Allstates Prediction Analysis

Predicting of policy purchase

Executive Summary

- Decision tree:
  - the allocation of the largest % of purchase cases is mostly characterized by **State, Feature G, and Feature A**;
  - (1-6) range of Shopping Point is characterized by low percent of purchase cases;
  - the next states expose lower % of purchase cases than others: (FL, MT, UT, WA, NE, ME).
- Regression:
  - our most potential customer is thinking of purchasing the G-4 feature;
  - he has a record of the previous policy duration,
  - he lives in CT, MD, SD, RI, and WV;

Business Understanding

Goal:

- Improvement of the targeting of the audience;
- Increasing sales rates;

Objective:

- Creation of a predictive model for forecasting policy purchasing a on the other characteristics;

Questions:

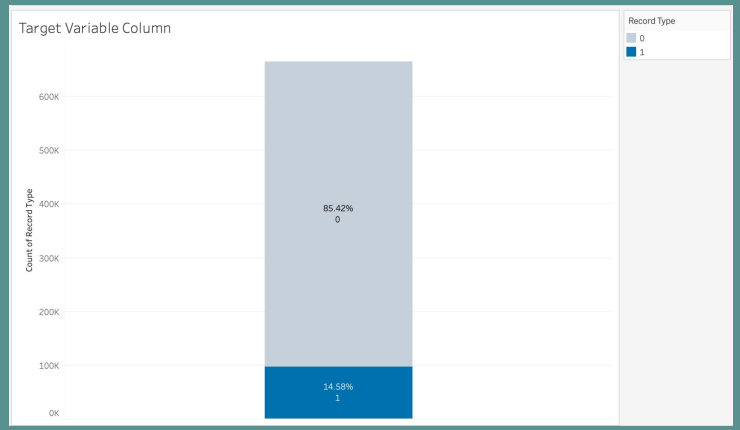
- What defines a potential customer mostly ?
- What are the features of the policy that could impact a customer's decision to purchase insurance?
- What increases the likelihood of purchase mostly?
- Who are our target potential customers in terms of the odds of purchase?

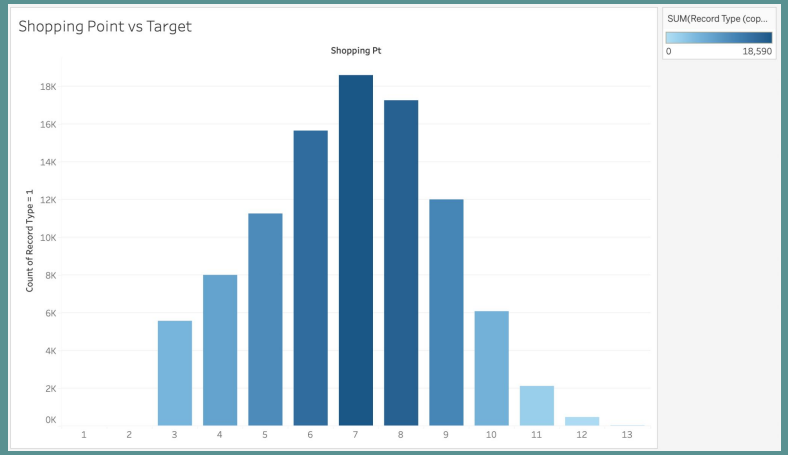
Data Preparation

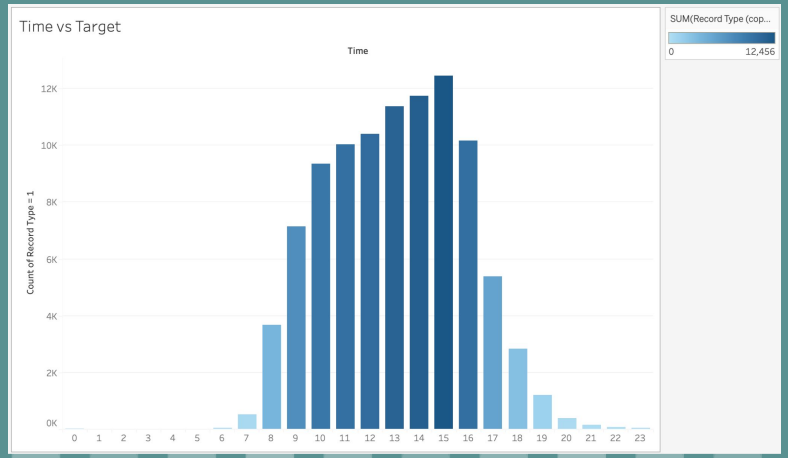
- Changing the Time format (HH:MM) -> (HH)

