Business Analytics Domain Task - Data Visualization

DAY 1 - Theoretical Analysis

1. What is Data Visualization?

Data visualization is the graphical representation of data and information using visual elements such as charts, graphs, maps, and diagrams. Its purpose is to present complex data sets in a visually appealing and easily understandable format, enabling users to extract meaningful insights, patterns, and trends from the data. Data visualization utilizes visual encoding techniques to represent data variables, such as using different colors, shapes, sizes, and positions to convey information.

By visually representing data, data visualization helps in exploring and understanding large datasets, identifying correlations, outliers, and patterns, and communicating insights effectively to a wider audience. It allows users to perceive information quickly, make comparisons, and discover relationships that might not be evident in raw data. Data visualization is widely used in various fields, including business, finance, science, journalism, healthcare, and many more, to support decision-making, storytelling, and data-driven analysis.

2. What are the different techniques involved in Data Visualization?

Data visualization employs a variety of techniques to represent data effectively. Some of the commonly used techniques include:

- i. Charts and Graphs: Charts and graphs are visual representations that use different types of visual elements such as bars, lines, points, and areas to display data. Examples include bar charts, line graphs, scatter plots, pie charts, and area charts. These techniques are useful for comparing values, showing trends over time, and displaying proportions.
- ii. Maps and Geospatial Visualization: Maps are used to represent data based on geographic locations. Geospatial visualization techniques utilize maps to display data such as population density, sales distribution, or any other spatially related information. Choropleth maps, heat maps, and bubble maps are common examples.
- iii. Infographics: Infographics combine text, images, and visual elements to present information and data in a visually appealing and concise manner. They often use icons, illustrations, and diagrams to convey complex concepts and statistics.
- iv. Dashboards: Dashboards provide a consolidated view of multiple visualizations and metrics on a single screen. They allow users to monitor key performance indicators (KPIs) and track the status of various metrics in real-time.
- v. Network Visualization: Network visualization techniques represent data that involves relationships and connections between entities. Network diagrams or graphs are used to visualize nodes (representing entities) and edges (representing relationships) between them. This technique is useful for analyzing social networks, communication networks, and other interconnected systems.
- vi. Tree Maps: Tree maps display hierarchical data structures using nested rectangles, where the size and color of each rectangle represent different attributes or values. They are useful for visualizing hierarchical relationships and comparing proportions within a hierarchy.
- vii. Interactive Visualization: Interactive visualization techniques allow users to manipulate and explore data dynamically. Users can interact with visualizations by filtering, sorting, zooming, and selecting data elements, enabling deeper exploration and discovery of insights.

These are just a few examples of the techniques involved in data visualization. The choice of technique depends on the nature of the data, the insights to be conveyed, and the target audience.

3. Which are the different softwares which helps in data Visualization.

There are numerous software tools available that assist in data visualization. Here are some popular options:

- i. Tableau: Tableau is a powerful and widely used data visualization tool that offers a user-friendly interface and a wide range of visualization options. It allows users to create interactive dashboards, reports, and charts, and supports various data sources.
- ii. Microsoft Power BI: Power BI is a business intelligence tool by Microsoft that enables users to visualize and analyse data through interactive reports, dashboards, and visualizations. It integrates well with other Microsoft products and offers robust data modelling and sharing capabilities.
- iii. QlikView/Qlik Sense: QlikView and Qlik Sense are data visualization and analytics platforms that allow users to create interactive visualizations, dashboards, and reports. They offer features like associative data model, data exploration, and collaboration capabilities.
- iv. Google Data Studio: Google Data Studio is a web-based data visualization tool that allows users to create interactive reports and dashboards. It integrates with various data sources and offers real-time collaboration and sharing features.
- v. D3.js: D3.js (Data-Driven Documents) is a JavaScript library that provides a framework for creating custom and interactive data visualizations on the web. It gives users complete control over the visualization design and is widely used for creating complex and dynamic visualizations.
- vi. Python Libraries (matplotlib, Seaborn, Plotly): Python, a popular programming language for data analysis, offers several libraries for data visualization. Matplotlib is a widely used library for creating static visualizations, Seaborn specializes in statistical visualizations, and Plotly allows for the creation of interactive and web-based visualizations.
- vii. R (ggplot2, Shiny): R is another programming language commonly used for statistical analysis and data visualization. The ggplot2 package provides an extensive range of static visualizations, while Shiny enables the creation of interactive web applications with visualizations.
- viii. Excel: Microsoft Excel, a widely available spreadsheet program, offers basic data visualization capabilities. It allows users to create charts, graphs, and pivot tables to represent data visually.
- ix. Adobe Illustrator: Adobe Illustrator is a vector graphics software that can be used for creating custom and visually appealing data visualizations. It provides extensive design and customization options for creating unique visual representations.

