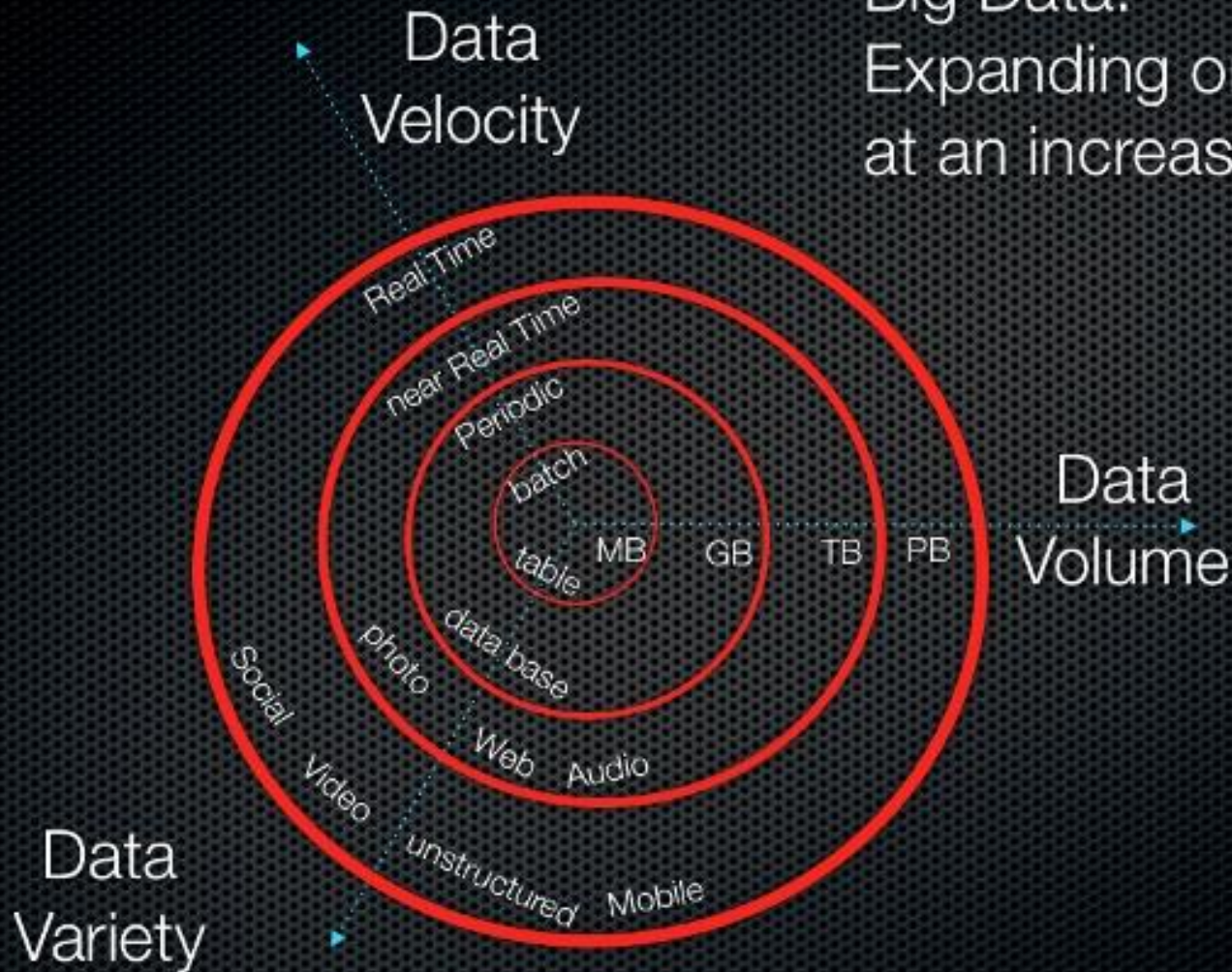


# Introduction to Big Data Science

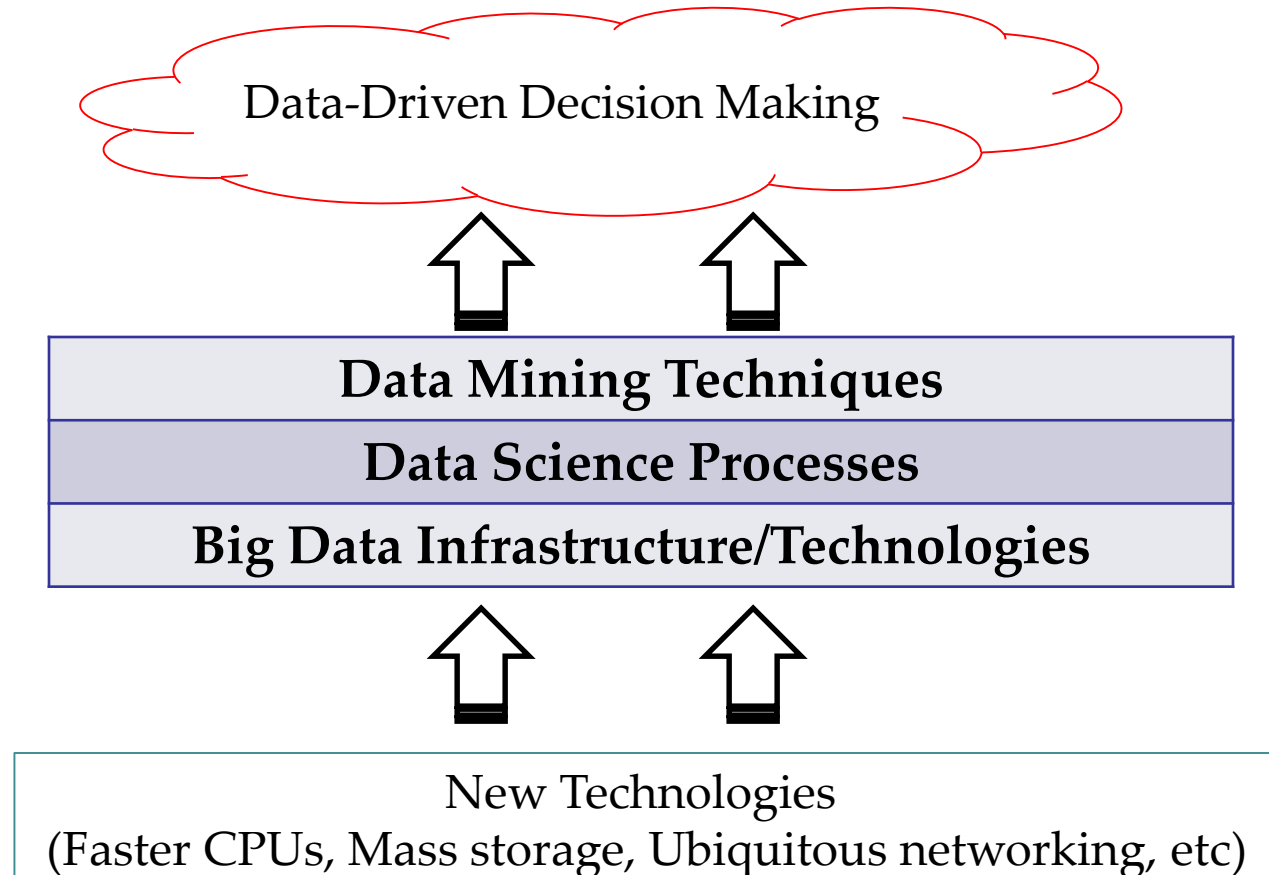
02<sup>nd</sup> Period

Data Science Process

Big Data:  
Expanding on 3 fronts  
at an increasing rate.



