**Project**

In this project, you will analyze a dataset and present a narrative of your recommendations and findings. The background of the dataset: Amrotel Inc has two spectrums of business: commercial business and government healthcare (Medicare & Medicaid). For Medicare, it is similar to car/home insurance in which you need to pay for your monthly premium. We collected some demographic information from our insured customer interactions via different channels (e.g. Online chat, email, mail, phone, etc.), along with their premium payment activities in the past 3 and 6 months. You are asked to use provided dataset to predict how likely the customer will renew his/her policy thru the phone call.

The description of each feature in the data set is detailed as follows:

|  |  |
| --- | --- |
| Column name | Description |
| RTD\_ST\_CD | rated state of policy, i.e. ST\_S0 = Alabama, … |
| CustomerSegment | segmentation of customer |
| Tenure | how long the policyholder stays with Anthem |
| Age | policyholder's age |
| MART\_STATUS | policyholder's marital status |
| GENDER | policyholder's gender |
| CHANNEL1\_6M | how many times the customer contacted Anthem thru channel 1 in the past 6 months |
| CHANNEL2\_6M | how many times the customer contacted Anthem thru channel 2 in the past 6 months |
| CHANNEL3\_6M | how many times the customer contacted Anthem thru channel 3 in the past 6 months |
| CHANNEL4\_6M | how many times the customer contacted Anthem thru channel 4 in the past 6 months |
| CHANNEL5\_6M | how many times the customer contacted Anthem thru channel 5 in the past 6 months |
| METHOD1\_6M | how many times the customer used the method 1 to make premium payments in the past 6 months |
| RECENT\_PAYMENT | binary indicator if making premium payment in the past 10 days |
| PAYMENTS\_6M | how many times the customer made premium payments in the past 6 months |
| CHANNEL1\_3M | how many times the customer contacted Anthem thru channel 1 in the past 3 months |
| CHANNEL2\_3M | how many times the customer contacted Anthem thru channel 2 in the past 3 months |
| CHANNEL3\_3M | how many times the customer contacted Anthem thru channel 3 in the past 3 months |
| CHANNEL4\_3M | how many times the customer contacted Anthem thru channel 4 in the past 3 months |
| CHANNEL5\_3M | how many times the customer contacted Anthem thru channel 5 in the past 3 months |
| METHOD1\_3M | how many times the customer used the method 1 to make premium payments in the past 3 months |
| PAYMENTS\_3M | how many times the customer made premium payments in the past 3 months |
| LATE\_PAYMENT\_3M | binary indicator of late payment in the past 3 months |
| LATE\_PAYMENT\_6M | binary indicator of late payment in the past 6 months |
| EVENT1\_30\_FLAG | binary indicator if Event 1 happened in the past 30 days |
| EVENT2\_90\_SUM | aggregated sum of how many times Event 2 happened in the past 90 days |
| LOGINS | how many times the customer log in his/her Anthem account |
| Call\_Flag | binary indicator if the customer will call us to renew his/her policy |

Please use Python to analyze the data, evaluate the performance of your model and write an explanation of your findings. As the final deliverable, please make at most 10 slides to summarize your model and send your code separately. The slides should include a short narrative for each of the following areas:

1. Explanatory data analysis
2. Data preprocessing and feature engineering
3. Modeling & evaluation
4. Final conclusion

Please also include plots and tables that highlight the insights you found.

**Project Group Requirements:**

1. Group project team members = 2
2. Presentation
3. Project Report (Includes code, graphs, your data analytics recommendations)
4. Final project weight = 150 Points
5. Class Presentations and report submission date = TBD