Loan Defaulter

By Ali Al Tamimi, Abdullah Al Huwaishel

CONTENTS

01
Introduction

1.Brief Description of the Project

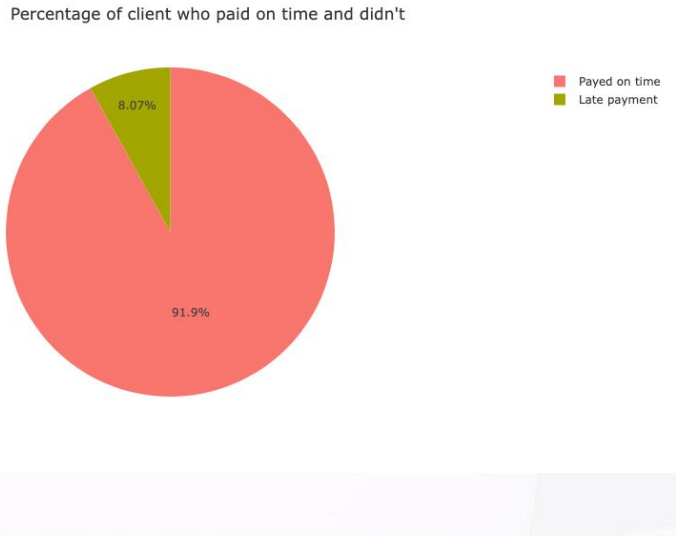
2.Project Objectives

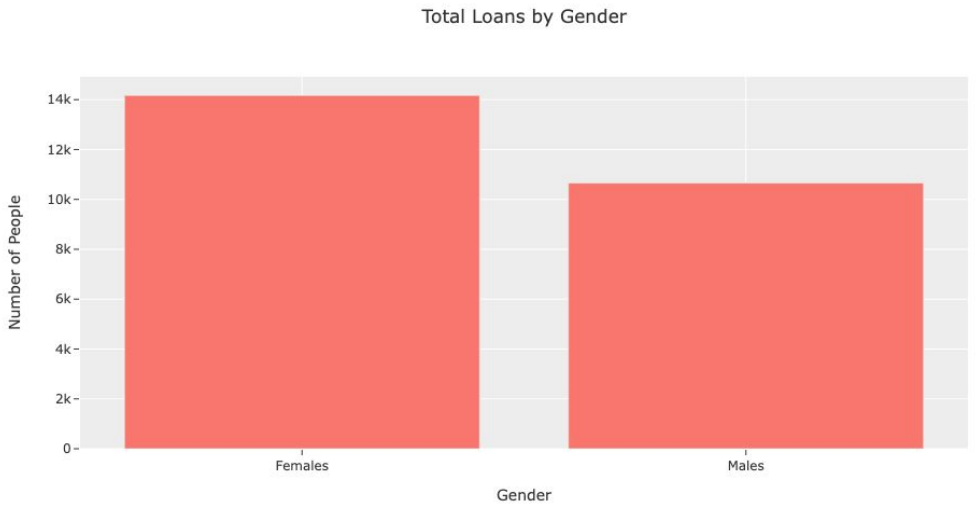
3.Data & Data source/s.

