

2015 *International Summer Semester Data Science and Management*

Factors deciding Student Performance

TEAM BBangya

Hansol_Kim/Youtaeck_Jeon/Youngha_Lee/Hyoshin_Soh

1 Motivation and
Background

2 Research
Question

3 Methods and
Data

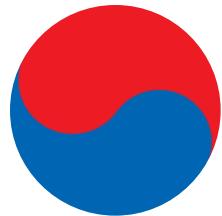
4 Results and
Interpretation

5 Conclusion and
Future Work

Factors deciding Student Performance

1

Motivation and Background



Environmental Factors



Gangnam

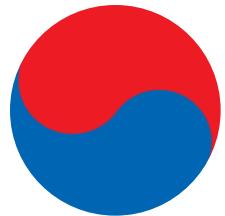


Portuguese student performance



1

Motivation and Background



Environmental Factors



Gangnam



Portuguese student performance



2

Research Question

Which factor affects
students' performance the best?

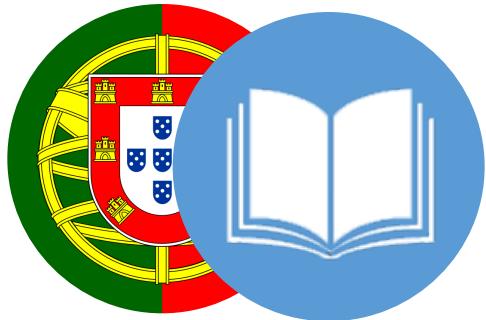
Can We Predict Korean Students'
Performance with This Data?



3

Methods and Data

DATA | Portuguese student performance

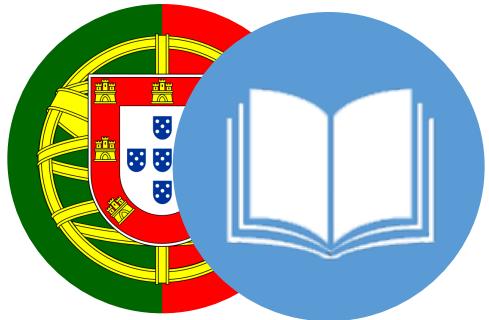


Portuguese

Mathematics

Methods and Data

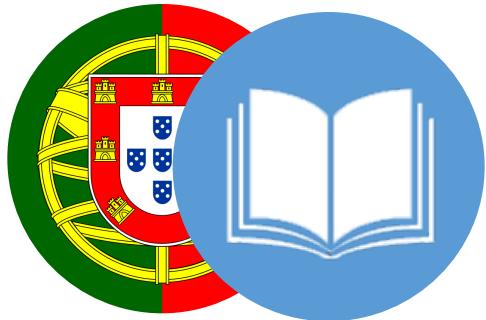
DATA | Portuguese student performance



“ Can we predict
Korean students'
with this data? ”

