IBM SkillsBuild

Data Analytics Internship Programme

ALL CHARTS, GRAPHS AND
IMPORTANT VISUALIZATIONS

PROJECT NAME :

A Comprehensive Analysis of Infrastructure and Technology in achieving Quality Education

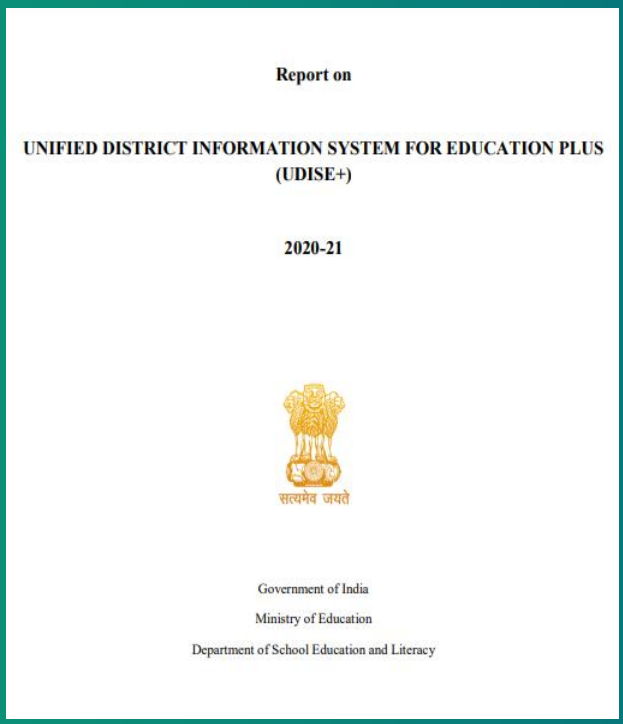
Unique ID : IBM3013

Team Name : AI Titans

College Name : Odisha University of Technology & Research, Bhubaneswar

DATA SOURCE

LINK TO SOURCE : CLICK HERE



Report on

UNIFIED DISTRICT INFORMATION SYSTEM FOR EDUCATION PLUS
(UDISE+)

2020-21

सत्यमेव जयते

Government of India

Ministry of Education

Department of School Education and Literacy

LINK TO FILE UPLOADED ON KAGGLE : CLICK HERE

Infrastructure_Technology_Education_India.xlsx (104.58 kB)

About this file

This file does not have a description yet.

Table	Total Rows	Total Columns
dig_lib_tbl	37	16
drop_tbl	37	10
enroll_tbl	37	8
infratech1_tbl	37	13

Data Explorer

Version 3 (104.58 kB)

- Infrastructure_Technology
  - dig_lib_tbl
  - drop_tbl
  - enroll_tbl
  - infratech1_tbl
  - infratech2_tbl
  - projector_tbl
  - prom_tbl
  - rep_tbl
  - school_no_tbl
  - smart_class_tbl
  - trans_tbl

DATA PREPROCESSING

Cleaning and Transformation

```
#school no tbl
school_no_tbl=school_no_tbl.drop(school_no_tbl.columns[2:],axis=1)
print('school_no_tbl:',school_no_tbl.columns.tolist())
#enroll_tbl
enroll tbl=enroll tbl.drop(enroll tbl.columns[2:],axis=1)
enroll_tbl.rename(columns={'Total':'Enrollment'},inplace=True)
print('enroll_tbl:',enroll_tbl.columns.tolist())
#prom_tbl
prom_tbl['Promotion Rate'] = prom_tbl[['P Total', 'UP Total', 'S
Total']].mean(axis=1)
prom_tbl=prom_tbl[['India/ State /UT','Promotion Rate']]
print('prom_tbl:',prom_tbl.columns.tolist())
#drop_tbl
drop tbl['Dropout Rate'] = drop_tbl[['P Total','UP Total','S
Total']].mean(axis=1)
drop tbl=drop tbl[['India/ State /UT','Dropout Rate']]
print('drop_tbl:',drop_tbl.columns.tolist())
#trans_tbl
trans tbl['Transition Rate'] = trans_tbl[['P Total','UP Total','S
Total']].mean(axis=1)
trans tbl=trans tbl[['India/ State /UT','Transition Rate']]
print('trans_tbl:',trans_tbl.columns.tolist())
#rep tbl
rep_tbl['Repetition Rate'] = rep_tbl[['P Total','UP Total','S
Total']].mean(axis=1)
rep_tbl=rep_tbl[['India/ State /UT','Repetition Rate']]
print('rep_tbl:',rep_tbl.columns.tolist())
#projector tbl
projector_tbl=projector_tbl[['India/ State /UT','All management (%)']]
projector tbl.rename(columns={'All management
(%)':'Projector'},inplace=True)
print('projector_tbl:',projector_tbl.columns.tolist())
#smart_class_tbl
smart_class_tbl=smart_class_tbl[['India/ State /UT','All management
```

```
(%)']]
smart class tbl.rename(columns={'All management (%)':'Smart
Class'},inplace=True)
print('smart_class_tbl:',smart_class_tbl.columns.tolist())
#dig lib tbl
dig lib tbl=dig lib tbl[['India/ State /UT','All management (%)']]
dig lib tbl.rename(columns={'All management (%)':'Digital
Library'},inplace=True)
print('dig lib tbl:',dig_lib_tbl.columns.tolist())
#infratech1_tbl
infratech1_tbl=infratech1_tbl.drop(infratech1_tbl.columns[[4,6,8,10,12
]],axis=1)
numeric cols =
infratech1 tbl.select dtypes(include=['number']).columns
infratech1 tbl[numeric cols] =
infratech1 tbl[numeric_cols].astype(float)
for i in range(2,8):

infratech1_tbl.iloc[:,i]=(infratech1_tbl.iloc[:,i]/infratech1_tbl.iloc
[:,1])*100
infratech1_tbl=infratech1_tbl.drop(infratech1_tbl.columns[[1]],axis=1)
infratech1 tbl.rename(columns={'Library/ Book Bank/ Reading
Corner':'Library',"Functional Girls' Toilet":"Girls'
Toilet","Functional Boys' Toilet":"Boys' Toilet","Functional
Electricity":"Electricity"},inplace=True)
print('infratech1_tbl:',infratech1_tbl.columns.tolist())
#infratech2 tbl
infratech2_tbl=infratech2_tbl.drop(infratech2_tbl.columns[[2,5,8,9]],a
xis=1)
numeric cols =
infratech2 tbl.select dtypes(include=['number']).columns
infratech2 tbl[numeric cols] =
infratech2 tbl[numeric_cols].astype(float)
for i in range(2,9):

infratech2_tbl.iloc[:,i]=(infratech2_tbl.iloc[:,i]/infratech2_tbl.iloc
[:,1])*100
infratech2 tbl=infratech2 tbl.drop(infratech2 tbl.columns[[1]],axis=1)
infratech2_tbl.rename(columns={'functional Computers used for
pedagogical purposes':'Computers',"Hand wash facility":"Hand
wash","Functional Drinking Water":"Drinking Water","Internet
Facility":"Internet","Schools with CWSN\nToilet facilities":"CWSN
Toilet"},inplace=True)
print('infratech2_tbl:',infratech2_tbl.columns.tolist())
```

FINAL MODIFIED DATA

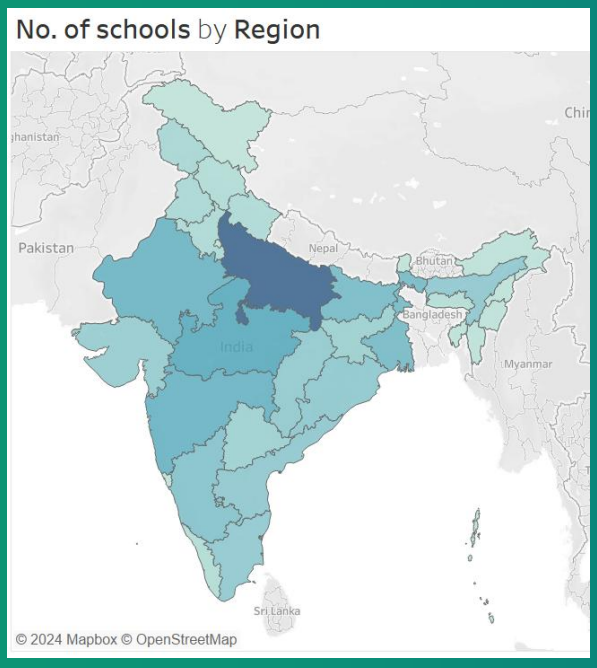
LINK TO FINAL DATA: CLICK HERE

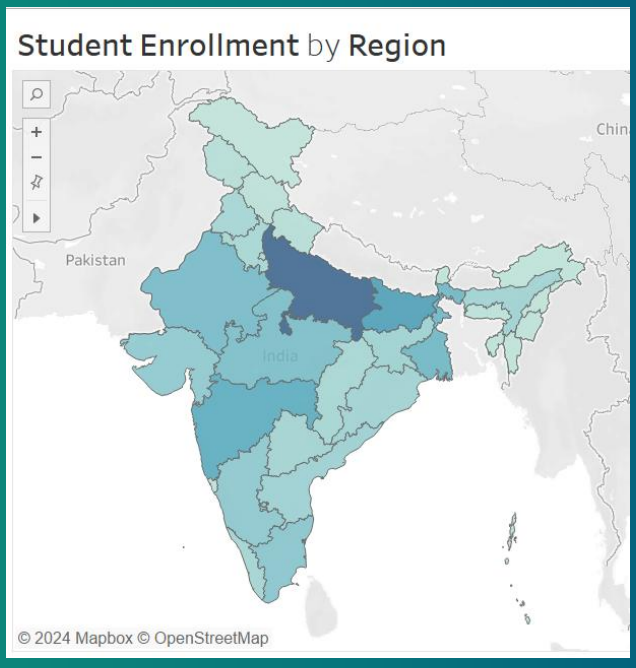
	A	B	C	D	E	F	G	H	I
1	India/ State /UT	Total	Enrollment	Promotion Rate	Dropout Rate	Transition Rate	Repetition Rate	Projector	Smart Cl
2	India	1489115	265235830	93.56666667	5.7	86.8	0.766666667	16.7	1
3	Andaman and Nicobar	416	73861	97.83333333	2.133333333	98.33333333	0.066666667	27.6	3
4	Andhra Pradesh	61948	8244647	94	5.966666667	89	0.166666667	23.8	
5	Arunachal Pradesh	3603	354382	89.1	9.233333333	89.6	1.7	22.4	
6	Assam	60859	7544960	87.76666667	11.7	82.26666667	0.533333333	5.9	
7	Bihar	93165	27472692	91.6	8.366666667	74.06666667	0.066666667	3.2	
8	Chandigarh	233	268627	100	0	107.6666667	0.366666667	85	4
9	Chhattisgarh	56512	5992197	94.93333333	4.866666667	90.9	0.233333333	9.3	
10	Delhi	5619	4572107	97.66666667	1.6	98.5	0.733333333	60.8	4
11	Goa	1510	304982	96.5	3	96.76666667	0.533333333	41.1	
12	Gujarat	53851	11542276	92.43333333	7.633333333	85.7	0	38.5	2
13	Haryana	23726	6035679	97.63333333	2.033333333	96.46666667	0.4	26.7	1
14	Himachal Pradesh	18028	1437022	99.3	0.7	97.93333333	0.033333333	16.4	1
15	Jammu and Kashmir	28805	2718644	95.53333333	4.333333333	91.86666667	0.166666667	12.6	
16	Jharkhand	44855	7970750	94.76666667	5	87.66666667	0.266666667	7.4	
17	Karnataka	76450	12092381	93.36666667	5.266666667	88	1.6	22.5	1
18	Kerala	16240	6423120	98.16666667	1.833333333	96.5	0	82.3	3
19	Ladakh	978	59788	95.6	4.166666667	94.4	0.233333333	19	1
20	Lakshadweep	38	13586	99	1.033333333	98.93333333	0	84.2	4
21	Madhya Pradesh	125582	16169265	91.93333333	7.333333333	87.16666667	0.7	6.9	
22	Maharashtra	109605	22586695	95.76666667	4.066666667	91.9	0.166666667	46.6	1
23	Manipur	4617	693194	93.16666667	6.733333333	91.66666667	0.133333333	10.8	
24	Meghalaya	14600	1169720	83.6	14.03333333	80.53333333	2.3	5	

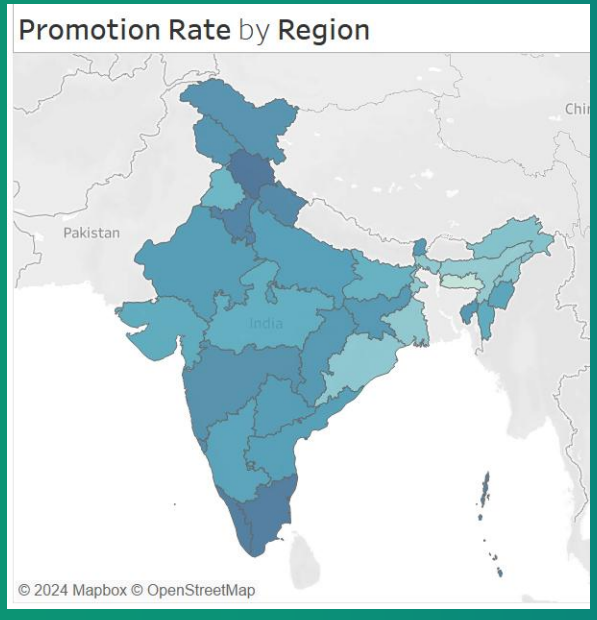
Sheet1 +

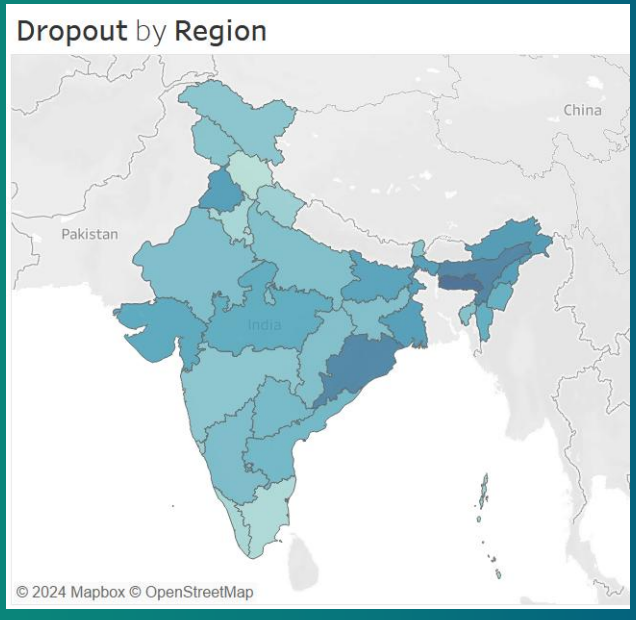
```
# Reviewing the final dataset characteristics
final_set.info()
final_set.describe()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 23 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   India/ State /UT    36 non-null     object
 1   Total               36 non-null     int64
 2   Enrollment          36 non-null     int64
 3   Promotion Rate      36 non-null     float64
 4   Dropout Rate        36 non-null     float64
 5   Transition Rate     36 non-null     float64
 6   Repetition Rate     36 non-null     float64
 7   Projector           36 non-null     float64
 8   Smart Class         36 non-null     float64
 9   Digital Library     36 non-null     float64
 10  Library             36 non-null     float64
 11  Playground          36 non-null     float64
 12  Kitchen Garden      36 non-null     float64
 13  Girls' Toilet       36 non-null     float64
 14  Boys' Toilet        36 non-null     float64
 15  Electricity         36 non-null     float64
 16  Computers           36 non-null     float64
 17  Internet            36 non-null     float64
 18  Drinking Water      36 non-null     float64
 19  Hand wash           36 non-null     float64
 20  Ramp                36 non-null     float64
 21  Ramp and Handrails  36 non-null     float64
 22  CWSN Toilet         36 non-null     float64
dtypes: float64(20), int64(2), object(1)
memory usage: 6.6+ KB
```

PRELIMINARY ANALYSIS ON TABLEAU

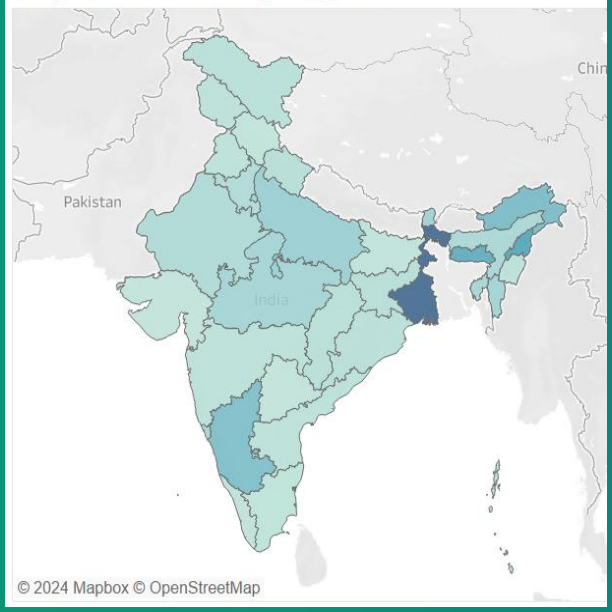




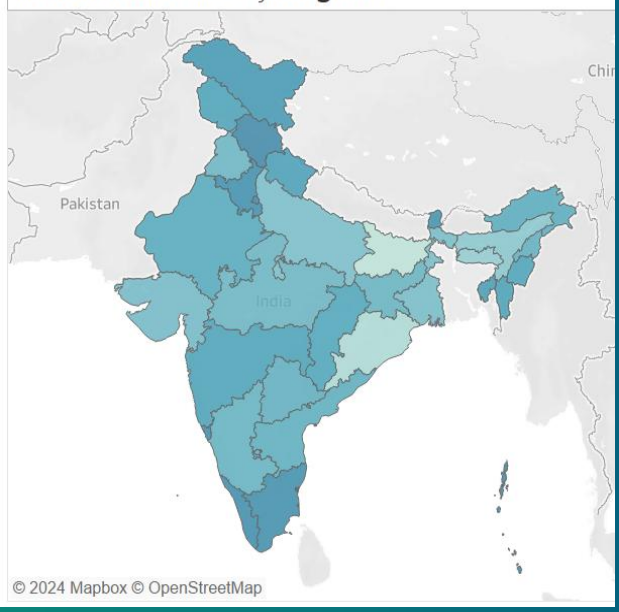




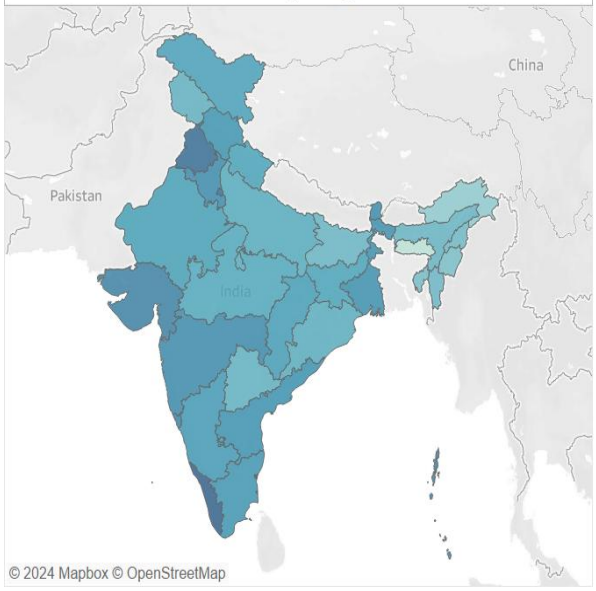
Repetition Rate by Region



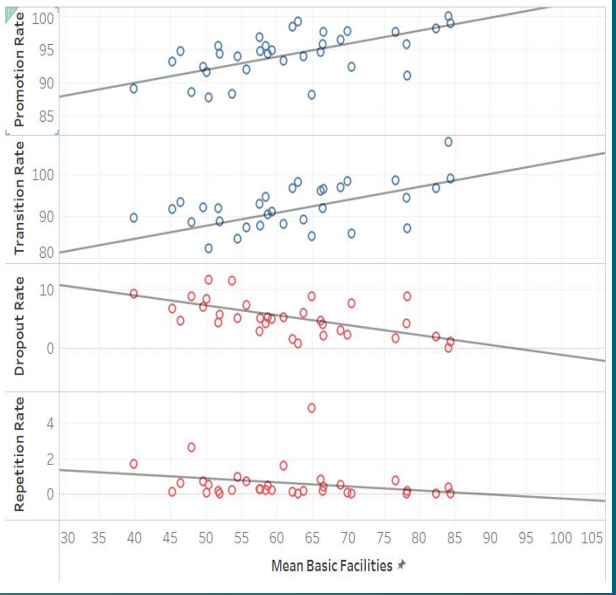
Transition Rate by Region



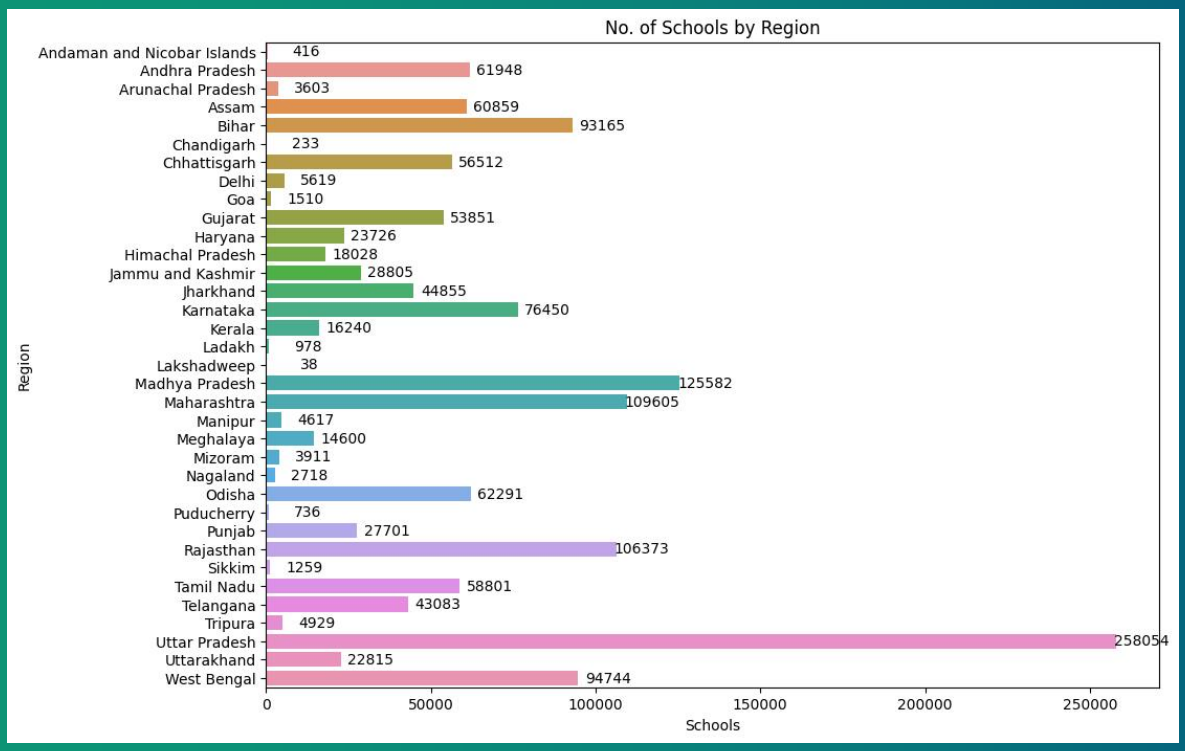
Mean Basic Facilities by Region

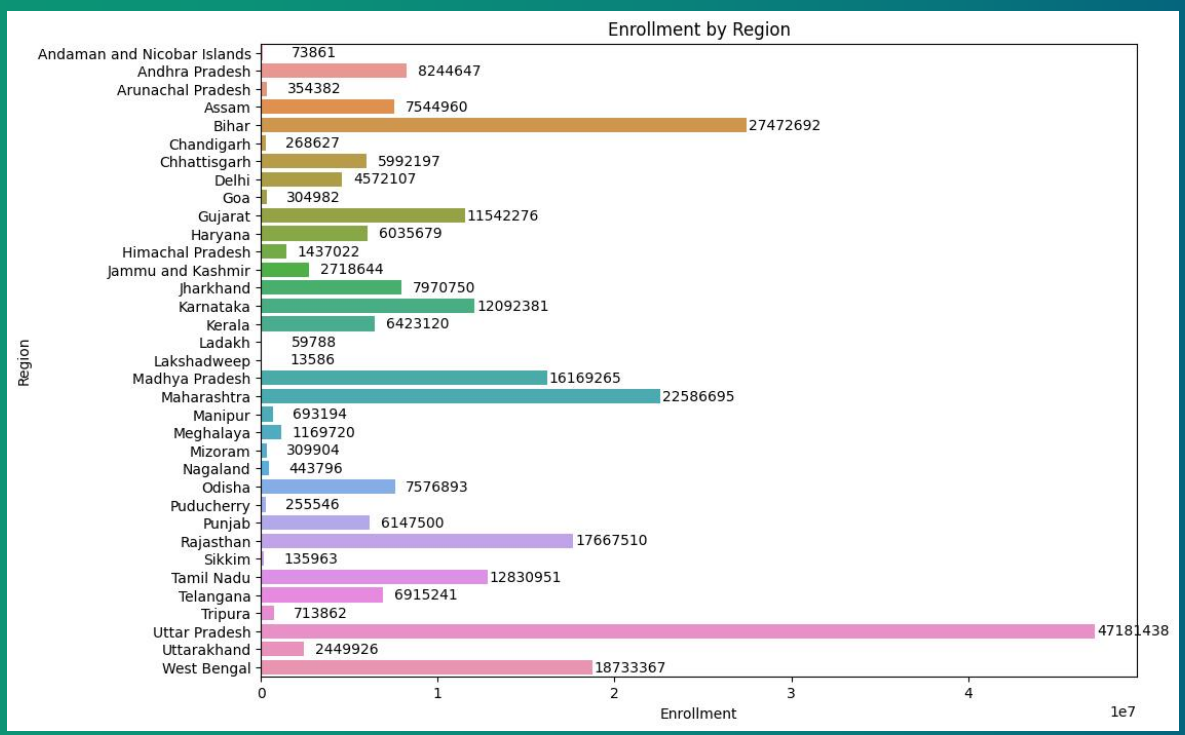


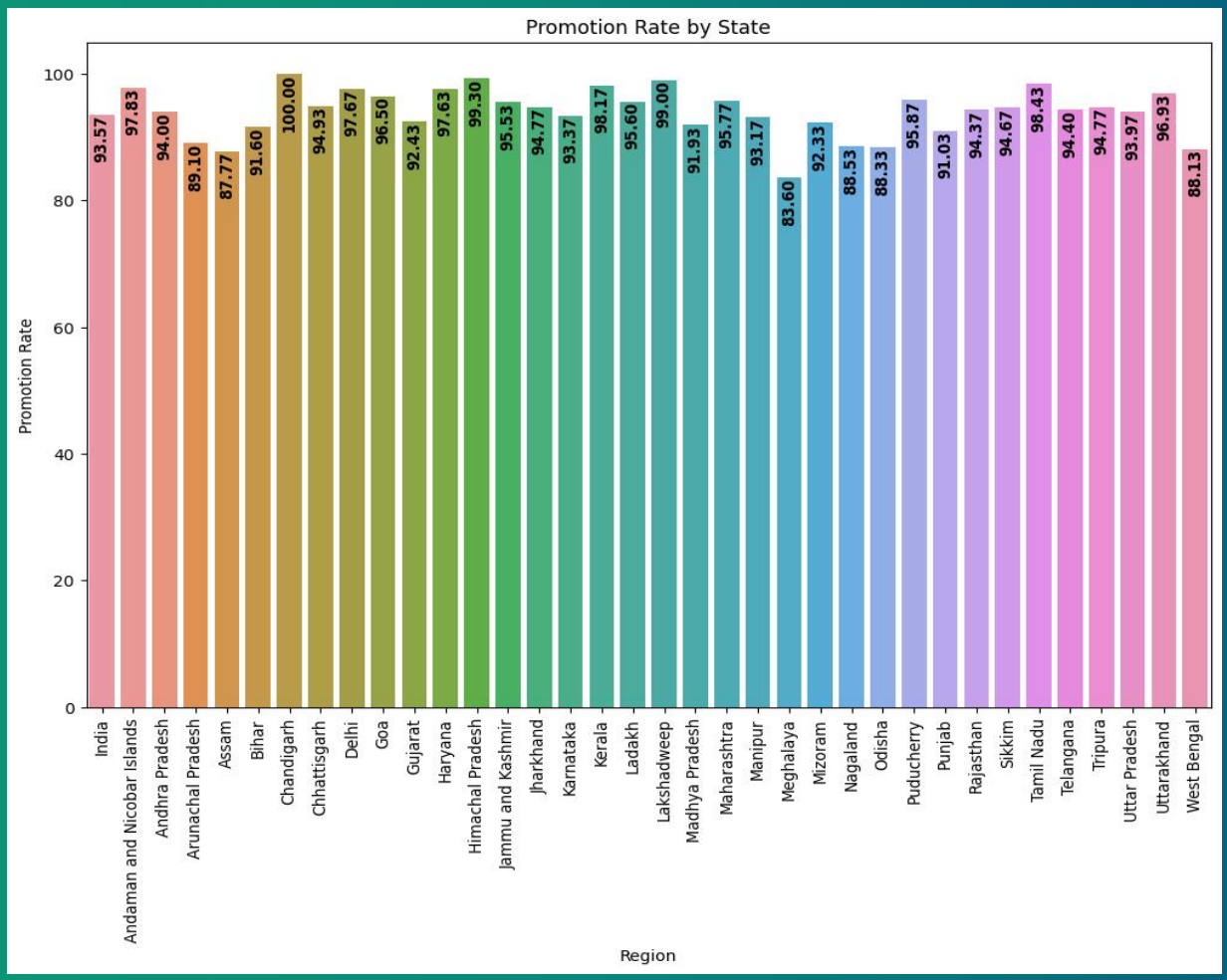
Performance Metrics VS Basic Facilities

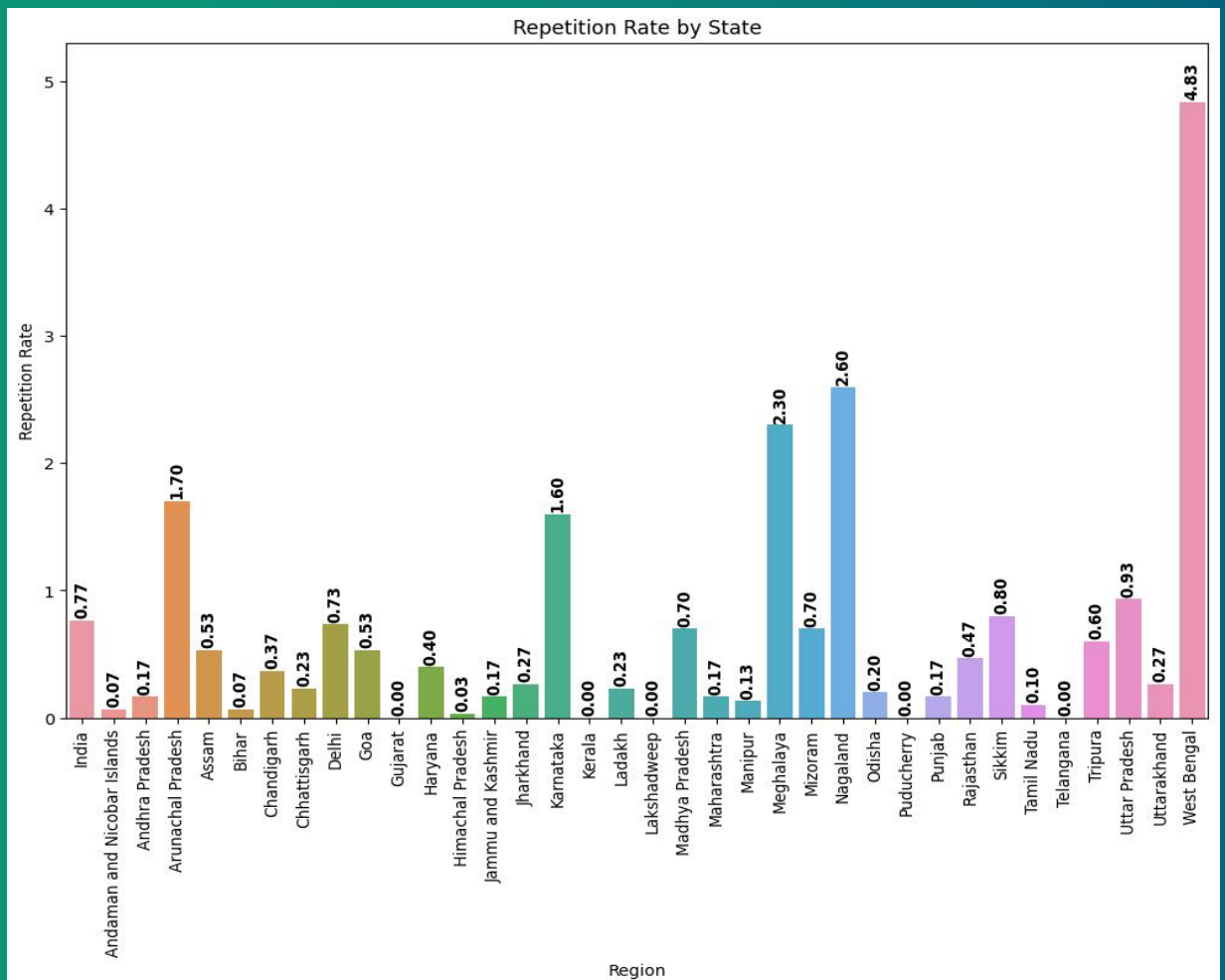


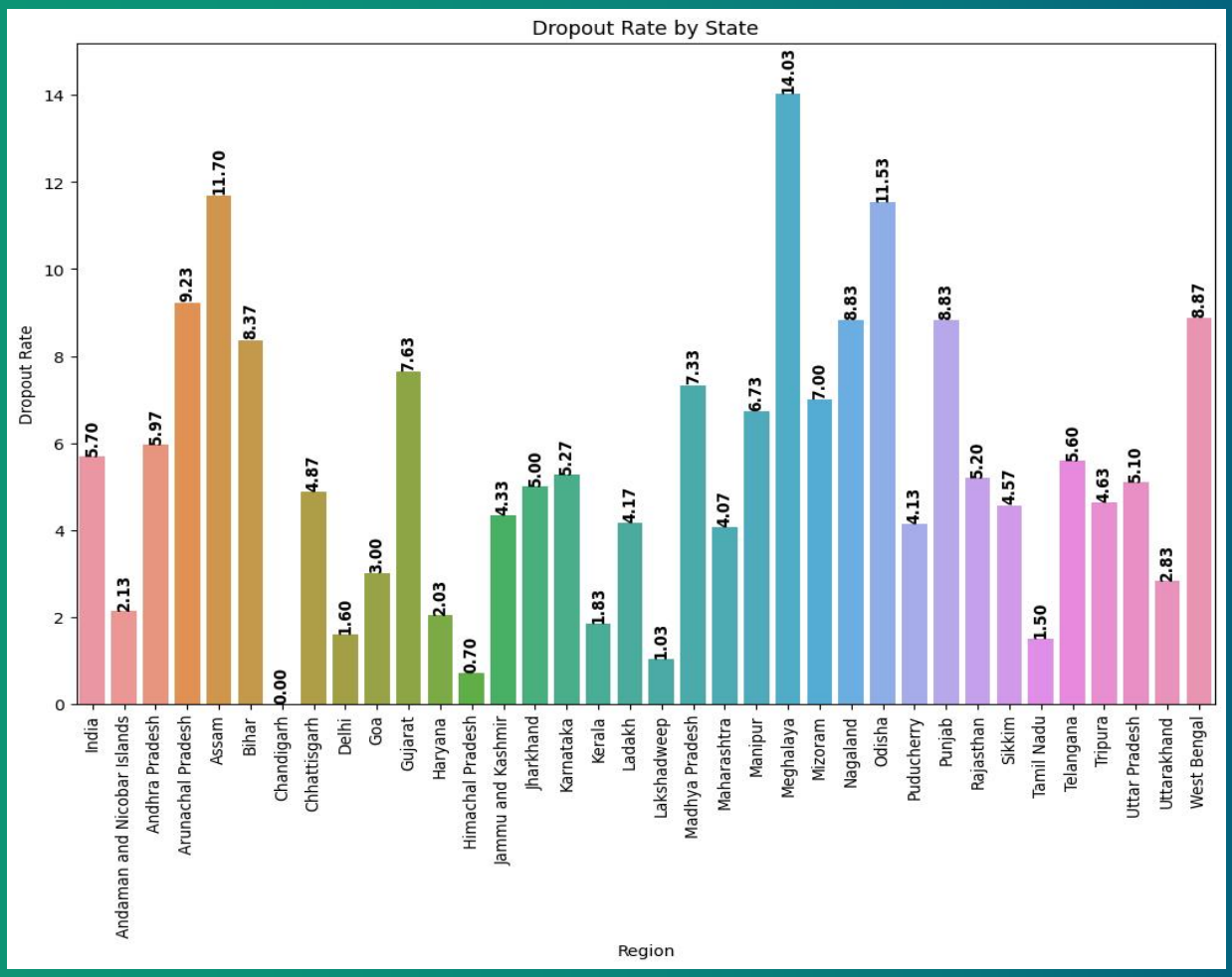
ANALYSIS USING PYTHON

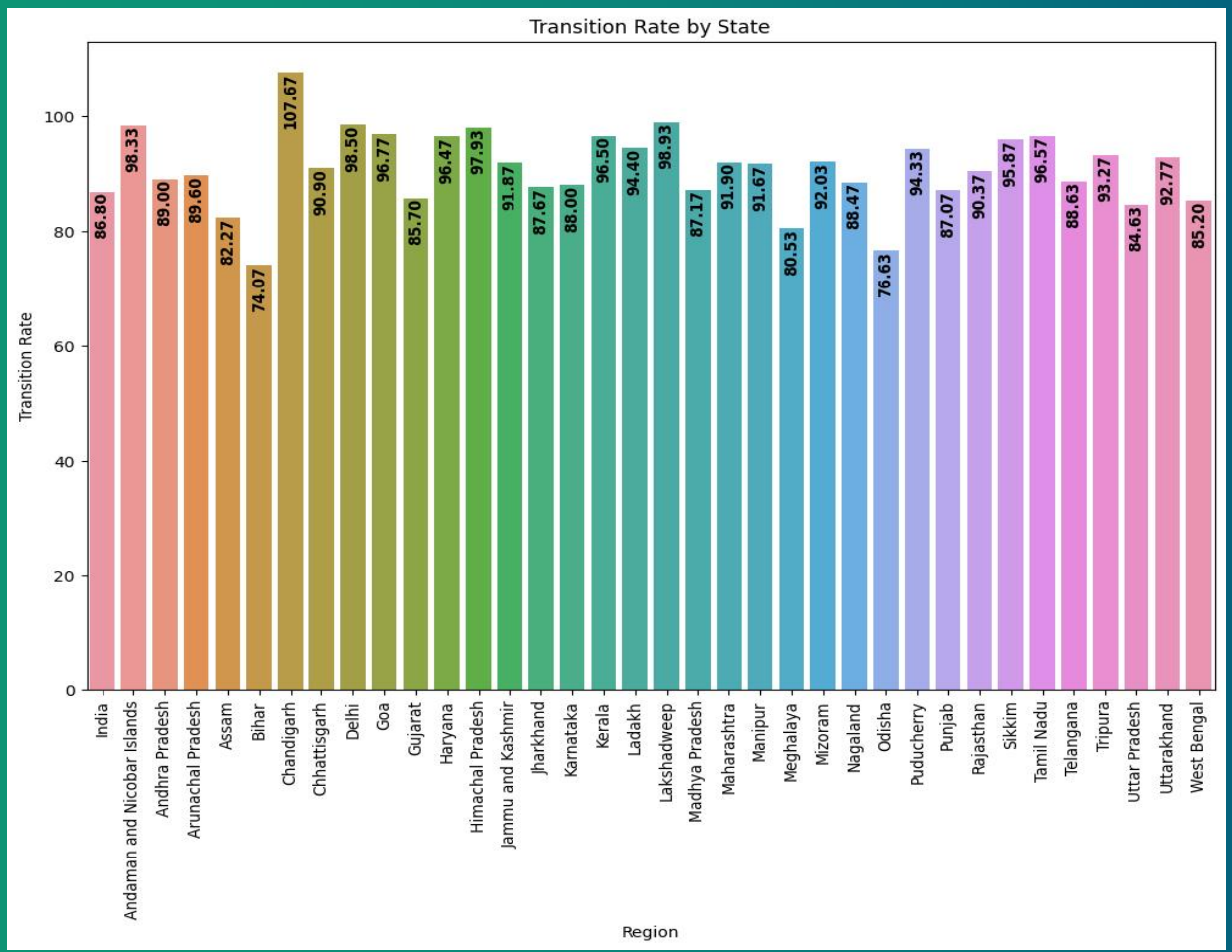




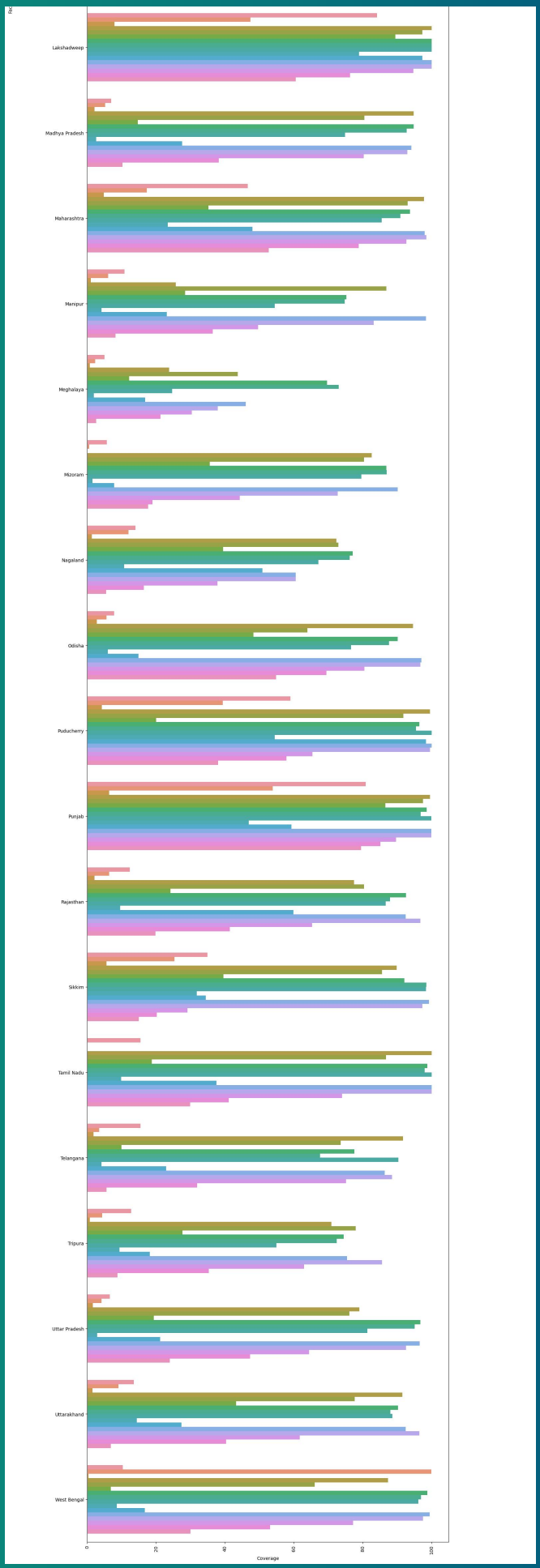

Promotion Rate by State
Promotion Rate
Region
India 93.57
Andaman and Nicobar Islands 97.83
Andhra Pradesh 94.00
Arunachal Pradesh 89.10
Assam 87.77
Bihar 91.60
Chandigarh 100.00
Chhattisgarh 94.93
Delhi 97.67
Goa 96.50
Gujarat 92.43
Haryana 97.63
Himachal Pradesh 99.30
Jammu and Kashmir 95.53
Jharkhand 94.77
Karnataka 93.37
Kerala 98.17
Ladakh 95.60
Lakshadweep 99.00
Madhya Pradesh 91.93
Maharashtra 95.77
Manipur 93.17
Meghalaya 83.60
Mizoram 92.33
Nagaland 88.53
Odisha 88.33
Puducherry 95.87
Punjab 91.03
Rajasthan 94.37
Sikkim 94.67
Tamil Nadu 98.43
Telangana 94.40
Tripura 94.77
Uttar Pradesh 93.97
Uttarakhand 96.93
West Bengal 88.13

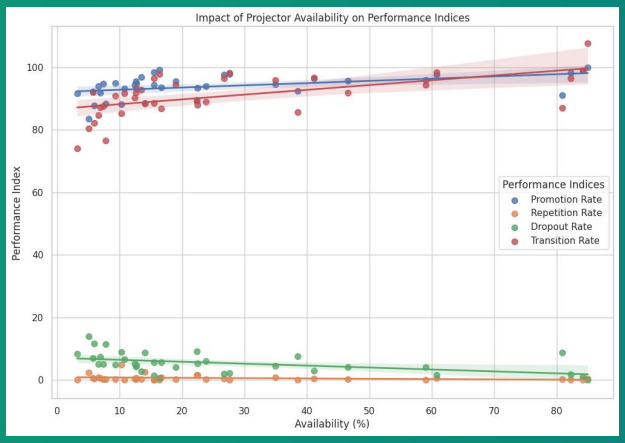

Repetition Rate by State
Repetition Rate
Region
India 0.77
Andaman and Nicobar Islands 0.07
Andhra Pradesh 0.17
Arunachal Pradesh 1.70
Assam 0.53
Bihar 0.07
Chandigarh 0.37
Chhattisgarh 0.23
Delhi 0.73
Goa 0.53
Gujarat 0.00
Haryana 0.40
Himachal Pradesh 0.03
Jammu and Kashmir 0.17
Jharkhand 0.27
Karnataka 1.60
Kerala 0.00
Ladakh 0.23
Lakshadweep 0.00
Madhya Pradesh 0.70
Maharashtra 0.17
Manipur 0.13
Meghalaya 2.30
Mizoram 0.70
Nagaland 2.60
Odisha 0.20
Puducherry 0.00
Punjab 0.17
Rajasthan 0.47
Sikkim 0.80
Tamil Nadu 0.10
Telangana 0.00
Tripura 0.60
Uttar Pradesh 0.93
Uttarakhand 0.27
West Bengal 4.83

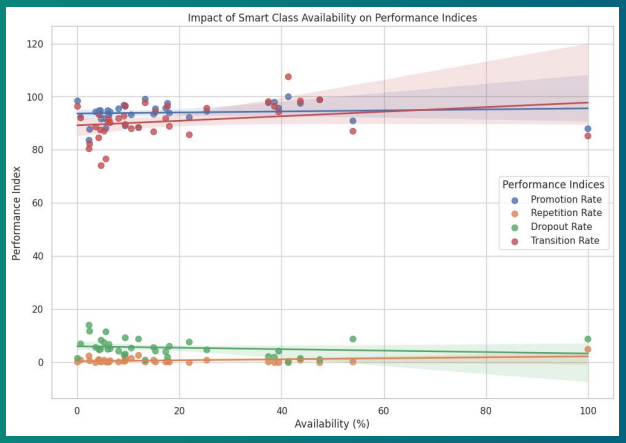

Dropout Rate by State
Dropout Rate
Region
India 5.70
Andaman and Nicobar Islands 2.13
Andhra Pradesh 5.97
Arunachal Pradesh 9.23
Assam 11.70
Bihar 8.37
Chandigarh 0.00
Chhattisgarh 4.87
Delhi 1.60
Goa 3.00
Gujarat 7.63
Haryana 2.03
Himachal Pradesh 0.70
Jammu and Kashmir 4.33
Jharkhand 5.00
Karnataka 5.27
Kerala 1.83
Ladakh 4.17
Lakshadweep 1.03
Madhya Pradesh 7.33
Maharashtra 4.07
Manipur 6.73
Meghalaya 14.03
Mizoram 7.00
Nagaland 8.83
Odisha 11.53
Puducherry 4.13
Punjab 8.83
Rajasthan 5.20
Sikkim 4.57
Tamil Nadu 1.50
Telangana 5.60
Tripura 4.63
Uttar Pradesh 5.10
Uttarakhand 2.83
West Bengal 8.87

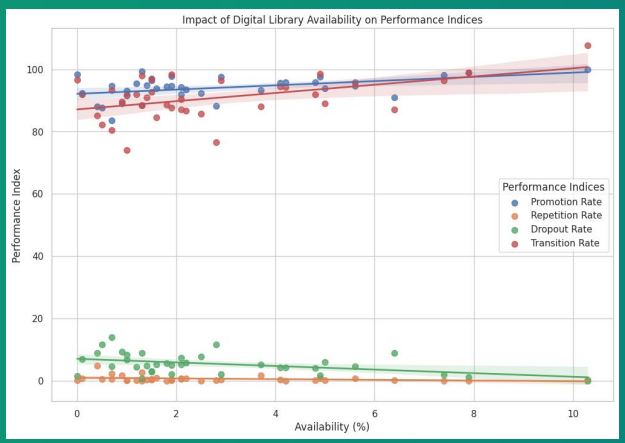

Transition Rate by State
Transition Rate
Region
India 86.80
Andaman and Nicobar Islands 98.33
Andhra Pradesh 89.00
Arunachal Pradesh 89.60
Assam 82.27
Bihar 74.07
Chandigarh 107.67
Chhattisgarh 90.90
Delhi 98.50
Goa 96.77
Gujarat 85.70
Haryana 96.47
Himachal Pradesh 97.93
Jammu and Kashmir 91.87
Jharkhand 87.67
Karnataka 88.00
Kerala 96.50
Ladakh 94.40
Lakshadweep 98.93
Madhya Pradesh 87.17
Maharashtra 91.90
Manipur 91.67
Meghalaya 80.53
Mizoram 92.03
Nagaland 88.47
Odisha 76.63
Puducherry 94.33
Punjab 87.07
Rajasthan 90.37
Sikkim 95.87
Tamil Nadu 96.57
Telangana 88.63
Tripura 93.27
Uttar Pradesh 84.63
Uttarakhand 92.77
West Bengal 85.20

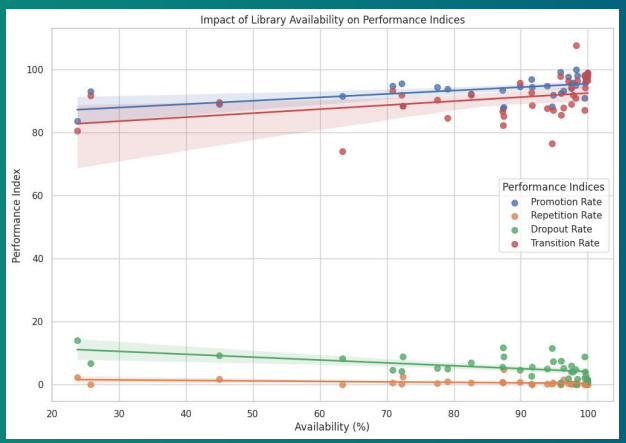

Facilities by State
India
Andaman and Nicobar Islands
Andhra Pradesh
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Projector
Smart Class
Digital Library
Library
Playground
Kitchen Garden
Girls' Toilet
Boys' Toilet
Electricity
Computers
Internet
Drinking Water
Hand wash
Ramp
Ramp and Handrails
CWSN Toilet

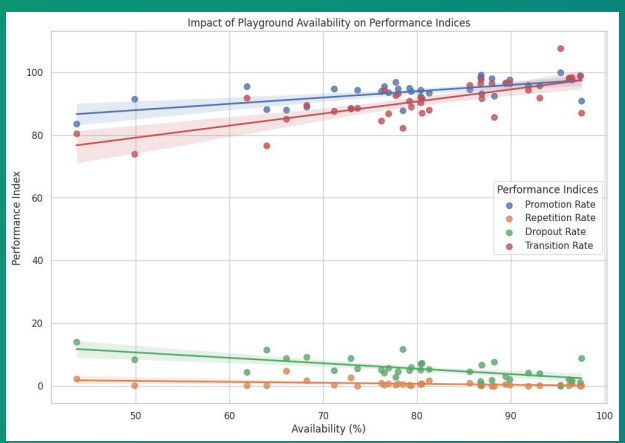

Lakshadweep
Madhya Pradesh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Telangana
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0
20
40
60
80
100
Coverage

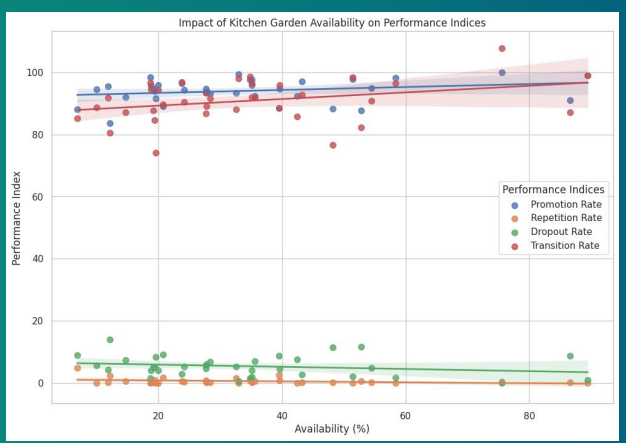

Impact of Projector Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

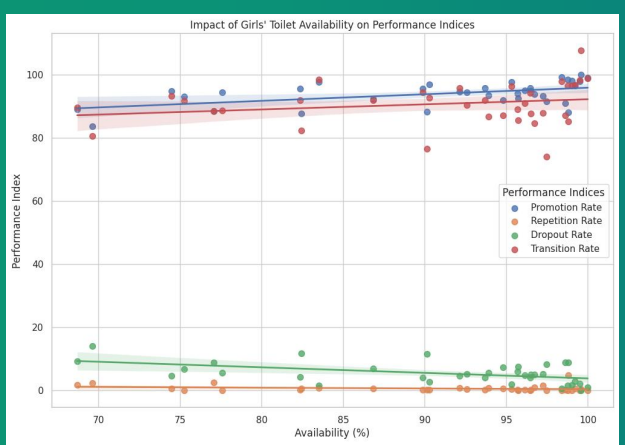

Impact of Smart Class Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

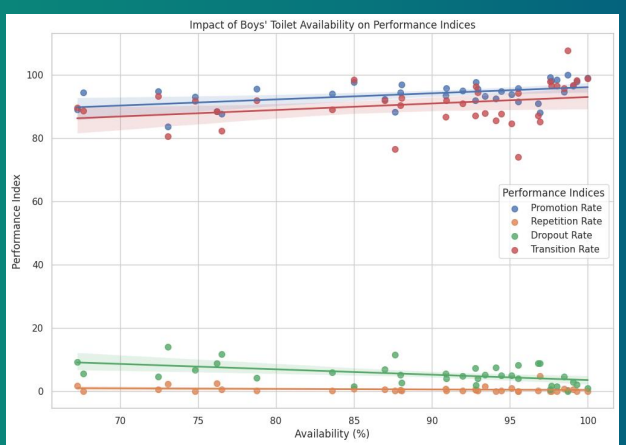

Impact of Digital Library Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

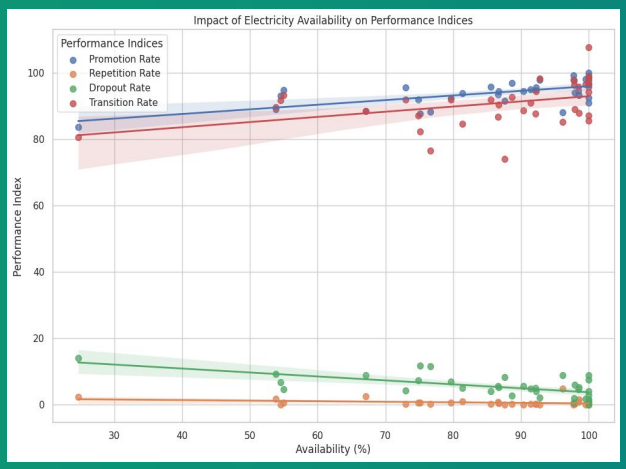

Impact of Library Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

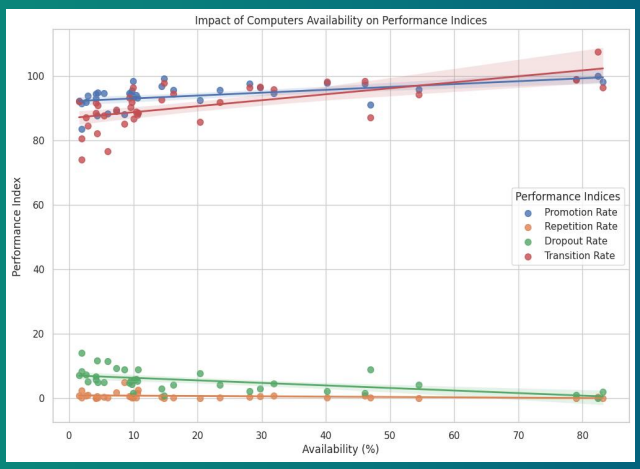

Impact of Playground Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

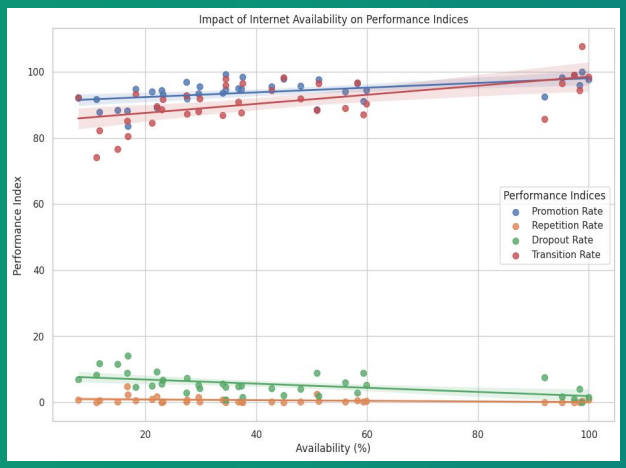

Impact of Kitchen Garden Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

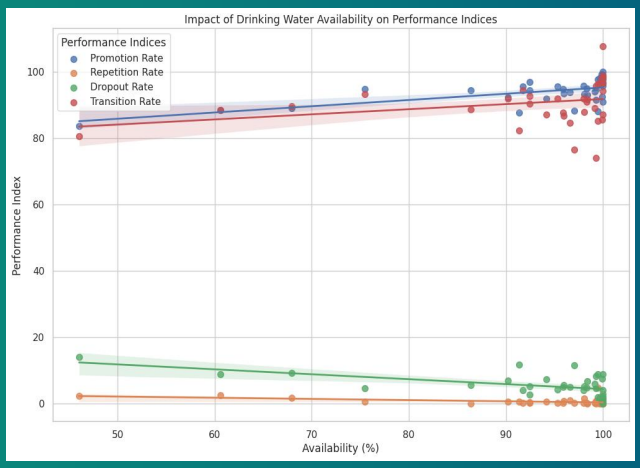

Impact of Girls' Toilet Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

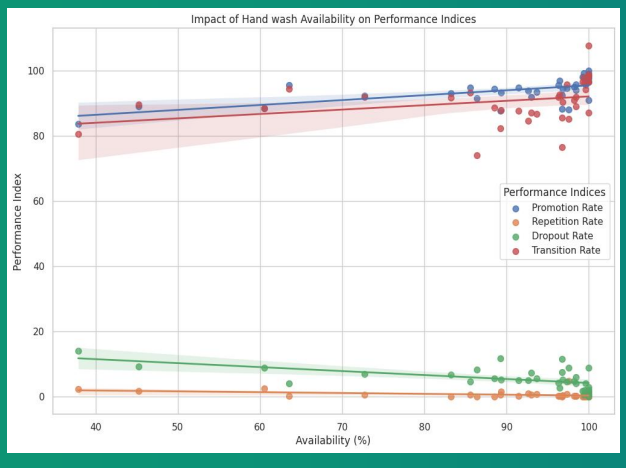

Impact of Boys' Toilet Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate

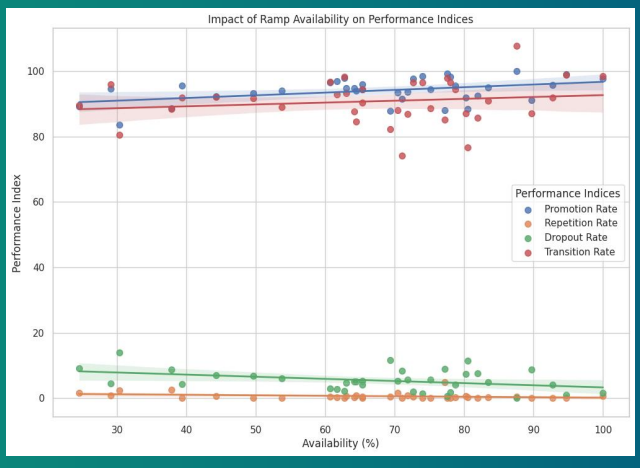

Impact of Electricity Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

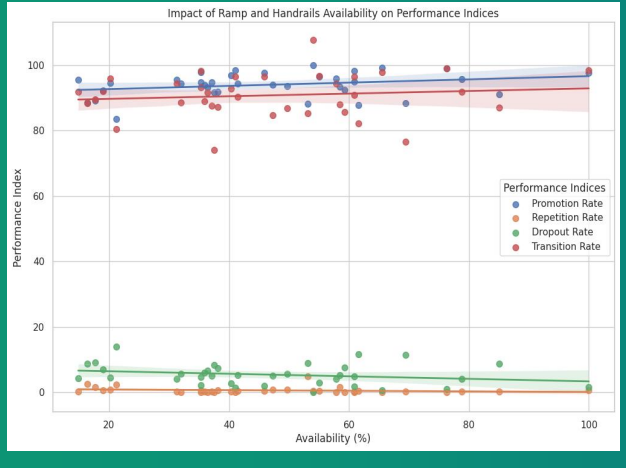

Impact of Computers Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

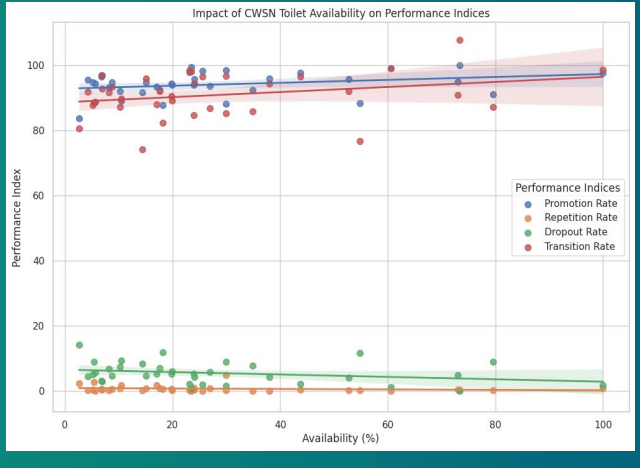

Impact of Internet Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

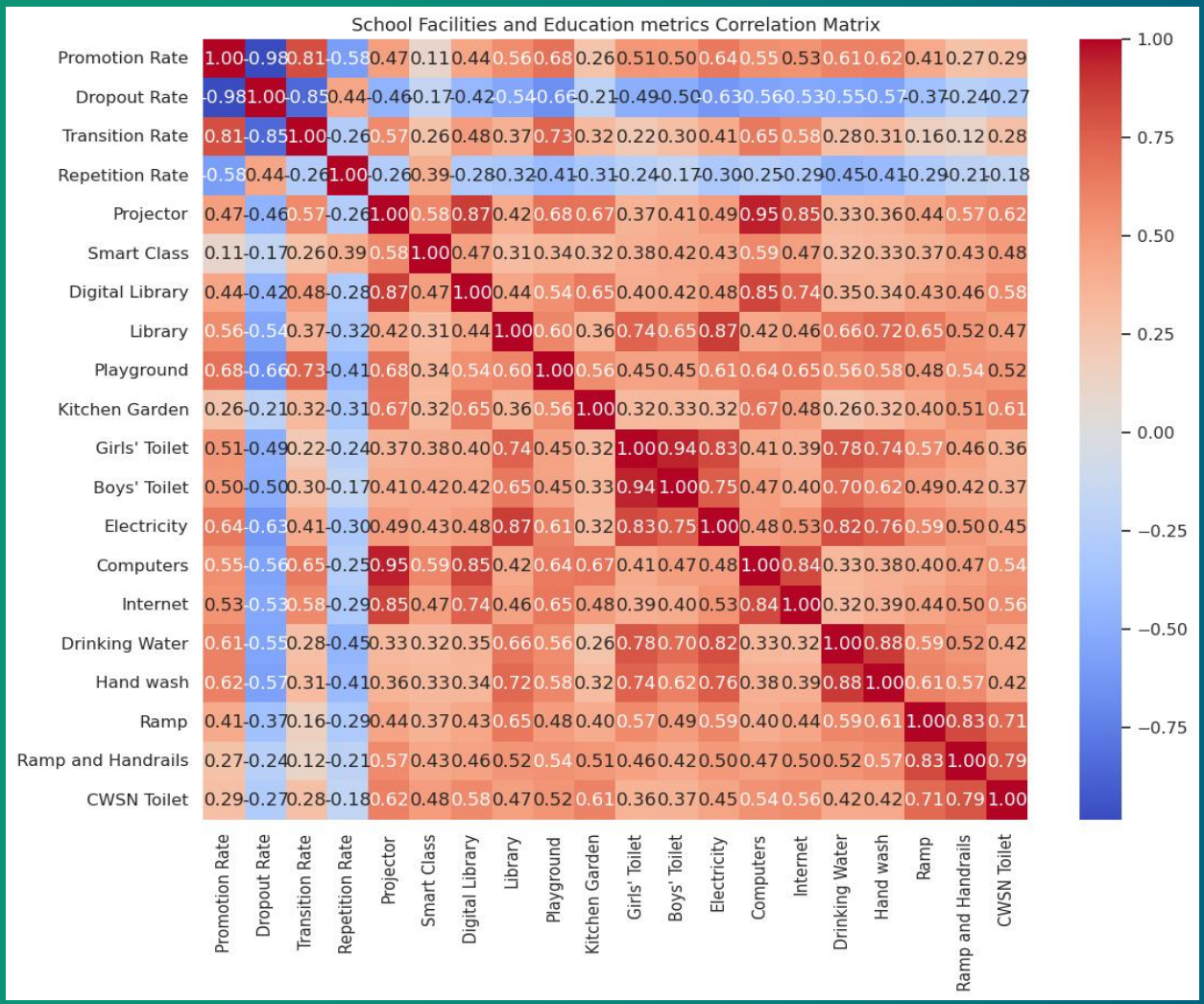

Impact of Drinking Water Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

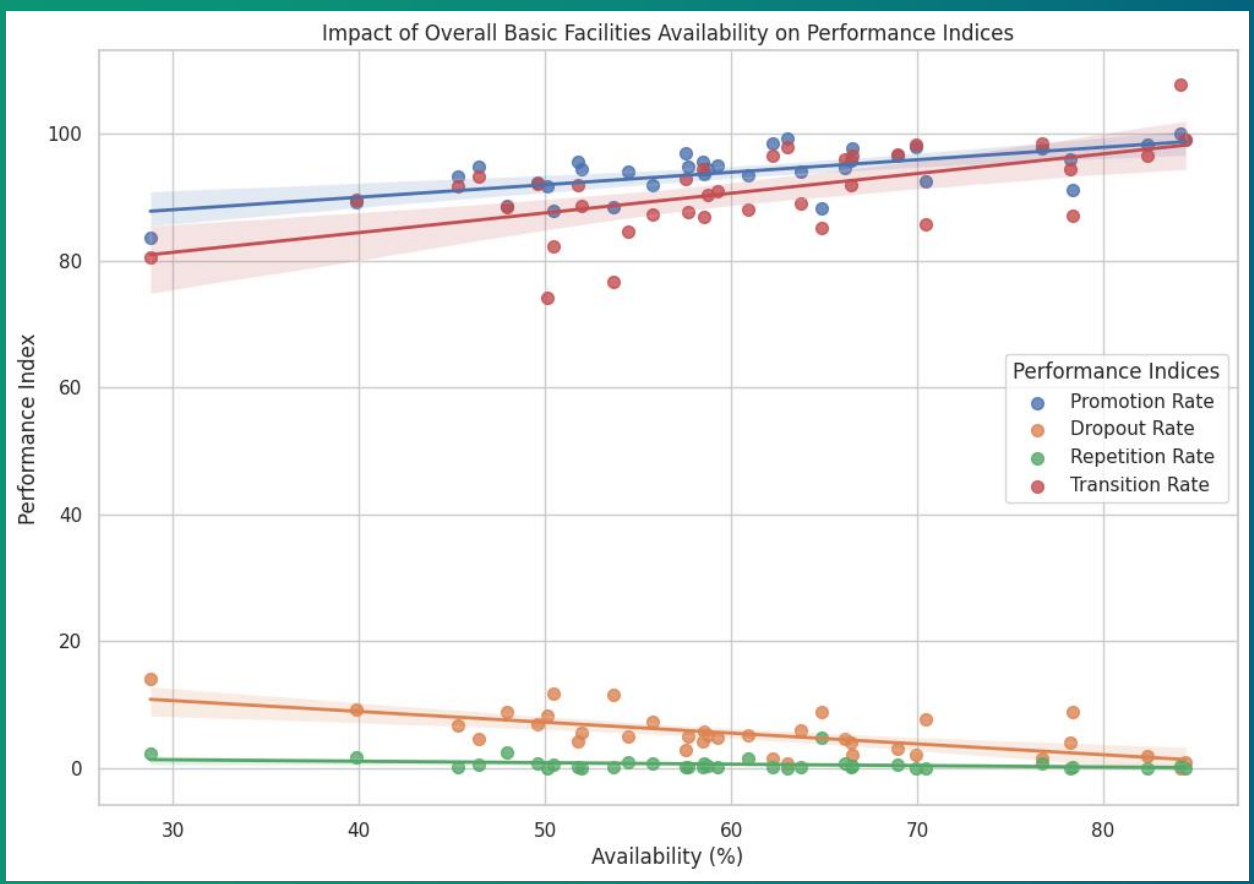

Impact of Hand wash Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)


Impact of Ramp Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

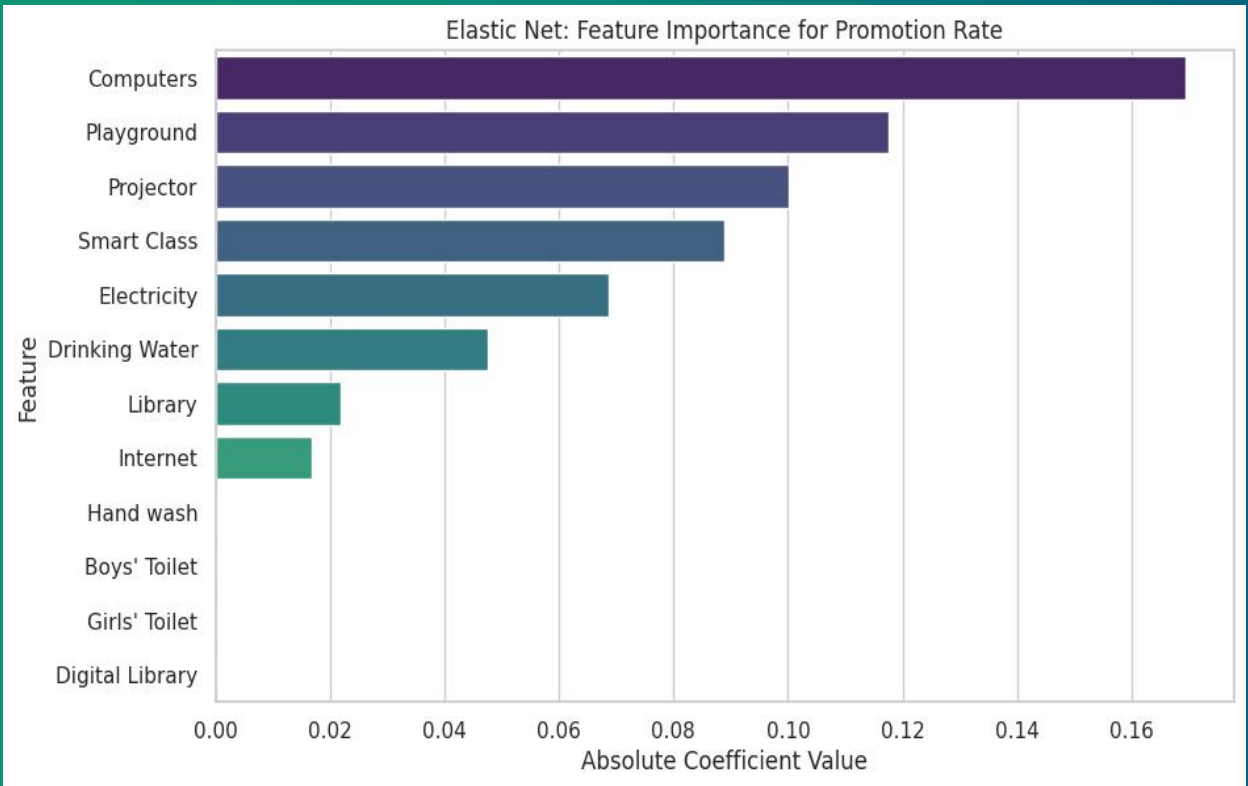

Impact of Ramp and Handrails Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)

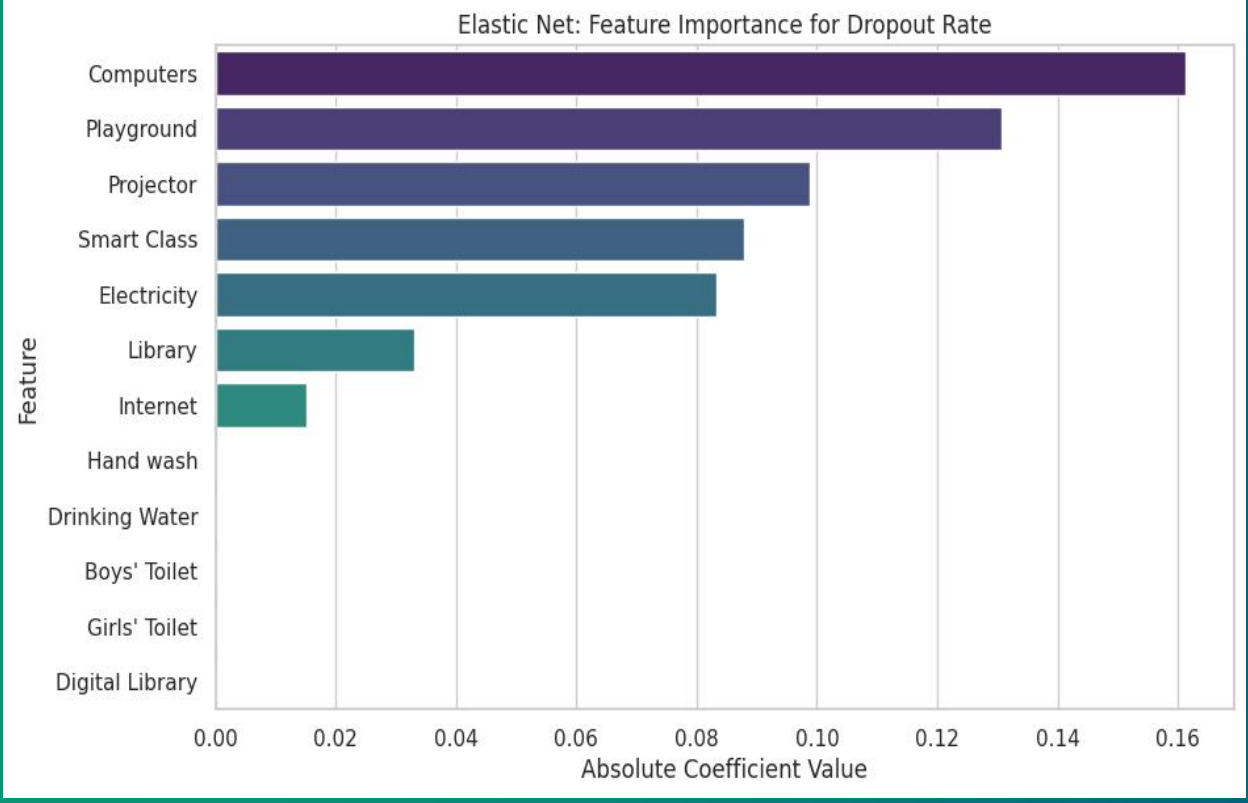

Impact of CWSN Toilet Availability on Performance Indices
Performance Indices
Promotion Rate
Repetition Rate
Dropout Rate
Transition Rate
Performance Index
Availability (%)


School Facilities and Education metrics Correlation Matrix
Promotion Rate
Dropout Rate
Transition Rate
Repetition Rate
Projector
Smart Class
Digital Library
Library
Playground
Kitchen Garden
Girls' Toilet
Boys' Toilet
Electricity
Computers
Internet
Drinking Water
Hand wash
Ramp
Ramp and Handrails
CWSN Toilet
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

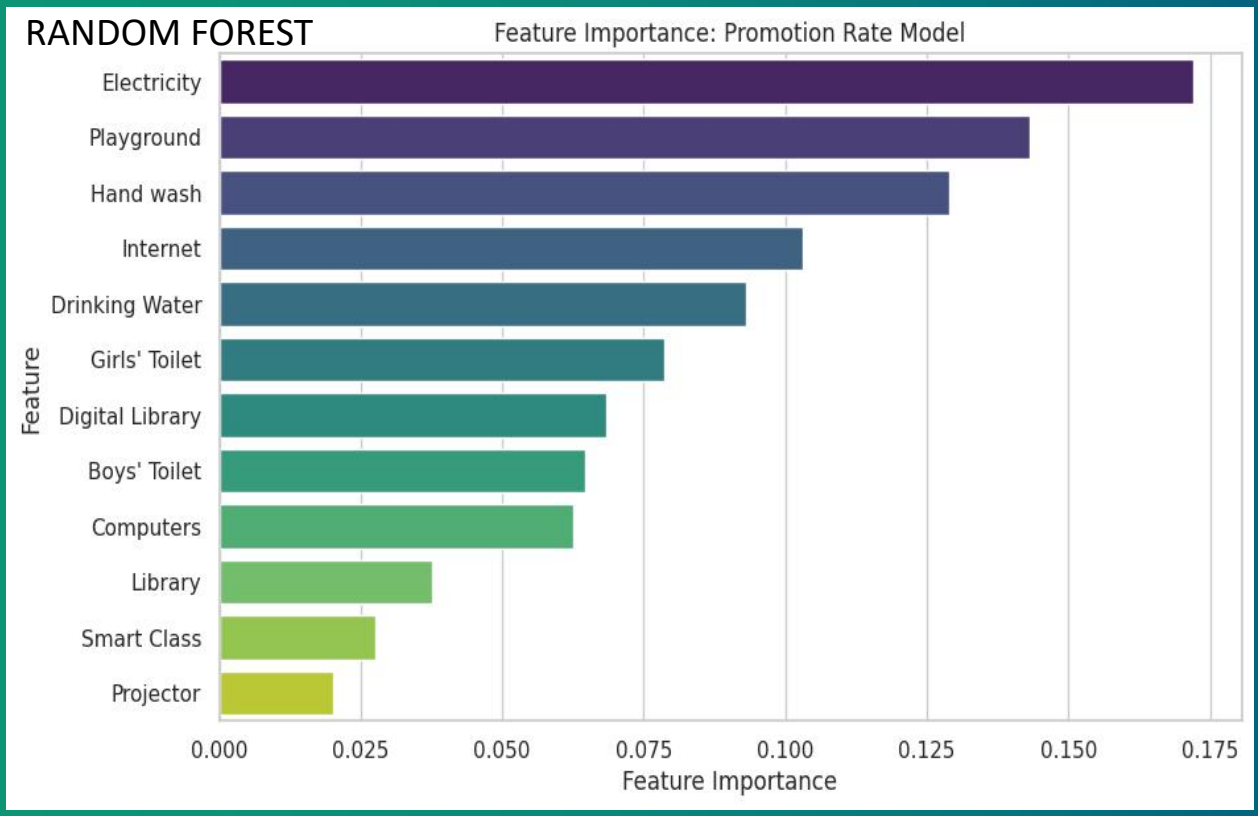

Impact of Overall Basic Facilities Availability on Performance Indices
Performance Index
Availability (%)
Performance Indices
Promotion Rate
Dropout Rate
Repetition Rate
Transition Rate
0
20
40
60
80
100
30
40
50
60
70
80

MODEL EVALUATION

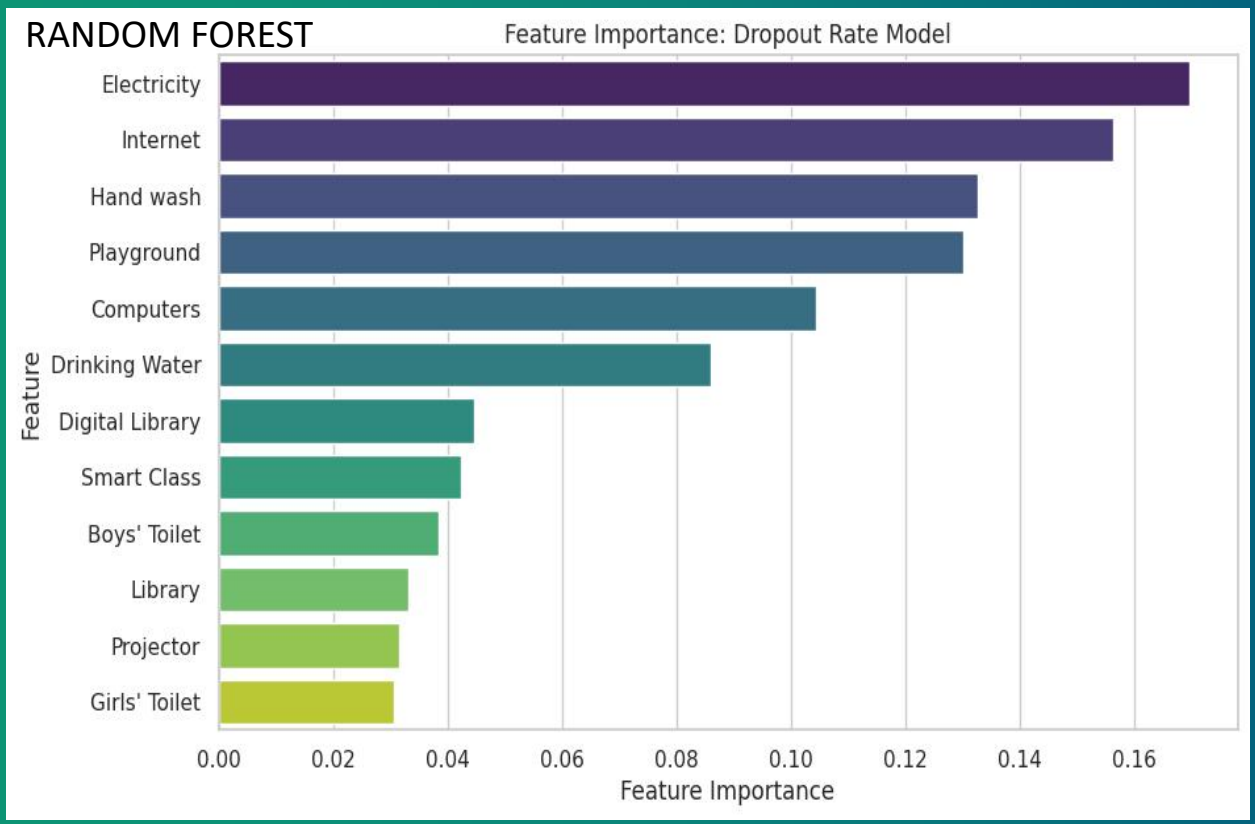


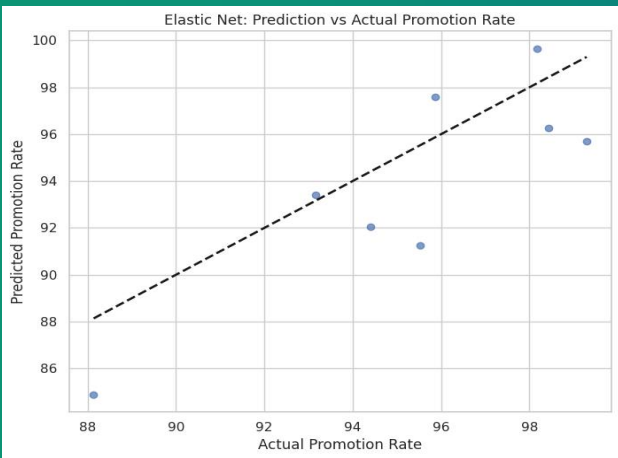


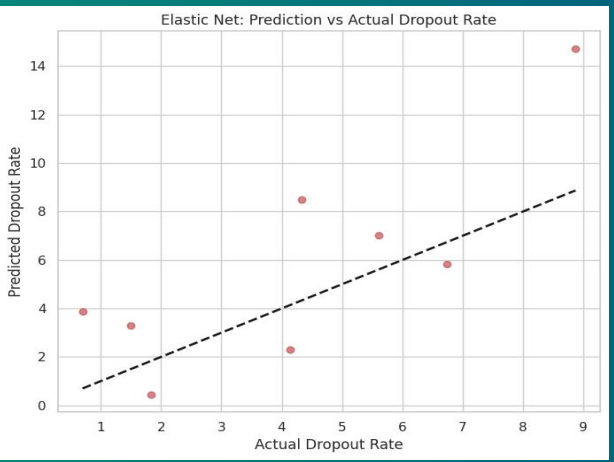
RANDOM FOREST

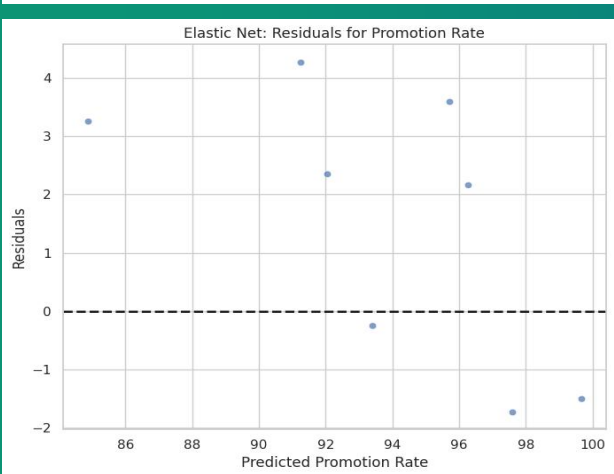


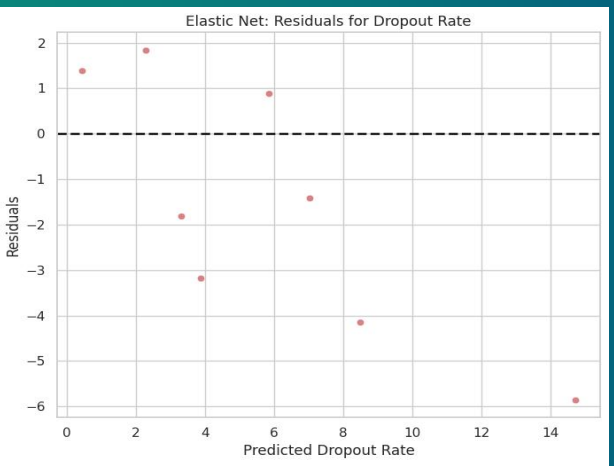
RANDOM FOREST



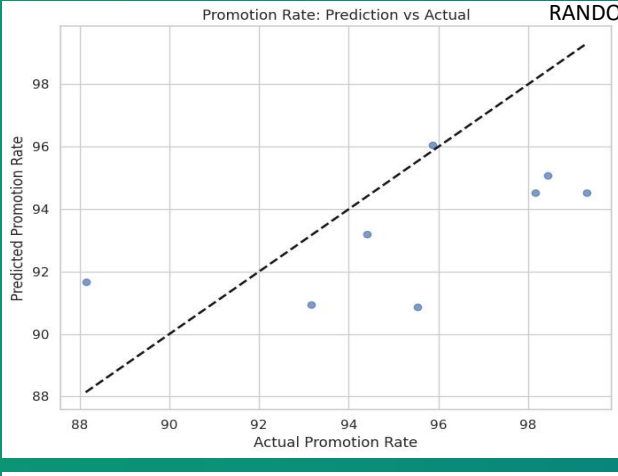


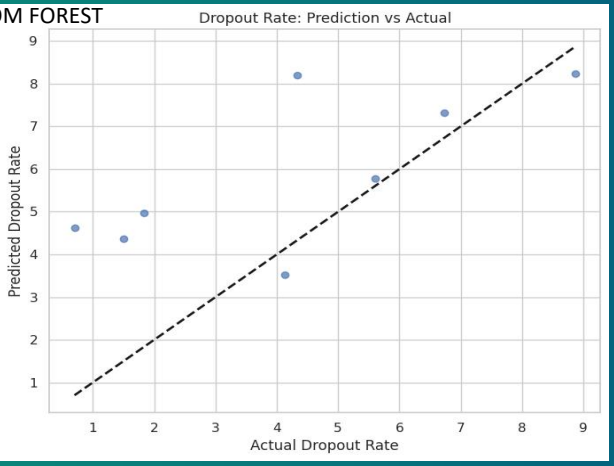




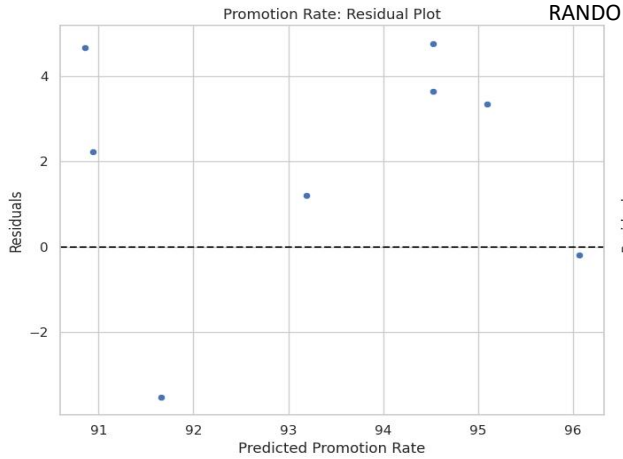


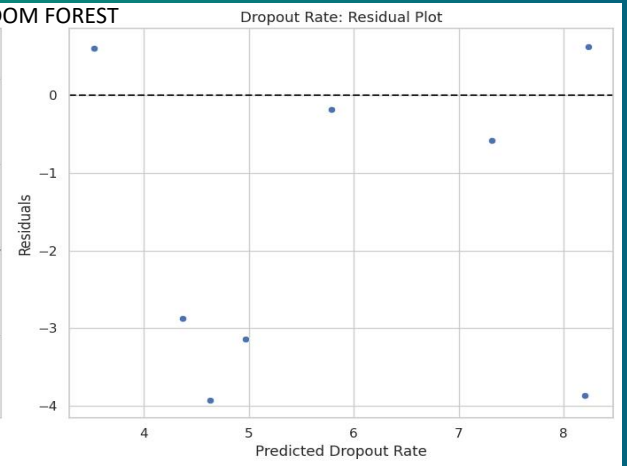
RANDOM FOREST





RANDOM FOREST





WORKING PREDICTION PAGE

LINK TO GITHUB REPO : CLICK HERE