CIND 820

Determinants of Financial Well-Being

Capstone Project Course
Ryerson University

Literature Review and Data Description

By

Zeynep B Cevik
501138112

Supervised By

Uzair Ahmad, Ph.D.

October 18, 2021

1. Literature Review

In the literature, there are various studies which define the Financial Well Being (FWB) from different perspectives. These different perspectives arise from its direct or indirect effects on different disciplines such as consumer behavior, overall individual well-being, economic growth etc. Although there is no consensus about the definition of a FWB, a formal definition, done by US Consumer Financial Protection Bureau (CFPB), is 'FWB is a state of being wherein a person can fully meet current and ongoing financial obligations, can feel secure in their financial future and is able to make choices that allow them to enjoy life.' (CFPB, 2017)

As the definition states, FWB is the individual's perception about his/her financial capacity/ limit, and ability to save and spend. This perception does not only depend on the financial environment but also the individual's own characteristics such as socioeconomic factors, financial knowledge, etc. In the literature, some empirical studies[1] found that the financial knowledge and their financial outcomes have a positive association because financially educated individuals make more rational decisions and have capacity to develop their financial earnings. Clark and D'Ambrosio (2008) stated that not enough financial knowledge may be an obstacle for reaching to the financial markets. Lee et. al. (2020) also examined the relationship between financial knowledge and FWB by using CFPB's 2016 National FWB Survey. This survey includes a FWB scale score, calculated from the answers of 10 questions and two financial knowledge scores: prepared from the answers of 3 objective questions and of 7 subjective questions. The bigger the scores are, the better the FWB and financial knowledge are. In addition to these variables, in this survey there are some attributes that give information about the respondents' socioeconomic structure such as age, gender, race, education level, etc. The dependent variable was FWB score, whereas financial knowledge variables and propensity to plan variable, which was constructed with Principal Component Analysis (PCA) by the authors, were independent variables. The other variables were defined as control variables. They used ordinary least squares (OLS) regression model and formed 3 regression models. These regressions differentiated in terms of including objective financial knowledge only, objective financial knowledge and propensity to plan and in addition to these two, their interaction term. Control variables and subjective financial knowledge score were included in all these three regressions. The results demonstrated that there is a positive significant relationship between financial knowledge and FWB, i.e., when the individual is financially more educated, the FWB score is higher. In addition, the propensity to plan variable is also statistically significant in improving FWB score.

CFPB (2017) prepared a technical report about the 2016 National FWB Survey. They examined the survey by dividing the attributes into six categories other than FWB's items: (1) individual characteristics; (2) household and family characteristics; (3) income and employment characteristics; (4) savings and safety nets; (5) financial experiences; and (6) financial behaviors, skills, and attitudes. They tried to find out the distribution patterns of FWB in these categories. Without making any kind of regression analysis, they only performed descriptive statistics for each category. They found out many valuable results. For instance, the results demonstrated that the higher education and age leads to higher FWB. In addition, married couples and employers provided any kind of benefits showed a higher FWB.

[1] Hilgert and Hogarth, 2003; Joo and Grable, 2004; Zulfiqar and Bilal, 2016

Joo and Grable (2004) focused on the determinants of financial satisfaction by using a random sample of white-collar clerical workers in United States. They defined the financial satisfaction, similar to FWB, so that in FWB literature this study was cited in most of the studies. As in other studies, their motivation lied on the effects of financial satisfaction onto individual's overall welfare. They calculated financial satisfaction score from a one-item 10-point stair-step question. They applied path analysis, which is a method for observing the logical and theoretical relationships between variables based on correlation analyses, by using SPSS. Since path analysis approach depends on multivariate regression, the authors categorized the dependent and independent variables. Dependent variable is financial satisfaction variable, whereas independent variables are demographic and socioeconomic characteristics, financial knowledge, financial stressors, financial stress level, financial risk tolerance, financial behaviors, and solvency level. They found out the financial literacy (i.e., financial knowledge) was one of the key variables to increase financial satisfaction. Aside from financial literacy, risk tolerance, financial behaviors and stress level are all statistically significant.

Iramani and Lutfi (2021) investigated the role of financial behavior on FWB by constructing an integrated family financial welfare model for Indonesia. They carried out an online survey in 20202 and there were more than 1000 respondents. Their exogenous variables were financial knowledge, financial experience, financial status, and internal locus of control. They used Likert scale model to create a FWB variable by using financial stress, financial satisfaction, financial comfort, financial worries, and financial confidence as in the studies of Archuleta et. al. (2013) and Ng and Diener (2014). They run three different regression models. The first model analyzes direct effects of all the exogenous variables. In the second model, financial behavior was introduced as a mediator variable. In the last model, marital status was used as a moderating variable. As consistent with the literature, they also found that the financial knowledge, financial status, and marital status directly affected FWB for Indonesia.

Fan and Henager (2021) studied not only direct and indirect determinants but also short and long-term determinants of FWB by using 2018 National Financial Capability Study. They applied Structural Equation Modelling (SEM) with Maximum Likelihood Estimation (ML). These methods are mostly used to identify the structural relationships simultaneously in a conceptional framework. They found out that short-term and long-term financial behavior affected the FWB in opposite directions. Short-term financial behavior and perceived financial capability positive affected the FWB, whereas long-term financial behavior and long-term financial behavior affected adversely.

The above-mentioned studies did not focus on specific age group or education group, however Gutter and Copur (2011) investigated the relationship between financial behaviors and FWB among college students in United States. They focused on three main questions: (1) what is the average level of FWB among college students? (2) How is it characterized in terms of demographic, financial characteristics, and financial behaviors? (3) If some demographic and financial characteristics are controlled, will the financial behaviors still affect the FWB significantly? They applied bivariate analyses and used t-tests, one-way analysis of variance (ANOVA), and Pearson correlation. They also applied OLS regression analysis. The results showed that the FWB was significantly affected by the various socioeconomic factors and financial behaviors. In addition, when demographic, financial characteristics, and financial behaviors were

controlled, some financial behaviors such as budgeting, saving, risky credit card behaviors, significantly influenced FWB.

Iannello et. al. (2021) also focused on a specific age group (20-27 years old) by using Italian and Portuguese participants. They applied a cluster analysis to classify FWB of these participants in terms of subjective and psychological factors such as uncertainty and ambiguity tolerance. They detected four clusters: (1) anxiety about ambiguity and uncertainty, (2) comfort with ambiguity and uncertainty, (3) black and- white thinking, (4) flexible thinking. They also performed a multi-group path analysis to determine individual FWB's effect onto individual overall well-being.

In short, on the one side of the FWB literature, the authors try to measure and quantify FWB, on the other side, there are some studies that tried to understand its determinants so that policy makers or individuals will take actions according to these factors in order to increase the social welfare of a society. This study also investigates significant factors that affect the FWB. which is already been calculated by CFPB. The main contribution of this study will be implementing machine learning tools to this dataset in order to find out the determinants of individual's financial well-being. As far as I know, there is no other study that uses machine learning tools

2. Data Description

In this study, 2016 National FWB Survey is used. There are 217 attributes and 6402 observations. 4 different scores were calculated: Financial Well-Being Score, Financial Skill Scale Score, Knoll and Houts Financial Knowledge Scale Score, Lusardi and Mitchell Financial Knowledge Scale Score.

In measurement of FWB, the literature is divided into three: objective scaling approach, subjective scaling approach and half-subjective half-objective scaling approach. In this survey, the FWB score was calculated from the answers of 10 subjective questions, so it can be said that subjective scaling approach was used. There were four main ideas in constructing this score: (1) management of daily and monthly finances, (2) ability to stay strong towards shocks, (3) on the way to meet the goals, (4) the ability to spend for enjoyment and to save. (Consumer Financial Protection Bureau, 2017) The higher the score is, the better the individual's financial well-being is. Likert-type scaling from 0 to 100 is applied.

There are two financial knowledge scores, differentiated in terms of subjectivity and objectivity. The objective financial knowledge variable was constructed from the answers of 3 objective questions, testing the participant's actual knowledge about the finance. It is also known as 3-item Lusardi and Mitchell Financial Knowledge Scaling. It is a categorical variable, can only take 0, 1, 2, 3 values. The subjective financial knowledge score was constructed from the answers of 7 subjective questions about the participant's perception about his/her financial knowledge. It is known as Knoll and Houts Financial Knowledge Scaling. It is a numeric variable.

Financial Skill Score is calculated from the answers of 7 subjective questions, focusing on the participant's ability to make financial decisions from his/her perspective. It is between 0 to 100. However, none of the participants got zero or hundred from FS score.

In this study, I did not use all 217 attributes since most of them were only survey questions. My aim is not to construct any measurement but use already calculated scores and other exogenous variables. All the other independent variables are categorical. For instance, CFPB did not prefer the participant's age directly, but asked which category is his/her age in 18-24, 25-34, 35-44, 45-54, 55-61, 62-69, 70-74, 75+. Since all the independent variables are similar to this idea, there are no numerical independent variable in this dataset for this study[2].

