Home Prices in Gainesville, FL

# **Nick Miller DSC580 Summer 2020** [github.com/nickmiller1023](https://github.com/nickmiller1023)

# Which Domain?

The domain I am going to research is the housing market. I am looking to move soon, and would like to know what type of home I can afford. So, I will be looking at home prices in the area and comparing rooms, lot size, and square footage, and other features that will help determine the price of the home. Here are some sites for reference:

<https://www.kaggle.com/PythonforSASUsers/hpindex>

<https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>

<https://www.zillow.com/research/data/>

<https://www.kaggle.com/alphaepsilon/housing-prices-dataset>

<https://github.com/datasets/house-prices-us>

<https://www.census.gov/topics/housing/data/datasets.html>

<https://www.kaggle.com/apratim87/housingdata>

<https://catalog.data.gov/dataset/characteristics-of-new-housing>

<https://blog.quandl.com/api-for-housing-data>

<http://neupy.com/2015/07/04/boston_house_prices_dataset.html>

# Which Data?

I will be using this dataset: (notes are included at the link)

<https://data.census.gov/cedsci/table?q=housing&hidePreview=true&tid=ACSDP5Y2018.DP04&t=Housing&vintage=2018>

# Research Questions? Benefits? Why analyze these data?

I would like to analyze this data by first filtering out unneeded features. This data set is large enough to include a nationwide baseline, and hopefully a smaller subset for my county or city. It seems to include enough features as well.

# What Method?

I will be using linear regression to determine cost of houses in my area, and possibly some other predictive analytics, once I do some more research.

# Potential Issues?

Challenges I foresee may be not enough data in my city/county. If so, the state of Florida is too large of a market to determine accurate costs in my area. Plus, some homes are new construction, some are not. Some homes may have add-ons like pools or solar power systems that add value, but may not be noted on this dataset. And, some homes may be in areas with homeowner association fees, which may not be listed here, but will add monthly cost.

# Concluding Remarks

Housing prices may be a difficult thing to predict in just one area. But if the dataset is big enough, we should be able to use regression to find out how much houses cost with certain attributes, such as bedrooms, bathrooms, square footage, lot size, and age of home. As I may be moving in a few months, this is a project I am invested in, and will hopefully come up with some information that I can use when the time comes to sell my current home and purchase a new one.