

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO MESTRADO PROFISSIONAL EM
ENGENHARIA DE PRODUÇÃO**

Renan Manhadosco Moraes

**APLICAÇÕES DE TÉCNICAS
MULTIVARIADAS NA ÁREA COMERCIAL
DE UMA EMPRESA DE COMUNICAÇÃO**

Porto Alegre

2017

Renan Manhabosco Moraes

Aplicações de técnicas multivariadas na área comercial de uma empresa de comunicação

Dissertação submetida ao Programa de Pós-Graduação Mestrado Profissional em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Profissional, na área de concentração em Sistemas de Produção.

Orientador: Prof. Michel José Anzanello, *Ph D.*

Porto Alegre

2017

Renan Manhabosco Moraes

Aplicações de técnicas multivariadas na área comercial de uma empresa de comunicação

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Profissional e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação Mestrado Profissional em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof. Michel José Anzanello, *Ph.D.*

Orientador PMPEP/UFRGS

Prof. Ricardo Augusto Cassel, Dr.

Coordenador PMPEP/UFRGS

Banca Examinadora:

Professor Guilherme Luz Tortorella, Dr. (PPGEP/UFSC)

Professor Ricardo Augusto Cassel, *Ph.D.* (PMPEP /UFRGS)

Alessandro Kahmann, Dr. (DEPROT)

DEDICATÓRIA

Aos meus pais, Renato e Sônia e ao meu irmão, Rafael e em especial, a minha esposa Mariele, que me apoiaram na busca desse objetivo.

AGRADECIMENTOS

Quero agradecer, em primeiro lugar, a Deus pela força e conforto nos momentos difíceis. A minha família pelo apoio irrestrito em todos os momentos dessa jornada. Em especial, agradeço a minha esposa Mariele, por ter me acompanhado, incentivado e me apoiado desde o início. Seu amor foi fundamental para essa conquista.

Gostaria de agradecer também aos professores do departamento de Engenharia de Produção e Transportes pelos aprendizados ao longo do Mestrado. Ao professor Michel Anzanello, em especial, pela dedicação, paciência, ensinamentos e por acreditar no meu projeto, me orientando de forma muito próxima em todos os momentos dessa formação.

RESUMO

A mudança de comportamento dos consumidores através do advento da tecnologia e das redes sociais gera um grande empoderamento dos mesmos, alterando substancialmente a forma de relacionamento das empresas com seu público final. Atentas a este mercado, as empresas de mídia passam por profundas mudanças, tanto do ponto de vista da entrega de conteúdo ao seu público, quanto no seu formato administrativo, estratégico e financeiro. Sendo assim, a presente dissertação apresenta abordagens apoiadas em técnicas multivariadas para composição de equipes comerciais e de remuneração dos times de venda de uma empresa de comunicação. No artigo 1, objetiva-se gerar um modelo para estimar a premiação comercial das equipes de venda das rádios do Grupo RBS. Para tanto, inicialmente geram-se agrupamentos das emissoras de rádio do Grupo RBS no estado do Rio Grande do Sul e de Santa Catarina com base nos seus perfis de similaridades. Para cada cluster gerado, gera-se uma regressão linear múltipla da premiação comercial validado através de validação cruzada por intermédio do R^2 ajustado e *Mean Absolute Percentage Error* (MAPE). O segundo artigo aborda a clusterização dos top clientes do Grupo RBS e o impacto na composição das equipes comerciais por meio do método da seleção de variáveis. As 7 variáveis originais foram avaliadas através do método de seleção de variáveis “Omita uma variável por vez”; o melhor *Silhouette Index* (SI) médio, métrica utilizada para avaliar a qualidade dos agrupamentos gerados, foi obtido quando 3 variáveis foram retidas. Os agrupamentos gerados por tais variáveis refletem o comportamento de compra de mídia dos clientes; os agrupamentos foram considerados satisfatórios quando avaliados por especialistas do Grupo RBS.

Palavras-chave: Clusterização, Regressão Linear, Modelo Preditivo, Premiação de Vendas Rádios do Grupo RBS, Seleção de Variáveis, Alocação das equipes comerciais, Top clientes.

ABSTRACT

The change in the behavior of consumers with the advent of technology and social networks generates a great empowerment of themselves, substantially altering the relationship form of companies to their final audience. Attentive to this market, media companies undergo profound changes, both from the point of view of delivering content to their audience, as well as in their administrative, strategic and financial format. Thus, the present dissertation presents approaches supported by multivariate techniques for the composition of commercial and remuneration teams of the sales group of a communication company. In article 1, the objective is to generate a model to estimate the commercial awards of the sales teams of the RBS Group radios. To do this, we initially generate groupings of radio stations from the RBS Group in the state of Rio Grande do Sul and Santa Catarina based on their profiles of similarities. For each cluster generated, a multiple linear regression of the commercial award is generated, validated through cross validation through the adjusted R² and Mean Absolute Percentage Error (MAPE). The second article addresses the clustering of RBS Group top clients and the impact on the composition of business teams through the variable selection method. The original 7 variables were evaluated through the variable selection method "Omit one variable at a time"; the best Silhouette Index (SI) average, metric used to evaluate the quality of the generated clusters, was obtained when 3 variables were retained. Clusters generated by such variables reflect customers' buying behavior of media; the clusters were considered satisfactory when evaluated by RBS Group experts.

Keywords: Clustering, Linear Regression, Predictive Model, Radio Sales Awards from the RBS Group, Variables Selection, Allocation of Sales Teams, Top Customers.

LISTA DE FIGURAS

Figura 1: Ilustração hipotética de análise de clusterização	20
Figura 2: Dendograma gerado na clusterização das rádios do Grupo RBS	28
Figura 3: Exemplo de dendograma gerado pelo método hierárquico de clusterização.....	39
Figura 4: Dendograma gerado na clusterização dos 91 clientes	47
Figura 5: SI médio método “omita uma variável por vez”	48

LISTA DE TABELAS

Tabela 1: Variáveis independentes para clusterização normalizadas	27
Tabela 2: Emissoras clusterizadas com k entre 02 e 06	28
Tabela 3: Emissoras dentro de cada cluster.....	29
Tabela 4: Alocação dos clientes/observações nos clusters.....	47
Tabela 5: Alocação dos clientes/observações nos clusters após a seleção de variáveis.....	50

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Considerações Iniciais	11
1.2 Objetivos	12
1.3 Justificativa do Tema	12
1.4 Procedimentos Metodológicos.....	14
1.5 Estrutura da Dissertação	14
1.6 Delimitações da Pesquisa.....	15
1.7 Referências.....	16
2 PRIMEIRO ARTIGO: PREDIÇÃO DA PREMIAÇÃO COMERCIAL DE VENDAS DO GRUPO RBS ATRAVÉS DE TÉCNICAS MULTIVARIADAS	17
2.1 Introdução	17
2.2 Fundamentação Teórica.....	19
2.2.1 Clusterização	19
2.2.2 Regressão Linear Múltipla	22
2.3 Método	23
2.4 Resultados e Discussão	26
2.5 Conclusões	31
2.6 Referências.....	33
3 SEGUNDO ARTIGO: ESTUDO DE CASO DA CLUSTERIZAÇÃO DE CLIENTES E O IMPACTO NA FORMAÇÃO DAS EQUIPES DE VENDAS NO GRUPO RBS ATRAVÉS DA SELEÇÃO DE VARIÁVEIS	35
3.1 Introdução	35
3.2 Fundamentação Teórica.....	37
3.2.1 Clusterização	37
3.2.2 Seleção de Variáveis	41
3.3 Metodologia	43
3.4 Resultados e Discussão	45
3.5 Conclusão.....	51
3.6 Referências.....	52
4 CONSIDERAÇÕES FINAIS.....	55

1 INTRODUÇÃO

1.1 Considerações Iniciais

A indústria da mídia passa por profundas transformações. A comunicação das empresas com seu público, antes restrita a poucos canais (rádio, televisão, jornal, revistas e outdoor), tornou-se extremamente complexa devido ao crescente número de canais de comunicação, à criação e interatividade das redes sociais, bem como ao domínio da informação por parte dos clientes. Nas empresas de mídia, há dois públicos distintos: B2C (leitores, internautas, telespectadores e ouvintes) e B2B (anunciantes, os quais são alvo deste estudo). Com vistas a realizar um atendimento qualificado e orientado as necessidades e ao comportamento de compra dos clientes, o papel do executivo de conta passa a ser um fator estratégico dentro da organização.

Como forma de se adequar à nova realidade global de transformação da mídia, as empresas estão em processo constante de modificação das suas práticas gerenciais para buscar formas de administração mais eficientes e que estejam em contato próximo com seu público. Na busca de uma administração mais eficiente, as organizações buscam implementar novos processos gerenciais, que garantam uma estrutura de pessoas enxuta e eficiente, a partir de análises detalhadas e profundas, muitas vezes realizadas com o auxílio de técnicas e ferramentas multivariadas. Mckenna (1993) afirma que uma das formas das empresas tornarem-se competitivas e apresentarem diferenciais perceptíveis ao seu cliente são o aprimoramento constante da qualidade, através da melhoria nas práticas de gestão e, principalmente, foco no atendimento às necessidades do consumidor. Neste contexto, técnicas estatísticas multivariadas podem oferecer importante subsídio na melhora da gestão dos times de venda.

Sendo assim, esta dissertação está apoiada em dois artigos. O primeiro artigo apresenta uma clusterização de emissoras de rádio de acordo com seus perfis de similaridade, Para cada cluster, gera-se uma regressão com vistas à premiação de vendas dos executivos de conta. O segundo artigo propõe a clusterização dos maiores clientes do Grupo RBS apoiada nas variáveis mais relevantes para geração dos agrupamentos. A obtenção de tais clusters permite a geração de estratégias conjuntas que podem ser aplicadas aos integrantes de cada grupo obtido, permitindo maior agilidade e controle na gestão dos clientes.

1.2 Objetivos

O objetivo geral dessa dissertação é propor metodologias de clusterização, seleção de variáveis e regressão linear múltipla na área comercial de uma empresa de comunicação com vistas a obter um novo conceito de gestão e de estímulos comerciais.

Como objetivos específicos, listam-se:

- A) Utilizar técnicas de clusterização através dos métodos propostos em um estudo de caso numa empresa de mídia;
- B) Desenvolver modelo preditivo de premiação comercial, dentro dos agrupamentos, através de uma regressão linear múltipla;
- C) Testar o modelo preditivo da premiação comercial com os resultados originais através de uma porção teste/treino;
- D) Identificar as variáveis descritivas mais relevantes para clusterização

1.3 Justificativa do Tema

As mudanças tecnológicas são percebidas de forma intensa nos grupos de comunicação. Os grupos de comunicação percebem a Era Digital como um grande marco em seus processos de produção. A chegada da Internet e das redes sociais resultaram em mudanças mais rápidas que as percebidas antes, com a substituição do computador pela máquina de escrever, por exemplo (DORNELES e FROEMMING, 2014). Conforme exposto anteriormente, a indústria de mídia passa por uma grande transformação em virtude do empoderamento do cliente/consumidor e ao advento das redes sociais, visto que não é mais necessário utilizar um único veículo de comunicação (rádio, tv, jornal, revista e outdoor) para realizar publicidade. Dados do Kantar IBOPE demonstram que houve uma redução de 6% no volume de inserções de rádio de 2015 para 2014 (IBOPE) e o jornal impresso viu seus gastos com anúncios nesse meio despencar de 28%, em 1995, para 16% em 2005.

A tecnologia fortemente se insere e cria alguns novos desafios para as empresas de mídia. O livre fluxo de informações na internet, bem como o estabelecimento de novas formas de se fazer negócio, faz com que a administração de vendas assuma um papel estratégico (GOLDBERG, 2007). A área de marketing apresenta campo fértil para utilização de técnicas

quantitativas, especialmente a análise multivariada, principalmente pela necessidade de conhecer-se o mercado consumidor a partir da mensuração das opiniões, atitudes, preferências, perfil e outras características dos consumidores (MALHOTRA, 2001). Isso faz com que os grupos tradicionais de comunicação busquem ferramentas que possibilitem maior eficiência em seus resultados, bem como um atendimento comercial personalizado, focado e direcionado às necessidades dos clientes.

Seguir uma orientação para o cliente traz vantagens competitivas para as empresas pois levam a um desempenho empresarial mais alto, na forma de um aumento de lucratividade e de um crescimento da receita (SHETH et al., 2001). No entanto, muitas vezes as empresas possuem estruturas organizacionais opostas ao seu objetivo de mercado, desenvolvendo modelos comerciais isolados para cada uma das mesmas. Tais modelos são tipicamente gerados de forma empírica, sem auxílio de técnicas ou abordagens estruturadas. Além disso, os gestores usualmente implementam estímulos comerciais aos vendedores de forma a privilegiar as menores unidades e os clientes com menor potencial de investimento em detrimento das maiores unidades e das grandes contas. Caso a organização opte por entender o comportamento específico de absolutamente todos os seus clientes, a mesma terá um custo excessivamente elevado de colaboradores e tecnologia da informação para suportar o processo. Desta forma, percebe-se que a ausência de sistemática estruturadas para compreensão da similaridade entre os perfis dos clientes, bem como para predição das premiações de vendas a tais clientes, justifica as abordagens aqui apresentadas.

A criação de clusters com emissoras de rádio e clientes com características semelhantes entre si permitirá a geração de políticas comerciais mais eficientes, seja de alocação de pessoas ou de estímulos comerciais. Além da gestão deixar de ter foco na unidade específica e passar a ter foco nos agrupamentos (aumentando a replicação de *cases* de sucesso dentro dos agrupamentos), também será possível utilizar a predição de custo dos modelos comerciais para fins orçamentários, bem como para fins comparativos de possíveis novos modelos comerciais.