# Big Data Science



# Contents

- ◆ CRISP-DM
- ◆ Business Understanding
- ◆ Data Understanding
- ◆ Data Preparation
- ◆ Modeling
- ◆ Evaluation
- ◆ Deployment

## ◆ CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM)

European Community funded effort to develop framework for data mining tasks

Goals:

- ◆ Encourage interoperable tools across entire data mining process
- ◆ Take the mystery/high-priced expertise out of simple data mining tasks

# ◆ CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM)

European Community funded effort to develop framework for data mining tasks

Goals:

- ◆ Encourage interoperable tools across entire data mining process
- ◆ Take the mystery/high-priced expertise out of simple data mining tasks

Troublesome if each phase of a process is done by a different tool...

Data Preparation	Modeling	Evaluation
SAS	SPSS	R

- Encourage a tool capable of the entire process.
- Provide an easy way for data mining.

# Why do we need a standard process?

- ◆ Framework for recording experience
  - Allows projects to be replicated
- ◆ Aid to project planning and management
- ◆ “Comfort factor” for new adopters
  - Demonstrates maturity of Data Mining
  - Reduces dependency on “stars”



(from wikipedia)



**Big Data Science**

After a star left, your team might not work any more...

Cavs lost 63 games, including a 26-game losing streak, which set an NBA record.

## ◆ CRoss Industry Standard Process for Data Mining

Initiative launched Sept.1996

SPSS/ISL (US/UK), NCR (US), Daimler-Benz (Germany), OHRA (Netherland)

Funding from European commission

Over 200 members of the CRISP-DM SIG worldwide

- DM Vendors - SPSS, NCR, IBM, SAS, SGI, Data Distilleries, Syllogic, Magnify, ..
- System Suppliers / consultants - Cap Gemini, ICL Retail, Deloitte & Touche, ...
- End Users - BT, ABB, Lloyds Bank, AirTouch, Experian, ...

Another standard process

## ◆ SEMMA

Sample, Explore, Modify, Model and Assess.

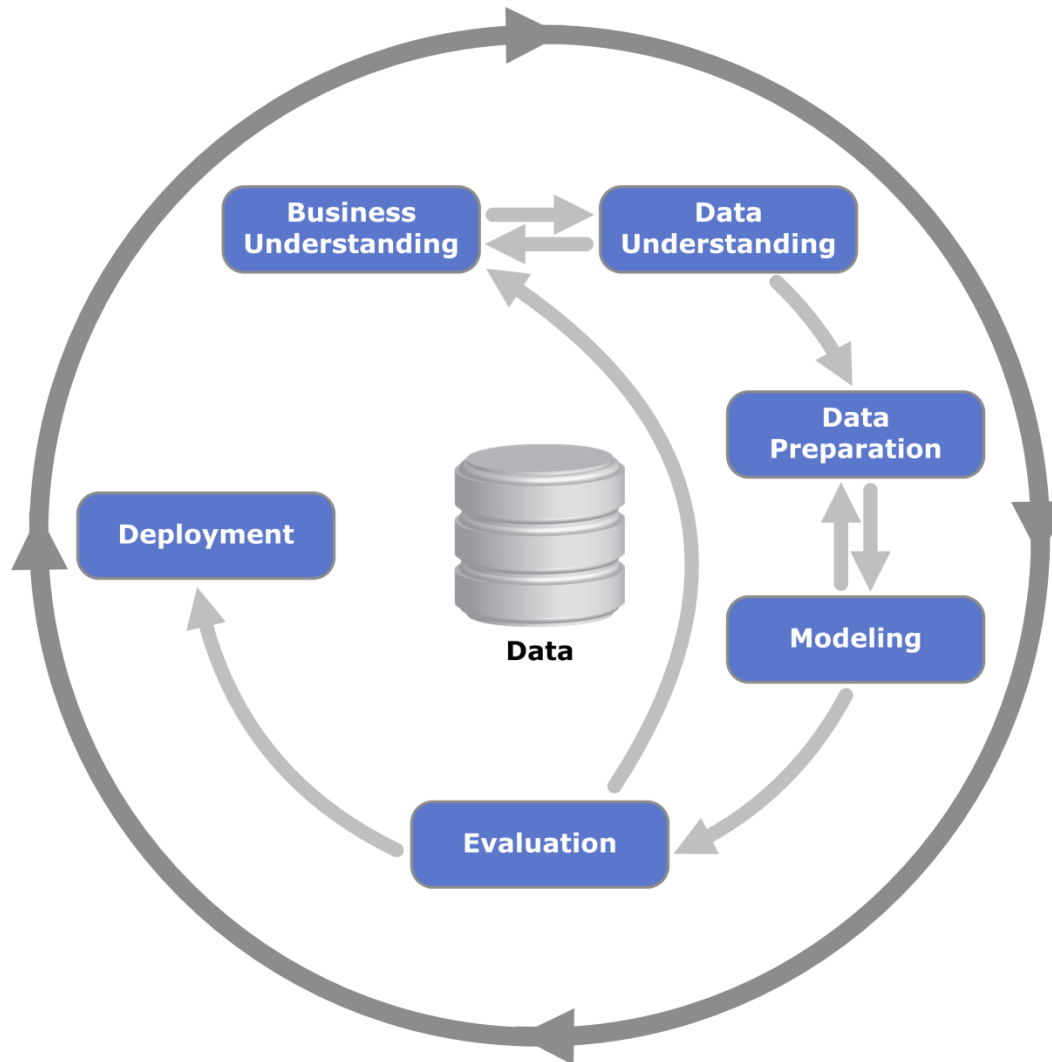
A list of sequential steps developed by SAS Institute Inc.

Focuses on the modeling tasks of data mining projects

(leave the business aspects out)



# CRISP-DM overview



Note: Iteration is the rule rather than the exception.

CRISP-DM figure (wikipedia.org)

# Case study: Canadian travel agency

(“Discovering Knowledge in Data” by D. Larose)

A travel agency, Travel Alberta in Alberta, Canada, wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists.

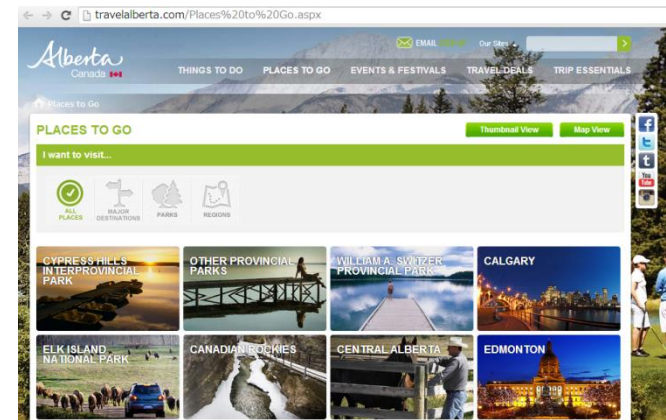
Let's apply CRISP-DM to Travel Alberta!