The independent variables are Financial Skill Score, Lusardi and Mitchell Financial Knowledge Scale Score, Knoll and Houts financial knowledge scale score, Age, Gender, Race/Ethnicity, Generation, Employment Status, Marital Status, Household Size, Education Level, Income, Military Status, MSA Status, Census Region, Census Division, Poverty Status, Presence of a Child between ages 0 and 1, Presence of a Child between ages 2 and 5, Presence of a Child between ages 6 and 12, Presence of a Child between ages 13 and 17, Presence of a Child whose age is above 18. (22 independent variables)

3. Explanatory Data Analysis

3.1.Key Observations

- There are 20 categorical variables, including the Lusardi and Mitchell Financial Knowledge Scale Score and 3 quantitative variables: one continuous (Knoll and Houts financial knowledge scale score) and two discrete variables (Financial Skill Scale Score, Financial Well-Being Score).
- There are 6402 participants, 163 of them refused to respond to some questions. This makes the dataset composed of 6239 observations. There are no N/A values.
- Dependent variable is Financial Well-Being Score, the others are independent variables.
- The average of financial well-being level is 56, however there is a slight difference between males and females. Male financial well-being is slightly higher than the female financial well-being
- The mode of FWB score is 49. The most frequent FWB scores are in between 48 and 65.
- Knoll and Houts financial knowledge scale score has left-skewed distribution (because abs (Mean) < abs (Median) and from its histogram) whereas Financial Well-Being Score and Financial Skill Scale have very close to symmetric distributions
- 'Generation' and 'Age' variables are highly correlated (above 0.95) and 'Census Region' and 'Census Division' variables are highly correlated (above 0.95)
- As the household size, if married, and the ages of the children in that household increase, the financial well-being decreases. (Negative correlation).
- As the individuals become more educated, the levels of financial well-being, financial skill, and financial knowledge increase. (Boxplots)

[2] The detailed explanation of an attribute for each category can be found in the following public use file codebook : https://files.consumerfinance.gov/f/documents/cfpb_nfwbs-puf-codebook.pdf

- The married individuals have the highest financial well-being score on average. (Boxplots)
- The black, non-hispanic individuals have the lowest financial well-being scores, whereas white, non-hispanic individuals have the highest financial well-being scores on average. The reason may be because that there is a huge difference between the income of black, non-hispanic individuals and white, non-hispanic individuals on average. (Boxplots)
- The more the older individual is, the higher the financial well-being is. (Boxplots)
- The highest financial well-being belongs to the 'retired' category in employment variable, on the other hand, permanently sick and unemployed individuals have the lowest financial well-being score on average. (Boxplots)
- There is no significant difference among regions in terms of financial well-being levels, financial knowledge scores and financial skill scores.
- In almost all variables, there are outliers.
- R-programming language was used throughout explanatory data analysis.

3.2.Descriptive Statistics

The minimum, maximum, mean values as well as first and third quartiles of all attributes are given in the below table (see Output 1.)

Output 1. Descriptive Statistics

```
      fpl            FWBscore        FSscore         LMscore        KHscore              EMPLOY       Military_Status
 Min.   :1.000   Min.   :-4.00   Min.   :-1.00   Min.   :0.00   Min.   :-2.05300   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000   1st Qu.:48.00   1st Qu.:42.00   1st Qu.:2.00   1st Qu.:-0.57000   1st Qu.:2.000   1st Qu.:5.000
 Median :3.000   Median :56.00   Median :50.00   Median :3.00   Median :-0.18800   Median :3.000   Median :5.000
 Mean   :2.663   Mean   :56.12   Mean   :50.77   Mean   :2.52   Mean   :-0.04322   Mean   :4.336   Mean   :4.528
 3rd Qu.:3.000   3rd Qu.:65.00   3rd Qu.:57.00   3rd Qu.:3.00   3rd Qu.: 0.71200   3rd Qu.:8.000   3rd Qu.:5.000
 Max.   :3.000   Max.   :95.00   Max.   :85.00   Max.   :3.00   Max.   : 1.26700   Max.   :8.000   Max.   :5.000
     agecat        generation        PPEDUC          PPETHM         PPGENDER        PPHHSIZE         PPINCIMP
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:3.000
 Median :4.000   Median :2.000   Median :3.000   Median :1.000   Median :1.000   Median :2.000   Median :6.000
 Mean   :4.463   Mean   :2.545   Mean   :3.165   Mean   :1.612   Mean   :1.476   Mean   :2.517   Mean   :5.527
 3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:8.000
 Max.   :8.000   Max.   :4.000   Max.   :5.000   Max.   :4.000   Max.   :2.000   Max.   :5.000   Max.   :9.000
    PPMARIT         PPMSACAT         PPREG4          PPREG9          PPT01            PPT25             PPT612
 Min.   :1.00   Min.   :0.0000   Min.   :1.000   Min.   :1.000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
 1st Qu.:1.00   1st Qu.:1.0000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
 Median :1.00   Median :1.0000   Median :3.000   Median :5.000   Median :0.0000   Median :0.00000   Median :0.0000
 Mean   :2.04   Mean   :0.8654   Mean   :2.644   Mean   :5.141   Mean   :0.0359   Mean   :0.07774   Mean   :0.1287
 3rd Qu.:3.00   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:7.000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
 Max.   :5.00   Max.   :1.0000   Max.   :4.000   Max.   :9.000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
    PPT1317          PPT18OV
 Min.   :0.0000   Min.   :1.000
 1st Qu.:0.0000   1st Qu.:2.000
 Median :0.0000   Median :2.000
 Mean   :0.1199   Mean   :2.083
 3rd Qu.:0.0000   3rd Qu.:2.000
 Max.   :1.0000   Max.   :4.000
```

