



Trabalho 2

Imobiliária Properati | Predição de preço por metro quadrado

Grupo 1 – Deusdedith, Eduardo Henrique,
Felipe Ferraz, Lucas Ribeiro, Thiago Ferro, Inês Dantas

Dados Imobiliários Properati

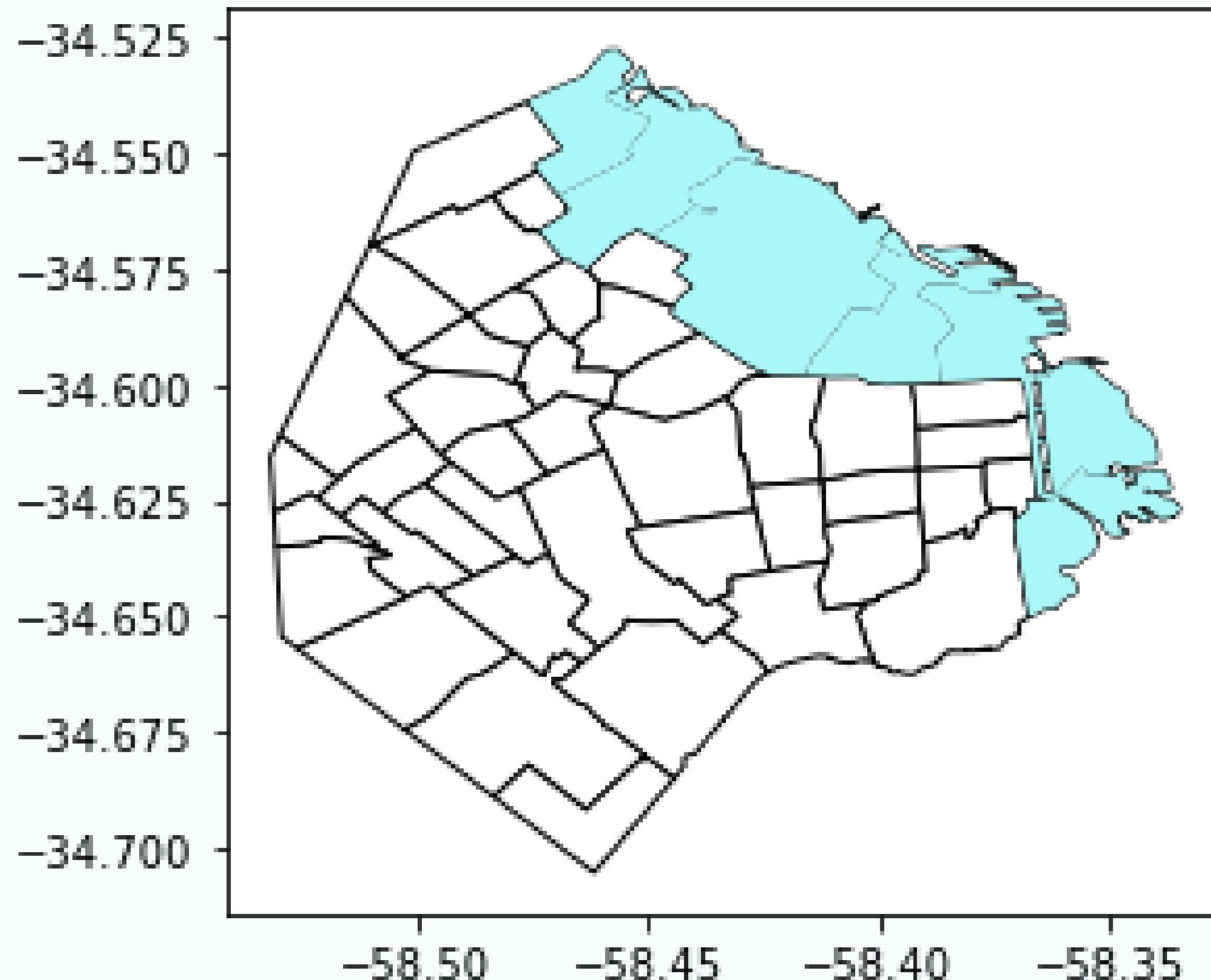
- Subconjunto escolhido
- Data Wrangling e Cleaning
- Análise de Modelos



Público Alvo - Apartamentos na zona litorânea da Argentina

Contém os bairros de:

- Puerto Madero;
- Palermo;
- Belgrano;
- Retiro;
- Boca;
- Rocola;
- Nuñez.



