rental_analysis

October 8, 2021

1 San Francisco Housing Cost Analysis

In this assignment, you will perform fundamental analysis for the San Francisco housing market to allow potential real estate investors to choose rental investment properties.

[ ]:
```
# imports
import panel as pn
pn.extension('plotly')
import plotly.express as px
import pandas as pd
import hvplot.pandas
import matplotlib.pyplot as plt
import numpy as np
import os
from pathlib import Path
from dotenv import load_dotenv

from panel.interact import interact
from panel import widgets

import hvplot.pandas
import warnings
warnings.filterwarnings('ignore')
```

[ ]:
```
# Read the Mapbox API key
load_dotenv()
map_box_api = os.getenv("mapbox")
px.set_mapbox_access_token(map_box_api)
```

1.1 Load Data

[ ]:
```
# Read the census data into a Pandas DataFrame
file_path = Path("Data/sfo_neighborhoods_census_data.csv")
sfo_data = pd.read_csv(file_path, index_col="year")
sfo_data.head()
```

[ ]:
```
          neighborhood  sale_price_sqr_foot  housing_units  gross_rent
year
```

2010	Alamo Square	291.182945	372560	1239
2010	Anza Vista	267.932583	372560	1239
2010	Bayview	170.098665	372560	1239
2010	Buena Vista Park	347.394919	372560	1239
2010	Central Richmond	319.027623	372560	1239

1.2 Housing Units Per Year

In this section, you will calculate the number of housing units per year and visualize the results as a bar chart using the Pandas plot function.

Hint: Use the Pandas `groupby` function.

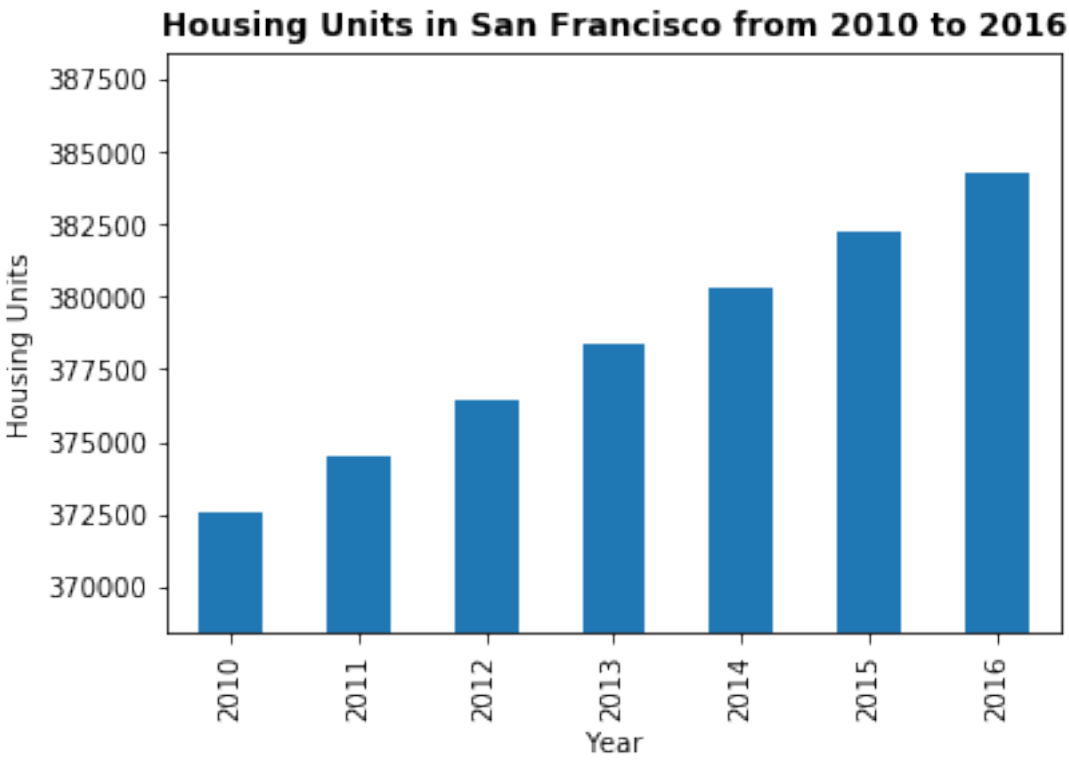
Optional challenge: Use the min, max, and std to scale the y limits of the chart.

