Datasheet for CFPB Financial Well-Being Scale Dataset*

Xinxiang Gao

December 3, 2024

1 Datasheet for CFPB Financial Well-Being Scale Dataset

1.1 Motivation

1. *For what purpose was the dataset created?*
   The dataset was created to develop and validate a scale that measures individual financial well-being, aiming to provide a standardized tool for assessing this concept across diverse populations. It serves researchers, policymakers, and practitioners in understanding and improving financial health.

2. *Who created the dataset?*
   The dataset was developed by the Consumer Financial Protection Bureau (CFPB) with contributions from experts in psychometrics and survey methodology. Consumer Financial Protection Bureau (2017)

1.2 Composition

1. *What do the instances represent?*
   Each instance represents an individual respondent's answers to the Financial Well-Being Scale items, along with demographic and other contextual information.

2. *How many instances are there in total?*
   The dataset contains data from 6,394 respondents who participated in the national survey.

---

*Code and data are available at: https://github.com/xgao28/financial_well_being_analysis.

3. *Does the dataset contain all possible instances?*
   The dataset includes a large representative sample of U.S. adults but does not cover every individual or every conceivable demographic subgroup.

4. *What data does each instance consist of?*
   Each instance includes responses to the Financial Well-Being Scale, demographic characteristics (e.g., age, income), and potentially other financial behavior indicators.

5. *Is there a label or target associated with each instance?*
   Yes, the primary target variable is the Financial Well-Being score derived from the scale items.

6. *Is any information missing from individual instances?*
   Missing data were handled through rigorous preprocessing steps, ensuring completeness where possible.

7. *Are relationships between instances made explicit?*
   Relationships between instances are not explicitly modeled within the dataset; each response is treated as independent.

8. *Are there recommended data splits?*
   The dataset may be split into training and testing sets depending on the analysis objectives, though specific recommendations depend on the research question.

9. *Are there errors, noise, or redundancies?*
   Efforts were made during data collection and preprocessing to minimize errors and noise, but some variability inherent to self-reported data may remain.

10. *Is the dataset self-contained?*
    The dataset includes all necessary variables for calculating the Financial Well-Being Score, but external validation might require additional datasets.

11. *Does the dataset contain confidential data?*
    No, the dataset anonymizes personal identifiers to protect respondent privacy.

12. *Does the dataset contain offensive or sensitive data?*
    No, the content focuses on financial behaviors and attitudes without including sensitive personal information.

13. *Does the dataset identify subpopulations?*
    Yes, it categorizes respondents by various demographic characteristics such as age, income level, and employment status.

14. *Is it possible to identify individuals?*
    No, all personal identifiers have been removed to ensure anonymity.

15. *Does the dataset contain sensitive data?*
    While it does not include directly sensitive attributes like race or religion, it does contain financial information which can be considered sensitive.

16. *Any other comments?*
    None.

1.3 Collection Process

1. *How was the data acquired?*
   Data were collected through GfK's KnowledgePanel using address-based sampling methods.

2. *What mechanisms were used to collect the data?*
   Mechanisms included online surveys and telephone interviews to accommodate different modes of participation.

3. *If the dataset is a sample, what was the sampling strategy?*
   The sampling strategy aimed for a nationally representative sample, employing stratified random sampling techniques.

4. *Over what timeframe was the data collected?*
   Data collection occurred over a specified period designed to capture seasonal variations in financial behavior.

5. *Were any ethical review processes conducted?*
   Ethical considerations were addressed through IRB approval and adherence to best practices in human subjects research.

6. *Did you collect the data directly?*
   The CFPB contracted with GfK to manage the data collection process.

7. *Were individuals notified about the data collection?*
   Participants were informed about the study's purpose and provided consent prior to participation.

8. *Did individuals consent to data collection?*
   Yes, all participants gave informed consent before taking part in the survey.

9. *Any other comments?*
   None.

1.4 Preprocessing, Cleaning, and Labeling

1. *Was any preprocessing done?*
   Yes, extensive preprocessing was conducted, including cleaning, coding open-ended responses, and computing composite scores.

2. *Was raw data saved?*
   Yes, raw data were preserved for future reference and verification purposes.

3. *Is the preprocessing software available?*
   Preprocessing details are documented in the technical report accompanying the dataset.

4. *Any other comments?*
   None.

1.5 Uses

1. *Has the dataset been used for any tasks already?*
   Yes, the dataset has been used to establish norms for the Financial Well-Being Scale and to conduct initial validity studies.

2. *What other tasks could the dataset be used for?*
   The dataset can support research into financial education, policy evaluation, and intervention design aimed at improving financial well-being.

3. *Are there tasks for which the dataset should not be used?*
   The dataset should not be used for identifying or targeting individuals due to anonymization measures.

4. *Any other comments?*
   None.

1.6 Maintenance

1. *Who will maintain the dataset?*
   The CFPB is responsible for maintaining and updating the dataset as needed.

2. *How can the dataset owner be contacted?*
   Contact information for inquiries regarding the dataset can be found on the CFPB website.

3. *Will the dataset be updated?*
   Updates may occur periodically to reflect changes in financial circumstances and new research findings.

4. *Any other comments?*
   None.

References

Consumer Financial Protection Bureau. 2017. "Financial Well-Being Survey Data." https://www.consumerfinance.gov/data-research/financial-well-being-survey-data/.