Features do Dataset

- Tipo de propriedade;
- Nome do bairro;
- Nome do país;
- Localização da propriedade;
- Latitude e longitude;
- Preço publicado na propaganda;
- Tipo de moeda utilizada no anuncio;
- Metragem da área do imóvel;
- Distância para o marco zero da cidade;
- População por bairro¹;
- Densidade populacional por bairro¹;
- Contém Universidades¹;
- Contém pontos turísticos¹.

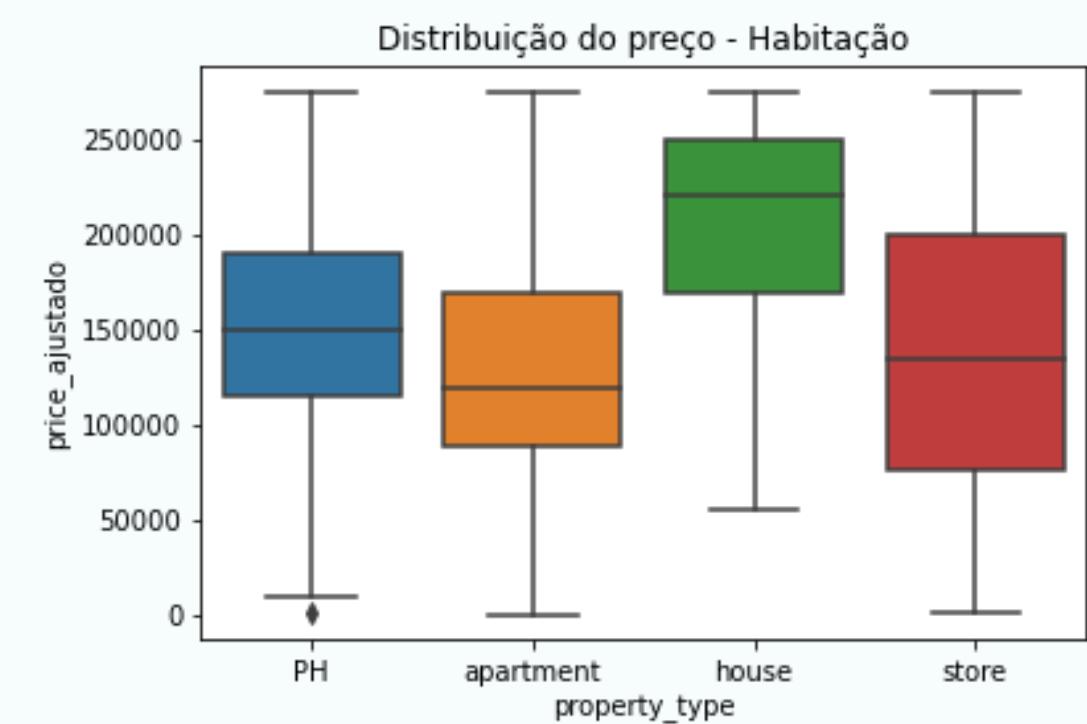
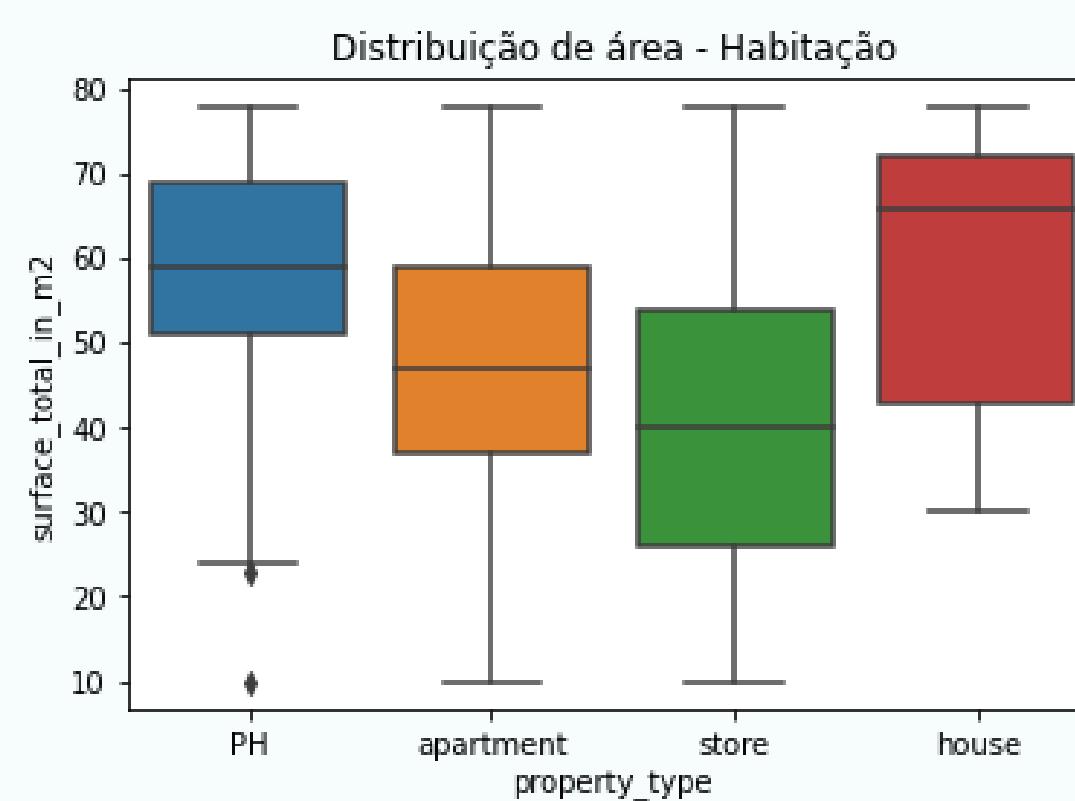
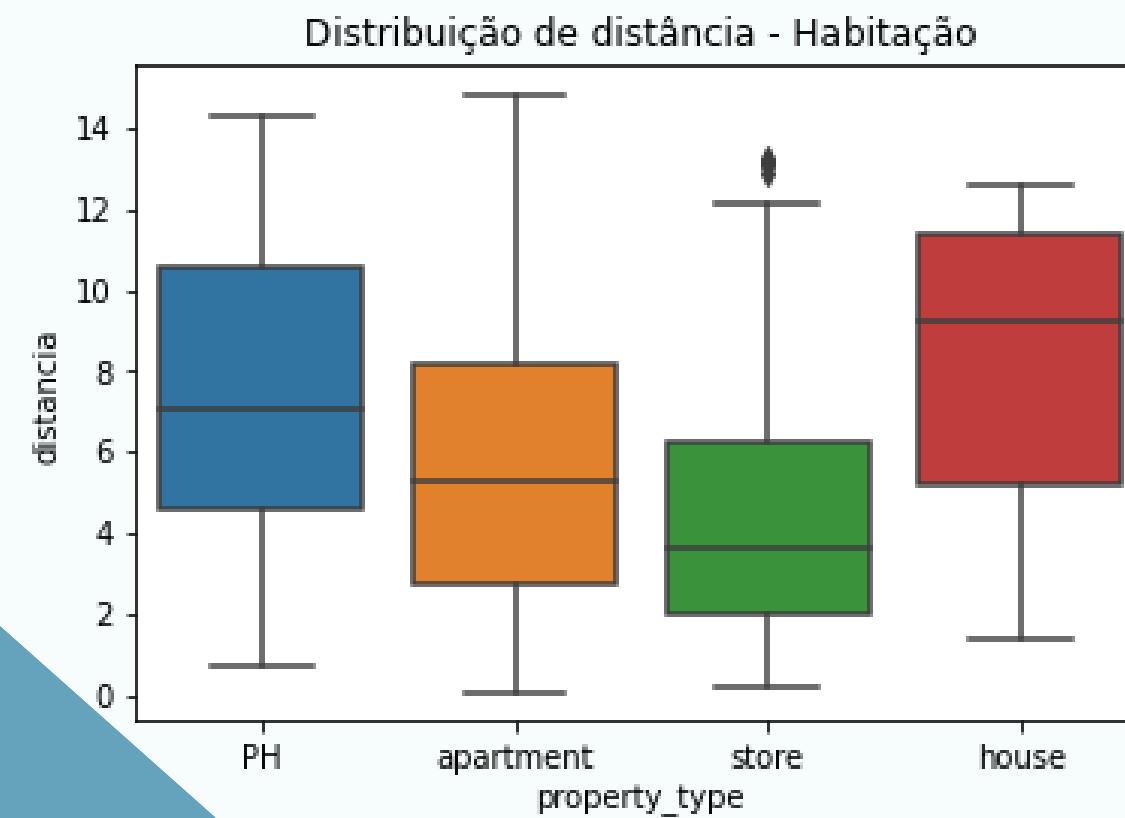


