have the potential to counteract cognitive biases by promoting transparency, enabling data-driven reasoning, and enhancing cognitive reflection.

2.4 Heuristics and Biases in decision making

The term heuristic refers to "a rule of thumb, or informal reasoning strategy, as opposed to a mathematical formula that can be calculated". (Marita Turpin, 2004)

Heuristics are mental shortcuts that we use to make quick and efficient decisions. They can be helpful in many situations, but they can also lead to biases in our decision-making. One common heuristic is availability bias, which is the tendency to judge the probability of an event based on how easily examples come to mind. For example, if you have recently heard about a plane crash, you might be more likely to think that plane crashes are common, even though they are actually very rare.

Another common heuristic is representativeness bias, which is the tendency to judge the likelihood of an event based on how similar it is to other events that we have experienced or heard about. For example, if you meet someone who is tall and handsome, you might be more likely to think that they are intelligent, even though there is no evidence to support this.

Heuristics and biases can lead to poor decision-making, but they are also a natural part of the human thought process. It is important to be aware of these biases so that we can try to avoid them when making important decisions.

Research papers mentioned above have highlighted cases where biases can be reduced by information systems. Data visualization can also help to reduce cues, and framing issue which will help in reducing biases and make informative decision.

2.5 Data visualization and Information Visualization

Data visualization is a key technique for representing complex datasets in a visual format, which can help people to better understand and analyze quantitative information. It involves transforming raw data into graphical elements such as charts, graphs, and maps, with the goal of revealing patterns, trends, and insights that might otherwise be hidden in the data (Tufte, 1983). By harnessing the power of visual representation, data visualization can facilitate data exploration, comparative analysis, and effective

communication of findings to a wide audience (Few, 2012). Some common examples of data visualization include bar charts, line graphs, scatter plots, and heatmaps.

Popular data visualization tools include Tableau, a versatile platform for interactive data exploration, and Microsoft Power BI, a powerful tool for creating compelling visual narratives. Open-source libraries such as Matplotlib (Python) and ggplot2 (R) provide researchers with customizable options to create tailored visualizations that meet their specific analytical needs (Heer & Shneiderman, 2012).

Data visualization is the graphical representation of data. It can be either static or interactive, and it is often used to identify important characteristics and patterns in large data sets. Data visualization can be used for statistical analysis and decision-making. It allows the designer (analyst) to explore, modify, consider, and experiment with a variety of different ways to visualize the data. Data visualization can be a powerful tool for communicating complex information in a clear and concise way. It can be used to help people understand data, make decisions, and solve problems.

Information visualization is the process of transforming complex information into visual representations that help people understand and make decisions (Spence, 2007). It can be used to visualize quantitative data, abstract concepts, relationships, and structures.

Information visualization draws on cognitive psychology and human-computer interaction to ensure that visual representations are clear and effective in conveying complex concepts (Card, et al., 1999). The goal is to empower users to derive insights and make informed decisions by engaging with intuitive visual depictions of information and knowledge.

Information visualization is the result of analyzing data and is intended to persuade the viewer to reach certain (and therefore subjective) conclusions. It is often created with a specific set of data in mind in order to draw the viewer's attention to the relationships that the designer wants to emphasize. As a result, it presents a well-developed narrative that guides the audience to "intended" conclusions. It can be seen as a message, an infographic, or a form of communication between the visualization designer and the viewer.

It's important to remember that when discussing manipulations and intentional distortions, we're primarily talking about the visualization of information, i.e., what the chart designer intended to communicate to the viewer. Any distortions in data visualization are likely unintentional, resulting from the analyst's lack of preparation or incorrect data collection. (Krok, 2021).

2.6 Power BI

Power BI is a unified, scalable platform for self-service and enterprise business intelligence (BI). Connect to and visualize any data, and seamlessly infuse the visuals into the apps you use every day. (Microsoft, 2020)

Power BI is a suite of software services, apps, and connectors that work together to turn your disparate data sources into coherent, visually immersive, and interactive insights. Your data could be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI makes it easy to connect to your data sources, visualize and discover what's important, and share it with anyone or everyone you want.

Basic standard layout of Power BI visuals can be seen in Figure 9, here multiple data set are taken as input and after processing into multiple visual types final dashboard is prepared.

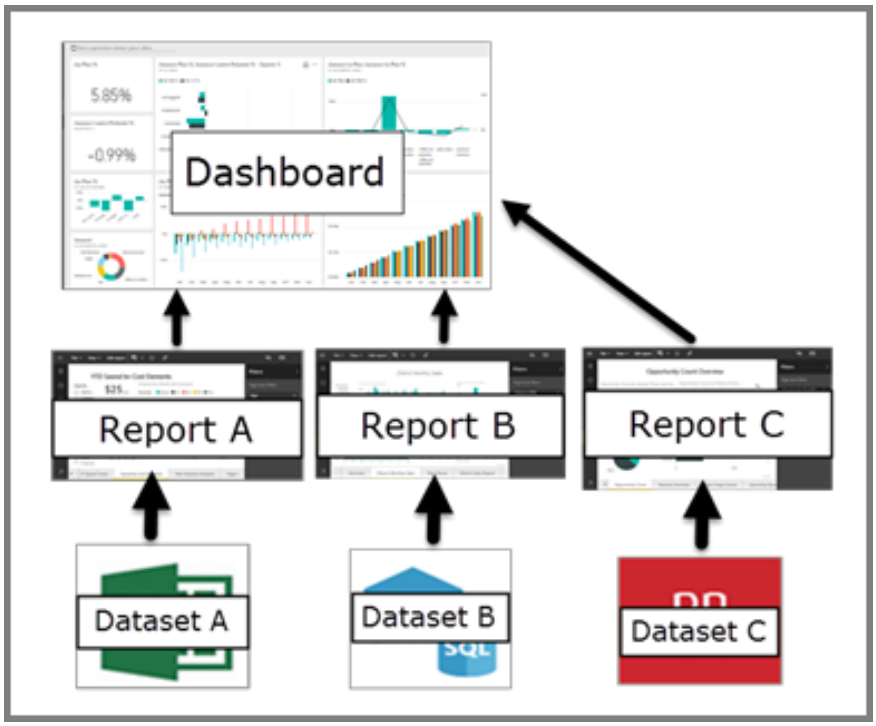

Figure 9 Dashboard overview

Source (Microsoft, 2023)

Power BI consist of several elements and on this basic classification can be done as:

- A Windows desktop application called Power BI Desktop.
- An online software as a service (SaaS) service called the Power BI service.
- Power BI Mobile apps for Windows, iOS, and Android devices.

Power BI Desktop, the service, and the mobile apps are designed to help you create, share, and consume business insights in the most effective way.

A typical workflow in Power BI begins by connecting to data sources in Power BI Desktop and creating a report. The report is then published from Power BI Desktop to the Power BI service, where it can be shared with business users who can view and interact with it on mobile devices. This workflow is common and demonstrates how the three main Power BI elements complement each other. Power BI Desktop is a desktop application that allows you to connect to data sources, create reports, and publish them to the Power BI service. The Power BI service is a cloud-based service that hosts your reports and allows you to share them with others. Mobile apps are available for iOS, Android, and Windows devices that allow you to view and interact with reports on the go. (Microsoft, 2023)

The three main Power BI elements are complementary in the following ways:

- Power BI Desktop allows you to connect to a wide variety of data sources and create reports.
- The Power BI service hosts your reports and allows you to share them with others.
- Mobile apps allow you to view and interact with reports on the go.

Power BI is a versatile tool that can be used by business analysts, salespeople, and developers to create reports, dashboards, and embed data into applications. It can be used for a variety of purposes, such as tracking sales progress, viewing inventory, and creating reports.

Parts of Power BI can be analyzed as shown in Figure 10, as highlighted by Learn Microsoft article,

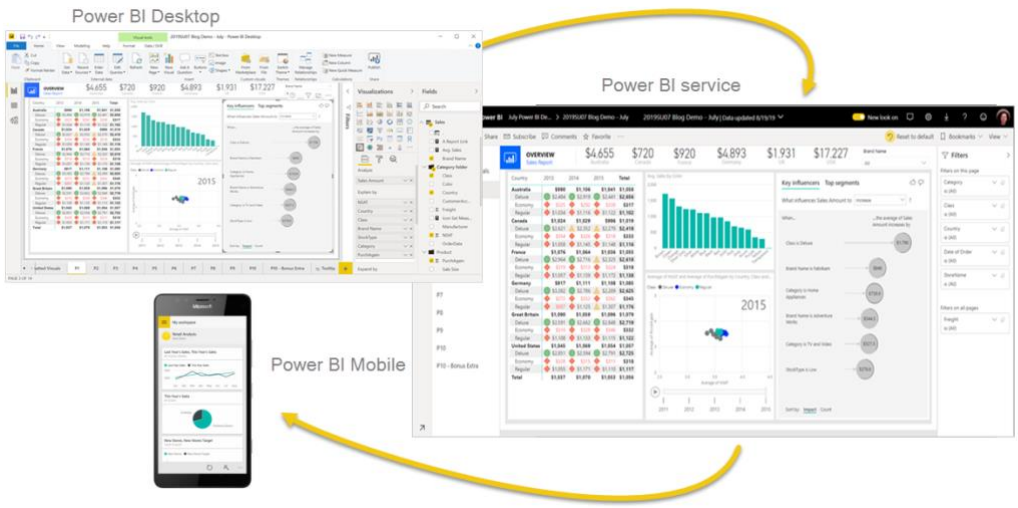

Figure 10 Power BI overview

Source (Microsoft, 2023)

2.7 Q&A in Power BI tools

Microsoft Power BI is one of the business intelligence tools that helps to connect to all sorts of data sources, turn that raw data into interactive graphs and charts, and share findings with the whole company. It's easy to use and has powerful analytical capabilities, which is why it's so popular with businesses and researchers of all kinds.

Power BI can be used as both data visualization and information visualization tool. User can use Power BI to make data visuals and also use it to manipulate the visual information which can manipulate in decision making.

Power BI's Q&A is a natural language query interface that allows users to interact with their data by asking questions in plain language. This innovative feature bridges the gap between data experts and non-technical users, allowing both groups to explore complex datasets without the need for specialized programming or query languages. Users can type questions into the Q&A box, just like they would in a search engine, and receive instant visual responses in the form of charts, graphs, or tables. (Richardson, n.d.)

Use of Q&A can help to remove biasness by user involved in making visualization as it is made by natural language input given to the processing engine thus making visuals free from biasness made by user.

2.8 Positive impact of Visualization

With proper training and extensive time given while creating visualization, biases in decision making can be improved and additionally decision making through visualization would be more effective. Study also have validated these cases to show positive impacts.

2.8.1 Positive Effect 1: Visualization improved faster data processing

From one research, Karin have highlighted that the role of visualization in computer graphics for decision support is complex and should not be underestimated. The cognitive fit theory's strong validity and the influence of user characteristics underscore the need for careful consideration during design. Clarity regarding the intended audience and purpose of the visualization is crucial. Additionally, while visualization can lead to overconfidence and automatic decision-making, it may not always be suitable, emphasizing the importance of balanced and thoughtful decision processes (Eberhard, 2023). After the study of more than 160 research paper, to analyze the effects of visualization on judgment and decision-making, research highlighted detailed audience analysis can help to make decision more constructive. And, when the clarification for whom and what the visualization is intended is not made then information may result in

decision-makers skipping on more elaborate thought, which may be desirable in some, but certainly not all situation.

There has been evidence that, applied principles of perceptual cognition to human-computer interface design to introduce uncertainty visualizations in an adaptive approach that improved the operator's decision-making process, without unduly burdening the operator's cognitive load (Block, 2013). Use of graphs lead to faster processing, learning, and decision-making as judgment and decision efficiency are measured and operationalized as the response time in various experiments.

2.8.2 Positive Effect 2: Visualization improves decision accuracy and quality

Several Research papers published back up the hypothesis that visualization is an effective tool to support decision making. Some of the research paper and their summary in approving the hypothesis.

Findings made during one study support “graphical representation had reduced risk aversion in gambling” (Michael Dambacher, 2023) . Researchers had agreed to cases of reduced risk aversion through a graphical display of outcome. In comparison of data vs graphs which are supported on behalf of visual elements.

Another research carried out in the public health domain for decision-making research, visualization appears to bring advantages by increasing the amount of information delivered and decreasing the cognitive and intellectual burden to interpret information for decision-making. (Seungeun Park, 2021)

Though for this data, it was seen that understanding regarding data visualization interventions specific to public health leaders’ decision-making still lacks. Research have not clearly mentioned cause of this but have suggested more research need in terms of user inclusion from starting phase and updating methodology for public health sector.

Several other studies are present to back up visualization effectiveness in decision making. But most importantly for the negative impact made by data visualization, there is not much research carried out.

One of the research projects carried out has highlighted output from other researchers and has stated that: “Producing a graph or a map and plotting data on it is one thing, but designing visualizations that actually support decision makers is something completely different which is far from being trivial.” All of the contributing research had shown visualization as an important parameter but has raised the research gap for its importance in decision making. (Corentin Burnay, 2019)

2.8.3 User Interaction and Adoption

The successful use of visualization tools to remove bias from decision-making processes depends on effective user interaction patterns and seamless adoption within organizational contexts. The way users interact with these tools and their willingness to use them have a significant impact on the extent to which biases can be identified and mitigated. Notable research findings emphasize the critical role of user interaction in improving bias detection. Active exploration, in which users engage in hands-on data manipulation and delve into subsets, has been shown to uncover subtle biases that might otherwise go unnoticed. (Ward, et al., 2015). Furthermore, visualization tools that provide multiple ways to visualize and manipulate data facilitate holistic exploration, enabling users to approach bias detection from diverse angles (Shneiderman, 1996). Collaborative decision-making environments, allowing stakeholders to engage with visualizations collaboratively, have also been highlighted as promoters of diverse perspectives that enhance overall bias detection effectiveness (Isenberg, et al., 2013). However, despite the potential benefits, user adoption of visualization tools is not without challenges. The introduction of new tools often entails changes in workflow and practices, which can encounter resistance within organizations (Muller, 1993) . Effective user training, focusing on enhancing data literacy and ethical decision-making, is pivotal for maximizing the benefits of visualization tools (Heer & Bostock, 2010)Balancing the advantages of automation facilitated by visualization tools with the need for thoughtful decision-making is a critical consideration. While visualization enables faster processing of information, users should remain cautious of overreliance on automatic decision-making, especially in contexts that require comprehensive analysis and mitigation of potential biases (Gleicher, 2018).

