

# **15CSE03 - CUSTOMER RELATIONSHIP MANAGEMENT**

## **Unit III**

**Collecting Customer Data**

# Collecting Customer Data

- Deals with customer data
- It discusses about
  - I. Classification of customers data
  - II. How to get customer data?
  - III. What can be derived from these data?

# **Types of customer data**

**Customer data comes from variety of sources.**

**The three main types of customer data include**

- Descriptive data
- Promotional data
- Transactional data

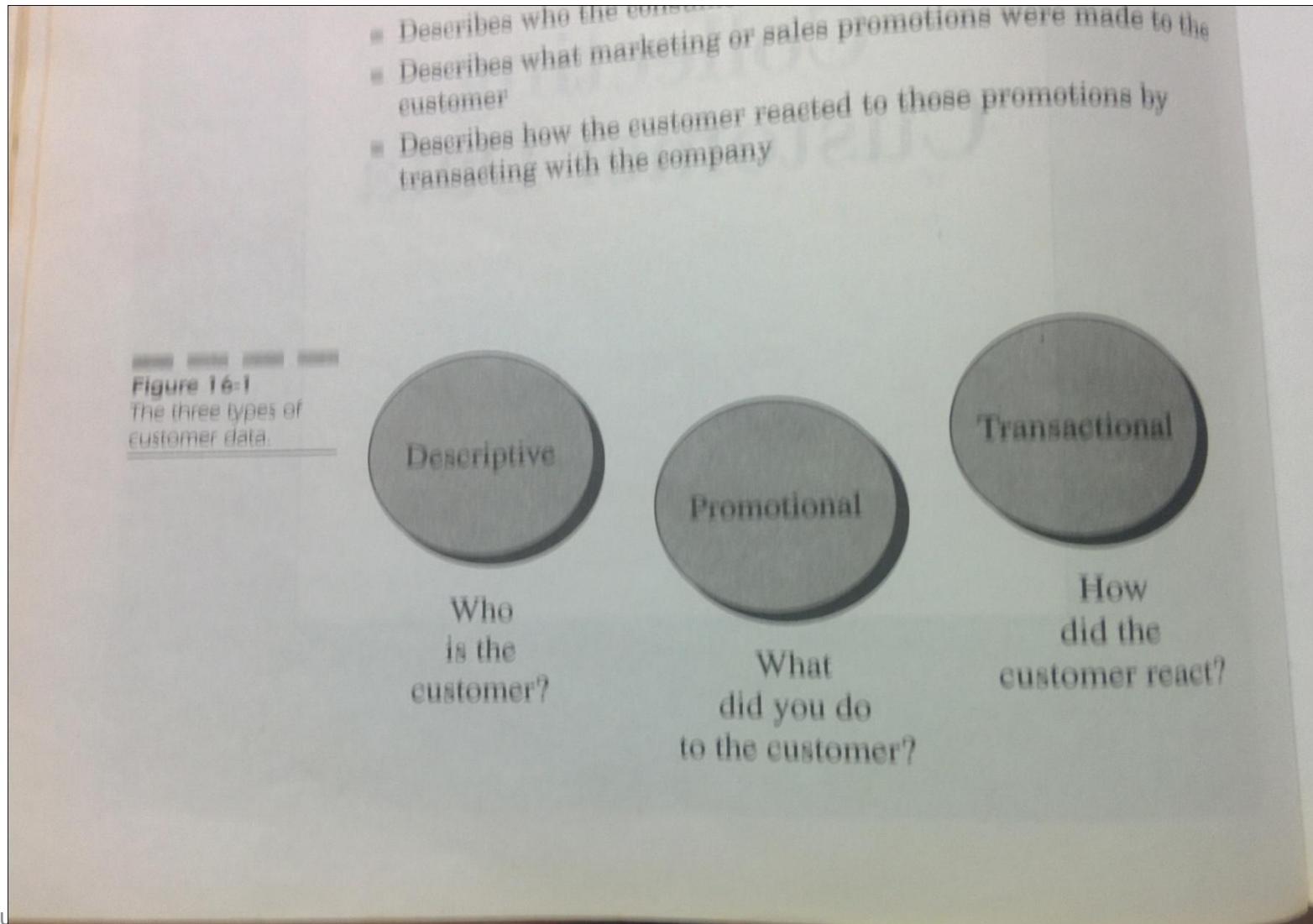
**This data Describes**

Who is the customer?

