Analyzing Student Salaries
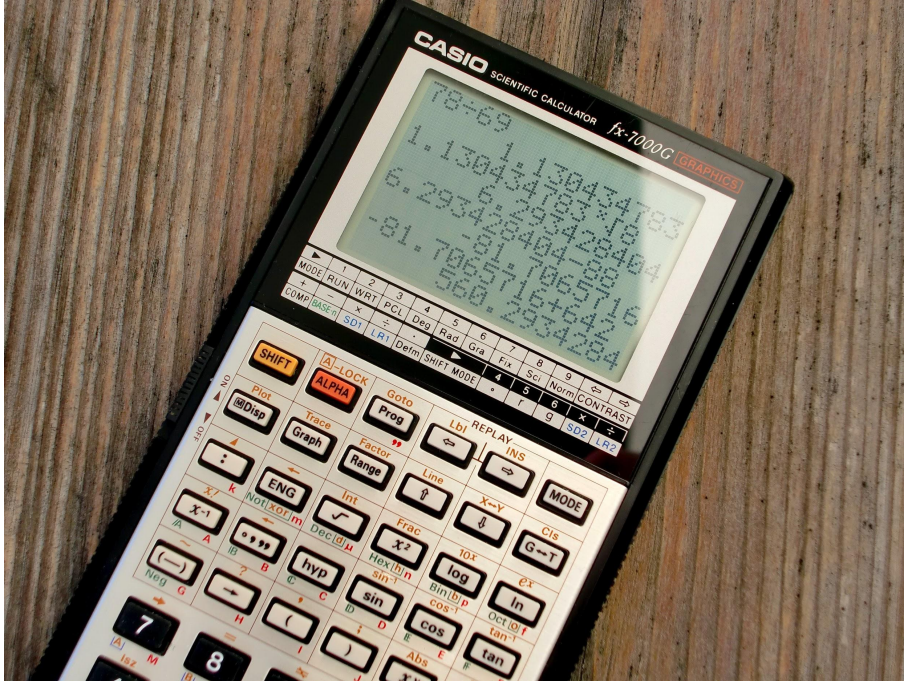
By: Gabe Arnold

What is the main purpose?

"Which data points will be the best predictor of a student's salary?"

My original hypothesis was that Undergraduate degree, and Postgraduate Specialization would have the largest effect on salary.

Findings

Based on the data, Employment Test scores, & MBA Percentage were the best predictors of student salary.
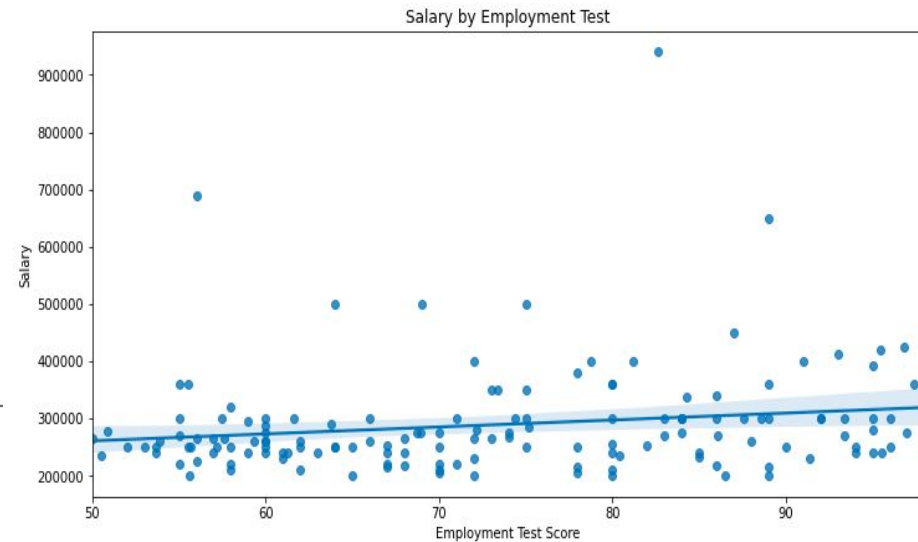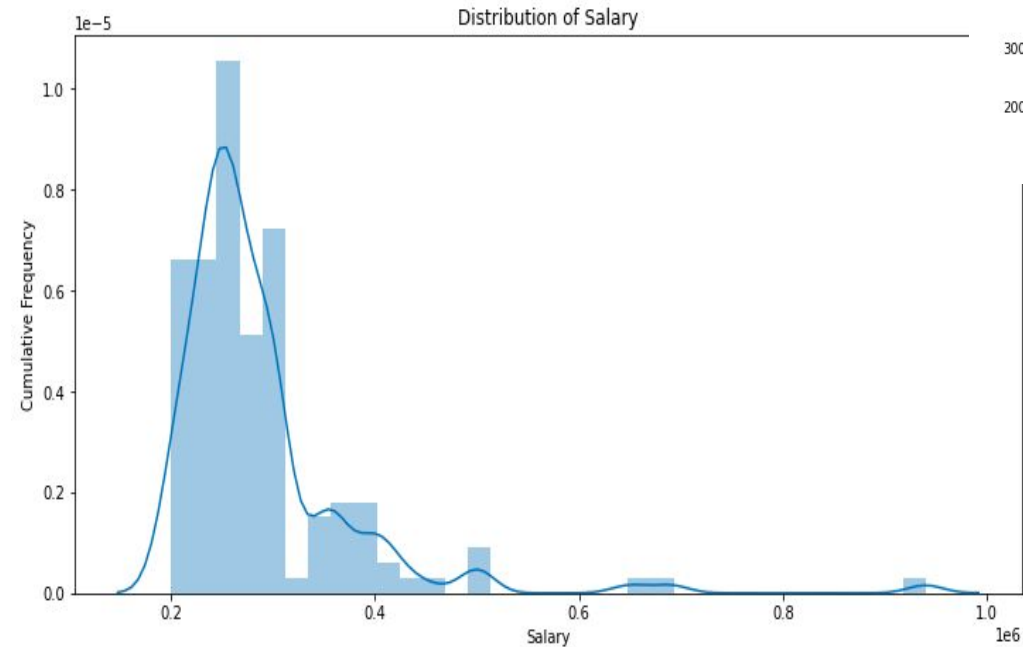
*Employment test are workplace readiness test that are given to job seekers in India.

Process

1. Large Scale Overview
2. Understanding current data via linear regression
3. Data cleanup and adjustments
4. Modeling

```
gender
F      267291.67
M      298910.00
```

Exploratory Data Analysis

Distribution of Salary

Salary by Employment Test

(Upper) Graph showing salary by employment test score.

(Left) Histogram of student salaries.

The average salary is 288655.41

Removing Data

Category	P-Score
Hsc_s - Grad School Specialization	.52008
Hsc_b - Board of Education Region	.92744
Degree_t - Undergrad Degree	.11183

Ran T-Test & ANOVA test on my categorical variables to see if there was a significant difference between the means in two unrelated groups.

Correlation Map
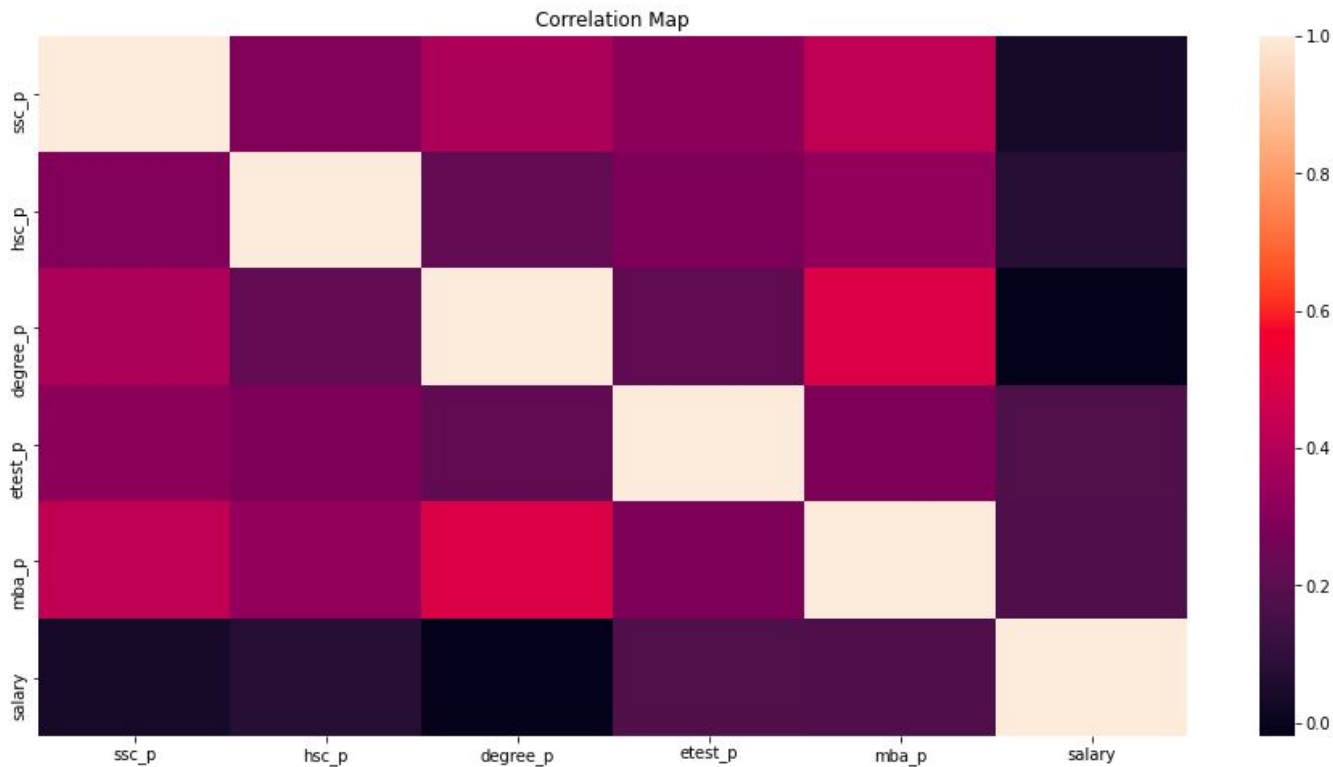
Modeling Results

Test Name	Training RMSE	Testing RMSE
Linear Regression Model	80450	88555
Lasso Testing	52536	83108
Select K-Best	96442	50428
RFECV	100746	55336

Takeaways

Using Select K-Best for feature selection, I found that Employment Test, MBA Percentage were some of the best predictors of student salary.