

ZILLOW

Zillow's Home Value Prediction

Presented by Sravya Bhaskara



INTRODUCTION

DATASET

- Zillow competition from kaggle
- Housing details of states in US
- Purpose: Analyze property trends, regional values, and patterns to gain insights into the U.S. housing market.

DATABASE

The dataset, sourced from Zillow, provides real estate property details, including tax, building features, and regional information, aimed at analyzing property trends, regional values, and characteristic patterns

TABLES

- 10000 rows in core tables
- ~ 40 columns combining all tables
- Two Groups
 - Core Tables
 - Look up Tables
- Primary Key for Core Tables: Parcelid

LIST OF TABLES



Core Tables

- Property Table
- Building Table
- Pool Table
- Region Table
- Price Table
- Yard Table

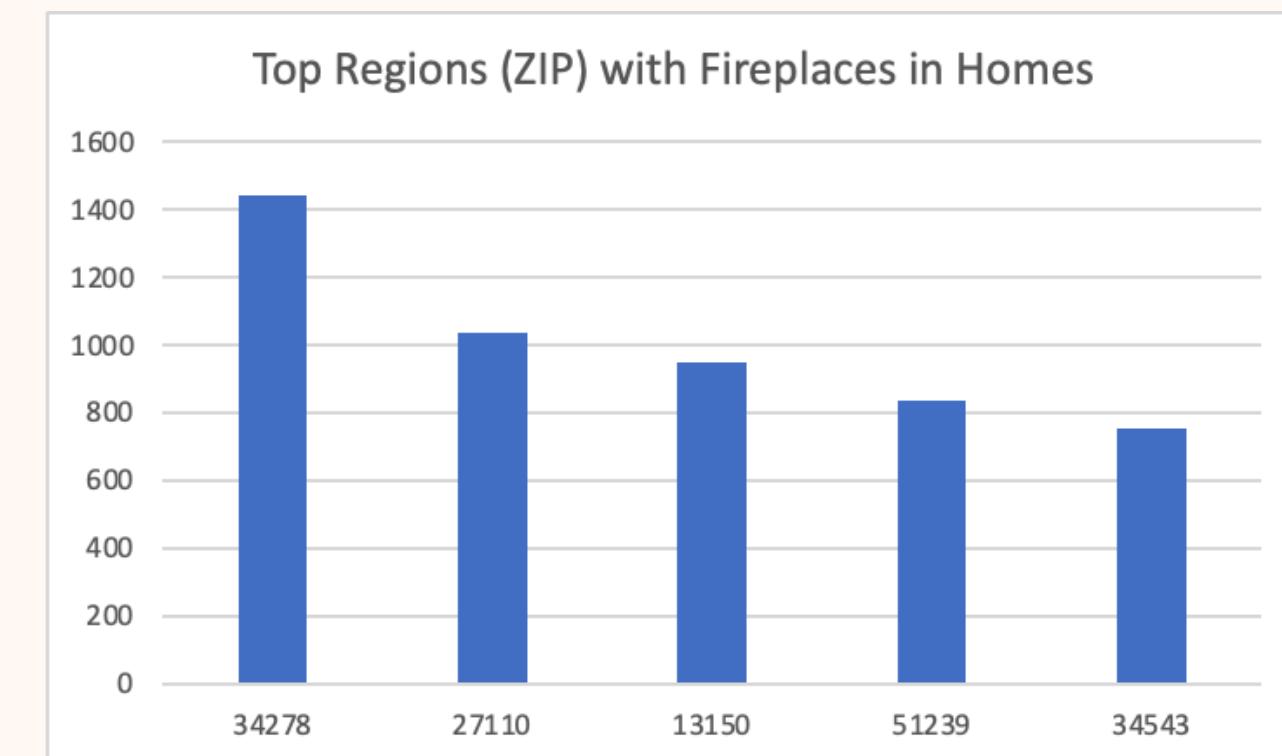
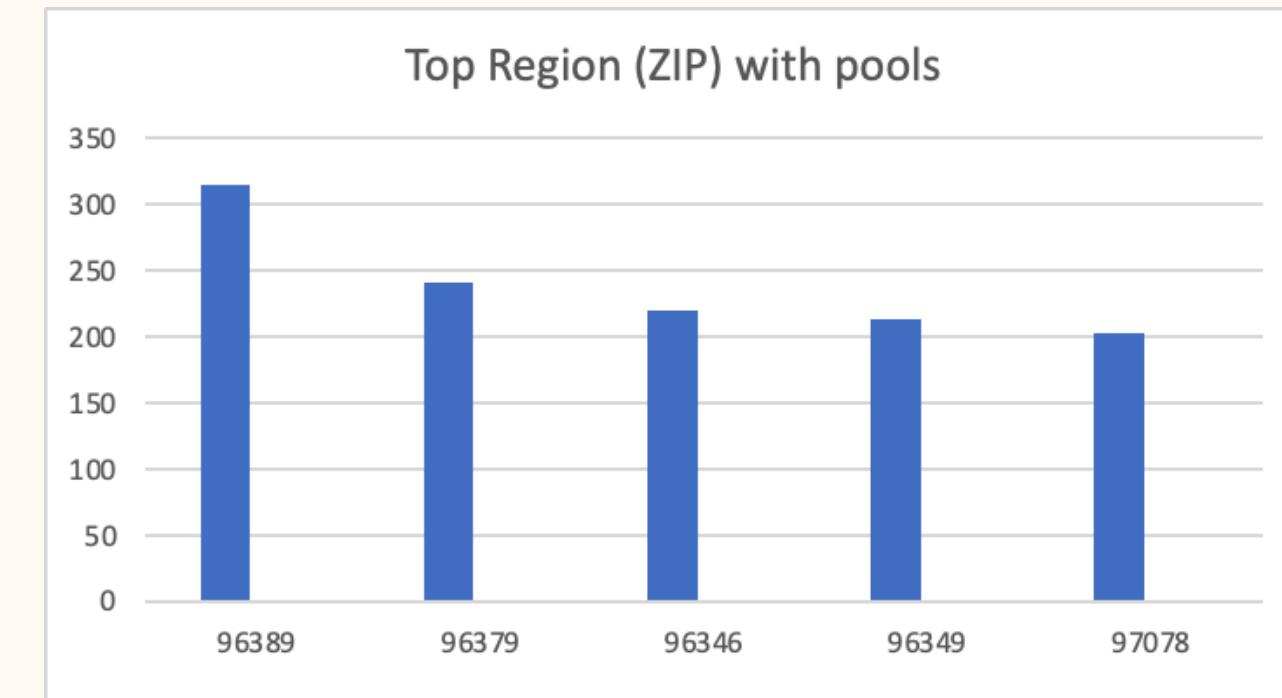
Look up Tables

