



Group Project

Applied Statistics

15 February, 2024



Prepared by:
G Madhava Asrith Murthy
Panshul Jindal
Sreenandan Babu

Introduction



This presentation covers two projects:

- **JEE Advanced 2023 Analysis:** Statistical insights into rank, marks, and seat allocation trends across IITs.
- **Statistics Project 2:** Parameter estimation, confidence intervals, and hypothesis testing using Indian rainfall and Cancer data
- **Objectives:** Apply statistical methods to real datasets, validate models, and extract actionable insights.



Gamma Parameter Estimation

Objective:

Model annual rainfall data using the Gamma distribution and estimate its parameters.

Result:

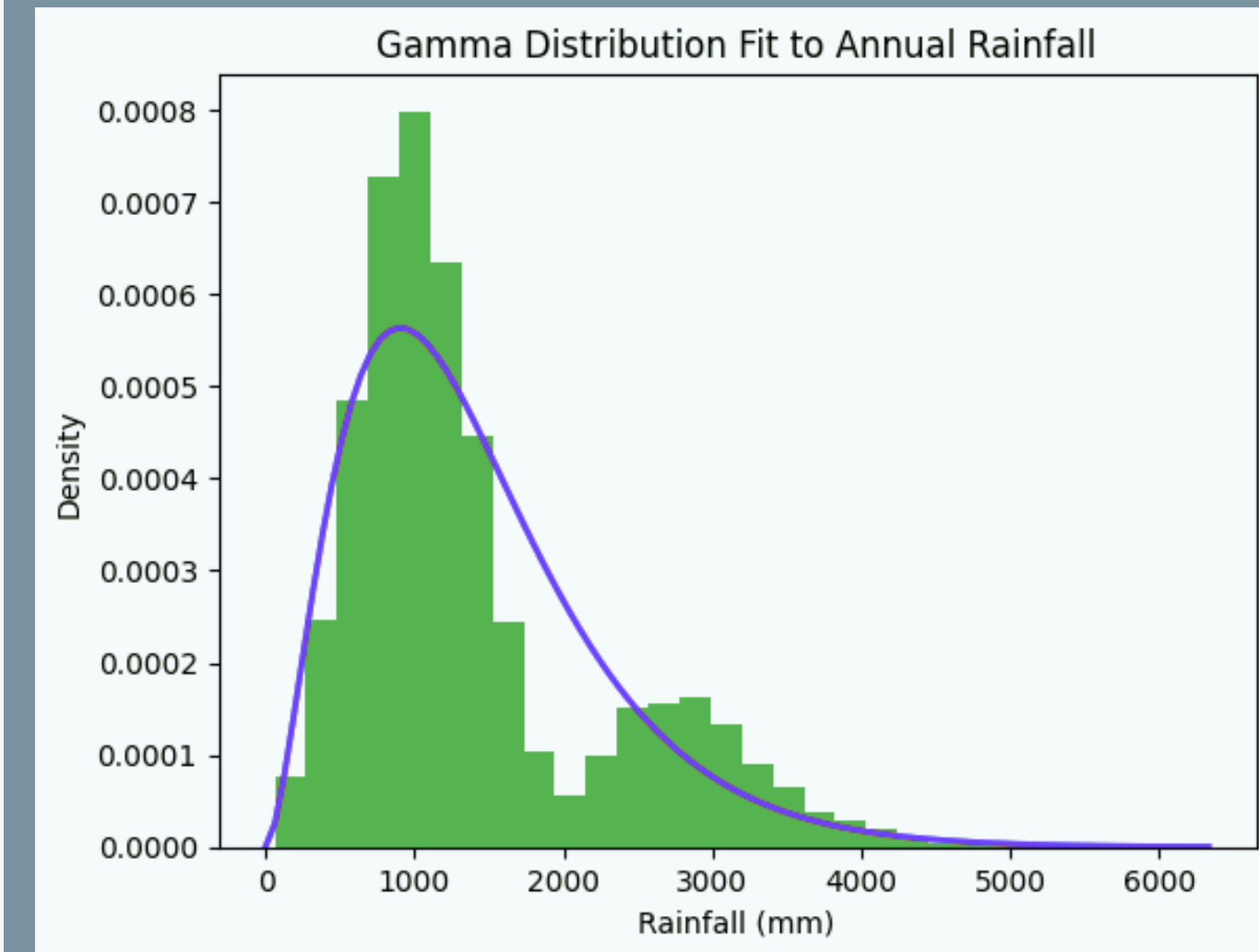
Method of Moments: Shape ≈ 2.46 , Scale ≈ 575

Maximum Likelihood: Shape ≈ 2.90 , Scale ≈ 523

Both methods yield similar results

Inference:

The Gamma model fits the rainfall data well, validating both estimation approaches



Confidence Interval for Variance

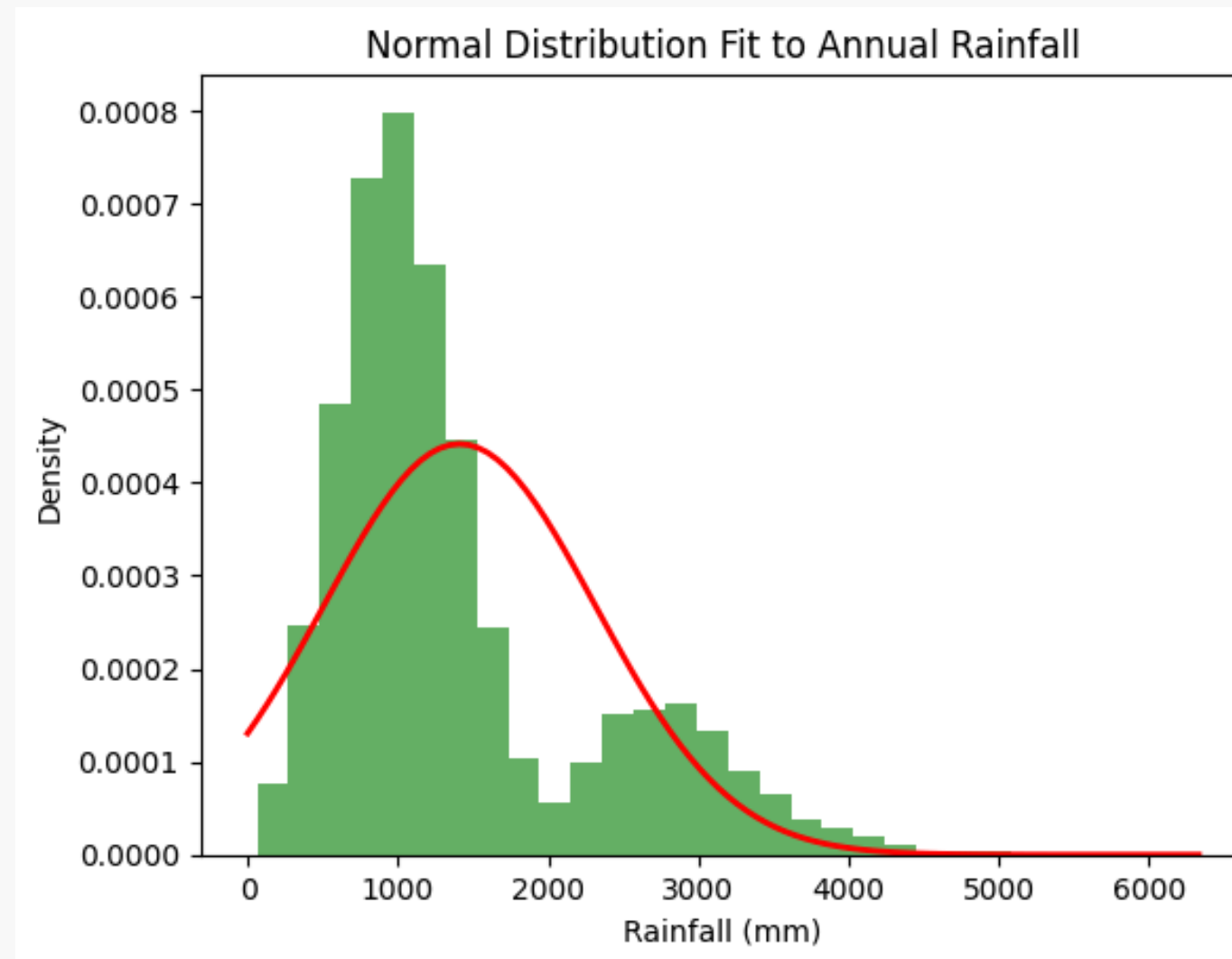


Objective:

Construct a 95% confidence interval to estimate the variance of annual rainfall.

