# Group Members

Nicholas Stout, Mason Kerns-Harper, Paribesh Chaudhary, Edoardo Martinelli

# Project title

Price is what you pay. Value is what you get.

We chose this title because it is short, sweet, and to the point. We feel it explains our project without giving away too much detail. It leaves the audience wanting more.

# Project pitch

We want to analyze and predict housing prices in the U.S by using a random forest model. These predictions will be based on several factors like crime rate, local assessment of education, risk of diseases, unemployment rate, and other factors that may occur throughout our project. We will be presenting our findings with plots, graphs, and an interactive map created from R Shiny.

# Names and general skills

Nicholas Stout: R coding for data visualization, Shiny for interactive maps and plots, and video making and editing for the final video presentation. I’ll be helping out with the implement ArcGIS via Jupyter notebooks.

Mason Kerns-Harper: Proficient in Python, SQL, and Pandas. We will be using these skills for predictive modeling based on the cleaned datasets.

Edoardo Martinelli: Python, pandas, and R. I will be using these languages to clean and aggregate datasets. Finally, I will be helping out with the predictive coding.

Paribesh Chaudhary: R, SQL, Python, and ArcGIS for making maps and analyze data.

# Data Ethics

We are collecting housing market data from DataPlanet. Yes, this dataset was obtained ethically through the University of Arizona Library database. Possibly the outcome of our research project could potentially hurt someone if used for criminal intent. We want to stress that our research project is not meant for unlawful purposes but rather for the user to get a much better user experience searching for housing market prices than they would if they use websites like Zillow.

# Literature/Market Review

**Peer Review Article:**

In this peer review article, we can see that the authors are looking at a specific location, trying to find the turning points for the housing market within their dataset. Along with finding the turning points in their dataset, they use various algorithms to predict if the turning points had a significant factor on the housing markets in Wuhan, China. The main thing that sets our capstone about from the peer review article is that we are not set to a specific location and that we are not just trying to predict housing prices overall for that particular location but show and predict housing prices for any area that the user might be interested in that we are displaying through our shinny interactive models.

**Source MLA format along with the link to the website:**

Dong, Shihai, et al. “Predicting the Turning Points of Housing Prices by Combining the Financial Model with Genetic Algorithm.” *PLOS ONE*, Public Library of Science, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0232478.

**Link:**

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0232478>

**Peer Review Article:**

The authors are developing what they perceive to be an accurate prediction model for housing prices for a specific location. Throughout their process, they take a few factors into account, such as age, sale price of the house, total area, and zip code, to name a few of the factors they considered. Some of the algorithms they used to predict the housing market are regression model, prediction model performance, coefficient of determination, mean of square error, mean of absolute error, and computational time, and a few others. Our findings will be from different years rather than just from 2015 to 2018, as they did in their finding. Two of our results will be put into interactive plots, charts, graphs, maps, and the user will be able to tune into some regions of the city, for example, a specific zip code or county.

**Source MLA format along with the link to the website:**

Jha, Shashi Bhushan & Babiceanu, Radu & Pandey, Vijay & Jha, Rajesh Kumar. (2020). Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study.

**Link:** <https://www.researchgate.net/publication/342302491_Housing_Market_Prediction_Problem_using_Different_Machine_Learning_Algorithms_A_Case_Study>

**Peer Review Article:**

In the article, researchers test whether neighborhoods with good accessibility, more attractive, slow higher housing prices and if these factors have any role. As for the result of the research, the authors performed the work and calculations using the GeoDa software, and they compared results from two different models, the OLS model, and a spatial lag model. There’s such as room size, number of rooms, and housing price were involved in predicting housing price.

**Source MLA format along with the link to the website:**

Wittowsky Dirk, Hoekveld Josje, Welsch Janina & Steier Michael. (2020). Residential housing prices: impact of housing characteristics, accessibility and neighboring apartments - a case study of Dortmund, Germany.

**Link:** <https://www.tandfonline.com/doi/citedby/10.1080/21650020.2019.1704429?scroll=top&needAccess=true>

**Peer Review Article:**

It was found that the impact of crime hotspots on the housing market depended on the type of home, whether it was a condo, flat, or single-family home. The article uses a hedonic approach which factors in the ‘negative externality of a crime hot spot, and its societal cost.’ not only the impact that crime has on a community which will bring down its value, but the impact that it has to everyone involved, including stakeholders, urban planners, housing developers, and safety experts.

