Prediction of placement and salary of Indian students

June 2021

Elizaveta German, Roberto Russo, Arvin Rastegar

Dataset

- Secondary School Education percentage
- Higher School Education percentage
- Specialization in Higher School Education
- Field of degree education
- Employability test percentage
- Specialization in MBA
- Degree percentage
- MBA percentage
- Board of Education
- Gender
- Work experience
- Salary
- Status of placement

215 students from India

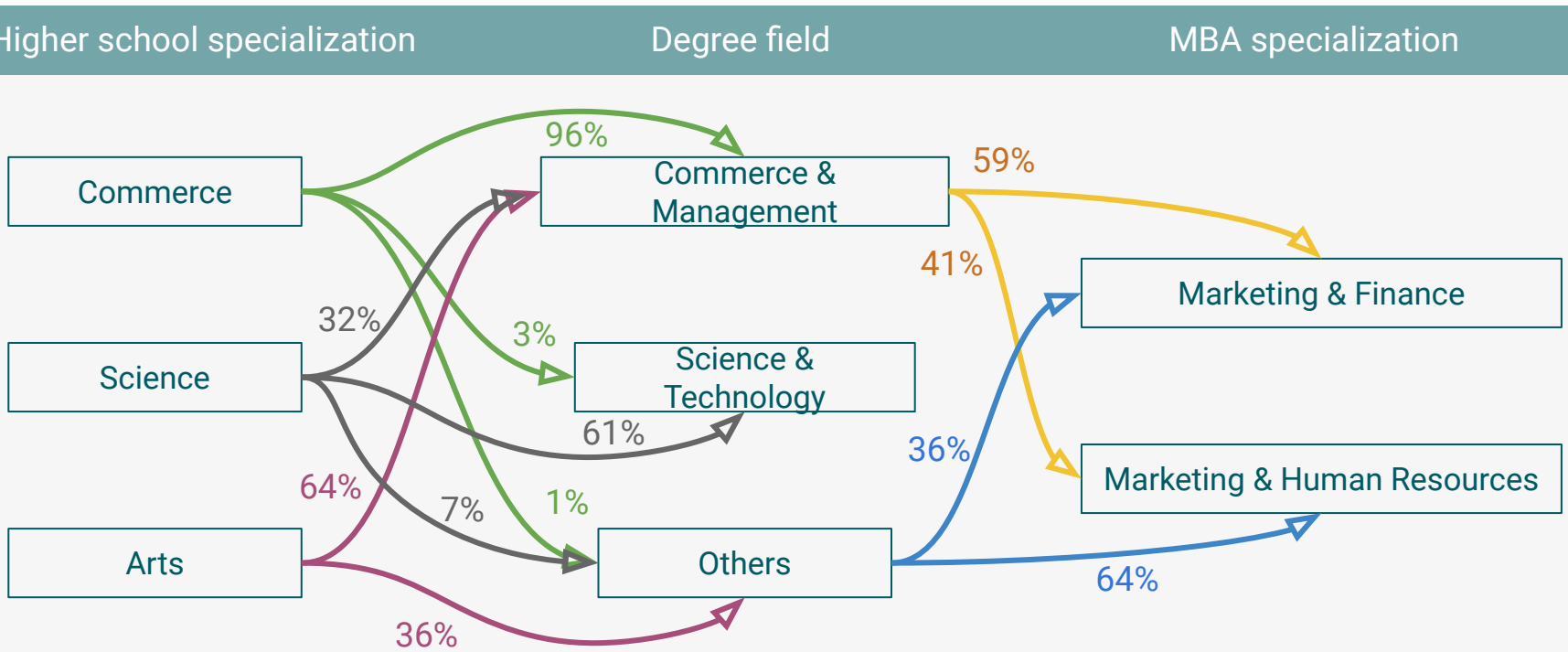
Source: https://www.kaggle.com/benroshan/factors-affecting-campus-placement

Board of Education in India

Each board has different methodology and different aims

Central	Others
– Central Board Secondary Education	– State Board – Cambridge Assessment International Education – International Baccalaureate – Council for the Indian School Certificate Examination

Educational path of Indian students



Methodology: objectives

- ➔ Clean and filter the data
- ➔ Perform exploratory analysis
- ➔ Draw hypotheses
- ➔ Build models capable to predict the placement and the salary of the students
- ➔ Interpret results

Exploratory dataset analysis (EDA)

1. Distribution of the numerical variables – scores through the whole education path
2. Distribution of the categorical variables
3. Splitted distributions (“placed” and “not placed”) to observe the differences
4. Splitted distributions of salary according to levels of each categorical variable
5. We see if the salary is affected by the interaction between the employability test and percentage of MBA

Distributions of the scores during the study path, employability test scores and salary

- ➔ The frequency for 4 education variables tends to be higher for the central values
- ➔ In all of them, both smaller and higher percentages are less frequent
- ➔ Employability test has a more uniform frequency, while salary shows higher frequency for lower values

