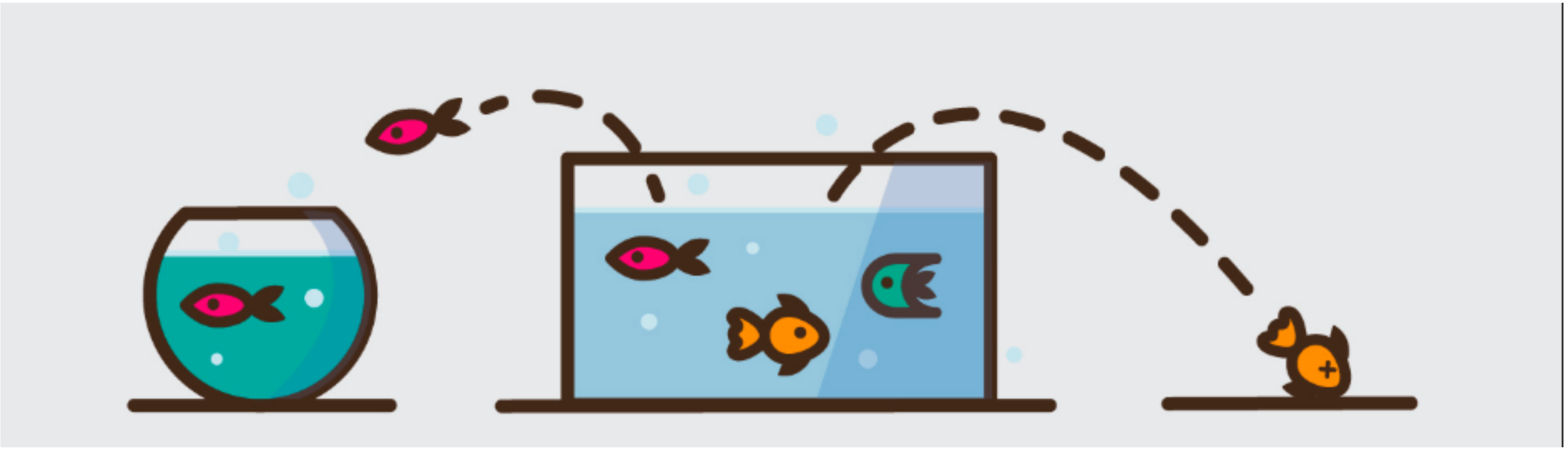
ReDI School of
Digital Integration

Predicting Churn for Bank Customers

Marina Trofimovich

Introduction

Who's going to leave?



Customer churn

Let's try to predict!

Profit

- predict future revenue;
- to identify, address, and get back customers that are likely to churn;
- identify and improve upon areas where customer service is lacking.

Problem

Bank customer churn dataset: 14 features, 10.000 customers.

