

# Health Insurance data

This project is designed to investigate patterns and relationships between the total amount of money charged by an insurance company to customers and other factors that might influence those charges. The data was taken from [www.kaggle.com](http://www.kaggle.com) (<http://www.kaggle.com>). I plan on using data visualization to ascertain a level of knowledge about feature behaviour. I then will investigate which features are important whilst also producing regression models for the prediction of charges given the input features.

```
In [3]: import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
# For Notebooks
```

```
In [4]: df = pd.read_csv('D:\Python\Insight\insurance project\insurance.csv')
```

```
In [5]: df.head()
```

Out[5]:

	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

```
In [6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
age          1338 non-null int64
sex          1338 non-null object
bmi          1338 non-null float64
children     1338 non-null int64
smoker       1338 non-null object
region       1338 non-null object
charges      1338 non-null float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.2+ KB
```

```
In [7]: df.describe()
```

```
Out[7]:
```

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

Data shows to be clean (ie no missing values)

## Client gender

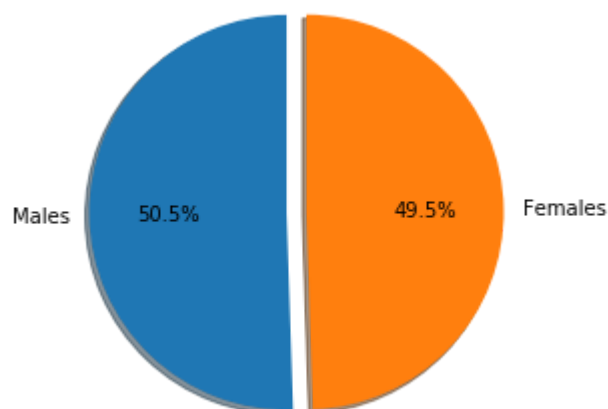
```
In [6]: labels = 'Males', 'Females'

malesn = len((df[df['sex']=='male'])) / len(df)
femalesn = len((df[df['sex']=='female'])) / len(df)

sizes = [malesn, femalesn]
explode = (0, 0.1) # only "explode" the 2nd slice (i.e. 'Hogs')

fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=90)
ax1.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.

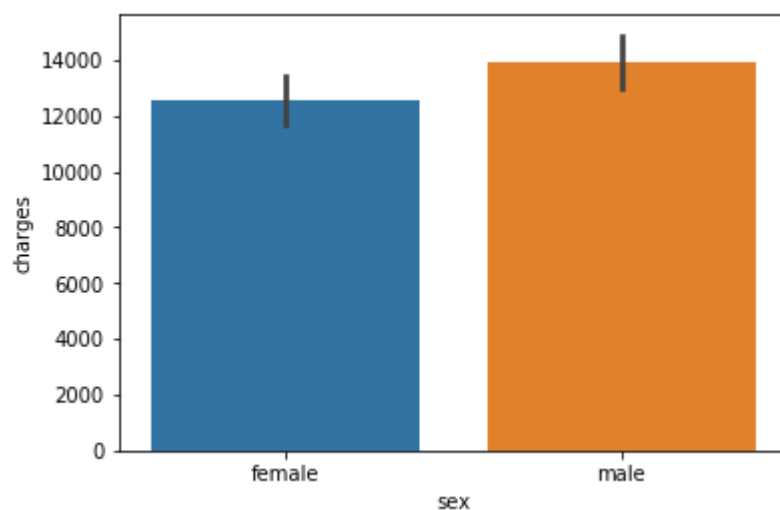
plt.show()
```



There seems to be a close 50-50 split in gender between males and females

```
In [7]: sns.barplot(x='sex',y='charges',data=df)
```

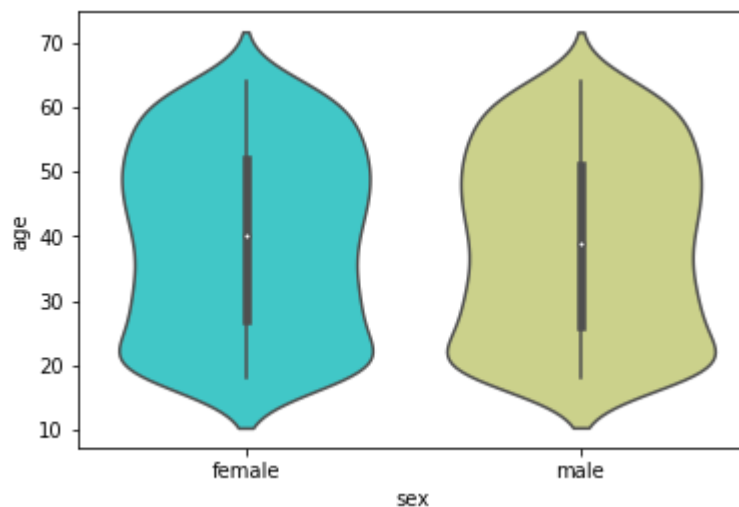
```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x294990236a0>
```



**Males paid more on average in charges than females.**

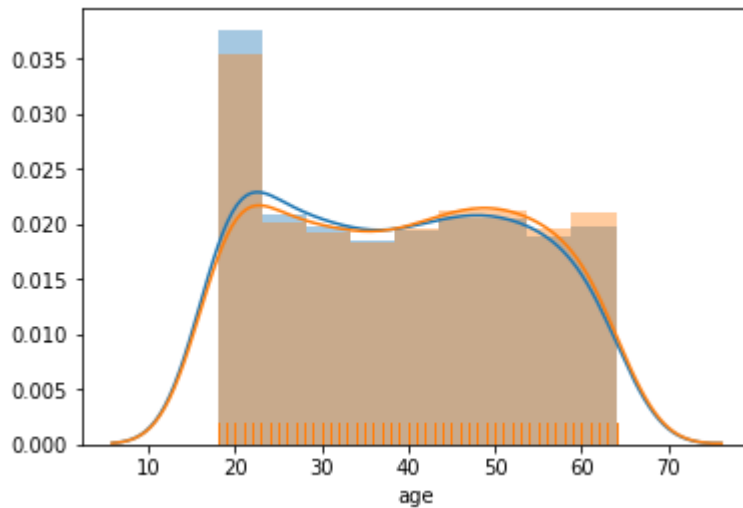
```
In [8]: sns.violinplot(x="sex", y="age", data=df,palette='rainbow')
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x29499016c88>
```



```
In [9]: sns.distplot(df.loc[df['sex'] == 'male', 'age'], rug=True)
sns.distplot(df.loc[df['sex'] == 'female', 'age'], rug=True)
```

