INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À ENGENHARIA E GESTÃO

DANIEL NOCERA DE CAMPOS

UTILIZAÇÃO DE ALGORITMO DE MACHINE LEARNING COMO FERRAMENTA PARA PREVER VENDAS EM UMA DISTRIBUIDORA DE BEBIBAS NA REGIÃO DOS LAGOS

Campos dos Goytacazes/RJ

2020

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À ENGENHARIA E GESTÃO

DANIEL NOCERA DE CAMPOS

UTILIZAÇÃO DE ALGORITMO DE MACHINE LEARNING COMO FERRAMENTA PARA PREVER VENDAS EM UMA DISTRIBUIDORA DE BEBIBAS NA REGIÃO DOS LAGOS

Cristine Nunes Ferreira

(Orientadora)

Rogerio Atem de Carvalho

(Coorientador)

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, no Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão (MPSAEG), como parte dos requisitos necessários à obtenção do título de Mestre em Sistemas Aplicados à Engenharia e Gestão.

Campos dos Goytacazes/RJ

2020

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGI FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À ENGENHARIA E GESTÃO

DANIEL NOCERA DE CAMPOS

UTILIZAÇÃO DE ALGORITMO DE MACHINE LEARNING COMO FERRAMENTA PARA PREVER VENDAS EM UMA DISTRIBUIDORA DE BEBIBAS NA REGIÃO DOS LAGOS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, no Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão (MPSAEG), como parte dos requisitos necessários à obtenção do título de Mestre em Sistemas Aplicados à Engenharia e Gestão.

Aprovado(a) em 17 de Outubro de 2020

Banca Examinadora:

_____

Cristine Nunes Ferreira,

Doutora em Física pelo (CBPF)

Instituto Federal de Educação, Ciência e tecnologia Fluminense

(Orientadora)

_____

Dalessandro Soares Vianna,

Doutor em Informática – PUC-RIO

Universidade Federal Fluminense (ICT/UFF)

(Externo)

_____

Rogerio Atem de Carvalho,
Doutor
Instituto Federal de Educação, Ciência e tecnologia Fluminense

ARTIGO 1: ANÁLISE BIBLIOMÉTRICA DE MACHINE LEARNING E SUAS APLICAÇÕES EM PREVISÃO DE VENDAS

RESUMO

Nos dias de hoje mantemos um estilo de vida, que a maior parte das interações que temos com as pessoas acontece online: WhatsApp, e-mail, mídias sociais, etc. Todas essas interações geram uma grande quantidade de dados, uma fonte valiosa de informações para as empresas, que pode estar sendo subutilizada. Essa fonte pode ser extremamente útil para a área de Vendas, que historicamente já utiliza métricas para medir o seu desempenho. Faz todo o sentido, que uma das atividades que mais exigem a interação entre pessoas – vendas – esteja passando por um renascimento digital. Para fazer uso de toda essa massa de dados e melhorar as vendas, as empresas estão começando a utilizar softwares de inteligência artificial que aprendem continuamente a partir da aprendizagem de máquina, mais como conhecido como em inglês *Machine Learninng,* que otimizam recomendações em tempo real para a equipe de vendas**.** O objetivo deste trabalho é propor um modelo de aprendizagem de máquina, para o estudo do comportamento das vendas e predições futuras utilizando como estudo de caso, o setor de bebidas em uma distribuidora situada na região dos lagos no Rio de Janeiro. Os dados utilizados são do período de 2018 à 2019. Com essa metodologia foi possível verificar o aumento ou diminuição das vendas no período e projeções futuras. Essa informação é utilizada para dimensionamento de estoque, equilíbrio da remuneração da equipe de vendas e recursos a serem investidos no mercado.

Palavras-chave: **Aprendizado de Máquina, Vendas no Setor de Bebidas, Região dos Lagos.**

ABSTRACT

Nowadays, we maintain a lifestyle, which is most of the interactions we have as people online: WhatsApp, email, social media, etc. All of these interactions generate a large amount of data, a valuable source of information for companies, which may be underutilized. This source can be extremely useful for the Sales area, which historically already uses metrics to measure its performance. It makes perfect sense, one of the activities that most involves interaction between people - sales - is going through a digital renaissance. To use all this mass of data and improve sales, like companies that are starting to use artificial intelligence software that learns from machine learning, better known as Machine Learningin, which optimizes tools in real time for a sales team. The objective of this work is to propose a machine learning model, to study sales behavior and future forecasts, using as a case study, or drinks sector in a distributor located in the region of the lakes of Rio de Janeiro. The data used are for the period from 2018 to 2019. With this methodology, it was possible to verify the increase or decrease in sales in the period and in future projections. This information is used for sizing inventory, balancing the remuneration of the sales team and resources to be invested in the market.

Key-Words: Machine Learning, Sales in the Beverage Sector, Cabo Frio

1. INTRODUÇÃO

A previsão de demanda é um dos maiores desafios em qualquer empresa no mercado global. Além disso sua precisão é fundamental para garantir vantagens competitivas como dimensionamento de estoque, lucro e promover o aumento das vendas. Para atingir a esses objetivos, precisamos estudar, quais os melhores métodos para tratar de forma eficaz esses problemas visado ampliar o setor de vendas e aumentar os lucros. Um exemplo das opções, mais interessantes é o setor de vendas eletrônicas. No atual momento que vivemos, onde a globalização pode parar o comercio presencial, devido a síndrome do pânico, atentados e pandemias, como a que vivemos nos dias de hoje, se torna um veículo interessante a compra pela internet. Neste caso o comércio se torna internacional, o que abrange um numero maior de compradores. A importância do estudo bibliométrico é entender a pesquisa científica sobre o assunto. Em Volles, Hoeltgebaum e Ammal, 2016 o estudo bibliométrico desse processo de internacionalização do comércio eletrônico, ajuda a entender o quanto o comércio eletrônico influencia a teoria tradicional da internacionalização. Diversos artigos descritos na base de dados Scopus moldaram este estudo, sob uma abordagem exploratória dedutiva. O comércio eletrônico pode ser considerado um modo de entrada para as PME e é um ambiente on-line para barreiras à exportação e estratégia de marketing. Os artigos pesquisados são teóricos e exploratórios, com diferentes modelos não replicáveis. No entanto, este artigo pode fornecer uma nova linha de perspectiva nas teorias da internacionalização, a fim de entender a importância desse fenômeno no ambiente internacional. Assim, a metodologia proposta visa criar a identificação e proposição de falhas, seguidas por diferentes etapas. Desta forma, o mapeamento científico, ou mapeamento bibliométrico, é um importante tópico de pesquisa no campo da bibliometria (Morris & Van Der Veer Martens, 2008; van Eck e Waltman, 2010). Tenta encontrar representações de conexões intelectuais dentro do sistema de conhecimento científico em mudança dinâmica (Small, 1997). Em outras palavras, o mapeamento científico visa exibir os aspectos estruturais e dinâmicos da pesquisa científica (Börner, Chen e Boyack, 2003; Morris & Van Der Veer Martens; Noyons, Moed e Luwel, 1999a). O fluxo geral de trabalho em uma análise de mapeamento científico etapas diferentes: recuperação de dados, pré-processamento, extração de rede normalização, mapeamento, análise e visualização. Ao final desse processo, o analista deve interpretar e obter algumas conclusões dos resultados.

Existem diferentes fontes bibliométricas em que os dados podem ser recuperados, como o ISI Web of Science (WoS) ou o Scopus. Além disso, uma análise de mapeamento científico pode ser realizada usando dados de patentes ou de financiamento. A etapa de pré-processamento é talvez uma das mais importantes uns. A bondade do resultado dependerá da qualidade dos dados. Vários métodos de pré-processamento podem ser aplicados,por exemplo, para detectar elementos duplicados e com erros ortográficos.

Diferentes abordagens foram desenvolvidas para extrair redes usando as unidades de análise selecionadas (autores, documentos periódicos, periódicos e termos). Análise de co-palavras (Callon, Courtial, Turner e Bauin, 1983) utiliza as mais importantes palavras ou palavras-chave dos documentos para estudar a concepção estrutura total de um campo de pesquisa. O co-autor analisa o autores e suas associações para estudar a estrutura social redes de colaboração (Gänzel, 2001; Peters & van Raan, 1991). Por fim, as referências citadas são utilizadas para lizar a base intelectual usada pelo campo de pesquisa ou para analise os documentos que citam as mesmas referências. Dentro Nesse sentido, o acoplamento bibliográfico (Kessler, 1963) analisa documentos citados, enquanto a análise de co-citação (Small, 1973) estuda os documentos citados. Outras abordagens como como acoplamento bibliográfico do autor (Zhao & Strotmann, 2008), co-citação do autor (White & Grifth, 1981), bibliografia da revista acoplamento gráfico (Gao & Guan, 2009; Small & Koenig, 1977) e cocitação de periódicos (McCain, 1991) são exemplosda análise macro usando dados agregados. Depois que a rede é construída, um processo de normalização é comumente realizada sobre a relação (arestas) entre seus nós usando medidas de similaridade. Uma revisão de similaridade. As medidas utilizadas no mapeamento científico foram realizadas em (van Eck & Waltman, 2009).

Com os dados normalizados, diferentes técnicas podem ser usado para construir o mapa (processo de mapeamento; Börner et al., 2003). Técnicas de redução de dimensionalidade, como princípios análise de componentes pal ou escala multidimensional (MDS), algoritmos de agrupamento e redes PathFinder (PFNETs) são amplamente utilizado. Os métodos de análise para mapeamento científico nos permitem extrair conhecimento útil dos dados. Análise de rede (Carrington, Scott e Wasserman, 2005; Cook & Holder, 2006; Skillicorn,2007; Wasserman & Faust, 1994) nos permite realizar uma análise estatística sobre os mapas gerados para mostrar diferentes medidas de toda a rede ou medidas de relacionamento navio ou sobreposição (o índice de Jaccard pode ser usado para) dos diferentes clusters detectados (se houver algo de cluster foi aplicado). Análise temporal (Garfield, 1994).

2. ANALISE BIBLIOMÉTRICA

O mapeamento da ciência, ou o mapeamento bibliométrico é um importante tópico no campo da ciência, onde utiliza métodos para estudar ou medir textos e informações de um grande conjunto de dados (Cobo et al., 2011). Basicamente é uma representação espacial de como as disciplinas, campos, especialidades, documentos ou autores estão relacionados um a outro baseados em indicadores de performance (Small, 1999).

Method	Description	Units of Analysis	Pros	Cons
Citation	Estimates influence of documents, authors, or journals through citation rates.	Document Author Journal	Can quickly find the important works in the field,	Newer publications had less time to be cited, therefore citation count as a measure of influence is biased toward older publications,
Co-citation	Connects documents, authors, or journals on the basis of joint appearances in reference lists.	Document Author Journal	It is the most used and validated bibliometric method. Connecting documents, authors, or journals with co-citation has been shown to be reliable. Since citation is a measure of influence, it offers a method to filter the most important works.	Co-citation is performed on cited articles so it is not optimal for mapping research fronts. Citations take time to accumulate, so new publications cannot be connected directly but only through knowledge base clusters. Several citations are needed to map articles, so it is impossible to map articles that are not cited much. When performing author co-citation analysis on SSCI (WOS) data, only first-author information is available.
Bibliographic Coupling	Connects documents, authors, or journals on the basis of the number of shared references.	Document Author Journal	Immediately available: does not require citations to accumulate. Can be used for new publications that are not cited yet, emerging fields, and smaller subfields.	It can only be used for limited timeframe (up to a five-year interval). It does not inherently identify the most important works by citation counts as co-citation; it is difficult to know whether mapped publications are important or not.
Co-author	Connects authors when they co-author the paper.	Author	Can provide evidence of collaboration and produce the social structure of the field.	Collaboration is not always acknowledged with co-authorship.
Co-word	Connects keywords when they appear in the same title, abstract, or keyword list.	Word	It uses the actual content of documents for analysis (other methods only use bibliographic meta-data).	Words can appear in different forms and can have different meanings.

Tabela 1 - Taxonomia das técnicas bibliométricas. Fonte: Zupic & Čater, (2015).

Uma análise de performance consiste em medir a contribuição relativa de temas para um campo de pesquisa especifico, enquanto o mapeamento cientifico vista focar na representação espacial de uma estrutura de temas, palavras e suas relações, isso se dá através do uso de várias técnicas como por exemplo a co-citação de documentos ou autores, e a análise de co-ocorrência de palavras (Cobo et al., 2011).

Zupic e Ater (2015) analisaram 81 trabalhos na área de *management* e *organization* que utilizam técnicas bibliométricas, com o objetivo em determinar quais métodos são mais utilizados por esta comunidade acadêmica, o resultado é apresentado na Figura 1, além disso o detalhamento dos métodos.

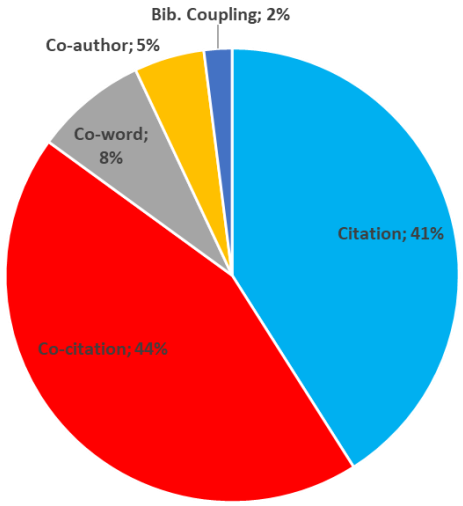


Figura 2 – Estratificação de métodos bibliométricos utilizados. Fonte: Zupic & Čater, (2015).

2.1. DETALHAMENTO DAS TÉCNICAS BIBLIOMÉTRICAS

Nesta seção serão apresentadas as técnicas bibliométricas mais utilizadas, bem como quais perguntas cada uma é capaz de responder.

Citation (Zupic & Čater, 2015) – Perguntas que pode ajudar a responder

Este método usa a citação como uma medida de **influência**, partindo do pressuposto que autores, trabalhos e periódicos mais citados são mais influentes.

- Autores que mais influenciaram a pesquisa em um período;
- Periódicos e disciplinas que tem mais impacto em uma vertente de pesquisa;
- Autores que são *experts* em uma área de pesquisa;
- O que ler de uma determinada área.

Co-Citation (Cobo et al., 2011) – Perguntas que pode ajudar a responder

Usa a citação a dois trabalhos como uma medida de **similaridade** entre trabalhos citados, autores e periódicos, partindo do pressuposto que quanto mais dois trabalhos são citados juntos, mais seu conteúdo está relacionado.

- Organizações centrais e periféricas em um determinado campo de pesquisa;
- Grupo de autores citados sistematicamente por um grupo determinado de trabalhos;
- Trabalhos que são referenciados conjuntamente.

Bibliographic Coupling (Cobo et al., 2011) – Perguntas que pode ajudar a responder

Usa o número de referências compartilhadas por dois trabalhos como uma medida de similaridade entre eles, partindo do pressuposto que quanto mais dois trabalhos citam trabalhos parecidos, mais seu conteúdo está relacionado.

- Organizações centrais e periféricas em um grupo de pesquisa emergente;
- Grupo de autores citados sistematicamente por um grupo determinado de trabalhos;
- Trabalhos que são referenciados conjuntamente.
  Obs: Faz sentido se os artigos pesquisados forem de um período temporal específico.

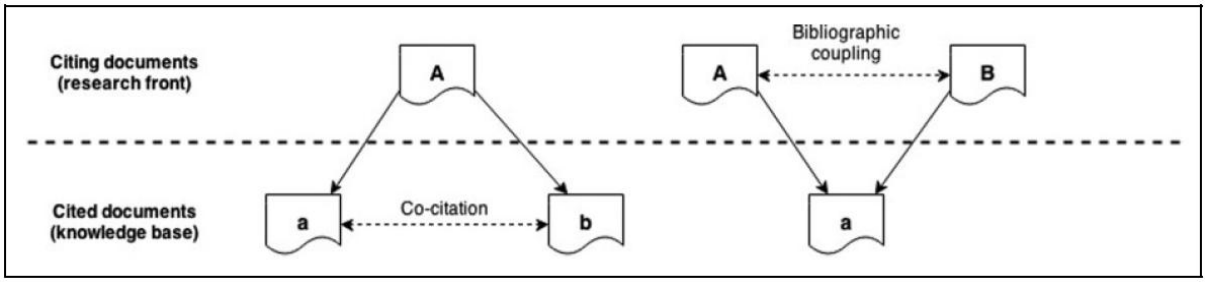


Figura 2 – Co-Citation vs Bibliogrraphic coupling. Fonte: Adaptado de Vogel &Guttel, 2013.

Co-Author (Zupic & Čater, 2015) – Perguntas que pode ajudar a responder

Usa uma medida de **colaboração** para examinar o aspecto social da pesquisa, ao invés do aspecto técnico. Pode ser utilizado para avaliar a colaboração em nível institucional e nacional. Uma relação entre dois autores é estabelecida quando publicam um artigo juntos.

- Autores que trabalham junto;

- Estrutura social de um campo de pesquisa (Bornmann et al., 2008);
- Instituições ou países que colaboram neste campo de pesquisa;
- Analisa as redes sociais que os cientistas criam colaborando em pesquisas cientificas;

Co-Word (Cobo et al., 2011) – Perguntas que pode ajudar a responder

É uma técnica de **análise de conteúdo** que usa as palavras nos documentos para estabelecer relacionamentos e construir uma estrutura conceitual do domínio, partindo do pressuposto que quando as palavras frequentemente co-ocorrem nos documentos, significa que os conceitos por trás dessas palavras estão intimamente relacionados

- Palavras-chave que vem sendo mais utilizadas em um determinado período de tempo;
- Palavras que são utilizadas em conjunto;
- Como o interesse de pesquisa mudou;
- Pode ser aplicado a título, resumo, palavras chave ou ao texto completo;

Co-Word (Cobo et al., 2011) – Limitações

- Com base na palavra chave:
- As bases não têm necessariamente as palavras chave corretamente definidas;
    - As palavras chave podem não representar adequadamente o tema.
- Com base no texto ou resumo:
    - Carrega ruído para análise.

Método (Unidades de Análise)	Pontos Fortes	Pontos Fracos
Citation (Trabalho, Autor, Periódico)	Acha facilmente trabalhos importantes de uma área	Novas publicações tem menos chance de serem consideradas importantes
Co-citation (Trabalho, Autor, Periódico)	Conectar trabalhos, autores e periódicos com co-citação é uma medida confiável	Não é ótimo para mapear "*reserach fronts* ", porque mapeia artigos citados
Bibliographic Couoling (Trabalho, Autor, Periódico)	Pode ser usado para achar áreas de pesquisas novas e sub áreas menores	Só pode ser utilizado em um *timeframe* de no máximo 5 anos
Co-Author (Autor)	Mostra a evidência de colaboração e a estrutura social do campo de pesquisa	Nem sempre a colaboração é apresentada com a co-autoria
Co-Word (Palavra)	Pode usar o conteúdo do artigo para análise	Palavras aparecem em diferentes formas e com diferentes significados

Tabela 2 – Tabela resumo das técnicas bibliométricas. Fonte: Elaborado pelo Autor.

2.2. SELEÇÃO DA BASE DE DADOS

As bases de dados do *Scopus* e *Web of Science* (WoS) são as mais utilizadas para buscar a literatura existente em um determinado campo de pesquisa, porém apesar de serem complementares apresentam ferramentas bem diferentes em termos de métodos e coberturas (HLWIKI INTERNATIONAL, 2015).

A base do WoS é atualizada semanalmente com mais de 30.000 artigos e 800.000 referências indexadas. Seu grande diferencial é a questão de abranger um ótimo período de tempo de cobertura e métodos de pesquisas únicos, porém não é considerada uma interface tão útil quando comparada ao *Scopus* (HLWIKI INTERNATIONAL, 2015).

A Elsevier tem a maior base de dados de citações que é o *Scopus*. Seu acervo abrange 16 milhões de autores, 25.000 trabalhos nas áreas de ciências sociais, físicas, saúde e tecnologia, com dados publicados desde 1995 (Elsevier B.V, 2020). A escolha desta base se dá principalmente pela maior representatividade, relevância e abrangência quando comparada ao WoS (Fahimnia et al., 2015).

Conclui-se então que uma pesquisa feita utilizando o *Scopus* como única base de dados contempla uma quantidade representativa de informações para se analisar um determinado campo da ciência.

2.3. SELEÇÃO DAS PALAVRAS-CHAVE

A primeira seleção de palavras-chave é selecionada com foco em serem suficientemente amplas para não restringir qualquer busca, e ao mesmo tempo serem específicas o suficiente para abordar apenas os temas desejados (Thomé et al., 2016).

A estratégia de pesquisa inicial contempla uma palavra e seus respectivos tesauros ou termos correspondentes. Foi selecionada a palavra-chave: *machine learning*. O motivo da escolha inicial ser uma única palavra é para criar um mapa onde pode-se analisar os principais focos de estudos de uma rede de trabalhos.

Figura 4 – Palavra-chave e seus respectivos tesauros. Fonte: Elaborado pelo Autor

A primeira estratégia resultou em um total de 685.792 trabalhos retornados, onde os trabalhos de Breiman (2001) e de Tamura (2011) são os mais citados.

A segunda busca, com 126.912 resultados encontrados, é adicionada a palavra-chave *Prediction* com seus tesauros. Esta adição fez com que a base de trabalhos encontradas fosse reduzida em aproximadamente 81% em relação a busca inicial, sendo os trabalhos mais citados são de Breiman (2001) e Haykin (2005).

Na terceira busca a pesquisa é direcionada para área de vendas, onde o termo *Sales* e seus correspondentes foram acrescidos a *string* de busca. Um total de 8.793 artigos foram encontrados. Os trabalhos mais citados são de Foley (2012) e Ngai (2009).

Na quarta busca o setor de bebidas foi adicionado com o termo *Beverage* como conceito principal. Um total de 44 trabalhos foram encontrados com uma crescente de publicação desde o ano de 2014, como pode ser observado abaixo. É importante ressaltar que nenhum autor da quarta pesquisa tem mais que 50 citações.

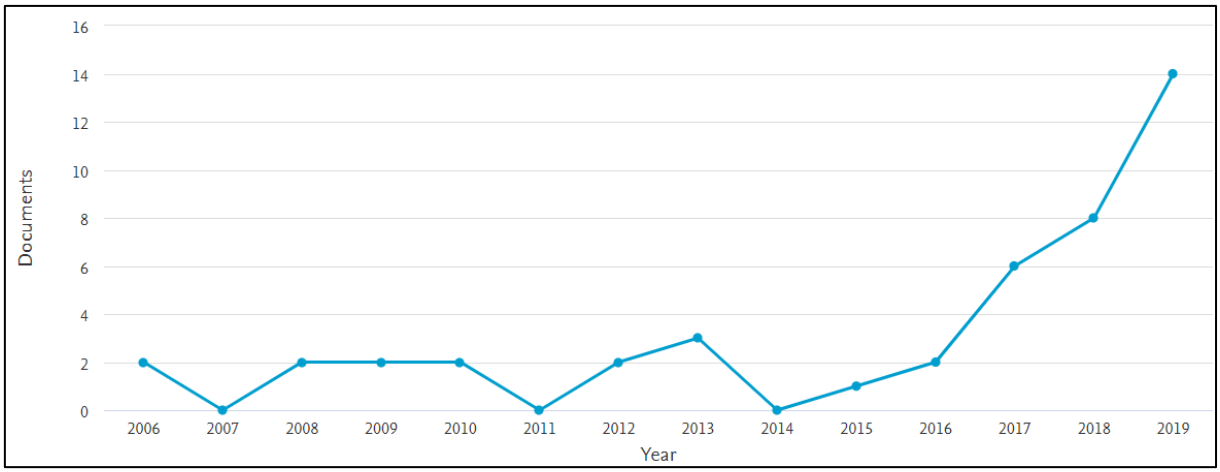


Figura 3 – Quantidade de documentos publicados por ano na quarta busca. Fonte: *Scopus*

Quadro 1 – Quadro resumo das buscas na base de dados *Scopus*. Fonte: Elaborado pelo autor.

	Palavras-chave	**Número de artigos**
1ª	TITLE-ABS-KEY ("Machine Learn*" OR "Data Mining" OR "Data-mining" OR "Deep Learn*" OR "Decision Tree" OR c45 OR "Random Forest" OR "Artificial Intel*" OR "Artificial Learn*") AND PUBYEAR < 2020	685.792
2ª	(TITLE-ABS-KEY ("Machine Learn*" OR "Data Mining" OR "Data-mining" OR "Deep Learn*" OR "Decision Tree" OR c45 OR "Random Forest" OR "Artificial Intel*" OR "Artificial Learn*") AND TITLE-ABS-KEY (predict* OR forecast*)) AND PUBYEAR < 2020	126.912
3ª	(TITLE-ABS-KEY ("Machine Learn*" OR "Data Mining" OR "Data-mining" OR "Deep Learn*" OR "Decision Tree" OR c45 OR "Random Forest" OR "Artificial Intel*" OR "Artificial Learn*") AND TITLE-ABS-KEY (predict* OR forecast*) AND TITLE-ABS-KEY (sales OR buying OR demand OR commercial OR trade)) AND PUBYEAR < 2020	8.793
4ª	(TITLE-ABS-KEY ("Machine Learn*" OR "Data Mining" OR "Data-mining" OR "Deep Learn*" OR "Decision Tree" OR c45 OR "Random Forest" OR "Artificial Intel*" OR "Artificial Learn*") AND TITLE-ABS-KEY (predict* OR forecast*) AND TITLE-ABS-KEY (sales OR buying OR demand OR commercial OR trade) AND TITLE-ABS-KEY (beverage OR liquor OR beer OR drink*)) AND PUBYEAR < 2020	44

Conceito	Aprendizado de Máquina	Previsão	Vendas	Bebida
Ingles	Machine Learning	Prediction	Sales	Beverage
Tesauros	Data Mining	Forecast	Buying	Liquor
	Deep Learning		Demand	Beer
	Decision Tree		Commercial	Drink
	C45		Trade	
	Random Forest			
	Artificial Intelligence			
	Artificial Learning			

Quadro 2 – Quadro das palavras-chave e seus tesauros. Fonte: Elaborado pelo autor.

2.4. SELEÇÃO DO SOFTWARE

Existem diversos *softwares* desenvolvidos especificamente para análise do mapeamento cientifico, assim como descrito em Cobo *et al.* (2012). São estes o Bibexcel, CiteSpace II, CoPalRed, IN-SPIRE, Loet Leydesdorff's software, Network Workbench Tool, Science of Science Tool, VantagePoint, VOSviewer e SciMAT.

Inicialmente os *softwares* escolhidos para este trabalho serão o VOSviewer e SciMAT. Ambos são de *download* gratuito e que incorporam diversos métodos, algoritmos e métricas de desempenho.

O *software* VOSviewer tem as seguintes vantagens em relação aos demais segundo Cobo *et al.* (2012):

- ✓ Ferramenta simples e de fácil entendimento;
- ✓ Ótimo para visualização de um mapa científico;
- ✓ Exige pouco tempo para o pré-processamento.

O SciMAT já é cosiderado uma ferramenta mais detalhada, onde se tem opção de escolha de diversos métodos e métricas. Suas principais qualidades segundo Cobo *et at.* (2012) são:

- ✓ Capacidade de escolha de unidade de análise (autores, palavras, etc) e medidas para desempenhar a análise;
- ✓ Possibilidade de alterar parâmetros durante os passos de mapeamento;
- ✓ Uso de métricas de desempenho;
- ✓ Forte base metodológica.

O presente trabalho fará a análise bibliométrica baseada em dois trabalhos que são complementares, Zupic & Čater (2015) e Cobo *et al.* (2011), com o propósito de analisar a evolução temática do tema abordado. Para isto ocorrerá uma análise de co-ocorrência de palavras e *Bibliographic Coupling* de autores mais citados.

2.5. ESTRATÉGIA UTILIZADA NA PESQUISA BIBLIOMÉTRICA

Para elaboração desta pesquisa é utilizado o método proposto por Cobo *et al*. (2011) que divide o mapeamento da ciência em oito passos apresentados na Figura 3. Além disto é utilizado software **SciMAT** (*Science Mapping Analysis Software Tool*)e **VOSviewer** proposto pelo autor por incorporar todos os passos, métodos e métricas para um mapeamento da ciência adequado, desde o pré-processamento até a visualização dos resultados (Cobo et al., 2012).

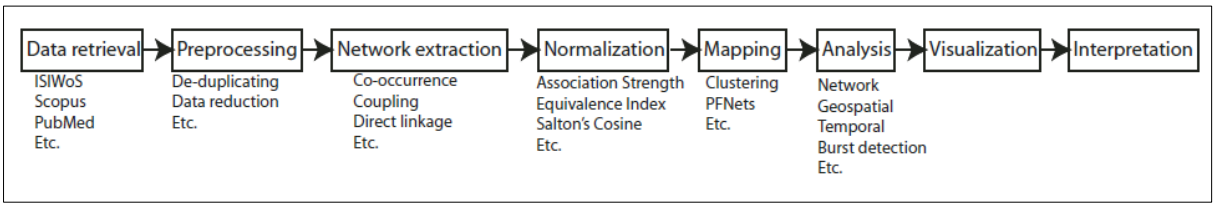


Figura 4 – *Workflow of Science Mapping*. Fonte: Cobo *et al. (2012)*.

Data retrieval – Busca de dados

Atualmente existem inúmeros banco de dados bibliométricos, onde documentos, trabalhos e citações são armazenados. Estas fontes de informações bibliométricas nos permite

pesquisar e recuperar informações sobre a maioria dos campos científicos. Os bancos de dados bibliográficos mais importantes são *Web of Science, Scopus, Google Scholar* e *Medline* (Cobo et al., 2011).

Preprocessing – Pré-processamento de dados

Os dados recuperados das fontes bibliográficas normalmente contêm erros, tais como, erros ortográficos no nome, no título da revista ou na lista de referências. Informações adicionais podem ser adicionadas aos dados originais, por exemplo, se o endereço do autor estiver incompleto ou errado (Cobo et al., 2011).

Por esse motivo, uma análise de mapeamento científico não pode ser aplicada diretamente aos dados recuperados das fontes bibliográficas, isto é, um processo de pré-processamento sobre os dados recuperados faz-se necessário. De fato, a etapa de pré-processamento é a mais importante para melhorar a qualidade das unidades de análise (principalmente autores e palavras) e, assim, obter melhores resulta na análise do mapeamento científico. Abaixo estão citados alguns processos de pré-processamento para se obter bons resultados.

