Business Analytics Consultant Project for UMD Alumni Association

UNIVERSITY OF MARYLAND

Team Information

Team 4 (Project_0506_04) Red Turtle

Adarsh Challa
George Puthean
Po-Han Yen (Team Leader)
Yi-Wei Sun

Dec. 3, 2021

Mission Statements

- Help the UMD alumni association find ways to attract more first time attendees and major prospects.

- Predict the percentage of first time attendees and major prospects based on the characteristics of the event that is going to be held.

Mission Objectives

- When do the events have higher percentage first time attendees or major prospects?
- What types of events have higher percentage first time attendees or major prospects?
- Which locations of the events have higher percentage first time attendees or major prospects?

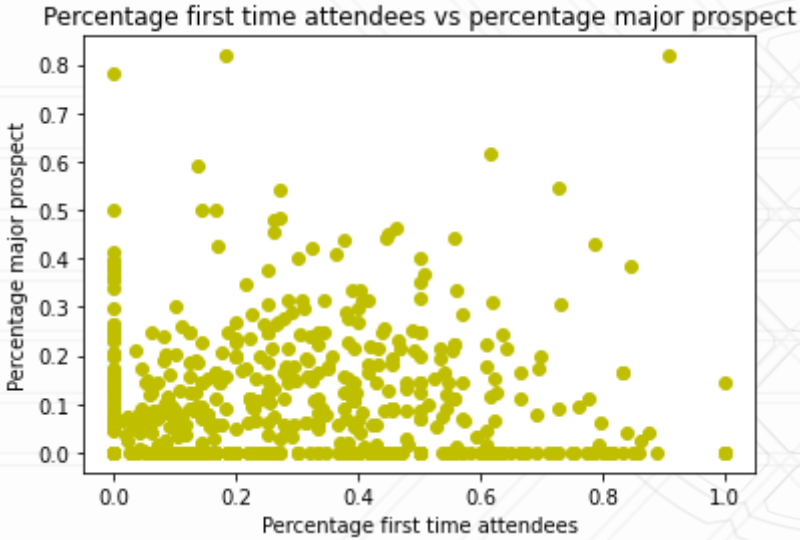
Methods - Outlines

Step 1: Divide the data based on the target value we want to analyze
Step 2: Perform exploratory data analysis respectively on numerical data, categorical data and time series data.
Step 3: Aggregate categorical data into smaller groups and analyze them.
Step 4: Identify relevant data, preprocess and split them for model training and evaluating.
Step 5: Build, train, test and evaluate the predictive model.

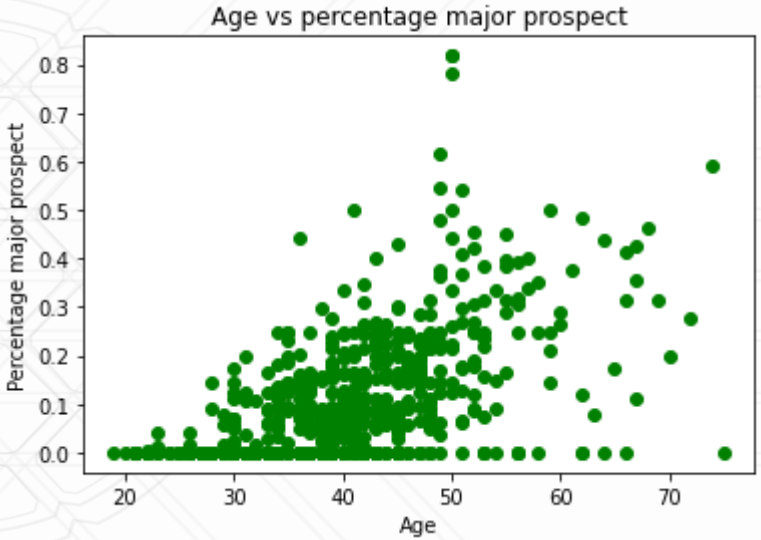
Methods - details

- Split the data into two, one for analyzing PFTA the other for PMP. PFTA one does not include the first year 2013.
- Adding month, year and weekday columns for time series analysis.
- Group by month, year, weekday, group code and location code respectively, and perform aggregate calculations.
- Further group location, group and activity into smaller groups based on similarities.