No contexto teórico, o tema abordado nessa dissertação se justifica pelo baixo número de estudos de aplicações de técnicas multivariadas em empresas de mídia, fazendo com que essa pesquisa venha a contribuir com a produção bibliográfica da área.

1.4 Procedimentos Metodológicos

O método utilizado na presente pesquisa é caracterizado, quanto à sua natureza, como pesquisa aplicada, visto que é aplicado dentro do Grupo RBS. Já quanto aos seus objetivos, é caracterizado como exploratório, pois através de análises das hipóteses será resolvido um problema prático. A abordagem é quantitativa através da utilização de técnicas multivariadas como clusterização, regressão linear múltipla e seleção de variáveis. Quanto aos procedimentos, será um estudo de caso, pois será realizada coleta de dados secundários para investigação real. O estudo de caso é uma investigação empírica que investiga um fenômeno contemporâneo em profundidade e em seu contexto de vida real, especialmente quando os limites entre o fenômeno e o contexto não são claramente evidentes (YIN, 2010).

A dissertação é desdobrada em dois artigos. O primeiro artigo apresenta a geração de modelos de regressão para predição da premiação de vendas das rádios do Grupo RBS com vistas a obter uma ferramenta eficiente para predição dos seus resultados. Este artigo foi dividido em duas etapas: clusterização das emissoras de rádio com base nas variáveis selecionadas e definição do número de clusters; e geração dos modelos preditivos em cada um dos clusters obtidos na etapa anterior.

O segundo artigo trata da formação de grupos de clientes de acordo com suas características similares com objetivo de estruturar e capacitar times comerciais que passem a atender os top clientes como consultores de mídia no formato multimídia além de possibilitar a replicação de estratégias comerciais de sucesso dentro do mesmo agrupamento. Este artigo foi desdobrado em duas partes: coleta de dados e definição do número de agrupamentos a serem gerados; e seleção de variáveis e posterior clusterização dos clientes com base nas variáveis selecionadas.

1.5 Estrutura da Dissertação

A dissertação está dividida em quatro capítulos. No primeiro são apresentadas as considerações iniciais, objetivo geral e específicos, justificativa do tema, procedimentos metodológicos adotados, estrutura da dissertação e as delimitações do estudo.

No segundo capítulo, que apresenta o primeiro artigo da dissertação, traz a geração de modelos de regressão para predição da premiação de vendas das rádios do Grupo RBS em cada um dos clusters definidos previamente via métodos hierárquicos e não hierárquicos. Os modelos de predição são avaliados e validados em porções de teste/treino através dos resultados de *Mean Absolute Percentage Error* (MAPE).

O terceiro capítulo apresenta o segundo artigo dessa dissertação, o qual utiliza metodologias de clusterização e seleção de variáveis nos top 93 clientes do Grupo RBS no estado do Rio Grande do Sul. Com os agrupamentos formados com base no comportamento de compra dos clientes, poderão ser identificadas as características dos executivos de conta que melhor possam atender aos mesmos juntamente com a possibilidade de desenvolver estratégias comerciais e programas de marketing para cada agrupamento.

Já o quarto capítulo é composto pelas considerações finais, além dos desdobramentos futuros deste trabalho dentro da empresa estudada.

1.6 Delimitações da Pesquisa

O estudo de caso dessa dissertação está delimitado ao Grupo RBS, sendo que o artigo 1 considera informações e dados dos veículos de comunicação e clientes dos estados do Rio Grande do Sul e de Santa Catarina enquanto que o artigo 2 considera apenas informações do Rio Grande do Sul. As operações de mídia de Santa Catarina foram comercializadas com outro grupo de comunicação durante a execução do trabalho.

O primeiro artigo foi limitado apenas às rádios do Grupo RBS, não sendo avaliados os negócios de jornal, digital e de televisão pertencentes a empresa. Além disso, não foram considerados neste estudo impactos de receita ocasionados pela sazonalidade de alguns grandes eventos. Contudo, como o processo de premiação de vendas é semelhante em todos os negócios e também é semelhante no novo formato multimídia, o presente estudo pode ser aplicado em todos os demais negócios.

No segundo artigo, as informações levantadas nos sistemas do Grupo RBS são referentes ao ano de 2016, havendo a possibilidade desses clientes não terem realizado qualquer

investimento em publicidade no ano de 2017. Por fim, não foi objeto do estudo o mapeamento do perfil dos atuais executivos de conta da empresa e a respectiva sugestão de alocação dos mesmos em cada cluster. Foi realizada apenas a identificação de elementos tidos como importantes nos profissionais que atenderão aos clientes dentro de cada agrupamento.

1.7 Referências

DORNELES, Felipe R.; FROEMMING, Lurdes M. **As tecnologias e os processos de produção em grupos de comunicação**. 5º Simpósio Internacional de Ciberjornalismo. Campo Grande: 27 a 29 de agosto, 2014.

GOLDBERG, Claudio. **A Estratégia E Objetivos Da Força De Vendas**. Disponível em <<http://www.institutomvc.com.br/artigos/post/a-estrategia-e-objetivos-da-forca-de-vendas>>. Acessado em novembro de 2017.

MALHOTRA, N.K. **Pesquisa de marketing**. Porto Alegre: Bookman, 2001.

MCKENNA, Regis. **Marketing de relacionamento: estratégias bem sucedidas para a era do cliente**. Rio de Janeiro: Campus, 1993.

YIN, Robert K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.

SHETH, J.N.; MITTAL, B.; NEWMAN, B.I. **Comportamento do cliente: indo além do comportamento de do consumidor**. São Paulo: Atlas, 2001.

2 PRIMEIRO ARTIGO

PREDIÇÃO DA PREMIAÇÃO COMERCIAL DE VENDAS DO GRUPO RBS ATRAVÉS DE TÉCNICAS MULTIVARIADAS

Resumo: *Modelos de premiação de vendas são importantes ferramentas utilizadas pelas organizações com vistas à motivação dos colaboradores e operacionalização da sua estratégia de negócio. Neste contexto, a predição de resultados futuros surge como uma importante ferramenta estratégica para tomada de decisão a curto, médio e longo prazo. Este artigo propõe a geração de modelos de regressão para predição da premiação de vendas das rádios do Grupo RBS. Para tanto, inicialmente geram-se agrupamentos das emissoras de rádio no estado do Rio Grande do Sul e de Santa Catarina com base nos seus perfis de similaridades; a geração de agrupamentos visa reduzir o número de modelos de regressão a serem gerados. Para cada cluster obtido, gera-se um modelo de regressão linear múltipla com vistas à predição da premiação de vendas dos executivos de conta e gerentes comerciais. Os modelos gerados foram testados em termos de sua coerência preditiva e considerados satisfatórios por especialistas de processo da empresa.*

Palavras-chave: Clusterização, Regressão Linear, Modelo Preditivo, Premiação de Vendas Rádios do Grupo RBS.

2.1 Introdução

Ao ingressar profissionalmente em uma organização, o interesse do colaborador não é apenas aumentar o lucro dessa organização ou empresa, mas satisfazer suas expectativas pessoais (as quais incluem realização com o que faz e remuneração adequada). Caso isso não ocorra, ele não sentirá uma relação de troca e sim de exploração (TAMAYO; PASCHOAL, 2003). De tal forma, as organizações têm percebido que a remuneração pode funcionar como um importante fator estratégico, pois pode alinhar os custos próximos aos fatores competidores e conduzir os seus empregados a um foco mais competitivo (apud ALMEIDA et al., 2007).

Se para os colaboradores em geral a remuneração é um fator estratégico, para a equipe de vendas é um fator ainda mais relevante e fundamental para o atingimento dos objetivos de resultado das organizações visto que os vendedores servem de elo pessoal da empresa com os clientes (KOTLER, 2000). Remuneração estratégica é aquela que representa um modelo de compensação que permite premiar aqueles colaboradores da empresa que, por uma razão ou outra, se destacaram em relação aos demais em determinado período (MARRAS, 2002).

Atualmente, o Grupo RBS trabalha com um modelo de premiação/remuneração comercial variável semelhante em todos os negócios (Rádio, TV e Jornal), abrangendo casos que vão desde um executivo de contas de televisão (que fatura milhões de reais ao mês) até outro de uma emissora de rádio que fatura centenas de reais ao mês. Esse modelo único de premiação comercial permite que o executivo com a menor meta, ao realizar uma grande venda, facilmente supere seus resultados comerciais. Tal modelo, no entanto, além de dificultar qualquer predição futura de custo comercial nas menores emissoras, também inibe uma gestão de carreira dos profissionais, visto que as possibilidades de ganhos com remuneração para esses profissionais podem ser superiores quando comparado aos dos profissionais que possuem um faturamento maior.

Tendo-se em vista que as informações que descrevem os perfis dos profissionais e seus ganhos tipicamente são armazenadas em bancos de dados, entende-se que a utilização de ferramentas multivariadas, como técnicas de clusterização e regressão, podem auxiliar na separação dos negócios semelhantes entre si e na realização de predições eficientes que norteiem e apoiem a gestão de premiação dos profissionais.

A clusterização de dados oferece uma maneira de entender e extrair informações relevantes de grandes conjuntos de dados. A ideia é que os dados inseridos em um grupo apresentem características em comum entre si e distintas em relação aos dados inseridos em outro grupo (OLIVEIRA, 2008). Técnicas de clusterização têm sido amplamente utilizadas em diversas áreas, tais como saúde, mineração de dados, segmentação de imagens, classificação de documentos, análises populacionais e de renda de países, entre outros (HAIR et al., 2005). Por sua vez, a análise de regressão é uma técnica estatística para investigar e modelar a relação entre variáveis independentes e dependentes. Em sua operacionalização, a análise de regressão requer que sejam estimados os parâmetros do modelo, os quais permitem ponderar as variáveis independentes para a predição de observações futuras (BARROS et al., 2008).

Este artigo propõe a aplicação de ferramentas multivariadas em dados de premiação/remuneração comercial variável no sistema de rádios do Grupo RBS. Para tanto, dados descrevendo características das diversas rádios são inicialmente analisados via técnicas de clusterização. Uma vez formados os grupos compostos por rádios que guardam similaridades entre si, geram-se modelos de regressão para cada cluster, de maneira a estimar a premiação a ser destinada às observações (representando colaboradores) de acordo com o volume e natureza das vendas realizadas. Pretende-se ainda avaliar a sensibilidade de cada uma das variáveis independentes que impactam no percentual de remuneração (variável dependente).

A sistemática proposta oferece contribuição real para o sistema em que será aplicado, visto que identificar o comportamento do sistema de remuneração variável atual é fundamental para a definição da estratégia comercial do negócio futuro a partir do projeto integrado de mercado e para a manutenção das suas margens de EBITDA/LAJIDA. A sigla EBITDA ou LAJIDA diz respeito ao Lucro Antes de Juros, Impostos, Depreciações e Amortizações, ou seja, reflete o resultado operacional/lucro operacional das empresas.

2.2 Fundamentação Teórica

Nesta seção são apresentados os fundamentos de clusterização e regressão linear múltipla, ferramentas sobre as quais o método proposto se apoia além da segmentação de mercado na área de marketing.

2.2.1 Clusterização

A análise de clusters é um procedimento multivariado que agrupa um conjunto de dados (observações) em subgrupos homogêneos, chamados de clusters. Essa análise inclui uma série de procedimentos estatísticos que podem ser usados para classificar objetos observando as semelhanças ou dissimilaridades entre eles, sem definir previamente critérios de inclusão em qualquer agrupamento (GUIDA, 2006). Os clusters formados apresentam homogeneidade entre os objetos que compõem um grupo e diferenças acentuadas entre os objetos pertencentes a diferentes grupos (JOHNSON; WICHERN, 1992; HAIR et al., 1995). Sistemáticas de clusterização têm sido amplamente utilizadas em variadas áreas do conhecimento, tais como ciências biológicas, ciências sociais e comportamentais, medicina, informática, TI, entre outras.

O objetivo da clusterização é permitir a separação dos dados com mais características em comum entre si do que analisadas com dados de outros grupos. Isso significa que dados e informações pertencentes a um mesmo cluster são mais similares, em diversos aspectos, do que dados e informações pertencentes a outro cluster. O agrupamento dos objetos permite analisar uma quantidade reduzida de dados sem haver, no entanto, perda significativa de informações (HAIR et al., 1995). As técnicas de clusterização, especificamente, procuram promover separação em forma de grupos – clusters – os quais sejam distintos (representativos) e que cada cluster contenha observações similares entre si e distintas em observações inseridas em outros clusters (KAUFMANN; ROUSSEEUW, 2005).

É importante observar que diversos fatores devem ser levados em consideração para uma eficiente definição de cluster: quantidade e qualidade dos dados, seleção das variáveis que descrevem as observações a serem agrupadas, medidas de similaridade, natureza e fundamentação teórica do algoritmo de clusterização e validação e análise dos resultados obtidos, dentre outros (JAIN; MURTY; FLYNN, 1999) (ver Figura 1).

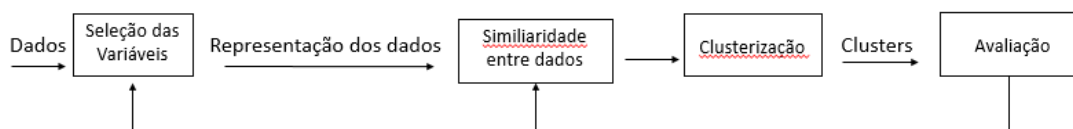


Figura 1: Ilustração hipotética de análise de clusterização

Uma análise de cluster criteriosa exige métodos que apresentem as seguintes características (ZAIANE, 2003): i - ser capaz de lidar com dados de alta dimensionalidade; ii - ser “escalável” com o número de dimensões e com a quantidade de elementos a serem agrupados; iii - ter habilidade para lidar com diferentes tipos de dados; iv - ser capaz de definir agrupamentos de diferentes tamanhos e formas; v - exigir o mínimo de conhecimento para determinação dos parâmetros de entrada; vi - ser robusto em relação à presença de ruído; e vii - apresentar resultado consistente independente da ordem em que os dados são apresentados.

