A Data-driven Approach to Sales Analysis

V. Vijayakumar[1,*], Priyadharshini Durga B.[2]

Abstract

Decisions made using data from digital sources are said to be data-driven when they are analysed and interpreted. Across many sectors, a data-driven approach is an effective technique for gaining insights, making wise choices, and guiding corporate strategy. This study covers the concept of data analytics in sales analysis of bakery and mess. It involves evaluating diverse types of information, including sales data, customer preferences, production costs, and supplier details. A summary of data analytics' methods, procedures, and applications is presented. The real time data set is collected from the bakery and mess for the sales analysis purpose. Data visualization techniques are frequently used in exploratory data analysis (EDA), which examines, analyses, and summarizes data sets' primary features. The data-driven strategy entails systematically gathering and examining a variety of data, from sales and inventories to defensive actions. The study presents findings and insights after the data preparation and methodology. Sales Performance Dashboard is prepared. Businesses may monitor and identify patterns and trends in their sales performance by using dashboard and also sales teams can monitor key performance metrics, identify problem areas, and make timely, data-driven choices to enhance their sales methods.

Keywords: Sales analysis, data analytics, excel analysis, business analytics, exploratory data analysis, descriptive data analysis, data-driven approach

INTRODUCTION

In order to provide better services to the customers and clients, businesses might use data-driven strategies for data analysis and management. By using data to inform its decisions, a company may use contextualization and/or personalization to communicate with prospects and customers in a way that is more focused on their needs. Data analytics is the process of examining unprocessed data to extract valuable, actionable insights that can be utilized to guide and advise smart business decisions. One may use it to interpret the past and forecast patterns and behaviours in the future [1].

Steps of Data Analytics

The steps of data analytics can vary depending on the specific problem or question being addressed, but generally, the process involves the following steps [2–4]:

Describe the Problem or Research Question

The data analyst is given a problem or business task in the first phase of the process. The analyst must comprehend the problem and the expectations of the stakeholders for the resolution. A stakeholder is an individual who has committed their financial resources and other assets to support a project. To identify the best solution for their issue, the analyst has to be able provide many questions. To know the problem completely, the analyst must identify its underlying cause.

*Author for Correspondence
V. Vijayakumar
E-mail: veluvijay20@gmail.com

[1]Head, Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India
[2]Student, Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India

Gather Data

Gathering or Preparing the Data is the second phase. This stage entails gathering information and keeping it for further study. The analyst is required to collect information from diverse outlets based on the assigned task. It is essential to acquire data from both internal and external sources. Internal data pertains to information available within the company, while external data refers to information outside the organization. First-party data is the data collected by an individual using their own resources. Second-party data is data gathered and sold by the provider. Third-party data is information collected from external sources. Surveys, questionnaires, feedback forms, and interviews are among the frequently used sources of data collection. A SQL database or spreadsheet can be used to store the gathered data.

Data Cleaning

Clean and Process Data is the third phase. Once information has been gathered from various sources, it is time to clean it up. Clear data lacks typos, repetitions, and unnecessary information. Data integrity is a major factor in clean data. The redundant and unneeded data is eliminated and cleaned since it may exist in several formats or be duplicated [5].

Data Analysis

Analysing is the fourth phase in the process. The cleaned data is employed in trend analysis and identification. To get better outcomes, it also aggregates data and does computations. Excel or SQL are the programs used to do computations. These tools provide in-built functions to perform calculations or sample code is written in SQL to perform calculations. Using Excel, we can create pivot tables and perform calculations while SQL creates temporary tables to perform calculations.

Data Visualization

In the fifth stage, the focus is on data visualization, recognizing its persuasive impact. The data, having undergone transformation, needs to be translated into a visual format, typically through charts or graphs. The objective of generating data visualizations is to facilitate comprehension, especially for non-technical stakeholders who are primarily individuals. Tableau and Looker stand out as widely utilized tools for crafting engaging data visualizations, with Tableau being a user-friendly tool employing a drag-and-drop approach for visualization creation.

Presenting the Data

Getting raw data into a format that may be readily understood and useful by a variety of stakeholders is the first step in presenting the data. In order to effectively convey patterns, trends, and insights discovered throughout the data analysis process, this step includes creating visual representations like charts, graphs, and tables. The goal is to facilitate a clear understanding of complex information, making it accessible to both technical and non-technical audiences.

Conclusion

In summary, data analysis processes enable companies to make well-informed decisions by reducing complicated information into concise, visually appealing narratives. The effective communication of data-driven insights is crucial for tackling business challenges and promoting ongoing enhancement across diverse fields.

Data Analytics Approaches

Data analysts work with data using a range of strategies, techniques, and resources. Data mining management and analysis can benefit greatly from the use of mathematical and statistical sciences. Actually, statistical data analysis tools are what most data mining approaches are. Certain approaches and strategies are well-known and very successful [6, 7].

