UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE

Master's thesis
(Magistrsko delo)

Imputing missing answers in the World Values Survey

(Imputiranje manjkajočih odgovorov v Svetovni Raziskavi Vrednot)

Ime in priimek: *Arsen Matej Golubovikj*
Študijski program: *Podatkovna znanost, 2. stopnja*
Mentor: *doc. dr. Branko Kavšek*
Somentor: *izr. prof. dr. Marko Tkalčič*

Koper, avgust 2022

Ključna dokumentacijska informacija

Ime in PRIIMEK: Arsen Matej GOLUBOVIKJ

Naslov magistrskega dela: Imputiranje manjkajočih odgovorov v Svetovni Raziskavi Vrednot

Kraj: Koper

Leto: 2022

Število listov: 112 Število slik: 8 Število tabel: 4

Število prilog: 2 Število strani prilog: 56

Število referenc: 44

Mentor: doc. dr. Branko Kavšek

Somentor: izr. prof. dr. Marko Tkalčič

UDK: 004.6:303.425(043.2)

Ključne besede: survey data imputation, imputation, matrix completion, collaborative filtering

Math. Subj. Class. (2020): 68T09, 62-08

Izvleček:

Na številnih področjih znanosti, zlasti na področju družboslovja, se za zbiranje podatkov uporabljajo vprašalniki. Postopek zbiranja podatkov z vprašalnikom, imenovan anketa, je pogosto glavno sredstvo za zbiranje podatkov neposredno od udeležencev, vendar je nagnjen k manjkajočim podatkom. Da bi ohranili celoten vzorec ankete, morajo raziskovalci za reševanje te težave pogosto uporabiti imputacijo. Metode za imputacijo lahko včasih ponudijo razumne ocene za manjkajoče podatke, vendar v primeru ankete: (i) imputacija lahko podatkom doda velik šum, ki vpliva na sklepanje, (ii) imputacija postane nezanesljiva, kadar manjka več kot 40 % podatkov. To magistrsko delo poskuša obravnavati ta vprašanja z oceno, ali lahko uporaba metod, ki izhajajo iz sodelovalnega filtriranja (CF) v priporočilnih sistemih, omogoči natančnejše pripisovanje manjkajočih vrednosti v anketnih podatkih. Razlog za uporabo teh metod je (i) podobnost med oblikovanjem problema, metodami in predstavitvijo podatkov, ki se uporabljajo pri CF in imputiranju vprašalnikov; (ii) učinkovitost metod, ki temeljijo na CF, v priporočilnih sistemih. Uporabljamo podatke iz Svetovne raziskave vrednot, dragocenega nabora podatkov v družboslovju z velikim obsegom in verodostojnostjo, za primerjavo (i) enega preprostega pristopa k imputaciji, (ii) dveh uveljavljenih pristopov k imputaciji (iii) dveh tehnik dopolnjevanja matrik, ki izhajata iz sodelovalnega filtriranja. Rezultati kažejo, da so naše izbrane tehnike dopolnjevanja matrik, ki izhajajo iz sodelovalnega filtriranja, v primeru raziskave primerljive, vendar ne boljše od obstoječih tehnik imputiranja. Prava tehnika za imputacijo je pogosto odvisna od problema, ti rezultati pa vabijo k upoštevanju tehnik, ki temeljijo na CF, v prihodnjih raziskavah o imputaciji anket.

Key document information

Name and SURNAME: Arsen Matej GOLUBOVIKJ
Title of the thesis: Imputing missing answers in the World Values Survey

Place: Koper
Year: 2022

Number of pages: 112 Number of figures: 8 Number of tables: 4
Number of appendices: 2 Number of appendix pages: 56
Number of references: 44
Mentor: Assist. Prof. Branko Kavšek, PhD
Co-Mentor: Assoc. Prof. Marko Tkalčič, PhD

UDC: 004.6:303.425(043.2)
Keywords: survey data imputation, imputation, matrix completion, collaborative filtering
Math. Subj. Class. (2020): 68T09, 62-08

Abstract:
Many areas of science, in particular social sciences, use questionnaires to gather data. The process of collecting data through a questionnaire, called a survey, is often the prime means of gathering data directly from participants, however, it's prone to missing data. In order to keep the full survey sample, researchers must often use imputation to deal with this problem. Methods for imputation can sometimes offer reasonable estimates for the missing data, however, in the case of the survey: (i) imputation can add high noise to the data, which influences the inference, (ii) imputation becomes unreliable when more than 40% of the data is missing. This thesis attempts to address these issues by evaluating if the usage of methods stemming from collaborative filtering (CF) in recommender systems can yield more accurate imputations of missing values in survey data. The rationale for the usage of these methods is (i) the similarity between the problem framing, methods and data representation used in CF and questionnaire imputation; (ii) the effectiveness of CF-based methods in recommender systems. We use data from the World Values Survey, a valuable dataset in social science of high volume and veracity, to compare (i) one simple approach to imputation, (ii) two established imputation approaches (iii) two matrix completion techniques stemming from collaborative filtering. The results show that our chosen matrix completion techniques stemming from collaborative filtering perform comparable, but not better than existing imputation techniques in the case of the survey. The right technique for imputation often depends on the problem, these results beckon the consideration of CF-based techniques in future research on survey imputation.

List of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

i.e.	that is
e.g.	for example
rf.	reference
etc.	et cetera
CF	collaborative filtering
RS	recommender systems
KNN	k-nearest neighbors
I2I	item to item collaborative filtering
MF	matrix factorization
CV	cross validation
MAR	missing at random
MCAR	missing completely at random
NMAR	not missing at random

Acknowledgments

I would like to thank the University of Primorska and Famnit for accepting me as foreign student and supporting me throughout my studies, their scholarship program helps a lot of young students, and it was a great honor to be selected as part of it.

I would also like to extend the deepest of gratitudes to my mentors doc. dr. Branko Kavšek and izr. prof. Marko Tkalčič, for their support, patience, perseverance, encouragement and guidance, during the writing of this thesis.

Lastly, I would like to thank my family for all the love and support that they have given me, seeing me through to this point in my studies. More than half of my success is due to them.

Arsen Matej Golubovikj,
August 17, 2020.

1 INTRODUCTION

For many areas of science, in particular social sciences, questionnaires are an essential tool for gathering data. A questionnaire is an instrument for measuring variables through a set of questions or prompts [39]. The process of gathering data through a questionnaire we call a survey [16]. Surveys have advantages, such as getting data directly from the participants, however, they often produce missing data [37] - data values which are unobserved [26], also called blanks. These blanks may be caused by many reasons: (i) the survey administrators might make an error [29] (ii) participants might miss or choose not to answer certain questions [29](iii) reasons beyond the influence of administrators or participants [29], e.g. acts of nature. No matter their cause, missing values in questionnaire-acquired data must be dealt with before researchers can make inferences from the data [7].

A common approach to dealing with missing values is to delete all entries which contain them [26]. The advantage this deletion is its simplicity [26], however, it forces the researcher to operate on a partial dataset, which might produce misleading results [26]. In order to operate on the whole data, missing values must be filled in with with replacement values. The process of doing this is called imputation [7]. Imputation techniques for handing missing data in the questionnaire case are no different than in other cases [29, 7]. Often used techniques for imputation include: (i) simple imputation cases [29], which replaces missing data in a variable with its average or most frequent value (ii) hot-deck imputation [29], which exploits the similarities between entries in the data to find suitable replacements (iii) model based approaches [29], which model each variable based on the available data and fill in missing values using the model for each variable. All of these techniques for imputation have their own advantages and disadvantages, which will be discussed in more detail the related work chapter.

The unifying advantage of imputation techniques is that they allow the researchers to use the larger data sample [7], with, to a degree, reasonable estimates of the missing values [29]. However, there are also disadvantages: (i) imputation methods can introduce high noise to the data, which influences the conclusion drawn from such data [26] (ii) in the case of the survey, imputation techniques are often not effective on data with high missingness, i.e. when more than 40% of the data is missing [29].

In our work we address these issues (i) noise introduction and (ii) imputing survey data with high missingness, by evaluating if the usage of alternative imputation methods that are commonly used in recommender systems can yield more accurate imputations of missing values, both in the case of low and high missingness. The rationale for the usage of these methods is (i) the similarity of the problem framing between questionnaires and recommender systems and (ii) the effectiveness of these methods in recommender systems, especially in the case when much of the data is unavailable (iii) a similarity between approaches in survey imputation and certain approaches in recommender system. We will discuss the similarity of approaches in the chapter covering the related work.

A similar framing of the questionnaire missing data happens in the domain of recommender systems. Broadly, the field of recommender systems [1] (RS) deals with recommending items to users. For the purpose of recommendation these systems usually collect ratings which users give to items that they are familiar with. These ratings contain the human user's opinion about the given items, usually expressed as a scalar value, for example 1 to 5 [1], where high values denote a high opinion and low ratings a low opinion. The ratings which the user has given to items he/she is familiar with are then utilized to predict the user's opinion of unfamiliar items, and, afterwards, items with high predicted ratings, i.e. opinion, are recommended to the user. We can see the similarity between the recommendation step and survey imputation since these systems are essentially predicting question answers of the form "What is your opinion of this item?".

Moreover, the collaborative filtering [1] technique from this field, bases its predictions on an item to user matrix where the i, j^{th} (row, column) entry denotes user i's rating for item j. Due to the large volume of items in such systems, users are usually familiar with only a fraction of the items, consequently, much of the entries in the ratings matrix are empty [1], i.e. missing. The prediction is then done by filling these missing using solely data from this matrix, through a process called matrix completion [1]. If we represent the questionnaire data as a matrix, where rows represent participants, columns represent questions and the i, j^{th} entry denotes the participant i's answer to question j, the problem of filling missing data is now similar to the problem of matrix completion. This thesis attempts to utilize this similarity.

We can now state the problem statements of this thesis. (i) How does the performance of matrix completion techniques stemming from collaborative filtering compare to classical survey imputation methods? (ii) What is the relation between the rate of missing data and the performance of the matrix completion and classical survey imputation techniques?

Our work address these statements using data from the World Values Survey [19] (WVS), which has gathered questionnaire data about the social, political, economic, religious, and cultural values of people in the world. It contains the answers of over 80 000 participants to more than 290 questions on many different scales and topics. For this reason the data from the WVS is highly valued, and is the topic of thousands of publications in many areas of research [27]. Such valuable data will prove an ample testing grounds for our methods. We will employ it in the comparison of (i) one simple approach to imputation, which we will use as a baseline for the other methods (ii) two established imputation approaches (iii) two matrix completion completion techniques stemming from collaborative filtering.

The limitation of this research is the fact that is based on one use instance and may not be fully representative of the general case. Nonetheless, we believe that results from data with such volume and veracity will indeed contribute to the research on imputation and matrix completion, concluding on the usefulness of our chosen alternate methods for the case of survey data imputation.

Continuing, the thesis will be structured as follows. Chapter 2 reviews the related work on this topic and reveals how our study contributes to the existing body of literature. Chapter 3 explains key terms and concepts on this subject, providing a theoretical background for the remained of our work. Chapter 4 describes the World Values Survey and analyzes its questionnaire data which we use in our trials. The study methodology, and the approaches we compare are described in Chapter 5, while its results are presented and discussed in Chapter 6. Finally, Chapter 7 lists our conclusions and provides recommendations for future work.

2 RELATED WORK

The structure of this chapter is as follows. First, we provide an overview of methods for survey data imputation. Then we cover techniques for matrix completion known in recommender systems research on collaborative filtering. We further present the existing work that applies CF-based matrix completion methods outside the field of recommender systems, and especially in survey data imputation problems. Finally, we identify the gaps in this research and provide the rationale for the proposed work.

2.1 SURVEY DATA IMPUTATION

Rubin [34] is arguably the first to express the need for explicit care when handling missing data. His work, and later the work of Heitjan and Basu [21] establish the well know consideration for handling missing data: (i) the method for handling missing data has an effect on the statistical inference [34, 21], (ii) when choosing how to handle missing data, considerations need to be made based on its reasons for missingness [21] or the "process that causes missing data" [34], (iii) three categories of missing data can be established based on "why its missing": missing at random (MAR) [34], missing completely at random (MCAR) [21] and not missing at random (NMAR) [34]. Broadly, MCAR variables are missing due to processes unrelated to both obeserved and unoberved data, missingness in MAR variables is related to other observed variables in the data, in NMAR variables the reasons and patterns for missingness are not captured by the observed data. Missing data which is NMAR requires the collection of additional data related to the missingness for the purpose of imputation. Imputation techniques usually work under the assumption that missing data is MCAR or MAR [26].

Older works on survey imputation, such as those of Giles [15], Brick and Calton [6], recommend the simple, hot-deck and regression based approaches. These methods are still relevant today.

Simple imputation handles missing data in each variable by replacing it with its mean or mode value [26]. Methods for simple imputation can be effective when data is divided into smaller parts containing small portions of missing data and each portion is imputed with simple imputation separately [15, 6, 26], however,(i) the choice of the correct division is often difficult [15, 6], (ii) at times, even with the right division simple methods can introduce monotony leading to bias [29].

Hot-deck [6] or donor [15] imputation aims to solve the problem of division, by for each entry finding other entries to impute from called donors [15]. K-nearest neighbors imputation (KNN) is a popular hot-deck method [6], which uses the K nearest , i.e. most similar, entries as donors [26]. Hot-deck methods have advantages, such as, exploiting and maintaining the similarities between entries in the data, however, (i) the choice of the donor group, for instance, the choice of the number K in KNN, is difficult [6] (ii) it is unbiased only under the MCAR assumption which is rarely holds in practical scenarios [26].

Regression [15, 6] based approaches handle both MCAR and MAR missing data [29], by building one regressor per feature, and imputing using this model. While regressors can be robust in their imputation of missing data [31], they cannot handle missing data on their own and often need (i) to be built on a subset with no missing data, which may be unrepresentative or nonexistent (ii) to be built after an initial imputation of the data, when there are many missing values this may heavily bias the regressors [7].

The biases which these existing imputations methods may produce are addressed specifically for the survey by later works. Random imputation found in [9, 8] attempts reduce variance bias in hot-deck methods by randomizing the donor selection process. This randomization manages to increase the variance in the resulting imputation while still achieving accurate imputations, however, it can overshoot introducing too much variance in the data [8]. A more stable approach in the case of the survey is achieved by focusing on proper post-stratification and clustering [24, 36]. Imputing for each cluster or strata separately greatly increases the accuracy of survey methods, however, the robust process of creating these partitions often requires external data sources or additional data to be collected.

Modern work on survey imputation [29] recommends similar methods as older studies. For the problem of variance they recommend multiple imputation which produces multiple possible imputations for the same missing value [29]. They are beneficial since it has been shown that they capture the variability nicely [29], however have the downsides of a more complicated data analysis phase now including a tensor instead of a matrix [31].

.

2.2 MATRIX COMPLETION IN RECOMMENDER SYSTEMS

Matrix completion techniques and collaborative filtering are made popular by the BellKor [3] solution to the Netflix prize [2, 11]. Collaborative filtering is the name used for techniques in recommender systems which utilize a ratings matrix, while matrix completion are the techniques used to fill blank spaces in a matrix [1]. Works which categorize CF based techniques include the works of Su and Khoshgoftaar [38], Chen et al. [10], and Bobadilla et al. [5].

These works make a division of collaborative filtering techniques based on the method used to predict ratings [5, 10, 38], i.e. the matrix completion algorithm [1]: (i) model-based colaborative filtering, which use machine learning modeling to predict ratings [5, 10, 38] (ii) and memory based techniques [5, 10, 38], popular among which are the neighbourhood based methods [10, 38] that utilize the similarity between entries in the ratings matrix to make predictions.

A further division of CF techniques can be made based on whether they utilize the interplay between items or users: (i) user to user [5] or user-based [10, 38] methods, consider the similarity between users based on their ratings of items, (ii) while item to item [5] or item-based [10, 38] methods, utilize the similarity between items based on how users have rated them.

We can see similar techniques in CF and survey imputation, (i) the user to user neighbourhood method is similar the K-nearset neigbours agorithm utilized in imputation, and (ii) regression based imputation is similar to the item-based modeling CF technique. However, there are also differences, for example, to our knowledge, the item-based neighbourhood methods are unique to the study of collaborative filtering in RS.

2.3 MATRIX COMPLETION AND SURVEYS

Studies have utilized matrix completion and collaborative filtering outside of the filed of recommender systems. Some of the fields which have used these techniques include medicine [20], bioinformatics [35], image processing [18], infrastructure [25] and security [33]. Many of these fields find favorable results in the use of collaborative filtering to their specific problems, especially in the case where large amounts of missing data. Moreover, the specific works of Saha et a [35] and Li et al. [25] have efficiently utilized matrix completion for imputation in DNA and highway traffic related data.

Two works examine the use matrix completion in a broad imputation scenario. (i) Wang et al. produce an ensemble-based imputation method, which includes an item to item collaborative technique in the ensemble. They show that their ensemble method outperforms KNN imputation, on common datasets from the UCI (University of California Irvine) data repository, however, do not evaluate the performance of the item to item collaborative technique on its own. (ii) Chi and Li [11] examines the use of low-rank matrix completion for general role of imputation. They use synthetic data to show that low-rank matrix completion techniques can operate on both the MAR and MCAR statistical assumptions for missing data.

For survey imputation, the case of matrix completion is also highlighted in some cases. Vozalis et. al. [40] make the connection between the often used 1 to 5 Likert style question ranges in surveys and the 1 to 5 stars rating used in recommender systems. They test the usage of a user based collaborative filtering technique in the imputation of a small transportation survey consisting of univariate question answers on the Likert scale. In their testing, 20% of the data is intentionally made missing and imputed using this technique, afterwards they measure the Mean Average Error (MAE) of this imputation by comparing the imputed values to the actual values, finding a MAE of 0.846 for this technique. Similarly, Oliveira et al. [30] compare matrix factorization and item to item collaborative filtering techniques for the purpose of predicting univariate Likert scale questionnaire responses in a large company survey. They find that, on 20% missing data, these techniques can distinguish between a positive and negative response with an Area Under the Curve (AUC) score of 0.80, with non-negative matrix factorization achieving the best result.

