

Sistemas Inteligentes

Projeto de Aprendizado de Máquina:
Aquisição e Visualização

Profa: Deborah Magalhães

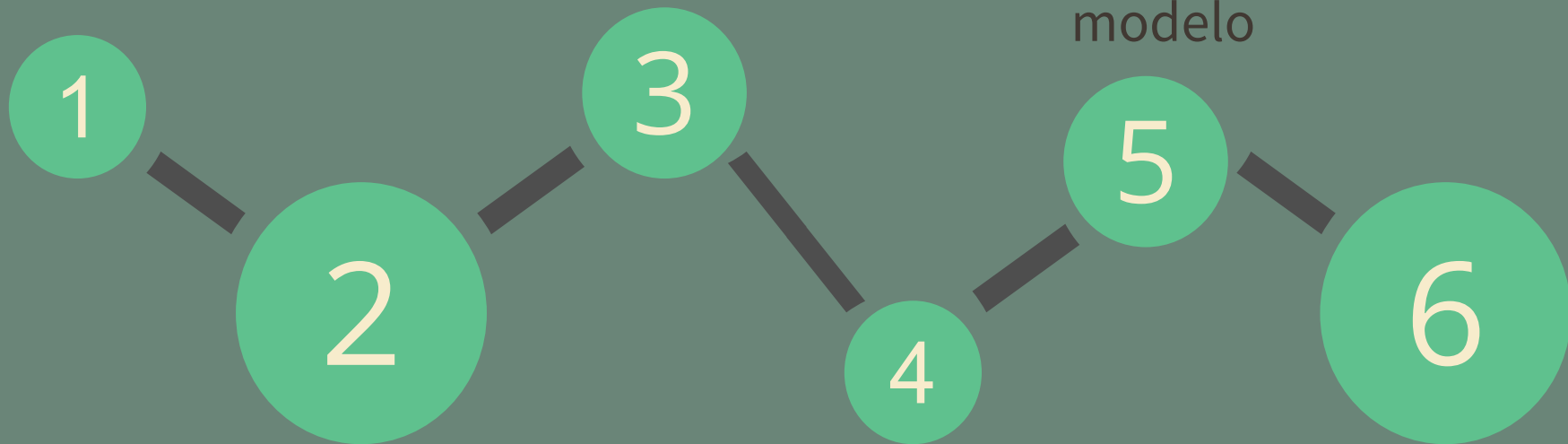


Processo

Aquisição

Pré-processamento

Ajuste do
modelo



Visualização

Seleção e
Treinamento

Resultados

1. Aquisição dos dados

Datasets públicos



[Reddit Datasets](#)



[UCI Machine Learning](#)



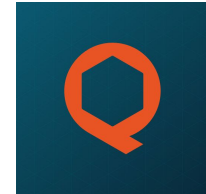
[Kaggle Datasets](#)



[AWS Amazon Datasets](#)



