Analytics and Econometrics for Business Decisions

Dr Kandrika Pritularga

Lancaster University & Universitas Diponegoro

Semarang, Indonesia, 2 September 2025

Copyright © 2025 CMAF. All rights reserved.

Outline

- Introduction
- Data, analytics, and insights
- Understanding business decisions
- Analytics in business decisions
  - Operations Decisions
  - Marketing Decisions
- Conclusions

CMAF

Objectives of the talk

By the end of the talk, participants are able to

- Recognise of the role of analytics to transform data into insights to support business decisions;
- Understand ways to define business problems as the basis of implementing analytics in business and management;
- Exemplify the roles of analytics in operations and marketing decisions.

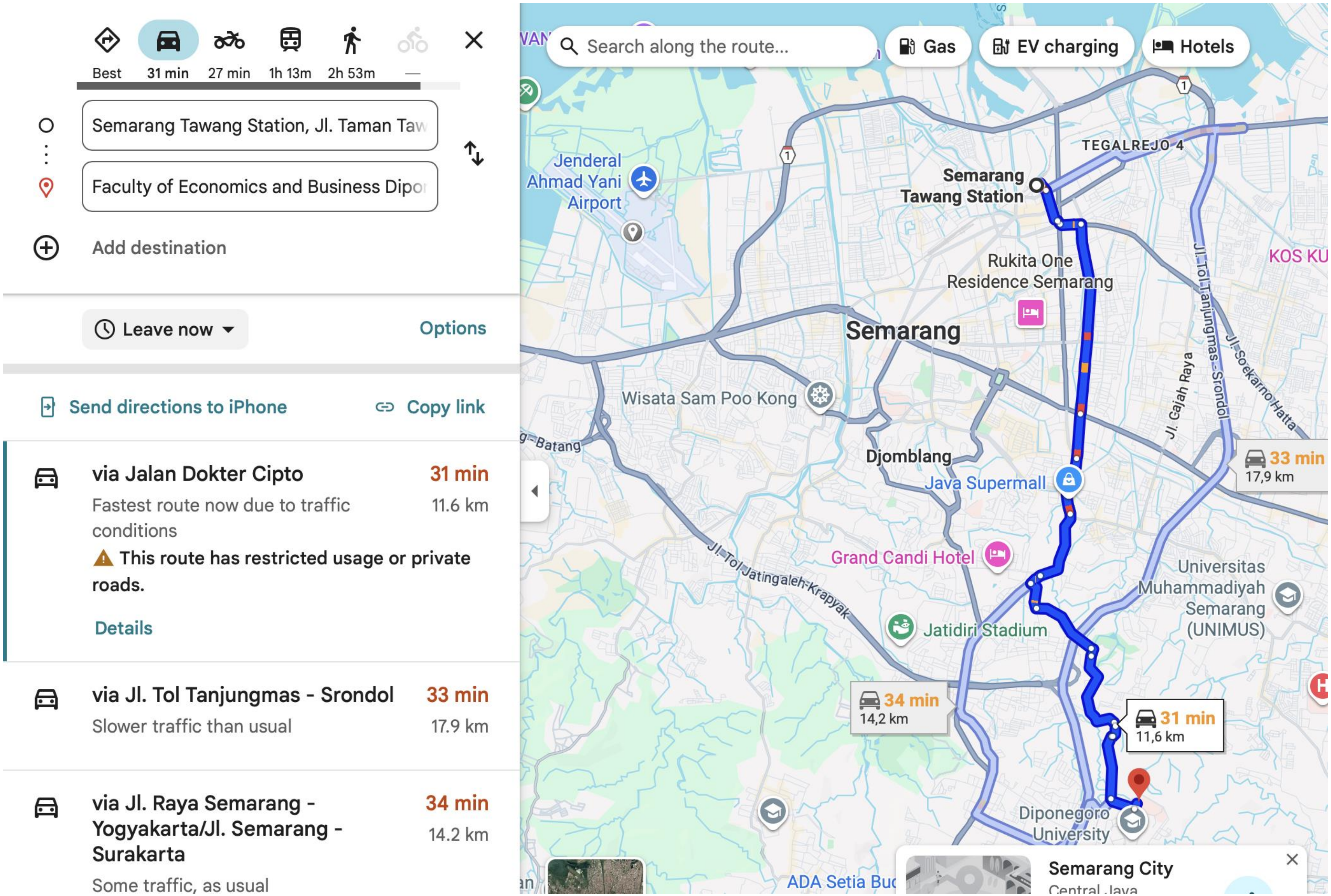
CMAF

Do we make decisions every day?

https://id.wikipedia.org/wiki/Stasiun_Semarang_Tawang#/media/Berkas:Stasiun_Semarang_Tawang_2024.jpg

- After this session, you need to take a train to Jakarta at 2pm
- The session finishes at 12pm
- What decisions do you need to make to get on the train on time?

Do we make decisions every day?



Decisions in retail industry

Think yourself as a manager of a retail store

What decisions do you need to make?

Decisions in retail industry

- How often do we need to replenish the store inventory?
- What promotions are we running for the next quarter?
- How do we schedule our cashier staff? When should we open/ close a till to avoid long queues?

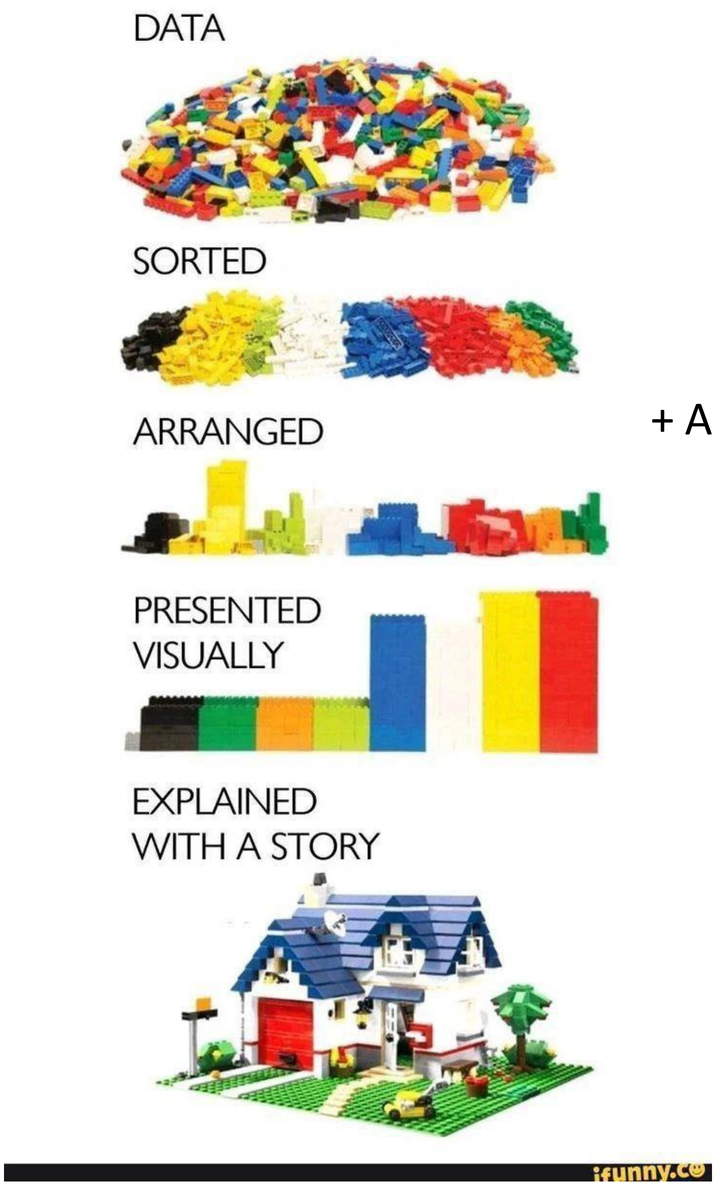
Analytics

- Analytics is the scientific process of transforming data into insight for making better decisions (INFORMS, 2023).

Data → Insight → Decision

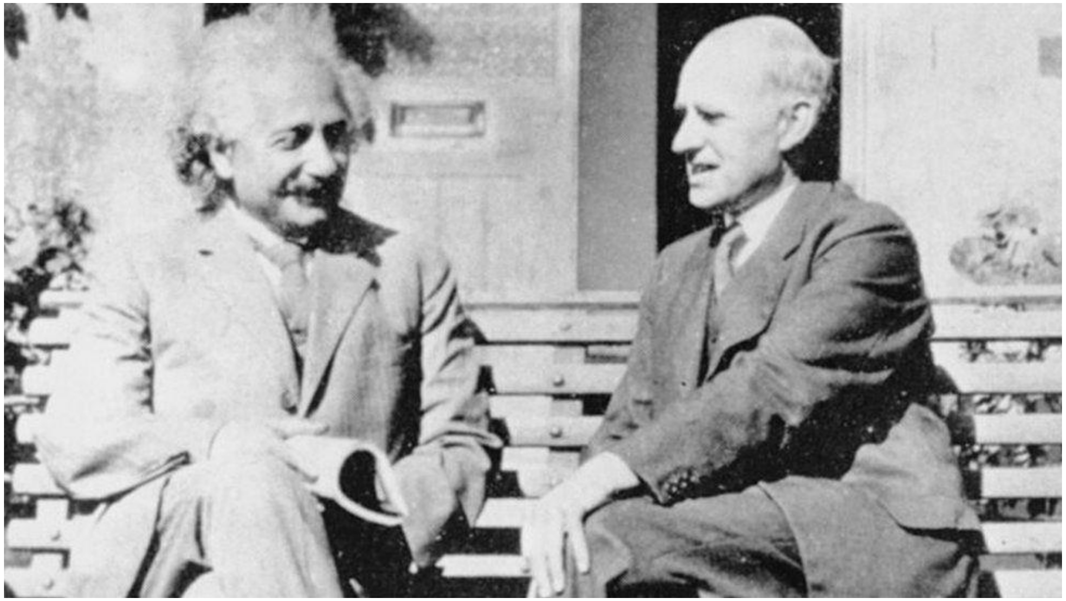
- We will discuss (1) data, (2) gaining insights from explanatory data analysis and modelling, (3) use insights to support decisions.

Analytics



+ Analysed!

In the past, it is difficult to obtain data



https://ichef.bbci.co.uk/news/976/cpsprodpb/3620/production/_107065831_6a2961c6-5860-4d9f-a669-fb93e60f95b4.jpg.webp

Einstein and Eddington

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > books & arts > article

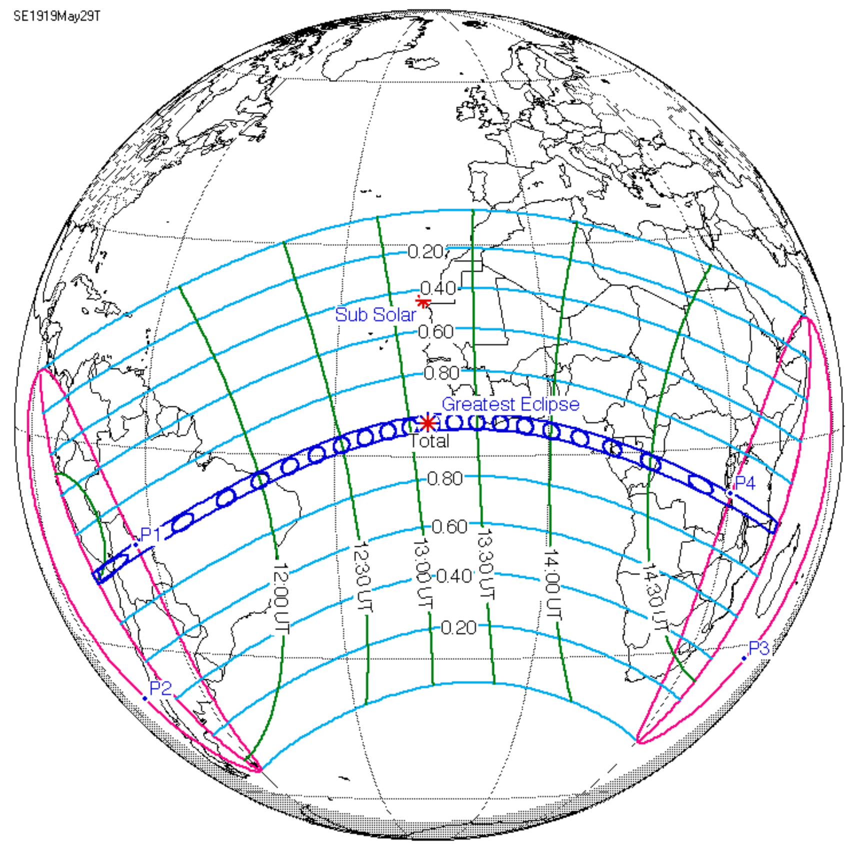
BOOKS AND ARTS | 15 April 2019 | Correction 17 April 2019

Einstein, Eddington and the 1919 eclipse

Peter Coles weighs up three books on the momentous expedition that proved the general theory of relativity.

