Sydney Harris
ISA 616 Data Selection

Beach Tan

Introduction: I have decided to work with the Beach Tan data set this semester because I used to work at a tanning salon so I am familiar with all the variables in the dataset. Additionally, I have a basic understanding of different tanning techniques, membership levels, and tanning retail products which I believe will help with my value propositions.

Data description: The Beach Tan data set has 11,041 observations which were compiled from ten different Beach Tan store locations. These observations include twelve variables. The twelve variables, a brief description, and their data types are listed below.

UIDClient: Client Identification Number
- Integer

UIDStoreLocation: Store Identification Code
- Integer

Gender: 0 = Female, 1 = Male
- Factor with two levels
- While females and males are only two levels, in r it says gender has four levels because some clients did not answer this question which show up as " ", or NULL. Moving forward, it is important to note that I may need to clean this up in r.

DateJoined: Date the client joined. If the client is a non-member, this is the date and time of their first tanning session.
- Integer
- DateJoined is showing up as a factor with 2840 levels in r. However, this variable should be an integer which I will need to change in r before moving forward.

DaySinceJoined: The number of days elapsed between DateJoined and 11/1/2014, which is when this data was probably compiled.
- Integer

MembershipType: 0 = Non Membership, 1 = Monthly Membership, 2 = Annual Membership
- Factor with 3 levels
- MembershipType is showing up as an integer in r. However, this variable is a factor since this data can be categorized. I will need to switch this in r as well.

MembershipLevel: Types of levels are based on the level of tanning equipment. The higher the membership level, the higher the tanning equipment level available to the client. Higher tanning equipment results in faster tanning results.
 0 = Non Membership, 1 = Fast, 2 = Faster, 3 = Fastest, 4 = Instant
- Factor with 5 levels
- MembershipLevel is also showing up as an integer in r. However, this variable is a factor since this data can be categorized. I will need to switch this in r as well.

Age: Age of clients in years
- Integer

UV Tans: The number of Ultra Violet Tans a client has taken since they joined Beach Tan. UV tans include tans in UV beds or tans in stand up booths.
- Integer

SunlessTans: The number of Sunless Tans a client has taken since they joined Beach Tan. Sunless tans are spray tans.
- Integer

UpgradeRevenue: Amount spent in dollars by members to tan in a higher level bed than what their regular membership includes.
- Integer
- This variable is an integer because no decimals were used

Retail Revenue: Amount spent in dollars on various skin care lotions and products that Beach Tan offers.
- Numeric

Amount of missingness in data: After running a profile missing report using the dataexplorer package, it seems like there is no missing data in this data set, which can be seen in the image below. However, the variable gender does include some NULL values. I assume that clients either forgot to fill in this question, refused to disclose what gender they identify as, or do not identify as a female or male.

```
          feature num_missing pct_missing
1        UIDClient           0           0
2  UIDStoreLocation          0           0
3           Gender           0           0
4       DateJoined           0           0
5    DaysSinceJoined         0           0
6    MembershipType         0           0
7   MembershipLevel         0           0
8              Age           0           0
9           UVTans           0           0
10      SunlessTans          0           0
11   UpgradeRevenue          0           0
12    RetailRevenue          0           0
```

High-level overview of planned analysis: When planning to analyze Beach Tan, first and foremost I would like to compare the revenue of the different store locations against each other. I can do this by comparing each location's retail revenue and upgrade revenue. Then, I want to dive further into each location and see how they are obtaining their revenue. I can do this by comparing the total amount of sunless tans, UV tans, membership levels and membership types. To help each location with potential marketing techniques and to create a value proposition, I can analyze the most common type of customer at each location by age and gender. I can then analyze which type of tan, membership level, and membership type different age groups and genders are most likely to purchase so these store locations can customize their marketing efforts towards their different target markets.

Here are some examples that could come from this analysis. If a certain age group is purchasing similar membership levels, Beach Tan could target this age group with promotions to get them to upgrade their membership level. Beach Tan could also see which membership levels are least likely to purchase retail , and then come up with promotions/discounts to get clients of a certain membership level to purchase more retail.