Harry Feldman

Raven Washington

Jessica Pardo

Andrey Tokarev

Team 2 ETL Project Proposal

We will be extracting data that pertains to the quality of life in different States of the USA. The purpose of extracting the data will be to create a database for potential future country-wide analysis of housing, healthcare, and other measure of quality of life. Because our data will come in the form of many separate tables (one for economic data, one for healthcare data, one for crime data, etc…), each containing a primary key (the state), we will use PostgreSQL, a relational database management system.

Our potential data sources are:

1) Zillow Economics Dataset<https://www.kaggle.com/zillow/zecon>  
2) Crime rate Dataset — US Crime rates by County  
<https://www.kaggle.com/mikejohnsonjr/united-states-crime-rates-by-county>  
3) School Dataset — USA Public Schools  
<https://www.kaggle.com/carlosaguayo/usa-public-schools>  
4) Income Level Dataset — USA Income Levels by United States Census Bureau <https://data.world/uscensusbureau/acs-2015-5-e-income/workspace/file?filename=USA_ZCTA.csv>  
5) Income Dataset by United States Census Bureau  
<https://www.census.gov/programs-surveys/acs/>  
6) Hospital Dataset  
<https://www.kaggle.com/carlosaguayo/usa-hospitals/version/1>  
7) Hospital Rating Dataset  
<https://www.kaggle.com/center-for-medicare-and-medicaid/hospital-ratings>  
8) Unemployment Dataset — USA Unemployment rate Dataset  
<https://www.kaggle.com/carlosaguayo/2018-unemployment-rate-by-county/version/1>