INTRODUCTION TO DATA MANAGEMENT

PROJECT REPORT

(Project Semester August-December 2021)

INDIAN STATE WISE SUICIDE DATA ANALYSIS

Submitted by

Qamber Hasnain

Registration No. 11907273

Section: K19BH

Course Code: INT217

Under the Guidance of

Ashu Ma'am

UID: 23631

Discipline of CSE/IT

School of Computer Science and Engineering

Lovely Professional University, Phagwara

INDEX

CERTIFICATE

This is to certify that **Qamber Hasnain** bearing Registration no. **11907273** has completed **INT217** project titled, **"Indian State wise Suicide Data Analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Mrs. Ashu

Assistant Professor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 15-12-2021

DECLARATION

I, Qamber Hasnain, student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 15-12-2021

Registration No. 11907273

Qamber Hasnain

Acknowledgement

I would like to express my special thanks of gratitude to my teacher *Mrs. Ashu* who gave me the golden opportunity to do this wonderful project on the topic Indian State wise Suicide Data Analysis which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to her.

Qamber Hasnain

11907273

Introduction

What is Analysis?

Analysis, refers to dividing a whole into its separate components for individual examination. Data analysis, is a process for obtaining raw data, and subsequently converting it into information useful for decision-making by users. Data, is collected and analyzed to answer questions, test hypotheses, or disprove theories.

Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.

Why we need it?

Data analysis is important in business to understand problems facing an organization, and to explore data in meaningful ways. Data in itself is merely facts and figures. Data analysis organizes, interprets, structures and presents the data into useful information that provides context for the data.

Steps of analysis of data:

a) Introduction.

b) Specific Requirements/Functions and Formulas.

i) Pivot table of the data.

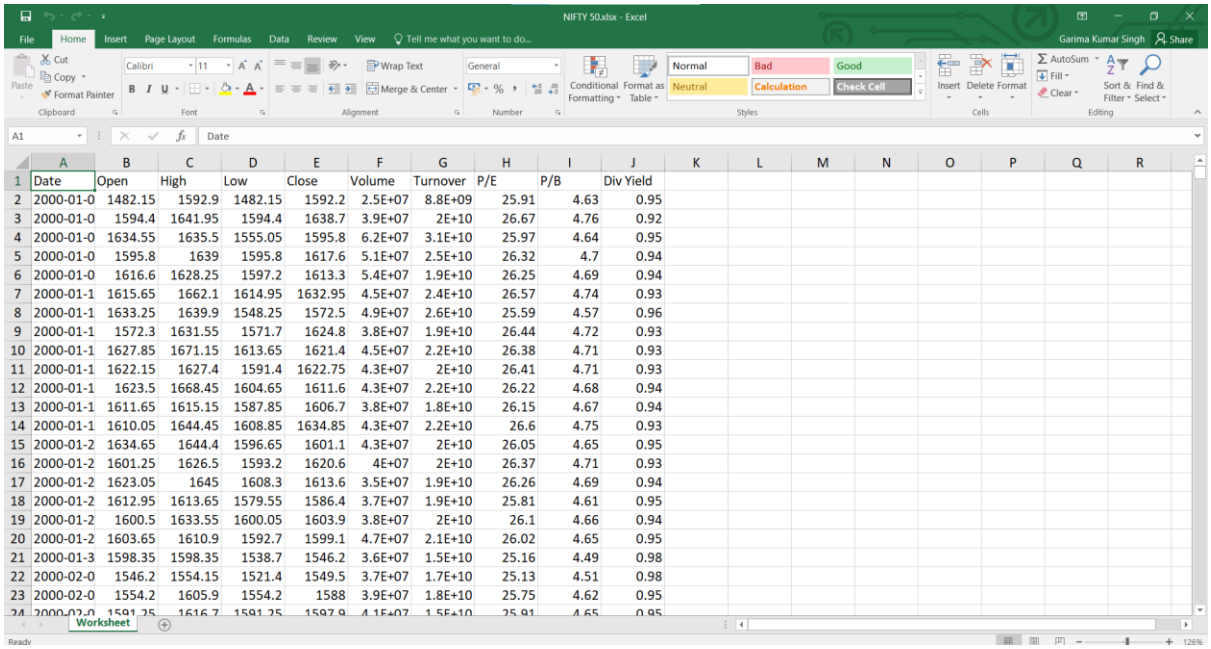
ii) With the help of data Clustered Chart is plotted.

c) Analysis Results.

d) Visualization.

Introduction to MS Excel:

The use of Excel is widespread in the industry. It is a very powerful data analysis tool and almost all big and small businesses use Excel in their day-to-day functioning. This is an introductory course in the use of Excel and is designed to give you a working knowledge of Excel with the aim of getting to use it for more advance topics in Business Statistics later. The course is designed keeping in mind two kinds of learners - those who have very little functional knowledge of Excel and those who use Excel regularly but at a peripheral level and wish to enhance their skills. The course takes you from basic operations such as reading data into excel using various data formats, organizing and manipulating data, to some of the more advanced functionality of Excel. All along, Excel functionality is introduced using easy to understand examples which are demonstrated in a way that learners can become comfortable in understanding and applying them.



