

CS 315 - Data Mining: Course Project Proposal

Amethyst Skye

October 23, 2022

1. **Data Mining Task:** *What is your data mining task? This task could be a series of exploratory questions that you want to investigate or analyze. What is your motivation behind choosing this task for your project?*

The data mining task I will be working on involves predicting the sale price of a home in Clark County, WA based on the properties of the home. It will be titled “Home Price Predictions for Clark County, WA”. I’ll be using supervised learning techniques like regression and classification. Since housing prices seem to have been rising quickly over the past decade, it would be useful to analyze what properties contribute to different purchase prices. Some of these properties include sale date, acreage, the style of home, and quality rating.

2. **Dataset:** *What is the source of your data? Provide a link to your data source if you acquired it online.*

The dataset I will be using is from the Clark County website. It is based on residential sales information that occurred most recently (2021 - 2022). There was a csv file available for download. Here is the link: <https://clark.wa.gov/assessor/residential-property-sales-information>.

3. **Methodology:** *How will you solve the data mining task? You should have some idea of the algorithms or software tools you plan to investigate.*

I plan on solving this task using various data science and machine learning libraries available for Python (pandas, numpy, matplotlib, sklearn, seaborn, etc.). I’d like to test both linear and multiple regression models using the sklearn library. There will also be iterations of cleaning that I will need to do to remove any unnecessary data such as the dollar sign and any commas in the column for listing/purchase price.

To do the majority of my programming work, I will use either Visual Studio Code, or Google Colab where I can create Python notebooks. Additionally, I plan on using excel to see if I can get any useful missing information (such as zip code) to add more dimensionality to the project.

4. **Final product:** *What will be the outcome of this project? How will you measure the success of your course project? Will this project help you explore or learn something new?*

The final outcome of the project will result in a better understanding of how to use machine learning models in Python. For this project, it will be a good exercise in using some data analysis knowledge to gain better insights and improve what information is used to train the machine learning model. I plan on comparing the outcomes of examining different housing properties against each other to see what results in the highest accuracy. This will be a wonderful opportunity to use the machine learning knowledge I have gained in a useful application.