Report on Medical Insurance Charges data

Executive Summary

In this report, it was found that there is a very strong positive correlation between age and medical insurance charges while there is also a moderately positive correlation between age and BMI. It was also found that the Southeast region of the UK is where the highest medical charges occur with both males and females incurring above average costs (3 million and 2.3 million respectively). With regards to smoking status, male smokers and female non-smokers had above average costs while female smokers had the lowest costs. Based on these findings, target groups are customers in the Southeast, male smokers and female non-smokers. For these customers, there should be a reduction in the number of claims accepted by implementing stricter protocols or their premiums should be increased in order to reduce the overall burden to the company.

Introduction

The aim of this report is to understand how predefined variables have affected the medical insurance charges of the company over the last six months. Compared to the previous year, the company has seen a significant rise in payouts and this needs to be addressed in order to maximise our net profit margin. From the analysis that will be conducted, key subsets of customers will be identified as areas of focus and then recommendations for future steps in order to maximise profits will be provided.

To do this, the previous six months of data was collected from our claims department, including demographic, healthcare and financial information from relevant customers. Once organised into a spreadsheet, the relationships between age, sex, BMI, region, smoking status and insurance costs will be investigated.

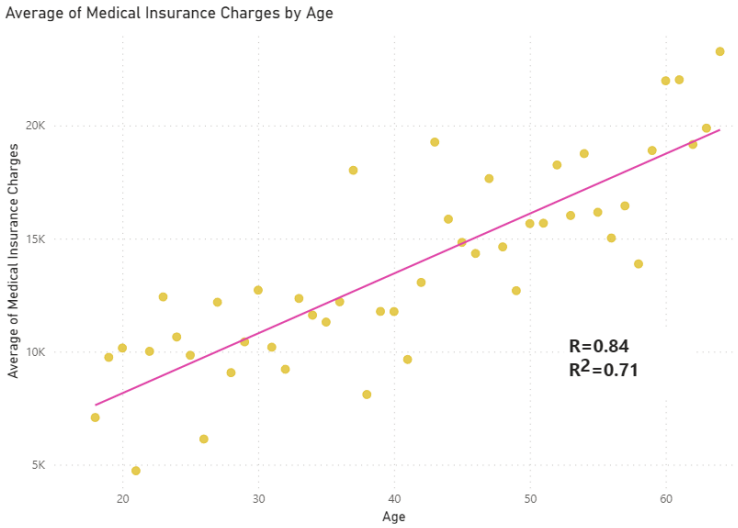
Findings & Analysis

The table below shows the summary statistics for the insurance claims data that was analysed. The average medical insurance charge was ~£13,000. There is a positive skew of the data suggesting an asymmetric distribution and also a negative kurtosis indicating there are less outliers than expected.

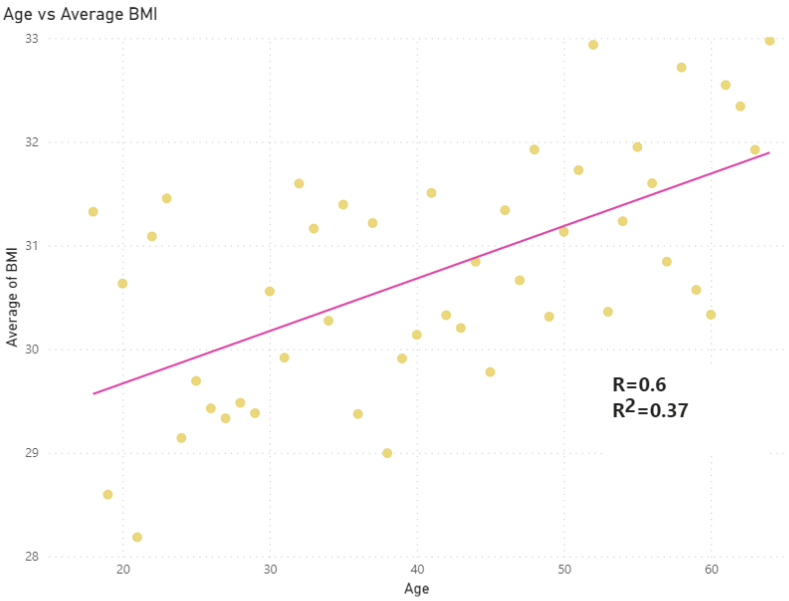
Mean	13270.42
Median	9382.03
Mode	1639.56
Standard Deviation	12110.01
Standard Error	331.07
Kurtosis	1.61
Skewness	1.52
Range	62648.56

