**Serenity Homes**

C964 Computer Science Capstone

Matthew Leyden

Western Governor’s University

Matthew Leyden

04/19/2021

[Version 1.0]

Contents

[Contents 2](#_Toc70163775)

[A.1 Letter of transmittal 3](#_Toc70163776)

[A.2 Project Recommendation 4](#_Toc70163777)

[A2.1 Problem Summary 4](#_Toc70163778)

[A2.2. Application Benefits 4](#_Toc70163779)

[A2.3 Application description 4](#_Toc70163780)

[A2.4 Data description 5](#_Toc70163781)

[A2.5 Objective and Hypotheses 5](#_Toc70163782)

[A2.6 Methodology 5](#_Toc70163783)

[A2.7 Funding requirements 6](#_Toc70163784)

[A2.8 Stakeholder impact 6](#_Toc70163785)

[A2.9 data precautions 6](#_Toc70163786)

[A2.10 DEVELOPERs expertise 6](#_Toc70163787)

[B. Project proposal 7](#_Toc70163788)

[B.1 Problem statement 7](#_Toc70163789)

[B.2 customer summary 7](#_Toc70163790)

[B.3 Existing System analysis 7](#_Toc70163791)

[B.4 Data 8](#_Toc70163792)

[B.5 Project methodology 9](#_Toc70163793)

[B.6 Project outcomes 11](#_Toc70163794)

[B.7 Implementation Plan 11](#_Toc70163795)

[B.8 evaluation plan 12](#_Toc70163796)

[B.9 Resources and costs 12](#_Toc70163797)

[B.10 Timeline and milestones 12](#_Toc70163798)

[D. Post Implementation Report 13](#_Toc70163799)

[D.1 Project Purpose 13](#_Toc70163800)

[D.2 Datasets 14](#_Toc70163801)

[D.3 Data product code 15](#_Toc70163802)

[D.4 Hypothesis Verification 17](#_Toc70163803)

[D.5 Effective Visualizations and reporting 17](#_Toc70163804)

[D.6 Accuracy Analysis 19](#_Toc70163805)

[D.7 Application Testing 20](#_Toc70163806)

[D.8 Application Files 20](#_Toc70163807)

[D.9 USER’s guide 21](#_Toc70163808)

[D.10 Summation of Learning Experience 21](#_Toc70163809)

# A.1 Letter of transmittal

April 19, 2021

Chanel Steiner, Owner

Serenity Homes LLC

14192 Keyler St.

Seattle, WA 98126

Dear Mrs. Steiner,

For the past year, Serenity Homes has been the #1 real estate company in the greater Seattle area. Serenity Homes is rapidly expanding in addition to the more than 10,000 homes already in their inventory. However, with increased growth comes increased costs. The average salary of a residential home estimator is 60,000/year, not to mention travel expenses. As more homes will need to be evaluated, more home estimators will need to be hired. What if you could cut out the middleman?

Our team at DP-Tech would like to suggest a complete overhaul of Serenity Home’s current system of home price estimation. Combining machine learning and a sophisticated graphical user interface, we can create an easy-to-use application that will estimate the home price for you! This product will benefit Serenity homes by saving hundreds of thousands of dollars each year that are currently spent on trained personnel. In addition, any employee of the organization with the proper credentials can get a quick and accurate home price estimate within seconds. This will allow Serenity Homes to get a home on the market more quickly.

The objectives of this project are to provide Serenity Homes with an all-inclusive stand-alone application that will allow representatives to get an accurate home price estimate. With just a few questions about the home’s features (like number of bedrooms, bathrooms, floors, etc.), a precise quote can be generated. In addition, and perhaps even more importantly, Serenity Homes will gain knowledge from the applications data visualization features, empowering them to make informed decisions in the future.

The project has a total of four phases: Developing a model that can generate accurate home price predictions, designing a user interface, testing, and deployment. As far as financing goes, we project the entire application from start to finish will cost a mere $13,500, most of that cost being human labor. This price is almost nothing compared to what you would need to pay professionals to do the same work. Our programmers have expertise in developing software solutions for business. They are well versed in creating effective machine learning models and they can incorporate data science into your business saving you time and money.

Look forward to hearing from you soon!

Sincerely,

Matthew Leyden, CEO DP-Tech

# A.2 Project Recommendation

# A2.1 Problem Summary

The problem is that Serenity Homes is a growing company that needs a more effective and cost-efficient method to set a baseline price for their new and used homes. Home price estimators are an expensive cost to any realty company. This incurred cost can be nearly eliminated with an application that can automatically suggest the best price for a home. Representatives of Serenity Homes will be able to price their homes accurately and immediately. Realtors can assist home sellers by informing them how much their home is worth on the spot, without having a professional take the time to do a home walkthrough. They can also assist home buyers by informing them about what features of a home they can afford.