Methods and Data

DATA | Portuguese student performance



G3

Final grade

Failure

The number of past class failures

Higher

Willingness to take higher education

Sex

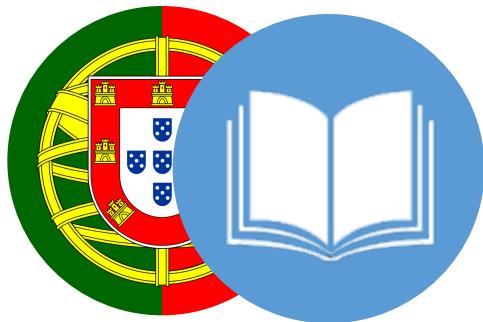
Whether male or female

Paid

Extra educational support from home

Methods and Data

DATA | Portuguese student performance



G3

Final grade

Failure

The number of past class failures

Absence

The number of school absences

Romantic

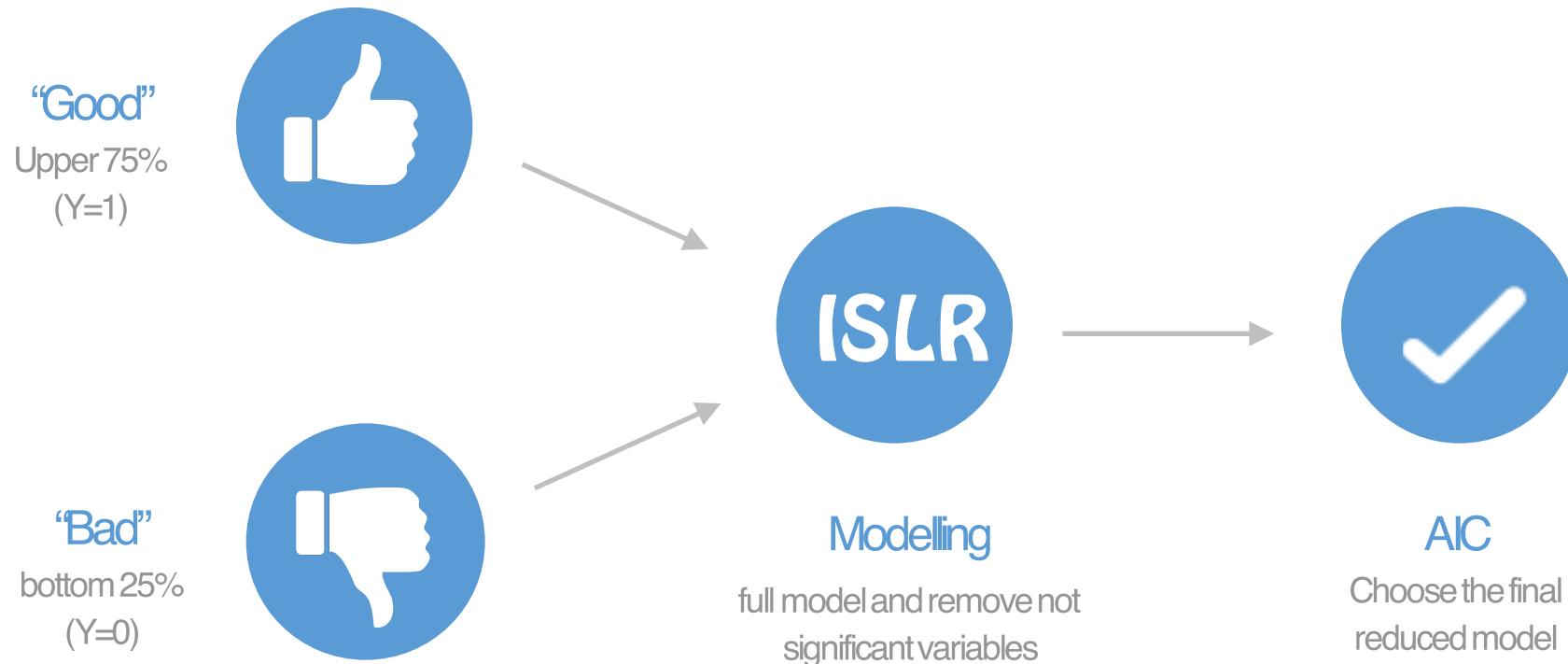
Whether a student is seeing someone or not

