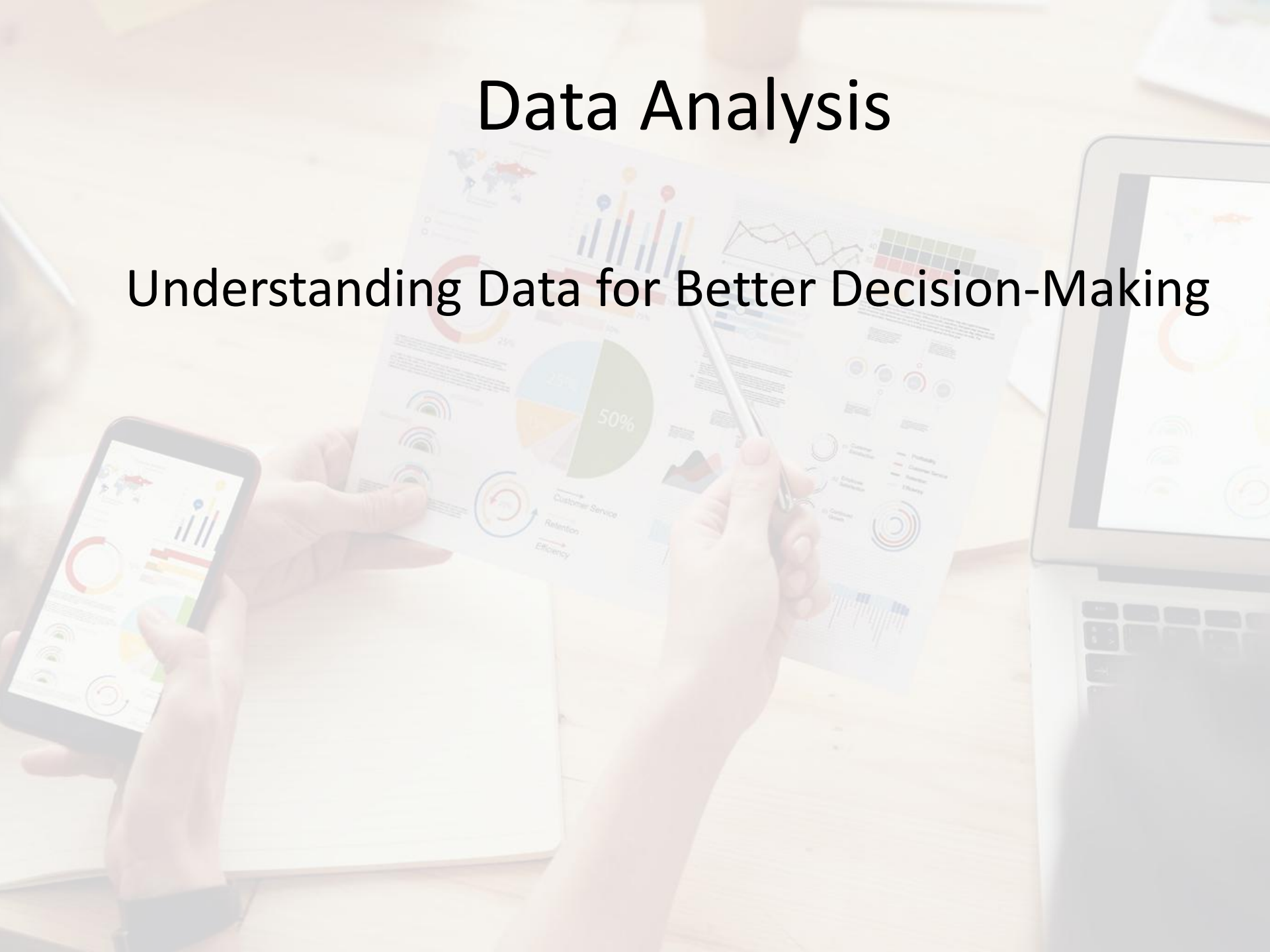


# Data Analysis

Understanding Data for Better Decision-Making



# Agenda

1. What is Data Analysis?
2. Importance of Data Analysis
3. Types of Data Analysis
4. Data Analysis Process
5. Data Collection & Cleaning
6. Exploratory Data Analysis (EDA)
7. Tools & Applications
- 8. Case Studies and Summary

# What Is Data Analysis?

- Data analysis is the process of collecting, cleaning, transforming, and interpreting data to discover useful insights.
- Example: Netflix uses viewer history to recommend movies.