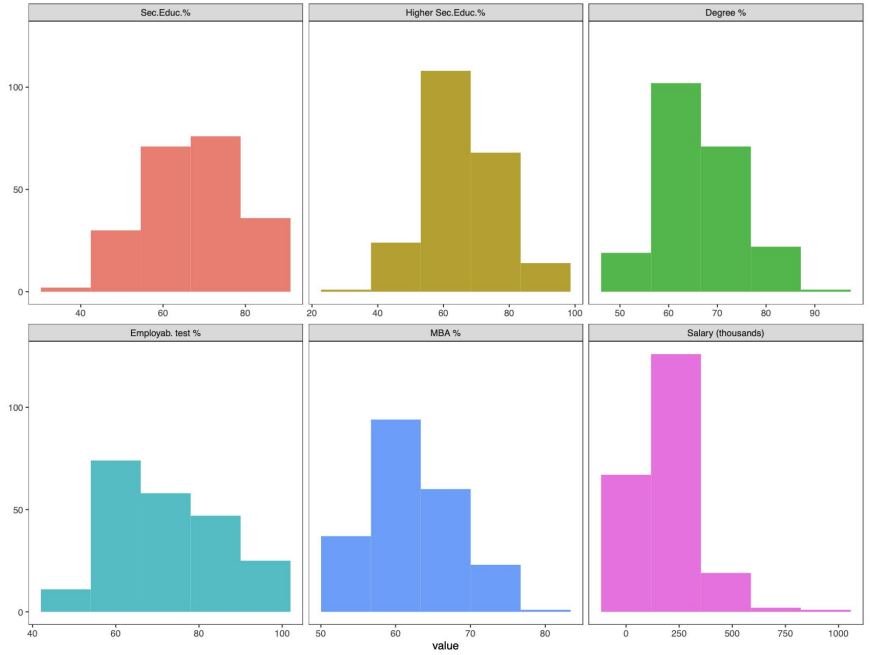
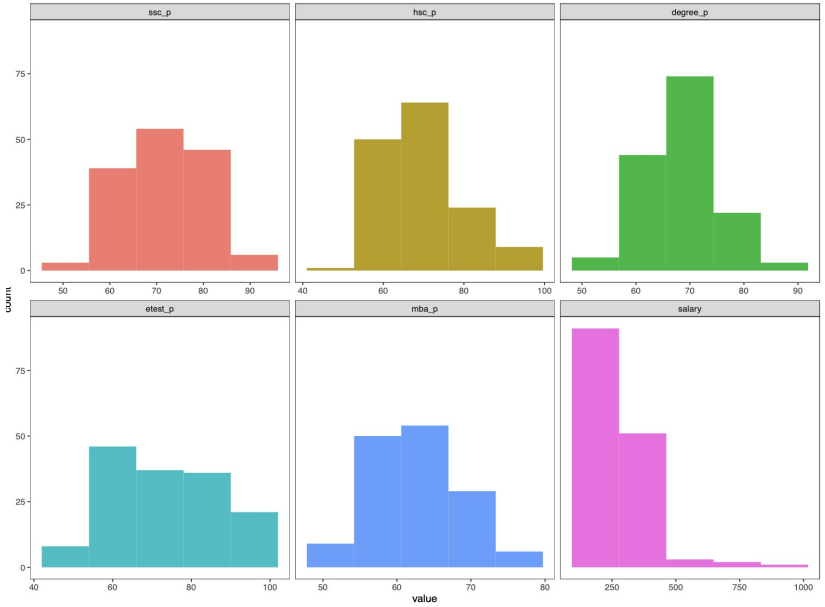


Table 1: Summary of numerical variables

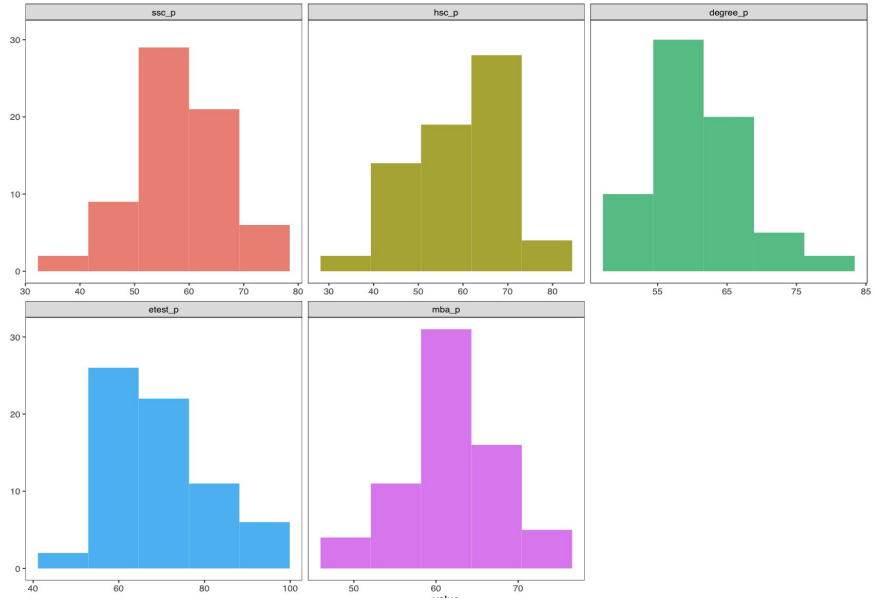
Sec.Educ.%	Higher Sec.Educ.%	Degree %	Employab. test %	MBA %	Salary (thousands)
Min. :40.89	Min. :37.00	Min. :50.00	Min. :50.0	Min. :51.21	Min. : 0.0
1st Qu.:60.60	1st Qu.:60.90	1st Qu.:61.00	1st Qu.:60.0	1st Qu.:57.95	1st Qu.: 0.0
Median :67.00	Median :65.00	Median :66.00	Median :71.0	Median :62.00	Median :240.0
Mean :67.30	Mean :66.33	Mean :66.37	Mean :72.1	Mean :62.28	Mean :198.7
3rd Qu.:75.70	3rd Qu.:73.00	3rd Qu.:72.00	3rd Qu.:83.5	3rd Qu.:66.25	3rd Qu.:282.5
Max. :89.40	Max. :97.70	Max. :91.00	Max. :98.0	Max. :77.89	Max. :940.0