Preparando dados para ML

Resumo - Remoção Outliers

Inspeção Boxplots e estatísticas (quartil):

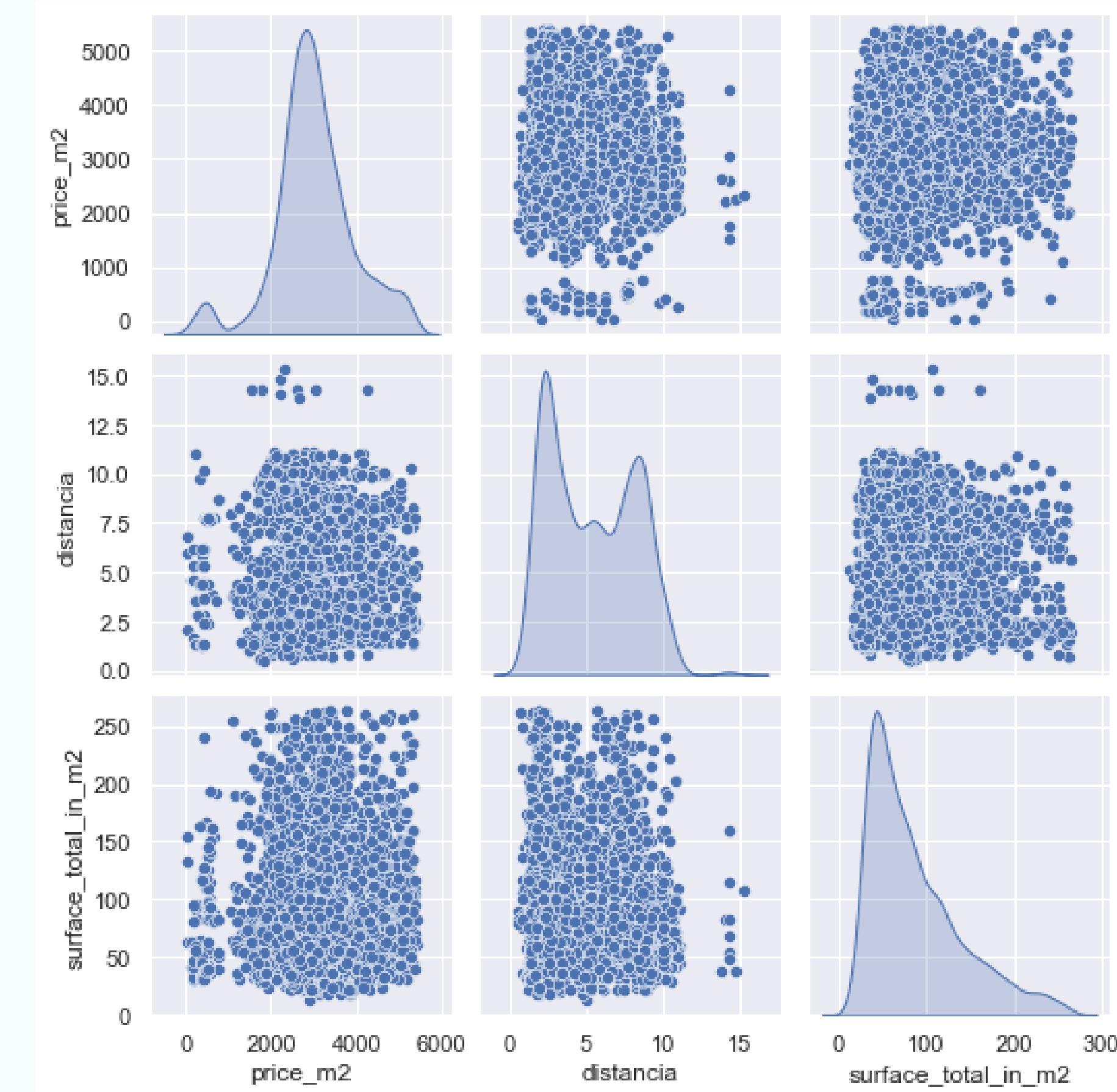
- Preço (convertido ARS → USD - 16/05/22);
- Área;
- Distância ao marco zero



