



UNIVERSIDADE FEDERAL DE MATO GROSSO
CAMPUS UNIVERSITÁRIO DO ARAGUAIA
Instituto de Ciências Exatas e da Terra
Curso de Bacharelado em Ciência da Computação

Trabalho Final 1 - Programação de computadores 2024/2

Tema: Análise exploratória de dados

Integrantes: Braz Amorim Campos, Gregório Tavares Matos

Professor: Dr. Ivairton Monteiro Santos

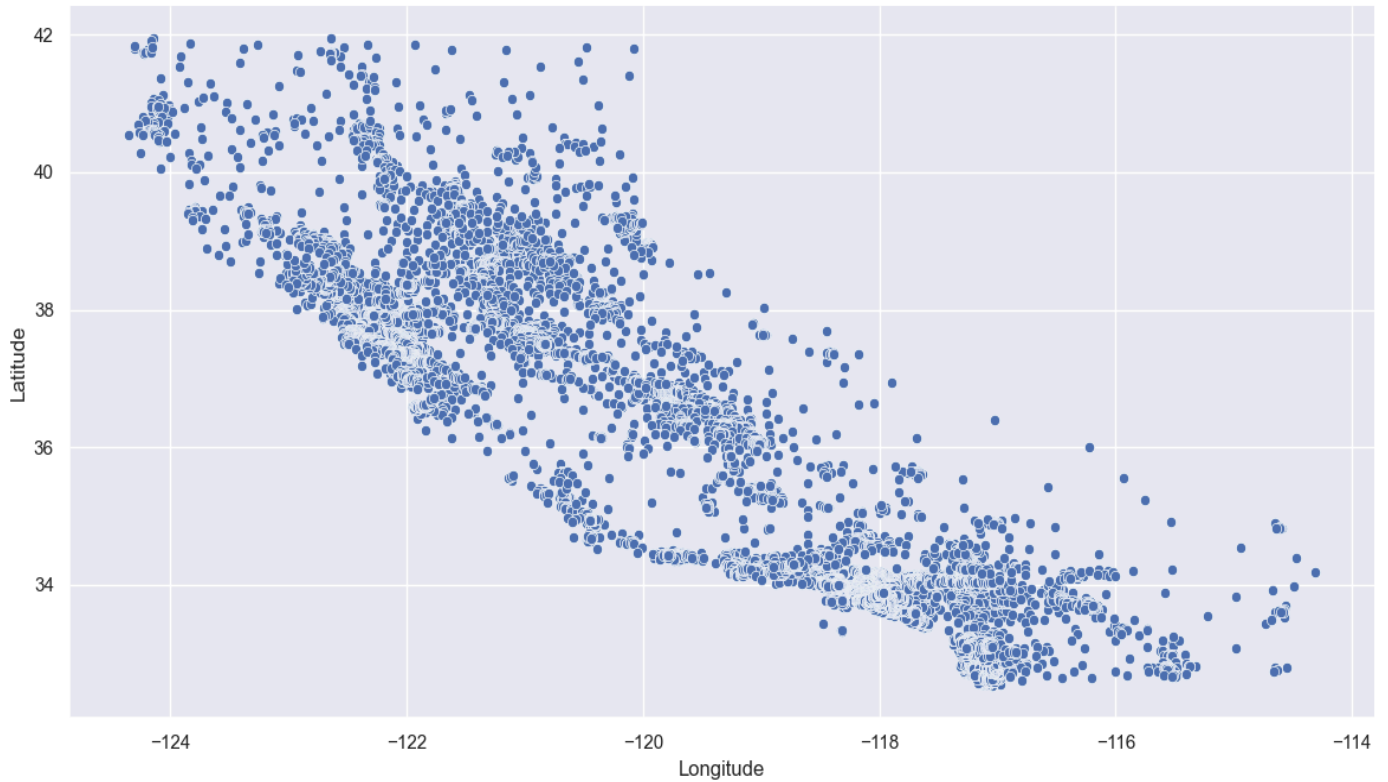
Requisito 1:

Carregue a base de dados, crie um `Pandas.DataFrame` com os dados e execute a função que descreve a base (`describe()`).

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
count	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000
mean	3.870671	28.639486	5.429000	1.096675	1425.476744	3.070655	35.631861	-119.569704
std	1.899822	12.585558	2.474173	0.473911	1132.462122	10.386050	2.135952	2.003532
min	0.499900	1.000000	0.846154	0.333333	3.000000	0.692308	32.540000	-124.350000
25%	2.563400	18.000000	4.440716	1.006079	787.000000	2.429741	33.930000	-121.800000
50%	3.534800	29.000000	5.229129	1.048780	1166.000000	2.818116	34.260000	-118.490000
75%	4.743250	37.000000	6.052381	1.099526	1725.000000	3.282261	37.710000	-118.010000
max	15.000100	52.000000	141.909091	34.066667	35682.000000	1243.333333	41.950000	-114.310000

Requisito 2:

Utilizando a biblioteca SeaBorn e dos campos 'Latitude' (eixo y) e 'Longitude' (eixo x) da base de dados, gere um gráfico que apresenta pontos representando os imóveis em sua localização espacial.



Calcule as métricas:

- Média**

- Para as 6 primeiras variáveis da base de dados (medinc, houseage, averooms, avebedrms, population, aveccup) e apresente os resultados de maneira adequada.

