

Logistic Regression, BNPL and Financial Wellbeing Among Students - Plan

Connor Smedley
Student ID: 710038386

1 Public Good

The first goal of this research is to empower broad public awareness and education on how usage of financing services may pose risk of developing poor financial and emotional wellbeing, thereby instilling greater financial resilience, particularly among students to avoid over-extension or debt cycling [5]. Additionally, it aims to foster understanding among financial institutions, by promoting the ethical delivery and marketing of services to keep the risk to customers and how that may be identified at the forefront. Finally, the modified data and findings will be made publicly available in a controlled access university repository to support and promote further research were deemed safe.

2 Data Description

The data collected reflects solely an individual level, primarily collecting quantitative data regarding individual financial circumstances and corresponding information concerning emotional wellbeing, level of financial assistance and self-assessed attitudinal measures. Importantly data regarding finances (expenditure and income ranges) or emotional wellbeing will primarily remain ordinal to aid in both comparability, privacy and model implementation [8]. Crucially, this will be done using a Likert scale of agree to disagree resembling 1–4 for all questions pertaining to finance or wellbeing, with the sole exclusion of income using a scale of 1–5 to allow for larger variance capture. By extension, data regarding demographics will be also categorized. Furthermore, raw data will be stored in CSV for Excel readability, with anonymized data being published in a similar format. Surveying will last a minimum of one month, beginning in February 2026 and running until March, unless the minimum level of respondents (200) has been reached.

3 Data Collection

As briefly mentioned prior, data collection will be online and distributed solely for anonymous response from students enrolled at university, following ISO 8000 standards of data collection [7]. Importantly, to ensure respondents are of the required student demographic, response forms will only be available for completion via institutional emails on a one-per-account basis, with emails being stored only temporarily during distribution phase. Prior to distribution a ten-person pilot will be run, shared privately by myself to trusted fellow students in an attempt to address any problematic wording and ensure validation mechanisms (mandatory fields and ranges) are well understood. Following the pilot and any consequential adaptations, distribution will occur via social media, posters, communication with module leaders to provide QR-links and student email lists. Lastly, data collection will follow a convenience sampling approach, due to the anticipated small sample size and time limitations. F [3]. Crucially, the ensure integrity of results sample size is a priority with the hopes of bias minimization and variance capture.

4 Data Storage and Security

Initially, while distribution is ongoing, data storage will primarily be through the secure University-approved storage system, OneDrive for Business. Immediately after responses are no longer available, data collected will be stripped of any identifiable information, such as email, and moved to both a local personal storage as well as a password accessed, encrypted cloud service which offers automated backup. Following this, responses will be removed from the survey platform to ensure

raw data no longer exists [6]. Finally, once data integrity and anonymisation has been ensured data will be published onto university repositories indefinitely [12].

5 Data Management, Documentation, and Curation

Throughout the management, documentation and curation process the FAIR data principles will be kept as a fundamental guideline [14]. Firstly, to ensure findability unique identification will be given to the data set on the university repository. Secondly, as discussed previously the use of a university repository will ensure selective accessibility, with sensitive data being completely inaccessible. Additionally, meta data will be provided describing data structure and variables to guarantee interpretability, as well as open CSV formatting, following Dublin Core metadata standards [2]. Similarly, rich meta data on data collection processes coupled with rigid version control will be the groundwork to ensure reusability of data.

6 Data Preservation

Data preservation will be ensured through version control, on a university platform. Consequently, preservation will ensure minimal requirements for manual maintenance as well as the capability to be stored near indefinitely. As previously discussed, the format will be anonymised, csv data alongside comprehensive documentation in accordance with University Research Data Management Policy [12].

7 Data Sharing and Access

Despite anonymisation precautions, to align with both GDPR and University ethics framework [4, 11, 13], access will be through a controlled university environment. However, due to the rigorous and rigid anonymity process data will follow an access upon request format where the primary researcher can monitor who and why data access is being requested will be utilized to ensure that public good is maximized through greater access to information. Moreover, this access will only be granted upon completion of a formal Data Access Agreement (DAA), to legally forbid re-identification attempts. Governance of the published dataset will be via the Creative Commons Attribution-Noncommercial-No Derivative Licence (CC BY-NC-ND) [1].

8 Ethics and Privacy

In accordance with the Confidentiality, Integrity, and Availability Triad [9] three risk levels have been identified for differing stages of research. Initially, risk must be identified as high, as information retaining to financial wellbeing as well as identifiable traits poses great risk to confidentiality and privacy. Consequently, mitigation of access is the key priority at this stage, ensuring data is kept solely in an encrypted environment. Moving on, the stage of pre-processing data introduces a move towards moderate risk as data becomes processed with the removal of easily identifiable traits, however still consistent with personal data and potentially identifiable through unique traits. As a result, data access remains strictly controlled throughout this stage, remaining only on an encrypted local storage. Lastly, through the process of generalization and suppression by applying k -anonymity to prevent information leak using $k = 3$ to mask any data retaining to financial or emotional wellbeing [10], the research moves into low risk as data cannot be identified back to individual. Overall, an important consideration for all stages is the requirement for informed consent, including participant awareness of how data will be transformed into the anonymized form. This will be ensured through a complementary consent form, with awareness and completion of this form being ensured as the first question within the survey.

9 Relevant Policies

First and foremost, research must adhere to the seven key principles of the UK General Data Protection Regulation Act (GDPR) [4] as well as the Data Protection Act 2018 [11]. Consequently, data collection has been designed in accordance with these key principles. To ensure transparency, purpose limitation and freedom to opt-out at any point, all individuals will be informed of their right to have data removed and to receive a copy of data currently held regarding themselves.

Additionally, survey design ensures a balance between minimization as well as accuracy, by small inclusion of validation questions. Furthermore, data storage and management have been proposed to ensure private data is only held for the necessary time before anonymisation and is then kept in a limited access environment, with access upon request approval.

Similarly, as a university researcher, this project must adhere to Exeter University Research Data Management Policy [12] as well as the Research Ethics Framework [13]. These frameworks also mandate adherence to FAIR data principles, controlled access, research identifiers (DOIs) and metadata documentation.

Bibliography

References

- [1] Creative Commons. (2024) *About CC Licences: CC BY-NC-ND 4.0*.
- [2] Dublin Core Metadata Initiative. (2022) *Dublin Core Metadata Element Set, Version 1.1*.
- [3] Etikan, I., Musa, S.A. and Alkassim, R.S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), pp.1–4.
- [4] European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*.
- [5] Financial Conduct Authority. (2020). *Financial Lives 2020 Survey: The Impact of Coronavirus*.
- [6] Information Commissioner's Office. (2021). *Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance*.
- [7] International Organization for Standardization. (2015). *ISO 8000-1:2015 — Data Quality — Part 1: Overview*.
- [8] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), pp.1–55.
- [9] Stallings, W. and Brown, L. (2018). *Computer Security: Principles and Practice*. Pearson.
- [10] Sweeney, L. (2002). k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), pp.557–570.
- [11] UK Parliament. (2018). *Data Protection Act 2018*.
- [12] University of Exeter. (2023a). *Research Data Management Policy*.
- [13] University of Exeter. (2023b). *Research Ethics Framework*.
- [14] Wilkinson, M.D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), p.160018.