

Relatório de Análise de Custos Médicos

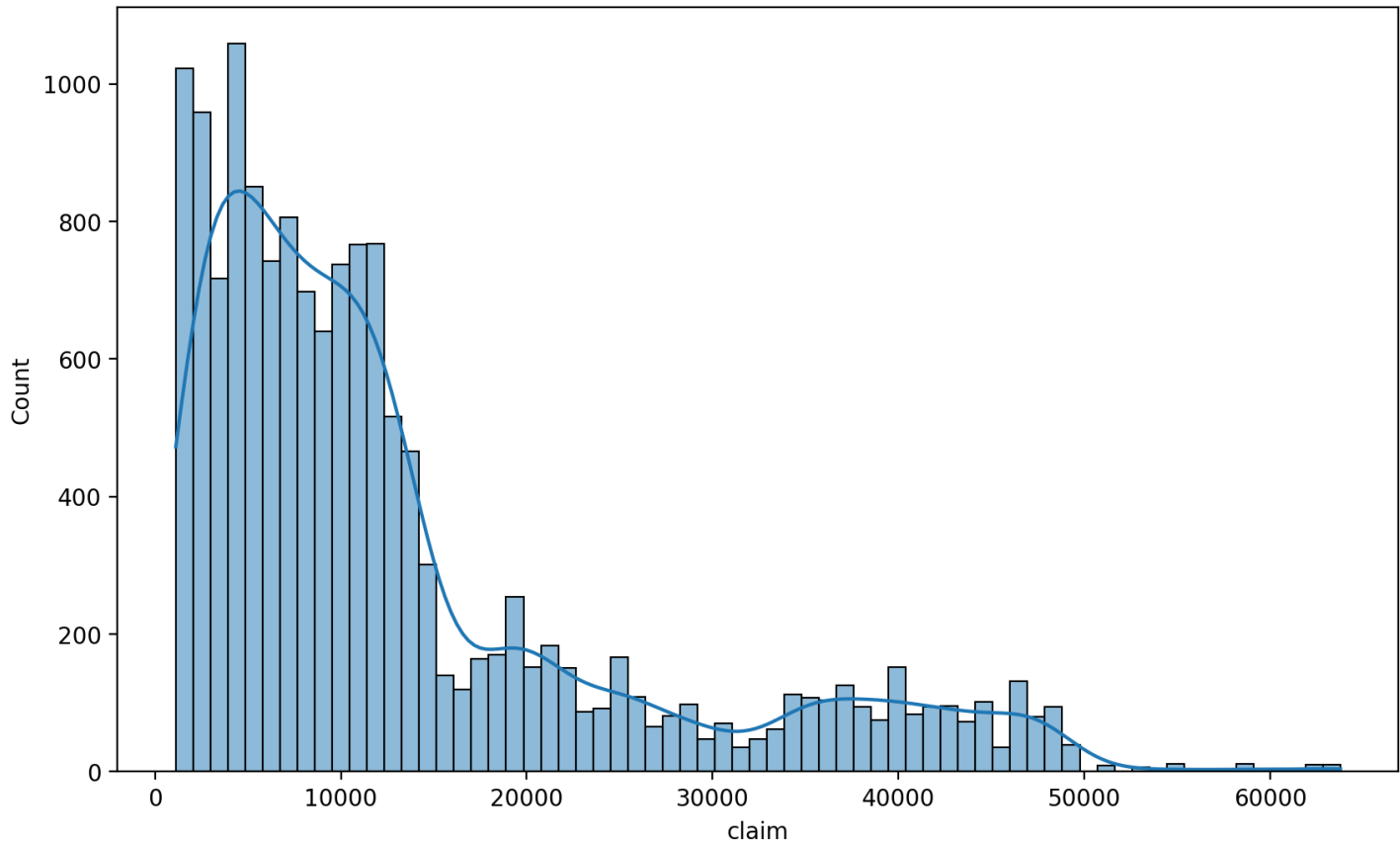
Data de geração: 02/06/2025 19:14:03

Análise de Custos Médicos Individuais

Distribuição dos custos médicos

Distribuição dos custos médicos

Distribuição dos custos médicos

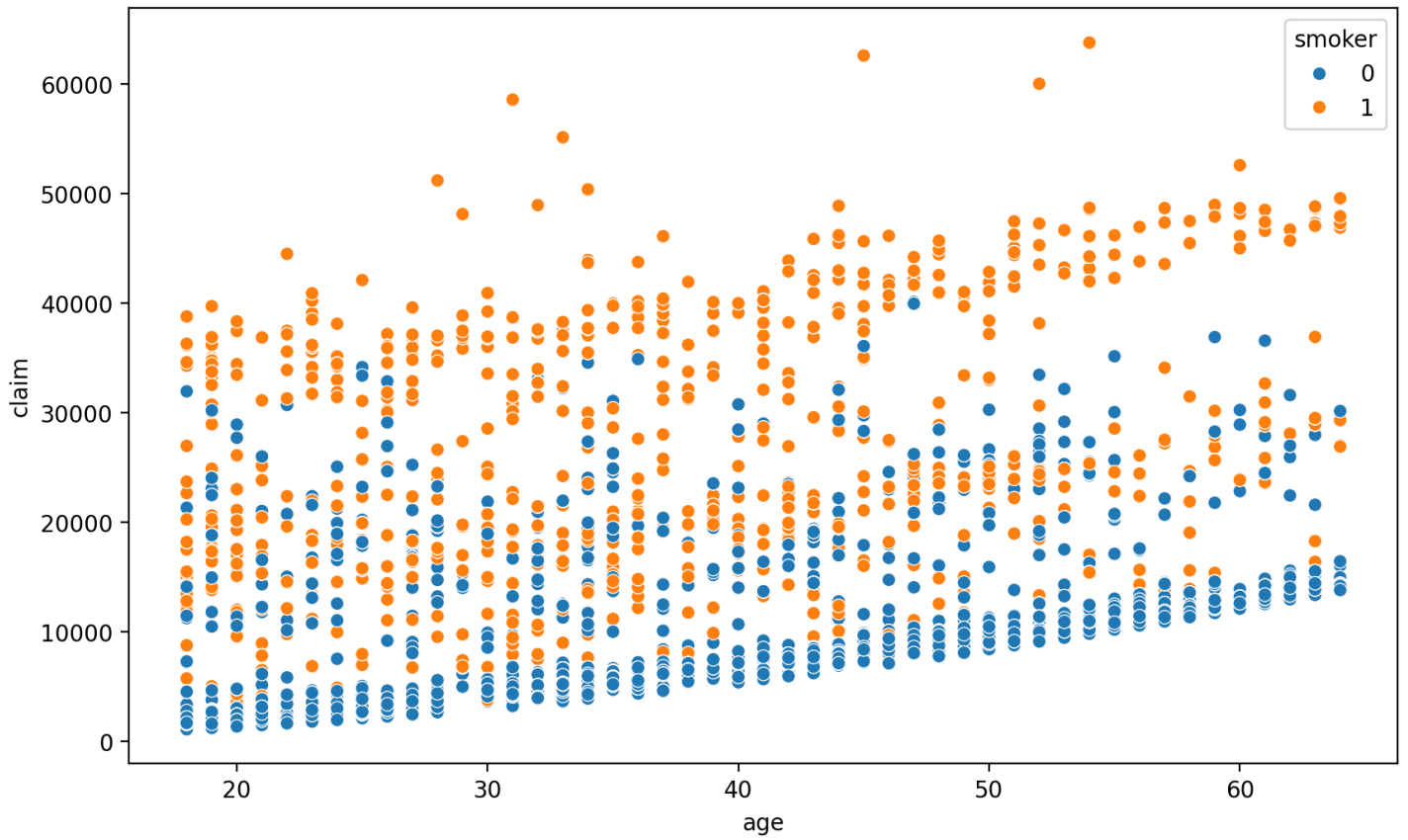


