

Enhancing Health Insurance Premiums Through Data- Driven Insights

AUTHOR: IRKING CHUA

Summary

Using linear regression from dataset from Kaggle titled 'Medical Cost Personal Dataset' so that we can reveal some key insights:

- How insurance companies calculate the number of medical charges a person has based on their demographic and lifestyle factors.
- Investigate the significance of smoking, BMI, and age in determining higher medical cost
- Identify Strategies for minimizing these expenses.



Outline

- Business Problem
- Data & Methods
- Results
- Conclusions



Business Problem



This project represents the culmination of our coursework, where we are applying the knowledge, we've gained to analyze real-world data to address a substantial business challenge. Our primary aim is to utilize data analysis and linear regression techniques to discover actionable insights that will directly impact how we calculate the medical expenses covered by our health insurance policies.

Our goal is to investigate the significance of smoking, BMI, and age in determining higher medical costs and to identify strategies for minimizing these expenses.

Data and Methods



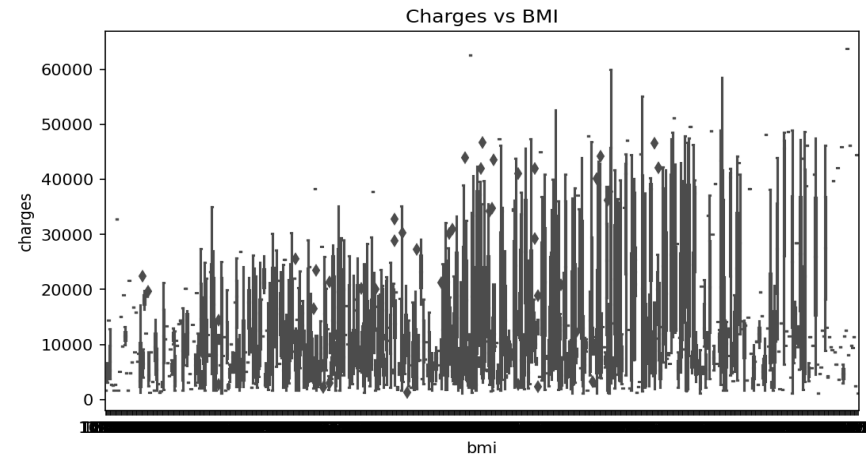
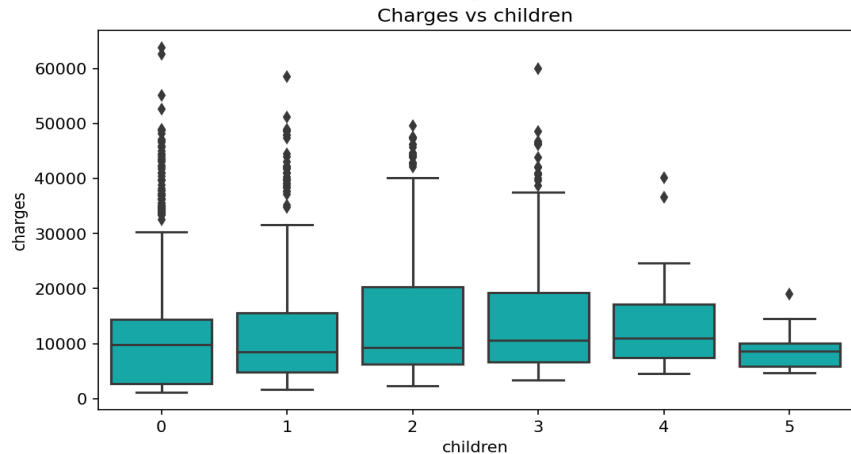
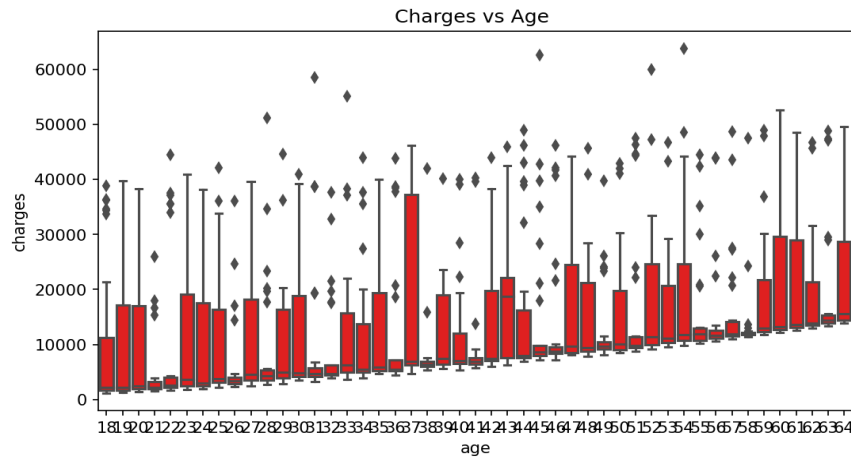
The dataset that we are using now is from [Kaggle](#), from the author, Miri Choi, where she states that the dataset comes from the book [Machine Learning with R](#) by [Brett Lanz](#). It contains columns like Age, Sex, Children and the likes

Data and Methods



It contains 1338 entires with 7 columns that includes how much the medical charges cost and the demographic and lifestyle factors, such as age, BMI, smoking habits and region, of the individual.

Data and Methods



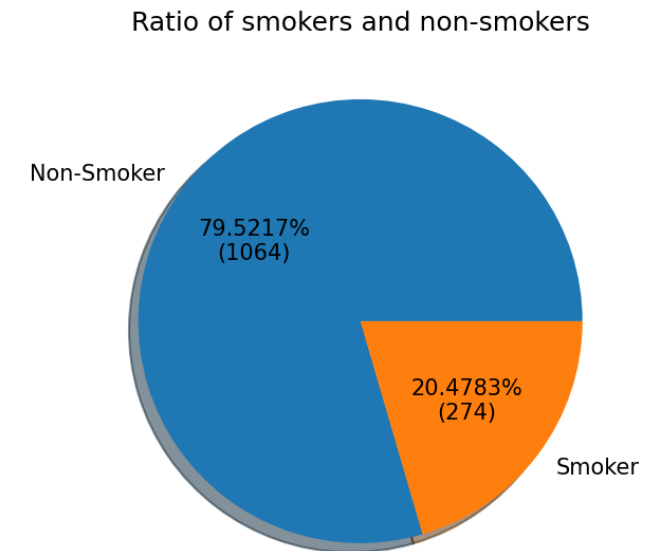
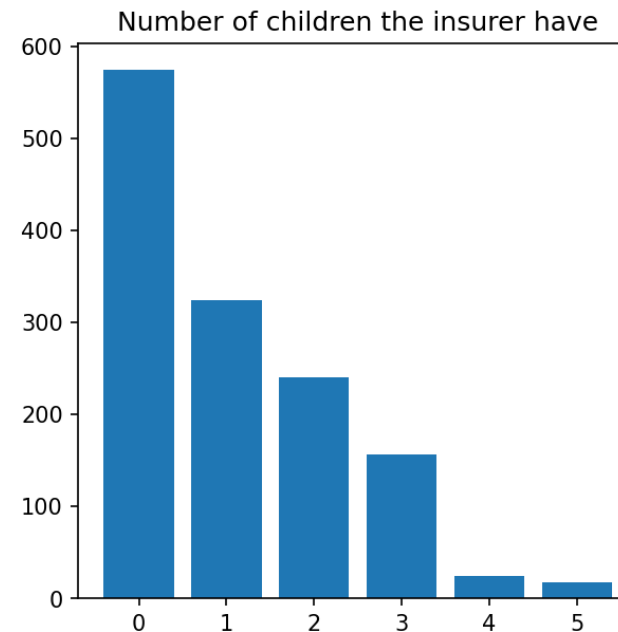
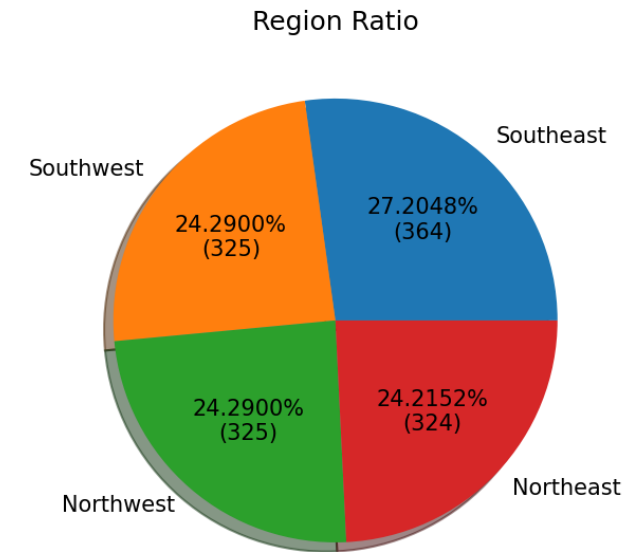
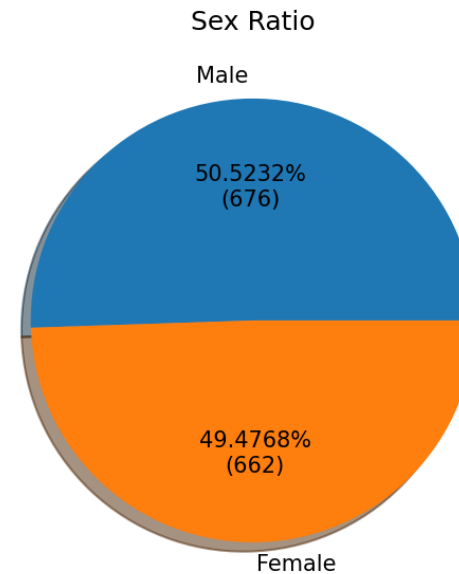
We had to exclude certain portions of the data because of the existence of larger numbers to make it more manageable.

