The Impact of Educational and Family Backgrounds on Juvenile Crimes.

Acknowledgement

I would like to express my deepest gratitude to all those who have supported me throughout this journey.

First and foremost, I would like to thank Visva-Bharati University for the great learning it provided which were required to accomplish this project work, I would also like to thanks my friends whose valuable guidance has been the ones that helped me throughout this project and make it a full-proof success. Their suggestions and instructions have served as the major contributor to the completion of the project.

Then I would like to thank my parents and siblings who have always created a positive surrounding near me throughout various phases of the completion of the project.

Finally, I would like to thank my classmates who have helped me a lot. In conclusion, the successful accomplishment of this project would not have been possible without the support and blessings of all these wonderful individuals. I am truly humbled and grateful for their unwavering support. Thank you all for being a part of this journey and for believing in me.

Introduction

Crime rate is increasing day by day in India and the most shocking trend that can be noticed is that it is up-heaving among Juveniles too. Criminal cases of all sorts ranging from thefts to murder or smuggling to sexual crimes is committed by them. The crimes committed by young people below a specific age (18 in most countries) are juvenile crimes. In India, juvenile crime is a grim reality. A juvenile is a child who has not reached the age at which they may be held accountable for their criminal activities in the same way that an adult can. When referring to a young criminal offender, the term juvenile is used. As a result, a juvenile is a child who is accused of doing certain acts or omissions that are illegal and have been classified as such by penal laws. The overall rate of juvenile arrests and prosecutions has increased in recent years. This may be the result of societal pressures like drugs or poverty, lack of education, abusive family or it may be due to more aggressive enforcement and zero-tolerance policies. The juvenile system differs from the adult criminal justice system in its greater emphasis on rehabilitation and prevention rather than punishment.

Some Examples of Juvenile Crimes

Most, but not all, juvenile crimes are less severe than those that would require them to be treated as an adult. Some of the more common crimes that juveniles commit includes:

- Vandalism
- Shoplifting
- Alcohol Infractions
- Drug Possessions
- rape

Reasons behind Juvenile Crimes in India

No one is born with the potential to be a criminal. Circumstances have shaped them into who they are. The socio-cultural environment, both within and outside of one's household, has a big influence on one's life and general personality. However, in India, it is poverty and the impact of the media, particularly social media, that encourages youths to engage in illegal activity. Poverty is one of the leading factors of a child's involvement in criminal activity. Also, the current function of social media, which has a more destructive impact on young brains

1. Abusive Family:
2. Lack of Education:
3. Poverty:
4. Social Evils:
5. Friends and Companions:
6. Role of Social Media

• Psychological Reasons:

1. Mental Illness
2. Personality traits:
3. Individualized Emotional Issues:

Project Objectives

In this project work our objective is to study whether the average proportions of juvenile crimes differ significantly over different Zones of India, the Educational Backgrounds of the candidate and their Family Backgrounds.

We also check for the significance of the interaction effects between these factors, if any.

About The Data

The data is collected from the website: https://ncrb.gov.in.

This is a data showing the total crimes(state/UT wise) for Educational Backgrounds and for different Family Backgrounds of juveniles. Educational Backgrounds are divided into four categories- Illiterate, upto Primary, Above Primary to Matric, Above Matric to Higher secondary and above Higher Secondary. For the data-set of 2016 The Educational Backgrounds has only four categories excluding the category Above Higher Secondary. Family Backgrounds are divided into three categories- Living with Parents, Living with Guardians and Homeless. In the data-set,

E1	Illiterate
E2	upto Primary
E3	Above Primary to Matric
E4	Above Matric to Higher Secondary
E5	Above Higher Secondary
F1	Living with Parents
F2	Living with Guardians
F3	Homeless

Here, the table shows State-wise Crimes Committed by Juveniles in 2016,2017,2018,2019 and 2020.

2016	EDUCATIONAL BACKGROUNDS					FAMILY BACKGROUNDS			
State/UT	E1	E2	E3	E4	TOTAL	F1	F2	F3	TOTAL
Andhra Pradesh	**228**	**427**	**475**	**99**	**1229**	**910**	**150**	**169**	**12292**
Arunachal Pradesh	3	10	64	3	80	73	7	0	80
Assam	70	216	182	12	480	337	125	18	480
Bihar	530	492	1216	388	2626	1910	648	68	2626
Chhattisgarh	135	857	1116	286	2394	2206	169	19	2394
Goa	2	7	17	0	26	24	1	1	26
Gujarat	160	1160	677	88	2085	1998	49	38	2085
Haryana	169	331	676	182	1358	1226	52	80	1358
Himachal Pradesh	7	59	158	39	263	259	1	3	263
Jammu & Kashmir	17	61	194	47	319	319	0	0	319
Jharkhand	40	42	49	11	142	101	30	11	142
Karnataka	18	163	369	77	627	604	20	3	627
Kerala	4	98	669	289	1060	917	124	19	1060
Madhya Pradesh	975	2730	3542	1217	8464	7291	889	284	8464
Maharashtra	420	2764	4129	399	7712	6814	739	159	7712
Manipur	0	1	11	0	12	12	0	0	12
Meghalaya	12	40	29	12	93	92	1	0	93
Mizoram	0	31	32	0	63	42	21	0	63
Nagaland	5	14	6	0	25	17	3	5	25
Odisha	270	624	336	55	1285	1218	65	2	1285
Punjab	25	40	82	10	157	151	4	2	157
Rajasthan	266	866	1411	400	2943	2703	188	52	2943
Sikkim	3	11	25	0	39	30	9	0	39
Tamil Nadu	223	783	1607	197	2810	2385	258	167	2810
Telangana	437	281	393	71	1182	1058	69	55	1182
Tripura	3	29	10	0	42	21	21	0	42
Uttar Pradesh	279	509	624	175	1587	1193	279	115	1587
Uttarakhand	24	51	31	45	151	99	36	16	151
West Bengal	52	236	534	16	838	674	143	21	838
A & N Islands	0	0	10	3	13	13	0	0	13
Chandigarh	7	52	78	4	141	103	32	6	141
D&N Haveli	0	0	10	0	10	10	0	0	10
Daman & Diu	0	0	8	0	8	8	0	0	8
Delhi UT	1008	1510	1185	105	3808	3159	405	244	3808
Lakshadweep	0	0	0	0	0	0	0	0	0
Puducherry	20	6	59	14	99	84	12	3	99
TOTAL (ALL INDIA)	**5412**	**14501**	**20014**	**4244**	**44171**	**38061**	**4550**	**1560**	**44171**

2017	Educational Backgrounds						Family Background			
State/UT	**E1**	**E2**	**E3**	**E4**	**E5**	**Total**	**F1**	**F2**	**F3**	**Total**
Andhra Pradesh	**267**	**395**	**626**	**120**	**33**	**1441**	**1139**	**119**	**183**	**14412**
Arunachal Pradesh	4	20	71	3	0	98	78	20	0	98
Assam	55	37	83	16	0	191	116	12	63	191
Bihar	163	212	456	333	183	1347	830	347	170	1347
Chhattisgarh	188	852	1149	204	24	2417	1979	325	113	2417
Goa	0	3	20	6	0	29	27	0	2	29
Gujarat	120	609	1398	207	29	2363	2290	33	40	2363
Haryana	125	367	489	146	69	1196	1073	37	86	1196
Himachal Pradesh	10	42	111	45	18	226	201	13	12	226
Jammu & Kashmir	5	27	143	80	7	262	244	9	9	262
Jharkhand	0	0	78	6	0	84	7	0	77	84
Karnataka	34	139	368	130	20	691	630	31	30	691
Kerala	1	54	327	247	53	682	613	51	18	682
Madhya Pradesh	704	2111	2953	1274	124	7166	5513	774	879	7166
Maharashtra	524	2234	3385	1312	316	7771	6620	593	558	7771
Manipur	0	0	7	7	0	14	13	0	1	14
Meghalaya	12	43	36	23	0	114	94	11	9	114
Mizoram	0	21	1	0	0	22	7	13	2	22
Nagaland	0	4	18	0	0	22	13	7	2	22
Odisha	483	468	232	49	10	1242	1242	0	0	1242
Punjab	35	59	123	47	6	270	247	1	22	270
Rajasthan	244	549	1198	455	246	2692	2232	341	119	2692
Sikkim	0	13	23	0	0	36	21	14	1	36
Tamil Nadu	123	601	1559	519	117	2919	2705	121	93	2919
Telangana	193	233	403	361	65	1255	998	98	159	1255
Tripura	1	21	15	0	0	37	37	0	0	37
Uttar Pradesh	82	158	282	276	91	889	582	76	231	889
Uttarakhand	59	48	51	22	2	182	142	22	18	182
West Bengal	72	173	381	98	9	733	549	113	71	733
A&N Islands	0	2	19	4	0	25	23	2	0	25
Chandigarh	10	46	106	2	0	164	116	46	2	164
D&N Haveli	0	3	6	6	0	15	15	0	0	15
Daman & Diu	0	2	12	0	0	14	14	0	0	14
Delhi	810	1217	1403	227	9	3666	3147	279	240	3666
Lakshadweep	0	0	0	0	0	0	0	0	0	0
Puducherry	0	27	34	35	49	145	137	5	3	145
TOTAL ALL INDIA	**4324**	**10790**	**17566**	**6260**	**1480**	**40420**	**33694**	**3513**	**3213**	**40420**

2018 State/UT	Educational Backgrounds						Family Background			
	E1	E2	E3	E4	E5	Total	F1	F2	F3	Total
Andhra Pradesh	**155**	**397**	**403**	**139**	**17**	**1111**	**903**	**75**	**133**	**11112**
Arunachal Pradesh	0	13	45	1	3	62	43	10	9	62
Assam	19	40	84	9	0	152	144	8	0	152
Bihar	82	113	177	187	55	614	454	112	48	614
Chhattisgarh	148	732	1194	210	40	2324	1993	116	215	2324
Goa	0	6	19	0	0	25	20	2	3	25
Gujarat	113	1062	1048	246	37	2506	2450	49	7	2506
Haryana	190	225	557	310	61	1343	1132	159	52	1343
Himachal Pradesh	10	39	141	89	8	287	283	4	0	287
Jammu & Kashmir	89	81	139	72	12	393	381	4	8	393
Jharkhand	12	17	60	0	0	89	58	17	14	89
Karnataka	30	207	401	95	11	744	672	46	26	744
Kerala	0	25	264	262	104	655	606	32	17	655
Madhya Pradesh	754	2055	2323	772	198	6102	4843	904	355	6102
Maharashtra	490	1840	4043	1193	152	7718	6844	522	352	7718
Manipur	0	1	15	0	0	16	16	0	0	16
Meghalaya	11	36	39	2	0	88	83	2	3	88
Mizoram	0	10	27	0	0	37	31	6	0	37
Nagaland	3	7	10	1	0	21	18	3	0	21
Odisha	8	169	799	179	0	1155	1147	0	8	1155
Punjab	25	91	136	41	4	297	287	6	4	297
Rajasthan	201	657	1042	553	157	2610	2087	309	214	2610
Sikkim	0	7	10	1	0	18	18	0	0	18
Tamil Nadu	212	789	1157	525	33	2716	2186	420	110	2716
Telangana	195	234	725	337	29	1520	1244	71	205	1520
Tripura	0	27	26	1	0	54	53	1	0	54
Uttar Pradesh	101	284	369	329	67	1150	859	100	191	1150
Uttarakhand	29	73	52	23	12	189	48	62	79	189
West Bengal	41	169	320	60	1	591	447	70	74	591
A&N Islands	6	11	14	4	0	35	26	9	0	35
Chandigarh	8	95	89	21	2	215	196	12	7	215
D&N Haveli	0	6	12	0	0	18	0	18	0	18
Daman & Diu	0	2	0	4	0	6	6	0	0	6
Delhi	678	1095	1261	248	33	3315	2775	283	257	3315
Lakshadweep	0	0	0	0	0	0	0	0	0	0
Puducherry	0	51	23	0	6	80	80	0	0	80
TOTAL ALL INDIA	**3610**	**10666**	**17024**	**5914**	**1042**	**38256**	**32433**	**3432**	**2391**	**38256**