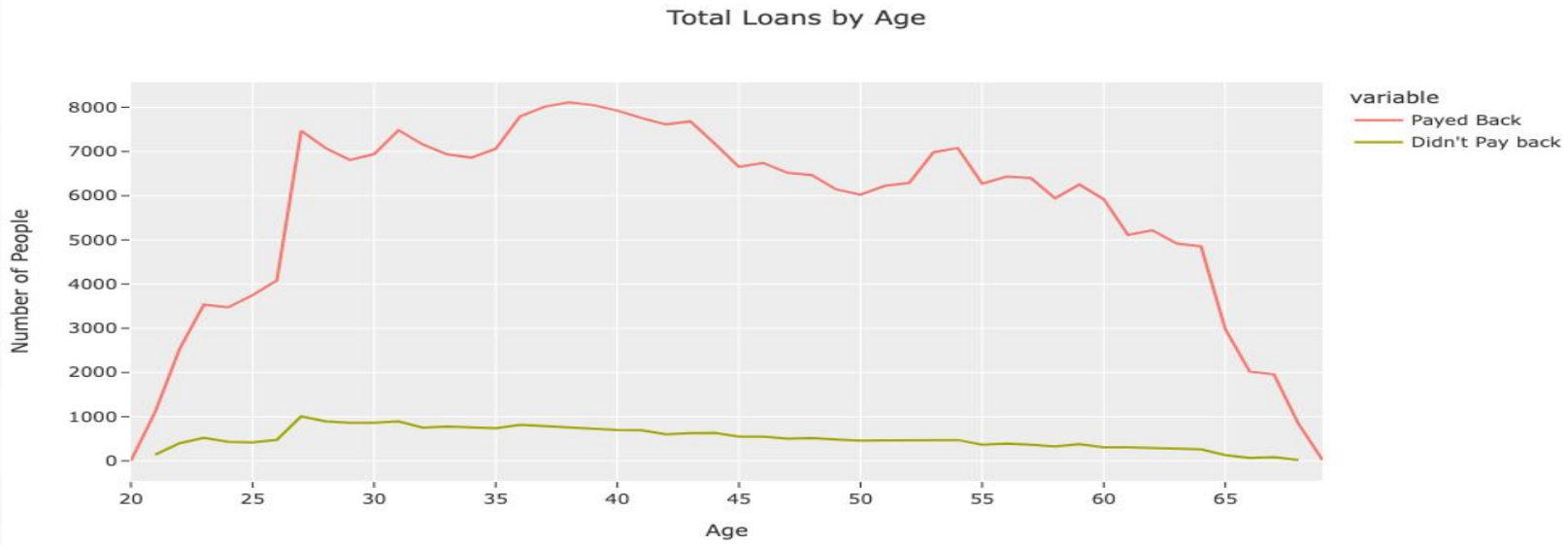
02

Exploratory Data Analysis

visualisations and analysis from the Visualizations







03
Data Preparation

1.Data Cleaning

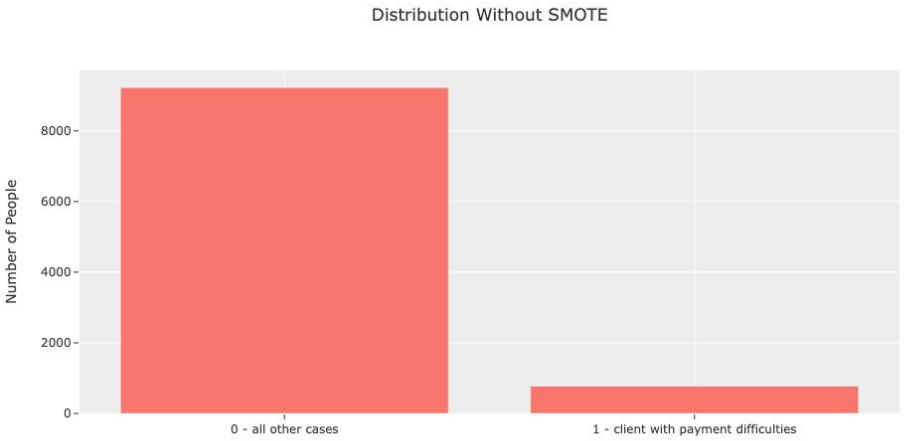
2.Feature Selection

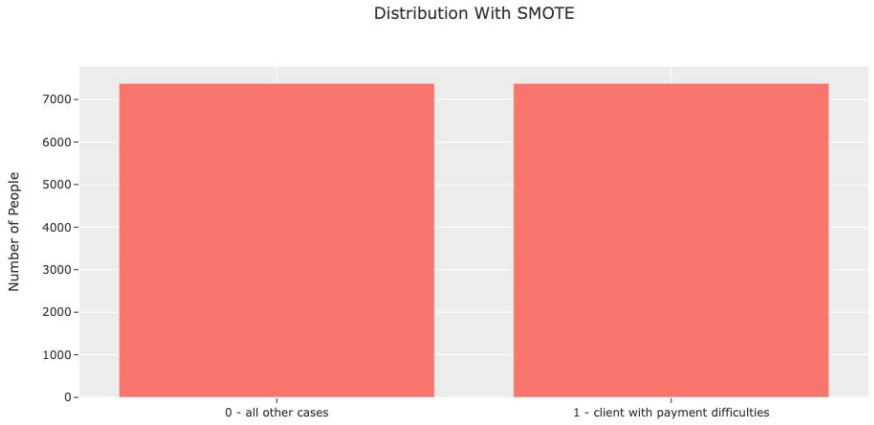
3.Feature Engineering

4.models & methodologies

5.The final dataset

Handling Imbalance Data





The final dataset chosen for building the model

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)
NAME_CONTRACT_TYPE	Identification if loan is cash or revolving
CODE_GENDER	Gender of the client
FLAG_OWN_CAR	Flag if the client owns a car
FLAG_OWN_REALTY	Flag if client owns a house or flat
CNT_CHILDREN	Number of children the client has
AMT_INCOME_TOTAL	Income of the client
AMT_CREDIT	Credit amount of the loan
AMT_ANNUITY	Loan annuity
AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given
NAME_TYPE_SUITE	Who was accompanying client when he was applying for the loan
NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,…)
NAME_EDUCATION_TYPE	Level of highest education the client achieved
NAME_FAMILY_STATUS	Family status of the client
NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)
REGION_POPULATION_RELATIVE	Normalized population of region where client lives (higher number means the client lives in more populated region)
DAYS_BIRTH	Client's age in days at the time of application
DAYS_EMPLOYED	How many days before the application the person started current employment
DAYS_REGISTRATION	How many days before the application did client change his registration

