



Credit Risk Classification

Shashin Sharan

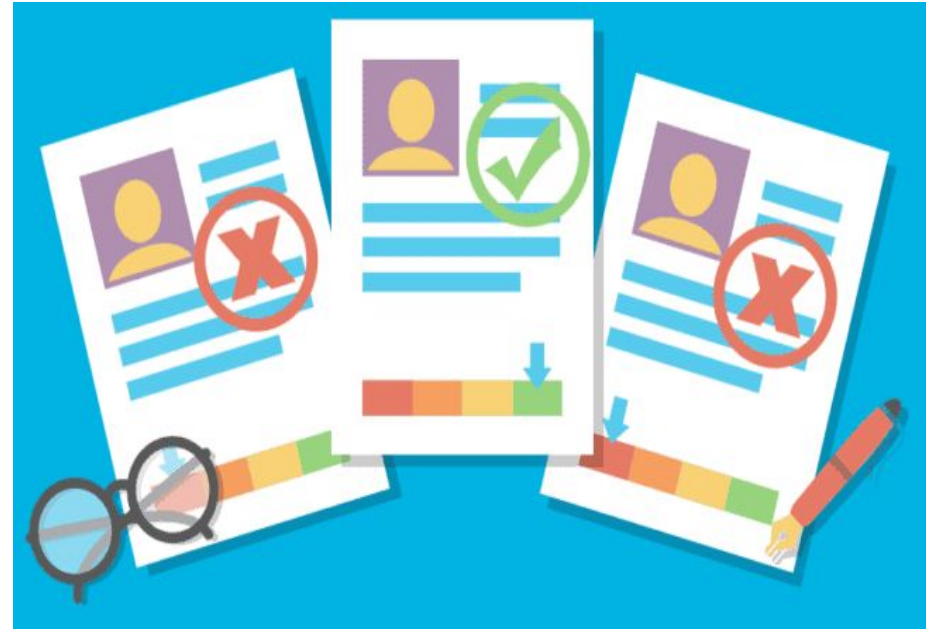
Problem Identification

- ❖ Every year many customers approach banks or financial institutions for loan



Problem Identification

- ❖ Banks need to know credibility of customers before approving loans



Who might care

Banks



Credit Unions



And many more....

What factors might affect?



- ❖ Number of Jobs
- ❖ Duration of loan
- ❖ Type of housing
- ❖ Purpose of loan
- ❖ Type of account
- ❖ Gender of customers



