Determine if local populations impact local rent prices in the Peninsula, by using One-Way ANOVA Test

Donald Cao

It is expensive to buy a home in the SF Bay Area. Even true for renting! The most expensive places to rent is the San Mateo County, situated along the peninsula. However, there are differences in the rental prices throughout that region. What is influencing the rental price? Local population? Let's find out!

Introduction

In SF Bay Area, the most expensive county to rent at is San Mateo County. It is also known as the peninsula leading up to S.F. Encompassing Mill Brae to the north and ends at Mountain view to the south, while aligning against the hills in the west and the bay to the east. At the north and south terminals, cities' rent prices are seen to be their least expensive, but when you travel towards the center of the peninsula, rent prices are assumed to be more expensive. But how did it become expensive? The Bay Area is a metropolis and many people resides in each city, so of course prices are considered expensive. However, prices are also seen least expensive in most the populated areas, especially in Redwood City. Redwood City is a city situated near the center of the Peninsula and most rent prices are considered least expensive.

Contact
Email:
cao.donald89@gmail.com

GitHub:
https://github.com/bellycose/Apartment_Prices

So, a null hypothesis is declared:

Local rental prices are not impacted by local populations. If local population is large, rent prices remain relatively the same or no changes and may be influenced by other factors.

While the alternative hypothesis:

At least two or more local rental prices are influenced by the population. If local population is large, rent prices changes, either increase or decrease depends on local population.

Methods

If visuals are difficult to see, please feel free to check my Git-Hub repository: Apartment_Prices/supplementals/imgs for a better experience to see the visuals.

Used python for web scraping on apartmentlist.com/ca/redwood-city [1]. Specifically, established TCP connection with Request framework and import sleep module for maintaining reliable message exchanges. And parsed objects from HTML with Beautiful Soup (BS4) framework. From *Figure 1*, after parsing objects, processed "Ask" into integer "0" under Price category to keep record of available sample population and maintain relevant data type. With Pandas and NumPy, converted data into data frames or arrays respectively for processing objects into data more efficiently for direct use. After tidying data, saved data as CSV as seen from *Figure 2*. Once finished for Redwood City, repeat the whole web scraping again for Menlo Park, Mountain View, and San Mateo.

```python
from bs4 import BeautifulSoup as bs
import requests
import re
from time import sleep
from random import randint

s=requests.Session()
def price():
    baseURL='https://www.apartmentlist.com/ca/redwood-city'
    r=s.get(baseURL)
    soup=bs(r.content,'html.parser')

    block=soup.find_all('div',class_='css-1u6cvl9 e1k7pw6k0')
    sleep(randint(2,10))

    result=[]
    for properties in block:
        priceBlock=properties.find_all('div',class_="css-q23zey e13lnafx0")
        price=[price.text for price in priceBlock]
        strPrice=''.join(price) #Change from list to string type
        removed=r'[$]' #Select and remove $ characters
        removed2=r'Ask' #Select and remove Ask
        removed3=r'[,]' #Select and remove comma
        modPrice=re.sub(removed,' ',strPrice)#Substitute $ for '_'
        modPrice2=re.sub(removed2,' 0',modPrice)#Substitute Ask for _0
        modPrice3=re.sub(removed3,'',modPrice2)#Eliminate space within price
        segments=modPrice3.split()#Change string with updates into list, remain

        result+=[insert for insert in segments]
    return result


price()
```

Fig 1. Web scraping and cleaning data for Rent Prices in Redwood City.

```python
from RESOURCE import ADDRESS as address
from RESOURCE import APARTMENTNAME as apartment
from RESOURCE import CITY as city
from RESOURCE import PRICE as price
from RESOURCE import BEDROOM as bedroom
import pandas as pd
import numpy as np

#DataFrames
rc=pd.DataFrame({
    'Apartment':apartment.apartmentName(),
    'Address':address.address(),
    'City':city.city(),
    })
#return rc
print(rc)

rcBed=pd.DataFrame({
    'Bedroom':bedroom.bedroom(),
    'Price':price.price(),
    })
#return rcBed
print(rcBed)

#CSV
df1rc=rc.to_csv(r'D:\Programming\Portfolio\Project 4 - Apartment\Redwood_City\Apartment(Redwood_City).csv')
df2rc=rcBed.to_csv(r'D:\Programming\Portfolio\Project 4 - Apartment\Redwood_City\Bedroom(Redwood_City).csv')
```

Fig 2. Convert data frame to CSV. Bedroom and Price data now grouped and saved in CSV with in respective cities. An example would be titled “Bedroom(Redwood).csv.