# A2.2. Application Benefits

As mentioned above, the primary benefit of this application would be to save money by not having to employ Home Residential Estimators. Other benefits include an easy-to-use interface, the ability to view all homes currently in the inventory along with pictures, and the convenience of being able to access the application from any laptop. This would be especially helpful and productive for realtors out in the field. Another very informative benefit would be the data visualization features. By exploring the data, representatives of Serenity Homes can gain insights, viewing all of their data through colorful visualizations. Most importantly, this can assist them in developing their business intelligence. Business Intelligence (BI) refers to the procedural and technical infrastructure that collects, stores, and analyzes data produced by a company’s activities. This will greatly support Serenity Home’s decision-making process.

# A2.3 Application description

The application we are proposing is a standalone application that can be run on any computer. It will feature a GUI developed with Kivy, a cross platform program library and written with the Python programming language. Because Kivy is cross-platform, employees will be able to use it on their specific computer, whether it be Windows, Mac, or Linux. The application will feature proper security controls (login screen) and a main screen that provides users the options of viewing homes currently in the inventory, exploring the data, or generating a price quote. If the user clicks the Properties button, they will be able to view Serenity Home’s properties along with pictures and details. If the user clicks the explore the data button, they will be able to view the data through different filters in order to gain a better understanding of the data. These will be processed using K-Means clustering and Matplotlib which will visually explain different features of the data and how they relate. The price quote feature will bring users to a page that will ask about some basic features of the home. When they press the generate price button, a saved classifier model will be queried and a suggested price will automatically be displayed. Users will also have the choice to retrain a new model if they wish and the accuracy of the new model will be displayed. We would encourage Serenity Homes to add to this dataset with information about their previously sold properties. As more homes are added, the model will become even more accurate.

# A2.4 Data description

The dataset used for this application contains over 20,000 entries. The data includes real home sales in the city of Seattle, the same city that Serenity Homes is currently based, and therefore should reflect realistic home prices specific for that city. It is in csv format and separated by commas. Each entry contains information about one home. The data types include real numbers and float values. For the purposes of this project, the dependent variable would be the price and the independent variables would include all of the other column headings except identification number and date, as those values are not necessary in determining the price of the home. There are very few anomalies in the data and it is pretty straight forward. The dataset is not limited in size but more data would likely lead to higher accuracy.

# A2.5 Objective and Hypotheses

The objective of this project is to provide a machine learning program that can give realistic home prices based on the most important features of a home. Access to the machine learning model should be through a user-friendly application that is as helpful and understandable as possible.

The hypothesis is that we can provide Serenity homes with home price suggestions that will assist them in determining an appropriate value for homes that they have acquired and are looking to sell. We believe the features of this application will allow Serenity homes to save money and put the house on the market as fast as possible.

# A2.6 Methodology

We at DP-Tech are big believers in the Agile Development Methodology. We employ this methodology for all of our projects. Agile is a type of iterative software development that emphasizes developing solutions through continuous testing and feature integration. The project will be broken down into phases called sprints. At the end of each sprint, we will carefully test the software to verify the goals of the sprint were achieved.

1. Preprocessing of data, training of model, and testing for accuracy.

* Train model and observe results.
* Test accuracy and stability of model.

1. Development of user interface.
2. Beta/acceptance testing with primary users.
3. Deployment

* Use the model to create home price suggestions/predictions
* Set up conditions for consistent monitoring and management

# A2.7 Funding requirements

We do our best to keep funding requirements as low as possible. Most of the data and libraries used to develop the application are open-source and therefore do not involve extra fees. As previously stated in our letter of transmittal, the total cost of the project is estimated to be $13,500. Our software engineers are paid an hourly rate of $150. A cost breakdown for the projects four distinct phases is below.

1. Phase 1: estimated at $3000 for 20 hours.
2. Phase 2: estimated at $6000 for 40 hours
3. Phase 3: estimated at $3000 for 20 hours
4. Phase 4: estimated at $1500 for 10 hours

# A2.8 Stakeholder impact

The success of this project will depend on stakeholder acceptance and approval. Our stakeholders support this project because the application of data science to business is steadily growing and has proven to be quite lucrative for all parties involved. The adoption of this project by Serenity Home’s stakeholders is suggested as success will lead to increased quarterly profits and faster home resales. In addition, Serenity Homes will gain a better understanding of what features increase the home’s value and where they should focus their efforts in the future. The application must adhere to the stakeholder’s ethical guidelines and privacy policies. All stakeholders will be informed of the project’s status before and after each phase. The next phase will only continue after stakeholder acceptance.

# A2.9 data precautions

The data from the initial dataset used to train the model is publicly available and does not contain sensitive material. It includes features of the specific homes sold, including date and price. It does not include information about the buyers or exact addresses of the properties. We would encourage Serenity Homes to add their own data in order to increase the datasets size and refine the model for highest possible accuracy. If they choose to do so, the data will contain proprietary and sensitive material. Appropriate login and security controls are present. The data doesn’t involve any human element (per HIPPA and FERPA regulations).

# A2.10 DEVELOPERs expertise

The developers of this application are college graduates from Western Governor’s University, a prestigious university. They have been extensively trained in the python and java programming languages. Some of their projects include the creation of complex inventory systems, disaster robot simulations, the application greedy algorithms and hash maps to fictional delivery systems, and working with databases. Through their work they have demonstrated professionalism, selection of appropriate algorithms for specific problems, and attention to detail. They will be able to meet the challenges of this application and create a suitable software product.

# B. Project proposal

# B.1 Problem statement

One of the difficulties for any real estate business is determining the market value for a home. Market value is what an informed and willing buyer will pay for a home. Many companies will use rough estimates of what a house is worth based on manual analysis of past sales of comparable houses. These are based on numerous factors including size, age, location, and features including comparable homes with similar amenities and level of finishes and updates.

Most real estate companies will hire residential home estimators to perform this work. These are professionals that are knowledgeable about residential construction and how to determine the price of a home. Not only do these professionals carry a substantial yearly cost to a company, they may take weeks to consider all of the factors before calculating a final estimate. The employer would likely accrue travel expenses as well. In addition, human error is always a possibility.

Our proposed solution is to develop an application that uses machine learning to determine a home’s market value. This would instantly analyze past trends to develop a reliable figure. The application would evaluate the most important elements of a home’s value such as square footage, exact location, the number of bedrooms and bathrooms, whether the home is by the water, etc. Using this application, a company will be able to determine with a high degree of accuracy exactly how much a home is worth. Because the data used to train a model is based on previous real sales, this generated price is likely to be attractive to buyers. The application will be informative to any real estate company and will save on the additional expenses that would be required to employ professionals to do the same work. It would also ensure the house gets on the market as soon as possible which means faster revenue generation.

# B.2 customer summary

This application will be used by representatives of the real estate company. This could include senior management all the way down to individual realtors. Senior management may find it very useful to examine the data visualizations of past sales trends. They can use this information to decide where properties have higher listings in order to determine where they should focus their efforts in home acquisition. The realtors of the company may find it very useful to see which attributes contribute to a properties price. By making individual adjustments to any of the features, they can view in real time what factors influence the sales price and by how much. They can also share this information with prospective buyers so that they can become aware of what they can afford. Realtors will be able to use this application from their laptops and it will work with any operating system. No special skill sets are required.

