# Project 4--Project Proposal Team 1

**Project title**

ML Application: Predicting Real Estate Sales in Connecticut

**Team Members**

Tori Paxton

Teresa Truong

John Weissmann

**Project Description**

The goal of this project is to create, train, optimize, and evaluate a supervised machine learning model that predicts the sale price of a real estate property in Connecticut based on location and property type with at least 75% accuracy.

**Data sets to be used**

[Real Estate Sales 2001-2020 GL](https://catalog.data.gov/dataset/real-estate-sales-2001-2018) from data.gov

**Rough Breakdown of Tasks**

Data preprocessing

* Obtain zip code based on address, town
* Clear unused columns
* Standardize sales prices with PCA
* One-hot encode categorical data
* Create bins for predictions(?)

Create model

* Train/test split
* Create labels and features
* Check balance of features 🡪 if unbalanced, use RandomOverSampler
* Fit model

Optimize model

* Record optimization trial conditions, accuracy ratings

Evaluate model

* Evaluate final accuracy

Create summary visuals

* Sales prices by location
* Sales price by property type

Create presentation