The data types of all attributes are given in the below table. (See Output 2) There are 20 categorical variables and 3 quantitative variables (1 continuous, 2 discrete variables).

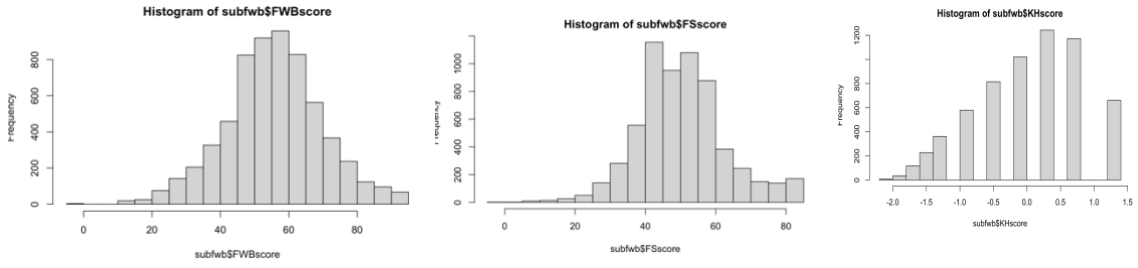
Output 2. Data types of attributes

```
'data.frame':   6239 obs. of  23 variables:
 $ fpl            : int  3 3 3 3 3 3 3 3 3 3 ...
 $ FWBscore       : int  55 51 49 49 67 51 47 58 78 62 ...
 $ FSscore        : int  44 43 42 42 57 54 35 42 66 57 ...
 $ LMscore        : int  3 3 3 1 3 3 3 3 2 3 ...
 $ KHscore        : num  1.267 -0.57 -0.188 -1.9 0.242 ...
 $ EMPLOY         : int  8 2 2 2 2 4 7 2 5 8 ...
 $ Military_Status: int  5 5 5 5 3 5 5 5 5 5 ...
 $ agecat         : int  8 3 3 2 2 3 2 4 1 6 ...
 $ generation     : int  1 3 3 4 4 3 4 3 4 2 ...
 $ PPEDUC         : int  4 2 3 2 4 4 1 2 3 4 ...
 $ PPETHM         : int  1 1 2 3 1 4 1 1 1 1 ...
 $ PPGENDER       : int  1 1 1 1 1 2 2 1 1 2 ...
 $ PPHHSIZE       : int  1 2 3 5 2 5 3 3 5 2 ...
 $ PPINCIMP       : int  7 6 6 7 7 7 4 8 6 8 ...
 $ PPMARIT        : int  3 3 3 1 1 1 4 1 1 1 ...
 $ PPMSACAT       : int  1 1 1 1 1 1 1 1 1 0 ...
 $ PPREG4         : int  4 2 4 2 2 2 2 1 4 4 ...
 $ PPREG9         : int  8 3 9 4 3 4 4 2 8 8 ...
 $ PPT01          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPT25          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPT612         : int  0 0 0 1 0 1 0 0 0 0 ...
 $ PPT1317        : int  0 0 1 0 0 1 0 0 1 0 ...
 $ PPT180V        : int  1 2 2 4 2 3 3 3 3 2 ...
```

3.3. Plots

There are different graph types that can be used for quantitative variables or qualitative variables. Histogram, one of these plotting methods, is a widely used graph type for quantitative variables because it gives an idea about the distribution of attributes. Since FWB score, Financial Skill (FS) score and Knoll and Houts Financial Knowledge (KHFK) scale score are quantitative variables, histograms are drawn for these variables. (See Output 3) The related observations are shared in the previous section. (See Section 3.1)

