**Final Project Problem Statement:**

Analyze and predict DC Metro area rental prices, given certain variables that I designate as pertinent for renters, from 2000-2017. Specific focus on professional class/millennials, analyze effects of certain features of rental properties and proximity to different amenities. Deliverable: to rank by geographic area the most expensive neighborhoods to rent for 2018.

**Data:**

Zillow Housing data, available here: <https://www.kaggle.com/zillow/zecon>

DC neighborhood/Ward definitions, available here: <http://opendata.dc.gov/datasets/neighborhood-labels-as-centroids>

**Hypothesis:**

To beat this hypothesis, as my “base”: <https://dc.curbed.com/2017/6/22/15854780/rent-most-expensive-neighborhoods-map>

**Variables:**

* Rental price
* Bedrooms
* Bathrooms
* Building size (# of floors)
* Distance away from closest Metro/Bus station
* Distance away from amenities
  + Groceries
  + Restaurants/Bars
  + Sports facilities
* Crime rate

**Research methodology:**

1. Gather and clean data relevant to DC Metro rental prices. Gather yearly rental prices for each type of rental property, and data on other variables as defined above.
2. Perform exploratory data analysis.
   1. Make color heat maps of average housing prices by property.
   2. Graph historical housing prices over time.
   3. Check for any anomalies/disruptions in data.
   4. Perform other necessary visualizations.
3. Perform baseline linear regression model.
4. Perform feature selection/tuning through lasso/ridge regressions.
5. Try variant modeling techniques, i.e. KNN, logistic regression, decision trees.
6. Present findings.