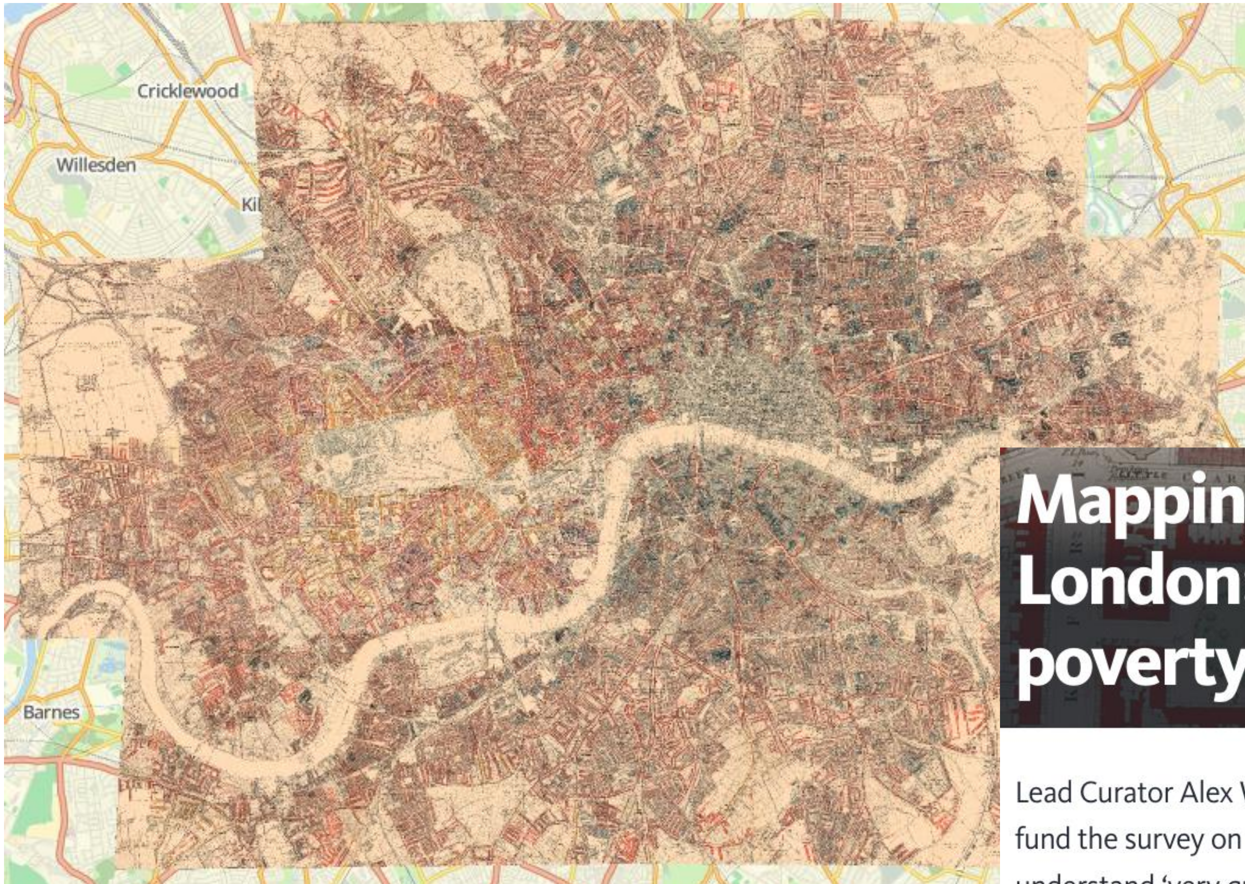
Peter Coles

Total solar eclipse in 1919



https://upload.wikimedia.org/wikipedia/commons/4/4d/SE1919May29T.png

In the past, it was difficult to obtain data



https://booth.lse.ac.uk/map/12/-0.1109/51.5080/100/0

Mapping wealth & poverty in London: Charles Booth's poverty map

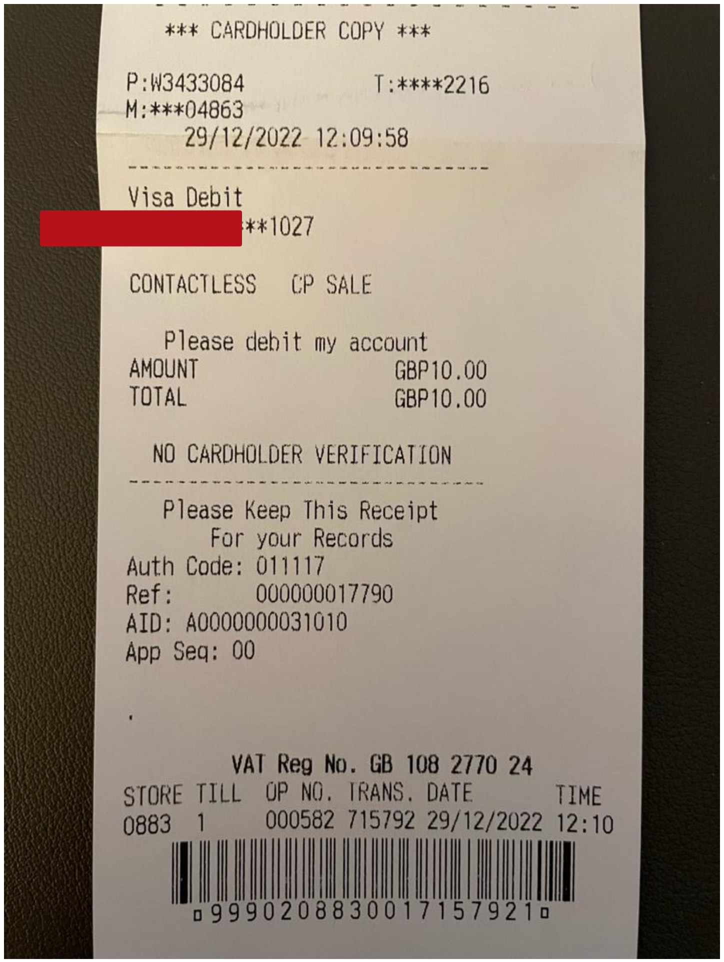
Lead Curator Alex Werner explains what prompted Charles Booth to self-fund the survey on which the map is based — which allows the onlooker to understand 'very quickly', where the wealthy and poor people of London lived.

22 September 2022

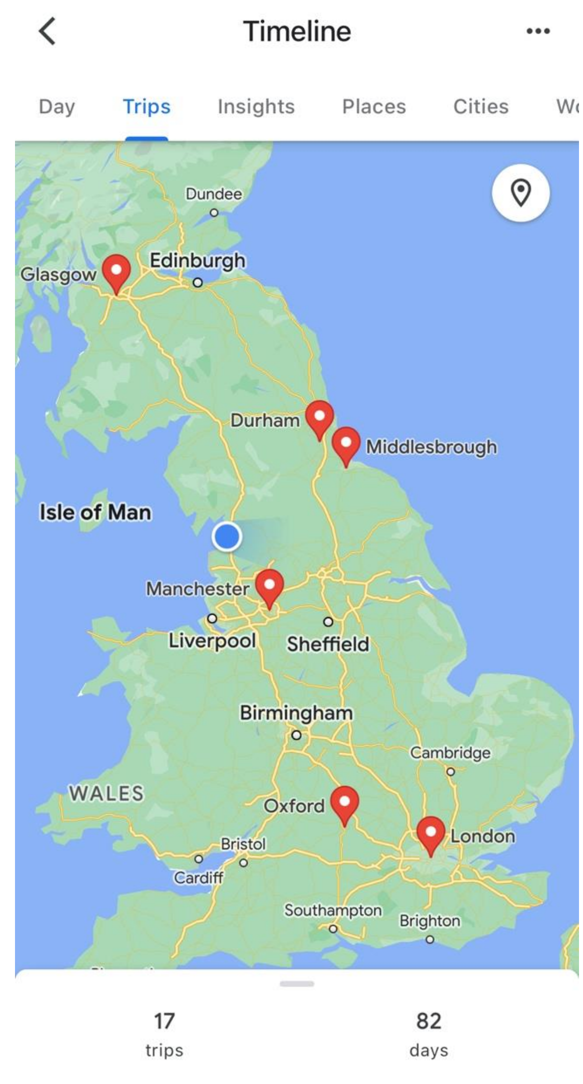
Charles Booth's poverty maps are colour-coded maps of Victorian London. They were funded by Booth — a Liverpool ship owner, who came down to London — and are based on a survey of the population between 1886 and 1903.

https://www.museumoflondon.org.uk/discover/mapping-wealth-poverty-london-charles-booths-poverty-map

Now data is easier to obtain and abundant!

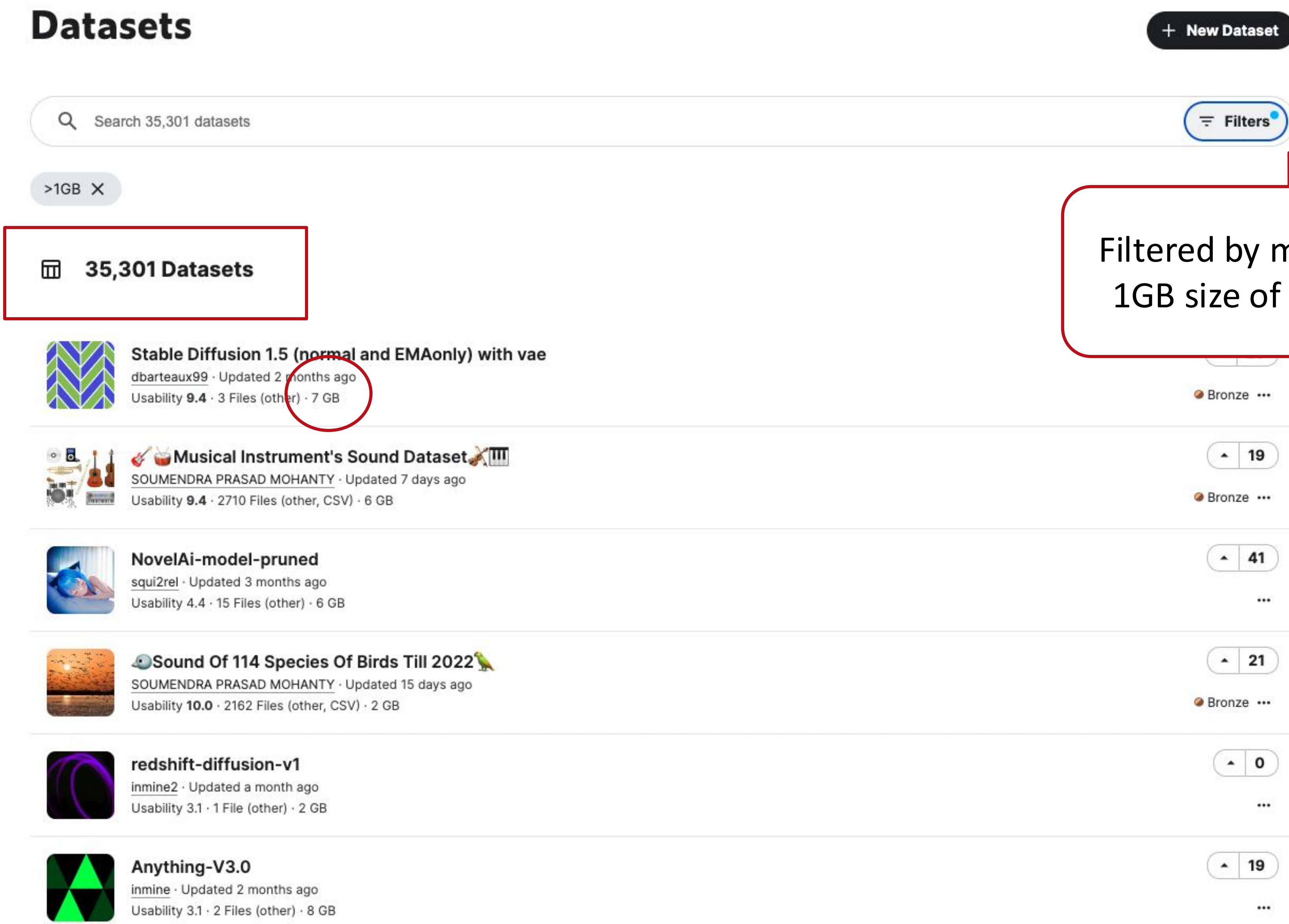


My transaction is recorded in a few second



My trips, recorded by Google

Now data is easier to obtain and abundant!



Filtered by minimum 1GB size of dataset

You can find thousands of datasets in internet, e.g., Kaggle

Technology is advancing…

We collect and access more data

We have more powerful computers (faster and better CPU and graphics)

These accelerate how we use data for analysis and supporting decisions

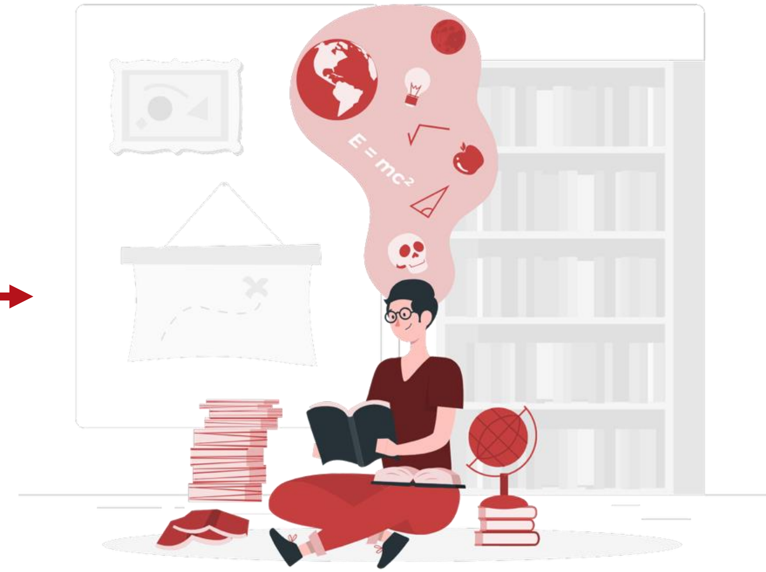
Too much data?

