SCORE-B *your best version of yourself for a better world*

Aline Vicente So, Eng., MBA

Diploma In Data Analytics Co-op, Toronto School of Management, Toronto

Business Problem

Credit score is a measure that reflects lenders' confidence in an individual's ability to repay their debts. Credit score is influenced by various factors, including payment history, debt levels, and credit history. When a person has a low credit score, lenders may consider them to be in a higher risk category and, as a result, may offer higher interest rates or even deny loan and financing approvals.

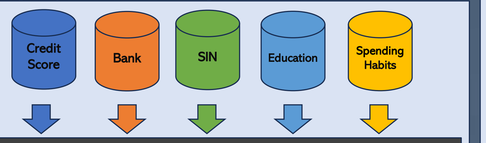
We know that misusing credit can delay or prevent opportunities for individuals to achieve their goals in life, as the money allocated to paying interest could be used for their own benefit, such as acquiring goods or services they may need or simply desire.

On the other hand, lack of access to credit increases the likelihood of individuals losing control over their personal finances, leading to an unsustainable debt burden. In a social context, this puts them in a vicious cycle of dependence on government financial assistance programs.

In Canada, many citizens are facing real challenges of affordability and experiencing the effects of rising food prices and housing costs, leaving them feeling "strapped for cash".

Such unfavorable conditions corner ordinary citizens and make them feel threatened. Withdrawn, they may drastically change their consumption patterns and, again, indirectly, and involuntarily contribute to the weakening of the economic ecosystem. Given the presented context, the business question is: ***What is the impact of an individual's consumption habits on their credit score?***

1. Business Understanding

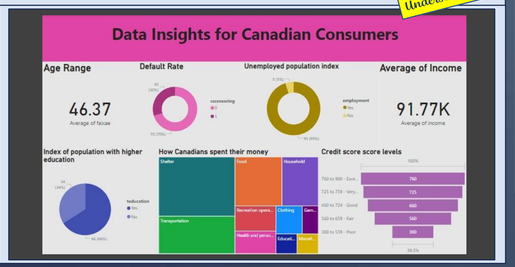


2. Data Understanding

```
#------------------------------------------------- getting the data
cscore <- read.csv("C:\\Users\\cscore\\cscore.csv");
csbank <- read.csv("C:\\Users\\bank\\bank.csv");
cssin <- read.csv("C:\\Users\\sin\\sin.csv");
cseducation <- read.csv("C:\\Users\\education\\education.csv");
csshabits <- read.csv("C:\\Users\\consuption\\consuption.csv");
#------------------------------------------------- knowing the data
str(cscore);
str(csbank);
str(cssin);
str(cseducation);
str(csshabits);
#------------------------------------- perfoming descriptive analysis
summary(cscore);
summary(csbank);
summary(cssin);
summary(cseducation);
summary(csshabits);
#------------------------------------------------- data integration
csintegrated_DF <- data.frame(cscore$cscore,
csbank$accounttype,
sin$sin,
sin$income,
cseducation$educationlevel,
csshabits$smiscellaneous,
csshabits$seducation,
csshabits$sgames,
csshabits$sclothing,
csshabits$shealth,
csshabits$shousehold,
csshabits$sfood,
csshabits$stransportation,
csshabits$sshelter);
```

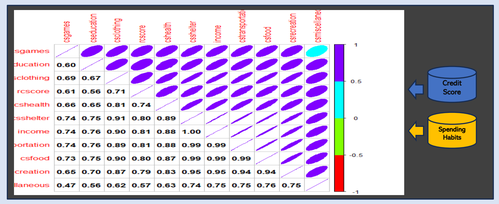
3. Data Preparation

CS Integrated



stack

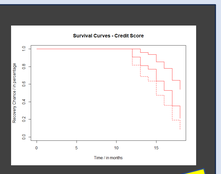
Performing Correlation Coefficient Analysis



Performing Survival Analysis

Insight: Identifying factors that influence the time it takes for a person to improve their credit score

