

ZILLOW PROJECT

- BAN 612-02
- Group 4:
 - Aman Solanki
 - Krishna Jaideep Patel
 - Farzad Emami
 - Deepak Dileepkumar

Libraries



Visualization

Pandas
Numpy
Seaborn
Matplotlib



Web Scraping

Requests
Datetime
Random
BeautifulSoup



Data Cleaning

KNNImputer
Regular Expression



Google Maps API



Regression Models

Ridge
LassoCV
AdaBoost Regressor
RandomForest Regressor
GridSearch CV
RegressionSummary

Data Collection



Data Collected in Batches.



Data Collected from each page is stores in a dictionary and then that dictionary is appended to a list.



Each batch has its own list of dictionaries.



The list of dictionaries is then converted into a Pandas DataFrame.

Data Collection - Code

```
city_list = ['San Jose', 'San Francisco', 'Oakland',
            'Fremont', 'Santa Rosa', 'Hayward',
            'Sunnyvale', 'Concord', 'Santa Clara',
            'Vallejo', 'Berkeley', 'Fairfield', 'Richmond',
            'Antioch', 'Daly City', 'San Mateo', 'Vacaville',
            'San Leandro', 'Livermore', 'Napa', 'Redwood City',
            'Mountain View', 'Alameda', 'San Ramon', 'Pleasanton',
            'Union City', 'Milpitas', 'Palo Alto', 'Walnut Creek',
            'South San Francisco', 'Pittsburg', 'Cupertino', 'Petaluma',
            'San Rafael', 'Novato', 'Brentwood', 'Gilroy', 'Dublin',
            'Danville', 'San Bruno', 'Rohnert Park', 'Campbell',
            'Morgan Hill', 'Pacifica', 'Martinez', 'Oakley', 'Pleasant Hill', 'Menlo Park']

batch1 = ['San Jose', 'San Francisco', 'Oakland',
          'Fremont', 'Santa Rosa', 'Hayward',
          'Sunnyvale', 'Concord', 'Santa Clara',
          'Vallejo']

batch2 = ['Berkeley', 'Fairfield', 'Richmond',
          'Antioch', 'Daly City', 'San Mateo', 'Vacaville',
          'San Leandro', 'Livermore', 'Napa']

batch3 = ['Redwood City', 'Mountain View', 'Alameda', 'San Ramon', 'Pleasanton',
          'Union City', 'Milpitas', 'Palo Alto', 'South San Francisco']

batch4 = ['Pittsburg', 'Cupertino', 'Petaluma', 'San Rafael', 'Novato', 'Walnut Creek']

batch5 = ['Brentwood', 'Gilroy', 'Dublin', 'Danville', 'Foster City']

batch6 = ['San Bruno', 'Rohnert Park', 'Campbell', 'Half Moon Bay', 'Morgan Hill']

batch7 = ['Pacifica', 'Martinez', 'Oakley', 'Pleasant Hill', 'Menlo Park']

test = ['Hayward']

common_url = 'https://www.zillow.com/homes/for_sale/'

url_list = []
for city in test:
    #url_list.append(common_url+city+"/ca") #for page 1
    url_list.append(common_url+city+"/ca/2_p/") #for page 2
```

```
req_headers = {
    'casca'
    'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8',
    'accept-encoding': 'gzip, deflate, br',
    'accept-language': 'en-US,en;q=0.8',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/61.0.3163.100'
}

s = requests.Session()
link=[] #list to store listing url
listingAgent=[] #list to store the listing company of the property
temp=[] #temporary list to append all the collected data

for url in url_list:
    r = s.get(url, headers=req_headers)
    soup = BeautifulSoup(r.content, 'html.parser')
    for i in soup.findAll(class_="list-card-link list-card-link-top-margin list-card-img"):
        link.append(i.get('href')) #appends urls to link list
    for i in soup:
        listingCompany = soup.find_all('div', attrs={'class': 'list-card-truncate'})
        for i in listingCompany:
            company = i.text
            temp.append(company)
            temp = temp[:len(link)] #appends listing companies for only the accounted links

listingAgent.extend(temp) #adds all the temp elements to listingAgent list

print(len(link), 'links generated.')

40 links generated.

chunks = [link[x:x+8] for x in range(0, len(link), 8)] #creates 5 lists with 8 links in each list

len(chunks)

5
```

Data Collection - Code

```
null_val = 'Not_Found'
start_time = datetime.datetime.now()

print(start_time)

zillowDataList = [] #list to store dictionary generated for each property data.

s = requests.Session()

for i in range(len(chunks)):
    print('\n')
    print('Sleeping Now')
    print('-----')
    print('\n')
    time.sleep(random.randint(26,30)) #sleeps after scraping 5 links
    for j in chunks[i]:
        zillowData = {}
        print('Sleeping Again....')
        time.sleep(random.randint(13,16)) #sleeps before scraping data from each link
        r = s.get(j, headers=req_headers)
        bs_html_data = BeautifulSoup(r.content, 'html.parser')

        print('Link: ', j)

        #price
        if bs_html_data.find('span', attrs={'class': 'ds-value'}) != None:
            price = re.sub("\D", "", bs_html_data.find('span', attrs={'class': 'ds-value'}).text)
            zillowData['Price'] = price #adds price to dictionary
        else:
            price = null_val
            zillowData['Price'] = price #adds null_val if price not found

        #address
        address_tags = bs_html_data.find_all('h1', attrs={'class': 'ds-address-container'})

        if address_tags != None:
            address_tags = bs_html_data.find_all('h1', attrs={'class': 'ds-address-container'})
            spans = address_tags[0].find_all('span')

            if spans != None:
                home_address = ' '.join([_.text for _ in spans])

                clean_home_address = re.sub(r'[^a-zA-Z0-9, ]', '', home_address)
                zillowData['Address'] = clean_home_address #full property address

                addr_line1 = clean_home_address.split(',')[0].strip() #address line 1

                home_city = clean_home_address.split(',')[2].strip() #city
                zillowData['City'] = home_city #adds city to dictionary

                zipCode = int(clean_home_address.split(',')[1].strip().split(' ')[-1].strip()) #zipcode
                zillowData['ZipCode'] = zipCode #adds zipcode to dictionary

                state = clean_home_address.split(',')[1].strip().split(' ')[0].strip() #state
                zillowData['State'] = state #adds state to dictionary

        #zestimate
        zestimate = bs_html_data.find("p", attrs = {"class": "Text-aiai24-0 sc-fzoxKX sc-oTpqt loFLRq"})
        value=re.sub("\D", "", zestimate.text) if zestimate != None else null_val
        zillowData['Zestimate'] = value #adds zestimate to dictionary
```

```
#Interior details
interior_details = bs_html_data.find_all('span', attrs={'class': 'Text-aiai24-0 czksOw'}) #get interior details

if interior_details != None:
    bedrooms = (interior_details[0].text.split(' ')[1] if interior_details[0] != None else null_val)
    zillowData['Bedrooms'] = bedrooms #adds bedrooms to dictionary

#adds the specified data to the dictionary, if data is not found, null_val is added instead
for i in interior_details:
    text = i.text
    if text.startswith('Total interior livable area:'):
        area = text.split()[4] if text.split()[4] != None else null_val
        zillowData['Living-Area'] = area
    if text.startswith('Bathrooms:'):
        tBathrooms = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Total-Bathrooms'] = (int(tBathrooms))
    if text.startswith('Full bathrooms:'):
        fBathrooms = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Full-Bathrooms'] = fBathrooms
    if text.startswith('1/2 bathrooms:'):
        hBathrooms = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Half-Bathrooms'] = hBathrooms
    if text.startswith('Fireplace:'):
        fireplace = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Fireplace'] = fireplace
    if text.startswith('Lot size:'):
        lotSize = text.split()[2] if text.split()[2] != None else null_val
        zillowData['Lot-Size'] = lotSize
    if text.startswith('Home type:'):
        homeType = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Home-Type'] = homeType
    if text.startswith('New construction:'):
        newConstruction = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['New-Construction'] = newConstruction
    if text.startswith('Year built:'):
        yearBuilt = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Year-Built'] = yearBuilt
    if text.startswith('Utilities for property:'):
        utilitiesProvider = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Utilities'] = utilitiesProvider
    if text.startswith('Sunscore: Great solar potentialSun Number='):
        sunScore = text.split(':')[2] if text.split(':')[2] != None else null_val
        zillowData['Sunscore'] = sunScore
    if text.startswith('HOA fee: $'):
        listingHOA = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['HOA'] = listingHOA
    if text.startswith('Tax assessed value:'):
        taxValue = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Tax-Value'] = taxValue
    if text.startswith('Annual tax amount:'):
        aTax = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Annual-Tax'] = aTax
    if text.startswith('Garage spaces:'):
        gSpace = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Garage-Spaces'] = gSpace
    if text.startswith('Stories:'):
        hStories = text.split(':')[1] if text.split(':')[1] != None else null_val
        zillowData['Stories'] = hStories

zillowDataList.append(zillowData) #appends the dictionary to the list

print('Total time:',datetime.datetime.now()-start_time, 'minutes')
```

