ISYE 6501, Week 15 HW

Question 1

Describe analytics models that could be used to help the company monetize their data: How could the company use these data sets to generate value, and what analytics models might they need to do it?

Response –

The problem statement is to identify how the data collected internally can generate value and how comb ining multiple data sets can result in better usage of data

1. **Using internal data**

Based on the previous search and purchase history (which web pages the customer looked at, how much time did the person spend on it, what the person clicked on the page and what the person bought), a linear regression model could be used to anticipate what products the customer might be interested to purchase next. Based on the linear regression model output the model could try a design of experiments approach to suggest the products to the customer. However, not all products predicted by the model will be what the customer is actually looking for. Hence, a multi-armed bandit approach could be used to determine the products which are of interest to the customer.

Based on purchase date and shipment location, the usage of a particular product can be anticipated. For example, purchases for daily essentials like paper towels and detergents for a period of time can be analyzed to approximate when does the customer need to re-stock. Sending a reminder through an email or mail can prompt the customer to purchase the product. Another way of applying costumer browsing history is to suggest alternatives. Since there are many e-commerce options available, customers tend to look around for couple of options before actually buying one. Depending on the time spent by the customer on the webpage and how frequently the product is searched, a classification model could be run to decide how interested the customer is to buy the product and a learning model could be used to suggest alternatives.

2. **Using multiple data sets**

The important constraint in being able to use multiple data sets is the ability to identify the information that belongs to a unique record from both the data sets. For example, in the data sets provided for this assignment, first name and last name in data sets #1 and #3 could be the same but they might have different middle name which is only captured in database #3 but not in #1. This could lead to incorrect predictions of customer behavior and not creating any value. A kernel based SVM approach could be used to find match between two records from different sets. If the value comes to be 1, then they belong to the same customer. If the value comes out to be 0, then they belong to different customers.

Once the customers who have data across different data sets are identified, data across different data sets could be used to improve the business with the customer. Data set#1 provides the interests of a

customer, data set #2 provides the credit worthiness or the spending capability of the customer. This information can be used to suggest the products that need to be displayed to the customer when designing the design of experiment or the multi-armed bandit approach as discussed above.

The financial capabilities of the customer from dataset #2 could also be used to suggest more activities which are similar to the ones mentioned in dataset #1 using clustering models.