[ ]:
```
# Calculate the mean number of housing units per year (hint: use groupby)
# YOUR CODE HERE!mean
mean_sfa=sfo_data['housing_units'].groupby(['year']).mean()
```

[ ]:
```
# Save the dataframe as a csv file
# YOUR CODE HERE!
mean_sfa.to_csv('mean housing per year.csv')
```

[ ]:
```
# Use the Pandas plot function to plot the average housing units per year.
# Note: You will need to manually adjust the y limit of the chart using the min␣
 ↪and max values from above.
# YOUR CODE HERE!
empty_canvas = plt.figure()
bar_hist = mean_sfa.plot.bar()
ymin=min(mean_sfa)-mean_sfa.std()
ymax=max(mean_sfa)+mean_sfa.std()

bar_hist.set_ylim(ymin=ymin,ymax=ymax)
bar_hist.set_xlabel('Year')
bar_hist.set_ylabel('Housing Units')
bar_hist.set_title("Housing Units in San Francisco from 2010 to 2016",␣
 ↪fontweight="bold")
```

[ ]:
```
Text(0.5, 1.0, 'Housing Units in San Francisco from 2010 to 2016')
```



1.3 Average Housing Costs in San Francisco Per Year

In this section, you will calculate the average monthly rent and the average price per square foot for each year. An investor may wish to better understand the sales price of the rental property over time. For example, a customer will want to know if they should expect an increase or decrease in the property value over time so they can determine how long to hold the rental property. Plot the results as two line charts.

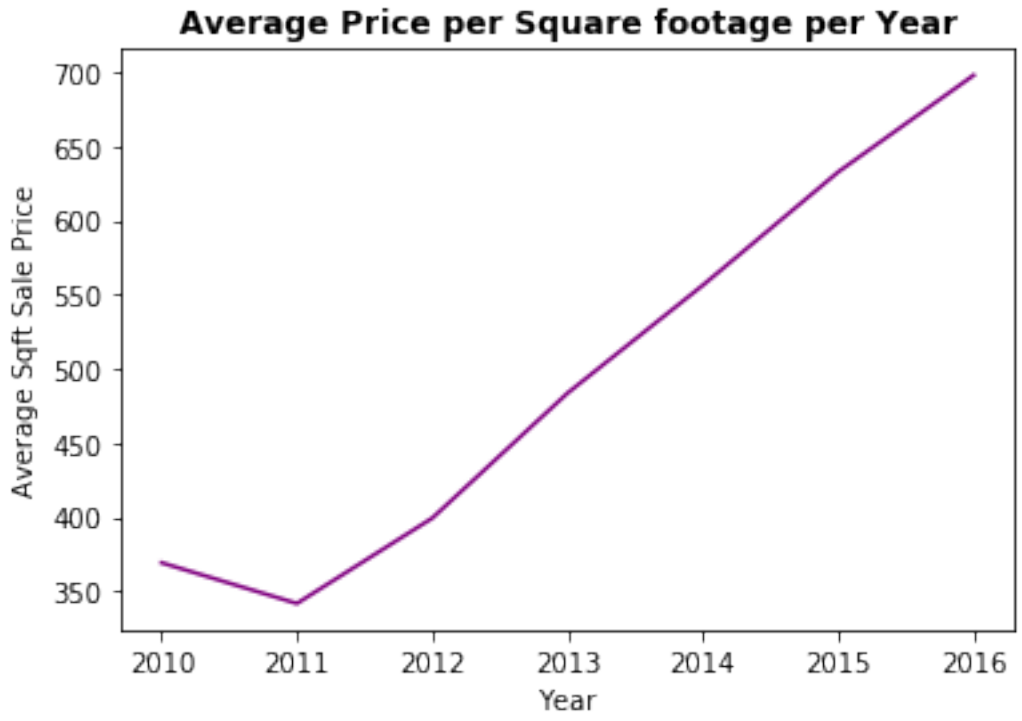
Optional challenge: Plot each line chart in a different color.

