

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

```
In [ ]: df1 = pd.read_csv('housing 2.csv')
print(df.info())
df1.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   longitude             20640 non-null  float64
1   latitude              20640 non-null  float64
2   housing_median_age    20640 non-null  float64
3   total_rooms           20640 non-null  float64
4   total_bedrooms        20640 non-null  float64
5   population            20640 non-null  float64
6   households            20640 non-null  float64
7   median_income         20640 non-null  float64
8   median_house_value    20640 non-null  float64
9   ocean_proximity       20640 non-null  object
dtypes: float64(9), object(1)
memory usage: 1.6+ MB
None
```

```
Out[ ]: longitude          0
latitude                 0
housing_median_age       0
total_rooms              0
total_bedrooms          207
population               0
households               0
median_income            0
median_house_value       0
ocean_proximity          0
dtype: int64
```

```
In [ ]: df1
```

Out[]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
...
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	78100.0	INLAND
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	77100.0	INLAND
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	92300.0	INLAND
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	84700.0	INLAND
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	89400.0	INLAND

20640 rows × 10 columns

In []:

```
df = df1.fillna(method='bfill')
print("_____")
df1.isnull().sum()
```

Out[]:

longitude

0

latitude

0

housing_median_age

0

total_rooms

0

total_bedrooms

207

population

0

households

0

median_income

0

median_house_value

0

ocean_proximity

0

dtype: int64

In []:

```
df.corr()
```

<ipython-input-75-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
df.corr()

Out[]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
longitude	1.000000	-0.924664	-0.108197	0.044568	0.069763	0.099773	0.055310	-0.015176	-0.045967
latitude	-0.924664	1.000000	0.011173	-0.036100	-0.067344	-0.108785	-0.071035	-0.079809	-0.144160
housing_median_age	-0.108197	0.011173	1.000000	-0.361262	-0.318956	-0.296244	-0.302916	-0.119034	0.105623
total_rooms	0.044568	-0.036100	-0.361262	1.000000	0.924855	0.857126	0.918484	0.198050	0.134153
total_bedrooms	0.069763	-0.067344	-0.318956	0.924855	1.000000	0.871378	0.972279	-0.006249	0.050665
population	0.099773	-0.108785	-0.296244	0.857126	0.871378	1.000000	0.907222	0.004834	-0.024650
households	0.055310	-0.071035	-0.302916	0.918484	0.972279	0.907222	1.000000	0.013033	0.065843
median_income	-0.015176	-0.079809	-0.119034	0.198050	-0.006249	0.004834	0.013033	1.000000	0.688075
median_house_value	-0.045967	-0.144160	0.105623	0.134153	0.050665	-0.024650	0.065843	0.688075	1.000000

In []:

```
df_corr=df.corr()  
print( df_corr[df_corr>=0.5])  
print("\n-----\n")  
print(df_corr[df_corr<= -0.5])
```

	longitude	latitude	housing_median_age	total_rooms	\
longitude	1.0	NaN	NaN	NaN	
latitude	NaN	1.0	NaN	NaN	
housing_median_age	NaN	NaN	1.0	NaN	
total_rooms	NaN	NaN	NaN	1.000000	
total_bedrooms	NaN	NaN	NaN	0.924855	
population	NaN	NaN	NaN	0.857126	
households	NaN	NaN	NaN	0.918484	
median_income	NaN	NaN	NaN	NaN	
median_house_value	NaN	NaN	NaN	NaN	

	total_bedrooms	population	households	median_income	\
longitude	NaN	NaN	NaN	NaN	
latitude	NaN	NaN	NaN	NaN	
housing_median_age	NaN	NaN	NaN	NaN	
total_rooms	0.924855	0.857126	0.918484	NaN	
total_bedrooms	1.000000	0.871378	0.972279	NaN	
population	0.871378	1.000000	0.907222	NaN	
households	0.972279	0.907222	1.000000	NaN	
median_income	NaN	NaN	NaN	1.000000	
median_house_value	NaN	NaN	NaN	0.688075	

	median_house_value
longitude	NaN
latitude	NaN
housing_median_age	NaN
total_rooms	NaN
total_bedrooms	NaN
population	NaN
households	NaN
median_income	0.688075
median_house_value	1.000000

	longitude	latitude	housing_median_age	total_rooms	\
longitude	NaN	-0.924664	NaN	NaN	
latitude	-0.924664	NaN	NaN	NaN	
housing_median_age	NaN	NaN	NaN	NaN	
total_rooms	NaN	NaN	NaN	NaN	
total_bedrooms	NaN	NaN	NaN	NaN	
population	NaN	NaN	NaN	NaN	
households	NaN	NaN	NaN	NaN	
median_income	NaN	NaN	NaN	NaN	
median_house_value	NaN	NaN	NaN	NaN	

	total_bedrooms	population	households	median_income	\
longitude	NaN	NaN	NaN	NaN	
latitude	NaN	NaN	NaN	NaN	
housing_median_age	NaN	NaN	NaN	NaN	
total_rooms	NaN	NaN	NaN	NaN	
total_bedrooms	NaN	NaN	NaN	NaN	
population	NaN	NaN	NaN	NaN	
households	NaN	NaN	NaN	NaN	
median_income	NaN	NaN	NaN	NaN	
median_house_value	NaN	NaN	NaN	NaN	

	median_house_value
longitude	NaN
latitude	NaN

housing_median_age	NaN
total_rooms	NaN
total_bedrooms	NaN
population	NaN
households	NaN
median_income	NaN
median_house_value	NaN