- Building Quality Table
- Air Conditioning Quality Table
- Heating System Quality Table
- Pool Quality Table

1. Give counts of following for the properties for each region, to understand which regions prefer pools and fireplaces

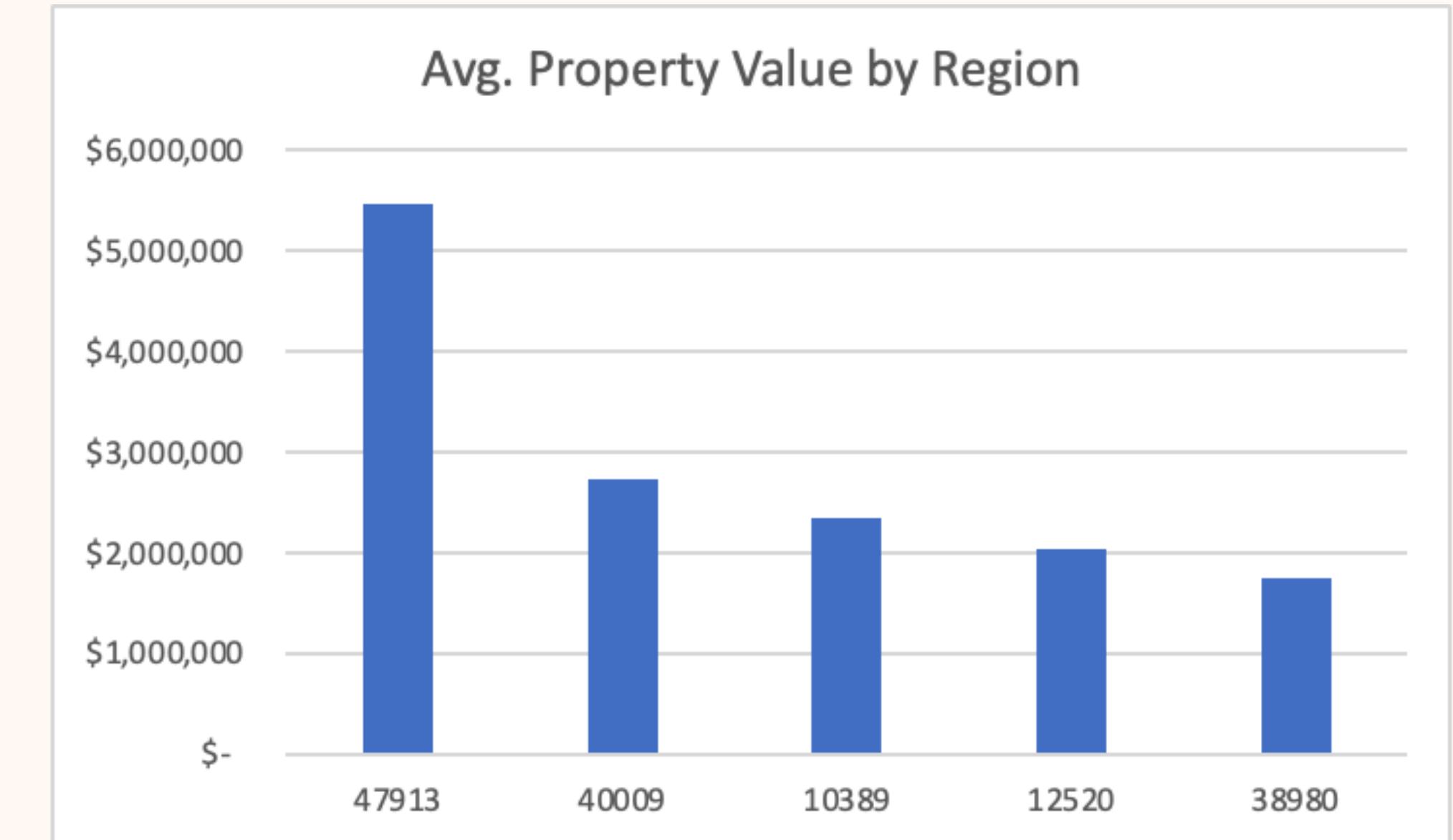
- a. Pool Count
- b. Fireplace Count

- Counties with zip 96389, 96379, and 96346 show more preference for pools.
- Counties with zip 34278 prefers fireplaces a lot
- Regions with both features may indicate a trend toward luxury preferences, though data is sparse.



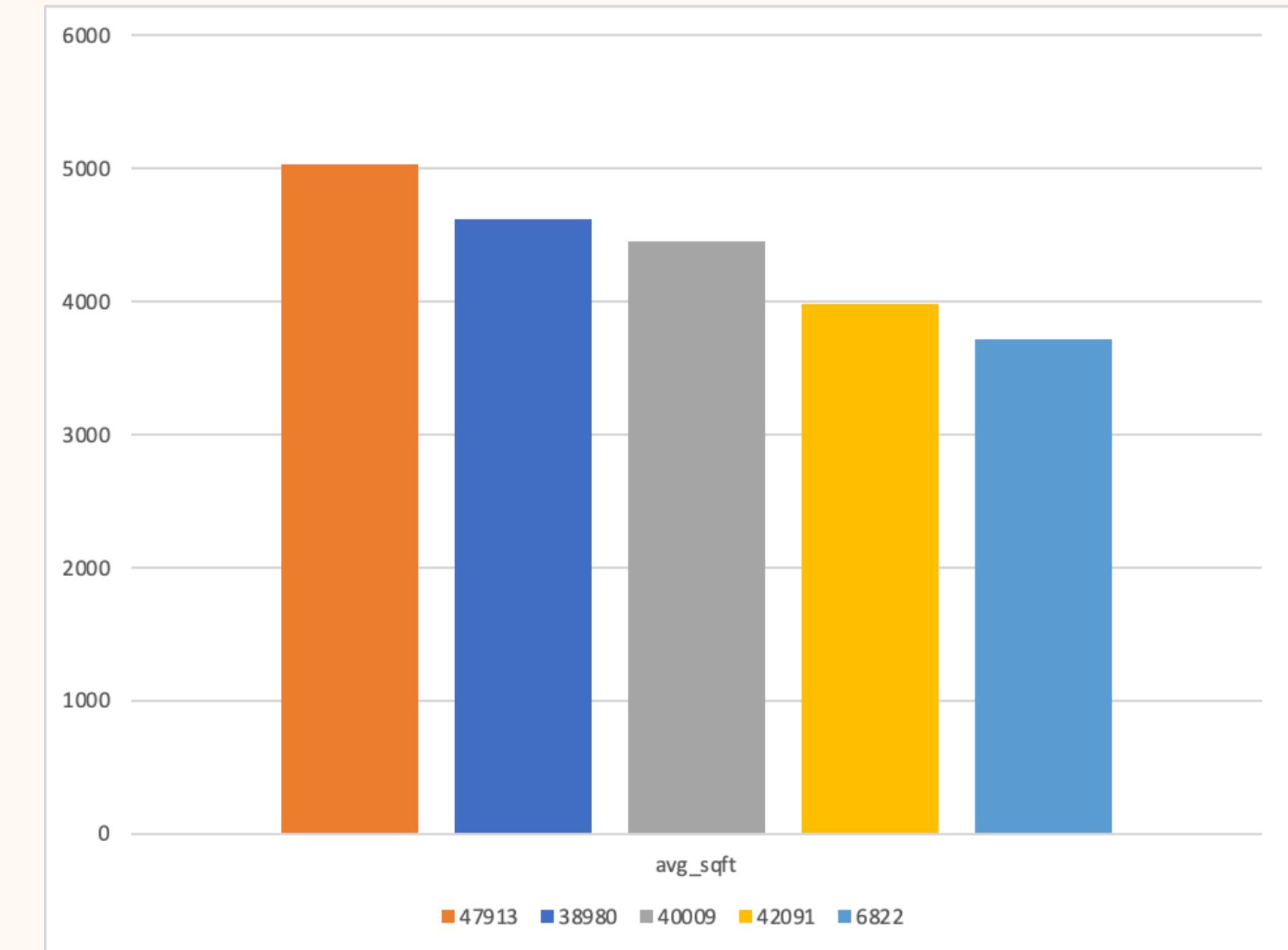
2. List Average Property Value for each city, to assess the living expenses