SchoolSup

Extra educational support

Methods and Data

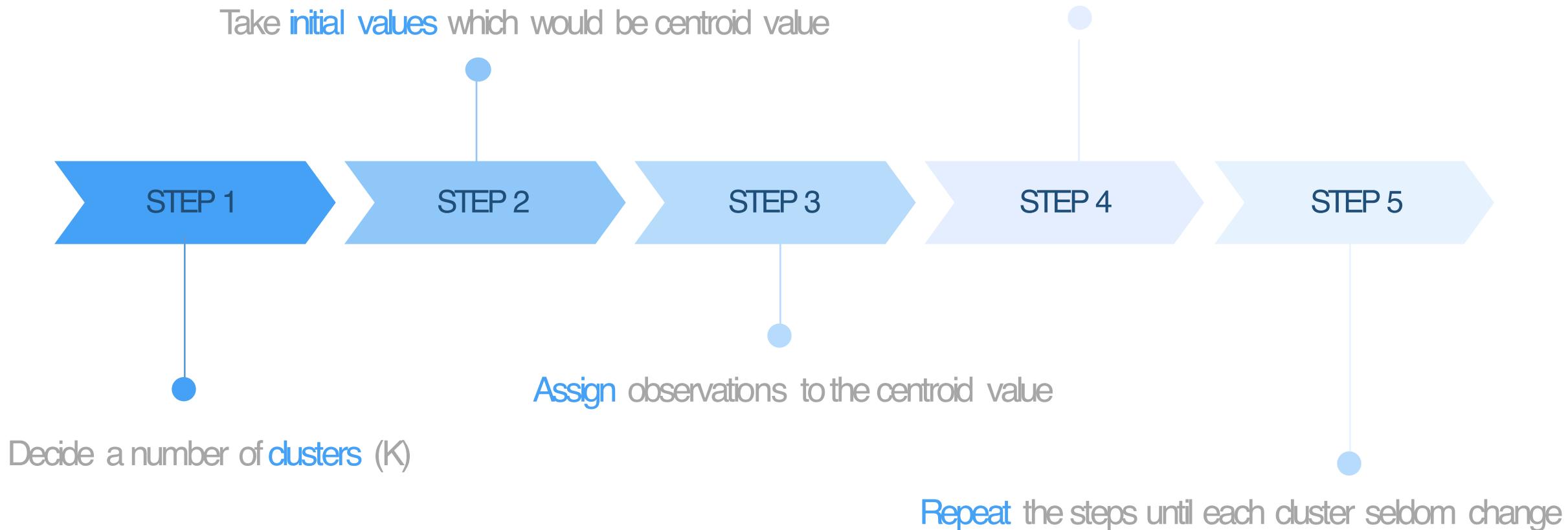
Logistic Regression



3

Methods and Data

Clustering Analysis (K-means Method)



Results and Interpretation

Logistic Regression

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.8388	0.2852	2.941	0.00328	**
failures	-1.0694	0.1691	-6.324	2.54e-10	***
higheryes	1.4774	0.3020	4.892	9.99e-07	***

Signif. codes:	0	'***'	0.001	'**'	0.01
			'*' 0.05	.	' 0

(Dispersion parameter for binomial family taken to be

Null deviance: 557.79 on 648 degrees of freedom
Residual deviance: 467.38 on 646 degrees of freedom
AIC: 473.38

