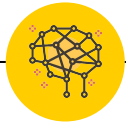


# Projeto **Aplicado**

Curso de Estatística para Ciência de Dados e Business Analytics





# Olá!

*Eu sou* **Liliane de Aquino**

In: <https://www.linkedin.com/in/liliane-l-de-aquino-a2999898/>



# Conjunto de dados

## Pesquisa Nacional por Amostra de Domicílios - 2015

**Fonte dos dados:**

<https://www.ibge.gov.br/estatisticas/sociais/educacao/9127-pesquisa-nacional-por-amostra-d-e-domicilios.html?=&t=microdados>



# Visão dos dados

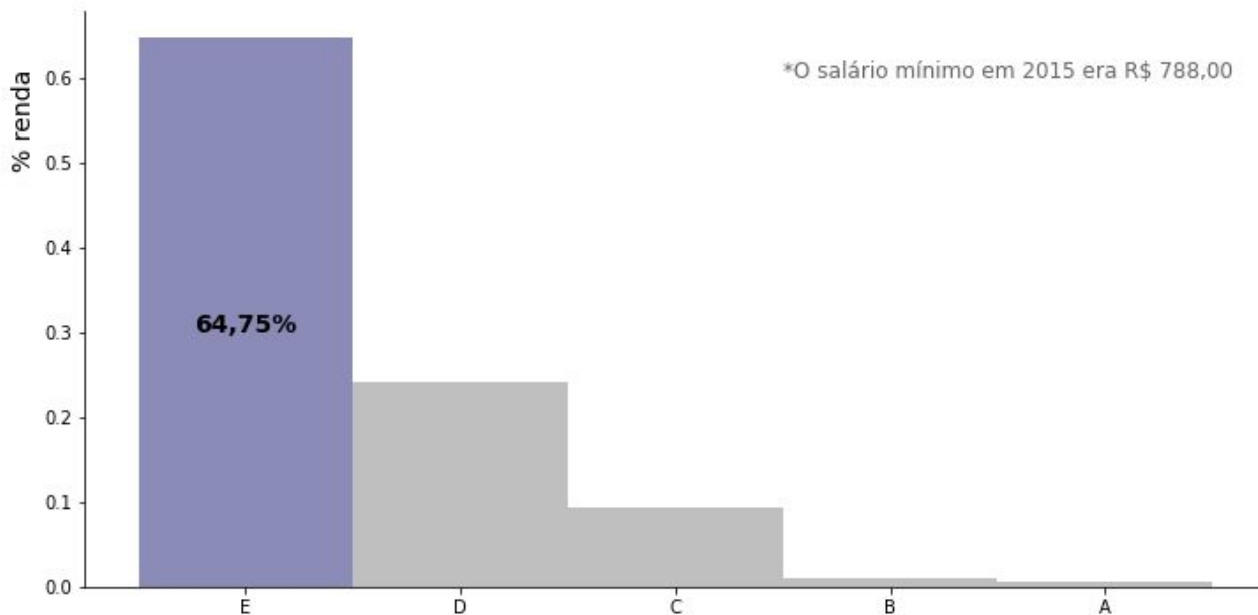
	Estado	Sexo biologico	Idade	Raca	Anos de Estudo	Renda
0	11	0	23	8	12	800
1	11	1	23	2	12	1150
2	11	1	35	8	15	880
3	11	0	46	2	6	3500
4	11	1	47	8	9	150
5	11	1	34	8	12	790
6	11	0	57	8	12	3150
7	11	1	60	8	12	1700
8	11	1	50	4	14	1800
9	11	0	26	8	12	1150



