Machine learning model that classifies Canadian citizens' financial well-being status and predicts the impact of global shocks

Huiqian Chen
Sprott School of Business
huiqianchen@cmail.carleton.ca

Aziz Al-Najjar
Department of Systems and Computer Engineering
azizalnajjar@cmail.carleton.ca

Zakaria Zoundi
Department of Economics
zakariazoundi@cmail.carleton.ca

Supervisor:
Prof. Ahmed El-Roby, School of Computer Science

Introduction

Financial well-being is an essential aspect of an individual's life. However, predicting financial well-being of Canadians has become more challenging with the recent global shock of the Covid-19 pandemic.

Using financial well-being data from both pre- and post-COVID-19 surveys, the project develops a machine learning model that can predict Canadians' financial well-being status and the potential impact of future global shocks.

This project's potential benefits include identifying vulnerable populations and helping decision-makers develop targeted support programs to mitigate negative global shocks on financial well-being.
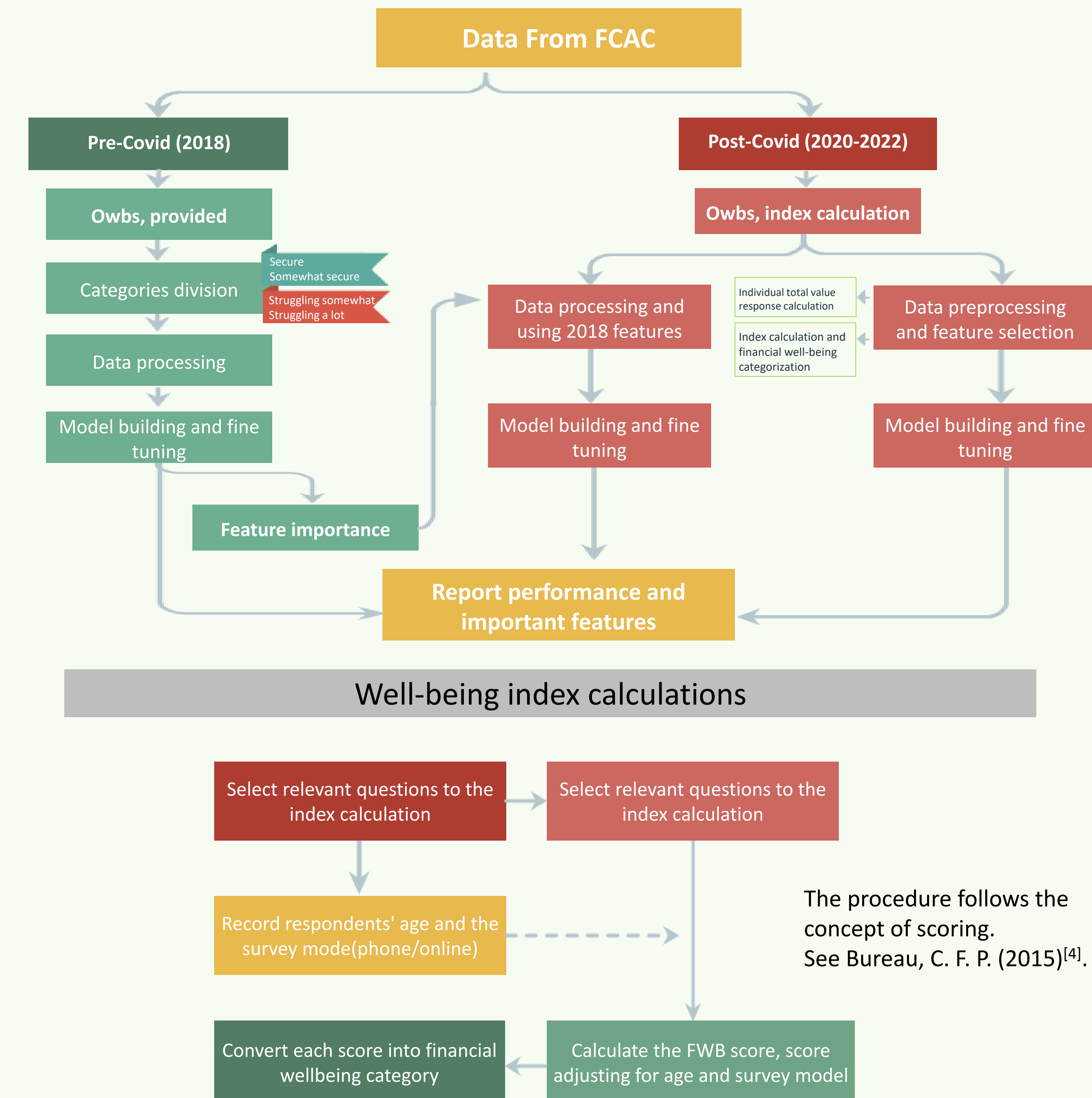
Datasets

