Data Analysis Report

Dataset Summary:

Number of Rows: 20640
Number of Columns: 10

Insights:

Detailed Insights Report

The provided dataset contains 20640 rows and 10 columns, offering a comprehensive overview of various demographic and socioeconomic factors. This report aims to identify key trends, anomalies, and correlations within the data, as well as detect outliers and unusual patterns.

Summary Statistics

The summary statistics reveal the following key insights:

* **Longitude**: The mean longitude is -119.57, with a standard deviation of 2.00. The minimum and maximum values are -124.35 and -114.31, respectively.
* **Latitude**: The mean latitude is 35.63, with a standard deviation of 2.14. The minimum and maximum values are 32.54 and 41.95, respectively.
* **Housing Median Age**: The mean housing median age is 28.64, with a standard deviation of 12.59. The minimum and maximum values are 1.0 and 52.0, respectively.
* **Total Rooms**: The mean total rooms is 2635.76, with a standard deviation of 2181.62. The minimum and maximum values are 2.0 and 39320.0, respectively.
* **Total Bedrooms**: The mean total bedrooms is 537.87, with a standard deviation of 421.39. The minimum and maximum values are 1.0 and 6445.0, respectively.
* **Population**: The mean population is 1425.48, with a standard deviation of 1132.46. The minimum and maximum values are 3.0 and 35682.0, respectively.
* **Households**: The mean households is 499.54, with a standard deviation of 382.33. The minimum and maximum values are 1.0 and 6082.0, respectively.
* **Median Income**: The mean median income is 3.87, with a standard deviation of 1.90. The minimum and maximum values are 0.50 and 15.00, respectively.
* **Median House Value**: The mean median house value is 206855.82, with a standard deviation of 115395.62. The minimum and maximum values are 14999.0 and 500001.0, respectively.

Correlation Matrix

The correlation matrix reveals the following key insights:

* **Longitude and Latitude**: The correlation between longitude and latitude is -0.92, indicating a strong negative relationship.

* **Housing Median Age and Total Rooms**: The correlation between housing median age and total rooms is -0.36, indicating a moderate negative relationship.
* **Total Rooms and Total Bedrooms**: The correlation between total rooms and total bedrooms is 0.93, indicating a strong positive relationship.
* **Population and Households**: The correlation between population and households is 0.91, indicating a strong positive relationship.
* **Median Income and Median House Value**: The correlation between median income and median house value is 0.69, indicating a strong positive relationship.

Key Trends and Anomalies

* The data suggests a strong negative relationship between longitude and latitude, indicating that as longitude increases, latitude decreases.
* The data also suggests a strong positive relationship between total rooms and total bedrooms, indicating that as the number of rooms increases, the number of bedrooms also increases.
* The median house value is strongly correlated with median income, indicating that as median income increases, median house value also increases.
* The data reveals a moderate negative relationship between housing median age and total rooms, indicating that as housing median age increases, the number of rooms decreases.

Outliers and Unusual Patterns

* The minimum and maximum values for total rooms (2.0 and 39320.0) and total bedrooms (1.0 and 6445.0) suggest the presence of outliers in the data.
* The minimum and maximum values for population (3.0 and 35682.0) and households (1.0 and 6082.0) also suggest the presence of outliers in the data.
* The correlation matrix reveals a strong positive relationship between population and households, but the data also suggests that there are some outliers in the population and households data.

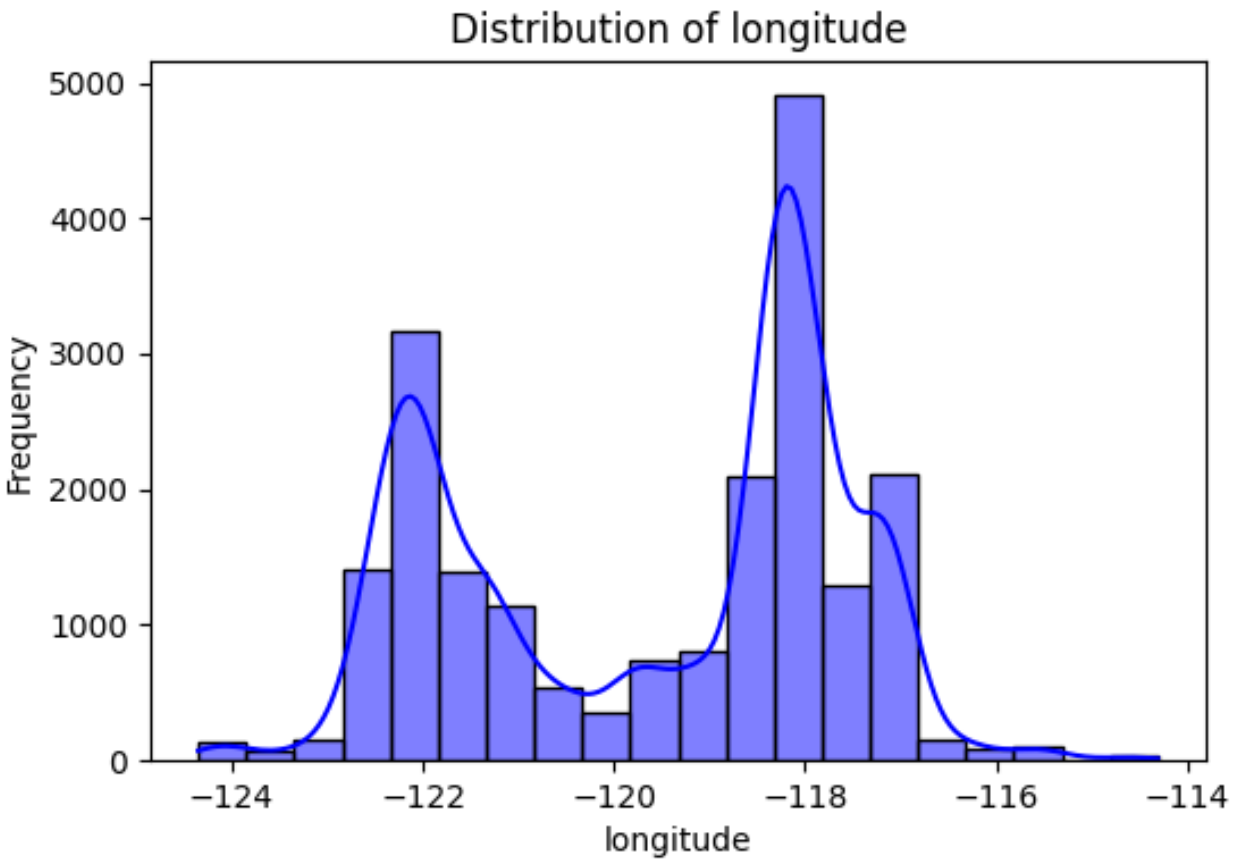
Data-Driven Actions

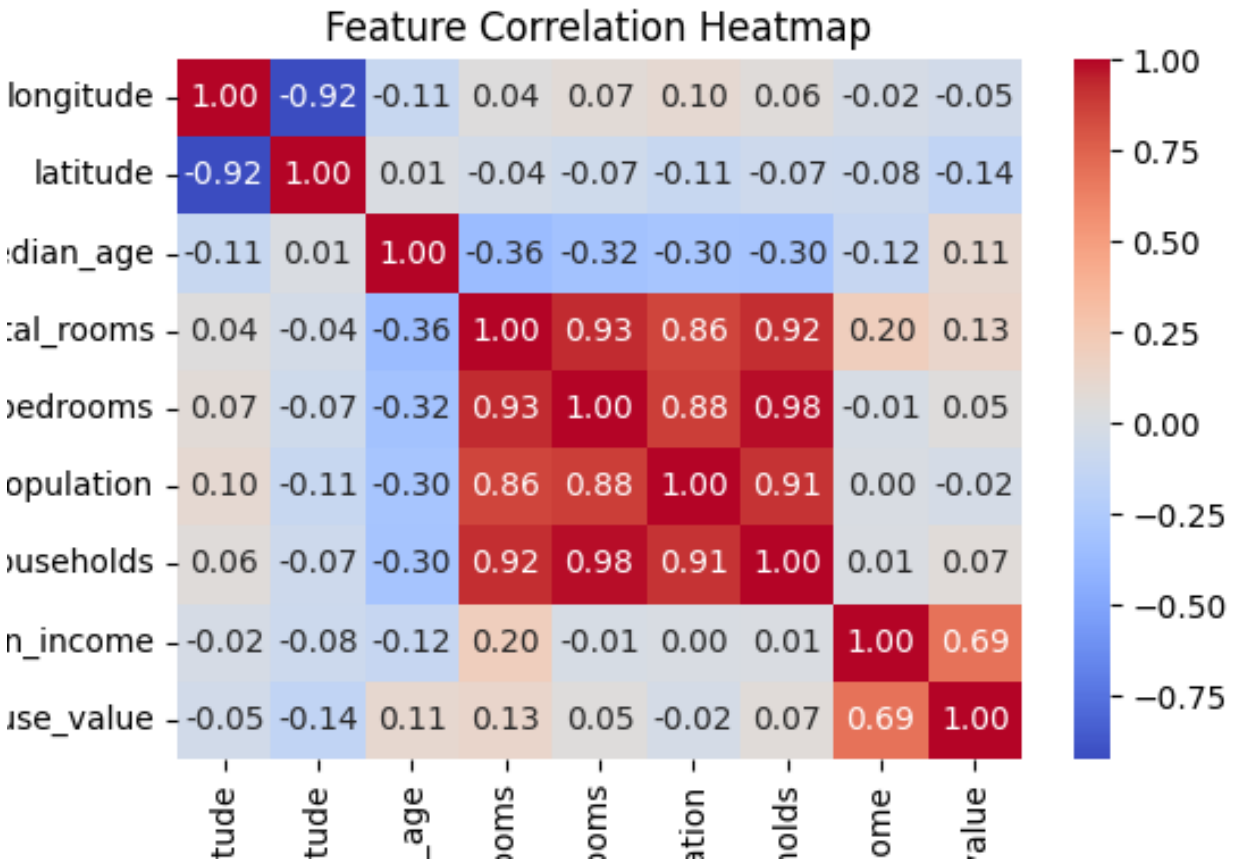
Based on the insights gained from the data, the following data-driven actions can be taken:

* **Targeted Marketing**: The strong positive relationship between median income and median house value suggests that targeted marketing efforts can be directed towards high-income households to increase the demand for high-value homes.
* **Urban Planning**: The strong negative relationship between longitude and latitude suggests that urban planning efforts can be focused on developing areas with high population density and limited land availability.
* **Housing Development**: The moderate negative relationship between housing median age and total rooms suggests that housing development efforts can be focused on building new homes with more rooms to meet the increasing demand for housing.
* **Outlier Analysis**: Further analysis is required to identify the causes of outliers in the data and to determine whether they are errors or legitimate values.

**Conclusion

Statistical Graphs:




Feature Correlation Heatmap
longitude 1.00 -0.92 -0.11 0.04 0.07 0.10 0.06 -0.02 -0.05
latitude -0.92 1.00 0.01 -0.04 -0.07 -0.11 -0.07 -0.08 -0.14
dian_age -0.11 0.01 1.00 -0.36 -0.32 -0.30 -0.30 -0.12 0.11
al_rooms 0.04 -0.04 -0.36 1.00 0.93 0.86 0.92 0.20 0.13
edrooms 0.07 -0.07 -0.32 0.93 1.00 0.88 0.98 -0.01 0.05
opulation 0.10 -0.11 -0.30 0.86 0.88 1.00 0.91 0.00 -0.02
ouseholds 0.06 -0.07 -0.30 0.92 0.98 0.91 1.00 0.01 0.07
n_income -0.02 -0.08 -0.12 0.20 -0.01 0.00 0.01 1.00 0.69
use_value -0.05 -0.14 0.11 0.13 0.05 -0.02 0.07 0.69 1.00
tude
tude
_age
oms
oms
ation
holds
ome
alue
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75