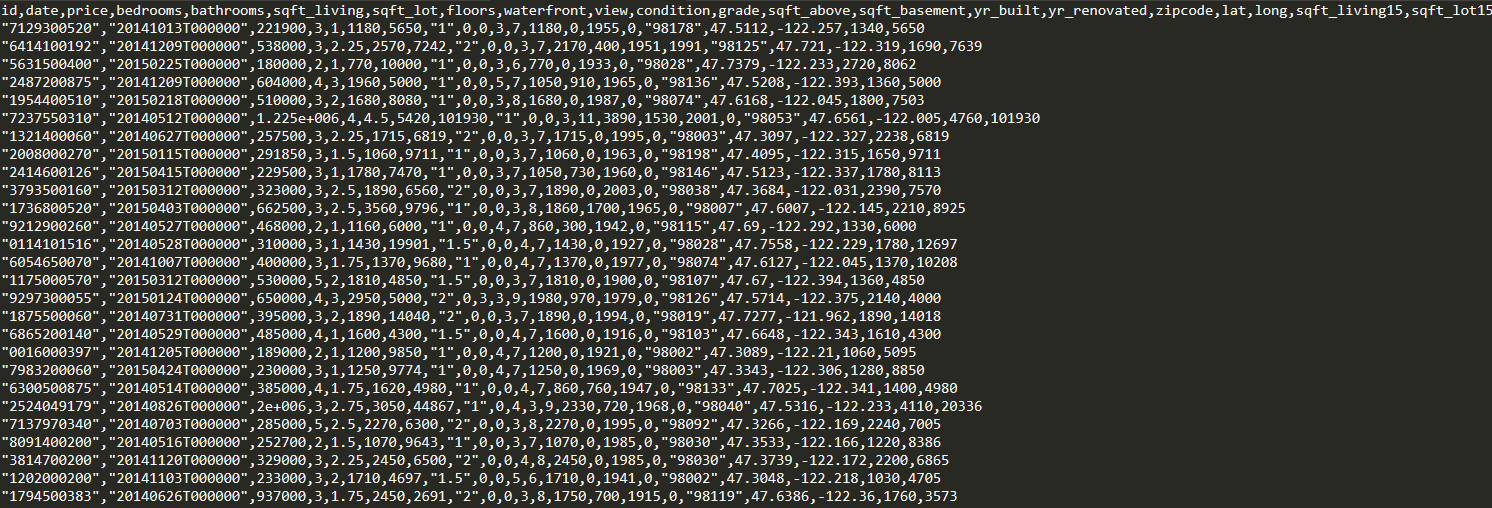
# B.3 Existing System analysis

The target platform can be any of the major operating systems including Windows, Mac, and Linux. The application uses Python and Kivy, significantly reducing the cost of overhead. The client would only need access to a computer to run the application. We can easily customize the look and feel of the program according to the client’s wishes. Because Kivy is cross-platform, the application can be easily adapted to work as a mobile application as well if the client chooses to go this direction in the future.

# B.4 Data

The data is in CSV format. We will use the pandas library to parse and modify the dataset. Pandas is a very useful tool in machine learning and is used for data cleaning and analysis. It has many features which are used for exploring, cleaning, transforming, and visualizing data. The original dataset can be accessed from <https://www.kaggle.com/swatchiachath/kc-housesales-data>. This dataset includes information about 21,613 home sales in the Seattle area. The column headings include an identification number, date the home was sold, the price of the home, number of bedrooms, number of bathrooms, floors, the square footage of interior living space, the square footage of the land space, whether or not it is overlooking the waterfront, whether or not there is a view and how good that view is, the grade (an index value that represents the level of design of the home), the square feet above ground level, the square feet below ground level, the year the home was built, the year it was renovated (if applicable), the zip code, latitude, longitude, and the interior and exterior living space of the nearest 15 neighbors. In order to simplify the data while still maintaining accuracy, we will remove some of the column headings that would be considered extraneous. We will remove latitude and longitude and keep zip code as the primary location factor. Also, because the application is meant to be as easy to use as possible and the exact square footage of the nearest 15 other neighbors (known as comparable houses) is not immediately known without thorough research, we will remove these from the data as well. We believe that we will still achieve a high degree of accuracy from our model without these variables. When the application is launched, the client will have the opportunity to add their own data in order to compensate for any loss in accuracy because the model will become more refined with the addition of more data.

Original Data



# B.5 Project methodology

We will be using the Agile methodology for this project. Agile development emphasizes iterative development with continuous testing and feedback. We feel this type of development is the most effective and inclusive. The shareholders opinions and suggestions are valued during each phase of the process. Below is a more detailed overview of how we will apply this methodology for this particular project.

Requirements:

* The application will be developed as closely as possible with stakeholders' desires.
* The objectives of each sprint will be elucidated beforehand and further broken down into manageable deliverables.
* The application should achieve an accuracy of at least 85%.

Development:

* Once all the requirements have been assessed, we will create a development plan. This will involve the breakdown of each requirement.

Testing:

* To ensure the project meets the requirements, regular testing will be done after each sprint.
* We will be performing both White box and Black box testing to ensure acceptable functionality.

Delivery:

* The application will be delivered to the client for their approval and feedback.
* Acceptance testing will be performed.
* Any new features the client would like added or fixed will be taken into consideration.

Feedback:

* Feedback is highly encouraged. We want to deliver a product that matches stakeholder expectations as closely as possible.
* Feedback is welcome at any stage during the development process.

# B.6 Project outcomes

Product deliverables include a trained model and python code with all the algorithms necessary to launch a fully functional application graphical user interface (GUI) which can use the model to generate accurate home prices. This application should also include a way to view the data through different filters and a way to retrain a new model within the applications interface if the user chooses to do so.

# B.7 Implementation Plan

The implementation of this project will have a defined set of tasks to be accomplished. There are four basic phases. During the first phase, we will collect and preprocess the data. This phase will involve modifying the data into a format that is used for training. This may include modifying all data points into float values and removing unnecessary column headings. We will be training the model with a couple of different forms of regression testing and choosing the method that produces the best results.

The second phase will involve the development of the user interface. During this time, we will collect feedback from all stakeholders involved as to the design implementation. Blueprints and visuals will be drawn up that match stakeholders' expectations as closely as possible. The design will include the fonts, graphics, usability, and layout. Development will be programmed using Python and Kivy. We would like to make the user interface as appealing as possible. Access to the data through the user interface should be seamless. It will allow the ability to parse user input and display all data. The user will be able to query the model through the user interface and the generated prices for homes will be displayed.