Findings - Numerical data

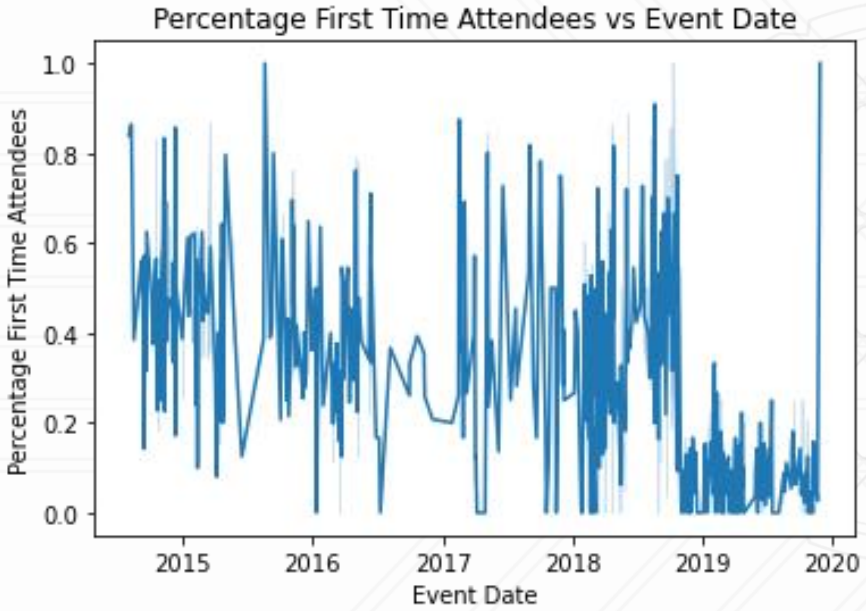


Percentage major prospect & percentage first time attendees does not highly correlated.

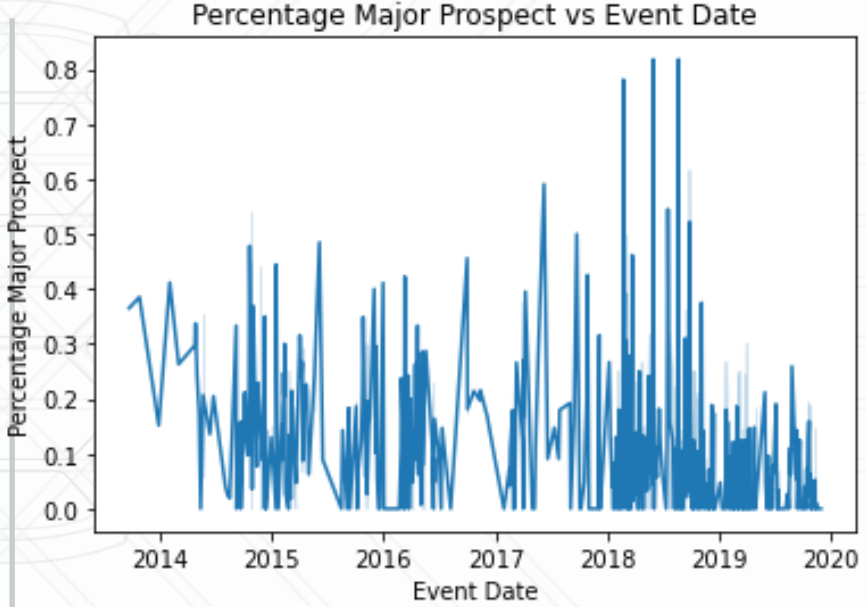


linear relationship between Average Age and percentage major prospect.