# Comportamento **renda**

## Distribuição da renda por classe

Comportamento da renda considerando a divisão por classe de salário mínimo

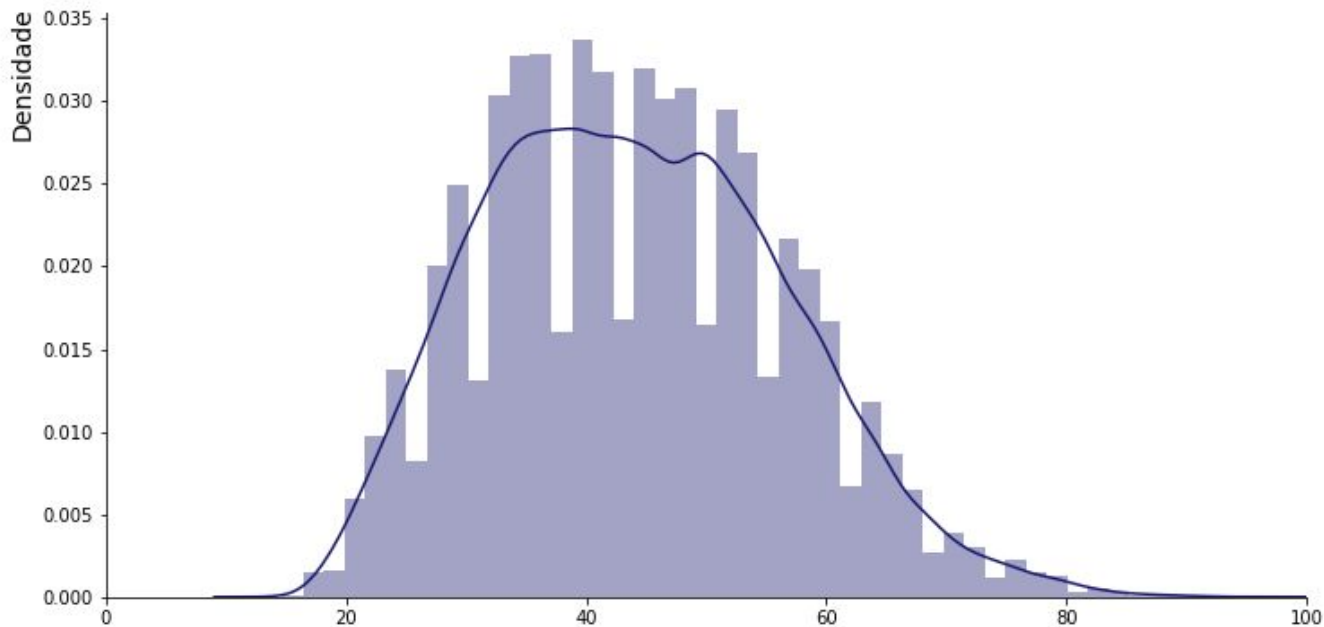




# Análise histogramas

## Histograma para idade

Análise da distribuição para a variável idade em anos

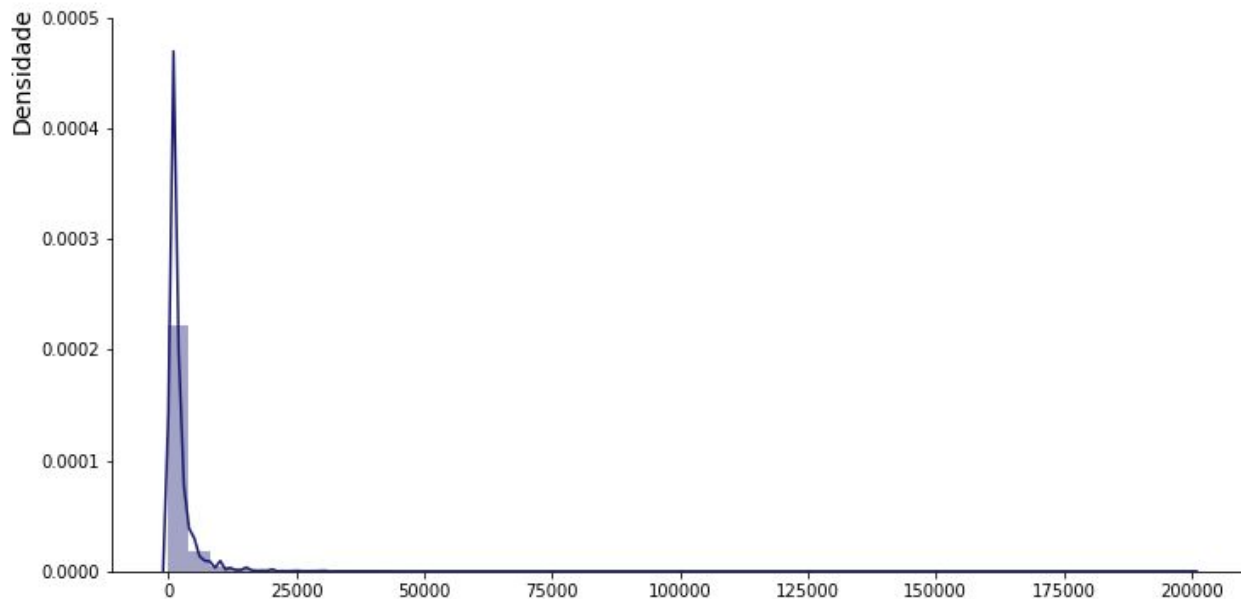




# Análise histogramas

## Histograma para renda

Análise da distribuição para a variável renda em reais





# Análise **descritiva**

	<b>Idade</b>	<b>Anos de Estudo</b>	<b>Renda</b>
<b>mean</b>	44.07	8.43	2000.38
<b>std</b>	12.48	4.54	3323.39
<b>min</b>	13.00	0.00	0.00
<b>25%</b>	34.00	5.00	788.00
<b>50%</b>	43.00	9.00	1200.00
<b>75%</b>	53.00	11.00	2000.00
<b>max</b>	99.00	15.00	200000.00



# Renda vs. raça e sexo

Sexo biologico	max		mean		median		std	
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino
Raca								
Indígena	3000.0	120000.0	911.10	3350.96	788.0	788.0	782.60	16531.48
Branca	80000.0	100000.0	2919.97	2139.75	1700.0	1200.0	4492.59	3384.96
Preta	48000.0	15000.0	1620.00	1173.74	1200.0	800.0	2178.64	1420.79
Amarela	30000.0	13000.0	5154.06	3019.50	2750.0	1800.0	6463.06	3252.39
Parda	90000.0	24500.0	1642.98	1179.74	1100.0	800.0	2423.05	1562.94