Descriptive Statistics

Summarized data with describe method as seen in *Figure 3*.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math
import statistics

#Imports
#Redwood City
dfrc=pd.read_csv(r'Redwood_City\Apartment(Redwood_City).csv',index_col=0)
bdrc=pd.read_csv(r'Redwood_City\Bedroom(Redwood_City).csv',index_col=0)
#Menlo Park
dfmp=pd.read_csv(r'Menlo_Park\Apartment(Menlo_Park).csv',index_col=0)
bdmp=pd.read_csv(r'Menlo_Park\Bedroom(Menlo_Park).csv',index_col=0)
#Mountain View
dfmv=pd.read_csv(r'Mountain_View\Apartment(Mountain_View).csv',index_col=0)
bdmv=pd.read_csv(r'Mountain_View\Bedroom(Mountain_View).csv',index_col=0)
#San Mateo
dfsm=pd.read_csv(r'San_Mateo\Apartment(San_Mateo).csv',index_col=0)
bdsm=pd.read_csv(r'San_Mateo\Bedroom(San_Mateo).csv',index_col=0)
'''
This file will manage data from list to dataframe and analyze with statistics.
'''

#Population Bedroom Available
frame=[bdrc,bdmp,bdmv,bdsm]
result=pd.concat(frame)
#print("Vacancy in all selected Apartments:\n",result)
popPrices=result["Price"]
pop=popPrices.describe()
#print(f'Population:\n{pop}\n')

#Sample Bedroom Available -
bdrcPrice=bdrc.describe()
bdmpPrice=bdmp.describe()
bdmvPrice=bdmv.describe()
bdsmPrice=bdsm.describe()

print(f'Redwood City:\n{bdrcPrice}\n')
print(f'Menlo Park:\n{bdmpPrice}\n')
print(f'Mountain View:\n{bdmvPrice}\n')
print(f'San Mateo:\n{bdsmPrice}\n')
```

Fig 3. Using describe method to acquire data summary for descriptive statistics.

Focused visualization code of price for each city in *Figure 4 and 5*. Bedroom was removed due to being unnecessary onward.

Import the matplotlib library was essential because matplotlib utilized arrays only, a data type by NumPy. In figure 4, under “Create Arrays” was used to change data type into array follow with plotting boxplot. Result can be seen for *figure 8*.

In boxplots, the inclusion of a fifth entity, “All”, which represented the total list of observed samples from each city. By having “All”, would have given other samples’ characteristics to be compared with. Characteristic would consist of but not limited to visual ques about variance between sample and means comparable to each cities’ properties (mean, median, population standard dev., and IQR). And these characteristics would be further calculated for F-test, Z-test, and P-value (P-test).

Boxplot

```python
#Population from each cities
frame=[bdrc,bdmp,bdmv,bdsm]
result=pd.concat(frame)

#Create Arrays
dfpop=np.array(result['Price'])
dfRC=np.array(bdrc['Price'])
dfMP=np.array(bdmp['Price'])
dfMV=np.array(bdmv['Price'])
dfSM=np.array(bdsm['Price'])
#print(dfpop)
#print(dfRC)
#print(dfMP)
#print(dfMV)
#print(dfSM)

#Boxplot
fig1,ax1,ax2,ax3,ax4,ax5=plt.boxplot([dfpop,dfRC,dfMP,dfMV,dfSM],
                                     labels=['All',
                                             'Redwood City',
                                             'Menlo Park',
                                             'Mountain View',
                                             'San Mateo'],
                                     vert=True,
                                     showmeans=True,
                                     sym='b.',
                                     showfliers=True,
                                     autorange=False,)

plt.ylabel("Price in USD($)\n",
           fontweight='bold',
           fontsize=12,
           )
plt.title("Summary of Prices in Each Cities",
          fontweight='bold',
          fontsize=16)

plt.show()
```

Fig 4. Boxplot code of the entire "Price" from all cities and "All" where the total sample population is seen as a fifth entity.

Histogram & Stack Histogram

In histogram and stacked histogram, "All" entity was removed as it distorted the graph. Results seen in *Figure 9*.

```python
#Population from each cities
frame=[bdrc,bdmp,bdmv,bdsm]
result=pd.concat(frame)

#Create Arrays
dfpop=np.array(result['Price'])
dfRC=np.array(bdrc['Price'])
dfMP=np.array(bdmp['Price'])
dfMV=np.array(bdmv['Price'])
dfSM=np.array(bdsm['Price'])

