

/kaggle/input/insurance/insurance.csv

+ Markdown

[67]:	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

[69]:	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	0	26.340	0	0	1	1716.69920
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520

```
[70]: <AxesSubplot:>
```



```
## we see no correlation
```

```
[72]: features.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  --
 0   age         1338 non-null   int64   
 1   sex         1338 non-null   int64   
 2   bmi         1338 non-null   float64  
 3   children    1338 non-null   int64   
 4   smoker      1338 non-null   int64   
 5   region      1338 non-null   int64   
 6   charges     1338 non-null   float64  
dtypes: float64(2), int64(5)
memory usage: 73.3 KB
```

Linear regression

```
[73]: from sklearn.linear_model import LinearRegression
      from sklearn.pipeline import Pipeline, make_pipeline
      from sklearn.linear_model import Ridge

      model = LinearRegression()
```

```
[74]: features.head()
```

```
[74]:
```

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520

```
[75]: target=features["charges"]
      target.head()
```

```
[75]: 0    16884.92400
      1     1725.55230
      2     4449.46200
      3    21984.47061
      4     3866.85520
      Name: charges, dtype: float64
```

```
[76]: lr=model.fit(features, target)
```

```
[77]: predictions = model.predict(features)
      predictions = pd.Series(predictions)
      predictions.head()
```

```
[77]: 0    16884.92400
      1     1725.55230
      2     4449.46200
      3    21984.47061
      4     3866.85520
      dtype: float64
```

```
[78]: from sklearn.metrics import mean_absolute_error as mae
      print("Mean Absolute Error:", mae(target, predictions))
```

```
Mean Absolute Error: 5.47566113708914e-12
```

```
[ ]:
```