

Data Quality & Governance Report

Dataset: CFPB Consumer Complaints

Prepared by: Adithya Prahasith

Records Analyzed: 62,517

1. Executive Summary

This report outlines the data quality assessment and governance checks performed on the CFPB Consumer Complaints dataset. The objective was to ensure the dataset meets the required standards of completeness, accuracy, consistency, and integrity before further analysis or building dashboards.

2. Data Dictionary

Field	Description
Complaint ID	The unique identification number for a complaint
Submitted via	How the complaint was submitted to the CFPB
Date submitted	The date the CFPB received the complaint
Date received	The date the CFPB sent the complaint to the company
State	The state of the mailing address provided by the consumer
Product	The type of product the consumer identified in the complaint
Sub-product	The type of sub-product the consumer identified in the complaint (not all Products have Sub-products)
Issue	The issue the consumer identified in the complaint (possible values are dependent

	on Product)
Sub-issue	The sub-issue the consumer identified in the complaint (possible values are dependent on Product and Issue, and not all Issues have corresponding Sub-issues)
Company public response	The company's optional, public-facing response to a consumer's complaint. Companies can choose to select a response from a pre-set list of options that will be posted on the public database. For example, "Company believes complaint is the result of an isolated error."
Company response to consumer	This is how the company responded. For example, "Closed with explanation."
Timely response?	Whether the company gave a timely response (Yes/No)

3. Data Quality Rules & Findings

Validation rule	Records Failing	% of Dataset	Action Taken
Unique Complaint ID	0	0%	No duplicates found
Missing State	0	0%	No Missing values
Invalid Timely Response? (Yes/No)	1494	2.3%	Dropped Missing Values
Date Submitted >= Date Received	0	0%	All date pairs are valid
Missing Company response to consumer	0	0%	No missing values
Invalid Product or category issues	0	0%	No invalid values

3. Actions Taken

- Converted date columns (Date submitted, Date received) into valid datetime format using Python.
- Verified chronological consistency of complaint submission and receipt.
- Removed missing entries in Timely response? for clarity and completeness.
- Ensured all categorical fields (e.g., Product, Sub-product, Issue, Sub-issue) conformed to valid domain values.
- Verified uniqueness of the primary key (Complaint ID) to ensure reliable joins.

Please Note: This Report is only for simulating a DQ report , and only be referred for educational purposes.