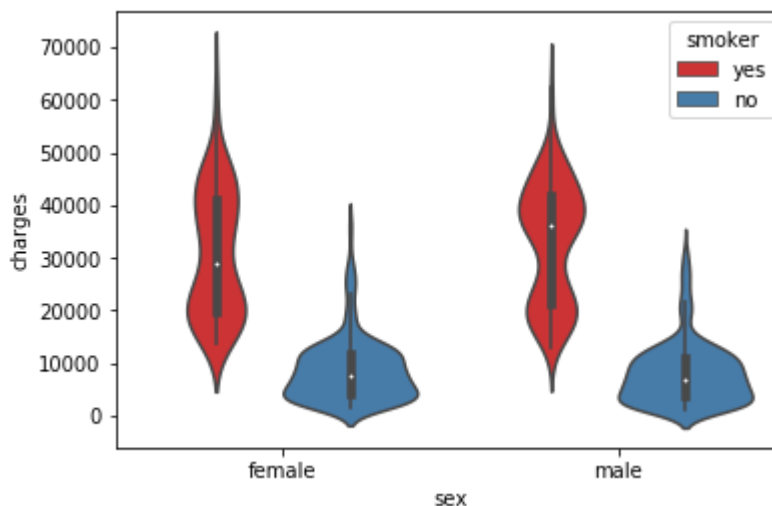
Out[9]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2949b07ae80>



Similar average age between sexes as seen by the above two charts. The distribution is very similar between males and females but is not gaussian and shows a tendency towards a bi-modal distribution with most clients being less than 25 and a second maximum around 50 years of age

```
In [10]: sns.violinplot(x="sex", y="charges", data=df, hue='smoker', palette='Set1')
```

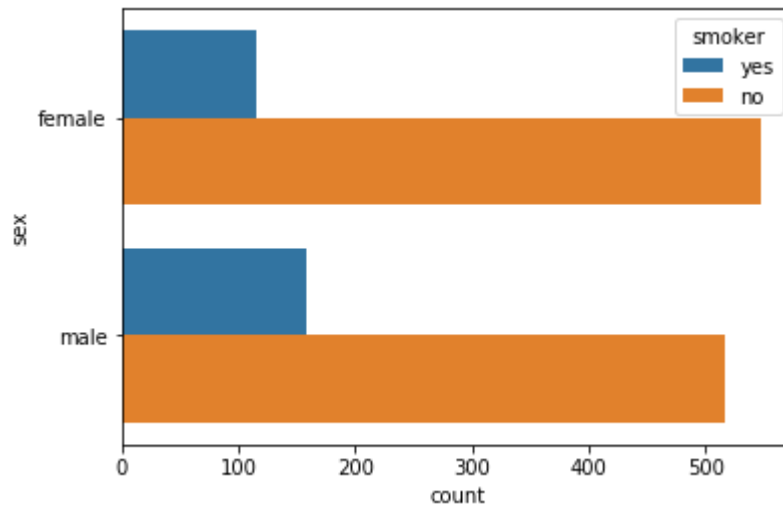
Out[10]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2949c197e48>



Our first inclination of how the relationship between gender and charges with respect to smoking is shown above. It shows regardless of gender, you are more likely to pay higher fees if you smoke compared to if you do not. Furthermore, there are slight differences in the distributions between male and female smokers, where on average males are more likely to pay higher fees if they smoke

## Smokers be warned

```
In [45]: ax = sns.countplot(y="sex", hue="smoker", data=df)
```

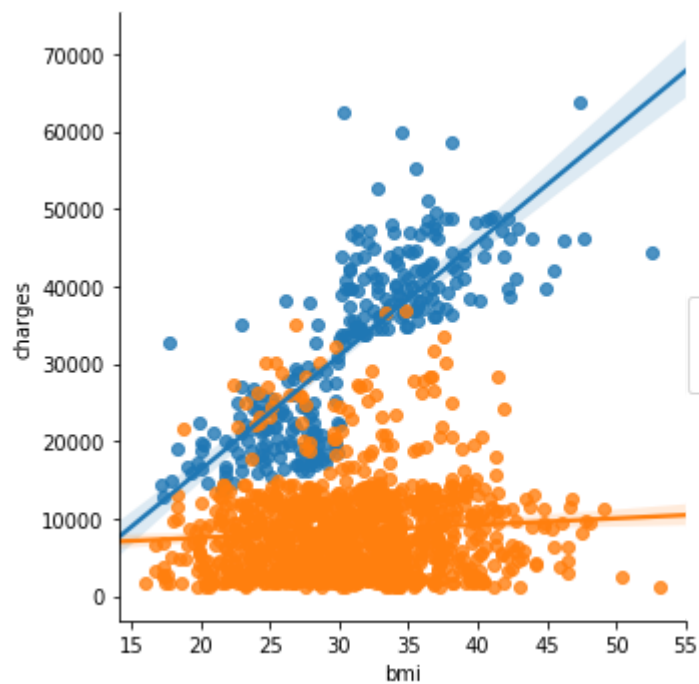


**More men smoke compared to woman. Below shows the breakdown percentage wise.**

```
In [40]: fs = len(df[(df['sex'] == 'female') & (df['smoker'] == 'yes')]) / len(df[df['sex']  
ms = len(df[(df['sex'] == 'male') & (df['smoker'] == 'yes')]) / len(df[df['sex'] =  
print('% number of females that smoke: ' + str(fs*100))  
print('% number of males that smoke: ' + str(ms*100))  
  
% number of females that smoke: 17.371601208459214  
% number of males that smoke: 23.5207100591716
```

```
In [41]: sns.lmplot(x='bmi', y='charges', hue='smoker', data=df, fit_reg=True)
```

```
Out[41]: <seaborn.axisgrid.FacetGrid at 0x294a2104518>
```

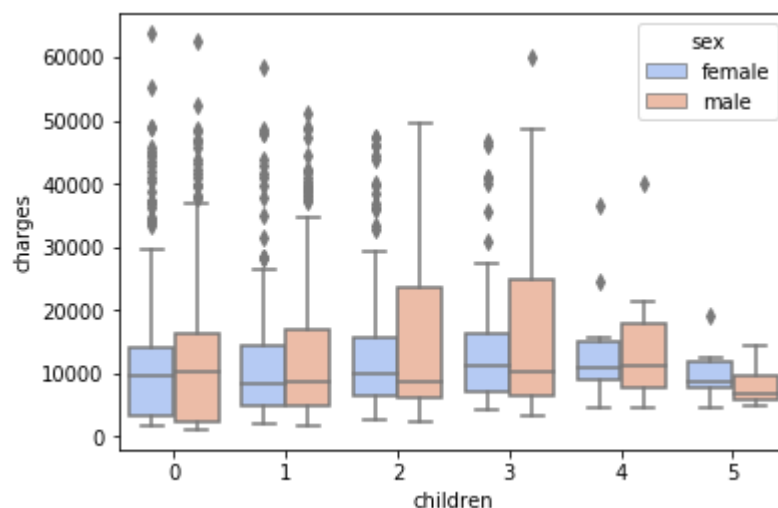


It can be seen that if you are a smoker, you will be charged higher. There is a strong linear trend in the data for smokers in regards to BMI. You smoke and your obese, you're gonna pay higher charges, plain and simple. More men smoke more than woman (although we are not told how much they smoke) and will have a higher total charge than females as seen by violin plot in previous section

