

Universidade Federal do Triângulo Mineiro

João Jacob de Ávila Neto

Análise de dados de sistemas de despacho

Uberaba - MG

2021

João Jacob de Ávila Neto

Análise de dados de sistemas de despacho

Trabalho de conclusão apresentado ao Programa de Mestrado Profissional em Inovação Tecnológica da Universidade Federal do Triângulo Mineiro, como requisito para obtenção do título de mestre.

Orientador: Prof. Dr. Mário Sérgio da Luz

Uberaba - MG

2021

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

A972a Ávila Neto, João Jacob de
 Análise de dados de sistemas de despacho / João Jacob de Ávila
 Neto. -- 2021.
 78 p. : il., graf., tab.

 Dissertação (Mestrado Profissional em Inovação Tecnológica) --
 Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2021
 Orientador: Prof. Dr. Mário Sérgio da Luz

 1. Análise de dados. 2. Despacho - Sistema. 3. CAN. I. Luz,
 Mário Sérgio da. II. Universidade Federal do Triângulo Mineiro.
 III. Título.

CDU 656.03:004.451.5

JOÃO JACOB DE ÁVILA NETO**ANÁLISE DE DADOS DE SISTEMAS DE DESPACHO**

Dissertação apresentada ao Programa de Pós-Graduação Profissional em Inovação Tecnológica, da Universidade Federal do Triângulo Mineiro como requisito parcial para obtenção do título de mestre.

Uberaba, 18 de fevereiro de 2021

Banca Examinadora:

Dr. Mário Sérgio da Luz – Orientador
Universidade Federal do Triângulo Mineiro

Dr. David Calhau Jorge
Universidade Federal do Triângulo Mineiro

Dr. Kléber Lopes Fontoura
Centro Federal de Educação Tecnológica



Documento assinado eletronicamente por **MARIO SERGIO DA LUZ, Professor do Magistério Superior**, em 18/02/2021, às 11:34, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#) e no art. 14 da [Resolução nº 34, de 28 de dezembro de 2017](#).



Documento assinado eletronicamente por **DAVID CALHAU JORGE, Professor do Magistério Superior**, em 18/02/2021, às 11:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#) e no art. 14 da [Resolução nº 34, de 28 de dezembro de 2017](#).

Documento assinado eletronicamente por **Kleber Lopes Fontoura, Usuário Externo**, em 26/03/2021, às 10:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de](#)



[2015](#) e no art. 14 da [Resolução nº 34, de 28 de dezembro de 2017](#).



A autenticidade deste documento pode ser conferida no site http://sei.uftm.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0473535** e o código CRC **C8B38AB6**.

Resumo

O uso do acompanhamento eletrônico nos equipamentos, utilizados no transporte de material em mineração, possibilita a obtenção de diversas informações sobre o processo, atuando como ferramenta de auxílio na tomada de decisão. As características da mineração acabam por torná-la cara e complexa, de forma que este tipo de solução age como aliada do setor. Diante dessa situação, propõe-se o desenvolvimento de pesquisas que possam contribuir para obtenção de estimativas e indicadores, permitindo uma melhor compreensão das variáveis envolvidas. A metodologia utilizada é a do estudo de caso, utilizando diferentes variáveis, produzidas por um veículo de transporte, utilizado no processo de lavra e obtidas pelo uso de um sistema de despacho eletrônico. Para isso, foi implementado utilizando diferentes abordagens de análise de dados, entre elas a utilização de clusters implementado com o algoritmo *k-means* e o de aprendizado de máquina em *KNN* (Algoritmo dos vizinhos mais próximos), as quais permitem obter uma compreensão melhor do processo. Também foi possível perceber correlações entre as variáveis utilizadas, assim com as possibilidades de realizar previsões de consumo de combustível, variável muito significativa dentro do processo de lavra, através de dados comuns em sistemas de despacho.

Palavras chave: Sistema de despacho, análise de dados, CAN

Abstract

The use of electronic monitoring in equipment, in the transport of material used in mining, makes it possible to obtain various information about the process, acting as an aid tool in decision making. The characteristics of mining end up making it expensive and complex, so that this type of solution acts as an ally of the sector. Given this situation, it is proposed to develop research that can contribute to the identification of indicators and indicators, allowing a better understanding of the variables involved. The methodology used is that of the case study, using different variables, produced by a transport vehicle, used in the mining process and basic by the use of an electronic dispatch system. For this, it was implemented using the different approaches of data analysis, among them the use of clusters in k-means and machine learning in KNN (Algorithm of the closest neighbors), as which allows to obtain an understanding of the best process. It was also possible to perceive the correlations between the variables used as well as the possibility of realizing fuel consumption, the variable very much within the washing process, through common data in dispatch systems.

Keywords: Dispatch system, Data analysis, CAN

LISTA DE FIGURAS

Figura 1 - Mina a céu aberto.....	12
Figura 2 – Exemplo de caminhão fora de estrada	13
Figura 3 - Formação de fila para carregamento.....	14
Figura 4 - Rede CAN dentro do formato OSI (ISSO 11898)	16
Figura 5 - Formato padrão CAN (2.0A).....	16
Figura 6 - Formato estendido da mensagem CAN (2.0B).....	16
Figura 7 – Representação de uma constelação de satélites	18
Figura 8 - Exemplo de gráfico Box-Plot	21
Figura 9 - Separação em grupos com características mais próximas	24
Figura 10 - Algoritmo K-means	25
Figura 11 - Verificação de linha vazias ou nulas.....	30
Figura 12 - Verificação de outlier	31
Figura 13 - Relação entre frequência e número de ocorrências	33
Figura 14 - Frequência de ocorrência de diferentes valores de velocidade.....	34
Figura 15 - Comportamento da temperatura.....	35
Figura 16 – Comportamento da variável RPM.....	36
Figura 17 - Comportamento da variável acelerador.....	36
Figura 18 - Comportamento da variável velocidade	38
Figura 19 - Comportamento de variáveis dentro de uma viagem	40
Figura 20 – Correlação (Spearman) de dados de uma viagem	41
Figura 21 - Correlação (spearman) entre todos os dados	42
Figura 22 - Parâmetros utilizados na criação do KMeans.....	43
Figura 23 - Cluster entre os dados de Altitude e Velocidade (Normalizado).....	44
Figura 24 - Correlação entre as variáveis do cluster 1	44
Figura 25 - Cluster entre os dados de Temperatura e Velocidade	45
Figura 26 - Correlações do cluster 3 entre as variáveis temperatura e velocidade ..	46
Figura 27 - Comportamento de variáveis dentro de uma viagem do veículo	47
Figura 28 - Correlação (Spearman) da variável taxa	48
Figura 29 - Parâmetros utilizados na predição KNN	48
Figura 30 - Parâmetros utilizados na predição KNN (Radios=1.0)	50
Figura 31 - Predição da variável taxa knn (radios=2)	50

SUMÁRIO

1. INTRODUÇÃO	9
2. REVISÃO BIBLIOGRÁFICA	11
2.1. MINERAÇÃO	11
2.2. ROTEAMENTO DE VEÍCULOS	12
2.3. SISTEMAS ELETRÔNICOS DE DESPACHO	14
2.4. PROTOCOLO CAN	15
2.4.1. Padrão J1939	17
2.5. GPS	18
2.6. ANÁLISE DE DADOS E TÉCNICAS EMPREGADAS	19
2.6.1. Gráfico box-plot	19
2.6.2. Coeficiente de correlação de <i>Spearman</i>	21
2.6.3. Aprendizado de máquina	22
2.6.3.1. K-means	23
2.6.3.2. KNN	26
2.7. FERRAMENTAS	27
2.7.1. Linguagem de programação	27
2.7.2. Bibliotecas	28
3. METODOLOGIA	28
4. RESULTADOS E DISCUSSÕES	33
4.1. ALOCAÇÃO INCORRETA DE EQUIPAMENTOS	33
4.2. AUMENTO DA DISPONIBILIDADE DA FROTA	34
4.2.1. Verificação do cumprimento de normas de segurança	37
4.2.2. Correlação entre as variáveis	40
4.2.3. Análise por clusters	43
4.2.4. Estimativa de consumo de combustível	46
CONCLUSÃO	51
TRABALHOS FUTUROS	52
REFERÊNCIAS	53
APÊNDICE – ANÁLISE POR CLUSTERS	56
1. ATITUDE E VELOCIDADE	57
2. TEMPERATURA E VELOCIDADE	61

3.	TEMPERATURA E RPM.....	64
4.	ALTITUDE E RPM	69
5.	VELOCIDADE E RPM	75
6.	ACELERADOR RPM.....	78

1. INTRODUÇÃO

A adoção de ferramentas, baseadas em análise de dados, para o auxílio nos processos gerenciais de tomada de decisão vem crescendo nos últimos anos, ainda assim a maior parte das empresas acaba por não aproveitar todo o potencial das suas informações, perdendo oportunidades de melhoria e de reduções de custos nos seus processos. Muitas dessas empresas apesar de serem ricas em dados, informações brutas produzidas por seus equipamentos e sensores sem nenhum tipo de tratamento ou análise, não conhecem realmente todas as particularidades dessas informações, e é por meio da análise desses dados que podem ser obtidos conhecimentos que realmente podem contribuir como ferramenta para o auxílio na tomada de decisão (FREITAS, 2000).

A análise de dados compreende o uso de diversas técnicas que vão desde os já conhecidos métodos clássicos como a análise de dados via estatística até soluções modernas baseadas em Inteligência Artificial (IA) e Machine Learning (ML). A adoção dessas ferramentas no auxílio de atividades de gerenciamento dentro de empresas cresce cada vez mais (DE AMO, 2004).

O setor de mineração acaba também por passar pelos problemas envolvendo a tomada de decisão, principalmente pelo fato que os empreendimentos desse setor geralmente envolvem grande dispêndio de recursos financeiros e materiais, por tratar-se de um processo caro e complexo. Ainda assim, o setor é estratégico e tem um papel importante para o país, pois é o segmento responsável por fornecer matéria-prima para diversos outros tipos de indústrias. Na economia brasileira, segundo o Instituto brasileiro de Mineração (IBRAM, 2019), no Relatório de atividades no período de julho de 2018 a junho 2019, a produção mineral brasileira representa 1,4% de todo o PIB Brasil, e segundo o IBGE emprega cerca de 195 mil trabalhadores diretamente. A importância do setor ainda é mais evidenciada quando é observada a sua relevância nas exportações brasileiras. O país exportou, em 2018, mais de 409 milhões de toneladas de bens minerais, e gerou divisas de US\$ 29,9 bilhões, o que representou 12,5% das exportações totais do Brasil, e 36,6% de saldo comercial.

A indústria mineradora avança na utilização de novas tecnologias, dentre elas da análises de dados, diante dessa situação, o desenvolvimento de técnicas e

pesquisas que podem contribuir para obtenção de estimativas e indicadores de produção que possibilitam planejamentos e previsões, a fim de estimar e comparar diferentes cenários, é de grande auxílio nas atividades do setor (RODRIGO; DE SOUZA; COSTA, 2019).

Na mineração uma parcela importante dos gastos para manutenção da operação vai para a alocação de máquinas, isso ocorre já que na maioria dos casos a área da extração e a área de beneficiamento, na qual o minério começa a ser transformado em produto, são separadas por distâncias consideráveis, sendo necessário então equipamentos para realização do transporte entre essas duas áreas. (ALMEIDA; DE CASTRO NEVES; DE FIGUEIREDO, [s.d.]).

Para auxiliar nesse processo são utilizados os sistemas eletrônicos de despacho, compostos por diversas ferramentas específicas para auxiliar a alocação de veículos e controle da produção. Apesar de serem voltados principalmente no desempenho da produção, estes sistemas fornecem e possibilitam a obtenção de uma grande quantidade de informação que muitas vezes não são aproveitadas pela empresa (ALMEIDA; DE CASTRO NEVES; DE FIGUEIREDO, [s.d.]).

Assim, este trabalho tem como objetivo geral avaliar os dados produzidos por sistemas eletrônicos de despacho, em específico o sistema de despacho Easymine, buscando técnicas e formas de visualização gráficas, que permitam a análise de dados, de forma a estudar as variáveis fornecidas por este sistema e então obter possíveis ganhos sob dois principais aspectos, representando os objetivos específicos:

- a) Aprendizado com os dados: contribuir para um melhor entendimento dos possíveis ganhos ao utilizar a análise de dados aplicada em sistemas de despacho.
- b) Aumento da produtividade: buscar diferentes análises e representações gráficas que auxiliem na realização de planejamentos, tornando mais eficiente a tomada de decisão no gerenciamento de veículos de transporte utilizados na mineração;

2. REVISÃO BIBLIOGRÁFICA

2.1. MINERAÇÃO

A mineração, juntamente com a agricultura é uma das atividades mais antigas realizadas pelo homem. Especificamente, no caso da mineração, acredita-se que seu início aconteceu tanto por aspectos estéticos (busca por materiais para criação de ornamentos como no caso do ouro e da prata) como pela exploração das características físicas dos materiais. Representa um importante segmento da cadeia produtiva já que é fornecedor de matéria-prima para outros tipos de indústrias existentes no país (CURI, 2014).

A mineração engloba diversas de atividades necessárias à obtenção de um produto mineral bruto como também processos como o de exploração, beneficiamento e processamento de recursos minerais que se encontram em estado natural, cujas as características fazem com que sua extração possa ser viável economicamente (MESQUITA; CARVALHO; OGANDO, 2016).

As minas produzem basicamente dois tipos de materiais: o minério e o estéril; o minério é composto por um ou mais minerais, que segundo Curi (2014) pode ser definido como um sólido homogêneo, cristalino, inorgânico de composição química definida, propriedades físicas características e ocorrência natural, esse material bruto é enviado para o tratamento e então transformando em produto final. O estéril, por sua vez, representa o material sem valor econômico que geralmente é retirado para dar acesso ao minério.

Como qualquer outro tema a mineração pode ser dívida com base em diversos aspectos. Pelo modo de escavação, as minas podem dividir-se em dois tipos principais, as minas subterrâneas e as a céu aberto. Subterrâneas são aquelas em que os minerais explorados se encontram muito abaixo da superfície, de forma que, muitas vezes, são criados túneis para exploração do minério. A mineração a céu aberto (Figura 1) já é caracterizada pela exploração em áreas em que o material de interesse está próximo à superfície.

Uma mina possui uma vasta gama de operações e uma larga área operacional. A atividade de lavra, pode ser entendida, como todas as atividades que coordenadas para o bom aproveitamento da jazida, vão desde a extração do minério até o início do beneficiamento. As características da mina determinam a escolha do método de lavra,

sendo o principal aspecto a viabilidade econômica. Dentro do processo de lavra o transporte é o processo responsável por levar o material da frente de lavra, área onde o material de interesse está sendo explorado, até as pilhas intermediárias, pilhas de estéril ou a britagem (CURI, 2014).

Figura 1 - Mina a céu aberto



Fonte: Do autor (2017)

2.2. ROTEAMENTO DE VEÍCULOS

As unidades de transporte são os equipamentos responsáveis por transportar o material, minério ou estéril, sendo o transporte feito do ponto de carregamento até o ponto de descarga, os principais representantes desse grupo são os caminhões. Esse podem ser subdivididos em caminhões convencionais e os caminhões fora de estrada (Figura 2), estes específicos para mineração podem transportar grande quantidade de material, no Brasil a maior parte deste tipo de veículo trabalha com cargas de 190 toneladas. (GERMANI, 2002)

Figura 2 – Exemplo de caminhão fora de estrada



Fonte: Do autor

No ciclo de extração mineral os equipamentos de transporte são alocados conforme as necessidades da produção na frente de lavra, sendo importante ressaltar que o transporte concentra grande parte do custo operacional da mina, com o combustível a principal variável dentro custo (GERMANI, 2002).

O processo começa com o posicionamento dos equipamentos de carga nas frentes de lavra, estes irão abastecer os caminhões, conforme a sua capacidade, assim que carregado os caminhões irão percorrer o caminho até o seu ponto de descarga, que pode ser uma britagem, pilha de estéril ou qualquer tipo de ponto de alimentação. Todo processo de gerenciamento da alocação é bastante complexo já que envolve um grande número de equipamentos além do grande dinamismo das minas, principalmente na exploração a céu aberto, que possui grandes modificações no terreno de forma bastante rápida (ALVARENGA, 1997).

O processo de exploração mineral exige em suas diferentes etapas uma grande diversidade de equipamentos e a integração destes com as operações da mina garante uma otimização da produção, a escolha das características dos equipamentos empregados diferem de uma empresa para outra dependendo de diversos fatores baseados na estratégia escolhida pela empresa gerando variações na alocação dos equipamentos como, por exemplo, na quantidade de equipamentos para transporte material e a sua capacidade de carga (GERMANI, 2002).

O processo de gerenciamento dos veículos (*dispatching* - despacho) pode ser feito de modo manual, baseado apenas na decisão dos controladores da mina, mas também pode ser realizado com auxílio de programas e equipamentos específicos

para auxiliar a alocação dos caminhões, maximizando a produção, diminuindo o tempo de fila e garantindo os índices desejados de teor do minério, esse tipo de solução é chamada de sistemas eletrônicos de despacho. A maior dificuldade na alocação dos caminhões é a redução do tempo perdido nas filas que se formam nos pontos de carga (Figura 3) e de báscula, o que acaba por diminuir a produção e o rendimento da exploração.

Figura 3 - Formação de fila para carregamento



Fonte: Autor (2017)

2.3. SISTEMAS ELETRÔNICOS DE DESPACHO

O controle das atividades de despacho dos veículos de carga de forma manual ainda é utilizado, mas a diminuição do custo das tecnologias e os ganhos proporcionados pelo uso de sistemas eletrônicos para acompanhar os equipamentos vêm contribuindo para a informatização desse processo. Os sistemas de despacho eletrônicos são formados por *Hardwares* (equipamentos eletrônicos) e *softwares* (programas de computador) voltados para o auxílio nas atividades de gerenciamento de veículos e da produção dentro das mineradoras (GERMANI, 2002)

Tendo em vista que o problema do roteamento dos veículos em mineração é bastante complexo e cresce proporcionalmente ao tamanho da mina e ao número de equipamentos empregados. Essas características acabam por tornar o uso de sistemas específicos um grande aliado para auxiliar nesse processo. Apesar do foco destes sistemas geralmente estar associado a produção da mina, de forma a garantir metas de produção, esses sistemas são compostos por diferentes tecnologias que

agregam diversas informações, indo muito além de apenas dados de produção (KOPPE, 2007).

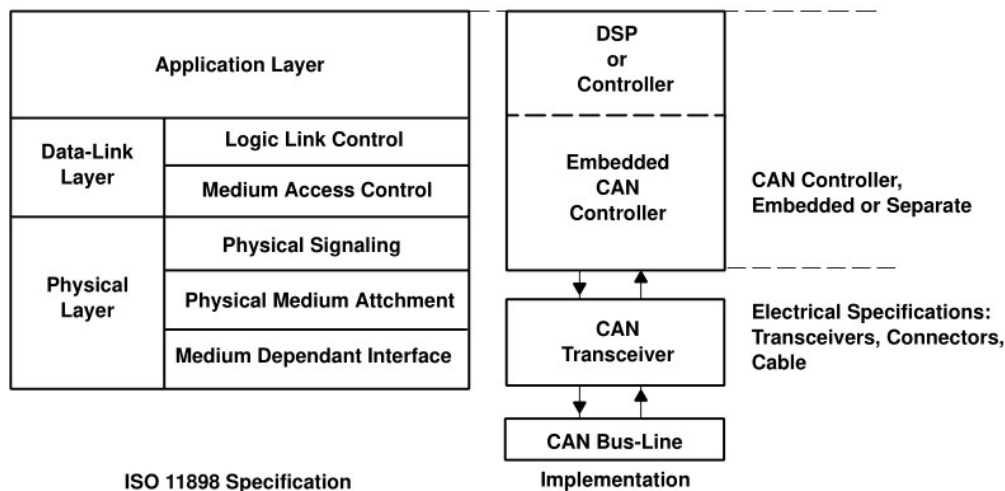
Para que os sistemas de despacho possam funcionar são utilizadas várias tecnologias para a aquisição de informações que possam contribuir no gerenciamento de todo o processo de extração mineral. As variadas informações podem ser obtidas pelos apontamentos dos motoristas no sistema, por exemplo, dados sobre condições da pista e climáticas, mas também e principalmente pela utilização de sensores que possibilitam a coleta de uma diferente gama de dados, que pode ser obtidos pela rede CAN (*Controller Area Network* - CAN). Outra fonte de informações é o uso do sistema de posicionamento global GPS (*global positioning system*) que, além de fornecerem a localização (latitude e longitude), também fornecem dados como a altitude, a velocidade do veículo, e o seu curso (KOPPE, 2007).

2.4. PROTOCOLO CAN

Entre as tecnologias que podem contribuir com dados para os sistemas de despacho uma das principais é o protocolo CAN (*Controller Area Network* - CAN) responsável por obter diversos valores de sensores presentes nos veículos. CAN é uma rede especificada pela Organização Internacional de Normalização (do inglês: *International Organization for Standardization* - ISO), originalmente desenvolvido para a indústria automotiva, depois ganhando popularidade em setores como marinha, medicina, manufatura e aeroespacial. (INSTRUMENTS, 2016)

O protocolo CAN Bus foi criado pela BOSCH e disponibilizado, em meados dos anos 80, como um sistema de transmissão serial síncrono de mensagens multi-mestre com taxa de transmissão máxima de 1 M bit por segundo (Mbps). Ao contrário de outras redes de transmissão, o CAN Bus não envia grandes blocos de dados, e não trabalha com um dispositivo mestre sob a supervisão do barramento central. Em uma rede CAN, as mensagens são curtas e transmitidas para toda a rede (*broadcast*) o que permite a consistência dos dados em cada nó do sistema. (INSTRUMENTS, 2016)

A norma ISO 11898 descreve como as informações são transmitidas entre dispositivos em uma rede de acordo com o Modelo OSI (acrônimo do inglês Open System Interconnection) definido em termos de camadas. A comunicação CAN Bus está dentro desse padrão, conforme a Figura 4 (INSTRUMENTS, 2016).

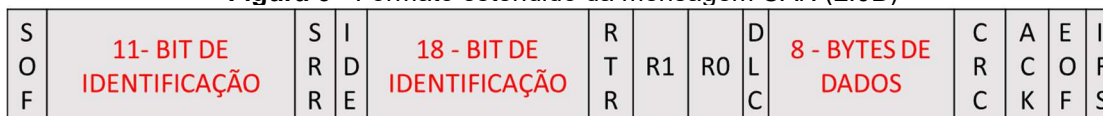
Figura 4 - Rede CAN dentro do formato OSI (ISSO 11898)

Fonte: Instruments (2016, p. 3)

Uma mensagem CAN Bus pode ser criada em dois padrões, o formato padrão conforme a Figura 5, e o formato estendido na Figura 6, variando, entre outras coisas, o número de dispositivos possíveis de utilização.

Figura 5 - Formato padrão CAN (2.0A)

Fonte: Adaptado de Instruments (2016, p. 4)

Figura 6 - Formato estendido da mensagem CAN (2.0B)

Fonte: Adaptado de Instruments (2016, p. 5)

Existem diversos padrões fundamentados no CAN, dentre os quais podem-se destacar, segundo GUIMARÃES e SARAIVA (2003):

- NMEA 2000: baseado no CAN 2.0B e utilizado em aplicações navais e aéreas.
- SAE J1939: baseado no CAN 2.0B e utilizado em aplicações automotivas, especialmente ônibus e caminhões.
- DIN 9684 – Ibs: baseado no CAN 2.0A e utilizado em aplicações agrícolas.
- ISO 11783: baseado no CAN 2.0A e também utilizado em aplicações agrícolas.

Uma rede CAN pode ser implementada utilizando 1, 2 ou 4 fios. Nas redes com 1 fio, esse é utilizado como meio de transmissão de dados. As redes com 2 fios trabalham com os sinais de dados CAN_H (CAN High) e CAN_L (CAN Low). No caso dos barramentos feitos com 4 fios além dos sinais de dados também são utilizados um fio com o VCC (alimentação) e outro com o GND (GUIMARÃES; SARAIVA, 2003).

A conexão com o meio físico, na rede CAN, é implementada através de um componente eletrônico responsável por intermediar a comunicação do frame CAN a outro protocolo, como por exemplo o RS232 (GUIMARÃES; SARAIVA, 2003).

2.4.1. Padrão J1939

Este trabalho foi baseado na obtenção de dados em um veículo que utilizava a especificação J1939. Esse é um padrão desenvolvido pela *Society of Automotive Engineers* (SAE), inicialmente destinado a comunicações de veículos pesados tornando posteriormente popular em outros setores, incluindo sistemas marinhos, instalações e fábricas ao controle. A especificação J1939 não é gratuita, podendo ser adquirida pela SAE. (DE ALMEIDA GUIMARÃES, 2001)

As características do protocolo J1939 incluem o uso do padrão extenso e de comunicação ponto a ponto (*peer-to-peer*). Nesta especificação, cada nó de comunicação CAN é chamado de Unidade de Controle Eletrônico (ECU). Cada ECU possui pelo menos um Nome e um Endereço de nó, mas também pode ter mais de um na mesma aplicação. A Tabela 1 mostra um resumo das características da especificação J1939 (DE ALMEIDA GUIMARÃES, 2001; GUIMARÃES; SARAIVA, 2003).

