# *WHO WILL LEAVE ?*

CAN YOU PREDICT IF BANK CUSTOMERS WILL TURNOVER NEXT CYCLE?

羅健華

# *Target*

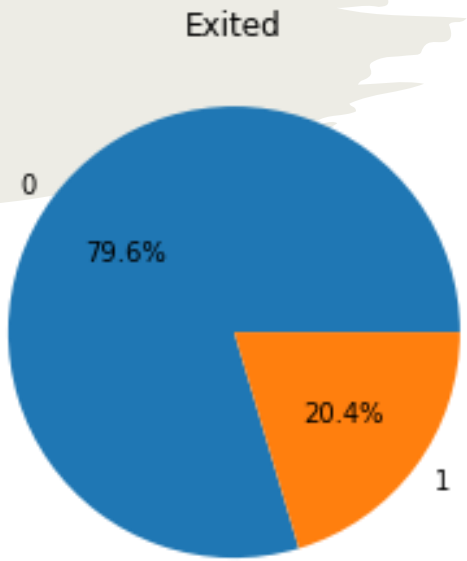
Given a Bank customer, can we build a classifier which can determine whether they will leave in the next 6 months or not?

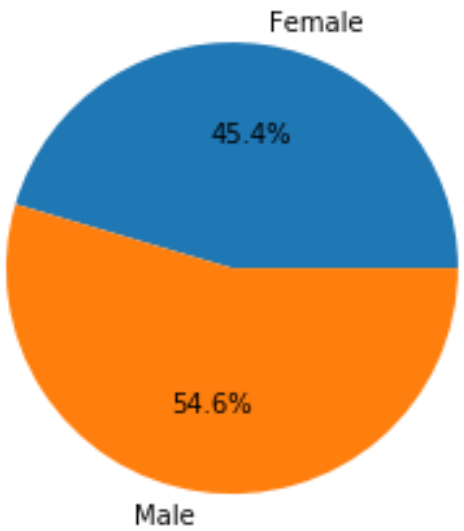
# *Data Source*

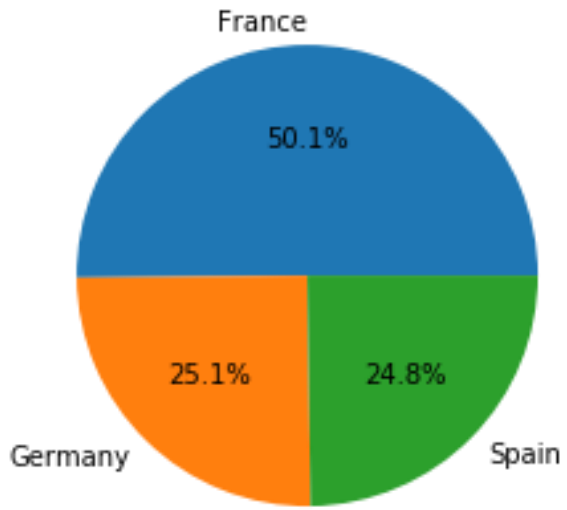
https://www.kaggle.com/barelydedicated/bank-customer-churn-modeling

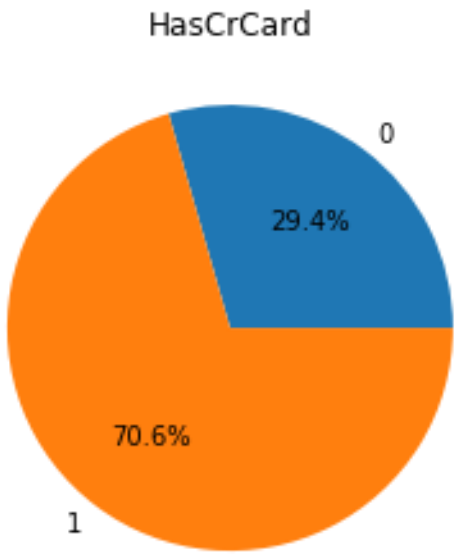
kaggle

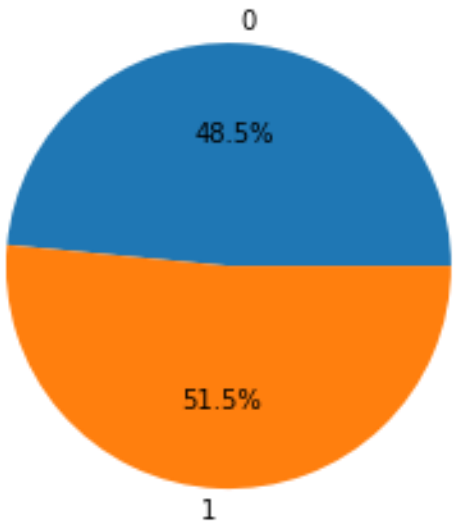
| Valuables | Definition |
| --- | --- |
| **dependent variable** | |
| Exited | = 1 if customer leaves |
| **independent variable** | |
| Age | |
| Gender | Male, Female |
| Geography | France, Germany, Spain |
| Estimated Salary | |
| Tenure | |
| Balance | |
| Credit Score | |
| Number Of Products | |
| Has Credit Card | = 1 if customer has |
| Is Active Member | = 1 if customer is |


