

Medical insurance cost analysis

Methodology document

1. Data sourcing

- The dataset is downloaded from [Kaggle](#).
- Saved the metadata in a separate sheet of the workbook.
- Got familiarized with the columns and metrics that are in the data set

2. Data preparation

- This dataset is imported to Jupyter notebook.
- Data wrangling:
 - Creation of categories for age, BMI.
 - Converted string values to binary values for the variables – sex, smoker.
 - Converted the column children into int.
 - Scaled variables for building linear regression model

3. Data analysis

- The modified data set is imported in tableau for further analysis.
- Linear regression model is built for predicting charges.
 - The final model has 78% R-square value.

4. Report and recommendation

- Rising rate of old teen and young adult smokers should be given awareness.
- Beneficiaries having more dependencies incur heavy bills.
- Unhealthy lifestyle leads is an early indicator of financial burden on the health front.