DATA SCIENCE 2024

EXPLORATORY DATA ANALYSIS

Villamor , Kurt Russsel
KurtyMittens – GITHUB

DATASET

▲ 278

New Notebook

⬇ Download (781 kB) ▼

Housing Price Prediction Data

Unleash Your Data Science Skills with a Diverse Dataset

Data Card   Code (57)   Discussion (6)   Suggestions (0)

About Dataset

Explore the fascinating world of housing price prediction with this synthetic dataset. Perfect for data science enthusiasts, machine learning practitioners, and Kaggle learners, this dataset offers a diverse collection of features, including square footage, bedrooms, bathrooms, neighborhood types, and the year of construction. Immerse yourself in the challenge of predicting house prices and enhance your skills in regression analysis.

Usability ⓘ
10.00

License
Other (specified in description)

Expected update frequency
Annually

Tags

Computer Science   Text

Classification

Cities and Urban Areas

DATASET

	SquareFeet	Bedrooms	Bathrooms	Neighborhood	YearBuilt	Price
0	2126	4	1	Rural	1969	215355.283618
1	2459	3	2	Rural	1980	195014.221626
2	1860	2	1	Suburb	1970	306891.012076
3	2294	2	1	Urban	1996	206786.787153
4	2130	5	2	Suburb	2001	272436.239065
...	...	...	...	...	...	...
49995	1282	5	3	Rural	1975	100080.865895
49996	2854	2	2	Suburb	1988	374507.656727
49997	2979	5	3	Suburb	1962	384110.555590
49998	2596	5	2	Rural	1984	380512.685957
49999	1572	5	3	Rural	2011	221618.583218

50000 rows × 6 columns

END GOAL

1. Clean The data for it to be ready for analysis

2. Run down a Descriptive Statistics
   - Mean, Median, Mode
   - Standard Variation
   - Range, Variance
   - Correlation Coefficient

3. Provide Analysis
   - Any Patterns?
   - What Housing Neighborhood is High demand?

Housing Prices Data frame

```
1 housing_price_df # Checking dataframe with the Price per square feet
```

	SquareFeet	Neighborhood	YearBuilt	Price	Price/SqFeet
0	2126	Rural	1969	215355.283618	101.295994
1	2459	Rural	1980	195014.221626	79.306312
2	1860	Suburb	1970	306891.012076	164.995168
3	2294	Urban	1996	206786.787153	90.142453
4	2130	Suburb	2001	272436.239065	127.904338
...	...	...	...	...	...
49995	1282	Rural	1975	100080.865895	78.066198
49996	2854	Suburb	1988	374507.656727	131.222024
49997	2979	Suburb	1962	384110.555590	128.939428
49998	2596	Rural	1984	380512.685957	146.576535
49999	1572	Rural	2011	221618.583218	140.978743

50000 rows × 5 columns

DATAFRAMES

DATAFRAMES

01 **Rural Housing Price Growth**

```
1 rural_price_df # data of Rural Houses
```

	SquareFeet	Neighborhood	YearBuilt	Price	Price/SqFeet
0	2126	Rural	1969	215355.283618	101.295994
1	2459	Rural	1980	195014.221626	79.306312
7	2044	Rural	1957	184992.321268	90.505050
16	2123	Rural	1956	190773.148563	89.860174
19	1130	Rural	1962	143050.201782	126.593099
...	...	...	...	...	...
49987	2237	Rural	2011	227359.621341	101.635951
49992	1818	Rural	1983	163683.675434	90.035025
49995	1282	Rural	1975	100080.865895	78.066198
49998	2596	Rural	1984	380512.685957	146.576535
49999	1572	Rural	2011	221618.583218	140.978743

16676 rows × 5 columns

O2 Suburbs Housing Price Growth

DATAFRAMES

```
1 suburb_price_df # Data of Suburb Houses
```

	SquareFeet	Neighborhood	YearBuilt	Price	Price/SqFeet
2	1860	Suburb	1970	306891.012076	164.995168
4	2130	Suburb	2001	272436.239065	127.904338
5	2095	Suburb	2020	198208.803907	94.610408
6	2724	Suburb	1993	343429.319110	126.075374
10	1466	Suburb	1951	191113.768679	130.364099
...	...	...	...	...	...
49989	2613	Suburb	1992	233026.843614	89.179810
49990	2724	Suburb	1978	395026.724290	145.017153
49991	1952	Suburb	1981	200237.485880	102.580679
49996	2854	Suburb	1988	374507.656727	131.222024
49997	2979	Suburb	1962	384110.555590	128.939428

