ADS NUTRITION LABEL: HOME CREDIT DEFAULT RISK

Responsible Data Science Final Project Report Draft

BY JOSEPH SCHUMAN AND KRISTIN MULLANEY

Table of Contents

1. Background: General information about your chosen ADS?

1a. What is the purpose of this ADS? What are its stated goals?

Discriminatory lending practices have existed in the United States for hundreds of years. The effects of discriminatory lending have likely been most felt by the African American community [1]. However, these practices have also harmed other minorities, the poor, females, the young, and the old. Discriminatory lending practices have the power to damage a targeted individual's chances of being economically successful and therefore, can hinder their ability to be socioeconomically mobile. In this way, discriminatory lending helps the privileged class to remain privileged, while ensuring that the unprivileged class remains unprivileged. Over the years, several laws have been passed in an effort to curb discriminatory lending, such as the Equal Credit Opportunity Act and the Fair Housing Act. However, the problem still exists, albeit in less obvious forms than before [2,3].

For our course project we will look at an ADS which assesses the likelihood that borrowers (or prospective borrowers) will have difficulties in repaying their loans. We will examine the system for bias and report on our findings. In particular, we have selected one of the solutions from a 2018 Kaggle competition hosted by a company called Home Credit that was titled "*Home Credit Default Risk. Can you predict how capable each applicant is of repaying a loan?*" [5]. The competition attracted 7,176 teams and had 131,888 entries. The total prize money was $70,000. The link to the Kaggle competition is https://www.kaggle.com/competitions/home-credit-default-risk/overview. In the competition overview, Home Credit states the following about their goals:

> *Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience, Home Credit makes use of a variety of alternative data--including telco and transactional information--to predict their clients' repayment abilities.*

Interestingly, Home Credit states that their goal is to provide a "positive and safe borrowing experience" to the "unbanked population". Home Credit does not state that their goal is provide more loans to the "unbanked population". The goal posed to competition applicants was to build a model that can most accurately predict when someone will default on a loan payment. Importantly, the sole standard for determining the competition winner was the AUC score. The team with the highest AUC score was declared the winner (winning score = 0.80570).

1b. If the ADS has multiple goals, explain any trade-offs that these goals may introduce?

The ADS in the Kaggle solution that we examined followed the competition guidelines, and was designed with the sole goal of accurately predicting when someone would default on their payments. This goal is not in any way related to Home Credit's stated goal of promoting financial inclusion for the unbanked population. In fact, the ADS takes in a loan applicant's personal information and uses the repayment histories of similar applicants to judge if they are at high risk for default. This process inherently promotes the idea that you can judge an individual's actions by the actions of others in a similar demographic. In terms of containing and promoting bias, we will examine whether this particular ADS was 'healthy' (low bias) or 'unhealthy' (high bias) in the coming sections.

2. Input and Output

2a. Describe the data used by this ADS. How was this data collected or selected?

Home Credit provided the data for the Kaggle competition. The data provided consisted of 10 comma separated files, each serving a different purpose. According to Home Credit Group, the files contain information related to “application, demographic and historical credit behavior data.” The data does not contain information on the race of the applicant. It should be noted that none of the provided data contains any information about the date of the loan application. Therefore, we do not know the timeframe that was covered.

In general, the datasets provided are robust and appear to require minimal sanitizing. A number of competitors commented on the quality of the data. The training data includes information on 307,511 borrowers. The testing data includes information on 48,744 borrowers. The training and testing datasets contain 122 and 121 columns, respectively. None of the datasets include the borrower’s race or zip code. However, they do contain information such as: age, gender, education, and housing situation. Additionally, the datasets contain information that could be used as a proxy for race, such as “REGION_RATING_CLIENT”, which is a “rating of the region where client lives (1,2, 3)”. After joining data from the various datasets provided, there is a maximum of 220 columns available to build a model upon. One of the ten files provided is a data dictionary to help with data navigation. Below are brief descriptions for each of the provided datasets:

1. **application_trian.csv** – This is one of the two main tables. This is the table used for training. It contains the target value. It contains static data for all loan applications. One row represents one loan. The file contains 307,511 rows and 122 columns.

2. **application_test.csv** - This is one of the two main tables. This is the table to be used for testing. It does not contain the target value. It contains static data for all loan applications. One row represents one loan. The file contains 48,744 rows and 121 columns.

3. **bureau.csv** – Contains details about previous borrowings from other financial institutions to the applicant. The borrowings listed in the file were all reported to the “Credit Bureau”, and all the borrowings occurred prior to an application for a loan being filed with Home Credit Group. A definition of “Credit Bureau” was not provided. Each borrowing in this dataset is reported on a separate row. If an applicant had more than one previous borrowing, the file will contain more than one row for the applicant. The file contains 1,716,428 rows and 17 columns.

4. **bureau_balance.csv** – This file contains the monthly balances of previous credits that were reported to the Credit Bureau. This table has one row for each month of history for every previous credit reported to the Credit Bureau – i.e. the table has (#loans in sample * # of relative previous credits * # of months where there is an observable history) rows. The file contains 27,299,925 rows and 3 columns.

5. **POS_CASH_balance.csv** – This file contains monthly balance snapshots of previous loans that the applicant had with Home Credit Group. This table has one row for each month of history of every previous credit with Home Credit Group (consumer credit and cash loans) related to loans

in the sample – i.e. the table has (#loans in sample * # of relative previous credits * # of months in which there is a history observable for the previous credits) rows. The file contains 2,400,953 rows and 8 columns.

6. **credit_card_balance.csv** – This file contains monthly balances for previous credit cards loans that each applicant had with Home Credit Group. The table contains one row for each month where a previous credit (consumer credit and cash loans) was outstanding. Each credit card would count as a separate line of credit. In total, the number of rows in the table equates to the number of applicants * the number of credit cards for each applicant * the average number of months that there was an outstanding balance on a card. The file contains 3,009,991 rows and 23 columns.

7. **previous_application.csv** – This file contains information on any previous loan applications that an applicant in the dataset made with Home Credit Group. There is one row for each previous loan application. The file contains 1,670,214 rows and 37 columns.

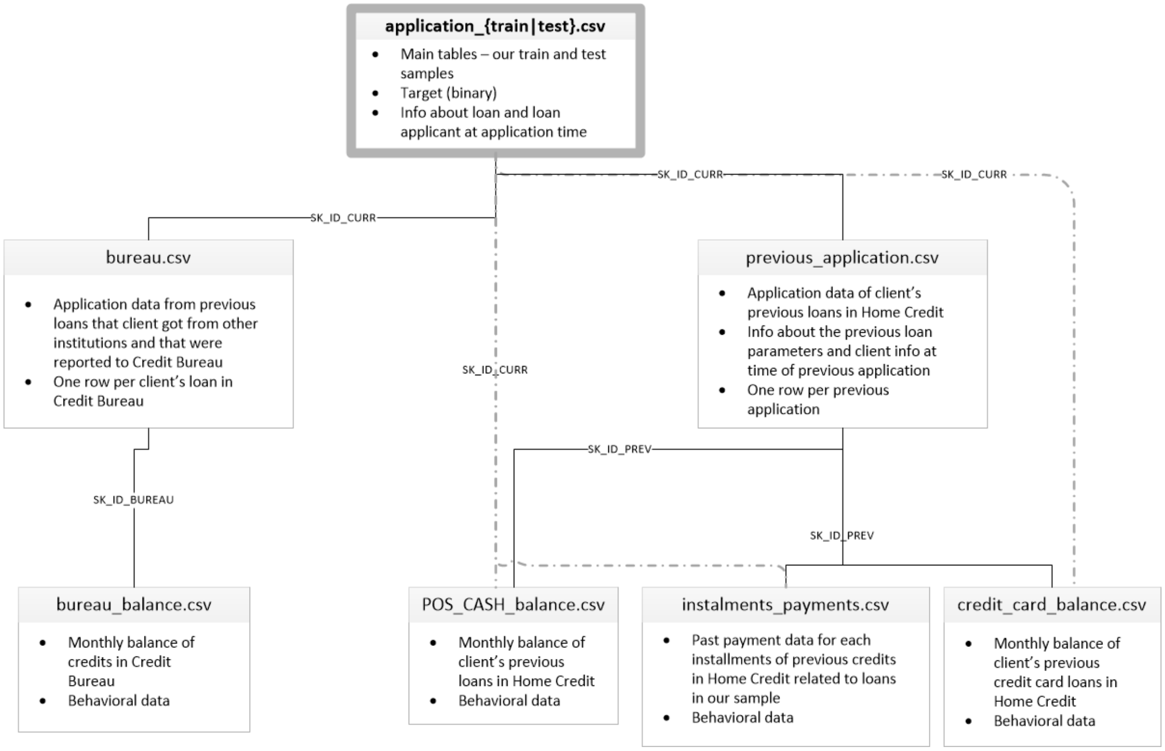
8. **installments_payments.csv** – This file contains repayment history for the previously disbursed credits by Home Credit related to the loans in the sample. There is: a) one row for every payment that was made, plus b) one row each for missed payment. One row is equivalent to one payment of one installment. This file contains 13,605,401 rows and 8 columns.

9. **HomeCredit_columns_description.csv** - This file contains descriptions for the columns in the various datasets. This file contains 220 rows and 5 columns.

10. **sample_submission.csv** – This file demonstrates the form that must be used for a submission to the Kaggle competition. This file contains 48,744 rows and 2 columns.

Below is a flow chart that was provided by Home Credit Group as part of the Kaggle competition. It shows the interaction of eight of the files. The flow chart does not have information on the sample submission (sample_submission.csv) or the data dictionary (HomeCredit_columns_description.csv).

In the diagram below, SK_ID_CURR connects the dataframes application_train and application_test with bureau, previous_application, POS_CASH_balance, installments_payment and credit_card_balance. SK_ID_PREV connects the dataframe previous_application with POS_CASH_balance, installments_payment and credit_card_balance. SK_ID_BUREAU connects the bureau with bureau_balance.


application_{train|test}.csv
Main tables – our train and test samples
Target (binary)
Info about loan and loan applicant at application time
SK_ID_CURR
SK_ID_CURR
SK_ID_CURR
SK_ID_CURR
bureau.csv
Application data from previous loans that client got from other institutions and that were reported to Credit Bureau
One row per client's loan in Credit Bureau
previous_application.csv
Application data of client's previous loans in Home Credit
Info about the previous loan parameters and client info at time of previous application
One row per previous application
SK_ID_PREV
SK_ID_BUREAU
SK_ID_PREV
bureau_balance.csv
Monthly balance of credits in Credit Bureau
Behavioral data
POS_CASH_balance.csv
Monthly balance of client's previous loans in Home Credit
Behavioral data
instalments_payments.csv
Past payment data for each installments of previous credits in Home Credit related to loans in our sample
Behavioral data
credit_card_balance.csv
Monthly balance of client's previous credit card loans in Home Credit
Behavioral data

In the diagram above, SK_ID_CURR connects the dataframes application_train and application_test with bureau, previous_application, POS_CASH_balance, installments_payment and credit_card_balance. SK_ID_PREV connects the dataframe previous_application with POS_CASH_balance, installments_payment and credit_card_balance. SK_ID_BUREAU connects the bureau with bureau_balance.

2b. For each input feature, describe its datatype, give information on missing values and on the value distribution. Show pairwise correlations between features if appropriate. Run any other reasonable profiling of the input that you find interesting and appropriate.

In this section we show a variety of information about the underlying data supplied by Home Credit Group. The goal is to give the reader an overview of the scope of the data, and an understanding of some of the important features that we will examine in detail as we create our nutritional label.

Datatypes:

The ADS used 8 of the 10 files provided by Home Credit Group. Two of the files were not used. One is "HomeCredit_columns_description.csv". This is the data dictionary. The other is "sample_submission.csv". This is a file that shows an example of the structure of the Kaggle submission. The 8 files that are used by the ADS contain 196 unique columns. The table below shows a summary of the datatypes.

Datatype	Count
float64	109
int64	54
object	33
Total	**196**

Appendix 1 is a complete data dictionary that includes the datatype for all 196 attributes. Below we show a table that provides a subset of the attributes. This group of attributes is of interest due to their potential for bias or likely high significance to the model. Consequently we show them individually here.

Selected Items from Data Dictionary

Index	Column Title	Description	Datatype
17	AMT_INCOME_TOTAL	Income of the client	float64
39	CNT_CHILDREN	Number of children the client has	int64
45	CNT_FAM_MEMBERS	How many family members does client have	float64
50	CODE_GENDER	Gender of the client	object
59	DAYS_BIRTH	Client's age in days at the time of application	int64
113	FLAG_OWN_CAR	Flag if the client owns a car	object
114	FLAG_OWN_REALTY	Flag if client owns a house or flat	object
142	NAME_EDUCATION_TYPE	Level of highest education the client achieved	object
143	NAME_FAMILY_STATUS	Family status of the client	object
145	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)	object
146	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,…)	object
166	OCCUPATION_TYPE	What kind of occupation does the client have	object
167	ORGANIZATION_TYPE	Type of organization where client works	object
177	REGION_POPULATION_RELATIVE	Normalized population of region where client lives (higher number means the client lives in more populated region)	float64
178	REGION_RATING_CLIENT	Our rating of the region where client lives (1,2,3)	int64
179	REGION_RATING_CLIENT_W_CITY	Our rating of the region where client lives with taking city into account (1,2,3)	int64

Missing values:

Below, we provide tables showing input features that were particularly sparse. We provide tables pertaining to 8 of the 10 data files. We do not provide a table for "HomeCredit_columns_description.csv" or for "sample_submission.csv".

File: application_train.csv

Column Title	Total Missing Values	Percent of Total
COMMONAREA_MEDI	214,865	69.87%
COMMONAREA_AVG	214,865	69.87%
COMMONAREA_MODE	214,865	69.87%
NONLIVINGAPARTMENTS_MODE	213,514	69.43%
NONLIVINGAPARTMENTS_MEDI	213,514	69.43%
NONLIVINGAPARTMENTS_AVG	213,514	69.43%
FONDKAPREMONT_MODE	210,295	68.39%
LIVINGAPARTMENTS_MEDI	210,199	68.35%
LIVINGAPARTMENTS_MODE	210,199	68.35%
LIVINGAPARTMENTS_AVG	210,199	68.35%

File: application_test.csv

Column Title	Total Missing Values	Percent of Total
COMMONAREA_MEDI	33,495	68.72%
COMMONAREA_AVG	33,495	68.72%
COMMONAREA_MODE	33,495	68.72%
NONLIVINGAPARTMENTS_MODE	33,347	68.41%
NONLIVINGAPARTMENTS_MEDI	33,347	68.41%
NONLIVINGAPARTMENTS_AVG	33,347	68.41%
FONDKAPREMONT_MODE	32,797	67.28%
LIVINGAPARTMENTS_AVG	32,780	67.25%
LIVINGAPARTMENTS_MEDI	32,780	67.25%
LIVINGAPARTMENTS_MODE	32,780	67.25%

File: bureau.csv

Column Title	Total Missing Values	Percent of Total
AMT_ANNUITY	1,226,791	71.47%
AMT_CREDIT_MAX_OVERDUE	1,124,488	65.51%
DAYS_ENDDATE_FACT	633,653	36.92%
AMT_CREDIT_SUM_LIMIT	591,780	34.48%
AMT_CREDIT_SUM_DEBT	257,669	15.01%
DAYS_CREDIT_ENDDATE	105,553	6.15%
AMT_CREDIT_SUM	13	0.00%

File: bureau_balance.csv

Column Title	Total Missing Values	Percent of Total
*** No Missing Values ***		

File: POS_CASH_balance.csv

Column Title	Total Missing Values	Percent of Total
CNT_INSTALMENT_FUTURE	4,117	0.17%
CNT_INSTALMENT	4,117	0.17%
SK_DPD_DEF	1	0.00%
SK_DPD	1	0.00%
NAME_CONTRACT_STATUS	1	0.00%

File: credit_card_balance.csv		
Column Title	Total Missing Values	Percent of Total
AMT_PAYMENT_CURRENT	612,362	20.34%
CNT_DRAWINGS_POS_CURRENT	599,146	19.91%
CNT_DRAWINGS_OTHER_CURRENT	599,146	19.91%
CNT_DRAWINGS_ATM_CURRENT	599,146	19.91%
AMT_DRAWINGS_ATM_CURRENT	599,145	19.91%
AMT_DRAWINGS_OTHER_CURRENT	599,145	19.91%
AMT_DRAWINGS_POS_CURRENT	599,145	19.91%
CNT_INSTALMENT_MATURE_CUM	226,127	7.51%
AMT_INST_MIN_REGULARITY	226,127	7.51%
AMT_PAYMENT_TOTAL_CURRENT	1	0.00%
AMT_TOTAL_RECEIVABLE	1	0.00%
SK_DPD	1	0.00%
NAME_CONTRACT_STATUS	1	0.00%
CNT_DRAWINGS_CURRENT	1	0.00%
SK_DPD_DEF	1	0.00%
AMT_RECIVABLE	1	0.00%
AMT_RECEIVABLE_PRINCIPAL	1	0.00%