<ipython-input-76-8f9de22b732a>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

df_corr=df.corr()

```
In [ ]: # High Correlations:
# ('total_rooms', 'total_bedrooms'),
# ('total_rooms', 'population'),
# ('total_rooms', 'households'),
# ('median_income', 'median_house_value'),
# Low Correlations:
# ('Longitude', 'Latitude')
```

CORELATION IN THIS DATA

1.Longitude and Latitude:

___There is a strong negative correlation (-0.92) between longitude and latitude.

2.Total Rooms :

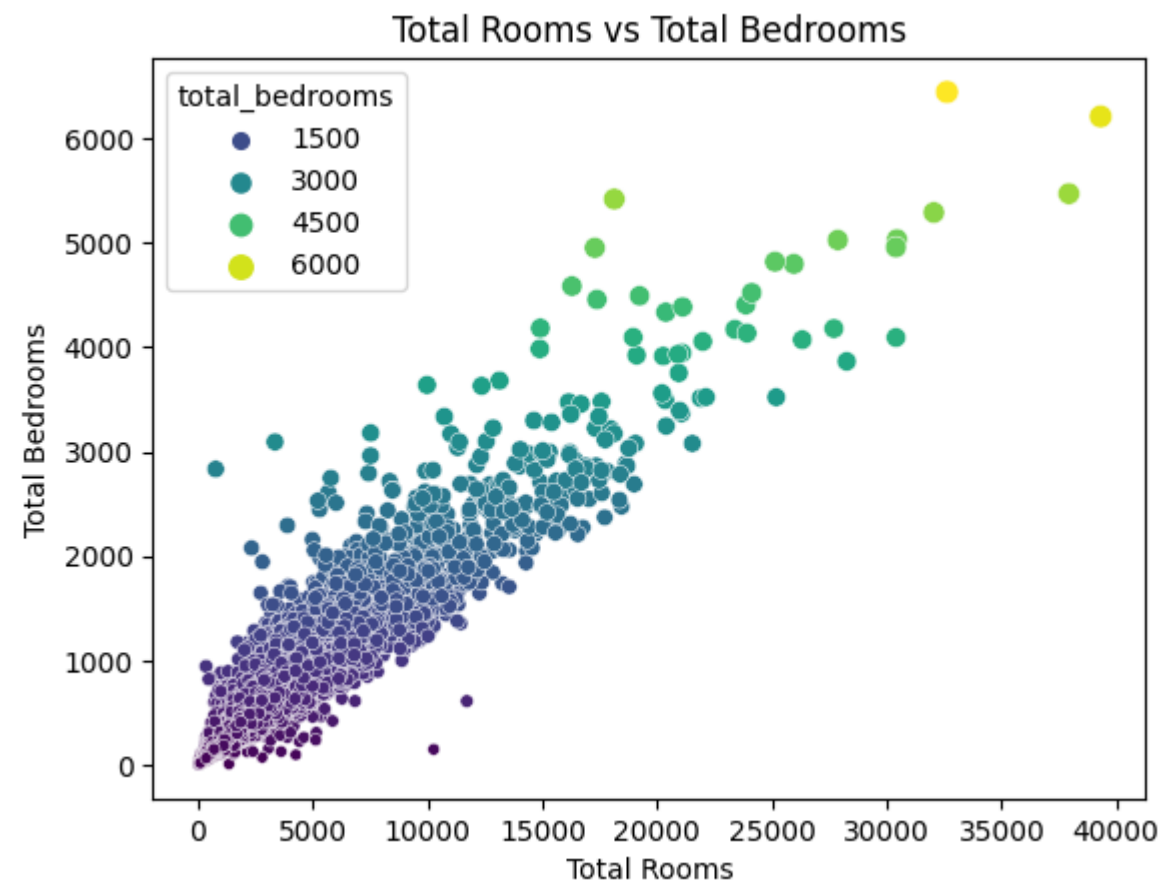
___Total_rooms has strong positive correlations with total_bedrooms, population, and households.

