



PKDD'99 DISCOVERY CHALLENGE

PROJETO INFERÊNCIA ESTATÍSTICA

AUTORES

MATHEUS VINICIUS FELIPE VITOR

A55409541

LUCIANA APARECIDA GONÇALVES FABBRIS

A58216033

CAMILA NASCIMENTO DE ALMEIDA

A57729771

SUMÁRIO

1.	Objetivo	5
2.	Métodos utilizados	6
2.1.	Ferramentas e bibliotecas utilizadas.....	6
3.	Análise do Mercado Alvo	7
3.1.	Perfil dos clientes.....	7
3.2.	Avaliação dos dados de cartões de crédito	9
3.3.	Incidência de transações.....	13
3.4.	Frequencia de extrato	14
3.5.	Receita média dos clientes.....	15
3.6.	Empréstimos e negativados	17
4.	Avaliação de concorrência	22
5.	Conclusão do estudo - Plano de ação	24
5.1.	Recomendações.....	26
5.1.1.	Diversificação de clientes	27
5.1.2.	Melhoria no algoritmos de risco de empréstimo	27
5.1.3.	Reorganização da segmentação de cartões de créditos	28
5.1.4.	Expansão de títulos.....	29
5.1.5.	Aumento de ofertas de Crédito	30

ÍNDICE DE FIGURAS

Figura 1 - Contas abertas por ano.....	7
Figura 2 - Emissão de Cartões por ano.....	7
Figura 3 - Contas abertas por Região	8
Figura 4 - Treemap contas por Distrito.....	8
Figura 5 - Número de Cartões por gênero.....	9
Figura 6 - Donut Cartões por Sexo	10
Figura 7 - Cartões por tipo.....	10
Figura 8 - Cartões Emitidos Dados Gerais.....	11
Figura 9 - BoxPlot tipo de cartões por receita mensal	12
Figura 10 - Total de transações por tipo (Débitos).....	13
Figura 11 - Total de transações por tipo Créditos	13
Figura 12 - Transações por Instituição.....	14
Figura 13 - Frequência de extrato	14
Figura 14 - Boxplot Receita média por região.....	15
Figura 15 - Receita Média por gênero	16
Figura 16 - Empréstimos e negativados	17
Figura 17 - Proporção de contas Negativas	18
Figura 18 - Negativados por tipo de cartão e sexo.....	18
Figura 19 - Negativados receita mensal e idade	19
Figura 20 - Status dos Empréstimos ativos.....	19
Figura 21 - Status por porcentagem de parcela.....	20
Figura 22 - Status, Porcentagem de Parcela, Idade e Amount.....	20
Figura 23 - BoxPlot status e amout	21
Figura 24 - Número de transferências para outros bancos	22
Figura 25 - Valores de transferências para outros bancos.....	22

Figura 26 - Valor de Empréstitmo por status	24
Figura 27 - Matriz de correlação de variáveis	26
Figura 28 - Quantidade de Clientes por Região	27
Figura 29 - Status de Pagamento de Empréstimo por Região	27
Figura 30 - Matriz de Correlação.....	28
Figura 31 - Boxplot tipo de cartão e receita mensal	29
Figura 32 - Títulos pagos em bancos	29
Figura 33 - Proporção de contas negativadas	30

1. OBJETIVO

Este estudo teve por objetivo analisar as características dos clientes e dos produtos do Banco Czech de maneira a identificar oportunidades de melhoria na oferta dos produtos

O estudo visa utilizar modelos estatísticos para construção de projeções de acordo com o perfil dos clientes do banco identificando tendências de consumo para a consolidação de um gerenciamento assertivo que possibilite explorar o potencial lucrativo de cada produto. Bem como propor alternativas para minimizar o risco de perda.

2. MÉTODOS UTILIZADOS

Para as inferências utilizamos a linguagem R e modelos de regressão linear. Realizamos a análise exploratória das bases disponibilizadas para a geração dos insights. Assumimos a data 01/01/1999 para iniciarmos os estudos propostos.

2.1. FERRAMENTAS E BIBLIOTECAS UTILIZADAS

Para execução do estudo utilizamos o R-studio e Jupyter notebook para compilar os scripts gerados na linguagem R além dos pacotes :

- Tidyr
- Dplyr
- corrplot
- lubridate
- ggplot
- treemap

3. ANÁLISE DO MERCADO ALVO

3.1. PERFIL DOS CLIENTES

De 1993 a 1998 houve um considerável aumento de contas e cartões, alcançando 4500 contas. Com isso, a análise foi iniciada para traçar o perfil dos clientes do Banco Czech. E foi identificado que:

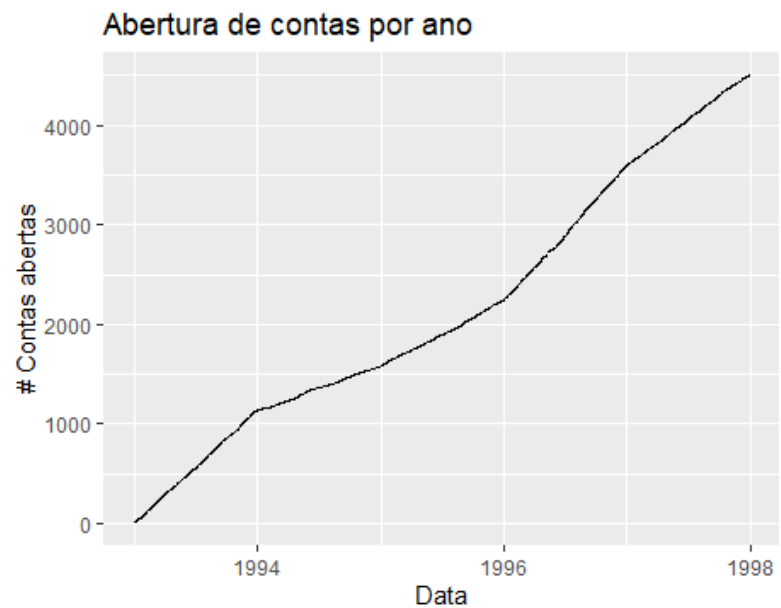


Figura 1 - Contas abertas por ano

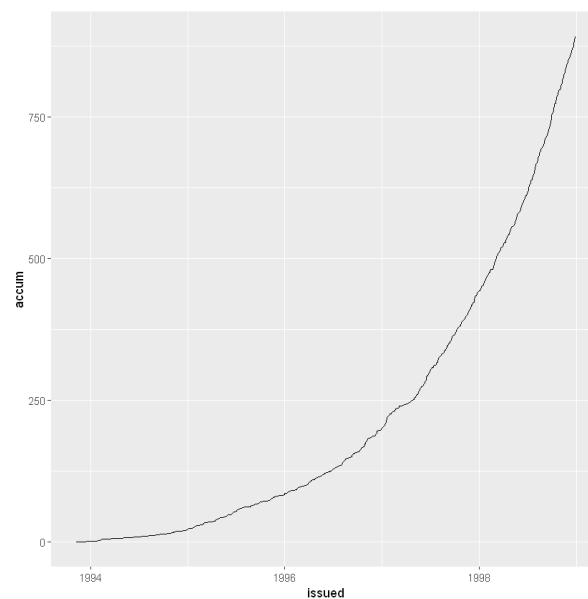


Figura 2 - Emissão de Cartões por ano

- Notável igualdade de clientes em relação ao gênero.

- Um considerável equilíbrio entre a distribuição de contas e cartões por região.

Contas Abertas por Região

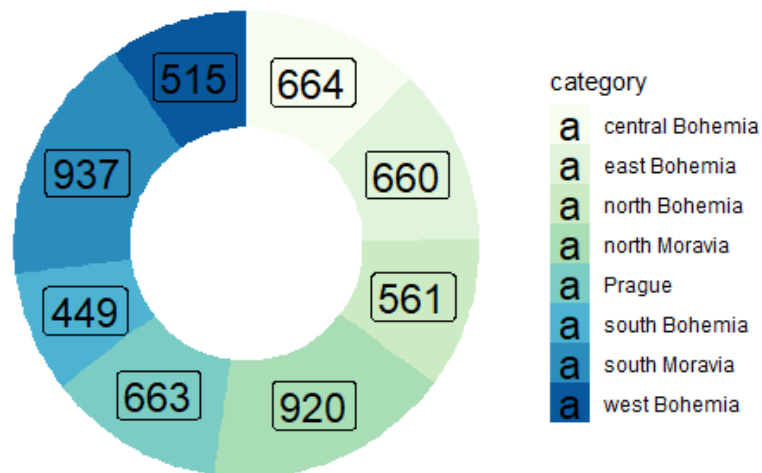


Figura 3 - Contas abertas por Região

Accounts per District

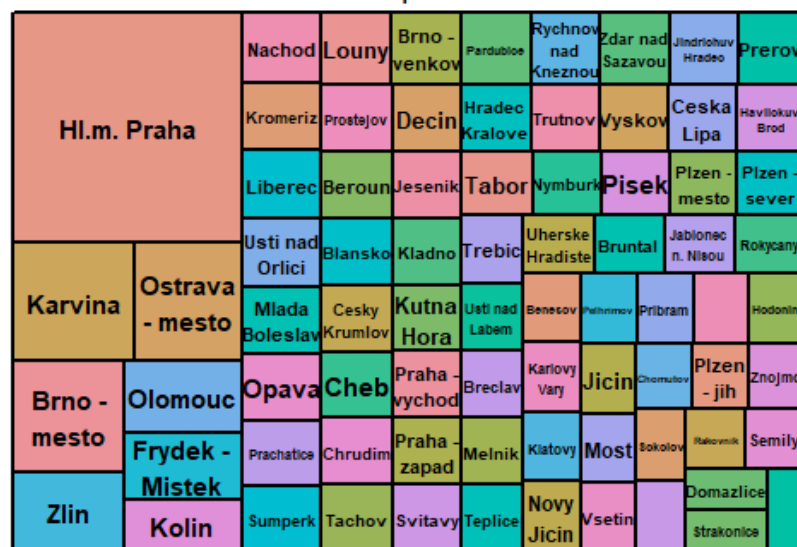


Figura 4 - Treemap contas por Distrito

- O distrito Hl.m.Praha se destacou por ter o maior número de contas.