## The impact of children

```
In [11]: sns.boxplot(x="children", y="charges", hue="sex", data=df, palette="coolwarm")
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x2949c22c7f0>
```



Here we can see the impact of children on gender wrt charges. On average there is little change in the median (per sex) wrt to the number of children in the family. If you are a male, you are more likely to pay on average more than say a male with two children. However, there seem to be more extreme cases (shown by the longer tails) of paying more charges if you have 2 or 3 children than compared to 0 and 1 or > 3 children

```
In [68]: for i in range(1,6):

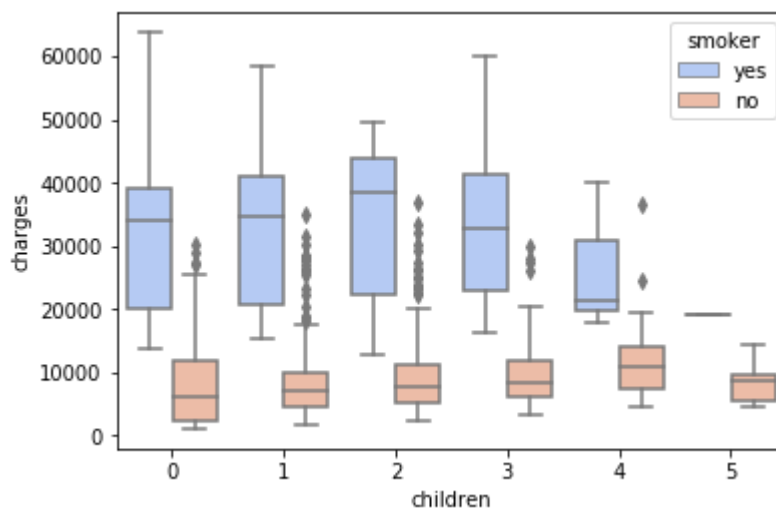
          print(str(i) + ' child: ' + str(len(df[(df['sex']=='male') & (df['children']
```

1 child: 166  
2 child: 121  
3 child: 80  
4 child: 14  
5 child: 10

There are few samples for 4 or more kids so statistics are not very robust. Typically want 30 or more samples before robustness starts to be more impactful

```
In [45]: sns.boxplot(x="children", y="charges", hue="smoker",data=df, palette="coolwarm")
```

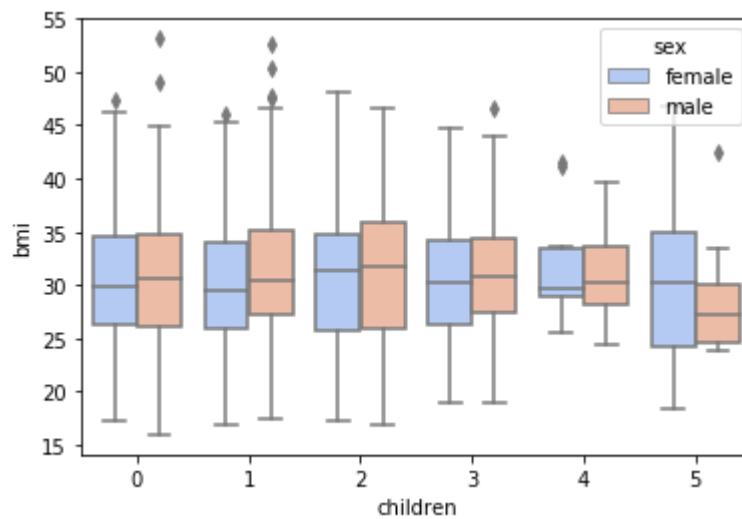
```
Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x25dc2abdfd0>
```



Here i show the effect of children and charges but splitting data in terms of smokers or no smokers. In all cases smokers are paying more

```
In [12]: sns.boxplot(x="children", y="bmi", hue="sex", data=df, palette="coolwarm")
```

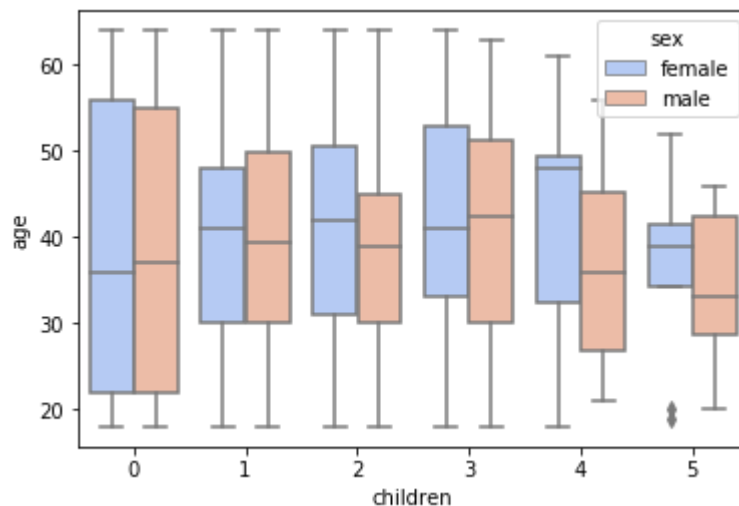
```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x2949c39cf28>
```



Here is how BMI is related to the number of children and is split by gender. Parents are generally in poorer health when they get two kids. This will also help partially explain why men are paying most charges when having two kids

```
In [13]: sns.boxplot(x="children", y="age", hue="sex", data=df, palette="coolwarm")
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x2949c4c5be0>
```



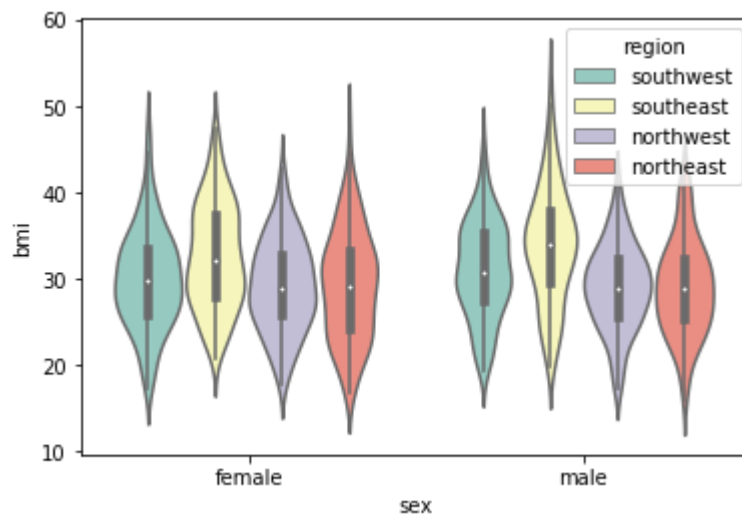
Age as a function of number of children split on gender. No obvious results here yet