These are just a few examples, and there are many other data visualization tools and software available, each with its own features, strengths, and target audiences. The choice of software depends on factors such as the complexity of the data, specific visualization requirements, user skill level, and budget considerations.

4. What is Big Data? What are it's characteristics?

Big data refers to extremely large and complex sets of data that exceed the processing capabilities of traditional data management and analysis tools. It encompasses data from various sources, including structured data (e.g., databases), unstructured data (e.g., text, images, videos), and semi-structured data (e.g., XML, JSON). Big data is characterized by the "Three V's":

- i. Volume: Big data involves a massive volume of data that exceeds the capacity of conventional systems to store, process, and analyze. It can range from terabytes (10^12 bytes) to petabytes (10^15 bytes) and beyond.
- ii. Velocity: Big data is generated at a high velocity or speed. Data is continuously streaming in from various sources, including sensors, social media, online transactions, and more. Real-time or near-real-time processing is often required to extract timely insights from the data.
- iii. Variety: Big data encompasses diverse types and formats of data, including structured, unstructured, and semi-structured data. It includes text, audio, video, social media posts, clickstream data, geospatial data, and more. Managing and analyzing this heterogeneous mix of data is a key challenge in big data processing.

Additional characteristics of big data include:

- i. Veracity: Big data often suffers from issues of veracity, referring to the trustworthiness and reliability of the data. It may contain inaccuracies, errors, or inconsistencies that need to be addressed during data processing and analysis.
- ii. Variability: Big data exhibits variability in terms of its arrival rate, structure, and format. The data may arrive in bursts or uneven patterns, and its structure may change over time. Adapting to this variability is crucial for effective big data management.
- iii. Complexity: Big data is complex not only due to its volume and variety but also because of the relationships, dependencies, and correlations within the data. Extracting meaningful insights requires advanced analytical techniques and tools.
- iv. Value: Big data has the potential to unlock valuable insights and opportunities for organizations. By analyzing large and diverse datasets, businesses can gain deeper insights, make data-driven decisions, and discover patterns and trends that were previously hidden.

To harness the power of big data, specialized technologies and approaches, such as distributed computing frameworks (e.g., Hadoop, Spark), NoSQL databases, data streaming platforms, machine learning algorithms, and data visualization techniques, are employed to store, process, analyze, and derive actionable insights from the vast amount of data.

5. Which all Data Visualization techniques and softwares you are aware of?

I'm aware of a wide range of data visualization techniques and software tools. Here are some commonly used techniques:

- i. Bar Charts
- ii. Line Charts
- iii. Area Charts
- iv. Scatter Plots
- v. Bubble Charts
- vi. Pie Charts
- vii. Tree Maps
- viii. Word Clouds
- ix. Gauge Charts
- x. Box Plots
- xi. Histograms

And here are some popular data visualization software tools:

- i. Tableau
- ii. Microsoft Power BI
- iii. Python libraries (matplotlib, seaborn, Plotly)
- iv. Excel

The choice of technique and software depends on factors such as the type of data, visualization requirements, user skill level, and specific needs of the project.

6. What is a pivot table?

A pivot table is a data summarization tool used in spreadsheet software, such as Microsoft Excel, Google Sheets, or similar applications. It allows you to extract and analyze large amounts of data from a table or dataset and provides a way to reorganize, summarize, and aggregate the data based on different criteria.

The main purpose of a pivot table is to provide a flexible and interactive way to explore and understand data by creating a summarized view. It enables users to quickly generate meaningful insights from raw data without the need for complex formulas or manual calculations.

Here's how a pivot table works:

- i. Input Data: You start with a dataset or table containing rows of data, typically with multiple columns representing different variables.
- ii. Selection: In a pivot table, you select the columns or fields that you want to include in the analysis. These columns become the "source data" for the pivot table.
- iii. Row and Column Fields: You choose one or more fields from the dataset to be the row and column fields in the pivot table. These fields determine how the data will be organized and displayed.
- iv. Value Field: You select a specific field that contains numerical data for aggregation, such as sums, averages, counts, or other calculations. This field provides the values that will be summarized and analyzed.
- v. Aggregation: The pivot table automatically groups the data based on the row and column fields you've chosen. It aggregates the values in the selected value field(s) based on these groups.
- vi. Interactivity: Pivot tables offer interactivity, allowing you to easily change the organization and summarization of the data. You can rearrange the row and column fields, add or remove fields, apply filters, and perform calculations on the fly.
- vii. Summary: The pivot table displays a summarized view of the data, presenting the aggregated values in a tabular format. It provides subtotals, grand totals, and often enables you to drill down into the underlying details for further analysis.

