MediTech is one of the top pharmaceutical laboratories of the US. They are working on a new medical procedure to treat diabetes and have partnered with major hospitals to target potential patients. They approach every patient and set up sessions to make them aware of this procedure, but it is expensive.

They have collected data from various hospitals and thousands of patients they approached. Using machine learning, build a model to predict if a patient will try their procedure and analyze what factors affect it.

Files

- train.csv
- test.csv
- sample_submission.csv

Problem

Perform an analysis to determine how different features are related to the target variable i.e. *opted*. Build a machine learning model that can predict whether the patient will opt for the procedure.

For each record in the test set (test.csv), predict the value of the *opted* variable. Submit a CSV file with a header row plus each of the test entries on its own line. The file (submissions.csv) should have exactly 2 columns:

- id
- opted (Whether the patient opts for the procedure or not)

Evaluation Metric

The evaluation metric is Accuracy.

```
Accuracy = number of correct predictions/total number of predictions
```

Deliverables

- Well commented Jupyter notebook
- “submissions.csv”

Explore the data, make visualizations, and generate new features if required. Make appropriate plots, annotate the notebook with markdowns and explain necessary inferences. A person should be able to read the notebook and understand the steps taken and the reasoning behind them. **The solution will be graded based on the usage of effective visualizations to convey the analysis and the modeling process.**