2019 State/UT	Educational Backgrounds						Family Background			
	E1	E2	E3	E4	E5	Total	F1	F2	F3	Total
Andhra Pradesh	**149**	**454**	**342**	**104**	**23**	**1072**	**897**	**53**	**122**	**10722**
Arunachal Pradesh	0	9	27	3	0	39	25	12	2	39
Assam	13	30	91	2	0	136	125	11	0	136
Bihar	385	390	518	279	48	1620	937	454	229	1620
Chhattisgarh	130	779	817	262	43	2031	1577	243	211	2031
Goa	1	9	19	4	0	33	32	0	1	33
Gujarat	124	1108	964	176	12	2384	2326	34	24	2384
Haryana	256	446	601	249	44	1596	1301	171	124	1596
Himachal Pradesh	5	59	123	36	3	226	214	11	1	226
Jammu & Kashmir	21	53	255	74	4	407	399	3	5	407
Jharkhand	0	20	57	0	0	77	72	5	0	77
Karnataka	27	163	313	65	18	586	473	27	86	586
Kerala	0	36	302	267	73	678	577	88	13	678
Madhya Pradesh	774	1877	2565	756	214	6186	4816	1022	348	6186
Maharashtra	283	1514	3360	1125	172	6454	5834	416	204	6454
Manipur	0	0	6	2	0	8	8	0	0	8
Meghalaya	7	40	41	1	0	89	87	1	1	89
Mizoram	6	13	4	2	0	25	22	3	0	25
Nagaland	2	3	1	0	0	6	4	2	0	6
Odisha	126	411	848	27	0	1412	1299	2	111	1412
Punjab	21	67	150	47	8	293	267	12	14	293
Rajasthan	181	751	1267	638	185	3022	2581	372	69	3022
Sikkim	0	0	4	0	0	4	4	0	0	4
Tamil Nadu	97	658	1812	636	101	3304	2899	302	103	3304
Telangana	266	140	488	663	29	1586	1274	87	225	1586
Tripura	0	27	21	0	0	48	48	0	0	48
Uttar Pradesh	139	209	458	282	41	1129	923	74	132	1129
Uttarakhand	18	30	29	13	11	101	40	17	44	101
West Bengal	41	169	320	60	1	591	447	70	74	591
A&N Islands	1	9	10	4	0	24	19	5	0	24
Chandigarh	0	22	178	5	3	208	191	9	8	208
D&N Haveli	0	4	8	8	0	20	7	13	0	20
Daman & Diu	1	3	10	5	0	19	16	2	1	19
Delhi	646	826	1623	152	9	3256	2616	203	437	3256
Lakshadweep	0	0	0	0	0	0	0	0	0	0
Puducherry	3	8	0	0	4	15	2	0	13	15
TOTAL ALL INDIA	**3723**	**10337**	**17632**	**5947**	**1046**	**38685**	**32359**	**3724**	**2602**	**38685**

2020 State/UT	Educational Backgrounds						Family Background			
	E1	E2	E3	E4	E5	Total	F1	F2	F3	Total
Andhra Pradesh	**122**	**240**	**424**	**105**	**9**	**900**	**722**	**65**	**113**	**9002**
Arunachal Pradesh	0	2	23	8	0	33	25	8	0	33
Assam	41	100	98	15	0	254	226	14	14	254
Bihar	218	213	311	93	34	869	654	147	68	869
Chhattisgarh	159	830	1346	203	15	2553	2157	284	112	2553
Goa	0	10	15	3	1	29	27	1	1	29
Gujarat	86	933	924	197	31	2171	2099	25	47	2171
Haryana	127	322	614	159	49	1271	1013	130	128	1271
Himachal Pradesh	2	55	149	39	0	245	242	3	0	245
Jharkhand	0	31	13	15	0	59	59	0	0	59
Karnataka	12	98	332	86	3	531	497	18	16	531
Kerala	0	20	224	166	24	434	402	20	12	434
Madhya Pradesh	465	1808	2463	611	150	5497	3520	1166	811	5497
Maharashtra	179	1120	2774	936	71	5080	4307	686	87	5080
Manipur	5	8	13	2	0	28	26	2	0	28
Meghalaya	4	22	46	0	0	72	69	3	0	72
Mizoram	7	3	12	0	0	22	22	0	0	22
Nagaland	0	2	7	2	0	11	5	3	3	11
Odisha	11	262	1009	78	0	1360	1279	53	28	1360
Punjab	23	64	193	40	12	332	296	14	22	332
Rajasthan	178	666	1261	605	220	2930	2618	252	60	2930
Sikkim	0	2	10	1	0	13	13	0	0	13
Tamil Nadu	60	712	2339	487	245	3843	3588	170	85	3843
Telangana	174	233	422	355	82	1266	1120	56	90	1266
Tripura	1	6	18	4	0	29	18	1	10	29
Uttar Pradesh	144	293	575	382	64	1458	1081	208	169	1458
Uttarakhand	20	13	28	17	7	85	40	36	9	85
West Bengal	49	111	401	62	7	630	440	79	111	630
A&N Islands	5	5	12	3	0	25	24	1	0	25
Chandigarh	2	12	58	6	0	78	68	6	4	78
D&N Haveli and Daman & Diu	0	3	26	0	0	29	27	2	0	29
Delhi	406	995	1265	238	36	2940	2396	289	255	2940
Jammu & Kashmir	11	45	87	48	24	215	199	0	16	215
Ladakh	0	0	0	0	0	0	0	0	0	0
Lakshadweep	0	0	2	0	0	2	2	0	0	2
Puducherry	41	13	0	0	4	58	4	0	54	58
TOTAL ALL INDIA	**2552**	**9252**	**17494**	**4966**	**1088**	**35352**	**29285**	**3742**	**2325**	**35352**

For working with the data it is divided into six Zones and for analysis the proportion of the no. of crimes and the crime totals for each sub category are taken.

Zones	States
Eastern	Odisha
	Bihar
	West Bengal
	Andaman & Nicobar Island
	Jharkhand
Western	Maharashtra
	Gujarat
	Goa
	Lakshadweep
	D& N Haveli and Daman & Diu
Northern	Rajasthan
	Chandigarh
	Himachal Pradesh
	Uttarakhand
	Punjab
	Haryana
	Jammu & Kashmir
	Delhi
Southern	Andhra Pradesh
	Karnataka
	Kerala
	Telangana
	Tamil Nadu
	Puducherry
Central	Chhattisgarh
	Madhya Pradesh
	Uttar Pradesh
North Eastern	Tripura
	Arunachal Pradesh
	Assam
	Manipur
	Meghalaya
	Mizoram
	Nagaland
	Sikkim

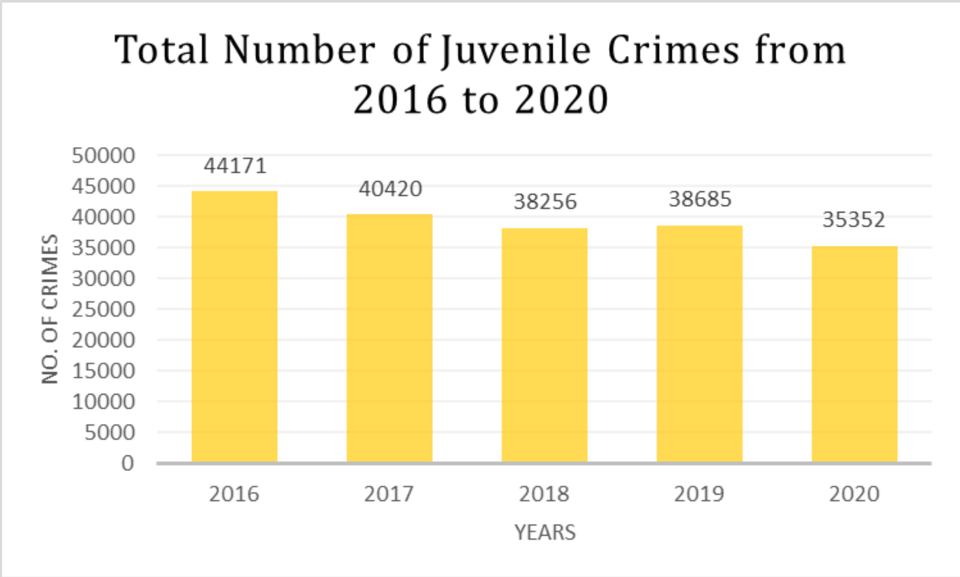
The final table is as follows:

Table Showing the Zones and The Corresponding Crime Rates

	2020	Educational Backgrounds					Family Background		
Zones	State/UT	E1	E2	E3	E4	E5	F1	F2	F3
Northern	Rajasthan	0.060751	0.227304	0.430375	0.206485	0.075085	0.893515	0.086007	0.020478
	Chandigarh	0.025641	0.153846	0.74359	0.076923	0	0.871795	0.076923	0.051282
	Himachal Prac	0.008163	0.22449	0.608163	0.159184	0	0.987755	0.012245	0
	Uttarakhand	0.235294	0.152941	0.329412	0.2	0.082353	0.470588	0.423529	0.105882
	Punjab	0.069277	0.192771	0.581325	0.120482	0.036145	0.891566	0.042169	0.066265
	Haryana	0.099921	0.253344	0.483084	0.125098	0.038552	0.79701	0.102282	0.100708
	Jammu & Kasl	0.051163	0.209302	0.404651	0.223256	0.111628	0.925581	0	0.074419
	Delhi	0.138095	0.338435	0.430272	0.080952	0.012245	0.814966	0.098299	0.086735
Southern	Karnataka	0.022599	0.184557	0.625235	0.161959	0.00565	0.93597	0.033898	0.030132
	Andhra Prade	0.135556	0.266667	0.471111	0.116667	0.01	0.080204	0.007221	0.012553
	Tamil Nadu	0.015613	0.185272	0.608639	0.126724	0.063752	0.933646	0.044236	0.022118
	Telangana	0.137441	0.184044	0.333333	0.280411	0.064771	0.884676	0.044234	0.07109
	Kerala	0	0.046083	0.516129	0.382488	0.0553	0.926267	0.046083	0.02765
	Puducherry	0.706897	0.224138	0	0	0.068966	0.068966	0	0.931034
Eastern	Odisha	0.008088	0.192647	0.741912	0.057353	0	0.940441	0.038971	0.020588
	Bihar	0.250863	0.245109	0.357883	0.10702	0.039125	0.752589	0.16916	0.078251
	West Bengal	0.077778	0.17619	0.636508	0.098413	0.011111	0.698413	0.125397	0.17619
	Andaman & N	0.2	0.2	0.48	0.12	0	0.96	0.04	0
	Jharkhand	0	0.525424	0.220339	0.254237	0	1	0	0
Western	Maharashtra	0.035236	0.220472	0.546063	0.184252	0.013976	0.847835	0.135039	0.017126
	Gujarat	0.039613	0.429756	0.42561	0.090742	0.014279	0.966836	0.011515	0.021649
	Goa	0	0.344828	0.517241	0.103448	0.034483	0.931034	0.034483	0.034483
	Lakshadweep	0	0	1	0	0	1	0	0
	D&N Haveli ar	0	0.103448	0.896552	0	0	0.931034	0.068966	0
North-East	Arunachal Pra	0	0.060606	0.69697	0.242424	0	0.757576	0.242424	0
	Assam	0.161417	0.393701	0.385827	0.059055	0	0.889764	0.055118	0.055118
	Meghalaya	0.055556	0	0.638889	0	0	0.958333	0.041667	0
	Manipur	0.178571	0.285714	0.464286	0.071429	0	0.928571	0.071429	0
	Mizoram	0.318182	0.136364	0.545455	0	0	1	0	0
	Nagaland	0	0.181818	0.636364	0.181818	0	0.454545	0.272727	0.272727
	Tripura	0.034483	0.206897	0.62069	0.137931	0	0.62069	0.034483	0.344828
	Sikkim	0	0.153846	0.769231	0.076923	0	1	0	0
Central	Madhya Prade	0.084592	0.328907	0.448063	0.111152	0.027288	0.640349	0.212116	0.147535
	Chattisgarh	0.06228	0.325108	0.527223	0.079514	0.005875	0.844888	0.111242	0.04387
	Uttar Pradesh	0.098765	0.20096	0.394376	0.262003	0.043896	0.741427	0.142661	0.115912

This is the data-set for only 2020 and this is only a part of the original dataset. Like 2020, for the years 2019, 2018, 2017, 2016 the data is divided into six Zones and the proportion of the no. of crimes in each subcategory and the crime totals are taken.

GRAPHICAL REPRESENTATION OF THE DATA



we see that the total number of juvenile crimes are gradually decreasing.

