

# BI Project: Bank Marketing

---

Group 1  
Kim Yong Jin  
Park Jang Won  
Jeen Vern Liew



# AGENDA

**1**

---

Data Description

**2**

---

Data Mining Model (1): Classification

**3**

---

Data Mining Model (2): Clustering



# 1

## Data Description

1.1 Data Description

1.2 Background: Bank Telemarketing and Financial Crisis

## 1.1 Data Description

### Data Description

#### Data Description

- The data is related with the **telemarketing campaign of a Portuguese retail bank** to sell a term deposit
- Retail bank: a bank whose customer targets are individuals, SMEs rather than large corporations

#### Data Information

##### Time

- from May 2008 to November 2010

##### # of records

- 41178

##### # of attributes

- 22 including key attribute (ID) that we put in



## 1.1 Data Background

### Source

The screenshot shows the UCI Machine Learning Repository homepage. The header features the UCI logo and the text "Machine Learning Repository" and "Center for Machine Learning and Intelligent Systems". Below the header, there's a search bar and navigation links for "About", "Citation Policy", "Donate a Data Set", and "Contact". A prominent yellow button at the bottom right says "View ALL Data Sets".

#### Bank Marketing Data Set

[Download: Data Folder](#), [Data Set Description](#)

**Abstract:** The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Data Set Characteristics:	Multivariate	Number of Instances:	45211	Area:	Business
Attribute Characteristics:	Real	Number of Attributes:	17	Date Donated	2012-02-14
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	379515

#### Source:

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, Elsevier, 62:22-31, June 2014

#### Data Set Information:

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

There are four datasets:  
 1) bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]  
 2) bank-additional.csv with 10% of the examples (4119), randomly selected from 1), and 20 inputs.  
 3) bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).  
 4) bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).  
 The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

#### Attribute Information:

Input variables:  
 # bank client data:  
 1 - age (numeric)  
 2 - job : type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')  
 3 - marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)  
 4 - education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')

- UCI Machine Learning Repository: Bank Marketing Data Set

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>



## 1.1 Data Description

### Attribute Information

Key Variable

Client ID

Output variable

Has the client subscribed a term deposit? (binary: 'yes','no')

Input Variables

1

Client Data

2

Previous telemarketing data

3

Social and economic context attributes



## 1.1 Data Description

### Input Variables Break-down

1

Client Data

2

Previous telemarketing  
data

3

Social and economic  
context attribute

1. **age** (numeric)
2. **job** : type of job  
(categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
3. **marital** : marital status (categorical: 'divorced', 'married', 'single', 'unknown')
4. **education** (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
5. **default**: has credit in default? (categorical: 'no', 'yes', 'unknown')
6. **housing**: has housing loan? (categorical: 'no', 'yes', 'unknown')
7. **loan**: has personal loan? (categorical: 'no', 'yes', 'unknown')



## 1.1 Data Description

### Input Variables Break-down

1

Client Data

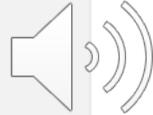
2

Previous telemarketing  
data

3

Social and economic  
context attribute

1. **contact:** contact communication type (categorical: 'cellular', 'telephone')
2. **month:** last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
3. **day\_of\_week:** last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
4. **duration:** last contact duration, in seconds (numeric)
5. **campaign:** number of contacts performed during this campaign and for this client (numeric)
6. **pdays:** number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
7. **previous:** number of contacts performed before this campaign and for this client (numeric)
8. **poutcome:** outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')



## 1.1 Data Description

### Input Variables Break-down

1

Client Data

2

Previous telemarketing  
data

3

Social and economic  
context attribute

1. **emp.var.rate**: employment variation rate - quarterly indicator (numeric)
2. **cons.price.idx**: consumer price index - monthly indicator (numeric)
3. **cons.conf.idx**: consumer confidence index - monthly indicator (numeric)
4. **euribor3m**: euribor 3 month rate - daily indicator (numeric)
5. **nr.employed**: number of employees - quarterly indicator (numeric)



## 1.1 Data Description

### Excel Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	ID	age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays	previous	poutcome	emp.var.ra	cons.price.con.s.conf.	euribor3m	nr.employ.y						
2	1	56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
3	2	57	services	married	high.school	unknown	no	no	telephone	may	mon	149	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
4	3	37	services	married	high.school	no	yes	no	telephone	may	mon	226	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
5	4	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
6	5	56	services	married	high.school	no	no	yes	telephone	may	mon	307	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
7	6	45	services	married	basic.9y	unknown	no	no	telephone	may	mon	198	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
8	7	59	admin.	married	profession	no	no	no	telephone	may	mon	139	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
9	8	41	blue-colla	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
10	9	24	technician	single	profession	no	yes	no	telephone	may	mon	380	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
11	10	25	services	single	high.school	no	yes	no	telephone	may	mon	50	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
12	11	41	blue-colla	married	unknown	unknown	no	no	telephone	may	mon	55	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
13	12	25	services	single	high.school	no	yes	no	telephone	may	mon	222	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
14	13	29	blue-colla	single	high.school	no	no	yes	telephone	may	mon	137	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
15	14	57	housemaid	divorced	basic.4y	no	yes	no	telephone	may	mon	293	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
16	15	35	blue-colla	married	basic.6y	no	yes	no	telephone	may	mon	146	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
17	16	54	retired	married	basic.9y	unknown	yes	yes	telephone	may	mon	174	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
18	17	35	blue-colla	married	basic.6y	no	yes	no	telephone	may	mon	312	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
19	18	46	blue-colla	married	basic.6y	unknown	yes	yes	telephone	may	mon	440	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
20	19	50	blue-colla	married	basic.9y	no	yes	yes	telephone	may	mon	353	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
21	20	39	managem	single	basic.9y	unknown	no	no	telephone	may	mon	195	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
22	21	30	unemploy	married	high.school	no	no	no	telephone	may	mon	38	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
23	22	55	blue-colla	married	basic.4y	unknown	yes	no	telephone	may	mon	262	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
24	23	55	retired	single	high.school	no	yes	no	telephone	may	mon	342	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
25	24	41	technician	single	high.school	no	yes	no	telephone	may	mon	181	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
26	25	37	admin.	married	high.school	no	yes	no	telephone	may	mon	172	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
27	26	35	technician	married	university	no	no	yes	telephone	may	mon	99	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
28	27	59	technician	married	unknown	no	yes	no	telephone	may	mon	93	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
29	28	39	self-emplc	married	basic.9y	unknown	no	no	telephone	may	mon	233	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
30	29	54	technician	single	university	unknown	no	no	telephone	may	mon	255	2	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
31	30	55	unknown	married	university	unknown	unknown	telephone	may	mon	362	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no					
32	31	46	admin.	married	unknown	no	no	no	telephone	may	mon	348	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
33	32	59	technician	married	unknown	no	yes	no	telephone	may	mon	386	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
34	33	49	blue-colla	married	unknown	no	no	no	telephone	may	mon	73	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
35	34	54	managem	married	basic.4y	unknown	yes	no	telephone	may	mon	230	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				
36	35	54	blue-colla	divorced	basic.4y	no	no	no	telephone	may	mon	208	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no				

## 1.2 Background

### What is Bank Telemarketing



The marketing of bank's products or services by means of telephone calls, typically unsolicited, to potential customers.

### Why Bank Telemarketing



After the 2008 Financial Crisis, banks try to attract more term deposits to recover its loss, so they use telemarketing as a way to reach out to potential customers.



## 1.2 Background

### Portugal during the 2008 Financial Crisis



- Before the Crisis: Cumulating government debt + lasting high unemployment+ high public spending+ low productivity
- GDP of Portugal contracted by 2.7% in 2008-2009 largely due to shrinking domestic demand
- Unemployment rose to 9.47% in 2009
- Budget deficit increased from 2.8% of GDP in 2008 to 9.4% in 2009

( Source: Euro Challenge 2012 )

Does this affect the banking choices made by Portuguese citizens?