	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	Volume	Turnover	P/E	P/B	Div Yield
2	2000-01-0	1482.15	1592.9	1482.15	1592.2	2.5E+07	8.8E+09	25.91	4.63	0.95
3	2000-01-0	1594.4	1641.95	1594.4	1638.7	3.9E+07	2E+10	26.67	4.76	0.92
4	2000-01-0	1634.55	1635.5	1555.05	1595.8	6.2E+07	3.1E+10	25.97	4.64	0.95
5	2000-01-0	1595.8	1639	1595.8	1617.6	5.1E+07	2.5E+10	26.32	4.7	0.94
6	2000-01-0	1616.6	1628.25	1597.2	1613.3	5.4E+07	1.9E+10	26.25	4.69	0.94
7	2000-01-1	1615.65	1662.1	1614.95	1632.95	4.5E+07	2.4E+10	26.57	4.74	0.93
8	2000-01-1	1633.25	1639.9	1548.25	1572.5	4.9E+07	2.6E+10	25.59	4.57	0.96
9	2000-01-1	1572.3	1631.55	1571.7	1624.8	3.8E+07	1.9E+10	26.44	4.72	0.93
10	2000-01-1	1627.85	1671.15	1613.65	1621.4	4.5E+07	2.2E+10	26.38	4.71	0.93
11	2000-01-1	1622.15	1627.4	1591.4	1622.75	4.3E+07	2E+10	26.41	4.71	0.93
12	2000-01-1	1623.5	1668.45	1604.65	1611.6	4.3E+07	2.2E+10	26.22	4.68	0.94
13	2000-01-1	1611.65	1615.15	1587.85	1606.7	3.8E+07	1.8E+10	26.15	4.67	0.94
14	2000-01-1	1610.05	1644.45	1608.85	1634.85	4.3E+07	2.2E+10	26.6	4.75	0.93
15	2000-01-2	1634.65	1644.4	1596.65	1601.1	4.3E+07	2E+10	26.05	4.65	0.95
16	2000-01-2	1601.25	1626.5	1593.2	1620.6	4E+07	2E+10	26.37	4.71	0.93
17	2000-01-2	1623.05	1645	1608.3	1613.6	3.5E+07	1.9E+10	26.26	4.69	0.94
18	2000-01-2	1612.95	1613.65	1579.55	1586.4	3.7E+07	1.9E+10	25.81	4.61	0.95
19	2000-01-2	1600.5	1633.55	1600.05	1603.9	3.8E+07	2E+10	26.1	4.66	0.94
20	2000-01-2	1603.65	1610.9	1592.7	1599.1	4.7E+07	2.1E+10	26.02	4.65	0.95
21	2000-01-3	1598.35	1598.35	1538.7	1546.2	3.6E+07	1.5E+10	25.16	4.49	0.98
22	2000-02-0	1546.2	1554.15	1521.4	1549.5	3.7E+07	1.7E+10	25.13	4.51	0.98
23	2000-02-0	1554.2	1605.9	1554.2	1588	3.9E+07	1.8E+10	25.75	4.62	0.95
24	2000-02-0	1591.25	1616.7	1591.25	1597.9	4.1E+07	1.5E+10	25.91	4.65	0.95

1. Introduction to Spreadsheets

The use of Excel spreadsheets and various basic data functions of Excel.

Topics covered include:

• Reading data into Excel using various formats

• Basic functions in Excel, arithmetic as well as various logical functions

• Formatting rows and columns

• Using formulas in Excel and their copy and paste using absolute and relative referencing

2. Spreadsheet Functions to Organize Data

Introduces various Excel functions to organize and query data. Learners are introduced to the IF, nested IF, VLOOKUP and the HLOOKUP functions of Excel.

Topics covered include:

• IF and the nested IF functions

• VLOOKUP and HLOOKUP

3. Introduction to Filtering, Pivot Tables, and Charts

Introduces various data filtering capabilities of Excel. You'll learn how to set filters in data to selectively access data. A very powerful data summarizing tool, the Pivot Table, is also explained and we begin to introduce the charting feature of Excel.

Topics covered include:

• VLOOKUP across worksheets

• Data filtering in Excel

• Use of Pivot tables with categorical as well as numerical data

• Introduction to the charting capability of Excel

4. Advanced Graphing and Charting

Explores various advanced graphing and charting techniques available in Excel. Starting with various line, bar and pie charts we introduce pivot charts, scatter plots and histograms. You will get to understand these various charts and get to build them on your own.

Topics covered include

• Line, Bar and Pie charts

• Pivot charts

• Scatter plots

• Histograms

Introduction to Tableau Prep Builder:

Data can be generated, captured, and stored in a dizzying variety of structures, but when it comes to analysis, not all data formats are created equal.

Data preparation is the process of cleaning dirty data, restructuring ill-formed data, and combining multiple sets of data for analysis. It involves transforming the data structure, like rows and columns, and cleaning up things like data types and values. The speed and efficiency of your data prep process directly impacts the time it takes to discover insights. Understanding the scope of data you are analyzing and seeing the changes you make to the data can accelerate the entire process.