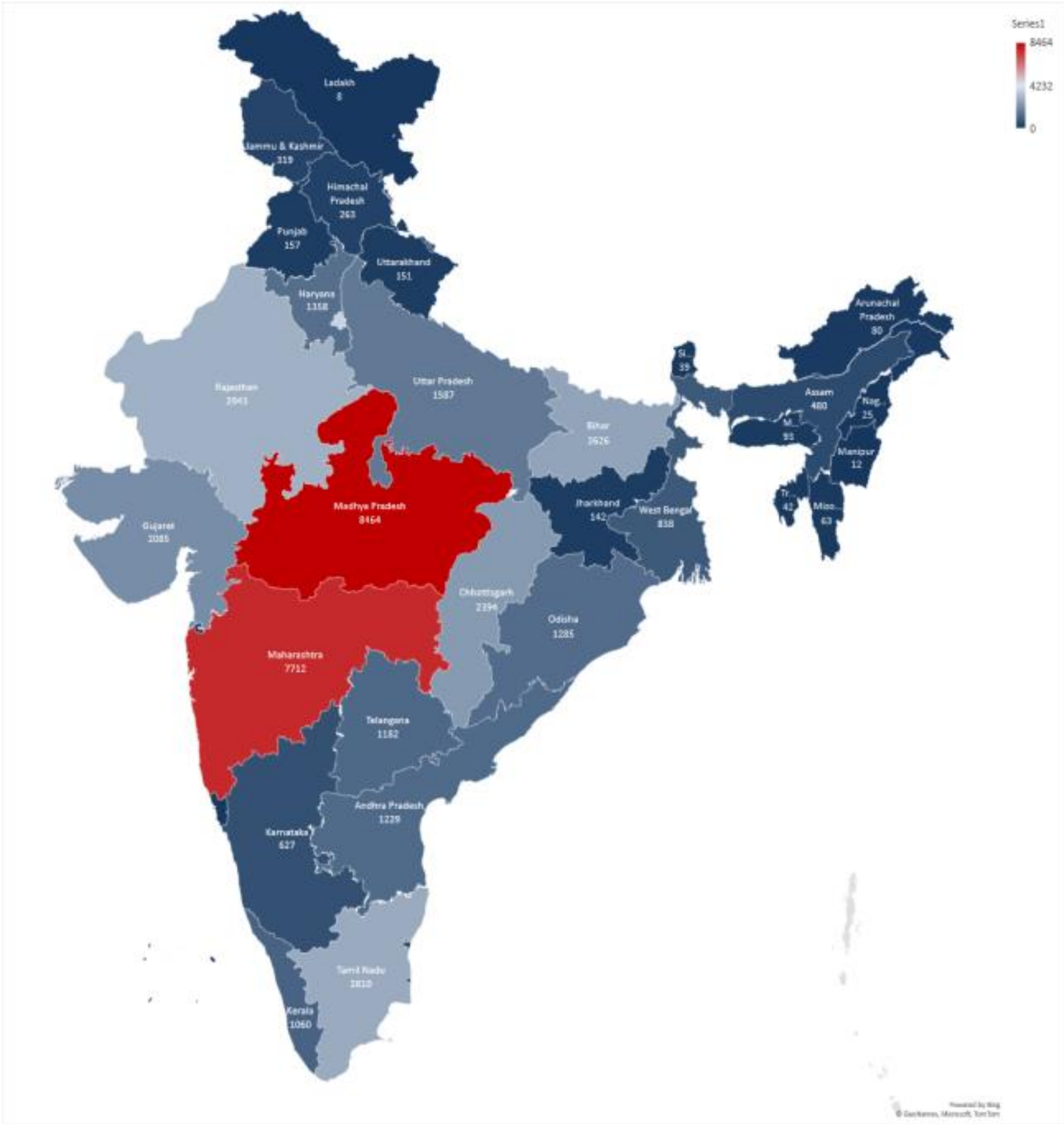


figure: Heatmap Showing State-wise Juvenile Crimes in 2016

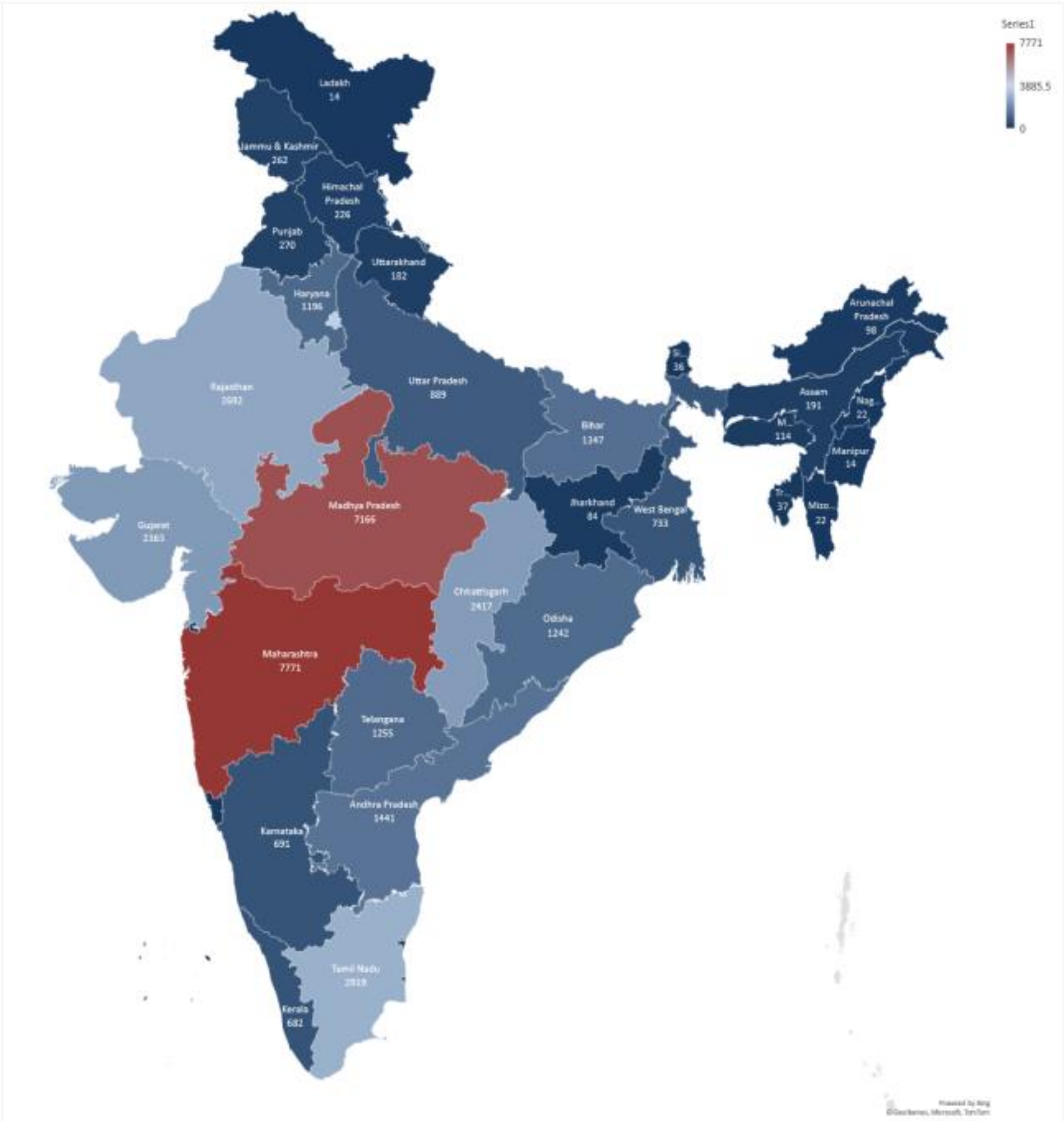


figure: Heatmap Showing State-wise Juvenile Crimes in 2017

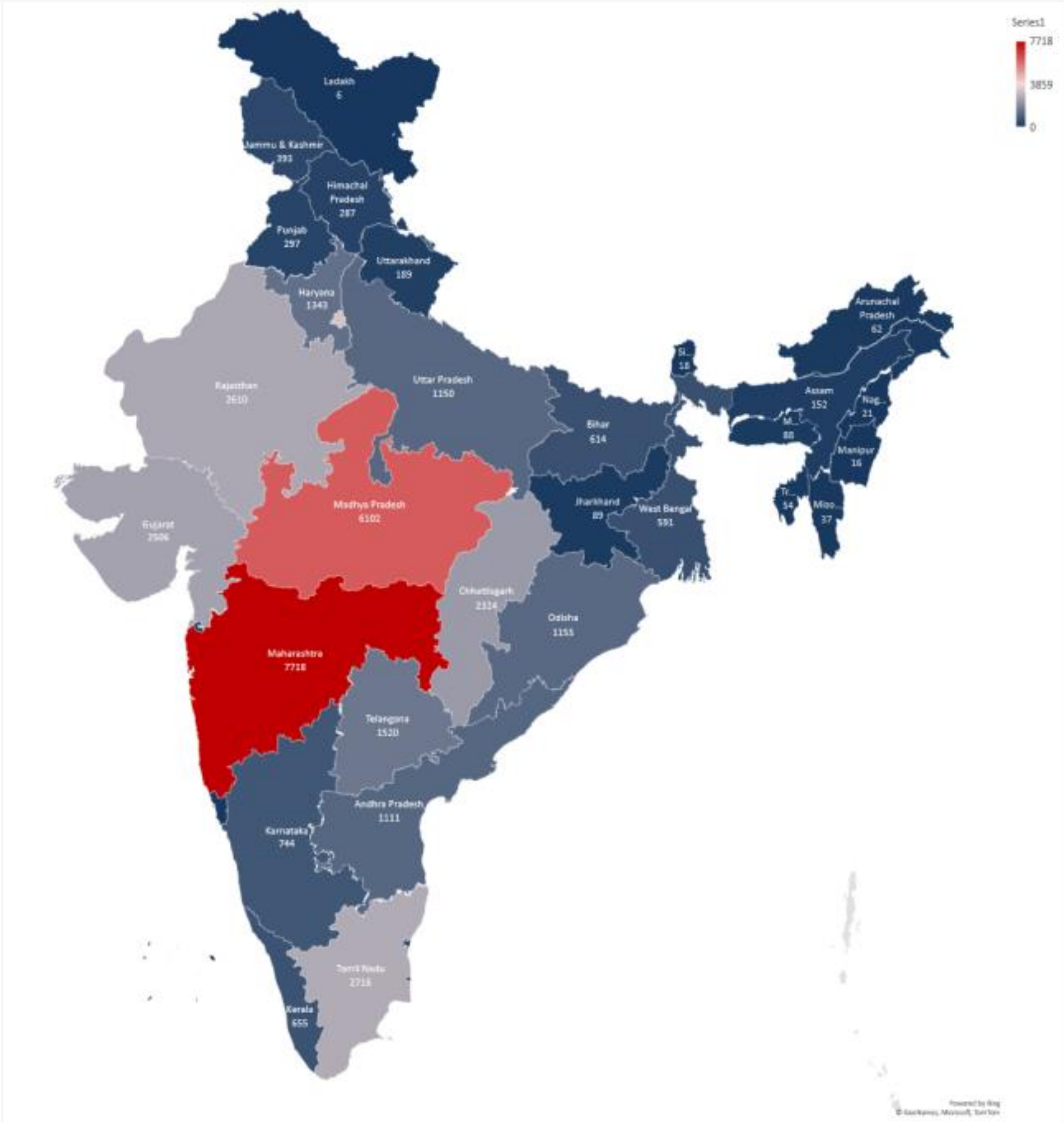


figure: Heatmap Showing State-wise Juvenile Crimes in 2018

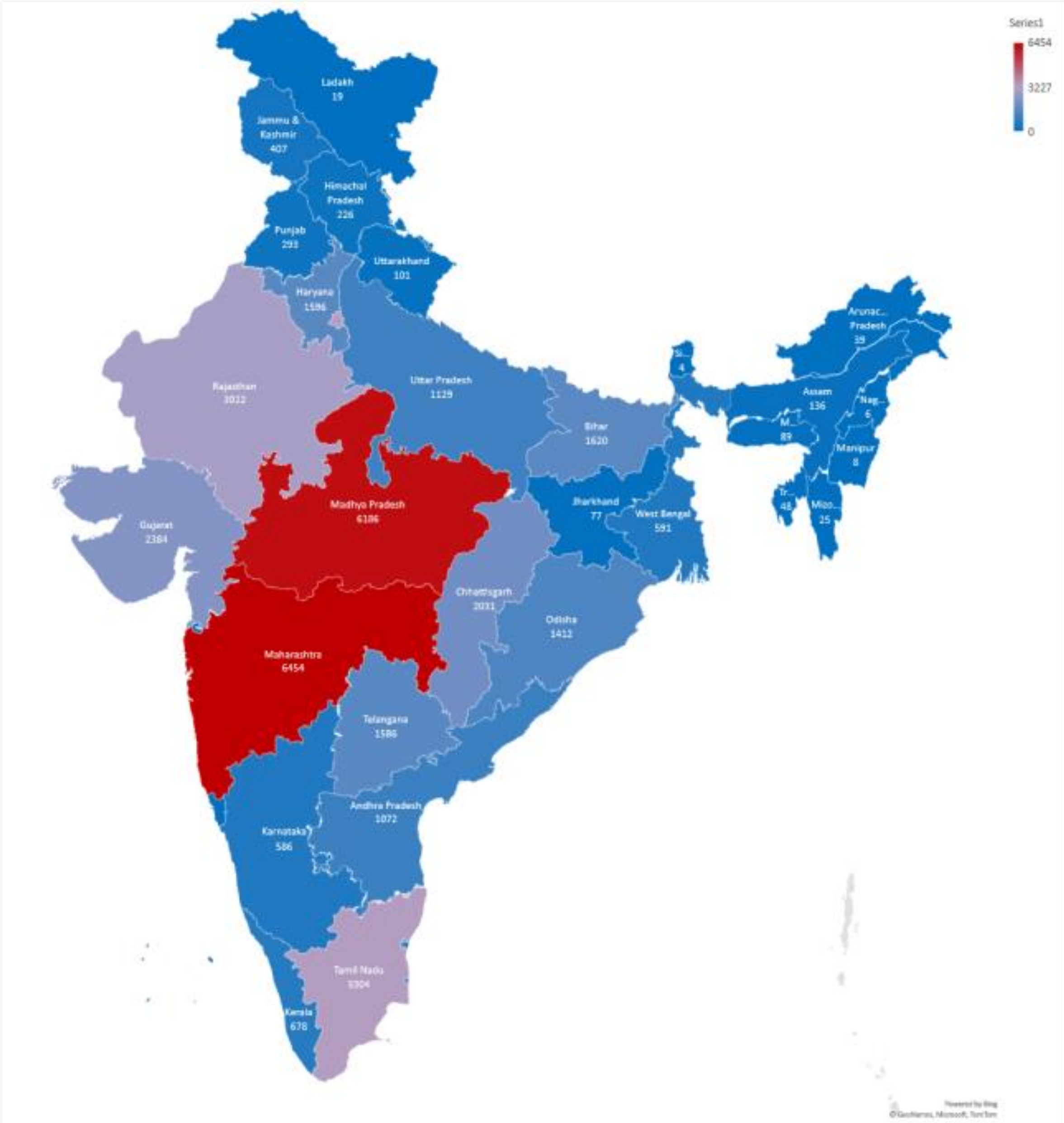


figure: Heatmap Showing State-wise Juvenile Crimes in 2019

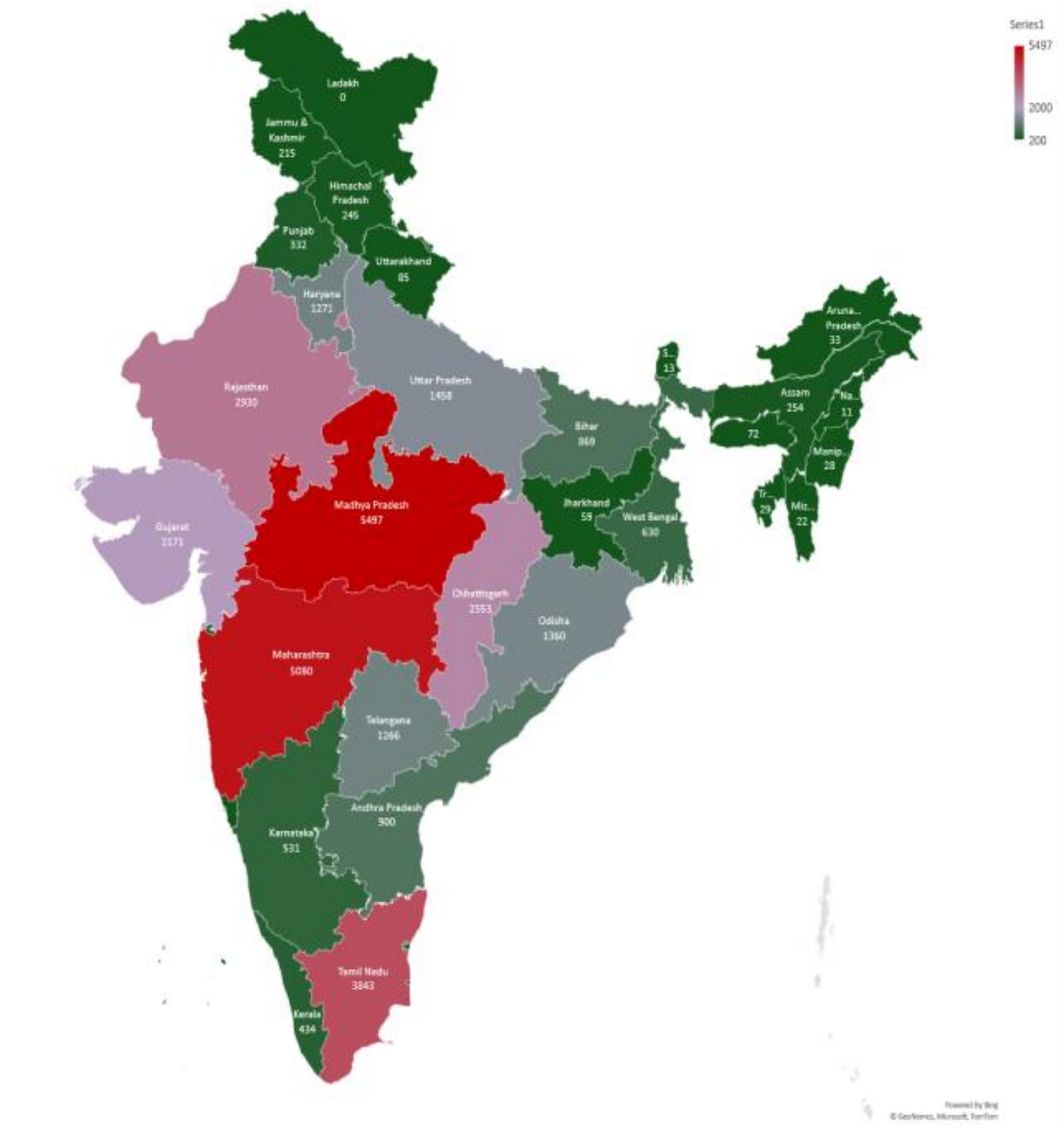
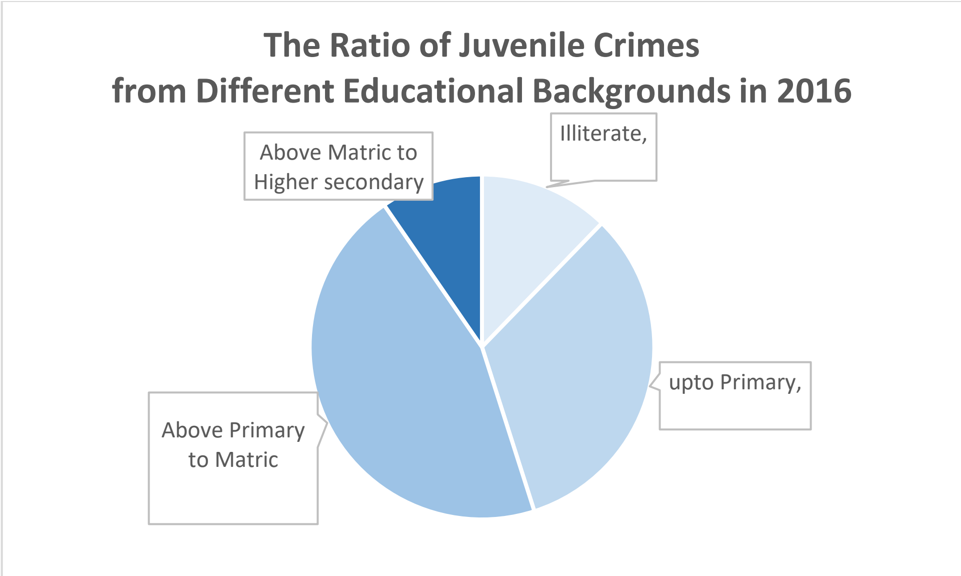
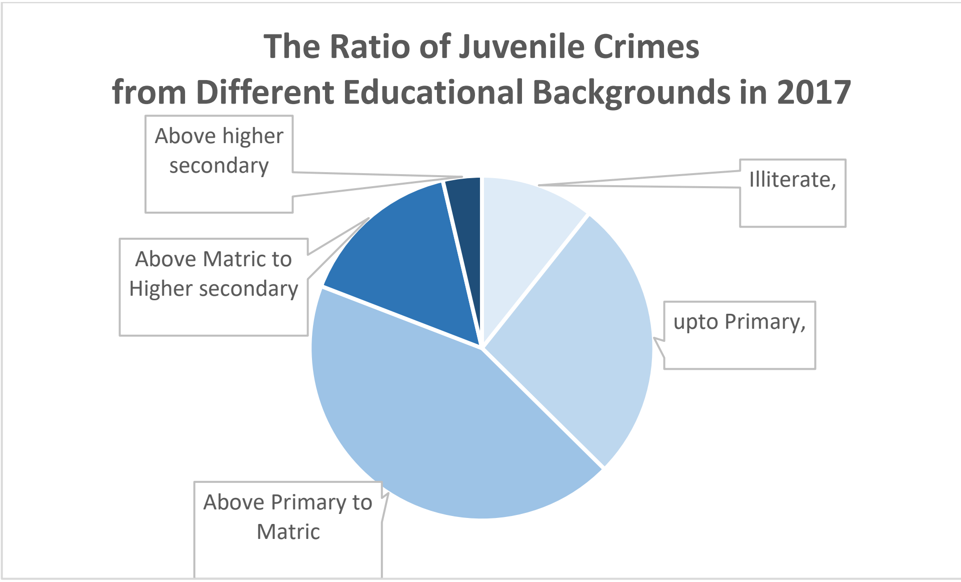
