**MGT-415 Executive Summary**

Authors: Sascha Frey, Jangwon Park, Malhotra Kuber

Outline:

Paragraph 1: brief introduction to the problem and what we aim to discover

Paragraph 2: key business insights in customer demographics

Paragraph 3: key business insights from t-tests and ANOVA

Paragraph 4: key business insights from TA’s statistical analysis (see TA’s notebook online)

Paragraph 5: conclusion – what type of customers are valuable to the company.

**Descriptive Analytics for Identifying Valuable Customers**

(Paragraph 1)

The aim of this exercise is to extract insights to understand which types of customers are most valuable to the company’s profits. The information here can be used optimize our customer retention efforts by focusing more on these valuable customers. The methodology in this exercise is mostly descriptive in nature.

(Paragraph 2)

* Remaining customers total 5,174 in numbers.
* A customer can be fully described by four attributes – gender, senior or not, partner or not, and whether he/she has dependents or not.
* Our largest segment is “non-senior, non-partnered males with no dependents” occupying nearly 19% of our entire customer base.
* Second largest segment is “non-senior, non-partner females with no dependents” at nearly 18%. From this, we can also realize that our male to female ratio is roughly equal.
* However, when it comes to partners, we have more customers that have dependents regardless of whether they are female or male. In fact, for female and male alike, “non-senior partners with dependents” occupy approximately 14% which happen to be the third and fourth largest segments.
* As a final note, while we have considerably fewer senior citizens in general at 13% of the entire customer base, there are 8 to 9 times as many customers if they do *not* have dependents, for females and males alike, as there are if they have dependents. For instance, there are 302 male senior citizens with no dependents whereas there are 36 male senior citizens with dependents.

(Paragraph 3)

* Customers with longer tenure or higher monthly charges are desirable for the company because they generate higher revenues over time.
* A number of hypothesis tests can reveal whether certain features will imply higher average tenure or monthly charges. If so, these features are informative to us in identifying more valuable customers.
* Of the four customer attributes described above, all except gender were found to be statistically significant at the 5% level. In other words, whether a customer is partner or not, for example, is statistically significant information for knowing the average monthly charges and tenure.
* On average, a senior citizen, a partner, or a customer who has no dependents have higher monthly charges as well as longer tenure, which is evident from the histograms. [Figure]
* Furthermore, the modes of the distribution give us useful information. Across all attributes, monthly charges between $20 - $30 is by far the most common. Fewest customers pay between $30 - $40, but after this, we can see a roughly normal distribution.
* On the tenure side, scenarios look different based on the attributes. If a customer is a partner, for instance, the modes ‘reverse’. We can see that most non-partners have stayed between 0 to 10 months after which the numbers steadily decrease. However, most partners have stayed for more than 67 months (close to the global maximum of the data), although there is a significant jump between 60 – 67 months and > 67 months blocks.
* Similarly, customers with dependents tend to stay long with our company while those without dependents are either very new or are long-time loyal customers. This scenario is very similar for senior vs. non-senior customers.

(Paragraph 4)

There are a number of key business insights one can gather from TA’s statistical analysis. One of the most important points would be measuring the right things. For instance, in the exploratory data analysis, we see that the value proposition in this particular case requires us to only measure those customers who are still active; and leave the inactive ones. Moreover, finding the right insights is of utmost importance. For example, if we end up with business insights which are already known or are not relevant to the case, then that can lead to a wastage of resources. Miss-classification and overfitting can come under this when we have redundant attributes which undermine the performance of regressors and classifications OR when we have irrelevant attributes which lead to overfitting. That is, when the model's accuracy is greater in the "training set" than in out-of-sample data, that is while looking at the 3D clustering of customers with respect to charges and tenure, from the actual data set must meet the expected requirements/or be close to reality, than just the result from a trial set used for inferencing and understanding the data. Consequently, we see that month-to-month contracts have the most feature importance amongst all the services that are present. And, one year-contracts are the most popular, with the order being one-year contracts > month-to-month > two-year contracts.