

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

EDMIR VIEIRA LIMA SOBRINHO

**CLUSTERIZAÇÃO DE DADOS DE ATENDIMENTO NO SETOR DE
SERVIÇOS PÚBLICOS: O CASO DE UMA EMPRESA DE
FORNECIMENTO DE ENERGIA ELÉTRICA**

**VITÓRIA
2020**

EDMIR VIEIRA LIMA SOBRINHO

CLUSTERIZAÇÃO DE DADOS DE ATENDIMENTO NO SETOR DE
SERVIÇOS PÚBLICOS: O CASO DE UMA EMPRESA DE
FORNECIMENTO DE ENERGIA ELÉTRICA

Trabalho de Conclusão de Curso
apresentado ao Departamento de
Engenharia de Produção do Centro
Tecnológico da Universidade Federal
do Espírito Santo, como requisito
parcial para obtenção do grau de
Bacharel em Engenharia de Produção.

Orientadora: Dr^a. Mirela Guedes Bosi

VITÓRIA
2020

RESUMO

O objetivo deste trabalho foi contribuir para o processo de tomada de decisão acerca da destinação de recursos humanos, financeiros e tecnológicos em uma distribuidora de energia elétrica que presta serviço de atendimento ao público. A tomada de decisão refere-se à melhor escolha de diversificação e destinação de recursos para 69 agências de atendimento presencial ao público, visando melhorar os níveis de qualidade da organização, que são retratados em indicadores de qualidade como TME (Tempo Médio de Espera na fila para ser atendido) e TMA (Tempo Médio de Atendimento no guichê após espera na fila). Foi verificado que o desafio do gestor operacional era saber para qual grupo de agências deveriam ser inseridos planos de ações para a melhoria desses indicadores. Portanto, foram coletados dados do sistema de ERP durante três meses. Os dados foram tratados manualmente para serem aproveitados no algoritmo de clusterização. O algoritmo selecionado foi baseado no método de agrupamento K-Means, que aplica cálculos sobre as variáveis de entrada e, assim, as agrupa por similaridade, que é calculada pela distância entre as mesmas, criando centroides de similaridade e, assim pode-se verificar como os dados se comportam. Os *clusters* (grupos) são formados ao redor dos centroides. O número de cinco grupos foi previamente estimado como valor inicial e inserido no algoritmo, porém após os cálculos foi plotado o gráfico *Elbow* que possibilitou a identificação do melhor número de *clusters* possível perante a diminuição da variação dos dados, deixando assim evidente que três grupos de agências são muito similares perante as variáveis propostas. Dessa forma, as agências foram classificadas em níveis A, B e C, que indicam respectivamente agências de grande porte, médio e de pequeno porte. A ferramenta de clusterização desenvolvida, portanto, tornou mais eficiente a tomada de decisão visto que as agências podem ser vistas como grupos com similaridade de atributos, o que possibilita a aplicação de recursos para um determinado grupo ao invés de focar em apenas uma localidade.

Palavras-chaves: Clusterização de dados, agrupamento de dados, tomada de decisão, serviços públicos, K-Means.

LISTA DE FIGURAS

Figura 1 – Exemplo de árvore de decisão	17
Figura 2 – Algoritmo k-means	19
Figura 3 – Exemplo de dendograma	20
Figura 4 – Fluxograma de atendimento ao cliente	21
Figura 5 – Etapas da pesquisa.....	25
Figura 6 – Código para implementação do gráfico <i>Elbow</i>	28
Figura 7 – <i>Dataset</i> gerado a partir dos dados do tratamento inicial	32
Figura 8 – Parâmetros de inicialização do algoritmo K-Means.....	33
Figura 9 – Distâncias entre variáveis de um centroide gerado em um <i>cluster</i>	33
Figura 10 – Rótulos de classificação de atributos (<i>labels</i>) de um dos <i>clusters</i>	34
Figura 11 – Gráfico <i>Elbow</i>	35
Figura 12 – Centroides e <i>clusters</i>	36

LISTA DE GRÁFICOS

Gráfico 1 – Demanda mensal média de cada agência	30
Gráfico 2 – TME (tempo médio de espera, em minutos) por agência	31
Gráfico 3 – TMA (tempo médio de atendimento, em minutos) por agência	31

LISTA DE QUADROS

Quadro 1 – Demanda mensal média de cada agência	37
---	----

LISTA DE TABELAS

Tabela 1 – TME (tempo médio de espera, em minutos) por agência	38
--	----

Sumário

1	INTRODUÇÃO	9
1.1	CONTEXTUALIZAÇÃO	9
1.2	OBJETIVOS	10
1.2.1	Objetivo geral	10
1.2.2	Objetivos específicos	10
1.3	JUSTIFICATIVA	11
2	Referencial teórico	12
2.1	TECNOLOGIAS DE INFORMAÇÃO	12
2.1.1	Sistemas ERP - Enterprise Resource Planning	12
2.1.2	Big Data	13
2.2	EXTRAÇÃO DE CONHECIMENTO EM BASES DE DADOS.....	14
2.2.1	Seleção de dados	14
2.2.2	Processamento de dados	15
2.2.3	Transformação de dados	15
2.3	DATA MINING (MINERAÇÃO DE DADOS)	16
2.3.1	Estatística	16
2.3.2	Indução	17
2.3.3	Algoritmos genéticos	18
2.3.4	Classificação	18
2.3.5	Análise de <i>cluster</i>	18
2.3.5.1	Método K-Means	18
2.4	SERVIÇOS PÚBLICOS	12
3	METODOLOGIA	22
3.1	CARACTERIZAÇÃO DA EMPRESA	22
3.2	IDENTIFICAÇÃO DO PROBLEMA.....	23
3.3	CARACTERIZAÇÃO DA PESQUISA	24
3.4	COLETA DE DADOS.....	25
3.5	ANÁLISE DE DADOS E CLUSTERIZAÇÃO DE DADOS.....	26
4	RESULTADOS E DISCUSSÃO	29
4.1	EXTRAÇÃO DE DADOS	29
4.2	TRATAMENTO DE DADOS	32
4.3	APLICAÇÃO DO MÉTODO DE CLUSTERIZAÇÃO K-MEANS.....	32
5	CONCLUSÕES	39
6	REFERÊNCIAS BIBLIOGRÁFICAS	40
	APÊNDICE 1 – TABELA COM DADOS DAS AGÊNCIAS	44

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Ao longo dos últimos anos, nota-se acentuado crescimento da quantidade de dados traduzidos em informações e armazenados em meios digitais (MAGLIO; LIM, 2016; SONKA, 2016; WOLFERT et al., 2017). Conforme tal cenário, as tecnologias de tratamento e armazenamento de dados são constantemente submetidas à inovação e, ao passo que essa evolução é notada, o barateamento de equipamentos e dispositivos de armazenamento de dados tem tornado os sistemas de gerenciamento de bancos de dados e *data warehouse* eficientes e populares em diversos âmbitos da sociedade (RIBEIRO, 2014; SONKA, 2016).

Entretanto a configuração dos dados armazenados em grande volume e escala não são propícios a ocasionar *insights* em pesquisadores (GAMARRA; GUERRERO; MONTERO, 2016; OLIVEIRA et al., 2017), uma vez que métodos tradicionais como planilhas e cálculos simples não têm efeito de extrair informações implícitas dos dados em questão (VALLI, 2002; OLIVEIRA, 2018). A necessidade de obter informações relativas e importantes de bases de dados é evidente. Portanto, esse cenário faz com que os métodos de exploração de dados se tornem mais complexos, permitindo que tais conhecimentos sejam obtidos por processos organizados de transformação de dados em informações, o que é chamado de *Knowledge Discovery in Databases* (GAMARRA; GUERRERO; MONTERO, 2016; OLIVEIRA et al., 2017).

Dessa forma, a fim de buscar padrões nos dados, é necessária a utilização de diversas ferramentas computacionais, que são utilizadas para a etapa de extração de conhecimento, transformando dados em informações. Tais ferramentas baseiam-se em técnicas como análise de *cluster*, regressão, indução, classificação, redes neurais, e outros diversos tipos de algoritmos para a extração de informações implícitas (KAVAKIOTIS et al., 2017). Esse processo de descoberta é chamado de mineração de dados (*Data Mining*) (LU; SETIONO; LIU, 1996; SHU et al., 2017; DUTT; ISMAIL; HERAWAN, 2017).

No presente estudo foi realizada a análise de dados de uma empresa de prestação de serviço público, aplicando uma das técnicas de *Data Mining*, a

clusterização de dados, para a extração de *insights* perante a complexidade das regiões onde estão os locais de atendimento ao público da empresa.

A clusterização de dados possui variadas aplicações, sendo útil em outros setores, além do setor de prestação de serviço público. Martins (2017) exemplificou a adoção de um modelo de clusterização de dados, por meio da identificação de *clusters* de opiniões sociais, que são usualmente formadas por interações em redes sociais e em “fóruns/*threads*”. Cohrs et al. (2013) utilizaram a análise de *cluster* para identificar grupos de indivíduos com características socioeconômicas e agravos ou vulnerabilidades semelhantes, o que possibilitou realizar o desenho de uma melhor promoção de saúde na localidade de estudo. Já Silva et al. (2019a) realizaram a clusterização de dados para dar apoio à gestão pesqueira em diferentes escalas em Tocantis. A partir da aplicação do algoritmo *k-prototypes*, foi evidenciado que as variáveis numéricas mais expressivas no estudo foram o rendimento da pesca e potência do motor da embarcação.

A análise de *cluster* é, portanto, uma técnica comumente utilizada em várias situações para a descoberta de grupos e identificação de padrões e de distribuições, buscando o entendimento dos dados para a melhor correlação entre os mesmos (KAVAKIOTIS et al., 2017; RISTOSKI; PAULHEIM, 2016; FENG; BARBOSA; TORRES, 2016).

1.2 OBJETIVOS

1.2.1 Objetivo geral

O objetivo geral deste trabalho é desenvolver e aplicar um algoritmo de clusterização de dados baseado no método K-Means, contribuindo para a melhoria do processo de tomada de decisão em uma empresa de fornecimento de energia elétrica.

1.2.2 Objetivos específicos

- i. Extrair e tratar dados do sistema ERP da organização, possibilitando sua posterior utilização no algoritmo de clusterização;
- ii. Construir algoritmo em Python (Linguagem de programação) baseado no método K-Means e usá-lo para agrupar os dados;

- iii. Desenvolver gráficos e imagens, a fim de contribuir, futuramente, para a melhoria da gestão à vista e do processo de tomada de decisão da equipe.

1.3 JUSTIFICATIVA

A prestação de serviços públicos objetiva satisfazer as necessidades coletivas da população para quem presta serviço e conforme crescimento da mesma, suas necessidades crescem simultaneamente, gerando assim um aumento das atividades essenciais voltadas para o atendimento desse público (REZENDE; RIBEIRO, 2018).

Mediante tal cenário, as organizações prestadoras de serviços públicos oferecem uma grande gama de serviços, que podem ser realizados em diversos canais, entre eles, canais digitais (sites e aplicativos), canais telefônicos e canais presenciais. Nessa ótica os canais presenciais, tratados nesta pesquisa como sendo agências de atendimento, têm grandes fluxos diários de clientes e, assim, requerem uma enormidade de opções de serviços disponíveis (FERLIN; REZENDE, 2019).

Esses serviços são transformados em dados que são armazenados no sistema da organização, a fim de serem tratados e executados posteriormente. Um exemplo é o sistema ERP, que é comumente utilizado (SOUZA, 2000). Cada tipo de serviço apresenta particularidades e, como a meta geral das organizações é obter um nível de serviço adequado, existem métricas que calculam a qualidade de tais dados (OLIVEIRA; TAMAYO, 2004).

Em vista de se obter resultados expressivos em relação aos objetivos da organização, os serviços tendem a ser otimizados, entretanto devido à grande quantidade de diferentes serviços, torna-se difícil estabelecer planos de ações específicos, conseqüentemente os estudos e métodos de agrupamentos de dados são frequentemente utilizados (GAMARRA; GUERRERO; MONTERO, 2016; OLIVEIRA et al, 2017).

As tomadas de decisão podem ser otimizadas conforme são utilizados métodos de agrupamentos de dados, sendo o K-means, utilizado neste trabalho, um método extremamente eficiente em vários tipos de situações vivenciadas no ambiente corporativo (ARORA et al., 2016; YANG et al., 2017a).