Tabela 1 - Características da especificação J1939

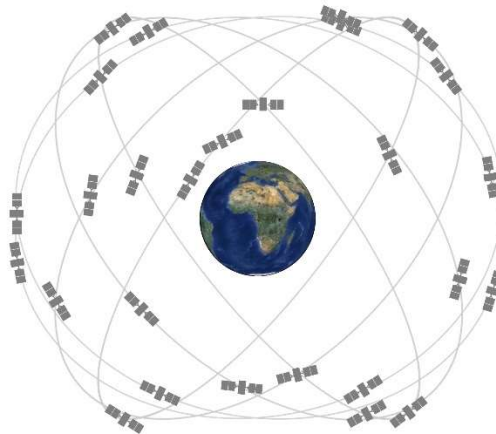
Item	Requerimento
Fabricação	2 fios
<i>Baud-rate</i>	250kb/s
Máximo de nós	30
Topologia	Linear
Comprimento máximo	40m
Resistor terminador	2 (120Ω)
Cabo	Blindado e par trançado

Fonte: Do autor

2.5. GPS

O posicionamento dos equipamentos dentro da mina é de extrema importância para o planejamento das operações, sendo indispensável a qualquer sistema de despacho eletrônico. O posicionamento pode ser feito de diversas formas sendo a mais comum pelo uso de posicionamento por satélite. O sistema de posicionamento global (*global positioning system* - GPS) é um sistema de navegação, feito por uma rede de satélites conforme observado na Figura 7, desenvolvido no início dos anos 70. O GPS foi criado pelo departamento de defesa dos EUA, inicialmente apenas para o uso militar, com o tempo começa a ser disponibilizado também para o uso civil (EL-RABBANY, 2002).

Figura 7 – Representação de uma constelação de satélites



Fonte: ("gps.gov", [s.d.])

O sistema GPS fornece posicionamento e outros dados, como por exemplo hora, altitude e velocidade, continuamente em qualquer posição da terra de forma passiva, enfim, os dados são apenas recebidos, isso acontece por razões de segurança. As informações são transmitidas via sinal de rádio sendo necessário teoricamente o mínimo de três satélites diferentes para obter uma posição correta. Recebido os dados do satélite o receptor converte as informações disponibilizando para uso geralmente em forma de comunicação serial RS-232 (EL-RABBANY, 2002).

A comunicação dos GPS é definida na especificação da National Marine Electronics Association (NMEA), sendo este o padrão mais utilizado. É através desta especificação que pode ser obtidas as variáveis de posição, velocidade e tempo. A ideia da NMEA é a formação "Frames" blocos constituídos por vários caracteres,

abaixo um exemplo de um frame, os blocos são iniciados pelos caracteres \$GP seguido do nome específico do frame, neste caso RMC, a vírgula delimita as diferentes variáveis disponíveis (ASSOCIATION, 2002).

\$GPRMC,123519,A,4807.038,N,01131.000,E,022.4,084.4,230394,003.1,W*6A

Existem diversos blocos que disponibilizam uma grande variedade de dados, especificamente para os sistemas de despacho os principais são o GGA e o RMC, já que estes possuem os principais dados utilizados por esses sistemas, como as variáveis de posição, latitude e longitude, assim como também a velocidade e a altitude do equipamento com o GPS, conforme podem ser visto na Tabela 2. Apesar da popularização do GPS atualmente existem outros conjuntos de redes de satélites, também chamadas de constelações, que também fornecem o posicionamento global, são exemplos as constelações GLONASS, BEIDU e a Galileu.

Tabela 2 - Parâmetros mais utilizados em sistemas de despacho

Bloco	Parâmetro	Posição específica no bloco
GGA	Latitude	3
	longitude	4
	altitude	9
	Diluição horizontal (HDOP)	8
	Data fixa	6
	Número de satélites usados	7
	Velocidade	7
RMC	Norte/Sul	4
	Leste/Oeste	6
	Curso	8
	Data	9

Fonte: Do autor (2019)

2.6. ANÁLISE DE DADOS E TÉCNICAS EMPREGADAS

2.6.1. Gráfico box-plot

Um gráfico box-plot (Figura 8) é geralmente usado para resumir várias informações de uma distribuição de uma amostra de dados. Para compreender este tipo de gráfico é necessário primeiramente conhecer a separação de dados por quartis, o primeiro quartil (Q1) representa o maior valor dos 25% menores dados de

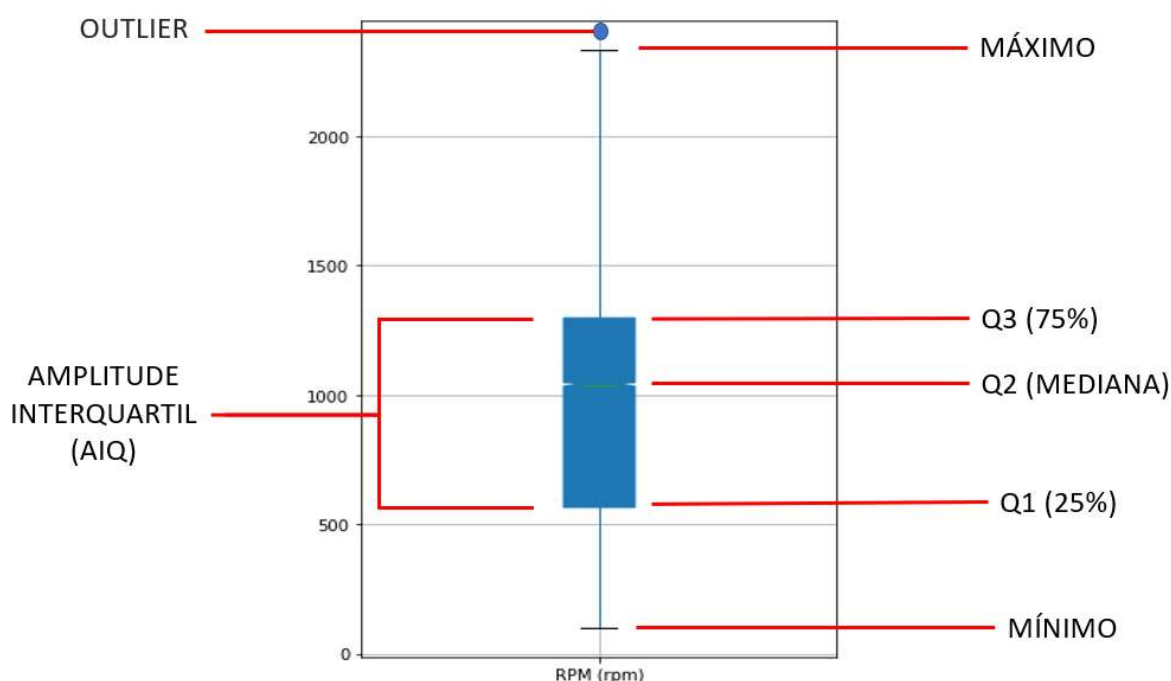
um conjunto de dados, o segundo quartil (Q2) separa os dados em 50%, sendo representado pelo valor da mediana, por último o terceiro quartil, o maior valor dos 75% menores valores (FARBER; LARSON, 2009).

Outra medida feita utilizando os quartis é a amplitude interquartil (AIQ) definida como a diferença entre o terceiro e o primeiro quartil, conforme pode ser vista na equação 1 abaixo:

$$AIQ = Q3 - Q1 \quad (1)$$

Além de fornecer o valor que 50% dos dados variam a AIQ também é utilizada para obter valores discrepantes, comumente chamados para este tipo de aplicação de outliers. Qualquer valor que seja $1,5 \cdot AIQ$ abaixo do Q1 ou acima Q3 é considerado um valor discrepante.

Em um box-plot linha inferior representa o menor valor em uma amostra de dados, desta linha a base da primeira caixa está o Q1. O Q2 representa a caixa, inferior, até a linha central, essa por sua vez possui o valor igual ao da mediana da amostra. Na sequência, a segunda caixa, tem-se o Q3, por fim, do final da segunda caixa até a linha superior o terceiro quartil. A linha superior corresponde pelo maior valor da amostra, valores fora dos limites, linha superior e inferior representam outliers. Começando no primeiro quartil (Q1) e terminando no terceiro (Q3), estão resumidos 50% dos dados centrais do conjunto. (FARBER e LARSON, 2009, p.86)

Figura 8 - Exemplo de gráfico Box-Plot

Fonte: Autor (2020)

2.6.2. Coeficiente de correlação de *Spearman*

O coeficiente de correlação de Pearson pode ser usado para resumir a força da relação linear entre duas amostras de dados. É calculado como a covariância das duas variáveis dividida pelo produto do desvio padrão de cada amostra de dados, como visto na equação 2 (BROWNLEE, 2018a).

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

A covariância e a correlação de Pearson são utilizados para identificar relações lineares entre as variáveis, como também assume que os dados estão em uma distribuição gaussiana. Nos casos em que isso não é verdade, ou não existe certeza, pode ser utilizado o coeficiente de correlação de *Spearman*, conforme a equação 3. Assim como no coeficiente de correlação de Pearson, o resultado também varia de -1 e 1, a diferença é que em vez de calcular o coeficiente usando covariância e desvios padrão das próprias amostras, são calculadas a partir da classificação relativa dos valores em cada amostra. Esse novo relacionamento também pode ser usado em

casos onde a relação é linear entre as variáveis, apenas é menos efetivo neste caso.(BROWNLEE, 2018a)

$$S_{X,Y} = \frac{con(rank(X),rank(Y))}{(\sigma(rank(X)) * \sigma(rank(Y)))} \quad (3)$$

A análise de dados dificilmente é feita apenas com uma variável, a partir de duas já podemos construir gráficos no plano cartesiano. Esses gráficos permitem identificação de tendências e de associação entre as variáveis, sobre as representações gráficas elas serão abordadas em sequência no texto (MAGALHÃES; DE LIMA, 2008).

2.6.3. Aprendizado de máquina

O volume de dados gerado dentro de empresas, instituições financeiras, instituições de pesquisas científicas entre diversas outras, cresce de forma acelerada. Dentro deste cenário a capacidade de produção de dados ultrapassa a capacidade de análise e extração de conhecimento sem ajuda de ferramentas computadorizadas. A Mineração de Dados (*Data Mining*), surge como uma ferramenta para a resolução deste problema, compõe-se de um campo de pesquisa amplo, utilizando para isso desde as ferramentas clássicas da estatística até modernas técnicas de aprendizado de máquina, todas reunidas para a obtenção de conhecimento a partir dados brutos. (Fontana e Naldi, 2009, p. 1 e seg.)

O aprendizado de máquina (*Machine Learning* - ML) é composto por diversas ferramentas com diversas inspirações em que cada uma oferece diferentes possibilidades e também possui diferentes limitações, cabendo ao responsável pela implementação analisar qual técnica responde melhor ao seu problema. Uma das categorias de ML, por exemplo, é a do aprendizado por agrupamentos, neste tipo de implementação são utilizados algoritmos com o propósito de agrupar dados segundo alguma regra que permita obter associação entre os dados.(RUSSEL; NORVIG, 2013)

Algoritmos de ML podem ser organizadas de acordo com diferentes critérios, um deles diz respeito ao paradigma sobre a forma como acontece o aprendizado, essa informação é importante já que a escolha da melhor ferramenta de ML, entre as várias

técnicas disponíveis, é dependente da forma que estão os dados. Dentro deste aspecto pode-se dividir em:

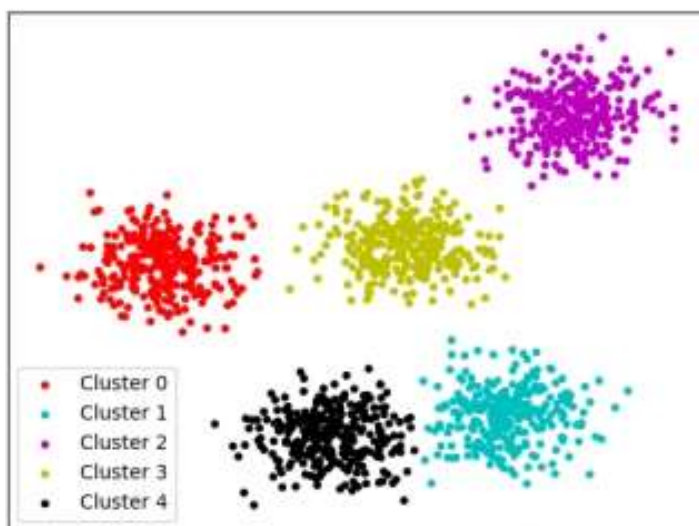
- a) **Aprendizado supervisionado:** O termo parte da ideia de simulação de um “supervisor externo” o qual conhece a saída para um determinado conjunto de entradas e a partir de um conjunto desses exemplos obtém a capacidade de prever a saída de novos exemplos.(FACELI et al., 2011)
- b) **Aprendizado não supervisionado:** neste paradigma os dados analisados não precisam estar rotulados, o software é exposto a grandes quantidades de informações sem saber do que se trata os dados, e durante o treinamento o programa encontra padrões de comportamentos entre as informações, terminando com sistema desenvolvendo suas próprias conclusões.(HARVEY, 2018)
- c) **Aprendizagem Semi-Supervisionada:** Como o próprio nome diz, é uma combinação de aprendizado supervisionado e não supervisionado, a base de dados utilizada no treinamento possui tanto dados rotulados com não rotulados. Neste sistema as informações já rotuladas são usadas para fazer novas inferências sobre os dados não rotulados, com objetivo de facilitar a obtenção dos melhores resultados.(HARVEY, 2018)
- d) **Aprendizagem Por Reforço:** este é uma implementação do aprendizado supervisionado ou semi-supervisionado, em que é utilizado sistemas de punições e recompensas. São testadas diversas combinações de entrada, as que obtiverem melhores resultados são recompensadas frente as combinações menos produtivas. Com o fim do treinamento o sistema acaba por encontrar padrões de funcionamento, e passa então ser possível a utilização dos parâmetros da melhor forma, mesmo em condições não treinadas.

2.6.3.1. K-means

O *Clustering* (Figura 9) é uma técnica de ML não supervisionado que funciona dividindo uma base de dados em grupos que apresenta características mais similares

entre si, utilizando para isso alguma medida que possa permitir revelar diferentes relacionamentos entre os dados.(SAYAD, [s.d.]

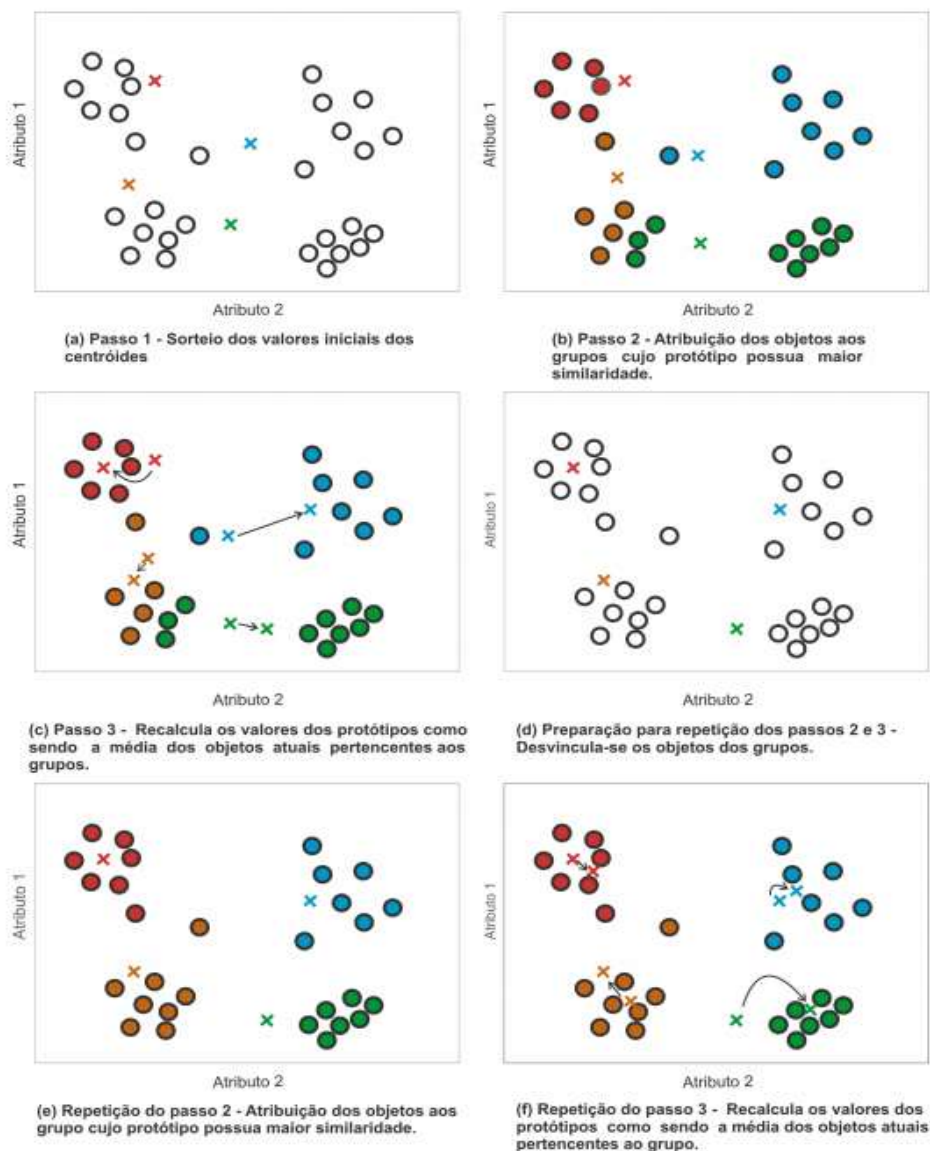
Figura 9 - Separação em grupos (agrupamentos) com características mais próximas, aqui representado pelas cores



Fonte: DataCoreSystem, [s.d.]

Um dos algoritmos utilizados em *cluster* é algoritmo *K-means* que pode ser descrito pelos passos apresentado na figura 10, de acordo com Fontana e Naldi (2009):

Figura 10 - Algoritmo K-means



Fonte: Fontana e Naldi, 2009, pag 22.

A determinação do número correto de núcleos de agrupamentos pode ser feita de forma experimental ou utilizando algum método, como por exemplo, o método *elbow*, que basicamente testa a variância dos dados em relação ao número de clusters, o valor ideal é obtido quando o aumento no número de clusters não representa um valor significativo de ganho. (NOGARE, 2015)

É importante ressaltar que as técnicas de agrupamento como o *K-means* não determinam o significado dos grupos formados, sendo necessária uma análise posterior feita de pôr um especialista sobre o tema do problema. Sendo uma aplicação

não supervisionada, o algoritmo também não possui qualquer informação sobre a base de dados da qual deve gerar partições. (Fontana e Naldi, 2009, p. 3)

2.6.3.2. KNN

Outra categoria de ML é a Classificação. Essa trata-se de uma técnica de aprendizado supervisionado, como objetivo de classificação por características comuns dentro da base de dados, utilizando para isso de classes anteriormente definidas. A classificação inicia com dados já rotulados, sendo estes inseridos em um algoritmo de classificação, este irá abstrair quais dados estão presentes em uma mesma classe e quais possuem características distantes. (Fontana e Naldi, 2009, p. 3)

Um dos algoritmos utilizados para classificação é o KNN (*K-Nearest Neighbor*). É um dos algoritmos de classificação mais simples, que trabalha basicamente classificando os objetos pela proximidade em um espaço de características, com base em exemplos passados ao programa na fase de treinamento. (“Nearest Neighbors”, [s.d.])

A proximidade é calculada pela distância entre a classe e o valor desconhecido. Existem várias formas diferentes de calcular essa distância. A mais simples é a distância euclidiana, equação 4:

$$d(x, y) = \sqrt{(x - y)^2} \quad (4)$$

Outra medida também utilizada é a distância Minkowsky, equação 5:

$$dw(p, q) = \left(\sum_{n=1}^n |x - y|^q \right)^{\frac{1}{q}} \quad (5)$$

Além da classificação o algoritmo KNN pode ser utilizado também para a obtenção da regressão. Neste caso pode ser usada em dados rotulados que são contínuos, em vez de variáveis discretas. O valor de um ponto obtido com base na média dos rótulos dos vizinhos mais próximos. Pela biblioteca utilizada neste trabalho, a scikit-learn, a regressão pode ser implementada de duas formas:

- a) `KNeighborsRegressor`: implementa o aprendizado com base em um valor inteiro, geralmente representado por k , em que este valor representa o número de vizinhos mais próximos de cada ponto; (“Nearest Neighbors”, [s.d.])
- b) `RadiusNeighborsRegressor`: A base nos vizinhos é obtida a partir de um raio fixo a partir do ponto de consulta. (“Nearest Neighbors”, [s.d.])

A regressão pode ser implementada dando pesos uniformes aos pontos da vizinhança, neste caso, cada ponto tem a mesma contribuição para a classificação de um ponto de consulta. A segunda forma de utiliza pesos proporcionais ao inverso da distância do ponto de consulta. Também existe a possibilidade desta implementação ser definida pelo usuário. (“Nearest Neighbors”, [s.d.])

2.7. FERRAMENTAS

Para a realização deste trabalho será utilizado algumas ferramentas que possibilitam a análise dos dados e a apresentação dos resultados, estas ferramentas são apresentadas neste capítulo.

2.7.1. Linguagem de programação

O Python foi a linguagem de programação escolhida para a utilização neste trabalho, devido a essa possuir uma sintaxe simples, muito associada a ganhos de produtividade, ter uma grande quantidade de bibliotecas com grande emprego destas em aplicações de análise de dados.

O Python foi criado em 1990 na Holanda, por Guido van Rossum, no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI). Sua criação foi feita a partir de outra linguagem, chamada ABC, e tinha originalmente o objetivo de atender usuário como físicos e engenheiros (BORGES, 2013, p. 14).

A linguagem é interpretada pela máquina virtual Python, o que torna os códigos desenvolvidos nela portáteis em diferentes sistemas operacionais. O Python é um software de código aberto, com licença *General Public License* (GPL), embora menos restritivo permitindo o uso para desenvolvimento de produtos proprietários. A especificação da linguagem é mantida pela Python Software Foundation 2 (PSF) (BORGES, 2013, p.14 e 15).

2.7.2. Bibliotecas

A vantagem mais importante na utilização da linguagem python é o grande número de bibliotecas disponíveis, as utilizadas neste trabalho foram:

- a) O NumPy: é uma biblioteca utilizada para aplicações científicas e numéricas que permite trabalhar com arranjos, vetores e matrizes de N dimensões, semelhante ao utilizado no Matlab, A estrutura de dados principal no NumPy é o ndarray, que é um nome abreviado para matriz N-dimensional (BROWNLEE, 2018b).
- b) Pandas: é uma biblioteca que fornece mecanismos para a análise de dados em python (MCKINNEY, 2014).
- c) Matplotlib: torna simples tarefas de plotagem de gráficos bidimensionais de vários tipos, pode ser utilizado tanto interativa quanto embutido em um código python, também é utilizado tanto em aplicações web com integrada a interfaces gráficas de vários tipos (HUNTER et al., 2018).
- d) Statsmodels: fornece classes e funções para a estimativa de diferentes modelos estatísticos, como também para a realização de testes estatísticos e exploração de dados estatísticos (PERKTOLD; SEABOLD; TAYLOR, [s.d.]).

3. METODOLOGIA

A metodologia desse trabalho divide-se em três etapas: a obtenção dos dados, a definição dos parâmetros e a análise dos dados. A primeira etapa iniciou-se com a retiradas dos dados, fornecidos pelo sistema de despacho EasyMine, solução da empresa Instale Tecnologia. A amostra foi feita de um veículo modelo Mercedes Actros 4844, no período de 01/03/2018 a 06/03/2018. O arquivo com os dados foi retirado no formato texto/CSV em um total de 25805 linhas, do sistema supervisor do sistema, contendo informações de diversas fontes, principalmente CAN e GPS.

Colunas vazias e com informações referentes a transmissão dos dados entre o computador embarcado e o servidor, sem utilidade para o trabalho, foram eliminadas, restando as colunas conforme a Tabela 3.

Tabela 3 - Colunas disponíveis

Variável	Fonte
Data	Computador embarcado
Hora	Computador embarcado
Latitude	GPS
Longitude	GPS
Altitude (m)	GPS
Ocorrência	Computador embarcado
KmHora	GPS
RPM	CAN
Acelerador	CAN
Odometro	CAN
Temperatura	CAN
TaxaCombustivel	CAN
Combustivel	CAN
Velocidade	CAN
Embreagem	CAN
Freio	CAN
HorimetroT	CAN

Fonte: Autor (2019)

Na segunda etapa, da definição dos parâmetros, foi verificado a consistência dos dados. Os parâmetros embreagem e freio foram eliminados, já que o veículo utilizado para retirada dos dados não fornecia ao sistema os valores corretos das duas variáveis. Também foram descartadas as variáveis de latitude e longitude, por opção do autor não serão utilizadas neste trabalho. Foi descartada a variável horímetro, já que seu valor pode ser obtido com a variável odômetro. Por último, a variável TaxaCombustivel também foi eliminada, já que representa o valor fixo de 0,5 litros, referente ao consumo do veículo analisado. As variáveis utilizadas estão presentes na Tabela 5, junto com seus respectivos cinco primeiros dados.