Data and Methods

We can see that the data that we have is quite balanced we have the sex the ratio at almost half, region is almost identical as well.

Number of children we have a lot of people with no child which make sense in real life

Also the number of smokers we have 79.52% non- smokers and 20.48% smokers



Data and Methods

It shows that the model appears to be a good fit, as indicated with the high R-squared and F-statistic, even considering model complexity. This suggests that the model explains a significant portion of the variation in "charges". However, it is important to be aware of the limitations, such as the non-normal distribution of residuals and potential issues related to skewness and kurtosis, which may warrant further investigation to ensure the reliability of the model's prediction.

OLS Regression Results

| | | | |
|--------------------------|------------------|----------------------------|----------|
| Dep. Variable: | charges | R-squared: | 0.774 |
| Model: | OLS | Adj. R-squared: | 0.770 |
| Method: | Least Squares | F-statistic: | 210.6 |
| Date: | Fri, 06 Oct 2023 | Prob (F-statistic): | 0.00 |
| Time: | 13:00:11 | Log-Likelihood: | -889.84 |
| No. Observations: | 1316 | AIC: | 1824. |
| Df Residuals: | 1294 | BIC: | 1938. |
| Df Model: | 21 | | |
| Covariance Type: | nonrobust | | |
| Omnibus: | 468.736 | Durbin-Watson: | 2.036 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 1690.390 |
| Skew: | 1.729 | Prob(JB): | 0.00 |
| Kurtosis: | 7.343 | Cond. No. | 12.6 |

Results



The resulting model that was made from the data indicates that

- Smoking is the biggest factor regarding higher medical charges followed by old age and high BMI (Body Mass Index).
- It also revealed that having more than 3 children can increase the value of the medical charges.
- I believe we will have a more accurate results if we have more data and if we have things like pre-existing conditions

Results



It revealed that

- Smokers pay \$23,790 additional in medical charges.
- Ages after 42 increases the medical charges greatly and the older you get the more the charges become. For context the model revealed that ages 63-66 pays \$11,010 in medical fees.
- BMI increases \$2,061 per BMI, for context the normal healthy weight is 18.5.

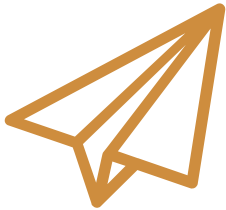
Conclusions



Here are some strategies to minimize medical charges:

- Do not smoke – smoking has a lot of bad effect in our body. It is a known carcinogen and can lead to different kinds of cancer.
- Try to maintain a healthy weight – controlling your BMI is hard and some articles indicate that it is misleading as it does not account for muscle mass and overall body composition. Irregardless, maintaining a good diet and balanced diet is good for your health.
- Having less children – having a family is a dream but you must also consider financially if it is viable. Having less kids can lessen the medical charges.

Thank You



- Email:
chuairking@gmail.com



- Github: @icchua



- LinkedIn:
www.linkedin.com/in/icchua