# The Goal of Data Analysis

- Convert raw data into meaningful information
- Support decision-making
- Identify trends and patterns
- Example: E-commerce companies analyze purchase data to plan discounts.

# Real-Life Examples

- Business: Amazon recommends products
- Health: Hospitals track patient symptoms
- Education: Universities monitor student grades
- Agriculture: IoT sensors measure soil moisture

# Why Data Analysis Matters

- Informed decision-making
- Improved productivity
- Cost reduction
- Customer insight and personalization
- Competitive advantage
- Example: Gojek uses ride data to optimize driver locations.

# Data-Driven Decision Example

- Case: Starbucks uses location and sales data to decide where to open new stores.
- Result: Higher profitability and better local engagement.

# Types of Data Analysis

- Descriptive: What happened?
- Diagnostic: Why did it happen?
- Predictive: What will happen?
- Prescriptive: What should we do?

# Descriptive – What happened?

- In the last quarter, our online sales increased by 25% compared to the previous quarter.
- (Focus: Summarizing past data to describe trends or outcomes.)

# Diagnostic – Why did it happen?

- Sales increased because we launched a targeted social media campaign and offered a 10% discount to new customers.
- (Focus: Finding the reasons or causes behind the trend.)

# Predictive – What will happen?

- Based on current trends, we predict that online sales will grow by another 15% next quarter.
- (Focus: Using data models to forecast future outcomes.)

# Prescriptive – What should we do?

- To sustain growth, we should expand the campaign to new markets and introduce loyalty rewards for repeat buyers.
- (Focus: Recommending actions to achieve desired results.)

# Example for Each Type

- Descriptive: Sales increased 20% last month
- Diagnostic: Sales rose due to social media ads
- Predictive: Predict sales for next quarter
- Prescriptive: Increase ads before weekends

# Data Analysis Process Overview



1. Define the problem
2. Collect data
3. Clean data
4. Explore and visualize
5. Analyze
6. Interpret and report

# Step 1 - Define the Problem

- Start by asking the right question.
- Example: Why are users uninstalling our app?
- Clear goals lead to focused analysis.

# Step 2 - Data Collection

- Sources of data:
- Surveys and questionnaires
- Databases and logs
- APIs (weather, finance)
- Open datasets (Kaggle, WHO, BPS)
- Example: A school collects attendance data to monitor performance.

# Step 3 - Data Cleaning

- Garbage in → Garbage out
- Remove duplicates
- Handle missing values
- Fix inconsistent formats
- Check for outliers
- Example: Remove duplicate sales entries before analysis.

# Step 4 - Exploratory Data Analysis (EDA)

- EDA helps understand data patterns and detect anomalies.
- Example: Plotting student grades reveals performance clusters.
- Tools: Excel, Python (Matplotlib, Seaborn), Power BI

# Step 5 - Analysis and Modeling

- Common techniques:
- Regression (Predict continuous values)
- Classification (Categorize outcomes)
- Clustering (Group similar data)
- Association rules (Market basket analysis)



# Common Machine Learning Techniques

Regression • Classification •  
Clustering • Association Rules

# 1. Regression (Predict Continuous Values)

- Example:
- Predicting house prices based on factors like square footage, number of rooms, and location.
- Input: Size = 150 m<sup>2</sup>, Rooms = 3, Location = Suburban
- Output: Predicted Price = \$120,000
- Techniques: Linear Regression, Random Forest Regressor

## 2. Classification (Categorize Outcomes)

- Example:
- Classifying emails as spam or not spam based on keywords and sender information.
- Input: Email text and sender domain
- Output: Spam / Not Spam
- Techniques: Logistic Regression, Decision Tree, SVM

# 3. Clustering (Group Similar Data)

- Example:
- Grouping customers based on their shopping behavior to identify market segments.
- Input: Purchase frequency, spending amount, product categories
- Output: Cluster 1 = Budget Shoppers, Cluster 2 = Premium Buyers
- Techniques: K-Means, Hierarchical Clustering

# 4. Association Rules (Market Basket Analysis)

- Example:
- Finding patterns in supermarket transactions such as:
  - Rule: If a customer buys bread and butter, they are likely to buy milk.
- Techniques: Apriori Algorithm, FP-Growth

# Example of Analysis

- Scenario: Predicting House Prices
- Features: Size, location, rooms
- Model: Linear Regression
- Output: Predicted price for each new house

# Step 6 - Interpretation

- Translate technical results into business meaning
- Connect findings with goals
- Use clear visuals and summaries
- Example: Instead of  $r = 0.84$ , say 'Strong relationship between ad spend and sales.'

