



IMPLEMENTAÇÃO DE LOJAS EM SÃO PAULO

PREDIÇÃO DE LOCAIS E FATURAMENTO

ESTUDO DA TABELA DE DADOS

- Essa foi a tabela fornecida para serem feitas previsões de onde implementar as lojas na cidade de São Paulo.
- Possui 456 linhas e 24 colunas, com dados sobre os bairros do Rio de Janeiro que possuem lojas e os bairros de São Paulo, cuja previsão será feita.

	codigo	nome	cidade	estado	população	popAte9	popDe10a14	popDe15a19	popDe20a24	popDe25a34	...
0	3304557060	Abolição	Rio de Janeiro	RJ	11676	1027	483	688	800	1675	...
1	3304557138	Acari	Rio de Janeiro	RJ	27564	5131	2188	2697	2630	4810	...
2	3304557057	Água Santa	Rio de Janeiro	RJ	9003	883	399	597	762	1755	...
3	3304557031	Alto Da Boa Vista	Rio de Janeiro	RJ	9606	1072	538	660	685	1317	...
4	3304557125	Anchieta	Rio de Janeiro	RJ	57222	7677	3774	4892	4600	8660	...
...
451	355030213	Vila Sônia	São Paulo	SP	34061	2908	1462	2253	2076	4579	...
452	355030207	Vila Suzana	São Paulo	SP	35403	4127	1890	2678	2433	5855	...
453	355030162	Vila Terezinha	São Paulo	SP	122359	18304	9304	13258	9965	19248	...
454	355030157	Vila Zatt	São Paulo	SP	125864	14670	7305	11225	9338	18841	...
455	355030164	Vista Alegre	São Paulo	SP	106	23	7	15	11	15	...

456 rows x 24 columns

TABELA QUE SERÁ TRABALHADA

- Seguindo as diretrizes da case, o público alvo são pessoas da faixa etária de 25 a 50 anos. A tabela ao lado está reduzida e renomeada para uma melhor a execução dos algoritmos.
- x_1 é a população total do bairro
- x_2 é a população de 25 a 34 anos
- x_3 é a população de 35 a 50 anos
- x_4 domicílios classe A1
- x_5 domicílios classe A2
- x_6 domicílios classe B1
- x_7 domicílios classe B2
- x_8 renda média
- y faturamento

	cidade	x1	x2	x3	x4	x5	x6	x7	x8	y	potencial
0	Rio de Janeiro	11676	1675	2300	0	145	715	1242	2501	932515.0	Médio
1	Rio de Janeiro	27564	4810	5308	0	0	82	506	931	588833.0	Baixo
2	Rio de Janeiro	9003	1755	2076	0	96	404	652	2391	874200.0	Baixo
3	Rio de Janeiro	9606	1317	2007	114	178	393	517	3727	912226.0	Médio
4	Rio de Janeiro	57222	8660	12272	0	0	1089	2821	1380	553020.0	Médio
...
451	São Paulo	34061	4579	7557	942	1089	2866	2219	5285	0.0	0
452	São Paulo	35403	5855	9107	1726	1522	2458	1186	7418	0.0	0
453	São Paulo	122359	19248	26592	0	0	1758	4517	1252	0.0	0
454	São Paulo	125864	18841	28500	8	872	5093	8063	1936	0.0	0
455	São Paulo	106	15	26	0	0	0	6	1288	0.0	0

- A seguir apresenta-se duas tabelas, uma com as informações do Rio de Janeiro, que será usada para treino e teste dos algoritmos e outra, com informações de São Paulo, que será usada para as previsões.
- A redução da tabela foi feita baseada nas orientações da case e no público alvo.
- Não foi utilizada a faixa de 50 a 59, pois ultrapassa além do público alvo e não era possível separar apenas as pessoas com 50 anos.



