Insurance price prediction analysis

Blood Pressure :

When the Insurance buyer doesn't have any blood pressure, the average premium is 23,357, average age = 38.5, average weight is 77.76 kg and average height is 168.5 and 524 applicants doesn't have any blood pressure.

When the Insurance buyer does have blood pressure, the average premium is 25,448, average age = 45.39 years, average weight is 76 kg and average height is 167.77 and 462 applicants does have any blood pressure.

We can say when that when the insurance purchaser has any blood pressure problems, he has to purchase the premium with avg of 2000 rupees higher.

Any Chronic Diseases:

When the Insurance buyer doesn't have any chronic diseases, the average premium is 23,725, average age = 41.41 years, average weight is 77.172 kg and average height is 168 cm and 808 applicants doesn't have any chronic diseases.

When the Insurance buyer does have any chronic diseases, the average premium is 27,112, average age = 43.26 years, average weight is 76 kg and average height is 169.2 and 178 applicants does have chronic diseases.

We can observe that 80 percentage of the applicants does not have any chronic diseases.

Any Transplant:

When the Insurance buyer doesn't have any transplant, the average premium is 23,898, average age = 41.77 years, average weight is 76.94 kg and average height is 168.26 cm and 931 applicants doesn't have any transplant.

When the Insurance buyer does have transplant, the average premium is 31,764, average age = 41.25 years, average weight is 77.07kg and average height is 166.87 and 55 applicants does have organ transplant.

We can observe that 93 percentage of the applicants does not have any organ transplant.

History of Cancer:

When the Insurance buyer doesn't have any history of cancer, the average premium is 24,147, average age = 41.88 years, average weight is 76.93 kg and average height is 168.14 cm and 870 applicants doesn't have any history of cancer.

When the Insurance buyer does have history of cancer, the average premium is 25,759, average age = 40.690 years, average weight is 77.08kg and average height is 168.47 and 116 applicants does have history of cancer.

We can observe that 87 percentage of the applicants does not have any organ transplant.

Known Allergies:

When the Insurance buyer doesn't have any Known Allergies, the average premium is 24,297, average age = 41.92 years, average weight is 76.67 kg and average height is 168.23 cm and 774 applicants doesn't have any Known Allergies.

When the Insurance buyer does have Known Allergies, the average premium is 24,481, average age = 41.09 years, average weight is 78kg and average height is 168 and 212 applicants does have Known Allergies.

We can observe that 77.4 percentage of the applicants does not have any organ transplant.

Premium Age distribution:

The average primum price a person pay between, Age 18 to 20 is 16114

Age 20 to 30 is 16518

Age 30 to 40 is 24143

Age 40 to 50 is 26638

Age 50 to 60 is 28608

Age 60 to 65 is 28859

From the above we can see they insurance purchaser pays a lower insurance premium if he starts purchasing insurance from a young age.

Number of major surgeries:

When the Insurance buyer doesn't have any major surgeries, the average premium is 22,969, average age = 37.85 years, average weight is 77.28 kg and average height is 167.83 cm and 479 applicants doesn't have any major surgeries.

When the Insurance buyer does have one major surgeries, the average premium is 24,742, average age = 40.82 years, average weight is 76.34 kg and average height is 168.39 cm and 372 applicants does have one major surgeries.

When the Insurance buyer does have two major surgeries, the average premium is 28,084, average age = 57.37 years, average weight is 77.23 kg and average height is 168.64 cm and 119 applicants does have two major surgeries.

When the Insurance buyer does have three major surgeries, the average premium is 28,000, average age = 63.50 years, average weight is 78.87 kg and average height is 170 cm and 16 applicants does have three major surgeries.

In summary

As the number of surgeries increase the average premium price a person pays increases 1800.

Aged person pay higher insurance premium compared to lower aged person as the aged person might have gone through surgeries.

First Dashboard

https://public.tableau.com/views/Insurancepriceprediction_Arivalagan_2/ICPDashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Second Dashboard

https://public.tableau.com/shared/6G69RBR3N?:display_count=n&:origin=viz_share_link

Now lets look into the Colab analytics and model building.

Analysis on the number of total counts on each unique values in a feature column.

- The unique counts of the age we could see that in most age there are atleast 15 people who have purchased a policy.
- When it comes to weight it seems that people in the weight between 65kg to 80 kgs are more in puchasing the policy.
- The most premium opted are 23000 with 249 people who had purchased it, 15000 by 202 people, 38000 by 132 people, 25000 by 103 people and 29000 by 72 people.

Insights from Boxplot

- We could see that there are more insurance premium purchaser between the age 30 to 53 with a median age as 42. No outliers observed on this category.
- We could see that the height of the premium purchaser lies between 161cm to 176cm with median height of 168. No outliers observed on this category.
- With respect to the weight feature we could see the weight range of most of the people purchasing the premium are between 67kg to 87kg with a median weight of 75 kg. There are many outliers here hence we might have to use either a minmax or a standard scaler wile trying out linear or logistic regression.
- Most of the premium purchased falls between 21000 to 28000 with a median of 23000. There are very few outliers.