Pivot tables are particularly useful when working with large datasets or when you want to quickly analyze data from multiple angles or dimensions. They can provide insights, identify patterns, and uncover trends in a structured and customizable manner, making them a valuable tool for data analysis and reporting.

7. How does dashboards aides Data Visualization?

Dashboards play a crucial role in data visualization by providing a consolidated and interactive view of multiple visualizations and key metrics in a single interface. They aid in data visualization in the following ways:

- i. Centralized Information: Dashboards bring together diverse visualizations and metrics into a centralized location. They provide a comprehensive snapshot of relevant data, enabling users to quickly assess the overall performance or status of a system, project, or organization.
- ii. Data Exploration: Dashboards allow users to interactively explore and drill down into data. They can apply filters, select specific metrics or dimensions, and zoom in on specific time periods or regions of interest. This interactive capability empowers users to investigate data patterns, identify outliers, and gain deeper insights.
- iii. Real-Time Monitoring: Dashboards can display real-time or near-real-time data, providing up-todate information on critical metrics. Users can monitor live data streams, track key performance indicators (KPIs), and respond to changes or anomalies promptly.
- iv. Visual Representation: Dashboards leverage data visualization techniques to present information in a visually appealing and easily understandable format. They utilize charts, graphs, gauges, maps, and other visual elements to represent data, making it easier to grasp complex information at a glance and spot trends, patterns, and correlations.

- v. Customization and Personalization: Dashboards often allow users to customize the layout, choose the specific visualizations to include, and tailor the metrics based on their needs. This flexibility enables individuals or teams to create personalized views that align with their unique requirements and preferences.
- vi. Communication and Collaboration: Dashboards facilitate data-driven communication and collaboration within organizations. They provide a common platform where stakeholders can access and share visualized data, making it easier to align goals, track progress, and make informed decisions based on a shared understanding of the data.

vii.	decisions based on a shared understanding of the data. Actionable Insights: Effective dashboards not only present data visually but also provide actionable insights. They highlight trends, exceptions, or threshold breaches, triggering alerts or notifications that prompt users to take appropriate actions based on the insights derived from the visualizations.				
By combining multiple visualizations, real-time data, interactivity, and customization options, dashboards enhance data visualization by delivering dynamic, informative, and actionable representations of data. They empower users to monitor, analyze, and respond to data efficiently, supporting data-driven decision-making across various domains and industries.					

Business Analytics Domain Task - DAY 2

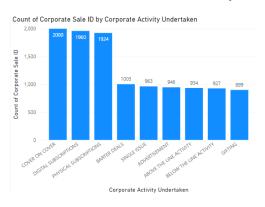
Using PowerBI

1. Most selling, least selling products



- From the above bar graph we can conclude that OUTLOOK INDIA is the most selling product with **3377** as the count of units purchased.
- While, OUTLOOK TRAVELLER is the least selling product with 1602 as the count of units purchased.

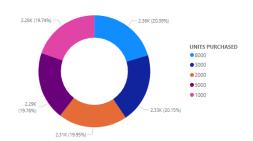
2. Most number, least number of Corporate activity undertaken



- From the above bar graph we can conclude that COVER ON COVER has the most no. of corporate activities undertaken with 2000 as the count.
- While, GIFTING has the least no. of corporate activities undertaken with 899 as the count.