Existem duas famílias de técnicas de clusterização amplamente utilizadas: hierárquicas e não hierárquicas (HAIR et al., 1995). Na clusterização hierárquica, são produzidas diversas partições do conjunto de dados com base na junção ou divisão dos clusters de acordo com a

medida de similaridade (OLIVEIRA, 2008). A medida é obtida através de procedimentos matemáticos sobre as variáveis de clusterização para todos os pares de objetos avaliados, possibilitando a comparação entre quaisquer objetos (JOHNSON; WICHERN, 1992). Quanto maior a similaridade (ou menor a dissimilaridade ou distância) entre dois elementos, mais próximos esses elementos encontram-se (EVERITT et al., 2001). A clusterização hierárquica tipicamente apoia-se em dendogramas. O dendograma é muito semelhante a uma árvore, no qual as folhas ou os ramos da árvore representam os clusters. De tal forma, tal recurso geralmente é utilizado para estimar um número adequado de agrupamentos a serem criados.

Na clusterização não-hierárquica, faz-se necessária a definição prévia da quantidade de agrupamentos/clusters a serem gerados. Geralmente, esse modelo é utilizado quando há fortes indícios para pré-definição do número de clusters. Os métodos mais conhecidos são o *k-means* (k-médias) e o método *k-medoid* (k-medóides).

Existem inúmeros métodos aglomerativos que se apoiam em distintas medidas de similaridade. Os métodos mais utilizados estão listados abaixo, sendo que os métodos i, ii e iii utilizam a matriz de similaridade inter-cluster; já o método iiiv apoia-se no conjunto de dados. Os métodos mais utilizados incluem o *i - simple linkage/nearest neighbor* (união simples): une clusters com a distância mínima entre pares de dados de clusters diferentes, gerando o primeiro cluster; *ii - complete linkage/furthest neighbor* (união completa): utiliza a distância entre os pares mais distantes, método oposto ao *simple linkage*; *iii - group average linkage* (união média): a distância entre os clusters é a média da distância entre todos os pares formados; *iiiv - centroid clustering* (método do centroide): a distância é medida pela diferença entre os centroides dos clusters.

As medidas de similaridade (ou distância) mais conhecidas no contexto de análise de cluster incluem distância Euclidiana, distância Euclidiana quadrática, distância de Manhattan e distância de Chebychev. A distância Euclidiana é a medida mais utilizada na definição de clusters, conforme a equação (1).

$$d_{ii'} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (1)$$

Na equação (1), d_{ij} é a distância entre os padrões, p é a dimensão dos vetores, X_{ik} é o componente k do padrão X_i e X_{jk} é o componente k do padrão X_j . A normalização dos valores de acordo com sua variância e média corrige uma das desvantagens do uso da distância Euclidiana, que é a possibilidade de dominância de um atributo sobre os demais.

A combinação dos métodos hierárquicos e não hierárquicos é muito utilizado na definição dos clusters. Segundo Hair et al. (1995), tipicamente utilizam-se os métodos hierárquicos para definição do número de clusters e para eliminação dos *outliers*; os métodos não-hierárquicos são então utilizados para efetiva geração dos clusters (inserção das observações nos clusters de destino).

Por fim, ressalta-se que, intuitivamente, pode-se considerar que um maior valor de dados, representado por um elevado número de variáveis descritivas da observação a ser alocada a um grupo, conduza a melhores agrupamentos. No entanto, a inclusão de variáveis irrelevantes e ruidosas no procedimento de clusterização pode reduzir a qualidade dos agrupamentos (ANZANELLO; FOGLIATTO, 2011), o que justifica o crescente número de estudos voltados à seleção de variáveis em contexto de clusterização.

2.2.2 Regressão Linear Múltipla

A regressão linear múltipla é uma técnica multivariada cuja finalidade principal é estabelecer uma relação matemática entre uma variável dependente (ou resposta) e o restante das variáveis que descrevem o sistema (variáveis independentes ou explicativas). De acordo com Hoffman (2006), gera-se uma regressão linear múltipla quando a variável dependente é função linear de duas ou mais variáveis exploratórias.

Segundo Sassi et al. (2011), na regressão linear múltipla, a variável determinada é aquela que tem correlação significativa com a variável a ser prevista. A apresentação genérica do modelo da regressão linear múltipla apoiado em k variáveis independentes e uma variável independente Y é trazida na equação (2):

$$Y = \beta_0 + \beta_{1x_1} + \beta_{2x_2} + \dots + \beta_{kx_k} + \varepsilon \quad (2)$$

onde existem k regressores ponderados por k coeficientes de regressão e ε é o erro aleatório (independente e que tendem a seguir distribuição normal com média zero e variância

desconhecida). O parâmetro β_j representa a variação esperada da resposta de Y para cada unidade na variação de x_j quando as demais variáveis da regressão são mantidas inalteradas.

A aderência do modelo de regressão aos dados é medida através do coeficiente de determinação (R^2), o qual representa a medida de variabilidade da variável dependente (Y) explicada pelo modelo gerado. O coeficiente R^2 é definido pela equação (3).

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \quad (3)$$

onde R^2 : coeficiente de determinação, SQ_R : soma dos quadrados da diferença, SQ_T : soma dos quadrados totais de y e SQ_E : soma dos quadros dos resíduos do conjunto das observações.

Um valor de R^2 próximo a 1, entretanto, não significa necessariamente uma boa adequação do modelo aos dados observados, visto que a inclusão de variáveis pouco significativas eleva o valor do coeficiente através do fenômeno de sobreajuste (JOBSON, 1991; MONTGOMERY; RUNGER, 1994). De forma a contornar tal problema, utiliza-se o R^2 ajustado (coeficiente de determinação ajustado), conforme equação (4):

$$R^2_{\text{ajust.}} = 1 - \frac{\frac{SQ_E}{n-1}}{\frac{SQ_T}{n-p}} = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2) \quad (4)$$

Dessa forma, quanto maior a diferença R^2 e R^2 ajustado, maior a hipótese que foram incluídas variáveis não significativas. Por fim, recomenda-se realizar testes de hipóteses com os coeficientes de regressão como forma de determinar sua real contribuição para o modelo de regressão (ou seja, se as variáveis devem ser mantidas ou não).

2.3 Método

O método de pesquisa utilizado é de natureza aplicada, pois tem como objetivo gerar conhecimento a partir de problemas já existentes dentro de uma organização privada. O objetivo é exploratório, pois através de análises das hipóteses resolve um problema prático. Além disso, a pesquisa apoia-se em uma abordagem quantitativa, pois são utilizadas

ferramentas matemáticas para a resolução dos problemas. A técnica de análise de dados a ser utilizada é multivariada, permitindo que as múltiplas relações possam ser adequadamente examinadas para se obter uma compreensão mais completa e realista na tomada de decisões (HAIR et al., 2005). Quanto aos procedimentos, trata-se de estudo de caso, o qual consiste em uma investigação empírica de um fenômeno contemporâneo dentro de um contexto da vida real, quando o limite entre o fenômeno e o contexto não estão claramente definidos (YIN, 2001). A vantagem notória do estudo de caso repousa na possibilidade de aprofundamento que oferece, pois os recursos se veem concentrados no caso visado, não estando o estudo submetido às restrições ligadas à comparação do caso com outros casos (LAVILLE; DIONNE, 1999).

A abordagem proposta se apoia nas seguintes etapas: 1) Coleta de dados de informações das rádios para clusterização; 2) Uniformização das variáveis; 3) Geração de agrupamentos através da clusterização (hierárquica e não hierárquica); e 4) Geração de modelos de regressão para predição das premiações. Tais etapas são detalhadas como segue.

1) Coleta de dados de informações das rádios para clusterização

Segundo Churchill e Peter (2000), ao escolherem a base para a clusterização, os profissionais de marketing se apoiam em seu conhecimento de mercado, nas tendências atuais das compras, na pesquisa de marketing e no bom senso. Dessa forma, a coleta de dados se caracteriza pelo levantamento das informações, de mercado e de negócio, que impactam na análise de resultado, junto aos clientes comerciais e ouvintes, das unidades de rádio do Grupo RBS. As variáveis escolhidas têm como objetivo abranger indicadores do negócio (tanto de resultado econômico financeiro quanto de audiência), incluindo receita anual, EBITDA anual e audiência digital, dentre outros. Também possibilitam mensurar o potencial de mercado da região onde a respectiva emissora está inserida.

A coleta de dados ocorrerá nos sistemas contábeis e de RH do Grupo RBS, podendo ainda utilizar informações extraídas do Censo 2010 realizado pelo IBGE para caracterização de aspectos demográficos das cidades em que as rádios estão sediadas. Os sistemas do Grupo RBS são auditados por uma consultoria externa, sendo estas bases confiáveis para utilizações como a proposta neste artigo. As variáveis selecionadas para a clusterização incluem: (i) faturamento anual total de cada rádio; (ii) faturamento anual local de cada emissora; (iii)

despesa anual de cada unidade; (iv) EBITDA anual de cada rádio; (v) margem de EBITDA da unidade; (vi) audiência digital da emissora; (vii) população da cidade sede da emissora; (viii) PIB total da sede da emissora; (ix) número de executivos de conta da unidade; (x) número de colaboradores da unidade; (xi) número de empresas da cidade sede da emissora; e (xii) número de clientes ativos da unidade.

2) Uniformização das variáveis

Tendo-se em vista que as variáveis apresentam medidas de apuração distintas em termos de magnitude (por exemplo, receita anual local, despesa local e PIB são medidos em milhões de reais, audiência digital em milhares, margem de EBITDA em % de atingimento e o número de executivos de conta e de colaboradores em unidades), faz-se necessária a uniformização das variáveis para evitar que magnitudes contrastantes influenciem negativamente na formação dos grupos (ou seja, uma variável de maior magnitude domine a clusterização). Para tanto, as variáveis são normalizadas (subtraídas da média daquela variável e divididas pelo desvio-padrão daquela variável) antes da formação dos grupos.

3) Geração de agrupamentos de rádios

Nessa etapa, inicialmente se define o número de clusters (k) a ser gerado através do método hierárquico. A decisão apoia-se na análise visual do dendograma gerado. Caso não seja possível definir claramente o número de cluster a ser gerado através do dendograma, recomenda-se testar um intervalo de k 's nas técnicas não hierárquicas (*k-means*, por exemplo). A definição de tais limites também pode se apoiar em aspectos subjetivos de interpretação dos clusters gerados (por exemplo, se o dendograma sugerir um intervalo de $k = 2$ a $k = 4$, pode-se optar por 4 caso existam justificativas subjetivas para a formação de tal número de grupos). Para o modelo hierárquico, a geração do dendograma utiliza o MATLAB; os clusters são gerados através do SPSS.

4) Geração de modelos de regressão para predição das premiações

As variáveis independentes X utilizadas são as duas únicas variáveis que compõem o cálculo da meta e também da premiação comercial do Grupo RBS. Neste estudo, a determinação da variável dependente Y (% de premiação de vendas) é composta por: x_1 : percentual de superação da receita praça (receita comercializada para a própria unidade de negócio); e x_2 :

percentual de superação da receita exportação (receita comercializada para as demais emissoras de rádio do Grupo RBS), conforme equação (5):

$$Y = a_1x_1 + a_2x_2 \quad (5)$$

Na sequência é gerada a regressão linear múltipla considerando as observações (informações acerca de executivos de conta e gerentes comerciais) inseridas em cada um dos clusters através das porções de teste e treino. A porção de treino compreende 75% do banco de dados e será utilizada na geração do modelo (determinação dos coeficientes da regressão). Com objetivo de validar quantitativamente os modelos gerados, os 25% restantes dos dados serão utilizados na porção de teste, ou seja, os valores originais de premiação são comparados com a predição dos modelos através da *Mean Absolute Percentage Error* (MAPE). Valores reduzidos de MAPE são desejados, visto que denotam boa capacidade preditiva dos modelos gerados quando aplicados a dados que não foram utilizados na geração do modelo preditivo.

2.4 Resultados e Discussão

O estudo de caso foi aplicado em um banco de dados de 5.121 informações das 20 emissoras de rádio do Grupo RBS nos estados do Rio Grande do Sul e de Santa Catarina, sendo: i: Rede Gaúcha (Rádio Gaúcha Porto Alegre, Rádio Gaúcha Serra, Rádio Gaúcha Zona Sul e Rádio Gaúcha Santa Maria); ii: Rede Atlântida (Atlântida Porto Alegre, Atlântida Florianópolis, Atlântida Serra, Atlântida Passo Fundo, Atlântida Zona Sul, Atlântida Santa Maria, Atlântida Tramandaí, Atlântida Blumenau, Atlântida Joinville, Atlântida Chapecó e Atlântida Criciúma); iii: Rede Itapema (Itapema Porto Alegre, Itapema Florianópolis e Itapema Joinville); iv: Rádio Farrroupilha; v: CBN Diário da Manhã. Foram excluídas da análise a CBN 1340 AM e Rádio Rural AM por possuírem baixo faturamento anual e por não fazerem parte da estratégia de desenvolvimento de negócio da empresa visto que estão no meio AM, meio que perde significativa audiência nos últimos anos pela má qualidade do sinal e também por não estar disponível para uso nos aparelhos de celular. A Rádio Cidade FM não faz parte da análise por não fazer mais parte do portfólio de marcas do Grupo, todas emissoras do Rio Grande do Sul.