- City 47913 has the highest average property value (\$5.5 million), suggesting higher living expenses.
- City 26521 has the lowest average value (\$141,000), indicating a more affordable cost of living.



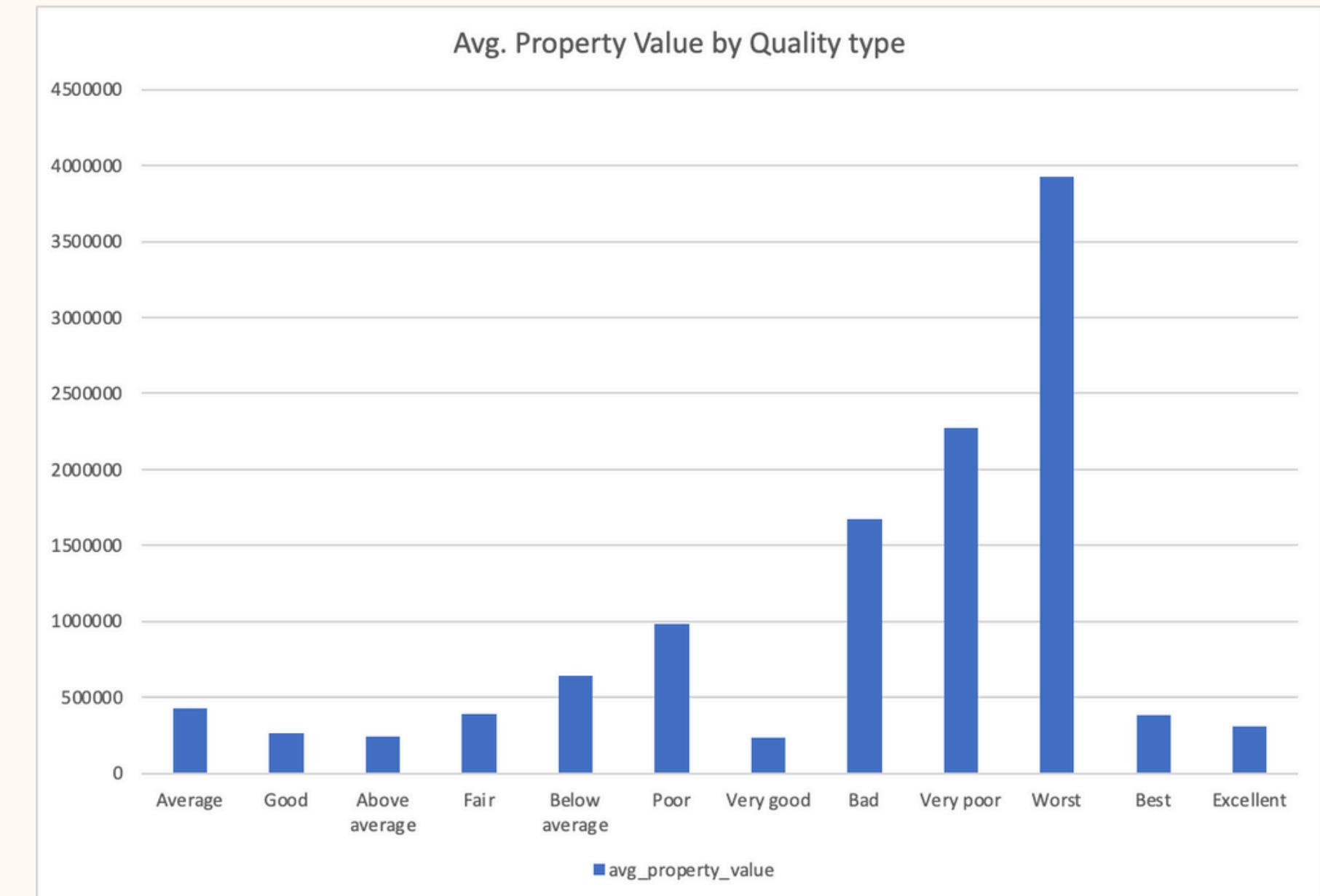
3. List Average Property Size for each region, what is the trend looking like?

- Region 479143 features the largest properties on average (5000 sq. ft.), appealing to families or those desiring more space.
- Regions like 41484 have smaller average property sizes (685 sq. ft.), possibly reflecting urban settings.



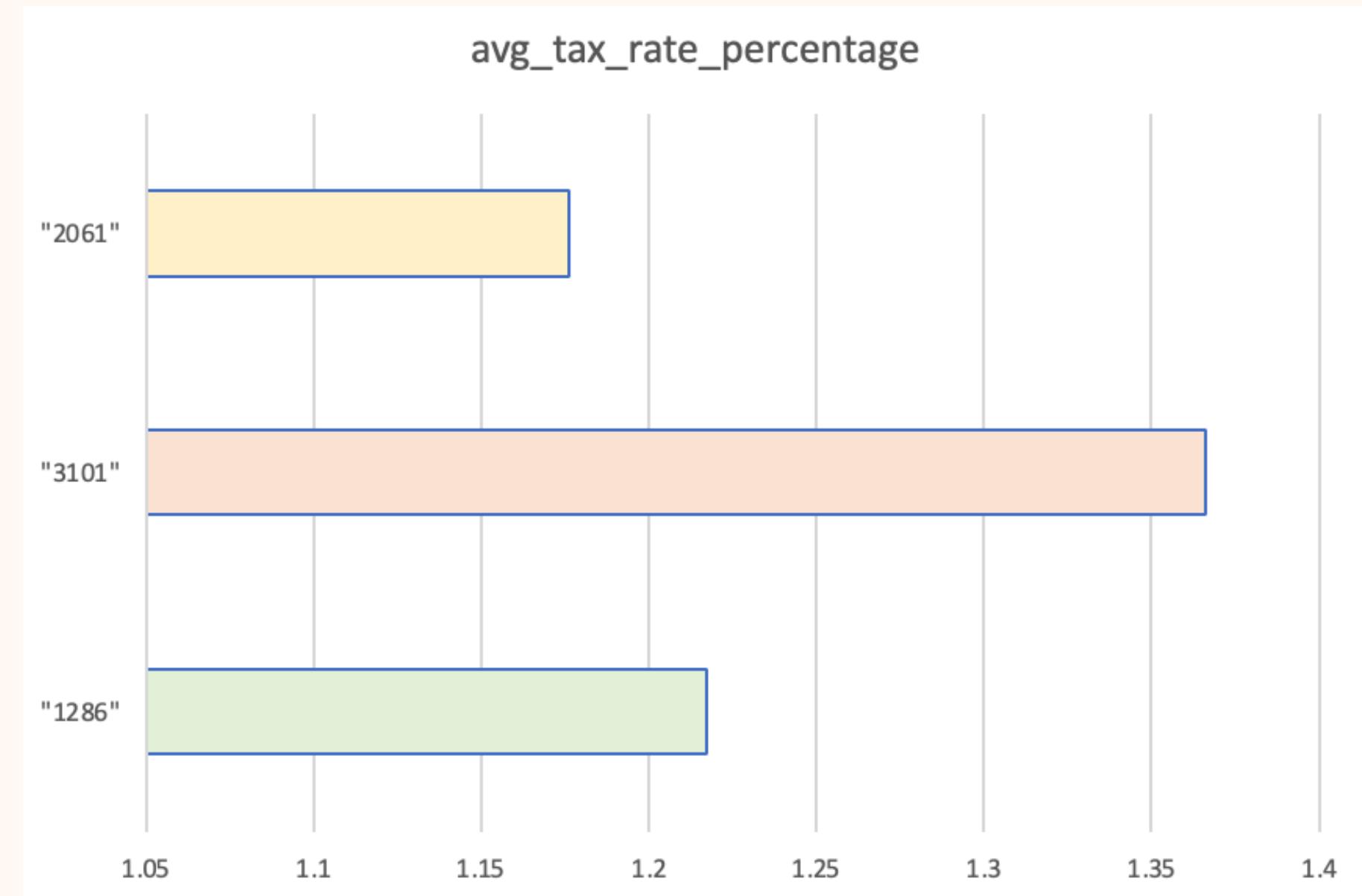
4. Find Average price for different property types and find premium property types

- Properties labeled "Worst" surprisingly have the highest average value (\$3.93M), possibly due to unique features or land value.
- Other high-value categories include "Very poor" (\$2.27M) and "Bad" (\$1.67M), suggesting specific demand for rare or premium features.
- Lower-quality descriptions like "Below average" are significantly more affordable (\$642K), catering to budget-conscious buyers.



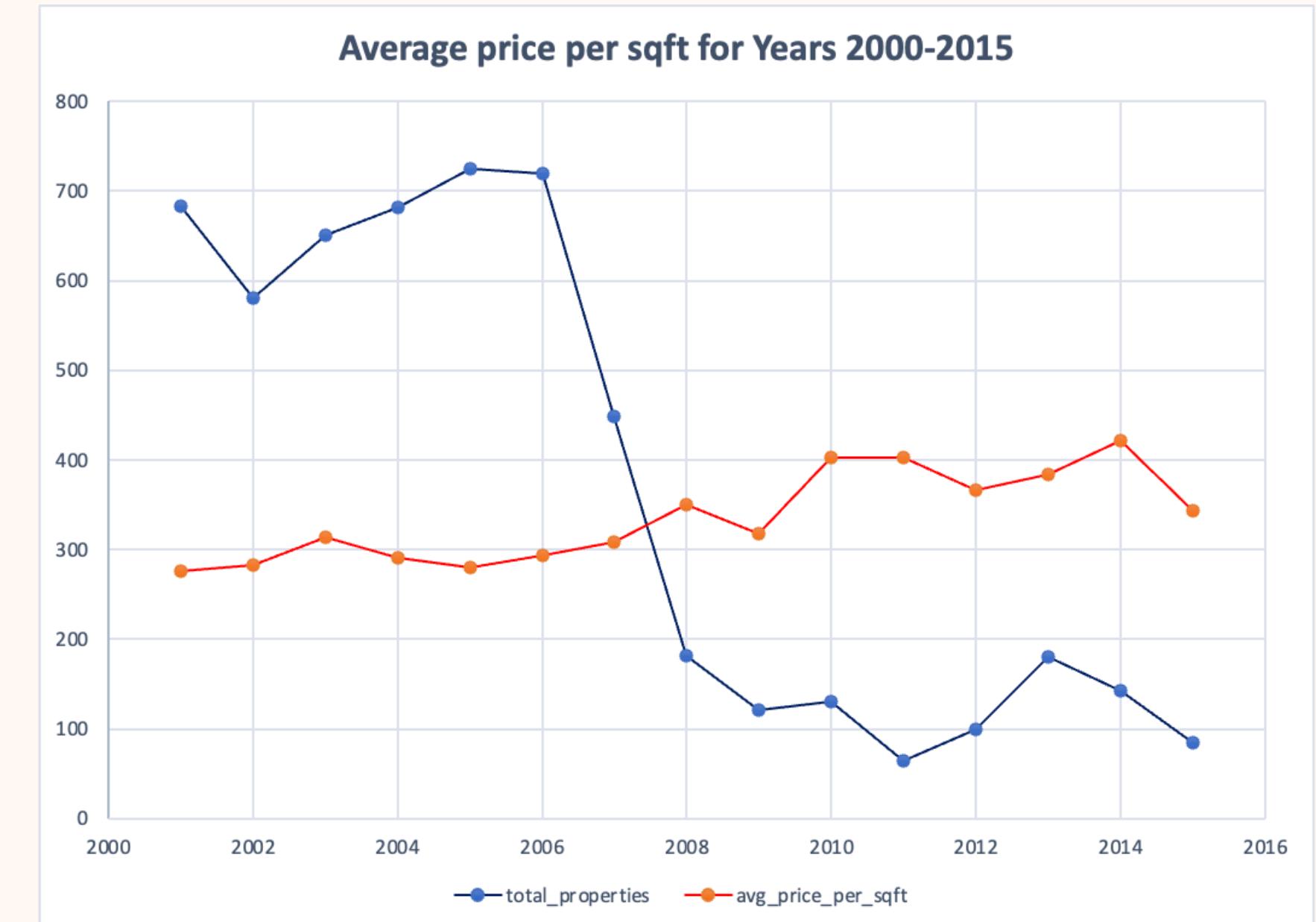
5. Average Property Tax Rate by County

- County 3101 have the highest average property tax rate (1.37%), indicating a potential influence on affordability
- County 2061 and 1286 are on lower end



