

QUAL O PERFIL DO CLIENTE QUE DEIXA EMPRESA ?

1. Apresentação do problema

O dataset que foi analisado representa um banco de dados de cliente de um banco comercial que atua em três países: França, Alemanha e Espanha.

A ideia desta análise é mapear o perfil dos clientes que saíram da empresa e apresentar estas informações para a equipe.

Para realizar a análise foi utilizada a linguagem de programação Python e as seguintes bibliotecas:

- Pandas
- Seaborn
- Matplotlib

Esta análise consistem em identificar os comportamentos das variáveis categóricas e numéricas, e a partir disso identificar o perfil do cliente que deixa a empresa.

2. Definindo o dataset

De acordo com a pergunta chave, foi estabelecido um filtro no dataset focando a maioria da análise em pessoas que tiveram o Exited = 1, ou seja, que deram Churn no negócio. Para isso foi definida uma variável saída_pessoal que nada mais é do que o dataset filtrado com apenas as pessoas que saíram.

A partir daí foi possível realizar as análises de mapeamento das pessoas que saíram. Além disso, foram removidas as seguintes colunas: RowNumber, CustomerId, Surname. Estas não exercem influência na análise.

Abaixo está a identificação do significado de cada coluna.

CustomerId: identificação do cliente.

Surname: sobrenome do cliente.

CreditScore: pontuação de crédito, 0 alto risco de inadimplência e 1000 clientes com baixo risco de inadimplência.

Geography: país que o serviço é oferecido.

Gender: sexo do cliente.

Age: idade do cliente.

Tenure: um indicativo de estabilidade no emprego, em que 0 significa pouca estabilidade e 10 muita estabilidade.

Balance: saldo da conta corrente.

NumOfProducts: número de produtos bancários adquiridos.

HasCrCard: se tem cartão de credito ou não, (Sim = 1 e Não = 0).

IsActiveMember: se é um cliente com conta ativa, (Ativo = 1).

EstimatedSalary: salário estimado.

Exited: cliente deixou de ser cliente do banco ou não (Churn = 1).

Também foram definidas as variáveis categóricas e numéricas para a análise.

Categóricas: 'Geography', 'Gender', 'HasCrCard', 'IsActiveMember', 'Exited', 'Tenure'.

Numéricas: 'CreditScore', 'Age', 'Balance', 'NumOfProducts', 'EstimatedSalary', 'Exited'

3. Iniciando a análise exploratória

Antes de iniciar a análise, não foi identificado nenhuma valor nulo ou faltante nas linhas e colunas do dataset e todos os tipos de variáveis estão corretas.

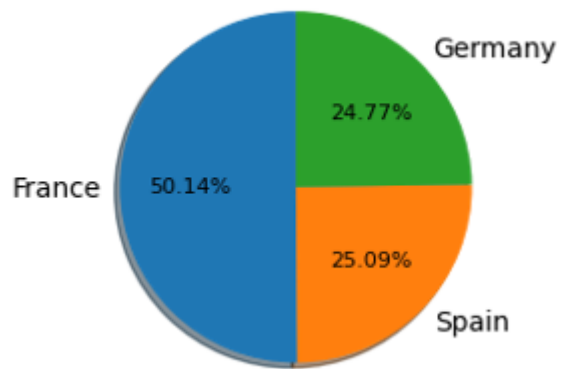
Para entender um pouco da base de dados, foi feita uma análise das métricas estatísticas com intuito de conhecer melhor o dataset, tirando as seguintes conclusões:

- A média de Score das pessoas que saíram é de 645.35
- A pessoa mais nova que saiu possui 18 anos e a mais velha, 84. A média de idade desse grupo é de 45 anos.
- O saldo médio em conta desse grupo é de 91108
- A média salarial das pessoas que saíram é de 101465.58

Com essas visões gerais já se pode especular um pouco sobre o perfil das pessoas que saíram.

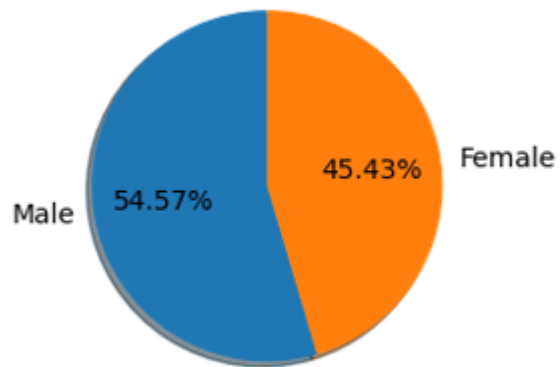
	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	2037.00	2037.00	2037.00	2037.00	2037.00	2037.00	2037.00	2037.00	2037.0
mean	645.35	44.84	4.93	91108.54	1.48	0.70	0.36	101465.68	1.0
std	100.32	9.76	2.94	58360.79	0.80	0.46	0.48	57912.42	0.0
min	350.00	18.00	0.00	0.00	1.00	0.00	0.00	11.58	1.0
25%	578.00	38.00	2.00	38340.02	1.00	0.00	0.00	51907.72	1.0
50%	646.00	45.00	5.00	109349.29	1.00	1.00	0.00	102460.84	1.0
75%	716.00	51.00	8.00	131433.33	2.00	1.00	1.00	152422.91	1.0
max	850.00	84.00	10.00	250898.09	4.00	1.00	1.00	199808.10	1.0