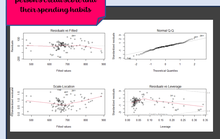
```
> summary(sobrevivecs);
Call: survfit(formula = Surv(recoveryt, csevent) ~ 1, data =
csdefaulters,
    type = "kaplan-meier")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     30       1    0.967  0.0328        0.905        1.000
   13     28       5    0.794  0.0750        0.660        0.955
   14     20       2    0.715  0.0860        0.565        0.905
   15     17       2    0.631  0.0942        0.471        0.845
   16     12       1    0.578  0.0999        0.412        0.811
   17      9       2    0.450  0.1116        0.276        0.731
   18      5       2    0.270  0.1191        0.114        0.641
```



4. Modeling

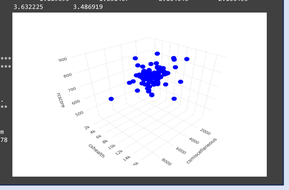
Performing Multiple Linear Regression Analysis

Insight: Examine the relationship between a person's credit score and their spending habits



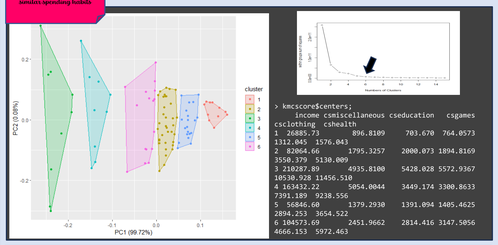

```
> summary(testcscore);
Call:
lm(formula = rcscore ~ employment + csmiscellaneous +
cseducation + csgames + csclothing + cshealth, data =
cscore)
Residuals:
      Min       1Q   Median       3Q      Max
-162.786  -26.829    2.977   31.064  150.986
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.548e+02  2.472e+01  18.395  < 2e-16 ***
employmentYes  1.166e+02  2.668e+01   4.372 3.21e-05 ***
csmiscellaneous 5.702e-03  4.324e-03   1.319  0.19046
cseducation    5.897e-04  4.779e-03   0.123  0.90206
csgames        7.203e-03  4.700e-03   1.532  0.12880
csclothing     6.857e-03  3.843e-03   1.784  0.07763 .
cshealth       1.001e-02  3.408e-03   2.937  0.00418 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 54.95 on 93 degrees of freedom
Multiple R-squared:  0.6691,  Adjusted R-squared:  0.6478
F-statistic: 31.34 on 6 and 93 DF,  p-value: < 2.2e-16
```

```
> testcscore <- lm(formula = rcscore ~ employment + csmiscellaneous +
cseducation + csgames + csclothing + cshealth, data = cscore);
> shapiro.test(testcscore$residuals);  ### testing the normality of
the residuals
Shapiro-Wilk normality test
data:  testcscore$residuals
W = 0.9837, p-value = 0.2551
> summary(rstandard(testcscore));      ### testing the outliers in
the residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-3.005688 -0.459576  0.056770 -0.004384  0.578287  2.826289
> bptest(testcscore);                  ### testing homoscedasticity
by Breusch-Pagan
studentized Breusch-Pagan test
data:  testcscore
BP = 4.3447, df = 6, p-value = 0.6301
> vif(testcscore);                     ### testing multicollinearity
     employment csmiscellaneous     cseducation         csgames
csclothing         cshealth
       1.119695       1.831407       2.104640       2.106455
3.632225       3.486919
```

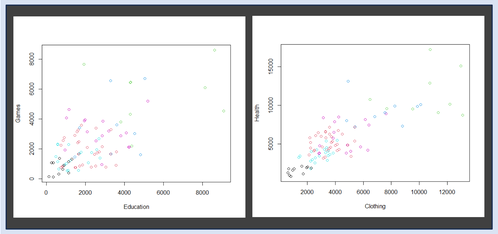


Performing Clustering Analysis

Insight: Identifying groups of consumers who have similar spending habits




```
> kmcscore$centers;
     income csmiscellaneous cseducation    csgames
csclothing  cshealth
1  26885.73         896.8109      703.670  764.0573
1312.045  1576.043
2  82064.66        1795.3257     2000.073 1894.8169
3550.379  5130.009
3 210327.89        4935.8100     5428.028 5572.9367
10530.928 11456.510
4 163432.22        5054.0044     3449.174 3300.8633
7391.189  9238.556
5  56846.60        1379.2930     1391.094 1405.4625
2894.253  3654.522
6 184573.69        2451.9662     2814.416 3147.5056
4666.153  5972.463
```



Acknowledgements

I would like to express my gratitude to ToSM for the opportunity to showcase my work. I extend my special thanks to **Gabriel Quiroz**, **Maddy Nanayakkara**, **Nikita Batta**, and everyone in the **Events** team for their support and assistance throughout my entire journey.