A usage of matrix completion in survey imputation with data of similar nature as the World Values Survey can be found in research for Adaptive Survey Design (ADS). Rather than having missing data by chance for item non-response, adaptive survey design attempts to control the process causing the missing data by administering partial questionnaires to each participant which they are more likely to answer fully. As such, it can generate a high volume of missing answers. Zhang et al. [43] test matrix completion techniques in the (ADS) setting on data from the 2016 CCES questionnaire. The CCES gathers political opinion from large number of participant using a questionnaire with questions on multiple scales, we note this dataset, since its has some of the questions in it are similar to those in the WVS. Using the CCES data, among others, they simulate an active learning framework which caters a unique survey for each participant. Answers in the existing data which are not present in this catered survey are deleted, hence simulating missing answers. Afterwards, they use matrix factorization to complete the data and they compare imputed answers to the original answers calculating the MAE. To make sense of the MAE in a multivariate scenario they scale all question answers to the range of -1 to 1. They find, that based on the number of questions in the catered survey, their technique achieves a MAE between 0.75 and 0.60 on the -1 to 1 scale, for the CCES data.

Although there have been research using matrix completion on survey data, to the best of our knowledge, there have been no attempts to compare the effectiveness of matrix completion techniques and classical survey imputation techniques. In this work we fill this gap by directly comparing both approaches on the scenario of World Values Survey data.

3 THEORETICAL BACKGROUND

This chapter details key concepts, techniques and terminology used in thesis, establishing a theoretical framework for the remainder of our work. We first present the surveys and questionnaires, as well as, discuss missing data in such scenarios. Following, we introduce techniques and concepts in (i) imputation, (ii) collaborative filtering, and (iii) other techniques which may aid our work. Finally, we detail methods and metrics which aid in our evaluation.

3.1 SURVEYS AND MISSING DATA

In this section, we will first discuss surveys, and questionares, and later examine missing data in the case of this type of data. A survey is defined as "a systematic method for gathering information from (a sample of) entities for the purposes of constructing quantitative descriptors of the attributes of the larger population of which the entities are members" [16]. While a survey is the method for gathering information, i.e. data, the instrument is often a set of questions in a form of a questionnaire [16]. Questions in questionnaires can be open-ended, allowing the participant to write his own answer to the question, or closed-ended, giving the participant a set of answers to choose from for each question [16].

Most often, questionnaires with closed-ended questions are preferred by researches [28]. This is due to the fact that, closed-ended answers are easier to record and interpret. For example, the set of answers can be numbered, and we can record the number relating to participant's choice from the answers, this is easier to interpreting than written text. Our thesis focuses on questionnaire data with close-ended questions.

Answers in close ended questionnaires can take many forms, and this needs to be taken into account when handling the data. We list some types of question answers applicable to this thesis:

- *Dichotomous questions* [16] - these are questions with binary answers which often refer to questions with two answer, e.g., Yes/No, Agree/Disagree questions. However, in the data, they can be used for other purposes as well, for example, if the participant can choose multiple answers for the same question, then, for each singe answer choice, we can record a binary variable denoting whether the single answer was included or excluded from the participant's multiple answers to the question.

- *Ordinal-Politomous questions* [28] - questions with a set of more than two answers which in themselves contain an ordering among answers. For example, questions with answer choices of the form, strongly disagree/disagree/agree/strongly agree, i.e. degree of agreement, or never/sometimes/often i.e. frequency of performing a task, fall under this category. In these example cases, the answers are ordered based on how much the participants agrees with a statement, in the former case, and how frequently something occurs, in the latter case.

- *Nominal-Politomous questions* [28] - questions with a set of more than two non-ordered answers. For example, "Which of the following location have you visited?", is a question whose answers cannot be ordered in magnitude, it's nonsensical to say that one location is of higher magnitude than the other.

Questions can also differ by range or scale of the answers. This scale comes from the encoding of the question answers. As previously mentioned, usually answers are encode with a number representing the the participants choice from the given answers, each possible answer has a unique number and the range of these numbers, gives the range and scale of the question. For example, dichotomous questions can only be recorded with two numbers, e.g. either with a 0 to 1 range, or 1 to 2 range, while nominal-politomous and ordinal-politomous can have different ranges, for example, the four strongly disagree/disagree/agree/strongly agree answers can be recorded in the range of 1 to 4, and never/sometimes/often can be in the range of 1 to 3. The scale or range of the questions gives us all the possible answers. We consider this range when imputing for missing values in a question, i.e. we draw the replacement values from the question's range.

Another consideration which must be made for our case of imputation is the reason for missingness. The most prominent cause for missingness in survey questionnaires is non-response, more specifically item non-response [16]. Non-response refers to missingness due to a individual not responding to the survey, there are two types of non-response: (i) unit non-response [16], which occurs when an individual selected in the sample refuses to respond the questionnaire in entirety, (ii) item non-response [16] occurs when a participant fail to or refuses to answer a specific subset of the items or questions in the questionnaire. This thesis focuses on imputation where the majority of answers are missing due to item non-response.

From the assumptions for the "process that causes missing data" [34], discussed in Section 2.1 of the previous work, non-response is usually treated as MCAR [7, 26], meaning that the reasons for missingness are considered to be completely due to chance, however, the MAR assumption is also, and maybe evn more [7, 26], appropriate for this type of data [7, 26]. MAR expects the missing data has some patterns, is not equally distributed among all questions and can be imputed using other data from the survey.

3.2 IMPUTATION

The methods for imputation which we will discuss are simple imputation, K-nearest neighbors imputation and model based imputation. In simple imputation when we impute missing answers for a given feature we fill its missing answers with its mean or most frequent value i.e. mode, we do such imputation for each feature, hence completing the data. Simple imputation is a trivial solution which can be beneficial in certain scenarios, it is fully elaborated on in Section 2.1, and referrer the reader for further comments. In this section we will focus on the K nearest neighbors and model-based imputation.

3.2.1 K Nearest Neighbors

K-Nearest Neighbors (KNN) is part of the family of hot-deck or donor based imputation methods [26]. In donor based methods we impute the missing answers for each entry separately, by finding related entries to impute from called donors [26]. K-Nearest Neighbors performs this selection of donors by finding the K closest entries to the entry we are imputing for. This value K is the number of donors we are selecting, and the closest entries are found by (i) first calculating the distance between the entry we are imputing for and all other entries in the data, (ii) then selecting the K entries with the smallest distance to the entry we are imputing for, i.e. the K nearest neighbors of our entry. This distance between entries is derived using a distance metric [44], which calculates the distance between entries based on the values which they have for each feature.

A common distance metric is Euclidean distance which takes the values of each feature as coordinates in a plane and calculates the distance between entry x and y

as [13],

$$Dist(x,y) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2} \tag{3.1}$$

where$d$ denotes the number of features in the data,i.e. its dimension, and x_i and y_i denote the values of the i^{th} coordinate or feature of x and y respectfully.

Once the K nearest neighbors are found, the algorithm needs to choose how to impute from these neighbors, i.e. a resolution needs to be achieved from the donors. For an entry E with a missing value for feature F, we usually impute, by taking either the mean or the mode of F from the K nearest neighbors of E [14]. However, this resolution does not take into account the relevance of each donor for entry E, an alternate methods would be to include the distances between the donors and the entry E in the resolution. In the mode resolution this can be done by giving more precedence , i.e. counting them as more frequent, to the entries which a closer to E. In the case of the mean, this can be done by assigning weights to donors, based on their closeness to the entry E and taking the weighted average, calculated as [14],

$$W = \frac{\sum_{i=1}^{K} w_i X_i}{\sum_{i=1}^{K} w_i} \tag{3.2}$$

where, K denotes the number of values we are averaging, X_i denotes the ith value in the average calculation, and w_i denotes the assigned weight for X_i.

From these methods of resolution, we can see that in the calculation of replacement values for missing data in a variable, we only utilize other values of that specific variable. This is falls under the assumption that the missingness in the specific variable is not related to other observed or unobserved data, hence, KNN falls strictly under the MCAR assumption [26].

3.2.2 Model Based Approach

Methods which utilize machine learning models in their imputation fall under the category of model based imputation [26]. The most often used method from this group is to build one regression model per feature based on the available data in the feature, and afterwards impute missing data using the model [7]. A regression model is any model which predicts a numerical prediction of a target.

A common method for regression modeling is Linear Regression. Linear Regression assumes a linear connection between the dependent variable, the variable we are predicting, and the independent variables, the variables which we use for the prediction [4]. Consequently, it assigns weights The result of this method is a line equation of the form [4],

$$\hat{y} = b + w_1x_1 + w_2x_2 + w_3x_3 + ... + w_nx_n \tag{3.3}$$

where, (i) y denotes the dependent variable and $\hat{y}$ our prediction for it, (ii) b is the line intercept of the line equation (iii) x_1 to x_n are the independent variables and n is the number of independent variables, (iv) w_i is the assigned weight for each x_i, respectively.

The method for selecting the intercept and weight is varied, and the mathematics behind it a beyond the scope of this thesis. We also mention ridge regression as an alternate to linear regression, which only differs in method and not final result.

The one regressor per feature approach imputes the missing data in each feature utilizing solely the values of the other features included in the data. Hence, this method operates under the MAR assumption for missing data.

3.3 COLLABORATIVE FILTERING

Collaborative filtering (CF) is a popular approach for recommendation in Recommender Systems (RS). An overview of collaborative filtering, and the techniques for prediction used in it, matrix completion, are examined in section 2.1 of the related work.

For the purpose of recommendation, collaborative filtering operates on a ratings matrix [1], an example of a ratings matrix is given in Figure 1. The rattings matrix is structured as follows, (i) each row in the matrix represents one user of the recommender system, (ii) each column denotes one item which the system can recommend, (iii) the entry at the i^{th} row and j^{th} column in the matrix contains the i^{th} users ratting for the j^{th} column [1].

As we can see from Figure 1, some of the ratings may be missing. If a user does not have a rating for a specific item, the system assumes that the user is not aware of that item and is a candidate for recommendation. Among these candidates, the system must produce a recommendation catered for the user. To cater this recommendation to the users, the system estimates the ratings which users would give to these items and recommends a number of those with the highest rating. The methods for this prediction are called matrix completion [1].

This thesis examines two matrix completion methods associated with collaborative

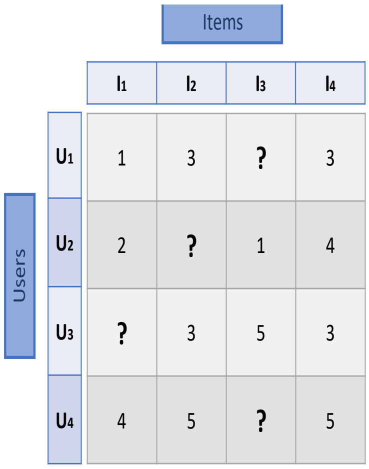

Figure 1: An example of a ratings matrix with four users and four items

filtering, namely, (i) item to item collaborative filtering, (ii) Matrix Factorization.

3.3.1 Item to Item Collaborative Filtering

Item to item collaborative filtering is a neighborhood based method in collaborative filtering [1], which works under the assumption that similar items will be rated similarly by users. Hence, to make a prediction for the rating a user would give to an item , item to item CF looks at the ratings the user has given to other items which are similar to the item being predicted, i.e. that item's neighborhood.

The neighborhood of an item is selected by considering that two items are similar if the rating the users have given to those items are similar [10]. The similarity between two items is measured through a metric, like in the case of k nearest neighbors imputation. The similarity between the given item and all others is measured, and a number, of the most similar items is selected to form the neighborhood. This size of the neighborhood is chosen by the user of the procedure. From this neighborhood, the rating which a user will give the item predicting can be approximated by the average or mode rating the user has given to the similar items. We can see that item to item collaborative filtering is almost identical to k-nearest neighbors imputation where the items are entries and the user's ratings are considered as features.

3.3.2 Matrix Factorization

For the purpose of prediction, matrix factorization (MF) divides and m x n ratings matrix R into a m x k matrix, which we will call U, and an n x k matrix which we, will call V, in the following way [1]:

$$R \approx UV^T \tag{3.4}$$

This new dimension k in U and V is contains the factors of the division [1], or more precisely, the factorization.

The U and V matrices approximate the original matrix and we can use then in the approximation of ratings which are missing from the original matrix [1]. We will further explain this through and example presented in figure 2. The figure show the factorization of a movie ratings matrix, where,(i) ratings are in the form of -1 to 1 where -1 denotes dislike, while 1 denotes liking of an item (ii) the columns are six movies, the first three fall under the genre of action, while the last three fall under the genre of history (iii) the rows denote users, the first three users like action movies, while the last three like history movies, and middle, i.e. 4^{th} user likes both genres. These genres of the movies become apparent when looking at the factors of U and V, and they show a new dimension which can be used to derive the preference of the user. For example in figure 2 we can see that the first user prefers the action over the history genre. In practice, these factors often do not have such nice interpretation such as genre, however, they can reveal underlying properties of the data useful in the prediction [1].

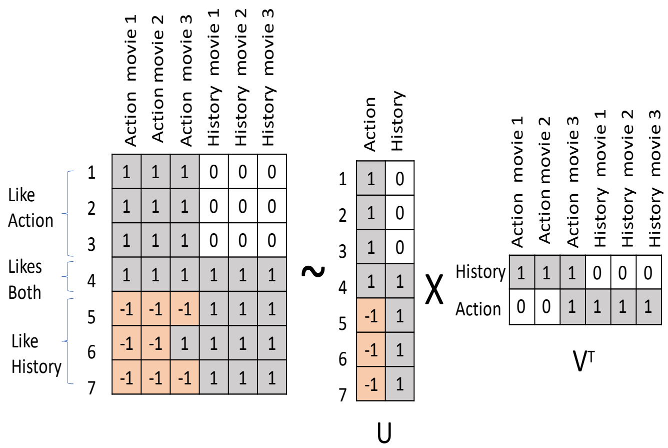

Figure 2: An example of matrix factorization in a movie ratings matrix

To estimate an entry in the original matrix R, i.e. predict a rating a user i would give to an item j, we multiply the row i denoting the user in the matrix U (r.f. Figure 2), and the column j denoting the item in the matrix V^T, i.e. if we see this rating as the entry $r_{i,j}$ in the matrix R, then [1],

$$r_{i,j} \approx \bar{u}_i \cdot \bar{v}_j \tag{3.5}$$

where u_i and v_j denote, the vector from the i^{th} row in matrix U, and v_j denotes the vector of the jth item in matrix V^T. If we view the factors as concepts in the data, we derive [1],

$$r_{i,j} \approx \sum_{s=1}^{k} u_{is} \times v_{js} =$$
$$= \sum_{s=1}^{k} (\text{Affinity of user i to concept s}) \times (\text{Affinity of item j to concept s}). \quad (3.6)$$

In order to derive an approximation of all the ratings in the original matrix we multiply the U and V^T matrices, thus arriving at the final result.

3.4 METRICS FOR EVALUATION

When we evaluate an estimator or predictor, we often compare the estimated or predicted values to the actual values, and derive a numerical indicator if its performance, called an evaluation metric [13]. There are many such metrics, capturing a different aspect of the results, in this section we list some which we will use in our thesis, (i) Mean Absolute Error, (ii) Mean Squared Error, (iii) Accuracy, (iv) Precision, (v) Redcall, (vi) F1 score.

Mean absolute error (MAE) calculates the average deviance from the predicted values and the actual values, as [32],

$$MAE = \frac{1}{n} \sum_{i=0}^{n} |y_i - \hat{y}_i| \quad (3.7)$$

where, n is the number of pairs of predicted and actual values being predicted, y_i is the actual value, and $\hat{y}_i$ is its predicted or estimated value. This metric provides an overall assessment of the error, however it may hide bias.

A similar evaluation metric to MAE is the Mean Squared Error (MSE), given as [32],

$$MSE = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2 \quad (3.8)$$

it captures the variance and the bias of the predictor, and punishes large errors.

The above errors are applicable to regression tasks, for the task of classification how well the predictor divides the data into classes, for each class we can derive, (i) True Positives (TP) rate [22]- the number of instances correctly identified as belonging to the class (ii) False Positives (FP) rate [22] the number of instances incorrectly identified as belonging to the class, (iii) True Negatives (TN) rate [22] - the number of instances correctly identified as not belonging to the class, (iv) False Negatives rate [22] - the number of instances incorrectly identified as not belonging to the class. From these we can obtain a number of important metrics for classification.

One such metric is the accuracy of model, we obtain it by calculating by calculating [22],

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.9}$$

i.e. the number of correctly identified instances, divided by the number of instances. This gives the percentage of correctly classified instances.

In the case where a large majority of the instances belong to one class, a model can always predict this majority class and achieve a high accuracy, therefore a better measure in that case would incorporate how well the data differentiates between classes, such as the Precision [22] and Recall [22] scores, given in equations (3.10) and (3.11), respectively. Precision measures how well the percentage of positive predictions which are correctly identified, while Recall measure the percentage of positive predictions which are correctly predicted over all positive instances.

$$Precision = \frac{TP}{TP + FP} \tag{3.10}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.11}$$

Precision and Recall are combined in the F1 score, defined as the harmonic mean of these two metrics [22], it is defined as,

$$Recall = \frac{2 \cdot Percision \cdot Recall}{Percision + Recall}. \tag{3.12}$$

4 DATA ANALYSIS AND PREPARATION

The World Values Survey (WVS) is an international research program devoted to the scientific and academic study of social, political, economic, religious, and cultural values of people in the world [19]. Seven waves have been conducted by the survey, covering over 120 countries representing 94.5% of the world's population [19]. Data from this survey has seen wide usage across the fields of social science [23, 17, 42, 12], being the subject of thousands of publications [27]. Such a valuable dataset beckons advanced approaches for dealing with its missing answers, and it will prove an ample testing grounds for our methods. In our study we will focus on the most recent, 7th wave, of the survey, conducted in the years 2017-2021. The next section gives an overview of the data.

4.1 OVERVIEW

The WVS's 7th wave dataset is extracted from the survey's website. At the time of writing this thesis it is composed of 84638 entries, one entry per participant, and 563 attributes (features). We summarize the attributes by dividing them into three categories:

- *Technical Variables* - This is data which is provided by the administrator of the survey outside of the participants answers. A total of 36 technical attributes are recorded including location, date, time, duration of the survey; as well as, judgment on the participant's interest in the survey, mode of data collection (phone, internet, in person), participant ID, etc.

- *Question Answers* - 326 attributes which describe the participant's answers to the questionnaire in the form of one attribute per question. The survey uses closed questions, meaning that the participants chooses from a list of answers rather that articulating the answers themselves. For each pair of question attribute and participant (entry) a number is recorded which denotes the participant's choice from the list.