Descriptive Analytics

Using historical data, descriptive analytics aims to shed light on past occurrences and comprehend what went wrong. Its main goal is to provide readers a comprehensive picture of the situation as it stands

right now by summarizing and visualizing facts. Data aggregation, data visualization, and elementary statistical analysis are examples of descriptive analytics approaches.

Diagnostic Analytics

The goal of diagnostic analytics is to examine the linkages and dependencies among the data in order to determine why specific patterns or occurrences happened. To identify the underlying reasons and contributing elements that resulted in particular results, it entails going deeper into the data. Root cause analysis, correlation analysis, and data profiling are examples of diagnostic analytics approaches.

Predictive Analytics

Predictive analytics forecasts future patterns or outcomes by using statistical models and past data. To predict what is likely to happen, patterns and trends in the data are analysed. Examples of methodologies in predictive analytics include regression analysis, time series forecasting, and the utilization of machine learning algorithms.

Prescriptive Analytics

Prescriptive analytics suggests certain actions or choices for optimum outcomes, going beyond simple prediction. To produce useful insights, it integrates historical data, optimization techniques, and prediction models. Techniques used in prescriptive analytics include decision analysis, simulation, and optimization modelling.

Cognitive Analytics

Cognitive analytics is an advanced technology that uses a variety of analytical methods to examine huge data sets and provides structure to unstructured data. A cognitive analytics system looks through the information in its knowledge base to identify answers that make sense for the queries that are asked.

Applications of Data Analytics

Security

Security staff members use data analytics, especially predictive analytics, to forecast future crimes or security breaches.

Transportation

Transforming the transportation sector may be greatly aided by data analytics. It is very helpful when transporting a large number of people to a place where smooth transit is required.

Risk Management

In the insurance industry, risk management is a key problem. Most people do not know that before a choice is made when insuring someone, the risk involved is determined using data that has been scientifically analysed.

Delivery

There are countless uses for data science and analytics. Several logistic corporations around the world, such as UPS, DHL, and FedEx, use data to improve their operational efficiency. Data analytics solutions have helped these organizations find the finest shipping routes, the most efficient delivery schedules, the most economical forms of transportation, and much more.

Healthcare

Managing financial difficulties while serving as many patients as possible and enhancing the quality of treatment is one issue that most hospitals encounter. Hospitals are using a great deal of machine and instrument data to monitor and improve patient flow, treatment, and equipment use. A 1% efficiency gain is estimated to result in approximately $ 63 billion in global healthcare services.

Education

Data analytics can be used by policymakers to enhance learning curriculum and management decisions. These programs would improve administrative management as well as learning. One may utilize student choice data to develop a curriculum that will enhance the program. Students would study the same material in numerous ways under a better system as a result of this.

Internet/Web Search

When the word "search" is mentioned, the first thing that comes to mind is "Google". Indeed, Google can be used in place of 'search the internet' by simply saying 'Google it' [8].

OVERVIEW OF THE PROBLEM
Problem Study

The problem study focuses on analysing the Bakery and Mess Data set of the day 06-03-2023 and 07-03-2023. The dataset may contain diverse types of information, including sales data, customer preferences, production costs, and supplier details. Integrating and analysing this heterogeneous data can be complex and time-consuming. Developing effective data processing pipelines and implementing suitable analytical tools are necessary to extract meaningful insights and make data-driven decisions. Analysing this data can help identify potential risks, improve quality control measures, and maintain customer satisfaction. Organizations must be able to support their sales personnel in adapting to the fast changes in the business environment. To cope with the obstacles of their everyday work, sales managers want easy access to published research that provides them with valuable insights into best practices and sound methodology [9].

Challenges/Need for the Study

The need of the study arises from the challenge of accurately predicting customer demands to optimize production and minimize waste. By analysing the dataset, patterns and trends can be identified to better understand customer preferences, seasonal variations, and popular items, enabling bakeries to streamline their operations and reduce costs. The primary obstacle encountered during the study focuses on the cleaning and modification of a dataset. So, analysing each column was a bit challenging but it is interesting to do. It helps in gaining more and new knowledge.

Goal

The goal of bakery and mess data set was to analyse the sales of bakery and mess products.

Hardware/System Requirements

In order for a hardware or software program to function properly and efficiently, a system has to be configured in a certain way:
System: Windows 11.
Installed RAM: 8.00 GB.
Processor: 11th Gen Intel(R) Core(TM) i3-1115G4@ 3.00 GHz.
System type: 64-bit operating system, x64-based processor.

Software, Tools and Libraries Requirements

The software used in this summer internship project is MS Excel which is 2016 version and the analysis is done using pivot table and forecast.

