APPLICATION SCORING MODEL

Korablina Maya
Ryabina Darya
Seraia Anastasia

PROJECT GOAL

Build an interpretable credit scoring model to predict default risk and compare different modeling approaches.

Key restrictions:
Target is only used in approved
Rejected is only used for analysis

DATA OVERVIEW

DATASETS USED:

- APPROVED – APPROVED BIDS WITH TGT_VAR TARGET
- REJECTED – REJECTED BIDS (WITHOUT TARGET)
- TOTAL – REFERENCE DATASET

Types of features:

1. Financial (amounts, income)
2. Behavioral (late payments)
3. Demographic
4. Counters

```
approved: (5252, 41)
rejected: (4749, 41)
total: (10001, 40)
```

- Merged into a single dataset, is_rejected flag added
- Sample sizes: Approved = 5252, Rejected = 4749 (almost 50/50)
- Target is only for approved

SANITY CHECK OF DISTRIBUTIONS

What was checked:
Correctness of data loading
Class balance
Presence of gaps

Result:
Default rate in approved ≈ 15%
It is typical imbalance for credit scoring tasks

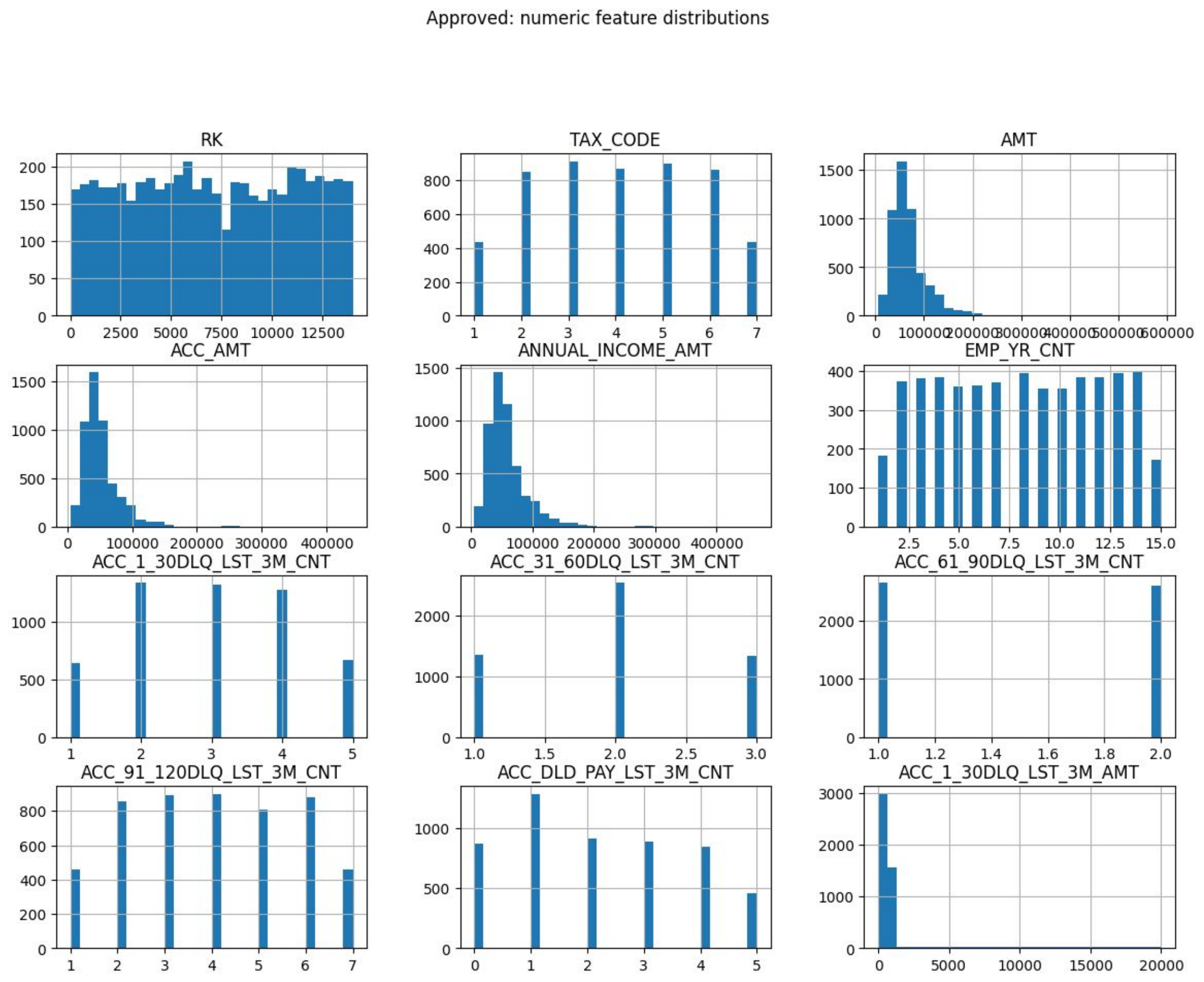
Why it is important: it
justifies the choice of ROC-AUC, KS, and AP metrics