#Histogram
plt.hist([dfRC,dfMP,dfMV,dfSM],
         10,
         range=(1000,10000),
         ec='k',
         label=['Redwood City',
                'Menlo Park',
                'Mountain View',
                'San Mateo'],
         color=['orange','limegreen','crimson','mediumpurple'],
         align='left',
         stacked=False)
plt.title("Summary Comparision in Histogram without ASK/$0 & Control",
          fontweight='bold',fontsize=16)
plt.xlabel("Price in USD($)",fontweight='bold',fontsize=12)
plt.ylabel("Frequency",fontweight='bold',fontsize=12)
plt.xticks(np.arange(1000,11000,1000))

plt.legend()
plt.show()
```

Fig 5. Code to plot Histogram with "All". Notice stacked=False (gold-yellow).

Like in *Figure 5*, but the argument stacked=False was changed to True to display stacked Histogram. See result from *Figure 10*.

Inferential Statistics by Test Hypothesis

From having analyzed the visualization of histograms and boxplots while only being analyzing "Price", had led to testing hypothesis by One-Way ANOVA. Since histogram visuals have shown favoring right tailed skewed graph, a F-test was necessary as it covers each graphs' variances and their means while accounting the total variance and mean to determine whether to reject hypothesis or not. This was done on python (*Fig.6*) and organized in Excels (*Fig.12*).

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math
import statistics

#Imports
#Redwood City
dfrc=pd.read_csv(r'Redwood_City\Apartment(Redwood_City).csv',index_col=0)
bdrc=pd.read_csv(r'Redwood_City\Bedroom(Redwood_City).csv',index_col=0)
#Menlo Park
dfmp=pd.read_csv(r'Menlo_Park\Apartment(Menlo_Park).csv',index_col=0)
bdmp=pd.read_csv(r'Menlo_Park\Bedroom(Menlo_Park).csv',index_col=0)
#Mountain View
dfmv=pd.read_csv(r'Mountain_View\Apartment(Mountain_View).csv',index_col=0)
bdmv=pd.read_csv(r'Mountain_View\Bedroom(Mountain_View).csv',index_col=0)
#San Mateo
dfsm=pd.read_csv(r'San_Mateo\Apartment(San_Mateo).csv',index_col=0)
bdsm=pd.read_csv(r'San_Mateo\Bedroom(San_Mateo).csv',index_col=0)

#Total Sum of Square
#Population
frame=[bdrc,bdmp,bdmv,bdsm]
result=pd.concat(frame)
popPrices=result["Price"]

#Grand Mean
xpop=popPrices.mean()
#print(xpop) #2601.07

#Sum of Square Total
ss=[math.pow(popsq-xpop,2) for popsq in popPrices]
#print(sum(ss)) #472349857.91


#Sum of Square Within
bdrcPrice=bdrc['Price']
bdmpPrice=bdmp['Price']
bdmvPrice=bdmv['Price']
bdsmPrice=bdsm['Price']
#Cities Mean Square
xrc=bdrcPrice.mean()
xmp=bdmpPrice.mean()
xmv=bdmvPrice.mean()
xsm=bdsmPrice.mean()
ssrc=[math.pow(poprc-xrc,2) for poprc in bdrcPrice]
ssmp=[math.pow(popmp-xmp,2) for popmp in bdmpPrice]
ssmv=[math.pow(popmv-xmv,2) for popmv in bdmvPrice]
sssm=[math.pow(popsm-xsm,2) for popsm in bdsmPrice]

print(sum(ssrc))#76228484.82
print(sum(ssmp))#166049835.64
print(sum(ssmv))#90451628.18
print(sum(sssm))#121274272.85
```

Fig 6. Calculation of Sum of Squares Within and Between for both cities and "All" for F-Statistics.

Then calculated a Z-test and compared to a Z-Score at significance level 5% and 10% to verify if rejecting null hypothesis. The Z score was founded on a t-table. Afterward, used Z-test value to convert p-value [4] to further verify both Z and F test results at significance level of 5% and 10% again. Any calculation and organizing was done on Excels seen in *Figure 13 and 14*.

Trimming Sample for Accuracy

Finally, trimmed samples of 5%, 10%, and 25% to verify statistic tests by ridding off extreme outliers. As seen from *Figure 7*, from arrays, then sorted, and finally trimmed after checking the lengths from a product of the trim to determine the amount of price to remove.