# Renda vs. estudo e sexo

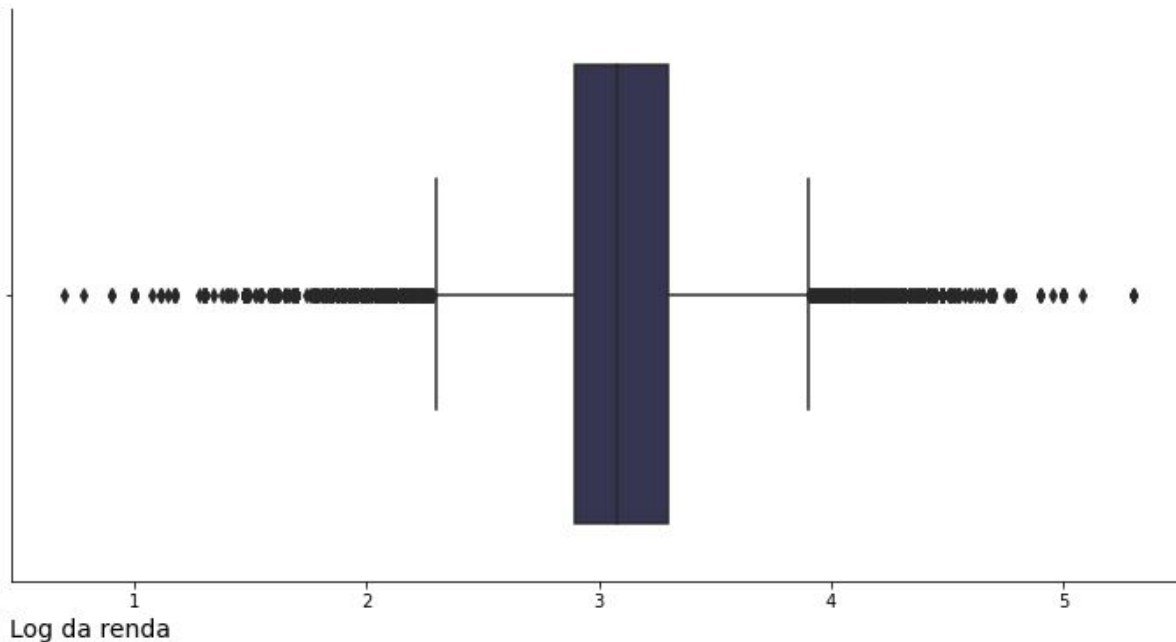
Sexo biológico	max		mean		median		std	
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino
Anos de Estudo								
<b>Sem instrução e menos de 1 ano</b>	30000.0	3000.0	803.34	494.46	700.0	300.0	1271.26	519.41
<b>1 ano</b>	6000.0	2000.0	834.33	480.58	788.0	390.0	776.10	427.68
<b>2 anos</b>	7000.0	2500.0	912.70	538.90	788.0	450.0	795.56	447.19
<b>3 anos</b>	5000.0	3500.0	945.70	531.80	788.0	455.0	785.12	441.07
<b>4 anos</b>	14000.0	10000.0	1317.07	682.29	1000.0	788.0	1184.52	614.74
<b>5 anos</b>	20000.0	8000.0	1288.29	792.16	1000.0	788.0	1229.48	703.63
<b>6 anos</b>	20000.0	3150.0	1526.37	801.37	1200.0	788.0	1650.07	504.00
<b>7 anos</b>	40000.0	4000.0	1631.95	848.70	1300.0	788.0	2395.48	546.38
<b>8 anos</b>	25000.0	18000.0	1654.93	972.26	1300.0	789.0	1576.35	1075.90
<b>9 anos</b>	15000.0	3500.0	1378.39	832.23	1100.0	788.0	1201.64	574.09
<b>10 anos</b>	18400.0	6000.0	1811.93	945.60	1300.0	806.0	1961.51	664.50
<b>11 anos</b>	50000.0	100000.0	2099.84	1314.50	1500.0	1000.0	2328.05	2161.17
<b>12 anos</b>	14000.0	120000.0	2355.41	1875.64	1800.0	1200.0	1931.42	6873.98
<b>13 anos</b>	18000.0	20000.0	3090.35	2112.41	2200.0	1500.0	2705.94	2608.09
<b>14 anos</b>	28000.0	20000.0	3700.14	2212.65	2500.0	1600.0	3586.95	2181.04
<b>15 anos ou mais</b>	90000.0	60000.0	6070.04	3939.51	4000.0	2800.0	7473.27	4120.78
<b>Não determinados</b>	3000.0	2000.0	1342.91	798.33	1285.0	788.0	903.73	448.49