Insights from bar plots

- From the above seven Barplots we could say that roughly 50 percentage of the Insurance purchaser are not have any health issues and had not undergone any surgeries.
- We could see that there are purchaser of the insurance who have diabetics and blood pressure problems around 42 percentages

Insights from Correlation plot

- We can see height is negatively correlated with BMI. BMI is positively correlated with weight.
- Age is positively correlated with premium price.
- Hence while creating Linear regression or logistic regression we will drop the weight feature.

Insights from qqplot

- From the above plot we could see that they are not normally distributed

KDE plot insights

- For all the numerical columns we could see that they are not normally distributed.
- We can see that age alone is falling a distribution similar to normal distribution.

Insights from scatter plot

- We could see that how the premiums are distributed across each categorical variables.

Feature Engineering

- We are creating 2 feature age_cat with 5 category and BMI cat with 4 category.

The average premium based on Age category and BMI category

Age Category

18-30 16443.514644

30-40 24142.857143

40-50 26638.297872

50-60 28607.734807

60+ 28859.259259

BMI Category

Underweight 22666.666667

Normal weight 23755.555556

Overweight 24236.760125

Obese 25237.942122

Insights on bar plots for Age Category and BMI category

- People of age between 18 to 30 are around 25 percentage who have purchased the insurance policy.
- People between 40 to 50 are around 23 percentage followed by people between 30 to 40 years of age with 20 percentage.
- People above 50 and less than 60 are around 18 percentage and people above 60 are around 13 percentage.
- 32 percentage each of Normal, obese and overweight people have purchased Health Insurance Premium.
- only 4 percentage of under weight people have purchased the Insurance premium.

Hypothesis Testing

- We will begin the hypothesis testing using ttest_ind for all the binaray categorical columns.

  Below is the result of NULL hypothesis.

Based on the t-test results, we can draw several conclusions regarding the average premium prices for various health conditions:

1. **Diabetes**:
   - Average premium price for individuals without Diabetes: ₹23,931.82
   - Average premium price for individuals with Diabetes: ₹24,896.14
   - **p-value**: 0.0167
   - Since the p-value is less than 0.05, we reject the null hypothesis, suggesting a significant difference in premium prices for individuals with and without Diabetes.
2. **Blood Pressure Problems**:
   - Average premium price for individuals without Blood Pressure Problems: ₹23,356.87
   - Average premium price for individuals with Blood Pressure Problems: ₹25,448.05
   - **p-value**: 1.31e-07
   - The p-value is extremely low, leading us to reject the null hypothesis. This indicates a significant difference in premium prices between those with and without Blood Pressure Problems.
3. **Any Transplants**:
   - Average premium price for individuals without Transplants: ₹23,897.96
   - Average premium price for individuals with Transplants: ₹31,763.64
   - **p-value**: 1.98e-20
   - With a very low p-value, we reject the null hypothesis. This confirms a significant disparity in premium prices for individuals with and without Transplants.
4. **Any Chronic Diseases**:
   - Average premium price for individuals without Chronic Diseases: ₹23,725.25
   - Average premium price for individuals with Chronic Diseases: ₹27,112.36
   - **p-value**: 3.71e-11

- The p-value being much lower than 0.05 allows us to reject the null hypothesis, indicating a significant difference in premium prices for those with and without Chronic Diseases.

5. **Known Allergies**:
    - Average premium price for individuals without Known Allergies: ₹24,297.16
    - Average premium price for individuals with Known Allergies: ₹24,481.13
    - **p-value**: 0.7043
    - Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This suggests no significant difference in premium prices between those with and without Known Allergies.

6. **History of Cancer in Family**:
    - Average premium price for individuals without History of Cancer in Family: ₹24,147.13
    - Average premium price for individuals with History of Cancer in Family: ₹25,758.62
    - **p-value**: 0.0090
    - The p-value being less than 0.05 leads us to reject the null hypothesis, indicating a significant difference in premium prices for individuals with and without a History of Cancer in Family.

These results highlight the impact of various health conditions on insurance premium prices, reflecting higher premiums for conditions like Diabetes, Blood Pressure Problems, Transplants, Chronic Diseases, and a Family History of Cancer. On the other hand, Known Allergies do not appear to significantly influence premium costs.

Based on the Chi-square test results, we can summarize the findings as follows:

1. **Blood Pressure Problems and Number of Major Surgeries**:
    - **Chi2 Statistic**: 86.12
    - **p-value**: 1.49e-18
    - The very low p-value indicates a significant association between Blood Pressure Problems and the Number of Major Surgeries. We reject the null hypothesis, suggesting that these variables are dependent.

2. **Blood Pressure Problems and Age Category**:
    - **Chi2 Statistic**: 62.04
    - **p-value**: 1.08e-12
    - With a low p-value, we reject the null hypothesis, indicating a significant relationship between Blood Pressure Problems and Age Category.

3. **Blood Pressure Problems and BMI Category**:
    - **Chi2 Statistic**: 3.83
    - **p-value**: 0.28
    - Since the p-value is greater than 0.05, we fail to reject the null hypothesis, suggesting no significant association between Blood Pressure Problems and BMI Category.

4. **Any Transplants and Any Chronic Diseases**:
    - **Chi2 Statistic**: 0.86

- **p-value**: 0.35
- The p-value is greater than 0.05, leading us to fail to reject the null hypothesis, indicating no significant relationship between Any Transplants and Any Chronic Diseases.