Path to dataset files:
C:\Users\niloa\.cache\kagglehub\datasets\sureshgupta\health-insurance-data-set\versions\1

Idade x Custo (fumantes)

Idade x Custo (fumantes)

Idade x Custo (fumantes)

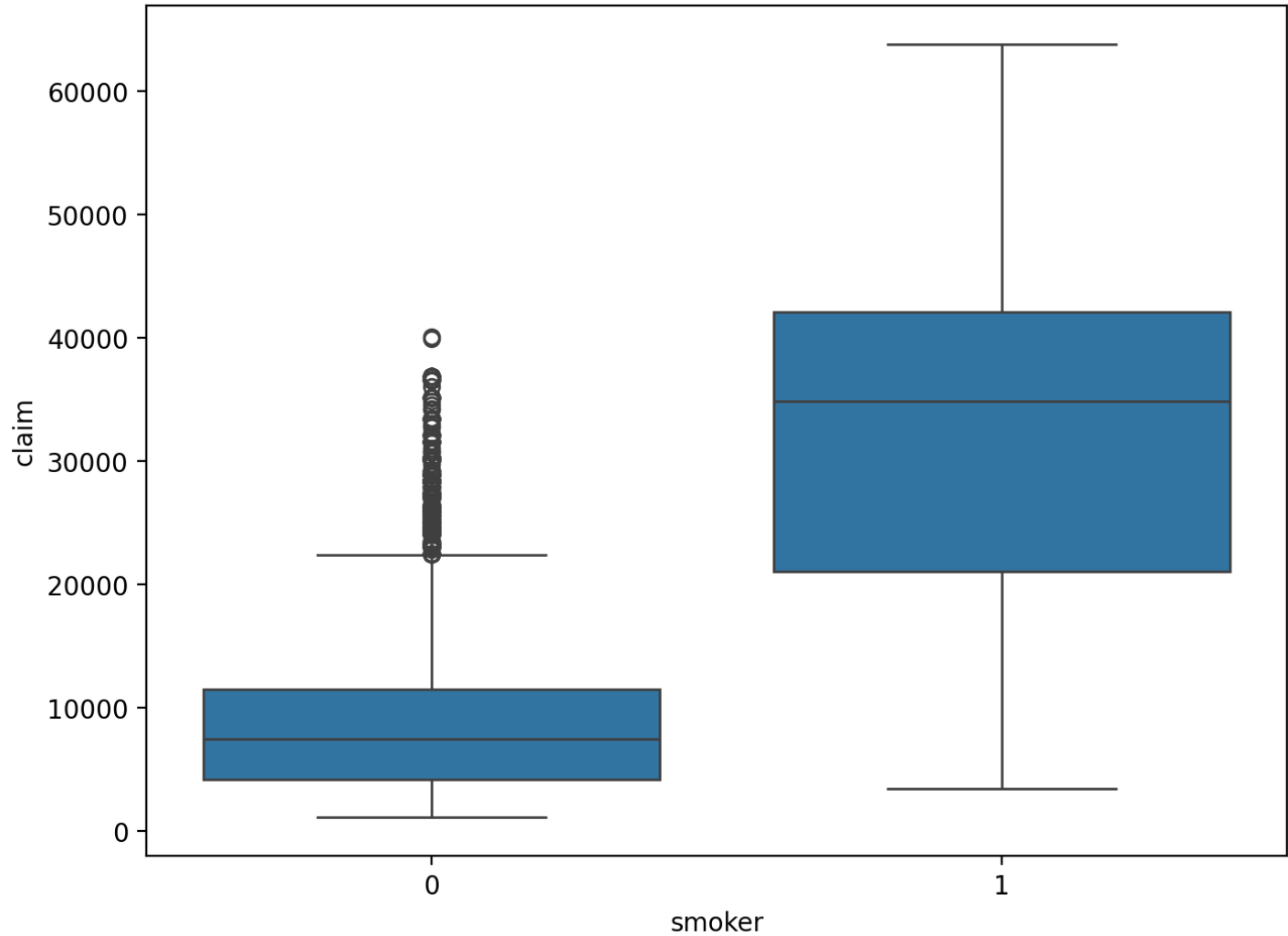


Arquivos baixados: ['1651277648862_healthinsurance.csv']