MS Excel
a. *Pivot table:* A pivot table is to take a dataset and rearrange its structure so that you can easily explore the relationships between different variables. By dragging and dropping fields or columns, you can group, sort, filter, and calculate data based on various criteria.
b. *Pivot chart:* A pivot chart is a visual representation. A pivot chart works hand-in-hand with a pivot table and is created based on the summarized data generated by the pivot table. By using a

pivot chart, we can transform the numerical data from a pivot table into various chart types such as bar graphs, line charts, pie charts, or scatter plots.

DATA PREPARATION
Data Collection Approaches
The real time data set is collected from the bakery and mess. When it comes to collecting data for bakery and mess datasets, there are several approaches that can be employed. One common method is through direct observation and manual data collection. This involves physically visiting bakeries and mess halls to record information such as the types of baked goods available, their prices, and customer preferences. This dataset contains 8 columns and 350 rows.

Data Method
Data scientists utilize exploratory data analysis (EDA) to examine and evaluate data sets and enumerate their primary attributes, frequently utilizing techniques for data visualization. Data scientists may find trends, identify anomalies, test hypotheses, and verify assumptions more easily when they know how to effectively manipulate data sources to acquire the answers they need [10].

EDA is mostly used to examine data before drawing any conclusions. It can assist in identifying obvious mistakes, better understanding data trends, and identifying outliers or unusual occurrences. Types of exploratory data analysis [11, 12]:
1. Univariate non-graphical data analysis is the most basic type of data analysis since it just uses one variable to get information. Knowing the underlying sample distribution and data, as well as making observations about the population, are the basic objectives of univariate non-graphical EDA.
2. *Multivariate Non-Graphical:* The multivariate non-graphical EDA approach is often used to display the relationship between two or more variables in a statistical or cross-tabulation context.
3. *Univariate graphical:* While non-graphical approaches are quantitative and objective, they are unable to provide a whole picture of the data. As a result, graphical methods are more frequently utilized since they necessitate some degree of subjective interpretation.
4. *Multivariate graphical data:* This type of data shows the connections between two or more sets of information via the use of visuals. A grouped bar plot is the only one that is frequently used, where each group represents one level of one of the variables and each bar in a gaggle represents the quantity of the opposing variable [11].

Performance may be significantly enhanced by using a data-driven approach to the sales collaboration cycle [13]. It is possible to adjust sales interaction to match market demands by employing an analytical method to ascertain customer wants and sales "signals" [14, 15]. Magee *et al*. discussed and presented how effective visualization is in driving end-user adoption in a sales environment [14].

Purpose of Data
The dataset can include information about the products sold by the bakery, such as item names, prices, quantities, and sales data. This data can be used for tracking inventory, analysing sales patterns, and making informed decisions about stock replenishment. The dataset can include data on production costs, including ingredient costs, labour expenses, and other overheads. Using this information, one may assess the profitability of various items, enhance pricing strategies, and decide on cost-cutting options with knowledge.

METHODOLOGY
Descriptive Analysis/Model Used
Descriptive analysis involves summarizing and interpreting data to gain insights into its key characteristics and patterns. It aims to provide a clear and concise understanding of the data without making any causal inferences or generalizations. Data aggregation and data mining are two techniques used in descriptive analysis to churn out historical data.

Methodology involved in Excel Data Analysis
Which date the products were sold the highest?
  1. *Step 1:* Highest sales on two days taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as cross tabulation, data summarization table, multidimensional analysis table, summary table, dynamic table, tabular analysis tool, spreadsheet analysis tool etc.
  4. *Step 4:* After the process, it was found that on 06-03-2023, the product sold the highest.

Which item sold the highest on 06-03-2023?
  1. *Step 1:* Highest sales of 06-03-2023 was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as cross tabulation, data summarization table, multidimensional analysis table, summary table, dynamic table, tabular analysis tool, spreadsheet analysis tool etc.
  4. *Step 4:* After the process, it was found that on 06-03-2023, chocolate cake are sold the highest.

Which item sold the highest on 07-03-2023?
  1. *Step 1:* Highest sales of 07-03-2023 was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as cross tabulation, data summarization table, multidimensional analysis table, summary table, dynamic table, tabular analysis tool, spreadsheet analysis tool etc.
  4. *Step 4:* After the process, it was found that on, plain cake are sold the highest.

Which cake sold the highest in the two days?
  1. *Step 1:* Highest cake sales was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as Cross tabulation, Data summarization table, Multidimensional analysis table etc.
  4. *Step 4:* After the process, it was found that plain cake is sold the highest.

Which day GST was highest?
  1. *Step 1:* GST was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as Cross tabulation, Data summarization table, Multidimensional analysis table etc.
  4. *Step 4:* After the process it was found that GST was highest on 07-03-2023.