3. Average no. of units per sale





From the above donut chart we can conclude that –
Highest average no. of units purchased = 8000
Lowest average no. of units purchased = 1000

<u>Maintaining CRM by studying consumer buying behaviour</u> <u>of Outlook Magazine</u>

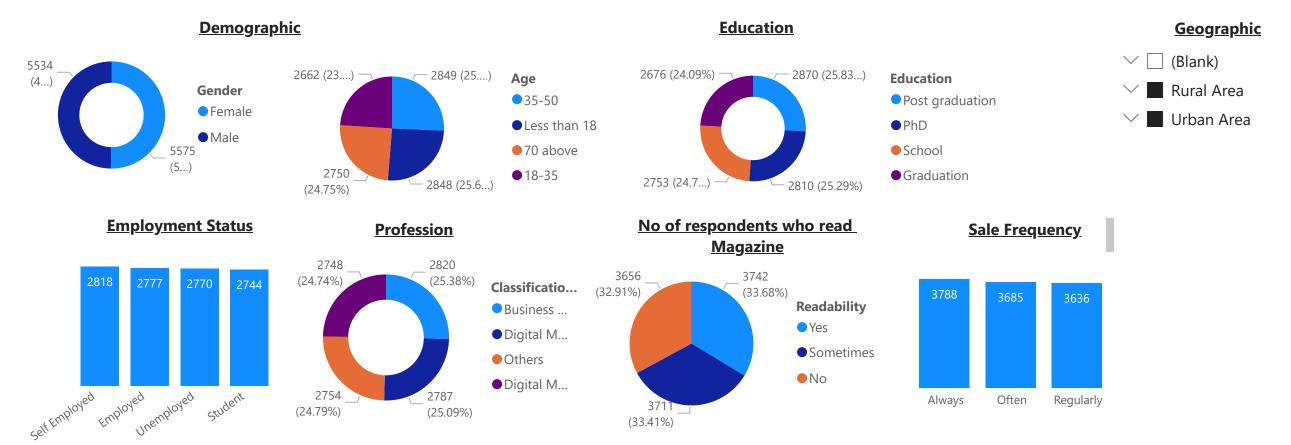
\rightarrow

Problem Statements:-

- 1) Allocation of advertising budget Demographic , Geographic , Digital Marketing Activities , Mode of awareness of Outlook Magazine , Social Media advertising
- **2) Customer Allocation based on their location** Geographic (More customer allocation in urban areas then in rural areas & More customer retention in rural areas then in urban areas)
- **3) Tie-ups with Institutions -** Profession (need for more corporate tie-ups then institutional tie-ups as majority of the respondents are office going people)
- **4) Marketing right products to the appropriate consumers -** Preference of product, Preferred genre while buying magazines, Sale Frequency by Demographic/Geographic, Awareness of Outlook magazine by Demographic/Geographic
- **5) Identifying various types of relationships that consumers have with different brands of outlook magazine -** Sale Frequency by Awareness of extensions of Outlook Magazine
- 6) Studying the spending habits of customers enables the firm to dedicate different resources and amounts of attention to different types of consumers Preference of product, Spending on products
- 7) Why consumers make the purchases that they make Monthly Income, Profession, Preference of product, Promotional Offers
- 8) What factors influence consumer purchases? Major deciding factor while buying magazine
- **9)** Buyers reactions to a firms marketing strategy has great impact in the firms success Most attractive strategy, Most influencial marketing program, Role of price in deciding which magazine to buy, Satisfaction with Pricing, Satisfaction with Pricing
- **10) Firm should create a Marketing Mix to analyze what, where, when and how consumers buy -** Preference of product, Preferred genre while buying magazines, Spending on products, Plans for subscription to Outlook, Location for buying, Frequency of usage, Mode of delivery