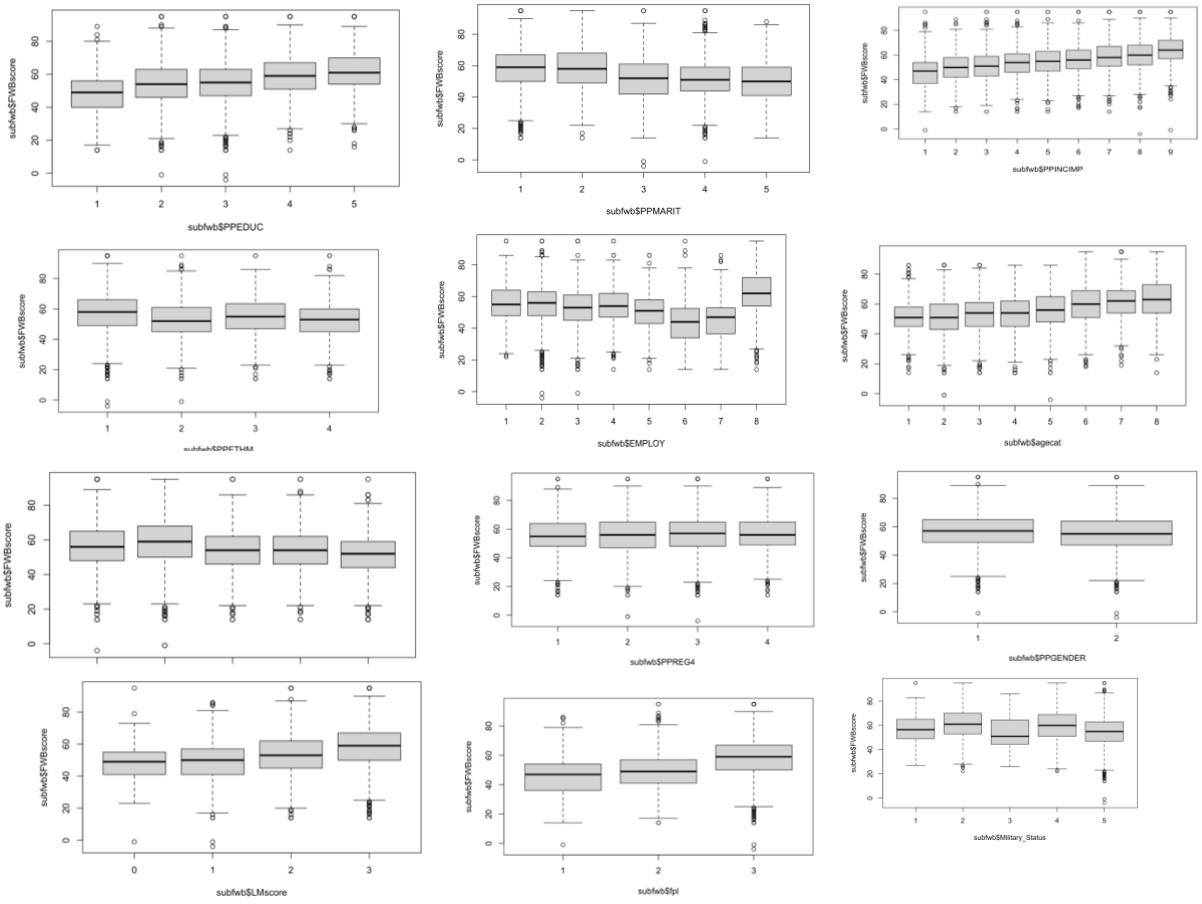
Output 3. Histograms (FWB, FS, KHFK respectively)



On the other hand, boxplots are preferred for categorical variables, so that we can observe outliers and some descriptive statistics such as minimum, maximum values, average, etc. In addition, if we

draw boxplot for two variables: one on x -axis, the other on y-axis, we can identify the relationship between these two attributes. Hence, to be able to see whether there is a pattern between each category for dependent variable, boxplots were used for one independent variable and one dependent variables, which is FWB. The observations are reported in the previous section. (See 3.1.) The categories of each independent variable were given in the public use file codebook of this survey

Output 4. Boxplots (x-axis: independent variable[1], y-axis: FWB-dependent variable)



1.1.Correlations

The correlation shows the direction and strength of the relationship between two variables. It can be between -1 to 1. It does not give any information about the causality. The closer the correlation is near to zero, the weaker the relationship is. If it is closer to one in absolute terms, it shows a strong relationship. The signs of the correlation give the direction of the relationship. Positive sign indicates positive relation, whereas negative sign indicates negative relation. If one is in interested in the correlation of more than two variables, correlation matrix should be constructed. In addition,

[1] The independent variables are education, marital status, income, ethnicity/race, employment status, age, household size, region census, gender, LM financial knowledge score, poverty status, military status respectively. (From left to right, from up to down)

there are different methods in calculating the correlation. For numerical variables, Pearson correlation method is appropriate, however, if the variables are categorical, Spearman rank correlation is more suitable. I will use correlation matrix for two purposes: (1) to choose one of the highly correlated variables in order to avoid multicollinearity problem which may lead to biased results, (2) to see the direction of the relationships.

