

Course: Introduction to Data Science (DS2006) - Laboratory 15

Task 1:

The first line in the insurance dataset describes the different data metrics collected. These metrics include each person's age, sex, body mass index, whether they smoke, region, and charges.

Task 2:

Line 2 until the last line contains insurance records structured by the metrics aforementioned, allowing streamlined analysis and monitoring of each person's insurance records.

Task 3: insurance.csv loaded

Task 4:

	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

Task 5: dataframe groups created

Task 6:

Features:

	age	sex	bmi	children	smoker	region
0	19	female	27.900	0	yes	southwest
1	18	male	33.770	1	no	southeast
2	28	male	33.000	3	no	southeast
3	33	male	22.705	0	no	northwest
4	32	male	28.880	0	no	northwest

Targets:

0	16884.92400
1	1725.55230
2	4449.46200
3	21984.47061
4	3866.85520

Name: charges, dtype: float64

Task 7: training and test data created for features and targets

Task 8: kNN regressor implemented

Task 9: model prediction implemented

Task 10:

```
ValueError: could not convert string to float: 'female'
```

Task 11: results where k=1

```
MAE: 6963.500  
MSE: 84459448.088  
RMSE: 9190.182
```

Task 12:

Value used for K	MAE	MSE	RMSE
2	6887.864	74221317.090	8615.179
3	6753.261	69883032.101	8359.607
5	6553.275	68106295.197	8252.654
10	6206.234	62057303.952	7877.646
30	5733.858	57193730.998	7562.654

Task 13: results where k=30

```
MAE: 4125.166  
MSE: 42667303.045  
RMSE: 6532.021
```

Task 14:

Value used for K	MAE	MSE	RMSE
2	3413.106	36120293.501	6010.016
3	3395.857	31864834.043	5644.895
5	3599.651	33220062.497	5763.685
10	3840.733	37255365.598	6103.717
30	4125.166	42667303.045	6532.021