In the third phase, we will conduct acceptance testing with the application's primary users. This will include the stakeholders and realtors who will actually be using the final product. We will produce questionnaires to get more feedback. Any bug fixes or user acceptance issues will be addressed.

The final phase will involve verification of functionality on all operating systems. Project ownership will be handed over to the client at this phase and the application will be considered complete.

# B.8 evaluation plan

The program will be evaluated after each phase to verify functionality and correctness. After phase one, we will need to verify the data has been processed correctly and that there are no null values for any element, as this may cause anomalies during training. Before the project can advance beyond phase one, an accuracy rate of at least 85% on the test data will need to be met.

Phase two will consist of the user interface development. The objective is to create an interface that is easy to use and straight forward. Testing during this phase will involve verification that all navigational buttons work, error checking and the information they display is particular to the error condition, and the UI design is acceptable by the stakeholders. In addition, graphs and other visual data should display correctly and with a response time of no more than 5 seconds. Access to the model and the ability to retrain and update the model will also need to be verified. By the end of phase two, complete functionality should be achieved.

During phase three, acceptance testing will commence. Feedback will be collected and any approved changes to the visual design will be applied.

During the final phase, the program will be tested on the different operating systems.

# B.9 Resources and costs

Since the programming environment is open source, there is no cost associated with it. No special hardware or software is required. The primary cost will involve the time spent creating the program, which will be charged at an hourly rate. The total human cost is estimated to be 13,500 dollars.

# B.10 Timeline and milestones

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Stage** | **Start** | **End** | **Duration** | **Dependencies** | **Resources Assigned** |
| Beginning of project | May 1, 2021 | May 1, 2021 | 0 days | N/A | Stakeholders |
| Phase one | May2, 2021 | May 5, 2021 | 3 days | N/A | Developers |
| Phase one complete | May 5, 2021 | May 5, 2021 | 0 days | Task 2 | Stakeholders, developers |
| Phase two | May 5, 2021 | May 12, 2021 | 1 week | task3 | Developers |
| Phase two complete | May 12, 2021 | May 12, 2021 | 0 days | Task 4 | Stakeholders, developers |
| Phase three | May 13, 2021 | May 16, 2021 | 3 days | Task 5 | Users, Stakeholders |
| Phase four | May 16, 2021 | May 19, 2021 | 3 days | Task 6 | Developers |
| Project Finish | May 20,2021 | May 20, 2021 | 0 days | All tasks | Stakeholders, Users |

# D. Post Implementation Report

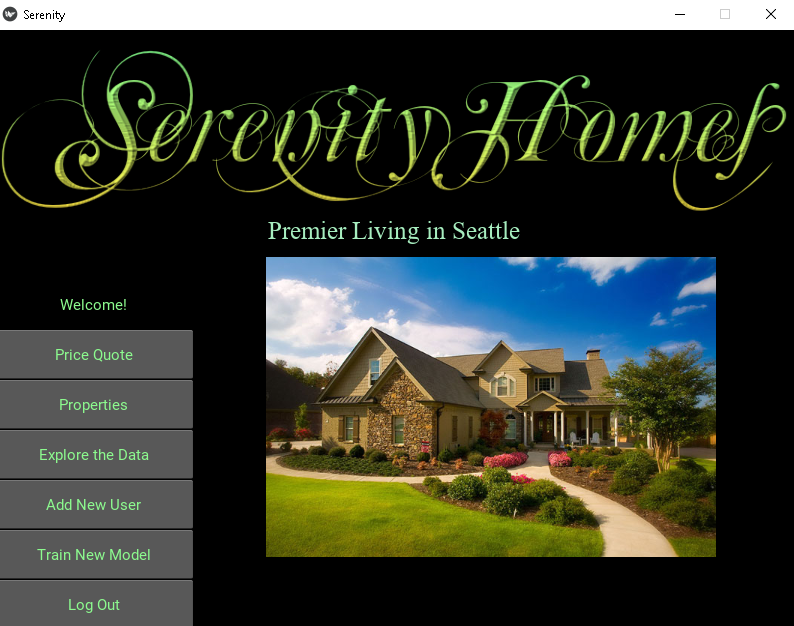
# D.1 Project Purpose

The problem we wanted to address is how a real estate company can accurately determine the price of an acquired home. We wanted to create a standalone application that allowed the user to input a number of features of the home and to immediately get an accurate baseline price. We also wanted this application to allow the realty company to review the data through colorful illustrations to gain knowledge of real property sales for the location where they are currently based. This information would enable them to make informed decisions in the future.