Exited
0
79.6%
20.4%
1

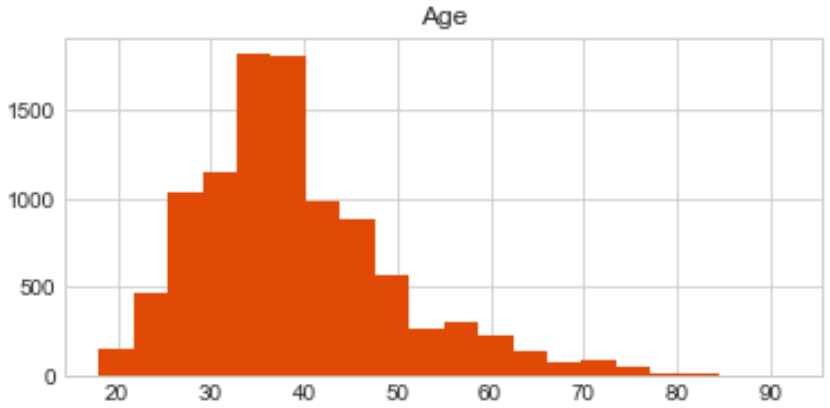

Gender
Female
45.4%
54.6%
Male

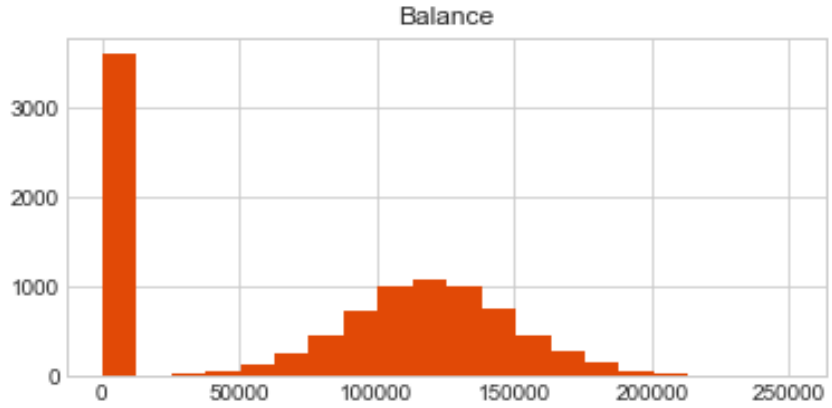

Geography
France
50.1%
25.1%
24.8%
Germany
Spain

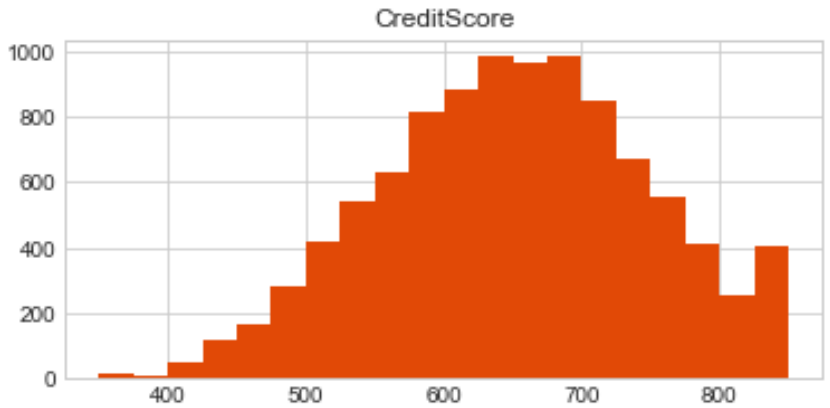

HasCrCard
0
29.4%
70.6%
1


IsActiveMember
0
48.5%
51.5%
1

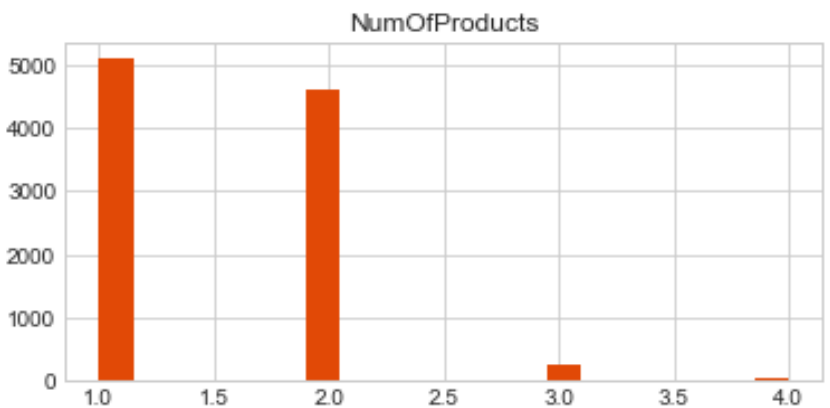
| | Credit Score | Age | Tenure | Balance | Number Of Products | Estimated Salary |
|---|---|---|---|---|---|---|
| **count** | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| **mean** | 650.53 | 38.92 | 5.012 | 76485.89 | 1.53 | 100090.24 |
| **Std** | 96.65 | 10.49 | 2.89 | 62397.41 | 0.58 | 57510.49 |
| **min** | 350.00 | 18.00 | 0.00 | 0.00 | 1.00 | 11.58 |
| **25%** | 584.00 | 32.00 | 3.00 | 0.00 | 1.00 | 51002.11 |
| **50%** | 652.00 | 37.00 | 5.00 | 97198.54 | 1.00 | 100193.91 |
| **75%** | 718.00 | 44.00 | 7.00 | 127644.24 | 2.00 | 149388.25 |
| **max** | 850.00 | 92.00 | 10.00 | 250898.09 | 4.00 | 199992.48 |