Para as 6 primeiras variáveis da base de dados (medinc, houseage, averooms, avebedrms, population, aveccup) e apresente os resultados de maneira adequada.

```

Modes:
      MedInc      HouseAge      AveRooms      AveBedrms      Population      AveOccup
0      3.1250      52.0      5.0      1.0      891.0      3.0
1     15.0001      NaN      NaN      NaN      NaN      NaN
-----
Quantile:
      1o quantile      2o quantile      3o quantile
MedInc      2.563400      3.534800      4.743250
HouseAge     18.000000     29.000000     37.000000
AveRooms      4.440716      5.229129      6.052381
AveBedrms      1.006079      1.048780      1.099526
Population    787.000000    1166.000000    1725.000000
AveOccup      2.429741      2.818116      3.282261
-----
Interquartile range = 0.8525197767200461
-----
Other info:
      MedInc      HouseAge      AveRooms      AveBedrms      Population      AveOccup
Mean      3.870671     28.639486     5.429000     1.096675     1.425477e+03     3.070655
Median     3.534800     29.000000     5.229129     1.048780     1.166000e+03     2.818116
Variance    3.609323    158.396260     6.121533     0.224592     1.282470e+06    107.870026
Standard deviation 1.899822    12.585558     2.474173     0.473911     1.132462e+03    10.386050

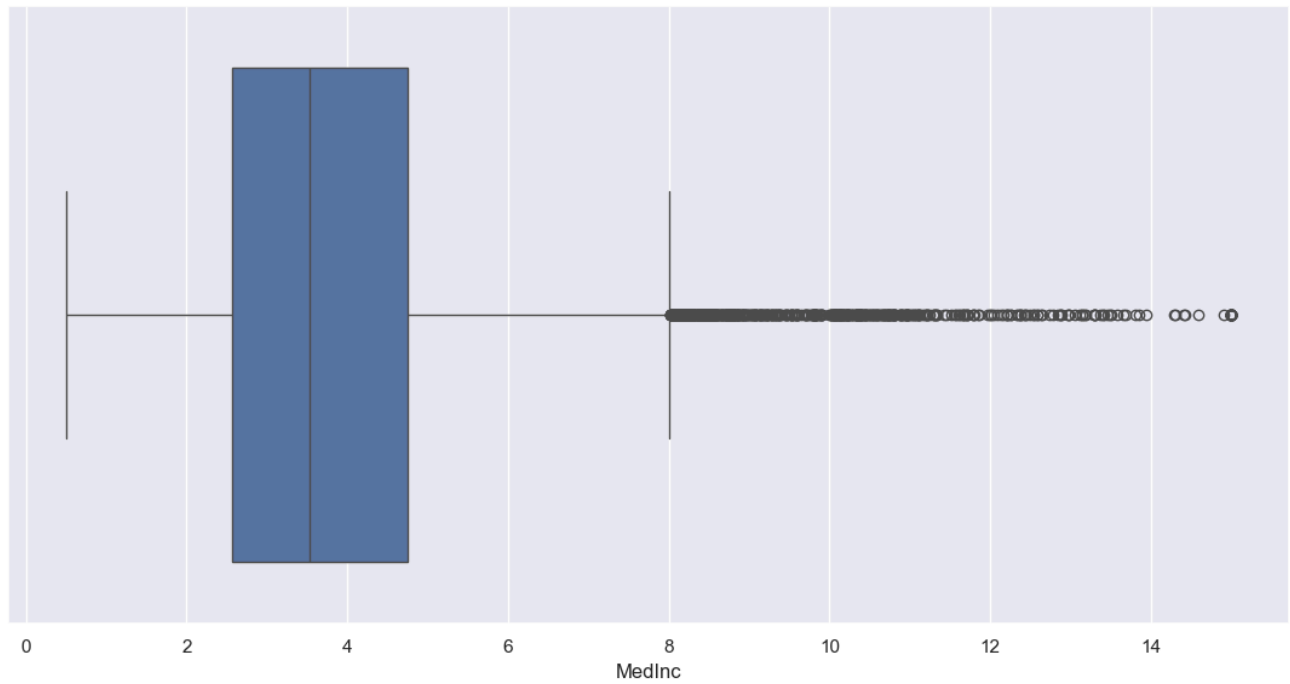
```

Requisito 4:

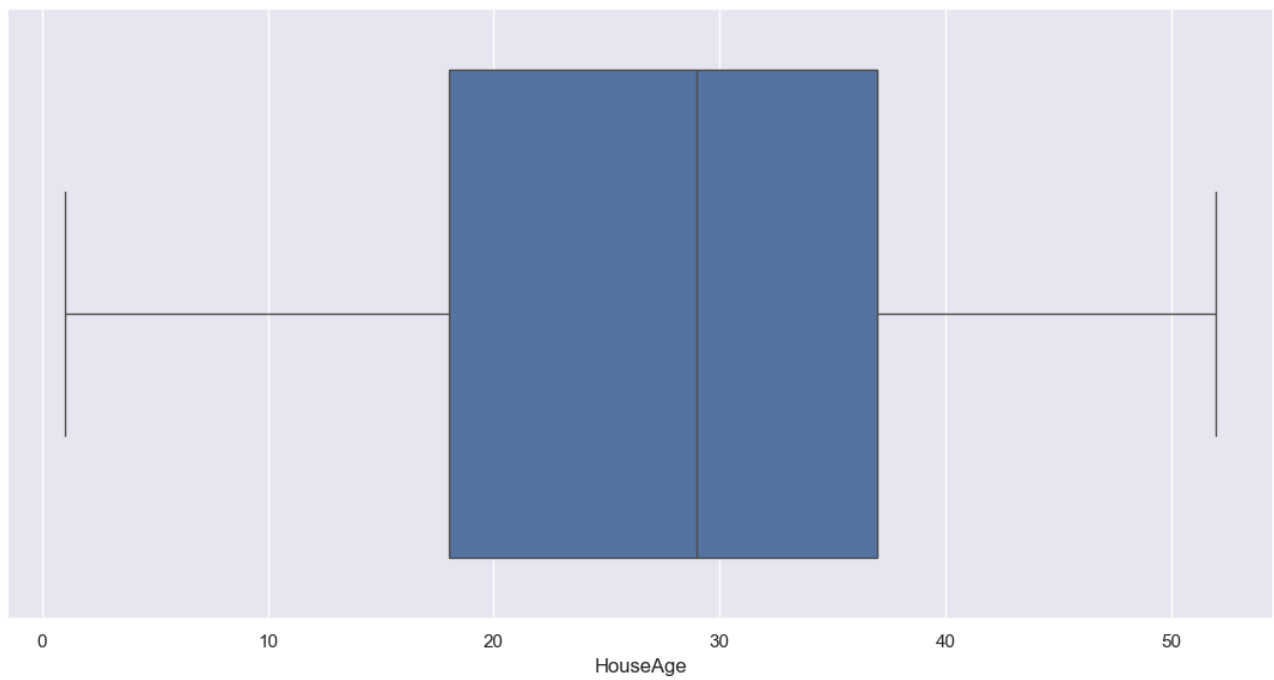
Gere o BoxPlot e o Histograma para as mesmas variáveis descritas no Requisito 3.

