DATA COLLECTION AND DATA EXTRACTION

PRESENTED BY – REMYA R S

Types of Data Sources

Types of Data Sources

Internal Data Sources

Data generated within the organization

External Data Sources

Data generated outside the organization

Types of Internal Data Sources

Company Databases

Centralized repositories for storing structured data to store and manage data from various business operations

Enterprise Systems

Manages company's interactions with customers, to improve customer relationships

Transactional Data

Data generated from business transactions, used for tracking operational efficiency

Types of External Data Sources

Market Research Reports

Helps In-depth analysis, expert insights, trend identification

Social Media

Real-time data to do customer sentiment analysis

Public Datasets

Data collected and made available by government agencies, non-profits, and academic institutions

Sources: Data.gov, World Bank, Kaggle

kaggle

UCI Machine Learning Repository

data.gov.in
Open Government Data (OGD) Platform India
सत्यमेव जयते

Data Collection Methods

Data Collection Methods

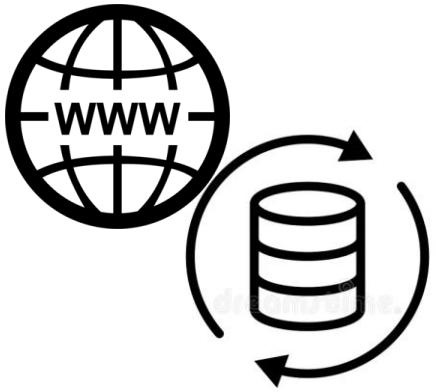
Aspect	Primary Data Collection	Secondary Data Collection
Definition	Gathering new, original data directly from sources for a specific purpose.	Using existing data collected and published by others.
Methods	Surveys, Questionnaires, Interviews, Observations, Experiments	Books, Articles, Government Reports, Online Databases, Company Records
Advantages	- Specific to the research needs - Up-to-date	- Cost-effective - Time-saving
Disadvantages	- Time-consuming - Can be costly	- Less control over data quality - May not perfectly fit research needs

Primary Data Collection Methods

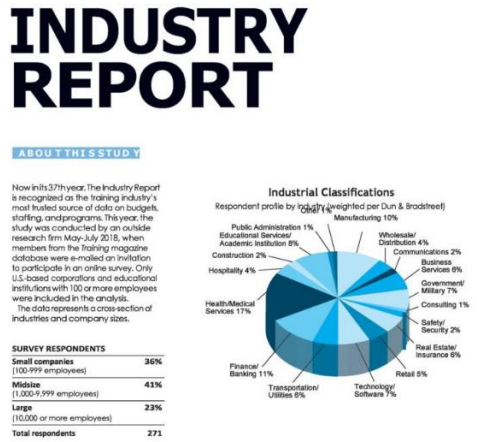
OBSERVATIONS

Secondary Data Collection Methods

Literature Review



DATABASE



Types of Business Analytics

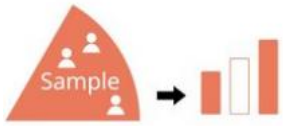
Aspect	Descriptive Analytics	Inferential Analytics	Predictive Analytics
Definition	Focuses on past data to understand what has happened.	Focuses on making generalizations or inferences about a population based on a sample of data.	Focuses on using historical data to predict future outcomes and trends.
Purpose	To provide insights and understand past behavior.	To draw conclusions and make inferences about a larger group from a smaller sample.	To forecast future events based on historical patterns and data.
Methods/Tools	Statistical measures (mean, median, mode), data visualization (charts, graphs)	Hypothesis testing, confidence intervals, regression analysis	Machine learning algorithms, time series analysis, predictive modeling
Examples	Reporting on last year's sales figures, summarizing customer demographics	Estimating the average income of a population based on a sample, testing the effectiveness of a new drug	Predicting next month's sales, forecasting customer churn

Analyzing Data for a Business

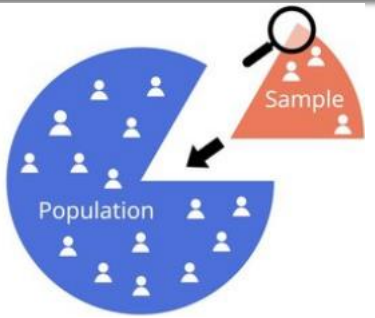
Understanding Problem Statement

Data Collection

Data Cleaning



Descriptive Analysis



Inferential Analysis

Data Visualization

Predictive Analysis


Thank
You