Collected Data

Address	City	ZipCode	State	Bedrooms	Price	Listing-Company	HOA	Zestimate	Link	Living-Area	Total-Bathrooms	Full-Bathrooms	Half-Bathrooms	Garage-Spaces	Stories	Fireplace	Lot-Size	Home-Type	New-Construction	Year-Built	Utilities	Sunscore	Annual-Tax	Tax-Value				
2863 S Bascom Ave APT 808, Campbell, CA 95008	Campbell	95008	CA	2	599000	Compass		\$445/mo	603718 https://www.zillow.ca	841	2	2	2	Not_Found	Not_Found	Yes	0.02	Condo	No		1984	Public Utilities	95.26	\$7,168	\$532,888			
6035 Admiralty Pl, San Jose, CA 95123	San Jose	95123	CA	3	895000	Golden Gate Sotheby's International Realty		\$117/mo	919331 https://www.zillow.ca	1,328	3	2	1	Not_Found	Not_Found	Yes	0.06	Single Family	No		1994	Public Utilities	81.65	\$8,062	\$590,585			
2491 Corriea Way, Fremont, CA 94539	Fremont	94539	CA	3	1E+06	Charles Roderick Baldwin		\$40/mo	1323250 https://www.zillow.ca	1,306	3	3	Not_Found		2	Not_Found	Yes	0.09	Single Family	No		1972	Not_Found	87.01	\$9,731	\$785,097		
48432 Silva Pomar Ter, Fremont, CA 94539	Fremont	94539	CA	4	2E+06	Re/Max Accord		\$231/mo	Not_Found https://www.zillow.ca	3,010	4	3	1		2	Not_Found	Not_Found	0.11	Single Family	No		2020	Not_Found	Not_Found	Not_Found	Not_Found		
34122 Asti Ter, Fremont, CA 94555	Fremont	94555	CA	3	1E+06	Star Realty		\$235/mo	1048392 https://www.zillow.ca	1,701	4	3	1		2	Not_Found	Not_Found	0.04	Townhouse	No		2013	Not_Found	93.96	\$8,467	\$675,713		
34903 Machado Cmn, Fremont, CA 94555	Fremont	94555	CA	2	978000	Connolly Real Estate Team		\$237/mo	Not_Found https://www.zillow.ca	1,441	3	3	Not_Found	Not_Found	Not_Found	Not_Found	0.02	Townhouse	No		2020	Public Utilities	Not_Found	\$4,484	\$369,043			
3924 Match Point Ave, Santa Rosa, CA 95407	Santa Ros	95407	CA	3	559000	Century 21 Alliance		\$61/mo	564751 https://www.zillow.ca	1,617	3	2	1		2	2	Yes	0.11	Single Family	No		1998	Not_Found	93.5	\$6,103	\$546,210		
2260 Knolls Hills Cir, Santa Rosa, CA 95405	Santa Ros	95405	CA	3	535000	Corcoran Global Living		\$296/mo	535878 https://www.zillow.ca	1,633	2	2	Not_Found	Not_Found		1	Yes	0	Condo	No		1979	Not_Found	93.87	\$3,794	\$342,061		
790 Shiloh Cyn, Santa Rosa, CA 95403	Santa Ros	95403	CA	5	5E+06	Sterling California Properties		\$295/mo	11000 https://www.zillow.ca	6,874	5	4	1		5	2	Yes	10.51	Single Family	No		2010	Not_Found	Not_Found	\$47,328	\$3,956,480		
6453 Meadowridge Dr, Santa Rosa, CA 95409	Santa Ros	95409	CA	3	849000	W Real Estate		\$110/mo	620456 https://www.zillow.ca	1,800	2	2	Not_Found		2	1	Yes	0.1	Single Family	No		1978	Not_Found	Not_Found	\$7,020	\$626,391		
1712 Las Raposas Ct, Santa Rosa, CA 95409	Santa Ros	95409	CA	3	385000	Keller Williams Realty		\$387/mo	359797 https://www.zillow.ca	1,220	2	2	1	1	Not_Found		2	Not_Found	0.02	Condo	No		1973	Not_Found	86.79	\$4,307	\$343,000	
2408 Lakeview Dr, Santa Rosa, CA 95405	Santa Ros	95405	CA	3	480000	NextHome Wine Country Premier		\$264/mo	480004 https://www.zillow.ca	1,309	2	2	Not_Found		1	2	Yes	0.03	Condo	No		1980	Not_Found	Not_Found	\$5,371	\$485,866		
1344 Woodhaven Dr, Santa Rosa, CA 95407	Santa Ros	95407	CA	3	519000	Century 21 Alliance		\$43/mo	519792 https://www.zillow.ca	1,245	3	2	2	1		2	2	Yes	0.07	Single Family	No		2003	Not_Found	94.3	\$4,743	\$436,804	
2719 Spindrift Ct, Hayward, CA 94545	Hayward	94545	CA	5	1E+06	Intero Real Estate Services		\$42/mo	1440874 https://www.zillow.ca	3,651	3	3	Not_Found		2	Not_Found	Yes	0.14	Single Family	No		2003	Not_Found	91.54	\$18,232	\$1,407,600		
2514 Admiral Cir, Hayward, CA 94545	Hayward	94545	CA	3	965000	Everhome		\$218/mo	925266 https://www.zillow.ca	2,047	3	2	2	1		2	Not_Found	Not_Found	0.06	Single Family	No		2016	Not_Found	89.14	\$12,033	\$891,830	
1421 Poppy Ln, Hayward, CA 94545	Hayward	94545	CA	3	829000	Alta Realty Group Ca, Inc.		\$229/mo	856796 https://www.zillow.ca	1,625	3	2	2	1		2	2	Not_Found	0.05	Single Family	No		2017	Not_Found	86.7	\$10,072	\$792,540	
2028 Oak Creek Pl, Hayward, CA 94541	Hayward	94541	CA	2	549950	Fohl and Hernandez RE		\$423/mo	565948 https://www.zillow.ca	1,542	3	2	2	1		2	Not_Found	Yes	0.02	Townhouse	No		1972	Not_Found	93.14	\$4,554	\$318,264	
1140 Walpert St, Hayward, CA 94541	Hayward	94541	CA	3	679000	Golden Gate Sothebys International Realty		\$505/mo	683797 https://www.zillow.ca	1,856	3	2	2	1		2	Not_Found	Not_Found	0.04	Townhouse	No		1985	Not_Found	93.14	\$7,320	\$405,678	
25850 Kay Ave APT 128, Hayward, CA 94545	Hayward	94545	CA	2	475000	Intero Real Estate Services		\$455/mo	486183 https://www.zillow.ca	1,007	2	2	Not_Found		1	Not_Found	Yes	Not_Found	Condo	No		1989	Not_Found	94.74	\$4,658	\$350,490		
29628 Desert Oak Ct APT 33, Hayward, CA 94544	Hayward	94544	CA	1	319999	UNITED R.E. - LOS ANGELES		\$364/mo	324373 https://www.zillow.ca	531	1	1	Not_Found		1	2	Not_Found	0.01	Condo	No		1985	Not_Found	85.1	\$4,209	\$321,300		
969 Cheryl Ann Cir APT 36, Hayward, CA 94544	Hayward	94544	CA	3	458888	BHG Reliance Partners		\$475/mo	Not_Found https://www.zillow.ca	1,245	2	2	Not_Found		1	Not_Found	Yes	Not_Found	Condo	No		1979	Not_Found	Not_Found	Not_Found	Not_Found		
28574 Starboard Ln, Hayward, CA 94545	Hayward	94545	CA	4	890000	Flat Rate Realty		\$202/mo	915958 https://www.zillow.ca	1,835	3	3	Not_Found		2	Not_Found	Yes	0.06	Single Family	No		2007	Not_Found	90.74	\$11,607	\$808,554		
448 Costa Mesa Ter APT C, Sunnyvale, CA 94085	Sunnyvale	94085	CA	2	788800	Bay One Real Estate Investment Corporation		\$399/mo	783839 https://www.zillow.ca	880	1	1	Not_Found		2	Not_Found	Not_Found	0.02	Condo	No		1985	Public Utilities	93.23	\$5,195	\$423,078		
1078 Doheny Ter, Sunnyvale, CA 94085	Sunnyvale	94085	CA	4	1E+06	Compass		\$250/mo	1477091 https://www.zillow.ca	1,947	4	3	2	1		2	Not_Found	Not_Found	0.02	Townhouse	No		2012	Public Utilities	94.86	\$11,258	\$925,877	