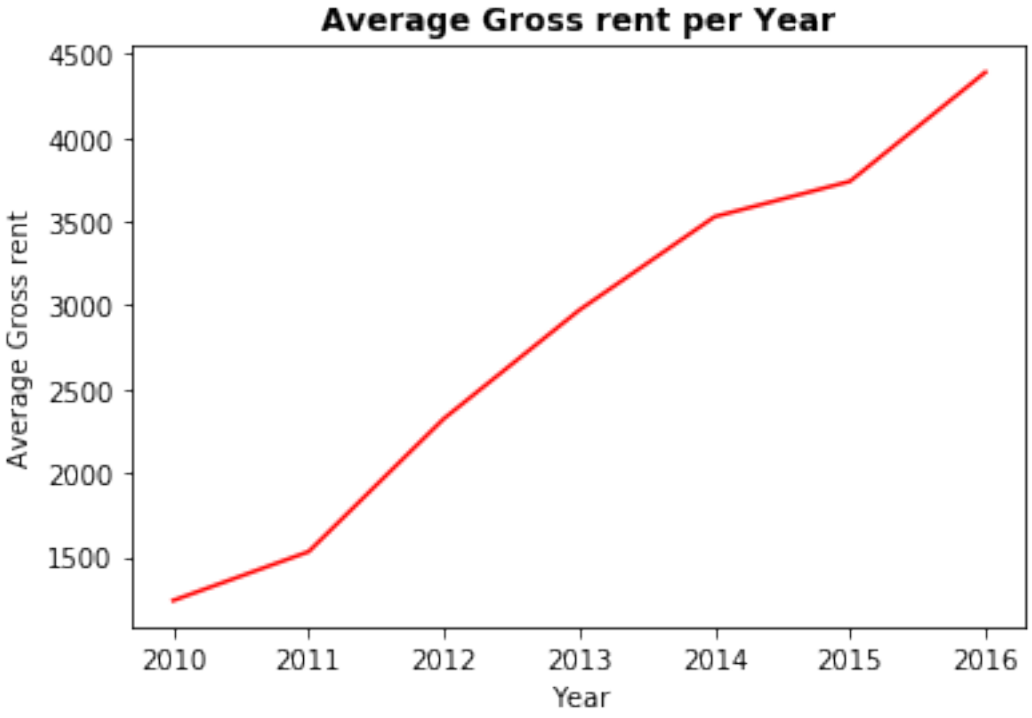
[ ]:
```
# Calculate the average sale price per square foot and average gross rent
# YOUR CODE HERE!
mu_sqrfootPrice_gross_rent=sfo_data[['sale_price_sqr_foot','gross_rent']].
 ↪groupby('year').mean()
```

[ ]:
```
def plot_line_chart(data,x_label,y_label,title,color):
    empty_canvas=plt.figure()
    data_plot =data.plot.line(color=color)
    data_plot.set_xlabel(x_label)
    data_plot.set_ylabel(y_label)
    data_plot.set_title(title, fontweight="bold")
```

```
    return plt.show()

def plot_hvplot_line(df,x_var,y_var,x_label,y_label,title,groupby):
   return df.hvplot.line(x=x_var,y=y_var,xlabel =x_label,ylabel␣
↪=y_label,title=title,groupby=groupby)

def␣
↪plot_hvplot_bar(df,x_var,y_var,x_label,y_label,title,rot=90,height=500,groupby=None):
↪
      if groupby==None:
         empty_canvas =df.hvplot.bar(x=x_var,y=y_var,xlabel =x_label,ylabel␣
↪=y_label,title=title,rot=rot,height=height)
      else:
         empty_canvas=df.hvplot.bar(x=x_var,y=y_var,xlabel =x_label,ylabel␣
↪=y_label,title=title,rot=rot,height=height,groupby=groupby)
      return empty_canvas
```

[ ]:
```
# Create two line charts, one to plot the average sale price per square foot␣
↪and another for average montly rent

# Line chart for average sale price per square foot
# YOUR CODE HERE!

plot_line_chart(␣
↪mu_sqrfootPrice_gross_rent['sale_price_sqr_foot'],'Year','Average Sqft Sale␣
↪Price',"Average Price per Square footage per Year",'purple')



# Line chart for average montly rent
# YOUR CODE HERE!
plot_line_chart( mu_sqrfootPrice_gross_rent['gross_rent'],'Year','Average Gross␣
↪rent',"Average Gross rent per Year",'red')
```


Average Price per Square footage per Year
Average Sqft Sale Price
700
650
600
550
500
450
400
350
2010
2011
2012
2013
2014
2015
2016
Year


Average Gross rent per Year
Average Gross rent
4500
4000
3500
3000
2500
2000
1500
2010
2011
2012
2013
2014
2015
2016
Year

1.4 Average Prices by Neighborhood

In this section, you will use hvplot to create two interactive visulizations of average prices with a dropdown selector for the neighborhood. The first visualization will be a line plot showing the trend of average price per square foot over time for each neighborhood. The second will be a line plot showing the trend of average montly rent over time for each neighborhood.