# 2

## Data Mining Model (1): Classification

- 2.1 Classification Problem
- 2.2 Building Mining Model
- 2.3 Result
- 2.4 Profit Chart
- 2.5 Implication



## 2.1 Classification Problem

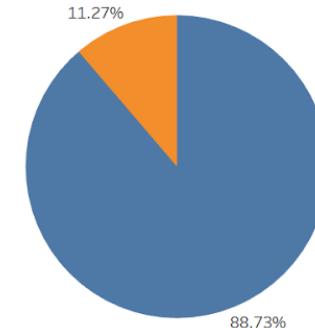
### Classification Problem

*“Whether the customer will positively respond to the telemarketing campaign or not”*

### Minimum Acceptable Performance Indicator (Accuracy-focused)

1 Random Probability = 50%

2



- Data Distribution: 'yes': 11.27% 'no': 88.73%
- Based on *Zero-R algorithm*,  
Prior Probability = 88.73%

$$\text{MAX}(50\%, 88.73\%) = 88.73\%$$

Minimum acceptable performance indicator = 88.73%



## 2.2 Building Mining Model

### Using NB

Bank\_Marketing\_Project - Microsoft Visual Studio

File Edit View Project Build Debug Team Database Mining Model Tools Test Analyze Window Help

Quick Launch (Ctrl+Q) Park Jang Won PW

Bank Marketing.dsv [Design] Classification.dmm [Design] Clustering.dmm [Design] NB.dmm [Design] ✎ Only\_Numerical\_DT.dmm [Design]

Mining Structure Mining Models Mining Model Viewer Mining Accuracy Chart Mining Model Prediction

Mining Model: NB Viewer: Microsoft Naïve Bayes Viewer

Dependency Network Attribute Profiles Attribute Characteristics Attribute Discrimination

All Links

Y

Cons#conf#idx  
Pdays  
Poutcome  
Euribor3m  
Previous  
Duration  
Marital  
Education  
Campaign  
Month  
Nr#employed  
Default  
Age  
Cons#price#idx  
Emp#var#rate  
Contact  
Job  
Cons#conf#idx

Select a node in the network to highlight its dependencies.

Strongest Links

- Selected node
- This node predicts the selected node
- Predicts both ways
- Selected node predicts this node

Solution Explorer

Search Solution Explorer (Ctrl+Shift+F)

Solution 'Bank\_Marketing\_Project' (

- Bank\_Marketing\_Project
  - Data Sources
    - Bank Marketing.ds
  - Data Source Views
    - Bank Marketing.dsv
  - Cubes
  - Dimensions
  - Mining Structures
    - Clustering.dmm
    - NB.dmm
    - Only\_Numerical\_DT.dmm

Properties

NB MiningStructure

ErrorConfiguration (default)  
Language

Basic

Description

ID NB  
Name NB

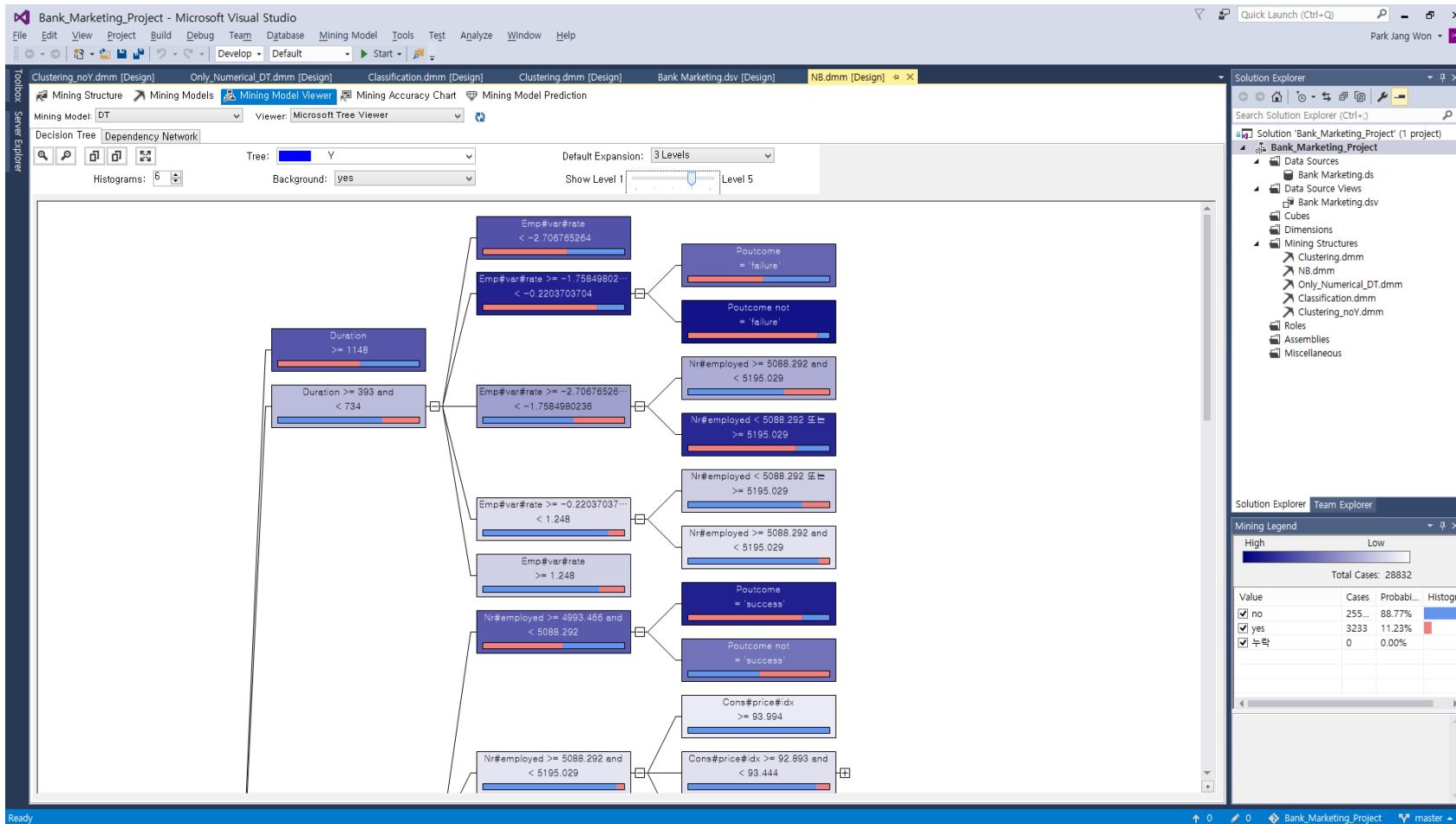
Name  
Specifies the name of the object.

Ready 0 5 Bank\_Marketing\_Project master



## 2.2 Building Mining Model

# Using DT



## 2.2 Building Mining Model

### Using ANN

The screenshot shows the Microsoft Visual Studio interface for a project named "Bank\_Marketing\_Project". The main window displays the "Classification.dmm [Design]" tab of the Mining Model Viewer. The "Input" section shows a table with columns "Attribute" and "Value", including rows for <All>, Month (oct), Duration (428.171 - 1,020.482), Marital (unknown), Month (sep), Housing (unknown), Duration (0.000 - 84.583), Pdays (396.196 - 834.641), Emp#var#rate (-3.400 - -0.983), Job (unemployed), Month (apr), Month (dec), Job (retired), Job (unknown), Job (student), Duration (256.377 - 428.171), and Duration (84.583 - 256.377). The "Output" section shows settings for Output Attribute (Y), Value 1 (yes), and Value 2 (no). Below these, a bar chart titled "Favors yes" vs "Favors no" visualizes the output probabilities for various input variables. The "Variables" table lists all input attributes and their corresponding values. The Solution Explorer on the right shows the project structure with files like "Bank Marketing.dsv", "Classification.dmm", "NB.dmm", and "Only\_Numerical\_DT.dmm". The Properties panel on the right shows the "Classification MiningStructure" properties, including Name (Classification) and Description (Classification).