Result:

95% CI for variance: (638,871, 1,118,374);
for standard deviation: (799, 1,058)



Inference:

The 95% confidence interval provides a range within which the true variance of annual rainfall is likely to lie. While the normal distribution was assumed for this calculation, the fit is not perfect, suggesting some caution in interpretation.



Confidence Interval for Difference of Means



Objective:

Compare mean annual rainfall between Lakshadweep and Andaman & Nicobar Islands and find the Confidence Interval.

Methodology using Welch's t-test

Result

95% CI for the difference of mean rainfall: (-1,432, -1,237).

Population mean difference: -1336.553

Experiment: Understanding the Confidence Interval

Sample Size: 30

75% Confidence → 58.00
80% Confidence → 67.30%
90% Confidence → 86.40%
95% Confidence → 94.70%
97.5% Confidence → 97.90%
100% Confidence → 100%

Sample Size: 50

75% Confidence → 63.80%
80% Confidence → 77.80%
90% Confidence → 92.00%
95% Confidence → 97.40%
97.5% Confidence → 99.30%
100% Confidence → 100%

Sample Size: 100

75% Confidence → 99.90%
80% Confidence → 100.00%
90% Confidence → 100.00%
95% Confidence → 100.00%
97.5% Confidence → 100%
100% Confidence → 100%

One-Sample Hypothesis Testing – Bernoulli Data

$$\alpha = 5$$

Objective:

In dataset containing the data of breast cancer being benign (B) or malignant (M), we test the null hypothesis that probability of a cancer being benign $\leq \frac{1}{2}$ vs probability of cancer being benign $> \frac{1}{2}$

Population Probability of cancer being Benign: $p = 0.6274$

Observation:

The test statistic is

$$Z = \frac{\bar{X} - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}$$

Rejection Rule : if $Z \geq z_{\alpha}$ $z_{\alpha} = 1.96$ for $\alpha = 0.05$

For $n = 30$, the test statistic was computed to be: $Z = 1.826 < 1.96$.