Alberta is about two times as large as Japan.



There are many activities in Alberta!  
That's why they think they should encourage  
intraprovince travels.

Aurora



Travelalberta.com

Ski/Snowboarding



Theater



River cruise



Natural scenery



## ◆ Business Understanding

A business project does not come  
as a prepackaged data mining problem...

Three things to do:

- State business objective
- State Data Mining objective
- State success criteria

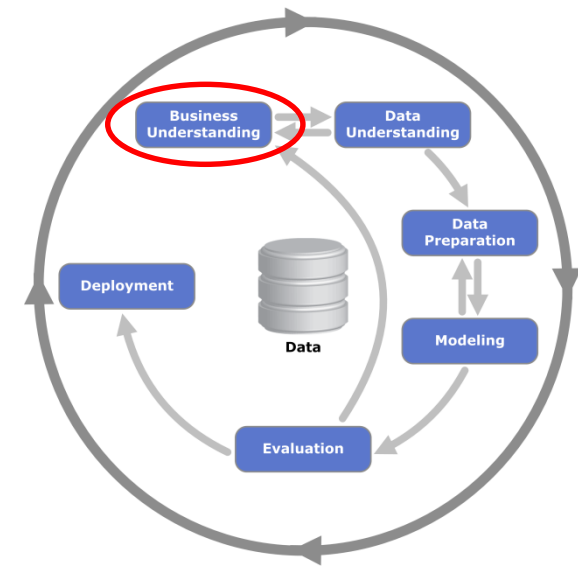
Iteratively recast the problem. (See the figure.)

The initial formulation: often not be optimal.

In cases, we need to go back from further steps.

Data Understanding → Business Understanding

Evaluation → Business Understanding



# Case study: Canadian travel agency

A travel agency, Travel Alberta in Alberta, Canada, wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists.

Travel Alberta sponsored two researchers, Simon Hudson and Brent Ritchie, of the University of Calgary to study intraprovince tourist behavior in Alberta.

## Business Understanding

- **Business goal**  
Develop an intraprovince marketing campaign.
- **Data Mining objectives**  
Toward this goal, the main objectives were
  - to determine which factors were important in choosing destinations in Alberta,
  - to evaluate the domestic perceptions of the “Alberta vacation product,” and
  - to attempt to comprehend the travel decision-making process.
- **Success criterion**  
We can form a quantitative basis for the development of the marketing campaign.

To perform this, the researchers decided to create profiles of domestic Albertan tourists based on the decision behavior of the tourists.

## ◆ Data Understanding

Data: available raw material

We need to build a solution from data

↓  
Your business problem

Data may not be exact match with the problem.

Because...

- Collected for different purposes
- No explicit purpose

Estimating costs and benefits

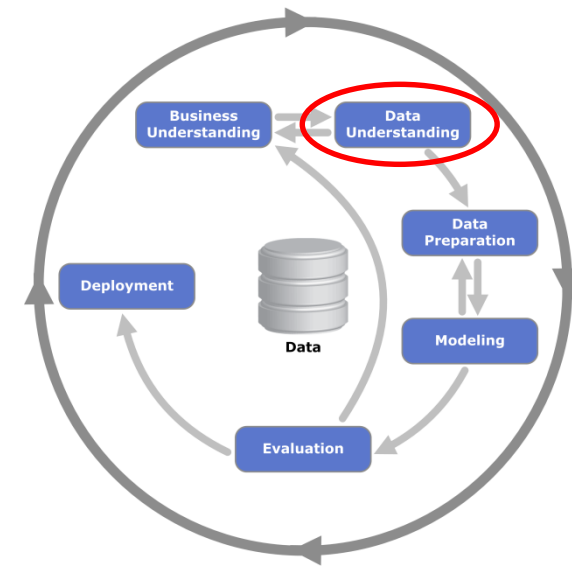
- Is data available free?

If not, estimate the cost to obtain.

- Is cleaning data costly?

E.g. noisy customer records.

Match customer records to ensure only one record per customer.



## Case study: Canadian travel agency

Travel Alberta wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists.

Travel Alberta sponsored two researchers, Simon Hudson and Brent Ritchie, of the University of Calgary to study intraprovince tourist behavior in Alberta.

### Data Understanding

The data were collected in late 1999 using a phone survey of 13,445 Albertans. The respondents were screened according to those who were over 18 and had traveled for leisure at least 80 kilometers for at least one night within Alberta in the past year. Only 3071 of these 13,445 completed the survey and were eligible for inclusion in the study.



## An example of data statistics

Note that this was obtained in a different year.

<http://tpr.alberta.ca/>

Table 1: Fielding Statistics

Status	TOTAL	Calgary & Area	Edmonton & Area	Alberta North	Alberta Central	Alberta South	Canadian Rockies
Completes	1404 18.7%	350 16.7%	350 18.9	202 19.5%	200 19.1%	200 18.8%	102 24.8%
Dead Numbers	6101	1749	1500	835	846	862	309
Language/Hearing Problems	195 3.2%	71 4.1%	75 5.0%	12 1.4%	8 0.9%	10 1.2%	19 6.1%
Respondent away for duration of study	5 0.8%	8 0.5%	21 1.4%	11 1.3%	5 0.6%	6 0.7%	0
Business/Fax Line	1112 18.2%	276 15.8%	299 19.9%	202 24.2%	138 16.3%	165 19.1%	32 10.4%
NIS	1036 17.0%	261 14.9%	179 11.9%	228 27.3%	154 18.2%	184 21.3%	30 9.7%
Survey Terminated	108 1.8%	25 1.4%	22 1.5%	18 2.2%	19 2.2%	15 1.7%	9 2.9%
Refused	571 9.4%	215 12.3%	212 14.1%	28 3.4%	54 6.4%	42 4.9%	20 6.5%
Ineligible –no leisure travel of 80km from home in past 12 months	2905 47.6%	851 48.7%	664 44.3%	328 39.3%	447 52.8%	425 49.3%	190 61.5%
Ineligible – Other	79 1.3%	30 1.7%	21 1.4%	2 0.2%	4 0.5%	14 1.6%	8 2.6%
Duplicate	44 0.7%	12 0.7%	7 0.5%	6 0.7%	17 2.0%	1 0.1%	1 0.3%
TOTAL SAMPLE USED	7,505	2,099	1,850	1,037	1,046	1,062	411
Eligible Exhausted (dialled 5 times)	849	221	223	103	109	114	79

NIS – Not In Service



A sample question: what activities would you likely participate in?  
 Responding from people in Canadian Rockies

Table 37

1pt.

2pt.

3pt.

4pt.