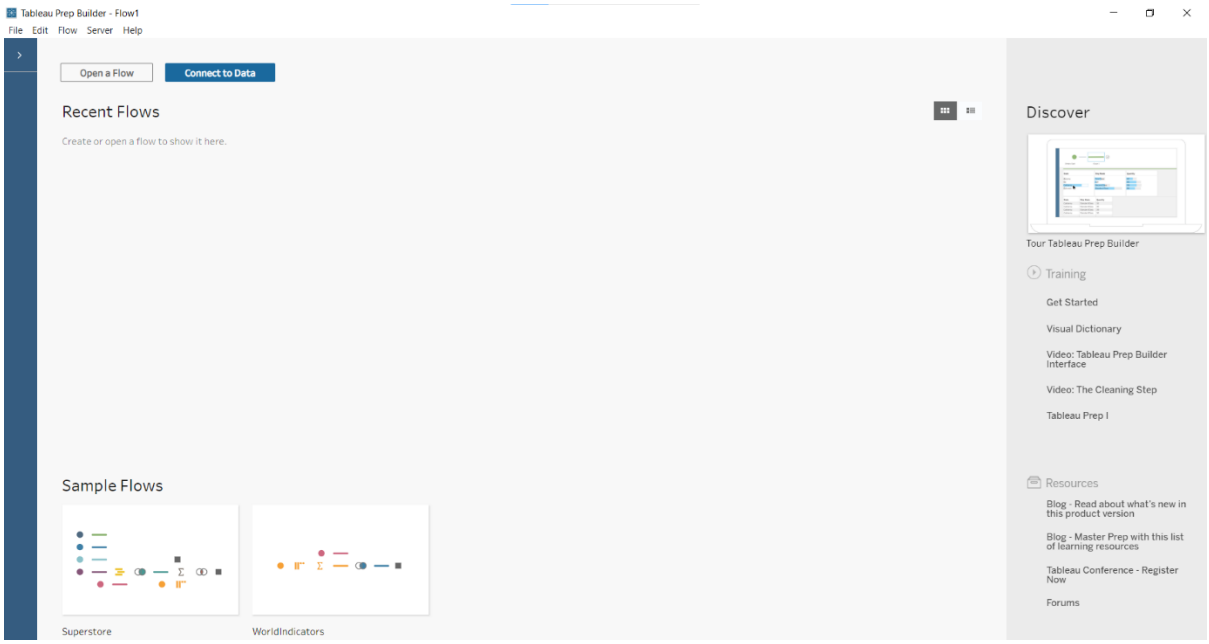
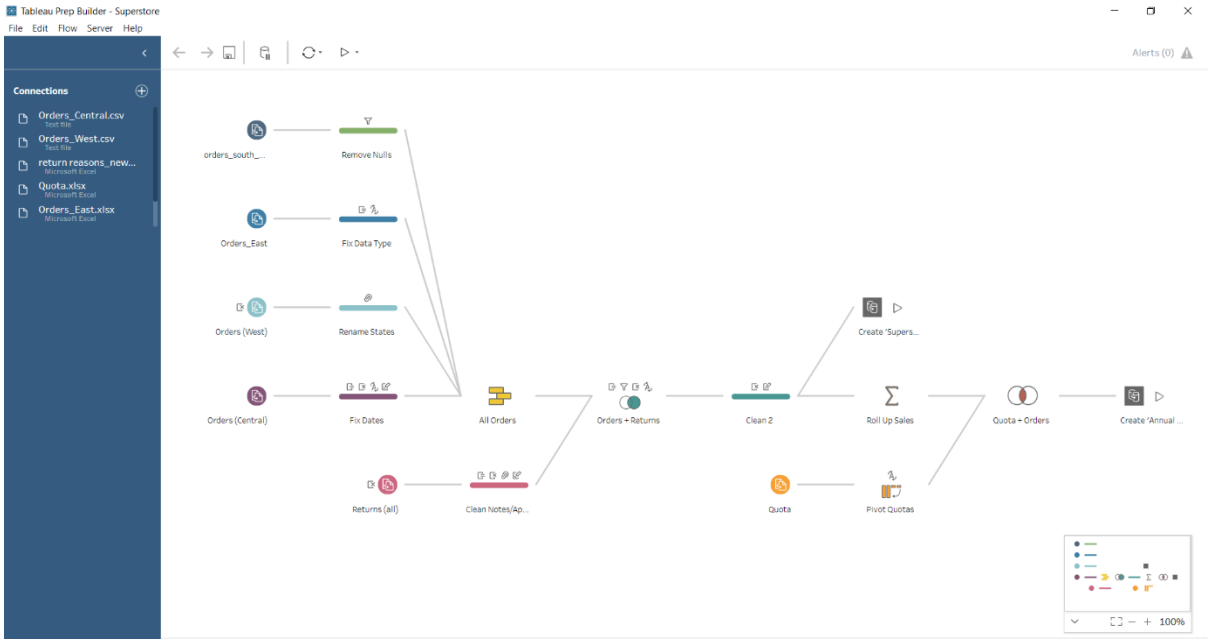


Tableau Prep Builder provides various cleaning operations that you can use out of the box to clean and shape your data. Cleaning up dirty data makes it easier to combine and analyze your data or makes it easier for others to understand your data when sharing your data sets.

You can also clean your data using a pivot step or a script step to apply R or Python scripts to your flow.