Fumante x Custo

Fumante x Custo

Fumante x Custo



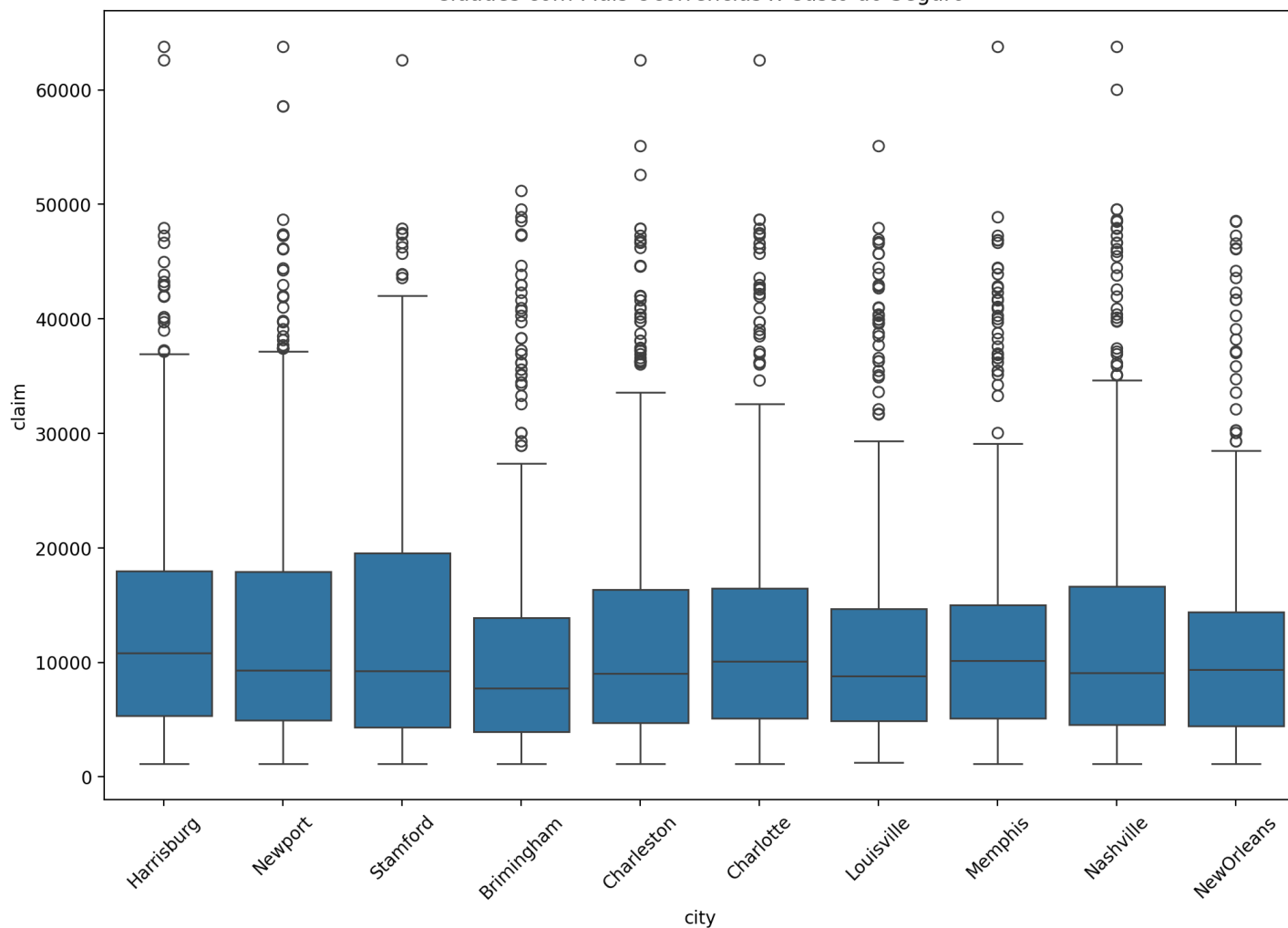
Primeiras linhas do DataFrame:

| | age | sex | weight | bmi | hereditary_diseases | no_of_dependents | smoker | city | bloodpressure |
|----------|------|--------|---------|------|---------------------|------------------|--------|-------------|---------------|
| diabetes | | | | | | | | | |
| 0 | 60.0 | male | 64 | 24.3 | NoDisease | 1 | 0 | NewYork | 72 |
| 0 | | 0 | 13112.6 | | | | | | |
| 1 | 49.0 | female | 75 | 22.6 | NoDisease | 1 | 0 | Boston | 78 |
| 1 | | 1 | 9567.0 | | | | | | |
| 2 | 32.0 | female | 64 | 17.8 | Epilepsy | 2 | 1 | Phildelphia | 88 |
| 1 | | 1 | 32734.2 | | | | | | |
| 3 | 61.0 | female | 53 | 36.4 | NoDisease | 1 | 1 | Pittsburg | 72 |
| 1 | | 0 | 48517.6 | | | | | | |
| 4 | 19.0 | female | 50 | 20.6 | NoDisease | 0 | 0 | Buffalo | 82 |
| 1 | | 0 | 1731.7 | | | | | | |

Cidades com Mais Ocorrências x Custo do Seguro

Cidades com Mais Ocorrências x Custo do Seguro

Cidades com Mais Ocorrências x Custo do Seguro



Informações do DataFrame:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 15000 entries, 0 to 14999

Data columns (total 12 columns):

| # | Column | Non-Null Count | Dtype |
|----|---------------------|----------------|---------|
| 0 | age | 14604 non-null | float64 |
| 1 | sex | 15000 non-null | object |
| 2 | weight | 15000 non-null | int64 |
| 3 | bmi | 14044 non-null | float64 |
| 4 | hereditary_diseases | 15000 non-null | object |
| 5 | no_of_dependents | 15000 non-null | int64 |
| 6 | smoker | 15000 non-null | int64 |
| 7 | city | 15000 non-null | object |
| 8 | bloodpressure | 15000 non-null | int64 |
| 9 | diabetes | 15000 non-null | int64 |
| 10 | regular_ex | 15000 non-null | int64 |
| 11 | claim | 15000 non-null | float64 |