O que faz sentido, pelo fato de ser o distrito mais industrializado e comercial da República Checa, onde se tem um maior número de indústrias e é além de capital é o centro financeiro da República Checa.

3.2. AVALIAÇÃO DOS DADOS DE CARTÕES DE CRÉDITO

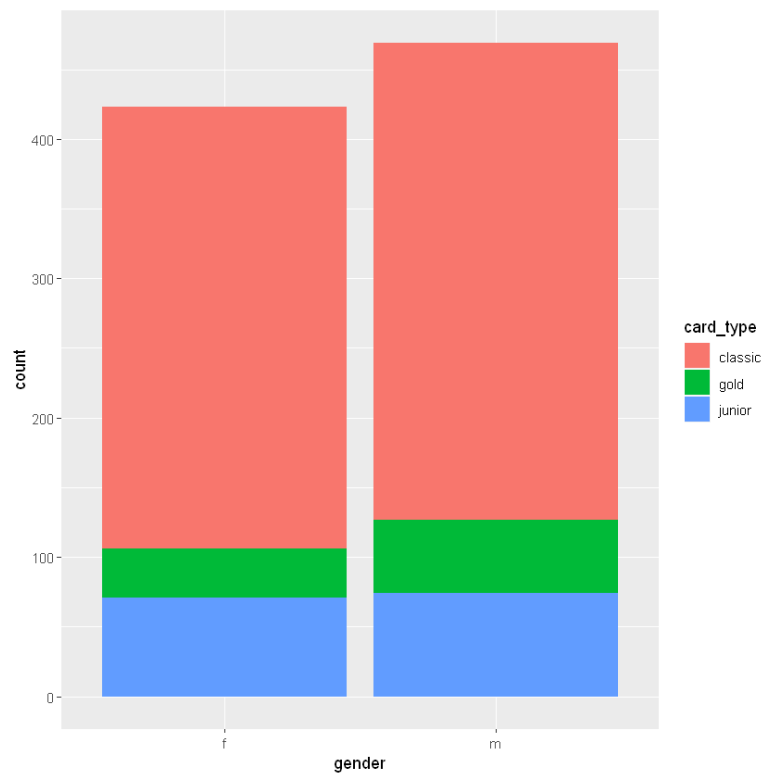


Figura 5 - Número de Cartões por gênero

- Ao avaliar os dados de emissão de cartões de crédito podemos perceber que a grande maioria deles são do tipo “Classic”, e existe um número maior de cartão para pessoas do sexo masculino, como fica comprovado no gráfico abaixo.

Total de cartões de crédito por sexo

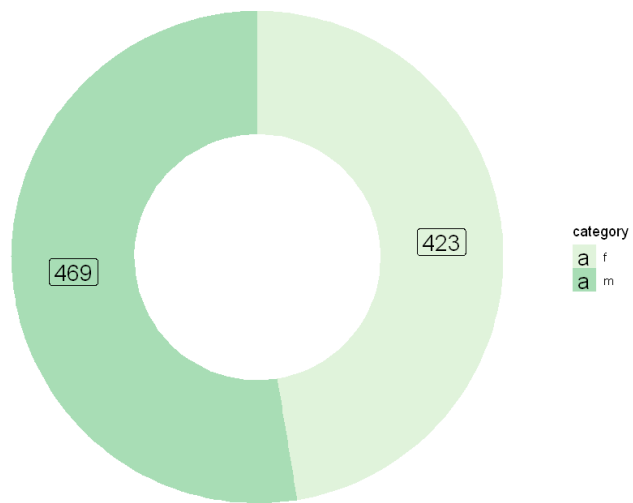


Figura 6 - Donut Cartões por Sexo

Total de Cartões por Tipo

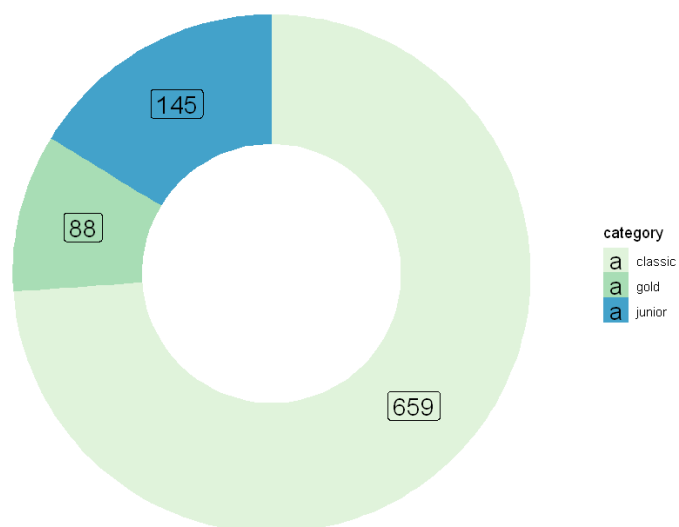


Figura 7 - Cartões por tipo

Como o banco tem mais clientes homens que mulheres é natural que mais homens tenham cartões de crédito apesar disso a proporção de cartões emitidos é :

- 17,2% dos homens que possuem conta, possuem cartão de crédito do banco

- 18,0% das mulheres que possuem conta possuem cartão de crédito.

Portanto, apesar do maior número de cartões serem de homens, as mulheres apresentam leve tendência a possuir mais cartões que os homens, ainda que o número de cartões emitidos a mulheres seja menor que os emitidos aos homens.

Ao avaliar a relação entre receita mensal, gênero, avaliarmos os dados de região e tipo de cartão temos:

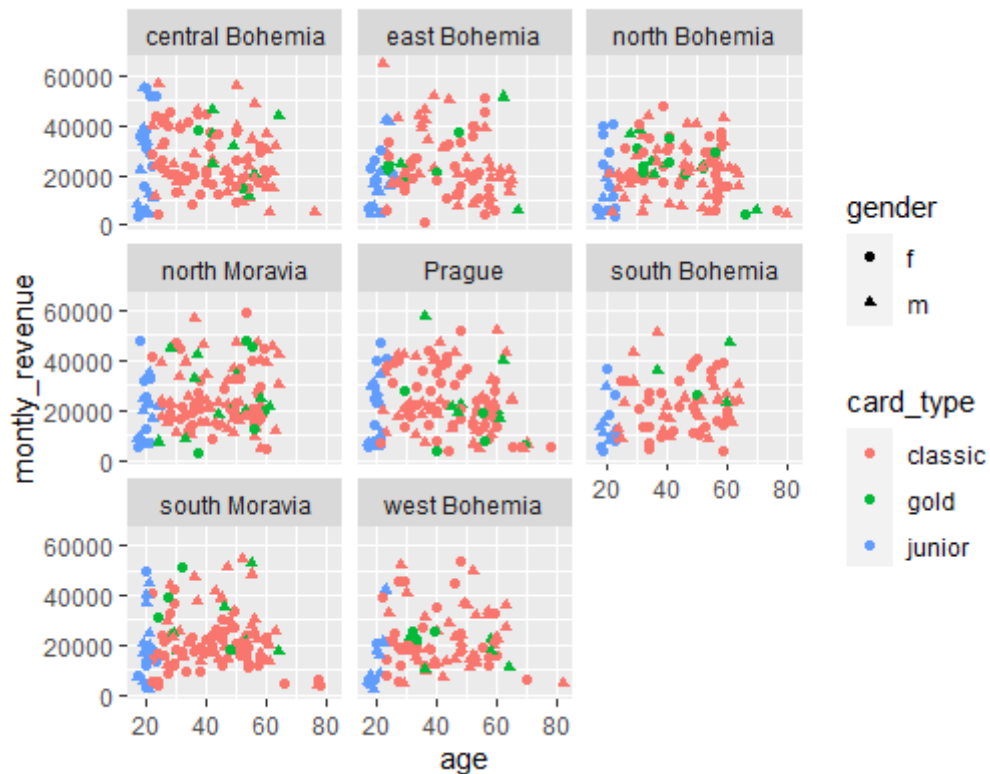


Figura 8 - Cartões Emitidos Dados Gerais

- A maior parcela dos nossos clientes possui o cartão classic.
- Não há diferenciação em relação ao saldo final em Jan/99 e a renda mensal estimada com relação ao tipo de cartão.
- Em específico para o cartão junior o público alvo se limita a variação da idade entre 20-25.
- Muitos clientes tipo clássico tem valores maiores do que o “Gold”.
- A média das receitas do “Gold” não é tão maior do que o “Classic”.