537 E Mc Kinley Ave APT A, Sunnyvale, CA 94086	Sunnyvale	94086	CA	3	Not_Fo	Sereno Group		\$430/mo	Not_Found https://www.zillow.ca	1,764	3	2	1		2	Not_Found	Yes	0.02	Townhouse	No		1981	Public Utilities	94.83	\$9,666	\$803,714		
903 Sunrose Ter APT 111, Sunnyvale, CA 94086	Sunnyvale	94086	CA	2	799500	Realty World-Residential Specialists		\$506/mo	834796 https://www.zillow.ca	1,086	2	2	Not_Found	Not_Found		Not_Found	Yes	0.03	Condo	No		1994	Individual Elec	94.83	\$3,567	\$289,470		
202 Peppermint Tree Ter UNIT 2, Sunnyvale, CA 94	Sunnyvale	94086	CA	2	1E+06	Compass		\$240/mo	1121270 https://www.zillow.ca	1,251	3	2	2	1		2	Not_Found	Not_Found	0.02	Townhouse	No		2010	Public Utilities	92.43	\$10,793	\$888,105	
1120 Karby Ter UNIT 102, Sunnyvale, CA 94089	Sunnyvale	94089	CA	2	949000	Compass		\$482/mo	1007084 https://www.zillow.ca	1,007	2	2	Not_Found		1	1	Not_Found	0.03	Townhouse	No		2014	Public Utilities	94.78	\$10,559	\$879,740		
574 Leyte Ter, Sunnyvale, CA 94089	Sunnyvale	94089	CA	3	1E+06	Compass		\$250/mo	1382721 https://www.zillow.ca	1,600	3	2	2	1		2	Not_Found	Not_Found	0.02	Townhouse	No		2004	Public Utilities	90.78	\$10,637	\$886,432	
1103 Munich Ter, Sunnyvale, CA 94089	Sunnyvale	94089	CA	3	1E+06	Sereno Group		\$253/mo	Not_Found https://www.zillow.ca	1,663	3	2	2	1	1	Not_Found		Not_Found	Not_Found	0.02	Townhouse	No		2004	Public Utilities	Not_Found	Not_Found	Not_Found
831 W California Ave UNIT R, Sunnyvale, CA 94086	Sunnyvale	94086	CA	2	699800	Compass		\$394/mo	706679 https://www.zillow.ca	760	2	2	1	1		1	Not_Found	Not_Found	0.05	Condo	No		1983	Public Utilities	86.83	\$7,854	\$649,458	
989 Asilomar Ter APT 6, Sunnyvale, CA 94086	Sunnyvale	94086	CA	3	1E+06	Compass		\$45/mo	1524171 https://www.zillow.ca	1,895	2	2	Not_Found		2	3	Yes	0.03	Townhouse	No		1991	Public Utilities	94.83	\$8,780	\$728,357		
1221 Pine Creek Way APT A, Concord, CA 94520	Concord	94520	CA	3	399000	Keller Williams Realty		\$315/mo	Not_Found https://www.zillow.ca	1,240	2	2	1	1		2	Not_Found	Not_Found	0.02	Townhouse	No		1971	Not_Found	94.81	Not_Found	Not_Found	
1212 Oak Knoll Dr, Concord, CA 94521	Concord	94521	CA	5	1E+06	RE/MAX Accord		\$189/mo	1065007 https://www.zillow.ca	2,923	3	3	Not_Found		3	2	Yes	0.18	Single Family	No		2003	Not_Found	80.29	\$10,373	\$846,161		
3525 Northwood Dr UNIT C, Concord, CA 94520	Concord	94520	CA	2	370000	RE/MAX Accord		\$395/mo	381127 https://www.zillow.ca	1,090	2	2	1	1		2	Not_Found	Yes	0.02	Townhouse	No		1971	Not_Found	Not_Found	Not_Found	Not_Found	
1095 Mohr Ln APT A, Concord, CA 94518	Concord	94518	CA	3	399950	Coldwell Banker		\$413/mo	Not_Found https://www.zillow.ca	1,039	2	2	Not_Found	Not_Found		Not_Found	Not_Found	Not_Found	Condo	No		1978	Not_Found	89.65	Not_Found	Not_Found		
952 Autumn Oak Cir, Concord, CA 94521	Concord	94521	CA	5	1E+06	BAY METROPOLITAN		\$85/mo	1154024 https://www.zillow.ca	3,168	4	3	3	1		3	Not_Found	Not_Found	0.23	Single Family	No		2000	Not_Found	72.29	\$11,317	\$936,249	
4336 Saint Charles Pl, Concord, CA 94521	Concord	94521	CA	3	499900	Keller Williams Realty Danville		\$125/mo	506409 https://www.zillow.ca	1,182	2	2	Not_Found		1	1	Not_Found	0.09	Townhouse	No		1980	Not_Found	80.25	\$4,710	\$354,899		
4888 Clayton Rd APT 4, Concord, CA 94521	Concord	94521	CA	1	240000	RICK FULLER INC.		\$420/mo	250439 https://www.zillow.ca	741	1	1	Not_Found		1	Not_Found	Not_Found	Not_Found	Condo	No		1969	Not_Found	94.69	Not_Found	Not_Found		
5455 Kirkwood Dr APT B8, Concord, CA 94521	Concord	94521	CA	1	275000	RICK FULLER INC.		\$375/mo	285330 https://www.zillow.ca	696	1	1	Not_Found	Not_Found		Not_Found	Not_Found	Not_Found	Condo	No		1980	Not_Found	74.69	\$3,711	\$259,038		
825 Oak Grove Rd APT 98, Concord, CA 94518	Concord	94518	CA	3	499000	Golden Gate Sotheby's Int'l Re		\$465/mo	521103 https://www.zillow.ca	1,447	2	2	Not_Found	Not_Found		Not_Found	Yes	0.03	Condo	No		1970	Not_Found	85.65	\$2,572	\$158,487		
60 Palm Ln, Concord, CA 94518	Concord	94518	CA	2	149500	Keller Williams Realty		\$692/mo	156882 https://www.zillow.ca	755	2	1	2	1		3	Not_Found	Not_Found	0.01	Mobile / Mani	No		1968	Not_Found	Not_Found	Not_Found	Not_Found	
3905 Clayton Rd APT 16, Concord, CA 94521	Concord	94521	CA	2	312000	Intero Real Estate Services		\$396/mo	314311 https://www.zillow.ca	1,050	2	2	1	1	Not_Found	Not_Found	Not_Found	Not_Found	Condo	No		1973	Not_Found	93.85	Not_Found	Not_Found		
3309 Benton St, Santa Clara, CA 95051	Santa Clar	95051	CA	3	1E+06	Tuscana Properties		\$520/mo	1171726 https://www.zillow.ca	1,325	2	2	Not_Found		1	1	Yes	0.06	Condo	No		1963	Public Utilities	90.06	\$3,376	\$278,044		
2980 Salem Dr, Santa Clara, CA 95051	Santa Clar	95051	CA	3	1E+06	Compass		\$320/mo	1157027 https://www.zillow.ca	1,630	3	2	2	1		2	Not_Found	Yes	0.06	Townhouse	No		1974	Public Utilities	94.86	\$7,841	\$647,998	
3033 Kaiser Dr UNIT H, Santa Clara, CA 95051	Santa Clar	95051	CA	1	495000	Coldwell Banker Realty		\$495/mo	503389 https://www.zillow.ca	680	1	1	Not_Found			1	Not_Found	Yes	0.02	Condo	No		1971	Master Meter	89.26	\$2,127	\$174,608	
2380 Homestead Rd UNIT 3103, Santa Clara, CA 95	Santa Clara	95050	CA	1	499000	Sela Homes		\$382/mo	499004 https://www.zillow.ca	624	1	1	Not_Found		1	Not_Found	Not_Found	0.02	Condo	No		2007	Public Utilities	92.46	\$3,279	\$270,000		
4942 Calle De Escuela, Santa Clara, CA 95054	Santa Clara	95054	CA	2	699800	8 Blocks Real Estate		\$377/mo	710957 https://www.zillow.ca	1,108	2	2	1	1	Not_Found	Not_Found	Yes	0.02	Townhouse	No		1978	Public Utilities	79.56	\$5,339	\$433,019		
1903 Miraplaza Ct APT 16, Santa Clara, CA 95051	Santa Clara	95051	CA	2	799950	Re/Max Accord		\$430/mo	810963 https://www.zillow.ca	1,041	2	2	Not_Found		2	Not_Found	Yes	0.02	Condo	No		1986						

