

FUNDAÇÃO GETULIO VARGAS - FGV

TRABALHO I – CZECH BANK

Ana Paula Pudo
Fábio Monteiro
Lucas Sena Alves
Marcos Soares

**MBA Business Analytics e Big Data | Banco de
Dados e Visualização**

Prof. Gustavo Mirapalheta

Brasília
2021

SUMÁRIO

LISTA DE FIGURAS	2
GRÁFICOS	2
1. Introdução.....	3
1.1 Objetivo Geral	3
1.2 Objetivos Específicos	3
2. Metodologia e Base de Dados	4
2.1 Clients.....	6
2.2 Account.....	6
2.3 Disp.....	6
2.4 Card.....	6
2.5 District.....	7
2.6 Loan.....	7
2.7 Order.....	8
2.8 Transaction	8
3. Análise I – Perfil da Carteira de Clientes.....	8
4. Análise II - Comportamento dos clientes.....	11
5. Conclusões.....	18

LISTA DE FIGURAS

Figura 1 – Estrutura da base de dados

Figura 2 – Library utilizadas no R

GRÁFICOS

Gráfico 1 – Clientes por sexo

Gráfico 2 – Faixa etária dos clientes

Gráfico 3 – Distribuição de clientes por distrito

Gráfico 4 – Distribuição de Renda por Região

Gráfico 5 – Distribuição total clientes x total habitantes por região

Gráfico 6 – Clientes que possuem cartão de crédito

Gráfico 7 – Clientes x Tipo de Cartão
Gráfico 8 – Empréstimos realizados
Gráfico 9 – Status dos Empréstimos
Gráfico 10 – Débito/Crédito por quantidade de transações
Gráfico 11 – Débito/Crédito por valor de transações
Gráfico 12 – Quantidade de transações por tipo
Gráfico 13 – Valor das transações por tipo
Gráfico 14 – Evolução das operações por tipo e ano
Gráfico 15 – Distribuição de quantidade de transferências por tipo
Gráfico 16 – Distribuição de valor de transferências por tipo

1. Introdução

O Czech Bank é uma fictícia instituição financeira privada situada na República Tcheca. A proposta desta empresa é oferecer os serviços mais comuns de um banco comercial tradicional, como empréstimos, cartões de crédito, contas de pagamento, entre outras.

Após anos de operação e com uma base de correntistas já estabelecida, a alta direção da companhia decidiu que era hora de melhorar seus serviços de análise da carteira de clientes. Os principais diretores e gerentes possuem apenas uma vaga ideia de quem são os bons e maus pagadores da instituição, o que limita ações promocionais para aqueles que possuem suas contas em dia, ao mesmo tempo que aumenta o risco de calote dos inadimplentes.

Com base nas informações armazenadas sobre transações bancárias, empréstimos concedidos e cartões de crédito emitidos, a instituição espera compreender melhor o perfil de seus clientes e promover melhorias nos serviços de promoções e vendas e de análise de risco.

1.1 Objetivo Geral

Compreender o perfil de bons e maus pagadores da instituição financeira.

1.2 Objetivos Específicos

- Sugerir decisões baseadas em dados
- Gerar análises e insights de ações para os bons pagadores

- Determinar perfis com alta probabilidade de inadimplência.

2. Metodologia e Base de Dados

A base de dados utilizada neste estudo foi somente a “Czech Data”. Contendo registro de transações bancárias, localidades, contas e clientes extraídas de uma instituição bancária situada na República Tcheca.

Foi realizada a leitura das 8 tabelas disponibilizadas em scripts no formato .asc. Importada a base como sendo dataframes para a ferramenta Rstudio, possibilitando a aplicação das investigações, análises exploratórias e análises estatísticas por meio da linguagem R.

Não foram removidas colunas dos conjuntos de dados e nem descartados dataframes, pois foram consideradas todas as informações pertinentes. Mesmo os dados que não auxiliaram diretamente nas análises deste estudo serviram indiretamente para o melhor entendimento da base fornecida.

Uma vez que se trata de uma cultura diferente e com aspectos e regramentos financeiros com diferenças dos adotados no Brasil, se fez prudente debruçar uma parte do tempo desta análise para melhor conhecer essas minúcias e jargões específicos adotados na base “Czech Data”.

A estrutura dos dados está elaborada da seguinte forma:

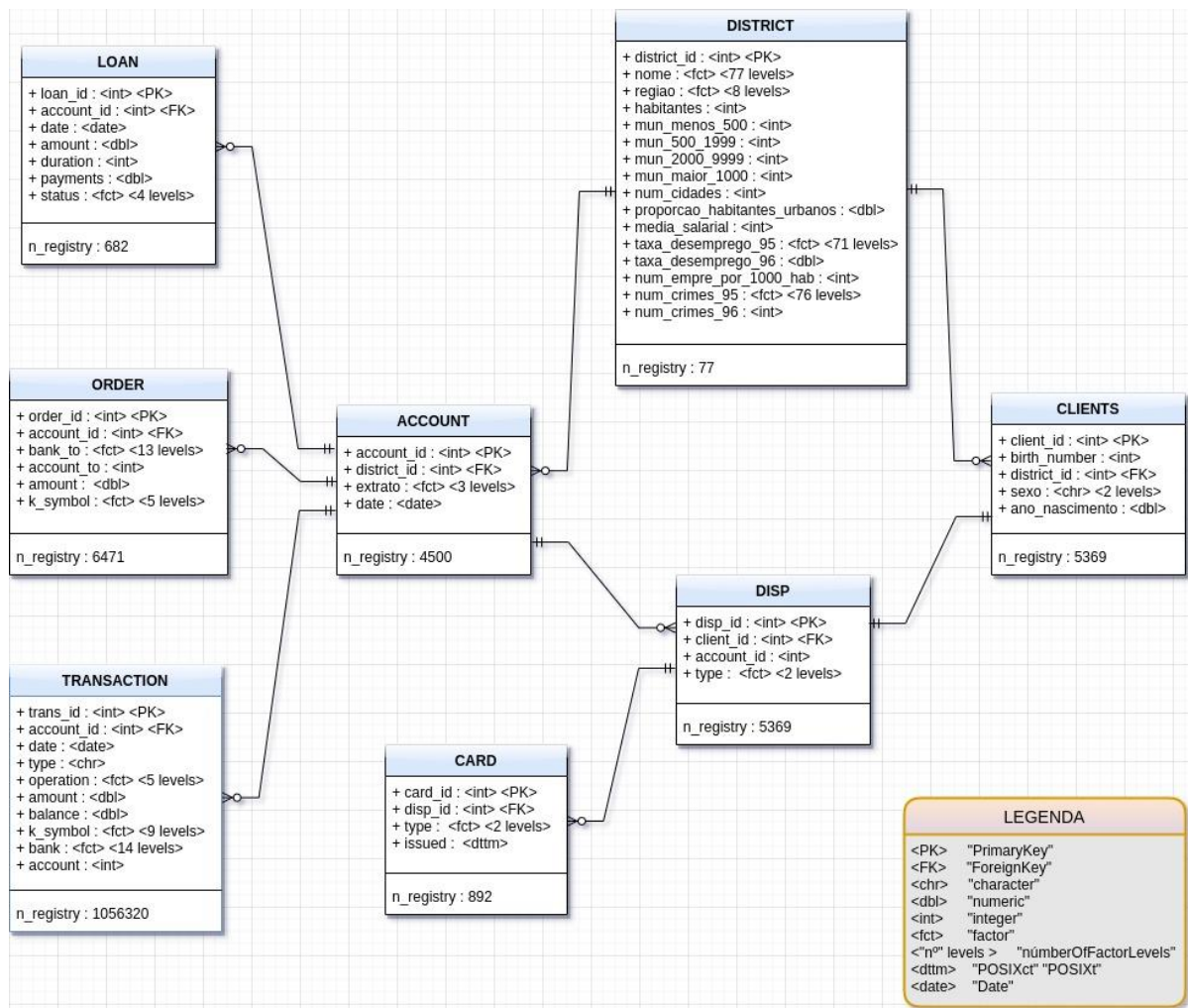


Figura 1: Estrutura da base de dados

Para a elaboração deste estudo, foram utilizadas as seguintes librarys disponíveis na linguagem R.

```
library(dplyr)
library(stringr)
library(gmodels)
library(ggplot2)
library(lubridate)
library(gridExtra)
library(grid)
library(lattice)
```

Figura 2: Librarys utilizadas no R

Fazendo o uso desses pacotes, fizemos ações e tratamentos nos dados necessários para as investigações aplicadas neste estudo em cada um dos dataframes, detalhadas no decorrer deste trabalho.

2.1 Clients

Visualização do banco de dados e separação por gênero e data de nascimento dos clientes.

```
View(data_client)
summary(data_client)
data_client = data_client %>% mutate(sexo = ifelse(as.numeric
  (substr(data_client$birth_number,3,4)) > 12, 'F', 'M'))
data_client$ano_nascimento = as.numeric(substr(data_client$birth_number,1,2))+1900
```

2.2 Account

Visualização do banco de dados, tradução livre das categorias, renomeação da terceira variável e organização do formato das datas.

```
View(data_account)
summary(data_account)
data_account$frequency = ifelse(data_account$frequency == 'POPLATEK MESICNE', 'MENSAL',
  ifelse(data_account$frequency == 'POPLATEK PO OBRATU', 'TRANSAÇÃO', 'SEMANAL'))
colnames(data_account)[3]='extrato'
data_account$extrato = factor(data_account$extrato)
levels(data_account$frequency)
data_account$date = ymd(data_account$date)
```

2.3 Disp

Visualização do banco de dados, categorização e tradução livre.

```
View(data_disp)
summary(data_disp)
data_disp$type = ifelse(data_disp$type=='DISPONENT', 'DEPENDENTE',
  ifelse(data_disp$type == 'OWNER', 'PROPRIETÁRIO', ''))
data_disp$type = factor(data_disp$type)
```

2.4 Card

Visualização do banco de dados, categorização. adicionamos 19 na data de emissão e organização do formato das datas.

```
View(data_card)
summary(data_card)
data_card$type = factor(data_card$type)
data_card$issued <- paste0("19", data_card$issued) #ADICIONANDO 19 NO ANO....
data_card$issued <- ymd_hms(data_card$issued)
```

2.5 District

Visualização do banco de dados e renomeação das variáveis categorização.

```
View(district)
summary(data_district)
colnames(data_district)[1]="district_id"
colnames(data_district)[2]="nome"
colnames(data_district)[3]="regiao"
colnames(data_district)[4]="habitantes"
colnames(data_district)[5]="mun_menos_500"
colnames(data_district)[6]="mun_500_1999"
colnames(data_district)[7]="mun_2000_9999"
colnames(data_district)[8]="mun_maior_1000"
colnames(data_district)[9]="num_cidades"
colnames(data_district)[10]="proporcao_habitantes_urbanos"
colnames(data_district)[11]="media_salarial"
colnames(data_district)[12]="taxa_desemprego_95"
colnames(data_district)[13]="taxa_desemprego_96"
colnames(data_district)[14]="num_empre_por_1000_hab"
colnames(data_district)[15]="num_crimes_95"
colnames(data_district)[16]="num_crimes_96"
```

2.6 Loan

Visualização do banco de dados, categorização e renomeação das categorias da variável status., organização do formato das datas

```
View(data_loan)
summary(data_loan)
data_loan$status = ifelse(data_loan$status=='A', 'FINALIZADO',
                           ifelse(data_loan$status=='B', 'INADIMPLENTE',
                                   ifelse(data_loan$status=='C', 'ADIMPLENTE', 'ATRASADO'))))
data_loan$status = factor(data_loan$status)
data_loan$date = ymd(data_loan$date)
data_loan$amount = as.double(data_loan$amount)
```

2.7 Order

Visualização do banco de dados, categorização, tradução das categorias da variável `k_symbol` e organização do formato das datas.

```
View(data_order)
summary(data_order)
data_order$amount = as.double(data_order$amount)
data_order$k_symbol = ifelse(data_order$k_symbol == 'POJISTNE', 'PAGAMENTO SEGUROS',
                             ifelse(data_order$k_symbol == 'SIPO', 'PAGAMENTOS DIVERSOS',
                                     ifelse(data_order$k_symbol == 'UVER', 'PAGAMENTO EMPRÉSTIMOS',
                                             ifelse(data_order$k_symbol == 'LEASING', 'LEASING', 'N.A.'))))
data_order$k_symbol = factor(data_order$k_symbol)
levels(data_order$k_symbol)
```

2.8 Transaction

Visualização do banco de dados, categorização, tradução das categorias da variável `k_symbol` e `operation` e organização do formato das datas.

```
View(data_trans)
summary(data_trans)
data_trans$type = ifelse(data_trans$type == 'PRIJEM', 'CRÉDITO', 'DÉBITO')
data_trans$k_symbol = factor(data_trans$k_symbol)
levels(data_trans$k_symbol)
data_trans$operation = ifelse(data_trans$operation == 'VYBER KARTOU', 'SAQUE CARTÃO DE CRÉDITO',
                              ifelse(data_trans$operation == 'VKLAD', 'DEPÓSITO EM DINHEIRO',
                                      ifelse(data_trans$operation == 'PREVOD Z UCTU', 'RECEBIMENTO DE BANCO',
                                              ifelse(data_trans$operation == 'VYBER', 'SAQUE EM DINHEIRO', 'REMESSA'))))
data_trans$operation = factor(data_trans$operation)
data_trans$date = ymd(data_trans$date)
data_trans$amount = as.double(data_trans$amount)
data_trans$balance = as.double(data_trans$balance)
```

Tomadas as ações iniciais, o presente estudo apresenta a seguir o conjunto de análises e criação de insights cabíveis.

3. Análise I – Perfil da Carteira de Clientes

Como abordagem inicial de análise buscamos compreender o perfil dos clientes que fazem parte da carteira do banco. Observamos que existe uma mistura homogênea de homens e mulheres e que a faixa etária média é de 47 anos, portanto uma clientela mais madura.


```

par(mar = c(5,3,2,3))
par(las = 1)
vrPorcentagemM <- round((nrow(subset(data_client, sexo == 'M')) / nrow(data_client))*100,digits = 2)
vrPorcentagemF <- round((nrow(subset(data_client, sexo == 'F')) / nrow(data_client))*100,digits = 2)
barplot(table(data_client$sexo),
        main = 'Distribuição de Clientes por Sexo.',
        ylab = 'Total de Clientes',
        ylim = c(0,3000),
        col = hcl.colors(2, palette = "Peach"),
        names.arg = c(str_c('',vrPorcentagemF,'% ',sep = ''),str_c('',vrPorcentagemM,'% ',sep = '')))
legend("topright", legend = c('Feminino', 'Masculino'), fill = hcl.colors(2, palette = "Peach"))

par(mar = c(5,4,7,5))
hist(data_client$ano_nascimento,
     main = 'Distribuição de Clientes por Ano de Nascimento.',
     xlab = 'Ano de Nascimento.',
     ylab = 'Total de Clientes',
     ylim = c(10,600),
     col = hcl.colors(20, palette = "Red-Blue"))

```

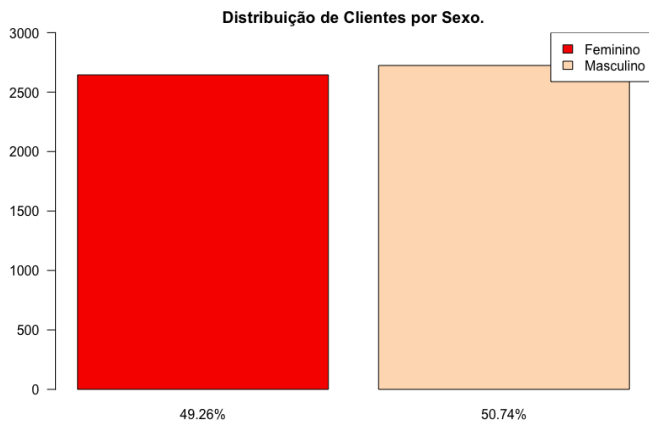


Gráfico 1 - Cliente por sexo

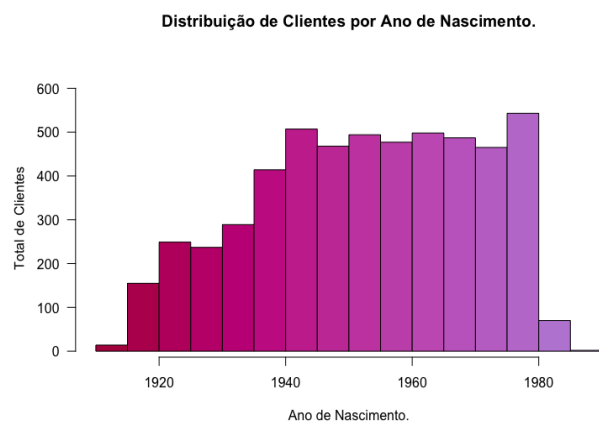


Gráfico 2 - Faixa etária dos clientes

Desta forma, um primeiro insight possível é de que o banco pode avaliar sua carteira de produtos e verificar quais são aqueles que possuem maior atratividade para um público +40 e pensar em ações de vendas/adesão para eles.

Mas onde estão esses clientes? Para verificar a localização dos clientes e como está distribuída a renda por região, foi realizada a seguinte verificação:

```

data_client_district = merge(data_client, data_district, all = T)
d1 <- data_client %>% group_by(district_id) %>% summarise(count = n())
d1$nome = data_district$nome
d1=head(d1,20)#TOP 20 dis
par(mar = c(8,6,3,3))
par(las = 2)
barplot(height=d1$count,
        main = 'Top 20 Distritos por Total de Clientes.',
        names=d1$nome,
        col= as.vector(terrain.colors(20)),
        yaxp=c(0, max(d1$count), 10))

```

```

par(mar=c(10,6,3,3))
par(las=2)
boxplot(media_salarial ~ regioao, data = data_district,
        main = 'Média de Renda Por Região',
        cex.main = 1.5,
        xlab = '',
        ylab = '',
        col = hcl.colors(10, palette = 'Cold'),
        border = 'gray20',
        cex.axis = 1.2,
        cex.lab = 1.2,
        ylim = c(min(data_district$media_salarial), max(data_district$media_salarial)))

```

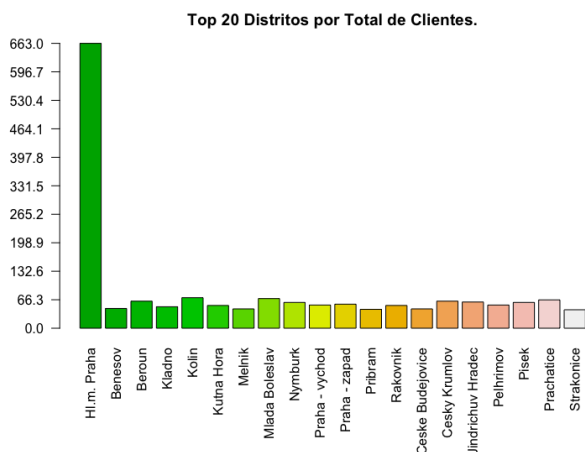


Figura 3. Distribuição de clientes por distrito

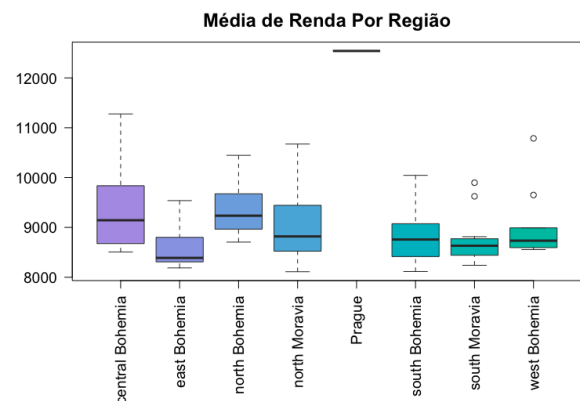


Figura 4 Distribuição de Renda por Região

Conforme a figura 3 e 4, observamos que a maior concentração de clientes está no distrito de HL.m. Praha, região de Prague, provavelmente justificado por ser a região que possui a maior média de renda (\$12.541). Em todas as outras regiões, a média salarial não ultrapassa os \$9.050.

```

d1 <- data_client_district %>% group_by(regiao) %>% summarise(clientes = n())
d2 <- data_district %>% group_by(regiao) %>% summarise(total_habitantes = sum(habitantes))
d1$habitantes = d2$total_habitantes
barplot(t(as.matrix(d1[, 2:3])),
        beside = TRUE,
        main = 'Total de Clientes x Total de Habitantes por Região.',
        cex.main = 1.2,
        names.arg = d1$regiao,
        col = hcl.colors(2, palette = "Temps"))
legend(x, y = NULL, legend, fill = NULL, col = par("col"), border = "black", lty, lwd, pch, angle
      = 45, density = NULL, bty = "o", bg = par("bg"), box.lwd = par("lwd"), box.lty = par("lty"),
      box.col = par("fg"), <...truncated...>)
legend("topleft", legend = c('Clientes', 'Habitantes'), fill = hcl.colors(2, palette = "Temps"))

```

Na comparação entre total de habitantes por região versus total de clientes do banco, conforme figura 5, observamos que seria importante fazer um estudo para se entender como é composta a população de cada região e se o banco poderia expandir sua carteira de clientes.

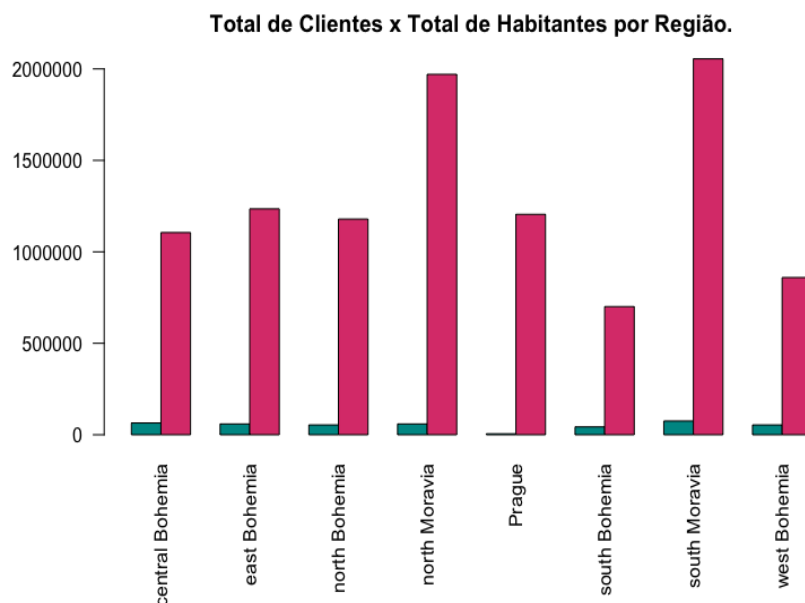


Figura 5 - Distribuição total clientes x total habitantes por região

4. Análise II - Comportamento dos clientes

Na sequência, o estudo buscou investigar como seriam os hábitos de consumo e os serviços utilizados pelos clientes do banco. Para isso, verificamos o seguinte:

```

dc1 <- data_client %>% inner_join(data_disp, by = "client_id") %>% left_join(data_card, by = "disp_id")
dc2 <- group_by(.data = dc1, POSSUI_CARTAO=is.na(card_id)) %>% summarise(count=n())
par(las = 1)
vrPorcentagemM <- round((nrow(subset(dc1, is.na(card_id))) / nrow(dc1))*100,digits = 2)
vrPorcentagemF <- round((nrow(subset(dc1, !is.na(card_id))) / nrow(dc1))*100,digits = 2)
par(las = 1)
barplot(table(dc2$POSSUI_CARTAO), height = dc2$count,
        main = 'Distribuição de Clientes com Cartão de Crédito.',
        xlab = 'Possui Cartão',
        ylab = 'Total de Clientes',
        ylim = c(0,5000),
        col = hcl.colors(2, palette = "Cividis"),
        names.arg = c(str_c('',vrPorcentagemF,'% ',sep = ''),str_c('',vrPorcentagemM,'% ',sep = '')))
legend("topleft", legend = c('NÃO', 'SIM'), fill = hcl.colors(2, palette = "Cividis"))

dc4 = filter(.data = dc1, !is.na(card_id))
dc5 <- filter(.data = dc4, !is.na(card_id)) %>% group_by(type.y) %>% summarise(count = n())
vrPorcentagemC <- round((nrow(subset(dc4, type.y == 'classic')) / nrow(dc4))*100,digits = 2)
vrPorcentagemG <- round((nrow(subset(dc1, type.y == 'gold')) / nrow(dc1))*100,digits = 2)
vrPorcentagemJ <- round((nrow(subset(dc1, type.y == 'junior')) / nrow(dc1))*100,digits = 2)
par(las = 1)
barplot(table(dc5$type.y), height = dc5$count,
        main = 'Distribuição de Clientes com Cartão por Tipo',
        xlab = 'Tipo de Cartão',
        ylab = 'Total de Clientes',
        ylim = c(0,800),
        col = hcl.colors(3, palette = "Emrld"),
        names.arg = c(str_c('',vrPorcentagemC,'% ',sep = ''),str_c('',vrPorcentagemG,'% ',sep = ''),str_c('',vrPorcentagemJ,'% ',sep = '')))
legend("topright", legend = c('Classic', 'Gold', 'Júnior'), fill = hcl.colors(3, palette = "Emrld"))

```

Percebe-se que, da carteira total de 5.369 clientes, 83% possuem cartão de crédito. A grande maioria desse recorte conta um cartão da categoria Classic.

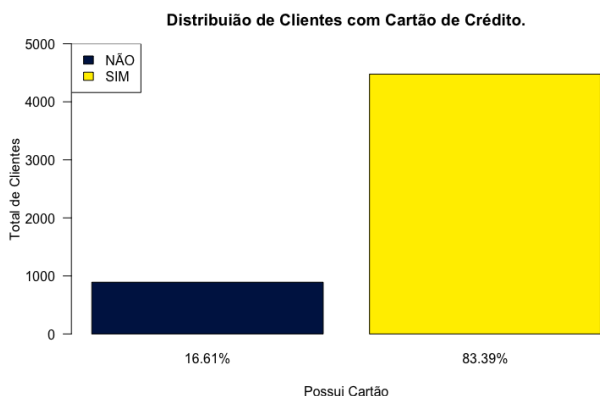


Figura 6 - Cliente que possuem cartão de crédito

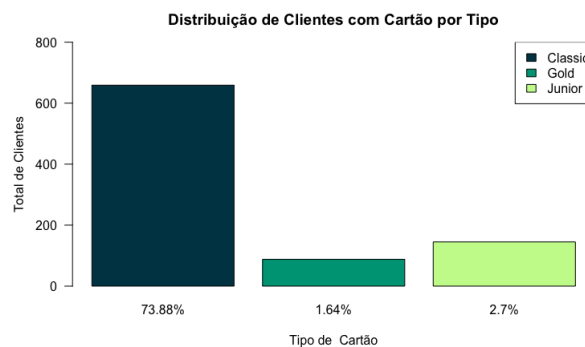


Figura 7 - Clientes x Tipo de Cartão

Com uma base de clientes madura e com grande adesão a esse tipo de meio de pagamento, é possível criar novos produtos considerando as faixas de rendas. Programas de milhagem e upgrade de categorias, por exemplo, devem estar no radar do banco. Já considerando o produto empréstimo, podemos dizer que este produto indica um nível de endividamento dos clientes relativamente baixo.

```

d1 <- data_client %>% inner_join(data_disp, by = "client_id") %>% left_join(data_loan, by = "account_id")
d1
vrPorcentagemP <- round((nrow(subset(d1, !is.na(loan_id))) / nrow(d1))*100,digits = 2)
vrPorcentagemN <- round((nrow(subset(d1, is.na(loan_id))) / nrow(d1))*100,digits = 2)
d2 <- group_by(.data = d1, POSSUI_EMPRESTIMO = !is.na(loan_id)) %>% summarise(count=n())
d2
par(las=1)
barplot(d2$count,
  main = 'Percentual de Clientes que possuem e/ou já realizaram empréstimo.',
  cex.main = 1,
  col = hcl.colors(2, palette = "Vik"),
  ylim = c(0,5000),
  ylab = "",
  names.arg = c(str_c('',vrPorcentagemN,'% ',sep = ''),str_c('',vrPorcentagemP,'% ',sep = '')))
legend("topright", legend = c('Não Possui', 'Possui/Já realizou'), fill = hcl.colors(2, palette = "Vik"))

par(las=1)
vrPorcentagemAT <- round((nrow(subset(data_loan, status == 'INADIMPLENTE')) / nrow(data_loan))*100,digits = 2)
vrPorcentagemF <- round((nrow(subset(data_loan, status == 'FINALIZADO')) / nrow(data_loan))*100,digits = 2)
vrPorcentagemAD <- round((nrow(subset(data_loan, status == 'ADIMPLENTE')) / nrow(data_loan))*100,digits = 2)
d1 <- data_loan %>% group_by(status) %>% summarise(count=n(), soma_valor=sum(amount)) %>% arrange(soma_valor)
par(mar=c(4,6,4,4))
barplot(d1$soma_valor,
  main = 'Distribuição de Empréstimos por Status.',
  cex.main = 1.2,
  col = hcl.colors(4, palette = "Temps"),
  ylim = c(0,70000000),
  ylab = "",
  names.arg = c(str_c('',vrPorcentagemI,'% ',sep = ''),
    str_c('',vrPorcentagemAT,'% ',sep = ''),
    str_c('',vrPorcentagemF,'% ',sep = ''),
    str_c('',vrPorcentagemAD,'% ',sep = '')),
  xlab = "")
legend("topleft", legend = d1$status , fill = hcl.colors(4, palette = "Temps"))

```

Apenas 15% possuem ou já realizaram alguma contratação no período analisado (figura 9), e entre os que realizaram, 59.09% são adimplentes e 30% já encerraram suas obrigações com o banco.

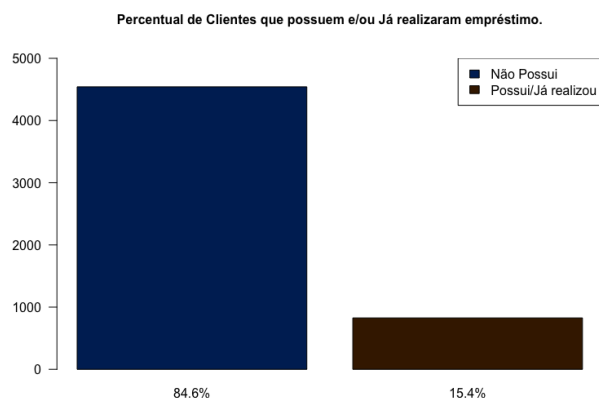


Figura 8 - Empréstimos realizados

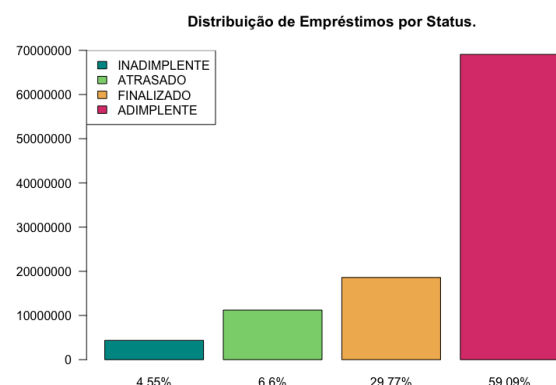


Figura 9 - Status dos Empréstimos

Na análise das transações do banco, identificamos as seguintes características:

- Apesar do nível de transação por quantidade a operação de débito seja maior, quando verificamos a grandeza de valores, os níveis são muito parecidos.
- As operações de depósito em dinheiro e saque com cartão de crédito apresentaram o maior número de transações;

```
d1 <- data_trans %>% group_by(type) %>% summarise(count = n(), valor = sum(amount))
d1 #TABELA EXPLICATIVA
```

```
par(mar=c(5,5,3,3))
par(las=1)
barplot(d1$count,
        beside = TRUE,
        main = 'Distribuição de Quantidade de Transações por Tipo.',
        cex.main = 1.2,
        names.arg = d1$type,
        col = hcl.colors(2, palette = "Blue-Red 3"),
        ylab = "",
        ylim = c(0,700000),
        xlab = "Tipo de Transação")
```

```
par(mar=c(5,7,3,3))
barplot(d1$valor,
        beside = TRUE,
        main = 'Distribuição de Valor de Transações por Tipo',
        cex.main = 1.2,
        names.arg = d1$type,
        col = hcl.colors(2, palette = "Blue-Red 3"),
        ylab = "",
        ylim = c(0,3500000000),
        xlab = "Tipo de Transação")
```

