

Predicting insurance cost based on many factors

Content

Python file in Jupyter Notebook:

- Imported all the libraries.
- Imported CSV .
- Cleaned data (removing missing values, removing duplicate values).
- Checked for correlation between prices and other factors.
- Plotted data for visualization and debugging.
- Checked for outliers and decide to remove outliers as they are only 4 outliers.
- Categorized BMI into tiers.
- Train, Test, Split data to be used in linear regression model.
- Encoded values into numbers to be used in machine learning.
- Developed some logic to input data into the linear regression equation and output predicted value.

Created Heroku account :

- Deployed UI web app to Heroku via Github connection.
- Attempted to connect the web app to the Python script via API's but I ran into many problems. I should have just done it using Python Flask and connected the entire Python to Heroku but I ran out of time.
- However, Nodes App have been deployed to Heroku (no interaction as ran out of time to build Flask).

<https://insurance-costing-ui.herokuapp.com/>