Sources
- ☐ The 2020-21 COVID-19 Financial Well-Being Survey [1]
- ☐ The 2018 Financial Well-being and Capability [2]
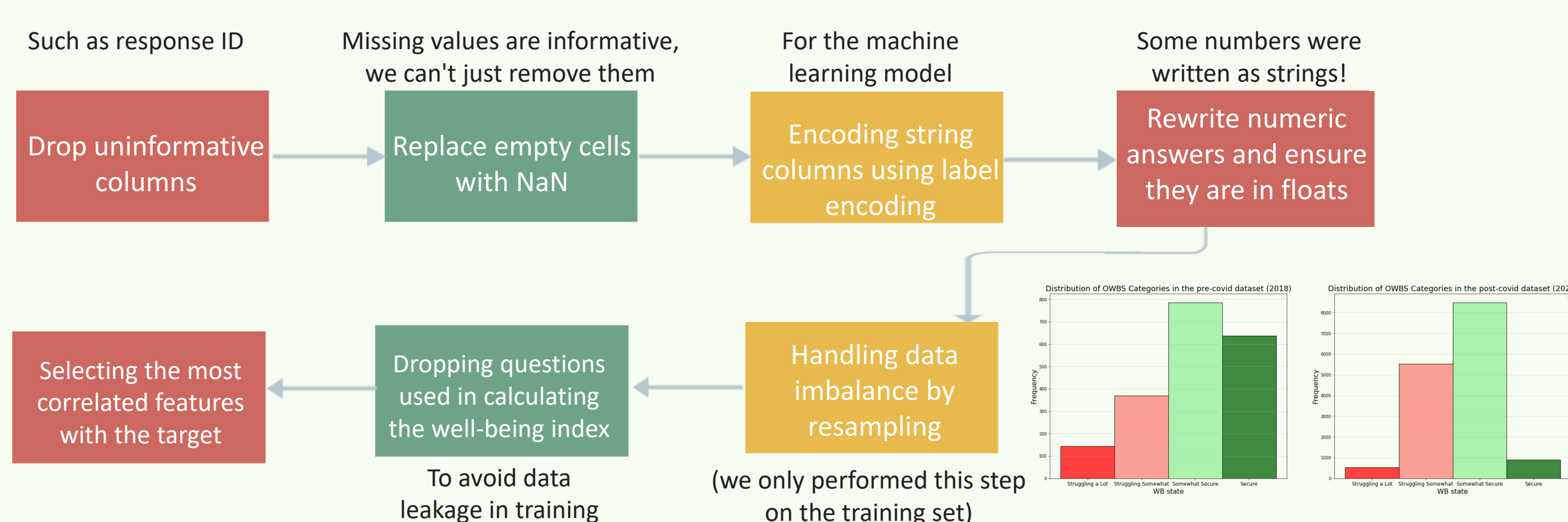
Characteristics
- ☐ The datasets aim to shed light on Canadians' financial decision-making capacities [3]
- ☐ The COVID-19 data : 9,394 interviews targeted toward adult Canadians
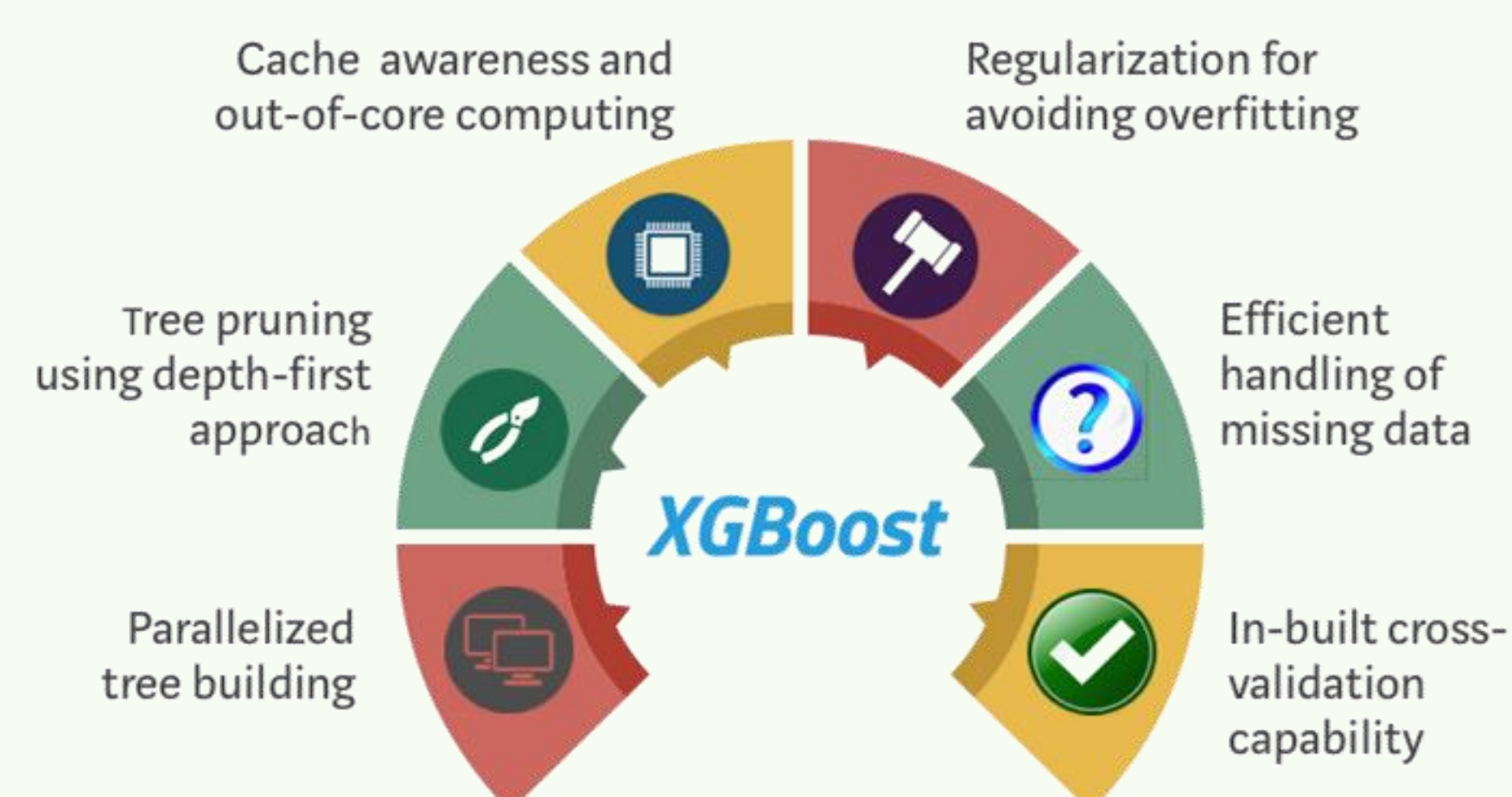- ☐ The 2018 data: 1935 observations