2 REFERENCIAL TEÓRICO

2.1 TECNOLOGIAS DE INFORMAÇÃO

O impacto das tecnologias de informação é evidenciado a cada dia, o que é sentido na casualidade do cotidiano da população mundial. Tal situação também é vista como sendo sinônimo de produtividade nas organizações globais. As novas tecnologias geram novos hábitos e novas formas de sociabilidade na população. Nas organizações, conforme as tecnologias são integradas, a rotina tende a mudar, seja na alta gestão ou no chão de fábrica (TRANTOPOULOS, 2017; AASHEIM; LI; WILLIANS, 2019).

Visto que os sistemas de informação têm se tornado mais populares, as empresas usam e buscam cada vez mais novos mecanismos de gestão e de criação de soluções para seus desafios e para o atendimento de suas metas. Nesse sentido é comum a busca por *softwares* e pessoal qualificado no ramo de dados (RONQUILLO; ZUCKERMAN, 2017).

2.1.1 Sistemas ERP - Enterprise Resource Planning

Os Sistemas Integrados de Gestão, conhecidos como ERP (*Enterprise Resource Planning*), podem ser definidos como sistemas de informação integrados, obtidos como pacotes de *software* comerciais (AYYAGARI, 2019). O objetivo principal desses sistemas é dar suporte às operações de uma organização nas áreas de manutenção, suprimentos, contabilidade e recursos humanos (CHIESA, 2004).

Os sistemas ERP possibilitam controle dos recursos da empresa utilizados na produção, comercialização, distribuição e gestão. Controlando assim a administração dos recursos utilizados nas áreas de materiais, pessoas e equipamentos (PADILHA; MARINS, 2005).

Esses sistemas são utilizados em conjunto por diferentes departamentos dentro de uma mesma organização, em oposição a um conjunto de sistemas que atendem isoladamente a cada um deles (AYYAGARI, 2019; PADILHA; MARINS, 2005). Essa integração oferecida permite a verificação cruzada de informações entre diferentes partes do sistema, e o compartilhamento de informações comuns entre os

diversos módulos, de maneira que cada informação seja alimentada no sistema unitariamente e sem perdas (CHIESA, 2014; AYYAGARI, 2019).

2.1.2 Big Data

A evolução gradual das tecnologias de informação e de comunicação é evidenciada ao longo das revoluções industriais, sendo a Quarta Revolução Industrial, a Indústria 4.0, a mais impactante no que tange o conceito de armazenamento e compartilhamento de dados (MAGLIO; LIM, 2016; WOLFERT et al., 2017). As tecnologias que surgiram nessa revolução mudaram o conceito dos sistemas empresariais e de automação, visto que conceitos como Internet das Coisas, Big Data, Computação em Nuvem e Fábricas Inteligentes surgiram e estão cada vez mais presentes no cotidiano das corporações (STERGIOU et al., 2018; BOTTA et al., 2016; VARGHESE; BUYYA, 2018).

Consequentemente à popularização e dependência desses novos recursos, o barateamento das máquinas computacionais e dispositivos de armazenamento e captura de dados tornou exponencial a criação de dados que são disponibilizados em diversos canais, resultando assim na formação da “Era do Big Data” (MAGLIO; LIM, 2016; WOLFERT et al., 2017).

“Big Data” é o conceito que se refere ao *dataset* (conjunto de dados) que tem grande volume de dados e também detém uma grande taxa de crescimento, fazendo com que tecnologias típicas sejam incapazes ou inviáveis, pelo tempo de construção de um determinado resultado, de gerenciar e analisar tais bases de dados (YANG et al., 2017b). O dimensionamento da quantidade de dados do “Big Data” é impreciso, visto que a geração de dados cresce de forma exponencial todos os anos (ZHANG et al., 2017; YANG et al., 2017a). Diversos setores da sociedade são afetados por tal transformação digital, as instituições bancárias, por exemplo, são constantemente envolvidas nesse fenômeno. Segundo Yang et al. (2017b), foi possível notar que nas últimas décadas a mudança no relacionamento do cliente com as instituições bancárias foi influenciada por essas inovações, visto que em 2016 as transações bancárias na internet e na plataforma *mobile* saltaram para 71% de representatividade.

Porém, a linha de produção da construção de *insights* e informações importantes depende da estrutura de como os dados são armazenados, uma vez que a quantidade exorbitante de dados não é passível de ser aproveitada se os

dados forem mal manuseados (ZHANG et al, 2017; YANG et al, 2017a). Nesta situação são necessárias tecnologias, métodos e ferramentas que propiciam o armazenamento, mineração e apresentação do enorme volume de dados, como por exemplo bancos de dados, métodos de agrupamento de dados (clusterização), métodos de *data mining*, *machine learning* e *softwares* de manipulação de dados.

2.2 EXTRAÇÃO DE CONHECIMENTO EM BASES DE DADOS

Proposta por Fayyad em 1989, a construção do conhecimento a partir de informações extraídas de dados, obtidos por técnicas mineração de dados, é constituída por um processo, conhecido como extração de conhecimento em bases de dados (*Knowledge Discovery in Databases – KDD*). Nesse processo são necessárias as seguintes etapas para a extração de dados corretamente estruturados: a seleção de dados, o processamento de dados e a transformação de dados (ZAHARIA et al., 2016; QIU et al., 2016).

O processo em questão é construído por diversas áreas presentes na computação, estatística e matemática, como *Machine Learning* (Aprendizado de Máquinas), *Artificial Intelligence* (Inteligência Artificial), estatística, visualização de dados, entre outras (ZAHARIA et al., 2016; QIU et al., 2016; FAYYAD et al., 1996). As técnicas comumente usadas são indução de regras, algoritmos genéticos e redes neurais artificiais, que objetivam construir conhecimento a partir de conjuntos de dados e utilizá-lo em processos de tomada de decisão (ZAHARIA et al., 2016; QIU et al., 2016; FAYYAD et al., 1996).

2.2.1 Seleção de dados

O processo de seleção de dados objetiva a identificação da informação, realizada pela verificação das origens internas e externas, extraindo assim um subconjunto de dados, o que é necessário para que a aplicação do método de mineração de dados seja corretamente orientada. Na seleção utilizam-se apenas os dados relevantes aos objetivos do processo de extração das informações para construção de conhecimento (GU et al., 2016; WANG et al., 2018).

Segundo Diniz e Louzada Neto (2000), existem dois tipos de classificação de variáveis: categórica e quantitativa. As variáveis categóricas representam valores finitos e são nominais ou ordinais. As ordinais são conhecidas por apresentarem

ordem em seus valores possíveis e podem ser citadas como grau de instrução (“primeiro grau”, “segundo grau”), já as nominais são classificadas como, por exemplo, altura (“baixo”, “alto”) e sexo (“masculino”, “feminino”). As variáveis quantitativas assumem valores numéricos e podem ser contínuas, valores reais, ou discretas, valores finitos ou infinitos. Como exemplo de variável discreta pode ser citado o número de filhos, e para variáveis contínuas, taxas.

É imprescindível que a etapa de seleção dos dados seja realizada de forma alinhada aos objetivos do processo de extração de informações para construção de conhecimento, portanto, o conjunto de dados gerado deve apresentar características necessárias para o alcance dos objetivos. Conclui-se que essa etapa deve otimizar o tempo de processamento da técnica de mineração (GU et al., 2016; WANG et al., 2018; DINIZ; LOUZADA NETO, 2000).

2.2.2 Processamento de dados

O processamento de dados inicia-se com a eliminação de dados duplicados, e também com a exclusão de *outliers* (valores significativamente fora do esperado para uma variável) e *missing values* (valores que não estão presentes no conjunto selecionado e valores inválidos que foram eliminados durante a detecção de *outliers*). Nessa etapa, portanto, utilizam-se métodos estatísticos e de visualização de dados para estruturá-los com qualidade (KWAK; KIM, 2017).

Para o processo final de tomada de decisão a qualidade do *dataset* é de extrema importância visto que as decisões devem ser acertadas o máximo possível, para isso a presente etapa deve assegurar a qualidade dos dados (YANG et al., 2017b).

2.2.3 Transformação de dados

A transformação de dados busca a redução do tempo de processamento na etapa de mineração de dados, objetivando a redução do número de exemplos e mantendo as características do *dataset* (GU et al., 2016; WANG et al., 2018; DINIZ; LOUZADA NETO, 2000; YANG et al., 2017a).

Podem ser utilizadas técnicas de discretização que buscam a redução do número de atributos, retirando subconjuntos de forma que não impacte na qualidade do conjunto final e na redução do número de valores de um atributo. A discretização

constitui a substituição de atributos contínuos por discretos e a suavização de valores agrupa atributos em um valor numérico (WANG et al., 2018; DINIZ; LOUZADA NETO, 2000; YANG et al., 2017a).

2.3 DATA MINING (MINERAÇÃO DE DADOS)

A mineração de dados visa à extração de informações de bases de dados, por meio de métodos computacionais. Esse processo tem crescido rapidamente ao longo dos anos motivado pela existência de necessidades práticas, sociais e econômicas de tomar decisões perante grande volume de dados (ZAHARIA et al., 2016; QIU et al., 2016; FAYYAD et al., 1996).

As próximas seções apresentam brevemente algumas técnicas utilizadas em processos de mineração de dados.

2.3.1 Estatística

Segundo Diniz e Louzada Neto (2000), a mineração de dados deve ser extremamente interligada com as técnicas de análises estatísticas a ponto de ser tratada como uma adequação das mesmas, sempre propiciando análise de grandes bases de dados.

Como exemplo de técnicas estatísticas podem ser citadas e brevemente descritas as seguintes.

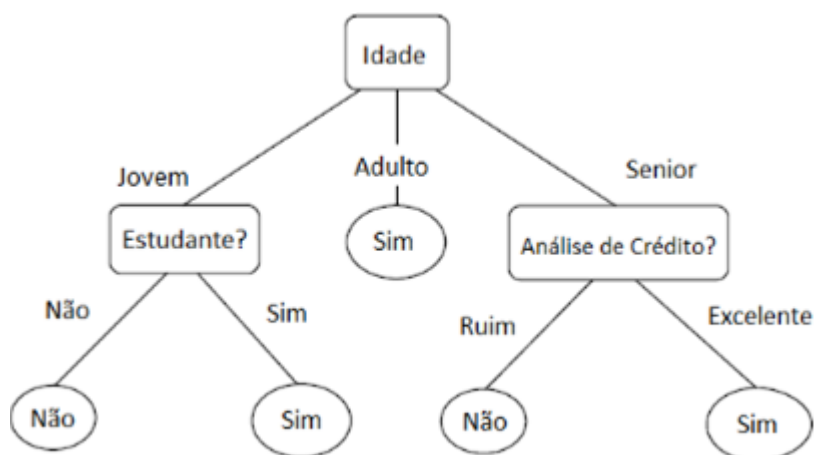
- i. Análise de regressão: modelo que prevê valores de variáveis conforme as relações do histórico de variáveis.
- ii. Estatística descritiva: são medidas de variabilidade, de média, de mediana, de frequências e porcentagens, tabelas de contingência e de coeficientes de correlação.
- iii. Análise fatorial: modelo que busca identificar relações estruturais entre variáveis e reduzir o seu número.
- iv. Técnicas de visualização: modelos com base em histogramas e gráficos de dispersão.

2.3.2 Indução

O modelo de indução é descrito como um processo de obtenção de hipóteses trabalhadas e ramificadas em uma árvore de decisão estruturada, que tem como entrada o *dataset* construído a partir de dados e fatos existentes (YOO et al., 2016; DINIZ; LOUZADA NETO, 2000).

O *dataset* inicial passa por uma sequência de pontos de decisão, representados como os nós de ramificação. Essa sequência é representada por uma árvore de decisão que consiste em uma sequência lógica, em que os dados são levados de nó em nó até que se estabeleça um resultado (YOO et al., 2016; DINIZ; LOUZADA NETO, 2000). Na Figura 1 é apresentado um exemplo de árvore de decisão.

Figura 1 - Exemplo de árvore de decisão



Fonte: elaborado pelo autor

2.3.3 Algoritmos genéticos

Os métodos fundamentados em algoritmos genéticos são elaborados a fim de obter segmentações e classificações apuradas, a partir de um processo iterativo que evolui uma população (*dataset*) inicial por meio de processos que simulam processos naturais de evolução (SUCH et al., 2017; DINIZ; LOUZADA NETO, 2000).

Os algoritmos utilizam processos de cruzamento, seleção e mutação entre variáveis, a fim de desenvolver gerações de soluções, que podem apresentar maior poder de previsão, produzindo uma solução ótima (SUCH et al., 2017; DINIZ; LOUZADA NETO, 2000).

2.3.4 Classificação

Segundo Rezende (2003), a classificação tem o objetivo de criar classes, por meio de um modelo criado e aplicado junto a dados não classificados. Exemplos dessa técnica são: classificação de pedidos de clientes de seguros de vida como baixo, médio e alto risco, renda de uma determinada população e criminalidade de bairros de determinadas cidades.