Final Report

The studies presented here were conducted using the **CRISP-DM** (Cross-Industry Standard Process for Data Mining) methodology and driven by the business question of how Canadian citizens' spending habits influence their credit score.

To achieve this, we initiated the project with a strong focus on **business understanding**, conducting an evaluation of the situation and producing a project plan in collaboration with the business area.

For the **data understanding** phase, we mined data from the original databases of Credit Score, Bank, SIN, Education, and Spending Habits. We explored business data views and compared the involved structures to ensure data quality. With data quality assured, we proceeded with the initial analysis as demonstrated in ***Analytical Approach 1***.

For the **data preparation** phase, we selected and shaped the data according to the project's initial needs – csintegrated database.

During the **modeling** phase, we obtained new insights from the business area to further unfold the initial business question and, based on the new inquiries, selected techniques to build models for the insights in collaboration with the business area – as demonstrated in ***Analytical Approach 2, 3, 4, and 5***.

During the results **evaluation** phase, we reviewed the process and determined the next steps, which involve the development project of the SCORE-B MVP.

The **implementation** phase is dependent on the approval of the actions taken in the previous phase.

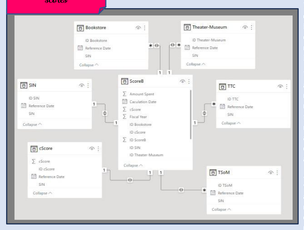
6. Deployment

Business Problem	What is the impact of a citizen's spending habits on their credit score?
Analytical Approach 1	Coefficient Correlation
Findings	
According to the correlation matrix using the Pearson coefficient, it was possible to observe strong and positive associations between the variable "rcscore" and the variables representing spending habits, except for the variables "cshealth", "csrecreation", "csmiscellaneous", "cseducation", and "csgames", which have moderate and positive associations. **We also examined the variable "income" in the same matrix, which conceptually has a direct impact on the formation of the credit score and found a strong and positive association with all the variables representing spending habits, except for the variable "csgames", which has a moderate and positive association.** **However, it is important to note that correlation does not imply causality. Although the study shows strong and moderate relationships between the variables, it does not provide evidence of a cause-and-effect relationship. Therefore, further analysis and investigation are necessary to uncover new insights and understand the underlying factors that contribute to this correlation.**	
Insight 1	Identifying factors that influence the time it takes for a person to improve their credit score
Analytical Approach 2	Survival Analysis
Findings	
According to the Kaplan-Meier survival rate calculation, at 12 months the survival rate is 0.967, indicating that the probability of an individual not having recovered from the debt is approximately 96.7%, meaning that there is a 3.3% chance (100% - 96.7%) that the individual has already recovered within this period. **As time goes on, the survival rate decreases, indicating a higher probability of debt recovery. When observing the 18th month, the survival rate is 0.270, indicating that the probability of an individual not having recovered from the debt is approximately 27%. This means that there is a 73% chance (100% - 27%) that the individual has already recovered within this period.**	
Insight 2	Examine the relationship between a person's credit score and their spending habits
Analytical Approach 3	Multiple Linear Regression
Findings	
The results of the multiple linear regression analysis showed that the credit score is significantly associated with some of the considered variables as follows: **() "cshealth" showed a significant positive association with the credit score, with a p-value of 0.00418. Increases in healthcare expenses are associated with an increase in the customers' credit score.** **() "csclothing" and "csgames": These two variables also exhibited a significant positive relationship, with p-values of 0.07763 and 0.12880, respectively, with the credit score. Increases in clothing and games of chance expenses are associated with an increase in the credit score, although the effect size is not large, it is still statistically significant.** **() "csmiscellaneous" and "cseducation": These variables showed a positive association with the credit score, but the level of statistical significance was not as high, as indicated by the higher p-values of 0.19046 and 0.90206, respectively. This suggests that increases in miscellaneous items and education expenses may have a modest effect on the customers' credit score.** **() "employment": This variable indicates whether the customer is employed or not. The presence of this factor had a significant impact on the credit score, with a low p-value of 3.21e-05. Employed customers generally had a higher credit score compared to those who are unemployed.** **Overall, the linear regression model showed a good fit to the data, explaining approximately 63.84% of the variation in the credit score. Assumption tests also indicated that the model residuals follow a normal distribution and there is no evidence of significant heteroscedasticity issues.** **However, it is important to note that this analysis is based on observed correlation and does not imply direct causality. Other factors not included in the model may influence the customers' credit score. Therefore, a more in-depth analysis and consideration of other relevant factors are recommended for a comprehensive understanding of the relationship between consumption habits and credit score.**	
Insight 3	Identifying groups of consumers who have similar spending habits
Analytical Approach 4	Clustering Analysis
Findings	
According to the results of the analysis conducted using the clustering technique for grouping and the Elbow Method for determining the number of clusters, we can observe the following patterns related to income and spending patterns for the identified groups: **() Overall, the 6 groups allocate equal shares of their income to education and games of chance, averaging around 2.5% of their income for each category. This means that for Canadians, investing in education has the same level of importance as spending on games of chance. Interestingly, the group with the highest income among the 6 groups invests the least in education, around 2.11% of their income.** **() Similarly, for the 6 groups, the share of income Canadians allocate to miscellaneous items is around 2.6% on average for this category.** **() In terms of overall averages for the 6 groups, clothing and healthcare are the categories that carry more weight. Across the 6 groups, clothing represents approximately 4.72% of their income, while healthcare represents 5.89%. However, it is noteworthy that the percentages for the Canadian middle-class income groups are higher in health expenditures, around 6.18%.** **It is important to note that these groups were formed based on observed characteristics and do not imply a direct cause-and-effect relationship. Cluster analysis is a powerful tool for identifying patterns and segmenting customers, but further analysis is needed to understand the underlying relationships between variables and the formed groups.**	
Insight 4	Creating an MVP to improve credit scores
Analytical Approach 5	Database Modeling

5. Evaluation

Insight: Creating an MVP to improve credit scores

Performing Data Model – SCORE-B



The SCORE-B MVP aims to accumulate points based on variables related to consumption categories and investments in education and incorporate them into the credit score calculation to reward or mitigate credit restrictions through the citizen's SIN (Social Insurance Number).

Thus, from the simple purchase of a book or a ticket to a cultural attraction, to the payment of a public transportation card for students, and even investments in training or professional development courses, whether for oneself or as a sponsor, in both B2C and B2B models, SCORE-B will have a database that accumulates points over a fiscal period based on the citizen's commercial transactions at establishments accredited by SCORE-B.

At the end of the fiscal period, the accumulated points will be grouped into categorical ranges with minimum and maximum score limits. The points within each range will be directly applied to the credit score calculation through the SIN, benefiting the citizen themselves.

Purpose

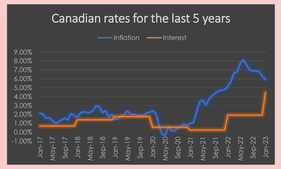
To be a financial industry product that adds value to society, based on the awareness of the best use of credit and the strengthening of the system through the principle of education

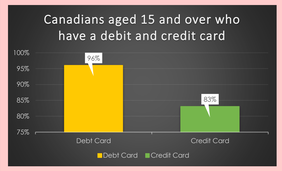
Objective

Accumulate points based on categories of consumption and investment in education and increment them to the credit score calculation to reward or mitigate credit restrictions by the citizen's SIN

Justification

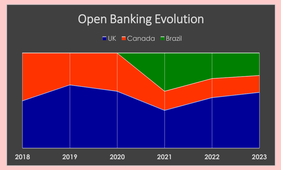
Sharp rise in inflation and interest rates since the post-pandemic

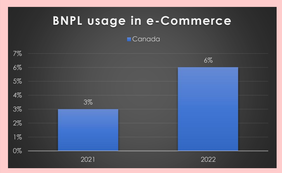




High rate of credit card use

Slowness in the implementation of Open Bank impacts on credit restrictions





"It has always been easy with credit cards, but BNPE gives the impression that it is almost free money in a sense, which is not the case."

