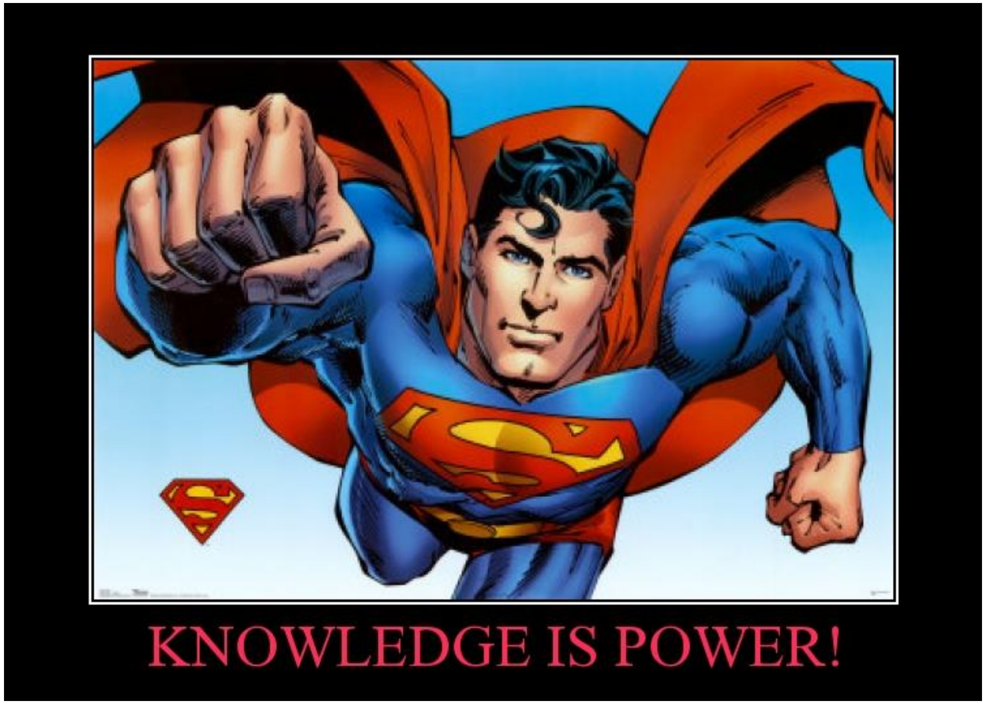
Data Analytics and Intelligence COMP8811

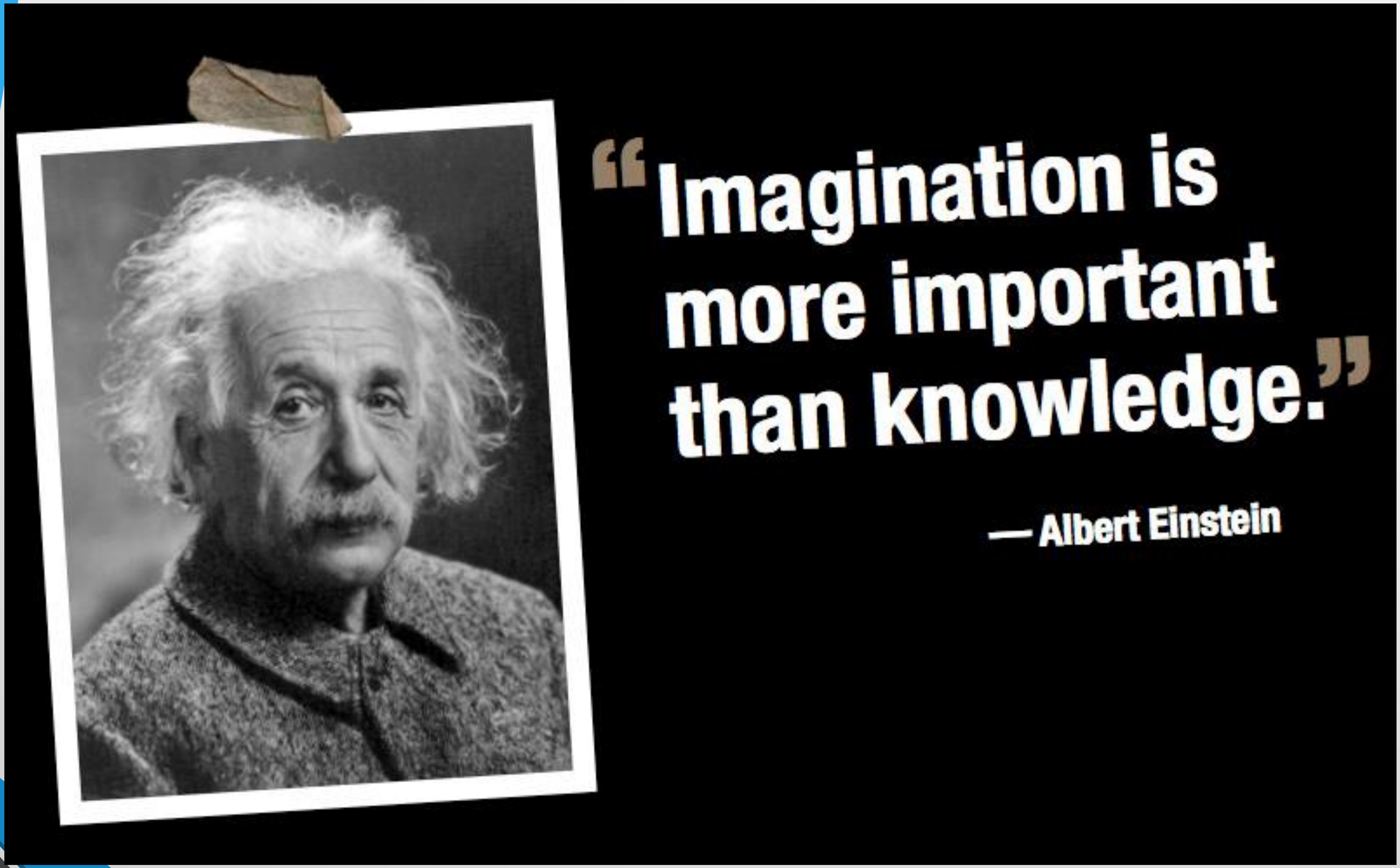
Introduction

Lecturer: Dr Neda Sakhaee

"The answer to my problem is hidden in my **data** but I cannot dig it up!"



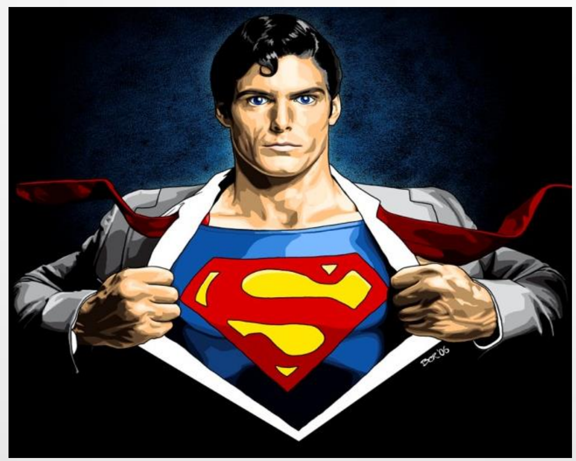
Because you don’t have enough **knowledge**!


