

Data Collection and Assumptions Documentation

1. Document Overview

This document provides a detailed account of data sources, methodologies, and assumptions used during the data collection process for Project XYZ. It aims to ensure transparency and reproducibility.

2. Data Sources

Source Name	Description	Data Collected	Date Accessed
Internal Database	Company's sales and inventory records	Sales volumes, product SKUs	2024-05-10
Customer Survey	Online survey of active customers	Satisfaction scores, feedback	2024-05-15
External Report	Industry growth projection 2024	Market trends, competitor analysis	2024-05-18

3. Data Collection Methodology

- Compiled sales and inventory figures from internal CRM using SQL queries.
- Distributed survey link via email to a random sample of 500 users.
- Extracted relevant tables from the external PDF industry report using OCR tools.

4. Assumptions

- All sales data recorded before 2023 is complete and accurate.
- Survey sample is representative of the customer base.
- Market trends noted in the external report are expected to continue for the next year.
- Data from all sources can be consolidated without significant duplication.

5. Data Limitations

- Survey response rate was 38%, possibly introducing participation bias.
- Some external report figures are estimates, not actuals.
- Data synchronization between systems is performed weekly; minor delays may exist.

6. Change Log

Date	Change Description	Author
2024-05-20	Initial document creation	A. Lee
2024-05-25	Updated survey response details	C. Smith

Important Notes

- This document should be kept updated as data sources or methodologies change.
- All assumptions must be regularly reviewed and validated.
- Documenting limitations increases the reliability of conclusions drawn.
- Transparency in changes ensures accountability and reproducibility.