Correlações

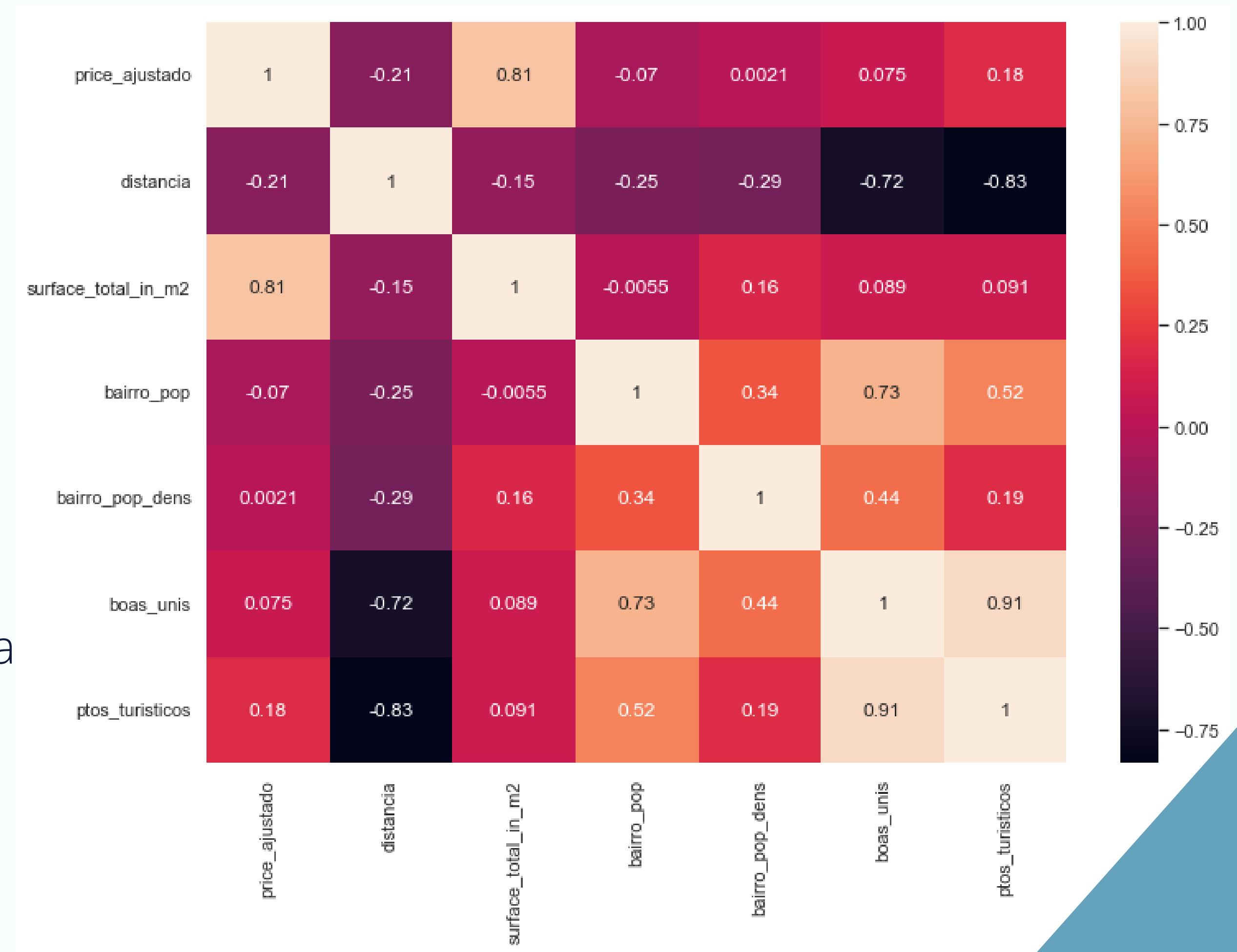
Variáveis numéricas

- Corr. de $$/m^2$ com
Distância e Área do imóvel



