



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

**Previsão de Vendas na Rede Alimentícia
Greenpeople Usando Modelo de Machine
Learning**

Priscila Santos Cardoso

TRABALHO DE CONCLUSÃO DE CURSO

CENTRO DE CIÊNCIAS SOCIAIS - CCS

DEPARTAMENTO DE ADMINISTRAÇÃO

Graduação em Administração de Empresas

Rio de Janeiro, novembro de 2021.



Priscila Santos Cardoso

**Previsão de Vendas na Rede Alimentícia Greenpeople
Usando Modelo de Machine Learning**

Trabalho de Conclusão de Curso

Trabalho de Conclusão de Curso, apresentado ao programa de graduação em Administração da PUC-Rio como requisito parcial para a obtenção do título de graduação em Administração.

Orientador(a): Evandro da Silveira Goulart

Rio de Janeiro, novembro de 2021.

“Ela os ouvia e surpreendia-se com a própria coragem em continuar. Mas não era coragem. Era o dom. E a grande vocação para um destino.”
(Clarice Lispector)

Agradecimentos

Primeiramente agradeço aos meus pais Dilza e Cardoso, que sempre me estimularam a estudar e sonhar grande, mesmo que não tivéssemos condições financeiras para que eu chegasse aqui. A dedicação diária em me fazer ler Machado desde bem nova, me apresentar museus e mostrar sempre que eu podia chegar longe me fizeram estar aqui hoje.

Ao meu noivo Douglas, que sempre me apoiou a procurar caminhos para o meu sucesso, que sempre deseja que eu siga meus sonhos. O amor que veio no momento certo na minha vida, aquele que não sufoca, mas te dá asas e a mão para que você chegue mais longe.

Agradeço também ao meu orientador Evandro, que mesmo com a distância do ensino remoto se fez presente semanalmente. Foi um verdadeiro orientador nos percalços que surgiram no meio do caminho, elogia e acalma, mas também sabe cobrar e aplicar a pressão certa quando é necessário.

Por último, agradeço a minha amiga e agora madrinha de casamento Dayane, por me tirar todas as dúvidas de uma forma excepcional que somente uma veterana na PUC poderia. Obrigada por sempre estar ao meu lado tanto no pessoal quanto na vida profissional e acadêmica, ter você na vida é um verdadeiro presente.

Resumo

Santos Cardoso, Priscila. Previsão de Vendas na Rede Alimentícia Greenpeople Usando Modelo de Machine Learning. Rio de Janeiro, 2021. Número de páginas p. Trabalho de Conclusão de Curso – Departamento de Administração. Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo deste artigo foi considerar conceitos de Business Intelligence na previsão de vendas da empresa Greenpeople. O trabalho é resultado de uma coleta documental secundária de dados referentes às vendas de cinco produtos e informações meteorológicas nas regiões de São Paulo e Rio de Janeiro. O uso de técnicas de Inteligência Artificial, precisamente Machine Learning dentro da ferramenta de Python com regressão linear, juntamente com a correlação presente em estatística inferencial. O resultado destacou uma correlação mais forte entre os preços e quantidades dos próprios produtos do que entre eles, além de apresentar uma assertividade maior do que a atual obtida pela organização.

Palavras- chave

Business Intelligence. Inteligência Artificial. Machine Learning.

Abstract

Santos Cardoso, Priscila. Previsão de Vendas na Rede Alimentícia Greenpeople Usando Modelo de Machine Learning. Rio de Janeiro, 2021. Número de páginas p. Trabalho de Conclusão de Curso – Departamento de Administração. Pontifícia Universidade Católica do Rio de Janeiro.

The purpose of this article was to consider Business Intelligence concepts in the sales forecast of the company Greenpeople. The work is the result of a secondary documentary collection of data referring to sales of five products and meteorological information in the regions of São Paulo and Rio de Janeiro. The use of Artificial Intelligence techniques, precisely Machine Learning within the Python tool with linear regression, together with the correlation present in inferential statistics. The result highlighted a stronger correlation between the prices and quantities of the products themselves than between them, in addition to being more assertive than the current one obtained by the organization.

Key-words

Business Intelligence. Artificial Intelligence. Machine Learning.

Sumário

1 O tema e o problema de estudo	1
1.1. Introdução ao tema e ao problema do estudo	1
1.2. Objetivo do estudo	2
1.3. Objetivos intermediários do estudo	2
1.4. Delimitação do estudo	2
1.5. Justificativa e relevância do estudo	2
2 Referencial Teórico	3
2.1. Importância da Previsão de Vendas	3
2.2. Business Intelligence	3
2.2.1. Análise Descritiva	4
2.2.2. Análise Preditiva	5
2.2.3. Análise Prescritiva	6
2.3. Inteligência Artificial e Machine Learning	6
3 Métodos e procedimentos de coleta e de análise de dados do estudo	11
3.1. Tipo de Pesquisa	11
3.2. Universo e Amostra	11
3.3. Procedimento da Coleta de Dados	14
3.4. Tratamento e Análise dos Dados	14
3.5. Limitação do Estudo	15
4 Análise	16
4.1. Rio de Janeiro	16
4.2. São Paulo	17
4.3. Previsão de Vendas Atual	18
4.4. Extração da Base	19
4.5. Variáveis	19
4.6. Correlação	21
4.7. Apresentação dos Resultados	24

5 Conclusão	26
5.1. Recomendação para Novos Estudos	27
6 Referências Bibliográficas	28
Anexo 1	32

Lista de figuras

Figura 1: Três Modelos.....	4
Figura 2: IA e ML.....	7
Figura 3: Suco de Laranja.....	12
Figura 4: Suco Verde.....	12
Figura 5: Suco de Abacaxi e Hortelã.....	13
Figura 6: Suco de Maçã.....	14
Figura 7: Suco Premium.....	16
Figura 8: Pontos de Venda RJ.....	17
Figura 9: Gráfico Grandes Setores.....	17
Figura 10: Pontos de Venda SP.....	17
Figura 11: Gráfico de Previsto x Realizado 2020.....	18
Figura 12: Gráfico de Previsto x Realizado 2021.....	18
Figura 13: Teste e treino.....	20
Figura 14: Entrada, input e saída.....	21
Figura 15: Gráfico de Correlação RJ.....	22
Figura 16: Gráfico de Correlação SP.....	22
Figura 17: Mapa de Calor RJ.....	23
Figura 18: Mapa de Calor SP.....	23
Figura 19: Tabela de Predição.....	24
Figura 20: Tabela de Predição.....	24

1 O tema e o problema de estudo

1.1.Introdução ao tema e ao problema do estudo

No ano de 2020 o mundo vivenciou uma mudança atípica e inesperada com a chegada da pandemia do covid-19 que impactou diretamente na venda das empresas. Dessa forma, o planejamento de vendas empresarial ampliou sua importância para diminuir impactos consequentes de modificações no ambiente econômico e social. As técnicas e ferramentas de Data Science e Machine Learning surgem como uma opção para o planejamento. Nesse contexto, os conceitos de Business Intelligence (BI) surgem para realizar a captura e o tratamento dos dados das empresas, além da capacidade de entender padrões comportamentais e prever as necessidades futuras, como exemplo temos a antecipação da demanda. As técnicas de BI executam a análise e a categorização dos dados na busca da informação estratégica que permitirá uma tomada de decisão mais assertiva nos mais diversos ambientes empresariais.

A antecipação da demanda decorre de um conhecimento das principais variáveis ambientais, econômicas, sociodemográficas e geográficas, por exemplo, que atingem a indústria e padrão dos dados, como sazonalidade, horizontalidade, se é cíclico e a tendência (Davis, 1997). A ocorrência de chuvas ou sol, faz parte de uma variável climática que pode afetar na venda dos produtos e serviços, bem como a perecibilidade dos estoques que podem afetar na oferta de produtos, armazenamento e distribuição. Dessa forma, ter a ciência do que atinge um negócio é essencial para a tomada de decisão corporativa.

1.2.Objetivo do estudo

Este trabalho tem como foco considerar conceitos do Business Intelligence na empresa Greenpeople em sua previsão de vendas usando técnicas de Inteligência Artificial (Machine Learning) para realizar a predição de resultados de vendas. Essas técnicas foram utilizadas com o intuito de melhorar a assertividade da gestão de vendas, com uma redução de estoques, diminuindo a perda de recursos.

1.3.Objetivos intermediários do estudo

É necessário identificar as variáveis que afetam a demanda e oferta de produtos para alcançar o objetivo final deste trabalho, bem como a definição e aplicação de técnicas de Machine Learning e suas métricas. Por último, os resultados encontrados do modelo proposto serão comparados com o atualmente praticado pela organização.

1.4.Delimitação do estudo

A empresa analisada no estudo será a Greenpeople, rede que oferta produtos como sucos, snacks e suplementos de origem vegana e natural. O material utilizado da organização são as previsões de vendas dos anos de 2019, 2020 e 2021, mais precisamente sua acurácia para os estados de São Paulo e Rio de Janeiro. Com isto, as variáveis de entrada, aquelas que afetam a variável de saída que no caso é a previsão de vendas, serão as climáticas, geográficas e econômicas como preço e quantidades vendidas.

1.5.Justificativa e relevância do estudo

O estudo será importante para as empresas do segmento de varejo e alimentos, visto que demonstrará índices essenciais para a previsão de vendas que podem impactar na gestão de estoques e planejamento das organizações a curto, médio e longo prazo.

2 Referencial Teórico

Neste capítulo são apresentados e discutidos aspectos conceituais e estudos relacionados ao tema e estudo em investigação e que servirão de base para a análise realizada.

