

FERNANDO DE OLIVEIRA LIMA XAVIER

**APLICAÇÃO DO CONCEITO DE *DATA SCIENCE* NA
MINERAÇÃO: ANÁLISE EXPLORATÓRIA DE CONDIÇÕES DE
MANUTENÇÃO DE EQUIPAMENTOS**

São Paulo

2018

FERNANDO DE OLIVEIRA LIMA XAVIER

**APLICAÇÃO DO CONCEITO DE *DATA SCIENCE* NA
MINERAÇÃO: ANÁLISE EXPLORATÓRIA DE CONDIÇÕES DE
MANUTENÇÃO DE EQUIPAMENTOS**

Trabalho de Formatura em Engenharia de
Minas do curso de graduação do Departamento
de Engenharia de Minas e de Petróleo
da Escola Politécnica da Universidade de São
Paulo

Orientador Dr. William Pratt Rogers
The University of Utah

Co-orientador Prof. Dr. José Renato Baptista de Lima
Escola Politécnica da Universidade de São Paulo

São Paulo

2018

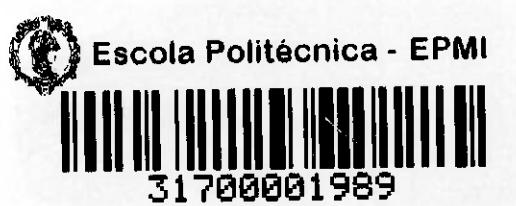
TF-2018

X18a

Syno 2918787

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

H 2018 L



Catalogação-na-publicação

XAVIER, FERNANDO

APLICAÇÃO DO CONCEITO DE DATA SCIENCE NA MINERAÇÃO:
ANÁLISE EXPLORATÓRIA DE CONDIÇÕES DE MANUTENÇÃO DE
EQUIPAMENTOS / F. XAVIER -- São Paulo, 2018.

49 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e Petróleo.

1.EQUIPAMENTOS DE MINERAÇÃO 2.BANCO DE DADOS
I.Universidade de São Paulo. Escola Politécnica. Departamento de
Engenharia de Minas e Petróleo II.t.

Ao meu pai, Filipe, e à minha mãe, Andrea,
por todo apoio e confiança que meu foi dado
ao longo dos anos.

E ao meu irmão, Pedro, pela amizade e
companheirismo inestimáveis.

AGRADECIMENTOS

Agradeço, primeiramente, ao Professor Dr. William Pratt Rogers, por todo o auxílio, direcionamento e conhecimento que me foram passados, e por ter sido essencial para que este trabalho se tornasse realidade.

Ao Professor Dr. José Renato Baptista de Lima, também por todo auxílio e conhecimento, e ainda por ter me apoiado em minha jornada para produzir um trabalho com tema inovador e internacionalizado.

À Escola Politécnica da Universidade de São Paulo e à *The University of Utah*, e a todos os integrantes de seus respectivos corpos docentes que foram de extrema importância para meu desenvolvimento como aluno, como engenheiro, e como cidadão.

Por fim, a todos os amigos e familiares que sempre estiveram ao meu lado nos momentos mais difíceis, tanto dentro como fora da universidade.

RESUMO

Com o avanço da tecnologia na mineração, o armazenamento, gerenciamento e análise de dados gerados nas operações são cada vez mais importantes para garantir o bom desempenho de um empreendimento. Desta forma, este trabalho busca aplicar o conceito de *Data Science* dentro do ambiente de mineração, utilizando como exemplo a questão envolvendo a manutenção de equipamentos, que pode representar mais de um terço dos custos de operações de minas ao redor do mundo. O objetivo é utilizar uma malha de dados extensa e completa e tratá-la de modo a evidenciar os problemas relacionados à manutenção de equipamentos, como intervalos entre falhas, duração de manutenções e custos. Este estudo utiliza-se de ferramentas de bancos de dados SQL e de análise estatística para identificar e agrupar equipamentos de acordo com seu desempenho em relação a manutenções no ambiente de mina. Os resultados mostram que é possível que equipamentos de categorias diferentes se comportem de maneiras similares e que equipamentos de uma mesma categoria podem demandar ações diferenciadas em relação à sua manutenção. Com este estudo, conclui-se não somente que pode haver uma relação direta entre intervalos de manutenção preventiva e gastos gerados com manutenção, mas também que a aplicação de conceitos de *Data Science* torna-se cada vez mais importante para os ambientes da mineração moderna.

Palavras-chave: *Mineração; Manutenção de equipamentos; Data Science*

ABSTRACT

With the constant advances of technology in mining, collecting, managing and analyzing data generated in a mine has become increasingly more important to ensure the quality and feasibility of the operation. This study seeks to apply the concept of *Data Science* within the mining environment, utilizing the problem of equipment maintenance as an example, as it can be responsible for more than a third of all operating costs in a mine. The main goal of this research is to treat the information in an extensive and complete dataset, in order to shed light to the main problems regarding equipment maintenance, like time between failures, duration of maintenance events and cost. This study utilizes tools such as SQL databases and statistical analysis identify and cluster equipment according to their performance in terms of maintenance. The results show that it is possible that equipment from different categories behave similarly, while equipment from within the same category might demand different actions in order to tackle maintenance issues. With this study, it is not only possible to understand that there is a probable relation between planned maintenance interval times and maintenance cost, but that the application of the concept of *Data Science* in the modern mining environment becomes more important every day.

Key words: *Mining, Equipment maintenance, Data Science*

LISTA DE GRÁFICOS

Gráfico 1 – O método Elbow para definição do número de <i>clusters</i>	25
Gráfico 2 – Distribuição dos equipamentos de acordo com os valores de MTTR e MTBF para o ano de 2013.....	26
Gráfico 3 – <i>Clusters</i> resultantes da aplicação do método K-Médias	27
Gráfico 4 – <i>cluster</i> 1 e equipamentos	28
Gráfico 5 – <i>cluster</i> 2 e equipamentos	28
Gráfico 6 – <i>cluster</i> 3 e equipamentos	29
Gráfico 7 – <i>cluster</i> 4 e equipamentos	29
Gráfico 8 – <i>cluster</i> 5 e equipamentos	30
Gráfico 9 – <i>cluster</i> 6 e equipamentos	30
Gráfico 10 – Número de equipamentos por cluster	31
Gráfico 11 – Custos totais de manutenção	32
Gráfico 12 – Custo médio de manutenção de um equipamento dentro de cada cluster, por categoria.....	33
Gráfico 13 – MTBF de escavadeiras, por grupo de clusters similares	34
Gráfico 14 – Intervalo médio entre manutenções de escavadeiras, por grupo de clusters similares	34

LISTA DE TABELAS

Tabela 1 - Quantidade de equipamentos por categoria	20
Tabela 2 – Valores de MTBF e MTTR, por equipamento.....	22
Tabela 3 – Custos de manutenção no ano de 2013, por equipamento.....	23

Sumário

1	INTRODUÇÃO	11
2	OBJETIVOS	12
3	REVISÃO BIBLIOGRÁFICA.....	12
3.1	Estratégias de manutenção.....	13
3.2	Parâmetros temporais e disponibilidade	14
3.3	<i>Data Science</i> e <i>Big Data</i>	15
3.4	Agrupamento e <i>clusterização</i>	16
3.4.1	Método K-Médias.....	16
3.4.2	Método Elbow	17
4	MÉTODO	17
4.1	Aquisição dos dados	18
4.2	Validação e triagem dos dados	18
4.3	Análise dos Dados	19
4.3.1	Montagem de códigos em SQL	19
4.3.2	Agrupamento em <i>clusters</i>	19
4.4	Análise e discussão.....	19
5	ANÁLISE DOS RESULTADOS	19
5.1	Aquisição das bases de dados.....	20
5.1.1	Número de equipamentos	20
5.1.2	Definição do MTBF e MTTR para cada equipamento.....	21
5.1.3	Definição dos custos de manutenção anual por equipamento	23
5.2	Análise estatística	24
5.2.1	Definição do número de <i>clusters</i> pelo método Elbow	24
5.2.2	Clusterização pelo método K-Médias	25
5.3	Análise comparativa de custos.....	31
5.3.1	Equipamentos por cluster	31
5.3.2	Análises de custo.....	32

5.3.3	Categoría Crítica.....	33
6	DISCUSSÃO	35
7	CONCLUSÕES.....	37
8	REFERÊNCIAS BIBLIOGRÁFICAS	39

1 INTRODUÇÃO

A modernização da atividade mineradora é um fenômeno de crescimento acelerado. Com os avanços nos métodos e tecnologias empregadas na extração e beneficiamento mineral, avançam também as preocupações com saúde, segurança, higiene e meio ambiente que envolvem as variadas atividades do setor mineral. O volume de informações que necessitam de monitoramento em operações de mineração se torna cada vez maior. A mineração moderna é capaz de gerar uma grande quantidade de dados a todo instante: condições de operação de equipamentos, teores de concentrados e rejeitos, valores de poluentes em solo e na atmosfera e incidentes de trabalho com funcionários são apenas alguns exemplos de todas as variáveis que devem ser controladas em uma atividade de mineração para garantir máxima eficiência, produtividade e segurança.

Para viabilizar o monitoramento e gerenciamento destas informações, as empresas de mineração tem implementado sistemas de bancos de dados em suas operações. Um sistema de bancos de dados, de maneira simplificada, consiste em uma plataforma virtual capaz de armazenar informações dos mais variados tipos dentro de tabelas e matrizes criadas pelo usuário de acordo com a sua necessidade. Esta plataforma permite o gerenciamento ágil e seletivo de grandes quantidades de informações, tornando possível o armazenamento de históricos de dados e até implementação de análises de previsibilidade.

Minas de grande porte geram uma grande quantidade de dados com grande diversidade. Sensores instalados em toda uma operação podem captar, interpretar e armazenar informações variadas, desde pressões instantâneas de pneus de caminhões, até movimentações e deslocamentos de taludes e pilares.

Além disto, uma mina é capaz de gerar dados não tão automatizados e que costumam ser armazenados na forma de relatórios. São informações derivadas de observações feitas por colaboradores ou eventos reportáveis, normalmente não atreladas a sensores. Podem incluir informações referentes a incidentes que fogem à normalidade das atividades do ambiente de mineração, como acidentes com danos a equipamentos ou pessoas, ou então eventos ligados a atividades

como manutenção de equipamentos, vias de acesso, reparos em geral, entre outros.

Obtenção, armazenamento e gerenciamento destes grupos de informação são essenciais para a realização de análises estratégicas sobre medidas de prevenção e correção de falhas nos ambientes de mineração. É possível, ainda, utilizar a base de dados para criar modelos de previsão de falhas e paradas mais coerentes com as realidades da mina em questão.

2 OBJETIVOS

O objetivo deste trabalho é realizar um estudo de caso sobre as condições de manutenção de equipamentos em um ambiente de mineração, do ponto de vista do conceito de *Data Science*. Para isto, uma base de dados real de uma mina de carvão subterrânea no Canadá foi utilizada como exemplo de obtenção de informações.

As análises desenvolvidas consideraram dados objetivos de equipamentos como caminhões, perfuratrizes, escavadeiras e tratores de terraplanagem. Procurou-se estabelecer correlações entre as informações características de cada equipamento, com o objetivo de definir agrupamentos que, por sua vez, permitirão avaliar o desempenho de diferentes modelos e entender a dinâmica de custos e tempos de manutenção de cada grupo.

Procurou-se, com este estudo, explorar as possibilidades analíticas que a aplicação do conceito de *Data Science* possibilita ao ambiente de mineração. Especificamente para este estudo, a análise exploratória dos dados procura identificar grupos críticos de equipamentos quanto aos seus tempos de manutenção e disponibilidade, além de estabelecer uma correlação entre estes fatores e os custos de manutenção gerados por equipamento.

3 REVISÃO BIBLIOGRÁFICA

A manutenção de equipamentos compõe uma parte importante do planejamento de operação de uma mina. As estratégias adotadas diante do tipo de

manutenção a ser priorizado para um determinado equipamento podem variar muito de acordo com fatores como modelo do equipamento, idade, tempo de uso, além de fatores de produção e localização, como qualidade das pistas de rolamento, cargas exigidas, dentre outras. A relevância da otimização dos processos de manutenção são de extrema importância não somente do ponto de vista econômico, mas também de segurança dos operadores e equipamentos (Sifferlinger & Rath, 2016).

No entanto, visualizar o impacto real destes fatores na mineração não é uma tarefa simples. Equipamentos de mesmo modelo podem demandar ações diferentes com relação à manutenção dependendo das condições de operação às quais estão submetidos.

Em minas subterrâneas de carvão no Canadá, o custo de manutenção de equipamentos costuma variar de 30% a 35% dos custos de operação da mina (Hall, 1997). Soma-se a isto o fato de que não só a maioria dos equipamentos de lavra de minas são móveis (aumentando a chance de falhas), como também, estes podem falhar em lugares inoportunos e de difícil acesso, o que torna os custos de manutenção ainda maiores. Assim, determinar ações corretas quanto à manutenção de equipamentos em uma mina é importante para minimizar os custos e perdas na operação.