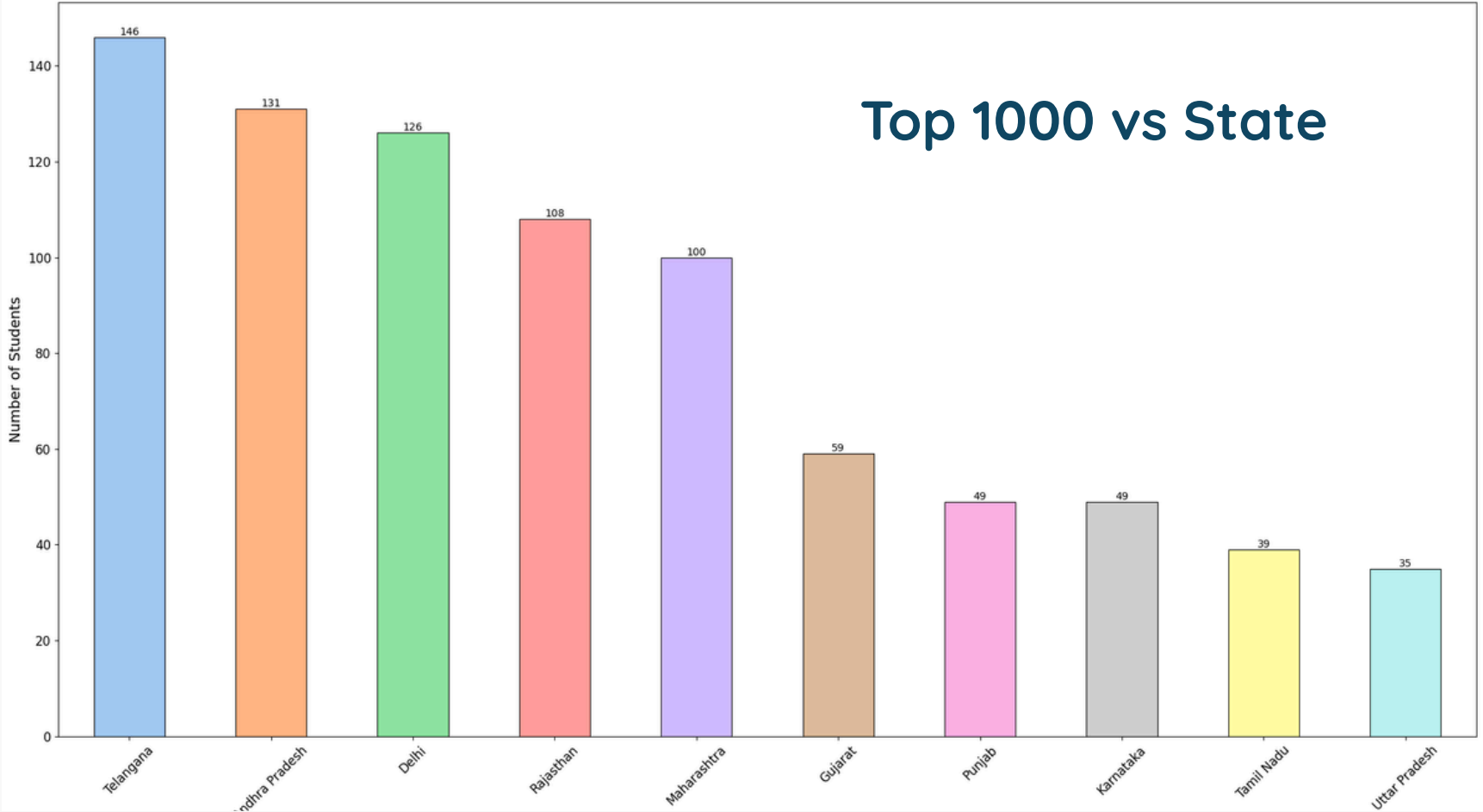
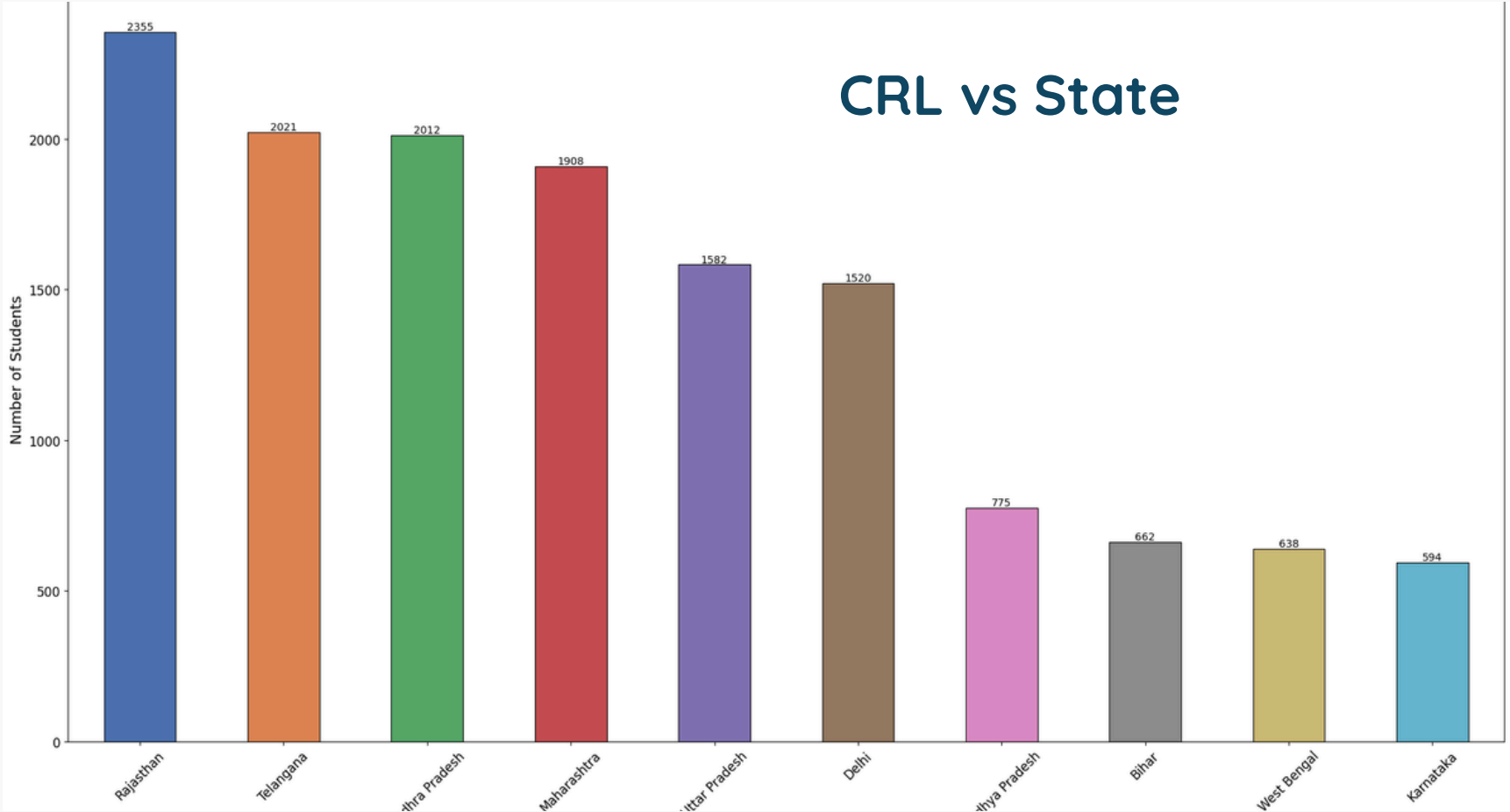
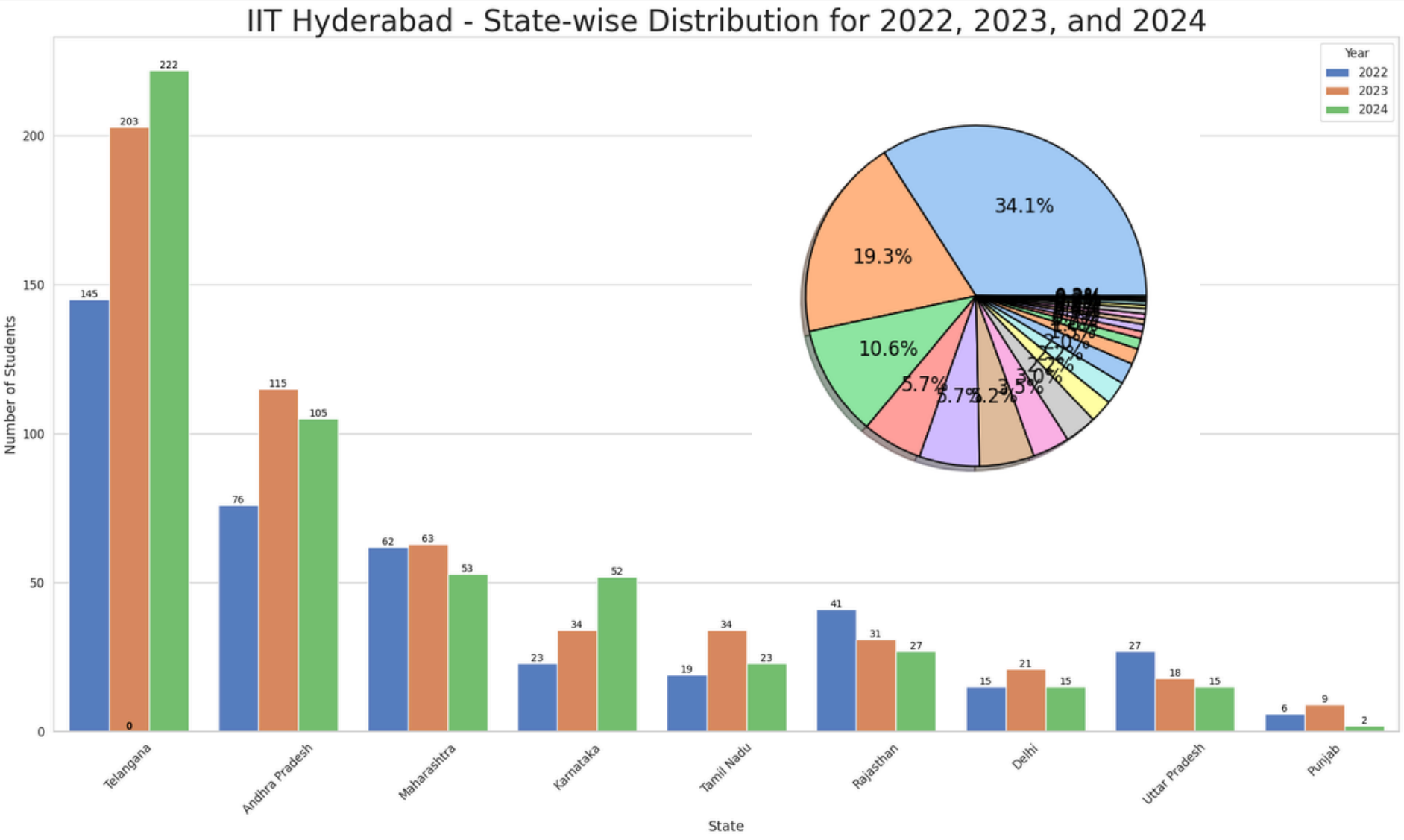
Experiment Effect of Sample Size on test statistic

```
n 1 value 1.00 Failed to Reject H0
n 11 value 2.11 Reject H0
n 21 value 1.96 Reject H0
n 31 value 1.98 Reject H0
n 41 value 1.09 Failed to Reject H0
n 51 value 3.78 Reject H0
n 61 value 1.66 Failed to Reject H0
n 71 value 2.25 Reject H0
n 81 value 2.11 Reject H0
n 91 value 0.52 Failed to Reject H0
n 101 value 3.88 Reject H0
n 111 value 3.13 Reject H0
n 121 value 2.45 Reject H0
```

```
n 131 value 3.58 Reject H0
n 141 value 3.96 Reject H0
n 151 value 3.50 Reject H0
n 161 value 3.70 Reject H0
n 171 value 3.75 Reject H0
n 181 value 2.90 Reject H0
n 191 value 4.99 Reject H0
n 201 value 3.32 Reject H0
n 211 value 5.30 Reject H0
n 221 value 3.97 Reject H0
n 231 value 3.62 Reject H0
n 241 value 3.80 Reject H0
n 251 value 2.59 Reject H0
```

We observe that after a certain sample size, the test consistently rejects H_0

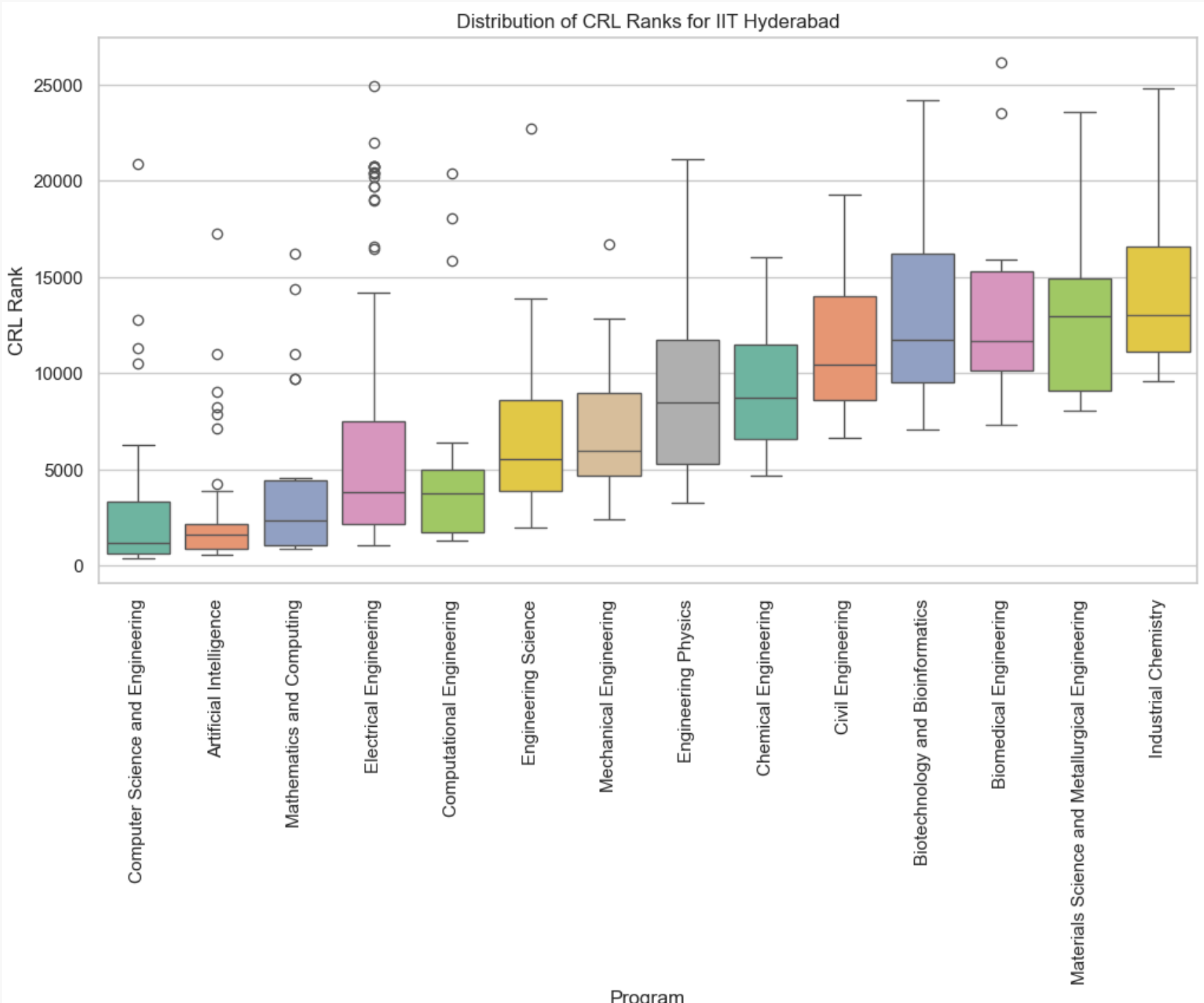
CRL Distribution



We observe that number of students from Telangana going to IITH increased over the last 3 years.

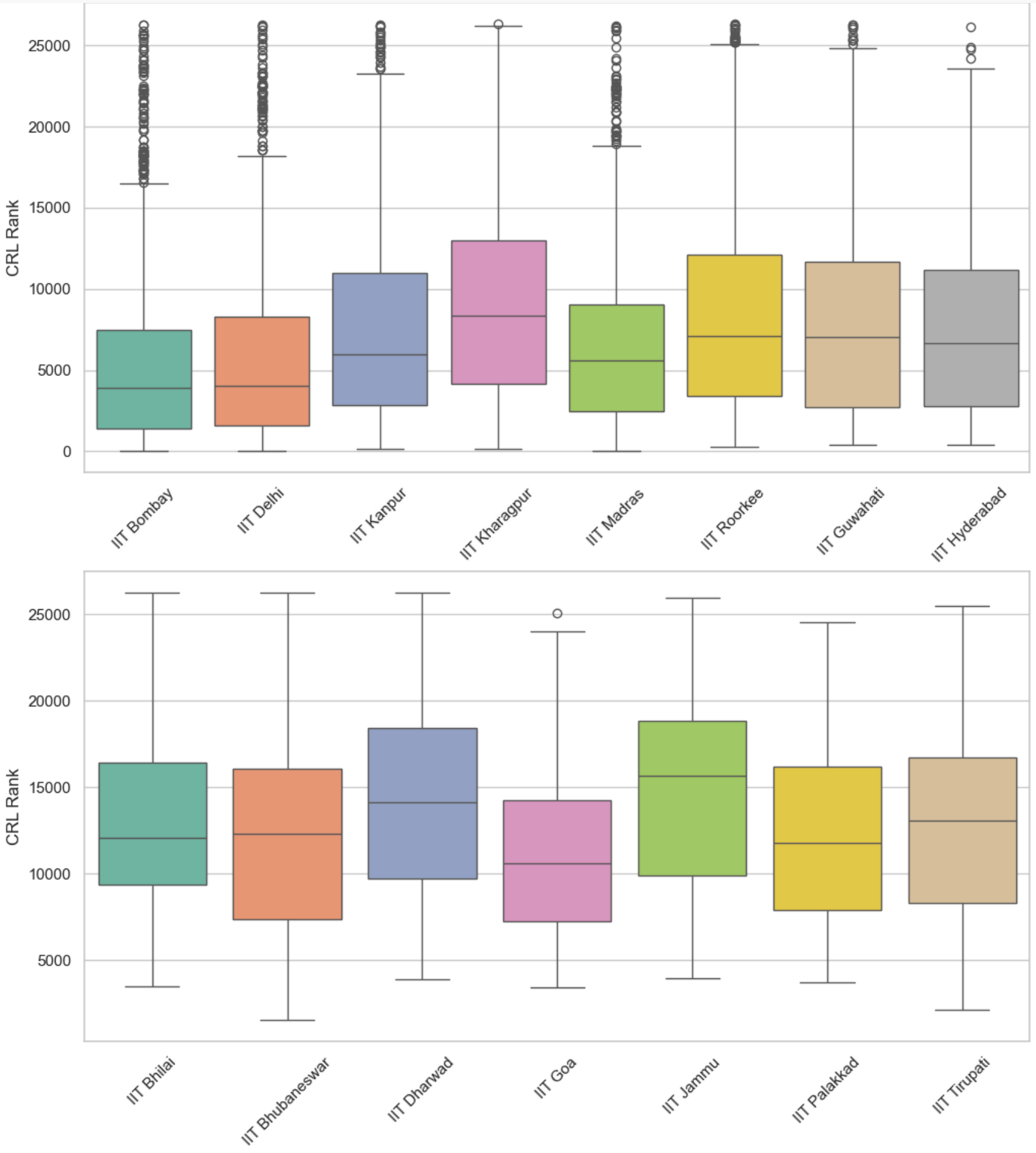


Comparision of Distribution of ranks across IITs

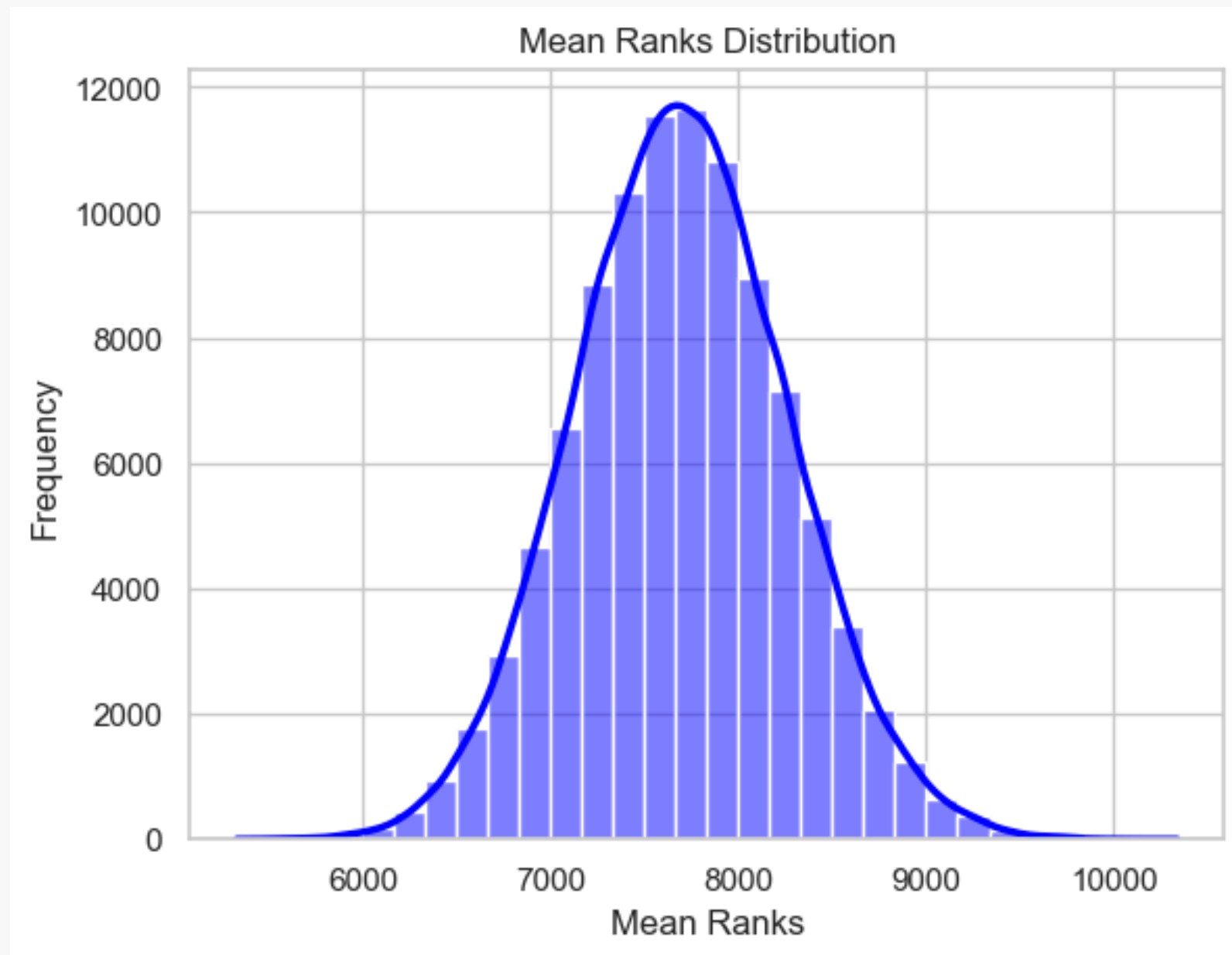


CRL vs Branch in IITH

Notice the variation in the Median Rank across the Branches in IITH



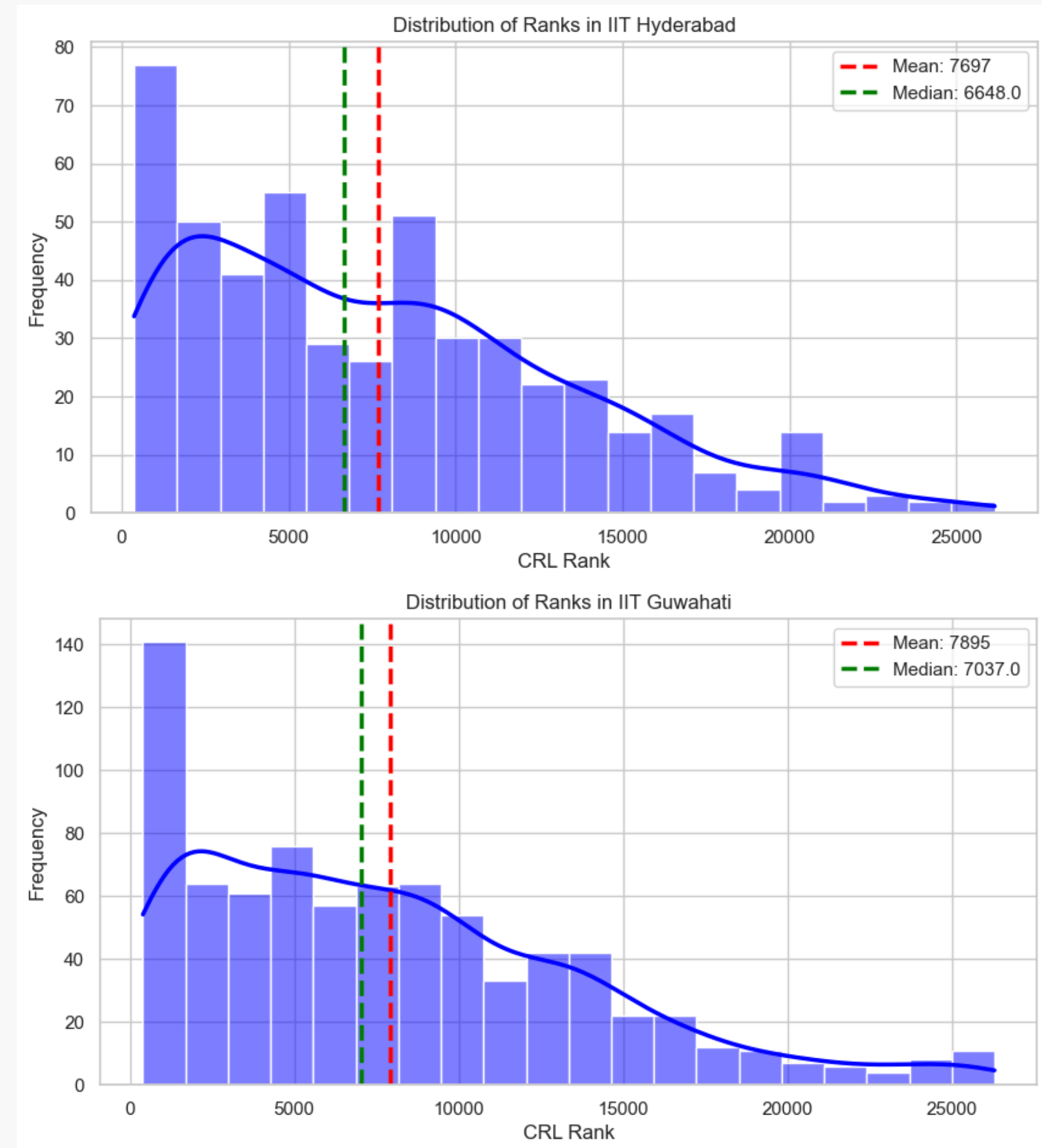
CRL Distribution



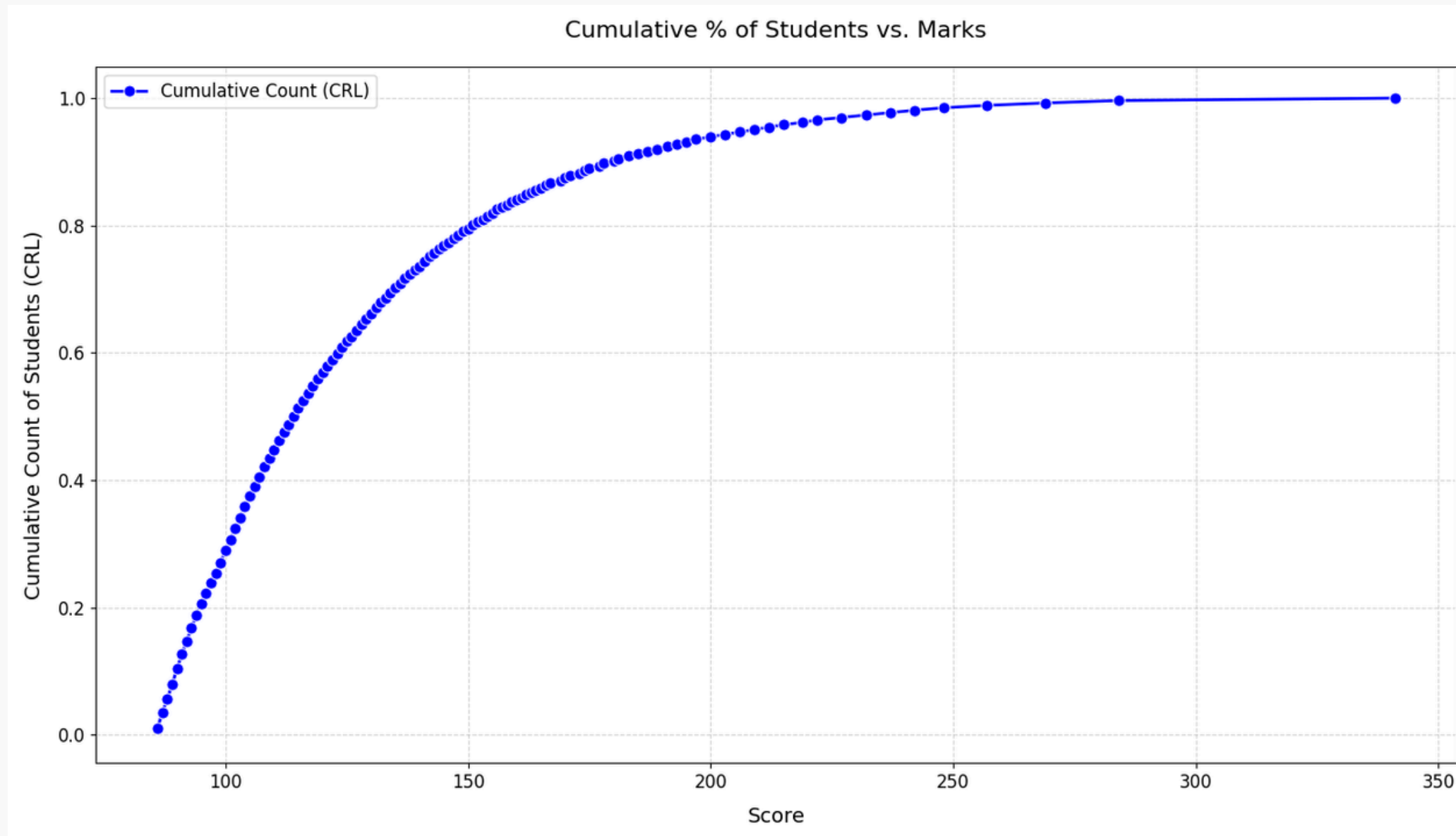
- The Approximate Mean Rank of IITH students is around 7500
- Notice the Median and Mean ranks of IITH and IITM



