

# Configuration Manual

MSc Research Project  
FinTech

**Rachita Patel**  
Student ID: x18121331

School of Computing  
National College of Ireland

Supervisor: Prof. Noel Cosgrave

National College of Ireland Project  
Submission Sheet School of  
Computing



<b>Student Name:</b>	Rachita Patel
<b>Student ID:</b>	x18121331
<b>Programme:</b>	FinTech
<b>Year:</b>	2019
<b>Module:</b>	MSc Research Project
<b>Supervisor:</b>	Prof. Noel Cosgrave
<b>Submission Due Date:</b>	12/08/2019
<b>Project Title:</b>	Configuration Manual
<b>Word Count:</b>	1890
<b>Page Count:</b>	9

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

<b>Signature:</b>	
<b>Date:</b>	11th August 2019

**1.1.1 PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

Attach a completed copy of this sheet to each project (including multiple copies).	Q
<b>Attach a Moodle submission receipt of the online project submission</b> , to each project (including multiple copies).	Q
<b>You must ensure that you retain a HARD COPY of the project</b> , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	Q

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Configuration Manual

Rachita Patel

X18121331

## 1. Introduction

The software and hardware requirements collectively termed as the system configuration which is required for the reproducibility of the thesis titled “*Understanding Saving Habit of Individuals for the Varied Financial Measurement Categories using Cluster Analysis*” is described in this configuration manual.

This manual details the data collection process in section 2. The hardware and software system setup is described in section 3 followed by the software and libraries used in section 4. Lastly, section 5 illustrates important results.

## 2. Data Collection

The financial well-being survey data published in September 2017 has been collected from the Consumer Financial Protection Bureau (CFPB) which is a .csv file. It has a total of 217 attributes and 6394 records. Taking into consideration the CFPB financial well-being survey report, this dataset was further divided into six financial measurement categories. So, this is being done using ‘R platform’.

## 3. System Configuration

### Hardware

#### Device Specifications

CPU: Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz

GPU: NVIDIA GeForce MX150 (2GB GDDR5) 69

RAM: 4.00GB

STORAGE: 1000GB HDD

SYSTEM TYPE: 64-bit operating system, x64-based processor

Windows Specifications  
EDITION: Windows 10 Home  
VERSION: 1803

### Software

Windows 10: Data processing  
Zotero: Centralised database for the peer-reviewed papers  
RStudio version 1.1.456: Data implementation platform (for exploratory data analysis, building several machine learning algorithms- supervised as well as unsupervised).  
R version 3.5.1: Programming language  
Microsoft Excel 365 version 1907: Data collection and extraction  
Microsoft Word 365 version 1907 and TeXworks version 0.6.2 (MikTeX 2.9.6930 64-bit)  
{LaTeX}: Writing drafts and final report

## **4. Software and Libraries Used**

'R platform' has been used for this research. Many packages and libraries had to be installed in order to carry out the analysis. As discussed in section 1, there are six financial measurement categories so corresponding six .R files are produced containing approximately 350 lines of code respectively.

Packages and Libraries used are as follows:

<b>Sr. No.</b>	<b>Package and version number</b>	<b>Library used</b>
1.	R.utils 2.7.0	read.table
2.	knitr 1.20	knitr
3.	corrplot 0.84	corrplot
4.	rcompanion 2.2.2	rcompanion
5.	ModelMetrics 1.2.2	Metrics
6.	randomForest 4.6-14	randomForest
7.	NbClust 3.0	NbClust
8.	factoextra 1.0.5	factoextra
9.	clvalid 0.6-6	clvalid, cluster
10.	fpc 2.2-3	fpc

## **5. Important Results**

This section illustrates the data quality report for the six code files or the financial measurement categories. It also enlists the description of selected attributes and the encoded values for all the six types. The encoding of values is done by CFPB.