Benefits

- Increased access to credit
- Promotion of financial education
- Valuing the entire educational ecosystem
- Economic strengthening for the benefit of society itself

Stakeholders

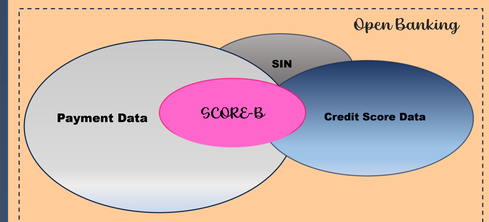
.Financial sector .Education and cultural commerce .NGOs and educational institutions .Government and civil society

References

- Bank of Canada. (2023). Canadian interest rates and monetary policy variables: 10-years lookup. https://www.bankofcanada.ca/rates/interest-rates/canadian-interest-rates/
- WANA. (2023). Canada Inflation Rate and CPI. https://www.ca/inflation-rate-canada-cpi
- The Global Economy. (2021). Canada: Percent people with debit and credit cards. https://www.theglobaleconomy.com/Canada/people_with_debit_cards/
- Government of Canada . (2023). Open banking. https://www.canada.ca/en/financial-consumer-agency/services/banking/open-banking.html
- Giovanna Gallo. (2022). Open Finance: compare como o sistema evolui nas 5 praias que estão nas quentas do Copa do Catar. Inja Money. https://www.injamoney.com.br/minhas-finanzas/open-finance-compare-como-o-sistema-evolui-nas-5-praias-que-estao-nas-quentas-da-copa-do-catar/
- Government of Canada. (2023). Buy now, pay later plans. https://www.canada.ca/en/financial-consumer-agency/services/loans/buy-now-pay-later.html
- Aliza Sackeroka. (2023). What you need to know about the rise of Buy Now, Pay Later in Canada. Yahoo! Finance. https://ca.finance.yahoo.com/news/rise-buy-now-pay-later-canada
- Government of Canada. (2023). Economic Overview. https://www.budget.canada.ca/2023/report-rapport/overview-apercu-en.html
- Pamela Heaton. (2023). Poshtale: One in six Canadians say they are likely to default on a major loan as rate hikes pinch. Financial Post. https://financialpost.com/executive/executive-summary/debt-default-rate-interest-rates
- Statista Canada. (2019). How Canadians spend their money in 2019. https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2021006-eng.htm
- de Vries, Andrie & Meys, Joris. R for Dummies. New Jersey: John Wiley & Sons, Inc, 2015.
- Schmuller, PhD, Joseph. Statistical Analysis with R for Dummies. New Jersey: John Wiley & Sons, Inc, 2017.
- Teng, J.D., Tetas, Paul. R Cookbook. Sebastopol: O'Reilly, 2019.
- Wickham, Hadley & Grolemund, Garret. R for Data Science. Sebastopol: O'Reilly, 2016.
- Wheelan, Charles. Naked Statistics. New York: W.W. Norton & Company, 2013
- Jamovi, version 2.3.25. https://dl.cloud.jamovi.org/
- Wankes Leandro Ribeiro & Helbert Ricardo Vieira. (2020). Canvas de Projeto. Como transformar ideias em projetos. https://www.canvasdeprojeto.com.br/

Characteristics

- ✓ 100% data-driven
- ✓ B2B and B2C model
- ✓ Aligned, yet independent of Open Banking implementation



Assumptions & Constraints

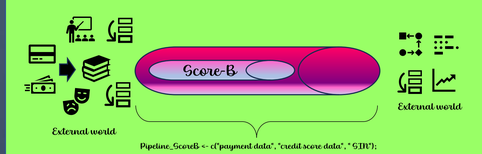
- The product must comply with PIPEDA regulations
- The product requires a data analysis intelligence environment in line with the CRISP-DM methodology
- The product needs to read the credit score database
- The product needs to be consumed by the credit score calculation system
- The product should be promoted and made available to society as a financial education incentive and tool

Risks

- Market rejection: Lack of recognition of the added value to society
- Regulation: Non-permission or delays in implementing coverage laws
- Operational risk: Errors or failures in the development and implementation process
- Financial risk: Loss of capital invested in MVP

Scope

- ❑ Pipeline Score-B:



Out of Scope

- application and maintenance of credit score

Deliverable

- ❑ Score-B generated and made available for application and maintenance of the credit score
- ❑ Insights engine for business decision making

Investment

The complete implementation and maintenance of the business chain: research and development / technology for product development and evolution

Dates: to evaluate

Data Analysis by CRISP-DM

SCORE-B - Canvas Project Design