- *Derived and External variables* - The final 185 attributes come from external datasets, or are derivation from the question answers, aimed at aiding inference.

For the purpose of this thesis we will be mainly concerned with the second attribute category - question answers. Technical variables will be examined for their utility in completing question answers data, while Derived and External variables will not be considered, since, they are not the result of the survey's own data collection (observation) process. We investigate the question answers in the coming section.

4.2 QUESTION ANSWERS ANALYSIS AND DATA PREPARATION

The questionnaire of the World Values Survey tries to capture data which draws a global picture of human values across the world. However, it also tries (i) to capture country or region specific data, to allow for closer analysis of interesting regions, (ii) the participants interest and attentiveness though repeated questions. Out of the 326 attributes describing questions answers, 291 capture data from the questionnaire which is comparable globally, comprising of questions which are constant throughout all countries, while 135 attributes bellong to the latter case, i.e. (i) country or region specific questions or (ii) repeated questions.

The set of opinion questions that are present in all variations of the World Values Survey across countries we call the "core questionnaire", our testing will be focused on this data. We outline the rationale, (I) missing answers due to variations of the questionnaire are systematic and do not fall under our MAR (Missing at Random) and MCAR (Missing completely at Random) assumption for non-response. Furthermore, attempting to predict the answers of a question globally, based on the answers to a question in a specific county or region, is prone to heavy bias stemming from the origin of the available answers. While an intra-country approach can be taken for the imputation of such cases, for the purpose of proper testing, evaluation, and explainability of the results, we avoid them. (II) Missing answers in repeated questions, aimed at measuring attentiveness, either contain already available data or their missingness is related to (II.i) other missing data, (II.ii) data which they have failed to observe and remains unobserved i.e. the attentiveness itself. Both of these reason fall under the NMAR assumption, and cannot be imputed from the data at hand.

We continue our analysis focusing on the core questionnaire. It contains 14 modules, each tackling a different aspect a human values. Topics include: (i) ethical values (ii) social values and perceptions (iii) political values and stances on various social and political questions (iv)postmaterialism, (v) etc. A full description of each of the 14 modules is presented in Appendix A.

The last module of the World Values Survey, contains social and demographic data. We will count it as separate from the opinion questions in the other modules, and will not be imputing missing data in it. We will still use the term "core questionnaire" to refer to the core opinion questions, however, we note the exclusion of socio-demographic data.

For considerations of their use in imputation of the core questionnaire, we analyzed the socio-demographic and technical variables (r.f. Section 4.1) , on a sample containing 25% of the data from each country. We measured the pearson correlation between these variables and question answers and found only 17 questions were correlated to these variables(with an absolute Pearson coefficient greater than 0.5). We tested their use on existing approaches, KNN and Regression-based, however we found they did not aid in the imputation, and excluded them. From the demographic data we only keep the country of origin of the participant.

Our analysis now considers only question answer from core questionnaire of the world values survey. Among those questions we find 8 question ranges: "1-2", "1-3","1-4","1-5","1-7","1-8","1-10" and "1-11". The "1-2" range denotes a binary yes/no, incude/exclude choice, we categorize these questions as dichotomous. Upon inspection of the "1-3" range we found that they include an unordered choice between three answers, these fall under the category of nominal-polytomous questions. Finnaly, all other ranges express degerees of agreeableness and frequency of doing a task, these are Ordinal-polytomous answers. A total 17 questions deviated from these established division of ranges or otherwise contained answers which required special consideration, for the sake of simplicity and correctness of the results, we excluded these questions from the data.

We note the "1-11" range, this range contains questions in the with a range of "1-10" for available answers, however the survey administer had the opportunity to record a special answer reserved for when the participant made a very negative comment about the question.

The number of questions in each question answer range is described in Figure 3, without the excluded questions. We can see that almost 100 questions belong to the '1-4' answer range, while there is only one question with a range of "1-7" and one with a range "1-8". We can see split among the ordinal-polytomous and other types of questions, where almost 77% of the questions belong to this type.

A distribution of answers for each question answer range is given by Figure 3. We can see that the answers in the majority "1-4" range are tending towards a normal distribution, while answers in other ranges are leaning left or leaning right.

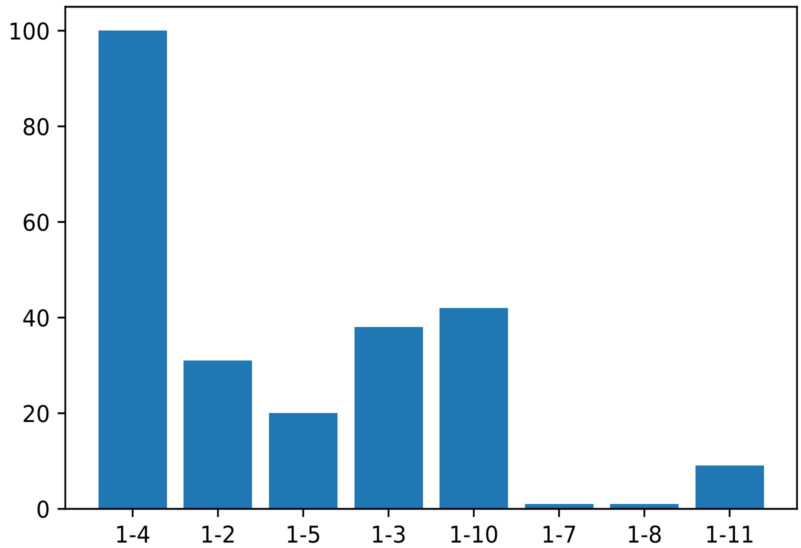

Figure 3: Number of questions in each question answer range

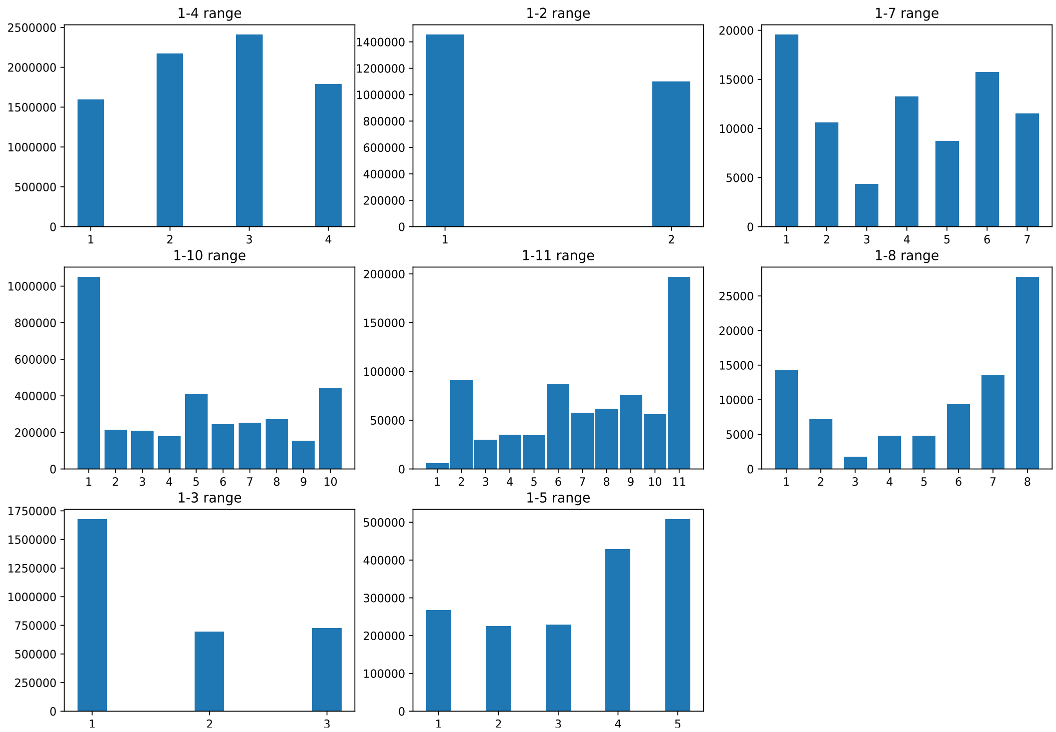

Figure 4: Distribution of answers per question answer range

4.3 OVERVIEW OF COUNTRIES

As mentioned in the related work section, large and complex data such that of the world values survey, often needs to be divided into smaller parts for the purpose of imputation. In our imputation we will use the division of participants into countries of origin.

The 7th wave of the world values survey covers 57 countries with at least one country from each continent. A detailed list of countries is given in Appendix A. The survey has collected at least 1000 participants from each country and at most 3500.

We summarize the relationships between the participants from included countries through clustering presented in Figure 5. Countries in the clustering were represented by their mean answer for each question. In order to be able to visually describe the clusters on a coordinate plane, we performed PCA to reduce the dimensionality of the data into 2 components. The components capture 50.33% of the variability in the data, with the first and second component each describing 38.39% and 11.94% of the variability, respectfully. While this may not paint a full picture, it will give an overview of the countries through clustering. We used hierarchical clustering with the ward linkage method and euclidean distance measure to find the clusters.

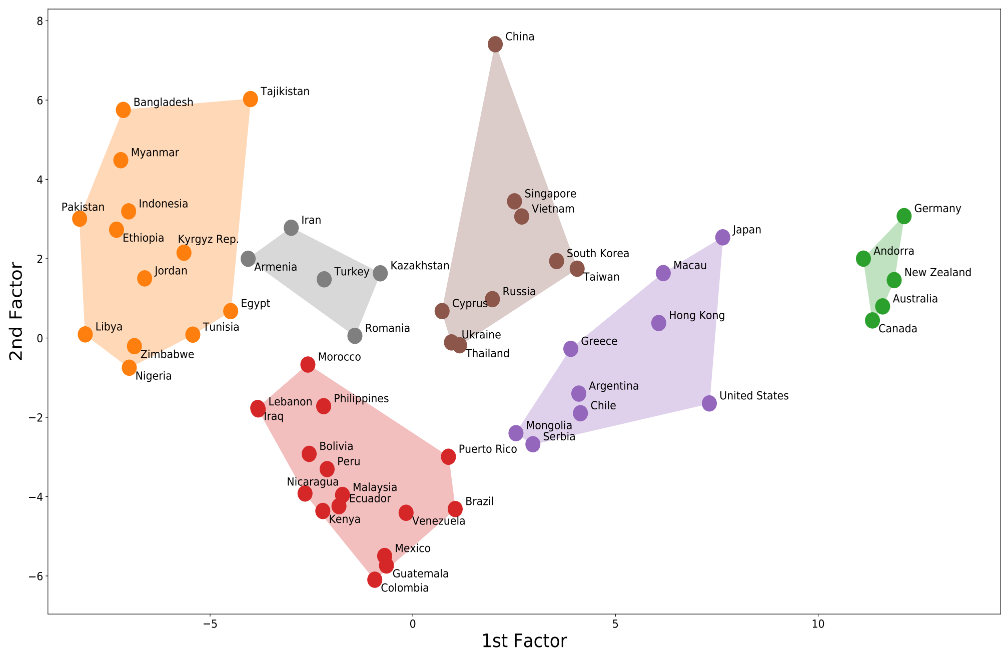

Figure 5: Country clusters based on two factor PCA of the mean WVS questionnaire answers per country

4.4 MISSING ANSWERS

In the core questionnaire data, 7% of the data values are missing. While the percentage may seem small, this still accounts for over 1,5 million answers and 84% of the participants have at least one question for which the answer is unavailable. Consequently, we cannot proceed by deletion in missing data handling and the data values need to be imputed.

Figure 6 describes the missing data in the questionnaire. It is generated by (i) plotting the question on the x-axis (ordered as they appear on the questionnaire), (ii)the participants row index on the y-axis (grouped by country), and(iii) placing a blue dot at the (x, y) position if the participant y's answer to question x is missing.

Such presentation of the missing data (r.f. Figure 6) allows us to see the distribution of missing data across the questionnaire. We first notice that there is missing data scattered across almost all of the questions in the survey. However, we can also see thatin the answers of some participants, missing data tends to be more dense in some questions than others. Questions which tend to have missing data a are grouped by the questionnaire, we can see interesting areas between questions (i) 80 and 90, (ii) 140 and 150, (iii) 200 and 225, where missing answers are more dense than other places.

From this analysis we can infer that, (i) mssingness in some cases is scattered around the questionnaire, hence invoking the MCAR assumption for missing data, however, in many instances we can also see grouping of missing data around particular questions, which falls under MAR assumption for missing data.

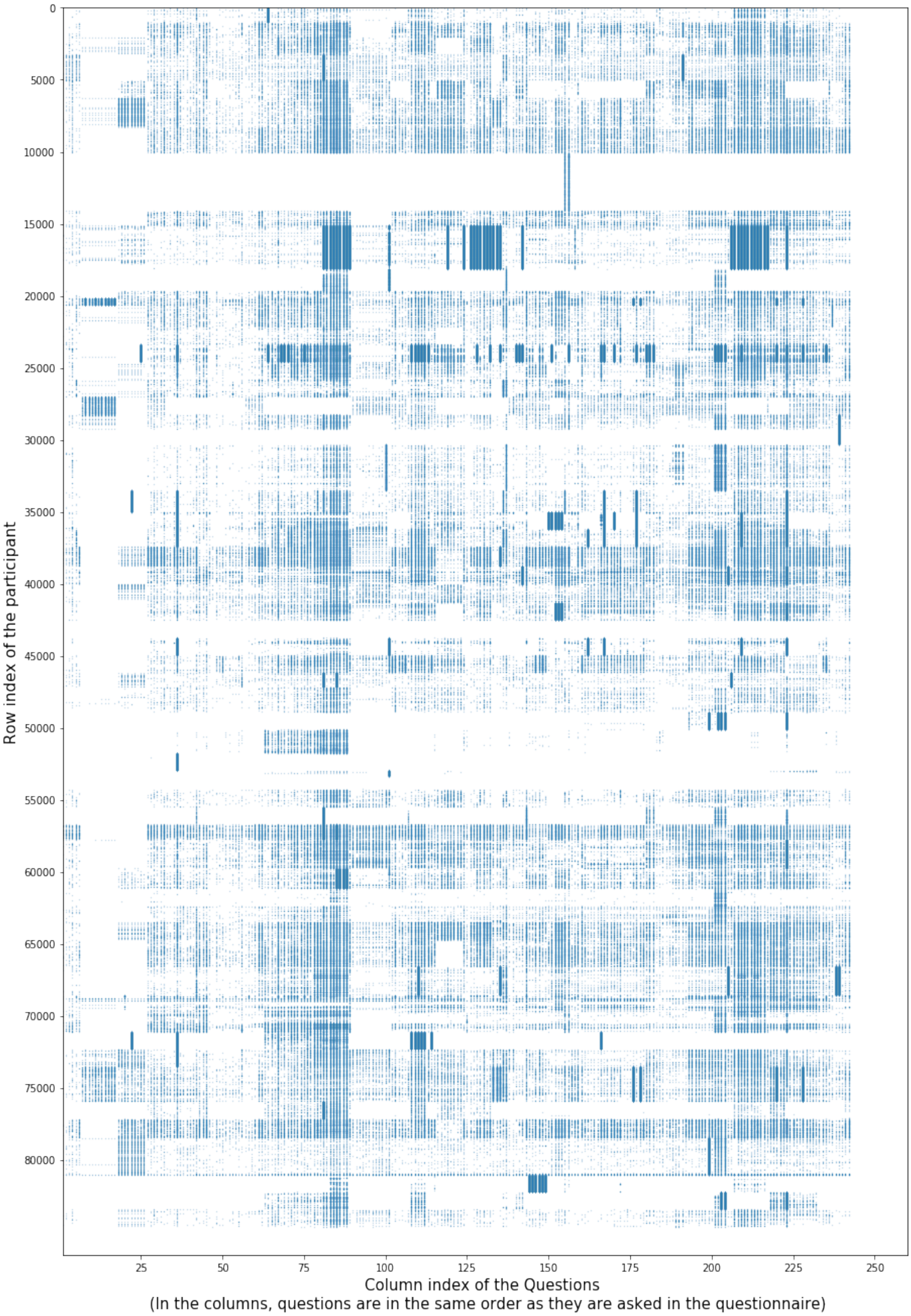

Figure 6: Trace of missing answers in the World Values Survey, generated by (i) plotting the question on the x-axis (ordered as they appear on the questionnaire), (ii)the participants row index on the y-axis, and(iii) placing a blue dot at the (x, y) position if the participant y's answer to question x is missing.

5 METHODOLOGY

5.1 OVERVIEW

This section provides an overview of the methodology of this thesis, while sections 5.2, 5.3, and 5.4, provide details on the techniques tested and the test design. The overall flow of the method is presented in Figure 7.

For the purpose of evaluation and comparison matrix completion and existing methods for survey imputation, we utilize the opinion based core questionnaire data of the world values survey, described in detail Chapter 4. The selected data contains the answers to the set of questions that are present in all variations of the World Values Survey across countries. Moreover in this data we found that, based on the types of question discussed in Chapter 3, Section 3.1, we can divide the questions into three types, namely (i) dichotomous questions, which have question answer range of 1 to 2 in the EN.tex WVS, (ii) nominal-polytomous questions questions, which have question answer range of 1 to 3 in the WVS, (iii) Ordinal-polytomous questions, with a question answer range of 1 to 4 and up (1-5, 1-10, etc.).

Among the three types of survey question answers, we find two imputation tasks, namely, a regression task and a classification task. Ordinal-polytomous answers are handled using regression, while classification is used to handle answers to dichotomous and nominal-polytomous questions. Methods for each task are detailed sections 3.2 and 3.3. We compare the methods for each task by simulating varying degrees of missingness in the data, from 10% to 50% and evaluating their performance in imputing the data. To cater our imputation to the data, hence producing more robust results, we perform the imputations per country.

The consideration for imputing per country comes from the previous work. It suggests that if clusters are apparent in the data, such as those born of demographics, better imputation results are achieved if data is imputed for each cluster separately [24, 36]. In international surveys, clustering participants based on their country of origin, i.e. imputing answers for each country separately, is an often taken and effective approach [41]. We will take advantage of this nature of the data in our predictions. From this point onward, when we talk about imputing the dataset, we will be referring to imputation per country. We note that, since the imputation for a single country is independent from the rest, with this approach we can also easily solve the problem of imputing missing answers with a varied questionnaire across couturiers. However, for the purpose of consistency in the results, we choose to avoid questions varying across countries.