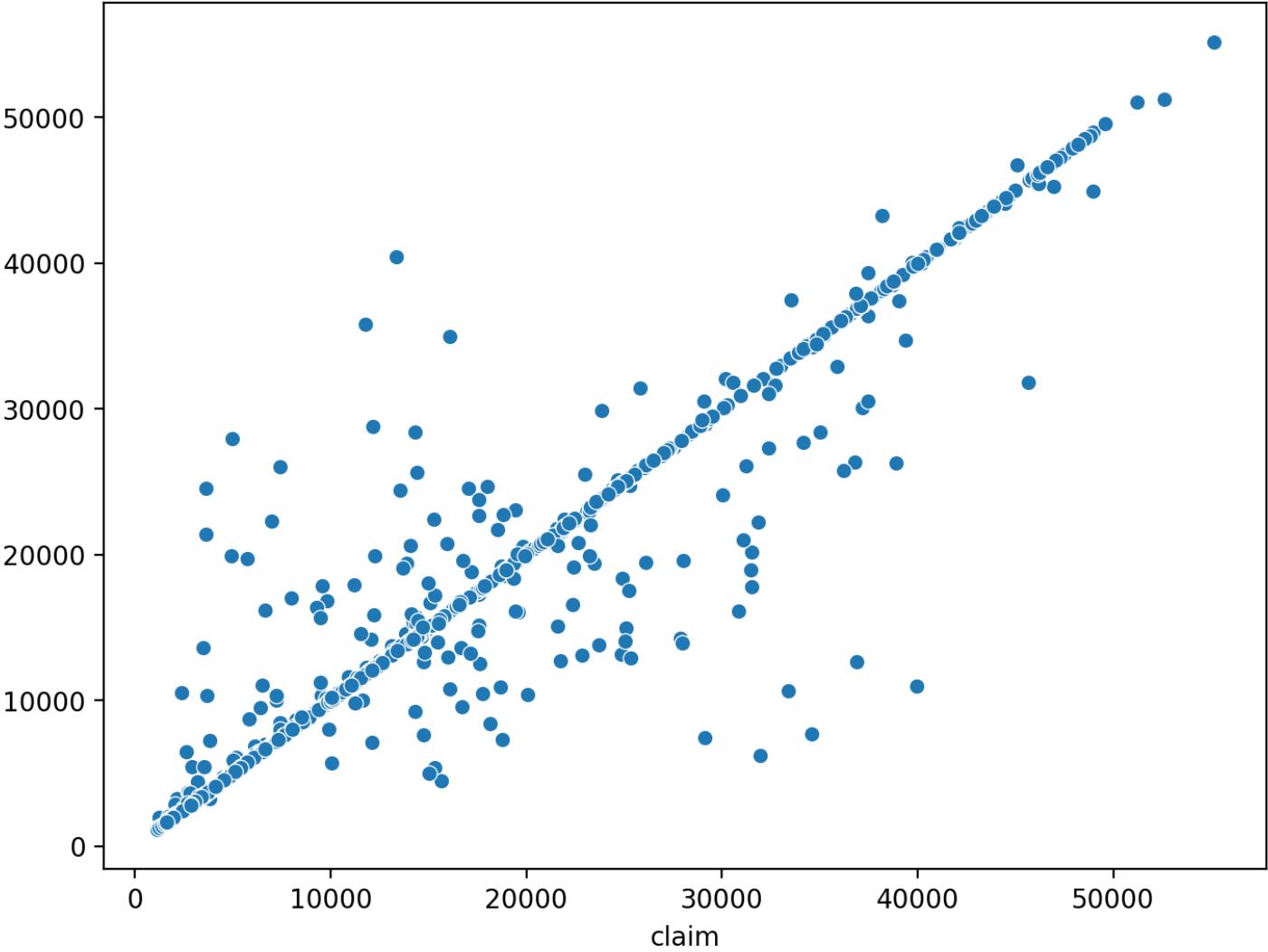
dtypes: float64(3), int64(6), object(3)