16721 rows × 5 columns

DATAFRAMES
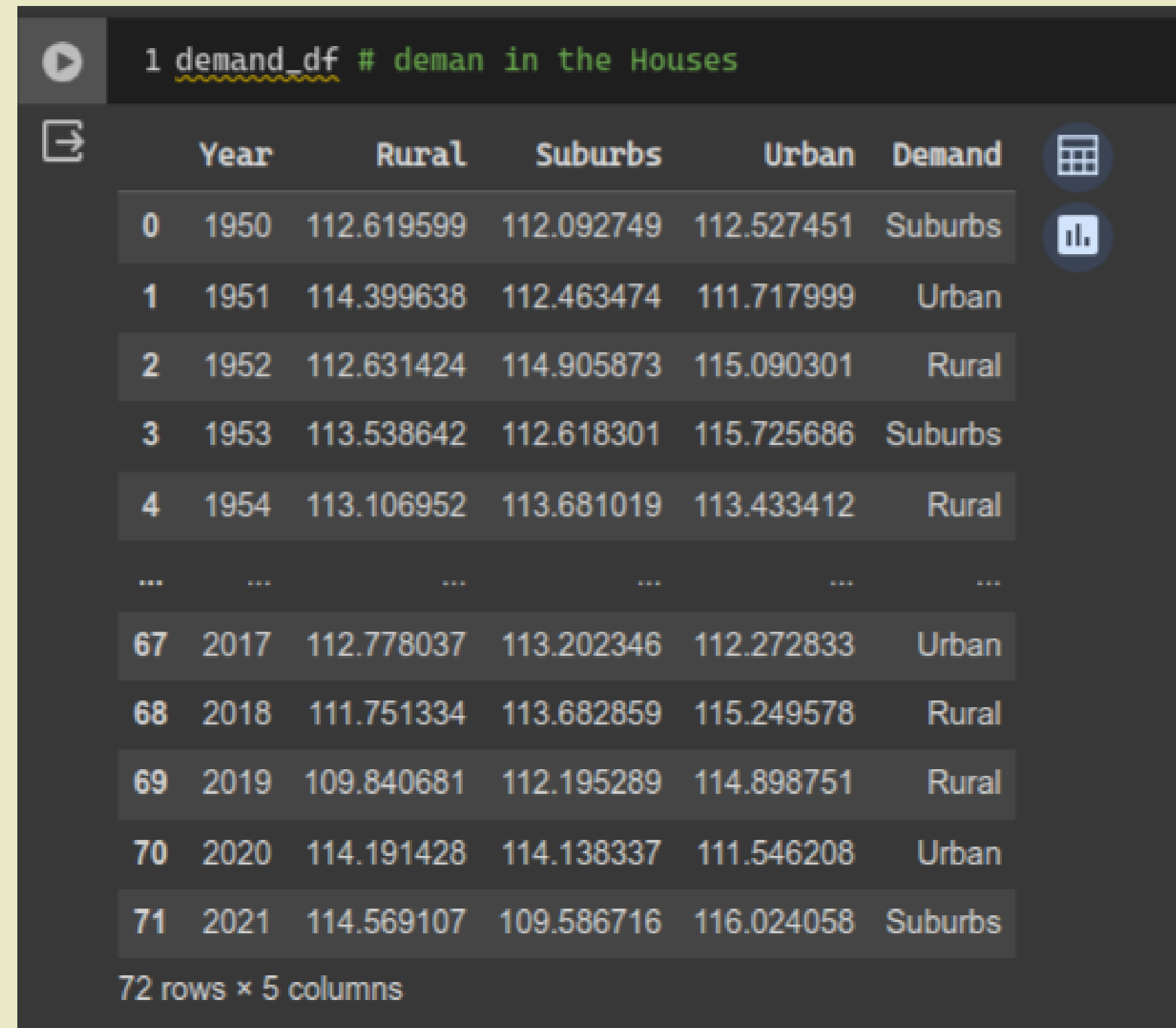
```
1 urban_price_df #Data of Urban Houses
```