Minimum	1121.87
Maximum	63770.43
Confidence Level (95%)	649.47

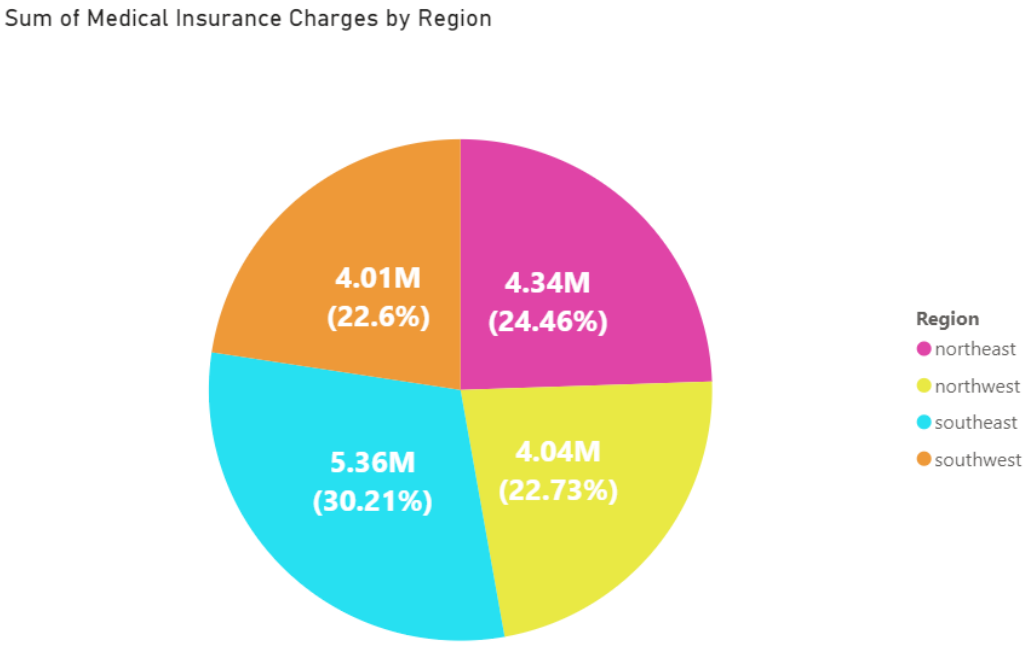
The first analysis that was conducted was regression analysis, comparing age with the average medical insurance charge and also with BMI. A strong positive correlation was identified between age and average cost with an R value of 0.84 and the R^2 value of 0.71 showed that the model which was used accounted for a large amount of the variation in the data.



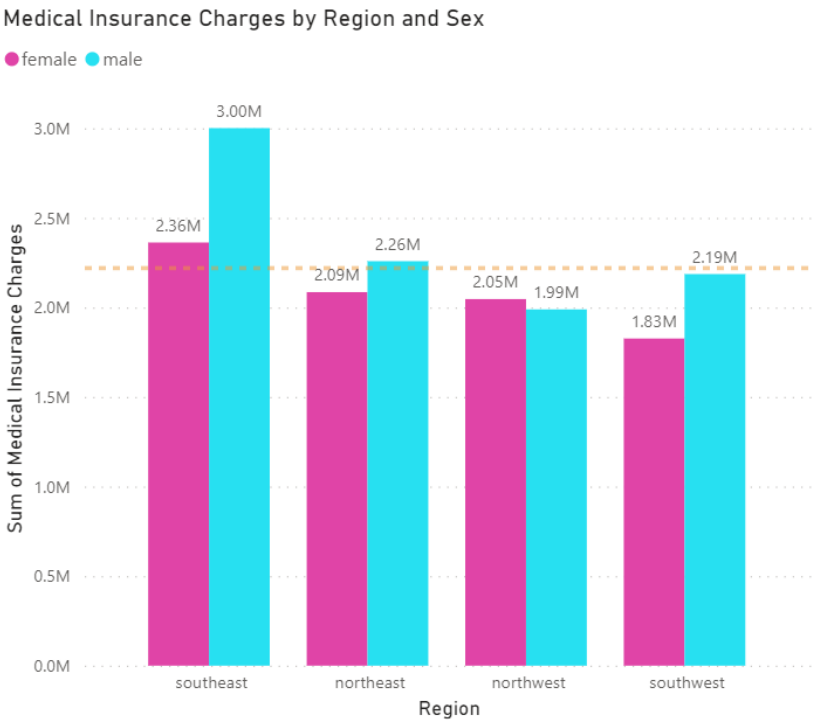
When comparing age and average BMI, a moderately positive correlation was identified with an R value of 0.6 however, an R^2 value of 0.37 was calculated, indicating that the model did not fit the dataset and couldn't account for the variance in the data.