<b>Q.10 Activities Canadian Rockies Respondents (n=20)</b>	<b>Likelihood of Participating in...</b>					<b>Mean</b>
	<b>Very Unlikely</b>	<b>Somewhat Unlikely</b>	<b>Somewhat Likely</b>	<b>Very Likely</b>	<b>Don't Know</b>	
Visit Friends & Relatives	5.9%	6.9%	39.2%	48.0%	1.0%	3.3
Outdoor Recreation	20.6%	5.9%	27.5%	46.1%	0	3.0
Outdoor Summer Sports	20.6%	5.9%	33.3%	40.2%	0	2.9
Indoor Leisure Activities	21.6%	9.8%	37.3%	30.4%	1.0%	2.8
Outdoor Leisure	34.3%	12.7%	38.8%	32.9%	0	2.4
Festivals & Cultural Activities	35.3%	10.8%	31.4%	21.6%	1.0%	2.4
Museums & Historical Sites	41.2%	9.8%	29.4%	18.6%	1.0%	2.3
Outdoor Winter Sports	44.1%	11.8%	13.7%	29.4%	1.0%	2.3
Attractions	37.3%	15.7%	34.3%	12.7%	0	2.2
Calgary Stampede/ Klondike Days	48.0%	12.7%	24.5%	14.7%	0	2.1
Outdoor Adventure	45.1%	11.8%	28.4%	13.7%	1.0%	2.1
Team Sports	55.9%	12.7%	18.6%	12.7%	0	1.9
Spectator Sports & Entertainment	53.9%	14.7%	19.6%	11.8%	0	1.9
Aboriginal Attractions	54.9%	20.6%	19.6%	4.9%	0	1.8
Organized Sports Events	55.9%	20.6%	14.7%	6.9%	2.0%	1.7
Spas & Other Health Club	65.7%	9.8%	16.7%	7.8%	0	1.7

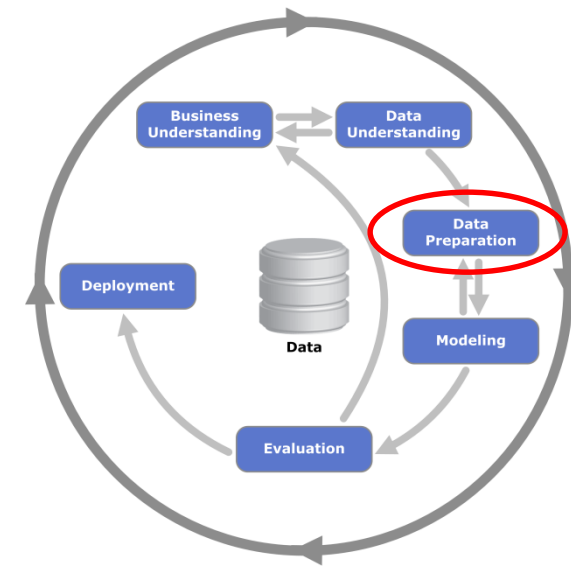
## ◆ Data Preparation

Analytic technologies are powerful..

But they impose certain requirements on data.  
e.g. Data should be in a particular form.

### Examples

- Convert data to a tabular format.
- Remove/infer missing values.
- Handle outliers.
- Convert data to different types.



## Example of data preparation

### Missing values

It often happens that values in some fields are missing.

1. Omit records with missing values.
2. Replace the missing value with some constant, chosen by analyst.
3. Replace the missing value with the field mean/mode.  
A standard choice:
  - Use mean for numerical values.
  - Use mode for categorical values.
4. Replace the missing value with a value generated at random from the observed variable distribution.

## Example of data preparation

### Outliers

Extreme values that lie near the limits of the data or go against the trend.  
Identifying them is important since they might be errors in data entry.

### Examples of identification:

1. Use a histogram of the variable. (graphical identification)
2. Use Z-Score standardization. (numerical identification)

How far an observation is from the field mean value.

$Z\text{-Score of } X = (X - \text{mean}(X)) / \text{SD}(X)$

SD(X): standard deviation of the field values.

e.g. view values as potential outliers if the absolute value of Z-Score exceeds 3.

Note: This course will not mainly focus on data preparation.

If interested, see ``Data Preparation for Data Mining'' by Pyle (1999).

## Case study: Canadian travel agency

Travel Alberta wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists.

Travel Alberta sponsored two researchers, Simon Hudson and Brent Ritchie, of the University of Calgary to study intraprovince tourist behavior in Alberta.

### Data Preparation

One of the survey questions asked the respondents to indicate to what extent each of the factors from a list of 13 factors most influence their travel decisions. These were then considered to be variables upon which the cluster analysis was performed, and included such factors as **the quality of accommodations, school holidays, and weather conditions.**

An example of data statistics

Note that this was obtained in a different year.

<http://tpr.alberta.ca/>

Table 1: Fielding Statistics

Status	TOTAL	Calgary & Area	Edmonton & Area	Alberta North	Alberta Central	Alberta South	Canadian Rockies
Completes	1404 18.7%	350 16.7%	350 18.9%	202 19.5%	200 19.1%	200 18.8%	102 24.8%
Dead Numbers	6101	1749	1500	835	846	862	309
Language/Hearing Problems	195 3.2%	71 4.1%	75 5.0%	12 1.4%	8 0.9%	10 1.2%	19 6.1%
Respondent away for duration of study	5 0.8%	8 0.5%	21 1.4%	11 1.3%	5 0.6%	6 0.7%	0
Business/Fax Line	1112 18.2%	276 15.8%	299 19.9%	202 24.2%	138 16.3%	165 19.1%	32 10.4%
NIS	1036 17.0%	261 14.9%	179 11.9%	228 27.3%	154 18.2%	184 21.3%	30 9.7%
Survey Terminated	108 1.8%	25 1.4%	22 1.5%	18 2.2%	19 2.2%	15 1.7%	9 2.9%
Refused	571 9.4%	215 12.3%	212 14.1%	28 3.4%	54 6.4%	42 4.9%	20 6.5%
Ineligible –no leisure travel of 80km from home in past 12 months	2905 47.6%	851 48.7%	664 44.3%	328 39.3%	447 52.8%	425 49.3%	190 61.5%
Ineligible – Other	79 1.3%	30 1.7%	21 1.4%	2 0.2%	4 0.5%	14 1.6%	8 2.6%
Duplicate	44 0.7%	12 0.7%	7 0.5%	6 0.7%	17 2.0%	1 0.1%	1 0.3%
TOTAL SAMPLE USED	7,505	2,099	1,850	1,037	1,046	1,062	411
Eligible Exhausted (dialled 5 times)	849	221	223	103	109	114	79

We use records of completed surveys

We omit These records. (missing values)

NIS – Not In Service

A sample question: what activities would you likely participate in?  
 Responding from people in Canadian Rockies

Table 37

1pt.

2pt.

3pt.

4pt.