Multicolinearidade com Dummies

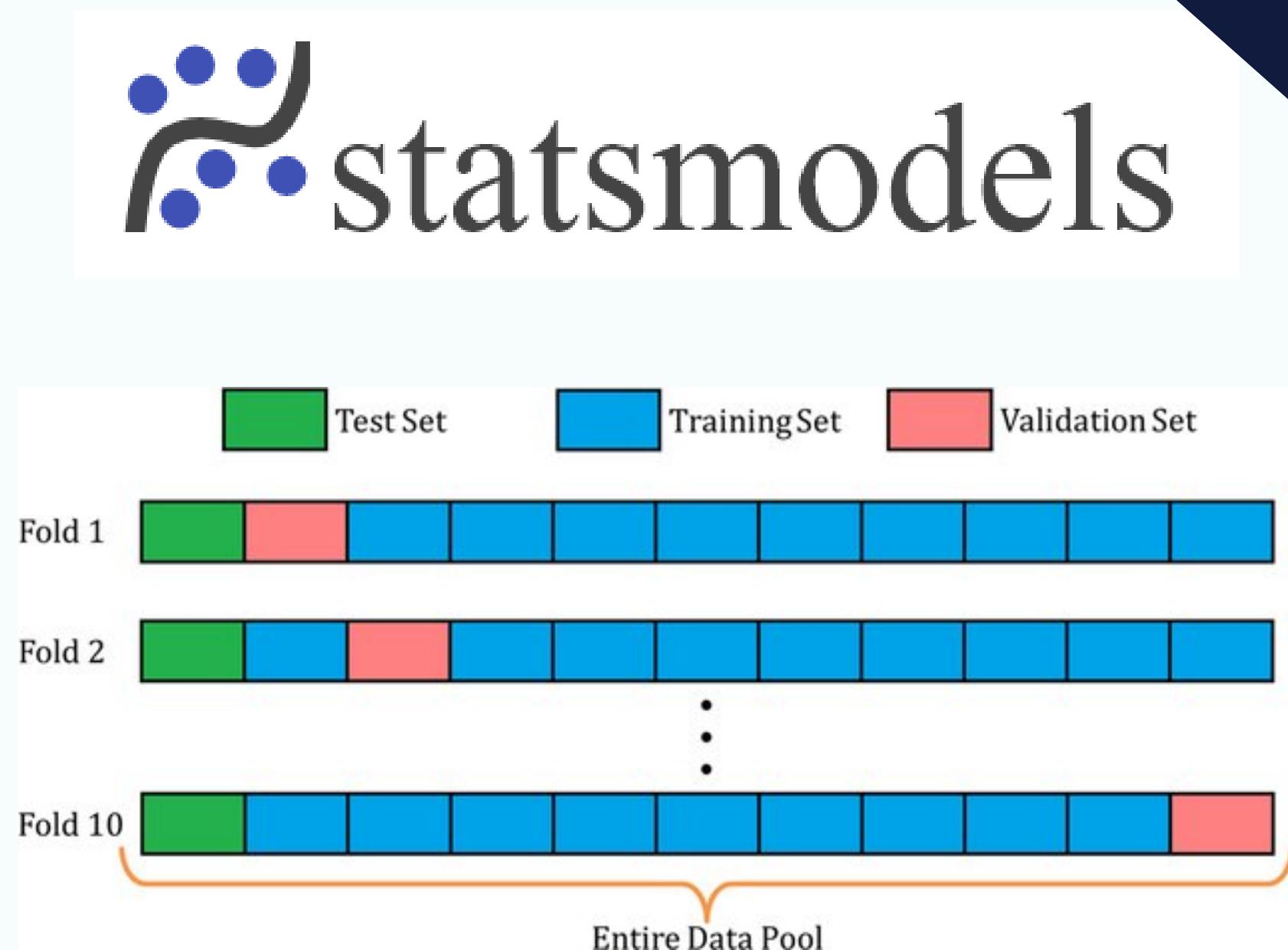
- Distância
- Pop. do Bairro
- Uma com a outra