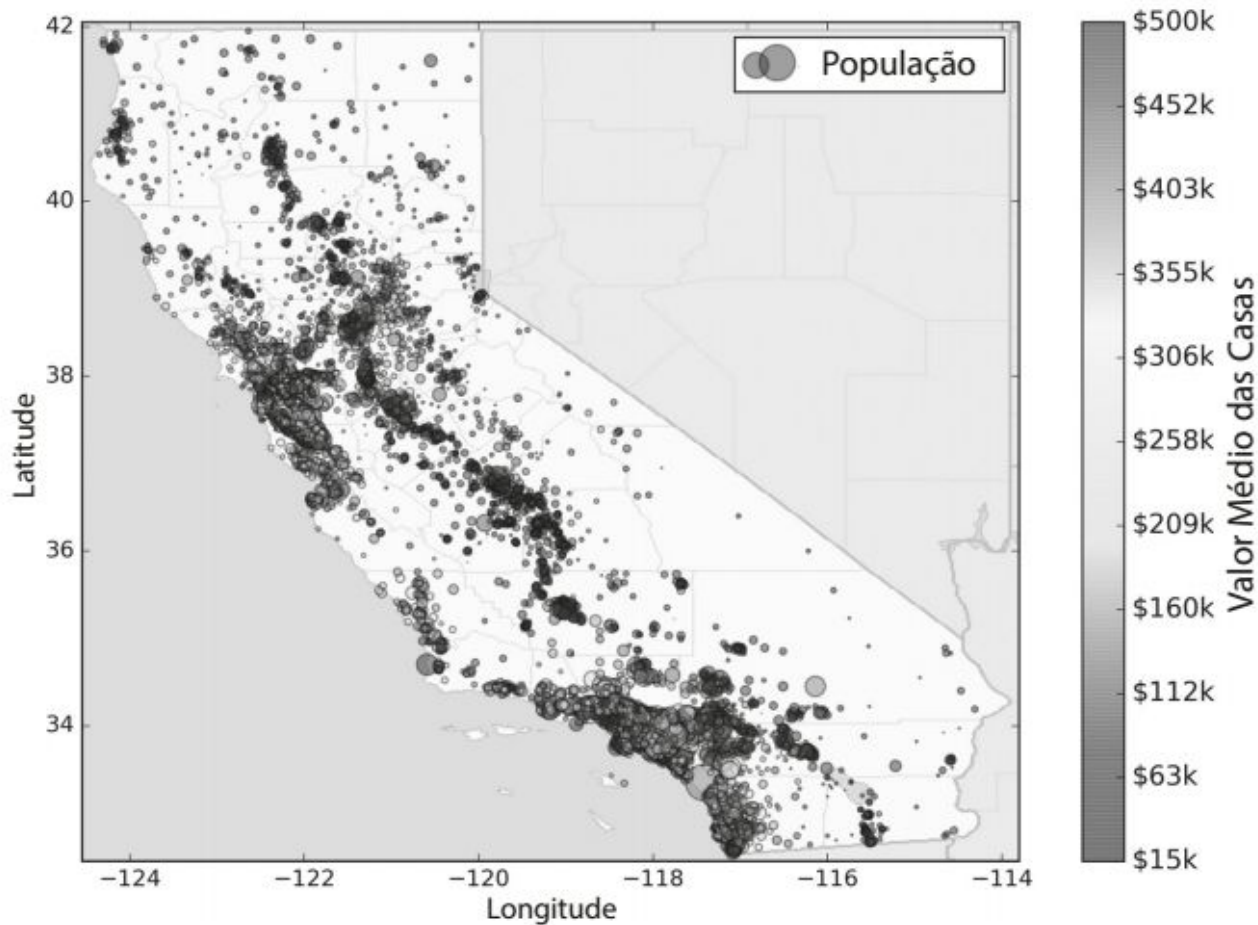
[Open Data Monitor*](#)



[Quandl*](#)

1. Obtendo os dados

Setor imobiliário na Califórnia (1990)



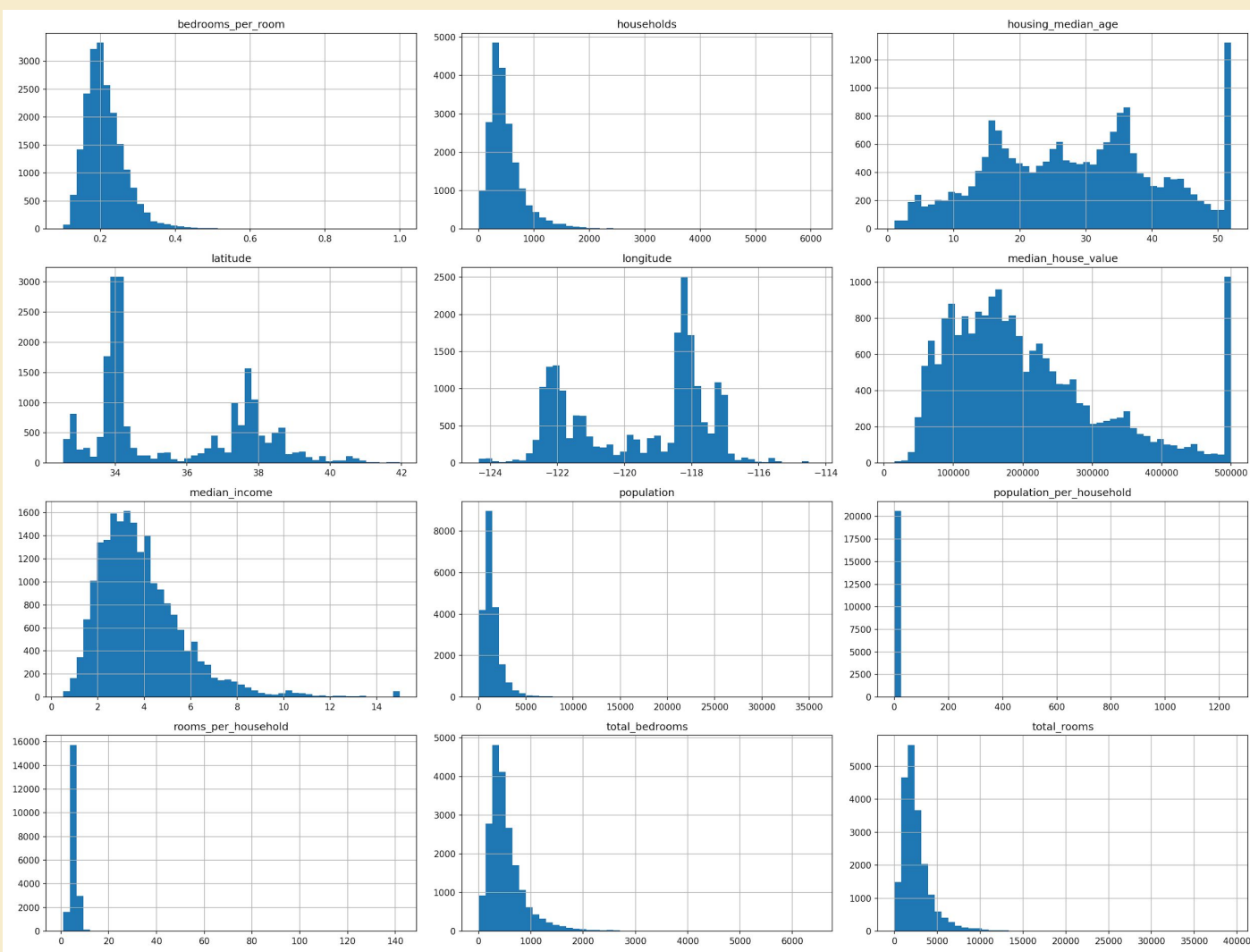
2. Visualizando os dados

Estrutura do dataset

	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

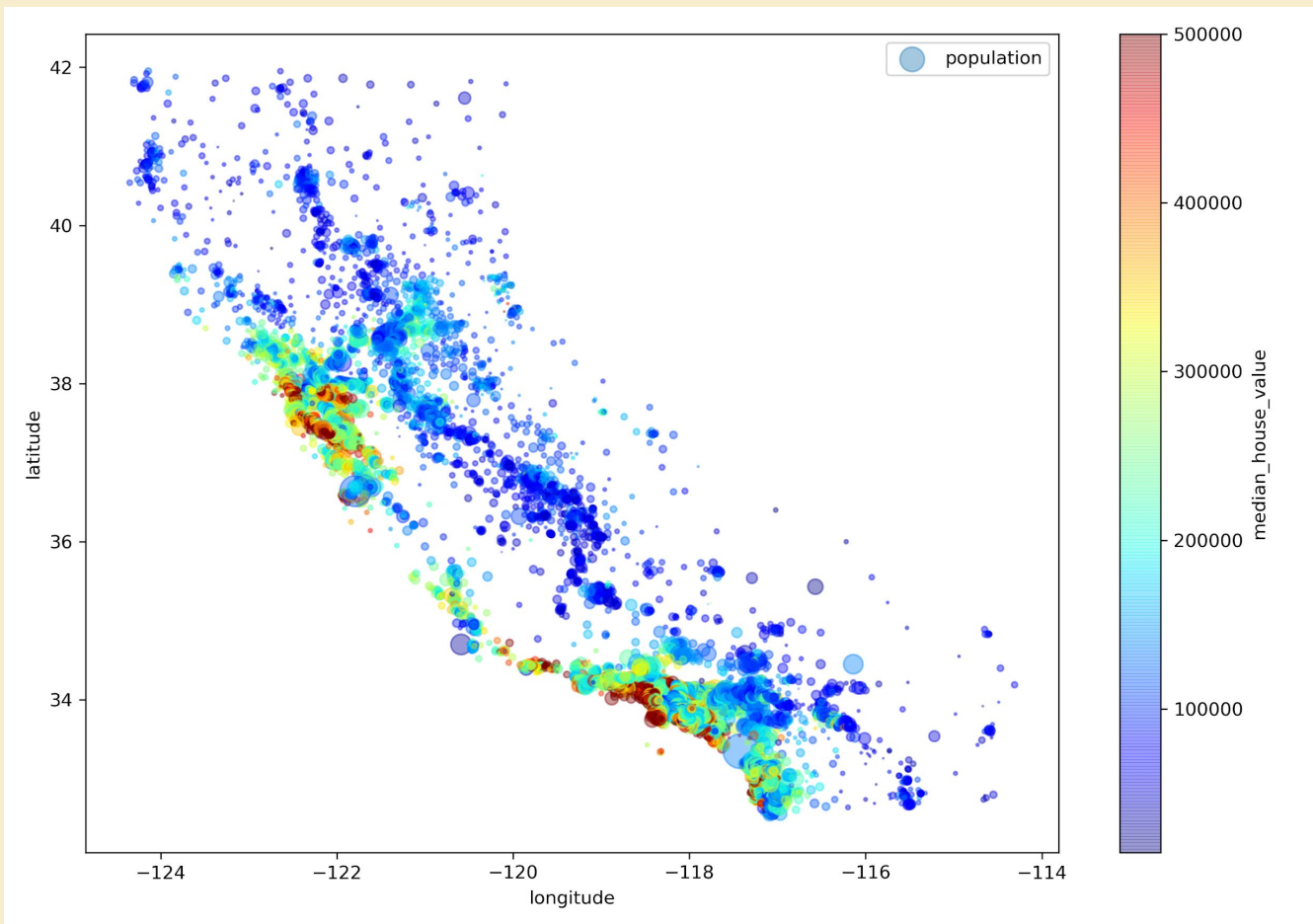
2. Visualizando os dados

Estrutura do dataset



2. Visualizando os dados

Estrutura do dataset



2. Visualizando os dados

Correlação dos dados

	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	0.069608	0.099773	0.055310	-0.015176	-0.045967
latitude	-0.924664	1.000000	0.011173	-0.036100	-0.066983	-0.108785	-0.071035	-0.079809	-0.144160
housing_median_age	-0.108197	0.011173	1.000000	-0.361262	-0.320451	-0.296244	-0.302916	-0.119034	0.105623
total_rooms	0.044568	-0.036100	-0.361262	1.000000	0.930380	0.857126	0.918484	0.198050	0.134153
total_bedrooms	0.069608	-0.066983	-0.320451	0.930380	1.000000	0.877747	0.979728	-0.007723	0.049686
population	0.099773	-0.108785	-0.296244	0.857126	0.877747	1.000000	0.907222	0.004834	-0.024650
households	0.055310	-0.071035	-0.302916	0.918484	0.979728	0.907222	1.000000	0.013033	0.065843
median_income	-0.015176	-0.079809	-0.119034	0.198050	-0.007723	0.004834	0.013033	1.000000	0.688075
median_house_value	-0.045967	-0.144160	0.105623	0.134153	0.049686	-0.024650	0.065843	0.688075	1.000000

Atividade (1.0)

Brasil.IO

dataset: dataset covid19

Data de captura: 2021-03-22

Colabore!