5. **Any Transplants and Known Allergies**:
   - **Chi2 Statistic**: 0.0
   - **p-value**: 1.0
   - The p-value is significantly greater than 0.05, showing no significant association between Any Transplants and Known Allergies.
6. **Any Transplants and History of Cancer in Family**:
   - **Chi2 Statistic**: 0.17
   - **p-value**: 0.68
   - We fail to reject the null hypothesis, suggesting no significant relationship between Any Transplants and History of Cancer in Family.
7. **Any Transplants and Number of Major Surgeries**:
   - **Chi2 Statistic**: 0.72
   - **p-value**: 0.87
   - The p-value indicates no significant association between Any Transplants and the Number of Major Surgeries.
8. **Any Transplants and Age Category**:
   - **Chi2 Statistic**: 2.00
   - **p-value**: 0.74
   - We fail to reject the null hypothesis, suggesting no significant relationship between Any Transplants and Age Category.
9. **Any Transplants and BMI Category**:
   - **Chi2 Statistic**: 0.70
   - **p-value**: 0.87
   - The p-value is greater than 0.05, showing no significant association between Any Transplants and BMI Category.
10. **Any Chronic Diseases and Known Allergies**:
    - **Chi2 Statistic**: 0.58
    - **p-value**: 0.45
    - The p-value indicates no significant relationship between Any Chronic Diseases and Known Allergies.
11. **Any Chronic Diseases and History of Cancer in Family**:
    - **Chi2 Statistic**: 0.02
    - **p-value**: 0.89
    - We fail to reject the null hypothesis, suggesting no significant association between Any Chronic Diseases and History of Cancer in Family.

12. **Any Chronic Diseases and Number of Major Surgeries**:
    - **Chi2 Statistic**: 6.11
    - **p-value**: 0.11
    - The p-value suggests no significant relationship between Any Chronic Diseases and the Number of Major Surgeries.
13. **Any Chronic Diseases and Age Category**:
    - **Chi2 Statistic**: 21.83
    - **p-value**: 2.16e-04
    - The low p-value indicates a significant association between Any Chronic Diseases and Age Category. We reject the null hypothesis.
14. **Any Chronic Diseases and BMI Category**:
    - **Chi2 Statistic**: 4.04
    - **p-value**: 0.26
    - We fail to reject the null hypothesis, indicating no significant relationship between Any Chronic Diseases and BMI Category.
15. **Known Allergies and History of Cancer in Family**:
    - **Chi2 Statistic**: 12.27
    - **p-value**: 4.60e-04
    - With a low p-value, we reject the null hypothesis, indicating a significant association between Known Allergies and History of Cancer in Family.
16. **Known Allergies and Number of Major Surgeries**:
    - **Chi2 Statistic**: 109.30
    - **p-value**: 1.55e-23
    - The extremely low p-value indicates a significant relationship between Known Allergies and the Number of Major Surgeries. We reject the null hypothesis.
17. **Known Allergies and Age Category**:
    - **Chi2 Statistic**: 2.30
    - **p-value**: 0.68
    - We fail to reject the null hypothesis, suggesting no significant association between Known Allergies and Age Category.
18. **Known Allergies and BMI Category**:
    - **Chi2 Statistic**: 3.67
    - **p-value**: 0.30
    - The p-value indicates no significant relationship between Known Allergies and BMI Category.
19. **History of Cancer in Family and Number of Major Surgeries**:
    - **Chi2 Statistic**: 160.28

- **p-value**: 1.59e-34
- The extremely low p-value suggests a significant association between History of Cancer in Family and the Number of Major Surgeries. We reject the null hypothesis.

20. **History of Cancer in Family and Age Category**:
    - **Chi2 Statistic**: 1.11
    - **p-value**: 0.89
    - The p-value indicates no significant relationship between History of Cancer in Family and Age Category.
21. **History of Cancer in Family and BMI Category**:
    - **Chi2 Statistic**: 4.44
    - **p-value**: 0.22
    - We fail to reject the null hypothesis, indicating no significant association between History of Cancer in Family and BMI Category.
22. **Number of Major Surgeries and Age Category**:
    - **Chi2 Statistic**: 348.56
    - **p-value**: 2.82e-67
    - The extremely low p-value indicates a significant relationship between the Number of Major Surgeries and Age Category. We reject the null hypothesis.
23. **Number of Major Surgeries and BMI Category**:
    - **Chi2 Statistic**: 7.53
    - **p-value**: 0.58
    - The p-value suggests no significant association between the Number of Major Surgeries and BMI Category.
24. **Age Category and BMI Category**:
    - **Chi2 Statistic**: 4.49
    - **p-value**: 0.97
    - We fail to reject the null hypothesis, indicating no significant relationship between Age Category and BMI Category.

These results highlight that certain health conditions and demographics are significantly associated with each other, while others are not. The significant associations, such as between Blood Pressure Problems and Number of Major Surgeries, provide valuable insights for understanding health patterns in the population.

Insights from the Hypothesis testing

- We could see the mean of the Premium prices of the binary categories are different.
- ANOVA cant be performed as the mean are different for each group which can be observed from the boxplot

Insights from Chi square test

- Few categories likes Diabetes and know allergies have some dependency with respect to other categorical columns hence for logistic and linear regression models we can remove them.