Findings - Time Series data

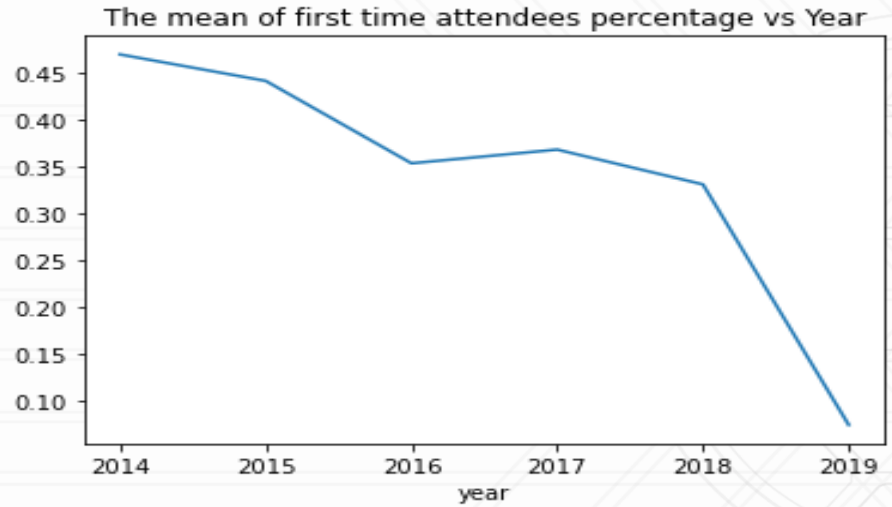


Lots of events with low PFTA occur during 19-20

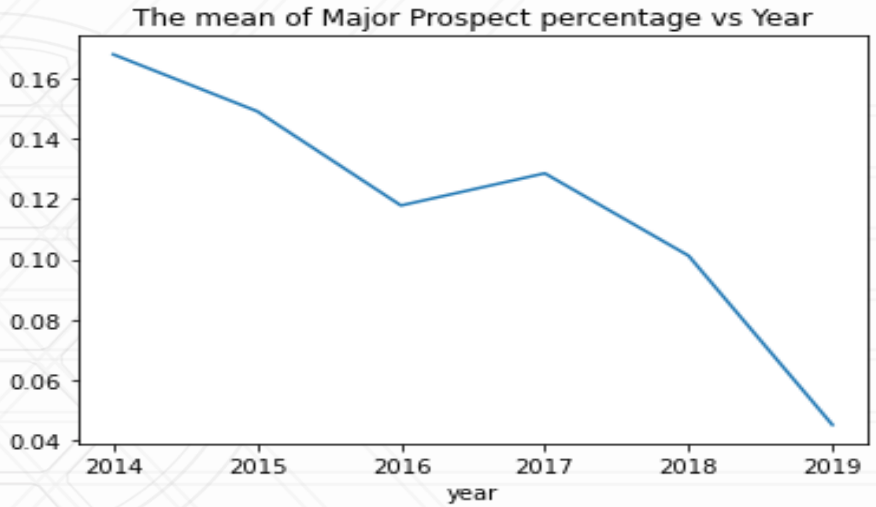


Events with the highest PMP occur during 18-19

Findings - Time Series data

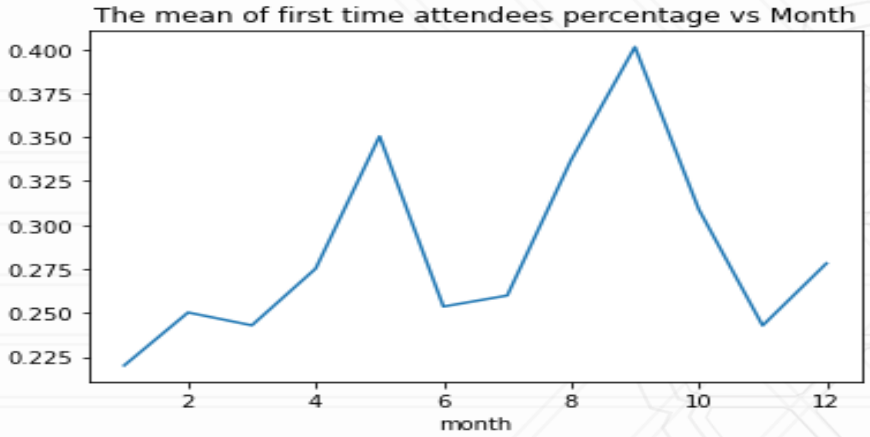


Concerning downward trend.

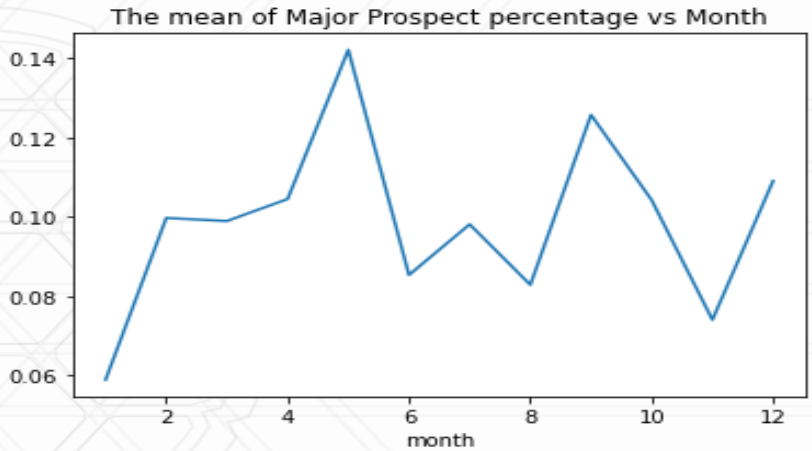


Concerning downward trend.

Findings - Time Series data

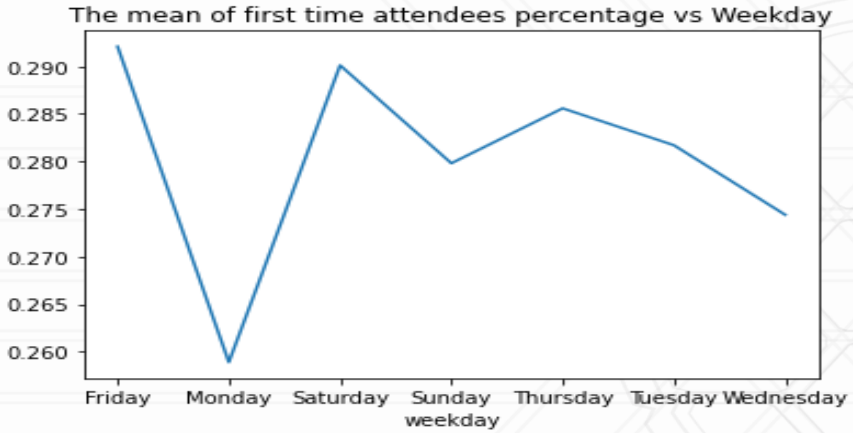


May and September have the highest percentage first time attendees!
January, March and November have the lowest percentage first time attendees!
The difference is over 10% which is significant.