	SquareFeet	Neighborhood	YearBuilt	Price	Price/SqFeet
3	2294	Urban	1996	206786.787153	90.142453
8	2638	Urban	1959	377998.588152	143.289836
9	1121	Urban	2004	95961.926014	85.603859
14	1087	Urban	1976	118393.823163	108.917961
18	2687	Urban	1979	239222.667797	89.029649
...	...	...	...	...	...
49972	2009	Urban	2000	251924.652151	125.398035
49975	2057	Urban	1973	194964.941703	94.781206
49984	2967	Urban	1960	345653.603641	116.499361
49993	1215	Urban	2017	126052.390562	103.746824
49994	2426	Urban	1950	331497.091307	136.643484

16603 rows × 5 columns

DATAFRAMES

04 **Demand of the Housing preferences base in the prices**

```
1 demand_df # deman in the Houses
```

	Year	Rural	Suburbs	Urban	Demand
0	1950	112.619599	112.092749	112.527451	Suburbs
1	1951	114.399638	112.463474	111.717999	Urban
2	1952	112.631424	114.905873	115.090301	Rural
3	1953	113.538642	112.618301	115.725686	Suburbs
4	1954	113.106952	113.681019	113.433412	Rural
...	...	...	...	...	...
67	2017	112.778037	113.202346	112.272833	Urban
68	2018	111.751334	113.682859	115.249578	Rural
69	2019	109.840681	112.195289	114.898751	Rural
70	2020	114.191428	114.138337	111.546208	Urban
71	2021	114.569107	109.586716	116.024058	Suburbs

72 rows × 5 columns

ANALYSIS

Rural Housing Prices

Mean (Sq Feet):

```
[52]    1 rural_price_df["SquareFeet"].mean() # Mean of the Square feet of Rural Houses

        2001.621132165987
```

Mean (price / Sq Feet):

```
[29]    1 rural_growth_df['GrowthMeanPriceSq'].mean() # Mean of Rural Prices per square feet

        113.23353993450549
```

Price Range:

```
[61]    1 print("Max:", rural_growth_df['GrowthMeanPriceSq'].max())
        2 print("Min:", rural_growth_df['GrowthMeanPriceSq'].min())
        3 rural_growth_df['GrowthMeanPriceSq'].max() - rural_growth_df['GrowthMeanPriceSq'].min()

        Max: 117.61810037786418
        Min: 108.84974916407697
        8.768351213787213
```

ANALYSIS

```
1 rural_growth_df # The Rural Growth Price DataFrame in 72 years
```

	Year	GrowthMeanPriceSq
0	1950	112.619599
1	1951	114.399638
2	1952	112.631424
3	1953	113.538642
4	1954	113.106952
...	...	...
67	2017	112.778037
68	2018	111.751334
69	2019	109.840681
70	2020	114.191428
71	2021	114.569107

72 rows × 2 columns

## Rural Growth Prices

- in this Forecast/Growth of price we can see a variety of prices that is not linear to positive or negative relationship

- Supported by the Low Correlation Coefficient

	Year	GrowthMeanPriceSq
Year	1.000000	-0.102064
GrowthMeanPriceSq	-0.102064	1.000000