Modelagem

Modelos

- Mínimos Quadrados da bib. StatsModels
- Modelos com CV:
 - OLS Regularizado
 - Lasso
 - Ridge
 - ElasticNet
 - ARD Regression



OLS

- Stats Models
- R^2 baixo
- Tabela de Parâmetros
 - Teste T-Student
 - P-valor

OLS Regression Results							
Dep. Variable:	price_m2		R-squared:	0.150			
Model:	OLS		Adj. R-squared:	0.149			
Method:	Least Squares		F-statistic:	81.25			
Date:	Sat, 02 Jul 2022		Prob (F-statistic):	7.48e-94			
Time:	13:25:12		Log-Likelihood:	-22565.			
No. Observations:	2760		AIC:	4.514e+04			
Df Residuals:	2753		BIC:	4.518e+04			
Df Model:	6						
Covariance Type:	nonrobust						
		coef	std err	t	P> t	[0.025	0.975]
	const	2658.0289	51.717	51.396	0.000	2556.621	2759.437
	distancia	216.2289	35.795	6.041	0.000	146.041	286.417
	surface_total_in_m2	39.9665	16.891	2.366	0.018	6.847	73.086
	bairro_pop	64.6044	28.510	2.266	0.024	8.701	120.508
	bairro_pop_dens	137.4332	24.097	5.703	0.000	90.183	184.684
	boas_unis	-2411.0922	158.701	-15.193	0.000	-2722.278	-2099.907
	ptos_turisticos	2845.3053	153.157	18.578	0.000	2544.992	3145.619
Omnibus:	160.386	Durbin-Watson:	1.974				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	377.747				
Skew:	-0.351	Prob(JB):	9.40e-83				
Kurtosis:	4.671	Cond. No.	20.1				

Regularização

- Lasso e Ridge: maior alfa, maior a penalização
 - Lasso $\rightarrow \alpha = 0.1$
 - Ridge $\rightarrow \alpha = 1e-5$ (aprox. zero)
- ElasticNet: modelo parecido com ridge, pois $\alpha = 0.12$

Conclusão

Tabela Comparativa - Parâmetros e Métricas

	Non Reg. OLS	Lasso Reg. OLS	Ridge Reg. OLS	ElasticNet Reg. OLS	ARD Regression
const	2658.028899	2561.755672	1994.425835	2562.546655	NaN
distancia	216.228856	185.815984	276.957536	184.349245	NaN
surface_total_in_m2	39.966455	47.839206	50.070013	47.777992	NaN
bairro_pop	64.604369	-67.850121	-299.353185	-68.273083	NaN
bairro_pop_dens	137.433184	20.868816	-109.682814	19.553623	NaN
boas_unis	-2411.092225	-1198.639306	508.312488	-1188.577782	NaN
ptos_turisticos	2845.305303	1856.422788	859.001521	1845.357301	NaN
MAE	638.773482	638.773496	638.887390	638.931454	639.199128
MAPE	0.483357	0.483357	0.483420	0.483453	0.483811
RMSE	879.032371	879.032372	879.035822	879.043082	879.444388
R2	0.132117	0.132117	0.132110	0.132096	0.131303

