

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:
 - o what type of payment it was – for credit cards, it would say “Visa”, “American express”, etc., not just “credit card”
 - o how much was owed

- o how much was paid
- o 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)

Problem 1:

For this problem, we are asked to use data set number 3 to help the company monetize their data.

Given: { data in dataset 3:

List of Products Purchased (with date of purchase and ship-to address)

Time Spent on Each Page

Clicked Elements on Each Page

Estimate of Eye Tracking Data (if available) }

Use: { K-means Clustering }

To : { Identify distinct customer segments based on purchasing behavior, web interactions, and eye-tracking data. }

Explanation: By applying K-means clustering to the specific columns from Data Set #3, including the list of products purchased, time spent on each page, clicked elements on each page, and the estimate of eye-tracking data, the company can derive valuable insights to enhance its operations. The first application of K-means clustering is to segment customers. This involves using the clustering model to identify distinct customer segments based on their purchasing behavior, web interactions, and eye-tracking data. This segmentation allows the company to understand different customer profiles and tailor its marketing strategies accordingly. The second application is to tailor marketing strategies by utilizing K-means cluster labels. By customizing promotions and content based on the shared characteristics of each customer segment, the company can optimize its marketing efforts to target specific groups more effectively. Lastly, leveraging K-means clusters enables the company to enhance user engagement by personalizing website content. By considering the preferences observed in each cluster, the company can create a more tailored and engaging online experience for users, ultimately improving customer satisfaction and increasing the likelihood of conversions. Overall, these applications

of K-means clustering empower the company to generate value by gaining a deeper understanding of its customer base, optimizing marketing strategies, and enhancing the overall user experience on its website.

Problem 2:

For this problem, we are asked to use all datasets to help the company monetize their data. Here I implement two models in an attempt to help the company generate value.

Given: { Demographic Information (Data Set #1):

- Year of graduation
- Number of children
- Marital status

Financial Information (Data Set #2):

- Whether they ever owned real estate
- List of monthly payment status over the last five years for credit cards, mortgages, rent, utility bills, etc.

Behavioral and Purchasing Patterns (Data Set #3):

- List of products purchased in the past
- Time spent on each page (potentially indicating engagement level)

AND

Demographic Information (Data Set #1):

- Marital status
- Number of children

Financial Information (Data Set #2):

- Whether they ever owned real estate
- List of monthly payment status over the last five years for credit cards, mortgages, rent, utility bills, etc.
 - Type of payment
 - Amount owed
 - Amount paid
 - Default status

Behavioral and Purchasing Patterns (Data Set #3):

- List of products purchased in the past (may indicate spending habits)
- Web pages viewed, time spent, and clicked elements (indicating online behavior) }

Use: { Linear AND Logistic Regression }

To : { predict the financial net worth of customers, as well as to predict the likelihood of a customer defaulting on payments }

For Linear Reg:

Example Variables:

- Number of Children:
 - Explanation: This variable may positively correlate with financial net worth, as individuals with more children might have greater financial responsibilities. Higher financial responsibilities could lead to increased income and assets, potentially contributing to a higher net worth.
- Real Estate Ownership:
 - Explanation: Individuals who own real estate may have accumulated significant assets, positively impacting their financial net worth. This variable can be indicative of stability and financial success, contributing to a higher predicted net worth.

For Log Reg:

Example Variables:

- Payment History for Credit Cards:
 - Explanation: Analyzing the payment history for credit cards, including factors such as the type of payment, amount owed, and whether the person was considered to be in default, can be crucial for predicting the likelihood of a customer defaulting on payments. For instance, consistent default status or a history of late payments may indicate an increased risk of future defaults.
- Real Estate Ownership:
 - Explanation: While real estate ownership is also used in linear regression, in logistic regression, it may help predict the likelihood of default differently. For example, individuals with real estate may be more likely to prioritize mortgage payments over other debts, reducing the risk of default. Conversely, if real estate ownership is associated with financial strain, it might increase the risk of default.

Explanation: Linear regression and logistic regression models offer the company valuable tools to monetize its data by leveraging key variables extracted from combined datasets. With the objective of predicting financial net worth, linear regression utilizes demographic information, financial behaviors, and online engagement patterns, such as year of graduation, number of children, marital status, real estate ownership, and payment histories. By tailoring campaigns to specific customer segments, the company can increase the effectiveness of its promotions and enhance the likelihood of conversion. Additionally, the prediction of financial net worth contributes to robust risk management practices, helping identify

resilient customers and mitigate potential credit risks. This predictive insight enables the company to tailor financial products and services, craft customized marketing campaigns, and segment customers effectively. On the other hand, logistic regression focuses on predicting default status, utilizing variables like marital status, number of children, real estate ownership, and payment histories. The model aids in risk assessment, credit risk mitigation, and optimized credit approvals, fostering responsible lending practices. Collectively, these regression models enhance decision-making, improve customer relationships through personalized approaches, optimize resource allocation, and identify monetization opportunities, contributing to a more secure and revenue-driven financial strategy.