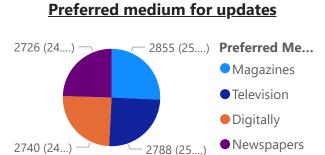


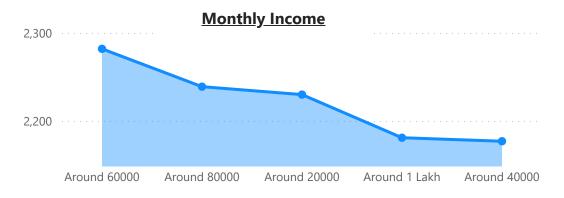






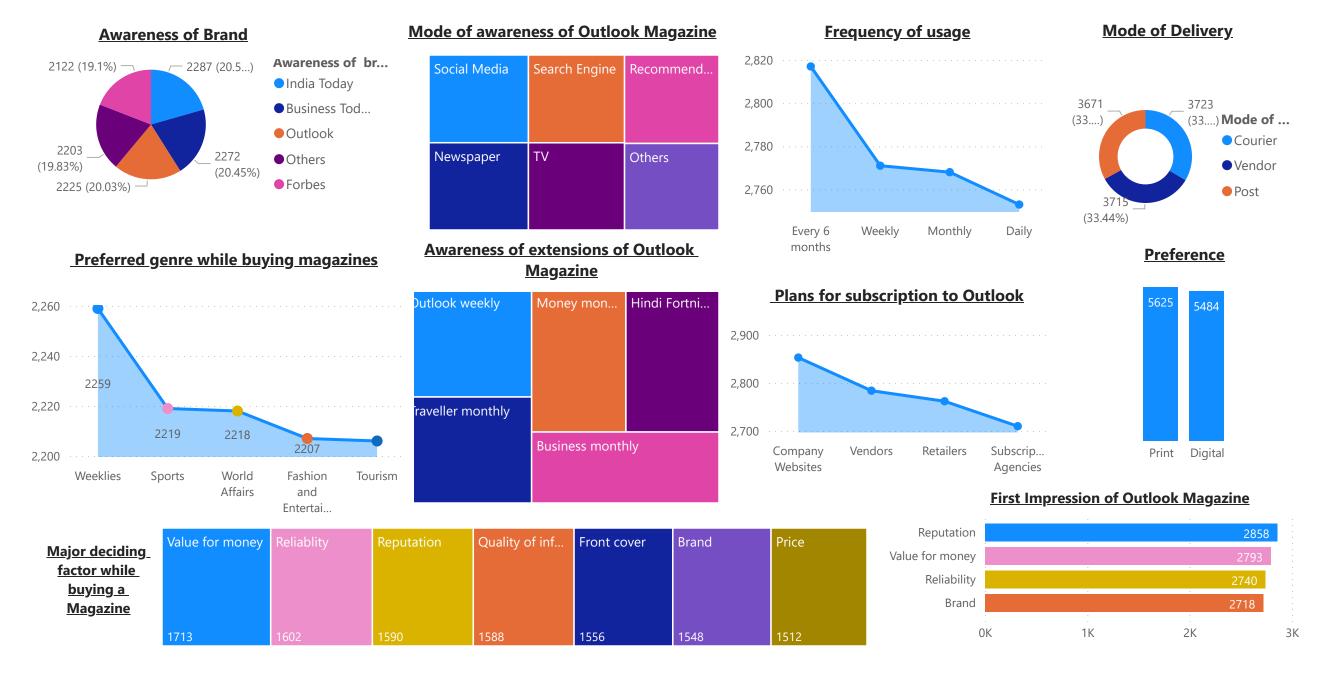








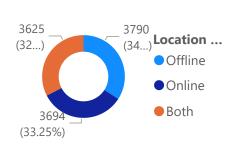












Duffle Strolleys &...

Digital Version Free

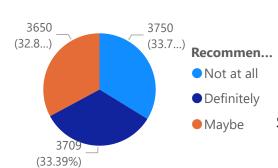
Extended Subscri..

Subscription period of Outlook Magazine



3707

Recommendation to others

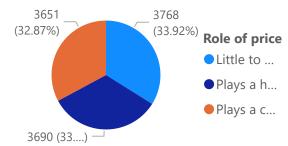


Digital Marketing Activities



(33.3970)