3.Median House Value :

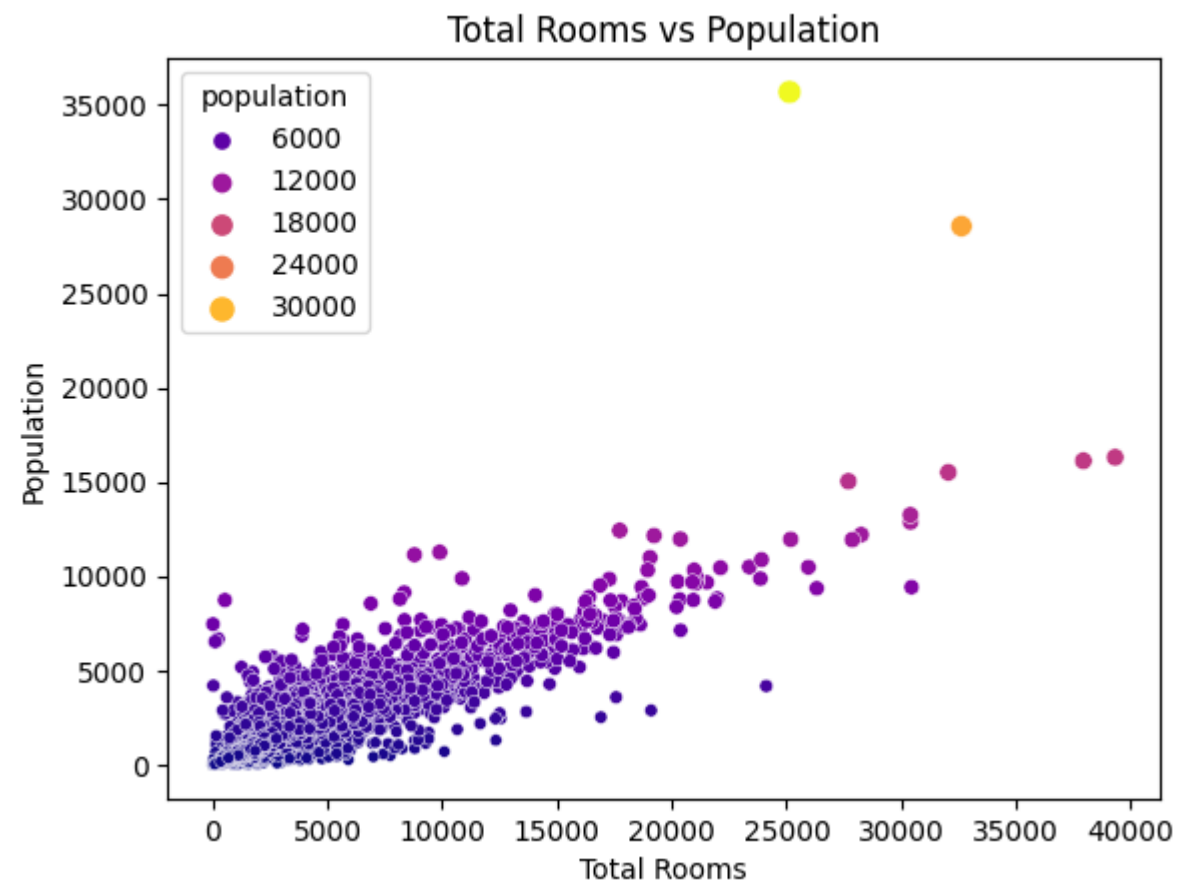
___Median_house_value has a strong positive correlation with median_income

```
In [ ]: import seaborn as sns
import matplotlib.pyplot as plt

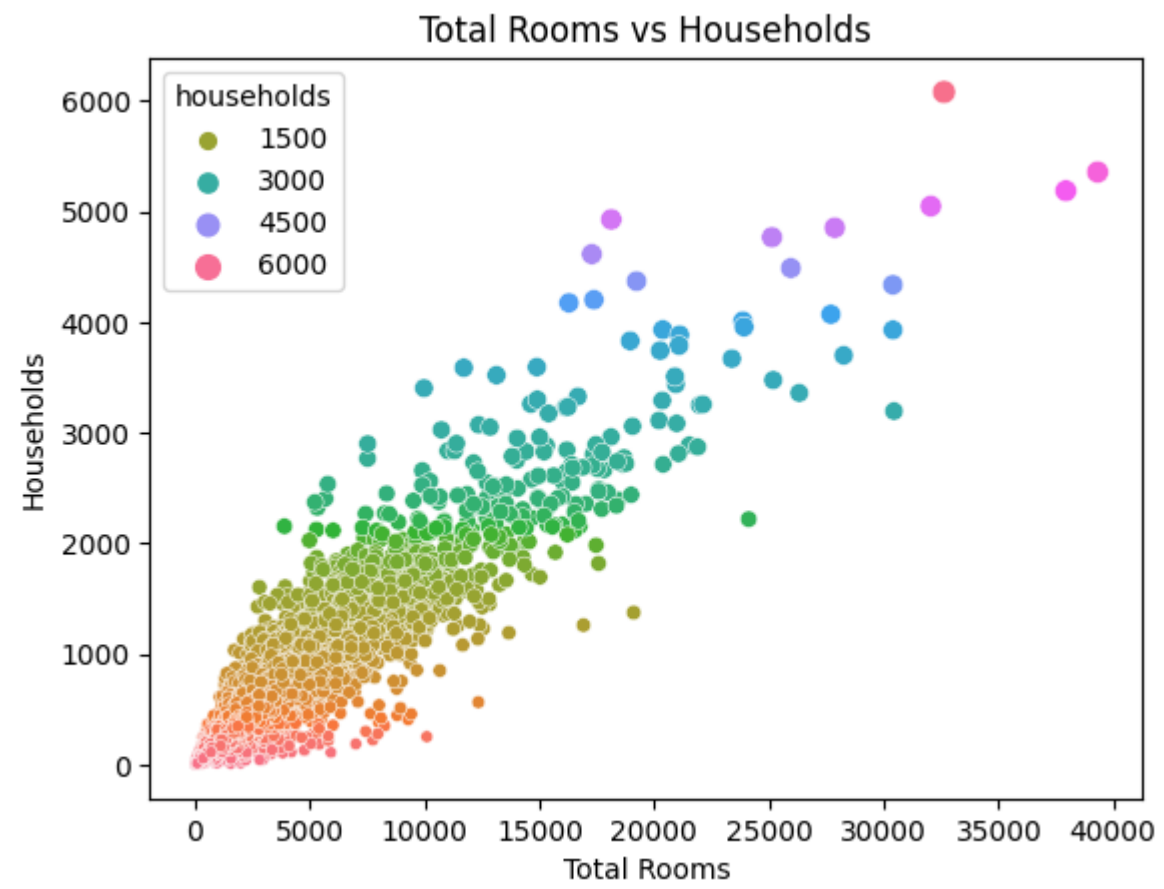
# Scatter Plot - Total Rooms vs. Total Bedrooms with gradient and color palette
sns.scatterplot(x='total_rooms', y='total_bedrooms', data=df, hue='total_bedrooms', palette='viridis', size='total_bedrooms')
plt.title('Total Rooms vs Total Bedrooms')
plt.xlabel('Total Rooms')
plt.ylabel('Total Bedrooms')
plt.show()
```



