INTRODUCTION TO DATA MANAGEMENT
PROJECT REPORT

DATA ANALYSIS ON SUICIDES IN INDIA

(Project Semester August-December 2019)

Submitted by

Adarsh Kumar Singh
Registration No. 11716553
Section: KM067
Course Code: INT217

Under the Guidance of

Assistant Professor
Ms. Sandeep Kaur
UID: 23614

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

INDEX

Contents **Page no.**

CERTIFICATE

This is to certify that **Adarsh Kumar Singh** bearing Registration no. **11716553** has completed **INT217** project titled, **"Data Analysis on Suicides in India"** 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.

Ms. Sandeep Kaur
Assistant Professor
School of Computer Science and Engineering

Lovely Professional University Phagwara,
Punjab.

Date: 20-11-2019

DECLARATION

I, Adarsh Kumar Singh, 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: 20-11-2019 Signature

Adarsh Kumar Singh

Registration No. **11716553** Name of the student

Acknowledgement

This report has been prepared for the project "**Data Analysis on Suicides in India"** that has been done in order to study the practical aspect of the course Introduction to Data Management and its implementation in the real-life based data analysis with the purpose of fulfilling the requirements of the course of B.Tech (Bachelor of Technology). This project aimed to be familiar with the practical aspect and uses of theoretical knowledge and clarifying the career goals, so I have completed the project and compiled this report as the summary and the conclusion that have drawn from the project. I would like to express my sincere gratitude to our project coordinator who has given their valuable time and given me the chance to learn something new despite having their busy schedule. I am thankful to **Ms. Sandeep Kaur** for her great guidelines for the analysis of data and co-operative support and also presented an opportunity for me to have practical experience in the field of data analysis. I am also grateful to all seniors for sharing their experience with me and teaching me different techniques to work with Excel and Tableau to operate effectively and efficiently. Thus, the time spent during this project was very audacious and supportive to my career. In this time I have gained valuable knowledge that will help me to create a great difference as a prospective Data Scientist in the near coming future.

Introduction

Suicide refers to the act or an instance of intentionally killing oneself. According to the World Health Organization, every 40 seconds, a person commits suicide somewhere in the world.

While most countries with high suicide rates are poor, there are also a surprising few, highly developed and rich nations which rank very high in this sad statistics. Generally, committing suicide is about three times more common in men than in women, with poisoning, hanging and firearms being the most common methods.

The official estimates of how many Indians take their lives annually vary greatly but according to some sources, it might be as many as 200,000 people. In India, poisoning is the most common suicide method, followed by hanging and self-immolation. Almost half of the suicides committed in this country are motivated by family and health reasons.

Until 2014, suicide was illegal in India and survivors would face jail term of up to one year. In 2014, however, this law was removed and suicide is no longer illegal in India. Gujarat has much to worry about on people valuing their lives as it recorded 32.6 per cent rise in the number of suicides in 2012 over the figures for the previous year. It is mainly because of growing urbanization and family pressure.

In today's society, where everyone, especially teenager and young adults, are under pressure to perform amazingly well in all fields, especially their career, and where every person wants to excel, suicides are not new. Every person, whether aged 22 or 12, has expectations to live up to – family, teachers and even peers have expectations of all sorts. This has led to rise in the stress levels among many teenagers (12-19 years) and even young adults (20-30 years).

There are many reasons for this stress – pressure to build a career, peer pressure, relationship issues, and family issues and so on.

Objectives: -

(1) The main objective of this study is to identify main causes of suicide among teenagers and young adults.

(2) This study focuses on the major factors of suicide as well as provides suggestions to avoid constantly rising rate of suicide.

Method: -

This study has been conducted by analyzing data of suicides between 2001 and 2012.

Results and Discussion: It was found that suicide was more common in males than females in both teenagers as well as young adults. The major issues which led to an individual committing suicide were failures in life issues such as relationships and career, the two major areas of life on which a person concentrates during this point in life. These are, however, not the only issues which lead to suicide.

Reasons of Suicide in India

Causes of Suicide in India	Total No. of Suicide
Not having Children (Barrenness/Impotency	766
Ideological Causes/Hero Worshipping	2118
Illegitimate Pregnancy	2494
Bankruptcy or Sudden change in Economic Status	2655
Physical Abuse (Rape/Incest Etc.)	3992
Divorce	4133
Paralysis	7286
Not having Children (Barrenness/Impotency)	7822
Illness (Aids/STD)	8723
Cancer	9058
Death of Dear Person	10321
Cancellation/Non-Settlement of Marriage	11296
Professional/Career Problem	12554

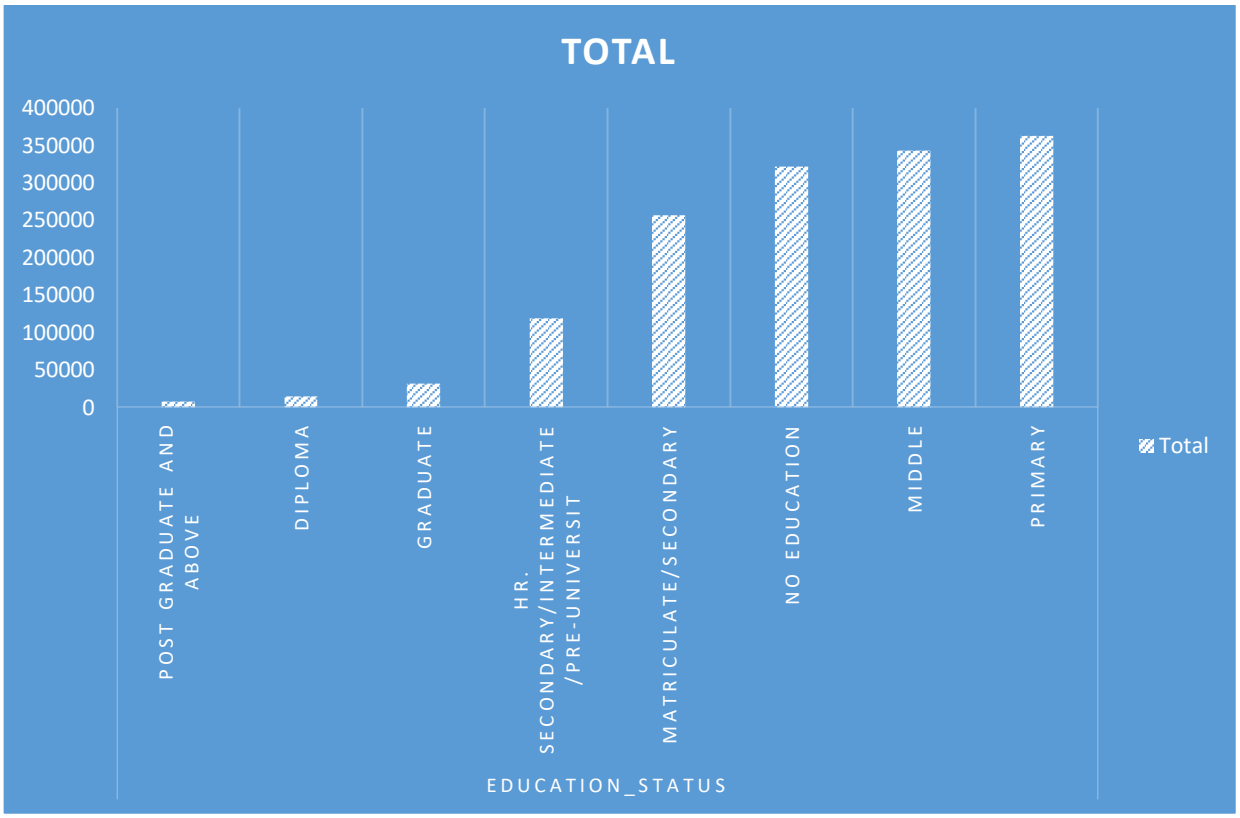
Fall in Social Reputation	13464
Suspected/Illicit Relation	14911
Property Dispute	18652

