Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente



MVP 3 – SPRINT ENGENHARIA DE DADOS

Aluno: Lucas Vital Alves da Silva

Pós-graduação em ciência de dados e Analytics

Professores: Vitor Almeida e Silvio Alonso

**SUMÁRIO**

[Introdução 3](#_Toc146818173)

[Busca e Coleta dos dados 4](#_Toc146818174)

[Modelagem dos dados 5](#_Toc146818175)

[Catálogo de dados 6](#_Toc146818176)

[Processamento de dados 7](#_Toc146818177)

[ocde\_carbon\_emission\_per\_country\_ocde 9](#_Toc146818178)

[ocde\_land\_cover 10](#_Toc146818179)

[mb\_avg\_preciptation\_per\_year 12](#_Toc146818180)

[emdat\_disasters\_impact\_uidds2nac 13](#_Toc146818181)

[gwis\_burned\_area\_full\_dataset\_2002\_2019 15](#_Toc146818182)

[Tabelas temperaturas locais 16](#_Toc146818183)

[Tabelas temperaturas globais 18](#_Toc146818184)

[Análise dos dados 19](#_Toc146818185)

[A evolução das queimadas ao longo do tempo e o perfil desse fenômeno nos países mais afetados 19](#_Toc146818186)

[Análise Localidades mais impactadas 21](#_Toc146818187)

[Impactos econômicas e perdas Humanas decorrentes das queimadas 23](#_Toc146818188)

[Análise das Características das localidades mais afetadas 25](#_Toc146818189)

[Características de evolução temperatura anual média ao logo do tempo 26](#_Toc146818190)

[Influência do aumento da temperatura global na temperatura local 27](#_Toc146818191)

[Como as emissões de carbono influenciam na temperatura global 28](#_Toc146818192)

[Análise regime de Chuvas Local 29](#_Toc146818193)

[Análise de característica de vegetação local 30](#_Toc146818194)

[Conclusões 31](#_Toc146818195)

[Bibliografia 32](#_Toc146818196)

# Introdução

Na última década os fenômenos de incêndios florestais, mas comumente mencionadas como queimadas aqui no brasil, tem chamado cada vez mais atenção mundialmente. As queimadas estão cada vez mais frequentes e intensas e tem causada devastações e perdas que tem quebrado recordes sobre tudo nos últimos anos. Tendo em vista as catástrofes ocorridas no Havaí e no Canadá esse ano tal fenômeno chamou minha atenção de modo a sentir-me engajado em tentar entender por meio dos dados o porquê tal fenômeno anda tão frequente e intenso e as verdadeira magnitudes de seu impacto na vida das pessoas que vivem nas áreas afetadas. Portanto, a ideia nesse trabalho é obter os dados necessários para responder as seguintes indagações:

* Sobre evolução das queimadas na última década:
* Como tem sido a evolução das queimadas nessa última década ?
* Como tem sido a evolução em intensidade dos incêndios nessa última década?
* Análise do perfil das localidades onde as queimadas ocorrem com maior frequência e intensidade:
* Onde tem acontecido os incêndios com mais intensidade ?
* Qual a característica desses locais em termos de localização espacial, temperatura, umidade, vegetação?
* Sobre Perdas em termos de vegetação e impactado na vida das pessoas:
* Perdas médias globais de vegetação
* Perdas média de vegetação por continente
* Perda média de vegetação nos países afetados
* Perdas econômicas ( média global, continental, e por país) devido as queimadas
* Impacto na qualidade do ar
* Sobre a relação entre as mudanças climáticas e o aumento das queimadas:
* Emissões de carbono
* Aumento da temperatura
* Relação entre aumento da temperatura global e local por país

Sendo assim o objetivo deste trabalho é responder as perguntas acima de modo que o compilado de resposta forme uma resposta consistente sobre a pergunta principal:

***Porque as queimadas tem aumentado tanto em frequência e intensidade na última década***

De modo a compreender todos os fatores a ela elencados. Para se chegar a esse compilado de respostas a ideia é usar os conhecimentos aprendidos essa Sprint realizando a coleta dos dados, a sua modelagem de forma adequada e construção de um catálogo de dados, a carga desses dados numa infraestrutura de nuvem e construção de um pipeline de tratamento, a análise dos dados para finalmente se obter a resposta e com isso se tirar uma conclusão final acerca da principal resposta buscada.

# Busca e Coleta dos dados

Buscando responde cada uma das perguntas acima, buscou-se dados referentes a cada um do aspecto em que se entendeu como parte da pergunta principal. Sendo assim dos dados que foram encontrados, os seguintes abaixo foram entendidos como relevates para as respostas buscadas:

* **Evolução das queimadas na última década:**

- [Monthly burned area [ha] by landcover class for years 2002-2019 - GWIS](https://gwis.jrc.ec.europa.eu/apps/country.profile/downloads)

* **Característica de vegetação por localização**

- [LANDCOVER DATASET - OCDE](https://stats.oecd.org/Index.aspx?DataSetCode=LAND_COVER)

* **Média temperatura por país**

- [Temperature all Countries](https://www.kaggle.com/datasets/subhamjain/temperature-of-all-countries-19952020/code)

- [Mean global temperature - Berkeley University](https://berkeleyearth.org/data/)

- [Mean Temperature Cites on Canada - climate data](https://climatedata.ca/download/#station-download)

- [Average monthly temperature by us state - kaggle](https://www.kaggle.com/datasets/justinrwong/average-monthly-temperature-by-us-state)

- [Hawaii monthly average temperatures Kahului - nceii noa](https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/city/time-series/USW00022516)

* **Média anual de precipitação**

[Precipitation per country - World Bank](https://data.worldbank.org/indicator/AG.LND.PRCP.MM?end=2020&start=1961)

* **Perdas em aspecto econômicos e mortes devido as queimadas**

[Global economic Losses and deaths due to wildfires - World Bank](https://data.worldbank.org/indicator/AG.LND.PRCP.MM?end=2020&start=1961)

[Disaster Database - EM Data](https://public.emdat.be/data)

* **Emissões Anuais de carbono por país**

[CO2 Emissions - Ourworld Data](https://ourworldindata.org/co2-dataset-sources)

[Greenhouse gas emissions - OCDE](https://stats.oecd.org/Index.aspx?DataSetCode=air_ghg)

[CO2 Emissions - World Bank](https://data.worldbank.org/indicator/EN.ATM.CO2E.PC)

# Modelagem dos dados

Tendo em vista que o que se deseja obter como fonte para camada de visualização é um Data Warehouse, fez-se então uma análise mais aprofundada da dos dados coletados selecionando os dados que de fato eram relevantes para a análise a ser realizada. Muitos dos dados encontrados eram redundantes, porém algumas tabelas apresentavam-se mais completas do que outras e foram justamente essas tabelas as quais foi dada a prioridade na seleção. Uma vez selecionadas as tabelas, quebrou-se essas tabelas em uma tabela fato e as tabelas dimensões correspondentes. Tal divisão fora feita considerando as principais métricas que deveriam ser obtidas as quais tinham que estar na tabela fato, já as tabelas dimensões foram definidas com campos que definiam algum aspecto de granularidade relativa as métricas principais que deveriam ser obtidas a partir da tabela fato. Muitas tabelas fatos acabaram por compartilha as mesmas dimensões então a mesma pode ser aproveitada para outra análise de métrica ao qual se desejou realizar. A partir desse processo obteve-se um modelo de DW abaixo, onde ao todo se obteve um total de 6 tabelas fato.

  
  
  
O esquema relacional pode melhor avaliado acessando-se o link abaixo:

[WildFire Analysis Data Warehouse Schema](https://drive.google.com/file/d/1L1eHUI49MDDmFsBoyL3AfaJzykHaSKl0/view)

# Catálogo de dados

Após a etapa de coleta dos dados chegamos na etapa de sua catalogação. Nessa etapa entendeu-se mais a fundo os dados que possuíamos a disposição, dado que se teve fazer uma investigação mais profunda para fazer sua catalogação. Sendo assim a partir do entendimento mais a fundo dos dados cada tabela obtida e investigações realizadas, definiu-se as restrições relativas a cada campo de cada tabela, seus o tipo do dado relativo a este campo e fez-se uma descrição do que significava esse campo. Em alguns casos( como no banco de dados [Disaster Database - EM Data](https://public.emdat.be/data)), pode-se contar com um catálogo próprio da base de dados para melhorar ainda mais a descrição dos dados. Em outros como o caso das médias das temperaturas locais, percebeu-se que era necessário obter mais dados, pois localidades de interesse( EUA, Havaí e Canadá) não possuíam dados registrados no banco de dados preliminarmente obtido. Sendo assim a catalogação não só permitiu um entendimento dos dados que possuíamos a disposição como também a verificação se de fato esses dados cobriam as linhas de análise desse trabalho. Assim sendo, foi-se gerado os catálogos relativos a cada uma tabela dos quais podem ser acessados pelo link abaixo:

[Catalogo de Dados - Github](https://github.com/lucasvittal2/wildfire_analysis_data_engineer_project/tree/main/Assets/Catalogs/tables/csv)

Como se pôde ver acima, temos definidas o nome das tabelas origem, as quais nesse trabalho estão simulando uma camada de “Staging” onde os dados estão em sua forma original. Os primeiros caracteres antes do hífen do nome de cada tabela origem fazem referência à sua fonte origem. Sendo assim segue abaixo a legenda abaixo:

|  |  |
| --- | --- |
| ocde | Tabelas obtida a partir do site da OCDE |
| mb | Tabelas a partir do site do banco mundial |
| Ncei | Tabelas obtidas a partir do site da ncei.nooa.gov |
| kaggle | Tabelas obtidas a partir do site do Kaggle |
| cd | Tabelas obtidas a partir do site climate data |
| emdata | Tabelas obtidas a partir do site do emdata |
| gwis | Tabelas obtidas a partir do site do gwis |

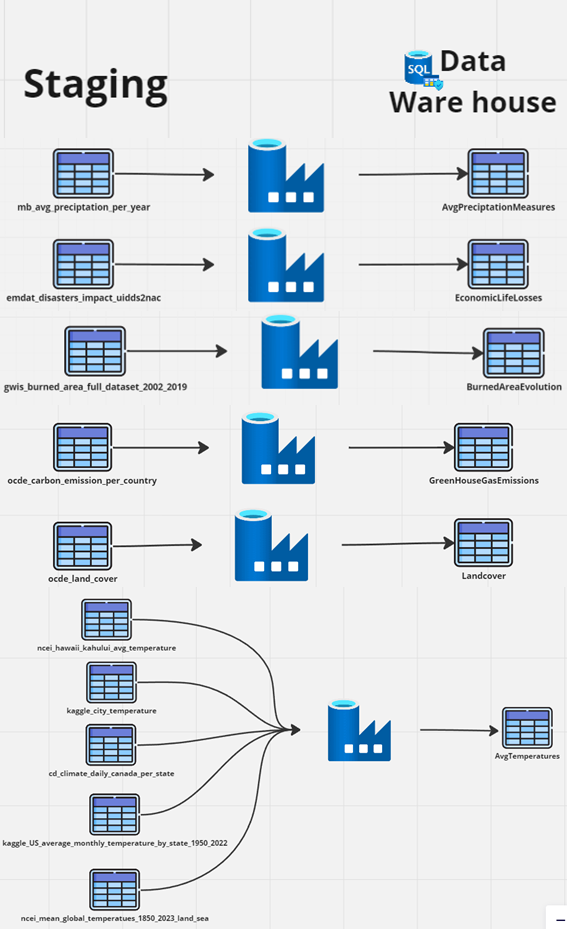
# Processamento de dados

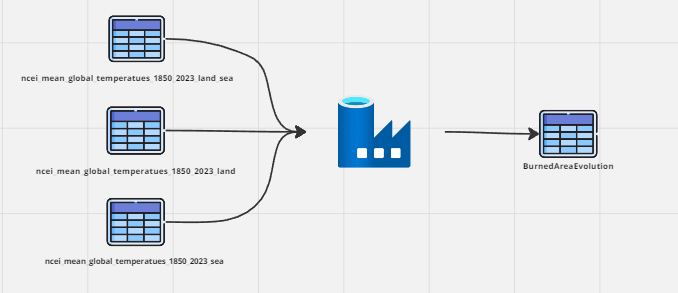
Uma vez que se realizou a catalogação e a devida documentação, se obteve um bom entendimento acerca dos dados em que se possuía em mãos. Com isso chegamos na etapa de processamento de dados. Contou-se com umas vastas quantidades de dados de diferentes fontes, sendo assim, por uma questão de organização, a primeira, coisa a ser realizada foi um mapeamento de que transformações seriam necessárias em cada campo de cada uma das tabelas obtidas.

Para a realização das transformações, inicialmente cogitou-se utilizar a plataforma da AWS, realizando os processos ETL por meio do “AWS Glue” que por sua vez carregaria os dados no “Amazon Redshift” onde se fariam as análises dos dados transformados. Isso fora feito para a primeira tabela, contudo a cota grátis na AWS demonstrou-se insuficiente considerando as cobranças pelos serviços utilizados nesse projeto, fazendo com que os custos alcançassem valores consideráveis para se testar um único pipeline. Sendo assim percebeu-se que a utilização da AWS como plataforma seria inviável devido aos altos custo financeiros que iriam implicar para a realização deste projeto.

Por conta disso decidiu-se aborta o usa da plataforma AWS e optou-se pela plataforma nuvem da Azure da Microsoft que por sua vez tem uma quantidade gratuita de créditos que atenderam muito bem para os fins desse projeto. Com isso terminamos por utiliza o serviço do “Azure Data Factories” para realização das transformações dos dados e o “Azure SQL Database” para tanto criar as tabelas do Data Warehouse projetado anteriormente na modelagem quanto realizar a análise dos dados em busca das respostas para as perguntas inicialmente elencadas nesse projeto.

Cada uma das tabelas teve seus dados tratados de acordo com o mapa de transformações onde a partir das quais foram criadas novas tabelas de acordo com a modelagem de dados anteriormente feita. A ilustração de cada tabela origem e tabelas geradas no Data Warehouse aqui projetado pode ser visto na página seguinte.