Which mess item sold the highest on 06-03-2023?
  1. *Step 1:* Highest sales of 06-03-2023 was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as Cross tabulation, Data summarization table, Multidimensional analysis table etc.
  4. *Step 4:* After the process it was found that Dosa (mess item) sold the highest.

Which mess item sold the highest on 07-03-2023?
  1. *Step 1:* Highest sales of 07-03-2023 was taken as input.
  2. *Step 2:* The pivot table is used for analysis.
  3. *Step 3:* Pivot table has many names such as Cross tabulation, Data summarization table, Multidimensional analysis table etc.
  4. *Step 4:* After the process it was found that Dosa (mess item) sold the highest.

Statistical Analysis Method: Forecasting

Statistical analysis is the process of collecting and analysing data in order to discern patterns and trends. It is a method for removing bias from evaluating data by employing numerical analysis. This technique is useful for collecting the interpretations of research, developing statistical models, and planning surveys and studies.

Table 1 and Figure 1 show the forecast with the Lower confidence bound and Upper confidence bound.

RESULTS, FINDING INSIGHTS AND DISCUSSION
Highest Sales

From the analysis, it is found that on 06/03/2023, the products were sold the highest (Table 2 and Figure 2).
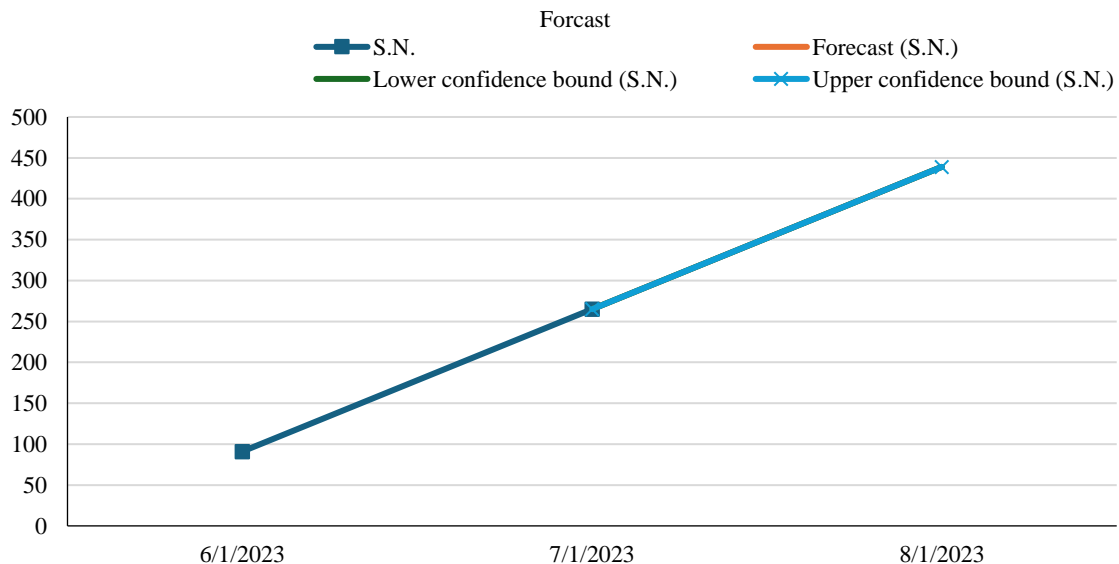
Highest Sales Item on 06-03-2023

From the analysis, it is found that on 06-03-2023, Chocolate cake is sold the highest (Table 3 and Figure 3).

Highest Sales Item on 07-03-2023?

From the analysis, it is found that on 07-03-2023, sweets are sold the highest (Table 4 and Figure 4).

Table 1. Forecast data.

Date	S.N.	Forecast (S.N.)	Lower confidence bound (S.N.)	Upper confidence bound (S.N.)
06/03/2023	91			
07/03/2023	265	265	265.00	265.00
08/03/2023		439	439.00	439.00

Figure 1. Forecast sheet.

Table 2. To find the highest sales.
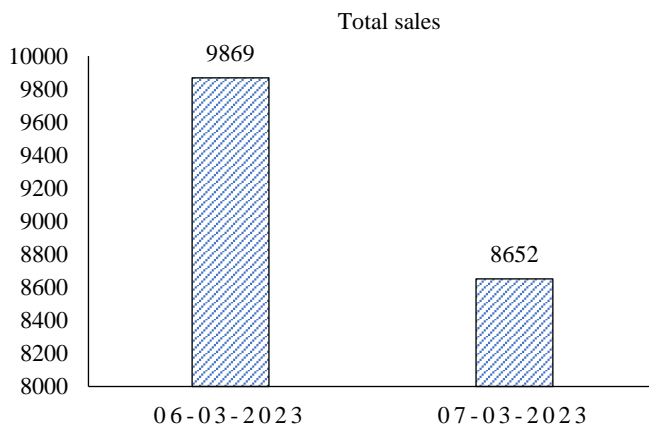
Row Labels	Sum of Price (Rs)
06-03-2023	9869
07-03-2023	8652

Figure 2. Highest sales.