Output 5. Pearson Correlation Matrix

	subfwb.FWBscore	subfwb.FSscore	subfwb.KHscore
subfwb.FWBscore	1.0000000	0.4983156	0.3330181
subfwb.FSscore	0.4983156	1.0000000	0.1926334
subfwb.KHscore	0.3330181	0.1926334	1.0000000

Output 6. Spearman Rank Correlation Matrix

	fpl	FWBscore	FSscore	LMscore	KHscore	EMPLOY	Military_Status	agecat	generation	PPEDUC	PPETHM	PPGENDER
fpl	1.0000	0.3454	0.1740	0.3097	0.3894	-0.0754	-0.1065	0.1581	-0.1499	0.3915	-0.2509	-0.0802
FWBscore	0.3454	1.0000	0.5196	0.2400	0.3517	0.1511	-0.1548	0.3135	-0.2962	0.2519	-0.1498	-0.0533
FSscore	0.1740	0.5196	1.0000	0.1487	0.2293	0.0094	-0.0823	0.0741	-0.0652	0.2278	-0.0728	-0.0723
LMscore	0.3097	0.2400	0.1487	1.0000	0.6056	-0.0113	-0.0954	0.1557	-0.1477	0.3161	-0.2240	-0.1674
KHscore	0.3894	0.3517	0.2293	0.6056	1.0000	0.0276	-0.1290	0.2351	-0.2267	0.4180	-0.3016	-0.2249
EMPLOY	-0.0754	0.1511	0.0094	-0.0113	0.0276	1.0000	-0.2360	0.5469	-0.5117	-0.1399	-0.0864	0.1184
Military_Status	-0.1065	-0.1548	-0.0823	-0.0954	-0.1290	-0.2360	1.0000	-0.2999	0.2826	-0.0355	0.0967	0.1379
agecat	0.1581	0.3135	0.0741	0.1557	0.2351	0.5469	-0.2999	1.0000	-0.9552	-0.0233	-0.1718	0.0366
generation	-0.1499	-0.2962	-0.0652	-0.1477	-0.2267	-0.5117	0.2826	-0.9552	1.0000	0.0365	0.1650	-0.0311
PPEDUC	0.3915	0.2519	0.2278	0.3161	0.4180	-0.1399	-0.0355	-0.0233	0.0365	1.0000	-0.1259	-0.1431
PPETHM	-0.2509	-0.1498	-0.0728	-0.2240	-0.3016	-0.0864	0.0967	-0.1718	0.1650	-0.1259	1.0000	0.0163
PPGENDER	-0.0802	-0.0533	-0.0723	-0.1674	-0.2249	0.1184	0.1379	0.0366	-0.0311	-0.1431	0.0163	1.0000
PPHHSIZE	-0.1300	-0.1307	-0.0484	-0.0550	-0.0758	-0.2166	0.1045	-0.3755	0.3610	-0.0306	0.1006	0.0066
PPINCIMP	0.7056	0.3928	0.2372	0.3089	0.4190	-0.1822	-0.0475	0.0112	-0.0148	0.5121	-0.2059	-0.0954
PPMARIT	-0.2376	-0.2519	-0.1243	-0.1690	-0.2220	-0.0917	0.1647	-0.3334	0.3115	-0.1012	0.1393	0.0324
PPMSACAT	0.0676	0.0252	0.0350	0.0121	0.0335	-0.0546	0.0361	-0.0401	0.0415	0.1003	0.1334	-0.0380
PPREG4	-0.0290	0.0243	0.0230	-0.0054	-0.0048	0.0101	-0.0638	-0.0084	0.0173	0.0230	0.1625	-0.0370
PPREG9	-0.0372	0.0197	0.0239	-0.0096	-0.0099	0.0054	-0.0549	-0.0159	0.0237	0.0149	0.1821	-0.0328
PPT01	-0.0707	-0.0400	0.0213	-0.0150	-0.0245	-0.0852	0.0501	-0.1733	0.1847	0.0461	0.0272	0.0008
PPT25	-0.1192	-0.0812	-0.0004	-0.0605	-0.0702	-0.1340	0.0749	-0.2484	0.2601	0.0123	0.0396	0.0159
PPT612	-0.1435	-0.1057	-0.0366	-0.0723	-0.0853	-0.1711	0.0773	-0.2492	0.2483	-0.0255	0.1108	0.0498
PPT1317	-0.1121	-0.1049	-0.0476	-0.0637	-0.0656	-0.1523	0.0774	-0.1773	0.1593	-0.0582	0.0970	0.0002
PPT18OV	-0.0359	-0.0588	-0.0378	-0.0091	-0.0259	-0.0964	0.0419	-0.2096	0.1934	-0.0349	0.0593	-0.0137