What sales promotions were made to the customer?

How the customer reacted to the promotions by transacting with the company?

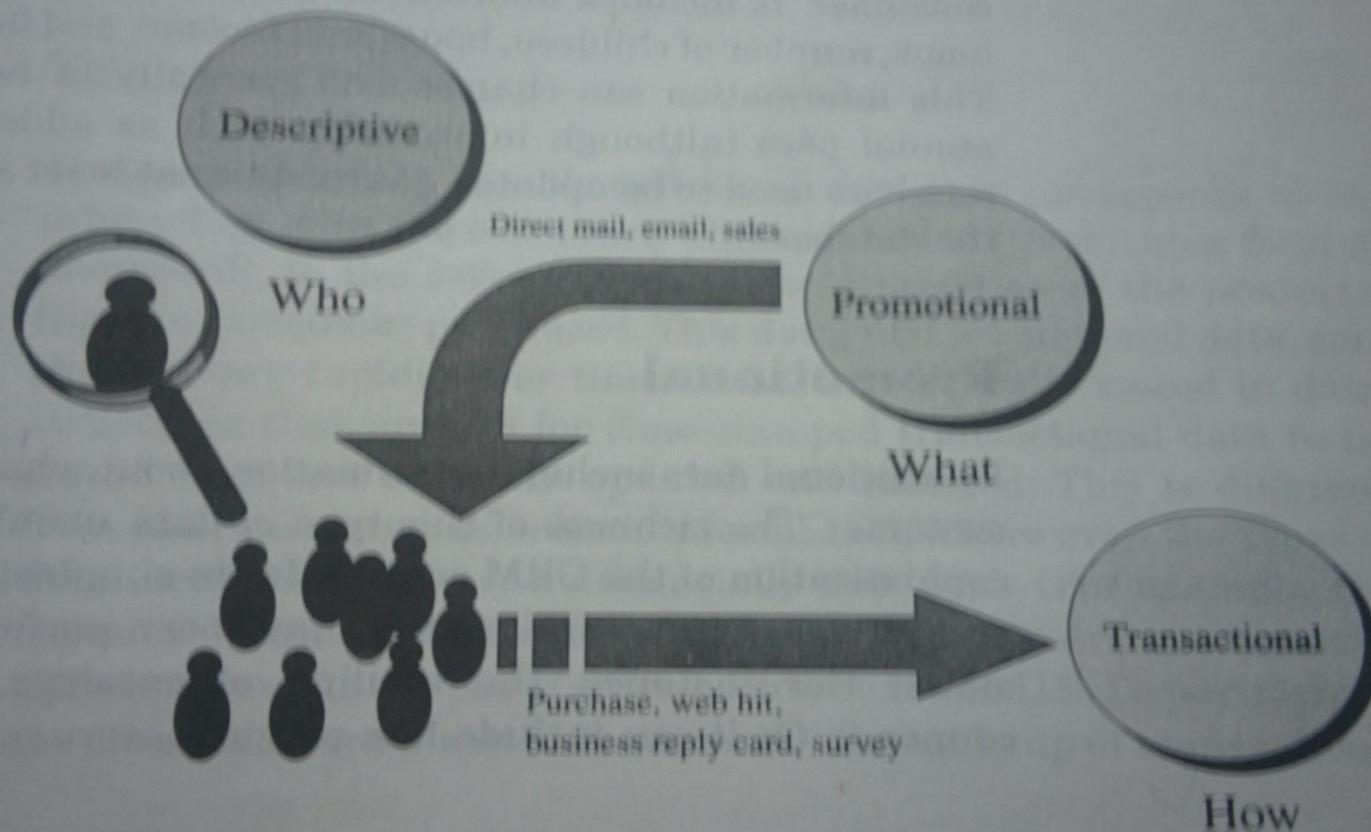
# The three types of customer data *contd..*



# The three types of customer data *contd..*

of what you did to each customer. You are creating many small experiments with your customer base, and you must know what is

**Figure 16-2**  
Knowing who, what, and how provide enough information to perform data mining.



## Descriptive data

- It is the data about the customer or consumer
- Summary of customer details – a relational table with multiple columns
- This data does not change quickly because this data is description about the consumer.