This followed with a repeat of the entire descriptive and inferential statistic process that were done previously. Results from *Figure 8 to 14* can be found.

```python
#Population from each cities
frame=[bdrc,bdmp,bdmv,bdsm]
result=pd.concat(frame)

#Create Arrays
dfpop=np.array(result['Price'])
dfRC=np.array(bdrc['Price'])
dfMP=np.array(bdmp['Price'])
dfMV=np.array(bdmv['Price'])
dfSM=np.array(bdsm['Price'])

#Sorted in Order
pop=sorted(dfpop)
RC=sorted(dfRC)
MP=sorted(dfMP)
MV=sorted(dfMV)
SM=sorted(dfSM)

#Length
#print(len(pop))#207-20
#print(len(RC))#50-4
#print(len(MP))#36-4
#print(len(MV))#60-6
#print(len(SM))#61-6

#Calculate for 5% Trimming
vpop=len(pop)*.05
vrc=len(RC)*.05
vmp=len(MP)*.05
vmv=len(MV)*.05
vsm=len(SM)*.05

#print(round(vpop))#10 both sides
#print(round(vrc))#2
#print(round(vmp))#2
#print(round(vmv))#3
#print(round(vsm))#3

#Trimming at 5%
'''
Test:
leftpop=pop[:10]#10
rightpop=pop[-10:]#10
trimpop=pop[10:-10]
print(trimpop)
'''
trimpop=pop[10:-10]
trimrc=RC[2:-2]
trimmp=MP[2:-2]
trimmv=MV[3:-3]
trimsm=SM[3:-3]

#Trimmed Length
#print(len(trimpop))
#print(len(trimrc))
#print(len(trimmp))
#print(len(trimmv))
#print(len(trimsm))

#Convert List to Data Frame
v=pd.DataFrame(trimpop).describe()
w=pd.DataFrame(trimrc).describe()
x=pd.DataFrame(trimmp).describe()
y=pd.DataFrame(trimmv).describe()
z=pd.DataFrame(trimsm).describe()

#Summary at 10% Trim
print(f'Population 5%:\n{v}\n\n\
Redwood City 5%:\n{w}\n\n\
Menlo Park 5%:\n{x}\n\n\
Mountain View 5%:\n{y}\n\n\
San Mateo 5%:\n{z}\n\n')
```

Fig 7. Trimming program code with python. *Allows sample collection to be trim at both terminal ends by the amount referring to the trim level for each entity, "All" and cities.*

Results and Evaluations

Boxplots (0%, 5%, 10%, and 25%)

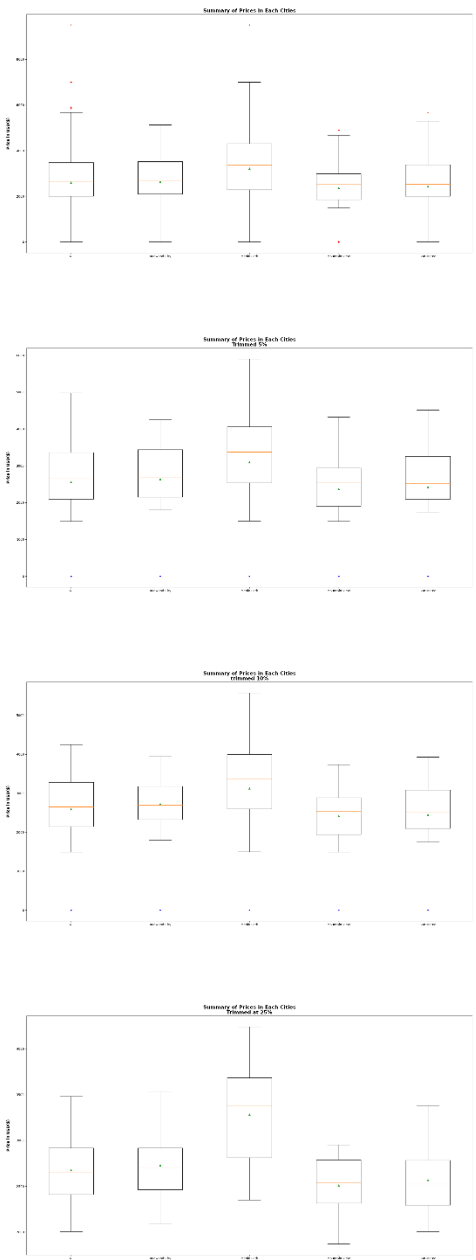

Fig 8. Shows boxplots with 0% - 25% trim levels from top to bottom respectively. *From left to right, All, Redwood City, Menlo Park, Mountain View, and San Mateo.*

These boxplots showed a quick overview of the samples' median, mean, and sample population variances. The purpose was to check if extreme outlier significantly changes plots after trimming and resulted insignificant changes.

Noted that 25% trim level, had shown all extreme minimum outliers, including zeros, has been removed at the expense of sample size. Thus, the histogram looked normalized.

While the rest looked relatively the same. Menlo Park having the highest price ¾ IQR and it's mean relatively behind the median at no trim level. And as trim level increase toward 25%, a slight increase at 5% and maintained the same distance in 15%. This had suggested that there were slightly more rental prices in the extreme, which held median price above the mean despite more rooms were priced within the box.

In San Mateo and Redwood City showed mean closely under median price at 0% trim. As the trim level increased to 25%, the mean will be over the median at 10% for Redwood City and 25% for San Mateo. Suggested that the with the extreme price did influence the graph and suggested that 10% and 25% trim would give a more accurate account when running a statistical test. And that there were more rental properties priced within the IQR for both cities, especially a bit more than the median in the 10%. Where the mean is above the median values.