2.9 Python and Matplotlib in visualization

Python is essential for ensuring fair and unbiased data analysis and visualization. It is a dependable tool for exploring data transparently and without hidden biases. Alongside Python's impartiality, tools like Pandas facilitate data cleaning and sorting, maintaining data impartiality throughout the analysis process (McKinney, 2011). Python also offers interactive data exploration tools like Plotly and Bokeh, which allow users to interact with data in real time and avoid jumping to conclusions (Sievert, 2020). The transparency of Jupyter Notebooks makes it easier to understand the data analysis process, which can help to ensure fairness.

Matplotlib, a powerful Python library, is particularly noteworthy for its use in data visualization. It allows for the creation of charts and graphs that accurately represent data, ensuring neutrality in visualizations (Hunter, 2007). Its objectivity and flexibility make it a popular choice for creating unbiased visual representations of data.

Python's commitment to impartiality extends to the realm of machine learning, where libraries like scikit-learn and TensorFlow prioritize fairness in model training and evaluation. (Pedregosa & Varoquaux, 2011). However, Python has some challenges in handling large datasets and integrating with external tools. Addressing these challenges in an impartial manner is crucial to maintaining its reputation as a reliable tool for honest and unbiased data-driven insights. In essence, Python, along with Matplotlib, acts as a fair arbiter in the data domain, ensuring that data exploration and analysis remain unbiased and trustworthy.

Figure 11, represents the types of plots which can be generated using Matlplotlib. Here based on user requirement and necessity the plot type can vary. (Nurhidayat, 2020).

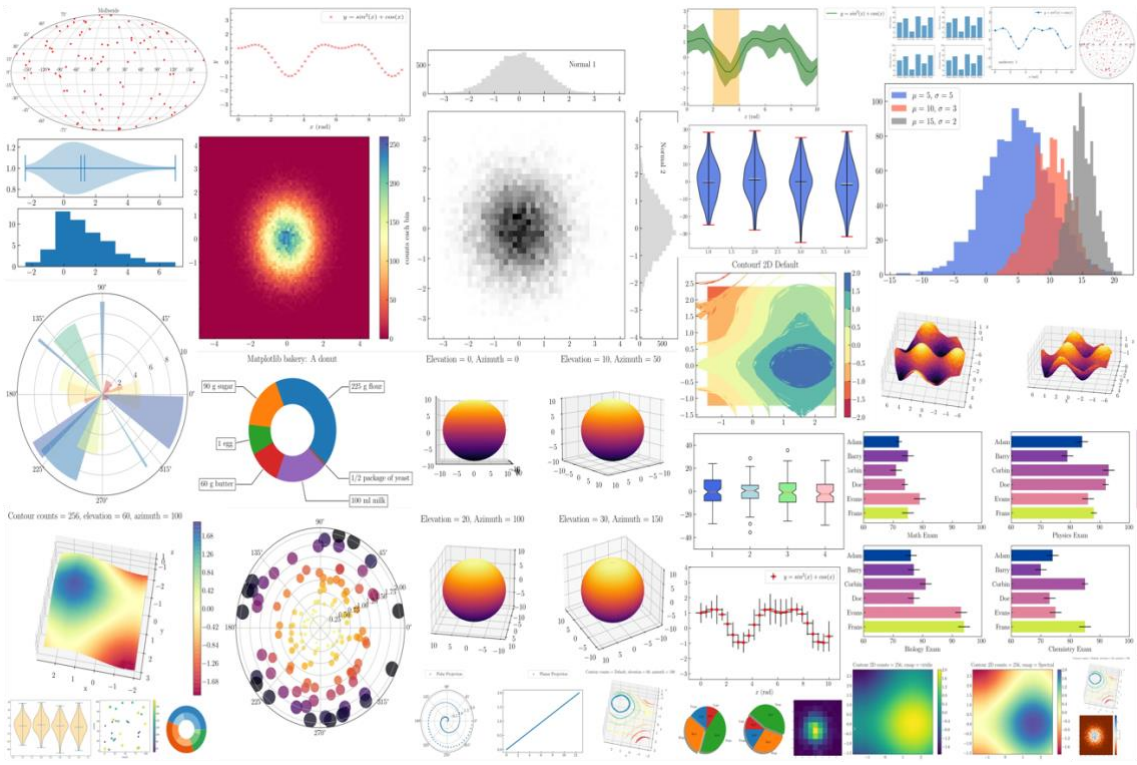

Figure 11 Various type of plot generated in Matplotlib

Source (Nurhidayat, 2020)

2.10 Pandas in Python

Pandas is a free and open-source Python library that is used for data science, data analysis, and machine learning. It is easy to use and can handle both relational and labeled data. (pandas, 2021) It offers a variety of data structures and operations for working with time series and numerical data. Pandas is built on top of the NumPy library, which supports multi-dimensional arrays. This makes Pandas fast and efficient. Pandas is one of the most widely used data-wrangling tools and integrates well with a variety of other data science modules in Python.

Pandas provides tools for loading data from various file formats into in-memory data objects. It also enables label-based slicing, indexing, and sub setting for large datasets. Pandas simplifies tasks like merging, joining, pivoting, and reshaping datasets. It excels at handling missing data, both in floating point and non-floating-point formats. Pandas represents data in tabular form. With size mutability, you can easily add or delete columns in Data Frames and higher-dimensional objects. Pandas also supports time-series functionality and offers effective grouping features for splitting, applying, and combining datasets. (Devi, 2022)

Pandas makes data visualization incredibly simple. This helps with improved data analysis and understanding. Data science projects produce better results when the data is represented more simply.

Another of Pandas' best features is that it can help you write less code. With the help of Pandas, multiple lines of Python code can be easily completed in one or two lines. This helps to reduce time and procedures while also speeding up the data-handling process. As a result, you can devote more time to data analysis algorithms.

2.11 Tkinter library for file opening

Tkinker library helps to create a simple graphical user interface (GUI) for selecting a CSV or Excel file. The ability to select and import data files is a common requirement in data analysis workflows. GUIs built using tkinter can facilitate this process, enhancing the user experience.

The Python tkinter library provides an efficient and user-friendly way to prompt users for file selections. This makes it a valuable tool in applications that require data input. The library allows for the creation of file dialogs that enable users to navigate their file systems and choose the relevant data file for analysis. (python, 2021)

Furthermore, tkinter offers capabilities for handling user interactions and providing feedback. This contributes to the overall usability of data analysis tools. Through error

handling and informative messages, tkinter-enhanced applications can guide users in selecting appropriate files and address any issues that may arise during the process.

In data analysis tasks, the integration of tkinter with libraries like pandas for data processing and manipulation is common practice. By coupling tkinter's GUI capabilities with pandas' data handling capabilities, analysts can develop user-friendly data importation tools that empower users to load and analyze data seamlessly (McKinney, 2017).

In conclusion, tkinter's role in simplifying user interactions and data file selection is pivotal in enhancing data analysis applications' usability. Its integration with other Python libraries, such as pandas, contributes to the development of efficient and user-friendly tools for data professionals and analysts. When designing data analysis software, considering the user experience and leveraging tkinter's capabilities can lead to more effective and accessible tools for data manipulation and interpretation.

2.12 Type of visualization that actually works

In the past, the ability to create smart data visualizations, or dataviz, was a desirable skill. For the most part, it benefited design- and data-minded managers who made a conscious decision to invest in acquiring it. However, this has changed. Now, visual communication is a necessary skill for all managers, because more and more often, it is the only way to make sense of the work they do. More and more, decision-making depends on data. This data comes at us at such a rapid pace and in such large quantities that we cannot comprehend it without some form of abstraction, such as a visual one.

With the growth of internet number of affordable tools are increasing. Because of which translating information into visual has become easier and any managers requiring information can easily generate required visuals.

Managers learning ideas about which color to use, which type of chart to use, knowing about the basis of zero level etc. is important and useful skill to have. However, knowing this tool does not guarantee that will make good chart. (Scott, 2016)

Scott, has highlighted to start thinking visually, nature and purpose of the visualization needs to be known.

- Is the information conceptual or data-driven?
- Am I declaring something or exploring something?

The first question is the easier of the two, and the answer is usually clear. We either are visualizing qualitative data or quantitative data: ideas or statistics. However, it needs to

be noticed that the question is about the data itself, not the formats that may be ultimately used to display.

Thus, with the use of the mentioned question, manager can get hands-on experience with the visualization tool. The use and learning of the tool are long and iterative process.

CHAPTER THREE: METHODOLOGY

3.1 Research Framework

The research framework for the study aims to investigate biases in decision-making processes and evaluate the effectiveness of visualization tools in reducing bias. It includes the following stages:

- Identifying biases
- Selecting suitable visualization tools based on predefined criteria
- Integrating visualization tools into decision-making scenarios
- Assessing the impact of visualization tools on bias reduction

The study uses a mixed-methods approach, combining qualitative observations with quantitative surveys, to gain insights into biases, measure the prevalence of biases, and gauge the efficacy of visualization tools. By integrating and analyzing these data sources, the study aims to offer a comprehensive understanding of biases, the utility of visualization tools, and their practical implications for enhancing decision-making processes. The framework concludes with the interpretation of findings, actionable recommendations, and suggestions for future research, thus contributing to the advancement of decision-making practices and strategies for bias mitigation. Addition to this, python-based program output will be presented which will help in producing visual elements easily without any biasness.

3.2 Research Design

The study titled "A Study on Usages of Visualization Tools to Remove Biases in Decision Making" uses a mixed-methods approach that combines both quantitative and qualitative research techniques. This design was chosen to provide a comprehensive understanding of how visualization tools can effectively mitigate biases in decision-making processes. The design includes exploratory and descriptive elements, allowing for the exploration of the phenomenon, the identification of biases, and the evaluation of the impact of visualization tools in real-world decision-making scenarios. And presenting solution for visualization using python.

Python based model is also created which can be helpful for managers or decision makers to create charts without much knowledge of software tools. Also, since the involvement of visual manipulation will be overridden manipulation and biasness will be less in this case.

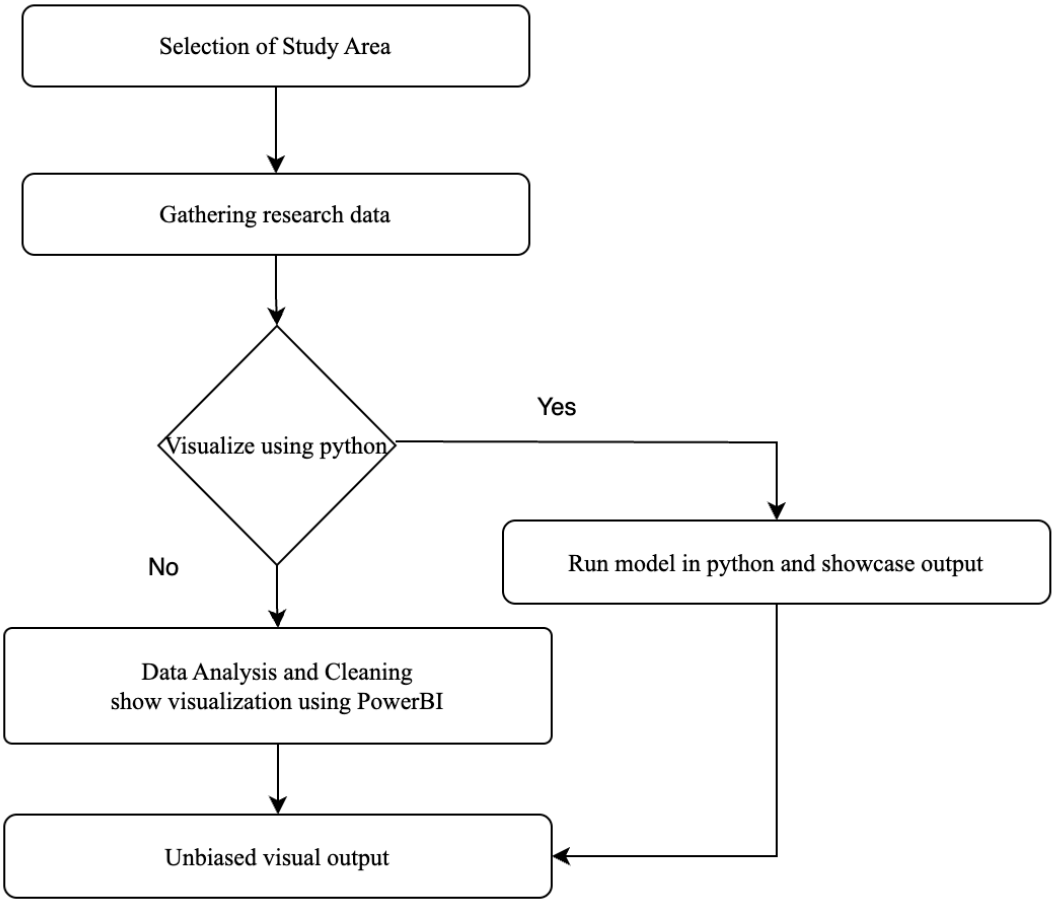

Figure 12 Research Methodology

3.3 Data selection

For Quantitative analysis data sample for ecommerce industry is used to show data deformation and manipulation. The showcase of biasness will be done with the use of ecommerce data as it contains many dynamics in itself. Since the data within the ecommerce domain contains all of the variables related to finance, commercial, operations and user details. However, the python model generated for visualization helps to create charts for most of the industries like IT, healthcare, Ecommerce, Airlines etc.

CHAPTER FOUR: ANALYSIS AND RESULTS

4.1 Types of Manipulation

Power BI was used to simulate and visualize potential manipulation scenarios in ecommerce data. Ecommerce data was used in analysis because of dynamic data nature of ecommerce industry. Commercial, financial or operational data mix can be found in ecommerce data sample. By creating synthetic datasets, Power BI highlighted different types of manipulations, such as conscious manipulation which visualizer can make. These type of distortion and manipulation needs to be considered well before using for data-based decision making. The sample rows of data used to show the types of manipulation is represented in Annex 2.

4.1.1 Changing Scales

Change of scale is common data distortion technique used commonly. Figure 13 represents the total quantity ordered which is plotted against each month. To hide the distraction changes, report can be distorted like in Figure 14, where axis level can be hidden which visually shows slow changes in the order level.