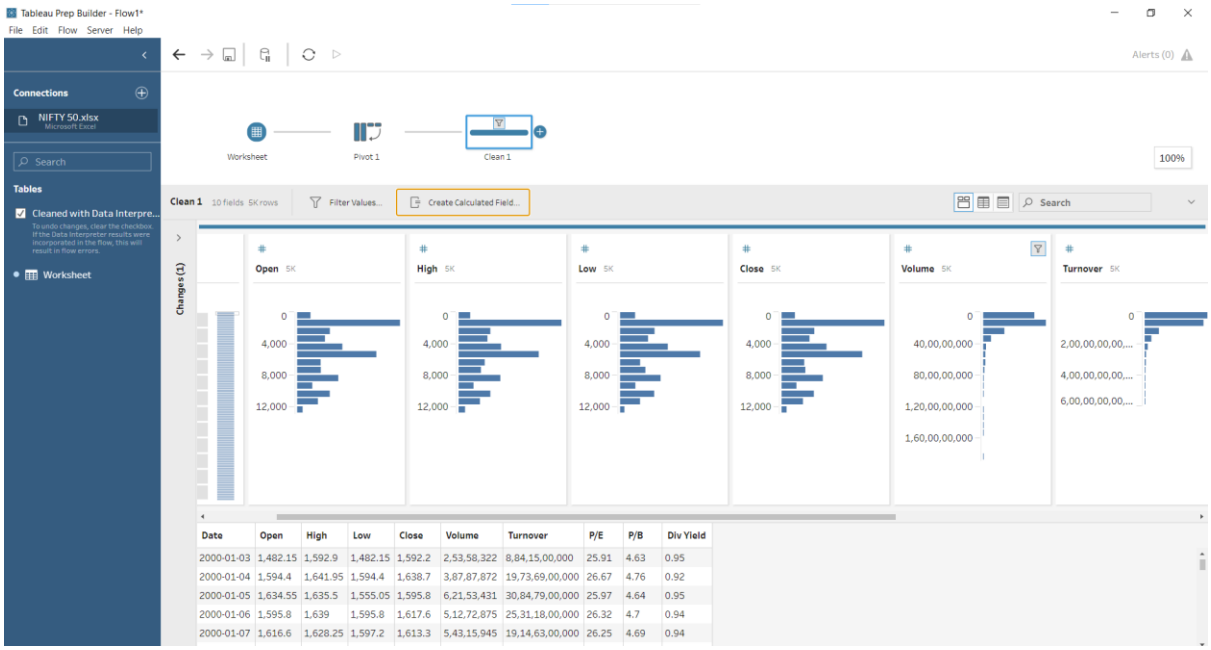
You clean data by applying cleaning operations such as filtering, adding, renaming, splitting, grouping, or removing fields. In prior releases, cleaning operations could only be performed in the clean step type. In Tableau Prep Builder version 2018.2.1 and later, you could also perform cleaning operations in the data grid in a cleaning step.



Apply cleaning operations:

To apply cleaning operations to fields, use the toolbar options or click more options on the field profile card, data grid or Results pane to open the menu.

In Aggregate, Pivot, Join and Union step types, the more options menu is available on the profile cards in the Results pane and corresponding data grid. If you perform the same cleaning operations or actions over and over throughout your flow, you can copy and paste your steps, actions or even fields.



Apply cleaning operations using recommendations:

Sometimes it can be hard to identify which cleaning operation you need to use to fix problems in your data. Tableau Prep Builder can analyze your data and recommend cleaning operations that you can apply automatically to quickly fix problems in your data fields or help to identify problems so you can fix them. This feature is available in all step types except Input, Output and Join step types.

Introduction to Project Topic:

Each suicide is a personal tragedy that prematurely takes the life of an individual and has a continuing ripple effect, dramatically affecting the lives of families, friends and communities. According to the data from 2001-2012, every year, on average 6,04,001 people commit suicide in our country. There are various causes of suicides like professional/career problems, sense of isolation, abuse, violence, family problems, mental disorders, addiction to alcohol, financial loss, chronic pain etc.

Majority of suicides were reported in Maharashtra (18,916) followed by 13,493 suicides in Tamil Nadu, 12,665 suicides in West Bengal, 12,457 suicides in Madhya Pradesh and 11,288 suicides in Karnataka accounting for 13.6%, 9.7%, 9.1%, 9.0% and 8.1% of total suicides respectively. These 5 States together accounted for 49.5% of the

total suicides reported in the country. The remaining 50.5% suicides were reported in the remaining 24 States and 7 UTs. Uttar Pradesh, the most populous State (16.9% share of country population) has reported comparatively lower percentage share of suicidal deaths, accounting for only 3.9% of the total suicides reported in the country.

Delhi, which is the most-populous UT, has reported the highest number of suicides (2,526) among UTs, followed by Puducherry (493). Remaining UTs together accounted for 2.2% of total suicides in the country. The States and UTs which have reported significant percentage increase in suicides in 2019 over 2018 were Bihar (44.7%), Punjab (37.5%), Daman & Diu (31.4%), Jharkhand (25.0%), Uttarakhand (22.6%) and Andhra Pradesh (21.5%) while highest percentage decrease was reported in Lakshadweep (100%), Himachal Pradesh (21.1%), Chandigarh (18.1%), Arunachal Pradesh (15.2%) and Jammu & Kashmir (13.9%).

