

House Prices – RNK

GitHub: <https://github.com/RNanWP/HousePrices-RNK>

Equipe Responsável:

Renan Santos de Oliveira – 2222200490

Rafael Silva e Silva – 2222102004

Curso: Análise e Desenvolvimento de Sistemas

Turma: 40

Semestre: 4-5

Junho de 2024

Índice

House Prices – RNK.....	1
Índice.....	2
Relatório de Análise de Preços de Casas.....	3
Resumo Estatístico do Conjunto de Dados.....	4
Metodologia de Análise.....	5
1. Exploração e preparação dos dados	5
2. Desenvolvimento do modelo de regressão linear	5
3. Avaliação do desempenho do modelo.....	5
4. Interpretação dos resultados	5
Resultados do Modelo	5
Exploração e Preparação dos Dados	5
Desenvolvimento do Modelo de Regressão Linear	5
Avaliação do Desempenho do Modelo	6
Coeficientes do Modelo.....	6
Gráficos e Visualizações.....	7

Relatório de Análise de Preços de Casas

Este relatório apresenta uma análise dos preços de casas utilizando um modelo de regressão linear. O conjunto de dados utilizado contém informações sobre diversas características das casas e seus preços. As colunas do conjunto de dados são:

- *Taxa_Criminalidade* (**CRIM**): Taxa de criminalidade per capita por cidade.
- *Zona_Residencial* (**ZN**): Proporção de terrenos residenciais com mais de 25.000 pés quadrados.
- *Proporcao_Comercial* (**INDUS**): Proporção de acres comerciais não comerciais por cidade.
- *Proximidade_Rio* (**CHAS**): Variável fictícia que indica se a área faz fronteira com o rio Charles (1 se limita; 0 caso contrário).
- *Concentracao_NOX* (**NOX**): Concentração de óxidos nítricos (partes por 10 milhões).
- *Media_Quartos* (**RM**): Número médio de quartos por habitação.
- *Acessibilidade_Rodovias* (**RAD**): Índice de acessibilidade às rodovias radiais.
- *Taxa_Imposto* (**TAX**): Taxa de imposto sobre propriedade de valor integral por \$10.000.
- *Proporcao_Aluno_Professor* (**PTRATIO**): Proporção aluno-professor por cidade.
- *Proporcao_Negros* (**B**): $1000(B_k - 0.63)^2$, onde B_k é a proporção de negros por cidade.
- *Proporcao_Status_Baixo* (**LSTAT**): Porcentagem da população com status socioeconômico baixo.
- *Valor_Medio* (**MEDV**): Valor médio das casas ocupadas pelos proprietários em \$1000s.

Resumo Estatístico do Conjunto de Dados

	<i>Contagem</i>	<i>Média</i>	<i>Desvio Padrão</i>	<i>Mínimo</i>	<i>25%</i>	<i>50%</i>	<i>75%</i>	<i>Max</i>
<i>CRIM</i>	<u>506.0</u>	3.613	8.601	0.006	0.082	0.256	3.677	88.97
<i>ZN</i>	<u>506.0</u>	11.36	23.32	0.000	0.000	0.000	12.50	100.0
<i>INDUS</i>	506.0	11.13	6.86	0.460	5.190	9.690	18.10	27.74
<i>CHAS</i>	506.0	0.069	0.253	0.000	0.000	0.000	0.000	1.000
<i>NOX</i>	506.0	0.554	0.115	0.385	0.449	0.538	0.624	0.871
<i>RM</i>	506.0	6.284	0.702	3.561	5.885	6.208	6.623	8.780
<i>RAD</i>	506.0	9.549	8.707	1.000	4.000	5.000	24.00	24.00
<i>TAX</i>	506.0	408.23	168.5	187.0	279.0	330.0	666.0	711.0
<i>PTRATIO</i>	506.0	18.45	2.164	12.60	17.40	19.05	20.20	22.00
<i>B</i>	506.0	356.67	91.29	0.320	375.3	391.4	396.2	396.9
<i>LSTAT</i>	506.0	12.65	7.141	1.730	6.950	11.36	16.95	37.97
<i>MEDV</i>	506.0	22.53	9.197	5.000	17.02	21.20	25.00	50.00

	Taxa_Crim	Z_Residencial	Propor_Comercial	Proxi_Rio	Concentracao_NOX	Media_Quartos	...
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	...
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	...
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	...
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	...
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	...
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	...
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	...
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	...