6. What is the average price per square foot for properties by year built (for properties built after 2000)

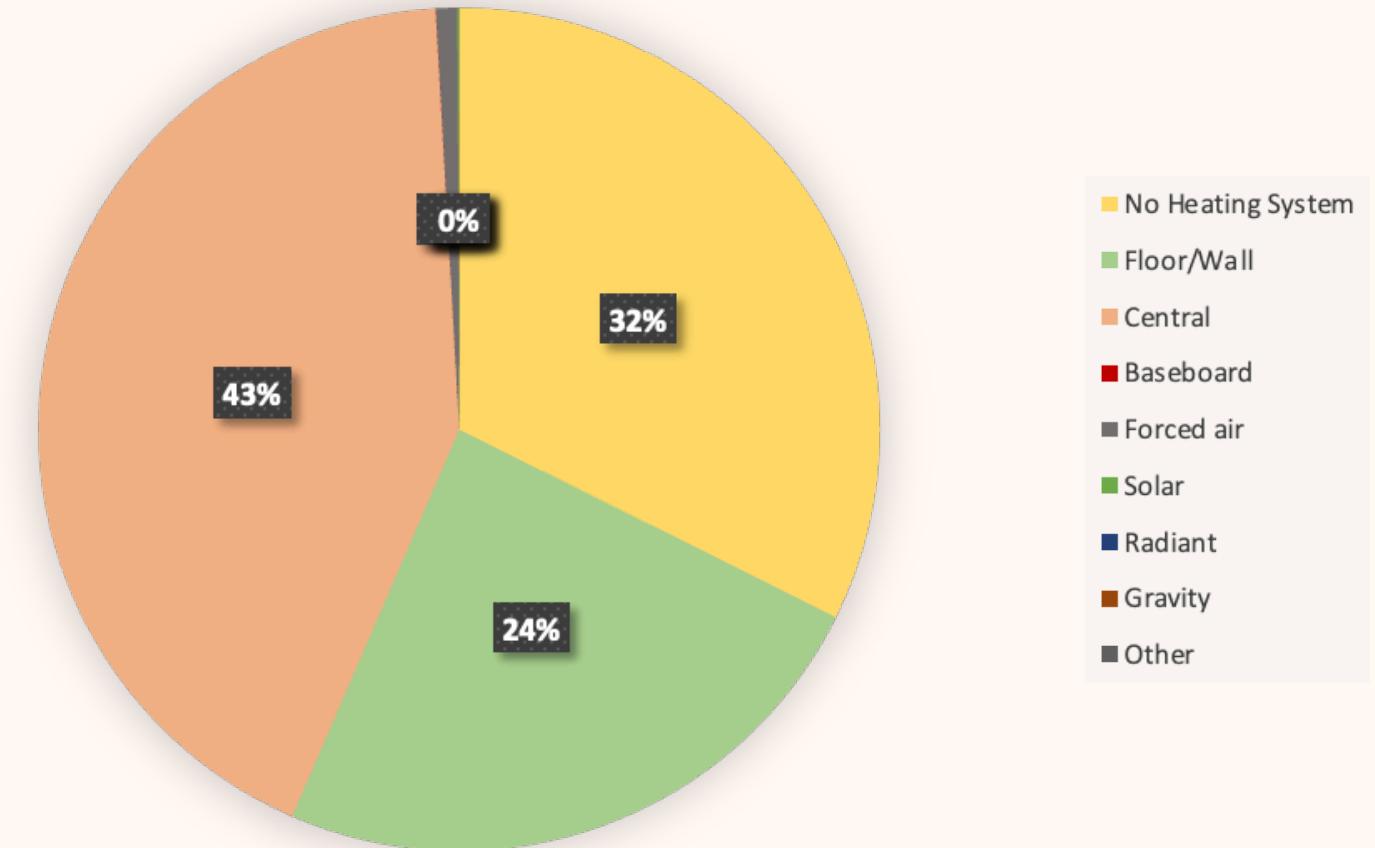
- Prices per square foot peaked in 2003 at \$313.36, followed by a constant trend in subsequent couple of years.
- All time high was in 2014 with \$420.87, followed by a decline again
- Number of total properties has declining trend after reaching its peak in 2006



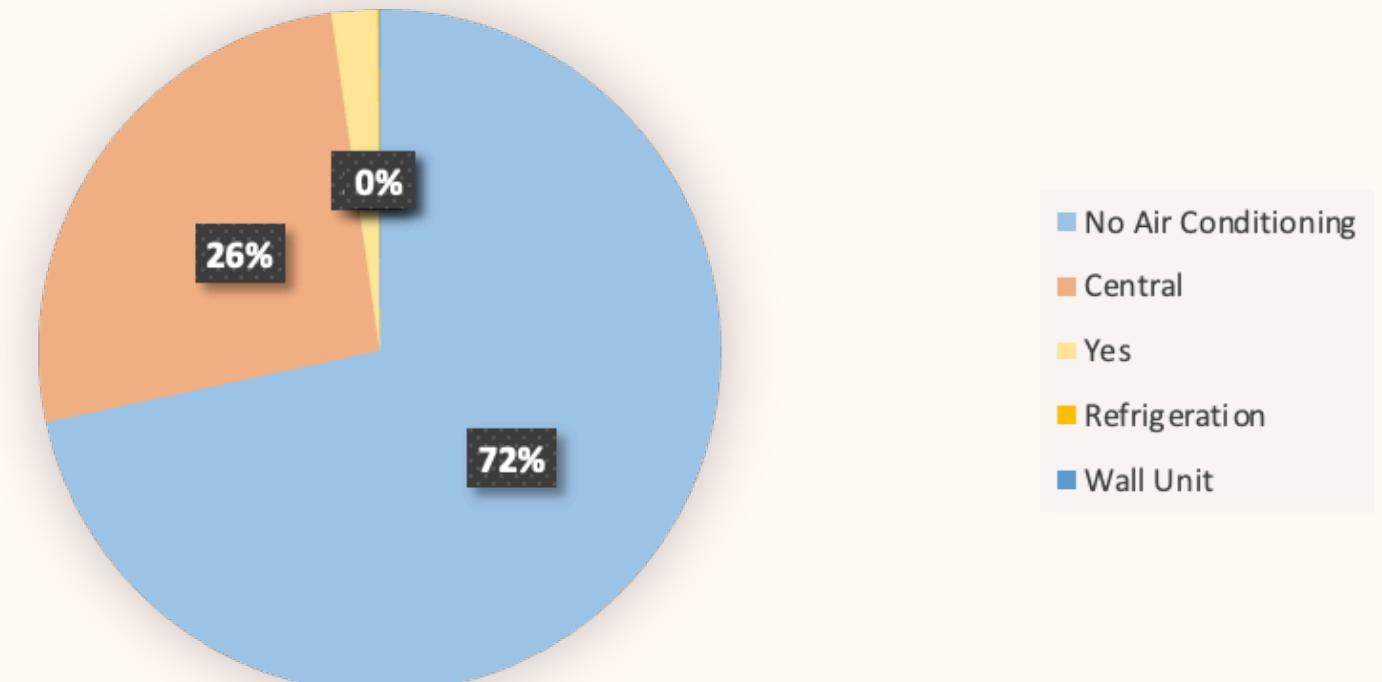
7. List most popular Heating System Type and Air condition quality type