Target

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

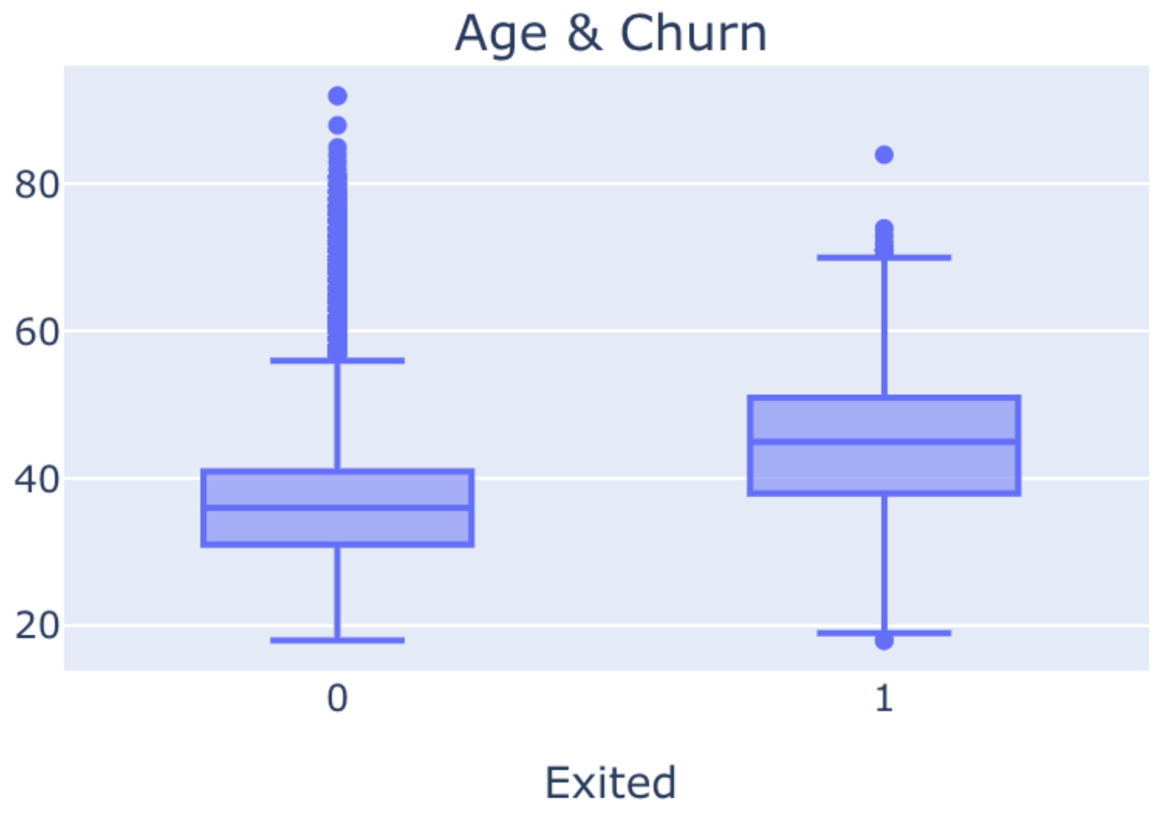
Independent variables

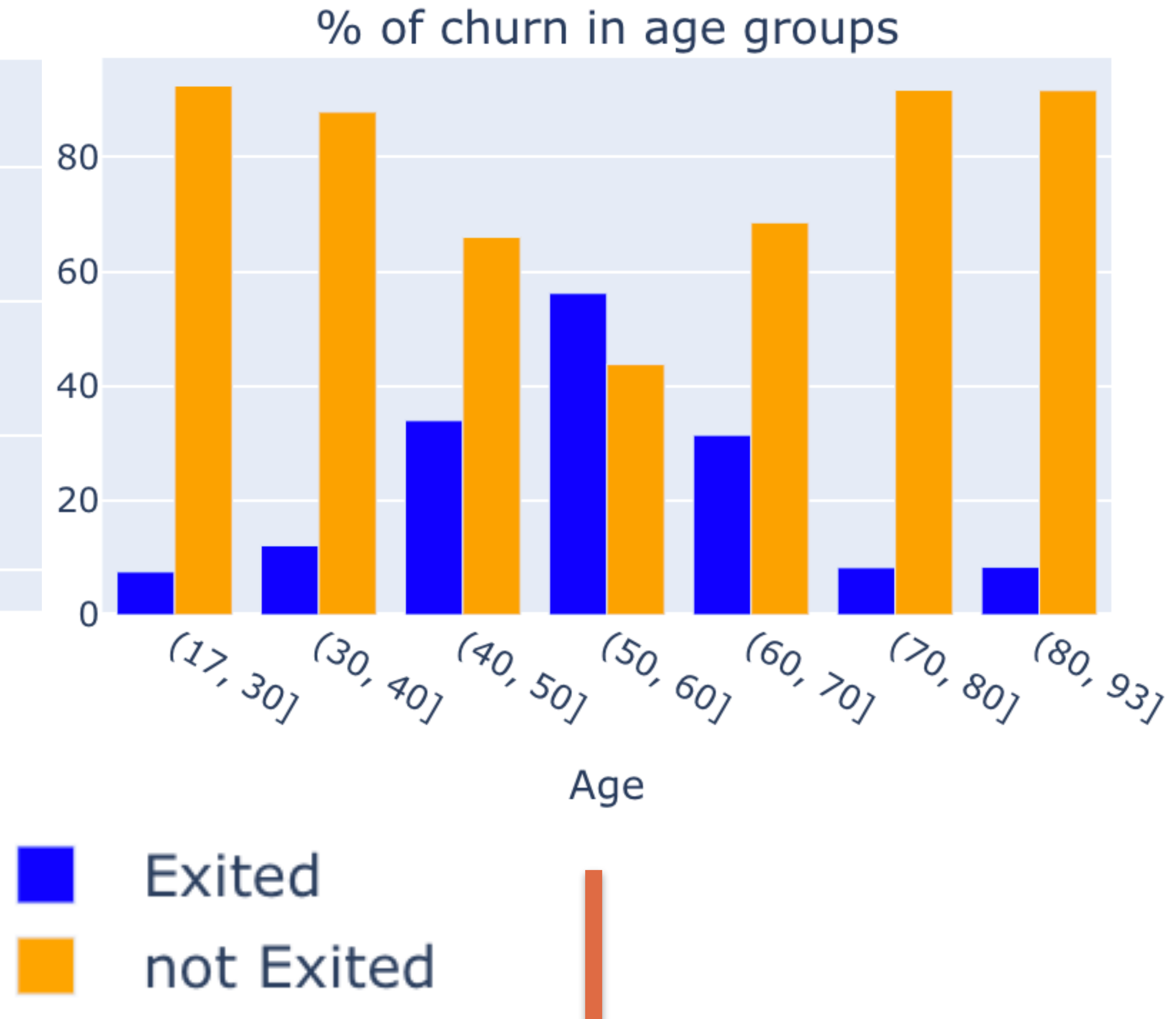
Objectives

- **identify and visualize which factors contribute to the customer churn;**
- **build a prediction model that will classify if a customer is going to churn or not.**