Table 3. Highest Sales on 06-03-2023.

Row Labels	Sum of Price (Rs)
06-03-2023	**3320**
Chocolate cake	1370
Chocolates	585
Sweets	495
Banana chips	470
Bread	400
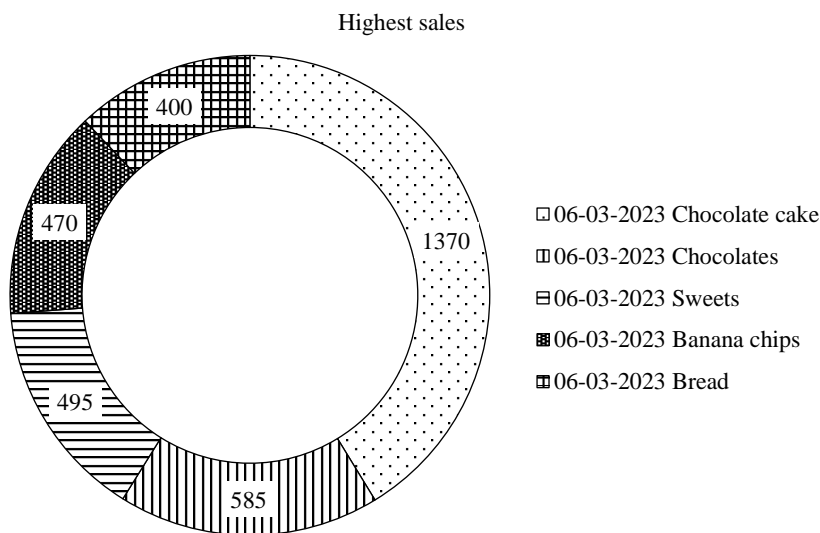
Figure 3. Graph of Highest Sales on 06-03-2023.

Table 4. Highest Sales on 07-03-2023.

| Row Labels	Sum of Price (Rs)
07-03-2023	**1880**
Sweets	480
Plum cake	450
Choco Truffle cake	400
Mixed chocolates	320
plain cake	230
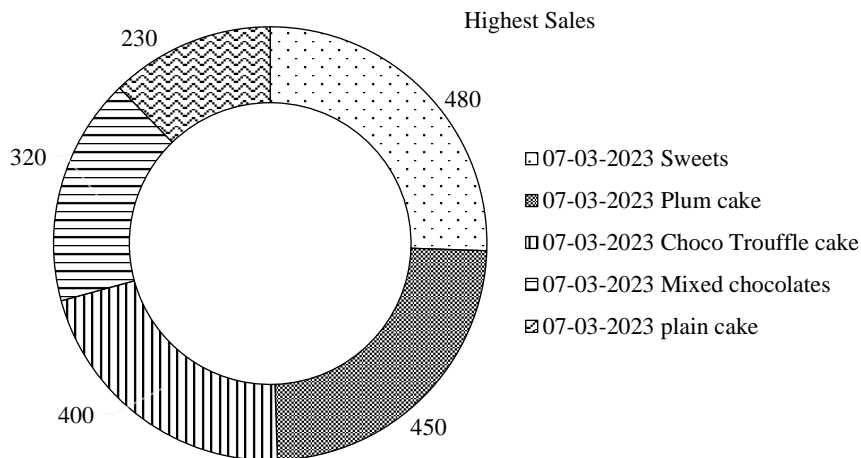
Figure 4. Graph of Highest Sales on 07-03-2023.

Table 5. Which cake sold the highest.

Row Labels	Count of quantity
Butter cake	1
Butterscotch cake	1
Choco almond cake	1
Choco truffle cake	1
Chocolate cake	5
Fruit cake	5
Jam cake	2
Orange cake	3
plain cake	10
Plum cake	5
Rose cake	1
Strawberry cake	3
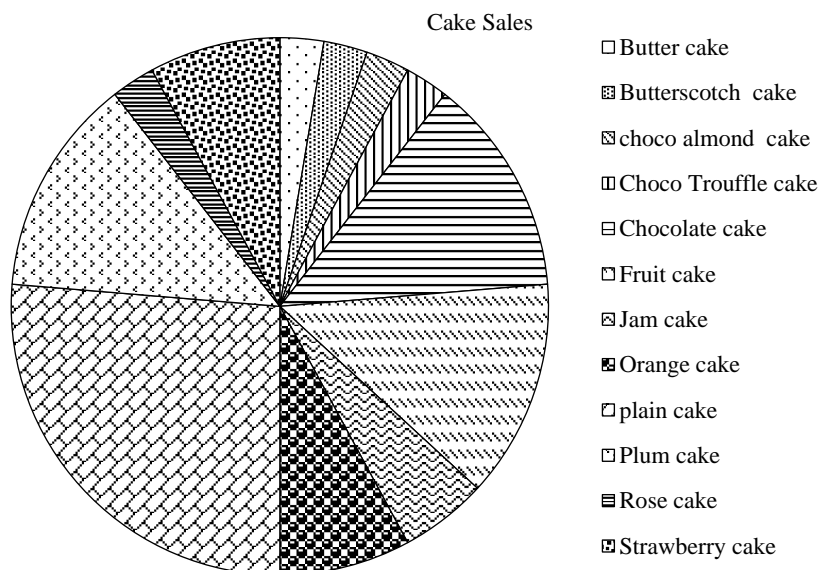
Figure 5. Which cake sold the highest.

