DS - 670 –Capstione Big Data & Business Analytics

Mir Ahmed, Kaushik Raj Devarashetty, Madhuri koganti

Project Proposal: Predicting House Prices and Identifying Ideal Locations in NJ

Saint Peter’s University | 2641 John F. Kennedy Blvd, Jersey City, NJ 07306

2023

# Project Proposal: Predicting House Prices and Identifying Ideal Locations in NJ

# Introduction

In this project, we aim to build two types of predictive models using various machine learning algorithms to predict the house price for purchase and identify ideal locations for purchase or rent in New Jersey. The project will use publicly available datasets from various sources such as Zillow, Freddie Mac, NJ Department of Education, and the Census Bureau. The project's primary goal is to help real estate investors and buyers as well as new residents and potential movers make informed decisions based on the models’ predictions.

# Project Group

Kaushik Raj Devarashetty

Madhuri Koganti

Mir Ahmed

# Problem Statement

Moving to a new state can be a challenging task, especially when trying to determine the ideal location to live. People who are planning to move to New Jersey may not have enough knowledge about the state and its counties, which can make it difficult to determine where to live. Additionally, there are various factors to consider, such as the cost of living, crime rate, schools, and accessibility to amenities, which can be overwhelming to evaluate.

To tackle this challenge, we aim to develop a recommendation system that will assist people who are moving to NJ from other states in finding the appropriate county to rent or own a home using predictive modeling. The system will collect and analyze data from various sources such as real estate listings, census data, and public records to identify key factors that influence the quality of life in each county. We will use machine learning algorithms to predict the best counties that match an individual's preferences and needs, such as number of bedrooms, quality of schools, crime rate, and cost of living.

Overall, our goal is to provide an easy-to-use and accurate recommendation system that will help people who are relocating to NJ find the ideal county to live based on their personal preferences and needs.

# Data Sources

## Zillow Home Value Index Data:

The dataset provides information on the Zillow Home Value Index by county, city, or ZIP code, which would be used as the target variable for building regression and/or classification models.

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

## Mortgage Value Data:

This dataset provides information on the average mortgage rates for different regions in NJ, which could be used as a predictor variable in the regression and/or classification models.

[http://www.freddiemac.com/pmms/#](http://www.freddiemac.com/pmms/)

## Tax Rate Data:

This dataset provides information on the tax rates for different regions in NJ, which could be used as a predictor variable in the regression and/or classification models.

<https://www.state.nj.us/treasury/taxation/lpt/statdata.shtml>

## County Lines and Shape Data (GeoJSON):

This dataset provides geographic boundaries for different counties in NJ and could be used to visualize location-based features.

<http://data.ci.newark.nj.us/dataset/new-jersey-counties-polygon/resource/95db8cad-3a8c-41a4-b8b1-4991990f07f3>

## NJ Department of Education Data for School Performance:

This dataset provides information on school ratings for different regions in NJ, which could be used as a predictor variable in the classification and/or regression models.

<https://www.schooldigger.com/go/NJ/schoolrank.aspx>

## Crime Data:

These datasets provide information on crime rates for different regions in NJ, which could be used as a predictor variable in the classification and/or regression models.

For 2017: <https://ucr.fbi.gov/crime-in-the-u.s>

For 2018,2019,2020: <https://www.njsp.org/ucr/current-crime-data.shtml>

## Poverty and Median Income Data:

This dataset provides information on poverty and median income rates for different regions in NJ, which could be used as predictor variables in both the regression and classification models.

<https://www.census.gov/programs-surveys/saipe/data/api.html>

## NJ Population History by County

These datasets will give us historical population counts (estimated) from 2010 till 2021.

<https://www.nj.gov/labor/labormarketinformation/demographics/population-household-estimates/>

## NJ Municipalities by County

These datasets will give us all municipalities by county for NJ.

<https://data.nj.gov/Reference-Data/Municipalities-of-New-Jersey/k9xb-zgh4>

## NJ Food Desert – USDA Foot Atlas

This dataset will provide us flags for low access areas (food deserts) by county, which could be used as a predictor variable in the classification and/or regression models.

<https://www.ers.usda.gov/data-products/food-access-research-atlas/>

## NJ Area Deprivation Index – Neighborhood Atlas

The Area Deprivation Index (ADI) is based on a measure created by the Health Resources & Services Administration (HRSA) over three decades ago, and has since been refined, adapted, and validated to the Census Block Group neighborhood level by Amy Kind, MD, PhD and her research team at the University of Wisconsin-Madison. It allows for rankings of neighborhoods by socioeconomic disadvantage in a region of interest (e.g., at the state or national level). It includes factors for the theoretical domains of income, education, employment, and housing quality. It can be used to inform health delivery and policy, especially for the most disadvantaged neighborhood groups. "Neighborhood" is defined as a Census Block Group.

<https://www.neighborhoodatlas.medicine.wisc.edu/download>

# Methodology

## Regression Model:

For the regression model, we will use the Zillow Home Value Index as the target variable and use different algorithms such as Linear Regression, Random Forest, K-Nearest Neighbor, and Support Vector Machines to predict the house prices for different regions in NJ. The regression model will use predictor variables such as mortgage rates, tax rates, poverty rates, and median income rates. We will use the Root Mean Squared Error (RMSE) to evaluate the performance of the regression models.

## Classification Model:

For the classification model, we will use various algorithms such as Logistic Regression, Random Forest, K-Nearest Neighbor, Support Vector Machines, and Neural Networks to identify ideal locations for purchase or rent in NJ. The classification model will use predictor variables such as school performance metrics, crime rates, poverty rates, and median income rates. We will use accuracy, precision, recall, and F1-score to evaluate the performance of the classification models.

# Expected Outcome

The expected outcome of this project is to develop an app with most possible accurate predictive models for real estate investment/ rentals in New Jersey. These models will provide insights and recommendations to potential investors/ new residents to make informed decisions about property investments/ rentals in the state. This project will also provide an opportunity to understand the impact of various factors such as school performance, crime rates, and poverty levels on housing prices and rental values in New Jersey.

# Conclusion

In this project, we aim to build an app with two types of predictive models using various machine learning algorithms to predict house prices and identify ideal locations for purchase or rent in NJ. The project will use publicly available datasets from various sources such as Zillow, Freddie Mac, NJ Department of Education, and the Census Bureau etc. . The project's primary goal is to help real estate investors and buyers make informed decisions based on the predicted house price and ideal locations for purchase or rent in NJ.

**References**

“Moving to New Jersey - A Complete Guide 2023.” *Movingist*, 2023, Retrieved March 10, 2023. <https://movingist.com/moving-to-new-jersey/>

“Moving to New Jersey.” *Moving to New Jersey - A Complete Relocation Guide 2023*, Retrieved March 10, 2023. <https://www.moverjunction.com/moving-to-new-jersey>