Features that contribute the most

Age

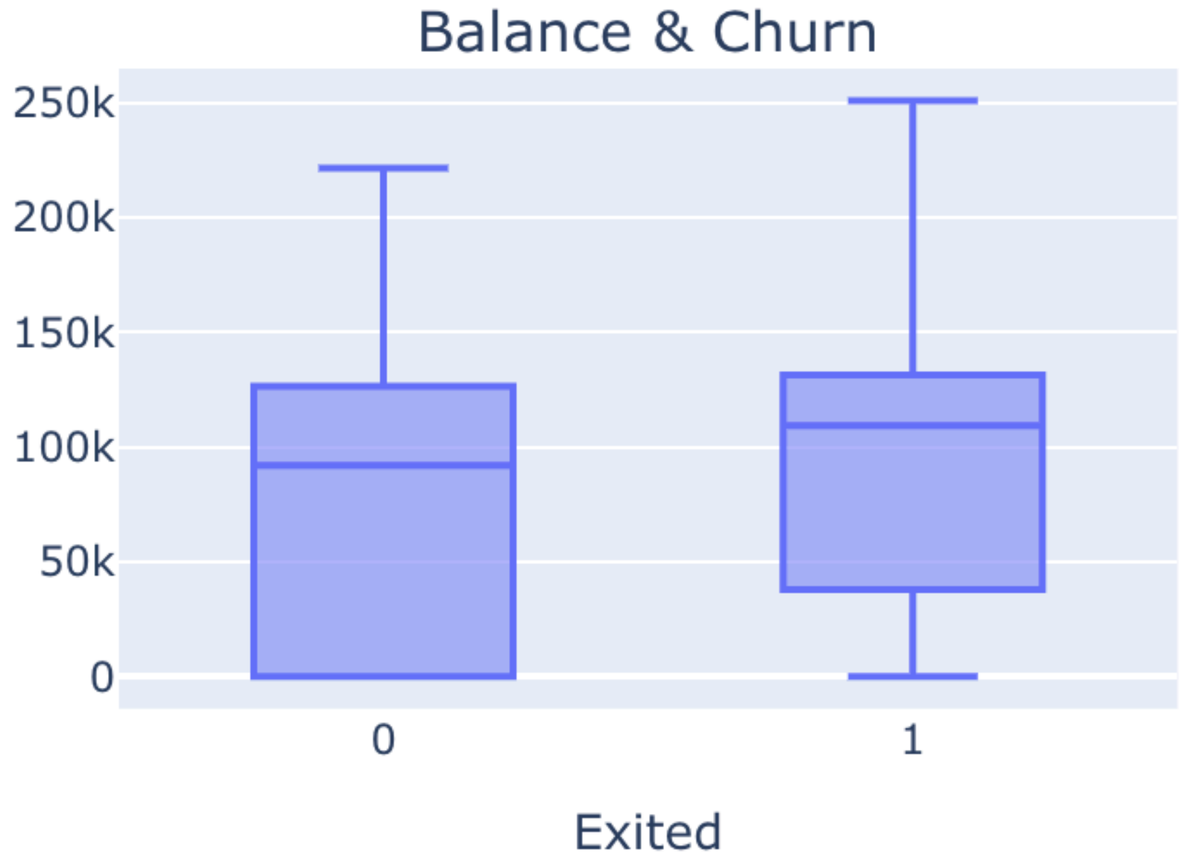


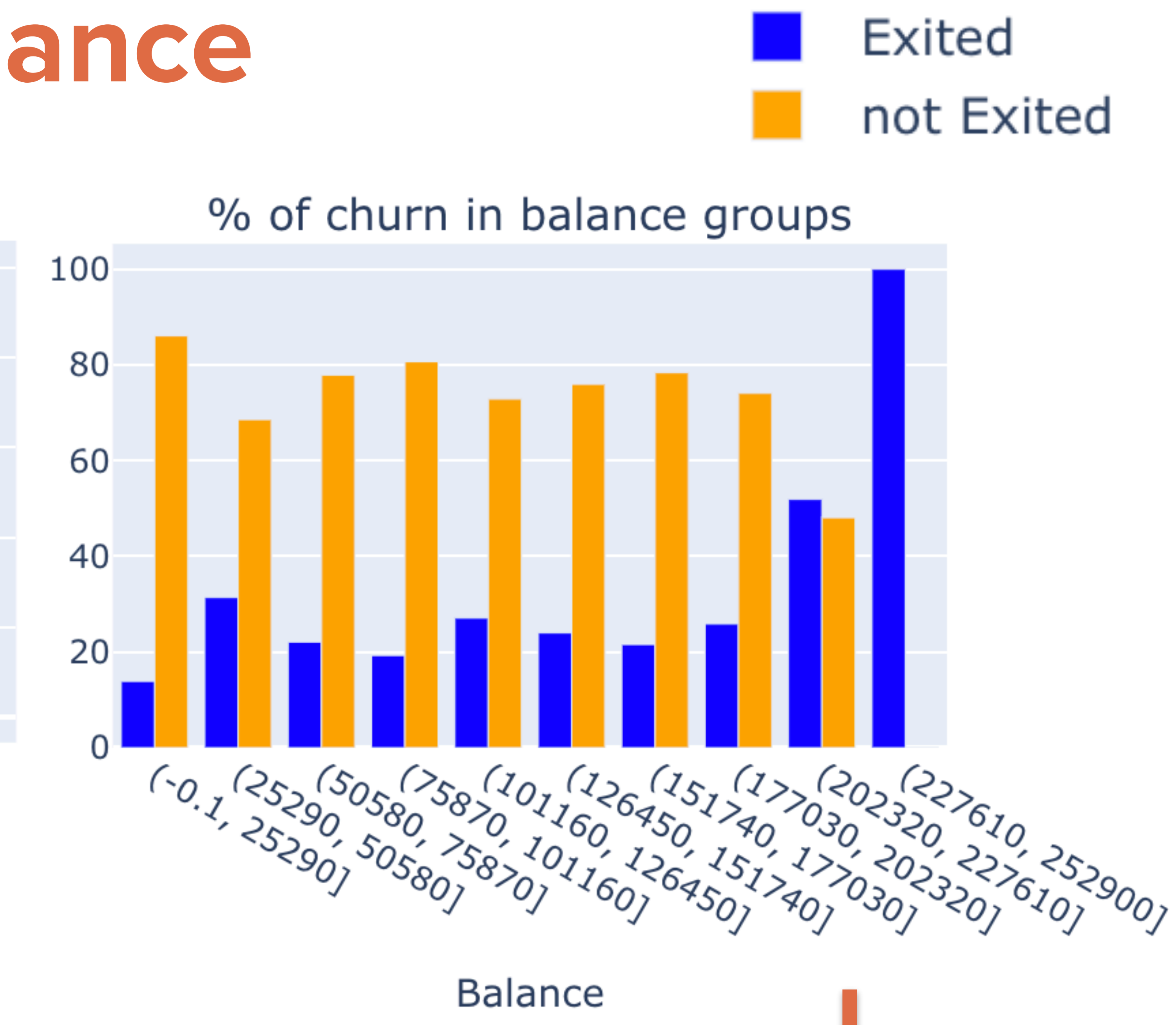


mean age
for “Exited” - 45 years,
“not Exited” - 37 years.

