Customer relationship management

Contents

1. The customer lifecycle

- Most companies have customers. The aim is to develop long-term relationships, loyalty and long-term customer value. We need to:
  - Gather and store data about customer interactions over the long-term
    - Operational and warehousing systems
  - Learn from the data via data mining
  - Act on the results, measure the benefits, and finally learn from the endeavour

The customer’s real lifecycle

- … based upon key events: graduating, first employment, significant promotion, getting married, having kids, buying house, moving house, retiring, …
- In many cases, difficult to handle in the context of CRM because the events are rare and difficult to predict

The customer’s business lifecycle

- More usually we’ll look at the lifecycle as defined by the business relationship
  - It is this relationship that is of value to the business
  - This relationship generates data that can be used for prediction and leverage
- The phases of this lifecycle suggest possible mining applications

Phases of the customer lifecycle

- *Prospects* are individuals or businesses in the target market, but who are not yet customers
- *Responders* are prospects who have expressed some interest
- Prospects become *new customers* when they first make some commitment
- *Established customers* are those who return for repeat business & whose value we wish to maximise
- *Former customers:* either voluntary, expected, forced

Customer acquisition

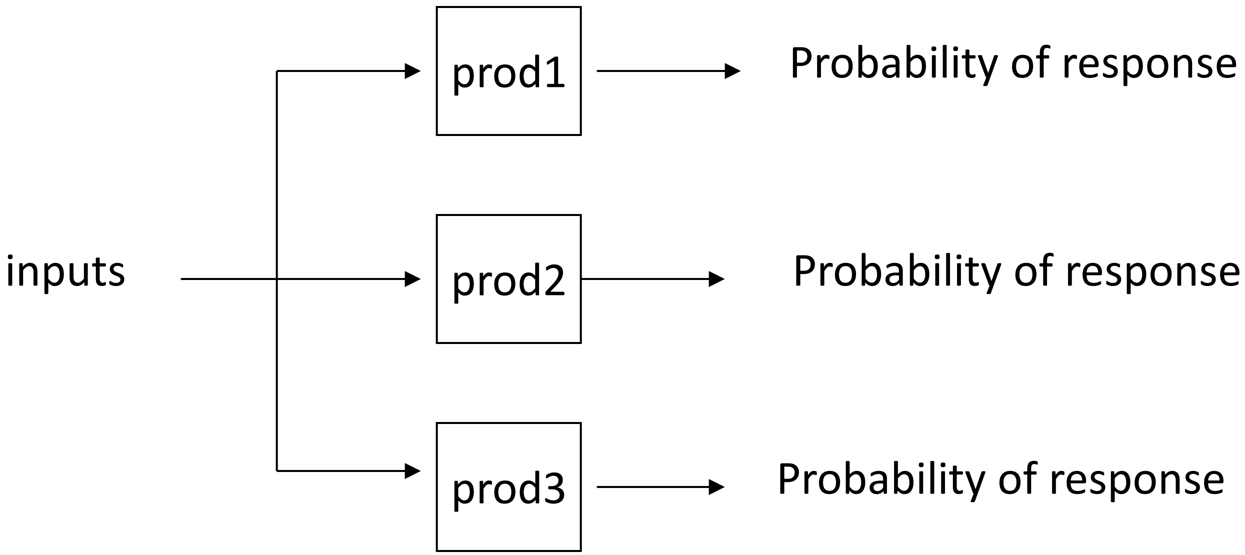
- Our aim is to turn prospects into responders and ultimately customers
  - No doubt some responders will fail to become customers – if significant, investigate (possibly using data mining)
- We can apply data mining to help determine the target market (i.e., who are the prospects) - using characteristics of past and existing customers to gain insight, but:
  - The past is not always a good predictor of the future
  - The market may vary significantly with geography, or be affected by new products or new competitors

Customer acquisition

- Marketing channels might include mass marketing/email/direct mail/… depending upon whether we have contact information
- May employ purchased data lists and/or targeted marketing/data mining to identify the most likely responders, most sensible channel and best marketing message
  - particularly when the cost of contact is high
  - provided we have the input data
  - may use reference data if available

Choosing the marketing message

Using multiple models to make a single decision



Suppose we have multiple products that might be included in our marketing message – we develop separate propensity models for the differing products and then use the resultant probabilities to decide upon the best

Acquisition information

- Information available at the time of acquisition has predictive capability and should be gathered and kept
- Conversely, understanding which of these characteristics is associated to long-term profitability (say) can influence the acquisition strategy

Established customers

… are those who return for repeat business

- Our aim is to develop the relationship and ultimately maximise their value
- Customer value prediction/estimation
- Up-selling/cross-selling/usage stimulation
- Retention … is important too …
  - it is expensive to acquire new customers
  - existing customers are of higher value than new

Relationship depth

- A customer transaction yields simple transaction data
  - product, price, time, location
  - but in the worst case it may leave behind little "customer information"
- In the absence of customer information, marketing is biased towards mass-marketing
- Developing the relationship suggests encouraging the gathering of information