Eg. age, gender, location of home, No. of children, household income, individual income

- Certain data like phone number, location and the address may be updated quarterly or semi annually.

# Promotional data

- It includes information about what was done to each customer.
- The richness of data usually depends on the sophistication of the CRM System.
- It is a list of promotion of interventions performed with the customer.
- The types of information that could be collected include
  - ✓ Type of intervention - Sales, telemarketing, print advertising, broadcast advertising, web advertising
  - ✓ Description of intervention - colour of the post card
  - ✓ Media
  - ✓ Timing - The data and time of the day
  - ✓ Description of intent - justification for the promotion (eg. Background music)
  - ✓ Financial - Fixed and variable costs of the intervention

## Transactional data

- It broadly defines all data that corresponds to an interaction with the consumer.
- It includes a phone call to the service desk or description of the products that the consumer purchased.
- This data can change very rapidly over time.
- Typically stored in a Database so that it can be easily updated.
- It is different from descriptive data since descriptive data does not change frequently.
- Transactional data can vary within short time.
- This data has time stamp – can be easily updated and changed

- Without the three piece of information,
  - Who the customers are,
  - What was done to them,
  - How the customers reacted?
- It is not possible to improve or optimize the system
- In order to optimize and improve the system, **we must interact with the customer.**
- We must differentiate good and bad customers, profitable and unprofitable customers.
- We must know who they are and also differentiate them.

- Create small experiments with your customer base to identify what is working and what is not working.
- To judge the actual working of the system, we must measure the results.
- Customer data can be collected from different sources.  
(From our historical data or from some of the large data providers)

The source will help us to classify the data into descriptive, promotional or transactional

# *Collecting customer data*

- Data can be from

1. Internal sources
2. Web data

## 1. Internal sources

- Marketing team- promotional activities – from the mail house
- Existing customer lists

Collecting data from existing process like existing customer list is sometimes hard because

1. The data owner group may not want to disclose the data to CRM system.(eg. Savings account division of a branch will not provide the customer details to the brokerage arm of the bank.)
2. Not shared due to customer privacy. the customers feel that their information is being mishandled and not carefully monitored.

## *Collecting customer data*

### **2. Web data**

- Web data is more important because promotion and transactional events take place in the web- ecom and mcom.
- The data is stored as web log files. These files give info on every minute buying decision and influence can be measured. (used for analysis)

# *Connecting Customer Data*

- *Data Warehouses and Data Marts*
- *Data Pumps and connectors*
- *Long distance connections*

## *Connecting customer data*

- **Data warehouse and data marts**

- Single localized data store for all customers with data relevant to making business decisions.
- They are fed by transactional systems.
- It is one of the many sources of data that support the DM system for the CRM.
- Data changes frequently and so the need for Data warehouse.
- Parts of the DW will be dynamic and some may be static. They cannot be stored within one monolithic structure and consistently up to date.
- Defined technically about how data is stored by different schemas – star, snowflake or galaxy

# *Connecting customer data*

- **Data pumps and connectors**
  - They are pieces of software that glue the entire system together.
  - They regularly connect desperate data sources together.
  - They provide a layer of abstraction between the way the data exists and the way the data is needed to support the Data mining for CRM.
  - This abstraction allows the system to
    - Accommodate the new and changing data.
    - Rapidly change to new data and data structures.
    - Perform a consistent way of moving and processing the data that is repeatable and possible to validate.
    - Create an optimized data structure to support data mining and CRM without forcing the design constraints.
- Does ETL job – Data is abstracted for mining purposes