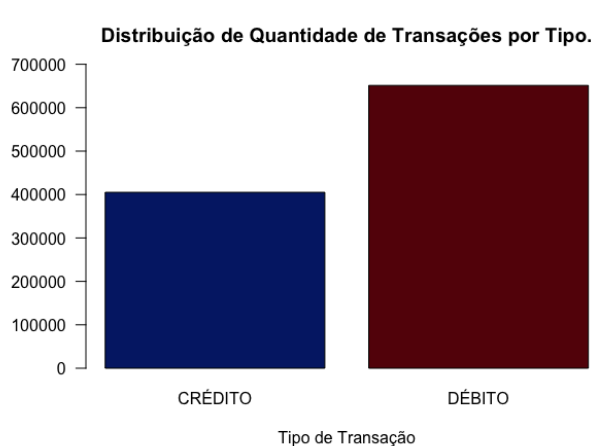


Figura 10 - Débito/Crédito por qtd de trans

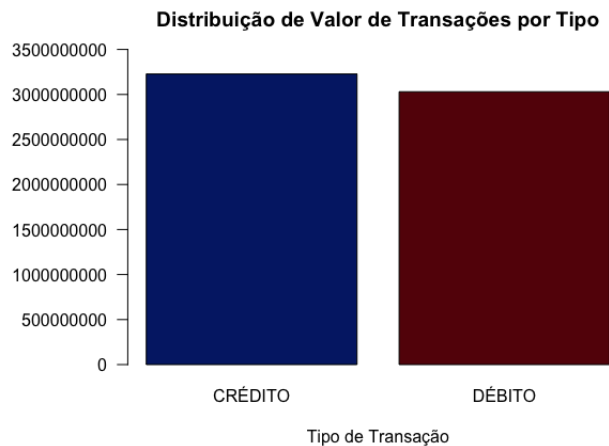


Figura 11- Débito/Crédito por valor de transações

```

d1 <- data_trans %>% group_by(operation) %>%
  summarise(count = n(), valor = sum(amount)) %>% arrange(count)
d1 #TABELA EXPLICATIVA
par(mar=c(5,5,2,3))
par(las=2)
barplot(d1$count,
  main = 'Distribuição de Quantidade Transações por Operação',
  cex.main = 1.2,
  ylim = c(0,500000),
  col = hcl.colors(5, palette = "Blue-Red 3"),
  xlab = "")
legend("topleft", legend = c(
  'Depósito em Dinheiro', 'Recebimento de Banco',
  'Remessa', 'Saque Cartão de Crédito', 'Saque em Dinheiro'),
  fill = hcl.colors(5, palette = "Blue-Red 3"))

d1 <- d1 %>% arrange(valor)
d1
par(mar=c(5,8,2,3))
par(las=2)
barplot(d1$valor,
  main = 'Distribuição de Valor Transações por Operação',
  cex.main = 1.2,
  yaxp=c(0, max(d1$valor), 10),
  col = hcl.colors(5, palette = "Blue-Red 3"),
  xlab = "")
legend("topleft", legend = c(
  'Depósito em Dinheiro', 'Recebimento de Banco',
  'Remessa', 'Saque Cartão de Crédito', 'Saque em Dinheiro'),
  fill = hcl.colors(5, palette = "Blue-Red 3"))

```

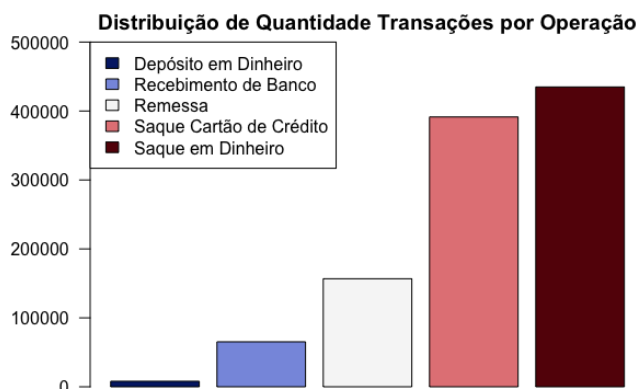


Figura 12 - Qtd de transações por tipo

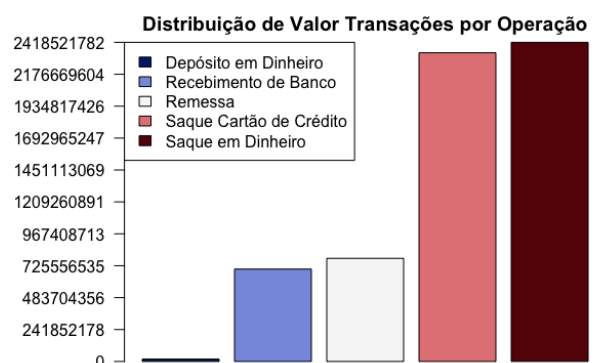
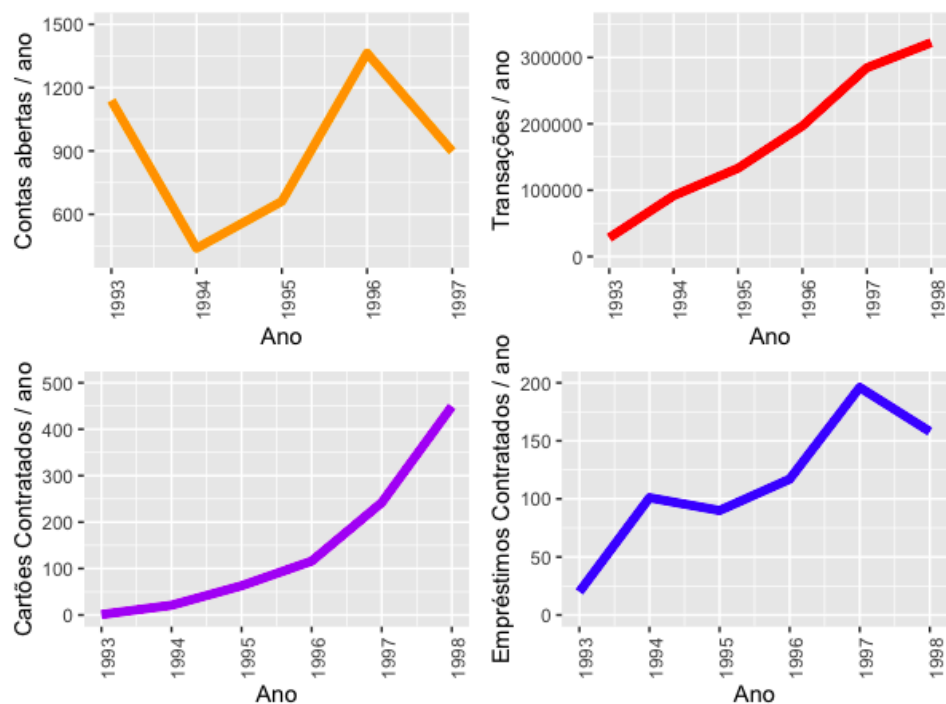


Figura 13 - Valor das transações por tipo

Os dados também permitiam analisar como foi a evolução da carteira de produtos ao longo dos anos. A base de dados analisada trouxe informações referentes aos anos de 1993 até 1998. Eles foram trabalhados da seguinte forma:

```
#DESEMPENHO DO BANCO NO DECORRER DOS ANOS
d1 <- group_by(.data = data_account, data_abertura=year(date)) %>% summarise(contas_abertas=n()) %>% arrange(data_abertura)
d1
d2 <- group_by(.data = data_trans, ano_transacao=year(date)) %>% summarise(num_transacoes=n()) %>% arrange(ano_transacao)
d2
d3 <- group_by(.data = data_loan, ano_emprestimo=year(date)) %>% summarise(num_emprestimos=n()) %>% arrange(ano_emprestimo)
d3
d4 <- group_by(.data = data_card, ano_cartao=year(issued)) %>% summarise(num_cartoes=n()) %>% arrange(ano_cartao)
d4
lsize <- 2.0
plot_contas_ano <- d1 %>% ggplot(aes(x = data_abertura, y = contas_abertas)) +
  geom_line(color = "orange", size = lsize) +
  labs(x = "Ano", y = "Contas abertas / ano") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE),
    limits = c(400, 1500)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
    legend.position = "none")
plot_transacoes_ano <- d2 %>% ggplot(aes(x = ano_transacao, y = num_transacoes)) +
  geom_line(color = "red", size = lsize) +
  labs(x = "Ano", y = "Transações / ano") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE),
    limits = c(0, 350000)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
    legend.position = "none")
plot_loans_ano <- d3 %>% ggplot(aes(x = ano_emprestimo, y = num_emprestimos)) +
  geom_line(color = "blue", size = lsize) +
  labs(x = "Ano", y = "Empréstimos Contratados / ano") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE),
    limits = c(0, 200)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
    legend.position = "none")
plot_cards_ano <- d4 %>% ggplot(aes(x = ano_cartao, y = num_cartoes)) +
  geom_line(color = "purple", size = lsize) +
  labs(x = "Ano", y = "Cartões Contratados / ano") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE),
    limits = c(0, 500)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
    legend.position = "none")
grid.arrange(plot_contas_ano, plot_transacoes_ano,
  plot_cards_ano, plot_loans_ano,
  nrow = 2, ncol = 2)
```



14: Evolução das operações por tipo e ano

Existem informações bastante relevantes para a tomada de decisões do banco. Os principais insights deste estudo apontam:

- Um pico de contas abertas em 1996, e um declínio no período seguinte, o que pode indicar a perda de espaço para concorrentes;
- Um volume de transações crescente desde 1993, com pico no ano do último dado analisado, o que infere uma base de clientes fiéis e cada vez mais bancarizados;
- Um volume de crescimento no número de cartões de crédito emitidos, o que indica uma forte tendência de crescimento do segmento e, portanto, uma oportunidade para realizar ações rentáveis para o banco com este produto;
- Um volume histórico ainda grande de empréstimos concedidos, embora haja um potencial bastante elevado de crescimento deste produto, já que o nível de endividamento total dos clientes já foi demonstrado como baixo.

Por fim, a base de dados order faz referência basicamente as operações de remessa/transferências que o banco fez para outros bancos. Fizemos uma distribuição por quantidade de transferências e também por valor. Foi identificado um número considerável de transações, no qual não se atribuiu a uma ou mais categorias específicas.

```
d1 <- data_order %>% group_by(k_symbol) %>%  
  summarise(count=n(), valor_total=sum(amount)) %>% arrange(count)  
d1  
par(las=2)  
par(mar=c(9,6,6,2))  
bar <- barplot(d1$count,  
  main = 'Distribuição de Qtd. de Transferências por Tipo.',  
  cex.main = 1.2,  
  names.arg = c('Leasing','Seguros', 'Empréstimos', 'NA', 'Pagto. Diversos'),  
  col = hcl.colors(5, palette = "Tems"),  
  ylab = "",  
  xlab = "")  
text(bar, 100, paste(d1$count), cex=0.8)  
legend("topleft", legend = c(  
  'Leasing','Seguros', 'Empréstimos',  
  'NA', 'Pagto. Diversos'), fill = hcl.colors(5, palette = "Tems"))  
  
barplot(d1$valor_total,  
  main = 'Distribuição de Valor de Transferências por Tipo.',  
  cex.main = 1.2,  
  names.arg = c('Leasing','Seguros', 'Empréstimos', 'NA', 'Pagto. Diversos'),  
  col = hcl.colors(5, palette = "Tems"),  
  ylab = "",  
  xlab = "")
```

Distribuição de Qtd. de Transferências por Tipo

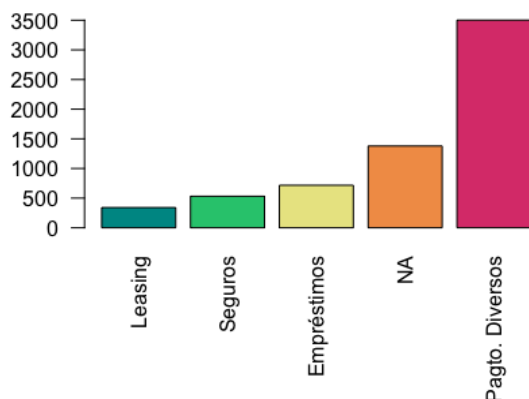


Figura 15 – Dist. de quantidade de transferências por tipo

Distribuição de Valor de Transferências por Tipo.

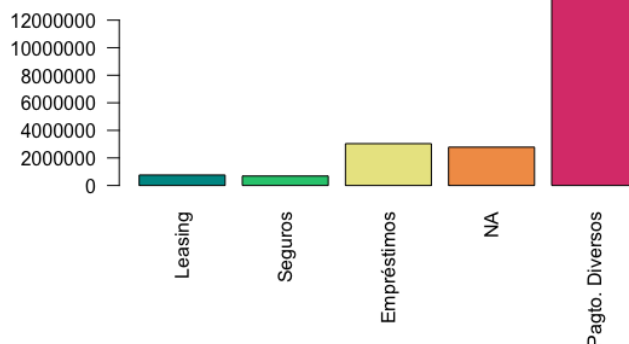


Figura 16 – Dist. De Valor das transações por tipo

5. Conclusões

Com as análises verificadas acima, percebe-se que existem oportunidades de negócio relevantes para o banco que podem ser exploradas. O objetivo da instituição era verificar como estavam suas bases de bons e maus pagadores, e percebe-se que há um espaço para concessão de novos financiamentos, uma vez que os índices de inadimplência da instituição parecem bastante controlados.

Há uma demanda crescente por cartões de crédito por parte de seus clientes, então uma forma de manter essa base satisfeita com os produtos ofertados nesse segmento, bem como atrair novos clientes, seria a criação de programas de recompensa ou de benefícios exclusivos para os detentores de cartões.

O banco conta com mais pagadores adimplentes do que maus clientes. Isso quer dizer que o nível de aversão ao risco da instituição pode ser baixo, e com isso promover ações de crédito e de financiamento com oferta de valores maiores que os atuais.