# Análise outliers Renda

## Boxplot renda

Análise boxplot para a variável renda com transformação de Log

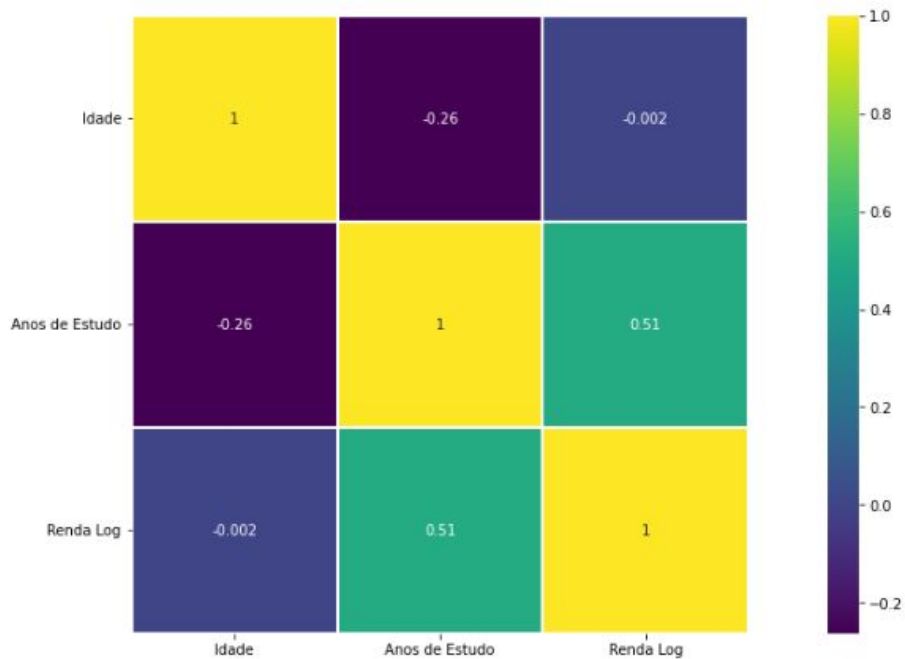




# Análise correlação

## Correlação de Pearson

Análise da correlação entre as variáveis de interesse

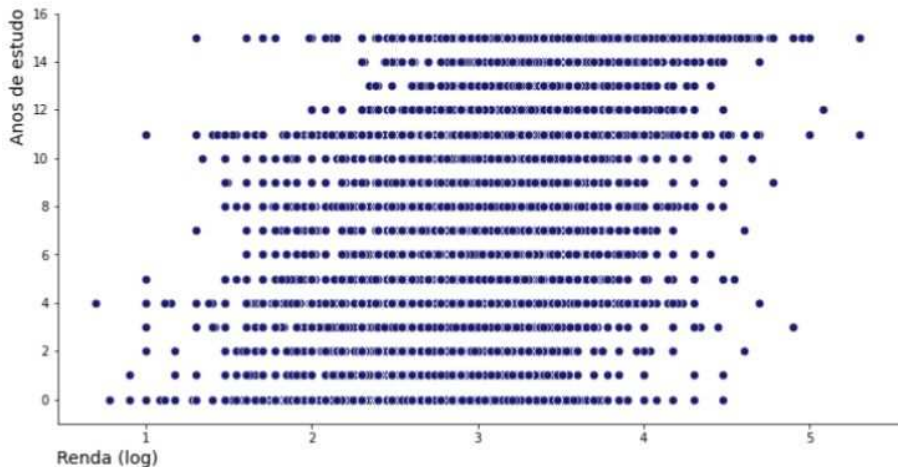




# Análise correlação

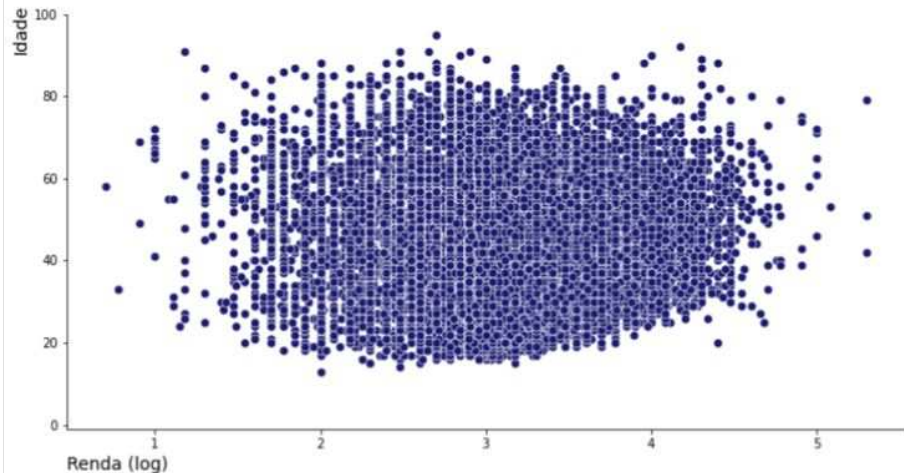
## Renda vs. anos de estudo

Análise da correlação entre a renda transformada e os anos de estudos



## Renda vs. Idade

Análise da correlação entre a renda transformada e idade em anos





# Análise Regressão

## OLS Regression Results

```
=====
Dep. Variable:          Renda      R-squared:                0.157
Model:                 OLS         Adj. R-squared:           0.157
Method:                Least Squares  F-statistic:              7157.
Date:                  Sun, 07 Nov 2021  Prob (F-statistic):       0.00
Time:                  11:13:26      Log-Likelihood:          -7.2554e+05
No. Observations:     76840         AIC:                     1.451e+06
Df Residuals:         76837         BIC:                     1.451e+06
Df Model:              2