Placed



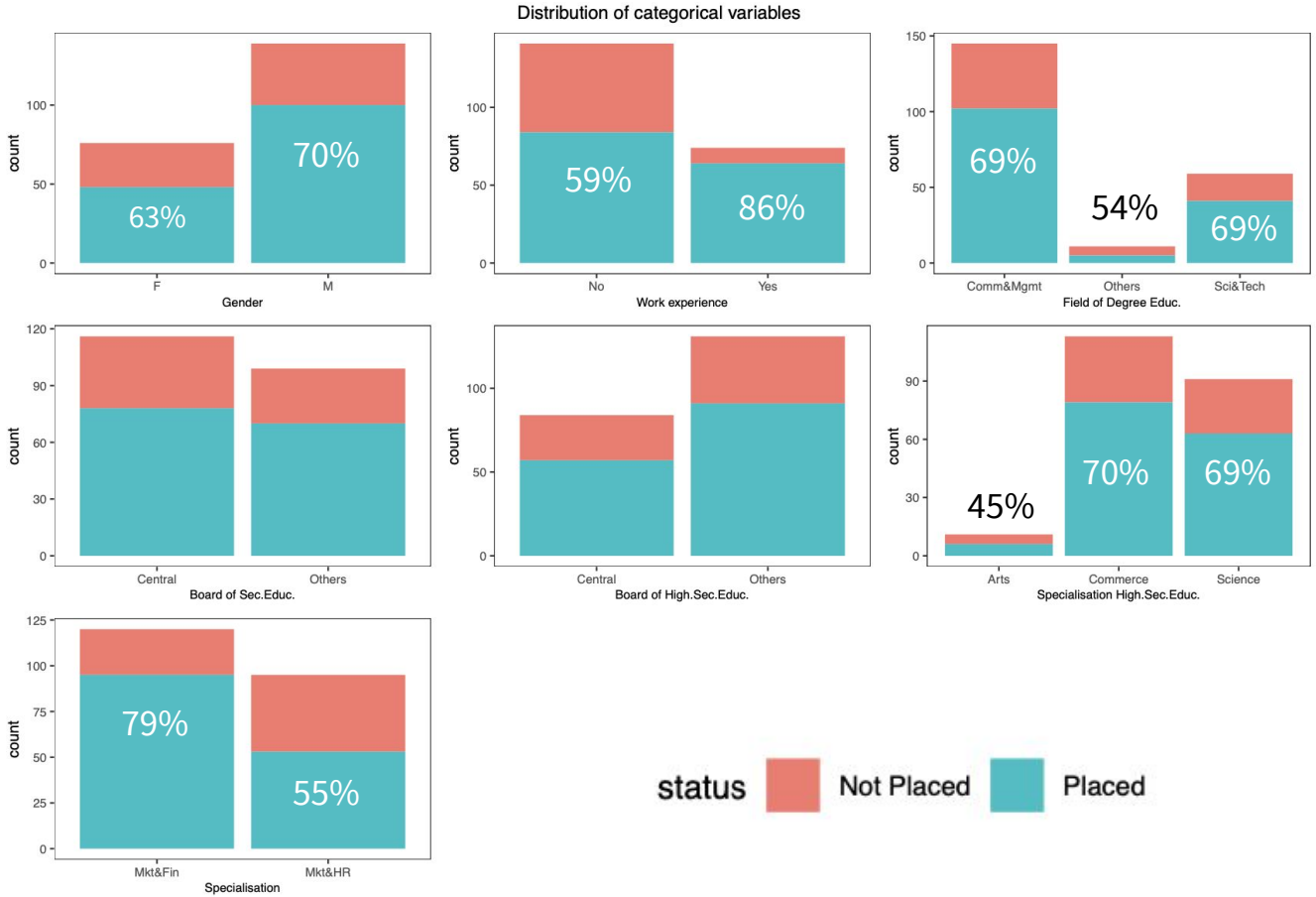
Sec.Educ.%	Higher Sec.Educ.%	Degree %	Employab. test %	MBA %	Salary (thousands)
Min. :49.00	Min. :50.83	Min. :56.00	Min. :50.00	Min. :52.38	Min. :200.0
1st Qu.:65.00	1st Qu.:63.00	1st Qu.:65.00	1st Qu.:60.00	1st Qu.:57.77	1st Qu.:240.0
Median :72.50	Median :68.00	Median :68.00	Median :72.00	Median :62.24	Median :265.0
Mean :71.72	Mean :69.93	Mean :68.74	Mean :73.24	Mean :62.58	Mean :288.7
3rd Qu.:78.12	3rd Qu.:75.25	3rd Qu.:72.42	3rd Qu.:85.00	3rd Qu.:66.76	3rd Qu.:300.0
Max. :89.40	Max. :97.70	Max. :91.00	Max. :98.00	Max. :77.89	Max. :940.0