A coleta de dados para a clusterização ocorreu no Censo de 2010, no BI Comercial e no sistema contábil JDEdwards. Os dados foram então normalizados para evitar que algumas variáveis se sobrepusessem às demais, conforme apresentado na Tabela 1:

Tabela 1: Variáveis independentes para clusterização normalizadas

	Faturamento Anual Total	Faturamento Anual Local	Despesa Anual	Ebitda Anual	Margem Ebitda	Audiência Digital	População Total	PIB Total	Nº Vendedores	Nº Colaboradores	Nº Empresas	Nº Clientes Ativos
RÁDIOS												
RD GAÚCHA	4,4	4,3	4,3	4,4	0,8	4,0	1,7	1,7	3,0	4,1	-0,6	4,2
RD GAÚCHA SM	-0,3	-0,4	-0,3	-0,3	0,8	-0,3	-0,6	-0,8	-0,4	-0,2	-0,5	0,1
RD GAÚCHA SERRA	-0,3	-0,3	-0,2	-0,2	1,6	-0,3	-0,3	-0,1	0,4	-0,1	1,5	-0,2
RD GAÚCHA ZONA SUL	-0,3	-0,4	-0,3	-0,3	2,0	-0,3	-0,7	-0,5	-1,3	-0,4	-1,0	-0,1
RD ATL POA	0,4	0,7	0,3	0,6	-1,0	1,2	1,7	1,7	1,3	1,0	-0,6	-0,3
RD ATL ZONA SUL	-0,3	-0,4	-0,3	-0,2	0,3	-0,4	-0,5	-0,7	-0,4	-0,4	-0,3	-0,1
RD ATL PASSO FUNDO	-0,3	-0,3	-0,3	-0,3	-0,1	-0,4	-0,8	-0,7	-0,4	-0,4	-0,6	0,2
RD ATL CAXIAS	-0,2	-0,3	-0,3	-0,1	1,0	-0,3	-0,3	-0,1	0,4	-0,4	1,5	0,2
RD ATL STA MARIA	-0,3	-0,4	-0,3	-0,3	0,5	-0,3	-0,6	-0,8	-0,4	-0,5	-0,5	-0,1
RD ATL TRAMANDAÍ	-0,4	-0,5	-0,4	-0,4	-0,8	-0,4	-1,1	-1,0	-2,1	-0,7	-1,6	-0,3
RD ATL STA CRUZ	-0,3	-0,3	-0,3	-0,4	-1,5	-0,4	-0,9	-0,7	-0,4	-0,4	-1,0	0,1
RD ITAP POA	-0,0	0,1	0,1	0,0	-1,0	-0,1	1,7	1,7	1,3	-0,1	-0,6	0,4
RD FARROUPILHA	0,3	0,4	0,7	-0,2	-0,9	-0,2	1,7	1,7	0,4	0,6	-0,6	-0,9
RD CIDADE	-0,2	-0,1	-0,0	-0,3	-0,1	-0,1	1,7	1,7	-0,4	0,0	-0,6	-0,4
RD CBN DIÁRIO DA MANHÃ	-0,1	-0,0	-0,2	-0,3	-0,6	-0,4	-0,3	-0,3	0,4	0,5	1,5	0,2
RD ATL BLUMENAU	-0,2	-0,3	-0,3	-0,1	1,1	-0,4	-0,5	-0,8	-0,4	-0,4	0,6	-0,6
RD ATL CHAPECO	-0,3	-0,4	-0,3	-0,3	0,3	-0,4	-0,8	-0,7	-0,4	-0,4	-0,7	0,2
RD ATL CRICIUMA	-0,3	-0,4	-0,3	-0,3	0,2	-0,4	-0,8	-0,7	-0,4	-0,4	-0,7	-0,4
RD ATL FLORIANÓPOLIS	-0,2	-0,1	-0,4	-0,2	-1,8	1,2	-0,3	-0,3	0,4	0,0	1,5	-0,8
RD ATL JOINVILLE	-0,3	-0,3	-0,3	-0,2	0,5	-0,4	-0,1	-0,0	-0,4	-0,4	0,8	-0,2
RD ITAP FLORIANÓPOLIS	-0,2	-0,3	-0,2	-0,2	-0,8	-0,4	-0,3	-0,3	0,4	-0,5	1,5	-0,4
RD ITAP JOINVILLE	-0,4	-0,4	-0,4	-0,4	-0,4	-0,5	-0,1	-0,0	-0,4	-0,5	0,8	-0,7

Na análise da Tabela 1, mesmo após a normalização das variáveis, percebe-se que a Rádio Gaúcha, com exceção do Nº Empresas, possui variáveis que se destacam quando comparadas às demais unidades de rádio, fazendo com que, provavelmente, a mesma seja inserida isoladamente em um cluster.

A Figura 2 traz o dendrograma gerado, o qual corrobora a exclusividade de um cluster para a Rádio Gaúcha POA (nº 1). Embora o dendrograma sugira a formação de somente dois clusters (Rádio Gaúcha POA e as demais rádios), entende-se que o teste de um maior intervalo de k's pode ajudar a desdobrar, de forma mais detalhada, as rádios que foram inseridas nos demais clusters.

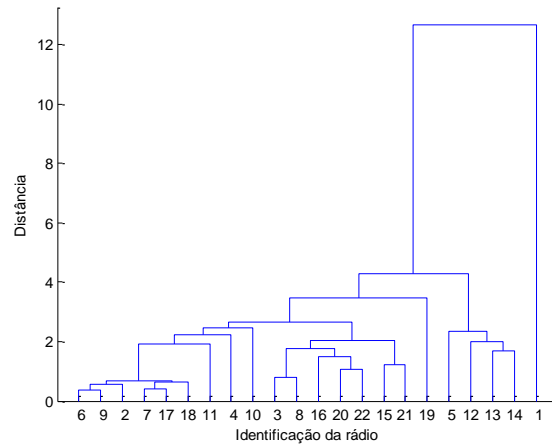


Figura 2: Dendrograma gerado na clusterização das rádios do Grupo RBS

De tal forma, a clusterização foi realizada utilizando o método hierárquico através do *software* SPSS; a medição de similaridade usada foi a distância Euclidiana. Conforme apontado acima, foram realizados agrupamentos variando k no intervalo $[2,6]$; o limite superior do intervalo foi definido com base no conhecimento de especialistas de negócio do Grupo RBS (gerentes comerciais), que não entendem como interpretável a geração de mais do que 6 clusters em uma análise composta por 20 rádios. A Tabela 2 aponta o cluster de destino das rádios para os valores testados de k .

Tabela 2: Emissoras clusterizadas com k entre 2 e 6

RÁDIOS	k=2	k=3	k=4	k=5	k=6
RD GAÚCHA	2	3	4	3	6
RD GAÚCHA SM	1	1	2	1	3
RD GAÚCHA SERRA	1	2	1	4	1
RD GAÚCHA ZONA SUL	1	1	2	1	3
RD ATL POA	1	2	3	5	5
RD ATL ZONA SUL	1	1	2	1	3
RD ATL PASSO FUNDO	1	1	2	1	3
RD ATL CAXIAS	1	2	1	4	1
RD ATL STA MARIA	1	1	2	1	3
RD ATL TRAMANDAÍ	1	1	2	2	2
RD ATL STA CRUZ	1	1	2	2	2
RD ITAP POA	1	2	3	5	5
RD FARROUPILHA	1	2	3	5	5
RD CBN DIÁRIO DA MANHÃ	1	2	1	4	4
RD ATL BLUMENAU	1	1	1	1	1
RD ATL CHAPECO	1	1	2	1	3
RD ATL CRICIUMA	1	1	2	1	3
RD ATL FLORIANÓPOLIS	1	2	1	4	4
RD ATL JOINVILLE	1	2	1	4	1
RD ITAP FLORIANÓPOLIS	1	2	1	4	4
RD ITAP JOINVILLE	1	2	1	4	4

Em todos os agrupamentos gerados, a Rádio Gaúcha ficou isolada das demais emissoras. Após a análise dos clusters obtidos, optou-se pela manutenção de 4 clusters, visto que este arranjo isola a Rádio Gaúcha em um cluster exclusivo e desdobra em outro cluster as demais emissoras de Porto Alegre (Atlântida Porto Alegre, Itapema Porto Alegre e Rádio Farroupilha), as quais são caracterizadas por forte audiência digital, elevado número de colaboradores, grande mercado consumidor, engajamento na sociedade e representativo faturamento. Nos dois clusters remanescentes, algumas situações chamam a atenção, tais como o agrupamento das rádios de Florianópolis (rádios consideradas de grande porte) junto a rádios de pequeno faturamento e engajamento, como a Itapema Joinville. Muito provavelmente, as variáveis de mercado e da região (PIB Total, N° de Empresas e População Total) apresentaram-se como mais relevantes do que as variáveis de negócio na formação dos agrupamentos. A composição final dos clusters ficou conforme a Tabela 3:

Tabela 3: Emissoras dentro de cada cluster

Cluster	Emissoras
1	Rádio Gaúcha Serra, Atlântida Caxias, CBN Diário da Manhã, Atlântida Blumenau, Atlântida Florianópolis, Atlântida Joinville, Itapema Florianópolis e Itapema Joinville
2	Rádio Gaúcha Santa Maria, Rádio Gaúcha Zona Sul, Atlântida Zona Sul, Atlântida Passo Fundo, Atlântida Santa Maria, Atlântida Tramandaí, Atlântida Santa Cruz, Atlântida Chapecó e Atlântida Criciúma
3	Atlântida Porto Alegre, Itapema Porto Alegre e Rádio Farroupilha
4	Rádio Gaúcha Porto Alegre

Após a geração dos clusters, realizou-se a coleta de informações referentes à premiação das equipes de vendas no sistema JDEdwards. Foram geradas regressões testando 3 formatos de variáveis de entrada no período de jun./2014 até jun./2015: i - regressão linear considerando como base todos os meses do período para cada variável independente x ; ii - regressão linear considerando a média trimestral de cada variável independente x no período; e iii - regressão linear após a normalização de todas as variáveis considerando a média trimestral de cada variável independente x . Para definição do formato com melhores resultados, realizou-se treino com 75% das amostras e teste em 25% das mesmas. Na comparação dos resultados originais com os da porção de teste através do MAPE, optou-se pela opção i (regressão linear considerando como base todos os meses do período para cada x), responsável pelo menor MAPE. Os resultados serão considerados satisfatórios para R^2 ajustado acima de 0,55 e MAPE abaixo de 10%.

Para o cluster 1, a regressão gerou o modelo $y = 0,6481x_1 + 0,0587x_2$. Na análise desse cluster, identifica-se que x_1 é fator preponderante na definição de y e, por isso, fundamental para análise e definição das estratégias comerciais. O desvio médio do cluster foi de 13% e o R^2 ajustado foi de 0,67, fato este que corrobora a validação da significância das variáveis. O resultado do MAPE para a porção de teste foi de 5,69%. O modelo gerado para esse cluster é muito satisfatório para predição de resultados futuros da premiação de vendas com base nos x_1 e x_2 mencionados acima, visto que o mesmo possui boa aderência aos dados e adequada precisão para predição (a julgar pelo MAPE 5,69%).

No cluster 2, a equação da regressão é $y = 0,3227x_1 + 0,123x_2$. Em comparação com o cluster 1, há uma redução na preponderância de x_1 , fato este explicado pois o resultado de receita dessas unidades é composto, em média, de 75% x_1 e de 25% x_2 . O desvio médio entre os resultados da porção de teste e de treino foi de 17%. O R^2 ajustado foi de 0,63 enquanto que o MAPE para a porção de teste é de 6,21%. Apesar de ser um pouco inferior ao R^2 do cluster 1, o modelo preditivo gerado pela regressão é muito satisfatório. É importante mencionar também que os clusters 1 e 2 contêm mais de 80% das observações da base de dados e do total de emissoras analisadas, sendo estes de grande relevância para futuras tomadas de decisão de orçamento de vendas para o Grupo RBS.

A equação de regressão para o cluster 3 é $y = 0,4453x_1 + 0,1664x_2$. O R^2 ajustado, assim como no cluster 2, foi de 0,63 e a média do MAPE foi de 1,02%. Importante destacar que esse

resultado de MAPE indica que os dados gerados na regressão da porção de treino foram muito próximos aos originais. Este cluster é estratégico para o Grupo RBS, pois representa 30% do custo de total com premiação comercial, o que reforça a importância dos resultados obtidos com a regressão, visto que o modelo é satisfatório para 63% dos resultados. Além disso, para esse cluster, está sendo estudado um estudo de Plano de Incentivo Comercial, o qual o modelo preditivo será de suma importância para calcular expectativa de receita *versus* expectativa de custo para cada incremento % de superação da meta.

Com relação ao cluster 4, referente aos dados de premiação de vendas da Rádio Gaúcha, a equação da regressão ficou $y = -0,0534x_1 + 0,2414x_2$. O resultado negativo para x_1 sugere que o vendedor que comercializar apenas este tipo de receita terá sua premiação com resultado negativo. Contudo, isso não ocorre porque, obrigatoriamente, há receita para x_2 nas comercializações do cluster 4 através da exportação de publicidade para as unidades do interior. Por exemplo, durante a transmissão de um jogo de futebol, uma publicidade feita pelo narrador será transmitida na Rádio Gaúcha Porto Alegre (x_1 nesta análise) e também na Rádio Gaúcha Santa Maria, Rio Grande e Caxias do Sul (x_2 nesta análise). Apesar desse cluster conter apenas 15% do número total de executivos de conta o mesmo é responsável por 40% do custo da premiação total de vendas em função do salário dos colaboradores, sendo este um cluster estratégico na projeção de custo do Grupo RBS. O R^2 ajustado foi de 0,49 enquanto o MAPE para a porção de teste foi de 0,92%. Na avaliação do R^2 ajustado, o mesmo é considerado insatisfatório ($<0,55$) pois a equação explica menos de 50% das observações do respectivo cluster.

2.5 Conclusões

Este artigo identificou o comportamento da premiação comercial de vendas das rádios do Grupo RBS através de métodos quantitativos dentro de agrupamentos gerados das respectivas emissoras. Após a coleta da base de dados realizada em diferentes sistemas da empresa, além de banco de dados internos e externos, identificou-se a necessidade de normalização das variáveis. Na sequência, ocorreu a clusterização utilizando 12 variáveis que impactam diretamente nos resultados do negócio rádio dentro do Grupo RBS ou que representam dados de mercado da respectiva região da emissora. O método testou diferentes agrupamentos ao variar k entre 2 e 6; em todos eles, devido a representatividade da Rádio Gaúcha, a mesma

permaneceu isolada em um cluster específico. Sendo assim, optou-se, através de uma combinação de métodos hierárquicos e não hierárquicos, por trabalhar com $k = 4$.

O comportamento da premiação comercial dentro de cada um dos clusters foi avaliada através de uma regressão linear. Identificou-se que a regressão linear aplicada dentro dos agrupamentos gerou melhores resultados do que quando aplicada no conjunto total das emissoras. Como resultado da regressão, o cluster 4, este composto apenas pelos dados da Rádio Gaúcha e sendo o mais representativo em termos de custo de premiação de vendas, apresentou os melhores resultados de MAPE. Justifica-se pelo fato da base de dados conter informações apenas da mesma unidade de negócio e também por não conter quaisquer outros dados de indicadores de mercado de regiões distintas à de Porto Alegre.

Para o Grupo RBS, este estudo é plenamente justificado pela possibilidade de aplicação prática do trabalho também no meio digital, televisão e jornal, visto que a previsibilidade assertiva dos custos de venda da empresa é fundamental para o atual modelo de operação integrada dos meios na venda de publicidade, principalmente para as emissoras do interior. A utilização de métodos estatísticos para construção de modelos preditivos e de cenários é uma tendência dentro das organizações, mas ainda não é uma realidade no Grupo. Sendo assim, este trabalho também será utilizado como projeto piloto para outras iniciativas neste contexto.

Além disso, com a predição gerada dentro de cada um dos clusters, o Grupo RBS passa a contar com uma ferramenta muito mais assertiva para predição dos seus resultados de custo gerado com a premiação comercial bem como utilizará os resultados deste trabalho para comparações futuras de estudos de implementação de novos modelos de premiação. Tal informação é fundamental para definição do orçamento de vendas por unidade de rádio dentro de cada cluster juntamente com o acompanhamento das métricas de resultado por cada um dos executivos de conta. Necessário também mencionar a representatividade que este trabalho refletirá no acompanhamento do projeto integrado de mercado implementado pela empresa no ano de 2017. Será possível, com o *input* da previsão anual de receita, calcular a margem comercial de cada cluster e comparar os mesmos como forma de medir a eficiência de cada equipe de venda.

2.6 Referências

- ANZANELLO, M. J.; FOGLIATTO, F. S. Selecting the best clustering variables for grouping mass-customized products involving workers learning. **International Journal of Production Economics**, v. 130, n. 2, p. 268-276, 2011.
- BARROS, E. A. C; SIMÕES, P.A; ACHCAR, J.A; MARTINEZ, E.Z; SHIMANO, A.C. Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos. **Revista Colombiana de Estatística**, jun. 2008.
- CHURCHILL, G. A. e PETER, J. Paul. Marketing: criando valor para os clientes. São Paulo: Saraiva, 2000.
- FURASTÉ, P. A. **Normas técnicas para o trabalho científico**: explicitação das normas da ABNT. 17. ed. Porto Alegre: Dáctilo-Plus, 2014.
- HAIR, J. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005.
- HAIR, J. et al. **Multivariate data analysis with readings**. 4. ed. New Jersey: Prentice-Hall, 1995.
- HOFFMANN, R. Regressão linear múltipla. In: **Análise de regressão**: uma introdução à econometria. São Paulo: Hucitec, 2006.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, v. 31, n. 3, p. 264-323, 1999.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. New Jersey: Prentice-Hall, 1992.
- KAUFMANN, L.; ROUSSEEUW, P. **Finding groups in data**: an introduction to cluster analysis. New Jersey: Wiley Interscience, 2005.
- KAZNAR, I. K.; GONÇALVES, B. M. L. **Regressão múltipla**: uma digressão sobre seus usos. Rio de Janeiro: IBCI, 2007.
- KOTLER, P. **Administração de marketing**. 10. ed. São Paulo: Prentice Hall, 2000.
- LAVILLE, C.; DIONNE, J. **A construção do saber**: manual da metodologia da pesquisa em ciências humanas. Belo Horizonte: Ed. UFMG, 1999.
- MARRAS, J. P. **Administração da remuneração**: remuneração tradicional e estratégica, elementos de estatística aplicada, normas legais e benefícios. São Paulo: Thomson, 2002.

RIBEIRO JÚNIOR, J. I. **Análises estatísticas no SAEG**. Viçosa: UFV, 2001.

SASSI, C. et al. **Modelos de regressão linear múltipla utilizando os softwares e estatística**: uma aplicação a dados de conservação de frutas. São Paulo, 2011.

YIN, R. K. **Estudo de caso**: planejamento e métodos. 2. ed. Porto Alegre: Bookman, 2001.

ZAIANE, O. R. et al. **On data clustering analysis**: scalability, constraints and validation. Edmonton Alberta: University of Alberta, 2003.

3 SEGUNDO ARTIGO

CLUSTERIZAÇÃO DE CLIENTES COM VISTAS À FORMAÇÃO DE EQUIPES DE VENDAS ATRAVÉS DA SELEÇÃO DE VARIÁVEIS

Abstract: *O rápido crescimento do volume de informações nos bancos de dados das organizações, aliado ao elevado nível de competitividade entre as empresas e à crescente exigência do consumidor, tem gerado uma necessidade de entendimento aprofundado e detalhado do comportamento de compra dos clientes. Para entender esse público, as organizações têm utilizado técnicas estatísticas que possam auxiliar na extração de informações de bancos de dados que descrevem o comportamento dos clientes, transformando tais informações em diferenciais mercadológicos. Este artigo propõe o agrupamento de clientes privados do Grupo RBS no estado do Rio Grande do Sul através de uma abordagem de clusterização apoiada em seleção de variáveis. As 7 variáveis inicialmente utilizadas refletem o porte do cliente e seu perfil de compra de mídia; os resultados gerados pela sistemática proposta mostraram a divisão dos clientes em três grandes agrupamentos, os quais dependem somente de 3 das 7 variáveis para sua geração. Os clusters gerados são majoritariamente compostos por clientes do mesmo segmento de mercado, o que contribui positivamente para a formação de equipes comerciais.*

Palavras-chave: Clusterização, Seleção de Variáveis, Top clientes Grupo RBS, Alocação equipes comerciais.

3.1 Introdução

Verifica-se atualmente um cenário de transformação na relação entre consumidores e empresas: os consumidores possuem acesso fácil à informação e buscam de forma muito mais rápida organizações e provedores que possam atender às suas necessidades. Neste cenário de mudança, as empresas que trabalham com um universo composto por elevado número de clientes possuem um nível adicional de dificuldade no entendimento dos mesmos. Muitas destas já buscam agrupar, através de técnicas multivariadas, os seus clientes em clusters com

perfil semelhante de compra (TODESCHINI; ANZANELLO; RODRIGUES, 2014). Tais técnicas visam à criação de agrupamentos de clientes que tenham perfis semelhantes de consumo, localidade geográfica, renda, investimento em mídia, forma de abordagem comercial e necessidade de customização nas suas propostas, dentre outros. Tal curso de ação permite a geração de estratégias comerciais de forma agregada, de maneira a atender um grupo de clientes com medidas semelhantes (evitando a necessidade de geração de medidas customizadas, tipicamente mais onerosas) além de designar o profissional de atendimento com o perfil mais adequado para cada agrupamento.

Trazer referências de outros trabalhos. O que são lacunas?

O Grupo RBS atende ao seu mercado B2B (*Business to Business*), mercado este composto apenas dos seus clientes de publicidade, tanto no interior do estado quanto na região metropolitana de forma fragmentada em rádio, televisão e jornal; dentro de cada meio de comunicação, o desdobramento é feito por agência de publicidade. Tal modelo fragmentado inviabiliza, para o Grupo RBS, a elaboração de estratégias e planos de ação focados no cliente, pois ao atender clientes de um único meio de forma especializada, a equipe de vendas possui o entendimento do comportamento de compra do cliente de apenas um meio e não da estratégia completa do mesmo e do problema que ele quer solucionar através da mídia. Além disso, a agência de publicidade possui em seu portfólio clientes de diversas magnitudes, desde os que investem milhões de reais de forma recorrente, até os que investem valores pouco representativos uma vez ao ano. De tal forma, o executivo de contas não possui a possibilidade de aproveitar *cases* de sucesso realizados com outros clientes da sua carteira, pois ele atende clientes de todos os portes, de todas as regiões do estado, de públicos-alvo e de comportamento de compra totalmente distintos.

Segundo Oliveira (2008), a clusterização de dados em grupos pode oferecer uma maneira de entender e extrair informações relevantes de um grande conjunto de dados. A proposta é que observações pertencentes a um mesmo grupo apoiem-se em características comuns, diferentes das características de observações inseridas em outros grupos. Oliveira (2008) também reforça a diferença entre classificação e clusterização. Enquanto classificação é a atribuição de dados em grupos previamente definidos, a clusterização busca descobrir os grupos através de técnicas exploratórias (que demandam análise e interpretação de especialistas). As técnicas de clusterização, no entanto, podem ser afetadas pela existência de variáveis pouco relevantes e

ruidosas, fazendo com que se faça necessária a aplicação de sistemáticas de identificação das variáveis mais relevantes (as variáveis menos relevantes são então removidas da análise). Tal curso de ação traz diversos benefícios, dentre os quais aumento da precisão da clusterização gerada e redução da necessidade de coleta de dados (HAIR et al., 2009).

Este artigo propõe a clusterização de clientes que possuem semelhanças em seus comportamentos de compra através de ferramentas multivariadas. Para tanto, dados relativos a compras dos clientes e características dos mesmos no mercado são coletados e analisados através de técnicas de clusterização. Com vistas ao incremento do processo de geração dos grupos, a sistemática de seleção de variáveis “omita uma variável por vez” é aplicada aos dados com o objetivo de manter no modelo somente as variáveis mais relevantes; os grupos gerados são então analisados na busca de medidas que atendam às particularidades de cada grupo.

A sistemática proposta oferece real contribuição para o Grupo RBS, visto que o agrupamento de clientes com características semelhantes contribui para a criação de estratégias internas e de programas de relacionamento de marketing voltadas a esses clientes. Além disso, a geração de uma abordagem estruturada para formação de grupos de clientes tem como objetivo substituir sistemáticas empíricas de agrupamento, as quais atualmente encontram-se em utilização. Por fim, o foco no cliente e não na agência e a designação de profissionais de venda com perfil semelhante ao perfil de compra do cliente são fundamentais para crescimento de receita e alavancagem de resultado através da venda de projetos customizados.

3.2 Fundamentação Teórica

3.2.1 Clusterização

A clusterização é uma técnica de análise multivariada que tem por objetivo fundamental alocar observações (clientes, no caso deste estudo) em grupos, de forma que as similaridades entre observações dentro de um mesmo cluster sejam elevadas, enquanto os clusters devem ser distintos entre si (HAIR et al., 1995). O objetivo da clusterização é encontrar uma organização válida e conveniente dos dados (JAIN; DUBES, 1988).

O agrupamento de dados e informações em clusters tem sido utilizado em diversas áreas da sociedade e organizações com impactos importantes para fins de planejamento e tomada de

decisão, as quais incluem: *i*: análise de dados espaciais, com a criação de mapas temáticos; *ii*: em hospitais e clínicas médicas através do agrupamento de pacientes com sintomas e características semelhantes; *iii*: nas áreas de marketing das organizações, com objetivo de segmentar e agrupar clientes bem como em pesquisas de mercado; *iv*): em bibliotecas, no agrupamento de livros por perfil de leitor; *v*: em seguradoras, para agrupamento de clientes com nível de risco semelhante; *vi*: nas áreas de recursos humanos das organizações para segmentação dos colaboradores; *vii*: identificação de terras semelhantes para fins agrários através do mapeamento via satélite (LEVIA; PAGE, 2000; e *viii*: análise de dados reais e sintéticos de terremotos para extração de características que permitam a previsão de eventos precursores de abalos sísmicos (DZWINEI et al., 2005).

Os algoritmos de clusterização são classificados em dez principais métodos: *i*: métodos hierárquicos; *ii*: métodos não hierárquicos; *iii*: métodos baseados em densidades; *iv*: métodos baseados em grade; *v*: métodos baseados em modelos; *vi*: métodos baseados em redes neurais; *vii*: métodos baseados em lógica *fuzzy*; *viii*: métodos baseados em kernel; *ix*: métodos baseados em grafos; *x*: métodos baseados em computação evolucionária. Os métodos hierárquicos e não-hierárquicos, quando utilizados de forma combinada, geram abordagens tradicionais para a clusterização (CASSIANO, 2014).

O método hierárquico agrupa observações através da distância entre os indivíduos. Uma distinção entre essas abordagens e as demais é que o resultado obtido não é constituído de apenas uma partição do conjunto inicial de dados, mas sim de uma hierarquia que descreve um particionamento diferente a cada nível analisado. O método hierárquico agrupa os dados de modo que se dois exemplos são agrupados em algum momento, nas próximas iterações eles continuam fazendo parte do mesmo grupo, mesmo se eles forem agrupados em outros clusters mais gerais, caracterizando assim uma hierarquia entre os clusters (FREITAS, 2006). Normalmente, o processo de agrupamento é representado por um dendograma (Figura 3).