the highest churn (> 56 %)
for age range (50,60]

Balance



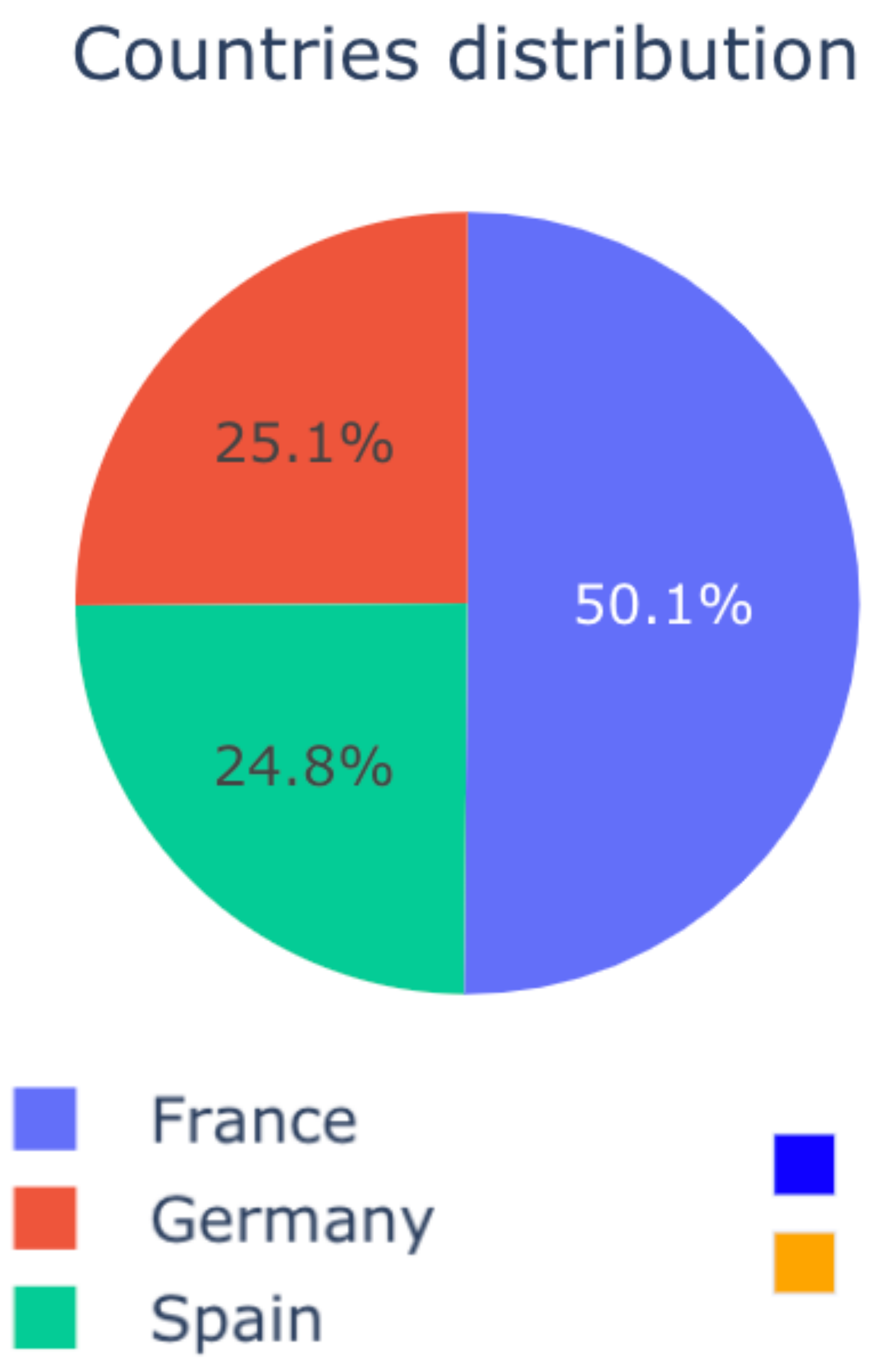