- Central heating is overwhelmingly preferred (43% properties), indicating a standardization of heating systems.
- Air conditioning is not available for majority of the homes.
- This could reflect regional climates where heating is prioritized over cooling.

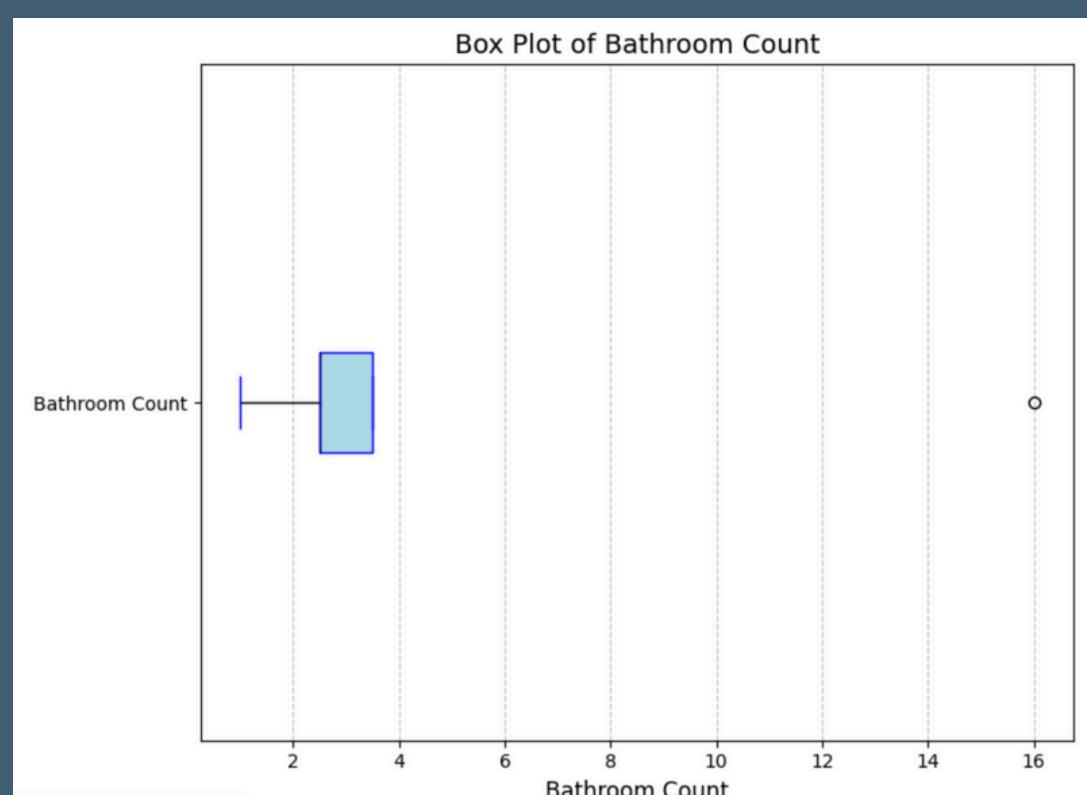
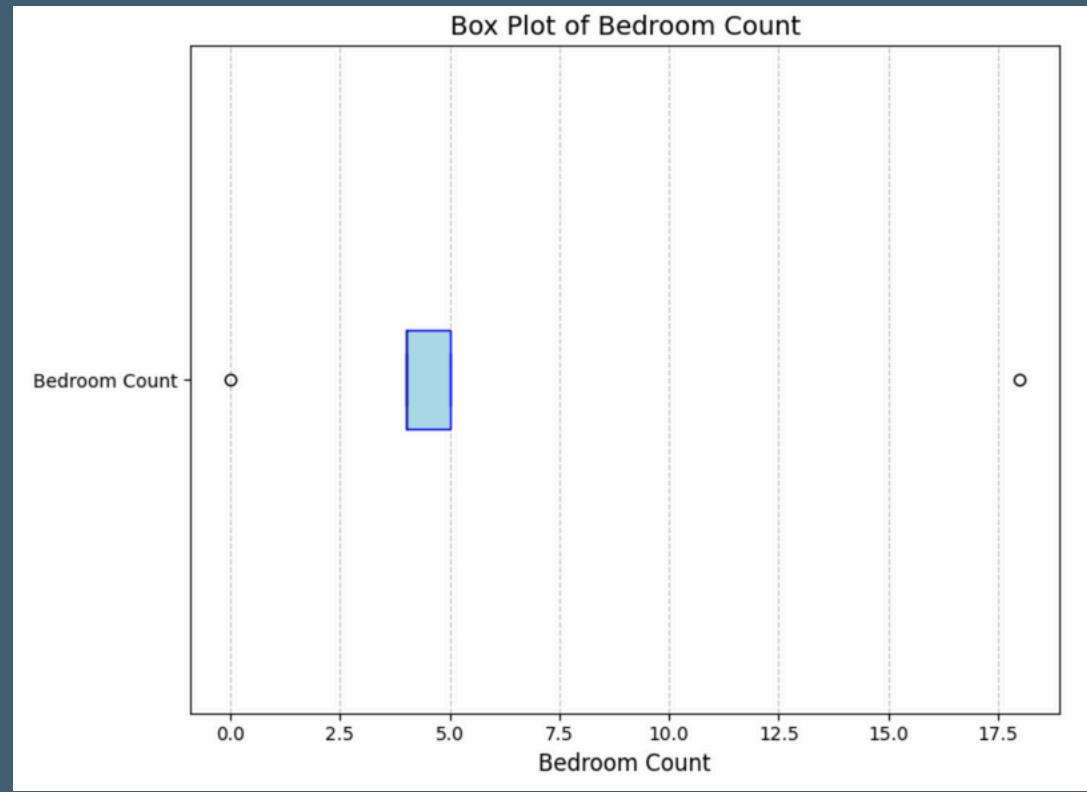
Most Popular Heating System



Most Popular AC System



8. What are the distribution statistics of bedroom and bathroom counts in the building table (Min, Q1, Q2, Q3, Max)



Metric	Min_Value	Q1	Q2	Q3	Max_Value
Bedroom Count	0	4	4	5	18
Bathroom Count	1	2.5	2.5	3.5	16

- Median bedroom and bathroom counts are 4 and 2.5, respectively, suitable for medium-sized families.
- Outliers exist with properties having up to 18 bedrooms and 16 bathrooms, indicating luxury or multi-family homes.
- Quartile values suggest a skew toward properties with moderate room counts.

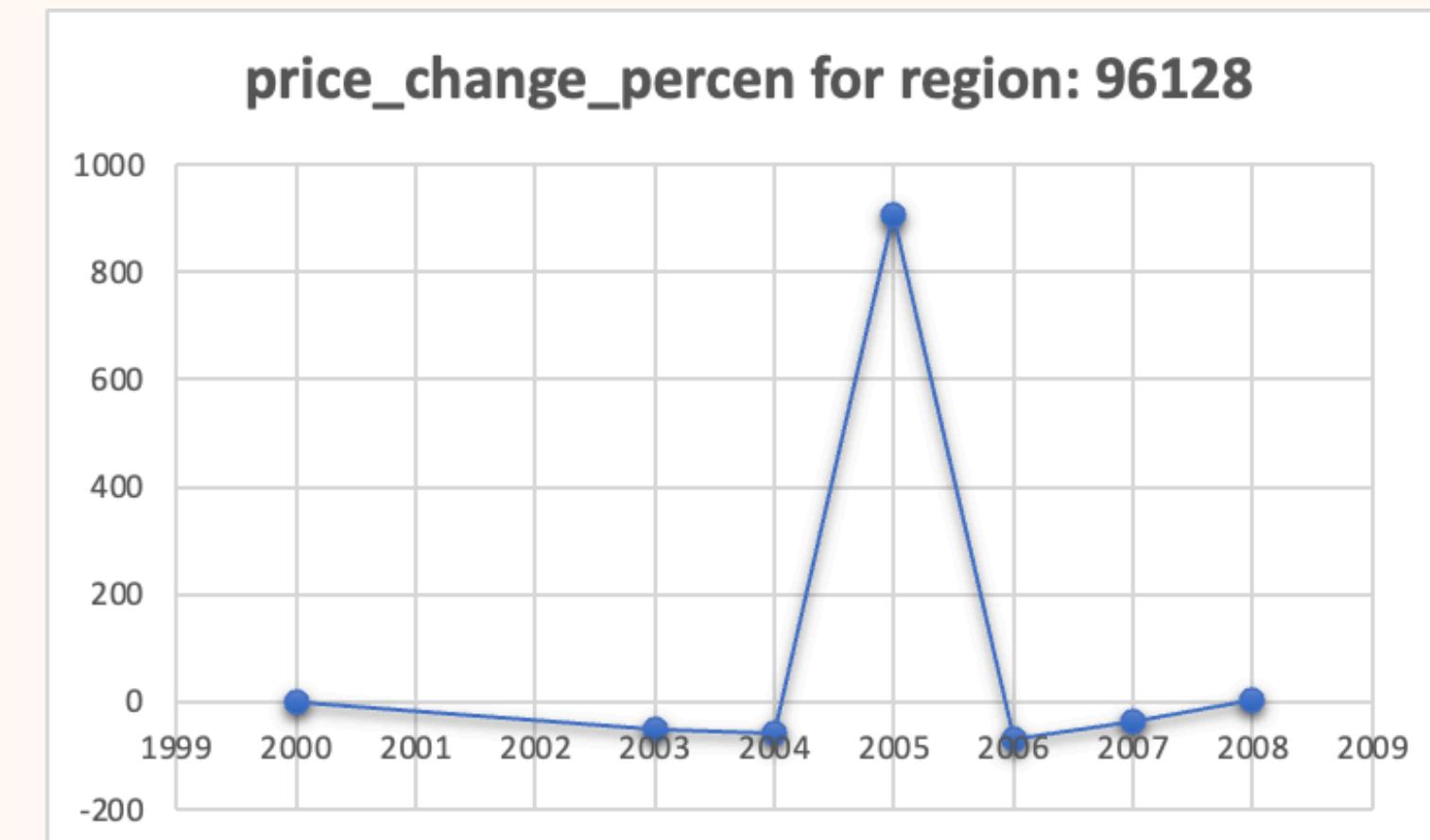
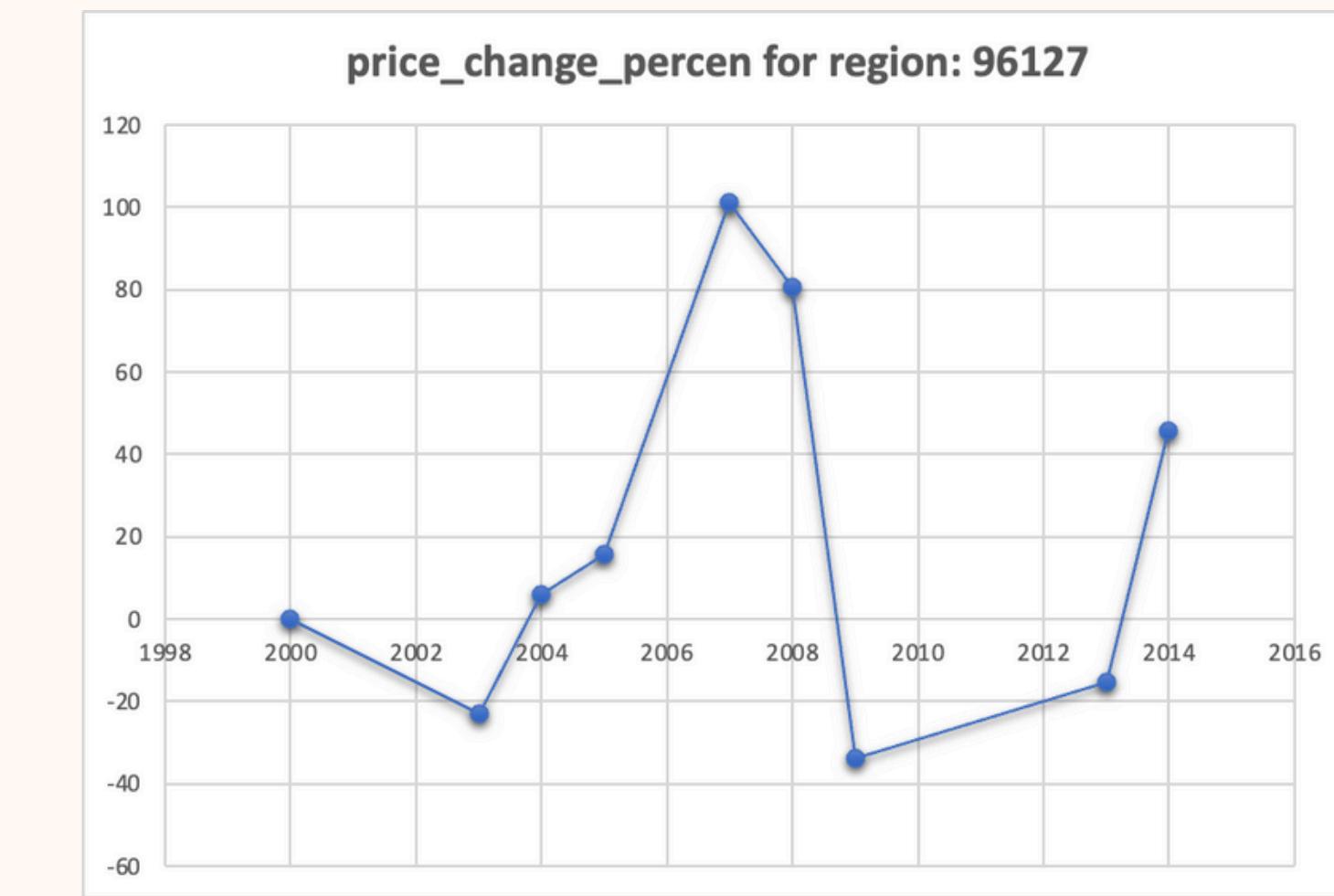
9. Find number of properties and their average price in the top 5% of their city by price

- Region 10389 leads with an average top-tier property price of \$16.7M, indicating extreme high-value properties
- Other high-value regions include 12520 (\$11.99M) and 38980 (\$6.26M), reflecting their premium housing markets.



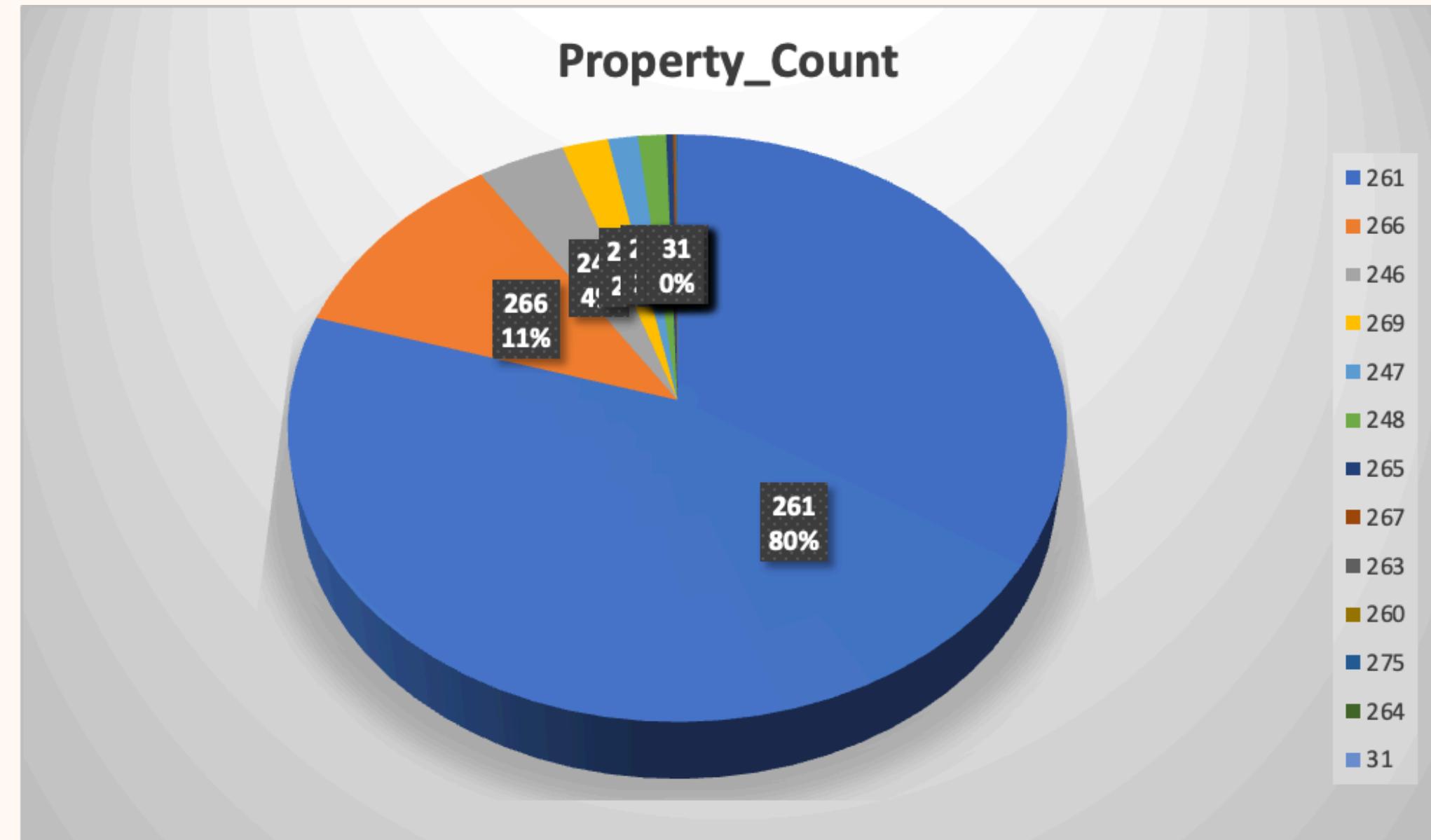
10. Calculate year-over-year price changes for any two cities in California. Consider years from 2000 to 2015

- Region 96127 experienced a 101.32% price increase in 2007, reflecting a boom period.
- In 2003, prices in the same region dropped by 23.02%, showing significant volatility.
- These fluctuations highlight market risks and the importance of timing in property investments.



11. What is the percentage share of properties in each property type for the entire dataset, and how does it rank among other types?

- Property type 261 dominates the dataset (79.69%), indicating a strong market preference.
- Secondary types like 266 (11.09%) and 246 (3.99%) cater to niche demands.
- The dominance of a few property types suggests limited diversity in housing options across the dataset.



THANK YOU!