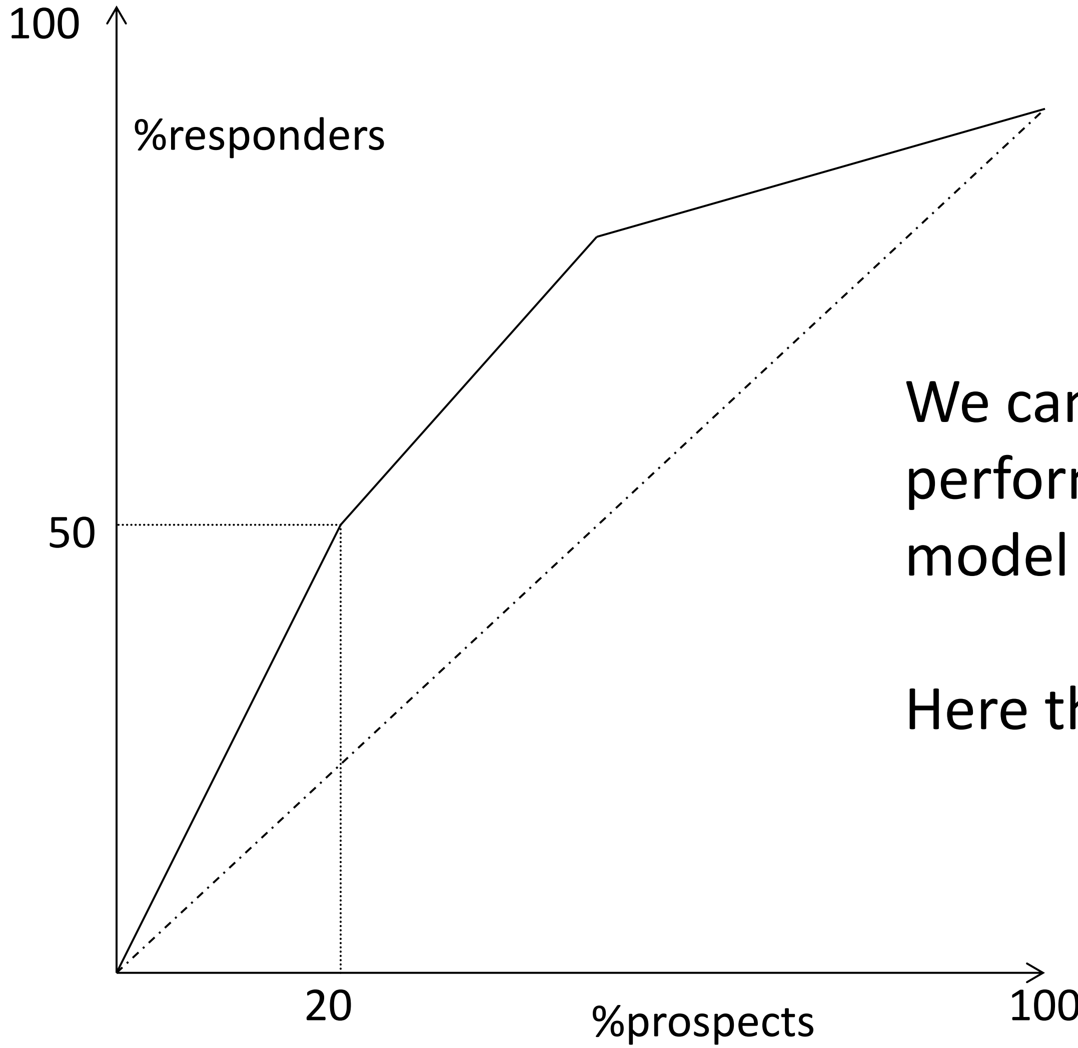
Relationship depth

- Other cases may yield more information:
  - loyalty card
  - credit card
  - online account
  - catalog account
  - subscription id
- Greater relationship depth facilitates wider applications: customer profiling, product associations, targeted marketing, …

2. Measuring the impact: The costs of targeted marketing

- Suppose we have a limited budget – and mailing has a real cost
- The value of a marketing campaign would seem to depend upon its success in reaching responders
- Suppose that our data mining model ranks prospects according to likelihood of response

The lift chart



We can evaluate the performance of a predictive model using a *lift chart*

Here the initial lift = 50/20

Costings

- Suppose that:
  - there are 100 000 prospects & we believe that 1% of these would respond (if we contacted them)
  - the cost of a mailing is 50p
  - we mail the top 20% of prospects (as a consequence of the model)
  - the revenue generated by a sale is £50
- Is the campaign predicted to be worthwhile?

Predicted net profit if we contacted them all

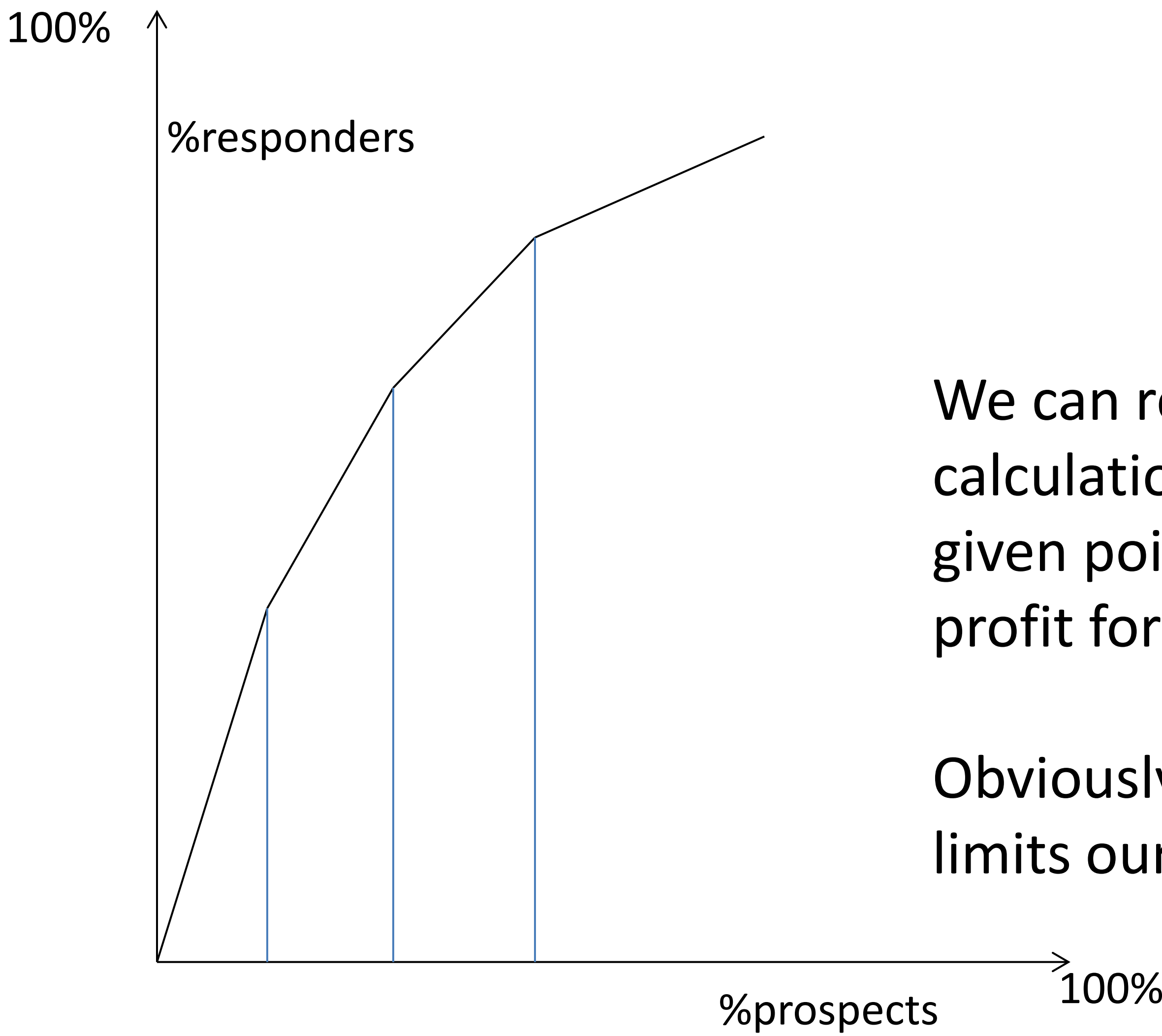
- We believe there to be 1% responders in the population as a whole
  - 1% of 100 000 is 1000
  - Yielding a sales profit of 1000 * £50 = £50 000
- Minus the cost of mailing them all

  = 100 000 * £0.5 = £50 000

Predicted net profit of campaign

- The “top 20%” of prospects (consists of 20 000 prospects) contains half the responders
  - Half of 1000 is 500
    - 500 = 2.5% of all prospects contacted (vs. 1% responders in the entire population)
    - In the target sample, responder density has increased by a factor of 2.5 = 5/2
  - Yielding a sales profit of 500 * £50 = £25 000
- Minus the cost of the mailing = 20 000 * £0.5 = £10 000

How many prospects should we contact?