1. O processo de intervalo de tempo é útil para dividir os dados em diferentes subperíodos de tempo, ou intervalos de tempo, para analisar a evolução do campo de pesquisa em estudo. Esse processo é necessário apenas se a análise do mapeamento científico for feita no contexto de um estudo longitudinal (Osinska & Malak, 2016);

2. A redução de dados visa selecionar os dados mais importantes. Normalmente, temos uma grande quantidade de dados. Com essa quantidade de dados, pode ser difícil obter resultados bons e claros na análise de mapeamento científico. Por esse motivo, é normalmente realizado usando uma parte dos dados. Essa parte pode ser, por exemplo, os artigos mais citados, os autores mais produtivos e os periódicos com as melhores métricas de desempenho (Cobo et al., 2011);

3. O pré-processamento de redes pode ser usado para selecionar os nós mais importantes da rede de relacionamentos entre as unidades de análise (redes

bibliométricas) de acordo com diferentes medidas, removendo os nós isolados, removendo os menos links importantes entre nós etc;

Network extraction – Extração da rede

Basicamente toma-se duas decisões nesta etapa, onde a primeira se dá pela escolha da unidade de análise, sendo elas: Autor, Periódicos, Trabalhos, etc. A segunda decisão está em selecionar a relação entre os nós, ou seja, o tipo de método bibliométrico que irá ser escolhido, sendo eles: *Citation, Co-Author, Co-Word, etc.*

Normalization – Normalização

Quando a rede de relacionamentos entre as unidades de análise é construída, uma transformação é primeiro aplicada aos dados para obter similaridades dos dados ou, mais especificamente, para normalizar os dados (Eck & Waltman, 2009).

Diferentes medidas de similaridade têm sido utilizadas na literatura, o mais popular é o *Cosine de Salton* (Salton & McGill, 1983), *Jaccard's Index* (Peters e van Raan, 1993), *Equivalence Index* (Callon, Courtial & Laville, 1991) e *Association Strenght* (Coulter, Monarch e Konda, 1998; van Eck & Waltman, 2007), também conhecido como *Proximity Index* (Peters & van Raan, 1993; Rip & Courtial, 1984) ou *Probabilistic Affinity Index* (Zitt, Bassecoulard e Okubo, 2000).

Mapping – Construindo o mapa

O uso do algoritmo *clustering* para detectar os temas: algoritmo dos centros simples (*simple centers algorithm*) de Cobo *et al.* (2011) que retorna automaticamente *clusters* etiquetados. Este algoritmo usa duas rodadas através dos dados para produzir as ligações desejadas. A primeira rodada constrói as ligações descrevendo as associações fortes, e as ligações adicionadas nesta rodada são chamadas ligações internas. A segunda rodada adiciona a essas redes, ligações de forças mais fracas que formam associações entre redes.

Analysis – Método de analise

Para análise da rede o **SciMat** utiliza a métrica de *Callon's density and centrality* (Cobo et al., 2011) como medida para determinar o grau de interação do *cluster* n com os outros *clusters.* Para tal existem duas medidas que são adotadas, a de centralidade que mede a interação entre *clusters* e a de densidade que mede a força e a coesão interna do *cluster*.

$$c = 10 * \Sigma e_{kh}$$

Equação 1 – Cálculo de centralidade de Callon. Fonte: (Cobo et al., 2011).

$$d = 100\frac{\Sigma e_{ij}}{w}$$

Equação 2 – Cálculo de densidade de Callon. Fonte: (Cobo et al., 2011).

Onde,

e_{kh}: $Índice\ de\ equivalência\ da\ palavra\ k\ do\ cluster\ com\ as\ palavras\ h\ do\ outro\ tema$;

e_{ij}: $Índice\ de\ equivalência\ das\ palavras\ i\ e\ j\ do\ cluster$;

w: $Número\ de\ palavras\ do\ cluster$.

O diagrama de Callon possui quatro quadrantes principais que são utilizados para classificação de um *cluster.*

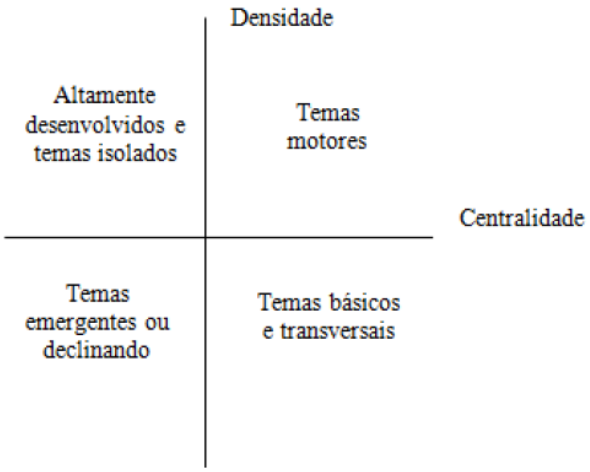


Figura 5 – Diagrama de Callon. Fonte: Adaptado de Cobo *et al*. *(2011)*.

Da figura 5, encontram-se, em sentido horário, primeiro quadrante com temas motores, que são considerados temas maduros, bem consolidados e importantes na comunidade científica. No segundo quadrante, os temas básicos e transversais são importantes, porém não estão totalmente desenvolvidos. O terceiro quadrante abraça temas que estão surgindo ou que estão desaparecendo. E por fim o quarto quadrante são temas muito específicos (normalmente periféricos na rede) e bem desenvolvidos.

Também neste passo é utilizado o *Inclusion Index (índice de inclusão)* que basicamente analise a evolução temática. Este índice é utilizado para medir a similaridade

entre temas e segundo Cobo *et al.* (2011) é o índice mais utilizado para pedir conjuntos similares.

Na evolução dos temas, figura 6, a linha sólida que une dois *clusters* representa temas muito conectados que compartilham do mesmo nome, já a linha tracejada são temas que compartilham o mesmo elemento, porém com nomes diferentes. O índice de inclusão é representado pela espessura da linha, onde para o valor de um considera-se que todas as palavras-chave do período 1 estão contidas no período 2. O volume das esferas é proporcional a quantidade de documentos publicados ligados ao tema (Cobo et al., 2011).

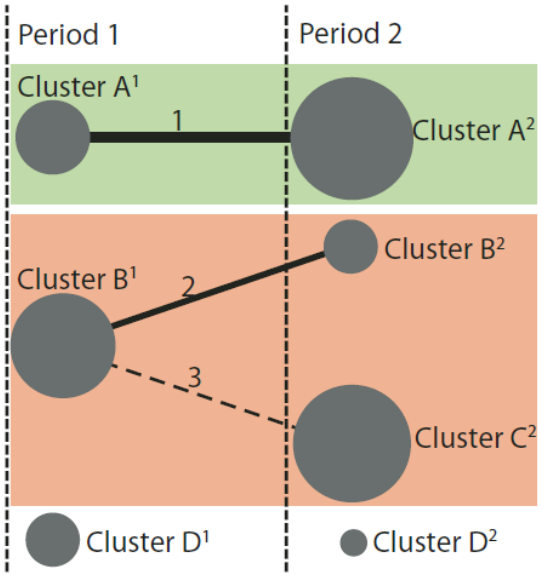


Figura 6 – Evolução temática. Fonte: Cobo *et al*. *(2011)*.

Outra visualização que é gerada neste passo é o gráfico de sobreposição (figura 7), que tem como objetivo medir o número de palavras-chave que são compartilhadas entre os períodos selecionados.

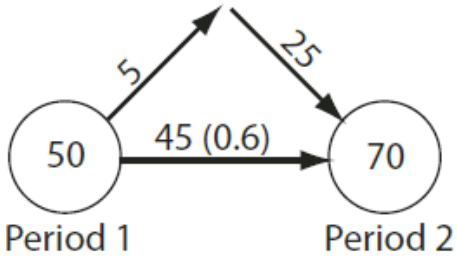


Figura 7 – Gráfico de sobreposição de palavras-chave. Fonte: Cobo *et al*. *(2011)*.

Cada círculo representa um período com seus respectivos números de palavras-chave associadas. A linha horizontal representa o número de palavras-chave que são compartilhadas em ambos os períodos, e entre parênteses, a proporção de palavras que passaram para o período seguinte (na figura 7 60% de 45 palavras-chave passaram para o período 2). A linha entrando em cada círculo (período de tempo) constitui novas palavras-chave e a linha saindo o número de palavras-chave que está presente no período atual, porém não estará no próximo (Cobo et al., 2011).

2.6. DEFINIÇÃO DOS PARÂMETROS

Nesta seção os parâmetros utilizados no VOSviewer e SciMAT serão abordados.

Para se obter uma melhor visualização da rede e mapeamento dos temas centrais e periféricos, no primeiro momento utiliza-se do VOSviewer para construção do mapa científico.

A calibração do VOSviewer consiste em 10 passos:

1. Definição da unidade de análise e técnica bibliométrica: a unidade de análise serão as palavras com a técnica de co-ocorrência;
2. Fonte de dados: leitura a partir da extração de arquivo (.csv) da base *Scopus* com **utilização da primeira busca como estratégia de busca (685.792 artigos)**;
3. Importação de arquivo: selecionado arquivo (.csv);
4. Seleção dos campos utilizados: títulos e abstratos;
5. Método de contagem: contagem binária;
6. Importação do arquivo de tesauros;
7. Definição do número mínimo de ocorrência de um termo para entrar na rede: 10;
8. Definição do número de termos a serem exportados para o mapa: 60%;
9. Verificação e seleção das palavras que serão utilizadas;
10. Execução do algoritmo e análises: etapa em que se gera o mapa para análise.

A calibração do SciMAT consiste em 12 passos:

1. Fonte de dados: leitura a partir da extração de arquivo (.csv) da base *Scopus* com **utilização da terceira busca como estratégia de busca (8.793 artigos)**;
2. Seleção dos períodos: todos os períodos foram selecionados;
3. Unidade de análise: palavras, da fonte e adicionadas;
4. Redução de dados: não houve redução de dados;
5. Técnica bibliométrica: co-ocorrência;
6. Redução da rede: calibrou-se o valor para ocorrência mínima de 2 para todos os períodos;
7. Medida de similaridade para normalização da rede: índice de equivalência;
8. Seleção do algoritmo de *clustering:* o algoritmo utilizado foi o de centro simples com tamanho máximo da rede de 110 e mínimo de 2;
9. Seleção do mapeador de documentos usado na análise de performance: o *core mapper* deve ser selecionado se tem no mínimo duas palavras-chave na rede e o *secondary mapper,* se tem somente uma palavra-chave associada com a rede temática.
10. Seleção de medidas de desempenho e qualidade bibliométrica: para Cobo *et al.* (2011), a medida de desempenho *h-index* é confiável para analisar o impacto bibliométrico dos temas e áreas temáticas. Além disso como medida básica foi selecionado a *Sum citations,* que mede a soma das citações da unidade de análise;
11. Seleção da medida de similaridade: selecionado índice de inclusão para ambos;
12. Execução das análises: etapa de conclusão do processo de definição de parâmetros.

3. RESULTADOS E DISCUSSÃO

3.1. ESTATÍSTICAS DESCRITIVAS

A estratégia de busca considerada para esta análise foi a busca que resultou em 8.793 artigos. Através da figura 8 é possível notar que o número de publicações sobre o tema de previsão de venda com a utilização de inteligência artificial vem em uma crescente, principalmente após o ano de 2015.

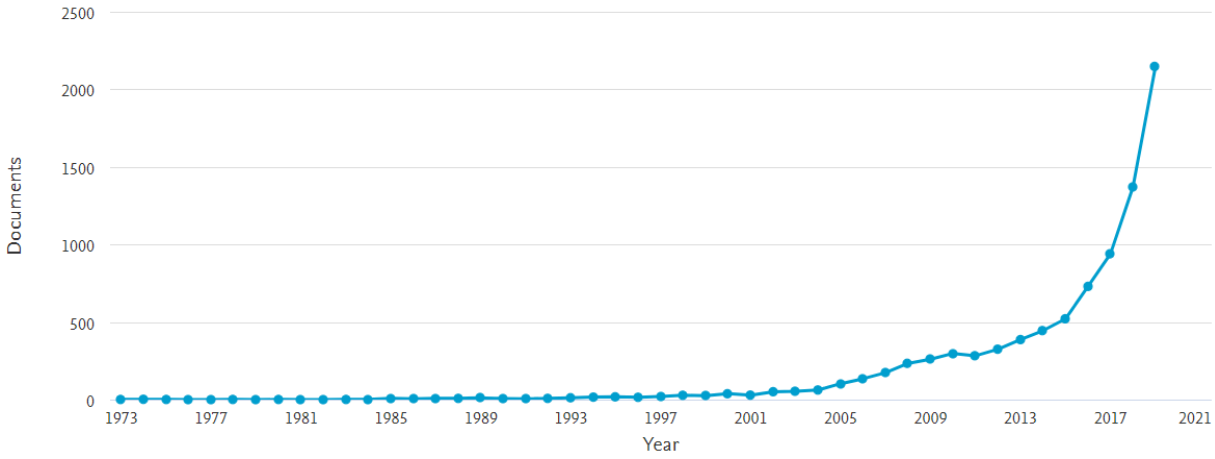


Figura 8 – Gráfico de quantidade de publicação por ano. Fonte: *Scopus*.

A figura 9 representa os dez autores com mais publicações na base *Scopus*, eles representem 1,5% do total de publicações, fato é que como este é um tema emergente a quantidade de autores tende a aumentar cada vez mais, com um número real de 6255 autores diferentes publicando na base.

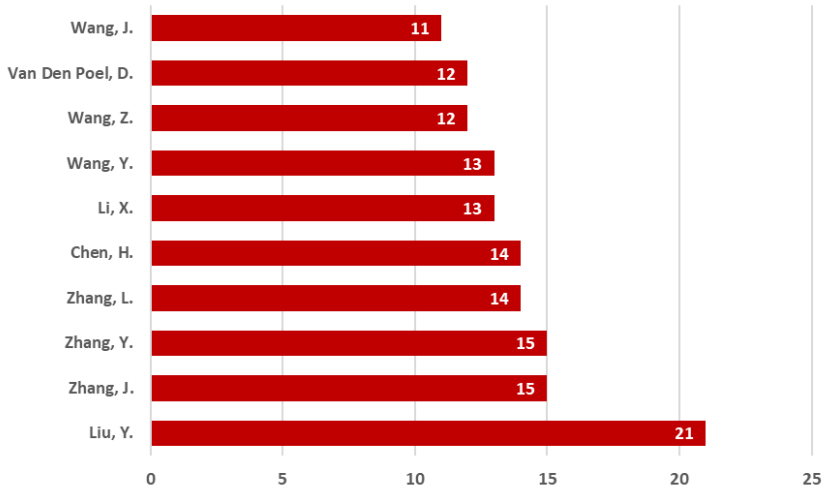


Figura 9 – Autores com maior número de documentos publicados. Fonte: *Scopus*.

A figura 10 representa os dez jornais que mais publicaram e que representam 20% do total de publicações (1404). Em temas não emergentes existe uma concentração alta de artigos publicados em um número reduzido de periódicos, fato que não ocorre nesta análise por se tratar de um tema emergente.

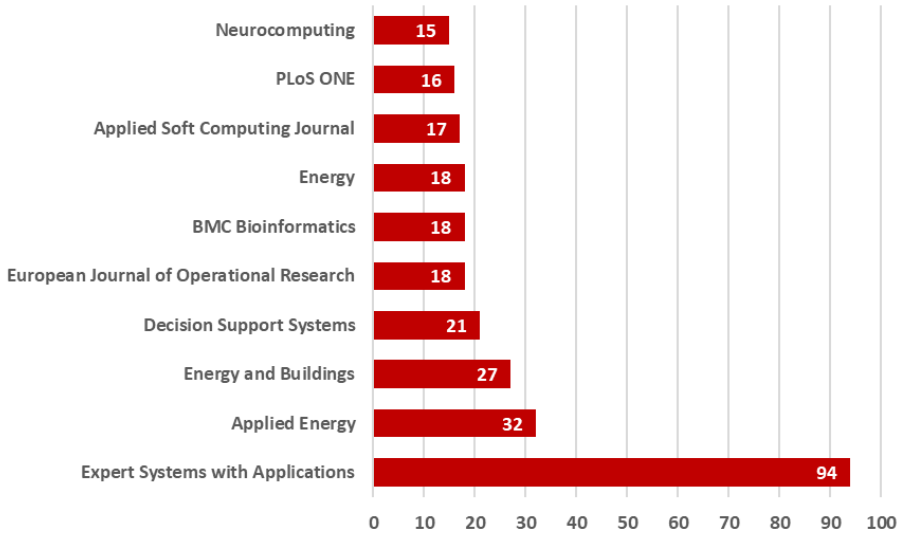


Figura 10 – Jornais com maior número de documentos publicados. Fonte: *Scopus*.

Na figura 11, observam-se os dez países com maior número de publicações que representam 75% do total (6630) de publicações no campo de pesquisa. Desses países, os Estados Unidos e a China representam 40% das publicações.

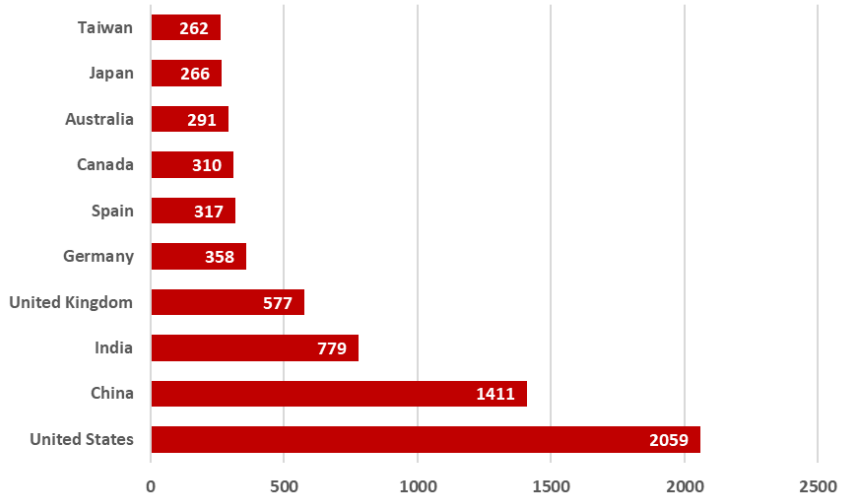


Figura 11 – Países com maior número de documentos publicados. Fonte: *Scopus*.

Na tabela 3 são apresentados os dez artigos mais citados, nota-se que não há autores repetidos nesta lista.

Item	Artigo	Autor	Número de citações	Ano Publicação
1	Current methods and advances in forecasting of wind power generation	Foley, A.M., Leahy, P.G., Marvuglia, A., McKeogh, E.J.	624	2012
2	Application of data mining techniques in customer relationship management: A literature review and classification	Ngai, E.W.T., Xiu, L., Chau, D.C.K.	593	2009
3	Discretization: An enabling technique	Liu, H., Hussain, F., Tan, C.L., Dash, M.	587	2002
4	Process mining: Data science in action	Van der Aalst, W.	556	2016
5	Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics	Ghose, A., Ipeirotis, P.G.	548	2011
6	Knowledge graph embedding by translating on hyperplanes	Wang, Z., Zhang, J., Feng, J., Chen, Z.	460	2014
7	A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip	Chang, R.W., Pellissier, J.M., Hazen, G.B.	402	1996
8	Multiobjective intelligent energy management for a microgrid	Chaouachi, A., Kamel, R.M., Andoulsi, R., Nagasaka, K.	359	2013
9	Empirical prediction models for adaptive resource provisioning in the cloud	Islam, S., Keung, J., Lee, K., Liu, A.	354	2012
10	Vital nodes identification in complex networks	Lü, L., Chen, D., Ren, X.-L., (...), Zhang, Y.-C., Zhou, T.	353	2016
		Total de citações:	4836	

Tabela 2 – Tabela com as 10 publicações mais citadas. Fonte: *Scopus.*

3.2. ANÁLISE NÃO DESCRITIVA

Foram identificados com o VOSviewer 397 termos com 27.150 ligações entre eles. Estes foram agrupados em forma de rede, com o objetivo de melhorar a visualização da conexão entre os termos apresentados. Nota-se cinco cores diferentes de grupos de termos, o software utiliza-se destas cores para diferenciar *clusters* de pesquisa diferentes.

Zupic & Čater (2015) considera estes *clusters* como possíveis diferentes grupos de pesquisa consolidados. Na análise nota-se que três dos cinco grupos são mais dominantes e com termos centrais bem definidos, são eles representados pela cor vermelha (*cluster* 1), verde (*cluster* 2) e azul (*cluster* 3). Os demais grupos como amarelo (*cluster* 4) e roxo (*cluster* 5) são menores e sem termos centralizados.

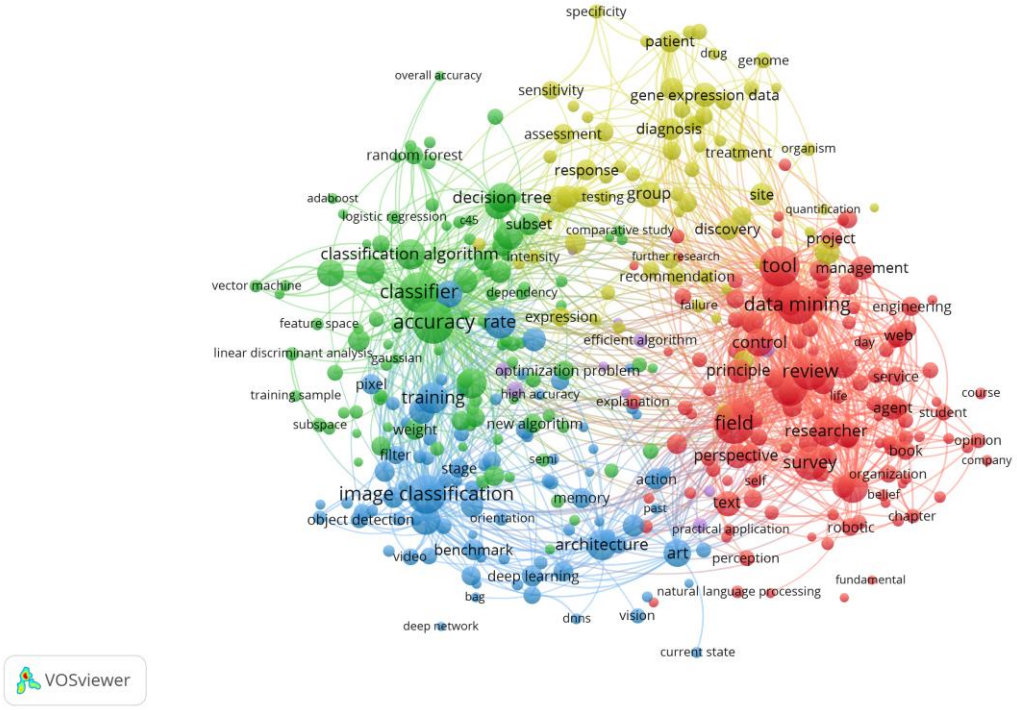


Figura 12 – Rede de palavras com método de co-ocorrência de palavras. Fonte: VOSViwer

A visualização do termos centrais é de suma importância para mapear os temas chave da pesquisa em um determinado grupo de autores em um período de tempo, porém para explorar temas emergentes é necessária a visualização da periferia da rede bibliométrica (van Eck & Waltman, 2010).

O *cluster 1* tem um total de 127 termos em sua totalidade, tendo palavras centrais como *data mining, engeneering, development, organization.* Na análise periférica obtém-se termos como *management, twitter, business, robotic, industry, project,* etc. Isso nos leva a crer que grande parte do esforço da comunidade acadêmica está concentrada na área de engenharia, desenvolvimento e administração de organizações. Já nos temas emergentes a aplicação de inteligência artificial em assuntos de gerenciamento de projetos e robótica.

O *cluster 2* tem um total de 106 termos em sua totalidade com palavras centrais como *image classification, face recognition,* com links muito fortes com algoritmos de *deep learning.* Em sua periferia palavras como *deep neural network, convolutional neural network, pixel e object detection* surgem como as principais. Neste campo de pesquisa os autores estão focados em aplicação de inteligência artificial para detecção de imagem utilizando métodos

emergentes como rede convolucional, que é um método criado com base no funcionamento dos neurônios para reconhecimento de linhas, curvas e bordas.

No *cluster 3* existem 87 termos no total, com palavras centrais como *classifier, decision tree* e *support vector machine, random forest logistic regression* em sua periferia. Nota-se uma forte correlação do *cluster 2 e 3* principalmente pelo fato dos algoritmos contidos no *cluster 3* serem muito utilizados para detecção de imagem. Isto pode ser percebido na visualização mais detalhada da rede na Figura 6.

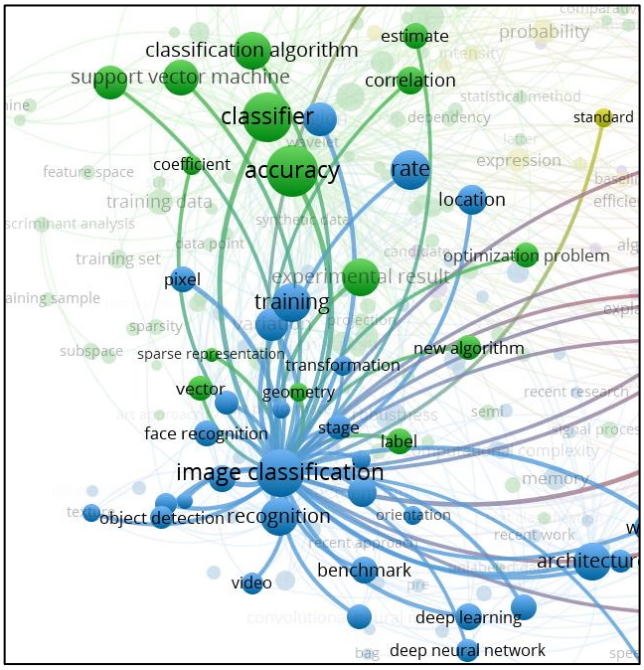


Figura 13 – Relação entre *cluster 2 e cluster 3*. Fonte: VOSViwer

O *cluster 4* tem um total de 67 termos em sua totalidade sem a presença de termos centrais nesta rede. Para Zupic & Čater (2015) a ausência de termos centrais sugere que uma rede seja emergente, visto que ainda não há citações de palavras suficientes para tornar esta rede madura no âmbito da ciência. Existem diversas palavras como *cancer, genome, diagnosis, organism, patient, group*. A rede tem aplicação de inteligência artificial em áreas da medicina como, mapeamento genético (Depristo et al., 2011), diagnostico de pacientes (Fiehn, 2002) e mapeamento de sequência genética de pacientes com câncer (Huang et al., 2009).

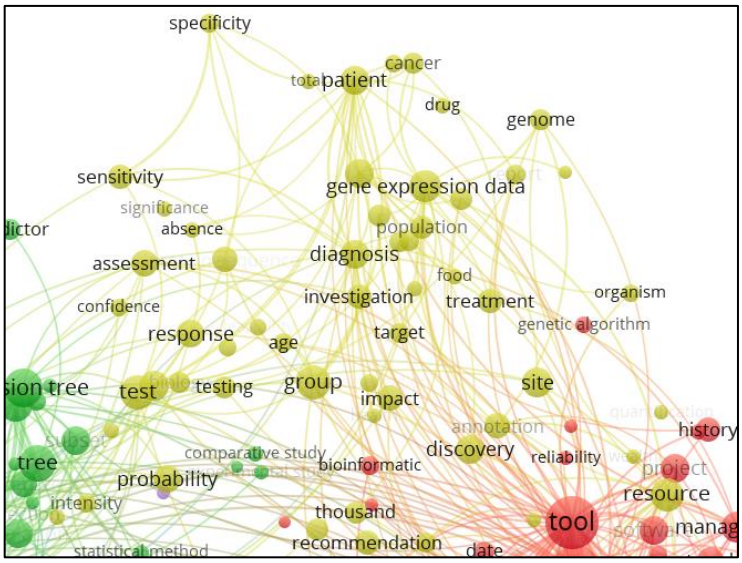


Figura 14 – Análise detalhada *cluster 4*. Fonte: VOSViwer

O *cluster 5* tem um total de 10 itens em sua totalidade sem a presença de termos centrais nesta rede. Seus termos são pouco relacionados com as demais e entre si, portanto será desconsiderado para este estudo por ter baixa relevância quando comparado aos demais termos bibliográficos da rede.

Na figura 15, observa-se que houve pouco crescimento do número de palavras-chave nos períodos selecionados. Além disso 68% das palavras-chave do primeiro período se mantiveram no segundo, e 80% do segundo período se mantiveram no terceiro. Este fator é um indicativo que apesar do crescimento em número de publicações, o tema está relativamente maduro considerando este campo de pesquisa. Outro fato que nos leva a esta conclusão a é a baixa quantidade de palavras-chave que são introduzidas em cada período.

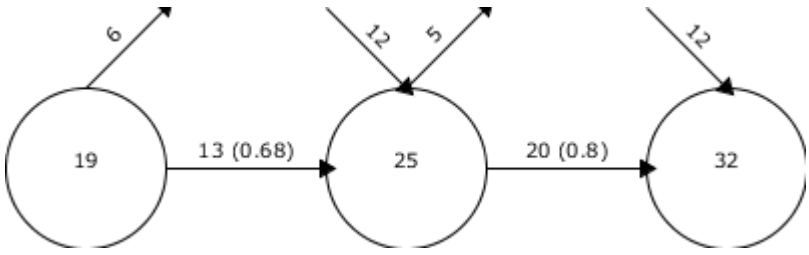


Figura 15 – Gráfico de sobreposição. Fonte: SciMAT

Na figura abaixo observa-se a evolução temática do campo de pesquisa ao longo dos três períodos selecionados. No período de 2005-2010 o tema com maior número de publicação foi *supervised machine learning*, e que os outros *clusters* estão uniformemente distribuídos.

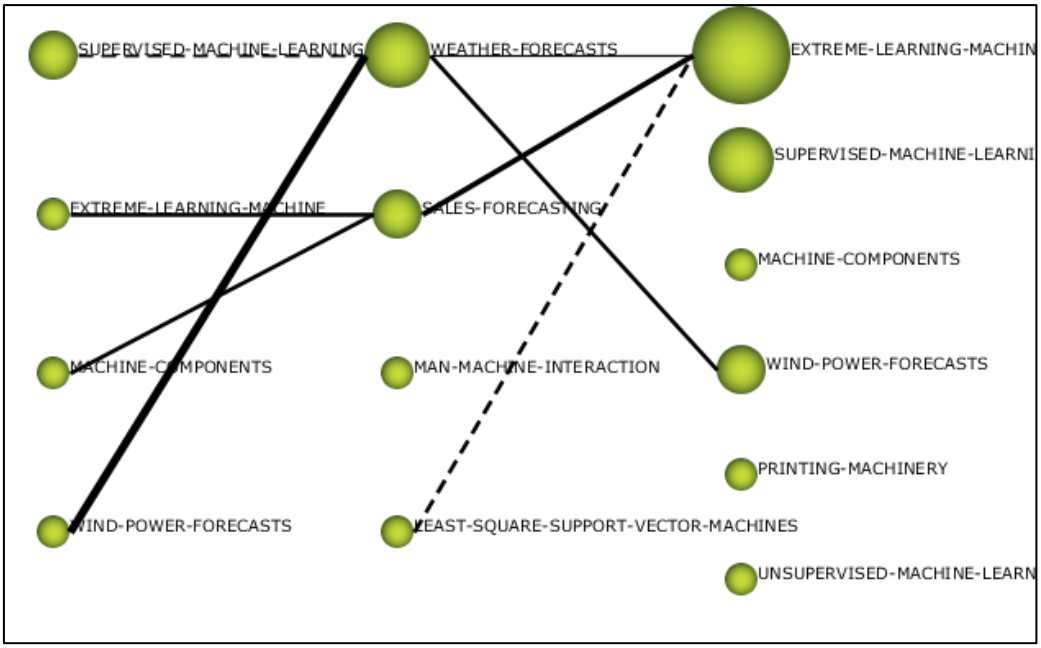


Figura 16 – Evolução temática. Fonte: SciMAT

Algo similar ocorre no período de 2011-2015, onde o tema de previsão do clima foi o que obteve maior número de publicações. Nota-se também que previsão de vendas é o segundo *cluster* com mais publicação, e tem forte relação direta com um *cluster* do próximo período.