Spending on products by Awareness of extensions

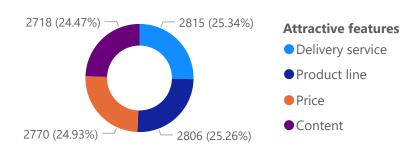


Sale Frequency by Awareness of extensions

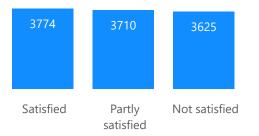


Attractive Features of Outlook Magazine

Promotional Offers



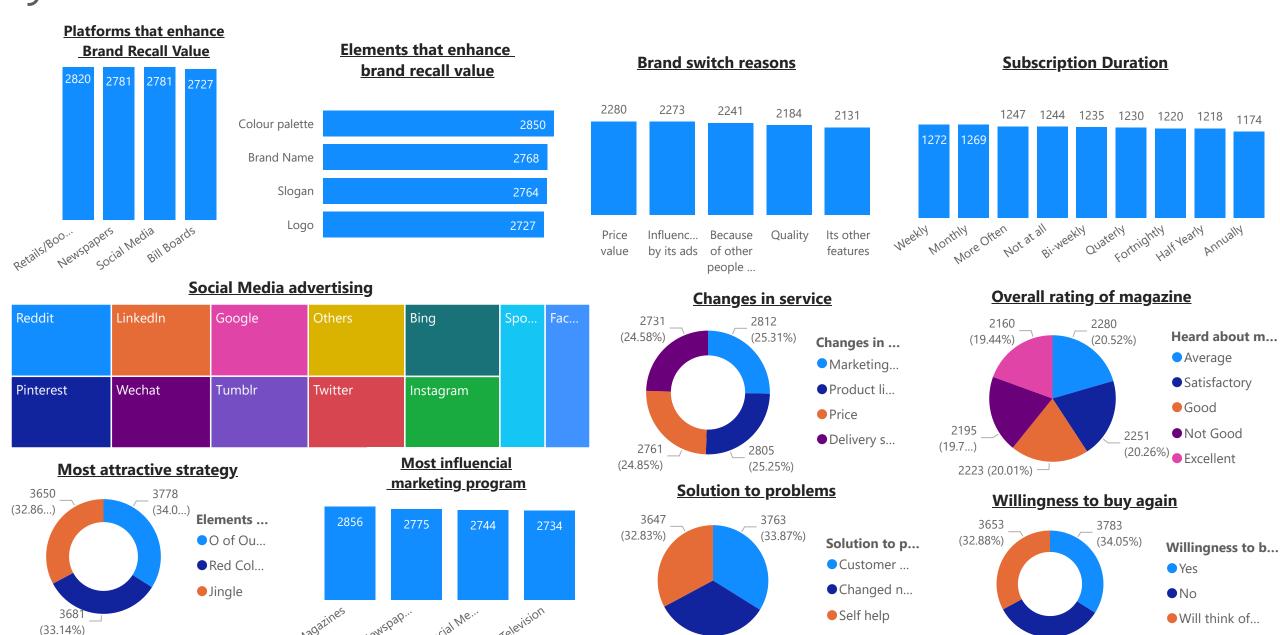
Satisfaction with Pricing



Brand Personality of Outlook

Ruggedness	Competance	Trendy Smart	Business Minded	Gentle
Excitement	Robust & Rugged	Sincerely	Sophitication	





3699 (33.3%)

3673 (33.06%)

Business Analytics Domain Task 4 - Data Forecasting

DAY 1 - Theoretical Analysis

1. What do you mean by data forecasting?

Data forecasting refers to the process of using historical data to make predictions or forecasts about future trends, patterns, or events. It involves analyzing past data and identifying patterns, relationships, and trends in order to make informed predictions about what is likely to happen in the future.

Data forecasting is commonly used in various fields, including finance, economics, business, weather forecasting, supply chain management, and sales forecasting, among others. It utilizes statistical models, mathematical algorithms, and machine learning techniques to analyze historical data and generate forecasts.

The primary goal of data forecasting is to leverage past data to estimate or predict future values, such as sales figures, stock prices, demand for a product, or weather conditions. By understanding and predicting future trends, organizations can make informed decisions, plan resources, optimize operations, and mitigate risks.

Different methods and techniques can be used for data forecasting, including time series analysis, regression analysis, exponential smoothing, moving averages, and more advanced approaches like ARIMA (Autoregressive Integrated Moving Average) and machine learning algorithms. The choice of method depends on the nature of the data, the available historical information, and the specific goals of the forecasting task.

2. What are the different data forecasting techniques you are aware of?

There are several data forecasting techniques commonly used in practice. Here are some of the most popular ones:

- i. Time Series Analysis: This technique focuses on analyzing data points collected over time to identify patterns and trends. It includes methods such as moving averages, exponential smoothing, autoregressive integrated moving average (ARIMA), and seasonal decomposition of time series.
- ii. Regression Analysis: Regression analysis involves identifying and quantifying the relationship between a dependent variable and one or more independent variables. It can be used to forecast future values by fitting a regression model to historical data and using it to make predictions.
- iii. Exponential Smoothing: Exponential smoothing is a time series forecasting method that assigns exponentially decreasing weights to past observations. It gives more weight to recent data points, enabling it to capture short-term trends and variations in the data.
- iv. Neural Networks: Neural networks, particularly recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, are powerful machine learning models that can capture complex patterns in sequential data. They are widely used for time series forecasting tasks.
- v. ARIMA (Autoregressive Integrated Moving Average): ARIMA models are a class of statistical models that combine autoregressive (AR), moving average (MA), and differencing components. They are useful for forecasting time series data with trends, seasonality, and random variations.
- vi. Prophet: Prophet is a forecasting tool developed by Facebook that utilizes an additive model to capture seasonality, trends, and holiday effects in time series data. It is designed to be user-friendly and can handle various types of time series patterns.