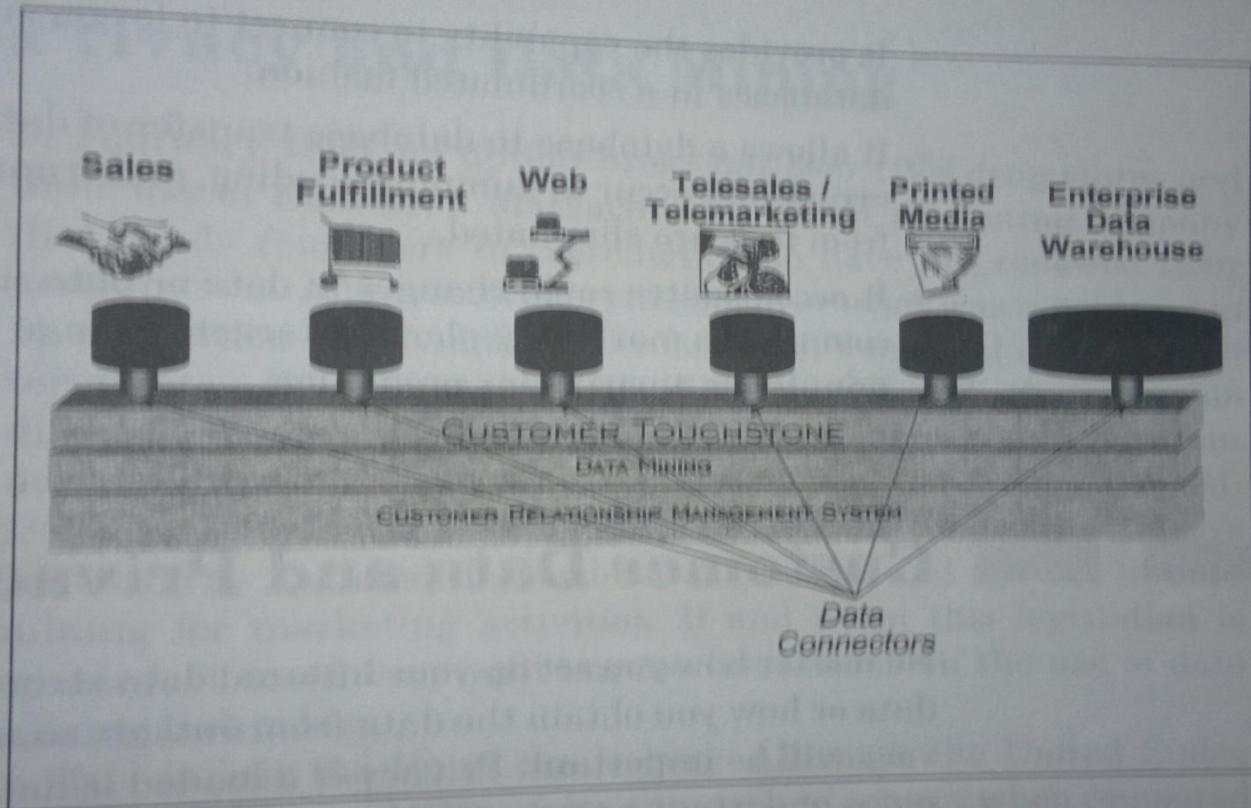
# *Connecting customer data*

## Chapter 16: Collecting Customer Data

369

**Figure 16-3**

An abstract view of a data mining system with data connectors.



... can vendors and partners. This can be

# *Connecting customer data*

- **Long distance connections**
  - Used for connecting data sources outside your organization.
  - This is done by secure internet transfer technology.
  - The virtual private network technology provides a secure virtual extension using the encryption technology and internet.
  - The advantages of long distance connectors are.
    - Capability to connect to large number of databases.
    - Allows Database to Database transfer and eliminates the errors due to dumping, loading and transforming the data.
    - It accommodates the rapid and last minute changes to data.
- **Remote systems, VPN, cloud, Edge, Far-Edge**

## *Customer data and privacy*

- No matter how you set up your internal data structures, how the data is obtained from other sources, CUSTOMER PRIVACY is important.
- Privacy is a loaded issue and is important when CRM becomes powerful and incorporates a DM engine.
- Privacy plays a role in Insurance companies, government agencies, etc.
- In addition to privacy, there are a number of legal issues that must be dealt with.