memory usage: 1.4+ MB

Previsão x Real

Previsão x Real

Previsão x Real

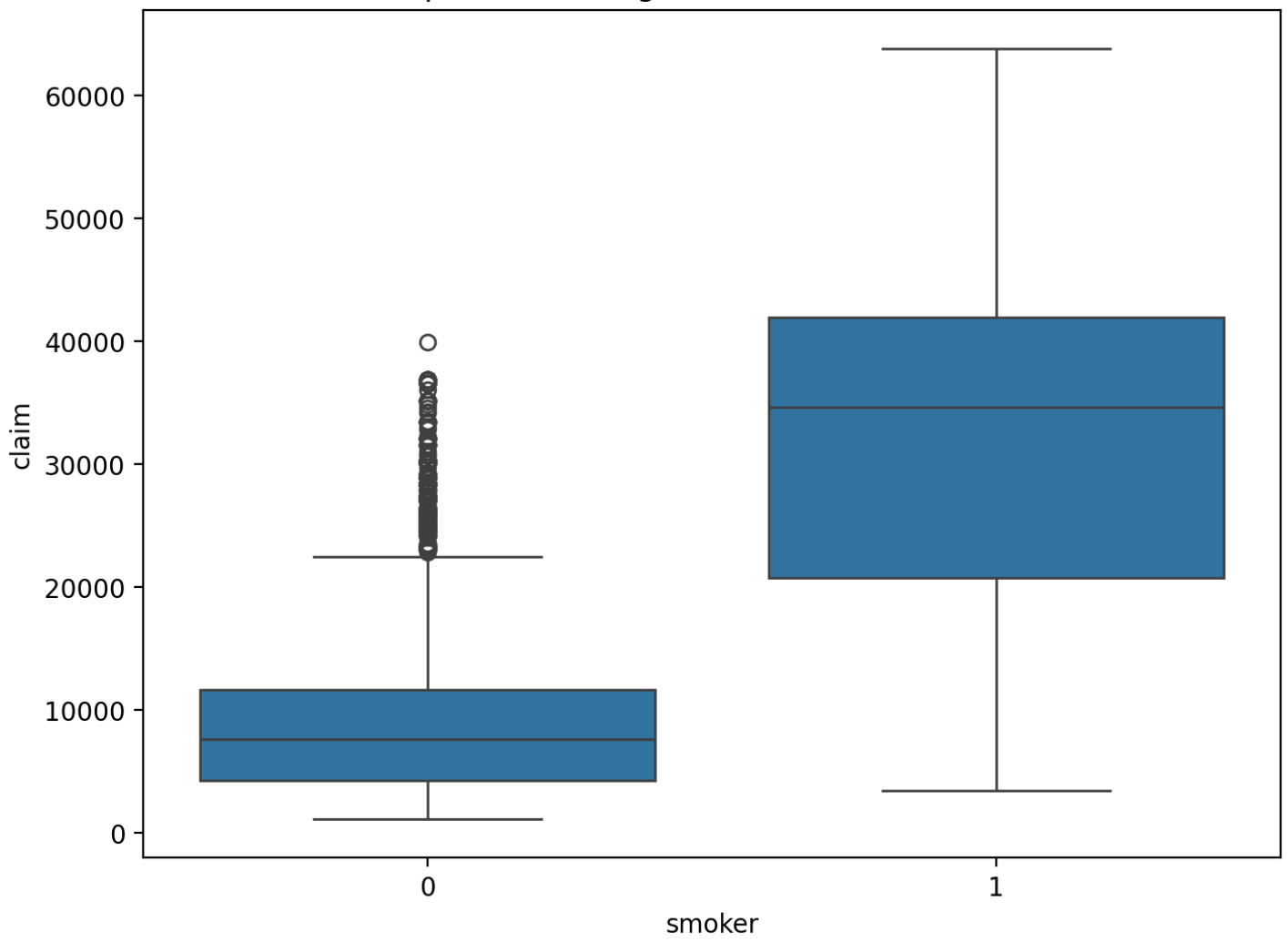


| | | | | | | | |
|---------------------------|--------------|--------------|--------------|------------------|--------------|---------------|--------------|
| Estatísticas descritivas: | | | | | | | |
| | age | weight | bmi | no_of_dependents | smoker | bloodpressure | diabetes |
| regular_ex | claim | | | | | | |
| count | 14604.000000 | 15000.000000 | 14044.000000 | 15000.000000 | 15000.000000 | 15000.000000 | 15000.000000 |
| 15000.000000 | 15000.000000 | | | | | | |
| mean | 39.547521 | 64.909600 | 30.266413 | 1.129733 | 0.198133 | 68.650133 | 0.777000 |
| 0.224133 | 13401.437620 | | | | | | |
| std | 14.015966 | 13.701935 | 6.122950 | 1.228469 | 0.398606 | 19.418515 | 0.416272 |
| 0.417024 | 12148.239619 | | | | | | |
| min | 18.000000 | 34.000000 | 16.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.000000 | 1121.900000 | | | | | | |
| 25% | 27.000000 | 54.000000 | 25.700000 | 0.000000 | 0.000000 | 64.000000 | 1.000000 |
| 0.000000 | 4846.900000 | | | | | | |
| 50% | 40.000000 | 63.000000 | 29.400000 | 1.000000 | 0.000000 | 71.000000 | 1.000000 |
| 0.000000 | 9545.650000 | | | | | | |
| 75% | 52.000000 | 76.000000 | 34.400000 | 2.000000 | 0.000000 | 80.000000 | 1.000000 |
| 0.000000 | 16519.125000 | | | | | | |
| max | 64.000000 | 95.000000 | 53.100000 | 5.000000 | 1.000000 | 122.000000 | 1.000000 |
| 1.000000 | 63770.400000 | | | | | | |

