

Data Quality Assignment

Hameed Jasim



APPENDIX

- Source for missing Data
- Process for performing data analysis

SOURCE FOR MISSING DATA

STEP 1: IDENTIFY MISSING DATA

- Start by analyzing your dataset and identifying the startup funding rounds that have missing values for the number of investors and amount raised.

STEP 2: PUBLIC DATABASES AND PLATFORMS

1. **Crunchbase:** Crunchbase is a popular platform that provides information about startups, investments, funding rounds, and investors. You can search for each startup in your dataset and find historical funding data.
2. **AngelList:** AngelList is another platform that focuses on startup investments. It allows startups to connect with investors and showcase their funding rounds.
3. **PitchBook:** PitchBook is a financial data platform that offers comprehensive information about private and public companies, including funding data.

STEP 3: STARTUP WEBSITES AND PRESS RELEASES

–Visit the official websites of the startups in your dataset and look for press releases, news articles, or blog posts that announce funding rounds. Companies often provide information about their funding activities on their websites

STEP 4: NEWS ARTICLES AND MEDIA COVERAGE

– Search for news articles, blog posts, and media coverage related to the startups. News outlets often report on funding rounds and provide details about the investment amounts and investors involved.

STEP 5: SOCIAL MEDIA AND INDUSTRY REPORTS

– Check the startups' social media profiles (e.g., Twitter, LinkedIn) for announcements about funding rounds. Additionally, industry reports and market research might provide insights

STEP 6: INVESTOR WEBSITES

- Some investors maintain portfolios on their websites, showcasing the startups they've invested in and the funding rounds they've participated in.

STEP 7: CONTACTING STARTUPS AND INVESTORS DIRECTLY

– If you're unable to find the required information from public sources, consider reaching out to the startups or investors directly. Startups might be willing to share their funding details, especially if they're interested in being featured in your dataset.

STEP 8: DATA ENRICHMENT SERVICES

- There are data enrichment services and APIs available that can provide additional information about startups, including funding details.

STEP 9: DATA CLEANING AND INTEGRATION

- Once you've gathered the missing data, integrate it into your dataset and ensure consistency in formatting and accuracy.

STEP 10: DOCUMENTATION AND ATTRIBUTION

- Document the sources of the newly collected data for transparency and proper attribution. This is important for maintaining the integrity of your dataset.

STEP 11: VALIDATE AND CROSS-CHECK

- Cross-check the newly collected data with existing data to ensure accuracy and consistency.

STEP BY STEP PROCESS FOR COLLECTING DATA

1. DEFINE THE PURPOSE AND RESEARCH QUESTIONS:

Clearly articulate the objectives of your data collection. What are you trying to achieve?
What specific questions do you need to answer using the collected data?

2. IDENTIFY DATA SOURCES:

Determine where and how you will gather your data. Possible sources include surveys, observations, existing databases, interviews, experiments, online sources, and more.

3. DESIGN DATA COLLECTION METHODS:

Choose appropriate methods for data collection based on your research questions and available resources. Common methods include:

- Surveys/questionnaires
- Interviews (structured, semi-structured, or unstructured)
- Observations (participant or non-participant)
- Experiments (controlled settings to test hypotheses)
- Existing datasets (secondary data)

4. DEVELOP DATA COLLECTION TOOLS:

Create the necessary tools for data collection, such as survey forms, interview guides, observation checklists, or experiment protocols. Ensure that these tools align with your research objectives.

5. PILOT TESTING:

Before conducting full-scale data collection, perform a pilot test on a small sample. This helps identify any issues with your data collection tools, instructions, or procedures and allows for necessary adjustments.

6. SAMPLING STRATEGY:

Decide on the sampling technique that best suits your research, whether it's random sampling, stratified sampling, convenience sampling, etc. This will help ensure your collected data is representative of the target population.

7. DATA COLLECTION:

Execute your data collection methods according to the established plan. Pay attention to detail, maintain consistency, and follow your data collection tools precisely.

8. DATA VALIDATION AND QUALITY CONTROL:

Implement measures to ensure data quality and reliability:

- Check for errors, missing values, and outliers.
- Cross-verify responses if possible.
- Conduct periodic checks during data collection to catch issues early.

9. DATA ENTRY AND STORAGE:

If you're using paper-based methods, ensure accurate data entry. If digital tools are used, establish a secure and organized data storage system that follows privacy regulations.

10. DATA CLEANING AND PREPROCESSING:

Prepare the collected data for analysis by cleaning and preprocessing it:

- Handle missing values.
- Standardize formats and units.
- Correct errors and inconsistencies.

11. DATA DOCUMENTATION:

Maintain thorough documentation about your data collection process, including methods, tools, potential biases, and any changes made along the way.

12. DATA ETHICS AND PRIVACY:

Respect ethical considerations and privacy regulations while collecting and using data. Obtain necessary permissions, inform participants about the purpose of data collection, and ensure data security.

13. DATA ANALYSIS AND INTERPRETATION:

Once your data is cleaned and ready, perform the analysis required to answer your research questions. Choose appropriate statistical or qualitative techniques based on your data type and research objectives.

Thank you!

Hameed Jasim
Data Enthusiast