Failure in Examination	27005
Unemployment	27365
Drug Abuse/Addiction	30046
Dowry Dispute	31970
Poverty	32684
Bankruptcy or Sudden change in Economic	32755
Love Affs	45039
Insanity/Mental Illness	94229
Other Prolonged Illness	194565
Other Causes (Please Specify)	216050
Causes not known	237069
Family Problems	341952

Means adopted for Suicide in India

Means adopted to Suicide in India	Total No. of Suicide
By Machine	1661
By Self Infliction of injury	5093
By Fire-Arms	6294
By Jumping from (Building)	7871
By Jumping off Moving Vehicles/Trains	8116
By jumping from (Other sites)	8127
By other means	9238
By Overdose of sleeping pills	9960
By touching electric wires	10816
By Over Alcoholism	15973
By coming under running vehicles/trains	45299
By Drowning	96711
By Fire/Self Immolation	128006
By other means (please specify)	135132
By Consuming Other Poison	231178
By Consuming Insecticides	275501
By Hanging	460955

Relationship between suicide and Education



The dataset contains the following features

State: It represents the states whose suicide data was measured.

Year: The year when the data was recorded.

Type: The type or causes of suicide where data measurement was made.

Type Code: The reasons of suicide where the data measurement was made.

Gender: The gender of people whose suicide data was recorded.

Age Group: It represent the age group of people whose data was measured.

Total: The total value of suicide committed.

Why these features?

Age Group: Approximately half of India's 1.2 billion people are under the age of 26, and by 2020 we are forecast to be the youngest country in the world, with a median age of 29 years. With this tremendous forecast, it becomes imperative to ensure an environment which promotes positive wellbeing. Unfortunately, India has the highest suicide rate in the world among the youth standing at 35.5 per 100,000 people for 2012. Age group is a major parameter in measuring and analyzing suicide data of India because it tells us which age group of people is committing maximum suicide and how to prevent it. India is quoted to experience the highest rate of suicide among the age bracket of 15-29 years. The reason for such high numbers can be attributed to lack of economic, social, and emotional resources. More specifically, academic pressure, workplace stress, social pressures, modernization of urban centers, relationship concerns, and the breakdown of support systems. Some researchers have attributed the rise of youth suicide to urbanization and the breakdown of the traditional large family support system. The clash of values within families is an important factor for young people in their lives. As young Indians become more progressive, their traditionalist households become less supportive of their choices pertaining to financial independence, marriage age, premarital sex, rehabilitation and taking care of the elderly.

Gender: There is a notable gender difference in the suicidal attempts and completion of suicide. Women are four times more likely than men to attempt suicide (make an attempt but not complete), whereas, men are twice more likely than women to complete the act of suicide. According to WHO data, the age standardized suicide rate in India is 16.4 per 100,000 for women (6th highest in the world) and 25.8 for men (ranking 22nd).

Type Code: Domestic violence is a major risk factor for suicide in a case study performed in Bangalore. However, as a fraction of total suicides, violence against women – such as domestic violence, rape, incest and dowry – accounted for less than 4% of total suicides. India's economy vastly depends on agriculture with around 60% of its people directly or indirectly depend upon it. Different reasons like droughts, low yield prices, exploitation by middlemen and inability to pay loans lead Indian farmers to die by suicide. That's one student every hour. In spite of being the advanced states of India, Maharashtra holds the first place with 921 of 16112 suicides (5.72%) and Tamil Nadu holds the second place with 795 of 16927 suicides (4.70%) and

Objectives/Scope of the Analysis

- State wise Suicide data analysis of India
- Age group wise Suicide data analysis of India
- Cause wise Suicide data analysis of India
- Profession wise Suicide data analysis of India
- Year wise Suicide data analysis of India
- Union Territories wise Suicide data analysis of India
- Current and Future Predication (till 2020) Suicide of India

Source of datasets

All the dataset used in this is collected from very well-known online website www.kaggle.com and data.gov.in.

ETL (Extract, Transform, and Load) Process: -

What is ETL?

ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL full-form is Extract, Transform and Load.

Why do you need ETL?

There are many reasons for adopting ETL in the organization: -

- It helps companies to analyse their business data for taking critical business decisions.
- Transactional databases cannot answer complex business questions that can be answered by ETL.
- A Data Warehouse provides a common data repository
- ETL provides a method of moving the data from various sources into a data warehouse.
- As data sources change, the Data Warehouse will automatically update.
- Well-designed and documented ETL system is almost essential to the success of a Data Warehouse project.
- Allow verification of data transformation, aggregation and calculations rules.
- ETL process allows sample data comparison between the source and the target system.
- ETL process can perform complex transformations and requires the extra area to store the data. ETL helps to Migrate data into a Data Warehouse. Convert to the various formats and types to adhere to one consistent system.
- ETL is a predefined process for accessing and manipulating source data into the target database.
- ETL offers deep historical context for the business.
- It helps to improve productivity because it codifies and reuses without a need for technical skills.

1) Extraction: -

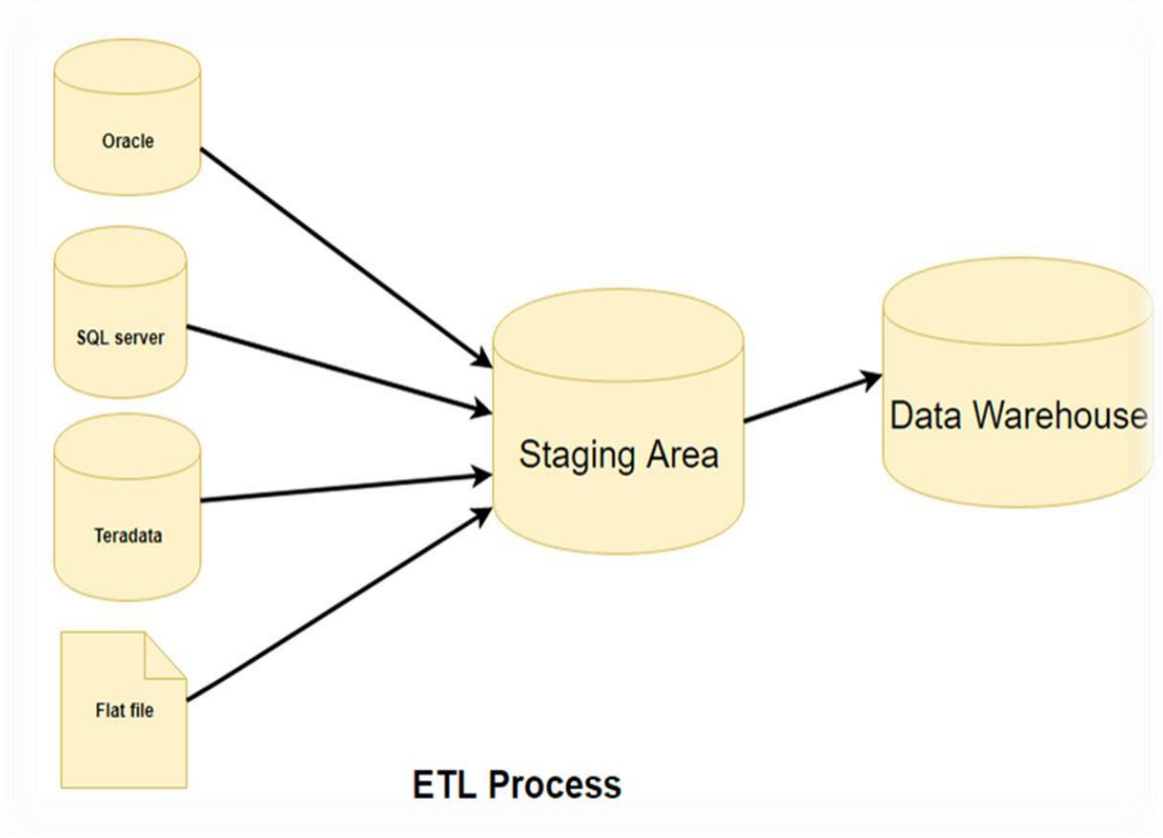
In this step, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system is not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, rollback will be a challenge. Staging area gives an opportunity to validate extracted data before it moves into the Data warehouse.

