Predict Market competitiveness

Problem statement

Predict market share of an insurance product using its features and current number of customers enrolled in each plan.

Description

A major criterion for a person choosing an insurance plan is the benefits it offers and the cost at which they are offered. However, it is hard to determine the impact of each benefit on the market and the enrolments it drives. The product design helps insurance providers to design their plans to capture the maximum market share. The objective of this problem is to predict the market share of all plans using their features, to predict the performance of each plan.

The data used for this problem statement is the publicly available data of the Medicare industry in the US. Medicare is the federal insurance provided to people aged 65 and above. All individuals above 65 years of age are eligible to enrol into an insurance plan. This data is released yearly by the Center for Medicare and Medicaid Services (CMS) which contains the description of plans, the regions it is offered in and the benefits each plan offers. Each plan offers certain benefits like hospital admission, skilled-nursing, etc. and services like X-Ray, Chiropractor, drugs etc.

The definition of domain specific terms that you will come across can be found here:
https://www.medicare.gov/glossary/a

Data:

The data is structured in the following files:

1. Landscape file: Contains the high-level information of the all the plans, overall cost features, Counties it is offered in and the enrolments of a plan in each market (County). The Bid ID equivalent can be created by concatenating "Contract ID_Plan ID_Segment ID"

2. Benefit files: All the files with filename format pbp_b* contain the benefits for each plan. This file is at a 'Bid ID' level and it describes the benefits that the insurance plan offers.

Note:

*   A plan is identified by the Bid ID, and the plan features in the benefit files will be unique for a Bid ID

*   A plan (Bid ID) can be offered in multiple Counties across the US

*   Two counties in different states can have the same name, so keep County coupled with the State to maintain the integrity of the data and your analysis.

*   In the benefits file, the Bid ID is the primary key.

End of Note

Example of a benefit file:

The Benefit dictionary provides the explanation of all the fields in each benefit file. To get an overview of the fields in the dictionary, consider benefit number 10 which is 'Ambulance and Transport'.

In the dictionary,

- FILE corresponds to the name of the benefit file (pbp_b10_amb_trans->Ambulance and Transport)

- NAME consist of all the variables in that benefit.

- Field _Title gives the description of the column.

- Type gives the data type of the field

- Length is the length of the value in that field.

- Codes is the variable whose values are shown in the columns of the corresponding benefit.

- Code_value explains the meaning of each code value.

Example-

- The column 'pbp_b10a_maxenr_yn'(NAME) can have 1 or 0 as value where:

1 -> Yes

2 ->No

- Similarly, for column ' pbp_b10a_maxenr_gas_per' can have values from 1 to 6 where-

1---->Every Three Years

2---->Every two Years

3---->Every Year

4---->Every six months

5---->Every Three months

6---->Other,Describe

- For column 'pbp_10a_ded_ehc', signifies the deductible for enhanced benefits. There are 2 such enhanced benefits, each signified by a truthness flag.

In this instance,

- 1 in 1 corresponds to 1 being at first index, which means 'there is a deductible for Medicare covered Ground Ambulance'.

- 1 in 2 corresponds to 1 being at second index, which means 'there is a deductible for Medicare covered Air Ambulance'.

These are indicator variables clubbed together to form the field 'pbp_10a_ded_ehc'.

- There are certain variables which have certain characteristics as well. For instance:

Pbp_b10b_bendesc_trn: This can either have value as 1 (corresponding to Plan-Approved Location) or 2 (corresponding to Any Health-related Location). The variables ending in 'pal' are for Plan-Approved Location while variables ending in '_al' are for Any Health-relation Location.

pbp_b10b_bendesc_amo_pal

pbp_b10b_bendesc_lim_pal

pbp_b10b_bendesc_amt_pal

pbp_b10b_bendesc_per_pal

pbp_b10b_bendesc_tt_pal

pbp_b10b_bendesc_amt_pal_days

pbp_b10b_bendesc_mt_pal

pbp_b10b_bendesc_amo_al

pbp_b10b_bendesc_lim_al

pbp_b10b_bendesc_amt_al

pbp_b10b_bendesc_per_al

pbp_b10b_bendesc_tt_al

pbp_b10b_bendesc_amt_al_days

pbp_b10b_bendesc_mt_al

Tools:

Any open-source tool can be used. No extra weightage will be given for using a certain tool.

Judging Criteria:

Your performance will be judged based on accuracy of market share prediction and the quantified variable impact on the enrolment.

Market share is defined as the percentage of enrolments each insurance plan captures in each market. For this problem statement, consider County as a market. The participants are expected to predict the market share for each plan in each county (independent market).

$$\text{Market share} = \text{plan enrolment (county)/total county enrolment}$$

Mean Absolute Percentage Error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$\text{MAPE\_Plan} = (PredictedMarketShareOfPlan*TotalCountyEnrolments - AcutalEnrolmentForPlan)/ActualEnrolmentForPlan$$

$$\text{MAPE\_County} = (SumOf(PredictedMarketShareOfPlan*TotalCountyEnrolments) - AcutalEnrolmentForCounty)/ActualEnrolmentForCounty$$

$$\text{MAPE\_State} = (SumOf_{forEachState}(PredictedMarketShareOfPlan*TotalCountyEnrolments) - AcutalEnrolmentForState)/ActualEnrolmentForState$$

Submission files:

- Submission_plan.csv: State, County, Bid_ID, MAPE

- Submission_county.csv: State, County, MAPE

- Submission_state.csv: State, MAPE

The judgement would be based on 2 levels:

1. MAPE (85%): We would be looking at the average MAPE at the three levels, namely Plan, County and State and your final score would be an weighted average with the following weightages:

- MAPE by plan size (45%):

  o Average of MAPE of plans (0 <= enrollment < 1000): 15%

  o Average of MAPE of plans (1000 <= enrollment < 5000): 15%

  o Average of MAPE of plans (5000 <= enrollment < 10000): 10%

  o Average of MAPE of plans (10000 <= enrollment): 5%

- Average of MAPE of County: 20%

- Average of MAPE of State: 10%

- Average of MAPE of Parent Organization: 10%

2. Quantified variable impact/importance on the enrollment (15%)