Com isso é possível perceber uma possibilidade de melhor segmentação do tipo de cartão uma vez que conseguimos saber os clientes com maior poder aquisitivo e assim oferecer serviços customizados “premium” como por exemplo um upgrade no tipo de cartão.

Reforçando o posicionamento exposto acima temos o boxplot a seguir.

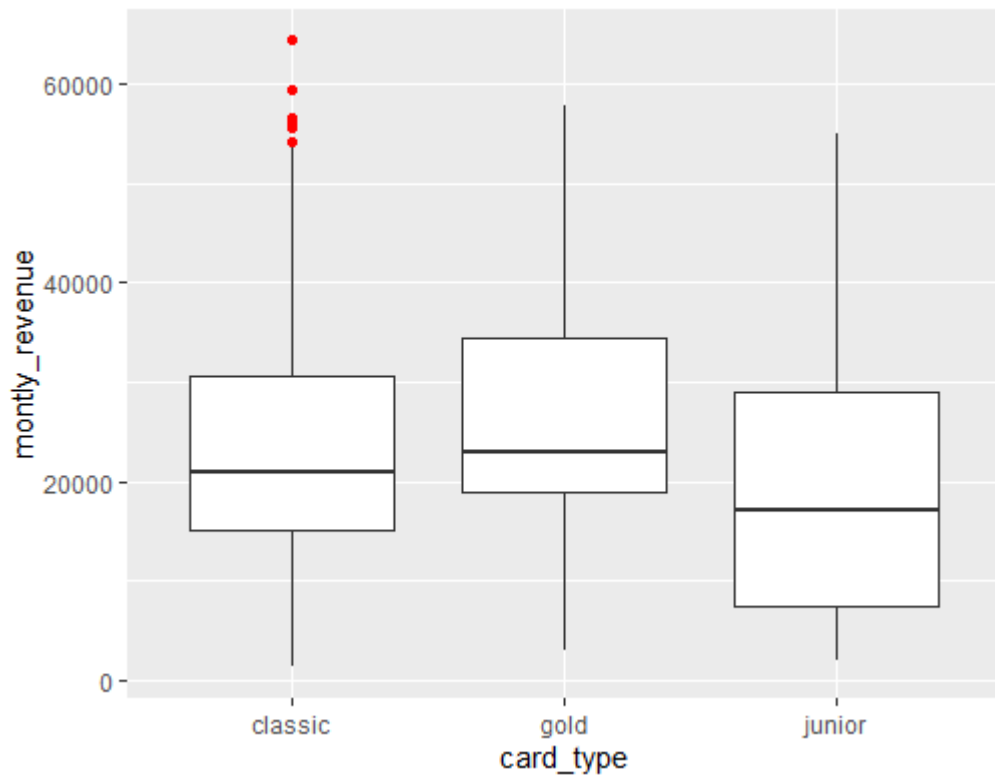


Figura 9 - BoxPlot tipo de cartões por receita mensal

Nesse gráfico fica claro que existem diversos outliers com cartão classic no que diz respeito a sua renda mensal, ou seja, são clientes que possuem uma alta renda e poderia ter um upgrade no tipo de cartão.

3.3. INCIDÊNCIA DE TRANSAÇÕES

- A maior parte das transações de débito e crédito não foram identificadas (NA).

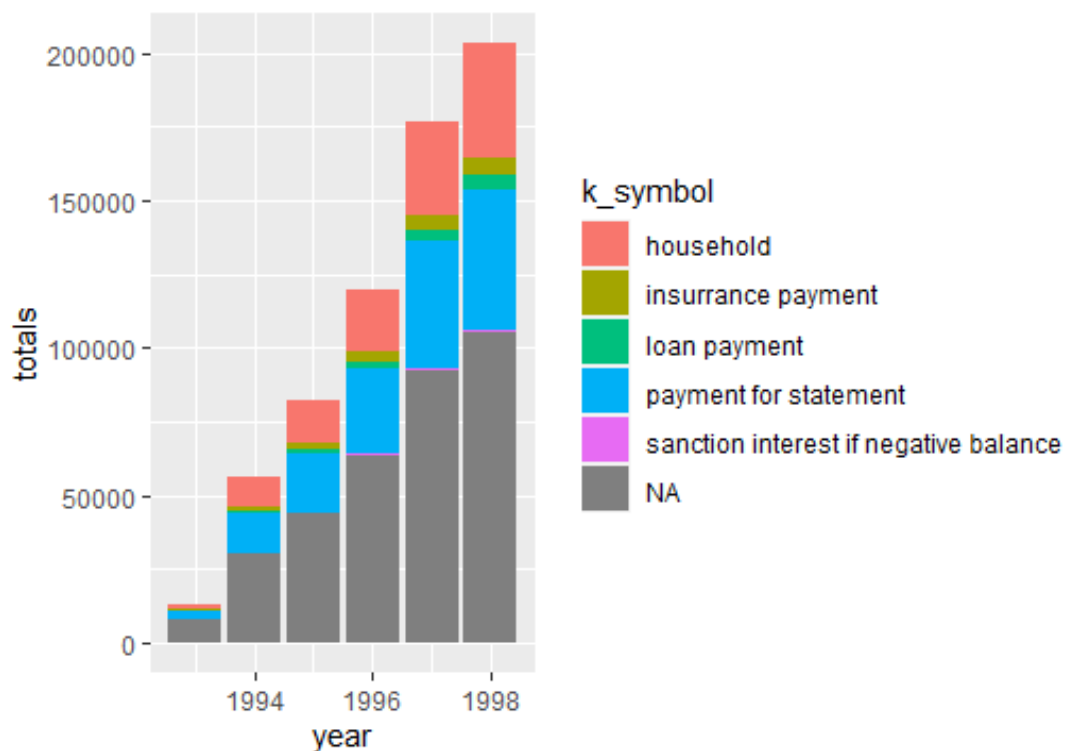


Figura 10 - Total de transações por tipo (Débitos)

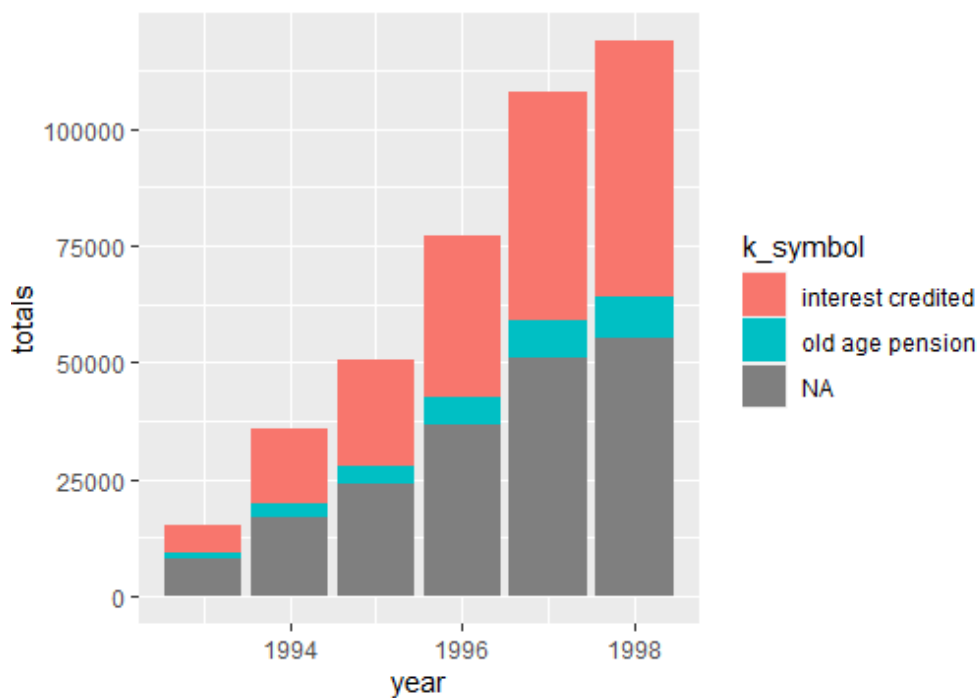


Figura 11 - Total de transações por tipo Créditos

Porém ao retirarmos os valores NA dos cálculos temos o seguinte resultado:

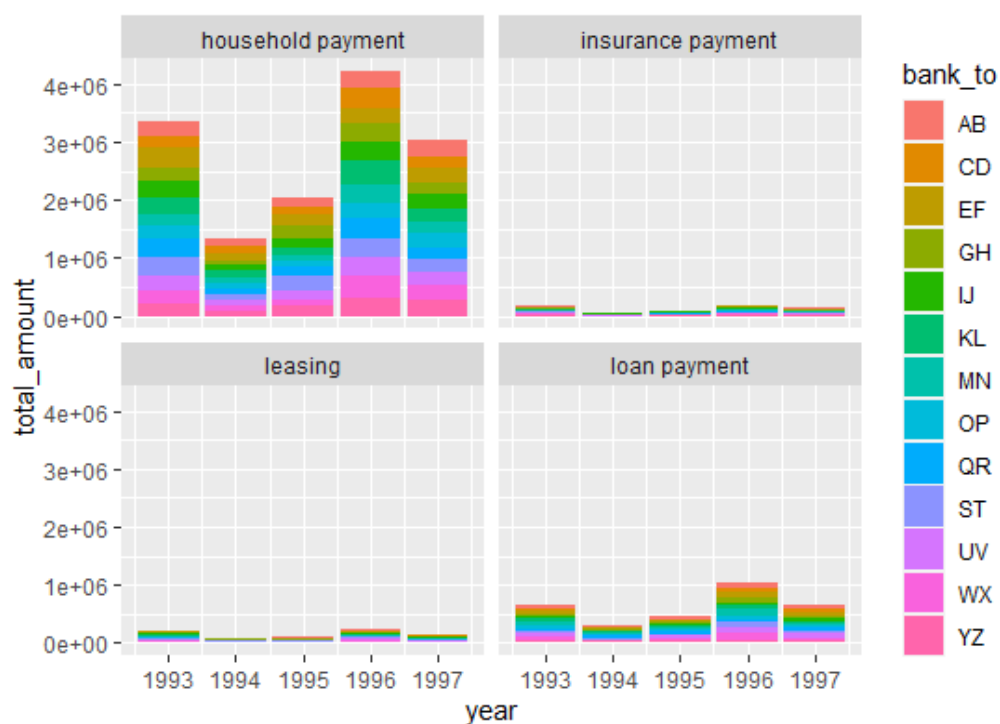


Figura 12 - Transações por Instituição

- Household é o tipo de transação credora com maior incidência das contas ativas.