3.1 Primeiras impressões de comportamento das variáveis categóricas



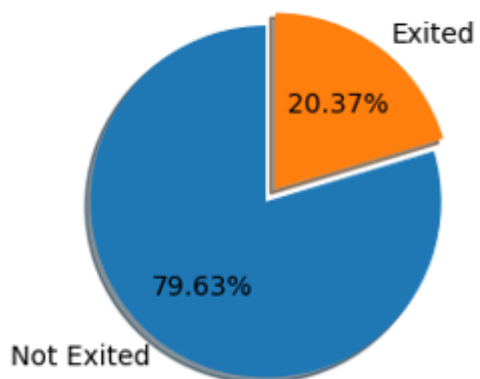
Dos 10000 clientes:

- 5014 são da França.
- 2477 são de Alemanha.
- 2509 são da Espanha.



Dos 10000 clientes:

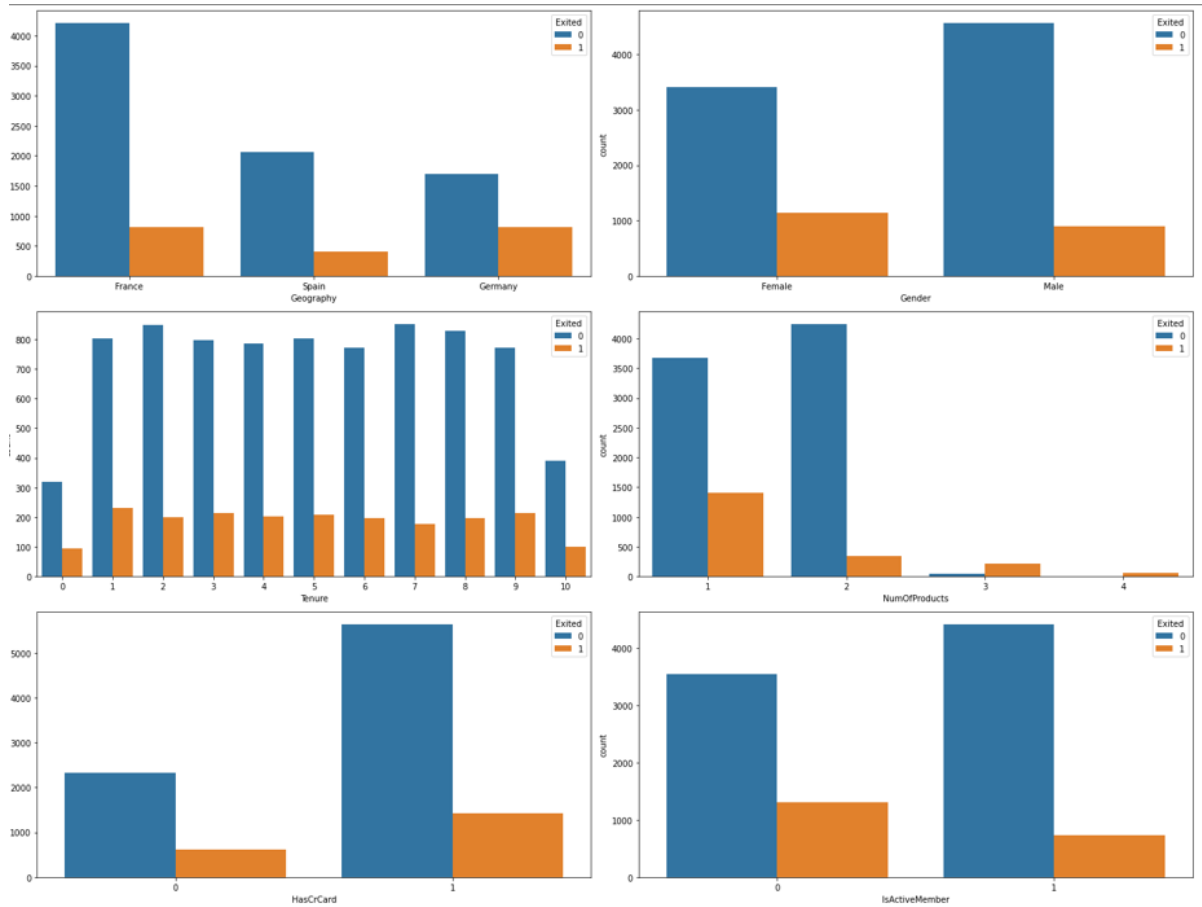
- 5457 são do gênero Masculino.
- 4543 são do gênero Feminino.



Dos 100000 clientes:

- 2037 saíram da empresa.
- 7963 continuam na empresa.