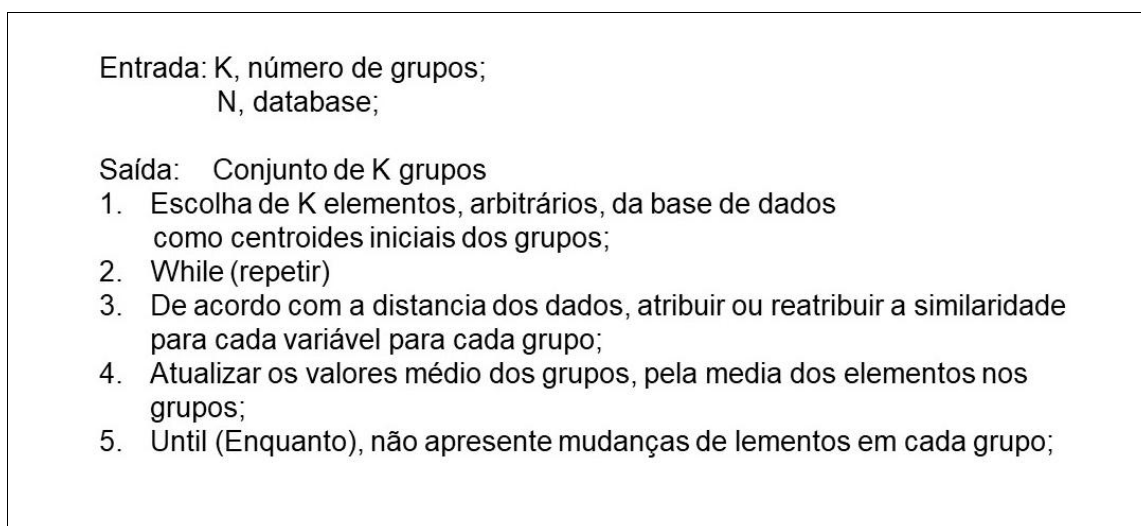
2.3.5 Análise de *cluster*

A análise de *cluster* objetiva a segregação por similaridade de variáveis de uma população heterogênea, formando assim diversos grupos que podem ser definidos antes da aplicação da técnica (SILVA et al., 2019b, DINIZ; LOUZADA NETO, 2000).

2.3.5.1 Método k-means

O método K-means toma como parâmetro de entrada uma variável de representação da quantidade de agrupamentos dos dados, tal variável é representada por K. Já a variável I é a quantidade de iterações que o algoritmo necessita para dar uma resposta conclusiva (JAIN et al., 1999). Nesse sentido, o conjunto de dados de N elementos é particionado em K grupos (Figura 2).

Figura 2 - Algoritmo k-means



Fonte: elaborado pelo autor

A medida de distância é feita de cada variável para a média do grupo em que está contida, sendo calculada pela distância Euclidiana, de acordo com a Equação 1 (JAIN et al., 1999; DINIZ; LOUZADA NETO, 2000).

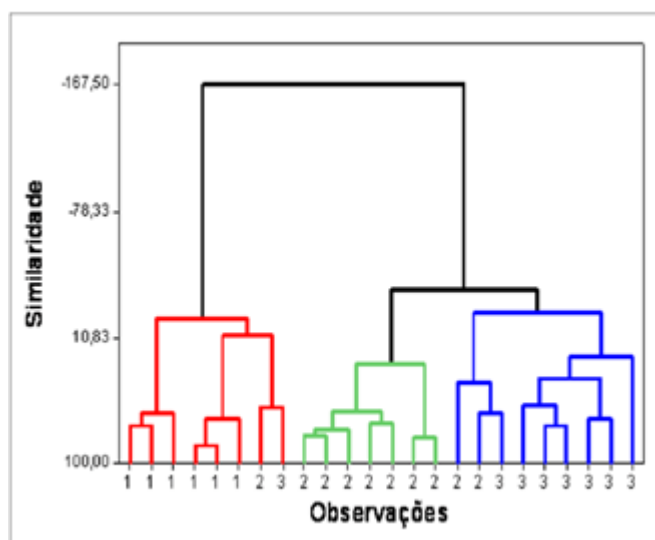
$$D = \sqrt{\sum_i^w (p_{ai} - p_{ji})^2}$$

(1)

Onde p_{ai} e p_{ji} representam as variáveis comparadas, sendo elas a média do grupo e a respectiva variável testada. Sendo D o somatório dos quadrados das distâncias realizadas para todas as variáveis do grupo, com $i = 1, 2... w$. Sendo w o número total de variáveis presentes no grupo.

O cálculo é realizado até que a verificação da distância seja não representativa e, assim, não exista a necessidade de realocação das variáveis nos grupos. Existem muitas formas de se representar graficamente os *clusters* formados, as principais formas são os gráficos de dispersão, dendogramas e os gráficos de correlação. A Figura 3 apresenta um exemplo de dendograma.

Figura 3 - Exemplo de dendograma



Fonte: elaborado pelo autor

Esse método apresenta algumas características chave como sensibilidade a ruídos, visto que caso um valor extremamente alto seja inserido no algoritmo, o método pode distorcer o resultado e formar grupos esféricos (DINIZ; LOUZADA NETO, 2000). O número de grupos é o mesmo durante todo o processo, entretanto pode-se mudá-lo inicialmente (DINIZ; LOUZADA NETO, 2000).

2.4 SERVIÇOS PÚBLICOS

Em benefício da coletividade e da utilização da arrecadação de tributos, o serviço público é o modo como a população tem suas necessidades sociais atendidas. Os direitos sociais e a dignidade do cidadão envolvem o fornecimento de energia elétrica, saneamento básico, transporte e outros. Nesse sentido o fornecimento de energia elétrica é essencial (SANTIN, 2019).

Essa prestação de serviços pode ser feita por meio de uma economia exclusivamente estatal ou privada, ou ainda mista e simultânea entre setores privado e público, este último com uma participação acionária relevante (CESAR et al., 2012). Para a regulação da qualidade desses serviços existem órgãos reguladores, como a Agência Nacional de Energia Elétrica (ANEEL), que é o órgão que regula as concessões de distribuição e geração de energia.

O contexto da presente pesquisa é de uma empresa privada que fornece serviços públicos e tem a regulação de um órgão estatal. Para uma melhor

produtividade e lucratividade existe a constante necessidade da adição das melhores práticas e dos melhores profissionais no setor. Sendo assim o tratamento e análise de dados é essencial neste ramo.

3 METODOLOGIA

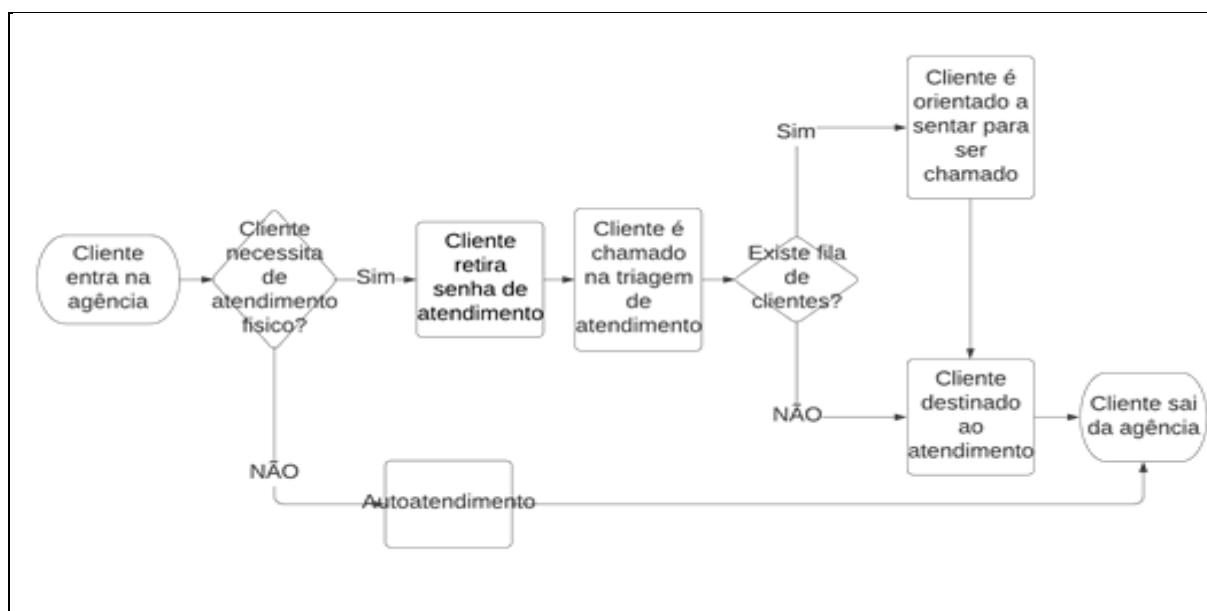
3.1 CARACTERIZAÇÃO DA EMPRESA

A empresa é uma distribuidora de energia elétrica que opera no estado do Espírito Santo, é uma empresa de capital aberto, operando na bolsa de valores. A empresa realiza a distribuição de energia para 70 municípios do estado. Dessa forma é regulado que devem existir agências de atendimento nessas localidades. Segundo a estimativa do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), o estado do Espírito Santo, em 2019, tinha população estimada de 4.018.650. Portanto tornou-se um desafio para a gestão da concessionária fornecer um serviço de qualidade para todos os seus clientes.

A empresa tem seus serviços de atendimento presencial terceirizados e conta com uma base centralizada com o objetivo de controlar os indicadores de qualidade de atendimento. Nessa ótica, o sistema de gestão é peça fundamental. Para tal gestão utiliza-se um sistema ERP que permite que os atendentes terceirizados nas agências gerem notas de atendimento para cada serviço prestado ao cliente. Os serviços são efetuados no prazo estabelecido, após aprovação.

Juntamente com o ERP, existe outra empresa terceirizada que fornece o sistema de controle de filas das agências. Essa empresa controla o fluxo do cliente dentro das agências, com início no momento em que o cliente retira uma senha de atendimento até o momento em que é atendido e sai da agência. Essa empresa disponibiliza todas as variáveis que são produzidas no processo de atendimento como TMA, TME, demanda mensal e quantidade de atendimento por cada tipo de serviço. A Figura 4 apresenta o fluxograma de atendimento ao cliente.

Figura 4 - Fluxograma de atendimento ao cliente.



Fonte: elaborado pelo autor.

3.2 IDENTIFICAÇÃO DO PROBLEMA

A identificação do problema deu-se a partir da dificuldade, por parte do gestor operacional e da equipe de trabalho, no processo de tomada de decisão, que consiste na escolha e no direcionamento de recursos, conforme alinhado com a alta gestão, para a intensificação da melhoria dos indicadores de qualidade das agências de atendimento.

Os recursos são limitados, perante as necessidades das 69 agências de atendimento, e, somado às particularidades de cada agência, cada público forma uma persona de cliente. Tais particularidades são representadas por meio de dados, como quantidade de atendentes, TME, TMA, e tipos de serviços. Cada uma das agências, portanto, necessita de recursos específicos e diferenciados para que a meta de qualidade seja batida.

Os dados são armazenados no sistema de gestão integrado em grande escala e não totalmente classificados. Assim verificou-se a necessidade de coletar, transformar e agrupar tais dados a fim de criar um mecanismo que melhorasse a gestão à vista, para assim, otimizar a tomada de decisão da equipe acerca da destinação de recursos para as agências de atendimento ao público.

3.3 CARACTERIZAÇÃO DA PESQUISA

Esta pesquisa foi realizada no período de estágio do ano de 2019. A caracterização das etapas realizadas na pesquisa está representada na Figura 5.

Quanto à abordagem, a pesquisa classifica-se como quantitativa, pois para atingir seu objetivo, foi necessário agrupar os dados referentes ao atendimento presencial das agências, a fim de detalhar como uma agência funciona no âmbito de atendimento ao público. Segundo Rossatto et al. (2008), a pesquisa quantitativa tem o caráter de controlar ao máximo possível o contexto, minimizando variáveis interferentes e irrelevantes. A neutralidade e objetividade do pesquisador são evidenciadas juntamente com a análise dos dados coletados por meio de métodos matemáticos.

A pesquisa também se define como descritiva, uma vez que buscou-se realizar a descrição da realidade vivenciada no local de pesquisa, e, também, agregar valor pelo cruzamento de dados existentes objetivando gerar informações estratégicas para fundamentar tomadas de decisão. Para Vieira (2002), uma pesquisa descritiva pretende analisar e interpretar a realidade sem nela interferir para modificá-la, visando à descoberta de fenômenos presentes e sua respectiva frequência e estrutura.