Model Building

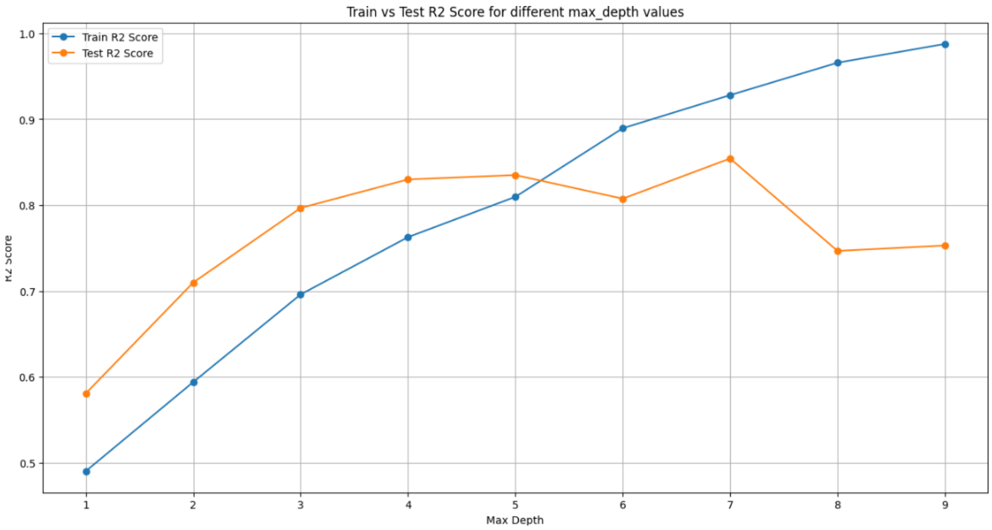
Columns to be dropped for model Building

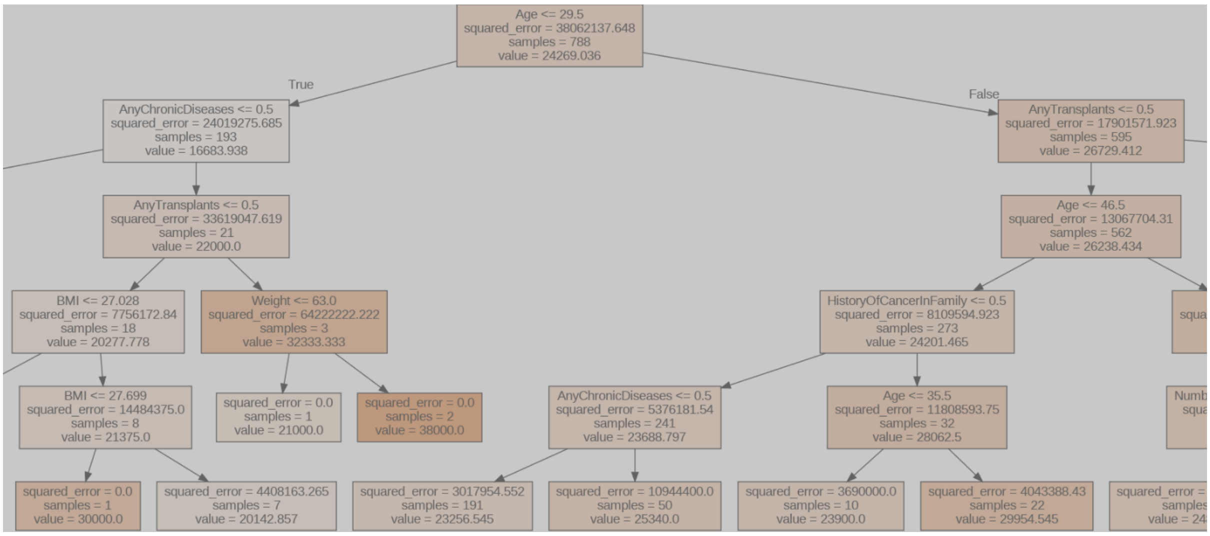
- We will be dropping the columns Diabetes and KnownAllergies from categorical columns and we will drop age and weight from the numerical columns
- Since we are first trying out linear Regression we will use minmax scaler for the numerical columns as the data is not normally distributed in the numerical columns.
- We will now proceed with the model building