Methodology

Project Flowchart

Data From FCAC

Pre-Covid (2018) → Owbs, provided → Categories division (Secure, Somewhat secure, Struggling somewhat, Struggling a lot) → Data processing → Model building and fine tuning → Feature importance

Post-Covid (2020-2022) → Owbs, index calculation → Data processing and using 2018 features / Individual total value response calculation, Index calculation and financial well-being categorization → Data preprocessing and feature selection → Model building and fine tuning

Report performance and important features

Well-being index calculations

Select relevant questions to the index calculation → Select relevant questions to the index calculation → Record respondents' age and the survey mode(phone/online) → Calculate the FWB score, score adjusting for age and survey mode → Convert each score into financial wellbeing category

The procedure follows the concept of scoring.
See Bureau, C. F. P. (2015).[4].

Preprocessing

Drop uninformative columns (Such as response ID) → Replace empty cells with NaN (Missing values are informative, we can't just remove them) → Encoding string columns using label encoding (For the machine learning model) → Rewrite numeric answers and ensure they are in floats (Some numbers were written as strings!) → Handling data imbalance by resampling (we only performed this step on the training set) → Dropping questions used in calculating the well-being index (To avoid data leakage in training) → Selecting the most correlated features with the target

Model Description

Selected model: XGBoost [5]
Reasons:
- Performs well on tabular data with missing values [6]
- Data contain informative missing values

XGBoost: Cache awareness and out-of-core computing, Regularization for avoiding overfitting, Efficient handling of missing data, In-built cross-validation capability, Parallelized tree building, Tree pruning using depth-first approach

Results

- Three models were developed: pre-Covid-19, post-Covid-19 without feature matching, and post-Covid-19 with feature matching using XGBoost.
- XGBoost algorithm was used to predict the target variable "owbs_category" after preprocessing.
- LabelEncoder method was used to convert owbs_category into numerical labels.
- Feature selection was performed using SelectKBest [7] method with f_regression function for correlation analysis.
- Grid search [7] was used to find the best hyperparameters for each model, focusing on learning_rate, max_depth, n_estimators, min_child_weight, subsample, colsample_bytree, gamma, and scale_pos_weight