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 3722 entries, 0 to 3842  
Data columns (total 26 columns):
```

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	3722 non-null	int64
1	Address	3722 non-null	object
2	City	3722 non-null	object
3	ZipCode	3722 non-null	int64
4	State	3722 non-null	object
5	Bedrooms	3722 non-null	object
6	Price	3701 non-null	object
7	Listing-Company	3722 non-null	object
8	HOA	1319 non-null	object
9	Zestimate	3202 non-null	object
10	Link	3722 non-null	object
11	Living-Area	3535 non-null	object
12	Total-Bathrooms	3491 non-null	object
13	Full-Bathrooms	3193 non-null	object
14	Half-Bathrooms	1160 non-null	object
15	Garage-Spaces	2938 non-null	object
16	Stories	1641 non-null	object
17	Fireplace	1794 non-null	object
18	Lot-Size	2900 non-null	object
19	Home-Type	3555 non-null	object
20	New-Construction	3458 non-null	object
21	Year-Built	3346 non-null	object
22	Utilities	1309 non-null	object
23	Sunscore	2486 non-null	object
24	Annual-Tax	2954 non-null	object
25	Tax-Value	2983 non-null	object

```
dtypes: int64(2), object(24)
```

```
memory usage: 785.1+ KB
```

Collected Data Summary



Data Cleaning



Deleted Records for listings where State != California. For example, Data for Dublin was scraped for Dublin, OH instead of Dublin, CA



Replaced 'Not_found' with null value (np.nan)



Dropped unwanted columns and columns with more than 60% missing data.



Imputed Missing values for Sunscore and Tax value using KNNImputer.



Cleaned Listing Company Column. Chose to manually go through the distinct company names and rename common companies. For example, Compass SF could be considered as Compass.



In the end, dropped a record with any missing value. [.dropna()]

Data Exploration (Cleaned Data)