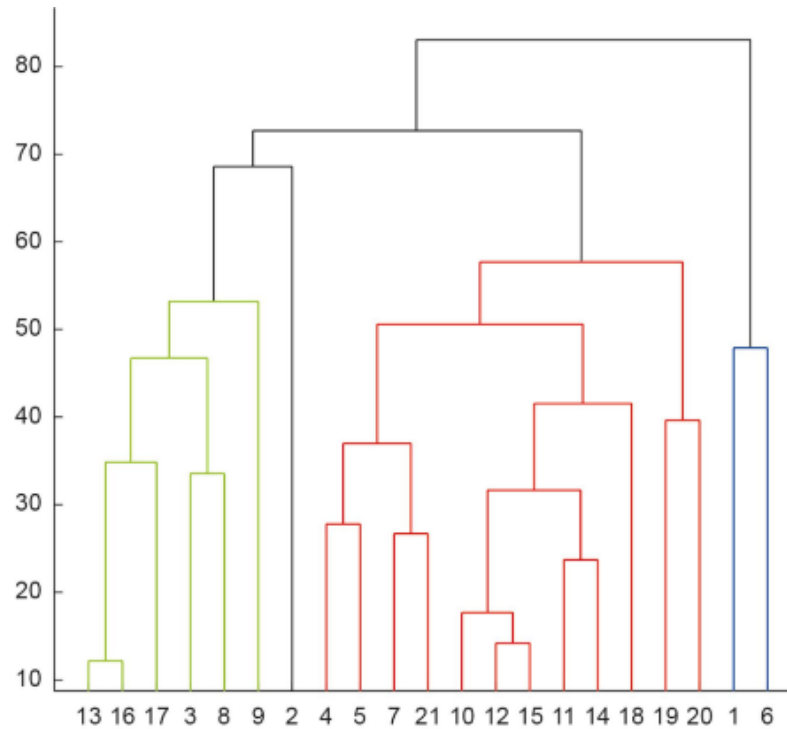


Figura 3: Exemplo de dendrograma gerado pelo método hierárquico de clusterização

A raiz do dendrograma representa o conjunto de dados inteiros, enquanto que os nós representam os indivíduos. A definição do número de clusters (k) a ser gerado pode ser obtido cortando-se o dendrograma em diferentes níveis; níveis com elevadas distâncias sugerem bons valores de k . Esta forma de representação fornece descrições informativas e visualização para as estruturas de grupos em potencial, especialmente quando há relações hierárquicas nos dados como, por exemplo, dados de pesquisa sobre evolução de espécies (CASSIANO, 2014).

O cluster hierárquico pode ser gerado através da forma aglomerativa ou forma divisiva. A primeira considera inicialmente todos os objetos como um cluster unitário e em seguida os agrupa através da medida de similaridade até a formação de um único cluster contendo todas as observações. A forma divisiva realiza o processo ao contrário, uma vez que parte-se de um único cluster que vai se dividindo em clusters menores até o número de cluster ser o mesmo número de objetos da porção, ou até o número de clusters atingir o número pré-determinado escolhido inicialmente. O método mais utilizado no processo de clusterização é o aglomerativo.

Por sua vez, o método não hierárquico agrupa os dados em um número k de clusters pré-determinado pelo decisor. Os objetos são divididos entre os k clusters de acordo com a medida de similaridade adotada, de modo que cada objeto fique no cluster que forneça o menor valor de distância entre o objeto e o centro do mesmo. Então, o algoritmo utiliza uma estratégia iterativa de controle para determinar como objetos devem mudar de cluster, de forma que a função objetivo usada seja otimizada (CASSIANO, 2014). Apresenta alta variância, pois a seleção dos exemplos representantes afeta, de forma significativa, o resultado do *clustering* e pode fazer com que a solução vá a um máximo local em função da avaliação. Para minimizar esse efeito negativo, usualmente os algoritmos são executados diversas vezes com exemplos iniciais diferentes e, então, a melhor solução é atribuída ao resultado do processo de clusterização (BERRY; LINOFF, 2002).

Han e Kamber (2001) observam que os mais bem conhecidos e geralmente usados métodos de particionamento são o *k-means*, o *k-medoids*, e suas variações. O *k-means*, mesmo tendo sido proposto há mais de 50 anos, ainda é um dos algoritmos mais utilizados para clusterização devido à facilidade de implementação, simplicidade, eficiência e sucesso empírico e possui várias extensões desenvolvidas em várias formas (JAIN, 2010).

Há quatro etapas principais que devem ser realizadas no processo de clusterização: *i*: pré-processamento dos dados – etapa de preparação e transformação dos dados, sendo muitas vezes necessária a normalização ou uniformização dos dados (subtraídas da média daquela variável e divididas pelo desvio-padrão daquela variável); *ii*: seleção da medida de similaridade – definição da medida de similaridade a ser utilizada na construção dos clusters; *iii*: avaliação do cluster – execução do processo de clusterização; *iv*: interpretação do cluster – é realizada por um especialista e consiste no processo de validação dos clusters.

A medida de similaridade mais utilizada no processo de clusterização é a distância Euclidiana. Ela expressa a distância entre os exemplos em um espaço multifuncional (FREITAS, 2006) conforme a equação (1), na qual $d_{ii'}$ é a distância entre os padrões, p é a dimensão dos vetores, X_{ik} é o componente k do padrão X_i e X_{jk} é o componente k do padrão X_j .

$$d_{ii'} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (1)$$

É importante mencionar que a avaliação do cluster é realizada com base em índices estatísticos que julgam, de uma maneira quantitativa, as estruturas encontradas no conjunto de dados. Um critério de validação expressa a estratégia utilizada para validar uma estrutura de agrupamento, enquanto que um índice é um valor estatístico pelo qual a validade é testada (FREITAS, 2006). Existem três tipos de critérios para investigar a validade de um agrupamento (HALKIDI et al., 2002): *i*: critérios relativos; *ii*: critérios internos; *iii*: critérios externos. Por exemplo, para validar a qualidade de um dendograma, se utiliza de um critério de validação interna conhecido coeficiente de correlação de Cophenético (EVERITT, 1993).

A qualidade dos agrupamentos gerados pode ser feita através do índice Silhouette (SI – *Silhouette Index*), que busca medir o grau de similaridade das observações dentro do seu próprio grupo em comparação com as dos demais grupos mais próximos a ela (ROUSSEUW, 1987). O SI é calculado conforme equação (2).

$$SI = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}} \quad (2)$$

Na equação (2), $a(i)$ é a média da distância da observação i dentro do grupo ao qual ela está alocada e $b(i)$ é a média da distância da observação i em comparação a todas as observações do grupo vizinho mãos próximo a ela. O valor de SI varia de -1 a +1, sendo que -1 indica que a observação foi alocada erroneamente no cluster de destino; valores próximos a 0 indicam que a observação poderia estar neste ou em outros clusters; + 1 indica que a observação está alocada no cluster correto.

3.2.2 Seleção de Variáveis

A seleção de um subconjunto de variáveis relevantes da totalidade das variáveis candidatas tem sido uma preocupação constante na literatura (ROSSINI et al., 2011), e visa à identificação de um subconjunto reduzido de variáveis que levam a resultados satisfatórios no reconhecimento de padrões em bases de dados consistindo de elevado volume de informações (VILLANUEVA, 2006).

Há muitos benefícios na utilização de métodos de seleção de variáveis, como maior facilidade na visualização dos dados, redução das dimensões do banco de dados para análise e armazenamento das informações e melhora no desempenho das previsões (GUYON;

ELISSEEFF, 2003). Para Friedman e Meulman (2004) e Huang et al. (2005), a sistemática de atribuição de pesos para identificar o subgrupo das variáveis mais relevantes é a melhor estratégia na definição das variáveis a serem consideradas para a clusterização.

Intuitivamente, pode-se considerar que um maior volume de dados, representado por elevado número de variáveis descritivas da observação a ser alocada a um grupo, conduza a melhores agrupamentos. No entanto, percebe-se que a inclusão de variáveis ruidosas e irrelevantes no procedimento de clusterização reduz a qualidade dos agrupamentos (ANZANELLO; FOGLIATO, 2011). A seleção de variáveis elimina o efeito indesejável das variáveis que mascaram a definição da estrutura dos clusters (BRUSCO, 2004).

Guyon e Elisseeff (2003) registram dois métodos clássicos para seleção de variáveis: *Forward Selection* e *Backward Elimination*. Enquanto o método *Forward* busca incorporar variáveis ao longo do processo, o método *Backward* elimina as variáveis irrelevantes para o processo de clusterização.

3.2.3 Estratégias de Vendas e Segmentação de Mercado

Estratégia é o padrão ou plano que integra as principais metas, políticas e sequência de ações de uma organização em um todo coerente. Uma estratégia bem formulada ajuda a ordenar e alocar os recursos de uma organização para uma postura singular e viável, como base em suas competências e deficiências internas relativas, mudanças no ambiente antecipadas e providências contingentes realizadas por oponentes inteligentes. (MINTZBERG e QUINN, 2001). A segmentação de mercado é um processo de identificação de subgrupos de clientes cujas necessidades, desejos e/ou recursos são diferentes, de maneira que os fazem responder de forma diferente a um dado composto de marketing. (SHETH, MITTAL E NEWMAN, 2001).

Um mercado é formado por compradores que diferem entre si em um ou mais aspectos. Eles podem diferenciar em seus desejos, recursos, localização geográfica, atitudes e práticas de compra. Quaisquer destas variáveis podem ser usadas para segmentar o mercado. Dividir um mercado em grupos distintos de compradores, que podem exigir mix de produtos e/ou de marketing distintos. A empresa identifica maneiras diferentes para segmentar o mercado e desenvolver os perfis dos segmentos resultantes (KOTLER, 1993).

As variáveis de segmentação são identificadas primeiramente através das características dos consumidores, utilizando-se variáveis geográficas, demográfica, comportamentais e psicográficas (ALVES, 2006). A segmentação geográfica refere-se à segmentação de mercados por região do país ou do mundo, tamanho de mercado, densidade de mercado ou clima. A segmentação geodemográfica agrupa clientes em categorias de estilo de vida próximas. A segmentação demográfica refere-se aos segmentos com base em variáveis demográficas porque elas estão amplamente disponíveis e, normalmente, relacionam-se com o comportamento de compra dos consumidores (LAMB, 2004). Na segmentação psicográfica, os compradores são divididos em diferentes grupos, com base em seu estilo de vida, sua personalidade e seus valores. Na segmentação comportamental, os compradores são divididos em grupos com base em seus conhecimentos de um produto, em sua atitude com relação a ele, no uso dele ou na resposta a ele (KOTLER, 2002).

3.3 Metodologia

O método de pesquisa é de natureza aplicada, visto que o objetivo é vinculado a uma situação real dentro de uma organização. O objetivo consiste em uma pesquisa exploratória, já que, segundo Mattar (2001), os métodos empregados compreendem levantamento em fontes secundárias, levantamento de experiências, estudos de casos selecionados e observações informações. Este trabalho, com objetivo de resolver um problema prático da organização, foi realizado através de técnicas de clusterização de clientes. A abordagem do presente trabalho foi quantitativa, visto que utilizou-se de ferramentas multivariadas de clusterização e seleção de variáveis para resolução de problemas. A abordagem quantitativa preocupa-se com quantificação dos dados, utilizando para isto recursos e técnicas estatísticas; é muito utilizada em pesquisas descritivas onde se procura descobrir e classificar a relação entre variáveis ou em pesquisas conclusivas, onde se buscam relações de causalidade entre eventos (OLIVEIRA, 1997). Em relação ao procedimento, o mesmo se caracteriza como um estudo de caso, visto que será realizada uma coleta de informações e análise de uma situação específica dentro do Grupo RBS. A vantagem mais marcante dessa estratégia de pesquisa repousa na possibilidade de aprofundamento que oferece, pois os recursos se veem concentrados no caso em análise (LAVILLE; DIONNE, 1999).

A operacionalização da proposta apoia-se em quatro etapas: 1) Coleta de dados de informações do perfil de compra dos clientes para clusterização; 2) Uniformização das

variáveis que descrevem o perfil dos clientes; 3) Definição do número de clusters; 4) Seleção das variáveis através do método “Omita uma variável por vez”. Tais etapas são detalhadas a seguir.

1) Coleta de dados de informações do perfil de compra dos clientes para clusterização

A primeira etapa deste estudo é a coleta de dados acerca do perfil de compra dos maiores clientes B2B que no somatório representaram mais de 50% da receita do Grupo RBS no ano de 2016. As informações levantadas, que incluem recorrência de compra, monoveículo ou multiveículo, customizado ou *standard* e receita anual, dentre outras, tem como objetivo o mapeamento completo do comportamento de compra, desde dados financeiros até informações qualitativas do tipo de mídia ou projeto comercializado. Cada informação é quantificada através de uma variável apropriada.

A coleta de dados para o presente estudo ocorre nos sistemas comerciais de rádio, televisão e jornal, no BI Comercial corporativo do Grupo e em revistas especializadas em economia para mensurar potencial de receita em não clientes. As informações coletadas proveem de sistemas validados pela área de Inteligência de Mercado, a qual valida se as mesmas estão iguais aos sistemas contábeis que são auditados por consultorias externas. As variáveis selecionadas para a clusterização, conforme segmentação de mercado, incluem: (i) faturamento anual do cliente; (ii) potencial de compra; (iii) número de colaboradores; (iv) investimento em mídia no Grupo RBS; (v) frequência de compra; (vi) número de projetos customizados; e (vii) número de veículos de mídia comprados dentro do Grupo RBS.

2) Uniformização das variáveis

A segunda etapa consiste em realizar o tratamento dos dados coletados em termos de valores espúrios (que devem ser auditados em termos de sua pertinência) e eventuais diferenças de magnitude. Variáveis com magnitudes contrastantes tendem a impactar negativamente no cálculo da distância Euclidiana, levando a clusterizações distorcidas. De tal forma, propõe-se o escalonamento das variáveis, no qual divide-se a variável em questão pela variável com valor máximo. Assim, as variáveis ficam restritas ao intervalo [0,1].