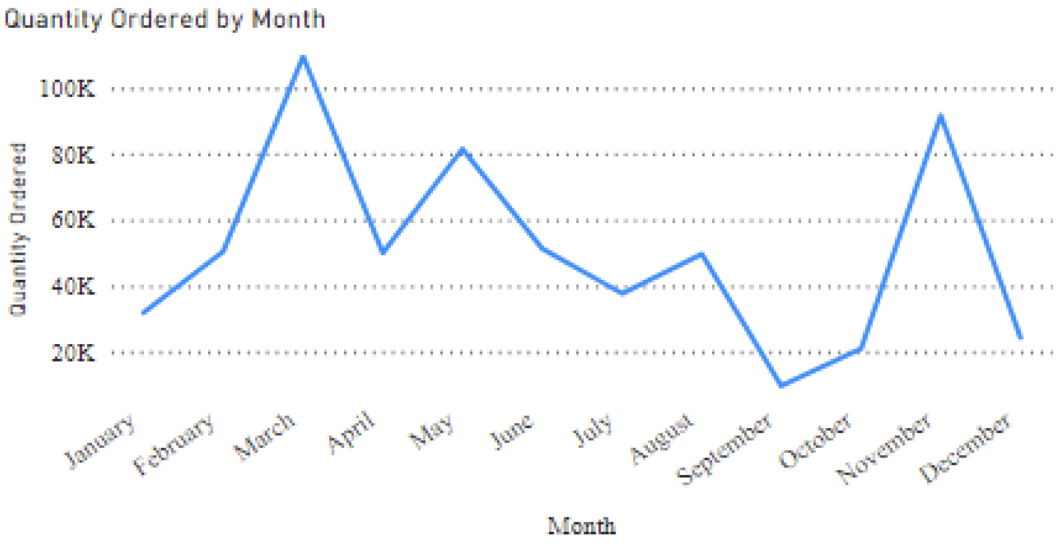

Figure 13 Order variation

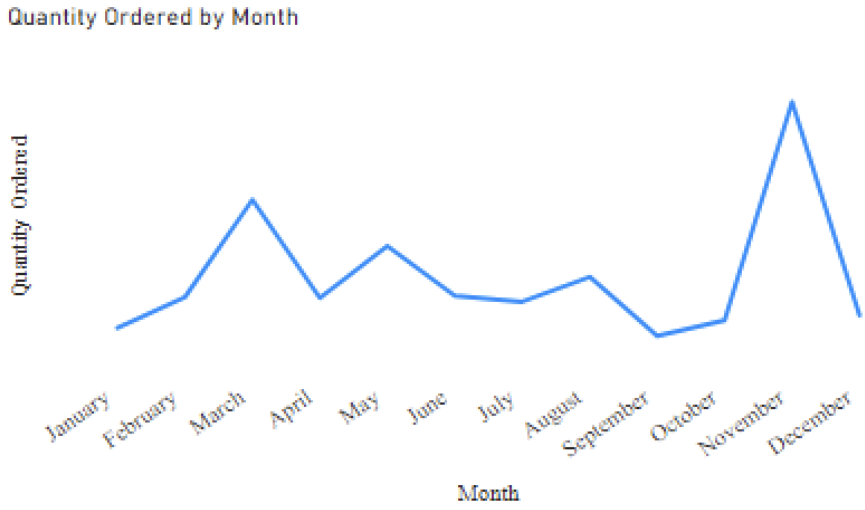

Figure 14 Deformation Axis

Scale and axis label are hidden in the figure to manipulate visual criteria.

This distortion helps to hide sudden changes in the data and also in reverse case can be used to show more variation. To cover slow changes in data, these manipulations are used. Figure 15 shows changes in scale technique which helps to hide order data which is below the required level. This makes order data to always be above the required level.

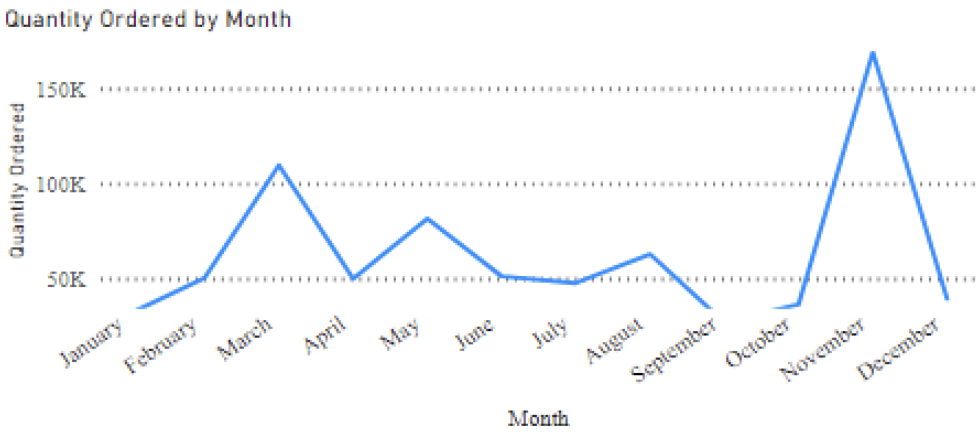

Figure 15 Change of scale

Another form of scaling can be found for figure 16, here the category wise order are presented. For manipulation purpose, category order for books which is ~1.5% can be scaled up by removing data labels and distortion.

Figure 17 and figure 18 represent such manipulation where user will be encouraged to believe that category wise segregation for book is not so bad. While making decision related to category level, chart can manipulate significance.

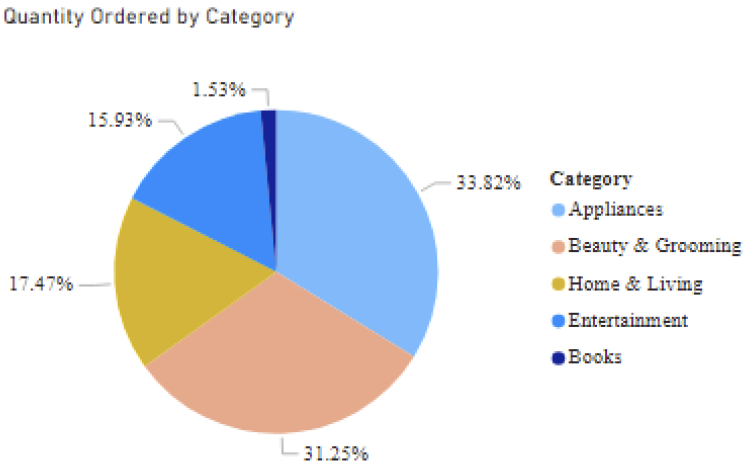

Figure 16 Order by category

Small difference in color will not make book category significant thus, while presenting category while comparing with other, book category can be made larger which make user perspective change towards quantity order of book.

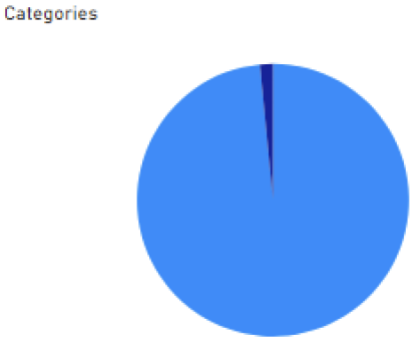

Figure 17 Category actual

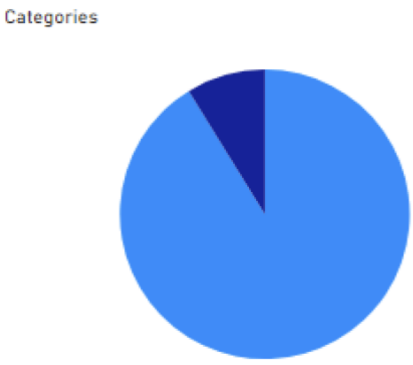

Figure 18 Category distorted

When same quantity order is being represented in bar chart, based on BI status of Net and Valid variation of quantity order can be observed as shown in figure 19. Here Net and Valid order segregation with each order quantity is represented.

A common deceiving method used by designers who create chart is to start the y-axis at a value other than zero. This makes it appear as though the bars are much different in size than they actually are. If one bar is twice as long as the other, it may appear to be four times as long if the y-axis starts at two instead of zero. This can be used to mislead people about the data being presented. In figure 19 both bar starts at zero, whereas compared to figure 20, bar starts at 150K, which makes recipient get a contrasting view of the two compared values.

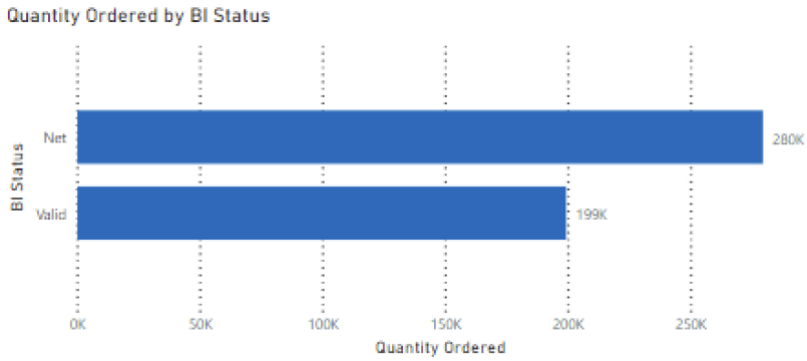

Figure 19 BI status wise order

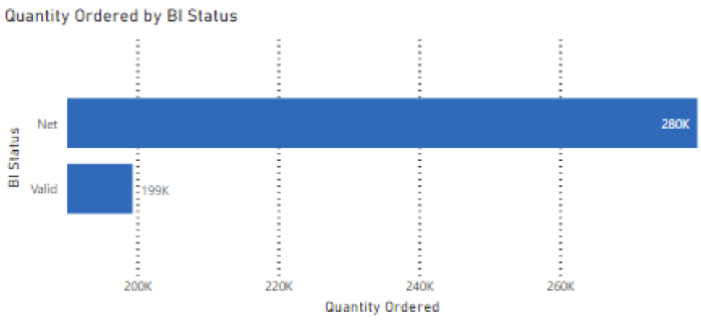

Figure 20 Scale changed BI status

A lie factor greater than one indicates that the graph exaggerates the relationship between data points, while a lie factor less than one indicates that the graph understates the relationship. The lie factor can be used to identify graphs that are misleading or intentionally deceptive.

Lie factor = (size of the effect in the visual) / (size of effect in the data)

Where size of effect = (second value – first value) / (first value)

To ensure that the graphics do not mislead the audience, the lie factor should be between 0.95 and 1.05 (Tufte, 2001). According to the creator of the coefficient, such small variations in the visual message will not affect the interpretation. However, if the lie factor value is lower or higher, it suggests that the data has been intentionally manipulated to understate or exaggerate it.

4.1.2 Changing perspective views and colors parameter

Another type of manipulation is by use of perspective and colors. When order is shown on the basis of category, pie chart segregation. Here category of entertainment, have order share of 16%.

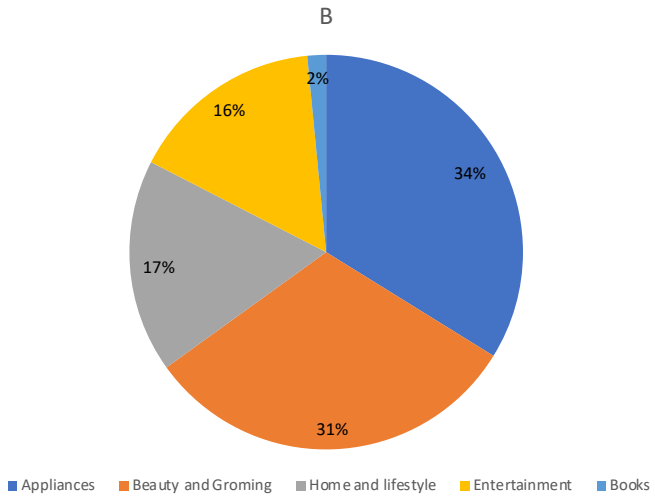

Figure 21 Category segregation

For reporting cases, if the design of chart is made according to highlight entertainment category and make beauty and grooming category to look performing poor then figure 22 like chart can be prepared. Here category beauty and grooming is shown in red color where as intended category Entertainment is shown as dominating category by use of perspective tool.

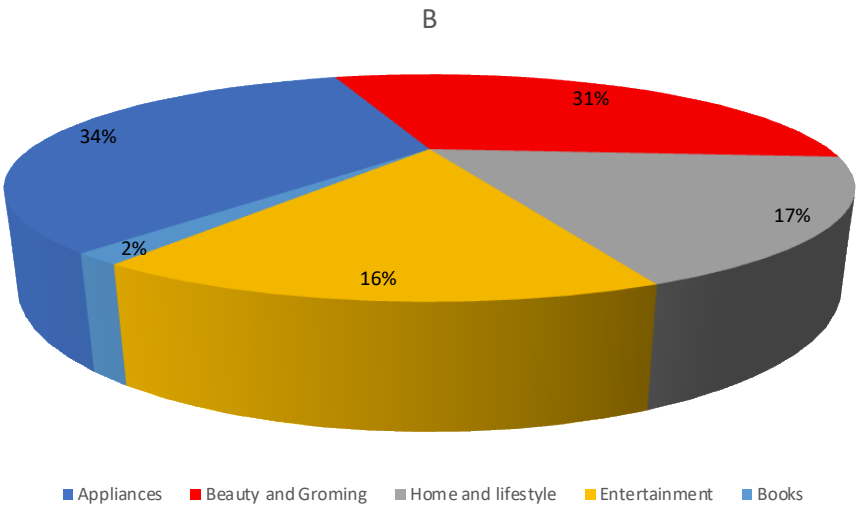

Figure 22 Perspective and color

4.1.3 Hiding trends and changing data view

Figure 23 represent variation of grand total, price, market value is shown month over month. Here the variation of cash flow at each month can be observed. Clearly for the month of January and April flow is low compared to other month.

