

**São Paulo, 23 de Fevereiro de 2021**

|  |  |  |  |
| --- | --- | --- | --- |
| **CONTROLE DE VERSÕES** | | | |
| **Autor** | **Versão** | **Data** | **Descrição** |
| Greyce Silva | 0 | 23/02/2021 | Criação do documento. |
| Greyce Silva | 1 | 01/03/2021 | Atualização |

1. **Introdução**

Este documento tem o intuito de apresentar o projeto Integrado proposto pelo programa Blueshift Academy.

1. **Solicitação**

O projeto abrange os dados referentes a Emissões de Gases Poluentes e Causadores do efeito estufa, e tem o intuito de analisar quais países e/ou continentes mais emitem esses gases, em quais períodos, se os tratados como Kyoto e o acordo de Paris são de fatos efetivos e Possíveis relações entre Crescimentos Populacionais e PIB com estes países.

A base principal do projeto contém dados de 1850 a 2019, portanto todas as demais bases serão normalizadas para conter dados referente a este período de tempo, exceto por algumas visualizações genéricas.

O projeto será desenvolvido na Plataforma GCP, utilizando a linguagem Python para consumo de dados, tratativas e normatizações, assim como a inserção dos dataframes no bigquery, os Dashboards serão apresentados no Power Bi através da conexão com o Bigquery.

1. **Premissas da Solução**
   1. **Origem dos dados e especificações**

Os arquivos foram consumidos de diversas formas distintas que devem ser citadas:

* Consumo de arquivo .csv encontrado no site ou no github e realizado a leitura através do pandas;
* Consumo de arquivo .csv no Kaggle através da API e realizado a leitura através do pandas;
* Não foi encontrado nenhum arquivo .csv ou de qualquer outra natureza sobre os dados demandados, porém foi encontrado tabelas no próprio site da UNTC(United Nations Treaty Collections), foi realizado Web sCraping utilizando a biblioteca Beautifulsoup para consumo dos dados.

Para informações mais detalhadas verificar especificações abaixo:

* + 1. **Emissão de CO2 e CNH4 e outros gases por ano e Países (**Fonte: <https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>**)**,

Modo de consumo de dados: Pandas, Read and Load Functions

Notebook:

1. kyoto\_Emission\_country\_year.ipynb, 2. kyoto\_Paris\_Emission\_country\_year.ipynb

Dataframes: Emission\_Country\_Year

Arquivos: Emission\_Country\_Year.csv

* + 1. **Países que assinaram o tratado de kyoto (**Fonte: <https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-a&chapter=27&clang=_en>**)**,

Modo de consumo de dados: Web Scraping, Beautifulsoup;

Notebooks: 1. kyoto\_Emission\_country\_year.ipynb

Dataframes: Emission\_Country\_Year, Kyoto\_Agreement\_Countries

* + 1. **Países que assinaram o Acordo de Paris (**Fonte: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg\_no=XXVII-7-d&chapter=27&clang=\_en**)**,

Modo de consumo de dados: Web Scraping, Beautifulsoup;

Notebook: 2. kyoto\_Paris\_Emission\_country\_year.ipynb

Dataframes: Emission\_Country\_Year, Paris\_Agreement\_Countries

* + 1. **Países por continente e Região (**Fonte: Kaggle Datasets**)**, Modo de consumo de dados: Pandas, Read and Load Functions;

Notebook: 3. Country\_Region\_Continent.ipynb

Dataframes: Country\_Continent\_Region

* + 1. **Temperatura por País e Ano(**Fonte: Kaggle Datasets**)**, Modo de consumo de dados: Pandas, Read and Load Functions;

Notebook: 4. Temperatura.ipynb

Dataframes: MediaTempPais, TempPais, MediaTempGlobal, TempPais

* + 1. **Crescimento Populacional por país e ano (**Fonte:https://ourworldindata.org/world-population-growth**)**, Modo de consumo de dados: Pandas, Read and Load Functions;

Notebook: 5. GrowthxEmission.ipynb

Dataframes: MediaGrowthPais, GrowthPais

* + 1. **Crescimento PIB por país e ano**(Fonte: https://ourworldindata.org/economic-growth),

Modo de consumo de dados: Pandas, Read and Load Functions;

