Question 20.1

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?

There are lots of good answers, and I want you to think about two types – at least one of your answers should be based on just one data set, the one they've collected internally on customer browsing patterns on the web site; and at least one of your other answers should be based on combining more than one of the data sets.

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

Here are the three data sets to consider:

DATA SET #1 (purchased from an alumni magazine publisher)

- first name
- last name
- college or university attended
- year of graduation
- major or majors
- marital status
- number of children
- current city
- email domain
- financial net worth
- binary variables (one for each interest in the publisher's long list of various sports, activities, hobbies, games, etc.) showing whether each one was or wasn't listed by each person

DATA SET #2 (purchased from a credit bureau)

- first name
- middle name
- last name
- marital status
- sex
- year of birth
- current city
- whether they ever owned real estate
- email domain

- list of monthly payment status over the last five years for credit cards, mortgages, rent, utility bills, etc. – for each month and each payment:
    - what type of payment it was – for credit cards, it would say “Visa”, “American express”, etc., not just “credit card”
    - how much was owed
    - how much was paid
    - whether the person was considered to be in default

DATA SET #3 (collected by the company using web site tracking code)

- title
- first name
- middle initial
- last name
- credit card type
- credit card number
- list of products purchased in the past, with date of purchase and ship-to address
- which web pages the person looked at
- how long the person spent on each page
- what the person clicked on each page
- estimate of how long the user’s eyes spent on each page viewed (for customers where the software was able to take over the device’s camera)

Single data set (#3)

The objective is to learn the users' specific browsing behavior to curate tailored marketing.

Given their basic personal information (names, titles, credit card information, purchase history) and engagement on the websites, K-means clustering can be **used** to segment users based on their browsing behavior. Data like pages viewed, time spent on the website, and purchase history can be fed into the K-means to create "buckets" (or clusters) of distinct personas of users. **As a result** of clustering, the business will have a high-level understanding of their prospective clients and marketing strategy.

Now **given** the clustering, a nearest neighbor algorithm can be **used** to personalize recommendations within each bucket of personas. By analyzing the preferences of users within the same cluster. **As a result** the algorithm suggests products or services tailored to similar interests and behaviors. In short this will look like how Spotify or Netflix recommend their music and movies.

Multiple data sets

The objective is to capitalize off of the multiple data sets, we can triangulate the multiple perspectives from the three sources of data to gain insights, leverage in robust (and potentially annoying) marketing strategy, and monetize off of the human experience.

Given the multiple sources of data we will need to sift through all the noise and extract the most useful information. We may also anticipate using imputation techniques. Data analysis is very important with bigger volumes of data. **To** execute the data analysis we will need to find similarities amongst the three tables as they will become the primary and foreign keys to join the three tables to make the data meaningful. For example, a person's full (first, middle, and last) name can be a unique identifier. Their demographic information and webpage engagement information from data set #3 can be linked to their purchase history with the bureau and/or alumni magazine. This three-way link can give us a comprehensive profile of someone. The next step is to take that data insight and monetize (exploit) off that individual. **As a result** of learning people's behaviors, we can suggest products based on their own biases, preferences, and potentially emotional sentiments. I imagine people only purchase things from an alumni magazine because they or someone they know are related to that school somehow. The marketing campaign can come in the forms of phone calls, emails, snail mails, SMS, or apps.