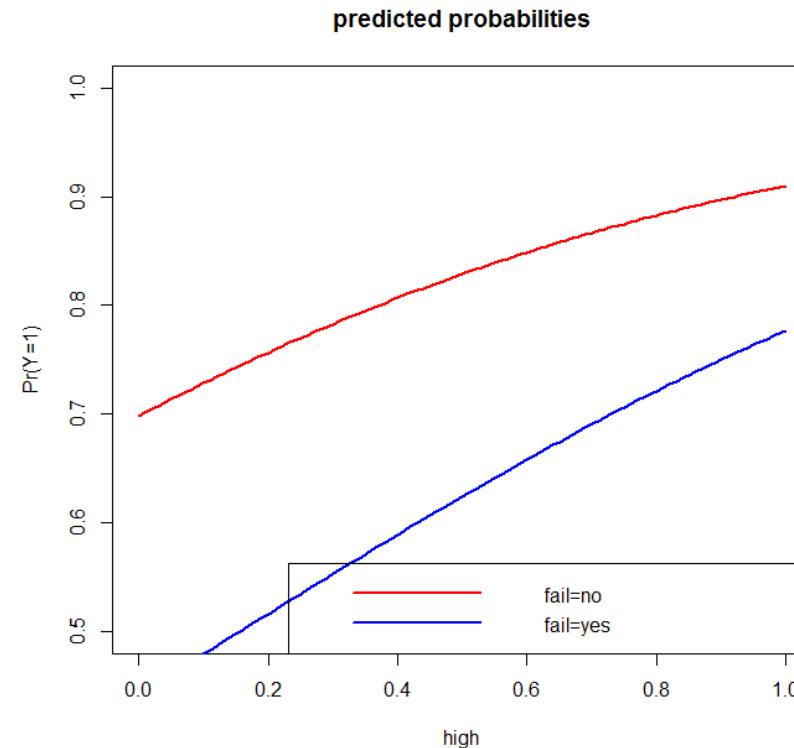
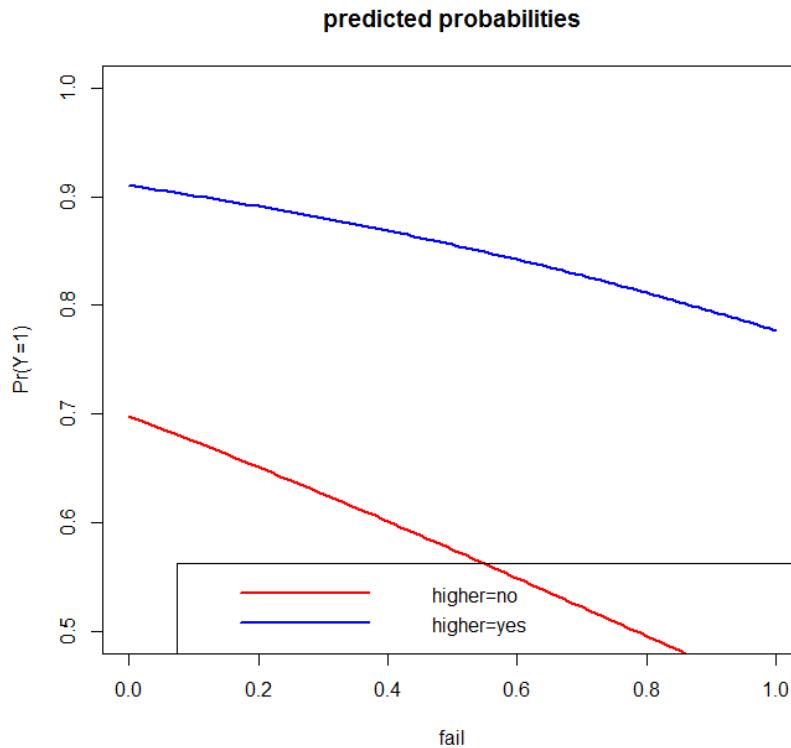
Under H0,

$90.41 \sim \chi^2(2)$, p-value = 2.33174E-20

“ Chosen model is significant ”

Results and Interpretation

Logistic Regression



Results and Interpretation

Logistic Regression

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.2160	0.2305	5.276	1.32e-07	***
sexM	0.7854	0.2971	2.644	0.00820	**
failures	-0.7691	0.1559	-4.933	8.11e-07	***
paidyes	0.8985	0.3098	2.900	0.00373	**

signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ (

(Dispersion parameter for binomial family taken to be

Null deviance: 369.04 on 394 degrees of freedom
Residual deviance: 326.42 on 391 degrees of freedom
AIC: 334.42