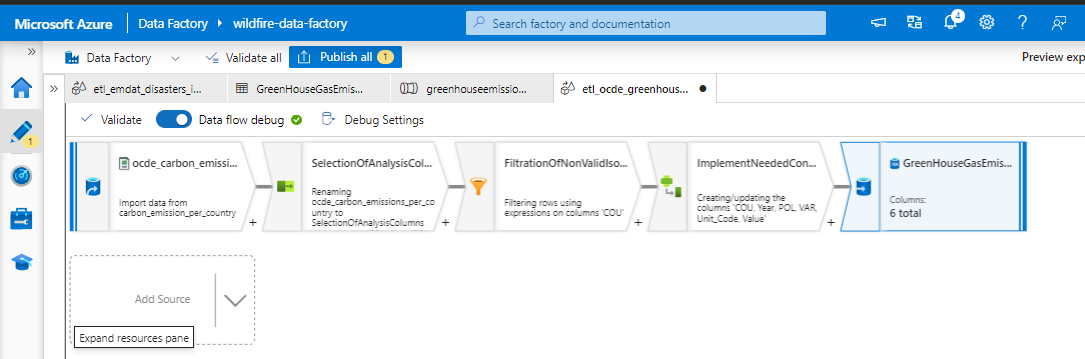


A seguir detalharemos com mais detalhes como foi realizado as devidas transformações em cada tabela

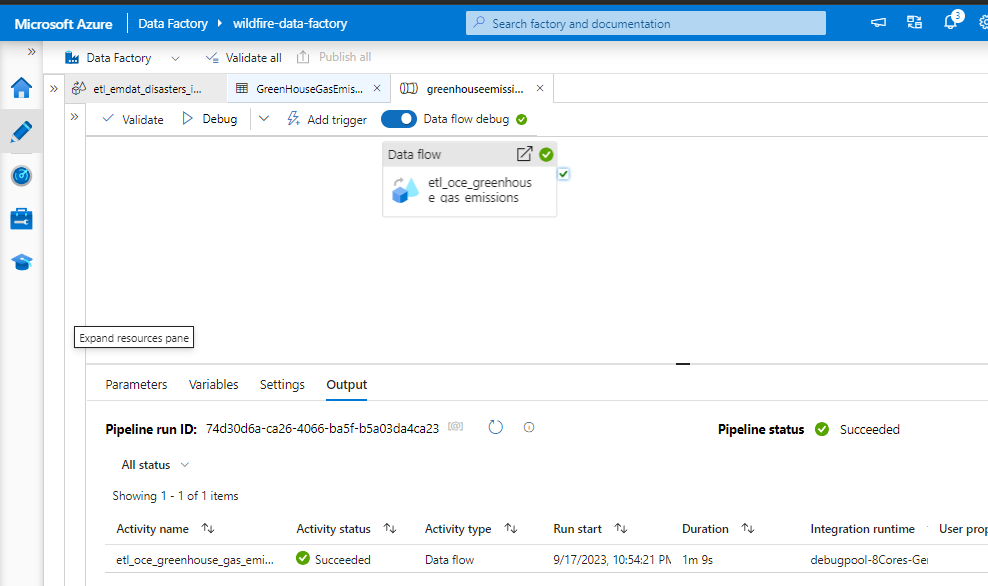
## ocde\_carbon\_emission\_per\_country\_ocde

Essa tabela foi designada para avaliar a relação entre as emissões de gases de efeito estufa e as queimadas. Sendo assim para essa tabela selecionou-se as colunas “COU”, “Country”, “POL”, “Pollutant”, “Year”, ”Unit”, ”Value”. Como não estamos interessados em analisar especificamente um gás de efeito estufa e sim as emissões como um todo, estão usou-se o campo “POL” filtrando os valores que não são “GHG”, afim de se pegar emissões relativas a todos os gases. O campo “Unit” também foi utilizado como filtro, filtrando valores diferentes de “Tonnes of CO2”, dado que queremos avaliar as emissões em valores absolutos e não relativos. Por fim criou-se a chave primária relativa a essa tabela concatenando o campo “Year” com uma string ‘01/01/’ e convertendo a concatenação para date e os valores das emissões para “float”.

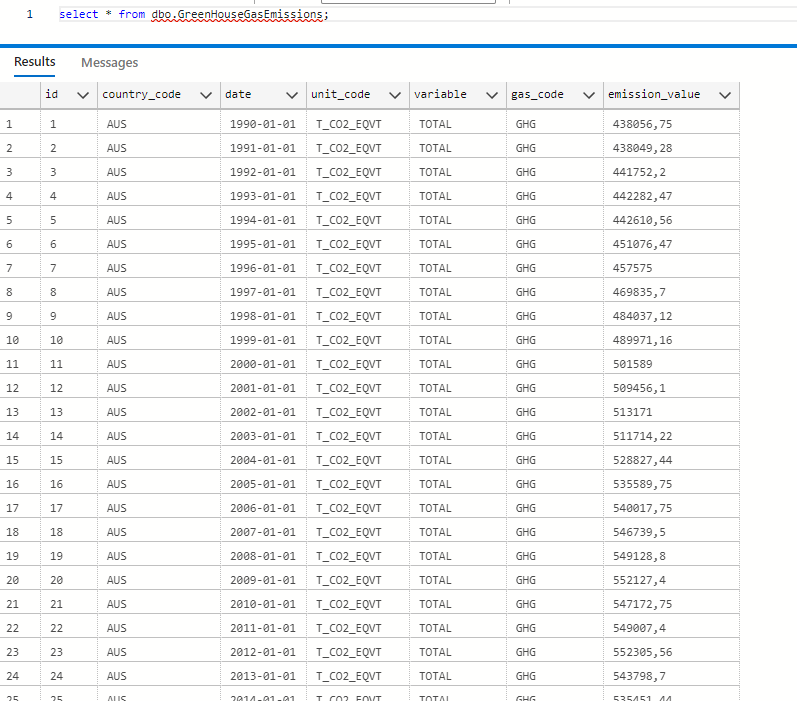
A seguir o “dataflow” utilizado no Azure Data Factories para realizar cada uma das transformações acima para gerar a tabela “*GreenHouseGasEmissions*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



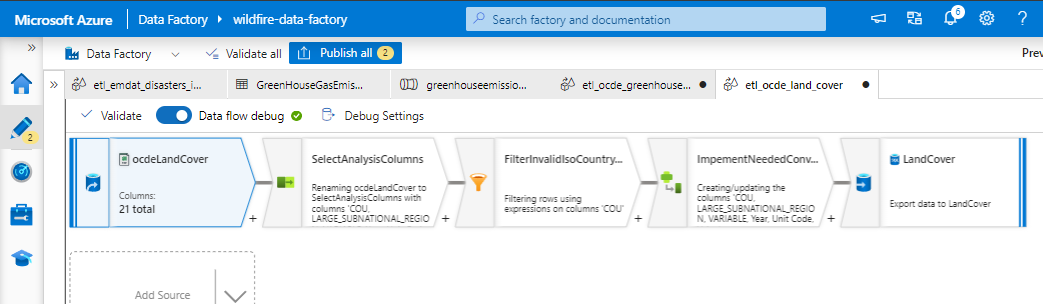
E os dados carregados no Azure Database podem ser vistos na imagem abaixo:



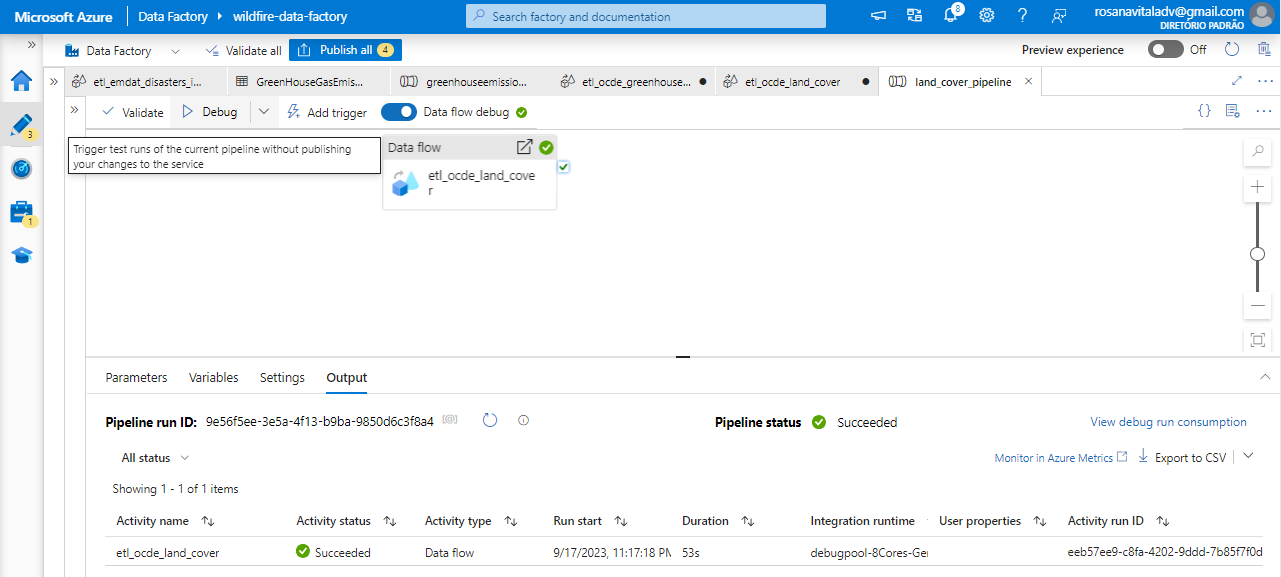
## ocde\_land\_cover

Essa tabela foi criada para avaliar as características de vegetação de cada país, especialmente os mais afetados pelas queimadas para se entender se existe algum tipo de característica de vegetação que propicie incêndios de grandes magnitudes. Sendo assim se selecionou os seguintes campos para essa tabela “COU“, “Country”, “Large subnational region”, “Variable”, “Land Cover Classe“, “Year“, “Unit code“, “Unit“, “Value“. Um único filtro fora implementado no campo “COU”, somente para garantir que haja valores válidos de código ISSO para cada país.

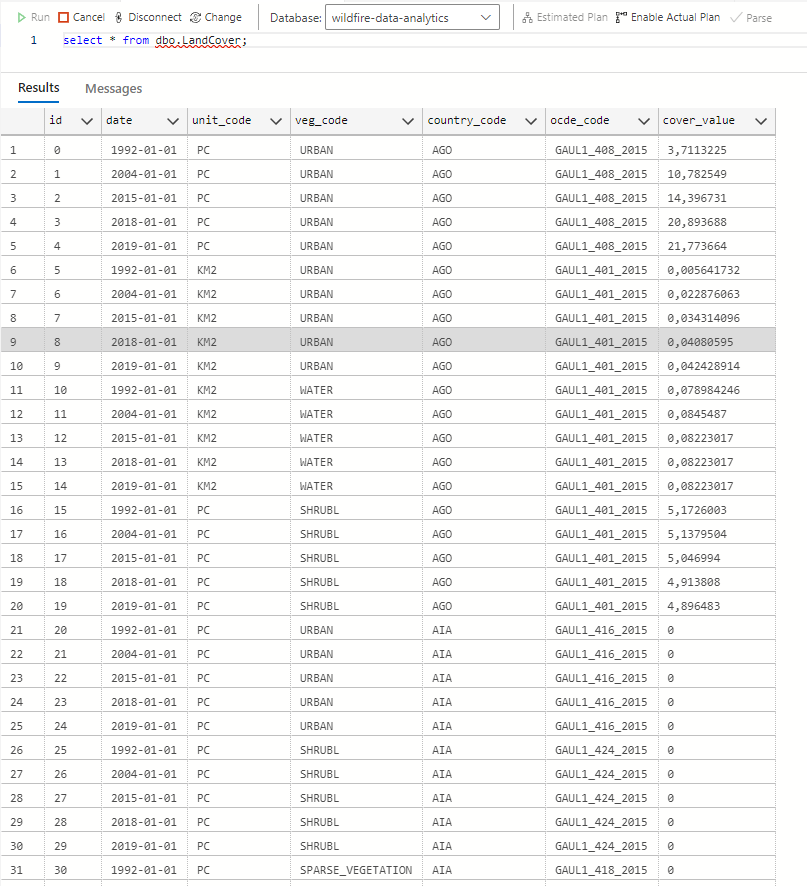
A seguir o “dataflow” utilizado no Azure Data Factories para realizar cada uma das transformações acima para gerar a tabela “*Landcover*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



E os dados carregados no Azure Database podem ser vistos na imagem abaixo:

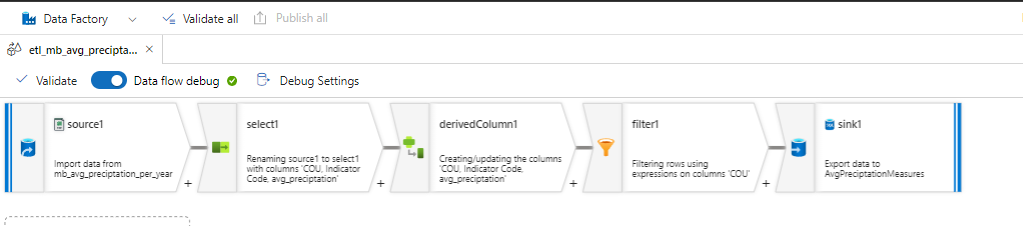


## mb\_avg\_preciptation\_per\_year

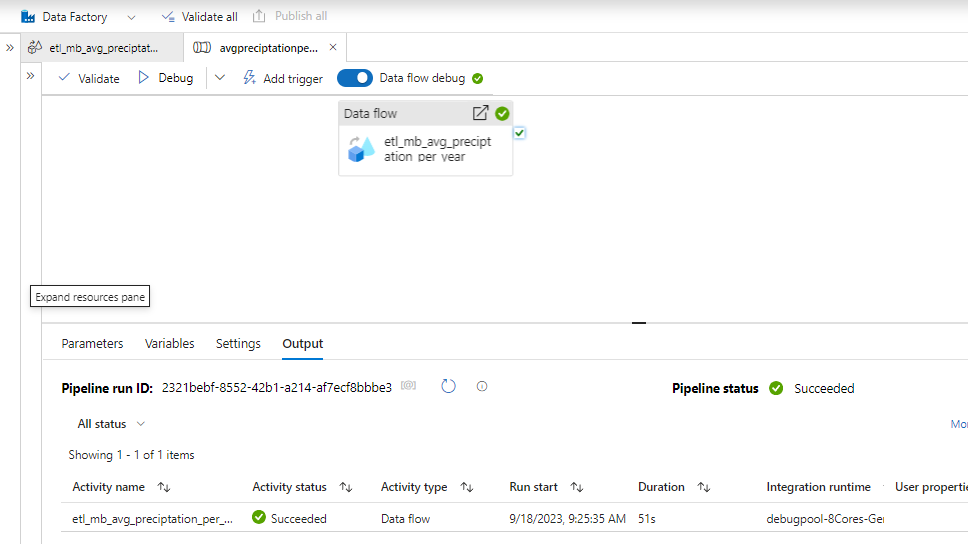
Essa tabela foi criada com o intuito de avaliar as características de precipitação das localidades mais afetadas pelas queimadas, afim de tentar-se descobrir alguma relação entre níveis de chuva e o fenômeno das queimadas. Sendo assim selecionou-se

As colunas “COU”, “Indicator code” e “2010”. A última coluna em especial representa o numérico médio de precipitação no ano de 2010 e foi escolhida especificamente porque percebeu que todos os registros de precipitação não nulos de 1960 a 2022 eram exatamente os mesmo para os demais países e pelo fato de todos os países que possuem registros nessa faixa de tempo tem um registro em 2010. Portanto, considerou-se que a média é exatamente o mesmo valor dos registros individuais não nulos, o que simplificadamente se resume a tomar o calor da coluna “2010”. Contudo, é válido reconhecer que seriam precisos dados mais precisos construir essa tabela. Dada a dificuldade e o limite de tempo isso vai ficar para uma interação posterior.