```
<class 'pandas.core.frame.DataFrame'>
```

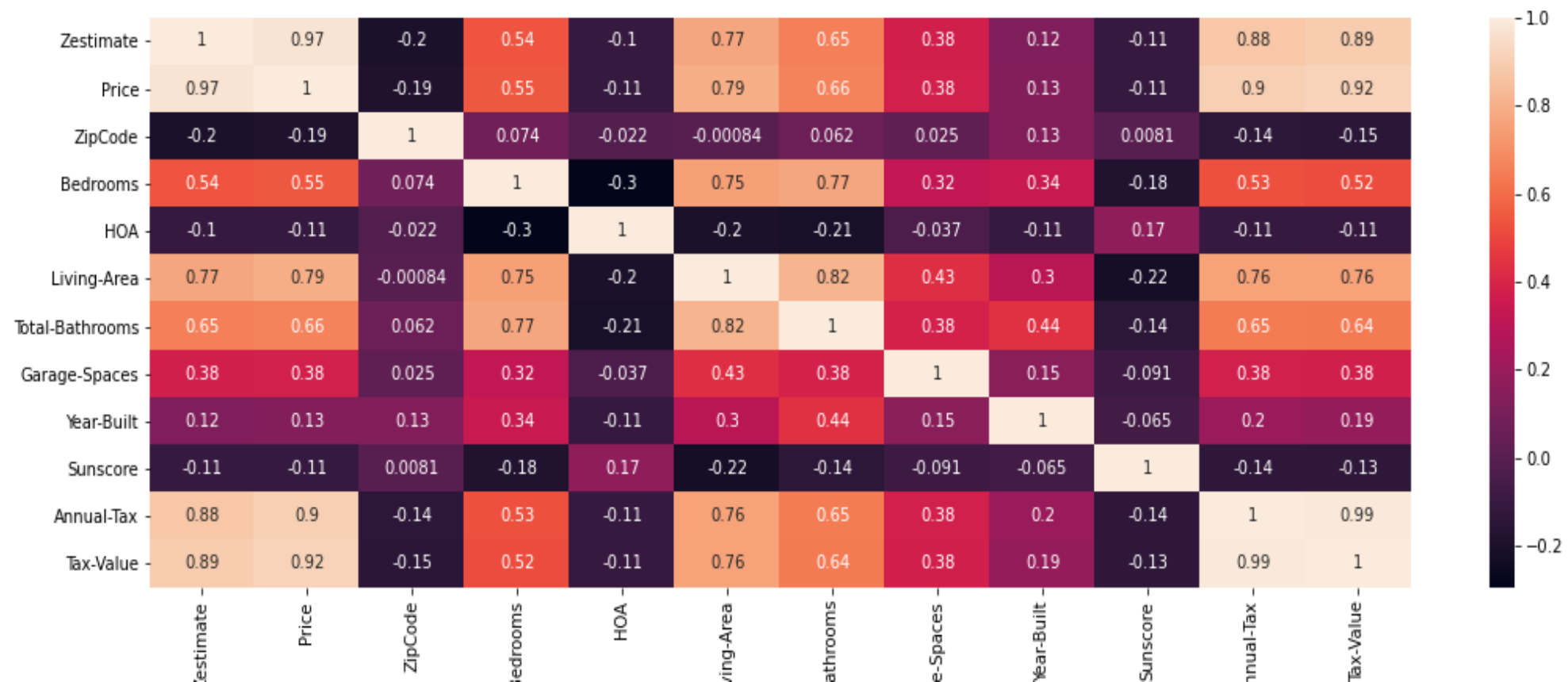
```
RangeIndex: 1072 entries, 0 to 1071
```

```
Data columns (total 19 columns):
```

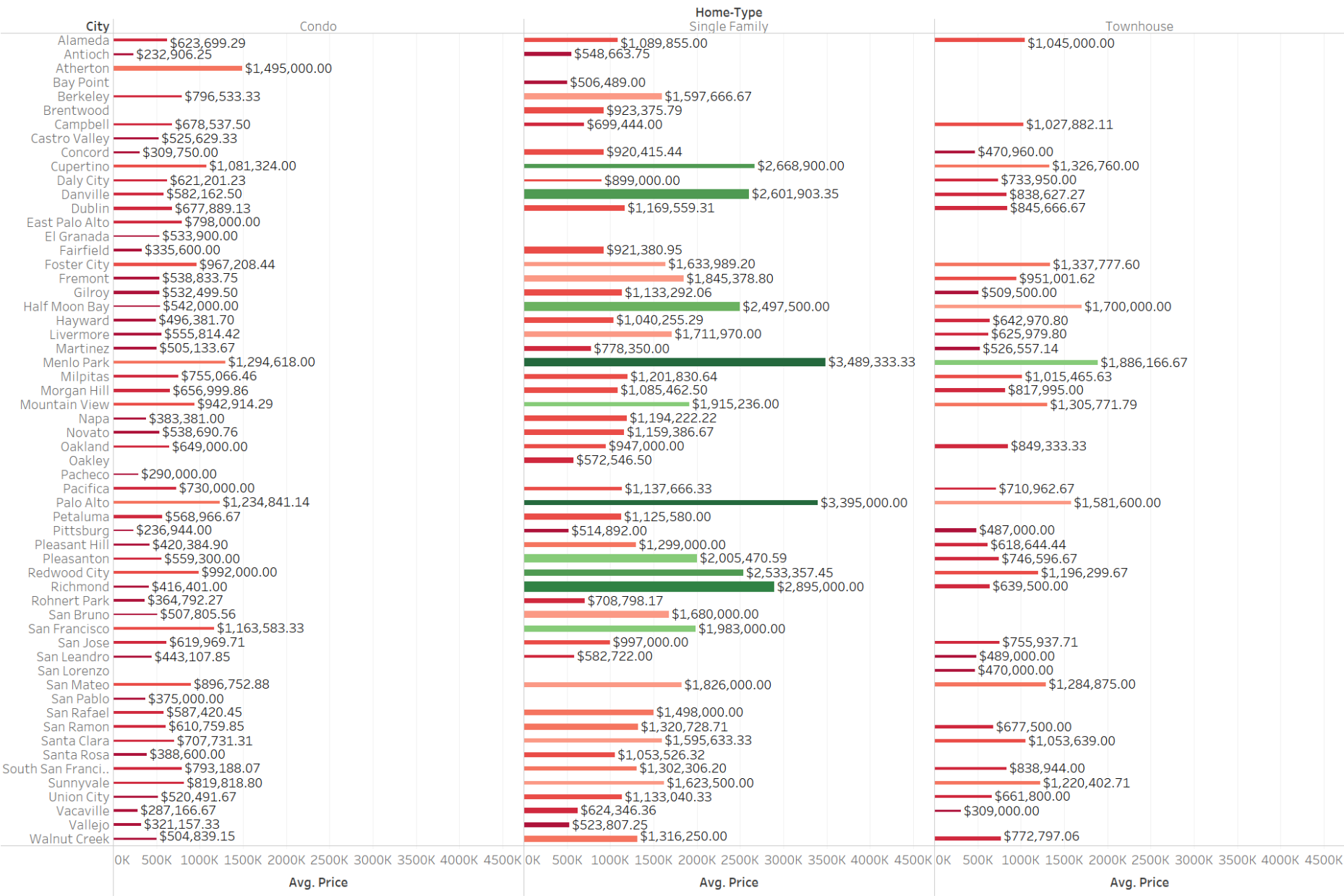
#	Column	Non-Null Count	Dtype
0	Address	1072 non-null	object
1	State	1072 non-null	object
2	Zestimate	1072 non-null	int64
3	Listing-Company	1072 non-null	object
4	Link	1072 non-null	object
5	Price	1072 non-null	int64
6	City	1072 non-null	object
7	ZipCode	1072 non-null	int64
8	Bedrooms	1072 non-null	int64
9	HOA	1072 non-null	float64
10	Living-Area	1072 non-null	float64
11	Total-Bathrooms	1072 non-null	int64
12	Garage-Spaces	1072 non-null	int64
13	Home-Type	1072 non-null	object
14	Year-Built	1072 non-null	int64
15	Utilities	1072 non-null	object
16	Sunscore	1072 non-null	float64
17	Annual-Tax	1072 non-null	float64
18	Tax-Value	1072 non-null	float64