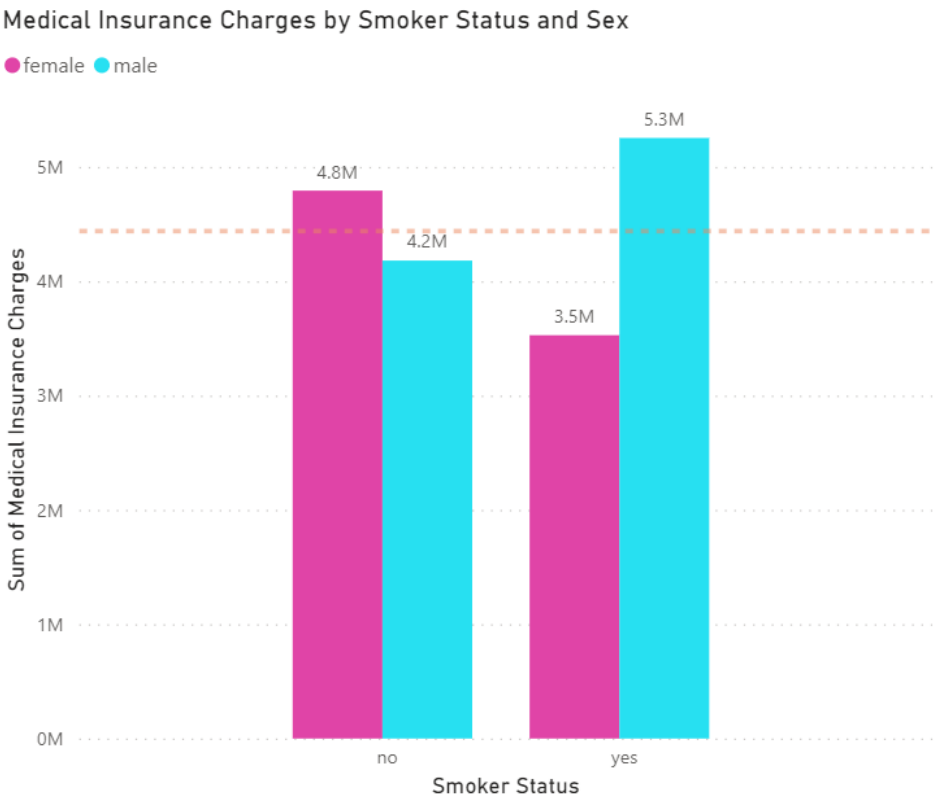
The next analysis determined the distribution of insurance charges regionally. Over 50% of the costs incurred are in the South with ~30% in the Southeast (5.36 million). The lowest costs are observed in both the Southwest and Northwest with ~4 million.



The final analysis investigated the effect region, smoking and gender had on insurance costs. The first bar graph once again highlighted that the Southeast region of the UK is where the highest medical charges occur with both males and females incurring above average costs (3 million and 2.3 million respectively). It also shows that the lowest costs are incurred by females in the Southwest (1.83 million).



The second bar graph indicates male smokers have the highest insurance costs while female smokers have the lowest. Interestingly, female non-smokers also have above average charges at £4.8 million compared to female smokers who sit at £3.5 million.



Recommendations

Firstly, the results need to be further analysed to see if they are statistically significant. The relationship between age and BMI then needs to be investigated further as the model used here is not a good fit for the data and does not explain the variation (R^2 value of 0.37). Something of note that was found in the analysis was that female non-smokers had higher medical charges (4.8 million) when compared to female smokers (3.5 million). This is not what you would expect so further analysis should be done to investigate this particular result.

In addition to this, I believe it would be beneficial to expand the dataset by also including the types of claims that are made and any previous medical conditions the individual has. These variables will then allow analysis to focus on specific conditions/groups of conditions that incur higher costs and changes that can be implemented in these key areas.

From the analysis conducted, target groups appear to be customers in the southeast, male smokers and female non-smokers. The company needs to find a way to either reduce the number of claims accepted by these particular customers i.e. by implementing new and stricter protocols with which medical payouts are given or increase their premiums. This will then reduce the burden these groups place on the company overall.