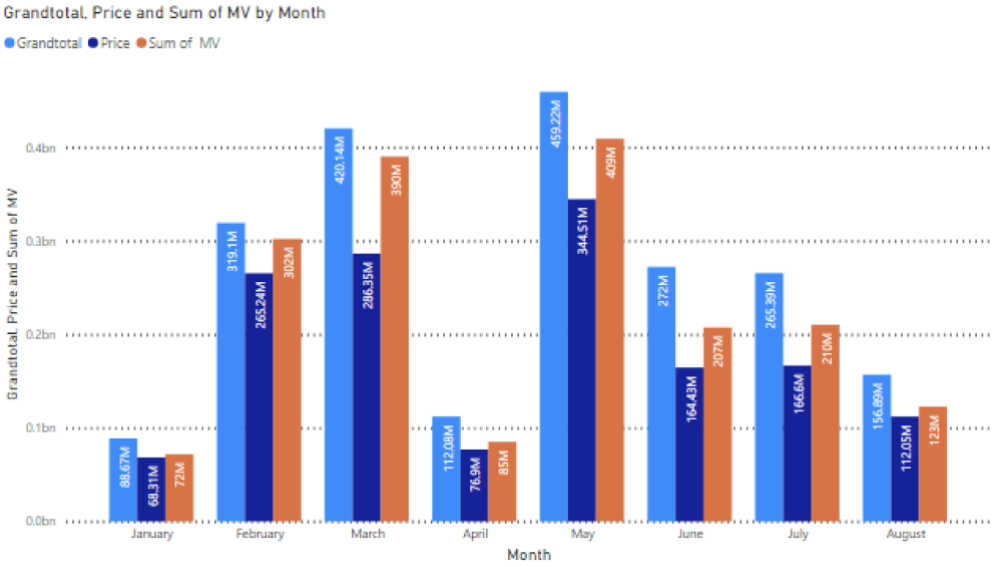

Figure 23 Amount shown MOM

Figure 24 shows how the variation in the values can be manipulated in the graphical formatting. Deliberately inconvenient data can be concealed which have decreasing trends. Cumulative chart can be used for such scenarios, where total data can be presented making up to 100%.

The visualization makes it difficult to see the company's poor performance. Analyzing the percentage share of each category requires more attention and interpretation. A decrease in revenue in January and April is also a sign of trouble, although in theory

(When a smaller number of clients generate the same amount of revenue as a larger number), it is not a clear sign of a problem. The cumulative chart design, therefore, is meant to hide data and give the recipient a different view of the company.

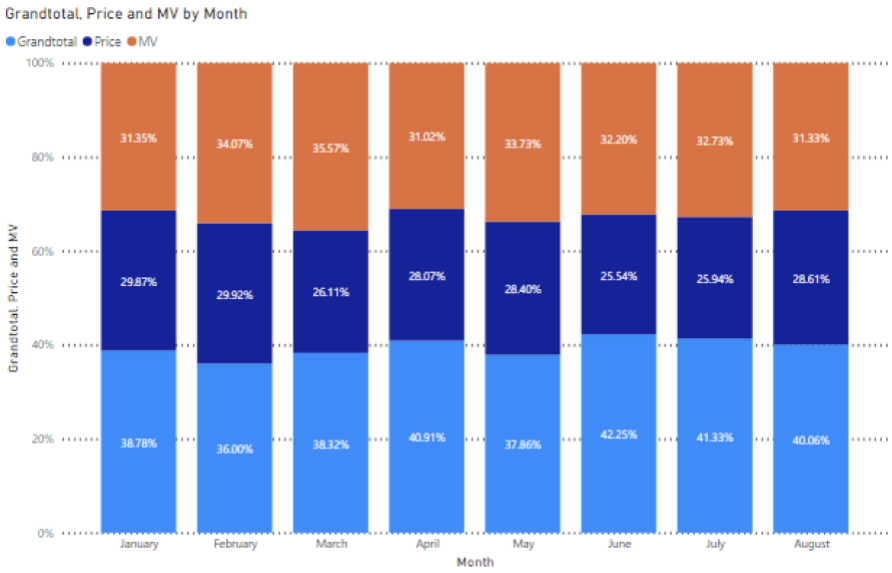

Figure 24 Low cash flow hidden

4.1.4 Showing absolute data in percentage form

Figure 25 represent amount changes with respect to BI status of the data. Here grand total, price and discount amount total are shown for each BI status. BI status for Net and Valid seems to have less amount resulting in total. However, to encourage Valid and Net status amounts chart can be represented in percentage form instead of absolute data.

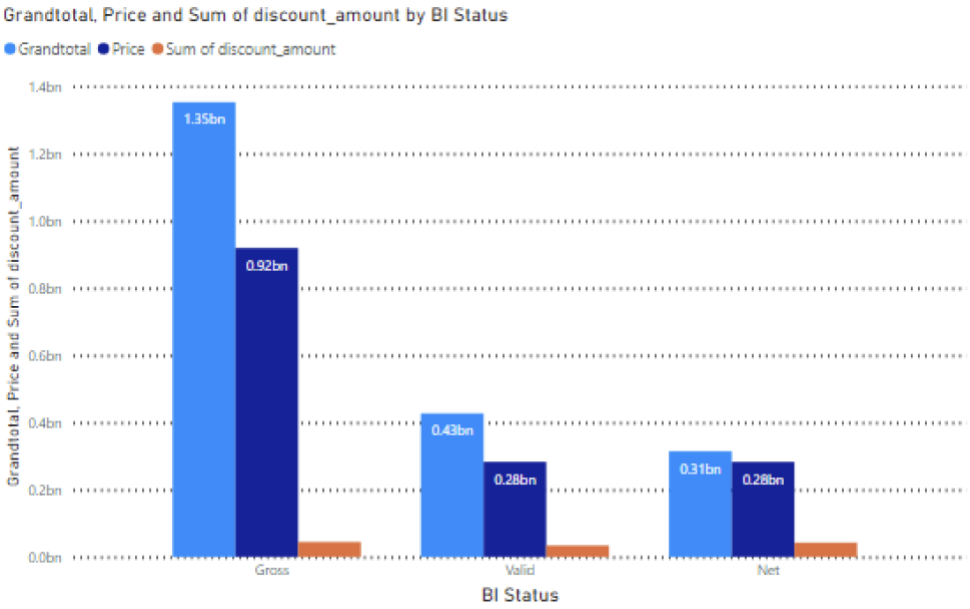

Figure 25 Amount with BI status

Figure 26, represent percentage data of thse variation in BI status. For convincing purpose, the percentage data will show manipulate result amount variation and make people to believe discounts can be seen equally distributed in all BI status.

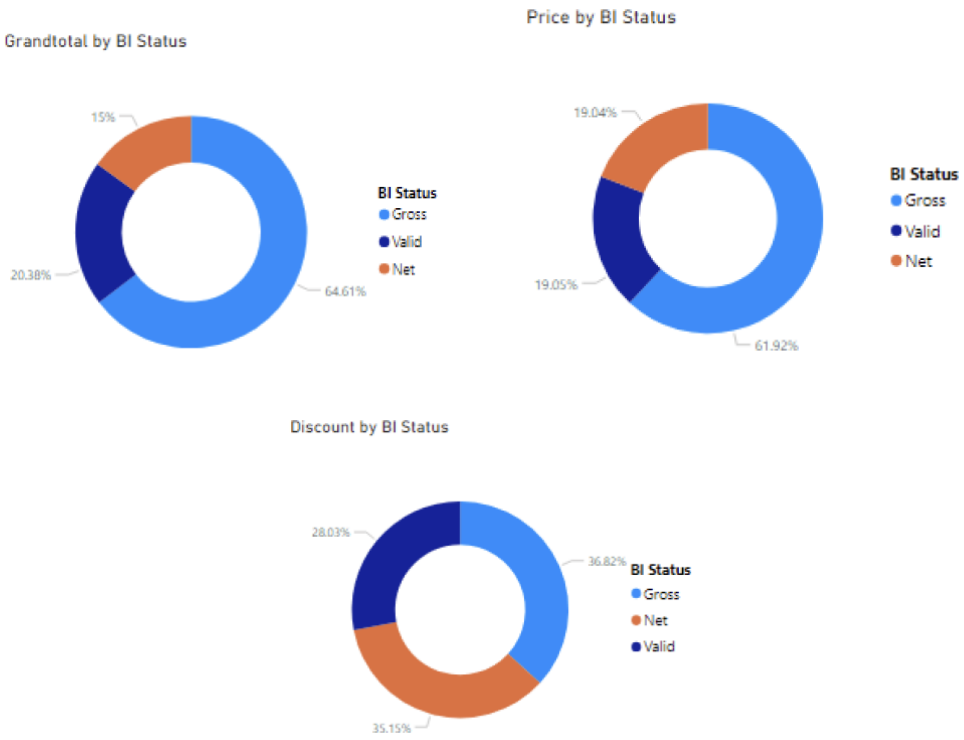

Figure 26 Percentage show

4.1.5 Cherry picking

Another form of manipulation visualizer can make is by highlighting cherry picking techniques. For the below categorical distribution of the order, Mobiles and tablet order has been seen significant whereas the order of book is the least.

Cherry picking for this visualization can be considered in a case where the category of less significance than Mobiles and tablet is shown to be high order quantity. Other category with high order will be removed from the chart and figure after the desired category will be shown. In this highlight case, Beauty and Grooming is shown as top ordered quantity by cherry picking this category and making this as significantly important compared to actual highly ordered quantities than Beauty and Grooming.

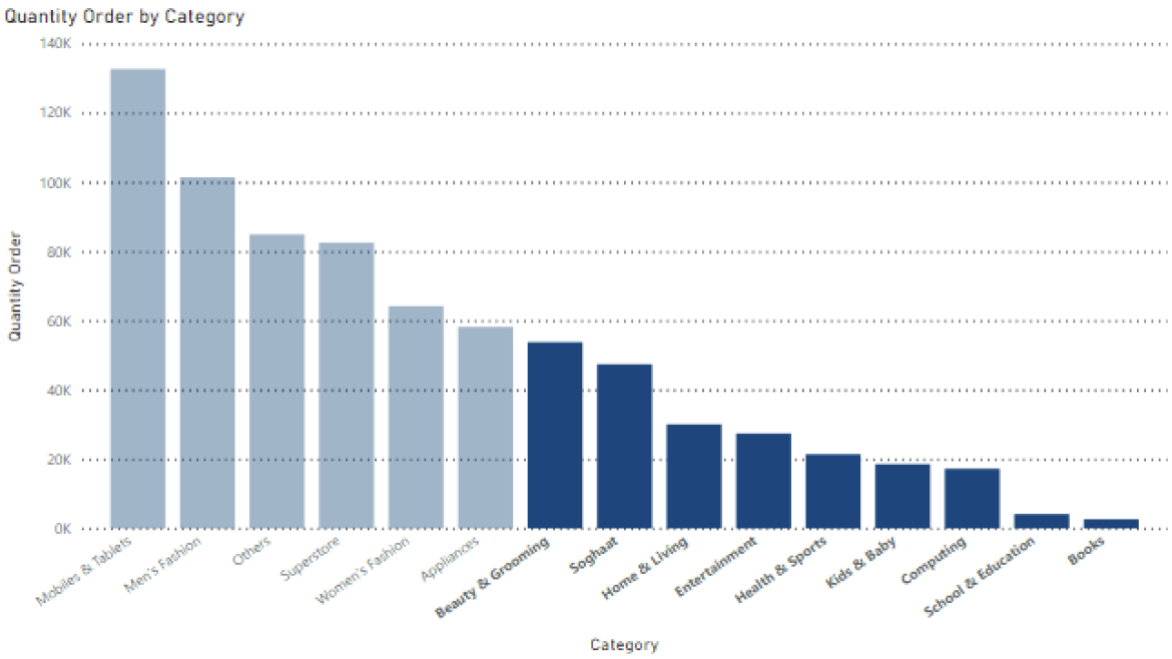

Figure 27 Quantity order for category

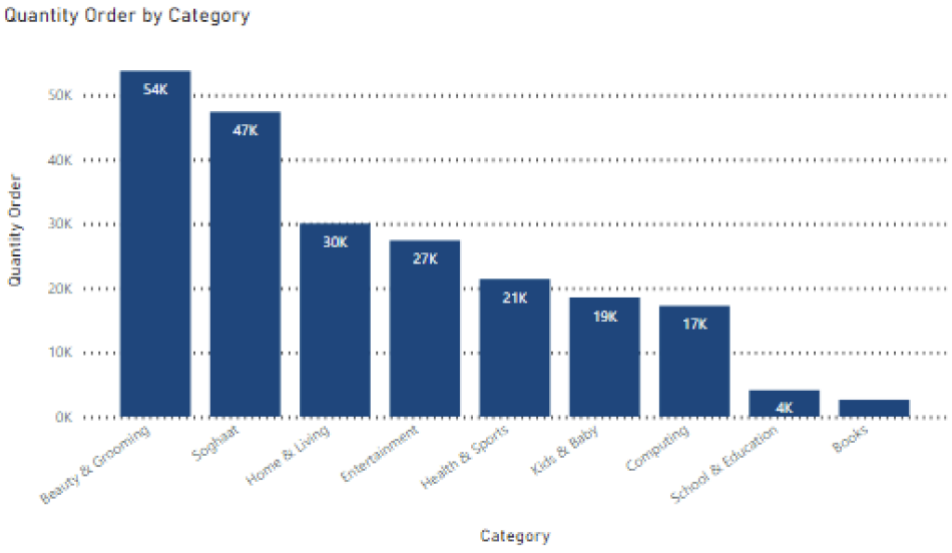

Figure 28 Cherry picking of category

4.1.6 Reference point

Use of reference point technique reports can be manipulated by highlighting different element than desired. For Average of grand total distribution, category Soghaat is seen as least compared to another category. But while presenting the data in visual form, reference of different parameter, quantity order, can be taken and presented in chart. This make manipulation of not showing bad performing category but instead visualization will be showing differently.

