

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)

I. Problem#1:

Based on data collected internally on customer browsing patterns on the web site

Company: ABC Corp, a diverse online retailer store

- Inventory management by understanding the Demographics of Interest:

- 1) Collect Top Customers for rolling 3 months;
- 2) For top customers identify the product categories of the product type.

Pre-requisite: Clusters of each product by product category is available

- Use the heavily purchased category to expand the inventory of this product category

How to achieve

- 1) Clustering K-means to create clusters of product category
- 2) Use CUSUM Change detection , identify the beneficial customers by setting a threshold percent with the existing time series data
- 3) Use Regression to identify the relationship of product categories for the Top Customer to expand the inventory of top purchased product category

More use cases...

II. Products of least and most interest

By using the time spent in each page assuming each page corresponds to the products.

The customer who spent less time in a page is least interested and more time is clearly more interested.

Using this information of pages with most spent time, the customer can be targeted to send complementary or similar products as home page or interested/ suggestion to motivate purchase

III. Develop data store of Complementary products:

By using the page navigation from current page to next page along with combination of the page spend more time on will indicate complementary products. This information can be collected and the complementary or similar products can be displayed together for marketing to other customers.

Problem#2: and at least one of your other answers should be based on combining more than one of the data sets.

I. Potential candidates for real estate marketing

Source:

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

DATA SET #2 (purchased from a credit bureau)

Solution:

Use random trees to identify combination path for potential real estate buyers.

One of the paths is listed above

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

- Marital status, number of children
- Financial net worth

DATA SET #2 (purchased from a credit bureau)

Customer with “NO” Default

Real estate owned: No

If amount owned is not above a threshold

Use combination of above data with the filters in place to identify potential real estate buyers.