Comparison of ranks across IITH vs IITG



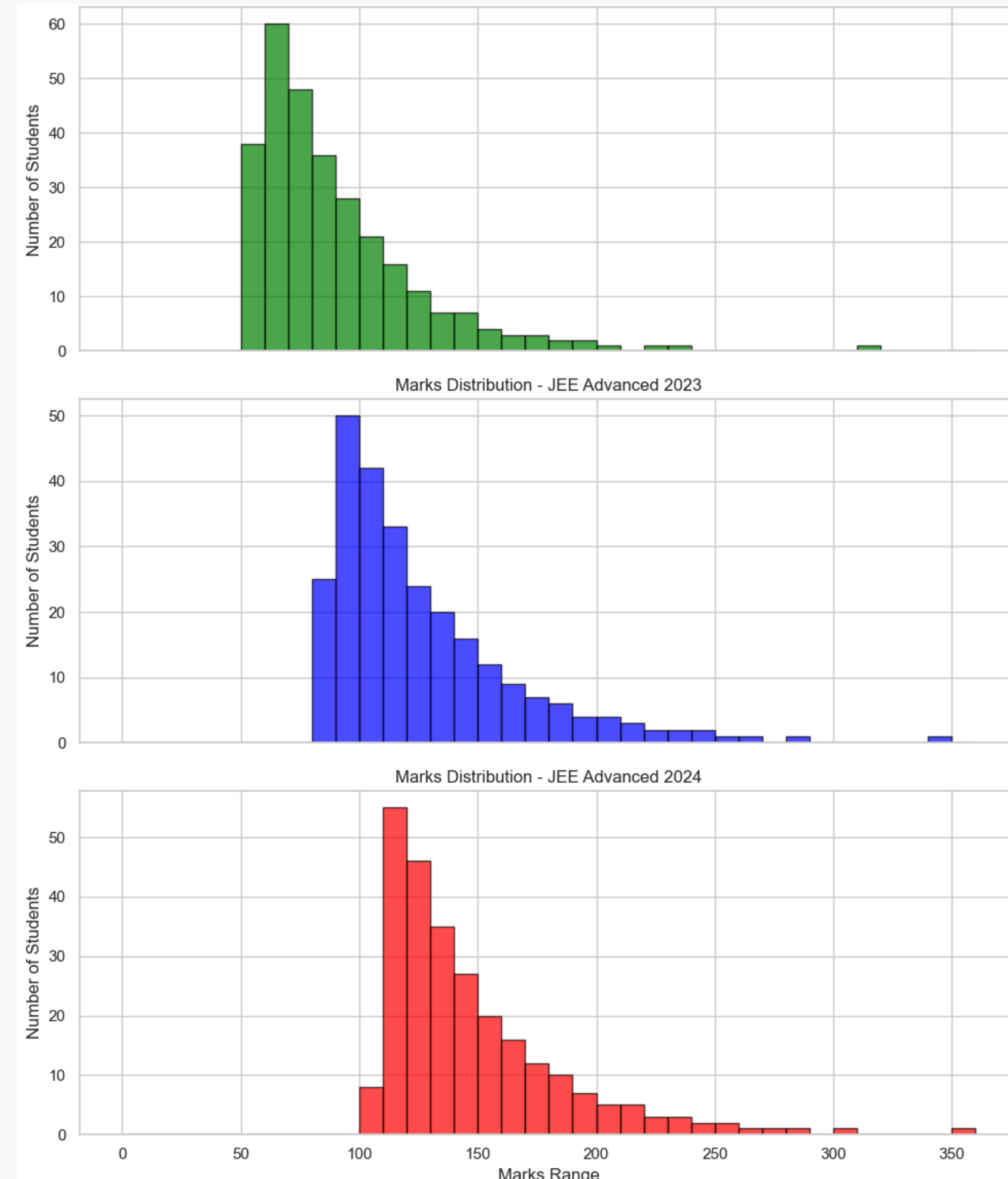
Analysis of Distribution of Marks



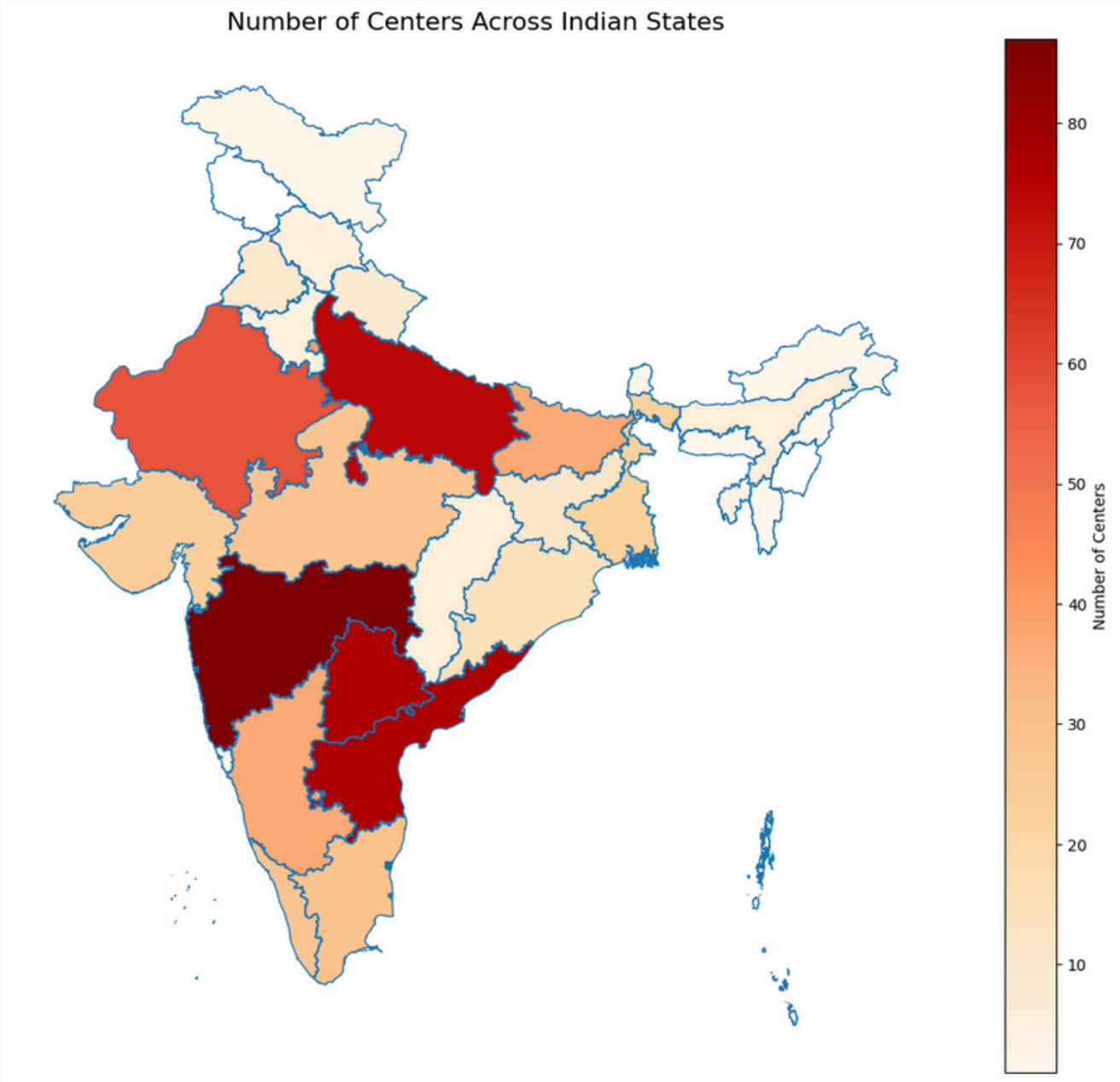
Observe the number of students obtaining
higher scores vs lower scores



