



**Business Intelligence**

PUC  
RIO

*Diego dos Santos Mello*

*Análise de agrupamentos para identificação de características de clientes inadimplentes*

*Monografia de Final de Curso*

*10/08/2017*

***Monografia apresentada ao Departamento de Engenharia Elétrica da PUC/Rio como parte dos requisitos para a obtenção do título de Especialização em Business Intelligence.***

***Orientadores:***

***Leonardo Mendonça***

---



## Dedicatória

Dedico esta monografia aos meus colegas de trabalho que me ajudaram com conhecimento e opiniões pertinentes. Aos professores e colegas de classe que sempre incentivaram e ajudaram com material didático excelente, sempre ajudando com as dúvidas que surgiam ao longo do trabalho. A minha família que sempre esteve ao meu lado nessa longa caminhada e me deu todo suporte emocional para conseguir terminar este trabalho.

## Agradecimentos

Agradeço ao meu supervisor por me incentivar com os dados e conhecimento necessário para finalizar esta etapa da minha vida. A minha avó que sempre esteve presente no decorrer da criação desta monografia. Aos meus pais que ajudaram a possibilitar o fechamento de mais um ciclo na minha formação.

## RESUMO

Esta monografia apresenta um modelo de agrupamento para identificar comportamentos de inadimplência de clientes de uma base de faturas de telefonia fixa. Neste trabalho primeiro faremos uma sintetização dos dados que originalmente estavam em duas tabelas distintas e com a utilização de linguagem de programação conseguimos juntar essas informações e classifica-las de forma que facilite o uso da mesma. Na etapa seguinte faremos uma classificação e agrupamento desses clientes em grupos distintos para poder inclusive oferecer regras de negócio específicas para cada tipo de cliente. E no final do trabalho poderemos identificar características do comportamento de inadimplência desses clientes baseado em uma amostra da base de dados.

## ABSTRACT

This monograph presents a grouping model to identify customer default behavior from a fixed telephony invoice base. In this work we will first synthesize the data that originally were in two distinct tables and with the use of programming language we were able to gather this information and classify them in a way that facilitates its use. In the next step we will classify and group these clients into distinct groups so that we can even offer different business rules for each type of customer. And at the end of the work we will be able to identify characteristics of the default behavior of these clients based on a sample of the database.

## Sumário

1.	INTRODUÇÃO.....	7
1.1.	MOTIVAÇÃO .....	7
1.2.	OBJETIVOS DO TRABALHO .....	7
1.3.	DESCRIÇÃO DO TRABALHO .....	8
1.4.	ORGANIZAÇÃO DA MONOGRAFIA .....	8
2.	DESCRIÇÃO DO PROBLEMA .....	8
3.	METODOLOGIAS.....	10
3.1.	DESCOBERTA DO CONHECIMENTO.....	10
3.2.	MINERAÇÃO DE DADOS .....	11
3.3.	CLASSIFICAÇÃO .....	12
3.4.	PREPARAÇÃO DE DADOS .....	12
3.5.	NORMALIZAÇÃO .....	12
3.6.	CLUSTERIZAÇÃO.....	12
3.7.	K-MEANS .....	13
3.8.	MAPAS AUTO AJUSTÁVEIS DE KOHONEN.....	13
4.	ARQUITETURA DO SISTEMA PROPOSTO.....	14
5.	CONCLUSÕES.....	17
6.	REFERÊNCIAS BIBLIOGRÁFICAS .....	19

## **LISTA DE FIGURAS**

<b>Figura 1- O processo de descoberta do conhecimento .....</b>	<b>11</b>
<b>Figura 2 - Plot de nr_dias_atraso x valor da fatura (soma) .....</b>	<b>15</b>
<b>Figura 3 - faixa_de_renda_estimada x valor da fatura (soma) .....</b>	<b>16</b>
<b>Figura 4- Training process utilizando SOM kohonen .....</b>	<b>16</b>
<b>Figura 5- Codes plot utilizando clusterização por SOM .....</b>	<b>17</b>

## **LISTA DE TABELAS**

<b>Tabela 1- Clusterização em 3 clusters por kmeans .....</b>	<b>14</b>
<b>Tabela 2- Comparação de kmeans 3 clusters com a variável de inadimplência .....</b>	<b>15</b>
<b>Tabela 3 - Categorização dos grupos de clientes inadimplentes. ....</b>	<b>18</b>

# 1. INTRODUÇÃO

Os cenários de inadimplência geram perda de receita para a companhia, o objetivo deste trabalho é demonstrar que técnicas de Data mining podem ser justificadas no quesito financeiro e criar modelos analíticos para melhorar os processos operacionais de entrada de clientes na base e articular melhor as regras de crédito visando diminuir a entrada de clientes possivelmente inadimplentes.

Na maioria das vezes a taxa de abandono dos clientes está relacionada a fatores externos como falta de dinheiro para pagar as contas devido a outros problemas financeiros (como perda de emprego). Mas também existem os clientes que abandonam voluntariamente, pois são clientes fraudadores que não tem intenção de realmente pagar a fatura para a empresa.

Na telefonia fixa existem regulamentações por parte da ANATEL que obrigam as empresas a manterem o serviço de telefone durante um tempo mesmo que o cliente esteja inadimplente, acarretando aumento da perda da receita.

Existem diversos tipos de fraude, a mais comum é a de subscrição, que é quando o cliente usa uma documentação falsa para adquirir um serviço sem a intenção de cumprir com suas obrigações com a empresa. Pelo fato de os fraudadores estarem sempre em busca de novas maneiras de burlar o sistema, os processos contra estes tipos de inadimplência são reativos.

As ações de inadimplência involuntárias são mais fáceis de serem analisadas, classificadas e controladas. Por isso o objetivo deste trabalho é tentar prever estas características na base de clientes da telefonia fixa, criando modelos para reconhecer esses padrões. Para isso vamos usar a metodologia de análise de dados baseada em inteligência de negócios para fazer os agrupamentos e para fazer as previsões usando redes neurais.

No primeiro momento vamos usar linguagem de programação em SQL para transformar os dados de diferentes tabelas em uma fonte única para facilitar o tamanho dos dados tratados. Depois da base de dados pronta vamos classificar e agrupar esses dados de clientes em grupos para identificar algumas características de inadimplência na telefonia fixa.

## 1.1. MOTIVAÇÃO

A motivação deste trabalho se deu pelo fato do autor trabalhar com este tipo de problema e já gerar estes tipos de informação, fazendo agrupamentos e classificações para tentar entender e identificar os comportamentos dos clientes da empresa.

Com a utilização de data mining pode se conseguir informações cruciais para o desenvolvimento de planos de ação para tentar descobrir a causa raiz ou tentar melhorar os indicadores de inadimplência e assim gerar receita para a empresa.

## 1.2. OBJETIVOS DO TRABALHO

O objetivo deste trabalho foi a aplicação de mecanismos de segmentação e classificação dos dados para que possam ser identificadas características de clientes inadimplentes a partir dos modelos de data mining criados, ajudando a minimizar um problema

que afeta todos os segmentos e também influência no aspecto financeiro como a obtenção de receita.

### **1.3. DESCRIÇÃO DO TRABALHO**

O desenvolvimento dessa monografia envolveu três etapas:

A primeira foi a utilização de SQL para criação de uma fonte unificada de dados, utilizando técnicas de data-mining para sintetizar diferentes tabelas numa visão mais simplificada e mais agrupada.

A segunda foi a sintetização da parte teórica de análise de dados e datamining.

E em seguida foi utilização de métodos de clusterização em Rstudio para agrupar os clientes de telefonia fixa.

### **1.4. ORGANIZAÇÃO DA MONOGRAFIA**

Esta monografia está dividida em quatro capítulos adicionais, descritos a seguir:

O capítulo 2 apresenta um pouco do universo da análise de crédito mostrando conceitos sobre clientes inadimplentes e como isto afeta as empresas.

O capítulo 3 apresenta a parte teórica do trabalho, explicando sobre análise de dados e métodos de agrupamento.

O capítulo 4 detalha os resultados obtidos com a utilização de métodos de agrupamento em cima de uma base de clientes de telefonia fixa de uma empresa brasileira de TELECOM.

Finalmente, o capítulo 5 descreve as conclusões do trabalho.

## **2. DESCRIÇÃO DO PROBLEMA**

Vivemos em um mundo onde grandes quantidades de dados são coletadas diariamente. Analisar esses dados é uma necessidade importante. A mineração pode ser vista como resultado da evolução natural da tecnologia da informação.

Neste trabalho vamos estudar o caso de clientes que se inscrevem em um serviço de telefonia fixa por meio de um formulário padrão, neste formulário temos várias variáveis sobre este cliente, com estas variáveis é possível identificar quais clientes possivelmente podem virar inadimplentes. Identificamos esses grupos de clientes por meio de análises feitas em cima de variáveis pontuais, que podem ser regionais, renda, pontuação no serasa entre outras. Na telefonia fixa que é o tema deste trabalho, o maior motivo de negação de proposta de serviço para o cliente é o débito antigo com a empresa ou o nome sujo no serasa.

Depois que o cliente passa pelo filtro da proposta inicial é feito uma série de análises da base de clientes para melhoria das regras de crédito e para descobrir possíveis problemas pontuais em regionais ou pacote de produtos específicos.

Este trabalho visa diminuir a perda de receita relacionada a insolvência e poder oferecer melhores pacotes de telefonia para cada grupo específico.

Para determinar as chaves críticas para ganhar a competição entre empresas, os analistas de crédito devem se concentrar não apenas nos principais fatores competitivos, mas mais importante, sobre os fatores subjacentes que influenciam. Se uma empresa compete com o preço, o fator subjacente será seus custos. O custo mais baixo irá fazer com que a empresa lucre mais. O fornecedor pode melhor reduzir seus preços e manter margens. O custo com a área de cobrança sendo muito alto por não saber como abordar corretamente os clientes pode prejudicar a lucratividade da empresa.

A chave para o sucesso é criar um produto valorizado que os consumidores desejam, precisam, estão dispostos a comprar e podem pagar.

O objetivo deste trabalho é estabelecer um padrão de comportamento dos clientes inadimplentes de uma empresa de telefonia brasileira, por meio de características comuns, hábitos de consumo, frequências de pagamento, tempo que estão na base de clientes, renda domiciliar e valor da fatura. Foram definidas as variáveis mais relevantes e uma janela de tempo para este conjunto de dados que servirá de base para geração de conhecimento sobre o comportamento dos clientes.

O modelo utilizado para identificar padrões de comportamento dos clientes inadimplentes com a empresa foi o agrupamento utilizando k-means e outro modelo de clusterização conhecido como mapas auto ajustáveis de Kohonen. Este tema está ligado diretamente a área financeira da empresa e existe um interesse grande de áreas de marketing, crédito e relacionamento com os clientes. Especialmente em empresas com grande quantidade de dados, ou seja, empresas de grande porte, é preciso a definição de alertas para prever movimentos de inadimplência de certos grupos de clientes específicos. Estes alertas estão associados a criação de indicadores de crédito ou mesmo até eventos de fraude.

Neste estudo o cliente será entendido como um registro de entrada, ou seja, uma fatura que indica uma linha em nossa base de dados. Assim desta forma o conjunto de informações pertencentes a cada um dos clientes da nossa amostra, serão agrupadas por registro. Algumas variáveis tiveram seu conteúdo alterado, numa tentativa de normalizar a distribuição dos seus dados.

O aspecto financeiro pela atual conjuntura do país prejudica os indicadores de inadimplência de todas as empresas brasileiras como um todo, mas a regulamentação do mercado de telefonia também atrapalha a tomada de ações por parte da empresa para tentar conseguir receita em cima desses clientes que não pagaram a fatura. Os modelos desenvolvidos neste trabalho são baseados em métodos com base em reconhecimento de padrões, em modelos de agrupamento e classificação. Este sistema de diagnóstico pode ser utilizado para descobrir conhecimento para identificação de eventos de não pagamento de fatura e assim ajudar a empresa a combater a inadimplência e melhorar os processos operacionais de faturamento e cobrança.

Os aspectos que ajudam a empresa neste relacionamento com o cliente são cadastro, crédito e cobrança. Para cada etapa do processo uma estratégia deve ser desenvolvida. Quanto mais parâmetros você conseguir incluir no planejamento dessas fases, maior qualidade terá o crédito concedido, reduzindo assim os esforços com a cobrança.

O cadastro de clientes visa colher dados que permitirão conhecê-lo melhor e realizar uma análise de crédito consciente utilizando as informações que formam o seu histórico. Um cadastro incompleto ou desatualizado pode causar problemas para a empresa e gerar um índice de inadimplência insustentável.

Uma ficha cadastral deve possuir campos disponíveis a serem preenchidos que sejam suficientes para uma análise de crédito eficaz. O excesso de informações irrelevantes,

muitas vezes, pode tornar a ficha cansativa de se preencher e analisar. Por isso, é importante avaliar quais informações são realmente importantes.

Geralmente, um cadastro acompanhado de cópias de documentos pode trazer mais qualidade à análise de crédito e evitar fraudes. Documentações de identificação, comprovação de renda e endereço, contracheque, declaração de Imposto de Renda, escritura de imóveis e até mesmo o IPVA são importantes, que ajudam a determinar o poder de compra que o cliente tem.

A concessão de crédito é o ápice de uma transação. Significa que já houve uma análise e foi determinado qual o limite de compra pode conceder. Nesse processo, é sempre bom lembrar que crédito nunca pode ser concedido sem uma análise apenas por ser um cliente conhecido ou que aparenta ser financeiramente equilibrado. Mesmo que ele seja antigo na base pode ser que ele venha a ser inadimplente adquirindo outros serviços da empresa. Dessa forma, é possível evitar riscos de inadimplência na operação.

Um limite de crédito concedido diferentemente para cada cliente é o valor que a empresa está disposta a correr o risco para não perder uma venda. Por isso, além da análise de crédito detalhada com informações atuais, é preciso deixar o cliente ciente das políticas de crédito antes da análise, para que não haja constrangimentos em uma possível rejeição do pedido.

Mesmo com uma ficha de cadastro corretamente preenchida e uma análise de crédito bem-feita, o risco de clientes pagarem em atraso é grande. Para não colocar a saúde financeira da empresa em risco quando isso ocorrer, alguns procedimentos devem ser colocados em prática rapidamente.

