Peter Fortunato

10/12/2020

Shining a Light on the Drivers of REtail And Upgrade Revenue for Beach Tan, Inc.

# Abstract

This section will contain the executive summary of the technical report.

# Background

This section will contain an introduction of the client, Beach Tan, Inc and their Marketing & Promotions team, as well as a brief overview of the products the company offers. This section will also contain a summary of Customer Segment section of the Business Value Proposition created for this client: the job currently at hand, their pains in trying to accomplish this job, as well as what they have to gain from the completion of this job.

# Methods

This section will explain the analytical methods used in the model, specifically the multiple linear regression analyses. This will also include a discussion of the creation of the dummy variables used in the models.

# Data

The data used in this analysis was provided by Beach Tan, Inc and the data was collected in 2014. It is contained in a data frame with dimensions of 11,041 observations and 12 variables. A data dictionary is provided in the table below:

Table

Description automatically generated

Variable names with an (F) next to them are factors; those with an (N) are either numeric or integer. “DateJoined” is a date factor.

The degree of missingness in this data set is miniscule; only “DateJoined” and “Gender” have any missing values, yet both of those variables are missing less than 0.5% of their rows. However, the rows with missing values for “DateJoined” have values in the variable of “DaysSinceJoined” over 40,000, which would mean that those particular customers joined the membership program or first purchased at Beach Tan, Inc. over 100 years ago. This anomaly would have negatively affected the models, so those rows were removed from the data set. The rows in which the values for “Gender” was missing were also removed so that the models didn’t create another dummy variable for the N/A’s in “Gender”.

In order to reduce the dimension of this data set, the variables “UIDClient” and “DateJoined” were removed from the data set. Each customer has their own unique identification number stored in “UIDClient”, so this variable would not provide valuable information in the model. “DateJoined” would have been treated as a factor in the models.

# Results

This section will report the model outputs and the variables that are associated with a statistically significant change in the response variables.

# Discussion/Conclusion

This section will include a discussion of the model output as well as the conclusions derived from the outputs. This will be the resolution part of the storytelling aspect of this analysis.

# Limitations

This section will describe the limitations encountered when conducting the model analysis.