Impacto do tabagismo nos custos médicos

Impacto do tabagismo nos custos médicos

Impacto do tabagismo nos custos médicos

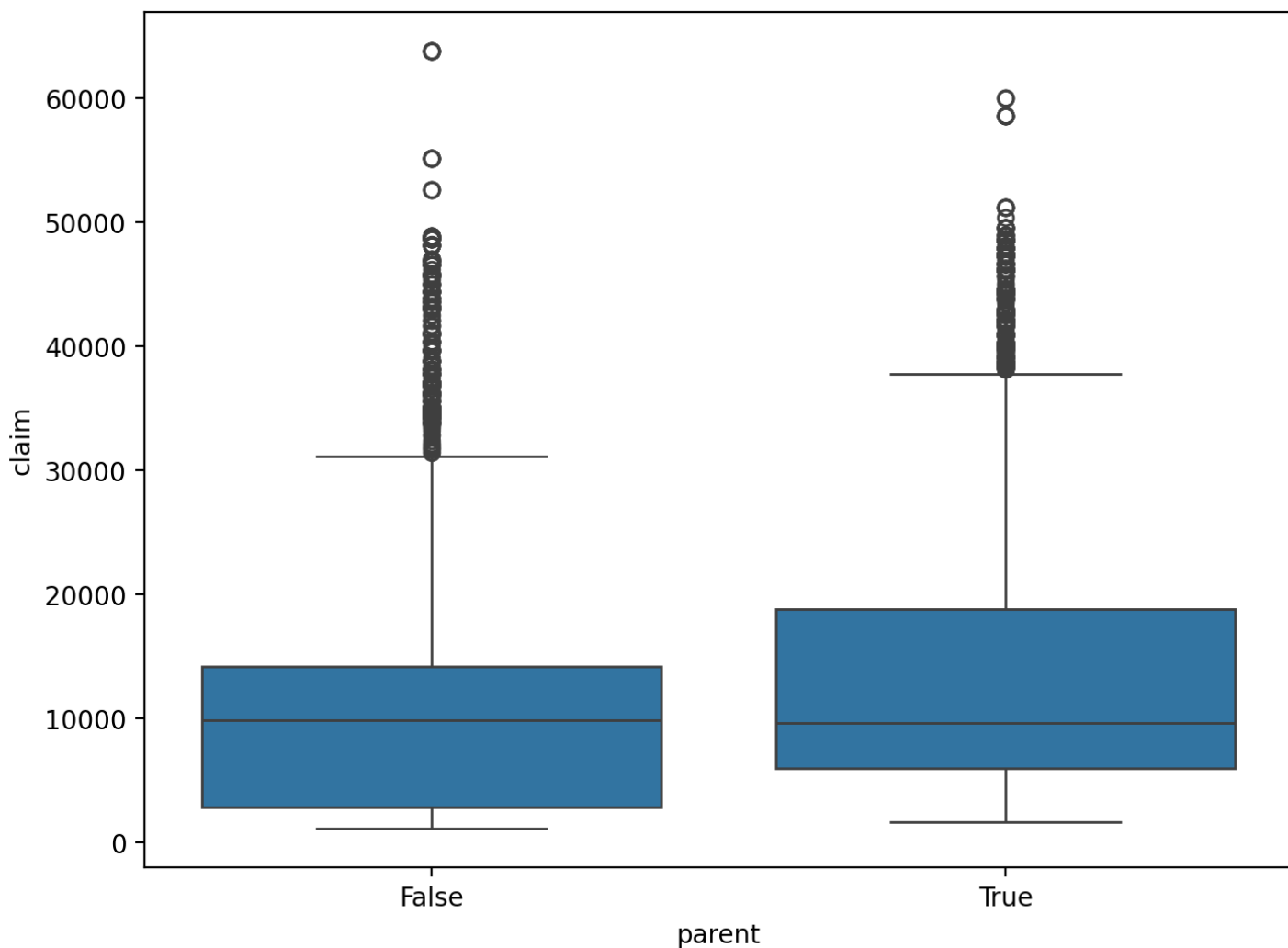


R²: 0.9609072515534081
MAE: 482.1655767914533
RMSE: 2377.8222727712287

Custos: Com filhos vs Sem filhos

Custos: Com filhos vs Sem filhos

Custos: Com filhos vs Sem filhos



OLS Regression Results

| | | | | | | |
|-------------------|------------------|---------------------|-------------|-------|----------|----------|
| Dep. Variable: | claim | R-squared: | 0.716 | | | |
| Model: | OLS | Adj. R-squared: | 0.716 | | | |
| Method: | Least Squares | F-statistic: | 2833. | | | |
| Date: | Mon, 02 Jun 2025 | Prob (F-statistic): | 0.00 | | | |
| Time: | 19:13:59 | Log-Likelihood: | -1.0303e+05 | | | |
| No. Observations: | 10113 | AIC: | 2.061e+05 | | | |
| Df Residuals: | 10103 | BIC: | 2.062e+05 | | | |
| Df Model: | 9 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| const | 1.354e+04 | 90.044 | 150.415 | 0.000 | 1.34e+04 | 1.37e+04 |
| age | 3734.0438 | 67.735 | 55.127 | 0.000 | 3601.269 | 3866.818 |
| weight | -649.5132 | 70.003 | -9.278 | 0.000 | -786.733 | -512.293 |
| bmi | 1694.6359 | 67.036 | 25.279 | 0.000 | 1563.231 | 1826.040 |
| no_of_dependents | 502.7146 | 64.796 | 7.758 | 0.000 | 375.701 | 629.728 |
| smoker | 9208.5277 | 64.382 | 143.030 | 0.000 | 9082.326 | 9334.729 |
| bloodpressure | 220.5152 | 64.139 | 3.438 | 0.001 | 94.791 | 346.240 |
| diabetes | 641.3938 | 64.587 | 9.931 | 0.000 | 514.790 | 767.998 |
| regular_ex | -485.8539 | 64.204 | -7.567 | 0.000 | -611.706 | -360.001 |
| sex_male | -304.6471 | 130.555 | -2.333 | 0.020 | -560.560 | -48.734 |
| ===== | | | | | | |
| Omnibus: | 1751.448 | Durbin-Watson: | 1.989 | | | |

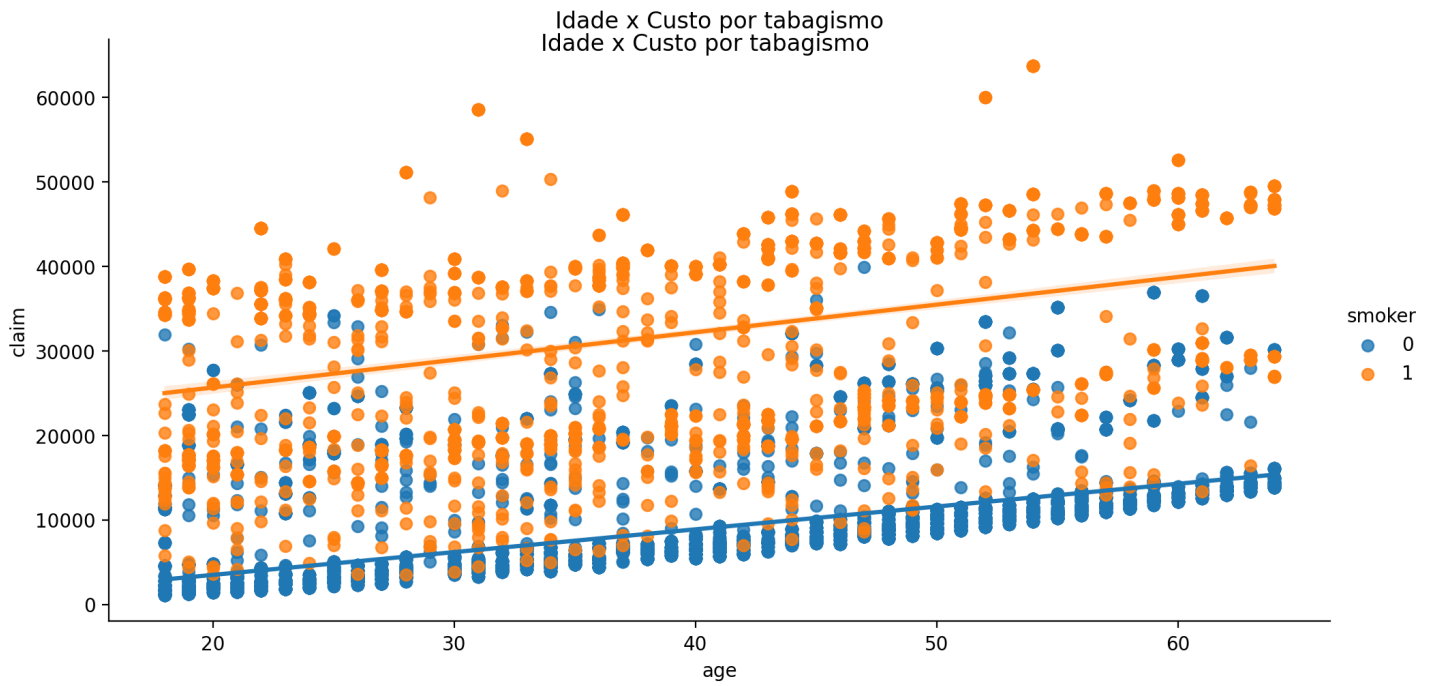
| | | | |
|----------------|-------|-------------------|----------|
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 4197.697 |
| Skew: | 0.978 | Prob(JB): | 0.00 |
| Kurtosis: | 5.477 | Cond. No. | 2.91 |

=====

Notes:

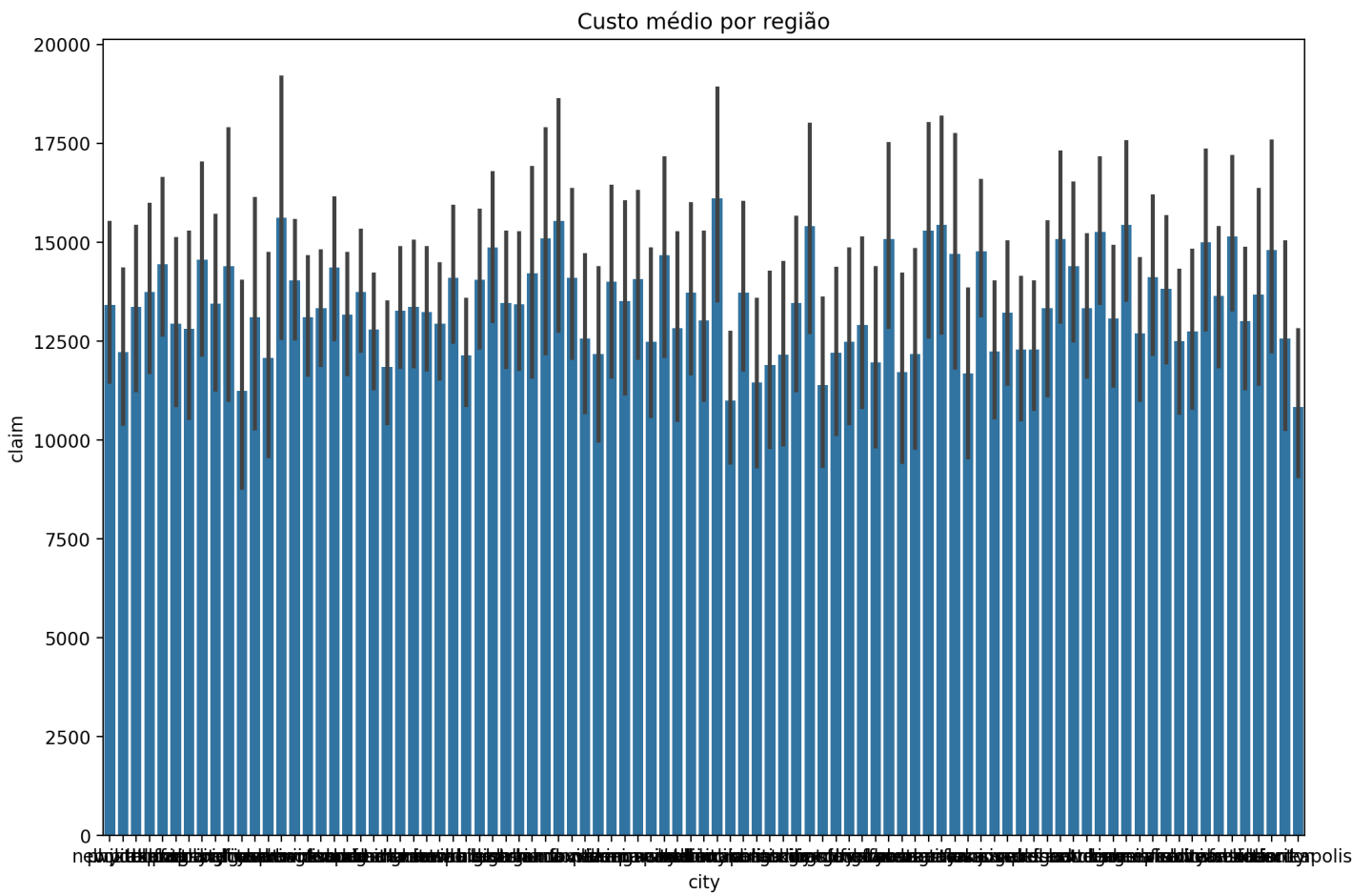
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Idade x Custo por tabagismo