## How does BMI impact?



```
In [19]: sns.violinplot(x="sex", y="bmi", data=df, hue='region', palette='Set3')
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x2949f13d2b0>
```

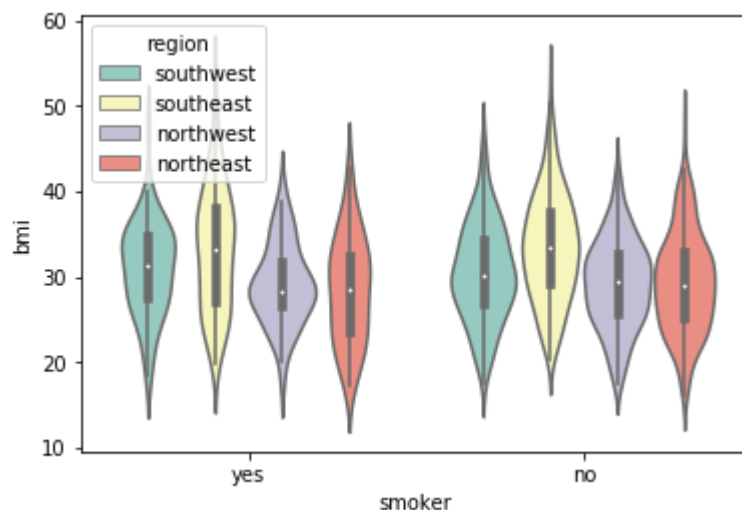


Here, BMI and gender are shown but are split based on region. For both sexes, people from the southeast have the largest BMI. Men are overall more over weight than woman. People from the northwest and northeast show the lowest BMIs. The distributions are fairly evenly spread around median values although some (like femal southeast) show a slight bimodal distribution.

```
In [ ]:
```

```
In [20]: sns.violinplot(x="smoker", y="bmi", data=df, hue='region', palette='Set3')
```

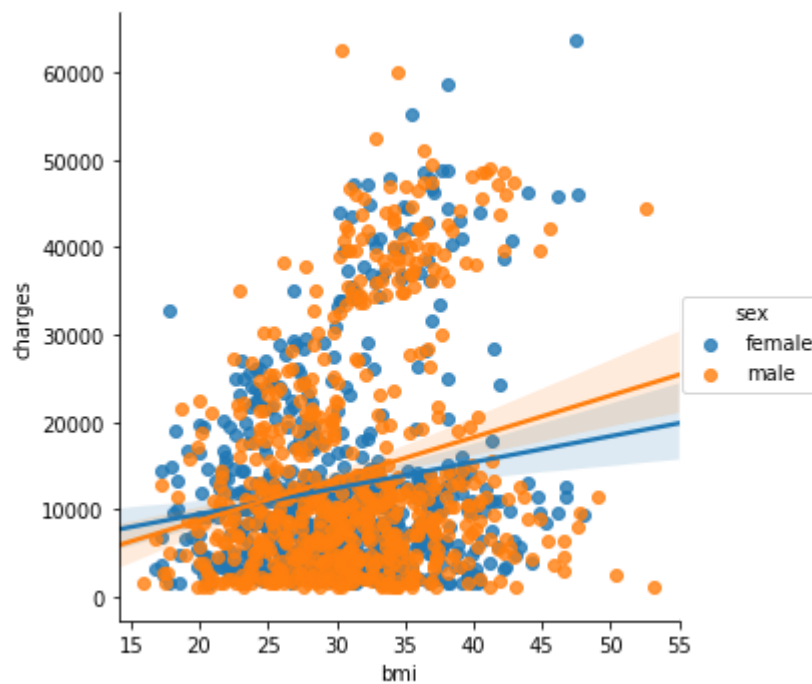
```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x2949f1bab00>
```



Similar to the above, except this is in comparison to smokers or not. This plot shows a similar result to the BMI vs sex plot

```
In [42]: sns.lmplot(x='bmi', y='charges', hue='sex', data=df, fit_reg=True)
```

```
Out[42]: <seaborn.axisgrid.FacetGrid at 0x294a1ebe080>
```

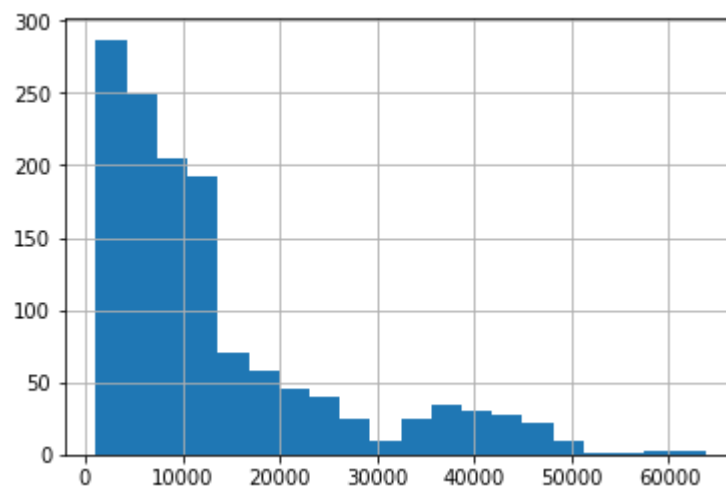


no clear clustering here, but based on BMI and gender, men are typically paying higher charges per unit change in BMI

## More on charges

```
In [4]: df['charges'].hist(bins=20)
```

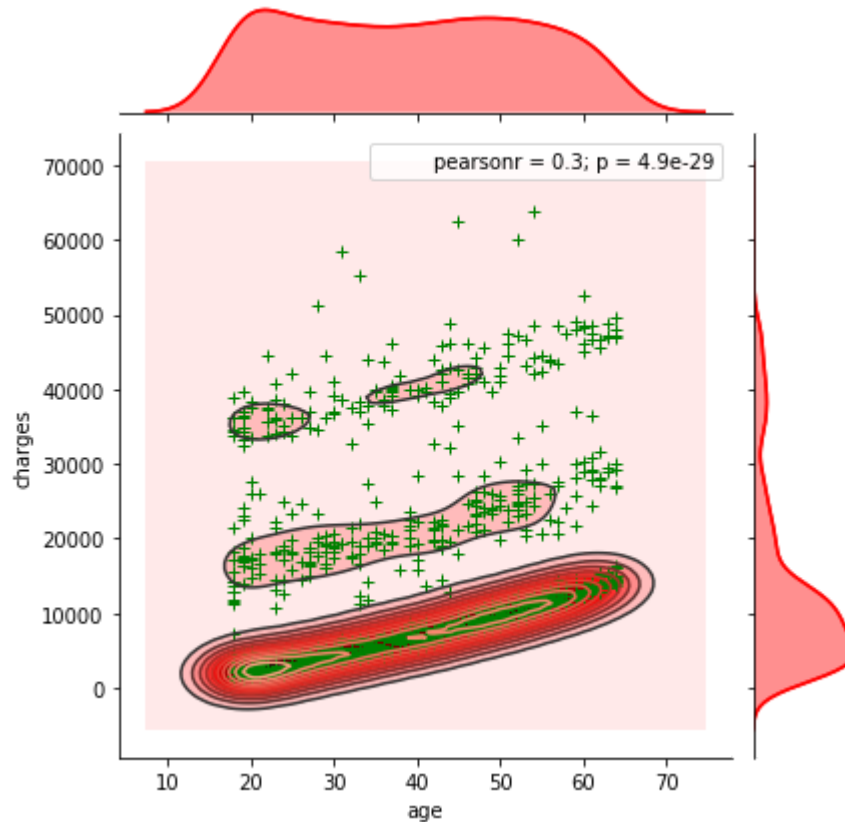
```
Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x259269ccb38>
```



