HOME PRICE OVER YEARS IN THE US

•••

KOJI SERA

PROJECT: CREATE CHARTS TO VISUALIZE US HOUSING PRICES OVER YEARS

Kaggle.com lists US housing market index data available to the public.

This project will use this data for my data extraction and data visualization project.

1. Data cleaning: Extracting data for visual presentation
2. Visualization: Presenting the housing price chart for a better understanding

1. DATA CLEANING (data columns)

The original data is composed of by the following columns;

1. hpi_type (area type: developmental, traditional, non-metro, distress-free)
2. hpi_flavor (transaction type: all-transactions, expanded-area, purchase-only)
3. frequency (data frequency: monthly, quarterly)
4. level (MSA, Puerto Rico, State, USA or Census Division)
5. place_name (major city in each state. format = ‘Akron, OH’, ‘Asheville, NC’, etc)
6. place_id (state level information: Zip code or state abs)
7. yr (year)
8. period (month)
9. index_nsa (non-seasonally adjusted price)
10. index_sa (seasonally adjusted price)

1. DATA CLEANING (python coding)

This project needed only state name, year, housing price (python code shown below)

```
houseprice_ave = []
year_of_ave = []
state_name = []

for i in sorted(HPI.place_id.unique()):
    for j in sorted(HPI.yr.unique()):
        ave_price = HPI.index_nsa[HPI.level
=='State'][HPI.place_id == i][HPI.yr ==j].mean()
        houseprice_ave.append(ave_price)
        year_of_ave.append(j)
        state_name.append(i)

df = pd.DataFrame([houseprice_ave,
year_of_ave,state_name]).T
df.columns = ['price', 'year','state']
df.to_csv('House Price List.csv')
```

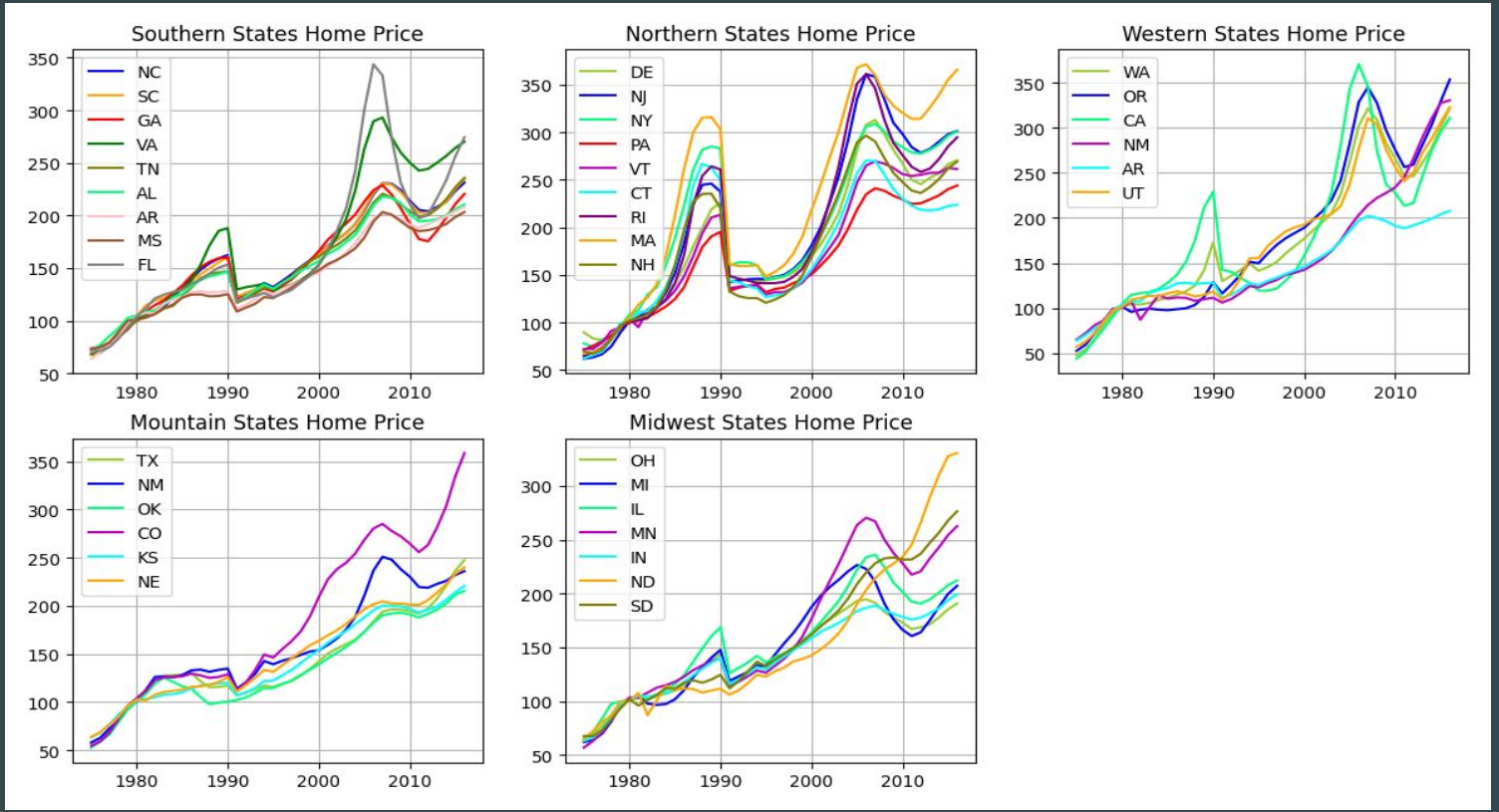
2. DATA VISUALIZATION (python coding)

After cleaning data, visualization is done by the following code:

```
df = pd.read_csv('House Price List.csv')
df['price'].replace('', np.nan, inplace = True)
df.dropna(subset = ['price'], inplace = True)

table = df.pivot(index = 'state', columns = 'year', values = 'state')

plt.figure(figsize = (6,4))
plt.subplot(2,3,1)
plt.plot(df.year[df.state == 'NC'], df.price[df.state == 'NC'],color='blue')
plt.plot(df.year[df.state == 'SC'], df.price[df.state == 'SC'],color ='orange')
plt.plot(df.year[df.state == 'GA'], df.price[df.state == 'GA'],color ='r')
plt.plot(df.year[df.state == 'VA'], df.price[df.state == 'VA'],color ='g')

plt.grid()
plt.legend(['NC','SC','GA','VA','TN','AL','AR','MS','FL'])
plt.title('Southern States Home Price')

plt.show()
plt.savefig('US home price.png')
```

Note: the actual code includes more states

2. DATA VISUALIZATION (output)



2. DATA VISUALIZATION (findings)

Findings:

1. Housing market has grown over time
2. The entire US market was hit by 2 economic crises: recession in 1990 and financial crisis in 2007-8
3. Impacts from recession in 1990 were different depending on area (northern states received the biggest impact)
4. Financial crisis in 2007-8 had bigger impacts on the entire US

For future analysis:

Housing market growth differs in each state, and multiple economic and other factors such as population transitions can be introduced and matched for better understanding