The application addresses the expectations of the client by providing these functionalities in an all-in-one dashboard. A login page is used to ensure security. From the main page, the user has the options to get a price quote, explore the data through statistical analysis and apply k-means clustering on any aspects of the data they desire, add a new user who then may access the application in the future, and train a new model. They can also view their own properties with pictures and important details.

We believe this will result in faster home sales and better decisions in the future.

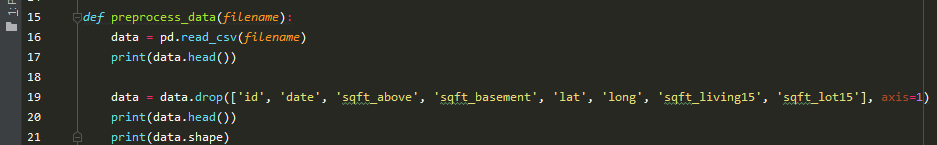
Main Screen Image

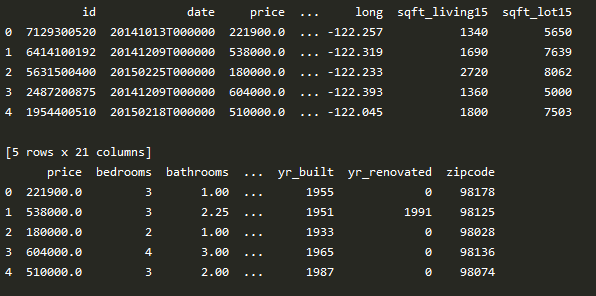


# D.2 Datasets

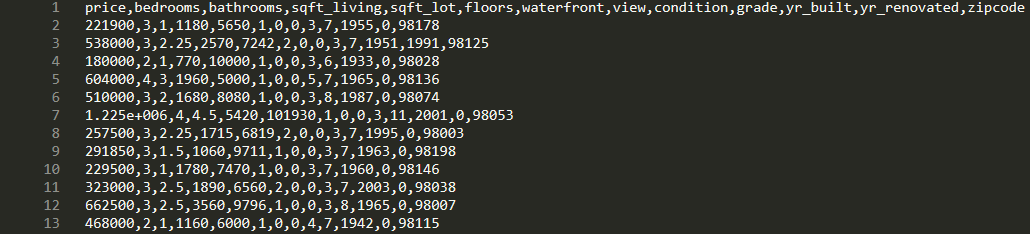
The dataset used was taken from kaggle.com and consists of 21,613 home sales in King’s County, Seattle. The original dataset is can be found in the folder “data/kc\_house\_data.csv”. It is a csv file with 21 column headings. For our project, we removed some of the columns in order to make the application as simple as possible, as the user may not have detailed knowledge of data like latitude and longitude, square area of nearby houses, etc. We preprocessed and cleaned the dataset to create a new file that can be found under the folder “data/modified\_house\_data.csv”. There were very few anomalies with the data. About 13 of the homes were above 3 million dollars. However, removing high priced homes from the data yielded very little difference in training accuracy. Because most of the dataset was already in numeric form, preprocessing only required removal of the column headings we weren’t going to use. The new file has 13 column headings. An example of the changes is below.

Removal of unused column headings





Modified data



# D.3 Data product code

The code used to preprocess the data and train the model can be found in the train\_model.py file. We tried a few different machine learning methods when training the model including state vector machine (SVM), linear regression, and a Decision Tree Regressor. We found that the SVM was far too slow during training and therefore was not an acceptable method of training. Training with Linear Regression produced an accuracy of 71 – 73%, which was an unacceptable percentage. We found the Gradient Boosting Decision tree regression method to be the most precise.

Preprocessing the data was very minimal as most of it was already in the format we needed. We used Pandas for processing, which automatically converted the integer and integer string variables to float values. After training with different feature sets, we removed the column headings that were not necessary.

The descriptive methods can be found in the file create\_visualizations.py. We used seaborn and Matplotlib to create the data visuals. Seaborn offered a better variety of options in displaying the data and we were able to produce some nice graphs. In order to create some of the graphs, we isolated the two features we wanted to work with into a separate dataframe. We found the unique values of the first feature and then determined the average and totals of the other feature in relation to the first. This allowed us to directly compare two features in the data and then plot them in a bar graph. For example, the average price of all the homes with a poor, fair, or very good condition.

We implemented K-means clustering in the file k\_means\_clustering.py. We set up a method that allowed for the user to input two variables and then generate cluster images. To do this, we would again create separate dataframes and then plot them using different colors. This allows the user to discover what data points belong to which cluster and where the centroid value lies. The user is able to do this directly from the program.