Objective of Analysis

- OBJ 1 - Top 10 states with high suicide rate.
- OBJ 2 - Suicides due to Failure in Examination with respect to states.
- OBJ 3 - Female suicides due to Illegitimate Pregnancy with respect to age groups.
- OBJ 4 - Male suicides of age below 30 due to Love Affairs.
- OBJ 5 - Suicides due to Unemployment with respect to states.

Project: Data Analysis

The dataset that I have chosen for this project is a census data of the Suicides happened between 2001-2012 arranged according to states of India. (source: https://www.kaggle.com/rajanand/suicides-in-india). Below are the field names of the dataset and its description [Table1.0]:

1. 'State': Name of the state
2. 'Year': Year of suicides
3. 'Type_code': Major category of suicide
4. 'Type': Specific category of suicide(reason)
5. 'Gender': Male or Female
6. 'Age_group': Groups of age (0-14,15-29…)
7. 'Total': Number of suicides reported

	A	B	C	D	Gender	F	G
1	State	Year	Type_code	Type	Gender	Age_group	Total
2	A & N Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0
3	A & N Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
4	A & N Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
5	A & N Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0

Table 1.0

OBJ 1 - Top 10 states with high suicide rate

The Excel workbook's sheet named "OBJ 1" has a pivot table [Table 1.1] which show the data for top 10 states with high suicide rate with respect to male and female separately. This data includes all the suicides happened between 2001-2012 only. Then I used pivot line chart [Chart 1.1] to visually understand the differences which can be found in Dashboard sheet.

Sum of Total	Column Labels		
Row Labels	Female	Male	Grand Total
Andhra Pradesh	2,71,939	5,42,120	8,14,059
Chhattisgarh	98,574	2,03,780	3,02,354
Gujarat	1,32,493	1,98,365	3,30,858
Karnataka	2,42,870	4,91,955	7,34,825
Kerala	1,45,153	3,93,793	5,38,946
Madhya Pradesh	2,03,150	2,48,385	4,51,535
Maharashtra	2,93,175	6,08,770	9,01,945
Odisha	1,09,605	1,57,629	2,67,234
Tamil Nadu	3,06,485	5,12,206	8,18,691
West Bengal	3,65,241	4,84,695	8,49,936
Grand Total	**21,68,685**	**38,41,698**	**60,10,383**

Table 1.1

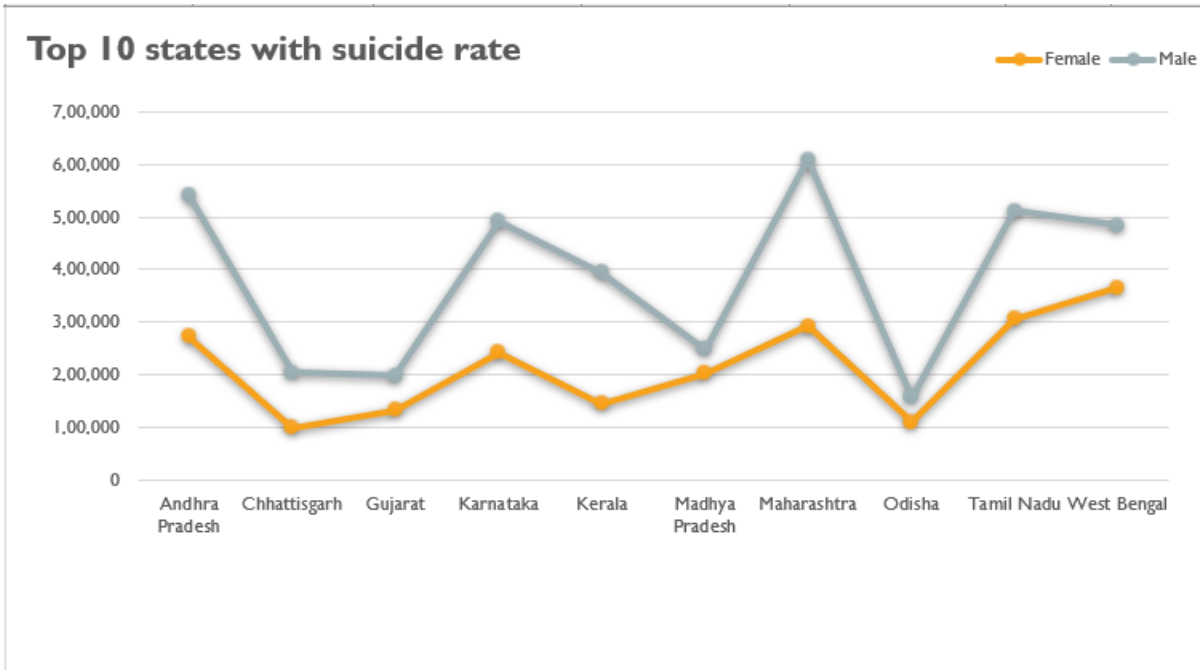


Chart 1.1

In this chart States are arranged in alphabetical order and legends already mentions that female census has orange line and male has grey. But as we can see male suicides are higher with respect to females. Another noticeable point is Maharashtra, Tamil Nadu, West Bengal, Karnataka and Andhra Pradesh contributes most male suicides than any of the other 34 States and UTs.