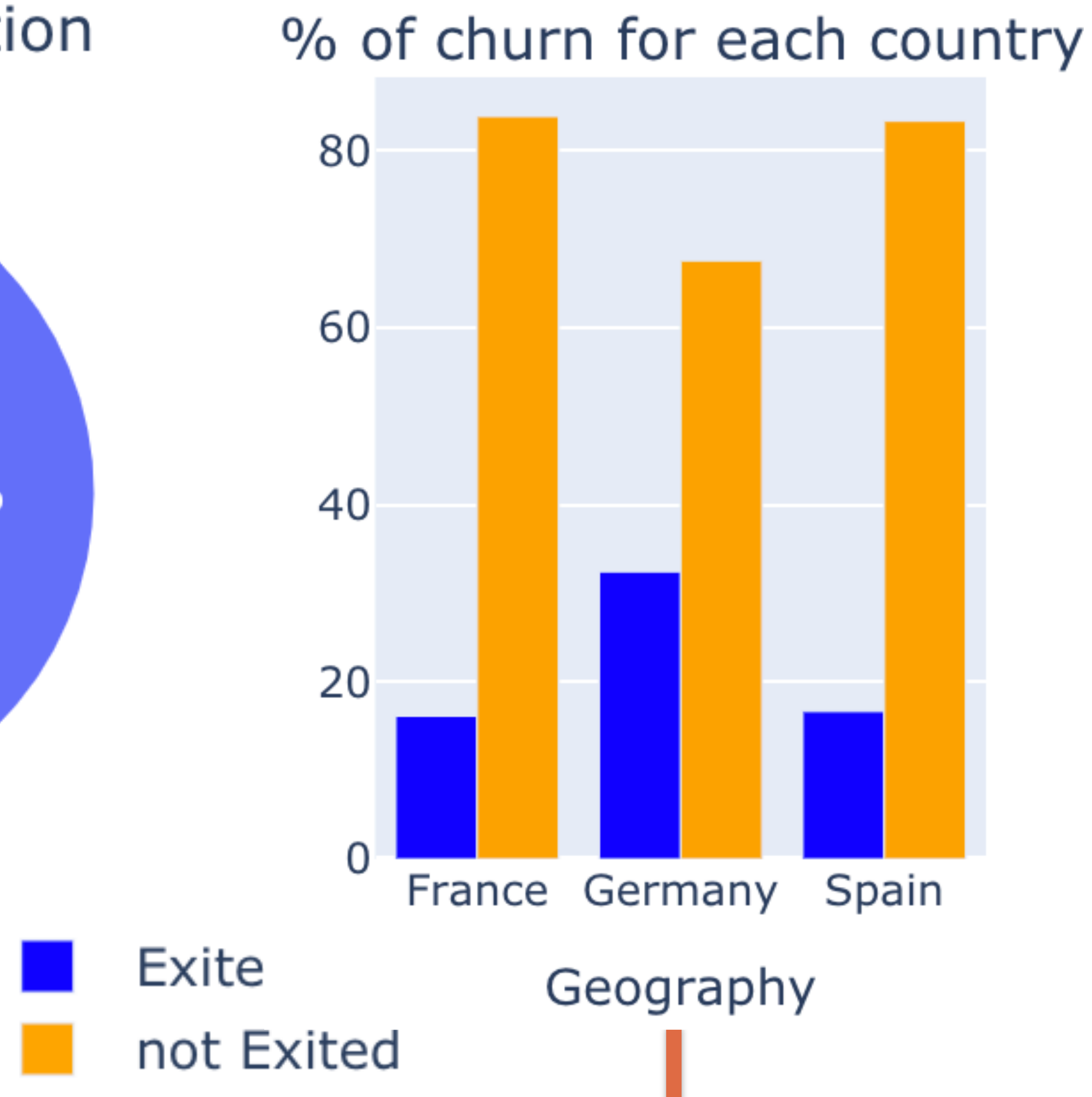
mean balance
for “Exited” - 91.109,
“not Exited” - 72.745.