Most charges are less than ~ 1300 USD. There is a second minor distribution of charges around 4000 USD owing to those that smoke and have a larger BMI. Few have charges

around 3000 USD. From previous plots and from here, we can say that if you are going to smoke, to keep charges down you need to have a BMI of less than ~30!

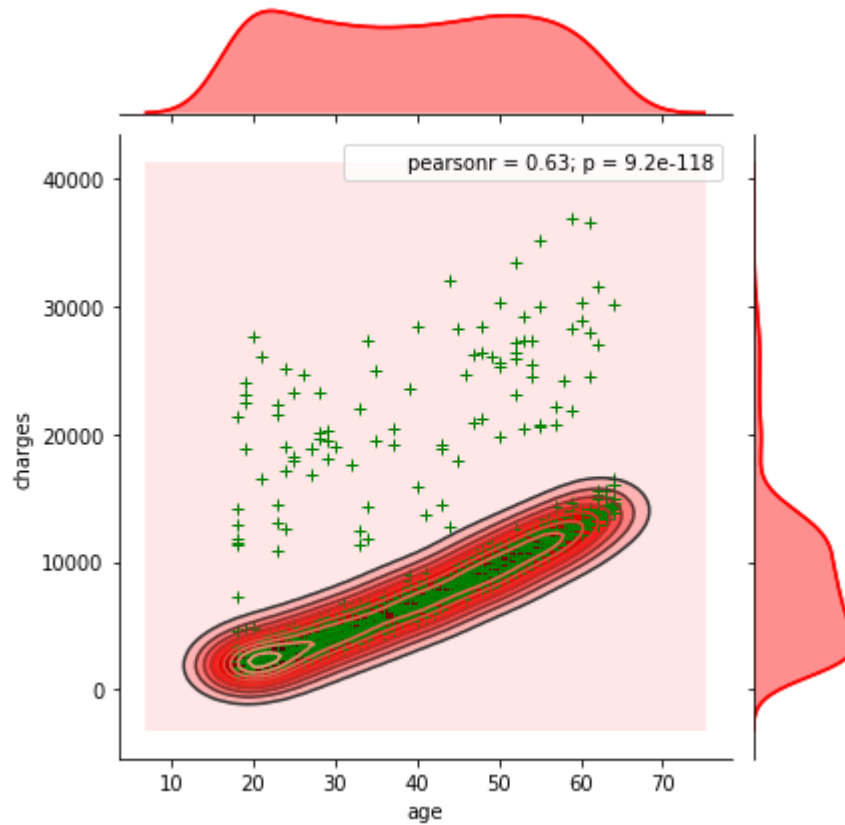
```
In [28]: g = sns.jointplot(x="age", y="charges", data = df, kind="kde", color="r")
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)
```

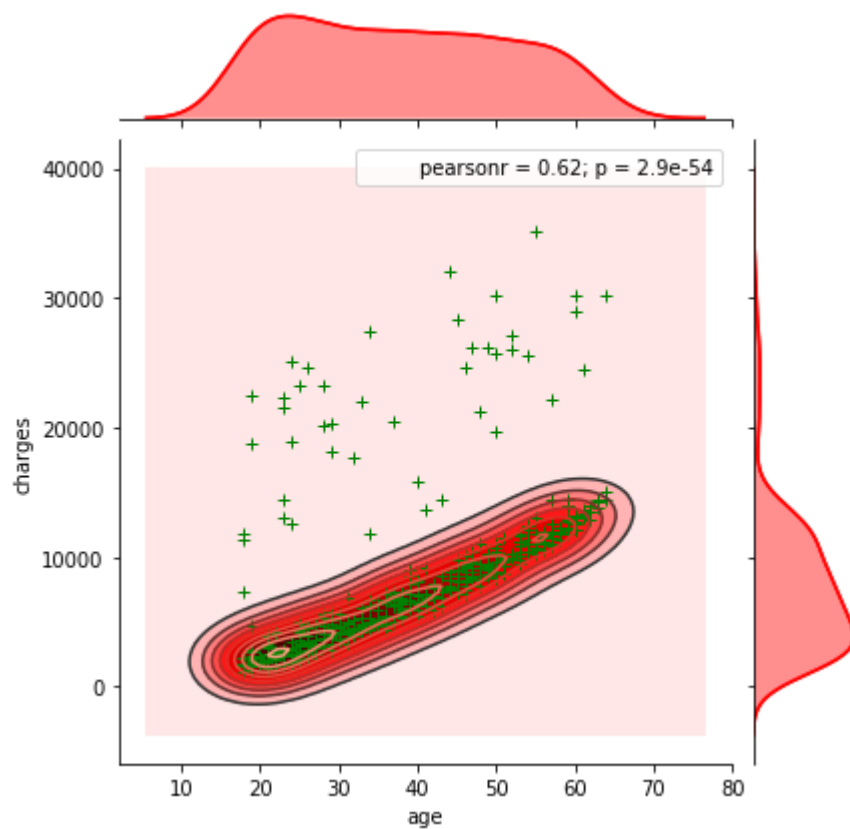


Here I am comparing age and see how charges are related. There appear to be three different clusters of data, with the most dense being the bottom level. So what are these clusters and are they related to smokers, BMI, children etc..

```
In [45]: g = sns.jointplot(x="age", y="charges", data = df[(df.smoker == 'no')], kind="kde",
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)

g = sns.jointplot(x="age", y="charges", data = df[(df.smoker == 'no') & (df.bmi < 3)], kind="kde",
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)
```