The next three sections elaborate further on the methodology. Sections 4.2, 4.3 present our choice for the imputes for each imputation task, Regression and Classification, respectively, while Sections 4.4 details the evaluation process. Figure 7 gives an overview of the testing pipeline.

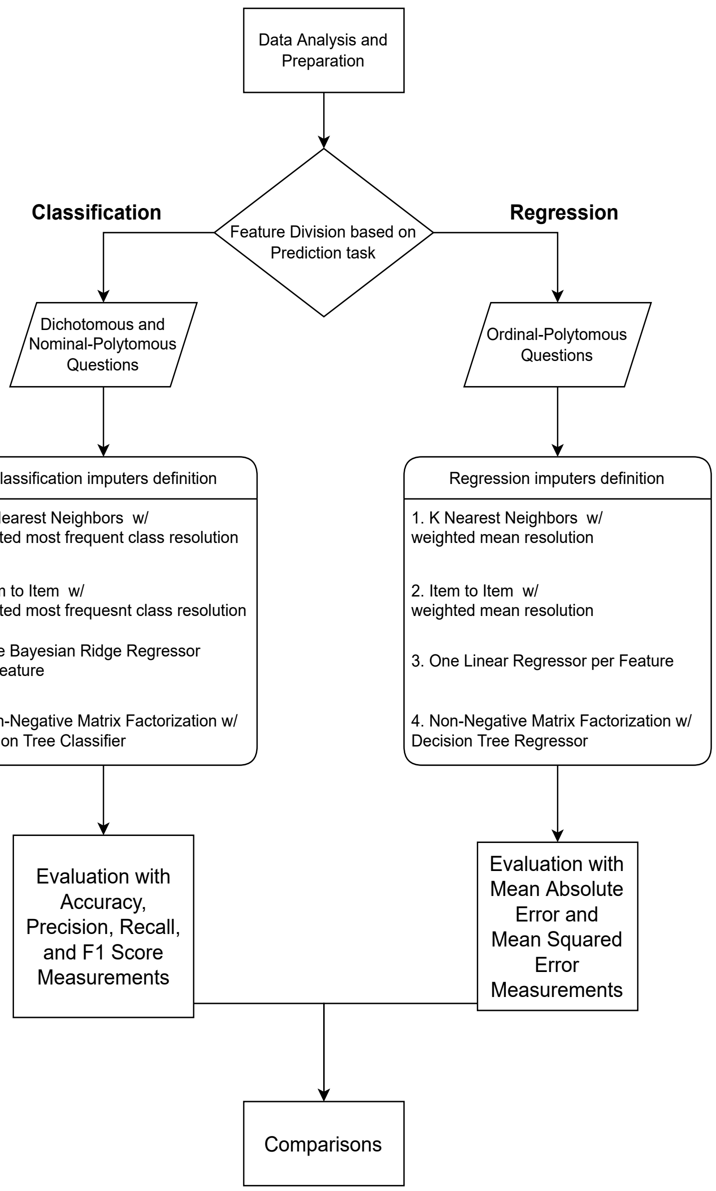

Figure 7: Flow of the imputation and evaluation procedures

5.2 REGRESSION APPROACHES

We compare five approaches for the task of regression. Three approaches originate from imputation techniques, these are: simple mean imputation, k-nearest neighbours (KNN) imputation and imputation using one model per feature (Iterative Imputation). The remaining two methods stem from other research in matrix completion, namely, non-negative matrix factorization (MF) and item to item collaborative filtering.

For this task we use only the set of ordinal-polytomous answers. All values on the data are scaled to the range of 1-3. This emulates the -1 to 1 scale used by previous work to handle the multivariate scenario, but allows for non-negative matrix factorization.

Simple mean imputation fills in the missing answers for each question with the average of the available answers for that specific question. We consider both imputing the mean per country, and imputing the global mean i.e. considering the mean answer in the whole dataset. This is the only point at which we will depart from our per country imputation established in the previous section.

Given a participant P with a missing answer for a specific question Q, K-nearest neighbors (KNN) imputation fills this missing answer in two steps. First, based on the answers to the other questions, it finds the K closest participants to P which have given an answer to Q, afterwards it predicts what participant P's answer to Q based on the answers of these closest participants. Our implementation uses the weighted average of the nearest neighbors, defined in chapter 3 (Theoretical Background). We found that this imputer works best when $K = 64$ and nearest neighbors are found using euclidean distance. We remind the reader that we will be imputing per country, and therefore the KNN algorithm will only consider the closest neighbors residing in the same country as the participant.

Maybe the most refined among the established approaches for imputation is fitting one regressor for each question. Here, an ordinary linear regressor is trained for each question with missing answers, the target is the question, and the independent variables are all other questions. Since the independent variables may also have missing answers, during training they are handled with Mean Imputation so we can fit the regressor. The model for each question is fit on available answers and then the missing answers are predicted from the model. For a lack of better wording, hence forward we will call this model based imputation.

Item to Item collaborative filtering is similar to the K-nearest neighbours imputation, only, the roles of the question and the participant are reversed. For each question, we impute a participant's missing answer by taking the weighted average of his answers from the 7 closest questions to it, based on euclidean distance.

In our matrix factorization (MF) approach, missing answers are substituted by a value of 0, we then perform Non-negative Matrix Factorization minimizing the Frobenious norm using Coordinate Descent with L2 regularization on both the component and the transformation matrices (W and H) using an alpha of 0.9. The MF algorithm initialized with Non-negative Double Singular Value Decomposition. We take the dot product of the resulting matrices to get an estimation of the original matrix. Let Q a column of the original matrix (containing all answers to a given question) and let Q' be the estimation of Q in the resulting matrix from matrix factorization, for each pair of Q and Q' we use the available data in Q to train a Decision Tree Regressor which predicts Q from Q', and use this model to predict the remaining missing answers in Q from Q'.

Hyper parameters for these methods were tuned on a randomly selected subset of the data containing 25% of the samples from each country. For scoring in this parameter selection step, we used the metrics and augmented cross-validation, detailed in section 2.4 (Test Design). In this step, the augmented cross-validation technique had 5 folds.

5.3 CLASSIFICATION APPROACHES

The approaches for classification and regression are similar. We again use the same five methods with a few alterations. The KNN imputer predicts the values of nominal attributes by taking the weighted majority answer the K closest participants, similarly, the Item to Item approach takes the weighted majority answer of K nearest questions. The best performing values for K were 16 and 5, for the KNN and Item to Item approach respectively. The matrix factorization method is refined with a Decision Tree Classifier instead of a regressor, other parameters remain the same. The Regresssion Based approach is adjuster to the case of classification by equipping it with a Bayesian ridge regressor, which predicts the closest whole number to its estimation, denoting the class. Finally, the simple imputer predicts the mode (most frequent) value instead of the mean.

We note the use of a bayesian ridge regressor over other methods, such as a logistic regressor for the task of classification. Logistic regression can only predict binary variables, for the questions on the 1 to 3 scale, this implies the fitting of at least two models in order to differentiate the classes, which increases the complexity of the imputation problem. On a subset of the data containing 25% of the participants from each country, we compared our bayesian ridge regressor approach to the multnomial logistic regressor and found it performs comparable to logistic regression, and is more efficient approach than multnomial logistic regression.

5.4 TEST DESIGN

We evaluate each method by generating missing answers and testing its efficacy in filling them. Different degrees of missingness are tested, from 20% to 50%, and for each we calculate evaluation statistics. For regressors we calculate MAE and MSE, while classifiers are evaluated with their accuracy, precision, recall and F1 scores.

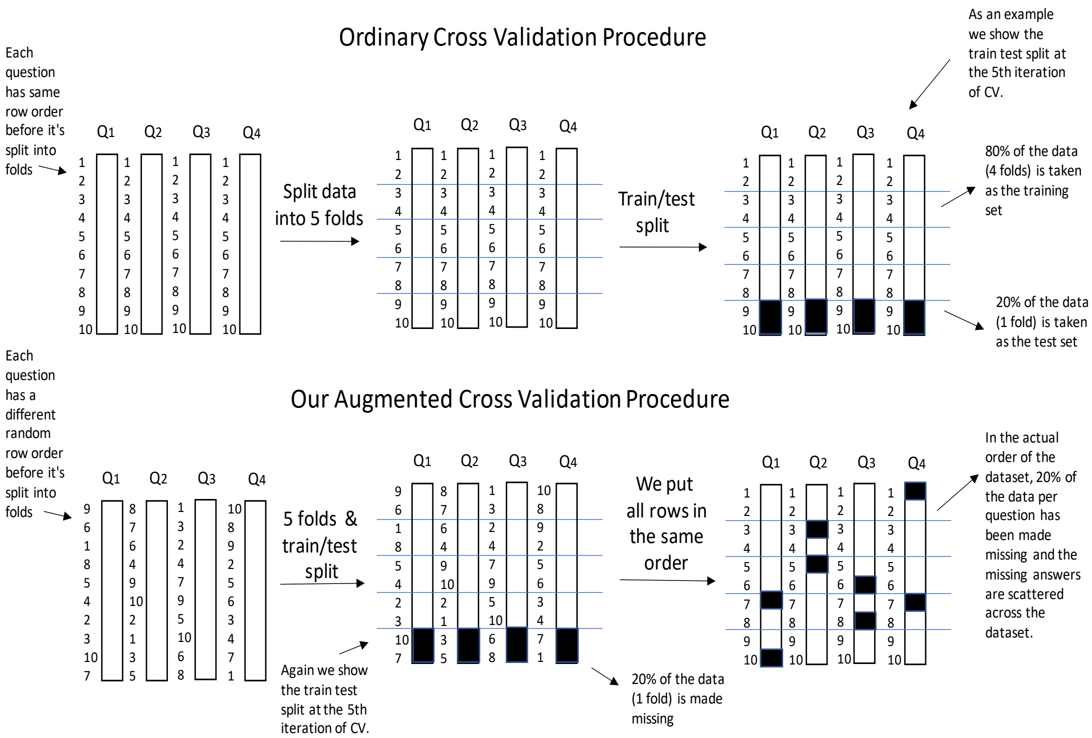

Figure 8: Differences between our augmented cross validation procedure and ordinary cross validation

To generate missing answers we randomly select a subset of the answers and make it missing. A single such selection of missing answers will only show the performance of the methods on a subset of the answers, to evaluate the methods across the whole data we use a method similar to cross validation (CV). Ordinary cross validation divides the entries in the data into folds of equal size, then, for each fold, removes the fold from the data, trains a model on the reaming entries, and tests it on the fold. Thus, the cross validation procedure test a model on the whole dataset. The classical train/test split, however, is not suitable for our case, since, imputation and matrix completion methods operate on the whole data, where some values, but not whole entries like in the train/test split, are missing. Instead making one fold division on the whole data, we adjust the CV method making a unique random division of the rows into folds for each question. Then we iterate through the folds, at each iteration taking on fold from each question as a test set and making it into missing values. We save the values in the test set, impute, and the compare imputed and actual values. Notice that the rows in the test set are different for each question, and this makes the missing answers scattered around the dataset while still having the same % of added missing answers for each feature. A visual explanation of the augmented CV technique is given in Figure 8.

The augmented cross-validation procedure is performed per country, the evaluation statistics for each country are computed by taking average of the evaluation statistics of each fold, i.e. iteration in the CV. The final scores for each method are acquired by taking the average of the scores of its performance across countries. We repeat the procedure for groups of four folds, six folds, eight folds and ten folds, i.e. testing our methods on data with 20%, 30%, 40% and 50% missing answers.

Implementation of the testing procedure and methods tested was conducted using the Python programming language. The methods were implemented using the Scikit Learn library with modifications when needed. The item to item collaborative filtering was implemented by performing KNN imputation on a transposed set, while iterative imputation was implemented using the IterativeImputter class of the package but by only allowing one iteration and removing the additional MICE step. MICE is part of a different family of imputers doing multiple imputation, which are beyond the topic of this thesis. The results of our testing are detailed in the next chapter.

6 RESULTS

This chapter details and discusses the results form our evaluation. We will first present the results from our regression task, and later our results for the classification task. We then continue to discuss the findings and the implications of our results.

6.1 REGRESSION RESULTS

Table 1 shows the Mean Absolute Error and Mean Squared error of our regression approaches, for added missing data magnitudes of 10%, 20% , 30%, 40% and 50%. The final errors are calculated by taking the average across all questions and all countries. We remind the reader that, for the task of regression, all values are scaled from to the 1 to 3 scale (which is similar to the -1 to 1 scale used in related work). We keep this scale in our final results to make sense of average error in the multivariate scenario. The best performing imputer is marked with bold.

We note that the mean squared error is smaller than the mean absolute error. This is due to the fact that the individual errors are smaller than 1 and therefore the squared error is smaller than the absolute error. The small errors are due to the scaling of larger ranges (1-4, 1-10 etc.) into a smaller range (i.e. 1-3).

Overall, the methods are consistent in their evaluation results across the varying degrees of missingness, with only slight increases in errors as the threshold rises. This development for the imputation methods is interesting due to the fact that the related work mentions problems when data is missing above 40%. We except this to be possibly, due to the size of the data. The matrix completion techniques, however, do not overcome the established imputation techniques when performing on high messiness, nor do they show resilience as the number of missing data rises.

We will first compare the models to global (whole dataset) and per country mean baselines, we will call these baseline 1 and baseline 2, respectfully. Mean imputation per country performs better than mean imputation on the whole dataset, we note the possible implication that imputing per country aided our methods. All methods pass baseline 1, and all but the matrix factorization technique pass the per contry base-

Table 1: MAE and MSE scores for each regression imputation method considered, when missing data is above the 10%, 20%, 30%, 40%, and 50% threshold. The errors are presented on scale of 1 to 3 (we would achieve similar answers on the -1 to 1 scale, as well). The best imputer is marked with bold.

Metric per %	Matrix Factorization (w/ Decision Tree)	Item to item CF	KNN	Model Based Imputation	Mean Imputation (whole dataset)	Mean Imputation (per country)
10%						
MAE	0.4075	0.3686	0.3835	**0.3412**	0.5076	0.4488
MSE	0.3655	0.2565	0.24899	**0.2105**	0.3757	0.3171
20%						
MAE	0.4099	0.3739	0.3876	**0.3492**	0.5076	0.4488
MSE	0.3685	0.2614	0.2531	**0.2183**	0.37579	0.3171
30%						
MAE	0.4120	0.3799	0.3925	**0.3581**	0.5077	0.4489
MSE	0.3713	0.2669	0.2581	**0.2275**	0.3759	0.3172
40%						
MAE	0.4142	0.3869	0.3983	**0.3691**	0.5078	0.4489
MSE	0.3741	0.2732	0.2640	**0.2394**	0.3759	0.3172
50%						
MAE	0.4165	0.3955	0.4050	**0.3834**	0.5077	0.4489
MSE	0.3774	0.2806	0.2714	**0.2557**	0.37585	0.31737

line 2. The matrix factorization technique overcomes baseline 1 in the terms of MAE, however, not in MSE statistic. MSE punishes high errors, consequently, this shows possible bias in the matrix factorization technique where some values are being more constantly predicted than others.

The Item to Item, KNN and Regression based methods all perform comparable to eachother, with the regression based approach being the best over all imputer in the case of regression. The Item to item method outperforms KNN in terms of MAE, however, is siligthly worse in terms of MSE, performs compareble in MAE to the regression based method while being slightly more biased in terms of MAE. We also note the similarity in MAE between the Matrix factorization and KNN approach.

While table 1 gives an overview of the results, specific questions may give us further insight into the performance. Table 2 presents the question specific results for each of the methods, in the case of questions Q1, Q39, Q172, Q195, Q253, i.e. the 1st, 39th,172nd, 195th and 253rd question of the world values survey. Since we observed little variability as the % of missing data changes in the case of regression, for the sake of comparison, we present the average result across all thresholds of missingness. Result for all questions are present in Apendix B.

We will refer to questions in the world values survey as 'Q' followed by the number of the question as it appears in the world values survey, e.g. Q5 refers to the 5th question of the survey. Questions Q39, Q172, Q195, Q253 are randomly selected from survey such that they all have different range. Interestingly, while the measurement do differ among these four questions, based on the data at hand, we cannot say that there is high difference among the performance of the imputers across these scales. This is highlighted by Q172 which is the only question in the range of 1 to 8, despite this sparcity, the imputation methods perform similarly on it, as they do in the case of Q195, in the second most frequent 1 to 10 range (r.f Chapter 4, Section 4.2).

Moreover, the matrix completion techniques tend to follow the trends of the established imputation technique in terms of MAE and MSE. There are two deviations in this regard, (i) the first is the matrix factorization handling questions in the 1 to 4 and 1 to 5 range with a much smaller MSE than in the case of the larger ranges such as 1 to 8 and 1 to 10. (ii) the deviation of item to item collaborative filtering in the case of Q1, this question is interesting because most of the participants have given an identical answer, the other methods are capable of identifying this, having very small error, however item to item CF cannot accurately estimate this case since it does not consider the item it is imputing.

Table 2: Average MAE and MSE over all % missing value thresholds of the regression imputation methods in the prediction of the 1st, 39th,172nd, 195th and 253rd question of the world values survey. The names of the imputation methods are abbreviated here, MF, I2I, KNN, MB, Mean, refer to matrix factorization, item to item, K-nearest neighbours, model based imputation, and simple mean imputation per country respectively.

	MF	I2I	KNN	MB	Mean	Range
Q1						1-4
MAE	0.067561	0.246952	0.102909	0.112958	0.116139	
MSE	0.056308	0.109641	0.046125	0.045151	0.049201	
Q39						1-5
MAE	0.349569	0.352861	0.357039	0.338780	0.379611	
MSE	0.276264	0.233756	0.219115	0.202527	0.243970	
Q172						1-8
MAE	0.434384	0.401889	0.408198	0.371339	0.474862	
MSE	0.419840	0.296632	0.284783	0.249140	0.352546	
Q195						1-10
MAE	0.444069	0.398048	0.413786	0.389999	0.483784	
MSE	0.446009	0.325774	0.303147	0.279956	0.372683	
Q253						1-4
MAE	0.386603	0.373819	0.372460	0.347178	0.420266	
MSE	0.329811	0.246285	0.234036	0.215373	0.282917	

6.2 CLASSIFICATION RESULTS

The results for our classification task are given in Table 3. Similarly as in the previous case, the evaluations statistics presented are the average across all questions that fall under this task. We measured the perfomace of each aproach Acuracy, F1 Score, Percision and Recall metrics of each model over a similar The best performing technique for each evalutaion statisti is marked with bold.