3.1 Estratégias de manutenção

Existem três tipos principais de abordagem de manutenção de equipamentos de mineração (Gölbaşı & Demirel, 2017):

- manutenção corretiva: a manutenção corretiva consiste na realização de paradas em resposta a falhas que necessitam de atenção em um determinado equipamento. Quebra de peças, vazamentos e mal funcionamentos são exemplos de eventos que costumam levar a paradas para manutenção corretiva;
- manutenção preventiva sistemática: a manutenção preventiva sistemática envolve a realização de paradas programadas para manutenção preventiva de equipamentos. Tais paradas seguem um cronograma específico pré-determinado pelos gerentes de manutenção

e/ou recomendadas pelos fabricantes de cada equipamento e buscam prevenir a ocorrência de paradas não programadas; e

- manutenção preventiva de condição: também chamada de “manutenção preditiva”, a manutenção preventiva de condição consiste na realização de paradas para manutenção preventiva em resposta ao histórico de falhas e às condições específicas nas quais o equipamento está inserido. Condições das pistas de rolagem e carga média são exemplos de indicadores que podem alterar a vida útil de equipamentos ou de partes específicas e que podem levar à decisão de realizar uma manutenção preventiva fora do cronograma.

Não há um consenso sobre a melhor ação a ser tomada. Embora em um primeiro momento as manutenções corretivas pareçam ser menos desejadas, o alto custo de realização de paradas para manutenção preventiva podem ser suficientes para não justificar a realização das mesmas. Por outro lado, para o caso de equipamentos essenciais e de difícil substituição, como britadores primários e skips, os custos atrelados a longas paradas não programadas podem comprometer a produtividade de uma operação. Assim, é necessária uma avaliação caso a caso para se determinar a melhor estratégia de manutenção para cada equipamento.

3.2 Parâmetros temporais e disponibilidade

Dois parâmetros são relevantes para a avaliação do desempenho de manutenção dos equipamentos (Dhillon, 2006):

- tempo entre falhas, ou “*time between failures*” (TBF): o TBF consiste no intervalo de tempo entre falhas reportadas de um equipamento, ou seja, a quantidade de tempo entre o final de um evento de manutenção até o início do próximo evento de mesma categoria. A média de tempo entre falhas é indicada pela sigla MTBF, ou “*mean time between failures*”;
- tempo em manutenção, ou “*time to repair*” (TTR): o TTR leva em consideração o tempo no qual um equipamento ficou efetivamente em manutenção, desde a data e horário de seu início até o final do evento. A média de tempo em manutenção é indicada pela sigla MTTR, ou “*mean time to repair*”;

Com base nestes dois parâmetros, é possível ainda definir a disponibilidade dos equipamentos (DF) como ilustrado na Equação I (Sundararajan, 1991):

$$DF = \frac{MTBF}{MTBF+MTTR} \quad (I)$$

3.3 *Data Science e Big Data*

Em se tratando de dados e informações obtidos em alguma operação de caráter industrial, o mais comum na atualidade é o conceito de *Data Science*, que pode ser caracterizado como um conjunto de dados homogêneo e unidimensional, contendo dados estáticos e estruturados voltados para análises descritivas. Recentemente, o conceito de *Big Data* vem ganhando força, pois está atrelado a utilização de dados de maneira mais versátil e robusta, permitindo análises de dados multidimensionais, dinâmicos e semiestruturados ou até desestruturados e é geralmente orientado para criação de modelos de previsão que necessitam da correlação entre um volume muito grande de variáveis (Ouyang, Wu, Huang, 2018).

Análises de caráter de *Data Science* operam de uma maneira na qual dados históricos são compilados para se descrever desempenhos e identificar padrões, e geralmente são apresentadas de maneira a unir modelagem estatística, gerenciamento de dados, computação, otimização, comunicação e visualização (Reid, 2018). Já no âmbito de *Big Data*, sua aplicação é mais específica. Nele, os dados são dinâmicos, sendo que dados coletados no passado podem ainda ser alterados no presente e futuro. *Big Data* é um conceito que costuma ser definido por parâmetros como volume, variedade e velocidade dos dados (Larson, Chang, 2016) e suas ferramentas são utilizadas para criação de modelos preditivos multivariados e de alta complexidade, para torná-los mais simples, mas mantendo a capacidade de analisar a probabilística de modelos espaciais, temporais e de diferentes dimensões (Lau, Adams, Girolami, Butler, Elshafie, 2018).

No entanto, para sistemas mais simples, modelos de previsão multivariados podem ser simplificados por modelos derivados de análises descritivas, por meio da utilização de métodos estatísticos, de padronização e linhas de tendência (Ceri, 2018).

No que diz respeito à mineração, é possível reduzir sistemas complexos como a previsibilidade de falhas de equipamentos e ocorrência de acidentes a modelos derivados de análises históricas descritivas, às custas de perder universalidade e adaptabilidade por não poder incluir a análise dinâmica dos dados. Como apresentado por Wang e Gao (2011), é possível criar modelos de previsão de acidentes em ambientes de mineração por análise descritiva, desde que algum algoritmo de previsão seja empregado e que se tenha noção das limitações do modelo. Esta ideia pode ser estendida à manutenção de equipamentos.

3.4 Agrupamento em *clusters*

A análise de dados históricos com foco em tomadas de decisões, sejam elas de negócios ou de operações, geralmente passa pelo conceito de agrupamento em *clusters*, que consiste em unir elementos de acordo com características específicas, não necessariamente intrínsecas à elementos de uma mesma categoria.

No caso da manutenção de equipamentos de mineração, Dindarloo (2017) expôs que é possível utilizar conceitos de *Data Science* e *Big Data* para criar modelos de manutenção preventiva mais adequados a equipamentos diferentes – ou até equipamentos individuais de mesmo modelo – em uma mina, de acordo com a análise do histórico de falha e magnitude do impacto das paradas não programadas dos mesmos na produção.

3.4.1 Método K-Médias

O método de clusterização K-Médias é um método simples, que procura encontrar um número ótimo K de agrupamentos para os quais a soma dos desvios quadrados de cada agrupamento seja a menor possível. Seja $X = \{x_i\}$, $i = 1, \dots, n$, o conjunto de n pontos a serem agrupados. Seja K um número fixado de agrupamentos e μ_k a média dos valores dos pontos contidos no agrupamento c_k . O erro quadrático $J(c_k)$ entre a média μ_k e os pontos em c_k pode ser dado de acordo com a Equação II.

$$J(c_k) = \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \quad (\text{II})$$

Assim, o objetivo da média K é determinar o número ótimo de agrupamentos K para o qual a somatória dos erros quadráticos de todos os agrupamentos é a menor possível, como ilustrado na Equação III.

$$J(c) = \sum_{k=1}^K \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \quad (\text{III})$$

O número de agrupamentos não pode ser determinado previamente, portanto uma série de tentativas serão necessárias, variando o número K de agrupamentos. Para um conjunto de n pontos, o primeiro valor ideal de K pode ser tomado como n/2, pois está mais distante das extremidades. Um conjunto K=1 agrupamentos coloca todos os pontos juntos, o que descharacterizaria o objetivo inicial do agrupamento, da mesma maneira que um conjunto onde K=n colocaria cada ponto em um grupo unitário distinto, também fazendo com que o método K-médias perca o sentido (Zhao, Deng, & Ngo, 2018).

3.4.2 Método Elbow

Alternativamente, uma análise da progressão das médias dos desvios pode ser obtida analisando-se a curva dos índices de variância obtidos da análise K-Médias em função da variação do número K de *clusters*. Idealmente, os valores médios da soma das variâncias tende a decrescer exponencialmente com o aumento do número de grupos. O cenário ideal para a escolha do número de *clusters* está atrelada à escolha do menor número K dos mesmos para qual a variação do índice de variância entre K e K+1 não seja mais significativa. É importante lembrar que fatores qualitativos podem interferir na escolha do número de agrupamentos, bem como fatores de forma das distribuições podem influenciar no resultado, pois não são levadas em consideração na aplicação do método K-Médias.

4 MÉTODO

O método adotado para realização do estudo pode ser dividido em quatro partes:

4.1 Aquisição dos dados

A etapa de aquisição dos dados consiste em obter a maior quantidade possível de dados sobre manutenção de equipamentos, tipos de equipamentos, razões de parada, tempos de manutenção dentre outros. Não somente é importante a variedade dos dados, como a quantidade. Quanto mais informações sobre um mesmo equipamento em um determinado intervalo de tempo, mais precisas se tornam as análises de comportamento histórico do mesmo. Diferenças entre idade e porte dos equipamentos podem resultar em diferentes MTBF e MTTR dentro de uma mesma categoria, mas devido às limitações da base de dados, estes tiveram que ser desconsiderados. Uma das premissas adotadas para este estudo é a de que não há diferenças entre o desempenho de equipamentos nos quais uma manutenção foi realizada de modo a substituir completamente determinadas partes ou peças de outro cujos reparos foram feitos mantendo as peças que apresentaram falhas (Barberá, Crespo, Viveros, & Kristjanpoller, 2012).

4.2 Validação e triagem dos dados

Antes da realização de análises sobre os dados obtidos, é necessária uma validação cuidadosa e criteriosa das informações para garantir representatividade e acurácia. Equipamentos que gerem poucas informações em relação à média devem, possivelmente, ser desconsiderados. Da mesma maneira, tabelas que contenham informações inválidas, como datas e intervalos de tempo negativos, campos de chave única duplicados ou inexistentes, entre outros valores incorretos, devem ser eliminados e descartados. Muitas vezes, as informações são adicionadas no banco de dados de maneira incorreta e esta possibilidade se torna muito mais provável quando a inserção destas informações depende de um operador ou intermediador humano. A etapa de validação é a etapa mais importante do estudo, uma vez que se os dados iniciais forem de baixa qualidade, os resultados decorrentes perderão validade. Entender, identificar e rearranjar dados tendenciosos, imprecisos ou faltantes é uma tarefa de extrema importância para garantir a aplicabilidade e confiabilidade de métodos estatísticos (Olhede & Wolfe, 2018).

4.3 Análise dos Dados

4.3.1 Montagem de códigos em SQL

O Microsoft SQL Server foi a ferramenta utilizada para viabilizar as análises do estudo. Assim, o cuidado com a programação cuidadosa dos códigos empregados nos modelos é essencial. Foi preciso conhecer o banco de dados para poder realizar a junção de tabelas por meio de chaves, assim como todas as condições iniciais de análise (como período, equipamentos e outros) que tiveram que ser organizadas previamente, de maneira clara e unificada.

4.3.2 Agrupamento em *clusters*

A identificação de *clusters* e consequente agrupamento dos diferentes equipamentos foi feita com base no método K-Médias. Inicialmente, foi necessário determinar quais os tipos de equipamentos que deveriam ser agrupados, bem como quais as variáveis seriam utilizadas para agrupá-los. Depois, aplicou-se o método para que fosse encontrado o número ótimo de *clusters* e seus integrantes. Esta parte do estudo foi realizada com o auxílio do programa desenvolvido em Python.

4.4 Análise e discussão

A análise dos resultados teve como foco a identificação das características específicas de *clusters* de equipamentos, bem como a avaliação do desempenho das estratégias de manutenção adotadas para os mesmos, através das análises de MTBF, MTTR, DF e custos de manutenção.

Por fim, buscou-se identificar quais os grupos críticos de equipamentos representados por um determinado cluster e discutir hipóteses para possíveis soluções para este cenário.

5 ANÁLISE DOS RESULTADOS

O processo de aquisição e tratamento dos dados se deu com o auxílio de três ferramentas básicas:

- Microsoft SQL Server: O banco de dados fornecido pela empresa encontra-se em formato SQL. Foram montados códigos para possibilitar o tratamento e a extração dos dados brutos a serem analisados.

- Python 3.7: Python foi a linguagem de programação utilizada para a execução das análises estatísticas dos métodos Elbow e K-Médias, bem como a elaboração dos gráficos de agrupamento.
- Microsoft Excel: O Excel foi escolhido para realizar o tratamento final dos dados, principalmente no cruzamento dos dados de *clusters* gerados pelo Python com os dados de custos de manutenção obtidos do SQL.

5.1 Aquisição das bases de dados

A análise proposta para este estudo foi feita com base nos dados da empresa no ano de 2013. Como o banco de dados não fornecido de maneira integral – por motivos de sigilo empresarial – o ano de 2013 foi escolhido como intervalo de estudo por conter a maior quantidade de informações inalteradas e representativas da base. Os códigos gerados para as análises estão disponíveis no Apêndice 1.

5.1.1 Número de equipamentos

Para definição do número de equipamentos contidos no estudo, inicialmente foram determinadas quais as categorias de equipamentos que seriam estudadas.

Com base na qualidade e relevância das informações, foram selecionadas 4 categorias de equipamentos:

1. caminhões (*truck*);
2. escavadeiras (*shovel*);
3. perfuratrizes (*drill*); e
4. tratores de terraplanagem (*dozer*).

A Tabela 1 indica quantos equipamentos de cada categoria foram analisados.