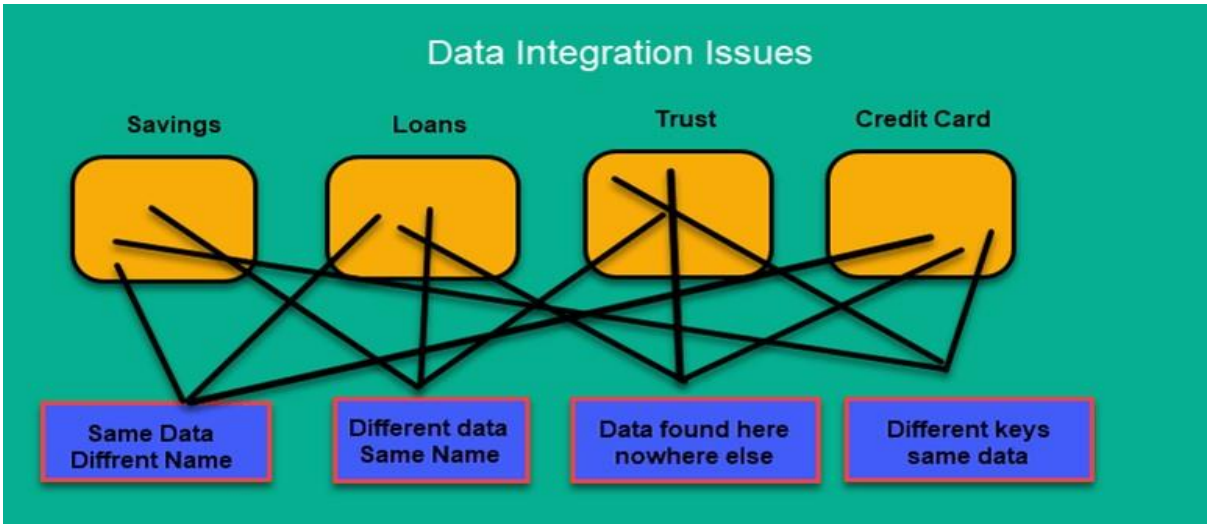
Some validations are done during Extraction: -

- Reconcile records with the source data
- Make sure that no spam/unwanted data loaded
- Data type check
- Remove all types of duplicate/fragmented data
- Check whether all the keys are in place or not

2) Transformation: -

Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful BI reports can be generated.