Figure 1: performance summarization, where **a.** pre-Covid 19 model where we will use the FCAC 2018 dataset. **b.** post-Covid 19 model with no feature matching. **c.** post-Covid 19 model with feature matching

Table 1: Top 3 questions that helped in predictions for each model
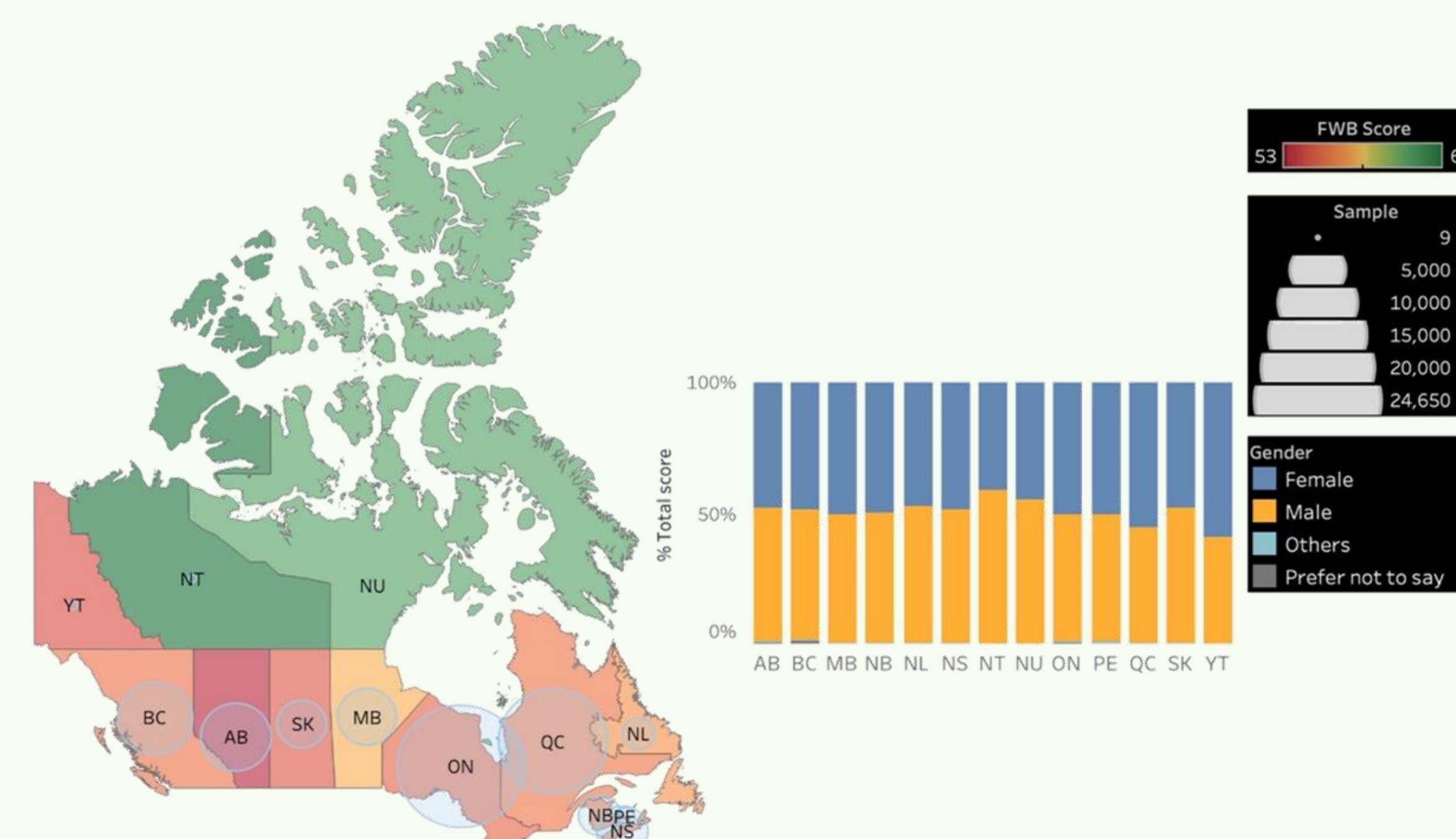
a. pre-Covid 19 model where we will use the FCAC 2018 dataset
Will you be able to meet an unexpected expense that is equivalent to a month's income?
How would you describe your household's/your current financial situation?
How does 'Our/my finances allow me to do the things I want and enjoy life' describe you?

b. post-Covid 19 models with no feature matching
In the past 12 months, have you run short of money and had to use a credit card, overdraft, or borrow to buy food or to pay monthly expenses
How would you describe your bills and other financial commitments?
How would you rate your current credit record?