Table 1 Average of Grandtotal for categories

Category	Average of Grandtotal	Quantity Order
Mobiles & Tablets	21094.03	132695
Entertainment	20475.88	27419
Computing	12712.34	17251
Appliances	12532.19	58203
Others	6587.23	84916
Women's Fashion	4731.63	64216
Home & Living	3328.44	30065
Kids & Baby	2863.28	18556
Superstore	2571.04	82542
Health & Sports	2388	21420
Beauty & Grooming	2342.05	53790
Men's Fashion	2105.16	101423
School & Education	1792.28	4135
Books	1431.12	2641
Soghaat	1362.6	47418
Total	8586.65	746690

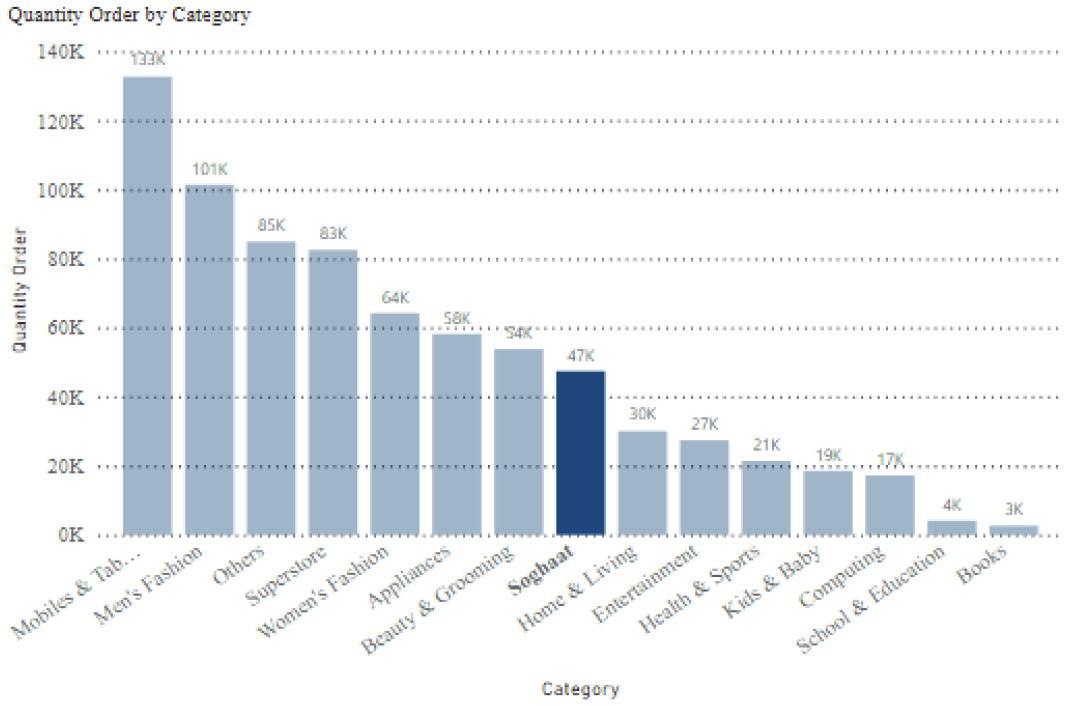

Figure 29 Category reference point

4.2 Result from use of Q&A tool

All of the discussed manipulation techniques and ways to solve then in Power BI can be helpful for user creating reports and managers using to create decision from the visuals. For effectiveness of this user needs to follow each and every precaution to make sure none of the visuals are manipulated. Only after following this the result drawn from the visualization can be believed to have effectiveness as required.

Addition to using and considering each manipulation techniques, with the use of Q&A feature built in Power BI can make task much simpler. User can provide natural language inputs and can use ML feature to get charts and visuals.

By default some suggestion is shown by Power BI itself about the useful metrics chart can be drawn from the data. Upon asking question in the question bar, visualization gets created in Power BI report section.

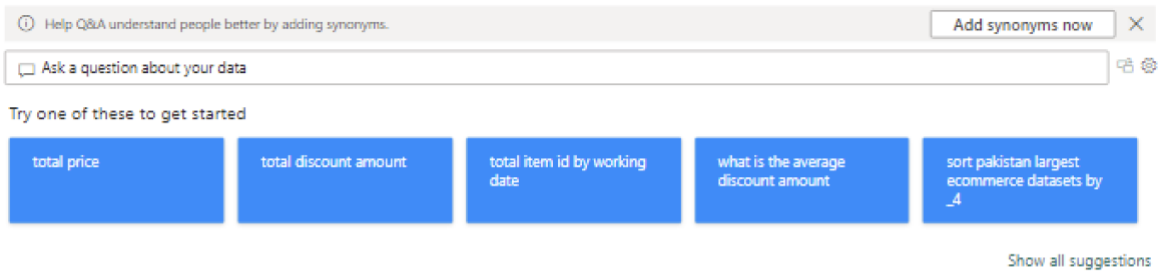

Figure 30 Q&A dialog box

Query about the average discount amount creates the widget value showing average of discount amount from all of the data loaded in the model. Adding more complex prompt creates more complex and useful visuals

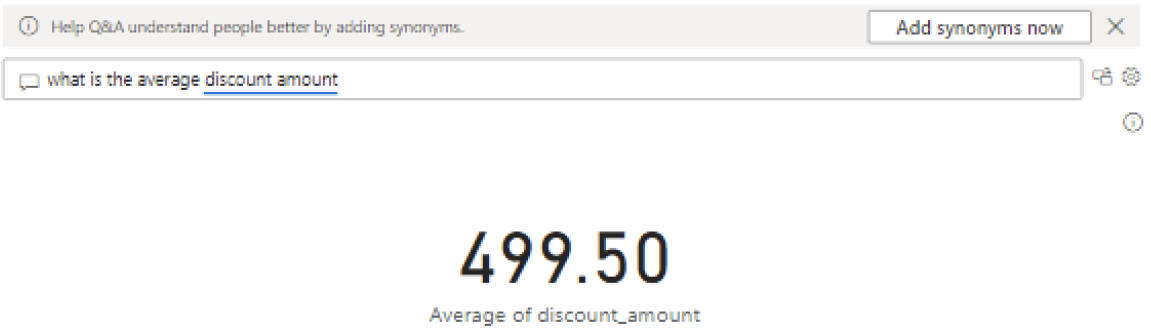

Figure 31 Widget from Q&A

When discount amount needs to be known for categorical cases, then prompt highlighted below showcased the bar chart of average discount amount by categories in the data. These charts are usually unbiased compared to manipulated by the user or visual creator. Managers or decision maker can use this feature without detail knowledge of the power BI and can get desired result output by use of natural language input and using Power BI processing power to generate outputs.

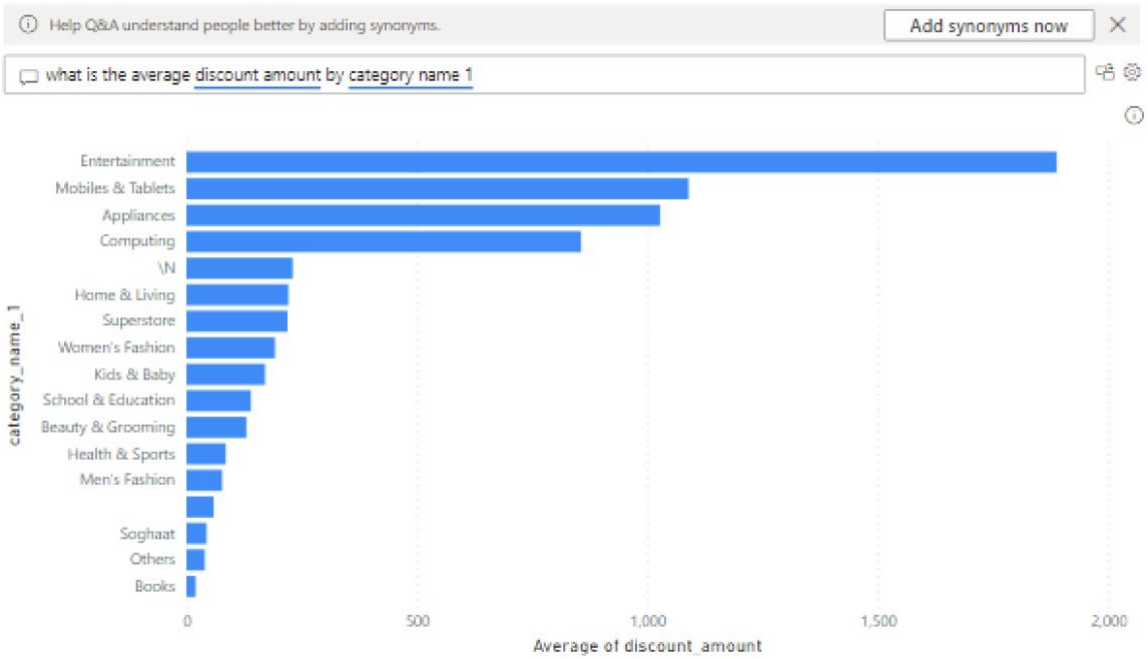

Figure 32 Visual generated by prompt

Using feature of Turn this Q&A result into standard visuals can be used which converts Charts like above mentioned figure into reports. Once visuals are converted into the charts, all of the changes in charts can be made. This includes: font changes, data label modification, chart color changes.

Without detailed knowledge of tool can also get benefits from the feature and make use of the visual reports. This helps all of the business users who are interested in making visual outputs as expected.

4.3 Visualization tools using python

The Data Visualization Application is created in Python program that helps data analysts and researchers visualize data from CSV or Excel files. This interactive tool uses the power of Pandas, Matplotlib, and Tkinter libraries to provide users with a versatile platform for creating various types of data visualizations.

The Data Visualization Application offers a variety of powerful features to help users visualize data:

File selection: When the application is launched, users are prompted to select a data file from their local storage. Supported file formats include CSV and Excel. The selected file is then loaded into the application for further analysis.

- Visualization types: Users can choose from a variety of visualization types to meet their data presentation needs. These options include line plots, bar charts, scatter

plots, pie charts, histograms, and box plots, ensuring versatility in visualizing datasets.

- Data customization: The application allows users to customize their chosen visualization type to suit their preferences. Customization options include specifying the X-axis and Y-axis data columns, providing labels for the axes, setting a chart title, and selecting a color palette for the chart. For scatter plots, users can also specify the marker type.
- Data presentation: The application enhances user-friendliness by presenting the top five rows of the loaded dataset in a well-structured table format. This helps users identify and select the most appropriate columns for visualization purposes.
- Interactive visualizations: After user input and customization, the application quickly generates the selected visualization type with the specified customizations. The resulting chart is displayed interactively using the Matplotlib library, allowing users to explore and analyze their data visually.
- Exit option: Users can choose to exit the program gracefully at any time during their interaction with the application.

Program uses user inputs for each of the parameters and uses such inputs in creation of visual output.

The Data Visualization Application requires users to provide the following inputs:

- File Selection: Users select a data file in CSV or Excel format. The file must contain at least one column of numeric data.
- Visualization Type: Users select a visualization type (line plot, bar chart, scatter plot, pie chart, histogram, or box plot). Each visualization type has its own strengths and weaknesses, so users should select the type that best suits their needs.
- Data Customization: Users provide specific inputs depending on the visualization type, such as the data columns to use, labels, and titles. For example, if users select a line plot, they will need to specify the two data columns that will be used to create the x-axis and y-axis.

Once users have provided all of the required inputs, the Data Visualization Application will generate a visualization of the data. Users can then interact with the visualization to explore the data and gain insights.

Once data in loaded into the mode, user is asked to input alternatives as:

Select a visualization type:

1. Line Plot
2. Bar Chart
3. Scatter Plot
4. Pie Chart
5. Histogram
6. Box Plot
7. Heatmap
8. Quit

Enter your choice (1/2/3/4/5/6/7/8):

User can select required chart type among six alternatives. Each of the chart will again ask for required parameter of x-axis, y-axis and label or legend requirement. Once chart type is selected user is represented with data sample for top five rows.

"Here is the top five data from file, select x-axis, y-axis, labels accordingly", sample of table output from created data set is shown as:

Table 2 Heading data

	Char	**Price**
0	A	$ 48.36
1	B	$ 45.45
2	C	$ 13.38
3	D	$ 98.23
4	C	$ 57.162

From the presented sample table user can accurately add information asked to create visuals. For each of the visual element's inputs are asked which represent variable as,

Enter the X-axis data column name: Value to show in X-axis of visuals

Enter the Y-axis data column name: Value to show in Y-axis of visuals

Enter X-axis label: Data Label for X-axis

Enter Y-axis label: Data Label for Y-axis

Enter chart title: Title of the visual chart.

After all of these inputs are asked from the user, the visual output is plotted using matplotlib in python and plot is created.

For Pie chart and Bar chart, specifically group by summarization is used. This is done to cater the cases where in larger data set, the value of label column and data variable needs to be summarized before displaying in the chart. The use cases, where the detail of Table 2, needs to be presented in Pie chart or Bar chart is needed. Here in the Table the Char data column have multiple rows for same value. Thus, when creating Pie chart and Bar chart for this type of data, this needs to be first aggregated and only after this visualization is possible. Grouping help to aggregate value, for this type of data repeating row values gets summed up for total price and collective value will be shown for repetitive data samples.

Once the data file is loaded in to the model, model ask user to select type of visualization. User can select type of visualization by providing input number in the box. The representation of this is presented in Figure 33.

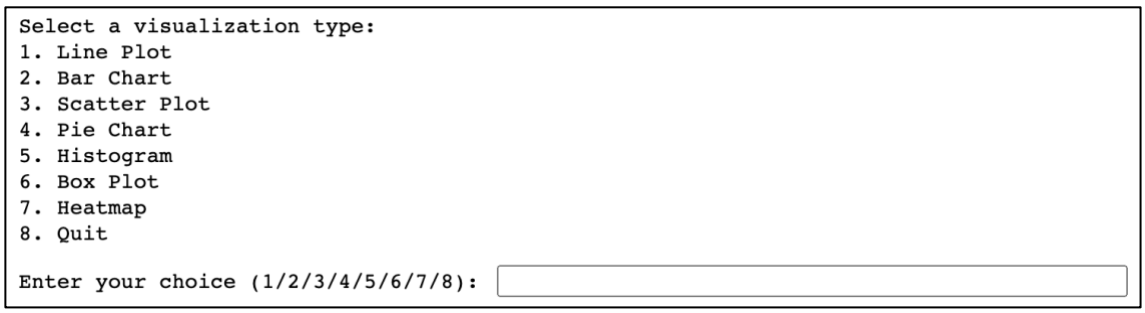

Figure 33 Input from user

Here after the type of chart is selected, the top five heading data sample is presented to the user. This is represented in Figure 34. The Output shows the data sample of the file and can give idea about the data without entirely going through the data in entirety. Another use of showing the data is to use exact heading details in further uses of the visual output processing.