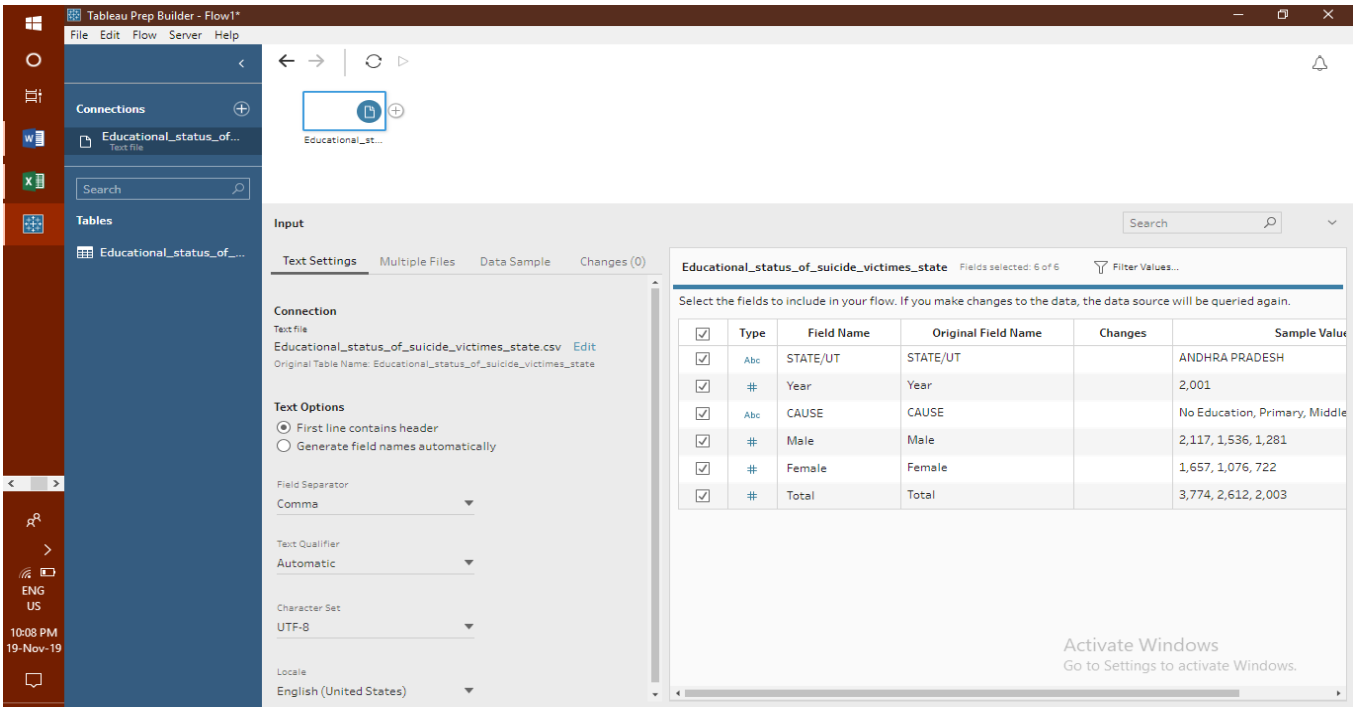




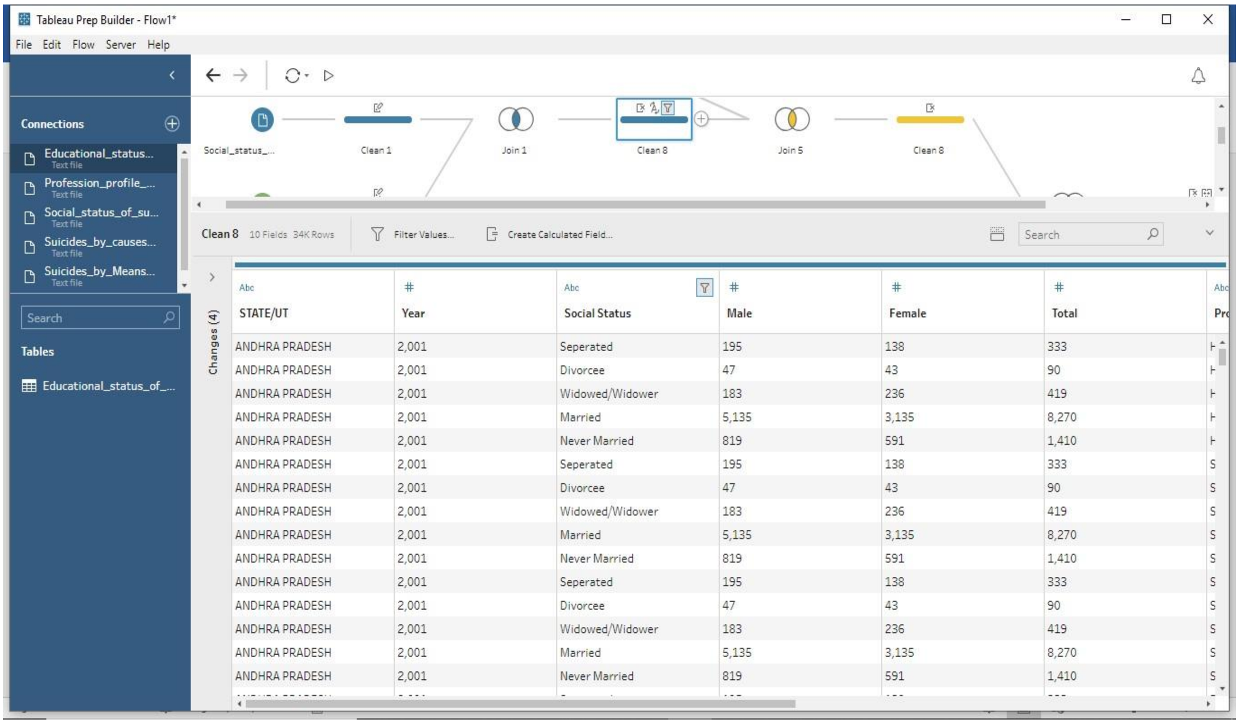
3) Loading: -

Loading data into the target data warehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period (nights). Hence, load process should be optimized for performance.

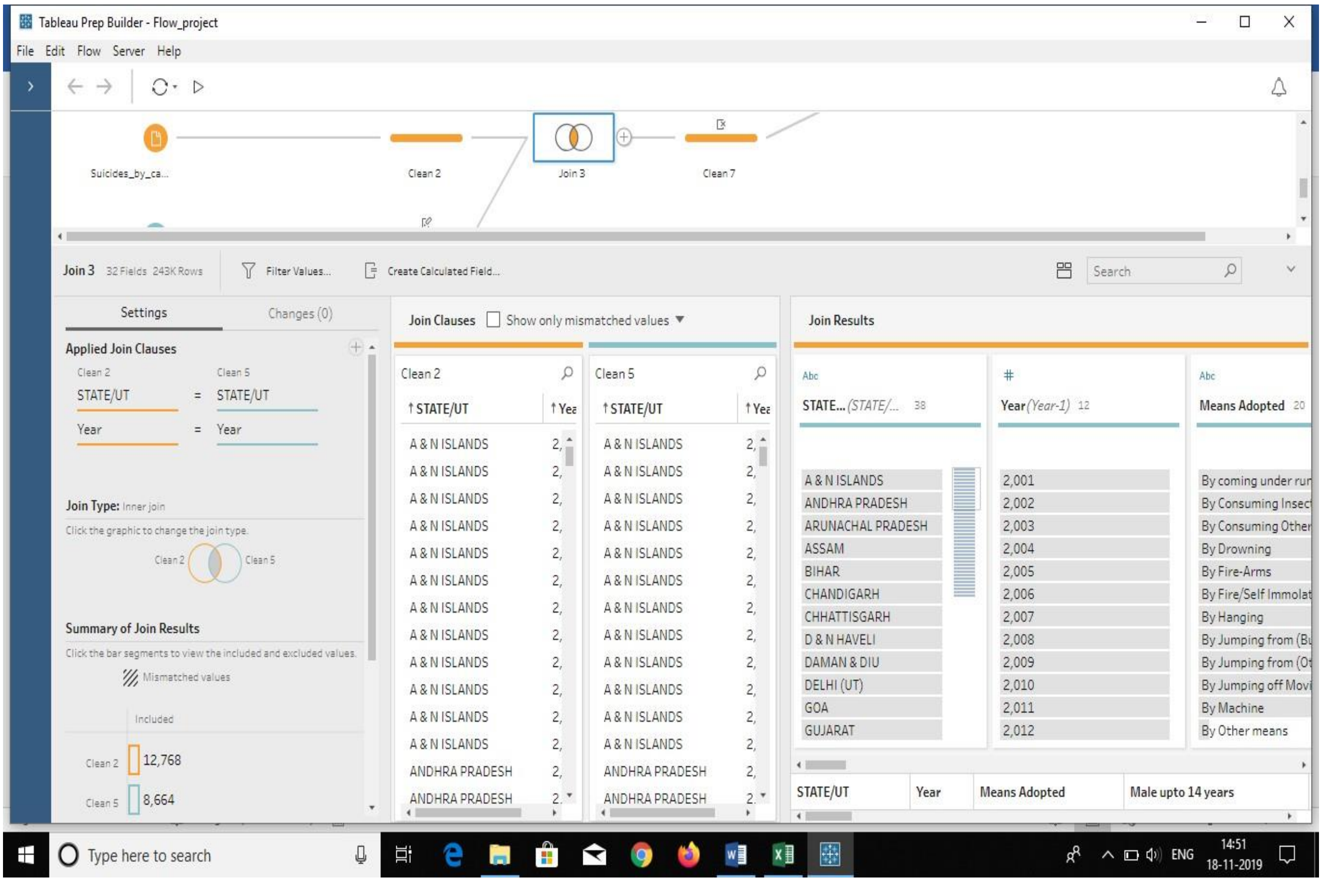
In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.