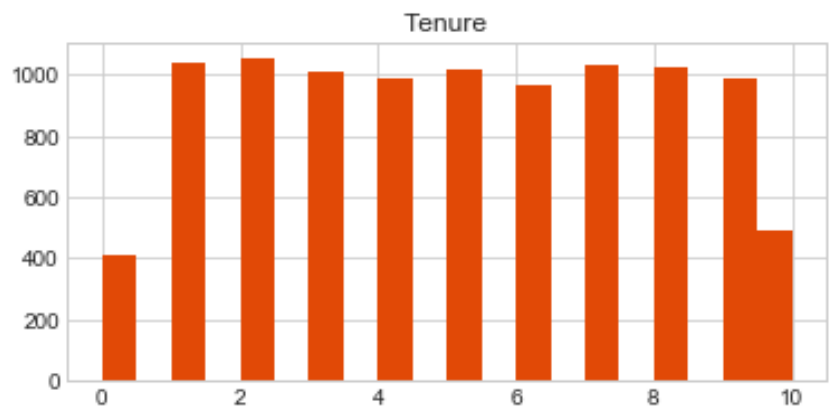

Age
1500
1000
500
0
20
30
40
50
60
70
80
90

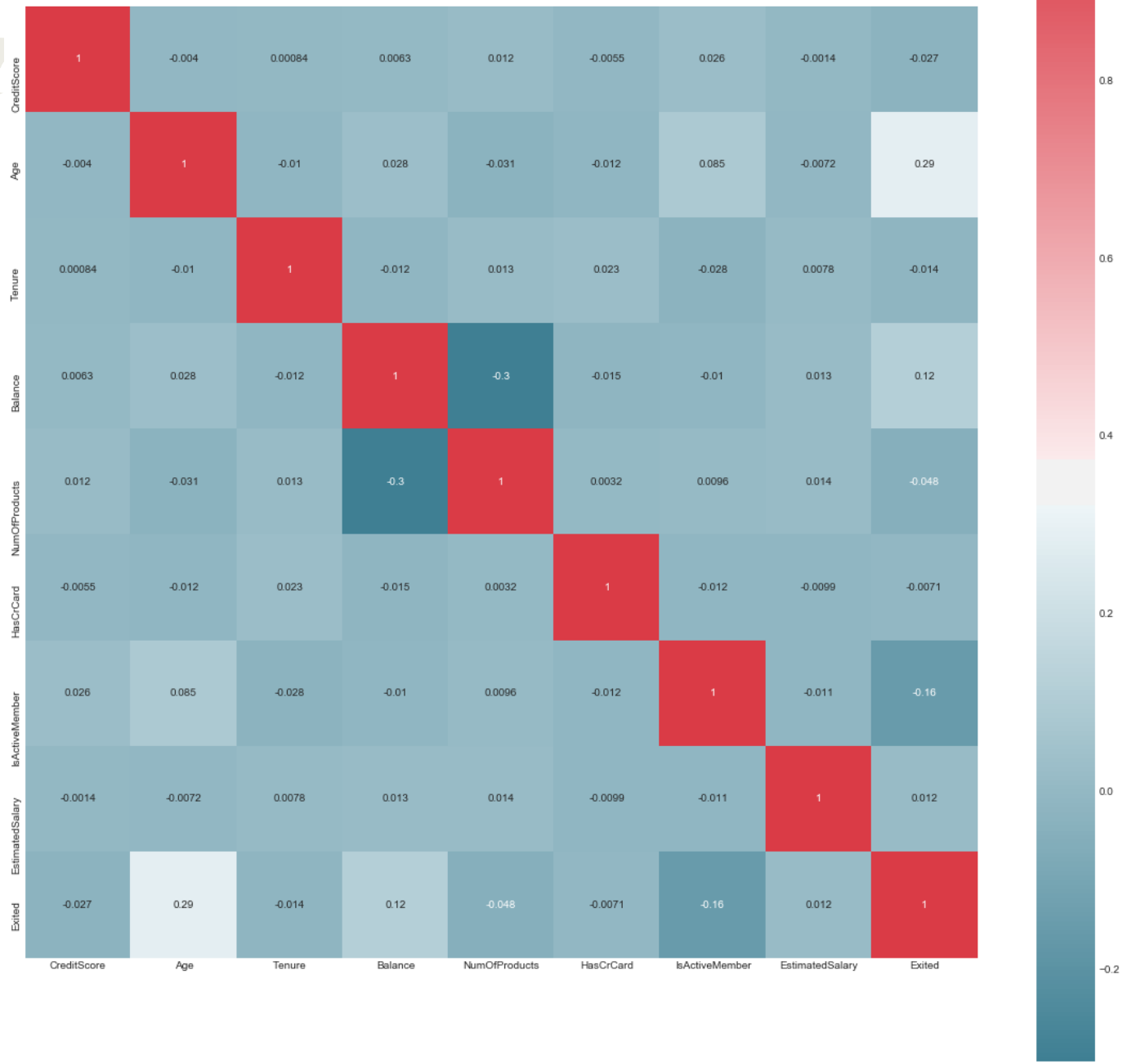

Balance
3000
2000
1000
0
0
50000
100000
150000
200000
250000

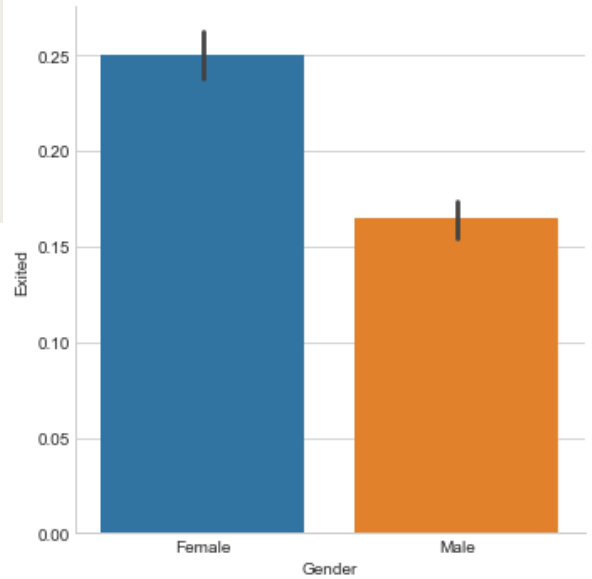

CreditScore
1000
800
600
400
200
0
400
500
600
700
800

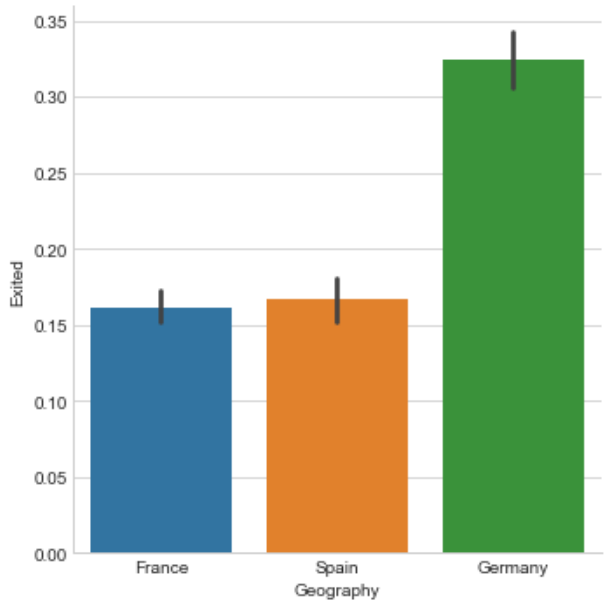

EstimatedSalary
500
400
300
200
100
0
0
25000
50000
75000
100000
125000
150000
175000
200000

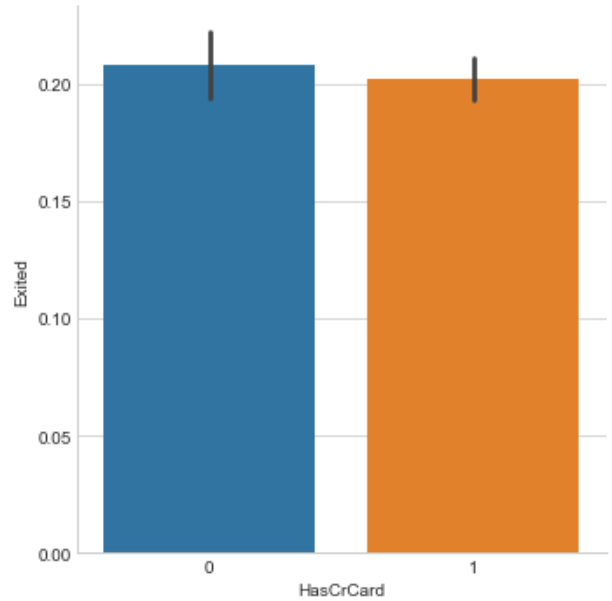

NumOfProducts
5000
4000
3000
2000
1000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0

