SOURCES AND METHODS OF COLLECTING DATA

1. PRIMARY VS. SECONDARY DATA

Primary data refers to information gathered by the researcher specifically for the research project at hand. This data is original and collected directly from the source to address the immediate marketing problem or objective. Secondary data is information that has been previously collected and published for a purpose other than the current research project. It is readily available from internal or external sources.

Feature	Primary Data	Secondary Data
Collection Time	Long	Short (Instant)
Collection Cost	High	Low or Free
Specificity/Relevance	High (Perfect Fit)	Low (May not fit the problem)
Control	High (You control methodology)	Low (No control over collection)
Availability	Not available yet	Readily available
Examples	Focus group transcripts, custom survey results	Census reports, competitor sales data, academic studies

1.1 Advantages and Disadvantages of Primary Data

Category	Advantage	Disadvantage
Specificity	High relevance, tailored to exact needs.	Time-consuming to collect and analyze.
Control	Full control over methodology and data quality.	High cost for labor, materials, and incentives.
Currency	Data is current and timely.	Requires specialized skills and expertise.

1.2 Advantages and Disadvantages of Secondary Data

Category	Advantage	Disadvantage
Cost & Speed	Quick access and low cost (often free).	Lack of fit; may not perfectly align with the research problem.
Foundation	Provides context and helps define the problem.	Accuracy is often unknown or unverifiable.
Availability	Easy to access through internal or external sources.	Data may be outdated (poor currency).

2. CRITERIA FOR EVALUATING SECONDARY DATA

Because secondary data was collected for a different purpose, researchers must evaluate its quality and fit before use.

Criteria	Question to Ask	Simpler Explanation
Error/Accuracy	How was the original data collected? What were the sample size and response rate?	Is the data believable? Check the source's methodology to see if they made mistakes (e.g., poor sampling, leading questions).
Currency	When was the data collected? Is the information still relevant today?	If the market changes quickly (e.g., technology trends), five-year-old data is useless. Ensure it's recent enough.
Objective	Why was the data originally collected? Does that objective align with our needs?	Data collected by a government for tax purposes might define 'business' differently than you need for a consumer study. Check the original goal.
Nature	How were key variables, units of measure, and classifications defined?	If you need data on "millennials," but the source groups them as "under 35," the data's definition doesn't fit your needs. Check the *fine print* on definitions.
Dependability	Who collected the data? Are they a credible, reputable, and unbiased source?	Always check the author or publisher. Is it a respected university (high dependability) or a blogger with a clear bias (low dependability)?

3. CLASSIFICATION OF SECONDARY DATA

Secondary data is broadly categorized by its origin: inside the company (Internal) or outside the company (External).

3.1 Internal Secondary Data

This data is generated within the organization during the routine course of business operations (e.g., sales, accounting, marketing).

3.1.1 Customer Databases

Systematic records containing purchase history, customer demographics, transaction dates, contact information, and service inquiries. This is the foundation for understanding who buys what and when.

3.1.2 Data Warehousing and Data Mining

- **Data Warehousing (Textbook):** A centralized, integrated database containing historical, aggregated data from various internal sources (sales, logistics, inventory) used for management decision-making and analysis.
  - **Simpler Explanation:** It is a massive, organized digital archive that stores all of a company's past data in one place, making it easy to analyze trends over long periods.
- **Data Mining (Textbook):** The process of automatically discovering hidden patterns, predictive information, and valuable relationships in large databases using advanced statistical and mathematical techniques.
  - **Simpler Explanation:** Using powerful software to automatically find non-obvious secrets (e.g., "Customers who buy Product A also tend to buy Product B three weeks later") within the data warehouse.

3.1.3 CRM and Database Marketing

- **Customer Relationship Management (CRM):** A system and strategy for managing a company's interactions with current and potential customers, primarily using the database to personalize communications.
- **Database Marketing:** Utilizing the customer database to execute targeted marketing campaigns, such as direct mail or personalized emails, based on segmentation (e.g., targeting high-value customers with loyalty offers).

3.2 External Secondary Data

This data originates **outside the organization** and is available publicly or commercially.

3.2.1 Business/Non-government Sources

Data provided by commercial organizations, often published for profit, including industry reports, trade journals, professional organization reports, and specialized market forecasts (e.g., reports from Gartner, Nielsen, or trade associations).

3.2.2 Government Sources

Official data collected by local, state, or federal government agencies. This is often the most reliable source for population demographics, economic activity, and industry statistics (e.g., Census data, labor statistics, import/export records).

3.2.3 Syndicated Services

Research organizations that collect standardized information (often continuous, subscription-based data) and sell it to multiple clients. Instead of paying for a custom report, you "subscribe" to a shared, standardized report that tracks a defined market area (e.g., grocery sales, TV viewing habits). Syndicated Services source their data from households and institutions.

Household/Consumers

- **Surveys (Psychographics and Lifestyles):** Measuring consumer attitudes, interests, opinions (AIOs), and consumption habits (e.g., large annual studies of general population behavior).
- **Panels (Purchase & Media):** A group of households that agree to provide information regularly over time.
  - *Purchase Panels:* Recording every purchase made (e.g., keeping a diary or submitting receipts).
  - *Media Panels:* Recording media exposure (e.g., TV viewership, website visits).
- **Electronic Scanner Services:** Data collected from retail store checkout scanners.
  - *Volume Tracking Data:* Provides sales, price, and inventory information.
  - *Scanner Panels:* Tracks purchase behavior over time from specific households whose purchases are linked to their demographic data via loyalty cards.
  - *Scanner Panels with Cable TV:* Links household purchases to the specific TV commercials they were exposed to, allowing for sophisticated advertising effectiveness analysis.

Institutions

- **Retailers and Wholesalers (Auditors):** Syndicated firms physically audit stores to check inventory levels, shelf facings, pricing, and point-of-sale display usage. This tracks product movement *out* of the store (sell-through).
- **Industrial Firms (Direct Inquiries, Corporate Reports):** Data on competitors' industrial products, market size for specific machinery, and information gathered from annual reports, press releases, and trade inquiries.

3.3 Social Media (Internal as well as External)

Social media data is unique as it serves as both internal (data from a company's own Facebook or LinkedIn pages) and external (public conversations about the brand or competitors).

- **Social Listening (External):** Using specialized tools to monitor and analyze millions of online conversations, mentions, and sentiment across platforms (Twitter, blogs, forums) to gauge public attitude toward a brand or topic.
- **Engagement Metrics (Internal):** Data related to user interactions with the company's own content (likes, shares, comments on owned pages).