3.4. FREQUENCIA DE EXTRATO

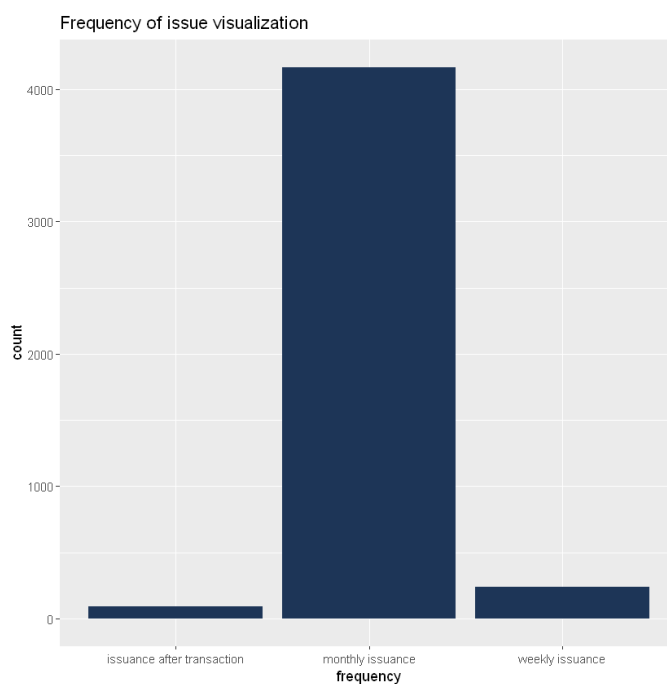


Figura 13 - Frequência de extrato

- 93% dos clientes possuem o hábito de retirar o extrato numa frequência mensal.

Podemos perceber que grande parte dos clientes retiram extratos mensais, isso pode estar atrelado ao fato de na década de 90 a existência de um custo para impressão do extrato, limitava as pessoas na visualização do seu saldo.

3.5. RECEITA MÉDIA DOS CLIENTES

Avaliando os montantes de valor creditados na história da conta foi possível avaliar as receitas médias de cada um dos correntistas do banco Czech.

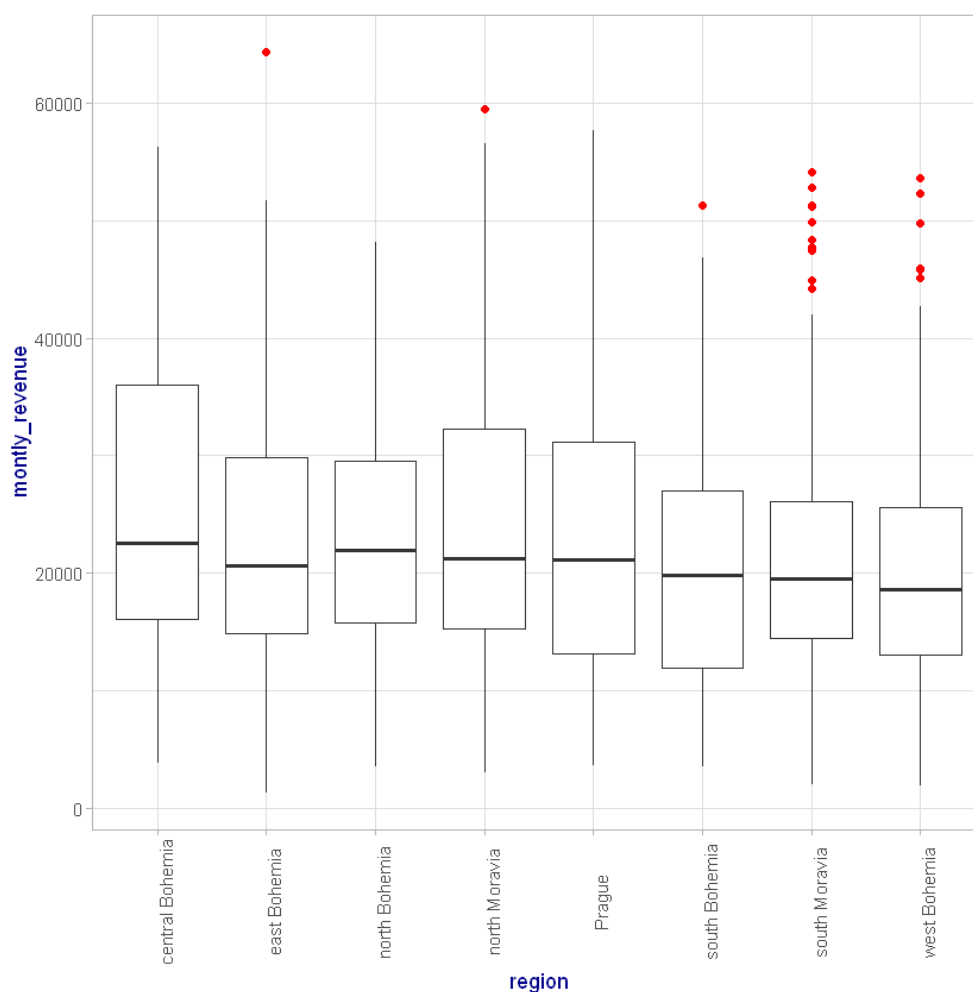


Figura 14 - Boxplot Receita média por região

Avaliando a receita média por região, podemos perceber que existe uma dispersão maior nos grandes polos industriais como em Prague e North Moravia.

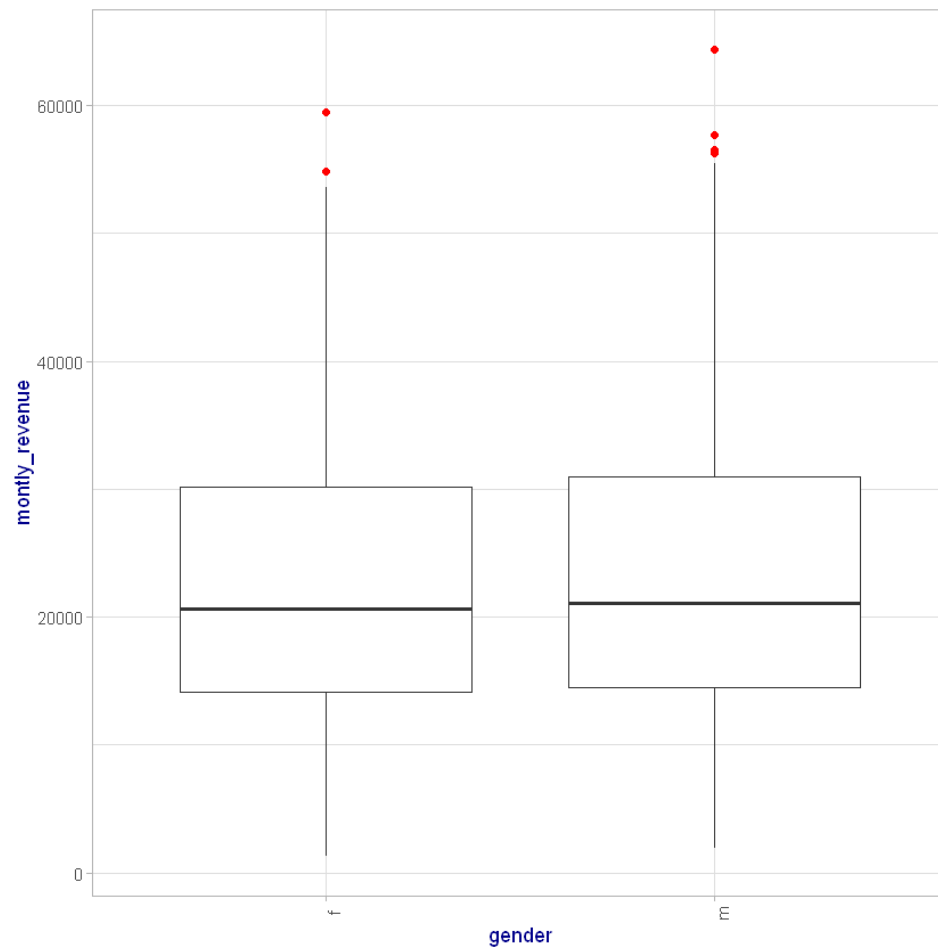
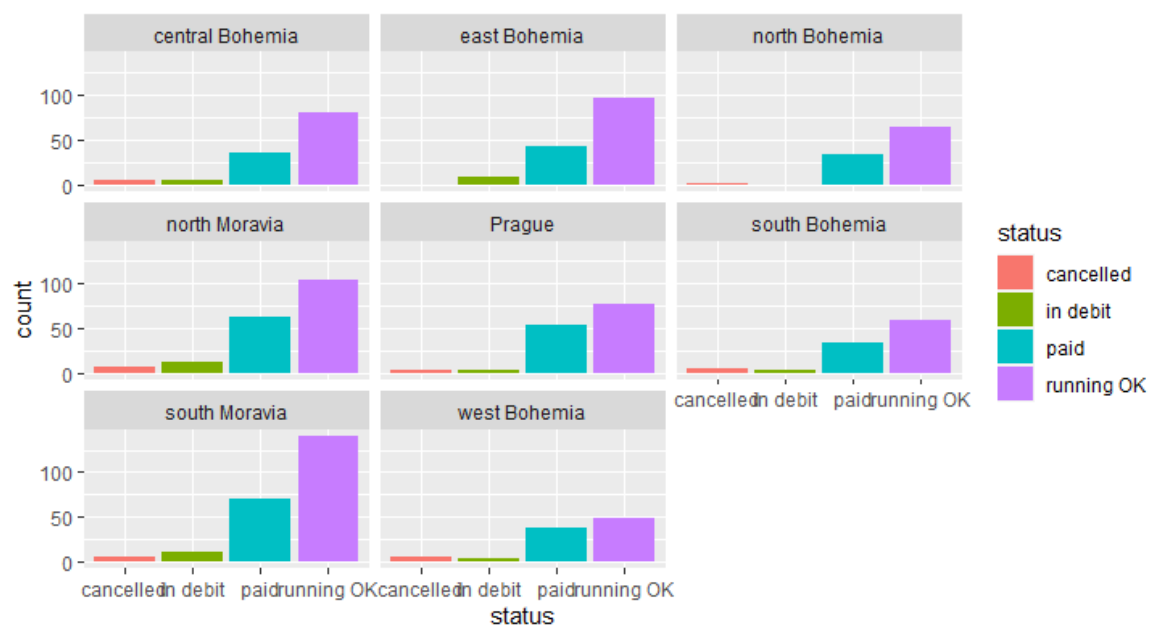