	proportion
TGT_VAR	
0	0.852056
1	0.147944

EDA: NUMERICAL FEATURE DISTRIBUTIONS

We analyzed: distribution shapes, asymmetry and outliers

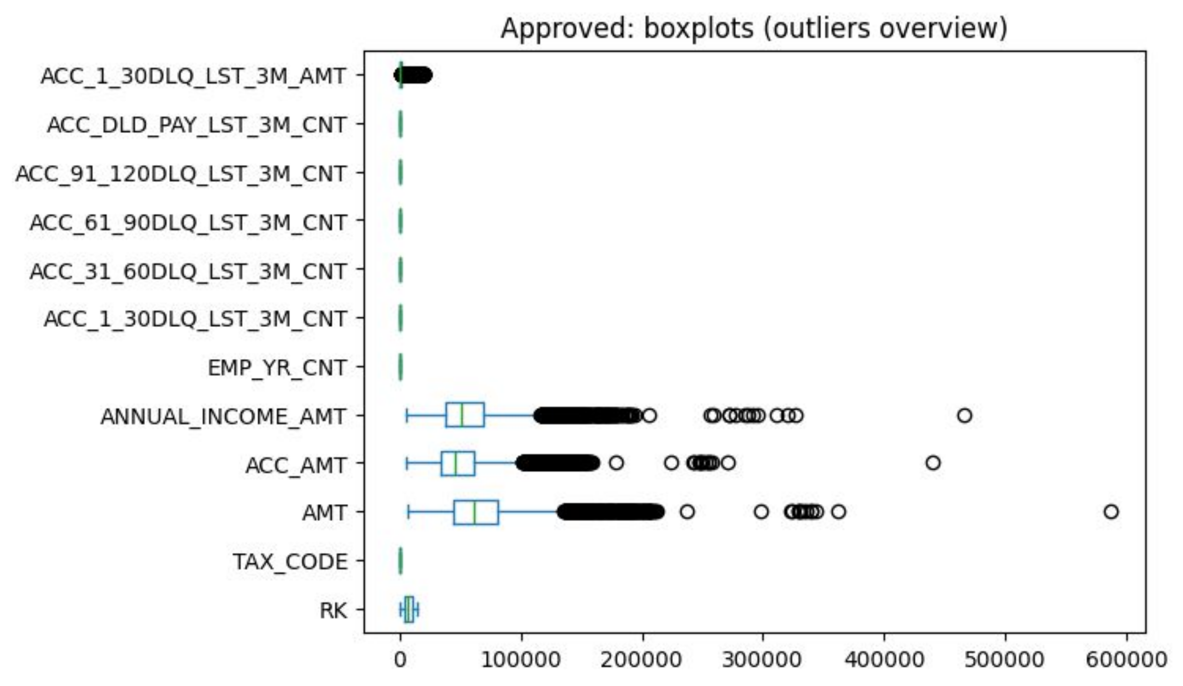
Histograms of numerical features show distributions that are logical for credit scoring, with Strong right-sided asymmetry, outliers, and a discrete structure, which confirms the quality of the data and the validity of using binning and WOE transformation.



EDA: OUTLIERS (BOXPLOTS)

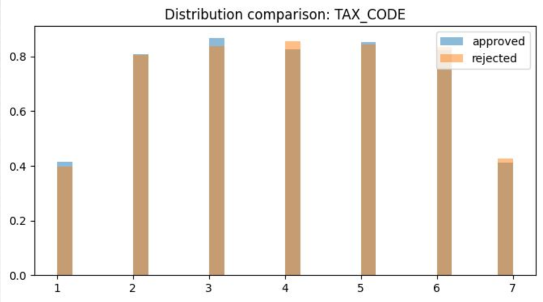
Observations: Significant outliers in monetary and behavioral features, no anomalies or logical errors