Acessi_Rodovias	Taxa_Imposto	Propor_Aluno_Professor	Propor_Negros	Propor_Status_Baixo	Valor_Medio
506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
9.549407	408.237154	18.455534	356.674032	12.653063	22.532806
8.707259	168.537116	2.164946	91.294864	7.141062	9.197104
1.000000	187.000000	12.600000	0.320000	1.730000	5.000000
4.000000	279.000000	17.400000	375.377500	6.950000	17.025000
5.000000	330.000000	19.050000	391.440000	11.360000	21.200000
24.000000	666.000000	20.200000	396.225000	16.955000	25.000000
24.000000	711.000000	22.000000	396.900000	37.970000	50.000000

Metodologia de Análise

A metodologia utilizada para esta análise envolve os seguintes passos:

1. **Exploração e preparação dos dados:** Análise inicial dos dados para identificar características importantes e preparação dos dados para o modelo.
2. **Desenvolvimento do modelo de regressão linear:** Divisão dos dados em conjuntos de treinamento e teste, e treinamento do modelo de regressão linear.
3. **Avaliação do desempenho do modelo:** Utilização de métricas adequadas para avaliar o desempenho do modelo.
4. **Interpretação dos resultados:** Interpretação dos coeficientes do modelo e análise da relação entre as variáveis predictoras e a variável alvo.

Resultados do Modelo

Exploração e Preparação dos Dados

A análise inicial dos dados revelou que não havia valores ausentes, o que simplificou o processo de limpeza dos dados. Em seguida, os dados foram explorados através de estatísticas descritivas e gráficos de dispersão para identificar relações entre variáveis importantes, como a relação entre a média de quartos (RM) e o valor médio das casas (MEDV).

Desenvolvimento do Modelo de Regressão Linear

Os dados foram divididos em conjuntos de treinamento (80%) e teste (20%). Em seguida, um modelo de regressão linear foi treinado utilizando o conjunto de treinamento. As variáveis predictoras incluíram todas as colunas, exceto Valor_Medio, que foi a variável alvo.

Avaliação do Desempenho do Modelo

O desempenho do modelo foi avaliado utilizando o erro quadrático médio (MSE) e o coeficiente de determinação (R^2). Os resultados foram os seguintes:

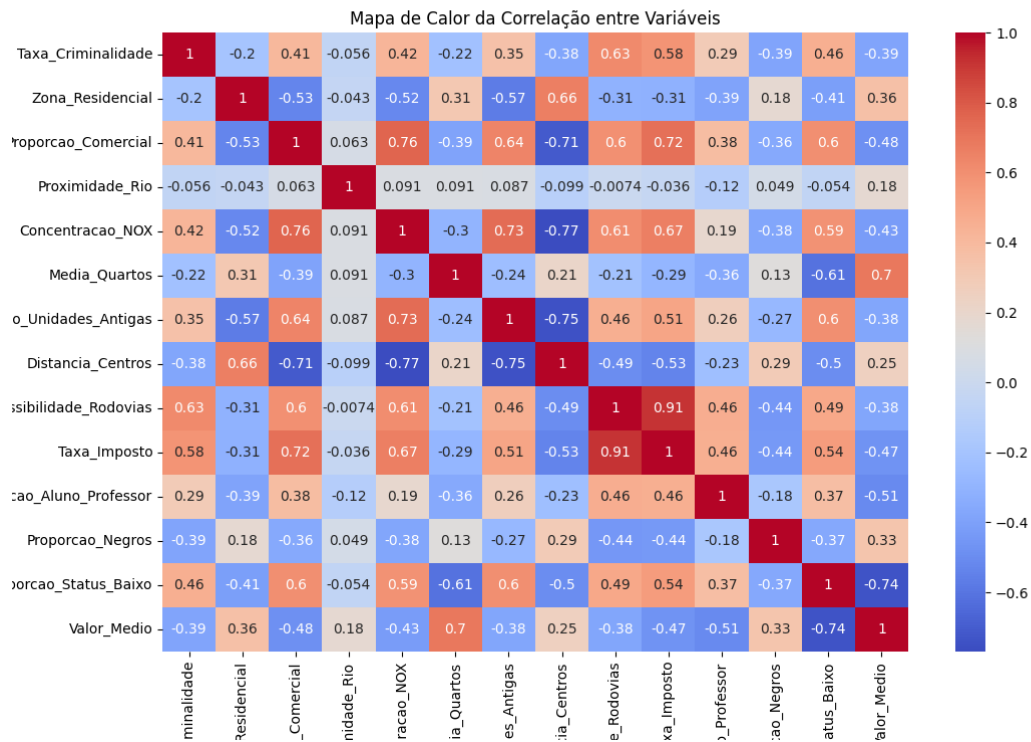
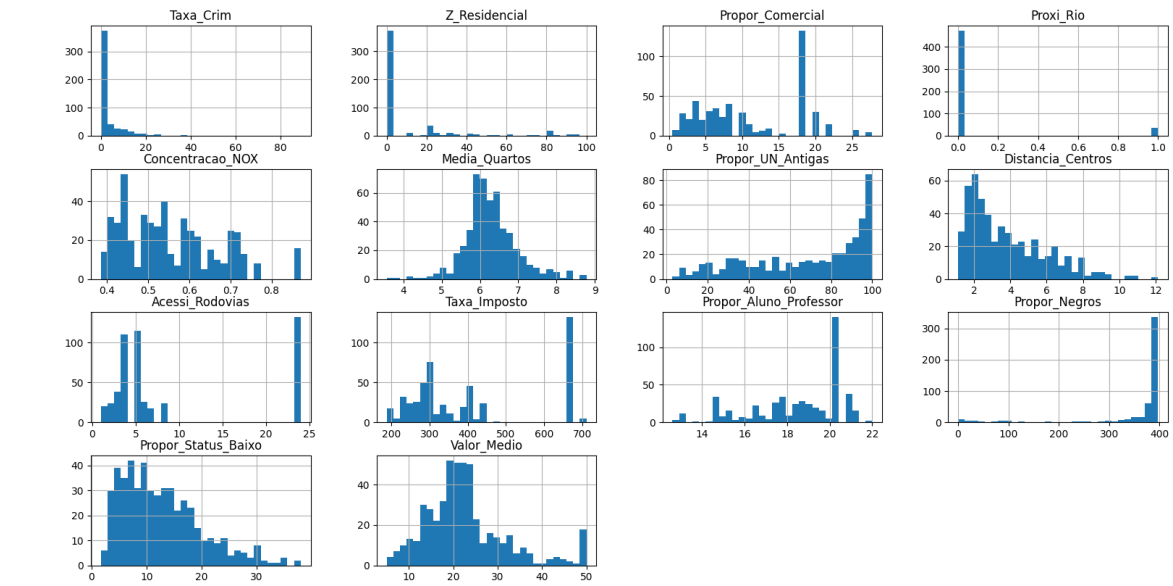
- **Erro Quadrático Médio (MSE):** 24.291119474973538
- **Coeficiente de Determinação (R^2):** 0.6687594935356317

Coeficientes do Modelo

Os coeficientes do modelo de regressão linear são apresentados na tabela abaixo:

Variável	Coeficiente
Taxa_Criminalidade	-0.113056
Zona_Residencial	0.030110
Proporcao_Comercial	0.040381
Proximidade_Rio	2.784438
Concentracao_NOX	-17.202633
Media_Quartos	4.438835
Proporcao_Unidades_Antigas	-0.006296
Distancia_Centros	-1.447865
Acessibilidade_Rodovias	0.262430
Taxa_Imposto	-0.010647
Proporcao_Aluno_Professor	-0.915456
Proporcao_Negros	0.012351
Proporcao_Status_Baixo	-0.508571

