

## 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)

### Just One Data Set:

After looking at the three data sets, I believed that it made the most sense to create a single data set model using Data Set 3 (web site tracking). This was because in my head it gave the most complete data out of all of them. This data set contains all personal information, address, what they have purchased, and how long they look at products. This one data set can give the company a full picture of which customer has which behaviors. Below is how I would use the data set:

Given {Data Set 3}, use {k-means clustering and predictive modeling} to {specify marketing and estimate who potential buyers would be}.

K-means clustering can put customers into groups based on browsing similarities, and regression analysis can predict future purchases based on how the users have previously interacted and other factors the data set gives.

### Two Data Sets:

Looking at the data sets, a model for something that came to mind needed the information in Data Set 2 and Data Set 3 to be made. This model is as follows:

Given {Data Set 2 and Data Set 3}, use {logistic regression} to {see which customers are most likely to make large purchases}. Using these two data sets with a logistic regression model, the company can put customers into categories of high-value, mid-value, and low-value customers based on their probabilities of making large purchases. The company could then focus on targeting their preferred (most likely high-value) customers in specific marketing campaigns, specialized coupons, etc.