## 2.3 Result

### Classification Matrix (focused on Accuracy)

Counts for NB on Y:

Predicted	yes (Actual)	no (Actual)
yes	837	1092
no	570	9857

Counts for DT on Y:

Predicted	yes (Actual)	no (Actual)
yes	670	340
no	737	10609

Counts for NN on Y:

Predicted	yes (Actual)	no (Actual)
yes	823	871
no	584	10078

$$\text{NB Accuracy} = (837+9857) / (837+1092+570+9857) \\ = 86.55\%$$

$$\text{DT Accuracy} = (670+10609) / (670+340+737+10609) \\ = 91.28\%$$

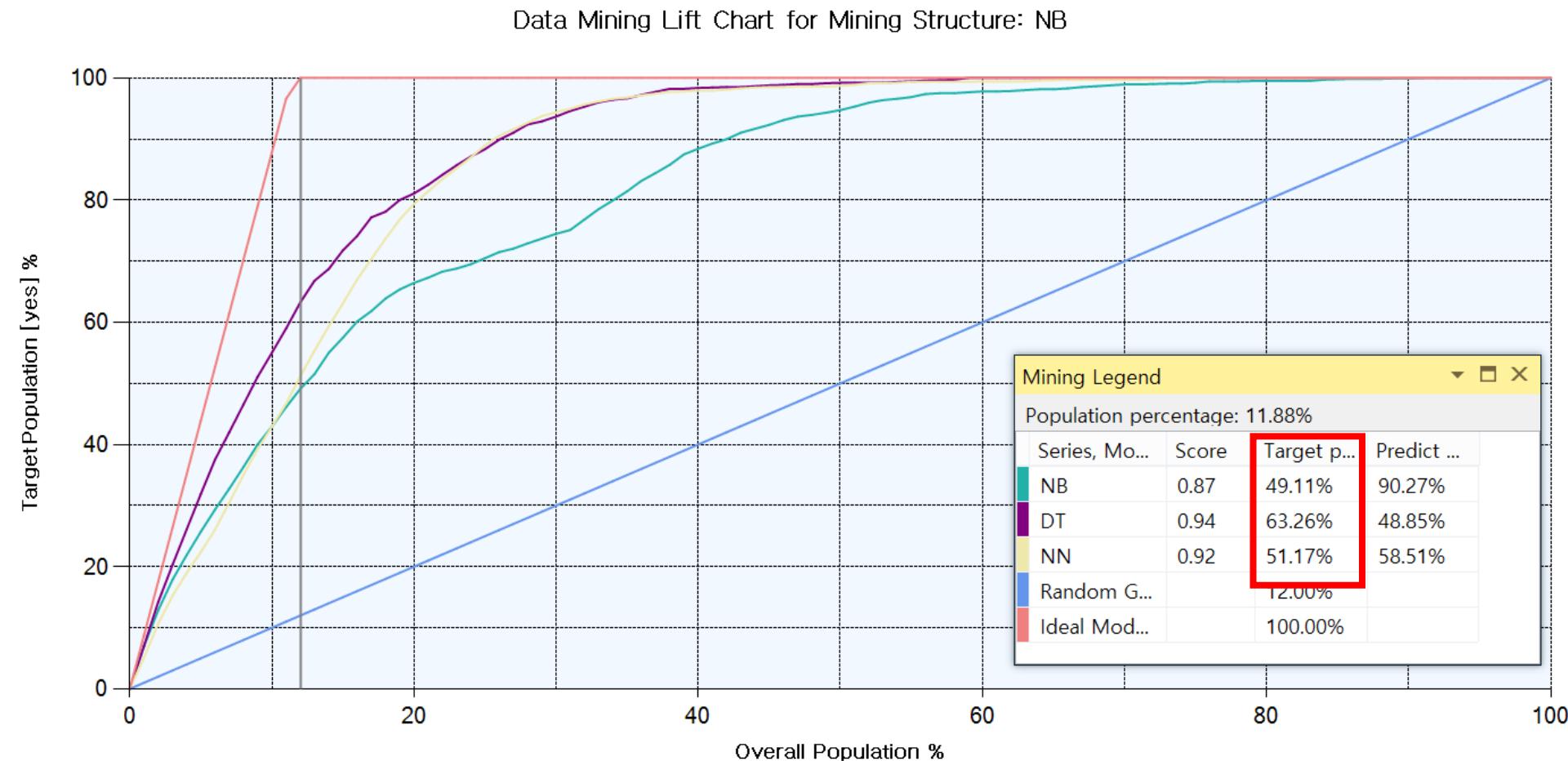
$$\text{ANN Accuracy} = (823+10078) / (823+871+584+10078) \\ = 88.22\%$$

Highest Accuracy : Decision Tree  
Lowest Accuracy : Naïve Bayesian



## 2.3 Result

### Lift Chart



## 2.3 Result

Q1. Is the difference 3% meaningful?

DT Accuracy = 91.28%  
Accuracy based on Prior Probability = 88.73%

Difference = 91.28% - 88.73% ≈ 3%

*“Is this difference really meaningful?”*

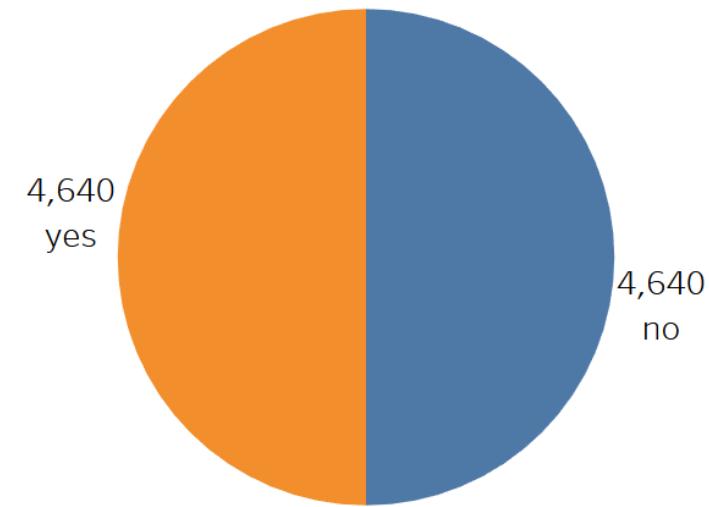


## 2.3 Result

### Verification by Undersampling Technique

New Dataset with same proportion of “yes” and “no” => “ Undersampling”

- Yes : No = 50 : 50
- Keep the data with “yes”
- Randomly choose 4,640 data out of 36,548 data with “no”



## 2.3 Result

### Conclusion

Counts for DT50to50 on Y:

Predicted	yes (Actual)	no (Actual)
yes	1304	304
no	80	1096

- Accuracy based on prior probabilities = 50%
- DT Accuracy =  $(1304+1096) / (1304+304+80+1096) = 86.21\%$
- $86.21\% - 50\% = \underline{36.21 \%}$

Decision Tree is still a highly predictive and accurate tool in this data set



## 2.3 Result

### Q2. Why is the Decision Tree better than Artificial Neural Network?

- ANN shows higher accuracy with continuous data
- In this data set, there are 10 categorical data attribute and 10 continuous data attribute