No período de 2016-2019 houve uma grande quantidade de publicações no tema relacionado a *extreme learning machine*, e este com forte ligação a previsão de vendas do período passado.

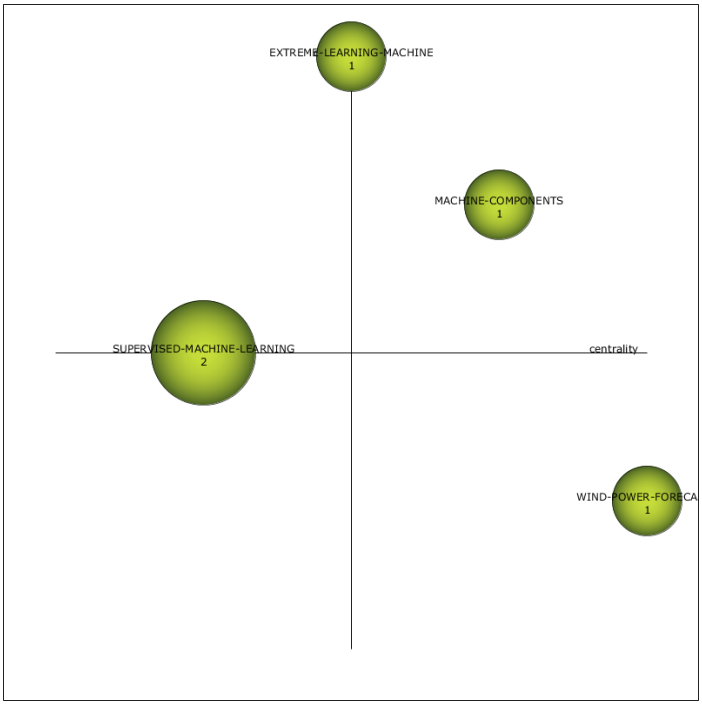


Figura 17 – Diagrama estratégico do período de 2005-2010. Fonte: SciMAT

A figura 17 contempla o diagrama estratégico do período de 2005-2010. Neste gráfico pode ser verificado que *supervised machine learning* é um tema com alta centralidade e baixa medida de densidade, porém com um grande número de publicações quando comparado a outros temas.

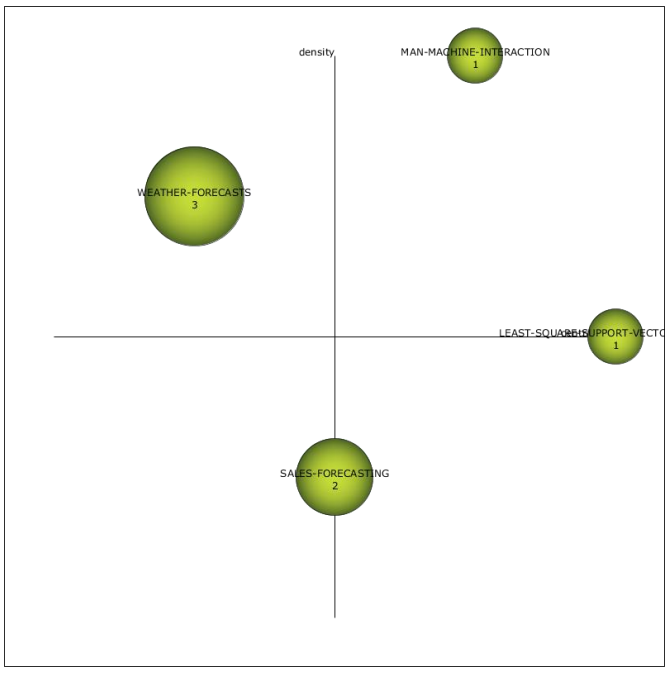


Figura 18 – Diagrama estratégico do período de 2011-2015. Fonte: SciMAT

A figura 18 contempla o diagrama estratégico do período de 2011-2015. Neste gráfico pode ser verificado que *weather forecasting* é um tema especializado e periférico na rede com média centralidade e média densidade, porém com um grande número de publicações juntamente com *sales forecasting.*

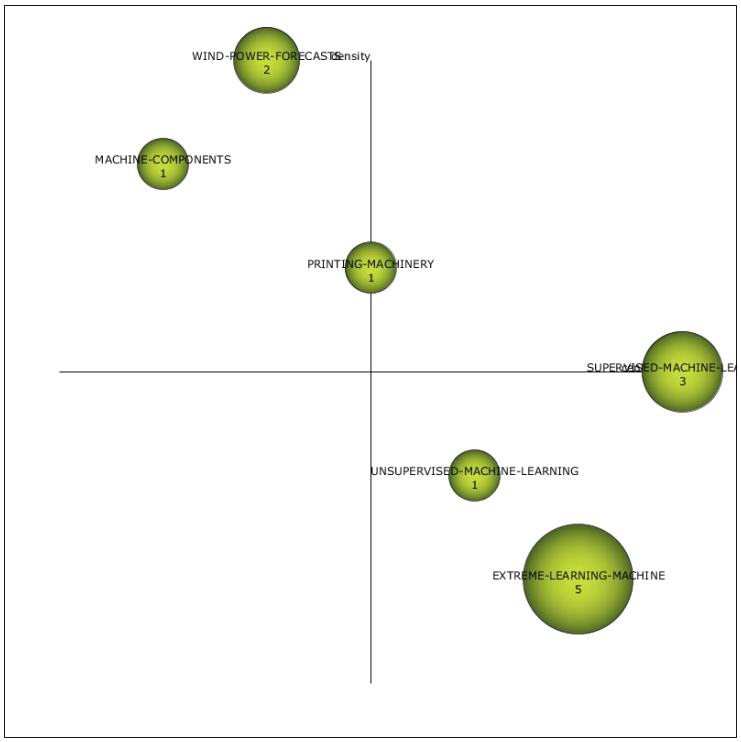


Figura 19 – Diagrama estratégico do período de 2016-2019. Fonte: SciMAT

A figura 19 contempla o diagrama estratégico do período de 2016-2019. Neste gráfico pode ser verificado que *extreme learning machine* é considerado um tema básico e transversal na rede com baixa densidade e alta centralidade, porém com um grande número de publicações, comparado a outros *clusters* da rede.

Referências Artigo 1

Bakker, B., Linaker, F., & Schmidhuber, J. (2002). Reinforcement learning in partially observable mobile robot domains using unsupervised event extraction. *IEEE/RSJ International Conference on Intelligent Robots and System*, *1*, 938–943. https://doi.org/10.1109/IRDS.2002.1041511

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Białkowski, J., Bohl, M. T., Stephan, P. M., & Wisniewski, T. P. (2015). The gold price in times of crisis. *International Review of Financial Analysis*, *41*, 329–339. https://doi.org/10.1016/j.irfa.2014.07.001

Carrion-i-Silvestre, J. L., Sansó-i-Rosselló, A., & Ortuño, M. A. (2001). Unit root and stationarity tests' wedding. *Economics Letters*, *70*(1), 1–8. https://doi.org/10.1016/S0165-1765(00)00348-7

Chang, T. (2002). An econometric test of Wagner's law for six countries based on cointegration and error-correction modelling techniques. *Applied Economics*, *34*(9), 1157–1169. https://doi.org/10.1080/00036840110074132

Chen, S., Mihara, K., & Wen, J. (2018). Time series prediction of CO2, TVOC and HCHO based on machine learning at different sampling points. *Building and Environment*, *146*, 238–246. https://doi.org/10.1016/j.buildenv.2018.09.054

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Data vs Instinct: Perfecting Global Sales Performance. (2012). 25.

Dewes, R., Viero, C. F., & de Lima Nunes, F. (2018). Dimensionamento de estoques: Uma análise em uma empresa varejista de peças em alumínio. *Revista Liberato*, *19*(31), 117–131. https://doi.org/10.31514/rliberato.2018v19n31.p117

Gers, F. A., Eck, D., & Schmidhuber, J. (2001). *Applying LSTM to Time Series Predictable Through Time-Window Approaches*. 8.

Giraitis, L., Kokoszka, P., Leipus, R., & Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, *112*(2), 265–294. https://doi.org/10.1016/S0304-4076(02)00197-5

Gonzalez-Dominguez, J., Lopez-Moreno, I., Sak, H., Gonzalez-Rodriguez, J., & Moreno, P. J. (2014). *Automatic Language Identification using Long Short-Term Memory Recurrent Neural Networks*. 5.

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, *23*(2), 201–220. https://doi.org/10.1109/JSAC.2004.839380

Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. 1735–1780.

Hussain, M., & Mahmud, I. (2019). pyMannKendall: A python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software*, *4*(39), 1556. https://doi.org/10.21105/joss.01556

Lazarini, J. (2019). *IBGE*. https://www.ibge.gov.br/estatisticas/economicaS

Libiseller, C., & Grimvall, A. (2002). Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. *Environmetrics*, *13*(1), 71–84. https://doi.org/10.1002/env.507

Ling, C. H., Ahmed, K., Binti Muhamad, R., & Shahbaz, M. (2015). Decomposing the trade-environment nexus for Malaysia: What do the technique, scale, composition, and

comparative advantage effect indicate? *Environmental Science and Pollution Research*, *22*(24), 20131–20142. https://doi.org/10.1007/s11356-015-5217-9

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Portal Action. (2019). http://www.portalaction.com.br/series-temporais/11-estacionariedade

Qi, C., & Tang, X. (2018). Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Computers & Industrial Engineering*, *118*, 112–122. https://doi.org/10.1016/j.cie.2018.02.028

SEBRAE. (2019). Comunidade Sebrae. https://comunidadesebrae.com.br/blog/passo-a-passo-para-implementar-um-plano-de-negocios

Shi, H., Xu, M., & Li, R. (2018). Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. *IEEE Transactions on Smart Grid*, *9*(5), 5271–5280. https://doi.org/10.1109/TSG.2017.2686012

SINDICERV. (2020). https://www.sindicerv.com.br/o-setor-em-numeros/

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799–813. https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Taylor, S. J., & Letham, B. (2017). *Forecasting at scale* [Preprint]. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3190v2

Thomé, A. M. T., Scavarda, A., Ceryno, P. S., & Remmen, A. (2016). Sustainable new product development: A longitudinal review. *Clean Technologies and Environmental Policy*, *18*(7), 2195–2208. https://doi.org/10.1007/s10098-016-1166-3

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Wang, H., Xu, H., Yuan, Y., Sun, X., & Deng, J. (2019). Balancing exploration and exploitation in multiobjective batch bayesian optimization. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 237–238. https://doi.org/10.1145/3319619.3321962

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, *93*, 41–56. https://doi.org/10.1016/j.energy.2015.08.045

Wöllmer, M., Marchi, E., Squartini, S., & Schuller, B. (2011). Multi-stream LSTM-HMM decoding and histogram equalization for noise robust keyword spotting. *Cognitive Neurodynamics*, *5*(3), 253–264. https://doi.org/10.1007/s11571-011-9166-9

Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629

Garfield, E. (1994). Scientography: Mapping the tracks of science. Current Contents: Social & Behavioural Sciences, 7(45), 5–10

Skillicorn, D. (2007). Understanding complex datasets: Data mining with matrix decompositions (Chapman & Hall/Crc Data Mining and Knowl- edge Discovery Series). London: Chapman & Hall

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge, UK: Cambridge University Press

Cook, D.J., & Holder, L.B. (2006). Mining graph data. Hoboken, NJ: John Wiley & Sons, Inc

McCain, K. (1991). Mapping economics through the journal literature: An experiment in journal co-citation analysis. Journal of theAmerican Society for Information Science, 42(4), 290–296

Carrington, P.J., Scott, J., & Wasserman, S. (2005). Models and methods in social network analysis. Structural Analysis in the Social Sciences. New York: Cambridge University Press

van Eck, N.J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. Journal of the American Society for Information Science and Technology, 60(8), 1635– 1651

Gao, X., & Guan, J. (2009). Networks of scientific journals: An exploration of Chinese patent data. Scientometrics, 80(1), 283–302

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24(4), 265–269.

Small, H. (1997). Update on science mapping: Creating large document spaces. Scientometrics, 38(2), 275–293.

Small, H. (1999). Visualizing science by citation mapping. Journal of the American Society for Information Science, 50(9), 799–813.

Small, H. (2006). Tracking and predicting growth areas in science. Sciento- metrics, 68(3), 595–610.

Small, H., & Garfield, E. (1985). The geography of science: Disciplinary and national mappings. Journal of Information Science, 11(4), 147–159.

Small, H., & Koenig, M.E.D. (1977). Journal clustering using a bibliographic coupling method. Information Processing and Management, 13(5), 277– 288.

Small, H., & Upham, S.P. (2009). Citation structure of an emerging research area on the verge of application. Scientometrics, 79(2), 365–37

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA
E TECNOLOGIA FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À ENGENHARIA E GESTÃO

DANIEL NOCERA DE CAMPOS

ARTIGO 2: ANÁLISE DO BANCO DE DADOS PARA APLICAÇÃO DE ALGORITMO DE APRENDIZAGEM DE MÁQUINA APLICADO A PREVISÃO DE VENDAS

Campos dos Goytacazes/RJ

(2020)

ARTIGO 2: ANÁLISE DO BANCO DE DADOS PARA APLICAÇÃO DE ALGORITMO DE APRENDIZAGEM DE MÁQUINA APLICADO A PREVISÃO DE VENDAS

RESUMO

A aplicação de modelos para previsão de vendas futuras é um processo que vem sendo aplicado cada vez mais pelas empresas, baseado principalmente em dados de vendas históricas e demais dados complementares. Especificamente no setor de varejo uma boa previsão de vendas é utilizada para dimensionamento de estoque, gerenciamento da cadeia de suprimentos, e dimensionamento de equipes, influenciando diretamente no seu lucro fnal e automaticamente em sua vantagem competitiva no mercado empresarial. É fundamental que se tenha dados de extrema qualidade, uma vez que define a boa aplicabildiade do modelo de *machine learning*, são os dados históricos, portanto o objetivo deste trabalho é realizar analises no banco de dados com foco principal na verificação da qualidade das informações e analisar sua aplicabilidade a modelos de predição de séries temporais já existentes. Com a aplicação desta metodologia foi possível verificar a estacionaridade do banco de dados, sua decomposição em séries temporais e analise de sazonalidade e complementarmente a visualização de sua tendência.

Palavras-chave: **Aprendizado de Máquina, Vendas no Setor de Bebidas, Banco de dados.**

ABSTRACT

The application of models for forecasting future sales is a process that has been applied more and more by companies, based mainly on historical sales data and other complementary data. Specifically in the retail sector, a good sales forecast is used for inventory sizing, supply chain management, and team sizing, directly influencing your final profit and automatically your competitive advantage in the business market. It is essential to have data of extreme quality, since it defines the good applicability of the machine learning model, it is historical data, so the objective of this work is to carry out analyzes in the database with the main focus on verifying the quality of information and analyze its applicability to prediction models of existing time series. With the application of this methodology, it was possible to verify the database's stationarity, its decomposition in time series and seasonality analysis and complementarily the visualization of its trend..

Key-Words: Machine Learning, Sales in the Beverage Sector, database

1. INTRODUÇÃO

A previsão de vendas é um processo para estimar as vendas futuras de acordo com as vendas históricas e outras informações relevantes, o que permite que uma empresa preveja negócios de curto e longo prazo e tome as decisões correspondentes. No setor de varejo, a previsão de vendas é importante para o controle de estoque, o gerenciamento da cadeia de suprimentos, o reabastecimento etc.

Uma boa previsão que atenda às demandas de consumo pode ajudar os varejistas a aumentar os lucros, promover produtos com relação aos padrões de consumo e controlar o estoque de segurança sem ser excessivo. No entanto, metodologias de previsão de vendas são tipicamente mais científicas do que um gráfico intuitivo, embora diferentes modelos de previsão estejam disponíveis em vários softwares, os executivos de previsão nas empresas não correm o risco de ajustar os parâmetros para uma previsão potencialmente melhor (*Data vs Instinct: Perfecting Global Sales Performance*, 2012).

Por outro lado, o aprendizado de máquina se desenvolve dramaticamente nos últimos anos e contribui em ambos campos da indústria e a vida cotidiana das pessoas, com configurações de parâmetros relativamente mais simples como alternativa aos métodos tradicionais baseados em modelos. Um grande ramo do aprendizado de máquina são as redes neurais com décadas de desenvolvimento e muitas variações. A rede neural de memória de longa duração (LSTM) é um tipo de rede neural recorrente (RNN) proposta em 1997 para abordar o problema do refluxo de erro em decomposição insuficiente no treinamento da RNN (Hochreiter & Schmidhuber, 1997).

No entanto, os interesses de pesquisa em LSTM têm aumentado lentamente até 2014, como mostra a Figura 1. As publicações são aumentadas drasticamente junto com a aplicação do LSTM no processamento de linguagem natural por empresas conhecidas como o Google (Gonzalez-Dominguez et al., 2014). Desde então, o LSTM tem sido aplicado campos internos, incluindo controle de robôs (Bakker et al., 2002), reconhecimento de fala (Gers et al., 2001), previsão de séries temporais (Wöllmer et al., 2011), etc. Devido à estrutura especial, um LSTM é capaz de aprender com informações históricas para classificar, processar e prever séries temporais e refletir eventos importantes. Este artigo testa uma rede LSTM com dados reais de vendas de 66 produtos coletados de um varejista durante 45 semanas. Como um experimento geral, sazonalidade e promoções na prática não estão sendo consideradas. Os resultados refletem o potencial da implementação da rede LSTM nas previsões de vendas no varejo.

Neste artigo foi realizado um estudo detalhado do banco de dados utilizado para construção do modelo de inteligência artificial. Tal estudo constitui uma análise de tendência, sazonalidade, resíduo, interdependência das variáveis, importância de períodos futuros, dentre outros. Para Barboza *et al*. (2017) uma boa análise do banco de dados é fundamental para verificar se através de seus dados é gerado um modelo de alta confiabilidade e que retrate algo bem próximo a realidade do problema em questão.

O banco de dados é composto por dados de vendas de cerveja e refrigerante do ano de 2015 até final do ano de 2019, de uma distribuidora de bebidas situada na cidade de São Pedro da Aldeia no estado do Rio de Janeiro. Sua área de atendimento consiste em cidades do norte fluminense (Quissamã, Macaé, Rio da Ostras, etc) e região dos lagos (Cabo Frio, Búzios, Saquarema, etc). O detalhamento é realizado abaixo.

- Unidade de Medida: Hecto litro (hcl ou hl);
- CORE: Volume venda de cervejas tradicionais em hecto litros;
- CORE_PM: Volume venda de cervejas categorizadas como “puro malte” em hecto litro;
- HE (*High End)*: Volume de venda de cervejas categorizadas como *premium* em hecto litro;
- Refri: Volume de venda de refrigeirantes em hecto litro;
- OW (*One Way)*: Volume de venda de cervejas categorizadas como descartáveis em hecto litro;
- RGB (*Retornable glass bottle)*: Volume de venda de cervejas categorizadas como retornáveis em hecto litro.
- Inteira: Volume de venda de cervejas 600ml retornáveis em hecto litro ;
- Litro: Volume de venda de cervejas 1000ml retornáveis em hecto litro;
- Meia: Volume de venda de cerejas 300ml retornáveis em hecto litro;
- Barril: Volume de venda de cerveja *“Chopp”* em hecto litro.

Data	CORE	CORE_PM	HE	REFRI	OW	RGB	INTEIRA	LITRO	MEIA	BARRIL
2015-01-01	0	0	0	0	0	0	0	0	0	0
2015-01-02	3141	37	286	447	1828	1662	929	581	47	106
2015-01-03	2790	10	71	182	1347	1545	765	718	26	35
2015-01-04	0	0	0	0	0	0	0	0	0	0
2015-01-05	3771	26	133	711	2081	1890	1047	690	101	53

Figura 20 – Visualização do banco de dados. Fonte: Elaborado pelo autor.

O banco de dados tem um total de 18.250 dados de venda nos quais são realizadas as análises situadas abaixo:

- Venda por categoria por ano;
- Distribuição normal das séries;
- Visualização de *outliers*;
- Disperção de valores;
- Decomposição das séries temporais;
- Verificação de estacionariedade das séries.
  - ✓ Observação gráfica;
  - ✓ Aplicação do teste de *Dickey Fuller e KPSS*;
- Verificação de tendência das séries.

2. ANÁLISE DE VENDA POR ANO POR CATEGORIAS

Neste capítulo é analisada as vendas ao longo do tempo das categorias mencionadas acima na estratificação do banco de dados. Esta análise é importante pois de uma forma visual é possível identificar tendências, sazonalidades, média de vendas por mês, média devendas por trimestre, vendas acumuladas por ano, etc.

2.1 ANÁLISE DE VENDA DE CERVEJA CORE

Esta categoria de cerveja é considerada a principal linha de venda e consequentemente a principal fonte de faturamento da empresa (média de 75% do total). São consideradas as marcas principais dentre todo o portifólio existente, e tem uma fatia de mercado (*market share)* aproximada de 80% quando comparada no mercado nacional total das cervejarias.

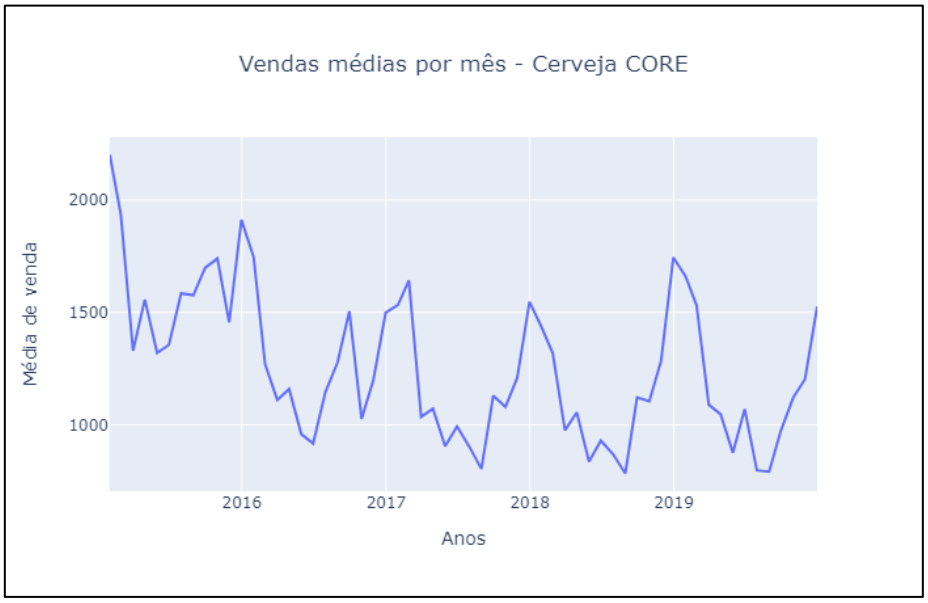


Figura 21 – Vendas médias cerveja core por mês. Fonte: Elaborado pelo autor.

Foram consideradas as médias de venda de cada mês nos anos utilizados na análise (2015-2019). Uma tendência de baixa é percebida com picos e vales bem definidos, e com uma recuperação crescente de venda do ano de 2018 quando em comparação com o ano de 2019. A existência de picos e vales ao longo da série temporal e haver grande variação do volume médio de venda entre os meses do ano, sugere a existência de sazonalidade no modelo, que será comprovada a frente com testes estatísticos de hipótese.

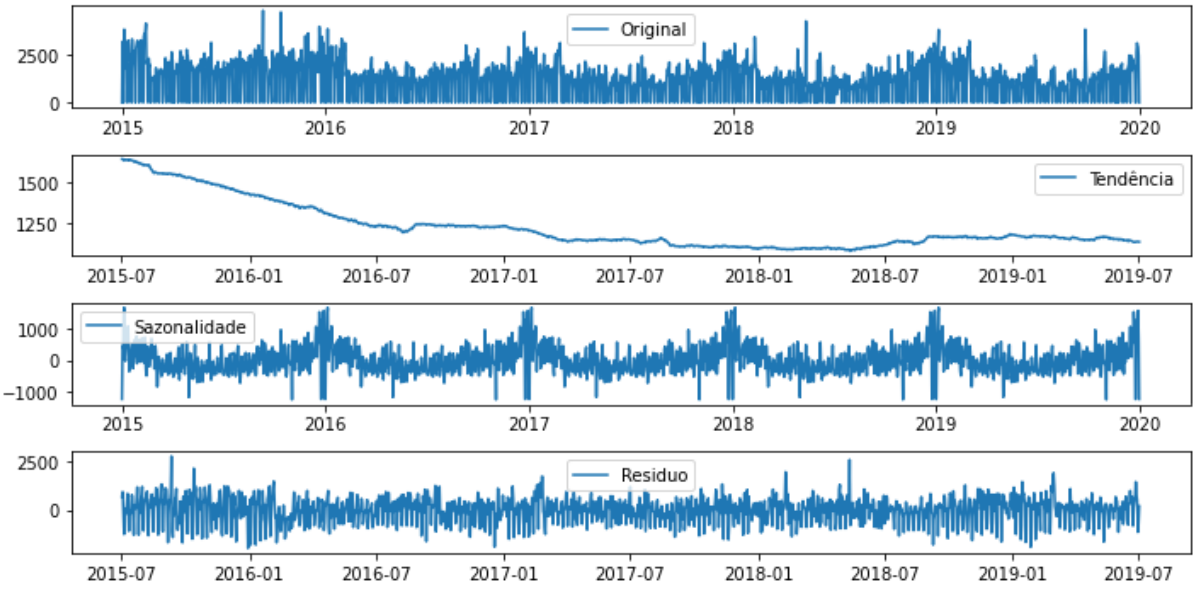


Figura 22 – Decomposição da série cerveja *core*. Fonte: Elaborado pelo autor.

A decomposição da série comprova a tendência de baixa com recuperação do volume de venda entre os anos de 2018 e 2019. O elemento de sazonalidade também pode ser observado com picos e vales ao longo de toda série, mais uma vez indicando forte sazonalidade presente no modelo em questão. Os resíduos são os dados na qual o modelo de decomposição não conseguiu interpretar, quanto menor o resíduo melhores resultados serão obtidos com aplicação de previsibilidade.

A distribuição da série temporal é apresentada abaixo, com inclinação positiva a direita e os dados não estão normalmente distribuídos **(teste estatístico para verificar normalidade)**. O maior volume de observações está na região de 1000 à 2000 hecto litros e baixo número de observações na região mais a direita de 3000 à 4000 hecto litros.

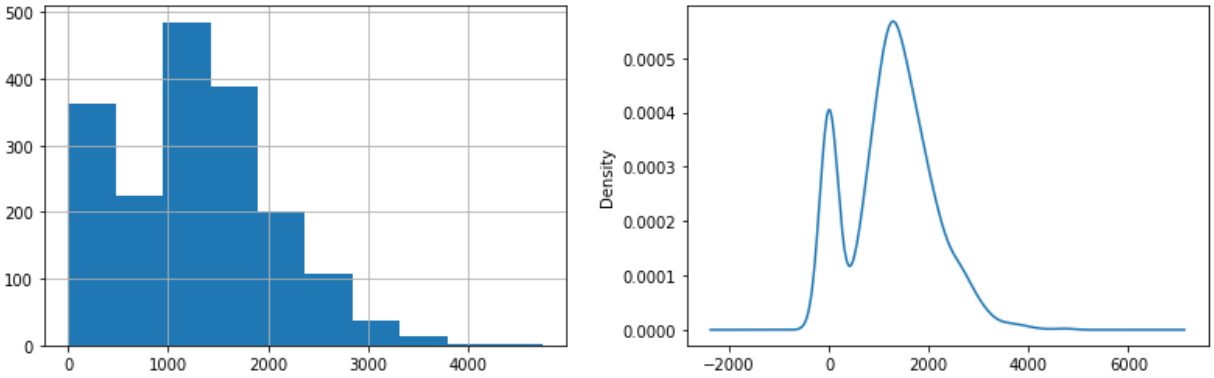


Figura 23 – Histograma e distribuição normal da série. Fonte: Elaborado pelo autor.

Estes dados com poucas observações são considerados *outliers*, que em outras palavras representam dados com um padrão muito fora da curva quando comparado com os demais. Com o gráfico de *Boxplot* é possível verificar as médias, quartis e *outliers* de toda a série temporal por ano de venda.

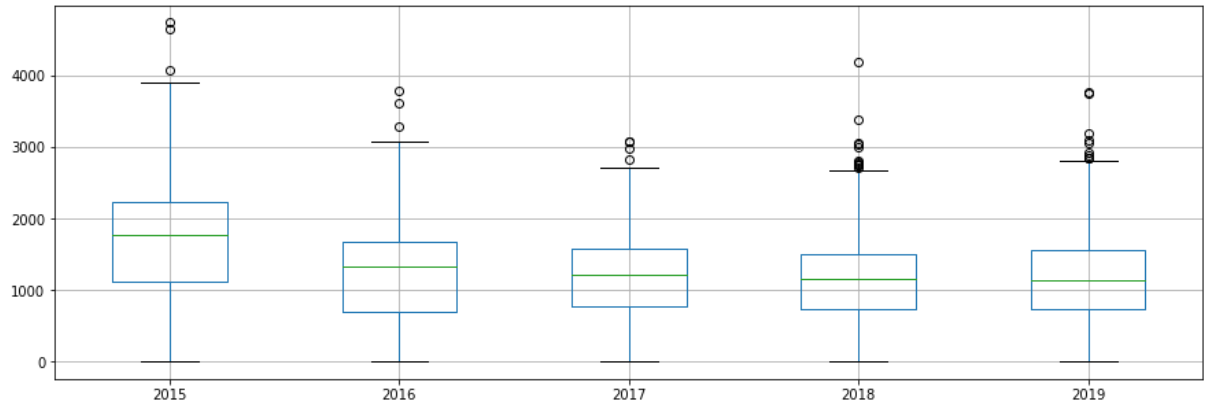


Figura 24 – *Boxplot* com visualização de outliers por ano. Fonte: Elaborado pelo autor.