Notebook: 6. PiBxEmission.ipynb

Dataframes: PibPais, MediaPibPais

* 1. **Desenvolvimento**

O desenvolvimento se deu na plataforma GCP (Google Cloud Platform), utilizando os acessos que foram disponibilizados pela Blueshift. Segue abaixo todas as ferramentas utilizadas para o projeto:

**Cloud Schedule**: Agendamento de atividades;

**Cloud Pub/Sub**:

**Cloud Functions**:

**Compute Engine**: Criação da VM Debian 10;

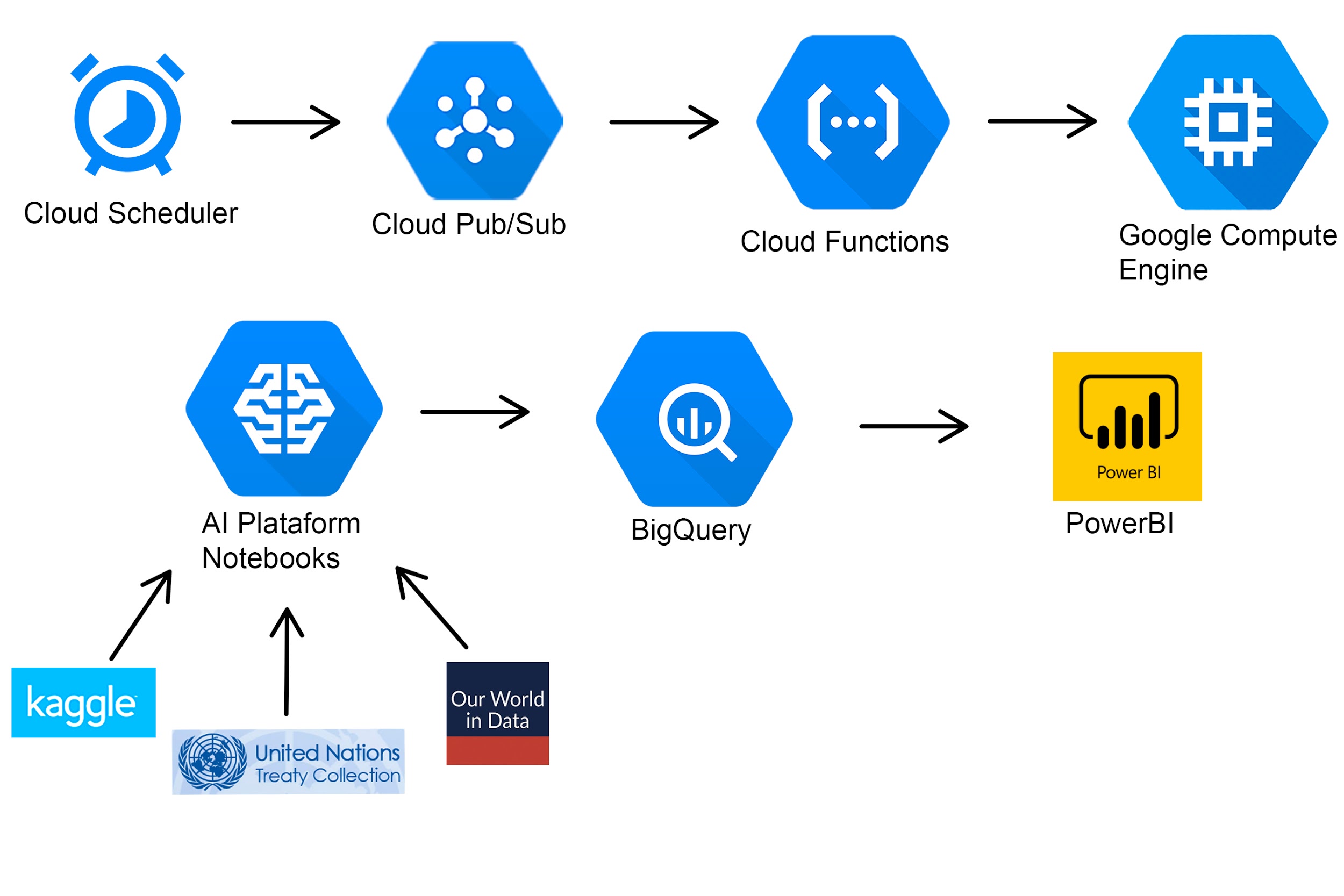
**AI Platform**: Utilização do Jupyter Notebook para Elaboração de códigos em Python;