"Imagination is more important than knowledge."
—Albert Einstein

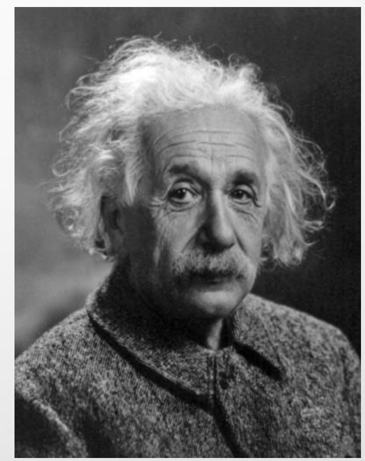
In decision making:

What do you prefer to look like?

Goofy?



Superman?



Einstein?

Let's try to be a SUPERMAN in Business environments!
We need "Data Analytics" to turn data into knowledge.

Data Analytics

- Q: How to make a good business?
- A: Making right ***decisions*** at the right time.

- Q: How to make right decision?
- A: Using ***knowledge***.

- Many businesses view ***knowledge*** as their final objective.

Data Analytics

- For being a Superman in business environment, we just need to turn ……**data**…… into **knowledge**……………. on a continuous base.

How?

Data Analytics

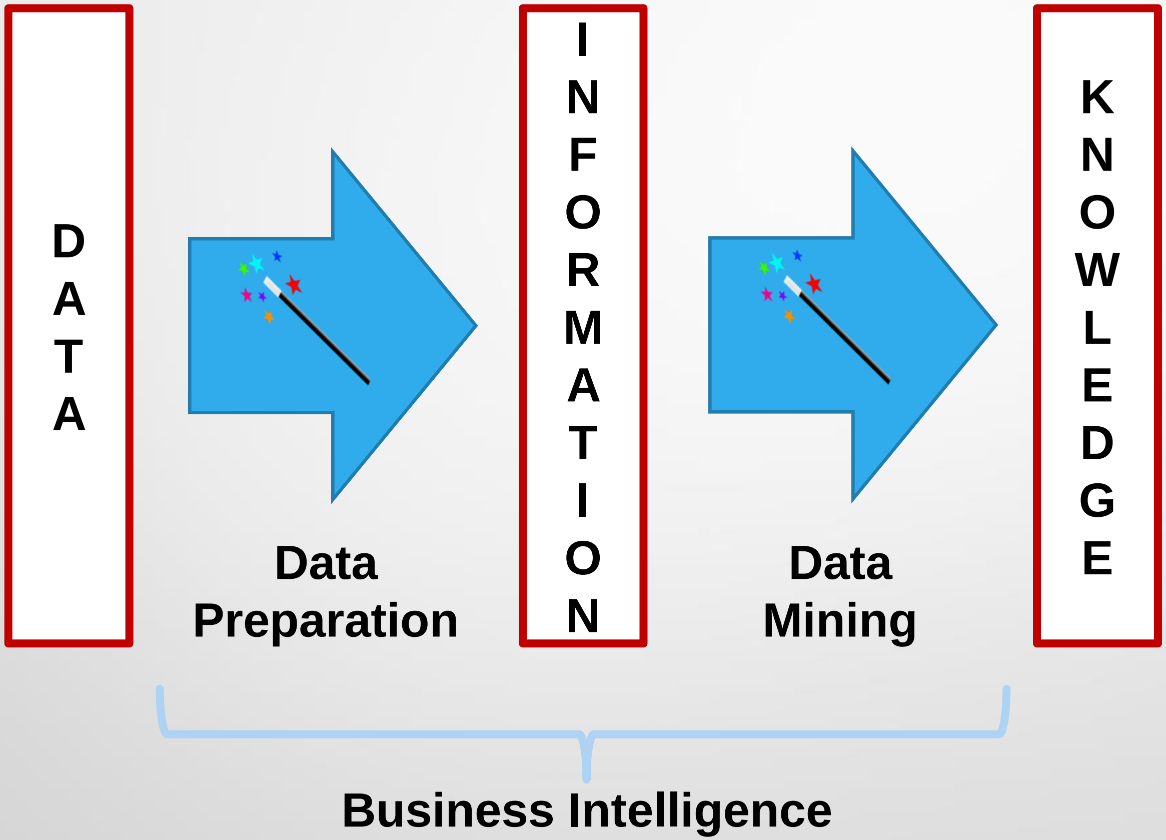
- ***Data*** are collected on a regular basis in the form of bits, numbers, symbols, and objects.

- ***Knowledge*** is integrated **information**, which includes facts and relationships that have been perceived, discovered or learned.

Data Analytics

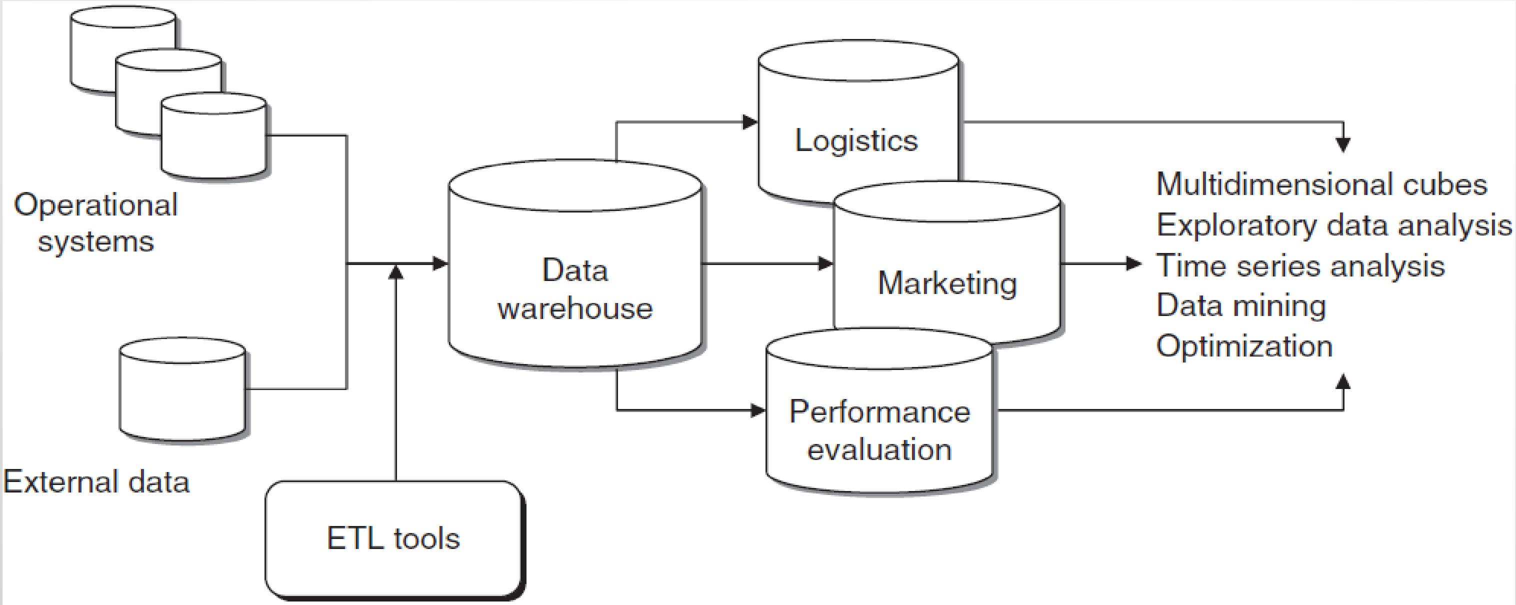
- ***Data*** are collected on a regular basis in the form of bits, numbers, symbols, and objects.
- **Information** is organized ***data*** , which pre-processed, cleaned, arranged into structures and stripped of redundancy.
- ***Knowledge*** is integrated **information** , which includes facts and relationships that have been perceived, discovered or learned.

