DMDA ASSIGNMENT

Types of Data Analytics

1. **Descriptive Analytics** – *What happened?* Summarizes past and current data.

2. **Diagnostic Analytics** – *Why did this happen?* Drills down to identify causes.

3. **Predictive Analytics** – *What might happen in the future?* Uses trends to forecast future events.

4. **Prescriptive Analytics** – *What should we do next?* Recommends actions based on predictions.

Descriptive Analytics

1. Definition – Descriptive analysis is a type of data analysis that summarizes and interprets historical data to understand past events, trends, and patterns.

2. Purpose – It focuses on showing "what happened" in the past and sometimes "what is happening now", without predicting the future.

3. Foundation for other analyses – It is the first step in data analysis, forming the base for diagnostic, predictive, and prescriptive analysis by providing a clear picture of past performance.

4. Methods used – Simple statistics (mean, median, totals), charts (bar, line, pie), graphs, and tables are common tools for descriptive analysis.

Steps for Descriptive Analytics Work:

Steps of Descriptive Analytics working.

Data Collection → Cleaning the Data and Preprocessing → Data analysis → Compilation & Summary → Visualization → Fiction Creation → Interpretation → Testing Actively

1. Data Collection

The analysis begins by systematically gathering raw data from the e-commerce business's transactional systems. Two structured data files were utilized:

Order Details Data (Details.csv):

- Contains transactional details for each order, including Order ID, Amount (sales revenue), Profit, Quantity, Category (Clothing, Electronics, Furniture), Sub-Category (e.g., Saree, Phones, Chairs), and PaymentMode (e.g., UPI, COD, Credit Card, EMI).

Order & Customer Data (Orders.csv):

- Records order-level and customer-related information, mapping each Order ID to Order Date, Customer Name, State, and City.

2. Data Cleaning and Preprocessing

Ensuring analysis accuracy required careful data preparation, working with the actual structure of the files:

Merging Datasets:

- Using the shared Order ID column, Details.csv and Orders.csv were merged into a unified dataset. This process linked transactional sales amounts/profits with each customer's location and transaction date.

Handling Missing Values:

- Critical columns such as Amount, Profit, and State were checked for missing or null entries. Based on the data, any problematic rows (such as those missing core identifiers or sales figures) were either flagged for review or excluded to preserve integrity.

Validating Data Types:

- Columns were carefully checked and converted:
    - Order Date from text to date/time,
    - Amounts and Profits to numeric types,
    - Quantities as integers.

Duplicate Removal:

- The dataset was scanned for duplicate Order IDs to ensure that no sales or profit numbers were double-counted.

---

3. Data Analysis

Key business metrics were then calculated from the cleaned, merged data:

Key Performance Indicators (KPIs):

- Total Sales Revenue:
  $\text{Total Amount} = \sum \text{Amount}$

- Total Profit:
  $\text{Total Profit} = \sum \text{Profit}$

- Total Quantity Sold:
  $\text{Total Quantity} = \sum \text{Quantity}$

Segmented Aggregation:

- Grouped and summed data by Category and Sub-Category to identify best-performing product lines.

- Calculated state-wise sales totals to reveal geographical revenue hotspots.

- Aggregated Amount by PaymentMode (e.g., COD, UPI) to see customer payment preferences.

- Grouped sales by month using Order Date to uncover time-based trends or seasonality.

---

4. Compilation & Summary

All calculated results were organized and compiled for executive reporting:

Headline KPIs:

- Total Sales Amount: ₹438,000

- Total Profit: ₹37,000

- Total Quantity Sold: 5,615 units

Summary Tables:

- A breakdown of sales, profit, and quantity by Category (Clothing, Electronics, Furniture).

- A granular list of Sub-Category metrics revealed that items like Printers, Phones, and Bookcases led in profit and sales.

- Geographic summary highlighted Maharashtra and Madhya Pradesh as top revenue-generating states.

- PaymentMode summary showed Cash on Delivery (COD) as the dominant method, followed by UPI and Credit Card.

---

5. Visualization

Data was brought to life with Power BI, transforming complex findings into clear visuals:

- Donut Chart: Displayed percentage of total quantity sold by Category, confirming Clothing as the volume leader.

- Bar Charts: Compared sales and profit across Sub-Categories, highlighting top and bottom performers.

- Bar Chart (by State): Visualized revenue concentration geographically, with Maharashtra clearly ahead.

- Line/Bar Chart (by Month): Showed monthly profits, indicating periods of loss and peak performance.

- Pie Chart (by Payment Mode): Revealed COD as the preferred payment option among customers.

- Card Visuals: Placed KPIs such as revenue, net profit, and total quantity front and center for at-a-glance decisions.

---

6. Data Storytelling

The combined visuals paint a coherent story reflecting overall business performance:

"In the analysis period, the e-commerce business achieved revenue of ₹438,000 and a net profit of ₹37,000. The main value drivers were high-ticket Electronics and Furniture—especially Printers, Phones, and Bookcases. Maharashtra stood out as the strongest market. Customer trust in COD dominated, making up nearly half the transactions. However, some popular Clothing sub-categories, like Sarees and Handkerchiefs, were found to be operating at a loss—highlighting issues that demand urgent review."