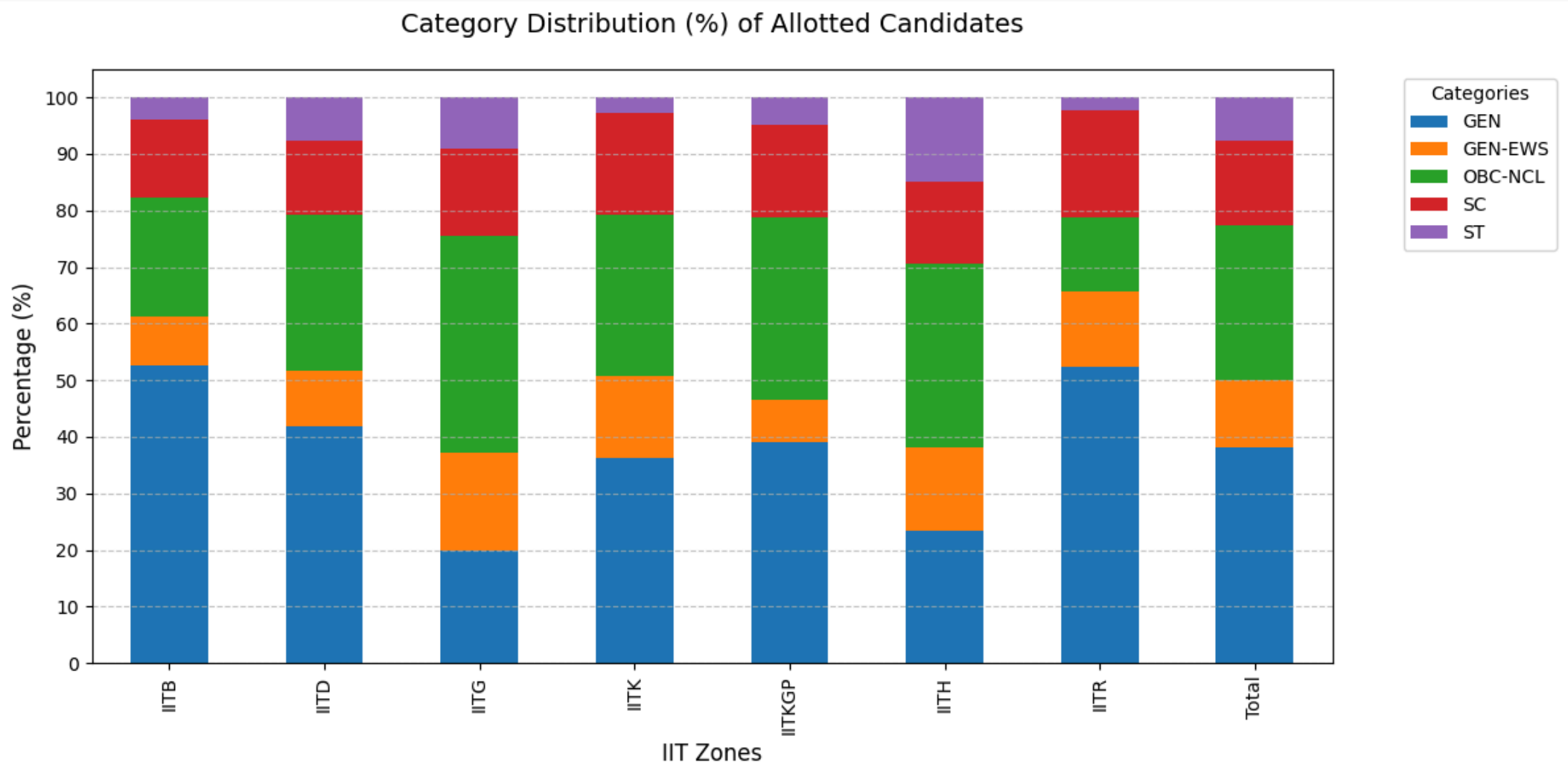
Comparision of Marks across 3 Years →



Centers and Category Distribution



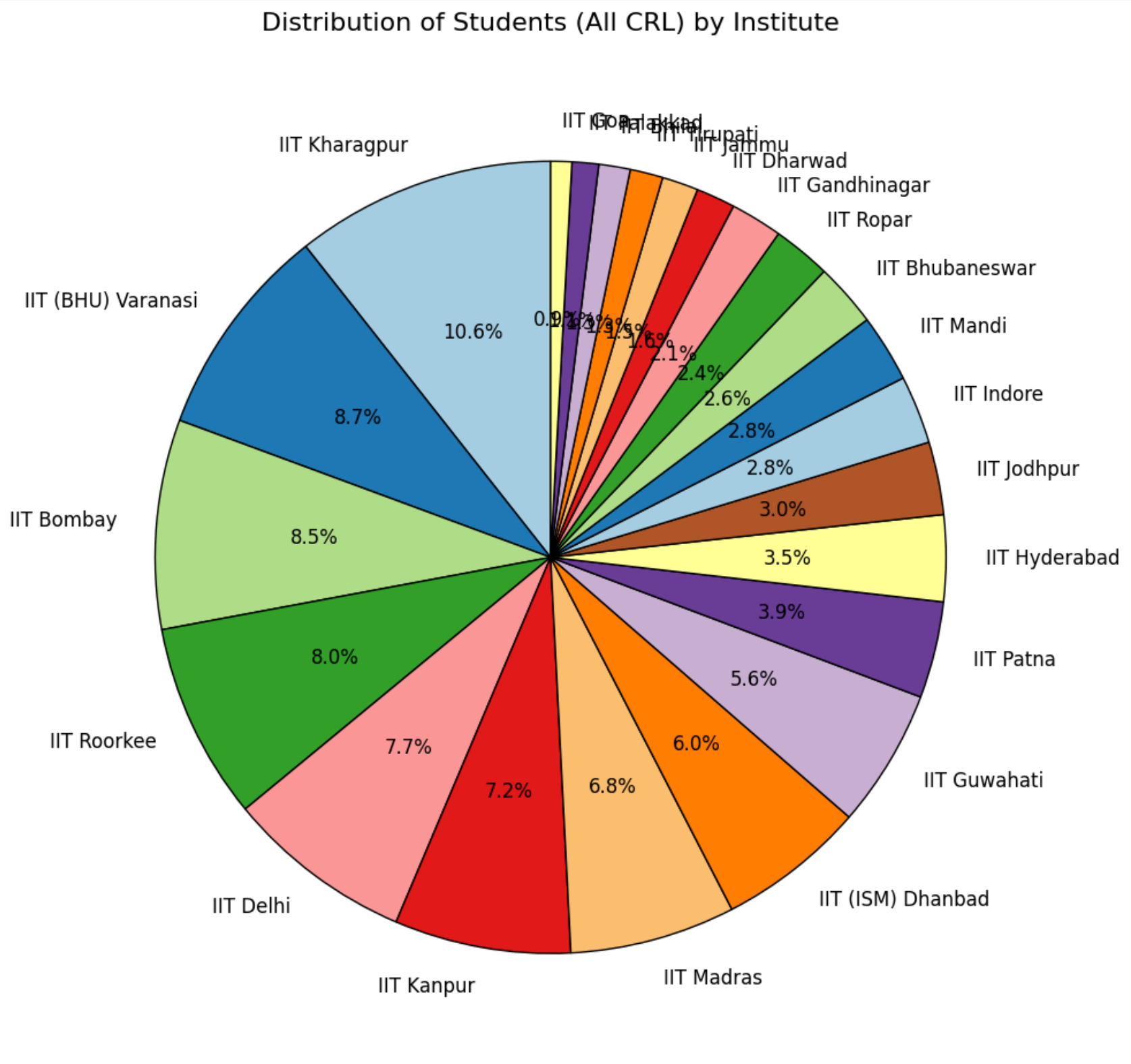
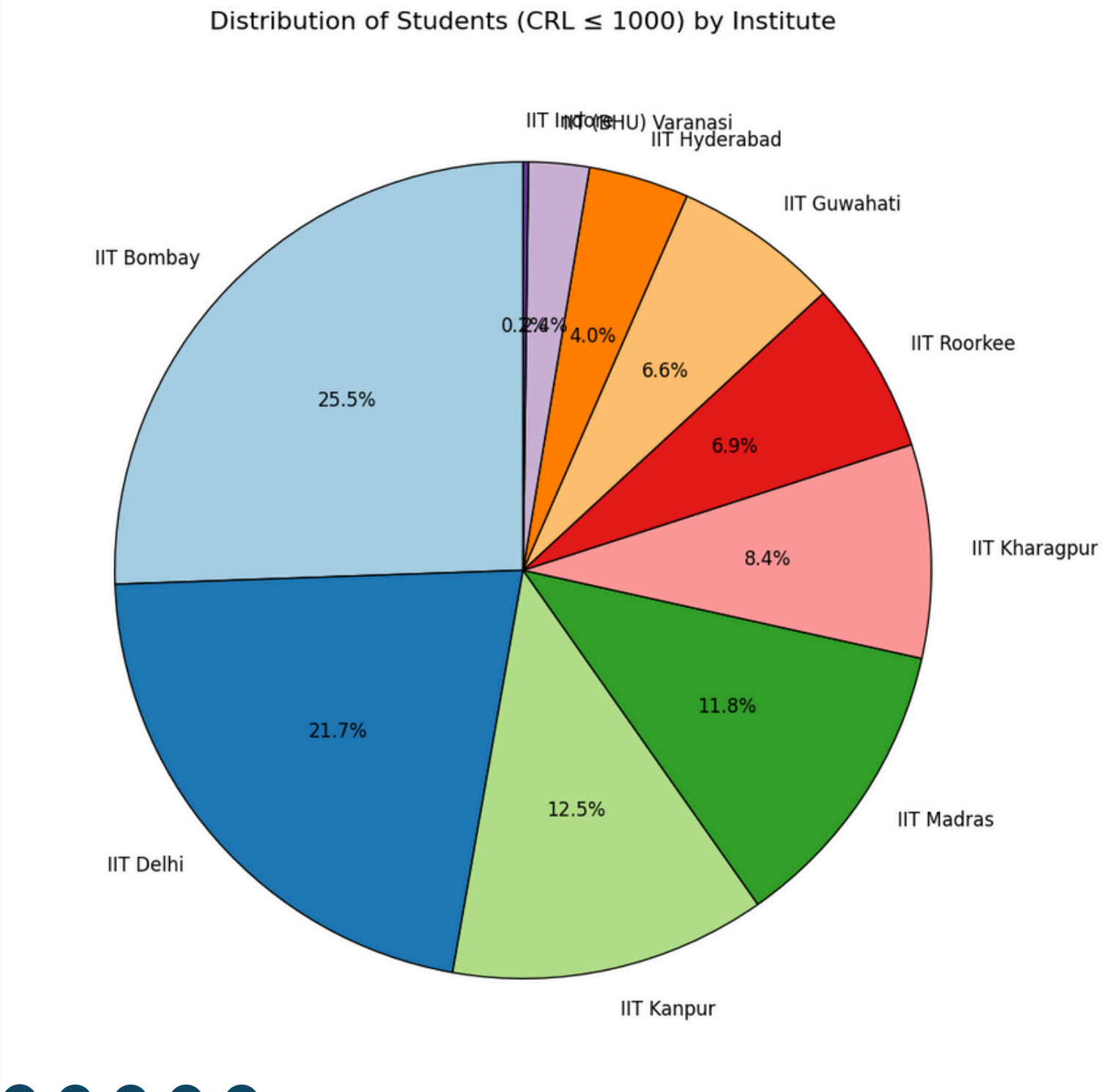
Maharashtra, Telangana, Andhra Pradesh, and UP have the most