Data Analytics



Business Intelligence

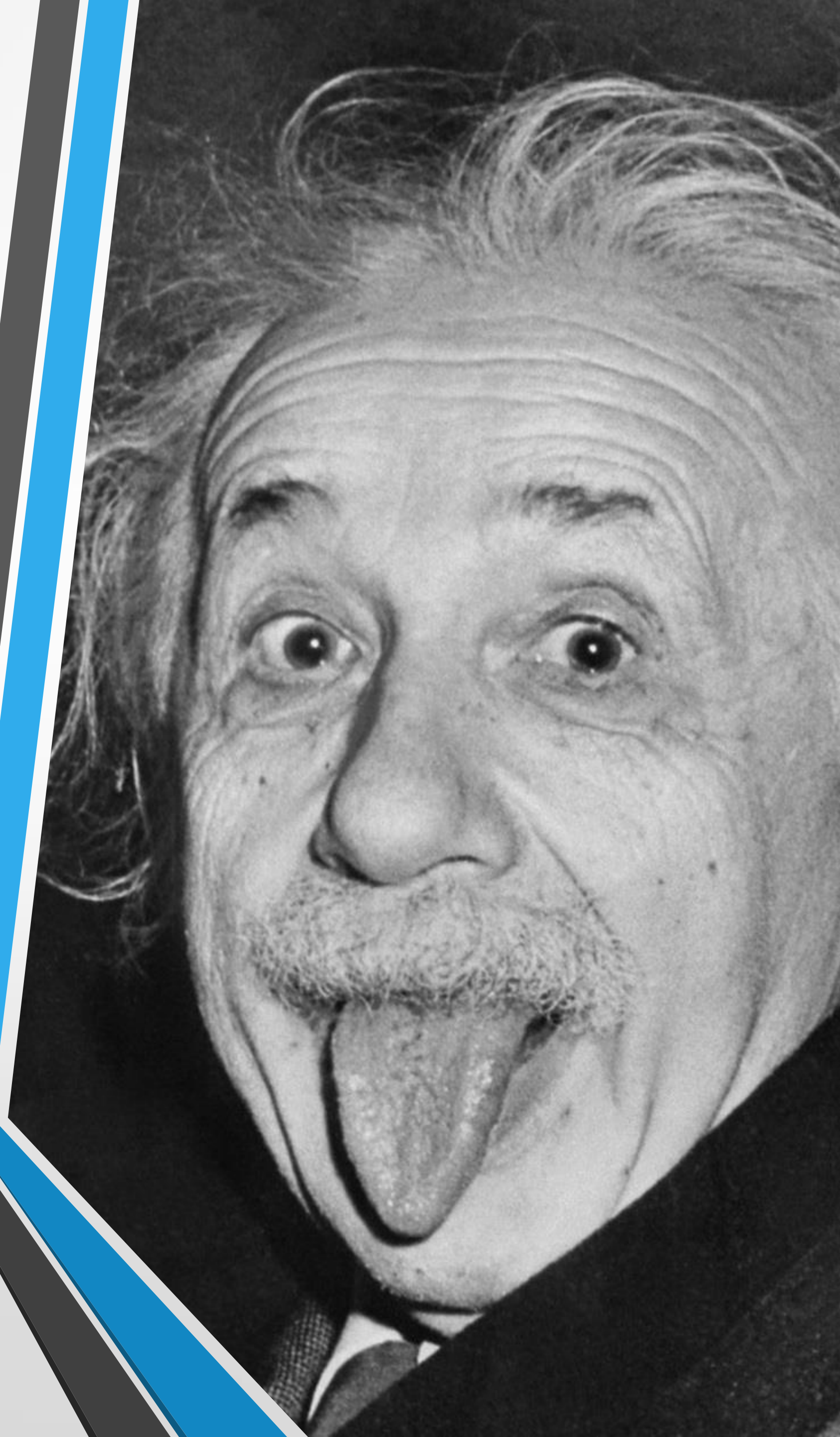
- Typical BI architecture



Data Analytics

- Is knowledge enough to make a right decision?
- Today, most business managers realize that a gap exists between having the right knowledge and making the right decision.
- Traditional BI cannot go further!

- We need to be more intelligent! **How?**
- We need to predict future! **How?**
- We need to use our imagination, **as he said!**



Data Analytics and Intelligence

- The future of BI lies in systems that can provide answers and recommendations, rather than mounds of knowledge in the form of reports.

Data Analytics and Intelligence

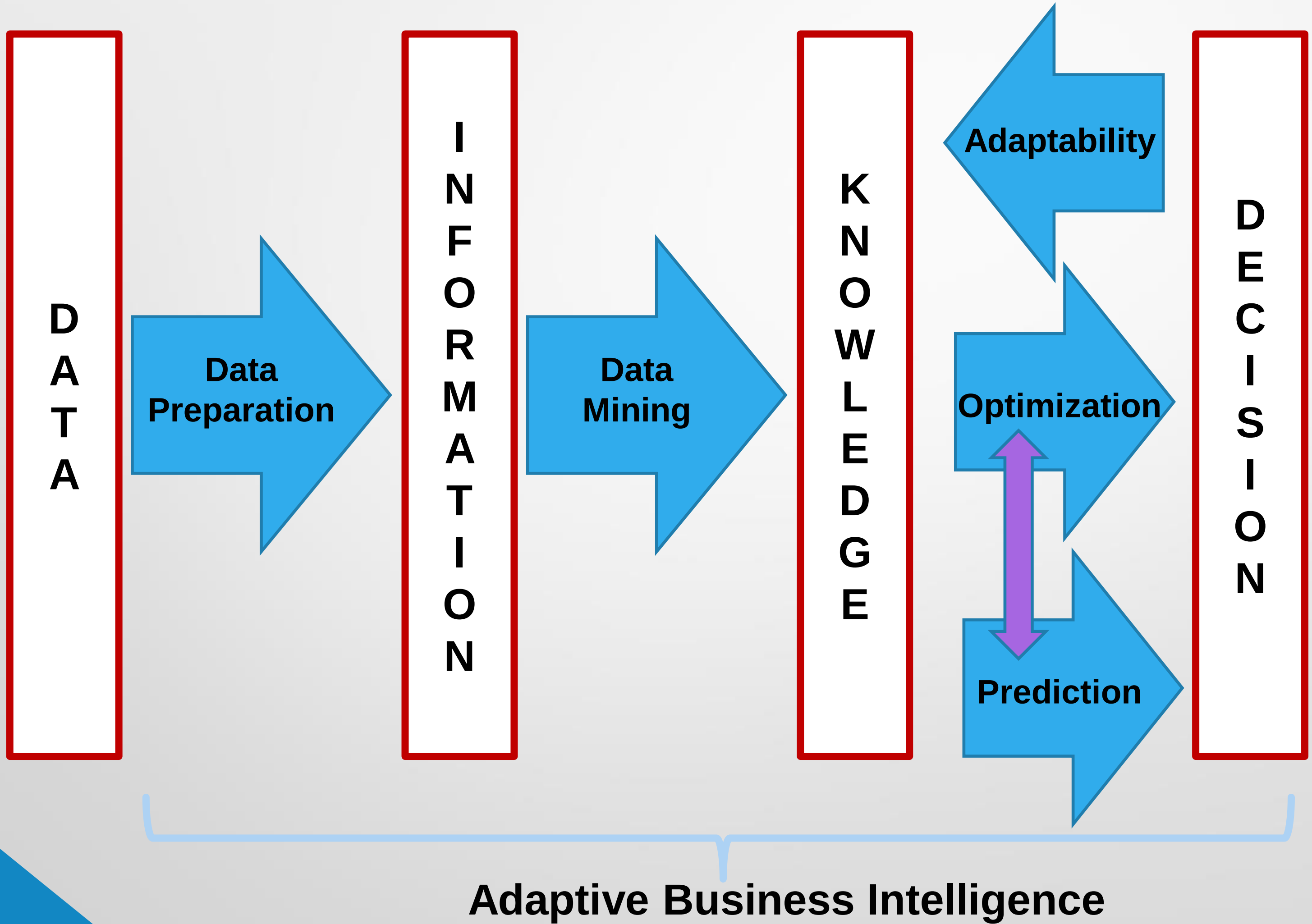
- As a result, there is a new trend emerging in the marketplace called:

Adaptive Business Intelligence.