Verify through new dataset whose categorical data are excluded



## 2.3 Result

### Verification by new data set

Counts for Only\_Numerical\_DT on Y:

Predicted	yes (Actual)	no (Actual)
yes	533	275
no	825	10723

Counts for Only\_Numerical\_ANN on Y:

Predicted	yes (Actual)	no (Actual)
yes	283	190
no	1075	10808

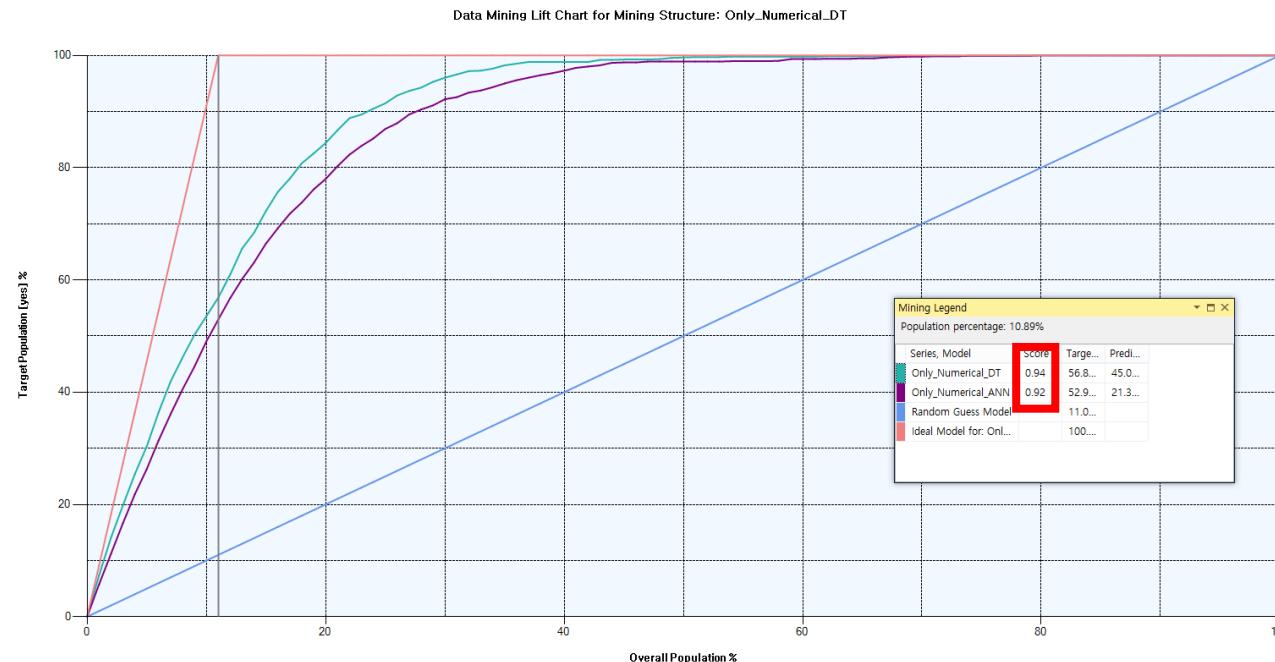
	DT	ANN
Original Data Set	91.28%	88.22%
Only continuous	91.25%	89.76%
Difference	-0.03%	1.54%

Still, the Decision Tree is better than Artificial Neural Network...



## 2.3 Result

### Conclusion



Still, DT is better than ANN  
Why? Probable reasons:  
- continuous data not dispersed enough  
- Replacing them with 'Discretized' does not make big difference



## 2.4 Profit Chart

### Basic Assumption



[Telemarketing Campaign]

- Duration: Telemarketing for 3 month
- Execution Method : Outsourcing the Telemarketing-specialized company

### Population Assumption

#### Population

Portugal Total Population	10,000,000
Above 25	73.1%
	7,310,000

### Revenue Assumption

#### Revenue

Avg GDP/person	\$ 24,000.00
Avg household Saving Rate	4%
Avg saving amount/person	\$ 960.00
Net Interest Margin	3%



## 2.4 Profit Chart

### Fixed Cost Assumption

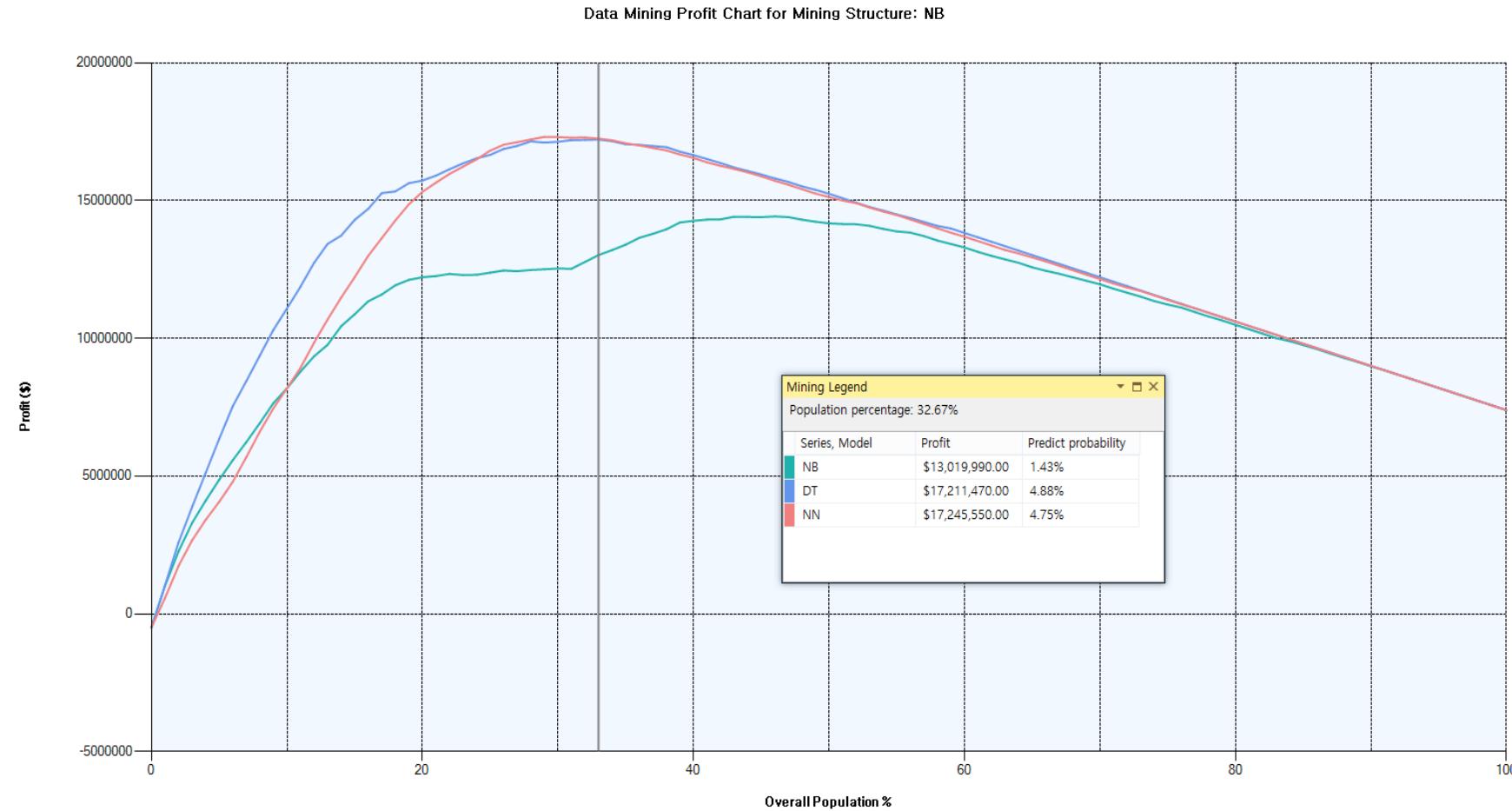
Fixed Cost	Price	Quantity	
Training Cost	\$ 322.72	1	\$ 322.72
Campaign Management(3 month total)	\$ 64.54	3	\$ 193.62
Marketing Gift(First-Come-First-Served)	\$ 5.00	100,000	\$ 500,000.00
			\$ 500,516.34

### Variable Cost Assumption

Variable Cost			
Avg Agent Cost/hour		\$	15.00
Avg Duration(min)/person			4
Avg Contact person/hour			15
Avg Agent Cost/person		\$	1.00
Avg Cost of Calls/person	\$ 0.30	4 \$	1.20
		\$	2.20



## 2.4 Profit Chart



Should target 32.67% of the population



## 2.4 Implication

Which factor is the most influential?

