

DATA SCIENCE

24 February 2022 , Arun P M 2021DDS6005

Data processing & visualization for the Employment rate depending on elementary and secondary education.

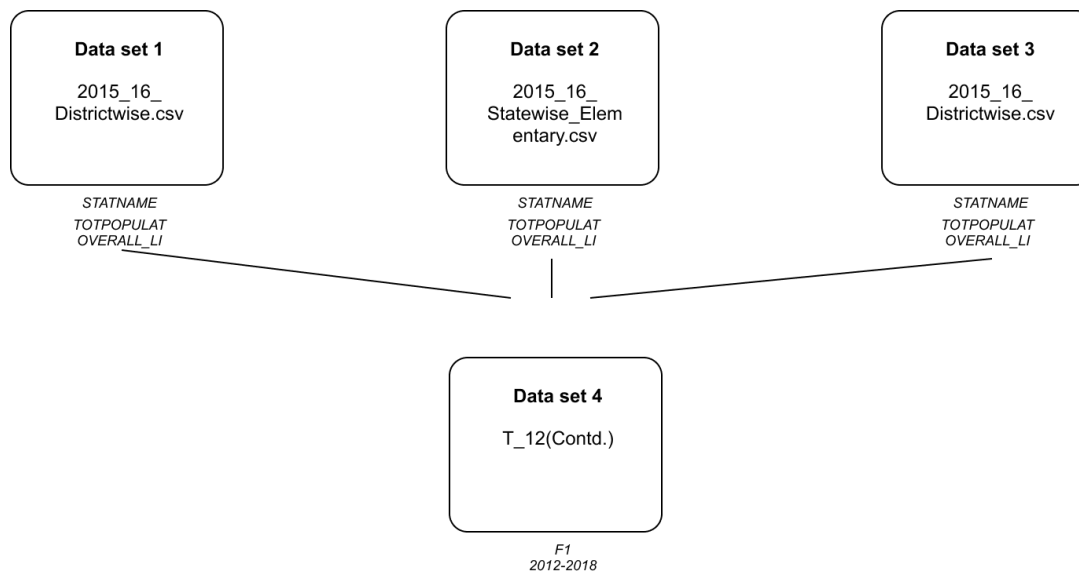
SOFTWARES USED

- **MATLAB is a programming and numeric computing platform used by millions of engineers and scientists to analyze data, develop algorithms, and create models.**
- **Tableau can help anyone see and understand their data. Connect to almost any database, drag and drop to create visualizations.**

OBJECTIVES

For this assignment, I closely wanted to look at how unemployment can be related to education in India. To compare these data sets of district wise education data and unemployment data.

ROUGH SCHEMA



SELECTION OF DATASET

The dataset was sourced from “www.kaggle.com” about the “**employment according to Indian Education**”. It contains 3 csv CSV files which could be used for data processing as well as visualization.

The screenshot shows the Tableau interface with a data source named '2015_16_Statewise_Secondary'. The fields list includes: ac_year, statcd, statname, area_sqkm, tot_population, urban_population, growth_rate, sexratio, sc_population, st_population, literacy_rate, male_literacy_rate, and female_literacy_rate. The preview table shows data for 15 states in 2015-16.

ac_year	statcd	statname	area_sqkm	tot_population	urban_population	growth_rate	sexratio	sc_population	st_population	literacy_rate	male_literacy_rate	female_literacy_rate
2015-16	1	Jammu And Kashmir	222,236	12,548,93	20,050	23,710	883	7,400	11,900	68,740	78,260	58,010
2015-16	2	Himachal Pradesh	55,673	6,856,51	8,690	12,810	974	25,200	5,700	83,780	90,830	76,600
2015-16	3	Punjab	50,362	27,704,24	29,820	13,730	893	31,900	0,000	76,680	81,480	71,340
2015-16	4	Chandigarh	114	1,054,69	76,660	17,100	818	18,900	0,000	86,430	90,540	81,380
2015-16	5	Uttarakhand	53,483	10,116,75	21,540	19,170	963	18,800	2,900	79,630	88,330	70,700
2015-16	6	Haryana	44,212	25,353,08	24,120	19,900	877	20,200	0,000	76,640	85,380	66,770
2015-16	7	Delhi	1,483	16,753,24	77,030	20,960	866	16,800	0,000	86,340	91,030	80,930
2015-16	8	Rajasthan	342,239	68,621,01	19,260	21,440	926	17,800	13,500	67,060	80,510	52,660
2015-16	9	Uttar Pradesh	240,928	199,581,48	17,310	20,090	908	20,700	0,600	69,720	79,240	59,260
2015-16	10	Bihar	94,163	103,804,64	8,360	25,070	916	15,900	1,300	63,820	73,390	53,330
2015-16	11	Sikkim	7,096	607,69	9,850	12,360	889	4,600	33,800	82,200	87,290	76,430
2015-16	12	Arunachal Pradesh	83,743	1,382,61	16,480	25,920	920	0,000	68,800	66,950	73,690	59,570
2015-16	13	Nagaland	16,579	1,980,60	17,310	-0,470	931	0,000	86,500	80,110	83,290	76,690
2015-16	14	Manipur	22,327	2,721,76	21,160	18,660	987	3,800	35,100	79,850	86,490	73,170
2015-16	15	Mizoram	21,081	1,091,01	40,420	22,780	975	0,100	94,400	91,580	93,720	89,400

Data sets for the Statewise_education data

The screenshot shows the Kaggle Data Explorer interface for the dataset '2015_16_Districtwise.csv' (1.9 MB). The table view shows columns for AC_YEAR, STATCD, DISTCD, STATNAME, and DISTNAME. Summary statistics indicate 1 unique value for AC_YEAR, 35 for STATCD, 101 for DISTCD, and 680 unique values for DISTNAME. The table lists districts across various states like Jammu & Kashmir, Himachal Pradesh, and Uttarakhand.

AC_YEAR	STATCD	DISTCD	STATNAME	DISTNAME
2015-16	01	0118	JAMMU & KASHMIR	RAMBAN
2015-16	01	0119	JAMMU & KASHMIR	GANDERBAL
2015-16	01	0120	JAMMU & KASHMIR	KULGAR
2015-16	01	0121	JAMMU & KASHMIR	REASI
2015-16	01	0122	JAMMU & KASHMIR	SAMBA
2015-16	02	0201	HIMACHAL PRADESH	CHAMBA
2015-16	02	0202	HIMACHAL PRADESH	KANGRA
2015-16	02	0203	HIMACHAL PRADESH	LAHUL & SPITI
2015-16	02	0204	HIMACHAL PRADESH	KULLU
2015-16	02	0205	HIMACHAL PRADESH	HANDI
2015-16	02	0206	HIMACHAL PRADESH	HAMIRPUR (H.P.)
2015-16	02	0207	HIMACHAL PRADESH	UNA

Data sets for the Statewise_district data

Kaggle interface showing the 'Unemployment in India.csv' dataset. The interface includes a search bar, navigation tabs (Data, Code, Discussion, Activity, Metadata), and a 'Download (66 kB)' button. The 'Data Explorer' section displays the dataset name and size (65.32 kB). Below this, there are tabs for 'Detail', 'Compact', and 'Column'. The 'Detail' view shows a table with columns: Frequency, # Estimated Unemp..., # Estimated Employ..., # Estimated Labour..., and Area. The table includes a summary row and a detailed data table with columns: Frequency, # Estimated Unemp..., # Estimated Employ..., # Estimated Labour..., and Area. The 'Area' column lists 'Urban' and 'Rural' with their respective percentages. The 'Summary' section shows 2 files and 16 columns.

sets for the unemployment data

PROCESSING THE DATASET ON TABLEAU

Selecting the specific data sheet makes it easy to relate the

I started cleaning the data as to my requirements. using libre office.

LibreOffice Calc interface showing a spreadsheet titled 'Unemployment.XLSX - LibreOffice Calc'. The spreadsheet displays a table with columns for 'State/Union Territory' and years from 1993-94 to 2018-19. The table is titled 'TABLE 12: STATE-WISE UNEMPLOYMENT RATE: USUAL STATUS (ADJUSTED) (Concid.) (Rural Over all) (Per 1000)'. The data is organized into rows for each state/territory, with the final row for 'ALL INDIA'.

State/Union Territory	1993-94	1999-00	2004-05	2009-10	2011-12	2017-18	2018-19
ANDHRA PRADESH	4	8	7	12	12	36	45
ARUNACHAL PRADESH	10	5	9	13	17	53	73
ASSAM	52	39	26	38	45	83	63
BIHAR	16	18	15	20	32	70	102
CHHATTISGARH	-	47	19	17	78	35	5
DELHI	-	6	6	6	8	25	10
GOA	90	93	111	47	51	139	80
GUJARAT	9	4	5	8	3	52	33
HARYANA	11	8	22	18	24	93	85
HIMACHAL PRADESH	5	12	18	16	10	52	48
JAMMU & KASHMIR	7	11	15	25	25	42	39
JHARKHAND	1	14	38	211	71	46	46
KARNATAKA	7	7	7	5	9	38	27
KERALA	69	82	107	75	68	100	84
MADHYA PRADESH	5	5	5	7	4	36	24
MAHARASHTRA	8	14	10	6	7	33	42
MANIPUR	10	19	11	38	26	116	98
MIZORAM	2	4	3	4	4	6	20
NAGALAND	10	9	3	13	18	65	52
NAGALAND	14	24	18	106	151	216	162
ODISHA	14	19	58	38	22	68	61
PUNJAB	13	18	38	28	19	78	77
RAJASTHAN	3	4	7	4	7	45	46
SIKKIM	7	28	24	43	10	27	25
TAMIL NADU	13	20	12	15	20	78	64
TRIPURA	23	12	133	92	105	63	93
UTTARAKHAND	3	13	15	16	25	69	72
UTTAR PRADESH	7	8	6	10	9	55	43
WEST BENGAL	18	28	23	19	27	38	35
ANDAMAN AND NICOBAR ISLANDS	21	34	62	88	54	147	146
CHANDIGARH	28	7	26	247	-	35	16
DADRA & NAGAR HAVELI	8	10	33	48	-	7	11
DAMAN & DIU	12	10	3	40	62	-	-
LAKSHADWEEP	169	194	75	97	160	133	480
PUDUCHERRY	24	40	70	30	8	104	116
ALL INDIA	12	15	17	16	17	53	50

Since this was having a different case structure of the State column which doesn't not match with the education data from kaggle.

DATA VISUALIZATION USING TABLEAU

Insights from the process:

1. Import the data in tableau
2. Use sheet view to come to place column and rows in graph view.
3. Check the number of persons and compare it with the data of the unemployment
5. Comparative analysis of statewise education vs unemployment visualisation.
6. Comparative analysis of literacy rate of the people in a particular state.