NIRF Rank Prediction

Team Number 4

- Yachika N Kumar, B.Tech IT, Sem:2
- Srishti, B.Tech IT, Sem:2
- Sakshi Gupta, B.Tech MAE, Sem:2
- Amitasha Tembulkar, B.Tech CSE, Sem:4
- Unnati, B.Tech IT, Sem:2

Name of Mentor: Rishabh Kaushal Sir

MOTIVATION

Every year, academic institutions invest considerable effort and substantial resources to influence, predict and understand the decision-making choices of applicants who have been offered admission and the NIRF Rank of a particular institute plays a major role in helping students decide the ideal university for them. In this study, we applied several supervised machine learning techniques to four years of data on top 200 ranked engineering colleges in india , each with 31 associated features, to predict the NIRF Rank of the University.

By treating the question of what will be the rank of a university with given parametres as a regression problem, we implemented a number of different regression algorithms and then evaluated the performance of these algorithms using the metrics of Mean Absolute Error (MAE) , Root Mean Squared Error (RMSE) . The results from this study indicate that the Gradient Boosting Regressor performed best in modeling the college ranking problem, i.e., predicting the NIRF Rank of college, with an RMSE Value of 1.2. The significance of this research is that it demonstrates that many institutions could use machine learning algorithms to improve the accuracy of their estimates of entering class sizes, Teaching and Learning Resources , Research, professional practice and collaborative performance , fee, infra, sex ratio thus allowing more optimal allocation of resources and better control over revenue.

NIRF- A brief overlook

The National Institutional Ranking Framework (NIRF) was approved by the MHRD and launched by Honourable Minister of Human Resource Development on 29th September 2015.

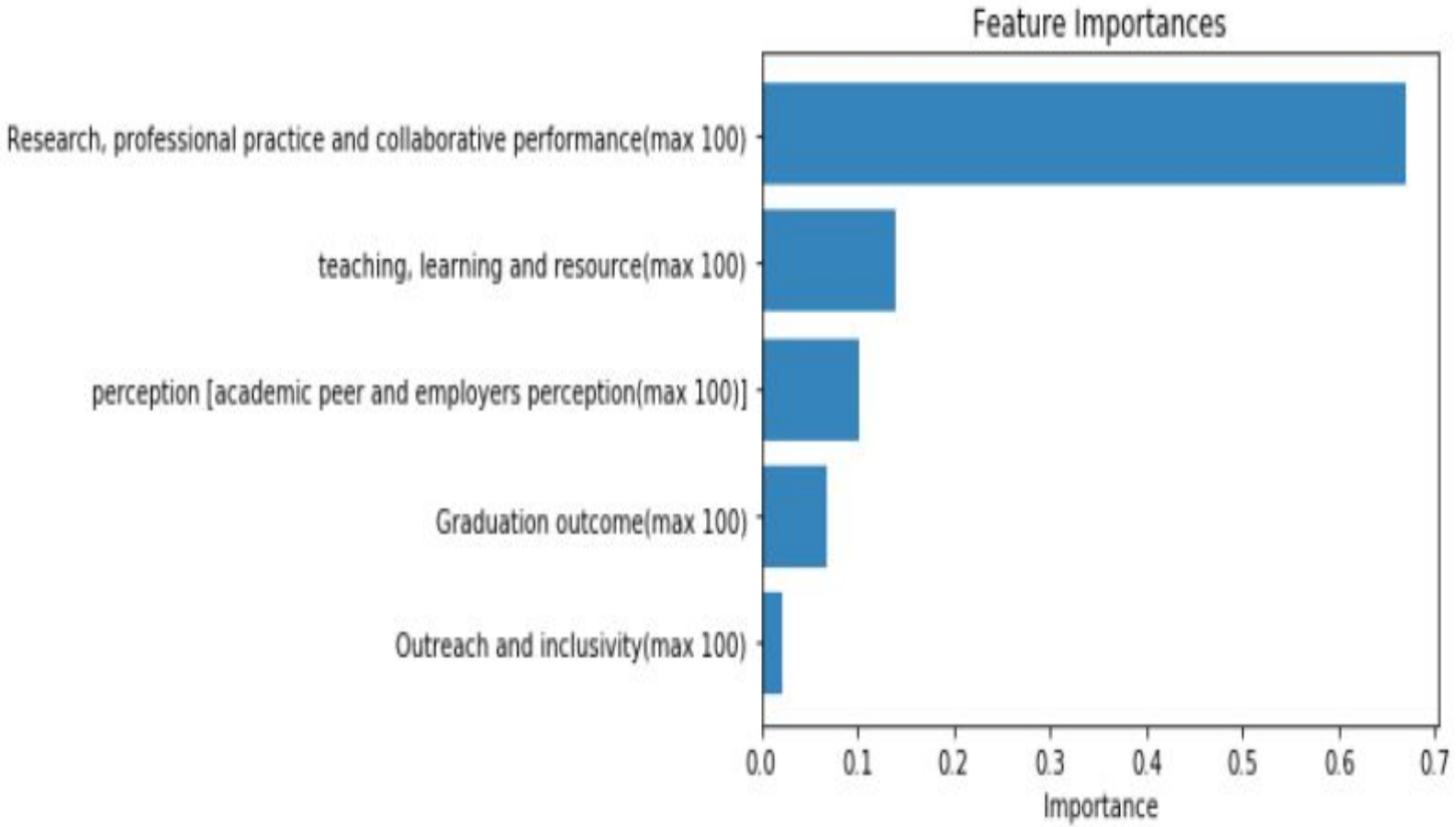
This framework outlines a methodology to rank institutions across the country. The methodology draws from the overall recommendations broad understanding arrived at by a Core Committee set up by MHRD, to identify the broad parameters for ranking various universities and institutions. The parameters broadly cover “Teaching, Learning and Resources,” “Research and Professional Practices,” “Graduation Outcomes,” “Outreach and Inclusivity,” and “Perception”.

India Rankings – 2016 based on this framework were released on 4th April 2016.

How it is determined?

