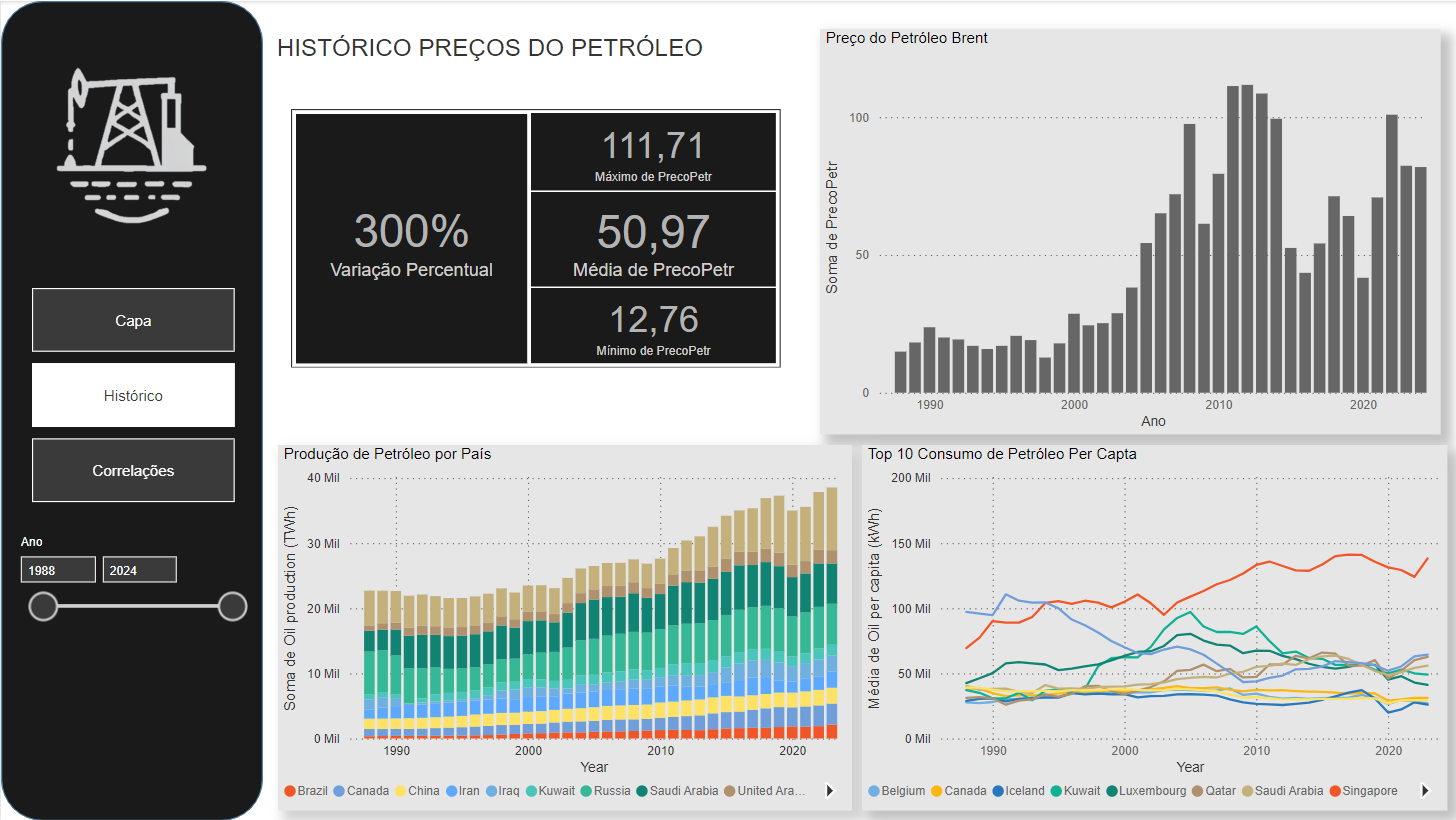
**ESTUDO DA VARIAÇÃO DO PREÇO DO PETRÓLEO**

De acordo com a solicitação do cliente do segmento foi desenvolvido um dashboard interativo para gerar insights relevantes para tomada de decisão.