- We know that data is a lot easier to collect and more abundant due to technology advancement
- Having a lot of data does not mean we solve our problem
- **Data is not knowledge!**
- We need to extract ‘knowledge’ from our data via data analysis techniques
- Data may contain ‘knowledge’ and much ‘noise’

From data to insights

Data Analysis is part of a larger process that extracts knowledge from data

Analytical tools



Data

Single instances
Describe individual properties
Available in large amounts
Easy to obtain
Unable to make predictions

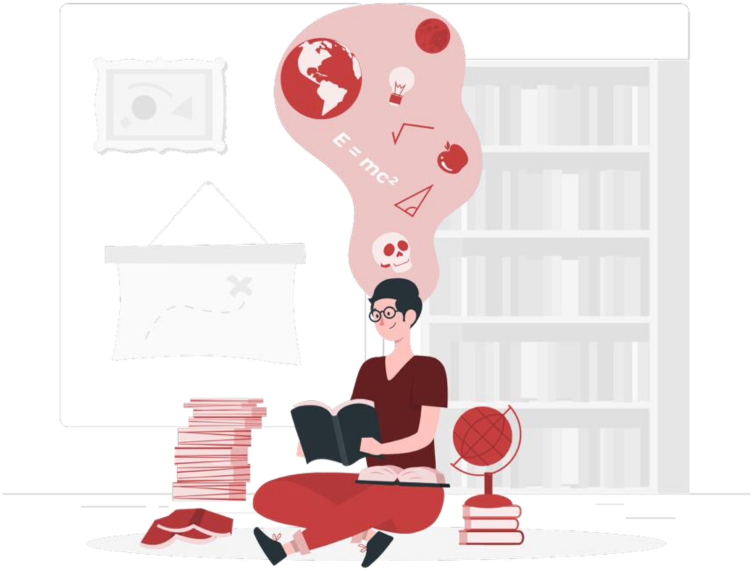
Insights

Classes, instances
General patterns, laws, structures
Available in short sentences
Time consuming
Able to make predictions

From data to insights

Insights

Correctness
Probability, success in tests



Usefulness
Relevance, predictive power

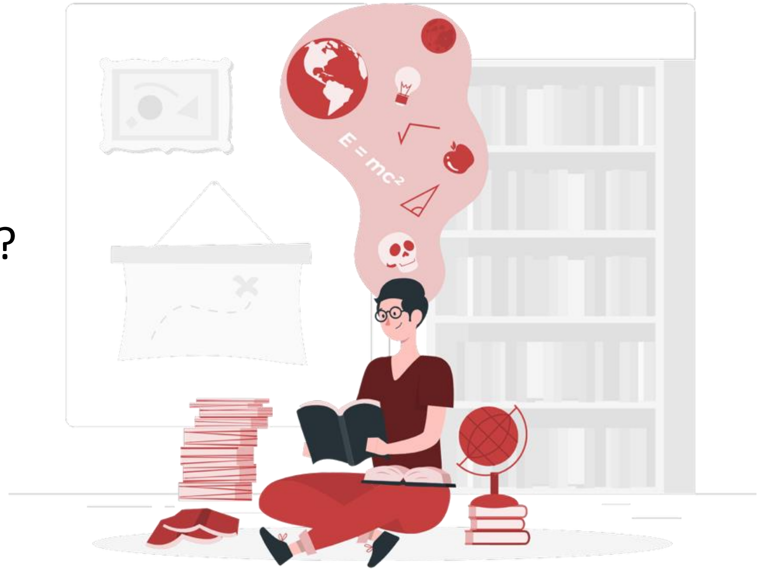
Generality
Domain, condition of validity

Comprehensibility
Simplicity, clarity, parsimony

Novelty
Previously unknown

From data to insights

People tend to buy beers and nappies at the same time



Correctness
Is this statement correct?
How much is the probability of this statement being true?
Can we test it?
Hypothesis testing

Usefulness
Can we use this knowledge increase our sales?
Can we predict this behaviour to happen in the other supermarket?
Our prediction, our predicted behaviour

Generality
Does this statement apply to all supermarkets in Lancaster?
Does this apply to all days of the week?
Scope of our problems

Novelty
Is it a new finding?

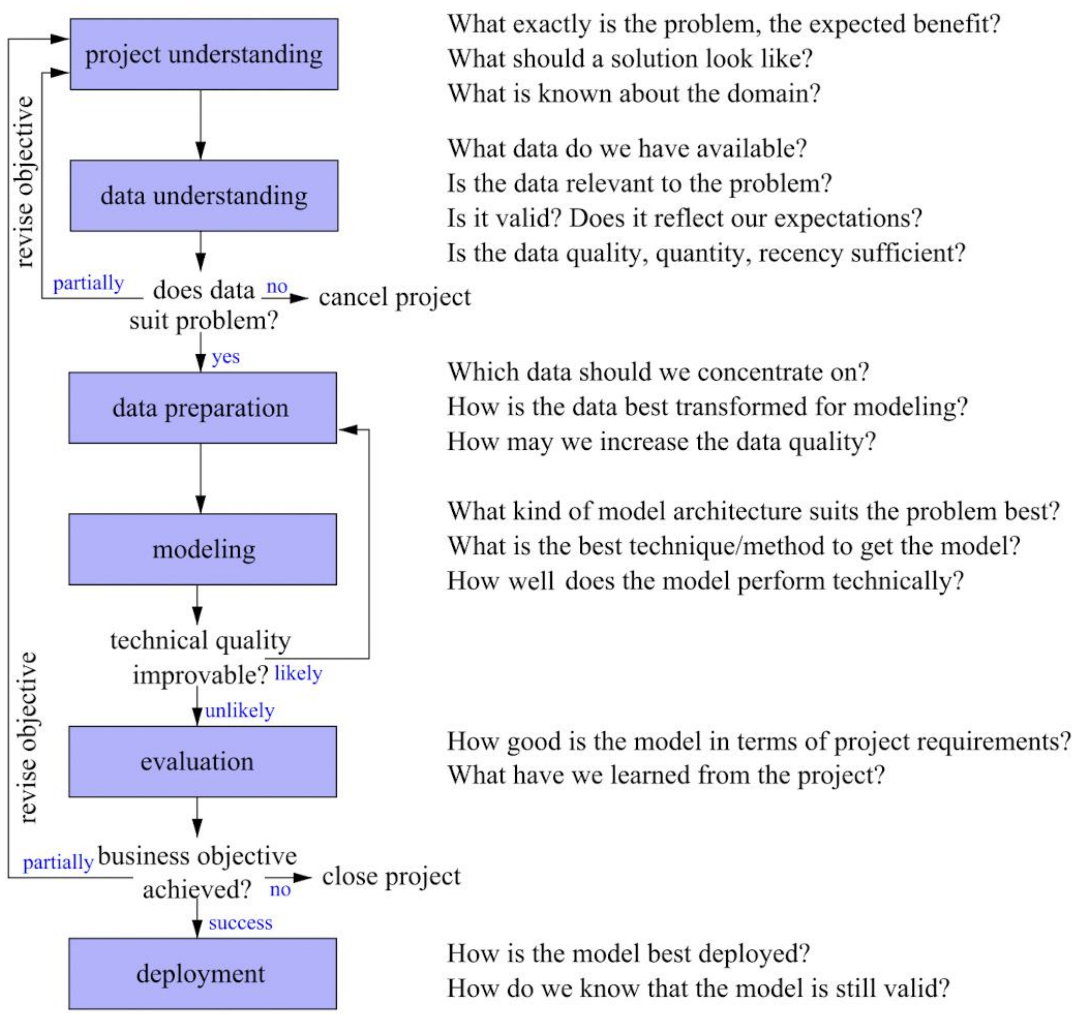
Comprehensibility
Can you explain the reason why customers buy beers and nappies at the same time?
Long story to tell…

Source: https://tdwi.org/articles/2016/11/15/beer-and-diapers-impossible-correlation.aspx

Cross-Industry Standard Process for Data Mining

CRISP-DM Framework

Berthold (2010) Chapter 1 Introduction



Data Analysis Problems

- **Classification**
  - Predict the outcome of an experiment with a finite number of possible results
  - Eg: Is this customer credit-worthy?
- **Regression**
  - Predict the outcome of an experiment with a numerical variable
  - Eg: How much money will the customer spend for vacation next year? Refer to MSCI521
- **Both are <u>supervised learning</u> tools**
  - We have a target variable that needs explanation

Data Analysis Problems

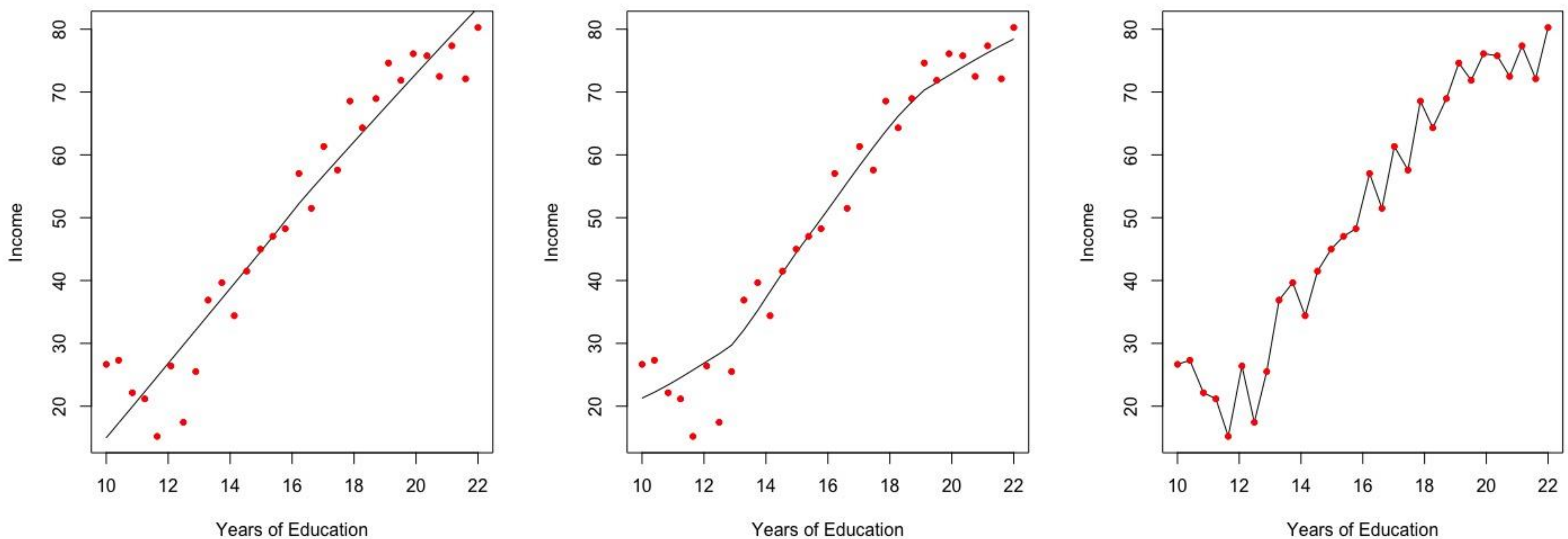
- **Clustering**
  - Forming groups of similar cases
  - Eg: Do my customers divide into different groups?
  - This is often called an **unsupervised learning** tool; we do not have a target variable that needs explanation
- **Association**
  - Describing the interdependencies of variables
  - Eg: How do the various qualities influence each other?

