

Marketing Intelligence Application

Yosuke Katada

2014/5/6

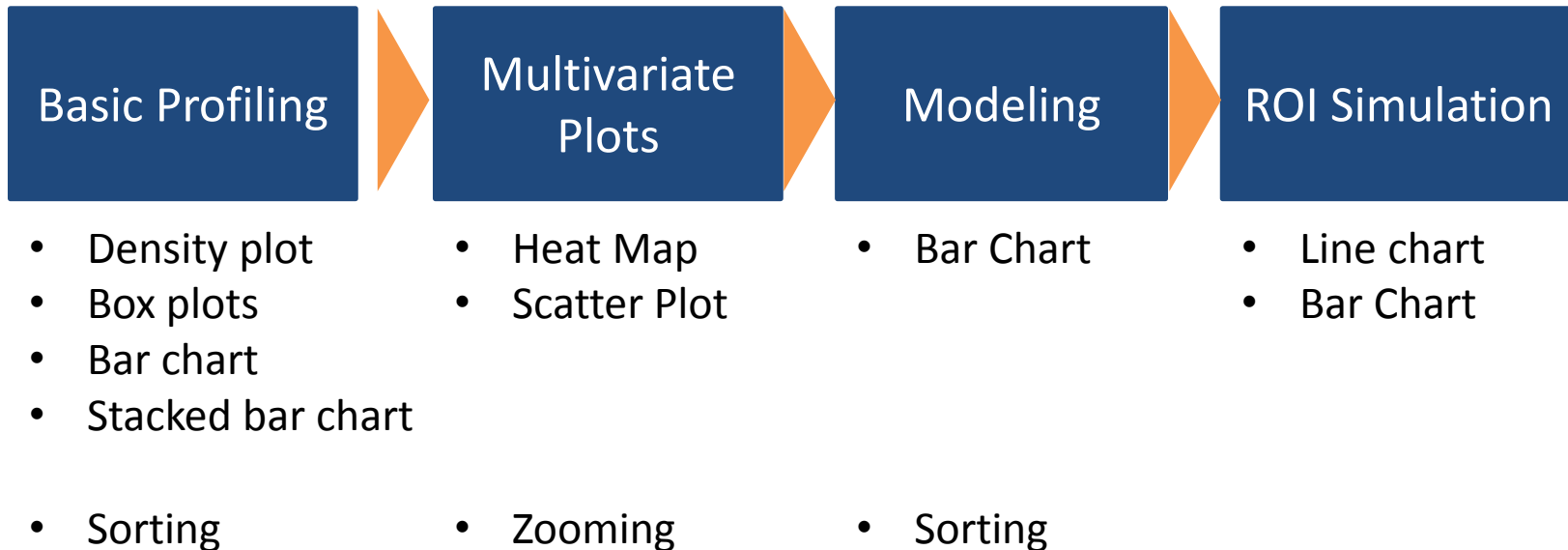
Data

- Marketing campaigns' results about **whether a customer subscribed term deposit or not** at a Portuguese bank.
- 45,211 rows and 17 variables
 - Each row represents each customer response
 - Dependent variable: whether a customer subscribed term deposit or not
 - Independent variable: 7 numeric variables /9 categorical variables