Customer Information

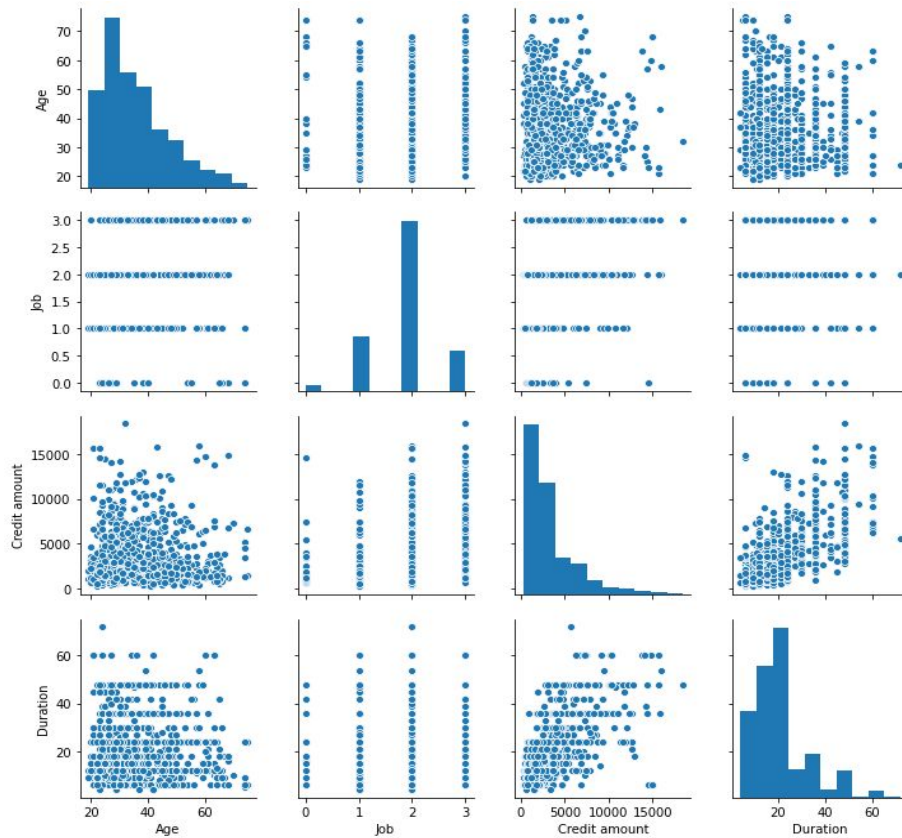
Data Information



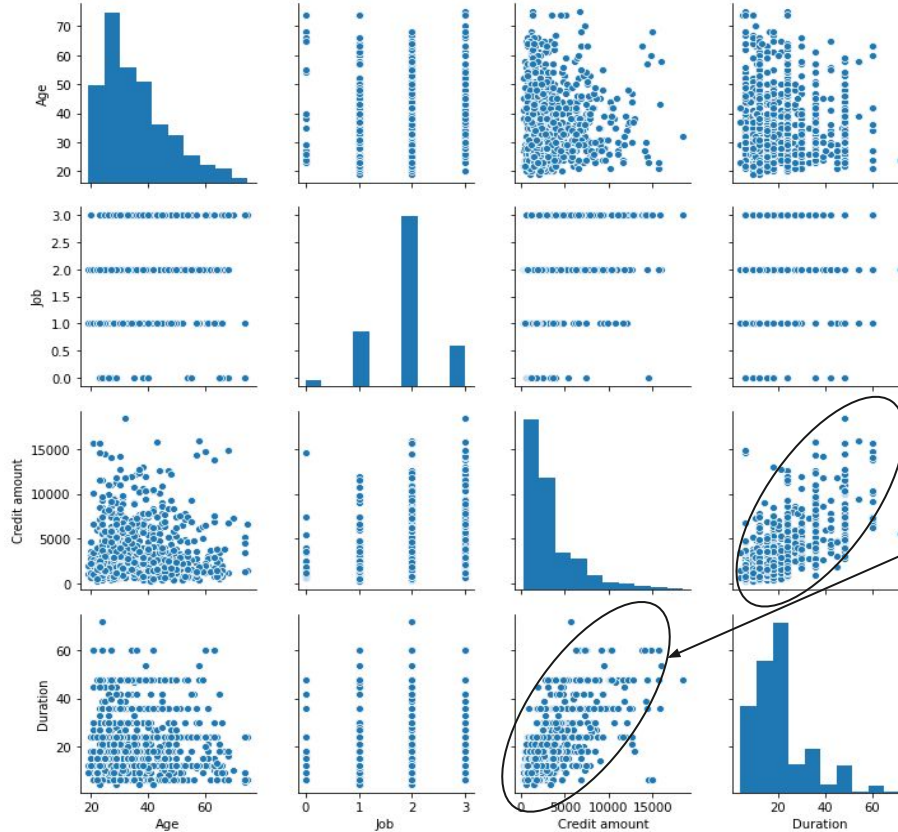
- ❖ Source: UCI ML Repository
- ❖ Number of customers: 1000
- ❖ Number of fields per customer: 10

Age	Sex	Job	Housing	Saving accounts	Checking account	Credit amount	Duration	Purpose	Risk
67	male	2	own	No Account	little	1169	6	radio/TV	good
22	female	2	own	little	moderate	5951	48	radio/TV	bad
49	male	1	own	little	No Account	2096	12	education	good
45	male	2	free	little	little	7882	42	furniture/equipment	good
53	male	2	free	little	little	4870	24	car	bad