centers; Northern and North-East have very few centers.



Category distribution of allotted candidates by zones

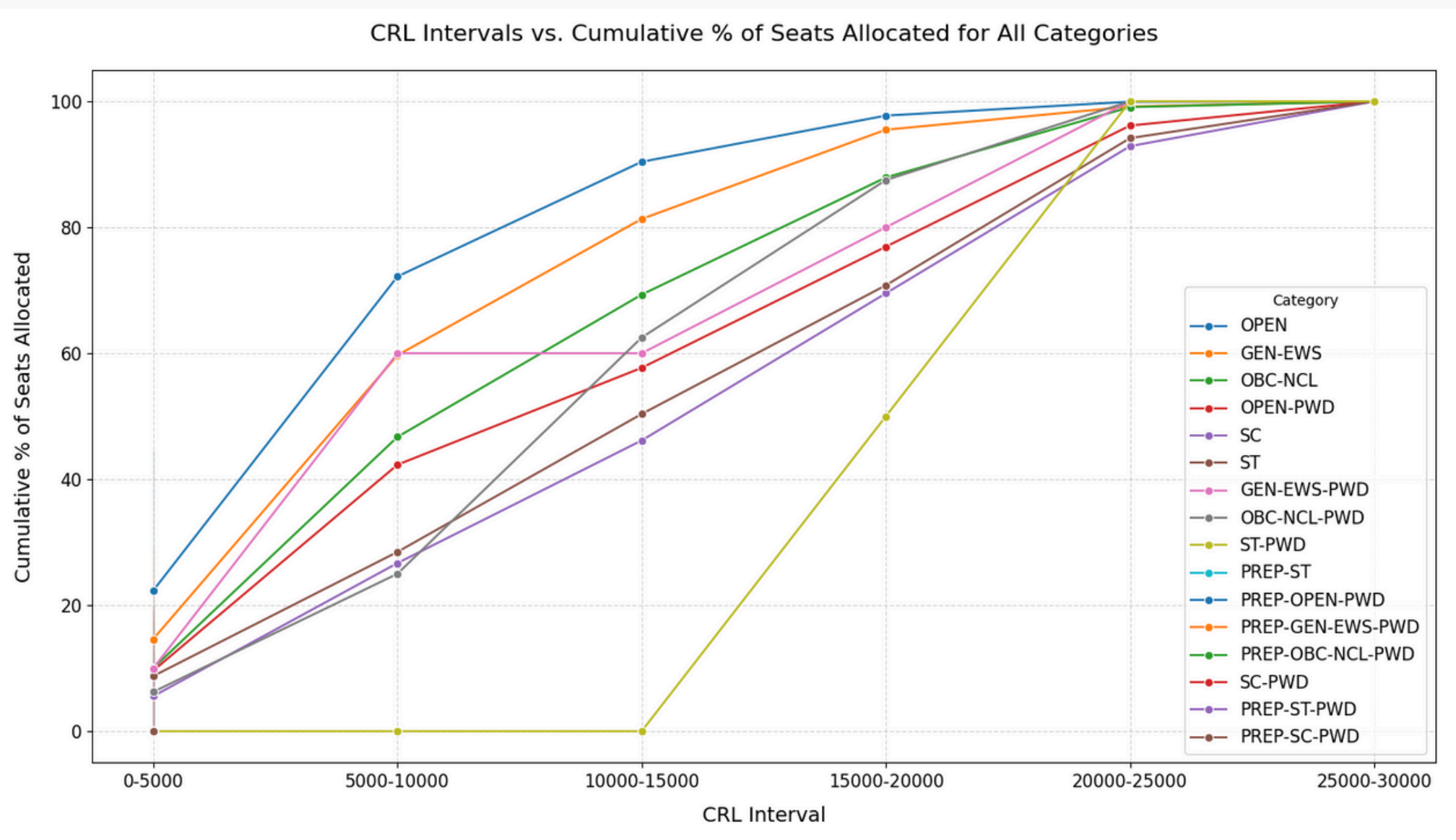
Seats Distribution

By Institute (CRL)



Seats Distribution

Among Categories



Among Various Engineering Disciplines