Hint: It will be easier to create a new DataFrame from grouping the data and calculating the mean prices for each year and neighborhood

[ ]:
```
# Group by year and neighborhood and then create a new dataframe of the mean␣
↪values
# YOUR CODE HERE!
mean_neighbor_year=sfo_data.groupby(['year','neighborhood']).mean().
↪reset_index()
```

[ ]:
```
mean_neighbor_year.loc[mean_neighbor_year['year']==2010].head(10)
```

[ ]:

	year	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
0	2010	Alamo Square	291.182945	372560	1239
1	2010	Anza Vista	267.932583	372560	1239
2	2010	Bayview	170.098665	372560	1239
3	2010	Buena Vista Park	347.394919	372560	1239
4	2010	Central Richmond	319.027623	372560	1239
5	2010	Central Sunset	418.172493	372560	1239
6	2010	Corona Heights	369.359338	372560	1239
7	2010	Cow Hollow	569.379968	372560	1239
8	2010	Croker Amazon	165.645730	372560	1239
9	2010	Diamond Heights	456.930822	372560	1239

[ ]:
```
# Use hvplot to create an interactive line chart of the average price per sq ft.
# The plot should have a dropdown selector for the neighborhood
# YOUR CODE HERE!
plot_hvplot_line(df=mean_neighbor_year,x_var='year',y_var='sale_price_sqr_foot',x_label='Year'
↪ Sale Price per SQFT',title='Average price per SQFT',groupby='neighborhood')
```

```
BokehModel(combine_events=True, render_bundle={'docs_json':␣
↪{'61013aa4-93e5-4173-a54a-ce4e446fb01c': {'defs': …
```

[ ]:
```
:DynamicMap   [neighborhood]
   :Curve   [year]   (sale_price_sqr_foot)
```

[ ]:
```
# Use hvplot to create an interactive line chart of the average monthly rent.
# The plot should have a dropdown selector for the neighborhood
# YOUR CODE HERE!
```

```
plot_hvplot_line(df=mean_neighbor_year,x_var='year',y_var='gross_rent',x_label='Year',y_label=
 ↪ gross rent',title='Avg. gross rent per year',groupby='neighborhood')
```

```
BokehModel(combine_events=True, render_bundle={'docs_json':␣
 ↪{'e331c836-4333-46f2-a2b5-d89991fe7985': {'defs': …
```

[ ]:
```
:DynamicMap   [neighborhood]
   :Curve   [year]   (gross_rent)
```

1.5 The Top 10 Most Expensive Neighborhoods

In this section, you will need to calculate the mean sale price per square foot for each neighborhood and then sort the values to obtain the top 10 most expensive neighborhoods on average. Plot the results as a bar chart.

[ ]:
```
# Getting the data from the top 10 expensive neighborhoods to own
# YOUR CODE HERE!
most_expensive_neighborhood=mean_neighbor_year[['neighborhood','sale_price_sqr_foot','housing_
 ↪groupby('neighborhood').mean().reset_index().
 ↪sort_values('sale_price_sqr_foot',ascending=False).head(10).
 ↪reset_index(drop=True)
most_expensive_neighborhood
```

[ ]:

	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
0	Union Square District	903.993258	377427.50	2555.166667
1	Merced Heights	788.844818	380348.00	3414.000000
2	Miraloma Park	779.810842	375967.25	2155.250000
3	Pacific Heights	689.555817	378401.00	2817.285714
4	Westwood Park	687.087575	382295.00	3959.000000
5	Telegraph Hill	676.506578	378401.00	2817.285714
6	Presidio Heights	675.350212	378401.00	2817.285714
7	Cow Hollow	665.964042	378401.00	2817.285714
8	Potrero Hill	662.013613	378401.00	2817.285714
9	South Beach	650.124479	375805.00	2099.000000

[ ]:
```
# Plotting the data from the top 10 expensive neighborhoods
# YOUR CODE HERE!
plot_hvplot_bar(df=most_expensive_neighborhood,x_var='neighborhood',y_var='sale_price_sqr_foot
 ↪ Sale Price per SQFT',title='TOP 10 expensive neighborhoods')
```

[ ]:
```
:Bars   [neighborhood]   (sale_price_sqr_foot)
```

1.6 Comparing cost to purchase versus rental income

In this section, you will use `hvplot` to create an interactive visualization with a dropdown selector for the neighborhood. This visualization will feature a side-by-side comparison of average price per square foot versus average montly rent by year.