- Rejecting the Location

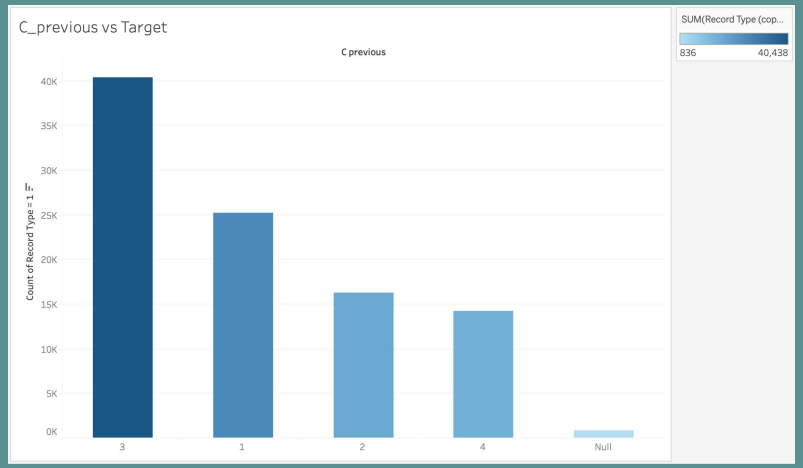
Preliminary Examination

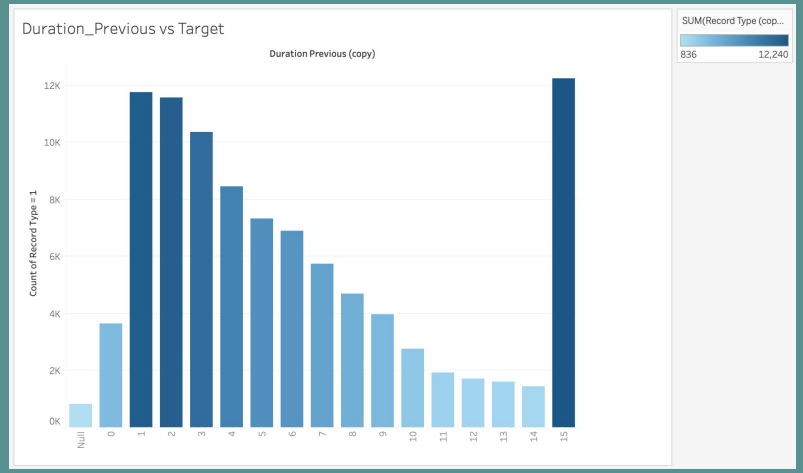


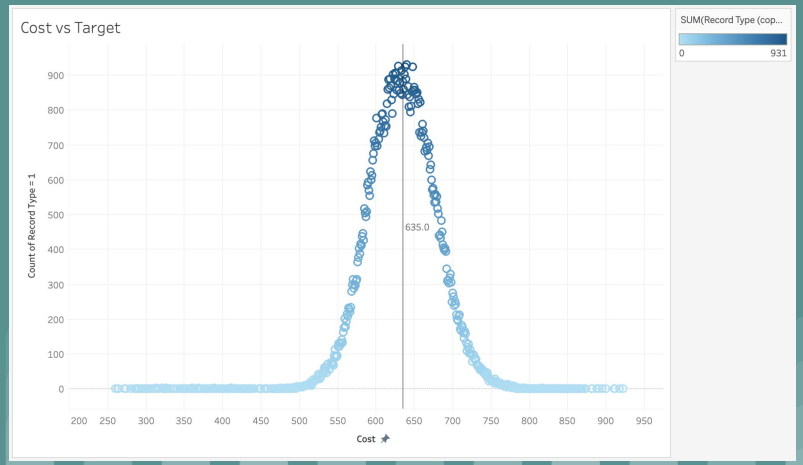




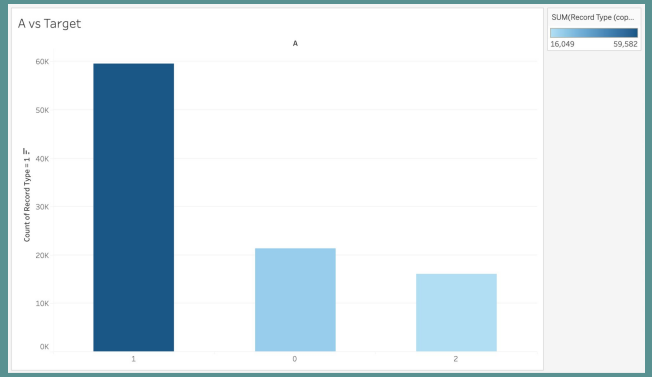
Preliminary Examination

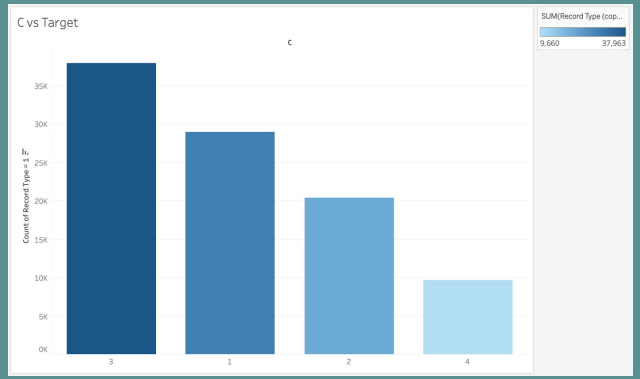


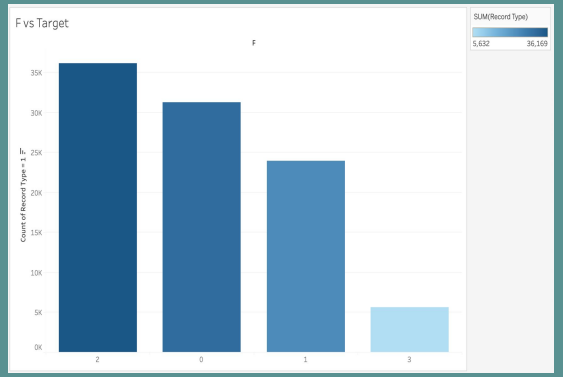


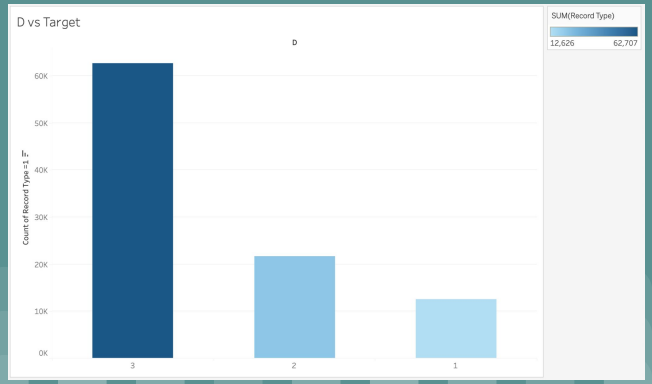


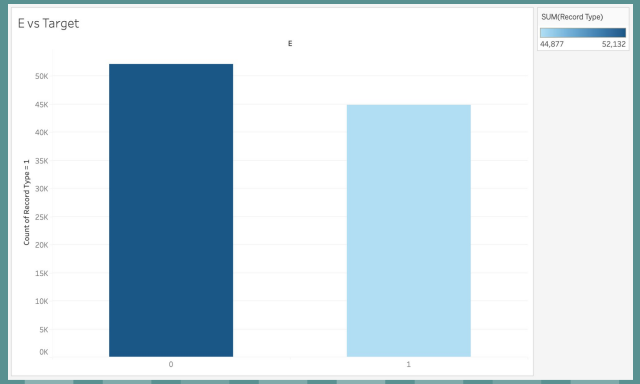
Preliminary Examination

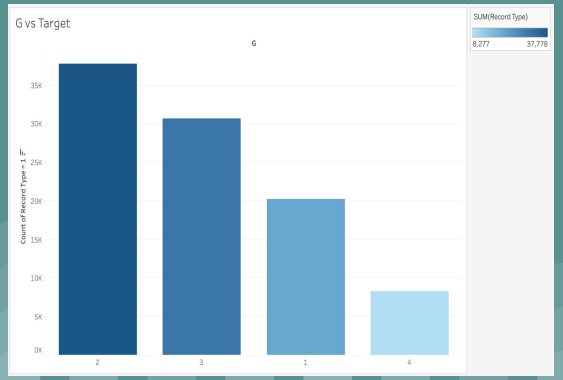












Decision Tree

Model Comparison

Technique	Sub-technique	Validation Misclassification Rate	Average Square Error Rate	Number of Leaves	Relative Accuracy (1-4) (higher the better) - based on Validation Misclassification & AVG Sq Er	Relative Complexity (1-4) (higher the better) - based on Number of Steps
Decision Tree	Interactive DT, Assessment measure: Default	0.132458	0.100453	6	1	4
Decision Tree	Interactive DT, Assessment measure: Decision	0.132458	0.100453	7	1	4
Decision Tree	Interactive DT, Assessment measure: AVG Square Error	0.132458	0.100453	7	1	4
Decision Tree	Maximal Tree, Assessment measure: Decision	0.132458	0.097437	36	4	1
Decision Tree	Maximal Tree, Assessment measure: AVG Square Error	0.132458	0.097437	36	4	1
Decision Tree	Optimal Tree, Assessment measure: Decision	0.132458	0.100453	7	1	4