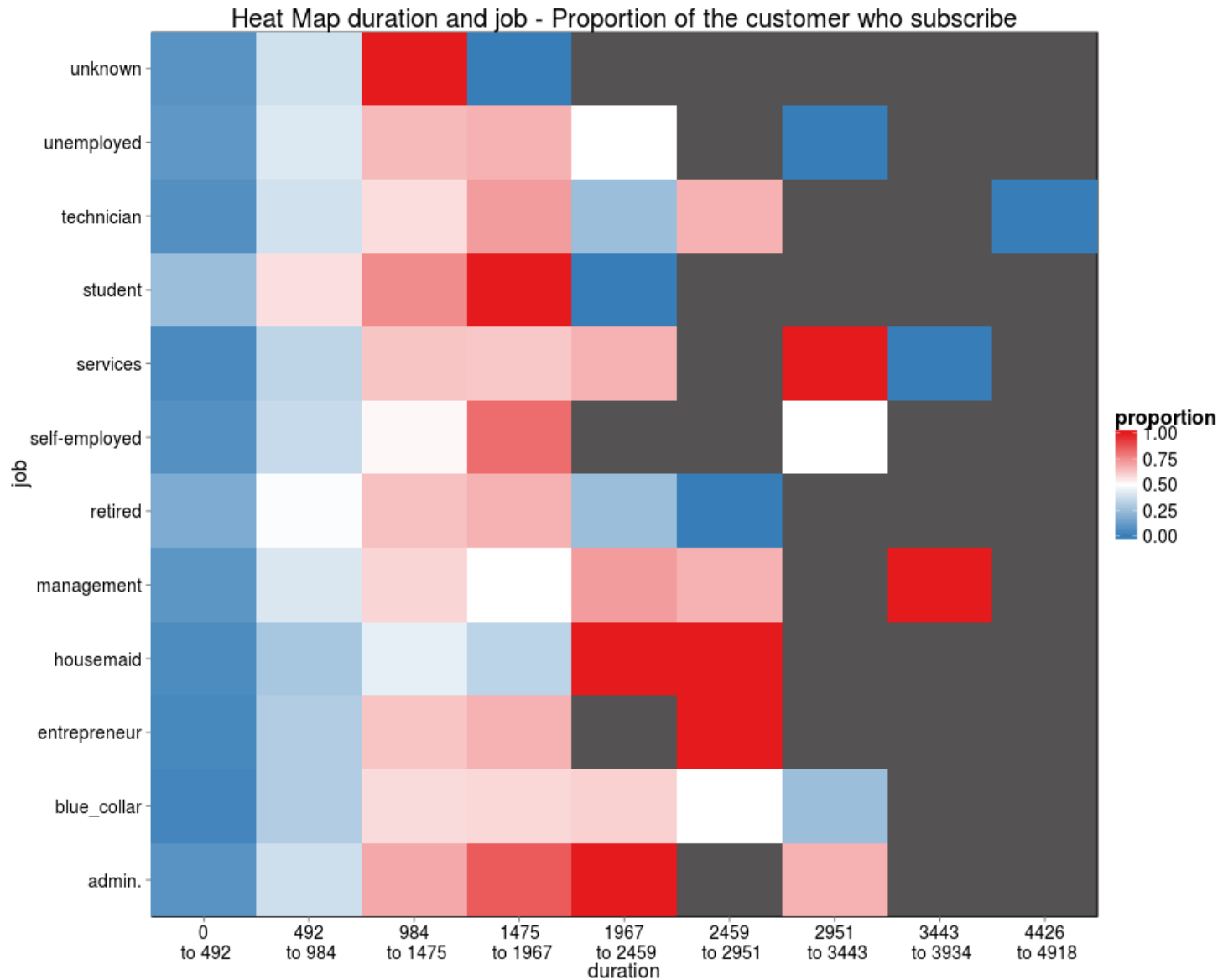
45,211 observations of 17 variables																	
	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
1	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
2	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
3	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
4	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
5	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
6	35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
7	28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
8	42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	0	unknown	no
9	58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	0	unknown	no
10	43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	0	unknown	no

Application components

- Target user: Marketing analytics personnel at the bank
- Target audience: Business managers
- Simple and easy plots



Demo1 – Heat Map



Demo2 – Logistic Regression

Bank Marketing Data Analysis

Basic Profiling

Heat Map

Scatter Plot

Logistic Regression

ROI Simulation

Independent Variable:

- ☒ age
- ☐ job
- ☐ marital
- ☐ education
- ☐ default
- ☒ balance
- ☒ housing
- ☒ loan
- ☐ contact
- ☒ day
- ☐ month
- ☒ duration
- ☐ campaign
- ☐ pdays
- ☐ previous
- ☒ poutcome

Training Size

Probability threshold

Sort Key:

Coefficient

Sorting

- ☒ decreasing
- ☐ increasing

Run

Model

Standardized Coefficients

Logistic Regression

```
Call:
glm(formula = subscribed ~ ., family = "binomial", data = df[idx,
])
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.4860  -0.4002  -0.3023  -0.2110   3.0025
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.111e+00  1.017e-01 -20.765  < 2e-16 ***
age           -5.421e-03  1.732e-03  -3.129  0.00175 **
balance       2.431e-05  5.367e-06   4.529  5.92e-06 ***
housingyes    -1.044e+00  4.111e-02 -25.392  < 2e-16 ***
loanyes       -5.793e-01  6.290e-02  -9.210  < 2e-16 ***
day           -6.156e-03  2.332e-03  -2.640  0.00830 **
duration      3.895e-03  6.687e-05  58.245  < 2e-16 ***
poutcomeother 2.563e-01  9.447e-02   2.713  0.00667 **
poutcomesuccess 2.307e+00  8.450e-02  27.298  < 2e-16 ***
poutcomeunknown -7.284e-01  5.857e-02 -12.436  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 26120  on 36167  degrees of freedom
Residual deviance: 18944  on 36158  degrees of freedom
AIC: 18964
```