ANALYSIS
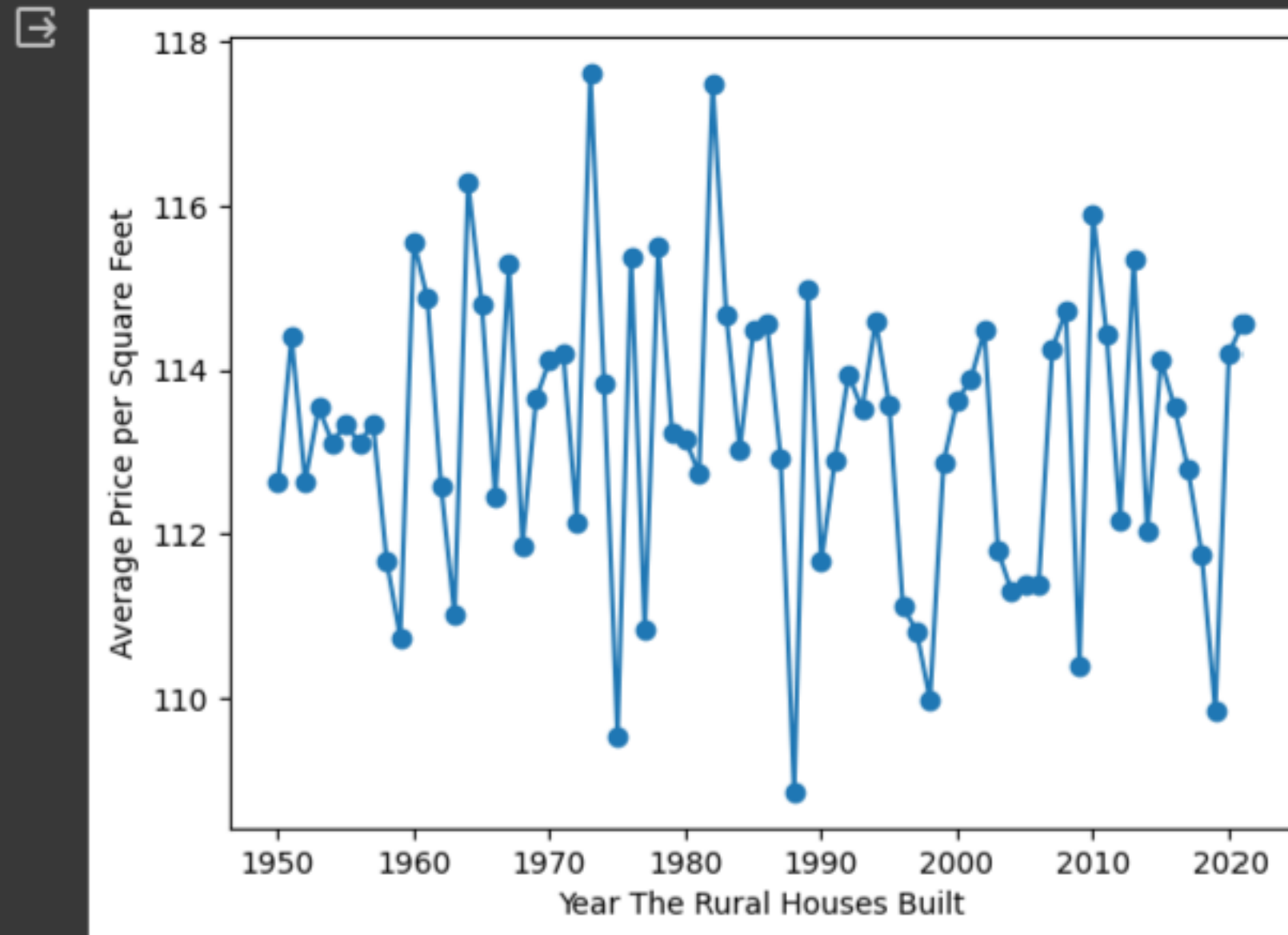
Suburbs Housing Prices

Mean (Sq Feet):

```
[53]    1 suburb_price_df["SquareFeet"].mean() # Mean of the Square feet of Rural Houses

     2000.1511273249207
```

Mean (price / Sq Feet):

```
[30]    1 suburbs_growth_df['GrowthMeanPriceSq'].mean() # Mean of Suburb Prices per square feet

     112.79938183777159
```

Price Range:

```
[63]    1 print("Max:", suburbs_growth_df['GrowthMeanPriceSq'].max())
        2 print("Min:", suburbs_growth_df['GrowthMeanPriceSq'].min())
        3 suburbs_growth_df['GrowthMeanPriceSq'].max() - suburbs_growth_df['GrowthMeanPriceSq'].min()

     Max: 116.44782522079875
     Min: 108.62092932567131
     7.82695895127436
```

ANALYSIS

```
1 suburbs_growth_df # The Suburbs Growth Price DataFrame in 72 years
```

	Year	GrowthMeanPriceSq
0	1950	112.092749
1	1951	112.463474
2	1952	114.905873
3	1953	112.618301
4	1954	113.681019
...	...	...
67	2017	113.202346
68	2018	113.682859
69	2019	112.195289
70	2020	114.138337
71	2021	109.586716

72 rows × 2 columns

Suburbs Growth Prices

- in this Forecast/Growth of price we can see a variety of prices that is not linear to positive or negative relationship

- Supported by the Low Correlation Coefficient

	Year	GrowthMeanPriceSq
Year	1.000000	-0.097245
GrowthMeanPriceSq	-0.097245	1.000000

ANALYSIS

Suburbs Growth Prices

```
1 plt.plot(suburbs_growth_df["Year"], suburbs_growth_df['GrowthMeanPriceSq'], marker = "o")
2 plt.xlabel("Year The Suburbs Houses Built")
3 plt.ylabel("Average Price per Square Feet")
4 plt.show()
```

- Supported also in this graph.