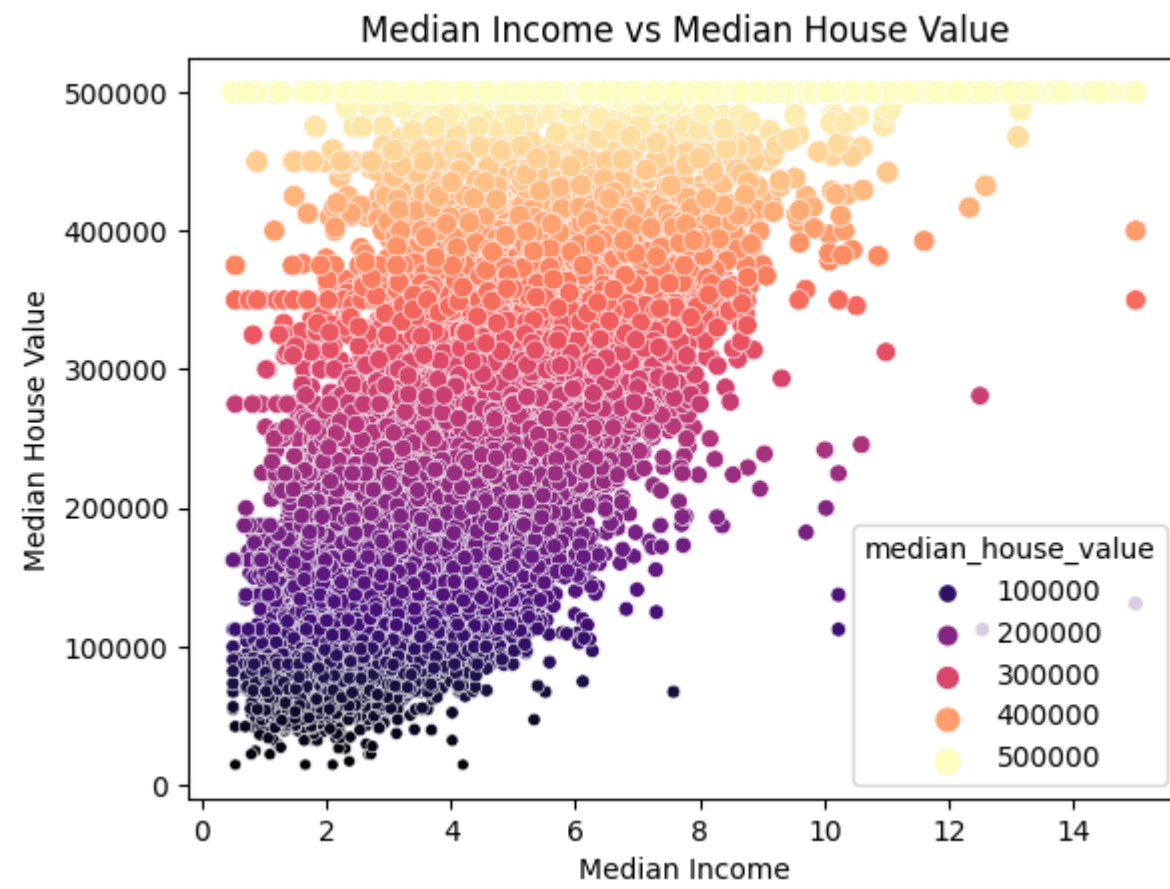
```
In [ ]: # Scatter Plot - Total Rooms vs. Population with gradient and color palette
sns.scatterplot(x='total_rooms', y='population', data=df, hue='population', palette='plasma', size='population')
plt.title('Total Rooms vs Population')
plt.xlabel('Total Rooms')
plt.ylabel('Population')
plt.show()
```