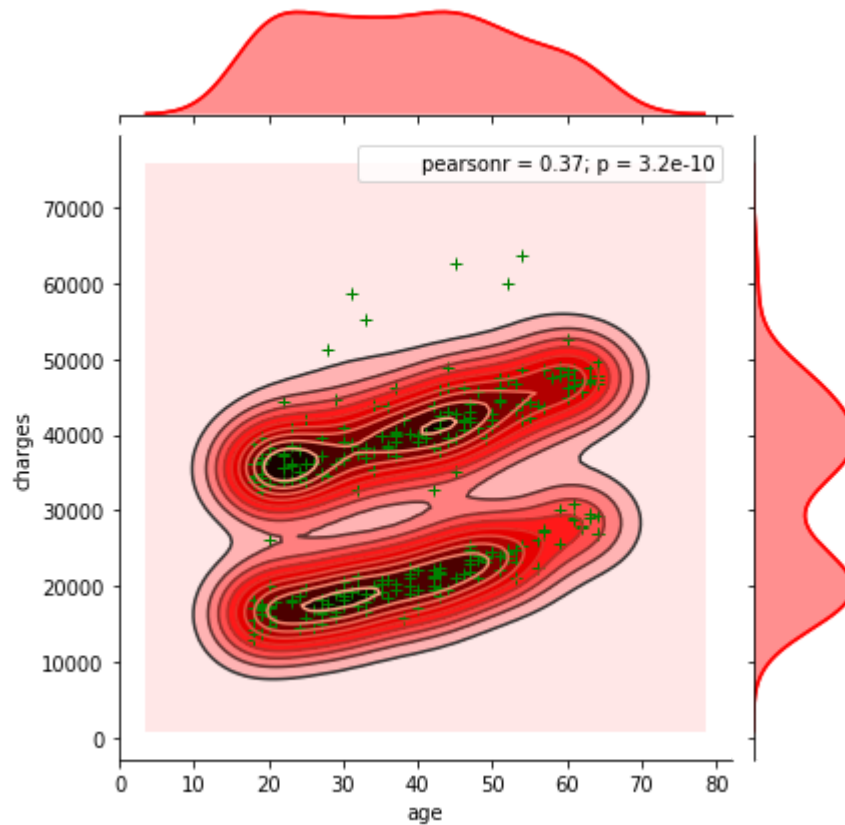


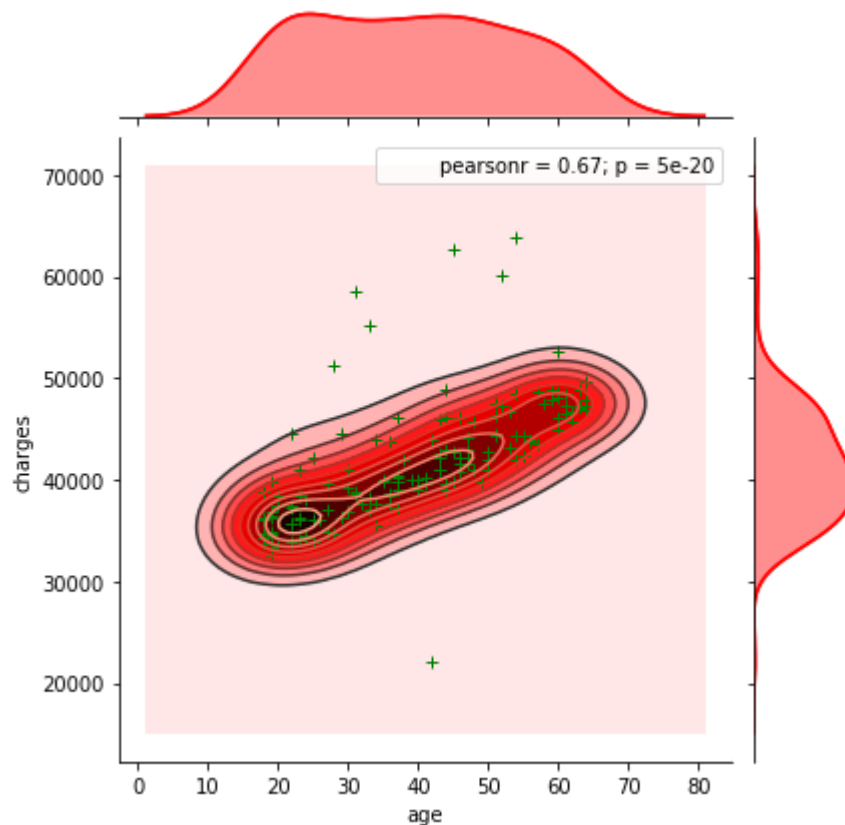
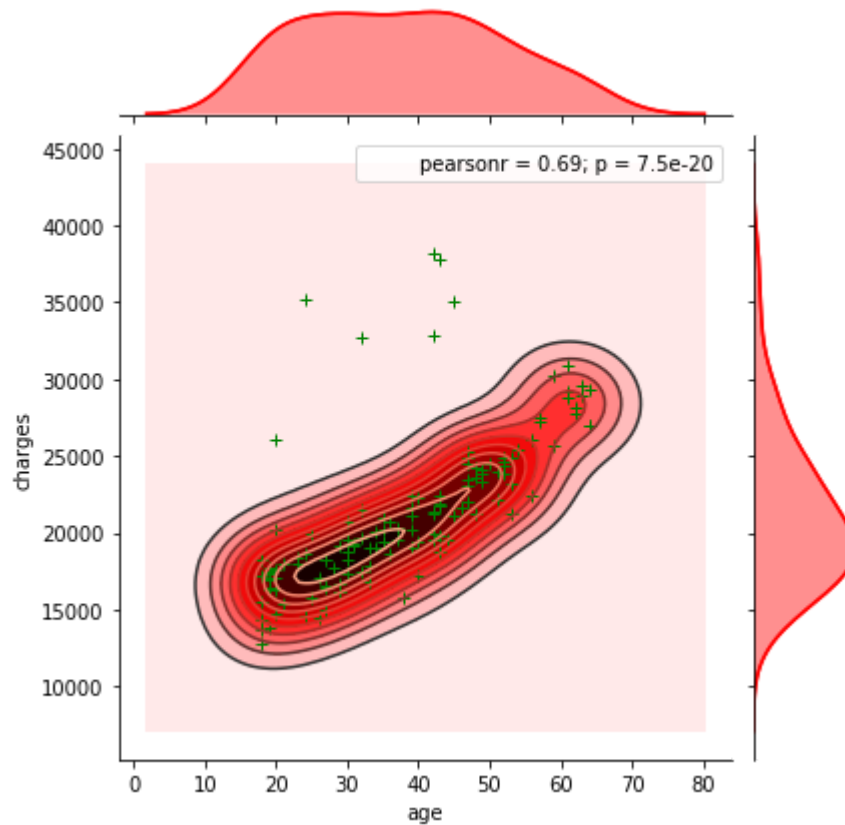
Data for non smokers (top) and non smokers with BMI < 30 (bottom) are shown above. The data is densist between 0 and ~1600 USD. There are some cases up as high as ~ 3500 USD. With the bottom plot there are cases that are due to other health reasons other than smoking and BMI

```
In [48]: g = sns.jointplot(x="age", y="charges", data = df[(df.smoker == 'yes')], kind="kde")
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)

g = sns.jointplot(x="age", y="charges", data = df[(df.smoker == 'yes') & (df.bmi <
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)

g = sns.jointplot(x="age", y="charges", data = df[(df.smoker == 'yes') & (df.bmi >
g.plot_joint(plt.scatter, c="g", s=30, linewidth=1, marker="+")
g = g.plot_joint(sns.kdeplot, cmap="Reds_d")
g = g.plot_marginals(sns.kdeplot, color="r", shade=True)
```