Tabela 4 – Cinco primeiras linhas dos dados utilizados

	Data	Hora	Altitude	Ocorrência	Km/Hora	RPM	Acelerador	Odômetro
0	01/03/2018	08:42:10	1096.6	deslocamento	0	556.88	13.2	169154315
1	01/03/2018	08:42:20	1096.5	deslocamento	1	573.38	2.0	169154315
2	01/03/2018	08:42:30	1096.3	deslocamento	2	652.13	4.8	169154320
3	01/03/2018	08:42:40	1095.8	deslocamento	2	569.25	0.0	169154325

4 01/03/2018 08:42:50 1095.4 deslocamento 2 822.88 14.8 169154330
Continua

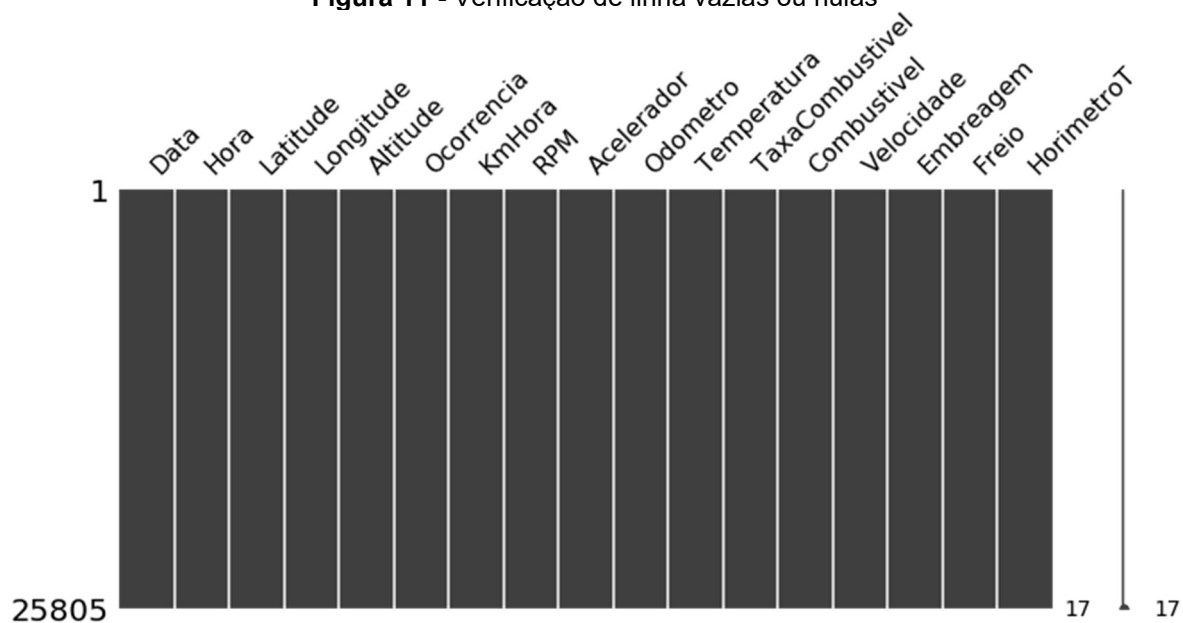
Temperatura(°C)	Combustível	Velocidade
29	106312.5	0.0
29	106312.5	0.0
29	106312.5	0.0
30	106312.5	2.0
30	106312.5	2.0

Fonte: Do Autor (2019)

A variável ocorrência representa o apontamento da atividade que o motorista está realizando no momento. RPM é o número de rotações por minuto do motor do veículo. O acelerador corresponde a porcentagem, de zero a cem, que o motorista está pisando no acelerador. Odômetro é a distância total percorrida pelo veículo desde a sua saída da fábrica. A variável combustível representa o valor consumido pelo veículo, é fornecida na forma de um valor ponto flutuante, um número com uma casa decimal. Essa casa decimal varia nos valores de 0 e 5, essa mudança acontece a cada consumo de 0,5 litro de combustível pelo veículo

A verificação de linhas vazias dentro da base de dados foi feita utilizando a biblioteca *missingno*, a Figura 11 mostra o total de dezessete variáveis com um total de 25805 linhas de dados, nenhuma apresentou valor nulo.

Figura 11 - Verificação de linha vazias ou nulas



Fonte: Do autor (2020)

A contagem também foi feita pelo número de linhas por coluna, de forma separada, apresentado o resultado na Tabela 5 também confirmando os resultados anteriores.

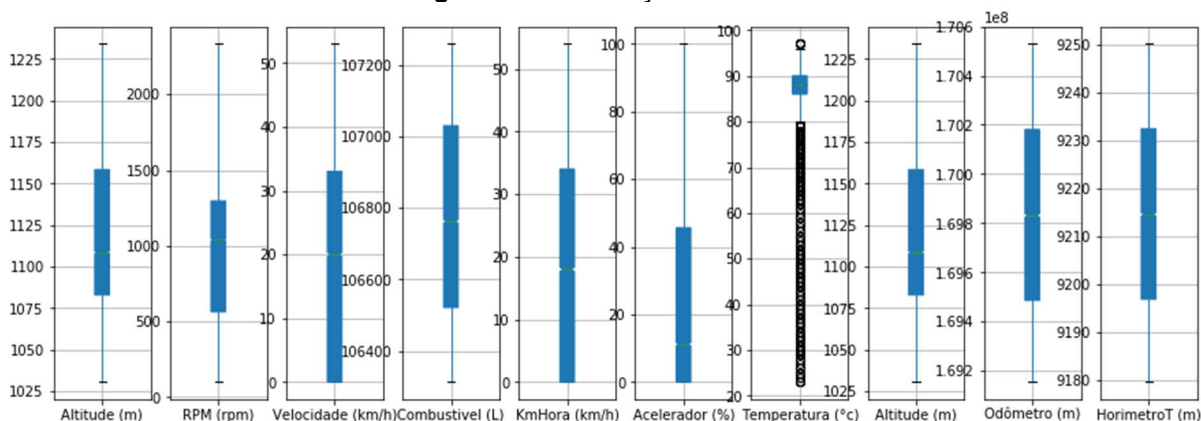
Tabela 5 – Número de linha em relação a cada variável

Variável	Número de linhas
Data	25805
Hora	25805
Latitude	25805
Longitude	25805
Altitude	25805
Ocorrência	25805
KmHora	25805
RPM	25805
Acelerador	25805
Odometro	25805
Temperatura	25805
TaxaCombustivel	25805
Combustível	25805
Velocidade	25805
Embreagem	25805
Freio	25805
HorimetroT	25805

Fonte: Do autor (2020)

Para verificar a existência de alguma incongruência nos dados foi utilizado o gráfico box-plot (Figura 12), para cada uma das variáveis. O valor de RPM apresentou um outlier, esse valor está fora do máximo possível para o veículo em análise, como ocorreu apenas uma vez, foi então assumido aqui que se trata de um erro e a opção feita foi pela sua retirada, no caso foi eliminado todos os dados da linha 21075.

Figura 12 - Verificação de outlier



Fonte: Do autor (2020)

O segundo valor a apresentar *outlier* foi a variável temperatura isso se deve ao fato dessa subir gradativamente quando o equipamento é ligado até ocorrer a estabilização. Apesar desses valores estarem fora do interquartil, não serão eliminados, como aconteceu na variável RPM, já que são realmente parte do processo

A etapa 3 trata-se da análise das variáveis, da decomposição e agregação destas quando necessário, afim de encontrar novas relações. Esta etapa tem com subdivisões:

- a) Análise de cada uma das variáveis de forma isolada, determinando os valores de máximo, mínimo, valor médio e quartis utilizando para isso o gráfico box-plot.
- b) Análise das correlações entre as variáveis,
- c) Uso de algoritmos de Clusters para tentativa de obter correlações antes não identificadas,
- d) Desenvolver metodologia para a análise do consumo de combustível,
- e) Analisar o consumo de combustível junto as outras variáveis,
- f) Desenvolver aplicações que permitem a predição do consumo em certas condições.

Foi utilizado a linguagem python para a criação do software, junto com algumas bibliotecas, as versões destas ferramentas utilizadas foram:

- a) scipy: 1.2.1;
- b) numpy: 1.15.2;
- c) matplotlib: 3.0.0;
- d) pandas: 0.23.4;
- e) sklearn: 0.21.3;
- f) statsmodels: 0.10.1.

Todos os experimentos foram executados em um mesmo computador suas principais características são: processador AMD Ryzen 2600, memória Ram de 8,0Gb, sistema operacional Windows 10 64 bits.

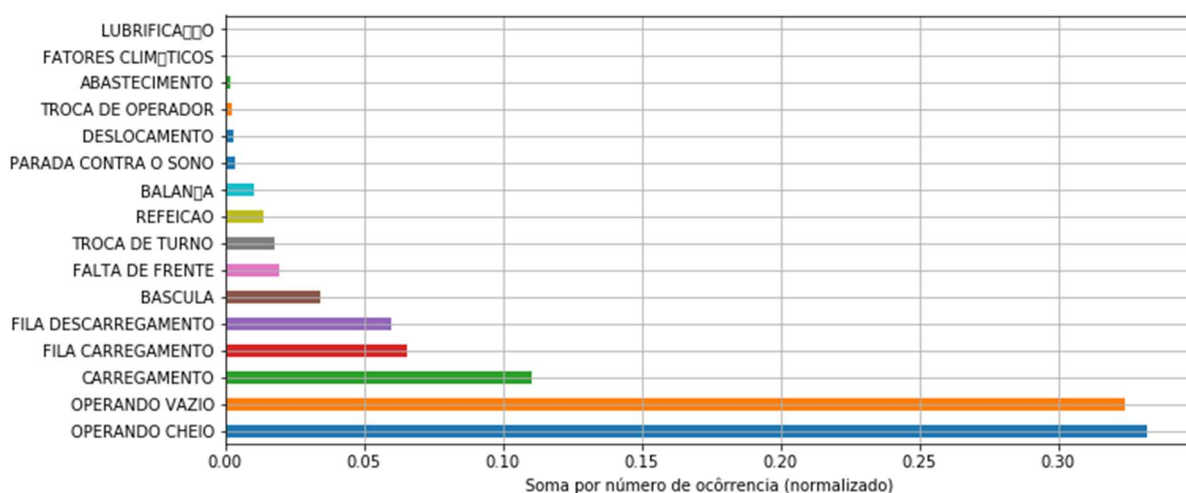
4. RESULTADOS E DISCUSSÕES

Retomando os objetivos de obter contribuições através da análise de dados vindos de sistemas de despacho, a partir de agora esse processo começa iniciando pela análise das variáveis de forma isolada e posteriormente a análise dos seus diferentes relacionamentos.

4.1. ALOCAÇÃO INCORRETA DE EQUIPAMENTOS

Assim como descrito anteriormente os gastos com combustível representam um valor considerável no custo da operação de uma mina, evitar consumos desnecessários é uma ferramenta que pode contribuir em um melhor desempenho financeiro. Dentro desse contexto a formação de filas no processo de mineração pode ser um indicativo de alocamento incorreto de veículos de carga. A Figura 13 mostra informações de apontamento provenientes de um veículo em relação ao número de ocorrências de cada desses apontamentos, as variáveis como “fila de carregamento”, “fila de descarregamento” e “carregamento” com valores altos são indicativos de alocamento incorreto do veículo na frente de trabalho. Isso acontece já que os veículos em filas estão consumindo combustível de forma improdutiva.

Figura 13 - Relação entre frequência e número de ocorrências apresentado de forma normalizada

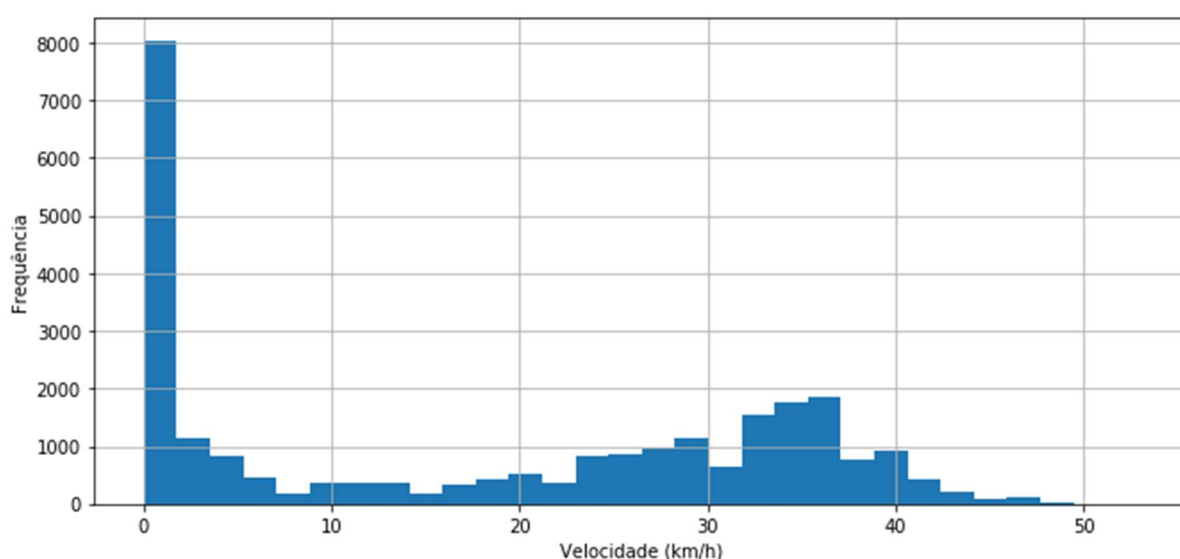


Fonte: Do autor (2020)

É importante ressaltar que as ocorrências apresentadas anteriormente são imputadas de forma manual pelo motorista do equipamento de carga, sendo assim

pode existir desvios em relação ao real tempo em que o equipamento passou em uma determinada ocorrência, já que nem sempre o motorista realiza os apontamentos no tempo correto, para conferir o tempo gasto em formação de filas pode ser utilizado o histograma da velocidade (Figura 14) que apresenta a frequência de valores de velocidade que o caminhão realizou dentro do período analisado, aqui os valores são obtidos via rede CAN do equipamento, nele é possível perceber novamente que o veículo passou tempo considerável parado.

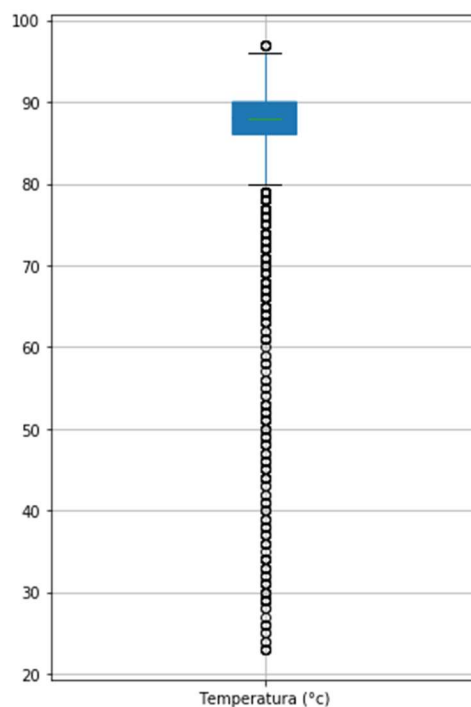
Figura 14 - Frequência de ocorrência de diferentes valores de velocidade



Fonte: Do autor (2020)

4.2. AUMENTO DA DISPONIBILIDADE DA FROTA

O controle de algumas variáveis pode contribuir no processo de tomada de decisão. Este é o caso das variáveis temperatura e RPM. Ambas podem ser um indicativo de problemas do veículo, sua análise permite antever possíveis manutenções corretivas, auxiliando na prevenção de ocorrências deste tipo. Apesar de não ser encontrado na literatura os valores ótimos de operação, destas variáveis, para o veículo aqui utilizado, a verificação do comportamento das mesmas pode possibilitar a identificação de anomalias. Como pode ser visto nas figuras 15, onde é possível observar o comportamento da temperatura do motor do veículo analisado, a tabela 6 descreve de forma mais detalhada os dados desta variável.

Figura 15 - Comportamento da temperatura

Fonte: Do autor (2020)

Pela figura 15 e a tabela 6 a seguir, é possível observar que em 50% do tempo a temperatura está com certeza entre 86°C e 90°C, respectivamente o quartil 1 e 2, e praticamente 100% entre 80°C e 95°C, com o valor máximo de 97°C.

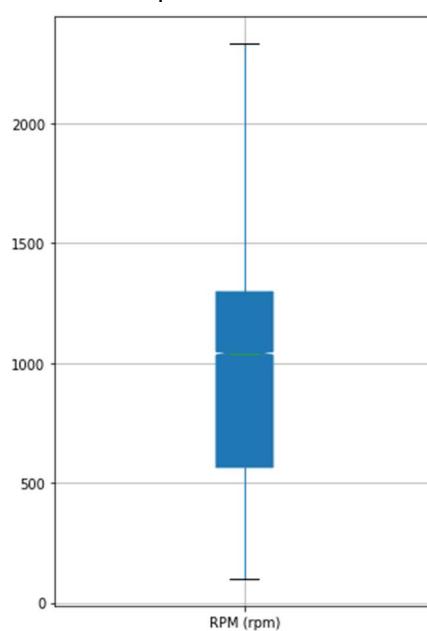
Tabela 6 - Comportamento da temperatura

Medição	Valor
Número de linhas	25804.00
Média	87.396528
Desvio padrão	5.222591
Mínimo valor	23.000000
Q1 (25%)	86.000000
Q2 (50%)	88.000000
Q3 (75%)	90.000000
Máximo valor	97.000000

Fonte: Do autor (2020)

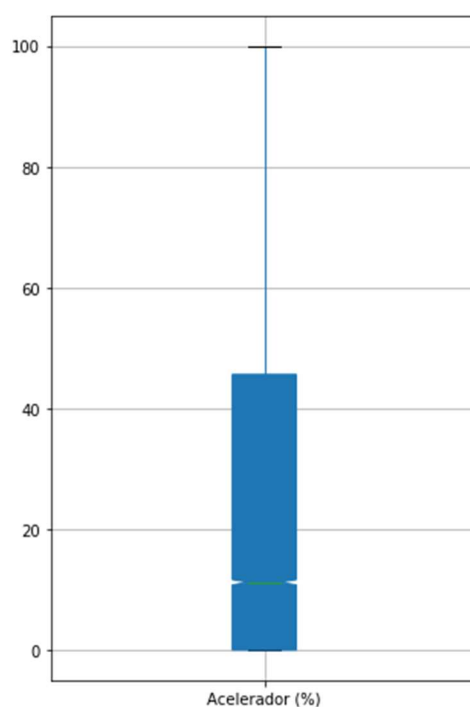
Outras duas variáveis que a sua análise pode auxiliar a identificação de anomalias são o RPM, Figura 16 e a aceleração, Figura 17, a última estando mais ligado ao comportamento do motorista.

Figura 16 – Comportamento da variável RPM



Fonte: Do autor (2020)

Figura 17 - Comportamento da variável acelerador



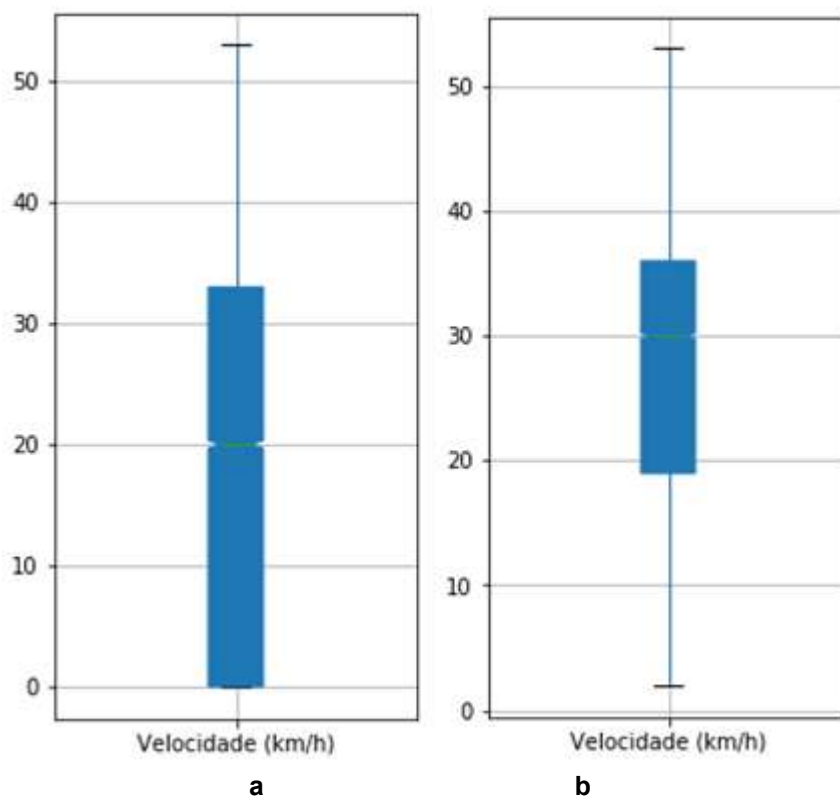
Fonte: Do autor (2020)

. Na figura 16 é possível perceber que a distribuição da variável RPM, em 75% do tempo abaixo de 1500 rotações por minuto e um máximo próximo a 3000. A figura 17 apresenta a distribuição da variável acelerador, esta por sua vez representa o quanto, em porcentagem, o acelerador do veículo está sendo acionado pelo motorista, 75% do tempo abaixo de 45%.

4.2.1. Verificação do cumprimento de normas de segurança

Como qualquer ambiente industrial as minas possuem regras e normas que regem o trabalho neste ambiente. A velocidade é uma delas, os veículos em minas percorrem distancia consideráveis em terrenos acidentados e com grande quantidade de carga, neste contexto uma alta velocidade pode significar um aumento significativo de risco de acidentes.

Analisando de forma global, com os dados de toda a amostra, a velocidade média do veículo foi de 18km/h, sendo 75% dos dados com valores abaixo de 33km/h conforme pode ser visto no gráfico na Figura 18a e com mais detalhes na tabela 7. Conforme visto anteriormente o veículo passa uma quantidade de tempo significativa parado. Analisando apenas os dados de velocidade diferente de zero obteve-se como resultado o gráfico da Figura 18b e a tabela 8.

Figura 18 - Comportamento da variável velocidade

Fonte: Do autor (2020)

Tabela 7 - Comportamento da velocidade, incluído as ocorrências em que o veículo permaneceu parado

Medição	Valor
Número de linhas	25804.00
Média	18.013060
Desvio padrão	15.678862
Mínimo valor	0.000000
Q1 (25%)	0.000000
Q2 (50%)	20.000000
Q3 (75%)	33.000000
Máximo valor	53.000000

Fonte: Do autor (2020)

Tabela 8 - Comportamento da velocidade sem o tempo parado

Medição	Valor
Número de linhas	17772.000
Média	26.154006
Desvio padrão	12.000135
Mínimo valor	2.000000
Q1 (25%)	19.000000
Q2 (50%)	30.000000
Q3 (75%)	36.000000
Máximo valor	53.000000

Fonte: Do autor (2020)

Ainda analisado o comportamento da velocidade do caminhão, esse teve como valor máximo 53 km/h de velocidade, apesar que dentro da mineração o máximo recomendado gira em torno de 30 a 40 km/h. O valor máximo de velocidade aconteceu dentro do apontamento de operando vazio, conforme a tabela 9. Quando é observado o segundo maior valor, esse de 50km/h, e é verificado o apontamento neste momento, o resultado é operando cheio, o que significa que este equipamento estava carregado, esse fato torna mais importante esta análise já que veículo ultrapassou muito o limite de velocidade mesmo nessa condição. O não cumprimento da velocidade permitida ocorreu em mais 822 dados, do total utilizado nesta análise.

Tabela 9 – Valor médio e máximo da variável velocidade em cada ocorrência

Ocorrência	Média	Valor máximo
Abastecimento	1.304348	22.0
Balança	24.360902	46.0
Báscula	0.136674	8.0
Carregamento	1.221871	28.0
Deslocamento	18.612500	39.0
Falta de frente	9.962451	41.0
Fatores climáticos	5.470588	30.0
Fila carregamento	1.195511	37.0
Fila descarregamento	2.598058	47.0
Lubrificação	0.000000	0.0
Operando cheio	24.146990	50.0
Operando vazio	26.556127	53.0
Parada contra o sono	5.535354	37.0
Refeição	13.336111	42.0
Troca de operador	20.791045	47.0
Troca de turno	13.379085	43.0

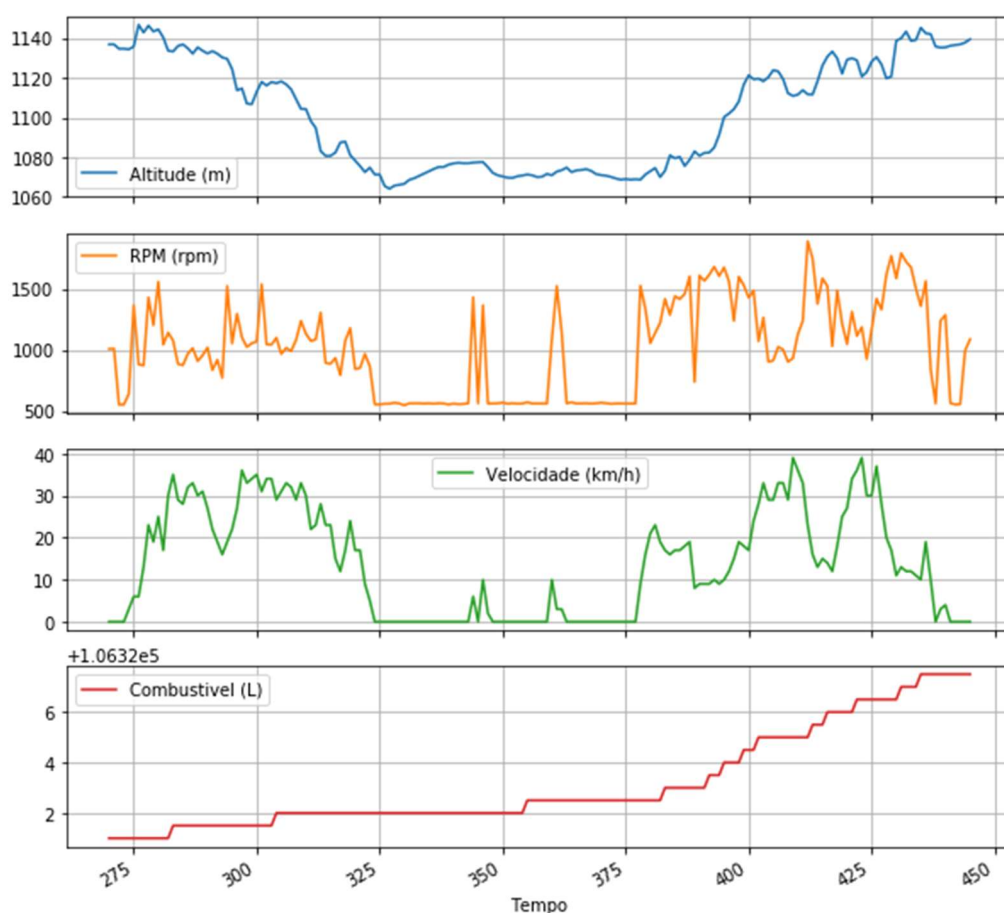
Fonte: Do autor (2020)

4.2.2. Correlação entre as variáveis

Até aqui a análise foi feita utilizando apenas as variáveis fornecidas pelo sistema de despacho de forma isolada, sem relacionar a influência entre as mesmas. A figura 19 mostra o comportamento de quatro variáveis em uma viagem do veículo analisado, do carregamento até uma báscula, na data 01/03 entre os horários de 12:48 e 13:17, é possível perceber graficamente algumas correspondências entre as variáveis altitude, velocidade e a aceleração.

O combustível obtido via CAN é fornecido de uma forma característica, de forma fica inviável a utilização da variável em análise de correlações com as outras variáveis, o comportamento dessa variável pode ser visto na figura 19.

Figura 19 - Comportamento de variáveis dentro de uma viagem do veículo de carga



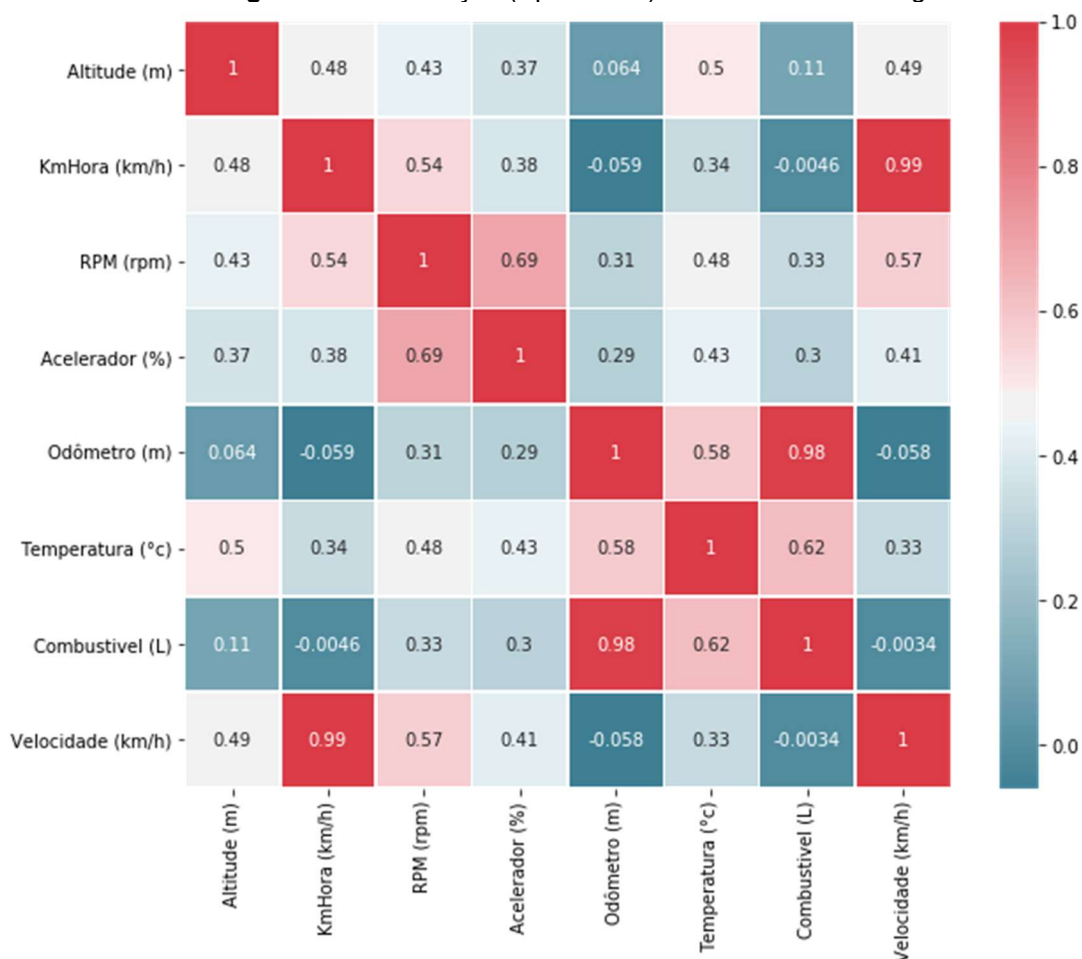
Fonte: Do autor (2020)

A variável combustível é representada por um número ponto flutuante com uma casa decimal, essa casa decimal assume apenas os valores de 0 e 5, e sua mudança

acontece quando o veículo em questão consome 0,5 litros de combustível, assim que isso acontece é incrementando o valor de 1 na parte inteira do número que representa o combustível, de forma está variável sempre cresça impedindo a utilização dela de forma direta em comparação com as demais.

Para transformar essa percepção visual de correspondência entre as variáveis, obtida na figura anterior, em “informação” foi utilizado a correlação de spearmam resumindo os resultados na figura 20, para os dados delimitados anteriormente, e na figura 21 para toda a base de dados utilizada neste trabalho. A intercessão entre duas variáveis no gráfico mostra o valor de correlação proporcional entre elas, pelo valor e pela cor de acordo com a escala ao lado da figura.

Figura 20 – Correlação (Spearmam) de dados de uma viagem



Fonte: Do autor (2020)

Figura 21 - Correlação (spearman) entre todos os dados

Fonte: Do autor (2020)

As principais correlações que aparecem nos dados são entre as variáveis:

- Velocidade e rpm;
- Horímetro, odômetro e combustível (estes três valores crescem continuamente);
- Acelerador e RPM;
- KmHora e velocidade (São a mesma variável, uma é obtida via CAN, inclusive mais precisa, enquanto a outra é fornecida pelo GPS);

4.2.3. Análise por clusters

Assim como discutido na revisão bibliográfica a utilização de cluster pode trazer a descoberta de novos relacionamentos entre as variáveis além das correlações anteriores obtidas. Na implementação utilizando os clusters, os parâmetros da biblioteca *sklearn* empregados foram utilizados conforme a Figura 22, já na determinação do número correto de núcleos de agrupamentos foi utilizado o método *elbow* para cada gráfico criado, os gráficos e dados gerados foram organizados no anexo 1 deste trabalho, apresentado aqui os principais resultados.

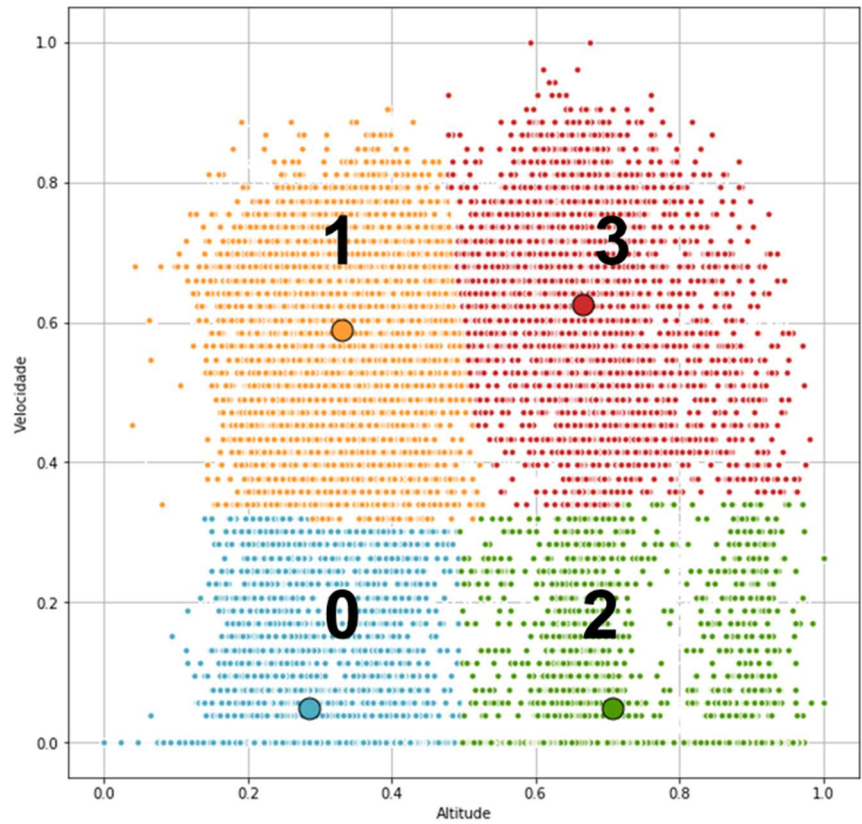
Figura 22 - Parâmetros utilizados na criação do KMeans

```
Kmean = KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,  
n_clusters=3, n_init=10, n_jobs=1, precompute_distances='auto',  
random_state=None, tol=0.0001, verbose=0)
```

Fonte: Do autor (2019)

O resultado a implementação do cluster entre as variáveis altitude e velocidade pode ser visto na Figura 23. Na sequência, Figura 24, o gráfico das correlações do cluster 1. Este cluster agrupa pontos de velocidade alta e altitude baixa, quando comparado ao gráfico com todos os pontos (figura 21), neste existe uma diminuição significativa da relação entre velocidade e rpm.

Figura 23 - Cluster entre os dados de Altitude e Velocidade (Normalizado)



Fonte: Do autor.

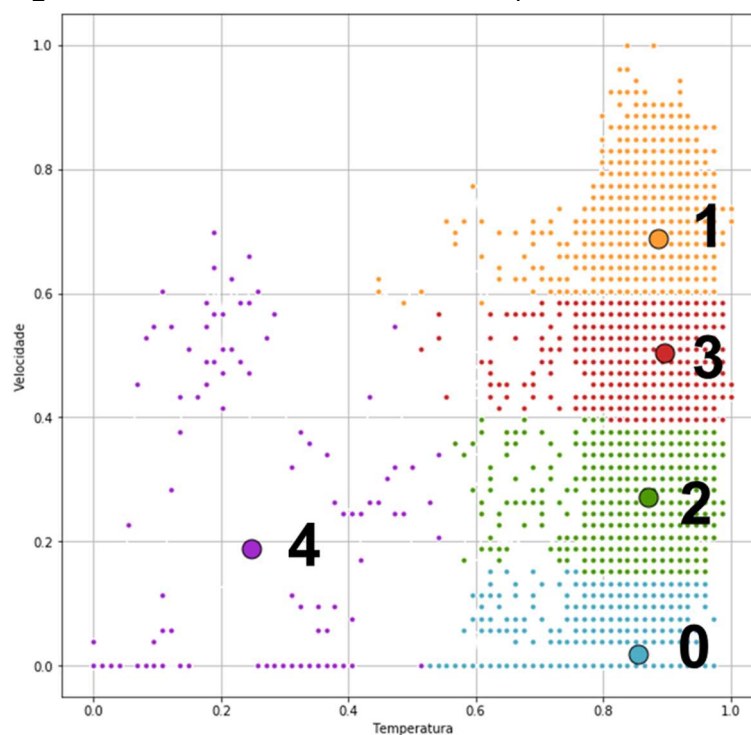
Figura 24 - Correlação entre as variáveis do cluster 1



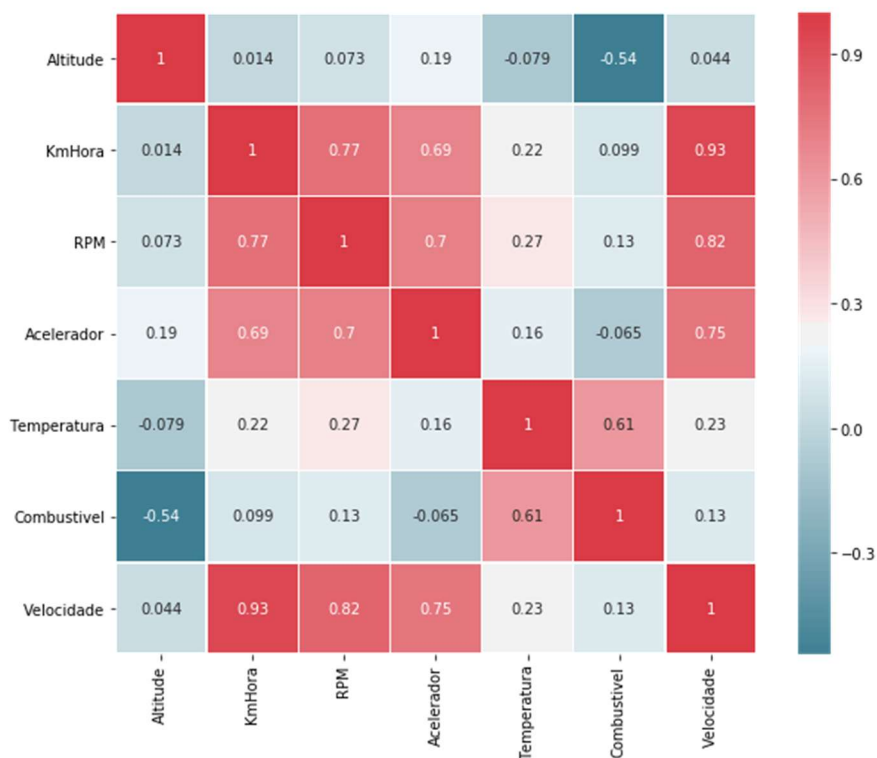
Fonte: Do autor

Na Figura 25 é apresentado o resultado a implementação das variáveis temperatura e velocidade. Na sequência, Figura 26, o gráfico das correlações referente ao cluster 4. Este cluster agrupa pontos de temperatura e velocidade visivelmente dispersos, o que explica as diferentes correlações, quando comparado aos outros clusters, da relação temperatura e velocidade, como também em relação a correlação com todos os dados. Este cluster poderia ser utilizado para eliminar dados pouco relevantes a análise total, sendo composto por 138 apontamentos do total.

Figura 25 - Cluster entre os dados de Temperatura e Velocidade



Fonte: Do autor (2020)

Figura 26 - Correlações do cluster 3 entre as variáveis temperatura e velocidade

Fonte: Do autor

4.2.4. Estimativa de consumo de combustível

Como anteriormente descrito a variável combustível, na forma que é obtida via CAN não é possível a utilização. Então, para a previsão do consumo de combustível, foi criado a variável taxa. Os dados foram agrupados pelos valores de combustível, então foi contado o número de dados dentro de cada um dos valores de combustível, conforme a tabela 10. Como cada variação no valor de combustível acontece com o consumo de 0,5 litro de combustível, esse valor foi então dividido pelo número em cada agrupamento, sendo assim possível visualizar onde o consumo de combustível aconteceu de maneira mais rápida. O valor obtido foi então associado aos dados globais.

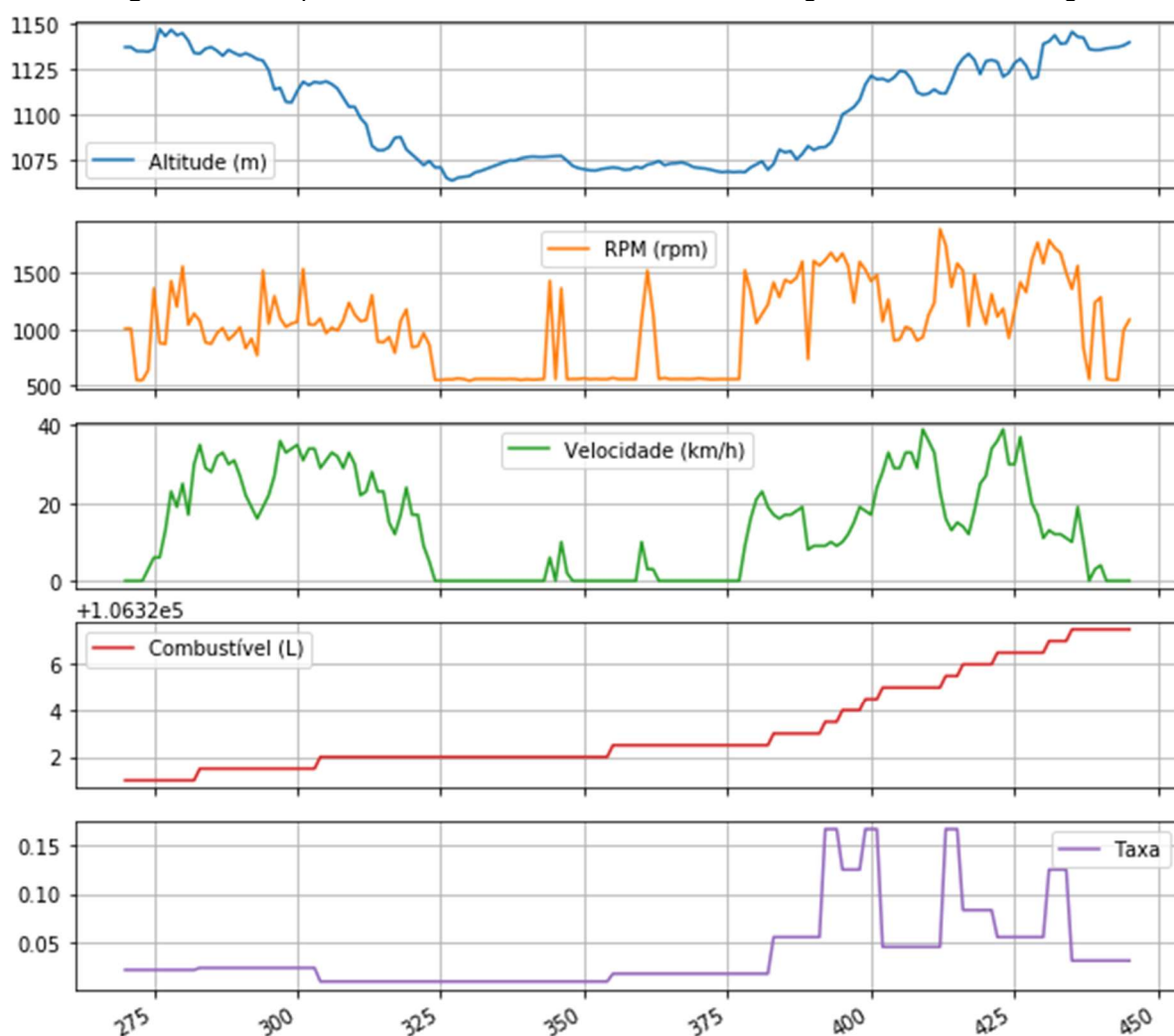
Tabela 10 - Número de ocorrências em diferentes valores de combustível

Combustível	Número de pontos agrupados
106312.5	7
106313.0	29
106313.5	9
106314.0	8

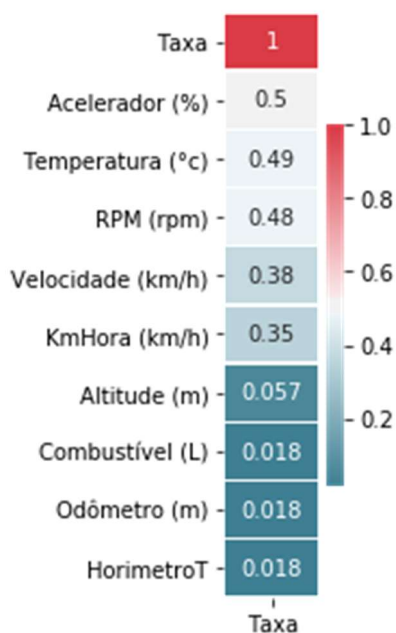
Fonte: Do autor (2020)

A variável taxa representa o aumento do consumo de combustível, com isso é possível verificar quais condições, valores das outras variáveis, podem influenciar no aumento do consumo do combustível. Na Figura 27 é possível visualizar a nova variável dentro da delimitação apresentada anteriormente na Figura 19. Ainda na Figura 27 é possível perceber que a subida do caminhão em altitude está relacionada em aumentos no valor da variável taxa. Na Figura 28 está apresentado a correlação (spearman) da nova variável com as demais, é possível perceber que nesta formatação o consumo de combustível relaciona-se melhor com as outras variáveis, quando comparado a variável que mede esse valor fornecido diretamente pela CAN.

Figura 27 - Comportamento de variáveis dentro de uma viagem do veículo de carga



Fonte: Do autor (2020)

Figura 28 - Correlação (Spearman) da variável taxa

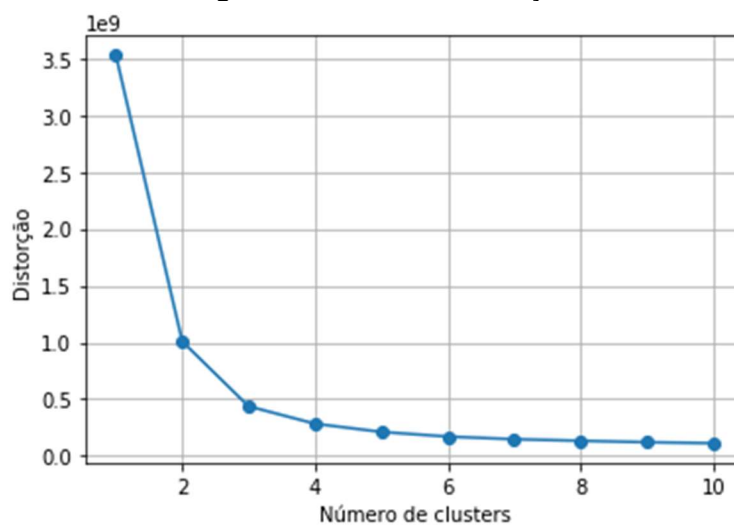
Fonte: Do autor (2020)

Tendo em vista a importância do controle do consumo do combustível no setor de mineração, e as possibilidades da utilização da variável taxa para fazer inferências, torna-se interessante a utilização de técnicas para a predição de diferentes cenários. Utilizando como entrada as variáveis altitude, RPM e velocidade e saída a variável taxa, foi implementado com a técnica k-nn um algoritmo para predição deste valor. Foi testado em duas implementações, a primeira utilizando KNeighborsRegressor, os parâmetros utilizados estão apresentados na figura 29 e o resultado da predição na figura 30. A determinação do número de cluster foi feito via gráfico de distorção figura 31. Nas duas implementações para treinamento foi utilizado os 25804 dados, e na previsão os últimos 445. Também nas duas implementações também foi utilizado a distância Minkowsky.

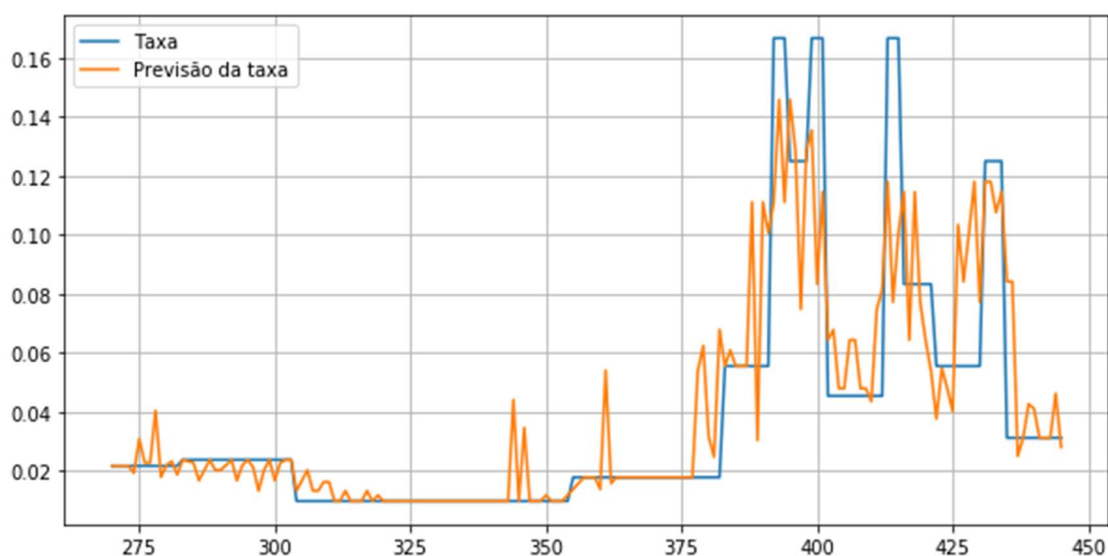
Figura 29 - Parâmetros utilizados na predição KNN

```
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=Ellipsis, p=2,
weights='uniform')
```

Fonte: Do autor (2020)

Figura 30 - Gráfico de distorção

Fonte: Do autor (2020)

Figura 31 - Predição da variável taxa knn (k=5)

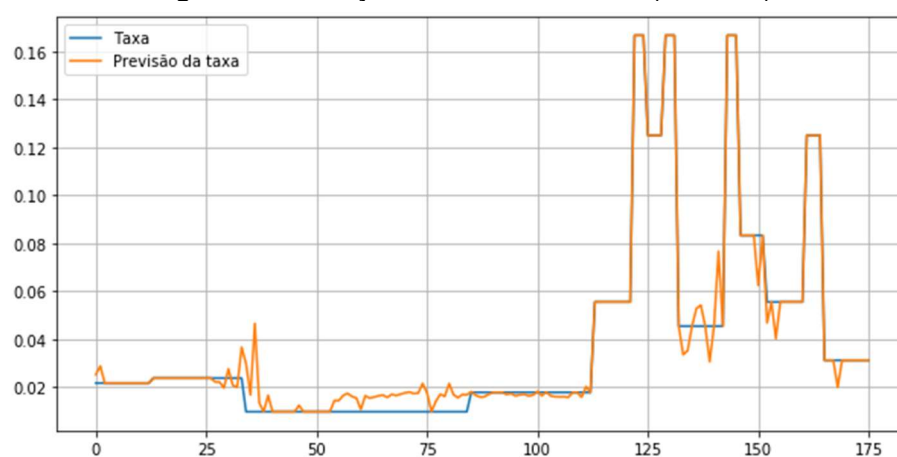
Fonte: Do autor (2020)

A segunda implementação utilizando RadiusNeighborsRegressor, os parâmetros utilizados então na Figura 30 e o resultado em sequência na Figura 31. A determinação do raio (variável *radius*) foi obtido experimentalmente, testando do número 7 até o número 1, o valor que apresentou o melhor resultado foi com o valor 2. A segunda implementação apresentou os melhores resultados, visualmente, quando comparado a primeira.

Figura 30 - Parâmetros utilizados na predição KNN (Radios=1.0)

```
RadiusNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',  
metric_params=None, n_jobs=None, p=2, radius=Ellipsis,  
weights='uniform')
```

Fonte: Do autor (2020)

Figura 31 - Predição da variável taxa knn (radios=2)

Fonte: Do autor (2020)

CONCLUSÃO

Este trabalho possibilitou entender a possibilidade de ganhos com o uso de análise de dados dentro da indústria da mineração, utilizando-se de informações obtidas por sistemas de despacho de caminhões que realizam o transporte de material dentro de uma mina. Dada à importância do setor da mineração, o desenvolvimento de formas para analisar os processos envolvidos permite a economia financeira, de tempo e de recursos, atuando como forte aliada no processo de tomada de decisão.

O objetivo geral de buscar técnicas e formas de visualização gráficas, para a análise de dados de sistemas de despacho, com objetivo de obter possíveis ganhos no processo foi cumprido, apresentando como resultados visualizações que permitem a redução de custos com a alocação, cumprimento de normas de segurança entre outras.

O trabalho iniciou com uma revisão de literatura que permitiu compreender sobre o processo de alocação de veículos utilizados na mineração e as principais tecnologias envolvidas nesta atividade. Na sequência, são apresentadas as técnicas utilizadas para a extração do conhecimento, a partir dos dados brutos. Como o objetivo do trabalho, de por meio dos dados trazer uma compreensão melhor do processo, o trabalho passou então para a análise das variáveis, primeiramente de forma separada e posteriormente a suas correlações.

A metodologia adotada foi a do estudo de caso, utilizando para isso um sistema específico de gerenciamento de despacho, mas que apresenta as principais características aplicadas no setor. Foi então realizada uma coleta de dados, de aproximadamente uma semana, por meio destes foi então aplicadas as técnicas descritas na revisão bibliográfica.

A utilização da correlação de *sperman* permitiu a confirmação e a quantificação de relações diretas entre variáveis analisadas, enquanto o uso de clusters, feito com a utilização da técnica *kmeans*, permitiu identificar relações entre as variáveis não diretamente proporcionais, permitindo uma análise mais completa das relações entre os dados analisados.

A principal contribuição do trabalho, no caminho de tornar mais eficiente a tomada de decisão no gerenciamento de veículos de transporte utilizados na mineração, foi da criação de um método para a tornar possível a utilização da variável

combustível no formato padrão fornecido por sistemas de despacho e posteriormente na previsão do consumo de combustível utilizando para isso a regressão feita com *knn*, em duas implementações com a *Radius Neighbors Regressor* com o melhor resultado.

A principal limitação do trabalho é referente a amostra, utilizando dados de apenas um caminhão, isso acontece pela dificuldade da obtenção destes dados junto às empresas. Outra limitação é a falta de material de referência bibliográfica específico sobre a análise de dados de sistema de despacho, este tema não é muito presente no meio acadêmico ficando mais ligado ao setor profissional.

Por fim, o estudo possibilitou então uma análise de diferentes dados obtidos via sistema de despacho eletrônico. Foram testadas técnicas de análise diversas e representações que permitiram uma compreensão mais consistente sobre as variáveis envolvidas no processo, como também desenvolvendo novas metodologias para a coleta das informações, como foi o caso da variável combustível, permitindo assim, que os objetivos propostos fossem alcançados.

TRABALHOS FUTUROS

Em trabalhos futuros seria interessante a análise, por meio dos algoritmos aqui desenvolvidos, em veículos com mais variáveis disponíveis, além da comparação das informações aqui trabalhadas em diferentes veículos em uma mesma situação. Outra sugestão, é a da utilização dos algoritmos em diferentes modelos de caminhões verificando se o comportamento se mantém ainda neste mesmo contexto.

Na metodologia da análise de combustível via CAN aqui desenvolvida, poderia verificar a aplicação prática deste algoritmo, comparando o consumo em diferentes situações, por exemplo variando a altitude.

No apêndice a este trabalho está o resultado da análise por cluster de todas as variáveis utilizadas neste trabalho, a análise destas pode vir a trazer novos relacionamentos não apresentados neste trabalho.

REFERÊNCIAS

ALMEIDA, C. M.; DE CASTRO NEVES, T.; DE FIGUEIREDO, A. C. L. ESTUDO DE VIABILIDADE ECONÔMICA DO USO DE CORREIAS TRANSPORTADORAS NAS OPERAÇÕES DE LAVRA DE UMA MINA DE MINÉRIO DE FERRO. [s.d.].

ALVARENGA, G. B. Despacho ótimo de caminhões numa mineração de ferro utilizando algoritmo genético com processamento paralelo. **Belo Horizonte, MG, Brasil: Master's thesis**, 1997.

ASSOCIATION, N. M. E. **NMEA 0183--Standard for interfacing marine electronic devices**. [s.l.] NMEA, 2002.

BORGES, L. E. **Python para Desenvolvedores. 2ª**, 2013.

BROWNLEE, J. **How to Calculate Correlation Between Variables in Python**. Disponível em: <<https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/>>.

BROWNLEE, J. **A Gentle Introduction to NumPy Arrays in Python**. Disponível em: <<https://machinelearningmastery.com/gentle-introduction-n-dimensional-arrays-python-numpy/>>.

CURI, A. **Minas a céu aberto: planejamento de lavra**. [s.l.] Oficina de Textos, 2014.

DA CUNHA, S. B.; CARVAJAL, S. **Estatística Basica-a Arte de Trabalhar com Dados**. [s.l.] Elsevier Brasil, 2009.

DATA CORE SYSTEM. Clustering and Dimensionality Reduction: Understanding the “Magic” Behind Machine Learning. [s.d.].

DE ALMEIDA GUIMARÃES, A. O Protocolo CAN Bus nas Aplicações Off-Road: Uma Análise Comparativa entre os Padrões Existentes. 2001.

DE AMO, S. Técnicas de mineração de dados. **Jornada de Atualização em Informatica**, 2004.

EL-RABBANY, A. **Introduction to GPS: the global positioning system**. [s.l.] Artech house, 2002.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [s.l.: s.n.].

FARBER, B.; LARSON, R. **Estatística Aplicada. 4ª Edição** Pearson Brasil, São

Paulo, , 2009.

FONTANA, A.; NALDI, M. C. **Estudo e comparação de métodos para estimação de números de grupos em problemas de agrupamento de dados**. São Carlos: ICMC, 2009. Disponível em: <http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_340.pdf>

FREITAS, H. M. R. Análise de dados qualitativos: aplicação e tendências mundiais em sistemas de informação. **Revista de Administração-RAUSP**, v. 35, n. 4, p. 84–102, 2000.

GERMANI, D. J. A mineração no Brasil. **Relatório Final, Brasil**, 2002.

gps.gov. Disponível em: <<https://www.gps.gov/systems/gps/space/>>. Acesso em: 10 mar. 2021.

GUIMARÃES, A. A.; SARAIVA, A. M. **Um roteiro de implementação de uma rede CAN (Controller Area Network)**. Conferência Internacional de Engenharia Automotiva-SIMEA. **Anais...**2003

HARVEY, C. **what is machine learning**. Disponível em: <<https://www.datamation.com/big-data/what-is-machin>>. Acesso em: 24 mar. 2019.

HUNTER, J. et al. Matplotlib-development-team. matplotlib. **See <https://matplotlib.org/>**. Retrieved, p. 5–21, 2018.

INSTITUTO BRASILEIRO DE MINERAÇÃO (IBRAM). **Relatório Anual de Atividades julho de 2018 -junho de 2019**. [s.l.: s.n.]. Disponível em: <<http://portaldamineracao.com.br/ibram/wp-content/uploads/2019/07/relatorio-anual-2018-2019.pdf>>.

INSTRUMENTS, T. Introduction to the controller area network (CAN). **SLOA101A. Dallas, TX**, 2016.

KOPPE, J. C. A lavra e a indústria mineral-estado da arte e tendências tecnológicas. 2007.

MAGALHÃES, M. N.; DE LIMA, A. C. P. **Noções de Probabilidade e Estatística (6a Edição Revista e Ampliada) Vol. 40**. [s.l.] Edusp, 2008.

MCKINNEY, W. Pandas, python data analysis library. 2015. **Reference Source**, 2014.

MESQUITA, P. P. D.; CARVALHO, P. S. L. DE; OGANDO, L. D. Desenvolvimento e inovação em mineração e metais. 2016.

Nearest Neighbors. Disponível em: <<https://scikit-learn.org/stable/modules/neighbors.html#regression>>. Acesso em: 3 jul. 2020.

NOGARE, D. **Entendendo como funciona o algoritmo de Cluster K-Means.** Disponível em: <<http://www.diegonogare.net/2015/08/entendendo-como-funciona-o-algoritmo-de-cluster-k-means/>>. Acesso em: 12 jun. 2019.

PERKTOLD, J.; SEABOLD, S.; TAYLOR, J. Statsmodels. **URL** <https://www.statsmodels.org/stable/index.html>. [Online, [s.d.].

RODRIGO, W. S. F. J. P.; DE SOUZA, V. DA C. V.; COSTA, O. E. F. A. UTILIZAÇÃO DA TELEMETRIA EMBARCADA PARA GERENCIAMENTO DO ABASTECIMENTO DE CAMINHÕES DE GRANDE PORTE NA MINERAÇÃO. **Tecnologia em Metalurgia, Materiais e Mineração**, v. 16, n. 2, p. 189–195, 2019.

RUSSEL, S.; NORVIG, P. Inteligência Artificial. 3a Edição. **Editora Campus**, 2013.

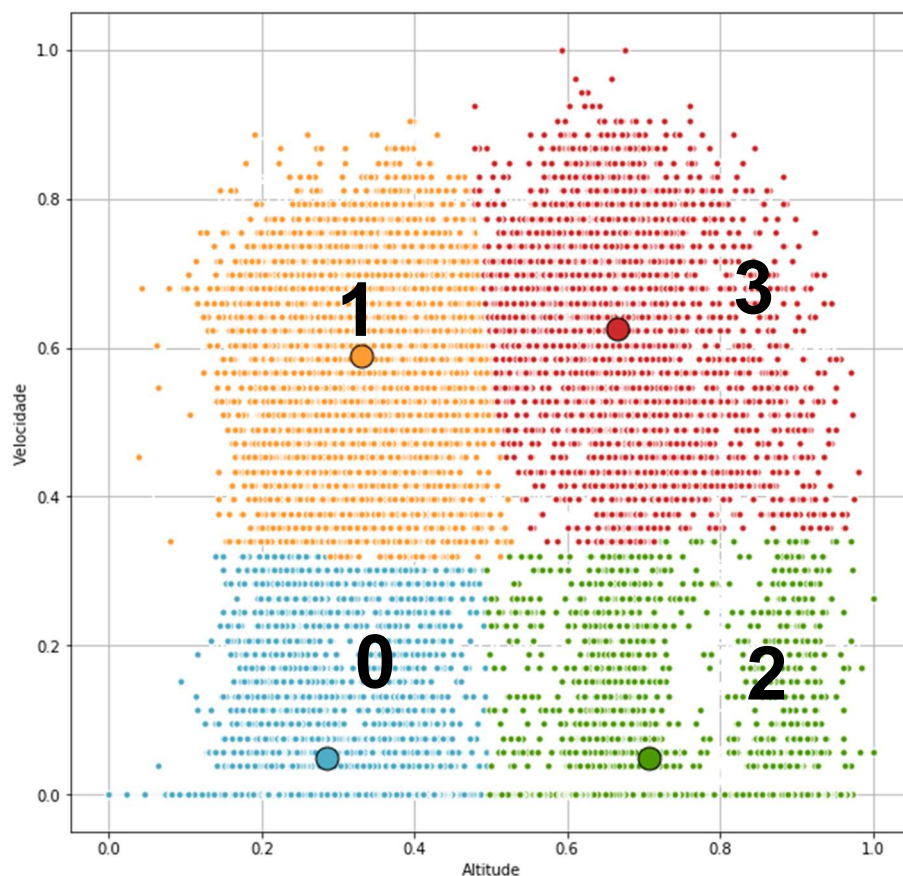
SAYAD, S. **Clustering.** Disponível em: <<https://www.saedsayad.com/clustering.htm>>. Acesso em: 25 mar. 2019.

APÊNDICE – ANÁLISE POR CLUSTERS

1. ALTITUDE E VELOCIDADE.....	61
2. TEMPERATURA E VELOCIDADE	61
3. TEMPERATURA E RPM	64
4. ALTITUDE E RPM	69
5. VELOCIDADE E RPM.....	75
6. ACELERADOR RPM.....	78

1. ATITUDE E VELOCIDADE

Figura 1 - Cluster entre os dados de Altitude e Velocidade



Fonte: Do autor (2020)

Tabela 1 - Coordenadas dos cluster dos dados de altitude e velocidade

Cluster	Centro	
	x	y
0	0.28404186657819874	0.04795396566252713
1	0.33107882145286766	0.589692700861205
2	0.7059918610379479	0.04948175660707266
3	0.6655704409244406	0.626405168206595

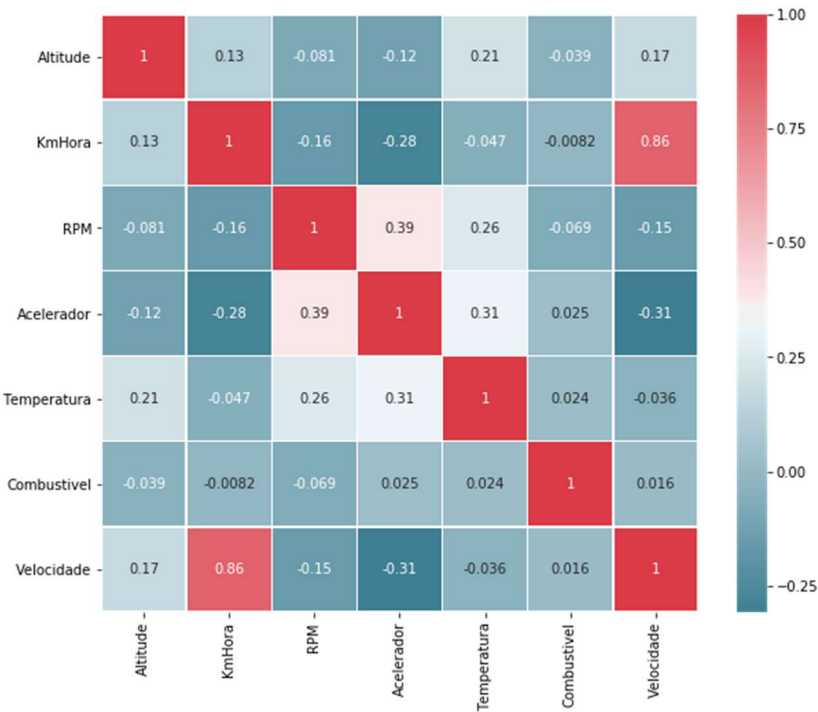
Fonte: Do autor (2020)

Tabela 2 – Número de pontos em cada cluster dos dados de altitude e velocidade

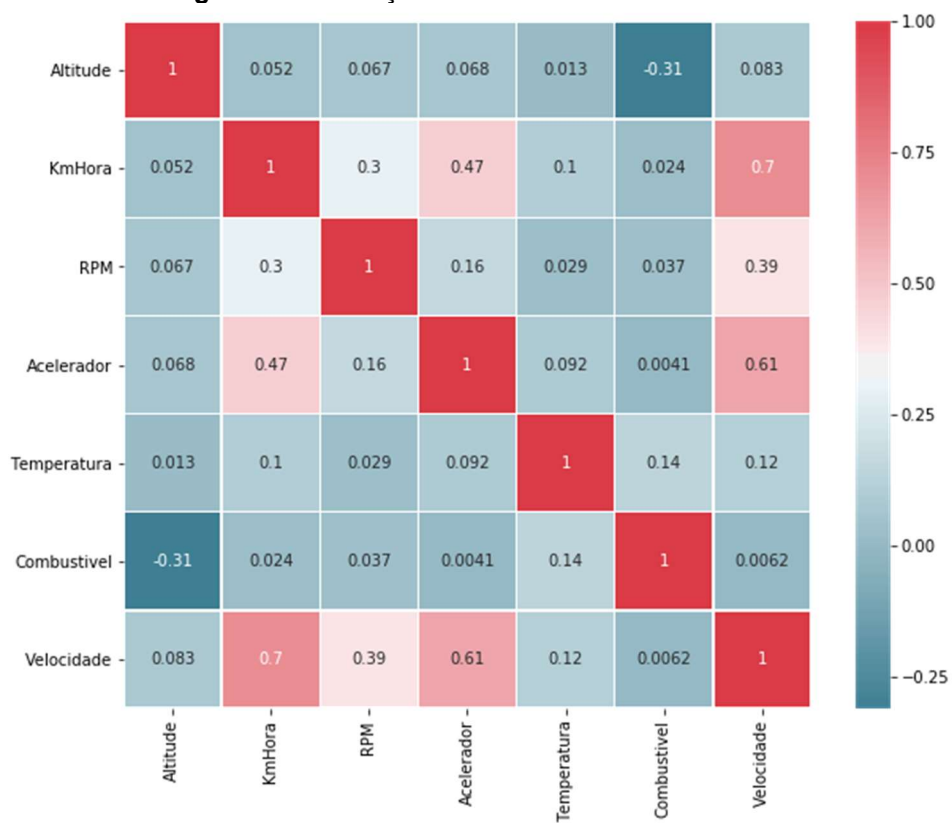
Cluster	Nº de pontos
1	9022
0	8219
3	4560
2	4003

Fonte: Do autor (2020)

Figura 2 – Correlação entre as variáveis do cluster 0

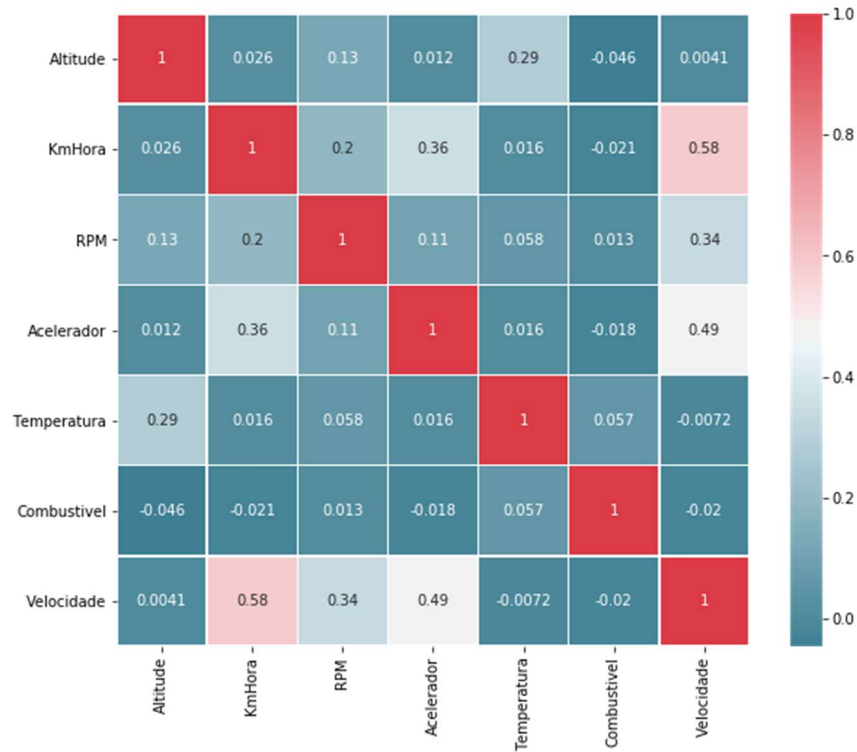


Fonte: Do autor (2020)

Figura 3 - Correlação entre as variáveis do cluster 1

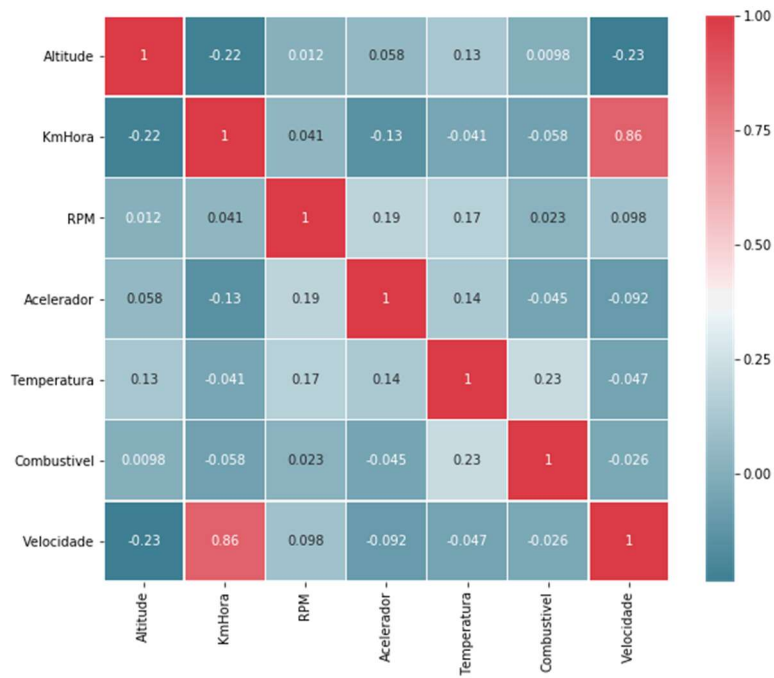
Fonte: Do autor (2020)

Figura 4 - Correlação entre as variáveis do cluster 2



Fonte: Do autor (2020)

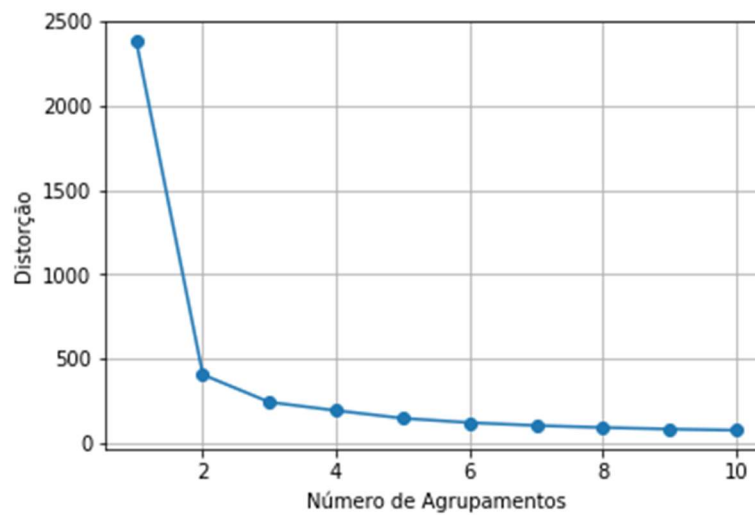
Figura 5 - Correlação entre as variáveis do cluster 3



Fonte: Do autor (2020)

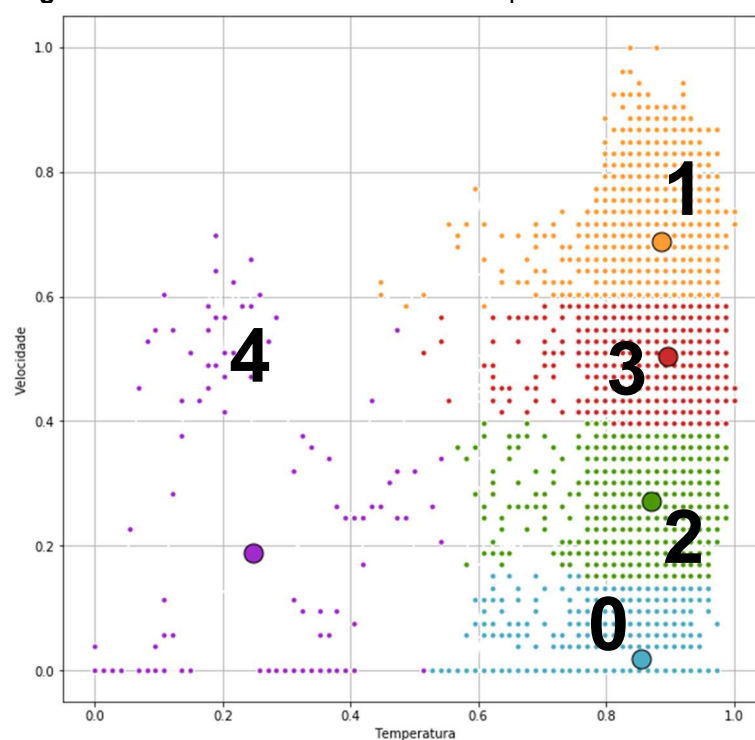
2. TEMPERATURA E VELOCIDADE

Figura 6 - Gráfico de distorção entre os dados de Temperatura e velocidade



Fonte: Do autor (2020)

Figura 7 - Cluster entre os dados de Temperatura e Velocidade



Fonte: Do autor (2020)

Figura 3 - Número de pontos em cada cluster dos dados de temperatura e velocidade

Cluster	tamanho
0	10398
1	7788
3	5027
4	2453
2	138

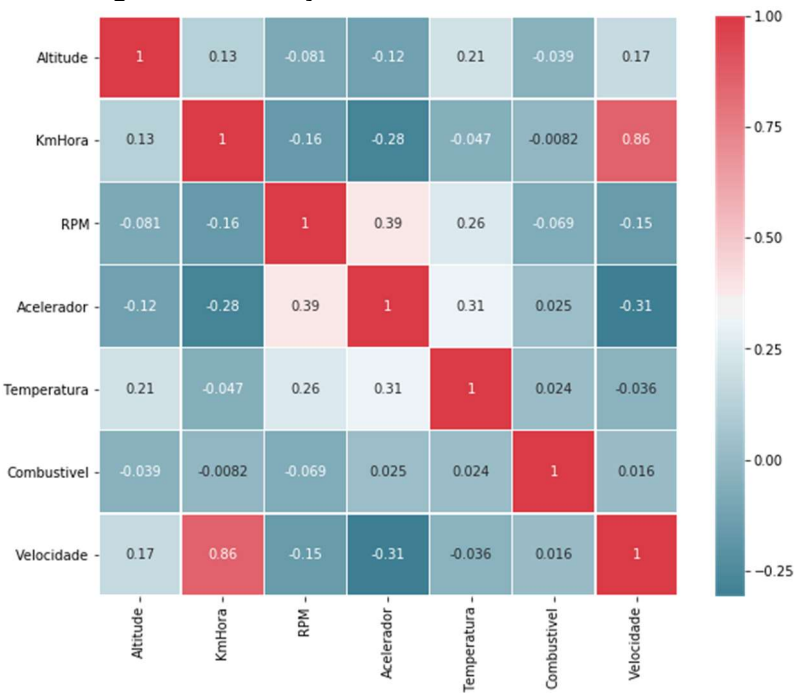
Fonte: Do autor (2020)

Tabela 4- Coordenadas dos cluster dos dados de temperatura e velocidade

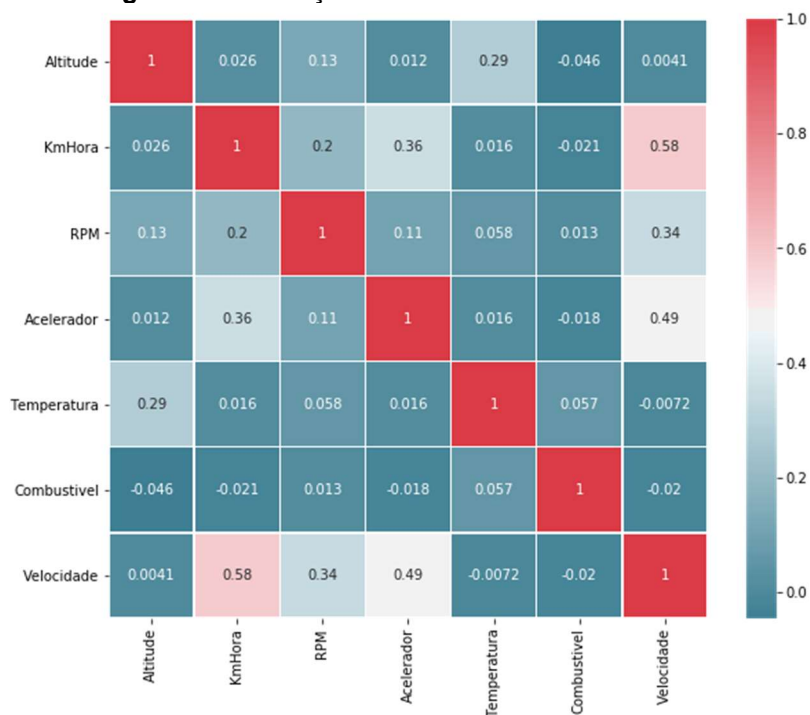
Centro		
Cluster	x	y
0	86.24782569712148	1.526136465521052
1	88.66523963662931	35.29643938602626
2	88.84972622186179	21.953559115798225
3	41.90909090909102	10.496503496503493

Fonte: Do autor (2020)

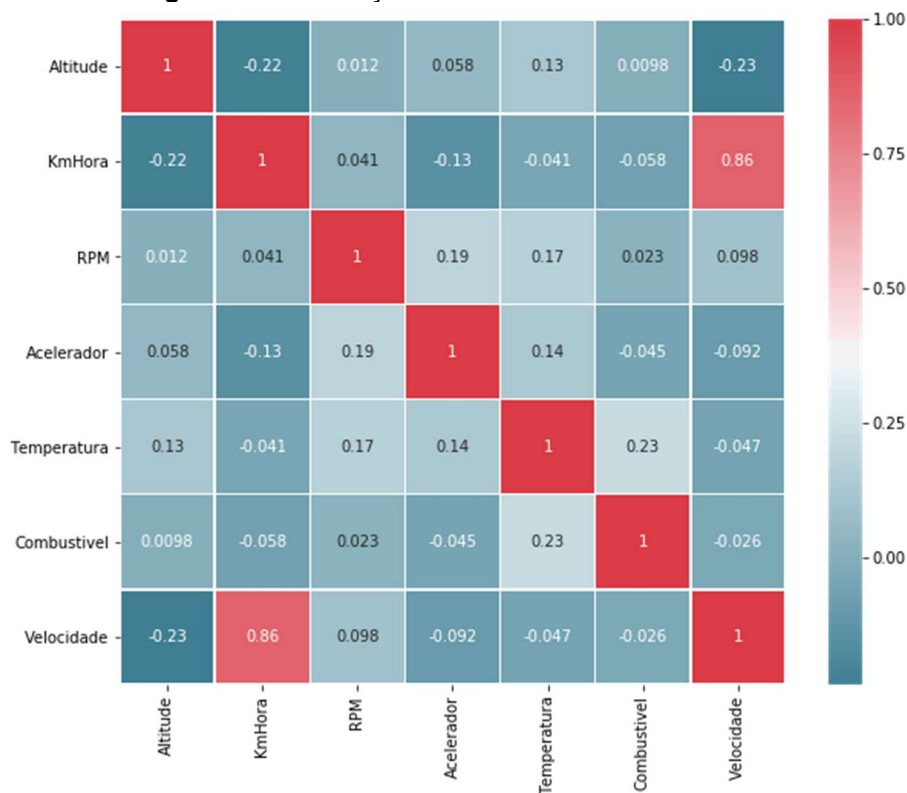
Figura 8 - Correlação entre as variáveis do cluster 0



Fonte: Do autor

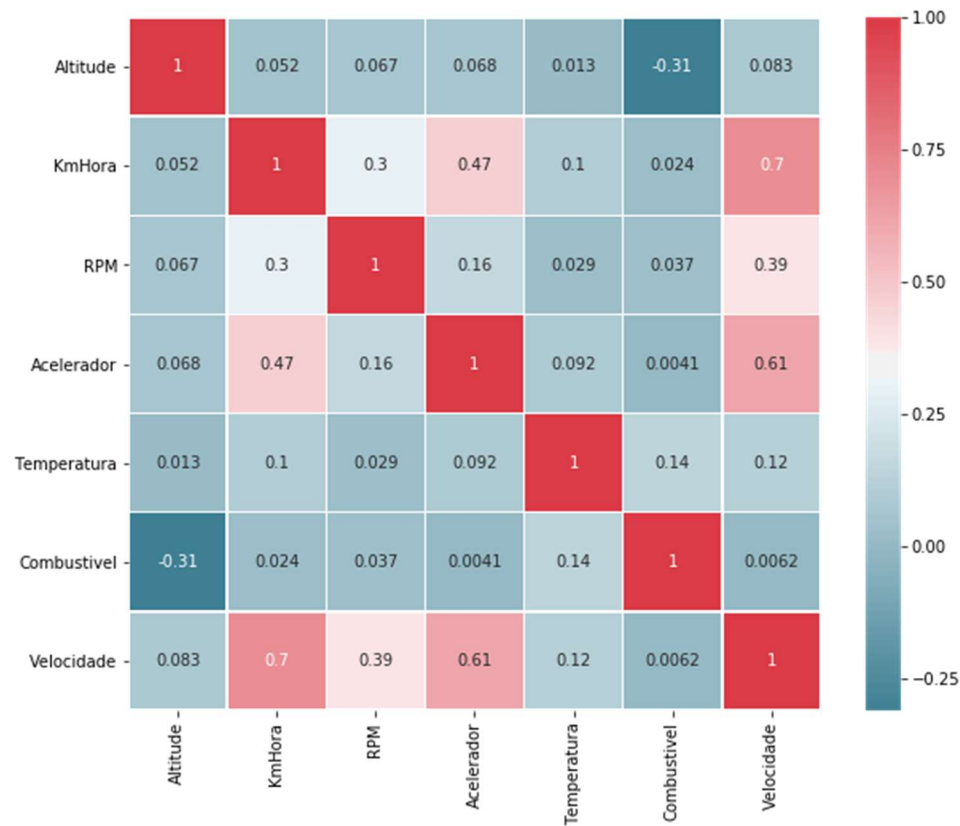
Figura 9 - Correlação entre as variáveis do cluster 1

Fonte: Do autor.

Figura 10 - Correlação entre as variáveis do cluster 2

Fonte: Do autor.

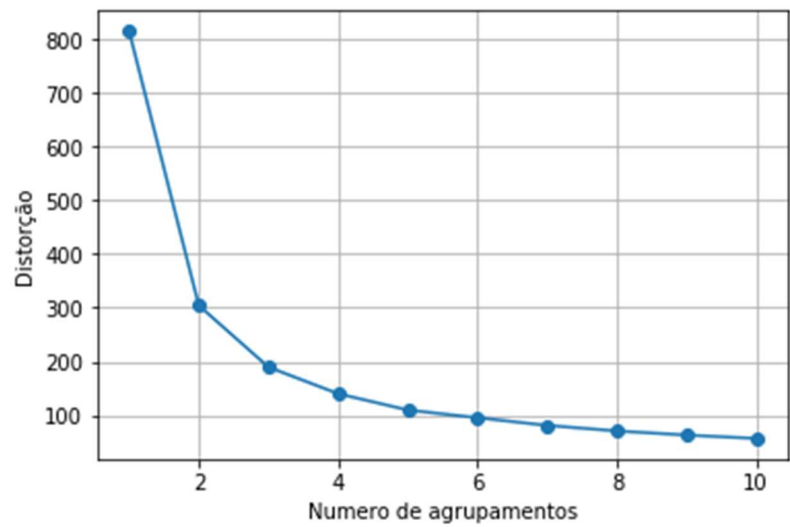
Figura 11 - Correlação entre as variáveis do cluster 3



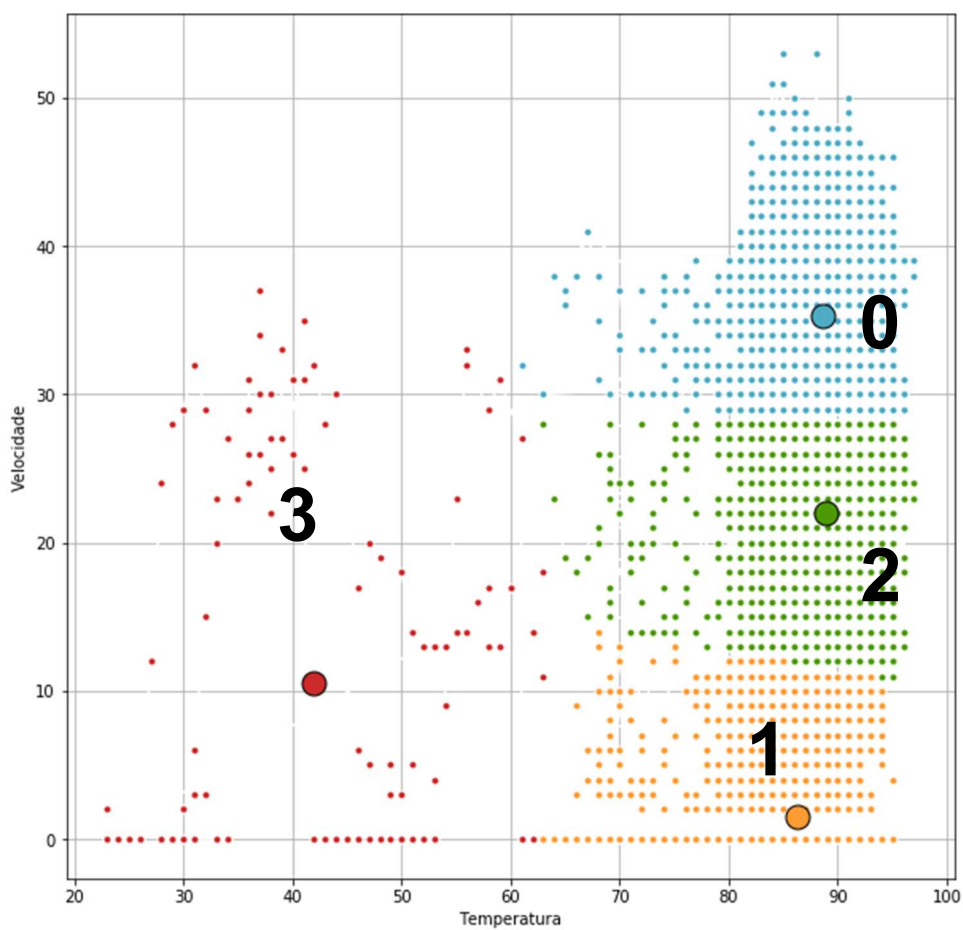
Fonte: Do autor.

3. TEMPERATURA E RPM

Figura 12 -



Fonte: Do autor

Figura 13 - Cluster entre os dados de Temperatura e Velocidade

Fonte: Do autor

Tabela 5 - Número de pontos em cada cluster dos dados de temperatura e RPM

Cluster	tamanho
1	8117
0	7035
2	6632
3	3871
4	149

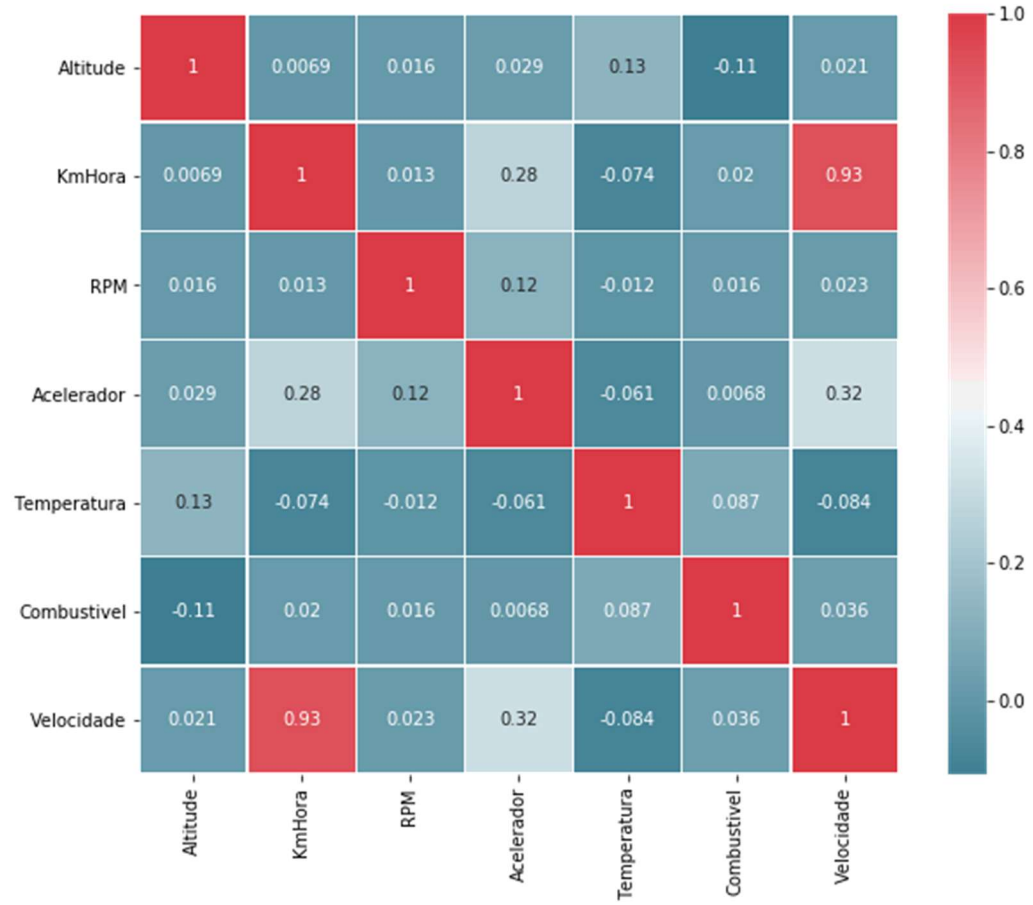
Fonte: Do autor

Tabela 6 - Coordenadas dos cluster dos dados de Temperatura e RPM

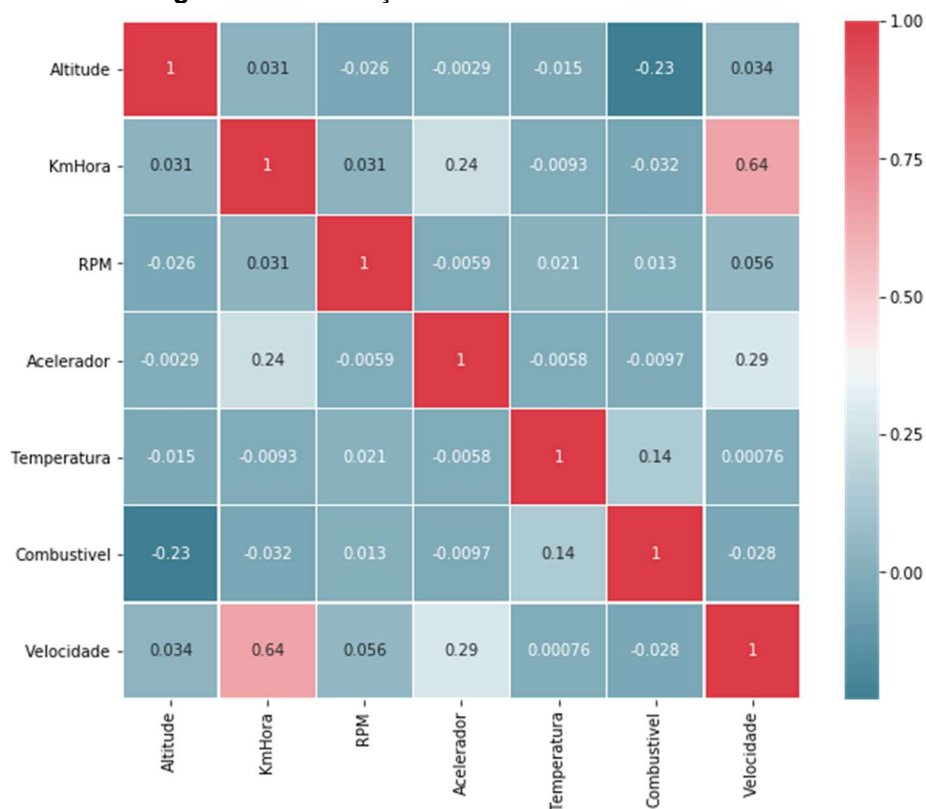
Centro		
Cluster	x	y
0	0.8724312798939665	0.39720268916892737
1	0.8894671209206793	0.5224068653677907
2	0.8947300439163006	0.6597612526937902
3	0.8519656776401879	0.2124196536773527
4	0.2673680391801194	0.344804049988843

Fonte: Do autor

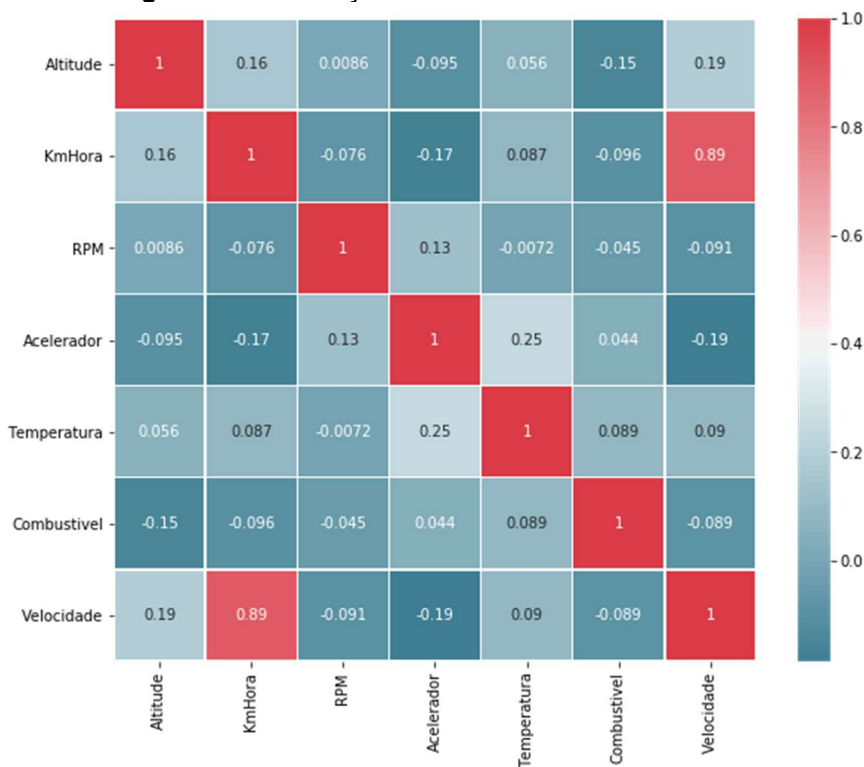
Figura 14 - Correlação entre as variáveis do cluster 0



Fonte: Do autor.

Figura 15 - Correlação entre as variáveis do cluster 1

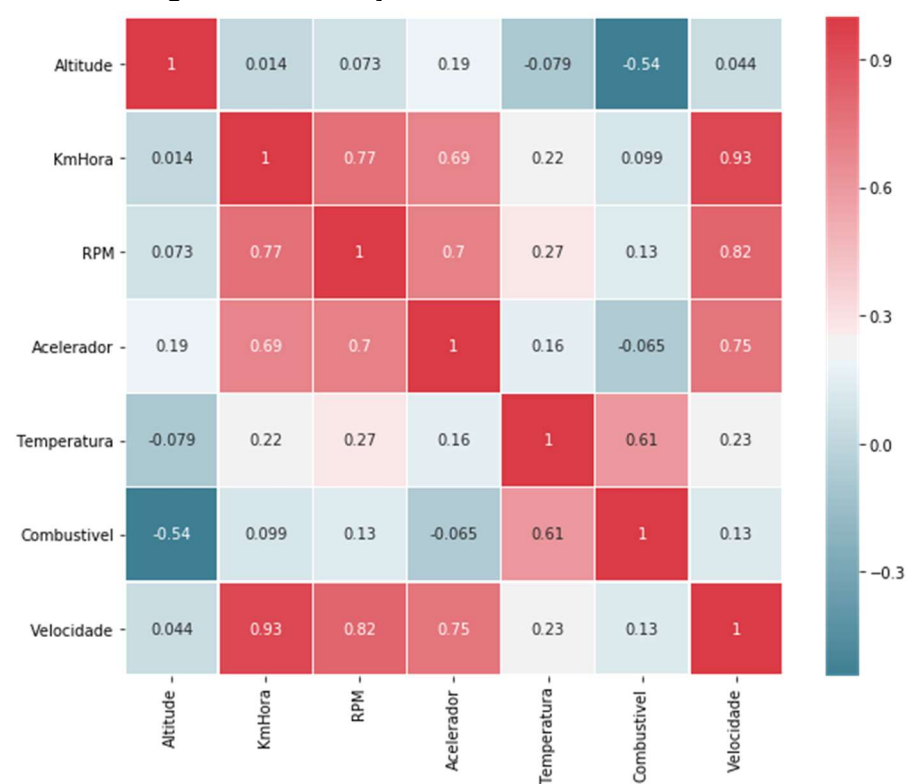
Fonte: Do autor.

Figura 16 - Correlação entre as variáveis do cluster 2

Fonte: Do autor.

Figura 17 - Correlação entre as variáveis do cluster 3

Fonte: Do autor

Figura 18 - Correlação entre as variáveis do cluster 4

Fonte: Do autor

3.1. ALTITUDE E RPM

Figura 18 -

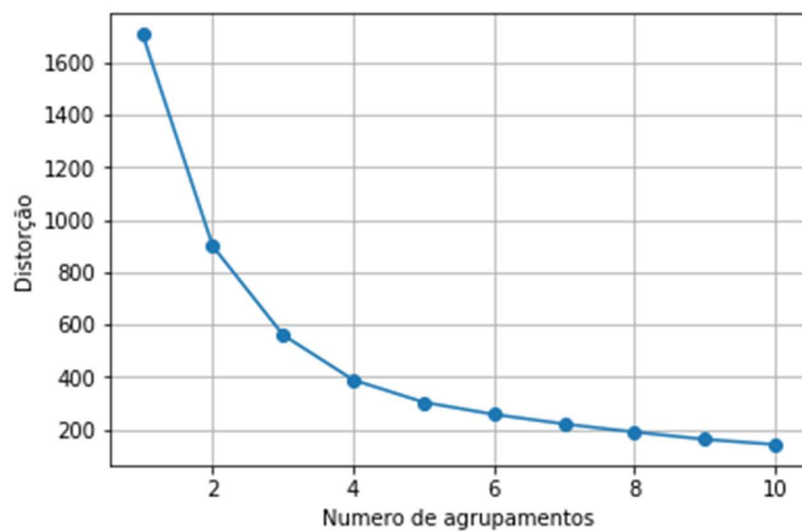
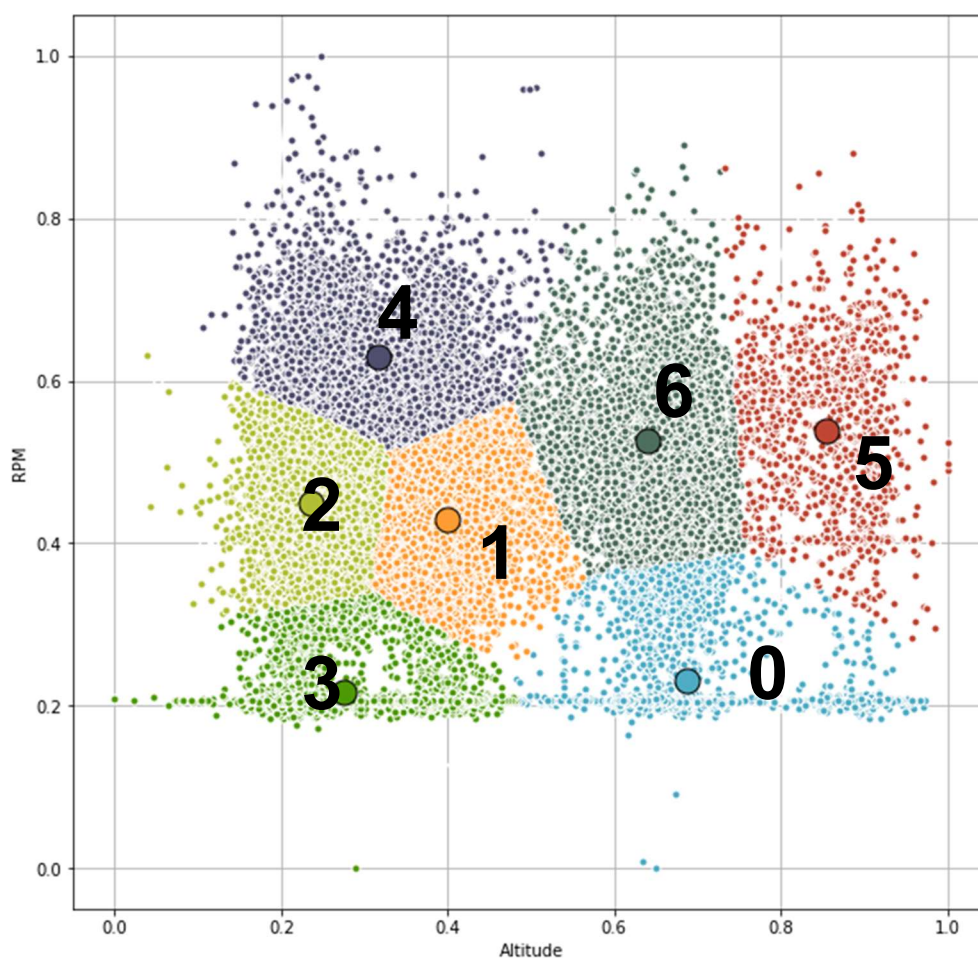


Figura 19 - Cluster entre os dados de altitude e RPM



Fonte: Do autor

Tabela 8 - Coordenadas dos cluster dos dados de altitude e RPM

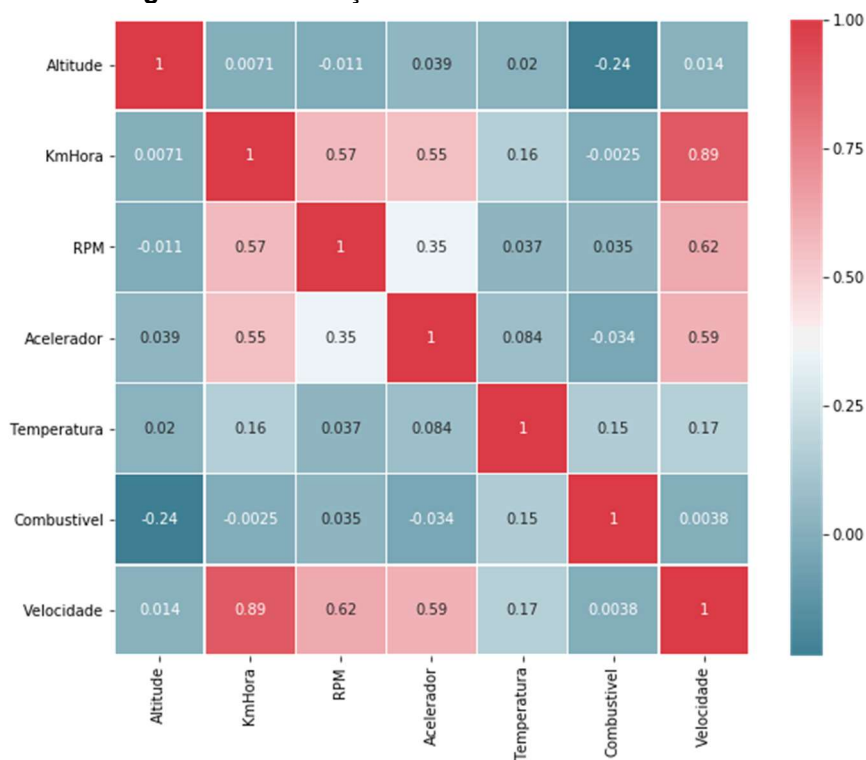
Centro		
Cluster	x	y
0	0.68708639	0.23185809
1	0.39839571	0.42846583
2	0.2741897	0.21622456
3	0.64069252	0.5258639
4	0.31558793	0.62912429
5	0.85513746	0.53894258
6	0.23523373	0.45032352

Fonte: Do autor

Tabela 9 - Número de pontos em cada cluster dos dados de altitude e RPM

Cluster	tamanho
2	5467
1	4803
3	3917
4	3909
0	3384
6	3218
5	1106

Fonte: Do autor. 1

Figura 20 - Correlação entre as variáveis do cluster 0

Fonte: Do autor.

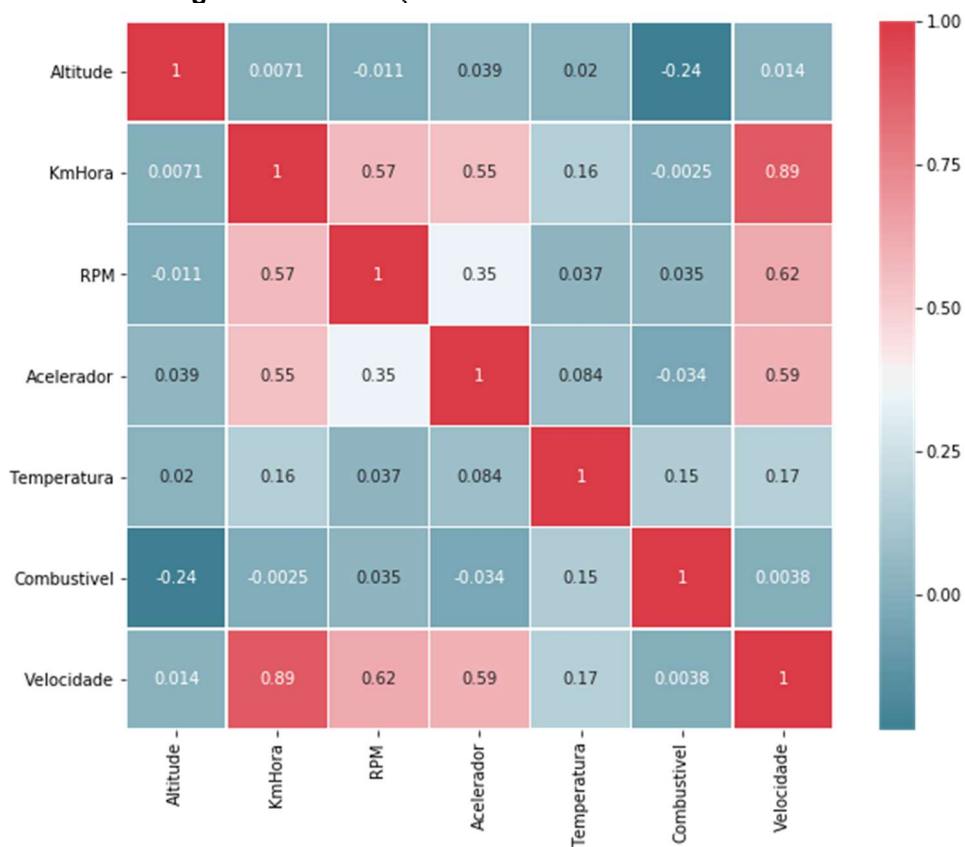
Figura 21 - Correlação entre as variáveis do cluster 1

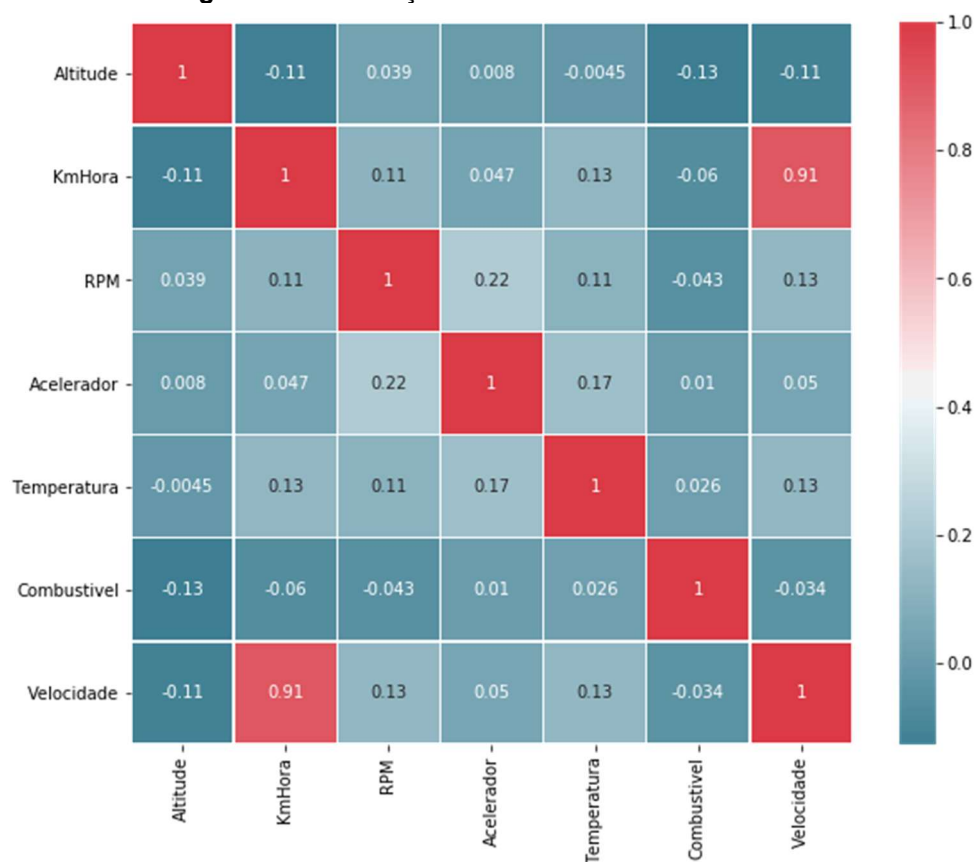
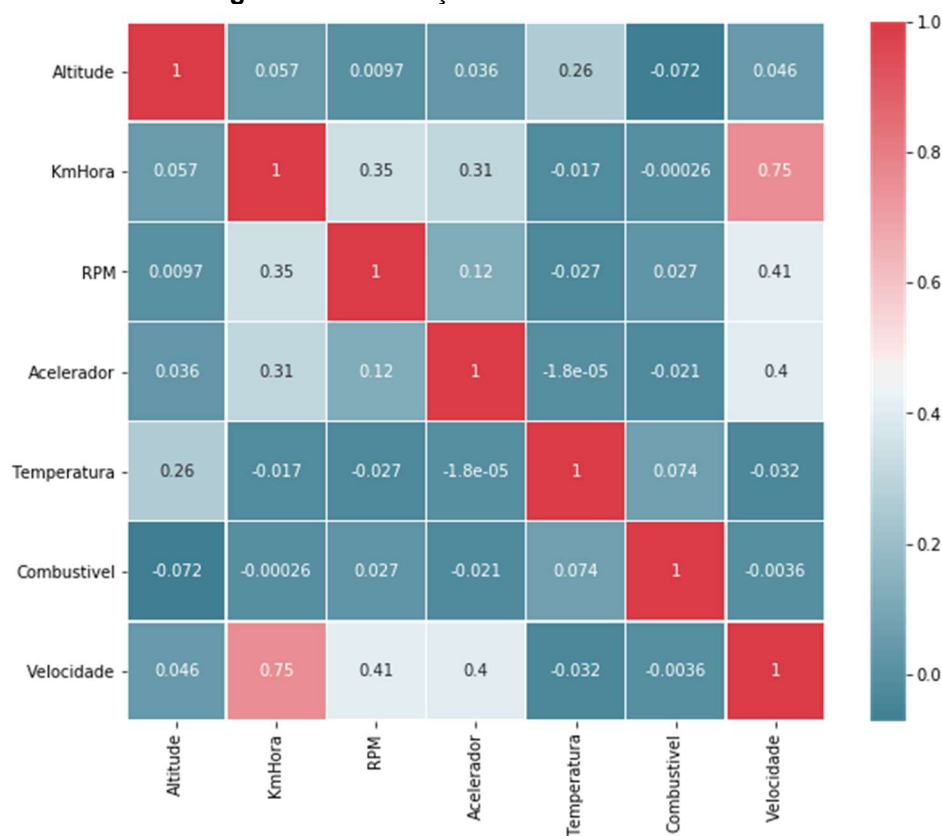
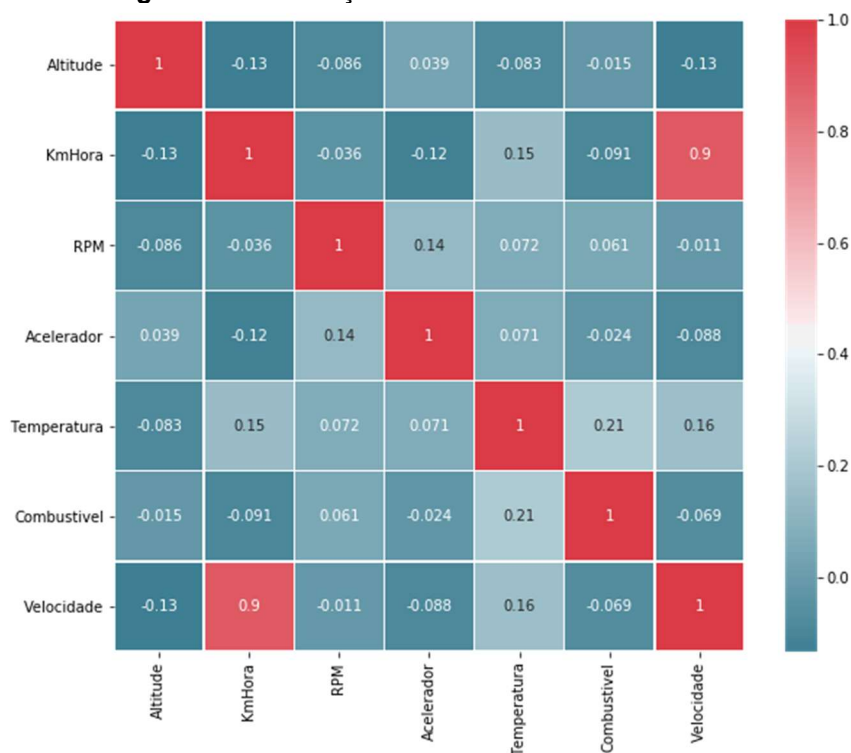
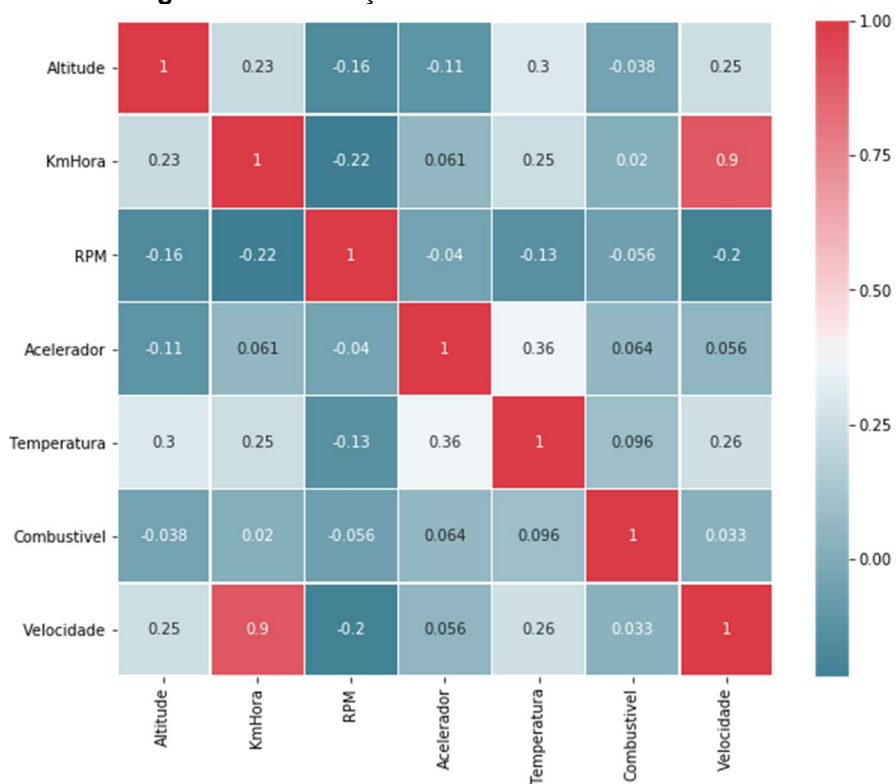
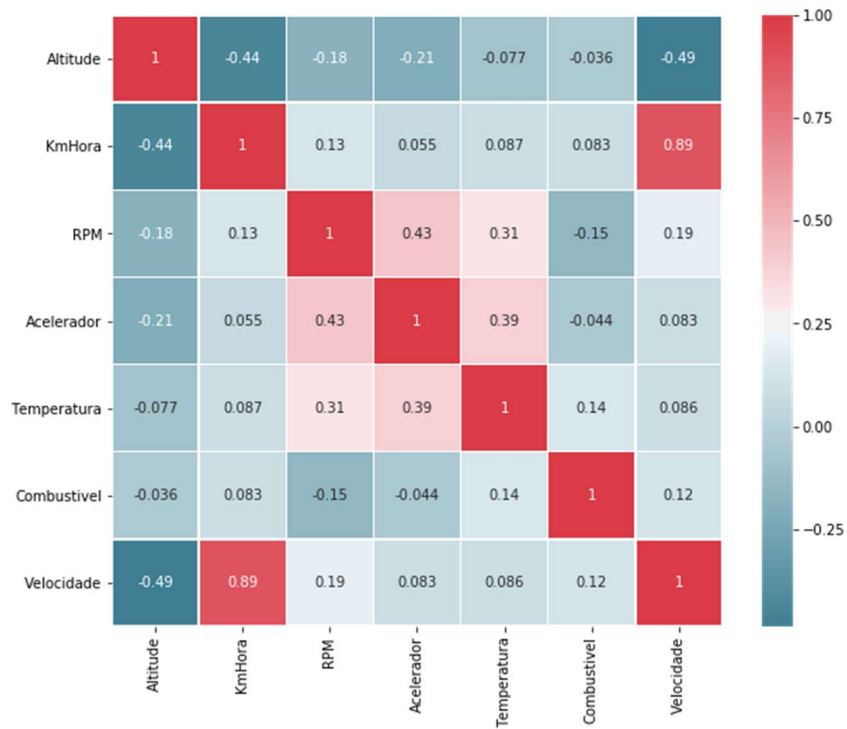
Figura 22 - Correlação entre as variáveis do cluster 2**Figura 23 - Correlação entre as variáveis do cluster 3**

Figura 24 - Correlação entre as variáveis do cluster 4**Figura 25 - Correlação entre as variáveis do cluster 5**

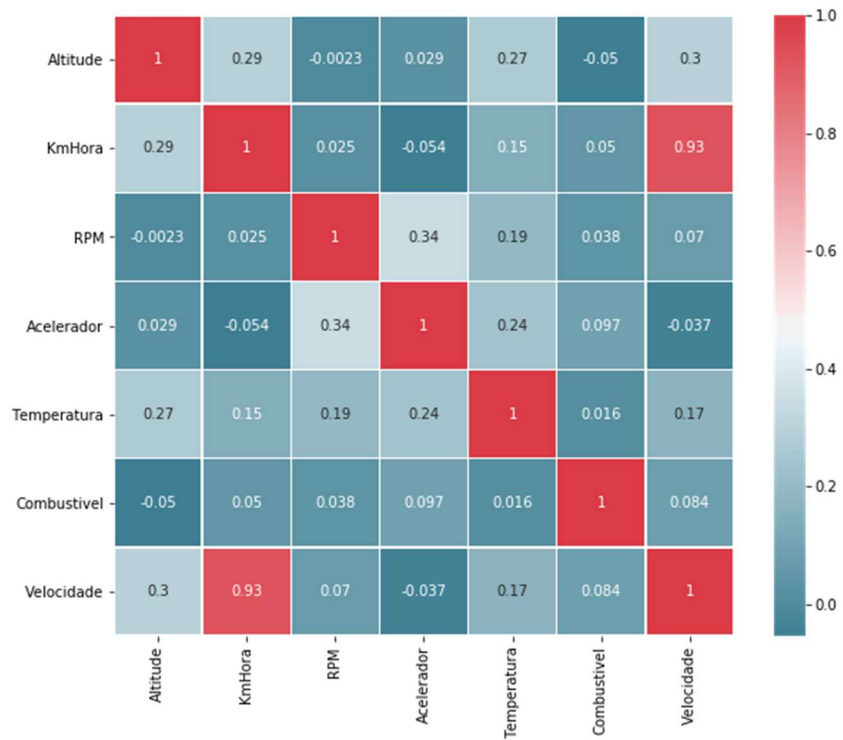
Fonte: Do autor

Figura 26 - Correlação entre as variáveis do cluster 6



Fonte: Do autor

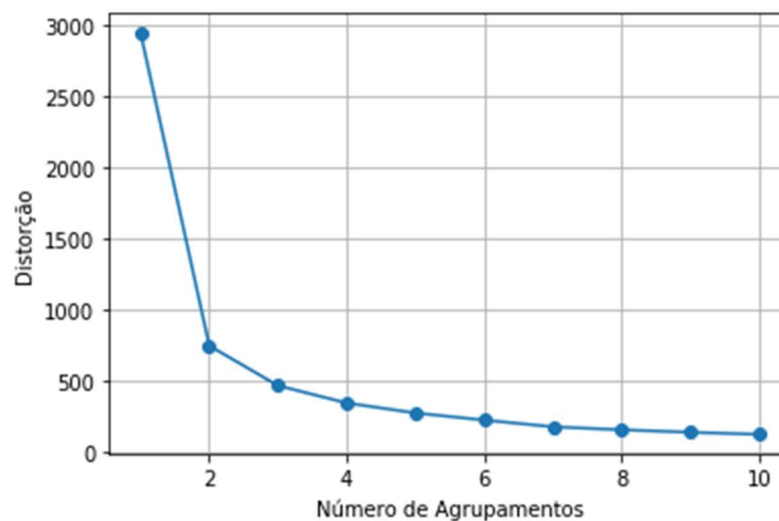
Figura 27- Correlação entre as variáveis do cluster 7