Classification Problem

- Information about different 'objects' encoded as feature vectors/ explanatory variables, denoted as X
- Qualitative variable of interest Y takes (unordered) values:
  - $email \in \{spam, no\ spam\}$
  - $debit\ card\ transaction \in \{legitimate, fraudulent\}$
- Classifier: Function $f(\cdot)$ that maps X to $P(Y|X)$
- Main Goals in Classification
  - Prediction
  - Assess uncertainty in prediction
  - Understand role of different variables/ predictors/ features

Regression versus Classification

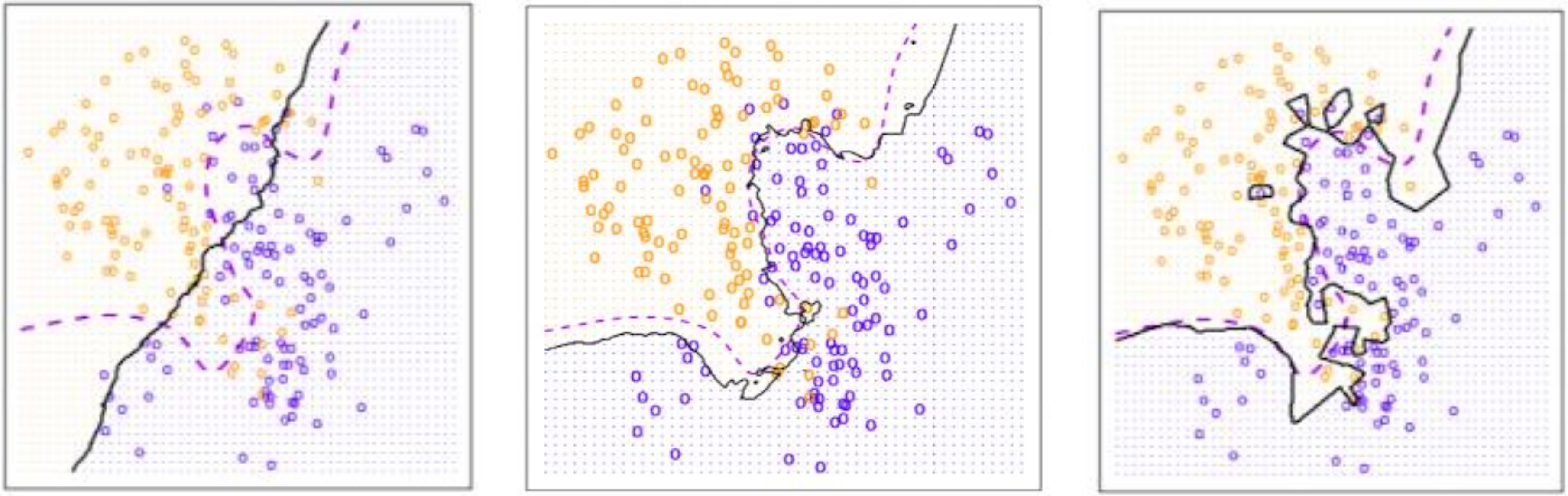
- Both problem involve finding a function to predict Y from a given set of pairs of $(x_i, y_i)_{i=1}^{n}$ for n observations
- Regression: Y is numerical
- Objective: Line/ surface of best fit



Source: James et al. (2013) An Introduction to Statistical Learning

Regression versus Classification

- Both problem involve finding a function to predict Y from a given set of pairs of $(x_i, y_i)_{i=1}^{n}$ for n observations
- Classification: Y is categorical
- Objective: Line/ surface of best discrimination



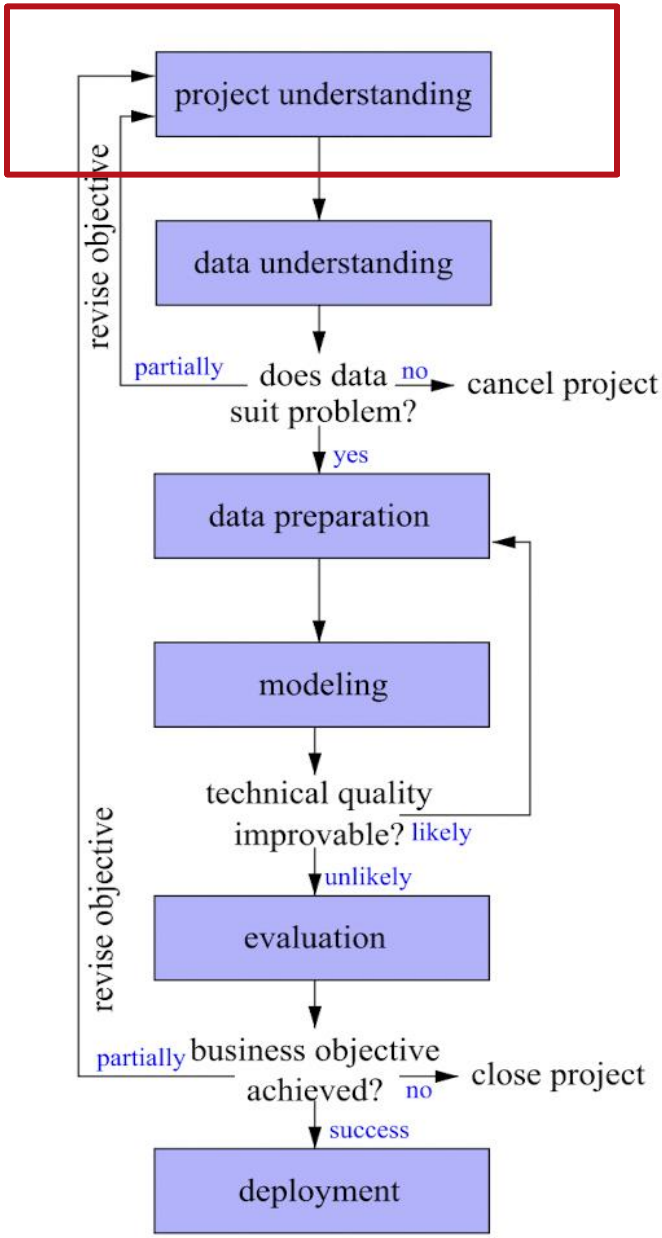
Source: James et al. (2013) An Introduction to Statistical Learning

UNDERSTANDING BUSINESS DECISIONS

- Determine the Business Objective
- Assess the Situation
- Determine Analysis Goals

Berthold (2010) Chapter 3 Project Understanding

Data Analysis Process



Business Decision Understanding

- We have to map a business decision onto one or many data analysis tasks
  - The dataset needs to align well to the problem
- Goals of understanding the business decision
  - to assess the main objective,
  - to identify potential benefits, constraints, assumptions, and risks,
  - to ensure that the tasks fulfil the main objective of the project

Business Decision Understanding

- Who are the stakeholders in our business decision?
- Depends on our role as an internal or an external analyst, potentially, we need to with:
  - Functional managers (marketing, procurement)
  - Data scientist, data analyst
  - Strategic managers, C-levels
- Each of them may have different views about a decision

Business Decision Objective(s)

- We need to set simple, clear, precise objectives and success criteria to **unite different views**
- Is it a clear objective?
  - Model profitable customers to increase sales
- What can be improved?
  - How much increase? What is the benchmark?
  - Who are the profitable customers?
  - Time frame? Increase sales in how many weeks?
  - Successful project criteria?
- Ensure all stakeholders speak the same language
  - Same assumptions, same definitions, etc.

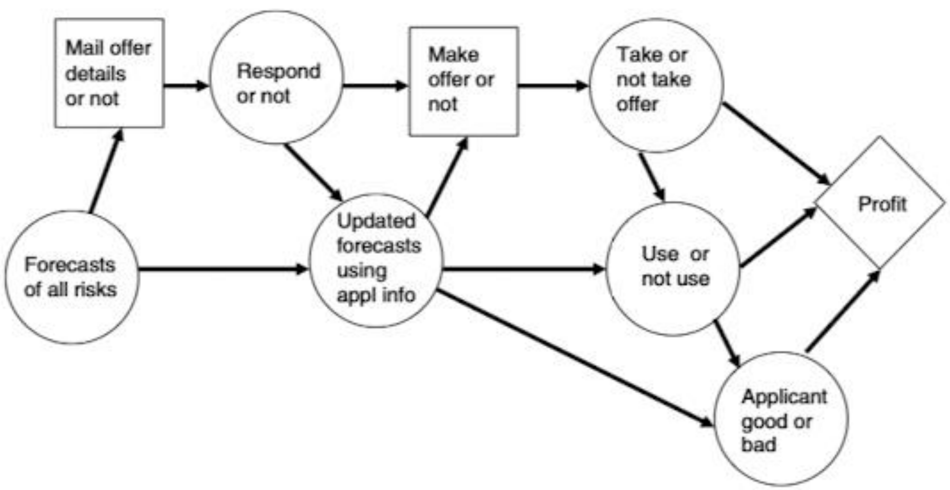
Domain/ Context Exploration

- What is the worst-case scenario?
  - Due to poor communication, each stakeholder creates their own assumptions and perspectives
  - At the project completion, all stakeholders realise that results do not meet expectations
- Issues from poor communication can be exaggerated by the fact that we as an analyst do not know the context of the problem (domain)
  - How can we deal with these issues?

Context Exploration

- We can use qualitative methods to understand the project better, using Problem Structuring, Soft System Methodology
- An example: Cognitive Map
  - Help understanding the problem more clearly
  - Help determining what variables are in the dataset

Talking to clients
Interviews



Source: Thomas (2009) Consumer Credit Models: Pricing, Profit and Portfolios

Analytics in business decisions

Business decisions as contexts

- The previous discussion implies the importance of understanding the business decision.
- Business decisions provide clear objectives, success criteria, stakeholders involved.
- Analytics extract insights from data to support decisions.
- Let's have a look at the examples below.

Decisions	Operations	Marketing
Strategic decisions	Supply chain design/ Buyer-seller contract mechanism	Brand positioning, customer segmentation
Tactical decisions	Inventory policies	Promotion campaigns
Operational decisions	Ordering decisions	Targeted emails

Analytics in business decisions

- How can we support decisions with analytics?
- We can split tasks in analytics into three tasks:
  - **Descriptive** analytics: focuses on summarising and interpreting data (summary statistics, association, clustering).
  - **Predictive** analytics: focuses on utilising historical data to predict the future (regression, classification)
  - **Prescriptive** analytics: focuses on recommending specific actions to achieve desired outcomes (beyond data analysis).

Decisions	Operations	Marketing
Strategic decisions	Supply chain design/ Buyer-seller contract mechanism	Brand positioning, customer segmentation
Tactical decisions	Inventory policies	Promotion campaigns
Operational decisions	Ordering decisions	Targeted emails

Analytics and Econometrics for Operations Decisions

Newsvendor problem

Newsvendor problem

Imagine a vendor selling newspapers for just one day.

- Too few copies: lose potential sales.
- Too many copies: leftover waste.

Decision: How many newspapers to order?

Why is the newsvendor problem important?

- How to make decisions under uncertainty.
- Why balance between shortage and excess is important.
- A simple yet powerful model for inventory decisions
- Applicable in many contexts, such as perishable products in a supermarket (bakery, vegetables, fruit), airline passengers

A simple newsvendor problem

Let’s use a simple hypothetical newsvendor problem

- Each newspaper costs $0.2 to buy.
- Sold at $1 to customers.
- Unsold copies have no value.

Profit depends on finding the right balance!

- Ordering too few: Missed profit from lost sales.
- Ordering too many: Money wasted on unsold copies.

Goal: Find the order quantity that balances these risks.

A simple newsvendor problem

- Objective: maximise the expected profit
  - Profit = Revenue – purchase cost + salvage value
  - We can calculate the optimal order quantity via calculus

Optimal solution:

- $Q^* = F_D^{-1}(CR)$, where F_D^{-1} is the inverse cumulative demand distribution.

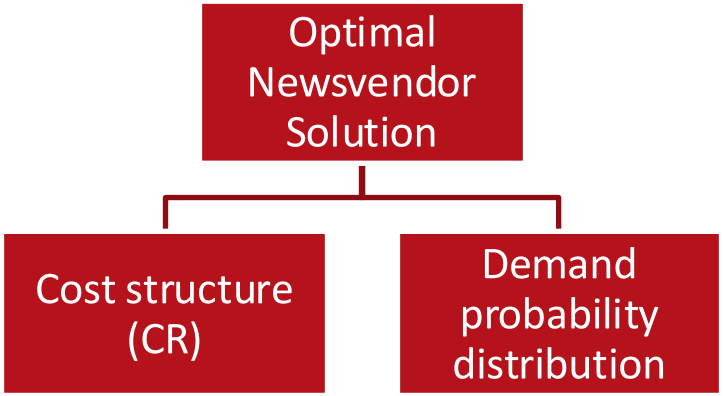
  The critical ratio:

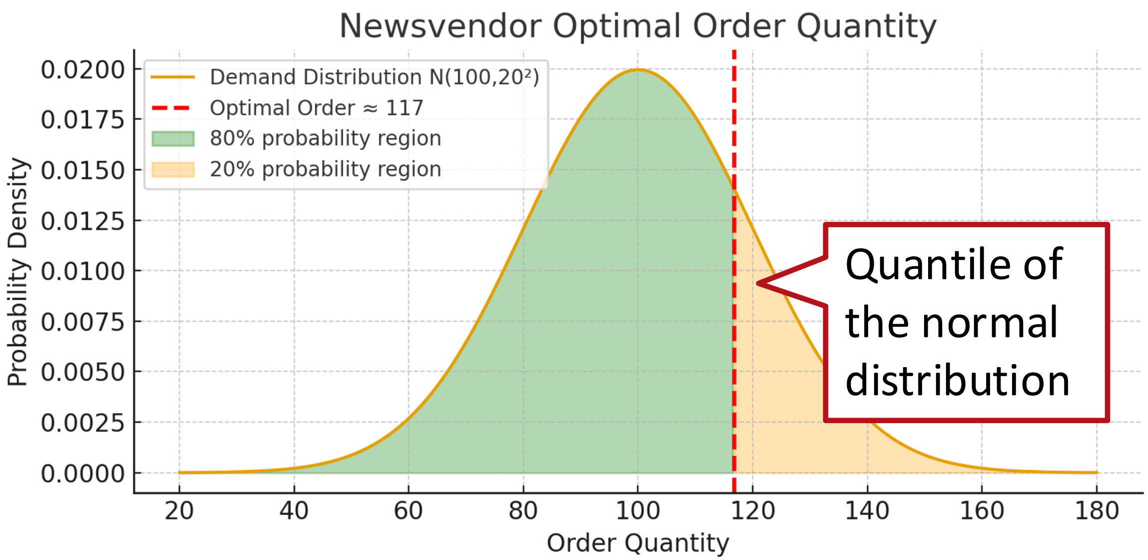
  $$CR = \frac{Cu}{Cu + Co}$$

  - Cu is lost margin when demand exceeds stock
  - Co is lost margin on unsold units

A simple newsvendor problem

- Demand follows a Normal distribution with the mean of 100 and the standard deviation of 20
- The optimal order is where the cumulative probability equals the critical ratio.



