



NAME: MUDASIR NAEEM

POSITION: TECH INTERN (AI/DATASCIENCE)

EMAIL: mudasirnaeem000@gmail.com

LINKEDIN: www.linkedin.com/in/mudasir-naeem-698679303

GITHUB:

<https://github.com/MudasirNaeem1/Kohaq-Tech-Internship>

**DATASET NAMED "CONSOLIDATED
(EDUCATIONAL DATASET).CSV" STATISTICS &
VISUALIZATIONS USING MATPLOTLIB AND
SEABORN**

Column Names:

=====

1. % Boys Enrolled
2. % Complete Primary Schools
3. % Girls Enrolled
4. % Primary Schools with single classroom
5. % Primary Schools with single teacher
6. All Four Facilities
7. Any One Facility
8. Any Three Facilities
9. Any Two Facilities
10. Area (km²)
11. Bomb Blasts Occurred
12. Boundary wall, Building condition satisfactory, Drinking water and 2 more (clusters)
13. Boundary wall
14. Building condition satisfactory
15. MeasureGroup 1 Measures
16. Color By Measure Value
17. Show Sheet
18. Table of Contents
19. Other Factors Measure Value
20. Analysis Level Selector
21. Color By Measure Name
22. Complete Primary Schools
23. Country
24. City
25. Drinking water
26. Drone attacks in Pakistan
27. Education score
28. Educational Budget Spend of GDP
29. Electricity
30. Enrolment score
31. Gender parity score
32. Global Terrorism Index - Pakistan
33. Learning score
34. MeasureGroup 2 Measures
35. No Facility
36. Number of Records
37. Number of primary schools as % of total schools
38. Number of primary schools
39. Number of secondary schools as % of total schools
40. Number of secondary schools
41. Pakistan Economic Growth
42. Population
43. Primary Schools with single classroom
44. Primary Schools with single teacher
45. Province
46. Retention score

47. School infrastructure score
 48. Terrorist Attacks Affectees
 49. Toilet
 50. Total number of schools
 51. Year

First 5 rows:

=====

	% Boys Enrolled	% Complete Primary Schools	% Girls Enrolled \
0	54.77%	0.936599	45.23%
1	62.50%	0.594203	37.50%
2	86.63%	0.939068	13.37%
3	60.77%	0.827225	39.23%
4	60.75%	0.592348	39.25%

	% Primary Schools with single classroom \
0	0.034582
1	0.359903
2	0.021505
3	0.041885
4	0.346966

	% Primary Schools with single teacher	All Four Facilities	Any One Facility \
0	0.028818	17.06%	
1	0.045894	15.25%	16.20%
2	0.039427	5.35%	20.21%
3	0.130890	1.57%	4.55%
4	0.060686	7.12%	22.44%
			23.06%

	Any Three Facilities	Any Two Facilities	Area (km ²)	...	Population
0	24.73%	24.95%	768.0	...	351415.0
1	17.38%	15.43%	1516.0	...	301633.0
2	16.31%	20.05%	NaN	...	NaN
3	7.09%	12.20%	600.0	...	150000.0
4	9.78%	13.60%	2162.0	...	834094.0

	Primary Schools with single classroom	Primary Schools with single teacher \
--	---------------------------------------	---------------------------------------

```

0          12.0
10.0
1          149.0
19.0
2           6.0
11.0
3           8.0
25.0
4          263.0
46.0

```

```

Province  Retention score  School infrastructure score \
0      AJK          14.8          24.929178
1      AJK          75.2          28.883827
2      AJK           0.0          22.463768
3      AJK           0.0          13.048128
4      AJK          73.2          20.184453

```

```

Terrorist Attacks Affectees  Toilet Total number of schools  Year
0          5379  37.393768          601  2013
1          5379  20.728929          595  2013
2          5379  43.840580          379  2013
3          5379  26.203209          262  2013
4          5379  14.888011         1036  2013

```

```
[5 rows x 51 columns]
```

Dataset Info:

```

=====
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 51 columns):
#   Column
Non-Null Count  Dtype
---  -
0   % Boys Enrolled
579 non-null    object
1   % Complete Primary Schools
576 non-null    float64
2   % Girls Enrolled
579 non-null    object
3   % Primary Schools with single classroom
576 non-null    float64

```

```
4 % Primary Schools with single teacher
576 non-null float64
5 All Four Facilities
576 non-null object
6 Any One Facility
576 non-null object
7 Any Three Facilities
576 non-null object
8 Any Two Facilities
576 non-null object
9 Area (km²)
512 non-null float64
10 Bomb Blasts Occurred
580 non-null int64
11 Boundary wall, Building condition satisfactory, Drinking water
and 2 more (clusters) 580 non-null object
12 Boundary wall
580 non-null float64
13 Building condition satisfactory
580 non-null float64
14 MeasureGroup 1 Measures
580 non-null float64
15 Color By Measure Value
580 non-null float64
16 Show Sheet
580 non-null object
17 Table of Contents
580 non-null object
18 Other Factors Measure Value
580 non-null object
19 Analysis Level Selector
580 non-null object
20 Color By Measure Name
580 non-null object
21 Complete Primary Schools
576 non-null float64
22 Country
580 non-null object
23 City
580 non-null object
24 Drinking water
580 non-null float64
25 Drone attacks in Pakistan
580 non-null int64
26 Education score
580 non-null float64
27 Educational Budget Spend of GDP
580 non-null object
28 Electricity
```



```
580 non-null    float64
   29 Enrolment score
580 non-null    float64
   30 Gender parity score
580 non-null    float64
   31 Global Terrorism Index - Pakistan
580 non-null    float64
   32 Learning score
580 non-null    float64
   33 MeasureGroup 2 Measures
580 non-null    float64
   34 No Facility
576 non-null    object
   35 Number of Records
580 non-null    int64
   36 Number of primary schools as % of total schools
580 non-null    float64
   37 Number of primary schools
580 non-null    int64
   38 Number of secondary schools as % of total schools
580 non-null    float64
   39 Number of secondary schools
580 non-null    int64
   40 Pakistan Economic Growth
580 non-null    object
   41 Population
512 non-null    float64
   42 Primary Schools with single classroom
576 non-null    float64
   43 Primary Schools with single teacher
576 non-null    float64
   44 Province
580 non-null    object
   45 Retention score
580 non-null    float64
   46 School infrastructure score
580 non-null    float64
   47 Terrorist Attacks Affectees
580 non-null    int64
   48 Toilet
580 non-null    float64
   49 Total number of schools
580 non-null    int64
   50 Year
580 non-null    int64
dtypes: float64(25), int64(8), object(18)
memory usage: 231.2+ KB
None
```

Basic Statistics:

=====

	% Complete Primary Schools	% Primary Schools with single classroom \
count	576.000000	
576.000000		
mean	0.592544	
0.135134		
std	0.302878	
0.121400		
min	0.000000	
0.000000		
25%	0.340517	
0.029740		
50%	0.691613	
0.096157		
75%	0.846848	
0.211669		
max	0.994764	
0.514768		

	% Primary Schools with single teacher	Area (km ²) \
count	576.000000	512.000000
mean	0.276365	6577.554688
std	0.212211	7086.477548
min	0.000000	0.000000
25%	0.098483	2310.000000
50%	0.196724	3948.000000
75%	0.436580	7801.750000
max	0.824257	44748.000000

	Bomb Blasts Occurred	Boundary wall	Building condition
satisfactory \			
count	580.000000	580.000000	
580.000000			
mean	26.750000	61.738460	
46.253208			
std	9.128591	27.633865	
29.428277			
min	16.000000	0.000000	
0.000000			
25%	21.250000	40.828917	
20.699990			
50%	25.000000	66.074143	
39.573685			
75%	30.500000	86.335951	
78.663448			
max	41.000000	99.930000	
99.910072			

Basic Statistics:

=====

	% Complete Primary Schools	% Primary Schools with single classroom \
count	576.000000	
576.000000		
mean	0.592544	
0.135134		
std	0.302878	
0.121400		
min	0.000000	
0.000000		
25%	0.340517	
0.029740		
50%	0.691613	
0.096157		
75%	0.846848	
0.211669		
max	0.994764	
0.514768		

	% Primary Schools with single teacher	Area (km ²) \
count	576.000000	512.000000
mean	0.276365	6577.554688
std	0.212211	7086.477548
min	0.000000	0.000000
25%	0.098483	2310.000000
50%	0.196724	3948.000000
75%	0.436580	7801.750000
max	0.824257	44748.000000

	Bomb Blasts Occurred	Boundary wall	Building condition
satisfactory \			
count	580.000000	580.000000	
580.000000			
mean	26.750000	61.738460	
46.253208			
std	9.128591	27.633865	
29.428277			
min	16.000000	0.000000	
0.000000			
25%	21.250000	40.828917	
20.699990			
50%	25.000000	66.074143	
39.573685			
75%	30.500000	86.335951	
78.663448			
max	41.000000	99.930000	
99.910072			

	MeasureGroup 1 Measures	Color By Measure Value \
count	580.000000	580.000000
mean	76.813786	76.813786
std	15.228615	15.228615
min	0.000000	0.000000
25%	68.315035	68.315035
50%	79.409258	79.409258
75%	88.970000	88.970000
max	99.300000	99.300000

	Complete Primary Schools ...	Number of secondary schools \
count	576.000000	580.000000
mean	470.756944	206.648276
std	383.146847	180.088696
min	0.000000	19.000000
25%	125.250000	72.000000
50%	435.000000	151.000000
75%	738.000000	289.000000
max	1786.000000	985.000000

	Population	Primary Schools with single classroom \
count	5.120000e+02	576.000000
mean	1.100493e+06	130.173611
std	1.475205e+06	209.484712
min	0.000000e+00	0.000000
25%	2.848275e+05	21.750000
50%	7.937310e+05	62.500000
75%	1.413893e+06	154.750000
max	1.321563e+07	1553.000000

	Primary Schools with single teacher	Retention score \
count	576.000000	580.000000
mean	246.138889	55.161128
std	342.356003	20.942687
min	0.000000	0.000000
25%	60.000000	40.090000
50%	125.500000	57.052500
75%	289.000000	71.151250
max	2076.000000	100.000000

	School infrastructure score	Terrorist Attacks Affectees
Toilet \		
count	580.000000	580.000000
580.000000		
mean	54.649542	3421.750000
57.289994		
std	25.597288	2037.380781
29.626153		
min	0.000000	1009.000000
0.000000		

25%	32.348388	1604.500000
32.538229		
50%	52.502500	3591.000000
59.935965		
75%	80.403750	5408.250000
85.117674		
max	98.600000	5496.000000
99.924127		

	Total number of schools	Year
count	580.000000	580.000000
mean	1048.889655	2014.500000
std	771.845995	1.118999
min	99.000000	2013.000000
25%	446.000000	2013.750000
50%	842.000000	2014.500000
75%	1487.000000	2015.250000
max	4008.000000	2016.000000

[8 rows x 33 columns]

Kohaq - Tech Intern - Week 1 Task Documentation

Missing Values:

=====

	Column	Missing Count	Missing Percentage
9	Area (km ²)	68	11.724138
41	Population	68	11.724138
1	% Complete Primary Schools	4	0.689655
5	All Four Facilities	4	0.689655
6	Any One Facility	4	0.689655
4	% Primary Schools with single teacher	4	0.689655
3	% Primary Schools with single classroom	4	0.689655
21	Complete Primary Schools	4	0.689655
34	No Facility	4	0.689655
8	Any Two Facilities	4	0.689655
7	Any Three Facilities	4	0.689655
42	Primary Schools with single classroom	4	0.689655
43	Primary Schools with single teacher	4	0.689655
0	% Boys Enrolled	1	0.172414
2	% Girls Enrolled	1	0.172414

Cleaning percentage columns...

Province Distribution:

=====

Province

Punjab	144
Balochistan	120
KP	100
Sindh	92
FATA	52
AJK	40
GB	28
ICT	4

Name: count, dtype: int64

Year Distribution:

=====

Year

2013	145
2014	145
2015	145
2016	145

Name: count, dtype: int64

Kohaq - Tech Intern - Week 1 Task Documentation

Gender Gap Analysis:

=====

Average Gender Gap: 45.14%

Standard Deviation: 27.50%

Max Gender Gap: 100.00%

Min Gender Gap: 1.40%

Gender Gap by Province:

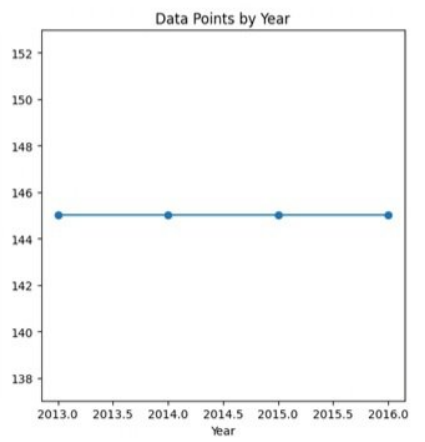
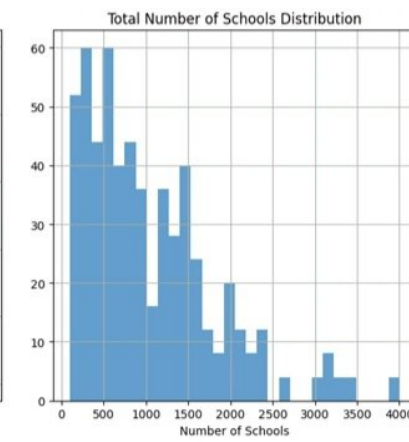
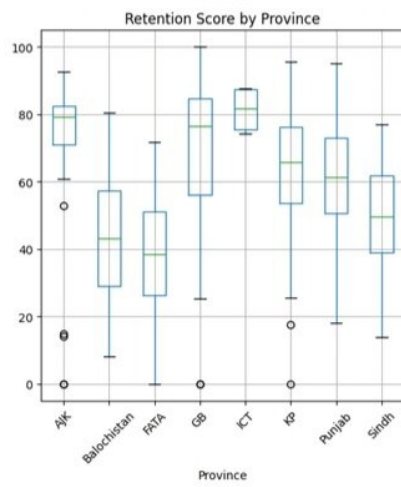
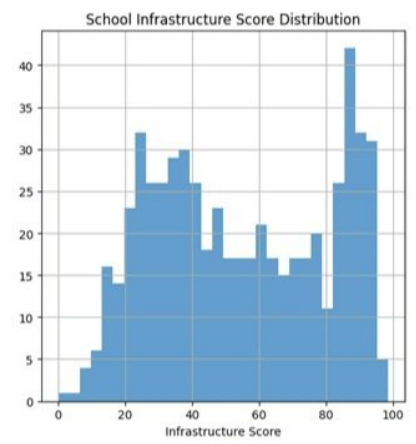
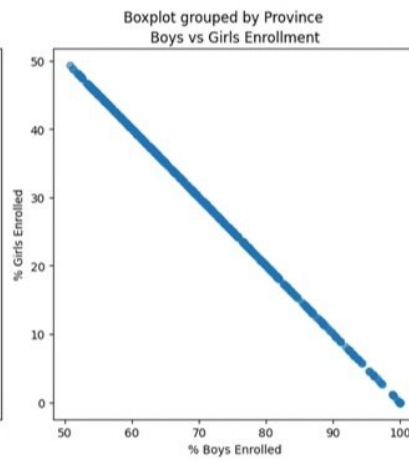
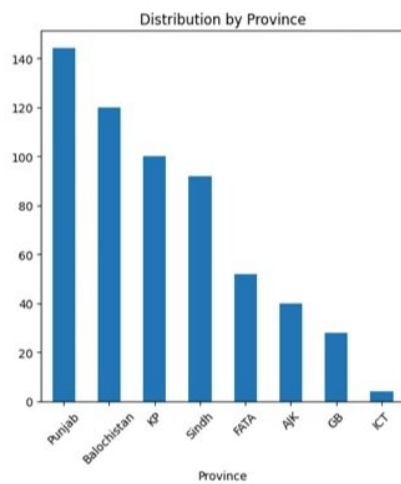
	mean	std
Province		
FATA	65.23	24.29
Balochistan	63.50	26.35
KP	48.85	24.00
Sindh	48.53	22.03
GB	39.79	31.57
Punjab	26.55	17.99
AJK	20.79	14.67
ICT	12.95	5.03

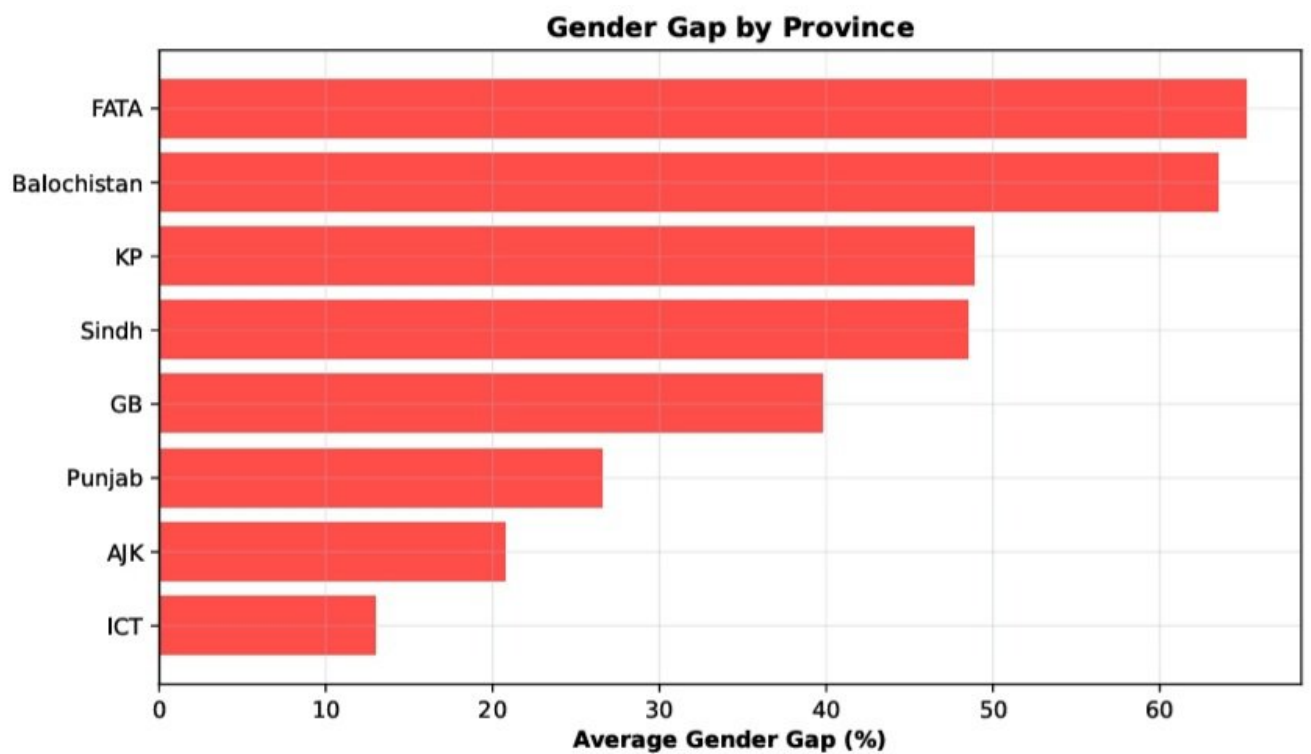
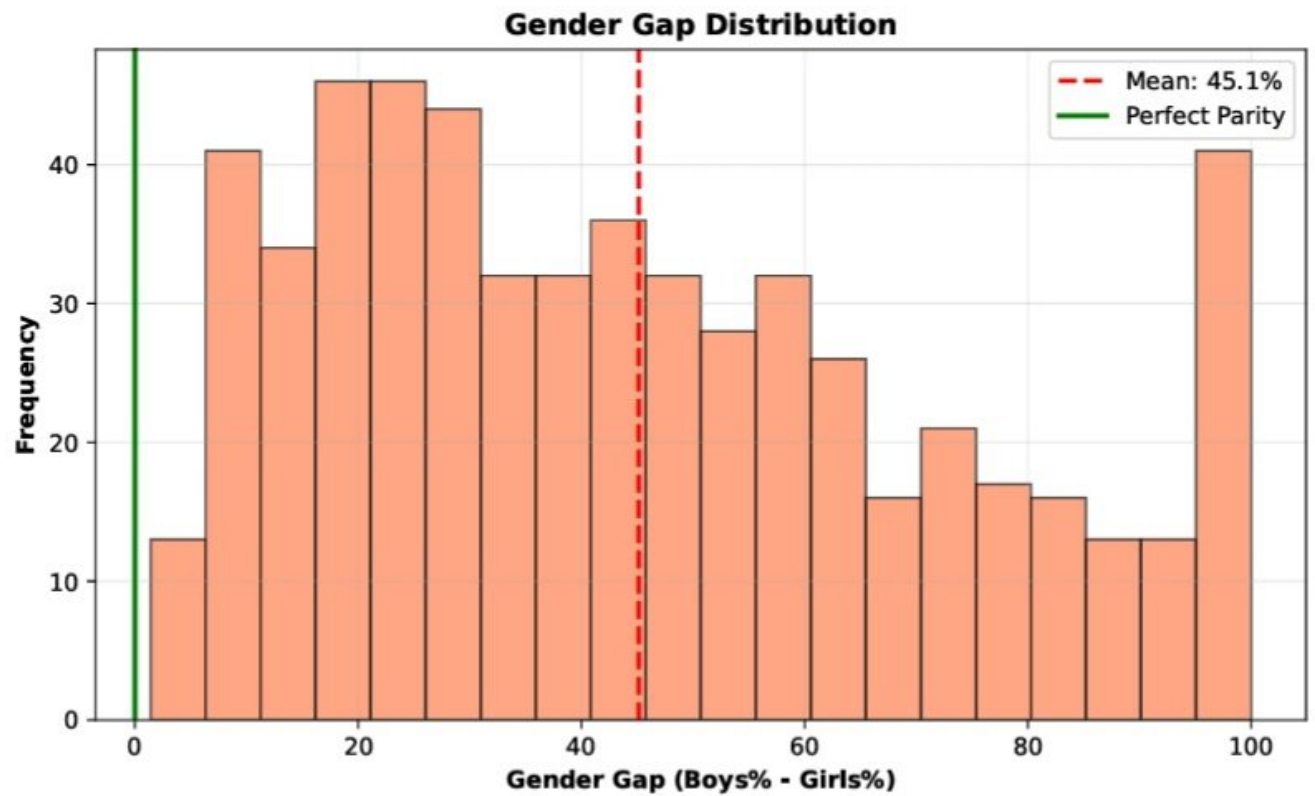
=====

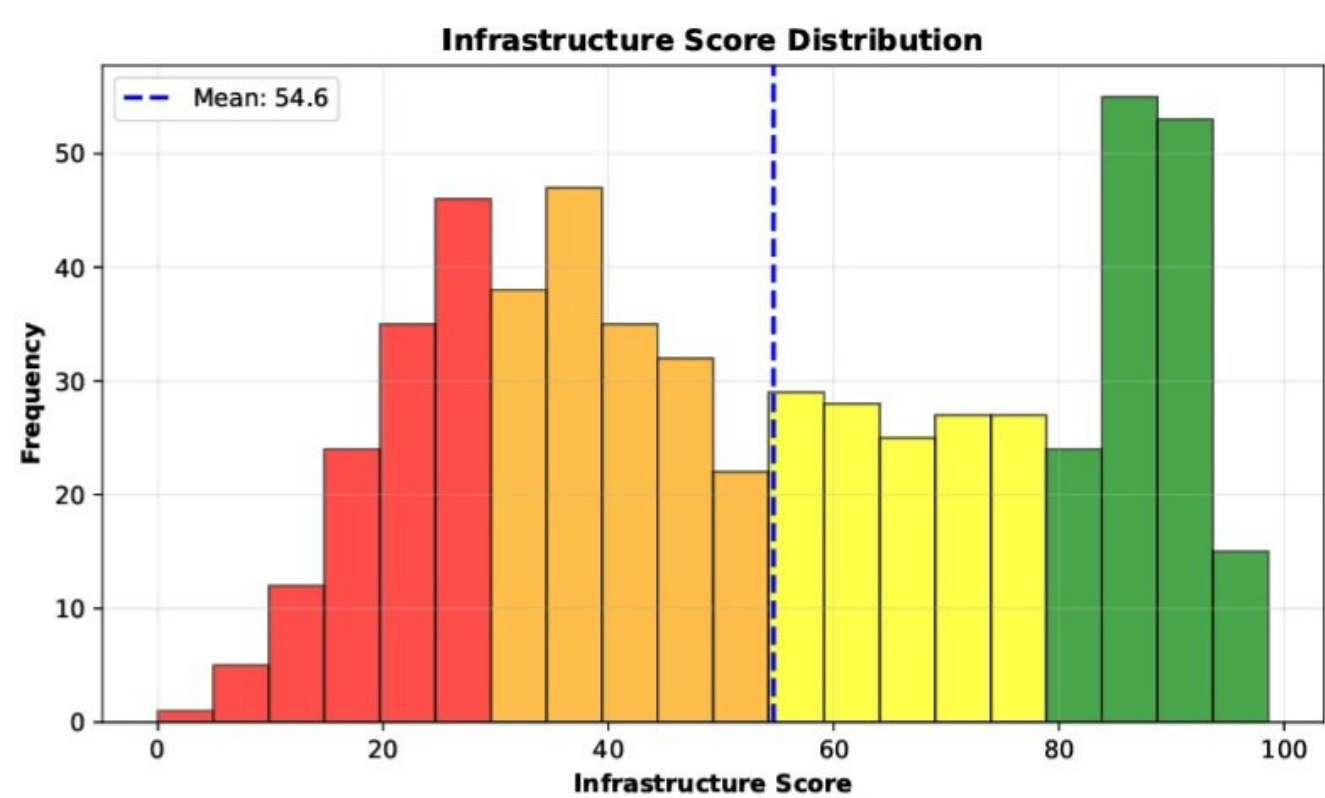
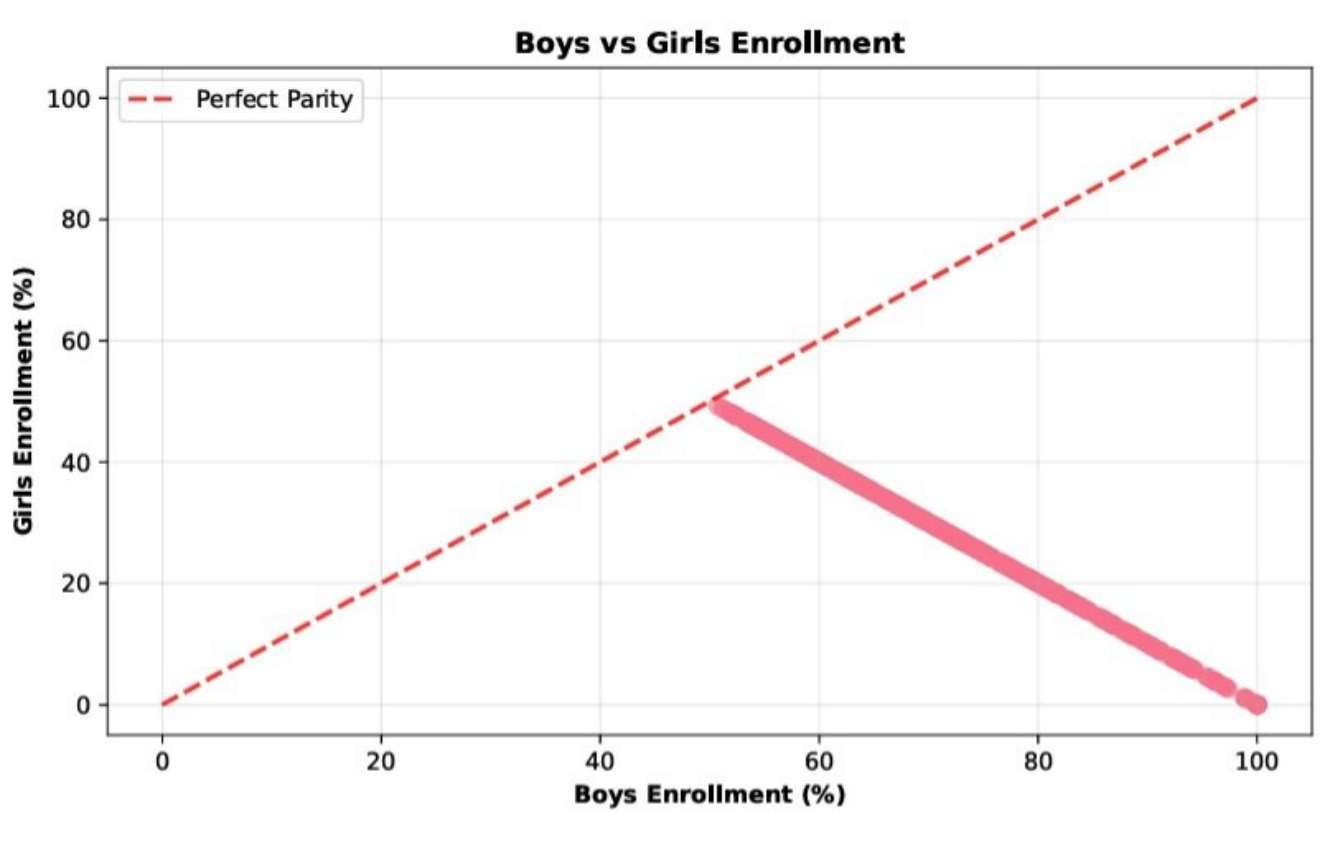
ANALYSIS COMPLETE!

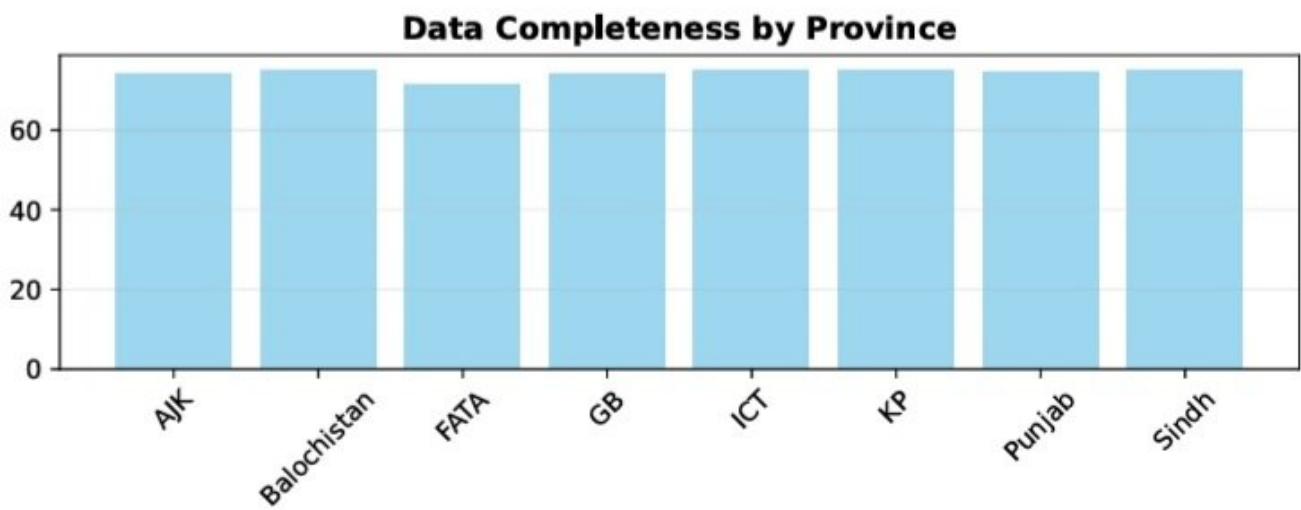
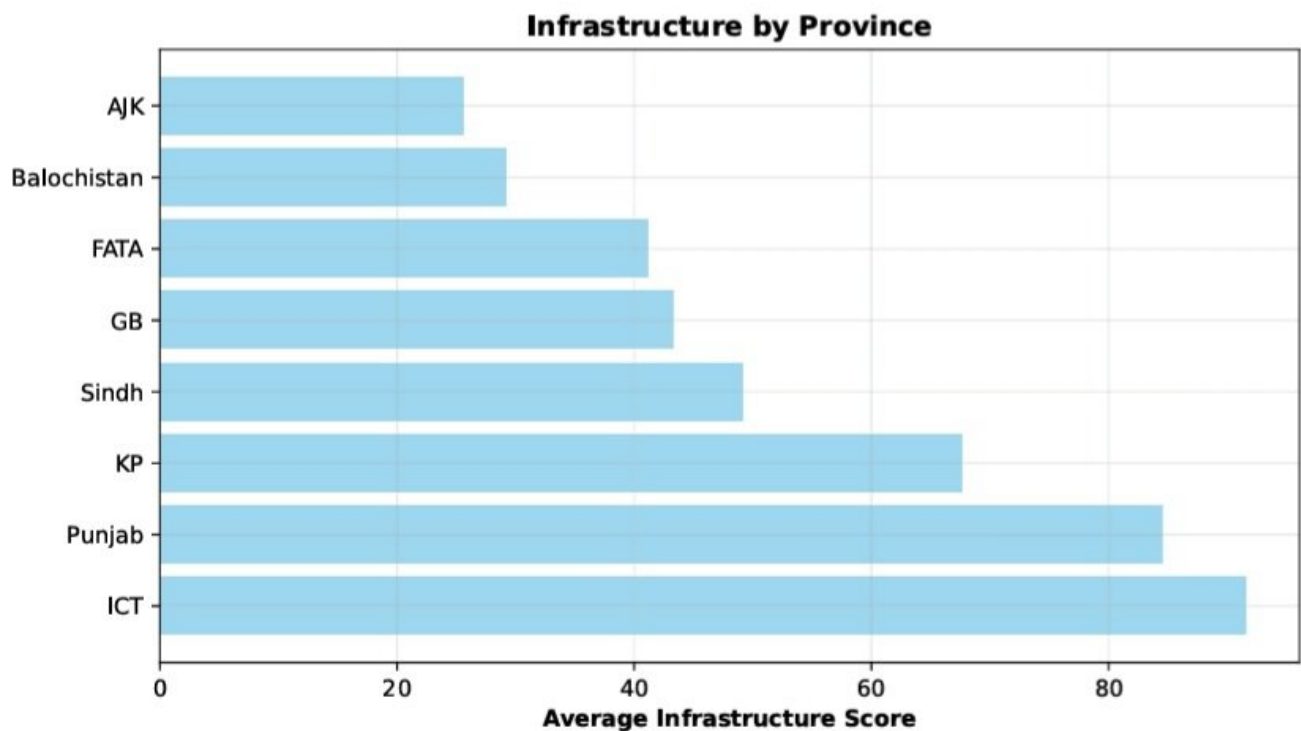
=====

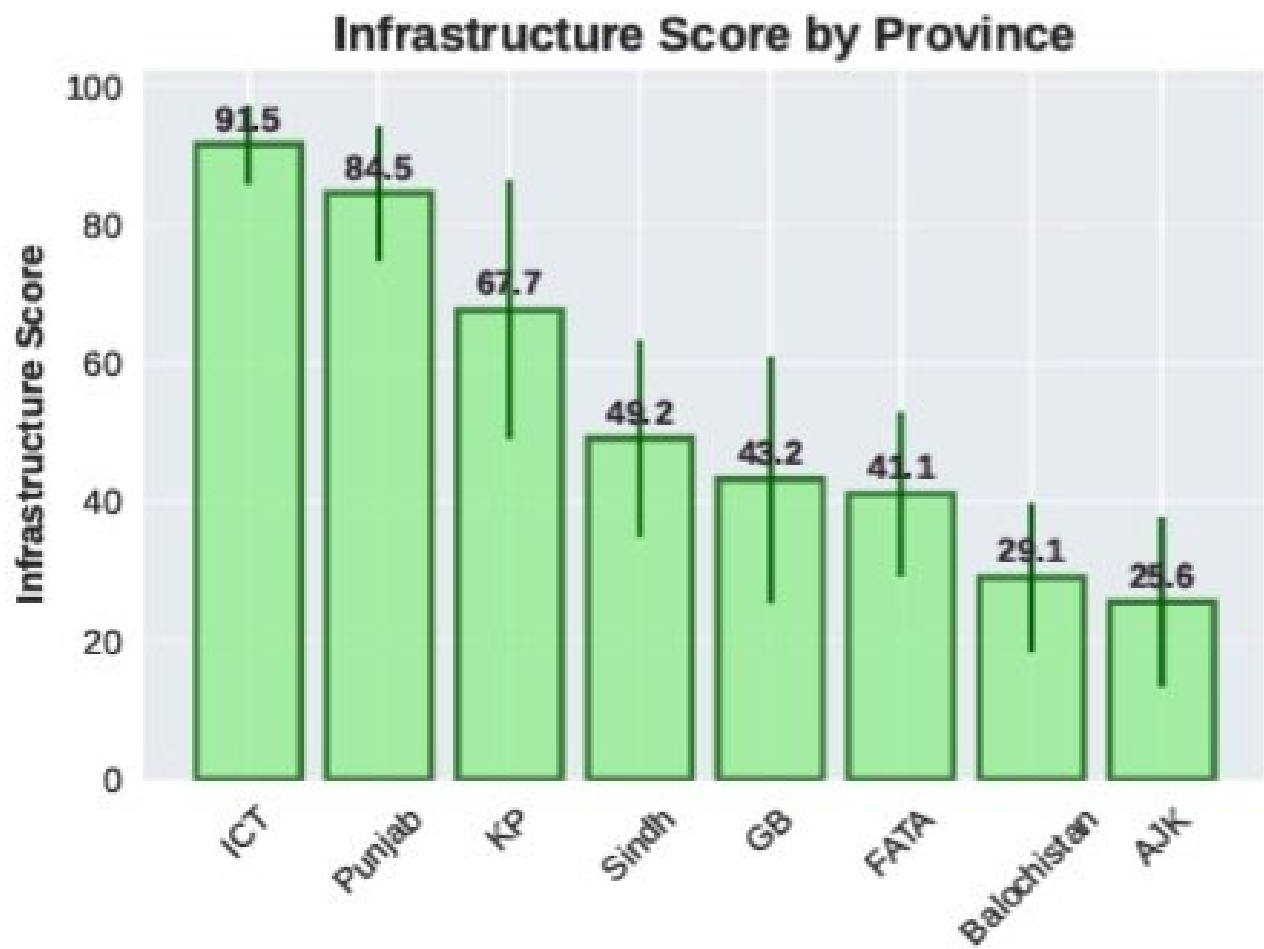
Kohaq - Tech Intern - Week 1 Task Documentation

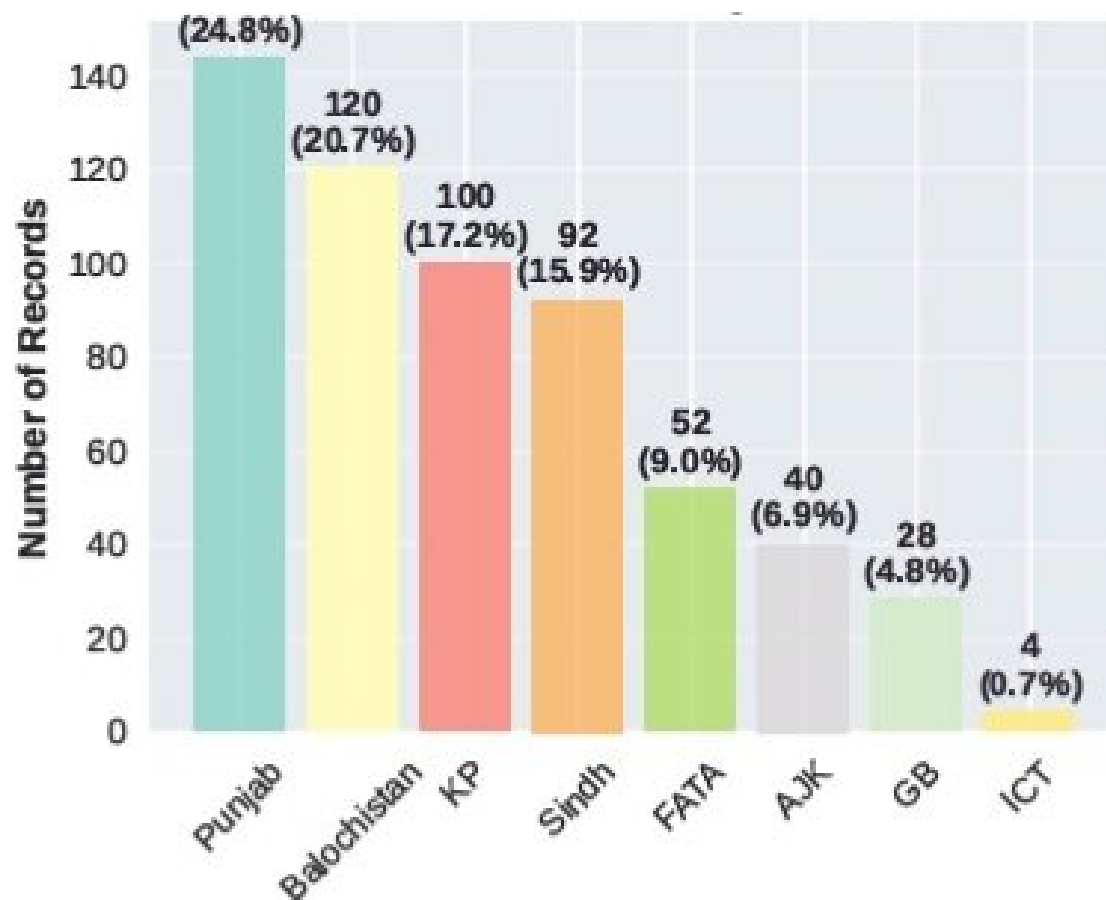
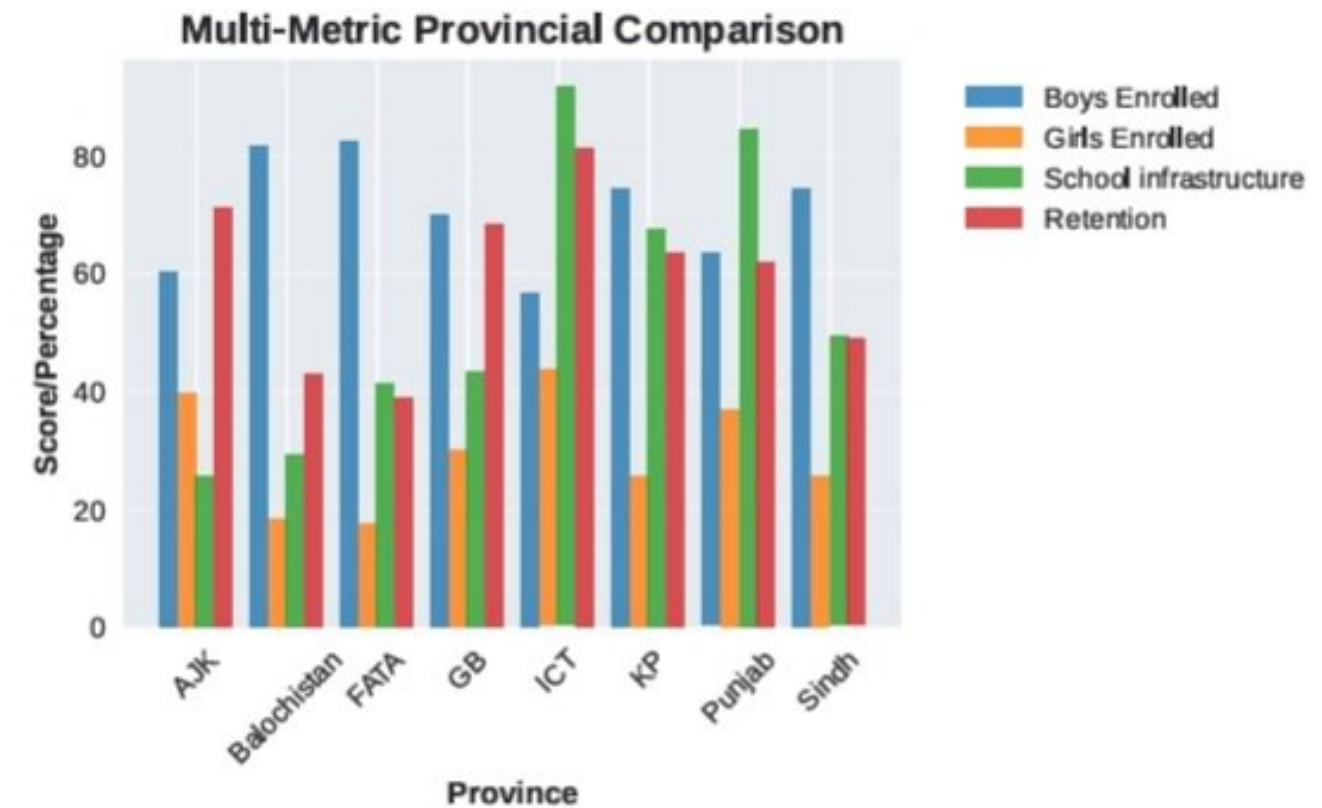




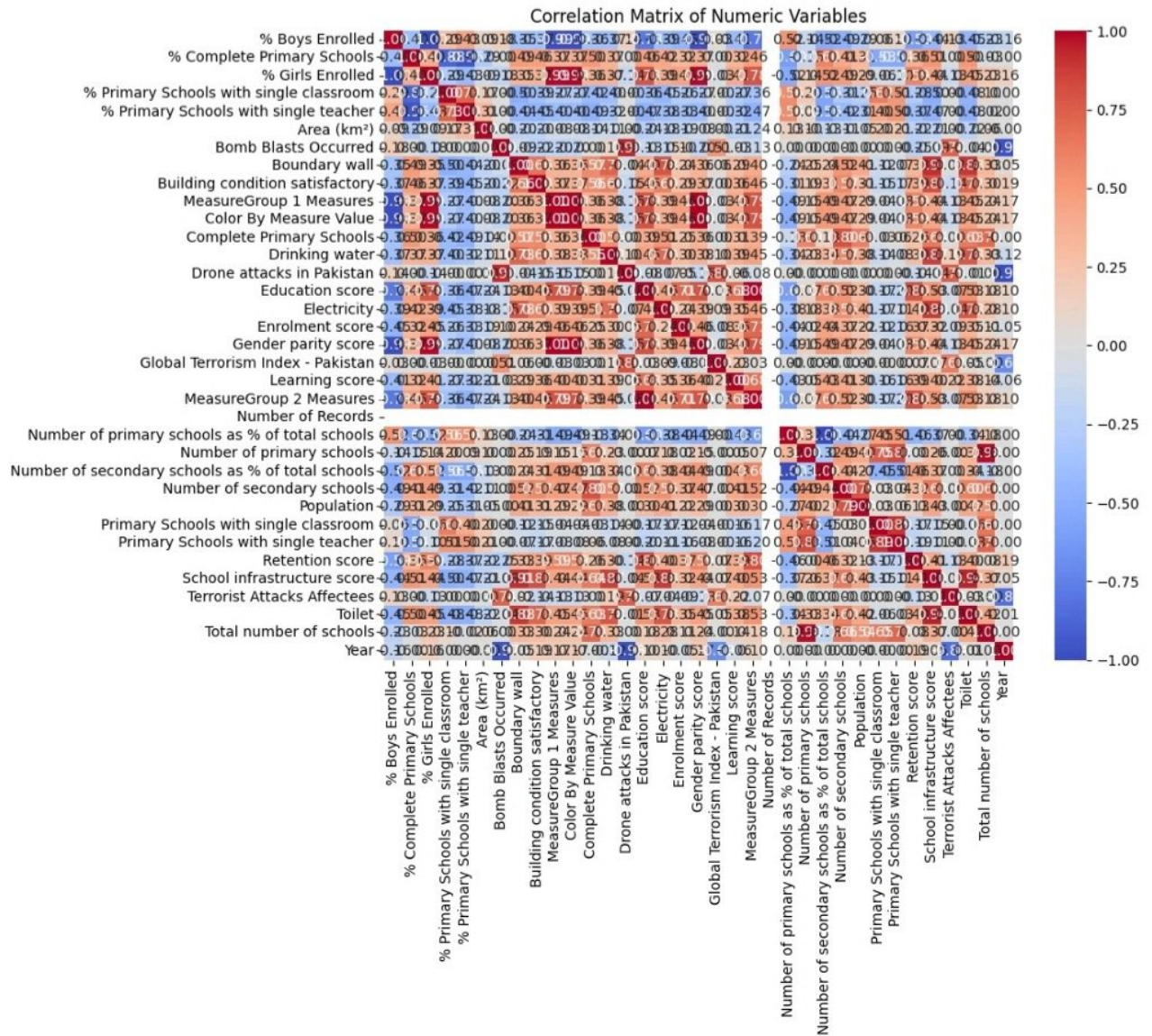








Kohaq – Tech Intern – Week 1 Task Documentation



Highly Correlated Pairs (>0.7 or <-0.7):

- % Boys Enrolled \leftrightarrow % Girls Enrolled: -1.000
- % Boys Enrolled \leftrightarrow MeasureGroup 1 Measures: -0.988
- % Boys Enrolled \leftrightarrow Color By Measure Value: -0.988
- % Boys Enrolled \leftrightarrow Education score: -0.785
- % Boys Enrolled \leftrightarrow Gender parity score: -0.988
- % Boys Enrolled \leftrightarrow MeasureGroup 2 Measures: -0.785
- % Complete Primary Schools \leftrightarrow % Primary Schools with single classroom: -0.880
- % Complete Primary Schools \leftrightarrow % Primary Schools with single teacher: -0.955
- % Girls Enrolled \leftrightarrow MeasureGroup 1 Measures: 0.988
- % Girls Enrolled \leftrightarrow Color By Measure Value: 0.988
- % Girls Enrolled \leftrightarrow Education score: 0.785
- % Girls Enrolled \leftrightarrow Gender parity score: 0.988
- % Girls Enrolled \leftrightarrow MeasureGroup 2 Measures: 0.785
- % Primary Schools with single classroom \leftrightarrow % Primary Schools with single teacher: 0.733
- Bomb Blasts Occurred \leftrightarrow Drone attacks in Pakistan: 0.923
- Bomb Blasts Occurred \leftrightarrow Terrorist Attacks Affectees: 0.739
- Bomb Blasts Occurred \leftrightarrow Year: -0.968
- Boundary wall \leftrightarrow Drinking water: 0.733
- Boundary wall \leftrightarrow Electricity: 0.782
- Boundary wall \leftrightarrow School infrastructure score: 0.909
- Boundary wall \leftrightarrow Toilet: 0.883
- Building condition satisfactory \leftrightarrow School infrastructure score: 0.808
- Building condition satisfactory \leftrightarrow Toilet: 0.725
- MeasureGroup 1 Measures \leftrightarrow Color By Measure Value: 1.000
- MeasureGroup 1 Measures \leftrightarrow Education score: 0.791
- MeasureGroup 1 Measures \leftrightarrow Gender parity score: 1.000
- MeasureGroup 1 Measures \leftrightarrow MeasureGroup 2 Measures: 0.791
- Color By Measure Value \leftrightarrow Education score: 0.791
- Color By Measure Value \leftrightarrow Gender parity score: 1.000
- Color By Measure Value \leftrightarrow MeasureGroup 2 Measures: 0.791
- Complete Primary Schools \leftrightarrow Number of secondary schools: 0.800
- Complete Primary Schools \leftrightarrow Total number of schools: 0.705
- Drinking water \leftrightarrow Electricity: 0.704
- Drinking water \leftrightarrow School infrastructure score: 0.867
- Drinking water \leftrightarrow Toilet: 0.775
- Drone attacks in Pakistan \leftrightarrow Global Terrorism Index - Pakistan: 0.796
- Drone attacks in Pakistan \leftrightarrow Terrorist Attacks Affectees: 0.754
- Drone attacks in Pakistan \leftrightarrow Year: -0.977
- Education score \leftrightarrow Enrolment score: 0.715
- Education score \leftrightarrow Gender parity score: 0.791
- Education score \leftrightarrow MeasureGroup 2 Measures: 1.000
- Education score \leftrightarrow Retention score: 0.796
- Electricity \leftrightarrow School infrastructure score: 0.878
- Electricity \leftrightarrow Toilet: 0.776
- Enrolment score \leftrightarrow MeasureGroup 2 Measures: 0.715

Gender parity score <-> MeasureGroup 2 Measures: 0.791
MeasureGroup 2 Measures <-> Retention score: 0.796
Number of primary schools as % of total schools <-> Number of
secondary schools as % of total schools: -1.000
Number of primary schools <-> Primary Schools with single classroom:
0.749
Number of primary schools <-> Primary Schools with single teacher:
0.823
Number of primary schools <-> Total number of schools: 0.979
Number of secondary schools <-> Population: 0.787
Primary Schools with single classroom <-> Primary Schools with single
teacher: 0.894
Primary Schools with single teacher <-> Total number of schools: 0.721
School infrastructure score <-> Toilet: 0.943
Terrorist Attacks Affectees <-> Year: -0.836