OBJ 2 - Suicides due to Failure in Examination with respect to states

The Excel workbook's sheet named "OBJ 2" has a pivot table [Table 1.2] which shows data for Top 10 states with suicides happened due to Failure in Examination and categorised with respect to gender. Then I used pivot Bar chart [Chart1.2] to visually analyse and understand the data, this chart can be found in Dashboard sheet.

Sum of Total	Column Labels		
	Female	Male	Grand Total
Row Labels	Failure in Examination	Failure in Examination	
Odisha	336	507	843
Gujarat	506	543	1049
Assam	554	798	1352
Uttar Pradesh	547	882	1429
Madhya Pradesh	721	940	1661
Karnataka	809	1060	1869
Andhra Pradesh	1028	1521	2549
Tamil Nadu	1468	1543	3011
Maharashtra	1578	1706	3284
West Bengal	2443	2928	5371

Table 1.2

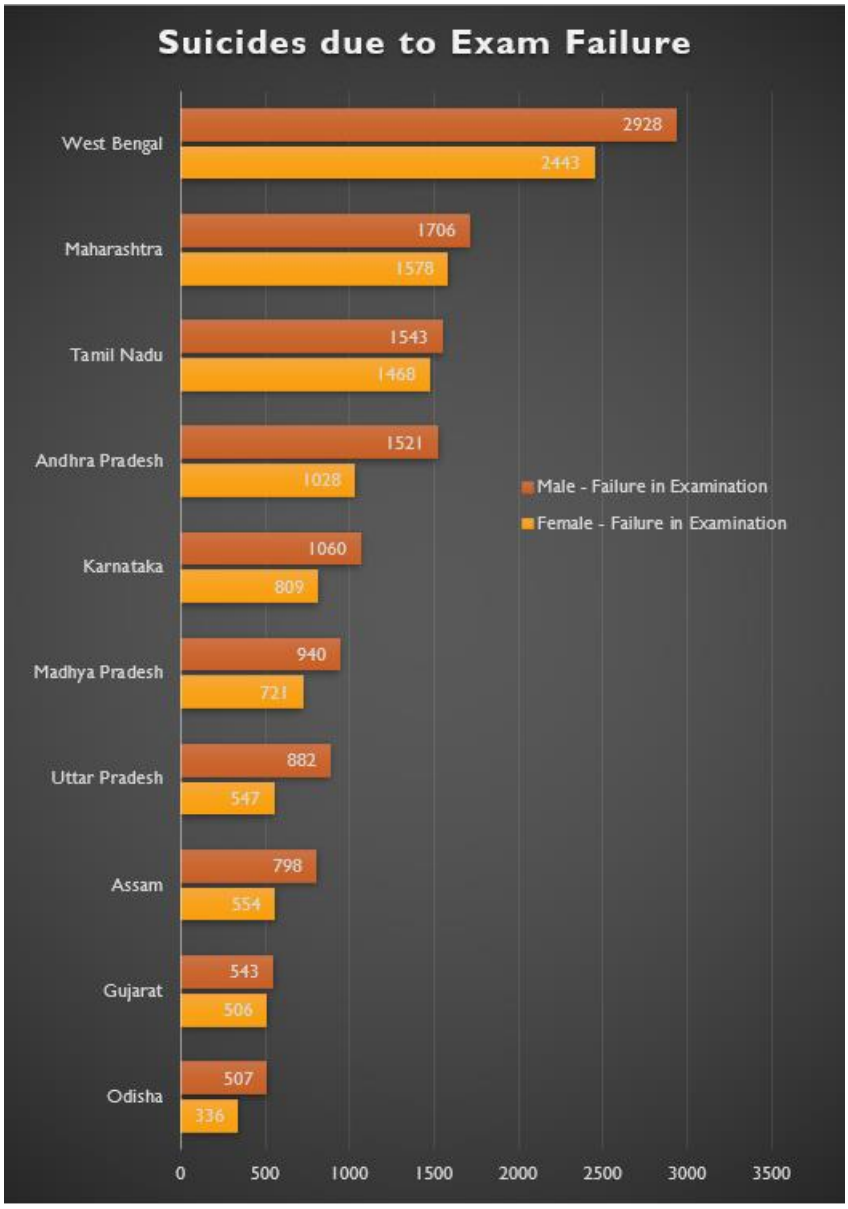


Chart 1.2

OBJ 3 - Female suicides due to Illegitimate Pregnancy with respect to age groups

The Excel workbook's sheet named "OBJ 3" has a pivot table [Table 1.3] which shows data regarding female suicides happened due to Illegitimate Pregnancy with respect to age groups and it also shows suicide percentage to understand which age group is crucial and fragile. Then I used pivot pie chart [Chart 1.3] to visual present and understand the data, this chart can be found in Dashboard.