Decision Tree	Optimal Tree, Assessment measure: AVG Square Error	0.132458	0.100453	7	1	4

Decision Tree

Results

The list of the variables that define the purchase case to a larger extent :

- Shopping point → (1-6) range is characterized by low percent of purchase cases;
- State → FL, MT, UT, WA, NE, ME expose lower % of purchase cases than others;
- G → option 1 for this product show higher % of purchase cases;
- A → option 0 for this product show higher % of purchase cases;

Regression

Model Comparison

Technique	Sub-technique	Validation Misclassification Rate	Number of Iterations	Relative Accuracy (1-4) (higher the better) - based on Validation Misclassification	Relative Complexity (1-4) (higher the better) - based on Number of Iterations
Logistic Regression	Forward Regression	0.131713	17	3	1
Logistic Regression	Forward Regression, Best Sequence Model	0.131517	13	4	2
Logistic Regression	Stepwise Regression	0.131944	9	1	3
Logistic Regression	Stepwise Regression with expanded Entry/Stay Levels	0.131944	9	1	3
Logistic Regression	Stepwise Regression with expanded Entry/Stay Levels, Best Sequence Model	0.131857	6	2	4

Regression

Results

Description of one who exposes the highest odds of purchase:

- This person is thinking of purchasing the option 4 of G feature;
- He has a record about the previous policy duration;
- He lives in CT, MD, SD, RI, and WV;

(In addition, he comes to purchase closer to the night, but it is a controversial point because there is no guarantee that one who purchases at night is not one who tries to buy a policy in a hurry right after the incident has happened.)

Conclusions

Actions Done

- 3 different techniques were used: **Decision Tree Modeling**, **Regression Modeling**.
- Different subtypes of modelings were conducted and examined for picking the most optimal model in terms of complexity and accuracy.
- The results were interpreted in a non-technical way so that Allstates can use them (the conclusions and the models themselves) for reaching the primary goal: targeting the audience and increasing sales rates.

Results

- Decision tree:
  - the allocation of the largest % of purchase cases is mostly characterized by **Shopping Point, State, Feature G, and Feature A**;
  - the next states expose lower % of purchase cases than others: (FL, MT, UT, WA, RI, NE, ME).
- Regression:
  - our most potential customer is thinking of purchasing the G-4 feature;
  - he has a record of the previous policy duration,
  - he lives in CT, MD, SD, RI, and WV;