Not placed



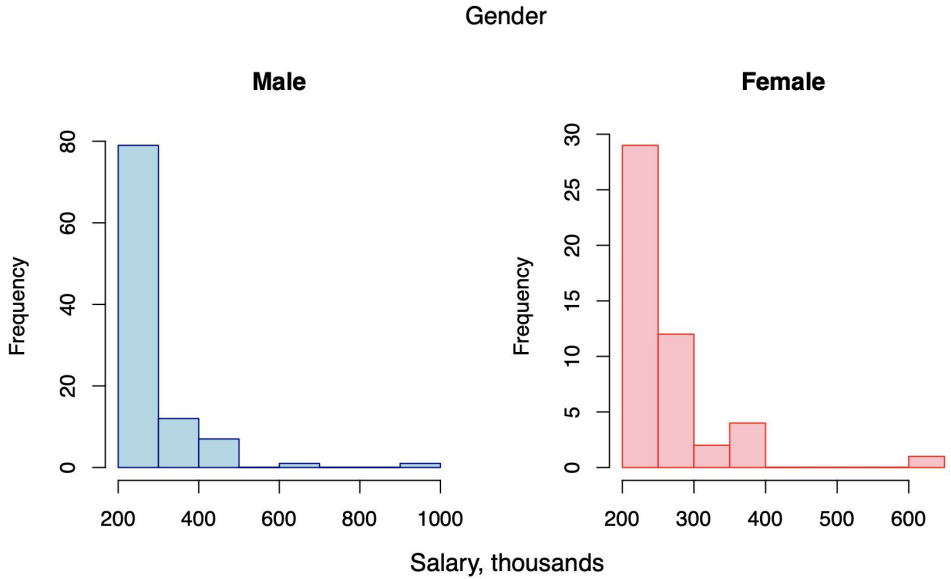
Sec.Educ.%	Higher Sec.Educ.%	Degree %	Employab. test %	MBA %
Min. :40.89	Min. :37.00	Min. :50.00	Min. :50.00	Min. :51.21
1st Qu.:52.00	1st Qu.:51.00	1st Qu.:57.00	1st Qu.:60.00	1st Qu.:58.48
Median :56.28	Median :60.33	Median :61.00	Median :67.00	Median :60.69
Mean :57.54	Mean :58.40	Mean :61.13	Mean :69.59	Mean :61.61
3rd Qu.:63.00	3rd Qu.:64.00	3rd Qu.:65.00	3rd Qu.:76.50	3rd Qu.:65.41
Max. :77.80	Max. :82.00	Max. :79.00	Max. :97.00	Max. :75.71

EDA: Distributions of the scores during the study path, employability test scores and salary



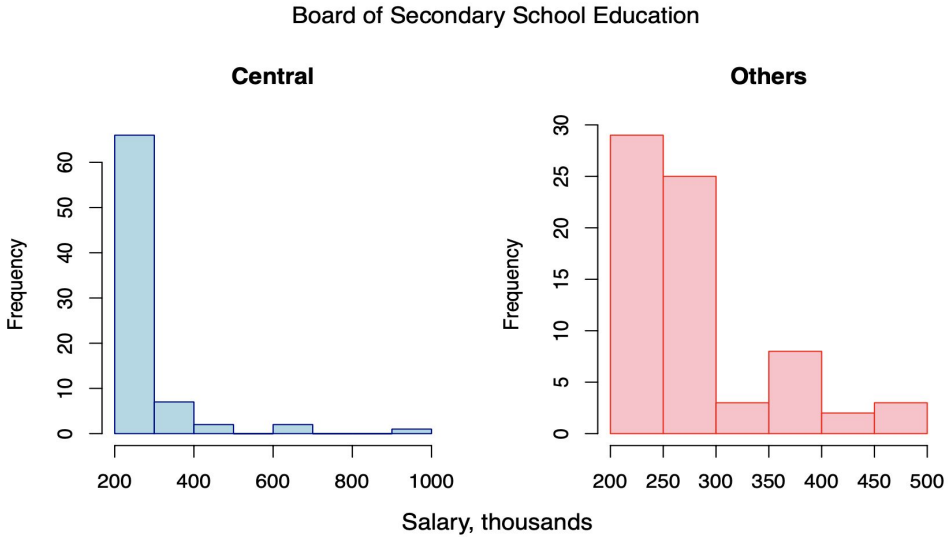
Those variables seem to affect the placement:

- ➔ Gender
- ➔ Work experience
- ➔ Field of Degree
- ➔ Specialization in H.S.
- ➔ MBA Specialization



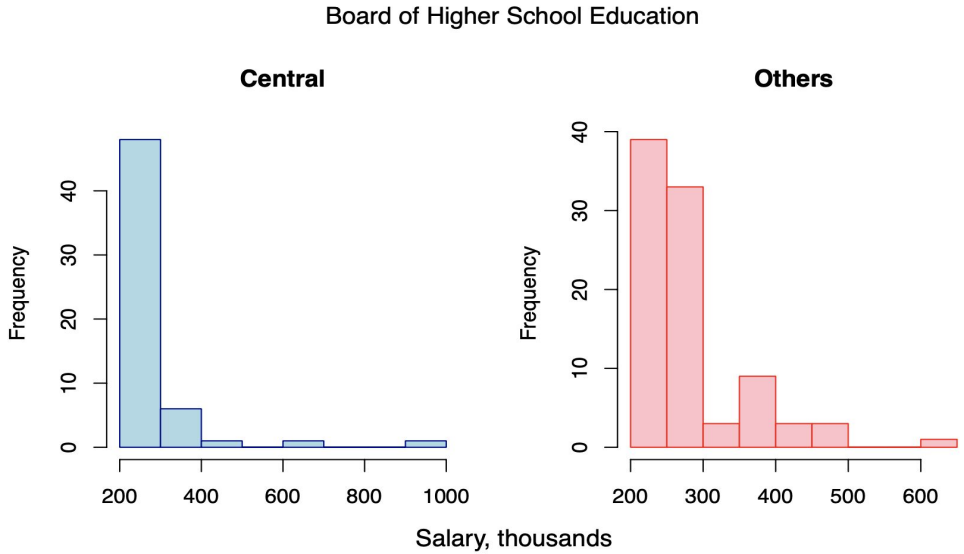
- ➔ Men earn on average more than women
- ➔ Less probability for men to earn less than 250000
- ➔ Less probability for women to earn more than 400000

	Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
	200	250	270	288.4	300	500
	200	219	250	259.1	293.5	400



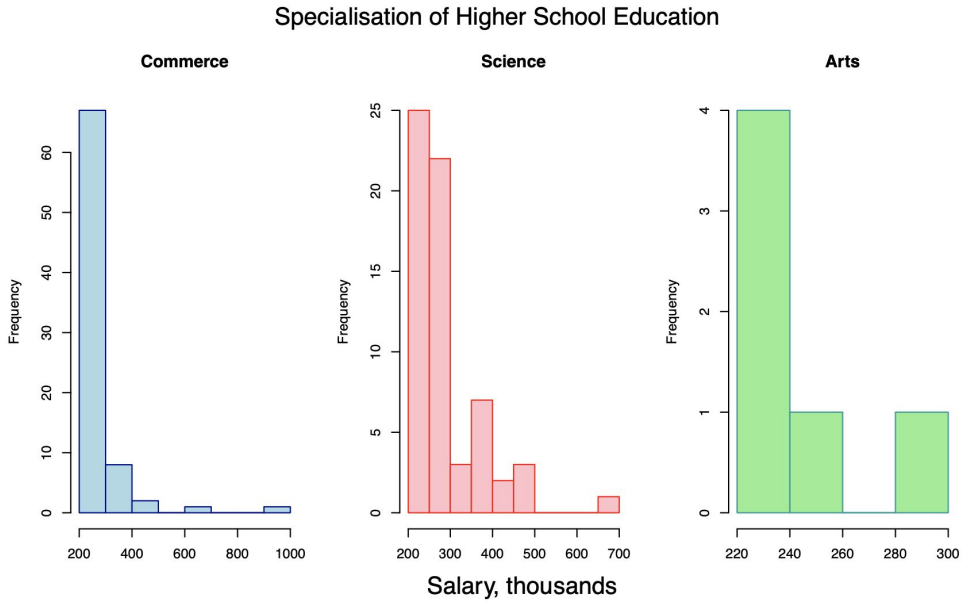
- ➔ People from other boards earn on average more than people from CBSE
- ➔ People from CBSE have less probability to earn more than 425000

Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
200	240	260	269.3	300	425
200	240	265	289.2	300	500



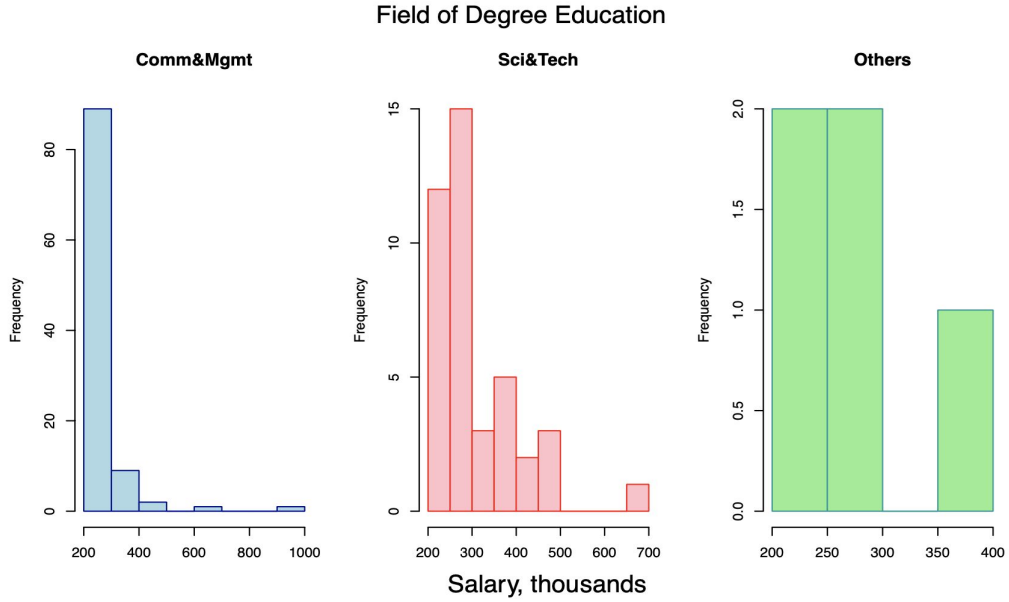
- People from other boards earn on average more than people from CBSE
- People from CBSE have less probability to earn more than 425000

	Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
	200	240	260	270.4	300	425
	200	240	265	284.1	300	500