Furthermore, 5% and 10% were seen as the more accurate trim level because Mountain View's lower extreme outlier was picked up as ¼ IQR in the 25% and 0% trim, which would had given a misinformed conclusion about the frequent priced bedrooms available.

With that, 5% and 10% would be considered an accurate trim level when observed during the hypothesis testing.

Histograms (0%, 5%, 10%, 25%)

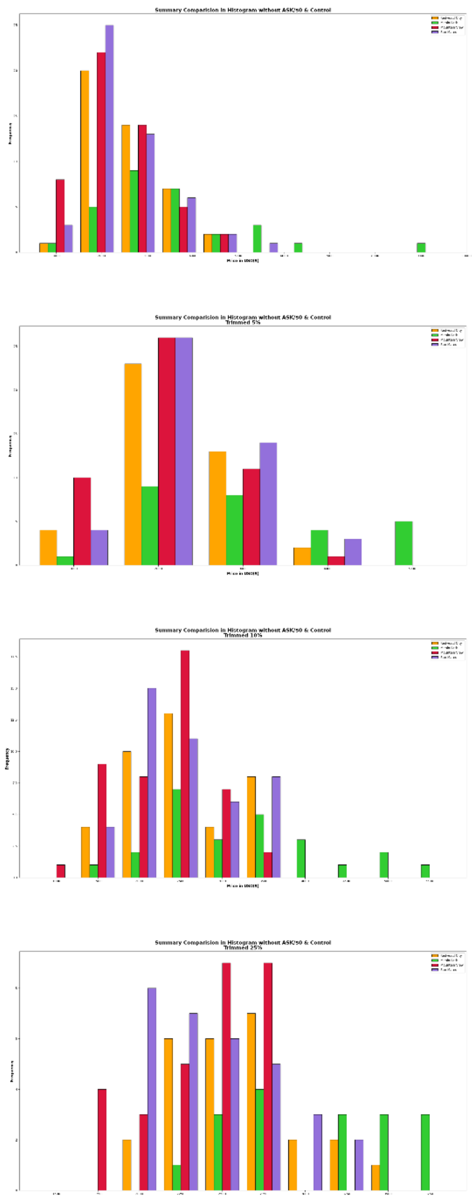

Fig 9. (Top to Bottom) Histogram with 0% - 25% trim levels, respectively. *The priced $0 was removed.*

Immediately, in Histogram with 0% trim shows right tail skewered histogram for each city. This helped with determining the statistic test necessary used to test the hypothesis: One-Way ANOVA test [3]. With 5% and 10% trim, skewed right tail remain intact. At 25%, with most of the extreme outliers had been removed, leaving an unrecognized graph. Correlation can be seen in the box plots (*Fig 8*).

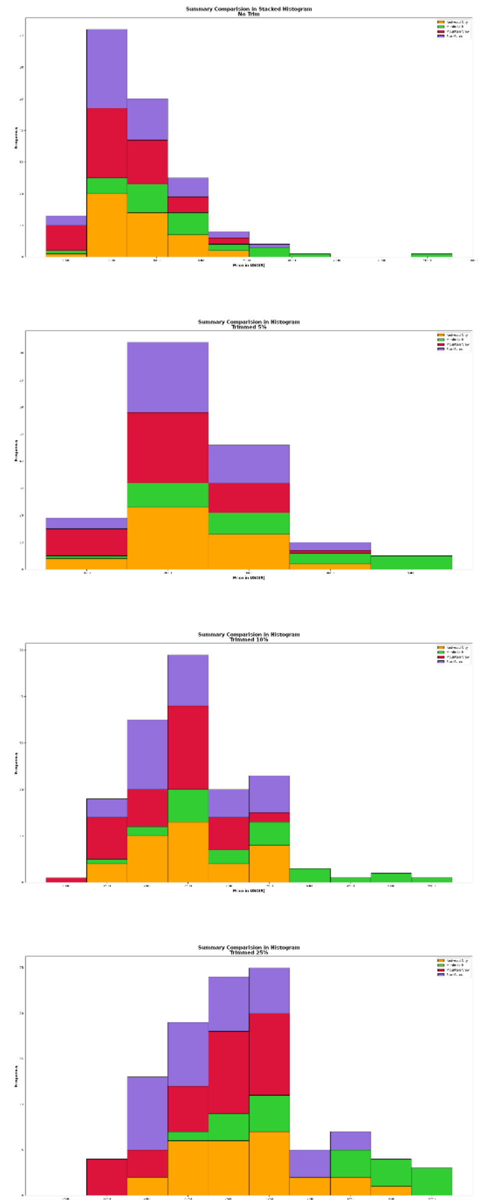

Fig 10. (Top to Bottom) Stacked Histogram with 0% - 25% trim levels, top to bottom respectively. *The price $0 was removed. Orange for Redwood City, Green for San Mateo, Purple for Menlo Park, Red for Mountain View.*