The Framework uses several parameters for **ranking** purposes like resources, research, fees, infrastructure and etc but the major 5 used for calculating it are :

1. Research,Professional practice and collaborative performance
2. Teaching ,learning and resources
3. Perception
4. Graduation Outcome
5. Outreach and Inclusivity


Feature Importances
Research, professional practice and collaborative performance(max 100)
teaching, learning and resource(max 100)
perception [academic peer and employers perception(max 100)]
Graduation outcome(max 100)
Outreach and inclusivity(max 100)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Importance

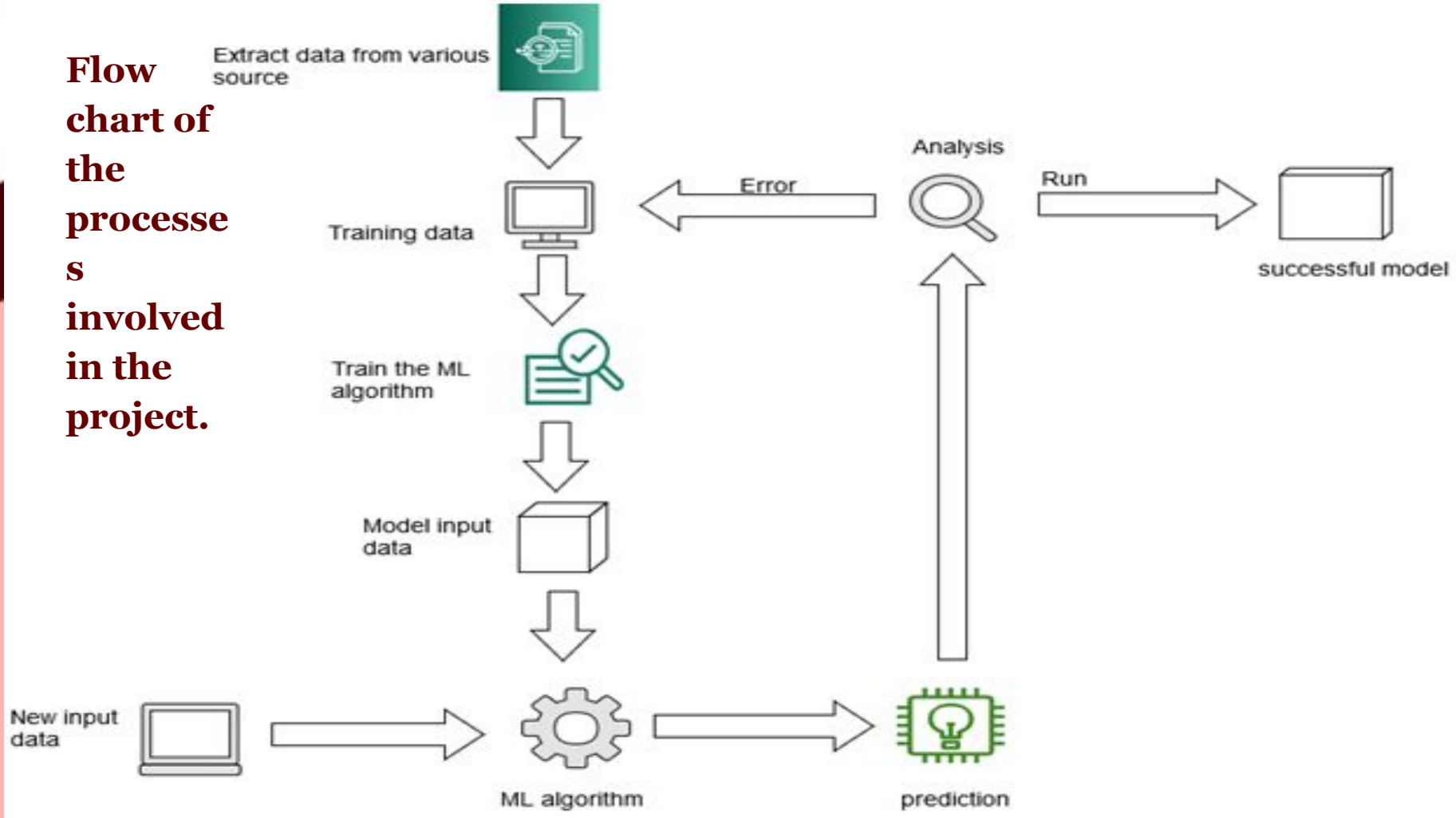
Problem Statement

The main objective of this project is to rank the engineering colleges in India on the basis of some parameters. It also ensures that there is no complacency in the colleges and they work harder each year to provide quality education to students.

Input:NIRF Parametres of 200 colleges

Output: Predicted NIRF Rank

Flow chart of the processes involved in the project.



Data Description

- We made our own dataset.This dataset contains 200 entries (200 rows and 31 columns) of different engineering colleges across India.

- We use data scraping, to extracts data from output generated from another program.

- We use the geocode in excel file to calculate the coordinates (longitude & latitude) with the help of physical addresses of colleges/universities.