# D.4 Hypothesis Verification

Our original hypothesis was that we could provide an accurate baseline home price 85% of the time. Using the Decision Tree Regression method, we were able to achieve an accuracy rate on the test data of 88.5%. This data included only the label (home sale price) and 12 features that were used to predict the label. When we included 20 features in the original dataset, the accuracy was closer 91.5%. However, because we wanted to create an application that asked a minimal number of questions and the user probably doesn’t have immediate knowledge of complex features, we trained with the original modified dataset.

The generated price could be anywhere from 100,000 to 3 million or more. The only way to determine accuracy would be to measure how close the model performed on test data. Within the application, the user can adjust individual features to see how much they effect the price. For example, if the user enters 2 bedrooms and the generated price is 436,000 and then increases the number of bedrooms, we would expect the newly generated price to be above 436,000. This is the primary way to test the model within the application. When we performed these tests within the application, the price adjusted in the direction we expected 9 out of 10 times. We consider this a success.

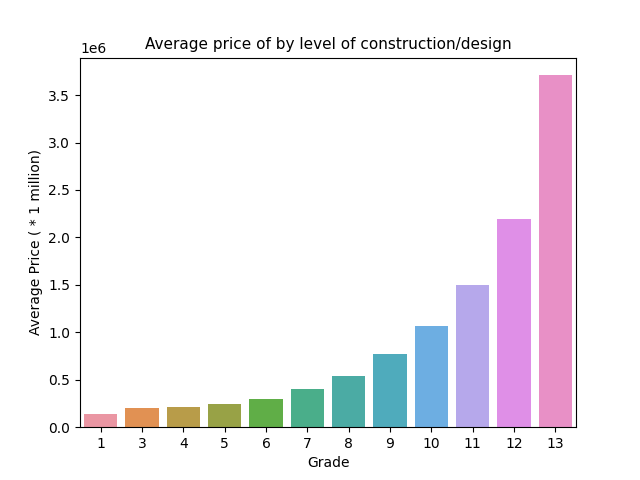
There were certain other input factors we considered. For example, if the user entered a number of bathrooms that was extraordinarily high (ex: 50), we thought the model would generate a ridiculous price. Interestingly, the model seemed to ignore this and still generate a price that was reasonable. It seemed as if the model would ignore these values and place a heavier emphasis on the other features.

# D.5 Effective Visualizations and reporting

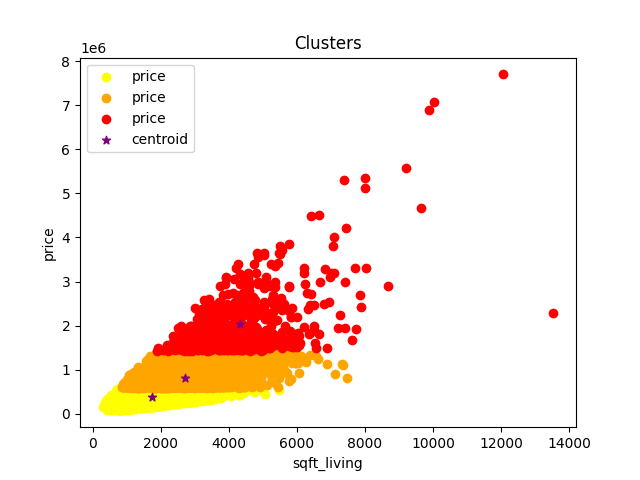
It was important for us to include a robust number of visualizations within the application. There is a button on the main screen titled “Explore the Data” that takes the user where they can view the datasets characteristics through a number of charts. These charts include:

* Average price of homes based on view
* Average price of homes based on the level of construction and design
* A plot of the price of the homes in the dataset based on the square footage of interior living space. Our research indicates this is the strongest factor that determines the price.
* Counts of all the unique elements of each feature like bathrooms, bedrooms, and grade.
* An implementation of K-means clustering.

The code is provided so the client can create updated descriptive data in the future. The visual data was created with the python libraries Matplotlib and Seaborn.