04
Results

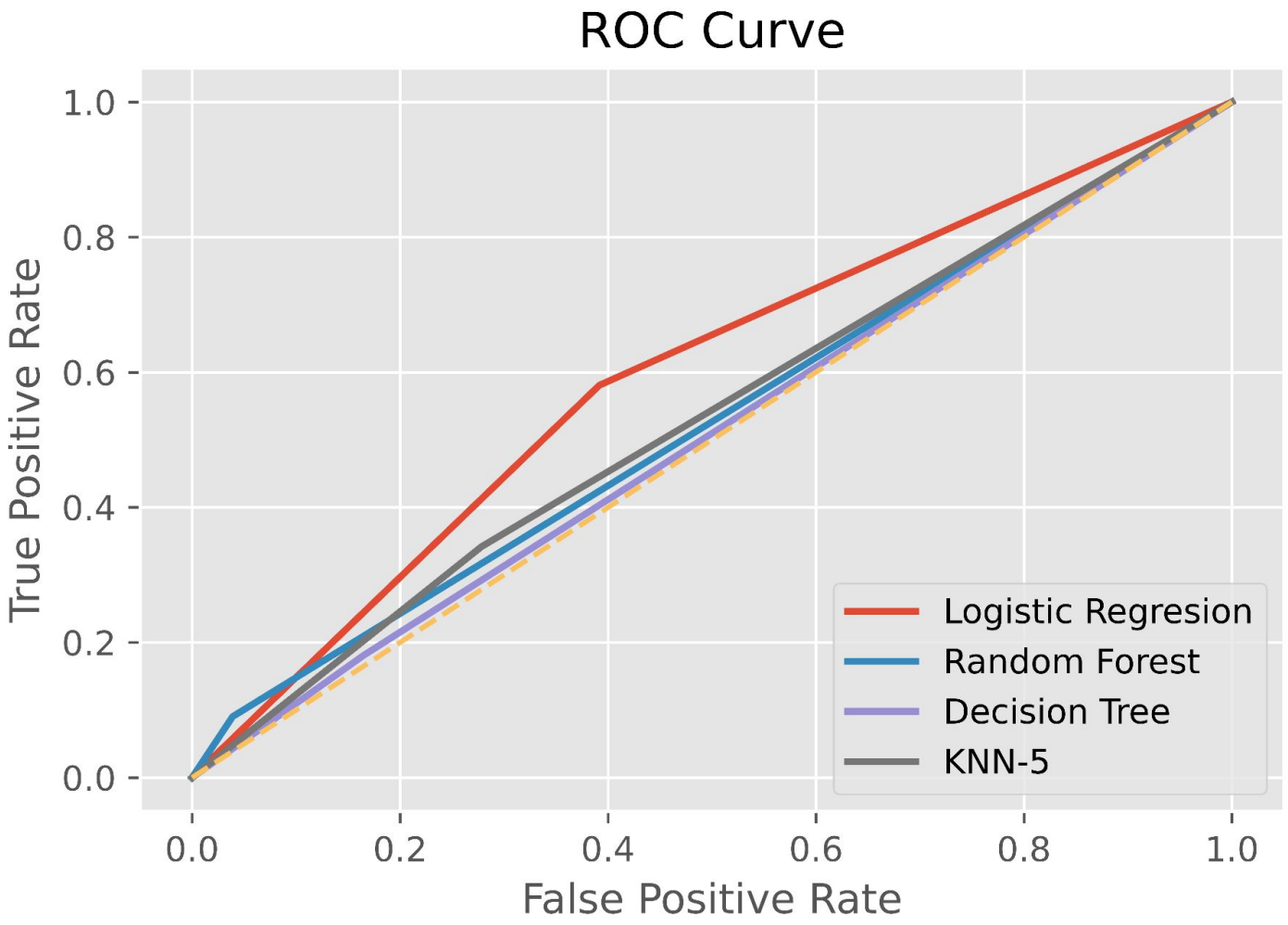
why a particular model is the best & Results model

	precision	recall	f1-score	accuracy
RandomForest	0.16	0.10	0.13	0.88
Logistic Regression	0.12	0.63	0.20	0.60
Decision Tree	0.10	0.16	0.12	0.82
KNeighbors	0.10	0.36	0.16	0.69