Quanto aos meios, a pesquisa caracteriza-se como um estudo de caso, pois se buscou realizar o entendimento da situação presente na empresa objeto de estudo, e assim levantar e estudar os dados pertinentes ao atendimento ao público. Objetivando a obtenção de *insights* que poderão ser utilizados em futuras tomadas de decisão, ao mesmo tempo que o processo de análise e estudo não interferiu no processo de estudo. Segundo Ponte (1994), um estudo de caso é uma estratégia que busca realizar a análise de fenômenos reais considerando o contexto e as variáveis que os influenciam.

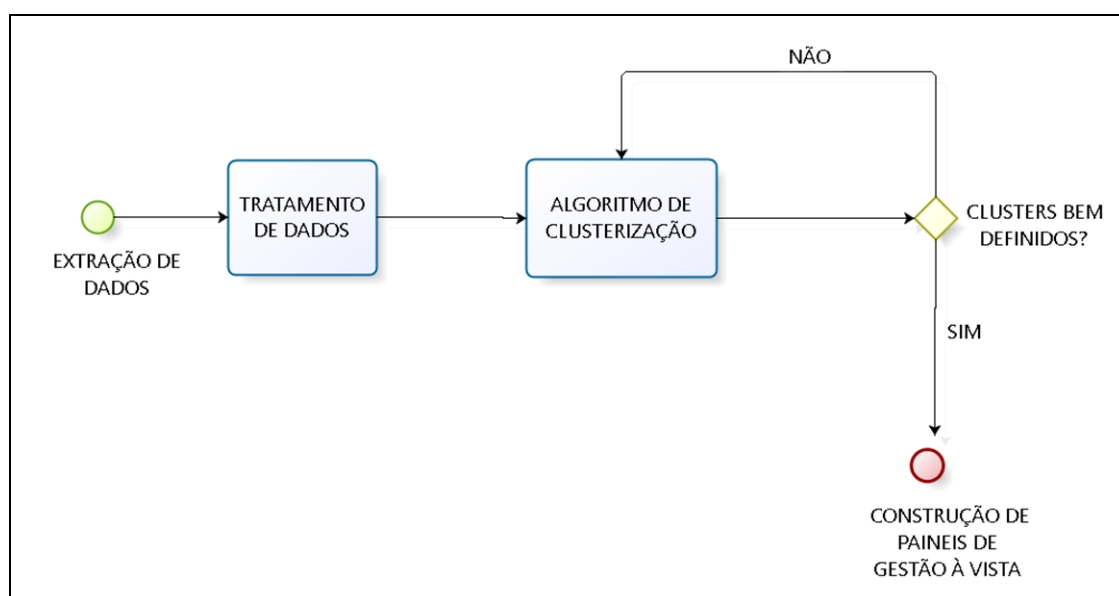
A pesquisa foi dividida em etapas, conforme a Figura 5, sendo o início a extração de dados, em que se buscou a coleta de dados não estruturados de todo o atendimento da empresa objeto de estudo. A segunda etapa é o tratamento de dados, em que foi necessária a estruturação dos dados iniciais a fim de organizá-los para etapas posteriores.

A terceira etapa foi a construção do código de clusterização e sua posterior utilização com base nos dados anteriormente coletados e tratados. Para a quarta

etapa foi analisada a eficiência do algoritmo perante os *clusters* formados, sendo que caso o número de *clusters* não estivesse bem definido, os dados seriam mais uma vez utilizados no algoritmo com ajustes nos parâmetros iniciais.

Para a quinta etapa foram propostas ferramentas de gestão à vista, como gráficos de *clusters*, que objetivam a visualização de como os *clusters* foram formados ao passo que as características similares foram evidenciadas.

Figura 5 - Etapas da pesquisa



Fonte: elaborado pelo autor

3.4 COLETA DE DADOS

A coleta de dados foi iniciada a partir da montagem da estrutura de dados desejada, selecionando os serviços e atributos de cada agência que era necessária para a montagem das variáveis propostas. As agências, por questões de anonimato foram numeradas de 1 a 69, e foram denominadas pelo respectivo número.

Dessa forma foi realizada a extração de relatórios mensais contendo os dados relativos à pesquisa, possibilitando sua posterior utilização no trabalho.

3.5 ANÁLISE DE DADOS E CLUSTERIZAÇÃO DE DADOS

Neste tópico é apresentado como foi utilizado o algoritmo k-Means com a linguagem de programação Python. É também apresentado como as aplicações de dados são benéficas para as organizações que buscam resultados consolidados.

Para o método requerido foi utilizada a biblioteca *Scikit-Learn*, que é uma biblioteca de aprendizado de máquina de código aberto para a linguagem de programação Python. Também foi utilizada a biblioteca de análise de dados pandas, que é uma biblioteca de *software* criada para a linguagem Python para manipulação e análise de dados.

Para a construção do algoritmo foi utilizado uma sequência de comandos, os comandos 1 e 2 tiveram como objetivo a construção do *dataset* inicial, conforme apresentado abaixo.

1. `dataset = pandas.read_excel('Caminho do Arquivo')`: importa os dados requeridos e os transforma em um *dataset*, propiciando um melhor manuseamento de dados.
2. `X = dataset.iloc[:, 0:42].values`: armazena os valores na variável X.

Inicialmente os dados apresentados não estavam de forma adequada para serem utilizados na técnica de clusterização proposta, pelo fato de não haver padronização dos dados, uma vez que não apresentavam uma escala definida. Portanto foi necessário utilizar um método de padronização de dados.

Para as próximas etapas foram utilizados alguns parâmetros importantes, que são brevemente explicados a seguir: i) o parâmetro `n_clusters` representa a quantidade de grupos em que os dados são agrupados, o padrão deste parâmetro, caso não seja especificado, é 8; ii) parâmetro `init` refere-se como o algoritmo é inicializado, podendo ser `K-means++`, `random` ou `ndarray`; iii) parâmetro `n_jobs` configura a quantidade de CPU's utilizadas; iv) parâmetro `algorithm` refere-se à versão do método K-Means utilizado.

O passo subsequente foi a inicialização do método k-means, configurado inicialmente com cinco *clusters*, podendo ser ajustado futuramente, e de forma a selecionar os centroides aleatoriamente, com o comando:

3. `kmeans = KMeans(n_clusters = 5, init = 'random')`.

Posteriormente o método `fit()` foi utilizado a fim de agrupar os dados e disponibilizar cada valor dos atributos especificados anteriormente, tais dados estavam armazenados na variável `X`, anteriormente construída. Esse passo foi representado pelo comando:

4. `kmeans.fit(X)`

Como próximo passo, por meio do comando 5, foi possível visualizar os centroides gerados por meio do atributo `cluster_centers_`. Conforme os centroides resultantes, foi gerado um *array* (vetor), com cinco instâncias e os valores de seus atributos.

5. `kmeans.cluster_centers_`

Pelo fato do método K-Means se basear em um mecanismo que agrupa os dados conforme a afinidade dos mesmos, foram realizados cálculos para a contabilização das distâncias entre variáveis, e para tal, o comando:

6. `distance = kmeans.fit_transform(X)`, disponibiliza as distâncias entre dados na variável *distance*.

Em razão do agrupamento ser feito com base em cinco *clusters*, foi disponibilizado, pelo comando anterior, cinco valores, que equivalem às distâncias entre os dados de cada *cluster*. Para a verificação das instâncias criadas pelo método de clusterização foi utilizado o comando:

7. `labels = kmeans.labels_`, que cria uma variável `labels`, que representa a instância das atribuições dos dados, variando de 0 a 4.

Como método de visualização de dados, para a melhor identificação do número de *clusters* corretos, foi utilizado o gráfico *Elbow*, traduzido como gráfico do cotovelo. Portanto foi utilizada a biblioteca `matplotlib`, que permite a construção de gráficos a partir de dados passados anteriormente.

No gráfico *Elbow* foi plotada, para cada número de *clusters*, a soma da variância dos dados, com o intuito de mostrar até qual *cluster* existem ganhos. Nesse gráfico deve-se procurar a transição do número de *clusters* que representa o começo da estabilidade da variância dos dados, e que pode ser visualizado como

sendo a protuberância de um cotovelo. Dessa forma, a partir do *cluster* de número três a variância já não é tão notada. O método para a implementação do gráfico é apresentado na Figura 6.

Figura 6 – Código para implementação do gráfico *Elbow*

```
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'random')
    kmeans.fit(X)
    print(i,kmeans.inertia_)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('O Metodo Elbow')
plt.xlabel('Numero de Clusters')
plt.ylabel('WSS') #within cluster sum of squares
plt.show()
```

Fonte: elaborado pelo autor.

4 RESULTADOS E DISCUSSÃO

4.1 EXTRAÇÃO DE DADOS

Foram coletados dados das 69 agências de atendimento durante três meses. Os dados foram classificados por meio de variáveis pré-definidas com base nos indicadores utilizados na empresa (demanda mensal média, tempo médio de espera, tempo médio de atendimento, contas faturadas, quantidade de equipamentos de vídeo atendimento, quantidade de atendentes, quantidade média de acordos de pagamentos, quantidade média de alteração/rescisão de contratos, quantidade média de alteração de endereço de entrega, quantidade média de alteração de dados de parceiros, quantidade média de análises de viabilidade de projetos, quantidade média de baixas provisórias, quantidade média de adesões a débito automático, quantidade média de ligações novas, quantidade média de declarações de quitações, quantidade média de desligamentos a pedido, quantidade média de emissão de faturas, quantidade média de pedidos de falta de energia, quantidade média de faturamentos indevidos, quantidade média de informações comerciais, quantidade média de inspeções de irregularidades, quantidade média de ligações com instalações inexistentes, quantidade média de ligações provisórias, quantidade média de análises técnicas, quantidade média de modificações em instalações, quantidade média de adesões a envio de contas por e-mail, quantidade média de ordens de restituições, quantidade média de pedidos de indenizações, quantidade média de podas de árvores, quantidade média de pedidos de verificações de qualidade recebida de serviço, quantidade média de religações, quantidade média de ressarcimentos de danos, quantidade média de manutenções de rede).

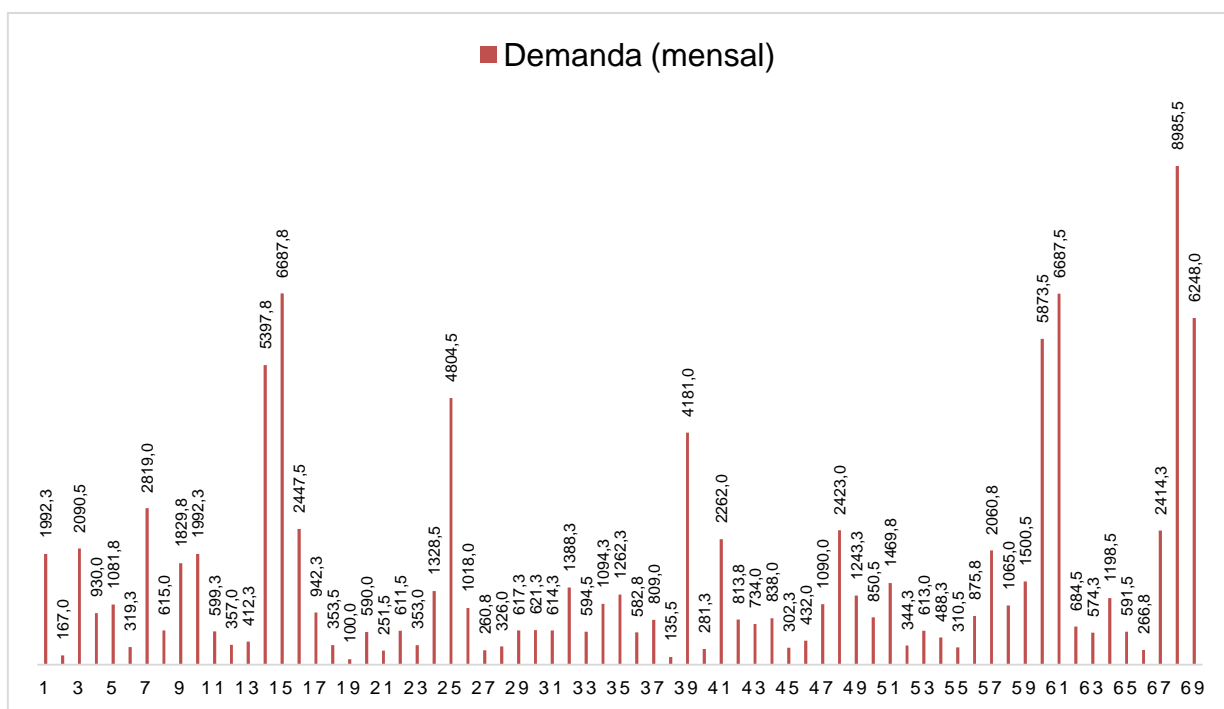
Como o presente trabalho visa realizar a clusterização de dados conforme todos os indicadores apresentados anteriormente, foram retirados do sistema ERP os dados de cada um dos indicadores durante três meses, resultando nas Tabelas apresentadas no Apêndice 1. A extração foi feita em bases de dados do sistema ERP da empresa.