- vii. Decision Trees: Decision tree algorithms, such as Random Forests and Gradient Boosting, can be applied to time series forecasting. They can capture nonlinear relationships and interactions between variables, making them suitable for complex forecasting problems.
- viii. Support Vector Machines (SVM): SVM is a supervised machine learning technique that can be used for time series forecasting. It aims to find a hyperplane that separates data points in a high-dimensional space to make predictions.
 - ix. Ensemble Methods: Ensemble methods combine the predictions of multiple forecasting models to improve accuracy and robustness. Techniques like model averaging, bagging, and boosting can be used to create ensemble forecasts.
 - x. These are just a few examples of data forecasting techniques. The choice of technique depends on the specific characteristics of the data, the problem at hand, and the available resources and expertise. It's common to experiment with multiple techniques and compare their performance to determine the most suitable approach.

3. Which of the data forecasting techniques have application in the print media industry?

Several data forecasting techniques can be applied in the print media industry to support decision-making, optimize operations, and enhance audience engagement. Here are a few techniques commonly used in the industry:

- i. Time Series Analysis: Time series analysis is widely used in the print media industry to forecast circulation numbers, subscription rates, and advertising revenues over time. It helps publishers understand historical patterns, seasonal variations, and trends in readership or advertising demand.
- ii. Regression Analysis: Regression analysis can be employed to forecast advertising revenues based on various factors such as circulation, audience demographics, market conditions, and ad pricing. By quantifying the relationships between these variables, publishers can estimate future advertising revenues and make informed decisions.
- iii. Machine Learning Algorithms: Machine learning algorithms, such as neural networks or ensemble methods, can be utilized to forecast readership trends, optimize content placement, and personalize recommendations. These techniques analyze historical readership data, user behavior, and content attributes to predict future engagement and tailor content strategies.
- iv. Customer Segmentation: Customer segmentation techniques, such as clustering or market basket analysis, can help identify distinct reader segments based on their preferences, behaviors, or demographics. This information can be leveraged to develop targeted marketing campaigns, subscription offers, or content strategies for different segments.
- v. Sentiment Analysis: Sentiment analysis techniques can be applied to analyze customer feedback, social media discussions, or online comments related to print media content or publications. By understanding the sentiment of readers and the public, publishers can anticipate audience preferences, evaluate content impact, and adjust their strategies accordingly.
- vi. Market Research and Surveys: Traditional market research methodologies, such as surveys and focus groups, play a significant role in forecasting print media trends. By collecting feedback, opinions, and preferences from readers and advertisers, publishers can gain insights into market demand, content preferences, and advertising needs.
- vii. Forecasting Ad Inventory: Print media publishers often need to forecast ad inventory availability and optimize its allocation to advertisers. By analyzing historical data on ad placements, demand patterns, and seasonal variations, publishers can estimate future ad space availability and plan their sales strategies accordingly.

These techniques provide insights and predictions that can guide strategic planning, content creation, revenue optimization, and audience engagement in the print media industry. However, the specific

techniques used may vary depending on the goals, available data, and resources of individual publishers or media organizations.

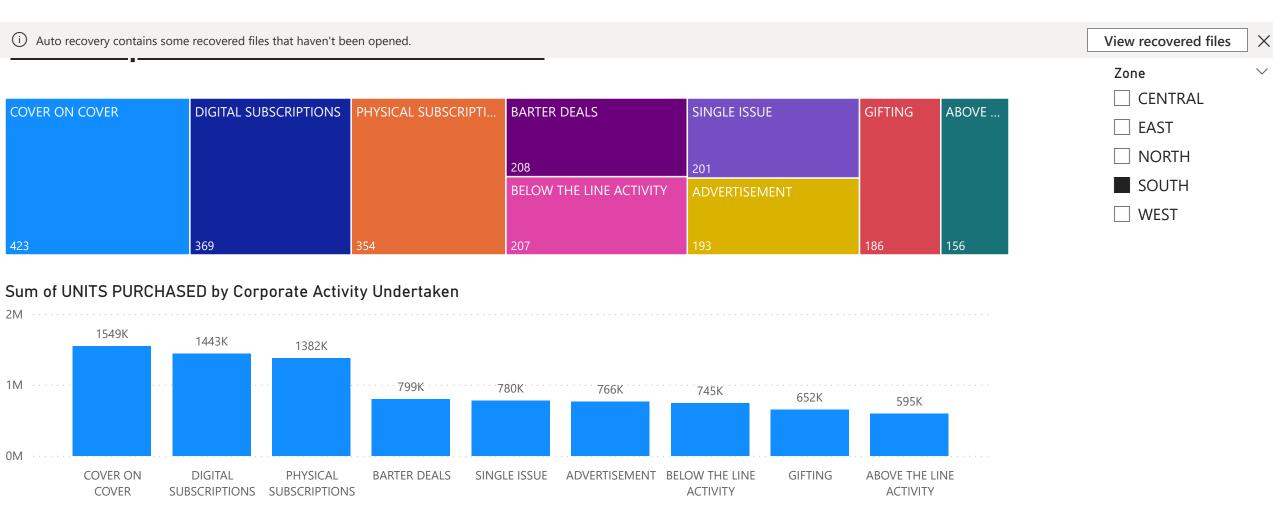
4. What is the significance of data forecasting in current business scenario?