Nessa etapa, é preciso definir de forma antecipada os critérios que serão utilizados para realizar uma cobrança eficiente, tendo sempre muito respeito com o cliente e com o Código de Defesa do Consumidor.

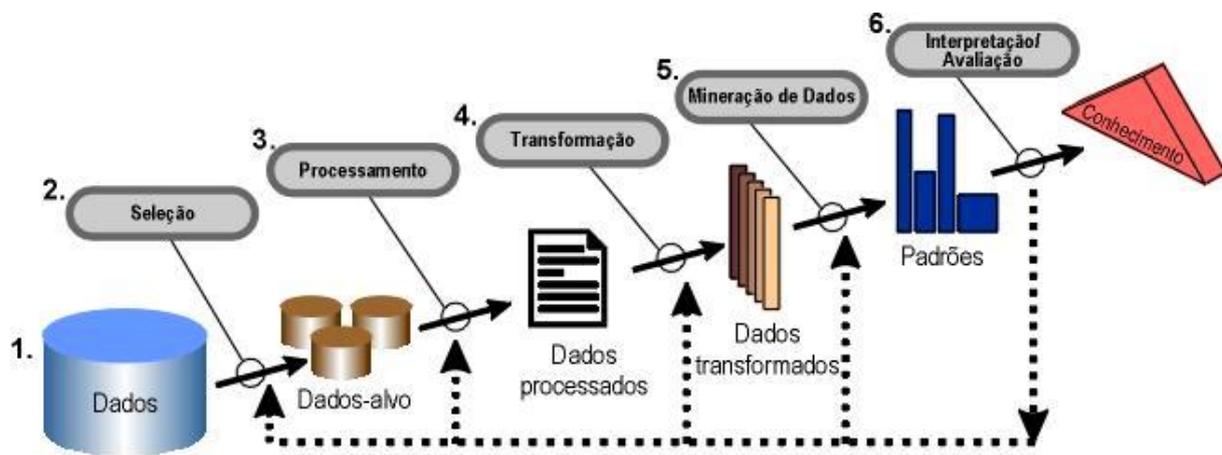
O tempo gasto com o preenchimento de uma ficha cadastral (formulário) permitirão uma análise de crédito eficiente e uma concessão com qualidade, além de reduzir os riscos de inadimplência da empresa.

### **3. METODOLOGIAS**

Neste capítulo, serão apresentados conceitos de data mining e descoberta do conhecimento. As aplicações de data mining têm grande utilização nas empresas, pois hoje em dia existe uma grande massa de dados para ser analisada, portanto precisa-se de ferramentas que consigam extrair informações desse volume extenso de dados para a tomada de decisão das ações que a empresa irá tomar em relação a diversos setores.

#### **3.1. *DESCOBERTA DO CONHECIMENTO***

A Descoberta de Conhecimento foi definida como a “extração não trivial e implícita de informações anteriormente desconhecidas e potencialmente úteis de dados”. Isto é um processo do qual a mineração de dados forma apenas uma parte, embora central.



Uma visão geral das etapas que compõem o processo de KDD

Figura 1- O processo de descoberta do conhecimento

Os dados entram, possivelmente de várias fontes. Está integrado e situado em alguns data marts. A partir disso é então pré-processada em um formato padrão. Estes "dados preparados" são então passados para um algoritmo de mineração de dados que produz uma saída na forma de regras ou algum outro tipo de "padrões". Estes são então interpretados para gerar conhecimento novo e potencialmente útil.

### 3.2. MINERAÇÃO DE DADOS

Data Mining, ou mineração de dados, é o processo de descoberta de padrões e tendências existentes em repositórios de dados (GROTH, 2000). Embora os algoritmos de mineração de dados sejam fundamentais para o conhecimento, o pré-processamento dos dados e a interpretação dos resultados são ambos importantes. São tarefas qualificadas que são muito mais uma arte (ou uma habilidade aprendida da experiência) do que uma ciência exata, embora ambos sejam abordados.

Aplicações de Data Mining

- Análise de compostos orgânicos
  - Resumo automático
  - Detecção de fraude de cartão de crédito
  - Previsão de carga elétrica
  - Previsão financeira
  - Diagnóstico médico
  - Previsão de participação no público da televisão
  - Design de produto
  - Avaliação imobiliária
- e muito mais.

### **3.3. CLASSIFICAÇÃO**

A classificação é uma das aplicações mais comuns para a mineração de dados. Isso corresponde a uma tarefa que ocorre frequentemente na vida cotidiana. Por exemplo, um hospital pode querer classificar pacientes médicos para aqueles que estão em alto, médio ou baixo risco de adquirir uma determinada doença, uma assembleia de opinião pode desejar classificar as pessoas entrevistadas para aqueles que são susceptíveis de votar em um partido político ou estão indecisos. Essencialmente envolve dividir objetos para que cada um seja atribuído a um número de categorias mutuamente exaustivas e exclusivas, conhecidas como classes. O termo “mutuamente exaustivo e exclusivo” significa simplesmente que cada objeto deve ser atribuído precisamente a uma classe, ou seja, nunca mais do que uma. Muitas tarefas práticas de tomada de decisão podem ser formuladas como classificação, ou seja, atribuir pessoas ou objetos a uma das várias categorias.

### **3.4. PREPARAÇÃO DE DADOS**

Para muitas aplicações, os dados podem simplesmente ser extraídos de um banco de dados talvez usando um método de acesso padrão como o ODBC. No entanto, para algumas aplicações, a tarefa mais difícil pode ser a de obter os dados em uma forma padrão em que ele possa ser analisado. Em algumas áreas de aplicação de mineração de dados, a disponibilidade de armazenamento é cada vez maior. As empresas têm capacidade hoje em dia de armazenar informações sobre todas as compras feitas por um cliente de supermercado por três meses ou uma grande quantidade de detalhes e informações sobre cada paciente em um hospital. Para alguns conjuntos de dados, pode haver substancialmente mais atributos do que necessário. Por isso é necessário também uma redução no número de atributos e retirar possíveis outliers que sejam tendenciosos e atrapalhem o modelo.

### **3.5. NORMALIZAÇÃO**

Um grande problema ao usar a fórmula de distância euclidiana (e muitos outras medidas de distância) é que os grandes valores frequentemente inundam os pequenos. Para superar este problema geralmente normalizamos os valores. A ideia é fazer com que os valores de cada atributo sejam executados de 0 a 1. Suponha que, para algum atributo A o menor valor encontrado nos dados de treinamento é -8,1 e o maior é 94,3. Primeiro ajustamos cada valor de A adicionando 8,1, então os valores agora são executados de 0 a  $94,3 + 8,1 = 102,4$ . A disseminação de valores do maior para o mais baixo é agora 102,4 unidades, então dividimos todos os valores desse número para fazer a propagação de valores de 0 a 1.

### **3.6. CLUSTERIZAÇÃO**

A Clusterização é preocupada em agrupar objetos que sejam semelhantes entre si e diferente dos objetos pertencentes a outros clusters. Existem muitos algoritmos para agrupar. Vamos descrever dois métodos para que a semelhança entre objetos se baseia em uma medida da distância entre eles.

