# Conclusions for Montgomery county Maryland

* Men get more traffic violations than women, 69% vs 59%.
* Black cars get more violations because there are more of them.
* Luxury vehicles have lower sales and higher prices.
* There is less of a correlation between affordable priced cars (Ford, Toyota, Honda) and their sales.

Steps for import

* Extract
  + Kaggle dataset for auto sales.
    - CSV file. Imported into a dataframe to read\_csv.
    - <https://www.kaggle.com/hsinha53/car-sales>
  + Government dataset for auto violations.
    - CSV file. Imported into a dataframe to read\_csv.
    - <https://catalog.data.gov/dataset/traffic-violations-56dda>
* Transform
  + Jupyter Notebook
    - Traffic Violations
      * Identified columns we needed.
      * Used count function and ‘groupby’ to count violations by Men, Women.
      * Filtered dataset by citations.
      * Rounded gender citation and converted gender citations decimals into percentage format.
    - Auto Sales
      * Identified columns we needed and separated details and price columns.
      * Used string strip function and the manufacturer and model columns to remove extra space.
      * Used ‘groupby’ function to group manufacturer and calculated average price and total sales for each manufacturer.
* Load
  + Converted the dataframes into JSON files, using the orient argument.
    - Used a different ‘orient’ argument for each dataframe.
  + Used PyMongo to connect to Mongo Client.
  + Inserted JSON files into the Mongo database using JSON.loads for each dataframe.
  + Confirmed the load and collections within Mongo.