O Brasil.io tem como objetivo facilitar o acesso a dados públicos brasileiros. O projeto é desenvolvido de forma colaborativa, [todo o código está disponível como software livre](#) e os custos são pagos através de [uma campanha de financiamento coletivo](#). Se você acredita nos ideais do projeto ou esses dados estão sendo úteis para você ou sua empresa, considere [fazer uma doação](#) ou [colaborar de outras formas](#).

Baixe os arquivos:

Arquivo	Tamanho	SHA512SUM
boletim.csv.gz	134.05kB	a29ed90ab7ebb27e28f2c089682e31192ed477948aba312341593ef6b410c9988f640a11db89a348eb3366d3dbc79f16133b1b57f8ca3ef75afb2638169a9aad
caso.csv.gz	17.35MB	a2c050dae558b894a4d94c12ec02216f1df707481ed24cbcd8763b4b1b392ca2ec8b0e35e2e9b6abae7a8388166675009cda1a5121839cfff1d94502d82e208
caso_full.csv.gz	39.53MB	75afb8b1349a30f5c2cbde19051892a63760a1036c0c20c36d202905aa6ec6232cc701069131a80a9f7ac69b4494127e557e24de174bc944a75363d200e988aa
obito_cartorio.csv.gz	434.18kB	94f91c9e05516b8f72419a4937d3d4aa27989d64e76b9badabec0b496e6ac8b96cdccf4ab116f79d175eae93ac93fb1436006013f515034feb8600dd8c940e0
SHA512SUMS	586B	80b00f14b7ff1b637e567494cc82cef16066b1d6e4823bd1d58166dcaa7d51fbb500e4252ca213116d28798c4ca8bda39625643eef89e0b801e5f608a72c0871

Atividade (1.0)

Dataset Casos COVID-19

1	date	state	city	place_type	confirmed	deaths	order_for_place	is_last	estimated_population_2019	estimated_population	city_ibge_code	confirmed_per_100k_inhabitant	death_rate
2	2021-03-22	AP		state	92462	1233	364	True	845731	861773	1610729.27557		0.0133
3	2021-03-21	AP		state	92010	1222	363	False	845731	861773	1610676.82557		0.0133
4	2021-03-20	AP		state	91515	1222	362	False	845731	861773	1610619.38585		0.0134
5	2021-03-19	AP		state	91077	1216	361	False	845731	861773	1610568.56040		0.0134
6	2021-03-18	AP		state	90597	1206	360	False	845731	861773	1610512.86128		0.0133
7	2021-03-17	AP		state	89822	1201	359	False	845731	861773	1610422.93040		0.0134
8	2021-03-16	AP		state	89118	1195	358	False	845731	861773	1610341.23835		0.0134
9	2021-03-15	AP		state	88363	1187	357	False	845731	861773	1610253.62828		0.0134
10	2021-03-14	AP		state	88047	1181	356	False	845731	861773	1610216.95969		0.0134
11	2021-03-13	AP		state	87792	1181	355	False	845731	861773	1610187.36953		0.0135
12	2021-03-12	AP		state	87479	1177	354	False	845731	861773	1610151.04906		0.0135
13	2021-03-11	AP		state	87095	1169	353	False	845731	861773	1610106.48976		0.0134
14	2021-03-10	AP		state	86674	1167	352	False	845731	861773	1610057.63699		0.0135
15	2021-03-09	AP		state	86346	1164	351	False	845731	861773	1610019.57592		0.0135
16	2021-03-08	AP		state	85879	1161	350	False	845731	861773	169965.38532		0.0135
17	2021-03-07	AP		state	85578	1156	349	False	845731	861773	169930.45732		0.0135
18	2021-03-06	AP		state	85375	1156	348	False	845731	861773	169906.90124		0.0135
19	2021-03-05	AP		state	85065	1156	347	False	845731	861773	169870.92889		0.0136
20	2021-03-04	AP		state	84808	1152	346	False	845731	861773	169841.10665		0.0136
21	2021-03-03	AP		state	84483	1146	345	False	845731	861773	169803.39370		0.0136



Muito Obrigada!

Se você tiver qualquer dúvida ou sugestão:

- deborah.vm@ufpi.edu.br