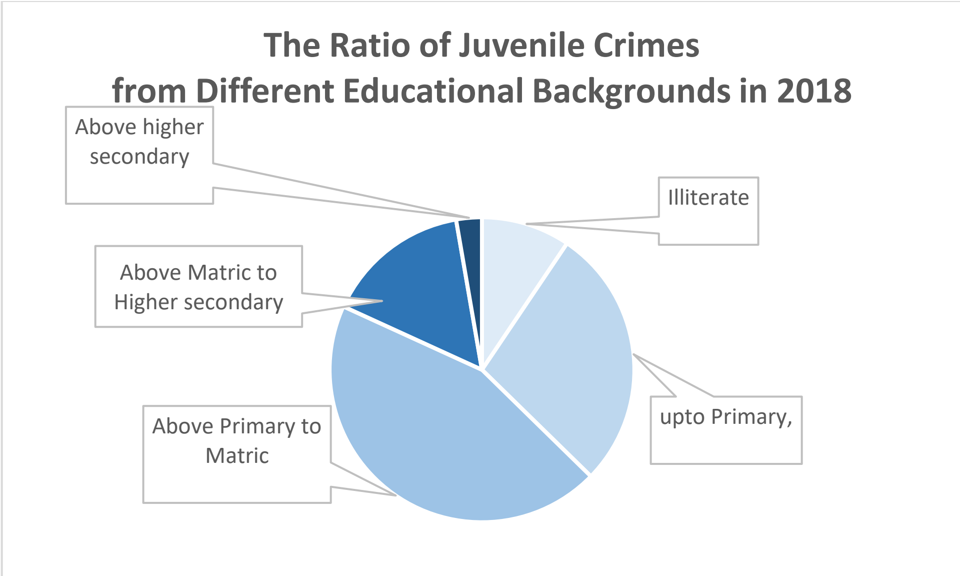
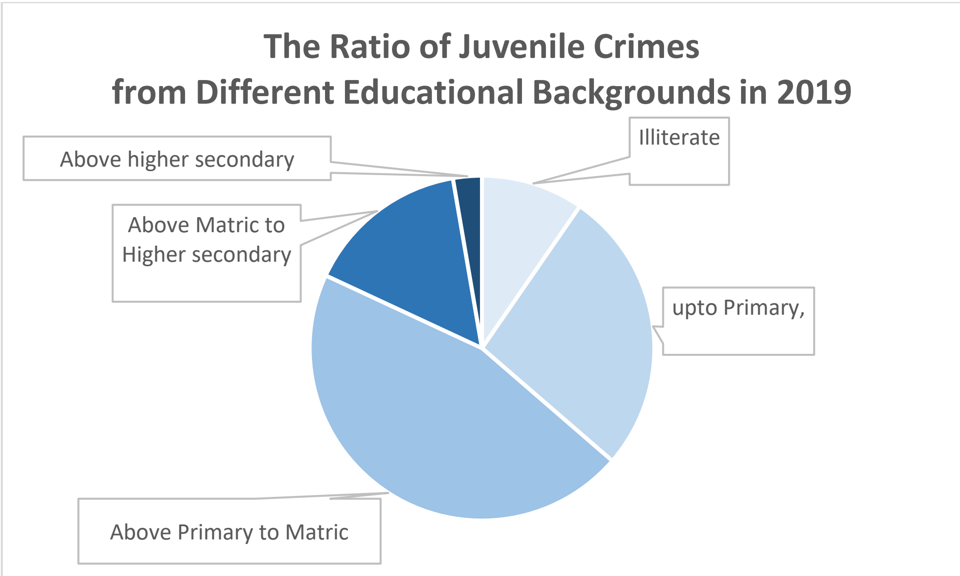


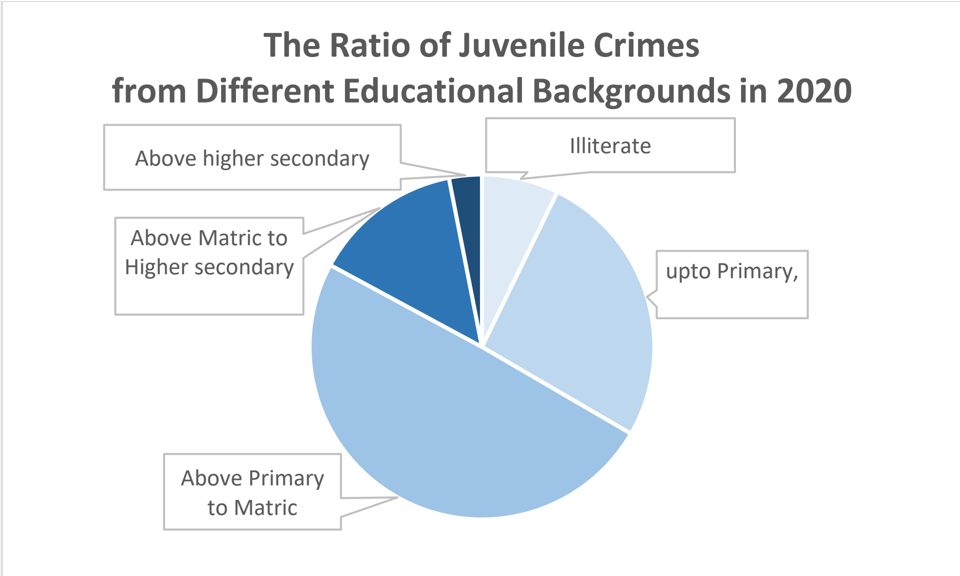
figure : Heatmap Showing State-wise Juvenile Crimes in 2020


The Ratio of Juvenile Crimes
from Different Educational Backgrounds in 2016
Above Matric to Higher secondary
Illiterate,
upto Primary,
Above Primary to Matric

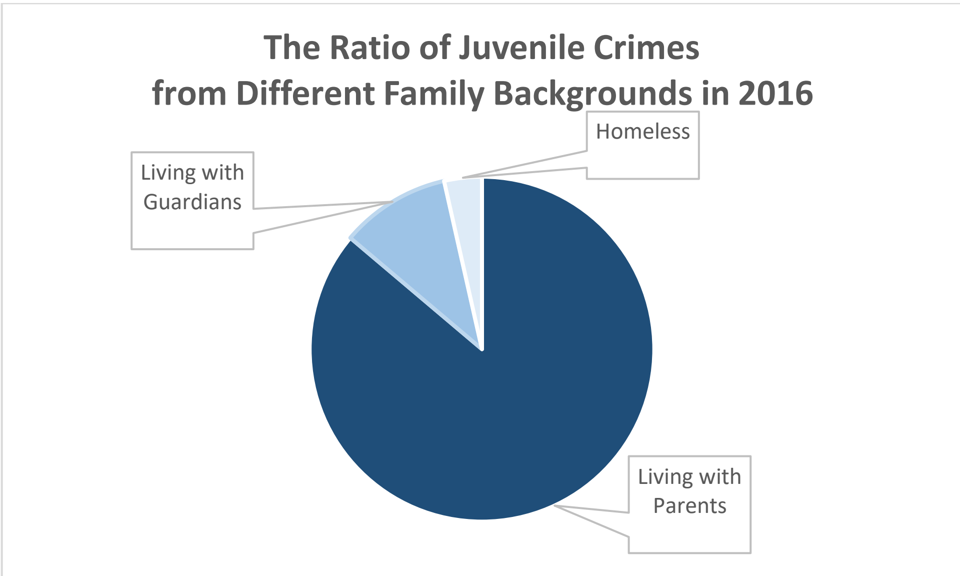

The Ratio of Juvenile Crimes
from Different Educational Backgrounds in 2017
Above higher secondary
Illiterate,
Above Matric to Higher secondary
upto Primary,
Above Primary to Matric