File: previous_application.csv		
Column Title	Total Missing Values	Percent of Total
RATE_INTEREST_PRIVILEGED	1,664,263	99.64%
RATE_INTEREST_PRIMARY	1,664,263	99.64%
RATE_DOWN_PAYMENT	895,844	53.64%
AMT_DOWN_PAYMENT	895,844	53.64%
NAME_TYPE_SUITE	820,405	49.12%
DAYS_TERMINATION	673,065	40.30%
NFLAG_INSURED_ON_APPROVAL	673,065	40.30%
DAYS_FIRST_DRAWING	673,065	40.30%
DAYS_FIRST_DUE	673,065	40.30%
DAYS_LAST_DUE_1ST_VERSION	673,065	40.30%
DAYS_LAST_DUE	673,065	40.30%
AMT_GOODS_PRICE	385,515	23.08%
AMT_ANNUITY	372,235	22.29%
CNT_PAYMENT	372,230	22.29%
PRODUCT_COMBINATION	346	0.02%
AMT_CREDIT	1	0.00%

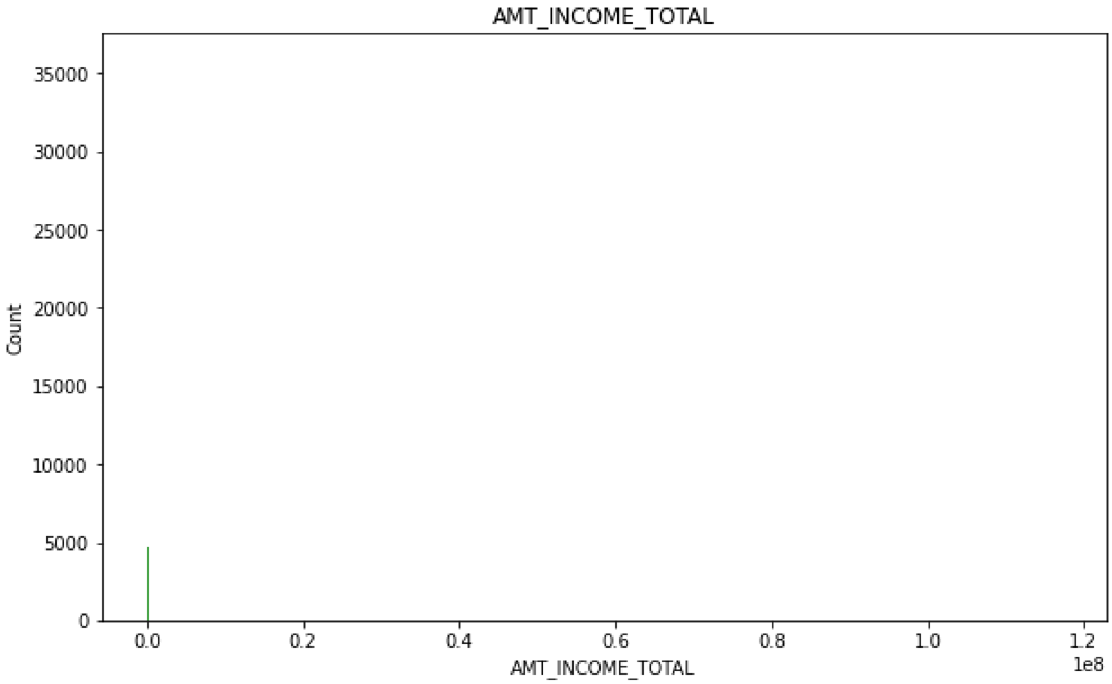
File: installments_payments.csv		
Column Title	Total Missing Values	Percent of Total
AMT_PAYMENT	2,905	2.14%
DAYS_ENTRY_PAYMENT	2,905	2.14%

Distribution of Values:

Below, we have created distribution plots for each of the 16 features that were previously identified as being the most significant features in the model. The plots are based on the distributions of the training data.

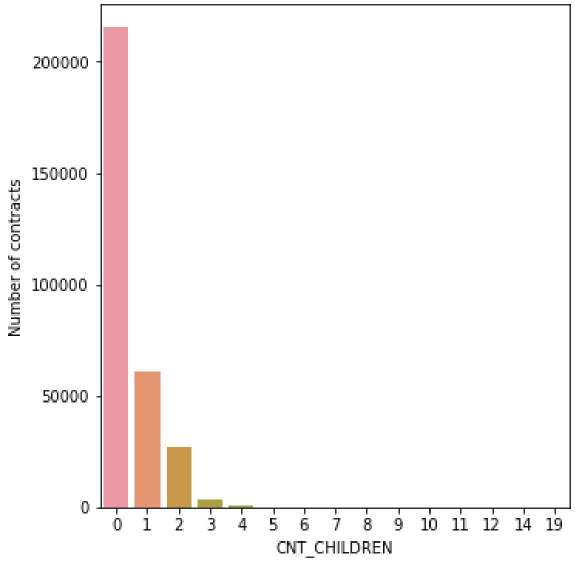
1. Total Income – Attribute name: AMT_INCOME_TOTAL

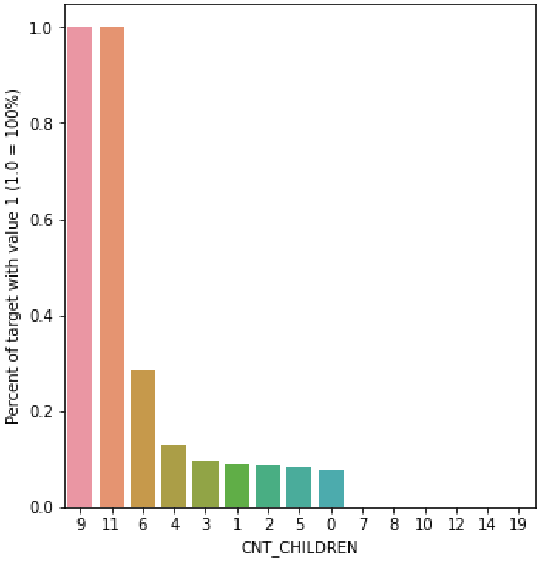
The distribution of total income is so broad that the histogram does not adequately portray the distribution of values. The median income is $147,150 and the mean is $168,798. However, the minimum income is $25,650 while the maximum income is $117 million.



2. Number of children – Attribute name: CNT_CHILDREN

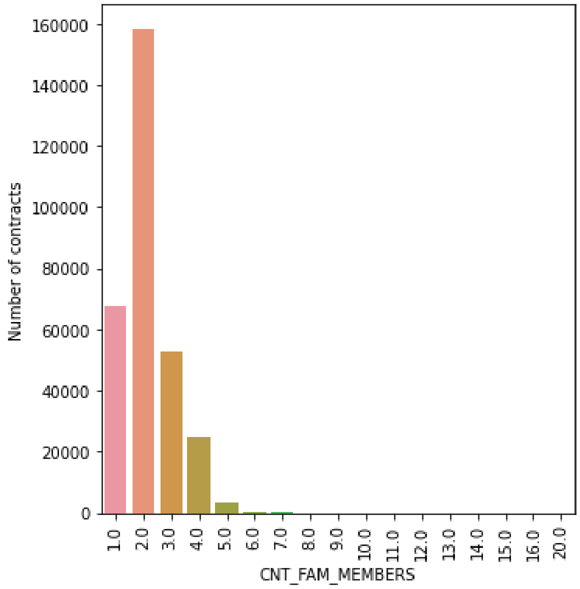
This attribute reports the number of children that the client has. The population is heavily skewed towards clients with 3 children or less. This group accounts for 99.8% of all clients. However, the training data contains information on 8 clients that have between 11 and 19 children. The population is heavily skewed towards clients with 3 children or less. This group accounts for 99.8% of all clients. We provided an additional table to provide more detailed information on this attribute. The table on the left demonstrates the distribution of past clients by number of children. The table on the right demonstrates what percentage of past clients with that number of children defaulted on their loans. 100% of the clients in the training data with 9 or 11 children defaulted on their loans. The final table illustrates the same information as the right histogram, but in a different way. This final table makes it clear that clients with more than three children actually had a low rate of default overall, and that the second histogram is simply overrepresenting outliers.

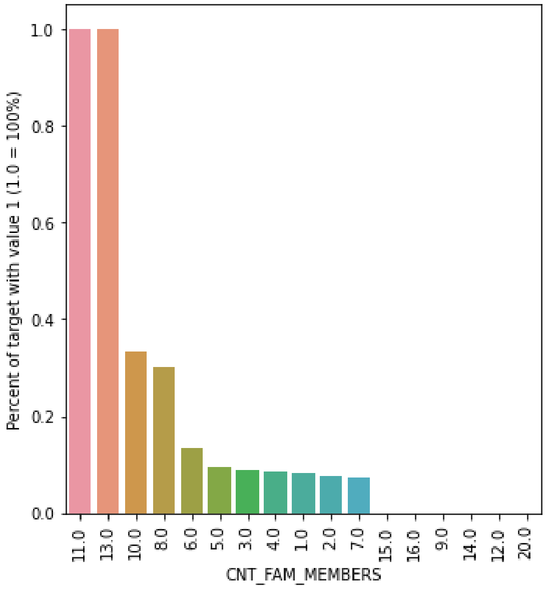




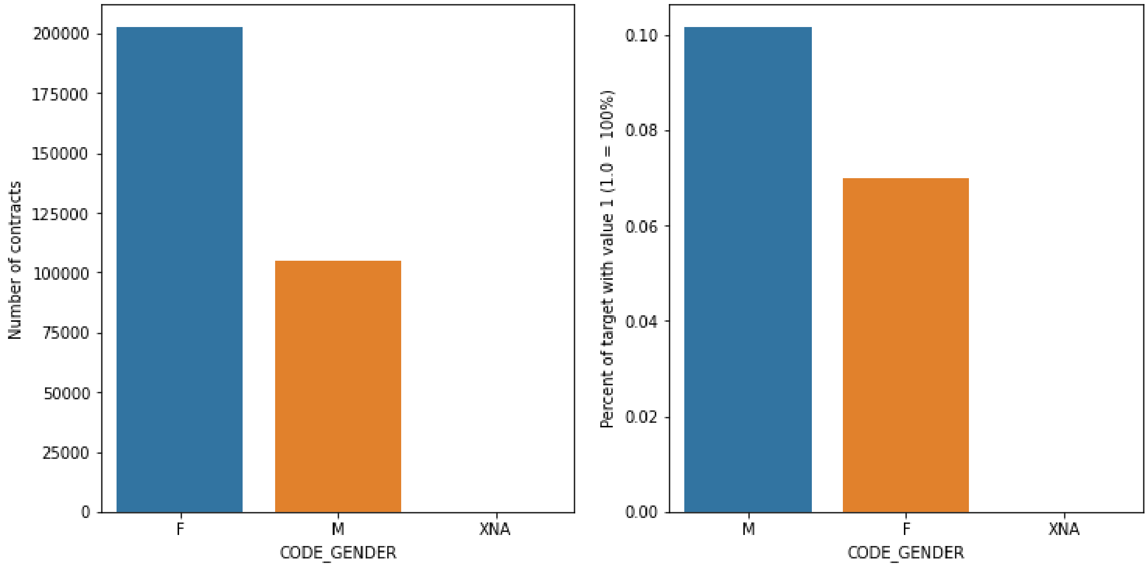
Count of Children	Count	Target = 1	%
0	215,371	16,609	7.71%
1	61,119	5,454	8.92%
2	53,498	2,333	4.36%
3	3,717	358	9.63%
>3	2,480	71	2.86%

3. Number of family members – Attribute name: CNT_FAM_MEMBERS



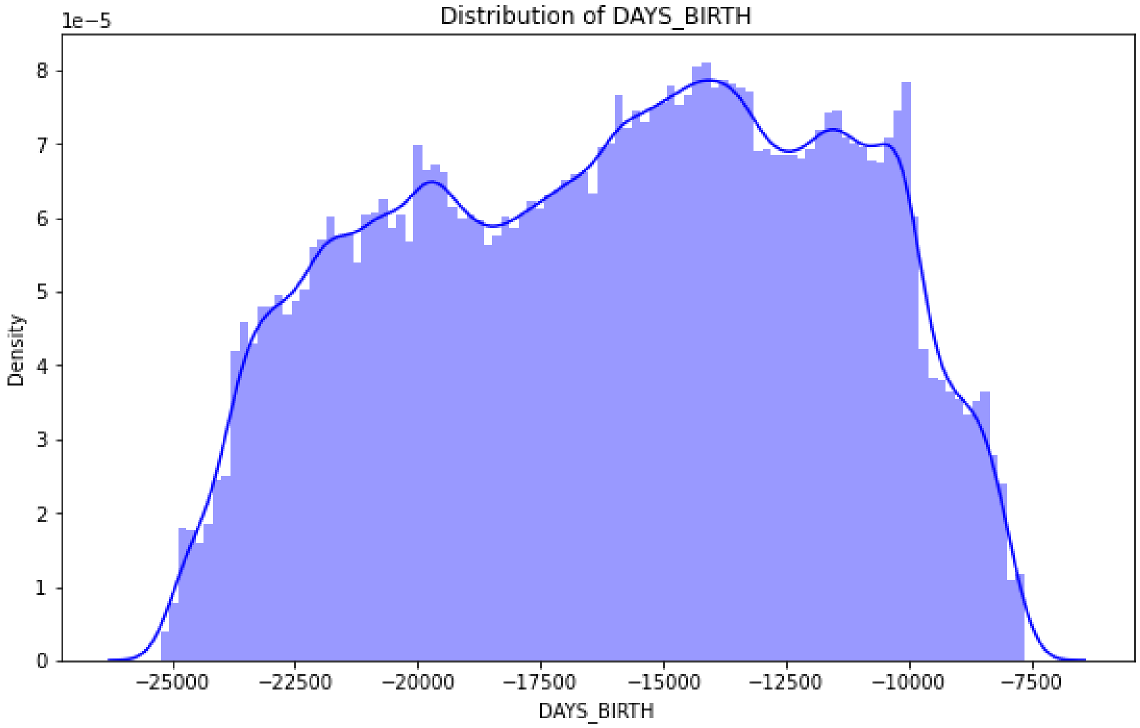


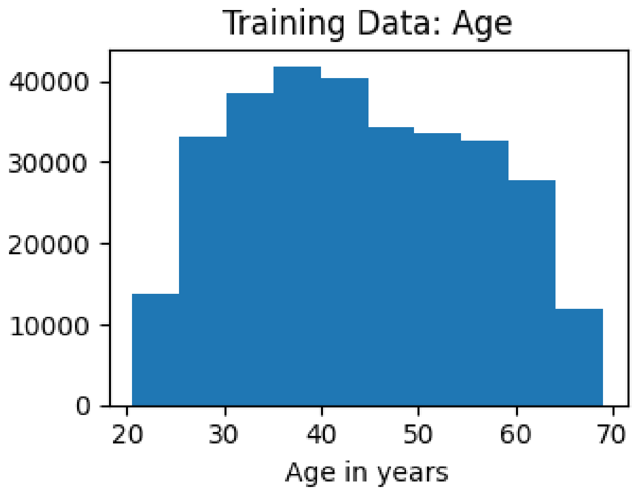
4. Gender – Attribute name: CODE_GENDER



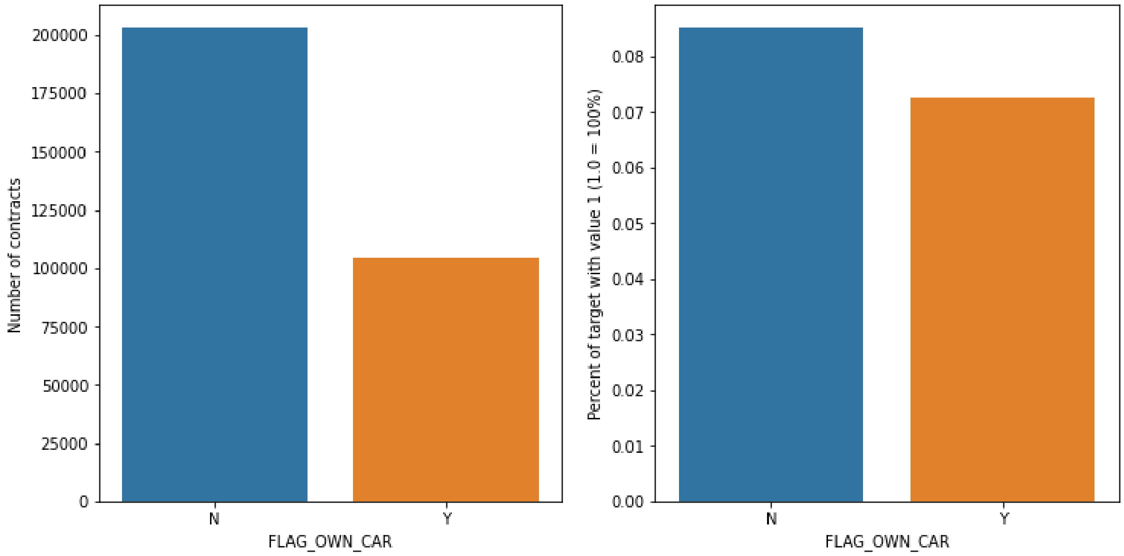
5. Age – Attribute Name: DAYS_BIRTH

The age of the client is expressed as the negative of the number of days that the client was alive as of the date of the loan application. We have included an additional table that converts the age from days into years and puts the age into buckets.

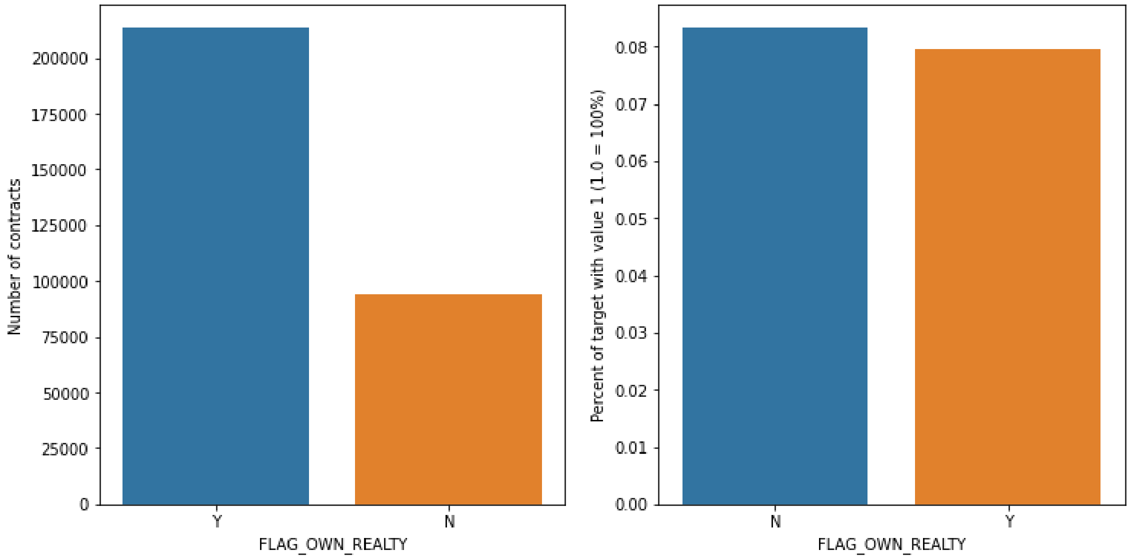




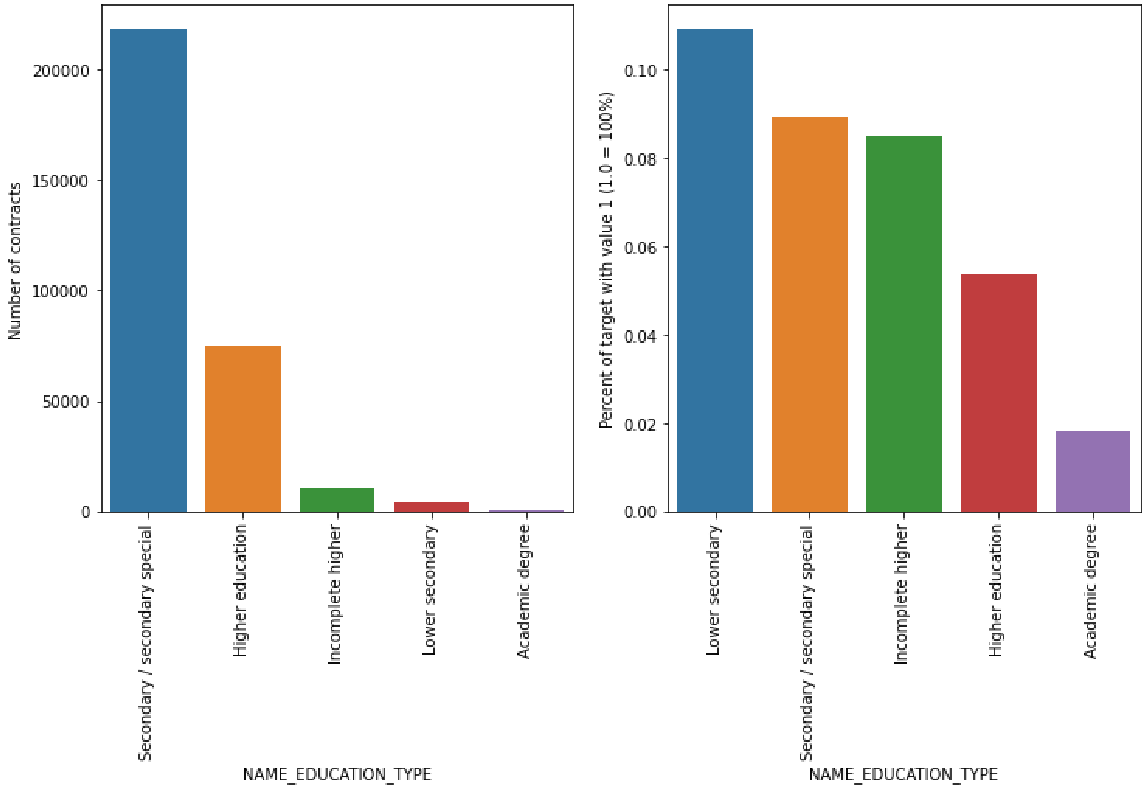
6. Does the client own a car – Attribute name: FLAG_OWN_CAR



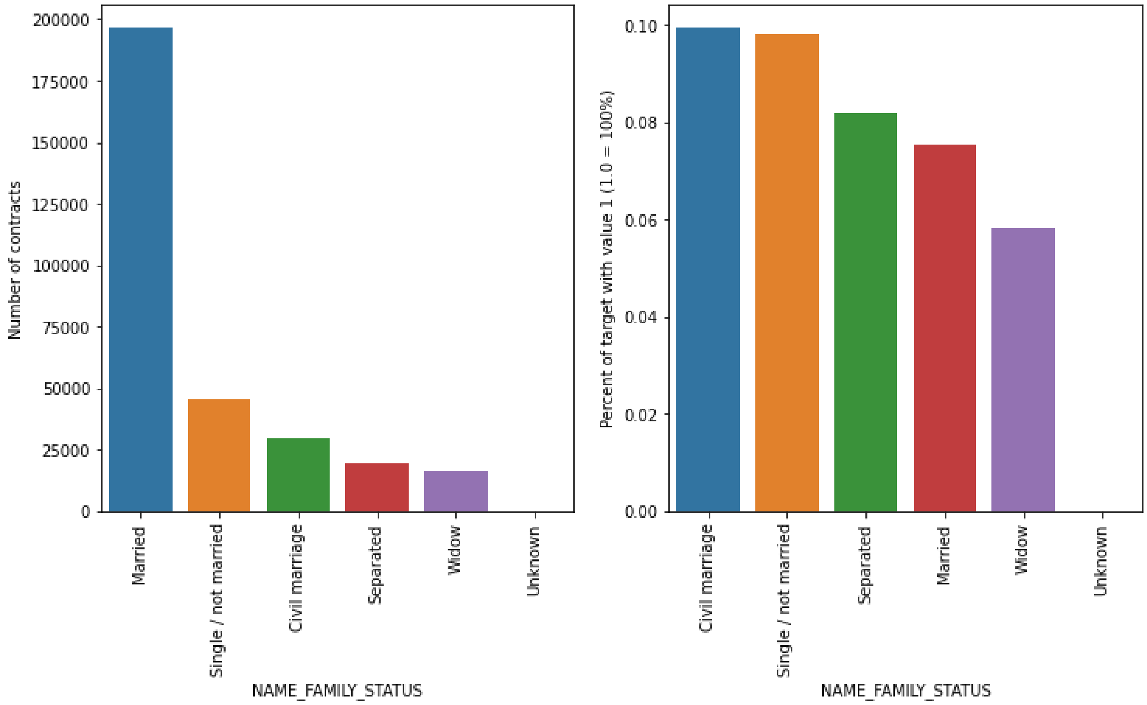
7. Does the client own property – Attribute name: FLAG_OWN_REALTY



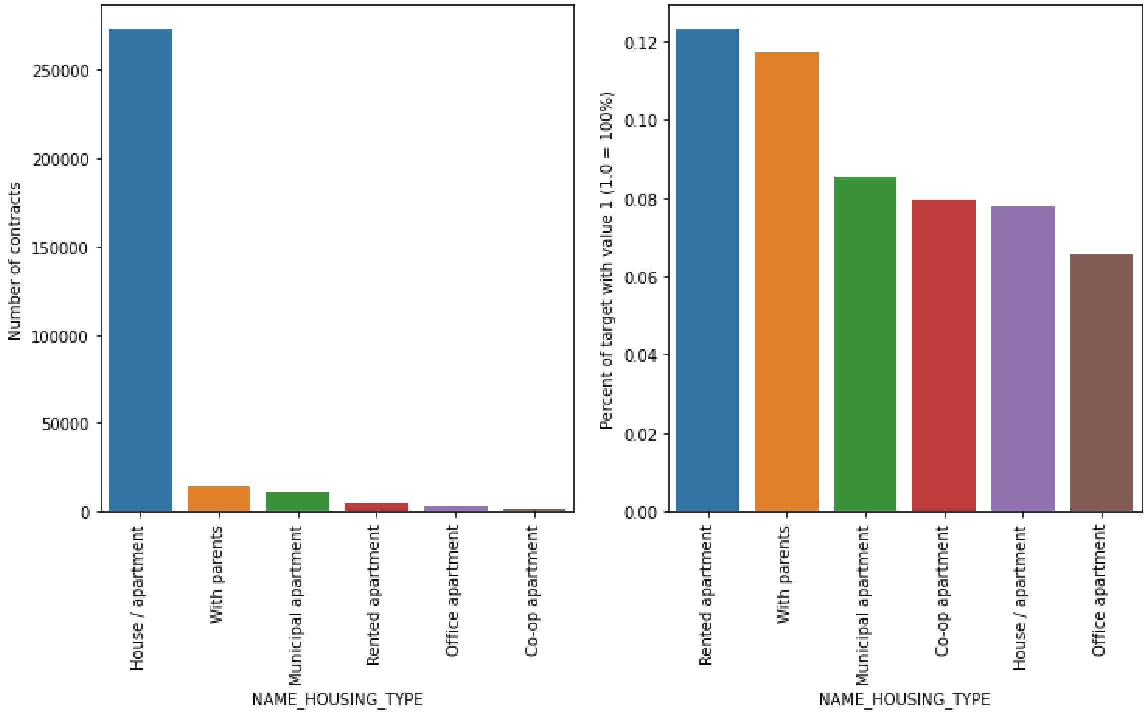
8. Education level of client – Attribute name: NAME_EDUCATION_TYPE



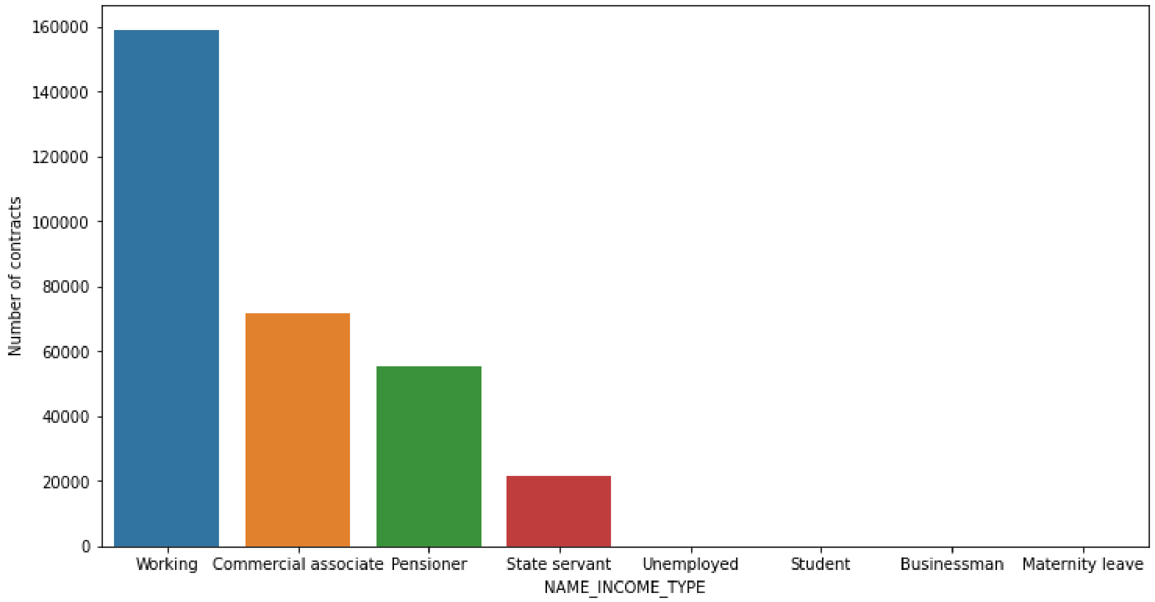
9. Family status of the client – Attribute name: NAME_FAMILY_STATUS