DATASET (DATABASE)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	NIRF Rank	institute_	College N	closing ra	City	State/UT	Latitude	Longitude	score	link	Mode of A	Fees	Infrastruc	no. of mal	no. of fem	total stud	Graduatio	Outreach	perceptio	Research,
2	1	IR-E-U-045	Indian Ins	188	Chennai	Tamil Nad	12.99149	80.23369	89.93	https://ni	JEE Advan	8.82	4.6	1814	272	2086	83.9	61.31	100	94.64
3	20	IR-E-U-043	Amrita Sch	19971	Amritapur	Kerala	34.36764	-89.5391	57.37	https://ni	AEEE	12.1	4.2	7011	1949	8960	59	60.07	28.01	56.76
4	75	IR-E-U-049	National I	439270	Agratala	Tripura	23.84088	91.42142	39.98	https://ni	JEE Mains	11.04	4.2	2172	488	2660	48.22	55.37	10.27	21.55
5	137	IR-E-U-038	Sant Long	25607	Longowal	Punjab	30.20649	75.68914	34.27	https://ni	JEE Mains	2.73	4.3	1725	469	2194	27.04	55.09	2.69	17.63
6	132	IR-E-C-141	JSS Scienc	6768	Mysuru	Karnataka	12.31327	76.61343	34.81	https://ni	JEE Mains	1.08	4.4	2364	1356	3720	54.88	51.44	8.05	3.86
7	35	IR-E-U-041	malaviya I	24422	Jaipur	Rajasthan	26.86397	75.81079	52.25	https://ni	JEE Mains	7.08	4.4	2214	477	2691	59.89	54.4	8.95	41.69
8	199	IR-E-C-481	BVRIT Hyd	132751	Hyderaba	Telangana	17.5263	78.37023	31.1	https://ni	JEE Mains	3.6	4.6	0	1557	1557	53.74	52.98	1.63	0.46
9	140	IR-E-C-182	Yeshwant	26662	Nagpur	Maharash	21.09519	78.97857	34.07	https://ni	JEE Mains	6.04	3.9	2985	1650	4635	53.65	52.42	0	8.58
10	102	IR-E-U-050	Dayalbagh	18000	Agratala	Uttar Prad	27.22792	78.01378	37.51	https://ni	JEE Mains	1.51	3.4	754	326	1080	52.51	53.47	8.95	7.89
11	169	IR-E-C-133	BMS Instit	12879	Bengaluru	Karnataka	13.13336	77.56738	32.36	https://ni	JEE Mains	2.13	4.5	1869	872	2741	51.43	53.36	11.13	4.04
12	177	IR-E-U-014	Pandit De	10126	Gandhina	Gujarat	23.15588	72.66488	32.08	https://ni	JEE Mains	8.2	4.8	2745	401	3146	50.34	33.8	3.2	11.97
13	51	IR-E-U-047	Sathyabar	645376	Chennai	Tamil Nad	12.8719	80.21822	46.77	https://ni	JEE Mains	3	3.6	6602	2967	9569	64.14	64.2	2.69	31.77
14	9	IR-E-U-046	National I	19626	Tiruchirap	Tamil Nad	10.75894	78.81322	64.1	https://ni	JEE Mains	8	4.3	2719	662	3381	74.71	61.49	63.68	50.04
15	115	IR-E-U-018	Jaypee Un	300000	Solan	Himachal	31.01655	77.07016	36.55	https://ni	JEE Mains	3.5	4.2	1229	380	1609	54.81	46.44	7.59	18.76
16	80	IR-E-U-046	Karunya In	62000	Coimbato	Tamil Nad	10.93616	76.74406	39.4	https://ni	JEE Mains	14.48	4.4	3968	1521	5489	50.75	54.83	6.65	20.5
17	57	IR-E-U-001	Jawaharla	378	Hyderaba	Telangana	17.49327	78.39139	44.97	https://ni	TS EAMCE	16	4	2683	2028	4711	40.88	53.9	11.13	29.24
18	122	IR-E-C-658	Pondicher	9453	Puducher	Pondicher	12.01189	79.85685	35.49	https://ni	JEE Mains	1.55	4.1	1396	717	2113	1	40.61	12.79	15.73
19	58	IR-E-U-002	Koneru La	81383	Vaddeswa	Andhra Pr	16.44192	80.62253	44.7	https://ni	JEE Mains	4.8	4	8895	4211	13106	44.25	54.5	6.18	28.57
20	166	IR-E-U-088	Harcourt B	10990	Kanpur Na	Uttar Prad	26.49345	80.30742	32.69	https://ni	JEE Mains	5.4	2.8	1596	327	1923	46.82	42.34	6.65	2.31
21	81	IR-E-U-028	Indian Ins	8777	Jabalpur	Madhya P	23.17914	80.02734	39.29	https://ni	JEE Mains	10.56	3.9	1047	141	1188	59.4	43.47	1.63	30.89
22	79	IR-E-U-020	National I	14897	Jamshedp	Jharkhand	22.77702	86.14412	39.44	https://ni	JEE Mains	14.32	3.4	2115	297	2412	64.87	45.15	11.13	12.85
23	160	IR-E-C-197	SR Engine	13850	Warangal	Telangana	18.08927	79.46726	32.95	https://ni	JEE Mains	3.8	4.1	1540	1338	2878	31.2	62.13	12.38	1.91
24	129	IR-E-U-077	DIT Univer	100000	Dehradun	Uttarakha	30.3983	78.07512	34.93	https://ni	JEE Mains	11.2	4.1	2677	1324	4001	43.77	59.09	3.71	7.54
25	104	IR-E-U-038	Punjab Te	22573	Kapurthal	Punjab	31.3534	75.45868	37.1	https://ni	JEE Mains	3.08	4.2	357	59	416	13.06	43.6	2.69	31

ALGORITHMS USED

1. Linear Regression
2. Support Vector Regressor
3. Decision Tree Regressor
4. Random Forest regressor
5. K Neighbours Regressor
6. AdaBoost Regressor
7. Gradient Boosting Regressor
8. Lasso Regression
9. Ridge Regression
10. Bayesian Ridge Regression
11. Elastic Net Regression
12. Huber Regressor
13. Artificial Neural Network

RMSE VALUES OF ALGORITHMS

```
Results...
DecisionTree : 2.779388421937459
Linear Regression : 226.98758143157002
RandomForest : 1.57284931464524
KNeighbours : 53.6631857049132
SVM : 48.294805061779385
AdaBoostClassifier : 2.3506943046824675
GradientBoostingClassifier:  1.2454620546251003
Lasso:  26.35109941441744
Ridge:  30.613707270286802
BayesianRidge:  30.489933140139836
ElasticNet:  30.593499386681298
HuberRegressor:  34.47847597343498
ANN : 14.757715274210618
```

Data Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps. It provides us with a quick, clear understanding of the information.

Libraries used here:

→Matplotlib

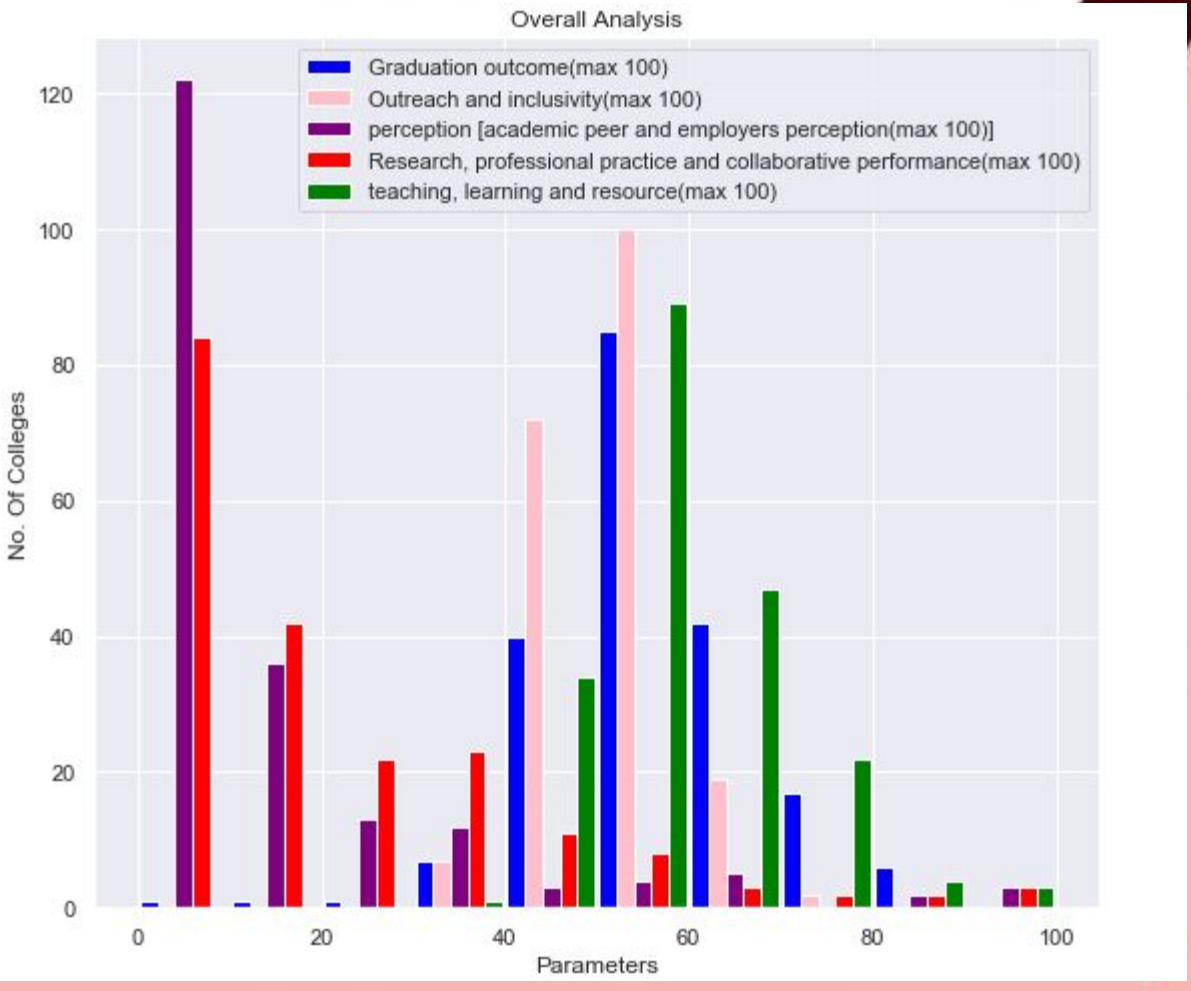
→seaborn

→geopandas

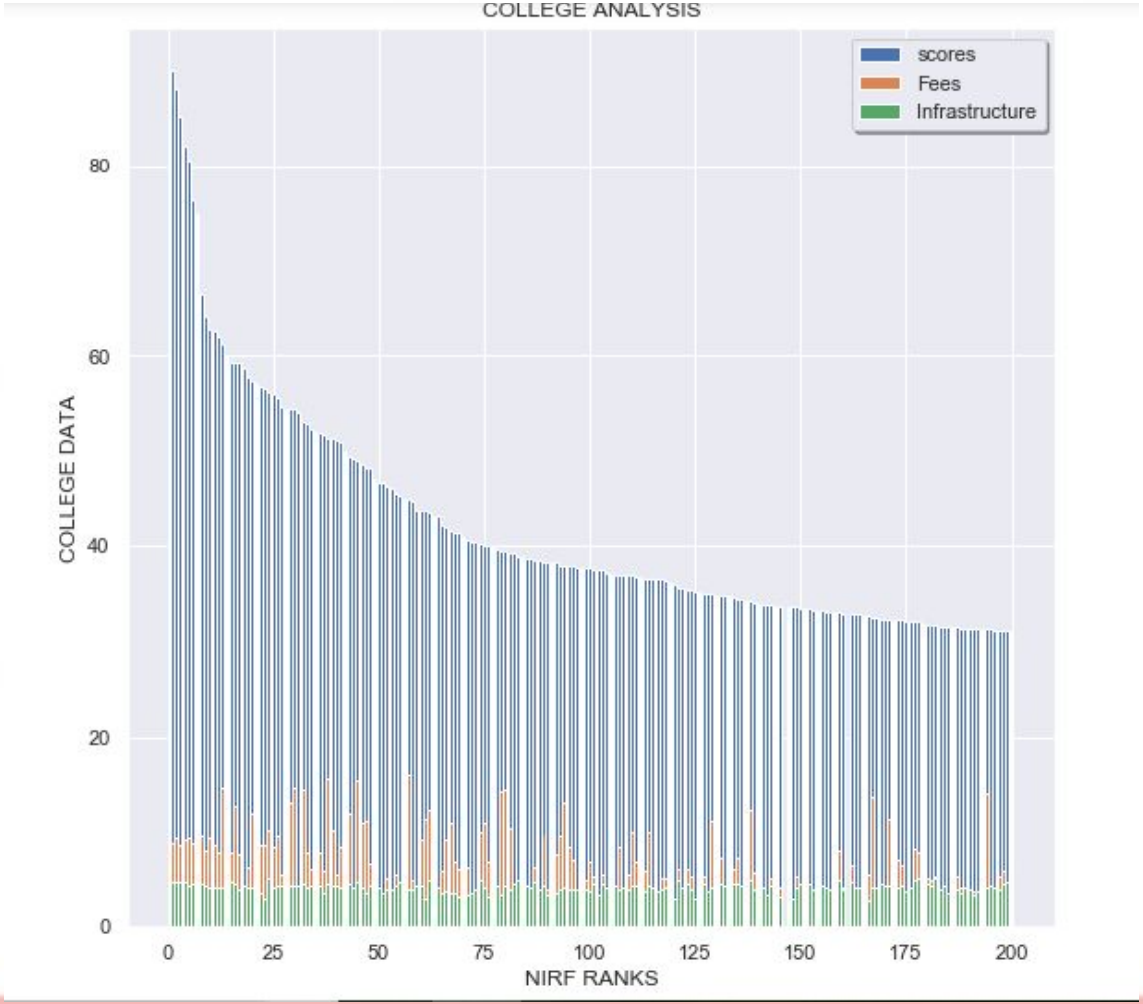
→folium

→geopy

♦ This histogram represents the important parameters which are required for nirf ranking. Because of this analysis colleges will improve themselves to get better nirf rank.