Gráficos e Visualizações



	Taxa_Crim	Z_Residencial	Propor_Comercial	Proxi_Rio	Concentracao_NOX	Media_Quartos	Propor_UN_Antigas
0	0.00632	18.0	2.31	0	0.538	6.575	65.2
1	0.02731	0.0	7.07	0	0.469	6.421	78.9
2	0.02729	0.0	7.07	0	0.469	7.185	61.1
3	0.03237	0.0	2.18	0	0.458	6.998	45.8
4	0.06905	0.0	2.18	0	0.458	7.147	54.2

	Distancia_Centros	Acessi_Rodovias	Taxa_Imposto	Propor_Aluno_Professor	Propor_Negros	Propor_Status_Baixo	Valor_Medio
	4.0900	1	296.0	15.3	396.90	4.98	24.0
	4.9671	2	242.0	17.8	396.90	9.14	21.6
	4.9671	2	242.0	17.8	392.83	4.03	34.7
	6.0622	3	222.0	18.7	394.63	2.94	33.4
	6.0622	3	222.0	18.7	396.90	5.33	36.2

Data columns (total 14 columns):

#	Column	Non-Null Count	Dtype
0	Taxa_Crim	506 non-null	float64
1	Z_Residencial	506 non-null	float64
2	Propor_Comercial	506 non-null	float64
3	Proxi_Rio	506 non-null	int64
4	Concentracao_NOX	506 non-null	float64
5	Media_Quartos	506 non-null	float64
6	Propor_UN_Antigas	506 non-null	float64
7	Distancia_Centros	506 non-null	float64
8	Acessi_Rodovias	506 non-null	int64
9	Taxa_Imposto	506 non-null	float64
10	Propor_Aluno_Professor	506 non-null	float64
11	Propor_Negros	506 non-null	float64
12	Propor_Status_Baixo	506 non-null	float64
13	Valor_Medio	506 non-null	float64

[8 rows x 14 columns]

Taxa_Crim	0
Z_Residencial	0
Propor_Comercial	0
Proxi_Rio	0
Concentracao_NOX	0
Media_Quartos	0
Propor_UN_Antigas	0
Distancia_Centros	0
Acessi_Rodovias	0
Taxa_Imposto	0
Propor_Aluno_Professor	0
Propor_Negros	0
Propor_Status_Baixo	0
Valor_Medio	0
dtype: int64	
Erro Quadrático Médio: 24.291119474973538	
Coefficiente de Determinação R²: 0.6687594935356317	

	Taxa_Crim	Z_Residencial	Propor_Comercial	Proxi_Rio	Concentracao_NOX	Media_Quartos	...
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	...
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	...
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	...
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	...
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	...
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	...
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	...
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	...

	Acessi_Rodovias	Taxa_Imposto	Propor_Aluno_Professor	Propor_Negros	Propor_Status_Baixo	Valor_Medio
	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
	9.549407	408.237154	18.455534	356.674032	12.653063	22.532806
	8.707259	168.537116	2.164946	91.294864	7.141062	9.197104
	1.000000	187.000000	12.600000	0.320000	1.730000	5.000000
	4.000000	279.000000	17.400000	375.377500	6.950000	17.025000
	5.000000	330.000000	19.050000	391.440000	11.360000	21.200000
	24.000000	666.000000	20.200000	396.225000	16.955000	25.000000
	24.000000	711.000000	22.000000	396.900000	37.970000	50.000000

	Coefficiente
Taxa_Crim	-0.113056
Z_Residencial	0.030110
Propor_Comercial	0.040381
Proxi_Rio	2.784438
Concentracao_NOX	-17.202633
Media_Quartos	4.438835
Propor_UN_Antigas	-0.006296
Distancia_Centros	-1.447865
Acessi_Rodovias	0.262430
Taxa_Imposto	-0.010647
Propor_Aluno_Professor	-0.915456
Propor_Negros	0.012351
Propor_Status_Baixo	-0.508571

Conclusão

Os resultados obtidos com o modelo de regressão linear indicam que certas variáveis têm um impacto significativo nos preços das casas. Em particular, a média de quartos (*Media_Quartos*) e a proximidade do rio Charles (*Proximidade_Rio*) foram encontradas como variáveis preditoras significativas.

Este modelo pode ser utilizado para prever preços futuros com uma precisão razoável. As implicações deste modelo são úteis para agentes imobiliários, compradores e investidores que desejam entender os fatores que influenciam os preços das casas.