1

Deposit tendency is substantially correlated with Macro Factor

- When 1) Interest rate, 2) Number of employed is low and 3) variation of employment is -, it indicates economic depression and people tend to increase deposit
- When interest rate is low, we usually think that people will decrease deposit but actually, in this data, people tend to increase deposit ← “psychological factor”

2

Deposit tendency is highly correlated with Previous telemarketing data

- “Duration” was the most powerful factor when it comes to predict the outcome

3

Deposit tendency has lowest correlation with Client Data

- Among client data, only default attribute was correlated
- It indicates that marketing campaign afterwards do not have to consider much about client data and this fact can contribute to lower search cost



## 2.4 Implication

### Other Findings

1

#### Months: October > March > December

- Since financial crisis arose around September (Bankruptcy of Lehman Brothers: 9/15/2008), it is likely that banks tried more to reach out customers right away
- We also hypothesized that since March and December is end of each quarter it is likely that “push sales” happened; since it is common practice in sales & marketing division to push sales at the end of quarter

### Limitation

- Since the company has less control over macro economic factors and the duration of the call, even if they know the result, there is no room for dealing with it
- Therefore, needs for clustering arises → Find out characteristics of customer groups and utilize them for marketing



# 3

## Data Mining Model (2): Clustering

3.1 Objective

3.2 Building Model

3.3 Result & Implication



## 3.1 Objective

### Clustering Objective



Customer Segmentation



Identify

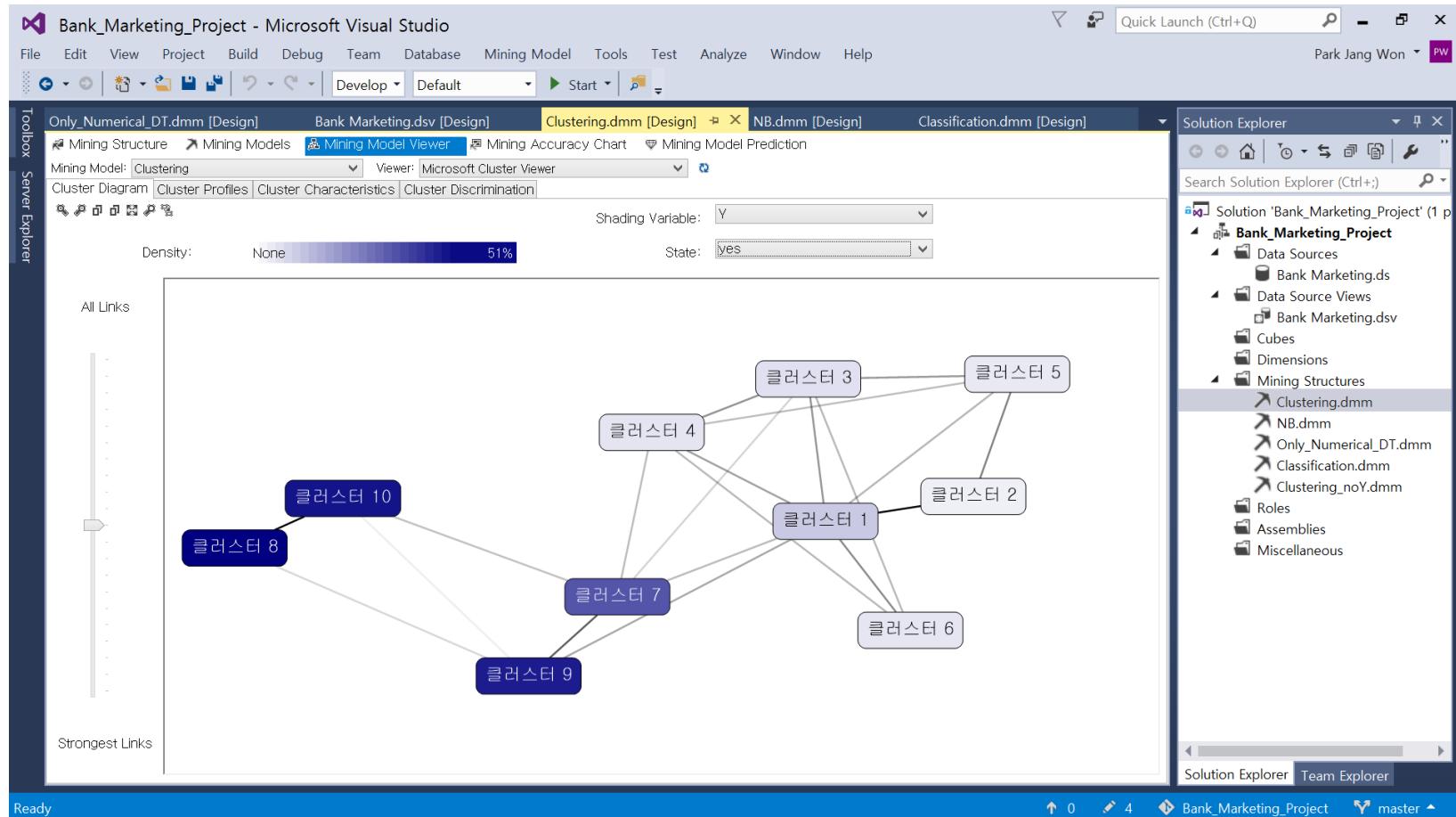


Target

- Identify the most unique characteristics of the clusters with the highest Yes ratio
- Design marketing strategies that concentrate on these targeted clusters