Figura 15 - Receita Média por gênero

- Ao avaliar a receita mensal e o gênero não é possível perceber significância no grau de desigualdade de receitas para os correntistas do banco Czech.

3.6. EMPRÉSTIMOS E NEGATIVADOS



- Das contas que possuem empréstimo, 59% estão com o pagamento em dia, 30% estão com o empréstimo pago, 7% estão devendo e 5% possuem o empréstimo cancelado.

Status com história ruim

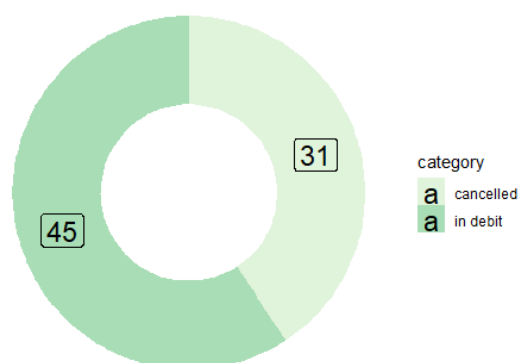


Figura 16 - Empréstimos e negativados

- Dentre os clientes que possuem débito ativo com o banco, temos 40,7 % com empréstimo cancelado e 59,3 empréstimos com empréstimos atrasados.

Apenas 0,87% das contas (39) estão negativadas na data 01/01/1999.

Proporção de contas negativas

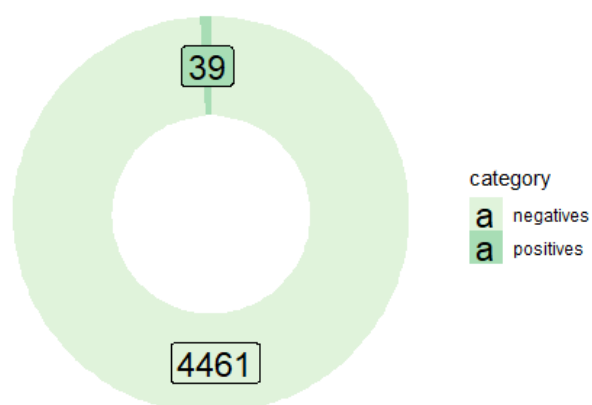


Figura 17 - Proporção de contas Negativas

- Em relação aos clientes devedores, identificamos que:
- Todos os clientes que já tiveram a conta negativa ao menos uma vez no histórico, estão com os contratos de empréstimo cancelados ou atrasados.
- A idade, região e ganho mensal dos clientes não apresentaram correlação com o status do empréstimo.

Pessoas que já ficaram devendo por região e sexo

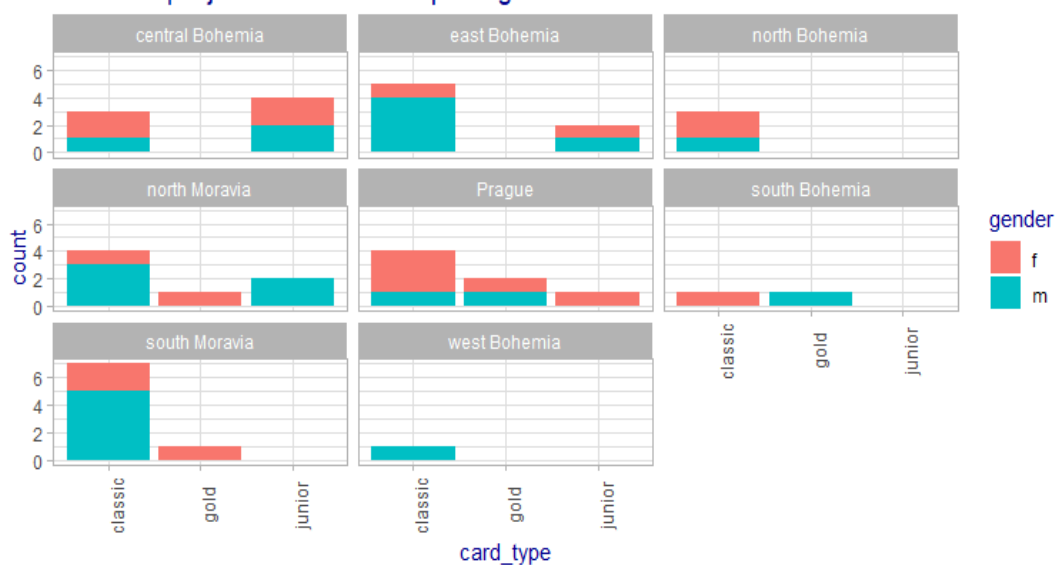


Figura 18 - Negativados por tipo de cartão e sexo

- Os empréstimos que estão cancelados são os que apresentam maior valor.
- O status dos clientes que pegaram empréstimo e não estão em dia :40% optou por cancelar e 60% ainda continua em débito.

- A maior parte dos clientes que estão devendo empréstimos, possuem o hábito de retirar extrato mensalmente.

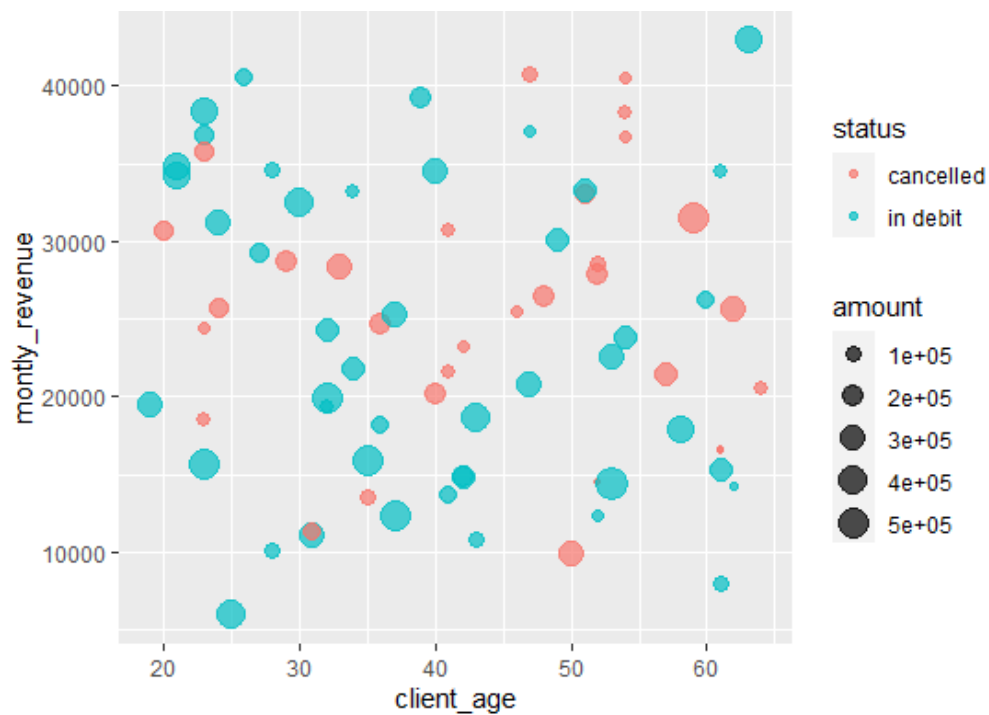


Figura 19 - Negativados receita mensal e idade

Avaliando os clientes com histórico ruim no banco podemos perceber que grande parte deles apesar de débito corrente possuem uma quantidade de receita que permite flexibilização de pagamentos, podendo estar dispostos a renegociar uma negociação dos valores emprestados.