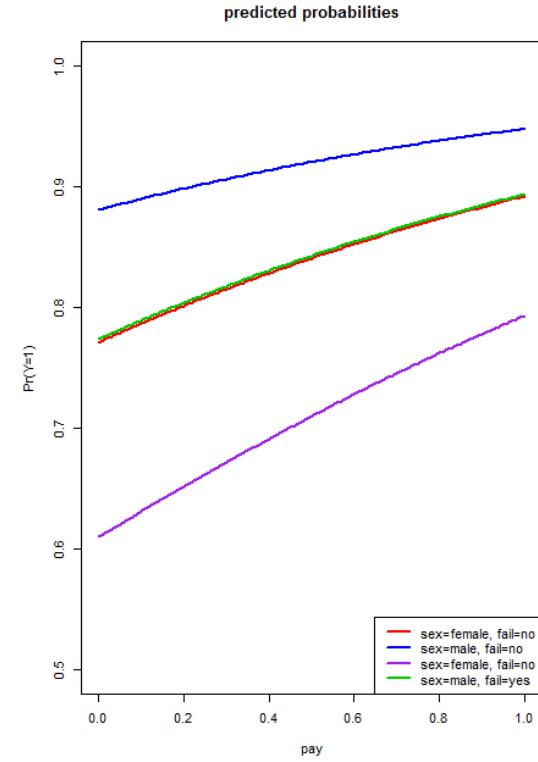
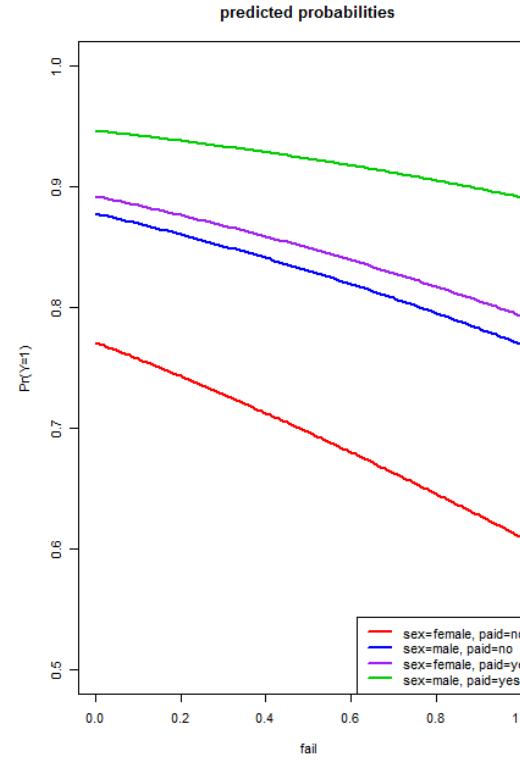
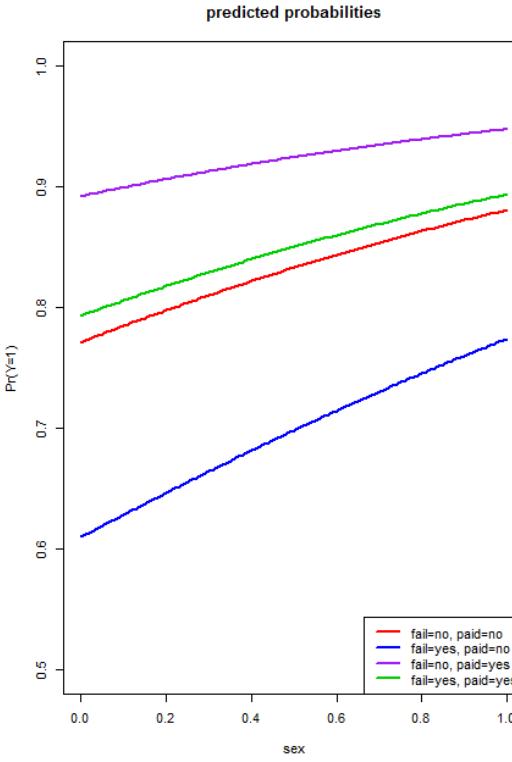
Under H₀,

42.62 ~ X₂ (3), p-value = 1.24075E-08

“ Chosen model is significant ”