```
dtypes: float64(5), int64(7), object(7)
```

```
memory usage: 159.2+ KB
```

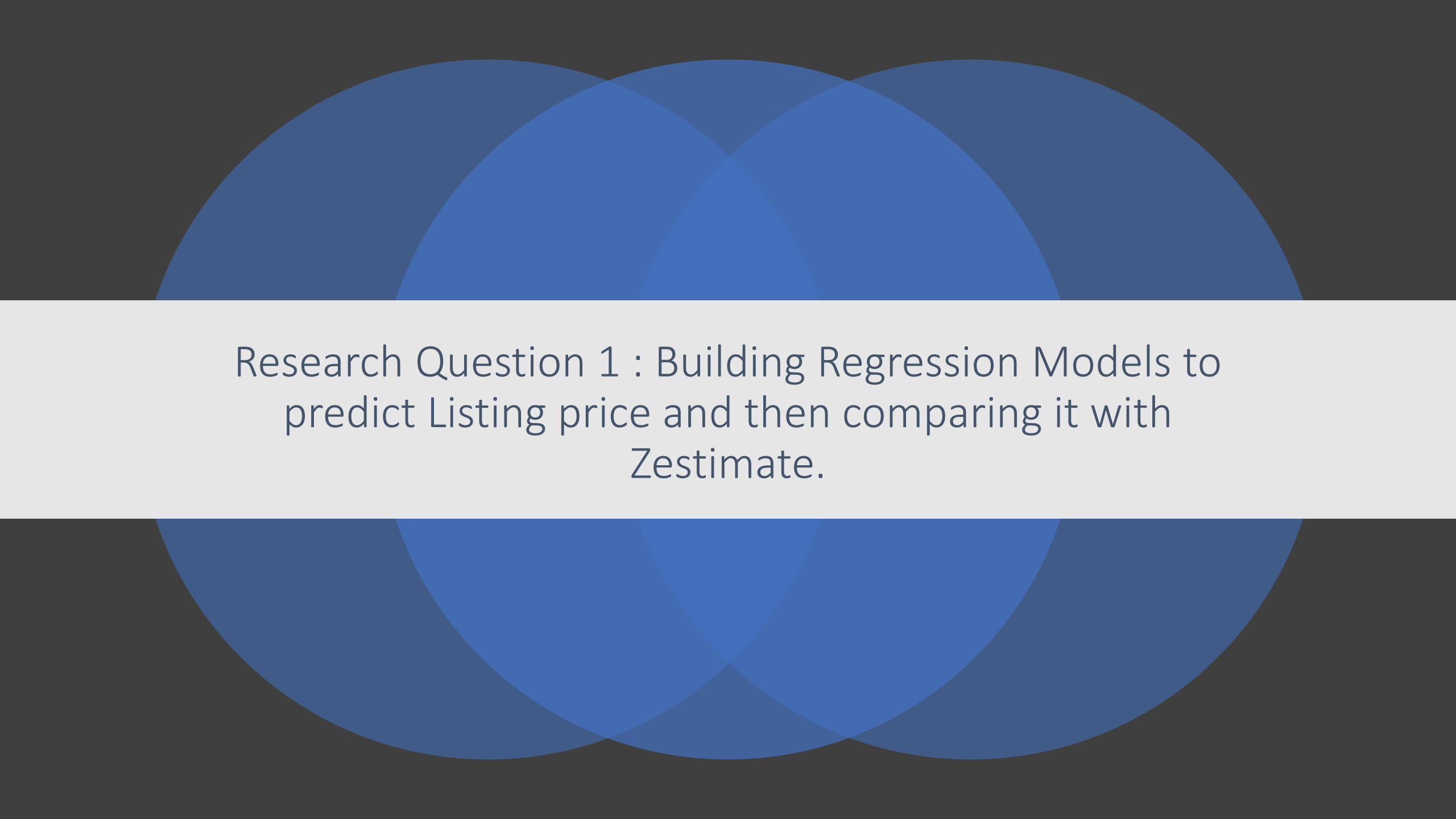
DataFrame Correlation Heatmap



Generated by
Tableau Desktop



Average of Price for each City broken down by Home-Type. Color shows average of Price. Size shows average of Living-Area. The marks are labeled by average of Price. The view is filtered on Home-Type, which keeps Condo,



Research Question 1 : Building Regression Models to predict Listing price and then comparing it with Zestimate.

Split Data for training, validation and testing

```
: scaler = preprocessing.MinMaxScaler()
scaler.fit(z4.iloc[:,6:])

df_scaled=pd.DataFrame(scaler.transform(z4.iloc[:,6:]), columns=z4.columns[6:])

X = df_scaled #predictor variables
y = z4['Price'] #outcome variable

#training (50%), validation (30%), and test (20%) partition
train_X, temp_X, train_y, temp_y = train_test_split(X, y, test_size=0.5, random_state=1)
valid_X, test_X, valid_y, test_y = train_test_split(temp_X, temp_y, test_size=0.4, random_state=1)

print('Training : ', train_X.shape)
print('Validation : ', valid_X.shape)
print('Test : ', test_X.shape)

Training : (536, 226)
Validation : (321, 226)
Test : (215, 226)
```

Preparing Data

Regression Models



Least Angle Regression



Random Forest: Used AdaBoost Regressor and GridSeachCV for tuning hyperparameters.



Ridge Regression

```

modelEvaluation['Price'] = test_y
modelEvaluation['Least Angle Regression'] = pred_leastAngle.astype(int)
modelEvaluation['Random Forest'] = pred_rfc.astype(int)
modelEvaluation['Ridge Regression'] = pred_ridge.astype(int)

```

```

zestimateModelComparison=[]

```

```

for i in modelEvaluation.index:
    zestimateModelComparison.append(z4['Zestimate'][i]) #get zestimate for the records in the test set

```