3.2 Visualização geral das categóricas



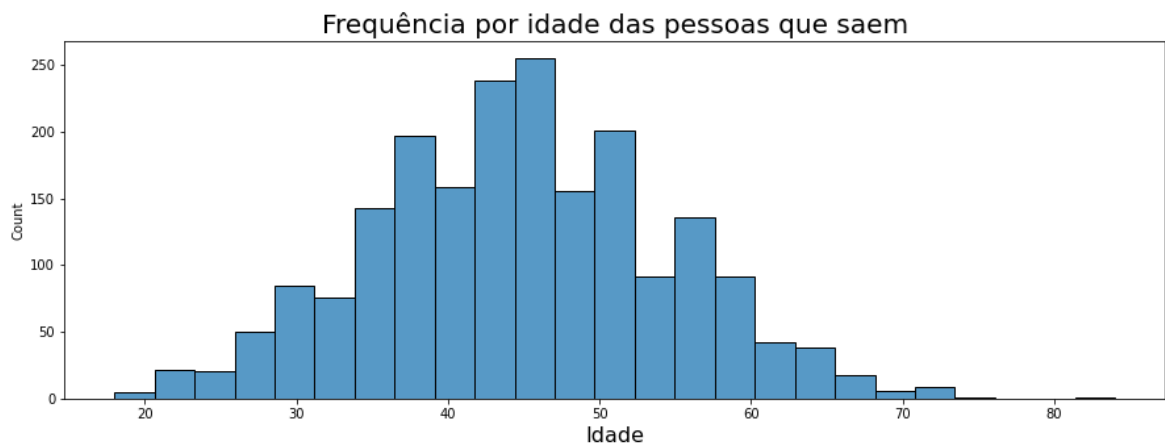
Algumas conclusões dos gráficos acima:

- A maioria dos clientes são da França, e 810 clientes deram Churn. Porém, o país com a maior número de saída de clientes foi a Alemanha com 814 clientes. A proporção de clientes que deram Churn na Alemanha também é a mais alta. O país que menos saiu foi a Espanha com 413.
- A proporção de mulheres que saem é maior do que a de homens.
- Pessoas com tenure entre 1 e 9 foram as que mais saíram da empresa.
- A maioria dos clientes possuem 1 ou 2 produtos, sendo as que possuem apenas um produto deram o maior número de Churn. Pessoas com três produtos deram mais churn em relação a proporção. Vale ressaltar que TODAS as pessoas que possuem 4 produtos deram churn.
- A maioria das pessoas que saíram não possuem cartão de crédito.
- Os membros inativos deram mais churn do que os ativos quem também deram churn. Entretanto, vale ressaltar o alto número de clientes inativos no banco

4. Aprofundando a análise exploratória

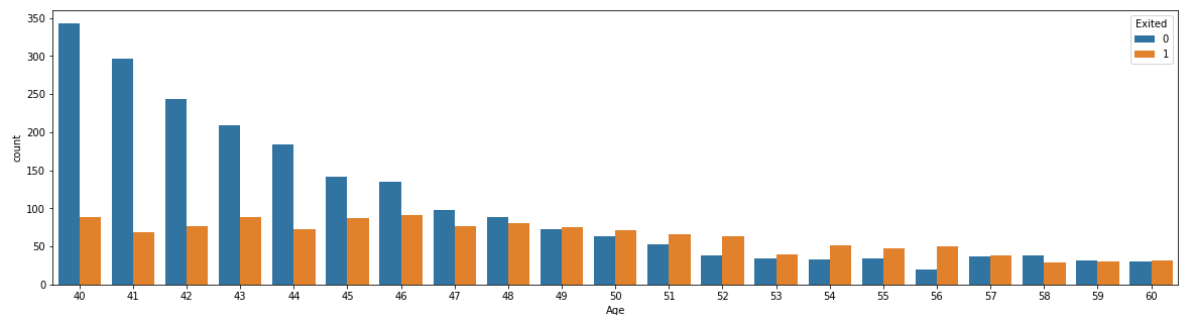
4.1 Observando a variável de idade

Na busca por uma identificação por um perfil de pessoa que, buscou-se entender a frequência de idade dessas pessoas, como é mostrado no histograma abaixo.

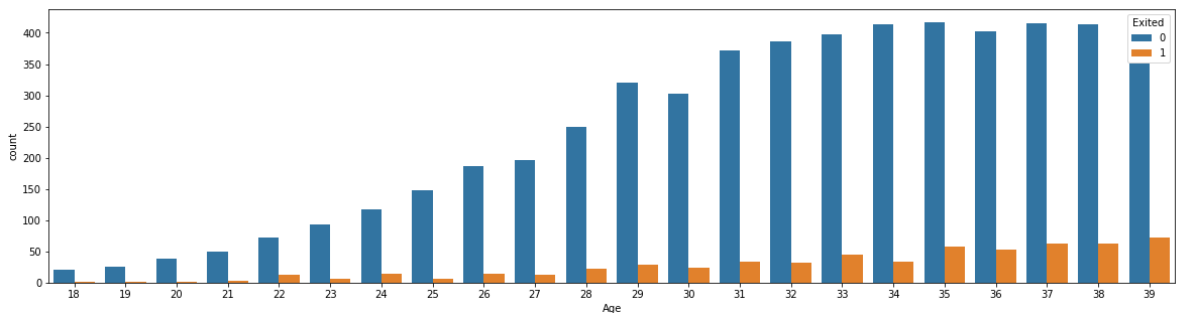


Nota-se que o grupo entre 40 e 55 anos representa as maiores quantidades de pessoas que saíram da empresa. Este grupo possui um total de 1146 pessoas, o que representa 56,25% da quantidade de pessoas que deram Churn no negócio. A proporção é grande se comparada a quantidade total de clientes dessa faixa etária.