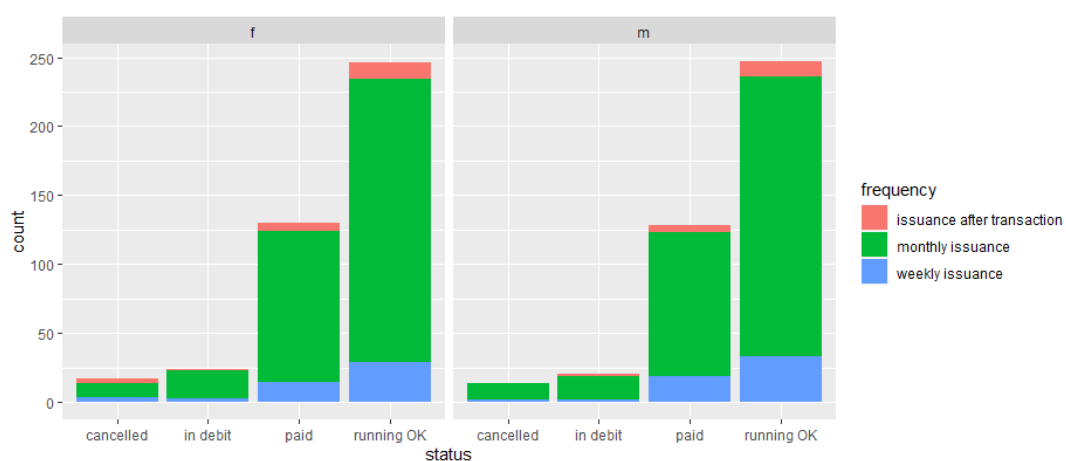


Figura 20 - Status dos Empréstimos ativos

Para se ter uma melhor visualização da hipótese levantada acima, criamos um indicador chamado “porc.parcela”, ou seja a porcentagem da parcela com relação às receitas mensais.

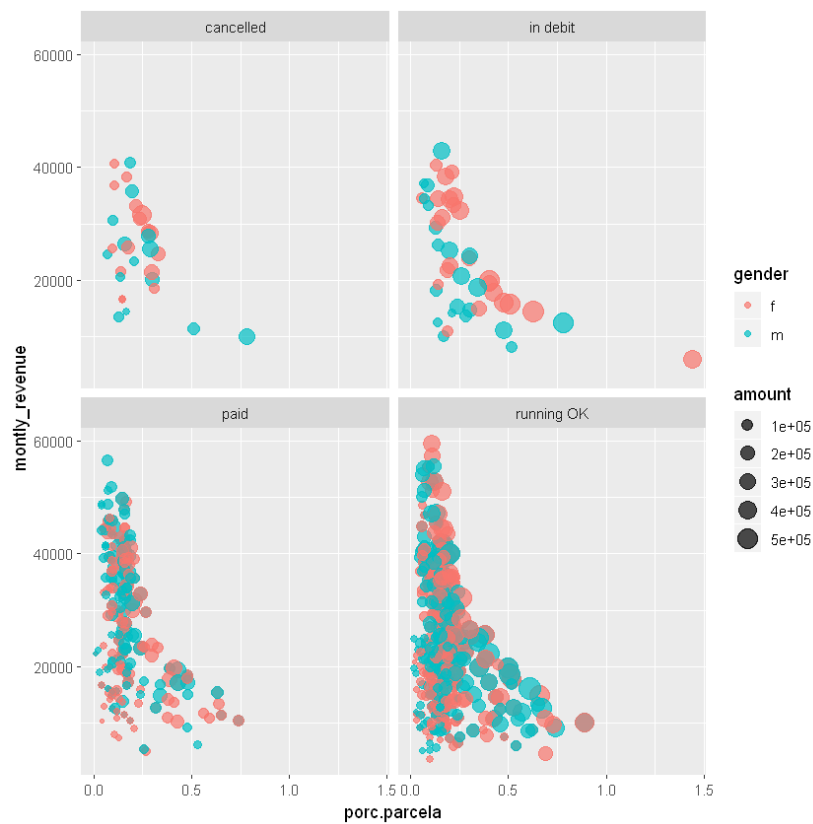


Figura 21 - Status por porcentagem de parcela

Avaliando a porcentagem da parcela com o status de pagamento e idade dos clientes temos:



Figura 22 - Status, Porcentagem de Parcela, Idade e Amount

- Aparentemente a porcentagem de parcela é um fator relevante no que diz respeito ao status do empréstimo, como pode-se observar no conjunto de gráficos acima nos empréstimos

cancelados e “em débito” tendem-se a estar um pouco mais a direita do gráfico, o que representa uma porcentagem maior de parcela.

Essa avaliação é pertinente uma vez que a porcentagem parcela é a relação do valor da parcela com o rendimento mensal de cada cliente, ou seja, quanto maior a parcela menor é a sua capacidade para reação sobre imprevistos.

Com isso podemos avaliar a situação corrente de todos os empréstimos como no gráfico abaixo:

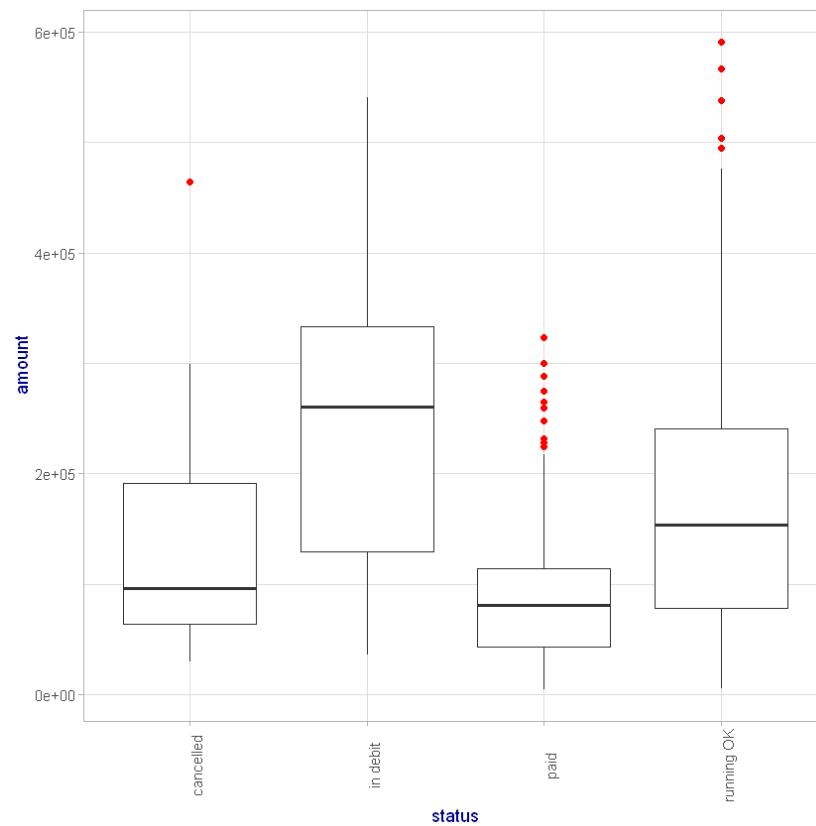


Figura 23 - BoxPlot status e amout

- Empréstimos cancelados tendem a ter uma dispersão de valores menores
- A média dos valores influencia nos empréstimos que estão em atraso.

4. AVALIAÇÃO DE CONCORRÊNCIA

Com os dados de transações realizadas pelos clientes foi possível notar para quais bancos se destinam a maior parte dos valores em pagamentos de títulos e transferências.