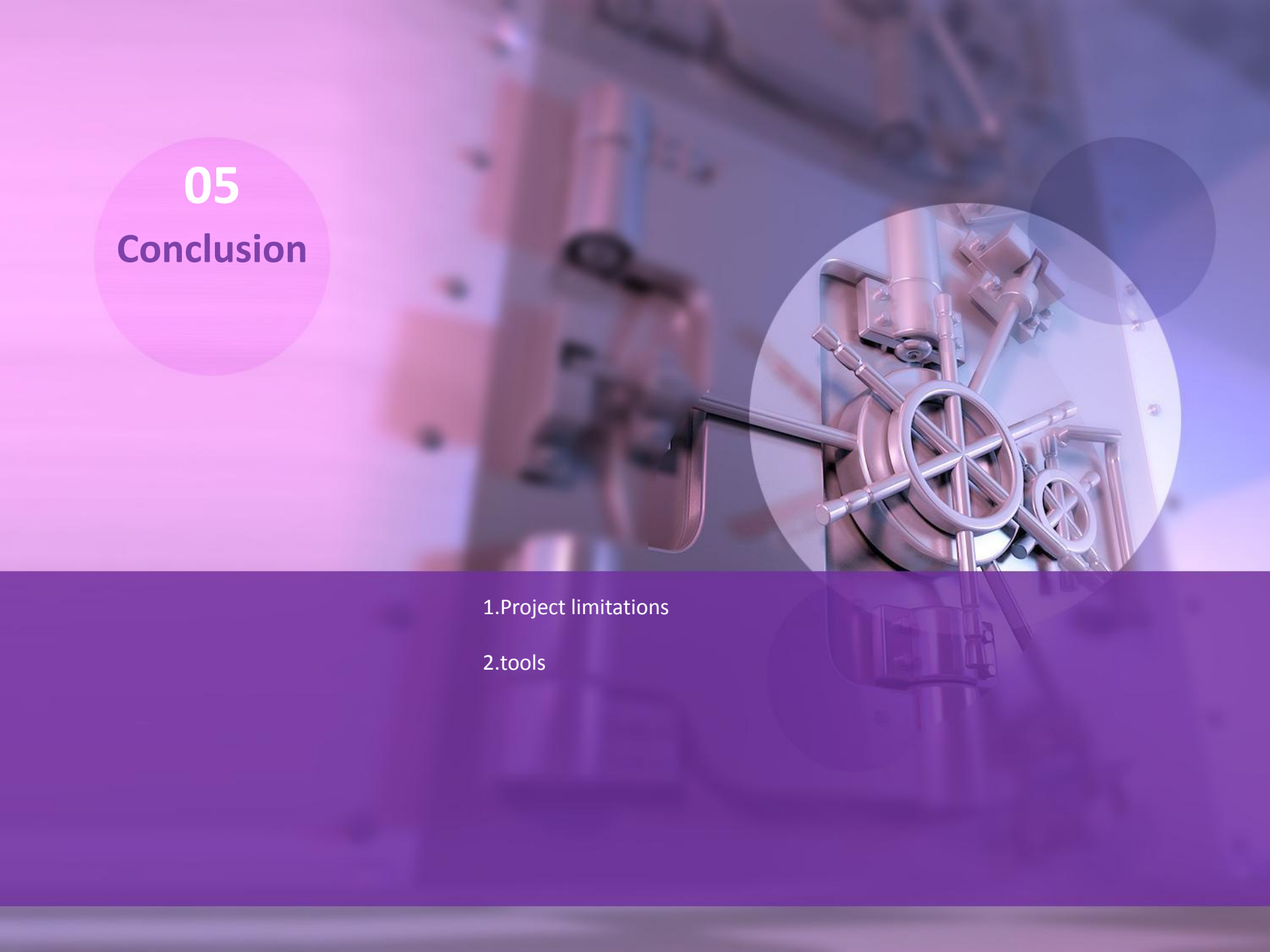
Confusion Matrix for Logistic Regression

Actual / Predicted	0	1	All
0	33832	22705	56537
1	1823	3142	4965
All	35655	25847	61502

ROC & AUC


ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regresion
Random Forest
Decision Tree
KNN-5

05
Conclusion



1.Project limitations

2.tools


THANK
YOU