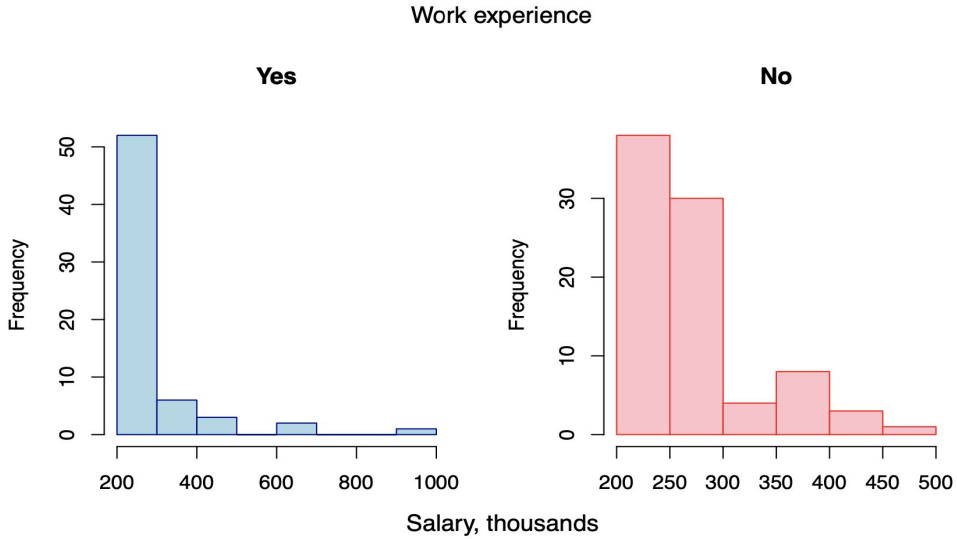
Min.	200	200	230
1st Qu.	240	240	236
Median	265	260	238
Mean	274.2	287.6	248.7
3rd Qu.	300	300	247.5
Max.	425.0	500	300

- ➔ People with H.S. specialization in Science earn on average more than the others
- ➔ They also have higher probability to earn more than 300000



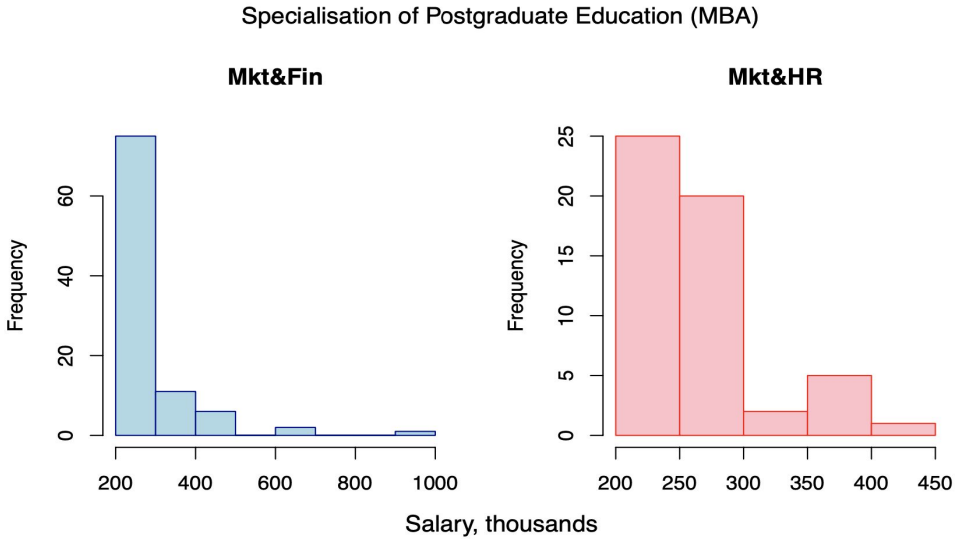
Min.	200	200	240
1st Qu.	236	250	250
Median	260	275	252
Mean	268.3	305.2	280.4
3rd Qu.	300	352.5	300
Max.	425	500	360

- ➔ People with degree field in Science and Technology earn on average more than the others
- ➔ They also have higher probability to earn more than 300000



➔ It seems that having work experience does not influence the salary

	Minimum	1st Qu	Median	Mean	3rd Qu.	Maximum
	200	250	265	280	300	500
	200	240	262	277.5	300	500



- People with MBA specialization in Marketing and Finance earn on average more than people specialized in Marketing and HR
- M & HR people have less probability to earn more than 450000

Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
200	240	266.5	283.8	300	500
200	240	255	270.4	300	450

Hypotheses

Placement

Final scores of school, degree and postgraduate education
Field of degree education
Specialisation of MBA
Work experience
Gender

Salary

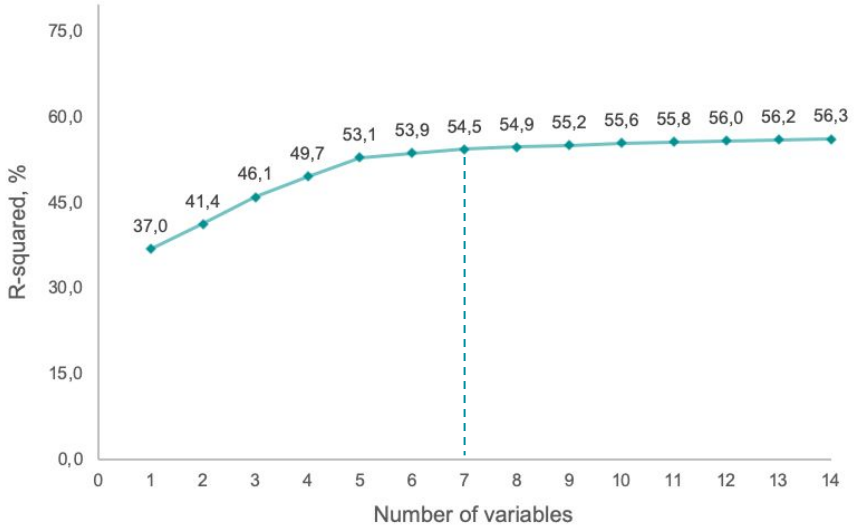
Gender
Higher School Specialization
Final scores of school, degree and postgraduate education
Field of degree education
Specialisation of MBA

