

HOMework 15

INSTRUCTIONS

- Every learner should submit his/her own homework solutions. However, you are allowed to discuss the homework with each other (in fact, I encourage you to form groups and/or use the forums) – but everyone must submit his/her own solution; you may not copy someone else's solution.
- The homework will be peer-graded. In analytics modeling, there are often lots of different approaches that work well, and I want you to see not just your own, but also others.
- The homework grading scale reflects the fact that the primary purpose of homework is learning:

Rating	Meaning	Point value (out of 100)
4	All correct (perhaps except a few details) <u>with</u> a deeper solution than expected	100
3	Most or all correct	90
2	Not correct, but a reasonable attempt	75
1	Not correct, insufficient effort	50
0	Not submitted	0

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)