### 3.7. K-MEANS

Proposto por J. MacQueen em 1967, este é um dos algoritmos mais conhecidos e utilizados, além de ser o que possui o maior número de variações (DINIZ e LOUZADA NETO, 2000). K-means é um algoritmo de cluster exclusivo. Cada objeto é atribuído precisamente a um dos conjuntos de clusters. Para este método de agrupamento, começamos por decidir quantos clusters nós gostaríamos de formar a partir de nossos dados. Chamamos esse valor de  $k$ . O valor de  $k$  é geralmente um inteiro pequeno, como 2, 3, 4 ou 5, mas pode ser maior. Há muitas maneiras pelas quais  $k$  clusters podem potencialmente ser formados. Nós podemos medir a qualidade de um conjunto de clusters usando o valor de uma função objetivo que consideramos ser a soma dos quadrados das distâncias de cada ponto do centroide do cluster ao qual é atribuído.

Pode se provar que o algoritmo k-means sempre terminará, mas não necessariamente encontrará o melhor conjunto de clusters correspondente a minimizar o valor da função objetivo. A seleção inicial de centroides pode significativamente afetar o resultado. Para superar isso, o algoritmo pode ser executado várias vezes para um dado valor de  $k$ , cada vez com uma escolha diferente dos  $k$  centroides iniciais, sendo então escolhido o conjunto de clusters com o menor valor da função objetivo. A desvantagem mais óbvia deste método de agrupamento é que não existe um princípio para saber qual valor  $k$  deve ter.

O algoritmo de *k-means* é bastante escalar e confiável, porém apresenta alguns problemas. Os dois principais problemas são:

- Exige que as variáveis sejam numéricas ou binárias (HUANG, 1997a).

Freqüentemente aplicações envolvem dados categorizados, neste caso, uma alternativa é converter os dados categorizados em valores numéricos.

- É sensível a valores *outliers*, um único objeto com valor muito extremo pode modificar, substancialmente, a distribuição dos dados (HAN e KAMBER, 2001).

### 3.8. MAPAS AUTO AJUSTÁVEIS DE KOHONEN

Os mapas de Teuvo Kohonen se caracterizam por serem uma das mais utilizadas em modelos não-supervisionados, em especial, naqueles voltados para a identificação de clusters característicos.

Os mapas aprendem por conta própria através de uma aprendizagem competitiva e sem supervisão. O mapa de Kohonen é criado utilizando-se técnicas baseadas em redes neurais. "Mapa" é porque eles tentam mapear seus pesos para se adequarem ao dado. Pode-se dizer que o SOM reduz as dimensões dos dados e exibe semelhanças entre os dados.

Os dados de entrada são um princípio fundamental dos mapas de Kohonen e uma das coisas que os torna tão valiosos. Os dados são preservados quando mapeados para uma rede de Kohonen. Isso tem um valor pragmático de representar dados complexos.

Um conjunto de vetores é utilizado como entrada para o mapa repetidas vezes, os quais são denominados unidades. Cada unidade é associada a um vetor de peso, o qual é inicialmente consistido por valores aleatórios. As unidades reagem, mais ou menos, aos vetores de entrada, de acordo com a correlação entre o vetor de entrada e o vetor de peso das unidades.

A unidade que possui a maior resposta para determinada entrada é indicada para o aprendizado, assim como, algumas outras unidades na "vizinhança". Essa vizinhança decresce

em tamanho durante o período de treinamento. O aprendizado é realizado pelo ajuste dos pesos das unidades, sempre por meio de pequenas quantidades, com o objetivo de tornar cada vez mais semelhante os valores das unidades com os valores do vetor de entrada.

O resultado do treinamento é a identificação no mapa de um padrão de organização. Unidades diferentes aprendem a responder a diferentes vetores no conjunto de entrada, e as unidades mais próximas fisicamente uma das outras tendem a responder aos vetores de entrada que mais se assemelham um ao outro. Quando o treinamento é finalizado, o conjunto dos vetores de entrada é aplicado ao mapa uma vez mais, realizando uma marcação em cada vetor de entrada, de qual unidade responde mais fortemente àquele vetor de entrada.

## 4. ARQUITETURA DO SISTEMA PROPOSTO

As empresas constroem um relacionamento com o cliente, este relacionamento precisa ter diálogo, confiança e reconhecimento. A empresa precisa entender melhor as necessidades de seus clientes pois muitas vezes nem mesmo o cliente sabe o que quer. Por isso é necessário desenvolver estratégias desenvolvidas em relação aos dados que a empresa detém sobre seus clientes. O conjunto de dados analisados foi uma amostra de 4 meses de uma base de dados de clientes de telefonia fixa, do mês de janeiro de 2016 até abril de 2016. A base de dados contém 227906 registros e selecionamos 5 variáveis para o estudo de caso. Foram removidos 30 outliers que tinham as faturas acima de R\$1000.

- A primeira variável é o flag\_90 que determina se o cliente já ficou entre 60 e 90 dias sem pagar uma fatura, os valores são 0 para adimplente e 1 para inadimplente.
- A segunda variável é a faixa de tempo que o cliente está na empresa variando 0 a 7, sendo que 0 é um cliente novo na base e 7 é um cliente com mais de 7 meses de base.
- A terceira variável é o número de dias de atraso, são quantos dias o cliente está em atraso com a fatura da empresa podendo chegar até 90 dias de atraso.
- A quarta variável é a faixa de renda do cliente, variando de 500 a 10000 reais.
- A quinta variável escolhida foi o valor da fatura em reais.

Para fazer a clusterização foi utilizado o software R studio e o pacote “cluster”. Neste trabalho foi utilizado primeiro o agrupamento pelo método kmeans separando em 3 grupos e utilizando nstart de 20, ou seja, isso significa que R tentará 20 atribuições de partida aleatórias diferentes e, em seguida, selecionará uma com a menor variação de cluster. Os grupos ficaram bem definidos.

	FAIXA_TEMPO_BASE	nr_dias_atraso	faixa_renda_estimada	soma
1	1.5766594	15.90160	3971.443	88.52244
2	0.6393668	22.85070	1206.085	87.99336
3	1.6582278	17.60077	9444.370	87.11875

Tabela 1- Clusterização em 3 clusters por kmeans

Os clusters ficaram definidos assim em quantidade:

K-means clusterização com 3 clusters de tamanhos (em clientes) 14147, 207754, 6004.

Comparando os clusters com situação de inadimplência temos:

Tabela	0	1	%
1	12207	1940	15,9%
2	158588	49166	31,0%
3	5094	910	17,9%

Tabela 2- Comparação de kmeans 3 clusters com a variável de inadimplência

O cluster mais antigo de tempo de base (cluster 3) tem uma taxa de 17,9% de inadimplência, o que reflete a informação da tabela 1 onde mostra um número de dias de atraso entre o cluster 1 que é a menor taxa de 15,9% de inadimplência e o cluster 2 que tem taxa de inadimplência de 31%. O cluster 2 é o maior de todos com quase a base inteira de clientes se firmando nele. O cluster 3 como tem uma concentração de renda alta, tem a menor concentração de clientes.