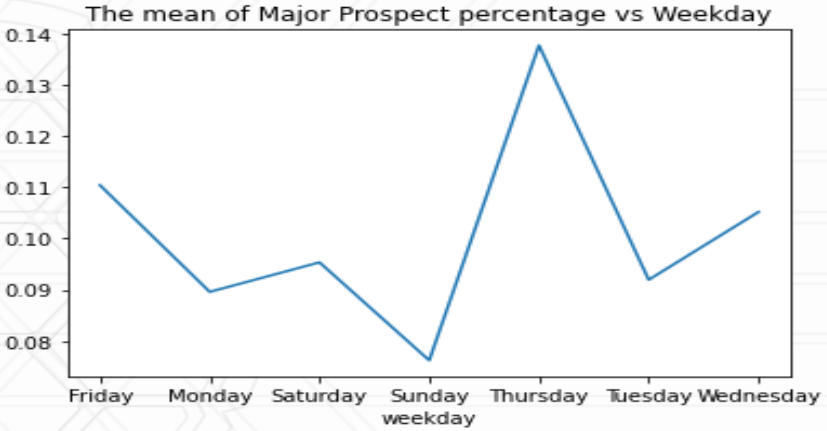


May and September also have the highest percentage **major prospect!**
January and November have the lowest percentage major prospect!
The difference is about 8%.

Findings - Time Series data

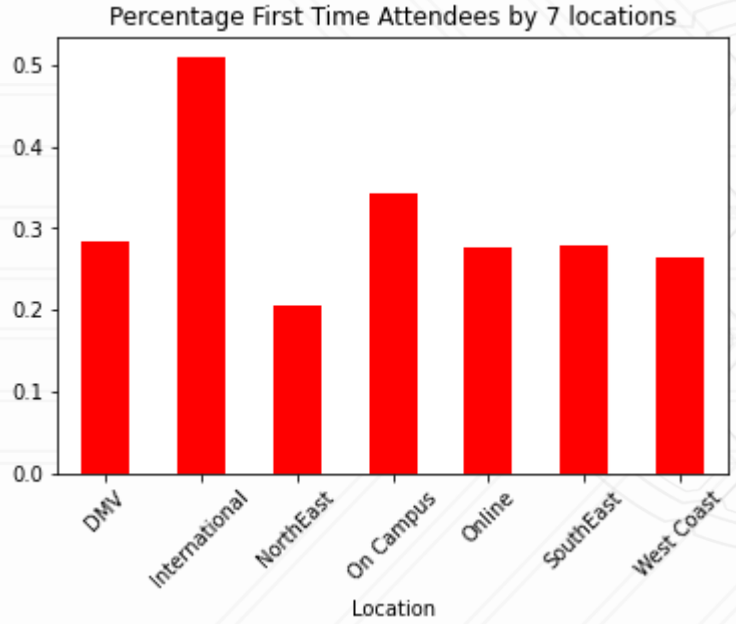


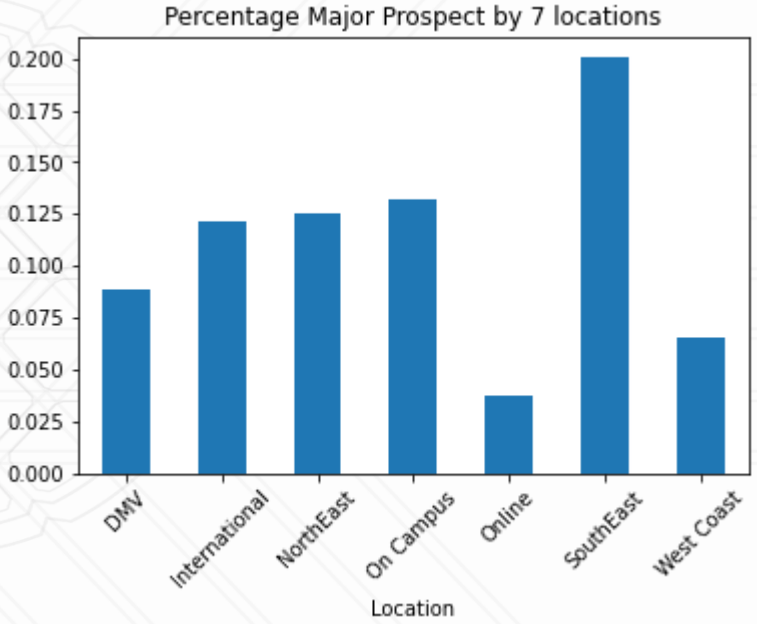
Friday and Saturday have the highest percentage first time attendees!
Monday has the lowest percentage first time attendees!
But the difference is only 3% which is not significant.