Which Cake Sold the Highest in the Two Days?
From Table 5 and Figure 5, in the two days' analysis on cake sales, plain cake was sold the highest.

Which Day GST was Highest?

From Table 6 and Figure 6 analysis, the GST was highest on 07-03-2023.
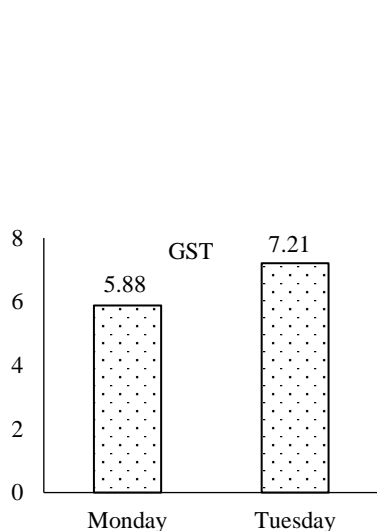
Which Mess Item Sold the Highest?

From Table 7 and Figure 7, the analysis on 06-03-2023, Dosa (Mess Item) was sold the highest.
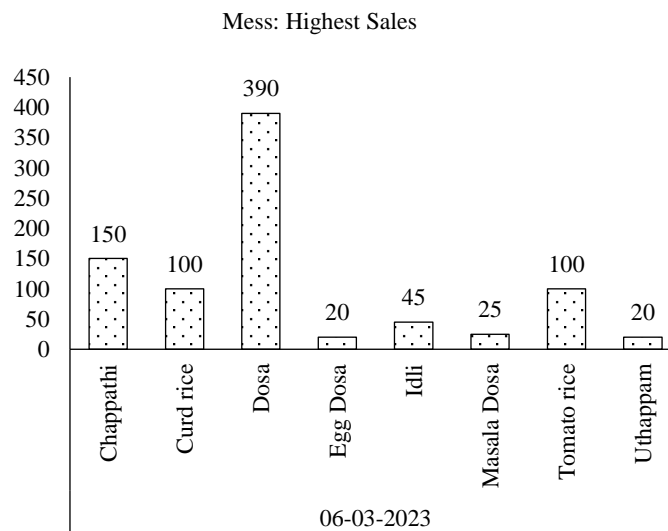
Table 6. Highest GST.

Row Labels	Sum of GST (Rs)
Monday	5.88
Tuesday	7.21
Grand Total	**13.09**

Table 7. Highest Sales of Mess Item.

Row Labels	Sum of Price (Rs)
06-03-2023	**850**
Chapati	150
Curd rice	100
Dosa	390
Egg Dosa	20
Idli	45
Masala Dosa	25
Tomato rice	100
Uthappam	20

Figure 6. Highest GST.

Figure 7. Highest sales of mess item.

Which Mess Item Sold the Highest?

From Table 8 and Figure 8, the analysis on 07-03-2023, Dosa (Mess Item) was sold the highest.

Dashboard

A dashboard is a data visualization that shows important metrics and information in a single view. It facilitates the tracking and analysis of data in an understandable manner so that decisions may be made. It shows information such as daily sales, popular items, revenue trends and marketing performance. The design of dash board is shown in the Figure 9.

Using Data Analysis Method for Price Column

A data analysis method is a set of procedures and methods for looking through and analysing data to find trends, patterns, and insights. It involves applying statistical, mathematical, or computational methods to analyse data, as shown in the Table 9.

Table 8. Highest sales of mess item.

Row Labels	Sum of Price (Rs)
07-03-2023	**945**
Curd rice	60
Dosa	330
Idli	45
Masala Dosa	40
Onion dosa	20
Parotta	210
Poori	150
Veg biryani	90

Table 9. Data analysis method for price column.