De 3213 pessoas nesta faixa etária, 1146 pessoas deram Churn, ou seja 35,66%.

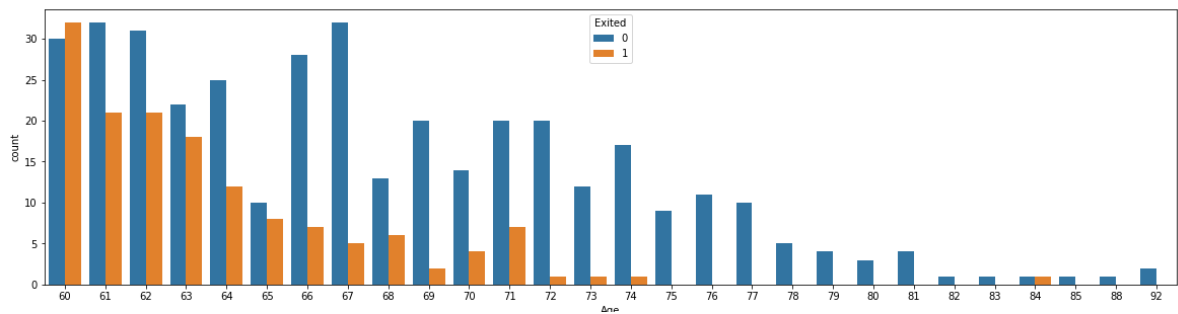


A proporção de saída aumenta conforme a idade, acontecendo mais com pessoas que possuem entre 49 e 56 anos.



Para fins de comparação nota-se uma baixa proporção de Churn entre pessoas de 18 e 40 anos, mostrando que a ampla maioria dessas pessoas continuam na empresa.

Com o público considerado idoso, a proporção dos que saíram permanece alta e tende a cair conforme a idade dos clientes aumentam.



4.2 Observando a variável de Credit Score

A coluna de Credit Score indica o grau de adimplência do cliente, sendo quanto mais alto mais confiável é esse cliente e quantos mais baixo, menos confiável ele é.

A média de score dos grupos que saem e ficam são parecidas:

- Saíram: 645,35
- Ficaram: 651,35

Como mostrado no início o menor score de alguém que saiu é de 350 e o maior é de 850. A partir disso foi criado faixas de score dividos em quartis.

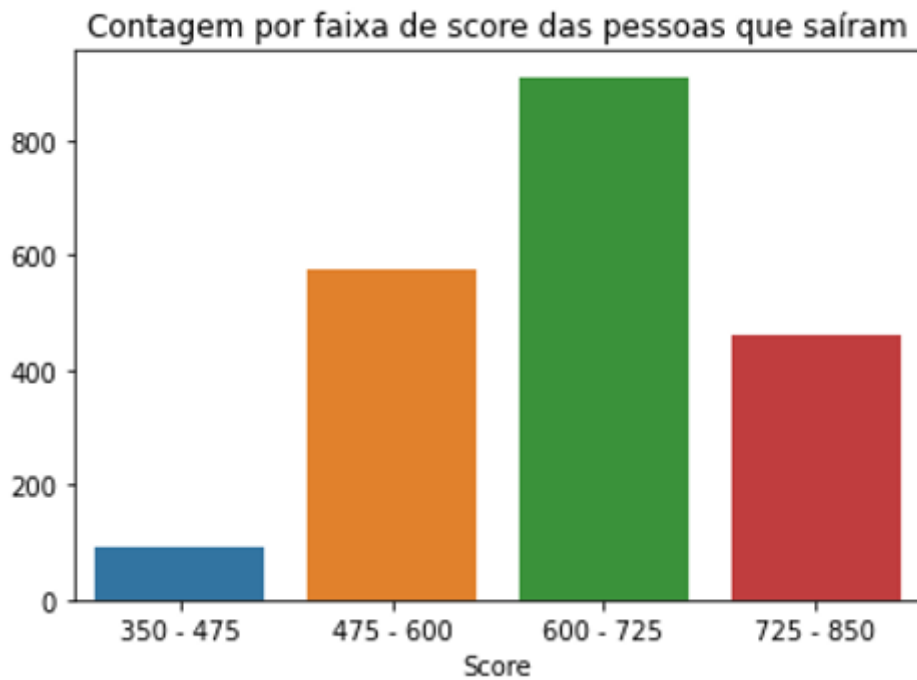
- 350 - 475
- 475 - 600
- 600 - 725
- 725 - 850

44,7% das pessoas que saíram possuem score entre 600-725.

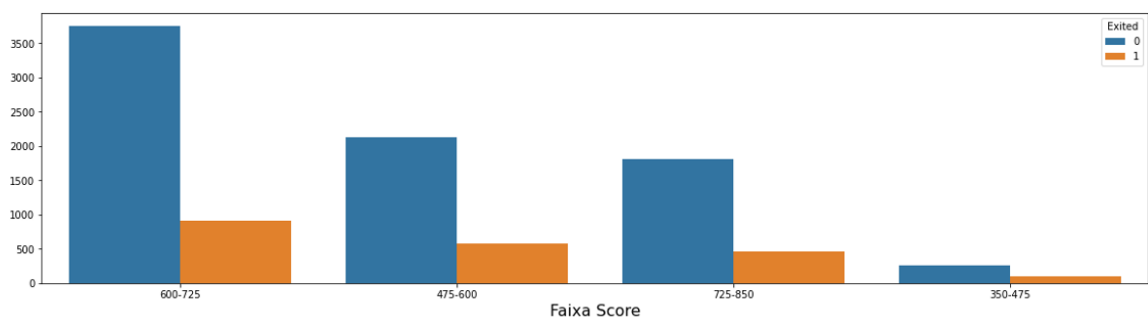
28,1 % das pessoas que saíram possuem score entre 475-600.