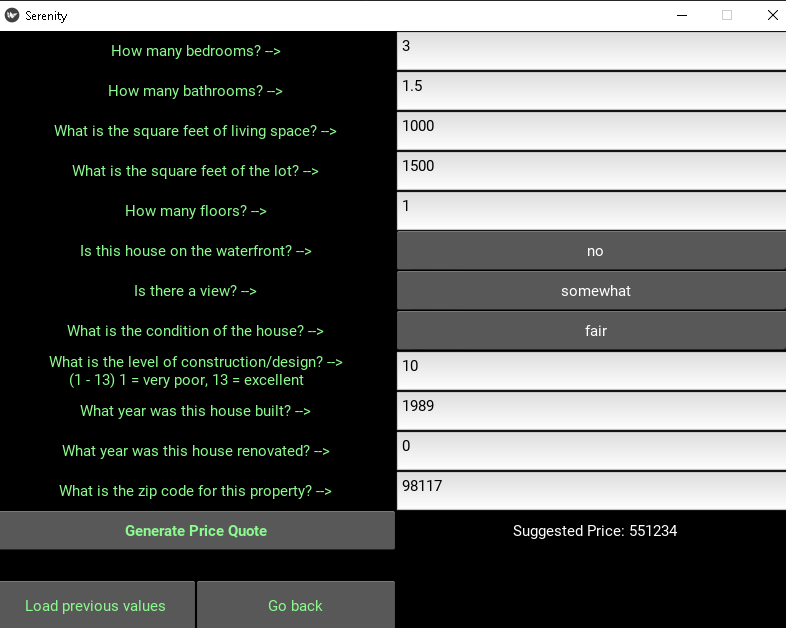


We implemented an entire screen for k-means clustering. K-means is a method used in clustering analysis, especially in data mining and statistics. It can be considered a method of finding out which group an object belongs to. On this screen, a user can choose any two features they want to compare and the number of clusters they want to use and four different images will be generated. The first image compares the value of K to the sum of squared error. This is a helpful graph because it can help the client identify the “elbow”, the point in the line graph where there is a noticeable bend. The elbow can assist in choosing the best value for the number of clusters that should be used. The other three images are clusters, scaled clusters, and an image that shows the centroid points for each cluster. The user can change any of these features and values to update all of the images in real time.



# D.6 Accuracy Analysis

Most of the testing occurred during phase 1 and phase 2. During phase 1, we created the backend for the program. This involved training the model and testing accuracy on the test set, which was 10% of the data. Testing during this phase yielded an accuracy rate of over 88%. This indicates that the model is effective at providing predictive outcomes. During phase 2, we tested the model through the GUI. The user enters a number of features of the home and is presented with a suggested price. Testing during this phase involved altering the features and verifying the suggested price changed in the direction we would expect. The success rate was 9/10 times. This shows the application is successful at presenting the data. An image of the Price Quote screen along with a generated price is below.



# D.7 Application Testing

During each phase, the application was tested to verify its integrity. We performed model testing during phase one. This was important because we needed to develop a strong backend for the program. During phase two we performed unit testing. We needed to verify that all methods in the code were accessible and produced the correct results. Every time the application was launched, we tested whether or not all the buttons were operational and they produced the desired result.

Acceptance testing was initiated during phase three. Realtors of Serenity Homes and stakeholders had a chance to try out the program. We provided each of them with a questionnaire and recorded the answers. Afterwards, we revisited the code and made some minor adjustments to the design. Functionally, the program behaved as expected.

Phase four mostly consisted of testing on different operating systems. This was important because most of the users of the application would be using their own computers and their preferences may vary. We tested the software on Windows, Mac, and Linux. It successfully passed all the tests.

# D.8 Application Files

All of the necessary files can be found in the submitted Capstone folder. These files include:

* home\_price\_generation.py - The application is launched from here. Contains most of the logic for the application.
* serenity.kv - This is the accompanying .kv file. Provides the design of the application.
* Train\_model.py - used for model training.
* K\_means\_clustering.py - contains the methods used to create k-means clustering.
* Create\_visualizations.py - contains methods to create different bar charts and plots.
* Folder db contains the database.py file and users.csv that instantiates a basic database for the login screen.
* The folders housepics and app\_images contain the images used in the application.
* The folders data\_images and k\_means\_images contain the descriptive images used in the application.

For further assistance, see User’s Guide below.

# D.9 USER’s guide

For users and evaluators of the application:

1. Open the Capstone folder
2. You may need to install dependency libraries sklearn, matplotlib, and kivy. You can do this with in the terminal using pip install sklearn, pip install matplotlib, and pip install kivy.
3. Launch the file house\_price\_generation.py located in the Capstone folder using your preferred IDE. Pycharm was used by the creators of this application.
4. From here the login page will open. If you are a first-time user, you can input the username “USER” and the password “PASSWORD”.

# D.10 Summation of Learning Experience

This project has been a learning experience in many ways. During this project, I learned how to train a machine learning model and access it through a graphical user interface. I enjoyed designing an application from scratch. While Java is my primary language, I decided to test myself by creating this project with python. I feel that python is certainly a more versatile language but I had no idea how to go about creating a visual application with it. I had to learn Kivy from scratch with numerous youtube videos to create the GUI.

I am very interested to work on a mobile application in the future. I believe my experience with Kivy will enable me to create a mobile application using the python programming language. I did not receive any assistance in creating this application and there was a big learning curve, but it was worth it. This experience has contributed to my concept of lifelong learning by providing me with the tools necessary to create programs that provide purpose and entertainment.