



Construisons ensemble votre
rêve de demain avec passion
et savoir-faire.



SOMMAIRE

1 Analyse

2 Correlation

3 Preprocessing

4 Test Features

5 Test Modele

6 Pipeline

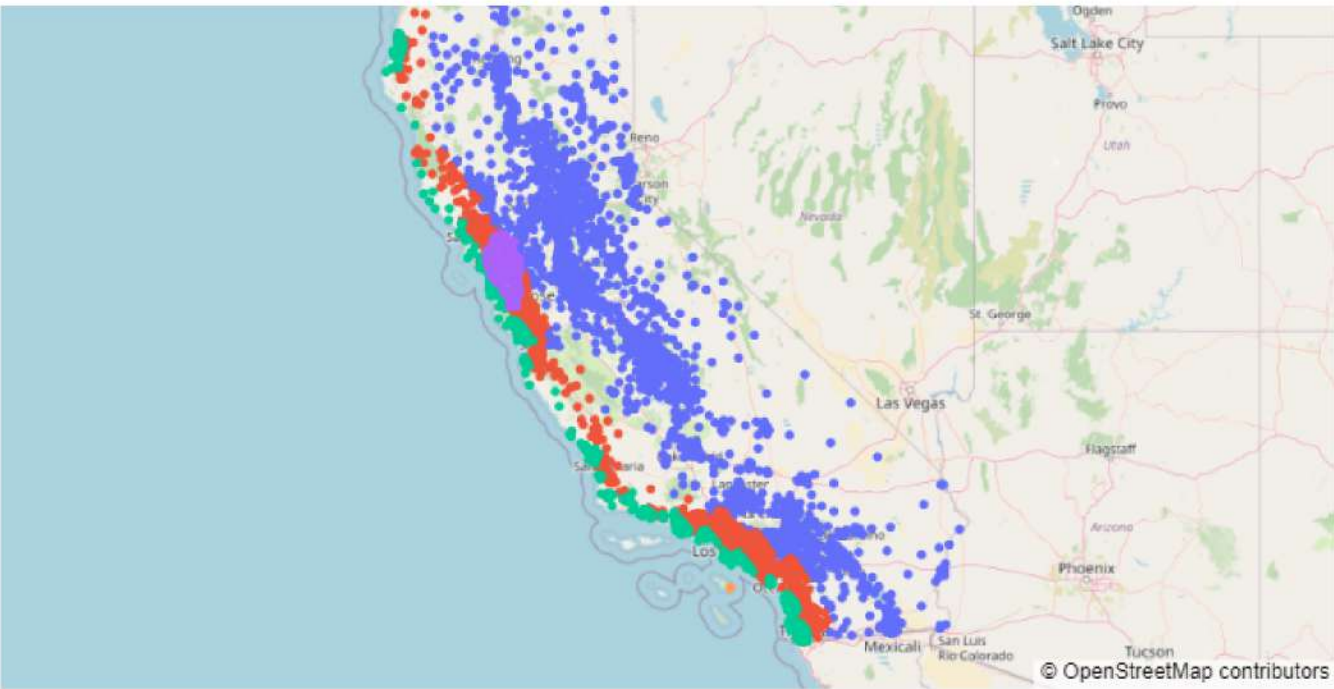