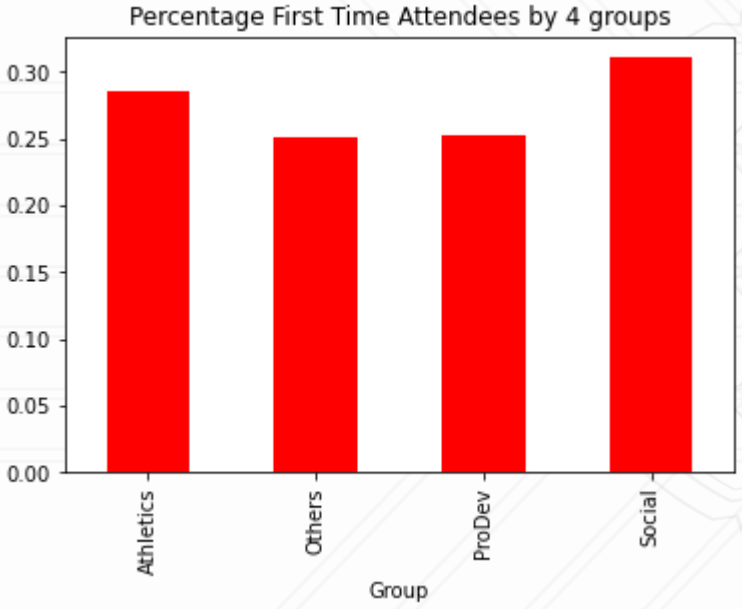
Thursday has the highest percentage major prospect!
Sunday has the lowest percentage major prospect!
The difference is about 6%.

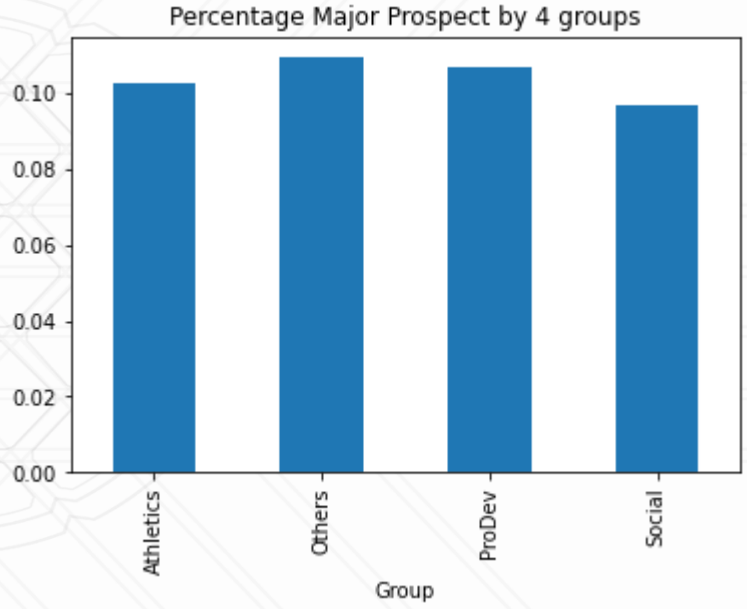
Findings - Categorical data I



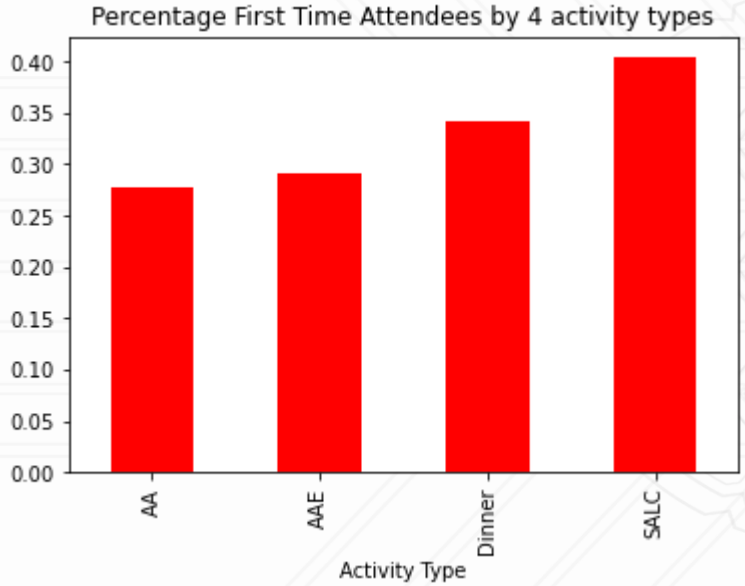


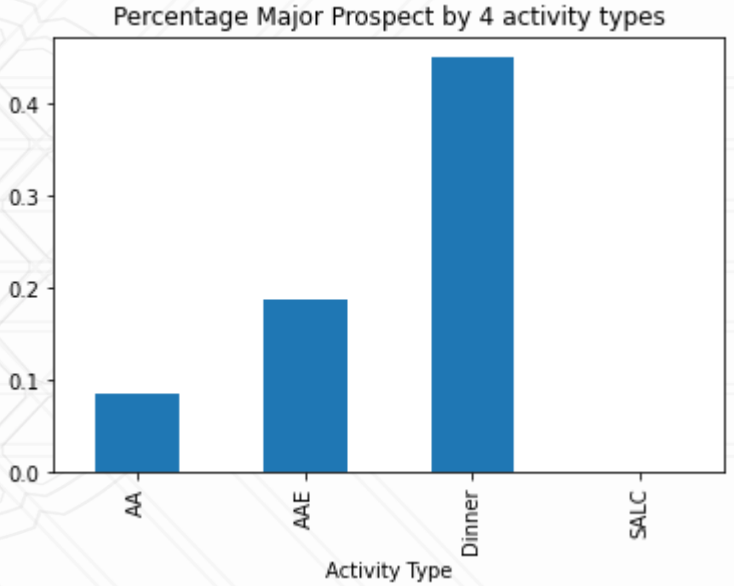
Findings - Categorical data II





Findings - Categorical data III





Recommendations

To increase percentage first time attendees:

I. Events should be held on May, August, September and October.

II. Preferable locations are international and on campus.

III. Preferable types are social and athletics.

IV. Preferable activity -- SALC.

Recommendations

To increase percentage major prospect:

I. Events should be held on May and September, preferably Thursday.

II. Try hold at South East region and prevent online events.

III. Try to attract alumni with higher age.

IV. Preferable activity -- Dinner. Unpreferable -- SALC, AA.

Linear Regression model

Identify the attributes that are highly correlated to our target value PFTA or PMP.

↓

Convert categorical data into dummy variables.

↓

Build and train two linear models respectively.

↓

Evaluate models based on MAE.

PFTA model :
['Participated', 'Average Age', 'year']
['Location', 'Group', 'month', 'Activity Type']

MAE: 16.09%

PMP model :
['Participated', 'Average Age', 'year']
['Location', 'month', 'weekday', 'Activity Type']

MAE: 7.12%

PMP model works better but both have potential improvements.

Future Work

Group activity codes into smaller groups based on similarities and domain knowledge.

If relevant, use them to improve the model performance.

Try different models to find the one with higher prediction accuracy.

Construct pipelines to simplify data cleansing and model training processes.

Thank you!

UNIVERSITY OF
MARYLAND

Questions?