We can repeat the calculation for each of the given points to calculate net profit for each

Obviously campaign budget limits our choices

Further considerations

- Contacting non-responders may not be a complete waste of time
  - It highlights the company and its products or services, which might bring in future trade
  - The loss may not be as high as thought
- Contacting responders may be a waste of time
  - They may have purchased the product anyway!
  - The profit may not be as high as thought
- Having undertaken the campaign we should analyse the actual results

False positives & negatives

- In a marketing campaign a false positive is someone predicted to respond – who does not
  - … and similarly a false negative is someone who is predicted not to respond who would have
- The cost of a false negative (the lost sale) is higher than that of a false positive (the wasted mailing)
  - This imbalance of costs needs to be taken into account in assessing a predictive model

False positives & negatives

- Similarly in a medical diagnosis, false negatives are potentially serious
  – In which case we miss the illness
- False positives are less serious, because a “positive” indicator would always be followed up/checked with further medical tests (although such tests do still have a cost)
  – Again this imbalance needs to be taken into account in assessing a predictive model

False positives & negatives
Confusion matrix

		Responder	
		Y	N
Contacted	Y	0.5%	19.5%
	N	0.5%	79.5%

False positives and false negatives may have differing costs. The percentages in the confusion matrix can be combined with these costs to evaluate a model.

3. Preliminary analysis

- Identify suitable business problem(s)
  - where data analysis might provide business value
- Transform data into actionable results
  - using data mining
- Act on these results
- Measure the impact of the actions

Obtaining actionable results

- Obtain, validate & clean data
- Preliminary analysis
- Choose modelling technique
- Pre-process data/ prepare the model set
- Build model & evaluate performance
- Pick “best” model and apply to score set

Case study

- Publisher sells magazines
  - 5 categories: cars, houses, sports, music, comics
- Aim: customer profiling & cross selling
- Typical questions might include
  - what's the profile of a reader of car magazines?
  - is there a correlation between readers of car magazines and comics?

Preliminary analysis

- To gain an initial insight into the basic structure
- Allows retrieval of lots of shallow information
  - SQL
  - graphical tools
  - visualisation tools
  - statistics

All customers	
Attribute	**Average**
Age	46.9
Income	25.8
Credit	34.9
Car Owner	0.55
House Owner	0.69

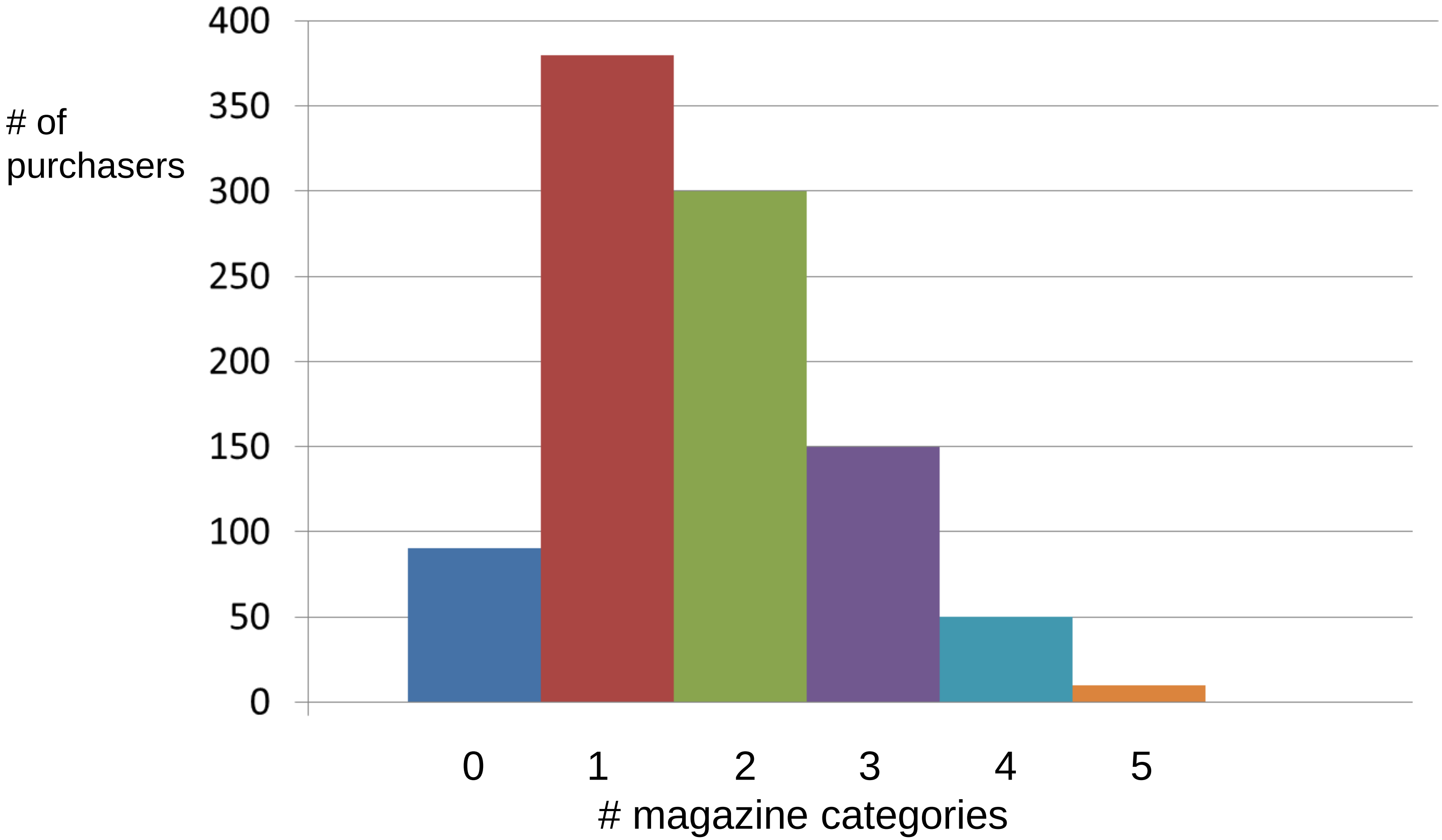
Current take-up

Magazine	% of clients	Naïve predictive capability
Car	32.9%	67.1%
House	70.2%	70.2%
Sports	44.7%	55.3%
Music	14.6%	85.4%
Comic	8.1%	91.9%