Placement prediction: full model

Variable	Coefficient	P-value < 0.1
Intercept	-18.4	0.001
Secondary Education %	0.23	0.000
Higher Education %‘	0.12	0.005
Degree %	0.19	0.001
Employability test %	-0.01	0.532
MBA %	-0.21	0.000
Gender (male)	1.19	0.082
Board of Secondary Education	0.23	0.751
Board of Higher Education	0.33	0.653
Higher school specialization (Commerce)	-1.50	0.271
Higher school specialization (Science)	-0.91	0.532
Degree field (others)	-1.12	0.470
Degree field (Science & Technology)	-1.73	0.029
Work experience (Yes)	2.08	0.003
MBA specialization (Marketing & HR)	-0.26	0.635

Placement prediction: best subset selection

R-squared for each model



Indicators for comparison of models

	# of variables
Mallow's Cp	7
Bayesian Information Criterion	5
Adjusted R-Squared	11

Placement prediction: final model

Binary classification

Variable	Coefficient	Exponential	P-value
Intercept	-20.9	0.00	0.000
Secondary Education %	0.22	1.25	0.000
Higher Education %‘	0.10	1.11	0.003
Degree %	0.19	1.21	0.001
MBA %	-0.20	0.82	0.000
Gender (male)	1.37	3.95	0.029
Degree field (others)	-0.48	0.62	0.681
Degree field (Science & Technology)	-1.43	0.24	0.020
Work experience (Yes)	2.29	9.88	0.001

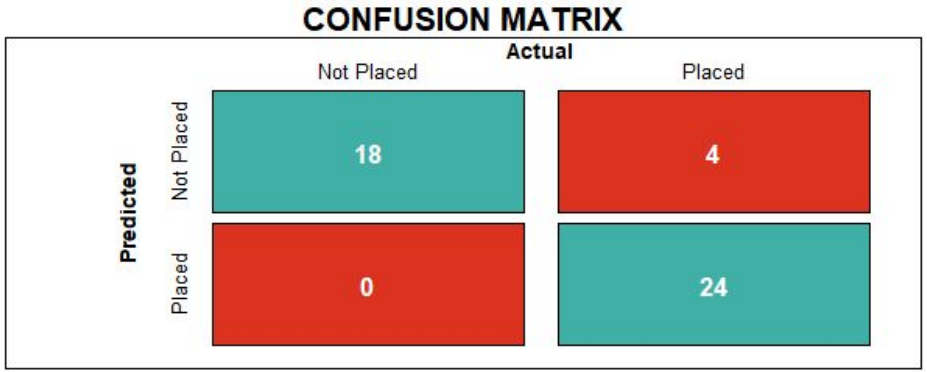
Placement prediction: performance measure

➔ Different performance measures’ errors

➔ Accuracy: Percentage of correct predictions = 91%

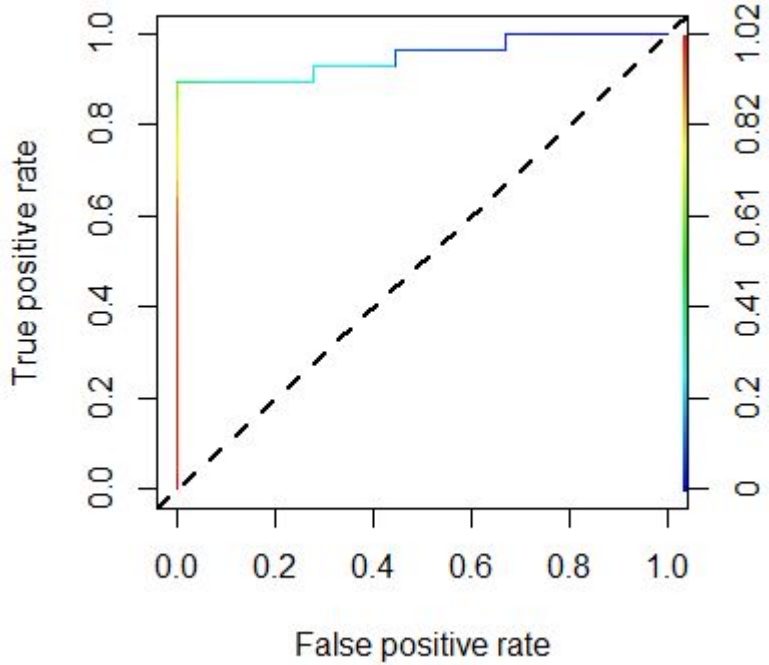
Confusion matrix

Different errors affect different groups of people differently



AUC

The area under ROC curve which can be used to determine the relationship between true positive rate and false negative rate