**BigQuery**: Armazenamento de Dados;

**PowerBi**: Elaboração de Dashboards;

1. **Arquitetura**

Segue representação da Arquitetura do Projeto:



* 1. **Compute Engine:**
     1. **VM Debian 10**

|  |  |
| --- | --- |
| **Propriedades** | **Valor** |
| Nome | [projetointegrado](https://console.cloud.google.com/compute/instancesDetail/zones/us-west1-b/instances/projetointregado?project=blueshift-academy) |
| Série | N1 |
| Tipo de Máquina | n1-standard-4 (4 vCPUs, 15 GB de memória) |
| Disco de Utilização | Debian 10 |
| Tamanho do Disco | 100 GB |
| Identidade e Acesso a API | Permitir acesso completo a todas as APIs do Cloud |
| Firewall | Padrão |

* 1. **AI Platform:**

Foi utilizado o Jupyter Lab ambiente TensorFlow:2.3 para desenvolvimento dos códigos em Python, foram utilizados os seguintes notebooks, conforme abaixo:

1. kyoto\_Emission\_country\_year.ipynb;

2. kyoto\_Paris\_Emission\_country\_year.ipynb;

3. Country\_Region\_Continent.ipynb;

4. Temperatura.ipynb;

5. GrowthxEmission.ipynb;

6. PiBxEmission.ipynb;

Link dos Notebooks: https://github.com/matheusgois97/projetointregado/tree/main/Scripts

* 1. **Bigquery**

Foram armazenados os dados nas tabelas do bigquery, tanto os dados consumidos quanto dados após a tratativas, todas a tabelas com o final “\_in” são dados consumidos diretamente das fontes sem nenhum tratamento, segue abaixo esquemas de dados tratados que serão utilizados para a exibição do dashboard proposto.

* + 1. **Esquemas de Tabelas**
       1. **Tabela Emission\_Country\_Year**

|  |  |  |
| --- | --- | --- |
| **Nome do campo** | **Tipo** | **Modo** |
| iso\_code | STRING | NULLABLE |
| country | STRING | NULLABLE |
| year | INTEGER | NULLABLE |
| co2 | FLOAT | NULLABLE |
| co2\_per\_capita | FLOAT | NULLABLE |
| share\_global\_co2 | FLOAT | NULLABLE |
| cumulative\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_co2 | FLOAT | NULLABLE |
| co2\_per\_gdp | FLOAT | NULLABLE |
| cement\_co2 | FLOAT | NULLABLE |
| coal\_co2 | FLOAT | NULLABLE |
| flaring\_co2 | FLOAT | NULLABLE |
| gas\_co2 | FLOAT | NULLABLE |
| oil\_co2 | FLOAT | NULLABLE |
| other\_industry\_co2 | FLOAT | NULLABLE |
| cement\_co2\_per\_capita | FLOAT | NULLABLE |
| coal\_co2\_per\_capita | FLOAT | NULLABLE |
| flaring\_co2\_per\_capita | FLOAT | NULLABLE |
| gas\_co2\_per\_capita | FLOAT | NULLABLE |
| oil\_co2\_per\_capita | FLOAT | NULLABLE |
| other\_co2\_per\_capita | FLOAT | NULLABLE |
| share\_global\_coal\_co2 | FLOAT | NULLABLE |
| share\_global\_oil\_co2 | FLOAT | NULLABLE |
| share\_global\_gas\_co2 | FLOAT | NULLABLE |
| share\_global\_flaring\_co2 | FLOAT | NULLABLE |
| share\_global\_cement\_co2 | FLOAT | NULLABLE |
| cumulative\_coal\_co2 | FLOAT | NULLABLE |
| cumulative\_oil\_co2 | FLOAT | NULLABLE |
| cumulative\_gas\_co2 | FLOAT | NULLABLE |
| cumulative\_flaring\_co2 | FLOAT | NULLABLE |
| cumulative\_cement\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_coal\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_oil\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_gas\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_flaring\_co2 | FLOAT | NULLABLE |
| share\_global\_cumulative\_cement\_co2 | FLOAT | NULLABLE |
| total\_ghg | FLOAT | NULLABLE |
| ghg\_per\_capita | FLOAT | NULLABLE |
| methane | FLOAT | NULLABLE |
| methane\_per\_capita | FLOAT | NULLABLE |
| nitrous\_oxide | FLOAT | NULLABLE |
| nitrous\_oxide\_per\_capita | FLOAT | NULLABLE |
| population | FLOAT | NULLABLE |
| gdp | FLOAT | NULLABLE |
| Kyoto | BOOLEAN | NULLABLE |
| Paris | BOOLEAN | NULLABLE |