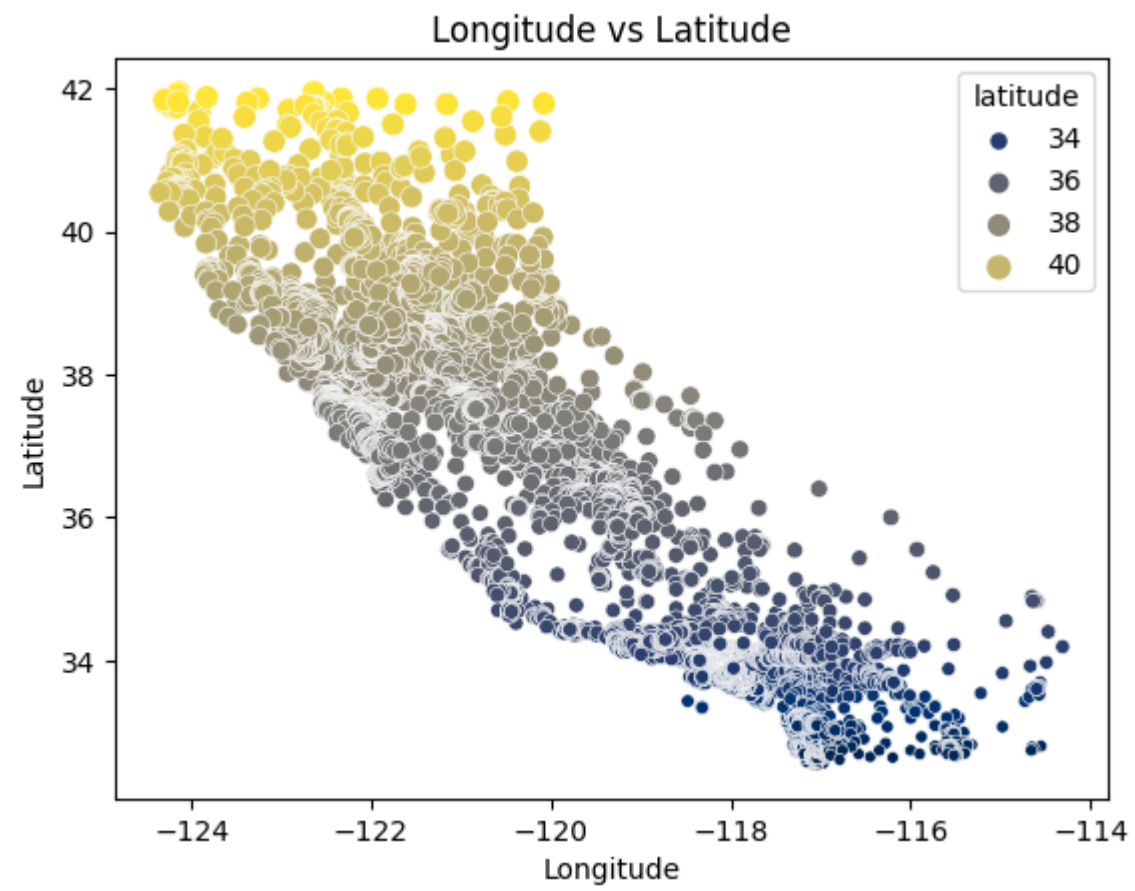
```
In [ ]: # Scatter Plot - Total Rooms vs. Households with gradient and color palette
sns.scatterplot(x='total_rooms', y='households', data=df, hue='households', palette='husl', size='households')
plt.title('Total Rooms vs Households')
plt.xlabel('Total Rooms')
plt.ylabel('Households')
plt.show()
```



```
In [ ]: # Scatter Plot - Median Income vs. Median House Value with gradient and color palette
sns.scatterplot(x='median_income', y='median_house_value', data=df, hue='median_house_value', palette='magma', size='median_house_value')
plt.title('Median Income vs Median House Value')
plt.xlabel('Median Income')
plt.ylabel('Median House Value')
plt.show()
```

```
In [ ]: # Scatter Plot - Longitude vs. Latitude with gradient and color palette
sns.scatterplot(x='longitude', y='latitude', data=df, hue='latitude', palette='cividis', size='latitude')
plt.title('Longitude vs Latitude')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.show()
```

