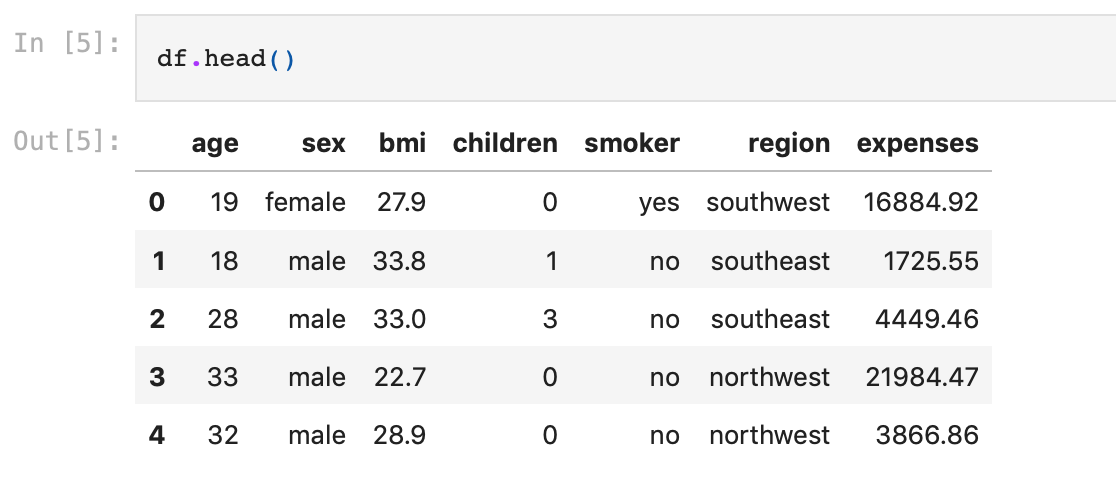
**HEALTH COST PREDICTION USING REGRESSION**

**OBJECTIVE:** To predict health cost for individuals based 6 features available for each person



Data used for training has the below statistical details:



**EXPLORATORY DATA ANALYSIS SUMMARY:**

After performing univariate and bivariate analyses, below results are obtained:

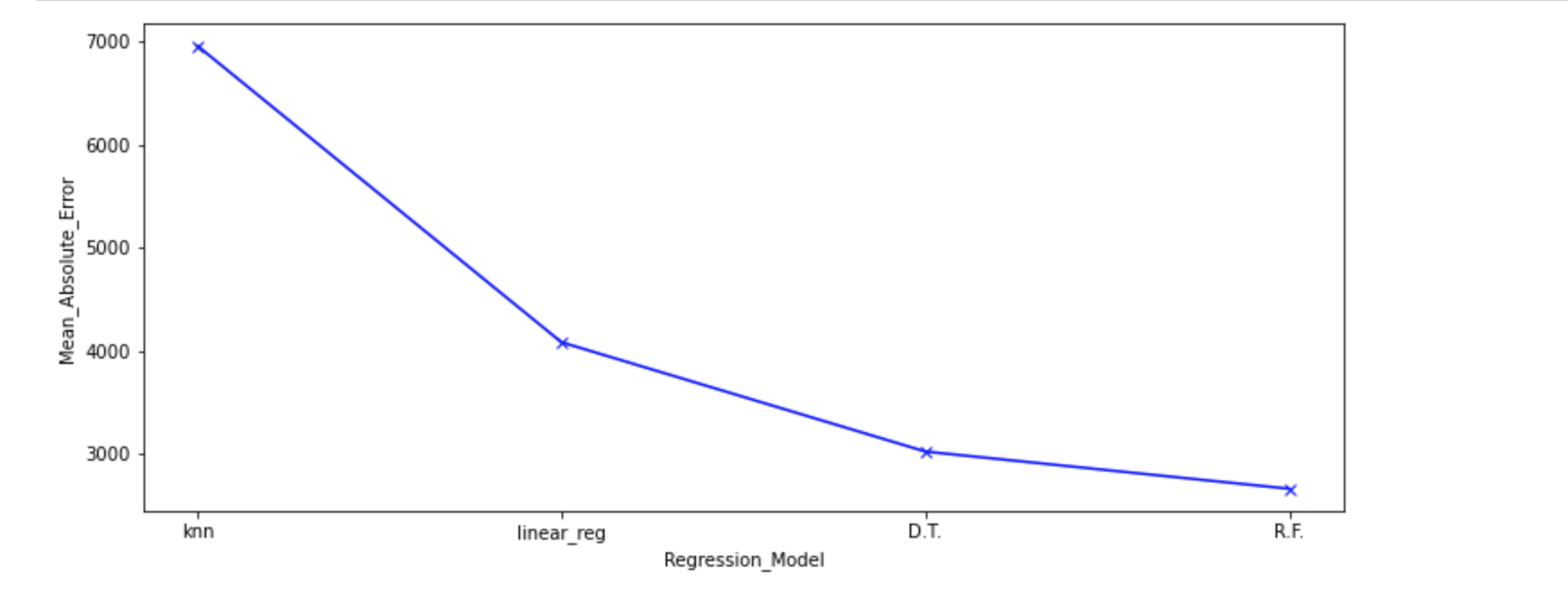
* Age of most number of customers are 18 and 19 years
* Customer count based on gender is similar
* Customer count based on region is similar
* Most of the customers don't have any children
* Most of the customers are non-smokers
* As the age increases, minimum health expenses increases
* As bmi increases, chances of higher health expenses can happen
* As compared to non-smokers, health expenses of smokers are higher
* Health expenses based on gender is similar for both male and female
* Health expenses based on region is similar for all different given regions
* The input variables have very less correlation between each other

**PREPROCESSING:**

* Performed one hot encoding on all categorical features
* Created a new feature based on standard values of bmi

**MODEL CREATION AND RESULT:**

After performing 10-fold cross validated KNN with 3 neighbours, linear regression, decision tree and random forest regression models, the least mean absolute error is obtained for random forest regression model.

****