Price (Rs)	
Mean	53.27874
Standard Error	3.302319
Median	35
Mode	30
Standard Deviation	61.60397
Sample Variance	3795.049
Kurtosis	39.34411
Skewness	5.340116
Range	595
Minimum	5
Maximum	600
Sum	18541
Count	348
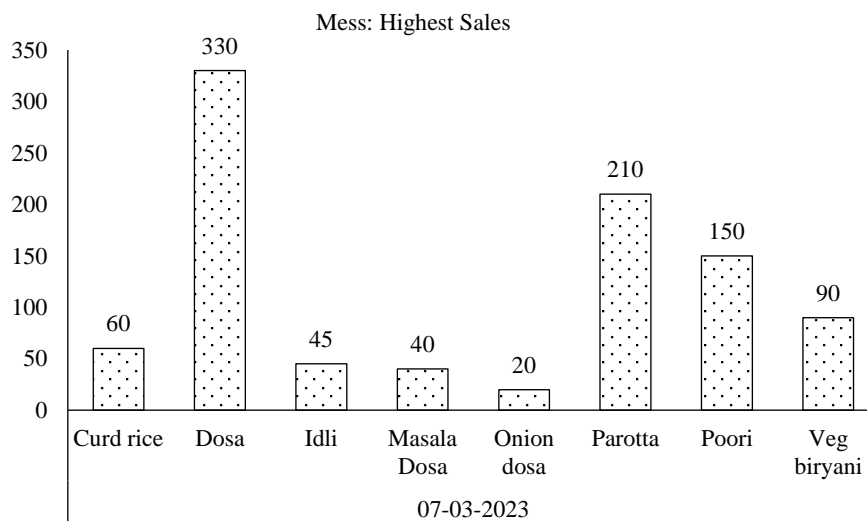
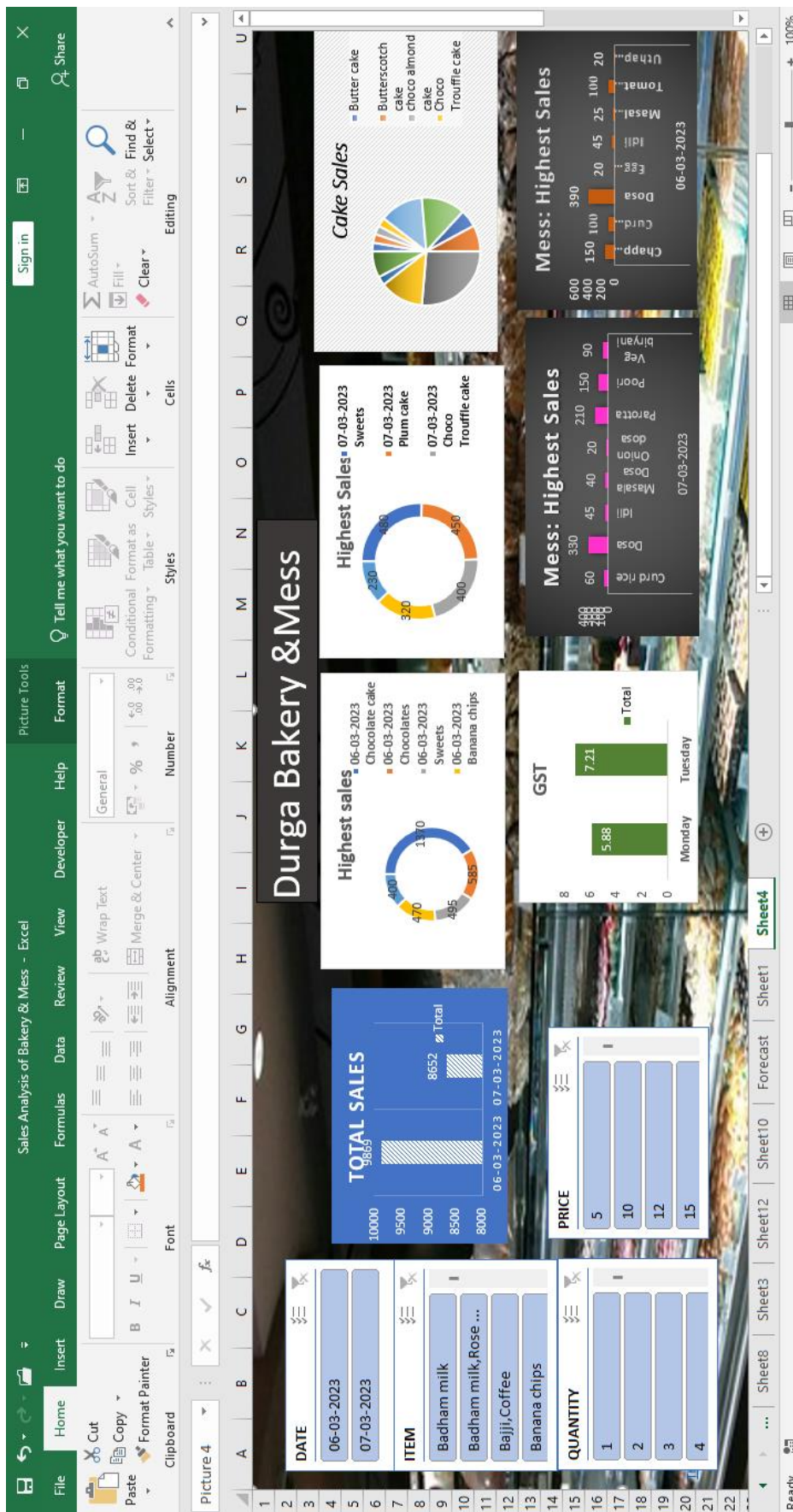**Figure 8.** Highest sales of mess item.

