**GROUP 11 ETL PROJECT**

**Group Members:**

Shahanwaz Kahn

Kandace Herbert

Nwabueze Lucas Okose

**Introduction/Pre-Processes**

In this project we Group 11 created an ETL database surrounding Gun Violence. Our reasoning, was due to the large amount of gun related incidents for the past few years, we felt that it would be ideal to analyze the data to see how many incidents have occurred over the years.

**Extraction**

In this project we extracted data from 2 websites but we were able to web scrape the data from one of the sites and the other site we were able to pull 3 datasets. The 3 datasets pulled included the following information:

* Accidental Death
* Accidental Injury
* Mass Shootings

And the data scraped and extracted was:

* Gun Laws

The following sources for our datasets used:

* <https://www.gunviolencearchive.org/accidental-deaths>
* <https://www.gunviolencearchive.org/accidental-injuries>
* <https://www.gunviolencearchive.org/mass-shooting>
* <https://en.wikipedia.org/wiki/Gun_laws_in_the_United_States_by_state>

**Transformation/User Case**

As said in our proposal a possible user case for this could be for people that are enforcing gun control laws, especially with the current environment of gun violence in the US. The completed database could be used to research gun violence data in their respective state in comparison to their states gun laws. The comparisons can show how useful the gun laws for the states are and how they affect the amount of gun incidents reported. In order to transform the public data and use it for our project/study the following process were established:

* Using the pandas in Jupyter Notebook to load up all three CSV files.
* Reviewing the files and transforming the data into dataframes
* Combining all 3 dataframes into 1 Dataframe
* Scraping data from the Wikipedia website into Jupyter Notebook
* Cleaning the data from the scraped websites to bring out readable dataframes
* Merging the data from the Scraped websites and the 3 combined csv files into one notebook
* Identifying and removing duplicates from the dataframes created.

**Load**

After pulling the csv files, scraping the websites and loading them all into data frames we did an initial connection to the Postgres database using PG admin. After running the queries and created the new tables with only the relevant information kept, we connected to the database and generated additional tables for the data fames.

**Summary**

As a group although we did work together, we each had our individual work allocated to ourselves and such was accomplished:

* Nwabueze Lucas Okose:
* Extracted CSV Files
* Clean and and arrange all 3 CSV Files
* Merging all 3 CSV Files into one single Data frame
* Shahanwaz Kahn:
* Web Scraped Wikipedia
* Clean and arrange scraped data from the website
* Brought out and simplified views on the laws from the dataset
* Kandace Herbert:
* Make an initial connection to Postgres database using PG admin
* Running queries and creating new tables to keep only relevant information
* Re connecting to the database and generating additional tables for the data frames
* Together as a group:
* Creating an ERD
* Combining scraped data and cleaned dataframes from the CSV files into one notebook