Linear Regression model

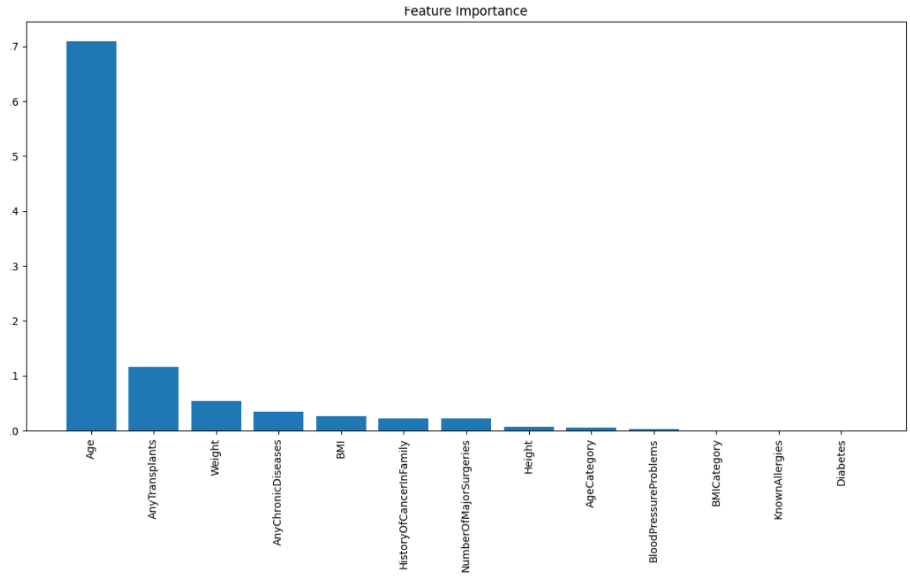
- The first model gave the below output
- Mean Absolute Error: 2602.2319853935646
- Mean Squared Error: 12278380.819083586
- R-squared: 0.7120643285244406
- Coefficient
- Age 10407.421805
- Diabetes -368.149975
- BloodPressureProblems 132.409950
- AnyTransplants 7495.977440
- AnyChronicDiseases 2601.856253
- Height 1128.141161
- Weight 3202.317998
- KnownAllergies 167.264366
- HistoryOfCancerInFamily 2067.559215
- NumberOfMajorSurgeries -2302.325042
- BMI -1023.990803
- AgeCategory 4598.297986
- BMICategory 2491.096366
- We could see that the Linear Regression model performs 71% r2 score.
- We could see that Age and AnyTransplants have higher coeff for model prediction.
- Train Score: 62.82%
- Test Score: 71.21%
- We tried another Linear regression model without the features of Age, wieght and Diabetes we have getting an r2 score of 70%.

- Hence we will go with the previous set of features.
- Now we will check the variance inflation factor
- Below were the outputs of the variance inflation factor
    - Dropping Age with VIF 227.22868793933873
    - Dropping Weight with VIF 129.87993946102108
    - Dropping BMI with VIF 87.30168753884766
    - Final VIF values:
    - Features VIF
    - 0 Diabetes 1.841852
    - 1 BloodPressureProblems 2.068747
    - 2 AnyTransplants 1.062261
    - 3 AnyChronicDiseases 1.243223
    - 4 Height 7.828096
    - 5 KnownAllergies 1.320343
    - 6 HistoryOfCancerInFamily 1.222034
    - 7 NumberOfMajorSurgeries 2.463259
    - 8 AgeCategory 3.566427
    - 9 BMICategory 5.012324
    - From the above VIF factors we can see that these 9 features Diabetes,BloodPressureProblems,AnyTransplant, AnyChronicDiseases, height, knownAllergies, History of Cancer in Family, Number of major Surgeries, Age Category, BMI Category are very important.
- Logistic Regression model:
    - Mean Absolute Error: 1222.2222222222222
    - Mean Squared Error: 8505050.505050505
    - R-squared: 0.8005512726646281
    - Train score DecisionTreeRegressor: 0.8121827411167513
    - Test score DecisionTreeRegressor: 0.7626262626262627
    - Train Score: 70.52%
    - Test Score: 80.06%
    - We could see that the model might slightly overfit for logistic regression.
- Decision Tree Regressor Model
    - Mean Absolute Error: 1383.8383838383838
    - Mean Squared Error: 17070707.07070707
    - Decision Tree R^2: 0.599681295490762

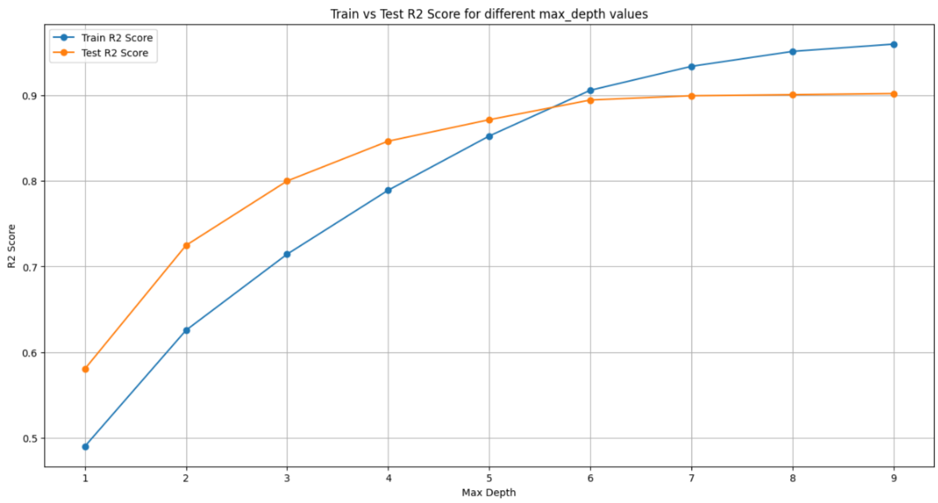
- Cross-validated R^2 scores: [0.54685803 0.62591385 0.72248682 0.69462981 0.67020115]
- Train score DecisionTreeRegressor 100.0 %:
- Test score DecisionTreeRegressor 60.0 %
- For Descision Tree model we could see that the model underfits as the test score is lessor comoared to the trainig score.
- 