Para o melhor entendimento de como os indicadores são tratados e como se comportam, foram construídos gráficos da média dos três meses de extração de dados. Mostrando assim como é difícil a sua classificação em grupos olhando individualmente os indicadores, mesmo que sejam de extrema importância. Os indicadores, demanda mensal média, TME e TMA são vistos como os principais

indicadores de qualidade, portanto são ilustrados como exemplos nos Gráficos 1, 2 e 3, respectivamente.

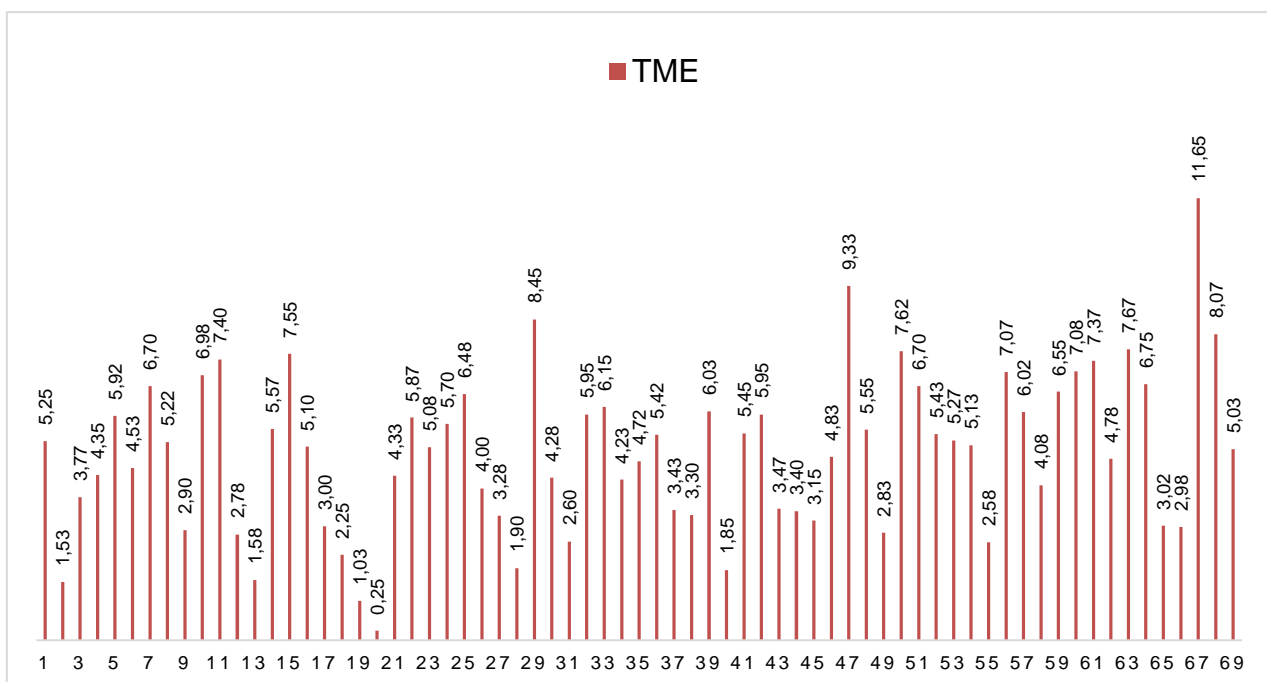
Graficamente os dados dos três principais indicadores não apresentam continuidade ou algum tipo de semelhança, tomando como exemplo, a agência número 1 tem demanda menor que a agência de número 69, mas a agência número 1 tem TME maior que a agência 69. Essas comparações podem ser feitas para todas as combinações de agências, entretanto não será facilmente encontrado algum padrão.

Gráfico 1 – Demanda mensal média de cada agência



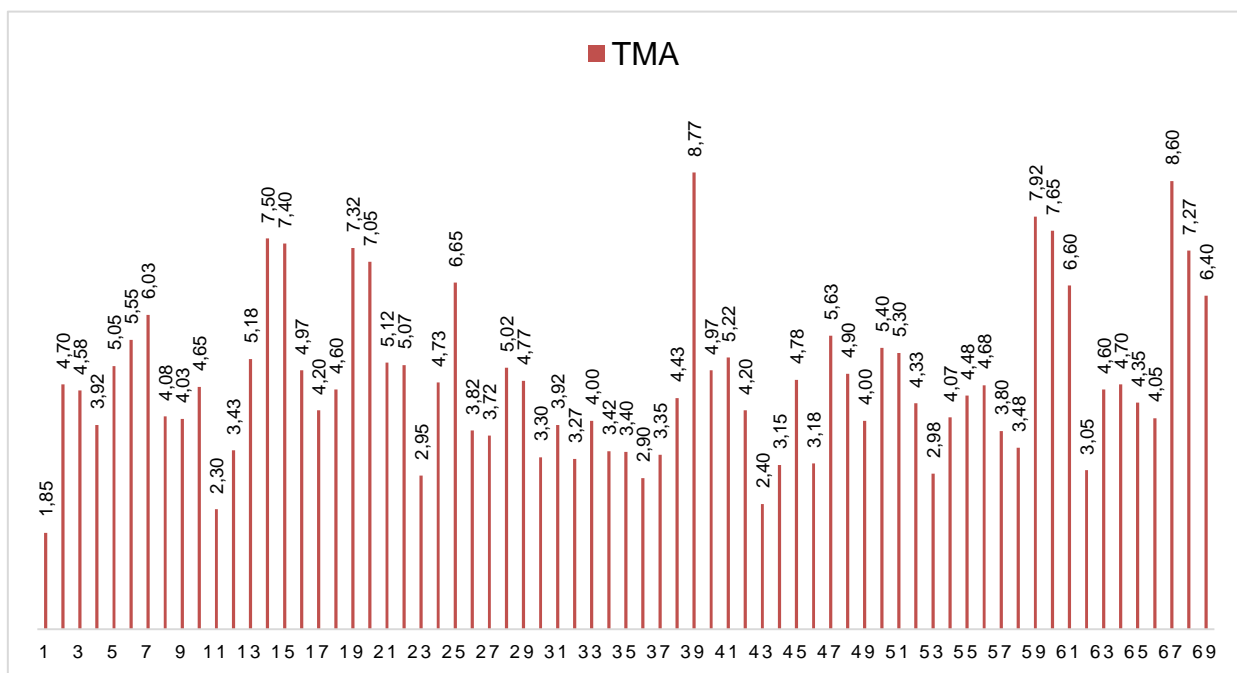
Fonte: elaborado pelo autor.

Gráfico 2 - TME (tempo médio de espera, em minutos) por agência



Fonte: elaborado pelo autor.

Gráfico 3 - TMA (tempo médio de atendimento, em minutos) por agência



Fonte: elaborado pelo autor.

4.2 TRATAMENTO DE DADOS

Para a utilização dos dados iniciais no algoritmo de clusterização foi necessária uma classificação inicial. Assim como os dados iniciais não apresentavam distinção de canais de atendimento, os dados também possuíam nomenclaturas diferentes para serviços prestados semelhantes. Como exemplo os dados “2ª via” e “2ª via de fat.” foram unificados em “Emissão de 2ª via de fatura”.

Dessa forma buscou-se o entendimento de todos os indicadores preliminares e, assim, foram renomeados os indicadores semelhantes e foi reduzida a quantidade de indicadores. Transformando assim a base bruta inicial de dados em um *dataset* apto a ser utilizado no algoritmo.

4.3 APLICAÇÃO DO MÉTODO DE CLUSTERIZAÇÃO K-MEANS

Diante do levantamento e tratamento inicial de dados, foi gerado um *dataset* base, apto para ser inserido no algoritmo de clusterização (Figura 7).

Figura 7 - *Dataset* gerado a partir dos dados do tratamento inicial

	Agências	Demanda (mensal)	TME	...	Ressarcimento Danos	Serviços Internos	Solic. Manut. Rede
0	1	1992.25	5.250000	...	0	8	0
1	2	167.00	1.533333	...	0	1	1
2	3	2090.50	3.766667	...	1	0	2
3	4	930.00	4.350000	...	0	33	1
4	5	1081.75	5.916667	...	0	66	0
...
64	65	591.50	3.016667	...	0	15	0
65	66	266.75	2.983333	...	0	0	0
66	67	2414.25	11.650000	...	0	0	0
67	68	8985.50	8.066667	...	2	322	0
68	69	6248.00	5.033333	...	1	585	0

[69 rows x 42 columns]

Fonte: elaborado pelo autor

Também foi gerado o conjunto de valores dos parâmetros necessários para o algoritmo, a fim de permitir sua correta configuração (Figura 8). Na Figura 8, toma-se o parâmetro *n_clusters* como sendo o número de *clusters* que o algoritmo irá formar, e que foi inicializado com 5. O parâmetro *init* se refere ao modo como o algoritmo será inicializado, sendo inicializado como “*random*” que indica que os centroides dos *clusters* serão gerados de maneira aleatória. O parâmetro *max-iter*, inicializado com 300, refere-se ao número máximo de iterações que o algoritmo poderá fazer. Já o

parâmetro `n_jobs` especifica quantos CPU's (*Central Processing Unit*) serão utilizados para executar o algoritmo. O parâmetro `algorithm` se refere à versão do algoritmo K-Means a ser utilizada. Os demais parâmetros são automáticos, "tol" é a tolerância e "verbose" é a visualização do processo do código em tela.

Figura 8 – Parâmetros de inicialização do algoritmo K-Means

```
KMeans(algorithm='auto', copy_x=True, init='random', max_iter=300, n_clusters=5,
        n_init=10, n_jobs=None, precompute_distances='auto', random_state=None,
        tol=0.0001, verbose=0)
```

Fonte: elaborado pelo autor

A partir da configuração inicial do K-Means (Figura 8), foram agrupados os dados na variável `X` por meio do método `fit(X)`, que recebe como parâmetro os dados a serem agrupados.

Nesta fase, foram agrupados os dados em cinco *clusters*. Para a verificação dos centroides gerados foi utilizado o atributo `cluster_centers_`. Assim foram retornados cinco *arrays* (vetores), cada um representando um *cluster* e cada um com seus respectivos valores de parâmetros iniciais. A representação de um deles está presente na Figura 9.

Figura 9 – Distâncias entre variáveis de um centroide gerado em um *cluster*

```
[[ 3.45000000e+01  5.06418750e+03  6.29166667e+00  7.64166667e+00
  6.92625000e+04  2.50000000e-01  6.00000000e+00  1.13175000e+03
  4.92750000e+02  1.19750000e+02  1.46500000e+03  4.50000000e+00
  1.66750000e+02  2.57500000e+02  4.75000000e+00  2.72500000e+01
  3.50000000e+00  1.06500000e+02  1.30350000e+03  3.00000000e+00
  1.00000000e+00  1.75000000e+01  1.04200000e+03  0.00000000e+00
  1.87250000e+02  1.86000000e+02  5.75000000e+00  6.32500000e+01
  1.24750000e+02  1.87750000e+02  1.82500000e+01  5.25000000e+00
  2.05000000e+01  3.75000000e+00  4.50000000e+00  0.00000000e+00
  1.89750000e+02  1.25000000e+00  4.75000000e+00  2.50000000e-01
  2.01500000e+02  2.25000000e+00]
```

Fonte: elaborado pelo autor

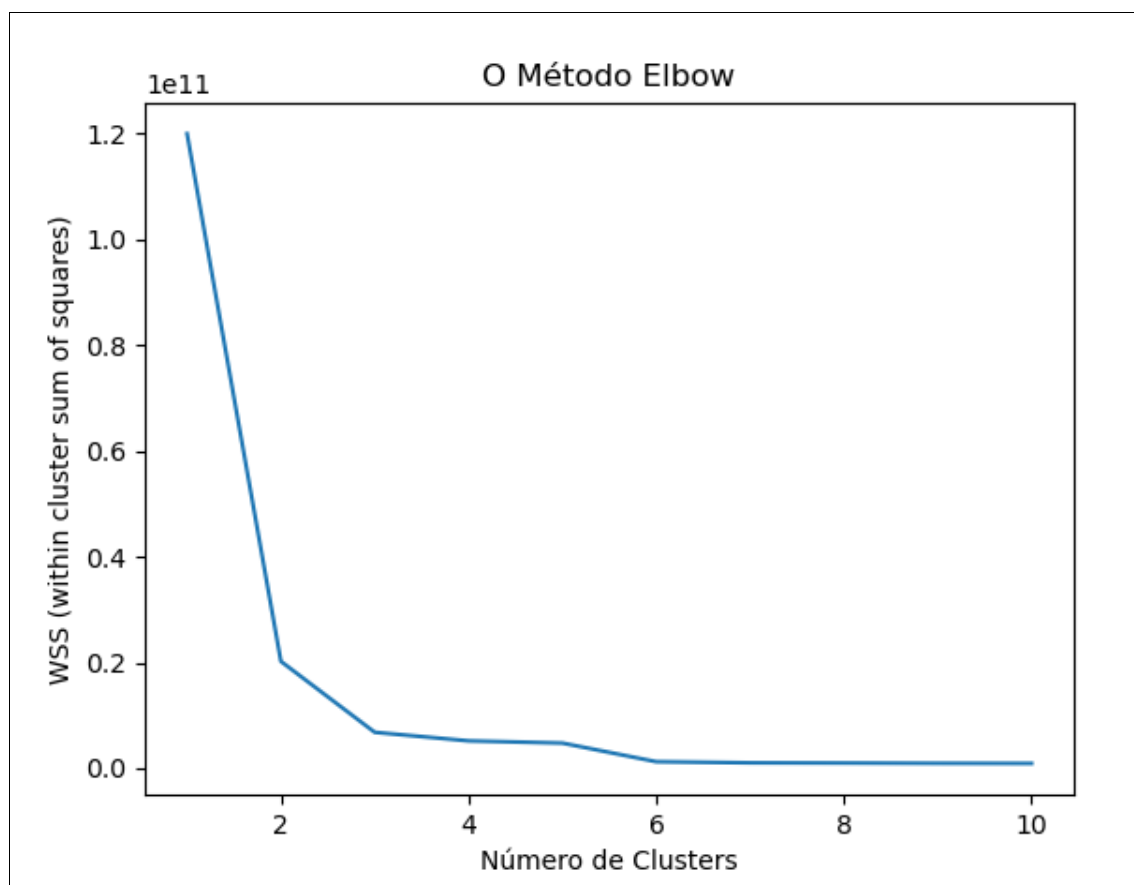
A partir dos *clusters* gerados e cada um com seus valores de parâmetros, foi possível classificá-los em *labels* (rótulos). A menor distância foi classificada com o número 0 e assim por diante, até a maior distância, classificada em 4 (Figura 10).

Figura 10 – Rótulos de classificação de atributos (*labels*) de um dos *clusters*

```
[2 0 2 0 2 0 1 0 2 1 0 0 0 3 4 1 2 0 0 1 0 2 2 2 3 2 0 0 0 0 0 1 0 2 2 0 0
0 3 0 1 2 2 2 0 0 0 1 2 2 2 0 0 2 0 0 1 2 0 3 4 2 2 2 1 0 0 4 4]
```

Fonte: elaborado pelo autor

Subsequentemente para a melhor clusterização de dados, objetivando a melhor escolha do número de *clusters* que se encaixam com os dados da pesquisa, foi construído o gráfico *Elbow*, traduzido como gráfico do cotovelo (Figura 11), sabendo que a protuberância que se assemelha ao cotovelo, representa a sequência da diminuição da variação dos dados. Foi possível identificar que a partir do *cluster* de número três, os dados apresentam uma menor variação. Na Figura 11 o eixo x (abscissas) representa o número de *clusters* formados e o eixo y (ordenadas) representa a variação da soma das distâncias entre variáveis dos *clusters*, que é a variância entre somas.

Figura 11 – Gráfico *Elbow*

Fonte: elaborado pelo autor

Segundo Varela e Quadrelli (2017), o número de *clusters* é mais bem identificado depois de alguns incrementos, até que os incrementos não sejam mais influenciadores da variação dos dados. Visto o gráfico *Elbow* e sua protuberância no ponto de número três, representando assim a pouca variação dos parâmetros dos *clusters* formados, foi feita a reutilização dos dados no algoritmo. Agora com o parâmetro de entrada $n_cluster = 3$, considerado o número ideal para a presente pesquisa. Os *clusters* foram classificados em agências de nível A, nível B e nível C, sendo que cada nível possui os parâmetros mais semelhantes.

A representação gráfica está demonstrada na Figura 12, em que os pontos vermelhos representam os centros de semelhanças entre distâncias dos dados: os centroides. Já os pontos amarelos são as agências do nível A, podendo ser classificadas como agências de grande porte; os pontos azuis são as agências de nível B, consideradas as de médio porte; e, por fim, os pontos roxos são as agências de pequeno porte, o nível C. Na Figura 12 o eixo y representa os valores dos

Quadro 1 – Distribuição das agências por *cluster*

Agência	Cluster	Agência	Cluster	Agência	Cluster	Agência	Cluster
14	1	8	3	29	3	51	3
15	1	9	3	30	3	52	3
25	1	10	3	31	3	53	3
39	1	11	3	32	3	54	3
60	1	12	3	33	3	55	3
61	1	13	3	34	3	56	3
68	1	16	3	35	3	57	3
69	1	17	3	36	3	58	3
3	2	18	3	37	3	59	3
7	2	19	3	38	3	62	3
41	2	20	3	40	3	63	3
43	2	21	3	42	3	64	3
48	2	22	3	44	3	65	3
1	3	23	3	45	3	66	3
2	3	24	3	46	3	67	3
4	3	26	3	47	3		
5	3	27	3	49	3		
6	3	28	3	50	3		

Fonte: elaborado pelo autor

Por fim as médias dos valores das variáveis, por *cluster*, são apresentadas na Tabela 1. Verifica-se assim como cada variável se comporta para cada perfil de *cluster* formado, o que pode viabilizar decisões futuras da organização, uma vez que deixa claro todos os valores de cada variável, facilitando a análise e a tomada de decisão.

Os resultados propiciaram que a escolha de determinada alocação de recursos ou de aplicação de soluções não seja vista de forma individual, uma vez que os aspectos determinantes possam ser enquadrados em demais locais. Uma solução de tecnologia de informação, por exemplo, pode ser proposta para todas as agências de determinado nível, pois os aspectos que as norteiam são similares. Assim torna-se mais eficiente a tomada de decisão visto que as agências podem ser vistas como grupos com similaridade de atributos.

Tabela 1 – Média das variáveis obtidas por *Cluster*

VARIÁVEIS	CLUSTER 1	CLUSTER 2	CLUSTER 3
Demanda.(mensal)	6108,188	2065,7	837,6071
TME	6,647917	4,986667	4,704167
TMA	7,279167	4,626667	4,38869
Contas.Faturadas	122546	21706,8	8520,839
Video.Aten.	0,875	0	0,05357143
Quant..PA	7,125	2,4	1,107143
Acordo.de.pagamento	1809,875	396,4	77,01786
Alt./Rescisão.Contr	632,5	237,2	76,39286
Alt.Endereço.entrega	180,75	55,6	26,25
Alter.Dados.Parceiro	1835,75	555,8	170,4643
Anal.Proj.Viabilidade	10,125	0	0,07142857
Atendimento	266,125	79,4	17,64286
Baixa.Provisória	339,25	96,2	19,67857
Débito.Automático	4	0	0,2321429
Declara.Ligação.Nova	33,75	22,2	1,678571
Declaração.Quitação	5,75	2	0,8035714
Desligamento.a.Pedid	135,375	50,8	12,60714
Envio/Emissão.Fatura	1804,5	1583,4	531,1964
Falta.Energia	4,5	8,8	1,035714
Fatura	0,625	0	0,05357143
Faturamento	24,75	32,4	12,14286
Informação.comercial	1612,75	331,8	96,71429
Insp..Irregularidade	0	0	0,01785714
Ligação.c/.Inst.Exis	314,75	90,4	22,21429
Ligação.Nova	216,25	86,6	32,96429
Ligação.Provisória	8,125	4,6	2,142857
Mod..Análise.Técnica	87,75	50,8	8,839286
Modificação	173,625	66,2	14,42857
Nota.de.Serviço	308,25	79,4	19,51786
Opt.in	12,375	1,8	8,839286
Ordem.de.Restituição	7,625	1,2	0,3035714
Pedido.Indenização	18,75	6	0,75
Poda.de.Arvore	3,5	2,8	0,9642857
Qualidade.Fornec.	4	6,4	0,7678571
Reimpressão.Carta	3	0	0,1071429
Religa	221,625	66,8	21,42857
Religação.Urgência	2,5	1	0,3928571
Remoção.Poste.:OV	7,125	3,2	1,267857
Ressarcimento.Danos	0,75	1,2	0,1071429
Serviços.Internos	296,75	44,2	19,53571
Solic..Manut..Rede	1,25	4	0,5714286

Fonte: elaborado pelo autor

5 CONCLUSÕES

Sabendo da grande dificuldade atual das organizações mundiais de definir estratégias e tomadas de decisão, os dados apresentam-se como munição para definição de planos de ação como solução. Porém sabe-se que é necessário conhecimento para traduzir os dados em informação, visto que sua taxa de crescimento é alta. Assim são necessários métodos e algoritmos específicos para tais situações.

Consequentemente os métodos de clusterização entram em cena, para assim, propiciar *outputs* representativos e assertivos para os tomadores de decisão. No presente trabalho, o método de clusterização K-Means foi utilizado para agrupar os dados de uma empresa de serviços públicos, que necessitava de conhecimento para a melhor destinação de recursos a fim de melhorar seus indicadores de qualidade.

Perante os objetivos propostos, a presente pesquisa buscou e teve êxito na extração de dados da empresa. Foi realizada a extração dos dados brutos do sistema ERP da organização, o que foi o início do processo de clusterização de dados.

Também conforme proposto na pesquisa foi viável a construção e o detalhamento de como codificar e executar um algoritmo de clusterização de dados na linguagem de programação Python, conforme o método K-Means. O método apresentou resultados objetivos e satisfatórios visto que a classificação final dos dados apresentou uma segmentação de níveis de agências bem organizada. Assim definiram-se agências de níveis A, B e C, que são respectivamente de grande porte, médio porte e de pequeno porte. Esses níveis seguem o agrupamento por similaridade de variáveis proposto pelo método de K-Means. Nesta ótica o presente estudo serve de embasamento futuro para gestores da área tomarem decisões assertivas e bem definidas para um determinado grupo, ao invés de olhar individualmente para cada agência.

Sabe-se que o número de *clusters* é modificado perante a similaridade de suas variáveis, e, assim, conforme a alimentação de dados no algoritmo, o número de *clusters* é passível de alteração. Se uma variável apresentar uma grande alteração, por exemplo, se a demanda mensal por um determinado serviço saltar de um mês para outro em uma porcentagem alta, o número de *clusters* pode ser alterado. Dessa forma, é recomendado que o algoritmo seja atualizado com novos dados, para assim propiciar resultados sempre atuais.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AASHEIM, C. L.; LI, L.; WILLIAMS, S. Knowledge and skill requirements for entry-level information technology workers: A comparison of industry and academia. **Journal of information systems education**, v. 20, n. 3, p. 10, 2019.

ARORA, P. et al. Analysis of k-means and k-medoids algorithm for big data. **Procedia Computer Science**, v. 78, p. 507-512, 2016.

AYYAGARI, R. Hands-on ERP learning: Using OpenERP®, an alternative to SAP®. **Journal of Information Systems Education**, v. 22, n. 2, p. 5, 2019.

BOTTA, A. et al. Integration of cloud computing and internet of things: a survey. **Future generation computer systems**, v. 56, p. 684-700, 2016.

CESAR, J. A. et al. Assistência pré-natal nos serviços públicos e privados de saúde: estudo transversal de base populacional em Rio Grande, Rio Grande do Sul, Brasil. **Cadernos de Saúde Pública**, v. 28, p. 2106-2114, 2012.

CHIESA, F. Metodología para selección de sistemas ERP. **Reportes técnicos en ingeniería del software**, v. 6, n. 1, p. 17-37, 2004.

COHRS, F. M. et al. Aplicação de Análise de Cluster em dados integrados de um estudo prospectivo: projeto epidioso como cenário. **Journal of health informatics**, v. 5, n. 1, 2013.

DINIZ, C. A. R.; LOUZADA NETO, F. **Data mining**: uma introdução. São Paulo: Associação Brasileira de Estatística, 2000.

DUTT, A; ISMAIL, M. A.; HERAWAN, T. A systematic review on educational data mining. **Ieee Access**, v. 5, p. 15991-16005, 2017.

FENG, J.; BARBOSA, L. A.; TORRES, V. Systems and methods for social media data mining. **U.S. Patent**, n. 9, p. 262-517, 2016.

FERLIN, E. P.; REZENDE, D. A. Big Data aplicado à cidade digital estratégica: estudo sobre o volume de dados das aplicações Smart City. **Revista Gestão & Tecnologia**, v. 19, n. 2, p. 175-194, 2019.

FAYYAD, U. M. Data mining and knowledge discovery: Making sense out of data. **IEEE expert**, v. 11, n. 5, p. 20-25, 1996.

GAMARRA, C; GUERRERO, J. M.; MONTERO, E. A knowledge discovery in databases approach for industrial microgrid planning. **Renewable and Sustainable Energy Reviews**, v. 60, p. 615-630, 2016.

GU, B. et al. Biscuit: A framework for near-data processing of big data workloads. **ACM SIGARCH Computer Architecture News**, v. 44, n. 3, p. 153-165, 2016.

IBGE. Coordenação de Trabalho e Rendimento. **Estimativa de população 2019**. Espírito Santo: IBGE; 2019. Disponível em: <<https://cidades.ibge.gov.br/brasil/es/panorama>>. Acesso em: 12 maio 2020.

- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys, New York**, v. 31, n. 3, p. 265-323, Sept., 1999.
- KAVAKIOTIS, L. et al. Machine learning and data mining methods in diabetes research. **Computational and structural biotechnology journal**, v. 15, p. 104-116, 2017.
- KWAK, S. K.; KIM, J. H. Statistical data preparation: management of missing values and outliers. **Korean journal of anesthesiology**, v. 70, n. 4, p. 407, 2017.
- LU, H.; SETIONO, R.; LIU, H. Effective data mining using neural networks. **IEEE transactions on knowledge and data engineering**, v. 8, n. 6, p. 957-961, 1996.
- LU, H.; SETIONO, R. R.; LIU, H. NeuroRule: a connectionist approach to data mining. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 21.1995. **Proceedings...** p. 11-15, 1995.
- MAGLIO, P. P.; LIM, C. H. Innovation and big data in smart service systems. **Journal of Innovation Management**, v. 4, n. 1, p. 11-21, 2016.
- MARTINS, T. G. **Modelo de clusterização de dados para identificação de grupos de opinião em uma ferramenta de participação social**. 2017. 67 f., il. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Software)—Universidade de Brasília, Brasília, 2017.
- OLIVEIRA, E. F. et al. Voltage thd analysis using knowledge discovery in databases with a decision tree classifier. **Ieee Access**, v. 6, p. 1177-1188, 2017.
- OLIVEIRA, F. A.; TAMAYO, A. Inventário de perfis de valores organizacionais. **Revista de Administração da Universidade de São Paulo**, v. 39, n. 2, p.129-140, 2004.
- OLIVEIRA, L. F. **Proposta de métodos de clusterização de dados com validação por testes de heterogeneidade e discordância aplicados à regionalização de bacias hidrográficas**. 2018. 63 f. Dissertação (Mestrado em Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2018.
- PADILHA, T. C. C.; MARINS, F. A. S. Sistemas ERP: características, custos e tendências. **Production**, v. 15, n. 1, p. 102-113, 2005.
- PONTE, J. P. O estudo de caso na investigação em educação matemática. **Quadrante**, v. 3, n. 1, p. 3-18, 1994.
- QIU, J. et al. A survey of machine learning for big data processing. **EURASIP Journal on Advances in Signal Processing**, v. 2016, n. 1, p. 67, 2016.
- REZENDE, D. A.; RIBEIRO, S. S. Análise das estratégias, informações, serviços públicos municipais e tecnologias da informação e suas relações com projetos de cidade digital estratégica nas capitais da região centro-oeste do Brasil. **Revista Brasileira de Gestão e Desenvolvimento Regional**, v. 14, n. 3, p.26-47, 2018.
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Editora Manole

Ltda, 2003.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**, v. 1, n. 1, p. 96-105, 2014.

RISTOSKI, P.; PAULHEIM, H. Semantic Web in data mining and knowledge discovery: a comprehensive survey. **Journal of Web Semantics**, v. 36, p. 1-22, 2016.

RONQUILLO, J. G.; ZUCKERMAN, D. M. Software-Related Recalls of Health Information Technology and Other Medical Devices: Implications for FDA Regulation of Digital Health. **The Milbank Quarterly**, v. 95, n. 3, p. 535-553, 2017.

ROSSATTO, D. R.; TSUBOY, M. S. F.; FREI, F. Arborização urbana na cidade de Assis-SP: uma abordagem quantitativa. **Revista da Sociedade Brasileira de Arborização Urbana**, v. 3, n. 3, p. 1-16, 2008.

SALVATIER, J.; WIECKI, T. V.; FONNESBECK, C. Probabilistic programming in Python using PyMC3. **PeerJ Computer Science**, v. 2, p. 55, 2016.

SANTIN, V. F. Serviço público e direitos humanos. **Revista Paradigma**, v. 28, n. 2, p. 134-153, 2019.

SHU, K. et al. Fake news detection on social media: A data mining perspective. **ACM SIGKDD Explorations Newsletter**, v. 19, n. 1, p. 22-36, 2017.

SILVA, P. B. et al. Identificação de outliers em dados de acidentes de trânsito no Brasil: análise de cluster versus métodos estatísticos. **Revista de engenharia e tecnologia**, v. 11, n. 1, p. 103-114, 2019b.

SILVA, R. S. et al. Clusterização de dados mistos para análise da atividade pesqueira artesanal na bacia Araguaia-Tocantins. **Revista Brasileira de Computação Aplicada**, v. 11, n. 3, p. 155-164, 2019a.

SONKA, S. Big data: fueling the next evolution of agricultural innovation. **Journal of Innovation Management**, v. 4, n. 1, p. 114-136, 2016.

SOUZA, C. A. **Sistemas integrados de gestão empresarial: estudos de casos de implementação de sistemas ERP**. 2000. 258f. Dissertação (Mestrado em Administração), Universidade de São Paulo, São Paulo, 2000.

STERGIOU, C. et al. Secure integration of IoT and cloud computing. **Future Generation Computer Systems**, v. 78, p. 964-975, 2018.

SUCH, F. P. et al. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. **arXiv preprint arXiv:1712.06567**, 2017.

TRANTOPOULOS, K. et al. External knowledge and information technology: **Implications for process innovation performance**. v. 41, n. 1, p. 287-300, 2017.

VALLI, M. Análise de cluster. **Augusto Guzzo Revista Acadêmica**, n. 4, p. 77-87, 2002.

VARGHESE, B.; BUYYA, R. Next generation cloud computing: New trends and research directions. **Future Generation Computer Systems**, v. 79, p. 849-861, 2018.

VIEIRA, V. A. As tipologias, variações e características da pesquisa de marketing. **Revista da FAE**, v. 5, n. 1, 2002.

WANG, L. et al. pipsCloud: High performance cloud computing for remote sensing big data management and processing. **Future Generation Computer Systems**, v. 78, p. 353-368, 2018.

WOLFERT, S. et al. Big data in smart farming: a review. **Agricultural Systems**, v. 153, p. 69-80, 2017.

YANG, B. et al. Towards k-means-friendly spaces: simultaneous deep learning and clustering. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 34, 2017, Sydney. **Proceedings...** v 70, n. 17, p. 3861-3870, 2017b.

YANG, C. et al. Big Data and cloud computing: innovation opportunities and challenges. **International Journal of Digital Earth**, v. 10, n. 1, p. 13-53, 2017a.

YOO, K. et al. Decision tree-based data mining and rule induction for identifying hydrogeological parameters that influence groundwater pollution sensitivity. **Journal of Cleaner Production**, v. 122, p. 277-286, 2016.

ZAHARIA, M. et al. Apache spark: a unified engine for big data processing. **Communications of the ACM**, v. 59, n. 11, p. 56-65, 2016.

ZHANG, T. et al. A survey on emerging computing paradigms for big data. **Chinese Journal of Electronics**, v. 26, n. 1, p. 1-12, 2017.

