**Presentation:**

**Selected Topic:**

For our final project, our group will be creating a machine learning model that will predict the next up-and-coming cities in the US using historical housing market data from 2012 to 2021.

**Reason why they selected their topic:**

The reason the team selected the topic is due to an inherent passion derived from living through current events. According to the Pew Research Center (https://www.pewresearch.org/social-trends/2020/01/09/trends-in-income-and-wealth-inequality/#:~:text=From%202015%20to%202018%2C%20the,era%20of%20the%20late%201990s.), “From 2000 to 2018, the growth in household income slowed to an annual average rate of only 0.3%. If there had been no such slowdown and incomes had continued to increase in this century at the same rate as from 1970 to 2000, the current median U.S. household income would be about $87,000, considerably higher than its actual level of $74,600.” Meanwhile, a CNBC 2017 report shows housing prices outpacing inflation (<https://www.cnbc.com/2017/06/23/how-much-housing-prices-have-risen-since-1940.html>). Thus, income has remained relatively stagnant but housing prices continue to increase at astronomical rates. For members of our team who are current home owners or even individuals looking to buy a home, we are living through this era where homes are quite factually unaffordable based on areas. For our team members living in the DMV area that fact is all too real.

This project is a means to shed light on such hardships currently faced and to potentially show an alternative path. Is there a city out there with an affordable home that will eventually be the next San Diego or Washington DC? Can we get it at an investment level cost? Can we live out the American Dream of one day owning an affordable home? These are some of the questions we hope to answer. Not just for this project but also for ourselves.

Description of their Source Data/Questions they hope to answer with the data:

Our team found redfin housing market information via Kaggle.com (<https://www.kaggle.com/thuynyle/redfin-housing-market-data?select=state_market_tracker.tsv000>). Utilizing data cleansing the team is analyzing the data based on factors such as region by month, region by year, median sale price versus median list price, the amount of homes sold vs new listings, the median amount of days on the market before a home is bought, and the amount of times a home is sold above list price. These are some data points that can help answer some of the questions above based on numerous hypotheses. Some of the following hypotheses and related data points are listed below:

1. Does the difference between the median sale price and the median list price appear to grow in cities that are up and coming? In other words, if we analyze the median sale price – median list price should we start to see exponential growth in cities that are about to become big?
2. Do the median number of days a home is on the market tend to indicate an increase in demand? In other words, if the homes are being sold faster once listed, does that mean more people are trying to get in on the market? If so, does that mean that the city is already a big city?
3. Do the number of homes sold mean that the properties are in more demand or that the people living there are trying harder to leave?

Finally, based on the above hypotheses what other data points could we potentially incorporate? Should we incorporate census data to see if there’s an inflow and outflow of people from one state to another? Should we see if there’s just a surge of wealthy individuals that happen to move into an area causing a drastic increase in sale prices? At the moment this is still something we are working on.