the highest churn (up to 100 %)
for balance > 202.320
only 29 customers

Features that contribute the most

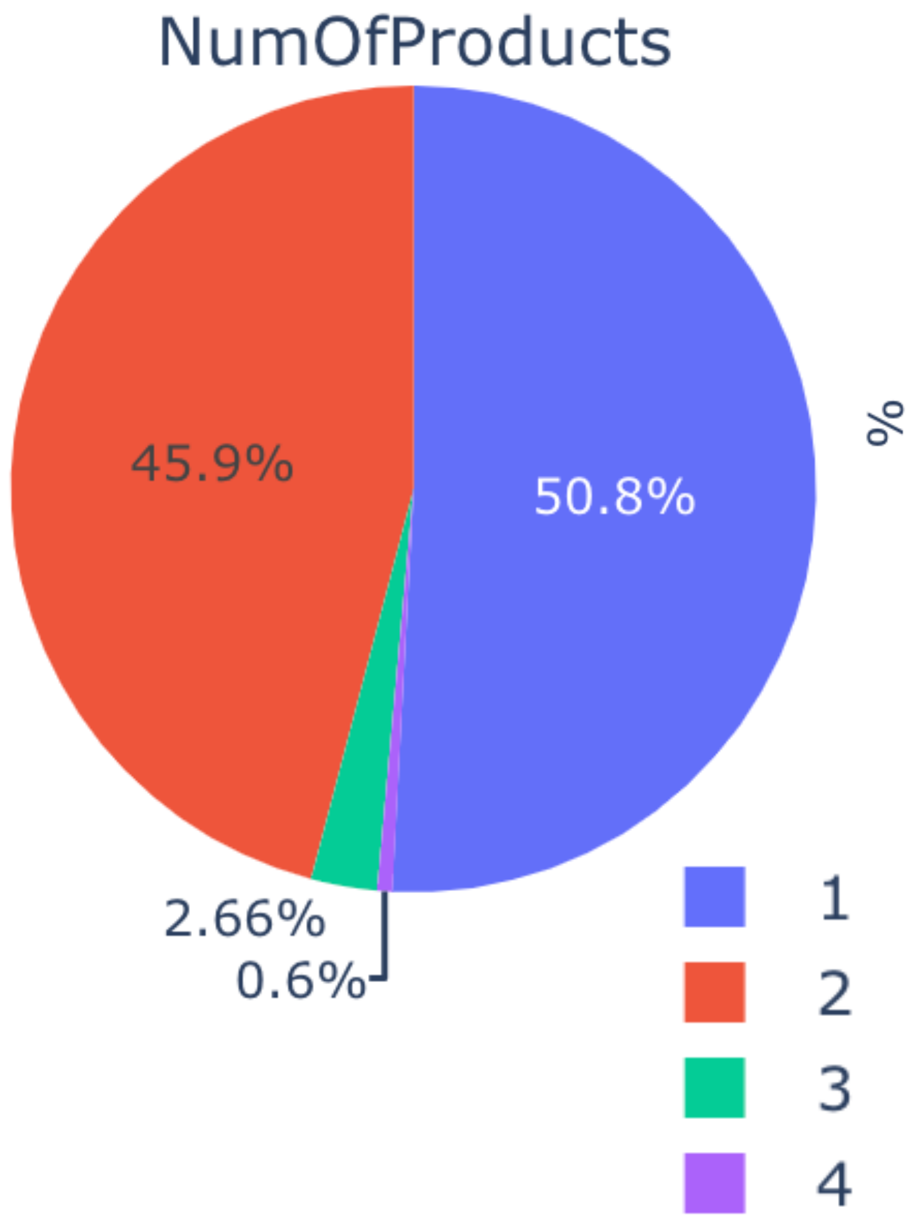
Geography

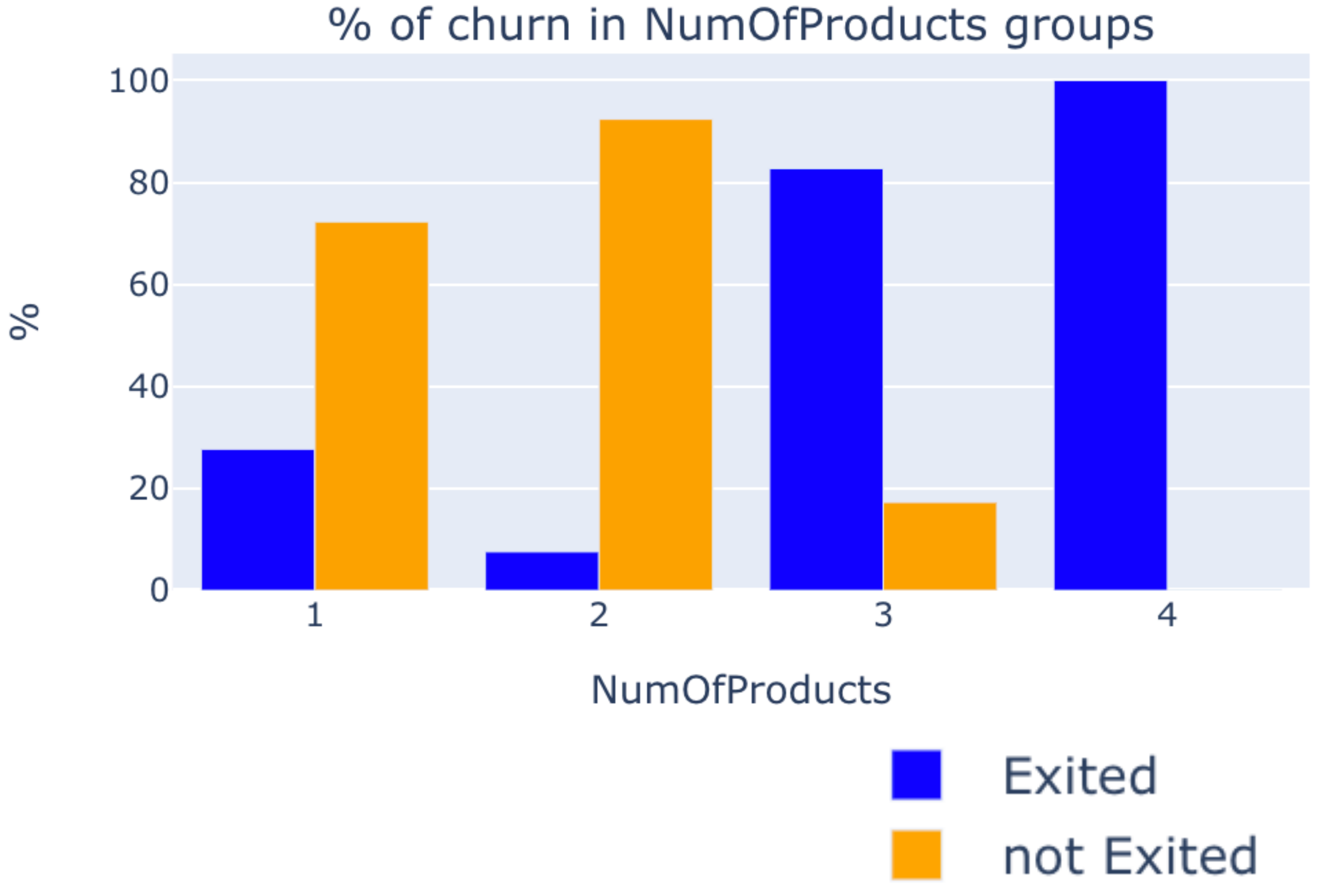




the highest churn (> 32 %) for Germany

NumOfProducts





the highest churn: 83% - 3 products,
100% - 4 products

Machine learning models

	Accuracy	Precision	Recall	F1
Logistic regression	0.85	0.74	0.50	0.60
K nearest neighbours	0.85	0.78	0.39	0.52
Support Vector Machine	0.86	0.84	0.43	0.57
Random Forest	0.87	0.78	0.52	0.63

52% of actual "Exited" customers are predicted correctly.

78% of predicted to be "Exited" customers are actual “Exited”.

Conclusions

Applications

- **developing retention programs for high-risk groups of customers;**
- **further research to identify reasons for high churn (for example, for Germany).**

ReDI School of Digital Integration

THANK YOU!