BoxPlot's

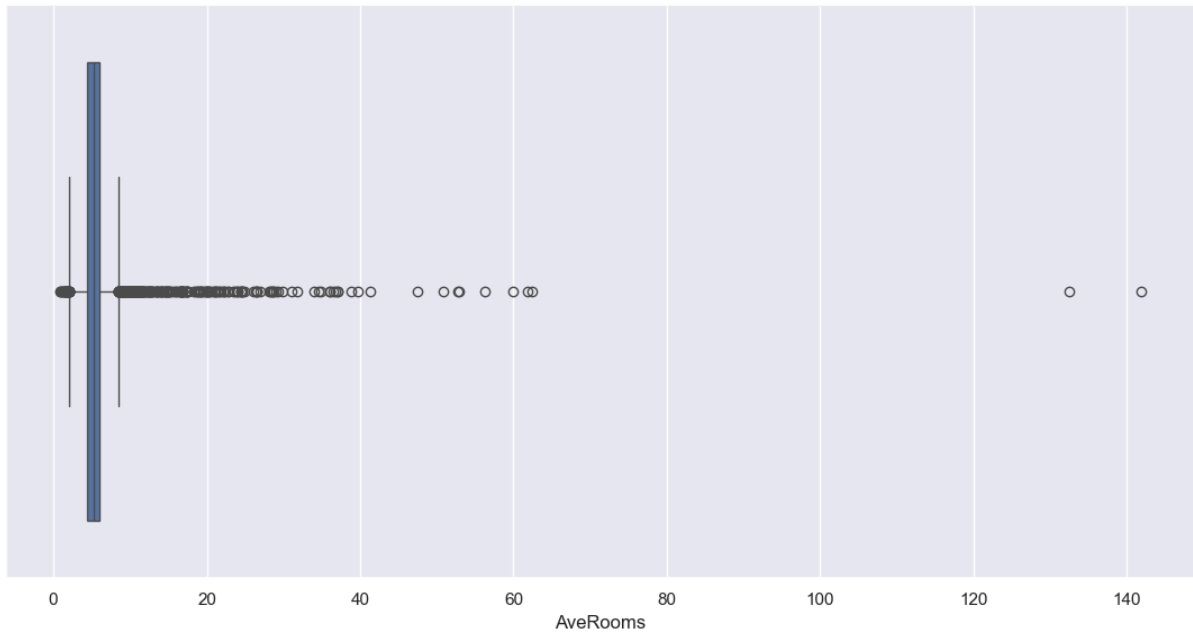
-> MedInc



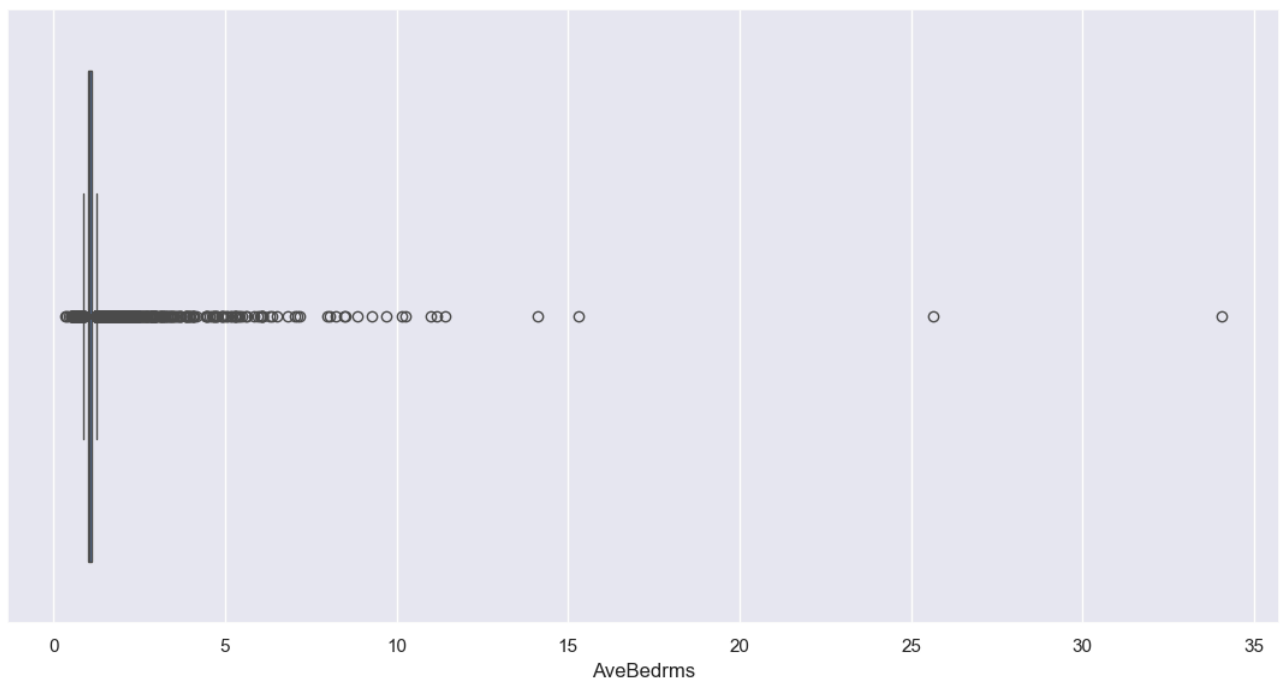
-> HouseAge



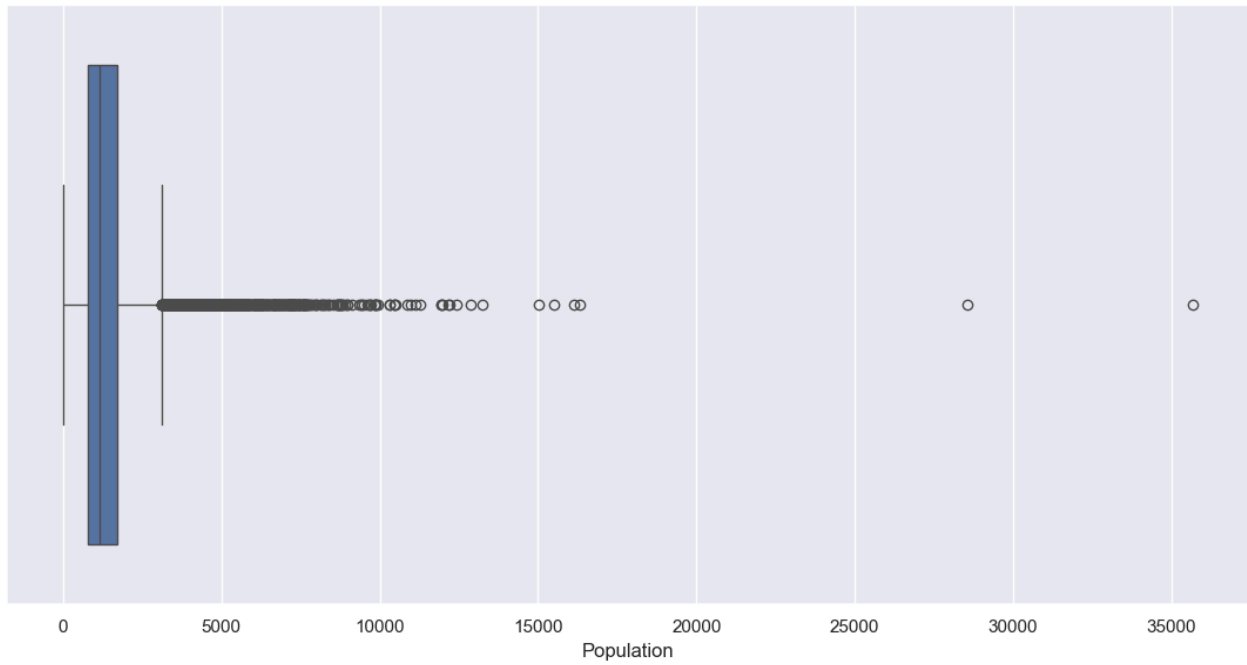
-> AveRooms



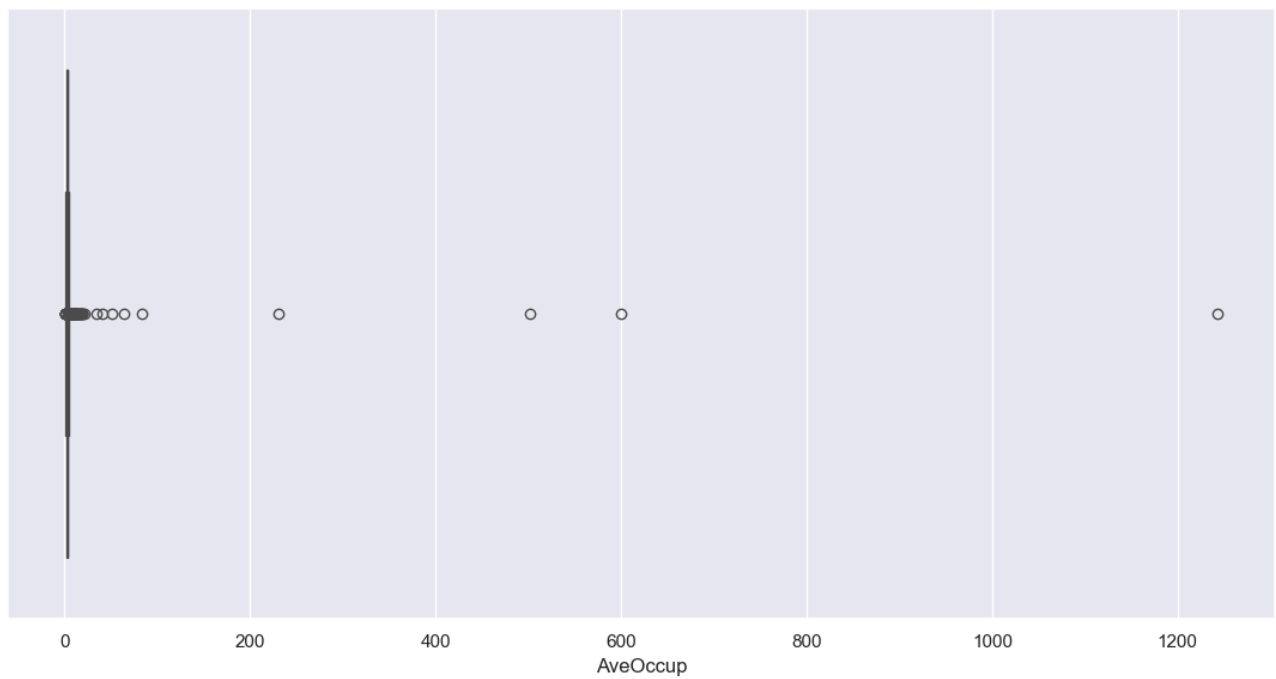
-> AveBedrms



-> Population

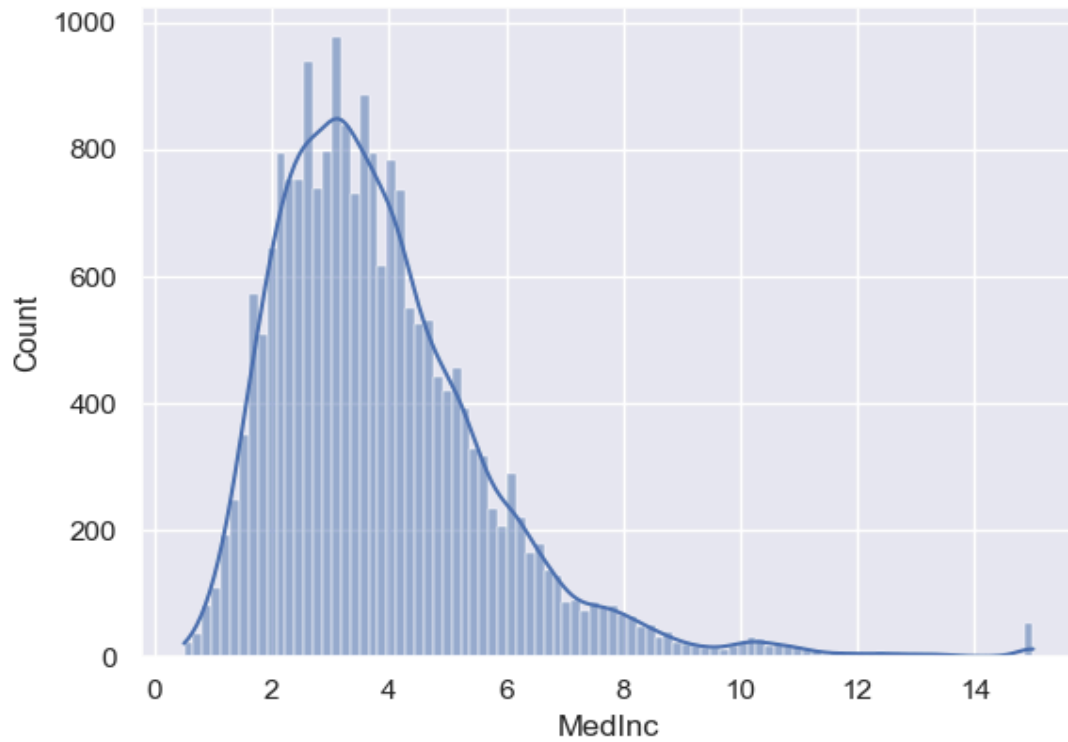


-> AveOccup

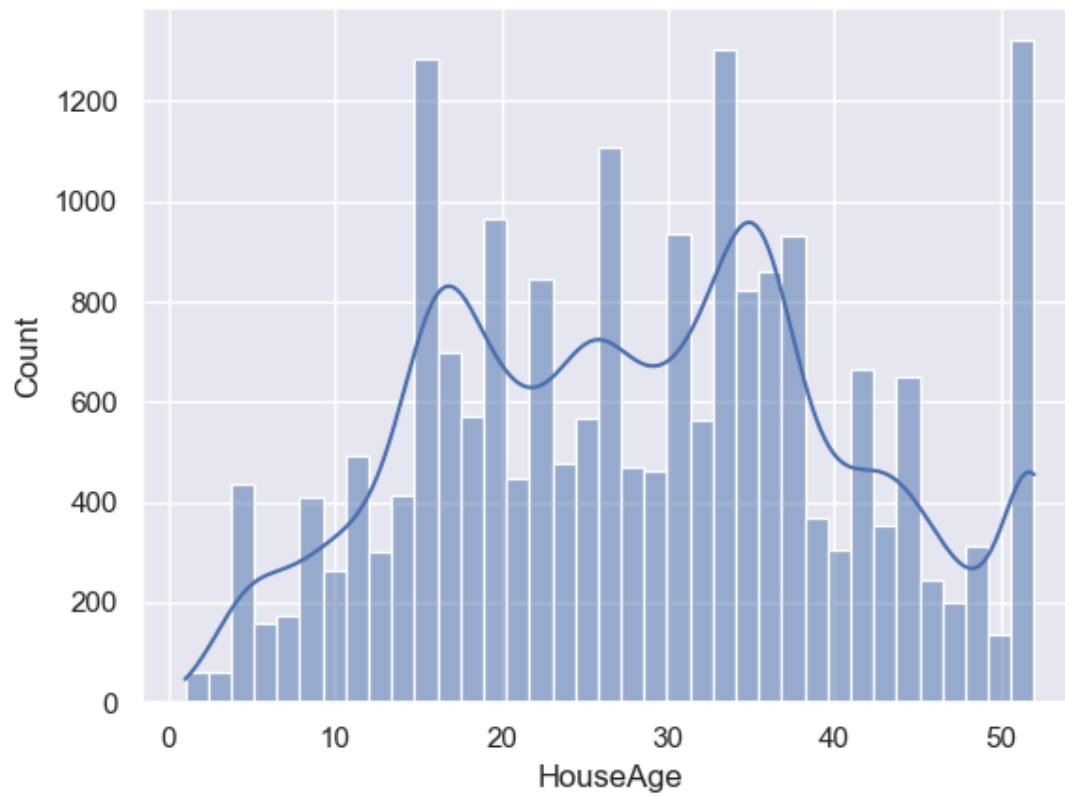