TABELA TREINO E TESTE

	cidade	x1	x2	x3	x4	x5	x6	x7	x8	y	potencial
0	Rio de Janeiro	11676	1675	2300	0	145	715	1242	2501	932515.0	Médio
1	Rio de Janeiro	27564	4810	5308	0	0	82	506	931	588833.0	Baixo
2	Rio de Janeiro	9003	1755	2076	0	96	404	652	2391	874200.0	Baixo
3	Rio de Janeiro	9606	1317	2007	114	178	393	517	3727	912226.0	Médio
4	Rio de Janeiro	57222	8660	12272	0	0	1089	2821	1380	553020.0	Médio
...
155	Rio de Janeiro	18791	2500	3944	0	259	774	1363	2316	901446.0	Médio
156	Rio de Janeiro	13556	2542	2908	0	219	714	828	2416	849731.0	Médio
157	Rio de Janeiro	33190	4752	7183	0	906	2866	2241	3189	1127042.0	Alto
158	Rio de Janeiro	8867	1205	1812	0	87	556	836	2418	921875.0	Baixo
159	Rio de Janeiro	2072	318	417	0	89	234	177	3791	919451.0	Baixo

160 rows x 11 columns

TABELA DE PREDIÇÃO

	cidade	x1	x2	x3	x4	x5	x6	x7	x8	y	potencial
160	São Paulo	94034	14535	21549	0	253	2197	4368	1501	0.0	0
161	São Paulo	32791	5341	7281	1413	1734	3704	2351	5920	0.0	0
162	São Paulo	104193	17749	23364	0	0	1423	4875	1284	0.0	0
163	São Paulo	12721	2142	2833	624	667	1558	1032	6278	0.0	0
164	São Paulo	48417	7388	10751	0	303	1794	2986	1905	0.0	0
...
451	São Paulo	34061	4579	7557	942	1089	2866	2219	5285	0.0	0
452	São Paulo	35403	5855	9107	1726	1522	2458	1186	7418	0.0	0
453	São Paulo	122359	19248	26592	0	0	1758	4517	1252	0.0	0
454	São Paulo	125864	18841	28500	8	872	5093	8063	1936	0.0	0
455	São Paulo	106	15	26	0	0	0	6	1288	0.0	0

296 rows x 11 columns

REGRESSÃO LINEAR MÚLTIPLA

R-squared (uncentered): 0.700

Adj. R-squared (uncentered): 0.685

F-statistic: 44.41

- A regressão múltipla fornece um relatório com informações importantes, aqui ao lado tem-se o *R-Squared* ou R^2 , que é o coeficiente de determinação da eficiência do modelo. Nesse caso, o modelo está funcionando com 70% de eficiência.
- Tem-se também o *F-Statistic*, que determina o quanto os dados estão dispersos, quanto maior o valor de F, mais dispersos os dados estão. Nesse caso o valor fornecido de 44.41 é um valor baixo, o que nos mostra a concentração dos dados.