3) Definição do número de clusters a serem gerados

Nessa etapa, o número de clusters (k) é definido pela geração de um dendograma (método hierárquico). Caso não seja possível definir um valor preciso de k através do dendograma, o mesmo ainda é entendido como útil visto que permite definir um intervalo adequado de clusters a ser testado (por exemplo, de $k=3$ a $k=5$). A definição do número de clusters a serem gerados pode ainda se apoiar em prerrogativas subjetivas (por exemplo, especialistas da área de inteligência de mercado do Grupo RBS com formação em estatística entendem que a geração de 5 ou mais clusters não é recomendada visto que dificulta a interpretação dos perfis de cada cluster). Uma vez definido k , o mesmo permanece constante durante o processo de seleção de variáveis, descrito no passo 4.

4) Seleção das variáveis através do método “Omita uma variável por vez”

A seleção das variáveis apoia-se na omissão momentânea de uma das variáveis do conjunto inicial e posterior clusterização. A cada supressão de variável, agrupamentos são gerados e a qualidade da clusterização é avaliada pelo indicador SI. A variável que gerar o maior SI no momento que for suprimida é eliminada em definitivo da base de dados, visto que obtém-se o melhor agrupamento quando tal variável é omitida do procedimento de clusterização. Na sequência, com as variáveis remanescentes, nova rodada de omissões e clusterizações é conduzida, removendo-se novamente a variável que quando ausente conduzir ao maior SI. Tal procedimento é repetido até restar somente uma variável.

3.4 Resultados e Discussão

A sistemática proposta foi aplicada aos 93 maiores clientes do Grupo RBS no estado do Rio Grande do Sul; tais clientes correspondem a apenas 0,7% do total de clientes da empresa, mas representaram 53% do faturamento total do Grupo no ano de 2016. Os 93 clientes selecionados são compostos por clientes de diversos segmentos e ramos de atuação, como segue: construtoras e imobiliárias: 10 clientes; lojas de varejo: 10 clientes; concessionárias de veículos: 9 clientes; alimentação: 9 clientes; vestuário: 9 clientes; educação: 9 clientes; bebidas e fabricação: 7 clientes; hospitais e clínicas médicas: 7 clientes; supermercados e hipermercados: 5 clientes; bancos e instituições financeiras: 4 clientes; farmácias: 4 clientes;

sindicatos: 4 clientes; shoppings centers: 3 clientes e agrícola: 3 clientes. Todos os procedimentos computacionais foram realizados em Matlab.

Com objetivo de selecionar variáveis que se enquadrem dentro da segmentação de mercado (geográficas, demográfica, comportamentais e psicográficas), optou-se por utilizar 7 variáveis que foram coletadas para a clusterização dos 93 clientes: (i) faturamento anual do cliente; (ii) potencial de compra; (iii) número de colaboradores; (iv) investimento em mídia no Grupo RBS; (v) frequência de compra; (vi) número de projetos customizados; e (vii) número de veículos de mídia comprados dentro do Grupo RBS.

Os projetos customizados, variável (vi), são caracterizados pela compra de algum evento ou projeto de mídia, ou seja, desconsidera a compra de mídia avulsa e patrocínio de linha que constam na tabela de preços. O cálculo do potencial de compra dos clientes, variável (iii), é estimada através da equação (3).

$$\textit{Potencial de compra} = \textit{Faturamento anual líquido do cliente} \times 0,5\% \quad (3)$$

A equação (3) é uma métrica comumente utilizada no mercado para medir o potencial de investimento do cliente em publicidade. As variáveis (iv), (v), (vi) e (vii) foram coletados no BI Comercial do Grupo RBS, sendo que esta ferramenta é atualizada diariamente pelos sistemas comerciais de rádio, televisão e jornal da organização. A variável (i) foi coletada no balanço patrimonial de 2016 das 93 empresas selecionadas, ao passo que a variável (iii) foi obtida em revistas especializadas e em questionários enviados aos respectivos clientes no ano de 2016, sendo que a mesma pode apresentar uma variação da realidade atual dos mesmos.

Durante o processo de coleta dos dados, verificou-se que algumas variáveis eram medidas em milhões de reais (i, ii, iv) e outras em unidades (iii, v, vi e vii). Para que as variáveis medidas em milhões não se sobrepusessem às variáveis medidas em unidades quando do cálculo da distância Euclidiana na clusterização, realizou-se o escalonamento das variáveis pelo valor máximo isolado de cada uma das mesmas.

Com os dados normalizados, gerou-se o dendograma apresentado na Figura 4. A mesma figura sugere a geração de 3 clusters na segmentação dos clientes, assim como foi esta a sugestão do Gerente de Inteligência Comercial do do Grupo RBS para os top clientes com base em referências de mercado. Com objetivo de validar os 3 clusters sugeridos pelo

dendograma, classificou-se, anteriormente ao processo de clusterização, os clientes em segmentos de mercado/ramos de atuação (supermercados e hipermercados, educação e etc). Cruzando essa classificação anterior com as observações pós-clusterização, verifica-se que, em média, 60% dos clientes permaneceram dentro dos seus segmentos de mercado iniciais. Como o comportamento de compra de mídia é semelhante entre os segmentos, este fato corrobora a utilização de 3 clusters.

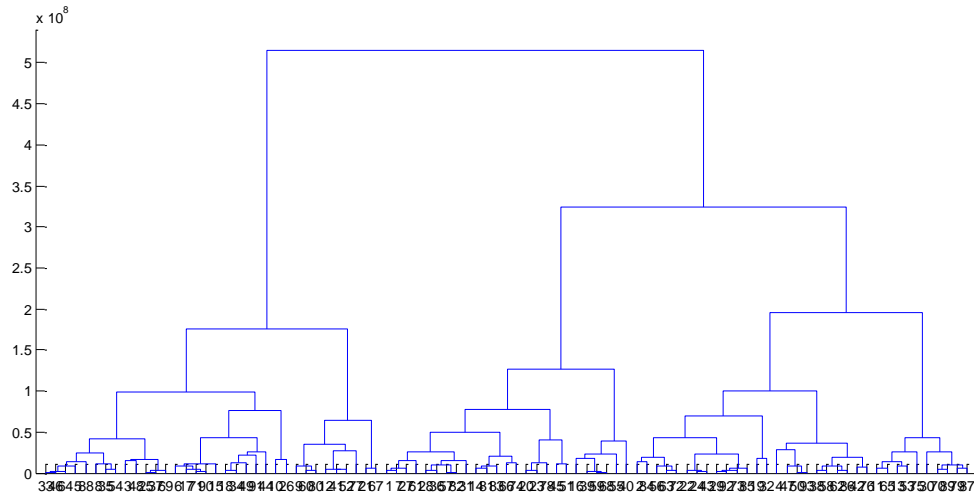


Figura 4: Dendograma gerado na clusterização dos 91 clientes

Os resultados da clusterização preliminar em 3 agrupamentos (utilizando todas as variáveis) são apresentados na Tabela 4.

Tabela 4: Alocação dos clientes/observações nos clusters

Cluster	Observação/Clientes															
1	1	3	5	6	7	9	12	16	17	20	26	28	31	32	33	44
	47	51	55	57	58	59	68	76	78	80	83	85	86	87	88	
2	4	8	10	13	14	15	18	19	25	27	36	37	38	39	40	41
	43	45	46	49	50	54	56	60	62	66	67	70	71	79	90	92
3	2	11	21	22	23	24	29	30	34	35	42	48	52	53	61	63
	64	65	69	72	73	74	75	77	81	82	84	89	91			

Na sequência, a sistemática de seleção de variáveis “Omita uma variável por vez” foi aplicada aos dados com o objetivo de excluir as variáveis menos relevantes e ruidosas com vistas ao

aumento da precisão da clusterização. O método foi realizado em fases, inicialmente com as 7 variáveis originais e então com a eliminação sucessiva de uma variável por fase até que somente uma variável restasse. O perfil de SI médio decorrente da eliminação das variáveis é apresentado na Figura 5.

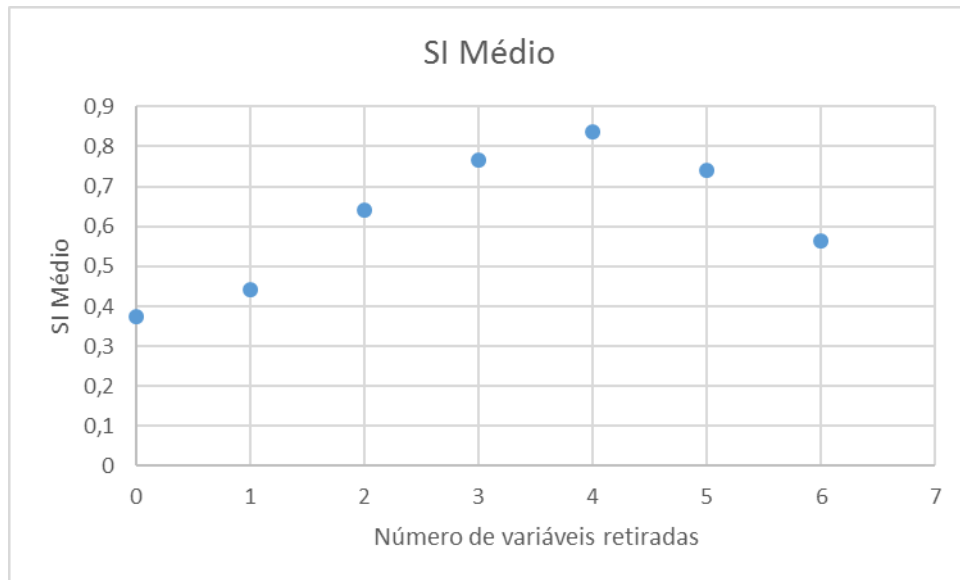


Figura 5: SI médio método gerado pelo método “omita uma variável por vez”

Na análise da Figura 5, identifica-se que o melhor SI médio é obtido quando 3 variáveis são utilizadas na clusterização, gerando um SI médio = 0,8358. Percebe-se ainda um incremento inicial no SI médio com a eliminação das variáveis menos relevantes, e então uma redução de desempenho por conta da remoção de variáveis contendo informação relevante. O maior SI médio foi obtido quando as seguintes variáveis foram mantidas: i (faturamento anual), iv (investimento em mídia no Grupo RBS) e v (frequência de compra). Tal resultado vai ao encontro de um modelo de segmentação de clientes já consagrado e comumente utilizado na gestão das áreas comerciais e de marketing que é o RVI, análise esta que já é utilizada há quase 50 anos por profissionais do marketing direto e que teve sua importância apontada pela primeira vez em 1961 por George S. Cullinan. O RVI utiliza faturamento e frequência de compra na formação de grupos; utiliza ainda a variável “recência” (data da última compra), a qual não foi analisada neste trabalho.

Enfatiza-se que, com as 7 variáveis originais, o SI médio foi de 0,3756. Na fase 1 (referente à primeira rodada de eliminação de variáveis), o SI médio foi de 0,4424 e a variável “número de veículos de mídia comprados dentro do Grupo RBS” foi excluída. Entende-se como

adequada a exclusão dessa variável, visto que a mesma acrescenta pouca informação na análise do perfil dos clientes em termos da formação de equipe de vendas, pois só identifica os tipos de mídia já adquiridos pelo cliente. Como o grupo RBS possui vendedor específico para cada veículo, é bem pertinente que, dentro dos maiores clientes, os mesmos já tenham realizado a compra de mídia com a grande maioria dos veículos.

Na segunda iteração (da qual resultaram 5 variáveis remanescentes), o SI médio foi de 0,6420 e a variável excluída foi “número de projetos customizados”. Apesar de ser um fator relevante na identificação do perfil de compra do cliente (e consequentemente na formação dos times comerciais, pois necessita de profissionais com caráter consultivo e criativo junto ao mesmo), a quase totalidade dos clientes da base analisada costuma comprar esse tipo de publicidade. Sendo assim, tal variável não é entendida como uma variável de diferenciação. Na terceira iteração, a variável excluída foi “número de colaboradores”. O SI médio, com as 4 variáveis restantes, foi de 0,7673. Esta variável, que ajuda a dimensionar o porte do cliente, não é fator relevante para definição da composição comercial de uma empresa de mídia.

Na quarta iteração descartou-se a variável “potencial de compra” e percebeu-se que a clusterização obteve o melhor SI médio = 0,8358. Na quinta iteração descartou-se a variável “investimento em mídia no Grupo RBS”; percebeu-se, no entanto, que a clusterização perdeu precisão nos grupos formados (o SI médio foi reduzido a 0,7407) Em uma composição de equipes comerciais, é fundamental considerar a variável “investimento do cliente no Grupo RBS”, pois este é um fator que explica se o cliente cresce ou decresce dentro da organização, além de ser chave para análise de histórico. Na última fase, com apenas as variáveis (i) e (v), descarta-se a (v) e o SI médio cai para 0,5644; a partir deste momento, a remoção de variáveis conduz a valores mais reduzidos de SI médio.

Concluído o processo de seleção, foram retidas as variáveis (i), (iv) e (v) para nova geração dos agrupamentos, conforme apresentado na Tabela 5. Antes da seleção de variáveis, conforme mencionado anteriormente, os clusters eram compostos por 60% de empresas do mesmo segmento/ramos de atuação, enquanto que pós seleção esse índice foi para 72% das observações.