From Table 3 we can see that the mode per country is a powerful predictor in the case of the classification task. This implies that the data is unbalanced, hence, the F1 score, Precison, Recall, are better indicators of the performance in this imputation task. Since the F1 score is balanced measure for between the Precison, Recall, we will use it as the prime metric for comparison in the case of classification.

In terms of the F1 score, all techniques beat the mode baseline, except the model based technique which fails at the 50% and 20% thresholds. We suspect that the poor performance of our model based technique is due to the unbalanced nature of the data.

The most accurate method for the task of classification is the KNN approach, however, the less biased F1, Precison, Recall metrics reveal that our matrix factorization technique is more capable than KNN in differentiating between the classes, and performs best in this regard among the techniques considered. However, we can see that its performance dwindles as the percentage of missing data increases.

Item to item collaborative filtering performs comparable to the most accurate KNN technique in terms of F1, however, its performance starts to drop after the 40% threshold.

Similarly as in the case of the regression task, we examine the results more closely by considering the performance of the imputes in per question scale. Table 4 shows the average accuracy, precision, recall, and F1 score over all % missing value thresholds of the regression imputation methods in the prediction of the 144th, 168th,213th and 219th question of the world values survey. The names of the imputation methods are abbreviated in the table similarly as in the case of regression. In the table the questions are presented with a "Q" preceding the number of the question.

From the table we can see variations in performance across questions, and across ranges. However, we also notice the similarity between the vase Q168 and Q213, regardless of the different ranges of these questions, the techniques perform similarly. The reason for these similarity we suspect is in the distribution of the data in these cases, we derive this insight on the distribution, by comparing the performance of the mode prediction in these cases.

On the same range, we can see difference of performance and distribution between the questions of Q144 and Q168, regardless of the same range. Performance of the matrix completion techniques is drastically different between these two cases which might indicate that the distribution has a high effect on the performance of these techniques. Similarly, Q219 is a question in which the matrix completion techniques perform considerably better than other methods.

Table 3: Accuracy, F1, Precision and Recall scores for each classification imputation method considered, when missing data is above the 10%, 20%, 30%, 40%, and 50% threshold. For each score, the best imputer is marked with bold.

Metric per %	Matrix Factorization (w/ Decision Tree)	Item to item CF	KNN	Model Based Imputation	Mode Imputation	Mode Imputation (per country)
10%						
Accuracy	0.7332	0.7082	**0.7505**	0.6368	0.6680	0.7081
F1	**0.5182**	0.4682	0.4832	0.3913	0.3261	0.3508
Precision	**0.5684**	0.5220	0.5344	0.4774	0.2761	0.3073
Recall	**0.5386**	0.4745	0.4991	0.3990	0.4082	0.4236
20%						
Accuracy	0.7205	0.6957	**0.7423**	0.6263	0.6680	0.7080
F1	**0.5020**	0.4401	0.4438	0.3287	0.3261	0.3398
Precision	**0.4984**	0.4603	0.4751	0.4069	0.2761	0.2997
Recall	**0.46677**	0.4151	0.4601	0.3378	0.4082	0.4100
30%						
Accuracy	0.7027	0.6784	**0.7345**	0.6176	0.6680	0.7079
F1	**0.4833**	0.4154	0.4393	0.3572	0.3261	0.3399
Precision	**0.5550**	0.4154	0.5007	0.4606	0.2761	0.2963
Recall	**0.5094**	0.4227	0.4612	0.3655	0.4082	0.4120
40%						
Accuracy	0.6758	0.6521	**0.7263**	0.6081	0.6680	0.7075
F1	**0.4542**	0.3849	0.4223	0.3415	0.3261	0.3386
Precision	**0.5367**	0.4628	0.4801	0.4446	0.2761	0.2950
Recall	**0.4913**	0.3998	0.4483	0.3511	0.4082	0.4109
50%						
Accuracy	0.6379	0.6163	**0.7188**	0.5988	0.6680	0.7073
F1	**0.4148**	0.3494	0.4072	0.3232	0.3261	0.3378
Precision	**0.5144**	0.4393	0.4565	0.4240	0.2761	0.2942
Recall	**0.4700**	0.3782	0.4374	0.3339	0.4082	0.4100

Table 4: Average Accuracy, Precision, Recall, F1 over all % missing value thresholds of the regression imputation methods in the prediction of the 144th, 168th,213th and 219th question of the world values survey. The names of the imputation methods are abbreviated here, MF, I2I, KNN, MB, Mean, refer to matrix factorization, item to item, K-nearest neighbours, model based imputation, and simple mean imputation per country respectively.

	MF	I2I	KNN	MB	Mode	Range
Q144						1-2
Accuracy	0.467432	0.461166	0.686142	0.639724	0.686999	
F1	0.402585	0.379577	0.384966	0.294290	0.251690	
Precision	0.402230	0.374450	0.381296	0.332562	0.363382	
Recall	0.324554	0.302061	0.342476	0.284429	0.294887	
Q168						1-2
Accuracy	0.741049	0.717415	0.783929	0.663262	0.719507	
F1	0.659818	0.543163	0.653794	0.514006	0.342173	
Precision	0.605733	0.466552	0.572311	0.355735	0.473222	
Recall	0.575190	0.463917	0.559478	0.390361	0.392767	
Q213						1-3
Accuracy	0.745339	0.700641	0.782495	0.602878	0.768492	
F1	0.516015	0.386337	0.535333	0.363247	0.371725	
Precision	0.497130	0.347180	0.510731	0.290596	0.478244	
Recall	0.454327	0.330771	0.482402	0.293094	0.415305	
Q219						1-3
Accuracy	0.816393	0.813389	0.795623	0.774574	0.783028	
F1	0.611859	0.597691	0.418169	0.540050	0.283038	
Precision	0.519637	0.504412	0.388966	0.480126	0.359657	
Recall	0.523108	0.510183	0.368953	0.464367	0.313945	

6.3 DISCUSSION

In our results we found that for the case of regression matrix completion techniques perform comparable but not better than existing approaches in imputation. Of note is the item to item collaborative technique, in this regard. On scaled data, item to item CF managed to achieve similar MAE and MSE as the other approaches, while considering a completely differed aspect of the data. Moreover, in the case of regression we found that the mean is a powerful predictor, this favors the Regression based method which relies on mean imputation as staring point in its procedure, and might be the reason for its high performance in the imputation of data of this kind. On data in which this initial imputation is enviable or the building of one model per feature is expensive, the simpler item to item method might prove useful and efficient, while still considering the same MAR assumption as the regression based methods.

In the case of classification, we found the matrix factorization refined by a decision tree classifier proved to be less biased predictor when compared to other methods in the case of unbalanced data. This implies that it might be useful to consider this method for imputing values of underrepresented data in the case of the survey. The downsides of our matrix factorization approach on the other hand was its declining performance as the magnitude of missing data increased. This was contrary to our expectations of the matrix factorization technique and we expect that it is due to the decision tree classifier, as the number of missing values increases, the number of samples the decision tree can learn from decreases, which might lead to less accurate predictions. Nonetheless, the matrix factorization technique can be fitted with a plethora of models to refine its results, and it might prove beneficial in future work.

Of note are also the assumptions taken for these imputation techniques. In the case of regression, the best performing method is one under the MAR assumptions, while in the case of classification the KNN approach operating under the MCAR assumption is the most accurate. Similarly, matrix completion performs better in the case where an MCAR method is successfully and for item to item the opposite is true. It could be that the differences in performance of these methods between tasks is caused by the processes which cause missing data, rather than their predictive power. This implication about the MCAR assumption in matrix factorisation and its poor performance where a MAR method is successful challenges previous work which showed that matrix factorization can perform well in both scenarios. Further research is required on this subject, which might focus on unearthing the mechanisms of missinness under which matrix completion techniques stemming from collaborative filtering can be of use.

In comparison with other related work, we achieve similar results for MAE in terms of matrix completion to Vozalis et. al. [40], his MAE of 0.846 on unvarite 1 to 5 data 20% missing answer is comparable to our 0.40 MAE on the scale of 1 to 3, achieved under multivariate data. This raises the question of whether the scale effects the matrix completion techniques, collaborative filtering techniques in recomender systems usually operate on ratings all on the same scale. Can alterations of this techniques to fit multivarite data, be more beneficial in future work in survey imputation?

Furthermore, our MAE results perform better than those of Zhang et al. [43], which working on data with similar nature as ours. This may be due to their considerations of an alternate mthod for survey sampling, while it also may be due to (i) the higher volume of the data in our case, (ii) our approach to impute per cluster rather on the data a whole (iii) our use of the the decision tree regressor to refine the imputation. The implications here are that our alternate methods may be useful for techniques in adaptive survey design, as well.

7 CONCLUSION

In this thesis we examined the use of matrix completion techniques stemming from collaborative filtering for the purpose of imputing missing survey data, with the goal of overcoming the shortcomings of existing imputation techniques. For this purpose we compared two matrix completion techniques, namely, matrix factorization refined by a model, and item to item collaborative, to the established techniques of K nearest neighbours and regression based imputation, taking simple imputation as baselines. The comparison of these methods we performed utilizing the World Values Survey dataset, valuable dataset of high volume and veracity, which allowed us to compare these methods in the tasks of imputation under regression and classification.

We have shown that item to item collaborative filtering performs comparable to the KNN technique in both imputation tasks, only failing to match it on high ratios of missing data in the classification case. On the other hand, item to item fails to compare to model based imputation in the regression, however, performs better than it in the classification task. Moreover, we showed that the matrix factorization technique offers poor results in terms of MSE in the regression case, failing to match both exiting imputation techniques, however, in case of classification it outperforms all techniques tested with its F1 performance on unbalanced data. Finally, we demonstrate that as the ratio of missing data increases, the performance of all techniques considered decreased at a similar rate in the case of regression, however, in the classification case, as the ratio of missing data increase, we saw that, matrix completion techniques deteriorate at a more rapid pace then existing imputation techniques.

Future work on this subject should consider these techniques on a wider range of data and different scenarios, examining the effects that these techniques have on the statistical inference. Moreover, our evaluation included only simple techniques of matrix completion stemming from collaborative filtering, the study of collaborative filtering is vast and has many more advanced techniques utilized in recommender systems, other techniques might succeed where we have failed, and the considerations of such techniques in the study imputation may prove fruitful in future work.

8 DALJŠI POVZETEK V SLOVENSKEM JEZIKU

Na številnih področjih znanosti, zlasti na področju družboslovja, so vprašalniki bistveno orodje za zbiranje podatkov. Vprašalnik je instrument za merjenje spremenljivk z nizom vprašanj ali pozivov. Postopek zbiranja podatkov s pomočjo vprašalnika imenujemo raziskava. Ankete imajo prednosti, kot so pridobivanje podatkov neposredno od udeležencev, vendar pa pogosto prinašajo manjkajoče podatke - podatki vrednosti, ki niso opazovane, imenovane tudi prazni podatki. Ta prazna mesta so lahko posledica več razlogov: (i) administratorji raziskave lahko naredijo napako, (ii) udeleženci (iii) razlogi, na katere administratorji ali udeleženci ne morejo vplivati, npr. naravne nesreče. Ne glede na vzrok, manjkajoče vrednosti v podatkih, pridobljenih z vprašalnikom, je treba obravnavati, preden raziskovalci lahko raziskovalci na podlagi podatkov sklepajo.

Običajen pristop k obravnavi manjkajočih vrednosti je brisanje vseh vnosov, ki jih vsebujejo. Prednost tega izbrisa je njegova preprostost, vendar je zaradi njega treba raziskovalca, da deluje na delnem naboru podatkov, kar lahko prinese zavajajoče rezultate. Za delovanje na celotnih podatkih je treba manjkajoče vrednosti zapolniti z nadomestnimi vrednostmi. Ta postopek se imenuje imputiranje. Imputiranje tehnike za obdelavo manjkajočih podatkov v primeru vprašalnika se ne razlikujejo od teh pri kot v drugih primerih. Pogosto uporabljene tehnike za imputacijo vključujejo: (i) preprosto imputacijo primerih, ki nadomesti manjkajoče podatke v spremenljivki z njenim povprečjem ali najpogostejšim vrednostjo (ii) imputacija z vročim nizom, ki izkorišča podobnosti med vnosi v podatkih za iskanje ustreznih zamenjav (iii) pristopi na podlagi modela, ki modelirajo vsako spremenljivko na podlagi razpoložljivih podatkov in dopolnijo manjkajoče vrednosti z uporabo modela za vsako spremenljivko. Vse te tehnike za imputacijo imajo svoje prednosti in ki bodo podrobneje obravnavane v poglavju o sorodnem delu.

Enotna prednost tehnik imputiranja je, da raziskovalcem omogočajo uporabo večjega vzorca podatkov, pri čemer so do neke mere razumne ocene manjkajočih vrednosti. Vendar obstajajo tudi slabosti: (i) metode imputiranja lahko v podatke vnesejo veliko šuma, kar vpliva na sklepanje na podlagi takih podatkov. (ii) v primeru raziskave tehnike imputacije pogosto niso učinkovite pri podatkih z ko manjka več kot 40 %

podatkov.

V našem delu obravnavamo ti vprašanji (i) uvajanja šuma in (ii) imputiranja anketnih podatkov z veliko pogrešanostjo, tako da ocenjujemo, ali je uporaba alternativnih imputacij metod, ki se običajno uporabljajo v priporočilnih sistemih, lahko prinesejo natančnejše pripisovanje manjkajočih vrednosti, tako v primeru nizke kot visoke pogrešanosti. Razlog za uporabo teh metod je (i) podobnost okvira problema med dvema metodama, vprašalniki in priporočilnimi sistemi ter (ii) učinkovitost teh metod pri v priporočilnih sistemih, zlasti v primeru, ko večina podatkov ni na voljo (iii) podobnost med pristopi pri pripisovanju podatkov v anketah in nekaterimi pristopi v priporočilnem sistemu.

V ta namen smo primerjali dve tehniki dopolnjevanja matrik, in sicer faktorizacijo matrik, izboljšano z modelom, in sodelovanje med postavkami, z uveljavljenima tehnikama K najbližjih sosedov in imputiranja na podlagi regresije, pri čemer smo za izhodiščno tehniko vzeli preprosto imputiranje. Primerjavo teh metod smo izvedli z uporabo podatkovne zbirke World Values Survey, dragocene podatkovne zbirke velikega obsega in verodostojnosti, ki nam je omogočila primerjavo teh metod pri nalogah imputiranja na podlagi regresije in klasifikacije.

Pokazali smo, da je sodelovalno filtriranje od elementa do elementa primerljivo s tehniko KNN pri obeh nalogah imputiranja, le pri klasifikaciji se ji ne more primerjati z visokimi deleži manjkajočih podatkov. Po drugi strani pa se metoda od elementa do elementa ne more primerjati z imputiranjem na podlagi modela pri regresiji, vendar se bolje od nje obnese pri nalogi klasifikacije. Poleg tega smo pokazali, da tehnika faktorizacije matrik ponuja slabe rezultate v smislu MSE v primeru regresije in se ne more primerjati z obema obstoječima tehnikama imputiranja, vendar v primeru klasifikacije prekaša vse preizkušene tehnike z uspešnostjo F1 pri neuravnoteženih podatkih. Nazadnje smo dokazali, da se s povečevanjem deleža manjkajočih podatkov uspešnost vseh obravnavanih tehnik se v primeru regresije zmanjšuje podobno hitro, v primeru klasifikacije pa smo videli, da se s povečevanjem razmerja manjkajočih podatkov tehnike dopolnjevanja matrik slabšajo hitreje kot obstoječe tehnike imputiranja.

Prihodnje delo na tem področju bi moralo te tehnike preučiti na širšem naboru podatkov in različnih scenarijev ter preučiti učinke teh tehnik na statistično sklepanje. Poleg tega je naša ocena vključevala le preproste tehnike dopolnjevanja matrik, ki izhajajo iz sodelovalnega filtriranja, študij sodelovalnega filtriranja je obsežen in ima veliko naprednejših tehnik, ki se uporabljajo v priporočilnih sistemih, druge tehnike bi lahko uspele tam, kjer nam ni uspelo.

REFERENCES

[1] C. C. Aggarwal. *Recommender Systems*. Springer International Publishing, 2016. ISBN: 978-3-319-29657-9. DOI: 10.1007/978-3-319-29659-3.

[2] R. M. Bell and C. Koren Y.and Volinsky. “All Together Now: A Perspective on the Netflix Prize”. In: *CHANCE* 23.1 (2010), pp. 24–29. ISSN: 0933-2480, 1867-2280. DOI: 10.1080/09332480.2010.10739787.

[3] R. M. Bell, Y. Koren, and C. Volinski. “The BellKor solution to the Netflix Prize”. In: *AT&T Labs – Research* (2007).

[4] C. M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. New York: Springer, 2006. 738 pp. ISBN: 978-0-387-31073-2.

[5] J. Bobadilla et al. “Recommender systems survey”. In: *Knowledge-Based Systems* 46 (2013), pp. 109–132. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2013.03.012.

[6] JM Brick and G. Kalton. “Handling missing data in survey research”. In: *Stat Methods Med Res* 5.3 (1996). Publisher: SAGE Publications Ltd STM, pp. 215–238. ISSN: 0962-2802. DOI: 10.1177/096228029600500302.

[7] S. Buuren. *Flexible Imputation of Missing Data*. 2nd ed. New York: Chapman and Hall/CRC, 2018. ISBN: 978-0-429-49225-9. DOI: 10.1201/9780429492259.

[8] G. Chauvet and W. D. Paco. *Exact balanced random imputation for sample survey data*. Number: arXiv:1608.04867. 2017. arXiv: 1608.04867[stat].

[9] J. Chen, J. N. K. Rao, and R. R. Sitter. “EFFICIENT RANDOM IMPUTATION FOR MISSING DATA IN COMPLEX SURVEYS”. In: (2000), p. 17.

[10] R. Chen et al. “A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks”. In: *IEEE Access* 6 (2018). Conference Name: IEEE Access, pp. 64301–64320. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2877208.

[11] E. C. Chi and T. Li. “Matrix completion from a computational statistics perspective”. In: *WIREs Comp Stat* 11.5 (2019). ISSN: 1939-5108, 1939-0068. DOI: 10.1002/wics.1469.

[12] M. Dakhli and D. De Clercq. "Human capital, social capital, and innovation: A multi-country study". In: *Entrepreneurship and Regional Development* 16.2 (2004), pp. 107–128. ISSN: 0898-5626. DOI: 10.1080/08985620410001677835.