## 3.2 Building the Model



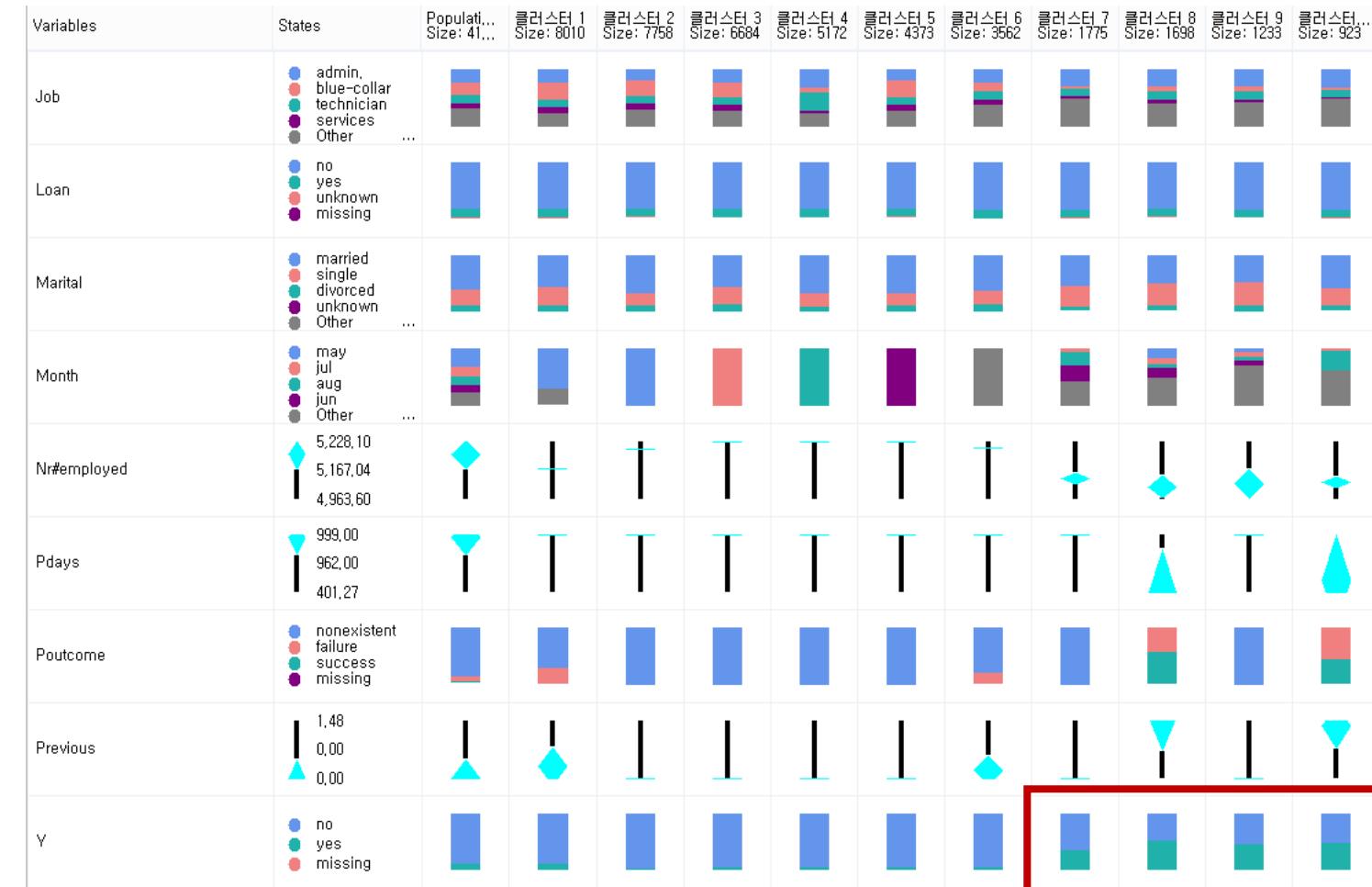
When we set the CLUSTER\_COUNT parameter to 0, 10 clusters were formed.



### 3.3 Results & Implication

1

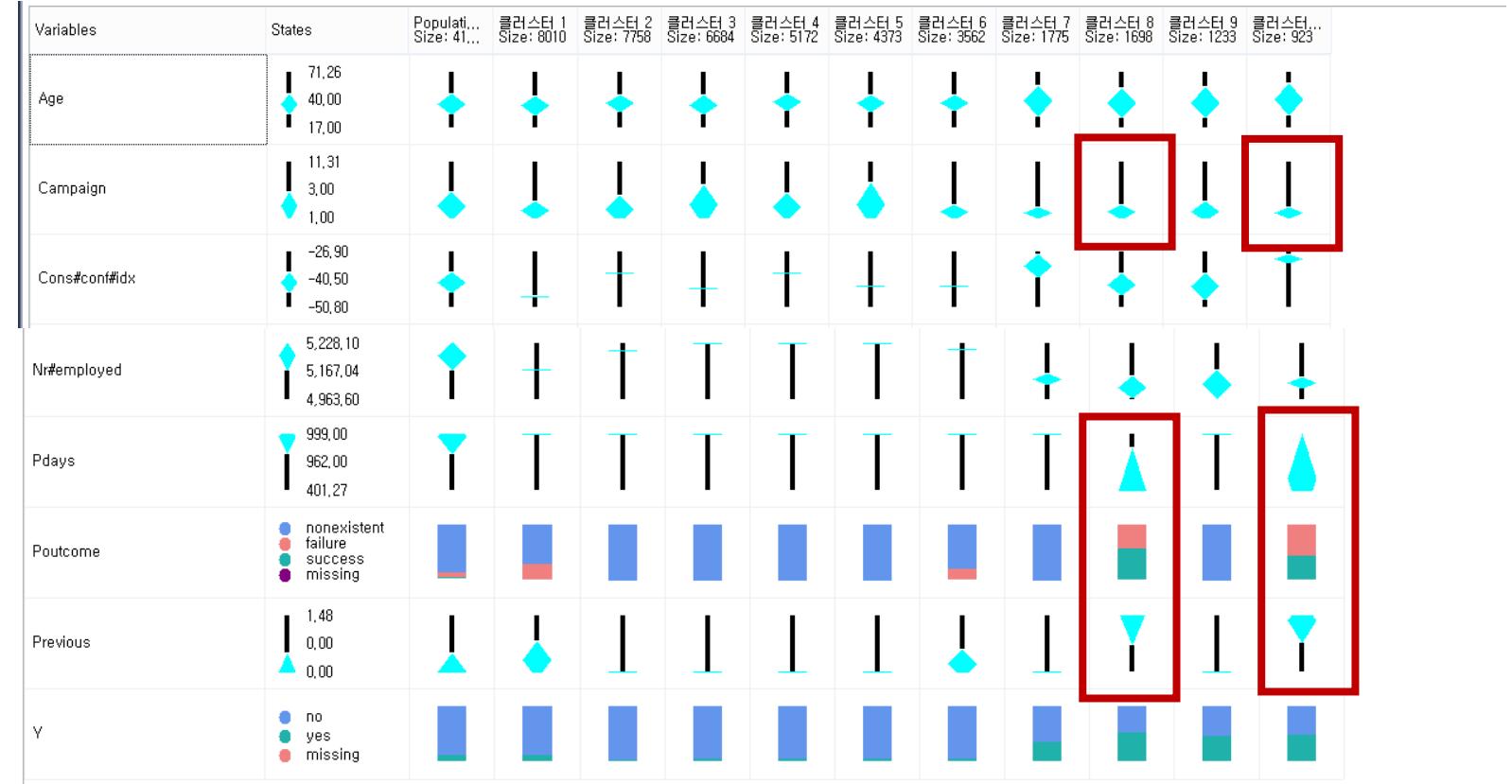
- Cluster 7, 8, 9, 10 has highest Yes ratio among the clusters, but there are still slight discrepancies among the clusters.



### 3.3 Results & Implication

1

- For further grouping based on similarities, can be paired as Cluster 7 & 9, and Cluster 8 & 10



Cluster 8 & 10 have higher values in attributes: campaign, pdays, previous, poutcome



### 3.3 Results & Implication

2

- Based only on the results of classification, we may fail to identify cluster-specific discrepancies → less efficient marketing

3

- From results of clustering, we can customize telemarketing strategy to attract more customers from Cluster 7,8,9,10
- Eg.: Contact customers in months that reflect higher Yes rate



## After Project..

- Classification and Clustering are complementary
- Classification is used as a tool to study the factors in a broad view
- Clustering extracts the customer segments' characteristic in a micro-perspective, helping to set practical action plans for the firm.