	PPHHSIZE	PPINCIMP	PPMARIT	PPMSACAT	PPREG4	PPREG9	PPT01	PPT25	PPT612	PPT1317	PPT18OV
fpl	-0.1300	0.7056	-0.2376	0.0676	-0.0290	-0.0372	-0.0707	-0.1192	-0.1435	-0.1121	-0.0359
FWBscore	-0.1307	0.3928	-0.2519	0.0252	0.0243	0.0197	-0.0400	-0.0812	-0.1057	-0.1049	-0.0588
FSscore	-0.0484	0.2372	-0.1243	0.0350	0.0230	0.0239	0.0213	-0.0004	-0.0366	-0.0476	-0.0378
LMscore	-0.0550	0.3089	-0.1690	0.0121	-0.0054	-0.0096	-0.0150	-0.0605	-0.0723	-0.0637	-0.0091
KHscore	-0.0758	0.4190	-0.2220	0.0335	-0.0048	-0.0099	-0.0245	-0.0702	-0.0853	-0.0656	-0.0259
EMPLOY	-0.2166	-0.1822	-0.0917	-0.0546	0.0101	0.0054	-0.0852	-0.1340	-0.1711	-0.1523	-0.0964
Military_Status	0.1045	-0.0475	0.1647	0.0361	-0.0638	-0.0549	0.0501	0.0749	0.0773	0.0774	0.0419
agecat	-0.3755	0.0112	-0.3334	-0.0401	-0.0084	-0.0159	-0.1733	-0.2484	-0.2492	-0.1773	-0.2096
generation	0.3610	-0.0148	0.3115	0.0415	0.0173	0.0237	0.1847	0.2601	0.2483	0.1593	0.1934
PPEDUC	-0.0306	0.5121	-0.1012	0.1003	0.0230	0.0149	0.0461	0.0123	-0.0255	-0.0582	-0.0349
PPETHM	0.1006	-0.2059	0.1393	0.1334	0.1625	0.1821	0.0272	0.0396	0.1108	0.0970	0.0593
PPGENDER	0.0066	-0.0954	0.0324	-0.0380	-0.0370	-0.0328	0.0008	0.0159	0.0498	0.0002	-0.0137
PPHHSIZE	1.0000	0.1439	-0.1863	0.0248	0.0418	0.0445	0.2264	0.3709	0.4913	0.4550	0.8058
PPINCIMP	0.1439	1.0000	-0.2913	0.1248	-0.0124	-0.0154	0.0048	-0.0105	0.0018	0.0210	0.1770
PPMARIT	-0.1863	-0.2913	1.0000	0.0379	-0.0270	-0.0212	-0.0514	-0.0687	-0.0860	-0.0634	-0.1717
PPMSACAT	0.0248	0.1248	0.0379	1.0000	0.0583	0.0505	0.0080	-0.0030	0.0170	0.0242	0.0256
PPREG4	0.0418	-0.0124	-0.0270	0.0583	1.0000	0.9718	0.0238	0.0269	0.0376	0.0238	0.0318
PPREG9	0.0445	-0.0154	-0.0212	0.0505	0.9718	1.0000	0.0273	0.0283	0.0428	0.0232	0.0343
PPT01	0.2264	0.0048	-0.0514	0.0080	0.0238	0.0273	1.0000	0.2593	0.0905	0.0004	0.0427
PPT25	0.3709	-0.0105	-0.0687	-0.0030	0.0269	0.0283	0.2593	1.0000	0.3049	0.0403	0.0696
PPT612	0.4913	0.0018	-0.0860	0.0170	0.0376	0.0428	0.0905	0.3049	1.0000	0.2737	0.0972
PPT1317	0.4550	0.0210	-0.0634	0.0242	0.0238	0.0232	0.0004	0.0403	0.2737	1.0000	0.1708
PPT18OV	0.8058	0.1770	-0.1717	0.0256	0.0318	0.0343	0.0427	0.0696	0.0972	0.1708	1.0000

1. Step-by-Step Overall Methodology:

1.1. Importing and Pre-Processing of Data Set

The first step is to import the dataset from the link: https://www.consumerfinance.gov/data-research/financial-well-being-survey-data/. This is a survey data for 2016. It means there are so many binary values for most of the attributes. I firstly imported all the dataset without eliminating any attributes or observations to see the data set as a whole.

1.2. Organizing Data Set

R Programming Language was used for this part.