- From the above graph we can see at a max depth of 5 we get a decent train and test score around 82% to 84% for Decision Tree Regressor.
- Train Score: 80.96%
- Test Score: 83.5%
- The above is the Decision tree that is built by the algorithm the visual representation is give to show on how the Algorithm has built the tree.
-
- 

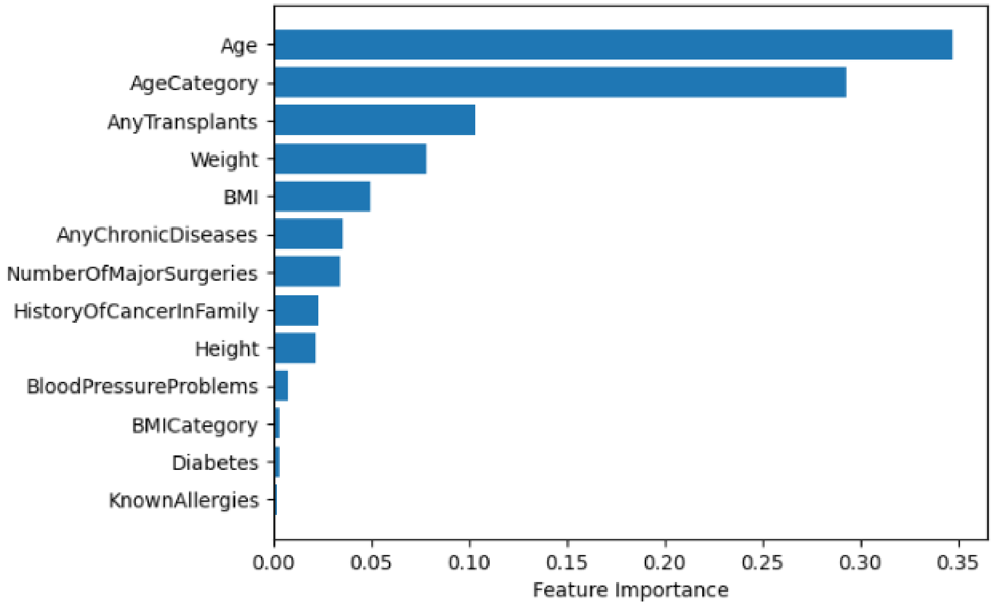
-
-
- From the above it can be see that the age and any transplant are the most important features that are used to build Descision Tree.
- From the above it can be seen that the age and any transplant are the most important features that are used to build Decision Tree.

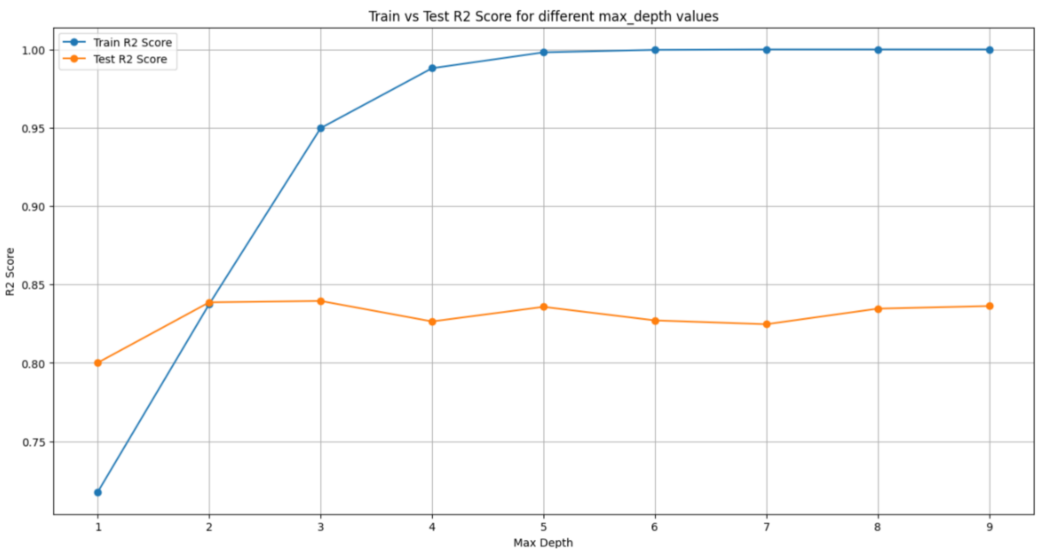


- Random Forest Regressor model
    - Mean Absolute Error: 964.2424242424242
    - Mean Squared Error: 4370732.323232323
    - Random Forest R^2: 0.8975036069598175
    - Cross-validated R^2 scores: [0.78925885 0.75856586 0.80607023 0.89850344 0.68441243]
    - Train Score: 96.71%
    - Test Score: 89.75%
    - From the Random Forest Regressor model we could see that it has 96% train r2 score and 89% r2 score, but we need to check the max depth to find the accurate or better r2 score for train and test.



