

Basic idea

1. Use code can be marked as Unit_countable and Unit_notcountable.

- By looking at some use codes, we can infer the number of units in a building. -- For example : 101 - Single family. - There are other usecodes that cannot tell much about the number of units within a building. -- For example: 109 - Multiple Houses on one parcel, 111 - 114 Apartments.

2. For those Unit_notcountable usecodes, we use address count directly.
3. For those Unit_countable usecodes, we need to do another classification. Classify those usecodes as unit_confident and unit_notconfident.

-- For example: 102 - Condo. We will use unit count directly. 102 is a unit_confident usecode. -- For other unit_notconfident usecodes, we will do data analysis. - -- If address count == unit count, use this value as prediction. - - - If address count != unit count, mark this data as anomaly. Use linear regression to make a good prediction.

All residential usecodes:

[[01x, 0x1], [101 - 11x], [945, 959, 970]], x in [0, 9]

Notice: all the proofs are in /note/proof/

TODO:

- May need to use town id - (Manually classification is still required or maybe use NLP)

Unit countable usecodes:

- 0101 - single fam
- 0102 - condo
- 0104 - two fam
- 0105 - three fam
- 010E - two fam
- 010G - three fam
- 010H - two fam
- 010I - single fam
- 010M - single fam
- 101 - single fam, official docs
- 102 - condo, official docs
- 104 - two fam, official docs
- 105 - three fam, official docs

Unit not countable usecodes:

- 0103 - mobile home
- 0107 - not clear
- 0108 - not clear
- 0109 - Multi houses on one parcel
- 010C - not clear
- 010F - not clear
- 010J - not clear
- 010Z - not clear
- 011 - apartments
- 012 - not clear
- 013 - not clear
- 014 - not clear
- 015 - not clear
- 016 - not clear, mainly apts
- 017 - not clear, mainly apts
- 018 - not clear, mainly apts
- 019 - not clear, mainly condos
- 021 - not clear
- 031 - not clear
- 041 - not clear, mainly apts
- 051 - No records
- 061 - not clear
- 071 - not clear, mainly apts
- 081 - not clear, mainly apts
- 091 - not clear, mainly two fam
- 103 - mobile home, official docs
- 107 - not clear, official docs
- 108 - not clear, official docs
- 109 - multi houses on one parcel, official docs\
- 11x - apts, official docs
- 12x - Non-Transient Group Quarters (SKIP)
- 945 - not clear
- 959 - not clear
- 970 - not clear