

seaborn_visualization

March 22, 2023

```
[3]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns

mydata="data_source/california_housing_test.csv"
data1=pd.read_csv(mydata)
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   longitude              3000 non-null   float64
1   latitude               3000 non-null   float64
2   housing_median_age     3000 non-null   float64
3   total_rooms            3000 non-null   float64
4   total_bedrooms         3000 non-null   float64
5   population             3000 non-null   float64
6   households             3000 non-null   float64
7   median_income          3000 non-null   float64
8   median_house_value     3000 non-null   float64
dtypes: float64(9)
memory usage: 211.1 KB
```

```
[ ]:
```

```
[4]: data1.describe()
```

```
[4]:
```

	longitude	latitude	housing_median_age	total_rooms	\
count	3000.000000	3000.000000	3000.000000	3000.000000	
mean	-119.589200	35.63539	28.845333	2599.578667	
std	1.994936	2.12967	12.555396	2155.593332	
min	-124.180000	32.56000	1.000000	6.000000	
25%	-121.810000	33.93000	18.000000	1401.000000	
50%	-118.485000	34.27000	29.000000	2106.000000	

75%	-118.020000	37.69000	37.000000	3129.000000
max	-114.490000	41.92000	52.000000	30450.000000

	total_bedrooms	population	households	median_income	\
count	3000.000000	3000.000000	3000.000000	3000.000000	
mean	529.950667	1402.798667	489.91200	3.807272	
std	415.654368	1030.543012	365.42271	1.854512	
min	2.000000	5.000000	2.000000	0.499900	
25%	291.000000	780.000000	273.000000	2.544000	
50%	437.000000	1155.000000	409.500000	3.487150	
75%	636.000000	1742.750000	597.250000	4.656475	
max	5419.000000	11935.000000	4930.000000	15.000100	

	median_house_value
count	3000.000000
mean	205846.27500
std	113119.68747
min	22500.000000
25%	121200.000000
50%	177650.000000
75%	263975.000000
max	500001.000000

```
[5]: data1.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	\
0	-122.05	37.37	27.0	3885.0	661.0	
1	-118.30	34.26	43.0	1510.0	310.0	
2	-117.81	33.78	27.0	3589.0	507.0	
3	-118.36	33.82	28.0	67.0	15.0	
4	-119.67	36.33	19.0	1241.0	244.0	

	population	households	median_income	median_house_value
0	1537.0	606.0	6.6085	344700.0
1	809.0	277.0	3.5990	176500.0
2	1484.0	495.0	5.7934	270500.0
3	49.0	11.0	6.1359	330000.0
4	850.0	237.0	2.9375	81700.0

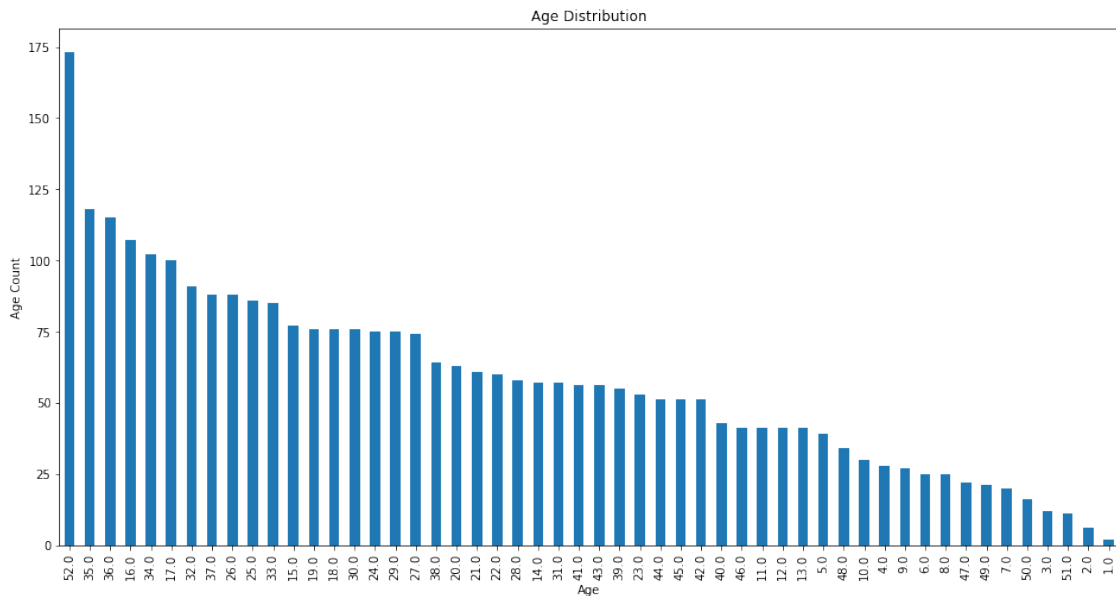
```
[6]: data1.tail()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	\
2995	-119.86	34.42	23.0	1450.0	642.0	
2996	-118.14	34.06	27.0	5257.0	1082.0	
2997	-119.70	36.30	10.0	956.0	201.0	
2998	-117.12	34.10	40.0	96.0	14.0	
2999	-119.63	34.42	42.0	1765.0	263.0	