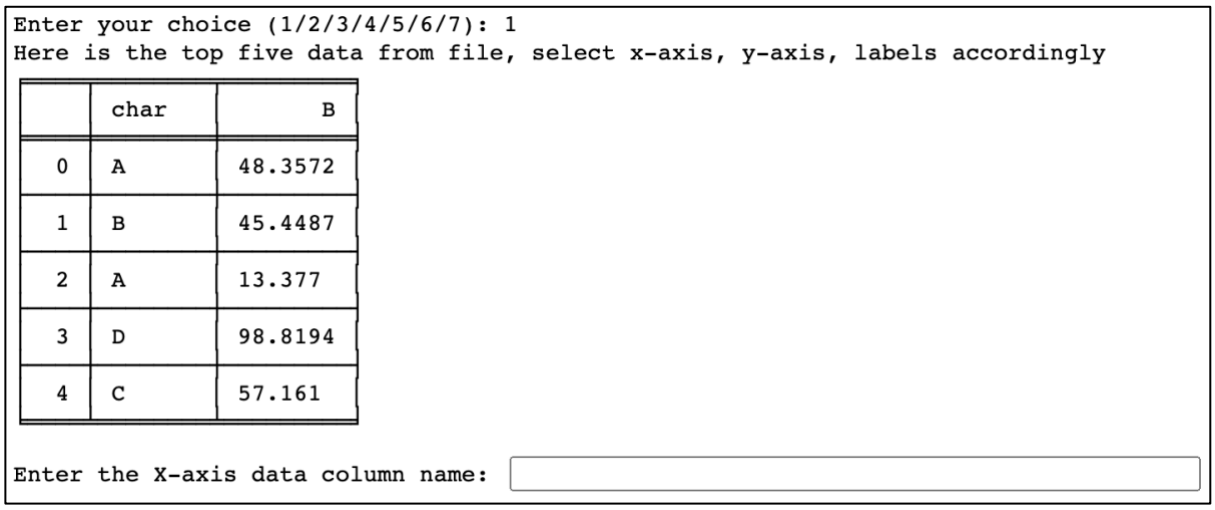

Figure 34 Top 5 data show

Figure 35, shows the input parameter seeking from the user. Here the details of chart detailing are asked from the user. User can add the lables, chart titles as desired for displaying the information.

```
Enter the X-axis data column name: category
Enter the Y-axis data column name: Profit made
Enter X-axis label: Category of product
Enter Y-axis label: Profit made from sales
Enter chart title: Category vs Profit
```

Figure 35 Visual element input

After the details as presented in Figure 35, is used from the user then chart is plotted and displayed to the user. Below figures are the output of the visualization made from the model prepared in the python. Sample of data set used in preparation of below figuresare presented in ANNEX 3.

Here in both of the figure, the output generated are not impacted by the manipulation techniques discussed earlier. The human manipulation is not done here. All of the charts or visualization made are by default and no user manipulation is created.

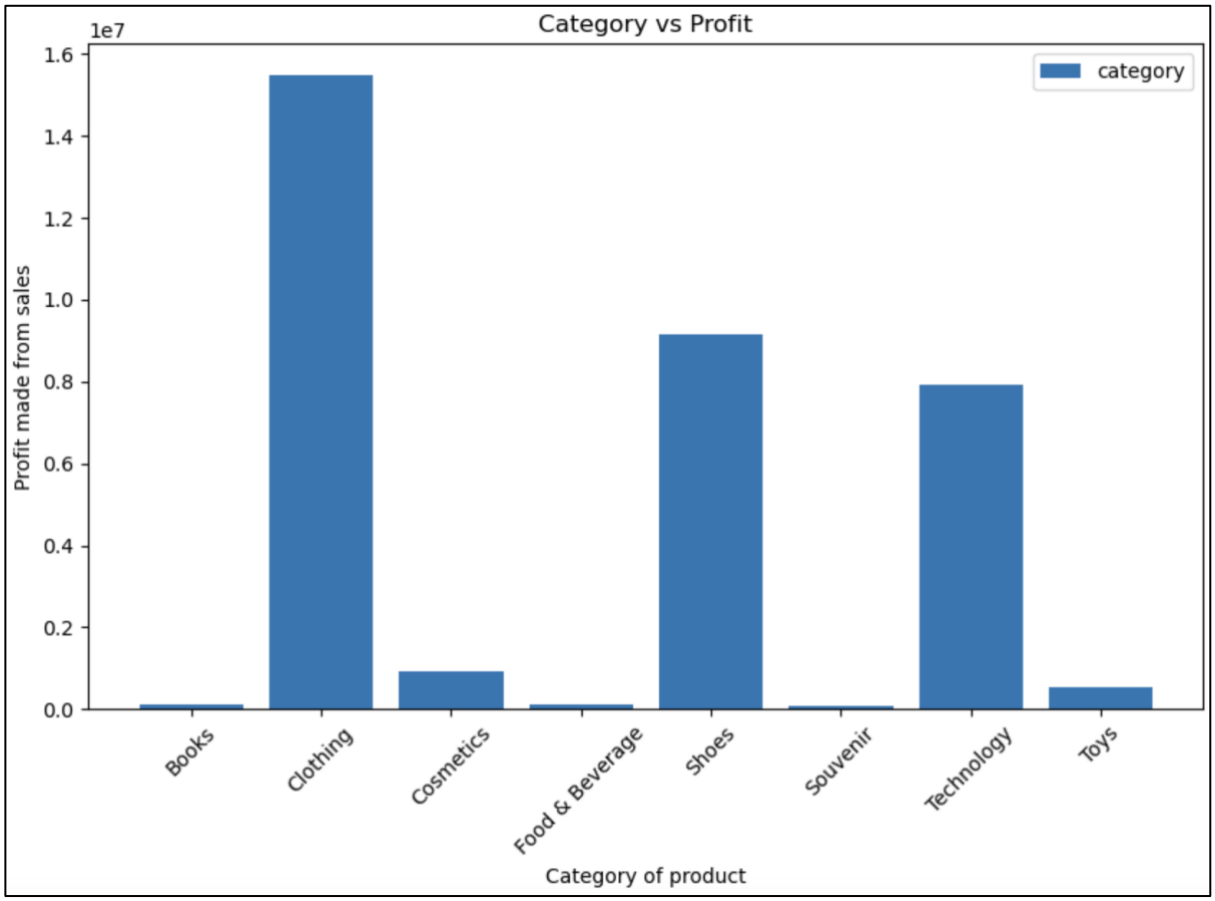

Figure 36 Bar graph as output

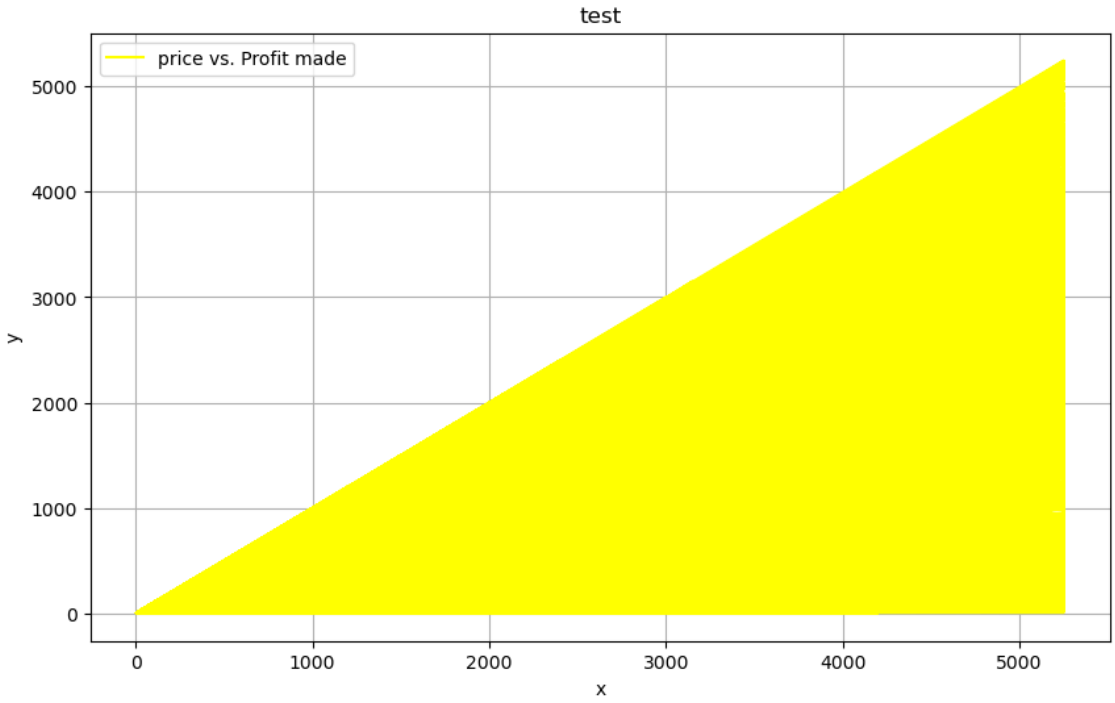

Figure 37 Line chart as output

Figure 37 shows the line graph of two parameter in the data, Price of the product and profit made on sales. With the use of this line chart, regression of two parameter can be drawn. This helps in generating conclusion about the nature of data.

Regarding the cases of data where multiple rows are presented as shown in Figure 38, pie chart can be developed which helps to aggregate function together and show combined result.

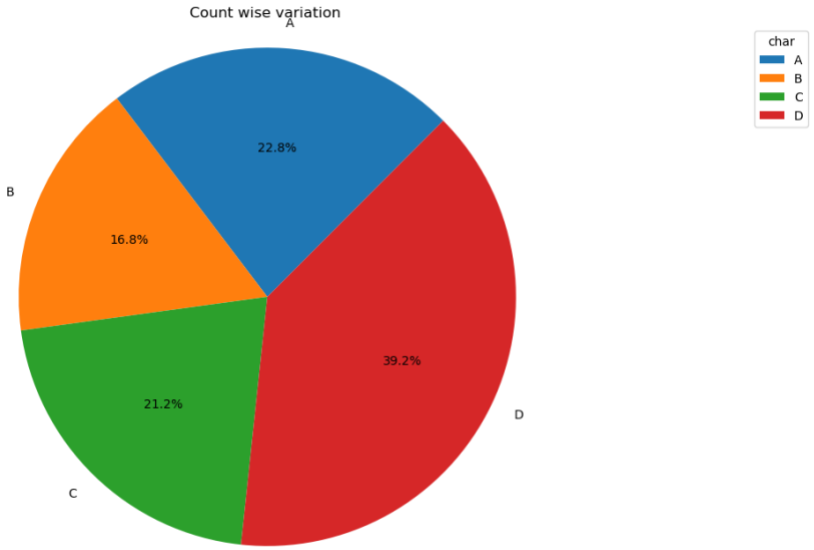

Figure 38 Pie chart from python

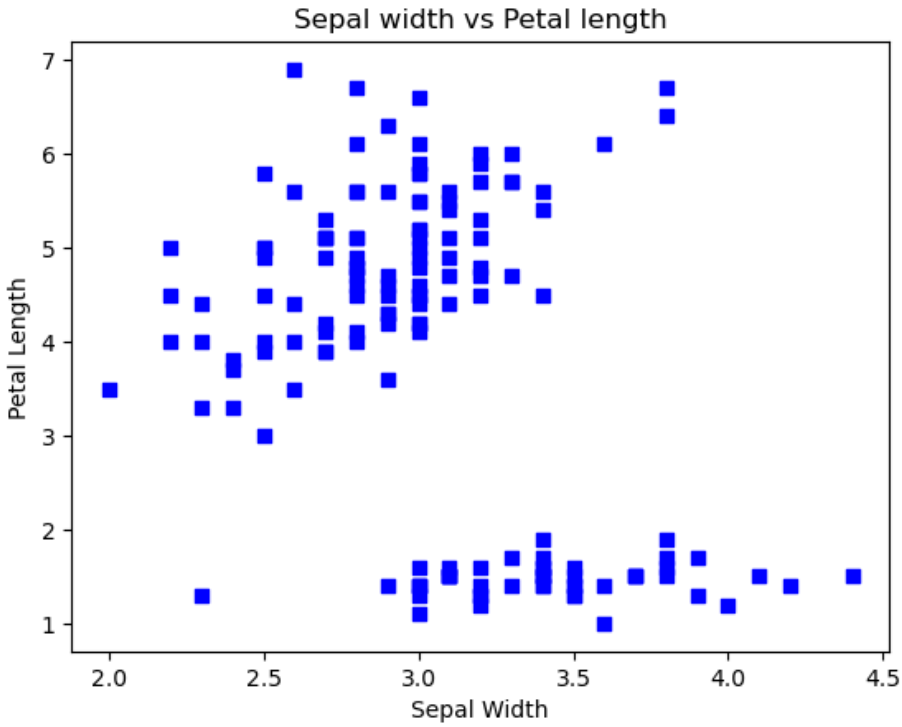

Figure 39 Scatter plot in python

With the Irish dataset of showing characteristics of sepal length, sepal width, petal length, petal width for multiple species is plotted in scatter plot as shown in figure 39. The marker of square is used in the result. This result can be used to make correlation analysis of scatterness in the parameter of sepal width and petal length. The sample data used in this analysis is highlighted in annex 4.

Another type of chart created from python is based on Figure 40 is Heatmaps. The heatmap visualization created by our data analysis code provides a comprehensive overview of the correlation values between the variables in our dataset. The correlation values are numerical representations of the strength and direction of the relationships between pairs of variables. In this context, positive values close to 1 indicate a strong positive correlation, meaning that as one variable increases, the other tends to increase as well. Conversely, negative values close to -1 suggest a strong negative correlation, where an increase in one variable is associated with a decrease in the other. Correlation values close to 0 indicate a weak or no linear relationship between the variables. By examining these values within the heatmap, we can quickly identify which variables are positively, negatively, or not correlated with each other. This insight is invaluable in understanding the interplay of variables and making data-informed decisions. Furthermore, it can guide

feature selection in modeling and reveal potential areas of further investigation, enhancing the depth and quality of our data analysis.

It is to be noted that for same value parameter correlation will be seen as 1. E.g.: Open-Open or any same parameter in x-axis and y-axis will have correlation as 1.

Sample data set used to create Heatmap is added in Annex 5 of the report.