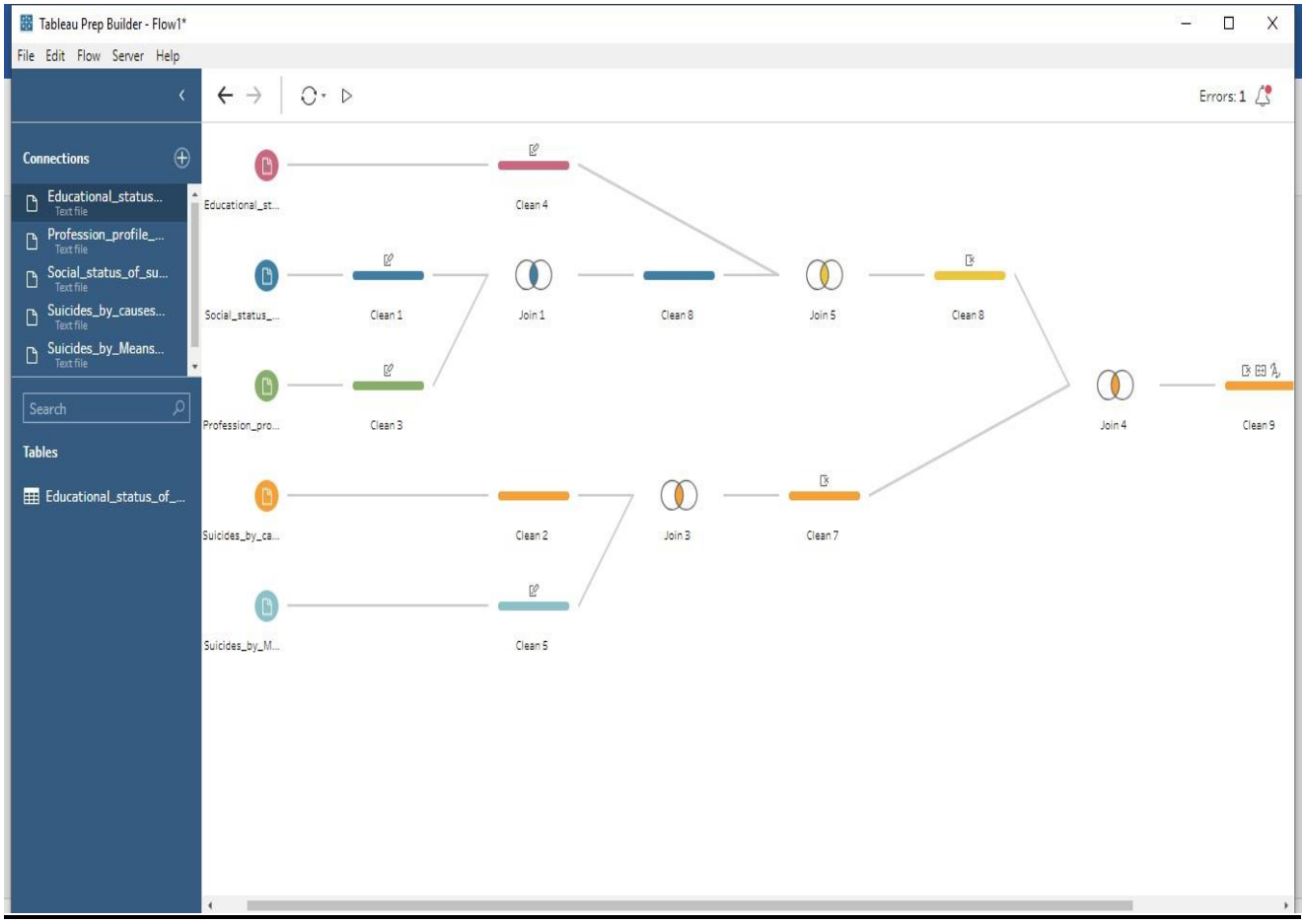


Loading Dataset in Tableau Prep Builder.



ETL process via Tableau Prep Builder where cleaning the data of different tables.

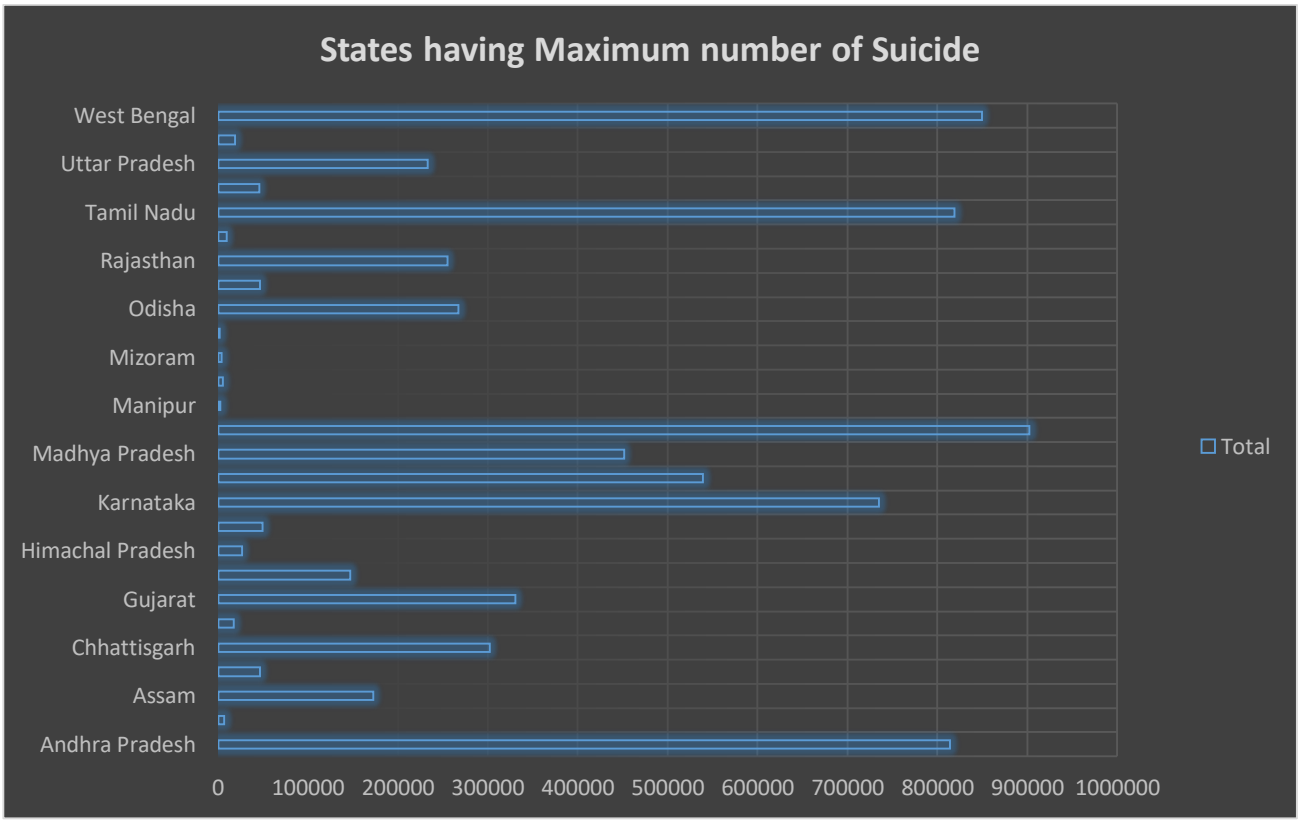




Complete flow of data in Tableau Prep builder.