Figure 9. Dash board.

CONCLUSION

In the bakery and mess data set, first the category of all columns were checked and changed accordingly. Indentation, font and column width were changed accordingly. A column name call day was newly created. The analysis was done in pivot table and different pivot charts were created for better understanding and the results were found accordingly. Then the forecast sheet is used to predict the future values. The above bakery and mess data set has data of 06-03-2023 and 07-03-2023. From the analysis it is found that on 06-03-2023 the products were sold the highest and, on this day, chocolate cake was sold the highest for Rs. 1370. On 07-03-2023, sweets are sold the highest for Rs. 480. In the two days' analysis on cake sales, plain cake was sold the highest. From this bakery and mess data set, GST was highest on 07-03-2023. In two days' analysis, dosa (mess item) sold the highest for above Rs. 300.

REFERENCES

1. CareerFoundry. (2021). What is Data Analytics? A Complete Guide for Beginners. [online] Available from: https://careerfoundry.com/en/blog/data-analytics/what-is-data-analytics/
2. GeeksforGeeks. (2021). Six Steps of Data Analysis Process. [online] Available from: https://www.geeksforgeeks.org/six-steps-of-data-analysis-process/.
3. Abdul-Jabbar Safa, Farhan Alaa. Data Analytics and Techniques: A Review. ARO-The Scientific Journal of Koya University. 2022; 10(2): 45–55. 10.14500/aro.10975.
4. Schnegg M, Möller K. Strategies for data analytics projects in business performance forecasting: a field study. J Manag Control. 2022; 33(2): 241–271. https://doi.org/10.1007/s00187-022-00338-7
5. Taherdoost Hamed. Different Types of Data Analysis; Data Analysis Methods and Techniques in Research Projects. Int J Acad Res Manag. 2022 Aug 1; 9(1): 1–9.
6. Ayush Singh Rawat. (2021). What is Descriptive Analysis? Types and Advantages | Analytics Steps. [online] AnalyticsSteps. Available from: https://www.analyticssteps.com/blogs/overview-descriptive-analysis
7. Nate Rosidi. (2023 Apr 20). Data analytics: Four approaches to analyzing data and how to use them effectively. [Online]. https://www.kdnuggets.com/2023/04/data-analytics-four-approaches-analyzing-data-effectively.html
8. Pushkar A. (2022). Top 10 Data Analytics Applications. [online] Intellipaat. Available from: https://intellipaat.com/blog/tutorial/data-analytics-tutorial/data-analytics-applications/
9. Ricky Fergurson J. Data-driven decision making via sales analytics: introduction to the special issue. J Market Anal. 2020; 8(3): 125–126.
10. Komorowski Matthieu, Marshall Dominic, Salciccioli Justin, Crutain Yves. Exploratory Data Analysis. 2016. In: MIT Critical Data, editor. Secondary Analysis of Electronic Health Records [Internet]. Cham (CH): Springer; 2016. Chapter 15. 10.1007/978-3-319-43742-2_15.
11. GeeksforGeeks (2021). Exploratory Data Analysis (EDA) Types and Tools. [online] GeeksforGeeks. Available from: https://www.geeksforgeeks.org/exploratory-data-analysis-eda-types-and-tools/
12. Tariq Hussain Sheikh, Aithal PS. Visualization and Explorative Data Analysis. Int J Enhanc Res Sci Technol Eng. 2023 Mar 26; 12(3): 11–21.
13. Elizabeth Fernandes, Sérgio Moro, Paulo Cortez, Fernando Batista, Ricardo Ribeiro. A data-driven approach to measure restaurant performance by combining online reviews with historical sales data. Int J Hosp Manag. 2021; 94: 102830.
14. Barry Magee, David Sammon, Tadhg Nagle, Paidi O'Raghallaigh. Introducing data driven practices into sales environments: examining the impact of data visualisation on user engagement and sales results. J Decis Syst. 2016; 25(Sup1): 313–328.