Q.10 Activities Canadian Rockies Respondents (n=20)	Likelihood of Participating in...					Mean
	Very Unlikely	Somewhat Unlikely	Somewhat Likely	Very Likely	Don't Know	
Visit Friends & Relatives	5.9%	6.9%	39.2%	48.0%	1.0%	3.3
Outdoor Recreation	20.6%	5.9%	27.5%	46.1%	0	3.0
Outdoor Summer Sports	20.6%	5.9%	33.3%	40.2%	0	2.9
Indoor Leisure Activities	21.6%	9.8%	37.3%	30.4%	1.0%	2.8
Outdoor Leisure	34.3%	12.7%	38.8%	32.9%	0	2.4
Festivals & Cultural Activities	35.3%	10.8%	31.4%	21.6%	1.0%	2.4
Museums & Historical Sites	41.2%	9.8%	29.4%	18.6%	1.0%	2.3
Outdoor Winter Sports	44.1%	11.8%	13.7%	29.4%	1.0%	2.3
Attractions	37.3%	15.7%	34.3%	12.7%	0	2.2
Calgary Stampede/ Klondike Days	48.0%	12.7%	24.5%	14.7%	0	2.1
Outdoor Adventure	45.1%	11.8%	28.4%	13.7%	1.0%	2.1
Team Sports	55.9%	12.7%	18.6%	12.7%	0	1.9
Spectator Sports & Entertainment	53.9%	14.7%	19.6%	11.8%	0	1.9
Aboriginal Attractions	54.9%	20.6%	19.6%	4.9%	0	1.8
Organized Sports Events	55.9%	20.6%	14.7%	6.9%	2.0%	1.7
Spas & Other Health Club	65.7%	9.8%	16.7%	7.8%	0	1.7

More important





A sample question: what activities would you likely participate in?  
 Responding from people in Canadian Rockies

Table 37

1pt.

2pt.

3pt.

4pt.

Q.10 Activities Canadian Rockies Respondents (n=20)	Likelihood of Participating in...					Mean
	Very Unlikely	Somewhat Unlikely	Somewhat Likely	Very Likely	Don't Know	
Visit Friends & Relatives	5.9%	6.9%	39.2%	48.0%	1.0%	3.3
Outdoor Recreation	20.6%	5.9%	27.5%	46.1%	0	3.0
Outdoor Summer Sports	20.6%	5.9%	33.3%	40.2%	0	2.9
Indoor Leisure Activities	21.6%	9.8%	37.3%	30.4%	1.0%	2.8
Outdoor Leisure	34.3%	12.7%	38.8%	32.9%	0	2.4
Festivals & Cultural Activities	35.3%	10.8%	31.4%	21.6%	1.0%	2.4
Museums & Historical Sites	41.2%	9.8%	29.4%	18.6%	1.0%	2.3
Outdoor Winter Sports	44.1%	11.8%	13.7%	29.4%	1.0%	2.3
Attractions	37.3%	15.7%	34.3%	12.7%	0	2.2
Calgary Stampede/ Klondike Days	48.0%	12.7%	24.5%	14.7%	0	2.1
Outdoor Adventure	45.1%	11.8%	28.4%	13.7%	1.0%	2.1
Team Sports	55.9%	12.7%	18.6%	12.7%	0	1.9
Spectator Sports & Entertainment	53.9%	14.7%	19.6%	11.8%	0	1.9
Aboriginal Attractions	54.9%	20.6%	19.6%	4.9%	0	1.8
Organized Sports Events	55.9%	20.6%	14.7%	6.9%	2.0%	1.7
Spas & Other Health Club	65.7%	9.8%	16.7%	7.8%	0	1.7

Choose important factors as  
determinants of destinations

More important



A total of 1,404 surveys were conducted with Alberta residents to gather information regarding their intentions to travel for leisure purposes, with a primary focus on travel within Alberta. The final results of the study were weighted to better reflect the actual geographical distribution of the population of the province of Alberta (source: [www.Albertafirst.com](http://www.Albertafirst.com)). The distribution of these surveys and their subsequent weighted totals are displayed in Table 3 below.

Table 3: Sample Distribution

<b>Tourism Destination Region</b>	<b>Actual Number of Surveys Completed</b>	<b>Weighted Totals</b>
Calgary & Area	350	482
Edmonton & Area	350	469
Alberta North	202	58
Alberta Central	200	232
Alberta South	200	144
Canadian Rockies	102	20
<b>TOTAL</b>	<b>1404</b>	<b>1405</b>

## ◆ Modeling

The primary stage:

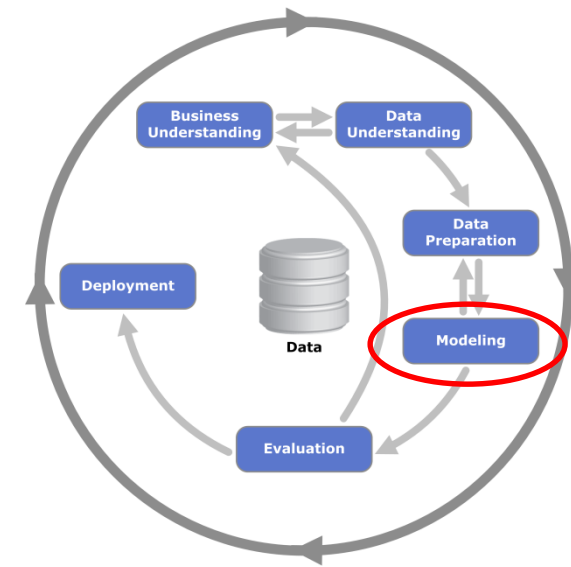
Data mining techniques are applied.

Output of 'Modeling':

Model or pattern capturing regularities.

Data mining techniques

- Summary Statistics (Lec. #7)
- Regression (Lec. #8)
- Classification (Lec. #10)
- Clustering (Lec. #11)
- Association (Lec. #11)



## ◆ Modeling

The primary stage:

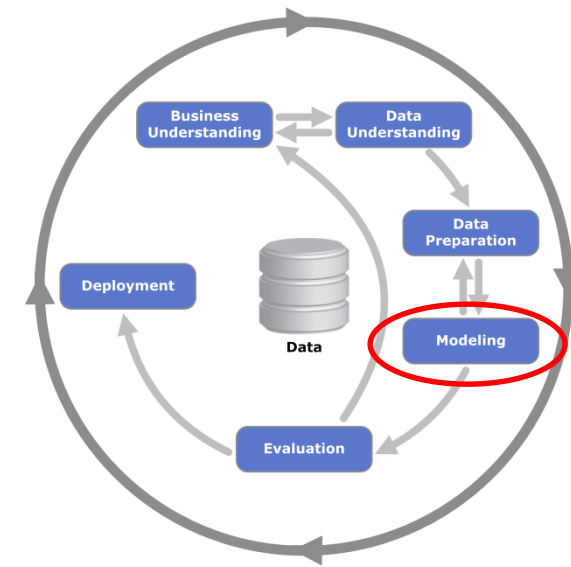
Data mining techniques are applied.

Output of 'Modeling':

Model or pattern capturing regularities.