### a. codeA\_indi.R – Individual Characteristics

#### I. DQR

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
Pareduc	6294	0	5	1	2	3	4
Generation	6294	0	4	1	2	2	3
Ppgender	6294	0	2	1	1	1	2
Savehabit	6294	0	6	1	4	5	6

#### II. Variable Information

Description	PAREduc	generation	PPGENDER	SAVE_HABIT
What?	Education (highest degree received)	Generation	Gender	Putting money into savings is a habit for me
How many?	-1 to 5	1 to 4	1 and 2	-1 to 6
-1	Refused	-	-	Refused
0	-	-	-	-
1	Less than high school	Pre-Boomer	Male	Strongly disagree
2	High school degree / GED	Boomer	Female	Disagree
3	Some college / Associate	Gen X	-	Disagree slightly
4	Bachelor's degree	Millennial	-	Agree slightly
5	Graduate / professional degree	-	-	Agree
6	-	-	-	Strongly Agree

## b. codeB\_hofa.R – Household and family Characteristics

### I. DQR

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
Housing	5750	0	3	1	1	1	2
Housesat	5750	0	4	1	3	3	4
Ppmarit	5750	0	5	1	1	1	3
Kids_nochildren	5750	0	2	0	0	1	1
Savehabit	5750	0	6	1	4	5	6

### II. Variable Information

Description	HOUSING	HOUSESAT	PPMARIT	KIDS_NoChildren	SAVE_HABIT
What?	Housing status	How satisfied are you with the place you currently live in?	Marital Status	Financially supporting children?	Putting money into savings is a habit for me
How many?	-1 to 3	-1 to 4	1 to 5	-1 to 1	-1 to 6
-1	Refused	Refused	-	Refused	Refused
0	-	-	-	Respondent financially supports children	-
1	I own my home	Not at all satisfied	Married	I have no children that I support financially	Strongly disagree
2	I rent	Not very satisfied	Widowed	-	Disagree
3	I do not currently own or rent	Somewhat satisfied	Divorced / Separated	-	Disagree slightly
4	-	Very satisfied	Never Married	-	Agree slightly
5	-	-	Living with partner	-	Agree
6	-	-	-	-	Strongly Agree

### c. codeC\_inex.R – Income and employment Characteristics

#### I. DQR

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
EMPLOY	6192	0	8	1	2	3	8
PPINCIMP	6192	0	9	1	3	6	8
VOLATILITY	6192	0	3	1	1	1	2
MILITARY	6192	0	2	0	0	0	0
Military_Status	6192	0	5	1	5	5	5
Savehabit	6192	0	6	1	4	5	6

#### II. Variable Information

Description	EMPLOY	PPINCIMP	VOLATILITY	MILITARY	Military_Status	SAVE_HABIT
What?	Employment status	Household income	Household income volatility	Current / former member of US Armed forces or spouse / dependent of service member	Military Status	Putting money into savings is a habit for me
How many?	1 to 9	1 to 9	-1 to 3	-1 to 1	-1 to 5	-1 to 6
-1	-	-	Refused	Refused	Refused	Refused
0	-	-	-	No	-	-
1	Self-employed	Less than \$20,000	Roughly the same each month	Yes	Active military	Strongly disagree
2	Work full-time for an employer or the military	\$20,000 to \$29,999	Roughly the same most months, but some usually high or low months during the	-	Veteran	Disagree

			year				
3	Work part-time for an employer or the military	\$30,000 to \$39,999	Often varies quite a bit from one month to the next	-		Spouse/dependent active	Disagree slightly
4	Homemaker	\$40,000 to \$49,999	-	-		Spouse/dependent veteran	Agree slightly
5	Full-time student	\$50,000 to \$59,999	-	-		Neither active nor veteran	Agree
6	Permanently sick, disabled or unable to work	\$60,000 to \$74,999	-	-	-		Strongly Agree
7	Unemployed or temporary laid off	\$75,000 to \$99,999	-	-	-		-
8	Retired	\$100,000 to \$149,999	-	-	-		-
99 or 9	Refused	\$150,000 or more	-	-	-		-