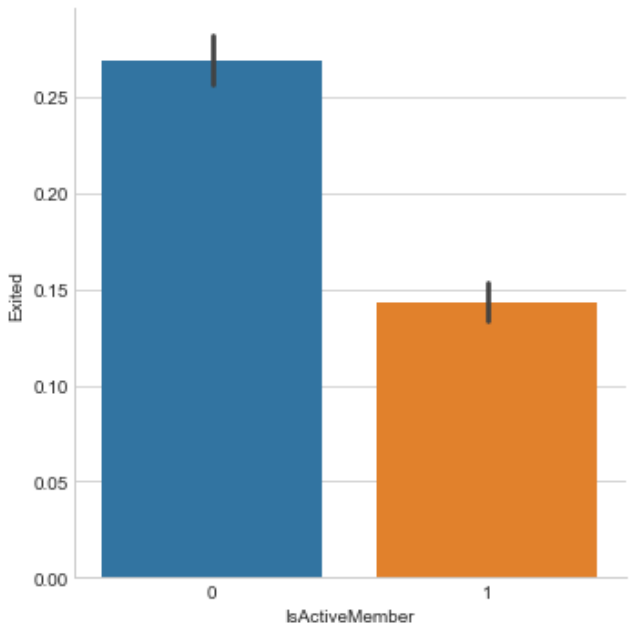

Tenure
1000
800
600
400
200
0
0
2
4
6
8
10

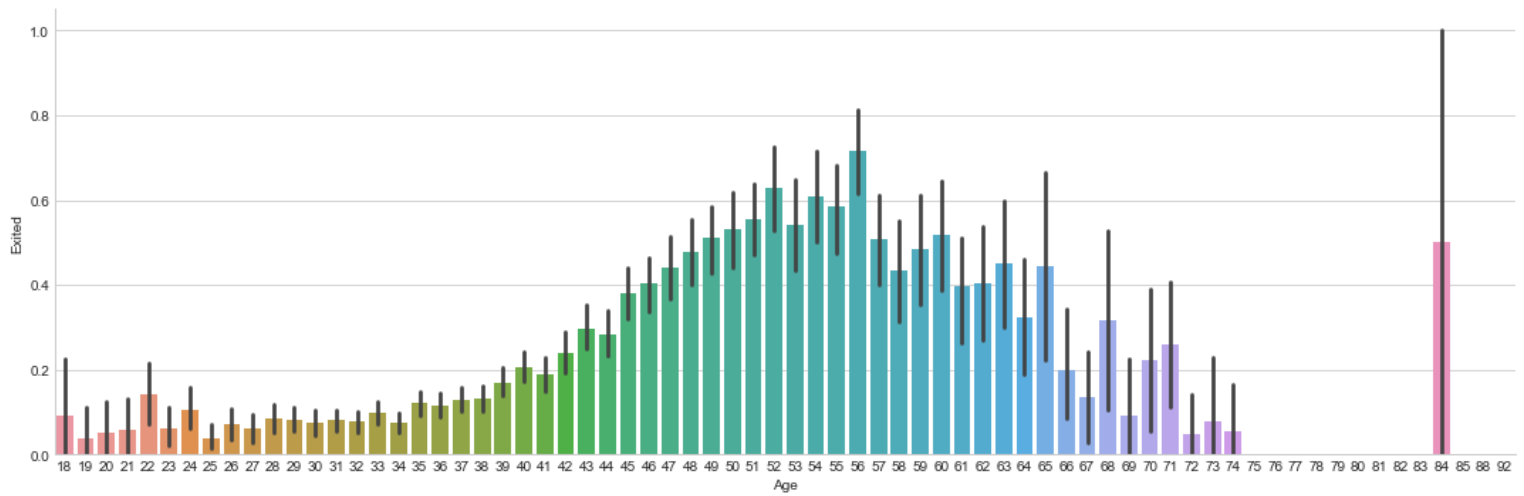

CreditScore
Age
Tenure
Balance
NumOfProducts
HasCrCard
IsActiveMember
EstimatedSalary
Exited
1 -0.004 0.00084 0.0063 0.012 -0.0055 0.026 -0.0014 -0.027
-0.004 1 -0.01 0.028 -0.031 -0.012 0.085 -0.0072 0.29
0.00084 -0.01 1 -0.012 0.013 0.023 -0.028 0.0078 -0.014
0.0063 0.028 -0.012 1 -0.3 -0.015 -0.01 0.013 0.12
0.012 -0.031 0.013 -0.3 1 0.0032 0.0096 0.014 -0.048
-0.0055 -0.012 0.023 -0.015 0.0032 1 -0.012 -0.0099 -0.0071
0.026 0.085 -0.028 -0.01 0.0096 -0.012 1 -0.011 -0.16
-0.0014 -0.0072 0.0078 0.013 0.014 -0.0099 -0.011 1 0.012
-0.027 0.29 -0.014 0.12 -0.048 -0.0071 -0.16 0.012 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