Demographic breakdown

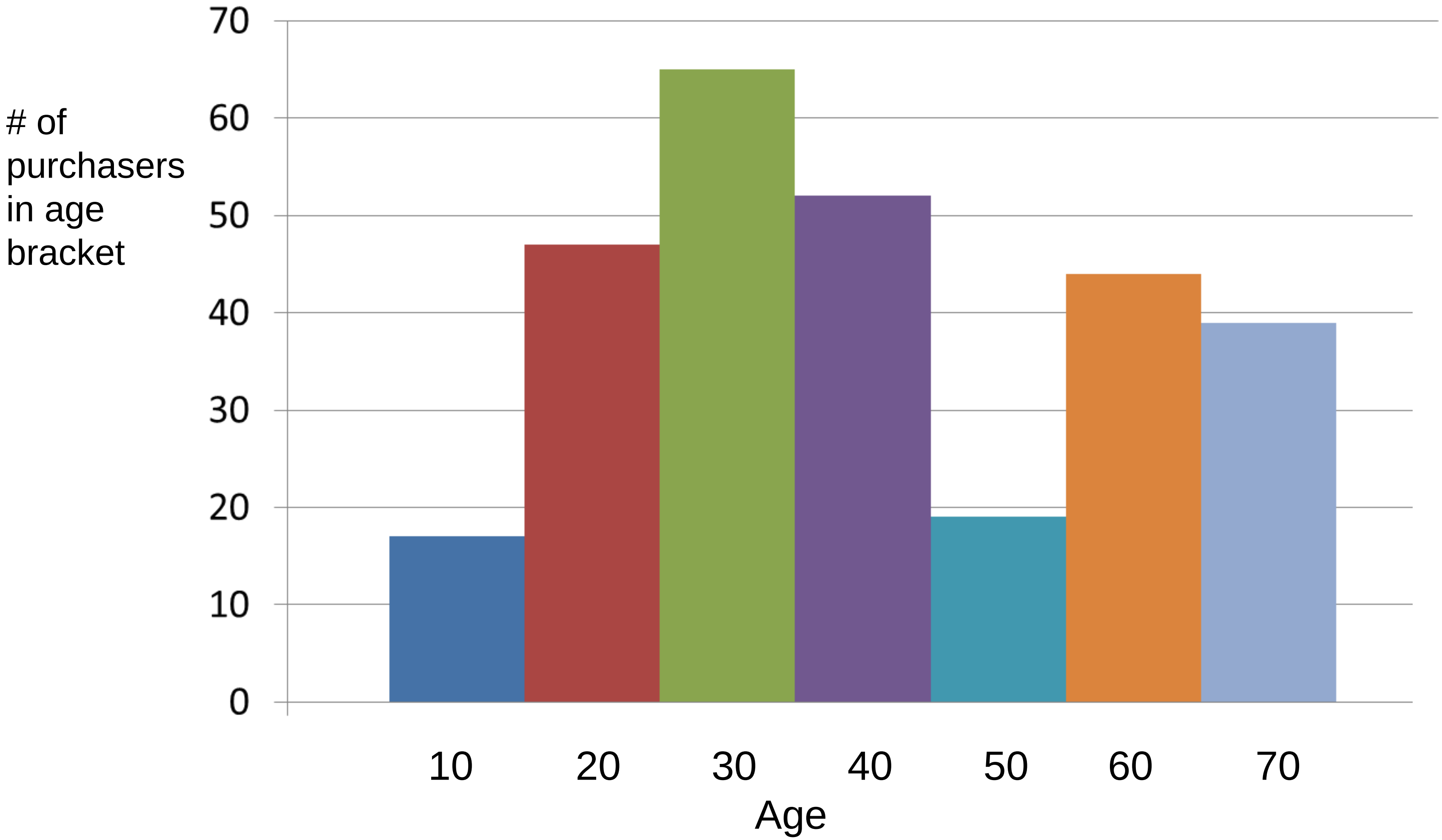
Magazine	Age	Income	Credit	Car owner	House owner
Car	39.3	27.1	27.3	0.78	0.52
House	48.1	51.1	45.1	0.58	0.84
Sports	39.2	24.3	31.2	0.7	0.61
Music	24.6	15.8	19.6	0.3	0.33
Comics	18.4	11.5	10.7	0.12	0.08

