Kenny Daniels

Eric Messerich

**ETL Project Report**

The data sets used in our ETL analysis project were extracted from Kaggle.com. All three data sets; Firearms Provisions in US States, Video Game Sales, Homicide Reports, 1980-2014 were in the .csv format.

Each data set contained columns with data that we didn’t have interest in so pandas was utilized to drop off those not needed columns. The ‘year’ column was not consistent between the data sets so they were re-casted as “int64”.

The data was joined using MySQL utilizing the inner join. The final database utilized to load the data into was a non-relational database.

Data sets used:

* Firearms Provisions in US States (.csv)
  + Presence of 133 Provisions in US States, 1991-2017
  + <https://www.kaggle.com/jboysen/state-firearms#raw_data.csv>
* Video Game Sales (.csv)
  + Analyze sales data from more than 16,500 games
  + <https://www.kaggle.com/gregorut/videogamesales>
* Homicide Reports, 1980-2014 (.csv)
  + Can you develop an algorithm to detect serial killer activity?
  + <https://www.kaggle.com/murderaccountability/homicide-reports>