```

modelEvaluation['Zestimate'] = zestimateModelComparison #add Zestimate to the dataframe

```

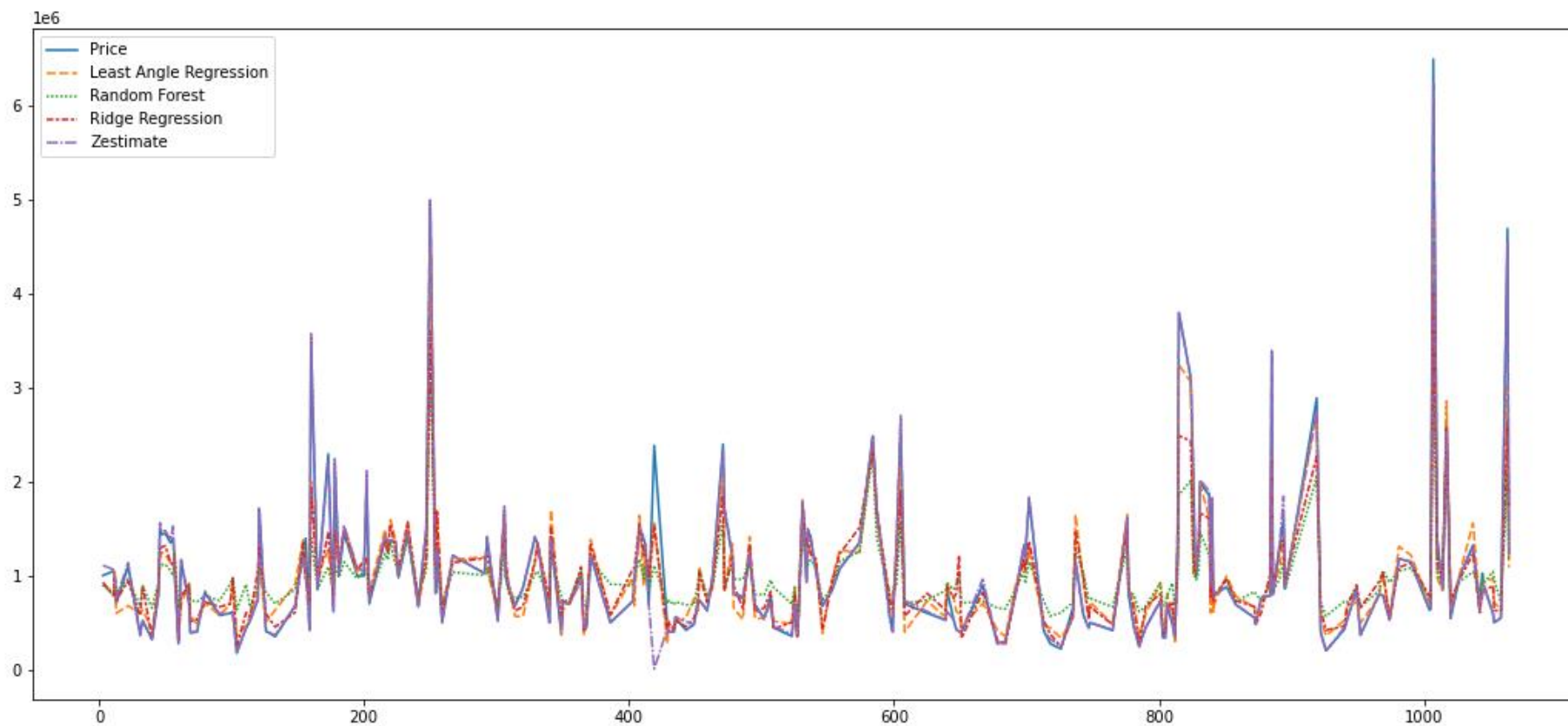
```

modelEvaluation.head()

```

	Price	Least Angle Regression	Random Forest	Ridge Regression	Zestimate
1017	2650000	2856645	2404015	2587368	2539235
411	1388000	889778	966880	1212489	1419589
1053	500000	715985	955293	627382	500375
12	949000	1050367	983655	962254	973567

Model Comparison

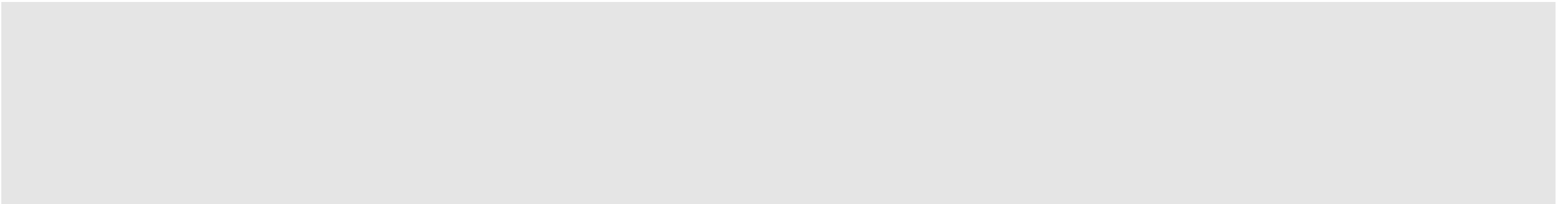


Model Comparison Graph

Accuracy

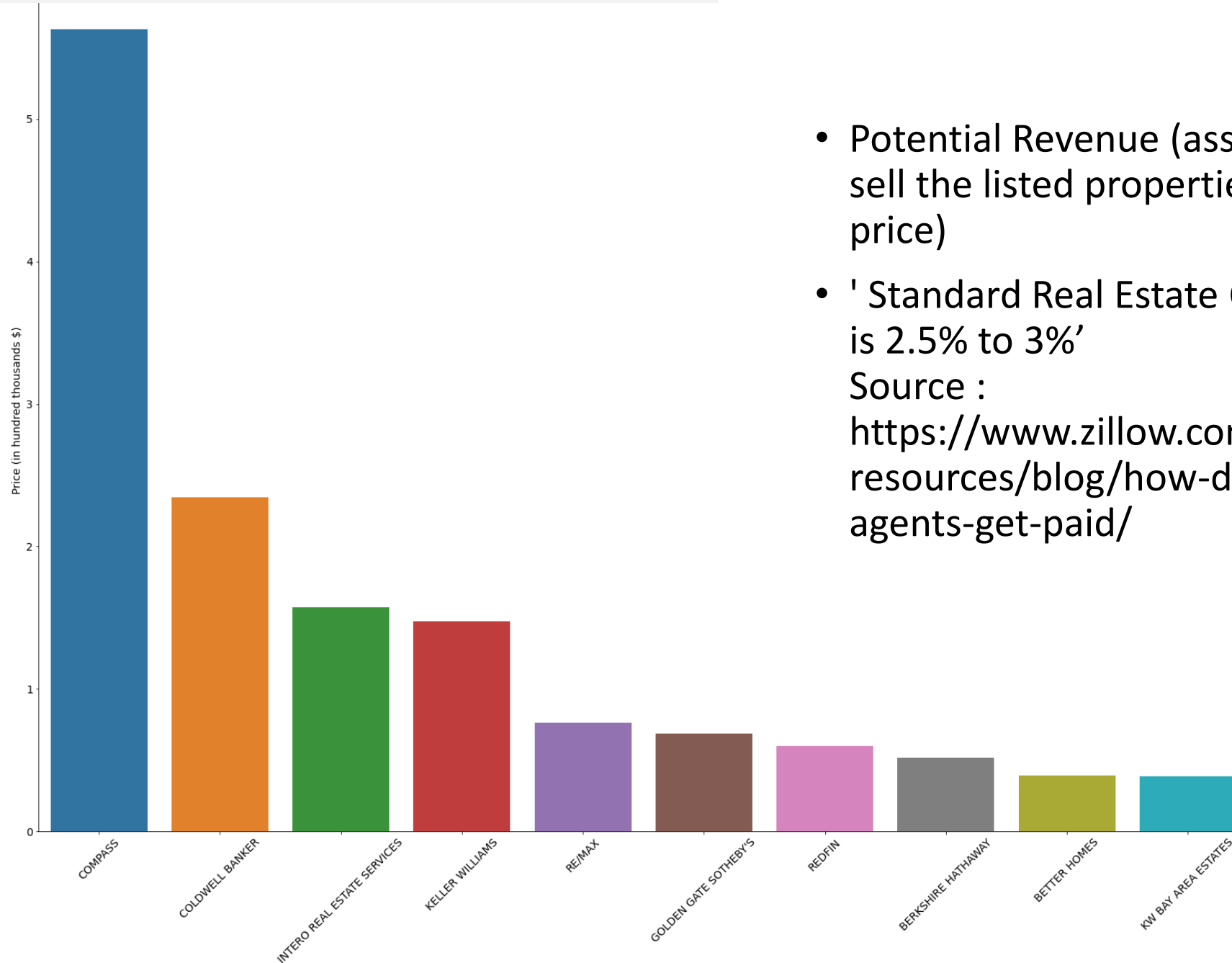
Model	Listing Price	Zestimate
Ridge	78.01 %	79.20 %
Least Angle Regression	83.16 %	84.77 %
Random Forest	60.84 %	61.05 %

Research Question 2 : How much money can a real estate company can potentially make in the Bay Area?



Listing-Company	Price	Properties
COMPASS	225307469	189
COLDWELL BANKER	93802173	104
INTERO REAL ESTATE SERVICES	62803490	58
KELLER WILLIAMS	59036205	59
RE/MAX	30388218	45
GOLDEN GATE SOTHEBY'S	27492900	23
REDFIN	23843552	27
BERKSHIRE HATHAWAY	20667521	22
BETTER HOMES	15560788	17
KW BAY AREA ESTATES	15420564	19

Top 10 Real Estate Companies In the Bay Area



- Potential Revenue (assuming they sell the listed properties at asking price)
- ' Standard Real Estate Commission is 2.5% to 3%'
Source :
<https://www.zillow.com/agent-resources/blog/how-do-real-estate-agents-get-paid/>

Research Question 3 : Is real estate market near San Francisco International Airport more expensive?

Data Collection



Google Maps API was used to obtain the distance and duration from the selected property address to San Francisco International Airport. (mode=driving)



Exported the new DataFrame to a CSV file.



Sorted the collected Data by Distance to the airport.



Generated a Correlation Heatmap to conclude the results.

```
gmaps = googlemaps.Client(key='myAPIkey') #removed intentionally

durationList=[]
distanceList=[]

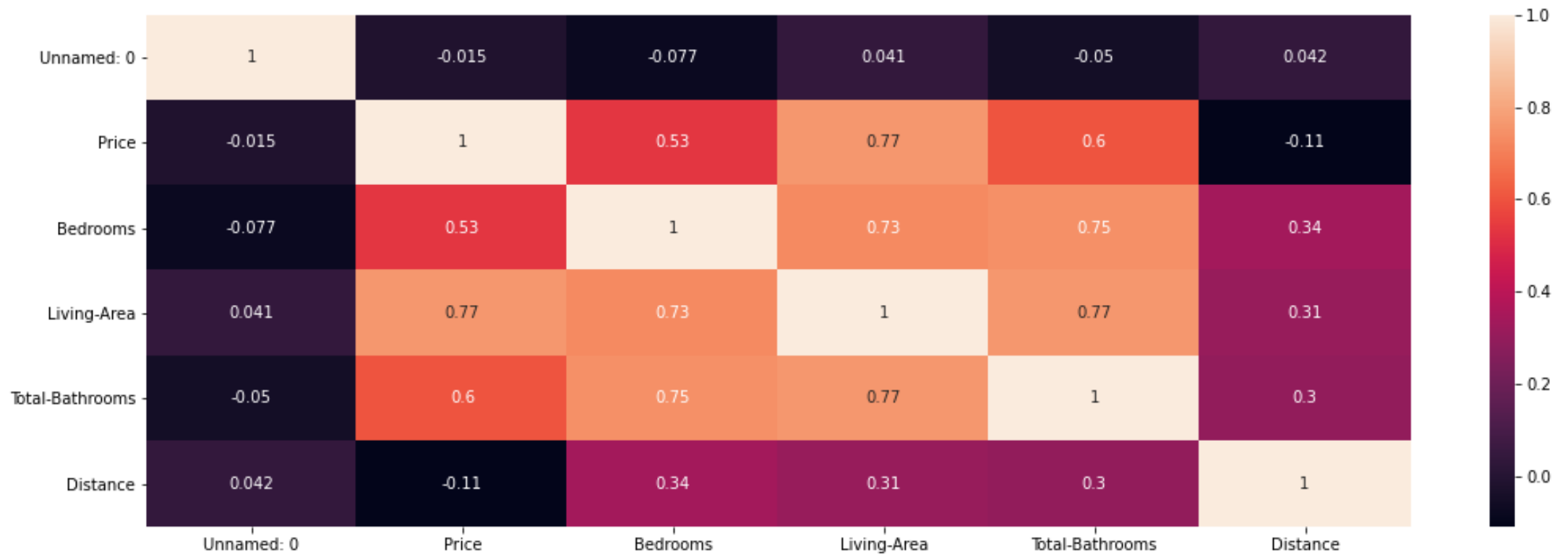
for i in range(len(z6['Address'])):
    directions={}
    now = datetime.now()
    directions = gmaps.directions(z6['Address'][i], "San Francisco International Airport",
                                   mode="driving",
                                   departure_time=now)
    distanceList.append(str(directions[0]['legs'][0]['distance']['text']))
    durationList.append(str(directions[0]['legs'][0]['duration']['text']))

z6['Distance'] = distanceList[:450]
z6['Duration'] = durationList[:450]
```

Code – Gather Duration and
Distance Data |

Address	Price	City	Bedrooms	Living-Area	Total-Bathrooms	Home-Type	Distance	Duration
2863 S Bascom Ave APT 808, Campbell, CA 95008	599000	Campbell	2	841.0	2	Condo	38.5	39 mins
6035 Admiralty Pl, San Jose, CA 95123	895000	San Jose	3	1328.0	3	Single Family	42.7	42 mins
761 Bonita Pl, San Jose, CA 95116	642888	San Jose	3	1231.0	3	Townhouse	36.5	37 mins
100 Ballatore Ct, San Jose, CA 95134	999900	San Jose	3	1405.0	3	Townhouse	30.4	32 mins
2664 Senter Rd APT 221, San Jose, CA 95111	550000	San Jose	2	1026.0	2	Condo	40.1	41 mins
3106 Capewood Ln, San Jose, CA 95132	1299000	San Jose	5	2110.0	4	Single Family	35.9	38 mins
3633 Jasmine Cir, San Jose, CA 95135	699000	San Jose	2	1037.0	2	Condo	44.4	46 mins
242 Jersey St, San Francisco, CA 94114	1249000	San Francisco	3	1232.0	1	Condo	11.1	16 mins
168 Dorantes Ave, San Francisco, CA 94116	2575000	San Francisco	4	3216.0	4	Single Family	11.7	17 mins
2257 Fulton St, San Francisco, CA 94117	1199000	San Francisco	2	1350.0	2	Condo	14.0	20 mins

Collected Data



Distance and Price have a negative correlation

Correlation Heatmap



THANK YOU