

**BLUE EDTECH**

**ANA CRISTINA CHAVES  
ANTONIO DUARTE MARCOS JUNIOR**

**BOOTCAMP - PESQUISA DE IMPACTO DO CORONAVÍRUS 2020  
SPRINT 4**

**BRASIL  
JANEIRO 2023**

## **1. Introdução**

Neste relatório são apresentados os principais resultados obtidos durante a quarta sprint do bootcamp. Para esta sprint os objetivos exigidos foram:

- Anotação das sugestões da pré-banca.
- Avaliar e aplicar as sugestões da pré-banca em relação ao projeto.
- Avanço do desempenho do modelo demonstrado pelo relatório.
- Ajustar a apresentação de acordo com as sugestões da pré-banca.
- Ajustar o dashboard de acordo com as sugestões da pré-banca.

Desta forma este relatório foi dividido em cinco seções que mostram os resultados de cada um dos itens exigidos.

## **2. Sugestões da Pré banca**

- Incluir nos slides a parte de extração dos dados, transformação, arquitetura e modelo.
- Mudar a cor da dica de ferramenta no dashboard.
- Ajustar os relacionamentos de tabelas.
- Harmonizar o gráfico nuvem de palavras com o plano de fundo.
- Apresentar a arquitetura dos dados e como foram salvos.

## **3. Aplicar as sugestões**

Foram alteradas conforme a orientação, a maioria das alterações estavam nos slides de apresentação e dashboard.

## **4. Avanço do modelo**

Modelo já está concluído, serão reavaliadas para melhor performance.

## **5. Apresentação**

Foram incluídas as partes de extração de dados, as transformações, os modelos de machine learning no slide, o dashboard será apresentado à parte.

Incluimos na apresentação as ferramentas utilizadas para todo o processo de arquitetura, Apache AirFlow, Apache PySpark, python, pandas, AWS, Statsmodel, mysql entre outras ferramentas.

Apresentamos também os locais que foram extraídos os dados, tanto os solicitados pela banca avaliadora quanto as tabelas auxiliares para enriquecimento do conteúdo.

Incluimos a estrutura de Datalake utilizada, modelo medallion, cujo as pastas bronze estão os dados coletados, na silver os dados processados e transformados e na gold os dados utilizados no dashboard.

Montamos um data warehouse, com os dados que serão utilizados direto no dashboard, a princípio ainda não conectamos o powerbi no AWS para esta sprint, mas para a próxima será atualizado em tempo real.

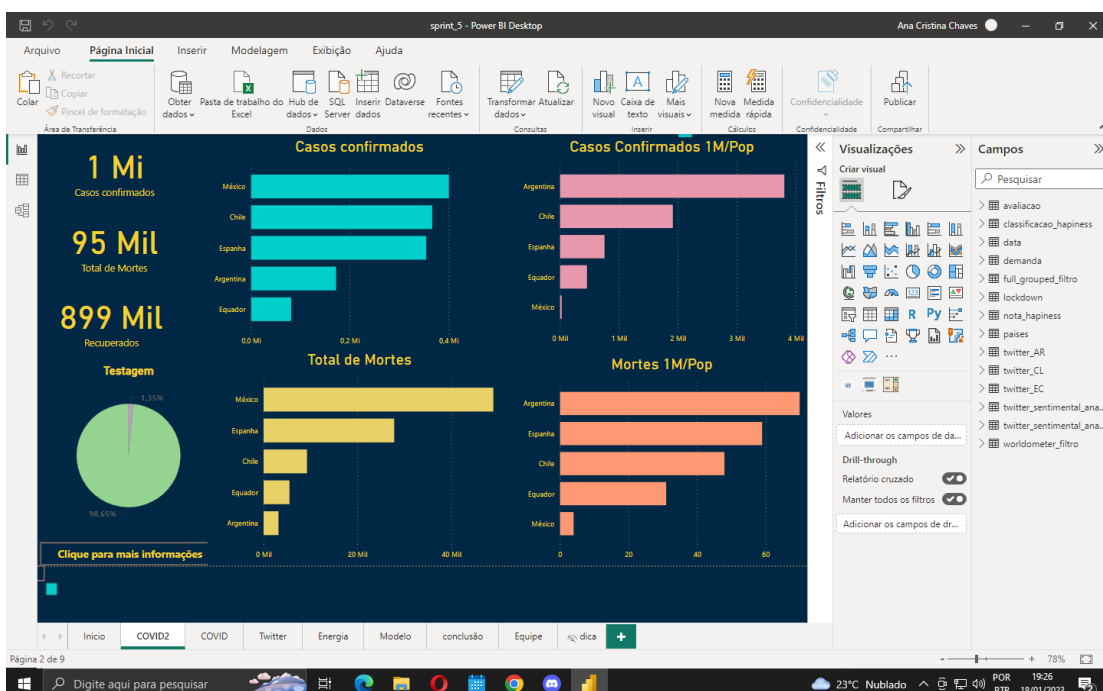
Demonstramos os modelos de machine learning utilizados para a análise de sentimentos, como foi feito a tradução.

## 6. Dashboard

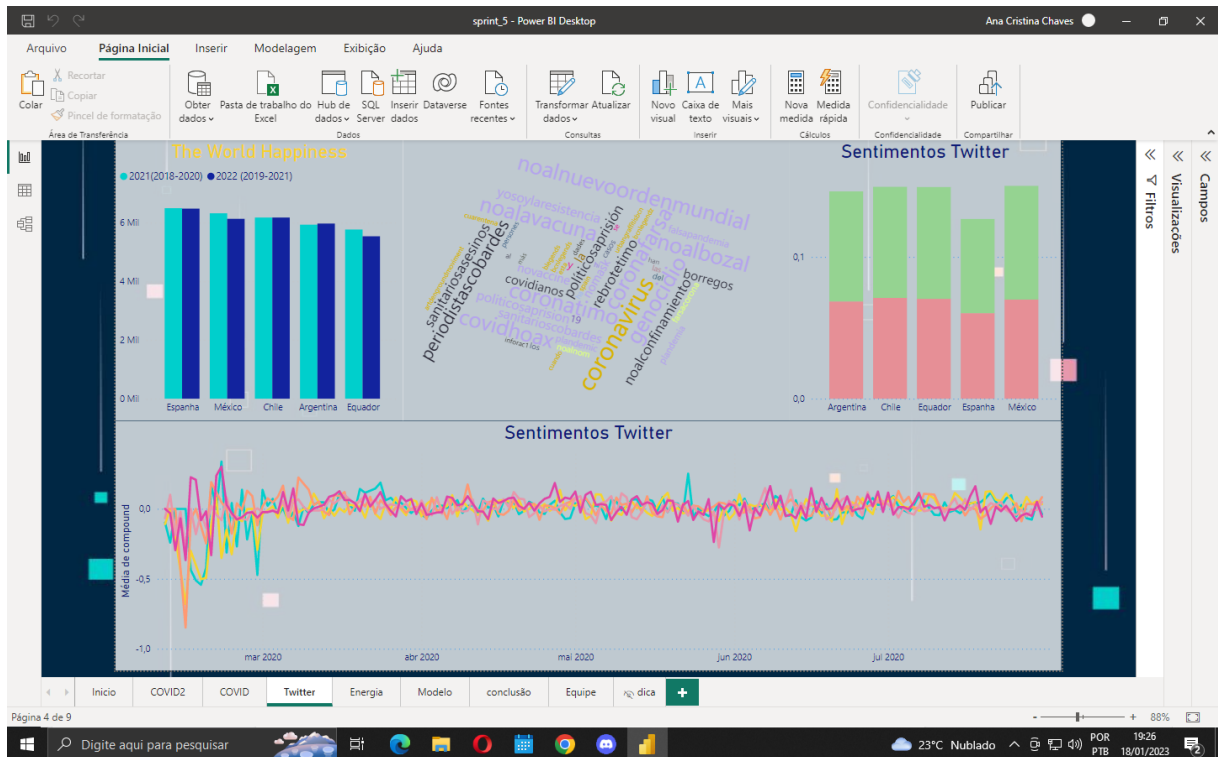
Mudamos a paleta de cores do dashboard, alteramos as cores solicitadas e reformulamos alguns gráficos.

Ajustamos a relação das planilhas e utilizamos como tabela de união, a tabela de países e de datas.

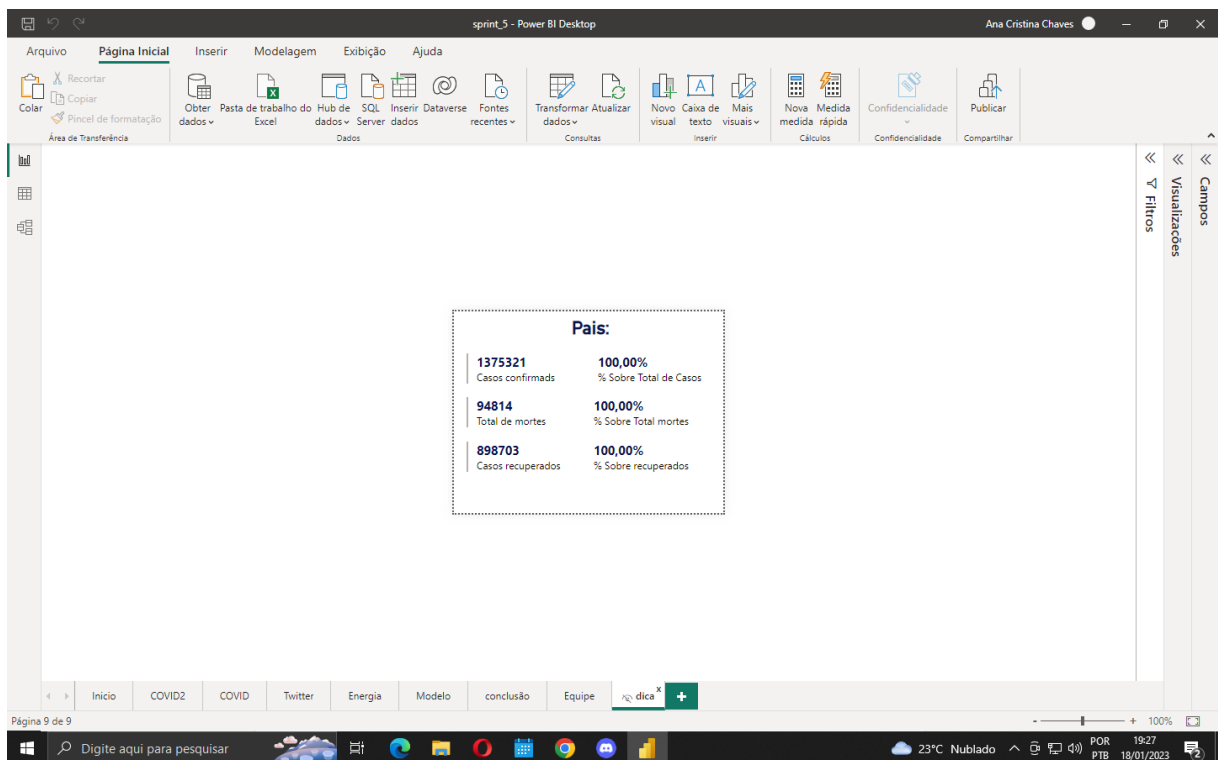
Paleta de cores:



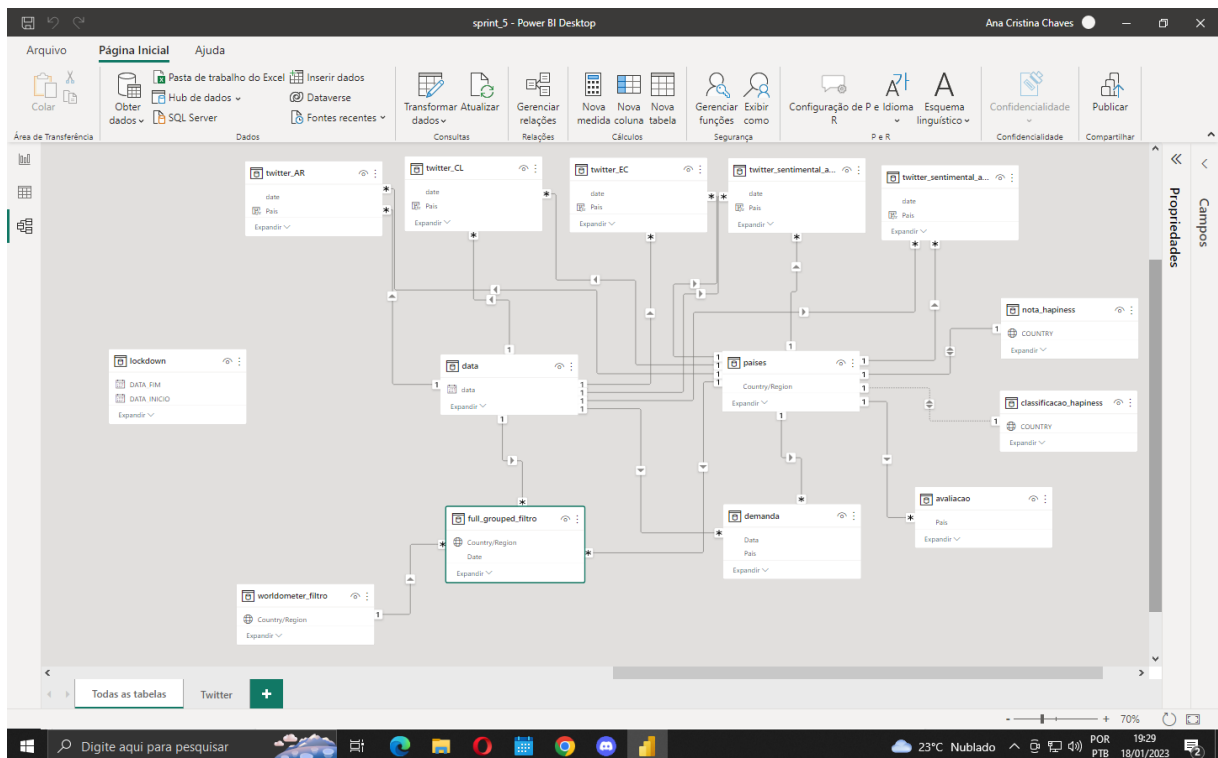
## Harmonização do gráfico nuvem de palavras:



## Ajuste das cores da dica de ferramenta, neste caso utilizei cores neutras:



Exluímos tabelas que não agregam valor ao dashboard, ajustamos os relacionamentos e conseguimos deixar todos na mesma página.



Estamos em desenvolvimento da apresentação e do dashboard.