2.1. Importância da Previsão de Vendas

O setor de alimentos e bebidas conta com 34.800 empresas, sem contar as padarias (Associação Brasileira da Indústria de Alimentos, 2015). Uma forma de se ter uma vantagem competitiva em um mercado que conta com tantas empresas é o planejamento com uma previsão de vendas bem-feita, pois isso é a base das decisões estratégicas da cadeia de suprimentos (Chopra 2003), para uma maior adaptação ao ambiente de incerteza. Além disso, são fatores que embasam a importância da previsão de vendas: a complexidade do ambiente organizacional aumentando a sistematização do processo decisório; o tamanho das organizações, quanto maiores, maior o impacto de suas decisões, e por último, a necessidade da manipulação de técnicas de previsão em níveis hierárquicos mais baixos e não mais a concentração de conhecimento em especialistas (Wheelwright e Makridakis, 1998).

2.2. Business Intelligence

O uso de dados para prever o comportamento do consumidor, prever seus desejos e gerar informação é de suma importância para as organizações trabalharem em seus lançamentos, prever demandas e oferecer produtos e serviços. Para implementar ações que gerem esse conhecimento existem diversos obstáculos como, a gestão e tratamento de base de dados volumosas que são geradas rapidamente. O Business Intelligence surge como um caminho para as organizações no auxílio da transformação de dados em conhecimento para tomada de decisões.

O Business Intelligence pode ser dividido em três categorias: Análise Descritiva, Análise Preditiva e Análise Prescritiva. A cada etapa, têm-se os níveis de perguntas referentes a cada análise, são elas respectivamente, “o que aconteceu?”, “o que está acontecendo?”, “o que vai acontecer?”, “o que devo fazer?”.



Figura 1: Três Modelos – Business Intelligence e Análise de Dados para gestão do negócio

2.2.1. Análise Descritiva

A análise descritiva é a base na qual o uso de métodos estatísticos se aplica, principalmente a estatística do tipo descritiva por meio de ferramentas como gráficos e dados de números para descrição deles. O surgimento desta tipologia advém da necessidade de se compreender tendências e padrões em fenômenos repetitivos, “A análise descritiva (ou extração de relatórios) diz respeito a conhecer o que está acontecendo na organização e entender tendências e causas subjacentes de tais ocorrências” (Sharda, Delen e Turban, 2019).

Na busca de bons resultados com análise descritiva, (Sahay, 2018) é de suma importância a compreensão dos tipos de dados, da representação dos gráficos e tabelas e ainda as técnicas gráficas utilizadas. No intuito de obter uma melhoria desses resultados, pode-se combinar a técnica a Big Data, ou seja, bancos de dados que trazem bastante informações que podem ser relacionadas e consolidadas de maneira que se alcance uma maior acurácia, “Em primeiro lugar, isso envolve a consolidação de fontes de dados e a disponibilidade de todos os dados relevantes de um modo que permite a extração e análise apropriadas de relatórios” (Sharda, Delen e Turban, 2019).

2.2.2. Análise Preditiva

Análises Preditivas usam dados, algoritmos estatísticos e técnicas de Inteligência Artificial para prever resultados a partir de dados históricos. As técnicas de Machine Learning são exemplos desse tipo de análise.

Existem dois tipos de modelos preditivos: Modelos de Classificação e Modelos de Regressão. Os modelos de Classificação basicamente preveem associações de classe enquanto os modelos de Regressão predizem um número.

Exemplos de técnicas de modelagem preditiva são: Árvores de Decisão, Regressão e Redes Neurais.

Árvore de Decisão são modelos de classificação que apresentam dados em subconjuntos com base em categorias de variáveis de entrada. Uma árvore de decisão parece uma árvore com cada ramo representando uma escolha entre um número de alternativas, e cada folha representando uma classificação ou decisão. Este modelo analisa os dados e tenta encontrar a variável que divide os dados.

Regressão é um método estatístico amplamente utilizado. Na análise de regressão nos Modelos de Machine Learning, busca-se estimar as relações entre as variáveis do modelo em questão. Esse método utiliza dados quantitativos para prever um número que chamamos de valor de resposta, ou variável Y, ou valor de saída. Além disso, uma variável independente X é usada para explicar ou prever o valor de saída Y. Podemos chamar os valores de X também de valores de entrada. A Regressão Múltipla usa duas ou mais variáveis independentes para prever o resultado.

As Redes Neurais são técnicas mais sofisticadas para modelar relações mais complexas. Foram originalmente desenvolvidas baseando-se no funcionamento do cérebro humano.

2.2.3. Análise Prescritiva

“A análise prescritiva recomenda às organizações as ações ideais para alcançar seus objetivos de negócios, como a satisfação do cliente, os lucros e a redução de custos” (IBM). A solução se utiliza de tecnologias de otimização advindas da pesquisa operacional e de gerenciamento para reconhecer a situação atual, prever o futuro e ajudar na melhor tomada de decisão, alocação de recursos e desempenho organizacional (Sharda, Delen e Turban, 2019). Os benefícios são principalmente na parte logística organizacional com prazos menores para entrega, custo de entrega para o cliente e organização e melhoria nos estoques, visto que os modelos dessa solução são voltados para isso, visto que são eles: de otimização linear, não linear, programação linear e inteira, problemas de agendamento, transporte, atribuição, programação 0-1, de simulação, entre outros (Sahay, 2018).

A análise prescritiva é uma abordagem que também tem como foco reconhecer uma situação atual e prever um resultado futuro visando uma melhor tomada de decisão, como por exemplo na alocação de recursos e melhoria do desempenho de uma organização (Sharda, Delen e Turban, 2019).

Esse tipo de análise trabalha com ferramentas de pesquisa operacional e de gerenciamento para alocar os recursos de maneira eficiente para otimizar e automatizar seus processos de negócios.

As ferramentas analíticas prescritivas comuns são modelos de otimização linear e de otimização não linear.

2.3. Inteligência Artificial e Machine Learning

Inteligência Artificial (IA) pode ser definida como capacidade que uma máquina tem de imitar as capacidades da mente humana, além de aprender com exemplos e experiências, conseguir reconhecer objetos, compreender e responder, tomar decisões e resolver os mais diversos problemas.

Podemos considerar quatro subconjuntos principais de IA:

- 1) Raciocínio de Máquina ou Machine Reasoning
- 2) Processamento de Linguagem Natural (NLP)
- 3) Planejamento Automatizado ou Planning
- 4) Aprendizado de Máquina ou Machine Learning

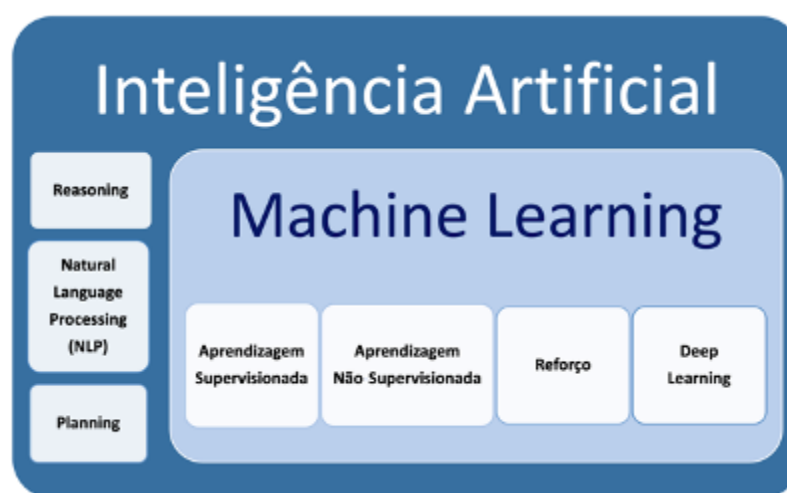


Figura 2: IA e ML. Fonte: HURWITZ, J. and KIRSCH, D., 2018.

Raciocínio de Máquina (Machine Reasoning) permite que um sistema faça inferências com base em dados, apresentando habilidade de preencher lacunas quando há dados incompletos (HURWITZ, J. e KIRSCH, D., 2018).

Processamento de Linguagem Natural (PLN) ou Natural Language Processing (NLP), em inglês como é mais conhecida, é a capacidade de treinar computadores para entender tanto o texto escrito quanto a fala humana. As técnicas de PNL são necessárias para capturar o significado do texto não estruturado de documentos ou da comunicação do usuário. (HURWITZ, J. e KIRSCH, D., 2018).

Planejamento Automatizado (Planning) é a capacidade de um sistema inteligente de agir de forma autônoma e flexível para construir uma sequência de ações para atingir um objetivo final. Em vez de um processo de tomada de decisão pré-programado que vai de A a B a C para chegar a um resultado, o planejamento automatizado é complexo e requer um sistema para se adaptar com base no contexto em torno do desafio dado. (HURWITZ, J. e KIRSCH, D., 2018).

Machine Learning ou Aprendizado de Máquina são necessárias para melhorar a precisão dos modelos preditivos. Dependendo da natureza do problema de negócios que está sendo tratado, existem diferentes abordagens com base no tipo e no volume dos dados. Machine Learning está dividida em quatro categorias:

- 1) Aprendizagem Supervisionada (Supervised Learning)
- 2) Aprendizagem Não Supervisionada (Unsupervised Learning)
- 3) Aprendizagem por Reforço (Reinforcement Learning)
- 4) Aprendizagem Profunda (Deep Learning)

Aprendizagem Supervisionada utiliza um conjunto de dados ou de exemplos estabelecidos e pré-processados para avaliar o desempenho do algoritmo de aprendizagem em descobrir padrões e fazer classificações, e fazer previsão ou regressão. Os modelos de Aprendizado Supervisionado normalmente começam com um conjunto estabelecido de dados e uma certa compreensão de como esses dados são classificados. Essa aprendizagem se destina a encontrar padrões em dados que podem ser aplicados em um processo analítico. Esses dados rotulam recursos que definem o significado dos dados. É dito Aprendizado “Supervisionado”, pois os dados de treinamento do modelo são conhecidos, isto é, para um conjunto de entrada de dados tem um conjunto de saída de dados (resultado) bem estabelecido e coerente com o problema. (HURWITZ, J. e KIRSCH, D., 2018).

Aprendizagem Não Supervisionada é mais adequada quando o problema requer uma enorme quantidade de dados que não são rotulados. Por exemplo, as redes sociais como o Twitter, Instagram, Snapchat, possuem uma enorme base de dados não rotulados. Compreender as informações por trás desses dados requer algoritmos que possam começar a entender os significados com base na capacidade de classificar os dados nos padrões encontrados. (HURWITZ, J. e KIRSCH, D., 2018).

Aprendizagem por Reforço é um modelo de aprendizagem comportamental. O algoritmo recebe feedback da análise dos dados para que o usuário seja guiado para o melhor resultado. Diferentemente das aprendizagens citadas anteriormente, nesse tipo de aprendizagem o sistema não é treinado com um conjunto de dados de amostra. Aprendizagem por Reforço aprende através de tentativa e erro. Exemplo de aplicação desse tipo de modelo é no treino de robô que precisa aprender a navegar (andar) em um terreno qualquer apenas com base nos dados resultantes das suas ações, ou seja, quando ele cai os dados são recalibrados para que as etapas sejam navegadas de maneira diferente de tal forma que o robô não caia novamente da mesma forma. (HURWITZ, J. e KIRSCH, D., 2018).

Deep Learning tem um modo de aprendizagem e um algoritmo de previsão mais interativos, pois emulam como o cérebro humano funciona permitindo lidar com vários tipos de dados e lidar com abstrações ou problemas mal definidos (HURWITZ, J. e KIRSCH, D., 2018). Nessa aprendizagem incorpora a ideia das Redes Neurais em camadas sucessivas para aprender com os dados de maneira iterativa. É especialmente útil quando se deseja aprender padrões a partir de dados não estruturados. Uma rede neural consiste em três ou mais camadas, sendo a primeira camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Nesse processo iterativo, os dados são modificados na camada oculta e na camada de saída conforme os pesos são aplicados em cada nó (neurônio). (HURWITZ, J. e KIRSCH, D., 2018).

Deep Learning é uma técnica de Machine Learning que usa Redes Neurais hierárquicas para gerar aprendizado considerando uma combinação de algoritmos de Aprendizagem Supervisionada com algoritmos de Aprendizagem Não Supervisionada.

Por ser um dos principais subconjuntos de IA, Machine Learning é comumente utilizada dentro das abordagens de Business Intelligence na etapa de análises de dados preditivas

A ferramenta de ML identifica padrões a partir de dados, automatiza a construção de modelos de previsão e, em alguns casos, podendo fazer melhorias nos modelos sem interação humana, o que permite aprender com tendências, padrões e relacionamentos de dados históricos e tomar decisões, (Sahay, 2018).

O Aprendizado de Máquina (Machine Learning) é um método de Análise de Dados que automatiza a construção de modelos analíticos. É um ramo da Inteligência Artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana. Enquanto a Inteligência Artificial (IA) pode ser definida como a ciência capaz de mimetizar as habilidades humanas, o Machine Learning é uma vertente específica de IA que treina “máquinas” para aprender com dados.

O processo básico do Machine Learning pode ser dividido em três partes: Entrada de Dados, Processamento ou Abstração e Saída de Dados.

A entrada de dados é a etapa que utiliza dados ou informações anteriores como base para tomadas de decisão.

Processamento ou Abstração é a etapa que recebe os dados de entrada, e por meio deles aprende e constrói a relação entre dados e o algoritmo subjacente de previsão. Essa etapa é importante pois ela é a responsável pela aprendizagem da máquina e melhoria do modelo de previsão (Chandramouli; Dutt e Kumar Das, 2018).

Na Saída de Dados surge a resposta que o modelo de ML gera após as etapas de treinamento e teste, ou seja, o valor da predição desejada.

3 Métodos e procedimentos de coleta e de análise de dados do estudo

3.1. Tipo de Pesquisa

Uma pesquisa pode ter fins do tipo exploratória, descritiva, explicativa, metodológica, aplicada ou intervencionista (Vergara, 2005). A investigação aplicada tem intuito de resolver problemas que existem na prática, de maneira imediata ou não, e a intervencionista não para somente na explicação do que se está sendo estudado, mas visa interferir no seu objeto de estudo. Dessa forma, esse estudo quanto aos fins é aplicado e intervencionista, pois além de querer melhorar a acurácia do modelo de previsão, isso tem intuito de melhorar os números da empresa usada.

Quanto aos meios, a autora diz que podem ser de campo, laboratório, documental, bibliográfica e experimental. A pesquisa experimental visa uma investigação empírica com a manipulação de variáveis e observação de resultados. Sendo assim, é a mais coerente para o estudo, visto que aqui foi realizada a manipulação de base de dados, uso de modelos de Machine Learning e análise da acurácia do que conseguimos como produto final.

3.2. Universo e Amostra

A amostra é composta por fatos que concernem a venda de cinco produtos: laranja (Figura 3), verde (Figura 4), abacaxi com hortelã (Figura 5), maçã (Figura 6) e premium (Figura 7), analisados durante as 53 semanas de 2019, 53 semanas de 2020 e 31 semanas de 2021, suas quantidades e preços nos estados do Rio de Janeiro e São Paulo. Já as amostras de temperatura e precipitação fornecem dados dos anos de 2019, 2020 e 2021 e suas respectivas semanas também no Rio de Janeiro e São Paulo.



Figura 3: Suco de Laranja - Greenpeople



Figura 4: Suco Verde - Greenpeople



Figura 5: Suco de Abacaxi e Hortelã - Greenpeople



Figura 6: Suco de Maçã - Greenpeople



Figura 7: Suco Premium - Greenpeople

3.3. Procedimento da Coleta de Dados

Os dados foram adquiridos por meio de coleta documental da empresa analisada, retirados de base de dados já existente, ou seja, secundária, bem como as informações de temperatura e precipitação obtidas de base de dados do Instituto Nacional de Meteorologia, pois são variáveis que podem afetar as vendas de produtos alimentícios. Os procedimentos visaram a manipulação de dados e a observação dos resultados.

3.4. Tratamento e Análise dos Dados

As bases de dados obtidas foram analisadas conforme a técnica de correlação presente em Análises Estatísticas, além de Modelos de Machine Learning utilizando bibliotecas do Scikit-Learn de Regressão Linear em Linguagem Python, associando os resultados às teorias presentes na bibliografia sobre Estatística Aplicada à Administração e Economia e de Introdução à Linguagem Python.

3.5.Limitação do Estudo

A pesquisa apresenta uma série de limitações. Primeiramente, a base de dados fornecida pela empresa contém apenas os anos de 2019, 2020 e 2021. Ou seja, três anos muito atípicos, visto que o primeiro é pré-pandêmico, o segundo pandêmico e o último pós-pandêmico, mas com barreiras sanitárias e econômicas diferentes do anterior, até por ser nesse ano (2021) que as vacinas foram distribuídas à população.

Além disso, dada a vasta gama de produtos e a cobertura que a organização tem em vários estados brasileiros, foram selecionados os sucos nos sabores Laranja, Verde, Abacaxi com Hortelã, Maçã e Premium apenas nos estados do Rio de Janeiro e São Paulo. Quanto aos dados climáticos, as temperaturas e a precipitação de cada cidade citada anteriormente, houve a seleção dos centros meteorológicos localizados em Interlagos, São Paulo, e Forte de Copacabana no Rio de Janeiro.

4 Análise

Neste capítulo serão apresentados os caminhos da aplicação da Aprendizagem Supervisionada de Machine Learning em ambiente de linguagem Python para obtenção dos melhores resultados possíveis. A Greenpeople é uma empresa que atende a nível Brasil, mas foram selecionados os principais estados da organização que contam com maiores vendas e fazem parte da mesma divisão regional, sudeste, são elas: São Paulo e Rio de Janeiro.

4.1. Rio de Janeiro

O Estado do Rio de Janeiro para a empresa é muito importante, visto que é onde possui sua fábrica na cidade de Três Rios e escritório principal na cidade do Rio de Janeiro, mais precisamente, no bairro da Gávea. Os seus pontos de vendas estão bastante concentrados principalmente na zona sul e o setor industrial vem tendo um aumento significativo pós-pandemia com um crescimento de 8,0% no segundo semestre de 2021 (Firjan, 2021). Ademais, o PIB do estado também vem demonstrando um crescimento de 10,3%, (Firjan, 2021) no mesmo período.

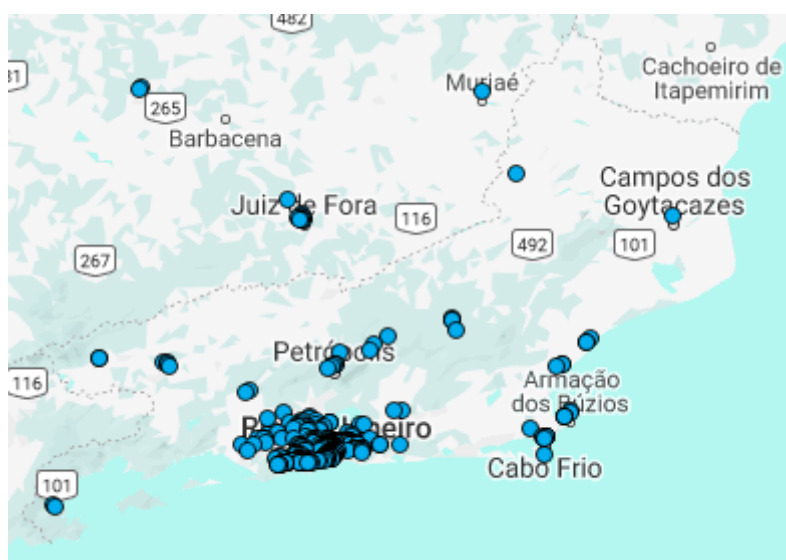
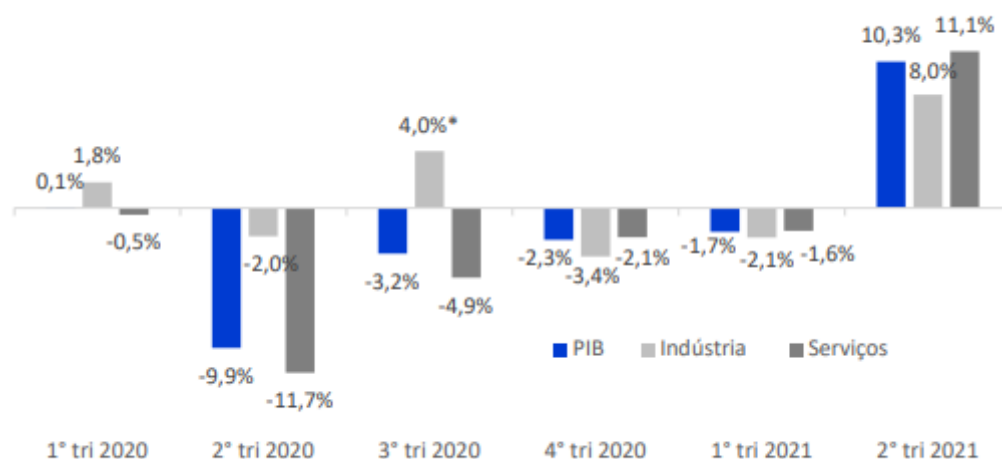


Figura 8: Pontos de Venda RJ - Greenpeople

Gráfico 1 - Taxa de variação interanual para o PIB do Rio de Janeiro - Grandes Setores



Fonte: Firjan

Figura 9: Gráfico Grandes Setores - Firjan

4.2. São Paulo

O Estado de São Paulo possui a 21ª maior economia do mundo (US\$ 603,4 bilhões), tendo um PIB maior que países como Polônia, Suécia, Bélgica, Argentina, Áustria, Noruega, Irlanda, Singapura e Dinamarca (Casa Civil de São Paulo, 2020). Os seus pontos de vendas estão bastante concentrados na zona oeste, por conter o público-alvo da organização. Entretanto, a empresa atua tanto na capital quanto no interior, visto que o estado detém fortes características econômicas como as descritas acima.



Figura 10: Pontos de Venda SP - Greenpeople

4.3. Previsão de Vendas Atual

Atualmente, a previsão de vendas da empresa não é realizada por tecnologia de Machine Learning, é feita por meio de base histórica de vendas e média do total vendido das últimas 4 semanas mais recentes. Utilizando esse modelo, foi obtido na primeira semana do ano de 2021 uma assertividade de 27% do previsto demandado comparado ao real vendido.

A equipe da empresa então decidiu adotar um modelo que usa pesos distribuídos por semanas juntamente ao modelo antigo com base histórica e médias semanais. O método empregado trouxe na semana trinta e dois de 2021 uma assertividade de 76%, porém o modelo não repetiu o mesmo desempenho nas semanas seguintes, caindo para menos de 30%.

Abaixo seguem os gráficos que demonstram a evolução do previsto e realizado em litros em milhares, os dados foram ocultados a pedido da organização para preservação de informações confidenciais. Os gráficos demonstram algumas rupturas quando há mais venda do que o previsto e quantidades de estoque mais elevadas quando o previsto é maior que o vendido. As duas situações são ruins para uma empresa do setor alimentício, visto que os produtos têm uma alta perecibilidade por se tratar de sucos naturais sem conservantes. Por exemplo, na semana 51 da Figura 11, há uma produção significativamente maior do que os litros vendidos.



Figura 11: Gráfico de Previsto x Realizado 2020 - Greenpeople



Figura 12: Gráfico de Previsto x Realizado 2021 - Greenpeople

4.4. Extração da Base

Para a escolha dos produtos foram feitos gráficos nas bases de vendas histórias da organização nos anos 2019, 2020 e 2021. Dada a vasta gama de produtos, foram eleitos os cinco produtos mais vendidos: Laranja, Verde, Abacaxi com Hortelã, Maçã e Premium nas regiões de São Paulo e Rio de Janeiro, as duas regiões foram escolhidas para o teste inicial, porém a empresa cobre todas as regiões do Brasil. Além disso, foram extraídas bases de dados do Instituto Nacional de Meteorologia para as semanas dos anos e áreas de estudo, as informações são dadas de uma em uma hora por dia, foram feitas então médias semanais para obtenção dos dados de temperatura e precipitação para servirem de entrada no modelo. Após a obtenção de dados, foi feita uma base de dados nova alimentada por todas as informações coletas que serviriam para as entradas do modelo: Semana, Ano, Temperatura, Precipitação, Média de Quantidade Vendida e Preços de suco de Laranja, Verde, Abacaxi com Hortelã, Maçã e Premium.

4.5.Variáveis

O modelo usado é o de Aprendizagem Supervisionada, e, como dito no referencial teórico, ela encontra padrões em dados. Para um conjunto de dados de entrada existe um conjunto de saída/resultado. No presente estudo, os dados de entrada são referente às 53 semanas de 2019 e 2020 (visto que a semana organizacional começa na quinta, por isso essa semana a mais) e 29 semanas de 2021 com as seguintes informações: Semana, Ano, Temperatura, Precipitação, Média da Quantidade de Laranja, Média da Quantidade de Verde, Média da Quantidade de Abacaxi com Hortelã, Média da Quantidade Maçã, Média da Quantidade Premium, Média do Preço de Laranja, Média do Preço Verde, Média do Preço de Abacaxi com Hortelã, Média do Preço de Maçã e Média do Preço de Laranja e um Índice, variável criada para obtenção de dados em sequência. Quanto às saídas, ou seja, o resultado que se quer obter no modelo, são as Quantidades Previstas de Laranja, Verde, Abacaxi com Hortelã, Maçã e Premium, em litros.

Na Figura 13 está exemplificado como funciona o modelo com as entradas e saídas, considerando, por exemplo, que 75% dos dados serão usados para treinamento (tanto da entrada como da saída) e 25% dos dados serão usados para teste. Em outras palavras, nesse exemplo citado anteriormente, ele vai usar a informação de 75% para treinar como ele pode adivinhar o restante dos 25% dos dados, ele vai entendendo como funciona com os 75% dos dados, usa os 25% de teste de x (valores de entrada) como input para adivinhar os valores dos 25% de teste de y (valores de saída). Nessa etapa de teste, será gerado um valor de acurácia do modelo com essa configuração. Valores de acurácia estarão sempre entre zero e um. Valores de acurácia próximos de zero indicam uma péssima acurácia, ou seja, o modelo não irá gerar um resultado confiável. Valores de acurácia próximos de um (1) indicam uma ótima acurácia, ou seja, o modelo poderá gerar um resultado confiável.

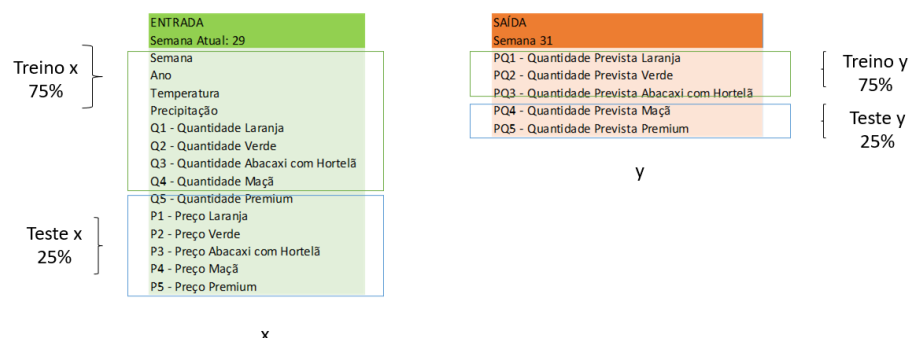


Figura 13: Teste e treino

Dessa forma, foram realizadas diferentes previsões no modelo que serão demonstrados no capítulo 4.7. A ideia foi semelhante ao exemplo dado anteriormente, usou-se como entrada todas as semanas de 2019, 2020 e 29 semanas de 2021. Como input para realizarmos um teste no modelo pronto foi considerada a semana 30 de 2021 e o resultado (ou saída) será comparado com os dados da semana 31 de 2021.

ENTRADA	INPUT	SAÍDA
Semana Atual: 29	Semana 30	Semana 31
Semana	Semana	PQ1 - Quantidade Prevista Laranja
Ano	Ano	PQ2 - Quantidade Prevista Verde
Temperatura	Temperatura	PQ3 - Quantidade Prevista Abacaxi com Hortelã
Precipitação	Precipitação	PQ4 - Quantidade Prevista Maçã
Q1 - Quantidade Laranja	Q1 - Quantidade Laranja	PQ5 - Quantidade Prevista Premium
Q2 - Quantidade Verde	Q2 - Quantidade Verde	
Q3 - Quantidade Abacaxi com Hortelã	Q3 - Quantidade Abacaxi com Hortelã	
Q4 - Quantidade Maçã	Q4 - Quantidade Maçã	
Q5 - Quantidade Premium	Q5 - Quantidade Premium	
P1 - Preço Laranja	P1 - Preço Laranja	
P2 - Preço Verde	P2 - Preço Verde	
P3 - Preço Abacaxi com Hortelã	P3 - Preço Abacaxi com Hortelã	
P4 - Preço Maçã	P4 - Preço Maçã	
P5 - Preço Premium	P5 - Preço Premium	

Figura 14: Entrada, input e saída

4.6. Correlação

Foi realizado dentro do próprio script na linguagem Python as correlações entre os preços e quantidades dos produtos. Decorrente dessas tentativas foram percebidos que as correlações entre os produtos são fracas. As correlações mais relevantes foram entre preços e quantidades do mesmo produto. Além disso, foi notado que para os anos e dados em questão, que a temperatura e a precipitação não tinham uma correlação significativa, pois esperava-se que apresentasse uma correlação positiva em relação às vendas, que, por se tratar de um suco, imaginava-se que em dias de maior temperatura a venda aumentasse ou ao chover poderia apresentar uma redução nas vendas.

Abaixo as Figuras 15 e 16, representam a correlação entre quantidade e preço do suco Verde nos estados de São Paulo e Rio de Janeiro, demonstrando uma correlação positiva no primeiro gráfico e uma correlação também positiva, porém com pontos um pouco mais dispersos em algumas semanas no segundo gráfico.

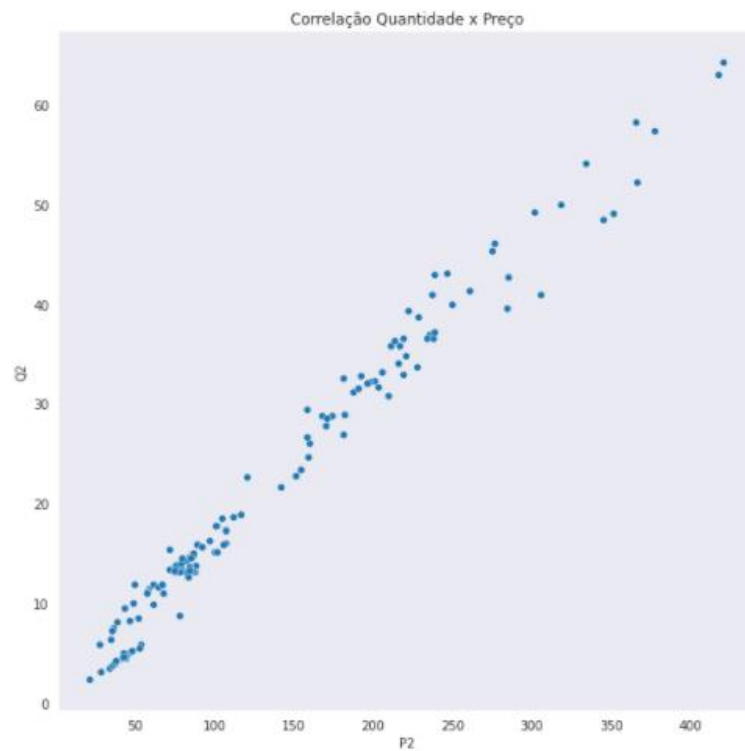


Figura 15: Gráfico de Correlação RJ

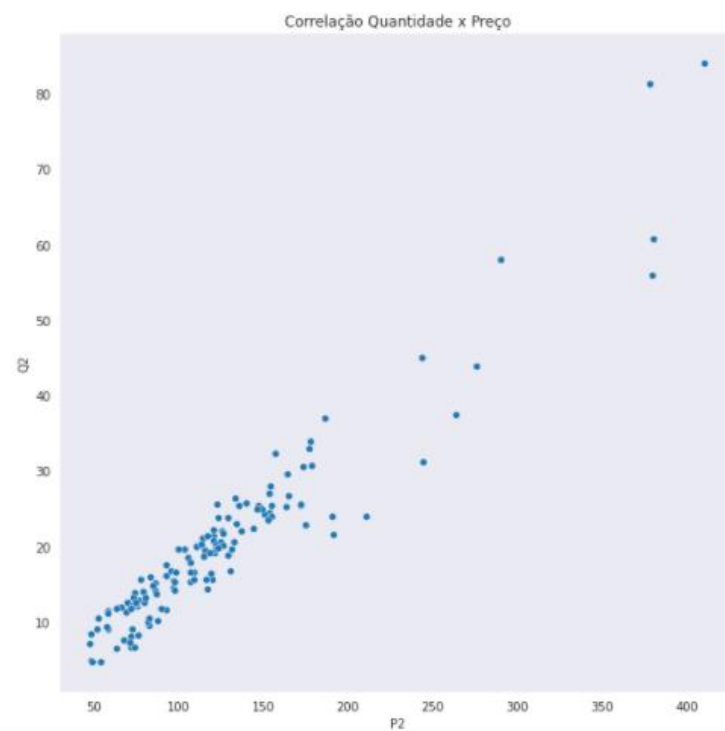


Figura 16: Gráfico de Correlação SP

O mapa de calor de correlação dos dados do RJ na Figura 17, confirma que o Rio de Janeiro possui correlações melhores do que as de São Paulo na Figura 18, visto que o mapa de calor de SP está com cores mais fortes mostrando uma correlação próximas de zero na maior parte da figura e o mapa de calor do RJ tem cores um pouco mais claras, indicando correlações com valores intermediários. Além disso, podemos comprovar que as melhores correlações, nos quadrados de cor branca (ou mais claras), são de informações dos produtos com eles mesmos (obviamente), mas também entre os preços e quantidades, por exemplo P1 com Q1 (P2 com Q2, ...), ou seja, média de preço do suco de laranja com média de quantidade do suco de laranja.

```
sns.heatmap(dataset.corr())
plt.title('HeatMap das Correlações')
plt.show()
```

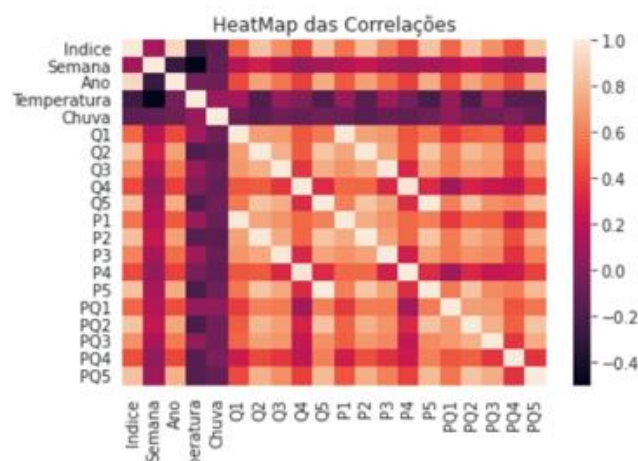


Figura 17: Mapa de Calor RJ

```
[ ] sns.heatmap(dataset.corr())
    plt.title('HeatMap das Correlações')
    plt.show()
```

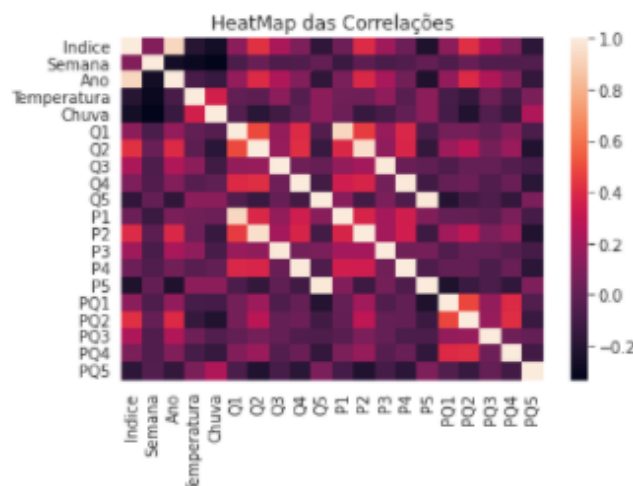


Figura 18: Mapa de Calor SP

4.7. Apresentação dos Resultados

Foram realizadas predições com diferentes quantidades para teste e treino para os estados do Rio de Janeiro e São Paulo. Observando a Figura 19 abaixo, pode-se observar que o primeiro conjunto de resultados (na linha “Predição 0,05”) levou-se em conta 95% dos valores para treino e 5% para teste, o segundo 10% teste e 90% de treino, até chegar a última predição de 0,30 (na linha “Predição 0,30”), que considera 30% dos dados para teste e 70% para treino. As melhores acurácias no modelo foram de 60% para o Rio de Janeiro e 58% para São Paulo, tendo uma assertividade melhor em São Paulo do que o modelo atual da organização que chega em 27% de acertos, comparando previsto com real.

No caso do Rio de Janeiro, não se obteve resultados comparativos muito bons, isso se deve às delimitações do estudo que poderia fornecer um melhor resultado tanto para São Paulo quanto para Rio de Janeiro se a coleta documental tivesse um período maior. Os anos de 2019, 2020 e 2021, diferem muito uns dos outros, o primeiro é pré-pandêmico, o segundo pandêmico e o terceiro pós-pandêmico.

RJ	Q1	Q2	Q3	Q4	Q5	Acurácia
Predição 0,05	32,6	49,9	20	29,5	21,9	60
Predição 0,10	38,4	49,6	21,2	29,6	22,5	52
Predição 0,15	30,5	50,7	20,6	21,97	23,1	48
Predição 0,20	30,8	49,99	20,7	30,7	21,7	43
Predição 0,25	25,2	46,8	20,2	29,95	22,1	47
Predição 0,30	38,0	52,03	21,3	29,8	22,5	41
Valores Reais	18,00	34,00	18,00	10,00	17,00	
Previstos/Reais	181%	147%	111%	295%	129%	

Figura 19: Tabela de Predição

SP	Q1	Q2	Q3	Q4	Q5	Acurácia
Predição 0,05	28,2	31,0	24,3	25,5	11,3	57
Predição 0,10	28,8	31,4	24,0	24,7	10,9	43
Predição 0,15	28,7	32,0	24,0	26,5	11,5	45
Predição 0,20	27,7	29,0	25,1	25,0	10,5	47
Predição 0,25	32,3	31,9	24,7	33,5	11,4	53
Predição 0,30	25,3	24,5	23,6	14,7	11,3	58
Valores Reais	28,38	20,89	20,52	22,00	19,22	
Previstos/Reais	89,3%	117,1%	115,2%	66,6%	58,9%	

Figura 20: Tabela de Predição

Na Figura 19 os valores da linha “Predição 0,05” foram os que apresentaram a melhor acurácia, quando comparados com os valores reais da semana 31 do ano de 2021 do RJ.

Na Figura 20 os valores da linha “Predição 0,30” foram os que apresentaram a melhor acurácia, quando comparados com os valores reais da semana 31 do ano de 2021 de SP.

5 Conclusão

A monografia teve como objetivo a aplicação de conceitos de Business Intelligence na previsão de vendas da rede de indústria alimentícia Greenpeople utilizando a linguagem Python para uso das técnicas de Machine Learning para predição de quantidades a serem vendidas. Com o modelo apresentado neste estudo, obteve-se valores melhores de assertividade se comparados com valores anteriores praticados na gestão de vendas, podendo reduzir estoques e perda de recursos. Dessa forma, não traz somente benefícios para a organização, mas para todo setor industrial, visto que dependem de uma previsão de vendas bem-feita para serem eficientes e eficazes.

A pesquisa indicou conforme foi colocado na apresentação dos resultados que a falta de uma base de dados com um intervalo temporal mais amplo e conseqüentemente mais representativo, levou a resultados não tão bons no modelo de predição com Machine Learning quando comparamos, por exemplo, o valor previsto com o valor real no estado do Rio de Janeiro. Apesar desse ponto, na região de São Paulo se obteve valores bem perto dos reais, o que demonstra um potencial enorme na previsão de vendas com ML para as organizações, mesmo que necessite de atenção a esses detalhes que foram citados nas limitações do nosso estudo. Apesar disso, o modelo utilizando Machine Learning alcançou resultados melhores do que a forma na qual é realizada as predições atualmente na organização.

Portanto, em linhas gerais, as implicações deste estudo para as empresas podem ser positivas. Podendo ser úteis para a formação de conhecimento sobre o assunto, possibilidade de tomada de decisões mais assertivas utilizando previsão de vendas com técnicas de Inteligência Artificial, mitigação de perda de recursos e melhora na gestão de estoques ou também para escolher o melhor método para a aplicação de Machine Learning dentro do setor industrial. Por fim, pode permitir que se originem outros estudos mais densos e consolidados sobre o assunto.

5.1.Recomendação para Novos Estudos

As delimitações do estudo, como por exemplo os dados referentes aos anos de 2019, 2020 e 2021, bem diferentes entre si com períodos pré-pandêmicos, durante a pandemia e pós-pandêmicos, podem levar a resultados bem distorcidos. O ideal para próximos estudos seria uma coleta documental com um intervalo de tempo maior para que a acurácia possa ser melhor. Além disso, poderia ser incluídos processos de Machine Learning por reforço para que não se precise alimentar a base manualmente quando novos dados forem incluídos, ele aprenda sozinho e consiga dar resultados cada vez melhores. Por último, podem ser incluídas mais variáveis como econômicas ou de perecibilidade de estoques para demonstrar uma correlação entre os dados e trazer maior assertividade ao modelo.

6 Referências Bibliográficas

ABIA. Indústria de alimentos cresce 6,7% em 2019. Disponível em: <https://www.abia.org.br/releases/industria-de-alimentos-cresce-67-em-2019>. Acesso em: 26 jun. 2021.

ANDERSON, D; SWEENEY, D; WILLIAMS, T. Estatística Aplicada à Administração e Economia 3.ed; São Paulo: Thomson Learning, 2013.

ALMEIDA, F.C, PASSARI, A.F. Aplicação de redes neurais na previsão de vendas no varejo, 2005. Disponível em: http://www.anpad.org.br/diversos/down_zips/9/enanpad2005-adid-0475.pdf. Acesso em: 01 jul. 2021

ANÔNIMO. A indústria de alimentos e bebidas na sociedade brasileira atual. Disponível em: <https://alimentosprocessados.com.br/industria-na-sociedade-brasileira.php>. Acesso em: 01 jul. 2021

ANÔNIMO. Amazon Forecast. Disponível em: <https://aws.amazon.com/pt/forecast/>. Acesso em: 27 jun. 2021.

ANÔNIMO. Análises Preditivas O que são e qual sua importância? Disponível em: https://www.sas.com/pt_br/insights/analytics/predictive-analytics.html. Acesso em: 5 nov.. 2021.

ANÔNIMO. Como identificar a sazonalidade das vendas. Disponível em: <https://www.sebrae.com.br/sites/PortalSebrae/ufs/ap/artigos/como-identificar-a-sazonalidade-das-vendas,f4b74341dedbc410VgnVCM2000003c74010aRCRD>. Acesso em: 26 jun. 2021.

ANÔNIMO. O que é análise prescritiva? Disponível em: <https://www.ibm.com/br-pt/analytics/prescriptive-analytics>. Acesso em: 01 nov. 2021.

ANÔNIMO. O que é análise preditiva? Disponível em: <https://www.ibm.com/br-pt/analytics/predictive-analytics>. Acesso em: 01 nov. 2021.

ANÔNIMO. PIB Brasil e Rio de Janeiro: resultados e projeções. Disponível em: <https://www.firjan.com.br/publicacoes/publicacoes-de-economia/pib-brasil-e-rio-de-janeiro-resultados-e-projecoes.htm>. Acesso em: 01 nov. 2021.

BESSA, A. B. Previsão de vendas no varejo de moda com modelos de redes neurais. [s. l.], 2018. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.8747144F&lang=pt-br&site=eds-live&scope=site>. Acesso em: 23 abr. 2021.

BUYA, R., CALHEIROS, R. and DASTJERDI, A. BIG DATA: Principles and paradigms. New York: Copyright Business Expert Press, 2016.

CHANDRAMOULI, S., DUTT, S. and KUMAR DAS. Machine Learning. New York: Copyright Business Expert Press, 2018

DOWNEY, A. Pense em Python: Pense Como um Cientista da Computação; Rio de Janeiro: Novatec, 2016.

ESCOVEDO, T. Machine Learning: Conceitos e Modelos - Parte I: Aprendizado Supervisionado*, 2020. Disponível em: <https://tatianaesc.medium.com/machine-learning-conceitos-e-modelos-f0373bf4f445>. Acesso em: 30 jun. 2021

FAVERO, E. P. Métodos de previsão de vendas: um estudo de caso numa rede varejista. [s. l.], 2015. Disponível em: <http://www.tcc.sc.usp.br/tce/disponiveis/18/18062300/tce-17112015-123704/publico/FaveroEduardoPerin.pdf>. Acesso em 23 abr. 2021.

FARIAS, D. M., KONZEN, P. H. A., SOUZA, R. R. Álgebra Linear - Um Livro Colaborativo. 2020. Pág. 96. Disponível em: <https://www.ufrgs.br/reamat/AlgebraLinear/livro/livro.pdf>. Acesso em: 20 out. 2021

FREUND, J; SIMON, G. Estatística Aplicada à Economia, Administração e Contabilidade; Porto Alegre: Bookman, 2004.

GOLFETTE, B. O que preciso saber sobre machine learning (com Python), 2019. Disponível em: <https://brunogolfette.com.br/artigos/o-que-preciso-saber-sobre-machine-learning-com-python/>. Acesso em: 02 jul. 2021.

HURWITZ, J. and KIRSCH, D. Machine Learning for dummies, IBM Limited Edition. New York: Copyright Business Expert Press, 2018.

LEMO, F.O. Metodologia para Seleção de Métodos de Previsão de Demanda. 2006. Disponível em: <http://www.producao.ufrgs.br/arquivos/publicacoes/FernandoOliveiraLemos.pdf>. Acesso em: 01 jul. 2021

LOUZADA, P. Previsão de Demanda: O que é? Qual é a sua importância?, 2020. Disponível em: <https://www.fm2s.com.br/previsao-de-demanda-o-que-e-qual-e-a-sua-importancia/>. Acesso em: 30 jun. 2021.

MANZANO, J. Introdução à Linguagem Python; Rio de Janeiro: Novatec, 2018.

MARTINS ASSIS, Pedro Miguel. Desenvolvimento de uma solução de Business Intelligence. Universidade de Lisboa, 2017

MENEZES, N. Introdução à Programação com Python: Algoritmos e Lógica de Programação Para Iniciantes; Rio de Janeiro: Novatec, 2019.

PAULA, E.S. Previsão de Vendas com Análise de Séries Temporais, 2019. Disponível em: <https://www.linkedin.com/pulse/previs%C3%A3o-de-vendas-com-an%C3%A1lise-s%C3%A9ries-temporais-everton-s-paula>. Acesso em: 02 jul. 2021.

PIMENTA, F.O. Previsão de demanda de uma empresa varejista de vinhos por meio de métodos quantitativos, 2016. Disponível em: http://pro.poli.usp.br/wp-content/uploads/2016/12/tf_completo_v13.pdf . Acesso em: 01 jul. 2021

QUEIROZ, A. A. CAVALHEIRO, D. Método de previsão de demanda e detecção de sazonalidade para o planejamento da produção de indústrias de alimentos. 2003. Disponível em: http://www.abepro.org.br/biblioteca/ENEGEP2003_TR0101_0801.pdf. Acesso em: 26 jun. 2021

SAHAY, A. Business Analytics, Volume I. New York: Copyright Business Expert Press, 2018.

SANTOS, A.C.L.S., NETO, M.S.O. Análise e previsão de demanda como ferramenta para controle de estoque: estudo de caso em indústria de equipamentos para piscinas e spas. Disponível em: [https://www.unifeso.edu.br/revista/index.php/revistacienciatecnologiainovacao/article/viewFile/939/483#:~:text=Chopra%20e%20Meindl%20\(2011\)%20enfatizam,desses%20desvios%20nos%20seus%20resultados.&text=Previs%C3%B5es%20de%20longo%20prazo%20%C3%AAm,que%20as%20de%20curto%20prazo](https://www.unifeso.edu.br/revista/index.php/revistacienciatecnologiainovacao/article/viewFile/939/483#:~:text=Chopra%20e%20Meindl%20(2011)%20enfatizam,desses%20desvios%20nos%20seus%20resultados.&text=Previs%C3%B5es%20de%20longo%20prazo%20%C3%AAm,que%20as%20de%20curto%20prazo). Acesso em: 27 jun. 2021.

SHARDA, R., DELEN, D., and TURAN, E. BUSINESS INTELLIGENCE E ANÁLISE DE DADOS. Porto Alegre: Bookman, 2019.

SILVA, A. F. Definição de um Modelo de Previsão das Vendas da Rede Varejista Alfabeto. [s. l.], 2008. Disponível em: https://www.ufjf.br/engenhariadeproducao/files/2014/09/2008_3_Andr%C3%A9.pdf. Acesso em 23 abr. 2021

STANFORD. Machine Learning. Disponível em: <http://mlclass.stanford.edu/>. Acesso em: 27 jun. 2021

SHARDA, R., DELEN, D., TURBAN, E. Business Intelligence e Análise de Dados para Gestão do Negócio 4.ed; Porto Alegre: Bookman, 2019

TUNES, S. Inteligência artificial contra a covid-19: especialistas da área do brasil e do mundo interrompem projetos para se dedicar à luta contra a pandemia. São Paulo: Fapesp, v. 291, 14 abr. 2020. Mensal. Disponível em: <https://revistapesquisa.fapesp.br/inteligencia-artificial-contra-a-covid-19/>. Acesso em: 20 jun. 2021.

VERGARA, S. C. Métodos de Pesquisa em Administração. São Paulo: Atlas, 2005. Disponível em: <https://madmunifacs.files.wordpress.com/2016/08/vergara-mc3a9todos-de-pesquisa-em-administrac3a7ao-sylvia-vergara.pdf>. Acesso em: 20 de out. 2021

WANG, C. B. Techno Vision II - Um Guia para Profissionais e Executivos Dominarem a Tecnologia e Internet. São Paulo: Makron Books, 1998. Disponível em:
<<https://repositorio.ufsc.br/bitstream/handle/123456789/79461/187911.pdf?sequence=1>>. Acesso em: 20 out. 2021.

WEBSTER'S NEW ENCYCLOPEDIC DICTIONARY. New York: Black Dog e Leventhal Publishers. 1961. Pág. 590.
Weiss, S. M. e Kulikowski, C. A. (1991). Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann Publishers Inc., São Francisco, CA, USA. [GS Search]. Disponível em:
<<https://dl.acm.org/citation.cfm?id=102700>>. Acesso em: 29 set. 2020.

Anexo 1

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
import pandas as pd
import numpy as np
from sklearn.svm import LinearSVC
from sklearn.model_selection import train_test_split
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.linear_model import LinearRegression
from sklearn.svm import LinearSVC
from sklearn.metrics import accuracy_score
import seaborn as sns
import matplotlib.pyplot as plt
```

```
dataset=pd.read_excel("/content/drive/MyDrive/Sucos/RioV1.xlsx")
x=pd.read_excel("/content/drive/MyDrive/Sucos/RioV1.xlsx")
y=pd.read_excel("/content/drive/MyDrive/Sucos/RioV1.xlsx")
x.tail()
```

	Indice	Semana	Ano	Temperatura	Chuva	Q1	Q2	Q3	Q4	Q5	
130	131	25	2021	20.375595	0.232143	23.142857	39.534247	23.012658	14.042553	18.117647	165.388
131	132	26	2021	21.483333	0.020238	24.464516	40.950000	19.500000	19.384615	17.204819	151.844
132	133	27	2021	20.238650	0.029448	26.478261	37.230769	16.936709	6.400000	16.981132	167.158
133	134	28	2021	20.211976	0.011976	31.384615	52.216216	20.640000	9.000000	28.285714	213.468
134	135	29	2021	22.346429	0.000000	44.400000	48.439024	20.896552	54.000000	16.380952	323.125

```
#dataset=pd.read_excel("/content/drive/MyDrive/Sucos/SP okok.xlsx")
#x=pd.read_excel("/content/drive/MyDrive/Sucos/SP okok.xlsx")
#y=pd.read_excel("/content/drive/MyDrive/Sucos/SP okok.xlsx")
#x.tail()
```

https://colab.research.google.com/drive/1z2YVgye0u2dTc2XemPCGf9i_18Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
x=x.drop('PQ1',axis=1)
x=x.drop('PQ2',axis=1)
x=x.drop('PQ3',axis=1)
x=x.drop('PQ4',axis=1)
x=x.drop('PQ5',axis=1)
x.head()
```

	Indice	Semana	Ano	Temperatura	Chuva	Q1	Q2	Q3	Q4	Q5	P1
0	1	1	2019	28.410833	0.070000	24.929348	4.555556	17.795745	9.752294	7.612156	137.319130
1	2	2	2019	26.433333	0.000000	18.771277	4.571429	14.008511	8.585586	7.406995	100.161596
2	3	3	2019	26.499405	0.000000	27.444444	4.000000	16.967742	7.080000	7.232283	149.406275
3	4	4	2019	28.484524	0.241667	23.705882	5.888889	13.669118	6.857143	7.719039	122.715049
4	5	5	2019	28.295833	0.000000	23.285714	3.571429	15.136150	9.983740	7.462199	123.115330

```
y=y.drop('Indice',axis=1)
y=y.drop('Semana',axis=1)
y=y.drop('Ano',axis=1)
y=y.drop('Temperatura',axis=1)
y=y.drop('Chuva',axis=1)
y=y.drop('Q1',axis=1)
y=y.drop('Q2',axis=1)
y=y.drop('Q3',axis=1)
y=y.drop('Q4',axis=1)
y=y.drop('Q5',axis=1)
y=y.drop('P1',axis=1)
y=y.drop('P2',axis=1)
y=y.drop('P3',axis=1)
y=y.drop('P4',axis=1)
y=y.drop('P5',axis=1)
#y
```

```
modelo=LinearRegression()
```

https://colab.research.google.com/drive/1z2YVgve0ii2dTc2XemPCGf9i_1t8Cgn-P#scrollTo=V8vbsS4yc0_9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
treino_x, teste_x, treino_y, teste_y = train_test_split(x, y, test_size=0.10)
modelo.fit(treino_x, treino_y)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
resultado=modelo.score(teste_x, teste_y)
print("Acurácia: ", resultado)
```

```
Acurácia: 0.5294038914281376
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:434: FutureWarning: The default value of multioutput (not exposed in score method) will change from 'best_averaging' to 'uniform_average'.
  "multioutput='uniform_average'").", FutureWarning)
```

```
# Dados Rio
dadosN2=pd.DataFrame({'Indice': [136], 'Semana': [30],
'Ano': [2021], 'Temperatura': [20],
'Chuva': [0], 'Q1': [27.57], 'Q2': [50.94],
'Q3': [22.84], 'Q4': [26.57], 'Q5': [22.62],
'P1': [199.31], 'P2': [388.07],
'P3': [150.16], 'P4': [233.54], 'P5': [242.14], })
```

```
# Dados SP
#dadosN2=pd.DataFrame({'Indice': [136], 'Semana': [30],
#'Ano': [2021], 'Temperatura': [13.85],
#'Chuva': [0], 'Q1': [32.72], 'Q2': [25.74],
#'Q3': [26.06], 'Q4': [20.11], 'Q5': [11.60],
#'P1': [178.46], 'P2': [177.96],
#'P3': [162.29], 'P4': [85.74], 'P5': [128.59], })
```

dadosN2

	Indice	Semana	Ano	Temperatura	Chuva	Q1	Q2	Q3	Q4	Q5	P1	P2	P3	P4	P5
0	136	30	2021	20	0	27.57	50.94	22.84	26.57	22.62	199.31	388.07	150.16	233.54	242.14

+ Código

+ Texto

```
previsoesN2=modelo.predict(dadosN2)
previsoesN2
```

https://colab.research.google.com/drive/1z2YVgye0u2dTe2XemPCGf9t_i8Cgn-P#scrollTo=V8vbs4yc0-9&printMode=true

3/18

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
# RIO 27.      50.      22.      26.      22.
# RIO 18.54    34.28    18.67    10.94    17.09    NOV 06
# SP  32.      25.      26.      20.      11.
# SP  28.38    20.89    20.52    22.      19.22    NOV 06
```

```
array([[38.42965618, 49.61284      , 21.19984088, 29.5853323 , 22.49895261]])
```

```
dadoscorr=dataset.corr()
dadoscorr
```

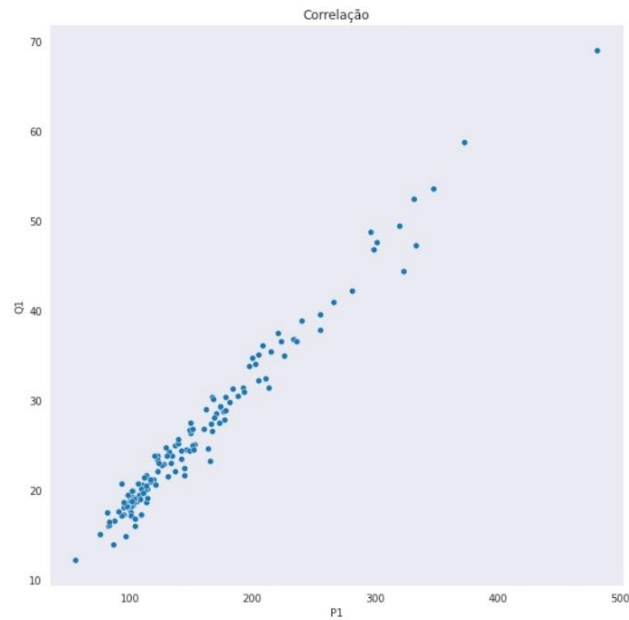
https://colab.research.google.com/drive/1z2YVgyc0ii2dTe2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

Índice Semana Ano Temperatura Chuva Q1 Q2 Q3 Q4

```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
plt.figure(figsize=(10,10))
plt.title('Correlação')
g=sns.scatterplot(x="P1",y="Q1",data=dataset)
```



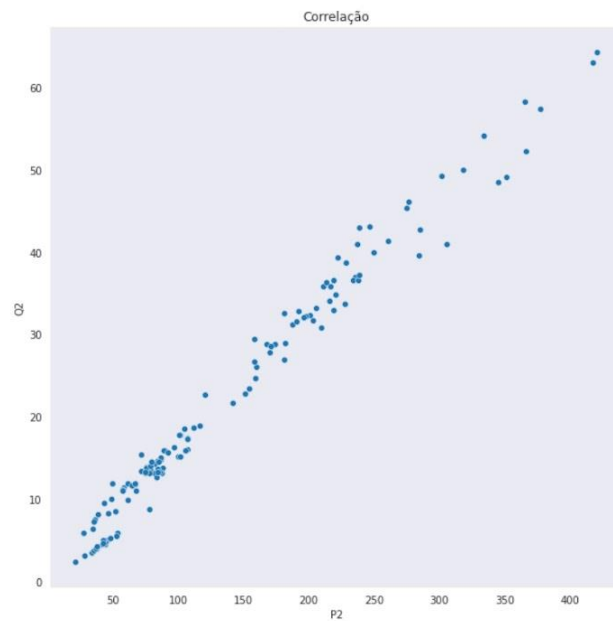
```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
```

https://colab.research.google.com/drive/1z2YVgye0ii2dTc2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
plt.figure(figsize=(10,10))
plt.title('Correlação')
g=sns.scatterplot(x="P2",y="Q2",data=dataset)
```

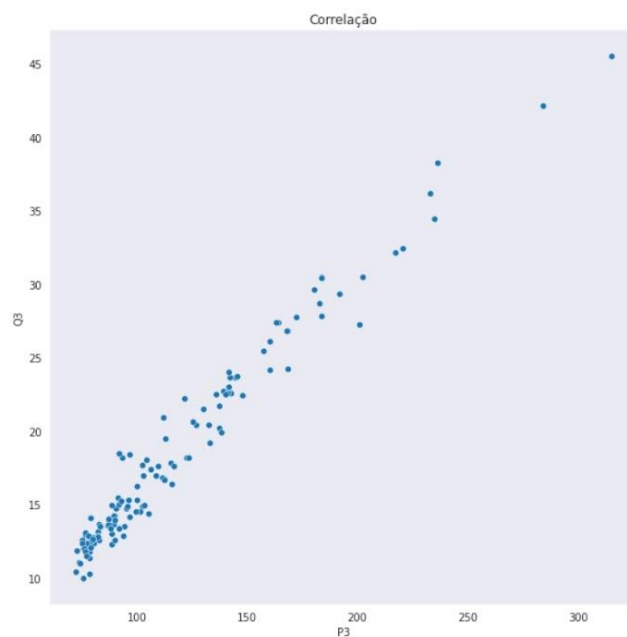


```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
plt.figure(figsize=(10,10))
plt.title('Correlação')
g=sns.scatterplot(x="P3",y="Q3",data=dataset)
```

https://colab.research.google.com/drive/1z2YVgye0i2dTc2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

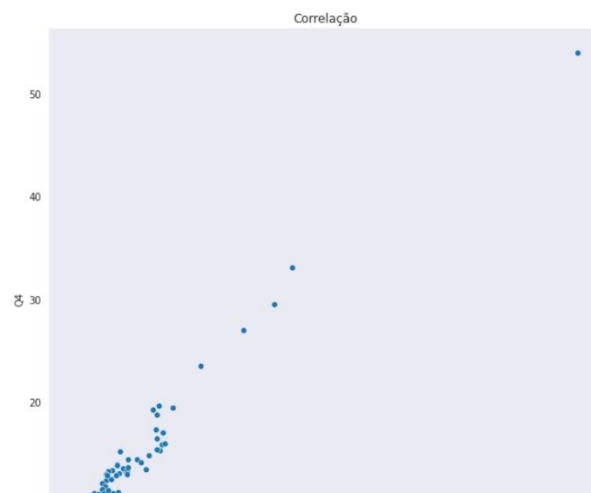


```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
plt.figure(figsize=(10,10))
plt.title('Correlação')
g=sns.scatterplot(x="P4",y="Q4",data=dataset)
```

https://colab.research.google.com/drive/1z2YVgve0ii2dTe2XemPCGf9i_18Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory



```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
plt.figure(figsize=(10,10))
plt.title('Correlação')
g=sns.scatterplot(x="P5",y="Q5",data=dataset)
```

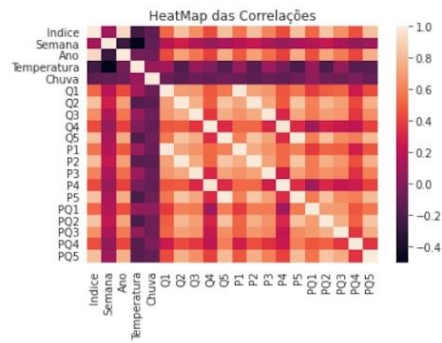
https://colab.research.google.com/drive/1z2YVgYe0u2dTe2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory



```
sns.heatmap(dataset.corr())
plt.title('HeatMap das Correlações')
plt.show()
```



```
fig=plt.figure()
```

https://colab.research.google.com/drive/1z2YVgye0ii2dT2XemPCGf9l_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```

ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Q1, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Quantidade 1", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.P1, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Pedidos 1', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Quantidade 1 e Pedidos 1',color="g")
plt.show()

```



```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

```

https://colab.research.google.com/drive/1z2YVgye0ii2dT2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4ye0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```

ax.plot(dataset.Indice, dataset.Q2, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Quantidade 2", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.P2, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Pedidos 2', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Quantidade 2 e Pedidos 2',color="g")
plt.show()

```



```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Q3, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Quantidade 3", color="C0")

```

https://colab.research.google.com/drive/1z2YVgye0u2dTe2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0.9&printMode=true

10/11/2021 02:42

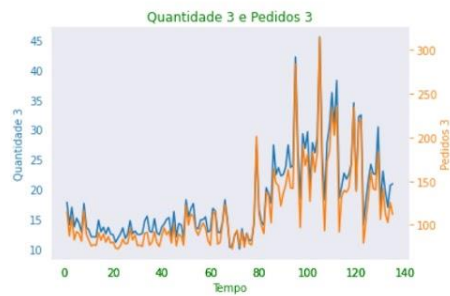
Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```

ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.P3, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Pedidos 3', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Quantidade 3 e Pedidos 3',color="g")
plt.show()

```



```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Q4, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Quantidade 4", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

```

https://colab.research.google.com/drive/1z2YVgye0ii2dT2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4ye0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```

ax2.plot(dataset.Indice, dataset.P4, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Pedidos 4', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Quantidade 4 e Pedidos 4',color="g")
plt.show()

```



```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Q5, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Quantidade 5", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.P5, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Pedidos 5', color="C1")

```

https://colab.research.google.com/drive/1z2YVgye0u2dTzXemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Quantidade 5 e Pedidos 5',color="g")
plt.show()
```



```
fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Temperatura, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Temperatura", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

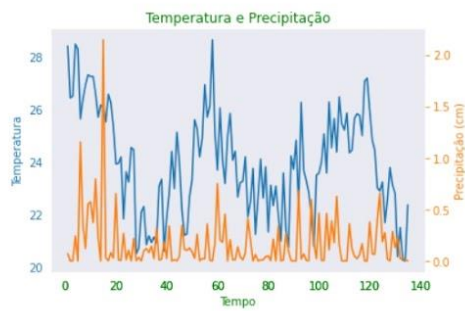
ax2.plot(dataset.Indice, dataset.Chuva, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Precipitação (cm)', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
```

https://colab.research.google.com/drive/1z2YVgye0i2dTe2XemPCGf9t_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

```
ax2.tick_params(axis='y', colors="C1")
plt.title('Temperatura e Precipitação',color="g")
plt.show()
```



```
# Criando o ambiente gráfico (Gráfico de Dispersão)
sns.set_style("dark")
plt.figure(figsize=(8,6))
plt.title('Correlação')
g=sns.scatterplot(x="Temperatura",y="Chuva",data=dataset)
```

https://colab.research.google.com/drive/1z2YVgye0ü2dTe2XemPCGf9i_18Cgn-P#scrollTo=V8vbsS4yc0_9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK.ipynb - Colaboratory

Correlação

2.0

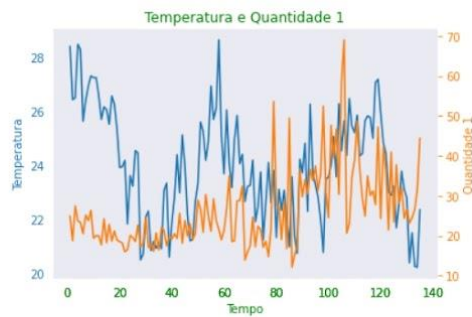
```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Temperatura, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Temperatura", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.Q1, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Quantidade 1', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Temperatura e Quantidade 1',color="g")
plt.show()

```



https://colab.research.google.com/drive/1z2YVgye0ü2dTe2XemPCGf9i_t8Cgn-P#scrollTo=V8vbs44yc0-9&printMode=true

10/11/2021 02:42

Sucos2021_V3_SPOKOK ipynb - Colaboratory

```

fig=plt.figure()
ax=fig.add_subplot(111, label="1")
ax2=fig.add_subplot(111, label="2", frame_on=False)

ax.plot(dataset.Indice, dataset.Chuva, color="C0")
ax.set_xlabel("Tempo", color="g")
ax.set_ylabel("Precipitação", color="C0")
ax.tick_params(axis='x', colors="g")
ax.tick_params(axis='y', colors="C0")

ax2.plot(dataset.Indice, dataset.Q1, color="C1")
ax2.yaxis.tick_right()
ax2.set_ylabel('Quantidade 1', color="C1")
ax2.xaxis.set_label_position('top')
ax2.yaxis.set_label_position('right')
ax2.tick_params(axis='x', colors="g")
ax2.tick_params(axis='y', colors="C1")
plt.title('Precipitação e Quantidade 1',color="g")
plt.show()

```



https://colab.research.google.com/drive/1z2YVgye0ii2dT2XemPCGf9i_t8Cgn-P#scrollTo=V8vbsS4yc0-9&printMode=true