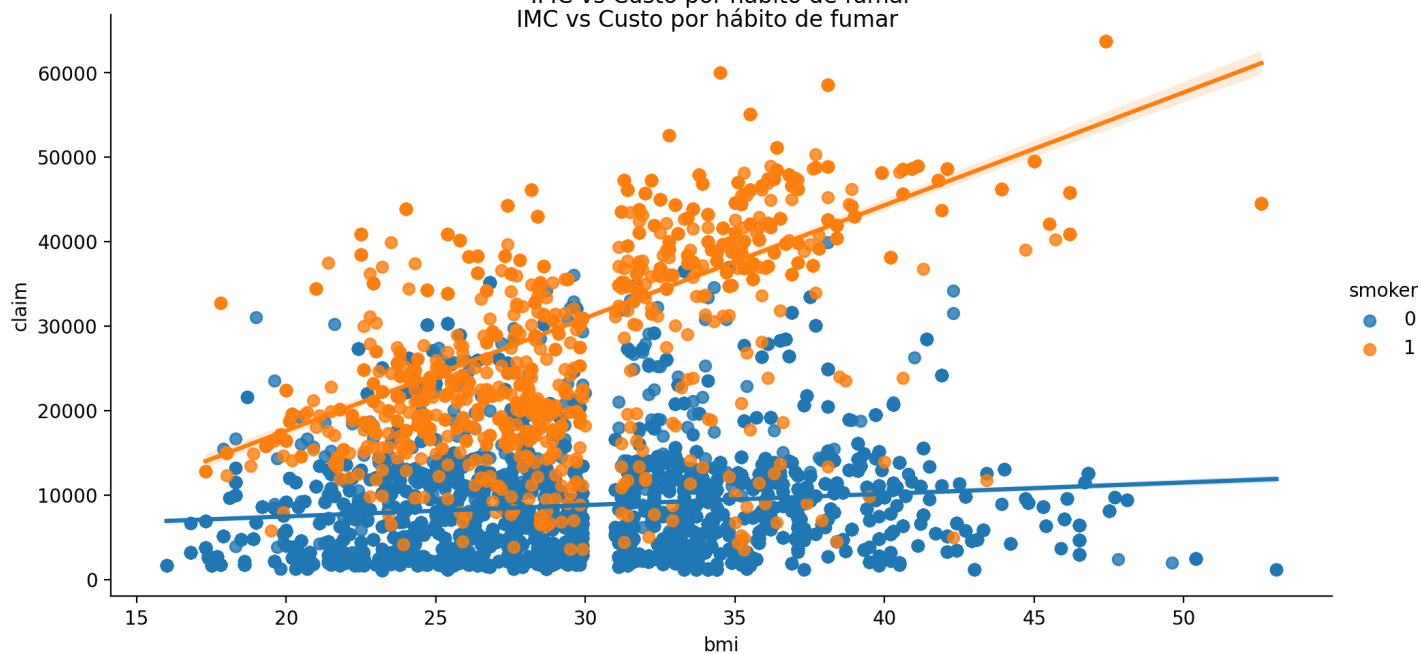
Custo médio por região

Custo médio por região



IMC vs Custo por hábito de fumar

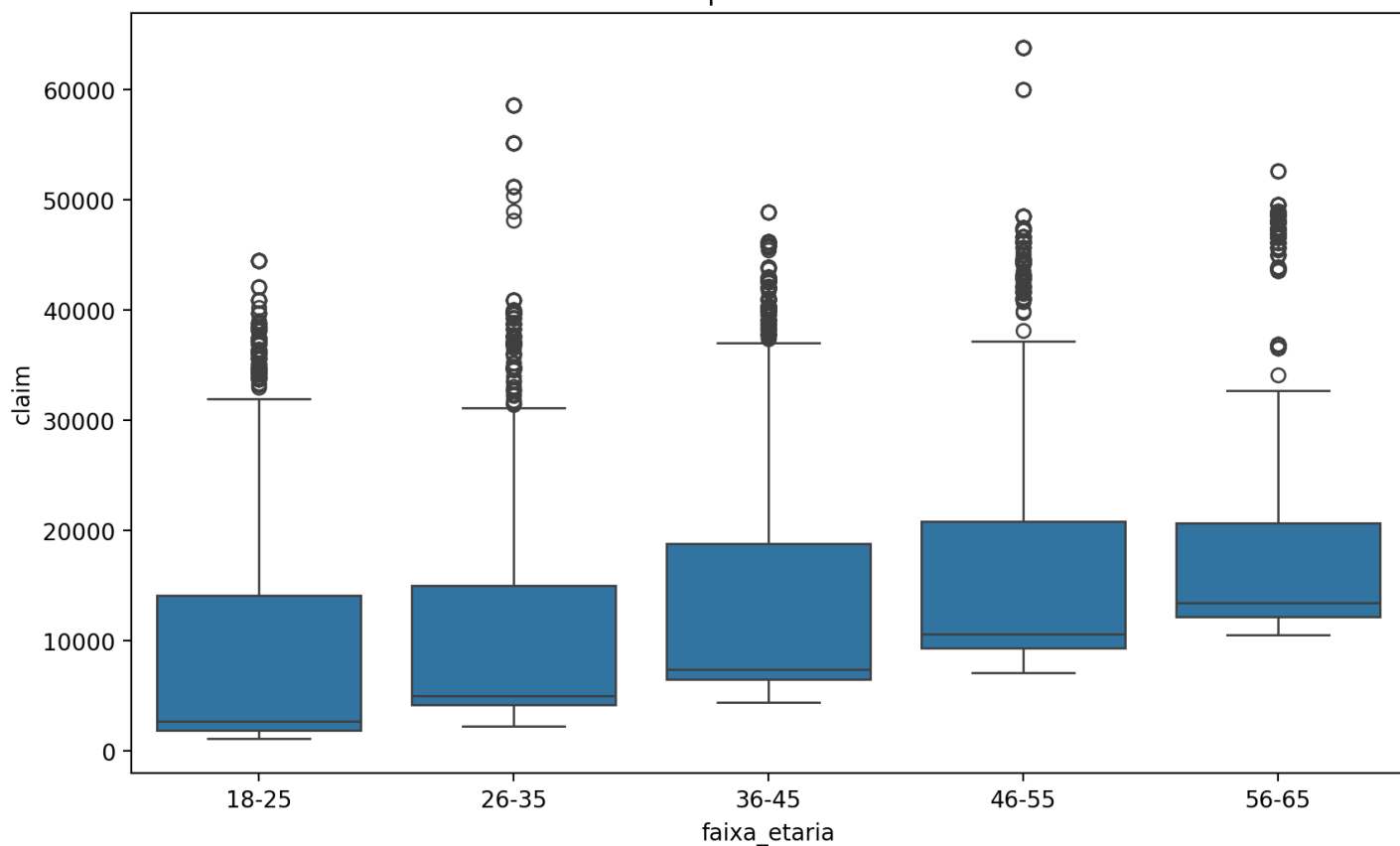
IMC vs Custo por hábito de fumar
IMC vs Custo por hábito de fumar



Custo por faixa etária

Custo por faixa etária

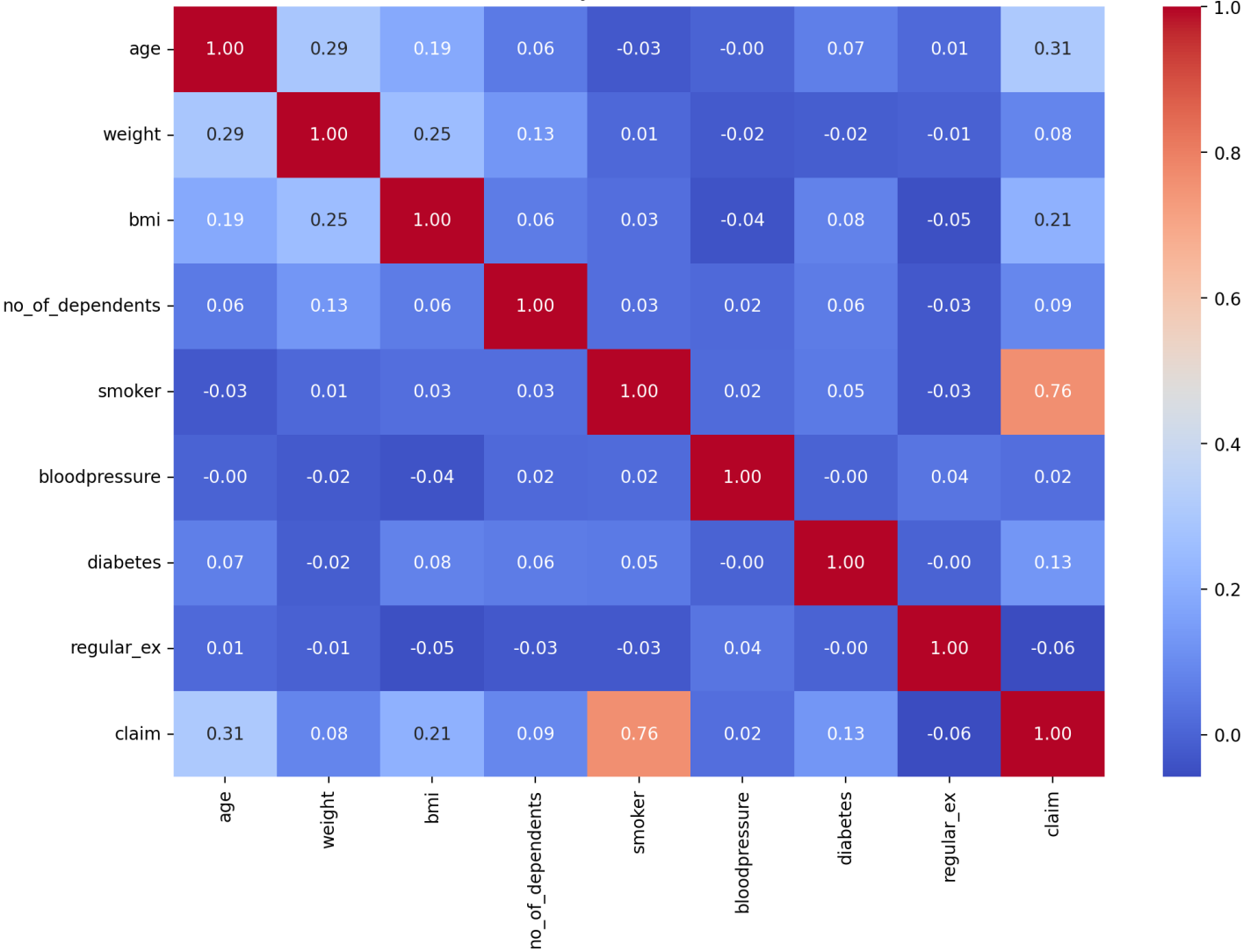
Custo por faixa etária



Correlação entre variáveis

Correlação entre variáveis

Correlação entre variáveis



Importância das variáveis no modelo

Importância das variáveis no modelo