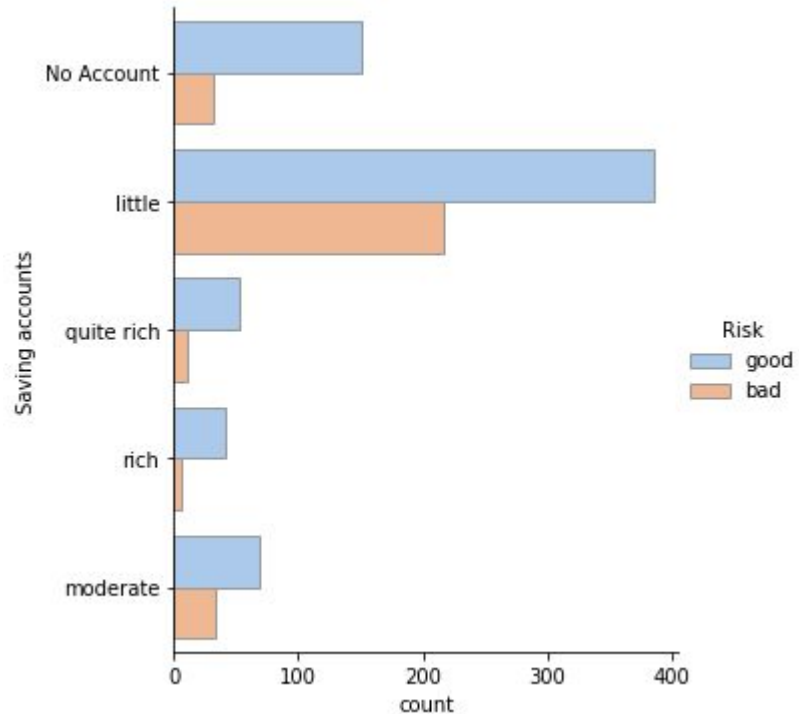
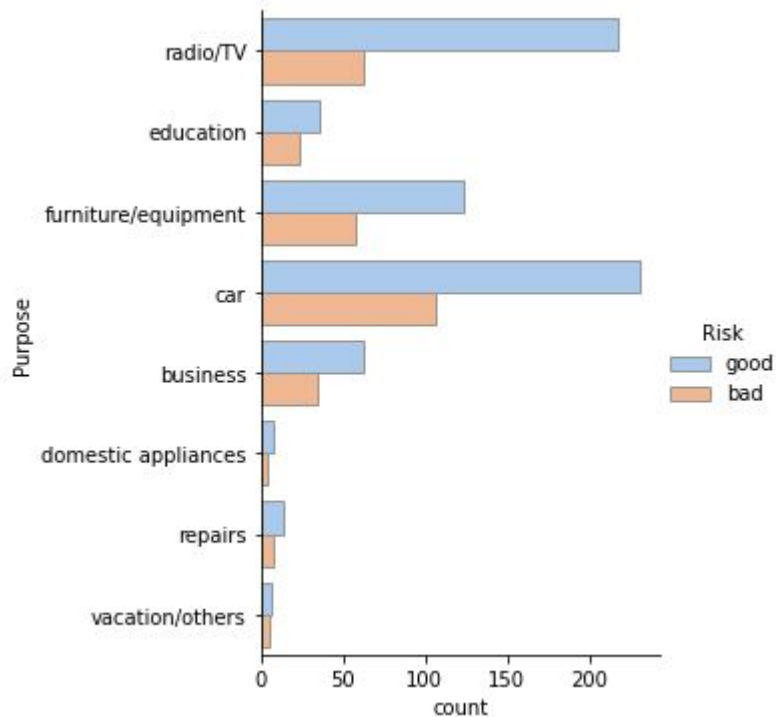
Exploratory Data Analysis



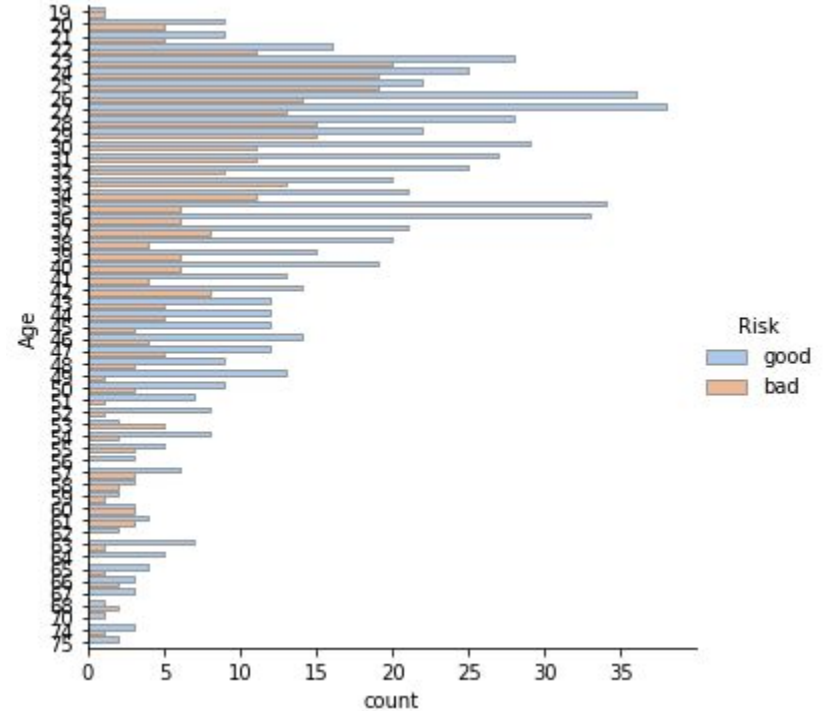
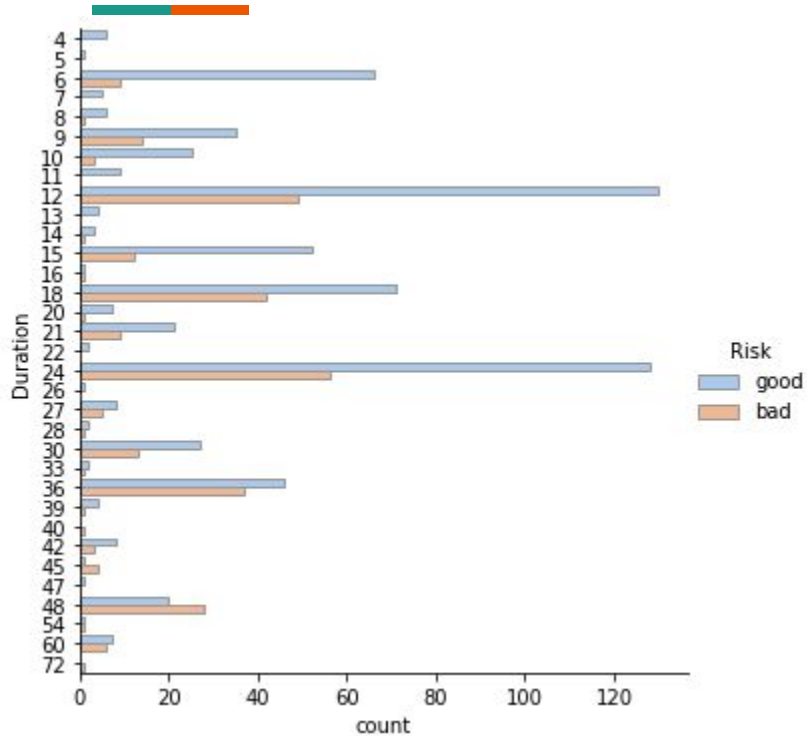
Exploratory Data Analysis



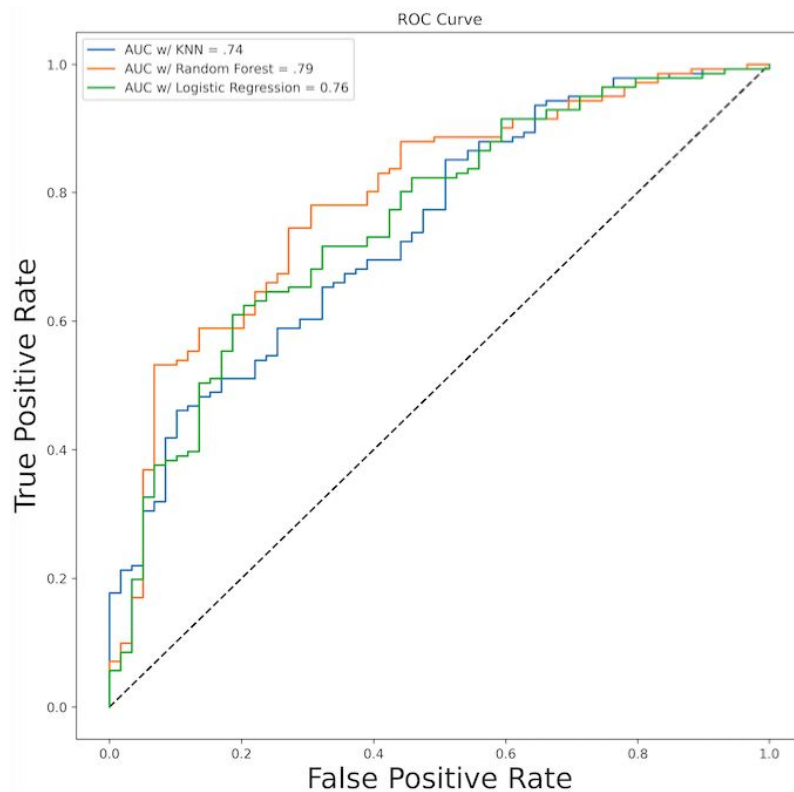
Different factors affecting Risk



Different factors affecting Risk



Data Modeling



Data Modeling



Model	Accuracy	AUC Score
KNN	.76	.74
Random Forest	.74	.79
Logistic Regression	.74	.76

Recommendations and key findings



- ❖ Random Forest seems to outperform other models
- ❖ Currently accuracy is around 75%
- ❖ More data can improve the accuracy