[13] P. A. Flach. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge ; New York: Cambridge University Press, 2012. ISBN: 978-1-107-09639-4.

[14] D. Forsyth. *Applied machine learning*. Cham, Switzerland: Springer, 2019. 494 pp. ISBN: 978-3-030-18113-0.

[15] P. Giles. "A model for generalized edit and imputation of survey data". In: *Canadian Journal of Statistics* 16.1 (1988), pp. 57–73. ISSN: 0319-5724. DOI: 10.2307/3315216.

[16] R. M. Groves, ed. *Survey methodology*. 2nd ed. Wiley series in survey methodology. Wiley, 2009. 461 pp. ISBN: 978-0-470-46546-2.

[17] L. Guiso, P. Sapienza, and L. Zingales. "People's opium? Religion and economic attitudes". In: *Journal of Monetary Economics* 50.1 (2003), pp. 225–282. ISSN: 0304-3932. DOI: 10.1016/S0304-3932(02)00202-7.

[18] N. Gupta K.and Goyal and H. Khatter. "Optimal reduction of noise in image processing using collaborative inpainting filtering with Pillar K-Mean clustering". In: *The Imaging Science Journal* 67.2 (2019), pp. 100–114. ISSN: 1368-2199, 1743-131X. DOI: 10.1080/13682199.2018.1560958.

[19] C. Haerpfer et al. *World Values Survey Wave 7 (2017-2022) Cross-National Data-Set*. In collab. with K. Kizilova et al. Version Number: 4.0.0 Type: dataset. 2022. DOI: 10.14281/18241.18.

[20] F. Hao and R. H. Blair. "A comparative study: Classification vs. user-based collaborative filtering for clinical prediction". In: *BMC Medical Research Methodology* 16.1 (2016). ISSN: 1471-2288. DOI: 10.1186/s12874-016-0261-9.

[21] D. F. Heitjan and S. Basu. "Distinguishing "Missing at Random" and "Missing Completely at Random"". In: *The American Statistician* 50.3 (1996). Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 207–213. ISSN: 0003-1305. DOI: 10.2307/2684656.

[22] M. Hossin and M. N. Sulaiman. "A Review on Evaluation Metrics for Data Classification Evaluations". In: *IJDKP* 5.2 (2015), pp. 01–11. DOI: 10.5121/ijdkp.2015.5201.

[23] R.T. Howell and C.J. Howell. "The Relation of Economic Status to Subjective Well-Being in Developing Countries: A Meta-Analysis". In: *Psychological Bulletin* 134.4 (2008), pp. 536–560. ISSN: 0033-2909. DOI: 10.1037/0033-2909.134.4.536.

[24] N. Karmitsa et al. "Missing Value Imputation via Clusterwise Linear Regression". In: *IEEE Transactions on Knowledge and Data Engineering* 34.4 (2022). Conference Name: IEEE Transactions on Knowledge and Data Engineering, pp. 1889–1901. ISSN: 1558-2191. DOI: 10.1109/TKDE.2020.3001694.

[25] L. Li et al. "Missing value imputation for traffic-related time series data based on a multi-view learning method". In: *IEEE Transactions on Intelligent Transportation Systems* 20.8 (2019), pp. 2933–2943. ISSN: 1524-9050. DOI: 10.1109/TITS.2018.2869768.

[26] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. 3rd. John Wiley & Sons, 2019.

[27] S. G. Ludeke and E. G. Larsen. "Problems with the Big Five assessment in the World Values Survey". In: *Personality and Individual Differences* 112 (2017), pp. 103–105. ISSN: 0191-8869. DOI: 10.1016/j.paid.2017.02.042.

[28] G. J. Mellenbergh. "Conceptual Notes on Models for Discrete Polytomous Item Responses". In: *Applied Psychological Measurement* 19.1 (Mar. 1995), pp. 91–100. ISSN: 0146-6216, 1552-3497. DOI: 10.1177/014662169501900110. URL: http://journals.sagepub.com/doi/10.1177/014662169501900110 (visited on 08/13/2022).

[29] A. Mirzaei et al. "Missing data in surveys: Key concepts, approaches, and applications". In: *Research in Social and Administrative Pharmacy* 18.2 (2022), pp. 2308–2316. ISSN: 15517411. DOI: 10.1016/j.sapharm.2021.03.009.

[30] M. Oliveira, M. Delgado, and R. Lüders. "Comparative Analysis of Collaborative Filtering-Based Predictors of Scores in Surveys of a Large Company". In: *Encontro Nacional de Inteligência Artificial e Computacional*. Encontro Nacional de Inteligência Artificial e Computacional. Vol. 18. Encontro Nacional de Inteligência Artificial e Computacional, 2021, pp. 739–750. DOI: 10.5753/eniac.2021.18299.

[31] S. Rana, A. H. John, and H. Midi. "Robust regression imputation for analyzing missing data". In: 2012 IEEE International Conference on Statistics in Science, Business and Engineering ICSSBE. 2012. DOI: 10.1109/ICSSBE.2012.6396621.

[32] "Regression model evaluation". In: P. Cichosz. *Data Mining Algorithms*. John Wiley & Sons, Ltd, 2015, pp. 295–310. ISBN: 978-1-118-95095-1. DOI: 10.1002/9781118950951.ch10.

[33] R. M. Rodríguez et al. "Using collaborative filtering for dealing with missing values in nuclear safeguards evaluation". In: *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems* 18.4 (2010), pp. 431–449. ISSN: 0218-4885. DOI: 10.1142/S0218488510006635.

[34] D. B. Rubin. "Inference and Missing Data". In: *Biometrika* 63.3 (1976). Publisher: [Oxford University Press, Biometrika Trust], pp. 581–592. ISSN: 0006-3444. DOI: 10.2307/2335739.

[35] S. Saha et al. "Missing value imputation in DNA microarray gene expression data: A comparative study of an improved collaborative filtering method with decision tree based approach". In: *International Journal of Computational Science and Engineering* 18.2 (2019), pp. 130–139. ISSN: 1742-7185. DOI: 10.1504/IJCSE.2019.097954.

[36] J. Shao and H. Wang. "Sample Correlation Coefficients Based on Survey Data Under Regression Imputation". In: *Journal of the American Statistical Association* 97.458 (2002), pp. 544–552. DOI: 10.1198/016214502760047078.

[37] G. N. Singh et al. "Some imputation methods for missing data in sample surveys". In: *Hacettepe Journal of Mathematics and Statistics* 45.6 (2016), pp. 1865–1880. ISSN: 2651-477X. DOI: 10.15672/HJMS.20159714095.

[38] X. Su and T. M. Khoshgoftaar. "A Survey of Collaborative Filtering Techniques". In: *Advances in Artificial Intelligence* 2009 (2009), pp. 1–19. ISSN: 1687-7470, 1687-7489. DOI: 10.1155/2009/421425.

[39] Gary R. VandenBos and American Psychological Association, eds. *APA dictionary of psychology*. 1st ed. Washington, DC: American Psychological Association, 2007. 1024 pp. ISBN: 978-1-59147-380-0.

[40] M. Vozalis, S. Basbas, and I. Politis. "Applying Collaborative Filtering Techniques In Transportation Surveys". In: *1st International Conference on Engineering and Applied Sciences Optimization*. 2014, pp. 1630–1638.

[41] M. Weber and M. Denk. "Imputation of Cross-Country Time Series: Techniques and Evaluation". In: (2010).

[42] C. Welzel, R. Inglehart, and H.-D. Klingemann. "The theory of human development: A cross-cultural analysis". In: *European Journal of Political Research* 42.3 (2003), pp. 341–379. ISSN: 0304-4130. DOI: 10.1111/1475-6765.00086.

[43] C. Zhang et al. *Active Matrix Factorization for Surveys*. 2019. DOI: 10.48550/arXiv.1902.07634.

[44] Z. Zhou. *Machine learning*. Singapore: Springer, 2021. ISBN: 9789811519673.

Appendices

APPENDIX A Appendix to the analysis of the World Values Survey

A.1 Countries included in the WVS

The following 57 countries are included in the 7th wave of the World Values Survey: Andorra, Argentina, Armenia, Australia, Bangladesh, Bolivia, Brazil, Canada, Chile, China, Colombia, Cyprus, Ecuador, Egypt, Ethiopia, Germany, Greece, Guatemala, Hong Kong, Indonesia, Iran, Iraq, Japan, Jordan, Kazakhstan, Kenya, Kyrgyz Republic, Lebanon, Libya, Macau, Malaysia, Mexico, Mongolia, Morocco, Myanmar, New Zealand, Nicaragua, Nigeria, Pakistan, Peru, Philippines, Puerto Rico, Romania, Russia, Serbia, Singapore, South Korea, Taiwan, Tajikistan, Thailand, Tunisia, Turkey, Ukraine, United States, Venezuela, Vietnam, Zimbabwe.

A.2 Contents of the WVS

The core questionnaire of the WVS is divided into 14 modules, ordered as they appear in the survey:

- Social Values, Norms, Stereotypes - 45 questions including questions regarding, (i) the participant life goals and priorities, and prospects (ii) important human and child qualities such as obedience, determination, etc. and (iii) perceptions towards marginal communities.

- Happiness and Wellbeing - 11 questions regarding the participants, perceptions of happiness, health and financial stability.

- Social Capital, Trust and Organizational Membership - 49 questions regarding the participant's involvement in society and societal organization, as well as the participant's trust in others.

- Economic Values - 6 questions regarding the participant's opinions regarding capital and economic equality .

- Perceptions of Corruption - 9 questions regarding the participants view on the presence of corruption in governmental and non-governmental institutions in his country.
- Perceptions of Migration - 10 questions regarding the participants opinion on migration and the positive and negative sides of it.
- Perceptions of Security - 21 questions regarding the participant's feeling of security.
- Index of Postmaterialism - 6 questions regarding what the regards places as important societal issues.
- Perceptions about Science and Technology - 6 questions, regarding the participants attitude towards science and technology
- Religious Values - 12 questions regarding the participant's beliefs, faith and participation in religious activities.
- Ethical Values - 23 questions on ethical and moral issues.
- Political Interest and Political Participation - 36 questions regarding the participant's activity and interest in politics.
- Political Culture and Political Regimes - 25 questions regarding the participant's political ideology and opinion about political issues.
- Demographic and Socioeconomic Variables - 31 questions regarding socio-demographic data.

A.3 Questions excluded from our evaluation

Some opinion questions were excluded from our analysis due to them, (i) deviating from the established question answer patterns in the WVS, (ii) or otherwise required special attention which might compromise our evaluation results. A total of 18 questions fall under this category. We list them using the coding "Q", followed by the number of the question as they are referenced in the WVS [19] :
"Q223", "Q56", "Q91", "Q92", "Q93", "Q111", "Q119", "Q149", "Q150", "Q174", "Q175", "Q223", "Q152", "Q153", "Q154", "Q155" ,"Q156" ,"Q157".

APPENDIX B Test Results per Question

This appendix shows the test result per question. The results shown are the average over each country and missing data % for each evaluation metric. The names of the imputation methods are abbreviated here, MAT, I2I, KNN, MB, Mean, refer to matrix factorization, item to item, K-nearest neighbours, model based imputation, and simple mean imputation per country respectively.

B.1 Regression task

This section lists the results per questions falling under the regression imputation task.
Q1 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.067561	0.246952	0.102909	0.112958	0.116139
MSE	0.056308	0.109641	0.046125	0.045151	0.049201

Q2 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.365643	0.383907	0.378665	0.364929	0.398669
MSE	0.287673	0.244051	0.213779	0.204313	0.229614

Q3 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.383674	0.408826	0.392727	0.388027	0.414070
MSE	0.303660	0.271645	0.238495	0.232700	0.255592

Q4 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.498760	0.494359	0.493248	0.456869	0.532394
MSE	0.450575	0.380633	0.355614	0.320787	0.402658

Q5 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.296079	0.341230	0.336147	0.326849	0.355663
MSE	0.270269	0.229564	0.202794	0.195011	0.220821

Q6 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.355562	0.336717	0.339191	0.307407	0.397211
MSE	0.334674	0.212331	0.212488	0.181688	0.277906

Q27 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.284736	0.317554	0.312150	0.306333	0.333606
MSE	0.225067	0.184133	0.162908	0.157492	0.181704

Q28 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.424506	0.438611	0.423244	0.415561	0.450672
MSE	0.364546	0.314248	0.284207	0.275548	0.315357

Q29 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.428815	0.396092	0.416459	0.376686	0.455397
MSE	0.369322	0.264845	0.275443	0.237666	0.318843

Q30 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.406847	0.398188	0.391414	0.361081	0.438176
MSE	0.357872	0.275860	0.260540	0.228180	0.314874

Q31 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419894	0.385442	0.406183	0.351740	0.446978
MSE	0.358992	0.250588	0.266588	0.216574	0.315336

Q32 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.436956	0.453538	0.439915	0.429424	0.464383
MSE	0.376295	0.335529	0.300018	0.292640	0.329769

Q33 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.442685	0.408400	0.423111	0.387809	0.473821
MSE	0.393790	0.280488	0.272991	0.243580	0.324498

Q34 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.366168	0.378628	0.370017	0.358124	0.400722
MSE	0.299589	0.265425	0.232346	0.221763	0.265879

Q35 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.458136	0.450121	0.447584	0.431661	0.474756
MSE	0.370095	0.311876	0.287266	0.274722	0.317536

Q36 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.432617	0.432971	0.409432	0.396679	0.453247
MSE	0.352219	0.319361	0.262821	0.252559	0.307460

Q37 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.397899	0.389117	0.391068	0.369144	0.427005
MSE	0.318783	0.261212	0.239162	0.222958	0.276489

Q38 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.333722	0.337183	0.333510	0.320610	0.358557
MSE	0.247029	0.211264	0.190478	0.179569	0.216764

Q39 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.349569	0.352861	0.357039	0.338780	0.379611
MSE	0.276264	0.233756	0.219115	0.202527	0.243970

Q40 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.354655	0.346557	0.358015	0.331557	0.386345
MSE	0.277400	0.220077	0.208951	0.189225	0.239121

Q41 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.371537	0.370840	0.370018	0.349557	0.402219
MSE	0.294577	0.239949	0.224759	0.205052	0.257687

Q46 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.319447	0.338862	0.332666	0.320008	0.351745
MSE	0.240197	0.194885	0.185590	0.169211	0.207259

Q47 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.329706	0.338162	0.325811	0.313460	0.351714
MSE	0.219124	0.190621	0.174145	0.160071	0.197911

Q48 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.372432	0.343963	0.351844	0.328187	0.381633
MSE	0.247030	0.207923	0.200508	0.181644	0.228073

Q49 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.375234	0.328430	0.349218	0.305576	0.383255
MSE	0.247242	0.191253	0.198865	0.164316	0.231567

Q50 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.402091	0.369318	0.377550	0.335084	0.418609
MSE	0.300177	0.238787	0.230667	0.191283	0.269314

Q51 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.328670	0.328610	0.334799	0.317657	0.398528
MSE	0.337318	0.236818	0.230414	0.196990	0.273724

Q52 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.373080	0.364031	0.379095	0.356637	0.439305
MSE	0.395875	0.276040	0.264440	0.233984	0.311572

Q53 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.399929	0.375950	0.403512	0.365208	0.465253
MSE	0.421055	0.290029	0.285568	0.242205	0.335447

Q54 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.398232	0.391852	0.401640	0.367417	0.464085
MSE	0.433674	0.308831	0.292255	0.245357	0.341355

Q55 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.218978	0.282651	0.243290	0.250605	0.300637
MSE	0.263561	0.209784	0.177239	0.156363	0.209854

Q58 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.216690	0.312293	0.257733	0.257819	0.293038
MSE	0.208782	0.185484	0.148392	0.137890	0.167403

Q59 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.326510	0.365167	0.336672	0.327101	0.366598
MSE	0.270721	0.231673	0.212905	0.188266	0.243872

Q60 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.353514	0.376850	0.356965	0.341203	0.382297
MSE	0.281677	0.238307	0.220317	0.194344	0.253189

Q61 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.384431	0.386952	0.381830	0.346697	0.411461
MSE	0.305889	0.245347	0.229033	0.195111	0.267955

Q62 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.409408	0.376170	0.394316	0.331389	0.436305
MSE	0.331113	0.232981	0.240984	0.186821	0.287934

Q63 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419277	0.384140	0.399866	0.33152	0.446340
MSE	0.343810	0.242451	0.247441	0.19147	0.302661

Q64 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.412911	0.388527	0.395030	0.359883	0.451114
MSE	0.359587	0.260032	0.252345	0.218856	0.315275

Q65 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419160	0.386658	0.404029	0.374230	0.456151
MSE	0.357762	0.262903	0.260550	0.230771	0.324979

Q66 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.402850	0.350834	0.386188	0.340360	0.438313
MSE	0.334869	0.223301	0.239495	0.197113	0.308395

Q67 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.391702	0.344037	0.37671	0.330327	0.425630
MSE	0.319104	0.217588	0.23024	0.188916	0.290836

Q68 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.400982	0.342688	0.380523	0.341939	0.442721
MSE	0.331113	0.216328	0.229619	0.197060	0.295341

Q69 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.424662	0.348059	0.391937	0.344452	0.462141
MSE	0.360140	0.220277	0.242082	0.199425	0.320887

Q70 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.413895	0.328632	0.382139	0.330881	0.451170
MSE	0.349923	0.202063	0.235864	0.188883	0.316911

Q71 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.415175	0.314798	0.374268	0.313481	0.455009
MSE	0.353465	0.185761	0.221397	0.170282	0.316833

Q72 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.372168	0.303299	0.349145	0.291106	0.418969
MSE	0.314765	0.174840	0.195377	0.148415	0.274205

Q73 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.390076	0.302060	0.358603	0.293767	0.435821
MSE	0.323319	0.170482	0.202056	0.151737	0.285526

Q74 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.393064	0.310959	0.360928	0.310928	0.437256
MSE	0.324080	0.181430	0.208425	0.167374	0.288131

Q75 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.387445	0.342191	0.366698	0.335662	0.427989
MSE	0.332988	0.215290	0.223268	0.192304	0.292659

Q76 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.422801	0.344685	0.387945	0.343185	0.458582
MSE	0.355283	0.216378	0.237655	0.197440	0.324251

Q77 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.405163	0.333975	0.384087	0.339043	0.448790
MSE	0.337154	0.207859	0.231006	0.192881	0.305652