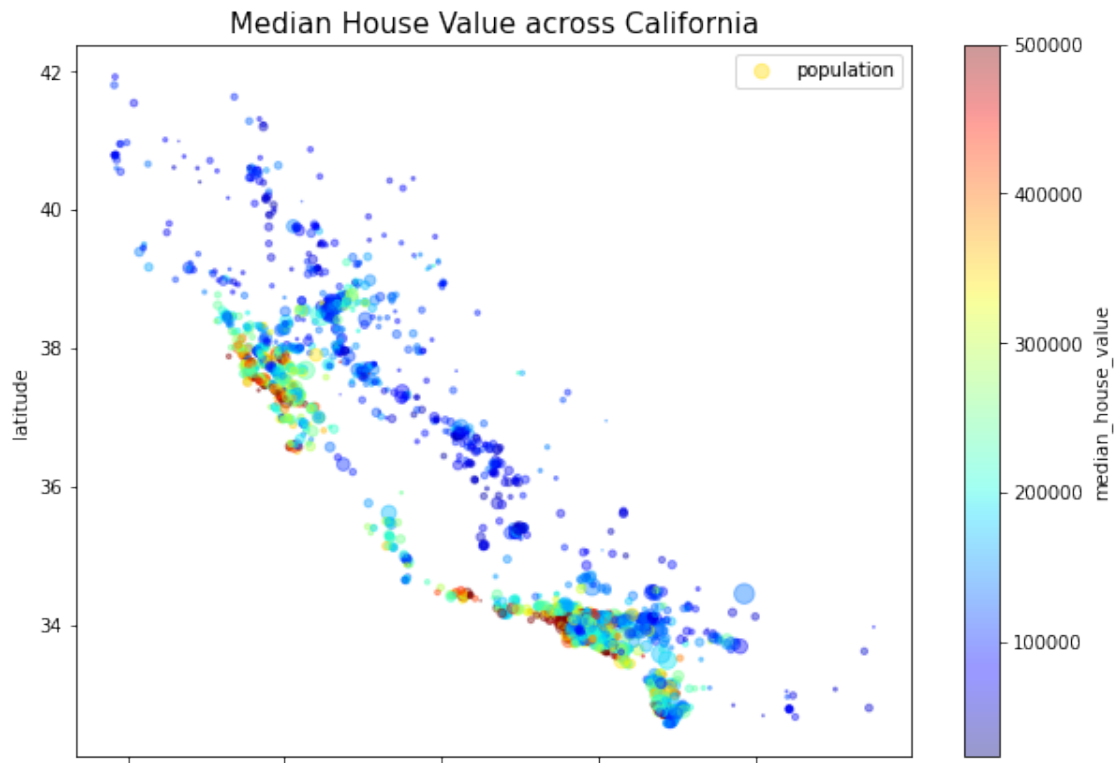
	population	households	median_income	median_house_value
2995	1258.0	607.0	1.1790	225000.0
2996	3496.0	1036.0	3.3906	237200.0
2997	693.0	220.0	2.2895	62000.0
2998	46.0	14.0	3.2708	162500.0
2999	753.0	260.0	8.5608	500001.0

```
[ ]: data1.set_index('housing_median_age').plot(kind='bar')
```

```
[ ]: plt.figure(figsize=(16,8))
data1['housing_median_age'].value_counts().plot(kind='bar')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Age Count")
plt.show()
```