APÊNDICE 1 – TABELA COM DADOS DAS AGÊNCIAS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1	1992,3	5,25	1,85	13258,00	1	1	30	88	31	143	0	10	14	0	1	0	15	919	0	0	3	16	0	33	57	1	20	35	49	0	0	0	0	0	0	8	0	1	0	8	0
2	167,0	1,53	4,70	5741,00	0	1	61	93	30	84	0	7	1	0	4	0	8	141	0	0	9	36	0	9	26	0	7	6	3	0	0	0	0	1	0	17	3	0	0	1	1
3	2090,5	3,77	4,58	14685,00	0	2	305	232	53	327	0	69	136	0	23	0	80	1203	1	0	15	144	0	142	140	10	115	77	62	1	2	0	3	1	0	57	1	2	1	0	2
4	930,0	4,35	3,92	7142,00	0	1	92	97	41	209	0	22	25	0	1	4	24	667	7	0	0	63	0	33	34	2	14	8	17	0	0	2	0	0	39	1	0	0	33	1	
5	1081,8	5,92	5,05	14650,00	0	1	56	197	45	355	3	43	48	2	8	29	0	517	0	0	2	478	0	29	38	1	14	28	13	58	0	0	4	0	0	30	0	2	0	66	0
6	319,3	4,53	5,55	3274,00	0	1	57	49	4	159	0	6	4	0	0	0	5	255	0	0	19	2	0	8	13	0	2	9	12	0	0	2	1	0	0	8	0	0	0	15	0
7	2819,0	6,70	6,03	37413,00	0	3	863	399	78	752	0	120	124	0	67	3	78	1875	11	0	24	248	0	132	122	5	59	109	155	6	1	16	3	9	0	70	3	4	4	77	7
8	615,0	5,22	4,08	4527,00	0	1	126	36	24	109	0	8	22	0	3	5	11	437	8	0	0	184	0	12	14	1	7	19	25	0	1	1	0	1	0	9	1	0	0	0	0
9	1829,8	2,90	4,03	12973,00	0	2	84	182	34	296	0	31	20	0	2	0	32	885	1	0	29	193	0	39	49	8	14	24	21	63	1	0	0	0	0	89	0	1	1	33	0
10	1992,3	6,98	4,65	18980,00	0	2	274	148	60	395	0	31	32	0	1	0	14	1362	3	0	22	637	0	43	60	14	15	27	31	27	0	1	0	1	0	75	0	2	0	49	1
11	599,3	7,40	2,30	6760,00	0	1	74	60	20	50	0	19	10	0	0	0	6	492	1	0	4	61	0	16	27	0	4	14	16	0	1	0	0	0	0	11	0	3	0	10	2
12	357,0	2,78	3,43	4620,00	0	1	116	51	12	150	0	35	12	0	0	0	7	589	1	0	41	16	0	21	100	46	5	15	22	0	0	0	0	0	10	0	0	0	27	0	
13	412,3	1,58	5,18	4433,00	0	1	53	52	4	59	0	10	25	0	2	0	10	323	3	0	0	42	0	24	26	0	4	2	19	1	0	0	0	0	0	6	2	1	0	5	0
14	5397,8	5,57	7,50	82290,00	0	7	1255	414	92	1446	7	155	202	0	59	3	99	1246	4	0	23	1171	0	186	170	11	50	93	192	0	3	12	8	1	0	237	3	5	0	211	1
15	6687,8	7,55	7,40	134884,00	2	6	1970	824	436	2982	16	235	248	1	35	1	137	1196	10	0	13	1569	0	271	202	11	105	247	359	13	11	10	1	4	17	239	4	6	0	338	1
16	2447,5	5,10	4,97	16387,00	0	2	70	76	19	131	0	14	19	1	0	0	14	61	0	0	0	167	0	21	20	2	2	4	19	1	0	6	0	0	0	20	1	1	0	68	1
17	942,3	3,00	4,20	11395,00	0	1	147	124	19	339	0	32	110	0	1	0	10	893	2	0	8	106	0	33	19	0	12	15	38	1	1	8	1	7	1	44	2	1	0	31	0
18	353,5	2,25	4,60	5390,00	0	1	28	59	26	56	0	17	12	1	3	0	8	436	1	0	17	78	0	31	34	0	18	16	5	3	0	0	0	0	6	0	0	0	1	0	
19	100,0	1,03	7,32	2143,00	0	1	23	25	9	34	0	3	0	0	0	0	2	102	0	1	1	2	0	3	10	1	0	0	1	0	0	0	0	0	0	0	0	0	0	15	0
20	590,0	0,25	7,05	17975,00	0	1	66	80	31	179	0	6	19	0	1	1	11	675	0	0	2	17	0	10	15	1	3	4	8	0	0	0	0	0	2	0	0	0	62	1	
21	251,5	4,33	5,12	2841,00	0	1	22	26	34	57	0	7	14	0	0	1	3	170	0	0	9	75	0	5	23	0	0	9	5	0	0	0	1	1	0	1	0	4	0	0	0
22	611,5	5,87	5,07	10682,00	0	1	104	104	58	373	0	33	6	0	17	0	12	494	0	0	9	219	0	17	22	0	7	16	10	30	0	0	0	0	0	41	0	1	0	0	4
23	353,0	5,08	2,95	9216,00	0	1	94	26	7	38	0	8	6	0	0	0	6	362	0	0	12	27	0	10	28	1	3	10	7	0	1	0	1	1	0	10	1	0	0	0	0
24	1328,5	5,70	4,73	12687,00	0	1	90	141	15	260	0	24	80	0	0	0	25	874	1	0	22	144	0	34	36	4	8	26	44	0	0	1	0	0	0	15	3	0	1	9	0
25	4804,5	6,48	6,65	74186,00	1	4	1297	568	135	1753	9	206	253	13	18	2	144	1858	1	0	14	974	0	192	173	2	64	139	170	13	1	18	2	8	0	195	0	3	0	150	0
26	1018,0	4,00	3,82	10306,00	0	1	43	172	43	420	0	18	3	1	0	0	20	849	0	0	6	268	0	49	49	3	17	18	14	95	0	2	1	2	0	22	0	2	1	45	0
27	260,8	3,28	3,72	4767,00	0	1	73	45	10	102	0	19	25	2	0	0	6	170	0	0	18	11	0	11	7	0	5	18	9	4	0	0	1	1	0	4	0	1	0	16	0
28	326,0	1,90	5,02	3822,00	0	1	29	51	25	73	0	1	0	0	0	0	5	315	0	0	20	6	1	20	16	0	6	5	14	0	0	0	1	1	0	6	1	2	0	23	0
29	617,3	8,45	4,77	5867,00	0	1	98	43	33	54	0	20	22	0	0	0	15	303	0	0	1	40	0	10	21	0	9	19	11	1	1	0	0	0	0	16	0	0	0	7	0
30	621,3	4,28	3,30	6039,00	0	1	46	46	28	238	0	6	0	0	0	0	9	688	0	0	49	13	0	20	15	0	2	3	7	0	2	0	0	1	0	9	1	2	0	19	1
31	614,3	2,60	3,92	6624,00	0	1	25	43	16	85	0	24	13	0	0	0	12	442	0	0	9	72	0	19	12	0	18	17	21	0	0	0	0	0	14	0	2	0	2	3	
32	1388,3	5,95	3,27	17877,00	0	1	44	70	28	216	0	14	15	0	0	15	998	0	0	23	39	0	28	32	0	8	22	39	0	0	0	1	0	0	44	1	1	0	37	0	
33	594,5	6,15	4,00	5109,00	0	1	78	49	31	86	0	12	3	0	1	0	9	335	0	0	0	45	0	16	38	2	8	10	6	0	0	0	2	3	0	12	0	1	0	0	0
34	1094,3	4,23	3,42	11947,00	0	1	151	131	40	235	0	20	8	4	0	0	11	896	5	0	56	44	0	48	67	1	9	29	31	20	0	3	5	2	0	34	1	2	0	57	4
35	1262,3	4,72	3,40	10902,00	0	1	128	182	46	171	0	31	23	0	3	0	35	1209	1	0	7	150	0	56	95	3	13	29	36	4	1	0	1	2	0	64	1	1	0	38	1
36	582,8	5,42	2,90	4968,00	0	1	81	53	12	61	0	5	13	0	5	0	10	906	0	0	63	12	0	25	19	0	3	9	15	0	0	3	4	1	0	14	0	2	0	5	0
37	809,0	3,43	3,35	7033,00	0	1	65	92	10	302	0	11	21	0	0	0	13	653	0	0	7	15	0	27	15	0	14	13	10	0	0	0	4	0	0	6	0	1	0	2	0
38	135,5	3,30	4,43	5263,00	0	1	33	33	16	251	0	13	6	0	0	0	3	225	1	0	7	44	0	12	51	1	3	4	7	1	0	0	0	0	2	0	1	0	7	1	
39	4181,0	6,03	8,77	67406,00	0	6	927	504	167	1479	0	157	310	5	8	7	100	722	7	4	19	1022	0	191	255	0	83	155	188	45	10	27	3	5	0	167	1	7	1	203	7
40	281,3	1,85	4,97	6211,00	0	1	55	64	5	107	0	4	0	0	5	1	9	342	1	0	2	7	0	16	19	2	14	12	36	16	0	0	0	1	0	12	0	1	0	13	1
41	2262,0	5,45	5,22	21608,00	0	3	395	209	53	791	0	105	55	0	11	6	41	1809	2	0	11	221	0	69	74	8	28	80	78	0	2	5	1	12	1	110	0	4	0	79	0
42	813,8																																								

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
43	734,0	3,47	2,40	11471,00	0	1	31	192	51	129	0	40	0	0	0	1	17	1116	29	0	102	11	0	41	42	0	15	20	23	0	0	2	6	4	0	19	1	1	1	7	8
44	838,0	3,40	3,15	8158,00	0	1	148	83	12	80	0	23	3	0	2	0	6	624	2	0	2	29	0	22	15	0	3	12	29	0	0	0	0	0	0	57	0	0	0	0	0
45	302,3	3,15	4,78	2652,00	0	1	54	31	1	66	0	3	3	0	0	2	4	201	0	0	1	9	0	1	5	0	4	3	6	6	0	0	0	0	0	4	0	0	0	4	0
46	432,0	4,83	3,18	7564,00	0	1	7	30	15	66	0	2	2	0	1	0	6	441	0	0	18	10	0	27	21	0	5	9	11	0	0	0	2	0	0	6	0	0	0	1	1
47	1090,0	9,33	5,63	5870,00	0	1	50	101	13	123	0	17	31	0	2	0	8	482	0	1	24	48	0	37	24	0	13	12	7	0	0	0	3	0	3	11	0	1	0	19	0
48	2423,0	5,55	4,90	23357,00	0	3	388	154	43	780	0	63	166	0	10	0	38	1914	1	0	10	1035	0	68	55	0	37	45	79	2	1	7	1	6	0	78	0	5	0	58	3
49	1243,3	2,83	4,00	8591,00	0	2	78	165	70	574	0	42	79	0	2	0	14	694	1	0	2	311	0	55	28	2	9	20	34	63	0	1	2	1	0	62	0	1	0	23	0
50	850,5	7,62	5,40	9471,00	1	1	127	79	25	152	0	27	21	0	1	0	17	413	1	0	5	42	0	27	45	0	4	12	20	4	0	0	0	1	0	26	1	0	2	11	0
51	1469,8	6,70	5,30	14002,00	0	2	312	126	46	395	1	75	76	0	4	0	87	897	1	0	2	134	0	53	49	2	24	54	96	0	3	3	3	3	2	98	0	5	0	125	0
52	344,3	5,43	4,33	3088,00	0	1	26	38	7	98	0	6	1	0	0	0	7	98	0	0	3	24	0	7	9	0	2	1	5	1	0	1	0	0	0	10	0	1	0	2	0
53	613,0	5,27	2,98	4973,00	0	1	84	35	21	54	0	8	23	0	2	0	2	434	0	0	5	39	0	4	19	1	6	11	27	0	0	0	0	0	0	16	0	0	0	13	0
54	488,3	5,13	4,07	9485,00	0	1	28	92	22	201	0	16	14	0	7	0	13	267	0	0	2	30	0	16	32	1	10	10	9	55	3	1	0	0	0	7	0	0	0	16	0
55	310,5	2,58	4,48	4620,00	0	1	82	33	22	111	0	24	13	1	0	0	5	301	0	0	0	32	0	5	18	1	3	11	20	1	0	1	1	1	0	13	0	0	0	6	0
56	875,8	7,07	4,68	5714,00	0	1	25	20	7	53	0	7	1	0	0	0	0	108	4	0	0	16	0	3	12	0	4	0	3	23	0	0	0	2	0	3	0	0	0	3	0
57	2060,8	6,02	3,80	17002,00	0	2	75	68	75	452	0	43	67	0	1	2	15	287	0	0	23	564	0	22	96	4	25	22	19	0	0	0	0	0	0	20	0	2	0	2	0
58	1065,0	4,08	3,48	9686,00	0	1	45	58	24	114	0	21	29	0	6	0	7	760	0	0	12	100	0	18	45	5	6	12	0	0	0	0	4	0	0	17	0	2	0	3	0
59	1500,5	6,55	7,92	5231,00	0	1	76	118	26	255	0	14	44	0	2	0	29	990	0	0	52	151	0	14	58	0	15	17	30	0	0	1	0	0	0	26	0	2	0	4	0
60	5873,5	7,08	7,65	53168,00	0	7	1048	485	85	1182	2	149	265	1	24	2	83	1388	0	0	15	1001	0	180	146	10	56	112	201	15	7	25	2	4	0	160	1	4	0	242	1
61	6687,5	7,37	6,60	203055,00	1	9	3020	693	189	2004	6	485	590	0	44	14	140	2942	12	0	15	2787	0	558	286	19	156	238	656	3	10	21	5	0	4	359	1	14	2	323	0
62	684,5	4,78	3,05	9439,00	0	1	42	45	38	308	0	13	16	0	1	0	13	624	0	0	3	302	0	27	31	2	13	30	52	0	0	1	0	0	0	45	1	2	0	40	2
63	574,3	7,67	4,60	8580,00	0	1	108	73	30	40	0	26	6	0	1	0	24	755	10	0	10	24	0	25	63	6	10	17	20	0	0	0	3	1	0	17	0	3	0	26	1
64	1198,5	6,75	4,70	10318,00	1	1	44	107	66	146	0	19	12	1	0	0	23	621	2	0	9	115	0	43	53	0	16	29	37	2	0	1	1	6	0	18	0	8	0	71	3
65	591,5	3,02	4,35	24645,00	0	1	177	51	24	120	0	11	9	0	3	0	9	541	0	0	9	22	0	18	22	2	11	18	29	0	0	4	1	0	0	14	0	1	0	15	0
66	266,8	2,98	4,05	4127,00	0	1	66	45	6	83	0	6	11	0	0	0	3	383	0	0	16	42	0	7	16	0	3	3	3	0	2	0	0	1	0	6	0	0	0	0	0
67	2414,3	11,65	8,60	3071,00	0	1	21	12	13	86	0	6	8	0	0	0	1	284	0	1	1	30	0	3	15	0	4	0	3	15	0	0	0	0	0	4	0	0	0	0	0
68	8985,5	8,07	7,27	207889,00	2	9	2686	775	210	2006	25	443	565	4	51	14	219	2954	2	0	56	2562	0	534	319	9	109	265	393	2	10	19	4	7	0	225	5	11	2	322	0
69	6248,0	5,03	6,40	157490,00	1	9	2276	797	132	1834	16	299	281	8	31	3	161	2130	0	1	44	1816	0	406	179	3	79	140	307	8	9	18	3	3	3	191	5	7	1	585	0

Enumeração de Agências: 1; Demanda (mensal): 2; TME: 3; TMA: 4; Contas Faturadas: 5; Cabines de Vídeo Atendimento.: 6; Quantidade atendentes: 7; Acordo de pagamento: 8; Alteração / Rescisão Contrato: 9; Alteração de Endereço entrega: 10; Alteração de Dados Parceiro: 11; Análise de Projeto e Viabilidade: 12; Quantidade de Atendimento: 13; Baixa Provisória: 14; Débito Automático: 15; Declara Ligação Nova: 16; Declaração Quitação: 17; Desligamento a Pedido: 18; Envio/Emissão Fatura: 19; Falta Energia: 20; Fatura: 21; Faturamento: 22; Informação comercial: 23; Inspeção de Irregularidade: 24; Ligação c/ Instalação Existente: 25; Ligação Nova: 26; Ligação Provisória: 27; Mod. Análise Técnica: 28; Modificação: 29; Nota de Serviço: 30; Opt in, conta por e-mail: 31; Ordem de Restituição: 32; Pedido Indenização: 33; Poda de Arvore: 34; Verificação de Qualidade Fornecimento: 35; Reimpressão Carta: 36; Religa: 37; Religação Urgência: 38; Remoção Poste: OV: 39; Ressarcimento Danos: 40; Serviços Internos: 41; e, Solicitação de Manutenção de Rede: 42.