Hint: Use the `hvplot` parameter, `groupby`, to create a dropdown selector for the neighborhood.

[ ]:
```
# Fetch the previously generated DataFrame that was grouped by year and␣
 ↪neighborhood
# YOUR CODE HERE!
mean_neighbor_year.head(10)
```

[ ]:

	year	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
0	2010	Alamo Square	291.182945	372560	1239
1	2010	Anza Vista	267.932583	372560	1239
2	2010	Bayview	170.098665	372560	1239
3	2010	Buena Vista Park	347.394919	372560	1239
4	2010	Central Richmond	319.027623	372560	1239
5	2010	Central Sunset	418.172493	372560	1239
6	2010	Corona Heights	369.359338	372560	1239
7	2010	Cow Hollow	569.379968	372560	1239
8	2010	Croker Amazon	165.645730	372560	1239
9	2010	Diamond Heights	456.930822	372560	1239

[ ]:
```
# Plotting the data from the top 10 expensive neighborhoods
# YOUR CODE HERE!
plot_hvplot_bar(mean_neighbor_year,x_var='year',
            y_var=['gross_rent','sale_price_sqr_foot'],
            x_label='Neighborhood',
            y_label='Num housing units',
            title='Top 10 Expensive Neighborhoods in SFO',
            groupby='neighborhood'
)
```

```
BokehModel(combine_events=True, render_bundle={'docs_json':␣
 ↪{'99c50bfc-237b-496d-b421-88aaa981509a': {'defs': …
```

[ ]:
```
:DynamicMap   [neighborhood]
   :Bars   [year,Variable]   (value)
```

1.7 Neighborhood Map

In this section, you will read in neighborhoods location data and build an interactive map with the average house value per neighborhood. Use a `scatter_mapbox` from Plotly express to create the visualization. Remember, you will need your Mapbox API key for this.

1.7.1 Load Location Data

[ ]:
```
%ls Data
```

```
 Volume in drive C is OS
 Volume Serial Number is 8A1A-7026
```

```
 Directory of c:\Users\ngond\python-homework\PyViz\Data

10/06/2021  09:23 PM    <DIR>          .
10/06/2021  09:23 PM    <DIR>          ..
10/06/2021  09:23 PM             2,635 neighborhoods_coordinates.csv
10/06/2021  09:23 PM            20,376 sfo_neighborhoods_census_data.csv
               2 File(s)         23,011 bytes
               2 Dir(s)  1,717,664,595,968 bytes free
```

[ ]:
```
# Load neighborhoods coordinates data
# YOUR CODE HERE!
file_path = Path("Data/neighborhoods_coordinates.csv")
sfo_coordinates = pd.read_csv(file_path)
sfo_coordinates.rename(columns={'Neighborhood':'neighborhood'},inplace=True)
sfo_coordinates.head()
```

[ ]:

	neighborhood	Lat	Lon
0	Alamo Square	37.791012	-122.402100
1	Anza Vista	37.779598	-122.443451
2	Bayview	37.734670	-122.401060
3	Bayview Heights	37.728740	-122.410980
4	Bernal Heights	37.728630	-122.443050

1.7.2 Data Preparation

You will need to join the location data with the mean values per neighborhood.

1. Calculate the mean values for each neighborhood.
2. Join the average values with the neighborhood locations.

[ ]:
```
# Calculate the mean values for each neighborhood
# YOUR CODE HERE!
mean_values_per_neighborhood=mean_neighbor_year[['neighborhood','sale_price_sqr_foot','housing
 ↪groupby('neighborhood').mean().reset_index()
mean_values_per_neighborhood.head()
```

[ ]:

	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
0	Alamo Square	366.020712	378401.0	2817.285714
1	Anza Vista	373.382198	379050.0	3031.833333
2	Bayview	204.588623	376454.0	2318.400000
3	Bayview Heights	590.792839	382295.0	3739.000000
4	Bernal Heights	576.746488	379374.5	3080.333333

[ ]:
```
# Join the average values with the neighborhood locations
# YOUR CODE HERE!
neighborhoods_with_loc=mean_values_per_neighborhood.
 ↪merge(sfo_coordinates,how='inner',on='neighborhood')
neighborhoods_with_loc.head()
```