Tabela 1 - Quantidade de equipamentos por categoria

CATEGORIA	EQUIPAMENTOS
Perfuratriz	6
Caminhão	27
Trator de terraplanagem	11
Escavadeira	11
TOTAL	55

5.1.2 Definição do MTBF e MTTR para cada equipamento

Para definição dos tempos médios entre falhas e de manutenção de cada equipamento, primeiramente foi necessário realizar uma triagem nos dados e consequente correção dos horários referentes ao início e término dos eventos. Registros com intervalos nulos ou muito pequenos em relação a eventos anteriores foram corrigidos, de modo que intervalos de manutenção inferiores a uma hora foram condensados e considerados como sendo um mesmo evento, o que exigiu extenso trabalho de tratamento e agrupamento dos dados fornecidos no banco.

Todos os eventos de manutenção reportados foram considerados, independentemente de sua categoria, desde que possuíssem intervalo de tempo superior a uma hora. O código SQL final fornecendo o MTBF e MTTR de cada equipamento para o ano de 2013 pode ser consultada no Apêndice 1. A

Tabela 2 fornece os dados referentes aos valores de MTBF e MTTR por equipamento, bem como seus respectivos códigos e categorias.

Tabela 2 – Valores de MTBF e MTTR, por equipamento

CATEGORIA	EQUIPAMENTO	MTTR (horas)	MTBF (horas)	DF (%)
Trator de terraplanagem	DZ501	9,47	57,18	86
Trator de terraplanagem	DZ502	10,15	55,20	84
Trator de terraplanagem	DZ503	10,35	49,77	83
Trator de terraplanagem	DZ504	10,56	44,88	81
Trator de terraplanagem	DZ505	11,69	50,90	81
Trator de terraplanagem	DZ507	10,77	34,86	76
Trator de terraplanagem	DZ511	11,60	27,86	71
Trator de terraplanagem	DZ514	9,19	58,68	86
Trator de terraplanagem	DZ515	10,78	55,39	84
Trator de terraplanagem	DZ520	12,94	21,84	63
Trator de terraplanagem	DZ521	14,58	36,27	71
Perfuratriz	DL101	7,98	52,70	87
Perfuratriz	DL110	6,10	23,63	79
Perfuratriz	DL111	5,88	25,89	81
Perfuratriz	DL130	6,90	31,81	82
Perfuratriz	DL131	5,99	35,53	86
Perfuratriz	DL160	6,95	24,97	78
Escavadeira	EX201	3,61	18,13	83
Escavadeira	EX205	9,85	41,13	81
Escavadeira	EX207	7,44	43,17	85
Escavadeira	EX213R	7,78	42,02	84
Escavadeira	LR301	6,83	48,24	88
Escavadeira	LR303	7,87	53,61	87
Escavadeira	LR305	4,84	22,34	82
Escavadeira	SH206	3,07	15,99	84
Escavadeira	SH210	3,72	19,39	84
Escavadeira	SH212	3,91	15,85	80
Escavadeira	SH215	2,38	19,38	89
Caminhão	HA402	5,99	29,37	83
Caminhão	HA403	7,87	29,54	79
Caminhão	HA404	4,91	53,16	92
Caminhão	HA405	8,62	41,48	83
Caminhão	HA420	6,91	22,34	76
Caminhão	HA421	6,41	29,38	82
Caminhão	HA422	6,80	19,84	74
Caminhão	HA423	6,18	34,22	85
Caminhão	HA424	5,23	26,80	84
Caminhão	HA425	4,94	43,62	90
Caminhão	HA426	4,95	33,20	87
Caminhão	HA427	6,88	23,55	77
Caminhão	HA452	7,23	31,58	81
Caminhão	HA456	7,87	30,25	79
Caminhão	HA457	9,02	21,47	70
Caminhão	HA460	8,49	31,44	79
Caminhão	HA471	6,61	43,69	87
Caminhão	HA472	6,03	32,75	84

CATEGORIA	EQUIPAMENTO	MTTR (horas)	MTBF (horas)	DF (%)
Caminhão	HA473	4,66	36,46	89
Caminhão	HA474	6,09	30,82	83
Caminhão	HA475	5,38	45,74	89
Caminhão	HA476	5,46	49,45	90
Caminhão	HA477	6,63	38,90	85
Caminhão	HA478	4,86	34,96	88
Caminhão	HA479	5,55	32,73	86
Caminhão	HA480	2,75	43,73	94
Caminhão	HA490	4,15	45,76	92

5.1.3 Definição dos custos de manutenção anual por equipamento

Por fim, a última base gerada por aquisição diretamente do banco de dados foi a base de custos de manutenção por equipamento para o ano de 2013. Foram considerados todos os custos ligados a um determinado equipamento, desde que relacionados a eventos de manutenção, incluindo, por exemplo, compra de peças de reposição e serviços de manutenção e reparo. Não foi computada nesta análise o valor indireto das perdas de produtividade por período ocioso.

O código SQL para obtenção dos custos pode ser conferido no Apêndice 1, enquanto que a Tabela 3 fornece os dados de custos de manutenção por equipamento.

Tabela 3 – Custos de manutenção no ano de 2013, por equipamento

CATEGORIA	EQUIPAMENTO	CUSTO (US\$)	CATEGORIA.	EQUIPAMENTO.	CUSTO (US\$).
Trator de terraplanagem	DZ501	287.114,89	Caminhão	HA402	168.409,32
Trator de terraplanagem	DZ502	281.032,59	Caminhão	HA403	383.158,20
Trator de terraplanagem	DZ503	286.573,07	Caminhão	HA404	186.748,58
Trator de terraplanagem	DZ504	479.196,51	Caminhão	HA405	617.712,63
Trator de terraplanagem	DZ505	365.310,38	Caminhão	HA420	1.569.115,83
Trator de terraplanagem	DZ507	464.706,32	Caminhão	HA421	614.830,68
Trator de terraplanagem	DZ511	592.191,28	Caminhão	HA422	678.372,45
Trator de terraplanagem	DZ514	222.840,47	Caminhão	HA423	562.004,92
Trator de terraplanagem	DZ515	407.717,02	Caminhão	HA424	530.727,66
Trator de terraplanagem	DZ520	83.219,14	Caminhão	HA425	310.354,31
Trator de terraplanagem	DZ521	30.120,48	Caminhão	HA426	925.910,65
Perfuratriz	DL101	148.279,61	Caminhão	HA427	613.471,42
Perfuratriz	DL110	507.626,34	Caminhão	HA452	159.018,35
Perfuratriz	DL111	661.377,33	Caminhão	HA456	862.166,95
Perfuratriz	DL130	1.344.762,53	Caminhão	HA457	109.043,03
Perfuratriz	DL131	935.632,13	Caminhão	HA460	286.977,34
Perfuratriz	DL160	961.659,90	Caminhão	HA471	584.642,66
Escavadeira	EX201	2.544.437,00	Caminhão	HA472	1.181.662,59

CATEGORIA	EQUIPAMENTO	CUSTO (US\$)	CATEGORIA.	EQUIPAMENTO.	CUSTO (US\$).
Escavadeira	EX205	362.065,95	Caminhão	HA473	1.071.911,82
Escavadeira	EX207	138.700,02	Caminhão	HA474	835.640,18
Escavadeira	EX213R	527.069,17	Caminhão	HA475	1.118.765,97
Escavadeira	LR301	416.422,18	Caminhão	HA476	1.222.323,50
Escavadeira	LR303	253.046,79	Caminhão	HA477	612.796,88
Escavadeira	LR305	1.966.785,77	Caminhão	HA478	977.480,82
Escavadeira	SH206	1.726.899,47	Caminhão	HA479	443.880,54
Escavadeira	SH210	3.502.912,45	Caminhão	HA480	233.138,91
Escavadeira	SH212	4.128.838,92	Caminhão	HA490	247.807,42
Escavadeira	SH215	999.843,00	-	-	-

5.2 Análise estatística

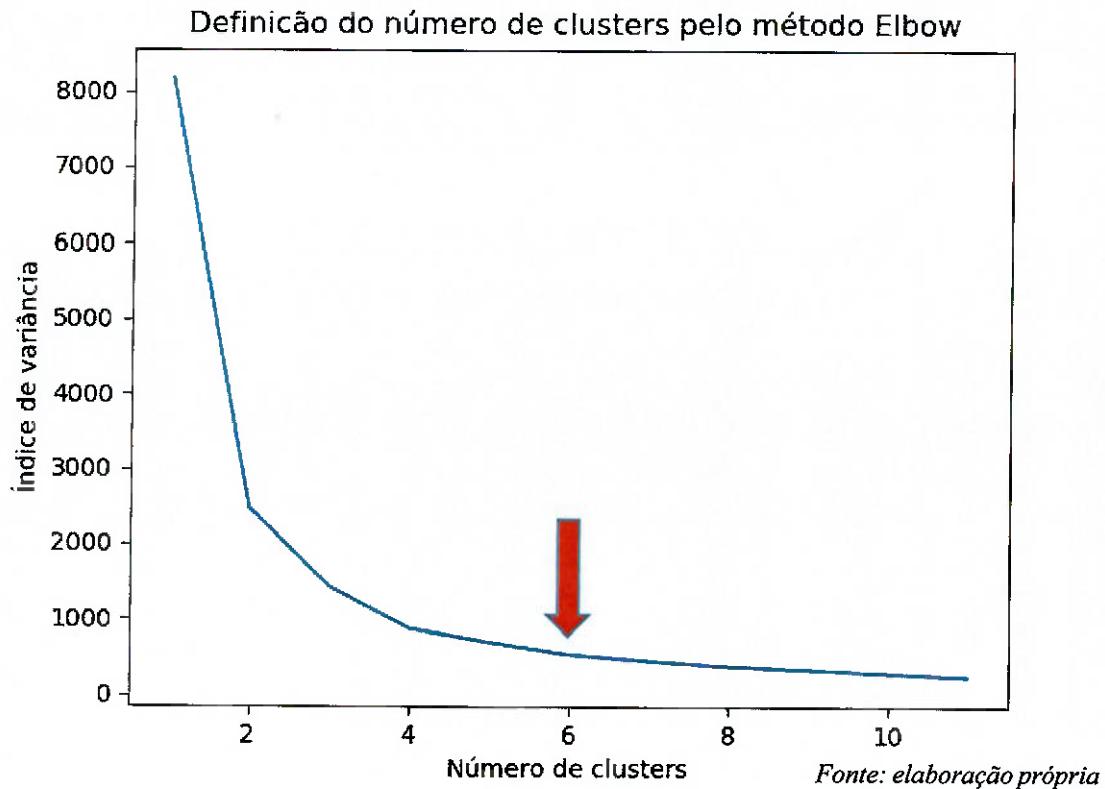
Nesta etapa, dois métodos analíticos foram utilizados:

- Método Elbow: método utilizado para definir o número ótimo de *clusters* a serem utilizados na análise estatística; e
- Método K-Médias: método estatístico de agrupamento em *clusters*.

5.2.1 Definição do número de *clusters* pelo método Elbow

O método Elbow, desenvolvido em linguagem Python, foi executado para avaliar as médias das variâncias entre os pontos e seus centroides para cenários entre 1 e 11 *clusters*. O Gráfico 1 indica a progressão destes valores.

Gráfico 1 – O método Elbow para definição do número de *clusters*



É possível verificar que o valor do índice de variância é máximo para o cenário com apenas um cluster. Nesta situação, as distâncias de todos os pontos são computadas em relação a apenas um único centroide, o que eleva estes valores. O valor do índice depois decresce exponencialmente até atingir um nível mais estável, próximo de 300. Para uma boa análise, deve-se optar, idealmente, pelo menor número de *clusters* que garantam um baixo valor para o índice de variância.

Assim, o número de *clusters* escolhido para a análise pelo método K-Médias foi 6.

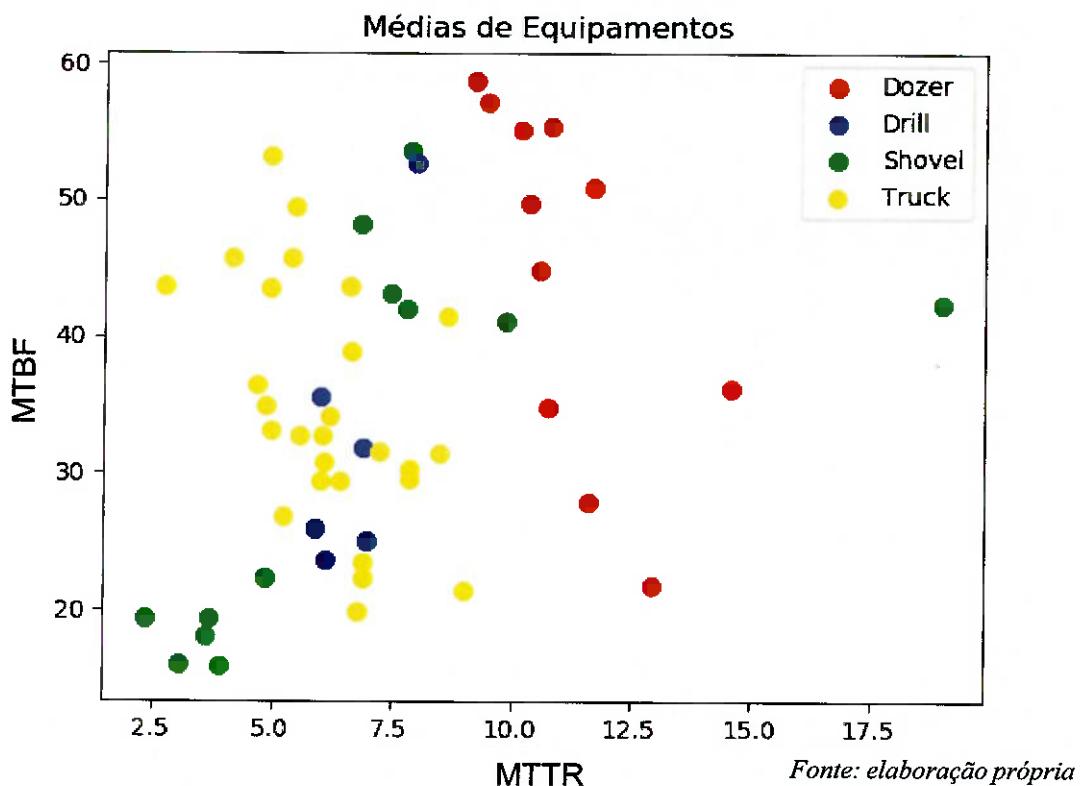
5.2.2 Clusterização pelo método K-Médias

Para a clusterização dos equipamentos, foi aplicado o método K-Médias com o auxílio da biblioteca de Python *sklearn*, enquanto que para a montagem dos gráficos, foi utilizada a biblioteca *matplotlib*. Os parâmetros de análise utilizados no método K-Médias foram:

- $n_clusters = 6$, para o número de *clusters*;
- $n_init = 50$, para o número de execuções com diferentes centroides iniciais, das quais será selecionada a melhor configuração inicial, com base na somatória das médias das distâncias, como descrito no item 3.3.1; e
- $max_iter = 500$, para o número de iterações realizadas em uma única execução.

A base com os dados médios de MTTR e MTBF foi utilizada como alvo do estudo de agrupamento. Inicialmente, os dados foram colocados no gráfico sem tratamento estatístico, para que se pudesse visualizar a distribuição dos equipamentos, como ilustrado no Gráfico 2. Para os gráficos de MTTR e MTBF, os valores são medidos em horas.

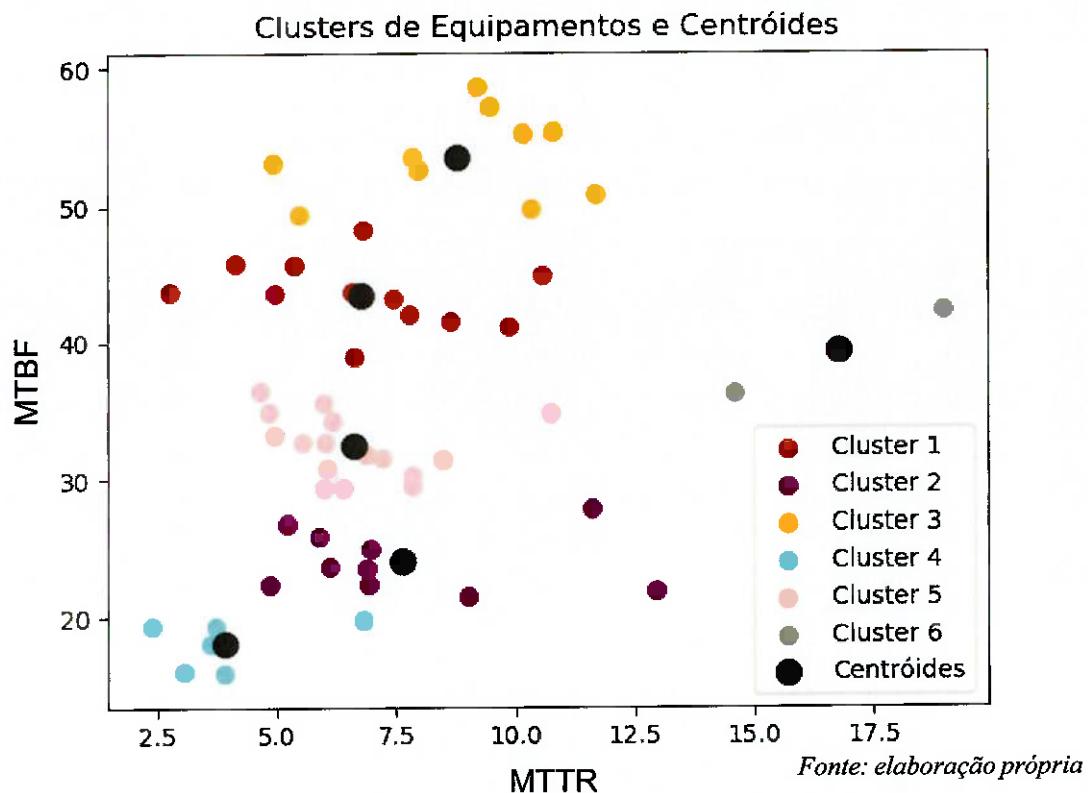
Gráfico 2 – Distribuição dos equipamentos de acordo com os valores de MTTR e MTBF para o ano de 2013.



Sobre estes dados foi aplicada a análise K-Médias, para identificação dos *clusters* de modo a minimizar a somatória das variâncias entre os pontos e os

centroides de seus respectivos *clusters*. O resultado da análise é apresentado no Gráfico 3.

Gráfico 3 – *Clusters* resultantes da aplicação do método K-Médias



Com os *clusters* já definidos, realizou-se uma análise reversa para identificar quais equipamentos compõe cada cluster. Estes resultados são apresentados nos Gráficos 4 a 9.