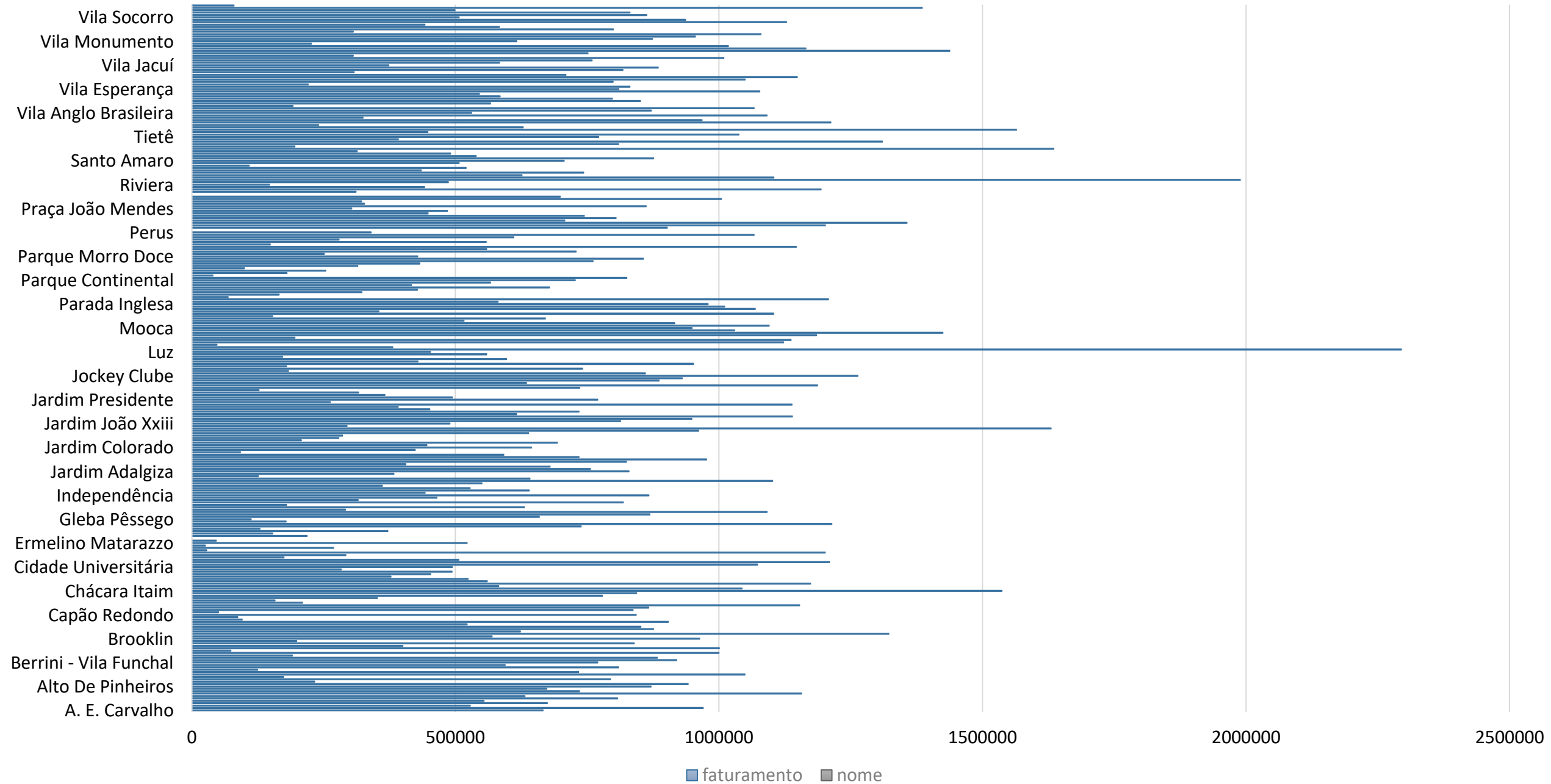
REGRESSÃO LINEAR MÚLTIPLA

	coef	std err	t	P> t	[0.025	0.975]
x1	-15.7486	4.910	-3.208	0.002	-25.449	-6.049
x2	8.4031	58.341	0.144	0.886	-106.860	123.667
x3	54.4932	47.039	1.158	0.248	-38.442	147.428
x4	112.5798	138.522	0.813	0.418	-161.096	386.256
x5	-360.5151	247.241	-1.458	0.147	-848.989	127.958
x6	272.6217	112.809	2.417	0.017	49.745	495.499
x7	57.2271	89.068	0.643	0.522	-118.743	233.198
x8	61.9321	7.714	8.028	0.000	46.691	77.173

Continuando o relatório da regressão, na tabela ao lado é possível encontrar:

- Coef: os coeficientes da regressão múltipla;
- *Std err*: os erros padrão (estimativa do desvio padrão do coeficiente);
- T: a significância individual dos parâmetros (testagem de parâmetro);
- $|P| > |t|$: P-valor, estatística que relata os resultados de um teste de hipótese, probabilidade exata de cometer um erro do tipo rejeitar a hipótese nula quando ela é verdadeira;
- [0.025 – 0.75]: intervalos de confiança.