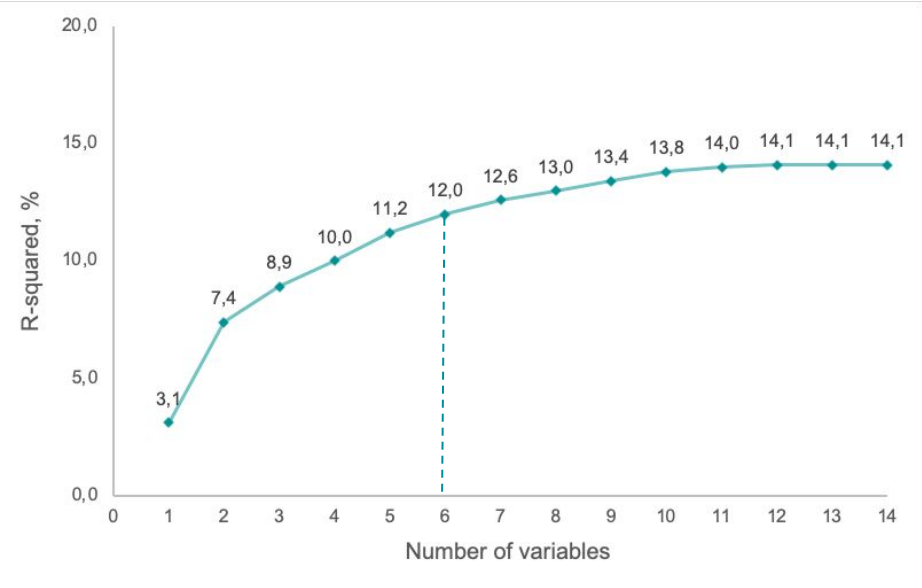


Salary prediction: assumptions

1. Linearity
2. Normality of standardized residuals
3. Heteroscedasticity
4. Multicollinearity
5. Outliers

Salary prediction: best subset selection

R-squared for each model



Indicators for comparison of models

	# of variables
Mallow's Cp	3
Bayesian Information Criterion	2
Adjusted R-Squared	6

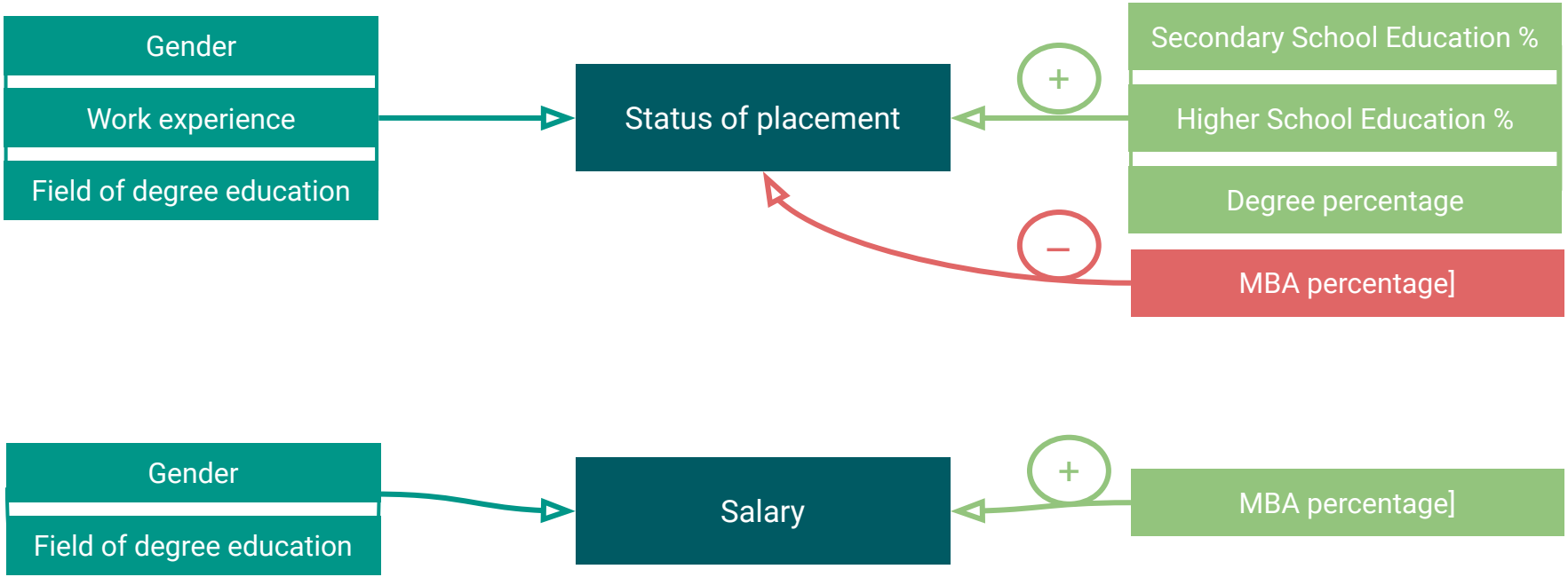
Salary prediction: linear regression

Variable	Coefficient	P-value
Intercept	67.6	0.529
Degree %	-2.1	0.132
Employability test %	0.7	0.221
MBA %	3.8	0.017
Gender (male)	29.5	0.092
Degree field (others)	28.8	0.535
Degree field (Science & Technology)	41.7	0.085
Higher School specialization (Commerce)	52.6	0.212
Higher School specialization (Science)	32.1	0.459

Salary prediction: exponential regression

Variable	Coefficient	P-value
Intercept	4.93	0.000
Degree %	-0.01	0.180
Employability test %	0.00	0.153
MBA %	0.01	0.012
Gender (male)	0.10	0.026
Degree field (others)	0.12	0.315
Degree field (Science & Technology)	0.13	0.036
Higher School specialization (Commerce)	0.15	0.166
Higher School specialization (Science)	0.09	0.430

Results


Gender
Work experience
Field of degree education
Status of placement
+
Secondary School Education %
Higher School Education %
Degree percentage
–
MBA percentage]
Gender
Field of degree education
Salary
+
MBA percentage]