Data mining techniques

- Summary Statistics (Lec. #7)
- **Regression** (Lec. #8)
- Classification (Lec. #10)
- **Clustering** (Lec. #11)
- Association (Lec. #11)



# 回帰分析： 原因と結果の結びつきの強さを推定

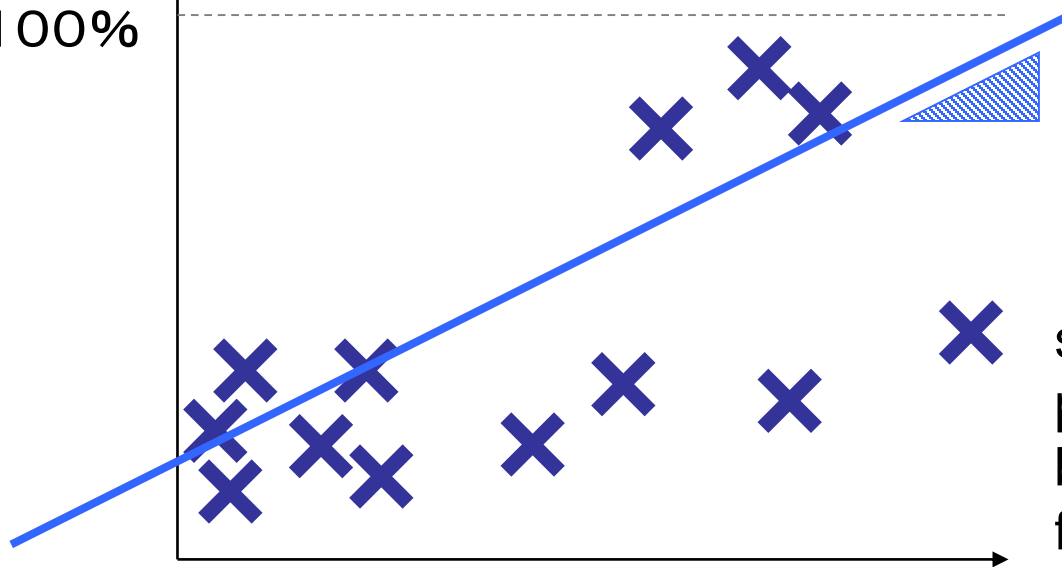


**Regression** analysis: estimates "strengths"  
between cause and result

キップをきられる確率  
probability to get tickets  
for overspeeding

**結果 result**

100%



0%

スピード speed

**原因 cause**

20km

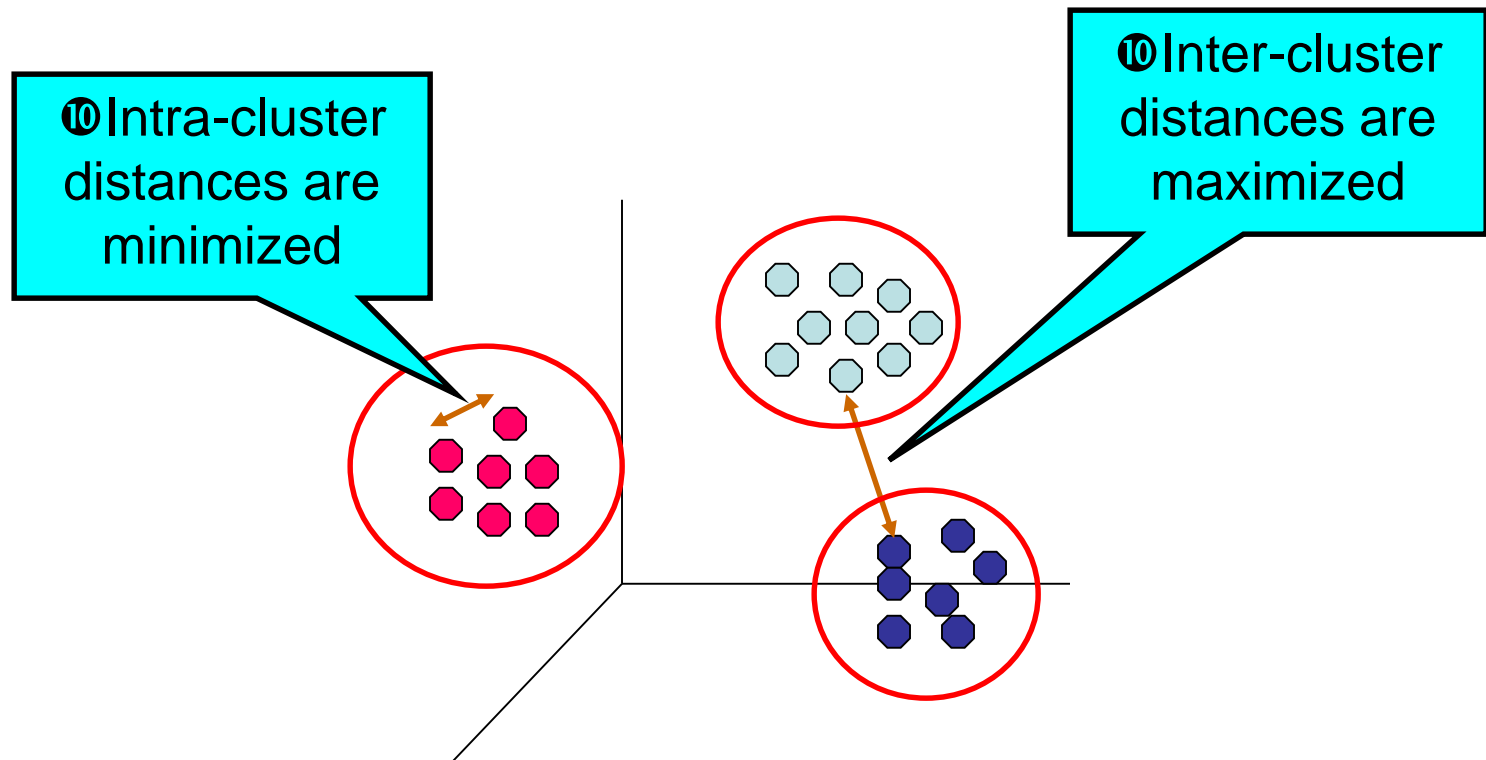
40km

80km

100km

# What is Cluster Analysis?

- ◆ Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



## Case study: Canadian travel agency

Travel Alberta wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists. Two researchers studied intraprovince tourist behavior in Alberta and found that key factors: the quality of accommodations, school holidays, and weather conditions, for example.

### Modeling

Clustering is a natural method for generating segment profiles. The researchers chose k-means clustering, since that algorithm is quick and efficient as long as you know the number of clusters you expect to find. They explored between two and six cluster models before settling on a five-cluster solution as best reflecting reality. Brief profiles of the clusters are as follows:

Cluster 1: the young urban outdoor market. Youngest of all clusters, equally balanced genderwise, with school schedules and budgets looming large in their travel decisions.