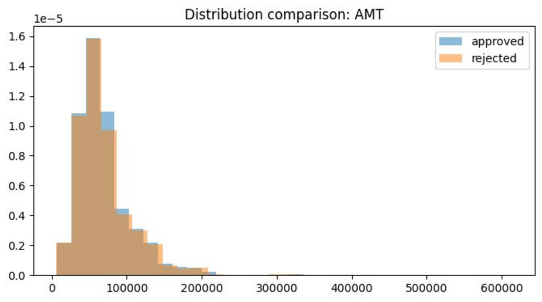
So outliers were not removed and bining + WOE is used for model stability

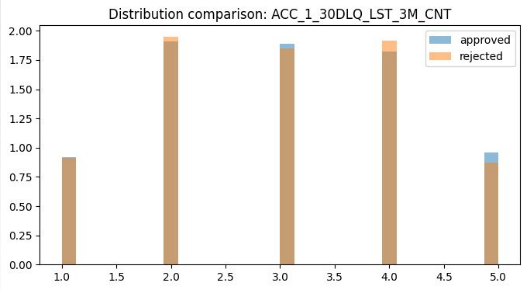


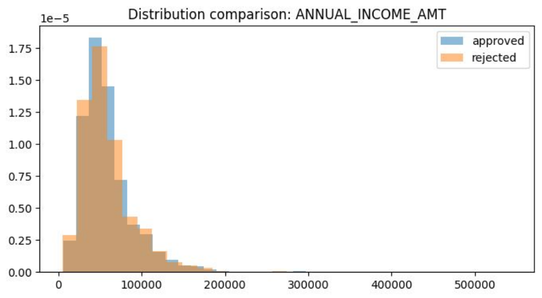
REJECTED ANALYSIS - COMPARE DISTRIBUTIONS

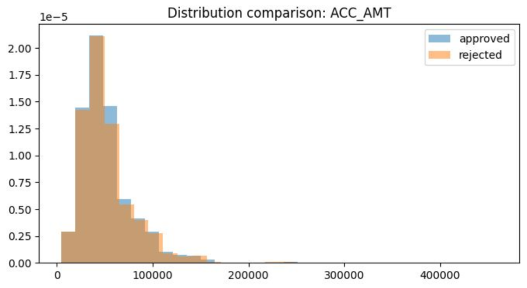
We compared the distributions of key numerical features between approved and rejected. It can be seen that rejected as a group differs in its risk/solvency profile in several features.
The model is trained on approved, so there may be selection bias relative to the entire flow of applications. Therefore, we perform a separate diagnostic analysis for rejected and carefully interpret the findings

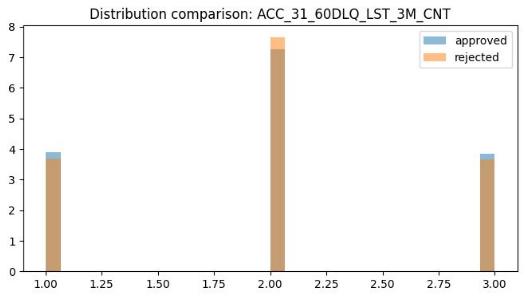


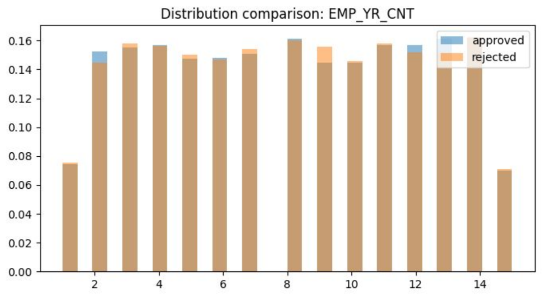












REJECTED ANALYSIS - KS-TEST

The KS-test results showed that the distributions of features for approved and rejected applications are almost identical (KS-stat < 0.03 for all features). This indicates that rejected applications in this dataset do not form a separate risk population based on the observed features. It is likely that the division into approved and rejected in this training dataset does not directly reflect credit risk.

	feature	ks_stat	p_value
24	APPL_PA_LBL_REST_AMT	0.030118	0.021098
18	ACC_APPL_PCL_VAL_AMT	0.025629	0.073817
21	APPL_PA_LQD_AST_AMT	0.022920	0.142554
26	APPL_APPT_MAX_LBL_AMT	0.022869	0.149974
23	APPL_PA_AST_OTH_AMT	0.020236	0.254234
4	ANNUAL_INCOME_AMT	0.019811	0.284860
15	SLN_DR_TRNS_LST_3M_CNT	0.019744	0.280352
12	TOT_OUTSTANDING_31_60_DAY_AMT	0.019011	0.322795
0	RK	0.018594	0.348759
13	ACC_61_90DLQ_LST_3M_AMT	0.018244	0.371617
2	AMT	0.018150	0.387100
3	ACC_AMT	0.018150	0.387100
22	APPL_PA_REST_AMT	0.017512	0.422403
20	APPL_PA_HHD_INC_AMT	0.017101	0.452545
11	ACC_1_30DLQ_LST_3M_AMT	0.015674	0.565091
16	ACC_91_120DLQ_LST_3M_AMT	0.014938	0.626383
7	ACC_31_60DLQ_LST_3M_CNT	0.014226	0.686368
9	ACC_91_120DLQ_LST_3M_CNT	0.013236	0.767749
6	ACC_1_30DLQ_LST_3M_CNT	0.011145	0.911326
8	ACC_61_90DLQ_LST_3M_CNT	0.011073	0.915184

MODEL-BASED RISK DIAGNOSTIC FOR REJECTED

```
Mean PD (approved, valid): 0.1625
Mean PD (rejected): 0.1652
```

We done baseline model applied to rejected (without training) and compared average predicted PD

Result:
Mean PD approved ≈ Mean PD rejected
Therefore decision to reject is not directly related to default risk

BINNING, WOE AND INFORMATION VALUE (IV)

Methodology:

- Supervised tree binning (on train)
- WOE transformation
- Application to valid and rejected

Why WOE:

- Interpretability
- Robustness to outliers
- Suitable for logistic regression

What the IV analysis showed:

- There are features with extremely high IV
- Possible information leakage or post-factum aggregates
- High-quality models require careful interpretation

	feature	IV
11	ACC_1_30DLQ_LST_3M_AMT	11.226807
10	ACC_DLD_PAY_LST_3M_CNT	5.442032
15	SLN_DR_TRNS_LST_3M_CNT	3.577736
0	RK	1.651800
2	AMT	0.236905
3	ACC_AMT	0.236905
4	ANNUAL_INCOME_AMT	0.222866
26	APPL_APPT_MAX_LBL_AMT	0.178451
13	ACC_61_90DLQ_LST_3M_AMT	0.061026
16	ACC_91_120DLQ_LST_3M_AMT	0.053721
20	APPL_PA_HHD_INC_AMT	0.053100
18	ACC_APPL_PCL_VAL_AMT	0.047813
23	APPL_PA_AST_OTH_AMT	0.042925
12	TOT_OUTSTANDING_31_60_DAY_AMT	0.041491
5	EMP_YR_CNT	0.036462
21	APPL_PA_LQD_AST_AMT	0.035732
30	APPL_PA_BUR1_CURR_LMT_AMT	0.035685
24	APPL_PA_LBL_REST_AMT	0.029985
17	APPL_SCR_NO	0.028505
25	APPL_APPT_MAX_AGE_NO	0.022797

BASELINE MODEL: LOGISTIC REGRESSION + WOE
VS
ALTERNATIVE APPROACH RANDOMFOREST

01

LOGISTIC REGRESSION + WOE

The model showed very high quality on validation (AUC/AP close to 1) → data analysis is critical

```
Baseline LogReg(WOE)
AUC(valid): 0.9968
AP(valid): 0.9865
```

02

RANDOMFOREST

The quality is comparable to LogReg but there is no improvement

```
RF AUC(valid): 0.9967
RF AP(valid): 0.9864
```

	model	AUC	KS	AP	bad_rate_true	avg_pred
0	Baseline LogReg (WOE)	0.996785	0.940349	0.986526	0.147843	0.162521
1	RandomForest (WOE)	0.996654	0.940483	0.986431	0.147843	0.146263

Comparison table shows that
A linear interpretable model is sufficient
A more complex model does not provide a gain

THE FINAL CONCLUSION

Baseline LR(WOE) is performed and gives very strong quality

Distribution analysis revealed missing values/tails → FE + WOE are justified

Rejected values differ significantly from approved values → it is important to consider selection bias

Alternatives (RF + another model) provide a comparison of quality/interpretation
Next steps: selecting a cut-off for the business metric, stability/PSI, and checking for leaks

Limitations:

- The training nature of the dataset
- Possible leakage of information in the features
- Lack of temporal validation

The project demonstrates a correct scoring pipeline. The results require careful interpretation when transferred to practice