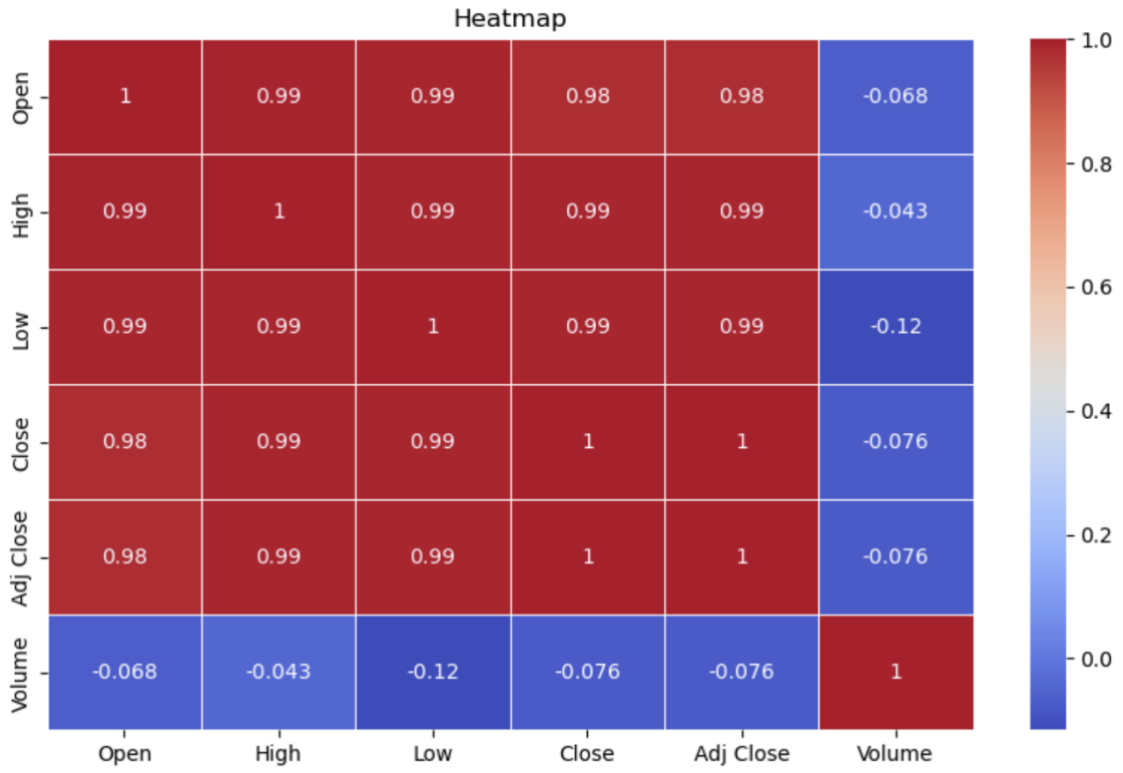

Figure 40 Heatmap

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The research analyses the technology used currently while doing data visualization. From the analysis of the reports and literatures regarding the visual biasness, it was found that generally the person who prepares reports and the one who uses it to make decision was different individual. The task performed by the information visualizer can greatly impact the data output. This leads to inaccurate output in the decision or any task performed by analyzing the visualization presented. Since while studying there are several positive impacts found which are contributed by visualization. Importantly it was found that, visualization improves decision accuracy and quality, visualization improved faster data processing. With the proper training of the user and showcasing effectiveness of the tools, more people are attracted towards it and more accurate result can be generated.

To showcase the biasness which can occur in the day-to-day visuals, data of Pakistan ecommerce was taken. Manipulation technique like change of scale, perspective and color, hiding trends, percentage data instead of actual data, cherry picking, reference points are highlighted using Power BI. Currently used trend and common types of visualization elements are also presented.

By study of this anyone who is preparing any data visualization or the one who uses visualization prepared by other can get benefits of this. And decision from the visualization can be free from any types of biasness. Also, the paper has highlighted the feature within Power BI by which user can generate visual result by giving natural language instruction in Power BI. The importance of learning visualization tool even by manager is found to be effective in current data trend. To cater this scenario, the use of Q&A feature in Power can help which is highlighted in the report. The use of Power BI based feature can play pivotable role while making decisions.

Python based basic visualization tool, with the Matplotlib, is also created which can be used by user without any prior tool or technique understanding. The model can work for data type of any industry and it can be used to generate charts from both excel and csv file format. This model serves basic purpose of visualization where six types of charts are included.

5.2 Recommendations

The research can be further broadened by analyzing the impact post implementation of Power BI feature. Also, by comparing the effectiveness of the decision-making impact

post uses of technique mentioned to remove visualization can be great learning for future analysis. Since the study is not able to analyze post impact assessment after implementation of the mentioned tool, this can be considered in future research activities. The aim should be to analyze impact faced by managers in data driven industry. Some of the important industry this can be analyzed can be highlighted as Ecommerce, Healthcare, Banks or Airlines.

Organization can also play important role in managing the common biasness scenario in the industry. Effective training of the manpower needs to focus on the effectiveness of the visualization. Since this approach is incremental improving which is based on continuous improvement strategy thus with patience and with some effort addition this can be further be improved.

The python model made is also limited to running only by using with necessary prior installed program. With future reference this program can be enhanced further to make it run in real time standalone application. Uses of enhanced API created by software company can be used in companies and get outputs in the real time.

REFERENCES

Michael Dambacher, P. H. ,. D. G. a. R. H., 2023. Graphs versus numbers: How information format affects risk aversion in gambling.

Seungeun Park, B. B. F. S., 2021. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review.

Corentin Burnay, F. D. P. Z., 2019. Special issue: Data visualization for decision-making: an important issue.

Eberhard, K., 2023. The effects of visualization on judgment and decision-making: a systematic literature review. *Karin Eberhard.*

Friendly, A. B. H. o. D. V. M., 2008. A Brief History of Data Visualization.

Insight Software, 2019. *Insightsoftware.* [Online]
Available at: https://insightsoftware.com/blog/a-brief-history-of-data-visualization/#:~:text=As%20the%20demand%20grew%20for,visual%20representation%20of%20statistical%20data

Marita Turpin, N. d. P., 2004. Decision-making Biases and Information Systems.

Salloum, S. A. & Al-Emran, M., 2017. Students' Attitudes Towards the Use of Mobile Technologies in e-Evaluation.

Nickerson, R. S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises.

Tversky, A. & Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases.

Tversky, A. & Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability.

Few, S., 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* s.l.:s.n.

Munzner, T., 2014. *Visualization Analysis and Design.* s.l.:s.n.

Block, G., 2013. Reducing Cognitive Load Using Adaptive Uncertainty Visualization.

Ward, M., Grinstein, G. & Keim, D., 2015. Interactive Data Visualization: Foundations, Techniques, and Applications.

Shneiderman, B., 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.

Isenberg, P., Isenberg, T. & Sedlmair, M., 2013. Think Before You Link: Controlling Visualizations via Hover Modals.

Muller, M. J. &. K. S., 1993. Participatory Design.

Gleicher, M., 2018. Navigating the Challenges of Collaborative Visualization Research.

Heer, J. & Bostock, M., 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design.

Tufte, E. R., 1983. The Visual Display of Quantitative Information.

Few, S., 2012. Show Me the Numbers: Designing Tables and Graphs to Enlighten.

Heer, J. & Shneiderman, B., 2012. Interactive dynamics for visual analysis. *Communications of the ACM.*

Spence, R., 2007. Information Visualization: Design for Interaction.

Card, S. K., Mackinlay, J. D. & Shneiderman, B., 1999. *Readings in Information Visualization: Using Vision to Think.* s.l.:s.n.

Krok, E., 2021. Visualization on charts – manipulations and distortions.

Tufte, E. R., 2001. *The Visual Display of Quantitative Information.* s.l.:s.n.

Richardson, B., n.d. *Acuity Training.* [Online]

Available at: https://www.acuitytraining.co.uk/news-tips/qa-visual-in-power-bi/

McKinney, W., 2011. pandas: a Foundational Python Library for Data Analysis and Statistics. Python for High Performance and Scientific Computing.

Sievert, C., 2020. Plotly. Journal of Open Source Software.

Pedregosa, F. & Varoquaux, G., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research.

Hunter, J. D., 2007. Matplotlib: A 2D graphics environment.

McKinney, W., 2017. Python for Data Analysis.

Microsoft, 2020. *Power BI.* [Online]

Available at: https://powerbi.microsoft.com/en-us/what-is-power-bi/

Devi, N., 2022. *Scalar Topics.* [Online]

Available at: https://www.scaler.com/topics/pandas/what-is-pandas-in-python/

Scott, B., 2016. *Harvard business review.* [Online]

Available at: https://hbr.org/2016/06/visualizations-that-really-work

Schetinger, M., Bartolomeo, S. D. & Assady, M. E., 2023. Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models. *Eurographics Conference on Visualization.*

Nurhidayat, R. M., 2020. *Towards Data Science.* [Online]

Available at: https://towardsdatascience.com/visualizations-with-matplotlib-part-1-c9651008b6b8

Anon., n.d. *Capturly.* [Online]
Available at: https://capturly.com/guides/ultimate-guide-to-heatmaps-everything-you-need-to-know-about-website-heatmaps/
Tran, T. L., 2023. *Vecteezy.* [Online]
Available at: https://www.vecteezy.com/vector-art/26117917-different-types-of-charts-and-graphs-vector-set-column-pie-area-line-graphs-data-analysis-financial-report-business-analytics-illustration-infographic-statistics-graph
Chintakindi, S., 2022. *Tableau.* [Online]
Available at: https://public.tableau.com/app/profile/santosh.chintakindi/viz/NetworkDiagramofHealthcareStakeholders/Sheet1
USAID, 2022. *USAID.* [Online].
Microsoft, 2023. *Learn microsoft.* [Online]
Available at: https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-treemaps?tabs=powerbi-desktop
Outlook India, 2023. *Outlook India.* [Online].
Ihi org, 2018. *Ihi.* [Online]
Available at: https://www.med.unc.edu/neurosurgery/wp-content/uploads/sites/460/2018/10/Flow-chart-Process-Flow.pdf
Capturly, 2023. *Capturly.* [Online]
Available at: https://capturly.com/guides/ultimate-guide-to-heatmaps-everything-you-need-to-know-about-website-heatmaps/
Spotfire, 2022. *TIBCO.* [Online]
Available at: https://www.tibco.com/reference-center/what-is-a-scatter-chart
Microsoft, 2023. *Learn microsoft.* [Online]
Available at: https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview
pandas, 2021. [Online]
Available at: https://pandas.pydata.org/docs/_sources/user_guide/index.rst.txt
python, 2021. *Python software foundation.* [Online]
Available at: https://docs.python.org/3/library/tkinter.html

ANNEX 1

Python code for creating visual from Matplot:

```
import pandas as pd
import matplotlib.pyplot as plt
import tkinter as tk
import seaborn as sns
from tkinter import filedialog
from tabulate import tabulate

# Create a Tkinter root window
root = tk.Tk()
root.withdraw()  # Hide the main root window

# Ask the user to select a file using a file dialog
file_path = filedialog.askopenfilename(title="Select a file (CSV or Excel)")

if not file_path:
    print("No file selected. Exiting...")
    exit()

# Determine the file format based on the file extension
file_extension = file_path.split(".")[-1].lower()

try:
    if file_extension == "csv":
        data = pd.read_csv(file_path)
    elif file_extension == "xlsx":
        data = pd.read_excel(file_path)
    else:
        print("Unsupported file format. Please provide a CSV or Excel file.")
        exit()
except Exception as e:
```

```
    print(f"Error reading the file: {str(e)}")
    exit()

# Print the column names for reference
print("Columns in the data:")
print(data.columns)

# Print the top 5 rows of the data for reference
print("Top 5 rows of the data:")
print(data.head())

# Enter the main loop for user interaction
while True:
    print("Select a visualization type:")
    print("1. Line Plot")
    print("2. Bar Chart")
    print("3. Scatter Plot")
    print("4. Pie Chart")
    print("5. Histogram")
    print("6. Box Plot")
    print("7. Heatmap")
    print("8. Quit")

    # Prompt the user to enter their choice
    choice = input("Enter your choice (1/2/3/4/5/6/7): ")

    # Display the top five data rows for reference when choosing columns
    print("Here is the top five data from file, select x-axis, y-axis, labels accordingly")
    print(tabulate(data.head(), headers='keys', tablefmt='fancy_grid'))

    if choice == "1":
        # Create a line plot
        x_column = input("Enter the X-axis data line name: ")
        y_column = input("Enter the Y-axis data line name: ")
```

```
    x_label = input("Enter X-axis label: ")
    y_label = input("Enter Y-axis label: ")
    chart_color = input("Enter chart color: ")
    chart_title = input("Enter chart title: ")

    x, y = data[x_column], data[y_column]
    plt.figure(figsize=(10, 6))
    plt.plot(x, y, color=chart_color, label=f'{x_column} vs. {y_column}')
    plt.xlabel(x_label)
    plt.ylabel(y_label)
    plt.title(chart_title)
    plt.legend()
    plt.grid(True)
    plt.show()

elif choice == "2":
    # Create a bar chart
    x_column = input("Enter the X-axis data column name: ")
    y_column = input("Enter the Y-axis data column name: ")
    x_label = input("Enter X-axis label: ")
    y_label = input("Enter Y-axis label: ")
    chart_title = input("Enter chart title: ")

    grouped_data = data.groupby(x_column)[y_column].sum().reset_index()

    x = grouped_data[x_column]
    y = grouped_data[y_column]
    plt.figure(figsize=(10, 6))
    plt.bar(x, y, label=x_column)
    plt.xlabel(x_label)
    plt.ylabel(y_label)
    plt.title(chart_title)
    plt.xticks(rotation=45)
    plt.legend()
```

```
    plt.show()

elif choice == "3":
    # Create a scatter plot
    x_column = input("Enter the X-axis data column name: ")
    y_column = input("Enter the Y-axis data column name: ")
    x_label = input("Enter X-axis label: ")
    y_label = input("Enter Y-axis label: ")
    chart_title = input("Enter chart title: ")
    chart_color = input("Enter chart color (e.g., 'red', 'blue'): ")
    chart_marker = input("Enter marker type(e.g., 'o','s', +):")

    x = data[x_column]
    y = data[y_column]

    plt.scatter(x, y, marker=chart_marker, color=chart_color)
    plt.xlabel(x_label)
    plt.ylabel(y_label)
    plt.title(chart_title)
    plt.show()

elif choice == "4":
    # Create a pie chart
    label_column = input("Enter the data column name for labels: ")
    size_column = input("Enter the data column name for sizes: ")
    chart_title = input("Enter chart title: ")

    grouped_data = data.groupby(label_column)[size_column].sum().reset_index()
    labels = grouped_data[label_column]
    sizes = grouped_data[size_column]

    plt.figure(figsize=(8, 8))  # Adjust figure size as needed for better distribution
    plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=45)
    plt.axis('equal')  # Equal aspect ratio ensures that the pie is drawn as a circle.
```

```
        plt.title(chart_title)
        plt.legend(labels, title=label_column, loc="best", bbox_to_anchor=(1, 0, 0.5, 1))  # Adjust legend position for best readability
        plt.show()

    elif choice == "5":
        # Create a histogram
        data_column = input("Enter the data column name for the histogram: ")
        x_label = input("Enter X-axis label: ")
        y_label = input("Enter Y-axis label: ")
        chart_title = input("Enter chart title: ")
        chart_color = input("Enter chart color (e.g., 'red', 'blue'): ")

        x = data[data_column]

        plt.hist(x, color=chart_color)
        plt.xlabel(x_label)
        plt.ylabel(y_label)
        plt.title(chart_title)
        plt.show()

    elif choice == "6":
        # Create a box plot
        data_column = input("Enter the data column name for the box plot: ")
        x_label = input("Enter X-axis label: ")
        y_label = input("Enter Y-axis label: ")
        chart_title = input("Enter chart title: ")
        chart_color = input("Enter chart color (e.g., 'red', 'blue'): ")

        x = data[data_column]

        plt.boxplot(x, vert=False, patch_artist=True, boxprops=dict(facecolor=chart_color))
        plt.xlabel(x_label)
```

```
    plt.title(chart_title)
    plt.show()

if choice == "7":
    # Create a heatmap
    heatmap_data = data.corr()  # Use the correlation matrix as heatmap data
    chart_title = input("Enter chart title: ")

    plt.figure(figsize=(10, 6))
    sns.heatmap(heatmap_data, annot=True, cmap="coolwarm", linewidths=.5)
    plt.title(chart_title)
    plt.show()

elif choice == "8":
    break
else:
    print("Invalid choice. Please enter a valid option.")
```