* + - 1. **Tabela Country\_Continent\_Region**

|  |  |  |
| --- | --- | --- |
| Nome do campo | Tipo | Modo |
| country | STRING | NULLABLE |
| code\_3 | STRING | NULLABLE |
| continent | STRING | NULLABLE |
| sub\_region | STRING | NULLABLE |
| country\_code | STRING | NULLABLE |

* + - 1. **Tabela Temperature\_Country\_Year (fAverageTemperatureByCountry)**

|  |  |  |
| --- | --- | --- |
| **Nome do campo** | **Tipo** | **Modo** |
| year | INTEGER | NULLABLE |
| iso\_code | STRING | NULLABLE |
| country | STRING | NULLABLE |
| AverageTemperature | FLOAT | NULLABLE |

* + - 1. **Tabela Growth\_Country\_Year (MediaGrowthPais)**

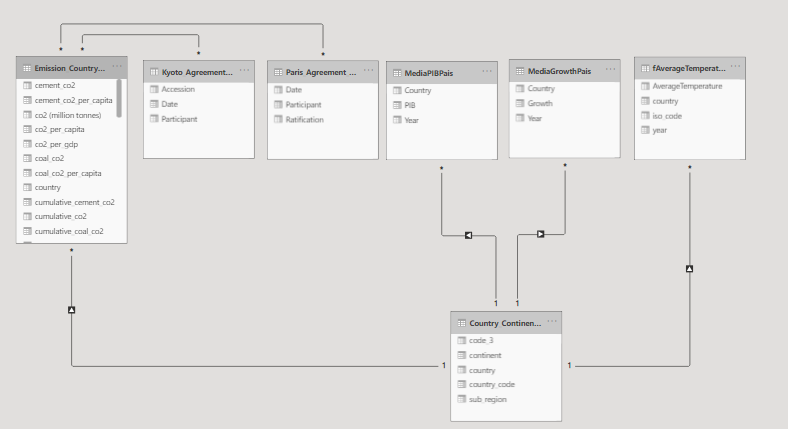
|  |  |  |
| --- | --- | --- |
| **Nome do campo** | **Tipo** | **Modo** |
| Year | INTEGER | NULLABLE |
| Country | STRING | NULLABLE |
| Growth | INTEGER | NULLABLE |

* + - 1. **Tabela PIB\_Country\_Year (MediaPIBPais)**

|  |  |  |
| --- | --- | --- |
| Nome do campo | Tipo | Modo |
| Year | INTEGER | NULLABLE |
| Country | STRING | NULLABLE |
| PIB | INTEGER | NULLABLE |

* 1. **Power BI**

Segue relacionamentos entre todas as tabelas no power BI:

****

* 1. **Schedule**
  2. **Pub/Sub**
  3. **Functions**

1. **DashBoards**

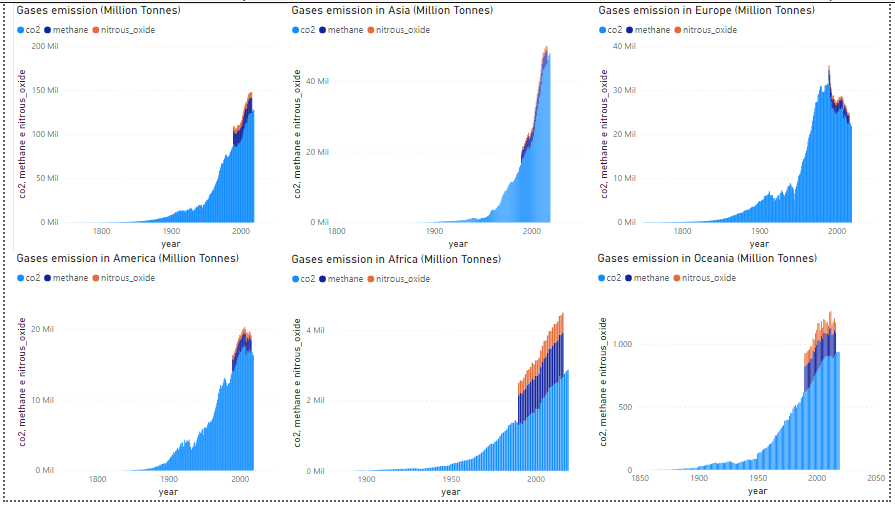
* 1. **Visualizações:**

As visualizações apresentadas no item relatórios:

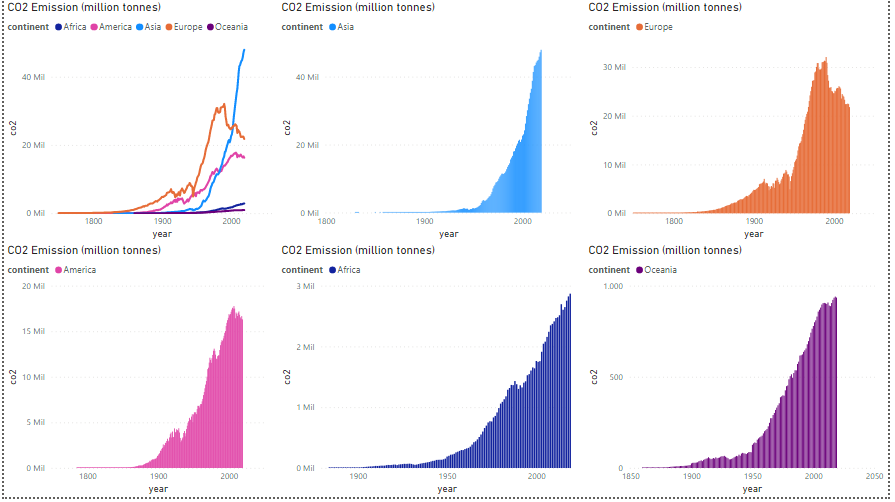
* Emissão de gases por País e Ano, segmentação por Região;
* verificação de impactos dos tratados de Kyoto e Acordo de Paris;
* Relação entre Emissões, PIB e Crescimento Populacional;
* Relação entre Emissão e Temperatura do Planeta;
* Emissão de gases por indústria e por País e ano.
  1. **Relatórios:**

É importante ter em mente que todos os valores na tabela Emission\_Country\_Year estão em milhões de Toneladas, ou seja, quando nos gráficos aparecer Mi o resultado real seria Bilhões.

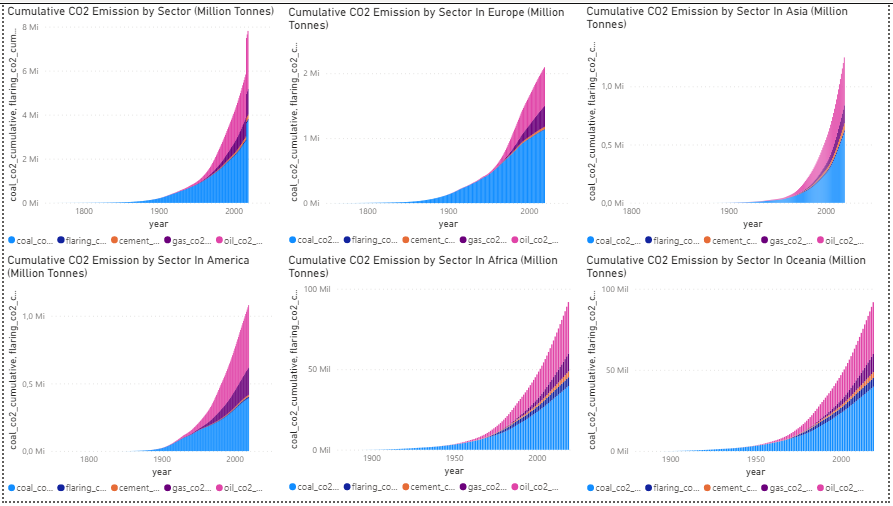
* + 1. **Emissão de CO2, Metano e Óxido Nitroso por ano e continente:**

****

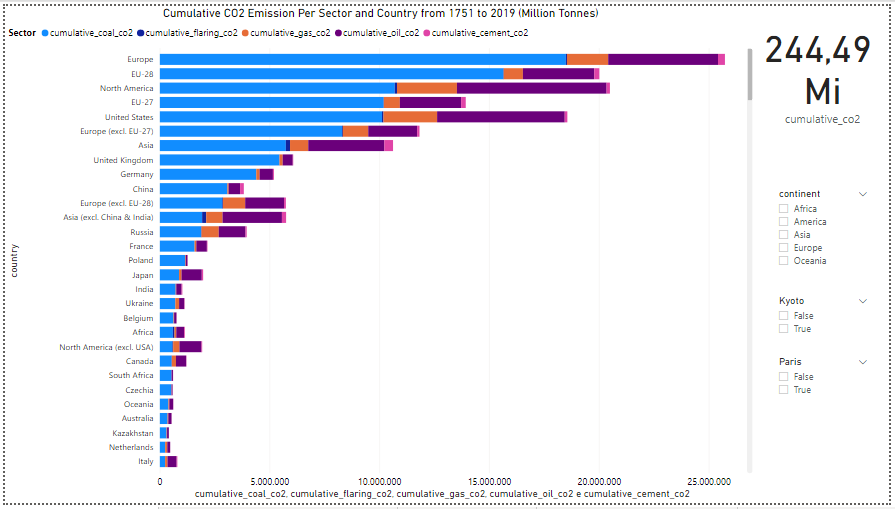
* + 1. **Emissão de CO2 por ano e continente:**

****

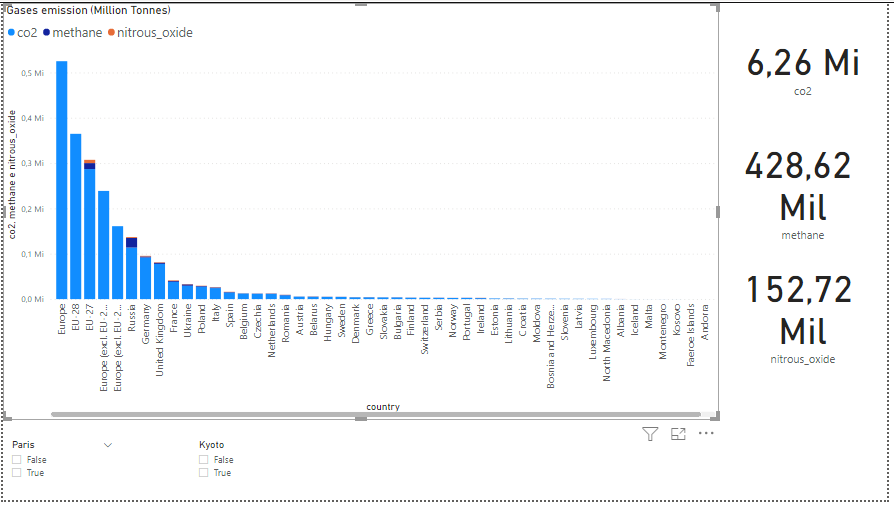
* + 1. **Emissão acumulada de CO2 por ano e Continente, para os setores conforme abaixo:**

****

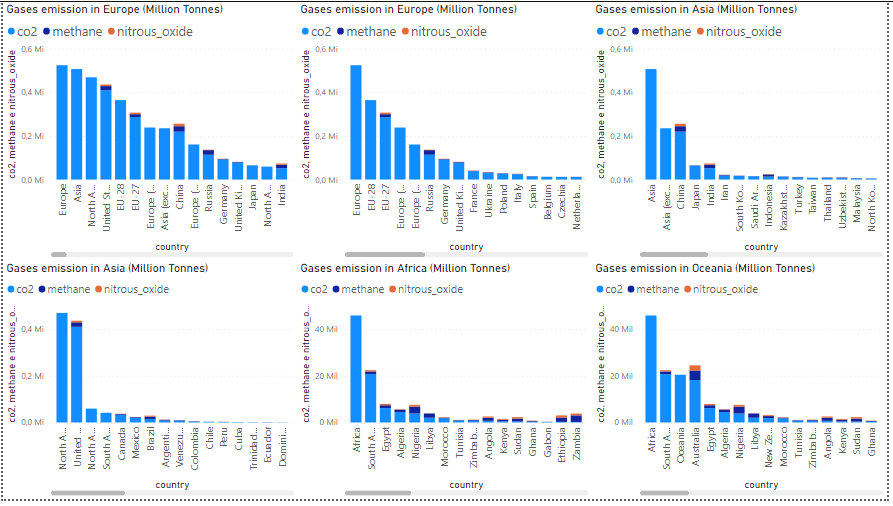
* + 1. **Emissão acumulada de CO2 por Setor e País/Região, Análise dos Protocolos de Kyoto e Acordo de Paris:**

****

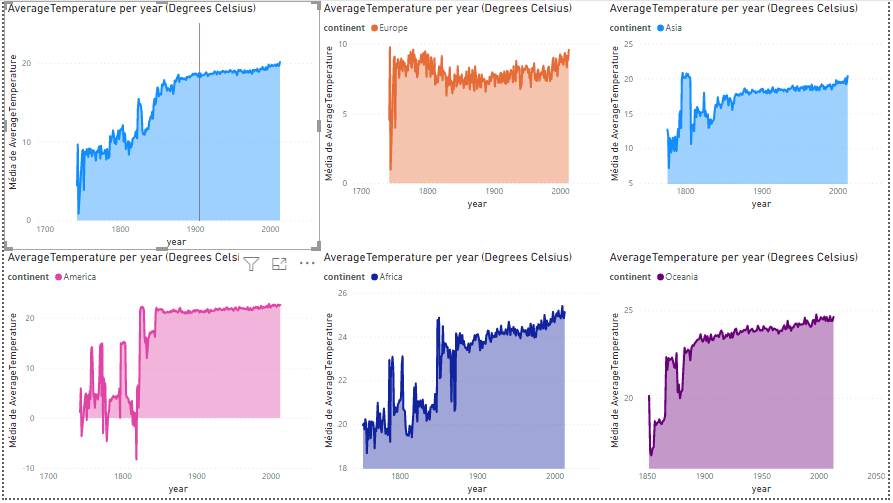
* + 1. **Emissão de CO2, Metano e Óxido Nitroso por ano e País/Região, Análise dos Protocolos de Kyoto e Acordo de Paris:**

****

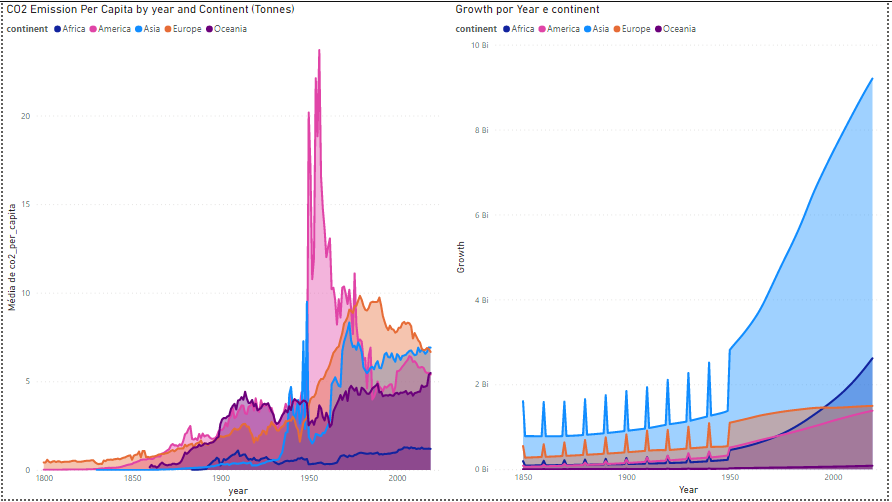
* + 1. **Emissão de CO2, Metano e Óxido Nitroso por País/Região:**

