Coding & STEM 4 Schools

2019 AI Workshop

Representing and Collecting Data

Presented by Mr Daniel Hickmott on 12th November 2019

Modelling a Solution

- When using Machine Learning, it is important to have:
  - An understanding of the appropriate data to collect
  - A variety of examples, including 'unusual' examples
- Having lots of data (examples) can be helpful too

Collecting the Appropriate Data

— We train Machine Learning models with observations (a data point - an album)

— These observations have different attributes (e.g. money spent on advertising)

— The attributes should some impact on the studied outcome (usually established through previous research)

— Bad example: Ice Cream Sales linked to Drowning Deaths

Identifying Attributes

— Can you think of other attributes for predicting album sales?

— Spend on advertising could have an impact on sales but there could be other factors

Collecting a Variety of Examples

- A wide variety of examples can improve our Machine Learning models' accuracy
- We aim to collect data from a representative sample, like a survey we want to collect responses across ages, genders, income etc
- A model that is trained on images could need a variety of images in a variety of lighting conditions
  - Mistakes could be made, e.g. animals in grass vs snow

Data for Student Projects

— Students could create data (text, sounds, images)
— Could use surveys (which we will do later)
— Publicly available data (e.g. data.gov.au)
— Could reach out to researchers (maybe?)
— Sample datasets in ML for Kids
— Could use hypothetical examples - without actually gathering data and training models

Hypothetical Example: Coffee Shop

— You own a coffee shop in a busy part of the city

— Your goal is to identify segments of customers (groups of 'similar' customers)

— The segmentation of customers involves an 'unsupervised learning' technique called clustering

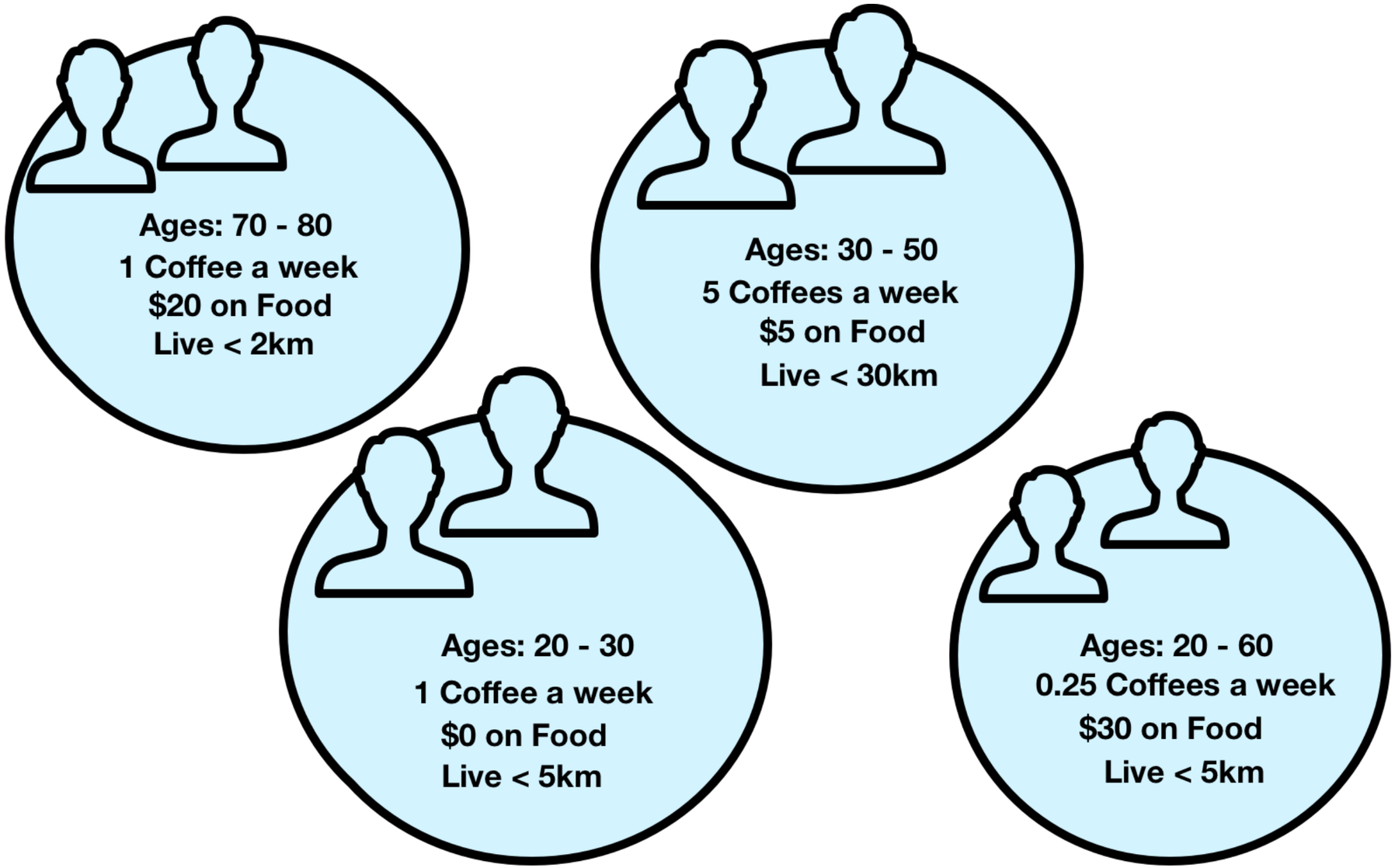
— Once you have the clusters you will develop targeted strategies to try to increase sales

Coffee Shop: Data to Collect

— Imagine that you can collect pretty much any information about the coffee shop's customers

— What attributes would you collect to use for grouping together 'similar' customers?

— Once you had these clusters, what strategies could you use to upsell to these customers?


Ages: 70 - 80
1 Coffee a week
$20 on Food
Live < 2km
Ages: 30 - 50
5 Coffees a week
$5 on Food
Live < 30km
Ages: 20 - 30
1 Coffee a week
$0 on Food
Live < 5km
Ages: 20 - 60
0.25 Coffees a week
$30 on Food
Live < 5km

Journey to School

- Next, we will work through an another Machine Learning for Kids activity: Journey to School
- Involves collecting data (observations) through a survey
- Each observation has attributes
- We will use these observations to train a model to predict whether a student travelled to school by car, bike, walking or bus