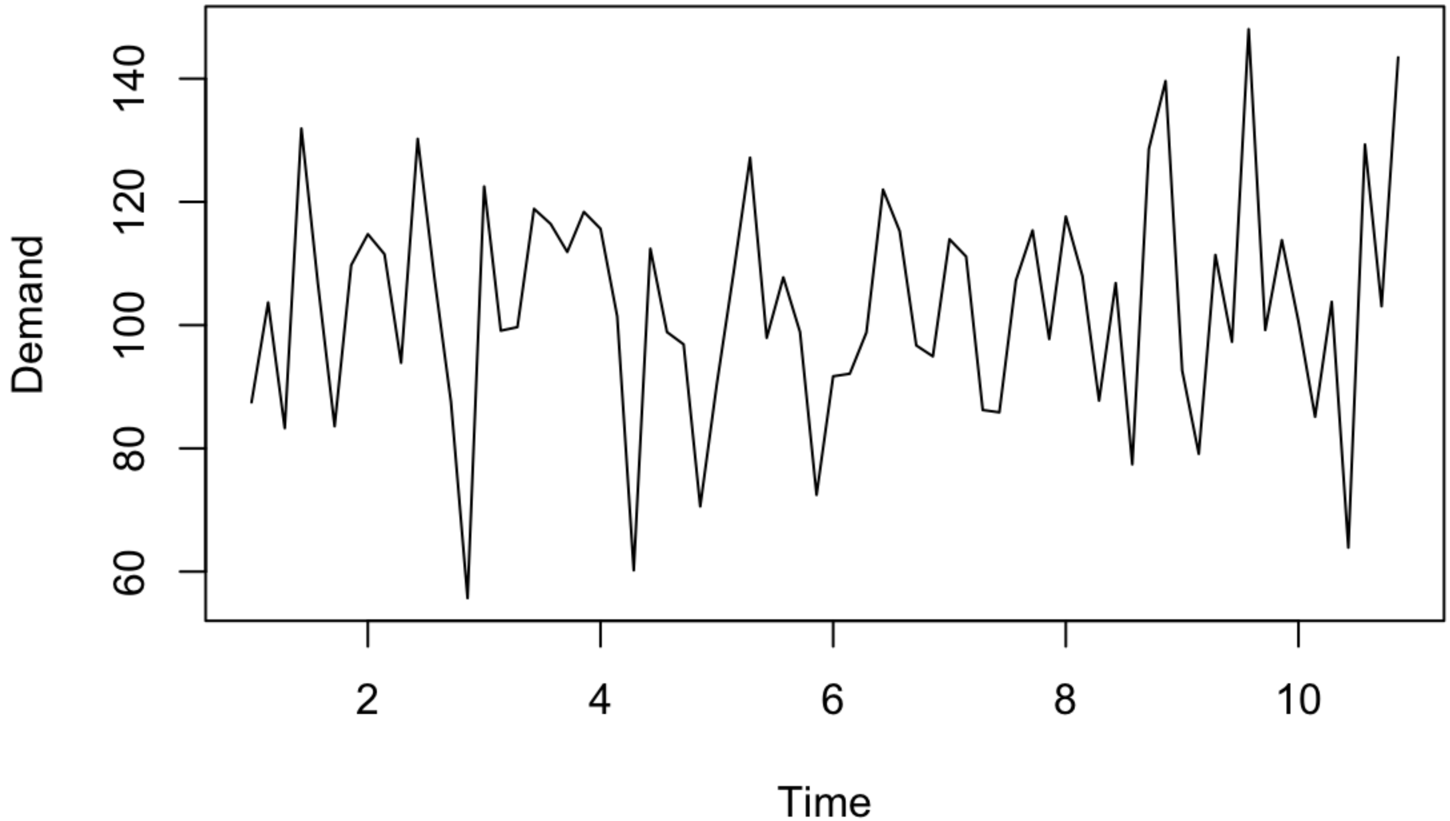


In practice…

- Demand distribution is unknown
- We need to estimate the predictive demand distribution using demand forecasting
- The alternative econometric models:
  - ARDL
  - ARIMA
  - Exponential smoothing
  - Kalman filter
  - …

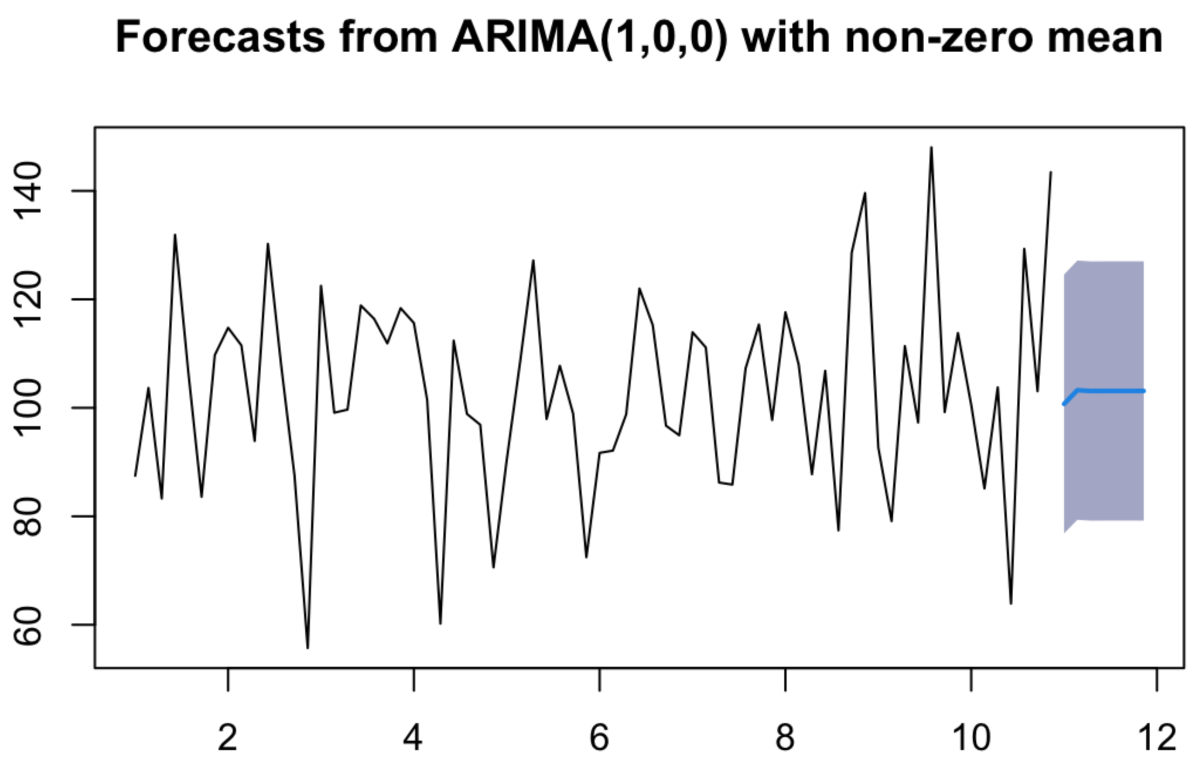
Demand forecasting

- Let D be the demand for the newspaper
- We have the historical data of D and it's shown as



Demand forecasting

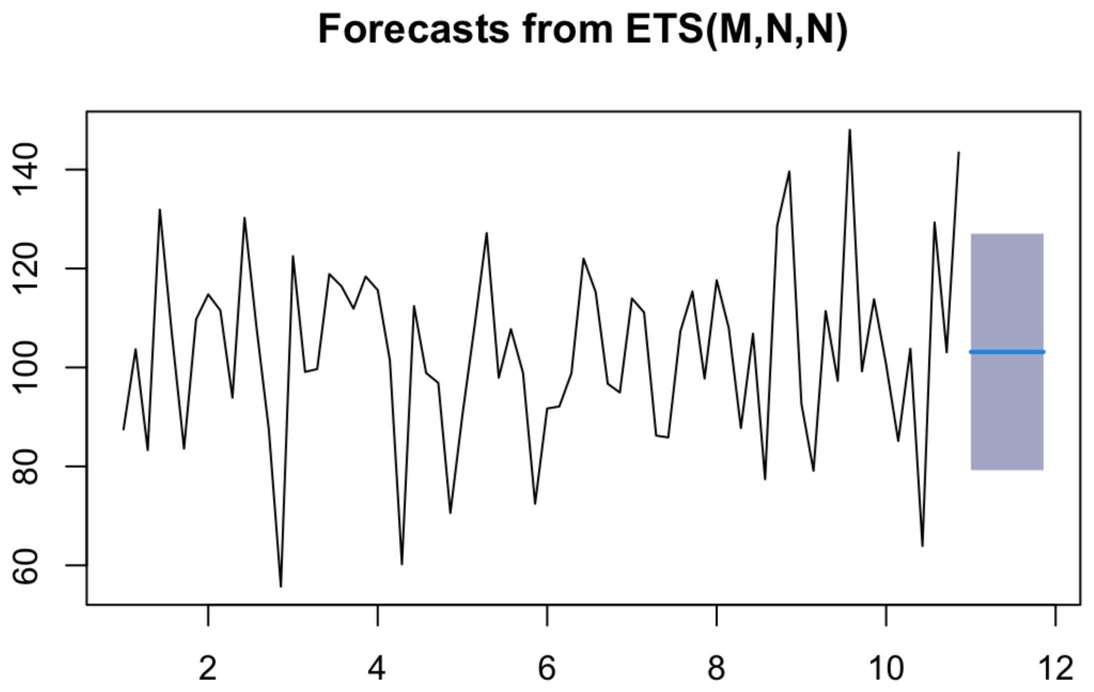
- We do not know D and attempt to forecast D using ARIMA(1,0,0)



The ‘optimal’ **order quantity** from the estimated demand (using ARIMA(1,0,0) is **125**

Demand forecasting

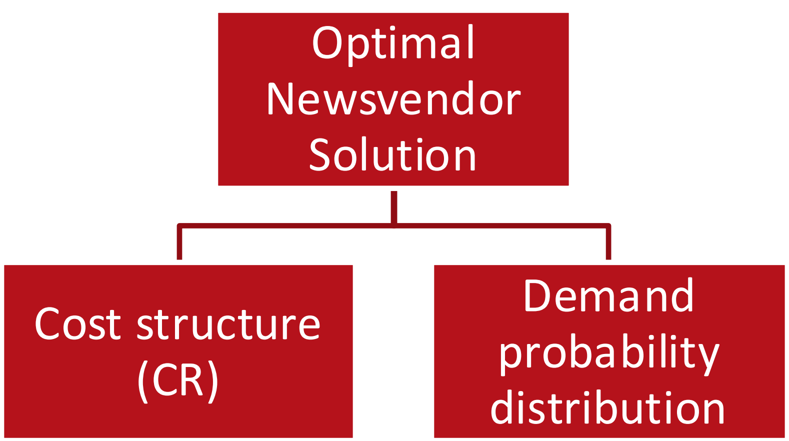
- We do not know D and attempt to forecast D using ETS(M,N,N)



The 'optimal' **order quantity** from the estimated demand (using ARIMA(1,0,0) is **127**

Forecasting and inventory problems

- If the future demand is known, we do not need forecasting
- However, in practice, we do not know the future demand
  - We need to predict the demand using statistical/ econometric models
  - In this example: we use ARIMA and Exponential smoothing
  - The ‘optimal’ order quantity depends on the forecasting models



Demand?	Optimal order if…
Demand is known	117
Estimated demand (ARIMA)	125
Estimated demand (ETS)	127

Analytics and Econometrics for Marketing Decisions

Perceptual maps

A young and ambitious company wants to conquer the world…

…by selling slippers.

- What slippers should they produce?
- Use Segmentation & Targeting to decide, which segment to aim for.

- What do we have in our segment?

They need our help!

Perceptual maps

We have selected a segment of home slippers for the middle class.

- Is there a niche for our product?
- Is there a demand for a product that is not yet sold?
- What should be the characteristics of the product?

- So, who are your competitors?
- What is the product mix of your competitors?
- What are the characteristics of the slippers?

Perceptual maps

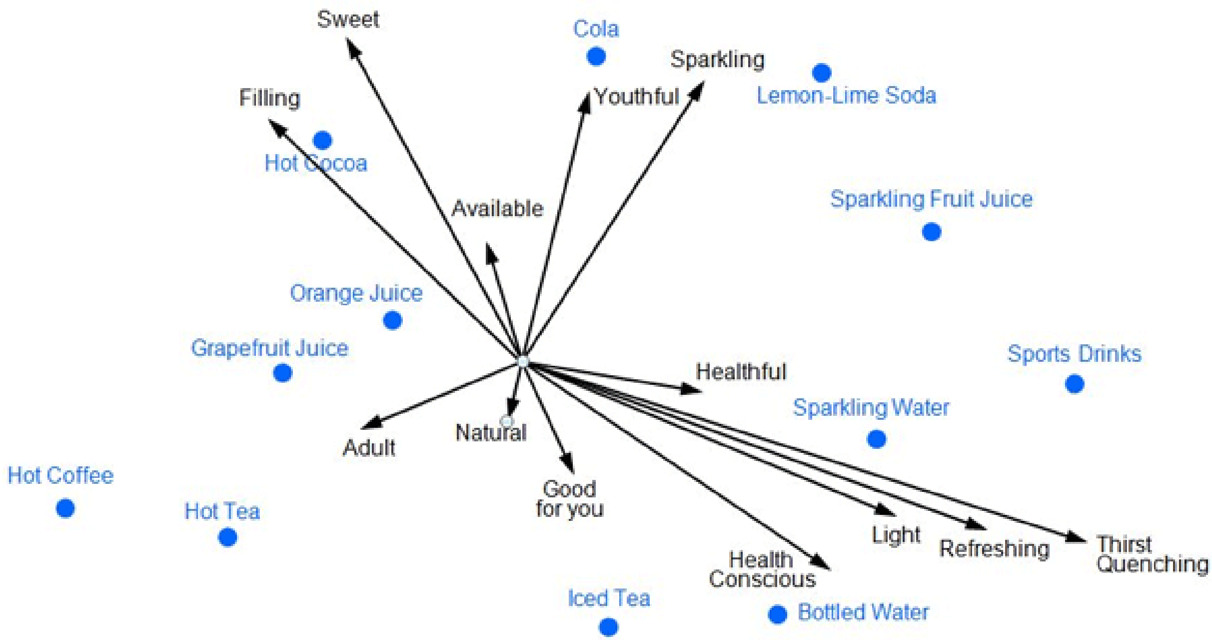
Name	ColourHue	ColourValue	Age	Sex	MaterialType	MaterialWarmth	Purpose	Feature
NordvekMen	Blue	Medium	Adult	Male	Soft	Warm	Home	General
NikeLadies	Red	Light	Teenager	Female	Medium	Cold	Pool	General
BlizzardYeti	Blue	Light	Junior	Unisex	Soft	Warm	Home	General
ASOSWomen	Blue	CLight	Adult	Female	Soft	Warm	Home	General
Hajj Flip Flop	Blue	CLight	Adult	Unisex	Hard	Cold	Bath	General
NikeJDI	Blue	Heavy	Adult	Male	Hard	Cold	Beach	General
ShiptonVelvet	Violet	CHeavy	Teenager	Female	Medium	Medium	Home	General
DancersPink	Red	Light	Juniour	Female	Medium	Cold	Dancing	General
OrthoNimble	Yellow	Heavy	Adult	Male	Medium	Cold	Beach	Ortho
IKEABath	Blue	Heavy	Adult	Unisex	Soft	Medium	Home	General

What are the most important characteristics of the slippers?

This would answer: what characters of slippers do we need to develop and think about more?

Perceptual maps

Multidimensional maps are not very helpful:



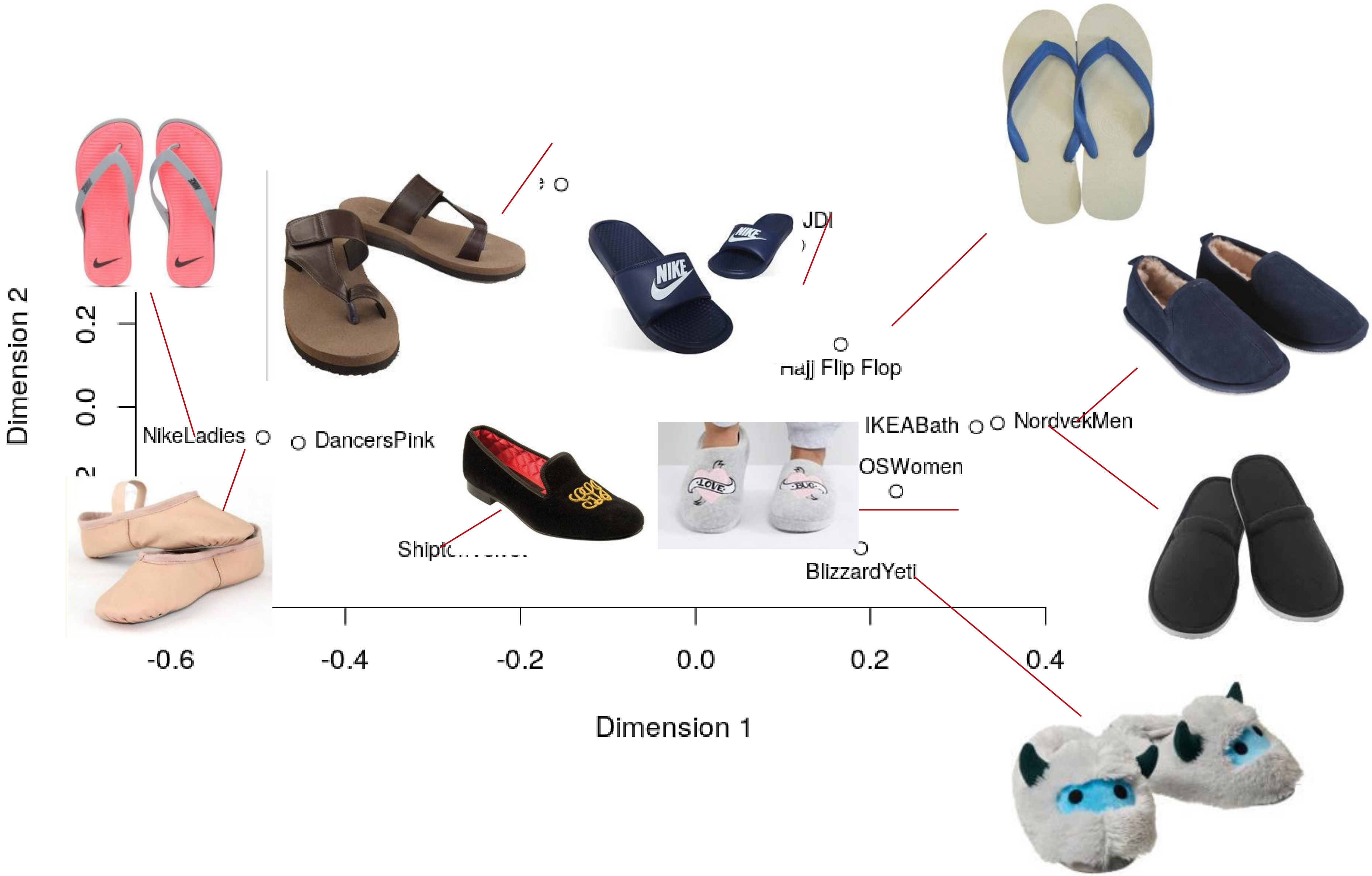
Perceptual Map example

Reducing the dimension helps us understand the problem

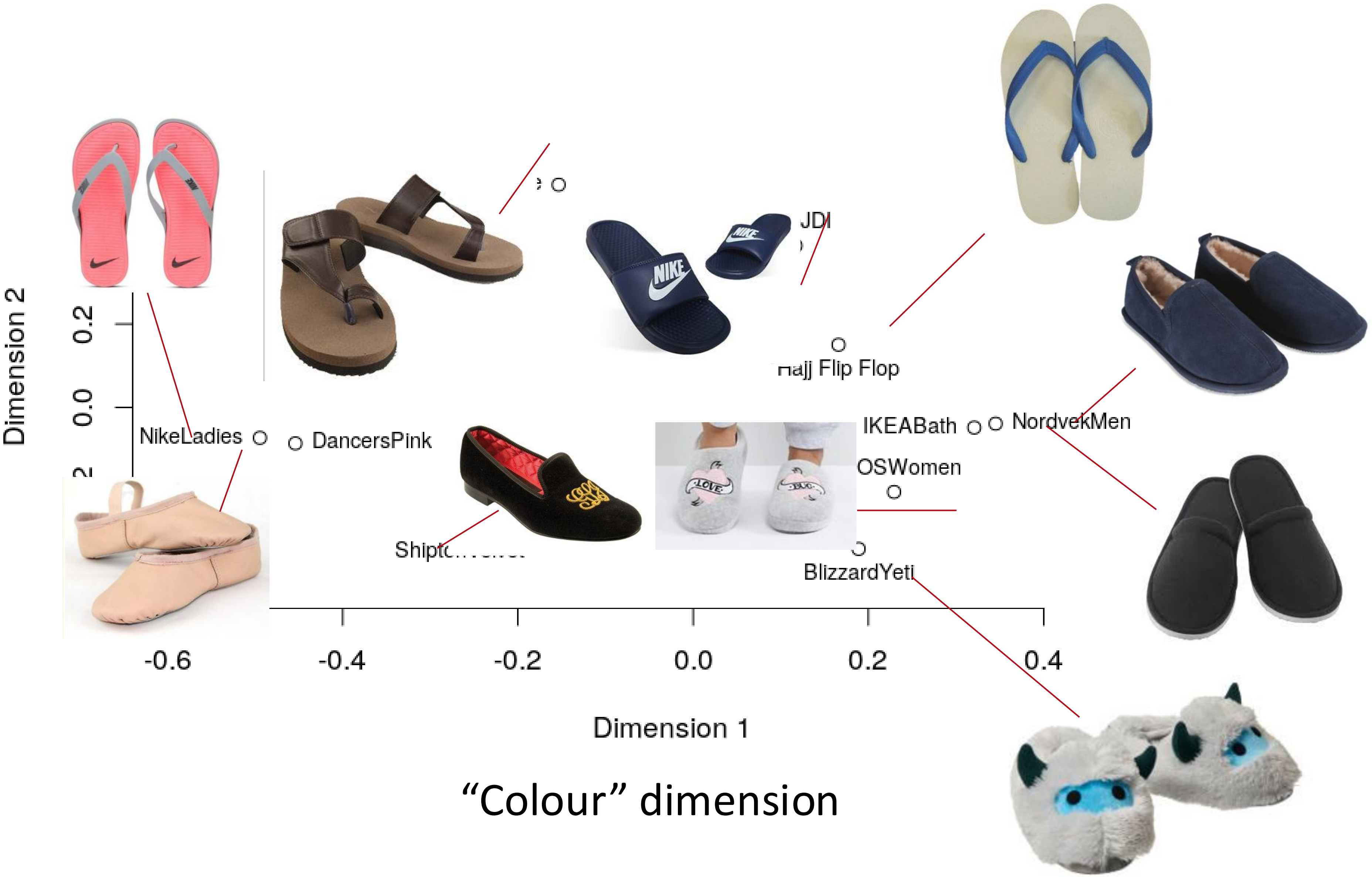
Dimensionality problem

1. Principle Components analysis;
   - Identify components that maximise the class-separation.
2. Multidimensional scaling;
   - Decrease dimensionality based on the proximities among the items.
   - These proximities are used to produce geometric configurations of the objects in lower dimensional spaces.

MDS: slippers example



MDS: slippers example



"Hardness" dimension

"Colour" dimension

Basics of promotional modelling

Promotional modelling is concerned with the impact of promotions on demand.

Promotions include any marketing instrument that can be applied short-term to affect consumer demand.

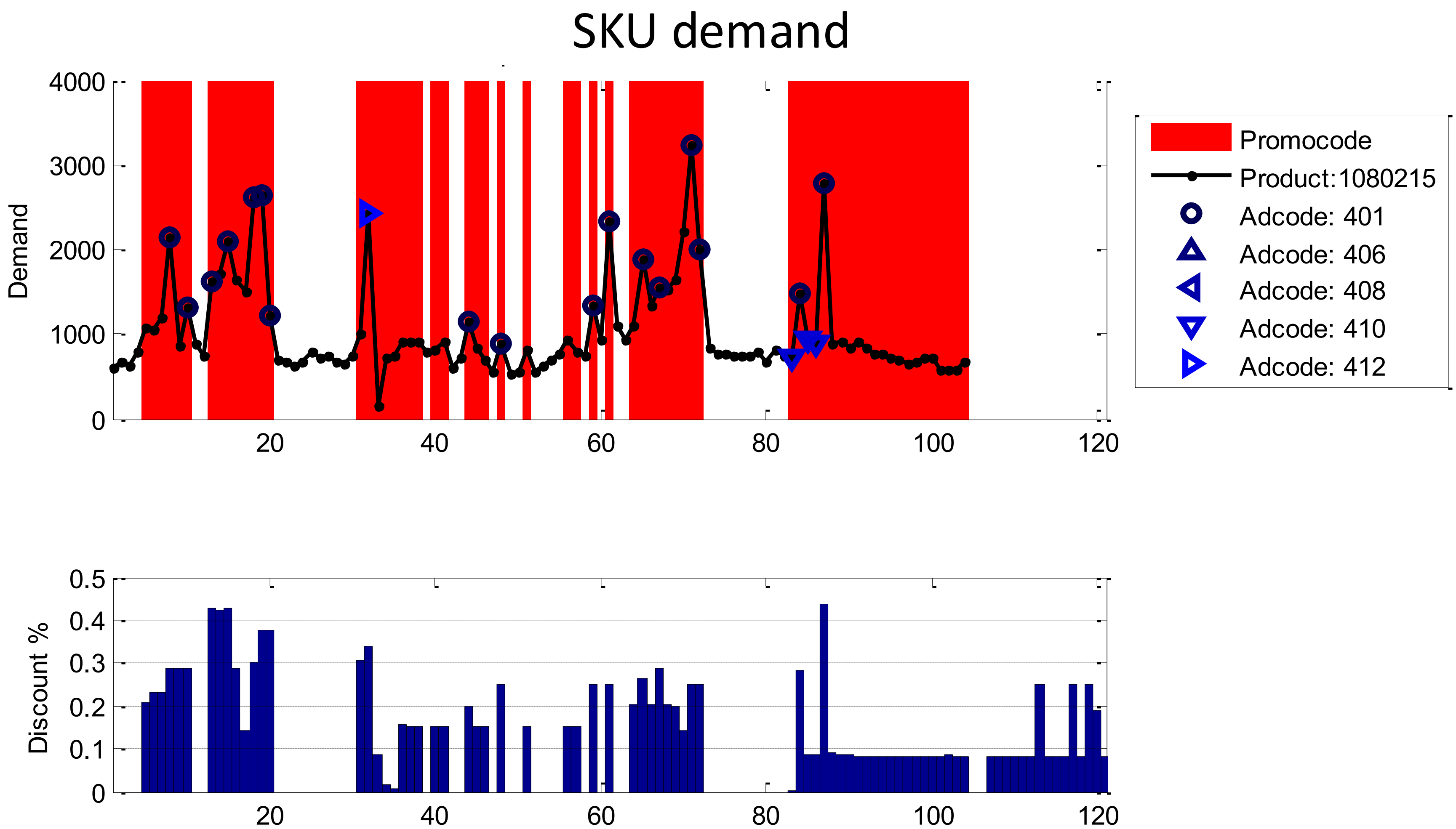
- Why do we need promotions?
- What types of promotions are there?

Basics of promotional modelling

- BOGOF – Buy one get one free;
- Discounts (price reduction);
- x% free;
- Coupons;
- Free samples;
- Buy x for y;
- Buy x get y;
- …

Basics of promotional modelling

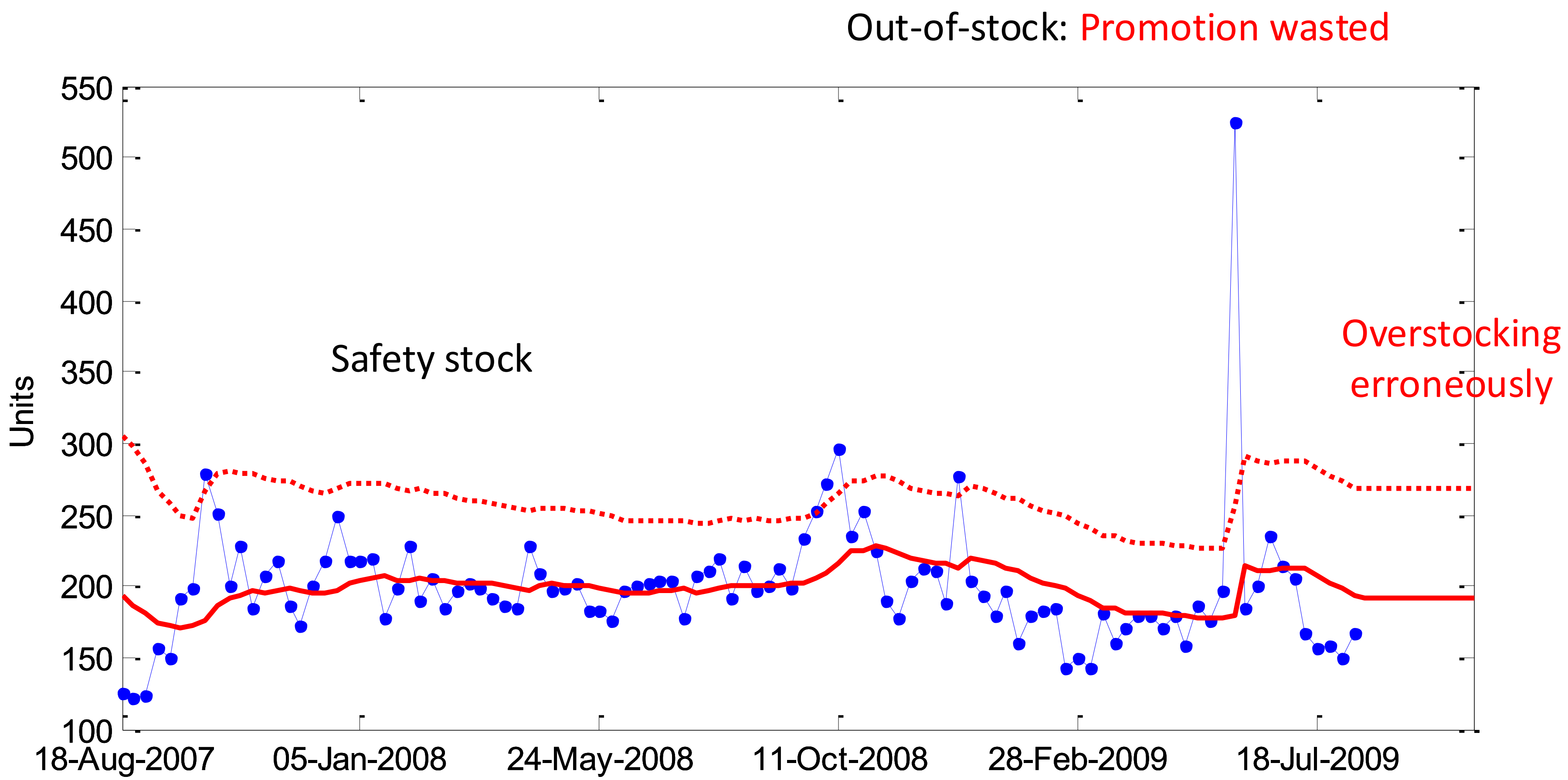
Different promotions, different impact:



Basics of promotional modelling

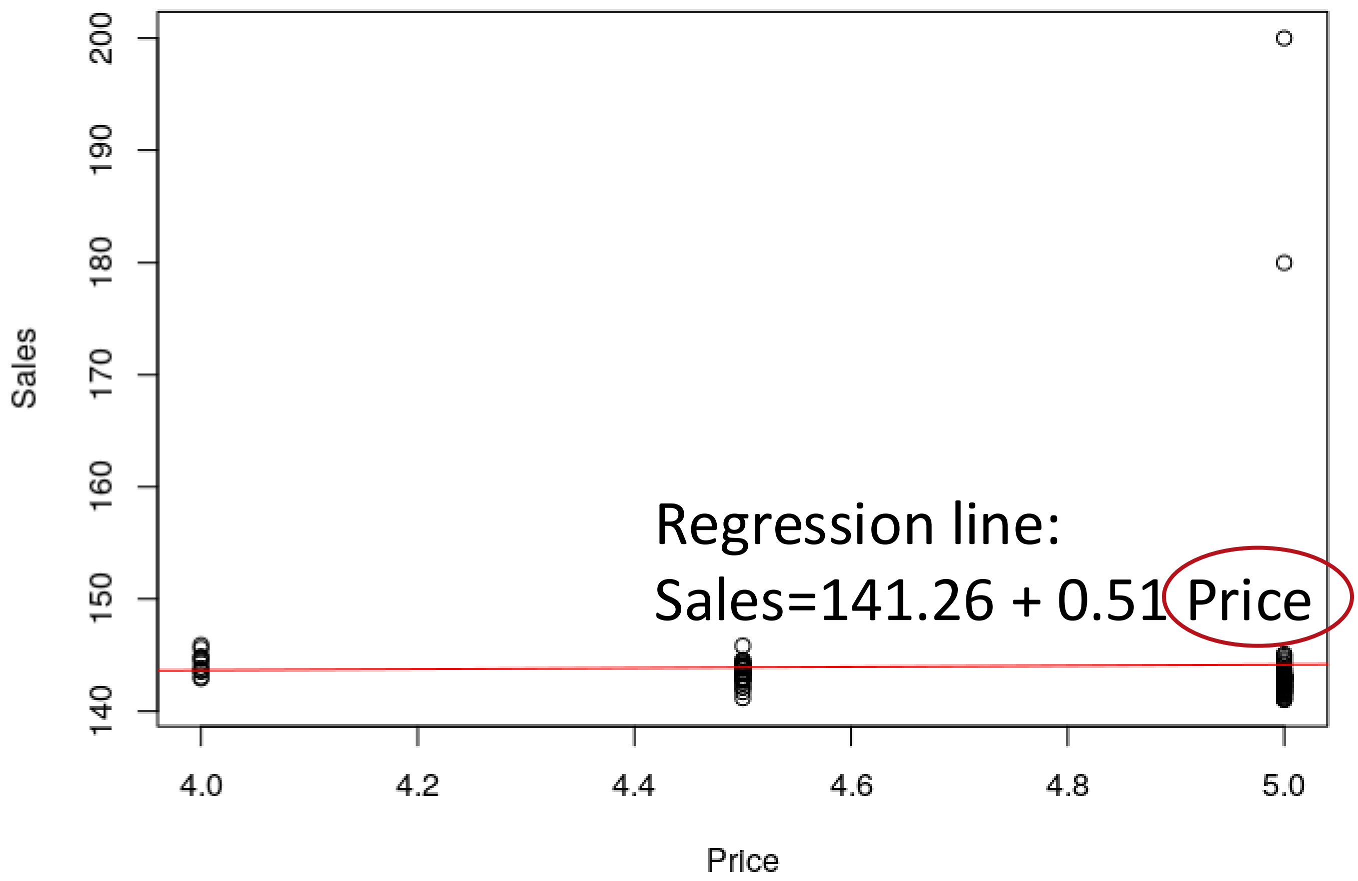
- What if we don’t model promotions correctly?
- What happens, when important variables are omitted?
- What happens when the model is misspecified?

Promotion not taken into account



Also: How do you forecast the impact of the next promotion?

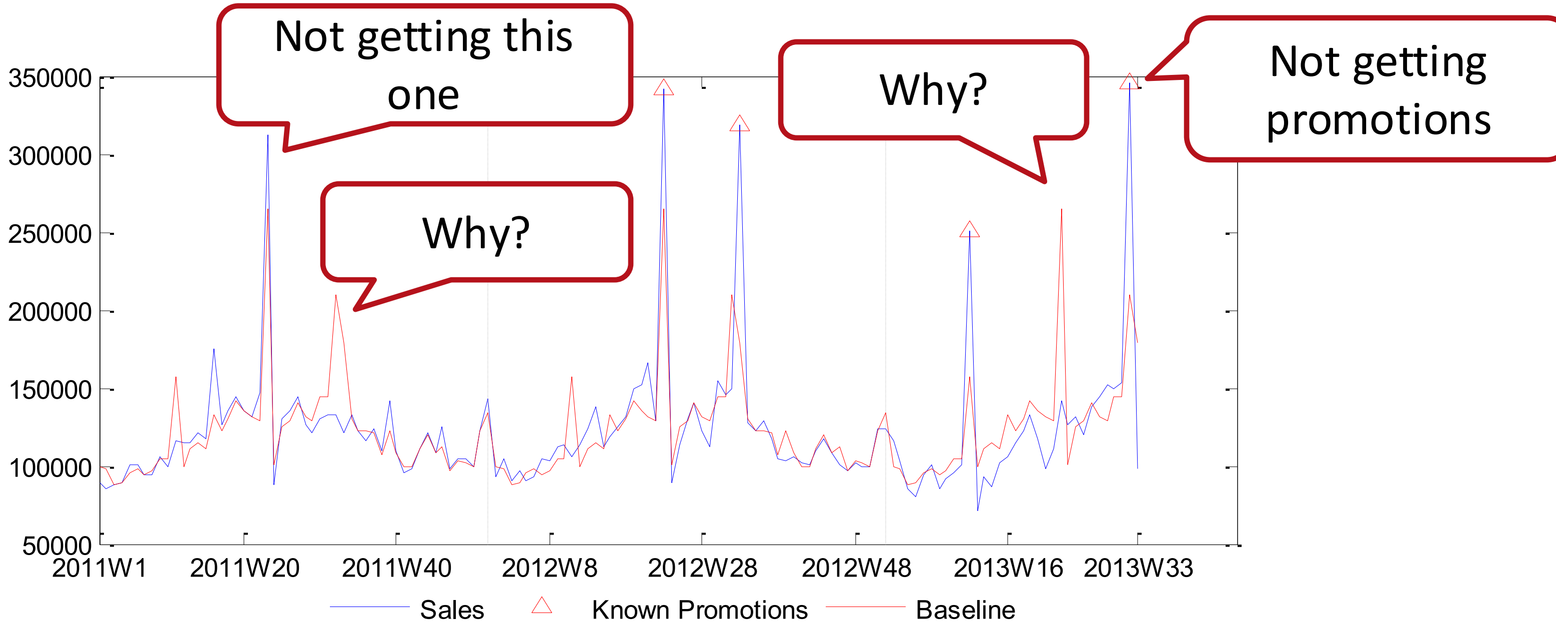
Promotion not taken into account



Basics of promotional modelling

So, how do we take promotions into account?

Regression modelling!



Let’s fit some baseline model to the data

$$\hat{y}_t = a_0 + \sum_{i=1}^{51} a_i S_i$$

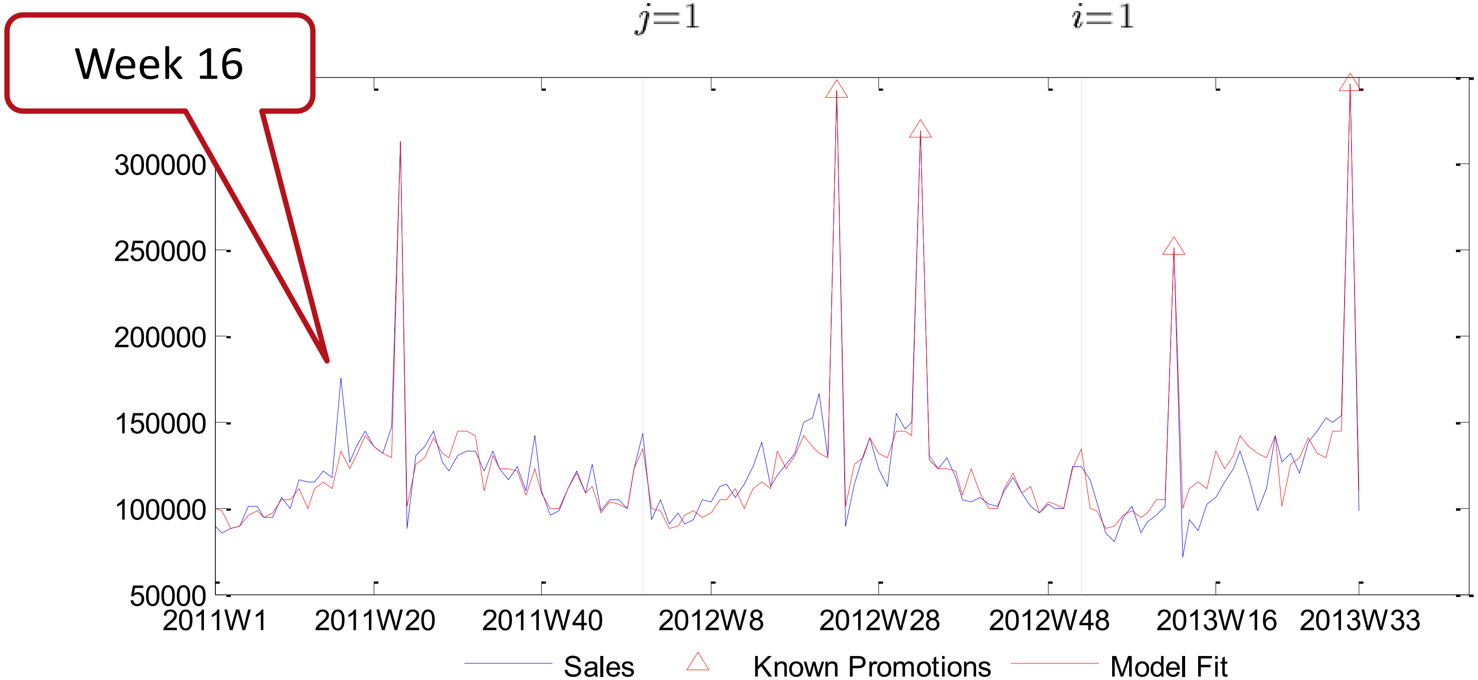
51 seasonal binary dummies

What are the problems here?

Basics of promotional modelling

Instead of a single input for the promotions, each promotion **type** is inputted as a different variable

$$\hat{y}_t = a_0 + a_1 D + \sum_{j=1}^{4} a_{j+1} Promo_j + \sum_{i=1}^{51} a_{i+5} S_i$$



What happened week 16 in 2011?

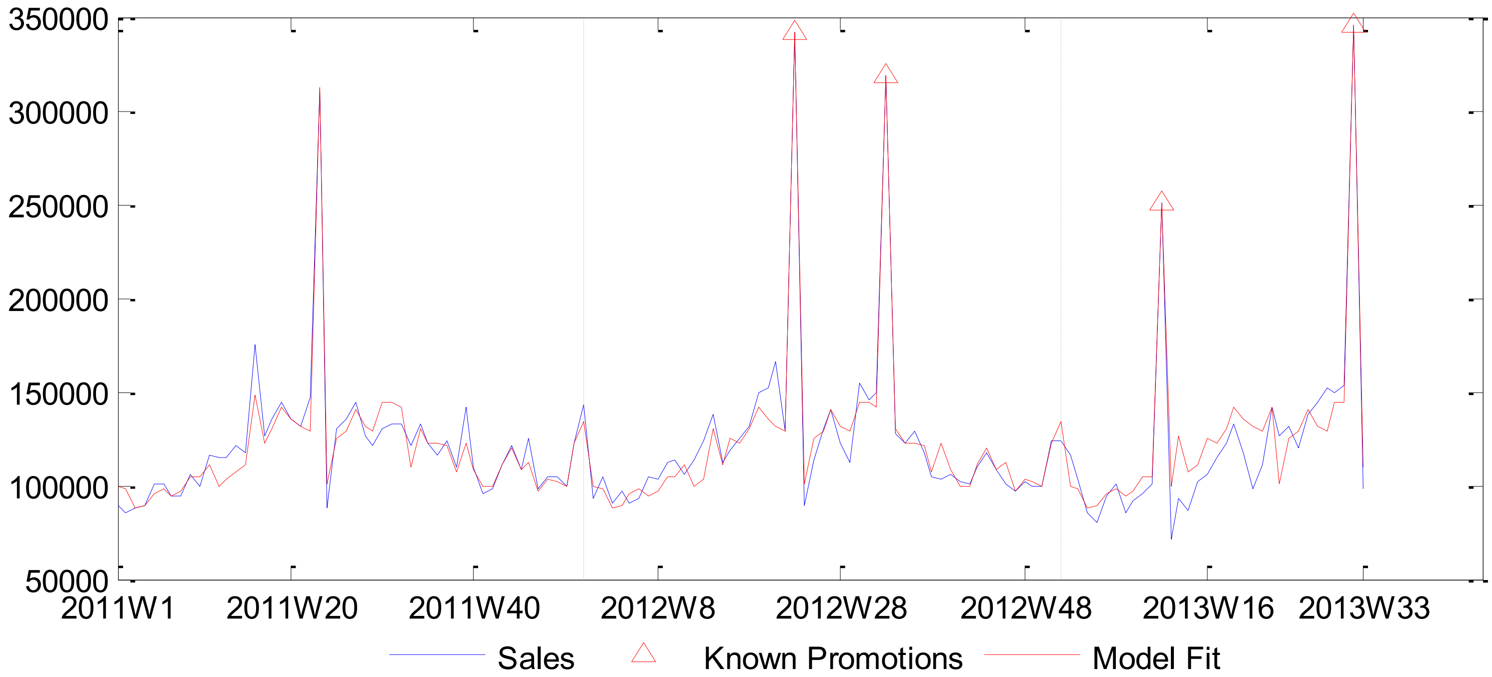
Is there a calendar event that is relevant? (Easter)

Basics of promotional modelling

The Easter dummy = 1 whenever it is Easter

Easter = 0,0,0,…,0,1,0,… …0,1,0,… …0,1,0,0,0,0,0,0…

$$\hat{y}_t = a_0 + a_1 D + a_2 Easter + \sum_{j=1}^{4} a_{j+2} Promo_j + \sum_{i=1}^{51} a_{i+5} S_i$$



Basics of promotional modelling

We can add known promotions in the model as additional variables

$$\hat{y}_t = a_0 + a_1 D + a_2 Easter + \sum_{j=1}^{4} a_{j+2} Promo_j + \sum_{i=1}^{51} a_{i+5} S_i$$

Variable	Coefficient
Constant	+133770.5
Outlier?	+170144.0
Easter	+23412.3
Promo1	+199007.0
Promo2	+209315.0
Promo3	+139347.5
Promo4	+204048.0
…	…

Effect on sales of Promo4

Do not forget to perform the usual residual diagnostics (Is your model valid?)

How can we do that?

Basics of promotional modelling

We need to know future values of variables in order to forecast using regression...

$$\hat{y}_t = a_0 + a_1 D + a_2 Easter + \sum_{j=1}^{4} a_{j+2} Promo_j + \sum_{i=1}^{51} a_{i+5} S_i$$

- D: Unless we know what this is, normally it is one-off event and the D = 0 for any future period.
- Easter: Its date is well known and the dummy will be equal to 1 when it occurs.
- Promo: Promotions are normally planned ahead. Based on these plans we can create the dummies. When a promotion is running its value = 1 and = 0 otherwise.
- S: These are the seasonal dummies, known values. Example: months of the year.

Basics of promotional modelling

$$\hat{y}_t = a_0 + a_1 D + a_2 Easter + \sum_{j=1}^{4} a_{j+2} Promo_j + \sum_{i=1}^{51} a_{i+5} S_i$$

The model assumes deterministic seasonality

An alternative is the model with stochastic seasonality:

- ARIMAX (use Arima() or msarima() or adam()):

$$\hat{y}_t = a_0 + a_1 D + a_2 Easter + \sum_{j=1}^{4} a_{2+j} Promo_j + \Phi_1 y_{t-m}$$

- ETSX or autoregressive distributed lags (ARDL)

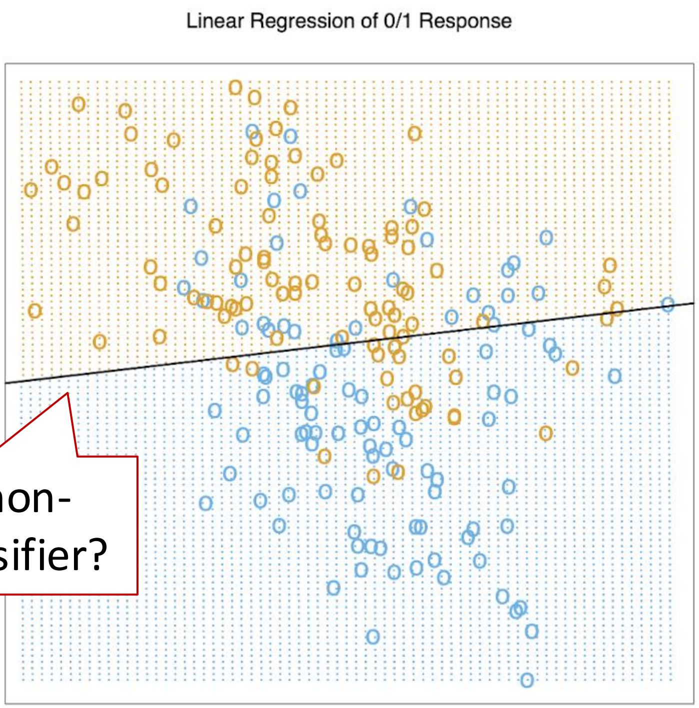
Whatever you use, compare it with regression!

An example of classification problems

- A bank would like to understand which factors can lead a customer to leave the bank
- Retaining its customers is cheaper than acquiring ones
- We can redefine this problem as a classification problem: whether a customer leaves or stays;
  - Binary option → 1 or 0
- Alternative classifiers:
  - k-NN (one of machine learning algorithms)
  - Logistic regressions (well-established econometric model)

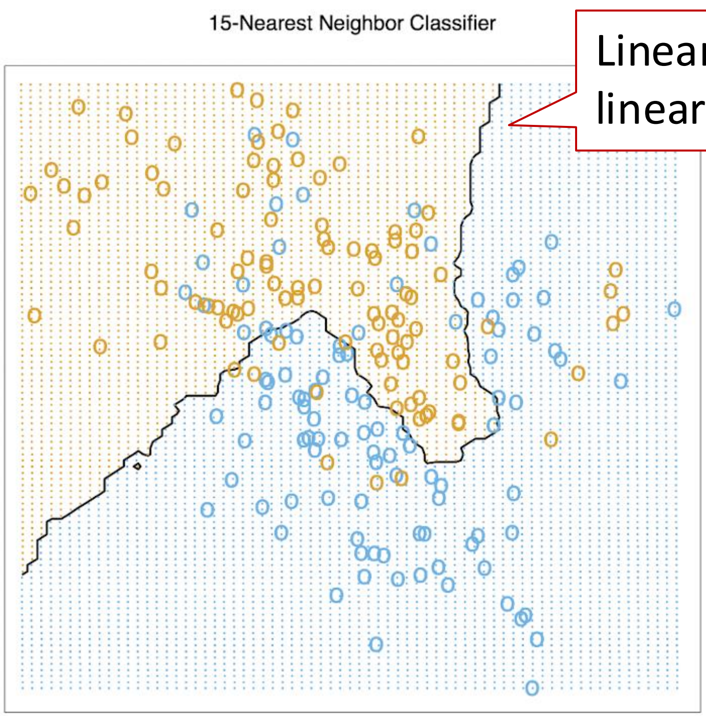
A classification problem

- Classification: Y is categorical
  - exited or stayed
  - 1 or 0
- Objective: Line/ surface of best **discrimination**



Linear or non-linear classifier?

Logistic regression



Linear or non-linear classifier?

k Nearest Neighbours

Conclusions

Conclusions

- Business decisions need insights from the data
  - We need to extract insights from the data using analytical tools

Ways to do analytics

Descriptive & Predictive

- Regression
- Classification
- Clustering
- Association

Prescriptive analytics

- Simulation and stochastic modelling
- Inventory optimisation
- Transportations and logistics

Conclusions

- We exemplify this connection using marketing and inventory decisions
  - Marketing:
    - Multidimensional scaling for product positioning
    - Regression models for promotional modelling
    - Logistic regression for marketing campaign
  - Operations:
    - Newsvendor problem for inventory policies
    - Demand forecasting for estimating the unknown demand

Thank you! Any questions?

Feedback form:
https://forms.gle/xSuzQe4xMhD5oYDA8

Kandrika Pritularga

kandrikapritularga.com

Marketing Analytics and Forecasting | Lancaster University Management School

For more information about our centre
For more information about our next Forecasting course