## *Privacy and data mining*

- Data privacy is important and many issues have arised in the past.
- Certain companies have given opt- out policy for the customer data.
  - Don't allow DM of customer data.
  - Allow DM for internal use.
  - Allow DM for both internal and External use.

Based on these options the customer data is used.

## **NEED FOR PRIVACY**

- ❖ Developments in communication and database technology have made resource sharing easy and simple.
- ❖ Sharing can be from centralized or distributed data source.
- ❖ Sharing of data for data mining can bring a lot of advantages for research and business collaboration.
- ❖ Centralized data source contains private and sensitive information such as social security number, income, credit rating, customer purchase details, etc.

## NEED FOR PRIVACY (CONTD.)

- ❖ Sharing of sensitive information causes serious privacy issues and must be properly protected.
  
- ❖ Privacy is defined as “The right of an entity to be secured from unauthorized disclosure of sensible information that are contained in an electronic repository or that can be derived as aggregate and complex information from data stored in an electronic repository”

## EXAMPLE

- ❖ Assume a departmental store is procuring products from three different suppliers. Based on some negotiations, every party comes to an agreement with the departmental store to access its data source for mutual benefits. The departmental store which is the owner of the data source feels that there are some sensitive information in the data source not to be disclosed to anyone else except the party concerned.

# PRIVACY PRESERVING DATA MINING (PPDM)

Two broad approaches

- ❖ Data Hiding : alters the data before it is delivered to the data miner so that real values are hidden.
- ❖ Secure Multiparty Computation (SMC) : This assumes that the data is distributed between two or more sites, and that these sites cooperate to learn the global data mining results without revealing the data at their individual sites.

## PPDM (CONTD.)

### ❖ Collaborating Parties

- ❖ Honest
- ❖ Semi-Honest - are honest but try to learn more from received information.
- ❖ Dishonest - are malicious and they do not follow the defined protocols.

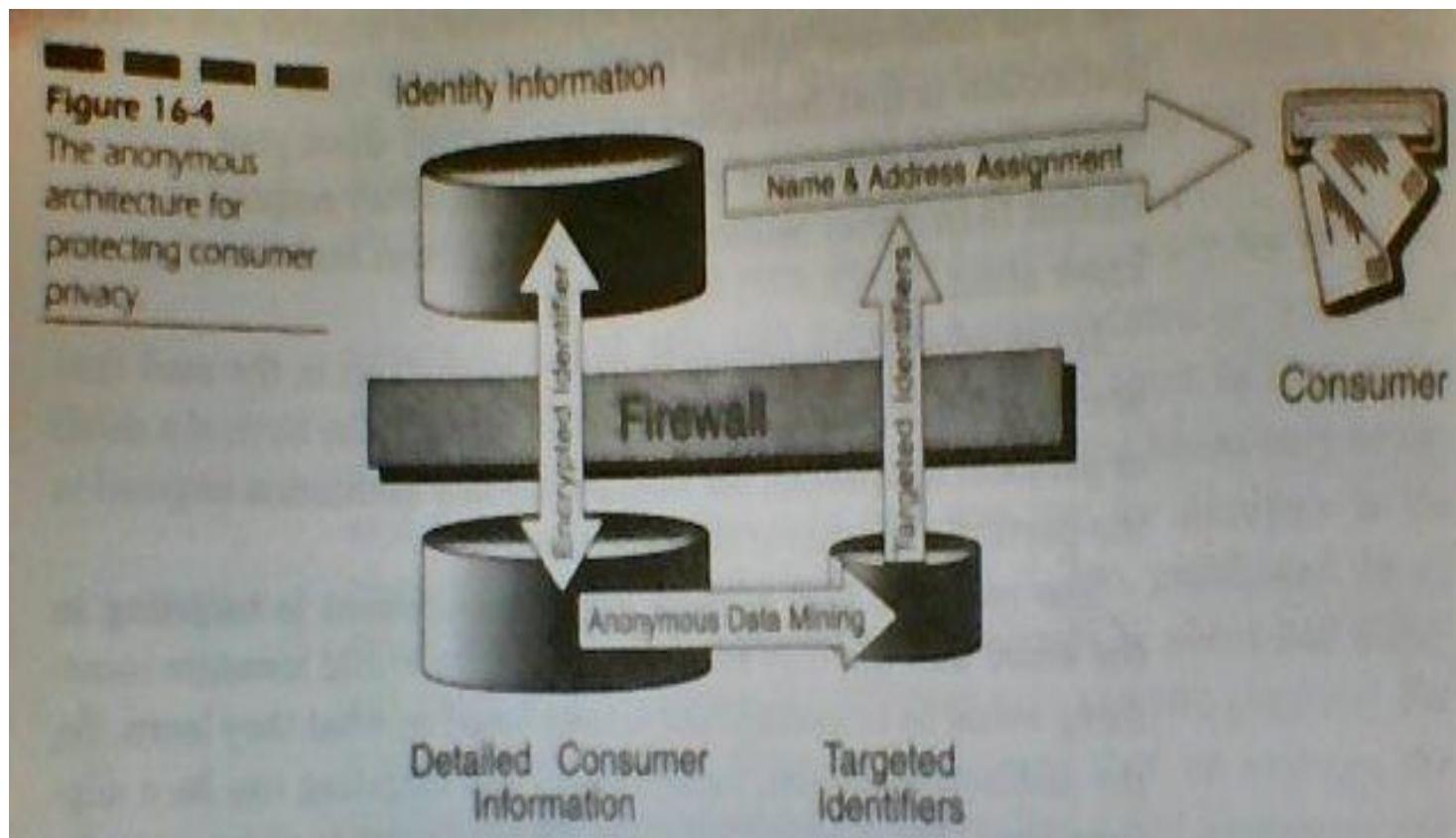
# *Guidelines for privacy*

- **Anonymity and identity information**
  - Distinguish between the data which give the identity of the customer and the data which does not provide that information.
  - We must abstract the customer data into an unique encrypted identifier.
  - Identifier must be random and independent of data about customer.
- **Detailed Vs consolidated data**
  - Even if the consumer data is anonymous, but it is still at the individual detailed level, there is a risk of customer being identified.
  - Eg: with Zip+4, age and type of car used, we can identify the customer.
  - For greater privacy, work only with data at a consolidated or rolled up level.

- **Information used for targeting or for measurement**
  - We must decide on using the customer data for targeting or only for the measurement.
  - Customers are less worried about people knowing what they are doing than the actions based on their information.
  - Measurement is also a level of targeting.
  - But measuring is a safer strategy.
- **Combined sources**
  - Now with the resources available, we can combine the data from various sources and build a complete picture of the customer.
  - This is done to identify the customer behavior for market and other products.

## *The anonymous architecture*

- This architecture has a firewall between the identity information and the consumer information.
- this provides privacy and helps the company to be out of legal issues.



## ***Legal issues associated with data mining***

- At almost every decision, there is a legal issue whenever a decision goes wrong.
- Eg : financial companies decisions on giving credit card to the customers.
- Now automatic process are available in the form of DM system to make sure the prohibited exclusions are not allowed.
- If you use DM on sensitive decision, make sure that there is no legal issue.
- There are many issues arising due to rapid development of the predictive models and large data.
- Care must be taken to make sure that the other variables are also considered for decisions.

## *Legal issues associated with data mining*

- We must check whether our model is both legal and desirable.
- Statistical profiling is done to check if the predictive model has excluded any group.
- We must also consider the regulatory environment and analyze the potential problems.
- In the end, competent legal counsel should review about the questionable data in a predictive model.