Data forecasting plays a crucial role in the current business scenario for several reasons:

- i. Strategic Planning: Data forecasting helps organizations make informed decisions and develop effective strategies. By predicting future trends, demand patterns, and market conditions, businesses can align their resources, investments, and operations accordingly. It enables them to identify growth opportunities, anticipate risks, and stay competitive in dynamic markets.
- ii. Demand and Supply Optimization: Forecasting allows businesses to optimize their demand and supply chain management. By accurately predicting customer demand, organizations can streamline their production, inventory management, and distribution processes. This helps reduce costs, minimize stockouts or overstock situations, and ensure efficient allocation of resources.
- iii. Financial Planning and Budgeting: Data forecasting plays a crucial role in financial planning and budgeting. It helps organizations estimate future revenues, expenses, and cash flows, enabling them to create realistic budgets, allocate resources effectively, and set achievable financial targets.

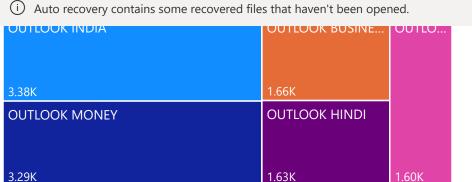
 Accurate forecasting improves financial stability, aids in capital allocation decisions, and supports long-term financial planning.
- iv. Marketing and Sales Optimization: Forecasting assists businesses in optimizing their marketing and sales efforts. By predicting customer behavior, market trends, and product demand, organizations can tailor their marketing campaigns, pricing strategies, and sales forecasts. This leads to improved customer targeting, increased sales effectiveness, and enhanced customer satisfaction.
- v. Risk Management: Data forecasting helps businesses identify and mitigate risks. By anticipating potential challenges, such as supply chain disruptions, market fluctuations, or changes in customer preferences, organizations can proactively develop risk management strategies. It enables them to minimize the impact of uncertainties, make contingency plans, and improve overall resilience.
- vi. Resource Planning and Optimization: Forecasting helps organizations plan and allocate resources effectively. By predicting future demand and operational requirements, businesses can optimize their workforce, inventory, production capacity, and logistics. This leads to improved resource utilization, reduced costs, and enhanced operational efficiency.
- vii. Performance Evaluation: Forecasting provides a basis for evaluating organizational performance. By comparing actual results with forecasted values, businesses can assess the accuracy of their predictions and identify areas for improvement. It facilitates data-driven performance evaluation, goal setting, and continuous improvement initiatives.

Overall, data forecasting enables businesses to make proactive, data-driven decisions, optimize resource allocation, enhance operational efficiency, and stay ahead of the competition. It helps organizations navigate uncertainties, seize opportunities, and achieve their strategic objectives in an increasingly dynamic and competitive business environment.



Count of Corporate Sale ID by Corporate Activity Undertaken



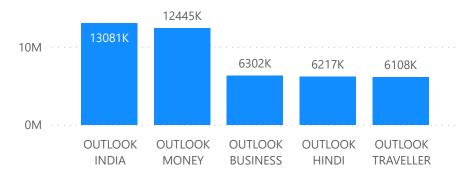




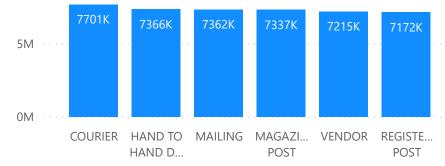
View recovered files × CENTRAL EAST NORTH SOUTH

WEST

Sum of UNITS PURCHASED by Purchased Product



Sum of UNITS PURCHASED by Delivery Routed Through



Count of UNITS PURCHASED by Purchased Product