Analysis on Dataset

1. State wise Suicide data analysis of India



Description:

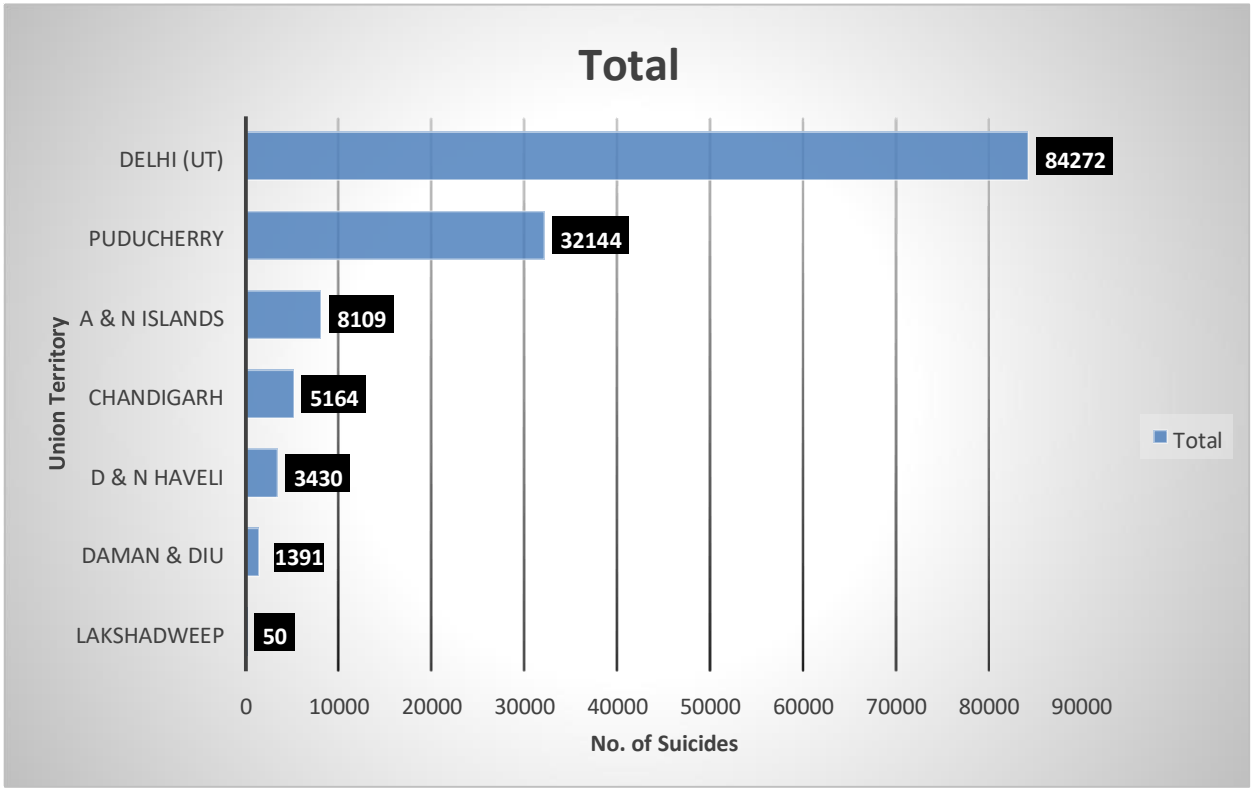
In this we are comparing number of suicide between different states and union territories of India. For that we have made the pivot table from the original dataset. Pivot table contains state in row and total number of suicide in each year. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- Suicide in Maharashtra is highest as compared to other states.
- Lakshadweep has least suicide among others.

2. Union Territory wise Suicide data analysis of India



Description:

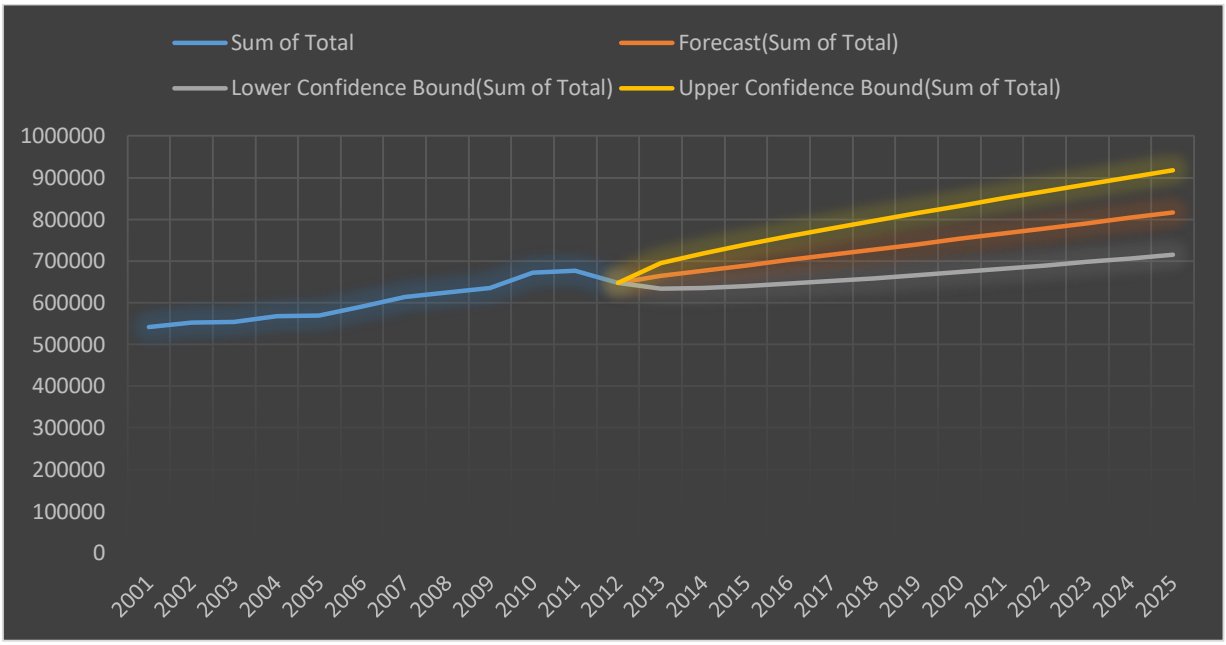
In this we are comparing number of suicide between different union territories of India. For that we have made the pivot table from the original dataset. Pivot table contains union territories in row and total number of suicide in each year. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- Suicide in Delhi is highest as compared to other union territories.
- Lakshadweep has least Suicide among others.