Histogramas

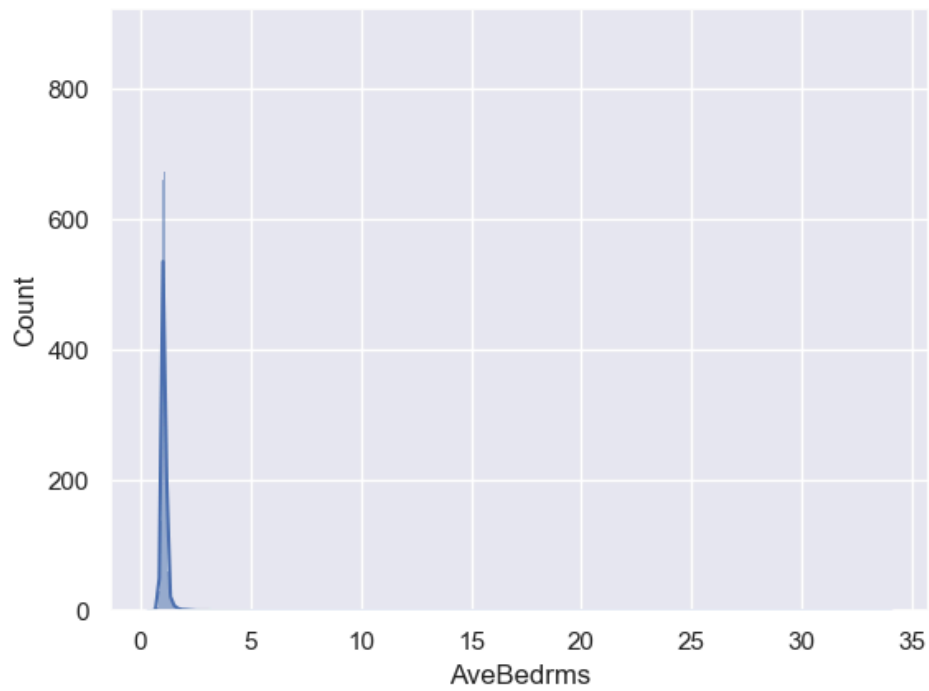
-> MedInc



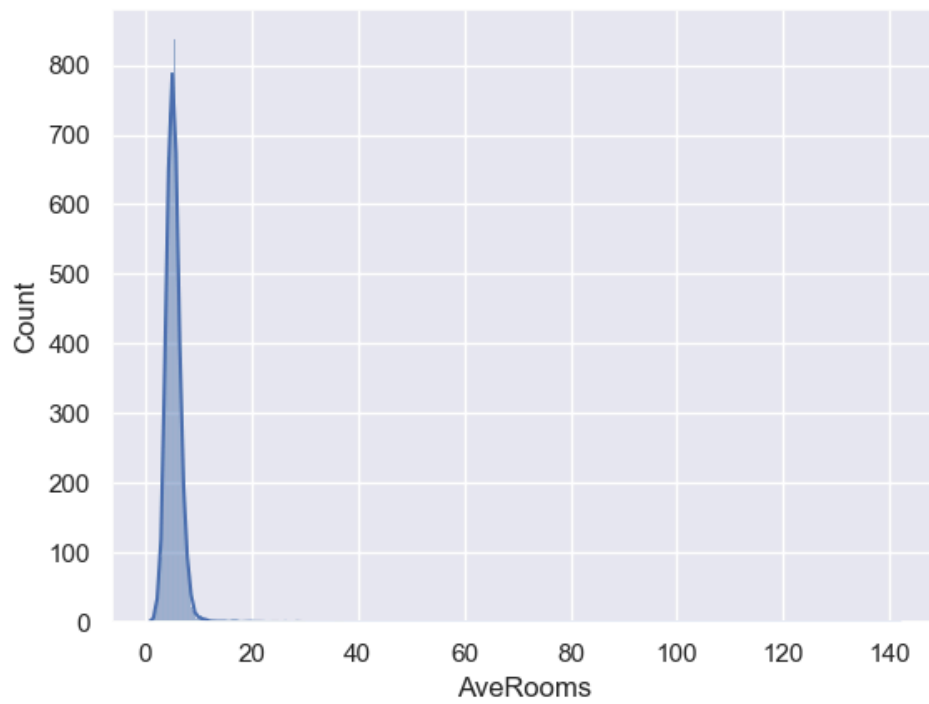
-> HouseAge



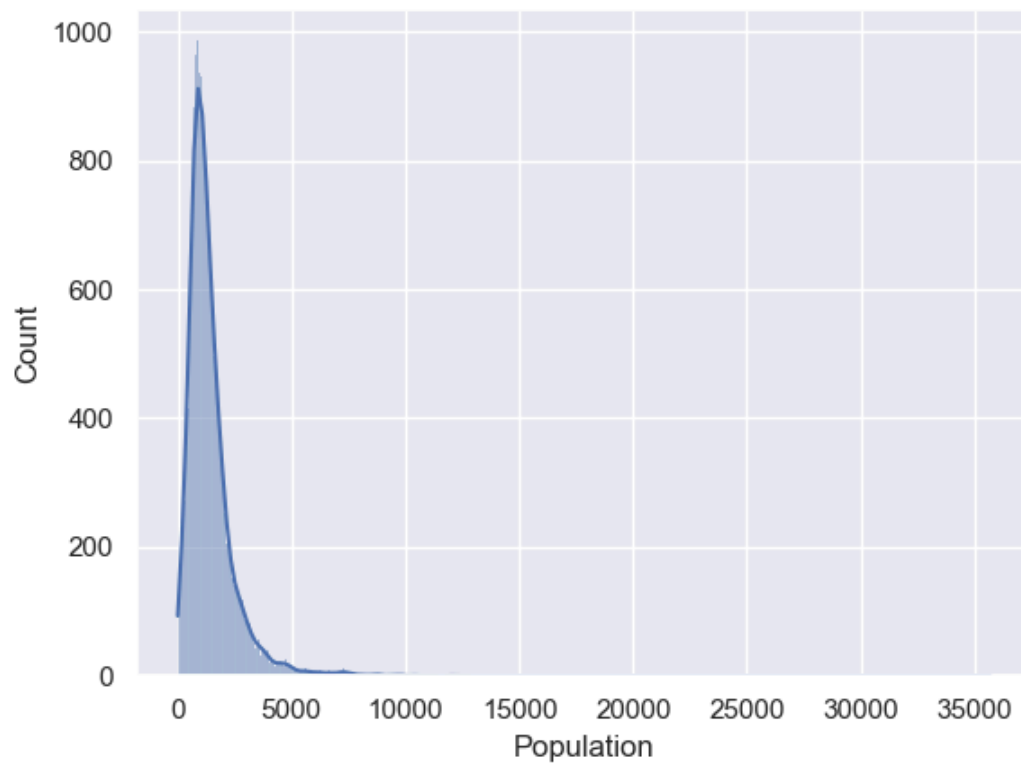
-> AveBedrms



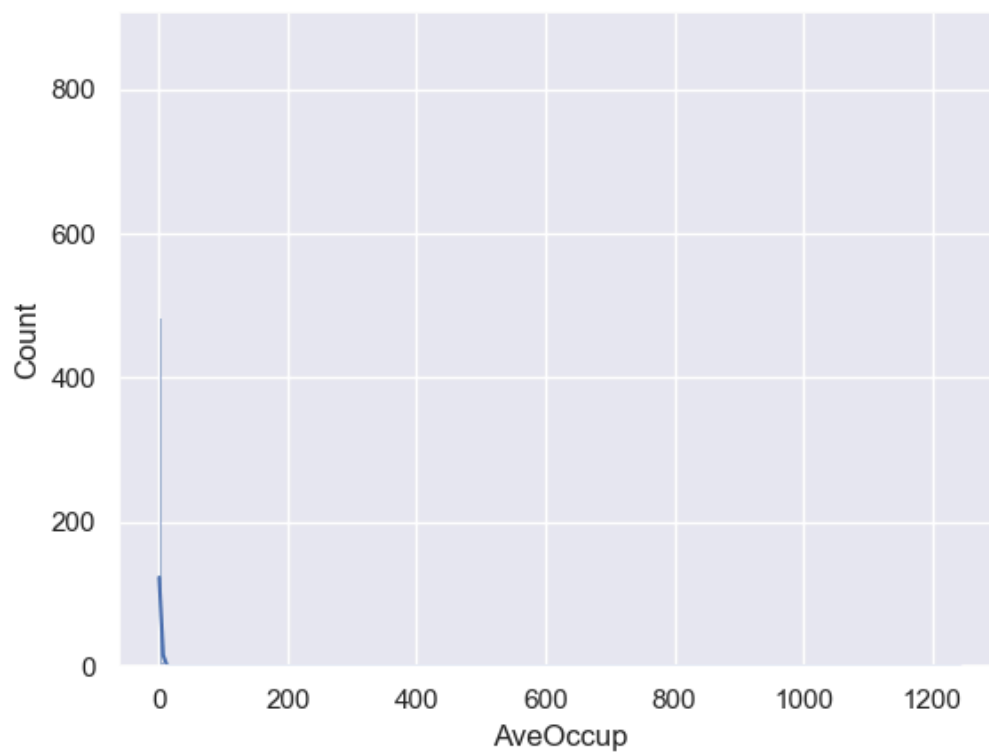
-> AveRooms



-> Population



-> AveOccup



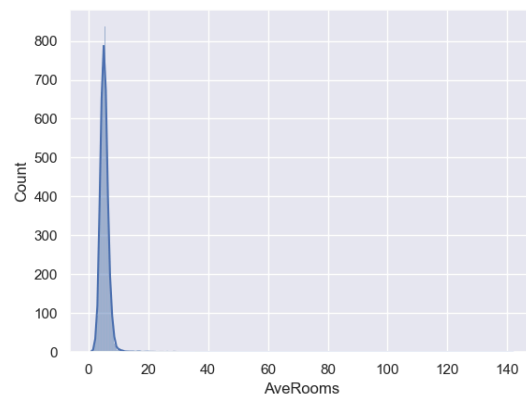
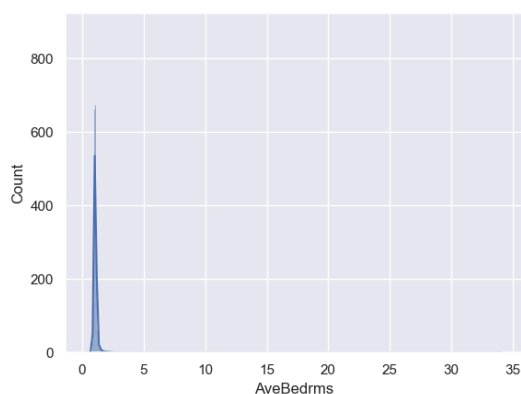
Requisito 5:

Identifique ao menos 2 pares de variáveis (dois conjuntos com 2 variáveis cada) que apresentam correlação entre si. Ou seja, identifique variáveis que seus valores estão correlacionados de uma tal maneira que juntos determinam uma variação no valor do imóvel.

Por exemplo, será que quanto menor a idade do imóvel e menor o número de membros da (correlação positiva), maior o valor do imóvel? Ou, por exemplo, quanto maior a renda familiar e menor a idade do imóvel (correlação negativa), maior o valor do imóvel?

1° Par:

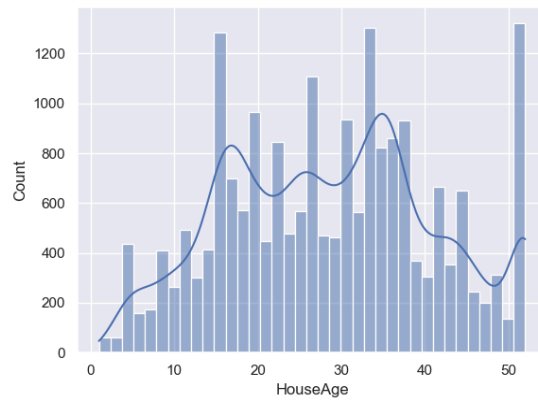
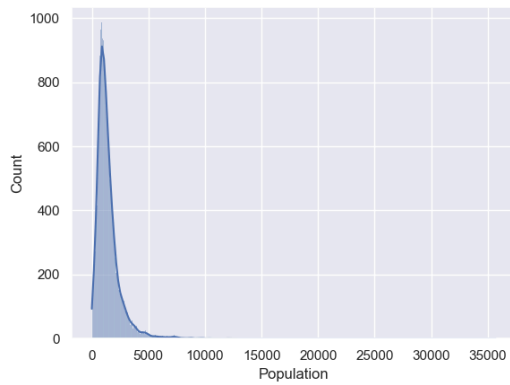
Quanto maior a quantidade de salas da casa (AveRooms) e maior quantidade de quartos (AveBedrms) maior o valor do imóvel.



	AveBedrms	AveRooms
Skewness	31.316956	20.697869
Kurtosis	1636.711972	879.353264

2º Par:

Quanto menor a população da região (Population) e maior idade da casa (HouseAge) menor o valor do imóvel.



	Population	HouseAge
Skewness	4.935858	0.060331
Kurtosis	73.553116	-0.800629