ANALYSIS
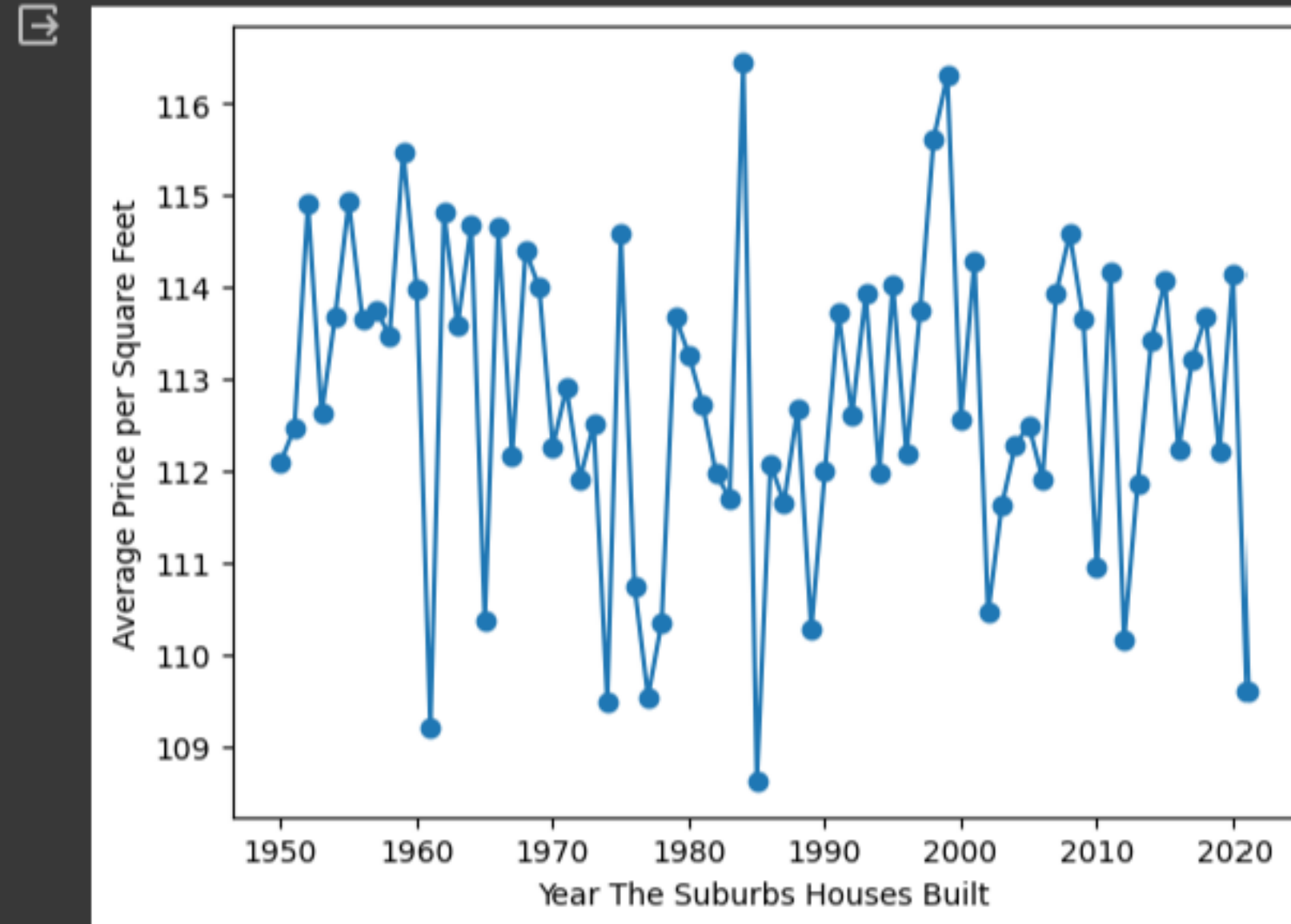
Urban Housing Prices

Mean (Sq Feet):

```
[54]    1 urban_price_df["SquareFeet"].mean() # Mean of the Square feet of Rural Houses

        2017.4169126061556
```