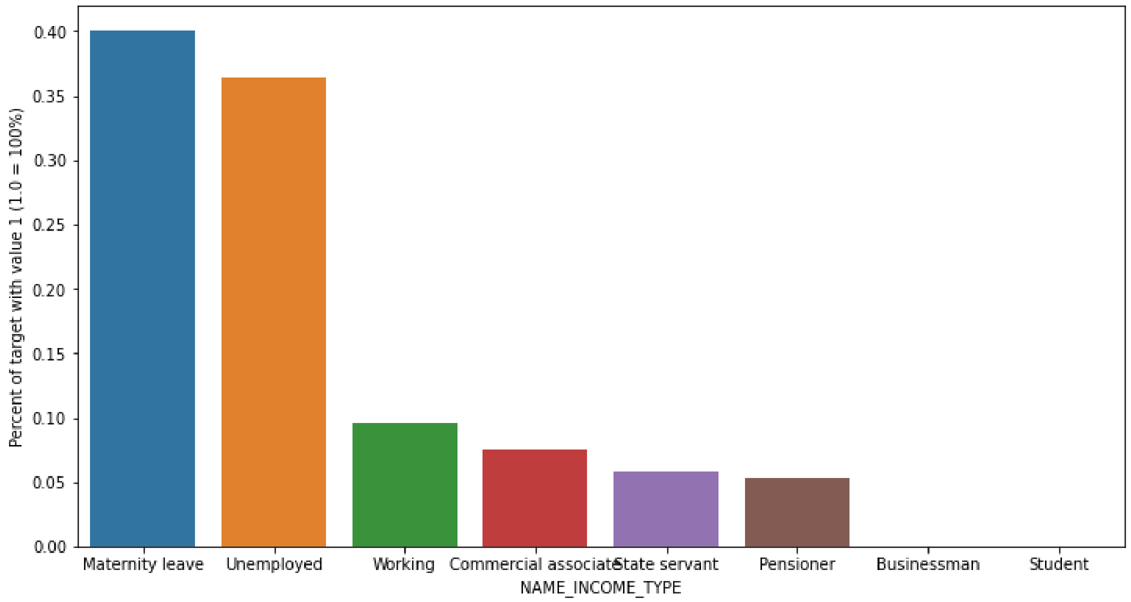


10. What type of home does the client live in – Attribute name: NAME_HOUSING_TYPE

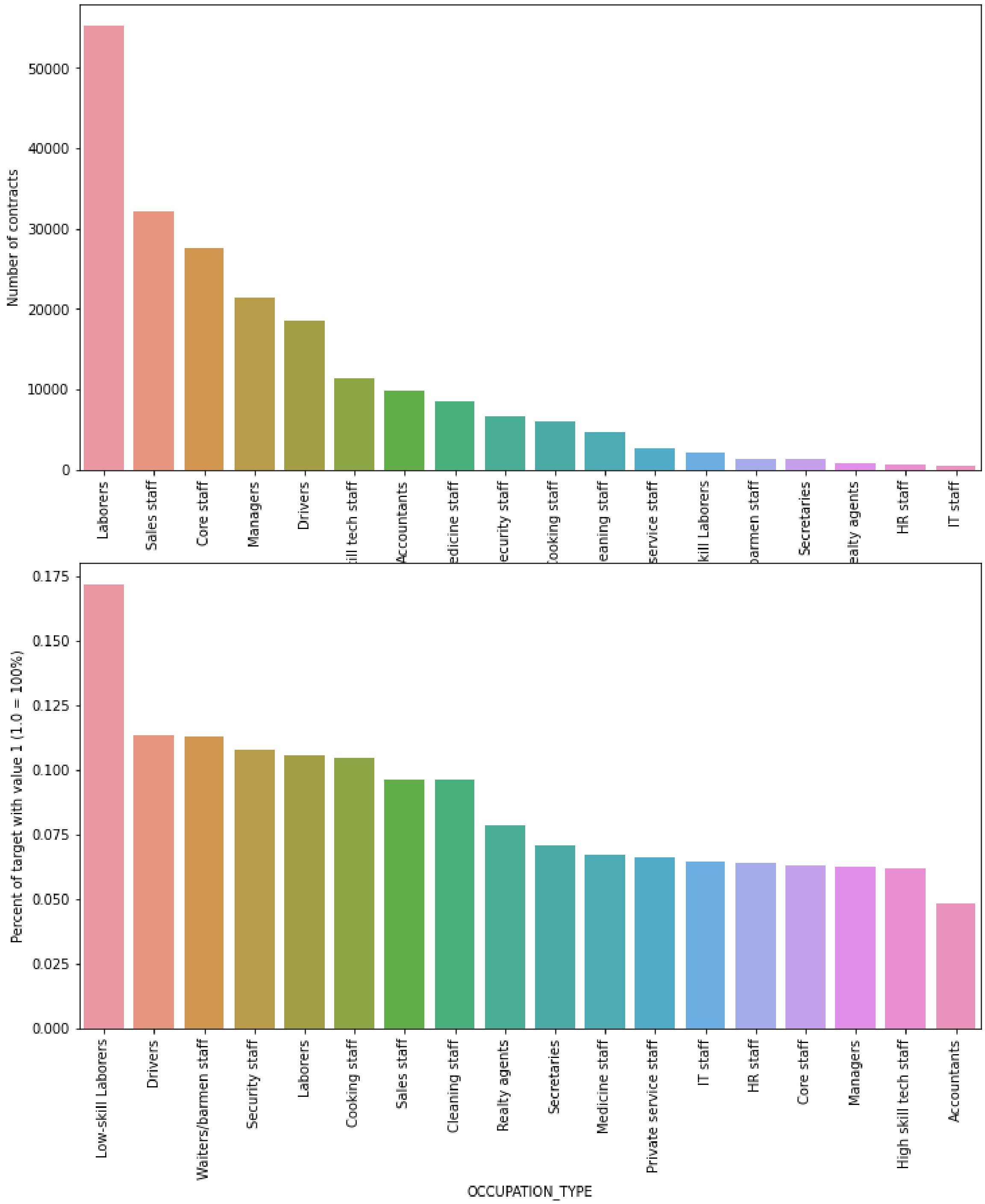


11. Source of client's income - Attribute name: NAME_INCOME_TYPE

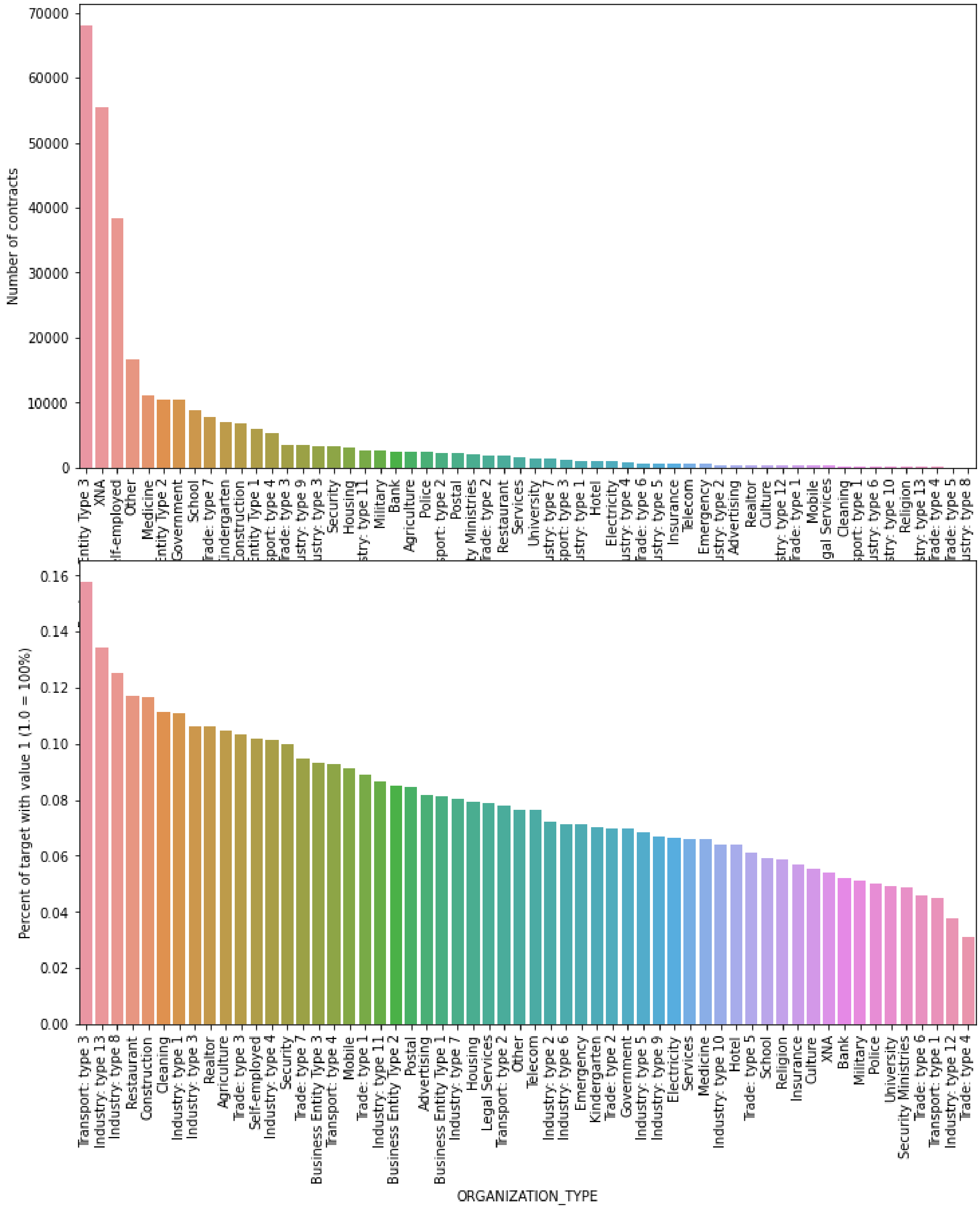




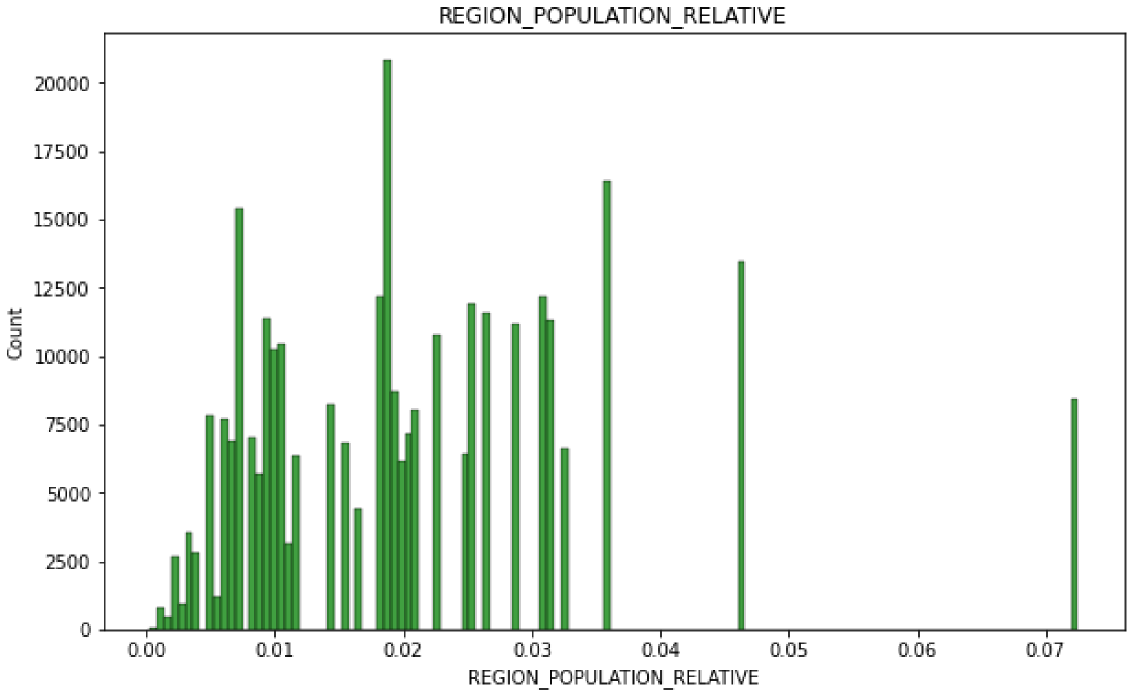
12. Client’s occupation: Attribute name - OCCUPATION_TYPE



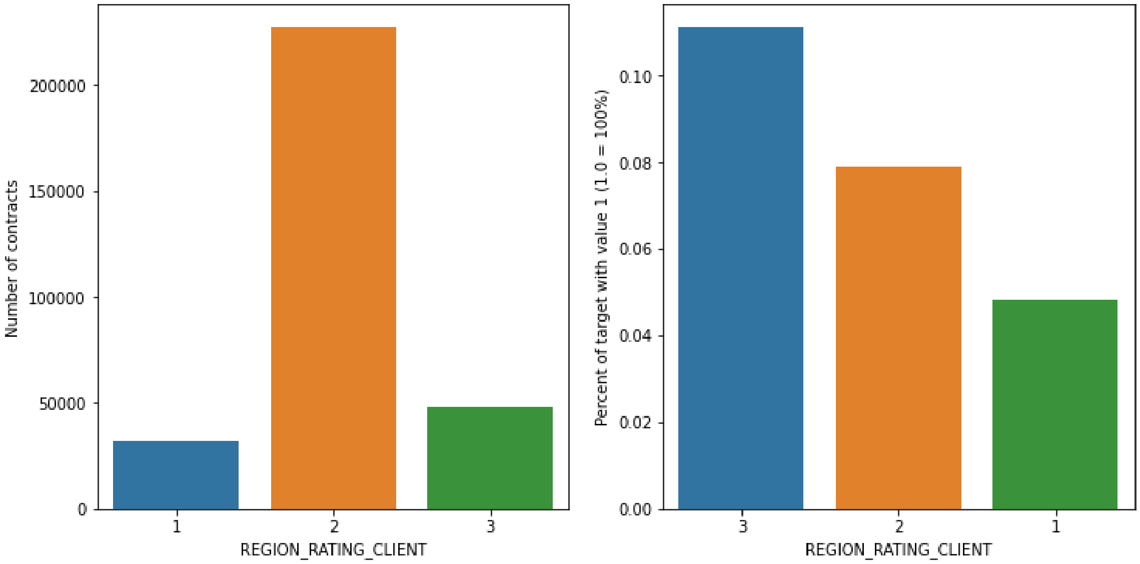
13. Type of business where client works – Attribute name: ORGANIZATION_TYPE



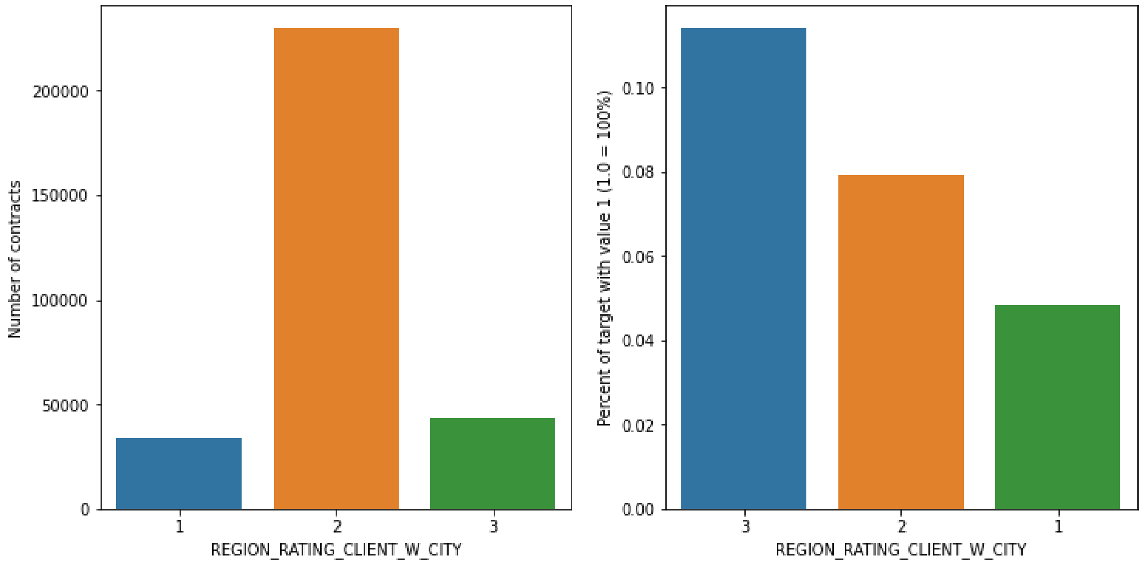
14. Population Density where client lives – Attribute name: REGION_POPULATION_RELATIVE



15. Rating of where client lives - Attribute name: REGION_RATING_CLIENT



16. More detailed rating of where client lives - Attribute name: REGION_RATING_CLIENT_W_CITY



2c. What is the output of the system (e.g. is it a class label, a score, a probability, or some other type of output), and how do we interpret it?

The goal of the ADS is to predict a “clients' repayment abilities”. It appears that Home Credit Group was attempting to improve their ability to: 1) asses the ability to repay of new loan applicants, and 2) assess the ability to repay of borrowers with outstanding loans. Specifically, for the Kaggle competition, the competitors were required “for each SK_ID_CURR“ to “predict a probability for the TARGET variable”. The “SK_ID_CURR” is the ID of the loan in the sample set. The “TARGET” is a field in the training data that has a datatype of int64. It contains values of 1 or 0. A value of 1 indicates that a “client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan” in the sample. A value of zero is used to indicate “all other cases”.

The specific probability that the ADS is predicting is the probability that the TARGET will be equal to 1.0. Essentially this means that if the model predicts that the client will have payment difficulties the model will assign a high probability (e.g. 0.9) that the TARGET will be 1.0. If the model predicts that the client will not have payment difficulties, the model will assign a low probability (e.g. 0.2) that the TARGET will be 1.0. The table below demonstrates the format of the required submission file. The ADS prediction is labeled as “TARGET” in this file. It has a datatype of float64.

SK_ID_CURR	TARGET
100001	0.1
100005	0.9
100013	0.2
etc.	

Unfortunately, the field name “TARGET” appears twice in the dataset. It appears in the required submission file, “sample_submission.csv”. It also appears in the training data, “application_trian.csv”. The TARGET field has two different meanings and two different datatypes in the two different files. In the training data, the value represents whether the client has had payment difficulties and has integer values of either 0 or 1. In the submission file, the TARGET field is an assessment of the likelihood that client will have payment difficulties. It had a datatype of float64 and takes on a value between 0.0 and 1.0.

3. Implementation and Validation: Present your understanding of the code that implements the ADS. This code was implemented by others (e.g. as part of the Kaggle competition), not by you as part of this assignment. You goal here is to demonstrate that you understand the implementation at a high level.

The submissions to the Kaggle were evaluated on area under the ROC curve between the predicted probability and the observed target. There were 7,180 submissions to Kaggle with the first place team achieving an ROC of 0.8057. The solution that we have implemented was submitted by a Kaggle user with the screen name "SANGSEOSEO". The model achieved an ROC of 0.79977 which placed within the top 100 submissions of the 7,180 submissions. The code is broken into sections and executed sequentially. We will describe each section.

a. Column data type conversions to reduce memory allocation – Numpy is used to convert datatypes. This initial work was done for performance reasons.

b. Feature engineering - Feature engineering is the process of selecting, manipulating, and transforming raw data into useable features. This section includes data cleansing. It is also the longest section of code. It involves multiple actions:

    1. Creation of Ratios - For example, an "income ratio" is created. The feature is named "APPS_ANNUITY_INCOME_RATIO" is added. This is equal to the amount of the loan payment ("AMT_ANNUITY") divided by the borrowers annual income ("AMT_INCOME_TOTAL").
    2. Management of missing values and bad data – This code does not insert values where values are missing. It does identify certain bad values and remove them. It also ensures that missing values will not prevent calculations and programs from running properly.
    3. Derive information and pre-process – This section derives information from the data. For example, it attempts to determine the interest rate on a borrower's loan from the loan balance, payment size, and number of payments. The code also creates many of the functions that will be used later to gather information from various files.

c. Aggregation of information regarding previous loans, installment payments, and credit cards – This section creates ratios and flags that will be used later. For example, the code assigns loans to different buckets based on whether the loan is past due and, if so, by how long it is past due. This section also establishes ranges for certain features by determining the mean, max, and min.

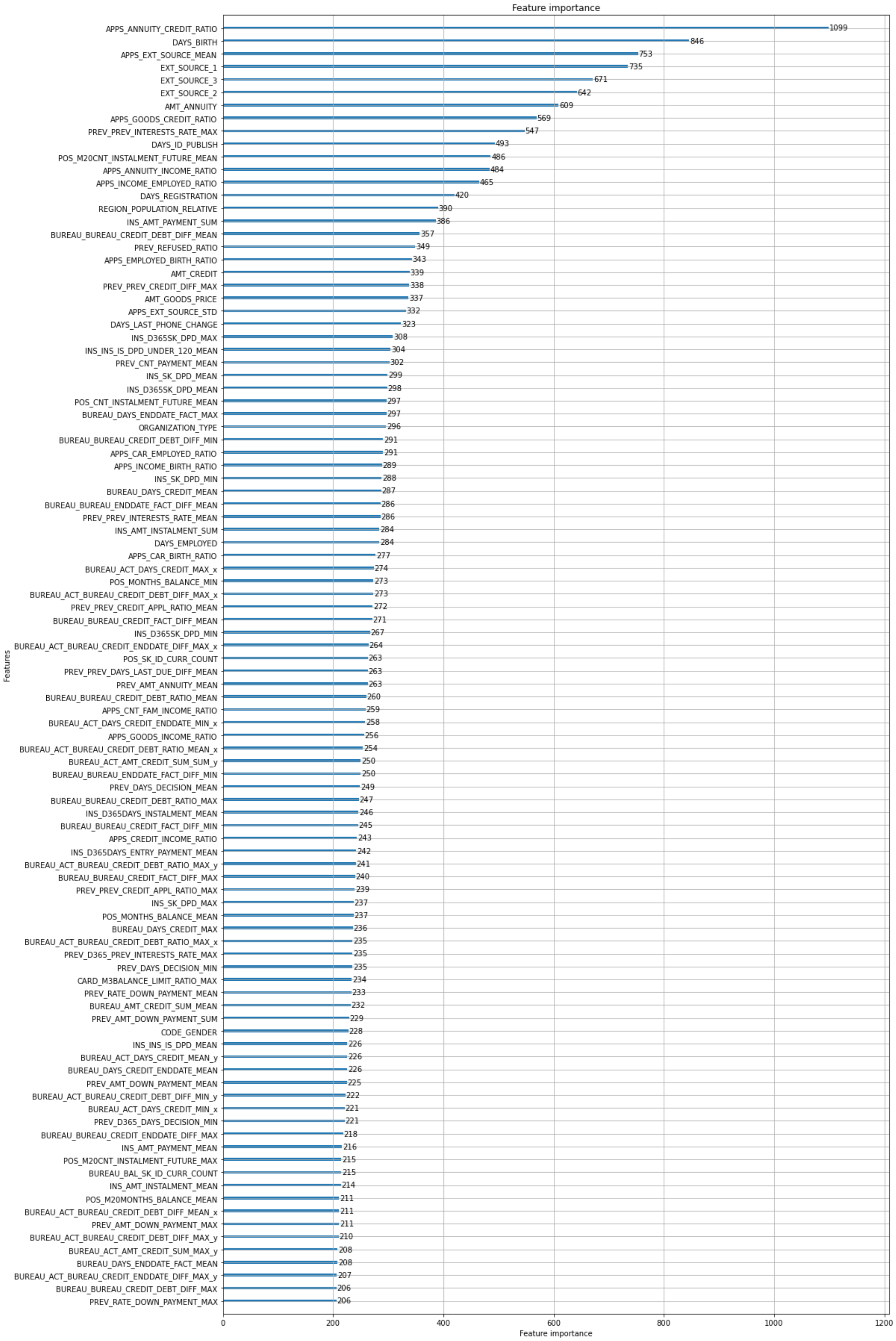
d. Datasets concatenation and Join – This is the final step in preparation of the data. All of the functions that have been created to process and transform the data are run. Then the data is joined based on the loan ID ("K_ID_CURR"). At this point the data is ready to be used by the models.

e. Data split and model fit - In this section the data is split and the model is trained. The code is very short and simply calls functions that were previously created. The train test split is 70/30. In this section a function called “train_apps_all” is run. This is the function that fits the model. The model that is used is a gradient boost decision tree. Specifically, the code imports the Python “lightgbm” package. This package contains functions to perform tree based learning algorithms based on gradient boosting. The specific function used is “LGBMClassifier”. This classifier is specifically designed for distributed processing with the advantages of being efficient, has low memory usage, good accuracy, and capable of handling large data sets. The model is not advised for use with small datasets because it is sensitive to overfitting. Below we have included the exact function and parameters that were implemented.

```
clf = LGBMClassifier(
        nthread=4,
        n_estimators=2000,
        learning_rate=0.02,
        max_depth = 11,
        num_leaves=58,
        colsample_bytree=0.613,
        subsample=0.708,
        max_bin=407,
        reg_alpha=3.564,
        reg_lambda=4.930,
        min_child_weight= 6,
        min_child_samples=165,
        silent=-1,
        verbose=-1,
        )
```

The classifier has multiple parameters, one of which is the method of gradient boosting. This model uses the parameter ‘gbdt’, which is describes as a “traditional Gradient Boosting Decision Tree”. The selection of “gbdt” is not specifically stated in the code above. It is the default for the model when no other selection is made. The model uses “n_estimators=2000”, this is the number of boosted trees to fit. The learning rate is 0.02, the default is 0.1. The learning rate is a hyper-parameter that controls how much the weights of the network are adjusted with respect the loss gradient. “reg_alpha” is the L1 regularization term on weights. “reg_lambda” is the L2 regularization term on weights. The model returns an AUC score.

f. Plot importance of features – The final step of the model is that creates a plot of feature importance. The resulting plot of feature importance is included below.


Feature importance
APPS_ANNUITY_CREDIT_RATIO 1099
DAYS_BIRTH 846
APPS_EXT_SOURCE_MEAN 753
EXT_SOURCE_1 735
EXT_SOURCE_3 671
EXT_SOURCE_2 642
AMT_ANNUITY 609
APPS_GOODS_CREDIT_RATIO 569
PREV_PREV_INTERESTS_RATE_MAX 547
DAYS_ID_PUBLISH 493
POS_M20CNT_INSTALMENT_FUTURE_MEAN 486
APPS_ANNUITY_INCOME_RATIO 484
APPS_INCOME_EMPLOYED_RATIO 465
DAYS_REGISTRATION 420
REGION_POPULATION_RELATIVE 390
INS_AMT_PAYMENT_SUM 386
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MEAN 357
PREV_REFUSED_RATIO 349
APPS_EMPLOYED_BIRTH_RATIO 343
AMT_CREDIT 339
PREV_PREV_CREDIT_DIFF_MAX 338
AMT_GOODS_PRICE 337
APPS_EXT_SOURCE_STD 332
DAYS_LAST_PHONE_CHANGE 323
INS_D365SK_DPD_MAX 308
INS_INS_IS_DPD_UNDER_120_MEAN 304
PREV_CNT_PAYMENT_MEAN 302
INS_SK_DPD_MEAN 299
INS_D365SK_DPD_MEAN 298
POS_CNT_INSTALMENT_FUTURE_MEAN 297
BUREAU_DAYS_ENDDATE_FACT_MAX 297
ORGANIZATION_TYPE 296
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MIN 291
APPS_CAR_EMPLOYED_RATIO 291
APPS_INCOME_BIRTH_RATIO 289
INS_SK_DPD_MIN 288
BUREAU_DAYS_CREDIT_MEAN 287
BUREAU_BUREAU_ENDDATE_FACT_DIFF_MEAN 286
PREV_PREV_INTERESTS_RATE_MEAN 286
INS_AMT_INSTALMENT_SUM 284
DAYS_EMPLOYED 284
APPS_CAR_BIRTH_RATIO 277
BUREAU_ACT_DAYS_CREDIT_MAX_x 274
POS_MONTHS_BALANCE_MIN 273
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MAX_x 273
PREV_PREV_CREDIT_APPL_RATIO_MEAN 272
BUREAU_BUREAU_CREDIT_FACT_DIFF_MEAN 271
INS_D365SK_DPD_MIN 267
BUREAU_ACT_BUREAU_CREDIT_ENDDATE_DIFF_MAX_x 264
POS_SK_ID_CURR_COUNT 263
PREV_PREV_DAYS_LAST_DUE_DIFF_MEAN 263
PREV_AMT_ANNUITY_MEAN 263
BUREAU_BUREAU_CREDIT_DEBT_RATIO_MEAN 260
APPS_CNT_FAM_INCOME_RATIO 259
BUREAU_ACT_DAYS_CREDIT_ENDDATE_MIN_x 258
APPS_GOODS_INCOME_RATIO 256
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MEAN_x 254
BUREAU_ACT_AMT_CREDIT_SUM_SUM_y 250
BUREAU_BUREAU_ENDDATE_FACT_DIFF_MIN 250
PREV_DAYS_DECISION_MEAN 249
BUREAU_BUREAU_CREDIT_DEBT_RATIO_MAX 247
INS_D365DAYS_INSTALMENT_MEAN 246
BUREAU_BUREAU_CREDIT_FACT_DIFF_MIN 245
APPS_CREDIT_INCOME_RATIO 243
INS_D365DAYS_ENTRY_PAYMENT_MEAN 242
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MAX_y 241
BUREAU_BUREAU_CREDIT_FACT_DIFF_MAX 240
PREV_PREV_CREDIT_APPL_RATIO_MAX 239
INS_SK_DPD_MAX 237
POS_MONTHS_BALANCE_MEAN 237
BUREAU_DAYS_CREDIT_MAX 236
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MAX_x 235
PREV_D365_PREV_INTERESTS_RATE_MAX 235
PREV_DAYS_DECISION_MIN 235
CARD_M3BALANCE_LIMIT_RATIO_MAX 234
PREV_RATE_DOWN_PAYMENT_MEAN 233
BUREAU_AMT_CREDIT_SUM_MEAN 232
PREV_AMT_DOWN_PAYMENT_SUM 229
CODE_GENDER 228
INS_INS_IS_DPD_MEAN 226
BUREAU_ACT_DAYS_CREDIT_MEAN_y 226
BUREAU_DAYS_CREDIT_ENDDATE_MEAN 226
PREV_AMT_DOWN_PAYMENT_MEAN 225
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MIN_y 222
BUREAU_ACT_DAYS_CREDIT_MIN_x 221
PREV_D365_DAYS_DECISION_MIN 221
BUREAU_BUREAU_CREDIT_ENDDATE_DIFF_MAX 218
INS_AMT_PAYMENT_MEAN 216
POS_M20CNT_INSTALMENT_FUTURE_MAX 215
BUREAU_BAL_SK_ID_CURR_COUNT 215
INS_AMT_INSTALMENT_MEAN 214
POS_M20MONTHS_BALANCE_MEAN 211
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MEAN_x 211
PREV_AMT_DOWN_PAYMENT_MAX 211
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MAX_y 210
BUREAU_ACT_AMT_CREDIT_SUM_MAX_y 208
BUREAU_DAYS_ENDDATE_FACT_MEAN 208
BUREAU_ACT_BUREAU_CREDIT_ENDDATE_DIFF_MAX_y 207
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MAX 206
PREV_RATE_DOWN_PAYMENT_MAX 206
Features
0
200
400
600
800
1000
1200
Feature importance

4. Outcomes

4a. Analyze the effectiveness of the ADS

The overall accuracy of the model was 79.31%. This is reasonably high and indicates that the ADS performs reasonably well. We also examined the effectiveness of the ADS by comparing its performance across different subpopulations. We wanted to get a sense of the degree of consistency or variability of the ADS across subpopulations. We had originally identified 16 features that were of interest. We have reviewed those subpopulations for accuracy and have included a table below with accuracy scores for a selection of the subpopulations.

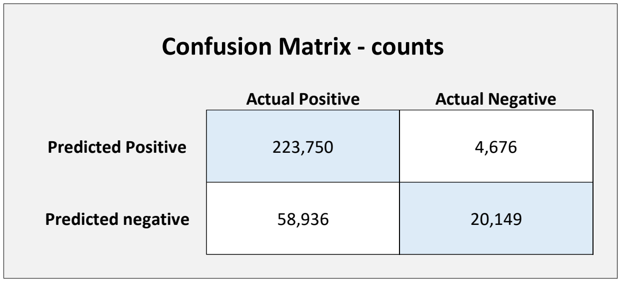
As we will discuss below, after further analysis, we have identified two features that are of particular interest due to the potential for bias. Those two features are CODE_GENDER and REGION_RATING_CLIENT. CODE_GENDER informs us as to the gender of the loan applicant. REGION_RATING_CLIENT is a subjective rating given by Home Credit Group to the area where the loan applicant lives. In the table below we have included in bold the accuracy measures for the subpopulations related to these two features. The table includes a total of 34 subpopulations. The lowest accuracy level was 62.4% for the 20 to 25 year old age bucket for loan applicants. It is not surprising that the 20 to 25 year old age group has the lowest accuracy. The model predicts that a higher percentage of applicants in that age group should be approved for loans. This is consistent with our concerns that there is bias against that age group.

The highest accuracy level was 93.3% for the over 65 year old age bucket for loan applicants. There is a wide range of accuracies amongst subgroups. However, the lowest accuracy level is still is 62.4%, and this is for the group where we believe bias may exist. Overall, the ADS appears to perform reasonably well across all of the examined subpopulations. However, accuracy does not address fairness. It also does not address whether the data was appropriate for the ADS.

Accuracy Score for Select Subpopulations

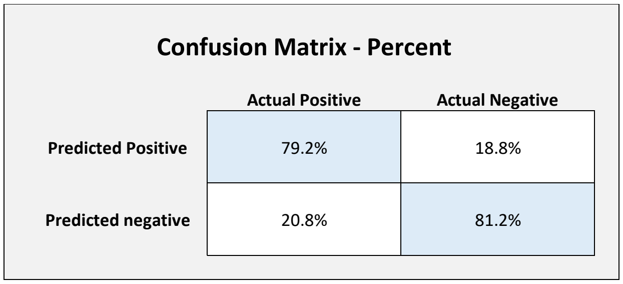
Category	Incorrect	Correct	Total	Accuracy
Overall	**63,612**	**243,899**	**307,511**	**79.31%**
Gender - Female	**35,771**	**166,677**	**202,448**	**82.33%**
Geneder - Male	**27,840**	**77,219**	**105,059**	**73.50%**
Region Rating - 1	**3,607**	**28,590**	**32,197**	**88.80%**
Region Rating - 2	**46,128**	**180,856**	**226,984**	**79.68%**
Region Rating - 3	**13,877**	**34,453**	**48,330**	**71.29%**
Total Income bucket: 0	13,447	50,251	63,698	78.89%
Total Income bucket: 100,000	34,362	121,536	155,898	77.96%
Total Income bucket: 200,000	12,659	52,517	65,176	80.58%
Total Income bucket: 300,000	2,228	12,448	14,676	84.82%
Total Income bucket: 400,000	624	4,737	5,361	88.36%
Total Income bucket: 500,000	142	953	1,095	87.03%
Total Income bucket: 600,000	82	787	869	90.56%
Count of Children - 0	42,243	173,128	215,371	80.39%
Count of Children - 1	14,052	47,067	61,119	77.01%
Count of Children - 2	6,202	20,547	26,749	76.81%
Count of Children - 3	968	2,749	3,717	73.96%
Count of Children - 4	112	317	429	73.89%
Age Bucket: 20 to 25	4,599	7,634	12,233	62.40%
Age Bucket: 25 to 30	10,272	22,681	32,953	68.83%
Age Bucket: 30 to 35	10,647	28,829	39,476	73.03%
Age Bucket: 35 to 40	9,581	33,274	42,855	77.64%
Age Bucket: 40 to 45	7,894	33,512	41,406	80.94%
Age Bucket: 45 to 50	6,570	28,623	35,193	81.33%
Age Bucket: 50 to 55	5,715	29,282	34,997	83.67%
Age Bucket: 55 to 60	4,437	28,660	33,097	86.59%
Age Bucket: 60 to 65	3,372	24,053	27,425	87.70%
Age Bucket: over 65	525	7,351	7,876	93.33%
Flag - Own Car - N	44,194	158,730	202,924	78.22%
Flag - Own Car - Y	19,418	85,169	104,587	81.43%
Education: Academic degree	28	136	164	82.93%
Education: Higher education	10,075	64,788	74,863	86.54%
Education: Incomplete higher	2,332	7,945	10,277	77.31%
Education: Lower secondary	987	2,829	3,816	74.14%
Education: Secondary / secondary special	50,190	168,201	218,391	77.02%

We also examined the model from the perspective of false positive and false negative rates. The confusion matrices below show the actual number counts and percentages. Overall, the model performed reasonably well at balancing the false positive and false negative rates. The false positive rate is 20.8% and the false negative rate is 18.8%. The false positive and false negative rates are both higher than we would like at approximately 20%



Confusion Matrix - counts

	Actual Positive	Actual Negative
Predicted Positive	223,750	4,676
Predicted negative	58,936	20,149



Confusion Matrix - Percent

	Actual Positive	Actual Negative
Predicted Positive	79.2%	18.8%
Predicted negative	20.8%	81.2%

4b. Fairness

In this section we examine of the ADS by looking at the Disparate Impact ratio for various subpopulations. We looked at Disparate Impact ratios for multiple features and have identified four features of interest. The table below shows the Disparate Impact ratios for select subpopulations involving the features of interest.

Select Disparate Impact Ratios

Unpriviliged Group	Privileged Group	Disparate Impact
Female	Males	1.12
Male	Female	0.89
Region Rating 3	Region Rating 2	0.90
Region Rating 3	Region Rating 1	0.80
Income: 0 to 100,000	Income: 600,000 to 700,000	0.87
Age: 20 to 25	Age: 55 to 60	0.72
Age: 20 to 25	Age: >65	0.67
Male in Region 3	Female in Region 1	0.70
Male, Region 3, Age: 20 to 25	Female, Region 1, Age: >65	0.46

Gender is of particular interest due to concerns about gender discrimination. However, the data indicates that males are the underprivileged group in this dataset with a Disparate Impact ratio of 0.89. This ratio is above 0.80 and consequently we do not have large concerns about fairness, particularly since men are the group that is experiencing less favorable outcomes and they would generally be considered the privileged group.

As mentioned earlier, the Region Rating is a subjective rating given by Home Credit Group to the area where the loan applicant lives. The ratings are 1, 2, or 3. The disparate impact ratio between Region Rating 3 and Region Rating 1, where Region Rating 1 is the underprivileged group, is 0.80. This is at a level where we have an increased level of concern regarding potential bias.

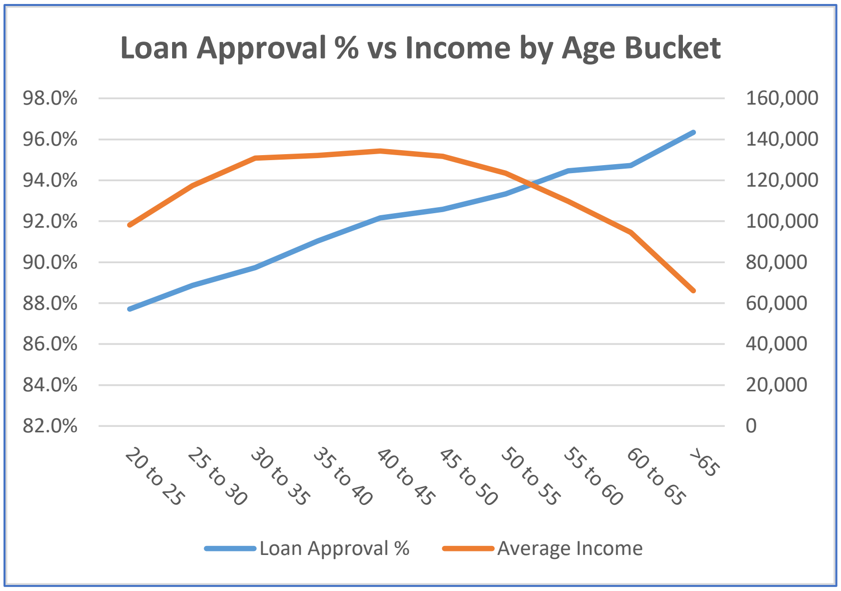
We expect total income to be a relevant factor in the decision of whether to grant a loan. The data supports this expectation. The disparate impact ratio for the lowest income group (0 to $100,000) versus a higher income level group where there was still a significant sample size ($600,000 to $700,000) is 0.87 with the 0 to $100,000 bucket being the underprivileged group. We note that the disparate impact between the low income and high income group is smaller than the disparate impact observed with regard to Region Rating.

We also examined the feature ‘age’ due to concern that age discrimination could be occurring. Above we show the disparate impact ratio for the youngest group (20 to 25 years old) versus the age 55 to 60 group and the over 65 group. We observed significant disparate impact ratios of 0.72 and 0.67, respectively. In general, each age group in the dataset appears to experience a degree of disparate impact versus groups that are older. This raises concern regarding whether age bias is occurring.

To further explore this concern, we looked at the average income for each age group. This is shown in the table below. As can be seen, the peak average income is the 40 to 45 year old bucket. The 20 to 25 year old bucket has a lower average income but it still exceeds the average income of the "over 65" bucket. The pattern for average income per age group is different than the pattern for average loan approval percentage. This can be seen in the plot below and increases our concern that age bias may be present.

Average Income per Age Bucket

Age Bucket			Average
20	to	25	98,185
25	to	30	117,325
30	to	35	130,844
35	to	40	132,209
40	to	45	134,374
45	to	50	131,779
50	to	55	123,453
55	to	60	109,672
60	to	65	94,582
		>65	66,125



Finally, we looked at the disparate impact for the intersection of certain subgroups. Our analysis has already identified differences in outcomes for subgroups related to gender, Region Rating, and Age. Given this, we looked at intersection of these features to see if potential bias was more apparent when an applicant belonged to multiple unprivileged groups. For the three features identified, the most extreme example would be to look at the disparate impact ratio for a male, between the age of 20 and 25, who lived in Region 3 (the underprivileged group) versus a female, over 65 years of age, who lived in Region 1 (the privileged group). The disparate impact ratio for these two subpopulations was 0.46, indicating a significant difference in outcomes for the underprivileged group.

4c. Additional methods for analyzing ADS performance

i. Underlying Data

Our analysis to this point indicates that the ADS is reasonably accurate but there are signs that bias may exist in the outcomes. This requires us to examine the underlying data more closely. For example, if the model had an accuracy of 100% but our fairness measures indicated bias, it is likely that our model learned to reproduce a bias that existed in the data from a pre-existing. To better understand this, we will now look more closely at the underlying data. This will be followed by a review of feature importance.

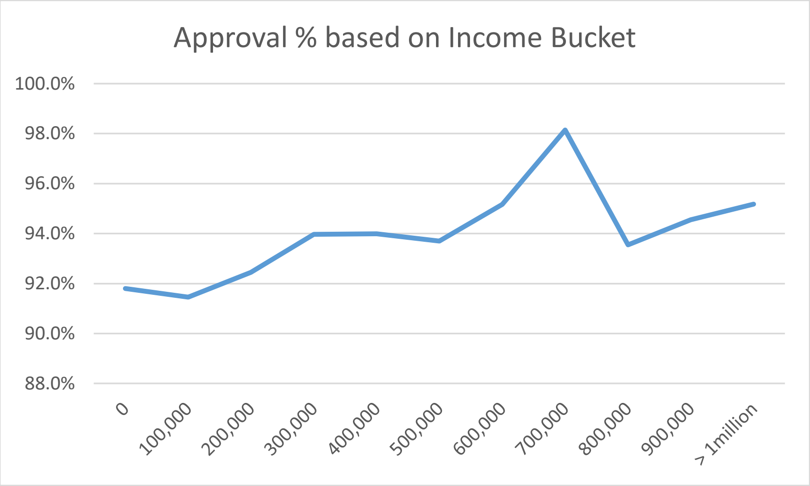
As explained above, the data set is large with numerous input features. In order to better direct our efforts, we initially reviewed the underlying data to look for features of concern. We identified 16 features of concern. In particular our concern has grown over the features age. There is a clear disparity of outcomes for different age groups. We also have concerns over Region Rating, which may be a proxy for race. Finally, we have also looked more closely at gender where the surprising result is that males are experiencing worse outcomes than females, although the differences are not large enough to be dispositive.

To be thorough, we will now briefly report on each of the 16 variables we initially identified. It is important to review each of the features to have confidence that the features that we are concerned about are truly the primary features that cause the differences in outcome.

AMT_INCOME_TOTAL

This feature informs us as to the income of the applicant. Our initial expectation was that the loan approval rate for the applicants would increase as income increased and this is what the data revealed. The correlation between income and approval percentage is 0.68. The general trend is that loan approval percentage increases as income increases.

Income							
			Counts			Percentages	
Bucket			Approved	Denied	Total	Approved	Denied
0	to	100,000	58,473	5,225	63,698	91.8%	8.2%
100,000	to	200,000	142,572	13,326	155,898	91.5%	8.5%
200,000	to	300,000	60,255	4,921	65,176	92.4%	7.6%
300,000	to	400,000	13,791	885	14,676	94.0%	6.0%
400,000	to	500,000	5,039	322	5,361	94.0%	6.0%
500,000	to	600,000	1,026	69	1,095	93.7%	6.3%
600,000	to	700,000	827	42	869	95.2%	4.8%
700,000	to	800,000	159	3	162	98.1%	1.9%
800,000	to	900,000	116	8	124	93.5%	6.5%
900,000	to	1,000,000	191	11	202	94.6%	5.4%
	>	1,000,000	237	12	249	95.2%	4.8%



CNT_CHILDREN and CNT_FAM_MEMBERS

CNT_CHILDREN is the number of children that the applicant has. CNT_FAM_MEMBERS is the number of family members that the applicant has. These two features are highly similar. The correlation between CNT_CHILDREN and CNT_FAM_MEMBERS is 0.88. Consequently, we will focus our reporting on CNT_CHILDREN. Additionally, 99.9% of the applicants have four children or less. The data indicates that the loan approval rating declines as the number of children increases from zero to four. The percentage approval declines from 92.3% for applicants with no children to 87.2% for applicants with four children. The differences in approval percentages is relatively small for families with 0 to 3 children. The approval percentage drops for families with 4 children but the sample size is significantly smaller. There is a possibility that this feature cold be a proxy for income or race, given that lower income families and minorities tend to have more children. However, the Disparate Impact ration for families with four children (the underprivileged group) versus families with no children is 0.95. This is not a strong indication of bias. Additionally, we note that the underlying data has certain anomalies which have reduced our confidence in this feature. For example, one applicant has 19 children.

CNT_CHILDREN						
	Counts			Percentages		
Children	Approved	Denied	Total	Approved	Denied	Total
0	198,762	16,609	215,371	92.3%	7.7%	100.0%
1	55,665	5,454	61,119	91.1%	8.9%	100.0%
2	24,416	2,333	26,749	91.3%	8.7%	100.0%
3	3,359	358	3,717	90.4%	9.6%	100.0%
4	374	55	429	87.2%	12.8%	100.0%
5	77	7	84	91.7%	8.3%	100.0%
6	15	6	21	71.4%	28.6%	100.0%
7	7	0	7	100.0%	0.0%	100.0%
>7	11	3	14	78.6%	21.4%	100.0%
Total	282,686	24,825	307,511			

CODE_GENDER

This is the gender of the applicant. As discussed above, this is one of the features that we are most interested in. There are three values for gender in the dataset: F = Female, M = Male, and XNA = Other. There is a clear difference between the loan approval percentage for women (93.0%) and men

(89.9%). This result was surprising due to the fact that it is men that are experiencing worse outcomes than women. Given the societal history of men experiencing better outcomes than women, it is difficult to conclude that a loan approval difference of approximately 3% (93.0% for women and 89.9% for men) is a major concern. This is particularly true given how high the overall loan approval rate is for both groups.

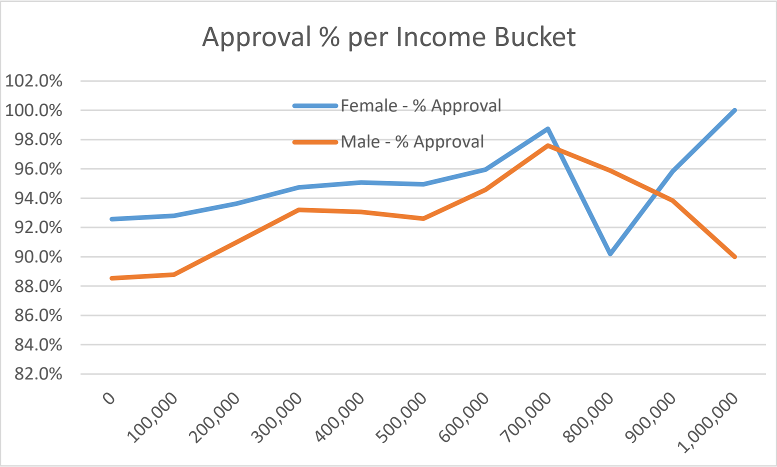
For completeness, we also reviewed the average total income for women versus men. From the table below we can see that the average income for females is approximately 20% lower than that of males. We can also see that there is no appreciable difference in the average income for women that are approved for a loan versus woman that are not. This is not true for the men, where the average income for men who are approved for a loan is significantly higher than the average income for men who are not approved for a loan. In fact, the average income for males being denied a loan is significantly higher than the average income for females who are approved for a loan. The average income data does not provide support for women being approved for loans at a higher percentages than men.

CODE_GENDER						
	Counts			Percentages		
Gender	Approved	Denied	Total	Approved	Denied	Total
F	188,278	14,170	202,448	93.0%	7.0%	100.0%
M	94,404	10,655	105,059	89.9%	10.1%	100.0%
XNA	4	0	4	100.0%	0.0%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

AMT_INCOME_TOTAL			
	Counts		
Gender	Approved	Denied	Total
F	155,984	156,671	156,032
M	195,190	177,502	193,396
XNA	186,750	n/a	186,750
Total	282,686	24,825	307,511

To further, explore the disparity in approval percentages between females and males, we examined the loan approval percentages for each gender based on income bucket.

	Female - Income					Male - Income				
	Counts			Percentages		Counts			Percentages	
Income Bucket	Approved	Denied	Total	Approved	Denied	Approved	Denied	Total	Approved	Denied
0 to 100,000	47,661	3,824	51,485	92.6%	7.4%	10,812	1,401	12,213	88.5%	11.5%
100,000 to 200,000	96,255	7,471	103,726	92.8%	7.2%	46,315	5,855	52,170	88.8%	11.2%
200,000 to 300,000	33,909	2,309	36,218	93.6%	6.4%	26,344	2,612	28,956	91.0%	9.0%
300,000 to 400,000	6,950	386	7,336	94.7%	5.3%	6,841	499	7,340	93.2%	6.8%
400,000 to 500,000	2,371	123	2,494	95.1%	4.9%	2,668	199	2,867	93.1%	6.9%
500,000 to 600,000	488	26	514	94.9%	5.1%	538	43	581	92.6%	7.4%
600,000 to 700,000	356	15	371	96.0%	4.0%	471	27	498	94.6%	5.4%
700,000 to 800,000	78	1	79	98.7%	1.3%	81	2	83	97.6%	2.4%
800,000 to 900,000	46	5	51	90.2%	9.8%	70	3	73	95.9%	4.1%
900,000 to 1,000,000	69	3	72	95.8%	4.2%	122	8	130	93.8%	6.2%
1,000,000 to 1,100,000	8	0	8	100.0%	0.0%	9	1	10	90.0%	10.0%
1,100,000 to 1,200,000	33	1	34	97.1%	2.9%	39	1	40	97.5%	2.5%
1,200,000 to 1,300,000	6	1	7	85.7%	14.3%	7	0	7	100.0%	0.0%
1,300,000 to 1,400,000	17	2	19	89.5%	10.5%	36	2	38	94.7%	5.3%
1,400,000 to 1,500,000	2	0	2	100.0%	0.0%	1	0	1	100.0%	0.0%
1,500,000 to 1,600,000	7	1	8	87.5%	12.5%	11	0	11	100.0%	0.0%
> 1,000,000	22	0	24	100.0%	0.0%	39	2	41	95.1%	4.9%



The loan approval percentage approval for females is consistently higher that of men up to and including the "700,000" bucket. This bucket represents incomes between $700,000 and $800,000. Female applicants with incomes less than $800,000 represent 99.9% of all female applicants. Male applicants with incomes less than $800,000 represent 99.7% of all male applicants. It appears that the lower loan approval percentages for males is not explained by income differences.

Finally, we looked at the average credit score for women and men for each age bucket. This is shown in the table below. There we noted that the average credit score for women is higher in every bucket. The correlation between credit score and the model output is -0.60, the largest absolute correlation that we noted. Additionally, features related to credit score are four of the six most important features. Credit score is a critically important feature and could explain the disparity in approval percentages between women and men. Ultimately, we do not conclude that this factor needs further attention.

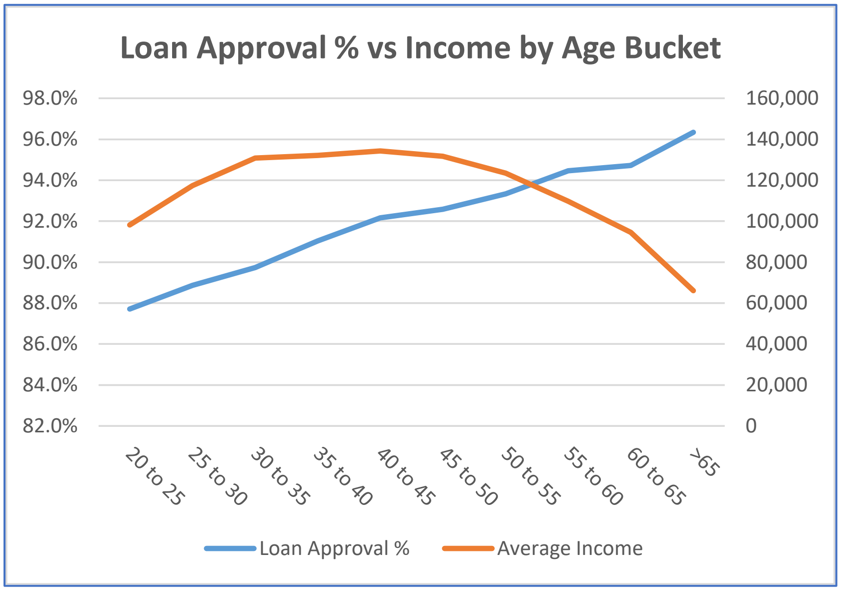
	Income		Credit Score	
Age	**Female**	**Male**	**Female**	**Male**
20 to 25	131,463	162,314	0.385	0.364
25 to 30	148,596	186,742	0.442	0.424
30 to 35	161,990	201,645	0.479	0.462
35 to 40	164,234	205,108	0.510	0.491
40 to 45	166,593	207,981	0.534	0.514
45 to 50	166,172	204,622	0.546	0.528
50 to 55	161,005	195,144	0.553	0.536
55 to 60	150,951	181,509	0.547	0.542
60 to 65	141,235	156,420	0.551	0.537
>65	117,140	124,349	0.564	0.551

DAYS_BIRTH

This feature tells us the how old the loan application was at the time of the application. The data is provided in days and shown as a negative number. For the report, we have converted the data from days into years and removed the negative sign. This variable is of interest for potential age bias. The data shows a steady increase in loan approval percentage as the age bucket increases.