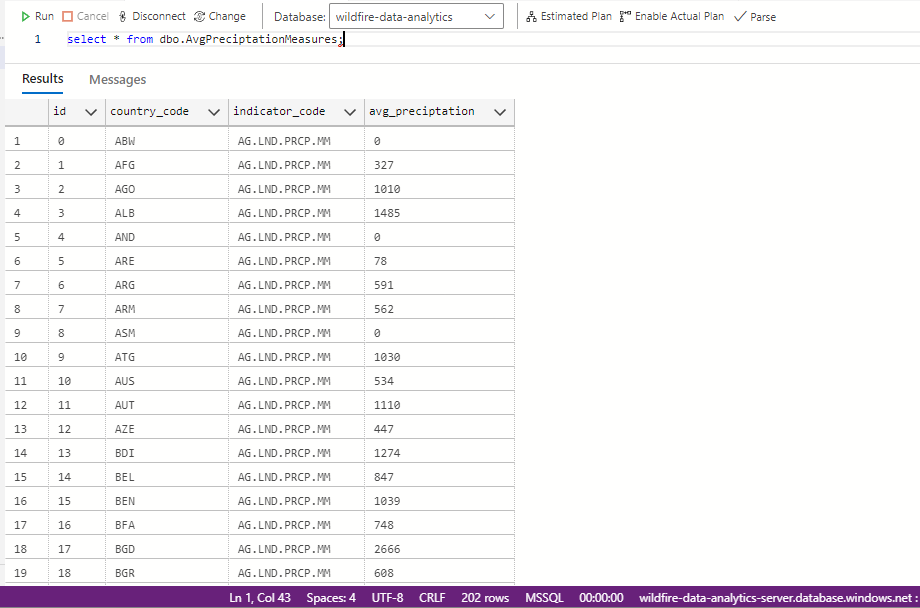
A seguir o “dataflow” utilizado no Azure Data Factories para realizar cada uma das transformações acima para gerar a tabela “*AvgPreciptationMeasures*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



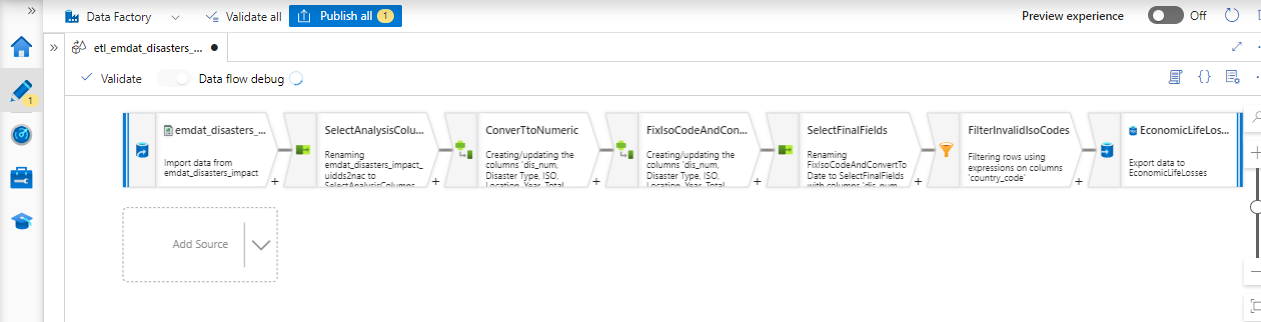
E os dados carregados no Azure Database podem ser vistos na imagem abaixo:



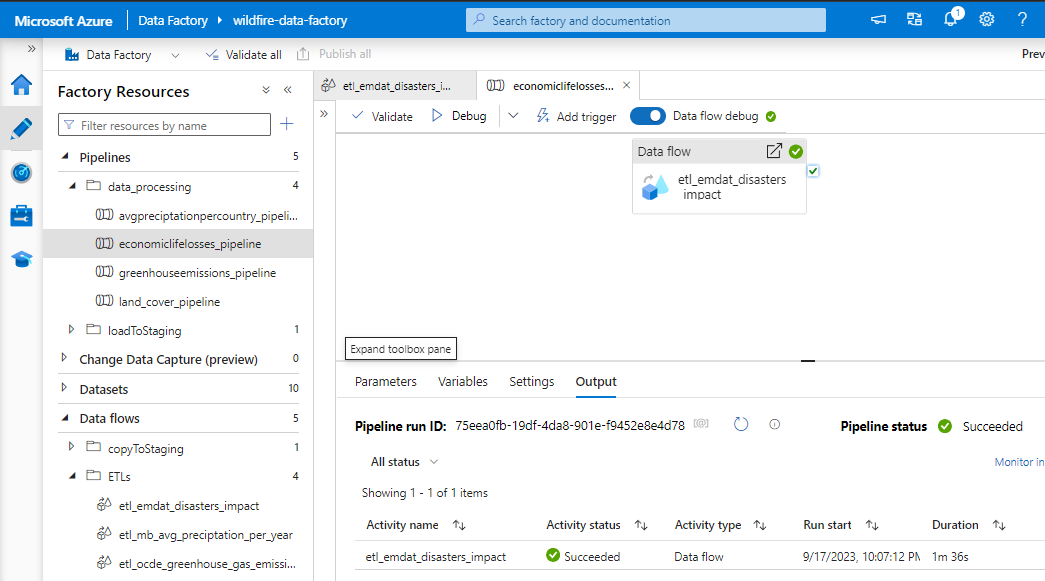
## emdat\_disasters\_impact\_uidds2nac

Essa tabela foi designada para avaliar os impactos econômicos e humanos acerca das queimadas. Portanto, para essa tabela, selecionou-se os campos “Dis No“ , “start Year“, “start month“, “start day“, “end year“, “end month“, “end day“, “Seq“, “Disaster Type“, “Country“, “ISO“, “Total Damages Adjusted ('000 US$)“ . Dos quai o campo “Disaster Type“ foi usado para filtrar todos as outras categorias de desastre que não fosse “Wildfire”. Já os campos “Total Damages Adjusted ('000 US$)“ e “Total Deaths” tiveram os valores nulos substituídos por 0, pois entende-se que no contexto de desastres naturais valores nulos para esses campos significam que não houve um registro e portando não houve nenhum registro. Outra transformação a ser feita foi no campo ISO do país referente a cada registro, alguns estavam com esse campo null. Como foi percebido que os três últimos caracteres do campo “Dis No” era o código ISO do país, então bastou extrair esses três dígitos e substitui-lo no campo “ISO” eliminando assim todos os valores nulos, já que o campo “Dis No” é uma chave primária que por definição não pode ter valor nulo.

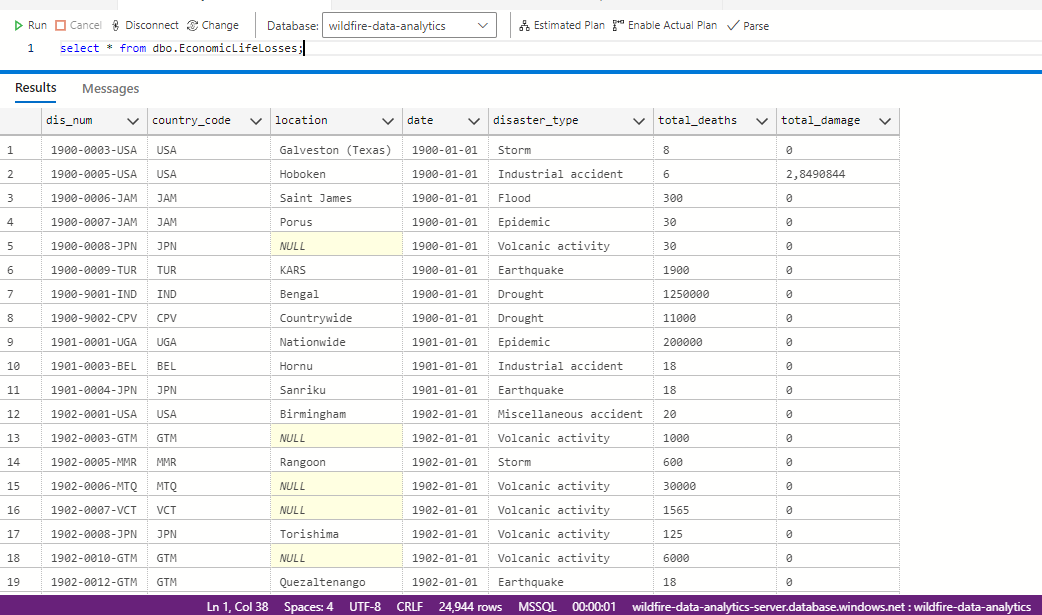
A seguir o “dataflow” utilizado no “Azure Data Factories para” realizar cada uma das transformações acima para gerar a tabela “*EconomicLifeLosses*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



E os dados carregados no Azure Database podem ser vistos na imagem abaixo:



O notebook com o qual se fez todos os experimentos de análise ao longo deste tópico pode ser acessado pelo link abaixo:

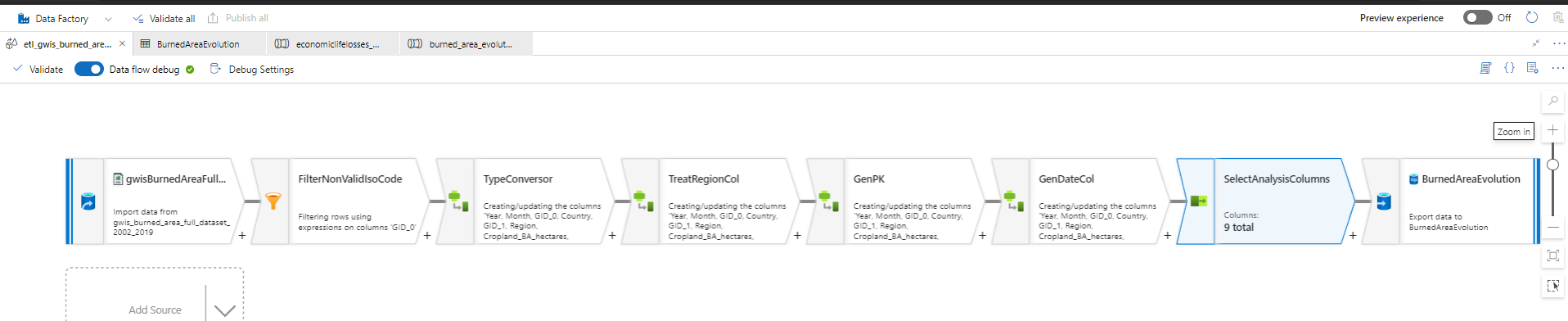
[Notebook para implementação de análises - Github](https://github.com/lucasvittal2/wildfire_analysis_data_engineer_project/blob/main/Notebooks/wildfire-analysis-implementation.ipynb)

## gwis\_burned\_area\_full\_dataset\_2002\_2019

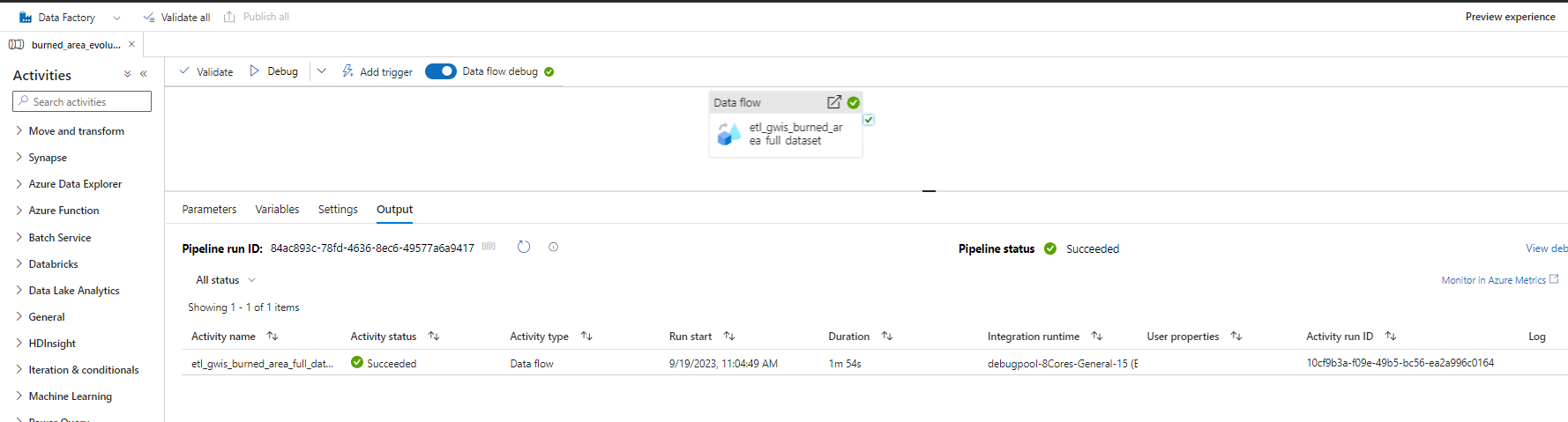
Essa tabela foi obtida com o intuito de se analisar a evolução das queimadas ao longo do tempo. Como todos os campos nela presente foram identificados como úteis à nossa análise eles inicialmente foram utilizados. Entretanto vale ressaltar que embora todas as informações tenham sido utilizadas alguns campos viraram outros campos. Especialmente os campos “Year”, “Month” e “GID\_0” , cuja combinação entre eles gerou o campo “id” na tabela final “*BurnedAreaEvolution”* ao qual fora usado como uma chave primária e que, portanto, ambos os três campos não poderiam ser nulos.

Além da combinação anteriormente citada, aplicou-se um filtro para retirar os códigos ISO’s de país não válidos. Fez-se as devidas conversões para valores numéricos dos quais originalmente eram representados como “string” devido terem sido extraídos de um arquivo csv. Ainda, também vale mencionar que foi necessário fazer um tratamento no campo “Region” pois alguns nomes de regiões continham aspas simples o que conflitava com o SGBD do “Azure SQL Database”.

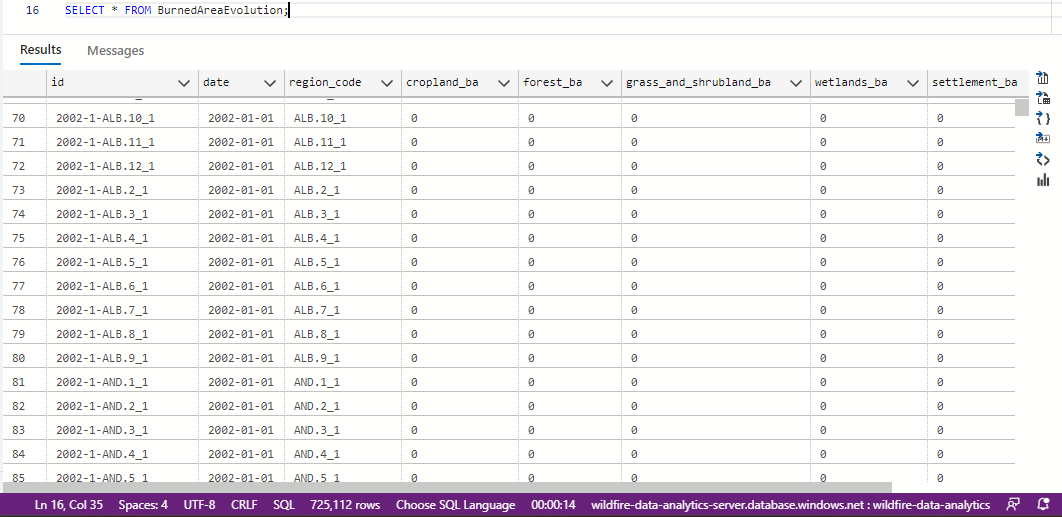
A seguir o “dataflow” utilizado no “Azure Data Factories” para realizar cada uma das transformações acima para gerar a tabela “*BurnedAreaEvolution*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



## Tabelas temperaturas locais

Com o intuito de poder analisar as como as temperaturas locais influenciam nas queimadas então coletou-se dados de diversas fontes. Objetivo inicial era ser ter uma única tabela com todos os dados de temperatura locais a nível de estado ou província de uma país. Contudo não foi possível obter um banco de dados com esse nível de abrangência e, portanto, foi-se necessário coletar dados de diversas fontes, que por sua fez diziam respeito a diferentes níveis de granularidade com respeito a localidade de onde os dados foram coletados (cidade e Estado). Além disso alguns possuíam dados de temperatura em Fahrenheit ou outros em Celsius, sendo necessária a conversão das unidades de temperatura. A obtenção desses dados não foi algo trivial, sendo necessário obter os dados por partes nas origens através da configuração de alguns parâmetros, pois alguns sites não permitiam baixar todos os dados de uma vez seja por políticas do site, seja pelo fato da base de dados se muito grande.

Os dados obtidos para essa etapa estão consolidados nas seguintes tabelas:

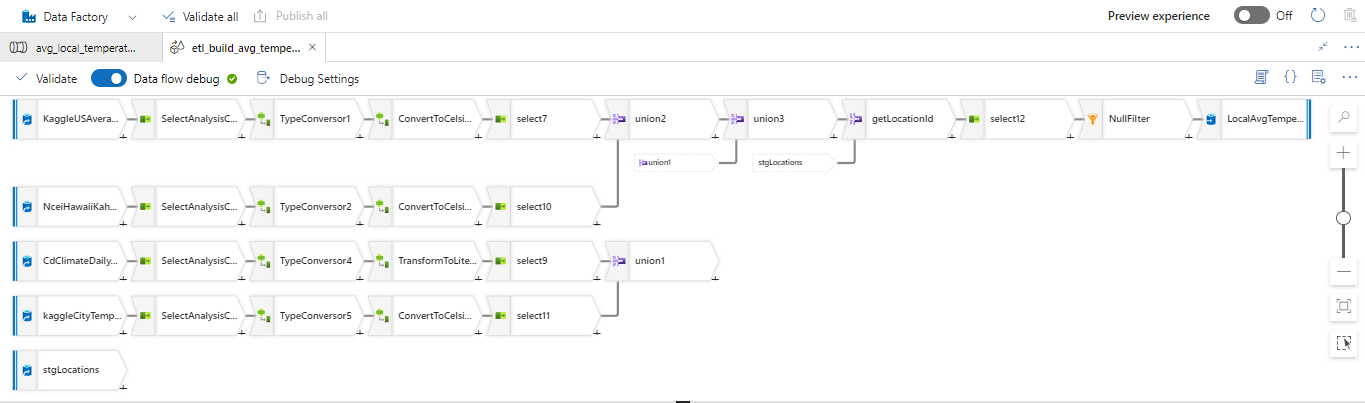
* ncei\_hawaii\_kahului\_avg\_temperature
* kaggle\_US\_average\_monthly\_temperature\_by\_state\_1950\_2022
* cd\_climate\_daily\_canada\_per\_state
* kaggle\_city\_temperature

As devidas transformações foram realizadas conforme o mapa de transformações com a finalidade de se obter tabelas que possuíssem um mesmo esquema de dados e mesmas unidade de medida para os dados de temperatura. Assim foi foi-se possível juntar todas essas tabelas em uma única tabela com o intuito de fazer análise no aspecto de locais temperatura acerca das queimadas. Com a geração de tabelas padronizadas as mesma foram concatenadas por meio da operação de união para que por seguinte se fizesse um “Join” com a tabelas “*Locations*”, uma tabela criada para ser dimensão de localização da tabela gerada pelos uniões. Essa tabela foi gerada a partir de um script Python que extraiu os nomes das cidades e estados e gerou identificadores únicos para cada localidade. Mais detalhes podem ser contemplados no arquivo abaixo:

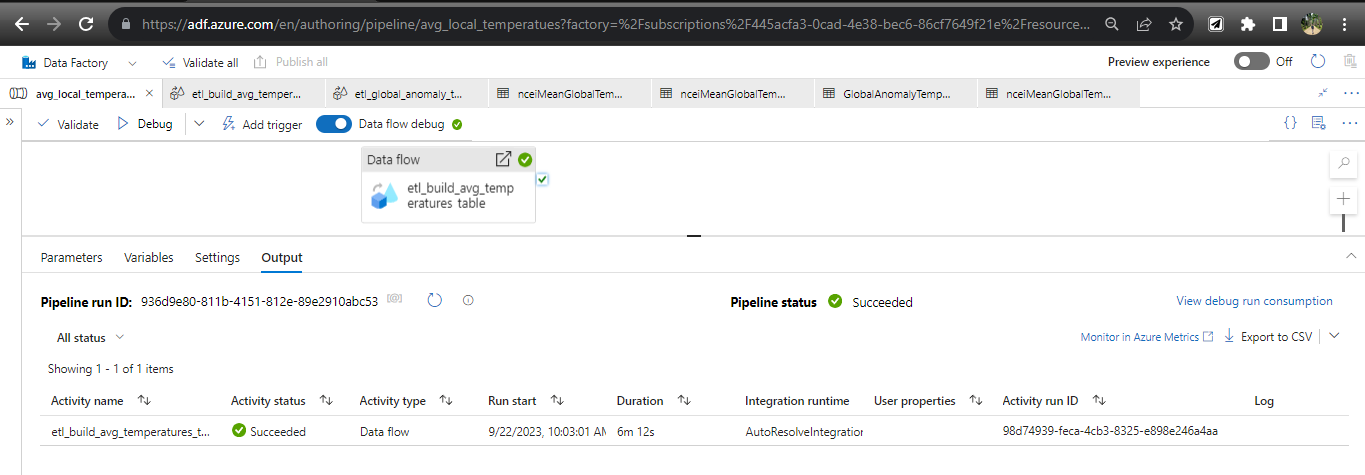
[Script para gerar dados das localidades](https://github.com/lucasvittal2/wildfire_analysis_data_engineer_project/blob/main/Assets/scripts/gen_location_table.py)

O resultado de todo o processo descrito acima foi a geração da tabela *“LocalAvgTemperatures”*, que usaremos posteriormente para análises de temperatura.

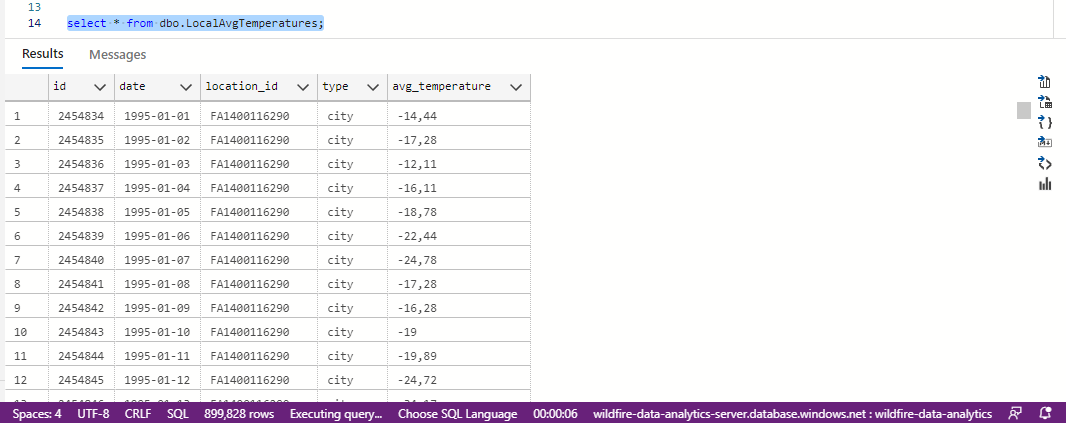
A seguir o “dataflow” utilizado no Azure Data Factories para realizar cada uma das transformações acima para gerar a tabela “*LocalAvgTemperatures*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



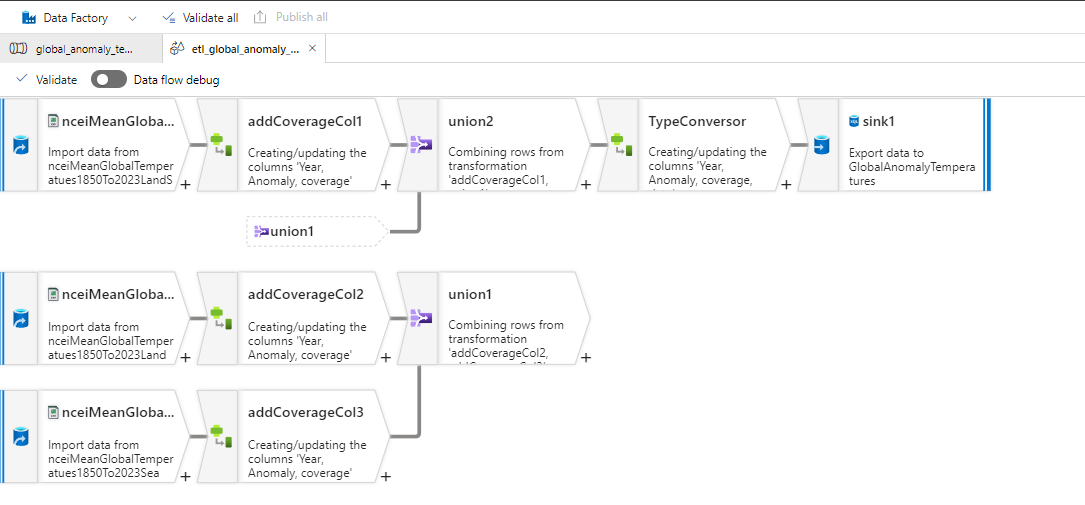
E os dados carregados no Azure Database podem ser vistos na imagem abaixo:



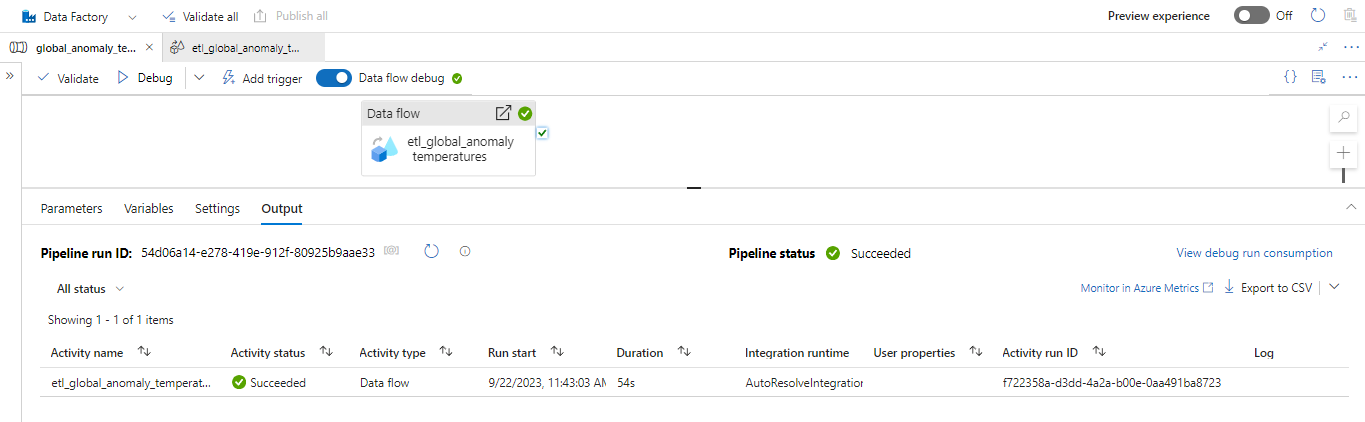
## Tabelas temperaturas globais

Como mencionado foram obtidos dados de temperaturas globais durante o processo de coleta de dados. Esses dados dizem respeito a anomalias de temperatura mês a mesma desde 1850. Anomalias em nosso contexto significam a diferença entre a média de temperatura de 1900 à 2000 e a medida de temperatura no ano corrente ao registro. Tais dados foram obtidos considerando três abrangências, terra, mar e ambos. Com o intuito de prover o máximo de flexibilidade as análises posteriores, uniu-se as três tabelas e fez-se os devidos tratamentos dos dados, o que teve como resultado a tabela “*GlobalAnomalyTemperatures*”. O tratamento dos dados foi apenas gerar o campo de date no formato de date e converter o campo “anomaly” para “float”, que originalmente estava em “string” devido a tabela original estar armazenada num arquivo csv.

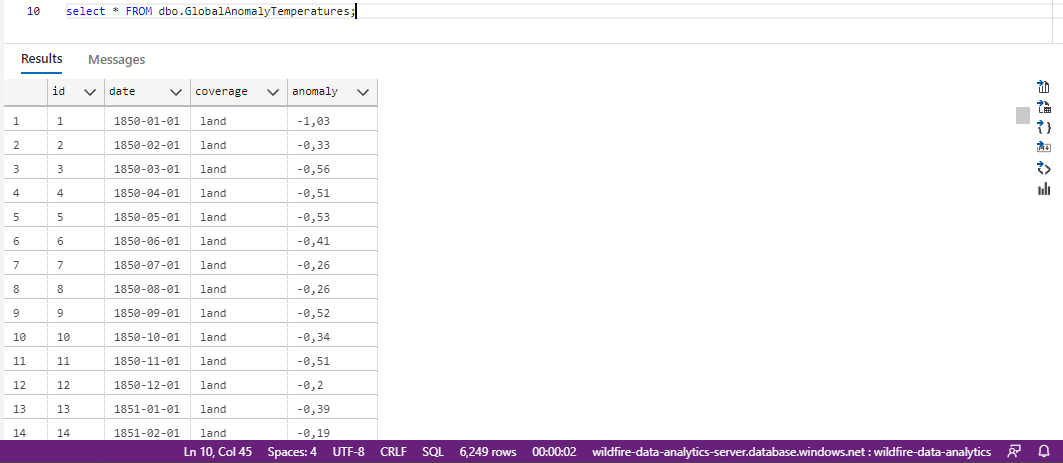
A seguir o “dataflow” utilizado no Azure Data Factories para realizar cada uma das transformações acima para gerar a tabela “*GlobalAnomalyTemperatures*”:



A execução do “dataflow” que implementa todas as transformações descritas pode ser vista abaixo:



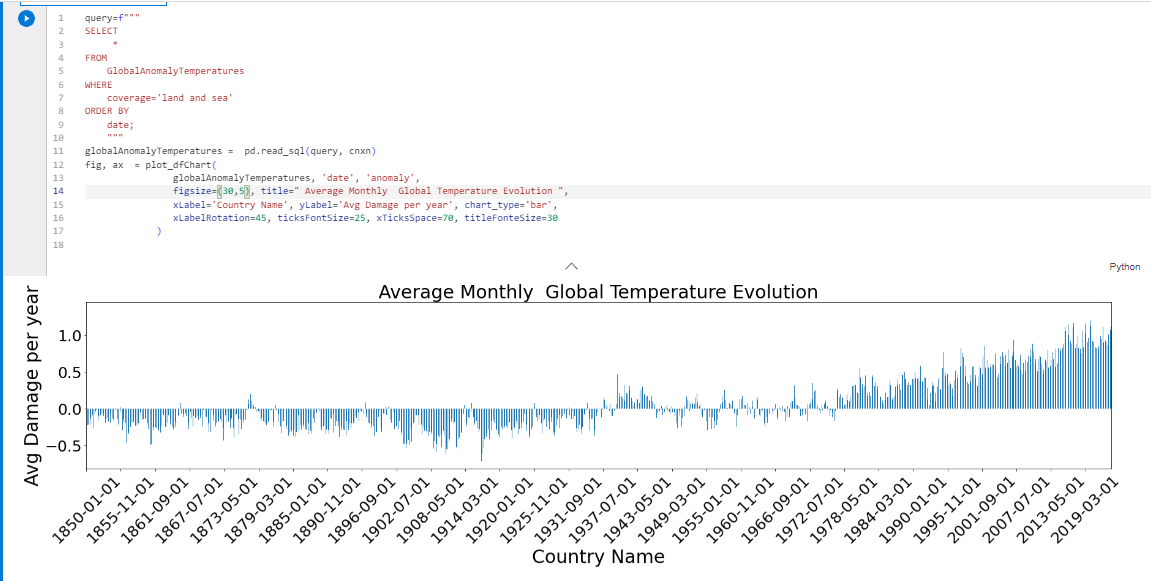
E os dados carregados no Azure Database podem ser vistos na imagem abaixo:



# Análise dos dados

Com os dados devidamente transformados e carregados, agora é possível fazer consultas no data Warehouse construído afim de se obter as respostas inicialmente buscadas nesse projeto. Basicamente seguiu-se 4 linhas de análise. A primeira diz respeito a evolução das queimadas ao longo do tempo e o perfil desse fenômeno nos países mais afetados. A segunda linha diz respeito a análise de perdas econômicos e humanas devido a esse fenômeno. A terceira linha de análise diz respeito as características locais dos lugares mais afetados em termos de temperatura, cobertura vegetal e níveis de precipitação para se tentar inferir o porquê tais localidades são as mais afetadas e como fatores relacionados a mudanças climática tal como emissões de carbono, aumento das temperaturas globais influenciam nos fatores locais para propiciar condições para ocorrência do fenômeno das queimadas.

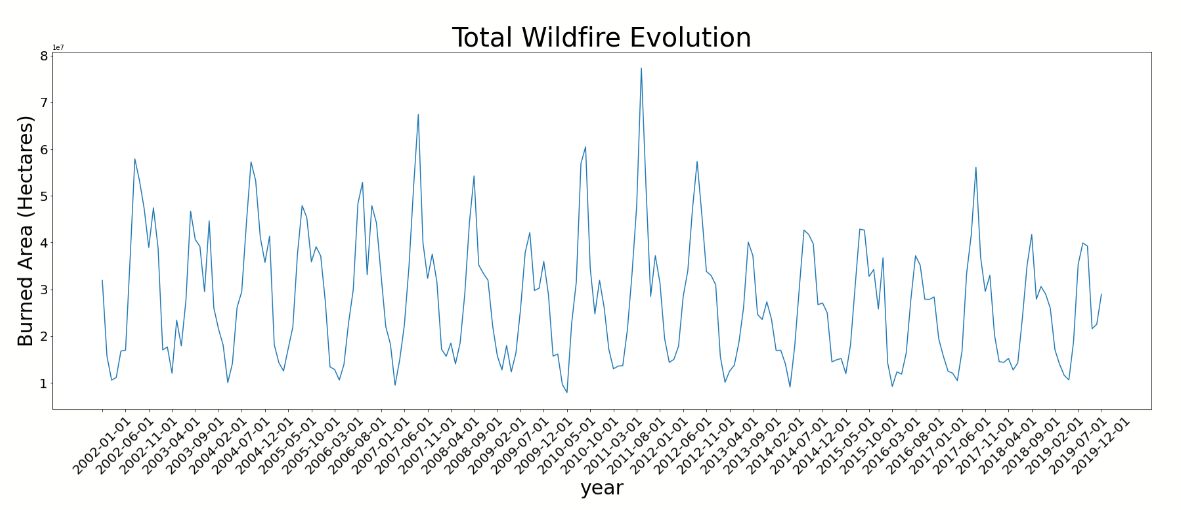
Para a realização das análise utilizou-se a biblioteca “pyodbc” para se realizar as consultas no Data Warehouse anteriormente projetado e as bibliotecas gráficas e numéricas do pandas, numpy, scipy e matplotlib tal como o exemplo abaixo:



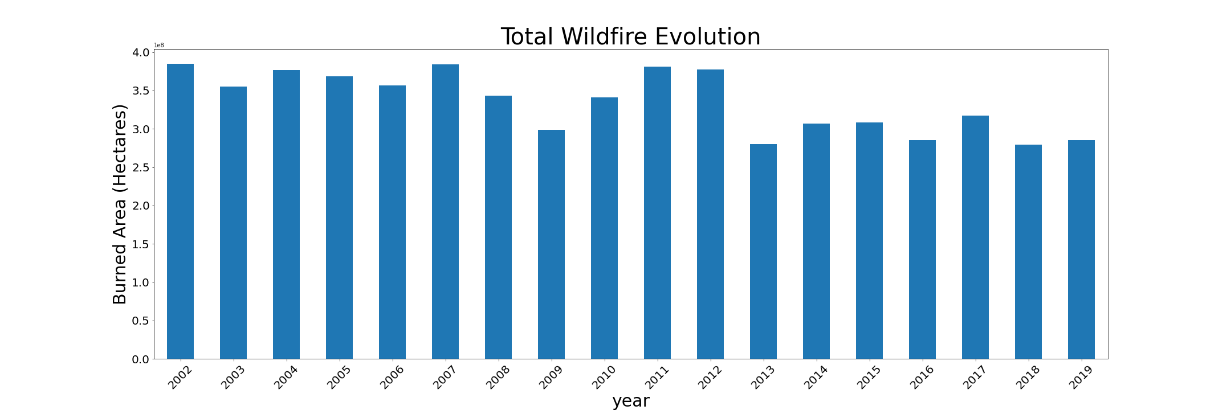
A seguir apresentaremos mais detalhadamente cada análise.

## A evolução das queimadas ao longo do tempo e o perfil desse fenômeno nos países mais afetados

Considerando essa linha de análise a primeira coisa a que se pensou em levar em consideração é como se dá o fenômeno das queimadas ao longo do tempo. Sendo assim fez-se um agrupamento dos dados pela data de registro a partir da soma da soma dos registros de queimadas, ou seja, somou-se os dados relativos a todos os campos de áreas afetadas, agrupou-se isso pela data de registro e somou-se todos os registros obtidos numa mesma data. A partir disso gerou-se o gráfico abaixo que diz respeito a como se tem dado os fenômenos das queimadas de 2002 à 2019:



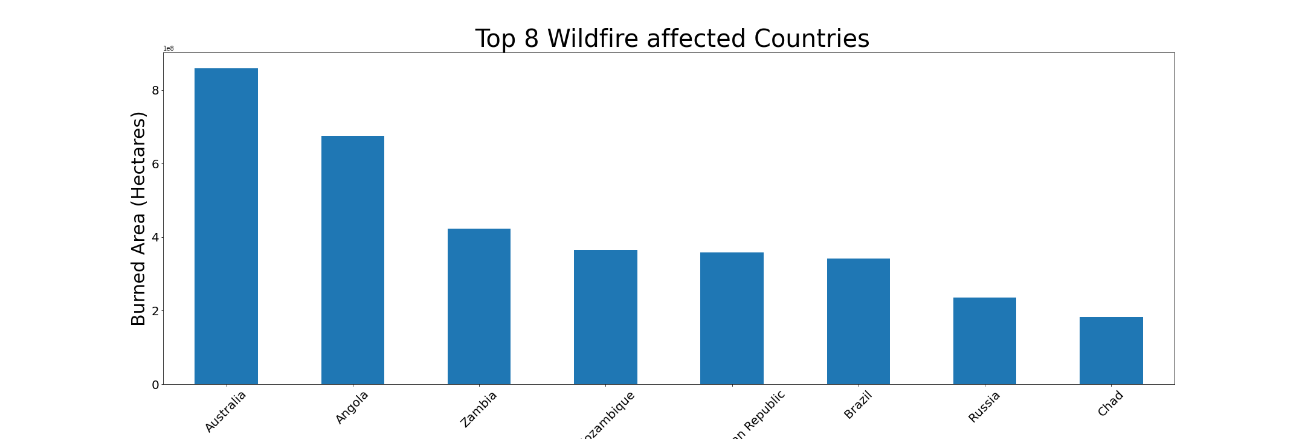
Ao se observar o comportamento do fenômeno no período citado, chama a atenção o fato de que globalmente não se tem observado uma tendência de crescimento na intensidade do fenômeno de acordo com os dados observados, e sim uma tendência de estabilidade. Ao mesmo tempo também é notório que o fenômeno das queimadas possui uma sazonalidade muito bem definida, tendo sua intensidade com período de pico e vales com períodos muito bem definidos.

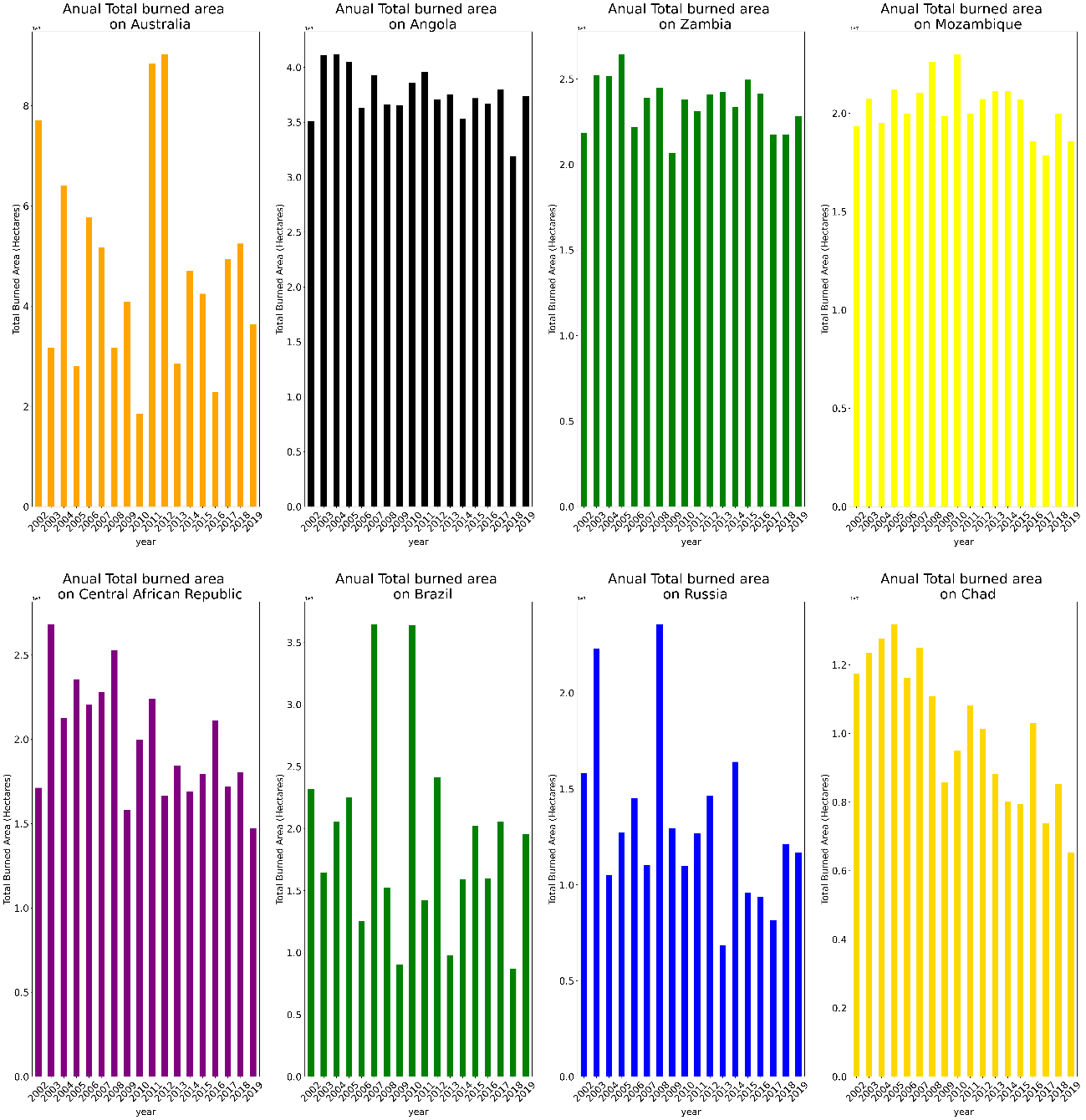
Para se ter uma ideia um pouco melhor sobre como se tem dada a evolução do fenômeno, também fez-se uma análise considerando o total de área queimadas por ano:se

Também não se observou comportamento diferente.

## Análise Localidades mais impactadas

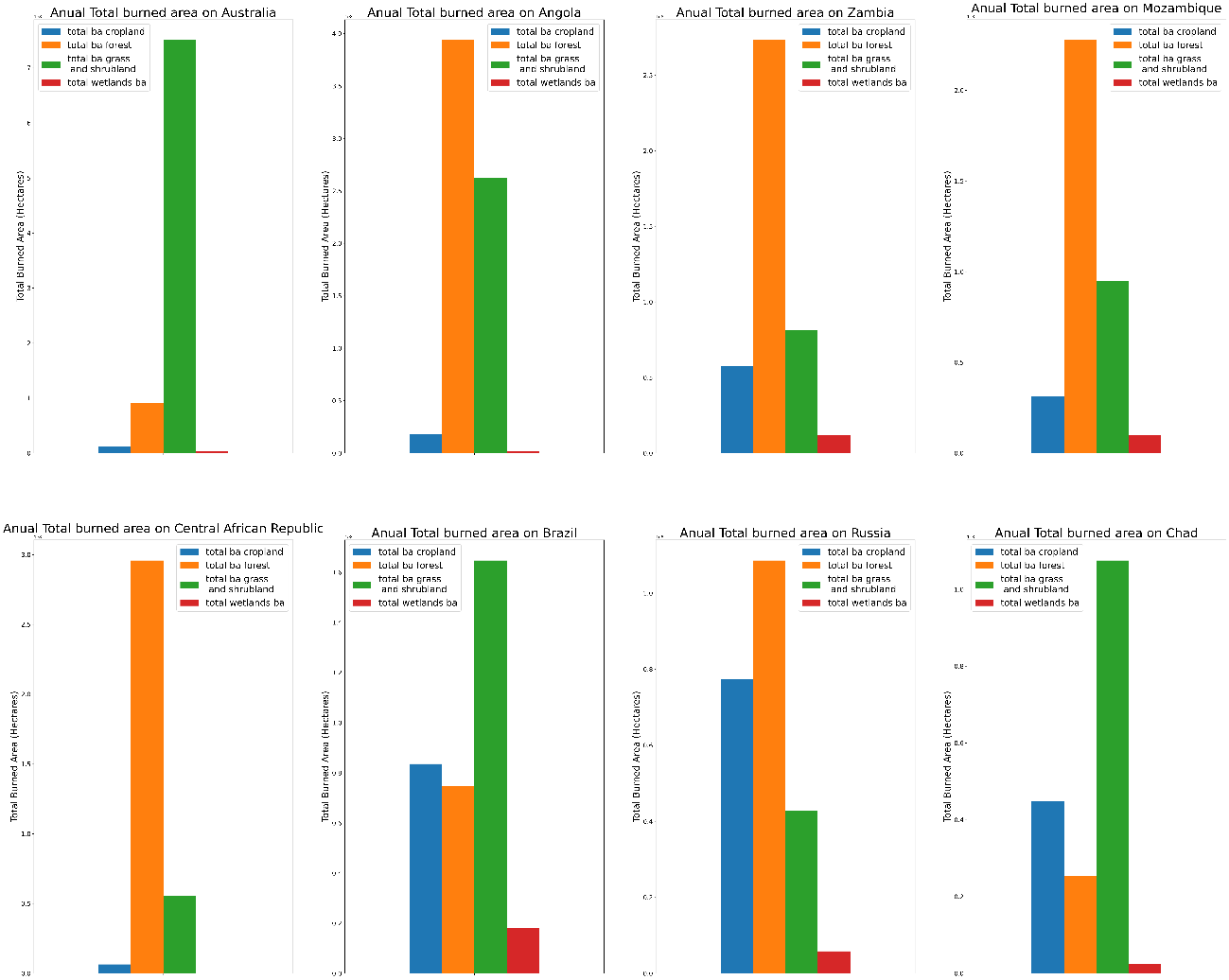
Imaginando que o comportamento seria diferente relativo a cada país então resolveu-se analisar localmente os países mais afetados. Sendo assim primeiramente tomou-se os países mais afetados para assim avaliar o total acumulado anualmente em cada país:





Tirando o Chad em que se observou uma tendência de queda, todos os demais países não demonstraram nenhuma tendência de mudança. Sendo assim grande parte dos mais afetados seguem a minha linha de comportamento global, ou seja, de acordo com os dados inferidos, uma tendência de estabilidade.

Além disso, também procurou-se entender o perfil dos países mais afetados pelas queimadas, e ao avaliar os perfis de cobertura vegetal percebeu-se uma particularidade comum entre todos os países afetados:

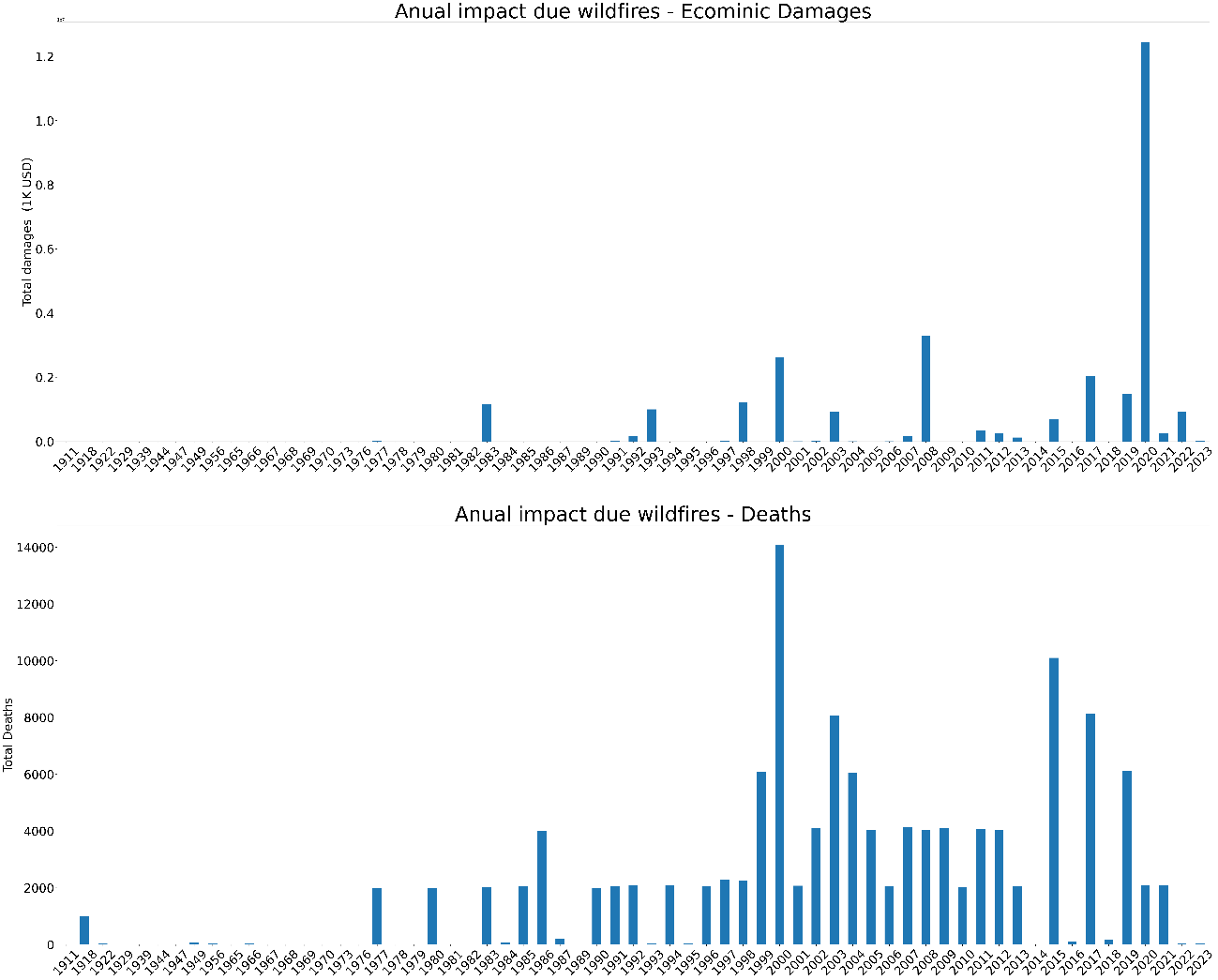


Percebeu-se que todos os países afetados tiveram como áreas mais afetadas as áreas de cobertura com gramíneas e arbustos ou com cobertura florestal. Com uma rápida pesquisa constata-se que essas áreas são mais susceptíveis a incêndios porque as áreas florestais tendem a acumular biomassa que tende a secar mais rápido e serve como combustível para os incêndios, já a área de gramíneas e arbusto é um tipo de vegetação típico de savana e ,portando, mais secas. Por isso tendem a incêndios com maior probabilidade. É importante citar que as causas citadas são uma das possíveis causas e que a ocorrência de incêndios é devido a combinação de fatores climáticos, de vegetação, geográficos e dentre outros.

## Impactos econômicas e perdas Humanas decorrentes das queimadas

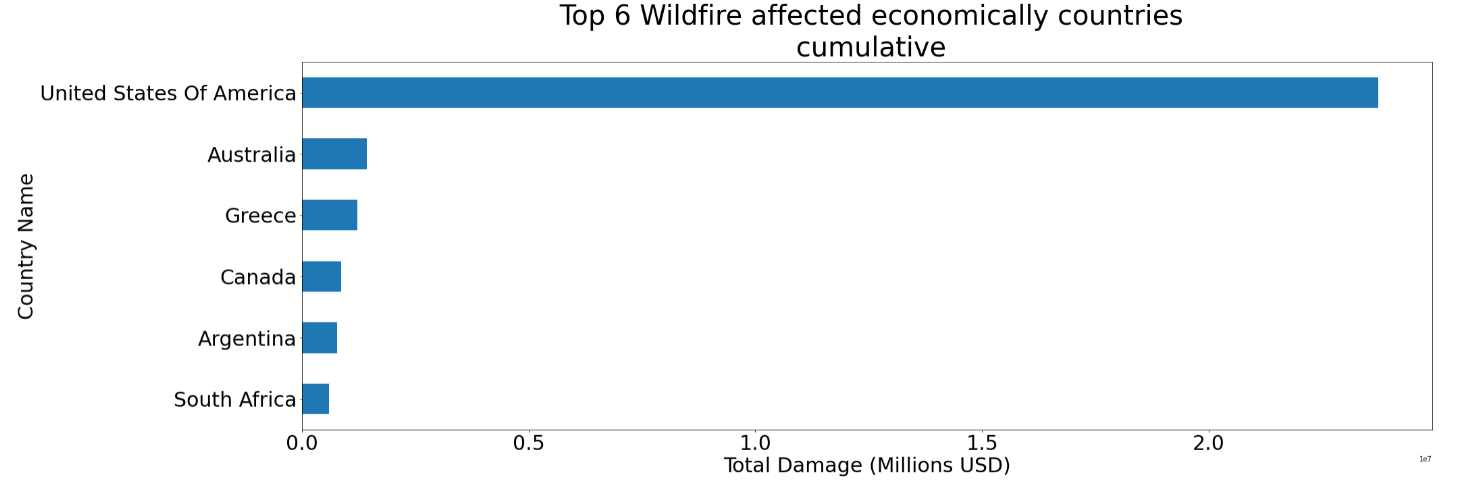
Na análise anterior foi feita uma análise de intensidade das queimadas considerando que a intensidade das queimadas está intrinsicamente ligada a quantidade de área queimada pelo fogo, e não se percebeu-se por essa linha de análise algum tipo de mudança significativa com o passar dos anos. Entretanto, não foi analisado anteriormente como esse fenômeno está afetando pessoas. Nesse sentido aqui vamos avaliar tais impactos em termos de peradas econômica e humanas, ou seja, em qual foi o prejuízo causado pelas queimadas e o número de mortes a ela associado.

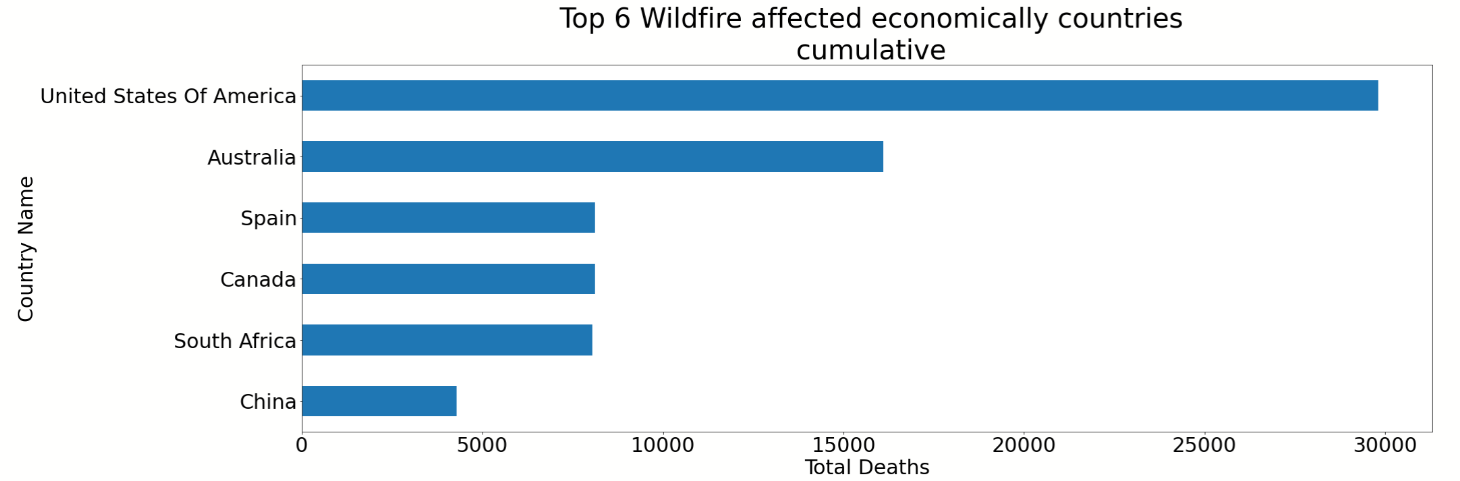
Primeiramente vamos analisar globalmente a evolução das perdas econômicas e humanas anuais de acordo com a abrangência dos dados disponíveis:



Ao se avaliar a magnitude das perdas econômica e humanas, percebeu-se que alguns anos houve picos e nesse sentido uma rápida pesquisa fora feita para se entender o porquê mediante a busca por notícias sobre grandes eventos que justifiquem o pico nesses anos. Nesse sentido achou-se alguns eventos que justificassem picos em prejuízos em alguns anos. A exemplo do “Black Summer” na Austrália e o “Servere August” no oeste dos Estados Unidos, eventos esse que ocorreram no ano de 2020 tendo os eventos ocorrido no oeste dos Estados unidos maior repercussão devido a terem afetados pessoa pertencentes ao meio artístico e cinematográfico. Também se encontrou grandes eventos no ano de 2018, 2009 e 2015, alguns dos quais não se refletiam nos dados ao qual temos disponíveis, o que talvez revele uma incompletude dos dados ou um investimento mais intenso em estudar os eventos ocorridos nos períodos em que os dados foram coletados, o que , por questões de limitação, de tempo não poderá ser feito nesta interação.

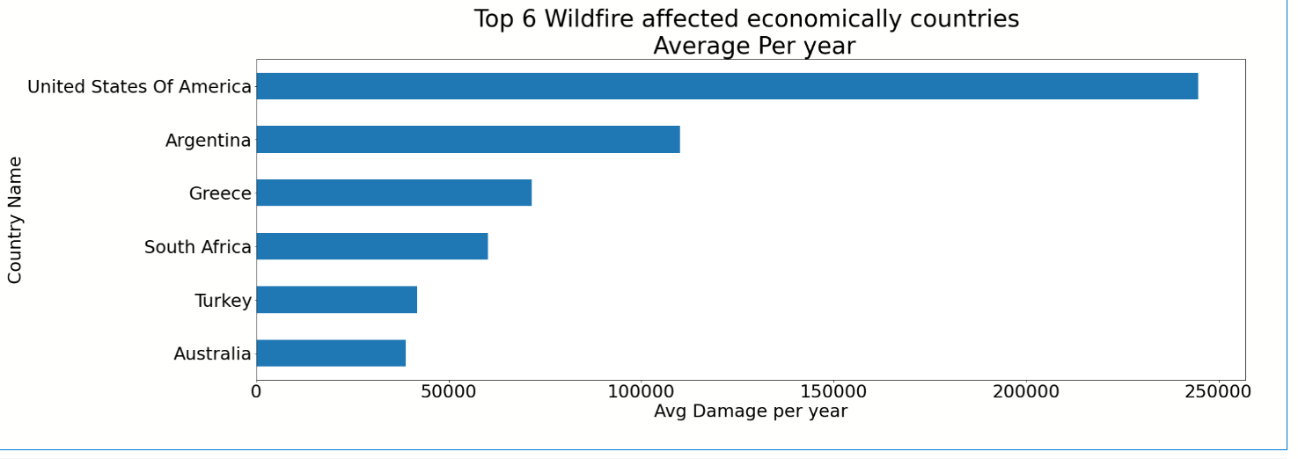
Por seguinte analisaremos os 6 países mais afetados em termos de perdas econômicas de humanas.

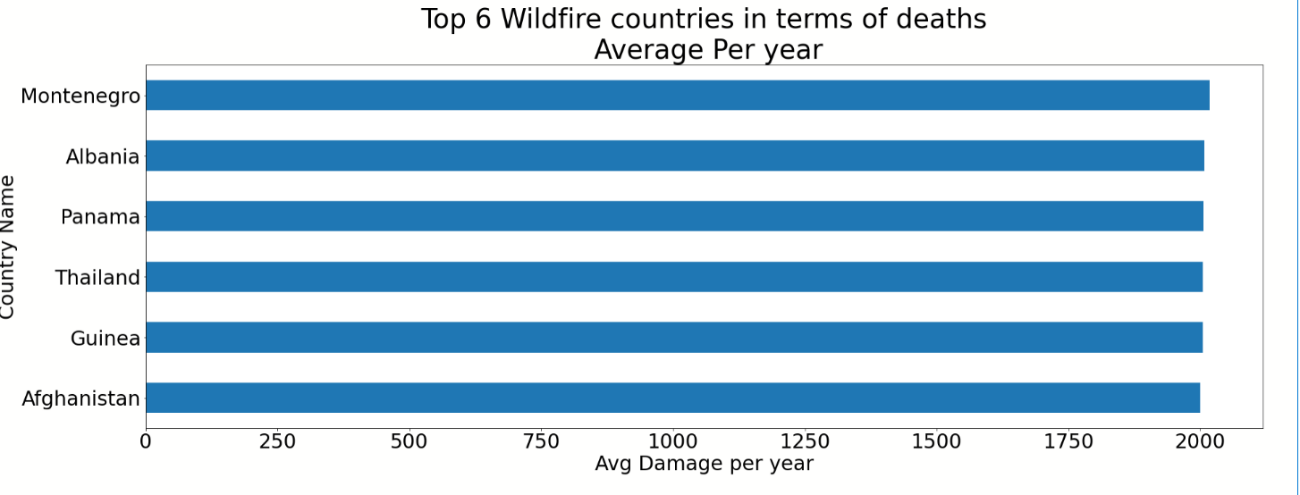




Se observa que os EUA têm a grande maioria tanto dos prejuízos econômicos como em termos de vidas perdidas devido às queimadas , e logo em seguida vem a Austrália. Os EUA e a Austrália junto detêm a maioria de todos os prejuízos em termos de vidas humanas que chama bastante atenção. Outro fato é que os EUA sozinho possuem a grande maioria dos prejuízos econômicos acumulados. O porquê dessa questão é um fator a ser olhado com mais calma posteriormente: será que é uma distorção dos dados que possuímos ou existe alguma característica nesses dois países que os fizeram deter a maior parte dos prejuízos econômico ou humano?

Outro fator que queríamos observa é quanto em média são os prejuízos econômicos e em termos de vidas devidos às queimadas:



Ao se avaliar a média os países mais afetados mudam, isso é um indicativo que nem sempre onde ocorrem as queimadas mais intensas é onde ocorrem mais frequentemente ocorrem mais queimada. Se observa que os países onde mais ocorrem mortes são países subdesenvolvidos e dentre uma das possíveis causas para a terem as maiores médias anuais é a falta de infraestrutura para socorro da população. Além disso, ao se comparar

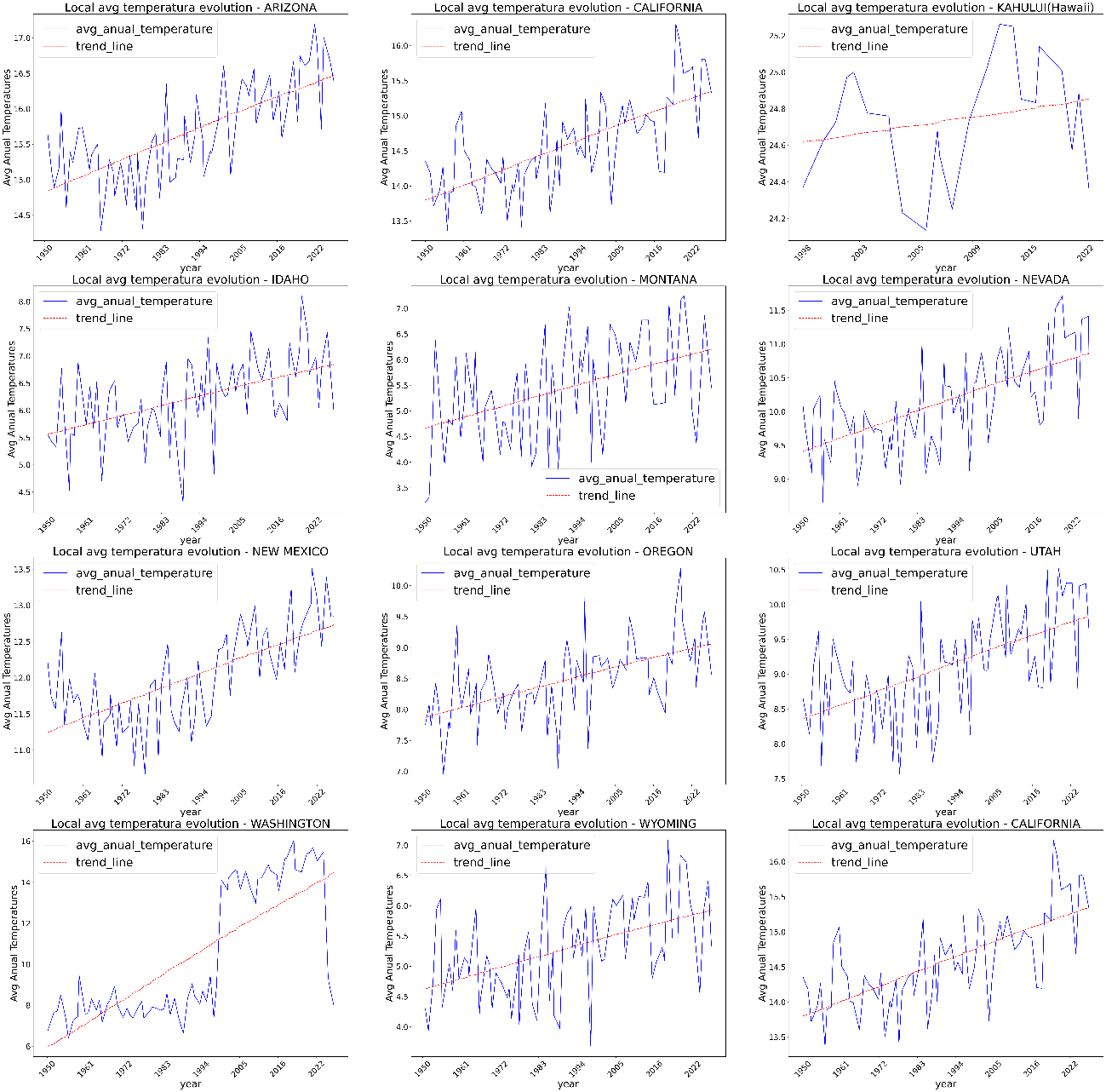
a análise desta etapa com a etapa anterior, se observa que boa parte das queimadas não implica em prejuízos humanos ou econômicos, pois os países onde mais tiveram hectares consumidos pelas queimadas não são os mesmos onde mais se teve prejuízos econômicos e humanos tanto no acumulado quanto na média. Outro fator que chama atenção é a avaliação da média anual de mortes devido às queimadas, praticamente todos os países com a maior média de morte são países subdesenvolvidos e com poucos recursos, o que mostra que a fata de recursos para prestar socorro a população está diretamente relacionada a um maior número de mortes num ano devido a ocorrência do fenômeno. Mas novamente vale ressaltar que se deve checar se tal conclusão é de fato plausível ou se isso seria alguma distorção dos dados, nesse sentido seria necessário um trabalho de validar os resultados.

## Análise das Características das localidades mais afetadas

Agora vamos avaliar as características locais dos locais mais afetados pelas queimadas. Nessa etapa a intenção inicial é avaliar características de temperatura, de precipitação e de vegetação das localidades mais afetadas com a finalidade de se determinar se existem características locais que propiciam os maiores intensidade no fenômeno das queimadas.

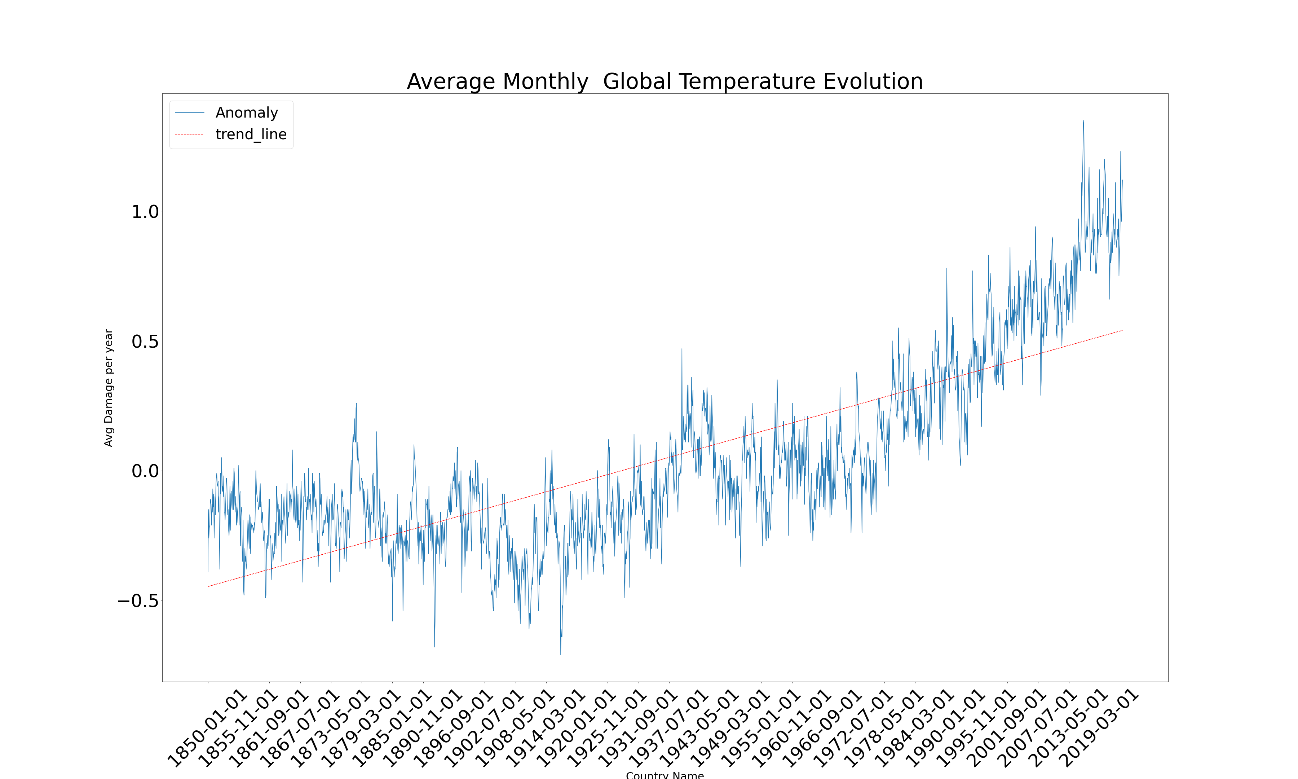
### Características de evolução temperatura anual média ao logo do tempo

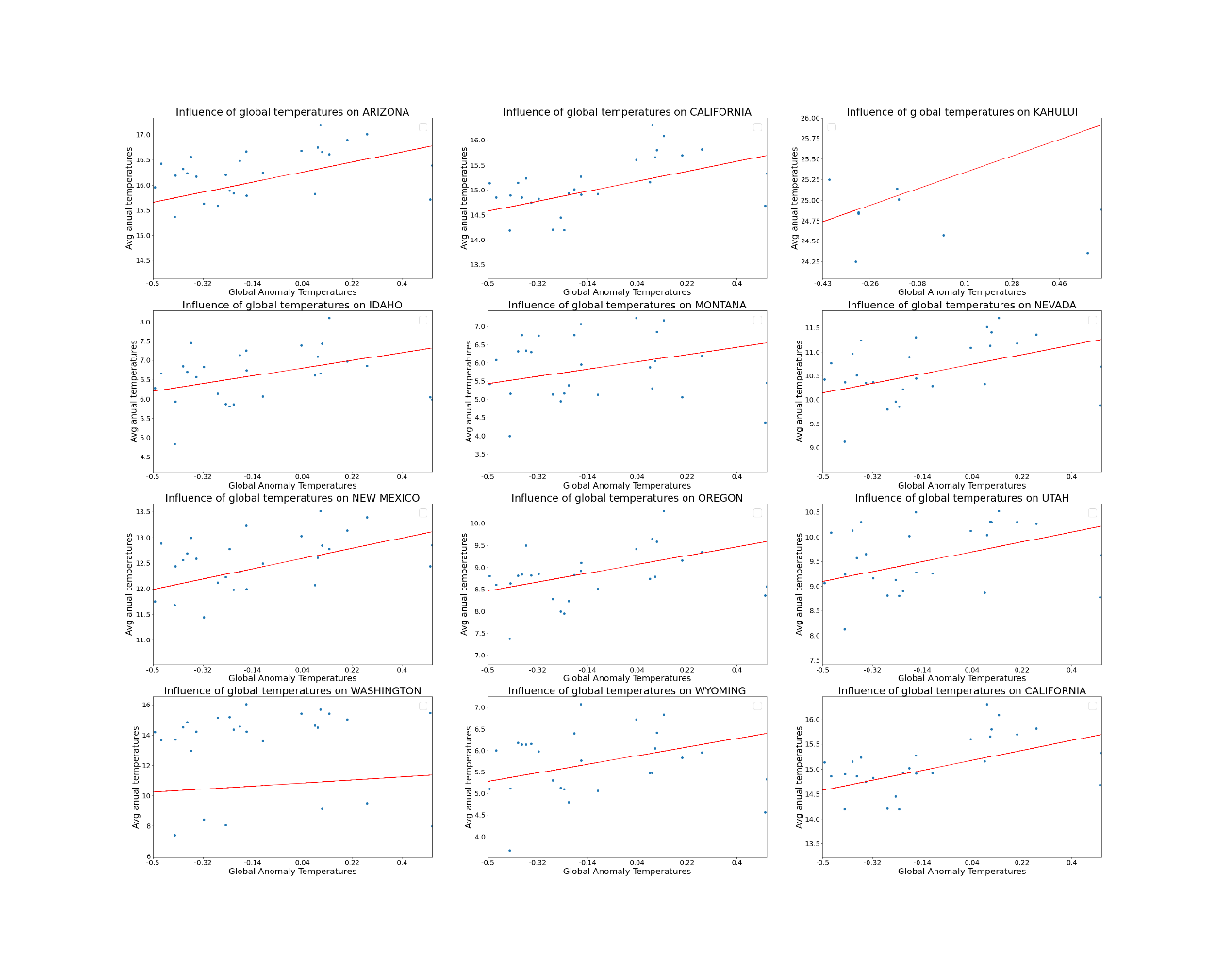
Primeiramente iniciamos com a análise da temperatura local anual média. A intenção inicial era avaliar as temperaturas médias em localidades dos países mais afetados. Contudo foi percebido que os dados de temperatura média global não contemplavam todos os países em que se desejava analisar. Por isso consideraremos somente os as temperaturas médias anuais dos estados ao oeste dos Estados unidos. Onde tipicamente ocorreram as queimadas de maiores intensidades e os a maioria das localidades estão contidas nos dados.



### Influência do aumento da temperatura global na temperatura local

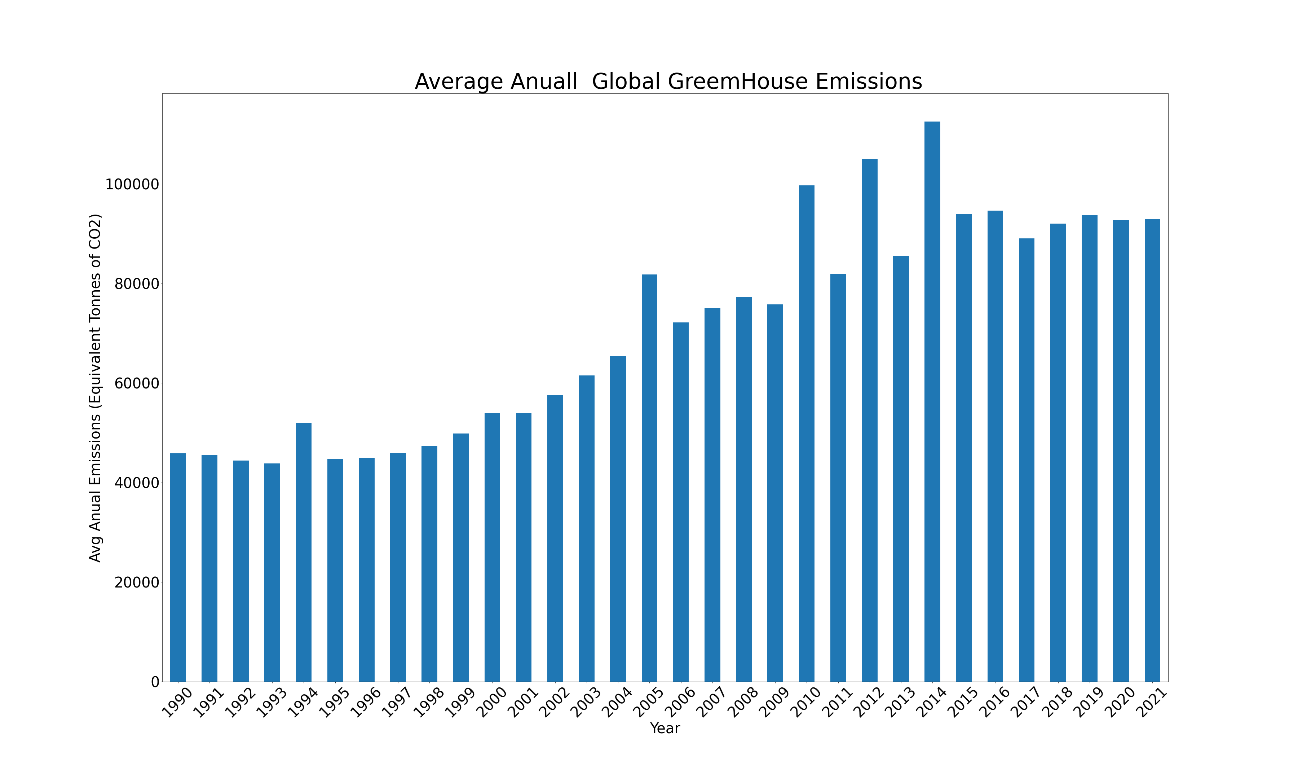
Ao se avaliar as temperaturas médias um ano nesses lugares, percebe-se que todas no geral estão com uma tendência de crescimento. O que de certa forma já esperado dado que a temperatura global também está crescendo tal como podemos ver a seguir:



Pois percebeu-se que todos as temperaturas locais tem algum grau de correlação com o aumento da temperatura global, alguns mais outros menos tal como se pode ver nos graficamente abaixo:

É sabido que o aumento das temperaturas globais está altamente relacionado ao as emissões de gases de efeito estufa, nesse sentido será que o aumento da temperatura global não estaria relacionado um suposto aumento das emissões dos gases e efeito estufa? Vamos verificar isso com os dados do DW anteriormente desenvolvido.

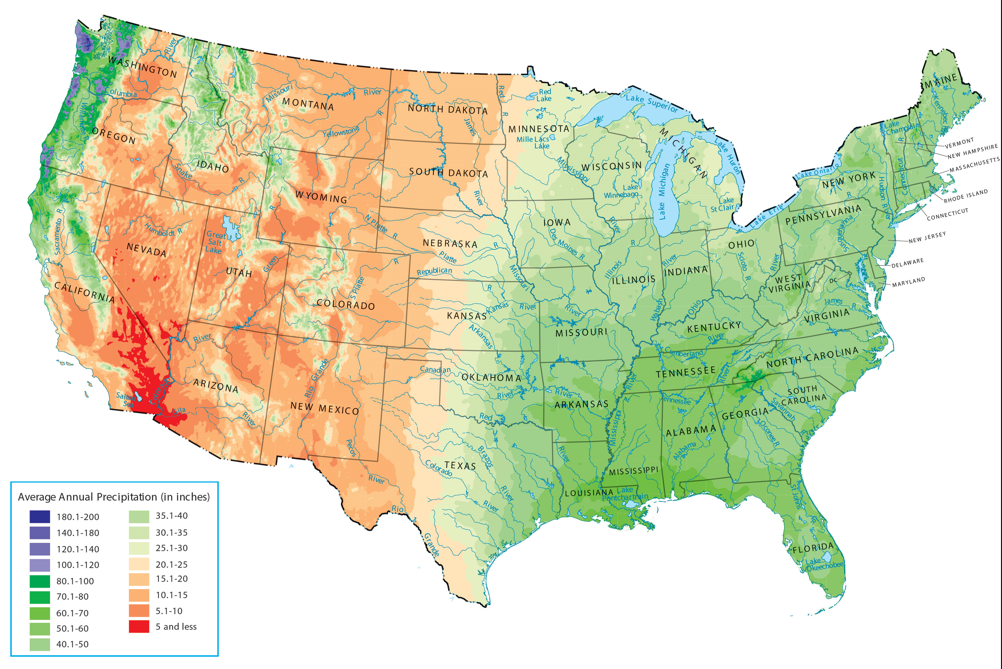
### Como as emissões de carbono influenciam na temperatura global



Ao se avaliar os dados percebe-se que durante 24 anos as emissões de carbono seguiram aumentando e que nos últimos anos ela segue uma tendência de estabilidade. Ao se avaliar a curva de temperatura global, ano após ano a temperatura não para de aumentar, mesmo que nos últimos anos as emissões não tenham sido aumentadas. Isso significa que a temperatura global possui um tempo de resposta muito grande em relação a quantidade de emissões de gases de efeito estufa, e que as emissões dos mesmos influenciam ao longo de muitos anos, o que torna perceptível porque em convenções do clima fala-se em redução das emissões de gás carbono para que a temperatura global não aumente tantos graus daqui a 30 ou 40 anos. Percebe-se então que as emissões de gases de efeito estufa estão relacionadas ao aumento da temperatura global à longo prazo e a mesma está atrelada ao aumento de temperaturas locais que por sua vez influenciam no regime das queimadas, logo isso significa que se as emissões de carbono não diminuírem à longo prazo a temperatura global irá cada vez mais aumentar, o que influencia nas temperaturas locais o que vai influenciar no aumento da incidência e da intensidade dos fenômenos das queimadas.

### Análise regime de Chuvas Local

Outro fator a ser analisado é o regime de chuvas, nesse caso focaremos na região oeste dos EUA. Infelizmente com os dados de precipitação obtidos não é possível se realizar uma análise consistente do regime de chuva, entretanto uma rápida pesquisa na internet permitiu obter um mapa de precipitação no território dos EUA o que serve muito bem como um dado complementar a nossa análise:

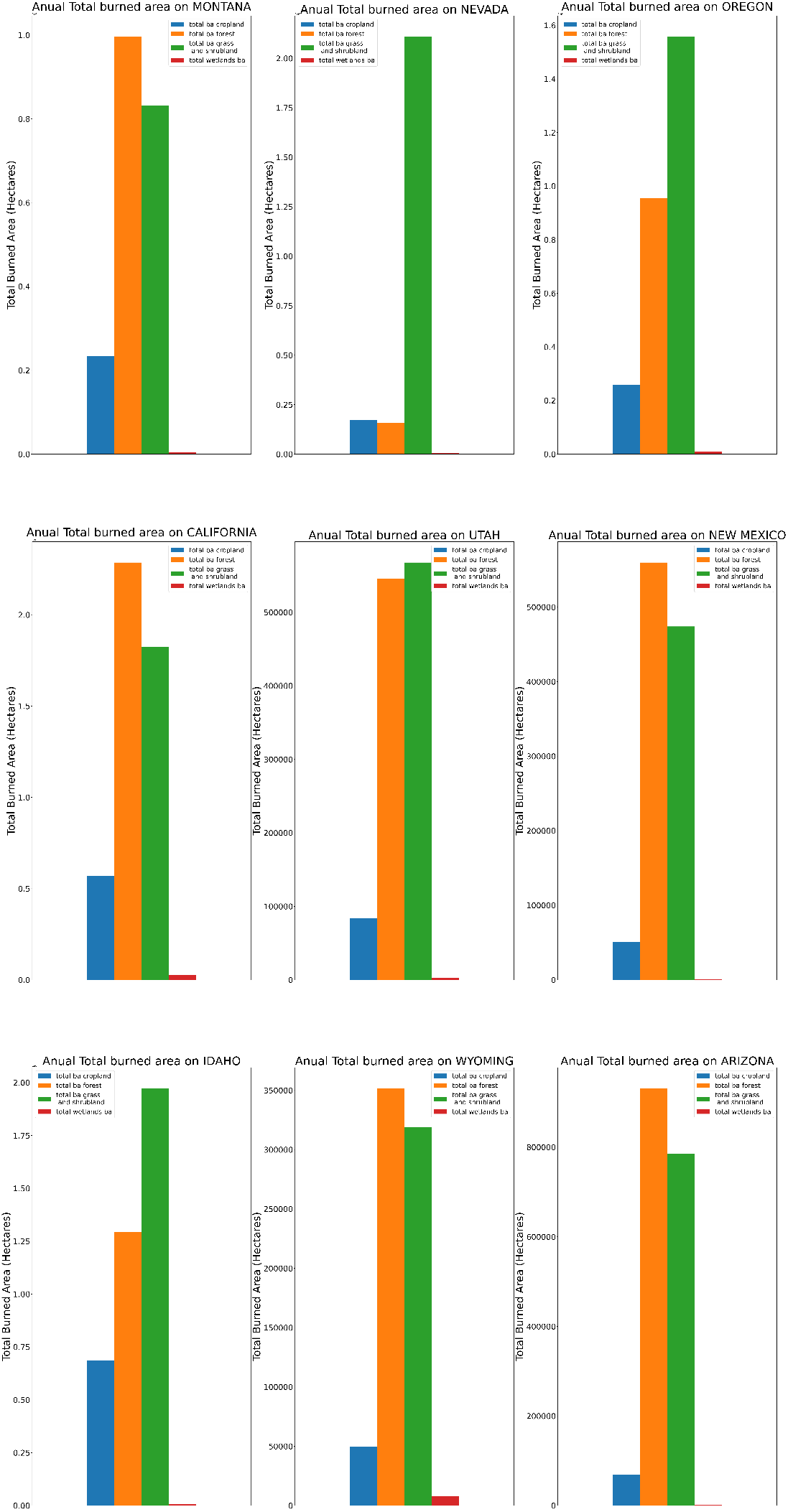


Fonte: <https://gisgeography.com/us-precipitation-map/>

E logo fica nítido porque os Estados da região oeste são os que mais sofrem com queimadas, são as regiões com os menores índices de precipitação do país. Por isso são as regiões mais propícias a ocorrência desses fenômenos

### Análise de característica de vegetação local

Por fim como último elemento de análise avaliaremos qual o tipo de área fora mais consumido pelas chamas nessas localidades, afim de se avaliar qual a composição de vegetação existe nessas áreas com de analisar com completude o porquê são as mais afetadas nos EUA:



Assim infere-se que a maior composição de vegetação desses locais é de florestas, gramíneas e arbustos. Exatamente o tipo de vegetação mais propícias as queimadas como foram observados em análises anteriores.

# Conclusões

A partir do trabalho que foi realizado neste projeto, alcançou-se boa parte dos objetivos estabelecidos inicialmente:

* Conseguiu-se determinar quais são os países mais afetados e como se deu a evolução das queimadas nesses países.
* Determinou-se quais são as características de vegetação das áreas mais afetadas.
* Inferiu-se o impacto econômico e os prejuízos em termos de vida causados pelas queimadas.
* Estabeleceu-se uma relação entre o aumento da temperatura global, as emissões de carbono e a incidência de queimadas.

O objetivo de determinar as características locais das localidades mais afetadas foi alcançado parcialmente, devido à incompletude dos dados, o que tornou necessário simplificar a análise.

O não cumprimento de todos os objetivos se deveu a algumas dificuldades, como a obtenção de todos os dados geográficos e climáticos relacionados a diversas localidades do globo, com agravante de tentar obter dados a nível de granularidade estadual em diversos países do mundo. Além disso, alguns dados só poderiam ser baixados de forma fragmentada, o que aumentou o tempo necessário para obtê-los. Também enfrentamos dificuldades com a utilização das plataformas de nuvem ao longo do projeto. Como mencionado anteriormente, devido ao alto custo da plataforma AWS, foi necessário reiniciar parte do trabalho na plataforma Azure, o que resultou em pelo menos uma semana a menos para a conclusão do projeto. Adicionalmente, o autor deste trabalho já havia utilizado a conta da Azure para outros projetos, tornando necessário procurar um cartão de crédito de um terceiro para obter créditos gratuitos na Azure.

Fazendo uma autoavaliação dos resultados de maneira geral, consideramos satisfatórios, embora haja alguns pontos a se observar. O escopo do objeto de análise escolhido foi muito grande, e embora desejássemos um desafio, isso limitou a profundidade das análises devido ao curto prazo para conduzir uma análise tão complexa. Além disso, ao longo deste trabalho, percebemos a importância de possuir domínio sobre a área em que o objeto de análise está inserido ou, pelo menos, envolver no projeto alguém com esse domínio, o que permitiria uma análise mais consistente do objeto em questão e decisões mais assertivas ao longo do projeto.

Por fim, mencionamos que há diversas possibilidades de evolução deste projeto, dada a riqueza dos dados obtidos. É possível, por exemplo, analisar os desastres naturais em diversas regiões do globo, estabelecer alguma relação entre mudanças climáticas locais e as emissões de carbono, analisar a relação entre a intensidade dos desastres naturais e o aumento da temperatura local, além de enriquecer os dados com a obtenção de informações adicionais para realizar outras análises.

Com isso, encerramos este trabalho, agradecendo a oportunidade de contribuir para a compreensão das questões relacionadas às queimadas e aos desastres naturais globais e pelo significativo aprendizado provido pela realização deste trabalho através da orientação dos professores Vitor Almeida e Silvio Alonso.

# Bibliografia

AVERAGE precipitation in depth (mm per year). **The World Bank**, 2020. Disponível em: https://data.worldbank.org/indicator/AG.LND.PRCP.MM?end=2020&start=1961. Acesso em: Agosto 2023.

BERKELEYEARTH. **Berkley Earth**, 2023. Disponível em: https://berkeleyearth.org/data/. Acesso em: Agosto 2023.

CLIMATEDATA. **climatedata**, 2023. Disponível em: https://climatedata.ca/download/#station-download. Acesso em: Agosto 2023.

ECONOMIC damage by natural disaster type, 1900 to 2023. **Our World Data**, 2023. Disponível em: https://ourworldindata.org/grapher/economic-damage-from-natural-disasters?time=1900.latest. Acesso em: Agosto 2023.

EMERGENCY Events Database. **EM-DAT**, 2023. Disponível em: https://public.emdat.be/data. Acesso em: Agosto 2023.

NOAA. **City Time Series**, 2023. Disponível em: https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/city/time-series/USW00022516/tavg/1/8/1895-2023?base\_prd=true&begbaseyear=1991&endbaseyear=2020. Acesso em: Agosto 2023.

OCDE. **OCDE.stat**, 2023. Disponível em: https://stats.oecd.org/Index.aspx?DataSetCode=LAND\_COVER. Acesso em: Agosto 2023.

SURANA, SUBHAM. Kaggle. **Temperature of All Countries (1995-2020)**, 2021. Disponível em: https://www.kaggle.com/datasets/subhamjain/temperature-of-all-countries-19952020/code. Acesso em: Agosto 2023.

THE Global Anomalies and Index Data. **NOAA**, 2023. Disponível em: https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies. Acesso em: Agosto 2023.

WONG, JUSTIN. Kaggle. **Average Monthly Temperature by US State**, 2022. Disponível em: https://www.kaggle.com/datasets/justinrwong/average-monthly-temperature-by-us-state. Acesso em: Agosto 2023.