****

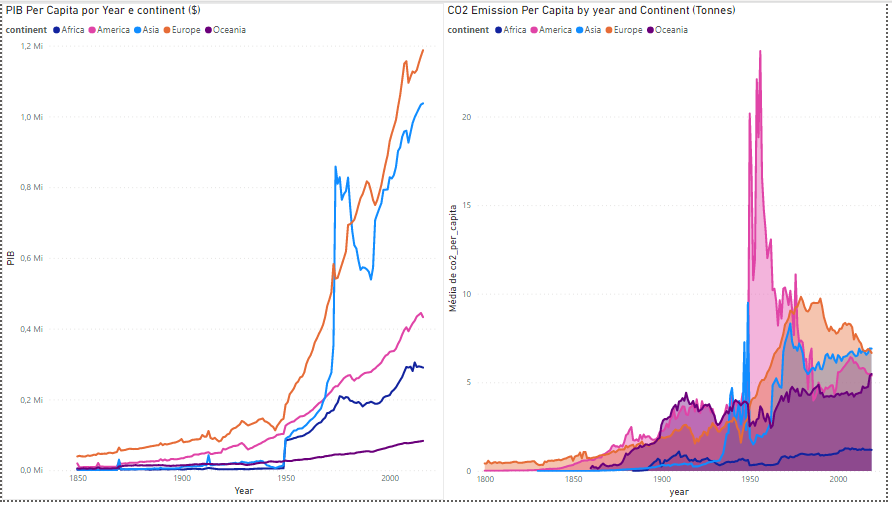
* + 1. **Variação de temperatura da terra.**

****

* + 1. **Emissão de CO2 per capita e comparação com crescimento populacional por Continente:**

****

* + 1. **Emissão de CO2 per capita e comparação com o PIB per capita:**

****

1. **Considerações finais**
   1. **Lições Aprendidas**

O projeto proposto inicialmente integrava na arquitetura o uso do Google Cloud Storage para o armazenamento dos arquivos tratados no Jupyter, porém foi concluído que não há a necessidade deste passo, uma vez que através da biblioteca pandas gbq e o pacote google.oauth2 é possível armazenar os data frames diretamente no Bigquery, sem a necessidade de códigos no bigquery.

* 1. **Propostas de Implementação Futura**

Automatização dos notebooks utilizando o Cloud Functions.

1. **Bibliografia**

<https://cloud.google.com/compute/docs/instances/transfer-files?hl=pt-br>

<https://thihenos.medium.com/compute-engine-upload-storage-9a5c0a8c5d2>

<https://googleapis.dev/python/storage/latest/client.html>

<https://data.giss.nasa.gov/gistemp/graphs/graph_data/Global_Mean_Estimates_based_on_Land_and_Ocean_Data/graph.txt>

<https://stackoverflow.com/questions/21546739/load-data-from-txt-with-pandas>

<https://cloud.google.com/ai-platform/notebooks/docs/create-new?hl=pt#:~:text=Acesse%20a%20p%C3%A1gina%20AI%20Platform%20Notebooks%20no%20Console%20do%20Google%20Cloud.,-Acessar%20a%20p%C3%A1gina&text=Clique%20em%20add_boxNova%20inst%C3%A2ncia%20e%2C%20em%20seguida%2C%20selecione%20Personalizar,nome%20para%20a%20nova%20inst%C3%A2ncia>.

<https://3a9700bde2b600a2-dot-us-west1.notebooks.googleusercontent.com/lab?authuser=0>

<https://googleapis.dev/python/storage/latest/blobs.html#google.cloud.storage.blob.Blob.download_as_bytes>

<https://colab.research.google.com/github/corrieann/kaggle/blob/master/kaggle_api_in_colab.ipynb#scrollTo=lLRgc90dojon>

<https://stackoverflow.com/questions/56721927/how-to-load-data-to-jupyter-notebook-vm-from-google-cloud>

<https://colab.research.google.com/github/corrieann/kaggle/blob/master/kaggle_api_in_colab.ipynb>

<https://medium.com/@charles2588/how-to-upload-download-files-to-from-notebook-in-my-local-machine-6a4e65a15767>

<https://towardsdatascience.com/running-jupyter-notebook-in-google-cloud-platform-in-15-min-61e16da34d52>

<https://stackoverflow.com/questions/61888615/how-to-connect-to-google-cloud-platform-data-storage-from-google-cloud-ai-jupyte>

<https://medium.com/data-hackers/como-fazer-web-scraping-em-python-23c9d465a37f>

<https://anderfernandez.com/en/blog/automate-python-script-google-cloud/>