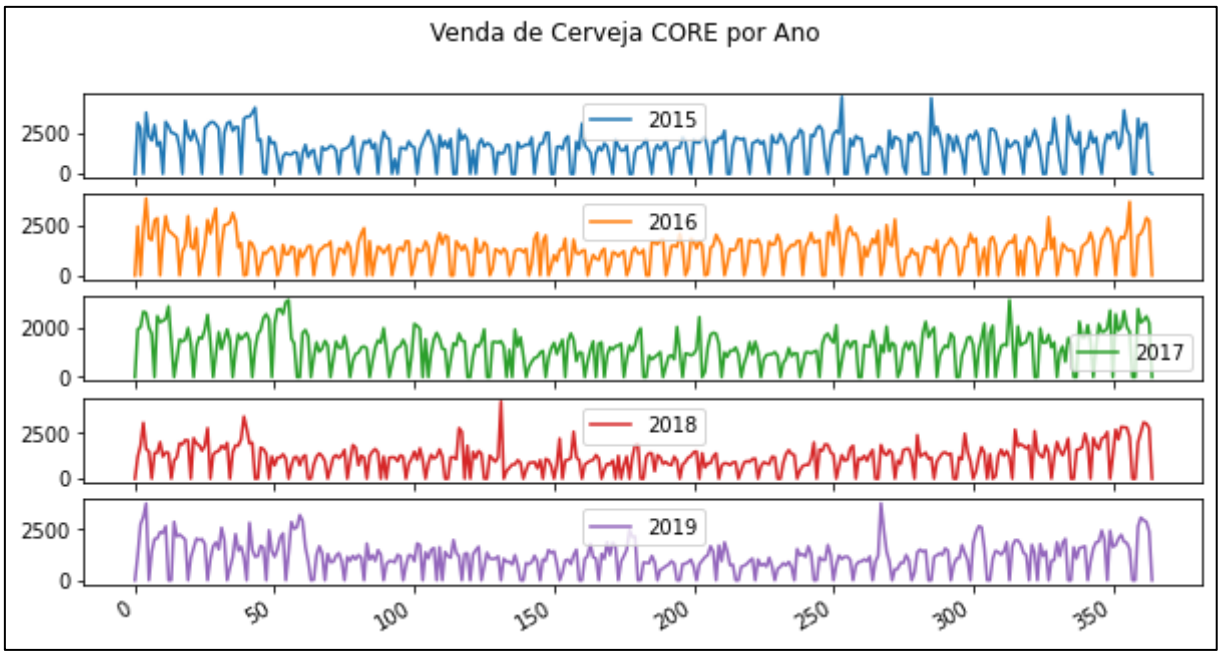


Figura 25 – Vendas médias cerveja core aberto por ano. Fonte: Elaborado pelo autor.

Nesta visualização pode-se analisar mais detalhadamente que os picos e vales são espelhados ao longo do ano, com maior volume concentrado nos meses iniciais do ano (janeiro, fevereiro e março) e com tendência positiva nos meses de novembro e dezembro, meses considerados como verão.

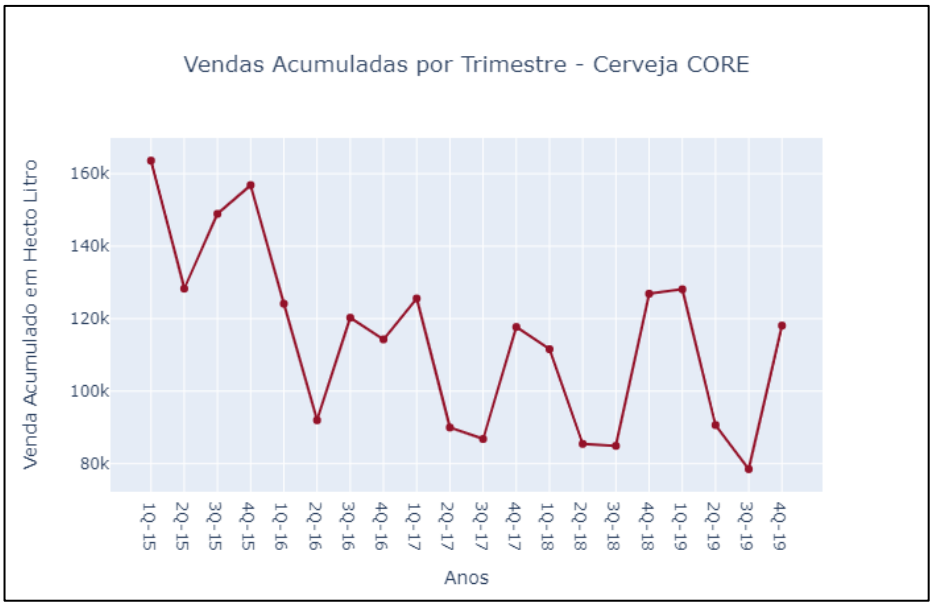


Figura 26 – Vendas acumuladas cerveja core por trimestre. Fonte: Elaborado pelo autor

Observa-se que ao longo dos trimestre houve queda no patamar de vendas acumuladas, quando comparado o primeiro e o quarto trimestre de cada ano. Esses trimestres tem um importância maior, uma vez que são os meses de maior volume de venda devido a estação do ano de maior consumo que é o verão e datas festivas de final de ano (dezembro). Esta queda neste segmento pode ser explicado pela mudança em hábitos de consumo de cerveja no Brasil. Em 2015 vendeu-se aproximadamente 600.000 hecto litros nesta categoria, e logo um ano depois houve queda de 25% no volume acumulado ano.

Segundo a Associação Brasileira de Supermercados (ABRAS) o mercado de cervejas puro malte teve um crescimento médio de 81% no ano de 2019 e o mercado de cervejas "*premium"* teve um crescimento aproximado de 45%. Vale ressaltar que houve crescimento em alguns trimestres devido a ajustes realizados nas árvores de preços de cervejas neste segmento em especifico, com foco principal em criar uma vantagem competitiva no mercado das cervejarias nacionais.

Figura 27 – Vendas acumuladas cerveja core por ano. Fonte: Elaborado pelo autor

Segundo a Associação Brasileira de Supermercados (ABRAS) o mercado de cervejas puro malte teve um crescimento médio de 81% no ano de 2019 e o mercado de cervejas "*premium"* teve um crescimento aproximado de 45%. Vale ressaltar que houve crescimento em alguns trimestres devido a ajustes realizados nas árvores de preços de cervejas neste segmento em especifico, com foco principal em criar uma vantagem competitiva no mercado das cervejarias nacionais.

2.2 ANÁLISE DE VENDA DE CERVEJA RETORNÁVEL

Esta é uma sub categoria de dentre todos os grandes segmentos considerados no modelo (*core*, puro malte, *premium* e refrigerante) atualmente são os produtos que tem a maior rentabilidade para o negócio. Dentre as principais embalagens estão: embalagens de um litro retornável (litro), embalagem de seiscentos mililitros retornável (inteira) e embalagem de trezentos mililitros retornável (meia).

Este sub segmento tem no mercado nacional um *market share* aproximado de 75% de acordo com as últimas leituras. Porém depara-se com um problema de perda de peso das embalagens retornáveis no volume total de venda, com migração de peso para embalagens descartáveis. Isto pode ser considerado um problema uma vez que a margem de lucro dos produtos descartáveis é muito menor quando em comparação a embalagem em análise.

Necessita-se de uma análise mais profunda desta categoria, uma vez que a perda de peso afeta diretamente na rentabilidade do negócio como um todo. Para tal as embalagens litro, inteira e meia que compõe 95% de todas as embalagens retornáveis, serão analisadas e detalhadas neste capítulo.

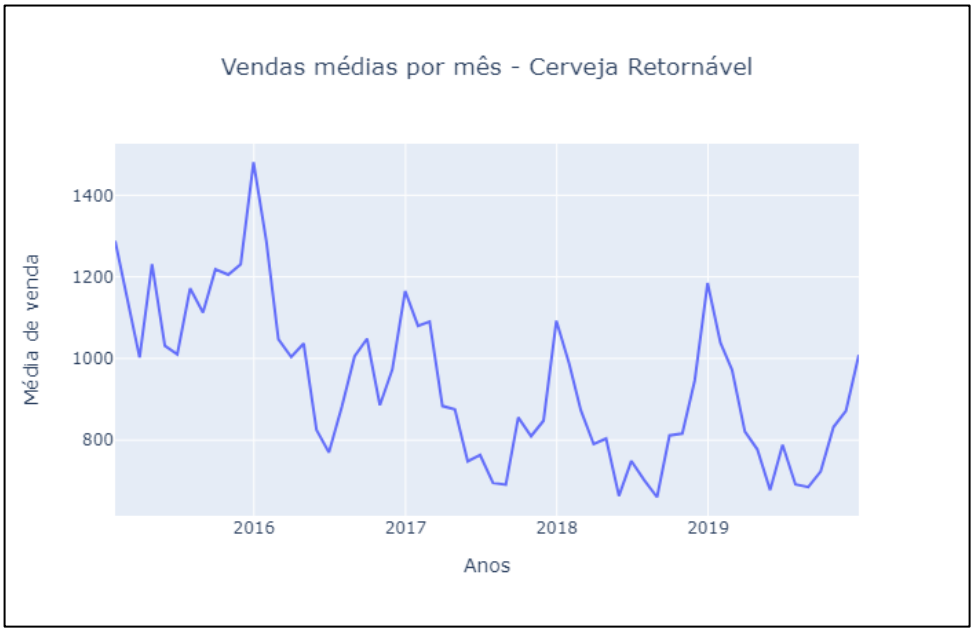


Figura 28 – Venda média por mês de cerveja retornável. Fonte: Elaborado pelo autor

Existe uma tendência de baixa que é facilmente observada no gráfico acima. Com uma leve recuperação do volume de venda de 2018 para 2019. Quando analisamos o picos de maior amplitude, percebemos que se trata de meses iniciais de cada ano. A partir de 2016 houve queda constante nesses picos.

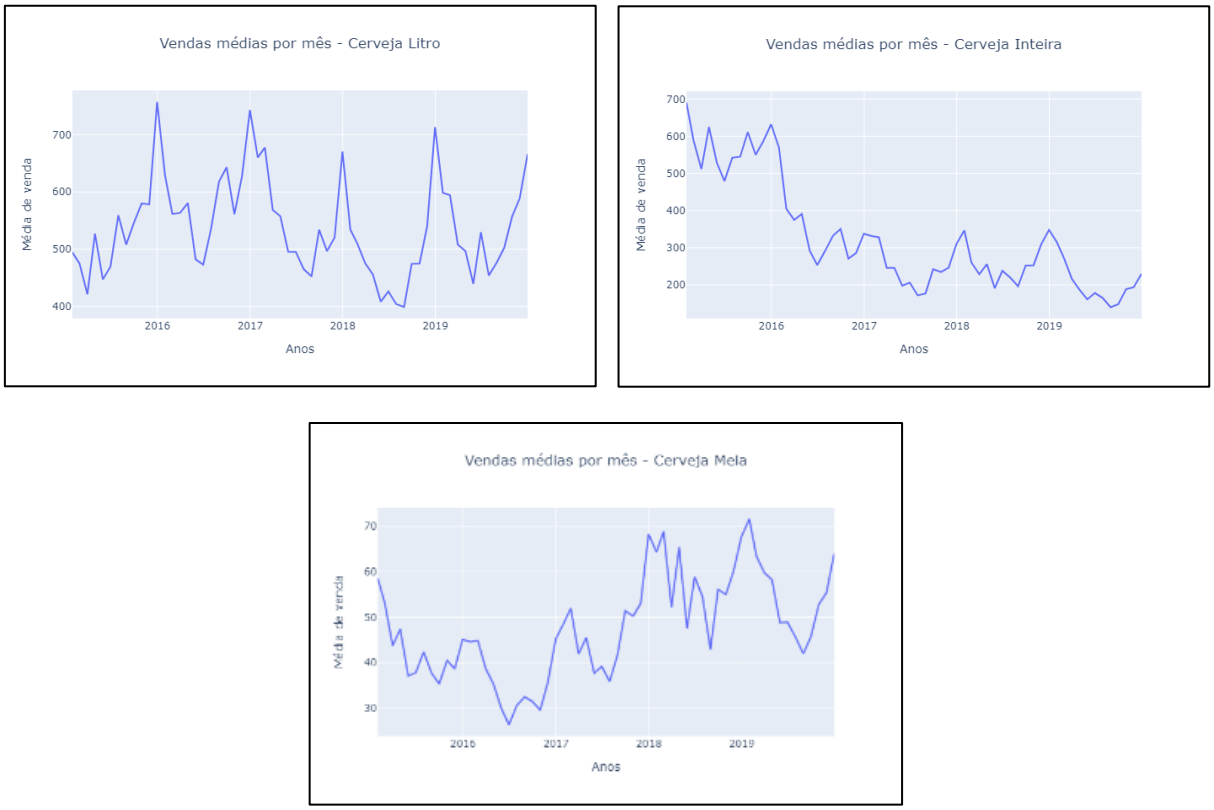


Figura 28 – Média de venda mensal estratificado por embalagem. Fonte: Elaborado pelo autor.

Na analise detalhada por sub categoria verifica-se que a embalagem litro teve queda a partir do ano de 2016 (volume acumulado ano de 213.000 hecto litro) com recuperação de das vendas no período de 2018 à 2019, chegando a igualar o mesmo patamar de vendas acumuladas do ano de 2015 (195.000 hecto litros). A média anual de venda por dia caiu 17% no período de 2016 à 2018, saindo de 585 hecto litro dia em 2016 para um patamar de 484 hecto litros dia em 2018.

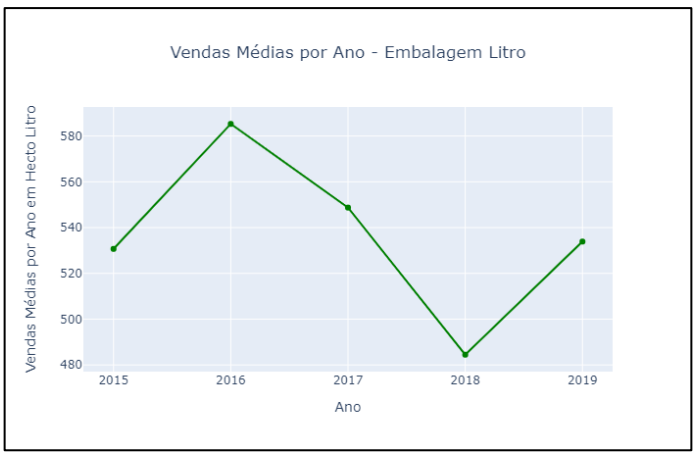


Figura 29 – Venda média por ano, embalagem litro. Fonte: Elaborado pelo autor.

A queda é maior na embalagem retornável inteira, apresentando uma tendencia de baixa desde o ano de 2015, e sem apresentar recuperação no período de 2018 à 2019 como ocorrido nas demais embalagens. Em 2015 a média de venda diária desta embalagem era 574 hecto litros, já em 2019 foi de 199 hecto litros, queda de 65% no período de 2015 à 2019.

No volume acumulado esta embalagem tinha uma representadade de 209.000 hecto litros em 2015, alcançando em 2019 um valor de 72.760 hecto litros. Isto por ser explicado pelo fato da caixa de cerveja inteira ser comercializada com vinte e quatro unidades, quando a embalagem litro é vendida com doze unidades por caixa. Para comparação de precificação das das duas embalagens, é necessária realizar a conversão para real por litro. Nota-se que na inteira o ponto de venda tem um preço de 75 reais por litro, quando no litro esse valor chega a ser de 70 reais.

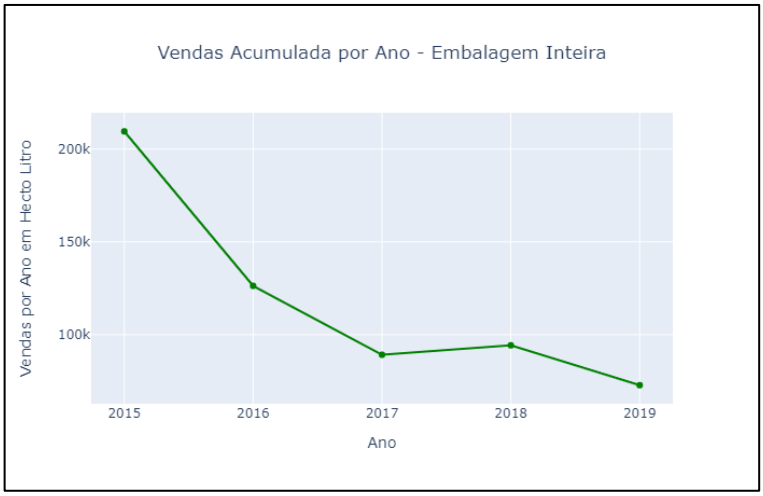


Figura 30 – Volume acumulado por ano, embalagem litro. Fonte: Elaborado pelo autor.

A embalagem meia tem pouca representatividade no volume total de venda de cervejas retornáveis, representando em 2015 um volume acumulado de 15.728 hecto litro, e apresentando crescimento de 26% no período 2015 a 2019 fechando em 19.946 hecto litro no ano de 2019. Em 2015 a venda média por dia era de 43 hecto litro, e em 2019 houve um crescimento para 54 hecto litro dia.

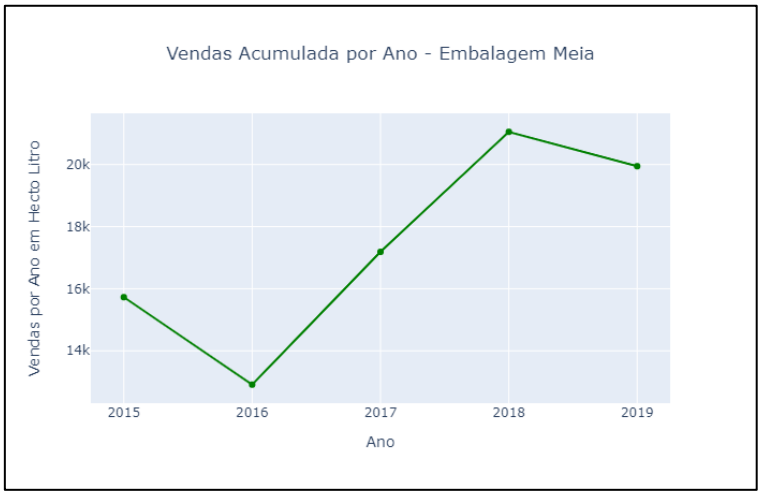


Figura 31 – Volume acumulado por ano, embalagem meia. Fonte: Elaborado pelo autor.

No geral quando se observa detalhadamente os volumes por tipo de embalagem no período de 2015 a 2019 percebe-se um crescimento não significativo de 0,6% nas vendas de embalagem litro, queda de 65% no volume da embalagem inteira e crescimento de 26% no volume de venda acumulado da embalagem meia. Como a embalagem meia tem pouca representatividade sob o volume total de retornável, seu crescimento não é capaz de suprir a queda na embalagem inteira, acarretando uma queda total de aproximadamente 30% no volume acumulado de retornável.

Figura 32 – Volume acumulado por ano, cerveja retornável. Fonte: Elaborado pelo autor.

Ambas as distribuições normal das embalagens analisadas são deslocadas a direita com concetração das observações em regiões especificas de volume de venda. Este fenomeno por ser observado nas figuras abaixo.

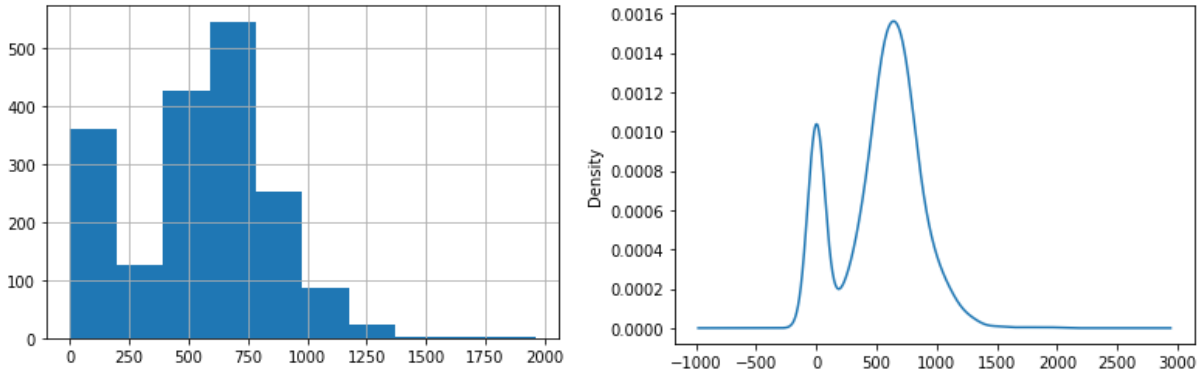


Figura 33 – Histograma e distribuição normal, embalagem litro. Fonte: Elaborado pelo autor.

Figura 34 – Histograma e distribuição normal, cerveja inteira. Fonte: Elaborado pelo autor.

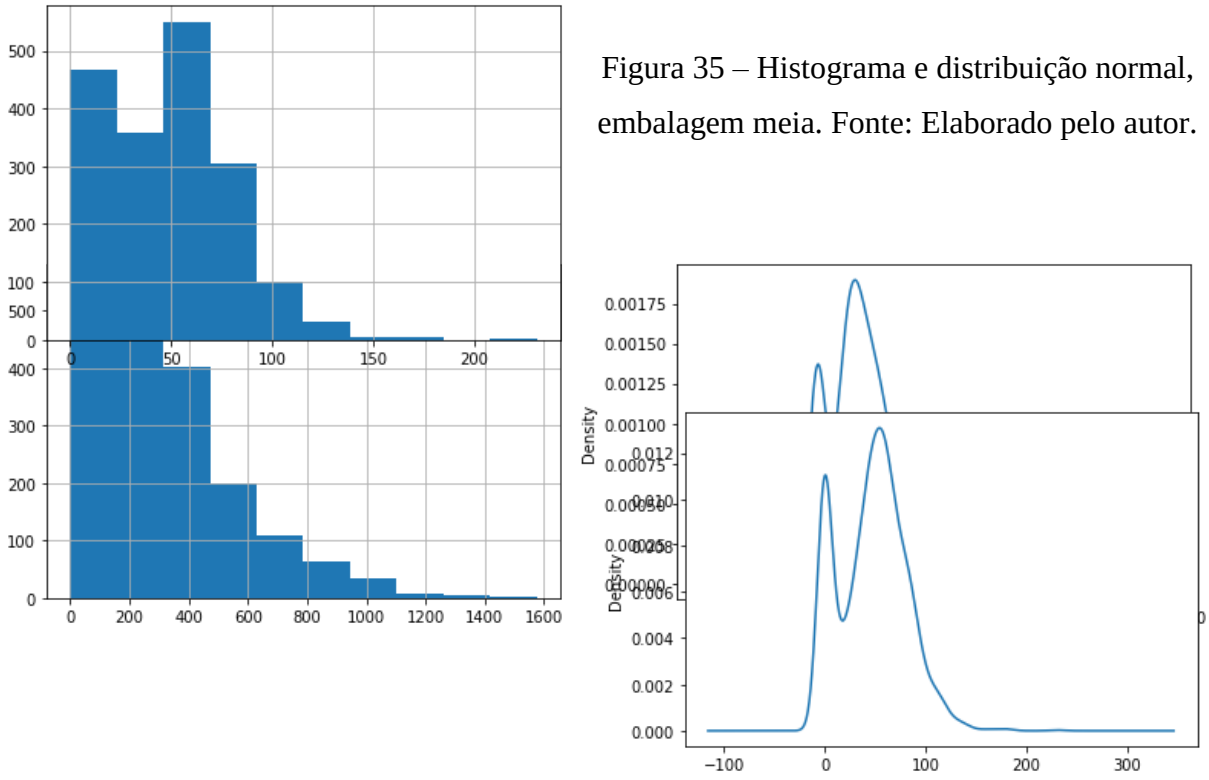


Figura 35 – Histograma e distribuição normal, embalagem meia. Fonte: Elaborado pelo autor.

Como a série temporal em questão tem forte sazonalidade, há momentos durante o ano em que alguns volumes de venda no dia estejam fora do padrão do respectivo ano. Estes *outliers* estão presentes em quase todo modelo de série temporal e podem ser visualizados nas análises abaixo.

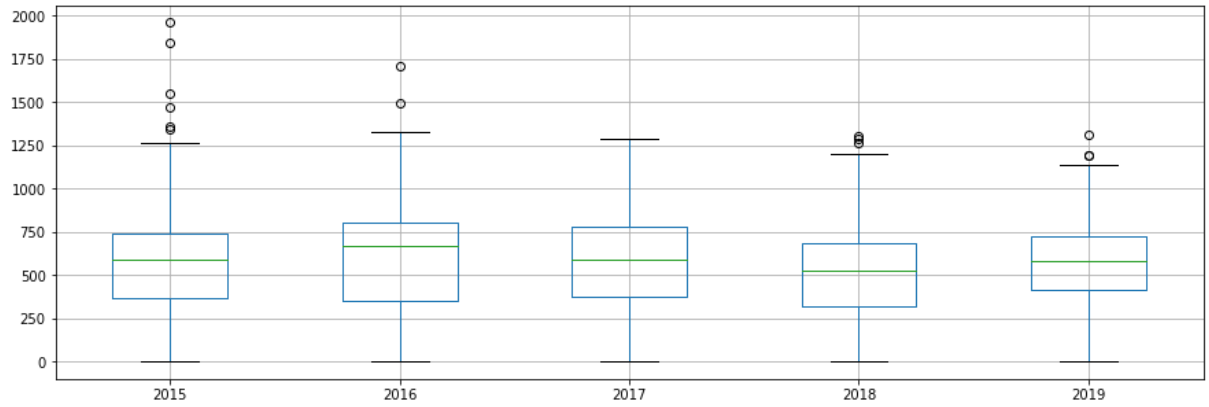
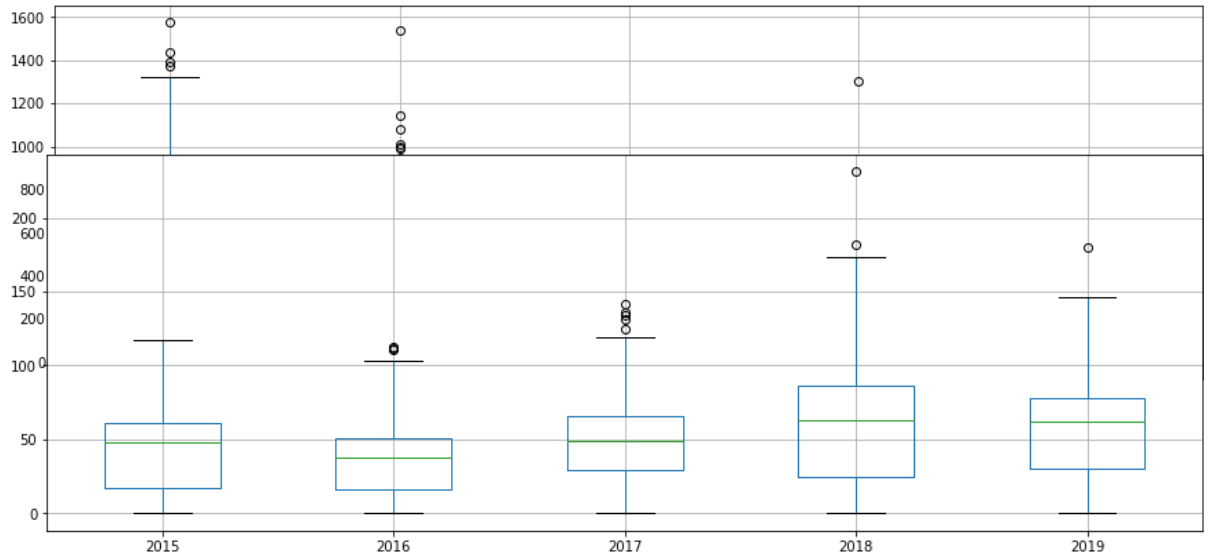


Figura 36 – *Boxplot* com visualização de outliers por ano, embalagem litro. Fonte: Elaborado pelo autor.

Figura 37 – *Boxplot* com visualização de outliers por ano, embalagem inteira. Fonte: Elaborado pelo autor.

Figura 38 – *Boxplot* com visualização de outliers por ano, embalagem meia. Fonte: Elaborado pelo autor.



2.3 ANÁLISE DE VENDA DE CERVEJA DESCARTÁVEIS

Assim como as cervejas retornáveis esta também é uma sub categoria de dentre todos os grandes segmentos considerados no modelo (*core,* puro malte, *premium* e refrigerante) e na visão empresarial são produtos com baixa margem de lucro. Dentre as principais embalagens estão: embalagens lata 350 mililitros, embalagens lata 473 mililitros, embalagens lata 310 e 410 mililitros e *long necks* 355 mililitros.

O *market share* desta categoria é aproximadamente de 55% de acordo com as últimas pesquisas. Este tipo de consumo vem aumentando muito nos últimos 3 anos, e como a fatia de mercado não é tão alta como as demais embalagens existe uma forte concorrência, principalmente na embalagem lata 473 mililitros, que é vulgarmente conhecida como “latão”. A queda no volume acumulado de venda da retornável e a não alteração do volume total, significa que houve migração de volume para as embalagens descartáveis. A soma de volume de cerveja retornável e descartável em 2016 atingiu 484.727 hecto litros, e em 2019 foi de 482.074 hecto litros, mesmo com tendência de queda na embalagem retornável.

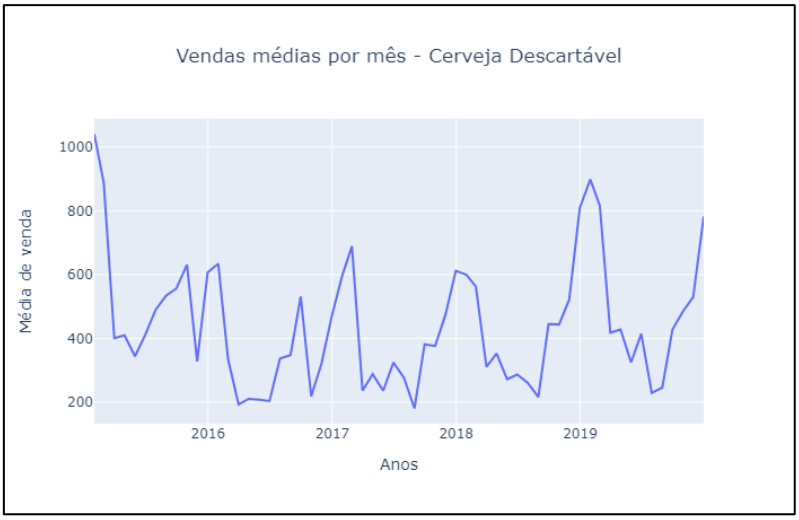


Figura 39 – Venda média por mês de cerveja descartável. Fonte: Elaborado pelo autor.

No período de 2015 a 2016 houve queda na média diária de venda, caindo de um patamar de 551 hecto litro por dia, para 333 hecto litro dia, queda de 40% em volume de descartáveis dia. Porém desde então, essa média teve aumento no período de 2016 a 2019, chegando ao valor de 497 hecto litro dia em 2019, crescimento de 49% em 3 anos.

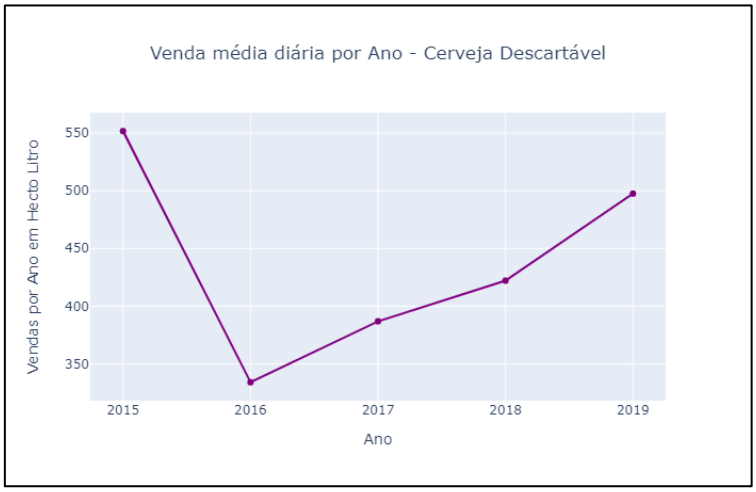


Figura 40 – Venda média por mês de cerveja descartável. Fonte: Elaborado pelo autor.

Esse crescimento por ser verificando também na série temporal por trimestres, na qual observa-se um aumento no patamar de venda média diária entres os primeiros trimestres de cada ano a partir de 2016.

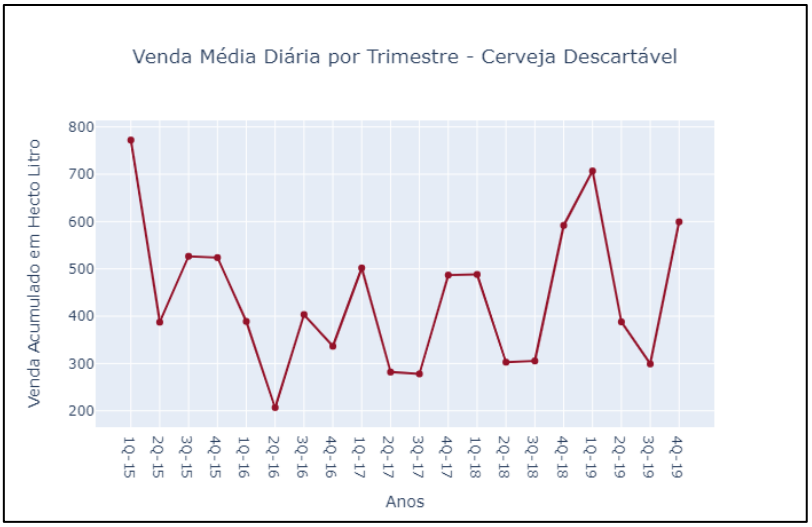


Figura 41 – Venda média diária por trimestre de cerveja descartável. Fonte: Elaborado pelo autor.

Como complemento da visualização por trimestre pode-se utilizar a analise de venda por ano separadamente, ressaltando mais uma vez para notar-se o maior patamar de venda dia nos 90 dias iniciais e 60 dias finais do ano correspondente.

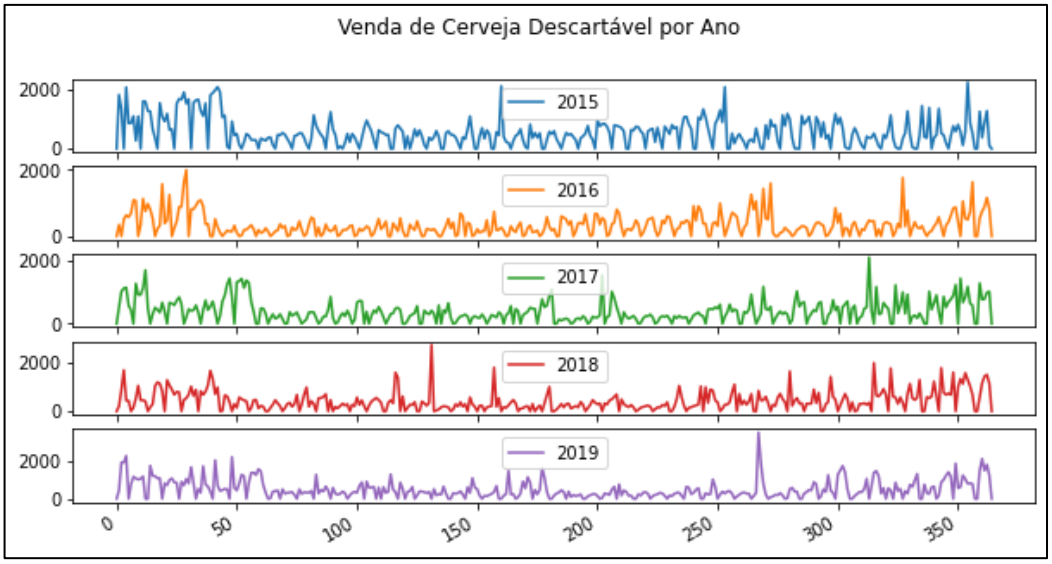


Figura 42 – Venda diária aberta por ano de cerveja descartável. Fonte: Elaborado pelo autor.

Através da decomposição da série temporal é possível observar o exatamo momento de recuperação e perda de tendência do objeto em questão. Além disso verifica-se também picos e vales presentes ao longo de toda série. A sazonalidade apresentada na decomposição será provada através de testes estatisticos apresentados em seção separada.

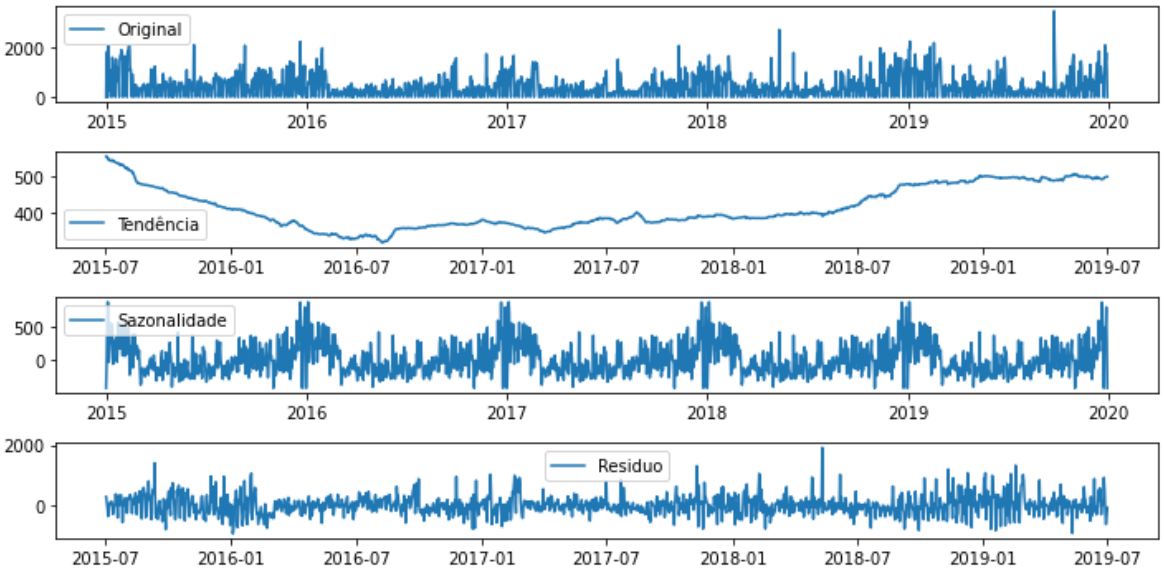


Figura 43 – Decomposição da série cerveja descartável. Fonte: Elaborado pelo autor.

Há grande concentração de observações no histograma no volume entre 0 e 650 hecto litro de venda por dia. Porém nota-se que existem valores observados até na ordem de 3.500 hecto litro dia. Isto pode ser explicado pela alta concorrência no mercado de descartáveis, onde utiliza-se de baixar o preço para aumentar o volume de venda, chegando a descontos no

patamar de R$5 por caixa vendida, fato tal que não ocorre nas embalagens retornáveis, com desconto médio de R$2 por caixa vendida.

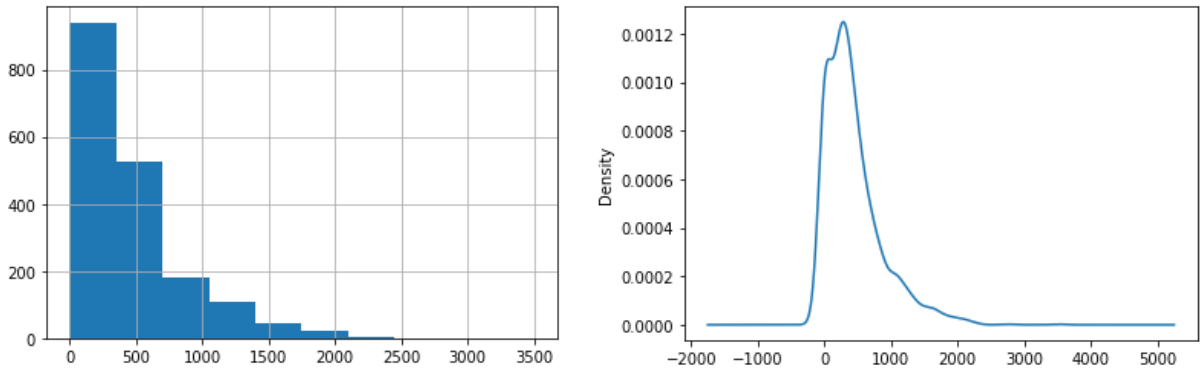


Figura 44 – Histograma e distribuição normal, cerveja descartável. Fonte: Elaborado pelo autor.

Alguns valores de venda diária acima de 1.000 hecto dia já são considerados *outliers* pelo modelo e podem ser visualizados através do *Boxplot* abaixo.

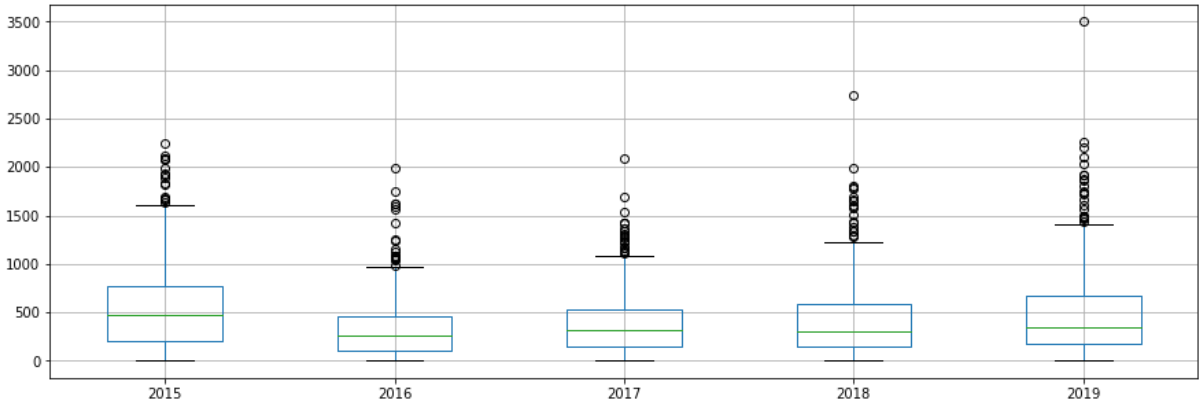


Figura 45 – *Boxplot* com visualização de outliers por ano, cerveja descartável. Fonte: Elaborado pelo autor.

Em resumo as embalagens descartáveis tem ganhado peso sob o volume total de venda no ano. Tal fato é preocupante principalmente pela rentabilidade de produto ser em média 50% menor quando em comparação com as embalagens retornáveis. O preço por litro das embalagens de 473 mililitros é maior que das cervejas retornáveis, chegando um preço médio de R$78 por litro.

Mesmo com as embalagens retornáveis serem o melhor custo benefico por litro de liquido vendido, quem regula o mercado é a caracteristica de consumo, onde cada momento

em que o consumidor se encontra pede um tipo de embalagem diferente. O cliente faz a conta de economia e versus comodidade, e nem sempre o preço vence.

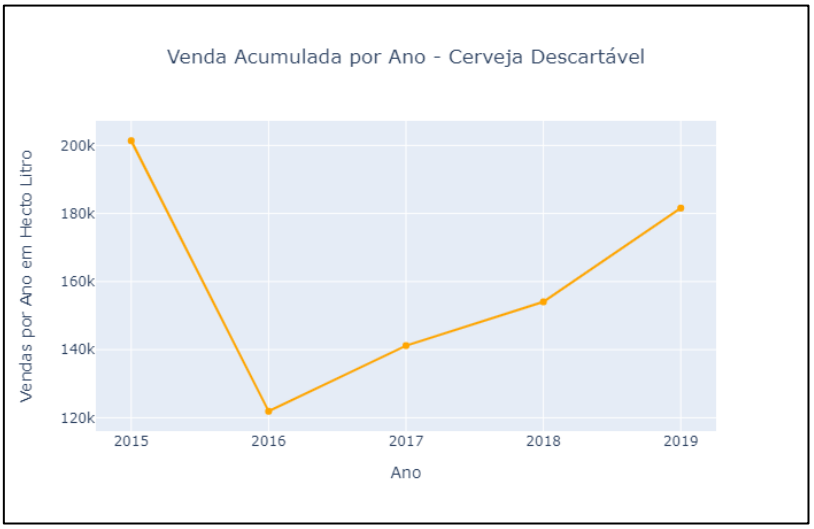


Figura 46 – Volume acumulado por ano, cerveja descartável. Fonte: Elaborado pelo autor.

2.4 ANÁLISE DE VENDA DE CERVEJA PURO MALTE

O mercado de cervejas categorizadas como puro malte vem crescendo ano após ano, principalmente a partir do ano de 2018. Estas cervejas não utilizam cereais não maltados em sua fabricação e recebem a denominação de “puro malte”. Segundo a ABRABE (Associação Brasileira de Bebidas), o consumidor brasileiro passou a observar mais atentamente as características de todas as bebidas, a harmonização e a presença de mais qualidade.

A tendência é que a indústria cervejeira sofra uma remodelação das marcas mais consumidas nos próximos anos, principalmente no que tange a rótulos puro malte. Para a ABRABE, no ano de 2019 foram lançadas 74 novas cervejas nesse mercado. Para tal observa-se abaixo o crescimento a nível exponencial nas vendas diárias desta categoria.

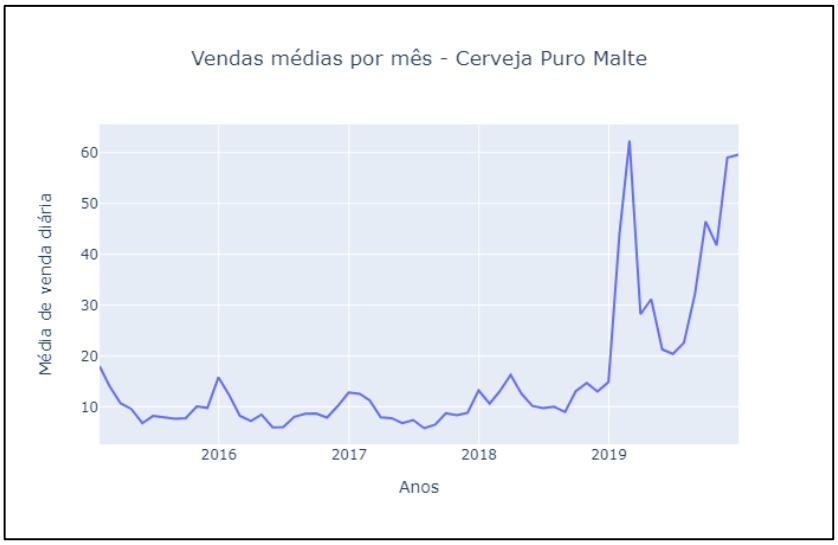


Figura 47 – Venda média por mês de cerveja puro malte. Fonte: Elaborado pelo autor.

Observa-se que a partir do ano de 2018 houve um forte crescimento na média de vendas diárias, quando em 2019 o patamar de 20 hecto litros dia consegue ser superado. Em 2015 a média de venda foi de 10,5 hecto litro dia e 2019 com uma média de 39 hecto litro dia, representando um crescimento de 270% no período.

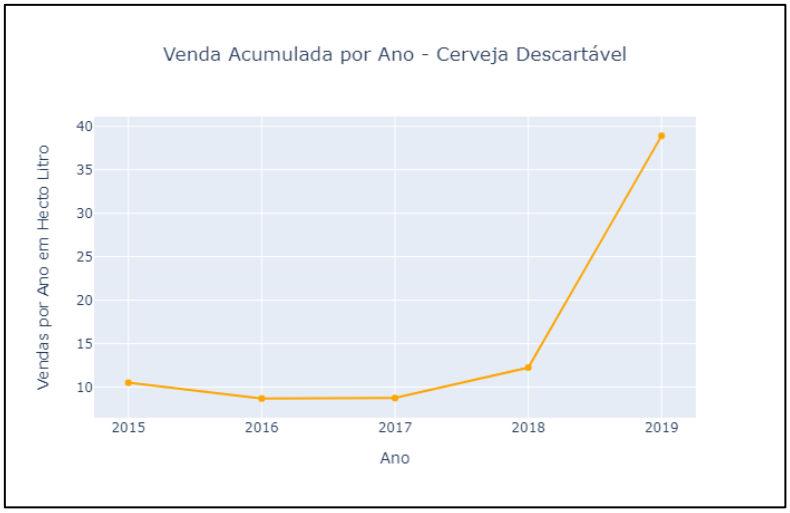


Figura 48 – Venda média diária por ano de cerveja puro malte. Fonte: Elaborado pelo autor.

A decomposição temporal mostra tendência positiva a partir de 2018 como mencionado anteriormente, além disso o elemento sazonal se faz presente, porém em menor intensidade quando comparado as análises anteriores. Outro ponto a observar é a baixa quantidade de resíduos da série temporal.

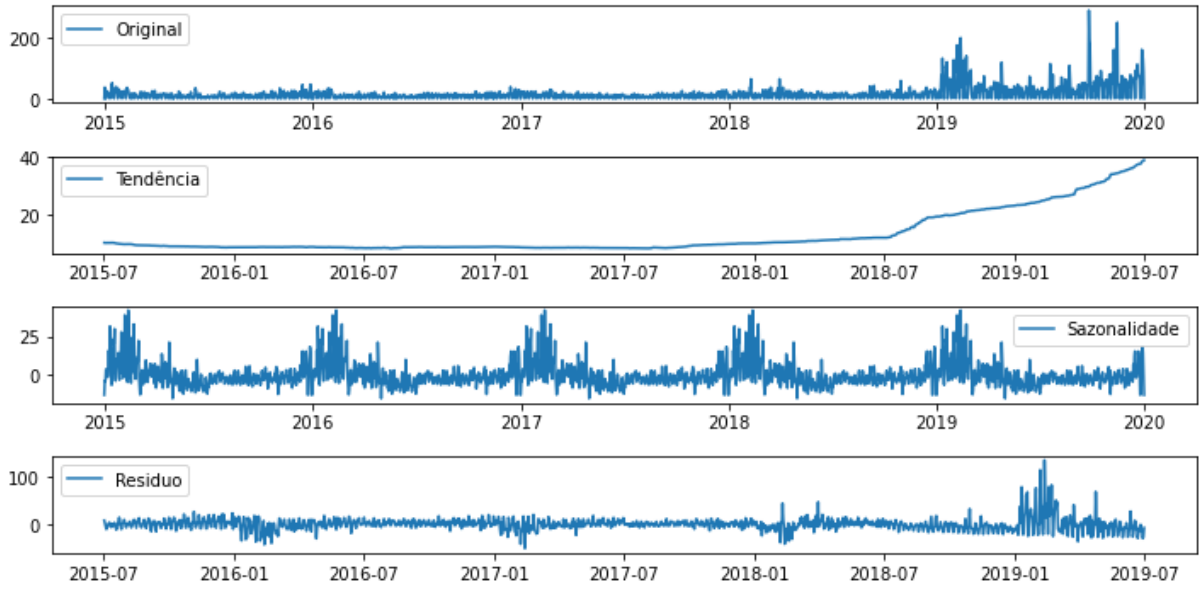


Figura 49 – Decomposição da série cerveja puro malte. Fonte: Elaborado pelo autor.

A distribuição das observações tem sua grande maioria localizada na região entre zero e cinquenta hecto litro dia. Como a partir de 2018 houve crescimento, a tendência é o modelo considerar que há mais *outliers* após esse período.

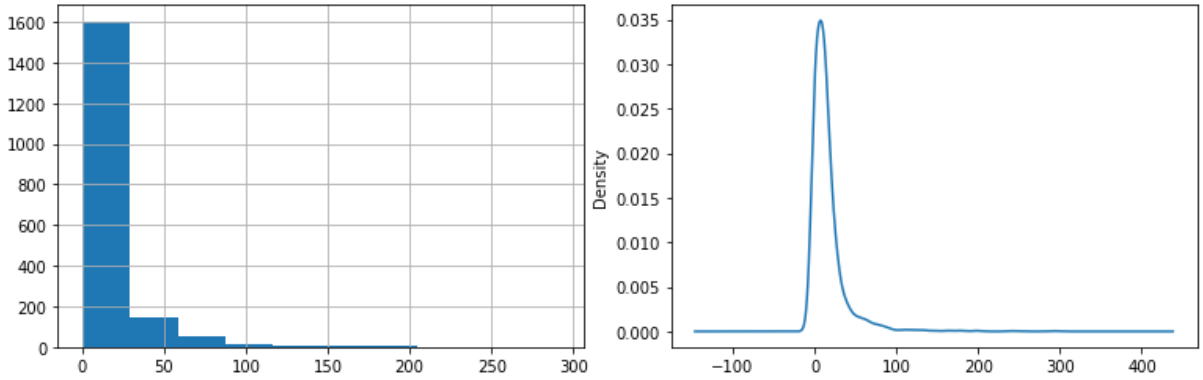


Figura 50 – Histograma e distribuição normal, cerveja puro malte. Fonte: Elaborado pelo autor.

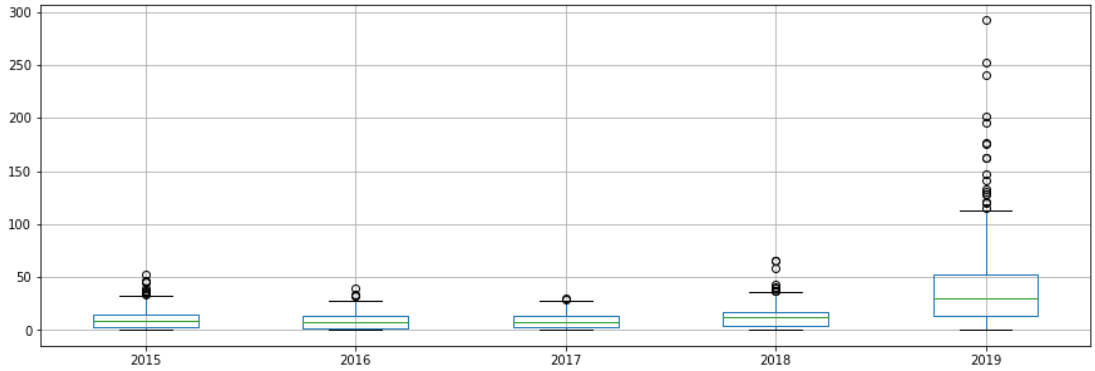


Figura 51 – *Boxplot* com visualização de outliers por ano, cerveja puro malte. Fonte: Elaborado pelo autor.

É importante ressaltar que esse o *market share* neste segmento é menor que 50%, ou seja, com um mercado emergente e com baixa performance de *share*, pela migração automática do consumidor para cervejas desse segmento perde-se volume de venda. Em outras palavras se um consumidor de cervejas tipo *core* passar a consumir no segmento de puro malte, existe mais chance do mesmo consumir cervejas concorrentes.

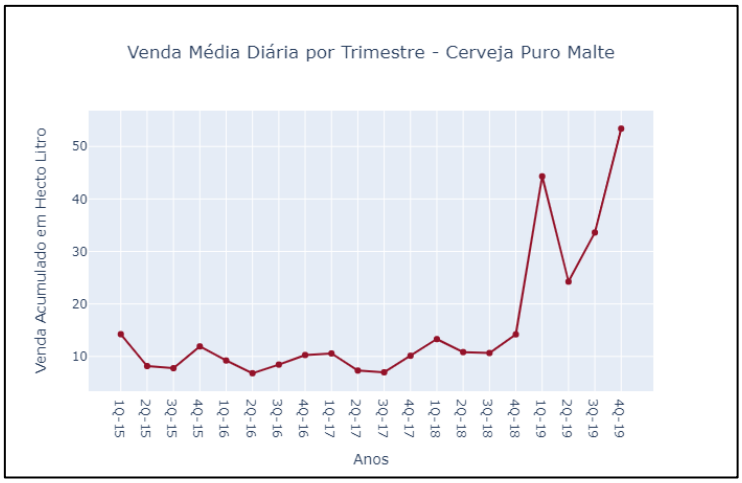


Figura 52 – Venda média diária por trimestre de cerveja puro malte. Fonte: Elaborado pelo autor.

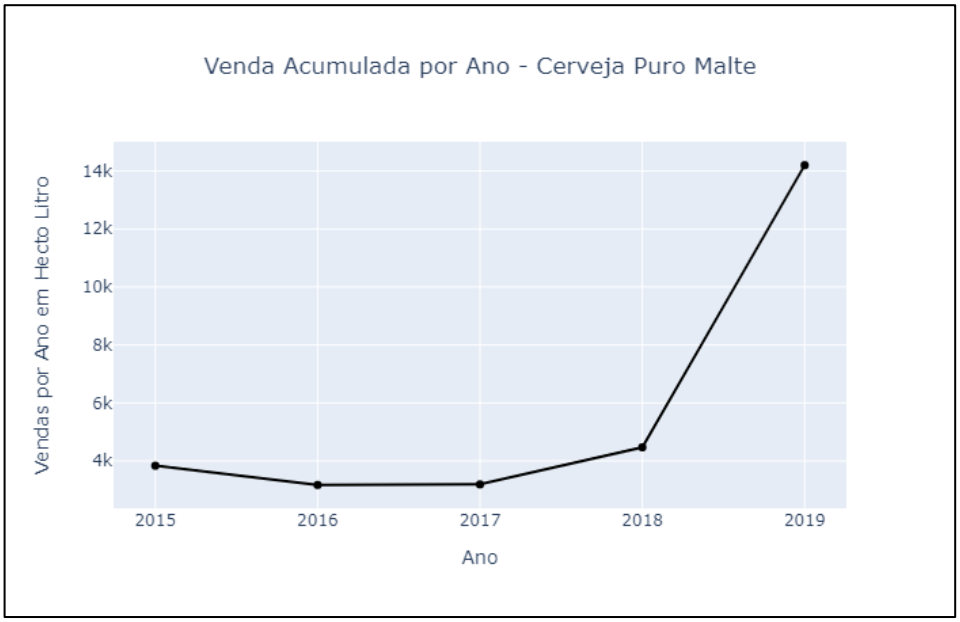


Figura 53 – Volume acumulado por ano, cerveja puro malte. Fonte: Elaborado pelo autor.

O crescimento do segmento puro malte vem tomando proporções cada vez maiores, visto que o consumidor cada vez mais busca cervejas desse tipo, há uma necessidade maior de se consolidar marcas sólidas no mercado nacional, uma vez que a fatia de mercado nesse segmento é baixa quando comparada aos demais.

Frente a um consumidor cada mais exigente e preocupado cada vez mais com os ingredientes do produto, experiencia de consumo e inovações de mercado, a estratégia é fortalecer todo o portifólio com lançamento de inovações cervejeiras, reforçando que para mercados onde se tem consumidores mais exigentes, faz-se necessario trabalhar com diversas marcas para momento diferentes.

Dados revelados pela empresa Nilsen, empresa considerada líder em pesquisa de *market share,* o Grupo Petrópolis e Heineken cresceram sua fatia de mercado em 1,6% e 1,9% respectivamente, principalmente pela elevação de consumo de marcas dos grupos neste segmento puro malte.

2.5 ANÁLISE DE VENDA DE CERVEJA PREMIUM

O mercado de cervejas segmentadas como *premium* é emergente assim como o mercado puro malte. Segundo a ABRABE a representatividade dessa categoria foi a que mais teve crescimento nos últimos dez anos. Este mercado se vincula marcas mais sofisticadas, além de nível superiores de preço quando em comparação com o mercado de cervejas *core*. Para a Associação Brasileira da Industria Cervejeira (CervBrasil), mudança de hábitos dos consumidores cervejeiros fez com que o mercado de cervejas *premium* crescesse no Brasil, atingindo um patamar de 15% do mercado.

Esta categoria caracteriza-se por ter rentabilidade similar a cervejas retornáveis do segmento *core*. O *market share* neste segmento é aproximadamente 57% com concorrência bem definida de marcas bem consolidadas no mercado nacional. A partir de 2016 houve crescente constante no volume de venda com crescimento de 86% em três anos.

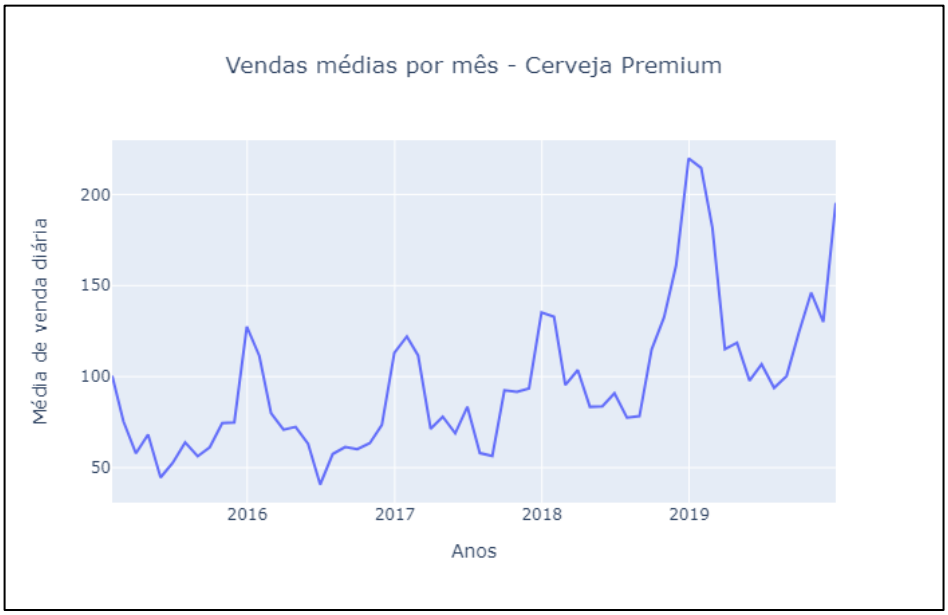


Figura 54 – Venda média por mês de cerveja *premium*. Fonte: Elaborado pelo autor.

Assim como os demais segmentos, a sazonalidade se faz presente na série temporal. Talz fato poderá ser melhor observado na decomposição do modelo. Nota-se tendência posivita de alta para os próximos anos. A média de venda em 2015 era de 71 hecto litro dia, e em 2019 foi de 135 hecto litro dia.

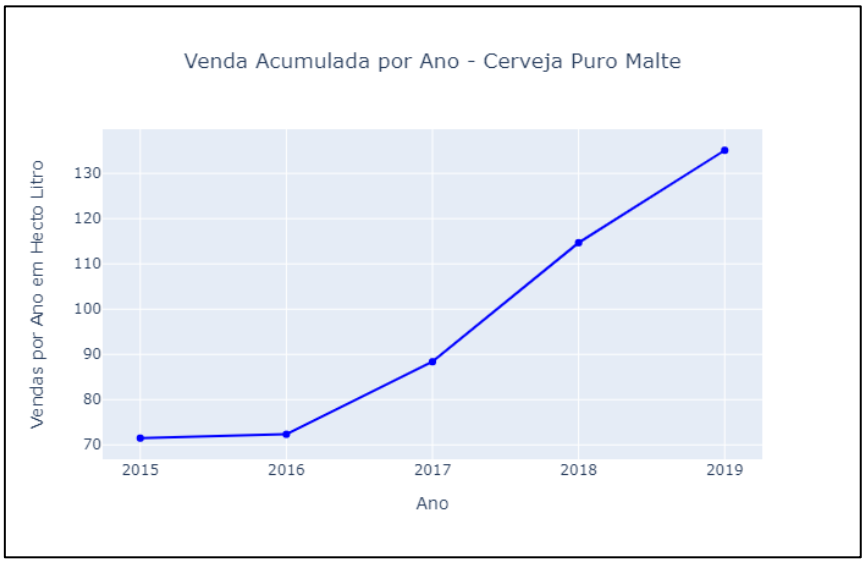


Figura 55 – Venda média diária por ano de cerveja *premium*. Fonte: Elaborado pelo autor.

A distribuição dos dados se dá principalmente no intervalo de 0 a 180 hecto litro, com distribuição não normal dos dados, como pode ser observado abaixo. Há observações na ordem de 300 a 400 hecto litro, este fato pode ser explicado pela sazonal de vendas em meses iniciais e finais de cada ano.

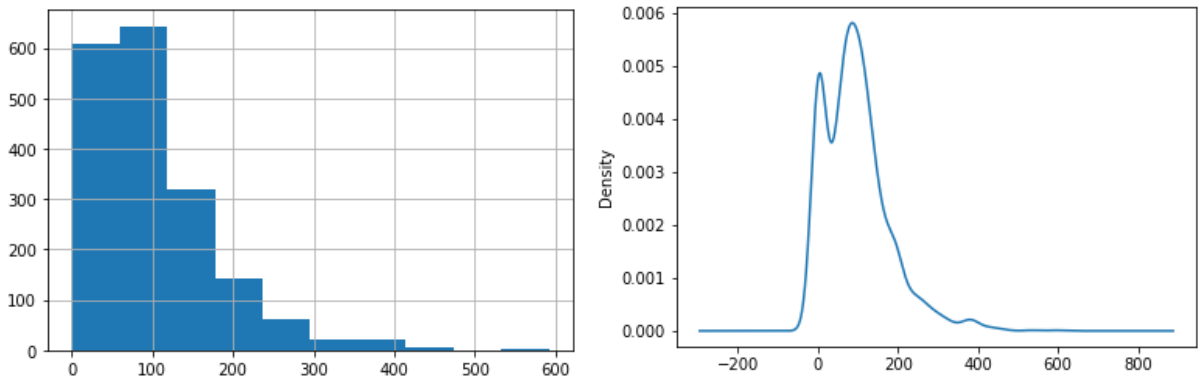


Figura 56 – Histograma e distribuição normal, cerveja *premium*. Fonte: Elaborado pelo autor.

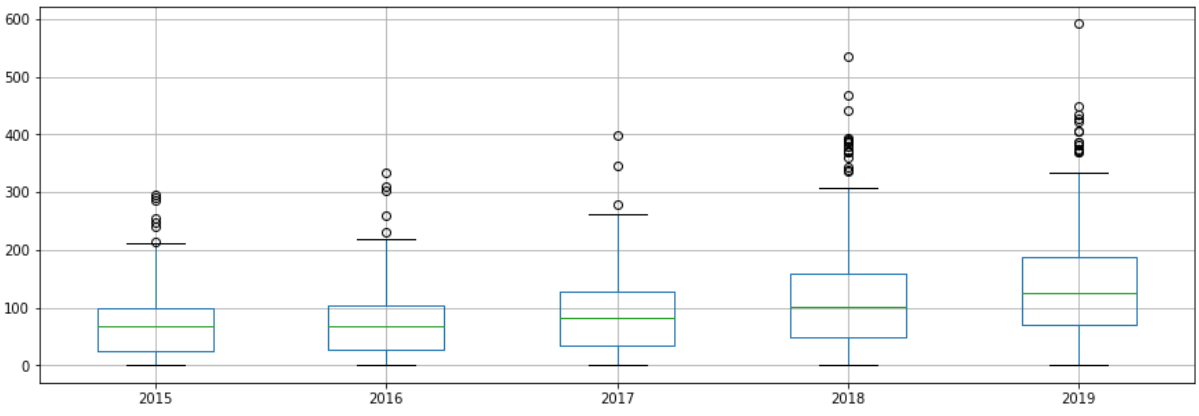


Figura 57 – *Boxplot* com visualização de outliers por ano, cerveja puro malte. Fonte: Elaborado pelo autor.

Em 2015 o volume de venda acumulado desta categoria de cerveja foi na ordem de 26.000 hecto litro ano, quando em 2019 houve crescimento de 88% representando 49.000 hecto litro ano. Vale ressaltar que o mercado de cervejas *premium* é considerado pela ABRAPE um mercado mais maduro, principalmente pelo surgimento de micro cervejarias com cervejas artesainais nos ultimos anos. Apesar de ser um mercado maduro, considera-se o mesmo emergente, com crescimento de 17% no periodo de 2018 a 2019.

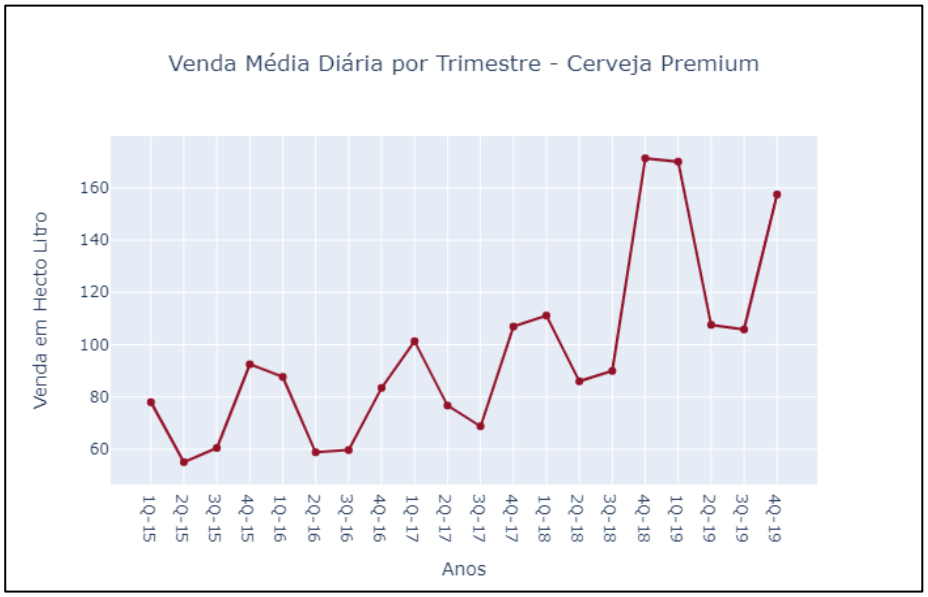


Figura 58 – Venda média diária por trimestre de cerveja *premium*. Fonte: Elaborado pelo autor.

Figura 59 – Volume acumulado por ano, cerveja *premium*. Fonte: Elaborado pelo autor.

3. TESTES ESTATÍSTICOS PARA DETECÇÃO DE ESTACIONAREIDADE E TENDÊNCIA

Uma série é dita estacionaria quando ela se desenvolve no tempo aleatoriamente ao redo de uma média constante, refletindo alguma forma de equilibrioestável. Na prática a princípio séries temporais com tendência e sazonalidade não são estacionárias. Existem casos que séries com elementos sazonais podem ser estacionárias, porém isto não é comum (*Portal Action*, 2019). Esta análise é importante pois existem técnicas analíticas de séries temporais que dependem da estacionariedade para terem um bom funcionamento, principalmente as ferramentas de previsões futuras (Ling et al., 2015).

Existem três tipos de séries estacionarias, são eles: estacionareidade estrita, estacionareidade com tendência e estacionareidade diferencial. A estacionareidade estrita é aquela na qual a série satifaz a definição matematica de processo estacionario, ou seja, média e variância não são função do tempo. A estacionareidade com tendência é uma série que não apresenta elemento de váriância, porém exibe tendência que pode ser removida pelo processo de diferenciação. A estacionareidade diferencial é aquela onde existem ambos os elementos de variância e média em função do tempo. Elas podem se torna estrita pelo processo de diferenciação (Giraitis et al., 2003).

Para determinar se uma série é estacionária é sugerida a aplicação de testes estatísticos como, tais como: *Dickey Fuller, KPSS e Philips-Perron.* Chang (2002) sugere a aplicação de *Dickey Fuller* e *KPSS* no modelo, uma vez que um teste pode considerar a série estacionária enquanto ou não. *Dickey Fuller* tem melhor aplicabilidade em séries estritas e diferenciais, enquanto o teste de KPSS tem boa performance em séries temporais com tendência.

3.1 APLICAÇÃO DE DICKEY FULLER

Dickey Fuller é utilizado para detecção de estacionareidade, isso se dá pela rejeição da hipotese nula para p-value maiores que 0.05, ou seja, a hipótese de não estacionareidade é

rejeitada. Diversos modelos não estacionarios foram testados em modelos de previsão de séries temporias, porém obtiveram baixo desempenho (Białkowski et al., 2015).

O teste de *Dickey Fuller* é aplicado a todas as séries temporais mencionadas no capítulo 2, caso haja necessidade utiliza-se da ferramenta de diferenciação para obter-se valores de p que rejeitem a hipótese nula. A biblioteca utiliza é a "adfuller" dentro do pacote "statsmodels" Tomam-se as seguintes hipóteses H0 e H1:

H0 (p-value > 0.05): Falha para rejeitar a hipótese nula, a série é não estacionária;

H1 (p-value ≤ 0.05): Rejeita-se a hipotese nula, a série é estacionária.

Obs: Não será aplicada diferenciação para o teste inicial.

Inicialmente os valores de p-value para as séries temporais em questão são apresentadas na tabela abaixo, vale ressaltar que os valores utilizados são os originais de cada série.

Série Temporal	p-value
Cerveja Core	0.000062
Cerveja Retornável	0.000693
Cerveja Litro	0.000012
Cerveja Inteira	0.051001
Cerveja Meia	0.001055
Cerveja Descartável	0.000008
Cerveja Puro Malte	0.332301
Cerveja Premium	0.114270

Figura 60– Aplicação de *Dickey Fuller* nas séries originais. Fonte: Elaborado pelo autor.

Como era de se esperar para as séries temporais onde quase não existe tendência o valor de p-value foi menor que 0.05, ou seja, rejeita-se a hipotese nula e toma-se como pressuposto que a série é estacionária. Porém as séries de Cerveja Inteira, Cerveja Puro Malte e Cerveja *Premium* apresentam p-values maior que 0.05, ou seja, não são séries estacionárias, sendo assim havendo necessidade de se aplicar a diferenciação no modelo em questão.

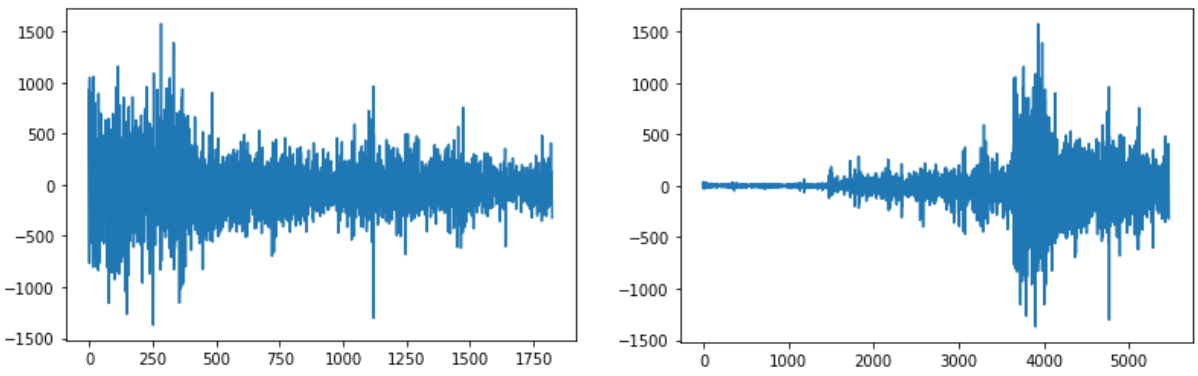


Figura 61– Cerveja Inteira, diferenciação remoção de tendência e sazonalidade respectivamente. Fonte: Elaborado pelo autor.

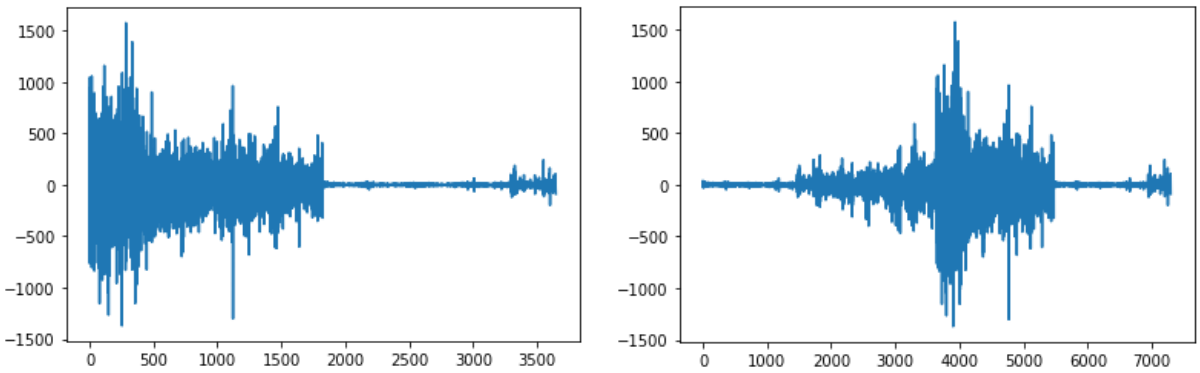


Figura 62– Cerveja Puro Malte, diferenciação remoção de tendência e sazonalidade respectivamente. Fonte: Elaborado pelo autor.

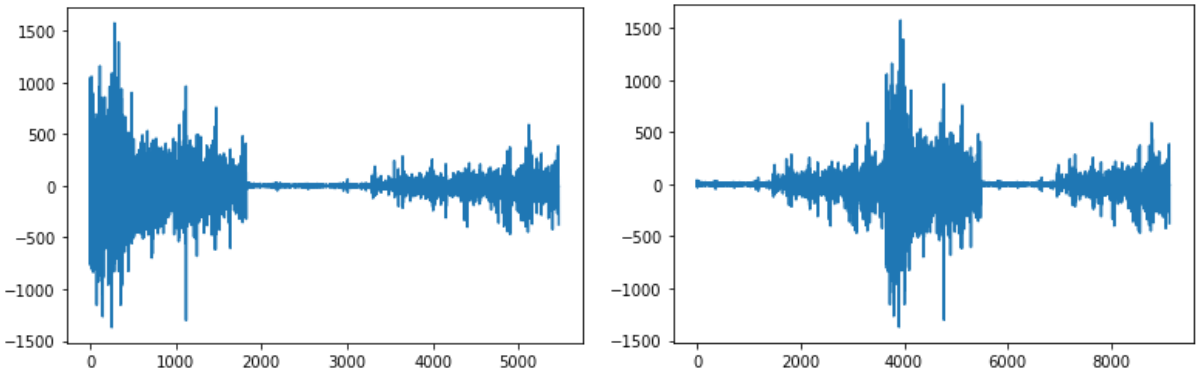


Figura 63– Cerveja *Premium*, diferenciação remoção de tendência e sazonalidade respectivamente. Fonte: Elaborado pelo autor.

A aplicação de diferenciação para remoção de tendência e sazonalidade nas tres séries com p-value maior que 0.05 fez com que visivelmente as séries fiquem lineares. Após tal transformação é aplicado novamente o teste de *Dickey Fuller* com os dados já diferenciados. Os novos resultados de p-value podem ser observados na tabela abaixo, na coluna “Com Diff”.

Série Temporal	Sem Diff p-value	Com Diff p-value
Cerveja Core	0.000062	-
Cerveja Retornável	0.000693	-
Cerveja Litro	0.000012	-
Cerveja Inteira	0.051001	0.000000
Cerveja Meia	0.001055	-
Cerveja Descartável	0.000008	-
Cerveja Puro Malte	0.332301	0.000000
Cerveja Premium	0.114270	0.000000

Figura 63– Aplicação de *Dickey Fuller* nas séries com diferenciação. Fonte: Elaborado pelo autor.

3.2 APLICAÇÃO DE KPSS

Teste criado por Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt e Yongcheol Shin, denominado teste KPSS devido a seus nomes, tem por finalidade determinar estacionariedade em uma série temporal. Sua aplicabilidade é melhor em séries temporias com tendência, obtendo melhor resultado quando comparado com o teste de *Dickey Fuller* (Carrion-i-Silvestre et al., 2001).

O teste de KPSS é aplicado a todas as séries temporais mencionadas no capítulo 2, caso haja necessidade utiliza-se da ferramenta de diferenciação para obter-se valores de p que aceitem a hipótese nula como verdadeira. A biblioteca utiliza é a "kpss" dentro do pacote "statsmodels" Tomam-se as seguintes hipóteses H0 e H1:

H0 (p-value > 0.05): Adoção da hipótese nula, a série é estacionária;

H1 (p-value ≤ 0.05): Rejeita-se a hipotese nula, a série não é estacionária.

	Dickey Fuller		*KPSS*	
	Sem Diff	**Com Diff**	**Sem Diff**	**Com Diff**
Série Temporal	**p-value**	**p-value**	**p-value**	**p-value**
Cerveja Core	0.000062	-	0,0135	0,1000
Cerveja Retornável	0.000693	-	0,0187	0,1000
Cerveja Litro	0.000012	-	0,0806	-
Cerveja Inteira	0.051001	0.000000	0,0100	0,1587
Cerveja Meia	0.001055	-	0,0100	0,2890
Cerveja Descartável	0.000008	-	0,1240	-
Cerveja Puro Malte	0.332301	0.000000	0,0100	0,1985
Cerveja Premium	0.114270	0.000000	0,0100	0,1000

Figura 64– Aplicação de KPSS nas séries. Fonte: Elaborado pelo autor.

Os valores destacados em verde na figura acima, significam que obteve-se valor nos testes estatisticos para considerar a série temporal estacionaria. Nota-se que nas séries cerveja inteira, puro malte e *premiu* houve em ambos os testes a conclusão de não estacionareidade, fato que ocorre por haver tendência positiva ou negativa bem presente nos modelos.

Outro ponto importante a ser destacado é o fato de nas séries cerveja *core*, cerveja retornável, cerveja meia houve divergência na conclusão dos testes, sendo as séries consideradas estacionárias por *Dickey Fuller* e não estacionárias por KPSS, quando isso ocorre considera-se série temporal de estacionareidade diferencial, ou seja, com necessidade de diferenciação para remoção de elementos de tendência e variância.

3.3 TESTE DE MANN-KENDALL

O teste de Mann-Kendall é utilizado para analisar dados de séries temporais para tendências consistentemente crescente ou decrescentes. É um teste não paramétrico, que significa que funciona para todas as distribuições, ou seja, os dados não necessitam atender à suposição de normalidade (Hussain & Mahmud, 2019).

Se os dados conterem correlação serial podem afetar o valor de p. Para superar este problema propõe-se testes de *Mann-Kendall* modificados. Diversos testes foram propostos por pesquisadores diferentes, para esta análise é proposto a utilização do método "*seasonal MK",* que foi desenvolvido para remover o efeito da sazonalidade (Libiseller & Grimvall, 2002). É realizada uma comparação de aplicação da ferramenta original versus a ferramenta de remoção de sazonalidade.

O teste de *Mann-Kendall* é aplicado a todas as séries temporais mencionadas no capítulo 2. A biblioteca utiliza é a "pymannkendall" dentro do pacote "numpy" Tomam-se as seguintes hipóteses H0 e H1:

H0 (p-value > 0.05): As observações da série são independentes e identicamente distribuídas. (Não há tendência);

H1 (p-value ≤ 0.05): As observações da série possuem tendência monotônica no tempo. (Há tendência).

Mann-Kendall

	Original Test		Seasonal Test	
Série Temporal	**p-value**	**Tendência**	**p-value**	**Tendência**
Cerveja Core	0.00000	decreasing	0.00000	decreasing
Cerveja Retornável	0.00000	decreasing	0.00000	decreasing
Cerveja Litro	0.20074	no trend	0.04828	decreasing
Cerveja Inteira	0.00000	decreasing	0.00000	decreasing
Cerveja Meia	6,66E-11	increasing	1,20E-11	increasing
Cerveja Descartável	0.29521	no trend	0.070838	no trend
Cerveja Puro Malte	0.00000	increasing	0.00000	increasing
Cerveja Premium	0.00000	increasing	0.00000	increasing

Figura 65– Aplicação de *Mann-Kendall* nas séries. Fonte: Elaborado pelo autor.

Os resultados apresentados acima que houve apenas uma modificação no resultado do teste na série temporal cerveja litro, onde no teste original a série foi considerada sem tendência, e após alteração para o teste com remoção de sazonalidade observa-se que obteve tendência negativa. A série cerveja descartável foi considerada pelo teste sem tendência em ambas as aplicações.

Referencias Artigo 2

Bakker, B., Linaker, F., & Schmidhuber, J. (2002). Reinforcement learning in partially observable mobile robot domains using unsupervised event extraction. *IEEE/RSJ International Conference on Intelligent Robots and System*, *1*, 938–943. https://doi.org/10.1109/IRDS.2002.1041511

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Data vs Instinct: Perfecting Global Sales Performance. (2012). 25.

Gers, F. A., Eck, D., & Schmidhuber, J. (2001). *Applying LSTM to Time Series Predictable Through Time-Window Approaches*. 8.

Gonzalez-Dominguez, J., Lopez-Moreno, I., Sak, H., Gonzalez-Rodriguez, J., & Moreno, P. J. (2014). *Automatic Language Identification using Long Short-Term Memory Recurrent Neural Networks*. 5.

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, *23*(2), 201–220. https://doi.org/10.1109/JSAC.2004.839380

Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. 1735–1780.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À ENGENHARIA E GESTÃO

DANIEL NOCERA DE CAMPOS

ARTIGO 3: APLICAÇÃO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINA APLICADO A PREVISÃO DE VENDAS FUTURAS

Campos dos Goytacazes/RJ

(2020)

ARTIGO 3: APLICAÇÃO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINA APLICADO A PREVISÃO DE VENDAS FUTURAS

RESUMO

Segundo a revista *Beer Art* o mercado de consumo de cerveja teve um crescimento aproximado de 36% no último ano, com esse segmento representando 2% do PIB. Segundo o IBGE 21% das empresas tem declarado falência após o primeiro ano em atividade e menos da meta sobrevivo após o quarto ano de existência. Esse alto percentual se dá principalmente pelo fato de haver falhas no planejamento de negócio, tais como planejamento de estoque. Essas empresas utilizam-se de médias e medianas para o dimensionamento de estoque, porem tais métricas não são recomendadas por não obterem um resultado satisfatório, principalmente pela não absorção de elementos sazonais e de tendência. Neste contexto este trabalho propõe a aplicação de diferentes modelos de inteligência artificial para obter-se uma previsão mais precisa dos valores de eventos futuros. O resultado final foi a comparação de tais modelos (janelas deslizantes, ARIMA e *Prophet)* com sugestão de trabalhos futuros para otimização do modelo.

Palavras-chave: **Aprendizado de Máquina, Estoque, Cerveja.**

ABSTRACT

According to Beer Art magazine, the beer consumption market grew by approximately 36% in the last year, with this segment representing 2% of GDP. According to the IBGE, 21% of companies have declared bankruptcy after the first year in activity and less than the surviving target after the fourth year of existence. This high percentage is mainly due to the fact that there are flaws in business planning, such as inventory planning. These companies use averages and medians for sizing inventory, but such metrics are not recommended because they do not obtain a satisfactory result, mainly due to the non-absorption of seasonal and trend elements. In this context, this work proposes the application of different models of artificial intelligence to obtain a more accurate forecast of the values of future events. The final result was the comparison of such models (sliding windows, ARIMA and Prophet) with suggestions for future work to optimize the model.

Key-Words: Machine Learning, Stock, Beer.

1. INTRODUÇÃO

Segundo a revista *Beer Art* o mercado de consumo de cerveja teve um crescimento de 36% no último ano, representando 2% do PIB nacional. Esta indústria gerou R$ 77 bilhões de reais de faturamento no último exercício fechado e contribuiu com R$ 25 bilhões em impostos. Pelas características inerentes ao negócio, as cervejarias impactam positivamente outros setores da economia como, agronegócio, transporte, energia, vidro etc. Gerando número de empregados diretos e indiretos de aproximadamente 2,7 milhões de pessoas (*SINDICERV*, 2020).

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), 21% das empresas quebram após primeiro ano em atividade e menos da metade sobrevive após completados quatro anos. Como um dos motivos principais está a falha no planejamento de negócio de empresas de médio e pequeno porte (Lazarini, 2019).

Um plano de negócio é um projeto de como sua empresa funcionará, desde do planejamento de *marketing*, logística, estoque e financeiro (*SEBRAE*, 2019). O dimensionamento e gestão do estoque é peça fundamental para garantir o sucesso de uma empresa. Isto se faz mais necessário com produtos de alto giro e empresas com um grande portifólio de produtos (Dewes et al., 2018).

Muitas empresas utilizam métricas de baixa performance para o dimensionamento de estoque, tais como médias e medianas. Tais métricas são pouco recomendadas por justamente obterem resultados satisfatórios geralmente em problemas com dados uniformes, e terem baixo desempenho em modelos que apresentam tendência e a variante sazonal. Técnicas de *machine learning* vem sendo amplamente utilizadas para previsão de vendas futuras e por consequência um dimensionamento mais preciso de estoque (Chong et al., 2017).

Neste contexto este artigo propõe a comparação de resultados de aplicações de algoritmos de *machine learning*, com o objetivo de prever vendas futuras de uma distribuidora de cerveja localizada na região dos lagos. Utiliza-se do método de janelas deslizantes como base de comparação com os demais métodos.

2. APLICAÇÃO DE ALGORITMOS DE MACHINE LEARNING

Um algoritmo de aprendizagem detecta automaticamente padrões em um conjunto de dados e pode ser utilizado para tomada de decisões em situações de incertezas (Murphy, 2012). Basicamente os algoritmos se dividem em supervisonados e não-supervisonados, aprendizagem reforçada e semi-supervionada. O principal objetivo de cada uma destas técnicas é derivar de um conjunto de dados, de forma indutiva, um modelo capaz de prever novos dados. É de suma importância definir informações de entrada e saída.

No trabalho em questão é definido um modelo inicial que tem como principal objetivo servir como comparação para aplicação de métodos de aprendizagem de máquina mais refinados. Além disso as os tipos de cervejas analisadas são: *core*, barril, retornável, descartável e puro malte.

2.1 PREVISÃO DE VENDA COM JANELAS DESLIZANTES (MÉDIA MÓVEL)

Na modelagem de algoritmos de previsão de vendas é muito importante a definição de um modelo que seja classificado como *baseline* para modelos mais complexos (Barboza et al., 2017). É importente que este modelo base seja preferencialmente o mais simples possível. Para construção deste modelo é utilizado o modelo de previsão com média móvel, ou previsão com janelas deslizantes.

A modelagem com o método de janela deslizante consiste em definir o tamanho da janela e calcular a média dos elementos contidos dentro da janela para realizar a previsão do próximo elemento. A idéia é que a esta janela consiga varrer todos os dados e realize as previsões.

Para a modelagem no trabalho em questão são relizados testes com diferentes tamanhos de janela, utilizando a métrica de RMSE para verificar se houve melhoria na disperção entre o valor predito e o valor real. Inicialmente foi definida uma janela com valor de dois. Como pode ser observado abaixo, o tamanho ideal da janela é de oito elemento, obtendo o menor RMSE dentre as janelas testadas.

Tamanho da Janela	Cerveja Genérica RMSE	Cerveja Barril RMSE	Cerveja Retor RMSE	Cerveja Desc RMSE	Cerveja PM RMSE
1	NA	NA	NA	NA	NA
2	955,29	30,1	703,69	444,65	9,9
3	931,38	29,29	677,17	437,51	10,2
4	917,72	28,72	661,03	435,72	10,2
5	900,67	28,23	654,03	425,52	9,9
6	855,8	26,82	618,29	413,08	9,28
7	781,78	24,73	559,31	393,81	8,48
8	772,17	24,28	552,31	390,01	8,26
9	789,86	24,84	568,02	393,39	8,45
10	803,38	25,31	576,04	398,59	8,72

Figura 66– Resultado de RMSE para tamanhos de janelas diferentes. Fonte: Elaborado pelo autor.

O código utilizado para realizar pode ser visualizado abaixo, com programação realizada em linguem *Python* e utilizando o *Google Colab*.

```
#Previsão com Média Móvel
X = cerveja_generica.values
window = 8
history = [X[i] for i in range(window)]
teste = [X[i] for i in range(window, len(X))]
predicoes = []

#Criando o Loop
for t in range(len(teste)):
  lenght = len(history)
  valor_predito = mean([history[i] for i in range (lenght -
 window, lenght)])
  valor_real = teste[t]

#Alimentando a lista de predições
  predicoes.append(valor_predito)

#Alimentando a lista de History
  history.append(valor_real)

#Imprimindo valor predito e real
  print('Valor Predito=%f, Valor Real=%f' % (valor_predito, valor_real)
)
rmse = sqrt(mean_squared_error(teste,predicoes))
print('Métrica RMSE: %3f' % rmse)
```

O gráfico do valor pedrito e valor real é plotado, sendo a linha azul o valor real e a linha vermelha o valor pedrito. Como se trata da metodologia mais simples de *machine learning* ela é definida como o modelo *baseline.* As demais modelagens que serão apresentadas no decorrer do trabalho tem o objetivo de obter melhores resultados quando comparados ao *baseline*.

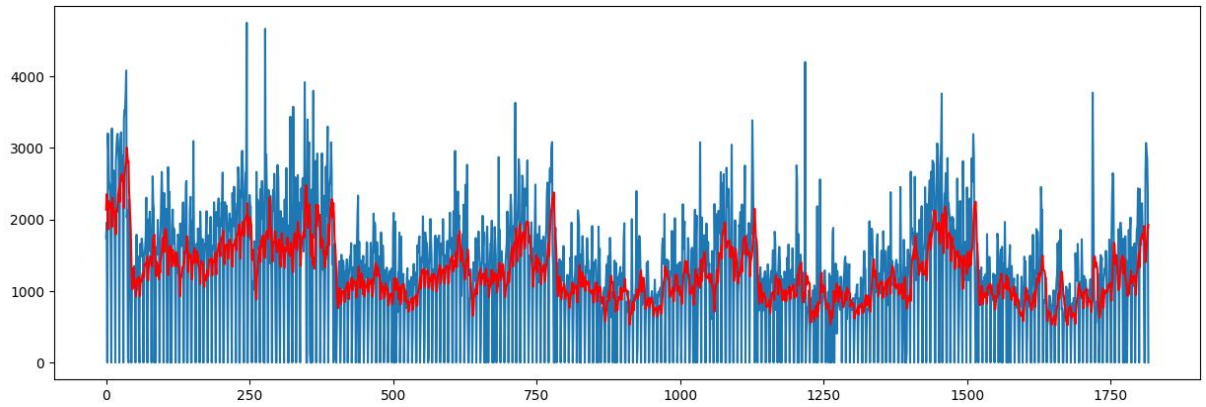


Figura 67– Comparação entre valor real e valor predito. Fonte: Elaborado pelo autor.

2.2 PREVISÃO DE VENDA COM ARIMA

A modelagem utilizando ARIMA (*Autoregression Integrated Moving Average)* é um método estatístico aplicado em um modelo linear que utiliza auto regressão e um modelo de média móvel para previsão de series temporais. Um modelo linear é construído incluindo um número especifico de termos e o dado é preparado por um nível de diferenciação afim de tornar este estacionário (Shi et al., 2018).

Este método é capaz de utilizar os erros residuais (diferença entre o valor real e o valor predito). Estes erros residuais são estruturas temporais que podem ser modeladas. O modelo utiliza essas informações para corrigir erros futuros e ajustar ainda mais o valor predito (J. Wang & Hu, 2015). É importante ressaltar que o modelo de média móvel contido no ARIMA é diferente do utilizado no *baseline*, pois utiliza-se dos erros residuais para corrigir suas previsões.

Dentro do modelo é possível ligar e desligar os parâmetros de auto regressão (p), integração (d) e média móvel (q). Para um bom funcionamento do método é necessário que a serie temporal seja estacionaria, por isso que o ARIMA contém o I para tentar diferenciar os dados analisados.

Parâmetros ARIMA

P: Número de *lags* que devem ser incluídos no modelo;

D: Número de vezes que as observações serão diferenciadas;

Q: O tamanho da janela da média móvel, ou também chamada de ordem da média móvel;

Para realizar o desligamento de algum dos três parâmetros basta colocar o valor de zero para o parâmetro em questão.

Inicialmente faz necessário realizar uma análise da auto correlação entre os *lags* das séries temporais. Para se obter um parâmetro P inicial é analisado nos cinco primeiros *lags* se o grau de auto correlação é o mais alto da série e acima do nível de confiança.

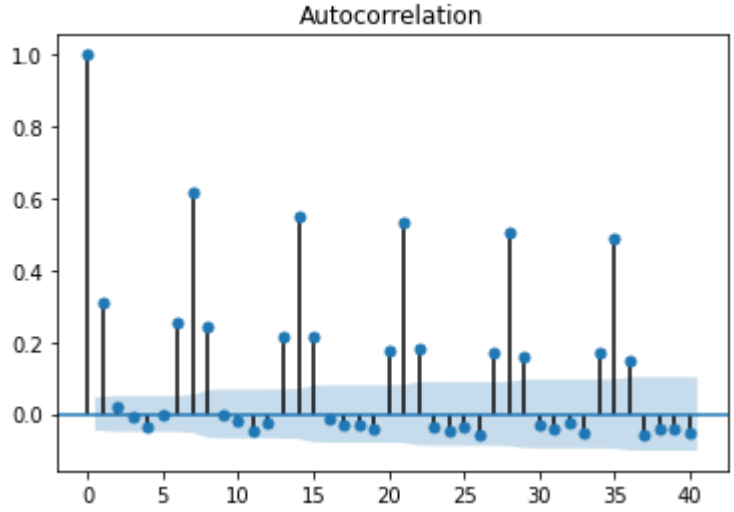


Figura 68– Gráfico de auto correlação entre os *lags* – cerveja core. Fonte: Elaborado pelo autor.

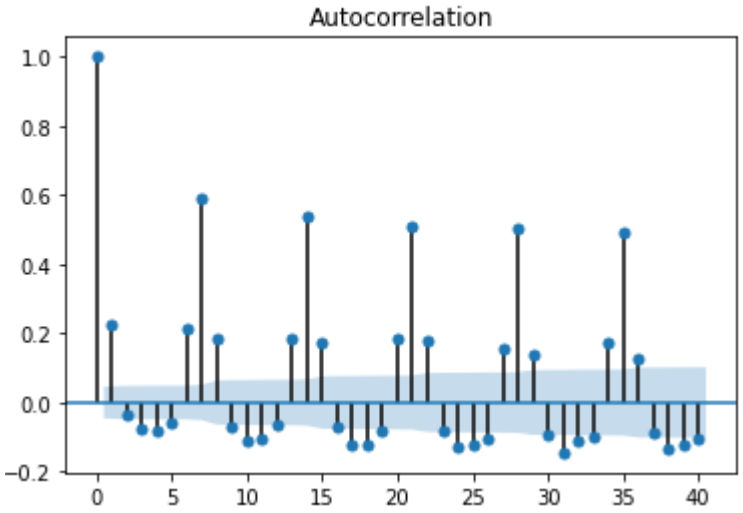


Figura 69– Gráfico de auto correlação entre os *lags* – cerveja barril. Fonte: Elaborado pelo autor.

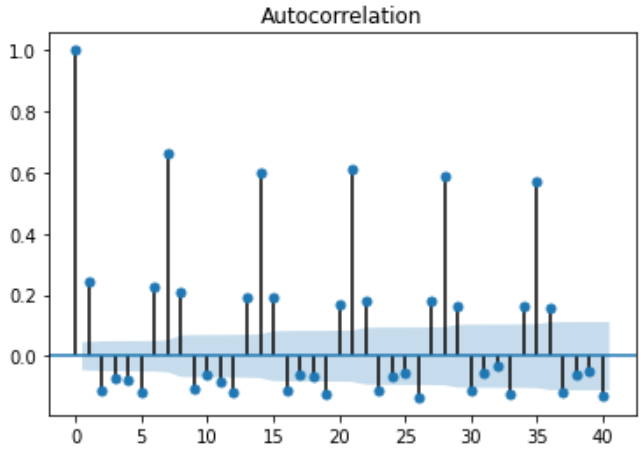


Figura 70– Gráfico de auto correlação entre os *lags* – cerveja retornável. Fonte: Elaborado pelo autor.

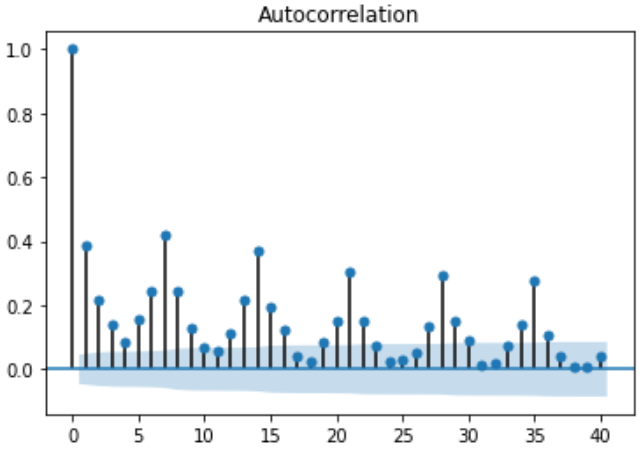


Figura 71– Gráfico de auto correlação entre os *lags* – cerveja descartável. Fonte: Elaborado pelo autor.

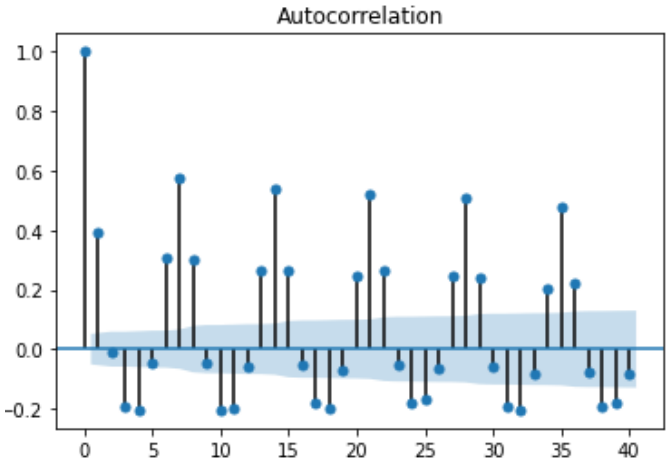


Figura 72– Gráfico de auto correlação entre os *lags* – cerveja puro malte. Fonte: Elaborado pelo autor.

Em todas as séries temporais os *lags* maior auto correlação estão entre o *lag* 1 e o *lag* 5, portanto inicialmente será escolhido cinco para o parâmetro P. Além disso o parâmetro D será mantido em um, ou seja, estará ligado. A definição do parâmetro Q será baseada no tamanho da janela com menor RMSE na modelagem da seção 2.1, ou seja, o valor de Q será oito.

Código de programação

```
import pandas as pd
from pandas.plotting import autocorrelation_plot
from matplotlib import pyplot

#autocorralation plot
autocorrelation_plot(cerveja_generica)
pyplot.show()
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
from matplotlib import pyplot
plot_acf(cerveja_generica, lags=40)
plot_acf(cerveja_barril, lags=40)
plot_acf(cerveja_retornavel, lags=40)
plot_acf(cerveja_descartavel, lags=40)
plot_acf(cerveja_puro_malte, lags=40)
pyplot.show()

#fit model
from statsmodels.tsa.arima_model import ARIMA
from matplotlib import pyplot
model = ARIMA(cerveja_generica, order=(5,1,8))
model_fit = model.fit()
print(model_fit.summary())

#analisando os residuos
from pandas import DataFrame
residuals = DataFrame(model_fit.resid)
plt.figure(figsize=(15,5), dpi=100)
pyplot.plot(residuals)
plt.show()
residuals.plot(kind='kde')
pyplot.show()
print(residuals.describe())
```

A análise de tendência dos resíduos é fundamental para verificar se os parâmetros iniciais foram bem escolhidos, uma vez que havendo tendência no residual significa que o modelo ARIMA não foi capaz de captura-la para a análise.

Na modelagem em questão observa-se que não há sinais de que os resíduos estejam com a variante de tendência, além disso é possível observar através da distribuição normal que não lateralidade, mas sim centralidade da série temporal, significando que o modelo em questão consegui absorver bem suas respectivas tendências, principalmente no grupo de cervejas puro malte. Esta por sua

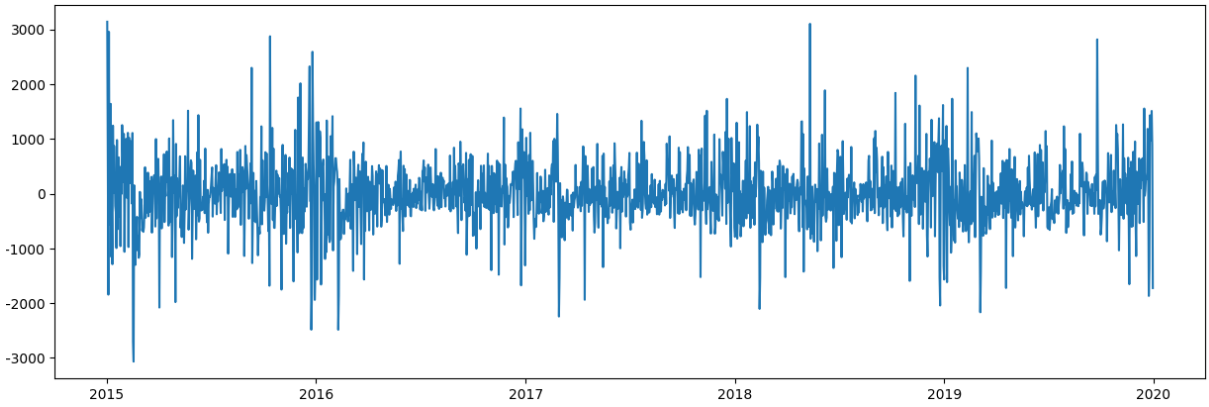


Figura 73– Gráfico dos erros residuais – cerveja *core*. Fonte: Elaborado pelo autor.

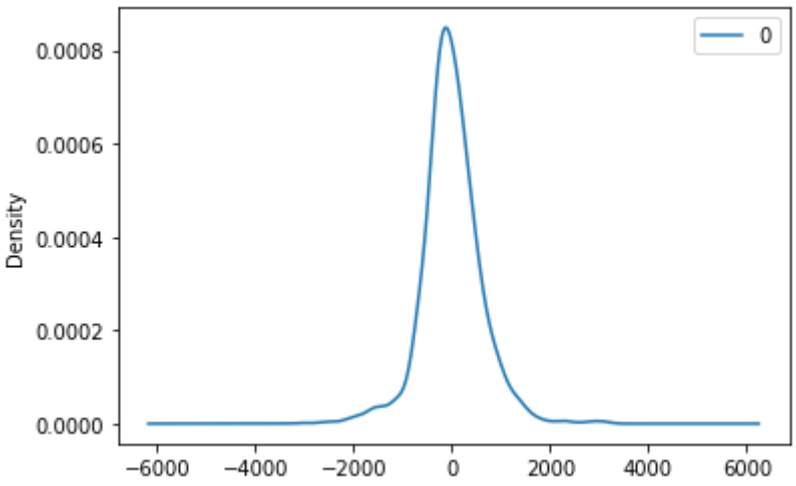


Figura 74– Distribuição normal dos resíduos – cerveja *core*. Fonte: Elaborado pelo autor.

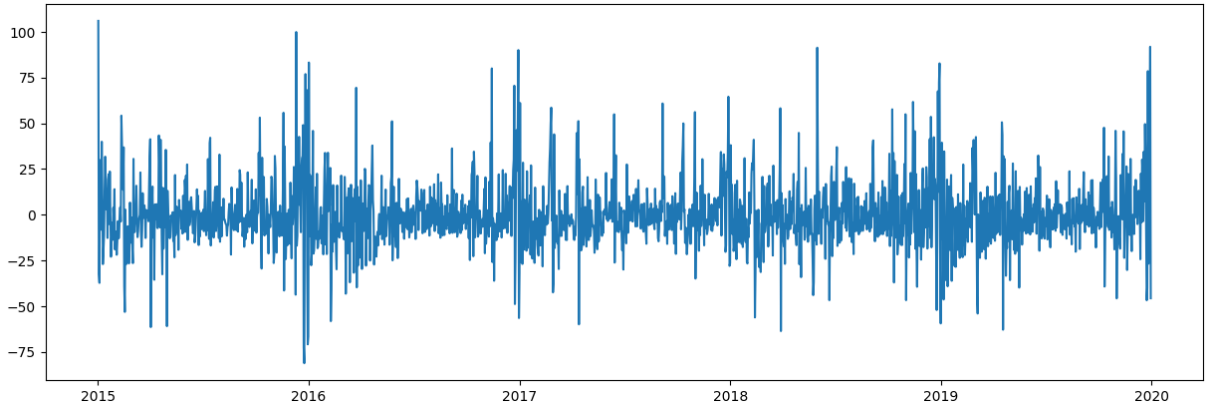


Figura 75– Gráfico dos erros residuais – cerveja barril. Fonte: Elaborado pelo autor.

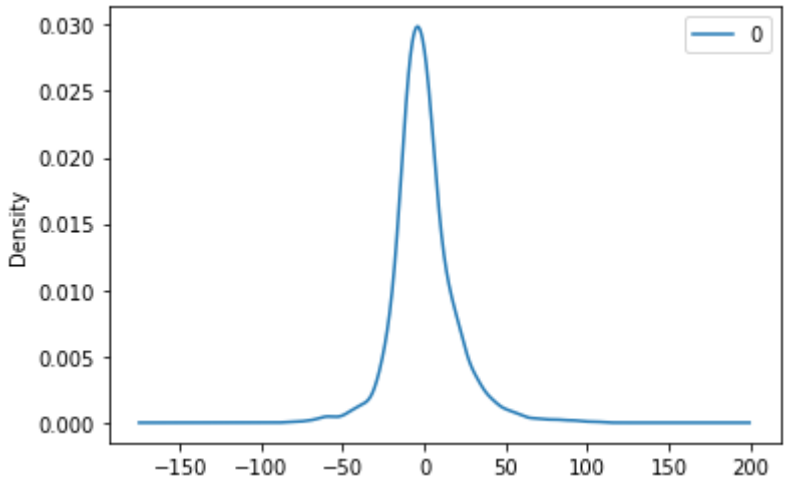


Figura 76– Distribuição normal dos resíduos – cerveja barril. Fonte: Elaborado pelo autor.

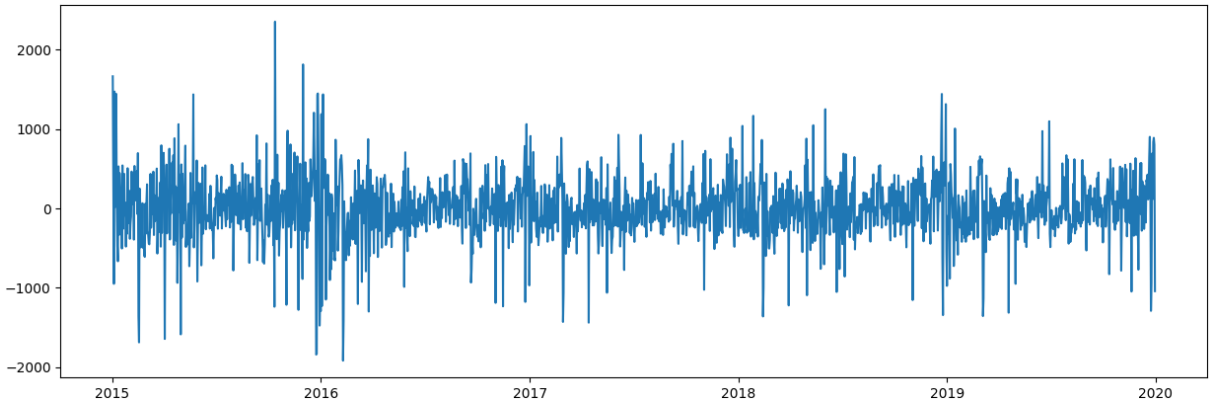


Figura 77– Gráfico dos erros residuais – cerveja retornável Fonte: Elaborado pelo autor.

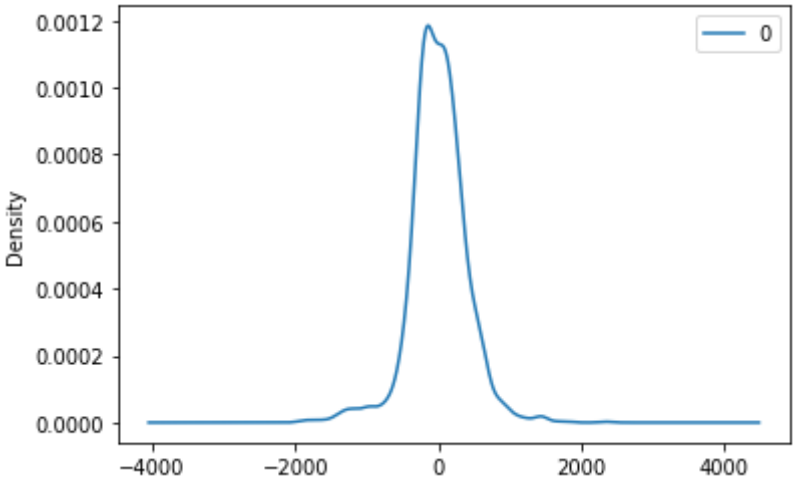


Figura 78– Distribuição normal dos resíduos – cerveja retornável. Fonte: Elaborado pelo autor.

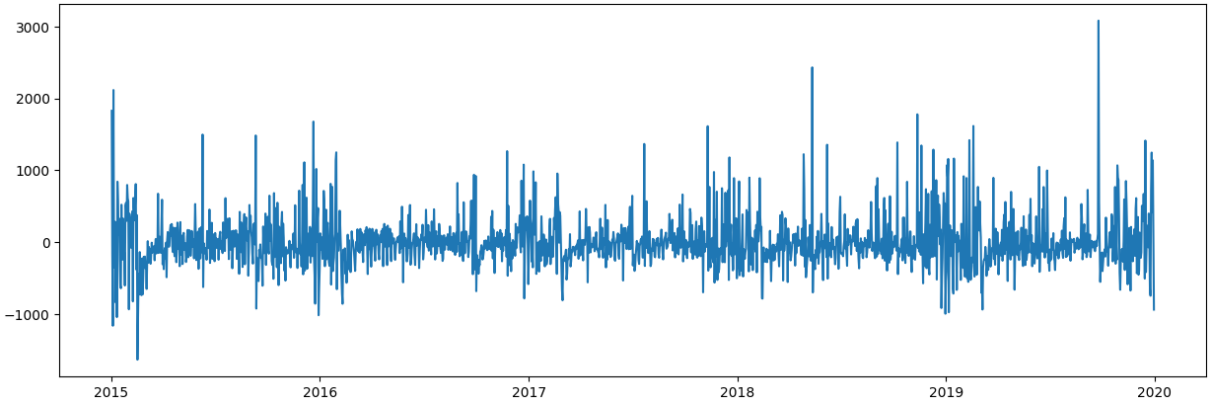
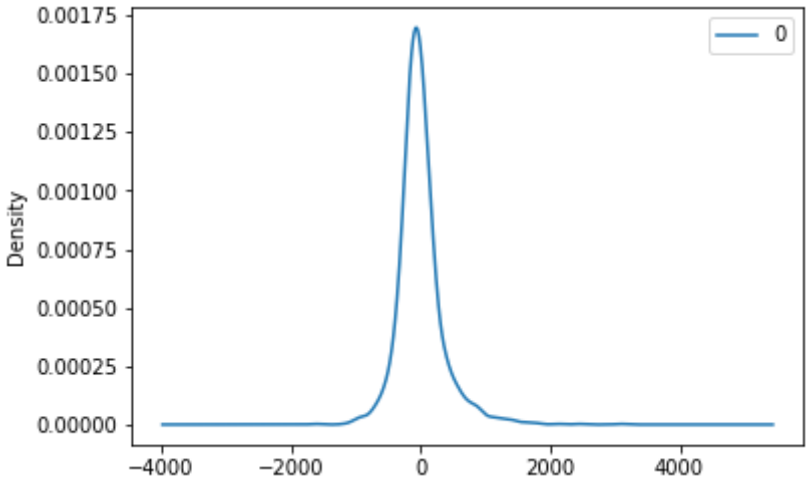


Figura 79 - Gráfico dos erros residuais – cerveja descartável. Fonte: Elaborado pelo autor.

Figura 80– Distribuição normal dos resíduos – cerveja descartável.



Fonte: Elaborado pelo autor.

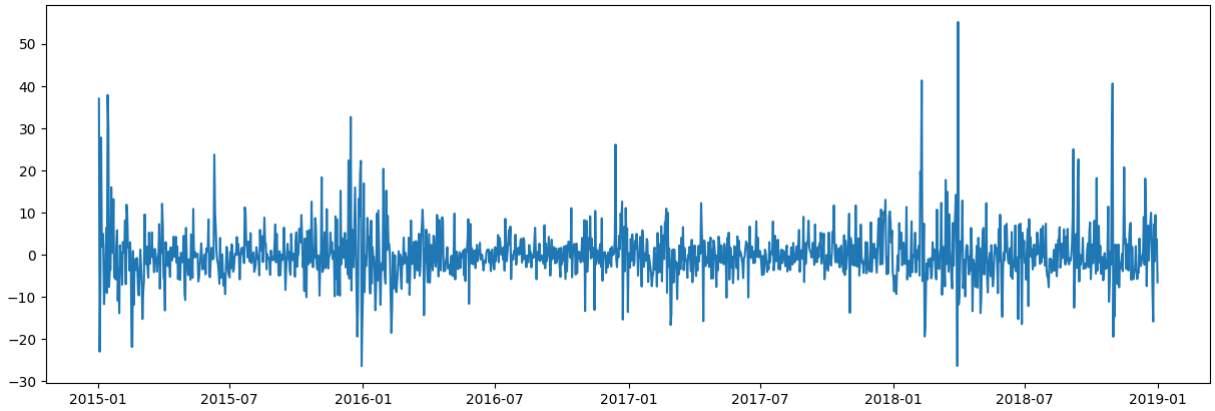


Figura 81 - Gráfico dos erros residuais – cerveja puro malte. Fonte: Elaborado pelo autor.

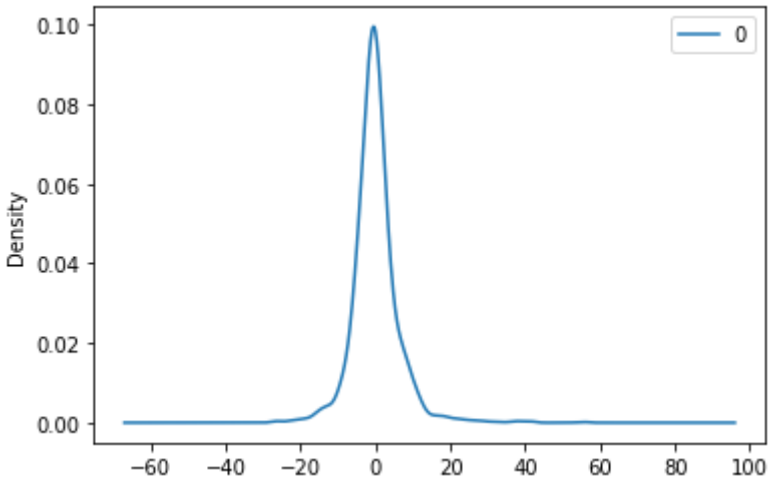


Figura 82– Distribuição normal dos resíduos – cerveja puro malte. Fonte: Elaborado pelo autor.

2.3 PREVISÃO DE VENDA COM ARIMA - CONCLUSÃO

A modelagem com utilização da metodologia ARIMA mostrou uma melhora média de 12,2% quando comparado ao modelo base com janelas deslizantes. Se trata de uma metodologia com mais rebuscamento para inserção de número de *lags* e quantidade de diferenciações. O resultado pode ser analisado na tabela abaixo.

RMSE			
Item	Modelo Base Média Móvel Janela = 8	ARIMA	% Melhora
Cerveja CORE	772,17	661,14	14,4%
Cerveja Barril	24,28	21,78	10,3%
Cerveja Retornável	552,31	456,39	17,4%
Cerveja Descartável	390,01	366,58	6,0%
Cerveja Puro Malte	8,26	7,21	12,7%

Figura 83– Tabela de comparação de desempenho do RMSE. Fonte: Elaborado pelo autor.

O grupo Cerveja Retornável foi o que obteve a maior melhoria percentual, resultando em uma melhora de 17,4%. Já o grupo Cerveja descartável foi o que obteve pior melhora percentual (6% em comparação ao modelo base).

Para trabalhos futuros, sugere-se utilizar a ferramenta de *tunning* para otimizar a configuração dos parâmetros p, d e q. A aplicação do *tunning* é realizada na modelagem a seguir, mostrando melhora dos resultados obtidos pelo modelo. Além disso existe a possibilidade de aplicação de uma metodologia similar ao ARIMA porém com a inclusão de elemento sazonal, neste caso SARIMA.

O SARIMA é utilizado para séries não estacionárias com facilidade em identificar tendência e sazonalidade, com possibilidade de modelar eventos de aumento ou diminuição das vendas devido a feriados, inverno, verão e dentre outras variáveis.

2.4 PREVISÃO DE VENDA COM PROPHET

O *Facebook* desenvolveu um algoritmo chamado *Prophet* que é uma ferramenta disponivel nas plataformas R e *Python.* Sua vantagem em comparação com a aplicação de métodos estatisticos é que não demanda conhecimentos profundos em modelagem de séries temporais. Produzir previsões de alta qualidade não é um problema fácil para as máquinas ou para a maioria dos analistas. Foram observados dois temas principais na prática de criar uma previsão de negócios:

- Técnicas de previsão completamente automáticas podem ser frágeis e muitas vezes são inflexíveis demais para incorporar suposições ou heurísticas úteis;
- Os analistas que podem produzir previsões de alta qualidade são muito raros, pois a previsão é uma habilidade especializada em ciência de dados que requer experiência substancial (Taylor & Letham, 2017).

O modelo *Prophet* utiliza-se de três componentes principais para realizar a modelagem de séries temporias seguindo a seguinte equação:

$$y(t) = g(t) + s(t) + h(t)$$

- g(t): curva de crescimento linear ou logística por partes para modelar mudanças não periódicas em séries temporais;
- s(t): mudanças periódicas (por exemplo, sazonalidade semanal / anual);
- h(t): efeitos de feriados (usuário fornecido) com horários irregulares.

O modelo pode ser considerado *overfitting* ou *underfitting* ao trabalhar com a componente de tendência. Neste ponto a grande vantagem em utilizar esta metodologia está em conseguir alterar um parãmetro com o objetivo de permitir que o *Prophet* se torne mais flexivel a bruscas variações de tendências. Um exemplo de aplicação no projeto é que em feriados e vésperas de feriado há um grande salto nas vendas.

Para haver uma base de comparação do erros dos modelos anteriores utilizaremos o RMSE (Root Mean Square Deviation) como métrica de medição de erro entre os valores previstos pelo modelo e os valores reais. Para Chen et al. (2018) esta métrica é comumente utilizada em previsões de séries temporias priorizando o menor valor possível de RMSE para o modelo.

A modelagem inicial será feita sem alteração dos parâmetros de sazonalidade diária, semanal e anual, métodos de interpetração sazonal (multiplicativo ou aditivo), flexibilização do modelo para mudanças brucas de tendência e sem definição de feriados e datas importantes. O código utilizado para realizar pode ser visualizado abaixo, com programação realizada em linguem *Python* e utilizando o *Google Colab*.

```
from fbprophet import Prophet
dataset = cerveja_generica
dataframe = pd.DataFrame({'ds': dataset.index, 'y': dataset.values})

#definindo modelo de previsão
model = Prophet(interval_width=0.95)
#treinando o modelo
model.fit(dataframe)

#criando dataframe para previsão dos próximos 90 dias com frequência di
ária
futuro = model.make_future_dataframe(periods = 365, freq='D')

#modelo de previsão de valores futuros
saida = model.predict(futuro)

#validandos os dados gerados com o modelo Prophet com os dados reais
validation_df = pd.DataFrame({'ds': dataset.index})

saida_validation = model.predict(validation_df)


#Calculando RMSE com Prophet
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(dataset.values,saida_validation.yhat))
print('Métrica RMSE: %3f' % rmse)
```

O resultado do modelo é fornecido em forma de limites médios, limite de alta e limite de baixa, de acordo com o valor de confiança estabelecido. Como pode ser visto abaixo foram

impressos os últimos sete valores da previsão da série temporal. O valor que é considerado para o cálculo do RMSE é *yhat.*

	ds	yhat	yhat_lower	yhat_upper
2183	2020-12-24	2064.660021	953.507085	3236.249762
2184	2020-12-25	2180.014626	1083.165960	3305.881864
2185	2020-12-26	1942.055600	840.771268	3068.979210
2186	2020-12-27	588.539725	-478.875073	1727.146985
2187	2020-12-28	1546.795500	349.200728	2674.073414
2188	2020-12-29	1912.317506	848.176968	3103.947434
2189	2020-12-30	1964.058644	909.810953	3034.67863

Figura 83– Valores preditos utilizando *Prophet* – cerveja genérica. Fonte: Elaborado pelo autor.

É possível plotar um gráfico para vizualizar as previsões de 365 dias para frente da última data. Os pontos em preto são os valores reais, a curva azul escuro é a predição do modelo. A seleção em vermelho se trata das predições futuras realizadas pelo modelo.

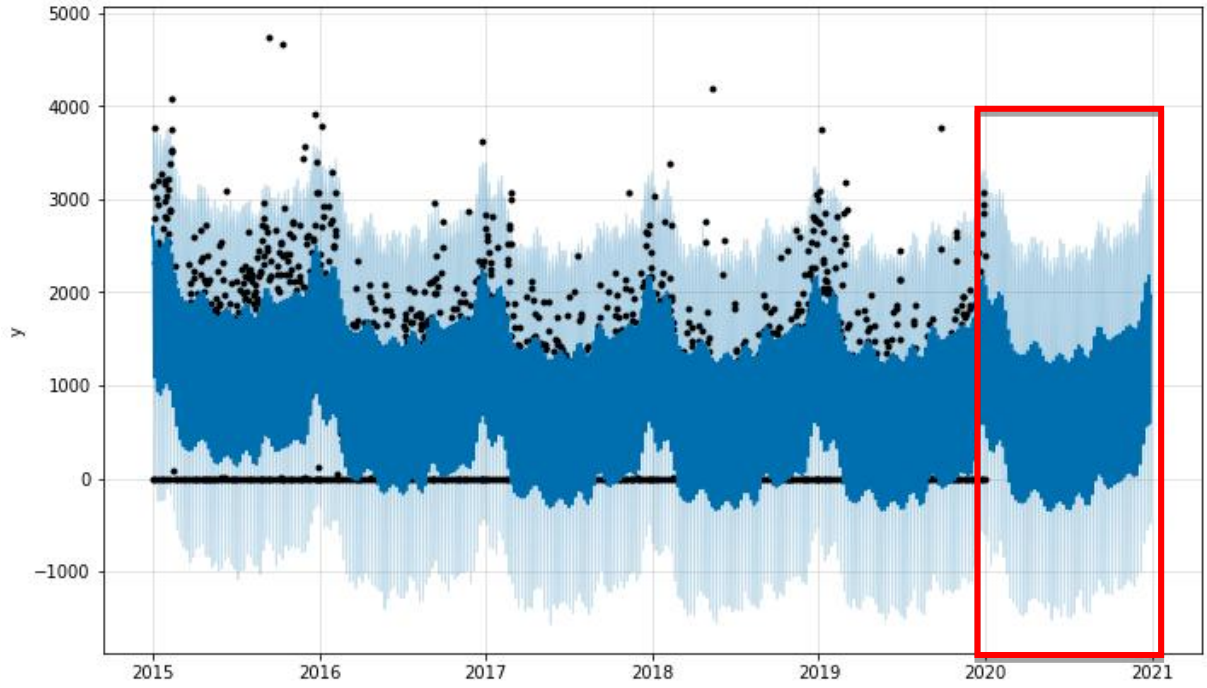


Figura 84– Valores preditos utilizando *Prophet* – cerveja genérica. Fonte: Elaborado pelo autor.

Outro dado importante é analisar as componentes de tendência, sazonalidade diária e sazonalidade anual. Esta análise é importante para verificar se o modelo foi capaz de capturar toda dinâmica da série temporal. Abaixo é plotado os gráficos de tendência e sazonalidade da série temporal cerveja genérica.

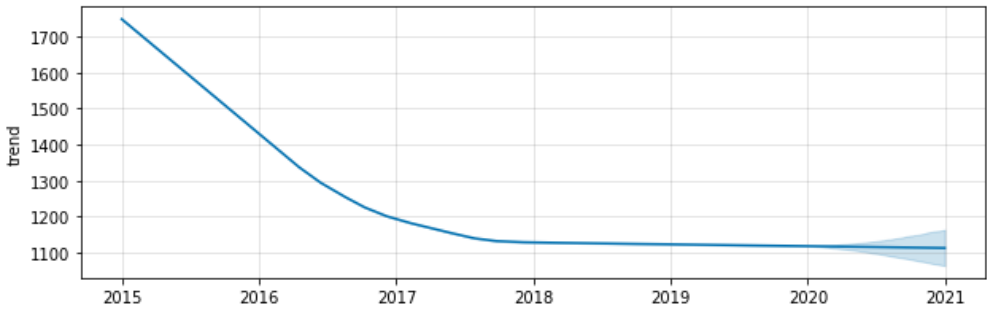


Figura 85– Decomposição da série temporal – cerveja genérica. Fonte: Elaborado pelo autor.

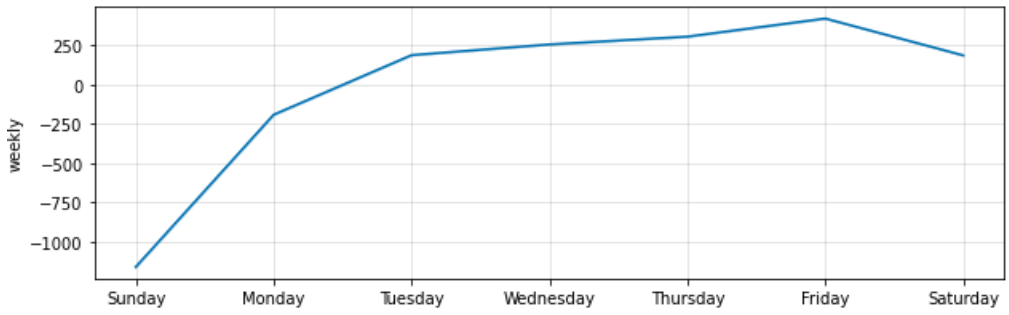


Figura 86– Decomposição da série temporal – cerveja genérica. Fonte: Elaborado pelo autor.

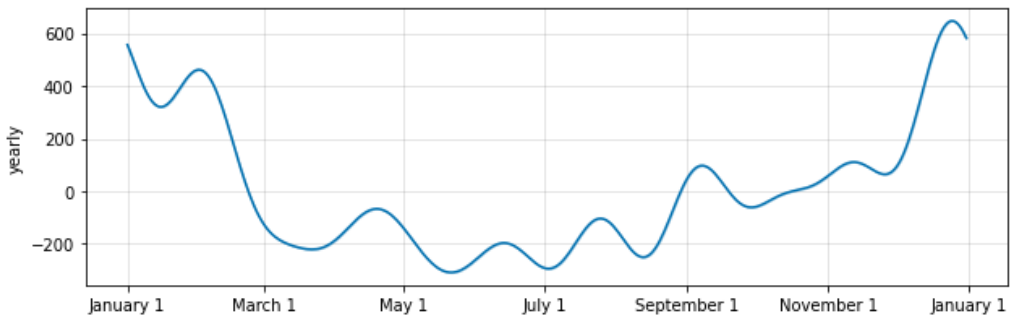


Figura 87– Decomposição da série temporal – cerveja genérica. Fonte: Elaborado pelo autor.

As modelagens de cerveja barril, cerveja retornável, cerveja descartável e cerveja puro malte são apresentadas abaixo de forma mais resumida.

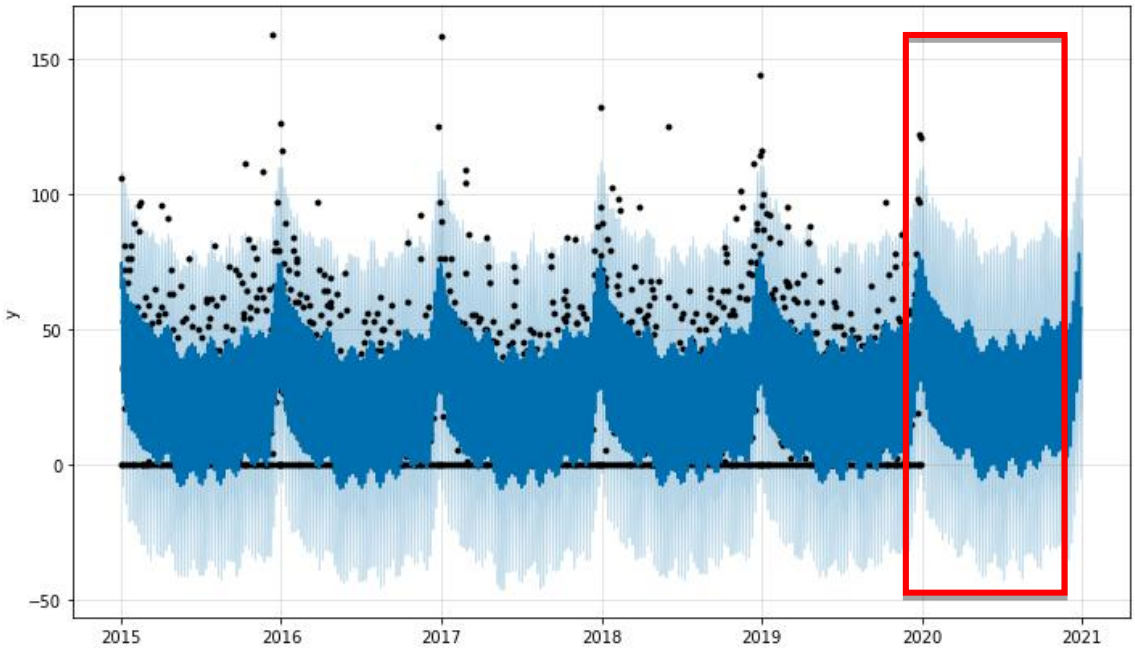


Figura 88– Valores preditos utilizando *Prophet* – cerveja barril. Fonte: Elaborado pelo autor.

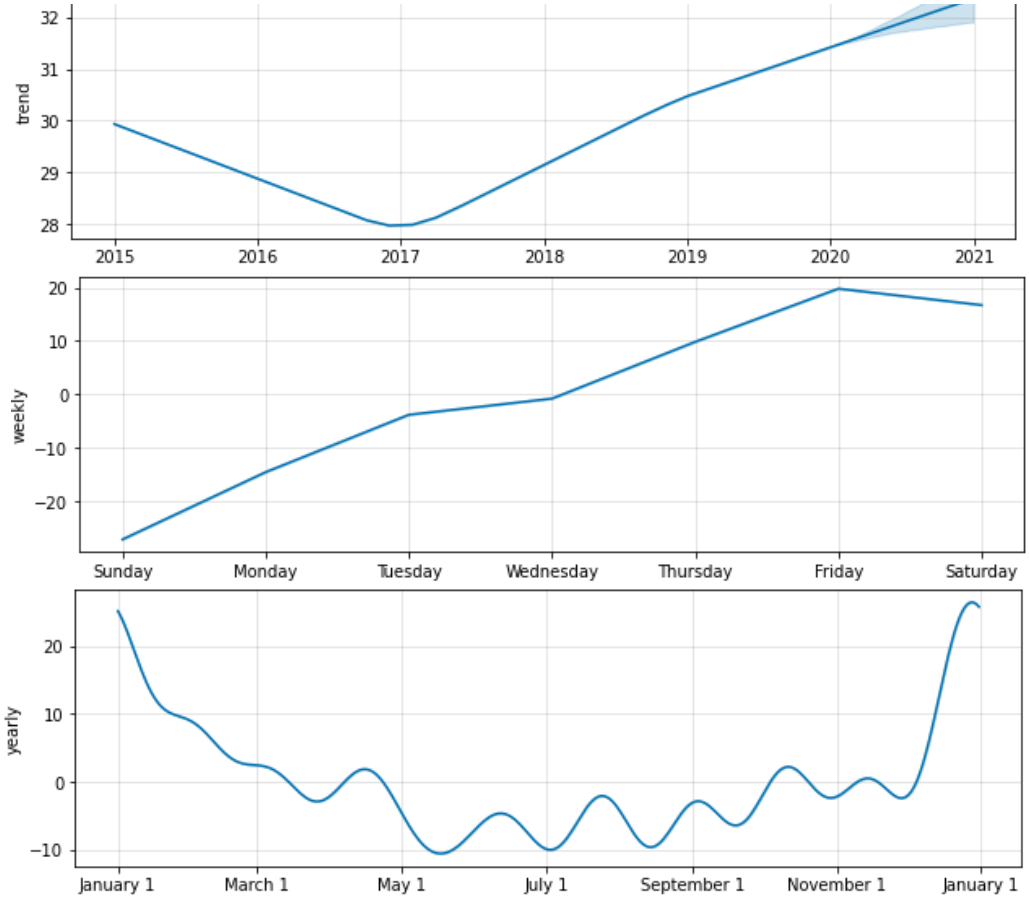


Figura 89– Decomposição da série temporal – cerveja barril. Fonte: Elaborado pelo autor.

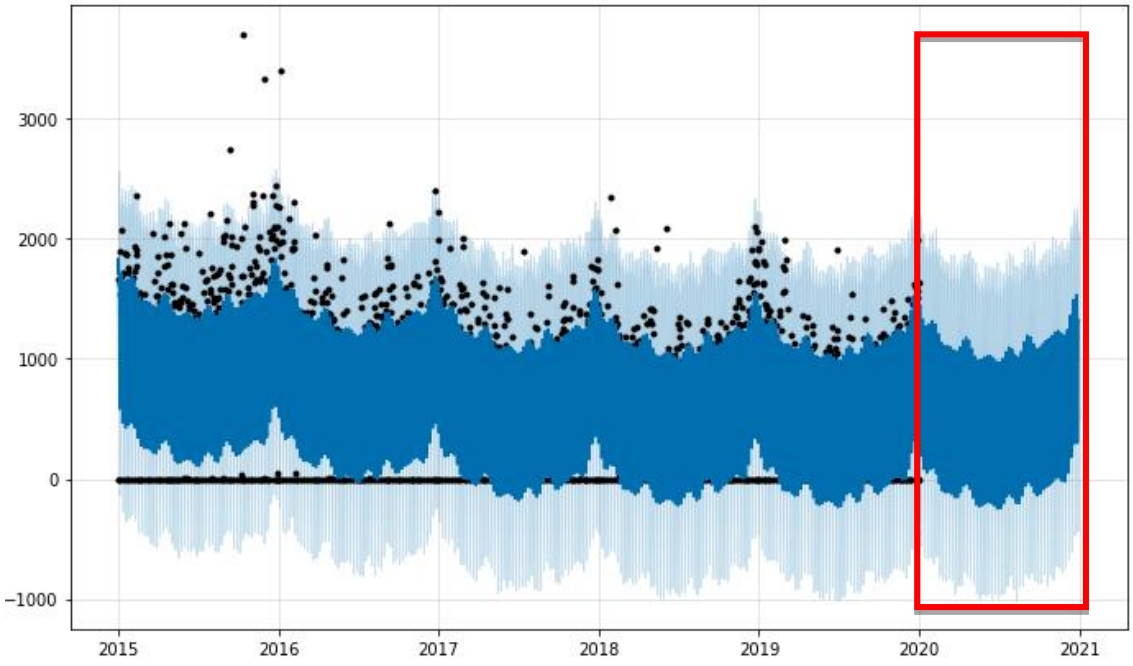


Figura 90– Valores preditos utilizando *Prophet* – cerveja retornável. Fonte: Elaborado pelo autor

Figura 91– Decomposição da série temporal – cerveja retornável. Fonte: Elaborado pelo autor.

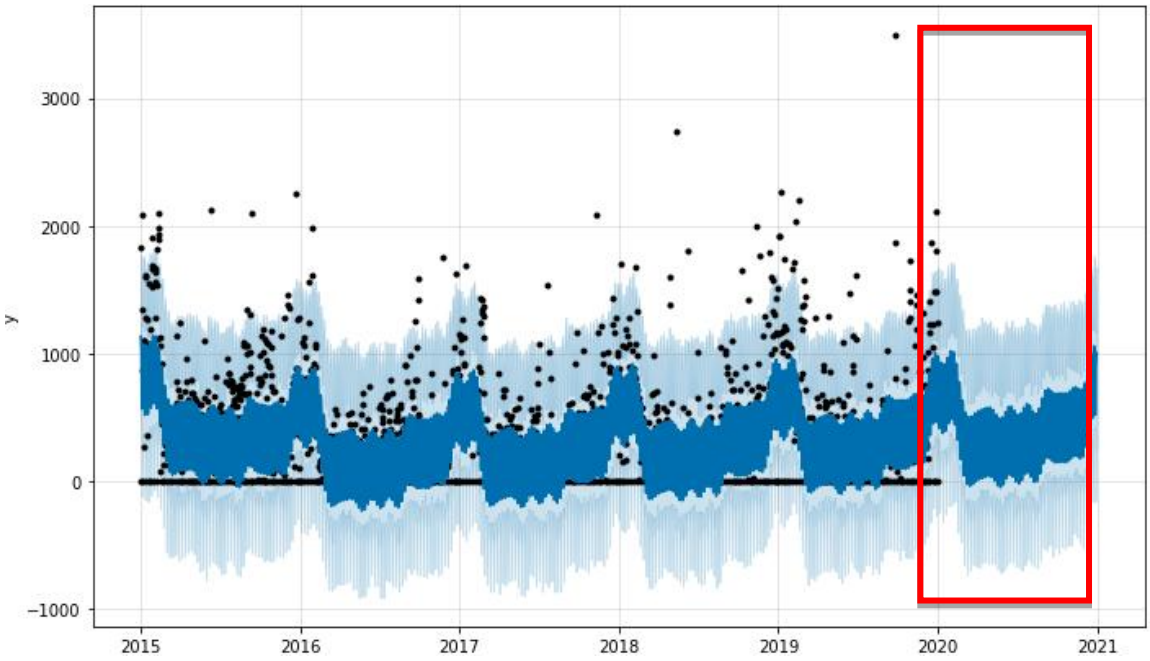


Figura 92– Valores preditos utilizando *Prophet* – cerveja descartável. Fonte: Elaborado pelo autor

Figura 93– Decomposição da série temporal – cerveja descartável. Fonte: Elaborado pelo autor.

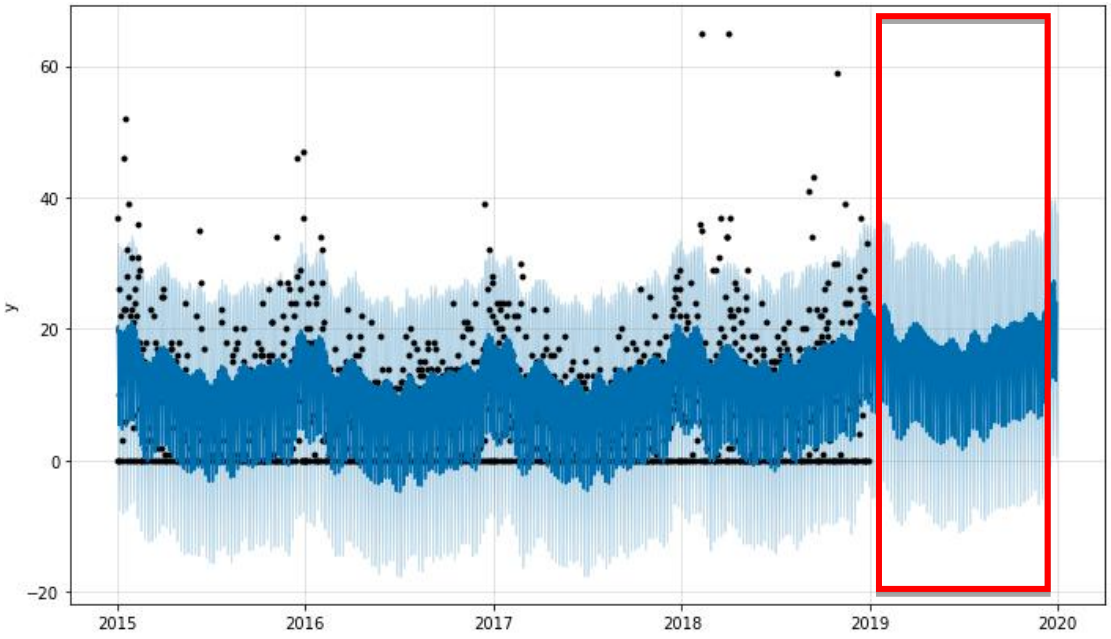


Figura 94– Valores preditos utilizando *Prophet* – cerveja puro malte. Fonte: Elaborado pelo autor

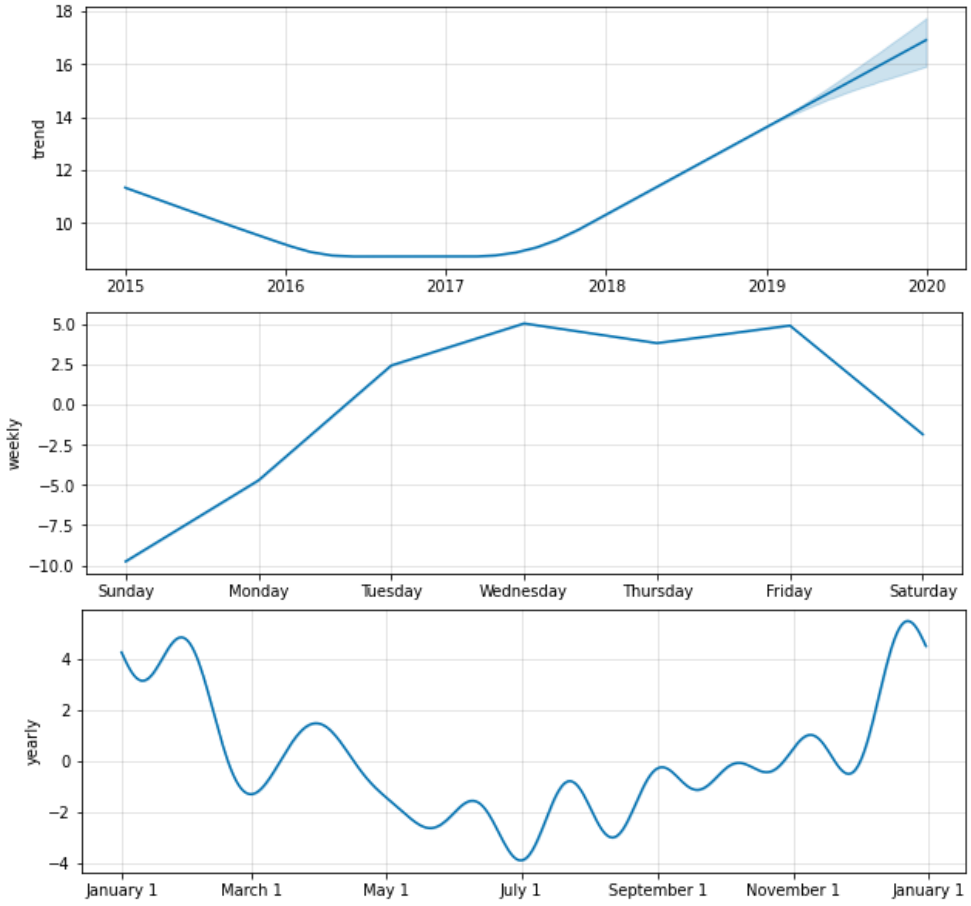


Figura 95– Decomposição da série temporal – cerveja puro malte. Fonte: Elaborado pelo autor.

2.5 PREVISÃO DE VENDA COM PROPHET COM APLICAÇÃO DE TUNNING

O *tunning* é um processo de maximização de desempenho de um modelo sem *overfitting* ou que cria variações muito altas. Este processo é realizado selecionando os "hiperparâmetros" mais adequados para cada modelo. Esta seleção é fundamental para se obter o máximo de precisão possível (Qi & Tang, 2018).

Existem diversos métodos para seleção dos hiperparâmetros, dentre eles: *Grid Search, Random Search, Bayesian Optimization*. Para o projeto em questão é aplicado o método de *Random Search,* que é um método mais rápido e menos caro porque não examinam todas as combinações possíveis. Para estabelecer parâmetros iniciais dentro de um modelo de série temporal o *Random Search* tem resultados satisfatórios (H. Wang et al., 2019).

Os parâmetros que serão alterados são:

- Seasonality_mode: Este parâmetro indica como os componentes de sazonalidade devem ser integrados às previsões. Por padrão o *Prophet* define como *additive,* tendo como opção o valor *multiplicative.* O parâmetro *additive* deve ser utilizado quando a tendência de sazonalidade for constante durante o período analisado. Já o parâmetro *multiplicative* deve ser utilizando quando se deseja aumentar a importância das sazonalidades ao longo do tempo. Para este projeto será utilizado o parâmetro *multiplicative*;

- holidays_prior_scale: Por padrão este parâmetro é 10, o que fornece pouca regularização, em outras palavras, reduzir este parâmetro amortece os efeitos dos feriados. É utilizado o valor de 20 no projeto em questão;

- yearly_seasonality: O valor padrão é 10, porém quando há necessidade de fazer com que a sazonalidade se ajuste mais rápido as mudanças de frequência pode-se aumentar o valor. O parâmetro será ajustado para 20;

- week_seasonality: O valor padrão é 1, porém quando há necessidade de fazer com que o modelo tenha maior adaptação as sazonalidades semanais pode-se aumentar o parâmetro. Neste projeto o valor a ser utilizado é de 5;
- change_prior_scale: Este é responsável por ajustar o quanto o modelo será flexível para alterar a tendência. Por padrão seu valor é de 0,05. O parâmetro tem valor ajustado para 2 no projeto.

O código utilizado para realizar pode ser visualizado abaixo, com programação realizada em linguem *Python* e utilizando o *Google Colab*.

```
from fbprophet import Prophet
dataset = cerveja_generica
dataframe = pd.DataFrame({'ds': dataset.index, 'y': dataset.values})

#definindo modelo de previsão
model = Prophet(interval_width=0.95, seasonality_mode='multiplicative',
 holidays= holiday, changepoint_prior_scale= 2, weekly_seasonality=5, h
olidays_prior_scale= 20, yearly_seasonality=20)
#treinando o modelo
model.fit(dataframe)

#criando dataframe para previsão dos próximos 90 dias com frequência di
ária
futuro = model.make_future_dataframe(periods = 365, freq='D')

#modelo de previsão de valores futuros
saida = model.predict(futuro)

#validandos os dados gerados com o modelo Prophet com os dados reais
validation_df = pd.DataFrame({'ds': dataset.index})

saida_validation = model.predict(validation_df)


#Calculando RMSE com Prophet
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(dataset.values,saida_validation.yhat))
print('Métrica RMSE: %3f' % rmse)
```

2.6 PREVISÃO DE VENDA COM PROPHET - CONCLUSÃO

O resultado da aplicação do *Prophet* pode ser considerado satisfatório, uma vez que se trata de uma biblioteca de média complexidade. No projeto em questão houve uma melhora média de 23,2% no resultado do RMSE quando comparado ao modelo com janelas deslizantes (n=8). O grupo de cerveja retornável foi onde se obteve mais melhora percentual de RMSE, chegando a um patamar de 30,7%, já o grupo de cerveja descartável teve um desempenho de 11,6%.

RMSE					
Item	Modelo Base Média Móvel Janela = 8	Prophet Padrão	% Melhora	Prophet Tunning	% Melhora
Cerveja CORE	772,17	579,97	24,9%	548,64	28,9%
Cerveja Barril	24,28	17,87	26,4%	16,95	30,2%
Cerveja Retornável	552,31	382,71	30,7%	363,07	34,3%
Cerveja Descartável	390,01	344,89	11,6%	328,19	15,9%
Cerveja Puro Malte	8,26	6,4	22,5%	6,08	26,4%

Figura 96 – Tabela de comparação de desempenho do RMSE. Fonte: Elaborado pelo autor.

Em todos os grupos foi aplicado o *tunning* nos parâmetros do *Prophet,* resultando em uma melhora média de 5% quando em comparação com o modelo sem *tunning.* Sendo assim o grupo de cerveja retornável obteve um desempenho 34,3% melhor que o modelo base e o grupo de cerveja descartável com um resultado de 15,9% melhor que o modelo base.

Como sugestão para trabalhos futuros com o *Prophet* no projeto faz a necessidade de se modelar mais datas especiais de venda de cerveja na região dos lagos, tais como feriados municipais e estaduais. Além disso pode-se sugerir a exclusão dos dias de domingo, por em sua grande maioria o valor da venda ser igual a zero.

Há a necessidade de aplicação de um algoritmo de *grid search* para exaurir todas as combinações possíveis dos hiperparâmetros, obtendo assim melhores resultado de RMSE.

Referências Artigo 3

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, *93*, 41–56. https://doi.org/10.1016/j.energy.2015.08.045

SINDICERV. (2020). https://www.sindicerv.com.br/o-setor-em-numeros/

Lazarini, J. (2019). *IBGE*. https://www.ibge.gov.br/estatisticas/economicaS

SEBRAE. (2019). Comunidade Sebrae. https://comunidadesebrae.com.br/blog/passo-a-passo-para-implementar-um-plano-de-negocios

Dewes, R., Viero, C. F., & de Lima Nunes, F. (2018). Dimensionamento de estoques: Uma análise em uma empresa varejista de peças em alumínio. *Revista Liberato*, *19*(31), 117–131. https://doi.org/10.31514/rliberato.2018v19n31.p117

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Portal Action. (2019). http://www.portalaction.com.br/series-temporais/11-estacionariedade

Shi, H., Xu, M., & Li, R. (2018). Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. *IEEE Transactions on Smart Grid*, *9*(5), 5271–5280. https://doi.org/10.1109/TSG.2017.2686012

Chen, S., Mihara, K., & Wen, J. (2018). Time series prediction of CO2, TVOC and HCHO based on machine learning at different sampling points. *Building and Environment, 146*, 238–246. https://doi.org/10.1016/j.buildenv.2018.09.054

Taylor, S. J., & Letham, B. (2017). *Forecasting at scale* [Preprint]. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3190v2

Qi, C., & Tang, X. (2018). Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Computers & Industrial Engineering, 118*, 112–122. https://doi.org/10.1016/j.cie.2018.02.028

Wang, H., Xu, H., Yuan, Y., Sun, X., & Deng, J. (2019). Balancing exploration and exploitation in multiobjective batch bayesian optimization. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 237–238. https://doi.org/10.1145/3319619.3321962