Gráfico 4 – cluster 1 e equipamentos

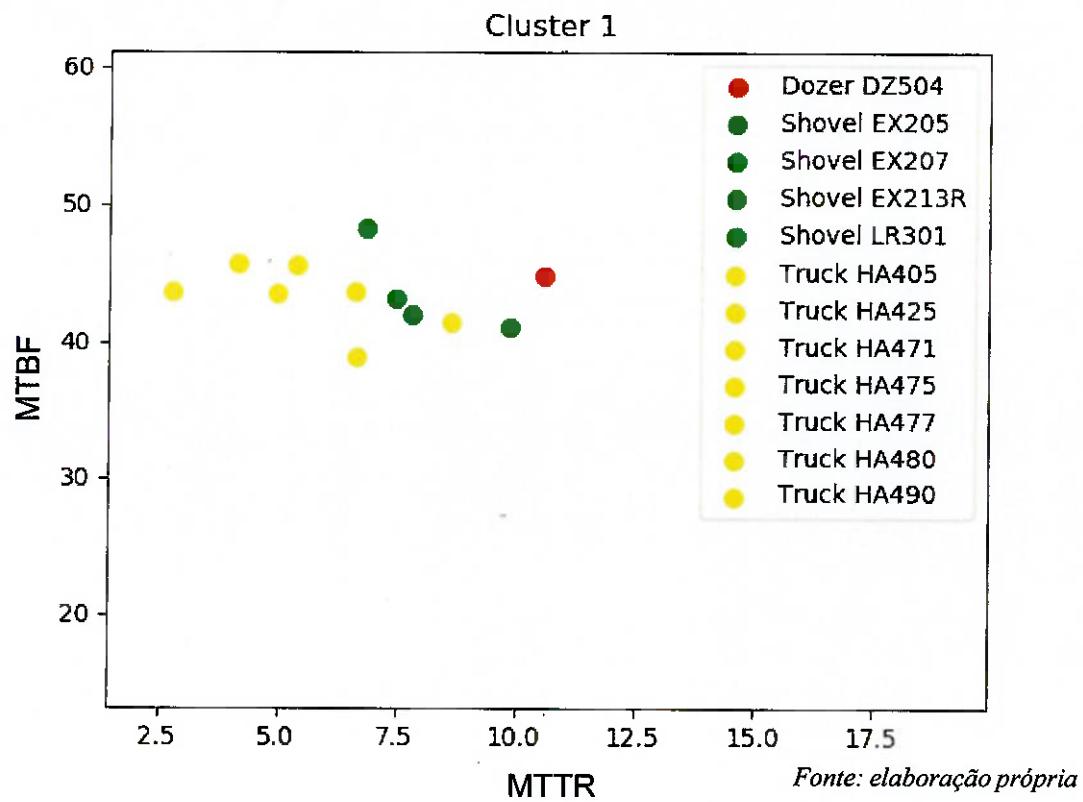


Gráfico 5 – cluster 2 e equipamentos

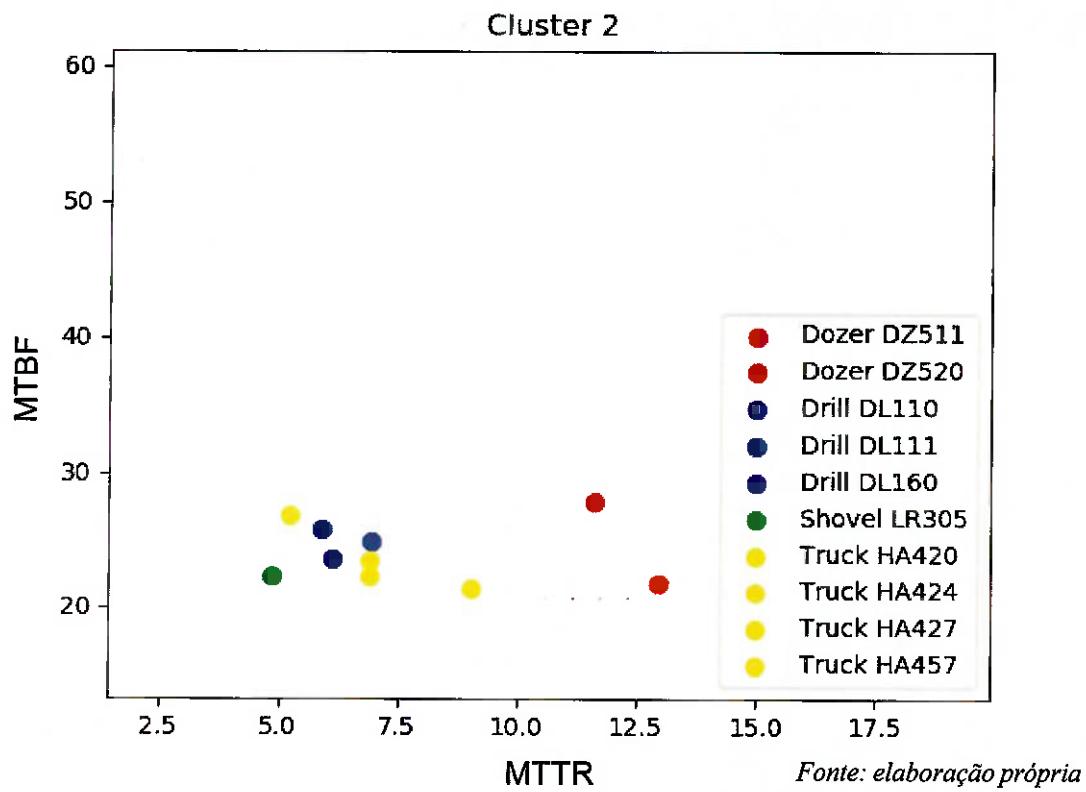


Gráfico 6 – cluster 3 e equipamentos

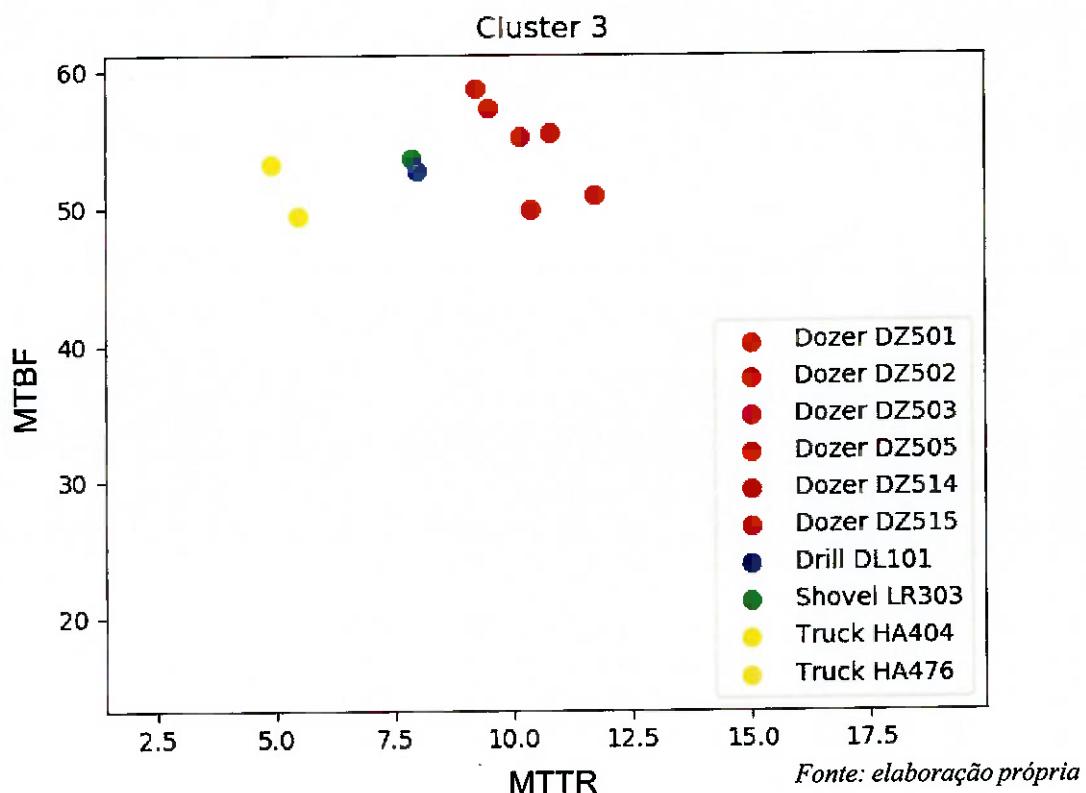


Gráfico 7 – cluster 4 e equipamentos

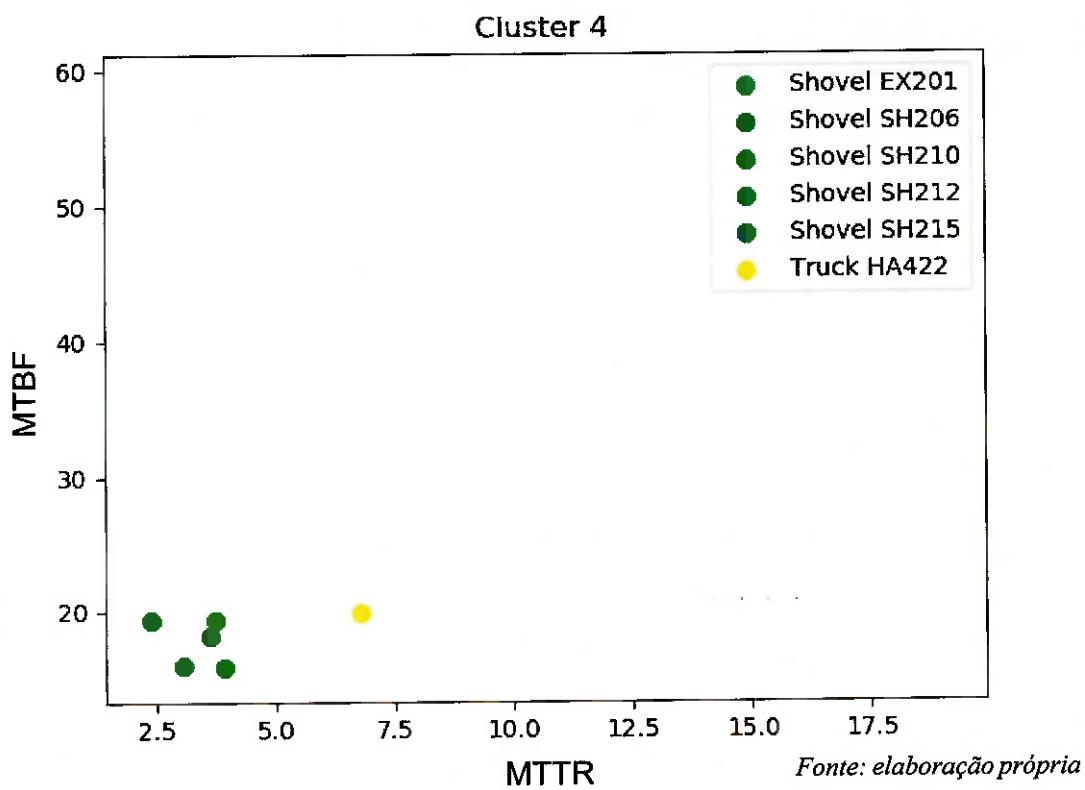


Gráfico 8 – cluster 5 e equipamentos

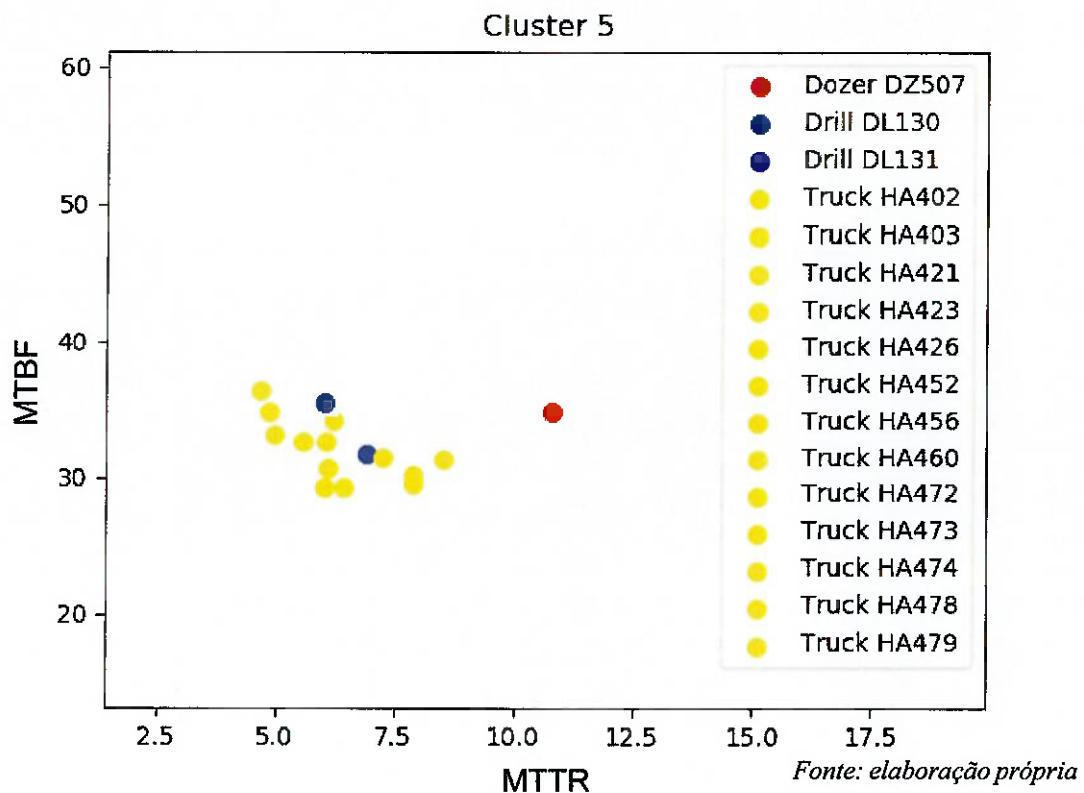
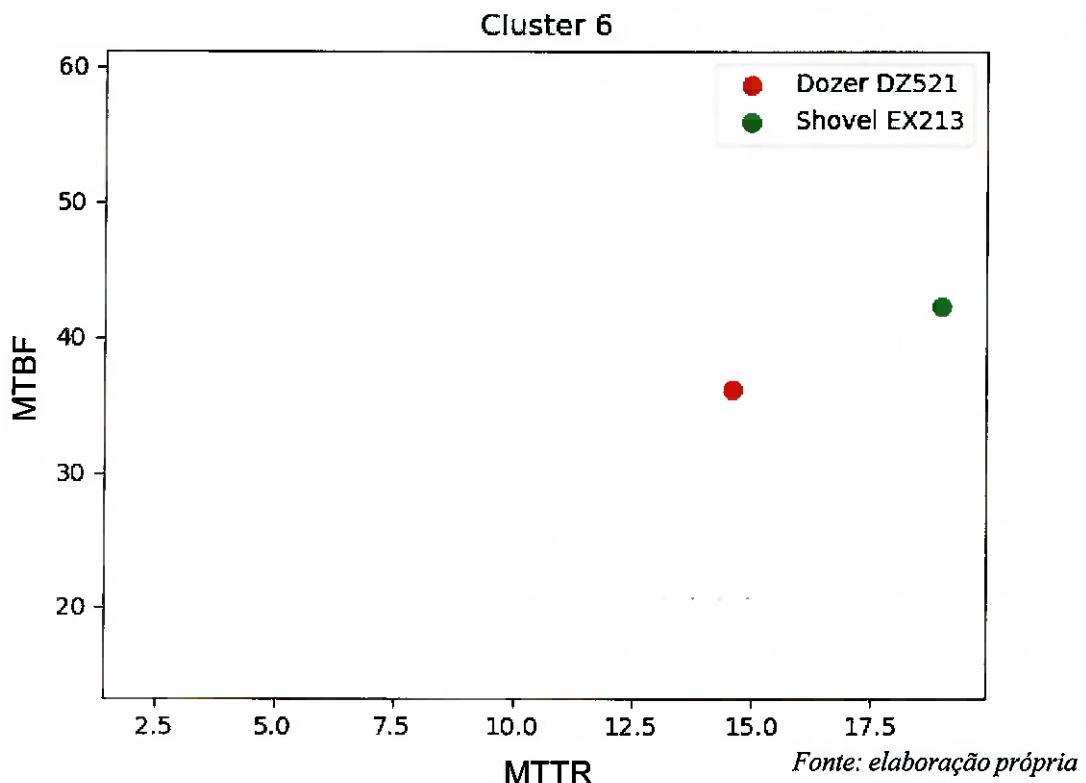


Gráfico 9 – cluster 6 e equipamentos



O código em Python criado para estas análises, bem como para a elaboração dos gráficos, pode ser conferido no Anexo 2.

5.3 Análise comparativa de custos

Com os *clusters* e seus componentes já definidos, foi realizada a análise comparativa dos custos de manutenção obtidos via banco de dados com o resultado do agrupamento.

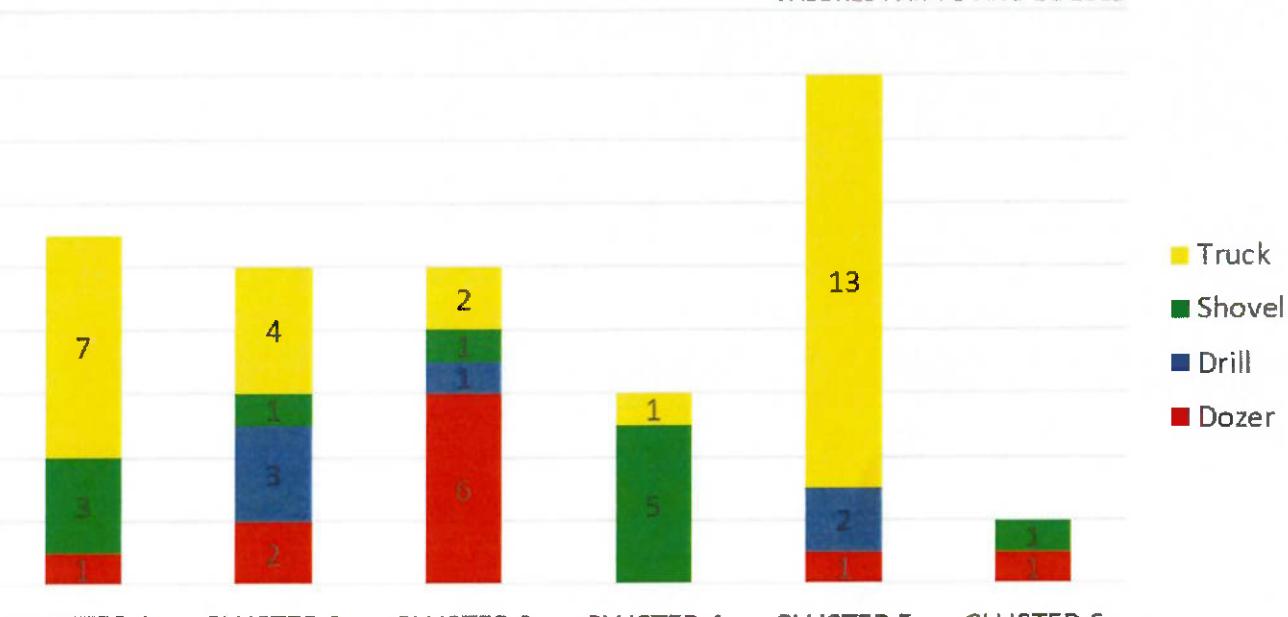
5.3.1 Equipamentos por cluster

Primeiramente, foi definida a quantidade e composição dos equipamentos dentro de cada cluster, ilustrados no Gráfico 10.