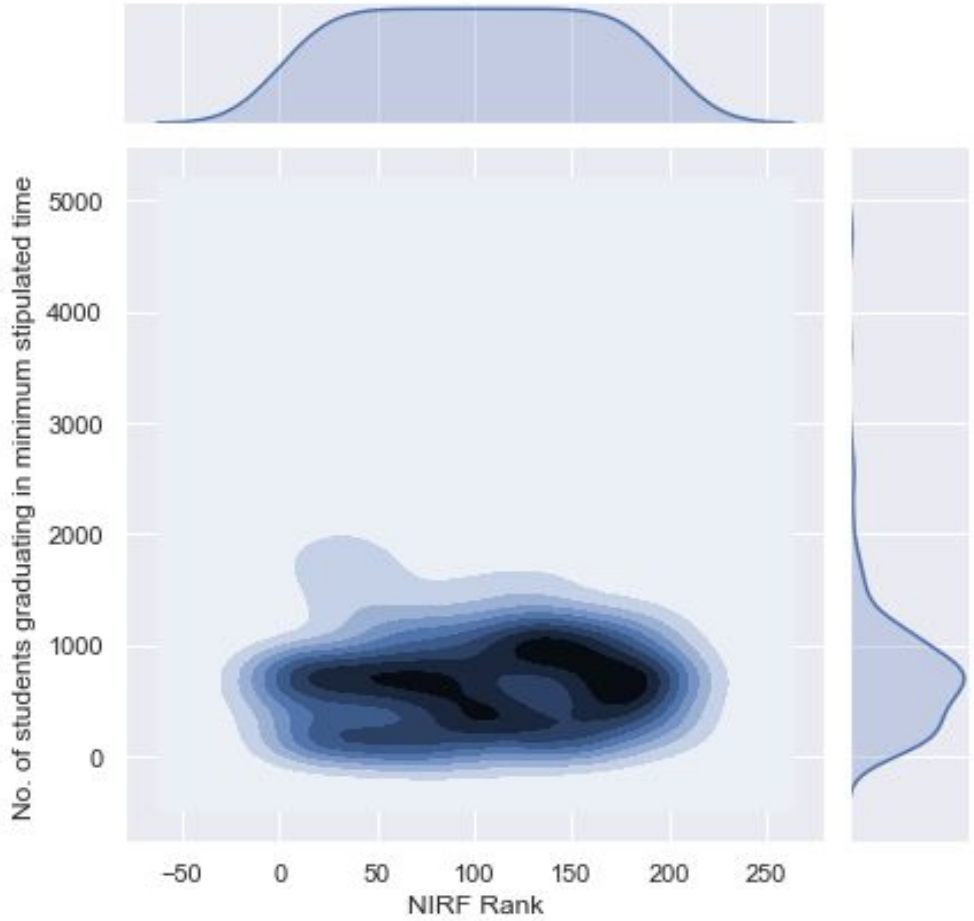


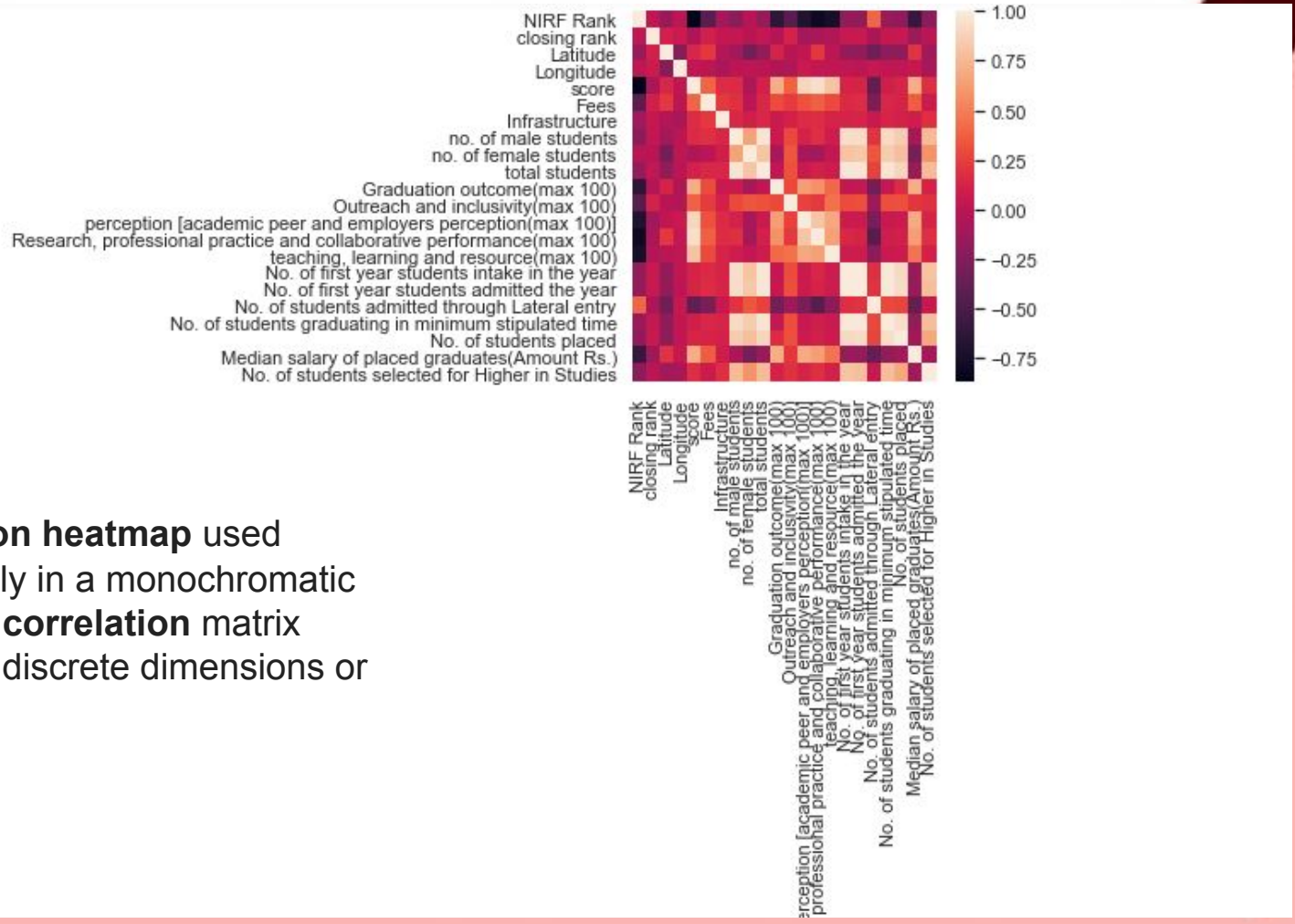
- This bar graph represents the basic details of colleges/universities in India such as scores, fees, infrastructure with nirf ranks. So that students will get to know more about their colleges/universities.



- **sns.jointplot** of **kind = ‘kde’** is used to show the joint distribution between two different parameters, along with the associated marginal distributions.

- shows the capabilities of no. of students graduating in minimum stipulated time.

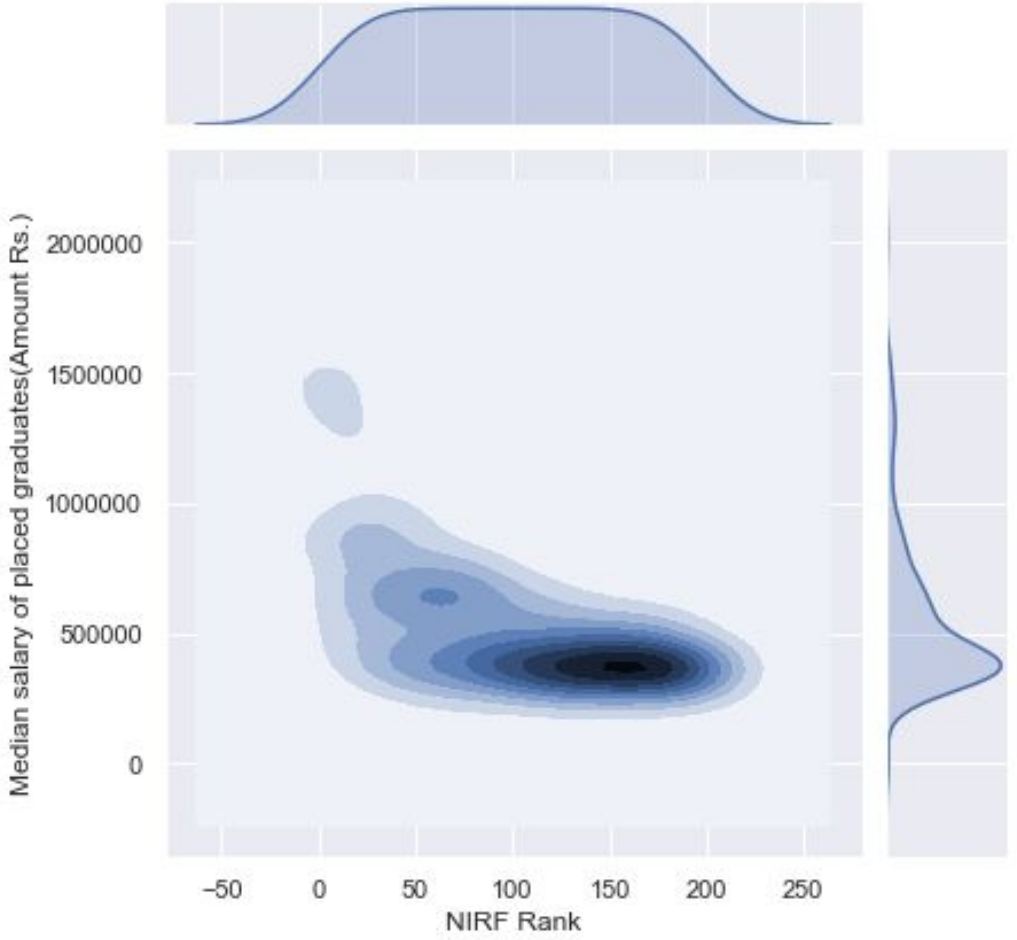




- This is **correlation heatmap** used colored cells, typically in a monochromatic scale, to show a 2D **correlation** matrix (table) between two discrete dimensions or event types.

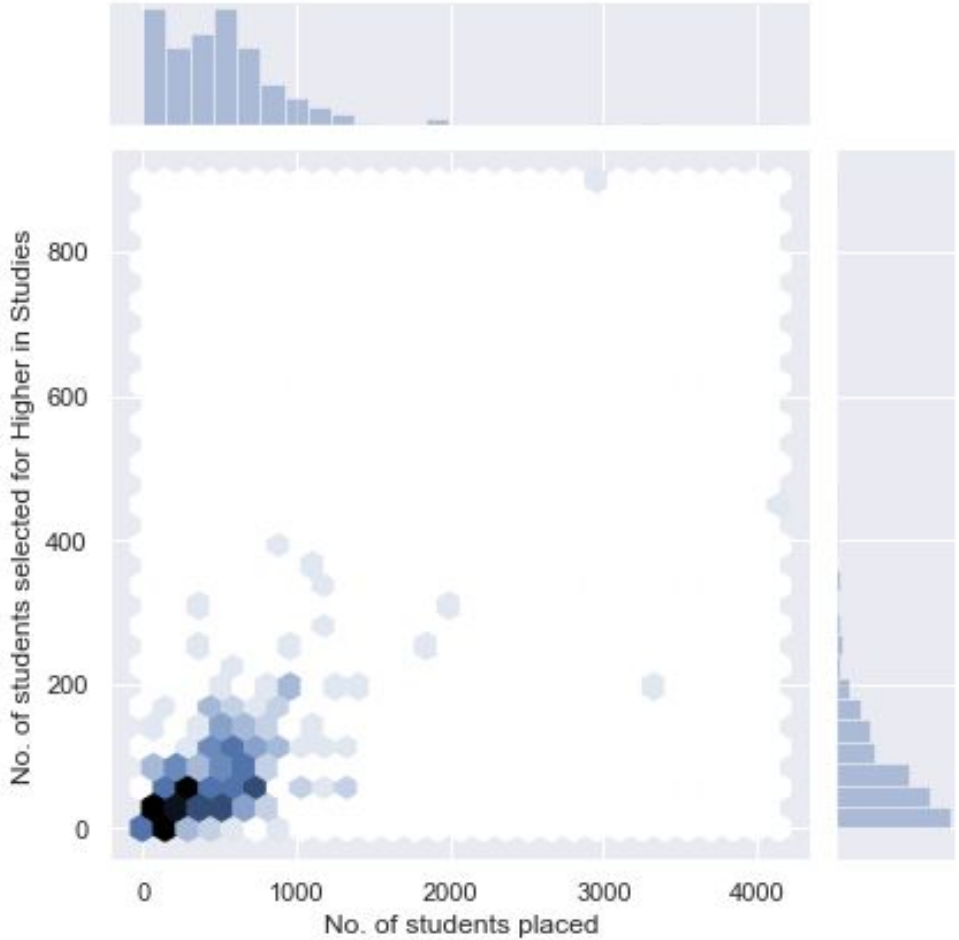
♦ **sns.jointplot** of **kind = ‘kde’** is used to show the joint distribution between two different parameters, along with the associated marginal distributions.

♦ shows the idea of salaries after placement as per nirf ranks.



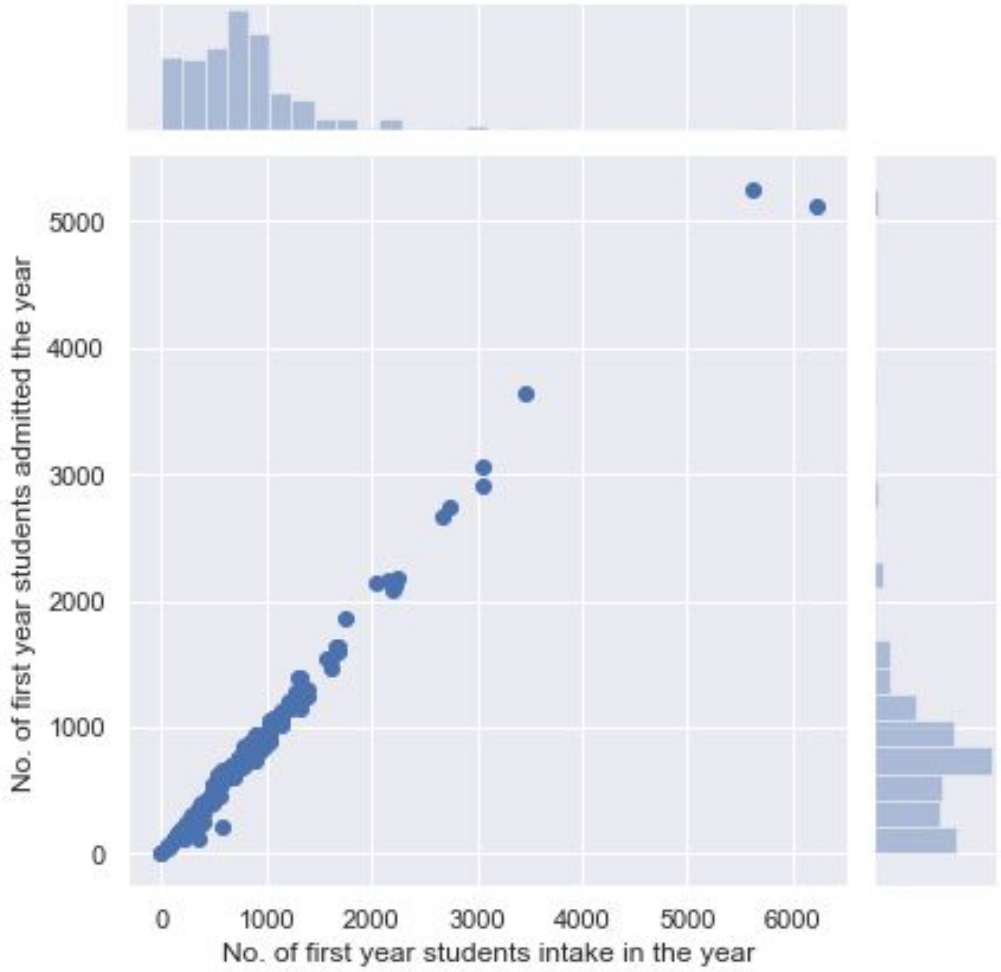
♦ **sns.jointplot** of **kind = ‘hex’** is used to show the joint distribution between two different parameters, along with the associated marginal distributions.

♦ shows the variation between no. of students placed in the company or selected for higher education from their respective colleges/universities.

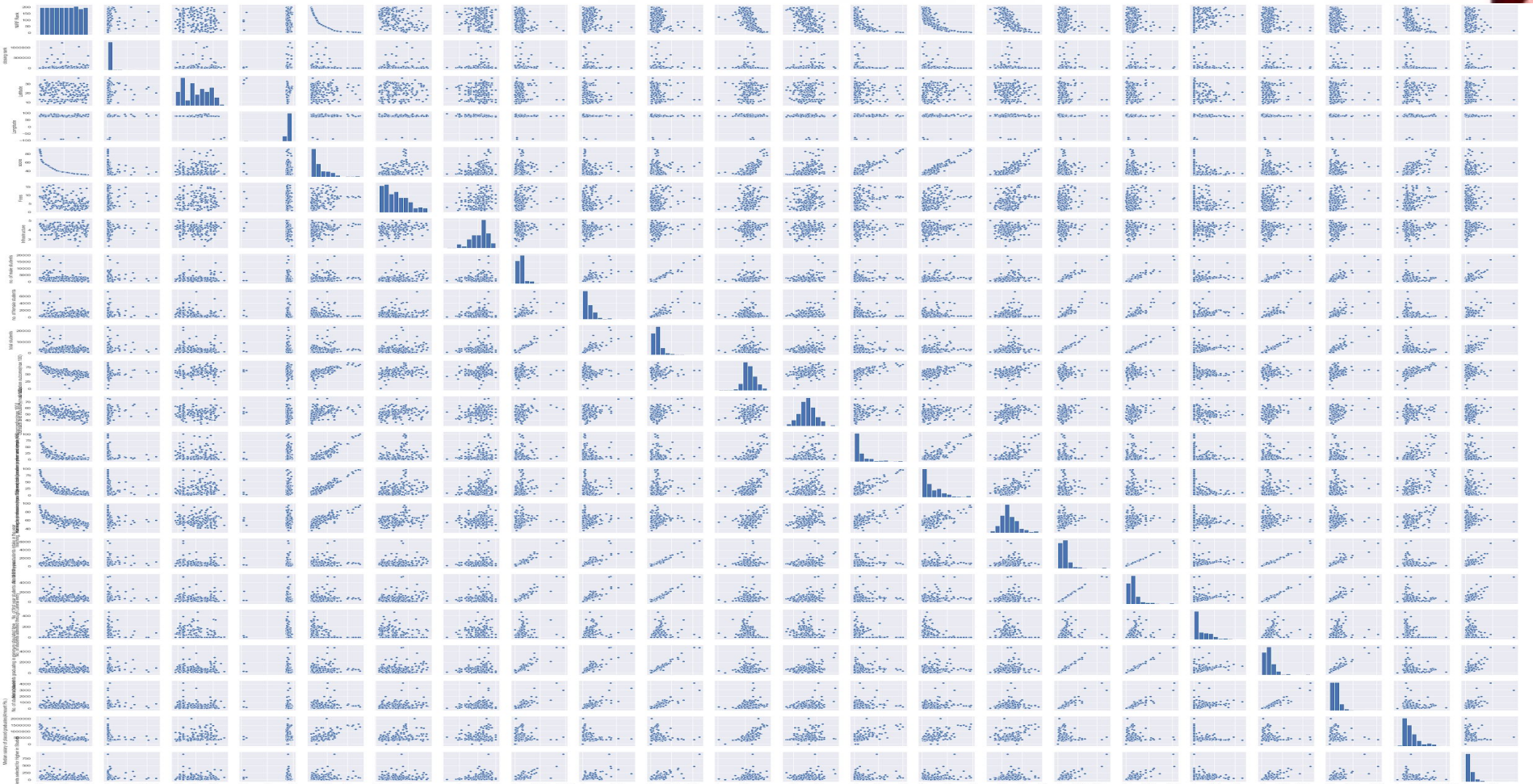


- **sns.jointplot** is used to show the joint distribution between two different parameters, along with the associated marginal distributions.

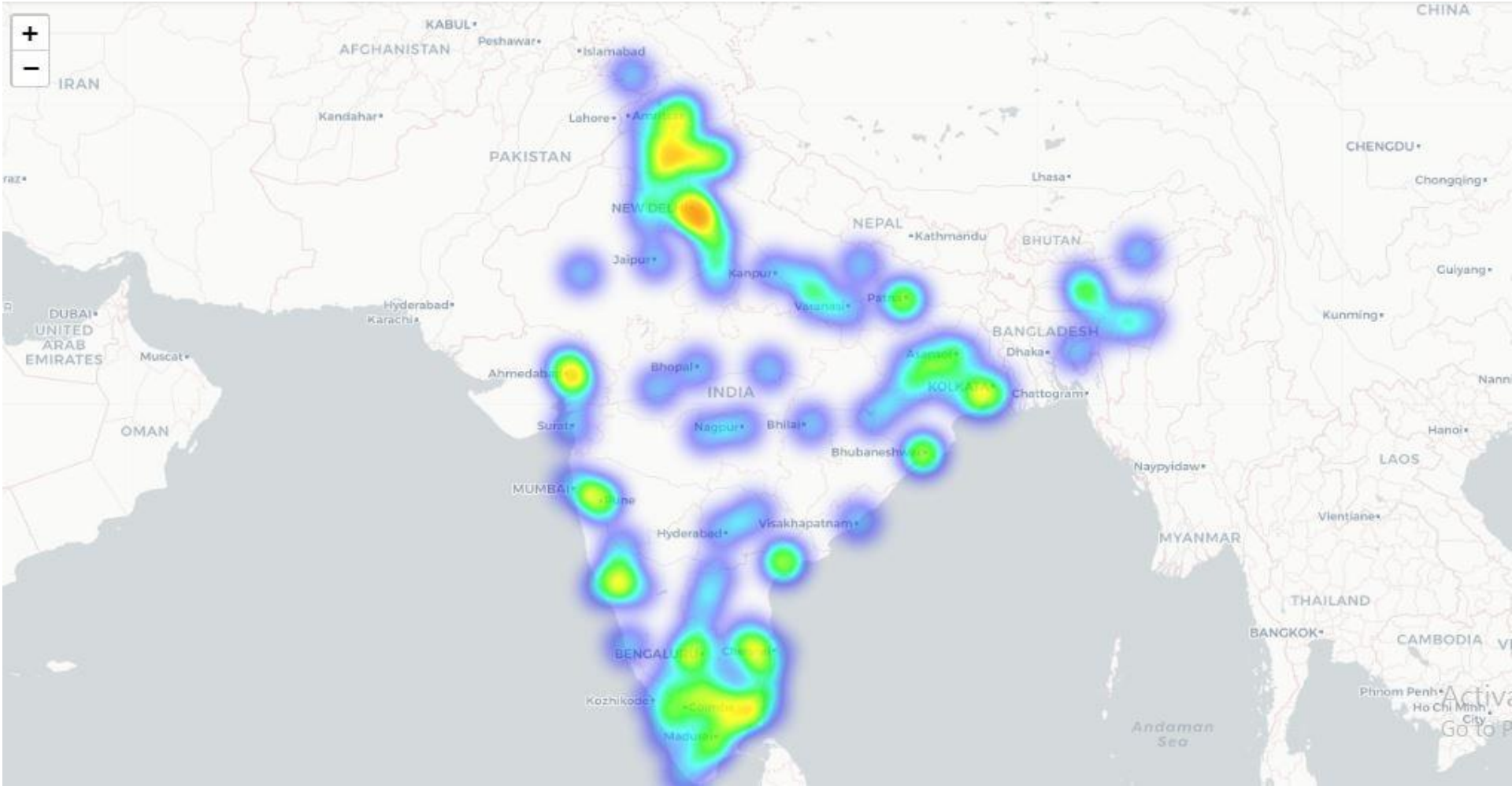
- shows how much students enrolled among all in their respective colleges.



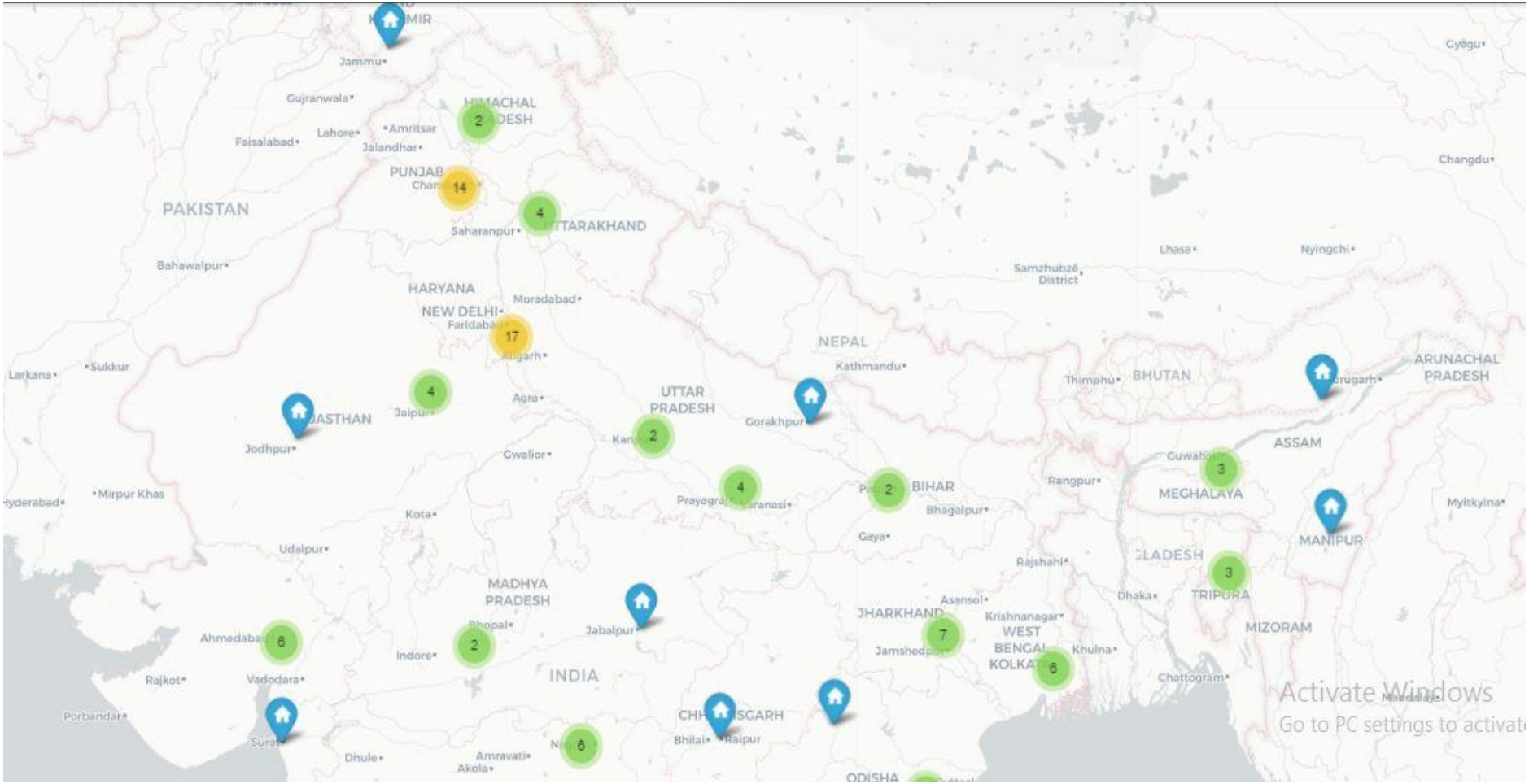
♦ sns.pairplot is used to plot a pairwise relationship in a dataset.



♦ Heatmap, is the system of color-coding used to represent the location of Indian engineering colleges.



♦ cluster map is used here to represent the no. of colleges/universities with their locations in India.



RESULTS

DecisionTree : 2.779388421937459
Linear Regression : 226.98758143157002
RandomForest : 1.57284931464524
KNeighbours : 53.6631857049132
SVM : 48.294805061779385
AdaBoostClassifier : 2.3506943046824675
GradientBoostingClassifier: 1.2454620546251003
Lasso: 26.35109941441744
Ridge: 27.747951134149346
BayesianRidge: 26.34191356074623
ElasticNet: 26.49396577365189
HuberRegressor: 99.2078402165999
ANN: 14.757715274

ACKNOWLEDGEMENT

We express our deepest gratitude to our mentor for his kind guidance during the entire period of project. His consistent support and advices have helped us refine our skills and successfully complete our Summer Training Project.

References

https://www.nirfindia.org/Parameter

https://www.nirfindia.org/2020/EngineeringRanking.html

https://www.youtube.com/watch?v=dKzpOqTRtkk&t=4s