Q78 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.406179	0.338624	0.382234	0.341146	0.441759
MSE	0.339581	0.212683	0.234668	0.197774	0.308202

Q79 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.409507	0.328192	0.381115	0.333136	0.454769
MSE	0.349051	0.201192	0.236441	0.191584	0.317100

Q80 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.411212	0.329336	0.382827	0.328055	0.454893
MSE	0.351311	0.202502	0.238018	0.189040	0.318238

Q81 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.417887	0.335804	0.390000	0.336138	0.464267
MSE	0.359549	0.209286	0.241097	0.195101	0.319614

Q82 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.426634	0.321881	0.391872	0.332153	0.470259
MSE	0.359803	0.200932	0.238701	0.191109	0.326534

Q83 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.422582	0.300674	0.383928	0.315719	0.464134
MSE	0.357110	0.180390	0.234044	0.177059	0.326212

Q84 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.420006	0.302071	0.384773	0.318332	0.463342
MSE	0.354650	0.181800	0.235590	0.181334	0.324126

Q85 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419038	0.304170	0.382943	0.313720	0.468269
MSE	0.358455	0.184351	0.231898	0.176771	0.321325

Q86 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.427876	0.312305	0.391564	0.319095	0.469567
MSE	0.364123	0.193591	0.240369	0.182715	0.327557

Q87 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419740	0.304408	0.386534	0.312342	0.461371
MSE	0.350907	0.185280	0.234482	0.175295	0.323286

Q88 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419727	0.323252	0.387544	0.324222	0.461928
MSE	0.352825	0.199292	0.237414	0.183435	0.322073

Q89 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.434656	0.340675	0.402908	0.339673	0.472887
MSE	0.358851	0.223125	0.253856	0.203689	0.331807

Q90 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.482452	0.447083	0.454350	0.430176	0.495295
MSE	0.398162	0.343870	0.322836	0.306332	0.367572

Q106 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.497036	0.478550	0.473277	0.459843	0.509494
MSE	0.423512	0.380555	0.344284	0.337794	0.382243

Q107 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.497464	0.491725	0.473535	0.464313	0.502410
MSE	0.420110	0.387584	0.344854	0.341896	0.375765

Q108 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.520237	0.502996	0.493862	0.48062	0.526719
MSE	0.443592	0.408701	0.363355	0.35869	0.398122

Q109 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.510173	0.489510	0.480549	0.463235	0.519556
MSE	0.442588	0.399242	0.349148	0.340256	0.390032

Q110 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.482244	0.465874	0.458422	0.437646	0.496383
MSE	0.418068	0.369461	0.324532	0.313866	0.364503

Q112 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.379729	0.382089	0.367164	0.342899	0.406665
MSE	0.307696	0.262348	0.225363	0.207446	0.265150

Q113 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.355056	0.339978	0.350308	0.310494	0.389685
MSE	0.280545	0.204951	0.205541	0.168774	0.248952

Q114 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.370154	0.342006	0.362331	0.316730	0.405747
MSE	0.299667	0.200936	0.210447	0.171886	0.252402

Q115 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.372880	0.337363	0.357721	0.305968	0.405783
MSE	0.300715	0.192385	0.208120	0.162581	0.255463

Q116 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.371967	0.347041	0.359377	0.314863	0.403953
MSE	0.298306	0.204072	0.212975	0.173217	0.256224

Q117 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.389327	0.381037	0.379806	0.349209	0.422079
MSE	0.332183	0.247094	0.237472	0.208418	0.276171

Q118 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.401964	0.404301	0.394471	0.380534	0.431240
MSE	0.339552	0.285279	0.250829	0.238595	0.287057

Q120 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.443072	0.445682	0.427713	0.416598	0.455325
MSE	0.337633	0.322544	0.280129	0.274661	0.308024

Q121 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.373944	0.408964	0.369375	0.363883	0.395051
MSE	0.283607	0.278633	0.229169	0.222439	0.255035

Q130 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.376544	0.411947	0.380743	0.372069	0.414686
MSE	0.318566	0.288403	0.234396	0.227768	0.262918

Q131 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.380739	0.391305	0.375227	0.358122	0.411139
MSE	0.308843	0.260388	0.229251	0.212649	0.264649

Q132 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.414988	0.379370	0.401700	0.366828	0.450706
MSE	0.375861	0.263432	0.258534	0.225774	0.310946

Q133 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.449187	0.403146	0.419550	0.376220	0.473624
MSE	0.396917	0.284655	0.277514	0.235896	0.342032

Q134 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.397348	0.349672	0.377167	0.341630	0.437620
MSE	0.372204	0.237901	0.239606	0.207129	0.298632

Q135 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.365553	0.332303	0.362799	0.324181	0.425104
MSE	0.355222	0.225214	0.228753	0.192728	0.288309

Q136 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.389063	0.33803	0.373803	0.323616	0.443131
MSE	0.385352	0.22722	0.248516	0.196146	0.315360

Q137 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.431567	0.357679	0.403606	0.340769	0.465600
MSE	0.406412	0.239771	0.266045	0.205933	0.332185

Q138 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.403133	0.348762	0.387497	0.333783	0.456696
MSE	0.399210	0.238404	0.256255	0.207632	0.321507

Q142 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.478914	0.439298	0.455695	0.409172	0.517853
MSE	0.477319	0.336875	0.328091	0.286031	0.397388

Q143 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.477179	0.437673	0.455345	0.403871	0.518452
MSE	0.486553	0.335589	0.331684	0.285862	0.405230

Q146 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.455391	0.395446	0.428629	0.355429	0.501815
MSE	0.459695	0.280989	0.296276	0.228784	0.381962

Q147 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.455399	0.381029	0.418075	0.330917	0.499419
MSE	0.456008	0.260729	0.281965	0.203313	0.373299

Q148 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.448651	0.380848	0.417791	0.329091	0.491042
MSE	0.435832	0.263711	0.280410	0.201204	0.365323

Q158 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419258	0.368117	0.393074	0.341121	0.446435
MSE	0.351002	0.249317	0.249122	0.210201	0.305429

Q159 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.414834	0.375156	0.388733	0.350473	0.432829
MSE	0.320394	0.254204	0.245305	0.220690	0.291859

Q160 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.476524	0.447059	0.444549	0.420125	0.488381
MSE	0.390561	0.331977	0.304864	0.292462	0.354011

Q161 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.483998	0.454834	0.451315	0.42704	0.494481
MSE	0.399169	0.339112	0.312169	0.29825	0.357804

Q162 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.493229	0.466855	0.461333	0.445788	0.505004
MSE	0.415564	0.361699	0.329432	0.322402	0.372697

Q163 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.425433	0.39845	0.402383	0.384217	0.444551
MSE	0.346672	0.28974	0.268480	0.259117	0.308000

Q164 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.380928	0.364079	0.357872	0.334021	0.418214
MSE	0.343487	0.253420	0.233836	0.212490	0.297973

Q169 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.387327	0.370550	0.369744	0.345994	0.423643
MSE	0.352063	0.257407	0.238635	0.216789	0.299855

Q170 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.414751	0.397832	0.396393	0.370932	0.451426
MSE	0.387075	0.287502	0.264226	0.243365	0.325513

Q171 (1-7 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.450394	0.417801	0.423396	0.389831	0.482137
MSE	0.410036	0.303696	0.286504	0.257629	0.348200

Q172 (1-8 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.434384	0.401889	0.408198	0.371339	0.474862
MSE	0.419840	0.296632	0.284783	0.249140	0.352546

Q176 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.433420	0.413656	0.414001	0.404366	0.455570
MSE	0.372971	0.325465	0.288817	0.283407	0.329133

Q177 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.399962	0.354517	0.380987	0.366296	0.439472
MSE	0.375740	0.288821	0.265734	0.253325	0.316076

Q178 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.379250	0.323767	0.361682	0.341759	0.424755
MSE	0.371372	0.263055	0.248986	0.230559	0.301778

Q179 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.273809	0.231593	0.26124	0.246716	0.336104
MSE	0.280314	0.157787	0.16585	0.142667	0.226319

Q180 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.307415	0.251885	0.290435	0.267081	0.364044
MSE	0.311805	0.178544	0.189770	0.160299	0.252035

Q181 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.275388	0.222257	0.261450	0.236662	0.335948
MSE	0.273027	0.143797	0.160426	0.130781	0.223124

Q182 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.367081	0.301040	0.329832	0.301171	0.410035
MSE	0.348759	0.216875	0.219345	0.185937	0.293690

Q183 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.372323	0.295120	0.329062	0.295720	0.409578
MSE	0.340540	0.203466	0.209621	0.176266	0.285837

Q184 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.390899	0.317732	0.349579	0.314843	0.431346
MSE	0.369342	0.229075	0.226471	0.192763	0.307105

Q185 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419414	0.346472	0.373286	0.336736	0.451990
MSE	0.381408	0.252109	0.247060	0.212131	0.328633

Q186 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.421826	0.340432	0.372702	0.333317	0.456002
MSE	0.389871	0.247195	0.249431	0.211563	0.333411

Q187 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.355994	0.283696	0.313694	0.285337	0.399458
MSE	0.337696	0.198036	0.201841	0.170870	0.283060

Q188 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.384361	0.315958	0.354322	0.325215	0.430283
MSE	0.376841	0.241084	0.240166	0.211009	0.312961

Q189 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.296204	0.246799	0.271625	0.253225	0.352214
MSE	0.301551	0.167084	0.172203	0.147253	0.244945

Q190 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.345691	0.304395	0.326324	0.308556	0.391991
MSE	0.325195	0.228270	0.218748	0.196588	0.276506

Q191 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.299388	0.250116	0.284743	0.259644	0.355463
MSE	0.294594	0.172012	0.177172	0.150490	0.237439

Q192 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.296617	0.253151	0.275575	0.256120	0.351127
MSE	0.287790	0.173594	0.176571	0.151341	0.241427

Q193 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.380519	0.324980	0.349056	0.323397	0.420110
MSE	0.361610	0.243514	0.239491	0.210982	0.307187

Q194 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.330605	0.276999	0.304514	0.277670	0.380929
MSE	0.332280	0.194199	0.195297	0.168823	0.265907

Q195 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.444069	0.398048	0.413786	0.389999	0.483784
MSE	0.446009	0.325774	0.303147	0.279956	0.372683

Q196 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.495820	0.477969	0.473347	0.444831	0.529766
MSE	0.496558	0.396081	0.358577	0.335467	0.418515

Q197 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.471748	0.451758	0.451663	0.409767	0.512781
MSE	0.482924	0.353330	0.324806	0.288784	0.389964

Q198 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.516241	0.490912	0.491087	0.446823	0.555655
MSE	0.535210	0.402402	0.371297	0.336393	0.441670

Q199 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.486586	0.472462	0.470287	0.434531	0.524417
MSE	0.477497	0.371676	0.341661	0.312510	0.407922

Q201 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.537197	0.523169	0.524017	0.490007	0.594255
MSE	0.606064	0.450550	0.406835	0.384539	0.487750

Q202 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.461621	0.472242	0.458491	0.440116	0.533979
MSE	0.551112	0.400752	0.357833	0.339696	0.434933

Q203 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.560457	0.551439	0.541471	0.510007	0.62092
MSE	0.667161	0.485932	0.441893	0.421738	0.52795

Q204 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.600980	0.557545	0.550105	0.499832	0.668959
MSE	0.779301	0.473850	0.447567	0.405342	0.591086

Q205 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.571025	0.549767	0.537016	0.489099	0.645959
MSE	0.733841	0.482599	0.435008	0.393828	0.566159

Q206 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.565872	0.510918	0.501946	0.434819	0.635798
MSE	0.738703	0.416765	0.394191	0.329290	0.560014

Q207 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.587280	0.523257	0.522184	0.437847	0.653698
MSE	0.774775	0.424975	0.409572	0.334147	0.569386

Q208 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.525409	0.498486	0.49165	0.447483	0.574559
MSE	0.605050	0.404016	0.37369	0.336964	0.475031

Q221 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.427195	0.397204	0.410690	0.352379	0.485512
MSE	0.444788	0.279997	0.277184	0.226728	0.351327

Q222 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.422327	0.400851	0.409921	0.347952	0.483330
MSE	0.433081	0.284407	0.277850	0.227162	0.348787

Q224 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.478377	0.448960	0.453239	0.415435	0.522354
MSE	0.475545	0.339456	0.322031	0.291620	0.397199

Q225 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.496531	0.475074	0.474318	0.442532	0.534993
MSE	0.480222	0.359128	0.339079	0.318492	0.407749

Q226 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.473749	0.456192	0.459083	0.429461	0.508142
MSE	0.438083	0.341431	0.323711	0.301806	0.379018

Q227 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.473308	0.442555	0.451928	0.415167	0.509918
MSE	0.447405	0.326831	0.315205	0.285526	0.387112

Q228 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.474995	0.458955	0.460831	0.431366	0.511254
MSE	0.443548	0.345008	0.324239	0.302980	0.378623

Q229 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.456571	0.435035	0.432900	0.408838	0.493297
MSE	0.421646	0.316634	0.301458	0.279874	0.370300

Q230 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.477256	0.445136	0.454286	0.421368	0.506326
MSE	0.441206	0.326940	0.319152	0.293153	0.381461

Q231 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.451234	0.433327	0.435763	0.411156	0.487107
MSE	0.416533	0.327624	0.300956	0.283461	0.359504

Q232 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.465308	0.447057	0.449158	0.428881	0.495635
MSE	0.418088	0.333007	0.317279	0.303153	0.371781

Q233 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.468196	0.451827	0.454118	0.429999	0.496775
MSE	0.426068	0.337329	0.314887	0.300202	0.368158

Q234 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.418469	0.414102	0.412200	0.394952	0.458998
MSE	0.388295	0.297794	0.272353	0.265781	0.322589

Q235 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.476474	0.457844	0.456715	0.437351	0.508589
MSE	0.444285	0.351374	0.324894	0.313922	0.381988

Q236 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.447483	0.438868	0.439836	0.419428	0.478144
MSE	0.396120	0.326085	0.303910	0.290293	0.349498

Q237 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.454525	0.427137	0.428752	0.403833	0.485823
MSE	0.408764	0.318314	0.292528	0.277064	0.351316

Q238 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.419873	0.400083	0.402770	0.385564	0.450700
MSE	0.363912	0.282999	0.262677	0.253183	0.311199

Q239 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.458786	0.431946	0.431523	0.410736	0.486007
MSE	0.412680	0.321277	0.291711	0.279991	0.353360

Q240 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.424013	0.412371	0.403197	0.394062	0.443514
MSE	0.346110	0.300597	0.270119	0.265724	0.309533

Q241 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.456469	0.422309	0.427783	0.40831	0.472233
MSE	0.367395	0.305951	0.284653	0.27654	0.328820

Q242 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.432303	0.394570	0.399533	0.376452	0.450564
MSE	0.348763	0.273726	0.257724	0.245622	0.305771

Q243 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.411984	0.370913	0.383408	0.359806	0.437992
MSE	0.345973	0.258352	0.247057	0.234632	0.296766

Q244 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.440959	0.402725	0.413164	0.390420	0.458285
MSE	0.353124	0.285955	0.269972	0.257949	0.312415

Q245 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.471646	0.417681	0.426581	0.402606	0.487386
MSE	0.387964	0.297936	0.281694	0.271128	0.338782

Q246 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.411290	0.363736	0.382541	0.355164	0.434267
MSE	0.333777	0.246803	0.238605	0.223729	0.287199

Q247 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.473376	0.425140	0.433788	0.404666	0.484131
MSE	0.374458	0.300927	0.287386	0.272513	0.338517

Q248 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.440873	0.398576	0.408355	0.381750	0.462577
MSE	0.362027	0.276376	0.263095	0.247913	0.318699

Q249 (1-11 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.382984	0.352110	0.362493	0.335528	0.420361
MSE	0.334036	0.234097	0.224018	0.207127	0.276410

Q250 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.391977	0.360232	0.364893	0.342665	0.418996
MSE	0.321821	0.241493	0.227507	0.213272	0.278840

Q251 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.410036	0.369221	0.379584	0.348248	0.433494
MSE	0.328636	0.245111	0.239875	0.214971	0.295683

Q252 (1-10 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.408474	0.369414	0.380239	0.349238	0.431633
MSE	0.323991	0.245624	0.237377	0.213063	0.290393

Q253 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.386603	0.373819	0.372460	0.347178	0.420266
MSE	0.329811	0.246285	0.234036	0.215373	0.282917

Q254 (1-5 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.336067	0.332969	0.330697	0.307747	0.371048
MSE	0.265046	0.204359	0.188937	0.172563	0.228357

Q255 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.358841	0.337882	0.351965	0.316931	0.395889
MSE	0.293612	0.208033	0.209656	0.181046	0.250699

Q256 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.359403	0.334385	0.352813	0.309652	0.398327
MSE	0.295753	0.202925	0.210312	0.172925	0.253959

Q257 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.385763	0.353475	0.377826	0.333146	0.420812
MSE	0.322500	0.222657	0.230103	0.191753	0.274636

Q258 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.391801	0.370660	0.385323	0.335889	0.428784
MSE	0.330453	0.237164	0.241855	0.198731	0.290856

Q259 (1-4 range)

Metric	MF	I2I	KNN	MB	Mean
MAE	0.44435	0.41394	0.429939	0.370709	0.481413
MSE	0.38953	0.28060	0.281157	0.231785	0.337770

B.2 Classification task

This section lists the results per questions falling under the classification imputation task.