    - From the above we could see at max depth = 6 we get a better Train r2 score 94% and test r2 score around 90%.
    - I have trained with some other variations but the above model becomes the best fit and we will use this to predict the output.
    - Mean Absolute Error: 1048.5304291278965
    - Mean Squared Error: 4294099.743993234

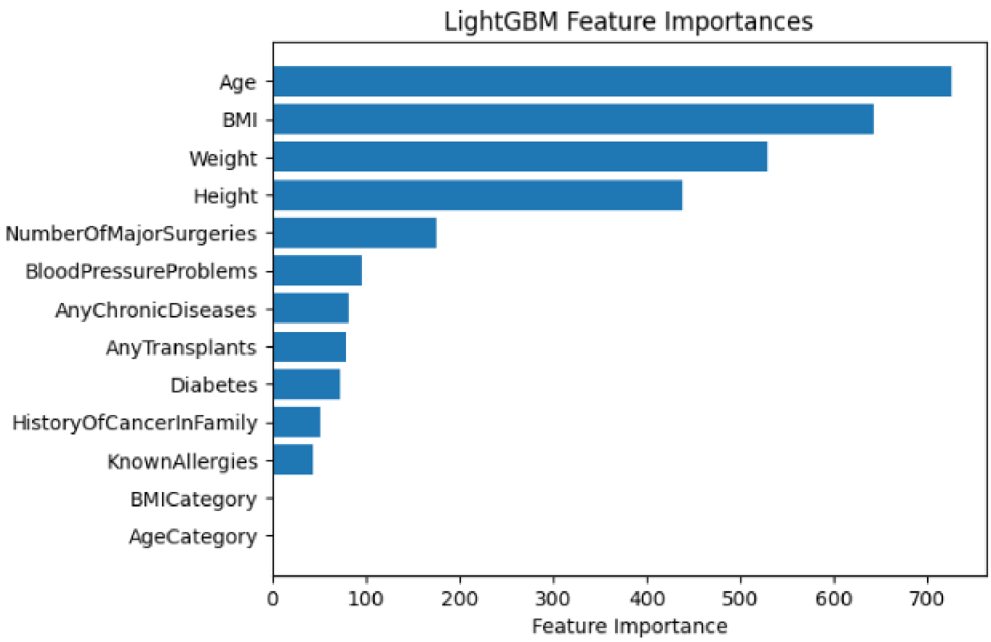
- Train Score: 93.36%
- Test Score: 89.93%
- From the above testing of various hyperparameters we could see that Random forest regressor performance works better with max depth 7
- 

- **XGB Regressor model**
  - XGBoost Regression R^2: 0.8357788324356079
  - Cross-validated R^2 scores: [0.78377521 0.7138021 0.77271008 0.85943568 0.64083159]
  - Mean Absolute Error: 1362.5696614583333
  - Mean Squared Error: 7002848.747049428
  - Train R^2 Score XGBRegressor:100.0%
  - Test R^2 Score XGBRegressor:84.0%
  - Now we built XGBoost Regression with train r2 score 100% and test r2 score of 84%. But we could see there is some overfit. lets check the depth at which we get a optimal performance.
  - 

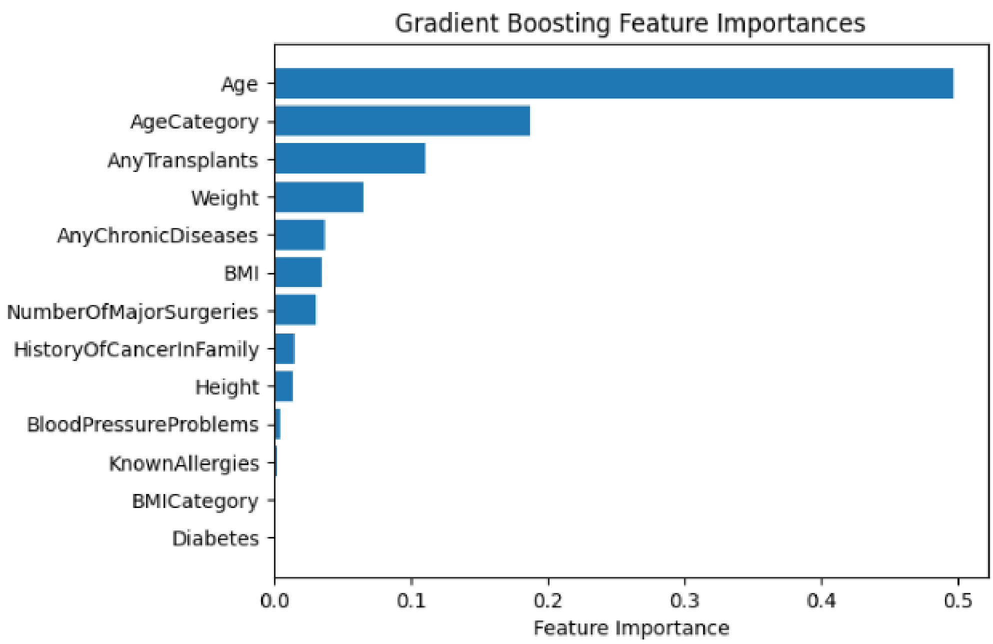
  - Train Accuracy: 83.75%
  - Test Accuracy: 83.87%

- From the above we can see we get an optimal performance with max depth 2 with a train and test r2 score around 84%. Still slightly lesser than Random Forest Regression.