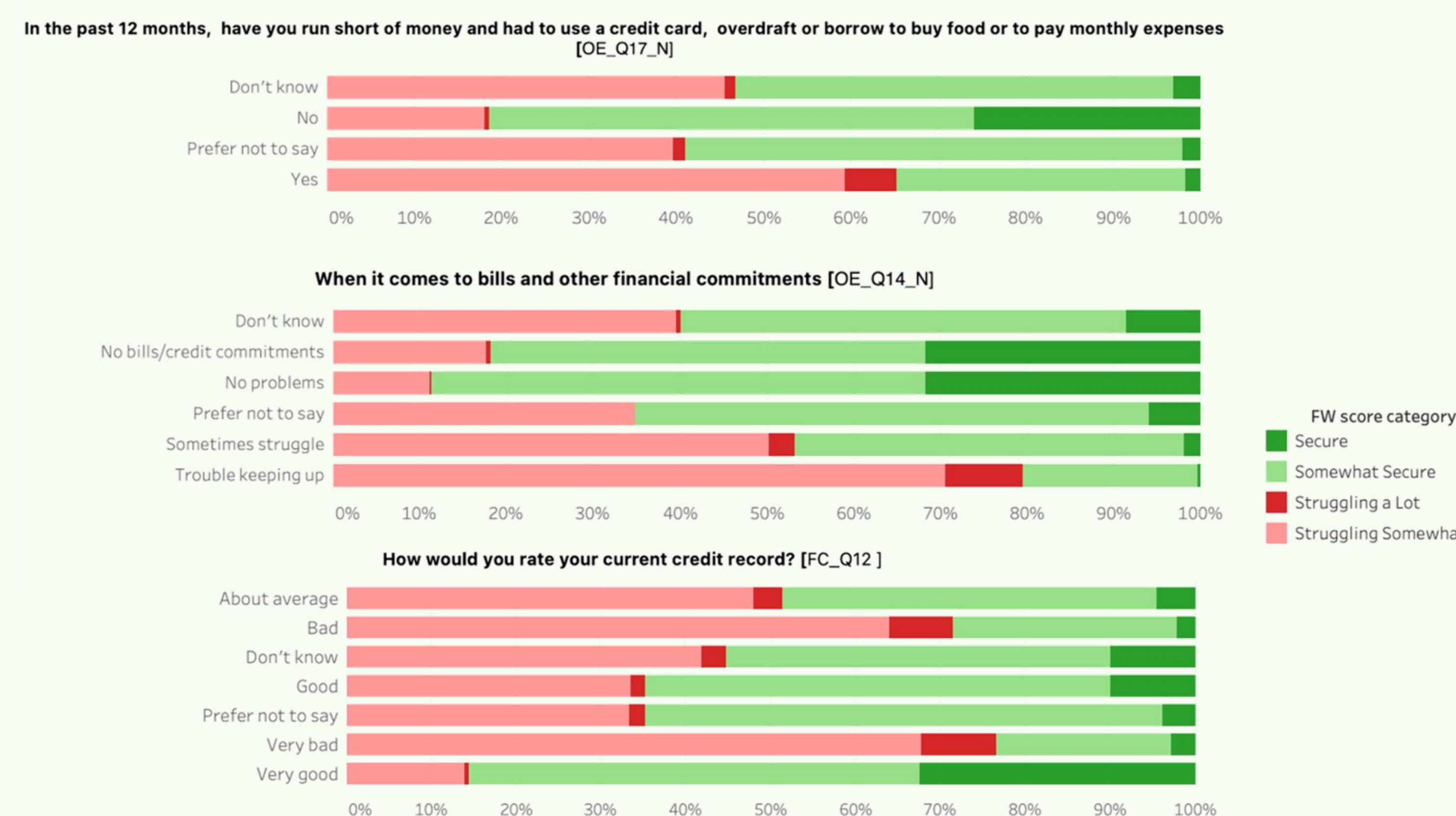| c. post-Covid 19 models with feature matching |
| --- |
| In the past 12 months, have you run short of money and had to use a credit card, overdraft, or borrow to buy food or to pay monthly expenses |
| How would you describe your bills and other financial commitments? |
| Over the past 12 months, have you used any of the following other methods to manage your day-to-day expenses? (Borrow, seek advice, use online lender) |
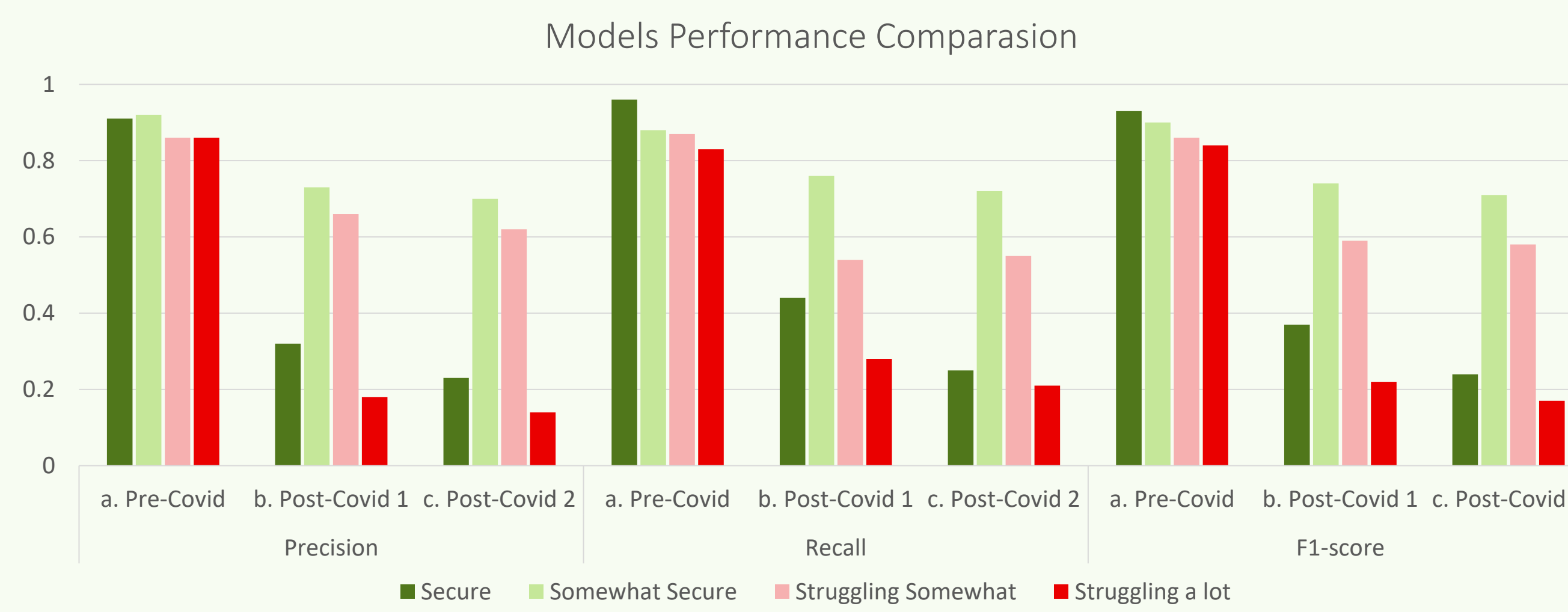
- Pre-covid survey-based model significantly outperformed other models in all metrics, possibly due to different question structures in the post-covid survey.
- Well-being score in 2018 dataset was provided by FCAC and calculated by a third party, making it easier for the model to predict the category.
- When we trained the model on post-covid dataset using the same questions used in index calculation, we achieved an accuracy of over 95%.
- Questions used to calculate well-being scores were dropped, and other factors were explored that could aid the prediction.
- Post-covid models may not perform as well, but they are still realistic and generate useful features.

Discussion

Unequal distribution of financial wellbeing across the country, with the NT exhibiting a higher score (sample size may have influenced the result to some extent).
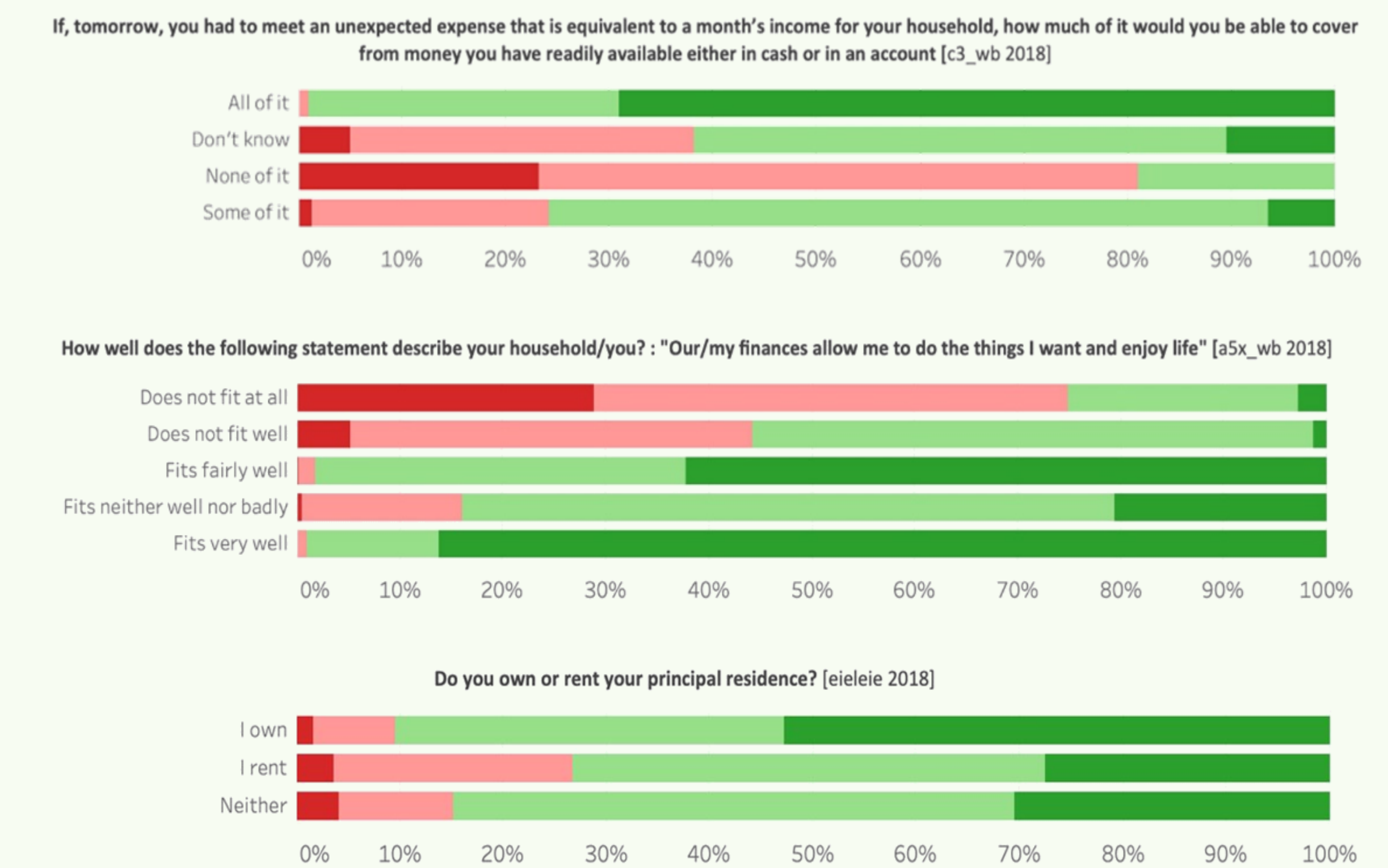

Figure 2 Financial well-being score along the Canadian provinces.

Pre-covid top features:

If, tomorrow, you had to meet an unexpected expense that is equivalent to a month's income for your household, how much of it would you be able to cover from money you have readily available either in cash or in an account [c3_wb 2018]

How well does the following statement describe your household/you? : "Our/my finances allow me to do the things I want and enjoy life" [a5x_wb 2018]

Do you own or rent your principal residence? [eieleie 2018]

- Key determinants before COVID-19: ability to meet expenses, capacity to enjoy life.
- Other important determinant: home ownership status

Post-covid top features:

In the past 12 months, have you run short of money and had to use a credit card, overdraft or borrow to buy food or to pay monthly expenses [OE_Q17_N]