Data Analytics and Intelligence

- In addition to performing the role of traditional Data Analytics (transferring data to knowledge), Data Analytics and Intelligence also includes the adaptive decision-making process which is based on **prediction** and **optimization**.

Data Analytics and Intelligence



Data Analytics and Intelligence

- We use **prediction** and **optimization** techniques to build self-learning decision systems.
- Prediction and optimization modules to recommend near-optimal decisions, and adaptability module for improving future recommendations.

Fun facts

IN 2020 EVERY PERSON GENERATED 1.7 MEGABYTES IN JUST A SECOND.

80-90% OF THE DATA WE GENERATE TODAY IS UNSTRUCTURED.

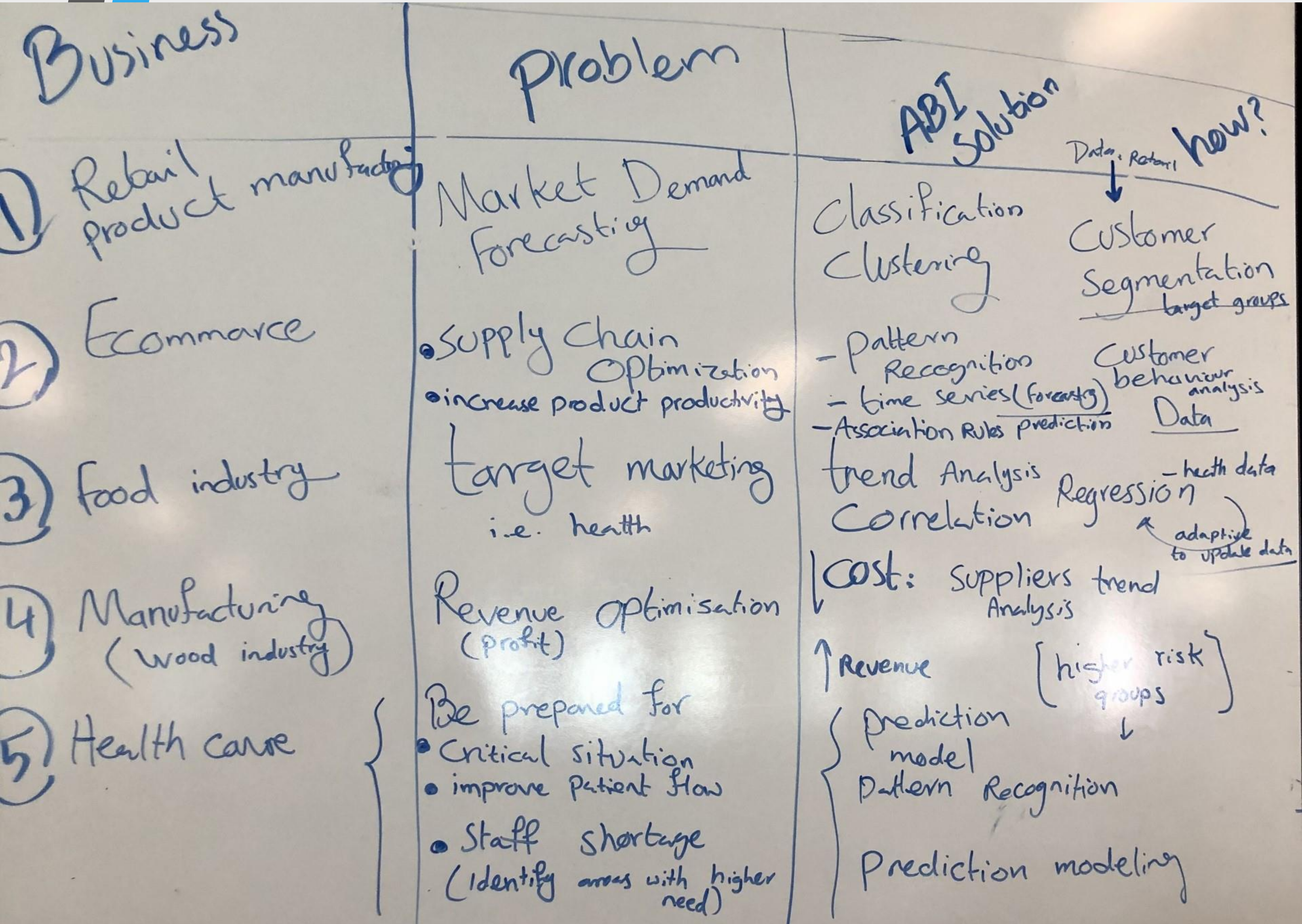
95% OF BUSINESSES CITE THE NEED TO MANAGE UNSTRUCTURED DATA AS A PROBLEM FOR THEIR BUSINESS.