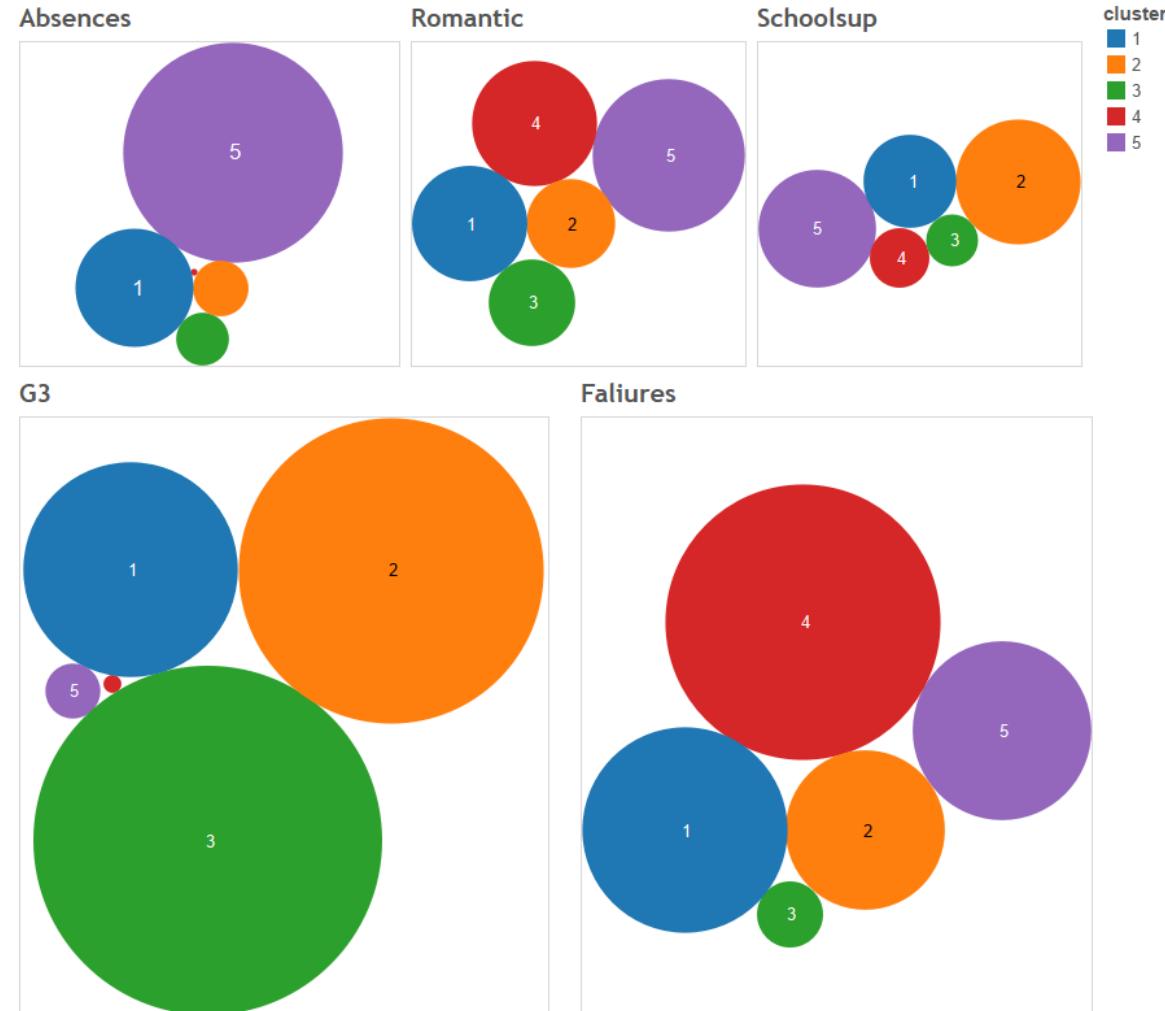
Results and Interpretation

Logistic Regression



Results and Interpretation

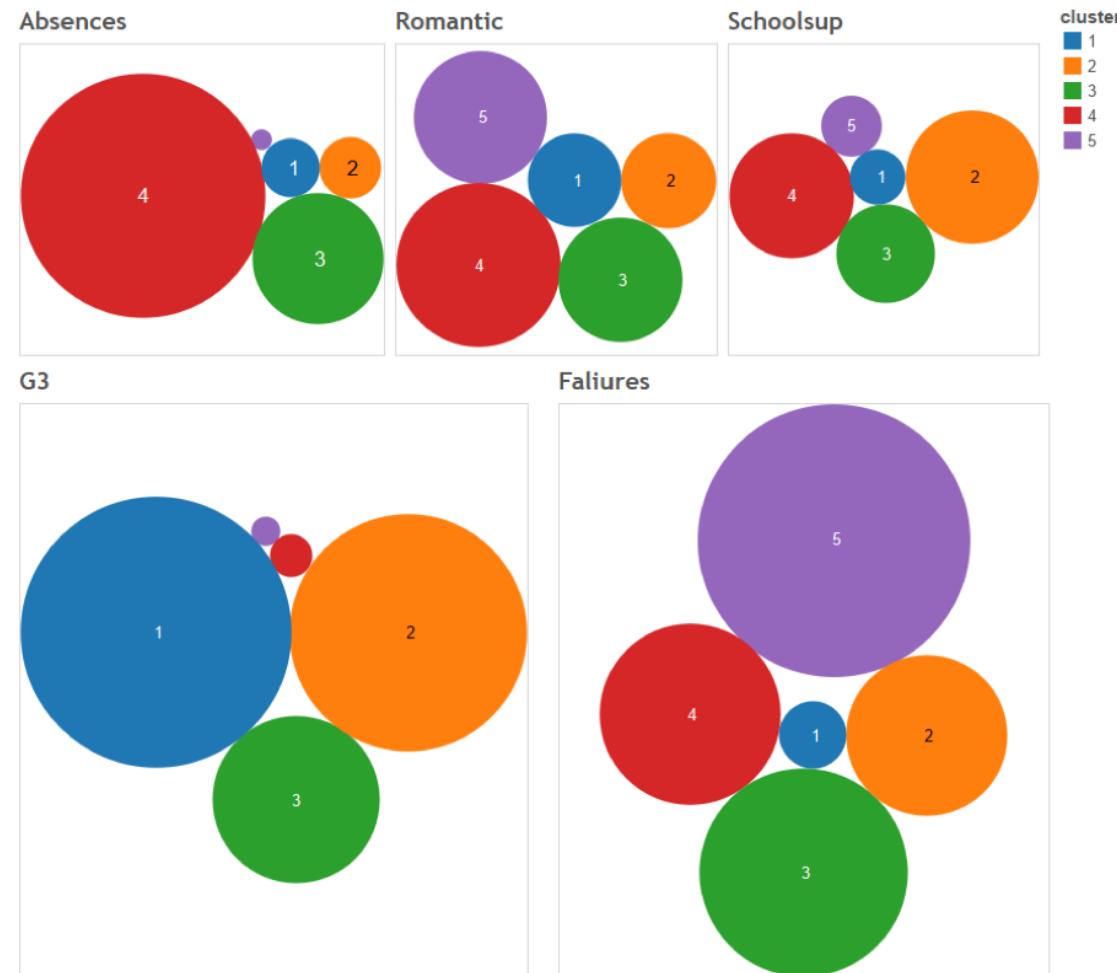
Clustering Analysis (K-means Method)



Portuguese

Results and Interpretation

Clustering Analysis (K-means Method)



5

Conclusion and Future Work



Which factor affects
students' performance the best?

“ *Extra curricular?
Study time?* ”



We
want to
know

5

Conclusion and Future Work



Which factor affects
students' performance the best?

“ Individual Features ”



We
want
to
know



2015 International Summer Semester Data Science and Management

Factors decoding Student Performance

TEAM BBangya

Hansol_Kim / Youtaek_Jeon / Youngha_Lee / Hyoshin_Soh