[ ]:
```
       neighborhood  sale_price_sqr_foot  housing_units   gross_rent  \
0      Alamo Square           366.020712       378401.0  2817.285714
1        Anza Vista           373.382198       379050.0  3031.833333
2           Bayview           204.588623       376454.0  2318.400000
3   Bayview Heights           590.792839       382295.0  3739.000000
4  Buena Vista Park           452.680591       378076.5  2698.833333

         Lat         Lon
0  37.791012 -122.402100
1  37.779598 -122.443451
2  37.734670 -122.401060
3  37.728740 -122.410980
4  37.768160 -122.439330
```

[ ]:
```
def␣
↪plot_mapbox_scatter(df,latitude,longitude,size_var,color_var,color_continuous_scale_var,␣
↪  hover_name_VAR,
   title,   size_max_var=15,   zoom_var=11):

  canvas= px.scatter_mapbox(
    ␣
↪df,lat=latitude,lon=longitude,size=size_var,color=color_var,color_continuous_scale=color_co
     size_max=size_max_var,zoom=zoom_var,hover_name=hover_name_VAR,title=title
  )
  canvas.update_layout(mapbox_style="carto-positron")
  return canvas.show()
```

1.7.3 Mapbox Visualization

Plot the average values per neighborhood using a Plotly express `scatter_mapbox` visualization.

[ ]:
```
# Set the mapbox access token
# YOUR CODE HERE!

# Create a scatter mapbox to analyze neighborhood info
# YOUR CODE HERE!
plot_mapbox_scatter(df=neighborhoods_with_loc,latitude='Lat',longitude='Lon',size_var='sale_pr
           color_continuous_scale_var=px.colors.cyclical.
↪IceFire,hover_name_VAR='neighborhood',
           title='Average Sale Price Per SQFT and gross rent in San francisco'
       )
```

[ ]:
```
df_expensive_neighborhoods_per_year =␣
↪mean_neighbor_year[mean_neighbor_year["neighborhood"].
↪isin(most_expensive_neighborhood["neighborhood"])]
df_expensive_neighborhoods_per_year.head()
```

[ ]:

	year	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
7	2010	Cow Hollow	569.379968	372560	1239
31	2010	Miraloma Park	680.608729	372560	1239
41	2010	Pacific Heights	496.516014	372560	1239
46	2010	Potrero Hill	491.450004	372560	1239
47	2010	Presidio Heights	549.417931	372560	1239

1.8 Cost Analysis - Optional Challenge

In this section, you will use Plotly express to create visualizations that investors can use to interactively filter and explore various factors related to the house value of the San Francisco's neighborhoods.

1.8.1 Create a DataFrame showing the most expensive neighborhoods in San Francisco by year

[ ]:
```
# Fetch the data from all expensive neighborhoods per year.
df_expensive_neighborhoods_per_year =␣
↪mean_neighbor_year[mean_neighbor_year["neighborhood"].
↪isin(most_expensive_neighborhood["neighborhood"])]
df_expensive_neighborhoods_per_year.head()
```

[ ]:

	year	neighborhood	sale_price_sqr_foot	housing_units	gross_rent
7	2010	Cow Hollow	569.379968	372560	1239
31	2010	Miraloma Park	680.608729	372560	1239
41	2010	Pacific Heights	496.516014	372560	1239
46	2010	Potrero Hill	491.450004	372560	1239
47	2010	Presidio Heights	549.417931	372560	1239

1.8.2 Create a parallel coordinates plot and parallel categories plot of most expensive neighborhoods in San Francisco per year

[ ]:
```
```

[ ]:
```
# Parallel Categories Plot
# YOUR CODE HERE!
px.parallel_categories(most_expensive_neighborhood,color='sale_price_sqr_foot',
    color_continuous_scale=px.colors.sequential.Inferno,)
```

[ ]:
```
# Parallel Coordinates Plot
# YOUR CODE HERE!
most_expensive_neighborhood
px.parallel_coordinates(most_expensive_neighborhood,color='sale_price_sqr_foot')
```

1.8.3 Create a sunburst chart to conduct a costs analysis of most expensive neighborhoods in San Francisco per year

[ ]:
```
# Sunburst Plot
# YOUR CODE HERE!
fig = px.sunburst(
    df_expensive_neighborhoods_per_year,
    path=["year", "neighborhood"],
    values='sale_price_sqr_foot',
    color='gross_rent',
    color_continuous_scale='Blues',
    title="Costs Analysis of Most Expensive neighborhoods in San Francisco per␣
 ↪Year",
    height=700
)
fig.show()
```

[ ]:
```

```