**Source MLA format along with the link to the website:**

Ceccato, Vania & Wilhelmsson, Mats. (2019). Do crime hot spots affect housing prices? Do crime hot spots affect housing prices?. Nordic Journal of Criminology. 21. 10.1080/2578983X.2019.1662595.

**Link:** <https://www.researchgate.net/publication/335773241_Do_crime_hot_spots_affect_housing_prices_Do_crime_hot_spots_affect_housing_prices>

**Peer-Reviewed Article:**

This article goes over how the housing market and housing prices react to differences in the quality of education. Compared to a service like Zillow, this is another aspect that dives into these factors’ impact. The article mentions that the ‘standard deviation in the increase in public school performance raises housing prices by 1.4 to 2.4%’. While Zillow may mention the schools nearby, it doesn’t mention the quality nor impact of the surrounding education’s quality.

**Source MLA format along with the link to the website:**

Fack, G., & Grenet, J. (2010). When do better schools raise housing Prices? Evidence from Paris public and private schools. *Journal of Public Economics*, *94*(1-2), 59–77. https://doi.org/10.1016/j.jpubeco.2009.10.009

**Link:**

<https://www.sciencedirect.com/science/article/abs/pii/S0047272709001388>

# Timeline

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | Nicholas | Mason | Paribesh | Edoardo |
| 5 | Set up a GitHub repository and a license for the group to be able to use and share files and code with one another. | Research the best possible analysis for the project. Including Random Forest Regression along with Keras for Python | Work in ArcGIS and functions to create and display the aspect and slope of a surface. (Map) | Searching for more predictive factors to use for our python scripts. |
| 6 | Determining what form of correlation analysis should be run, and how to run it in R | Determining and compiling a comparing dataset to ensure our predictability will be valid. | Setting up ArcGIS so it’s ready for base map selection (vector/raster) map. | Cleaning and aggregating house market data using python. |
| 7 | Data wrangling within R | Begin building the random forest regression model for use with data in python. | Work on ArcGIS to implement trails, roads, parks, and more. | Using predictive models to check which factors impact the housing market. |
| 8 | Once the data is wrangled, I will run our analysis test in a Rmarkdown file | Continue building out random forest regression model and testing it by using our dataset and the comparing dataset from Zillow | Using housing dataset variables such as county in ArcGIS map. | Researching and learning about Keras neural network. |
| 9 | Convert the Markdown to a Shiny Web APP | Begin building out Keras with Edoardo | Using housing dataset variables such as county in ArcGIS map. | Helping and implementing Keras neural network. |
| 10 | Improve upon the aspect of how the data looks within Shiny APP | Continue building Keras with Edoardo | Make relation between housing dataset and ArcGIS and move towards Shiny. | Continue working with Keras and debugging. |
| 11 | Work with group members to create a Shiny website based upon the already created Shiny APP, along with ArcGIS in shiny | Testing and comparing Keras vs Random Forest Regression Model | Work with Nicholas on implementing ArcGIS into Shiny via RStudio | If scripts work properly. Will be adding more states to our research project. |
| 12 | Add the filter functions to our Shiny | Implement python scripts with Shiny | Work with Nicholas on the Shiny app and check the dataset within the app. | Will be helping with the final touches with the Shiny graphs. |
| 13 | Make sure our GitHub has all the necessary documents and has a licensing along with a README. Make sure GitHub is up to standards. | Final touches on the predictive modeling | Checking if the project runs as it was designed and helps users with the housing market. | Start to document and organize all the scripts, graphs, and research. |
| 14 | Help with the video presentation of the final project delivery. | Finishing up the presentation for the final delivery of the project | Helping with the final presentation and finishing touches. | Going to join and help in the recording and editing of the video presentation. |

# Final Deliverable Product Specifications

The final deliverable product will be a presentation where each group member talks about what they did, and as a group, we all present our final shiny product with our code to the graders. Each member will talk about how their code was used and implemented into the interactive modules and filters via shiny.

## Final Presentation Format

The final presentation format will be delivered via shiny, where we will have interactive maps. These graphs will show off our data that allow users to implement any filter they want to help look at housing market prices for a specific area.

## What Analysis Is Being Run?

We will be using a random forest regression model and Keras as our predictive modeling. Using the two offers a comparison to determine the best model for predicting house prices based on the variables in the datasets.

## What if the Analysis doesn't work?

Our mentor Dylan concluded that our random forest analysis would work for predicting and representing in shiny.