Customer Segmentation for Fast Food Chain

Motivation: A private equity firm has just invested in a family-owned, local fast food chain, which operates 9 locations in the Chicago area. In preparation for their first board meeting, they have requested a data-driven review of marketing strategy. With the infusion of private equity cash, the new board's focus is now on rapid growth. This will be their first deep-dive on current customer consumption patterns as they get to know the business.

Instructions

0. Read "Segmentation and Targeting" posted in Classes/Resources. Note: this does not directly pertain to the tasks in this assignment, but provides helpful context on customer segmentation in the context of business strategy.
1. Use unsupervised learning on the transaction dataset to create clusters representing types of customers. Note: we do not have customer attributes, we are trying to infer customer types based on their purchase patterns. You may use the Python libraries to execute the clustering and do not need to write your own algorithm, though you are welcome to if you'd like to get more "in the weeds". The following python libraries may be useful:
    - sklearn.cluster.KMeans
    - sklearn.preprocessing.OneHotEncoder
    - sklearn.model_selection.train_test_split
2. Propose "profiles" for each of your clusters, e.g. does a cluster represent a young urban professional? Couples on romantic dates?
3. Qualitatively evaluate the attractiveness of the segments, thinking about where you would like to focus marketing resources.
4. Propose a total of 2-3 segment-specific marketing initiatives you could use to support the board's strategy focus. If you need help getting started, use the 4 P's of marketing to get you started: price, product, promotion, place. Feel free to be creative!

Submission

Submit 5-7 slides as a group to Assignment 2 on the following:

1. Descriptive data overview to the board – remember this is their first time seeing the marketing data as new board members
2. Results and discussion on the clustering model including high level summary of model robustness checks, e.g. how you selected the number of clusters
3. Discussion on instructions 2-4

On your title slide be sure to include the names of group members and a link to one GitHub repository containing a complete version of the code. Only one version will be reviewed. All are welcome to have the code in their repos – and it's encouraged – but it will not be reviewed. You can work in groups of up to 5 people. Fewer (or individual work) is ok; groups are for your convenience and learning.

3 groups will be selected on Monday to give ~5 minute presentation. Though the motivation for this case is a board meeting, board level presentation formatting is not necessary. The content and discussion are the priorities. The feature engineering may take a deceptively substantial portion of your time depending on your comfort with Python/pandas, so make sure you start early.

Submissions of presentation and code for all groups are due by Monday 4/13/2010 9am Shanghai time.