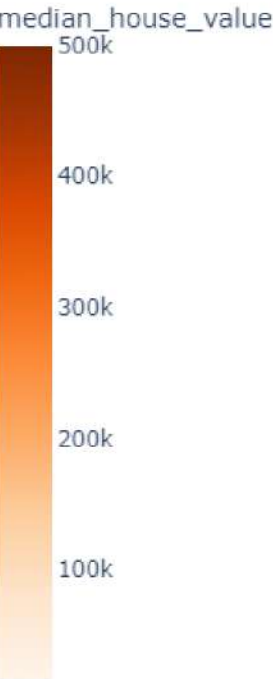
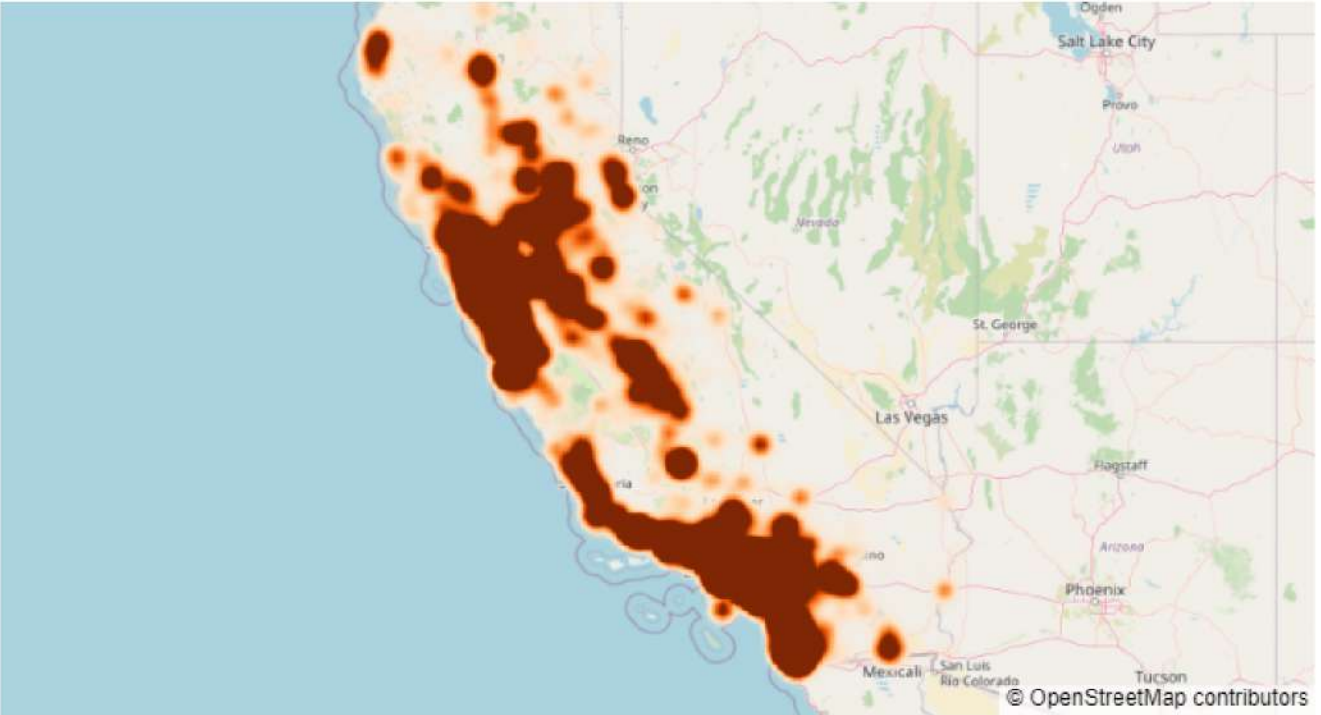
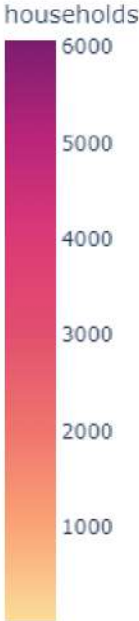
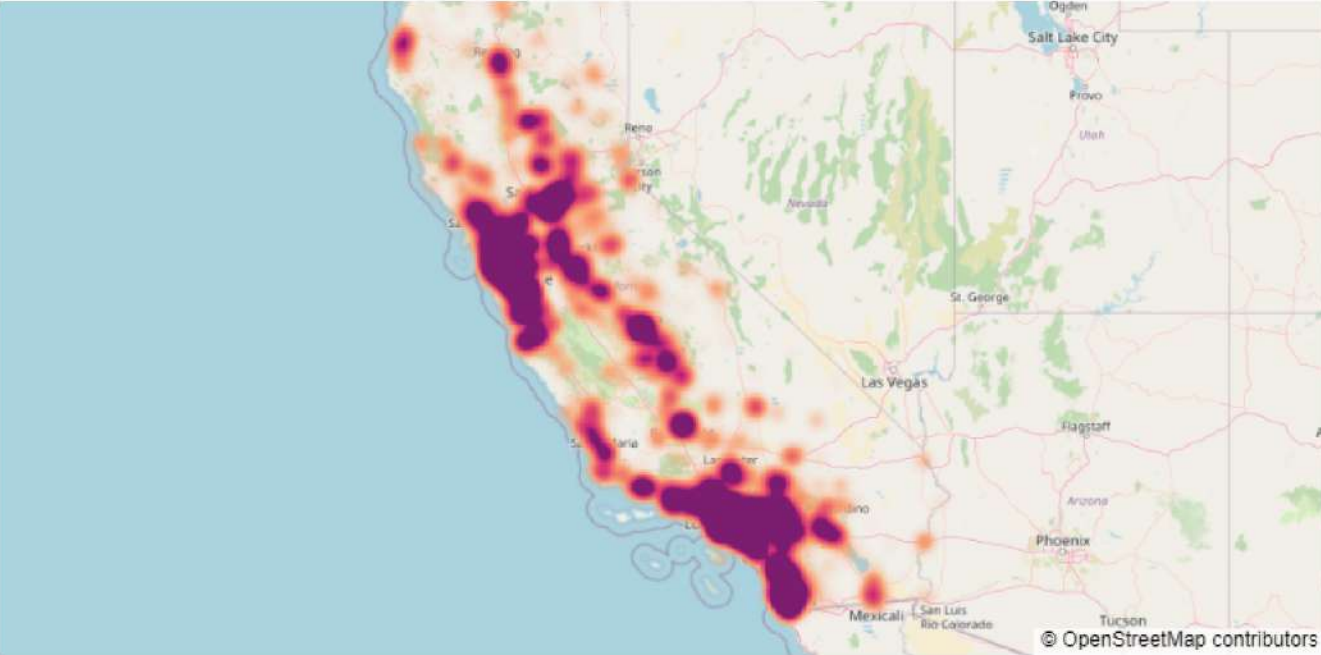
7 Streamlit



1

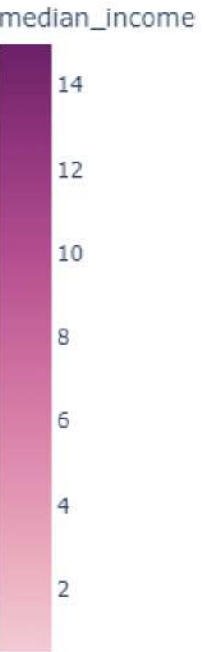
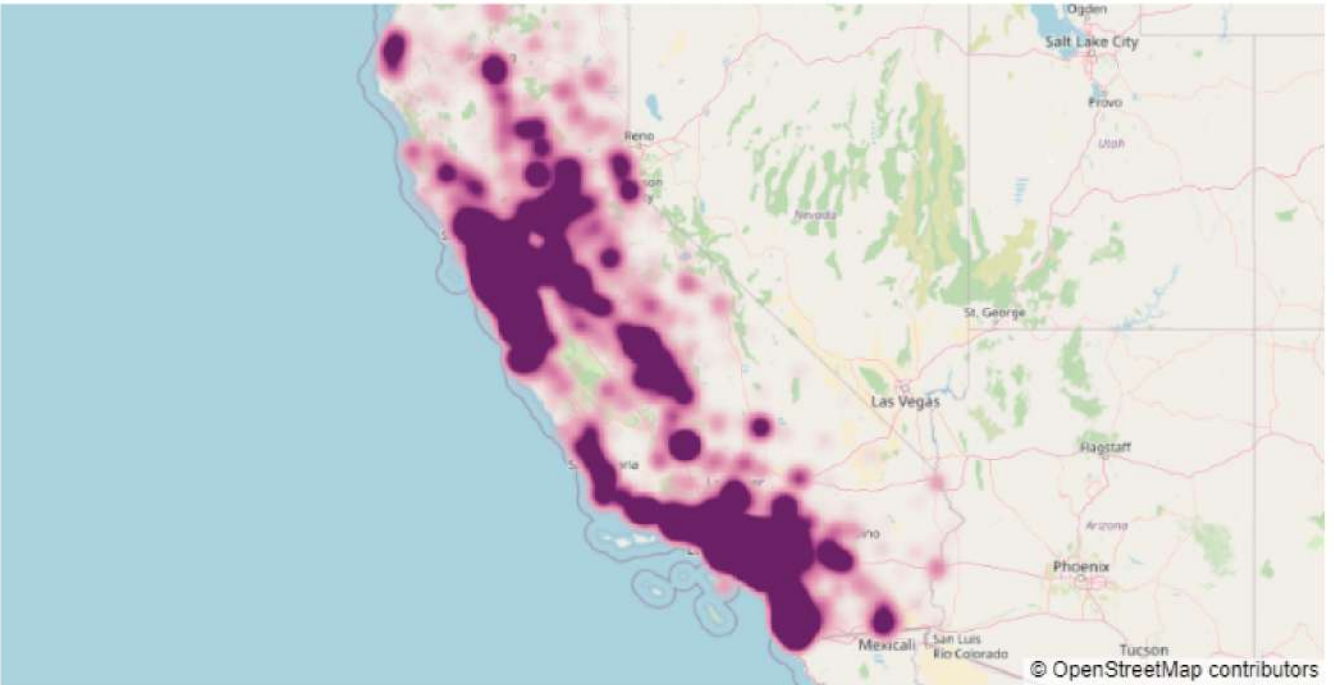
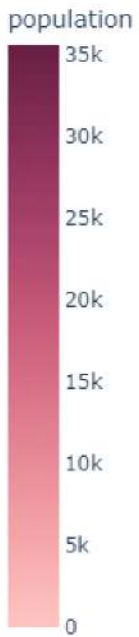
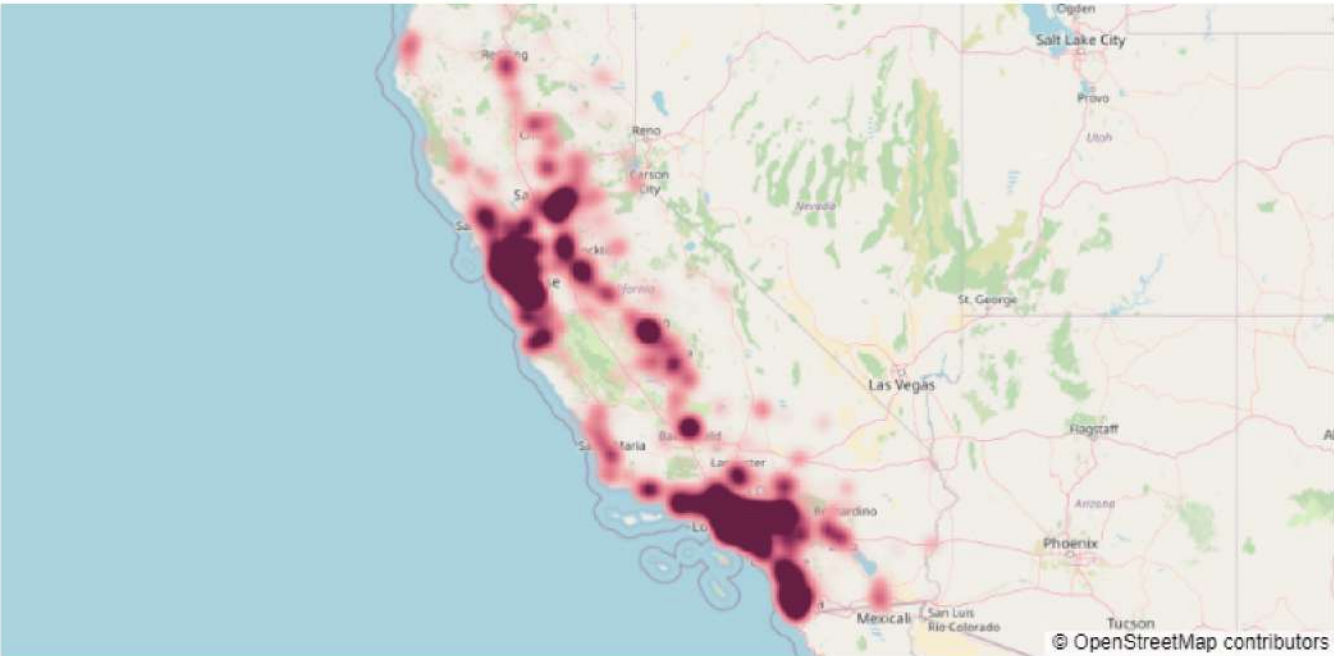
Analyse

Présentation



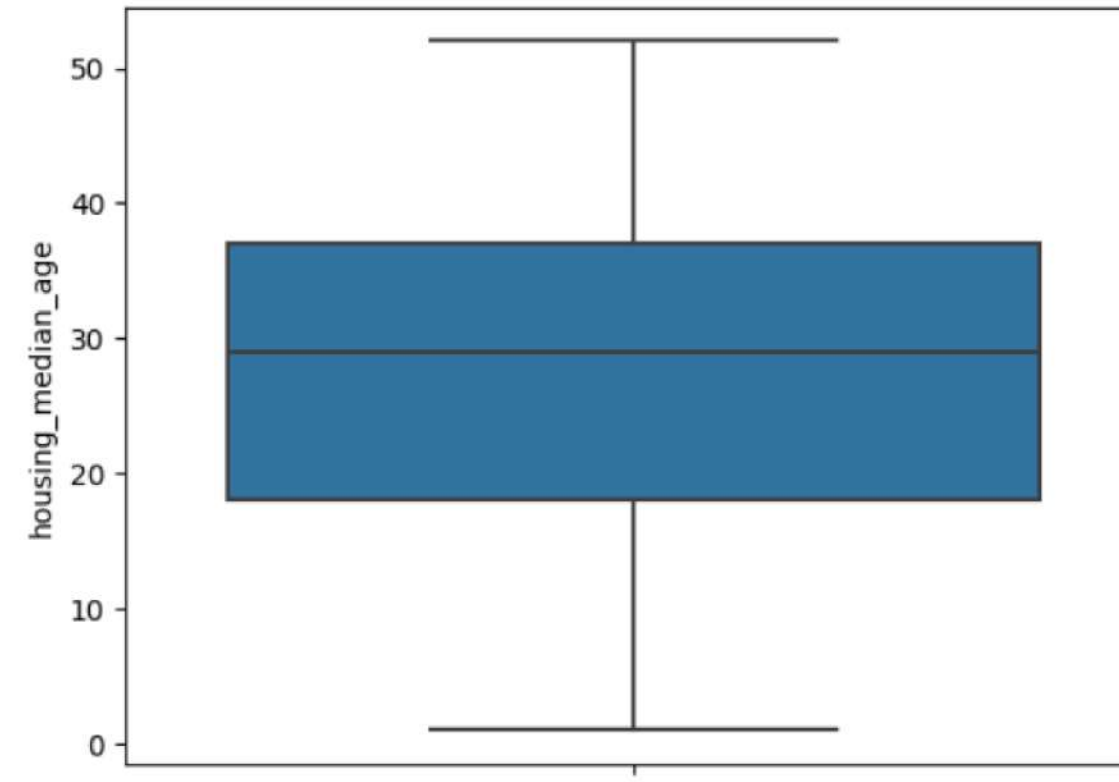
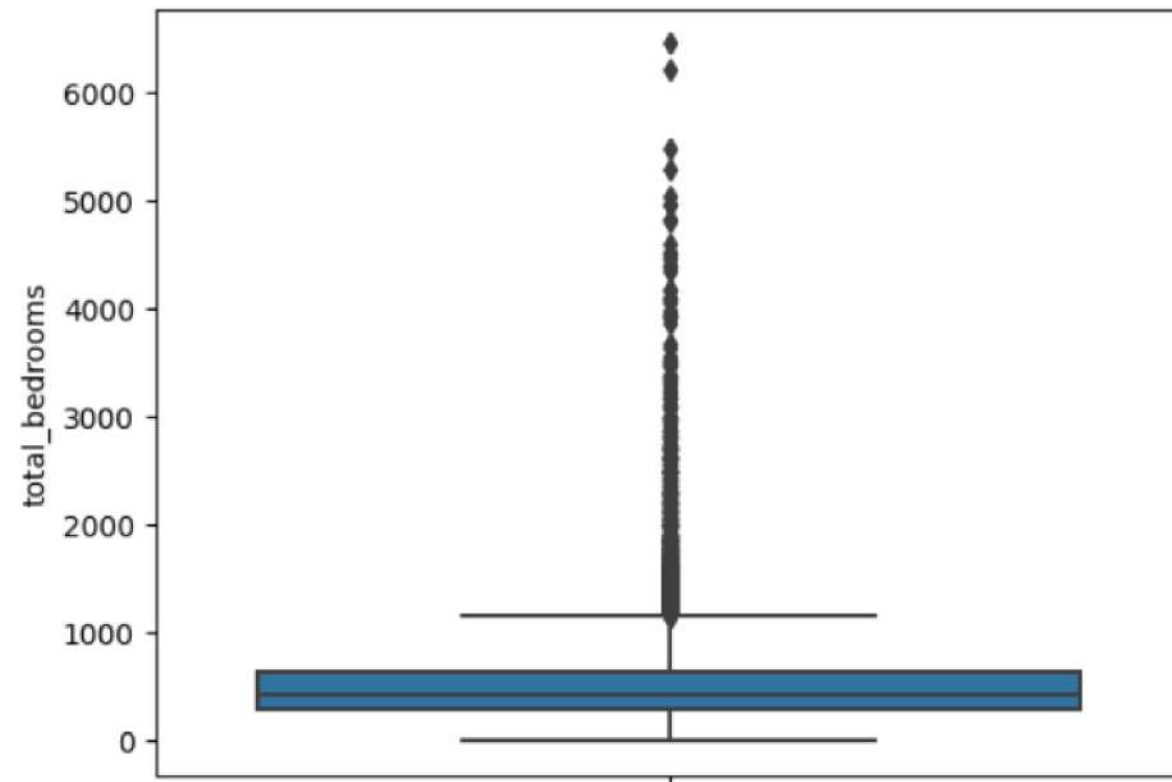
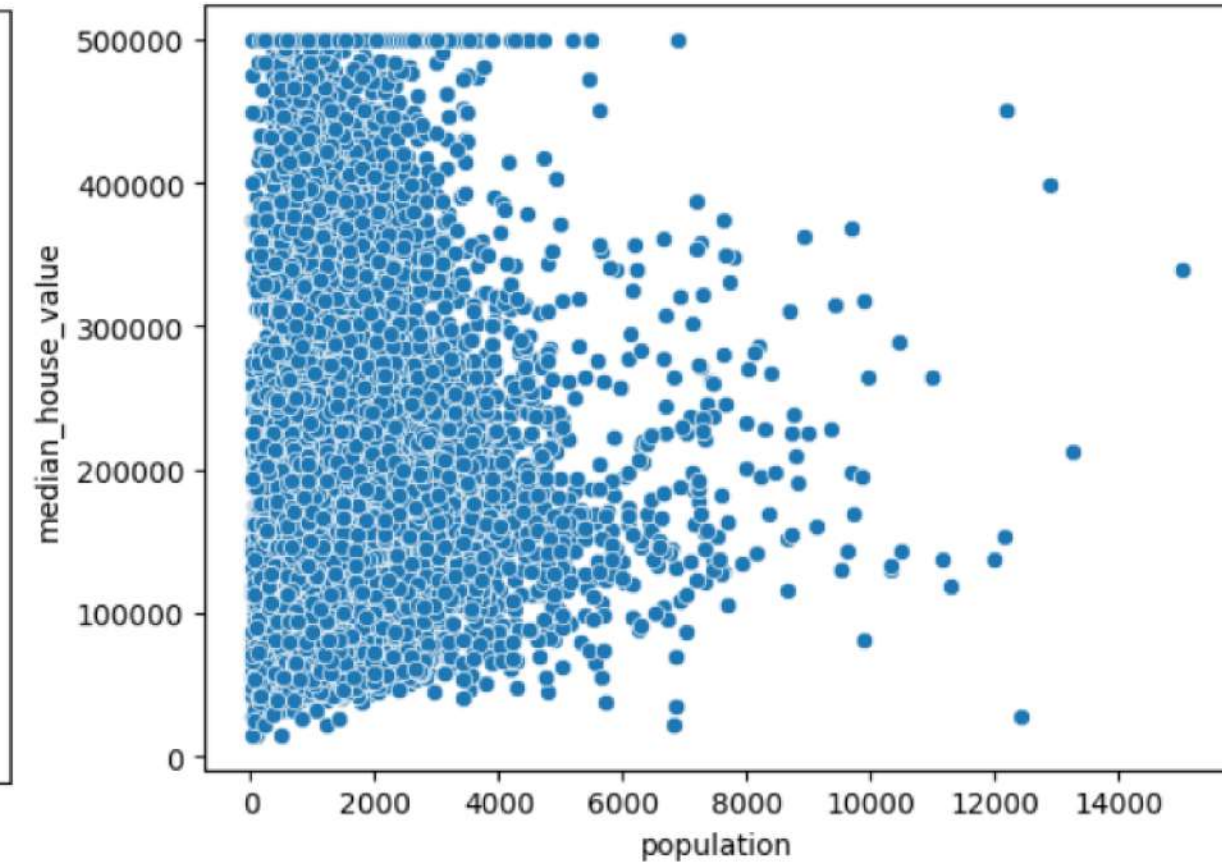
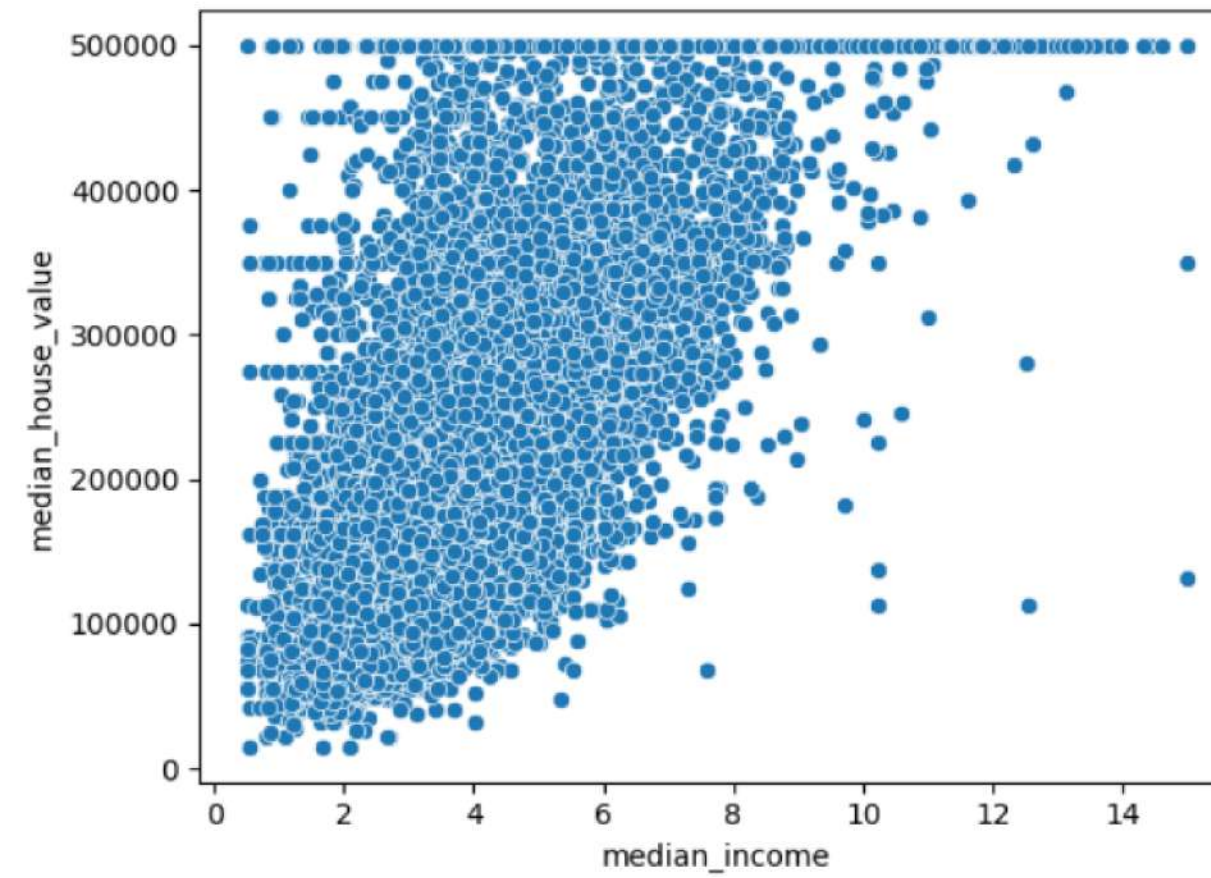
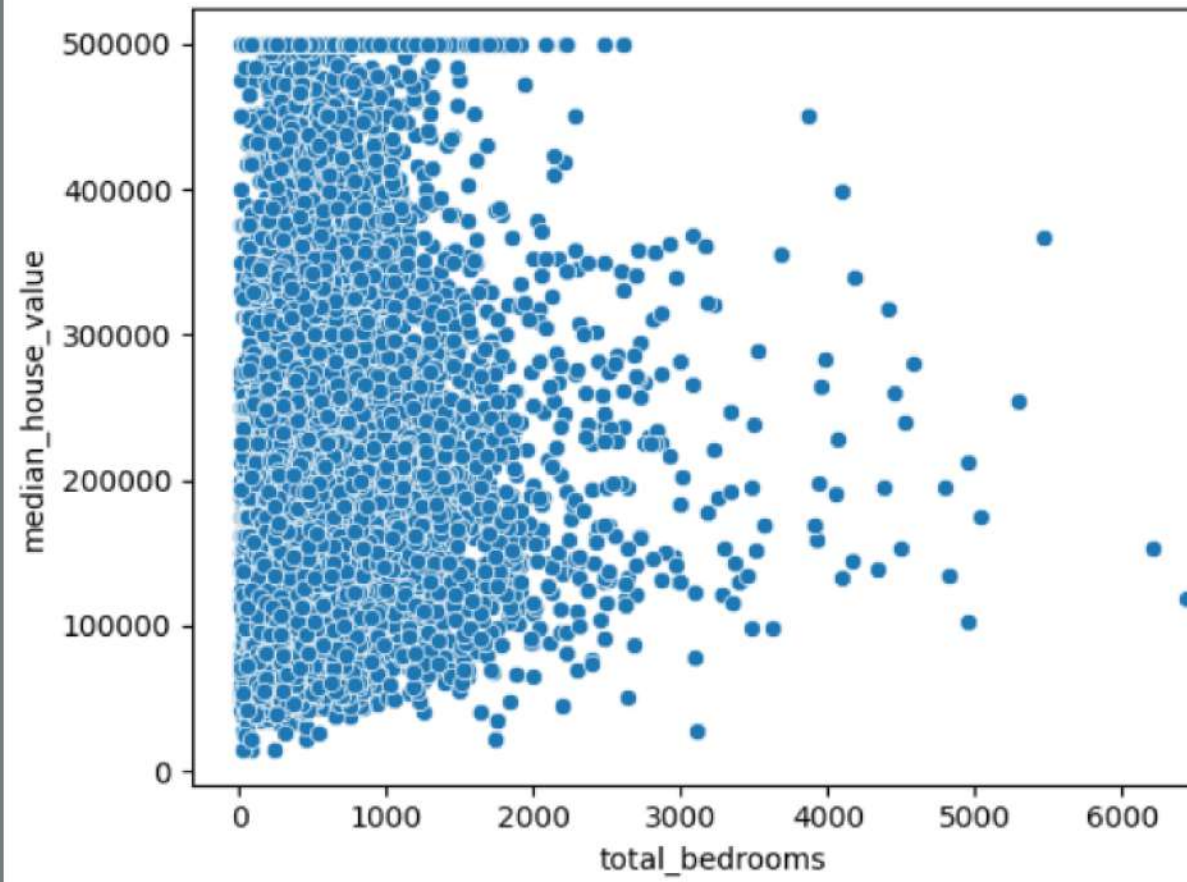
ocean_proximity

- INLAND
- <1H OCEAN
- NEAR OCEAN
- NEAR BAY
- ISLAND



1

Analyse



2

Corrélation

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
longitude	1.000000	-0.924753	-0.109873	0.044776	0.069623	0.099824	0.055629	-0.019501	-0.046754
latitude	-0.924753	1.000000	0.010653	-0.035012	-0.065210	-0.106416	-0.068901	-0.076643	-0.143970
housing_median_age	-0.109873	0.010653	1.000000	-0.362083	-0.324779	-0.295541	-0.307466	-0.113804	0.108626
total_rooms	0.044776	-0.035012	-0.362083	1.000000	0.932408	0.855772	0.920225	0.198840	0.133218
total_bedrooms	0.069623	-0.065210	-0.324779	0.932408	1.000000	0.878581	0.979994	-0.005643	0.047478
population	0.099824	-0.106416	-0.295541	0.855772	0.878581	1.000000	0.907647	0.004353	-0.027441
households	0.055629	-0.068901	-0.307466	0.920225	0.979994	0.907647	1.000000	0.014794	0.063245
median_income	-0.019501	-0.076643	-0.113804	0.198840	-0.005643	0.004353	0.014794	1.000000	0.689659
median_house_value	-0.046754	-0.143970	0.108626	0.133218	0.047478	-0.027441	0.063245	0.689659	1.000000

3

Preprocessing

distance_LA	distance_SF	distance_Ontario	<1H OCEAN	INLAND	ISLAND	NEAR BAY	NEAR OCEAN
331.215391	253.787363	359.902494	0	1	0	0	0
72.833475	617.562919	45.369747	1	0	0	0	0
327.527099	236.692058	363.628664	0	1	0	0	0
18.679677	550.840909	61.687520	1	0	0	0	0
488.188587	103.466026	524.340406	0	1	0	0	0
...
691.455141	205.155903	720.819055	0	1	0	0	0
568.222244	38.374421	610.296736	0	0	0	1	0
598.811547	162.984525	626.558166	0	1	0	0	0
36.302153	556.833241	46.124519	1	0	0	0	0
904.905600	366.235948	941.416422	0	0	0	0	1

- Haversine distance
- Calculate distance
- Hot One Encoding
- Drop "Unnamed: 0"
- KNNImputer pour les NaN

4

Test features

Les caractéristiques les plus importantes sont :
`Index(['longitude', 'latitude', 'median_income', 'INLAND', 'ISLAND'], dtype='object')`

```
Importance des features avec Lasso :
longitude          12788.709556
latitude           12547.869914
housing_median_age    803.282219
total_rooms          4.258114
total_bedrooms       81.714956
population           36.117512
households           54.714233
median_income       38500.627197
distance_LA         346.223599
distance_SF          30.014543
distance_Ontario     279.475669
<1H OCEAN           5552.651983
INLAND              33758.922113
ISLAND              160849.192888
NEAR BAY             18738.368644
NEAR OCEAN           24382.158563
dtype: float64
```

- Test RFE (Recursive Feature Elimination)

Principe : RFE est une technique qui permet de sélectionner les fonctionnalités les plus importantes en les éliminant progressivement.

- Test Lasso (Least Absolute Shrinkage and Selection Operator)

Principe : Lasso est une technique de régularisation qui permet de réduire les coefficients de régression et d'effectuer une sélection automatique des fonctionnalités.

Afin d'effectuer des tests de features en fonction de ce que nous donne les deux méthodes avec les tests de modèles.

5

Test Model & Selection

Methode	Score R ²
Multiple Linear - All columns	0,66
Multiple linear - without proximity ocean	0,65
Multiple linear - without proximity ocean + distance	0,64
Multiple linear - without proximity ocean + distance + latitude/longitude	0,57
Multiple Linear - without proximity ocean + median_income	0,37
Multiple Linear - without proximity ocean + households	0,64
Multiple Linear - without proximity ocean + population	0,61
Multiple Linear - without proximity ocean + total_bedrooms	0,63
Multiple Linear - without proximity ocean + total_rooms	0,64
Multiple Linear - without proximity ocean + housing_age_median	0,63
Multiple Linear - without proximity ocean + latitude	0,57
Multiple Linear - without proximity ocean + longitude	0,58
Multiple Linear - without longitude/latitude	0,66
Multiple Linear - without longitude/latitude + houshold	0,66
Multiple Linear - without longitude/latitude + houshold + total_rooms	0,66
Test Lasso qui donne un score aux les features	0,61
Test RFE qui donne les features	0,59
Multiple Linear - All columns (with standardisation)	0,66
Multiple Linear - All columns (with normalisation)	0,66
Multiple Linear - All columns - RandomForestRegressor	0,82
KNN - without latitude/longitude - total_rooms & total_bedrooms	0,7

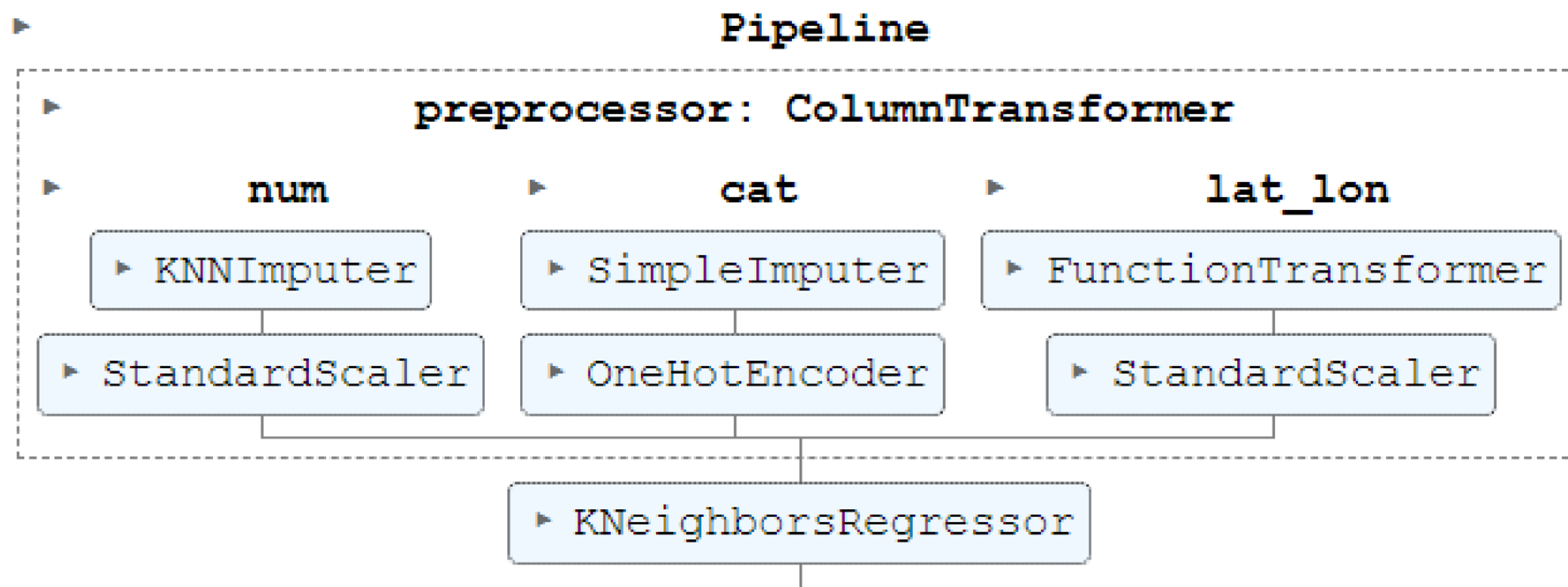
Modèle KNN

```
Meilleurs paramètres:
Score R2: 0.76
RMSE: 56616.02
```

- algorithm : ball_tree
- leaf_size : 10
- metric : manhattan
- n_neighbors : 9
- weights : distance

6

Pipeline

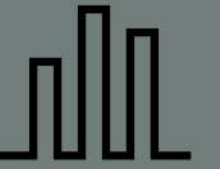


7

Streamlit



Streamlit





LOU&RAYAN
PREDICTION

THANK YOU!