Covariance Type:      nonrobust
=====
```



# Análise Regressão

	coef	std err	t	P> t	[0.025	0.975]
const	-374.3902	29.174	-12.833	0.000	-431.572	-317.208
Idade	0.3309	0.010	34.255	0.000	0.312	0.350
Anos de Estudo	18.3249	0.154	119.286	0.000	18.024	18.626



# Significância conjunta

## OLS Regression Results

```
=====
Dep. Variable:          Renda      R-squared:                0.157
Model:                  OLS        Adj. R-squared:           0.157
Method:                 Least Squares  F-statistic:              7157.
Date:                   Sun, 07 Nov 2021  Prob (F-statistic):       0.00
Time:                   11:13:26      Log-Likelihood:          -7.2554e+05
No. Observations:      76840         AIC:                     1.451e+06
Df Residuals:          76837         BIC:                     1.451e+06
Df Model:               2
Covariance Type:       nonrobust
=====
```



# Significância individual

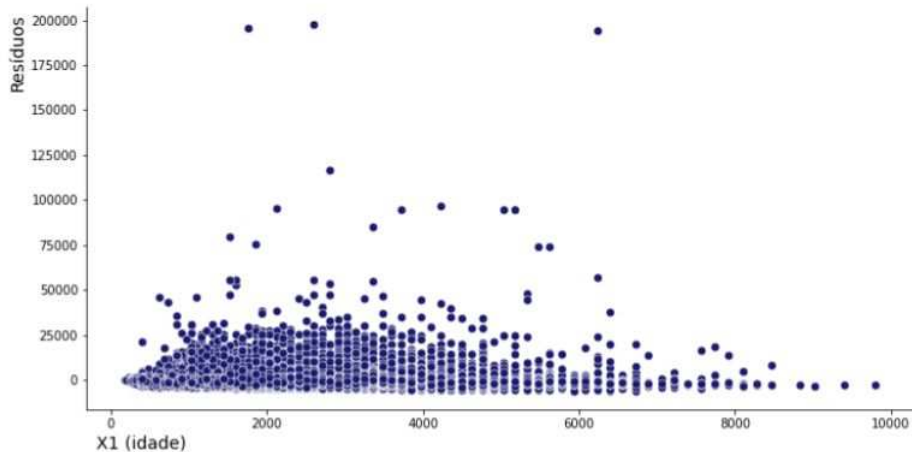
	coef	std err	t	P> t	[0.025	0.975]
const	-374.3902	29.174	-12.833	0.000	-431.572	-317.208
Idade	0.3309	0.010	34.255	0.000	0.312	0.350
Anos de Estudo	18.3249	0.154	119.286	0.000	18.024	18.626