Mean (price / Sq Feet):

```
[31]    1 urban_growth_df['GrowthMeanPriceSq'].mean() # Mean of Urban Prices per square feet

        113.85089125000715
```

Price Range:

```
[65]    1 print("Max:", urban_growth_df['GrowthMeanPriceSq'].max())
        2 print("Min:", urban_growth_df['GrowthMeanPriceSq'].min())
        3 urban_growth_df['GrowthMeanPriceSq'].max() - urban_growth_df['GrowthMeanPriceSq'].min()

        Max: 117.91030160912347
        Min: 109.59635845632064
        8.31394315280282
```

ANALYSIS

```
1 urban_growth_df # The Urban Growth Price
```

	Year	GrowthMeanPriceSq
0	1950	112.527451
1	1951	111.717999
2	1952	115.090301
3	1953	115.725686
4	1954	113.433412
...	...	...
67	2017	112.272833
68	2018	115.249578
69	2019	114.898751
70	2020	111.546208
71	2021	116.024058

72 rows × 2 columns

Urban Growth Prices

- in this Forecast/Growth of price we can see a variety of prices that is not linear to positive or negative relationship

- Supported by the Low Correlation Coefficient

ANALYSIS

Urban Growth Prices


```python
1 plt.plot(urban_growth_df["Year"], urban_growth_df['GrowthMeanPriceSq'], marker = "o")
2 plt.xlabel("Year The Suburbs Houses Built")
3 plt.ylabel("Average Price per Square Feet")
4 plt.show()
```
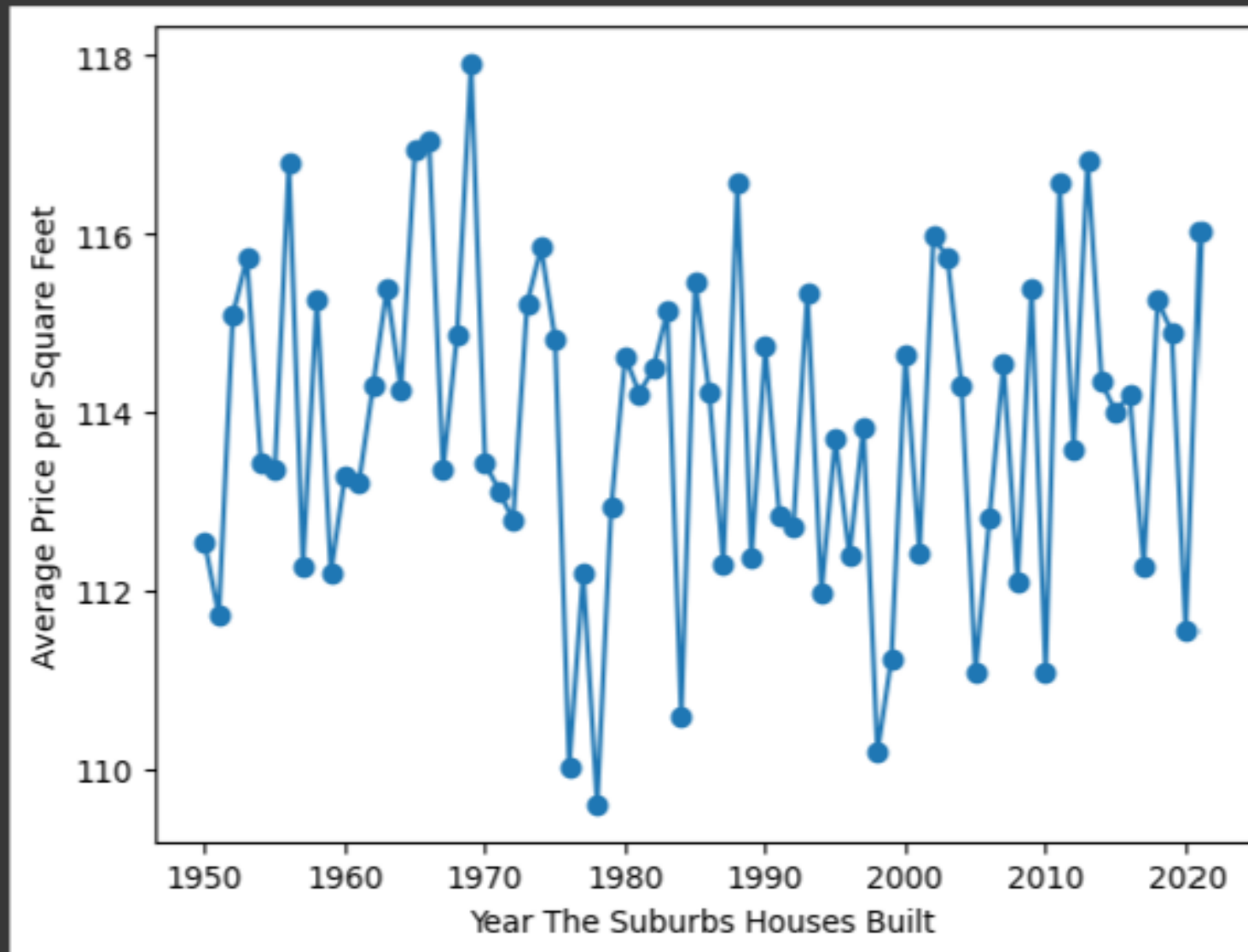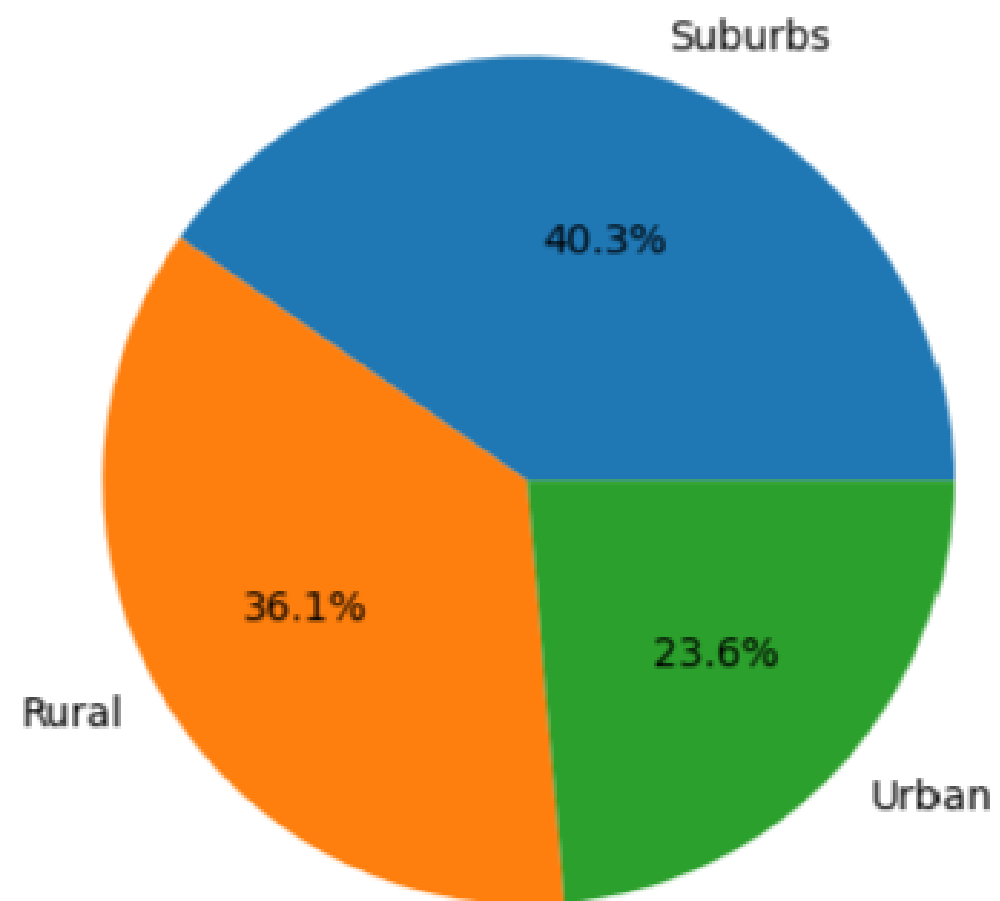
- Supported also in this graph.

THANK YOU

Villamor, Kurt Russel
KurtyMittens