# Data Visualization

- Purpose: Communicate complex data clearly
- Examples:
  - Bar chart: Compare sales across regions
  - Pie chart: Market share
  - Line chart: Trend over time
  - Heatmap: Correlation
- Tools: Power BI, Tableau, Python

# Storytelling with Data

- Combine visuals + narrative:
- 'Sales grew 25% after launching digital ads in April.'
- Helps audience understand why and how, not just what.

# Tools for Data Analysis

- Spreadsheet: Excel, Google Sheets
- Programming: Python, R
- Visualization: Power BI, Tableau
- Database: SQL, MongoDB
- Big Data: Spark, Hadoop

# Example - Python in Data Analysis

- `import pandas as pd`
- `df = pd.read_csv('sales.csv')`
- `print(df.describe())`
- Output: Summary statistics for sales data.
- Helps detect outliers and trends quickly.

# Applications by Field

- Business: Customer segmentation
- Healthcare: Predicting disease risk
- Agriculture: Smart irrigation with sensors
- Education: Student performance tracking
- Social Media: Sentiment analysis

# Mini Case Study

- Case: Amazon's Recommendation Engine
- Uses collaborative filtering
- Suggests products based on behavior
- Increases revenue by 35%

# Challenges in Data Analysis

- Poor data quality
- Lack of domain understanding
- Misinterpretation of results
- Ethical & privacy concerns
- Example: Misusing personal data can lead to GDPR violations.

A person's hands are shown holding a tablet displaying various data visualizations, including a world map, bar charts, and a donut chart. In the background, a laptop screen also displays similar data visualizations. A printed report with various charts and graphs is spread out on a wooden desk. The overall scene is a professional workspace focused on data analysis and visual communication design.

# Data Analysis in Visual Communication Design

Leveraging Data to Enhance Visual Communication Design

# Introduction to Data Analysis

- What is Data Analysis?
  - Data analysis involves systematically examining raw data to extract meaningful insights and support decision-making.
  - In design, it helps assess how visuals impact user behavior, engagement, and satisfaction.
- Importance of Data in Design
  - Data guides designers in making informed choices about color schemes, typography, layout, etc. to meet audience needs and business goals.

# Role of Data in Visual Communication Design

- Analyzing Audience Behavior
  - User interaction, engagement metrics, and feedback help improve design effectiveness.
- Improving Design Effectiveness
  - Data-driven decisions optimize designs for user engagement and actions.
- Personalization of Visual Content
  - Tailor visuals for specific audience segments based on demographic and behavioral data.

# Types of Data Used in Visual Communication

- Quantitative Data

- Engagement metrics (e.g., page views, conversion rates)
- Sales data to understand design impact on purchases

- Qualitative Data

- User feedback, reviews, and preferences to understand emotional responses to visuals.

# Tools and Techniques for Data Analysis in Design

- Analytics Software
  - Google Analytics, Adobe Analytics
- Heatmaps and Click Tracking
  - Tools like Hotjar to visualize user interactions with designs
- A/B Testing
  - Comparing design versions to determine which performs better.

# Data Interpretation for Design Improvements

- Identifying Trends and Patterns
  - Analyzing user behavior over time to identify successful design elements.
- A/B Testing Results
  - Evaluate design variations based on user interaction and performance.
- Optimizing Visuals Based on Data
  - Fine-tune design elements (e.g., colors, layout) based on real-time insights.

# Case Studies: Data Analysis in Design

- Example 1: Website Redesign
  - Using heatmaps and user feedback to streamline navigation and layout.
- Example 2: Ad Campaign Analysis
  - Data-driven adjustments to target specific user segments and improve ad performance.

# Challenges in Data-Driven Design

- **Balancing Data with Creativity**
  - Ensuring creativity isn't stifled by over-reliance on data.
- **Data Overload**
  - Managing large volumes of data to focus on key metrics.
- **Privacy and Ethical Concerns**
  - Responsible handling of user data while adhering to privacy laws.

# Conclusion

- Key Takeaways

- Data analysis helps optimize design for better user engagement and outcomes.

- Final Thoughts

- Data-driven design is essential for creating impactful visuals that resonate with audiences.

# Summary & Key Takeaways

- Data analysis transforms raw data into actionable insight
- Steps: Define → Collect → Clean → Analyze → Visualize → Report
- Tools: Excel, Python, Power BI, Tableau
- Always communicate results clearly and ethically.