ANNEX 2

Custome	1	2	3	4	5	6	7	6	8	9
FY	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17
M-Y	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16
Custome	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7
Month	7	7	7	7	7	7	7	7	7	7
Year	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016
MV	1,950	240	2,450	360	1,110	80	360	170	96,499	96,499
BI	Gross	Gross	Gross	Net	Valid	Gross	Net	Net	Gross	Gross
Workin	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
payment	cod	cod	cod	cod	cod	cod	cod	cod	ublcredit	mygatew
discount	0	0	0	300	0	0	300	0	0	0
sales_co	\N	\N	\N	R-FSD-	\N	\N	\N	\N	\N	\N
category	Women's	Beauty	Women's	Beauty	Soghaat	Soghaat	Beauty	Soghaat	Mobiles	Mobiles
increme	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474
grand_t	1950	240	2450	60	1110	80	60	170	96499	96499
qty_ord	1	1	1	1	2	1	1	1	1	1
price	1950	240	2450	360	555	80	360	170	96499	96499
sku	kreations	kcc_Buy	Ego_UP	kcc_kron	BK7010	UK_Na	kcc_kron	UK_Na	Apple	Apple
created_	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
status	complete	canceled	canceled	complete	order_ref	canceled	complete	complete	canceled	canceled
item_id	211131	211133	211134	211135	211136	211137	211138	211139	211140	211141

9	10	10	11	12	11	13	13	13	14	14	14
FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17
Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16
2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7
7	7	7	7	7	7	7	7	7	7	7	7
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016
5,500	210	156	120	320	1,550	420	360	490	899	899	320
Net	Valid	Valid	Net	Net	Gross	Net	Net	Net	Gross	Gross	Gross
07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
cod	cod	cod	ublcredit	customer	ublcredit	cod	cod	cod	cod	cod	cod
0	0	0	0	0	0	0	0	0	0	0	0
\N	\N	\N	105259	\N	105259	R-KHW-	R-KHW-	R-KHW-	\N	\N	\N
Applianc	Soghaat	Soghaat	Home &	Beauty	Men's	Soghaat	Soghaat	Beauty	Home &	Home &	Home &
1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474
5500	366	366	120	0	1550	1270	1270	1270	2118.25	2118.25	2118.25
1	1	1	1	1	1	1	1	1	1	1	1
5500	210	156	120	320	1550	420	360	490	899.25	899	320
GFC_Pe	BK1070	BK1130	kcc_Sult	kcc_gla	Assetme	cr_DAT	UK_Gift	itter_AB	RL_B00	bed&rest	L&L_LL
07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
complete	received	received	complete	complete	canceled	complete	complete	complete	canceled	canceled	canceled
211142	211143	211144	211145	211146	211147	211149	211150	211151	211152	211153	211154

15	15	16	17	17	16	16	18
FY17	FY17	FY17	FY17	FY17	FY17	FY17	FY17
Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16	Jul-16
2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7	2016-7
7	7	7	7	7	7	7	7
2016	2016	2016	2016	2016	2016	2016	2016
149	149	1,000	1,913	1,913	500	500	3,000
Net	Net	Valid	Valid	Valid	Net	Net	Gross
07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
cod	cod	customer	cod	cod	customer	customer	cod
0	0	0	0	0	0	0	0
\N	\N	\N	\N	\N	\N	\N	\N
Kids &	Kids &	\N	Men's	Men's	Others	Others	Men's
1001474	1001474	1001474	1001474	1001474	1001474	1001474	1001474
298	298	0	3826	3826	0	0	3000
1	1	1	1	1	1	5	2
149	149	1000	1913	1913	500	100	1500
J&J_JJR	J&J_JJR	D Lend a	Mochika	Mochika	SKMT_	SKMT_	sputnik_
07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20	07/01/20
complete	complete	order_ref	order_ref	order_ref	complete	complete	canceled
211155	211156	211157	211158	211160	211162	211163	211164

ANNEX 3

Profit made	1139	758	187	2587	24	555	26	439	540	463
Return flag	1	0	1	0	1	0	1	0	1	0
delivered date	08/08/2022	08/01/2022	19/11/2021	18/05/2021	21/11/2021	29/05/2022	30/03/2022	01/02/2021	05/12/2021	02/09/2021
Shipping charge	75.02	90.0255	15.004	150.0425	3.03	75.02	2.033	30.008	45.012	30.008
shopping_mall	Kanyon	Forum Istanbul	Metrocity	Metropol AVM	Kanyon	Forum Istanbul	Istinye Park	Mall of Istanbul	Metrocity	Kanyon
order date	8/5/2022	12/12/2021	11/9/2021	5/16/2021	10/24/2021	5/24/2022	3/13/2022	1/13/2021	11/4/2021	8/22/2021
Payment method	Credit Card	Debit Card	Cash	Credit Card	Cash	Credit Card	Cash	Credit Card	Credit Card	Credit Card
price	1500.4	1800.51	300.08	3000.85	60.6	1500.4	40.66	600.16	900.24	600.16
qty	5	3	1	5	4	5	1	2	3	2
category	Clothing	Shoes	Clothing	Shoes	Books	Clothing	Cosmetics	Clothing	Clothing	Clothing
age	28	21	20	66	53	28	49	32	69	60
Gender	Female	Male	Male	Female	Female	Female	Female	Female	Male	Female
Customer	C241288	C111565	C266599	C988172	C189076	C657758	C151197	C176086	C159642	C283361
Order	I138884	I317333	I127801	I173702	I337046	I227836	I121056	I293112	I293455	I326945

8	0	44	160	140	2667	19	9	232	19	242	2
0	0	1	0	0	1	0	0	1	1	0	0
24/01/2023	12/11/2022	04/08/2022	20/11/2022	28/06/2022	08/11/2021	10/02/2021	05/02/2022	13/08/2021	04/04/2023	28/02/2023	27/05/2021
0.523	0.7575	7.168	30.008	30.008	150.0425	1.515	0.7845	30.008	7.168	15.004	0.2615
Metrocity	Emaar Square	Metrocity	Cevahir AVM	Kanyon	Viaport Outlet	Metrocity	Zorlu Center	Metropol AVM	Cevahir AVM	Emaar Square	Cevahir AVM
12/25/2022	10/28/2022	7/31/2022	11/17/2022	6/3/2022	11/7/2021	1/16/2021	1/5/2022	7/26/2021	3/7/2023	2/15/2023	5/1/2021
Cash	Credit Card	Debit Card	Cash	Credit Card	Credit Card	Debit Card	Credit Card	Cash	Cash	Credit Card	Cash
10.46	15.15	143.36	600.16	600.16	3000.85	30.3	15.69	600.16	143.36	300.08	5.23
2	1	4	2	2	5	2	3	2	4	1	1
Food & Beverage	Books	Toys	Clothing	Clothing	Shoes	Books	Food & Beverage	Clothing	Toys	Clothing	Food & Beverage
36	29	67	25	67	24	65	42	46	24	23	27
Female	Female	Female	Male	Female	Male	Male	Female	Female	Male	Male	Female
C240286	C191708	C225330	C312861	C555402	C362288	C300786	C330667	C218149	C196845	C220180	C125696
I306368	I139207	I640508	I179802	I336189	I688768	I294687	I195744	I993048	I992454	I183746	I412481

ANNEX 4

Irish dataset

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5.0	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3.0	1.4	0.1	Iris-setosa
14	4.3	3.0	1.1	0.1	Iris-setosa
15	5.8	4.0	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.3	Iris-setosa
20	5.1	3.8	1.5	0.3	Iris-setosa
21	5.4	3.4	1.7	0.2	Iris-setosa
22	5.1	3.7	1.5	0.4	Iris-setosa
23	4.6	3.6	1.0	0.2	Iris-setosa
24	5.1	3.3	1.7	0.5	Iris-setosa
25	4.8	3.4	1.9	0.2	Iris-setosa
26	5.0	3.0	1.6	0.2	Iris-setosa

ANNEX 5

DOW dataset

Date	Open	High	Low	Close	Adj Close	Volume
Jun 23, 2023	33,835.66	33,835.66	33,646.49	33,727.43	33,727.43	379,800,000
Jun 22, 2023	33,900.47	34,003.56	33,835.39	33,946.71	33,946.71	275,600,000
Jun 21, 2023	33,990.56	34,097.93	33,876.17	33,951.52	33,951.52	323,690,000
Jun 20, 2023	34,206.66	34,206.66	33,915.93	34,053.87	34,053.87	334,810,000
Jun 16, 2023	34,464.02	34,588.68	34,285.69	34,299.12	34,299.12	595,230,000
Jun 15, 2023	33,945.98	34,488.98	33,945.98	34,408.06	34,408.06	361,400,000
Jun 14, 2023	34,044.70	34,151.42	33,783.55	33,979.33	33,979.33	354,910,000
Jun 13, 2023	34,111.08	34,310.28	34,107.98	34,212.12	34,212.12	308,790,000
Jun 12, 2023	33,906.80	34,077.84	33,878.46	34,066.33	34,066.33	308,130,000

ANNEX 6

A Study on usages of visualization tools to remove biases in decision making

ORIGINALITY REPORT

11%
SIMILARITY INDEX

PRIMARY SOURCES

1	link.springer.com Internet	117 words — 1%
2	bg.hbr.org Internet	105 words — 1%
3	www.dundas.com Internet	69 words — 1%
4	www.coursehero.com Internet	61 words — 1%
5	www.futurelearn.com Internet	58 words — 1%
6	learn.microsoft.com Internet	53 words — 1%
7	mite.ac.in Internet	47 words — < 1%
8	nsuworks.nova.edu Internet	43 words — < 1%
9	www.researchgate.net Internet	39 words — < 1%

10	www.link-de.com Internet	35 words — < 1%
11	V. Schetinger, S. Di Bartolomeo, M. El-Assady, A. McNutt, M. Miller, J. P. A. Passos, J. L. Adams. "Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models", Computer Graphics Forum, 2023 Crossref	29 words — < 1%
12	Karin Eberhard. "The effects of visualization on judgment and decision-making: a systematic literature review", Management Review Quarterly, 2021 Crossref	26 words — < 1%
13	emprovisetech.blogspot.com Internet	26 words — < 1%
14	www.ijcua.com Internet	26 words — < 1%
15	pt.scribd.com Internet	23 words — < 1%
16	www.medien.ifi.lmu.de Internet	21 words — < 1%
17	favtutor.com Internet	19 words — < 1%
18	glisc.info Internet	16 words — < 1%
19	Phi-Hung Nguyen, Thi-Ly Nguyen, Chia-Nan Wang, Minh-Duc Vu, Lan-Anh Thi Nguyen, Hong-Anh Pham, Mai-Anh Thi Pham, Hong- Quan Le. "Linking Investment Decisions-Based on Firm Performance and Open	15 words — < 1%

	Innovation Practices In Vietnam's Wire and Cable Market Using Data Envelopment Analysis Models", Journal of Open Innovation: Technology, Market, and Complexity, 2023 Crossref	
20	repository.enp.edu.dz Internet	15 words — < 1%
21	zebrabi.com Internet	15 words — < 1%
22	Michael Dambacher, Peter Haffke, Daniel Groß, Ronald Hübner. "Graphs versus numbers: How information format affects risk aversion in gambling", Judgment and Decision Making, 2023 Crossref	14 words — < 1%
23	ebin.pub Internet	14 words — < 1%
24	slideplayer.com Internet	14 words — < 1%
25	store.hbr.org Internet	14 words — < 1%
26	webapps.itc.utwente.nl Internet	14 words — < 1%
27	"Data Visualization: An Untapped Potential for Political Participation and Civic Engagement", Lecture Notes in Computer Science, 2015. Crossref	12 words — < 1%
28	pulmonary.ucsf.edu Internet	12 words — < 1%

29	1library.net Internet	11 words — < 1%
30	www.annauniv.edu Internet	11 words — < 1%
31	www.innovationjn.com Internet	10 words — < 1%
32	Nanddeep Sadanand Nachan, Smita Sadanand Nachan. "Understanding Hybrid Environments in SharePoint 2019", Springer Science and Business Media LLC, 2020 Crossref	9 words — < 1%
33	dr.ntu.edu.sg Internet	9 words — < 1%
34	repository.iaa.ac.tz:8080 Internet	9 words — < 1%
35	ts2.space Internet	9 words — < 1%
36	www.journalism.co.uk Internet	9 words — < 1%
37	automated-polishing.eu Internet	8 words — < 1%
38	dissertations.mak.ac.ug Internet	8 words — < 1%
39	hevodata.com Internet	8 words — < 1%
	medium.com	

40	Internet	8 words — < 1%
41	opus.govst.edu Internet	8 words — < 1%
42	webthesis.biblio.polito.it Internet	8 words — < 1%
43	www.atarimagazines.com Internet	8 words — < 1%
44	www.ncbi.nlm.nih.gov Internet	8 words — < 1%
45	www.udemy.com Internet	8 words — < 1%
46	Reza Rad. "Pro Power BI Architecture", Springer Science and Business Media LLC, 2023 Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF