Kamran Abdullah

Ahmed Lotfy

Tunde Adebayo

**Submission Procedures Accident ETL**

In order to run the Etl\_Proj\_Final.ipynb, you will need to have a config.py file and secret.py file.

In the config.py file, run the line below, and populate the api\_key with you census api key

api\_key = ' '

To search the api manually, use the link below, and manually enter your key after “key=”:

**US census api**: [api.census.gov/data/2016/acs/acs1?get=NAME,B01001\_001E&for=county:\*&key=](http://api.census.gov/data/2016/acs/acs1?get=NAME,B01001_001E&for=state:*&key=)

In the secret.py file, populate it with the variables username and password for postgres

username= ‘ ‘

password= ‘ ‘

The files necessary for the notebook to run can be found inside the following links from kaggle and the official fhwa:

**US accidents 2016-2020:** <https://www.kaggle.com/sobhanmoosavi/us-accidents>

**Number of license per State** (offered by Federal Highway Administration and US department of Transportation) **:** <https://www.fhwa.dot.gov/policyinformation/statistics/2016/dl201.cfm>

**State vehicle registration** (offered by Federal Highway Administration and US department of Transportation)**:** <https://www.fhwa.dot.gov/policyinformation/statistics/2016/mv1.cfm>

**Data Model Diagram** can be foundunder the image folder

If you would prefer, or are unable to find the information, the following resources link to google drive holds the relevant information. Please use this to find the relevant resources to run the notebook. You should need 4 csv files to properly run the notebook:

DL.csv

RMV.csv

state\_csvData.csv

US\_Accidents\_June20.csv

**Note: The state\_csvData.csv can only be found in the google drive link.**

Google drive link: <https://drive.google.com/drive/folders/1FCA2Ch_vn2JQRQN3mbt7tA06yVu_vivT?usp=sharing>

**Note: Ignore the output folder in Google Drive. These are csv outputs from running the notebook, and are NOT necessary in order to run the notebook.**

The file Accident\_Schema\_Final.sql has to be run before loading. A note is made in the notebook.

A limitation on the time constraint kept us from doing more advanced data modeling. Thank you.