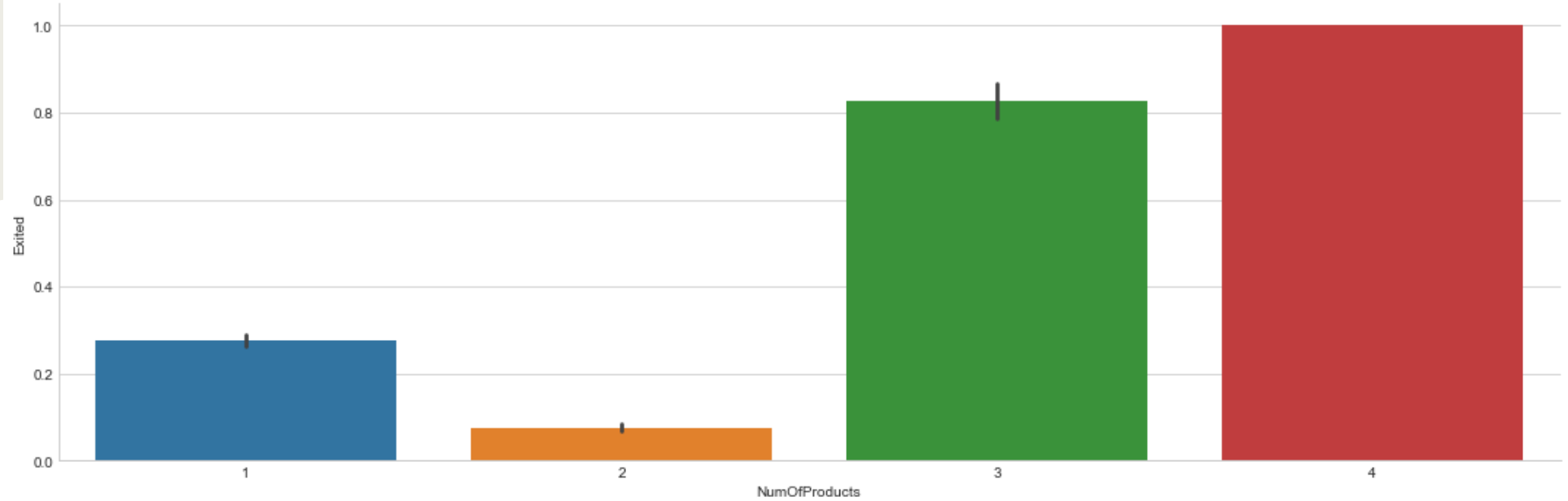

Exited
0.25
0.20
0.15
0.10
0.05
0.00
Female
Male
Gender

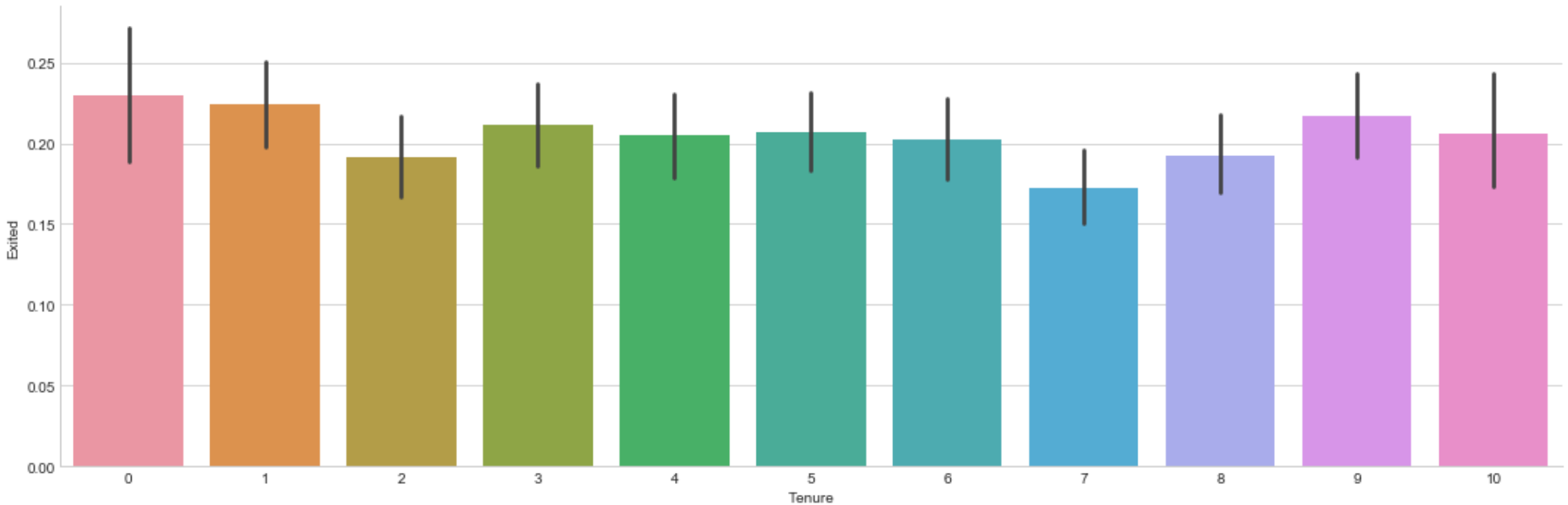

Exited
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
France
Spain
Germany
Geography

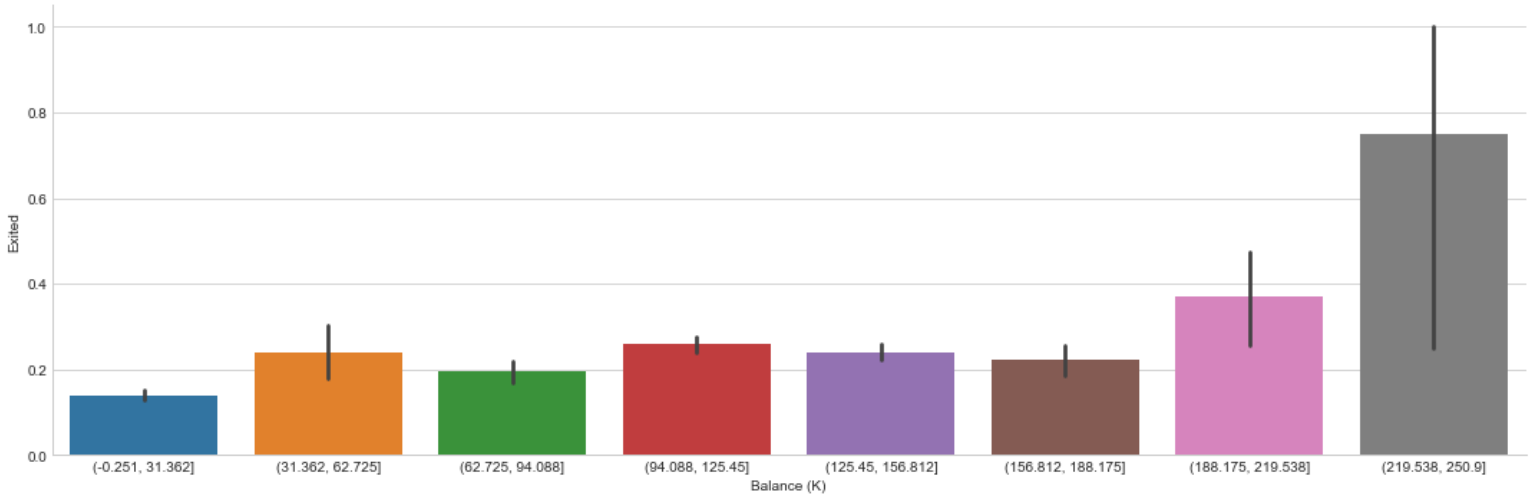

Exited
0.20
0.15
0.10
0.05
0.00
0
1
HasCrCard

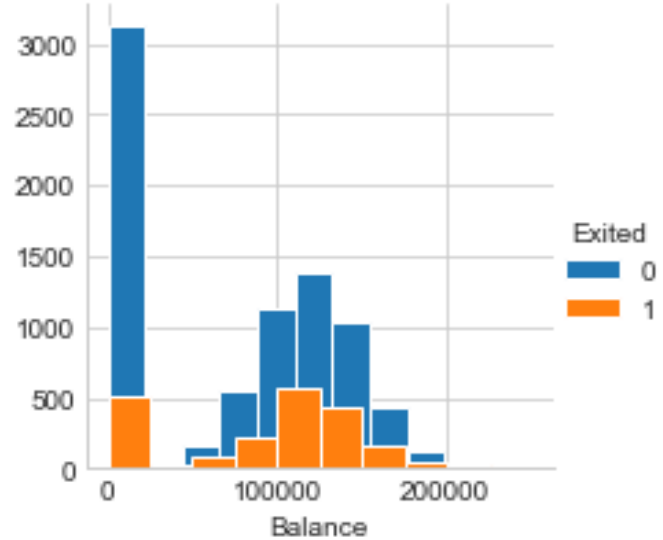

Exited
0.25
0.20
0.15
0.10
0.05
0.00
0
1
IsActiveMember

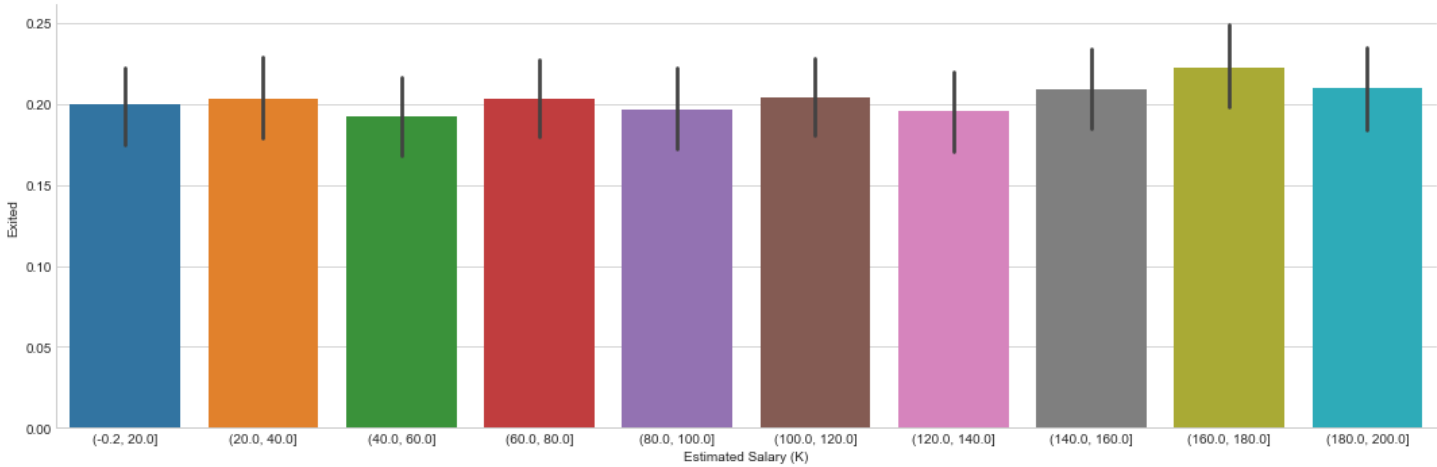

Exited
1.0
0.8
0.6
0.4
0.2
0.0
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 88 92
Age

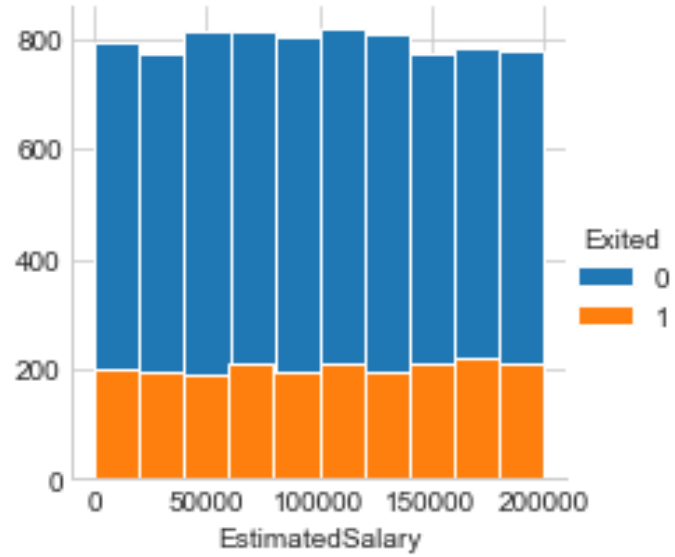

1.0
0.8
0.6
0.4
0.2
0.0
Exited
1
2
3
4
NumOfProducts

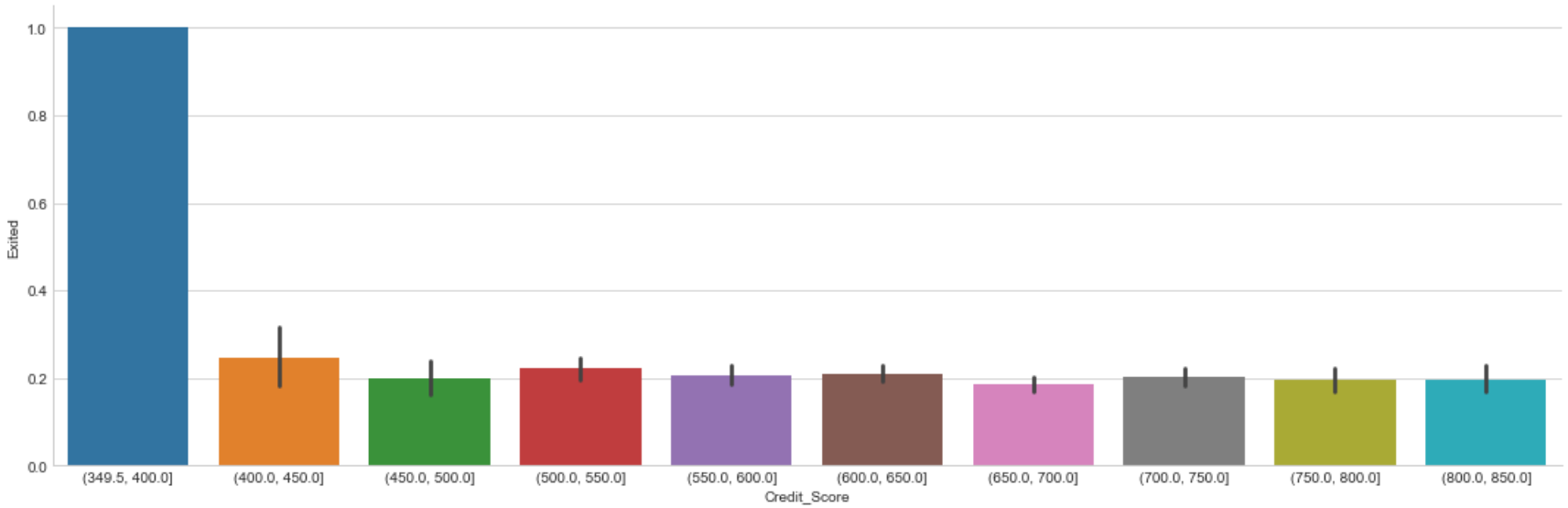

0.25
0.20
0.15
0.10
0.05
0.00
Exited
0
1
2
3
4
5
6
7
8
9
10
Tenure

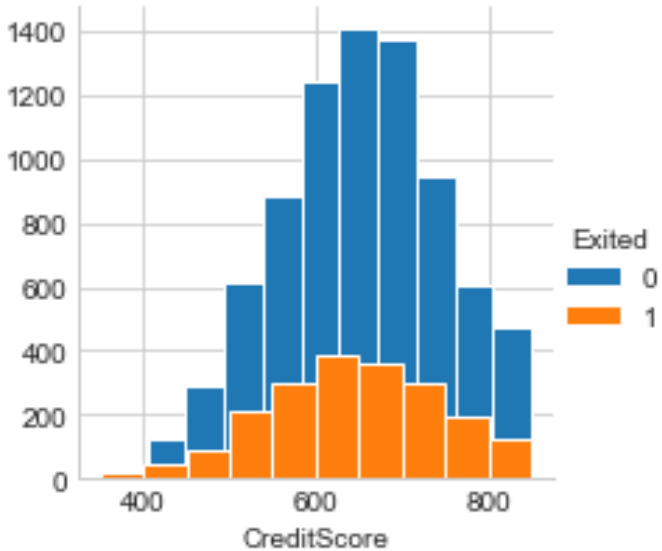

Exited
1.0
0.8
0.6
0.4
0.2
0.0
(-0.251, 31.362]
(31.362, 62.725]
(62.725, 94.088]
(94.088, 125.45]
(125.45, 156.812]
(156.812, 188.175]
(188.175, 219.538]
(219.538, 250.9]
Balance (K)


3000
2500
2000
1500
1000
500
0
0
100000
200000
Balance
Exited
0
1


Exited
0.25
0.20
0.15
0.10
0.05
0.00
(-0.2, 20.0]
(20.0, 40.0]
(40.0, 60.0]
(60.0, 80.0]
(80.0, 100.0]
(100.0, 120.0]
(120.0, 140.0]
(140.0, 160.0]
(160.0, 180.0]
(180.0, 200.0]
Estimated Salary (K)

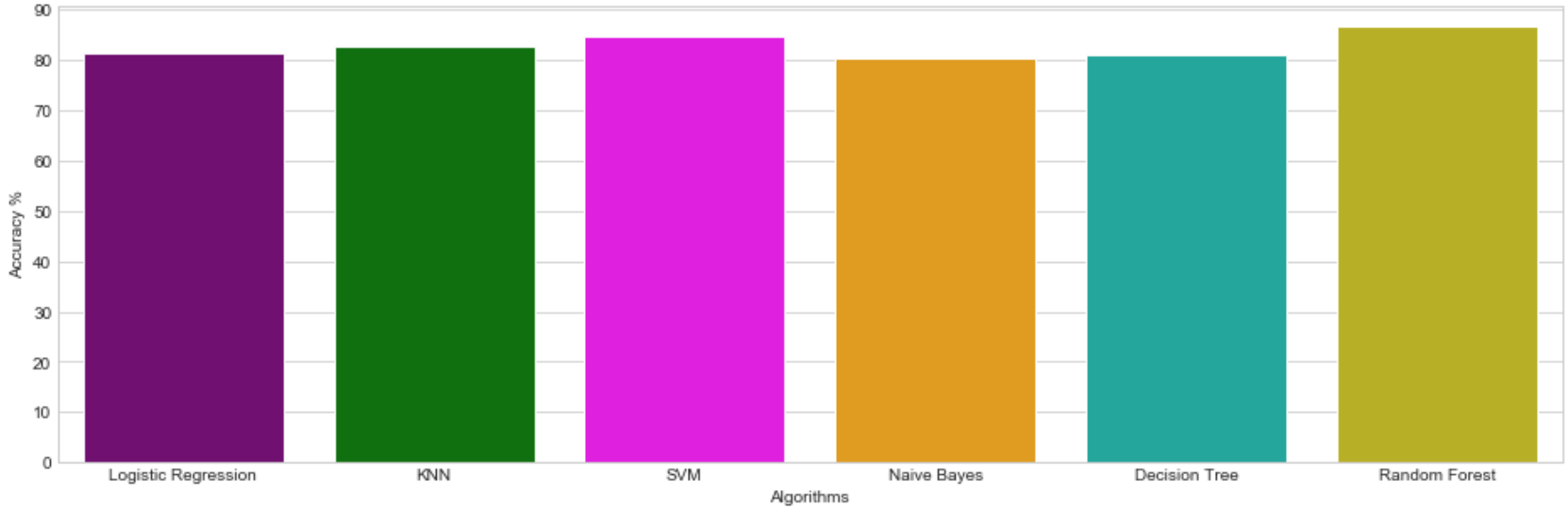

800
600
400
200
0
0
50000
100000
150000
200000
EstimatedSalary
Exited
0
1


1.0
0.8
0.6
0.4
0.2
0.0
Exited
(349.5, 400.0]
(400.0, 450.0]
(450.0, 500.0]
(500.0, 550.0]
(550.0, 600.0]
(600.0, 650.0]
(650.0, 700.0]
(700.0, 750.0]
(750.0, 800.0]
(800.0, 850.0]
Credit_Score

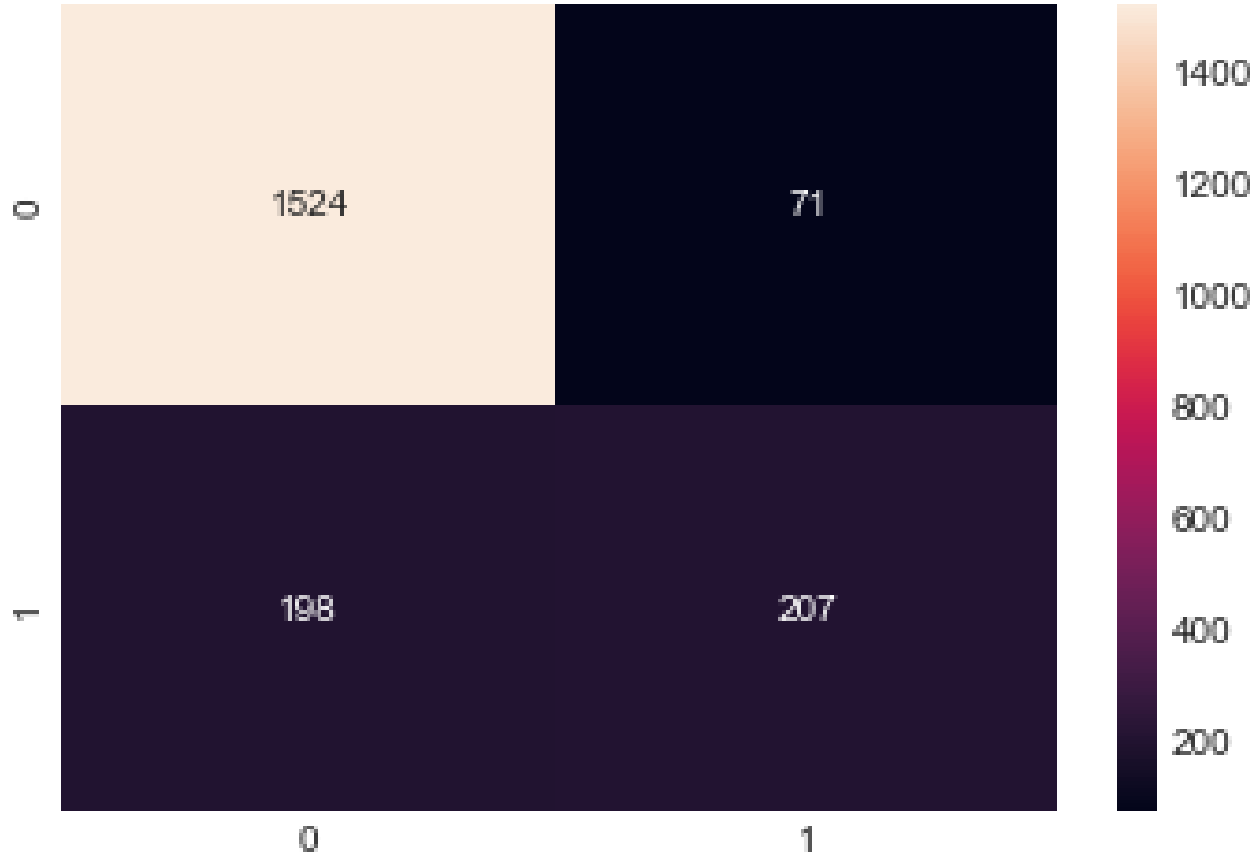

1400
1200
1000
800
600
400
200
0
400
600
800
CreditScore
Exited
0
1

# *Predict with Different Model*

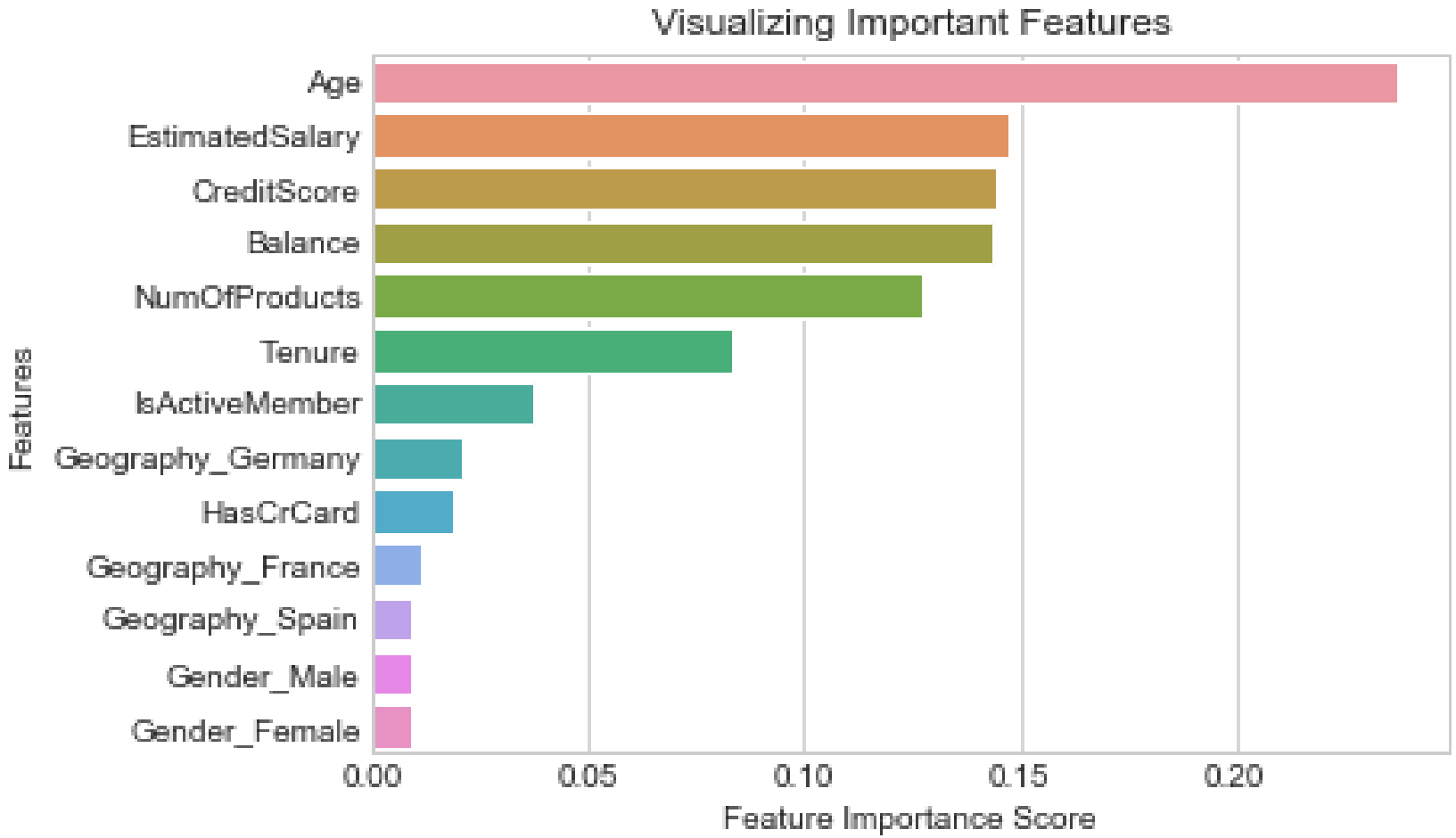
| Model | Accuracy |
|---|---|
| Logistic Regression | 81.20% |
| KNN (k=7) | 82.60% |
| SVM | 84.70% |
| Naive Bayes | 80.35% |
| Decision Tree | 81.15% |
| Random Forest | 86.55% |



# *Random Forest – Confusion Matrix*



# *Random Forest – Important Features*



# *Conclusion*

- limitation: imbalanced dataset
- Who is more likely to leave?
  - Female
  - Germany
  - Not active Member
  - 40 ~ 50 years old
  - who has less products
  - whose balance > 190000