Fonte: Do autor

5. VELOCIDADE E RPM

Figura 28 -



Fonte: Do autor

Figura 29 - Cluster entre os dados de RPM e Velocidade

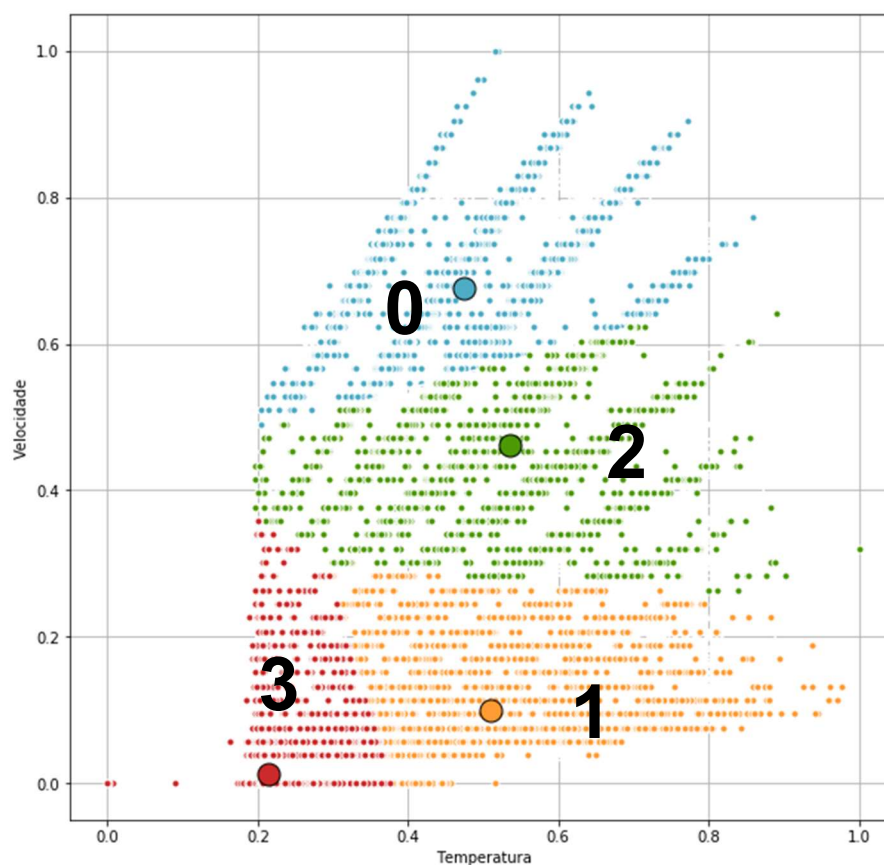


Tabela 11 - Coordenadas dos cluster dos dados de Velocidade e RPM

Centro		
Cluster	x	y
0	0.4738400597951976	0.6769068658830787
1	0.5099407261836002	0.0984547864118534
2	0.5351338732977652	0.46090698444223555
3	0.2136308258864271	0.012584058947979815

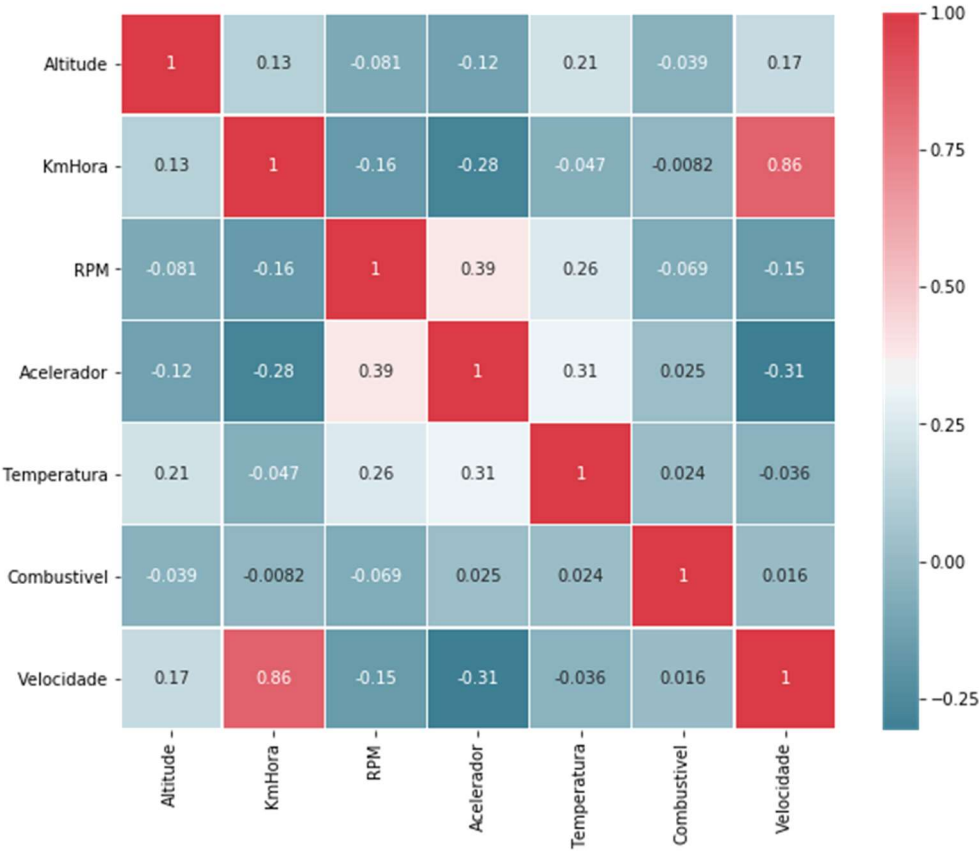
Fonte: Do autor

Tabela 12 - Número de pontos em cada cluster dos dados de temperatura e velocidade

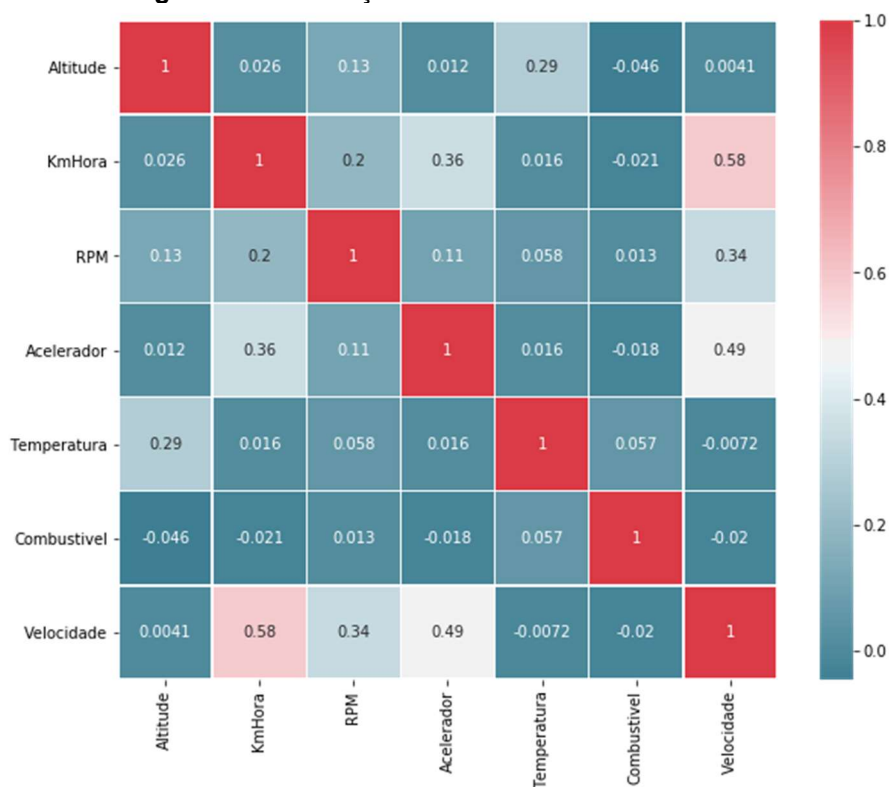
Cluster	tamanho
0	8571
3	8088
2	5421
1	3724

Fonte: Do autor. 2

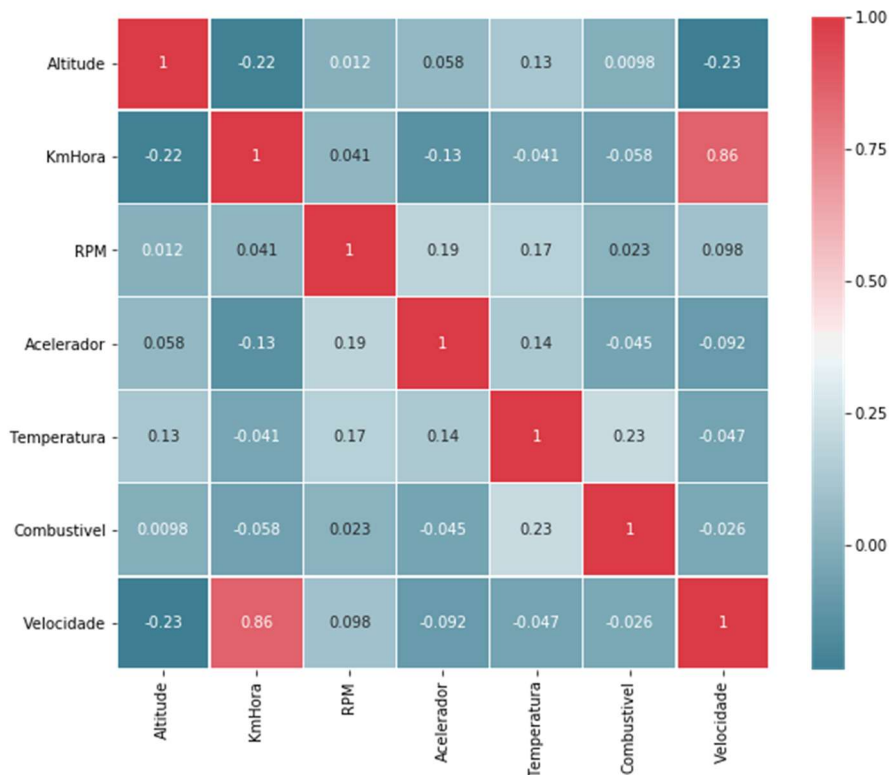
Figura 30 - Correlação entre as variáveis do cluster 0



Fonte: Do autor

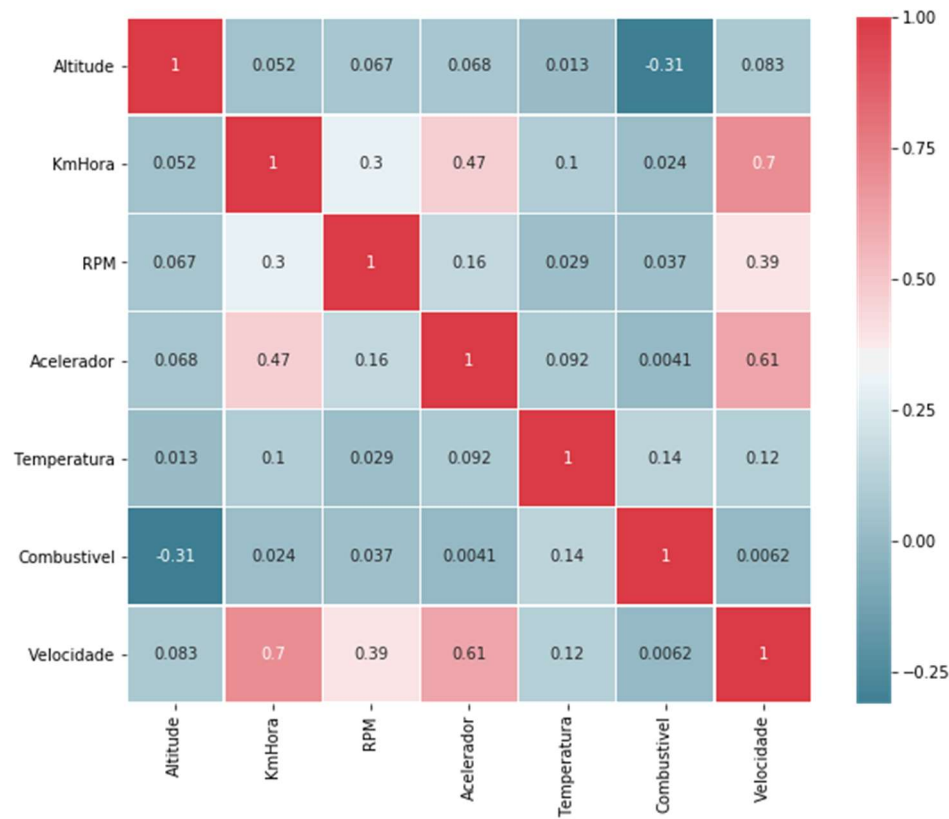
Figura 31 - Correlação entre as variáveis do cluster 1

Fonte: Do autor

Figura 32 - Correlação entre as variáveis do cluster 2

Fonte: Do autor

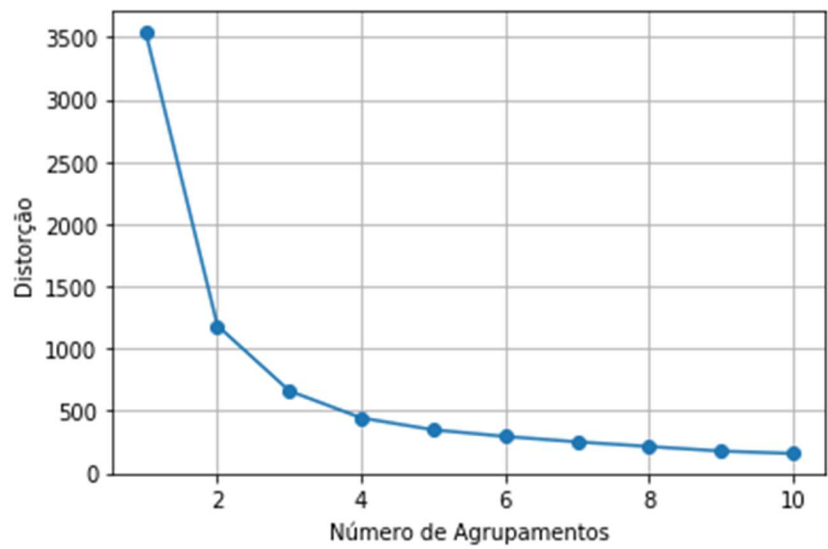
Figura 33 - Correlação entre as variáveis do cluster 3



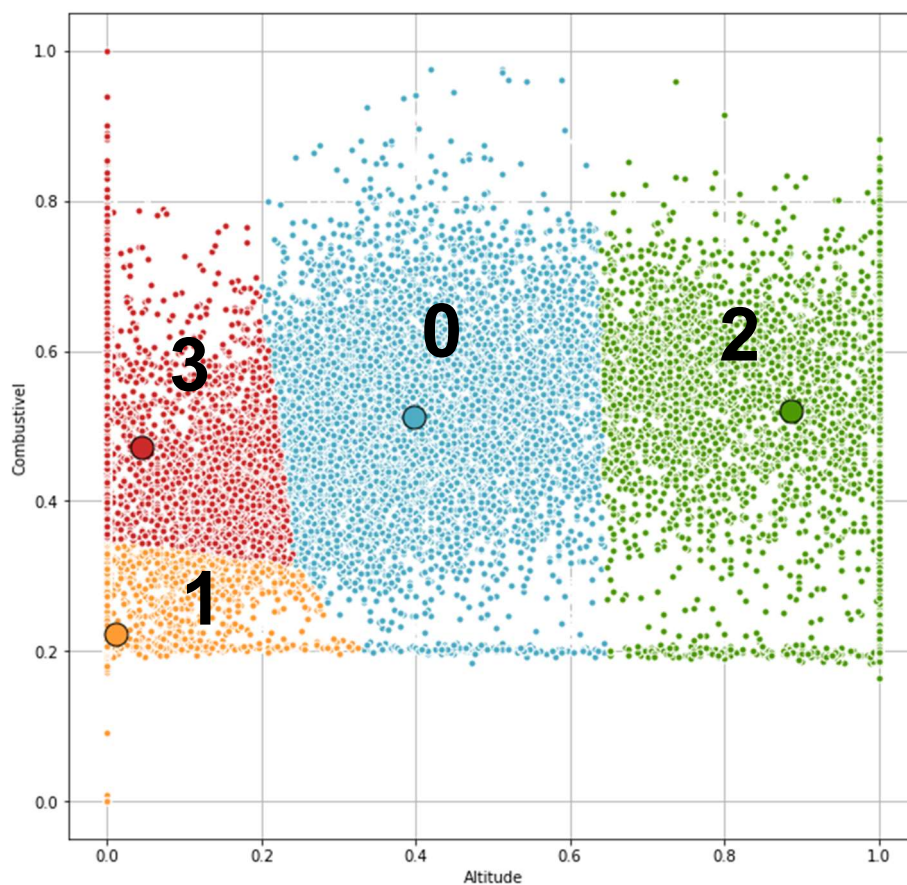
Fonte: Do autor

6. ACELERADOR RPM

Figura 34 -



Fonte: Do autor

Figura 35 - Cluster entre os dados de acelerador e RPM

Fonte: Do autor

Tabela 12 - Coordenadas dos cluster dos dados de acelerador e RPM

Centro		
Cluster	x	y
0	0.8865980414484138	0.52000777792429
1	0.01056005806216892	0.2215855994317414
2	0.39855031541003294	0.5131901284750942
3	0.04582120838470624	0.47108862703340315

Fonte: Do autor

Tabela 13 - Número de pontos em cada cluster dos dados de acelerador e RPM

Cluster	tamanho
1	8268
2	6661
3	6478
0	4397

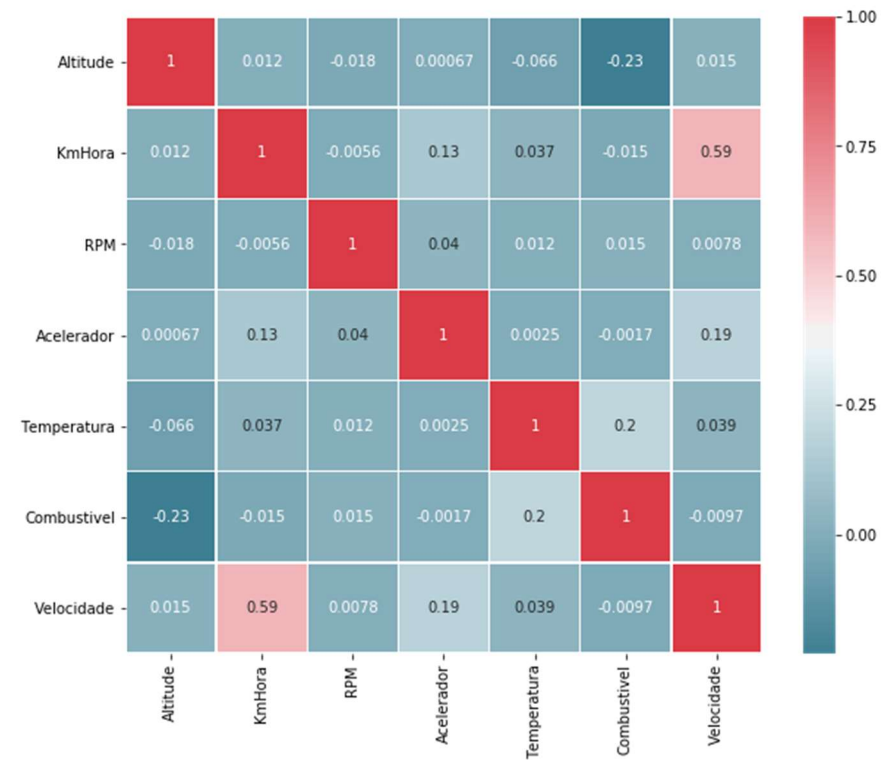
Fonte: Do autor.

Figura 36 - Correlação entre as variáveis do cluster 0

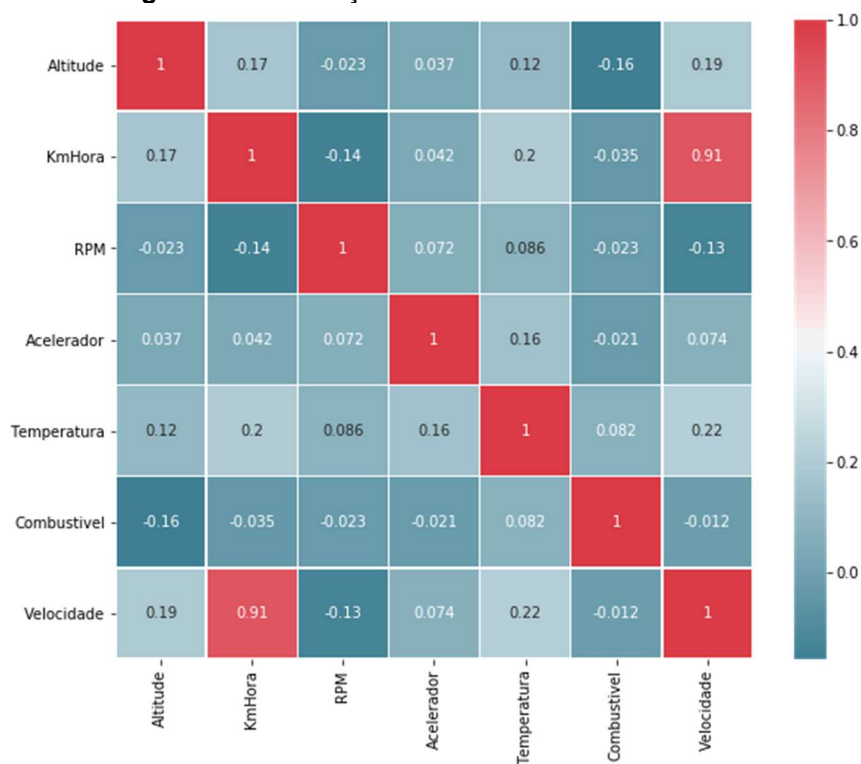


Fonte: Do autor

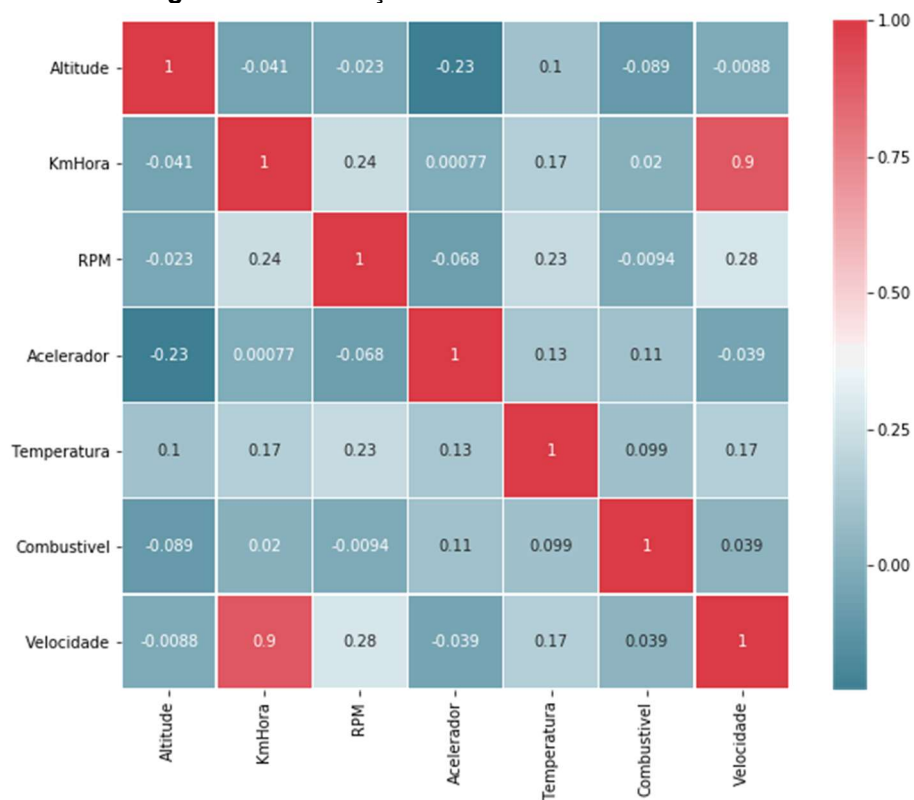
Figura 37 - Correlação entre as variáveis do cluster 1



Fonte: Do autor.

Figura 38 - Correlação entre as variáveis do cluster 2

Fonte: Do autor

Figura 39 - Correlação entre as variáveis do cluster 3

Fonte: Do autor