Passion Project Summary

I) Design

The project sought to determine whether or not the announcement of a major corporate expansion would have an effect on the change in housing prices in the geographic region that they expanded in after the expansion was announced.

The data used came from four major sources. The data on announced corporate expansions (and contractions) came from the New York State Department of Labor and the Virginia Economic Development Partnership. The data on housing prices came from Zillow. The data on the change in employment and income came from the Bureau of Labor Statistics, part of the US Department of Labor. The data on demographics came from the American Community Survey, part of the US Department of Commerce.

The time period chosen for the analysis was from 2010 onwards. The reason for the selection of this time period is that 2010 was the beginning of the period of economic expansion that followed the Great Recession that started in 2008. The geographic unit used was the county level because the data on corporate expansion was provided by state governments at that level and because of the abundance of relevant economic and demographic data present at that level.

II) Tools

A) Python — Programming Language
    1) Pandas (Feature Engineering)
    2) Numpy (Feature Engineering)
    3) Scikit-learn (Modeling)
        a) Random Forest
        b) Gradient Boosting
        c) Linear Regression
    4) Treeinterpreter (Data Analysis)
    5) Matplotlib (Graphic Design)
    6) Flask (Interactive Data Explorer)
B) Google Slides — Presentation

III) Data

FEATURE	**CONTEXT**	**SOURCE**
Y: future_price_change	The percentage change in median house price one year in the future	Zillow
X_0: new	The number of jobs created by corporate expansion at new site	VA Economic Development Partnership, NY Dept of Labor
X_1: expansion	The number of jobs created by corporate expansion at existing site	VA Economic Development Partnership, NY Dept of Labor

X_2: annual_new	The average number of jobs created per month by corporate expansion at new site over the prior year	VA Economic Development Partnership, NY Dept of Labor
X_3: annual_existing	The average number of jobs created per month by corporate expansion at existing site over the prior year	VA Economic Development Partnership, NY Dept of Labor
X_4: prev_price_change	The percentage change in median house price from one year prior	Zillow
X_5: price	The median price of a house in a county in that month	Zillow
X_6: unemp_rate	The unemployment rate in a county in that month	Bureau of Labor Statistics
X_7: annual_unemp_change	The annual change in the unemployment rate in a county	Bureau of Labor Statistics
X_8: foreclosure_rate	The foreclosure rate in a county	Zillow
X_9: inven	The housing inventory level in a county	Zillow
X_{10}: inven_change	The percentage change in housing inventory level in a county from the year prior	Zillow
X_{11}: inven_popul	The ratio of housing inventory and residents of a county	Zillow & US Census Bureau
X_{12}: annual_popul_change	The percentage change in the number of residents in a county from the prior year	US Census Bureau
X_{13}: inc_pc	The level of income per capita in a county	Bureau of Labor Statistics
X_{14}: inc_pc_change	The percentage change in the level of income per capita in a county from the prior year	Bureau of Labor Statistics
X_{15}: emp_popul	The ratio of employers and residents in a county	Bureau of Labor Statistics & US Census Bureau
X_{16}: emp_change	The percentage change in the number of employers in a county from the prior year	Bureau of Labor Statistics
X_{17}: inf_change	The percentage change in the inflation rate from the prior year	

IV) Results

The model that provided the best results was a Random Forest Regression model, which returns the consensus predicted value of the target variable of a group of Regression Decision Trees. The model explains 64% of the overall variance in the change in home prices. It was found that, on average corporate expansion increased home prices by 0.4% while other factors increased home prices by 1.3%. The range of the impact of corporate expansion on home prices was between -0.9% and + 5.0%.

V) Next Steps

To continue the analysis, there are several steps that I plan on taking. One of the steps I plan on taking is to search for data at the ZIP Code level and run the model again to search for a more granular effect. Another step that I plan on taking is to broaden the geographic scope of the analysis to search for states other than New York and Virginia for the data. An additional step that I plan on taking is to further refine the effect of government policies such as school funding and property taxes on the median house price to further ameliorate the overall accuracy of the model.