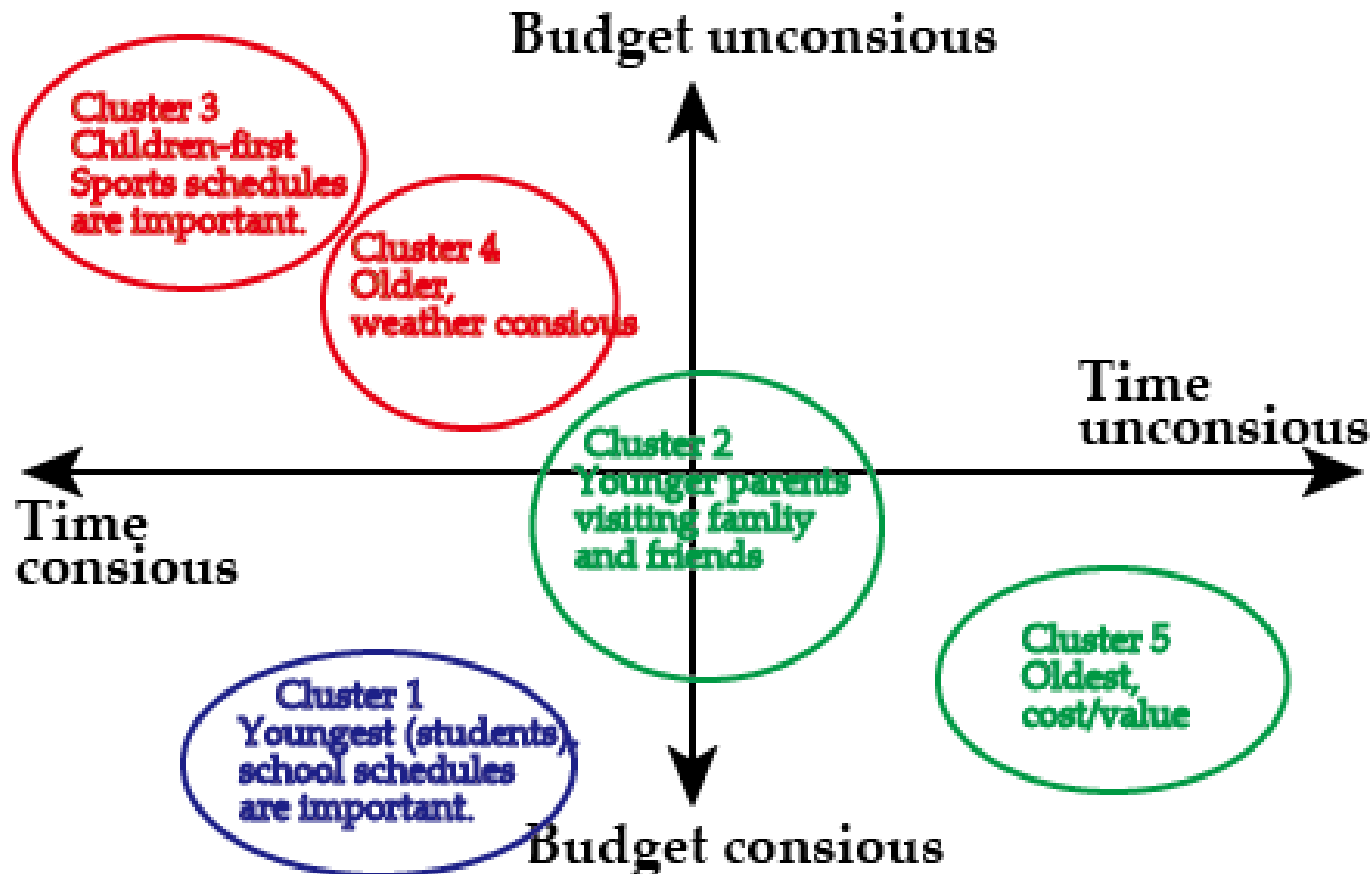
Cluster 2: the indoor leisure traveler market. Next youngest and very female, mostly married with children, with visiting family and friends a major factor in travel plans.

Cluster 3: the children-first market. More married and more children than any other cluster, with children's sports and competition schedules having great weight in deciding where to travel in Alberta.

Cluster 4: the fair-weather-friends market. Second-oldest, slightly more male group, with weather conditions influencing travel decisions.

Cluster 5: the older, cost-conscious traveler market. The oldest of the clusters, most influenced by cost/value considerations and a secure environment when making Alberta travel decisions.

- Cluster 1: Youngest with school schedules and budgets looming large in their travel decisions.
- Cluster 2: Next youngest with visiting family and friends a major factor in travel plans.
- Cluster 3: the children-first market. Children's sports and competition schedules are important.
- Cluster 4: Second-oldest, with weather conditions influencing travel decisions.
- Cluster 5: The oldest of the clusters, most influenced by cost/value considerations.





## ◆ Evaluation

Assess the data mining results rigorously.

Gain confident that the results are reliable.

Don't: deploy results immediately after data mining.

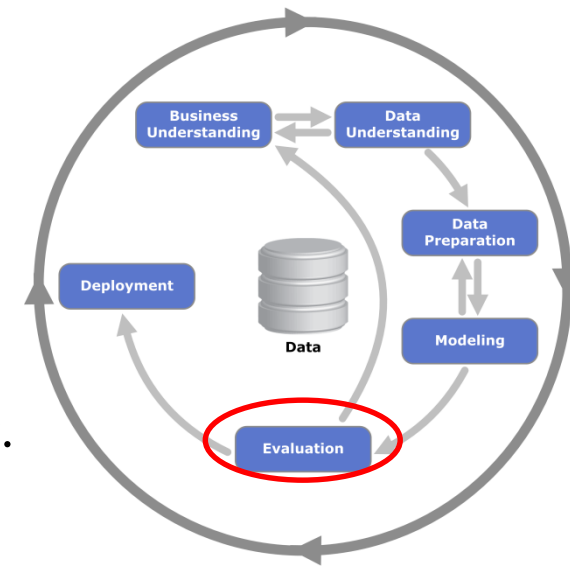
Do: test a model first.

Why?

- Ensure that patterns extracted from the data are true regularities.  
e.g. apply the model to some while keeping others unaffected.
- Ensure that the model satisfies the business goal.

Make sure there is no flaw in the actual business context.

e.g. In a fraud detection problem, a model which is accurate (>99%) in the lab may produce too many false alarms.



## Case study: Canadian travel agency

Travel Alberta wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists. Two researcher studied intraprovince tourist behavior in Alberta and found that key factors: the quality of accommodations, school holidays, and weather conditions, for example.

### Evaluation

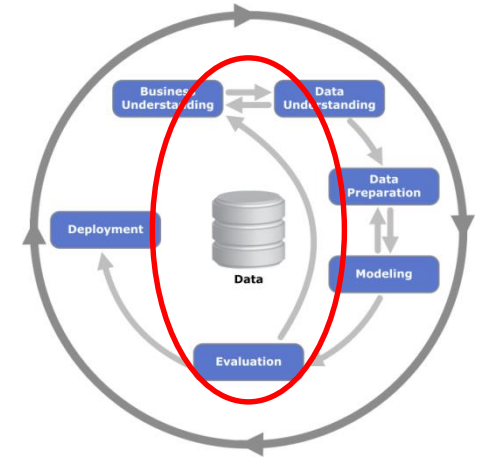
Discriminant analysis was used to verify the “reality” of the cluster categorizations, correctly classifying about 93% of subjects into the right clusters. The discriminant analysis also showed that the differences between clusters were statistically significant.