Thank you!  
Людмила



---

## Appendix

1. Assumption for Profit Chart
2. Prediction
3. YouTube Link
4. Peer Evaluation Table

# 1. Assumption for Profit Chart

Charge	What is it?	Expected Cost
Hourly call rate	Primarily the cost of wages of the staff employed to make the calls.	£10-£30 per hour, per operative
Set up fee	These are the expenses involved in setting up your campaign, such as training staff to make calls for your particular product or service (during which time they are unavailable for other campaigns), writing a script and refining the sales pitch.	£250
Cost of calls	Will usually be rolled into the hourly call rate but there may be a surcharge for international calls.	Depends on location and targeting
Cost of data	If the telemarketing company needs to source the lists of names and numbers used in the campaign then you'll pay an extra fee for this.	Varies by project
Campaign Management	Admin and managerial fees for setting up and overseeing your campaign	£50 per month

## Age structure:

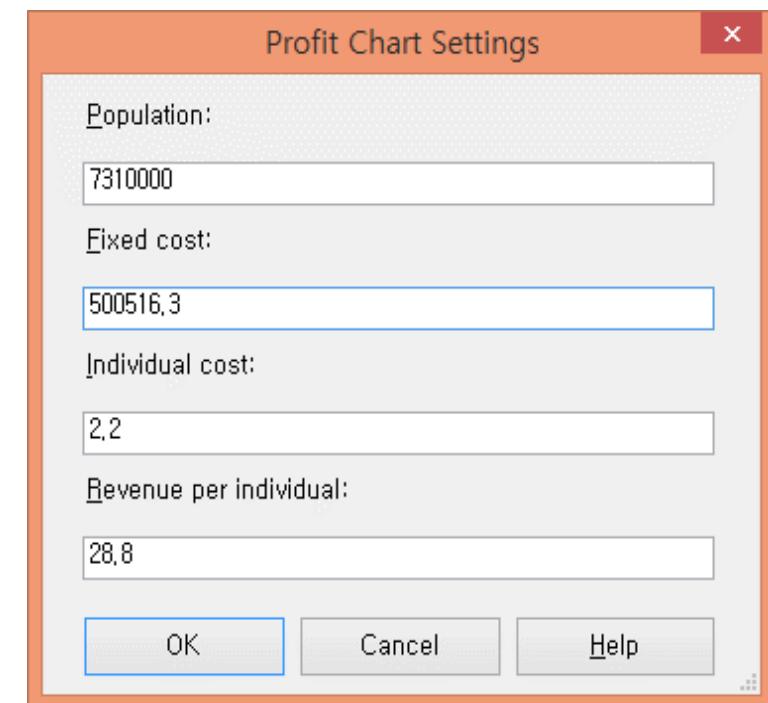
**0-14 years:** 15.5% (male 874,807/female 804,483)

**15-24 years:** 11.4% (male 655,234/female 579,669)

**25-54 years:** 41.88% (male 2,300,872/female 2,236,077)

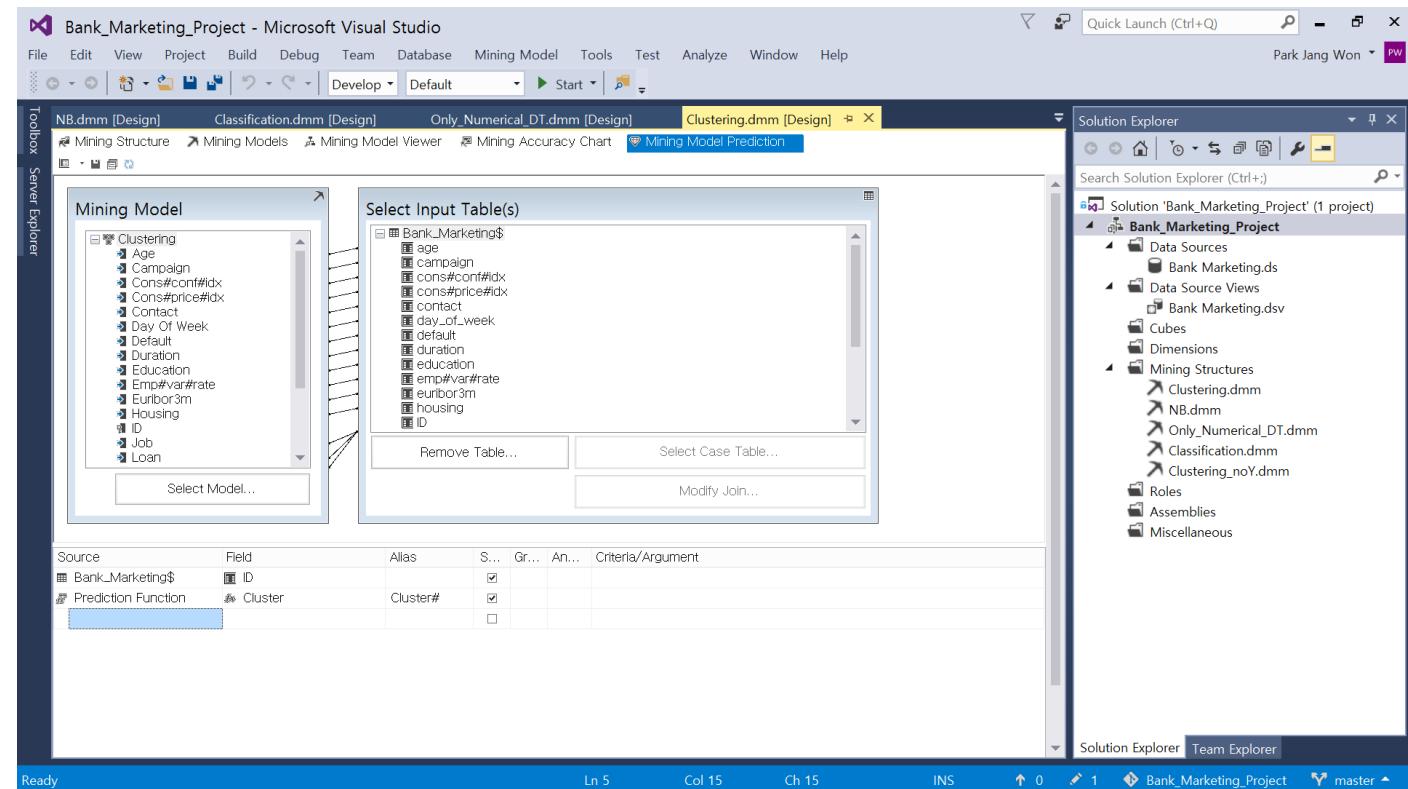
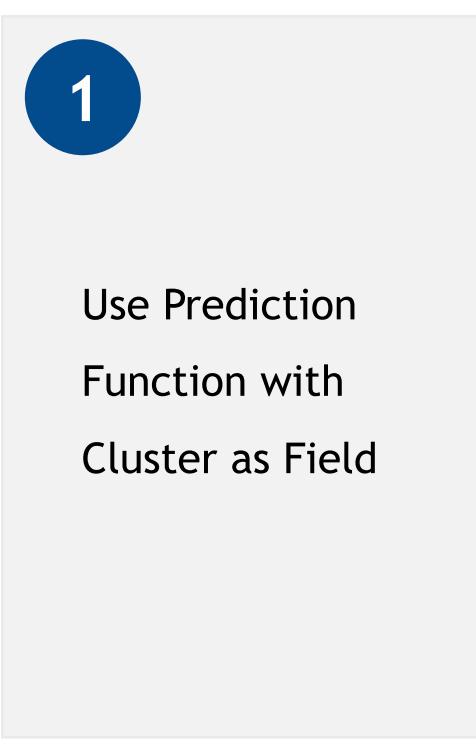
**55-64 years:** 12.07% (male 610,886/female 697,287)