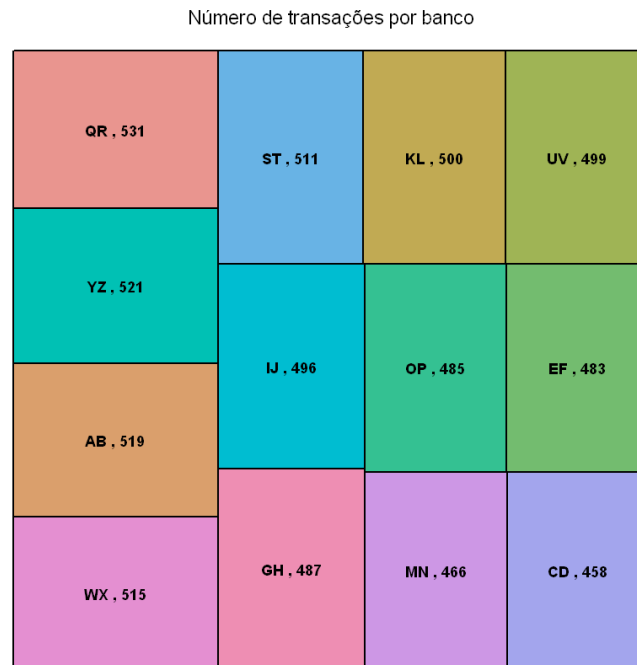


Figura 24 - Número de transferências para outros bancos

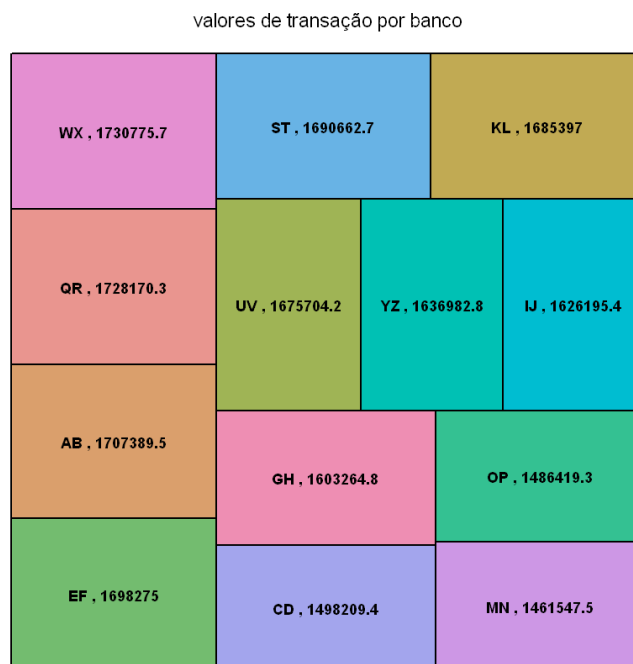


Figura 25 - Valores de transferências para outros bancos

Com a avaliação dos gráficos acima podemos verificar que o banco EF apesar de não figurar entre o banco que possui os maiores números absolutos de transações aparece como um dos que possuem maior valor absoluto isso se deve possivelmente ao valor agregado dos títulos possuídos pelo banco.

5. CONCLUSÃO DO ESTUDO - PLANO DE AÇÃO

Considerável aumento de contas ao longo dos anos, bem como da emissão de cartões. E baixa quantidade de contas em negativo no final de 1998 representando 0,86%, demonstrando um sólido crescimento ao longo dos anos.

Muitos clientes tipo clássico tem valores maiores do que o “Gold”. Com isso é possível perceber uma possibilidade de melhor segmentação do tipo de cartão uma vez que conseguimos saber os clientes com maior poder aquisitivo e assim oferecer serviços customizados “premium” como por exemplo um upgrade no tipo de cartão.

Das contas ativas 15% representam as que possuem empréstimo com o Banco Czeh. Embora 89% das contas ativas dos clientes estava em dia com os pagamentos. Sugerimos um aumento no valor dos juros para empréstimos de valores altos haja vista que notamos uma relação do status de pagamento e o valor de empréstimo, destacado pelo modelo como propensos a não pagamento.

Outra recomendação é a análise mais detalhada em relação a proporção da renda e o pagamento do empréstimo. Ao executar o modelo nota-se que as parcelas em relação ao salário influenciam no não pagamento por isso sugerimos uma análise criteriosa do quanto da renda será consumido na parcela do empréstimo.

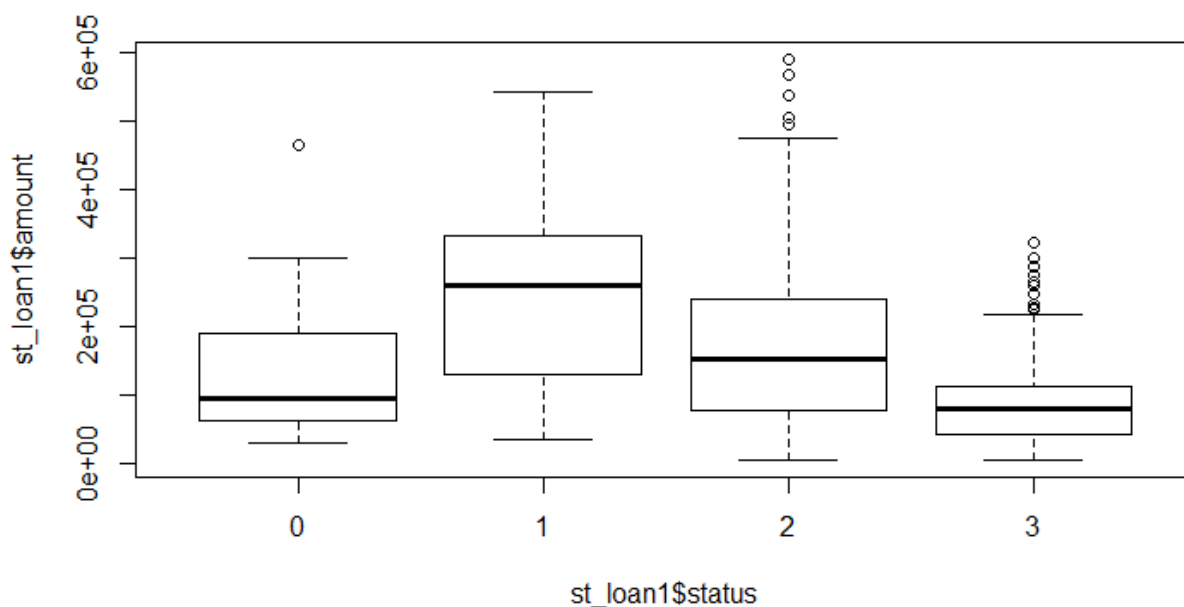


Figura 26 - Valor de Empréstimo por status

Para definição de correlação entre as variáveis utilizamos o método de regressão linear, conforme imagem abaixo, com ela retornamos o valor de P-value caso o valor seja maior que 0,5% consideramos que existe um alto valor de aleatoriedade nos resultados, ou seja, não existe comprovação por esse método que haja relação entre as variáveis.

```
lm(data = st_loan1, formula = status ~ amount ) -> modelo
summary(modelo)

Call:
lm(formula = status ~ amount, data = st_loan1)

Residuals:
    Min       1Q   Median       3Q      Max
-2.39820 -0.30303 -0.07872  0.61429  1.11994

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.450e+00  3.845e-02  63.726  <2e-16 ***
amount      -1.762e-06  2.022e-07  -8.715  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6658 on 825 degrees of freedom
Multiple R-squared:  0.08431,    Adjusted R-squared:  0.0832
F-statistic: 75.96 on 1 and 825 DF,  p-value: < 2.2e-16
```

Tabela Comparativa dos valores de P-value

Atributo	Descrição do Atributo	p-value
Frequency	Frequência em que o cliente retira um “extrato”	0,73
Payments	Quantidade de parcelas (pagamentos)	0,047
Amount	Valor total do Empréstimo	2e(-16)
Duration	Duração do Empréstimo	0,04
Average_salary	Salário médio na região	0,02
Montly_revenue	Receita mensal do cliente	0,001
Porc.Parcela	Porcentagem da Parcela em relação a receita mensal	0,001
Gender	Gênero do cliente	0,6792

Com isso podemos perceber que os atributos Região, Duração, Gênero, não tiveram correlação.

Essa correlação reforça a recomendação realizada acima de uma melhor análise no que diz respeito aos valores de empréstimos cedidos, receita mensal, porcentagem de comprometimento da parcela de financiamento e quantidade de meses em que ocorrerá o parcelamento do empréstimo.

Para visualizar de maneira gráfica a dependência de todos os atributos segue uma matriz de correlação com avaliando a dependência de todos os atributos.

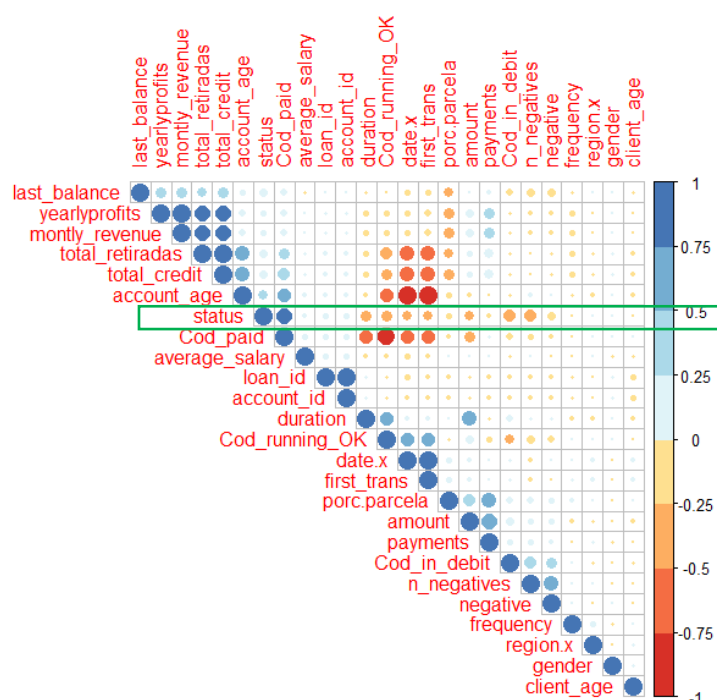


Figura 27 - Matriz de correlação de variáveis

Ao avaliar a matriz de correlação foi possível verificar que o valor que apresenta maior índice de correlação é o número de vezes que o cliente ficou negativo, ou seja, quanto mais o cliente ficou sem pagar maior é o risco de ele não realizar os pagamentos do empréstimo nos prazos estabelecidos. Sendo assim a recomendação é que esse seja um fator determinante no momento de decisão do risco de concessão de um empréstimo.

