|  |  |
| --- | --- |
|  |  |
| ETL Mini Project: Climate Data |  |
|  |  |
|  | 10.12.2020Data BootCamp |
|  | *Aaron Schwartzman*  *Benjamin Raehpour*  *Grecia Acosta*  *Kevin Athey* Ricardo Dominguez |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Technical Write-up report for ETL-Project We interested in creating a database of climate data to clean and normalize this data from multiple data sources. Because many experts agree that the rise of carbon emissions is a large contributor to climate change, we will include this data in this database as well. Our final database will include US climate and global emissions data for several decades that makes analyzing any trends that may be available that much easier. We will use a combination of multiple data sources and data science tools (Python, Jupyter Notebook, SQL) to help us extract, transform and load these data into a final production database. Data Sources | |  |
|  | For our data sources, we used Kaggle, we used three datasets, we used the following three files:  <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByCountry.csv>  <https://www.kaggle.com/yoannboyere/co2-ghg-emissionsdata>  <https://www.kaggle.com/headsortails/us-natural-disaster-declarations?select=us_disaster_declarations.csv>    Because there was a massive information on these files, we cleanup the columns and only keep the following: |  |  |

### Transformation

*In order to transform the data and use it in our study we performed the following:*

*\*Used Pandas functions in Jupyter*

*\*Notebook to load all three CSV files.*

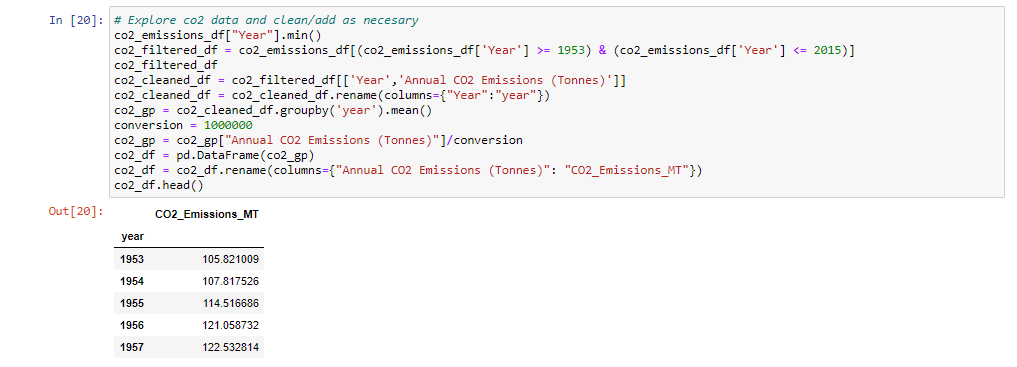
*\*Reviewed the files and transformed into data frames*

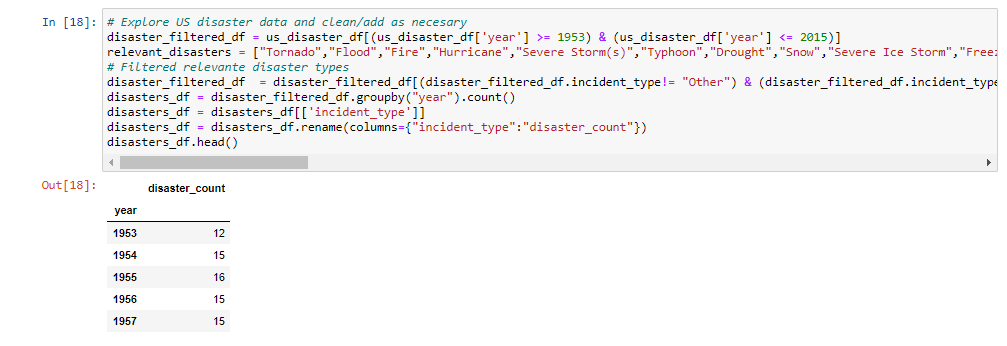
*\*Keep year for all three data sets, and temp average, number of disasters and tones columns.*

*\* Based on the data we decided to focus between 1953 and 2013 to the focus of this study.*

*\*Created queries to address our hypothesis by grouping the data by year and getting the average of the annual temperate and the CO2 emissions. We sorted the data so we could visually see if there were a big difference in the average temp per year.*







|  |  |  |  |
| --- | --- | --- | --- |
|  |  | LOAD: For the load process, we started by creating our ERD tables to show up the relationship between the datasets.  The last step was to transfer our final output into a Database. We created a database and respective table to match the columns from the final Panda's Data Frame using Postgres database using PG admin to store our original clean data sets. We reconnected to the database and generated additional tables for the data frames |  |
|  | Conclusion: Our plan take advantage of this ETL process and continue using this data for our final project.  Even with some limitations on our data, we were able to address our original question in our proposal for the ETL mini project. | |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Is Climate change real? A sunset over a city  Description automatically generated |
|  |