---

7. Interpretation

Data insights led to critical business conclusions:

- **Product Reliance**: Revenue and profit heavily depended on Electronics and Furniture, especially sub-categories like Printers and Bookcases. Retaining and growing these lines is crucial.

- **Geographic Opportunity:** High sales in Maharashtra can be used as proof-of-concept for targeting other states with similar strategies.

- **COD Risk:** High COD usage reveals customer habits but also adds financial risk (returns, cash handling) and operational cost.

- **Profit Leaks**: Sub-categories in Clothing, despite high volume, dragged profitability down—posing a challenge for pricing, returns, or sourcing.

---

8. Testing Actively

Findings guided actionable business experiments:

- **In-Depth Clothing Review:**
  Hypothesis: Losses in Saree and Handkerchief sales stem from pricing or high returns.
  Action: Audit return reasons, monitor competitor prices.

- **Geographic Expansion:**
  Hypothesis: Targeted Chair and Phone ads in Maharashtra will boost revenue by 15%.
  Action: Run campaign, compare actual vs. predicted sales uplift.

- **Promote Prepaid Payments:**
  Hypothesis: 2% UPI/Credit Card discount will shift COD to prepaid by at least 10%.
  Action: Launch A/B test at checkout, measure result.

- **Inventory Tuning:**
  Hypothesis: Pre-stocking bestsellers a month ahead of trend curve prevents out-of-stocks and maximizes sales.
  Action: Adjust procurement plan based on cyclical trends in the monthly data.

Advantages of Descriptive Analytics

1. Enables Fact-Based Decision Making
Descriptive analytics turns raw sales data into clear insights so decisions are grounded in evidence, not guesswork.
Example:
"Instead of guessing which customer to target, we can see from the 'Top Customers' bar chart that Harivansh is the highest spender. This fact-based insight allows us to focus loyalty offers on our most valuable customers."

---

2. Clarifies Complex Information
It transforms thousands of scattered data points into simple, meaningful visuals.
Example:
"The 'Quantity by Category' donut chart takes sales from hundreds of products and instantly shows that Clothing accounts for 63% of our sales volume. Without this, identifying the top category from raw data would be overwhelming."

---

3. Simplifies Communication Across Teams
Dashboards make sharing findings quick and universally understandable, keeping everyone on the same page.
Example:
"When coordinating with the sales team, we can just refer to the 'Revenue by State' bar chart. Seeing that Maharashtra leads in revenue helps the team align on focusing campaigns there, without wading through spreadsheets."

---

4. Highlights Key Performance Indicators (KPIs)
It presents essential metrics at a glance, making monitoring easy and continuous.
Example:
"The cards at the top of our dashboard display headline numbers — 438K in Revenue, 5615 in Total Quantity, and a Net Profit of 37K. With these metrics front and center, management instantly sees company performance."

---

5. Affordable and User-Friendly Implementation
Descriptive analytics can be performed using standard tools without costly investments.
Example:
"Our dashboard is built using standard CSV files and Power BI — no expensive software or advanced programming needed. This makes powerful analytics accessible to any team."

---

6. Fast, Automated Reporting
Modern BI tools generate fresh reports instantly with new data inputs.
Example:
"Once our Power BI dashboard is set up, we only need to upload new sales data and hit refresh. All visuals and KPIs are updated in seconds, keeping our business insights current without manual calculations."

Disadvantages of Descriptive Analytics

1. Only Shows "What Happened," Not "Why"

- Explanation: Descriptive analytics summarizes past results but cannot identify the reasons behind them.

- Example:
  Dashboard's KPI cards and graphs (like "Revenue by State" or "Profit by Month") show that profit fell in June and July (visible as negative bars), but you can't see *why* profits dropped—for example: Was it fewer orders, a specific state underperforming, high returns, new competitors, or failed promotions? To answer "why," you'd need more detailed analysis.

2. Analysis Can Be Overly Simplistic

- Explanation: It shows basic trends or totals, but may miss complex patterns created by combinations of factors.

- Example:
  The "Quantity by Category" donut chart simply shows Clothing as the top-selling category. But it won't reveal if a specific festival, an influencer campaign, AND a special discount together caused a sales spike—those linkage effects are hidden because the dashboard looks at only one or two variables at a time.

3. Not Suitable for Real-Time Decisions

- Explanation: Descriptive dashboards analyze data after the fact, not as it happens, making them unsuited for immediate responses.

- Example:
  Dashboard is likely updated daily or monthly from CSV files. If the website or payment system fails during business hours, you'll only see a dip in "Total Quantity" or "Revenue" after the fact. You miss the chance for real-time alerts or fixes, because insights are always backward-looking.

4. Struggles with Unstructured Data

- Explanation: It works best with numbers and organized tables; not effective for unstructured data like comments, reviews, or images.

- Example:
  Your dashboard shows structured data—totals, charts, and card KPIs. It can calculate the "best customer" by sales value, but if you had thousands of customer feedback emails, it could not automatically flag common complaints ("late delivery" or "damaged product"). Specialized text analysis tools, not standard descriptive analytics, would be required for that.