Tabela 5: Identificação dos clientes/observações inseridos nos clusters após a seleção de variáveis

Cluster	Observação/Cientes - pós seleção variáveis																
1	4	8	10	20	22	23	26	28	30	31	34	36	42	43	49	51	
	54	59	64	68	71	74	75	76	79	83	84	85	86	88			
2	2	3	6	7	9	11	13	16	18	19	24	25	27	29	33	39	44
	46	47	52	53	58	60	63	66	67	69	70	72	73	77	80	81	82
3	91	1	5	12	14	15	17	21	32	35	37	38	40	41	45	48	50
	55	56	57	61	62	65	78	87	89	90	92	93					

O resultado chama a atenção, pois sugere um modelo comercial semelhante ao já adotado pela Zero Hora (ZH) em um passado recente, modelo este que aloca executivos de conta exclusivos para determinado segmento/ramo de atuação (imobiliário, automotivo, supermercados, educação, etc). Este modelo, com exceção dos segmentos imobiliário e automotivo, não gerou o resultado esperado e foi substituído pelo modelo de atendimento por agência.

Ainda sobre o modelo já aplicado, como o executivo passa a atender clientes do mesmo segmento, ele irá conhecer de forma detalhada e aprofundada as estratégias comerciais de clientes concorrentes. De tal forma, é fundamental que apresente um perfil discreto e que inspire confiança nos clientes para que os mesmos possam dividir suas estratégias. Isso é premissa para que o mesmo crie planos e projetos que alavanquem o resultado de todos, sem poder abrir para os demais clientes a estratégia comercial dos seus concorrentes.

Contudo, é importante destacar que a grande diferença do modelo aplicado neste trabalho ao modelo de segmentos de atuação já aplicado na ZH é que este foi construído de acordo com o perfil de compra dos top clientes, sem qualquer ligação com o ramo de atuação dos clientes. Na visão do autor, como muitas das variáveis aplicadas neste trabalho eram relacionadas ao comportamento do investimento em publicidade e mais especificamente no Grupo RBS, faz sentido que clientes do mesmo segmento de mercado tenham sido alocados no mesmo cluster, visto que o Grupo RBS é líder de audiência em rádio, tv, jornal e digital e os top clientes concentram boa parte dos seus investimentos publicitários no mesmo.

3.5 Conclusão

Este trabalho realizou o agrupamento dos 93 maiores clientes do Grupo RBS através de métodos de clusterização e seleção de variáveis. A formação de grupos de clientes teve como objetivo formatar e desenvolver equipes comerciais que passem a atender esses clientes como consultores de mídia no formato multimídia. O método utilizado neste trabalho teve como base a formação de agrupamentos de clientes após a definição prévia do número de clusters com o apoio de um dendograma. Durante o processo de coleta de dados, tanto nos sistemas do Grupo RBS quanto na consulta de dados externos referente aos top clientes, identificou-se a necessidade de normalização dos mesmos visto que as variáveis apresentavam grandezas distintas. Na sequência, utilizou-se a seleção de variáveis através do método “omita uma variável por vez” para melhorar a eficiência dos agrupamentos gerados. A qualidade dos agrupamentos gerados durante o processo de seleção de variáveis foi monitorada através do *Silhouette Index* médio. O melhor agrupamento é gerado quando 3 variáveis são mantidas (gerando SI médio = 0,8358), frente ao SI médio de 0,3756 quando as 7 variáveis originais são mantidas. As variáveis retidas foram faturamento anual, investimento em mídia no Grupo RBS e frequência de compra, tidas como adequadas através de uma validação subjetiva.

De acordo com as variáveis retidas, a formação das equipes comerciais que irão atender aos top 93 clientes do Grupo RBS deverá ser conduzida por profissionais criativos, que inspirem confiança e que possuam habilidades voltadas ao entendimento do comportamento dos clientes desses 3 grandes agrupamentos. É fundamental ainda que tais equipes possam propor planos e projetos multimídia em linha com a estratégia dos mesmos.

Para o Grupo RBS, este trabalho possui grande relevância, visto que durante a execução do mesmo, a organização realizou substancial transformação do seu modelo de operação comercial, passando a atuar de forma integrada em todos os seus veículos de comunicação. Neste projeto, o Grupo teve o apoio de uma consultoria externa durante o período de 8 meses, a qual a sua metodologia não baseada em nenhum método estatístico estruturado ou teve a utilização de ferramentas multivariadas. Esse processo ocorreu tanto no interior do estado quanto em Porto Alegre, com formatos de atuação distintos.

Com base no exposto acima, entende-se que esse estudo traz contribuições significativas ao Grupo RBS, o qual atende os seus clientes de forma multimídia, fato este que no início do

trabalho era uma considerável barreira a ser superada. Sendo assim, o trabalho proporcionará dois grandes estudos na empresa: (i) utilizar o método deste estudo e ampliar para todos os clientes do Grupo RBS, com a geração de uma nova segmentação, esta baseada em ferramentas estatísticas, com objetivo de comparar com a realizada pela empresa com o apoio da consultoria externa; (ii) possibilidade de implementar este trabalho para os top 91 clientes do Grupo RBS com a alocação dos melhores profissionais de venda e comparar a performance comercial desse seleto grupo de clientes com todos os demais já desdobrados nos clusters da empresa.

O formato de atuação da equipe comercial no modelo multimídia e o agrupamento de clientes em clusters é um avanço nas práticas comerciais do Grupo RBS e reflete a mudança de comportamento do mercado B2B e a busca pelo entendimento pleno dos clientes com um atendimento único voltado a potencializar seus resultados através de uma grande estratégia de comunicação. Além disso, a redução de custos com a diminuição do número de executivos de conta é outro fator positivo do modelo. O que ainda falta para a empresa é a utilização de modelos estatísticos e ferramentas multivariadas para que suas tomadas de decisão sejam mais precisas.

3.6 Referências

- ALVES, L.A.X. **Segmentação de Mercado como estratégia de marketing**. 2006. Centro Universitário de Brasília, Brasília, 2006.
- ANZANELLO, M. J.; FOGLIATTO, F. S. Selecting the best clustering variables for grouping mass-customized products involving workers learning. **International Journal of Production Economics**, v. 130, n. 2, p. 268-276, 2011.
- BERRY, M. J. A.; LINOFF, G. S. **Mastering data mining: the art and science of customer relationship management**. 1. ed. New York: Wiley, 2002.
- BRUSCO, M. J. Clustering binary data in the presence of masking variables. **Psychological Methods**, v. 9, n. 4, p. 510-523, 2004.
- CASSIANO, K. M. **Análise de séries temporais usando análise espectral singular (SSA) e clusterização de suas componentes baseada em densidade**. 2014. 172 f. Tese de Doutorado

(Departamento de Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2014.

DZWINEL, W. et al. Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. **Nonlinear Processes in Geophysics**, v. 2, n. 1, p. 117-128, 2005.

EVERITT, B. S. et al. **Cluster analysis**. 3. ed. London: Arnold, 1993.

FREITAS, J. A. B. **Análise de cluster da Lisozima**. 2006. 46 f. Relatório Científico (Escola de Artes, Ciências e Humanidades) - Universidade de São Paulo, São Paulo, 2006.

FRIEDMAN, J. H.; MEULMAN, J. J. Clustering objects on subsets of attributes (with discussion). **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 66, n. 4, p. 815-849, 2004.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of Machine Learning Research**, v. 3, p.1157-1182, 2003.

HAIR, J. et al. **Multivariate data analysis with readings**. 4. ed. New Jersey: Prentice-Hall, 1995.

HAIR, J. F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HALKIDI, M. et al. Cluster validity methods: Part I. **ACM SIGMOD Record**, v. 31, n. 2, p. 40-45, 2002.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 1. ed. Burlington: Morgan Kaufmann, 2001.

HUANG, J.Z. et al. Automated variable weighting in k-means type clustering, 2005.

JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, p. 651-666, 2010.

JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. New Jersey: Prentice-Hall, 1988.

KOTLER, P. **Administração de Marketing - Análise, Planejamento, Implementação e Controle**. 3 ed. São Paulo: Atlas, 1993.

- LAMB, Charles W. Jr., et al. **Princípios de marketing**. São Paulo: Pioneira Thomson Learning, 2004.
- LAVILLE, C.; DIONNE, J. **A construção do saber: manual da metodologia da pesquisa em ciências humanas**. Belo Horizonte: Ed. UFMG, 1999.
- LEVIA, D. F.; PAGE D. R. The use of cluster analysis in distinguishing farmland prone to residential development: a case study of Sterling, Massachusetts. **Environmental Management**, v. 25, n. 5, p. 541-548, 2000.
- MATTAR, F. N. **Pesquisa de marketing**. 3. ed. São Paulo: Atlas, 2001.
- MINTZBERG, H; QUINN, J. B. **O Processo da Estratégia**. Porto Alegre: Bookman, 2001.
- OLIVEIRA, S. L. **Tratado de metodologia científica**. 1. ed. São Paulo: Pioneira, 1997.
- OLIVEIRA, T. B. S. **Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada**. 2008. 112 f. Dissertação de Mestrado (Instituto de Ciências Matemáticas e de Computação) - Universidade de São Paulo, São Carlos, 2008.
- ROSSINI, K. et al. Método de seleção de variáveis para minimização da variância de predição. In: ROSSINI, K. **Seleção de variáveis no desenvolvimento, classificação e predição de produtos**. 2011. 143 f. Tese de Doutorado (Programa de Pós-Graduação em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.
- ROUSSEUW, P. J. Silhouettes: a graphical aid to interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53-65, 1987.
- SHETH, J.N.; MITTAL, B.; NEWMAN, B.I. **Comportamento do cliente: indo além do comportamento de do consumidor**. São Paulo: Atlas, 2001.
- TODESCHINI, B. V. et al. **Utilização de ferramenta de clusterização para alinhar a estratégia de uma empresa às necessidades de gestão de seus talentos**. 2014. 18f. Trabalho de Conclusão de Graduação (Curso de Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.
- VILLANUEVA, W. J. P. **Comitê de máquinas em predição de séries temporais**. 2006. 178 f. Dissertação de Mestrado (Faculdade de Engenharia Elétrica e de Computação) - Universidade Estadual de Campinas, Campinas, 2006.

4 CONSIDERAÇÕES FINAIS

O modelo de premiação comercial para equipes de vendas e a alocação de executivos de conta com perfil adequado ao de compra dos clientes são fatores estratégicos para o atingimento dos resultados operacionais de qualquer organização. Além dos resultados financeiros que um bom modelo pode gerar para a empresa, o mesmo deve permitir avaliar o impacto nos custos e também deve estar alinhado aos objetivos estratégicos de longo prazo ao estimular a venda de determinados produtos e serviços para cada tipo de cliente.

O objetivo deste estudo foi propor sistemáticas de agrupamentos de clientes prioritários e de emissoras de rádio do Grupo RBS, com vistas a adequar, dimensionar e direcionar as equipes comerciais com base nos clusters gerados. Esse estudo é justificado pela quase inexistente bibliografia do uso de técnicas multivariadas em grandes clientes de empresas de mídia além de propor uma nova forma de abordagem comercial em empresas de comunicação do Brasil. Dessa forma, o primeiro artigo desenvolveu modelos de regressão, dentro de clusters gerados nas emissoras de rádio, para predição da premiação de vendas com objetivo de criar uma ferramenta eficiente para predição dos seus resultados. As etapas operacionais foram desdobradas em quatro fases: (i) coleta de dados de informações das rádios para clusterização; (ii) uniformização das variáveis; (iii) geração de agrupamentos de rádios; (iv) geração de modelos de regressão para predição das premiações. Foram selecionadas 12 variáveis que impactam diretamente nos resultados do negócio rádio, as quais foram testadas agrupamentos com k entre 2 e 6; optou-se, por utilizar $k = 4$. Dentro de cada um dos clusters gerados, foi realizada uma regressão linear com objetivo de avaliar o comportamento da premiação comercial. Os melhores resultados de MAPE foram obtidos relativamente ao cluster composto pela Rádio Gaúcha Porto Alegre. Ficou evidente que essa ferramenta será muito importante para o Grupo RBS avaliar cenários, implementar projetos pilotos futuros e mensurar custos.

O segundo artigo abordou a geração de agrupamentos dos top clientes do Grupo RBS com vistas a identificar e direcionar profissionais de venda para atendimento desses clientes como consultores de mídia no formato multimídia. O estudo foi realizado em quatro etapas: (i) coleta de dados de informações do perfil de compra dos clientes para clusterização; (ii) uniformização das variáveis; (iii) definição do número de clusters a serem gerados; (iv) seleção das variáveis através do método “omita uma variável por vez”. Foram coletados e normalizados os dados de 7 variáveis para a geração de 3 clusters, definidos através de um

dendograma. No método de seleção de variáveis, identificou-se que o melhor resultado de SI médio ocorreu com 3 variáveis no modelo. Após isso, geraram-se novos agrupamentos considerando somente as variáveis selecionadas. Enfatiza-se que os modelos gerados oferecem grande oportunidade para que a empresa identifique as melhores alocações dos melhores profissionais de venda.

Com base no exposto, entende-se que os objetivos descritos no capítulo 1 desse trabalho foram atingidos, uma vez que os modelos gerados em ambos os artigos podem ser desdobrados em projetos piloto na nova forma de atendimento comercial multimídia que o Grupo passou a utilizar a partir de 2017. A aplicação de ferramental multivariado mostrou-se fundamental para a estruturação de métodos mais objetivos na predição de premiação e formação de grupos de vendas alinhados com os perfis dos clientes.

Futuros desdobramentos deste estudo incluem: (i) para fins orçamentários, calcular a margem comercial de cada cluster e comparar os mesmos como forma de medir a eficiência de cada equipe de venda; (ii) aplicar o método aplicado no artigo 2 a todos os clientes do Grupo RBS, com a geração de uma nova segmentação que permita comparar as proposições com a sistemática utilizada pela empresa; (iii) aplicar o método desenvolvido no artigo 2 para alocar os melhores profissionais de venda nos top clientes e comparar a performance comercial desse seleto grupo de clientes com todos os demais já desdobrados nos clusters da empresa.