1.2.1. Removing Unnecessary Attributes

There are 217 attributes and 6402 observations. 4 different scores from were calculated some of these attributes. Since I am not interested in creating my own scaling or measurement, I excluded all the attributes that were used or can be used in these sorts of calculations.

Secondly, I will not use 'Generation' and 'Census Region' in my study, because they are highly correlated with 'Age' and 'Census Division' respectively. (Done after Explanatory Data Analysis-Correlation Section)

1.2.2. Removing Unanswered / NA Values from Data Set

There were 6402 participants in this survey, however not all the questions were answered, so I eliminated the 'refused to answer' observations from the updated data set. Then I checked whether there was any missing value or N/A value in the dataset. I could not find any.

1.3. Doing Explanatory Data Analysis

R Programming Language was used for this part.

1.3.1. Attribute Types

The first step is to analyze the attribute types. If any of the attribute types needs to be transformed, it should be done in this step. For instance, if it is a Yes/No answer, it should be changed into a binary variable to be able to use in further analysis. In this dataset, I did not do any transformation because even the categorical variables were in the 'integer' form.

1.3.2. Descriptive Statistics

The second step is to do descriptive statistics which tells us about minimum, maximum, mean, median, first and third quartiles of each variable. We can guess the skewness or the range of that variable.

1.3.3. Plots

The third step is to plot some graphs to examine the distributions and outliers if there exists. I used 'histogram' for quantitative variables and 'boxplots' for categorical variables. I mentioned their distributions. Also doing boxplots of dependent variable and independent variables' each category, I could observe some relational findings.

1.3.4. Correlations

The last step is to do correlation matrix between each of the variable to see the strength and direction of the variables. Some variables were found (mentioned above) highly correlated, so for further analysis, they will be removed. Also the direction of the relationships were interpreted.

1.4. Applying Machine Learning Methods

R and Python programming languages are planning to be used.

Since this data is labeled, I will use one of supervised learning algorithms: classification algorithms. To be more specific, I plan to use Naïve Bayes classifier, Decision trees and k-Nearest Neighbors (k-NN). I will do evaluation measurement performances to decide which method I should continue for interpretation.

2. REFERENCES

Clark, R., & D'Ambrosio, M. (2008). *Adjusting retirement goals and saving behavior: The role of financial education.* In: Lusardi, A. (Ed.), Overcoming the saving slump: How to increase the effectiveness of financial education and saving program. University of Chicago Press.

Consumer Financial Protection Bureau. (2017). CFPB financial wellbeing scale: scale development technical report. *New York: Consumer Financial Protection Bureau.* https://files.consumerfinance.gov/f/documents/201705_cfpb_financial-well-being-scale-technical-report.pdf

Fan, L. & Henager, R. (2021). A structural determinants framework for financial well-being. *Journal of Family and Economic Issues,Sep 16*:1-14 https://doi.org/10.1007/s10834-021-09798-w

Gutter, M. & Copur, Z. (2011). Financial behaviors and FWB of college students: Evidence from a national survey. *Journal of Family and Economic Issues*, *32*:699-714 https://doi.org/10.1007/s10834-011-9255-2

Hilgert, M., & Hogarth, J. (2003). Household financial management: The connection between knowledge and behavior. *Federal Reserve Bulletin, 89(7)*: 309–322. https://www.federalreserve.gov/pubs/bulletin/2003/0703lead.pdf

Iannello, P., Sorgente, A., Lanz, M. & Antonietti, A. (2021). Financial well-being and its relationship with subjective and psychological well-being among emerging adults: testing the moderating effect of individual differences. *Journal of Happiness Stud* (22): 1385–1411 https://doi.org/10.1007/s10902-020-00277-x

Iramani, R., & Lutfi, L. (2021). An integrated model of financial well-being: The role of financial behavior. *Accounting, 7(3):* 691–700. http://growingscience.com/beta/ac/4557-an-integrated-model-of-financial-well-being-the-role-of-financial-behavior.html

Joo, S., & Grable, J. E. (2004). An exploratory framework of the determinants of financial satisfaction. *Journal of Family and Economic Issues, 25(1*): 162–171. https://doi.org/10.1023/B:JEEI.0000016722.37994.9f

Lee, J.M, Lee, J. & Kim, K.T. (2020). Consumer financial well-being: Knowledge is not enough. *Journal of Family and Economic Issues, 41*:218-228 https://doi.org/10.1007/s10834-019-09649-9

Zulfiqar, M., & Bilal, M. (2016). Financial wellbeing is the goal of financial literacy. *Research Journal of Finance and Accounting, 7(11)*: 94–103. https://core.ac.uk/download/pdf/234631442.pdf