

«به نام خدا»

تکلیف سوم – سوال اول – مرضیه علیدادی – 9631983
(کد های مربوط، در دو فرمت py و ipynb. ضمیمه شده اند.)

1.

(a) قبل از حذف سطر های شامل مقادیر null ، تعداد سطر ها بدین صورت بود:

	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

و این تعداد از ستون ها به تفکیک دارای مقدار null بودند: (فقط یکی از ستون ها به تعداد 207 تا)

```
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
```

پس از حذف سطر های شامل مقادیر null ، تعداد سطر ها بدین صورت شد: (دقیقا 207 تا کمتر از قبل از حذف شد)

	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

20433 rows × 10 columns

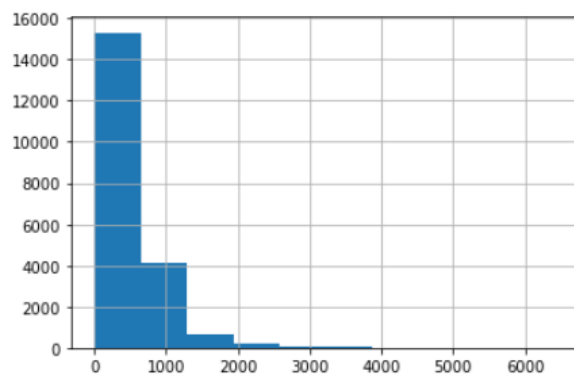
و دیگر هیچ کدام از ستون ها شامل null نیست:

```
longitude      0
latitude        0
housing_median_age  0
total_rooms     0
total_bedrooms  0
population      0
households      0
median_income   0
median_house_value  0
ocean_proximity 0
dtype: int64
```

(b) تعداد هر کدام از مقادیر یکتا در این ستون:

```
ocean_proximity
<1H OCEAN      9034
INLAND         6496
ISLAND           5
NEAR BAY       2270
NEAR OCEAN     2628
Name: ocean_proximity, dtype: int64
```

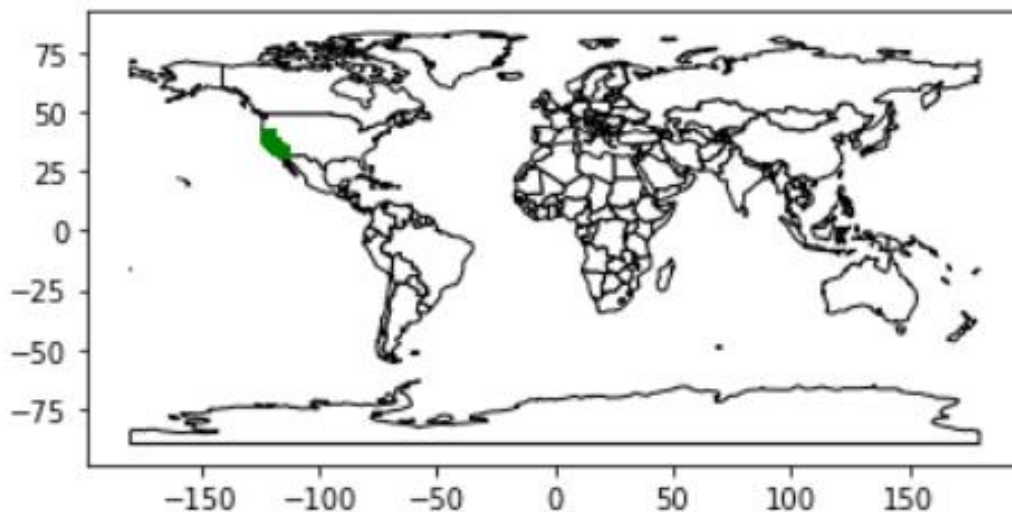
(c) نمودار هیستوگرام مربوط به همه ی ویژگی ها را رسم کردم. برای مثال، برای ویژگی `total_bedrooms` ، نمودار به شکل زیر است؛ که نشان می دهد بیشتر مقادیر در بازه ی 0 تا 1000 قرار دارند.



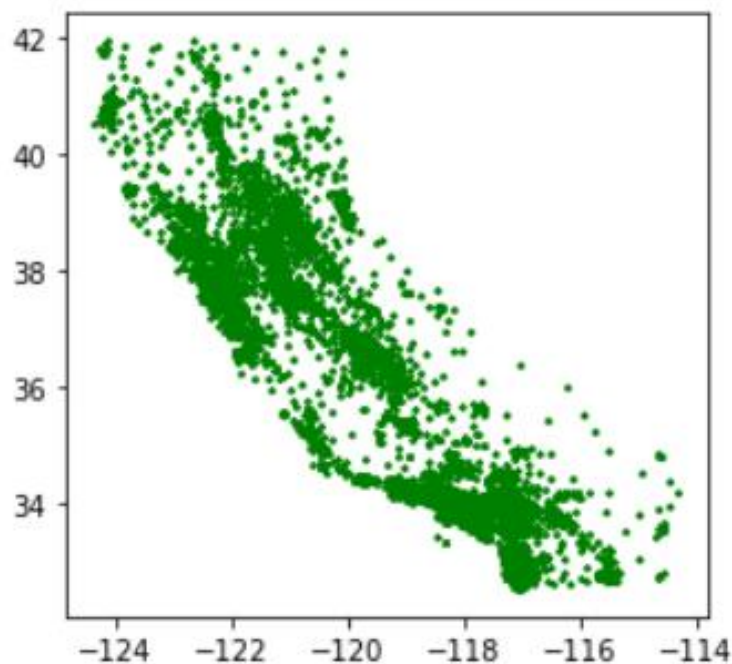
(d) ابتدا پارامتری برای نشان دادن پوینت های مکانی (geometry)، از روی دو پارامتر طول و عرض جغرافیایی ایجاد کردم:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity	geometry
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY	POINT (-122.23000 37.88000)
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY	POINT (-122.22000 37.86000)
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY	POINT (-122.24000 37.85000)
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY	POINT (-122.25000 37.85000)
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY	POINT (-122.25000 37.85000)

سپس این مکان ها را بدین صورت روی نقشه نشان دادم (با رنگ سبز):



سپس برای دیدن تراکم مکان ها، بدین صورت نشان دادم:



تمامی خانه های دیتاست، در این بازه ی جغرافیایی قرار گرفته اند. که نقاط سبز، هر کدام یک خانه را نشان می دهد. که با توجه به تراکم آن ها در نقشه و با توجه به طول و عرضی که رو نقشه مشخص شده است، می توان دید که در هر منطقه ی جغرافیایی، تراکم خانه ها به چه شکل است. برای مثال، وقتی همزمان طول جغرافیایی بین 117- تا 119- باشد و عرض جغرافیایی بین 33.5 تا 34.5 باشد، تراکم خانه ها در این منطقه خیلی بالاست.