DAYS_BIRTH						
	Counts			Percentages		
Age Bucket (years)	Approved	Denied	Total	Approved	Denied	Total
20	10,729	1,504	12,233	87.7%	12.3%	100.0%
25	29,286	3,667	32,953	88.9%	11.1%	100.0%
30	35,422	4,054	39,476	89.7%	10.3%	100.0%
35	39,012	3,843	42,855	91.0%	9.0%	100.0%
40	38,164	3,242	41,406	92.2%	7.8%	100.0%
45	32,582	2,611	35,193	92.6%	7.4%	100.0%
50	32,662	2,335	34,997	93.3%	6.7%	100.0%
55	31,264	1,833	33,097	94.5%	5.5%	100.0%
60	25,977	1,448	27,425	94.7%	5.3%	100.0%
65	7,588	288	7,876	96.3%	3.7%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

As discussed above, the pattern for average income per age group is different than the pattern for loan approval percentage. We repeat the plot from above to illustrate this point.



For completeness, we looked at the age distribution of loan applicants. This is shown in the table below. The peak age group for loan applicants is 35 to 40 years of age. The 20 to 25 year group and the over 65 year group have significantly fewer applications.

The disparate impact ratios indicate that the differences in outcomes are of sufficient magnitude as to be of concern. The Disparate Impact ratio for the 20 to 25 year old group (the underprivileged group) versus the over 65 year old group is 0.67. This increases our concern about the inclusion of age as a feature.

Finally, given the importance of credit score, we looked at the correlation between age and average credit score. The correlation is positive 0.28. This indicates that as age goes up, there is a modest correlation with credit score. This may provide a partial explanation but we would note that even if credit score did explain the anomaly, an argument could still be made to remove age since credit score would be dominant and provide the explanation and age would not be needed.

Unpriviliged Group	Privileged Group	Disparate Impact
Age: 20 to 25	Age: 55 to 60	0.72
Age: 20 to 25	Age: >65	0.67

FLAG_OWN_CAR and FLAG_OWN_REALTY

"FLAG_OWN_CAR" tells us if the loan applicant owns a car. "FLAG_OWN_REALTY" tells us if the loan applicant owns a home. These features have the potential to be proxies for race. Loan applicants who owned a car were approved for loans at a higher percentage (92.8%) than applicants who did not own a car (91.5%). Loan applicants who owned a home were approved for loans at a higher percentage (92.0%) than applicants who did not own a home (91.7%). Our concern is that these two features may be a proxy for race. However, the results are relatively close for both features are not indicative of strong bias, if any.

FLAG_OWN_CAR						
	Counts			Percentages		
Flag	Approved	Denied	Total	Approved	Denied	Total
N	185,675	17,249	202,924	91.5%	8.5%	100.0%
Y	97,011	7,576	104,587	92.8%	7.2%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

FLAG_OWN_REALTY						
	Counts			Percentages		
FLAG_OWN_REALTY	Approved	Denied	Total	Approved	Denied	Total
N	86,357	7,842	94,199	91.7%	8.3%	100.0%
Y	196,329	16,983	213,312	92.0%	8.0%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

NAME_EDUCATION_TYPE

This feature tells us the level of education of the loan applicant. We are curious if there is a bias against individuals based on their education level. Unfortunately, the information provided by Home Credit Group is sparse with regard to this feature. For example, it is likely that "secondary" school is the equivalent of a high school education. This is supported by the fact that it is also the category with the largest number of loan applicants. Home Credit Group did not provide additional information regarding the meaning of each category label. It is our assumption that "Higher Education" signifies that the application has a college or graduate degree. This would be the highest level of education for applicants. This group has a higher loan approval percentage than the other categories with significant population sizes. It is possible that education level is correlated with income and that could explain the different approval percentages we are seeing for different education levels.

NAME_EDUCATION_TYPE						
	Counts			Percentages		
Education Type	Approved	Denied	Total	Approved	Denied	Total
Secondary / secondary special	198,867	19,524	218,391	91.1%	8.9%	100.0%
Higher education	70,854	4,009	74,863	94.6%	5.4%	100.0%
Incomplete higher	9,405	872	10,277	91.5%	8.5%	100.0%
Lower secondary	3,399	417	3,816	89.1%	10.9%	100.0%
Academic degree	161	3	164	98.2%	1.8%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

NAME_FAMILY_STATUS

This feature tells us the “family status” of the client. This feature has categories such as “Married” and “Single / not married”. We are curious if there is a bias against individuals based on their family status. Additionally, this feature could be a proxy for race. The data does not show any clear indications of bias in the loan approval process. Married couples made up the largest group of applicants and had the second highest loan approval percentages. The differences in loan approval percentages are not sufficient to raise a high level of concern.

NAME_FAMILY_STATUS						
	Counts			Percentages		
Family Status	Approved	Denied	Total	Approved	Denied	Total
Married	181,582	14,850	196,432	92.4%	7.6%	100.0%
Single / not married	40,987	4,457	45,444	90.2%	9.8%	100.0%
Civil marriage	26,814	2,961	29,775	90.1%	9.9%	100.0%
Separated	18,150	1,620	19,770	91.8%	8.2%	100.0%
Widow	15,151	937	16,088	94.2%	5.8%	100.0%
Unknown	2		2	100.0%	0.0%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

NAME_HOUSING_TYPE

This feature tells us the “housing situation of the client”. This feature has categories such as “House / apartment” and “Municipal apartment”. We are curious if there is a bias against individuals based on their housing status. Additionally, this feature could be a proxy for race. The data does not show any clear indications of bias in the loan approval process.

NAME_HOUSING_TYPE						
	Counts			Percentages		
Category	Approved	Denied	Total	Approved	Denied	Total
House / apartment	251,596	21,272	272,868	92.2%	7.8%	100.0%
With parents	13,104	1,736	14,840	88.3%	11.7%	100.0%
Municipal apartment	10,228	955	11,183	91.5%	8.5%	100.0%
Rented apartment	4,280	601	4,881	87.7%	12.3%	100.0%
Office apartment	2,445	172	2,617	93.4%	6.6%	100.0%
Co-op apartment	1,033	89	1,122	92.1%	7.9%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

NAME_INCOME_TYPE

This feature tells us the source of the applicant’s income. We are curious if there is a bias against individuals based on the source of their income. Additionally, this feature could be a proxy for race. Unfortunately, the information provided by Home Credit Group is sparse with regard to this feature and the categories are broad. For example, almost 52% of the applicants fall into the category of “Working”. Not surprisingly, this category has the highest approval percentage of the categories with significant population sizes. Additionally, it is unclear what some of the other category labels represent. For example, it is unclear was a “State servant” is. Consequently, we have little guidance regarding the potential for this category to be a source of bias.

NAME_INCOME_TYPE						
	Counts			Percentages		
Category	**Approved**	**Denied**	**Total**	**Approved**	**Denied**	**Total**
Working	143,550	15,224	158,774	90.4%	9.6%	100.0%
Commercial associate	66,257	5,360	71,617	92.5%	7.5%	100.0%
Pensioner	52,380	2,982	55,362	94.6%	5.4%	100.0%
State servant	20,454	1,249	21,703	94.2%	5.8%	100.0%
Unemployed	14	8	22	63.6%	36.4%	100.0%
Student	18	0	18	100.0%	0.0%	100.0%
Businessman	10	0	10	100.0%	0.0%	100.0%
Maternity leave	3	2	5	60.0%	40.0%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

OCCUPATION_TYPE

This feature tells us the occupation of the applicant. There are 18 different categories. A review of the categories shows the potential for overlap between the categories. For example, “Managers” could also be “Medicine staff”. It is difficult to draw any conclusions from a review of this data.

OCCUPATION_TYPE						
	Counts			Percentages		
Category	**Approved**	**Denied**	**Total**	**Approved**	**Denied**	**Total**
Laborers	49,348.00	5,838.00	55,186.00	89.4%	10.6%	100.0%
Sales staff	29,010.00	3,092.00	32,102.00	90.4%	9.6%	100.0%
Core staff	25,832.00	1,738.00	27,570.00	93.7%	6.3%	100.0%
Managers	20,043.00	1,328.00	21,371.00	93.8%	6.2%	100.0%
Drivers	16,496.00	2,107.00	18,603.00	88.7%	11.3%	100.0%
High skill tech staff	10,679.00	701.00	11,380.00	93.8%	6.2%	100.0%
Accountants	9,339.00	474.00	9,813.00	95.2%	4.8%	100.0%
Medicine staff	7,965.00	572.00	8,537.00	93.3%	6.7%	100.0%
Security staff	5,999.00	722.00	6,721.00	89.3%	10.7%	100.0%
Cooking staff	5,325.00	621.00	5,946.00	89.6%	10.4%	100.0%
Cleaning staff	4,206.00	447.00	4,653.00	90.4%	9.6%	100.0%
Private service staff	2,477.00	175.00	2,652.00	93.4%	6.6%	100.0%
Low-skill Laborers	1,734.00	359.00	2,093.00	82.8%	17.2%	100.0%
Waiters/barmen staff	1,196.00	152.00	1,348.00	88.7%	11.3%	100.0%
Secretaries	1,213.00	92.00	1,305.00	93.0%	7.0%	100.0%
Realty agents	692.00	59.00	751.00	92.1%	7.9%	100.0%
HR staff	527.00	36.00	563.00	93.6%	6.4%	100.0%
IT staff	492.00	34.00	526.00	93.5%	6.5%	100.0%
Total	192,573.00	18,547.00	211,120.00	91.2%	8.8%	100.0%

ORGANIZATION_TYPE

This feature tells us the type of business where the applicant works. We had originally thought that this feature might give us information on the relationship between loan approvals and business type. However, after reviewing the data we no longer believe this to be the case. There are 58 different categories of businesses. Just over 18% of the loan applicants have an ORGANIZATION_TYPE of “XNA” which we believe means “not applicable”.

REGION_RATING_CLIENT and REGION_RATING_CLIENT_W_CITY

REGION_RATING_CLIENT tells us Home Credit Group's rating of "the region" where the applicant lives. REGION_RATING_CLIENT_W_CITY is Home Credit Groups rating of the region where the applicant lives "taking city into account". These feature appears to be subjective in nature. Each feature has possible ratings of 1, 2, or 3. In our initial review we identified these features as features of interest. After looking at the data more closely, they continue to be features of interest. The data from the features are highly similar with a correlation of over 0.95. For brevity, we will on discuss REGION_RATING_CLIENT but the conclusions are the same for both features.

There is a clear difference in loan approval percentages based on the rating of the region with the lowest rated region having the lowest loan approval percentage. There is a concern that these features could be used as a proxy for race.

REGION_RATING_CLIENT						
	Counts			Percentages		
Rating	Approved	Denied	Total	Approved	Denied	Total
1	30,645	1,552	32,197	95.2%	4.8%	100.0%
2	209,077	17,907	226,984	92.1%	7.9%	100.0%
3	42,964	5,366	48,330	88.9%	11.1%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

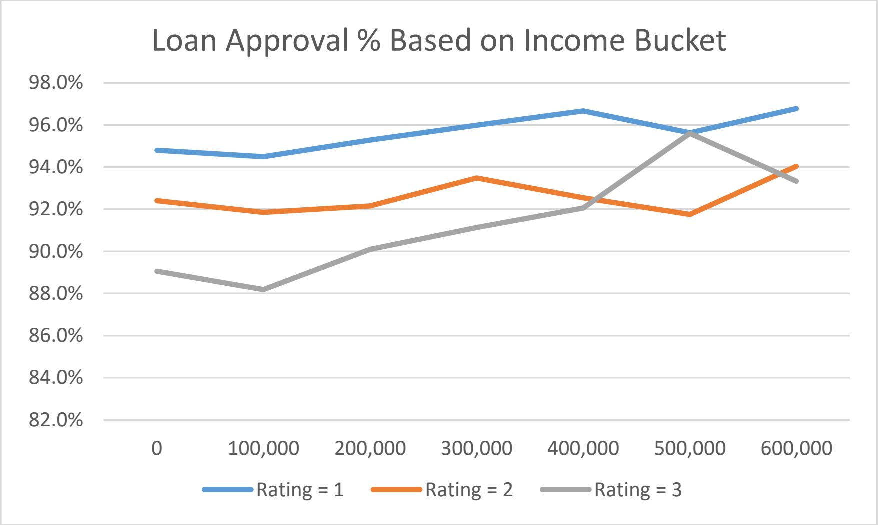
REGION_RATING_CLIENT_W_CITY						
	Counts			Percentages		
Rating	Approved	Denied	Total	Approved	Denied	Total
1	32,513	1,654	34,167	95.2%	4.8%	100.0%
2	211,314	18,170	229,484	92.1%	7.9%	100.0%
3	38,859	5,001	43,860	88.6%	11.4%	100.0%
Total	282,686	24,825	307,511	91.9%	8.1%	100.0%

To explore these features further, we broke down each of the ratings by income bucket.

Region Rating			
	Average Income		
Rating	Approved	Denied	Total
1	242,521	228,137	241,825
2	161,349	165,860	161,707
3	149,655	144,031	149,014

Region Rating Per Income Bucket									
	Rating = 1			Rating = 2			Rating =3		
Bucket	Approved	Denied	Total	Approved	Denied	Total	Approved	Denied	Total
0 to 100,000	1,493	82	1,575	45,804	3,769	49,573	11,176	1,374	12,550
100,000 to 200,000	11,003	642	11,645	109,291	9,696	118,987	22,278	2,988	25,266
200,000 to 300,000	10,999	544	11,543	41,872	3,565	45,437	7,384	812	8,196
300,000 to 400,000	4,157	174	4,331	8,196	571	8,767	1,438	140	1,578
400,000 to 500,000	1,887	65	1,952	2,665	215	2,880	487	42	529
500,000 to 600,000	438	20	458	501	45	546	87	4	91
600,000 to 700,000	361	12	373	410	26	436	56	4	60
700,000 and 800,000	84	1	85	68	2	70	7	0	7
800,000 to 900,000	60	4	64	46	4	50	10	0	10
900,000 to 1,000,000	78	3	81	95	7	102	18	1	19
> 1 million	85	5	90	129	7	136	23	1	24

REGION_RATING_CLIENT - loan approval percentage					
				Rating	
	Income Bucket		1	2	3
0	to	100,000	94.8%	92.40%	89.05%
100,000	to	200,000	94.5%	91.85%	88.17%
200,000	to	300,000	95.3%	92.15%	90.09%
300,000	to	400,000	96.0%	93.49%	91.13%
400,000	to	500,000	96.7%	92.53%	92.06%
500,000	to	600,000	95.6%	91.76%	95.60%
600,000	to	700,000	96.8%	94.04%	93.33%
700,000	to	800,000	98.8%	97.14%	100.00%
800,000	to	900,000	93.8%	92.00%	100.00%
900,000	to	1,000,000	96.3%	93.14%	94.74%
1,000,000	to	1,100,000	100.0%	85.71%	100.00%
1,100,000	to	1,200,000	96.7%	97.14%	100.00%
1,200,000	to	1,300,000	100.0%	87.50%	100.00%
1,300,000	to	1,400,000	93.8%	94.44%	80.00%
1,400,000	to	1,500,000	100.0%	100.00%	n/a
1,500,000	to	1,600,000	80.0%	100.00%	100.00%
> 1,500,000		> 1,500,000	90.9%	94.6%	100.0%



The graph above shows the loan approval percentages based on income bucket for each region rating. The graph only shows the first seven income buckets. The highest bucket is $600,000. This represents applicants with an income between $600,000 and $700,000. We chose to cutoff the graph at $600,000 because we had already accounted for over 99% of loan applicants in each ratings group, and the data became sparse after that. The graph above lends weak support to the proposition that Region Rating is correlated to income.

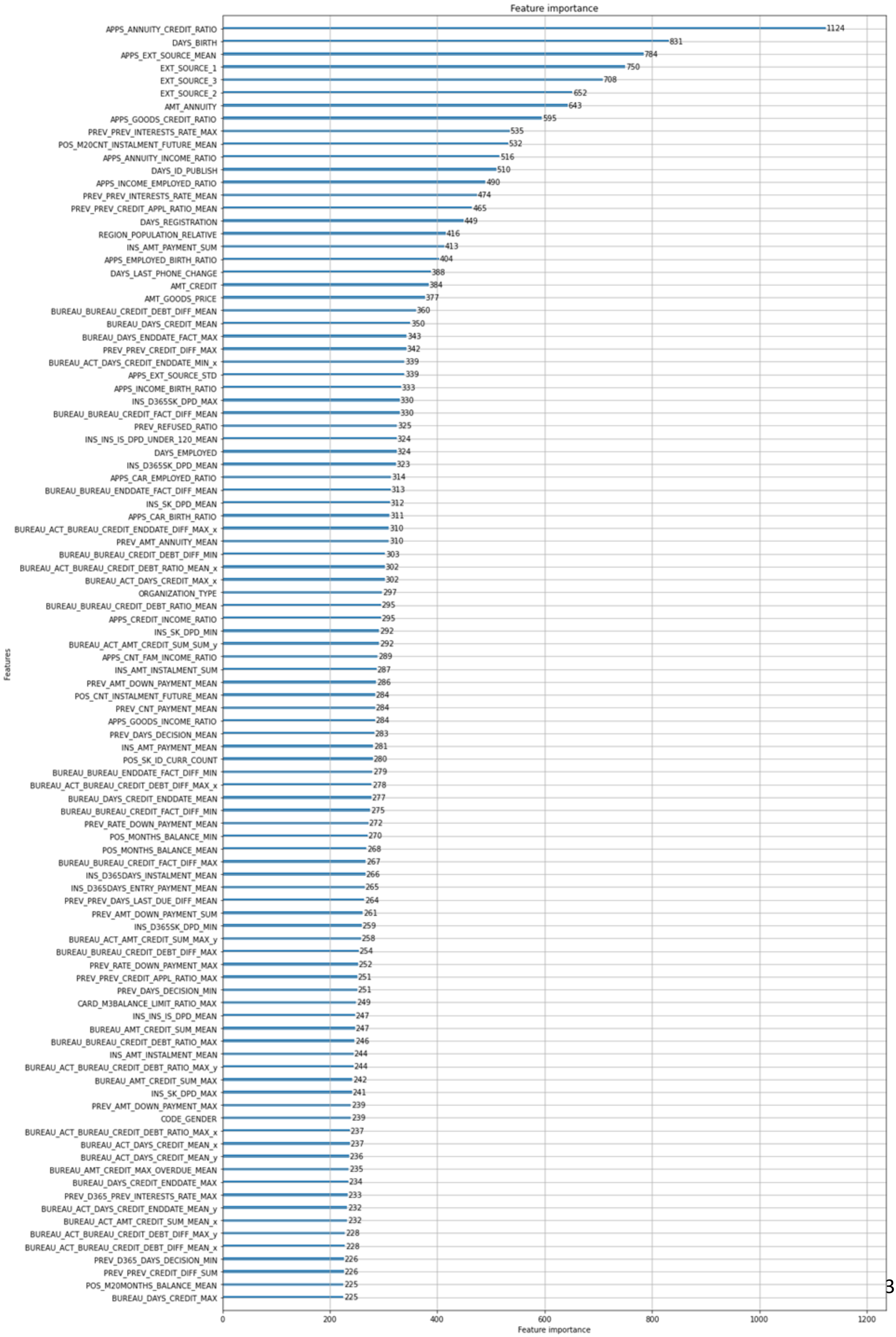
The disparate impact ration for Region 3 (the unprivileged group) versus Region 1 (the privileged group) is 0.80. This a level that raises our concern.

Disparate Impact Ratios		
Unpriviliged Group	Privileged Group	Disparate Impact
Region Rating 3	Region Rating 2	0.90
Region Rating 3	Region Rating 1	0.80

Given the potential that this feature could be a proxy for zip code or race, and that there is a significant disparate impact, we would recommend removing this feature. We would also recommend removing all other features related to region.

ii. Feature Importance

We implemented the ADS in a Google Colab notebook and created a plot of feature importance. That plot is below. A portion of that plot is below. There are approximately 200 features and we could not show all of them on the plot.


Feature importance
APPS_ANNUITY_CREDIT_RATIO 1124
DAYS_BIRTH 831
APPS_EXT_SOURCE_MEAN 784
EXT_SOURCE_1 750
EXT_SOURCE_3 708
EXT_SOURCE_2 652
AMT_ANNUITY 643
APPS_GOODS_CREDIT_RATIO 595
PREV_PREV_INTERESTS_RATE_MAX 535
POS_M20CNT_INSTALMENT_FUTURE_MEAN 532
APPS_ANNUITY_INCOME_RATIO 516
DAYS_ID_PUBLISH 510
APPS_INCOME_EMPLOYED_RATIO 490
PREV_PREV_INTERESTS_RATE_MEAN 474
PREV_PREV_CREDIT_APPL_RATIO_MEAN 465
DAYS_REGISTRATION 449
REGION_POPULATION_RELATIVE 416
INS_AMT_PAYMENT_SUM 413
APPS_EMPLOYED_BIRTH_RATIO 404
DAYS_LAST_PHONE_CHANGE 388
AMT_CREDIT 384
AMT_GOODS_PRICE 377
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MEAN 360
BUREAU_DAYS_CREDIT_MEAN 350
BUREAU_DAYS_ENDDATE_FACT_MAX 343
PREV_PREV_CREDIT_DIFF_MAX 342
BUREAU_ACT_DAYS_CREDIT_ENDDATE_MIN_x 339
APPS_EXT_SOURCE_STD 339
APPS_INCOME_BIRTH_RATIO 333
INS_D365SK_DPD_MAX 330
BUREAU_BUREAU_CREDIT_FACT_DIFF_MEAN 330
PREV_REFUSED_RATIO 325
INS_INS_IS_DPD_UNDER_120_MEAN 324
DAYS_EMPLOYED 324
INS_D365SK_DPD_MEAN 323
APPS_CAR_EMPLOYED_RATIO 314
BUREAU_BUREAU_ENDDATE_FACT_DIFF_MEAN 313
INS_SK_DPD_MEAN 312
APPS_CAR_BIRTH_RATIO 311
BUREAU_ACT_BUREAU_CREDIT_ENDDATE_DIFF_MAX_x 310
PREV_AMT_ANNUITY_MEAN 310
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MIN 303
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MEAN_x 302
BUREAU_ACT_DAYS_CREDIT_MAX_x 302
ORGANIZATION_TYPE 297
BUREAU_BUREAU_CREDIT_DEBT_RATIO_MEAN 295
APPS_CREDIT_INCOME_RATIO 295
INS_SK_DPD_MIN 292
BUREAU_ACT_AMT_CREDIT_SUM_SUM_y 292
APPS_CNT_FAM_INCOME_RATIO 289
INS_AMT_INSTALMENT_SUM 287
PREV_AMT_DOWN_PAYMENT_MEAN 286
POS_CNT_INSTALMENT_FUTURE_MEAN 284
PREV_CNT_PAYMENT_MEAN 284
APPS_GOODS_INCOME_RATIO 284
PREV_DAYS_DECISION_MEAN 283
INS_AMT_PAYMENT_MEAN 281
POS_SK_ID_CURR_COUNT 280
BUREAU_BUREAU_ENDDATE_FACT_DIFF_MIN 279
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MAX_x 278
BUREAU_DAYS_CREDIT_ENDDATE_MEAN 277
BUREAU_BUREAU_CREDIT_FACT_DIFF_MIN 275
PREV_RATE_DOWN_PAYMENT_MEAN 272
POS_MONTHS_BALANCE_MIN 270
POS_MONTHS_BALANCE_MEAN 268
BUREAU_BUREAU_CREDIT_FACT_DIFF_MAX 267
INS_D365DAYS_INSTALMENT_MEAN 266
INS_D365DAYS_ENTRY_PAYMENT_MEAN 265
PREV_PREV_DAYS_LAST_DUE_DIFF_MEAN 264
PREV_AMT_DOWN_PAYMENT_SUM 261
INS_D365SK_DPD_MIN 259
BUREAU_ACT_AMT_CREDIT_SUM_MAX_y 258
BUREAU_BUREAU_CREDIT_DEBT_DIFF_MAX 254
PREV_RATE_DOWN_PAYMENT_MAX 252
PREV_PREV_CREDIT_APPL_RATIO_MAX 251
PREV_DAYS_DECISION_MIN 251
CARD_M3BALANCE_LIMIT_RATIO_MAX 249
INS_INS_IS_DPD_MEAN 247
BUREAU_AMT_CREDIT_SUM_MEAN 247
BUREAU_BUREAU_CREDIT_DEBT_RATIO_MAX 246
INS_AMT_INSTALMENT_MEAN 244
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MAX_y 244
BUREAU_AMT_CREDIT_SUM_MAX 242
INS_SK_DPD_MAX 241
PREV_AMT_DOWN_PAYMENT_MAX 239
CODE_GENDER 239
BUREAU_ACT_BUREAU_CREDIT_DEBT_RATIO_MAX_x 237
BUREAU_ACT_DAYS_CREDIT_MEAN_x 237
BUREAU_ACT_DAYS_CREDIT_MEAN_y 236
BUREAU_AMT_CREDIT_MAX_OVERDUE_MEAN 235
BUREAU_DAYS_CREDIT_ENDDATE_MAX 234
PREV_D365_PREV_INTERESTS_RATE_MAX 233
BUREAU_ACT_DAYS_CREDIT_ENDDATE_MEAN_y 232
BUREAU_ACT_AMT_CREDIT_SUM_MEAN_x 232
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MAX_y 228
BUREAU_ACT_BUREAU_CREDIT_DEBT_DIFF_MEAN_x 228
PREV_D365_DAYS_DECISION_MIN 226
PREV_PREV_CREDIT_DIFF_SUM 226
POS_M20MONTHS_BALANCE_MEAN 225
BUREAU_DAYS_CREDIT_MAX 225
Features
0
200
400
600
800
1000
1200
Feature importance

The feature importance plot indicates that the most important feature is APPS_ANNUITY_CREDIT_RATIO. This is a feature that was implemented by the designer of the ADS. It is equal to AMT_ANNUITY divided by AMT_CREDIT. This is the loan payment divided by the loan amount. It is a proxy for interest rate. Most importantly, it incorporates the size of the loan payment which is a significant factor in determining the borrower's ability to repay.

The second most import feature is DAY_BIRTH. This is the age of the borrower. This tells us that the borrower's age is a significant factor in the decision of whether to approve the loan. This is concerning, particularly since we have already identified concerns about this feature related to potential bias.

The feature importance plot indicates that APPS_EXT_SOURCE_MEAN is the third most important feature. This is a feature that was implemented by the ADS designer. It represents the average of three normalized credit scores obtained from different credit rating agencies. Home Credit Group did not provide the credit ratings for the loan applicants due to contractual restrictions with the credit rating agencies. Instead Home Credit Group provided a normalized score for credit rating. It is clear that this feature should be important. However, it is concerning that this feature is less important than age.

The fourth, fifth, and sixth most important features are the individual's normalized credit scores from three credit rating agencies. This appears redundant with the APPS_EXT_SOURCE_MEAN feature discussed above. A solution to this would be to remove the individual normalized credit scores and leave the APPS_EXT_SOURCE_MEAN feature. The model is assigning weight to all four features. It is likely that if only APPS_EXT_SOURCE_MEAN were included the importance of that feature would increase.

The seventh most important feature is AMT_ANNUITY. This is the amount of the proposed loan payment. This feature is also incorporated in the APPS_ANNUITY_CREDIT_RATIO feature. The model is assigning weight to both of these features. It is reasonable to include both the AMT_ANNUITY and the APPS_ANNUITY_CREDIT_RATIO features in the model. The ratio is indicative of the interest rate on the loan, whereas the AMT_ANNUITY directly reflects the size of the payment.

We have also identified gender as a feature of interest. We were surprised to see that women experienced better outcomes than men. The feature importance plot assigned a low value to the importance of CODE_GENDER and it is near the bottom of the list.

5. Summary

5a. Do you believe that the data was appropriate for this ADS?

We do not believe that the data was appropriate for this ADS. The vast majority of the data is appropriate for the ADS. However, the ADS showed that bias may exist with regard to age and Region Rating. The use of the age and Region Rating appear problematic. Our concern over the inclusion of age is only reinforced by the feature importance plot which shows that age is the second most important feature, even more important than the borrower's credit rating. The inclusion of these features in the model is problematic.

The goal of the ADS is to predict a borrower's ability to repay a loan. However, the model attempts to do this without having any information about the potential borrower's assets or liabilities.

This is significant weakness in the model. We believe that data related to the applicants assets and liabilities should be included.

5b. Is the implementation robust, accurate and fair?

The question that the ADS was attempting to answer is whether a potential borrower would experience payment difficulties if they were granted a loan. The target variable is defined as:

> *Target : Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample, 0 - all other cases)*

The stated objective of the Kaggle competition was to predict the “clients' repayment abilities”. Despite the concerns with the data discussed above, the ADS was reasonably accurate. We measured the overall accuracy of the ADS based on how well it predicted the target variable. We also measured accuracy for a set of subpopulations and intersections of subpopulations. We used accuracy to assess performance because it is directly related to the question that we are asking. We wanted to know how well the model did at predicting the target variable. The accuracy is measured as the percentage of correct predictions.

The goal of the model was to assign a value of 1 to loan applicants if the model predicted that the applicant would experience payment difficulties. Conversely, the model would assign a value of 0 to loan applicants if the model predicted that the applicant would not experience payment difficulties. The overall accuracy of the model was almost 80%. This is a reasonable accuracy level. In addition, the accuracy of the model varied over the various subpopulations reviewed but the range of accuracies was not of particular concern other than the particular features of age and Region Rating. In addition, we looked at false positive and false negative rates. The model does achieve a relative balance between the false positive rate (18.8%) and the false negative rate (20.8%). However, the rates are still approximately 20% which is higher than we would like.

We have concerns regarding the features that were used to implement the model. These concerns can be seen by reviewing the feature importance plot. The designer of the ADS created ratios from features. The features created by combining other features all have names that begin with “APPS_”. Some of these “user defined” features appear to be reasonable and some do not. Including features that have high correlation with other features can cause the model problems, distort the true importance of features, and increase the difficulty of interpreting the model. The model designer did not remove any features that were highly correlated. A review of the feature plot make it clear that the normalized credit ratings is critically important. However, the true importance is difficult to assess because of the ratios included by the designer. Consequently, we believe that there is room for improvement in the model by improving feature selection.

The ADS is based on a gradient boosting model which appears suited for the purpose for which it is being used. In particular, the ADS uses the LGBMClassifier created by LightGBM. It is a commonly used gradient boosting framework based on decision tree algorithms, used for ranking, classification and other machine learning tasks. We believe the model choice is reasonable as is its implementation.

We used disparate impact as a measure fairness. The disparate impact ratio is a measure of the percentage of favorable outcomes for the underprivileged group divided by the percentage of favorable outcomes for the privileged group. It seems extremely well suited as a fairness measure for this ADS due to the large number of potential subpopulations. It gave us the ability to examine subpopulations from different features which permitted us to gain a deeper understanding of the data and the ADS. As discussed above, we have fairness concerns related to the model, in particular with regard to age and region rating.

The loan applicants are a stakeholder that would likely find the use of both disparate impact and accuracy as appropriate. The overall loan approval percentage for the training dataset was 92%. We have no experience in lending decisions but we were surprised that the approval rate was so high. We wondered whether there could be some type of intervention that occurs prior to the loan application being submitted which filters out applications that are less likely to be approved. If so, it is possible that bias exists in that process. Overall, it appears that loan applicants have a high probability of being approved. This high approval rate acts as a degree of mitigation against lack of fairness. For example, if 100% of the loans were approved it would be difficult to argue that the model was biased. Loan applicants who belong to underprivileged groups would likely be particularly interested in the disparate impact ratio as it relates to them. And, all applicants are likely concerned with the accuracy of the model.

Home Credit Group, as the lender, is a stakeholder that may have a greater concern with accuracy than disparate impact. Home Credit Group would like the ADS to accurately predict whether a borrower will experience payment difficulties. In this regard, model accuracy is extremely important. It is less clear that looking at disparate impact is beneficial to the lender. Examining disparate impact across subpopulations has the potential to reveal bias. The lender may or may not wish for this to be discovered. Increasing fairness would likely reduce precision and this would not be to Home Credit Groups advantage.

5c. Would you be comfortable deploying this ADS?

The ADS produces reasonably accurate results. However, as stated above, we believe the dataset is not appropriate for this ADS and we have concerns about bias. We would recommend that the features related to age and region be removed. We also have concerns with how the data was processed, in particular, the inclusion of a number of ratios that may cause instability in the model and decrease model interpretability. Weighing these factors, we would not be comfortable deploying the ADS in the public sector or industry. We commend the developer of the ADS for their work and we believe that the model has the potential to reach a level that it can be deployed to the public sector or industry, but it is not yet at that level.

5d. What improvements do you recommend?

We would remove age, gender and Region Rating as features. We believe the inclusion of and features related to Region Rating are problematic. Additionally we would review all of the features that were included by the ADS. For example, the table below shows the correlation between certain features and the model output. The correlation between the applicant's average credit score and the model output is -0.60. This is highest absolute correlation that we identified by a considerable amount. A number of features incorporate the credit score and it is likely that some of these features are unnecessary and lead to interpretability issues.

Correlation of Features with Model Ouput

Feature	Correlation
Credit Score	-0.60
Age	-0.21
Loan Size	0.14
Region Population Density	-0.10
Loan Size	-0.08
Income	0.03
Loan Payment	-0.03

The model is well suited for the purpose that it is being used. The model is being assigned a difficult task. It is assessing a large number of features where some have large variability. The model is then being asked to accurately predict a result that was reached by humans where there was likely different individuals making different subjective decisions in similar circumstances. Additionally the loan applications in the training data were from applications that occurred over a period of years. Lending criteria could have changed during this time. All of these issues added to the difficulty of designing the model. We are suggesting potential improvements but we are not criticizing the hard work of the designer.