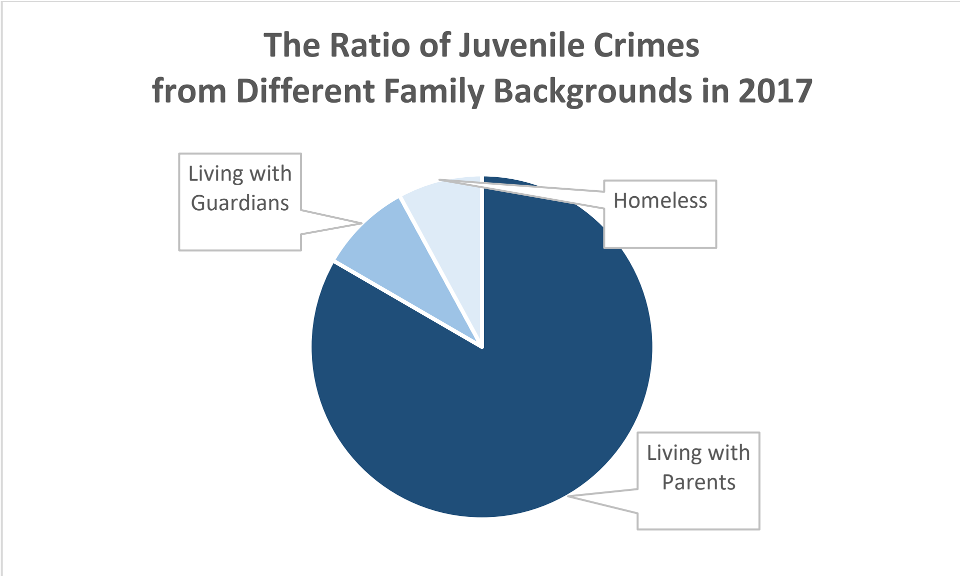

The Ratio of Juvenile Crimes
from Different Educational Backgrounds in 2018
Above higher secondary
Illiterate
Above Matric to Higher secondary
upto Primary,
Above Primary to Matric

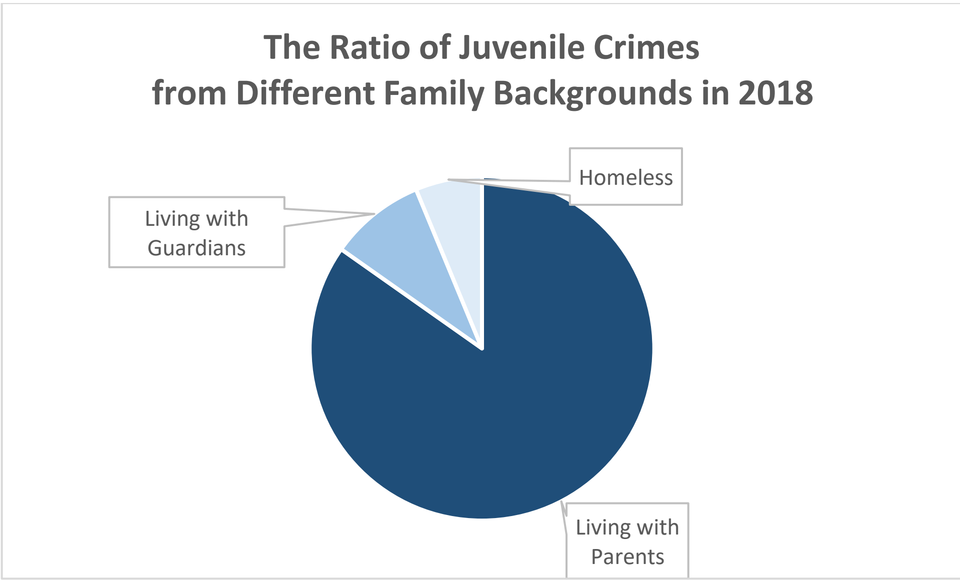

The Ratio of Juvenile Crimes
from Different Educational Backgrounds in 2019
Illiterate
Above higher secondary
Above Matric to Higher secondary
upto Primary,
Above Primary to Matric

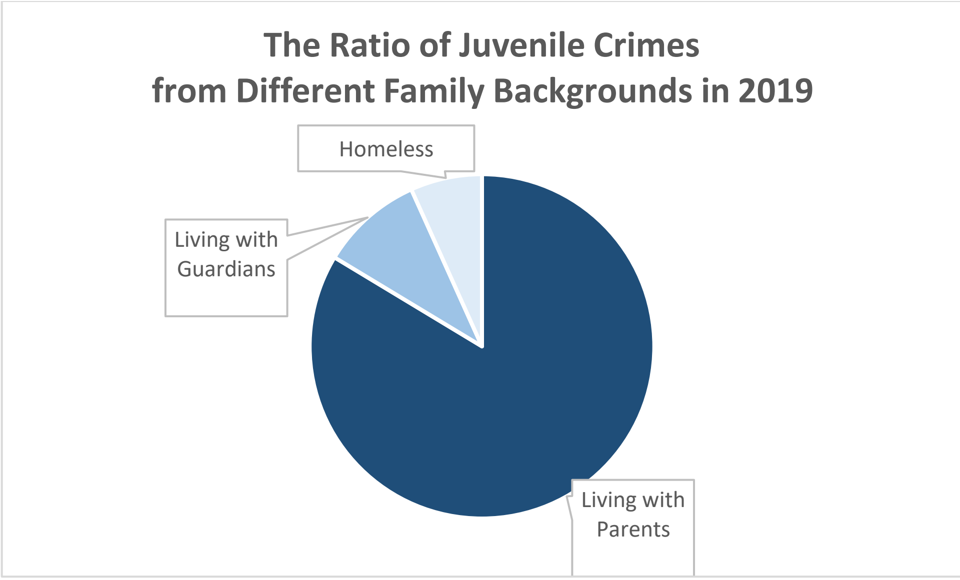


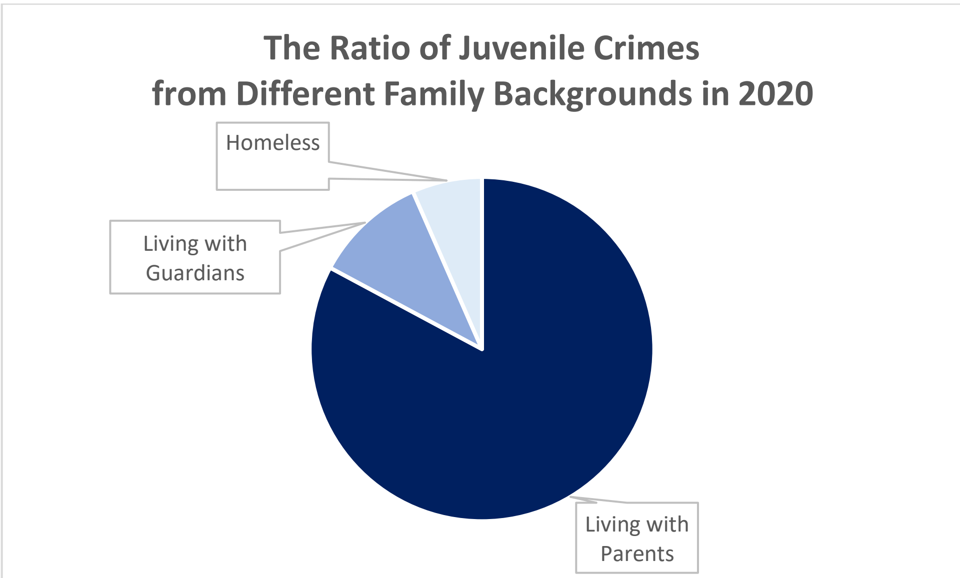
Interpretation: Most of the crimes are done by juveniles who are in intermediate stage between Primary and Matric. The number of crimes done by juveniles who have already passed High Secondary is very less.




The Ratio of Juvenile Crimes
from Different Family Backgrounds in 2017
Living with Guardians
Homeless
Living with Parents


The Ratio of Juvenile Crimes
from Different Family Backgrounds in 2018
Homeless
Living with Guardians
Living with Parents





<u>Interpretation</u>: There is a big difference between the different family backgrounds. The graphs show that the crime rate is high among the juveniles who are living with their parents.

Materials and Methods

We want to analyse the datasets to find out if there is any effect of various educational backgrounds and various family backgrounds. To fulfil this objective, we have to do two-way ANOVA with more than one and unequal observations per cell. So firstly, we will check the first assumption of ANOVA i.e., if the data is coming from Normal distribution or not through Normality Test (Shapiro-Wilk Test) and if the data is not coming from Normal distribution, we will perform an equivalent Non-parametric Test (Friedman Test).

SHAPIRO-WILK'S TEST

The Shapiro-Wilk's test is a way to tell if a random sample comes from a normal distribution. The test gives us a W value; small values indicate our sample is not normally distributed (we can reject the null hypothesis that our population is normally distributed if our values are under a certain threshold).

HYPOTHESIS: The null hypothesis is as follows:

H0: The data comes from a normal distribution

H1: The data does not come from a normal distribution

TEST STATISTIC:

The formula for the test statistic, W value is:

Where,

1. xi are the ordered random sample values

2. ai are constants generated from the covariances, variances and means of the sample (size n) from a normally distributed sample.

The test has limitations, most importantly that the test has a bias by sample size. The larger the sample, the more likely we'll get a statistically significant result.

H_0: A sample $x_1, x_2, \ldots x_n$ came from a normally distributed population.

The test statistics is w = $\frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

Where $x_{(i)}$ is the ith order statistics, $\bar{x}$ is the sample mean.

The coefficients a_i are given by $(a_1, a_2, \ldots, a_3) = \frac{m^T V^{-1}}{C}$, Where C is a vector norm.

C=||$V^{-1}m$|| = $(m^T V^{-1} m)^{\frac{1}{2}}$ and the vector m, $m = (m_1, m_2, \ldots, m_n)^T$ is made of the expected values of order statistics of independent and identically distributed random variables sampled from the standard normal distribution; finally, V is the covariance matrix of those normal order statistics.

ANOVA

The ANOVA (Analysis of variance) test compares the means of different samples and the significance of their differences. A two-way ANOVA test is a statistical analysis tool that determines the effect of two variables on an outcome, as well as testing how altering the variables will affect the outcome.

Assumptions of two-way ANOVA:

• Independence of variables:

The two variables for testing should be independent of each other. One should not affect the other, or else it could result in skewness. This means that one cannot use the two-way ANOVA test in settings with categorical variables.

• Homoscedasticity:

In a two-way ANOVA test, the variance should be homogenous. The variation around the mean for each set of data should not vary significantly for all the groups.

• Normal distribution of variables:

The two variables in a two-way ANOVA test should have a normal distribution. When plotted individually, each should have a bell curve. If the data does not meet this criterion, one could attempt statistical data transformation to achieve the desired result.

Let n_{ij} be the number of observations in the (i,j)th cell formed by the ith row classification(A) and jth column classification (B) for $i = 1(1)p;\ j = 1(1)q$. let y_{ijk} be the kth observation in the $(i,j)th$ cell, for $k = 1(1)n_{ij}$.

Then the model is:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk};$$
$$i = 1(1)p;\ j = 1(1)q, k = 1(1)n_{ij}.$$

μ=general effect

α_i=main effect due to factor A,

β_j=main effect due to factor B,

$(\alpha\beta)_{ij}$=interaction effect of A and B

e_{ijk}=random error

The errors e_{ijk} are assumed to be independently normally distributed with common mean zero and common variance σ^2

<u>HYPOTHESIS</u>: Null hypothesis and Alternative Hypothesis

The hypothesis of interest are about the main effects and interaction effects. The hypothesis of no interaction effects or of additivity is

$H_{AB} : \forall\ (\alpha\beta)_{ij} = 0$

The normal equations under H_{AB} are obtained by minimising $\sum_i \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j)^2$

With respect to μ, α_i and β_j and they are

$$\left.\begin{aligned} n\mu + \textstyle\sum_i \sum_j n_{ij}(\alpha_i + \beta_j) &= \textstyle\sum_i \sum_j \sum_k Y_{ijk}, \\ \textstyle\sum_i(\mu + \alpha_i) + \sum_j n_{ij}\beta_j &= \textstyle\sum_j \sum_k Y_{ijk}, \\ \textstyle\sum_j(\mu + \beta_j) + \sum_i n_{ij}\alpha_i &= \textstyle\sum_i \sum_k Y_{ijk}, \end{aligned}\right\} \star$$

Or,

$$\begin{rcases} n\mu + \sum_i n_{i0}\alpha_i + \sum_j n_{0j}\beta_j = \sum_i \sum_j \sum_k Y_{ijk} , \\ n_{i0}(\mu + \alpha_i) + \sum_j n_{ij} \ \beta_j = R_i , \\ n_{0j}(\mu + \beta_j) + \sum_i n_{ij} \ \alpha_i = C_j , \end{rcases} \star\star$$

Where $\sum_j n_{ij} = n_{i0}, \ \sum_i n_{ij} = n_{0j}, \ R_i = \sum_j \sum_k Y_{ijk}, \quad C_j = \sum_i \sum_k Y_{ijk}$

We obtain, $\beta_j = n_{0j}^{-1}(C_j - \sum_{i'} n_{(i'j)} \alpha_{i'} - \mu,$ ★★★

Substituting this value, we get

$$n_{i0}\alpha_i + \sum_i n_{ij} n_{0j}^{-1}(C_j - \sum_{i'} n_{i'j} \ \alpha_{i'}) = R_i$$

$$(n_{i0} - \sum_j n_{ij} n_{ij} n_{0j}^{-1})\alpha_i + \sum_j [n_{ij} n_{0j}^{-1} C_j - \sum_{i', i \neq i'} n_{ij} n_{0j}^{-1} n_{i'j} \ \alpha_{i'}] = R_i$$

Or, $(n_{i0} - \sum_j n_{ij} p_{ij})\alpha_i + \sum_j \sum_{i', i' \neq i} n_{ij} p_{i'j} \alpha_{i'} = R_i - \sum_j p_{ij} C_j, \qquad \forall \ i = 1(1)p,$ ★★★★

where $p_{ij} = \frac{n_{ij}}{n_{0j}}$.

Similarly, eliminating the α's from the normal equations in (★★), we can get the following q equations for determining the β's:

$(n_{i0} - \sum_i n_{ij} q_{ij})\beta_i - \sum_i \sum_{j', j \neq j'} n_{ij} q_{ij'} \beta_{j'} = C_i - \sum_i q_{ij} R_i$, $\forall \ j = 1(1)q$, Where $q_{ij} = \frac{n_{ij}}{n_{i0}}$

★★★★★

In other cases of the two-way model, we assumed $\sum_i$ αi = $\sum_j$ βj = 0. In solving the equations, it is simpler to assume for the present problem that $\alpha_p = 0, \beta_q = 0$. This will give exactly the same results for any comparison among the α's(β's).

Then ★★★★ becomes a set of (p-1) equations in $\alpha_1, \alpha_2, \dots, \alpha_{p-1}$ and ★★★★★

$\beta_1, \beta_2, \dots, \beta_{q-1}$.

The residual SS under H_{AB} is :

$S_1^2 = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \sum_k Y_{ijk} \hat{\mu} - \sum_i \sum_j \sum_k Y_{ijk} \widehat{\alpha_i} - \sum_i \sum_j \sum_k Y_{ijk} \widehat{\beta_j},$

Using ★★★ if We eliminate the $\hat{\beta}$'s, the SS reduces to,

$S_1^2 = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_j (R_i - \sum_j p_{ij} C_j)\, \widehat{\alpha_\iota} - \sum_j ({C_j^2}/{n_{0j}})$, ★★★★★★ and to

$S_1^2 = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i (C_j - \sum_j p_{ij} R_i)\, \widehat{\beta_\iota} - \sum_i ({R_i^2}/{n_{i0}})$ (★★★★★★ a)

If we eliminate the $\widehat{\alpha_\iota}$'s instead of the $\hat{\beta}$'s. S_1^2 has d.f.=(n − p − q + 1).

To obtain S_1^2 we solve (p−1) linear equations in the α's or (q −1) linear equations in the β's .

Thus the numerator SS for testing H_{AB} is

$SS(AB) = S_1^2 - SSE$, with d.f. (p − 1)(q − 1).

The test statistic for testing H_{AB} is:$\frac{SS(AB)}{SSE} \frac{n-pq}{(p-1)(q-1)}$

And under H_{AB} it has the central F-distribution with (p-1)(q-1)(n-pq) degrees of freedom.

It is easy to see that,

$$SS(AB) = \frac{\sum_i \sum_j (\sum_k Y_{ijk})^2}{n_{ij}} - \sum_i (R_i - \sum_j p_{ij} C_j)\, \widehat{\alpha_\iota} - \sum_j ({C_j^2}/{n_{0j}})$$

$$SS(AB) = \frac{\sum_i \sum_j (\sum_k Y_{ijk})^2}{n_{ij}} - \sum_j (C_j - \sum_i q_{ij} R_i)\, \widehat{\beta_J} - \sum_i ({R_i^2}/{n_{i0}})$$

In the unequal cell frequency case, the SS due to the main effects 'rows' and 'columns' cannot be added, for they are non-orthogonal. Thus the SS due to row effects(main effects of A) would be calculated by calculating the reduction in residual SS brought about by first eliminating all the parameters except α's and then estimating all including α's. This will provide the SS due to the main effects of A. Similarly, the SS due to B may be computed. These computations are simple if we assume additivity.

If there is interaction, we perform a one-way ANOVA for the main effects of A(at each level of B) and similarly for the main effect of B(at each level of A).

Here we consider the simple case of testing H_A: $\alpha_i = 0$ for all i and H_B : $\beta_j = 0$ ∀ j under additivity.

So now the valid error for the above two tests will be the S_1^2 obtained for performing the test of H_{AB}.

Under H_{AB} and under H_A, The model is:

$Y_{ijk} = \mu + \beta_j + e_{ijk}$, e_{ijk} are independently $N(0, \sigma^2)$

So, the residual SS under H_A (assuming H_{AB} is the SSE for a one-way classification) is:

$$\sum_i \sum_j \sum_k \left(Y_{ijk} - \frac{C_j}{n_{0j}}\right)^2$$

$$\sum_i \sum_j \sum_k \left(Y_{ijk}^2 - \frac{\sum_j \ C_j^2}{n_{0j}}\right)$$

With d.f. (n-q).

Thus,

$SSA^* \ = S_2^2 - S_1^2 = \sum_i (R_i - \sum_j p_{ij} \, C_j)\widehat{\alpha_i}$, with d.f. (p-1).

This is called the adjusted SS of A.

Under additivity, The test statistic for H_A is

$$\frac{SSA^*}{S_1^2} \cdot \frac{n-p-q+1}{p-1}$$

Which is also a central F under H_B with d.f. (q-1,n-p-q+1)

where $SSB^* = \sum_j (C_j - \sum_i q_{ij} R_i)\widehat{\beta_j}$

Source	d.f.	SS	SS	Source
Rows(A), Adjusted Columns(B), Unadjusted Interactions	(p-1) (q-1) (p-1)(q-1)	SSA* SSB SS(AB) ⟷	SSA SSB* SS(AB)	Rows(A) UnadjustedColumn(B) interactions
Between cells Within cells	pq-1 n-pq	SSBe ⟷ SSE ⟷	SSBe SSE	Between cells Within cells
Total	n-1	$\sum_i \sum_j \sum_k (Y_{ijk} - Y_{000})$^2		Total

⟷ Denote that the two end quantities are equal.

In fact,

SSA= The adjusted SS due to A = $\frac{\sum_i R_i^2}{n_{io}}$ - $\frac{(\sum_i \sum_j \sum_k Y_{ijk})\text{^}2}{n}$,

SSB= The adjusted SS due to B = $\frac{\sum_j C_j^2}{n_{oj}} - \frac{(\sum_i \sum_j \sum_k Y_{ijk})\text{^}2}{n}$,

SSBe= The between cells SS = $\frac{\sum_i \sum_j (\sum_k Y_{ijk})^2}{n_{ij}} - \frac{(\sum_i \sum_j \sum_k Y_{ijk})^2}{n}$,

and also, SSBe= the between cells SS= SSA* + SSB + SS(AB) = SSA + SSB* + SS(AB).

Friedman Test.

Friedman Test: A Non-parametric Test.

It is a non-parametric test alternative to the one-way ANOVA with repeated measures. It tries to determine if subjects changed significantly across occasions/conditions. It is used to test for differences between groups when the dependent variable is ordinal. This test is particularly useful when the sample size is very small.

Assumptions :

– The group is a random sample from the population.

– Samples are not normally distributed.

Hypothesis:

Null Hypothesis: There is no significant difference between the given conditions of measurement OR the probability distributions for all the conditions are the same. (Medians are same)

i.e., $H_0: M_1 = M_2 = M_3 = \ldots\ldots = M_k$; M= Median

Alternate Hypothesis, H_1: At least two of them show significant difference.

Test Statistic:

$$\frac{\mathbf{12}}{\boldsymbol{nk}(\boldsymbol{k}+\mathbf{1})} \cdot \sum_{\boldsymbol{i}} \boldsymbol{R}_{\boldsymbol{i}}^{\mathbf{2}} - \mathbf{3}\boldsymbol{n}(\boldsymbol{k}+\mathbf{1})$$

where,

n= total number of subjects/participants.

k = total number of blocks to be measured.

R_i= Ri = sum of ranks of all subjects for a block i

Decision Rule for Friedman Test:

We can make the decision on the basis of the below-mentioned rules-

1. Calculated Value vs Table Value: If FR is greater than the critical value limits reject the Null Hypothesis. Otherwise, accept the Null Hypothesis.

2. P-Value Approach: Compare the P-Value with (Level of Significance). If the p-value is less than or equal to α then reject the Null Hypothesis.

ANALYSIS

Shapiro-Wilk Test:

At first we test the normality of the data through Shapiro-Wilk Normality Test. Here,

Null Hypotheses, H_0: The data is normal (for all years).

Alternative Hypotheses, H_1: The data is not normal (for all years).

The results from Shapiro-Wilk Test are as follows:

	Educational Backgrounds		Family Backgrounds	
Year	Test Statistic	p-value	Test Statistic	p-value
2020	0.847301	3.019802×10^{-12}	0.749304	4.231827×10^{-12}
2019	0.856948	1.346829×10^{-11}	0.753740	1.012046×10^{-11}
2018	0.866408	3.851449×10^{-11}	0.747939	7.095002×10^{-12}
2017	0.8566418	1.302856×10^{-11}	0.767496	2.076415×10^{-11}
2016	0.916055	3.684827×10^{-07}	0.760003	1.288107×10^{-11}

Here, the p-values for all years are very low to accept the null hypotheses,H_0. So we reject the null hypothesis and conclude that the data does not follow normal distribution. To meet our objectives we have to go for Non-parametric Test, the Friedman Test.

However, we performed two way ANOVA as well to justify the synergy of the results of both the test procedures.

Friedman Test:

For Friedman Test, Null Hypotheses 1, H_{10}: Educational Backgrounds do not differ significantly (for all years). Null Hypotheses 2, H_{20}: Family Backgrounds do not differ significantly (for all years). Null hypotheses 3, H_{30}: The effect of Zones does not differ significantly (for all years). The results from Friedman Test are as follows:

	Educational Backgrounds		Zones	
Year	Test Statistic	p-value	Test Statistic	p-value
2020	99.91137	1.027338×10^{-20}	627.24566	2.62050×10^{-133}
2019	89.73858	1.496361×10^{-18}	602.04863	7.29844×10^{-128}
2018	102.76398	2.536695×10^{-21}	602.10781	6.23868×10^{-128}
2017	87.30189	4.925687×10^{-18}	603.40683	3.71332×10^{-128}
2016	66.13846	2.862901×10^{-14}	500.17241	7.32894×10^{-106}

	Family Backgrounds		Zones	
Year	Test Statistic	p-value	Test Statistic	p-value
2020	53.96992	1.908006×10^{-12}	362.64177	3.31867×10^{-76}
2019	45.1875	1.540490×10^{-10}	342.24172	8.18981×10^{-72}
2018	46.24806	9.064857×10^{-11}	346.40813	1.03845×10^{-72}
2017	45.68992	1.198283×10^{-10}	350.84758	1.1498×10^{-73}
2016	57.1811	3.830729×10^{-13}	337.01149	1.09403×10^{-70}

Here, in Friedman Test, for Educational Backgrounds, Family Backgrounds and Zones the p-values for all years are very small to accept the Null hypotheses, H_{10}, H_{20} and H_{30}. So we successfully reject the null hypotheses and conclude that the effects Educational Backgrounds and Family Backgrounds differ significantly and the effects of Zones are also signific

ANOVA:

The Null Hypotheses and Alternative Hypotheses For Educational Backgrounds:

Null Hypotheses 1,

H_{10}: The effect of Zones does not differ significantly(for all years).

Alternative Hypotheses 1,

H_{11}: The effect of Zones does differ significantly(for all years).

Null Hypotheses 2,

H_{20}: Educational Backgrounds does not differ significantly (for all years).

Alternative Hypotheses 2,

H_{21}: Educational Backgrounds does differ significantly(for all years).

Null Hypotheses 3,

H_{30}: There is no interaction effects between Zones and Educational Backgrounds(for all years).

Alternative Hypotheses 3,

H_{31}: There is some interaction between Zones and Educational Backgrounds(for all years).
The results from ANOVA are as follows:

Year	Source	DF	Sum of Squares	Mean Sum of Squares	F	pvalue
2020	Zones	5	3.659528×10^{-19}	7.319055×10^{-20}	5.057322×10^{-18}	1.000000
	Education	4	5.470106	1.367526	94.49336	2.172384×10^{-39}
	Interaction	20	0.3925526	0.01962763	1.356230	0.1539916
	Residual	145	2.098468	0.01447220	–	–
2019	Zones	5	0.044118	0.008824	0.474380	0.7948808
	Education	4	4.307075	1.076769	57.890336	9.471138×10^{-29}
	Interaction	20	0.315623	0.015781	0.848441	0.6514266
	Residual	140	2.604020	0.018600	–	–
2018	Zones	5	0.050431	0.010086	0.561553	0.8747808
	Education	4	4.097474	1.024369	80.575380	2.406768×10^{-35}
	Interaction	20	0.571904	0.028595	2.249257	3.281808×10^{-3}
	Residual	140	1.779844	0.012713	–	–
2017	Zones	5	0.044118	0.008824	0.4355404380	0.8231489
	Education	4	3.868114	0.967029	47.733686	2.909778×10^{-25}
	Interaction	20	0.327888	0.016394	0.809248	0.6991546
	Residual	140	2.836236	0.020259	–	–
2016	Zones	5	0.055147	0.011029	0.567148	0.7249952
	Education	3	3.387799	1.129266	58.068460	1.01465×10^{-22}
	Interaction	15	0.243037	0.016202	0.833153	0.6392224
	Residual	112	2.178082	0.019447	–	–

Here, the p-values for Zones for all years and interaction for all years except 2018 are greater than 0.05 so we fail to reject the null hypotheses for Zones, H10 and for interaction effects, H30. The p-values for Educational Backgrounds for all years are very small so we can reject the null hypothesis, H20. We conclude that there is no significant difference between the Zones and there is no interaction effects for the years 2016, 2017, 2019, 2020 but for the year 2018 there is some interaction effects between the Zones and the Educational Backgrounds and the Educational Backgrounds for all years differ significantly.

The Null Hypotheses and Alternative Hypotheses For Family Backgrounds:

Null Hypotheses 1

, H_{40}: The effect of Zones does not differ significantly(for all years).

Alternative Hypotheses 1,

H_{41}: The effect of Zones does differ significantly(for all years).

Null Hypotheses 2,

H_{50}: Family Backgrounds does not differ significantly (for all years).

Alternative Hypotheses 2,

H_{51}: Family Backgrounds does differ significantly(for all years).

Null Hypotheses 3,

H_{60}: There is no interaction effects between Zones and Family Backgrounds(for all years).

Alternative Hypotheses 3,

H_{61}: There is some interaction between Zones and Family Backgrounds(for all years).

The results from ANOVA are as follows:

Year	Source	DF	Sum of Squares	Mean Sum of Squares	F	pvalue
2020	Zones	5	6.305555×10^{-19}	1.261111×10^{-19}	4.957854×10^{-18}	1
	Family	2	12.97896	6.489482	255.1235	4.040083×10^{-37}
	Interaction	10	0.2691543	0.02691543	1.058136	0.4031596
	Residual	87	2.212987	0.02543663	–	–
2019	Zones	5	0.037099	0.007420	0.248149	0.9396294
	Family	2	11.547134	5.773567	193.090855	6.040686×10^{-32}
	Interaction	10	0.280096	0.028010	0.936750	0.5043936
	Residual	83	2.481765	0.029901	–	–
2018	Zones	5	0.036306	0.007261	0.385930	0.8571476
	Family	2	12.750708	6.375354	338.848194	1.17888×10^{-40}
	Interaction	10	0.172914	0.017291	0.919035	0.5200211
	Residual	83	1.561627	0.018815	–	–
2017	Zones	5	0.084021	0.016804	0.515839	0.7635924
	Family	2	10.076604	5.038302	154.660804	6.923443×10^{-29}
	Interaction	10	0.405548	0.040555	1.244911	0.2752188
	Residual	84	2.736423	0.032576	–	–
2016	Zones	5	0.079294	0.015859	0.906498	0.48099035
	Family	2	12.671254	6.335627	362.146655	5.033066×10^{-42}
	Interaction	10	0.124487	0.012449	0.711571	0.7112426
	Residual	84	1.469550	0.017495	–	–

Here, the p-values for Zones for all years and interaction for all years are greater than 0.05 so we fail to reject the null hypotheses for Zones, H_{40} and for interaction, H_{60}. The p-values for Educational Backgrounds for all years are very small so we can reject the null hypothesis, H_{50}. We conclude that there is no significant difference between the Zones and there is no interaction effects for all years but the Educational Backgrounds for all years differ significantly.

Conclusion

We can see from both the Parametric test and the Non-parametric test that the effect of Educational Background and the Family Background is significant. However, the effect of Zones is not significant in ANOVA but it is significant in Friedman Test.

As the data was not normal we will take the result we got from Friedman Test. Since the Educational Backgrounds, Family Backgrounds and society play a significant role in juvenile crimes, we can say that if a child gets proper education and proper care from his/her parents and society the rate of juvenile crimes will be lesser. The world will be a better place.

35 Nowadays people are working on juvenile crimes. As a result, we get fewer crimes committed by juveniles. In 2016 total juvenile crime was 44171 in India and in 2020 it was 35352.

Children are just like flowers, which need to be given a chance to blossom. Parents, guardians, and society as a whole have the duty to handle them with care. Alongside, providing ample opportunity for growth in a healthy environment would make them responsible citizens of the country.

References

1. https://blog.ipleaders.in/juvenile-crimes-india/-Juvenile Crimes in India

2. https://www.mapsofindia.com/my-india/society/juvenile-crimeon-the-rise-handle-children-with-care-Juvenile Crime on the Rise, Handle Children with Care

3. https://internationaljournalofresearch.com/2020/08/02/causes-of-juvenile-crimes/-Causes of juvenile crimes

4. https://timesofindia.indiatimes.com/city/indore/Lack-of-education/causes-delinquency/articleshow/55646736.cms-Times of India

5. An Outline of Statistical Theory- Gun, Gupta, Dasgupta

Appendix

Codes for Shapiro-Wilk Test of the data:

```python
from scipy.stats import shapiro
import numpy as np
import pandas as pd

mydf4=pd.read_csv("C:/Users/Lenovo/Documents/JuvenileCrime_Projectwork/2020 ed final.csv",header=0)
mydf.head()
shapiro(mydf4["Data"])
mydf4=pd.read_csv("C:/Users/Lenovo/Documents/JuvenileCrime_Projectwork/2020 fam final.csv",header=0)
mydf.head()
shapiro(mydf4["Data"])
```

Codes for Friedman Test:

```python
from scipy import stats
import pandas as pd

mydf=pd.read_csv("C:/Users/Lenovo/Documents/JuvenileCrime_Projectwork/2020 fri.csv",header=0)
mydf.head()
stats.friedmanchisquare(mydf["Illeterate"],mydf["Upto Primary"],mydf["Above Primary to Matric"],mydf["Above Matric to High Secondary"],mydf["Above High Secondary"])
stats.friedmanchisquare(mydf["Living with Parents"],mydf["Living with Guardians"],mydf["Homeless"])
```

Codes for ANOVA:

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

mydf=pd.read_csv("C:/Users/Lenovo/Documents/JuvenileCrime_Projectwork/2020 ed final.csv",header=0)
mydf.head()
model = ols('Data ~ C(Zones) + C(Education) + C(Zones):C(Education)',data=mydf).fit()
sm.stats.anova_lm(model, type=2)
mydf1=pd.read_csv("C:/Users/Lenovo/Documents/JuvenileCrime_Projectwork/2020 fam final.csv",header=0)
mydf1.head()
model = ols('Data ~ C(Zones) + C(Parents) + C(Zones):C(Parents)',data=mydf1).fit()
sm.stats.anova_lm(model, type=2)
```