|  |
| --- |
|  |
| Fatores macro econômicos no preço da gasolina para o consumidor final |
| Analise do preço da gasolina e dos fatores que influenciam seu valor |

Felipe Delamar

Mateus Paape

Kelvin Spaler

Componentes do Grupo

Felipe Delamar Magalhães

Kelvin Spaler

Mateus Paape

Sumário

[1. Introdução 3](#_Toc74237447)

[2. Objetivo 3](#_Toc74237448)

[3. Levantamento de dados 3](#_Toc74237449)

[4. Tratamento dos dados 4](#_Toc74237450)

[6. Machine learning 5](#_Toc74237451)

[11. Referências 9](#_Toc74237452)

1. Introdução

Nos tempos atuais, pós pandemia, podemos observar a reabertura gradativa dos mercados internacionais e nacionais. A crise gerada, devido a esta questão, ainda está sendo observado por estudiosos de diversas disciplinas distintas. No espectro da ciência de dados pode-se explicar diversas influências na macroeconomia geradas devido à crise.

Um dos assuntos que repercutiu bastante foram os reajustes no preço da gasolina tipo C no mercado brasileiro. Segundo informações vindas da Petrobras, no qual será detalhado no presente trabalho, o preço de seu combustível está atrelado diretamente a cotação do dólar e o preço do barril de petróleo.

1. Objetivo

O objetivo deste trabalho é, utilizando a linguagem de programação R, levantar e demonstrar as correlações de preços que impactam no preço final ao consumidor, no caso, preço na bomba dos postos de combustível. Também será demonstrado previsões de reajustes nos preços utilizando técnicas de machine learning.

1. Levantamento de dados

Para a coleta de informações foi necessário saber o que influencia o preço da gasolina tipo C nos postos de combustível brasileiro. Segundo (ANP, 2021), a gasolina tipo A é a gasolina pura, produzida diretamente pelas refinarias, ela se torna a gasolina tipo C após ser retirada das refinarias pelo distribuidor que por sua vez faz a mistura da Gasolina tipo A com álcool etílico anidro combustível, mais conhecido como etanol.

Este trabalho utilizou informações públicas divulgadas por sites confiáveis, a primeira fonte de dados foram as tabelas de composição do preço da gasolina (ANP, 2021).

Conforme é observado abaixo na figura 1, os dados extraídos do site da ANP estão separados por períodos mensais do ano e até a presente data do trabalho estavam disponíveis dados a partir de julho de 2018 até julho de 2020.

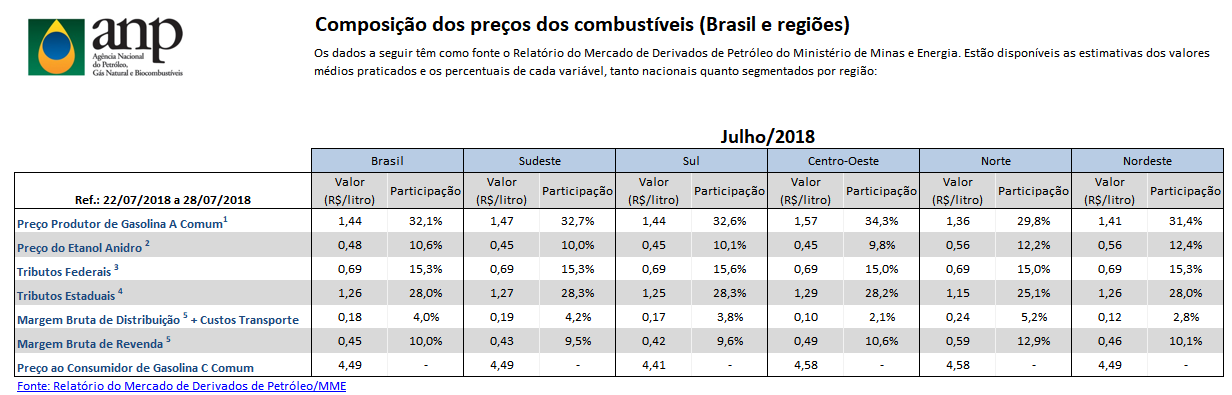


Figura 1 - Composição do preço da Gasolina tipo C

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

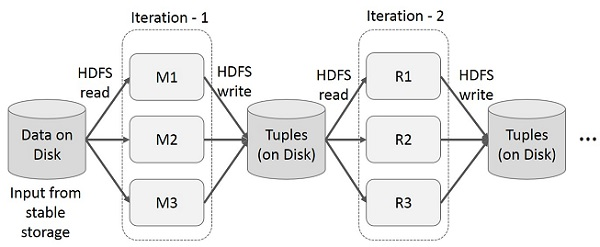


Figura 2 – Estrutura de

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

1. Tratamento dos dados

Mencionar o R, explicando que é a linguagem que será utilizada no trabalho responsável pela manipulação da informação, geração de dataframes e posteriormente utilizará bibliotecas de correlação, plotagem e machine learning.

1. Correlação de dados

Inicalmente será explicado o que é correlação e para que a utilizaremos.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

1. Machine learning

Aqui falaremos um pouco sobre o que é machine learning, qual técnica utilizaremos e o que queremos prever.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Segundo a postagem do site databrickso Spark SQL absorveu funções do Shark; Este no qual, originalmente, foi a primeira ferramenta rodando dentro do Spark trabalhos SQL de maneira interativa.

Na figura abaixo é demonstrado a arquitetura do Spark SQL:

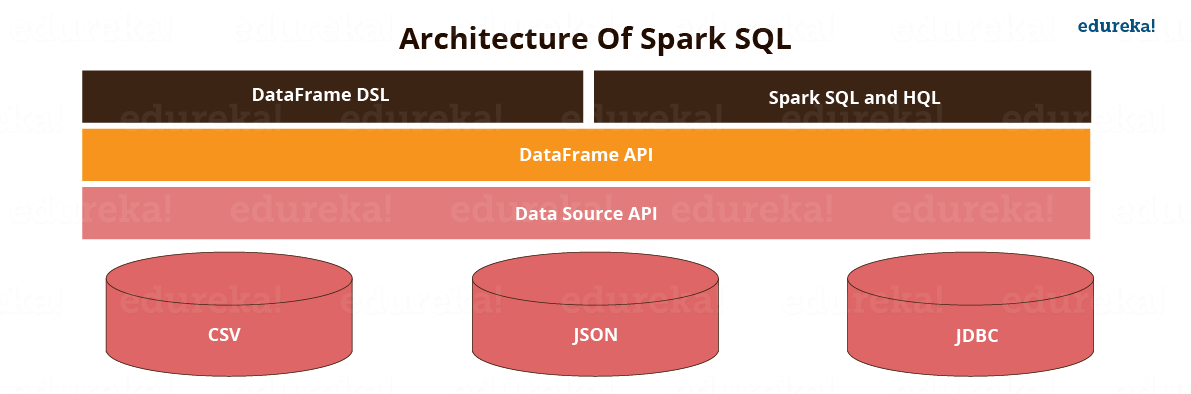


Figura 3 – Arquitetura do Spark

Segundo o site Spark.Apache, os dados importados ao Spark SQL podem ser separados de duas formas: DataSets e DataFrames

* Dataset:

O dataset é um conjunto de dados distribuídos, possuí uma interface que provê os beneficios do RDD com os beneficios do uso da engine de otimização de execução do Spark SQL.

A API do Dataset só está disponível em Scala e Java.

* DataFrame:

DataFrame é um dataset organizado em colunas nominares. Conceitualmente é equivalente a uma tabela em um relacionamento de tabelas ou data frame em python/R. DataFrames podem ser construídos de diversas fontes de dados como: Dados de arquivos estruturados, tabelas em Hive, tabelas externas e RDDs existentes.

DataFrame estão disponíveis em Python, a linguagem que iremos usar neste estudo para a execução dos trabalhos dentro do ambiente Spark SQL

1. Desenvolvimento

Aqui será detalhado o tratamento que utilizamos em cada dataframe, a onde foi utilizado cada biblioteca no código e o dataframe gerado para consumo de gráficos e machine learning.

Será explicado também o estudo da correlação de tudo o que foi levantado de informação.

Demonstrado como foi utilizado o machine learning, a biblioteca de R que foi utilizado e os parâmetros adotas e o porque

1. Apresentação e resultado

Aqui será apresentado todo o resultado obtido, explicando fatos relevantes que levou ao resultado.

\*Os resultados são produtos gerados durante o projeto. • Podem ser apresentados diversos pontos de execuções de sua aplicação, com a devida análise. • A análise está relacionada ao que é esperado do projeto, portanto, explique como foi obtido tal resultado e se ele é satisfatório. • Justifique eventuais problemas ou melhorias, se necessário. • Visualizações de dados são formas muito eficientes e esperadas de comunicação em apresentações deste tipo!

Limitações do Hive

1. Considerações finais

Aqui devemos falar sobre o resultado demonstrando se atingiu as expectativas, o que ele pode agregar para o futuro e as motivações apresentadas na introdução.

Inserir recomendações para melhoria do projeto e para projetos futuros.

1. Lista de Figuras

[Figura 1 – Estrutura de interações (Tutorialspoint, s.d.) 4](#_Toc27083231)

[Figura 2 – Arquitetura do Spark SQL (Dayananda, 2019) 5](#_Toc27083232)

[Figura 3 – Painel de administração do cluster site AWS 7](#_Toc27083233)

[Figura 4 - Configuração detalhada do cluster EMR site AWS 7](#_Toc27083234)

# Referências

ANP. (2021). Fonte: Agência Nacional de Petróleo, Gás Natural e Biocombustíveis.: https://www.gov.br/anp/pt-br