- **Ridge and Lasso Regression**
    - Ridge Regression R^2: 0.7109516446662018
    - Cross-validated R^2 scores: [0.62612757 0.6491237 0.65471917 0.75020615 0.4973608]
    - Mean Absolute Error: 2607.6006163653583
    - Mean Squared Error: 12325828.764913611
    - Train R^2 Score Ridge: 0.63 %
    - Test R^2 Score Ridge:0.71 %
    - Ridge is still showing a lower train and test r2 score. Hence we will not consider it for model building
    - Lasso Regression R^2: 0.7118915588787265
    - Cross-validated R^2 scores: [0.62638404 0.64948227 0.65364123 0.75149942 0.4957133]
    - Mean Absolute Error: 2603.0205678018433
    - Mean Squared Error: 12285748.199072268
    - Train R^2 Score Lasso:0.63 %
    - Test R^2 Score Lasso:0.71 %
    - We are seeing Lasso also performing train and test r2 score similar to Ridge.
- **SVR Regression**
    - SVR Regression R^2: -0.05844978479583074
    - Cross-validated R^2 scores: [-0.01124829 -0.10335441 -0.02909925 -0.07056746 0.0001163]
    - Mean Absolute Error: 5327.939509270694
    - Mean Squared Error: 45135253.541183464
    - Train R^2 Score SVR:-0.04 %
    - Test R^2 Score SVR:-0.06 %
    - SVR is performing the lowest hence we are neglecting it for model building.
- **LightGBM Regression**
    - LightGBM Regression R^2: 0.8771448566922515
    - Cross-validated R^2 scores: [0.8153393 0.73587743 0.81612337 0.86924939 0.67117122]
    - Mean Absolute Error: 1421.4233478279525
    - Mean Squared Error: 5238886.267148964
    - Train R^2 Score LightGBM:0.93 %
    - Test R^2 Score LightGBM:0.88 %
    - Light GBM Regression is performing with 93% r2 score for train and 88% r2 score for test.



- **Gradient Boosting Model**
    - Gradient Boosting R^2: 0.8679253236809571
    - Cross-validated R^2 scores: [0.72490308 0.72675346 0.77913785 0.88322191 0.65858112]
    - Mean Absolute Error: 1521.7125687486919
    - Mean Squared Error: 5632032.8915552925
    - Train R^2 Score GradientBoostingRegressor: 0.89 %
    - Test R^2 Score GradientBoostingRegressor:0.87 %
    - From all the above Regressor model testing we could see Random forest Regressor, Gradiant boosting regressor, lightGBM have given r2 score greater than 85% we will combine the output of the 3 model to predict the insurance price.



Insights From Shape Graph

- For the Random forest Regression model the SHAP graphs tells us that Age, Age category, Any transplants ,weight and BMI have the highest importances.
- For Light GBM regression model the SHAP graphs tells us that Age, BMI, weight, height, Number of Major Surgeries are given more importances to predict the model output.
- For Gradient Boosting Regression model the SHAP graphs tell us that Age, Age Category, Any transplants, weight.

Our model will be built in such a way we will combine the average output from random forest regressor and Gradient boosting regressor. We are using pickle to package the model. We will use streamlite to build the working app.

Recommendations for Insurance Companies

- With the findings from data analysis and model insights, here are some actionable recommendations for insurance companies:
- Target Healthier Applicants: A significant portion (50%) of applicants does not report major health issues, suggesting a stable premium segment. Encouraging younger, healthier individuals to purchase insurance could increase customer retention and reduce risk.
- Adjust Premiums Based on Transplants: Individuals with organ transplants bear significantly higher premiums. Implementing a dynamic premium adjustment could help insurers manage the risk associated with transplants better.
- Focus on High-Risk Age Groups: Targeting age segments with higher premiums, particularly individuals over 50, could optimize insurance offerings. Discounts for early purchase (e.g., ages 18-30) might also encourage long-term policy uptake.
- BMI-Linked Premium Adjustments: Since BMI influences premium amounts, insurers could adopt tiered premiums based on weight categories, rewarding healthier lifestyles.
- By refining these strategies, insurance companies can offer more personalized and accurate premium rates, ensuring both profitability and customer satisfaction.

Link for :tableau dash board: https://public.tableau.com/app/profile/arivalagan.raghavan/viz/shared/6G69RBR3N

https://public.tableau.com/app/profile/arivalagan.raghavan/viz/Insurancepriceprediction_Arivalagan_2/ICPDashboard1

Link for Git Hub: https://github.com/ArivalaganRaghavan/Insurance_price_predictor_model

Link for colab : https://colab.research.google.com/drive/1MKRAJJKtIDFIXIN4X81CPBnGoaehpOy1?usp=sharing

Link for Medium article: https://medium.com/@arivalagan.rxprism/insurance-price-prediction-analyzing-key-factors-and-building-an-effective-predictive-model-7f7b1dffe682

Link for loom video : https://www.loom.com/share/6bd98b15636e4939804326a1532ed424?sid=938a0c31-0dfe-429e-a91b-cc480a370b62