1. O primeiro cluster caracteriza-se por clientes com 1 mês e meio de base, o menor número de dias de atraso da fatura dentre os 3 grupos, a faixa de renda mediana levando em consideração que a maior faixa é R\$10000 e o valor da fatura é o maior dos 3 grupos.
2. O segundo cluster ficou definido com os clientes que são relativamente novos na base, ou seja, isso demonstrou um fator que já imaginávamos, eles são os que tem o maior número de dias de atraso, e isso também se confirmou com a renda baixa deste cluster apesar do valor da fatura ser menor em relação ao cluster 1.
3. O terceiro cluster é o mais antigo em relação ao tempo de base, apesar de ser o cluster mais antigo o número de dias de atraso é maior que o do primeiro cluster apenas 1,7 dias aproximadamente. A renda estimada quase atinge a maior faixa que é de R\$10.000 e o valor da fatura é o menor dentre os 3 clusters.

Foi feito um plot de nr\_dias\_atraso x valor da fatura (soma) podemos ver que com mais de 50 dias de atraso o cluster 2 que é o maior tem muitas faturas com valores altos.

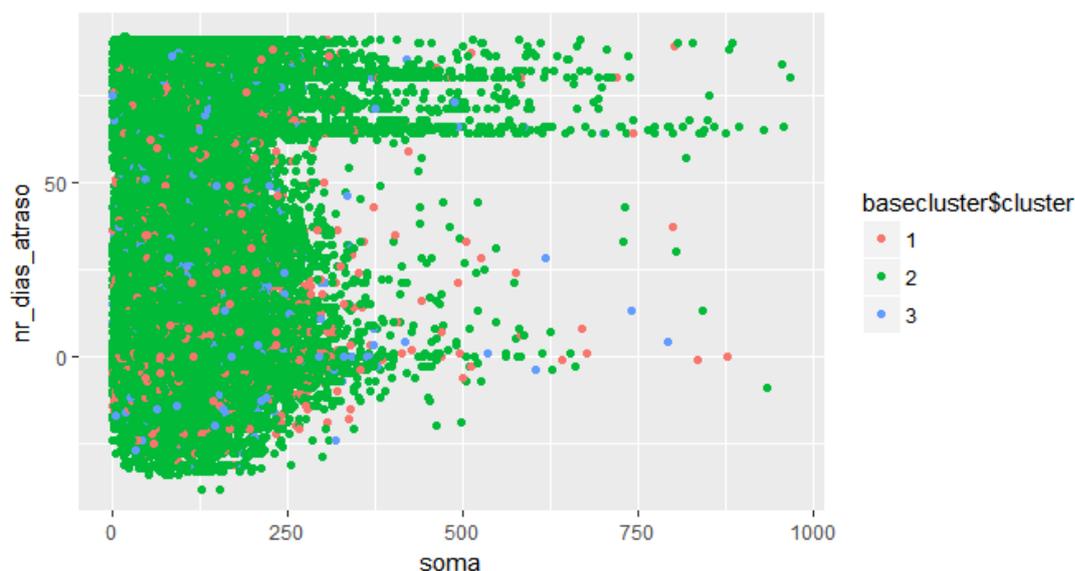


Figura 2 - Plot de nr\_dias\_atraso x valor da fatura (soma)

Neste outro plot foi colocado faixa\_de\_renda\_estimada x valor da fatura (soma) conseguimos ver claramente a divisão dos clusters por renda.

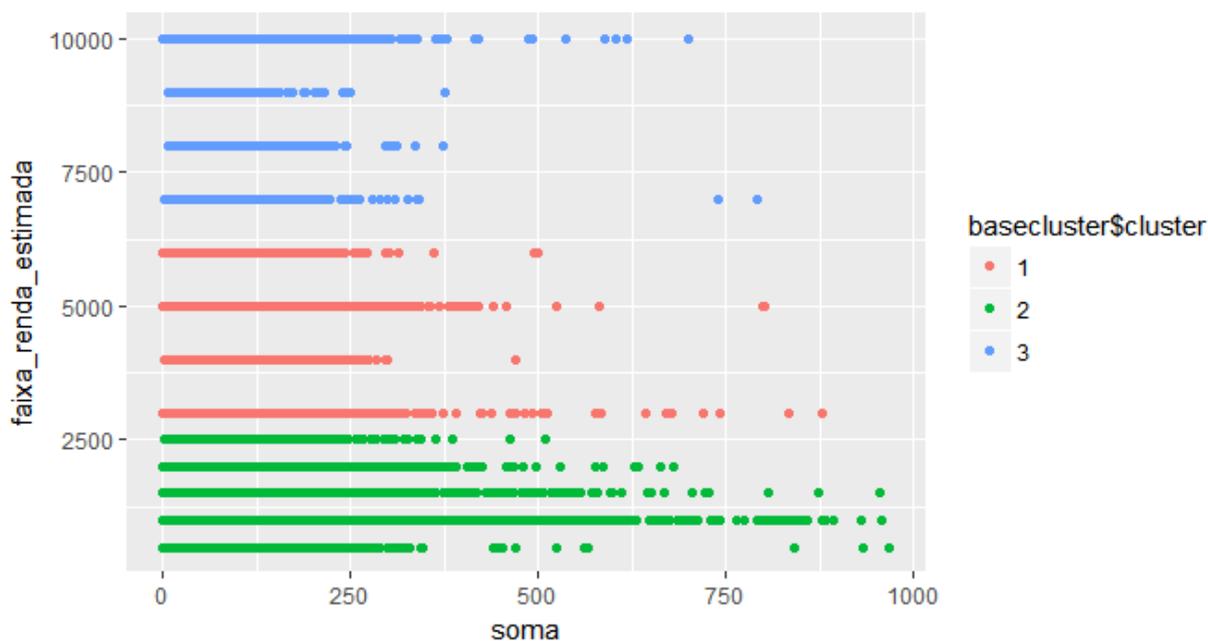


Figura 3 - faixa\_de\_renda\_estimada x valor da fatura (soma)

Em seguida foi feito o agrupamento por mapas auto ajustáveis de Kohonen para comparar com o kmeans.

Fizemos a normalização da base, logo após rodamos o kohonen com 100 iterações e a taxa de aprendizado utilizada foi de 0,05 decrescendo até 0,01.

Neste primeiro gráfico do training process podemos ver que em pouco mais de 40 iterações as distancias entre os nódulos estavam muito próximas.