Gráfico 10 – Número de equipamentos por cluster

NÚMERO DE EQUIPAMENTOS*

* VALORES PARA O ANO DE 2013



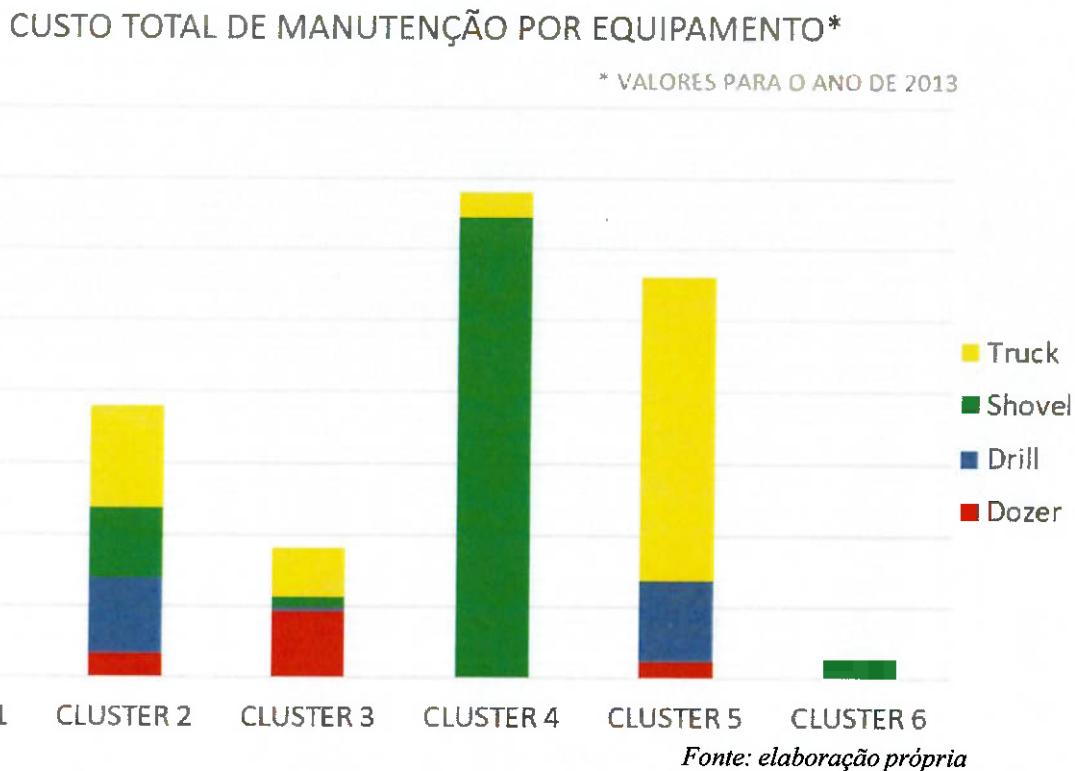
Fonte: elaboração própria

O gráfico ilustra a distribuição dos equipamentos em cada cluster. Vale notar a maior concentração de caminhões nos *clusters* 1 e 5, bem como o predomínio dos tratores de terraplanagem no *cluster* 3 e de escavadeiras no 4.

5.3.2 Análises de custo

Posteriormente, os valores de custos de manutenção dos equipamentos foram atribuídos aos *clusters*, de acordo com sua categoria, como mostrado do Gráfico 11.

Gráfico 11 – Custos totais de manutenção

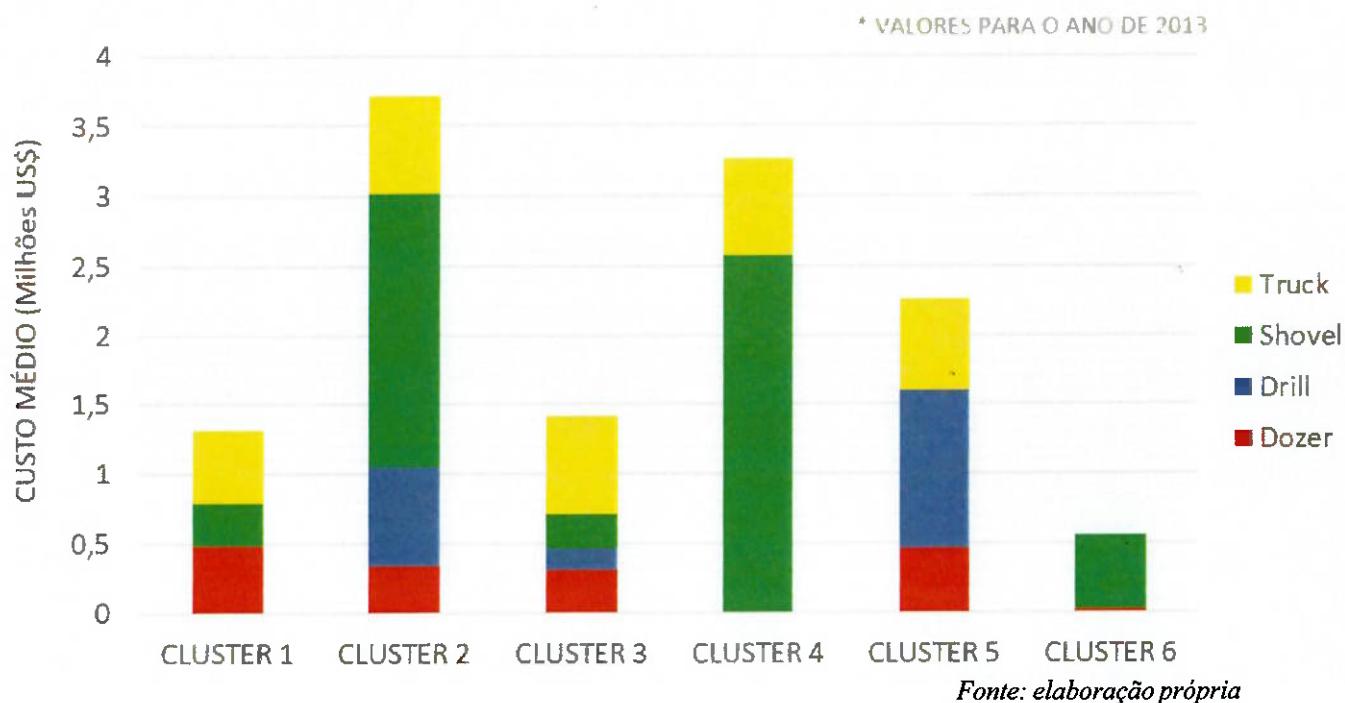


Fica evidente o elevado custo de manutenção gerado pelas escavadeiras do *cluster* 4, que mesmo sendo o segundo grupo com menos equipamentos, possui o maior valor total de gastos com manutenção.

Esta análise, porém, não fornece uma ideia real dos custos unitários com os equipamentos de cada *cluster* e pode esconder informações importantes. O Gráfico 12 busca balancear os diferentes tamanhos e composições dos grupos, fornecendo o custo unitário médio de manutenção de um equipamento de uma determinada categoria dentro de seu respectivo *cluster*.

Gráfico 12 – Custo médio de manutenção de um equipamento dentro de cada cluster, por categoria

CUSTO MÉDIO POR CATEGORIA DE EQUIPAMENTO*



5.3.3 Categoria Crítica

De acordo com os dados de custos e tempos médios obtidos anteriormente é possível estabelecer a categoria de escavadeiras como “Categoria Crítica”, ou seja, a categoria que mais demanda ações imediatas visando melhorar o desempenho de equipamentos deste grupo.

Foram analisados os tempos médios de duração de manutenções e intervalos entre estes eventos para os equipamentos desta categoria, agrupados nos clusters 2 e 4 e nos clusters 1, 3 e 6 por terem se comportado de maneira similar diante dos custos de manutenção por equipamento. Foi incluída nesta etapa a quebra em manutenções preventiva e corretiva, para se ter uma visão mais detalhada dos equipamentos.

O Gráfico 13 ilustra os tempos médios de duração de manutenções, enquanto o Gráfico 14 mostra os intervalos médios entre as mesmas.

Gráfico 13 – MTBF de escavadeiras, por grupo de clusters similares

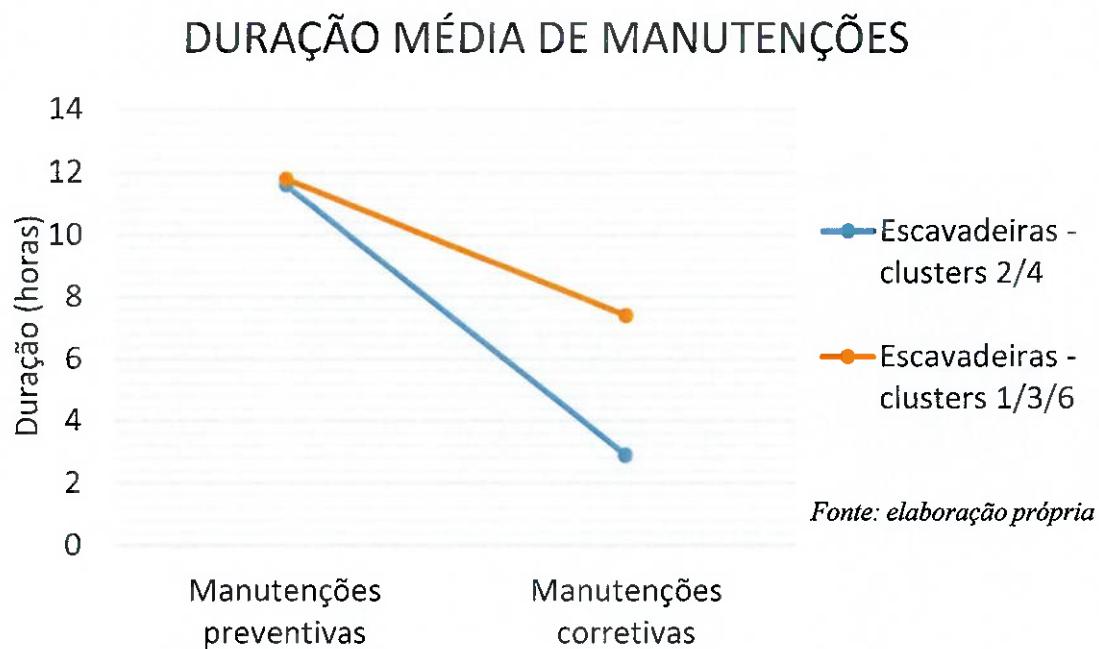
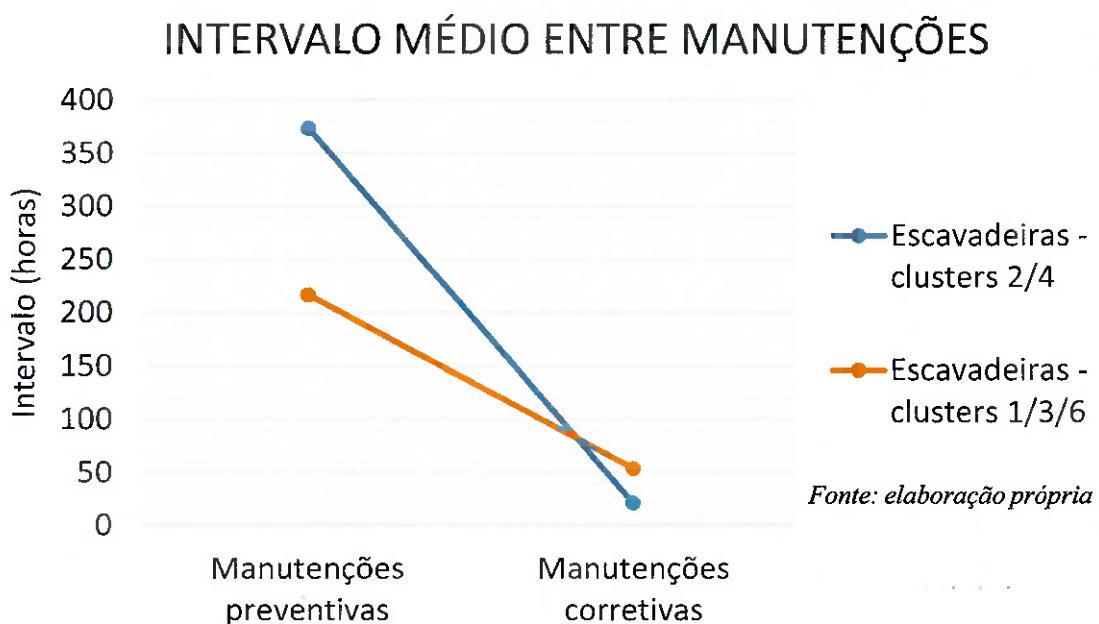


Gráfico 14 – Intervalo médio entre manutenções de escavadeiras, por grupo de clusters similares



6 DISCUSSÃO

Analizando o Gráfico 2, é possível verificar que a lógica de distribuição dos equipamentos não é a mesma para diferentes categorias. No entanto, é possível identificar alguns padrões intrínsecos a determinados tipos de equipamento. Ao comparar o Gráfico 2 com o Gráfico 3, nota-se que a clusterização não necessariamente reflete a segregação dos equipamentos por categoria, o que possibilita tratar equipamentos de categorias diferentes de maneiras similares.

Com os Gráficos 4 a 9 torna-se possível identificar as características de cada cluster. O *cluster* 1 é formado basicamente por caminhões e escavadeiras, com MTBF entre 40h e 50h e MTTR que variam de 2,5h a 11h. O *cluster* 2 é um grupo heterogêneo, contendo pelo menos um equipamento de cada categoria, no qual se destaca o TTR mais elevado dos tratores de terraplanagem em relação aos demais membros do grupo, mas todos possuindo TBF's entre 20h e 30h. O *cluster* 3 se destaca por conter predominantemente tratores – também de alto TTR – e por apresentar os maiores valores de MTBF dentre os outros *clusters*. O *cluster* 4 é composto basicamente por escavadeiras, contendo apenas um único caminhão, que nitidamente se destaca do grupo. Este *cluster* possui os menores valores de MTTR e MTBF e é o segundo menor grupo em número de equipamentos (apenas 6). O *cluster* 5 é marcado pela predominância dos caminhões. É também o *cluster* com o maior número de equipamentos totais, como ilustrado no Gráfico 10. Por fim, o *cluster* 6 agrupa apenas dois equipamentos – um trator e uma escavadeira – ambos caracterizados por possuírem os mais altos valores de TTR de toda a amostra, mas com valores de MTBF próximos de 40h.

Ao comparar os dados dos Gráficos 10 e 11, percebe-se que o *cluster* 4 apresenta uma clara discrepância em relação aos demais, na medida em que concentra cerca de 33% dos custos de manutenção do ano, concentrados em apenas 7,3% dos equipamentos, em contraste com o *cluster* 5 que concentra 26,7% dos gastos em aproximadamente 30% dos equipamentos. Nota-se que o MTBF dos equipamentos do *cluster* 5 chega a ser de 1,5 a 4 vezes maior do que o dos equipamentos do *cluster* 4.

Na comparação entre os custos médios por equipamento, dois pontos se tornam mais relevantes. Primeiramente, o custo unitário médio dos caminhões se mantém aproximadamente constante, em torno de US\$ 500.000, independentemente do *cluster* ao qual pertencem. No comparativo com o Gráfico 2, isso indica que para intervalos de TBF entre 20 e 55 h, e de TTR entre 2,5 e 9 h, os custos de manutenção dos caminhões variam muito pouco.

Por outro lado, o Gráfico 12 evidencia que os custos de manutenção das escavadeiras dos *clusters* 2 e 4 (entre US\$ 2.000.000 e US\$ 2.500.000) chega a ser até 10 vezes maior do que os custos unitários das escavadeiras dos *clusters* 1, 3 e 6. Em correlação com os dados do Gráfico 2, é possível observar que as escavadeiras dos *clusters* 2 e 4 possuem MTBF muito inferior às demais, indicando que falham com mais frequência. Dentre as demais, o MTBF é sempre superior a 40 h, enquanto que a variação do MTTR está entre 7,5 e 19 h. Desta forma, pode-se dizer que as escavadeiras com baixo TBF e baixo TTR geram muito mais custo de manutenção do que as de TBF alto e TTR mais elevado.

Após definição da categoria de escavadeiras como “Categoria Crítica” e a consequente comparação dos intervalos e durações de manutenções destes equipamentos nos *clusters* 2 e 4 e nos *clusters* 1,3 e 6, foi possível observar que a duração média das manutenções preventivas segue um padrão de aproximadamente 11,5 h para todos os equipamentos, independentemente do grupo de *clusters*, como mostrado pelo Gráfico 13. No entanto, a duração média das manutenções das escavadeiras dos *clusters* 2 e 4 chega a ser cerca de 2,5 vezes menor do que das demais, o que pode indicar a ocorrência de falhas de menor complexidade. Na comparação destes dados com os apresentados no Gráfico 14, verifica-se que o intervalo de tempo médio entre ocorrências com as escavadeiras dos *clusters* 2 e 4 que exigem manutenções corretivas é de aproximadamente 2,6 vezes menor do que este mesmo indicador para as demais, levantando a hipótese de que os equipamentos do primeiro grupo estão sujeitos a falhas de menor complexidade com alta frequência. Para complementar esta hipótese, observa-se que o intervalo médio entre manutenções preventivas dos equipamentos do primeiro grupo é cerca de 1,72 vezes maior do que das escavadeiras dos *clusters* 1, 3 e 6, o que pode indicar que a maior incidência de falhas nas escavadeiras dos *clusters* 2 e 4 esteja

relacionada ao maior intervalo entre manutenções preventivas destes mesmos equipamentos.

7 CONCLUSÕES

No âmbito pontual do estudo de caso realizado por este estudo, é possível concluir que equipamentos de uma mesma categoria podem apresentar características muito diferentes quanto ao seu desempenho e custo de manutenção. Assim, análises derivadas de agrupamentos mais abrangentes podem fornecer parâmetros para detectar problemas em grupos de equipamentos que fogem das classificações usuais. Foi possível identificar, por exemplo, que um tratamento unificado para todas as escavadeiras por meio de uma média do grupo não refletiria as heterogeneidades presentes dentro deste mesmo conjunto. Ao mesmo tempo, foi possível identificar que, apesar da grande variabilidade dos dados de desempenho de operação e manutenção, os caminhões apresentam custos médios de manutenção anual muito próximos uns dos outros, o que indica que ações tomadas de modo a melhorar MTBF e MTTR dos caminhões de melhor desempenho não surtiram tanto efeito do ponto de vista financeiro.

Quanto ao grupo crítico definido pelas escavadeiras, as análises mostram que, embora as causas não possam ser comprovadas sem uma análise qualitativa mais aprofundada, é possível que os intervalos de manutenções preventivas destes equipamentos nos *clusters* 2 e 4 devam ser revistos e reajustados para evitar a grande incidência de falhas que exigem manutenções corretivas. Do ponto de vista de produção, as manutenções corretivas podem custar ainda mais caro à operação pois a elas devem ser atrelados os custos de oportunidade decorrentes da interrupção não-planejada da produção.

As informações analisadas neste estudo são de extrema utilidade para uma empresa de mineração, pois auxilia na tomada de decisões de maneira a focar nas ações corretivas e de melhoria de desempenho que renderão maior impacto em cada *cluster* de equipamento, ao mesmo tempo que evidencia quais os grupos mais críticos, que demandam ações mais imediatas e maiores investimentos.

Numa escala macroscópica, este estudo também ilustra a importância das aplicações de análises de *Data Science* no ambiente de mineração. A aplicação sobre desempenho e custos de manutenção de equipamentos é apenas uma das várias possibilidades de estudos a serem realizados em relação à atividade mineradora com o emprego de análises de bancos de dados e abre portas para a conscientização da importância de se fomentar, incentivar e cultivar uma cultura de dados no meio da mineração.

8 REFERÊNCIAS BIBLIOGRÁFICAS

- Barberá, L., Crespo, A., Viveros, P., & Kristjanpoller, F. (2012). RAM analysis of mining process: a case study of a Copper Smelting Process in the field of mining, Chile. *IFAC Proceedings Volumes*, 45(31), 217–222. <https://doi.org/10.3182/20121122-2-ES-4026.00002>
- Ceri, S. (2018). On the role of statistics in the era of big data: A computer science perspective. *Statistics & Probability Letters*, 136, 68–72. <https://doi.org/10.1016/J.SPL.2018.02.019>
- Dhillon, B. S. *Maintainability, Maintenance, and Reliability for Engineers*. Taylor & Francis Group, LLC. Boca Raton, FL. pp 3, 2006.
- Dindarloo, S. R., & Siami-Irdemoosa, E. (2017). Data mining in mining engineering: results of classification and clustering of shovels failures data. *International Journal of Mining, Reclamation and Environment*, 31(2), 105–118. <https://doi.org/10.1080/17480930.2015.1123599>
- Gölbaşı, O., & Demirel, N. (2017). A cost-effective simulation algorithm for inspection interval optimization: An application to mining equipment. *Computers & Industrial Engineering*, 113, 525–540. <https://doi.org/10.1016/J.CIE.2017.09.002>
- Hall, R. (1997). Analysis of Mobile Equipment Maintenance Data In an Underground Mine.
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700–710. <https://doi.org/10.1016/J.IJINFOMGT.2016.04.013>
- Lau, F. D.-H., Adams, N. M., Girolami, M. A., Butler, L. J., & Elshafie, M. Z. E. B. (2018). The role of statistics in data-centric engineering. *Statistics & Probability Letters*, 136, 58–62. <https://doi.org/10.1016/J.SPL.2018.02.035>
- Olhede, S. C., & Wolfe, P. J. (2018). The future of statistics and data science. *Statistics & Probability Letters*, 136, 46–50. <https://doi.org/10.1016/J.SPL.2018.02.042>
- Ouyang, Q., Wu, C., & Huang, L. (2018). Methodologies, principles and prospects of applying big data in safety science research. *Safety Science*, 101(March 2017), 60–71. <https://doi.org/10.1016/j.ssci.2017.08.012>
- Reid, N. (2018). Statistical science in the world of big data. *Statistics & Probability Letters*, 136, 42–45. <https://doi.org/10.1016/J.SPL.2018.02.049>
- Sundararajan, C.R. *Guide to Reliability Engineering*. Van Nostrand Reinhold, 1991.
- Sifferlinger, N., & Rath, G. (2016). Functional Safety and Mean Time to Fail for Underground Mining Proximity Detection Device in No-Go-Zones. *IFAC-PapersOnLine*, 49(20), 31–36. <https://doi.org/10.1016/J.IFACOL.2016.10.092>

Wang, Z. Q., & Gao, W. M. (2011). The study of management information system for coal mine safety quality standardization. *Procedia Engineering*, 26, 1970–1976. <https://doi.org/10.1016/j.proeng.2011.11.2392>

Zhao, W.-L., Deng, C.-H., & Ngo, C.-W. (2018). k-means: A revisit. *Neurocomputing*, 291, 195–206. <https://doi.org/10.1016/J.NEUCOM.2018.02.072>

APÊNDICE A

Códigos em SQL

I. Query para obtenção do TEF e TEM

```

1. ****
2. ** FERNANDO DE OLIVEIRA LIMA XAVIER - 2018/09/16 ***
3. ** TCC - QUERY PARA DETERMINAÇÃO DO TEM E TEF ***
4. ****
5.
6. DROP TABLE IF EXISTS #TEMP1
7. SELECT
8.     CONVERT(DATE,A.SHIFTDATE,112) AS DATA_EVENTO
9.     ,ROW_NUMBER() OVER (PARTITION BY A.EQMT ORDER BY A.SHIFTDATE, A.TOD_EVENT_START_DATETIME) AS ROWN
UM
10.    ,EQ.DESCRIPTION
11.    ,A.EQMT AS EQUIPAMENTO
12.    ,B.TIPO
13.    ,A.TOD_EVENT_START_DATETIME AS [INICIO MANUTENCAO]
14.    ,A.TOD_EVENT_END_DATETIME AS [FIM MANUTENCAO]
15. INTO #TEMP1
16. FROM [dbo].[Fact_StatusEvents] A
17. INNER JOIN (
18.     SELECT DISTINCT
19.         DESCRIPTION
20.         ,SUBCATEGORY
21.         ,EquipmentID
22.     FROM [dbo].[Dim_Equipment]
23.     WHERE
24.         DESCRIPTION IN ('TRUCK','SHOVEL','DOZER','DRILL')
25.         AND SUBCATEGORY <> 'UNKNOWN'
26.     ) AS EQ
27.     ON A.EQMT = EQ.EQUIPMENTID
28. LEFT JOIN
29. (
30.     SELECT DISTINCT
31.         CATEGORY AS CATEGORIA
32.         ,TIME_CODE_PV_CATEGORY AS TIPO
33.     FROM [dbo].[Dim_Time_Code]
34.     ) AS B
35.     ON A.category = B.CATEGORIA
36. WHERE
37.     CATEGORY IN (4,5)
38.     AND YEAR(A.SHIFTDATE) = 2013
39.     AND DATEDIFF(MINUTE,A.TOD_EVENT_START_DATETIME,A.TOD_EVENT_END_DATETIME) > 0
40.
41. =====
42.
43. DROP TABLE IF EXISTS #TEMP2
44. SELECT
45.     A.DATA_EVENTO
46.     ,ROW_NUMBER() OVER (PARTITION BY A.EQUIPAMENTO ORDER BY A.DATA_EVENTO, A.START_ATUAL) AS ROWNUM
47.     ,A.DESCRIPTION
48.     ,A.EQUIPAMENTO
49.     ,A.TIPO
50.     ,A.START_ATUAL
51.     ,A.END_ATUAL
52.     ,A.START_PROXIMO
53.     ,A.END_PROXIMO
54. INTO #TEMP2
55. FROM (
56.     SELECT
57.         A.DATA_EVENTO
58.         ,A.ROWNUM
59.         ,A.[DESCRIPTION]