5.1. RECOMENDAÇÕES

Com base no estudo realizado faremos algumas recomendações de acordo com o relatório exposto acima.

5.1.1. DIVERSIFICAÇÃO DE CLIENTES

Como exposto no gráfico acima existem algumas regiões em que há margem clara para expansão da atividade do banco, essas áreas com grande potencial de expansão apresentam baixo grau de inadimplência o que pode aumentar a lucratividade do banco.

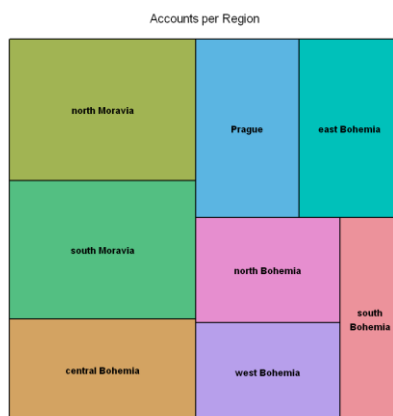


Figura 28 - Quantidade de Clientes por Região

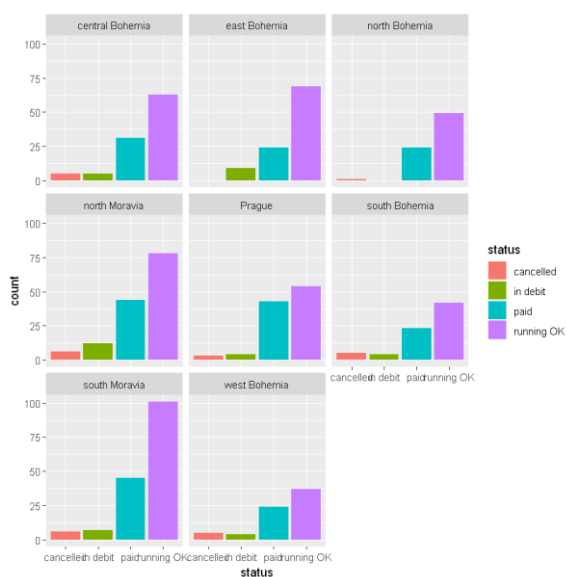


Figura 29 - Status de Pagamento de Empréstimo por Região

5.1.2. MELHORIA NOS ALGORITMOS DE RISCO DE EMPRÉSTIMO

Sugerimos uma atualização dos algoritmos que determina o risco de concessão de empréstimos aumentando o peso para os seguintes atributos:

- Número de vezes que o cliente ficou negativado
- Valor total solicitado
- Número de parcelas do Empréstimo
- Receita média mensal
- Tempo de conta
- Localização

Os atributos abaixo podem ter pesos reduzidos:

- Sexo
- Renda média pela localização
- Idade

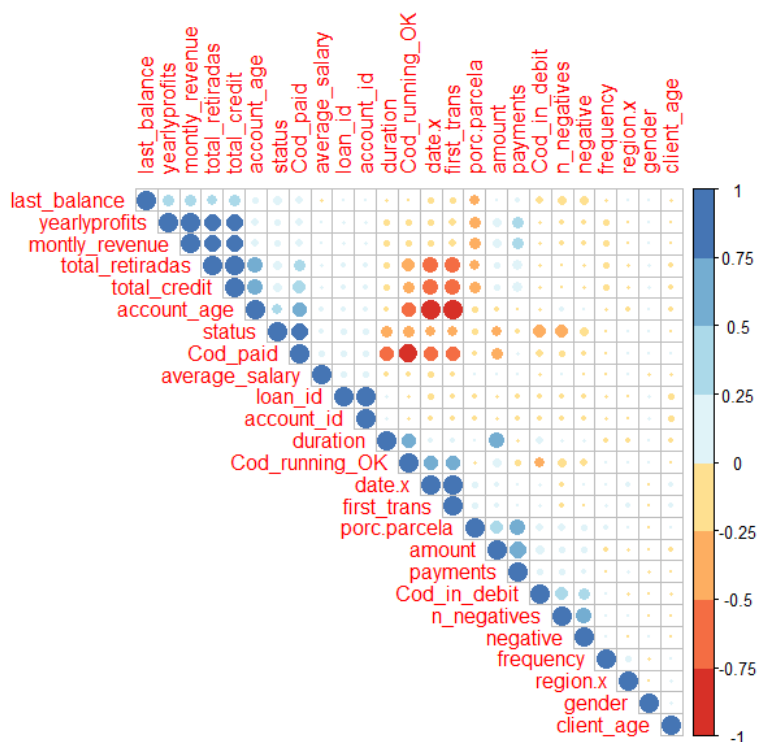


Figura 30 - Matriz de Correlação

5.1.3. REORGANIZAÇÃO DA SEGMENTAÇÃO DE CARTÕES DE CRÉDITOS

De acordo com as análises realizadas ficou claro que alguns clientes possuem uma alta renda que não é compatível com o seu tipo de cartão.

Dessa maneira sugerimos criar uma linha de corte para que o cliente que tenha uma renda mensal acima de um valor a ser especificado pelo banco, nossa sugestão é que seja acima de 50% da

renda média da região, receba ofertas de upgrades do cartão de crédito e benefícios adicionais em sua conta.

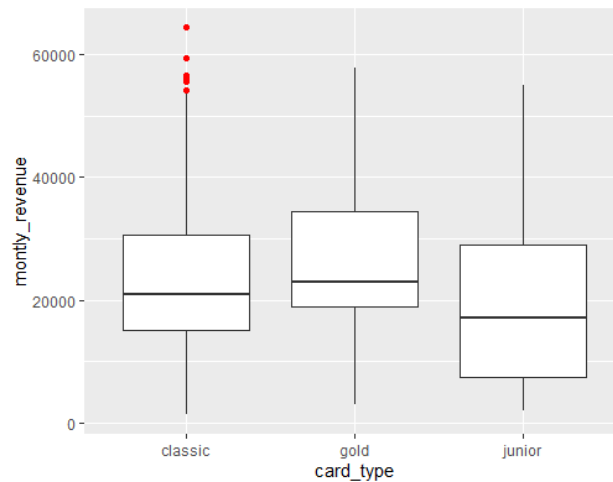


Figura 31 - Boxplot tipo de cartão e receita mensal

5.1.4. EXPANSÃO DE TÍTULOS

Ao analisar as transações dos clientes observa-se que muitos deles retiram dinheiro de sua conta para pagar títulos de “household” de outros bancos, então sugerimos que o Banco Czech expanda a quantidade de títulos de “household” e “insurance payment” para que o cliente não retire o dinheiro do Czech Bank.

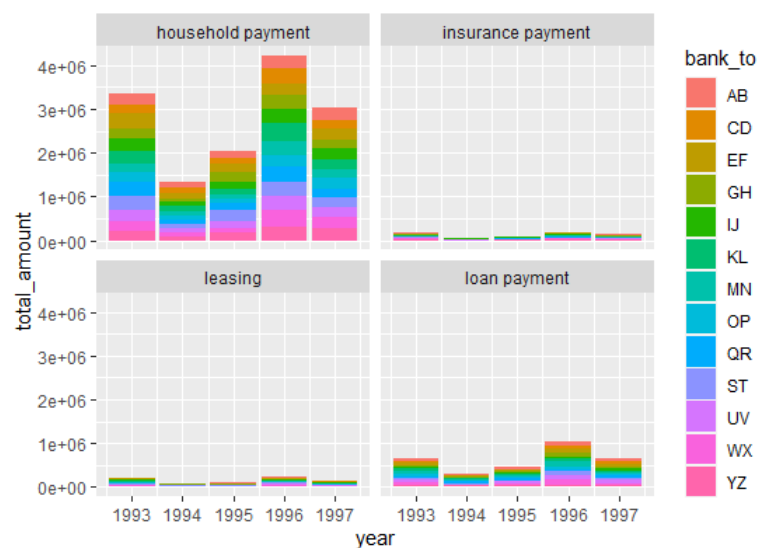


Figura 32 - Títulos pagos em bancos

5.1.5. AUMENTO DE OFERTAS DE CRÉDITO

Como o banco possui um índice de inadimplência baixo, pode-se aumentar a quantidade de oferta de empréstimos desde que respeitadas as características do algoritmo expostas acima.

Proporção de contas negativas

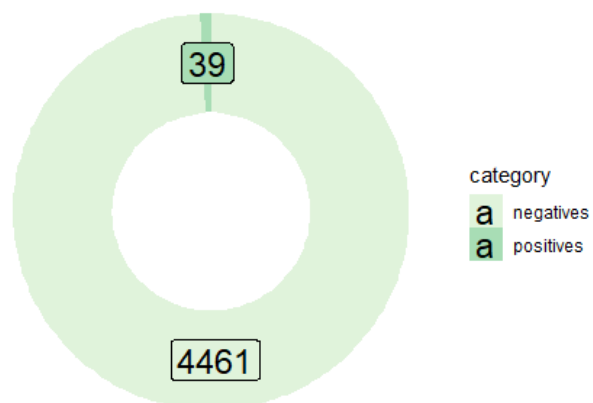


Figura 33 - Proporção de contas negativas