Predição do Faturamento

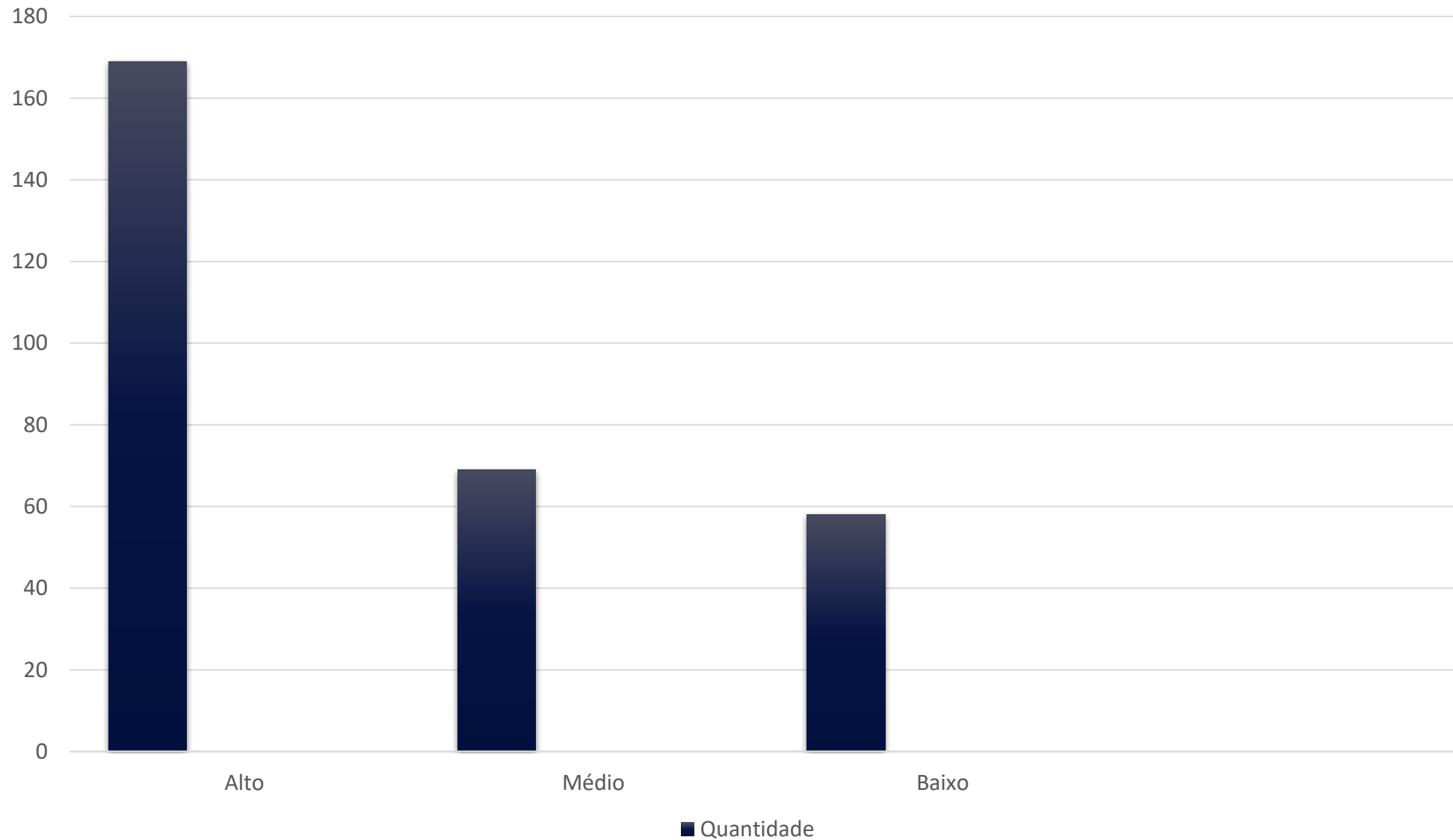


ÁRVORE DE DECISÃO

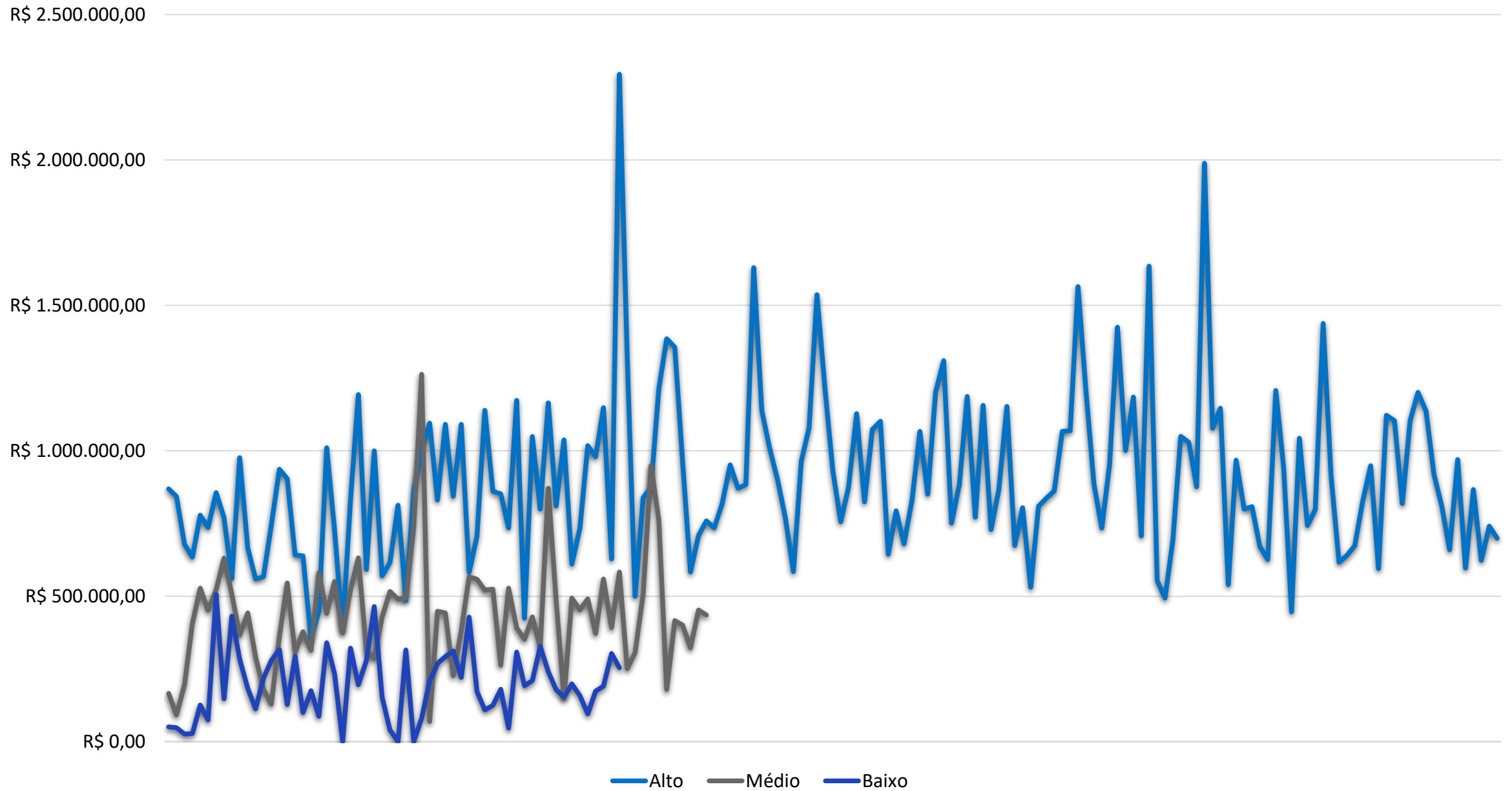
- A árvore de decisão foi usada para testar e treinar o modelo para predizendo o potencial dos bairros e classificando-os em “Baixo, Médio e Alto” potencial. Essa análise dos bairros, para implementação das lojas, foi feita de acordo com o público alvo e gerou a imagem a seguir .



Potencial dos bairros de São Paulo



Predição do potencial dos bairros e faturamento



Bairros	Faturamento	Potencial
Vila Matilde	R\$ 1.164.796,00	Alto
Vila Terezinha	R\$ 499.147,00	Alto
Vila Zatt	R\$ 1.385.515,00	Alto
Chácara Itaim	R\$ 1.536.597,00	Alto
Cidade Vargas	R\$ 1.073.043,00	Alto
Água Funda	R\$ 554.147,00	Alto
Ipiranga	R\$ 639.944,00	Alto
Luz	R\$ 452.461,00	Médio
Santa Efigênia	R\$ 435.166,00	Médio
Capivari - Monos	R\$ 50.691,00	Baixo
Cohab Jova Rural	R\$ 174.796,00	Baixo
Canindé	R\$ 95.216,00	Baixo

CONCLUSÃO

- Após feita a análise dos dados fornecidos, a tabela apresenta a predição final da implementação das novas lojas nos bairros da cidade de São, com possíveis faturamentos e a indicação do potencial de cada um.



OBRIGADA!!



MAILINEMORAES@GMAIL.COM