purchasers

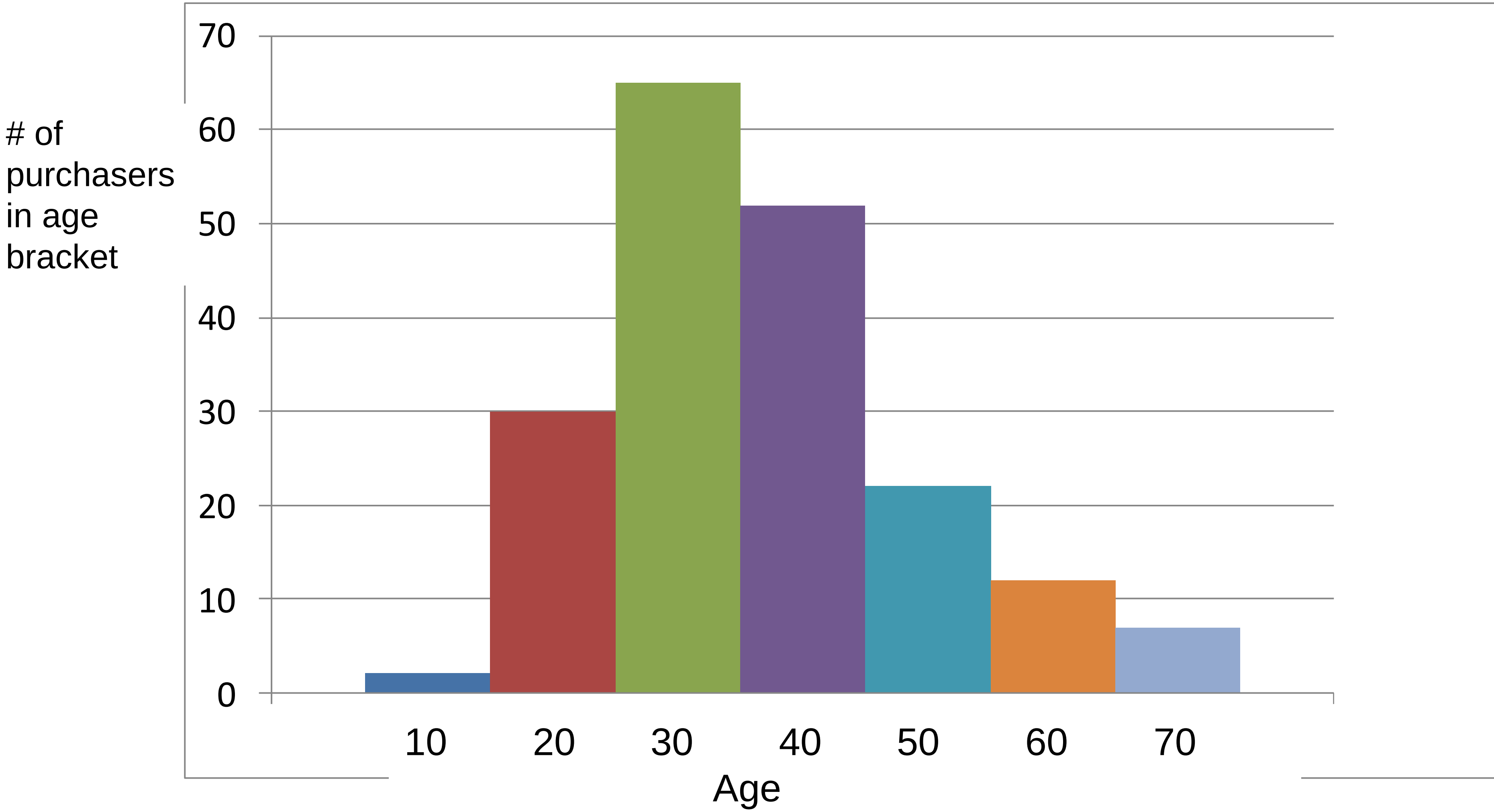


# of purchasers
400
350
300
250
200
150
100
50
0
0
1
2
3
4
5
magazine categories

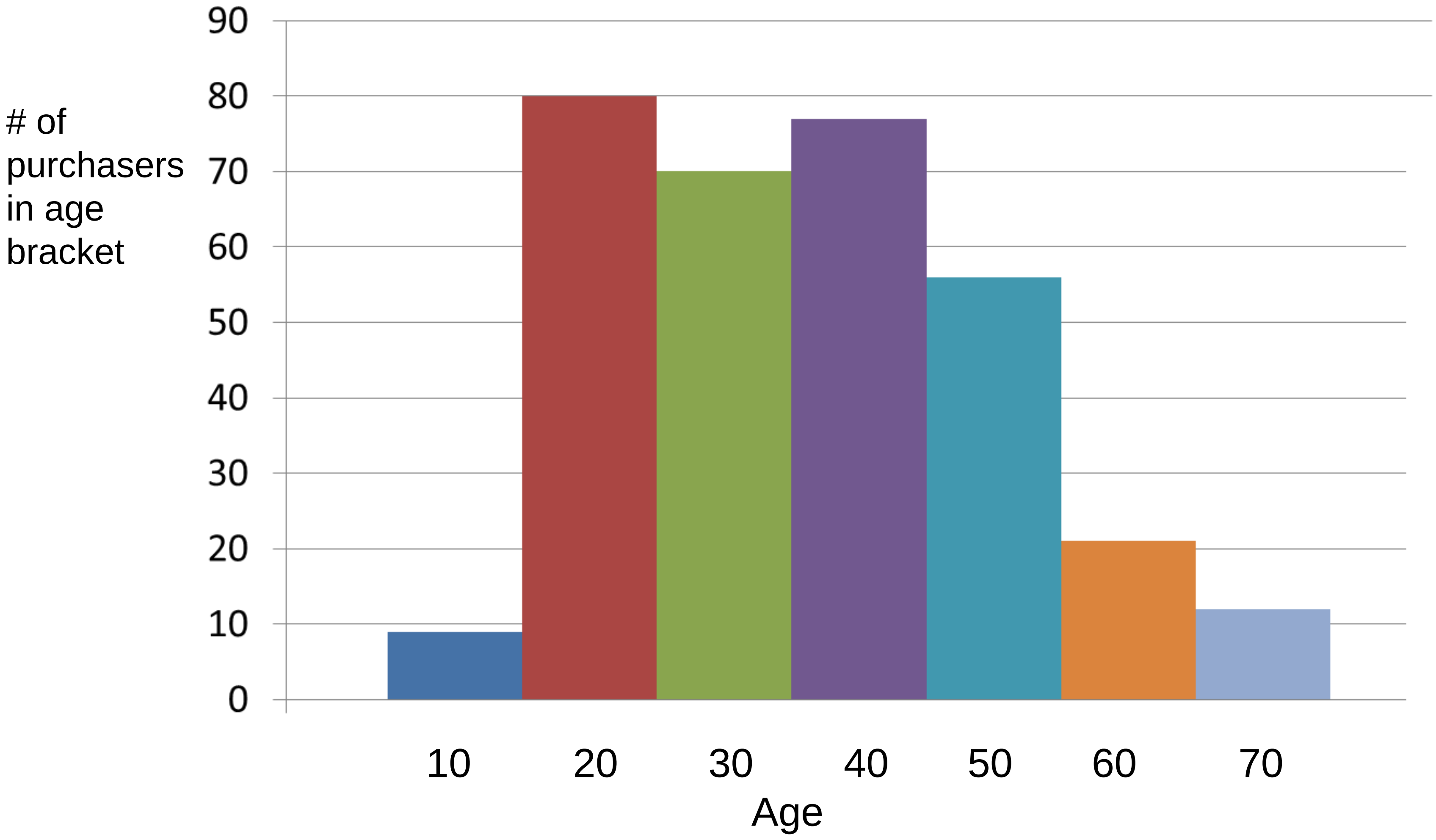
purchasers by Age



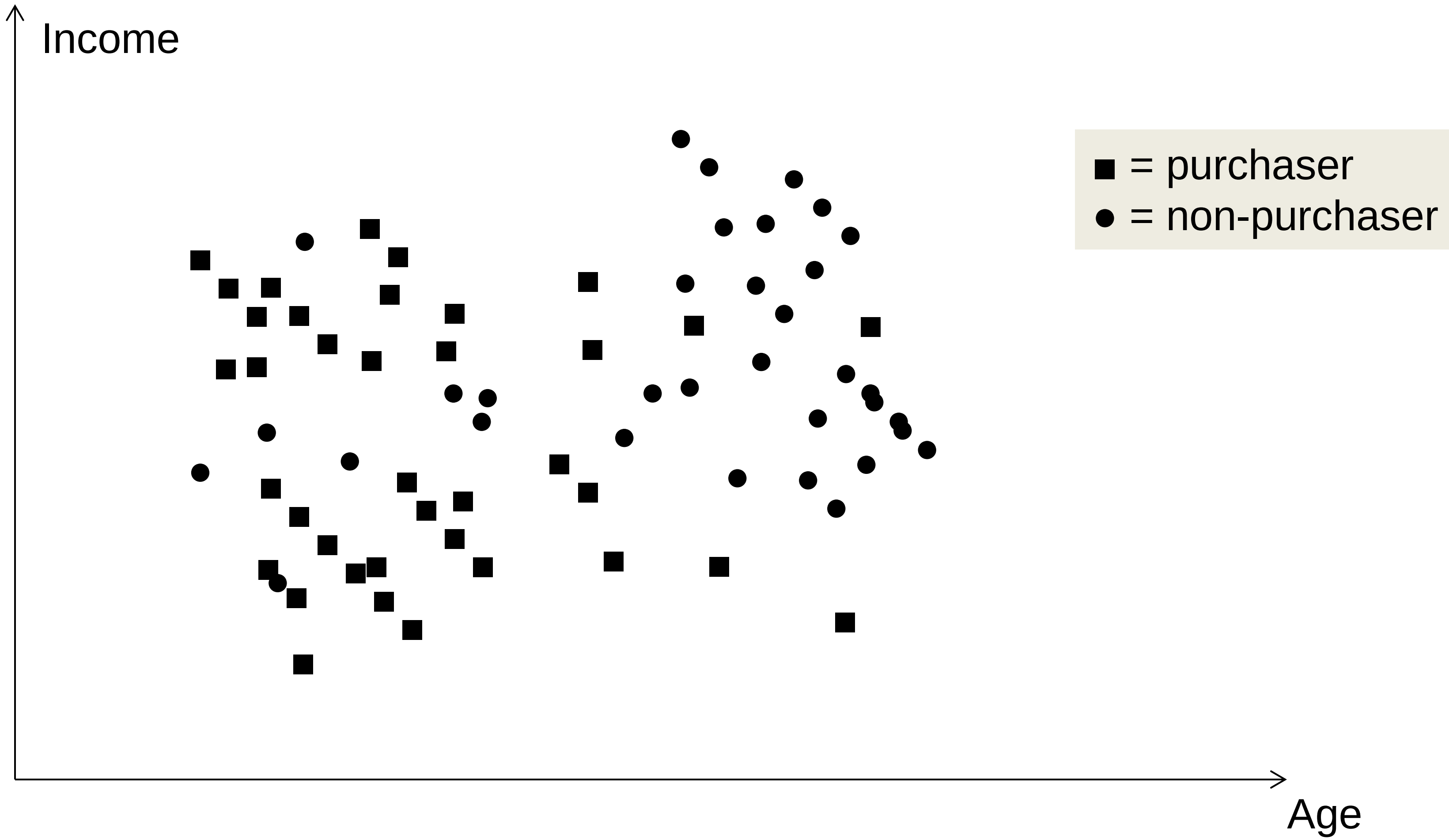
Car purchasers by Age



Sports purchasers by Age

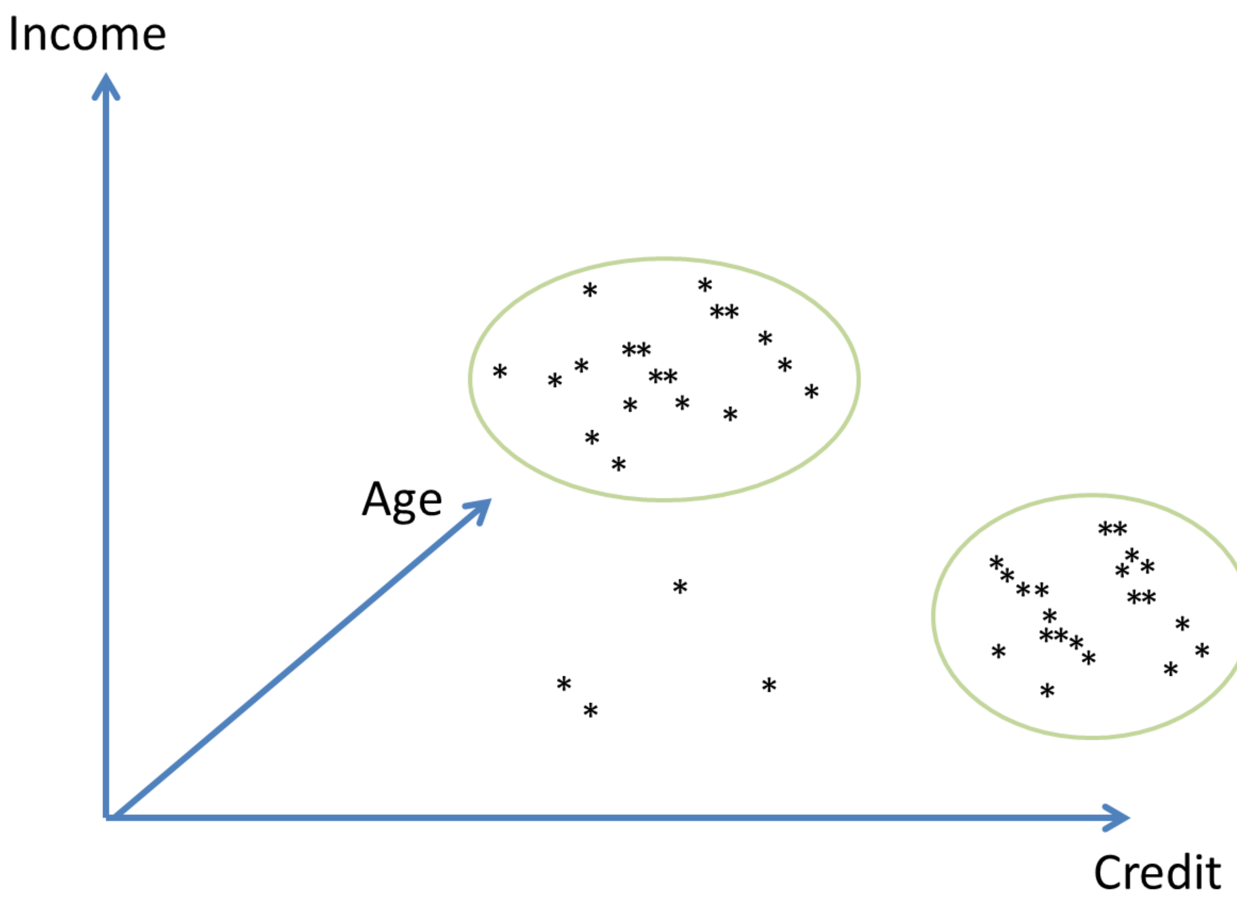


A scatter plot might reveal variables correlated with a target



Pursuit projection

Plotting the data may reveal some (demographic) clustering



4. Customer data and its information content

- Data for mining should be in a single table
- Each row should correspond to a single instance - a "unit of action"
  - e.g., customer, insurance policy, insurance claim, basket, household, PC, inventory item
  - So: need to understand how the results of the mining effort will be employed
- Suppose that the customer = unit of action

Customer data

- Financial/lifestyle/demographic data
- Present & historical transaction/ subscription/… data
- Acquisition data
- Historical marketing and response data
- Other interaction data
- Predicted and derived data

Customer id and address

- A customer id is unique - hence explicitly contains no patterns or information content
  - can delete *for the purposes of mining*
- A customer address is unique (no patterns) & far too detailed
  - extract derived information (e.g., region code)
    - a form of binning
  - extract any useful reference information
    - E.g., average sales in the given region
  - … and then delete the original address

Information content

- A customer will have many transactions, each containing a number of items purchased - this has to be aggregated:
  - one customer = one row
  - total number of transactions
  - total value of transactions
  - # of high value transactions
  - etc, etc, ...