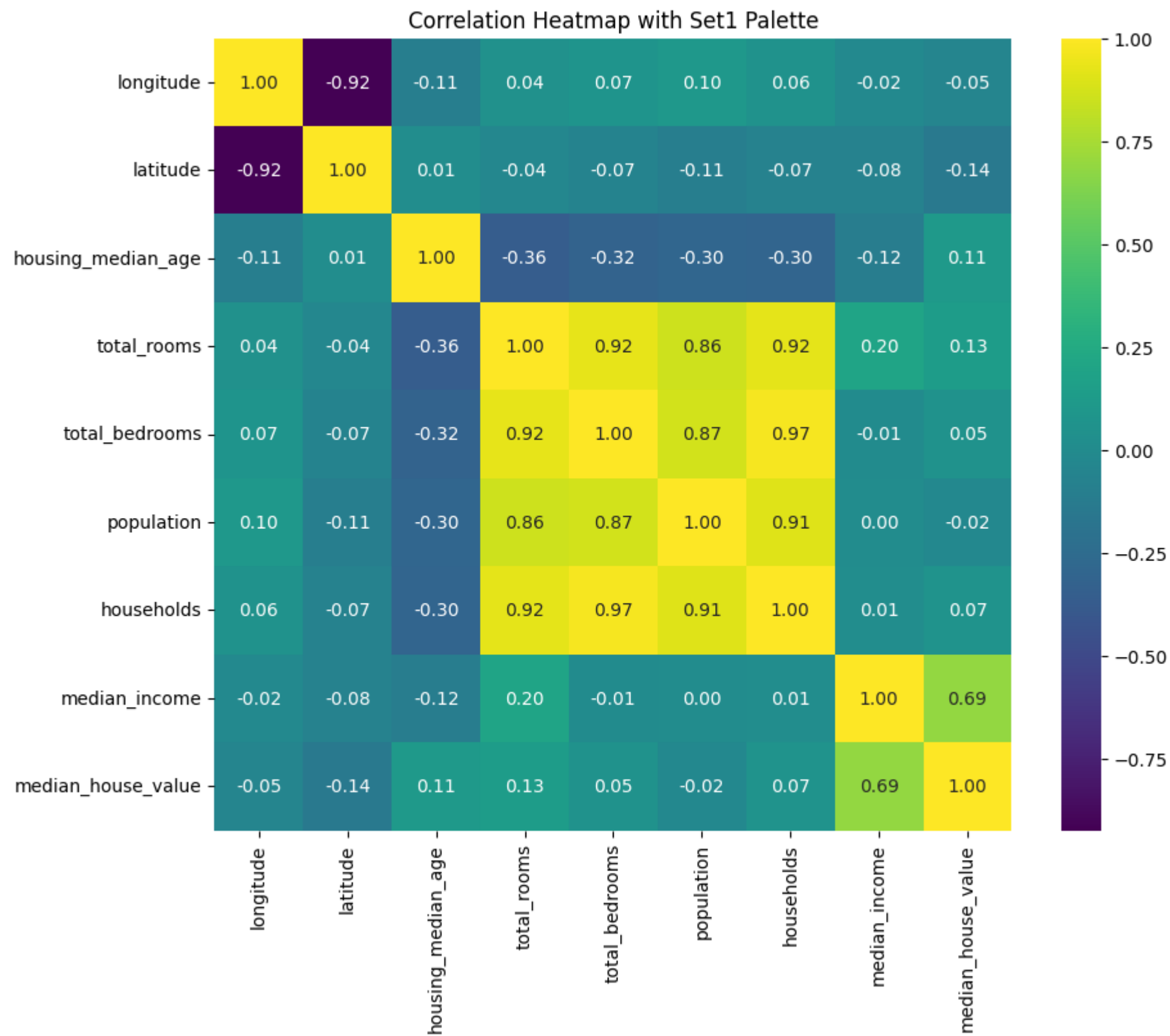


```
In [ ]: corr_matrix = df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(df_corr, annot=True, cmap="viridis", fmt='.2f')

plt.title('Correlation Heatmap with Set1 Palette')
plt.show()
```

