AD654: Marketing Analytics –Project
Boston University

The Disney Corporation is considering some tough strategic decisions at this time, and they heard that you knew about analytics, and about theme parks. They have come to you for help.

Note: Some of the data used in this project is real, while other data is simulated for the project.

Your team will submit a written report via Blackboard. This report should come in the form of one ZIP file. This ZIP file will include:

- A written document that includes a link to your Tableau dashboard and your write-up/description of the dashboard.
- A PDF with an accompanying .ipynb that includes all of your results for the following parts:
  - Summary Stats
  - Segmentation and Targeting

During the last week of the course, your will deliver a 15-minute presentation in Zoom to your Professor.

Some things to keep in mind:

- As you move through the various tasks, remember to "call 'em like you see 'em." If you see results that aren't pretty, it does not necessarily mean that you did something wrong, or that there is a problem with the dataset or the software.
- Some tasks in this project are very similar to things that we have done in homework assignments, whereas others are unique. Take advantage of *all* your available resources...but keep in mind that the teaching team will remain at "arm's-length" distance from all project teams.
- Assumptions are okay -- if you make an assumption anywhere along the way, you can just state it so that your facilitator can see why you took some particular step.
- A rubric will be available on Blackboard.

Data Visualization (10 points): Disney's movies are one of the most valuable jewels in its crown.

Disney knows that many of its films have been major commercial successes, but it wants to understand this success with even greater detail.

Using Tableau, build a dashboard that includes anywhere from 4 to 6 visualizations created from the variables in the *disney_movies.csv* dataset.

Include a 1-2 paragraph description of your dashboard that talks about the plots you made and some of the valuable takeaways and insights that they may provide for management.

Summary Stats (5 points): In Python, conduct some exploratory data analysis of the *disney_movies.csv* dataset.

You may want to present some summary statistics of the entire dataset, but should also consider some groupings of variables, using groupby() or pivot_table() from pandas. Create anywhere from 4 to 6 total summary statistics for this section.

In a paragraph, describe your findings. What did you learn about the dataset? Mention any insights that might be particularly useful or valuable for management. Use a markdown cell to write this paragraph.

Segmentation and Targeting (25 points): After a very rough period over the couple years, Disney is hoping to bring more families back to its parks.

Here's where your team can be a huge help to the company: Disney recently reached out to a large data aggregation company to obtain information about households but the park doesn't know how to begin analyzing it. The dataset is named *family_segments.csv*. There's a lot of data here, though! There are more than 9000 observations in the file. Disney is hoping that your team can help them to separate this huge group of people into distinct clusters, and then figure out how to reach out to each of these groups from a marketing perspective.

You may wish to use either k-means or hierarchical clustering for this task. To perform the actual clustering, use only your numeric variables (but when you analyze your clusters, you can include observations about the categorical factors).

Once you have built your clustering model, use anywhere from 4-6 visualizations that help to communicate information about your model. The visualizations should depict information about your clusters that you can clearly explain, and that park management can understand. So stay away from things like PCA and t-SNE!

Name each one of your clusters, and include a few sentences describing/explaining the name that you chose for each cluster.

Finally, for each of your clusters, talk about targeting. In a couple of sentences per group, how should park management reach each of these segments?

Use markdown cells for the cluster names and the targeting section. Also, use a markdown cell to describe the process that you used for arriving at the number of clusters for your model.

***Note**: As you explore, review, and analyze this dataset, keep in mind that it is a specific sample of household data that was delivered to Disney by a data aggregator. If the numbers do not seem representative of the entire population, that's because they aren't supposed to be!*