

CSE 4705: Dataset Definition & Question Ideation Submission

Tom McCarthy

Filip Graham

(We happened to pick the same real estate dataset in class as another group so we chose another dataset to supplement, per instructor request).

Datasets:

<https://catalog.data.gov/dataset/real-estate-sales-2001-2018>

This dataset contains information of real estate sales in Connecticut from 2001 to 2022.

Each row contains sale information including town, date of sale, property type, sales price, property assessment and sales ratio (assessed value divided by the sale price).

<https://catalog.data.gov/dataset/municipal-fiscal-indicators-economic-and-grand-list-data-2019-2024>

This dataset is a compilation of financial and demographic information on municipalities in Connecticut between the years 2019 and 2024. It includes aggregate information like net current education expenditures, and per capita values like median household income.

Some possible questions to explore using data:

- What is the relationship between education spending and the real estate market? Does a change in spending affect sale prices and sales ratio in the near future, or vice versa?
- Do household income statistics correlate with real estate prices (or are home buyers buying homes above/below their means)? Can we predict future real estate prices based on trends in income?
- Do changes in mill rates (tax rate) for property taxes affect real estate sales in surrounding municipalities, possibly indicating that property taxes play a factor in the decision to buy real estate?
- Can unsupervised learning find additional relationships between real estate sales and municipal statistics?
 - How does the amount of vacant land in a municipality correlate with property prices?
- Can anomaly detection find additional relationships between real estate prices and municipal statistics?
 - Are there municipalities with unusual real estate prices changes compared to their fiscal status (income levels, tax rates)