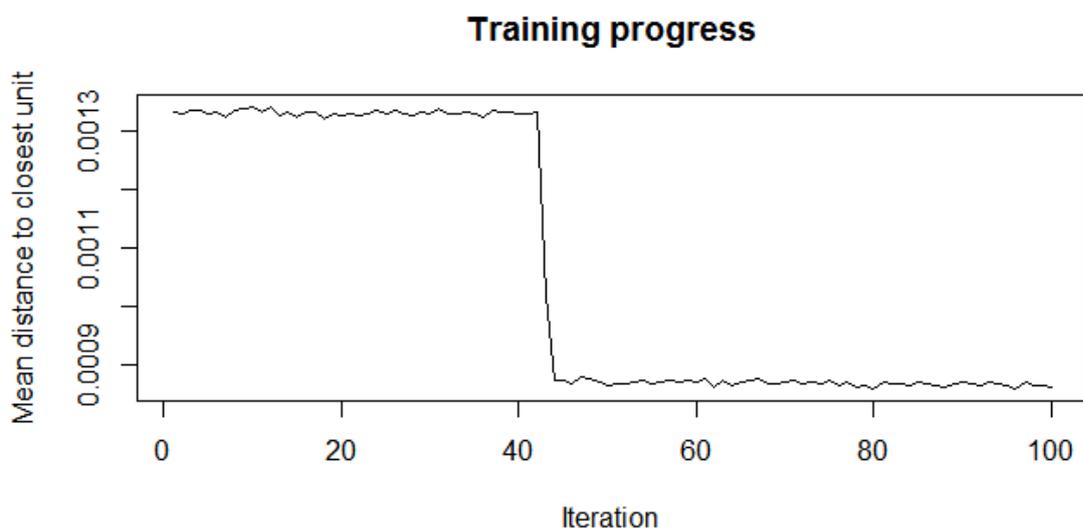
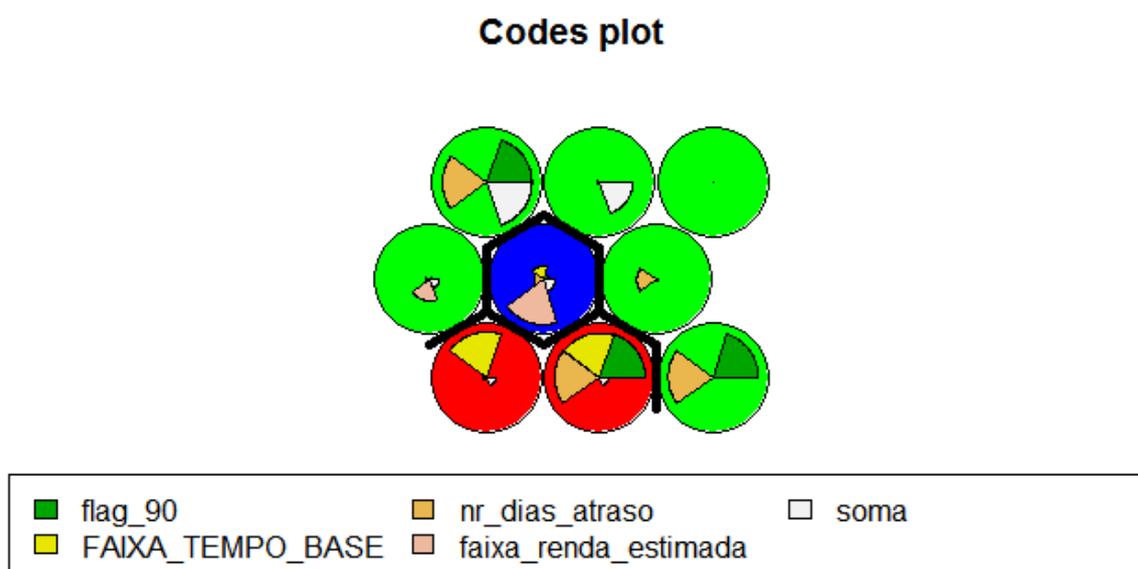


Figura 4- Training process utilizando SOM kohonen

Neste gráfico pode-se reparar que temos 3 clusters, sendo o cluster 3 seria o azul o qual é o menor, o cluster vermelho seria o cluster 1 e o cluster 2 seria o verde.



**Figura 5- Codes plot utilizando clusterização por SOM**

## 5. CONCLUSÕES

O trabalho foi desenvolvido com objetivo de utilizar os métodos de agrupamento K-means e mapas auto organizáveis de Kohonen para identificar características de clientes inadimplentes. O desenvolvimento de modelos de classificação permite a criação de conhecimento sobre o comportamento de clientes da empresa, sendo identificados padrões de características de clientes que podem ser agrupadas, permitindo que a empresa crie ações de relacionamento específicas como ofertar tipos de serviços e planos de telefonia mais baratos ou métodos de parcelamento da fatura para facilitar o pagamento.

Por outro lado, a empresa também pode oferecer descontos para a função de debito automático se prevenindo dessas faturas inadimplentes, pois muitos clientes podem esquecer de pagar a conta. Cada grupo identificado pode ter uma abordagem distinta pois os clientes têm características particulares. Esta inteligência analítica possibilita criação de regras de negócio específicas para cada cluster de cliente fazendo que a empresa possa alcançar resultados mais efetivos com menor custo e esforço.

Essas classificações da base de clientes podem ajudar a empresa a se antecipar em determinados acontecimentos. O modelo de segmentação comportamental da inadimplência identificou 3 grupos, separando os clientes em classes distintas. Apesar dos valores das faturas nos 3 grupos estarem equilibradas, pois no caso da telefonia fixa, hoje em dia não é mais tão utilizado quanto a telefonia móvel, portanto os valores das faturas não variam tanto em relação

aos grupos separados. Mas conseguimos separar o grupo por tempo de base, onde percebemos que a maioria dos clientes de nossa amostra são clientes novos e, portanto, que tem maior chance de serem inadimplentes pois não tem um relacionamento longo com a empresa podendo inclusive ser uma fraude, este seria o grupo classificado como péssimo.

O grupo com característica bem marcante de uma elevada quantidade de renda seria qualificado como o grupo dos moderados, pois apesar de terem a maior taxa de renda e serem o grupo mais antigo de clientes eles não tem a melhor performance no pagamento em relação ao tempo de atraso do pagamento da fatura.

E o outro grupo que se caracterizou por ter o menor número de dias de atraso no pagamento da fatura, ou seja, são os que resolvem mais rápido a situação de inadimplência com a empresa apesar de terem uma renda mediana em relação a escala total de renda entre zero e mil reais, seria o grupo dos bons.

Cluster	Classe	Resumo
1	Bom	Possui alto valor de fatura com inadimplência baixa (15,9%), e com o menor período de atraso médio. É a segunda maior população, com a renda média.
3	Moderado	Possui menor valor de fatura, com inadimplência média (17,9%), baixo tempo de atraso e a menor população e possui a renda mais alta.
2	Ruim	Possui médio valor de fatura e muito alto valor de inadimplência (31%), com o maior período de atraso médio, tendo a maior população e a maior inadimplência.

**Tabela 3 - Categorização dos grupos de clientes inadimplentes.**

O modelo de segmentação de todos os clientes da empresa de telefonia permitiu que fosse identificado o valor de cada cliente para a empresa possibilitando a definição de ações de relacionamentos mais eficientes. Cada grupo característico possui da mesma forma um comportamento particular no uso dos serviços da empresa e estabelecem eventos de pagamento das contas de telefone de formas distintas.

Para estes 3 grupos podemos criar produtos mais aderentes, ações de relacionamento mais focadas e gestão das atividades de cobrança e faturamento melhor articuladas. Os modelos de classificação permitem a criação de uma inteligência de negócio que capacita a empresa a tomar ações em relação aos clientes da empresa. Os clientes com características de pouco tempo na base teriam que ter regras de negócio que ofereçam maiores dificuldades no momento de adquirir um produto de telefonia fixa além de direcionar ações de cobrança para esses clientes que tem mais chance de se tornar inadimplentes proporcionando um menor custo na realização das ações e maior retorno em termos de receita.

A criação de políticas diferenciadas, de acordo com os grupos identificados, propicia a geração de uma régua de cobrança, contendo ações distintas para determinados períodos de atraso e de acordo com a classe de cliente identificada. Estas políticas podem trazer resultados satisfatórios não apenas da redução de despesas de cobrança, mas no processo de recuperação de receita como um todo.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ANDRITSOS, Periklis. Data Clustering Techniques. Canadá, Toronto, 2002.