**d. codeD\_sav.R – Savings and safety nets**

**I. DQR**

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
SAVINGSRANGES	5978	0	9	1	4	5	7
PRODHAVE_6	5978	0	2	0	0	0	1
BENEFITS_1	5978	0	2	0	0	1	1
BORROW_1	5978	0	2	0	0	1	1
BORROW_2	5978	0	2	0	0	0	1
SAVEHABIT	5978	0	6	1	4	5	6

**II. Variable Information**

Description	SAVINGSRANGES	PRODHAVE_6	BENEFITS_1	BORROW_1	BORROW_2	SAVE_HABIT
What?	How much money do you have in savings today?	Non-retirement investments (such as stocks, bonds or mutual funds)	Health insurance	My friends or family would lend me the money and expect me to repay them	My friends or family would lend me the money with no expectation of repayment	Putting money into savings is a habit for me
How many?	-1 to 9	0 and 1	-1 to 1	-1 to 1	-1 to 1	-1 to 6
-1	Refused	-	Refused	Refused	Refused	Refused
0	-	No	No	No	No	-
1	0	Yes	Yes	Yes	Yes	Strongly disagree
2	\$1-99	-	-	-	-	Disagree
3	\$100-999	-	-	-	-	Disagree slightly
4	\$1000-4999	-	-	-	-	Agree slightly
5	\$5000-19999	-	-	-	-	Agree
6	\$20000-74999	-	-	-	-	Strongly Agree
7	\$75000 or more	-	-	-	-	-
8	I don't know	-	-	-	-	-
99 or 9	Prefer not to say	-	-	-	-	-

**e. codeE\_fiex.R – Financial experiences**

**I. DQR**

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
PRODHAVE_1	6321	0	2	0	1	1	1
HOUSERANGES	6321	0	9	1	3	4	6
SHOCKS_1	6321	0	2	0	0	0	0
PRODHAVE_8	6321	0	2	0	0	0	0
SAVEHABIT	6321	0	6	1	4	5	6

**II. Variable Information**

Description	PRODHAVE_1	HOUSERANGES	SHOCKS_1	PRODHAVE_8	SAVE_HABIT
What?	Have checking or savings account	Housing cost burden	Experienced any negative financial shocks	Have student loan	Putting money into savings is a habit for me
How many?	0 and 1	-1 to 99	0 and 1	0 and 1	-1 to 6
-1	-	Refused	-	-	Refused
0	No	-	No	No	-
1	Yes	Less than \$300	Yes	Yes	Strongly disagree
2	-	\$300-499	-	-	Disagree
3	-	\$500-749	-	-	Disagree slightly
4	-	\$750-999	-	-	Agree slightly
5	-	\$1000-1499	-	-	Agree
6	-	\$1500-1999	-	-	Strongly Agree
7	-	\$2000 or more	-	-	-
98 or 8	-	I don't know	-	-	-
99 or 9	-	Prefer not to say	-	-	-

## f. codeF\_fibe.R – Financial behaviour, skills, and attitudes

### I. DQR

Feature	Instances	Missing	Cardinality	Min	FirstQuantile	Median	ThirdQuantile
FINGOALS	6289	0	2	0	0	1	1
PROPPLAN_4	6289	0	5	1	3	4	4
GOALCONF	6289	0	4	1	3	3	4
SAVEHABIT	6289	0	6	1	4	3	6

### II. Variable Information

Description	FINGOALS	PROPPLAN_4	GOALCONF	SAVE_HABIT
What?	Do you have a current or recent financial goal?	A clear plan of action with detailed steps to achieve my financial goals	Confidence in own ability to achieve financial goals	Putting money into savings is a habit for me
How many?	-1 to 1	-1 to 5	-1 to 4	-1 to 6
-1	Refused	Refused	Refused	Refused
0	No	-	-	-
1	Yes	Strongly disagree	Not at all confident	Strongly disagree
2	-	Disagree	Not very confident	Disagree
3	-	Neither agree or disagree	Somewhat confident	Disagree slightly
4	-	Agree	Very confident	Agree slightly
5	-	Strongly agree	-	Agree
6	-	-	-	Strongly Agree