	Column Labels Female	
Row Labels	Number of suicides	Suicides %
Illegitimate Pregnancy		
0-14	66	2.65%
15-29	1639	65.72%
30-44	625	25.06%
45-59	61	2.45%
60+	0	0.00%
Grand Total	**2391**	**95.87%**

Table 1.3

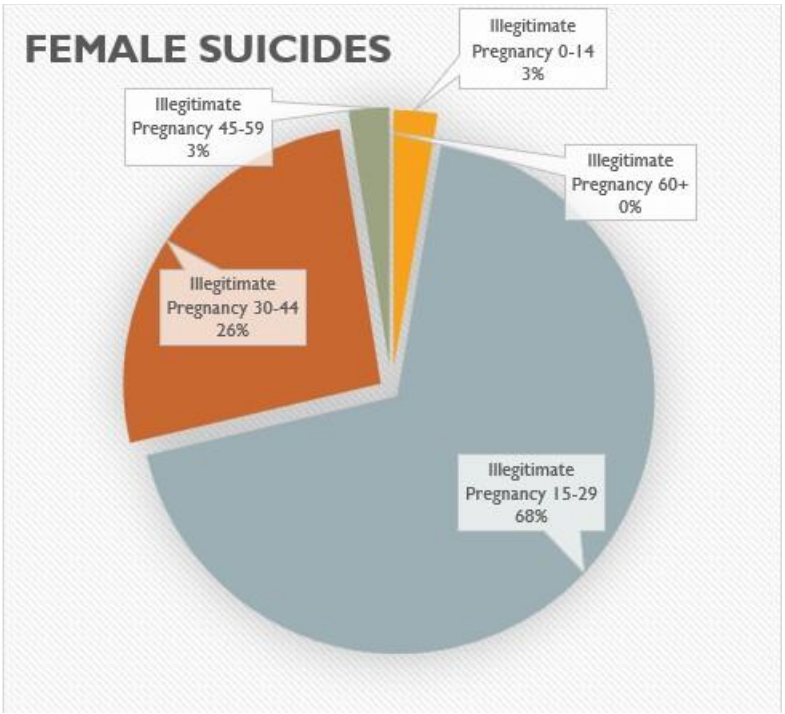


Chart 1.3

OBJ 4 - Suicides of age below 30 due to Love Affairs

The Excel workbook's sheet named "OBJ 4" has three pivot tables [Table 1.4.1, Table 1.4.2, Table 1.4.3] in which first pivot table shows data of Top 10 States having suicides due to Love Affairs with age below 30 and arranged according to years (2001-2012). Second and third pivot table shows data with respect to genders, male and female respectively. Then I used Spark line pivot chart [Chart 1.4] to show the differences visually in a single chart that can be found in Dashboard sheet.

Gender	(All)																									
Sum of Total	Column Labels																									
	2001		2002		2003		2004		2005		2006		2007		2008		2009		2010		2011		2012			
Row Labels	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29
Love Affairs																										
Andhra Pradesh	6	149	4	184	8	199	8	234	5	204	10	215	22	238	10	213	10	193	2	312	18	293	15	242		
Assam	1	248	7	170	4	290	1	287	4	249	2	277	0	228	4	233	1	212	4	251	1	270	1	271		
Gujarat	3	123	1	151	3	119	2	125	1	115	1	132	3	113	4	116	0	148	1	139	1	158	0	211		
Karnataka	4	105	2	88	6	103	4	75	0	79	3	83	1	74	0	87	8	83	21	92	113	148	16	276		
Madhya Pradesh	2	91	7	119	4	123	10	112	0	97	2	105	13	116	3	155	10	156	17	187	14	223	13	264		
Maharashtra	2	188	7	207	4	195	2	179	3	205	0	190	1	191	4	171	1	180	3	211	7	236	2	262		
Odisha	1	174	1	104	7	169	8	114	7	122	5	98	4	123	5	146	23	138	12	152	15	229	27	212		
Tamil Nadu	12	276	4	250	22	327	17	389	18	306	0	344	0	380	0	492	0	331	4	453	9	399	10	344		
Uttar Pradesh	5	121	4	113	8	120	0	96	4	98	0	105	5	135	3	104	5	184	6	166	5	201	9	168		
West Bengal	46	342	18	421	5	707	11	588	46	699	19	1070	30	478	44	569	50	575	36	542	29	635	0	0		

Table 1.4.1

Gender	Male																									
Sum of Total	Column Labels																									
	2001		2002		2003		2004		2005		2006		2007		2008		2009		2010		2011		2012			
Row Labels	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29
Love Affairs																										
Andhra Pradesh	3	84	0	109	6	119	1	139	2	97	7	128	6	135	0	120	4	112	0	203	3	151	0	138		
Assam	1	131	2	92	1	151	1	169	1	141	0	168	0	116	3	113	0	100	0	128	0	170	0	160		
Gujarat	0	56	0	73	1	56	0	70	0	61	0	64	0	59	0	56	0	67	0	66	0	80	0	119		
Karnataka	1	51	0	48	1	47	0	34	0	49	1	47	0	41	0	56	0	46	2	60	113	78	16	137		
Madhya Pradesh	0	45	0	49	0	62	2	58	0	40	0	48	0	52	0	82	3	83	7	82	4	126	4	142		
Maharashtra	0	72	0	106	0	108	0	94	0	129	0	96	0	108	0	88	0	92	0	114	0	128	1	145		
Odisha	0	65	0	57	2	90	1	54	2	48	1	40	0	48	1	68	11	64	2	66	6	109	4	87		
Tamil Nadu	10	161	1	123	7	161	6	211	10	163	0	173	0	200	0	197	0	164	1	241	0	218	0	171		
Uttar Pradesh	1	54	1	56	4	73	0	46	2	47	0	53	1	56	2	49	0	104	2	82	0	92	0	88		
West Bengal	5	179	0	219	0	362	4	320	17	286	4	515	9	200	9	272	13	281	14	273	10	290	0	0		