Stacked Histograms (0%, 5%, 10%, 25%)

Figure 10, helped to identify the density or number of bedrooms frequency aligned of the prices in the city

The area of the colored bar represented the density or number of vacant rooms within the city based on their colors (Orange for Redwood City, Green for San Mateo, Purple for Menlo Park, Red for Mountain View) and determined the position on prices. From 0% trim, it's easier to see the graph property, however as the level of the trim increased and sample size decreased, the graph becomes less recognizable to determine its graphs' properties.

In 5% trim, the use of less bins or columns helped coalesced the rental properties more cleanly, but doing so gives doubt about variances. This would mean if the data is accurate through generalization.

In 10% trim, the use of more bins had eliminated the doubts and showed variance from the positive skewed tail. This also meant that 5% trim could be considered as accurate without a doubt.

Table of No Trim													
City	Count (N)	k	dfb	dfw	Mean	SEM	St.Dev. (s)	Min.	Q2	Median	Q3	Max.	IQR
Redwood City	50	4	3	203	2632.940	176.391	1247.270	0	2098.00	2697.50	3522.50	5124.00	1424.50
Menlo Park	36	4	3	203	3213.690	363.023	2178.140	0	2286.25	3375.00	4305.50	9500.00	2019.25
Mountain View	60	4	3	203	2368.720	159.847	1238.170	0	1858.75	2537.50	2993.50	4899.00	1134.75
San Mateo	61	4	3	203	2441.950	182.033	1421.720	0	2005.00	2526.00	3380.00	5669.00	1375.00
Table of 5% Trim													
City	Count (N)	k	dfb	dfw	Mean	SEM	St.Dev. (s)	Min.	Q2	Median	Q3	Max.	IQR
Redwood City	46	4	3	183	2642.240	156.418	1060.880	0	2149.50	2697.50	3444.00	4250.00	1294.50
Menlo Park	32	4	3	183	3099.780	303.084	1714.500	0	2536.75	3375.00	4060.25	5900.00	1523.50
Mountain View	54	4	3	183	2371.350	322.700	2371.350	0	1903.00	2537.50	2940.25	4321.00	1037.25
San Mateo	55	4	3	183	2423.760	163.878	1215.350	0	2092.50	2526.00	3257.00	4513.00	1164.50
Table of 10% Trim													
City	Count (N)	k	dfb	dfw	Mean	SEM	St.Dev. (s)	Min.	Q2	Median	Q3	Max.	IQR
Redwood City	40	4	3	161	2768.170	120.502	762.120	0	2329.50	2697.50	3180.00	3960.00	850.50
Menlo Park	28	4	3	161	3122.960	273.348	1446.420	0	2600.50	3375.00	3996.25	5575.00	1395.75
Mountain View	48	4	3	161	2416.400	119.255	826.220	0	1937.50	2537.50	2889.25	3732.00	951.75
San Mateo	49	4	3	161	2446.650	145.076	1015.530	0	2095.00	2526.00	3083.00	3928.00	988.00
Table of 25% Trim													
City	Count (N)	k	dfb	dfw	Mean	SEM	St.Dev. (s)	Min.	Q2	Median	Q3	Max.	IQR
Redwood City	26	4	3	101	2729.460	74.242	378.560	2095.00	2464.75	2697.50	2918.25	3536.00	453.50
Menlo Park	18	4	3	101	3282.720	134.086	568.880	2350.00	2812.50	3375.00	3682.50	4241.00	870.00
Mountain View	30	4	3	101	2511.230	59.422	325.470	1870.00	2316.25	2537.50	2787.50	2953.00	471.25
San Mateo	31	4	3	101	2569.840	67.717	377.030	2005.00	2293.50	2526.00	2782.50	3380.00	489.00

Fig 11. Table of Price Summary in respects to Trim Level.

Rental Price Summary Table

Last of the descriptive statistics was the summary table. This provided sample's count, mean, standard error mean, standard deviation, minimum, maximum, and the IQR. Optionally, included the degree of freedom from the number of categories which was four. Four because of the number of cities. Calculated for standard error mean (SEM), degree of freedoms, and Interquartile range (IQR).

F Test

F-Distribution Test								
0% Trim	Grand Mean	SStotal	SSb	SSw	SSMb	SSMw	F statistic	F Critical
Redwood City	2601.07	472349857.9	396121373.1	76228484.82	132040457.7	375509.7774	**351.63**	2.60
Menlo Park	2601.07	472349857.9	306300022.3	166049835.6	102100007.4	817979.4859	**124.82**	2.60
Mountain View	2601.07	472349857.9	381898229.7	90451628.18	127299409.9	445574.5231	**285.70**	2.60
San Mateo	2601.07	472349857.9	351075585.1	121274272.9	117025195	597410.2111	**195.89**	2.60
5% Trim	Grand Mean	SStotal	SSb	SSw	SSMb	SSMw	F statistic	F Critical
Redwood City	2557.86	271623269.1	220977578.7	50645690.37	73659192.91	276752.4064	**266.16**	2.60
Menlo Park	2557.86	271623269.1	180498083.6	91125185.47	60166027.88	497951.8332	**120.83**	2.60
Mountain View	2557.86	271623269.1	214252174.8	57371094.31	71417391.6	313503.2476	**227.80**	2.60
San Mateo	2557.86	271623269.1	191861153.2	79762115.93	63953717.72	435858.557	**146.73**	2.60
10% Trim	Grand Mean	SStotal	SSb	SSw	SSMb	SSMw	F statistic	F Critical
Redwood City	2593.00	153683506	131031274.2	22652231.77	43677091.41	140697.0917	**310.43**	2.60
Menlo Park	2593.00	153683506	97195585.04	56487920.96	32398528.35	350856.6519	**92.34**	2.60
Mountain View	2593.00	153683506	121599762.5	32083743.48	40533254.17	199277.9098	**203.40**	2.60
San Mateo	2593.00	153683506	104180932.9	49502573.1	34726977.63	307469.3981	**112.94**	2.60
25% Trim	Grand Mean	SStotal	SSb	SSw	SSMb	SSMw	F statistic	F Critical
Redwood City	2678.76	14601392.93	11018636.47	3582756.46	3672878.823	35472.83624	**103.54**	2.60
Menlo Park	2678.76	14601392.93	9099755.32	5501637.61	3033251.773	54471.6595	**55.68**	2.60
Mountain View	2678.76	14601392.93	11529437.56	3071955.37	3843145.853	30415.3997	**126.36**	2.60
San Mateo	2678.76	14601392.93	10336894.74	4264498.19	3445631.58	42222.75436	**81.61**	2.60

Fig 12. Table about F-Distribution Test.

From the summary table of *Figure 12*, calculating for the Grand Mean, SS_{total}, SS_b, SS_w, SS_{mb}, and SS_{mw}, for the F-test in relation of the One-Way ANOVA. Since the Grand Mean is the average of the "All" entity and SS_{total}. Noted that each calculation must be aware of their own Trim level. F Critical can then be calculated from a table using only the level significant and both bounded and within Degree of Freedoms.

$$F_{Critical} = 2.60$$

The F-statistics was calculated with SS_b, SS_w, and SS_{total} which had led to SS_{mb} and SS_{mw}. These two values are divided to each other:

$$F_{Statistics} = \frac{SS_{mb}}{SS_{mw}}$$

Results from F-Statistics were larger than the F-Critical – meant rejection of null hypothesis. *Figure 12*, highlighted in yellow.

$$F_{Statistical} > F_{Critical}: Rej\ H_0$$

This also meant that when comparing "All" to each of the city within its own respective trim levels, that meant from the sample does not share any joint effects. So, prices were seen as bounded values to other cities weren't true and only the local prices within their own city were impacted.

Z Test

Z-Test			
0% Trim	**Z Value**	**Z Score (5%)**	**Z score (10%)**
Redwood City	13.75322604	1.645	1.282
Menlo Park	8.285114823	1.645	1.282
Mountain View	14.80239927	1.645	1.282
San Mateo	-0.87412917	1.645	1.282
5% Trim	**Z Value**	**Z Score (5%)**	**Z score (10%)**
Redwood City	15.69663663	1.645	1.282
Menlo Park	9.613781729	1.645	1.282
Mountain View	7.340412206	1.645	1.282
San Mateo	-0.81829285	1.645	1.282
10% Trim	**Z Value**	**Z Score (5%)**	**Z score (10%)**
Redwood City	21.60275635	1.645	1.282
Menlo Park	10.82489498	1.645	1.282
Mountain View	20.24073123	1.645	1.282
San Mateo	-1.00878359	1.645	1.282
25% Trim	**Z Value**	**Z Score (5%)**	**Z score (10%)**
Redwood City	35.35020275	1.645	1.282
Menlo Park	23.706523	1.645	1.282
Mountain View	42.21689371	1.645	1.282
San Mateo	-1.60846854	1.645	1.282

Fig 13. Z-Test

Z values was used to test for false positive and false negative rejections of the null hypothesis. Though Z test were known to work with normalized graph, or bell-curves. However, there was still relevancy in using it in a positive skewed graph. The criteria were that the sample sizes had to be larger than 30 in each city and both the variances and SEM were known. In this case, all were met and Z-test was used.

All resulted in greater than the Z score in all 5% and 10% significance level. Therefore, the null hypothesis was rejected and all observed samples were deemed significant in each city. This meant that each city's price was truly independent but influenced by population.

In relation to the F-distribution, the result from Z test reflected well with the F-test.

P-Value Test

P-Test		
P-Values	**α-level 5%**	**α-level 10%**
0.00001	0.05	0.10
0.00001	0.05	0.10
0.00001	0.05	0.10
0.191032	0.05	0.10
P-Values	**α-level 5%**	**α-level 10%**
0.00001	0.05	0.10
0.00001	0.05	0.10
0.00001	0.05	0.10
0.26621	0.05	0.10
P-Values	**α-level 5%**	**α-level 10%**
0.00001	0.05	0.10
0.00001	0.05	0.10
0.00001	0.05	0.10
0.156487	0.05	0.10
P-Values	**α-level 5%**	**α-level 10%**
0.00001	0.05	0.10
0.00001	0.05	0.10
0.00001	0.05	0.10
0.053918	0.05	0.10

Fig 14. P-Test. Every city showed insignificant values towards the alpha levels, thus confirmed with rejecting the null hypothesis. Except for San Mateo at 10% alpha level.

P-values had shown all values to be insignificant, except for one in red when compared to p-critical at 10% significance level. This could be ignored because sample size for San Mateo at 25% trim was less than 20. This meant it was giving false reading.

Because Redwood City, Menlo Park, and Mountain View had shown insignificant value close to zero in the p-value test; when compared to significance level at 5% and 10%, while in both Z and F-tests had shown over significant values from Z-score and F-Critical respectively. This meant there is a high level of acceptances that the null hypothesis was believed to be true to reject the null hypothesis confidently.

Comparing Population to Hypothesis

With these findings and verification by the Z-test and P-test, the null hypothesis can be rejected with a high level of confidence. And to further test, visited the US Census bureau [2] to check the cities' populations to see if the alternative hypothesis was true in *Figure 15*. In this case, it was true!

Interestingly, population of renters or homeowners for leasing properties were roughly 50% - 60% of the population.

Population of Locals San Mateo County (2019)			
Cities	Total Population	Rental Population	Rental Occupancy (%)
Redwood City	30829	15488	50.24%
Menlo Park	11906	5010	42.08%
Mountain View	33756	19700	58.36%
San Mateo	38549	17570	45.88%
**US Census Bureau*			

Fig 15. A summary about population census of Redwood City, Menlo Park, Mountain View, and San Mateo in 2019.

Conclusion

With trimming and statistical tests to verify and account for accuracy, all answers suggested that there was a strong relationship about local populations impacted rental prices. Therefore, this test rejected the null hypothesis – that *local rental prices are not impacted by local populations. If local population is large, rent prices remain relatively the same or no changes.*

And we accepted the alternative hypothesis – that *at least two or more local rental prices are influenced by the population. If local population is large, rent prices changes, either increase or decrease depends on local population.*

Having to know that local population influenced the rent price, potential renters may consider looking for a more populated cities to find the least expensive apartment to rent. Especially in San Mateo with the highest population of 38,549 and has more affordable selections of rent. That follows with Redwood City and then Mountain View. While Menlo Park, with the lowest population of 11,906 holds the highest rent cost.

With that said, the consideration to live in the peninsula through rent is not for someone who can't budget nor those who don't have the finance because San Mateo County is the most expensive county in the SF Bay Area. But if someone must reside in the Peninsula, please suggest high populated cities.

References and Notes

1. "Apartment List - More than 5 Million Apartments for Rent."

Apartment List, 2021, www.apartmentlist.com.
2. "Cities Populations: Redwood City, Menlo Park, Mountain View, San Mateo." *United States Census Bureau*, data.census.gov/cedsci/table?g=0400000US06_1600000US0660102&tid=ACSDP5Y2019.DP04. Accessed 24 Feb. 2021.
3. Devore, Jay L. "Chapter 10.1 Single-Factor ANOVA." *Probability and Statistics for Engineering and the Sciences*, Ninth Edition, Rev. 3, United States of America, Cengage Learning, 2015, pp. 410–20.
4. "Quick P Value from Z Score Calculator." Social Science Statistics, 2021, www.socscistatistics.com/pvalues/normaldistribution.aspx.
5. "F Statistic / F Value: Definition and How to Run an F-Test." *Statistics How To*, 1 Jan. 2021, www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/#:%7E:text=The%20F%20statistic%20just%20compares,is%20smaller%20than%20variance%202.