When it comes to bills and other financial commitments [OE_Q14_N]
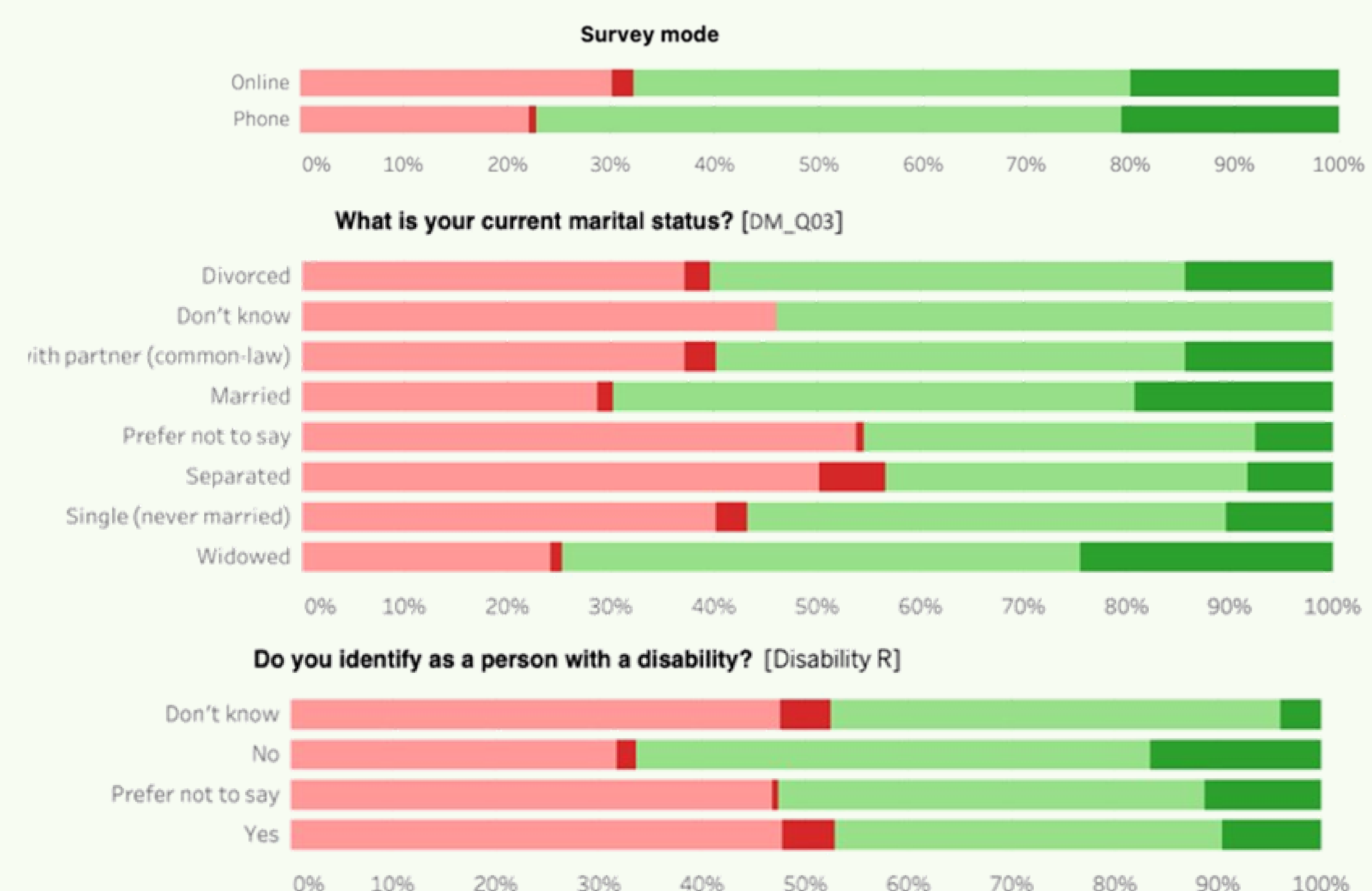
How would you rate your current credit record? [FC_Q12]

- Shift in the determinant of financial wellbeing post-Covid. Key determinant: credit score
- Other important determinants: reliance on credit cards, overdrafts, or borrowing money to meet basic, meeting financial commitments, including bills.

Interesting features that contributed to predictions across models:

Survey mode

What is your current marital status? [DM_Q03]

Do you identify as a person with a disability? [Disability R]

- Unexpected features that contribute to predicting financial security:
  - Mode of the survey (online vs phone), marital status, disability status, being a First Nation, and language learned as a child.

Conclusion

Key results:
- ☐ Covid 19 contributed to lowering Canadian's financial wellbeing
- ☐ Financial wellbeing is disproportionately distributed across the nation
- ☐ Main drivers: financial situation, capacity to meet expenses, saving, credit score.
- ☐ The prediction is more accurate when global shocks are not accounted for.

Policy implications:
- ☐ More support to Canadians across provinces to better cope with their finances.
- ☐ More quality jobs, inflation control, debt relief including mortgage, more assistance to the poor.
- ☐ Particular attention to vulnerable groups: women, persons with disability, indigenous communities.

References

[1] Financial Consumer Agency of Canada. 2018. Government of Canada. https://www.canada.ca/en/financial-consumer-agency/programs/research/ financial-well-being-survey-results.html
[2] Financial Consumer Agency of Canada. 2018. Government of Canada. https://www.canada.ca/en/financial-consumer-agency/programs/research/ financial-well-being-survey-results.html
[3] Leslie-Anne Keown. 2011. The financial knowledge of Canadians. https://www150.statcan.gc.ca/n1/pub/11-008-x/2011001/article/11413-eng.htm
[4] Bureau, C. F. P. (2015). Measuring financial well-being: A guide to using the CFPB financial well-being scale. Washington, DC: Consumer Financial Protection Bureau. https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-scale/
[5] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[6] Ravid Shwartz-Ziv, Amitai Armon, Tabular data: Deep learning is not all you need, Information Fusion, 81-2022-84-90 https://doi.org/10.1016/j.inffus.2021.11.011.
[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.