Liangrui (Larry) Lu

Jiashu Wang

Prince Nwankwoala

Nicholas Claud

Team: *py.smith*

BUDT 704

**Project\_0506\_02 Read Me File**

**Background**

*py.smith* is providing business analytics (BA) consulting to the University of Maryland Alumni Association, in order to:

* Inspire lifelong connections with a global network of alumni.
* Connect, cultivate and channel the power of alumni to enrich themselves and advance the university through membership, event engagement, and volunteerism.
* Strive to make data-driven decisions to continuously improve programs and events for students, alumni, faculty, and staff.
* To evaluate the return on investment for individual events.
* To analyze the aggregate data to draw conclusions that will inform future strategies.

**Mission**

To identify variables that are highly correlated with higher event attendance of first time attendees, and major gift prospect attendees so the Alumni Association can better engage these groups.

**Objectives**

* Based on the data determine how the Alumni Association can attract more for first time attendees.
* Based on the data determine how the Alumni Association can attract more major prospects.

**Data sources**

UMD Alumni Association Dataset.xlsx, which includes:

* The dataset provided is from 07/01/2013-11/30/2019 and contains seven subsets of data, arranged by fiscal year (July 1 – June 30).
* The data set outlines each program or event hosted by the Alumni Association and provides general information on the event as well as the event registrants.

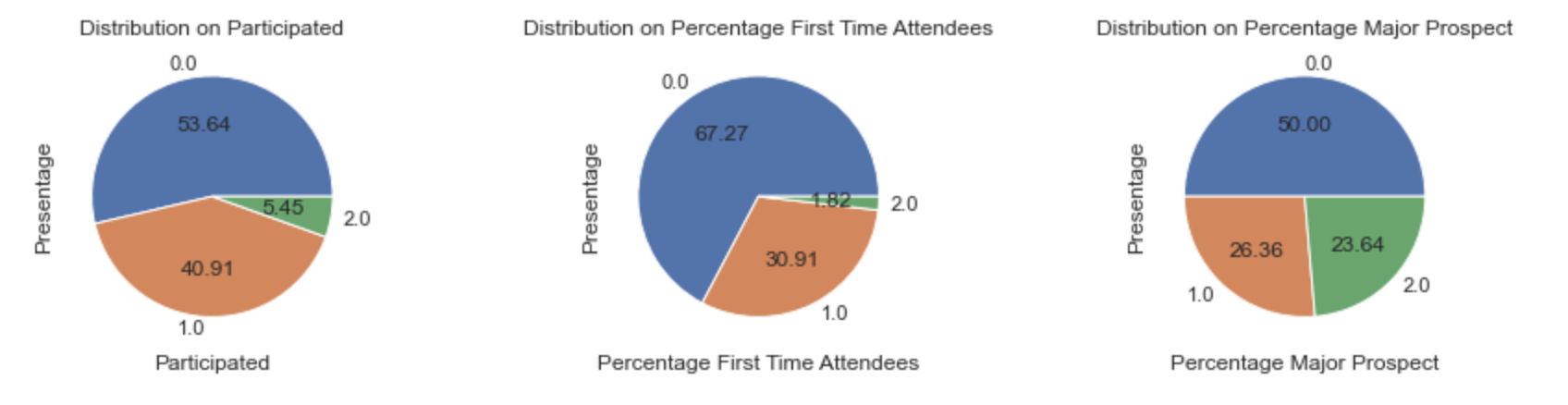
**Package Usage:**

* pandas, as pd
* numpy, as numpy
* pyplot from matplotlib, as plt, for plotting
* plotly.express as px, for treemaps plotting
* seaborn as sb, for heatmap plotting
* preprocessing, from sklearn, and then use LabelEncoder() function inside to encode the code attributes to categorical data

**Analysis Processes:**

**All our analysis is done in the jupyter notebook, using python.**

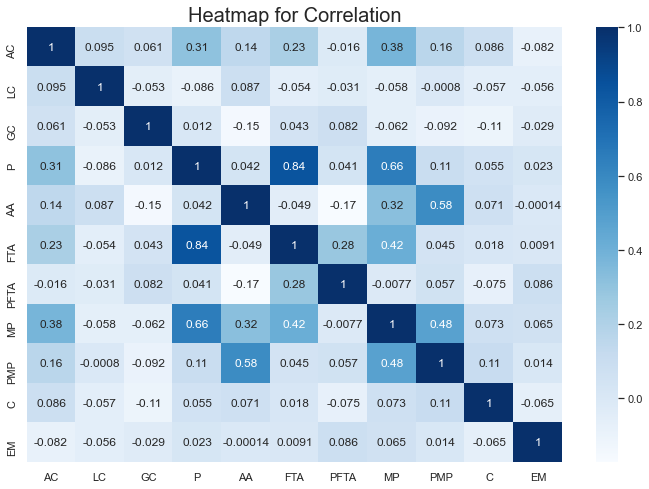
1. Data normalization & columns adjustment
   * Ensure that the same events have the same event names, using RE then manually check.
   * Add Calculated Fields, create 'combined' as identifier for events, EM for event month.
   * Column names abbreviation and description columns elimination, for events‘ better display.
   * Create count\_growing DataFrame Variable Init, for factor comparison.
   * This part is in queries 1-9.
2. Analysis for individual events, which are held multiple times.
   * Grouping for individual events, using ‘combined’ as identifier.
   * Display multi-time events one by one.
   * Plot for events that are held more than twice, by the time they are held.
   * Do factor comparison, comparing performance for existing events, when they were held for the first time, with the performance when they came back once more. Plot the results for 3 independent variables in pie charts:



Here 0 means the first-time stage performed better, 1 the opposite, 2 they are the same.

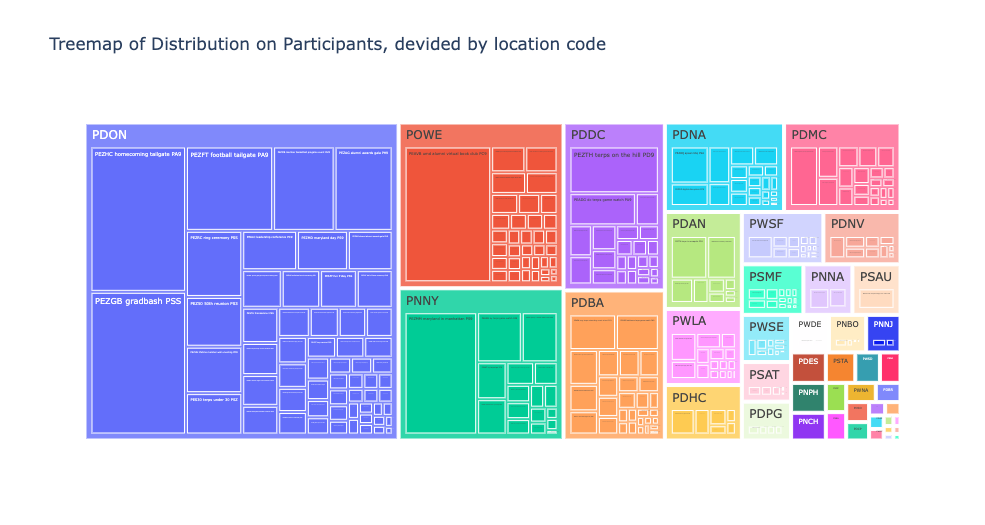
* + This part is in queries 10-14.

1. Filtering
   * Filter out events which are only held once, with participants no more than 5.
   * Filter standard:
     + Participant <= 5.
     + Just be held for once.
   * Filter Result: 622 -> 596
   * This part is in query 12.
2. Correlation among variables.
   * Use LabelEncoder to encode activity code, location code, group code to categorical variables, for correlation analysis.
   * Using heatmap to explore the relationship between variables:

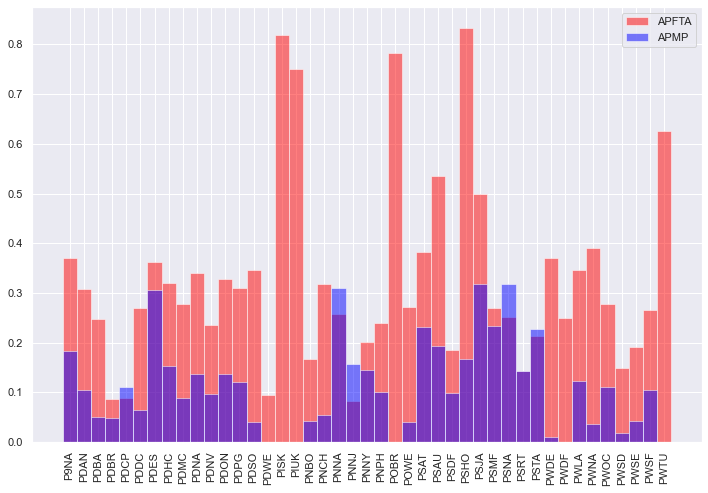


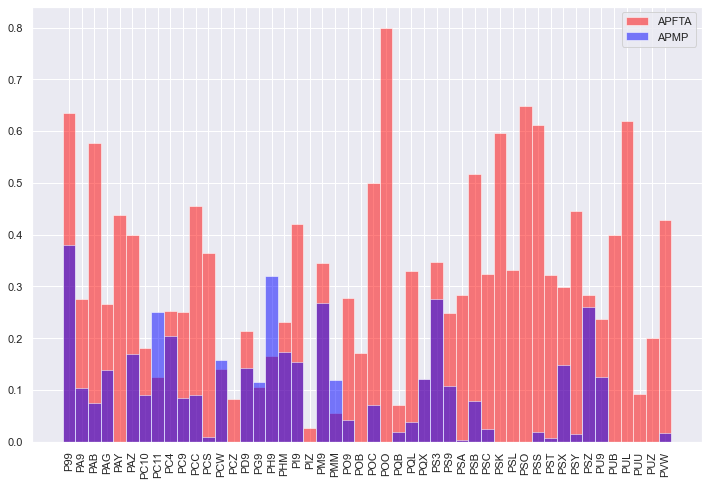
* + This part is in queries 15-16.

1. Finding good event locations and target groups
   * Using Hierarchy Treemap for Location Code and Group Code, then events identifier, by Participants.



* + Using bar charts for Location Code and Group Code with Average Percentages (APFTA and APMP) in events.

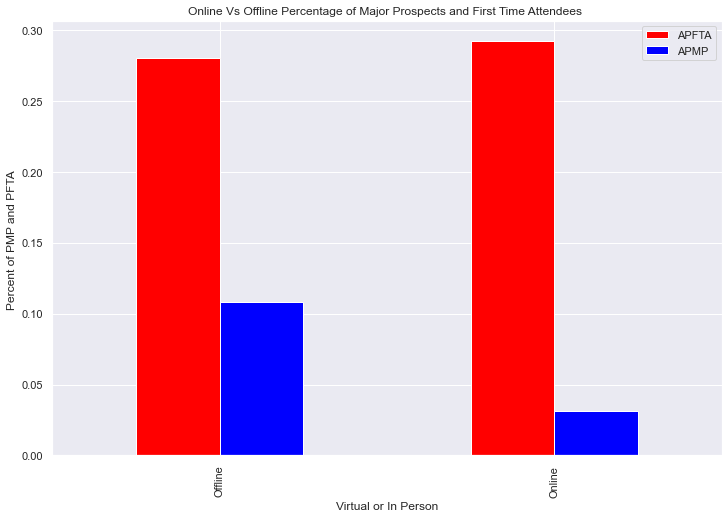




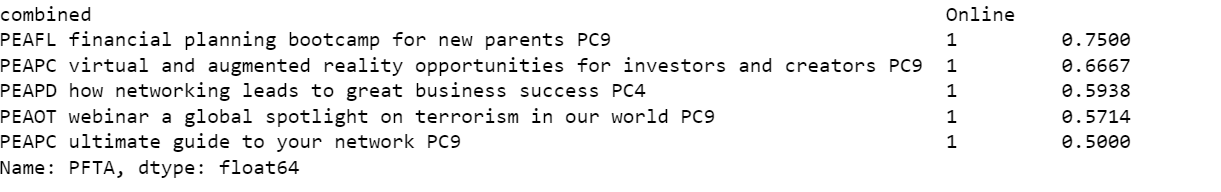
* + This part is in queries 17-20.

6. Comparing the performance of Offline and Online Activities

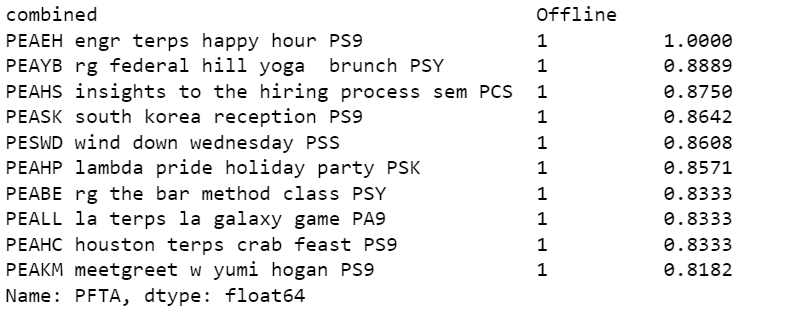
* Double bar plot showing the average percentage of major prospects and the average first time attendees for online and offline events



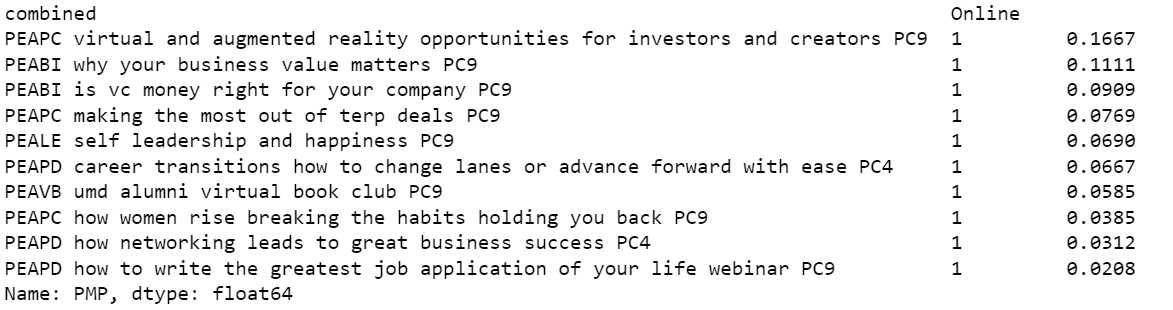
* Query Showing the top online events in terms of % of first time attendees for each month



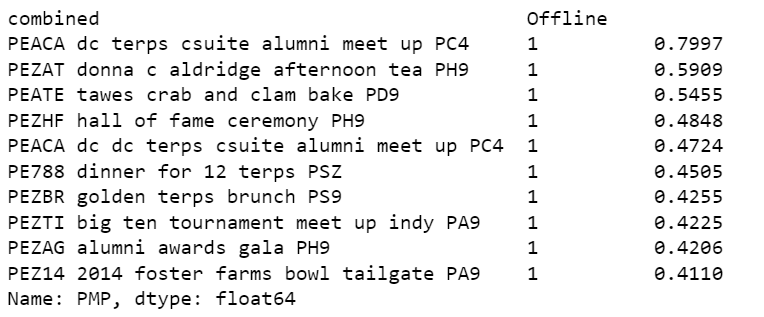
* Query Showing the top offline events in terms of % of first time attendees for each month



* Query Showing the top online events in terms of % of Major Prospects for each month

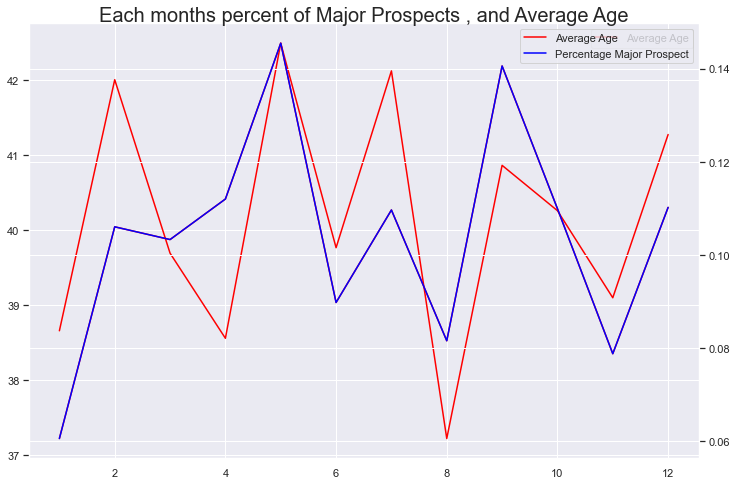


* Query Showing the top offline events in terms of % of Major Prospects for each month

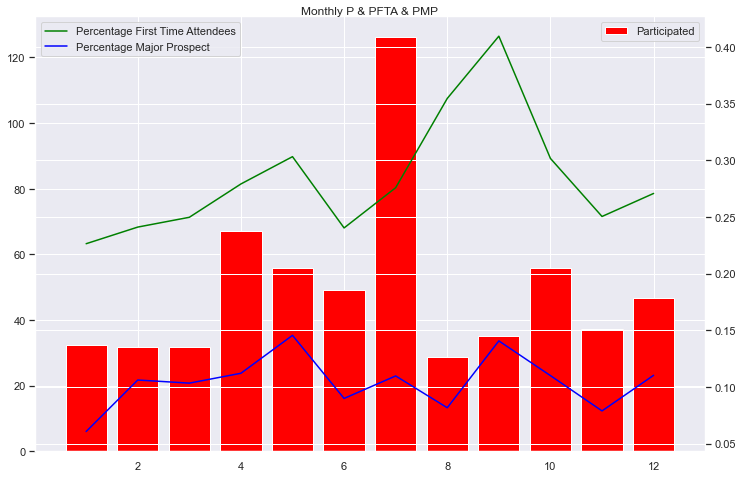


* + This part is in queries 23-29.

7.Looking at the relationship between Average Age and Percent major Prospects



* + Comparing the relationship between participants and Percent Major Prospects and Percent First Time Attendees



* + This part is in queries 30-33.

**Findings and Recommendations:**

* In analysis 2, the time series plots show highly random patterns of events across time, so we are not able to predict events’ performances based on their previous statistics.
* In analysis 2, all the three pie charts show that events’ first-stage perform better than the second, not only having a better percentage first time attendees, but also more participants and better percentage major prospects. So here I **recommend** our alumni association to **come up with new and attractive events**:
  + Holding a new event that has more advantages on enhancing the percentage of first time attendees and percentage of major prospects. While in promoting the number of participating alumni, holding new events is not very significant. So maybe we should put more effort in another way to improve the absolute value of participants.
* In analysis 5, the 2 treemaps reveal the way to increase the absolute value of participants: Alumni association should come up with **more online activities** for **specific populations** like PA9, PS9 and PC9.
* In analysis 6, the double bar plots shows that offline events significantly outperforms online events in terms of Percent Major Prospects, however online slightly outperforms offline events in Percent of First Time Attendees. **We** **recommend** that the alumni associations **comes up with online events that target alumni with a higher average age** because from analysis 4 we know those two variable are correlated
* In analysis 7 the plots show the average age and average Percent Major Prospects grouped by the event month. We see that when the average age of an event increases the average percent of major prospects also increases and when the average age decreases the percent of major prospects also decreases. **We recommend that the Alumni Associations create more events that target older alumni**.
* From the plot in analysis 7, we see that the month that garners the most participation is July, however the month that has the highest number of percentage first time attendees, and percentage of major prospects is september. **We Recommend that the alumni move some of its most popular events in July to September** to maximize the number of first time attendees and major Prospects it receives.
* In analysis 7, there seems to be a slight trend between the variables participants, percent major prospects and percent first time attendees
* For the event recommendation, we use the ratio of both average percentage of first time attendees, major prospects and percentage of first time attendees, major prospects respectively to calculate the coefficient as **evaluation criteria**. Sorting these criteria as descending order, we got top 20 recommendation events:
  + PEACA dc terps csuite alumni meet up PC4
  + PEATE tawes crab and clam bake PD9
  + PEZAT donna c aldridge afternoon tea PH9
  + PEAHH howard county happy hour P99
  + PEACA dc dc terps csuite alumni meet up PC4
  + PEAAR pg terps guided arboretum tour P99
  + PEZHF hall of fame ceremony PH9', 'PE788 dinner for 12 terps PSZ
  + PEZPT terps penn state tailgate PA9
  + PEZ14 2014 foster farms bowl tailgate PA9
  + PEZTI big ten tournament meet up indy PA9
  + PEASF sofl terps tailgate PA9
  + PEASB sofl terps boca museum of art tour PU9
  + PEAEH engr terps happy hour PS9
  + PEATW atlanta terps game watch PA9
  + PEAJK meet maryland life of anacting surgeon PC9
  + PEZBR golden terps brunch PS9
  + PEAEB engr terps bethesda event PS9
  + PEAHC houston terps crab feast PS9
  + PEALL la terps la galaxy game PA9
* Using the same criteria, we can also recommend top 10 places for holding event:
  + CP DMV- General
  + CP DMV-Eastern Shore
  + CP DMV- On Campus
  + CP DMV-Howard County
  + CP DMV-On Campus
  + CP Southeast- Fort Lauderdale
  + CP West Coast- San Francsico
  + CP West Coast- Los Angeles
  + CP Northeast- New York

**Future Work**

* Collect 2020 data to see if the performance of online events has changed.
* Improving criteria for event evaluation in ‘calculation’, e.g., involving ‘count’ into factors.