Number of Fisher Scoring iterations: 6

Confusion Matrix

	y_pred	
y_actual	no	yes
no	7804	185
yes	730	324

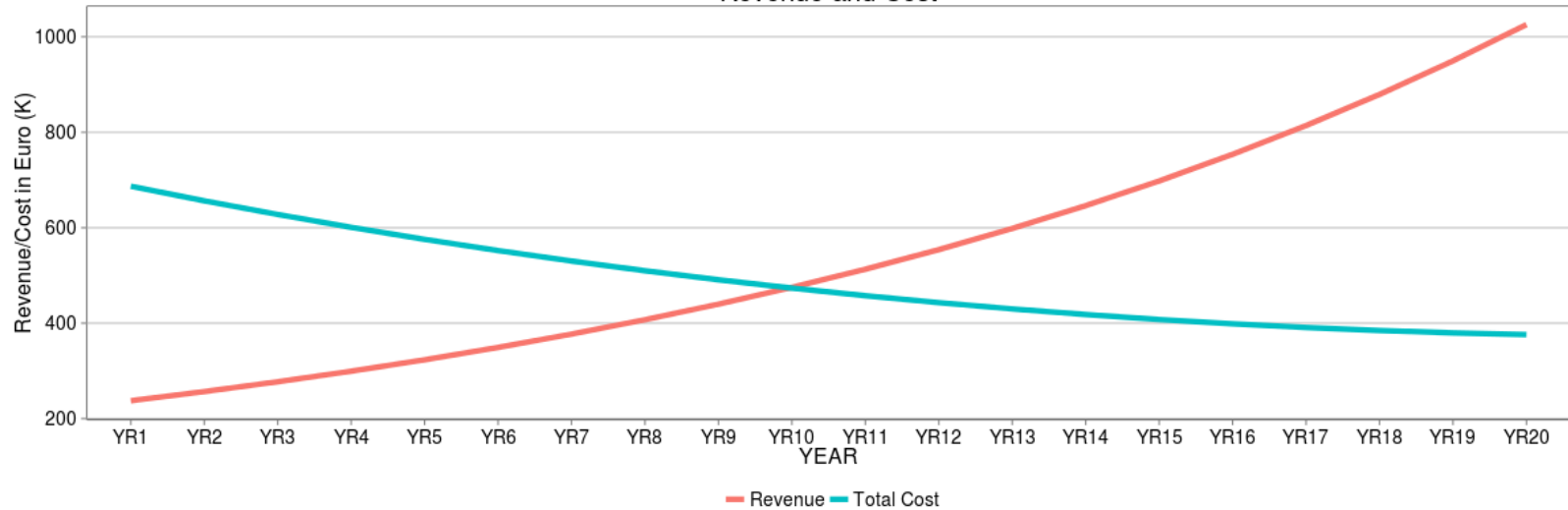
Accuracy Report

	Accuracy	Precision	Recall	F1_Score
1	0.90	0.64	0.31	0.41

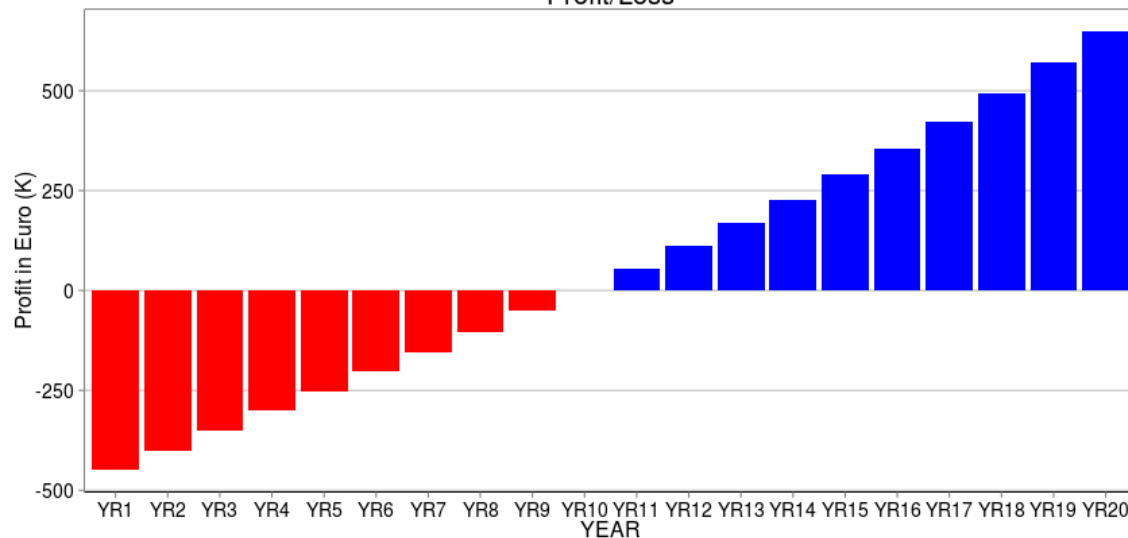
Demo3 – ROI Simulation

ROI Simulation

Revenue and Cost



Profit/Loss



ROI
(PV vs Campaign Cost)