3. Current and Future Predication (till 2030) Suicide data analysis of India



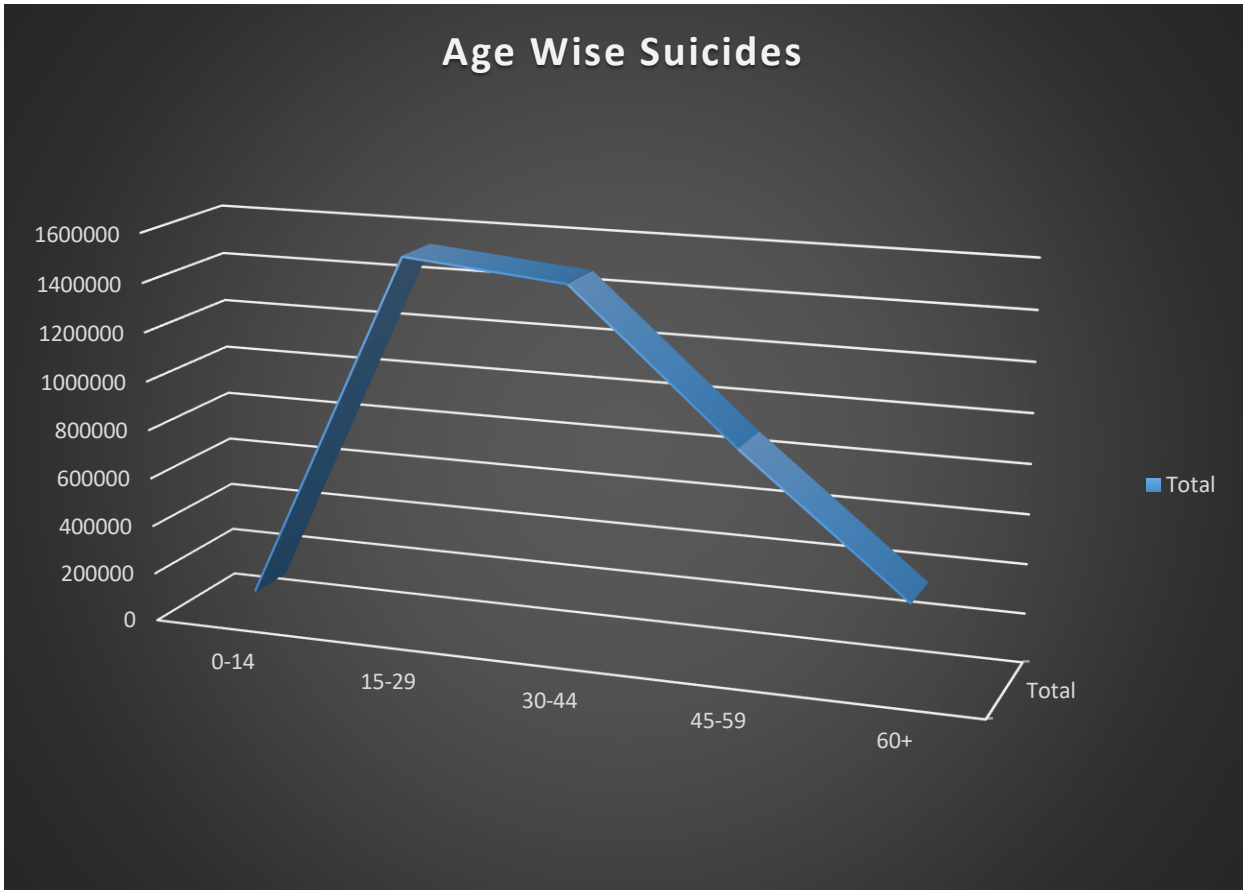
Description:

In this we are making the future predication based on previous data from 2001 to 2012. Pivot table contains year in row area and sum of total number of suicide in each year. In chart area we've added data labels to each value which makes easy for the viewer to see. Formula used in predication =TREND( )

Conclusion:

According to this pivot chart, Suicide is increasing rapidly year by year. It will be at peak in 2030 which is not good.

4. Age group wise Suicide data analysis of India



Description:

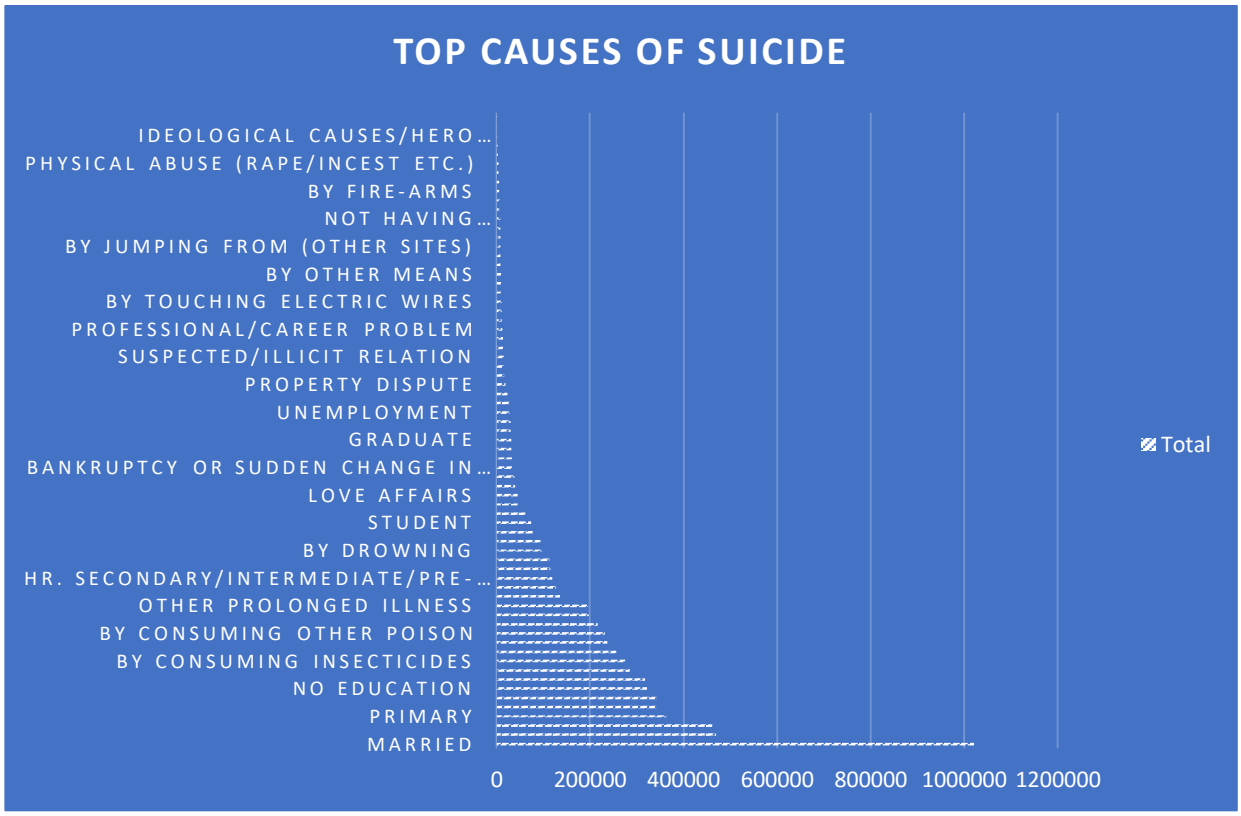
In this we are showing the Suicide of India age group wise by plotting the pie chart. For that we have made the pivot table from the original dataset. Pivot table contains age group in row area, Suicide values as columns. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- Age group between 15 to 29 are committing maximum number of suicides shown in the pie chart.
- Age group between 0 to 14 are committing minimum number of suicides shown in the pie chart.

5. Cause wise Suicide data analysis of India



Description:

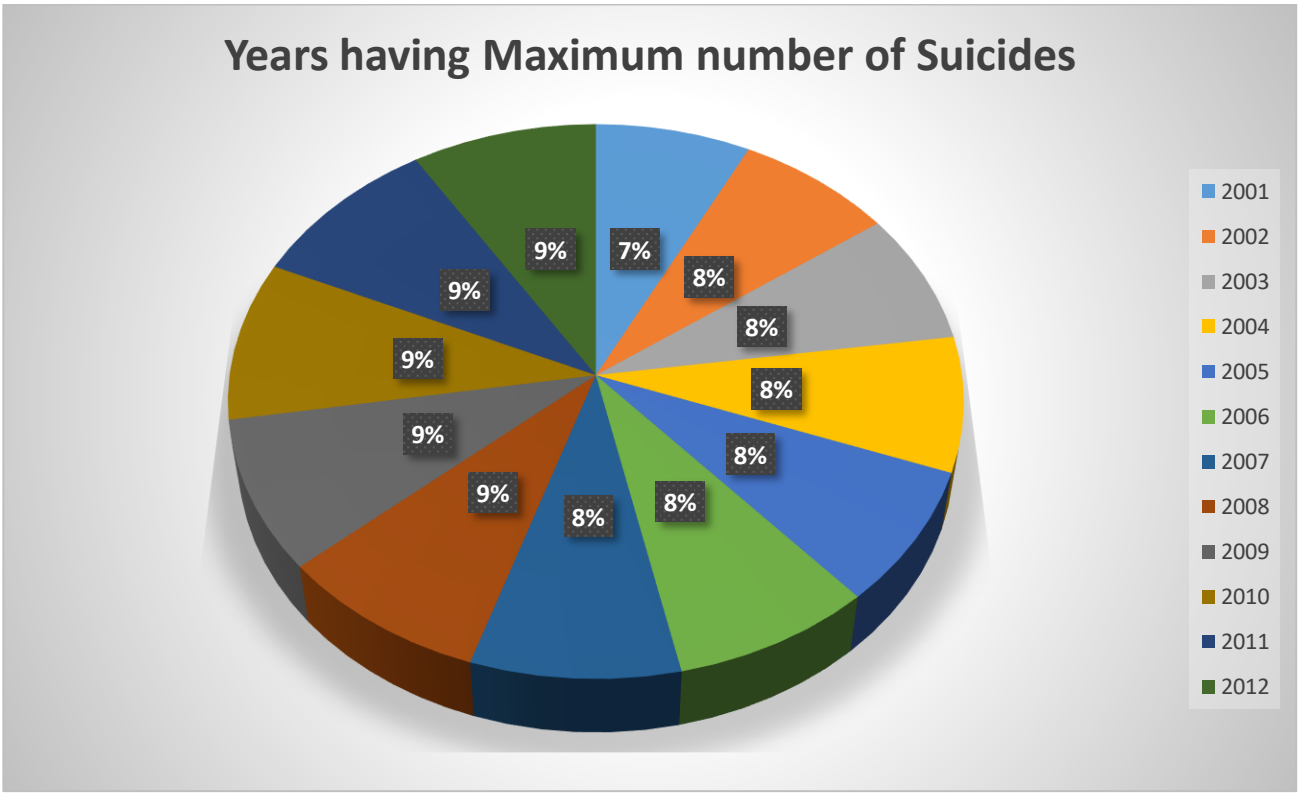
In this we are showing the Suicide of India cause wise by plotting the pie chart. For that we have made the pivot table from the original dataset. Pivot table contains cause in row area, Suicide values as columns. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- Around 3, 50, 0000 people commit suicide due to family problem which is the highest one.
- 766 Suicides are committed due to not having children.
- In around 2, 37,000 cases of suicide cause is not known.

6. Year wise Suicide data analysis of India



Description:

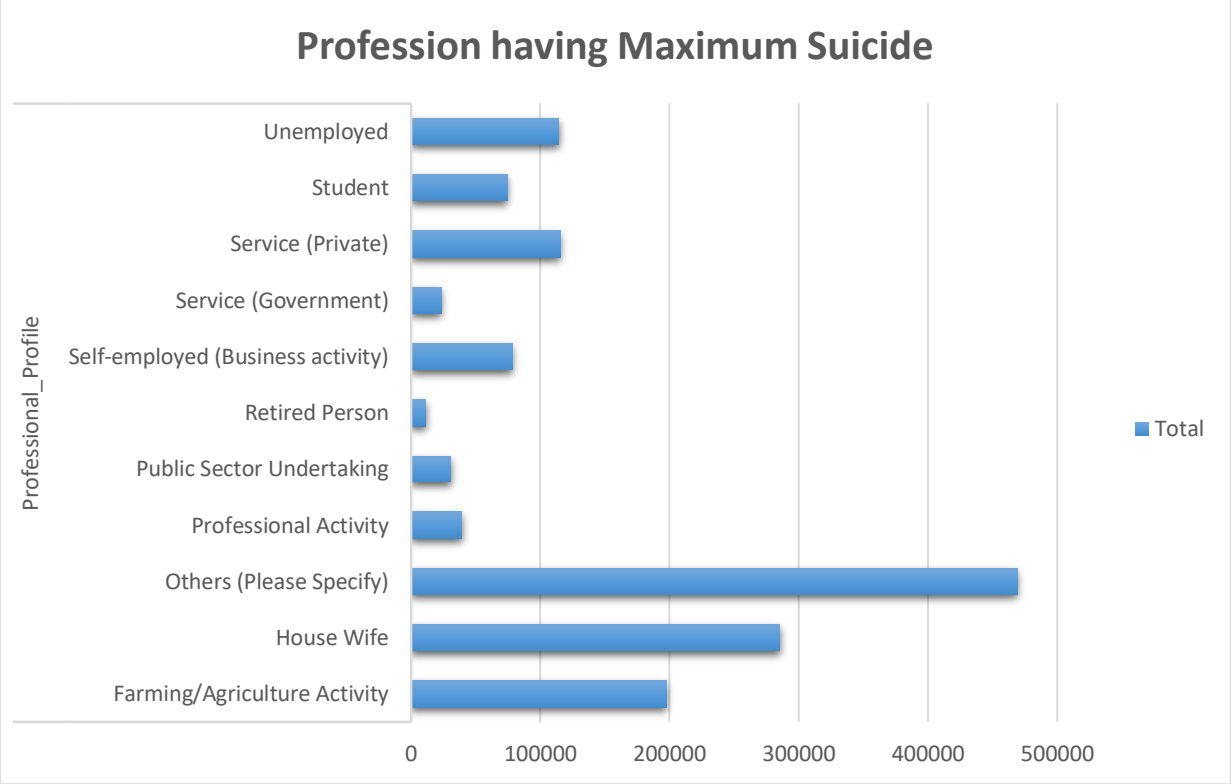
In this we are finding the year having highest Suicide. For that we have made the pivot table from the original dataset. Pivot table contains year in row area and sum of Suicide values as columns. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- Year 2011 is having the maximum number of suicide in India.
- Year 2001 is having the minimum number of suicide in India.

7. Profession wise Suicide data analysis of India



Description:

In this we are finding the top profession having highest Suicide. For that we have made the pivot table from the original dataset. Pivot table contains profession in row area and sum of Suicide values as columns. In chart area we've added data labels to each value which makes easy for the viewer to see.

Conclusion:

According to this pivot chart:

- In most of the suicide cases profession of the people is not known.
- Housewives have the second profession committing maximum number of suicide.
- Least number of suicide is committed by retired person.

Conclusion

Suicide is a very crucial topic that should taken very seriously when intervening in a people's life. People's suicide can be triggered my many things and it is important to be able to recognize the warning signs that can lead to an attempted or successful taking of one's life. People take their own lives for numerous reasons including feeling depressed and hopeless. There are several ways that you can help a people if you feel that they are at risk at attempting suicide: call the physician, remove any harmful items from the house, and talk to them openly about their problems! Teens suicide is the leading in the suicide data of India. Together, hopefully, we can decrease that statistic and save lives! For a country like India, suicide is an extremely worrying situation and it certainly is a national problem which demands immediate solution. Take an example of farmers, The government should run more effective welfare schemes for the poor and landless farmers, some of which may be like crop insurance and providing loans to the farmers on minimal interest rates. If such welfare schemes can be offered immediately and without losing time any further, only then the farmers can be prevented from committing suicides.

Bibliography

- Dataset from www.kaggle.com
- Dataset from www.data.gov.in
- Information from Wikipedia
- Used Tableau Prep
- Used MS-Excel 365 Pro Plus
- Used MS-Word 365 Pro Plus
- Information from Google.com
- Learnt some things from Tutorial Point YouTube channel