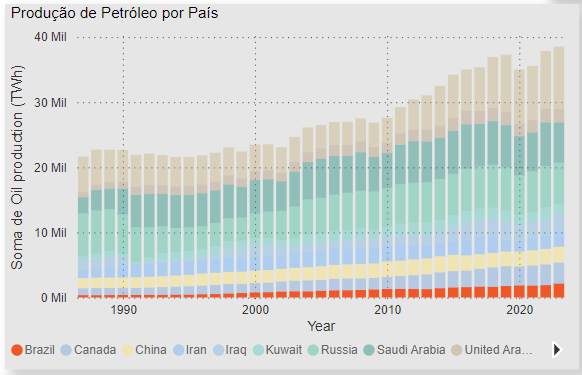
Como ferramenta para criação do Dashboard foi utilizado o Power BI. Essa é uma ferramenta de **business intelligence** (BI) da Microsoft que permite conectar, transformar e visualizar dados de maneira interativa. Ele é amplamente usado para criar relatórios e dashboards dinâmicos que auxiliam na análise de informações.

Foram criadas duas abas com informações no dashboard (Histórico e Correlações). O objetivo é gerar insights relevantes sobre a variação do preço do petróleo, como situações geopolíticas, crises econômicas, demanda global por energia e etc. Seguem os insights observados no texto abaixo.

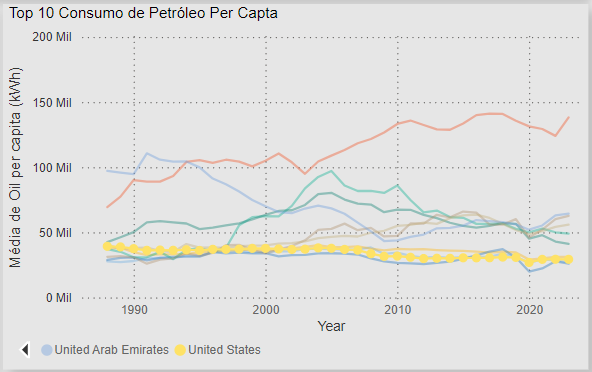
**Aba Histórico**

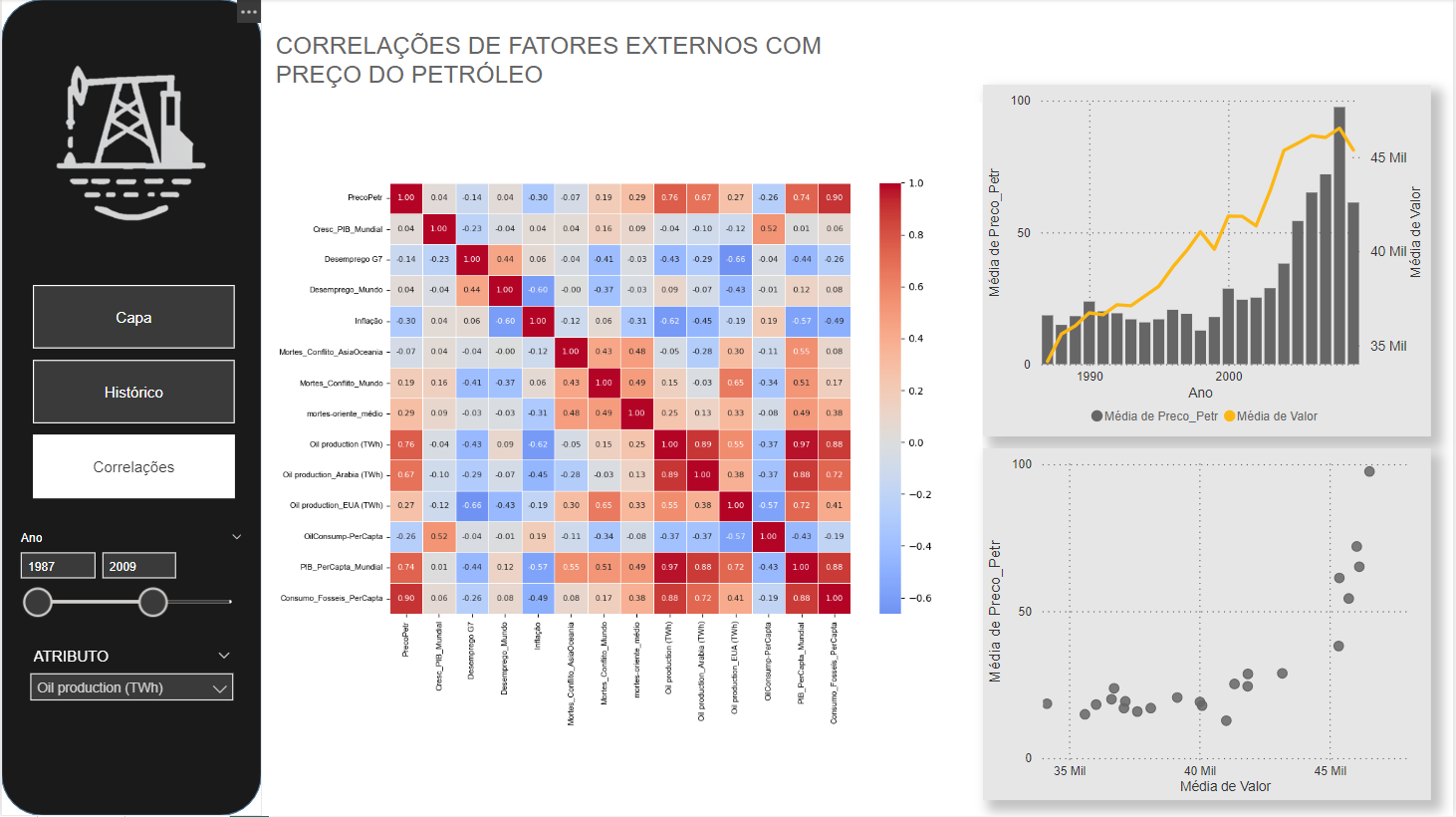


1. O período observado engloba os anos de 1987 a 2024, a variação no período foi de 300%. Nos primeiros anos observados, o preço se manteve relativamente estável, e passou a apresentar maiores flutuações a partir do ano de 2002. Além disso, o momento atual não representa o período de preço mais elevados, este ocorreu entre os anos de 2011 a 2013 ($111,29 – $108,57).
2. Produção de petróleo concentrada em alguns países. Dos quais, Estados Unidos, Arábia Saudita e Rússia são os 3 maiores produtores hoje. A produção total apresenta tendência incremental ao longo dos anos observados, impulsionada tanto pelo aumento da produção desses 3 países principais, quanto nos demais produtores.
3. O Brasil se encontra entre os 10 países que mais produzem petróleo no mundo. O gráfico abaixo mostra tendência de crescimento na produção de petróleo no Brasil, além de aumento da relevância do país neste contexto.



1. Entre os 10 países com maior consumo per capta, Singapura apresenta uma demanda crescente e muito acima dos demais países. Enquanto, outros países como Emirados Árabes apresentam uma redução na tendência de consumo.
2. Outro insight é que nas últimas décadas, o consumo per capta dos Estados Unidos se mantém no mesmo patamar, com uma leve redução no índice de consumo per capta.



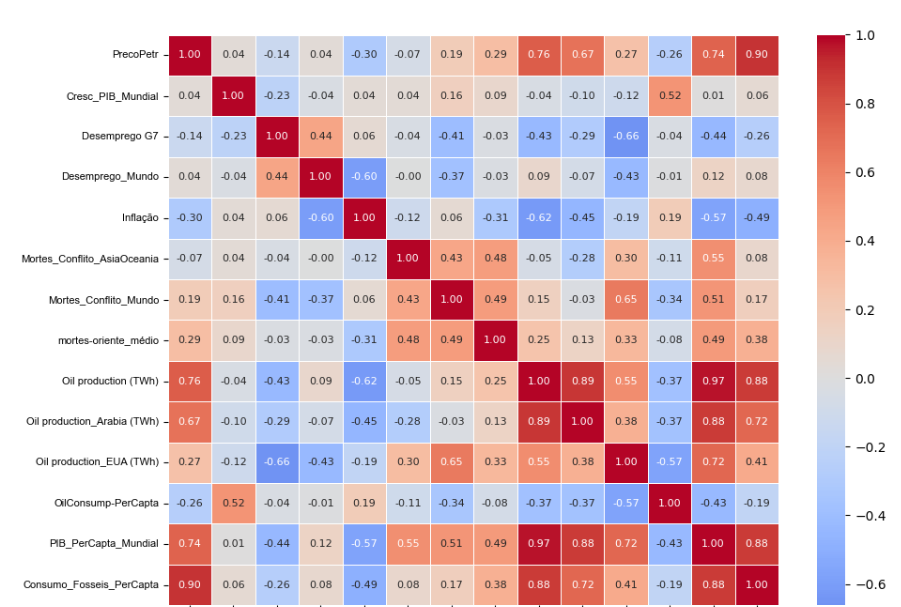
**Aba de Correlações**

Na aba de Correlações, foram avaliados 13 parâmetros em busca de correlações que pudessem trazer informações extras para compor a análise.

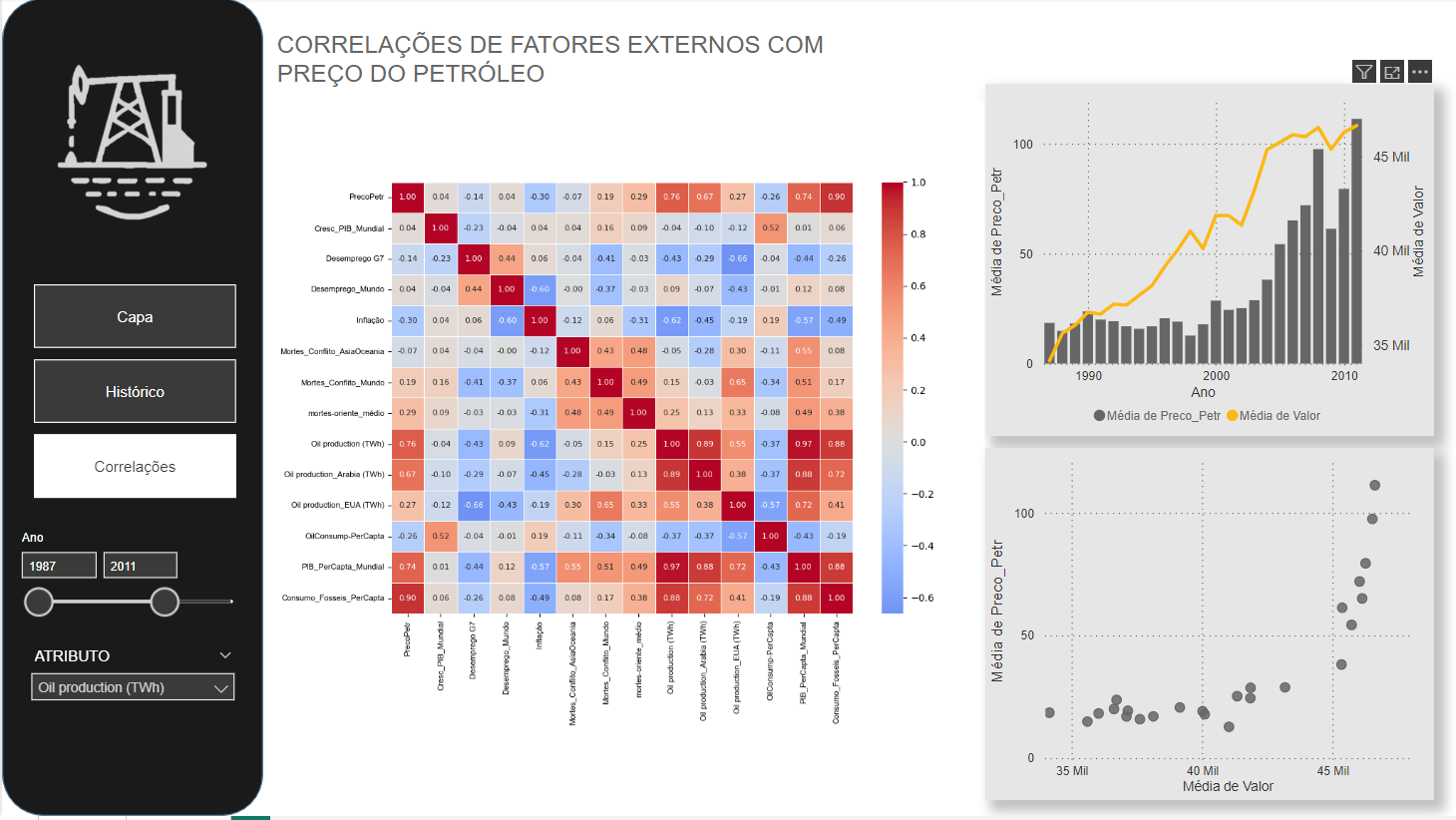
Aspectos e Parâmetros avaliados:

* **Desenvolvimento Econômico** – Crescimento do PIB Mundial, Índice de Desemprego no G7, Índice de Desemprego no Mundo, Inflação no Mundo, PIB per Capta Mundo
* **Fatores de Conflitos e Guerras** – Nº de Mortes em Conflito na Ásia e Oceania, Nº de Mortes em Conflitos no Mundo, Nº de Mortes no Oriente Médio
* **Demanda** – Consumo de Petróleo per Capta, Consumo de Combustíveis Fósseis per Capta
* **Oferta** – Produção de Petróleo no Mundo, na Arábia Saudita e Estados Unidos

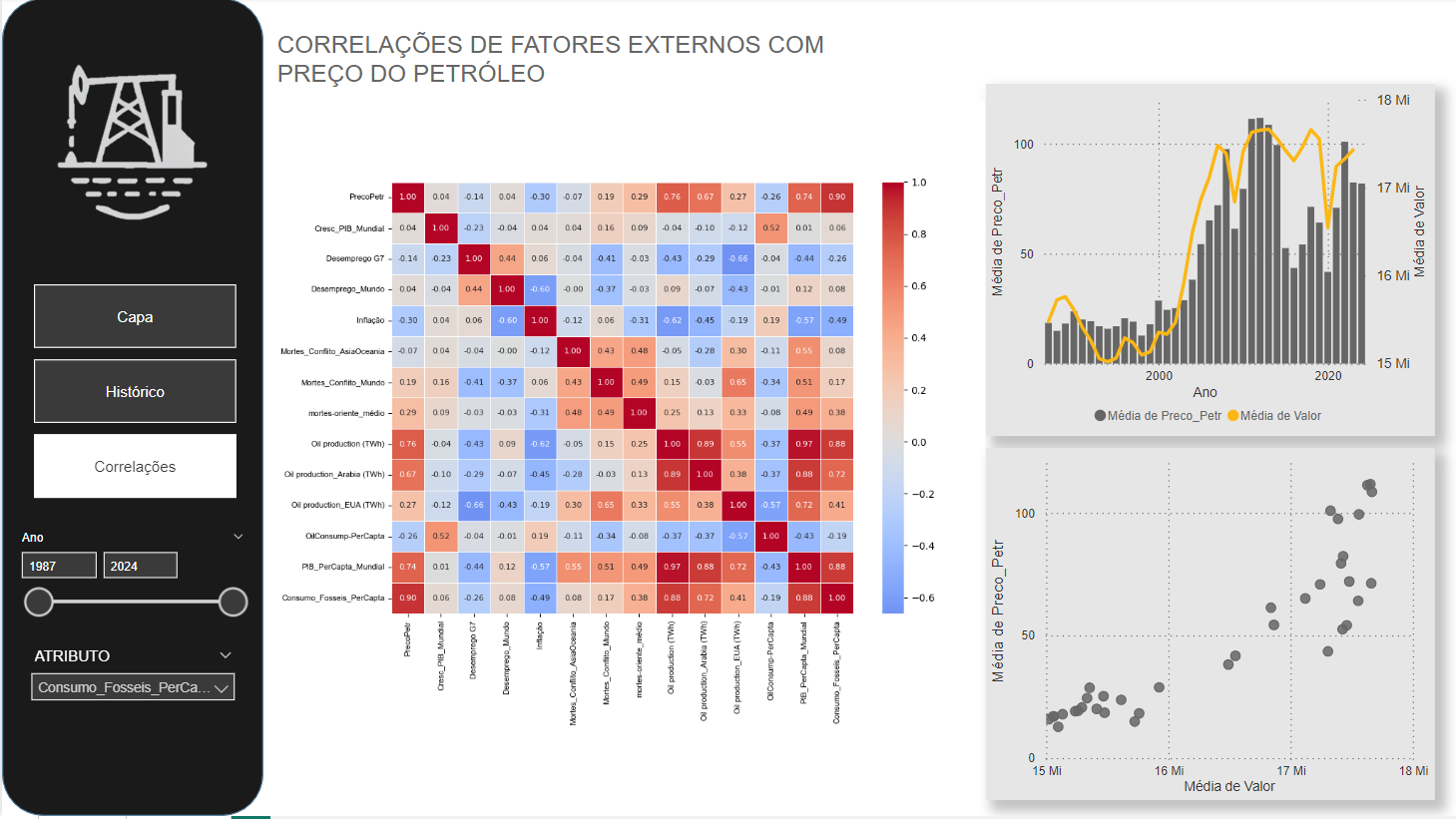
Segue abaixo heatmap de correlações, com destaque nos 3 itens com maior relevância:



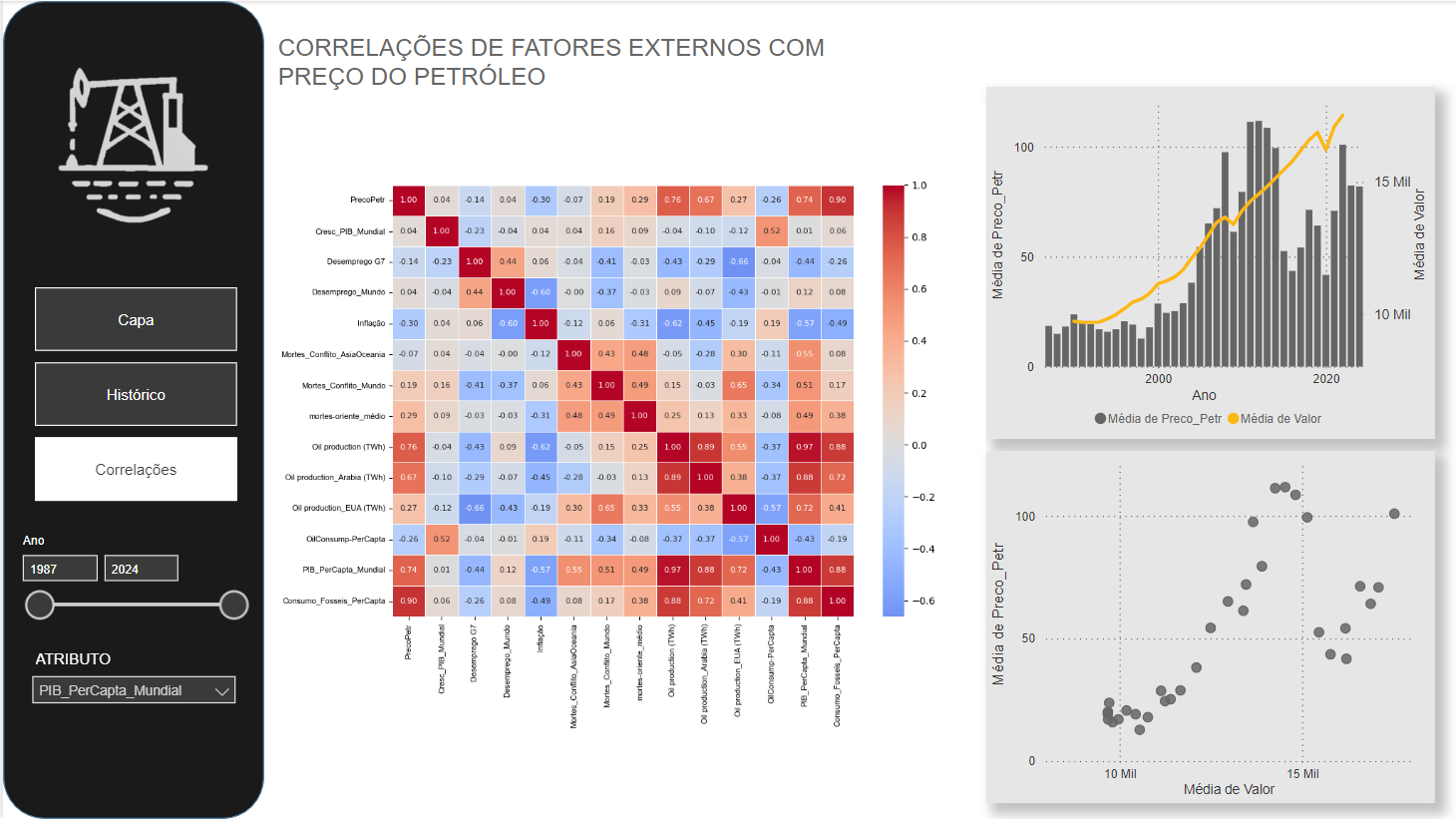
1. A produção total de petróleo no mundo e o seu preço apresentam correlação positiva. Conforme gráfico destacado abaixo, até o ano de 2011, o preço do petróleo apresenta um comportamento exponencial com relação à quantidade produzida.



1. Preço do petróleo e Consumo de Combustíveis Fósseis apresentam correlação de 0,90 e pelos gráficos podemos notar um comportamento similar de subidas e descidas, além de correlação próxima à linearidade no gráfico scatterplot.



1. O preço do petróleo e o indicador PIB Per capta Mundial (que representa o cenário econômico mundial) apresentam também correlação positiva. No gráfico scatterplot, aparenta haver duas sequencia lineares.



Utilizando os insights evidenciados neste texto, além de outros que podem ser obtidos diretamente do Dashboard, é possível obter uma compreensão ampla da variação do preço do petróleo ao longo dos anos, além de fatores que podem influenciar ou serem influenciados pelo seu comportamento.

**Código para previsão do preço do petróleo**

Realizado a leitura dos dados através:

# Leitura dos dados

url = "https://raw.githubusercontent.com/andremanhas/TechChallenge4/main/ipeadata.xlsx"

series = pd.read\_excel(url, engine='openpyxl')

Conseguimos uma série de dados com a informação diária do preço do petróleo:

Série Original

preco

data .

2024-01-01 NaN

2024-01-02 76.24

2024-01-03 77.18

2024-01-04 75.79

2024-01-05 78.31

Observa-se que na base contêm diversos valores NaN, ou seja, nulos, nos dias que não são úteis, logo eles foram excluídos para melhoria da previsão.

Os valores coletados se comportam da forma abaixo durante o período de 2020 a 2024 (tempo utilizado para predição)

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Analisando mais a fundo os componentes da série temporal utilizando a função sazonal\_decompose, conseguimos a informação de que estamos lidando com uma série que possui sazonalidade e com tendencia de crescimento.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

O mesmo se comprova ao analisarmos fatores como ACF, PACF e o teste de Dickey-Fuller Aumentado.

Gráfico

Descrição gerada automaticamente

Dickey-Fuller Aumentado

“Teste Estatístico: -1.4199

Valor-p: 0.5727

Number of lags used: 6.0000

Valores Críticos:

1%: -3.4357

5%: -2.8639

10%: -2.5680

O teste Dickey-Fuller Aumentado mostra que a série é: Não estacionária”

Para isso aplicamos uma diferenciação, para deixar a série estacionária, conseguindo novos valores para o teste de Dickey-Fuller.

“Teste Estatístico: -16.1833

Valor-p: 0.0000

Number of lags used: 5.0000

Valores Críticos:

1%: -3.4357

5%: -2.8639

10%: -2.5680

O teste Dickey-Fuller Aumentado mostra que a série é: Estacionária”

Foi então separado os dados entre teste e treino, com 80% para treino e 20% para teste.

Para os parâmetros do SARIMAX utilizamos auto\_arima que foi capaz de encontrar o melhor parâmetro a ser utilizado:

Arima\_stepwise = auto\_arima(train, seasonal=False, trace=True, error\_action='ignore', suppress\_warnings=True, stepwise=True)

Feito isso chegamos nos valores: Best model: ARIMA(2,1,2)(0,0,0)[0]

Com isso o modelo foi treinado:

model = SARIMAX(train, order=(2, 1, 2), seasonal\_order=(0, 0, 0, 0))

model\_fit = model.fit(disp=False)

E feito previsões passo a passo para melhor performance do modelo

predictions = []

history = train.tolist()  # Conjunto de dados para atualização iterativa

for t in range(len(test)):

    model = SARIMAX(history, order=(2, 1, 2), seasonal\_order=(0, 0, 0, 0))

    model\_fit = model.fit(disp=False)

    yhat = model\_fit.forecast(steps=1)[0]

    predictions.append(yhat)

    history.append(test.iloc[t])

Onde chegamos nesses resultados:

Gráfico, Gráfico de linhas, Histograma

Descrição gerada automaticamente

Por fim podemos dizer que a previsão tem uma grande chance de acerto de acordo com as métricas abaixo:

ARIMA Metrics:

MAE: 1.0969034436824738

MSE: 2.089692655692525

MAPE: 1.3618913682037124%

**Link Streamlit.io**

No link abaixo é possível utilizar do algoritmo que realiza a previsão do preço do petróleo

<https://pretoleo-forecast.streamlit.app/>

**Link GitHub**

No link abaixo se encontram todos os arquivos utilizados para este trabalho, incluindo o dashboard em power bi, as bases, arquivos python e etc

<https://github.com/andremanhas/TechChallenge4>