22,7 % das pessoas que saíram possuem score entre 725-850.

Apenas 4,6% das pessoas que saíram possuem score entre 350-475.

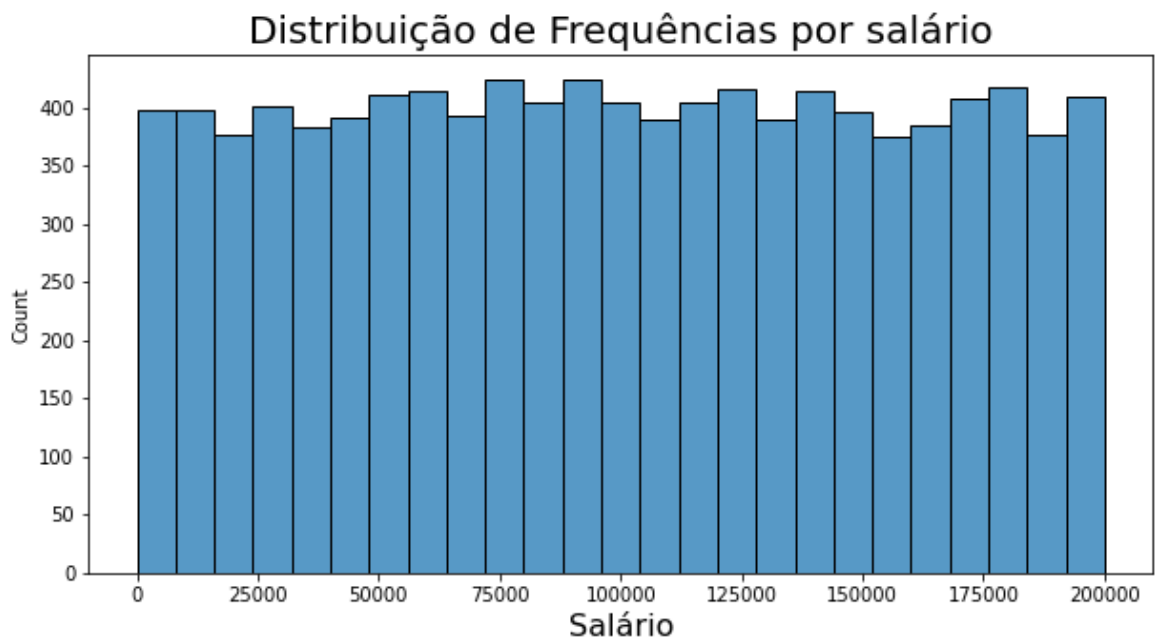


Os números apresentados refletem a proporção de clientes de cada faixa, o que pode ser mostrado abaixo.



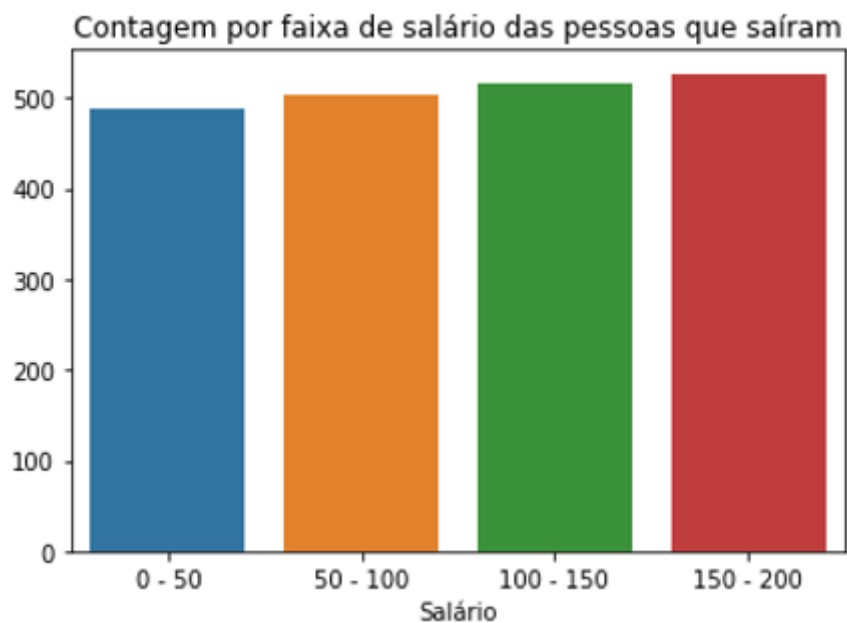
4.3 Observando a variável de salário

A coluna 'EstimatedSalary' representa o salário anual de cada cliente. Para realizar essa análise, a faixa salarial foi dividida em quatro partes iguais, visto que a primeira vista não foi exibida uma variação muito grande no histograma abaixo.