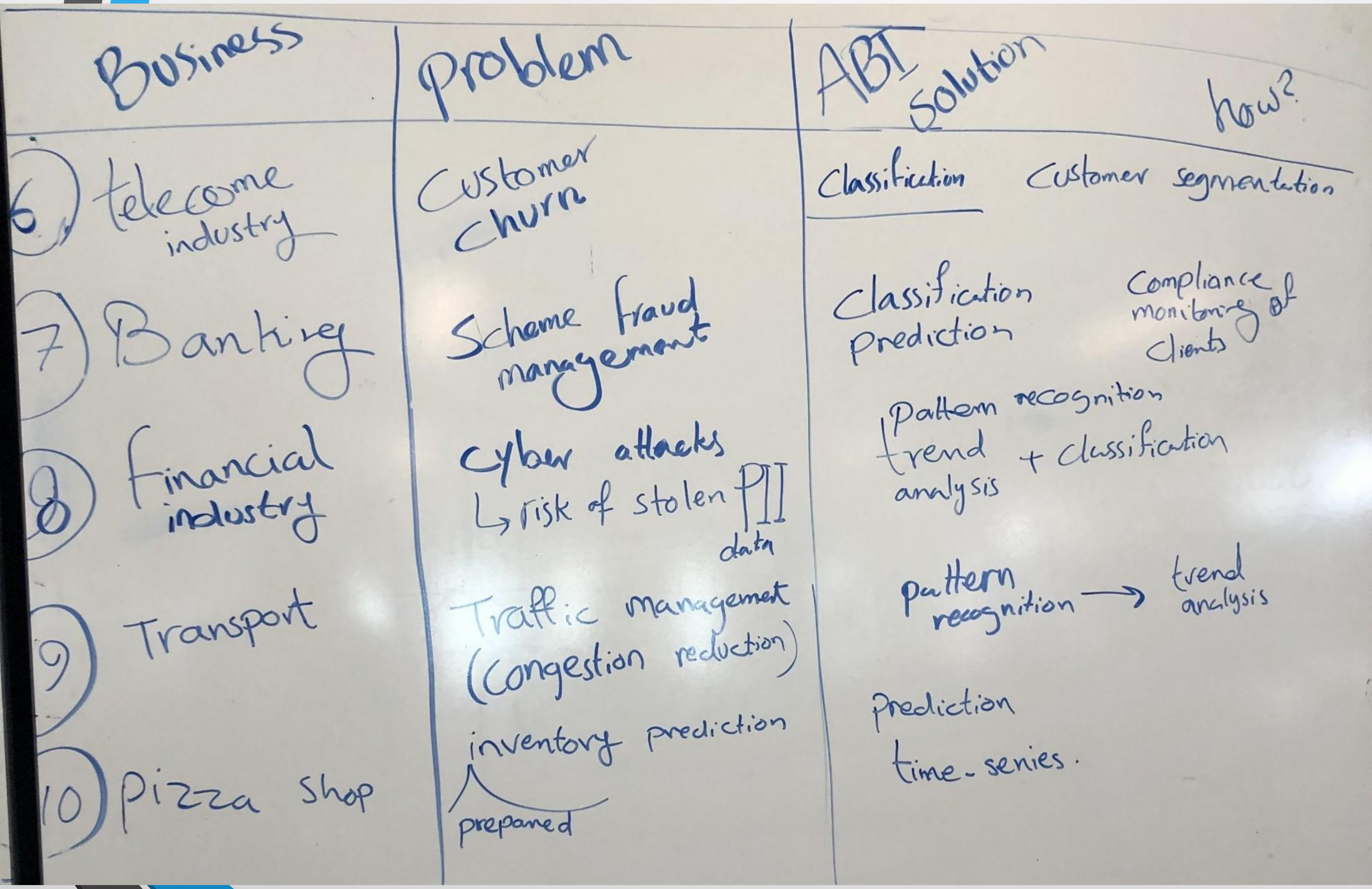
Class activity 30 min

THINK ABOUT EXAMPLES OF HOW DATA ANALYTICS AND INTELLIGENCE CAN HELP DIFFERENT BUSINESSES.

DISCUSS YOUR EXAMPLE S IN YOUR TEAM.

SHARE YOUR EXAMPLES WITH OTHERS WHEN YOU ARE READY.


Business
Problem
ABI Solution
how?
1) Retail product manufacturing
2) Ecommarce
3) food industry
4) Manufacturing (wood industry)
5) Health care
Market Demand Forecasting
• Supply chain Optimization
• increase product productivity
target marketing
i.e. health
Revenue optimisation (profit)
Be prepared for
• Critical situation
• improve Patient flow
• Staff shortage (Identify areas with higher need)
Classification
Clustering
Data, Retail
Customer Segmentation
target groups
- pattern Recognition
- time series (forecasting)
- Association Rules prediction
Customer behaviour analysis
Data
trend Analysis
Correlation
Regression
- heath data
adaptive to update data
cost: suppliers trend Analysis
Revenue
[higher risk groups]
prediction model
Pattern Recognition
Prediction modeling


Business
Problem
ABT Solution
how?
6) telecome industry
Customer churn
Classification
Customer segmentation
7) Banking
Scheme fraud management
Classification prediction
Compliance monitoring of clients
8) financial industry
cyber attacks
↳ risk of stolen PII data
Pattern recognition
trend analysis
+ classification
9) Transport
Traffic management (Congestion reduction)
pattern recognition → trend analysis
10) Pizza shop
inventory prediction
prepared
Prediction
time-series.

Pause

5 min

Why Data Analytics and Adaptive Business Intelligence?

- **Q: Why a complex method like ABI?**
- A: Complex business problems are difficult to solve.

- **Q: Why business problems are difficult to solve?**
- A: Because of the following characteristics:
- 1) the number of possible solutions is so large
- 2) time-changing environment
- 3) Problem-specific constraints
- 4) Multi objective problems (possibly conflicting)
- 5) other items e.g. noisy data, uncertainty and etc.

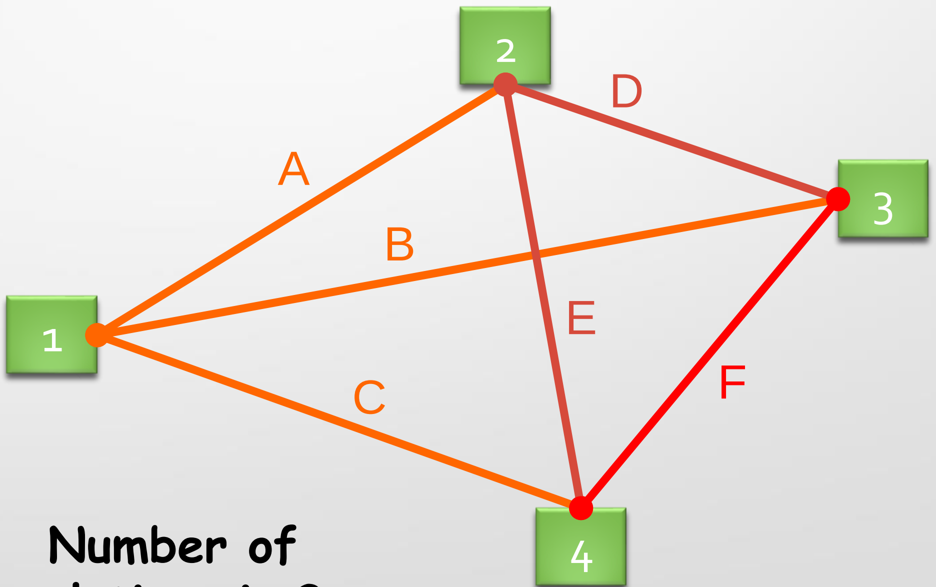
1) Number of possible solutions

Example: traveling salesman problem

The problem is very simple: traveling the shortest possible distance, the salesman must visit every city in his region once and then return home.

- for 4 cities:

Case 1	A	D	F	C
Case 2	A	E	F	B
Case 3	B	D	E	C
~~Case 4~~	~~B~~	~~F~~	~~E~~	~~A~~
~~Case 5~~	~~C~~	~~F~~	~~D~~	~~A~~
~~Case 6~~	~~C~~	~~E~~	~~D~~	~~B~~



Number of solutions is 3

1) Number of possible solutions

Example: traveling salesman problem

- for 4 cities:
  - 3 choices for the first trip
  - 2 choices for the second trip
  - 1 choice for the third trip
  - symmetric trips should be removed
  - $N = (3 \times 2 \times 1) \div 2 = 3$
- for 5 cities: $(4 \times 3 \times 2 \times 1) \div 2 = 12$
- for 6 cities: $(4 \times 3 \times 2 \times 1) \div 2 = 60$
- ...
- For 10 cities: 181440
- ...
- For 50 cities: **about 10^{62}**
- **New Zealand has 43 cities with more than 10,000 people!**

Our planet holds about 10^{21} liters of water!

Each year has about 3×10^{8} seconds!

Even if 1n Sec is needed for processing each case, we need more than our universe age to process all the possible solutions.