Information content

- Date of first purchase: too detailed - aggregate to month/year
  – facilitates time series analysis
  – again this is binning
- Other derived variables from dates are
  – weekday (or not), day of the week/month, week of the month/quarter/year, holiday month (or not), time between events, …
  – Similarly for times, phone numbers, web addresses, bar codes, car registration#, ...

Information content

- Profitability of customer
  – may be used to weight the importance of customers in a mining effort
- Predicted long-term profitability
  – as above – but in this case the value may come from some other mining effort

Taking the output from one mining effort and using it as an input in another is called *enhancing* the data … it is clearly intended to improve performance

Information content

- Type of contract, e.g., mobile phone
- Type of handset . . .
  - Both may be inserted as is - or (better) replaced by reference information, e.g., churn rate
- Variables synonymous or improperly correlated with the target

Rows with invalid data

We could:

- delete row?   can produce biased sample depending upon the size of the problem
- fix data, replace by null, ignore error
  - e.g., DoB = 11-11-11

Rows containing some null values

- ignore the nulls: OK for some techniques
- delete row? OK if there are just a few random cases – but otherwise creates bias
- delete variable? … but not if variable has predictive capability
- estimate value?
- don’t insert a typical or average value (bias)
- don’t use special code for numeric nulls
- consider segmenting the model set

Deleting rows

- Some rows may be ignored due to the aims of the mining
  - e.g., a marketing campaign in London should focus on customers in London
  - Having restricted the data to customers in London, the “City” field then becomes one-valued, thus contains no information, and can be ignored.

Row duplication

Client#	Name	Address	Subscription Date	Magazine
23003	Smith	11, Wilton …	4/6/2008	Car
23003	Smith	11, Wilton …	7/9/2010	House
23003	Smith	11, Wilton …	…	…
…	…	…	…	…
23019	Smit	11, Wilton …	3/2/2014	Sports
…	…	…	…	…

- Some clients will be incorrectly represented by duplicate records

5 Predicting churn

- Suppose that we wish to predict who will churn next month
  - churn = ditch your service
- We use historical purchase and interaction patterns to predict ahead

Time series data

E.g., Mobile phone usage

- avg call bill; total call bill; avg call length; local calls vs non local; number of calls; number of calls exceeding some threshold
- aggregation by month, quarter, …
- holiday month (or not)
- proportion of spend during holiday months
- rate of growth of …
- variability of …

Predicting churn

- Suppose that we wish to predict who will churn in August --- we intend to send out a mail shot to prevent
- It is now June and we have customer data for past 9 months (Sept – May)
- Model set data therefore has the form

(Sept data, …, April data, churn in May)

target

- i.e., we use the past 8 months to predict one month ahead

Deployment of model

- When we come to deploy the model we wish to predict churn in August
- To do this we need data for past 8 months:
  - July, June, …, Dec.
- The problem is that the data for July will not be available until (say) the 2nd week of August
- By the time we've cleaned the data, deployed the model, sent out the mail shot, August will be nearly over!

Latency

- Instead we use a model set of the form

(Sept data, …, March data, churn in May)

target

- April is the latent month – not used
- To deploy this model to predict churn in August you do not need July data --- only up to June
  - Can now apply the model and get mail shot to customers by end-July.

Latency and churn

- Customer becomes dissatisfied with service
- Decides to switch
- Eventually investigates other offers
- Takes up other offer
- Eventually cancels existing service

- The decision to switch often takes place weeks or months before the actual cancellation!

Shelf life of predictive models

- Predictive modelling is based upon the assumption that the past is a good predictor of the future
- A predictive model will typically behave less well than predicted since the score set is from a later time-period than the model set
- Its performance will degrade further with time - the model has a shelf life
- Predictive models can overfit (i.e., pick up idiosyncrasies of) the past

Overfitting the past

- Suppose we have 12 months historical data
- Model set records of the form

target

(Month1, Month2, … Month10, Month12)

will tend to overfit the particular time period

- Instead use records of the form

(Month1, Month2, …, Month6, Month8)

(Month2, Month3,…, Month7, Month9)

(Month3, Month4, …, Month8, Month10)

(Month4, Month5, …, Month9, Month11)

(Month5, Month6, …, Month10, Month12)

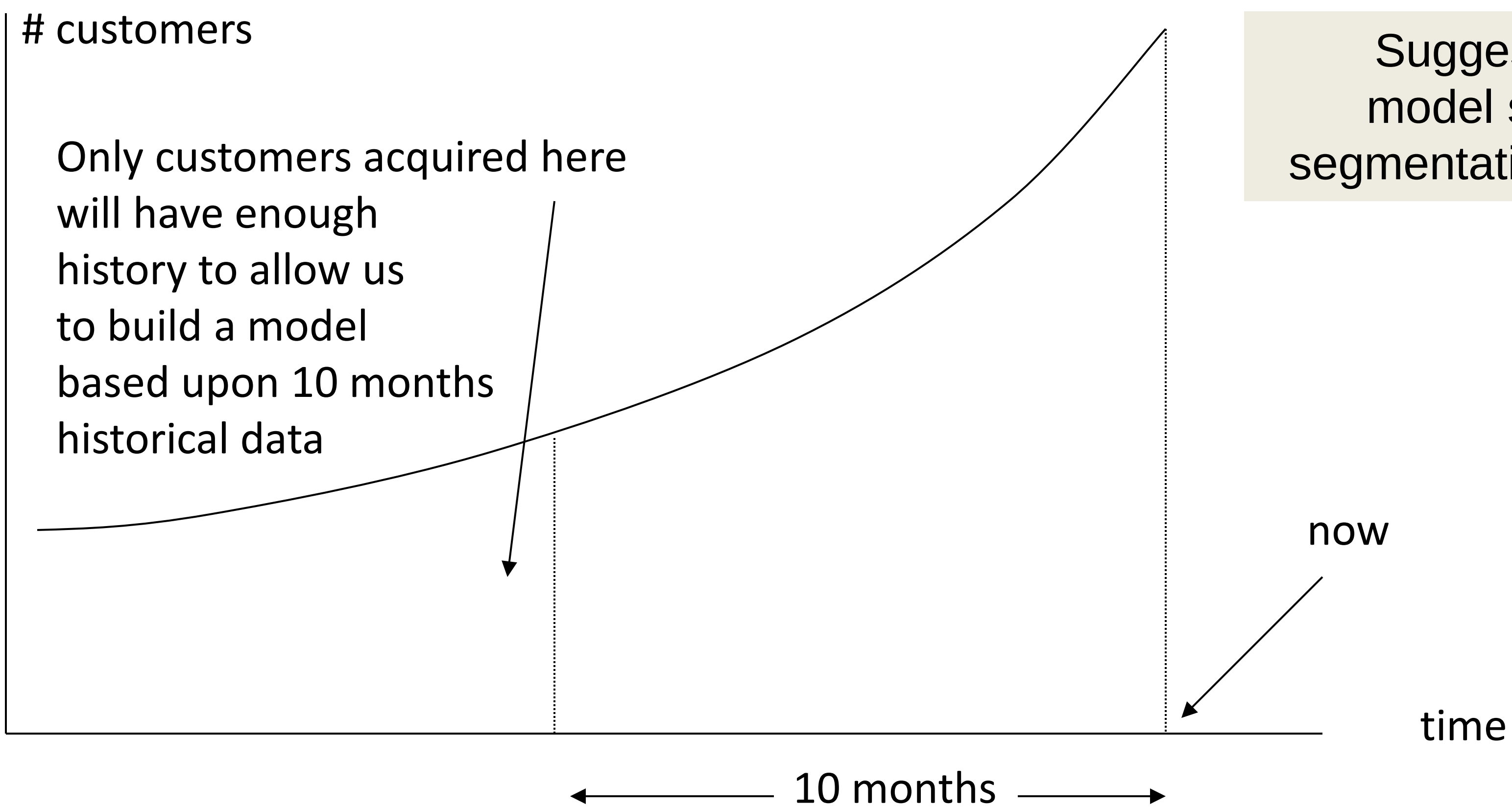
The new model set

- Does not overfit a particular time period
- Has 5 times as many records
  – although not as much history
- Can slide forward in time
  – as new data becomes available (for month 13) we can add records of the form

    (Month6, Month7, …, Month11, Month13)

  and then re-compute the model

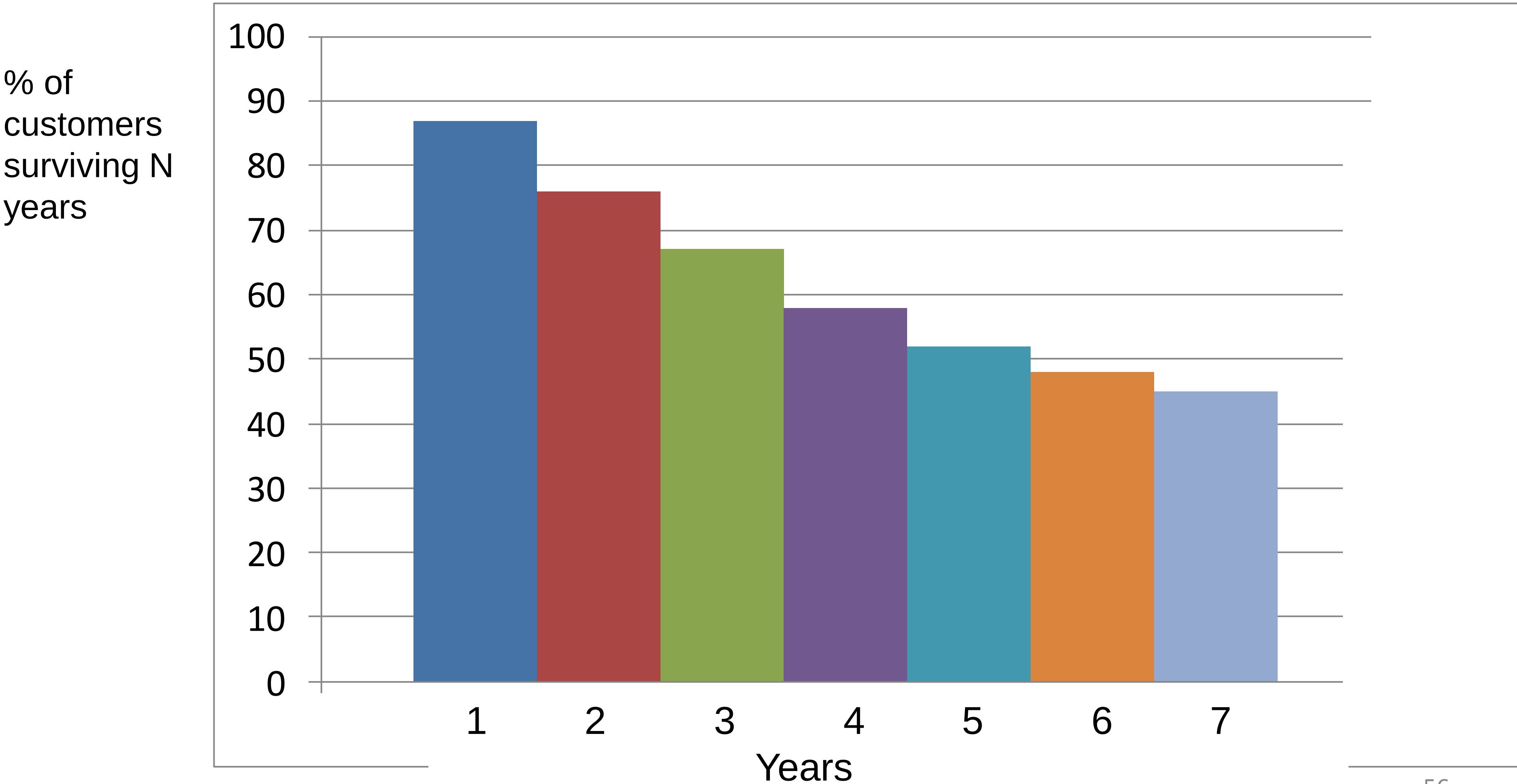
Predictive models based upon historical data will not apply to recently acquired customers



Suggests model set segmentation

6. Survival curves

For a given group of customers we can plot their longevity (as customers) in order to understand attrition rates and customer value



Analysing the survival curve

- Suppose that the values on the curve are

$$p_1, p_2, \ldots, p_7$$

- Suppose say 1000 customers initially, then these figures correspond to $n_1, n_2, \ldots, n_7$ customers where

$$n_i = p_i * 1000$$

The average survival length

- … is given by:

$$\frac{(n_1 - n_2)*1 + (n_2 - n_3)*2 + \ldots + (n_6 - n_7)*6 + n_7*7}{1000}$$

$$= (n_1 + n_2 + \ldots + n_7)/1000$$

$$= p_1 + p_2 + \ldots + p_7$$

Similarly we can compute (for example) the average customer value, or the value of retaining customers

When to worry about churn

- The *hazard* H_i at time i is the probability that an individual who has "survived" to time *i* will not survive to time *i*+1 (i.e., they'll churn)

$$H_i = 1 - (p_{i+1} / p_i)$$

- In some cases hazard fluctuation is well understood: E.g., for a contracted customer, hazard is initially high & then rises again at the end of the contract