A partir disso foram estabelecidas as seguintes faixas para os clientes:

- De 0 a 50 mil
- De 50 mil a 100 mil
- De 100 mil a 150 mil
- De 150 mil a 200mil



Entre os 2037 que saíram, a distribuição de salário também apresenta uma leve subida junto com o aumento das faixas salariais sendo os seus percentuais definidos entre 24% e 26%. Para uma análise mais profunda dos salários, e encontrar melhores insights este relatório abordará mais na frente a questão de salário de clientes por país.

4.4 Observando a variável de balance(saldo em conta)

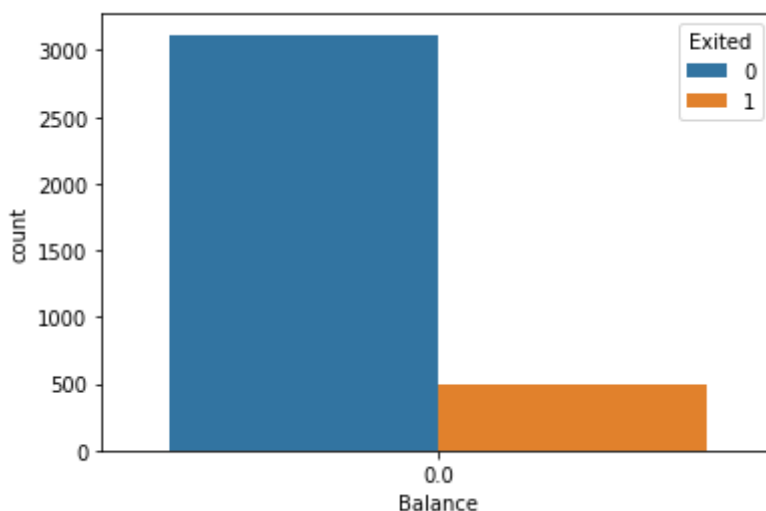
A coluna de balance apresenta o saldo em conta dos clientes, no momento. Para fins de análise, é necessário compreender alguns comportamentos desta variável:

Durante as análises, notou-se um alto número de pessoas SEM SALDO em conta no banco geral de clientes.

Em relação ao total de clientes, 3617 não possuem saldo em conta. Sendo:

- 3117 NÃO DERAM CHURN.
- 500 DERAM CHURN.

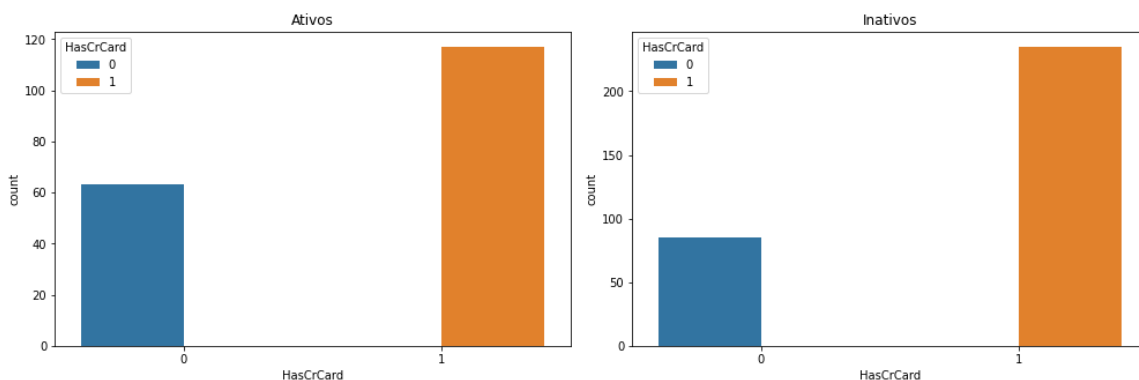
Os números de clientes que continuam no negócio e não possuem saldo em conta, merecem um pouco mais de atenção por partes dos gestores por representar 39% dos remanescentes.



4.4.1 Sem saldo em conta e com cartão de crédito.

Dos 500 clientes que deram Churn e não possuem saldo em conta, 64% são inativos e 36% ativos:

- DOS INATIVOS, 235 de 320 são detentores de cartão de crédito, ou seja 73%.
- DOS ATIVOS, 117 de 180 são detentores de cartão de crédito, ou seja 65%.



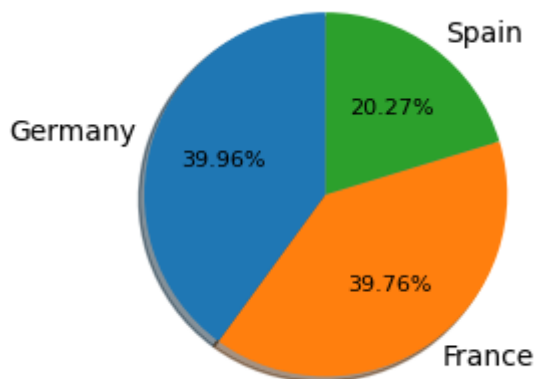
5. Mapeando o perfil dos clientes que saíram

Como abordado no início deste relatório, dos 10000 clientes:

- 5014 são da França.
- 2477 são da Alemanha.
- 2509 são da Espanha

De 2037 clientes que deram churn,temos as seguintes informações:

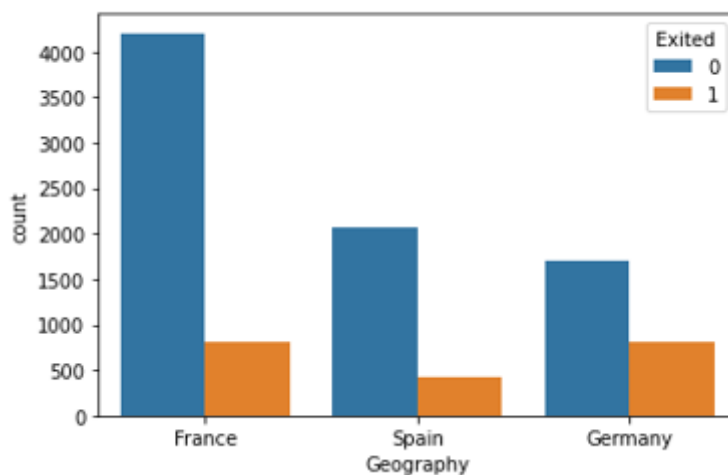
- 814 são da Alemanha
- 810 são da França
- 413 são da Espanha



Para aprofundar um pouco mais a análise, buscou-se identificar a proporção de Churn por país e este foi o resultado:

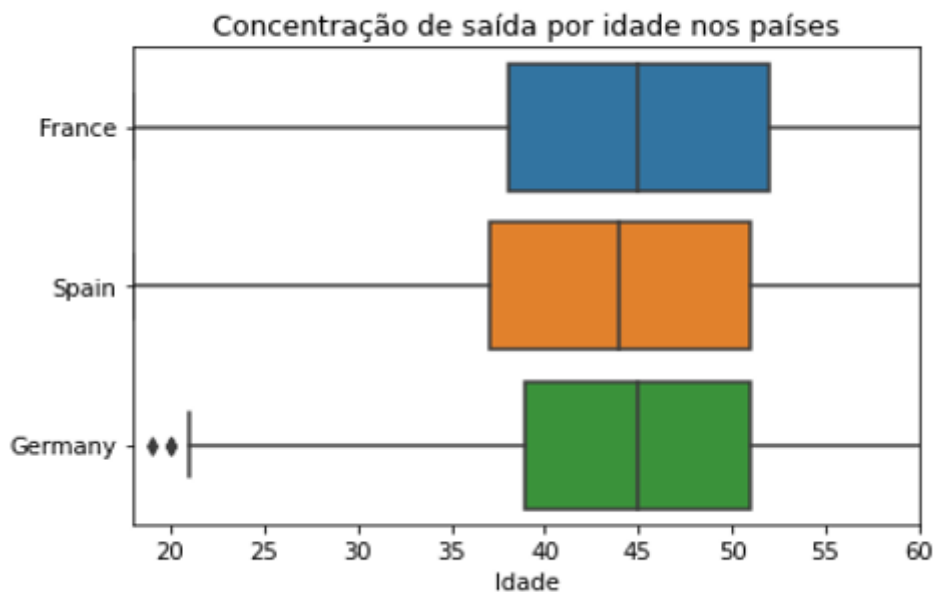
- Alemanha perdeu aproximadamente 33% de seus clientes.
- Espanha perdeu aproximadamente 16,5% de seus clientes
- França perdeu aproximadamente 16% de seus clientes

Com a maior taxa de proporção a alemanha merece uma atenção especial por ter perdido 1/3 de seus clientes.



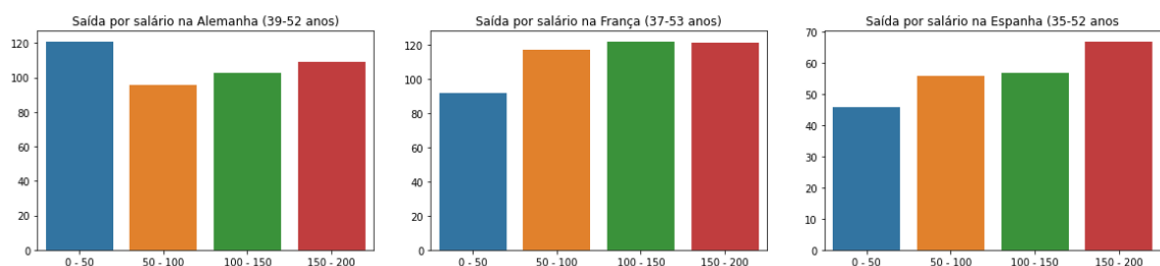
Para traçar um perfil dos clientes que saem por país, a primeira variável analisada foi a de idade e notou-se que:

- Na França, 55,80% das pessoas que deram Churn estão entre 37 e 53 anos. (452)
- Na Espanha, 54,72% das pessoas que deram Churn estão entre 35 e 52 anos. (226)
- Na Alemanha, 52,70% das pessoas que deram Churn estão entre 39 e 52 anos. (429)



A partir disso foi possível analisar a quantidade de clientes desta faixa etária baseado no seu salário anual e que deram churn:

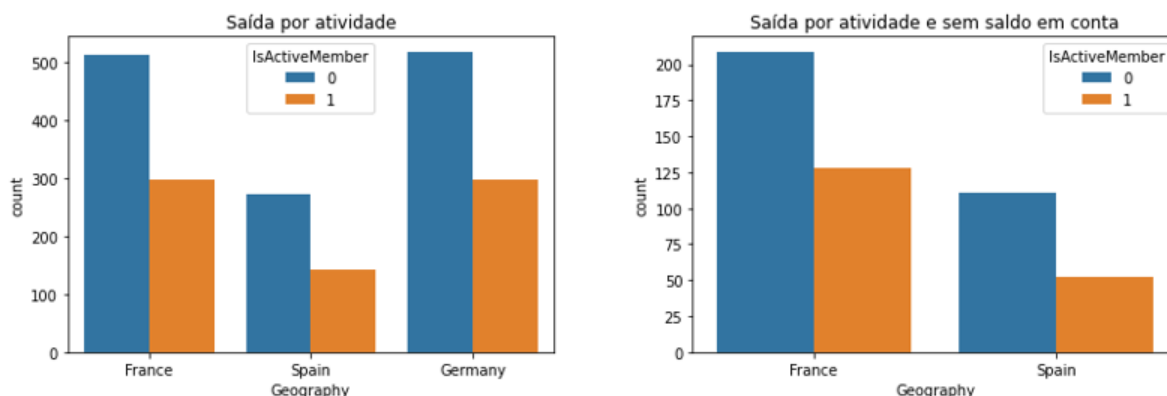
- Na França, os cliente que mais saem com idade entre 37 e 53 anos possuem salário entre 100 e 200 mil anuais, o que representa aproximadamente 54% do grupo.
- Na Espanha, os clientes que mais saem com idade entre 35 e 52 anos possuem salário entre 150 e 200 mil anuais, o que representa aproximadamente 30% do grupo.
- Na Alemanha, os clientes com idade entre 39 e 52 anos e com salário anual entre 0 e 50 mil, representam 28 % do grupo com a determinada faixa etária. Vale ressaltar que a Alemanha é o país que mais perdeu clientes.



O alto número de clientes sem saldo em conta(25%), foi considerada uma variável relevante para esta análise. Por ser uma variável volátil, o foco desta análise está nos clientes que não eram membros ativos e não possuíam saldo em conta, que por lógica esses tendem a dar mais Churn na empresa.

Analisando os gráficos abaixo, conclui-se:

- Apesar de ser o país que mais perdeu clientes, a Alemanha não apresenta nenhum desses clientes com conta zerada.
- Como era de se esperar os clientes que mais saem da empresas não são membros ativos.

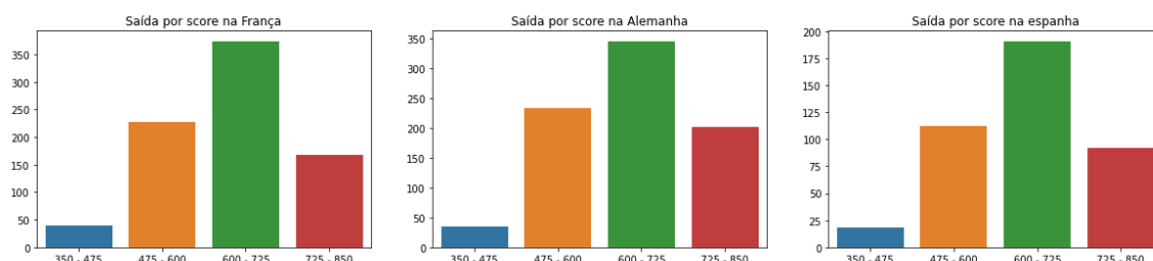


Por fim, uma variável a ser levada em consideração do cliente que sai é o seu Score.

Dos 2037 que deram Churn, 910 possuem score entre 600 e 725, ou seja aproximadamente 45%.

Esta proporção se mantém quando visto especificamente por país, como é mostrado no gráfico abaixo.

Nota-se que na Alemanha muitas pessoas com score excelente(725-850), deram Churn na empresa.



6. Conclusão

Este relatório se aprofundou no banco de dados dos clientes da empresa. Foram avaliadas as principais variáveis como idade,salário,balance,atividade e score para chegar a um mapeamento de perfil. Notou-se que há diferença nas variáveis analisadas por país, e a empresa não deve tratar todos os clientes como um padrão e sim avaliar os comportamento das pessoas em cada região.

Entretanto, analisando maneira mais generalizada notou-se uma grande concentração de Churn no Credit Score entre 600 e 725.

Notou-se também que os clientes com idade superior a 37 anos dão mais churn do que os mais novos. Esse números tende a cair ao chegar nos 60 anos visto que há menos clientes dessa idade em diante.

Além disso, 1/4 dos clientes que deram churn não apresentam saldo em conta, sendo este grupo concentrado na França e Espanha.

O país que merece uma atenção especial é a Alemanha, visto que 1/3 dos clientes deram Churn, mesmo apresentando um alto número de clientes ativos.

Com um olhar crítico para o dataset, algumas variáveis são mais difíceis de compreender como a de Tenure, por não ficar claro quais são suas métricas para atingir aquele valor.