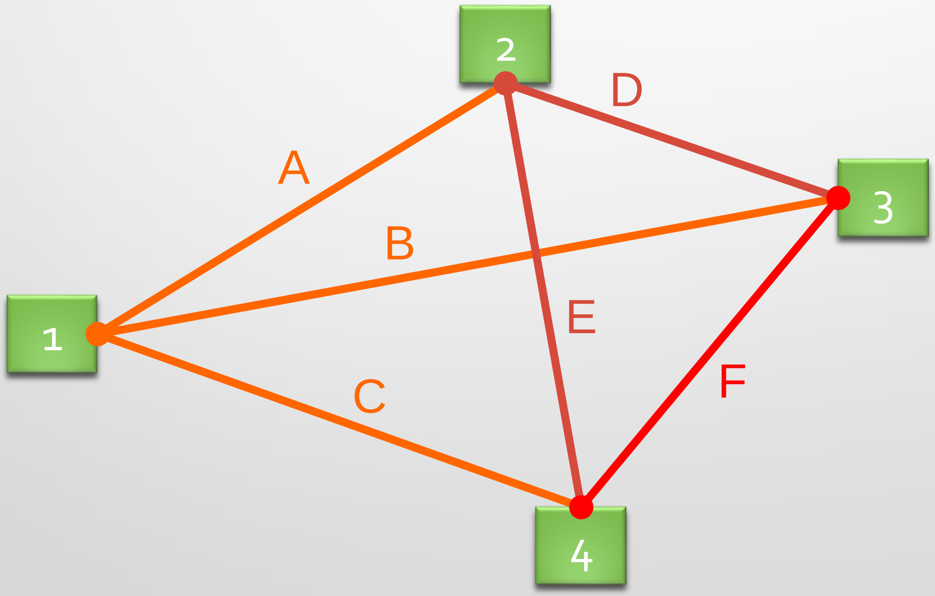
2) Time changing environment

Because real-world business problems are set in time-changing environments, it is important to address the time factor clearly and in detail.

The optimal solution at this time period may not be optimal for the next time period.

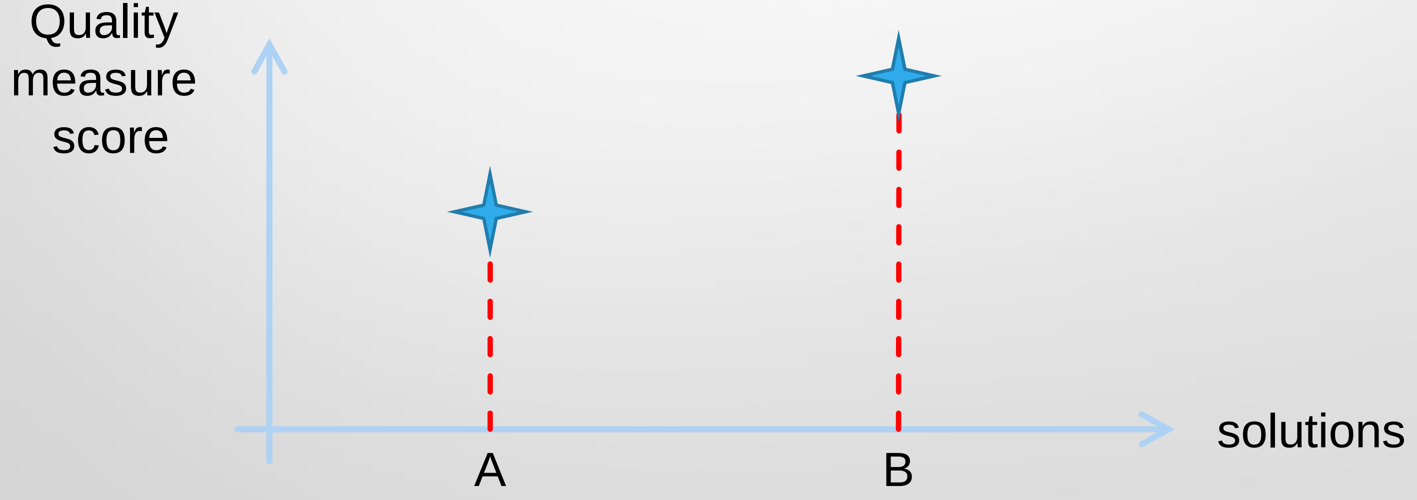
2) Time changing environment

- Examples of time changing environment factors for the traveling salesman problem:
  - Roads B and C are dangerous in Winter!
  - Construction on Road D from March 16, 2021.



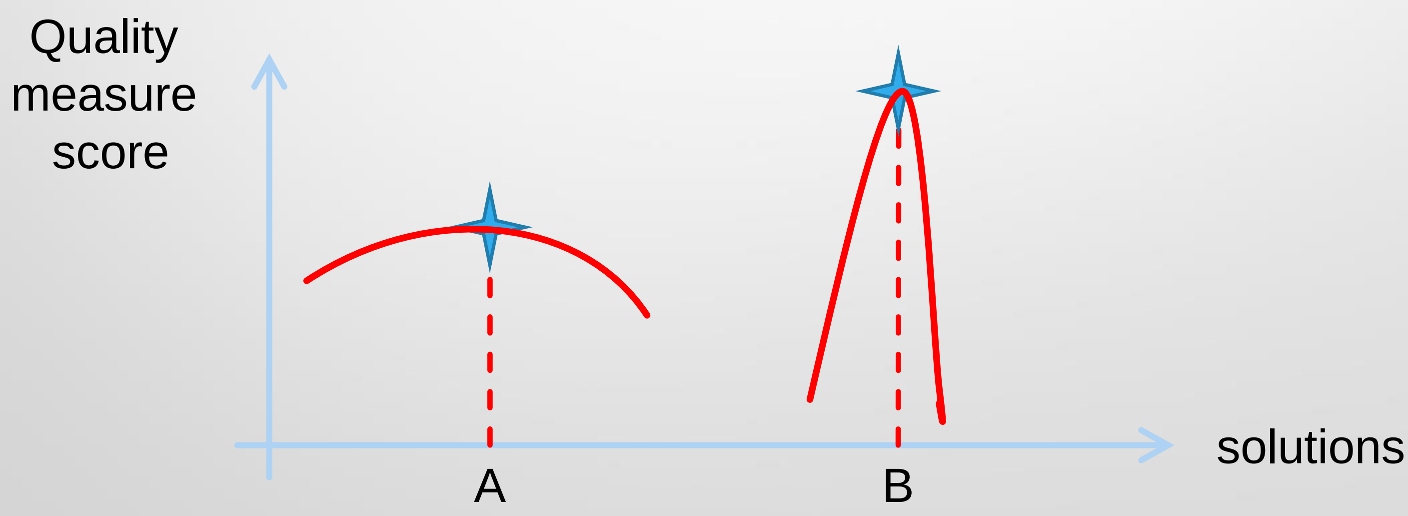
2) Time changing environment

- Imagine that we are considering the implementation of solution A or solution B. Which of these two solutions would we select?



2) Time changing environment

- if we are forced to modify solution B for any reason (equipment failure, bad weather, etc.), then the quality of solution B will deteriorate very quickly.
- Solution A, on the other hand, is much more “stable” in the sense that it can tolerate changes and modifications without a sharp drop in quality. Given that solution A is less risky than solution B,



3) Problem-Specific Constraints

- All real-world business problems have constraints of some sort, and if a particular solution does not satisfy these constraints, then we cannot consider this solution.

3) Problem-Specific Constraints

- Examples of problem-specific constraints for the traveling salesman problem:
  - capacity limits,
  - delivery time windows,
  - maximum driving time, etc.
  - not transporting chemicals and food together on the same truck
  - personnel preferences

3) Problem-Specific Constraints

It is necessary to assert the relative importance of each constraint (hard or soft) by assigning numeric weights to it.

When solving the problem, we can then use these weights to calculate a final quality measure score for each possible solution.