# Variância resíduos

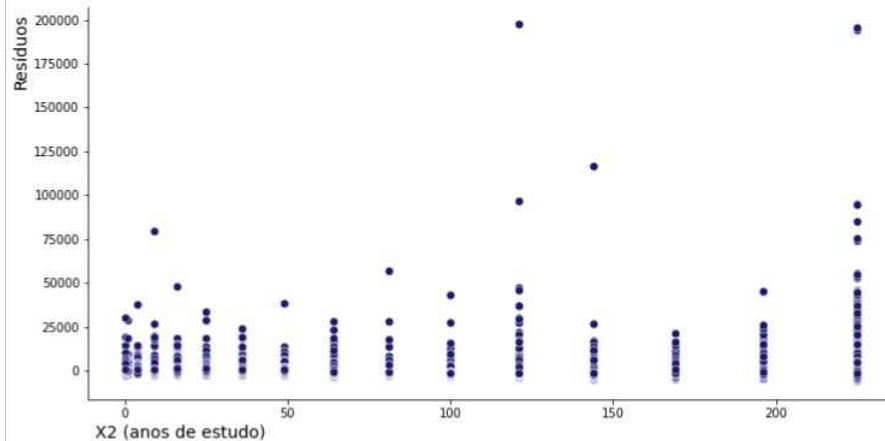
## Resíduos vs. X

Análise dos resíduos do modelo contra as variáveis independentes



## Resíduos vs. X

Análise dos resíduos do modelo contra as variáveis independentes

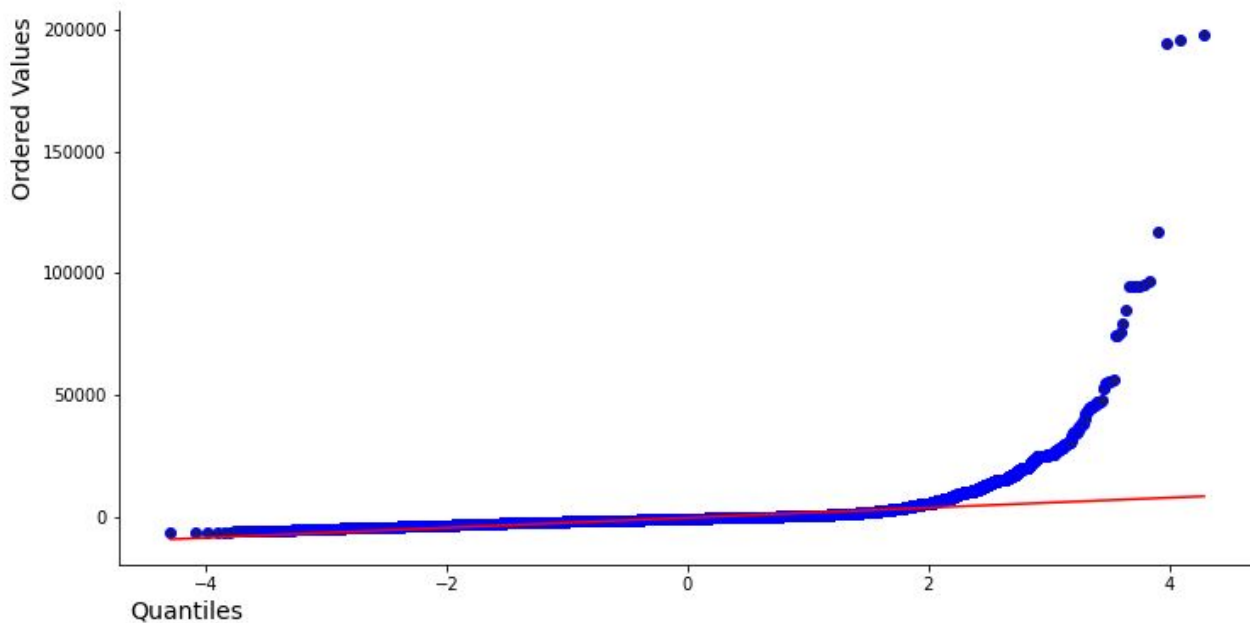




# Normalidade resíduos

## Q-QPlot dos resíduos

Análise da normalidade dos resíduos





## Considerações **finais**

A análise da regressão demonstrou que as suposições de linearidade e de variância constante não foram atendidas.



---

# Boa noite!

Obrigada!