These plots show charges as a function of smokers (top) and smokers with a BMI <30 (middle) and smokers with BMI >30 (bottom). The top plot shows that there is clearly a bimodal distribution of smokers when it comes to charges. Which has also been seen earlier and seems to be separated using a BMI of 30. These clusters are separated using BMI either larger or smaller than 30

From an insurers perspective, I would be wanting to help these higher risk groups (to help firstly them for health reasons), but also minimize the amount of money paid out. It would seem that both smoking, BMI are large factors to large payouts, but lets us see if we can capture the importance of each features. By doing this, we could focus on specific groups to help lead to a healthier life style

## Random forest regressor

Here, I want to be able to predict charges given our input features using multivariate random forest regressor

```
In [8]: df.smoker.replace(('yes', 'no'), (1, 0), inplace=True)
df.sex.replace(('male', 'female'), (1, 0), inplace=True)
regions= pd.get_dummies(df['region'])
df1 = pd.concat([df,regions],axis=1)
df1.drop(['region'],axis=1, inplace=True)
df1.head()
```

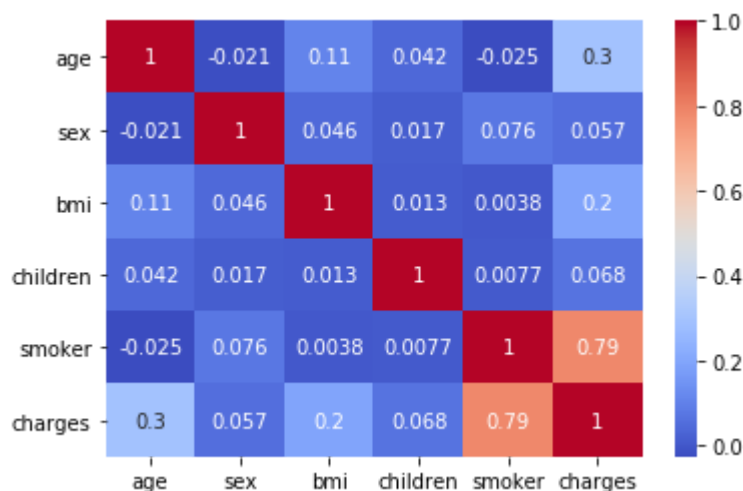
```
Out[8]:
```

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

hot encoded region, and also made sex and smoker binary

```
In [16]: sns.heatmap(df.corr(), cmap='coolwarm', annot=True)
```

```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x21e26bf2f28>
```



This correlation plot shows the correlations between each variable. We see that smokers, BMI, and age have the largest correlations with charges. We also can see that the



correlations with each other are low, which is a good sign concerning multi co-linearity purposes and their impact on regression.

```
In [9]: df2 = df1.copy()

targets = df1['charges']

df1.drop(['charges'],axis=1, inplace=True)

features = df1
```

### Set up features (independent variables) and target (dependent variable)

```
In [10]: from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(features, targets, train_size=

scaler = StandardScaler().fit(X_train)

X_train_scaled = pd.DataFrame(scaler.transform(X_train), index=X_train.index.value
X_test_scaled = pd.DataFrame(scaler.transform(X_test), index=X_test.index.values,
```

### Create a training and test set for the input data

```
In [48]: from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor(n_estimators=500, oob_score=True, min_samples_split=2,
rf.fit(X_train, y_train)
```

```
Out[48]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=500, n_jobs=1,
oob_score=True, random_state=0, verbose=0, warm_start=False)
```

Here, I set up a random forest state to predict charges based on the input features. I use 500 trees and use bootstrapping (ie with replacement to help reduce variance between trees). Maximum of two splits are used at each node. Oob\_score is to whether to use out-of-bag samples to estimate the R<sup>2</sup> on unseen data

```
In [49]: from sklearn.metrics import r2_score
from scipy.stats import spearmanr, pearsonr
predicted_train = rf.predict(X_train)
predicted_test = rf.predict(X_test)
test_score = r2_score(y_test, predicted_test)
spearman = spearmanr(y_test, predicted_test)
pearson = pearsonr(y_test, predicted_test)
print(f'Out-of-bag R-2 score estimate: {rf.oob_score_:>5.3}')
print(f'Test data R-2 score: {test_score:>5.3}')
print(f'Test data Spearman correlation: {spearman[0]:.3}')
print(f'Test data Pearson correlation: {pearson[0]:.3}')
```

```
Out-of-bag R-2 score estimate: 0.848
Test data R-2 score: 0.83
Test data Spearman correlation: 0.845
Test data Pearson correlation: 0.911
```

Our training and test scores are fairly similar which is good as it means that we are finding a good trade off between bias in the training set and overall variance between the training and test sets (ie, we are not over or underfitting the data). The  $R^2$  value is stating that the features we are using help explain about 80-85% of the variation of the overall charges. There is another 15-20% of variation that is not accounted for by these features.

```
In [55]: rf.feature_importances_
```

```
Out[55]: array([0.13286553, 0.00530178, 0.20026284, 0.02189229, 0.62383184,
0.00601176, 0.00365138, 0.00311414, 0.00306842])
```

```
In [59]: feature_importances = pd.DataFrame(rf.feature_importances_, index = X_train.columns
feature_importances
```

```
Out[59]:
```

	importance
smoker	0.623832
bmi	0.200263
age	0.132866
children	0.021892
northeast	0.006012
sex	0.005302
northwest	0.003651
southeast	0.003114
southwest	0.003068

As each independent variable was scaled when creating the random forest environment, we can compare the relative importance of each feature to each other directly (this is a relative value, where the sum = 1). Here, we see that (as hypothesized), smoker, bmi, and age have the largest impact on charges. smoker is the most influential and accounts for about 62%, bmi 20%, and age 13%. Hence, this can be thought of as given an  $R^2$  of 83%, smoking accounts for 62% of the total 83%

whilst we saw that children do have an influence on charges from the data visualization, especially on bmi, we can see that overall it can only be accountable for about 2%. Gender and location are fairly insignificant as weights,

However, which features are significant and which are not for the prediction of charges? We would need to look at P-values to find out

## Classic multi-variate least squares

```
In [17]: from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from sklearn.metrics import r2_score
import statsmodels.formula.api as sm

training = pd.concat([y_train, X_train], axis=1, sort=False)
testing = pd.concat([y_test, X_test], axis=1, sort=False)
```

### Training set model fit

```
In [16]: model1 = sm.ols(formula='charges ~ smoker + bmi + age + children + northeast + sex
fitted1 = model1.fit()
fitted1.summary()
```