Conclusão

Tabela Comparativa - Predições

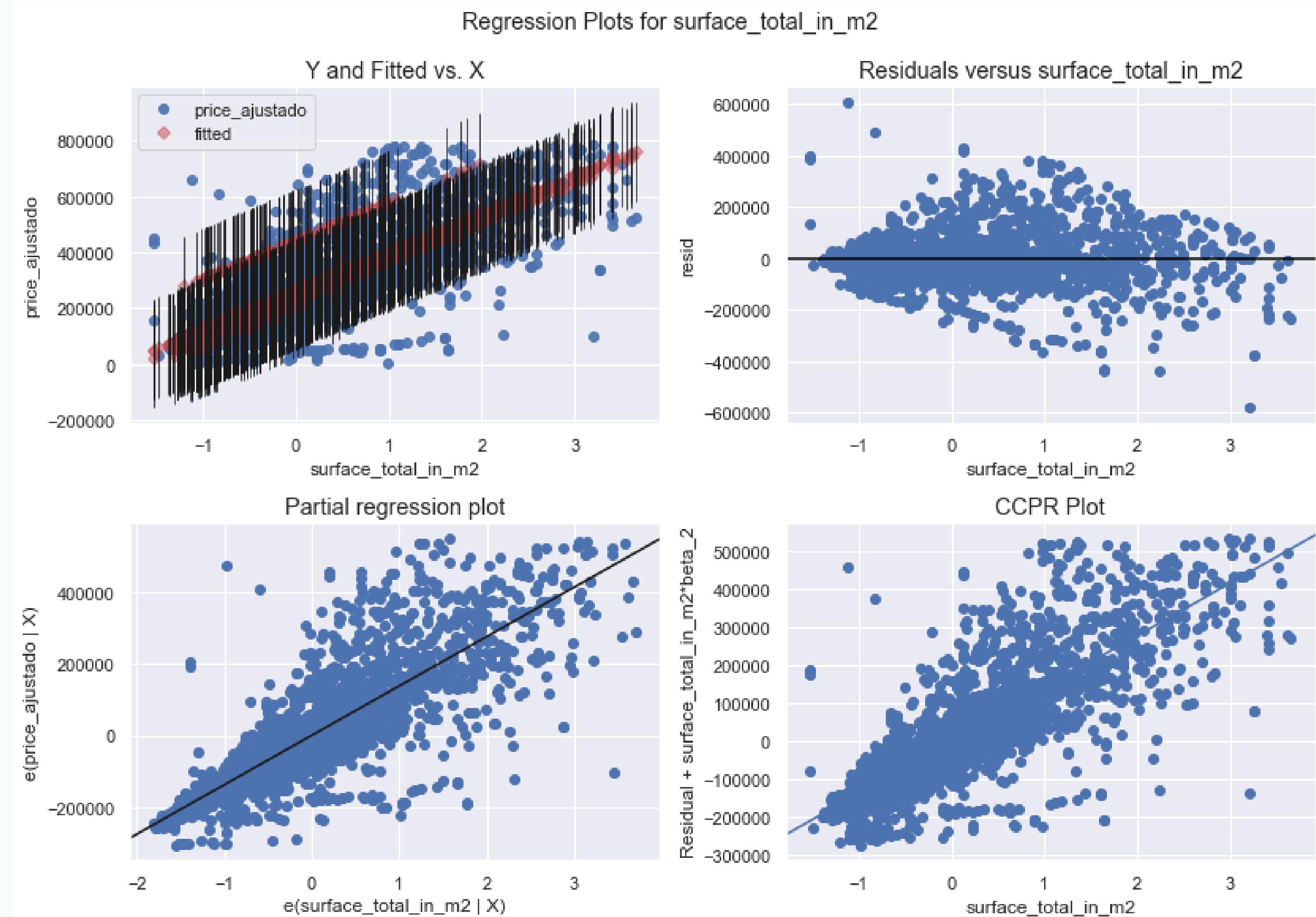
	y_test	Pred. non_reg_OLS	Pred. Lasso Reg	Pred. Ridge Reg	Pred. ElasticNet Reg	Pred. ARD Reg
0	265000.0	392491.288824	392433.671402	392412.477877	390898.401097	391247.266387
1	140000.0	124386.211624	124324.655194	124306.641057	122686.215642	122695.360893
2	145000.0	99020.347659	99046.903695	99057.580505	99758.049403	98661.112460
3	307650.0	250858.709117	250852.477831	250855.130938	250688.898276	250104.068379
4	319000.0	302960.408646	302921.253122	302913.690407	301876.011560	302182.508697

1191	187000.0	205374.421772	205390.943616	205400.280327	205834.210067	204936.273359
1192	128000.0	267750.757880	268170.721229	268326.665627	279358.774543	276379.273326
1193	170000.0	177016.145300	176987.507748	176981.369404	176226.688817	175866.444498
1194	135000.0	149311.807173	149274.899750	149265.219162	148290.063904	148460.330763
1195	600000.0	515103.231453	515073.655775	515073.504338	514286.808833	514176.819181

1196 rows × 6 columns

Resíduos

sm.graphics.
plot_regress_exog()





Obrigado pela atenção!