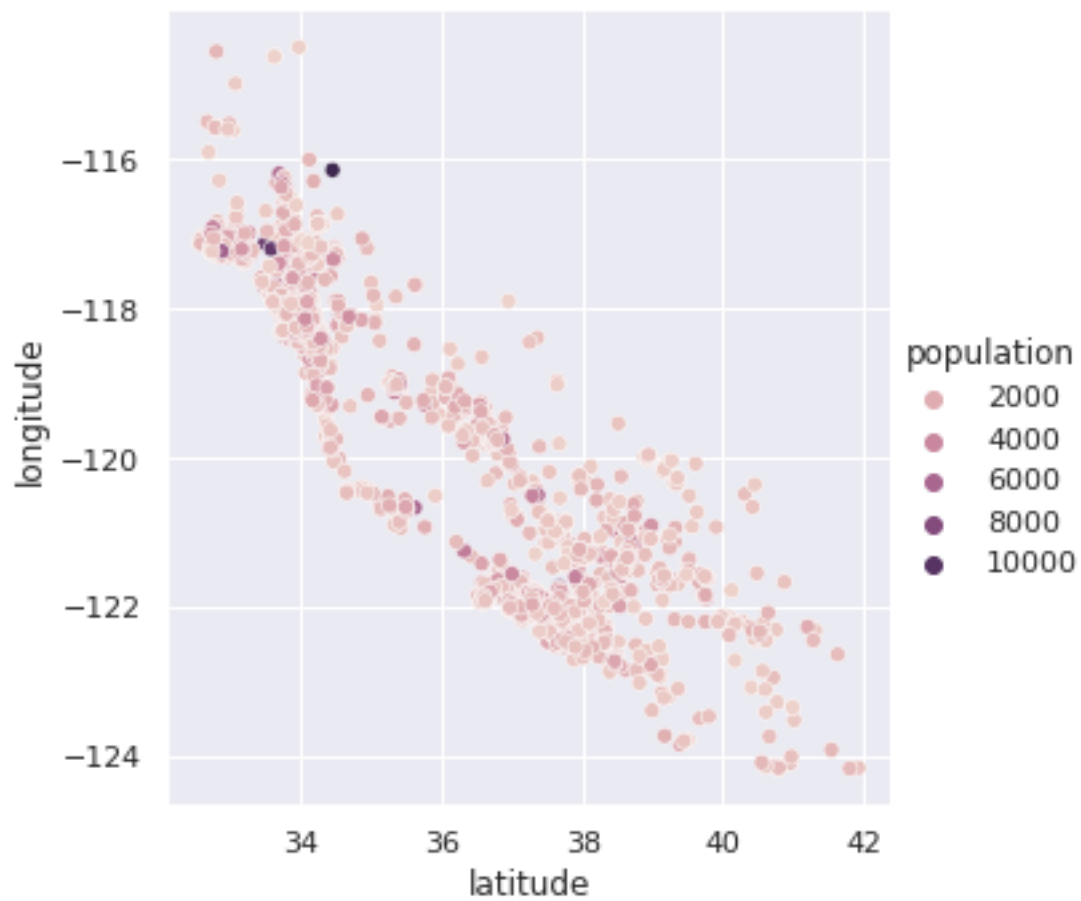
```
[ ]: data1.plot(kind='scatter', x='longitude', y='latitude', alpha=0.4,
s=data['population']/100, label='population', figsize=(10,7),
c='median_house_value', cmap=plt.get_cmap('jet'), colorbar=True)
plt.title('Median House Value across California',fontsize=15)
plt.legend()
plt.show()
```



```
[8]: sns.set_theme()  
      # apply the default theme
```

```
[9]: sns.relplot(data= data1, x="latitude", y="longitude", hue="population")
```

```
[9]: <seaborn.axisgrid.FacetGrid at 0x7f35109eba90>
```



[]: