

INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2020)

Indian Suicide Data Analysis

Submitted by

Karicharla GNV Swamy Naidu

11811841

Computer Science and Engineering

Section: KM079

Couse Code: INT217

Under the Guidance of

Ashu | 23631

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara



L OVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

CERTIFICATE

This is to certify that **Karicharla GNV Swamy Naidu** bearing Registration no. **11811841** has completed INT217 project titled, “**Indian 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.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer science and engineering

Lovely Professional University

Phagwara, Punjab.

Date: 05/12/2020

DECLARATION

I, **Karicharla GNV Swamy Naidu**, 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: 05/12/2020

Registration No. 11811841

Signature

Karicharla GNV Swamy Naidu

ACKNOWLEDGEMENT

A project work is a combination of views, ideas, suggestions and contribution of many people. Thus, one of the pleasant parts of writing the report is to thank those who have contributed towards its fulfilment.

I consider it as great privilege to have esteemed Lecturer **Ms. Ashu** as my project guide. I take this opportunity to express my sincere gratitude to him through constant advice and constructive criticism nourished my interest in the subject and provided a free and pleasant atmosphere to work against all odd situations. I avail this opportunity to extend my heart full thanks and deep respect to faculty member for their able guidance during this project.

My gratitude to all those, who responded to my questionnaire in a well-defined manner and helped me acquiring knowledge.

I would like to communicate a deep sense of gratitude to all these people without whom my project would not have been such a great learning experience.

Karicharla GNV Swamy Naidu

11811841

KM079

Lovely Professional University

TABLE OF CONTENTS

S.NO	CONTENTS
1	INTRODUCTION
2	SCOPE OF ANALYSIS /OBJECTIVES
3	SOURCE OF DATA SET
4	ETL PROCESS
5	ANALYSIS ON DATASET
6	LIST OF ANALYSIS WITH RESULTS
7	REFERENCES
8	BIBLIOGRAPHY

INTRODUCTION

Suicides are in fact a sensitive subject inherently. Countless great lives and potential is lost every year to the lie that ending your life is the solution to ending pain for yourself and others. Although suicide is a deeply personal and an individual act, suicidal behaviour is determined by a number of individual and social factors

More than one lakh (one hundred thousand) lives are lost every year to suicide in our country. In the last two decades, the suicide rate has increased from 7.9 to 10.3 per 100,000.

This analysis illustrates the quantifiable state of suicides in India between the years of 2001–2012. The data contains in state, Year, Type_code, Type, Gender, Age group, Total suicides.

Data Set

Column name	Column Data
Year	Year of suicide
Type_code	If it is Cause, Means or Profession
Type	Reason or How he adapted suicide or his/her social status
Gender	Gender of the suicide victim
Age_group	The age group of suicide victim
Total Suicides	Total Number of suicides

SCOPE OF ANALYSIS /OBJECTIVES

Suicide is an important issue in the Indian context. Suicide is perceived as a social problem in our country. As population raises every year, the number of suicides also raises, even though suicide of a person deals with his/her personal problems. There are lot more many reasons that community adds boost to ones though of self-assassination without knowing, this could bring up a question which gender/profession has suffered a lot with suicides?

India is a democratic country which has different religions and cultures. Every state in India has its own religion or culture at the same time every individual who are living in state have their own thought process and mindset. Though there is a wide variation in the suicide rates within the country. For example, the southern states of Kerala, Karnataka, Andhra Pradesh and Tamil Nadu have a suicide rate of > 15 while in the Northern States of Punjab, Uttar Pradesh, Bihar and Jammu and Kashmir, the suicide rate is < 3 . The suicide victim may have many problems such as Health, Social status, Lack of support but the variation in cases indicates that atmosphere they live in also have an impact.

Since such vast field of data present of the Indian Suicide and many reasons that we can get from the data there is wide range of scope of the analysis of date. For example:

- Suicides over years based on profession
- Which way victims are adapting the most?
- Which causes/reason leads to more suicides?
- Trendline to show suicides on years for every state?
- Area distribution based on total suicides

Source of Data set

The data was gathered from a public dataset made available through Kaggle

Dataset link : <https://www.kaggle.com/rajanand/suicides-inindia?select=Suicides+in+India+2001-2012.csv>

Author: Rajanand Ilangovan

ETL PROCESS

ETL Stands for Extract transform Load process.

The data is downloaded from Kaggle.com. the data is in formal of comma separated values (CSV)

I extracted the data from Kaggle and changed the format to excel formal and after I cleaned the data

There are there types of data in the given data

1. Causes
2. Means Adapted
3. Profession

I cleaned the data by generalizing data like for age group from 0-12 are kept as children and 51 Causes are generalized with 12 causes based on health, social status etc.

After I load the data into separate sheets and using Power Pivot, I created Relationships for each table and used for analysis

Analysis on data set

1. Suicides over years based on profession

Introduction

This analysis shows which profession's person makes more suicides and which profession is making a smaller number of suicides

General Description

In this analysis, the chart shows the partition of total suicides based on profession in a tree map. tree map is a basic graphical visualization that shows the share of each profession and can be easily compared with other. For the slicer I connected different tables using power pivot and have filters based on Gender, Age Group, Year, and state.

Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Functions

Analysis Result

When we compare Both male and female together, the victim's professions are unidentified, and later there are more housewives who committed suicides and later farmers in third place

Compared in Male, there are more victims whose profession is unknown. After Farmers are the one who committed more suicides

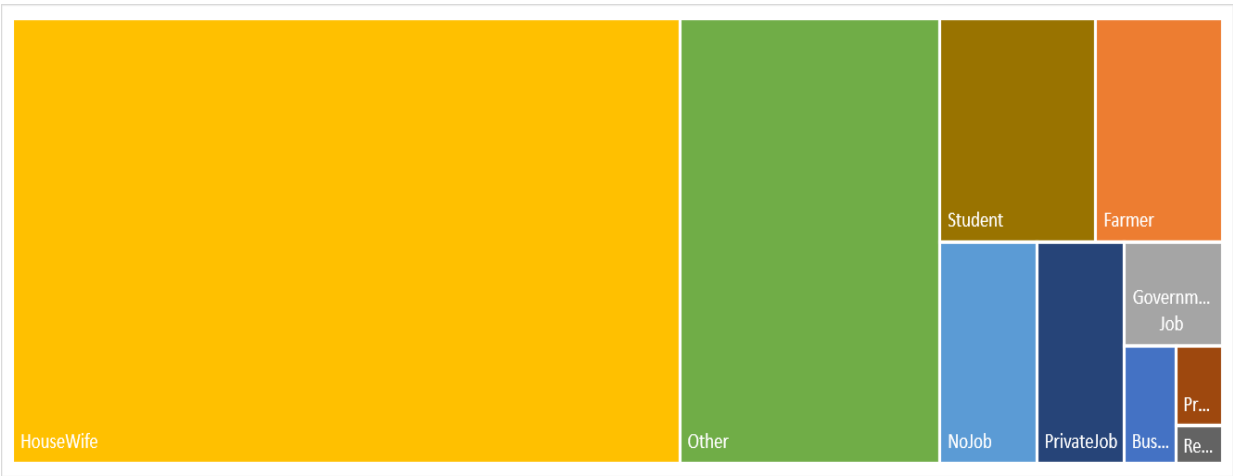
Compared in Female, there are more victims in Housewives, after others.

Visualizations

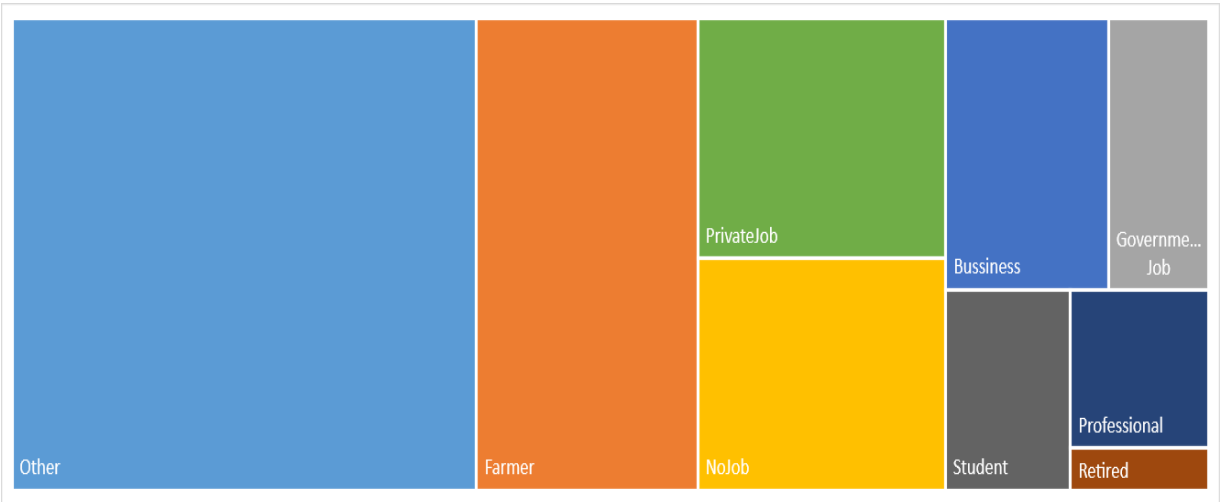
When compared both male and Female



When compared in Female



When compared in male



2. Which way victims are adapting the most?

Introduction

This analysis show how victims are committing suicides for example, whether they are physically harming themselves by hanging, electric or by poison or by unknown reason

General description

In this analysis, the chart shows the trend of total suicides based on means adapted in a line graph. Line graph is a basic graphical visualization that shows the trend over years or anything and can be easily compared with other for rise and loses. For the slicer I connected different tables using power pivot and have filers based on Gender, Age Group, Year, and state.

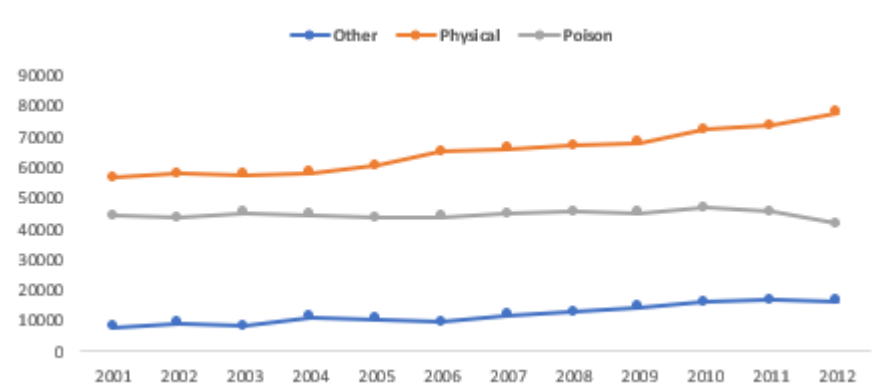
Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Functions

Analysis Result

From this analysis we came to know that major part of victims belongs to physical means for all age groups and genders, and physical harming cases are raised from 2001 to 2012 as like other.

Visualization



3. Which causes/reason leads to more suicides?

Introduction

This analysis shows why people are suicides more and which reason is less likely to cause suicide

General description

In this analysis, the chart shows the bar graph based on count of causes of a suicide. Bar Graph is a graphical visualization that is used to compare between items or anything and even can be used when we want to compare an item over a period of time. For the slicer I connected different tables using power pivot and have filters based on Gender, Age Group, Year, and state

Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Function

Analysis Result

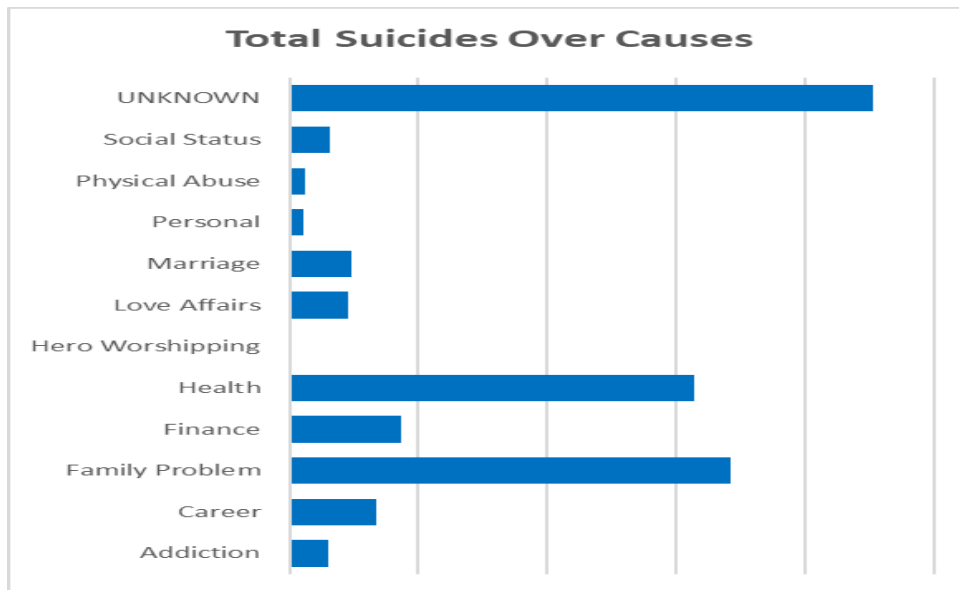
From this analysis we came to know that there are many causes for suicide in which unknown cause occupies the first place whereas family problems, marriage and love affairs comes next and hero-worshipping stands at the last. When it comes to male and female again the cause gets changes,

for female health, marriage and family problems occupies the major part and

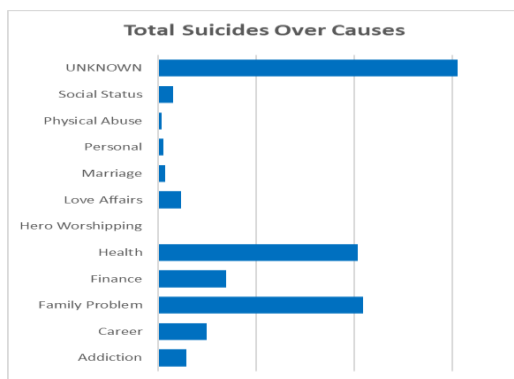
for male health, family problems and social status occupies major part than unknown

Visualization

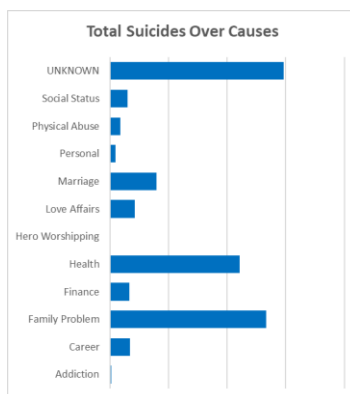
When compared Both Male and Female



When compared to Male



When compared to Female



4. Trendline to show suicides on years for every state

Introduction

This analysis shows the leaderboard of all states based on the total suicides and the line graph for a particular state from 2001 to 2012

General description

In this analysis, this is a leader board build in cells and sparklines to show the state performance in a state individually and can tell whether the suicides increase or decreases for a particular state. For the slicer I connected different tables using power pivot and have filters based on Gender, Age Group, Year, and state

Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Function

Analysis Result

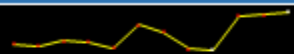




From this analysis we came to know that Maharashtra stands at the first place in India with total 180389 cases for both male and female and this graph is being increased from 2001 to 2012 whereas Tamilnadu with 163813 cases occupies the second place next comes Andhra Pradesh with 162820 cases and next West Bengal with 161030 cases and Karnataka with 146965 cases.

When it comes to female west bengal occupies the first place with 69299 cases and its graph being decreased from 2001 to 2012, second place is occupied by tamilnadu with 61324 cases and next comes Maharashtra with 58635 cases.

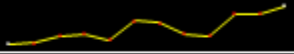




In male suicide cases Maharashtra stands at the first place with total 121754 cases, its graph being increased from 2001 to 2012 and Andhra Pradesh stands at second place with 108424 cases, Tamil Nadu at 3rd, Karnataka at 4th place and west Bengal stands at 5th place

Visualization

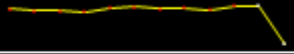




When both male and Female are compared

Rank	State	Total Suicides	SparkLines
1	Maharashtra	180389	
2	Tamil Nadu	163813	
3	Andhra Pradesh	162820	
4	West Bengal	161030	
5	Karnataka	146965	

When Male is compared

Rank	State	Total Suicides	SparkLines
1	Maharashtra	121754	
2	Andhra Pradesh	108424	
3	Tamil Nadu	102489	
4	Karnataka	98391	
5	West Bengal	91731	

When Female is compared

Rank	State	Total Suicides	SparkLines
1	West Bengal	69299	
2	Tamil Nadu	61324	
3	Maharashtra	58635	
4	Andhra Pradesh	54396	
5	Karnataka	48574	

5. Trendline to show suicides on years for every state

Introduction

This analysis shows the distribution of suicides over states from this we can know that which state has highest and lowest of suicides over states.

General description

In this analysis, the total no of suicides is represented graphically on the geo-graphical map. This type of graph is used when we have to represent numbers for a specific Geo location. For the slicer I connected different tables using power pivot and have filters based on Gender, Age Group, Year, and state

Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Function
6. Bing Map

Analysis Result

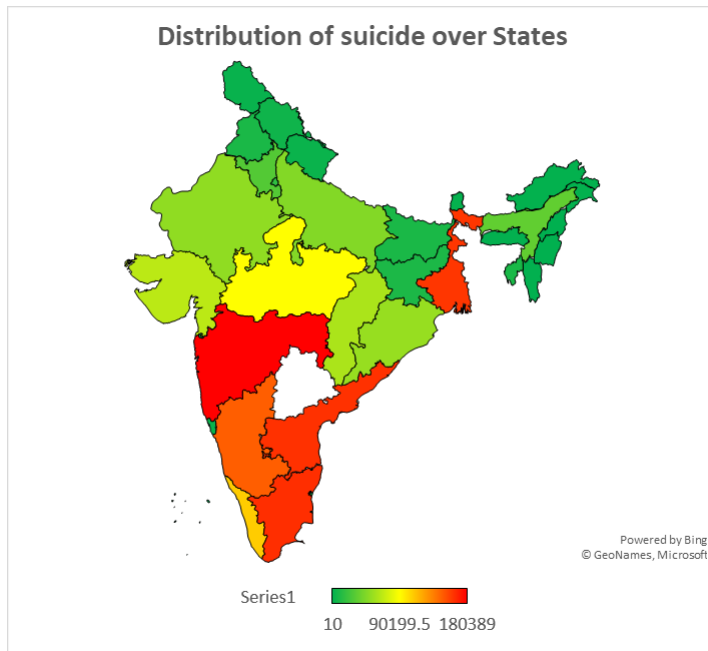
When compared both male and Female, the southern states like Maharashtra, Andhra Pradesh, Tamil Nadu, West **Bengal** have more suicides. And northern states like Jammu, Himachal Pradesh, Punjab Has Least.

When compared in children, West Bengal has Higher suicides and in Young Aged People, Maharashtra tops the rank.

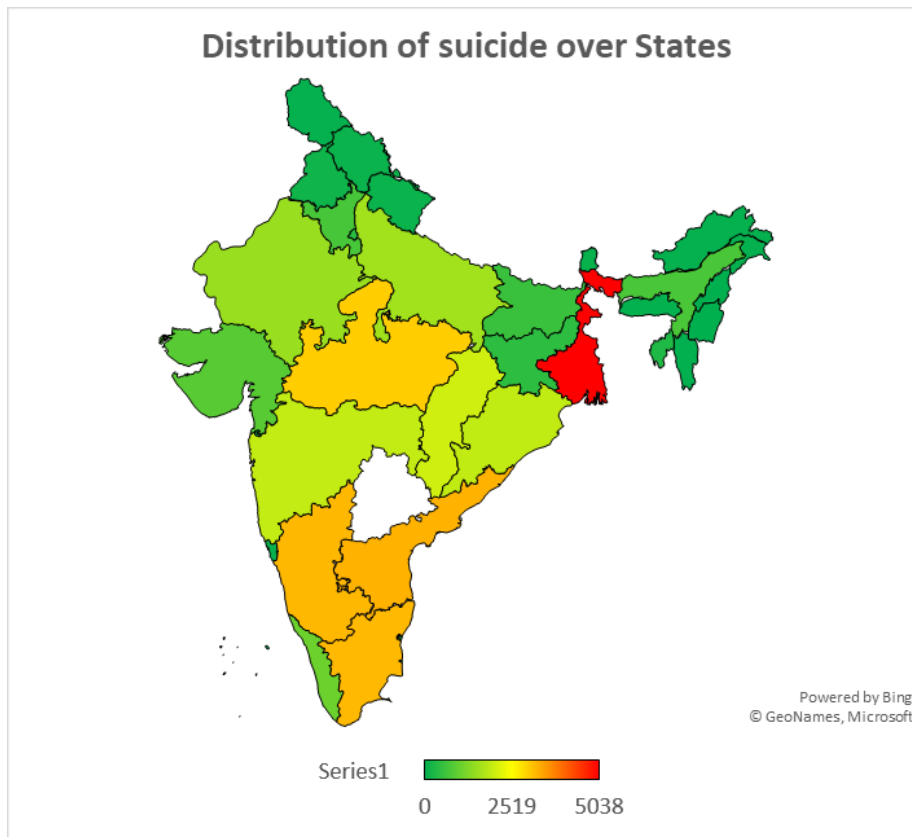
The senior aged and older people the figures have not much change as compared to all age groups

Visualizations

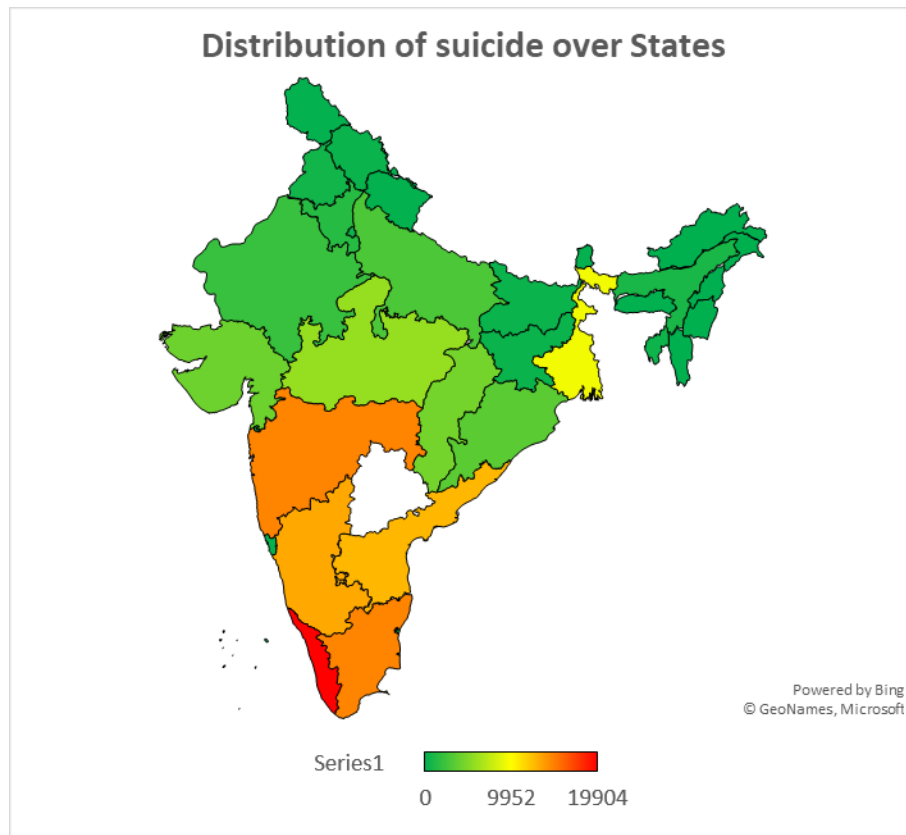
When compared in all aged groups



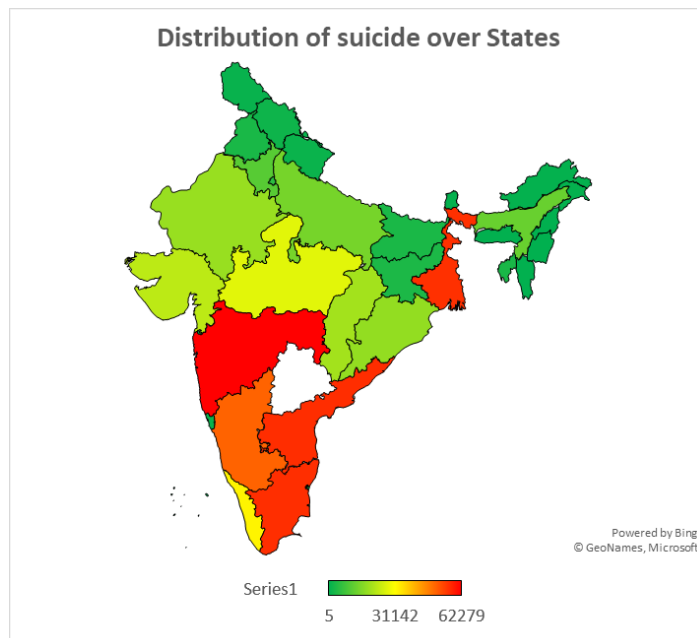
For children



For Senior old aged



For middle aged



6. Trend over genders over years?

Introduction

This analysis shows the trend over years for each male and female separated each other from 2001 to 2012

General description

In this analysis, the chart shows the bar graph based on count of causes of a suicide for each gender. Bar Graph is a graphical visualization that is used to compare between items or anything and even can be used when we want to compare an item over a period of time. For the slicer I connected different tables using power pivot and have filters based on Gender, Age Group, Year, and state

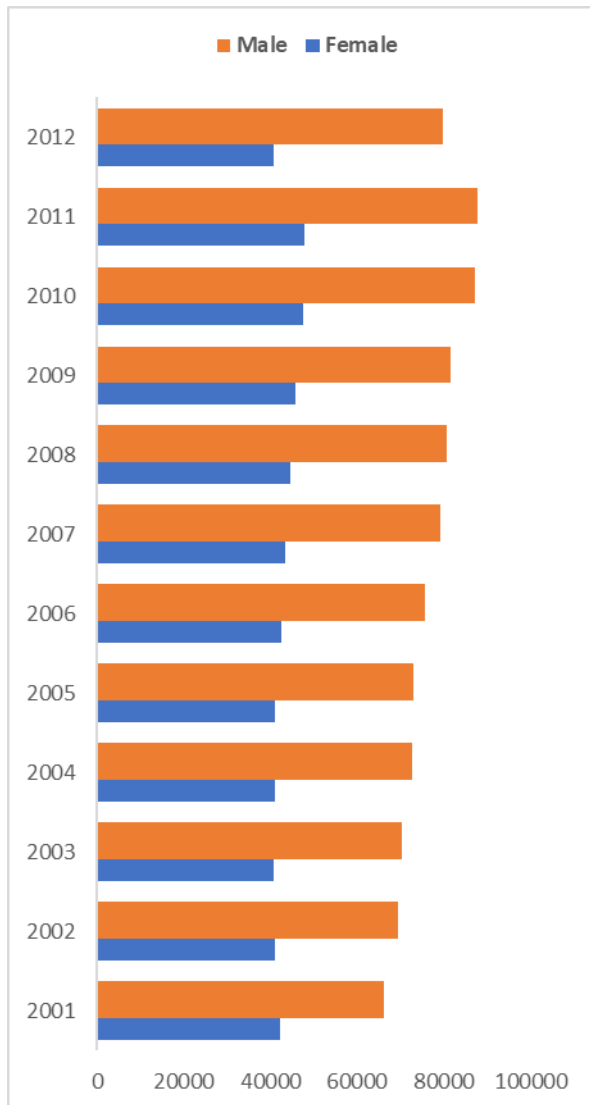
Specific Requirements

1. Pivot tables
2. Pivot charts
3. Slicers
4. Power Pivot
5. Lookup Function

Analysis Result

From this analysis, we can clearly state that male has 2 times more suicides than female suicides and the suicide cases slightly increases from 2001 to 2012.

Visualization



List of Analysis with results

- Nearly Male victims of suicides are 2/3 of the total suicides
- 25% of the suicides are committed by people of Young aged (15-29) from the states Maharashtra, Madhya Pradesh, West Bengal, Gujrat, Andhra Pradesh, Tamil Nadu, Karnataka.
- Most people are adapting the suicide by poison compared to all separately and physical when compared all physical harms
- Housewives are committing suicides in Female and number of suicides in children are as students
- More number of suicides occurred in 2011 in between 2001 to 2012
- Health, Career, and Family Problems are the top three reasons for committing suicide
- West Bengal has More children suicides
- Retired people are less Likely to cause Suicide
- Maharashtra has high number of suicides as Maharashtra has larger in area than others

References

1. www.kaggle.com
2. www.youtube.com
3. www.stackoverflow.com

Bibliography

1. Excel 2016 Bible: The Comprehensive Tutorial Resource by John Walkenbach, Wiley