Q7 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.873258	0.872532	0.874220	0.853803	0.869015
F1	0.544311	0.537593	0.403728	0.470996	0.330462
Precision	0.506310	0.499481	0.396425	0.470916	0.374786
Recall	0.502310	0.495554	0.383026	0.444768	0.349535

Q8 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.478965	0.478250	0.567161	0.442631	0.544507
F1	0.443986	0.434384	0.443258	0.448613	0.190797
Precision	0.434220	0.423254	0.403686	0.387430	0.344893
Recall	0.391976	0.382315	0.376290	0.302821	0.242244

Q9 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.731022	0.706502	0.796676	0.610537	0.736204
F1	0.648539	0.492969	0.649210	0.461620	0.371346
Precision	0.619119	0.439299	0.590742	0.294091	0.503109
Recall	0.582698	0.434320	0.574717	0.333878	0.422710

Q10 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.868853	0.841698	0.876022	0.804327	0.871425
F1	0.566192	0.431169	0.511822	0.390995	0.433906
Precision	0.526979	0.390187	0.507817	0.360312	0.492281
Recall	0.512086	0.384818	0.487361	0.358178	0.459363

Q11 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.568938	0.556622	0.608366	0.518636	0.554054
F1	0.531743	0.516757	0.494602	0.535571	0.196060
Precision	0.524103	0.497525	0.437785	0.437766	0.351238
Recall	0.484546	0.470163	0.417625	0.401547	0.248556

Q12 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.554835	0.458293	0.683231	0.319942	0.669766
F1	0.518120	0.372117	0.542313	0.344833	0.321152
Precision	0.505518	0.273936	0.518540	0.153942	0.484513
Recall	0.424863	0.275586	0.487068	0.180661	0.383066

Q13 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.829842	0.829842	0.807367	0.781776	0.797349
F1	0.587849	0.587961	0.374785	0.499328	0.271236
Precision	0.470244	0.470244	0.360480	0.469665	0.338996
Recall	0.482895	0.482924	0.335648	0.436819	0.299435

Q14 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.891937	0.891937	0.866476	0.867731	0.858167
F1	0.627927	0.627931	0.385124	0.526593	0.301664
Precision	0.525688	0.525688	0.371056	0.497717	0.349386
Recall	0.545600	0.545602	0.355562	0.480944	0.322618

Q15 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.852894	0.852743	0.836988	0.830834	0.826171
F1	0.566863	0.566928	0.393358	0.486694	0.291019
Precision	0.486239	0.485652	0.374932	0.463496	0.349951
Recall	0.494212	0.494120	0.356597	0.448278	0.315854

Q16 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.648481	0.644181	0.614861	0.598219	0.572062
F1	0.599665	0.584501	0.468015	0.557920	0.203784
Precision	0.514585	0.496680	0.422582	0.449042	0.351345
Recall	0.495200	0.480779	0.395149	0.434531	0.254491

Q17 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.734761	0.675644	0.751619	0.496415	0.736572
F1	0.507441	0.367661	0.577200	0.359508	0.367830
Precision	0.512405	0.338979	0.545234	0.226821	0.499055
Recall	0.453667	0.316866	0.519492	0.250405	0.421006

Q18 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.833881	0.816367	0.838859	0.722084	0.811478
F1	0.644017	0.530494	0.606117	0.449615	0.399863
Precision	0.580808	0.469517	0.541966	0.333435	0.486481
Recall	0.568306	0.465460	0.527993	0.349787	0.435530

Q19 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.600467	0.599339	0.583345	0.565948	0.536117
F1	0.569230	0.566179	0.456100	0.544123	0.182744
Precision	0.500923	0.497088	0.416690	0.445642	0.338645
Recall	0.476608	0.473949	0.391110	0.418466	0.232996

Q20 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.725820	0.710000	0.704203	0.691056	0.651533
F1	0.617859	0.597880	0.486998	0.597370	0.236049
Precision	0.530492	0.498629	0.438150	0.486644	0.360867
Recall	0.523473	0.502828	0.420720	0.490227	0.281625

Q21 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.600466	0.512610	0.698035	0.324700	0.686260
F1	0.528410	0.392262	0.549688	0.342103	0.330179
Precision	0.507193	0.301903	0.521894	0.159986	0.485244
Recall	0.427839	0.293976	0.492318	0.185639	0.389594

Q22 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.888142	0.887503	0.871129	0.856942	0.862692
F1	0.618097	0.615583	0.402018	0.528107	0.306581
Precision	0.538535	0.534987	0.378326	0.498455	0.353323
Recall	0.548427	0.545373	0.365876	0.476780	0.326972

Q23 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.859383	0.857796	0.842649	0.814738	0.822620
F1	0.631518	0.613667	0.428264	0.530074	0.300406
Precision	0.539237	0.519882	0.402291	0.477898	0.364678
Recall	0.547722	0.531302	0.386051	0.462636	0.326976

Q24 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.778928	0.768395	0.888079	0.803835	0.864951
F1	0.582677	0.451837	0.604289	0.484604	0.470654
Precision	0.598809	0.440190	0.577918	0.414009	0.535409
Recall	0.539297	0.416131	0.569248	0.425678	0.497044

Q25 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.698808	0.671239	0.830735	0.699915	0.811811
F1	0.563297	0.441998	0.540853	0.413743	0.393120
Precision	0.537889	0.374862	0.516443	0.340435	0.481746
Recall	0.497093	0.377778	0.494073	0.352177	0.429985

Q26 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.702696	0.698239	0.681366	0.678266	0.634723
F1	0.601995	0.584671	0.472620	0.562519	0.227093
Precision	0.519124	0.499260	0.427151	0.473994	0.353772
Recall	0.514813	0.499455	0.404599	0.469177	0.272819

Q42 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.631692	0.622095	0.618888	0.591078	0.566339
F1	0.562997	0.544249	0.475990	0.545095	0.202470
Precision	0.496695	0.471369	0.427013	0.435677	0.354230
Recall	0.476076	0.458686	0.402260	0.425110	0.254537

Q43 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.583432	0.518184	0.691844	0.439105	0.668328
F1	0.563685	0.429883	0.570416	0.438321	0.311997
Precision	0.531119	0.339501	0.522533	0.238027	0.465409
Recall	0.463195	0.335906	0.500571	0.265948	0.370528

Q44 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.674603	0.664687	0.670539	0.627066	0.611581
F1	0.574220	0.555590	0.474784	0.532623	0.218674
Precision	0.505500	0.479648	0.436928	0.446030	0.355624
Recall	0.489670	0.471852	0.415698	0.431775	0.267254

Q45 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.469833	0.450949	0.557227	0.372932	0.536183
F1	0.419428	0.393728	0.432917	0.386293	0.198007
Precision	0.418484	0.379445	0.415674	0.330681	0.362788
Recall	0.361733	0.335516	0.391290	0.240503	0.251903

Q57 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.639704	0.545676	0.694245	0.359771	0.677105
F1	0.496308	0.363856	0.538976	0.332893	0.324699
Precision	0.496536	0.301862	0.521474	0.175749	0.480390
Recall	0.424646	0.287800	0.489653	0.194173	0.384250

Q94 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.563487	0.487927	0.719616	0.385277	0.713069
F1	0.492525	0.367418	0.518897	0.315485	0.332803
Precision	0.488771	0.288166	0.496464	0.182443	0.470068
Recall	0.417318	0.288992	0.459431	0.199264	0.386980

Q95 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.578742	0.570639	0.702924	0.631419	0.700131
F1	0.410087	0.401668	0.372691	0.362823	0.241181
Precision	0.398735	0.383596	0.367846	0.370608	0.344805
Recall	0.364122	0.355309	0.336883	0.334952	0.281508

Q96 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.817704	0.817518	0.800796	0.783764	0.784260
F1	0.577698	0.575392	0.409264	0.517619	0.268818
Precision	0.502317	0.499843	0.376558	0.477826	0.342222
Recall	0.508724	0.506655	0.358935	0.450775	0.298111

Q97 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.608636	0.607074	0.628959	0.554104	0.577845
F1	0.526110	0.516682	0.477431	0.551330	0.204523
Precision	0.502807	0.492492	0.426880	0.448491	0.349776
Recall	0.474781	0.466326	0.405052	0.410159	0.253469

Q98 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.703052	0.698674	0.686290	0.695016	0.627696
F1	0.566426	0.548458	0.481083	0.554073	0.229812
Precision	0.516984	0.497052	0.435986	0.481188	0.360721
Recall	0.501518	0.485493	0.414435	0.474692	0.276434

Q99 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.774456	0.751187	0.794628	0.646769	0.760736
F1	0.621000	0.513262	0.597931	0.468281	0.365387
Precision	0.555546	0.444803	0.538351	0.326402	0.474844
Recall	0.534087	0.438010	0.523661	0.342131	0.408724

Q100 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.498759	0.490957	0.608637	0.438267	0.585017
F1	0.438581	0.405468	0.471380	0.435727	0.226808
Precision	0.428975	0.392313	0.432662	0.354270	0.375887
Recall	0.380625	0.350750	0.409501	0.284153	0.277807

Q101 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.701598	0.646265	0.754921	0.522426	0.733228
F1	0.517700	0.383842	0.565090	0.386856	0.359648
Precision	0.506579	0.336607	0.533452	0.237037	0.488411
Recall	0.452704	0.322584	0.510577	0.254204	0.409199

Q102 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.848050	0.822996	0.859095	0.759487	0.832961
F1	0.603925	0.487621	0.582796	0.431299	0.408492
Precision	0.555308	0.432434	0.543790	0.346202	0.487973
Recall	0.540190	0.430314	0.533649	0.359283	0.442039

Q103 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.787686	0.732123	0.796635	0.634753	0.791433
F1	0.536458	0.412488	0.514636	0.381371	0.374689
Precision	0.506355	0.353913	0.493426	0.313102	0.471754
Recall	0.472845	0.349279	0.467370	0.316292	0.414597

Q104 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.661020	0.654382	0.650116	0.644095	0.594983
F1	0.587653	0.570411	0.462552	0.527059	0.214141
Precision	0.509700	0.485588	0.422228	0.457007	0.352515
Recall	0.493114	0.475685	0.393966	0.442599	0.261241

Q105 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.730066	0.722693	0.734936	0.660429	0.720778
F1	0.555225	0.536506	0.416057	0.458262	0.258367
Precision	0.460850	0.435222	0.387346	0.422336	0.356232
Recall	0.449651	0.430686	0.357200	0.386851	0.296433

Q122 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.627238	0.590865	0.773039	0.601963	0.767578
F1	0.511800	0.395931	0.496959	0.346681	0.365923
Precision	0.504187	0.340568	0.487812	0.310184	0.469308
Recall	0.457395	0.340089	0.454001	0.297323	0.407847

Q123 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.588695	0.582986	0.644460	0.537904	0.609655
F1	0.501873	0.484697	0.470495	0.507232	0.227150
Precision	0.502953	0.479563	0.430649	0.416302	0.362914
Recall	0.463078	0.446008	0.408393	0.365462	0.274540

Q124 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.543785	0.529510	0.645737	0.534769	0.627803
F1	0.406015	0.376815	0.430927	0.380832	0.231324
Precision	0.405142	0.364999	0.408732	0.331467	0.367066
Recall	0.355642	0.326798	0.382084	0.294609	0.279401

Q125 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.619938	0.533849	0.689406	0.326542	0.666666
F1	0.518193	0.378253	0.592150	0.375999	0.326210
Precision	0.516628	0.312104	0.543103	0.160845	0.489187
Recall	0.438290	0.298942	0.516259	0.192323	0.388721

Q126 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.715874	0.677184	0.774787	0.554461	0.731593
F1	0.622070	0.493500	0.635487	0.461129	0.365542
Precision	0.571845	0.419273	0.576219	0.274504	0.497895
Recall	0.530058	0.411131	0.559425	0.310332	0.417716

Q127 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.773738	0.758273	0.836008	0.736668	0.760444
F1	0.683687	0.552699	0.676788	0.501507	0.376587
Precision	0.638197	0.492565	0.605545	0.367484	0.494084
Recall	0.606542	0.484590	0.599391	0.398754	0.422951

Q128 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.715032	0.670210	0.752417	0.545548	0.709667
F1	0.603672	0.501622	0.583963	0.471123	0.336808
Precision	0.539181	0.413821	0.539629	0.289816	0.471511
Recall	0.496811	0.404396	0.523603	0.313326	0.387757

Q129 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.744225	0.736487	0.742330	0.733658	0.708665
F1	0.553623	0.535508	0.458568	0.531728	0.263148
Precision	0.499954	0.475647	0.429910	0.470596	0.366979
Recall	0.484058	0.466446	0.411999	0.464495	0.303616

Q139 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.640265	0.635148	0.614725	0.580837	0.551770
F1	0.593182	0.576289	0.507357	0.570853	0.202978
Precision	0.522253	0.500905	0.452790	0.460575	0.358065
Recall	0.499721	0.483178	0.428409	0.447091	0.254879

Q140 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.641286	0.590916	0.701300	0.484508	0.679145
F1	0.488837	0.387432	0.540400	0.410323	0.317899
Precision	0.492559	0.354879	0.510491	0.258569	0.461310
Recall	0.428029	0.326203	0.486881	0.270417	0.371477

Q141 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.588951	0.581683	0.618011	0.518019	0.582730
F1	0.498293	0.481274	0.465755	0.492047	0.211837
Precision	0.491805	0.468544	0.421041	0.394986	0.358302
Recall	0.452287	0.434786	0.396472	0.352265	0.261625

Q144 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.473456	0.470398	0.691655	0.654993	0.689249
F1	0.407746	0.392570	0.378086	0.305176	0.247903
Precision	0.400774	0.383013	0.374441	0.343428	0.356326
Recall	0.326664	0.311684	0.336184	0.292621	0.289199

Q145 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.710514	0.658844	0.742004	0.502124	0.729563
F1	0.512785	0.417442	0.524501	0.359966	0.330023
Precision	0.500125	0.373946	0.493161	0.250975	0.455244
Recall	0.448472	0.352794	0.463822	0.260369	0.380048

Q151 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.842518	0.834450	0.831928	0.787583	0.821353
F1	0.579518	0.550340	0.411016	0.455980	0.306833
Precision	0.502105	0.470144	0.401121	0.423779	0.374337
Recall	0.502457	0.476180	0.381945	0.412923	0.335274

Q165 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.877848	0.874622	0.863859	0.839376	0.846817
F1	0.608687	0.583065	0.448600	0.512175	0.326583
Precision	0.550500	0.523382	0.421342	0.476546	0.385485
Recall	0.553238	0.529114	0.408697	0.467323	0.351445

Q166 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.878730	0.849814	0.890475	0.844593	0.880212
F1	0.562144	0.444122	0.522172	0.442793	0.425238
Precision	0.527263	0.400945	0.502876	0.412737	0.479505
Recall	0.517339	0.400039	0.487459	0.411936	0.449090

Q167 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.864163	0.849600	0.868640	0.788821	0.841256
F1	0.655595	0.548994	0.582246	0.479400	0.403958
Precision	0.574800	0.469759	0.530237	0.379510	0.477869
Recall	0.574224	0.476499	0.519661	0.393642	0.435556

Q168 (1-2 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.751462	0.728852	0.788652	0.675327	0.731317
F1	0.651856	0.537987	0.639917	0.512756	0.343372
Precision	0.601290	0.466819	0.560302	0.356744	0.468263
Recall	0.573052	0.464784	0.547392	0.389841	0.391986

Q173 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.721049	0.710988	0.719698	0.697775	0.673079
F1	0.559299	0.528245	0.473086	0.514512	0.250205
Precision	0.502846	0.463810	0.434404	0.441280	0.368246
Recall	0.488931	0.458409	0.412256	0.438904	0.294058

Q200 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.699543	0.693298	0.668890	0.679758	0.599451
F1	0.615928	0.587044	0.497671	0.580623	0.231146
Precision	0.536165	0.502819	0.455245	0.483765	0.371072
Recall	0.528070	0.500064	0.431850	0.484950	0.278427

Q209 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.675430	0.666523	0.699829	0.601365	0.682352
F1	0.531018	0.491638	0.461177	0.456264	0.269561
Precision	0.463794	0.414448	0.427858	0.387637	0.385448
Recall	0.436705	0.396746	0.401675	0.357039	0.312390

Q210 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.618239	0.558352	0.716480	0.469373	0.694122
F1	0.546817	0.421732	0.556166	0.411208	0.332356
Precision	0.521771	0.342847	0.522168	0.252324	0.474766
Recall	0.462190	0.337460	0.494306	0.271298	0.387348

Q211 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.640770	0.577043	0.719302	0.476090	0.690214
F1	0.541526	0.400003	0.576597	0.427787	0.335793
Precision	0.525417	0.333275	0.538416	0.241446	0.485443
Recall	0.462544	0.324065	0.513332	0.274303	0.393058

Q212 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.770177	0.729109	0.796842	0.647113	0.787789
F1	0.550893	0.420464	0.530669	0.403444	0.374076
Precision	0.500397	0.359212	0.497277	0.311471	0.469795
Recall	0.467114	0.347567	0.470438	0.314414	0.413607

Q213 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.733800	0.693102	0.779687	0.606728	0.766741
F1	0.516624	0.395274	0.529065	0.365930	0.366651
Precision	0.496232	0.355719	0.506221	0.299547	0.472049
Recall	0.455040	0.339772	0.477805	0.301878	0.409665

Q214 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.533774	0.515876	0.636236	0.472590	0.619166
F1	0.449014	0.409945	0.461073	0.386685	0.244696
Precision	0.443397	0.390396	0.430264	0.335806	0.384266
Recall	0.394663	0.354412	0.402974	0.275333	0.293865

Q215 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.578316	0.522696	0.701128	0.456748	0.696111
F1	0.504711	0.397316	0.517106	0.360886	0.317820
Precision	0.483367	0.323898	0.484129	0.246659	0.453272
Recall	0.408840	0.306052	0.456198	0.244251	0.370145

Q216 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.544719	0.532467	0.621528	0.425103	0.606918
F1	0.414923	0.379482	0.440185	0.371069	0.234906
Precision	0.403966	0.356478	0.417538	0.340375	0.376604
Recall	0.348624	0.315403	0.387896	0.256704	0.284698

Q217 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.641017	0.635656	0.707230	0.672136	0.681369
F1	0.410145	0.382889	0.453859	0.458631	0.263023
Precision	0.394579	0.364990	0.423151	0.414903	0.378077
Recall	0.358786	0.333419	0.400690	0.388978	0.304845

Q218 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.768889	0.747796	0.803503	0.705238	0.772922
F1	0.582663	0.492187	0.562936	0.452480	0.363678
Precision	0.540315	0.441226	0.521391	0.369786	0.459755
Recall	0.518236	0.432280	0.508376	0.374844	0.401469

Q219 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.805659	0.802264	0.796639	0.777497	0.782798
F1	0.612101	0.589365	0.431196	0.527774	0.291048
Precision	0.524049	0.499468	0.399476	0.479594	0.367241
Recall	0.523220	0.502221	0.379571	0.461167	0.321737

Q220 (1-3 range)

Metric	MF	I2I	KNN	MB	Mode
Accuracy	0.529380	0.529380	0.608656	0.610970	0.597911
F1	0.473737	0.473737	0.395502	0.459840	0.200798
Precision	0.429602	0.429602	0.372132	0.398018	0.335025
Recall	0.388056	0.388056	0.342488	0.364487	0.248847