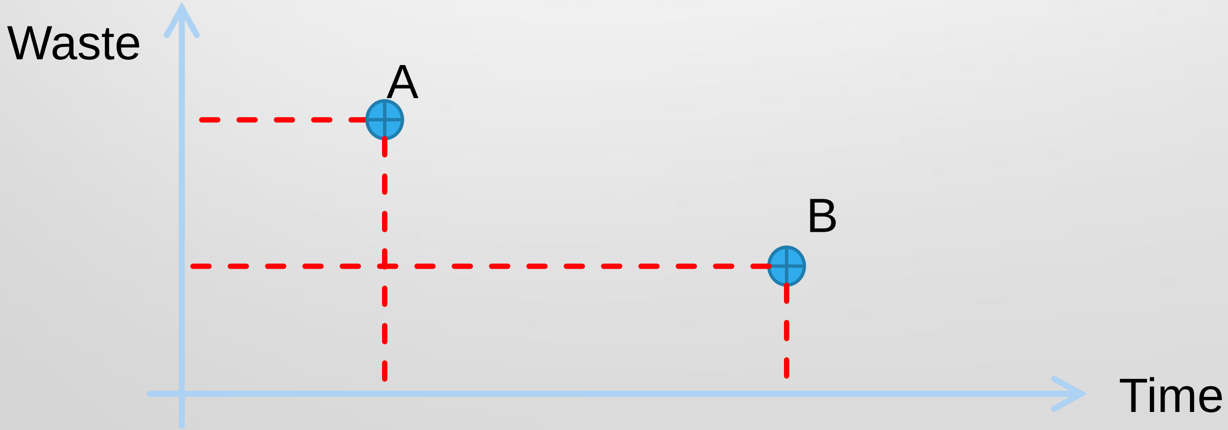
4) Multi-objective Problems

It is quite unusual for any real-world business problem to have only one objective.

For example: The objectives may include the **minimization of production time** and the **minimization of material waste**. These objectives might “work” against each other, as the minimization of production time may trigger an increase in material waste, and vice versa.

4) Multi-objective Problems

- Let us consider solutions A and B: Which of them is better? Solution A is faster, but the amount of material waste is higher, and vice versa.
- In problems with multiple objectives, it is possible to find a solution that is best with respect to the first objective, but not the second, and a different solution that is best with respect to the second objective, but not the first.



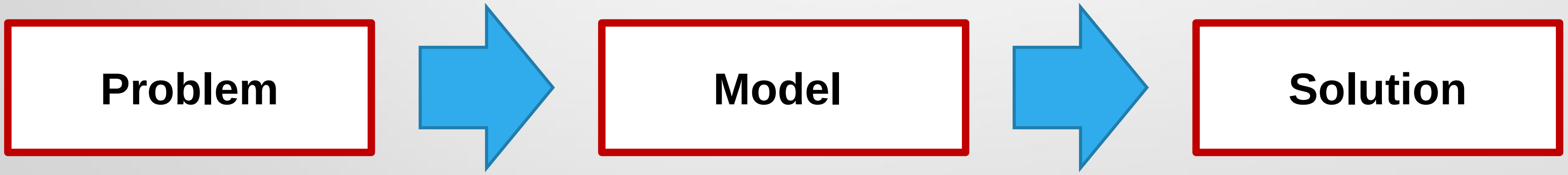
4) Multi-objective Problems

A or B?

It is impossible to answer this question without first agreeing on a common denominator for time and waste: we can translate both objectives into $ by calculating that five minutes of production time is worth $100, and each pound of material waste is worth $180. We can then calculate the merits (expressed in $) of both solutions, compare the numbers, and select the solution with the lowest dollar figure.

Problem Solving Process

- The problem-solving process consists of two separate steps:
  - Creating a model of the problem
  - Using the model to generate a solution



Problem Solving Process

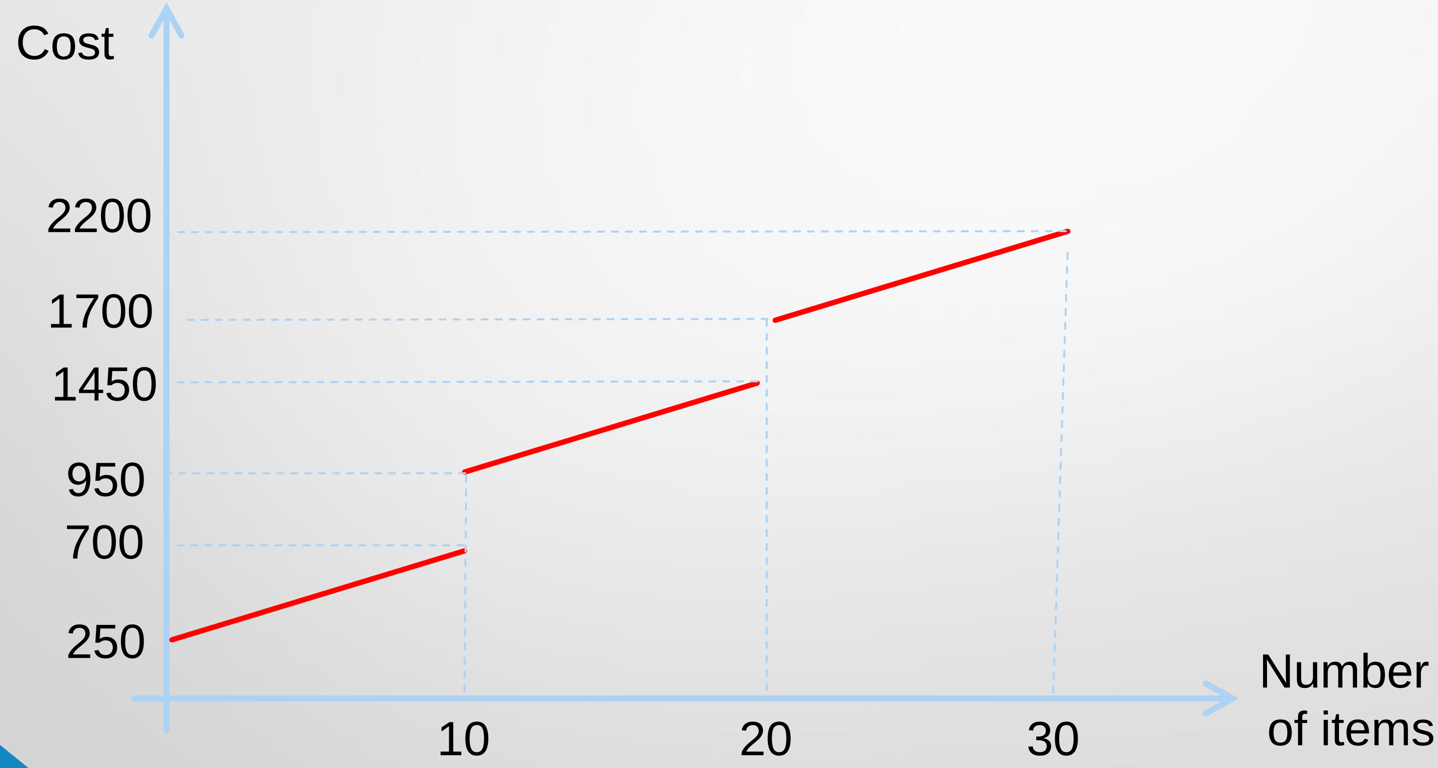
- We can only find a solution to the model; hence, the accuracy of the model is very important.
- Accurate model → meaningful solutions
- Vague model → meaningless solutions

Problem Solving Process

- Example: modelling cost for the transportation between a warehouse and a distribution centre given that
  - The cost is zero when there is no delivery.
  - Each truck can transfer up to 10 items.
  - Hiring a driver costs 250$ per truck
  - Each item costs an extra transportation 45$ (fuel and etc…)

Problem Solving Process

Model:

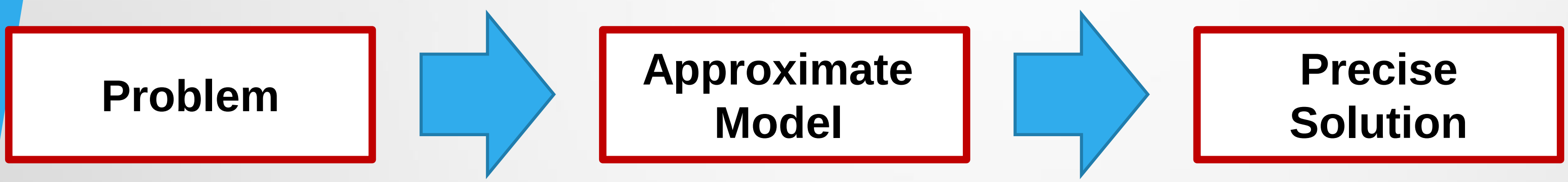


Problem Solving Process

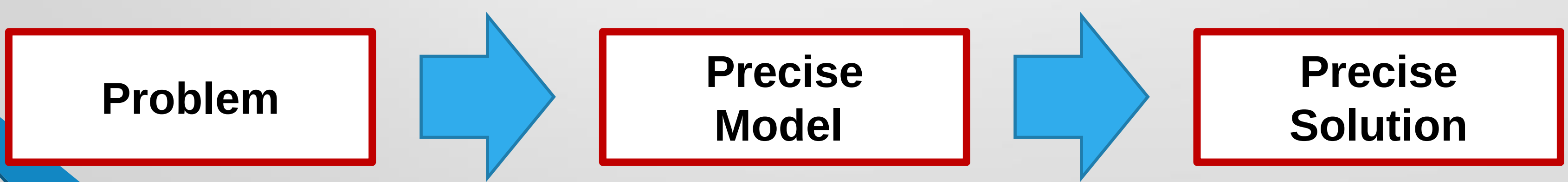
- In real-world situation, the problem and, thereby, the model is more complicated:
  - There are 80 warehouses and 5 distribution centers (80x5=400 variables)
  - Constraint (transportation law, environmental issue, driving regulations) should be considered.
  - The total transportation cost should be minimized.

Problem Solving Process

- We can simplify the model, so that traditional optimization techniques can be used.

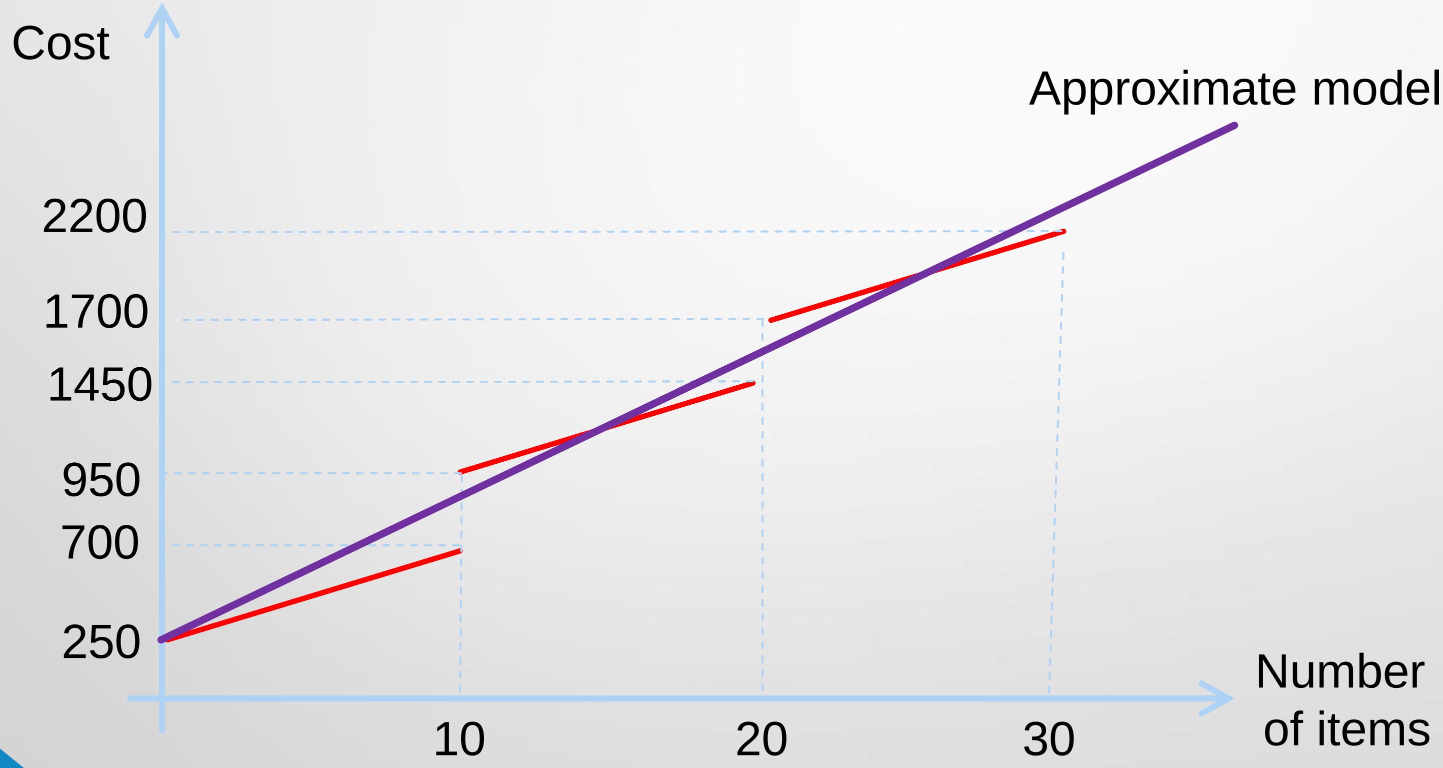


- Or we can leave the precise model unchanged and use untraditional optimization techniques.



Problem Solving Process

- Example for simplifying a model



How we implement Data Analytics and Adaptive Business Intelligence?

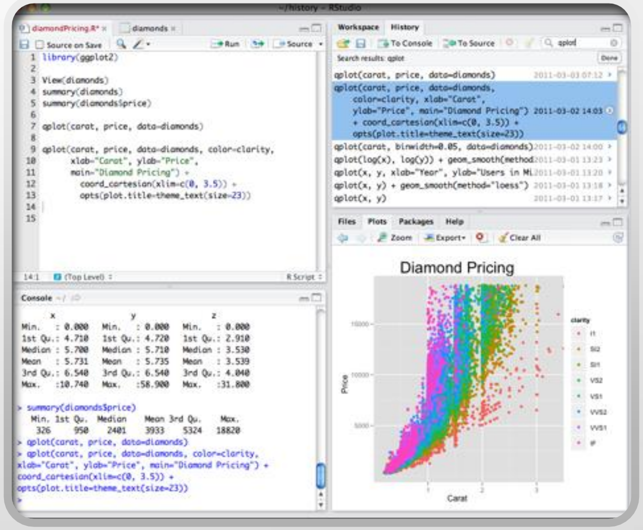
- Our system may include major components:
  - A data mining module (data preparation, visualization and analytics)
  - A prediction model (on the data mining results)
  - An optimization module (recommend the best solutions based on the prediction results)
  - An adaptability module (responsible for adapting the prediction module to the time-changing environment)

Q&A

Feedback

R

Data Mining Tools



https://www.r-project.org/

https://rstudio.com/products/rstudio/#rstudio-desktop/

Assignment 1