**65 years and over:** 19.15% (male 849,506/female 1,224,995) (2016 est.)



## 2. Mining Model Prediction

### Predicting Which Cluster Customer Belongs to



## 2. Mining Model Prediction

### Predicting Which Cluster Customer Belongs to

2

Prediction Results in Query View

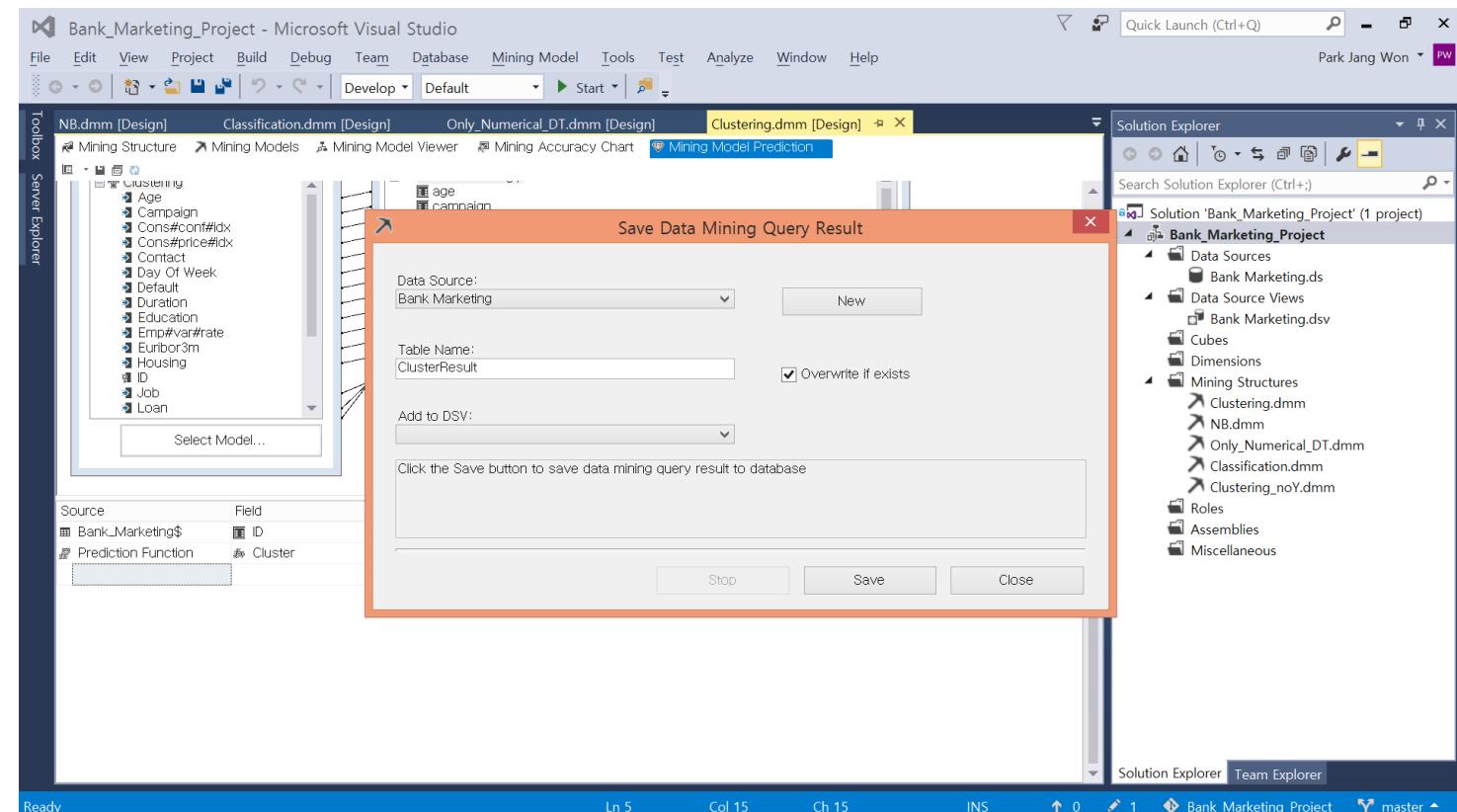
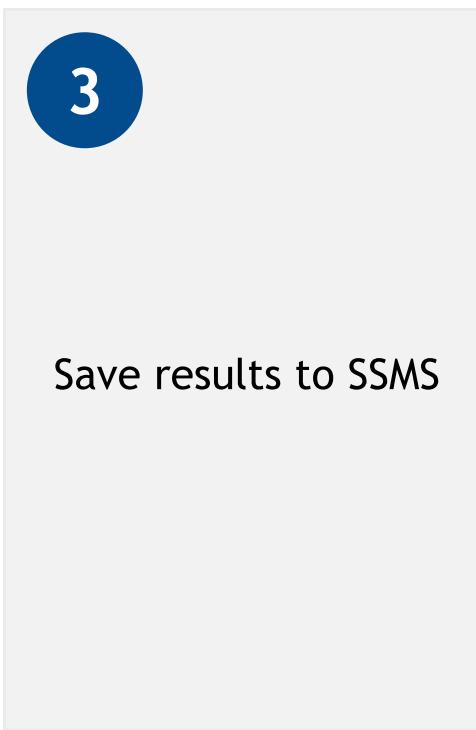
ID	Cluster#
4839	클러스터 2
4840	클러스터 2
4841	클러스터 2
4842	클러스터 2
4843	클러스터 2
4844	클러스터 2
4845	클러스터 2
4846	클러스터 2
4847	클러스터 2
4848	클러스터 2
4849	클러스터 2
4850	클러스터 2
4851	클러스터 2
4852	클러스터 2
4853	클러스터 2
4854	클러스터 2
4855	클러스터 2
4856	클러스터 2
4857	클러스터 2
4858	클러스터 2
4859	클러스터 2
4860	클러스터 2
4861	클러스터 2
4862	클러스터 2
4863	클러스터 2
4864	클러스터 2

Query run completed with 41188 rows fetched

Solution Explorer Team Explorer

## 2. Mining Model Prediction

### Predicting Which Cluster Customer Belongs to



## 2. Mining Model Prediction

### Predicting Which Cluster Customer Belongs to

4

Saved as Table  
'ClusterResult' in  
SSMS

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. In the Object Explorer, the database 'Bank\_Marketing' is selected, and its tables are listed, including 'ClusterResult'. In the center pane, a query window displays a SELECT statement from the 'ClusterResult' table. The results grid shows 11,188 rows, each containing an 'ID' and a 'Cluster#' value, all of which are '클러스터 2' (Cluster 2). A message at the bottom of the results grid indicates that the query was executed successfully.

ID	Cluster#
1	클러스터 2
2	클러스터 2
3	클러스터 2
4	클러스터 2
5	클러스터 2
6	클러스터 2
7	클러스터 2
8	클러스터 2
9	클러스터 2
10	클러스터 2
11	클러스터 2
12	클러스터 2
13	클러스터 2
14	클러스터 2
15	클러스터 2
16	클러스터 2
17	클러스터 2
...	...
11188	클러스터 2

## 2. Mining Model Prediction

### Predicting Which Cluster Customer Belongs to



ClusterResult.xlsx - Excel

ID	Cluster#
38259	24014 클리스터 8
38260	24020 클리스터 8
38261	24109 클리스터 8
38262	24265 클리스터 8
38263	24280 클리스터 8
38264	24398 클리스터 8
38265	24483 클리스터 8
38266	24614 클리스터 8
38267	24800 클리스터 8
38268	24851 클리스터 8
38269	24911 클리스터 8
38270	25054 클리스터 8
38271	25275 클리스터 8
38272	25294 클리스터 8
38273	25455 클리스터 8
38274	25498 클리스터 8
38275	25505 클리스터 8
38276	25520 클리스터 8
38277	25704 클리스터 8

준비 41188개 중 1698개의 레코드가 있습니다.

### 3. YouTube Link

---

Click Here:

<https://www.youtube.com/watch?v=-qhtlmt6k08&feature=youtu.be>

## 4. Peer Evaluation Table

Jeen Vern Liew	김용진	박장원
100%	100%	100%