```

```

60.      ,A.EQUIPAMENTO
61.      ,A.TIPO
62.      ,A.[INICIO MANUTENCAO] AS START_ATUAL
63.      ,A.[FIM MANUTENCAO] AS END_ATUAL
64.      ,B.[INICIO MANUTENCAO] AS START_PROXIMO
65.      ,B.[FIM MANUTENCAO] AS END_PROXIMO
66.  FROM #TEMP1 AS A
67.  LEFT JOIN #TEMP1 AS B
68.    ON A.EQUIPAMENTO = B.EQUIPAMENTO AND B.ROWNUM = (A.ROWNUM+1)
69.  ) A
70. WHERE DATEDIFF(HOUR,A.END_ATUAL,A.START_PROXIMO) > 1 OR A.ROWNUM = 1
71.
72. =====
73.
74. DROP TABLE IF EXISTS #TEMP3
75. SELECT *
76. INTO #TEMP3
77. FROM (
78.     SELECT
79.         A.DATA_EVENTO,
80.         A.ROWNUM,
81.         A.DESCRIPTION,
82.         A.EQUIPAMENTO,
83.         A.TIPO,
84.         A.START_ATUAL,
85.         B.END_ATUAL,
86.         B.START_PROXIMO
87.     FROM #TEMP2 AS A
88.     LEFT JOIN (
89.         SELECT
90.             ROWNUM,
91.             EQUIPAMENTO,
92.             END_ATUAL,
93.             START_PROXIMO
94.         FROM #TEMP2
95.     ) AS B
96.     ON A.EQUIPAMENTO = B.EQUIPAMENTO AND B.ROWNUM = (A.ROWNUM+1)
97.     WHERE A.ROWNUM = 1
98.
99. UNION ALL
100.
101.    SELECT
102.        A.DATA_EVENTO,
103.        A.ROWNUM,
104.        A.DESCRIPTION,
105.        A.EQUIPAMENTO,
106.        A.TIPO,
107.        A.START_PROXIMO AS START_ATUAL,
108.        B.END_ATUAL,
109.        B.START_PROXIMO
110.    FROM #TEMP2 AS A
111.    LEFT JOIN (
112.        SELECT
113.            ROWNUM,
114.            EQUIPAMENTO,
115.            END_ATUAL,
116.            START_PROXIMO
117.        FROM #TEMP2
118.    ) AS B
119.    ON A.EQUIPAMENTO = B.EQUIPAMENTO AND B.ROWNUM = (A.ROWNUM+1)
120.    WHERE A.ROWNUM <> 1
121. ) A
122. ORDER BY 4,2
123.
124.
125. =====

```

```

126.
127.
128. SELECT
129.     C.DESCRIPTION,
130.     C.EQUIPAMENTO,
131.     CAST(AVG(C.TEM) AS FLOAT) AS AVG_TEM,
132.     CAST(AVG(C.TEF) AS FLOAT) AS AVG_TEF
133. FROM (
134.     SELECT
135.         B.*,
136.         DATEDIFF(MINUTE,END_ATUAL,START_PROXIMO)/60.00 AS TEF
137.     FROM (
138.         SELECT
139.             A.*,
140.             DATEDIFF(MINUTE,START_ATUAL,END_ATUAL)/60.00 AS TEM
141.             FROM #TEMP3 AS A
142.         ) AS B
143.     ) AS C
144. WHERE EQUIPAMENTO <> 'DL140'
145. GROUP BY
146.     C.DESCRIPTION,
147.     C.EQUIPAMENTO
148. ORDER BY 1,2

```

II. Query para obtenção dos custos

```

1. ****
2. **          FERNANDO DE OLIVEIRA LIMA XAVIER - 2018/09/16      */
3. **          TCC - QUERY PARA DETERMINAÇÃO DOS CUSTOS DE MANUTENÇÃO  */
4. ****
5.
6. SELECT
7. *
8. FROM (
9.     SELECT
10.        EQ.DESCRIPTION
11.        ,PROJECT,
12.        CASE WHEN PROJECT IN ('DZ504','EX205','EX207','LR301','HA405','HA425','HA471','HA475','HA477'
13.              , 'HA480','HA490') THEN 'CLUSTER 1'
14.              WHEN PROJECT IN ('DZ511','DZ520','DL110','DL111','DL160','LR305','HA420','HA424','HA427'
15.              , 'HA457') THEN 'CLUSTER 2'
16.              WHEN PROJECT IN ('DZ501','DZ502','DZ503','DZ505','DZ514','DZ515','DL101','LR303','HA404'
17.              , 'HA476') THEN 'CLUSTER 3'
18.              WHEN PROJECT IN ('EX201','SH206','SH210','SH212','SH215','HA422') THEN 'CLUSTER 4'
19.              WHEN PROJECT IN ('DZ507','DL130','DL131','HA402','HA403','HA421','HA423','HA426','HA452'
20.              , 'HA456','HA460','HA472','HA473','HA474','HA478','HA479') THEN 'CLUSTER 5'
21.              WHEN PROJECT IN ('DZ521','EX213R') THEN 'CLUSTER 6'
22.        END AS cluster
23.        ,SUM([COST DEBIT]) as CUSTO
24.    FROM [dbo].[Fact_Cost_WH] AS FC
25.    INNER JOIN [dbo].[Dim_Equipment] AS EQ
26.        ON FC.PROJECT = EQ.EQUIPEMENTID
27.    WHERE YEAR(SHIFTDATE) = 2013
28.    GROUP BY PROJECT
29.        ,EQ.DESCRIPTION
30.    ) AS A
31. WHERE A.CLUSTER IS NOT NULL
32. ORDER BY 1,2

```

APÊNDICE B

Códigos em Python

I. Código Python para Elbow, K-Médias e plotagem dos gráficos

```

1. import pandas as pd
2. from sklearn.cluster import KMeans
3. import matplotlib.pyplot as plt
4.
5. ##### LEITURA DO ARQUIVO CSV E CRIACAO DA BASE #####
6. #          LEITURA DO ARQUIVO CSV E CRIACAO DA BASE      #
7. #####
8.
9. medias = pd.read_csv('C:\\\\Users\\\\Fernando Xavier\\\\Desktop\\\\MINAS\\\\TCC - 2018\\medias.csv')
10.
11. X = medias.iloc[:,2:4].values
12. Y = medias.iloc[:,0:4].values
13. Z = medias.iloc[:,1].values
14. print(X)
15. print(Y)
16.
17. def eqp(j):
18.     if j == 0:
19.         return('Dozer')
20.     elif j == 1:
21.         return('Drill')
22.     elif j== 2:
23.         return('Shovel')
24.     else:
25.         return('Truck')
26. #####
27. #          EXECUCAO ELBOW                                #
28. #####
29.
30. elb = []
31.
32. for i in range(1,12):
33.     km = KMeans(n_clusters=i,n_init=50,max_iter=500)
34.     km.fit(X)
35.     print(i,km.inertia_)
36.     elb.append(km.inertia_)
37.
38. plt.plot(range(1,12),elb)
39. plt.title('Definicão do número de clusters pelo método Elbow')
40. plt.xlabel('Número de clusters')
41. plt.ylabel('Índice de variância')
42. plt.show()
43.
44.
45. #####
46. #          EXECUCAO K-MEDIAS                            #
47. #####
48.
49. estimador = KMeans(n_clusters=6,n_init=50,max_iter=500)
50. #kmeans = estimador.fit(X)
51. kmeans = estimador.fit_predict(X)
52.
53. #####
54. #          PLOTAGEM DO GRÁFICO K-MEDIAS                #
55. #####
56.
57. plt.scatter(X[kmeans==0,0],X[kmeans==0,1], s=50, c = 'brown', label = 'Cluster 1')
58. plt.scatter(X[kmeans==1,0],X[kmeans==1,1], s=50, c = 'purple', label = 'Cluster 2')
59. plt.scatter(X[kmeans==2,0],X[kmeans==2,1], s=50, c = 'orange', label = 'Cluster 3')
60. plt.scatter(X[kmeans==3,0],X[kmeans==3,1], s=50, c = 'cyan', label = 'Cluster 4')

```

```

61. plt.scatter(X[kmeans==4,0],X[kmeans==4,1], s=50, c = 'pink', label = 'Cluster 5')
62. plt.scatter(X[kmeans==5,0],X[kmeans==5,1], s=50, c = 'grey', label = 'Cluster 6')
63. plt.scatter(estimador.cluster_centers_[:,0],estimador.cluster_centers_[:,1], s=100, c='black', label
   = 'Centróides')
64. plt.title('Clusters de Equipamentos e Centróides')
65. plt.ylabel('TEF médio')
66. plt.xlabel('TEM médio')
67. plt.legend()
68. plt.show()
69.
70. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS #####
71. #          PLOTAGEM DO GRÁFICO EQUIPAMENTOS          #
72. #####
73.
74. plt.scatter(Y[Z==0, 2], Y[Z==0, 3], s=50, c = 'red', label = eqp((Y[Z==0, 1]).any()))
75. plt.scatter(Y[Z==1, 2], Y[Z==1, 3], s=50, c = 'blue', label = eqp((Y[Z==1, 1]).any()))
76. plt.scatter(Y[Z==2, 2], Y[Z==2, 3], s=50, c = 'green', label= eqp((Y[Z==2, 1]).any()))
77. plt.scatter(Y[Z==3, 2], Y[Z==3, 3], s=50, c = 'yellow', label= eqp((Y[Z==3, 1]).any()))
78.
79. plt.title('Médias de Equipamentos')
80. plt.ylabel('TEF médio')
81. plt.xlabel('TEM médio')
82. plt.legend()
83. plt.show()
84.
85. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 1 #####
86. #          PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 1          #
87. #####
88.
89. for k in range(0,len(Y)):
90.     if kmeans[k] == 0:
91.         if Y[k,1] == 0:
92.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
93.         elif Y[k,1] == 1:
94.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
95.         elif Y[k,1] == 2:
96.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
97.         else:
98.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
99.
100.    else:
101.        plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')
102. plt.title('Cluster 1')
103. plt.ylabel('TEF médio')
104. plt.xlabel('TEM médio')
105. plt.legend()
106. plt.show()
107.
108. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 2 #####
109. #          PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 2          #
110. #####
111.
112. for k in range(0,len(Y)):
113.     if kmeans[k] == 1:
114.         if Y[k,1] == 0:
115.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
116.         elif Y[k,1] == 1:
117.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
118.         elif Y[k,1] == 2:
119.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
120.         else:
121.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+ ' '+str(Y[k,0]))
122.
123.    else:
124.        plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')

```

```

124.
125. plt.title('Cluster 2')
126. plt.ylabel('TEF médio')
127. plt.xlabel('TEM médio')
128. plt.legend()
129. plt.show()
130.
131.
132. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 3 #####
133. # PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 3 #
134. #####
135.
136.
137. for k in range(0,len(Y)):
138.     if kmeans[k] == 2:
139.         if Y[k,1] == 0:
140.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+' '+str(Y[k,0]))
141.         elif Y[k,1] == 1:
142.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+' '+str(Y[k,0]))
143.         elif Y[k,1] == 2:
144.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+' '+str(Y[k,0]))
145.         else:
146.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+' '+str(Y[k,0]))

147.     else:
148.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')
149.
150. plt.title('Cluster 3')
151. plt.ylabel('TEF médio')
152. plt.xlabel('TEM médio')
153. plt.legend()
154. plt.show()
155.
156.
157.
158. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 4 #####
159. # PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 4 #
160. #####
161.
162. for k in range(0,len(Y)):
163.     if kmeans[k] == 3:
164.         if Y[k,1] == 0:
165.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+' '+str(Y[k,0]))
166.         elif Y[k,1] == 1:
167.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+' '+str(Y[k,0]))
168.         elif Y[k,1] == 2:
169.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+' '+str(Y[k,0]))
170.         else:
171.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+' '+str(Y[k,0]))

172.     else:
173.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')
174.
175. plt.title('Cluster 4')
176. plt.ylabel('TEF médio')
177. plt.xlabel('TEM médio')
178. plt.legend()
179. plt.show()
180.
181.
182. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 5 #####
183. # PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 5 #
184. #####
185.
186. for k in range(0,len(Y)):
187.     if kmeans[k] == 4:

```

```

188.     if Y[k,1] == 0:
189.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+' '+str(Y[k,0]))
190.     elif Y[k,1] == 1:
191.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+' '+str(Y[k,0]))
192.     elif Y[k,1] == 2:
193.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+' '+str(Y[k,0]))
194.     else:
195.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+' '+str(Y[k,0]))

196. else:
197.     plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')

198.
199. plt.title('Cluster 5')
200. plt.ylabel('TEF médio')
201. plt.xlabel('TEM médio')
202. plt.legend()
203. plt.show()
204.
205.
206. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 6 #####
207. #          PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 6      #
208. ##### PLOTAGEM DO GRÁFICO EQUIPAMENTOS - cluster 6 #####
209.
210. for k in range(0,len(Y)):
211.     if kmeans[k] == 5:
212.         if Y[k,1] == 0:
213.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='red', label = eqp(Y[k,1])+' '+str(Y[k,0]))
214.         elif Y[k,1] == 1:
215.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='blue', label = eqp(Y[k,1])+' '+str(Y[k,0]))
216.         elif Y[k,1] == 2:
217.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='green', label = eqp(Y[k,1])+' '+str(Y[k,0]))
218.         else:
219.             plt.scatter(Y[k, 2], Y[k, 3], s=50, c='yellow', label = eqp(Y[k,1])+' '+str(Y[k,0]))

220.     else:
221.         plt.scatter(Y[k, 2], Y[k, 3], s=50, c='white')

222.
223. plt.title('Cluster 6')
224. plt.ylabel('TEF médio')
225. plt.xlabel('TEM médio')
226. plt.legend()
227. plt.show()

```