Table 1.4.2

Gender	Female																									
Sum of Total	Column Labels																									
	2001		2002		2003		2004		2005		2006		2007		2008		2009		2010		2011		2012			
Row Labels	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29	0-14	15-29
Love Affairs																										
Andhra Pradesh	3	65	4	75	2	80	7	95	3	107	3	87	16	103	10	93	6	81	2	109	15	142	15	104		
Assam	0	117	5	78	3	139	0	118	3	108	2	109	0	112	1	120	1	112	4	123	1	100	1	111		
Gujarat	3	67	1	78	2	63	2	55	1	54	1	68	3	54	4	60	0	81	1	73	1	78	0	92		
Karnataka	3	54	2	40	5	56	4	41	0	30	2	36	1	33	0	31	8	37	19	32	0	70	0	139		
Madhya Pradesh	2	46	7	70	4	61	8	54	0	57	2	57	13	64	3	73	7	73	10	105	10	97	9	122		
Maharashtra	2	116	7	101	4	87	2	85	3	76	0	94	1	83	4	83	1	88	3	97	7	108	1	117		
Odisha	1	109	1	47	5	79	7	60	5	74	4	58	4	75	4	78	12	74	10	86	9	120	23	125		
Tamil Nadu	2	115	3	127	15	166	11	178	8	143	0	171	0	180	0	295	0	167	3	212	9	181	10	173		
Uttar Pradesh	4	67	3	57	4	47	0	50	2	51	0	52	4	79	1	55	5	80	4	84	5	109	9	80		
West Bengal	41	163	18	202	5	345	7	268	29	413	15	555	21	278	35	297	37	294	22	269	19	345	0	0		

Table 1.4.3

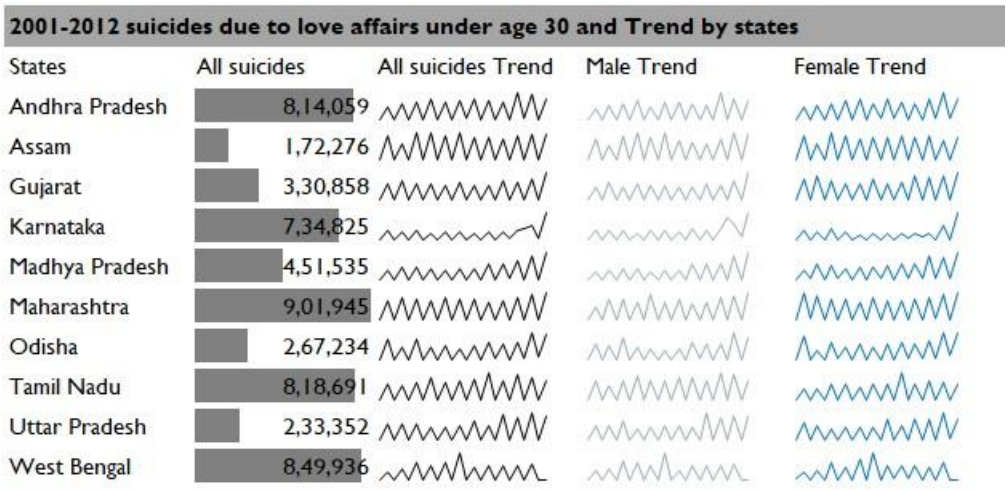


Chart 1.4

OBJ 5 - Suicides due to Unemployment with respect to states

The Excel workbook's sheet named "OBJ 5" has a pivot table [Table 1.5] which shows data for suicides happened in each state due to unemployment. According to this table most of the suicides happened in 3 states i.e., Tamil Nadu, Maharashtra and West Bengal. Then I used 3D map chart [Chart 1.5] to get the clear idea. As the map was 3D and couldn't be used in the same Dashboard sheet, I had to take the snapshot and put it in the dashboard that's why only map is not the interactive part of the dashboard.

States	Unemployed
A & N Islands	77
Andhra Pradesh	7504
Arunachal Pradesh	115
Assam	1666
Bihar	631
Chandigarh	151
Chhattisgarh	3637
D & N Haveli	31
Daman & Diu	13
Delhi (Ut)	3518
Goa	737
Gujarat	6123
Haryana	3137
Himachal Pradesh	465
Jammu & Kashmir	392
Jharkhand	853
Karnataka	7259
Kerala	13655
Lakshadweep	0
Madhya Pradesh	3595
Maharashtra	11389
Manipur	75
Meghalaya	171
Mizoram	205
Nagaland	48
Odisha	5324
Puducherry	1146
Punjab	761
Rajasthan	1918
Sikkim	267
Tamil Nadu	20770
Tripura	1125
Uttar Pradesh	4050
Uttarakhand	385
West Bengal	13181
Grand Total	**114374**

Table 1.5

Chart 1.5

Bibliography

- Dataset from Kaggle.com
- https://www.kaggle.com/rajanand/suicides-in-india
- Information from Wikipedia
- Information from www.Google.com
- Learnt some things from YouTube channels

----------END OF REPORT----------