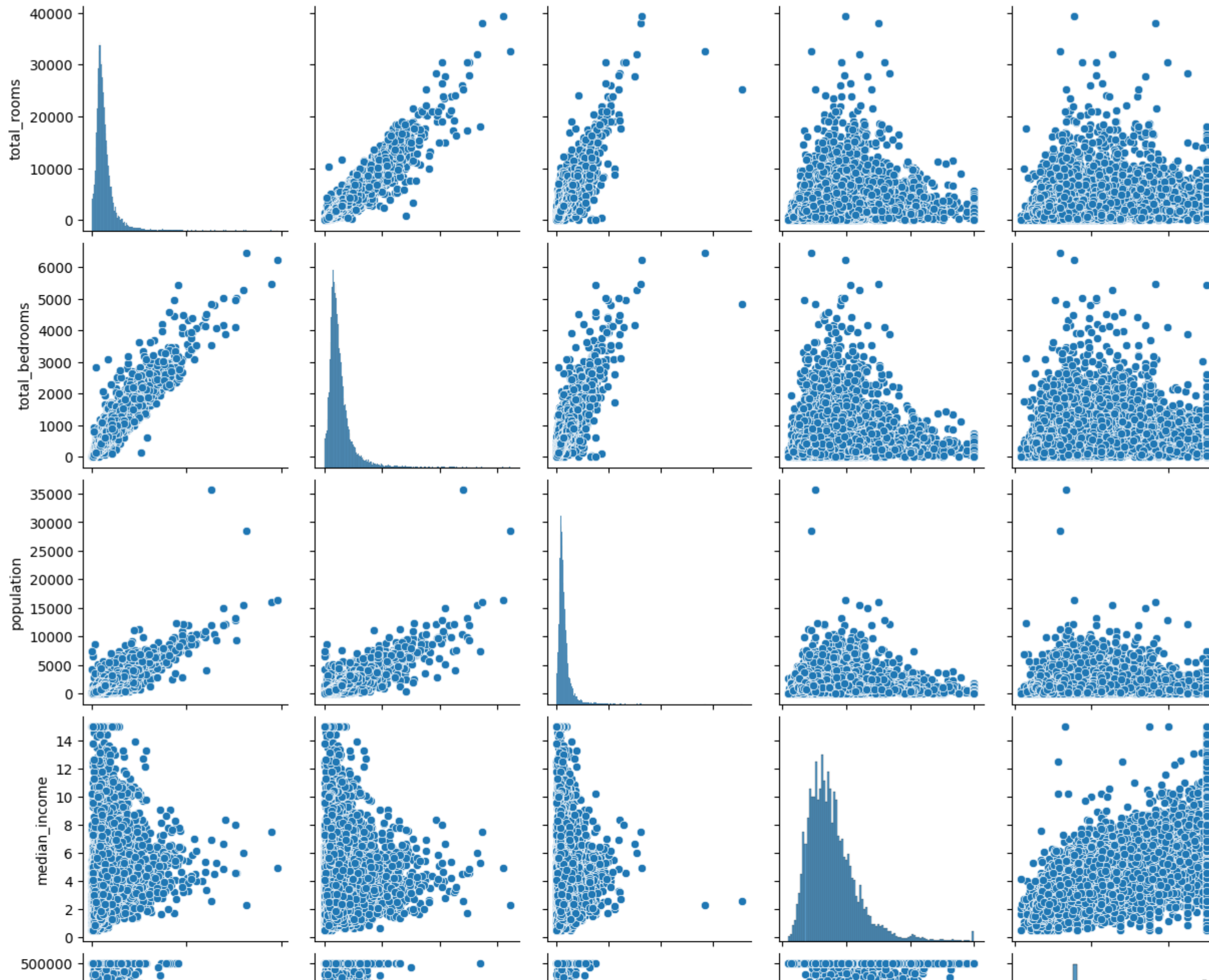
<ipython-input-94-da08ae9b2e6d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

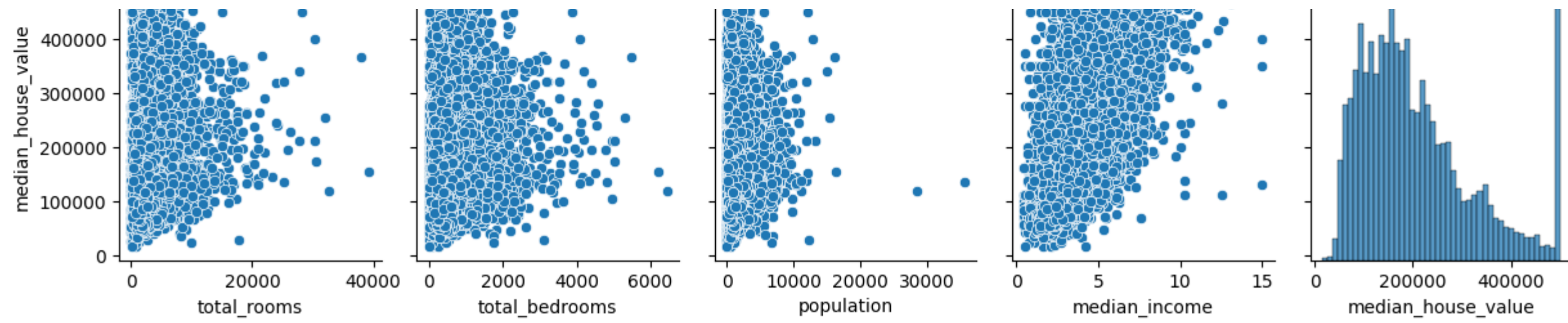
```
corr_matrix = df.corr()
```



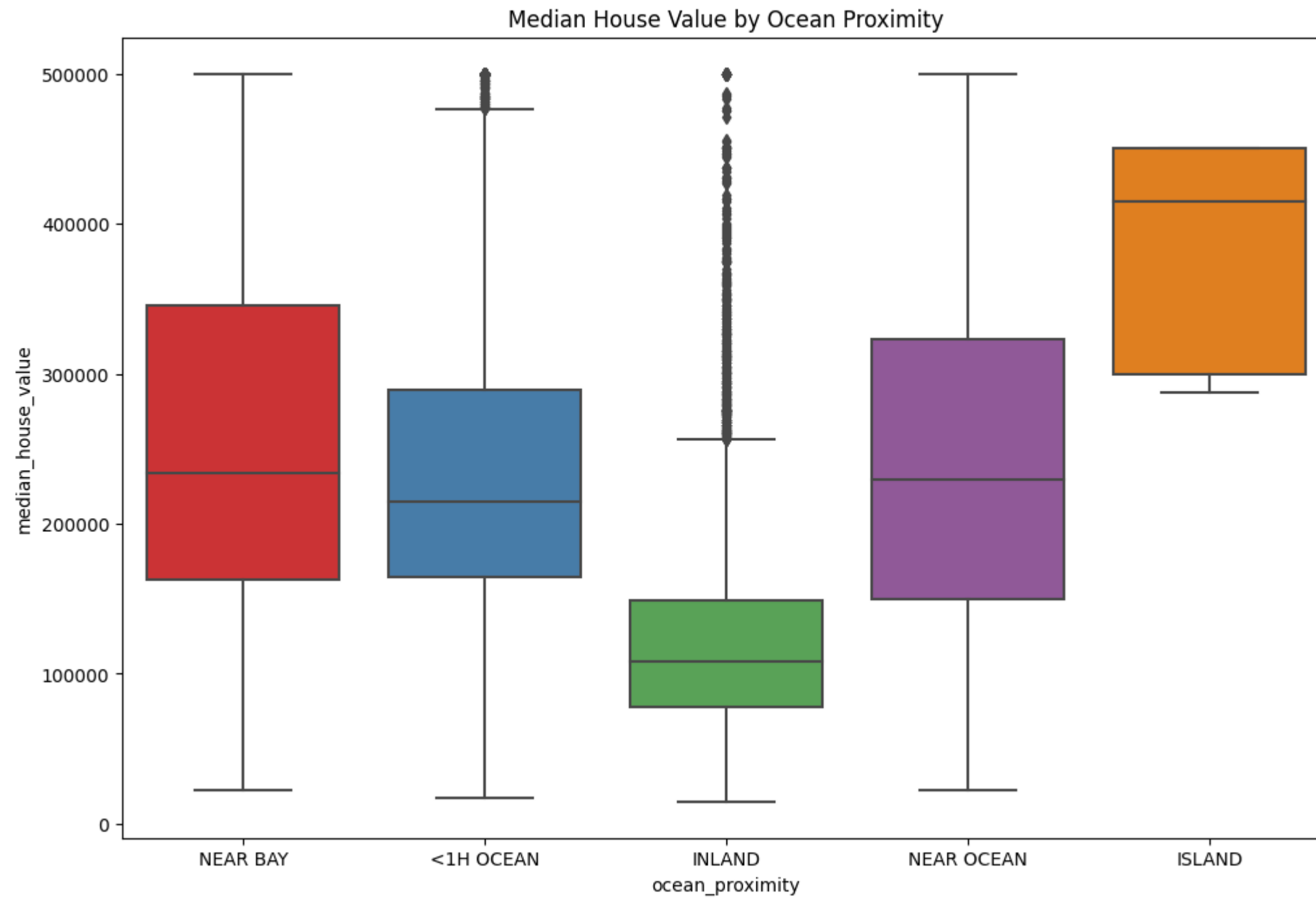
```
In [ ]: # Pair Plot for selected columns
sns.pairplot(df[['total_rooms', 'total_bedrooms', 'population', 'median_income', 'median_house_value']], height=2.5)
plt.suptitle('Pair Plot: Key Variables', y=1.02)
plt.show()
```

Pair Plot: Key Variables





```
In [ ]: # Box Plot for median house value by ocean proximity
plt.figure(figsize=(12, 8))
sns.boxplot(x='ocean_proximity', y='median_house_value', data=df, palette='Set1')
plt.title('Median House Value by Ocean Proximity')
plt.show()
```

Hence there is one more relation that median price is inversely proptnal to the ocean proximity

price increase with decrease in dist from ocean