BERRY, Michael J. A.; LINOFF, Gordon. Data Mining Techniques: For Marketing, Sales, and Customer Support. New York: Wiley Computer Publishing, 1997.

BUSSAB, Wilton de O.; MIAZAKI, Édina S; ANDRADE, Dalton F. de. Introdução à Análise de Agrupamentos. 9o Simpósio Nacional de Probabilidade e Estatística. São Paulo: ABE, 1990.

DINIZ, Carlos Alberto; LOUZADA NETO, Francisco. **Data Mining**: uma introdução. São Paulo: ABE, 2000.

GROTH, Robert, Data Mining – Building Competitive Advantage, Prentice Hall PTR, 2000.

HAN, Jiawei; KAMBER, Micheline. **Data Mining**: Concepts and Techniques. San Diego: Academic Press, 2001.

HOLLMÉN, Jaakko, User profiling and classification for fraud detection in mobile communications networks, Dissertation for the degree of Doctor of Science in Technology, University of Technology, Department of Computer Science and Engineering, Laboratory of Computer and Information Science, Helsinki, 2000.

HUANG, Zhexue. Clustering Large Data Sets with Mixed Numeric and Categorical Values. In: **Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'97)**. Singapore, 1997

A seguir, é apresentado o programa em código de programação SQL para criação da base, em formato SQL, utilizada na construção dos modelos de agrupamento e classificação do presente trabalho.

WITH

MAX\_PROPOSTA\_NOVO AS

(select ad\_tax\_id as cpf,

ap\_application\_date,

f.score\_cco1\_var,

f.cp\_classification\_code,

f.spare\_txt\_04,

```

f.ad_addr_postcode as cep,

f.pc_dt_oldest_act_line,

f.pcs_amount_arrears,

f.pcs_days_arrears,

f.faixa_renda_estimada,

f.faixa_rend_dom,

case when ap_application_Date >= '20151014'

    then case when flagbasona = '1' and score_cco1_var = 'NULL' or
score_cco1_var = "

        then dv_serasa_called

        else flagbasona

        end

    else dv_serasa_called

    end as alavancado,

Row_Number() Over(Partition BY ad_tax_id

ORDER BY ap_application_date DESC) as rownumber

from dbo.TB_TRS_OIFULL_FIXA F

left join dbo.TB_CRDT_FPD_VAR_FIXA_3rol A on a.nr_pessoa = f.ad_tax_id

AND datediff( d, cast (F.ap_application_Date as datetime ), cast (
A.DT_ATIVACAO_cadastro as datetime ) ) between 0 and 60

    where OP_FIN_DEC_REASON_CODE LIKE 'AA%' AND
CAST(AD_TAX_ID_TYPE AS int) = 1

),

```

FPD\_FIXA as (

```
select
f.ap_application_date
,cd_mes_vencto as vencto
,FL_90 as flag_90
, de_pdv_crm as nome_pdv
, de_canal_bov as Canal
, de_grupo_canal_bov as Resumo_Canal
, CD_GRUPO_UNIDADE as REGIONAL
, de_plano_emissao as produto
, CD_GRUPO_UNIDADE as UN
,cd_uf_filial_cobranca as UF
,f.score_cco1_var as score
,f.cp_classification_code as class_code
,f.spare_txt_04 as cluster
,f.cep
,f.alavancado
, f.pc_dt_oldest_act_line
,f.pcs_amount_arrears
,f.pcs_days_arrears
,cd_ciclo_fatura
,de_faixa_ativ_vencto
,nr_dias_atraso
, f.faixa_renda_estimada
,f.faixa_rend_dom
,replace(VL_EFETIVO,'!',',') as valor
```

```

from dbo.TB_CRDT_FPD_VAR_FIXA_3ROL

left join MAX_PROPOSTA_NOVO F

on cpf = nr_pessoa

and rownumber = 1

where cd_mes_vencto >= '201601'

and cd_mes_vencto <= '201606'

and fl_indicador = 'S'

and fl_ativacao_bov = 'S'

and cd_grupo_produto_emissao like 'VOZ B%'

)

select

VENCTO

, flag_90

, a.nome_pdv

, Canal

, Resumo_Canal

, case when REGIONAL like 'Var%' then B.CD_REGIONAL

else B.DE_REGIONAL_EMP

end as REGIONAL

, case when produto like 'MINHA LINHA%' then 'MINHA LINHA'

when produto like '%FALE 350%'

or produto like '%FALE 500%'

or produto like '%FALE 1000%' then 'FALE 350, 500 ou 1000'

```

```

when produto like '%CONTROLE%' then 'OI FIXO CONTROLE'

when produto like '%BASICO%'

    or produto like '%FALE 230%' then 'BASICO + FALE 230'

when produto like '%ECONOMIA%' then 'PLANO ECONOMIA'

when produto = 'OI FIXO A VONTADE' then 'OI FIXO A VONTADE'

when produto like '%OI FIXO 200%' then 'OI FIXO 200'

        else produto

end as produto

, UN

,case when alavancado = 'R' then 'Alavancado'

        else 'Nao Alavancado'

        end as alavancado

,cep

,case when convert(bigint,isnull(a.score,0))= 0 or a.score = '' then 'Zero'

        when convert(bigint,a.score) >= 1 and convert(bigint,a.score) <= 100 then '01 - Entre 1 e 100'

        when convert(bigint,a.score) >= 101 and convert(bigint,a.score) <= 200 then '02 - Entre 101 e 200'

        when convert(bigint,a.score) >= 201 and convert(bigint,a.score) <= 300 then '03 - Entre 201 e 300'

        when convert(bigint,a.score) >= 301 and convert(bigint,a.score) <= 400 then '04 - Entre 301 e 400'

        when convert(bigint,a.score) >= 401 and convert(bigint,a.score) <= 500 then '05 - Entre 401 e 500'

        when convert(bigint,a.score) >= 501 and convert(bigint,a.score) <= 600 then '06 - Entre 501 e 600'

        when convert(bigint,a.score) >= 601 and convert(bigint,a.score) <= 700 then '07 - Entre 601 e 700'

```

when convert(bigint,a.score) >= 701 and convert(bigint,a.score) <= 800 then '08 - Entre 701 e 800'

when convert(bigint,a.score) >= 801 and convert(bigint,a.score) <= 900 then '09 - Entre 801 e 900'

when convert(bigint,a.score) >= 901 and convert(bigint,a.score) <= 1000 then '10 - Entre 901 e 1000'

end as FAIXA\_SCORE

, case when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) is null then 'Cliente novo'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 0 then 'Zero mÍs'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 1 then 'Um mÍs'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 2 then 'Dois meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 3 then 'TrÍs meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 4 then 'Quatro meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 5 then 'Cinco meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 6 then 'Seis meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) > 6 then 'Mais de 6 meses'

end as FAIXA\_TEMPO\_BASE

, case

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) = 0 then '01 - Zero'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 0 and convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 10 then '02 - De 0 a 10'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 10 and convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 20 then '03 - De 10 a 20'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 20 and convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 30 then '04 - De 20 a 30'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 30 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 40 then '05 - De 30 a  
40'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 40 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 50 then '06 - De 40 a  
50'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 50 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 60 then '07 - De 50 a  
60'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 60 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 70 then '08 - De 60 a  
70'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 70 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 80 then '09 - De 70 a  
80'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 80 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 90 then '10 - De 80 a  
90'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 90 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 100 then '11 - De 90 a  
100'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 100 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 150 then '12 - De 100 a  
150'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 150 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 200 then '13 - De 150 a  
200'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 200 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 500 then '14 - De 200 a  
500'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 500 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 1000 then '15 - De 500  
a 1000'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 1000 then '16 -  
Mais de 1000'

end as FAIXA\_DEBITO

, case

```

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) = 0 then '01 - Zero'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 0 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 10 then '02 - De 0 a 10'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 10 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 20 then '03 - De 10 a 20'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 20 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 30 then '04 - De 20 a 30'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 30 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 40 then '05 - De 30 a 40'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 40 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 50 then '06 - De 40 a 50'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 50 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 60 then '07 - De 50 a 60'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 60 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 70 then '08 - De 60 a 70'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 70 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 80 then '09 - De 70 a 80'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 80 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 90 then '10 - De 80 a 90'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 90 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 100 then '11 - De 90 a 100'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 100 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 150 then '12 - De 100 a 150'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 150 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 200 then '13 - De 150 a 200'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 200 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 500 then '14 - De 200 a 500'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 500 and
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 1000 then '15 - De 500 a 1000'

when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 1000 then '16 - Mais de
1000'

end as FAIXA_ATRASO
, A.UF

```

```

,cluster

,class_code

,cd_ciclo_fatura

,de_faixa_ativ_vencto

,nr_dias_atraso

, faixa_renda_estimada

,faixa_rend_dom as faixa_rend_domiciliar

, sum(convert(float,replace(replace(valor,',',''),',','.'))) as soma

, count(*) as cont

from

        FPD_FIXA A

left join dbo.TB_CRDT_DIM_REGIONAL B on  A.UF = B.CD_UF

where A.UN like 'Var%'

and VENCTO >= CONVERT(char(6), DATEADD(MM, - 15, GETDATE()), 112)

AND VENCTO <= CONVERT(char(6), DATEADD(MM, - 2, GETDATE()), 112)

group by

        VENCTO

, flag_90

, a.nome_pdv

, Canal

, Resumo_Canal

, case when REGIONAL like 'Var%' then B.CD_REGIONAL

        else B.DE_REGIONAL_EMP

end

, case when produto like 'MINHA LINHA%' then 'MINHA LINHA'

        when produto like '%FALE 350%'

```

```

    or produto like '%FALE 500%'

    or produto like '%FALE 1000%' then 'FALE 350, 500 ou 1000'

when produto like '%CONTROLE%' then 'OI FIXO CONTROLE'

when produto like '%BASICO%'

    or produto like '%FALE 230%' then 'BASICO + FALE 230'

when produto like '%ECONOMIA%' then 'PLANO ECONOMIA'

when produto = 'OI FIXO A VONTADE' then 'OI FIXO A VONTADE'

when produto like '%OI FIXO 200%' then 'OI FIXO 200'

        else produto

end

, UN

,case when alavancado = 'R' then 'Alavancado'

        else 'Nao Alavancado'

        end

, A.UF

,cluster

,class_code

,case when convert(bigint,isnull(a.score,0))= 0 or a.score = '' then 'Zero'

        when convert(bigint,a.score) >= 1 and convert(bigint,a.score) <= 100 then '01 - Entre 1 e 100'

        when convert(bigint,a.score) >= 101 and convert(bigint,a.score) <= 200 then '02 - Entre 101 e 200'

        when convert(bigint,a.score) >= 201 and convert(bigint,a.score) <= 300 then '03 - Entre 201 e 300'

        when convert(bigint,a.score) >= 301 and convert(bigint,a.score) <= 400 then '04 - Entre 301 e 400'

```

when convert(bigint,a.score) >= 401 and convert(bigint,a.score) <= 500 then '05 - Entre 401 e 500'

when convert(bigint,a.score) >= 501 and convert(bigint,a.score) <= 600 then '06 - Entre 501 e 600'

when convert(bigint,a.score) >= 601 and convert(bigint,a.score) <= 700 then '07 - Entre 601 e 700'

when convert(bigint,a.score) >= 701 and convert(bigint,a.score) <= 800 then '08 - Entre 701 e 800'

when convert(bigint,a.score) >= 801 and convert(bigint,a.score) <= 900 then '09 - Entre 801 e 900'

when convert(bigint,a.score) >= 901 and convert(bigint,a.score) <= 1000 then '10 - Entre 901 e 1000'

end

,cep

, case when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) is null then 'Cliente novo'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 0 then 'Zero mÍs'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 1 then 'Um mÍs'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 2 then 'Dois meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 3 then 'TrÍs meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 4 then 'Quatro meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 5 then 'Cinco meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) = 6 then 'Seis meses'

when datediff(MM,pc\_dt\_oldest\_act\_line, ap\_application\_date) > 6 then 'Mais de 6 meses'

end

, case

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) = 0 then '01 - Zero'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 0 and convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 10 then '02 - De 0 a 10'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 10 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 20 then '03 - De 10 a  
20'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 20 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 30 then '04 - De 20 a  
30'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 30 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 40 then '05 - De 30 a  
40'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 40 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 50 then '06 - De 40 a  
50'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 50 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 60 then '07 - De 50 a  
60'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 60 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 70 then '08 - De 60 a  
70'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 70 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 80 then '09 - De 70 a  
80'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 80 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 90 then '10 - De 80 a  
90'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 90 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 100 then '11 - De 90 a  
100'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 100 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 150 then '12 - De 100 a  
150'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 150 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 200 then '13 - De 150 a  
200'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 200 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 500 then '14 - De 200 a  
500'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 500 and  
convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) <= 1000 then '15 - De 500  
a 1000'

when convert(float,replace(isnull(PCS\_AMOUNT\_ARREARS,0),',','.')) > 1000 then '16 -  
Mais de 1000'

end

, case

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) = 0 then '01 - Zero'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 0 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 10 then '02 - De 0 a 10'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 10 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 20 then '03 - De 10 a 20'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 20 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 30 then '04 - De 20 a 30'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 30 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 40 then '05 - De 30 a 40'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 40 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 50 then '06 - De 40 a 50'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 50 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 60 then '07 - De 50 a 60'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 60 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 70 then '08 - De 60 a 70'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 70 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 80 then '09 - De 70 a 80'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 80 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 90 then '10 - De 80 a 90'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 90 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 100 then '11 - De 90 a 100'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 100 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 150 then '12 - De 100 a 150'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 150 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 200 then '13 - De 150 a 200'

when convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) > 200 and  
convert(numeric,isnull(PCS\_DAYS\_ARREARS,0)) <= 500 then '14 - De 200 a 500'

```
when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 500 and  
convert(numeric,isnull(PCS_DAYS_ARREARS,0)) <= 1000 then '15 - De 500 a 1000'
```

```
when convert(numeric,isnull(PCS_DAYS_ARREARS,0)) > 1000 then '16 - Mais de  
1000'
```

```
end
```

```
,cd_ciclo_fatura
```

```
,de_faixa_ativ_vencido
```

```
,nr_dias_atraso
```

```
, faixa_renda_estimada
```

```
,faixa_rend_dom
```