From Lecture #1:

Example (Grouping customers):

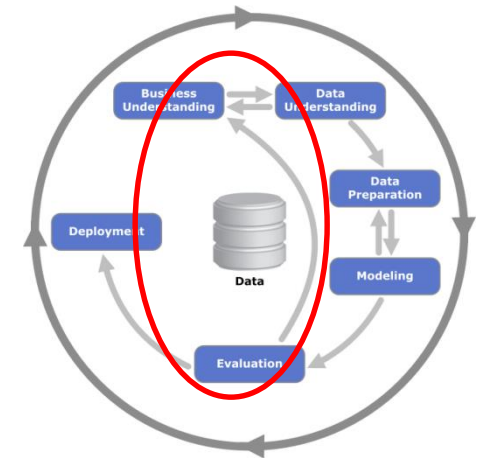
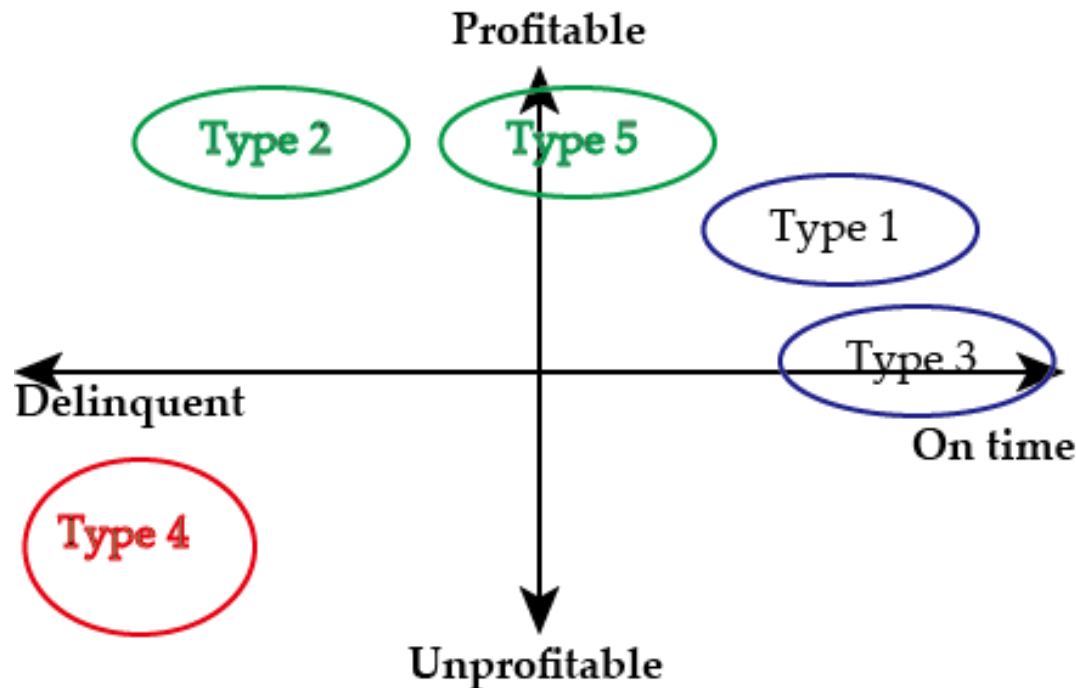
Ira Haimowitz and Henry Schwartz (1997) applied Data Science approach to a business problem of setting credit lines for new credit customers.

Data: GE Capital customers' use of their cards, payment of their bills, and profitability to the company.



GE Capital is a financial service unit of General Electric.  
GE Capital was a former owner of Lake. (It already sold Lake.)

Haimowitz and Schwartz clustered those GE Capital customers based on similarity. They settled on five clusters that represented very different consumer credit behavior (e.g., those who spend a lot but pay off their cards in full each month versus those who spend a lot and keep their balance near their credit limit).



*The problem* with using this clustering immediately for decision making is that *the data are not available when the initial credit line is set*. Haimowitz and Schwarz *took this new knowledge and cycled back to the beginning of the data mining process*. They used the knowledge to define a precise predictive modeling problem: using data that are available at the time of credit approval, predict the probability that a customer will fall into each of these clusters. *This predictive model then can be used to improve initial credit line decisions.*

(“Data Science for Business” by Provost&Fawcett)

## ◆ Deployment

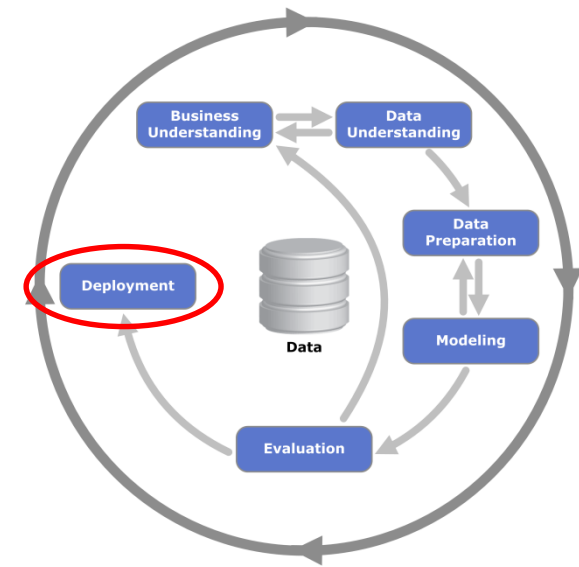
Put the results of data mining into real use.

Realize return on investment.

e.g. A new fraud detection system is built into an information system of a credit card firm.

Data mining techniques might be deployed.

- The world may change faster than data analysts can adapt.
- Reduce modeling tasks of data analysts.



## Case study: Canadian travel agency

Travel Alberta wants to develop an intraprovince marketing campaign to increase domestic Alberta tourists. Two researchers studied intraprovince tourist behavior in Alberta and found that key factors: the quality of accommodations, school holidays, and weather conditions, for example.

### Deployment

These study findings resulted in the launching of a new marketing campaign, “Alberta, Made to Order,” based on customizing the marketing to the cluster types uncovered in the data mining.

More than 80 projects were launched, through a cooperative arrangement between government and business. “Alberta, Made to Order,” television commercials have now been viewed about 20 times by over 90% of adults under 55.

Travel Alberta later found an increase of over 20% in the number of Albertans who indicated Alberta as a “top-of-the-mind” travel destination.

## Case study: Canadian travel agency

### *TRAVEL ALBERTA “MADE TO ORDER” SUMMER CAMPAIGN 2002*

Travel Alberta has launched its summer campaign theme for 2002 entitled “Made to Order”. This campaign is aimed at increasing In-Province tourism by Albertans. The goal of the tag line is to show people that there is more choices in terms of experiences and destinations in Alberta than their preconceptions have allowed them to visualize.

The campaign will appear in a variety of media vehicles including the usual television, radio, print and web as well as some unique opportunities using a provincial information cruiser that will tour around special events in Alberta and a Destination Awareness component that will provide tourism operators the opportunity to profile their packages and programs.

If your community would like to be involved in the “Made to Order” Summer Campaign 2002 please visit [www.travelalberta.com](http://www.travelalberta.com)

More than 80 projects were launched, through a cooperative arrangement between government and business. “Alberta, Made to Order,” television commercials have now been viewed about 20 times by over 90% of adults under 55.

Travel Alberta later found an increase of over 20% in the number of Albertans who indicated Alberta as a “top-of-the-mind” travel destination.

- ◆ **Business Understanding**  
State business objective, Data Mining objective and success criteria.
- ◆ **Data Understanding**  
Estimating costs and benefits of obtaining and cleaning data
- ◆ **Data Preparation**  
Convert data to a form suitable for data mining techniques.
- ◆ **Modeling**  
Data mining techniques, e.g. regression, are applied.
- ◆ **Evaluation**  
Ensure that the model and the result satisfy the business goal.
- ◆ **Deployment**  
Put the results of data mining into real use.