Out[16]: OLS Regression Results

<b>Dep. Variable:</b>	charges	<b>R-squared:</b>	0.753
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.751
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	353.1
<b>Date:</b>	Fri, 16 Nov 2018	<b>Prob (F-statistic):</b>	2.87e-275
<b>Time:</b>	22:54:23	<b>Log-Likelihood:</b>	-9499.8
<b>No. Observations:</b>	936	<b>AIC:</b>	1.902e+04
<b>Df Residuals:</b>	927	<b>BIC:</b>	1.906e+04
<b>Df Model:</b>	8		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	-9142.5392	948.959	-9.634	0.000	-1.1e+04	-7280.182
<b>smoker</b>	2.451e+04	507.088	48.344	0.000	2.35e+04	2.55e+04
<b>bmi</b>	317.1960	35.163	9.021	0.000	248.187	386.205
<b>age</b>	251.9676	14.591	17.269	0.000	223.333	280.602
<b>children</b>	464.1098	167.239	2.775	0.006	135.898	792.322
<b>northeast</b>	-1702.9143	405.474	-4.200	0.000	-2498.667	-907.162
<b>sex</b>	-306.0062	409.683	-0.747	0.455	-1110.021	498.008
<b>northwest</b>	-2109.4432	409.918	-5.146	0.000	-2913.918	-1304.969
<b>southeast</b>	-2624.2085	477.085	-5.501	0.000	-3560.500	-1687.917
<b>southwest</b>	-2705.9732	423.730	-6.386	0.000	-3537.554	-1874.392

<b>Omnibus:</b>	210.515	<b>Durbin-Watson:</b>	2.185
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	516.035
<b>Skew:</b>	1.185	<b>Prob(JB):</b>	8.80e-113
<b>Kurtosis:</b>	5.760	<b>Cond. No.</b>	3.57e+17

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 1.92e-29. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

## Test model fit

```
In [18]: model1 = sm.ols(formula='charges ~ smoker + bmi + age + children + northeast + sex',
                        data=fitted2.data)
fitted2 = model1.fit()
fitted2.summary()
```

Out[18]: OLS Regression Results

<b>Dep. Variable:</b>	charges	<b>R-squared:</b>	0.755
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.750
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	151.5
<b>Date:</b>	Fri, 16 Nov 2018	<b>Prob (F-statistic):</b>	4.52e-115
<b>Time:</b>	22:56:17	<b>Log-Likelihood:</b>	-4037.3
<b>No. Observations:</b>	402	<b>AIC:</b>	8093.
<b>Df Residuals:</b>	393	<b>BIC:</b>	8129.
<b>Df Model:</b>	8		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	-1.193e+04	1370.579	-8.706	0.000	-1.46e+04	-9237.345
<b>smoker</b>	2.234e+04	709.928	31.466	0.000	2.09e+04	2.37e+04
<b>bmi</b>	382.8733	48.645	7.871	0.000	287.235	478.511
<b>age</b>	269.4708	20.422	13.195	0.000	229.321	309.620
<b>children</b>	517.2834	241.066	2.146	0.032	43.344	991.223
<b>northeast</b>	-2331.4061	583.304	-3.997	0.000	-3478.192	-1184.620
<b>sex</b>	209.3892	563.250	0.372	0.710	-897.970	1316.749
<b>northwest</b>	-2491.8791	567.607	-4.390	0.000	-3607.804	-1375.954
<b>southeast</b>	-3673.9869	638.428	-5.755	0.000	-4929.147	-2418.826
<b>southwest</b>	-3434.6567	620.542	-5.535	0.000	-4654.654	-2214.659

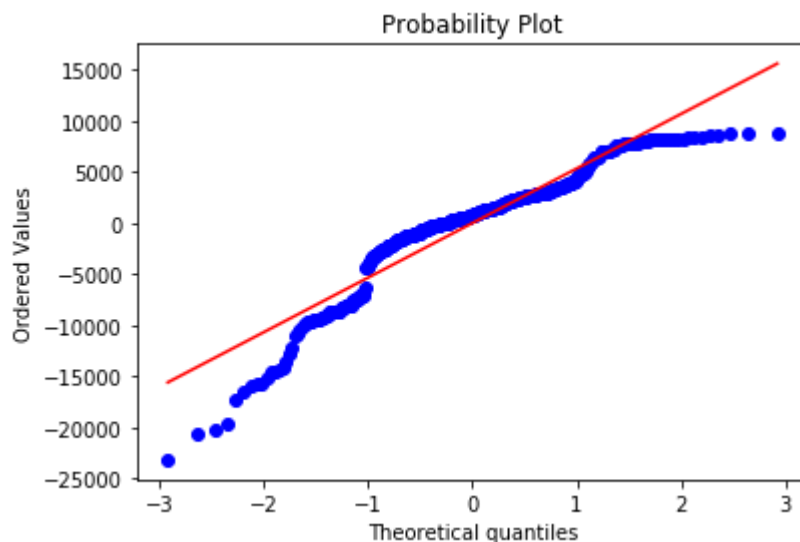
<b>Omnibus:</b>	85.316	<b>Durbin-Watson:</b>	2.081
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	156.490
<b>Skew:</b>	1.185	<b>Prob(JB):</b>	1.04e-34
<b>Kurtosis:</b>	4.931	<b>Cond. No.</b>	1.91e+17

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 2.95e-29. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

```
In [28]: import pylab
import scipy.stats as stats

stats.probplot((fitted2.predict() - y_test), dist="norm", plot=pylab)
pylab.show()
```



Q-Q plot to show how residuals lie when comparing the model predictions to the test data. If the predictions were perfect then they would line up along the red line. It can be seen that our residuals are mostly however not aligning.

Consulting the  $R^2$  values show the model to do a poorer a job at explaining the overall dependent variable variance ( $75\% < 83\%$ ). Along with the plot above, it could be feasible that our linear model contains features whose relationship with charges is non linear. It is possible to manipulate some of the features in order to engineer a more linear relationship, but this can be fairly exhaustive. A decision tree does not have issues with non linear effects and hence is a better predictor. However, our linear model does provide us with some other important information

The P-values tell us the level of feature importance (significance) to the overall fit by ways of testing the null hypothesis that the individual coefficients associated with each independent variable are equal to zero. Typically, we can use  $\alpha = 0.05$ , hence any P-value  $< 0.05$  we can reject the null hypothesis and consider that coefficient to be significant.

It can be seen, all but sex have P-values  $< 0.05$ , hence all (apart from sex) are important features in the prediction of charges.

## Summary

I have analyzed health insurance data in order to train and test various features as predictors of total charges paid by individual customers. Using various models and visualization, I have found that the most important features that predict charges are if clients are smokers or not, BMI, and age, but to a lesser degree, region and the number of children also have an impact. Gender was shown to have not significant impact on overall charges.

**It is in the best interest for the clients and the insurer that clients quit habits of smoking and to eat healthier. Perhaps as an incentive or product, the insurance company would be inclined to contact clients who are paying large charges to try help them make better everyday life decisions.**

In [ ]: