Good health and well being

Prepared By

Vardhana Sharma (2020BTechCSE081)

Project on

Of

Computational Data Analysis

Institute of Engineering and Technology

JK Lakshmipat University, Jaipur

February 2021

INDEX

Contents | Page No.

ABSTRACT

Recently, the United Nations adopted 17 sustainable development goals for 2030.The sustainable development goal (SDG) 3 “Ensuring a healthy life and promoting well-being for all ages” is one of the most intersecting goals, with other sustainable development goals. In this paper we are analyzing the barriers for this goal to achieve. Obseity, Alcohol consumption, how does variation in climate affects our concentration power, our health. How is our immune system to fight with any epidemic like covid-19.

Due to anxiety, hypertension, sugar, low-income, poor sanitization, poor quality of wheat, rice, flour ingredients, our mind didn’t work properly as it should works in ordinary conditions, a frequent number of suicides, crimes, robbery taking place, resulting increment in death rate, due to lots of stress while driving vehicles road accidents are often occurring.

INTRODUCTION

Health is a state of body, wellness is a state of being. For the development of all United Nations adopted sustainable development goals in 2015. The sustainable development are collection of 17 interlinked global goals, sustainable development goal 3 regarding “Good Health and Well Being” is one of the 17 sustainable development goals. The official wording is “To ensure healthy lives and promote wellbeing for all at all ages. SDG 3 has 13 targets and 28 indicators to measure progress towards targets. Those are reduction of maternal mortality, communicable disease, non-communicable

Disease, mental health, substance abuse etc. SDG 3 aims to achieve universal Health coverage, that seeks equity and equitable access of healthcare services to all men’s and womens.it focuses on broader economic and social inequalities etc. Due to the unprecedented situation of global pandemic COVID-19, there is a need to give significant attention towards the realization, of good health and wellbeing on global scale.

India has made some progress in reducing mortality rate ,which declined from 125 per 1000 live births in 1990-91 to 50 per 1000 live births in 2015-16.According to the magnitude of substance abuse(21019) survey, about 2.1% of the country's population uses opioids these psychoactive substances causes significant health and social problems for the people who use them. For the prevention of those people who use these substances government take significant initiative by sustainable development goals for good health and welfare of the society.

SDG 3 “Ensure healthy lives and promote wellbeing for all at all ages”

The goals within a goal: Health targets for SDG 3

3.1 By 2030, reduce the global maternal mortality ratio to less than 70 per 100 000 live births.

3.2 By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1000 live births and under-5 mortality to at least as low as 25 per 1000 live births.

3.3 By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases.

3.4 By 2030, reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being.

3.5 Strengthen the prevention and treatment of substance abuse, including narcotic drug abuse and harmful use of alcohol.

3.6 By 2020, halve the number of global deaths and injuries from road traffic accidents.

3.7 By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programs.

3.8 Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all.

3.9 By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water and soil pollution and contamination.

OBJECTIVE

1) To analyze road accidents in India.
2) To compare infant mortality rate with life expectancy from year 2011 to 2018.
3) To compare alcohol consumption in urban and rural areas.
4) To study lifestyle of different weight categories.

Literature review

Obesity can be defined as a condition of abnormal or excess fat accumulation in adipose tissue, to the extent that health may be impaired. Body Mass Index (BMI), which is calculated as [((weight in kg) / (height in m)) ^ 2], is considered to be the most useful population level measure of obesity, and it is a simple index to classify underweight, overweight and obesity in adults. The WHO has classified overweight obesity in adults based on various BMI cutoffs [5].
For instance, globally, there are more than 1 billion overweight adults, at least 300 million of whom are obese [1,2]. The prevalence of excess weight (overweight and obesity) among United Kingdom (UK) adults in 1996 was 59% in men and 49% in women while two-thirds of adults in the United States (US) in the year 2000 were overweight and 30.5% were obese [6]. In Iraq there is no national figure available, but one study carried out in Baghdad in 1997 showed that the prevalence of obesity was 23.16% in women aged 25 years old and above [7].

Hypothyroidism is a condition in which the thyroid gland doesn’t produce enough thyroid hormone. Hypothroidism’s deficiency of thyroid hormones can disrupt heart rate, body temperature and all aspects of metabolism. It is most occurring in older women.
Hyperthyroidism is condition when there is overproduction of a hormone by the butterfly-shaped gland in the neck(thyroid).due to this metabolism increases.

Recently, it was found statistically that there is positive association between specific behaviors, like-number of meals per day, eating fatty and fast food with obesity and overweight implementation of nutritional diet, and various weight-control strategies are highly recommended for these category of people.

Increasing heat exposure in many workplaces is posing a lot of challenges for working population. Sectors like agriculture, manufacturing, fisheries, forestry, SMEs, construction, some service sectors etc. are affected due to the climatic changes, [2] [3]. The outdoor workers and indoor workers without efficient cooling systems are exposed to heat waves. The continuous exposure to heat affects the employees physically and mentally. The work capacity for many working people is affected.

The workers with some previous heat illness were ready to adjust work habits to adapt themselves to increasing hot weather. Several studies have been undertaken to understand the impact of climate change on working population. Climate change has and will continue to exacerbate workplace heat as highlighted in the IPCC assessment [4] Workers use self-pacing mechanisms i.e. adjusting work rate to avoid physiological heat strain to keep themselves away from heat stress when working in hot environments. [5] [6]. The slowing down of work as a defense mechanism during severe heat exposure is labelled 'autonomous adaptation' by climate change researchers, [11]. Exposure to extreme heat conditions has been found to be hazardous to health, [11]. Workers frequently exposed to heat in their workplace have been found to suffer heat exhaustion, heat stroke, kidney disease, heart or lung disease, accidents, and injuries. [11] The body heat balance is determined by factors such as air temperature, radiant temperature, humidity, air movement (wind speed); clothing and the metabolic heat generated by human physical activity, [11]. It is well known that physical work creates heat inside the body and that this affects occupational health and performance when combined with excessive workplace heat [11]. People performing physical outdoor labor are exposed to higher temperatures combined with body heat generated by the jobs themselves, [11].

Road accidents in India - Overview

- Road traffic fatalities and injuries are, to a great extent, preventable, since the risk of incurring injury in an accident is largely predictable and many countermeasures, proven to be effective, exist. The most effective way to reduce fatalities and injuries would be through an integrated approach involving close collaboration of many sectors. Progress is being made in many parts of the world where multisectoral strategic plans are leading to incremental reductions in the number of road accidental fatalities and injuries. Such strategies focus on four key factors that contribute to the risk of occurrence of a road accident – exposure, behavioral factors, road environment, and vehicle factors.[1]

- Rapid growth of population along with numerous economic activities have acted as the major catalyst for the steady growth of the motor vehicle population and has subsequently resulted in the increase of road accidents.[2]
- "*Accidents, and any illnesses, may be seen as a legitimized form of reaction to stress in which the individual through no mindful fault of his own is forced to retreat from his ordinary social roles. Both accidents and illness differ from deliberately self-inflicted injuries or feigned illness in this respect*" [3]

Road Safety—A Public Health Issue

World's first RTA (Road Transport Accident) is supposed to have occurred in 1896. Everybody concerned at that time reported to have said, "this should never happen again." But more than a century later, 1.2 million people were killed on roads every year and up to 50 million more are injured. For every one killed, injured, or disabled by RTA, there are countless others deeply affected by the cost of prolonged medical care, loss of a family bread winner, or the extra funds needed to care for the people with disabilities. RTA survivors, their families, friends, and other care givers often suffer adverse social, physical, and psychological effects. If the current trends continue, the number of people killed and injured on the world's roads will rise by more than 60% by 2020.[4]

Infant Mortality –

Infant mortality rate is the probability of a child born in a specific year or period dying before reaching the age of one, if subject to age-specific mortality rates of that period. Infant mortality rate is strictly speaking not a rate (i.e. the number of deaths divided by the number of population at risk during a certain period of time) but a probability of death derived from a life table and expressed as rate per 1000 live births.[8]

An infant mortality rate (or IMR) is considered a primary and important indicator of a geographic area's (country, state, county) overall health status or quality of life.

• There are some concerns about the quality of reporting of infant mortality internationally and within states, especially in terms of defining a live birth and/or complete reporting of both birth and death certificates for very low birth weight babies.

. • The IMR is usually calculated using the annual number of resident infants who died during a year in the numerator and the total annual number of resident live births during the same year in the denominator.

Life Expectancy –

Life expectancy is the key metric for assessing population health. Broader than the narrow metric of the infant and child mortality, which focus solely at mortality at a young age, life expectancy captures the mortality along the entire life course. It tells us the average age of death in a population.[9]

Life expectancy in India was 25.4 in the year 1800, and over the course of the next 220 years, it has increased to almost 70. Between 1800 and 1920, life expectancy in India remained in the mid to low twenties, with the largest declines coming in the 1870s and 1910s; this was because of the Great Famine of 1876-1878, and the Spanish Flu Pandemic of 1918-

1919, both of which were responsible for the deaths of up to six and seventeen million Indians respectively; as well as the presence of other endemic diseases in the region, such as smallpox. From 1920 onwards, India's life expectancy has consistently increased, but it is still below the global average.[10]

people use of consumption of alcohol in India and what are the negative affect of the consumption of alcohol on the body and how it is available to college students easily which can ruin their life but there are many ways of preventing exercise of alcohol use.

According to sources in India alcohol consumption in India amount to at least 5.4 billion litters in 2016 and was estimated to reach about 6.5 billion litters by 2020. According to recent data published by the WHO, the total per capita consumption of alcohol by individual alone at 15 years of age is 6.2 L of pure alcohol per year which equals to 13.5 kg of pure alcohol per day, there is a little variation between the who region and member states. Nearly 6.1% of the global surrounding of decease is attributed to alcohol consumption and its causes nearly 1.3 million death every year.

There are many long tiring effects of alcohol consumption such as High blood pressure, heart disease, stroke, liver disease, liver cancer, mouth cancer and digestive problems.

According to drink aware, drinking alcohol as a teenager can affect memory function reaction, learning ability attention spam etc. this can create a huge menace for future generations but there are many preventive initiatives taken by the government to protect the society from this hazard like leadership, awareness and commitment, health services response, community action drink driving policies and counter measures, monitoring and surveillance etc.

Methodology

A statistical hypothesis is considered as an assumption made about a population which may or may not be true. Hypothesis testing is a set of formal procedures used to either accept or reject statistical hypotheses. Statistical hypotheses are of two types:

- **Null hypothesis,** H0 - represents a hypothesis of chance basis.
- **Alternative hypothesis,** Ha - represents a hypothesis of observations which are influenced by some non-random cause.

Hypothesis Tests:

The process used to determine whether to reject a null hypothesis, based on sample data is known as hypothesis testing.

Steps involved –

- Building Hypothesis
- Formulate an analysis plan
- Analyze sample population/data
- Interpret results

T–Test concerning difference of means (Two Populations):

- A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.
- The t-test is used for hypothesis testing purposein statistics.
- Calculating a t-test requires three key data values. They include the difference between the mean values from each data set

(called the mean difference), the standard deviation of each group, and the number of data values of each group.

Anova:

- Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests.
- A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.
- If ANOVA's F-ratio equal close to 1, then no true variance exists between the groups.

Eigen values and Eigen vectors:

- Almost all vectors change direction, when some operations(A) are applied on them. Certain exceptional vectors (v) are in the same direction as Av. Multiply an eigenvector by A, and the vector Ax is a number lambda times the original x.
- The eigenvalue lambda tells whether the special vector x is stretched or shrunk or reversed or left unchanged, when multiplied by A.
- we have apply the concept of eigen values and vectors to determine which types of food is mostly liked by which category of age in this survey.
- How we have find the elements of matrix
- We have use the mode of scale chosen by the users for that particular food.
- Method to find rank :
- Find out the largest dominant eigenvalue and their respective eigen vector

- Each element of vector represents the age category, type of food

Correlation –

- Correlation refers to some statistical relationships involving dependence between two data sets.
- Correlation values range between -1 and 1.
- There are two key components of a correlation value:
- Magnitude – The larger the magnitude (closer to 1 or -1), the stronger the correlation
- Sign – If negative, there is an inverse correlation. If positive, there is a regular correlation.

RESULTS AND DISCUSSION

OBJECTIVE 1 –

To analyze road accidents and person killed (in accidents) in India.

1.1) To compare total number road accidents in India(state wise) in different years.

Applying Anova (One-Way)

- Step 1 - Building Hypothesis
  Null Hypothesis H0: µ1 = µ2 = µ3
  Alternative Hypothesis: µ1 ≠ µ2 ≠ µ3
  Here, µ1 = road accident in year 2015, µ2 = road accident in year 2017, µ3 =road accident in year 2019
- Step 2 – Test Statistic
  Level of significance = 5% (0.05)

$$F = \frac{MS\ (Tr)}{MSE}$$

Treatment Mean Squaure $MS\ (Tr) = \frac{SS\ (Tr)}{k-1}$

Error Mean Squaure $MSE = \frac{SSE}{k\ (n-1)}$

Total sum of squares $SST = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}^2 - \frac{1}{N} T_{..}^2$

Treatment sum of squares $SS(Tr) = \frac{1}{n} \sum_{i=1}^{k} T_{i\cdot}^2 - \frac{1}{N} T_{..}^2$

Error sum of squares $SSE = SST - SS(Tr)$

Where $T_{i\cdot}$ is the total of i[th] row and $T_{..}$ is the grand total

- Step 3 - Criteria of Rejection

  Reject the null hypothesis if $F > F_{\alpha,k-1,N-k}$

- Step 4 – Calculation

Source of Variation	Degrees of freedom	Sum of Squares (SS)	Mean Square (MS)	F (Test Statistic)
Treatments	k - 1	SS (Tr)	MS (Tr)	$\frac{MS\,(Tr)}{MSE}$
Errors	N - k	SSE	MSE	
Total	N - 1	SST		

- Step 5 – Decision

  Fvalue = 0.0674, pvalue = 0.935
  Since the p-value is not less than 0.05(level of significance), we fail to reject the null hypothesis.
  Therefore, Null Hypothesis accepted.

  Here, we can conclude that there is no significant difference between number of road accidents in different years this may be due to increase in violation of traffic rule and under age driving.

PYTHON IMPLEMENTATION-

In [66]:
```
import pandas as pd
import matplotlib.pyplot as plt
road_accident = pd.read_csv("Total road acciedent in india(2015,17,19).csv", index_col="S.No.")
```

In [67]:
```
road_accident
```

Out[67]:

S.No.	States	2015	2017	2019
1	Andhra\nPradesh	24258	25727	21992
2	Arunachal\nPradesh	284	241	237
3	Assam	6959	7170	8350
4	Bihar	9555	8855	10007
5	Chhattisgarh	14446	13563	13899
6	Goa	4338	3917	3440
7	Gujarat	23183	19081	17046
8	Haryana	11174	11258	10944
9	Himachal Pradesh	3010	3114	2873
10	Jammu & Kashmir	5836	5624	5796

In [69]:
```
#naming coulomns of dataframes
coloumn_2015 = road_accident['2015']
coloumn_2017 = road_accident['2017']
coloumn_2019 = road_accident['2019']
```

In [70]:
```
#importing library
from scipy.stats import f_oneway

#perform one-way ANOVA
f_oneway(coloumn_2015, coloumn_2017, coloumn_2019)
```

Out[70]: F_onewayResult(statistic=0.06744279464116898, pvalue=0.9348317584544216)

In [72]:
```
pvalue = 0.935
if pvalue<0.05:
    print("Null hypothesis accepted")
else:
    print("Null hypothesis rejected")
```

Null hypothesis rejected

1.2) To check if the number of total person killed in road accident from year 2015 to 2019 decreased or not.

Applying T-test (difference of mean)

- Step-1 Building hypothesis

  Null Hypothesis H0: µ1 = µ2

  Alternative Hypothesis:µ1 ≠ µ2 (Two tailed)

  µ1 is the mean of the total number of person killed in road accident in 2015.

  µ2 is the mean of the total number of person killed in road accident in 2019.

  Level of significance – α = 0.05,

  Sample size -n1 = 29, n2 = 29

  Degree of freedom = n - 1= 28

- Step-2 Test statistics

  Here population is known so,

  Case 2: If population variance in unknown and n < 30: $t = \frac{\overline{x_1} - \overline{x_2} - \delta}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$,

  Where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$; s_1^2 and s_2^2 are sample variances of sample 1 and 2 respectively.

  Here, $s_p^2 = \frac{s_1^2+s_2^2}{2}$ if $n_1 = n_2$

- Step-3 Criteria of Rejection

; $|t| > t_{\alpha,n_1+n_2-2}$ (Two Tailed)

- Step-4 Calculation

  Pvalue = 0.8934102981029542

- Step-5 Decision

Here, the p-value of our t-test is greater than (level of significance) alpha = 0.05, we fail to reject the null hypothesis of the test. So we can conclude that there is no major difference and no decrement in number of deaths from 2015 to 2019.

PYTHON IMPLEMENTATION-

```
In [21]:
import pandas as pd
accident_deaths = pd.read_csv("Road accident deaths.csv", index_col="S.No.")
```

```
In [4]:
accident_deaths
```

Out[4]:

S.No.	States	2015	2016	2017	2018	2019
1	Andhra Pradesh	8297	8541	8060	7556	7984
2	Arunachal Pradesh	127	149	110	175	127
3	Assam	2397	2572	2783	2966	3208
4	Bihar	5421	4901	5554	6729	7205
5	Chhattisgarh	4082	3908	4136	4592	5003
6	Goa	311	336	328	262	297
7	Gujarat	8119	8136	7289	7996	7390
8	Haryana	4879	5024	5120	5118	5057
9	Himachal Pradesh	1096	1271	1203	1208	1146
10	Jammu & Kashmir	917	958	926	984	996
11	Jharkhand	2893	3027	3256	3542	3801
12	Karnataka	10856	11133	10609	10990	10958
13	Kerala	4196	4287	4131	4303	4440
14	Madhya Pradesh	9314	9646	10177	10706	11249
15	Maharashtra	13212	12935	12264	13261	12788
16	Manipur	139	81	136	134	156
17	Meghalaya	183	150	182	182	179
18	Mizoram	72	70	60	45	48

19	Nagaland	30	46	41	39	26
20	Odisha	4303	4463	4790	5315	5333
21	Punjab	4893	5077	4463	4740	4525
22	Rajasthan	10510	10465	10444	10320	10563
23	Sikkim	70	85	78	85	73
24	Tamil Nadu	15642	17218	16157	12216	10525
25	Telangana	7110	7219	6596	6603	6964
26	Tripura	158	173	161	213	239
27	Uttarakhand	913	962	942	1,047	867
28	Uttar Pradesh	17666	19320	20124	22256	22655
29	West Bengal	6234	6544	5769	5711	5500

In [7]:
```
#naming 2 coloumns of dataframe
column_2015=accident_deaths["2015"]
column_2019=accident_deaths["2019"]
```

In [9]:
```
import scipy.stats as stats
#perform two sample t-test
stats.ttest_ind(a=column_2015, b=column_2019, equal_var=True)
```

Out[9]:
```
Ttest_indResult(statistic=-0.13460057684149837, pvalue=0.8934102981029542)
```

1.3) To find correlation between Total deaths in road accidents and Total death rate of India.

In [58]:
```
import pandas as pd
correlation_data = pd.read_csv("Correlation between deaths and accident (per thousand).csv", index_col="S.No.")
```

In [59]:
```
correlation_data
```

Out[59]:

S.No.	States	Road accident	Total deaths (per thousand)	Unnamed: 4	Unnamed: 5	Road accident (per thousand)
1	Andhra\nPradesh	21,992	5.3	1.610000e-05	0.016095	0.016
2	Arunachal\nPradesh	237	6.7	1.730000e-07	0.000173	0.000
3	Assam	8,350	6.0	6.110000e-06	0.006111	0.006
4	Bihar	10,007	6.4	7.320000e-06	0.007324	0.007
5	Chhattisgarh	13,899	5.8	1.020000e-05	0.010172	0.010
6	Goa	3,440	4.3	2.520000e-06	0.002518	0.003
7	Gujarat	17,046	8.0	1.250000e-05	0.012475	0.012
8	Haryana	10,944	3.8	8.010000e-06	0.008009	0.008
9	Himachal Pradesh	2,873	4.5	2.100000e-06	0.002103	0.002
10	Jammu & Kashmir	5,796	3.3	4.240000e-06	0.004242	0.004
11	Jharkhand	5,217	5.9	3.820000e-06	0.003818	0.004
12	Karnataka	40,658	5.9	2.980000e-05	0.029755	0.030
13	Kerala	41,111	5.9	3.010000e-05	0.030087	0.030
14	Madhya Pradesh	50,669	6.9	3.710000e-05	0.037082	0.037

13	Kerala	41,111	5.9	3.010000e-05	0.030087	0.030
14	Madhya Pradesh	50,669	6.9	3.710000e-05	0.037082	0.037
15	Maharashtra	32,925	4.9	2.410000e-05	0.024096	0.024
16	Manipur	672	5.4	4.920000e-07	0.000492	0.000
17	Meghalaya	482	6.3	3.530000e-07	0.000353	0.350
18	Mizoram	62	6.9	4.540000e-08	0.000045	0.000
19	Nagaland	358	5.6	2.620000e-07	0.000262	0.000
20	Odisha	11,064	6.7	8.100000e-06	0.008097	0.008
21	Punjab	6,348	5.5	4.650000e-06	0.004646	0.005
22	Rajasthan	23,480	4.5	1.720000e-05	0.017184	0.017
23	Sikkim	162	5.8	1.190000e-07	0.000119	0.000
24	Tamil Nadu	57,228	4.1	4.190000e-05	0.041882	0.042
25	Telangana	21,570	3.5	1.580000e-05	0.015786	0.016
26	Tripura	655	7.3	4.790000e-07	0.000479	0.000
27	Uttarakhand	1,352	6.9	9.890000e-07	0.000989	0.001
28	Uttar Pradesh	42,572	6.6	3.120000e-05	0.031156	0.031
29	West Bengal	10,158	5.9	7.430000e-06	0.007434	0.007

[60]:

```
x=correlation_data["Road accident (per thousand)"]
y=correlation_data["Total deaths (per thousand)"]
```

In [60]:

```
#naming dataframe coloumn
x=correlation_data["Road accident (per thousand)"]
y=correlation_data["Total deaths (per thousand)"]
```

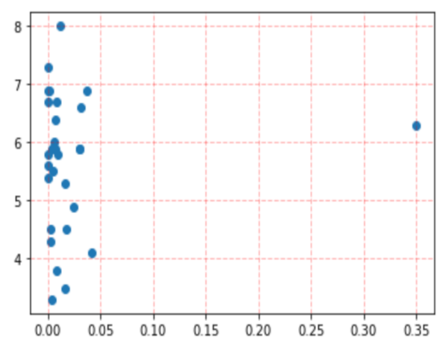
In [61]:

```
#finding correlation using corr()
correlation_coefficient=y.corr(x)
correlation_coefficient
```

Out[61]: 0.08372401110480497

In [62]:

```
from matplotlib import pyplot
pyplot.scatter(x, y)
plt.grid(color = 'red', alpha = 0.3, linestyle = '--', linewidth = 1)
pyplot.show()
```

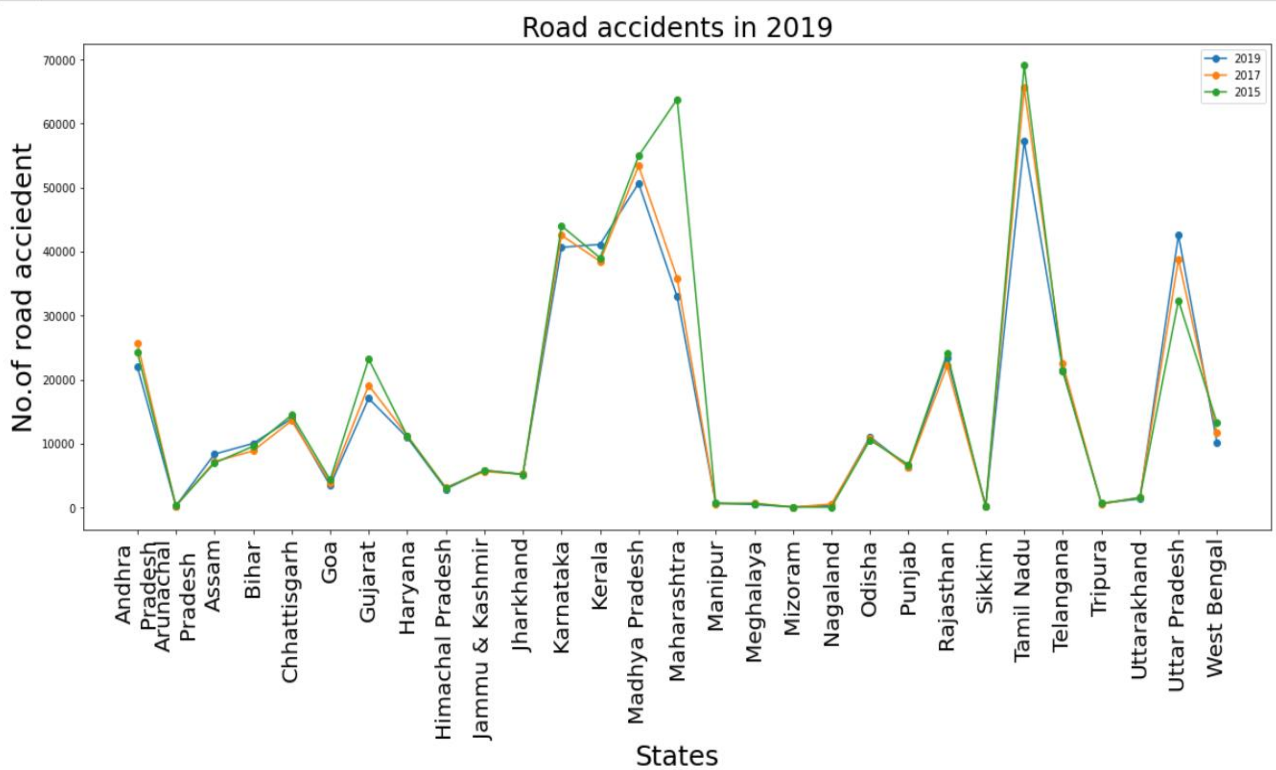


Result = The correlation coefficient is 0.084 which means that there is very slight positive correlation between Total deaths in road accidents and Total death rate of India.

GRAPHS –

In [68]:

```
# adding coloums as variable name
year=road_accident['2019']
year1=road_accident['2017']
year2=road_accident['2015']
#c as x-axis labels
c=road_accident['States']
#plotting figure
plt.figure(figsize=(20,8))
p2=plt.plot(c,year, marker='o')
p3=plt.plot(c,year1, marker='o')
p4=plt.plot(c,year2, marker='o')
#adding labels, title, legend
plt.xticks(rotation=90,fontsize = '20')
plt.xticks(fontsize = '20')
plt.xlabel('States',fontsize = '25')
plt.ylabel('No.of road acciedent',fontsize = '25')
plt.title('Road accidents in 2019',fontsize = '25')
plt.legend(['2019','2017','2015'])
plt.show()
```



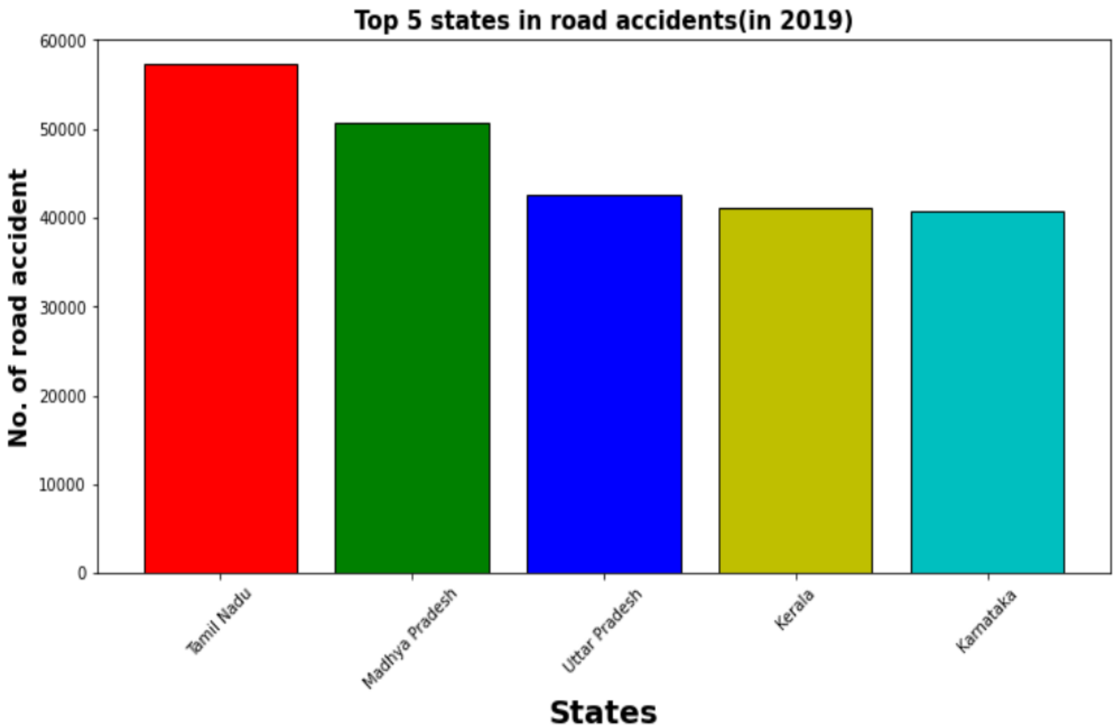
The above graph represents number of road accidents in different years. And the rate is not significantly different.

In [40]:

```
# finding top 5 states with high number of road accidents.
top_5=road_accident.sort_values(by=['2019'],ascending=False).head(5)
top_5
```

Out[40]:

S.No.	States	2015	2017	2019
24	Tamil Nadu	69059	65562	57228
14	Madhya Pradesh	54947	53399	50669
28	Uttar Pradesh	32385	38783	42572
13	Kerala	39014	38470	41111
12	Karnataka	44011	42542	40658



The above graph shows the top 5 states with highest number of road accidents.

In [7]:

```
#plotting bar graph
import matplotlib.pyplot as plt
plt.figure(figsize=(12,6))
plt.bar(top_5.States, top_5['2019'],color = 'rgbycm', edgecolor = 'black')
plt.xticks(rotation=45)
plt.xlabel('States',fontweight = 'bold', fontsize = '20')
plt.ylabel('No. of road accident',fontweight = 'bold', fontsize = '15')
plt.title('Top 5 states in road accidents(in 2019)',fontweight = 'bold', fontsize = '15')
plt.show()
```

In [11]:

```
#importing csv
import pandas as pd
type_of_road_accident = pd.read_csv("Type of road acciedent.csv")
```

In [12]:

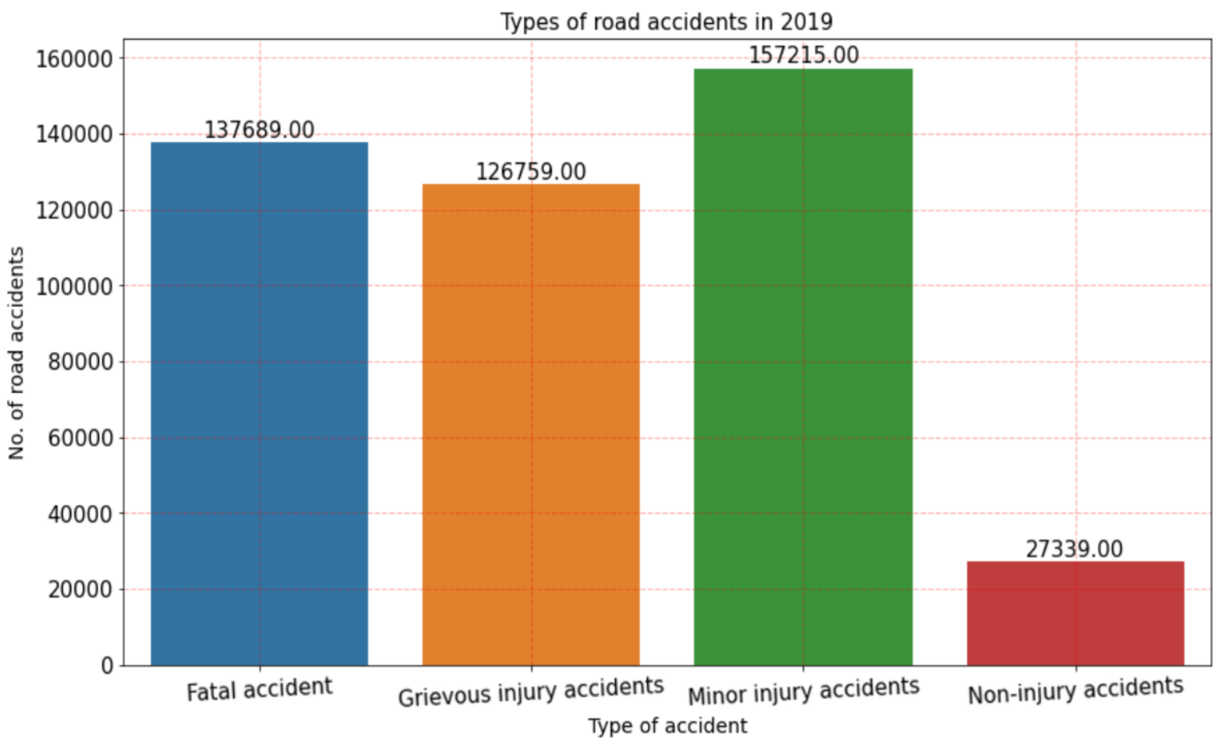
```
type_of_road_accident
```

Out[12]:

	Type_of_road_accident	2015	2016	2017	2018	2019
0	Fatal accident	131726	136071	134796	137726	137689
1	Grievous injury accidents	119,668	120848	120971	125311	126759
2	Minor injury accidents	192634	187642	174400	169920	157215
3	Non-injury accidents	57395	36091	34743	34087	27339

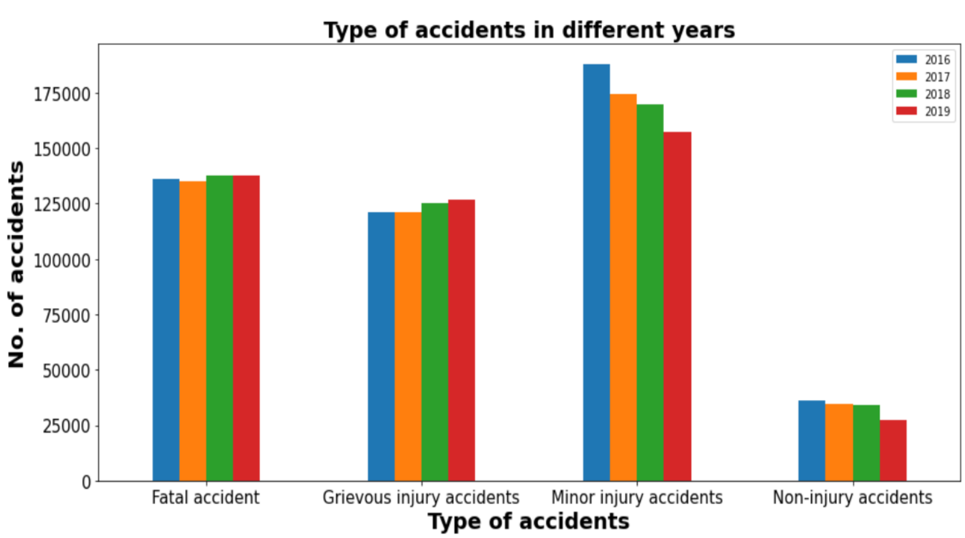
In [13]:

```
import seaborn as sns
# Defining the plot size
plt.figure(figsize=(14, 8))

# Defining the values for x-axis, y-axis
# and from which datafarme the values are to be picked
plots = sns.barplot(x="Type_of_road_accident", y="2019", data=type_of_road_accident)

# Iterrating over the bars one-by-one
for bar in plots.patches:

# Using Matplotlib's annotate function and
# passing the coordinates where the annotation shall be done
# x-coordinate: bar.get_x() + bar.get_width() / 2
# y-coordinate: bar.get_height()
# free space to be left to make graph pleasing: (0, 8)
# ha and va stand for the horizontal and vertical alig
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                       bar.get_height()), ha='center', va='center',
                   size=15, xytext=(0, 8),
                   textcoords='offset points')

# Setting the label for x-axis
plt.xlabel("Type of accident", size=14)

# Setting the label for y-axis
plt.ylabel("No. of road accidents", size=14)
plt.xticks(rotation=3,fontsize = "15")
plt.yticks(fontsize = "15")
# Setting the title for the graph
plt.title("Types of road accidents in 2019",fontsize = "15")
plt.grid(color = 'red', alpha = 0.3, linestyle = '--', linewidth = 1)
# Fianlly showing the plot
plt.show()
```



From the above graph we can conclude that there are very less accidents without any injury.

In [133]:
```
#plotting bar graph
type_of_road_accident.plot(x="Type_of_road_accident",y=["2016", "2017", "2018","2019" ], kind="bar",figsize=(15,7))
plt.xlabel("Type of accidents",fontweight = 'bold', fontsize = '20')
plt.xticks(rotation=0,fontsize = "15")
plt.yticks(fontsize = "15")
plt.ylabel("No. of accidents",fontweight = 'bold', fontsize = '20')
plt.title("Type of accidents in different years",fontweight = 'bold', fontsize = '20')
plt.show()
```



From above graph we can conclude that in 2016 there were more minor injury accidents. The number of fatal accidents in each year is almost same.

In [56]:

```python
#creating dataframe of types of road and accident
import pandas as pd
data=[["National highway",137191,53872,137549],
      ["State highway",108976,38472,111831],
      ["Other roads",202835,58769,201981]]
df=pd.DataFrame(data,columns=["Name","Accident","Person killed","Person injured"])
print(df)
```

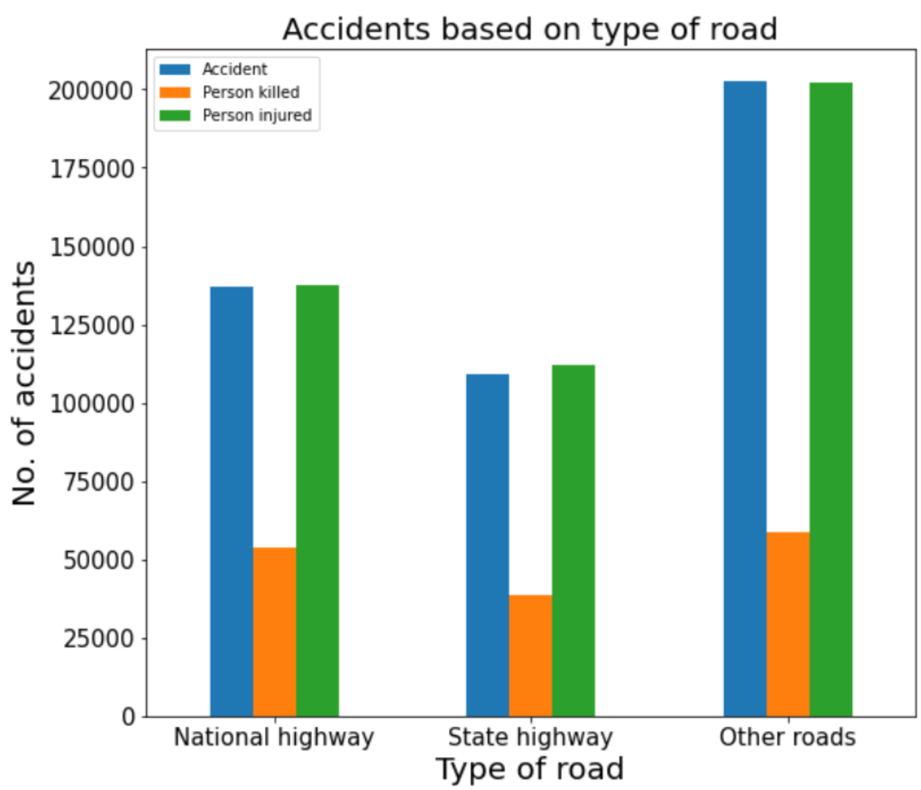
```
               Name  Accident  Person killed  Person injured
0  National highway    137191          53872          137549
1     State highway    108976          38472          111831
2       Other roads    202835          58769          201981
```

In [9]:

```python
#plotting graph
df.plot(x="Name", y=["Accident","Person killed","Person injured"], kind="bar",figsize=(9,8))
plt.xticks(rotation=0,fontsize = "15")
plt.yticks(fontsize = "15")
plt.show()
```

In [16]:

```python
#plotting graph
df.plot(x="Name", y=["Accident","Person killed","Person injured"], kind="bar",figsize=(9,8))
plt.xticks(rotation=0,fontsize = "15")
plt.yticks(fontsize = "15")
plt.xlabel("Type of road", fontsize = '20')
plt.ylabel("No. of accidents", fontsize = '20')
plt.title("Accidents based on type of road", fontsize = '20')
plt.show()
```



Form above graph it can be concluded that other roads having more accidents may be due to traffic and heavy, light vehicles on the same road.

In [17]:
```
import pandas as pd
age_and_gender_wise = pd.read_csv("age and gender wise(2019).csv",index_col="Age-group")
```

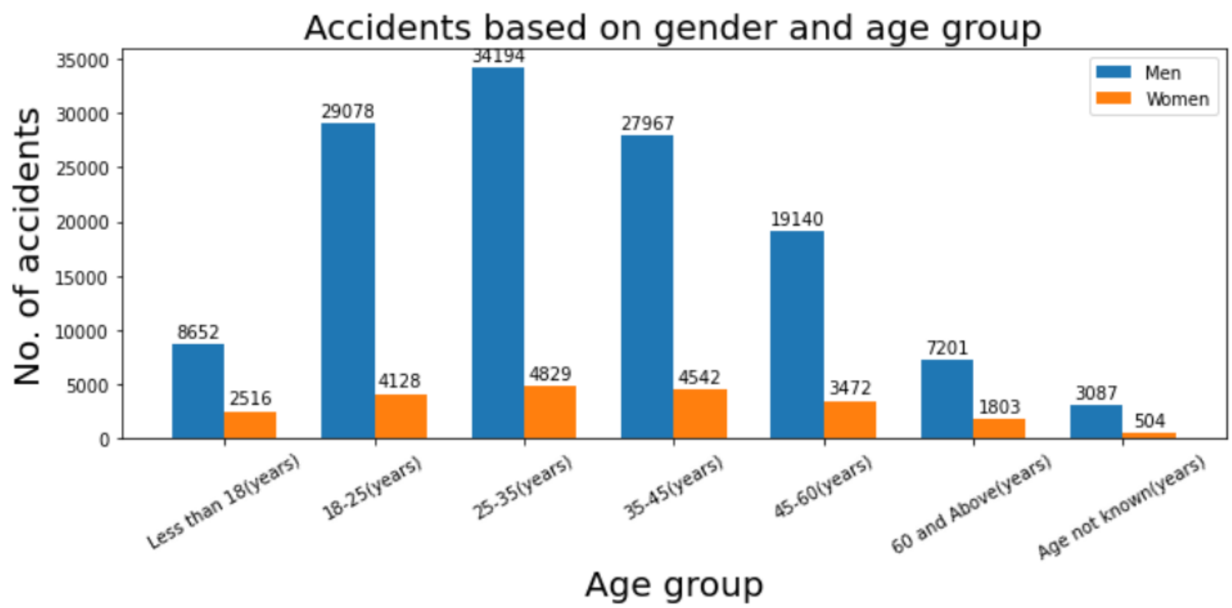
In [6]:
```
age_and_gender_wise
```

Out[6]:

	Male	Female	Total
Age-group			
Less than 18(years)	8652	2516	11168
18-25(years)	29078	4128	33206
25-35(years)	34194	4829	39023
35-45(years)	27967	4542	32509
45-60(years)	19140	3472	22612
60 and Above(years)	7201	1803	9004
Age not known(years)	3087	504	3591

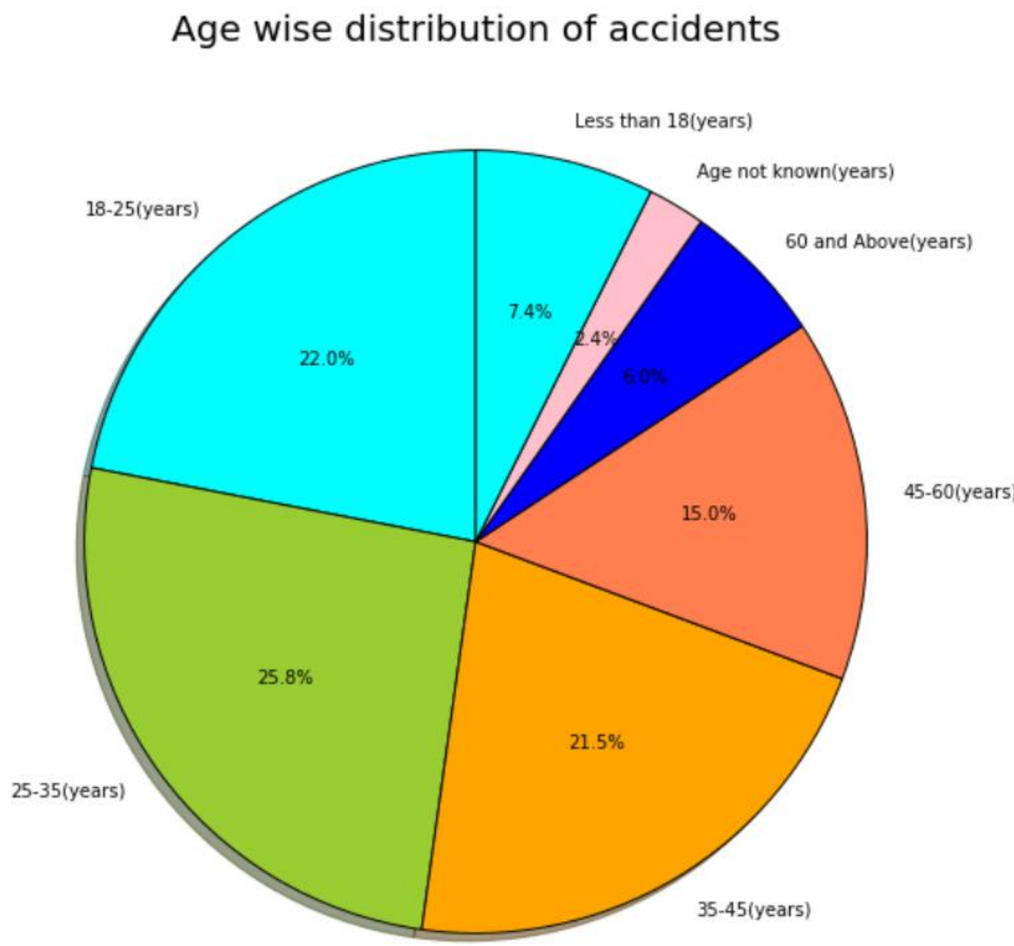
- In [19]:

```
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
plt.figure(figsize=(15, 8))

labels = age_and_gender_wise.index
men_means = age_and_gender_wise["Male"]
women_means = age_and_gender_wise["Female"]
x = np.arange(len(labels))  # the label locations
width = 0.35 # the width of the bars
fig, ax = plt.subplots(figsize=(10,5))
rects1 = ax.bar(x - width/2, men_means, width, label='Men')
rects2 = ax.bar(x + width/2, women_means, width, label='Women')
# Add some text for labels, title and custom x-axis tick labels, etc.
ax.set_xlabel('Age group', fontsize = '20')
ax.set_ylabel('No. of accidents', fontsize = '20')
ax.set_title('Accidents based on gender and age group', fontsize = '20')
ax.set_xticks(x)
ax.set_xticklabels(labels)
ax.legend()
plt.xticks(rotation=30)
def autolabel(rects):
    """Attach a text label above each bar in *rects*, displaying its height."""
    for rect in rects:
        height = rect.get_height()
        ax.annotate('{}'.format(height),
                    xy=(rect.get_x() + rect.get_width() / 2, height),
                    xytext=(0, 2),  # 3 points vertical offset
                    textcoords="offset points",
                    ha='center', va='bottom')
autolabel(rects1)
autolabel(rects2)

fig.tight_layout()

plt.show()
```



From the above graph it can be concluded that mostly males of 25-35 years of age are involved in road accidents.

In [9]:
```
import pandas as pd
from matplotlib import pyplot as plt
f = plt.figure()
f.set_figwidth(25)
f.set_figheight(10)
sales_data = age_and_gender_wise.groupby('Age-group')[['Total']].sum()

colors_list = ['cyan', 'yellowgreen', 'orange', 'coral', 'Blue', 'pink']

plt.pie(sales_data['Total'], labels = sales_data.index,
    autopct = '%1.1f%%', colors = colors_list,wedgeprops = {"edgecolor" : "black",
                     'linewidth': 1,
                     'antialiased': True}, shadow = True, startangle = 90,)
plt.title("Age wise distribution of accidents", fontsize = '20')
plt.show()

```



Above pie chart conclude that accidents involves person between 25-45 years of age.

In [106]:

```python
import pandas as pd
rule_voilation = pd.read_csv("traffic rule voilation.csv",index_col="Traffic rules violation")
```

In [107]:

```python
rule_voilation
```

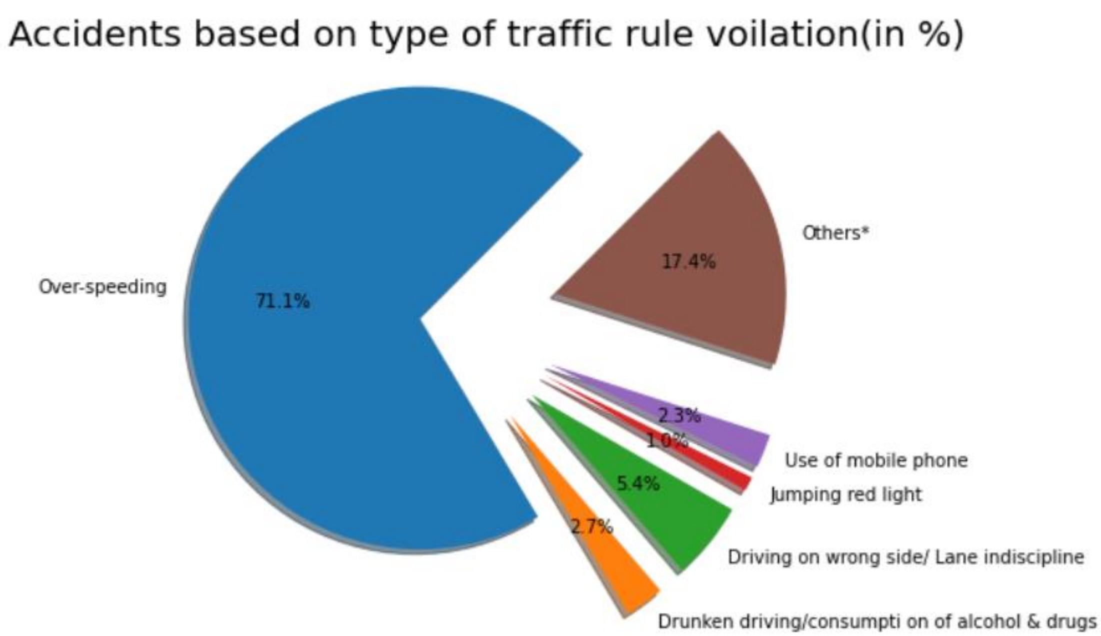
Out[107]:

Traffic rules violation	S.No.	Number of accidents(2018)	Persons Killed(2018)	Persons injured(2018)	Number of accidents(2019)	Persons Killed(2019)	Persons injured(2019)
Over-speeding	1	310612	97588	316421	319028	101723	326850
Drunken driving/consumpti on of alcohol & drugs	2	12018	4188	9944	12256	5325	10564
Driving on wrong side/ Lane indiscipline	3	24781	8764	24100	24431	9201	24628
Jumping red light	4	4441	1545	4126	4443	1797	4006
Use of mobile phone	5	9039	3707	7878	10522	4945	8144
Others*	6	106150	35625	106949	78322	28122	77169

-

In [24]:

```python
plt.figure(figsize=(15, 15))
labels = rule_voilation.index
sizes = rule_voilation["Number of accidents(2019)"]
explode = (0.1, 0.5, 0.5, 0.5,0.5,0.5)  # only "explode" the 2nd slice (i.e. 'Hogs')

fig1, ax1 = plt.subplots(figsize=(15,6))
ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=45)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title("Accidents based on type of traffic rule voilation(in %)", fontsize=20)
plt.show()
```



From the above graph we can conclude that over speeding is the reason for road accidents.

In [10]:
```python
import pandas as pd
Year_wise = pd.read_csv("india accident(1970-2019).csv",index_col="Years")
```

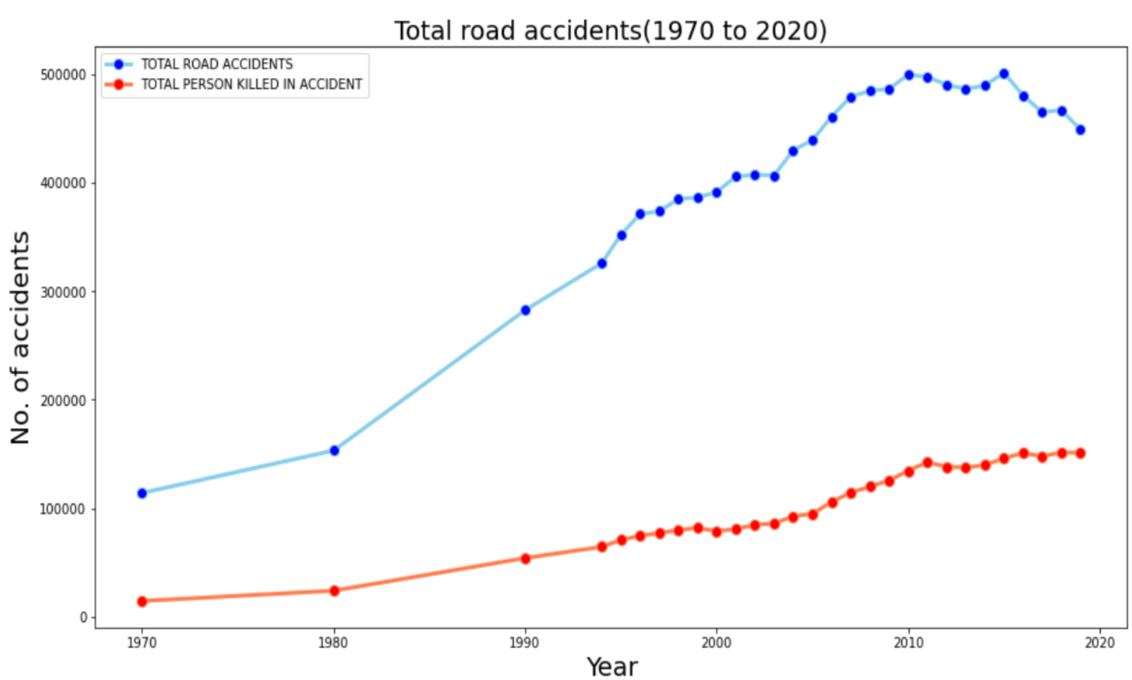
In [11]:
```python
Year_wise
```

Out[11]:

Years	Total Number of Road Accidents (in numbers)	Total Number of Persons Killed (in numbers)
1970	114100	14500
1980	153200	24000
1990	282600	54100
1994	325864	64463
1995	351999	70781
1996	371204	74665
1997	373671	76977
1998	385018	79919
1999	386456	81966
2000	391449	78911
2001	405637	80888
2002	407497	84674
2003	406726	85998
2004	429910	92618
2005	439255	94968

- In [28]:

```python
# libraries
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
plt.figure(figsize=(15, 8))
# Data
df=pd.DataFrame({'x': Year_wise.index, 'TOTAL ROAD ACCIDENTS': Year_wise['Total Number of Road Accidents (in numbers)'],
                 'TOTAL PERSON KILLED IN ACCIDENT': Year_wise['Total Number of Persons Killed (in numbers)']})
plt.xlabel("Year", fontsize = '20')
plt.ylabel("No. of accidents", fontsize = '20')
plt.title("Total road accidents(1970 to 2020)", fontsize = '20')
# multiple line plot
plt.plot( 'x', 'TOTAL ROAD ACCIDENTS', data=df, marker='o', markerfacecolor='blue', markersize=8, color='skyblue', linewidth
plt.plot( 'x', 'TOTAL PERSON KILLED IN ACCIDENT', data=df, marker='o', color='coral', linewidth=3, markerfacecolor='red', ma

plt.legend()

```



The above graph represents that the rate of road accidents and person killed in accidents are gradually increases year by year with very slight decrement between 2016-17 in India. Necessary measures should be taken to reduce the rate of accidents.

Objective 4–
To study lifestyle of different weight categories

4.1 To study lifestyle of different weight categories

1.Hypothesis1-From the survey of 84 females and 46 males.40 females,30 males claimed that in their family no one suffer from similar condition(obesity) is not transferred to him/her by heredity.

Null Hypothesis (H0)- p1=p2
Alternative Hypothesis: p1>p2 (Right Tailed)
Where, p1=proportions of sample1, male participants =0.65
P2=proportions of sample2, female participants =0.48
Sample 1(n1, sample size of male participants) =46
Sample 2(n2, sample size of female participants) = 84
Level of significance*: 0. 05*

Test statistics-

We will use z-test to test the hypothesis concerning difference of proportions and n>30

Z= (p1-p2)/sqrt(p*q(1/n1+1/n2))

Where p= (n1p1+n2p2)/(n1+n2)

P1=X1/ n1; p2=X2/n2 ; x1=26; x2=44

P1 =0.56

p2=0.52

z=1.89

Criteria of rejection-

Reject the null hypothesis if z>z0.05 (right tailed)

Z=1.89

Z0.05=1.645

Decision-Reject the null hypothesis which signifies that risk for obesity is through heredity.

Python implementation

```
In [1]: import pandas as pd
```

```
In [44]: food=pd.read_csv("obesity1.csv",index_col='ID')
         food
```

Out[44]:

ID	What is your gender ?	Select your age group ?	Do you try to maintain balance diet ?	How often do you eat junk food at your home or your workplace ?	Select the weight category you belong to ?	Which of the following meals,you are most likely to have after busy day schedule ?Rate from 1-5 (with 1 being least likely and 5 being most).\n\nChapati,Dal,Rice, vegetables.	Non-vegetarian meal.	Pizza,Burger,fries,cold drink.	How often you eat fruits and salad ?	What sort of baverages you take very often or on daily basis ?\n(Choose all options that apply).	How often you take alcoholic beverages ?	H
1	Male	20- 30 years old .	Yes	2-4 times per week	Fit	5	3	3	Often	Juices;Coffee;	Sometimes	
2	Male	20- 30 years old .	Yes	Daily	Fit	1	1	1	Always	Tea ;	Frequently	
3	Female	40 - 60 years	Maybe	Daily	Healthy	2	5	5	Always	Tea ;Coffee;Cold drinks;Juices;Milk;	Frequently	
4	Male	20- 30 years old .	Yes	Once a week	Obese	4	4	1	Always	Coffee;	Never	
5	Female	20- 30 years old .	Yes	Once a week	Overweight	5	2	1	Often	Tea ;Coffee;Milk;	Sometimes	
...	...	...	...	...	...	...	...	...	...	...	...	
126	Female	20- 35 years old .	Maybe	2-4 times per week	Obese	4	1	2	Always	Tea ;Milk;Coffee;	Sometimes	
127	Female	Above 60 years old.	Yes	Once a week	Healthy	4	3	1	Often	Tea ;Milk;	Never	
128	Male	20- 35 years old .	Yes	Once a week	Healthy	3	5	4	Often	Juices;	Never	
129	Male	20- 35 years old .	Yes	Once a week	Healthy	5	5	1	Sometimes	Cold drinks;Milk;	Sometimes	
130	Female	20- 35 years old .	Yes	Once a week	Healthy	5	5	1	Always	Green tea ;Juices;Milk;	Never	

130 rows × 15 columns

Python code

```python
#p1=proportion of population of mens claiming obesity is through heredity =0.65
p1=float(input(" proportion of population of mens claiming obesity is through heredity ,p1=  "))
#p2=proportion of population of females  claiming obesity is through heredity=0.48
p2=float(input(" proportion of population of females  claiming obesity is through heredity ,p2=  "))
q1=1-p1
q2=1-p2
sig=float(input("Level of Signifiance,alpha  = "))
n1=float(input("Sample size of  male,n1 = "))
n2=float(input("Sample size of female ,n2 =  "))
if(p1>p2):
    print("Right Tailed Hypothesis ")
elif(p1<p2):
    print("Left Tailed Hypothesis")
elif(p1!=p2):
    print("Two Tailed Hypothesis")

print("Test Statitics")
p=((n1*p1)+(n2*p2))/(n1+n2)
q=1-p
a=((1/n1)+(1/n2))*(p*q)
z=(p1-p2)/(math.sqrt(a))
print("The value of z= ", z)
za=1.645#a=aplha
zb=1.96# z(a/2) that is z(alpha/2)
print("Criteria of Rejection")
if(z>za):
    print("Right Tailed")
```

```python
    print("Rject the null Hypothesis")
    print("At 5% level,these samples will revelas that  risk of obesity is through heredity")
elif(z<(zc)):
    print("Left Tailed")
    print("Reject the null Hypothesis")
    print("At 5% level,these samples will revelas that  risk of obesity is through heredity")
elif((abs(z))>zb):
    print("Two tailed")
    print("Reject the null Hypothesis")
    print("At 5% level,these samples will revelas that  risk of obesity is through heredity ")
else:
    print("Hypothesis Accepted")
    print("At 5% level,these samples will claimed that risk of obesity is not through heredity")
```

Output of code

```
 proportion of population of mens claiming obesity is through heredity ,p1=  0.65
 proportion of population of females  claiming obesity is through heredity ,p2=  0.48
Level of Signifiance,alpha  = 0.05
Sample size of  male,n1 = 46
Sample size of female ,n2 =  84
Right Tailed Hypothesis
Test Statitics
The value of z=  1.859648176034905
Criteria of Rejection
Right Tailed
Rject the null Hypothesis
At 5% level,these samples will revelas that  risk of obesity is through heredity
```

Amongst the age categories of the above survey who eats the maximum number of chappties , rice, dal, vegetables, pizza, burger, cold drink, milk, tea, coeffe , green Tea, non-veg.

Types of food

Age-category	Meals	Non-Veg	Junk-Food	Fruit &Salad
Less than20	4	1	4	3
20-35	5	5	3	3
36-40	5	5	5	3
Above60	4	3	1	2

By using power method, we have calculate the eigen values and their corresponding eigen vectors by using power method.

Python code for power method

```
In [1]: import numpy as np
        import sys
        n = int(input('Enter order of matrix: '))
        a = np.zeros((n,n))
        print('Enter Matrix Coefficients:')
        for i in range(n):
            for j in range(n):
                a[i][j] = float(input( 'a['+str(i)+']['+ str(j)+']='))
        x = np.zeros((n))
        print('Enter initial guess vector: ')
        for i in range(n):
            x[i] = float(input( 'x['+str(i)+']='))
        tolerable_error = float(input('Enter tolerable error: '))
        max_iteration = int(input('Enter maximum number of steps: '))
        lambda_old = 1.0
        condition =  True
        step = 1
        while condition:
           x = np.matmul(a,x)
            lambda_new = max(abs(x))
            x = x/lambda_new
            print('\nSTEP %d' %(step))
            print('----------')
            print('Eigen Value = %0.4f' %(lambda_new))
            print('Eigen Vector: ')
```

```
print('Eigen Vector: ')
for i in range(n):
    print('%0.3f\t' % (x[i]))
step = step + 1
if step > max_iteration:
    print('Not convergent in given maximum iteration!')
    break
error = abs(lambda_new - lambda_old)
print('errror='+ str(error))
lambda_old = lambda_new
condition = error > tolerable_error
```

```
Enter order of matrix: 4
Enter Matrix Coefficients:
a[0][0]=4
a[0][1]=1
a[0][2]=4
a[0][3]=3
a[1][0]=5
a[1][1]=5
a[1][2]=3
a[1][3]=3
a[2][0]=5
a[2][1]=5
a[2][2]=5
a[2][3]=3
a[3][0]=4
a[3][1]=3
a[3][2]=1
```

```
Enter Matrix Coefficients:
a[0][0]=4
a[0][1]=1
a[0][2]=4
a[0][3]=3
a[1][0]=5
a[1][1]=5
a[1][2]=3
a[1][3]=3
a[2][0]=5
a[2][1]=5
a[2][2]=5
a[2][3]=3
a[3][0]=4
a[3][1]=3
a[3][2]=1
a[3][3]=2
Enter initial guess vector:
x[0]=1
x[1]=1
x[2]=1
x[3]=1
Enter precession : 0.00001
Enter maximum number of steps: 100
```

```
STEP 1
----------
Eigen Value = 18.0000
Eigen Vector:
0.667
0.889
1.000
0.556
errror=17.0

STEP 2
----------
Eigen Value = 14.4444
Eigen Vector:
0.638
0.862
1.000
0.515
errror=3.5555555555555554
```

```
STEP 3
----------
Eigen Value = 14.0462
Eigen Vector:
0.638
0.858
1.000
0.510
errror=0.39829059829059865

STEP 4
----------
Eigen Value = 14.0093
Eigen Vector:
0.638
0.857
1.000
0.510
errror=0.036843879012552705
```

```
STEP 5
----------
Eigen Value = 14.0074
Eigen Vector:
0.638
0.857
1.000
0.510
errror=0.0018825796271233486

STEP 6
----------
Eigen Value = 14.0075
Eigen Vector:
0.638
0.857
1.000
0.510
errror=0.00011606478040349089

STEP 7
----------
Eigen Value = 14.0076
Eigen Vector:
0.638
0.857
1.000
0.510
errror=4.415212115560507e-05

STEP 8
----------
Eigen Value = 14.0076
Eigen Vector:
0.638
0.857
1.000
0.510
errror=6.514726443640484e-06
```

Ranking foods

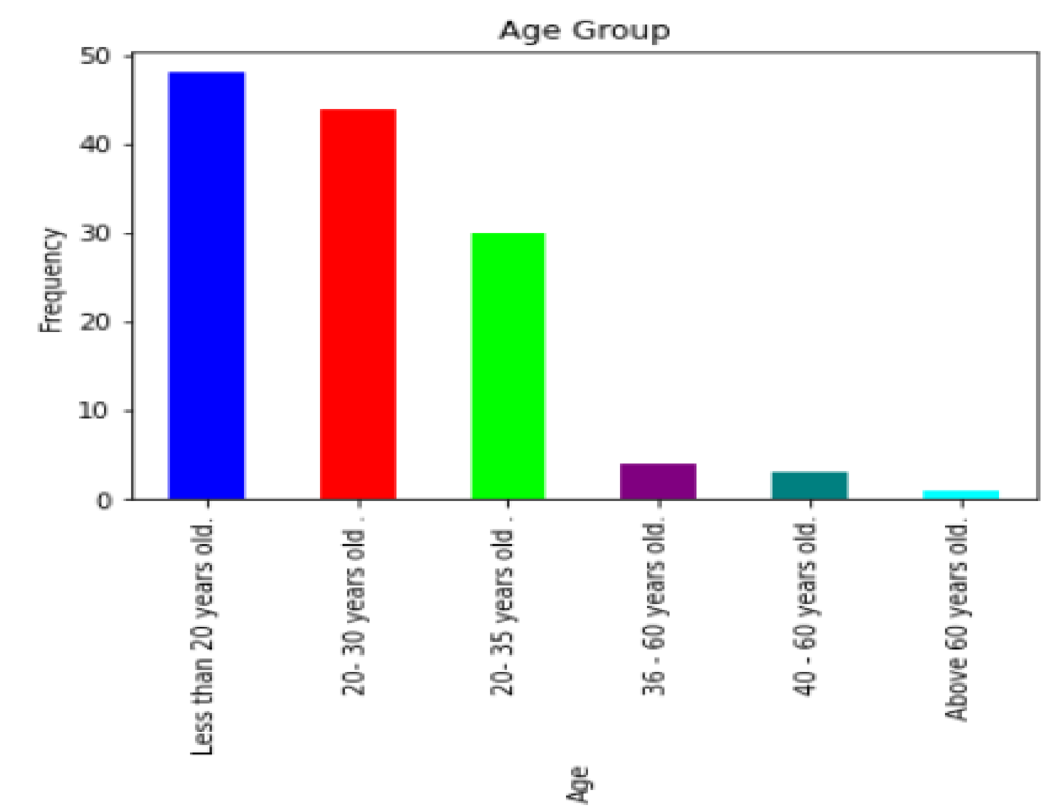
Junk-food

Non-veg

Chapattis , dal, vegetables, rice

Fruit and salad

```
import matplotlib.pyplot as plt
a=food['Select your age group ?'].value_counts()
a.plot.bar(color=['blue','red','lime','purple','teal','cyan'])
plt.title("Age Group")
plt.xlabel("Age",rotation=90)
plt.ylabel("Frequency")
plt.show()
```



Above graph shows that in our survey how many people and from which age category are participating.

```
b=food['Do you try to maintain balance diet ?'].value_counts()
b.plot.bar(color=['green','blue','red'])
plt.title("Do you try to maintain balance diet")
plt.xlabel("Responses")
plt.ylabel("Frequency of responses")
plt.show()
```

Objective 3-

To compare alcohol consumption in urban areas
Claim: Alcohol consumption in rural and urban area is equal

Testing Hypothesis:

Null Hypothesis H0: µ1-µ2=0(alcohol consumption in rural and urban area is equal)
Alternative Hypothesis: µ1-µ2≠0(alcohol consumption in rural and urban area is not equal)

Level of Significance(α):0.05

Sample Size: Sample1 -n1 and Sample2 -n2
n1=n2=22

n<30, so here we will apply t-test

$$t = \frac{\overline{x_1} - \overline{x_2} - \delta}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{\eta_2}}}$$

Where , $S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

S1^2 and S2^2 are sample variance of sample 1 and 2 respectively.

Here, n1=n2=22

$\overline{x_1}$=1.7864
$\overline{x_2}$=3.2727

```
Variance1=2.8693
Variance2=4.5298
Sp=3.7916
```

15	Ladakh	5.3	3.4
16	Andaman and Nicobar Islands	0.7	7.6
17	Goa	5.6	5.3
18	Tripura	0.8	8.4
19	Telangana	2.6	9.0
20	Assam	2.6	8.2
21	Sikkim	12.7	18.4

```
In [10]: data1['Urban'].mean()

Out[10]: 1.7863636363636366

In [11]: data1['Urban'].std()
Out[11]: 2.869261636320804

In [8]: data1['Rural'].mean()
Out[8]: 3.272727272727273

In [12]: data1['Rural'].std()
Out[12]: 4.529814651751303
```

```
|t|=1.3001(observed value)
```

Criteria of rejection:

Reject the null hypothesis if

$|t| > t_{\alpha, n1+n2-2}$

Now from t-table:

$t_{\alpha, n1+n2-2} = 1.645$

So hypothesis cannot be rejected because $|t| > t_{\alpha, n1+n2-2}$ is not satisfied.

Python implementation

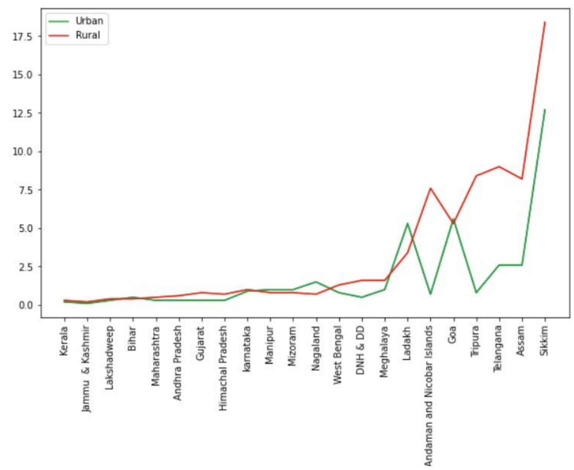
```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np

In [2]: data1=pd.read_csv("project--1(W).csv")
        del data1['Total']
        data1
Out[2]:
```

	State	Urban	Rural
0	Kerala	0.2	0.3
1	Jammu & Kashmir	0.1	0.2
2	Lakshadweep	0.3	0.4
3	Bihar	0.5	0.4
4	Maharashtra	0.3	0.5
5	Andhra Pradesh	0.3	0.6
6	Gujarat	0.3	0.8
7	Himachal Pradesh	0.3	0.7
8	karnataka	0.9	1.0
9	Manipur	1.0	0.8
10	Mizoram	1.0	0.8
11	Nagaland	1.5	0.7
12	West Bengal	0.8	1.3
13	DNH & DD	0.5	1.6
14	Meghalaya	1.0	1.6
15	Ladakh	5.3	3.4

15	Ladakh	5.3	3.4
16	Andaman and Nicobar Islands	0.7	7.6
17	Goa	5.6	5.3
18	Tripura	0.8	8.4
19	Telangana	2.6	9.0
20	Assam	2.6	8.2
21	Sikkim	12.7	18.4

```
In [10]: data1['Urban'].mean()

Out[10]: 1.7863636363636366

In [11]: data1['Urban'].std()
Out[11]: 2.869261636320804

In [8]: data1['Rural'].mean()
Out[8]: 3.272727272727273

In [12]: data1['Rural'].std()
Out[12]: 4.529814651751303
```

```
In [3]: plt.figure(figsize=(10,6))
        plt.plot(data1['State'],data1['Urban'])
        plt.plot(data1['State'],data1['Rural'])
        plt.xticks(rotation=90)
        p1=plt.plot(data1['Urban'])
        p2=plt.plot(data1['Rural'])
        plt.legend((p1[0],p2[0]),('Urban','Rural'))
        plt.show()
```



Objective - 4

To compare infant mortality rate in 2011 and 2018 and to compare infant mortality with life expectancy

4.1 To compare infant mortality rate in 2011 and 2018

Hypothesis-1-Infant mortality rate in India have decreased from 2011 to 2018

Applying t-test

Step-1 Null Hypothesis H0: $\mu_1 > \mu_2$ (Infant mortality rate has decreased from 2011 to 2018)

Alternative Hypothesis Ha: $\mu_1 < \mu_2$(Infant mortality rate has not decreased from 2011 to 2018)

Level of Significance: $\alpha = 0.05$

Sample Size: Sample 1 -n1=22

Sample 2 –n2=22

Step-2 Test Statistics

Since n<30 we apply t-test

$$t = \frac{\overline{x_1} - \overline{x_2} - \delta}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

$$\text{Here, } s_p^2 = \frac{s_1^2 + s_2^2}{2} \text{ if } n_1 = n_2$$

Step-3 Criteria of Rejection

It’s a left tail test.

$t < -t_{\alpha, n1+n2-2}$

$-t_{0.05,42}$=1.65(from statistical tables)

Step-4 Calculation

(from python code shown next)

$\overline{x1}$=39.2727

$\overline{x2}$=27.1364

Var1=156.4926

Var2=133.5527

Sp= Var1* Var2 =20900.009

$$t = \frac{\overline{X_1} - \overline{X_2} - \delta}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

t=0.001969(from calculation)

Step-4 Decision

$-t_{0.05,44}$=1.65(from statistical tables)

t=0.001969(from calculation)

It can be seen that t>-1.65 We cannot reject null hypothesis and it can also be concluded that that Infant mortality rate has decreased from 2011 to 2018 .

Python Implementation

In [12]:
```
import pandas  as pd              #importing necessary library
import matplotlib.pyplot as plt
import scipy
import seaborn as sns
```

In [13]:
```
infant=pd.read_csv("infant mortality__final1utf.csv",index_col='State/Union Territory') #reading csv file
```

In [14]:
```
infant #dataframe displaying
```

Out[14]:

State/Union Territory	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Andhra Pradesh	59	57	56	54	52	49	46	43	41	39	39	37	34	32	29
Assam	66	68	67	66	64	61	58	55	55	54	49	47	44	44	41
Bihar	61	61	60	58	56	52	48	44	43	42	42	42	38	35	32
Chhattisgarh	60	63	61	59	57	54	51	48	47	46	43	41	39	38	41
Delhi	32	35	37	36	35	33	30	28	25	24	20	18	18	16	13
Gujarat	53	54	53	52	50	48	44	41	38	36	35	33	30	30	28
Haryana	61	60	57	55	54	51	48	44	42	41	36	36	33	30	30
Himachal Pradesh	51	49	50	47	44	45	40	38	36	35	32	28	25	22	19
Jammu and Kashmir	49	50	52	51	49	45	43	41	39	37	34	26	24	23	22
Jharkhand	49	50	49	48	46	44	42	39	38	37	34	32	29	29	30
Karnataka	49	50	48	47	45	41	38	35	32	31	29	28	24	25	23
Kerala	12	14	15	13	12	12	13	12	12	12	12	12	10	10	7
Madhya Pradesh	79	76	74	72	70	67	62	59	56	54	52	50	47	47	48
Maharashtra	36	36	35	34	33	31	28	25	25	24	22	21	19	19	19
Odisha	77	75	73	71	69	65	61	57	53	51	49	46	44	41	40
Punjab	45	44	44	43	41	38	34	30	28	26	24	23	21	21	20
Rajasthan	67	68	67	65	63	59	55	52	49	47	46	43	41	38	37
Sikkim	32	30	33	34	33	34	30	26	24	22	19	18	16	12	7
Tamil Nadu	41	37	37	35	31	28	24	22	21	21	20	19	17	16	15
Uttar Pradesh	72	73	71	69	67	63	61	57	53	50	48	46	43	41	43
Uttarakhand	42	42	43	48	44	41	38	36	34	32	33	34	38	32	31
West Bengal	40	38	38	37	35	33	33	32	32	31	28	26	25	24	22

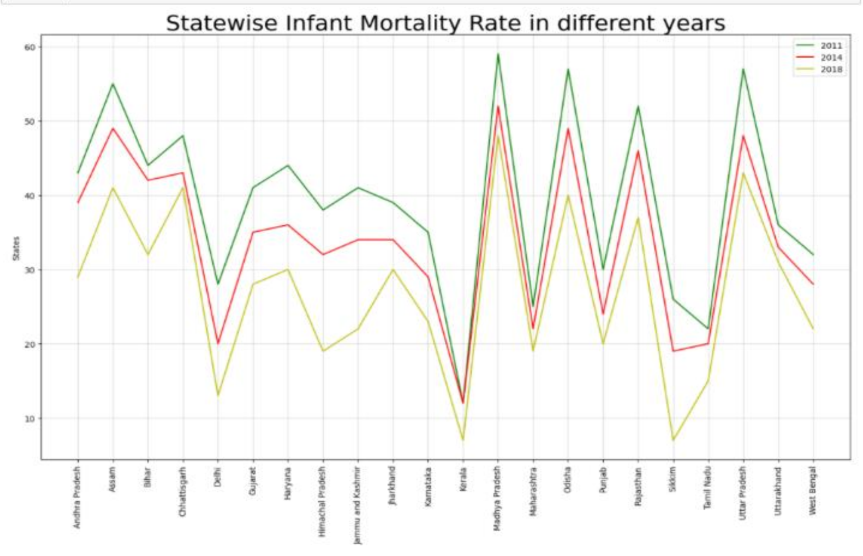
In [16]:
```
year_2011=infant['2011']  #taking columns
year_2018=infant['2018']
```

In [17]:
```
result = pd.concat([year_2011, year_2018], axis=1) #concating two columns of dataframe
result.describe()                                  #finding mean,standard deviation
```

Out[17]:

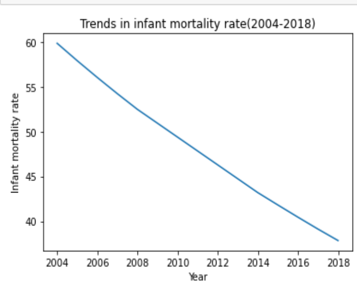
	2011	2018
count	22.000000	22.000000
mean	39.272727	27.136364
std	12.509736	11.556468
min	12.000000	7.000000
25%	30.500000	19.250000
50%	40.000000	28.500000
75%	47.000000	35.750000
max	59.000000	48.000000

```
In [68]: plt.figure(figsize=(17,10), dpi= 100)
         plt.plot(infant["2011"],'g',infant["2014"],'r',infant["2018"],'y')
         plt.xticks(rotation=90)
         plt.ylabel('States')
         plt.title("Statewise Infant Mortality Rate in different years", fontsize=27)
         plt.grid(axis='both', alpha=0.5)
         plt.legend(['2011','2014','2018'])
         plt.show()
```



The above graph represents trend of infant mortality rate in different years.It can be seen that rate was highest in 2011 then less in 2014 and then more decreased in 2018.

```
In [56]: import matplotlib.pyplot as plt

         Year = [2004,2005,2006,2007,2008,2014,2015,2016,2017,2018]
         infant_mort_india = [59.85,57.91,56.04,54.23,52.48,43.19,41.81,40.45,39.12,37.84]

         plt.plot(Year, infant_mort_india)
         plt.title('Trends in infant mortality rate(2004-2018)')
         plt.xlabel('Year')
         plt.ylabel('Infant mortality rate')
         plt.show()
```



The above graph represent trend of infant mortality rate from 2004-2018.

Python implementation

```
In [1]: import pandas  as pd              #importing necessary library
        import matplotlib.pyplot as plt
        import scipy
        import seaborn as sns
```

```
In [2]: data=pd.read_csv("infant mortality__final1utf.csv",index_col='State/Union Territory')
```

```
In [4]: data
```

Out[4]:

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
State/Union Territory															
Andhra Pradesh	59.0	57.0	56.0	54.0	52.0	49.0	46.0	43.0	41.0	39.0	39.0	37.0	34.0	32.0	29.0
Assam	66.0	68.0	67.0	66.0	64.0	61.0	58.0	55.0	55.0	54.0	49.0	47.0	44.0	44.0	41.0
Bihar	61.0	61.0	60.0	58.0	56.0	52.0	48.0	44.0	43.0	42.0	42.0	42.0	38.0	35.0	32.0
Chhattisgarh	60.0	63.0	61.0	59.0	57.0	54.0	51.0	48.0	47.0	46.0	43.0	41.0	39.0	38.0	41.0
Delhi	32.0	35.0	37.0	36.0	35.0	33.0	30.0	28.0	25.0	24.0	20.0	18.0	18.0	16.0	13.0
Gujarat	53.0	54.0	53.0	52.0	50.0	48.0	44.0	41.0	38.0	36.0	35.0	33.0	30.0	30.0	28.0
Haryana	61.0	60.0	57.0	55.0	54.0	51.0	48.0	44.0	42.0	41.0	36.0	36.0	33.0	30.0	30.0
Himachal Pradesh	51.0	49.0	50.0	47.0	44.0	45.0	40.0	38.0	36.0	35.0	32.0	28.0	25.0	22.0	19.0
Jammu and Kashmir	49.0	50.0	52.0	51.0	49.0	45.0	43.0	41.0	39.0	37.0	34.0	26.0	24.0	23.0	22.0
Jharkhand	49.0	50.0	49.0	48.0	46.0	44.0	42.0	39.0	38.0	37.0	34.0	32.0	29.0	29.0	30.0

```
In [8]: line=pd.read_csv('india_infant.csv')
        line
```

Out[8]:

	Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	...	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
0	India	68.26	66.06	63.92	61.85	59.85	57.91	56.04	54.23	52.48	...	47.57	46.07	44.6	43.19	41.81	40.45	39.12	37.84	36.6	35.39

1 rows × 22 columns

```
In [9]: data1=data.iloc[:,7:]
        data1
```

Out[9]:

	2011	2012	2013	2014	2015	2016	2017	2018
State/Union Territory								
Andhra Pradesh	43.0	41.0	39.0	39.0	37.0	34.0	32.0	29.0
Assam	55.0	55.0	54.0	49.0	47.0	44.0	44.0	41.0
Bihar	44.0	43.0	42.0	42.0	42.0	38.0	35.0	32.0
Chhattisgarh	48.0	47.0	46.0	43.0	41.0	39.0	38.0	41.0
Delhi	28.0	25.0	24.0	20.0	18.0	18.0	16.0	13.0
Gujarat	41.0	38.0	36.0	35.0	33.0	30.0	30.0	28.0
Haryana	44.0	42.0	41.0	36.0	36.0	33.0	30.0	30.0
Himachal Pradesh	38.0	36.0	35.0	32.0	28.0	25.0	22.0	19.0
Jammu and Kashmir	41.0	39.0	37.0	34.0	26.0	24.0	23.0	22.0

```
In [9]: data1=data.iloc[:,7:]
        data1
```

Out[9]:

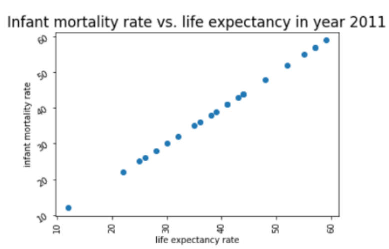
State/Union Territory	2011	2012	2013	2014	2015	2016	2017	2018
Andhra Pradesh	43.0	41.0	39.0	39.0	37.0	34.0	32.0	29.0
Assam	55.0	55.0	54.0	49.0	47.0	44.0	44.0	41.0
Bihar	44.0	43.0	42.0	42.0	42.0	38.0	35.0	32.0
Chhattisgarh	48.0	47.0	46.0	43.0	41.0	39.0	38.0	41.0
Delhi	28.0	25.0	24.0	20.0	18.0	18.0	16.0	13.0
Gujarat	41.0	38.0	36.0	35.0	33.0	30.0	30.0	28.0
Haryana	44.0	42.0	41.0	36.0	36.0	33.0	30.0	30.0
Himachal Pradesh	38.0	36.0	35.0	32.0	28.0	25.0	22.0	19.0
Jammu and Kashmir	41.0	39.0	37.0	34.0	26.0	24.0	23.0	22.0
Jharkhand	39.0	38.0	37.0	34.0	32.0	29.0	29.0	30.0
Karnataka	35.0	32.0	31.0	29.0	28.0	24.0	25.0	23.0
Kerala	12.0	12.0	12.0	12.0	12.0	10.0	10.0	7.0
Madhya Pradesh	59.0	56.0	54.0	52.0	50.0	47.0	47.0	48.0
Maharashtra	25.0	25.0	24.0	22.0	21.0	19.0	19.0	19.0
Odisha	57.0	53.0	51.0	49.0	46.0	44.0	41.0	40.0
Punjab	30.0	28.0	26.0	24.0	23.0	21.0	21.0	20.0

```
In [14]: life=pd.read_csv("life_expectancy_final2utf.csv",index_col='State/Union Territory')
         life
```

Out[14]:

State/Union Territory	2004-08(Male)	2004-08(Female)	2004-08(Total)	2008-12 (Male)	2008-12 (Female)	2008-12 (Total)	2012-16 (Male)	2012-16 (Female)	2012-16 (Total)
Andhra Pradesh	63.5	68.1	65.7	64.7	69.4	67.0	68.0	71.4	69.6
Assam	60.0	62.3	61.0	61.2	64.8	62.7	64.4	66.8	65.5
Bihar	64.2	64.7	64.4	66.7	67.6	67.2	68.9	68.5	68.7
Chhattisgarh	-	-	-	-	-	-	63.6	66.8	65.2
Delhi	-	-	-	-	-	-	72.7	75.9	74.2
Gujarat	64.3	68.7	66.4	65.5	70.1	67.7	67.4	71.8	69.5
Haryana	65.3	69.0	67.1	65.4	70.1	67.6	67.2	72.0	69.4
Himachal Pradesh	67.7	72.1	69.8	68.3	72.7	70.5	69.4	75.5	72.3
Jammu and Kashmir	69.0	70.6	69.8	69.9	72.4	71.0	71.6	76.2	73.5
Jharkhand	-	-	-	-	-	-	67.8	68.0	67.9
Karnataka	64.4	69.2	66.7	65.8	70.3	68.0	67.6	70.7	69.1
Kerala	71.2	77.4	74.3	71.6	77.7	74.7	72.2	77.9	75.1
Madhya Pradesh	60.3	62.7	61.4	61.9	65.0	63.3	63.7	67.2	65.4

```
In [17]: fig, axis = plt.subplots()
         # Grid lines, Xticks, Xlabel, Ylabel

         #axis.yaxis.grid(True)
         axis.set_title('Infant mortality rate vs. life expectancy in year 2011',fontsize=17)
         plt.xticks(rotation=90)
         plt.yticks(rotation=30)
         axis.set_xlabel('life expectancy rate',fontsize=10)
         axis.set_ylabel('infant mortality rate',fontsize=10)

         X = data1['2011']
         Y = data['2011']

         axis.scatter(X, Y)
         plt.show()
```



```
In [27]: data2=data1[['2016']]

In [28]: life1=life[['2012-16 (Total)']]

In [29]: result = pd.concat([life1, data2], axis=1)

In [30]: result
```

Out[30]:

	2012-16 (Total)	2016
Andhra Pradesh	69.6	34.0
Assam	65.5	44.0
Bihar	68.7	38.0
Chhattisgarh	65.2	39.0
Delhi	74.2	18.0
Gujarat	69.5	30.0
Haryana	69.4	33.0
Himachal Pradesh	72.3	25.0
Jammu and Kashmir	73.5	24.0
Jharkhand	67.9	29.0
Karnataka	69.1	24.0

4.2 Correlation between infant mortality rate and life expectancy-

```
In [31]: result.corr()
```

Out[31]:

	2012-16 (Total)	2016
2012-16 (Total)	1.000000	-0.898552
2016	-0.898552	1.000000

```
In [32]: result.cov()
```

Out[32]:

	2012-16 (Total)	2016
2012-16 (Total)	8.787532	-28.302143
2016	-28.302143	111.573123

CONCLUSION

Now, this can be concluded that the number of road accident and the number of persons killed in road accidents are increasing gradually per year in India. Tamil Nadu has highest number of road accidents. The major reason of road accidents is over-speeding (traffic rule violation).

Death rate and death due to road accidents are slightly correlated. Necessary and safety measures should be taken by government and citizens to decrease road accidents.

Obesity is not only through heredity but also with unhealthy lifestyle. Those who are maintaining their diet have lesser chance of hypothyroidism and hyperthyroidism Respective of age and gender.

REFERENCE

[1] Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.
10.1016/j.trpro.2017.05.484

[2] ROAD ACCIDENT MODELS FOR LARGE METROPOLITAN CITIES OF INDIA
Author links open overlay pane P. PramadaVALLIPh.D.(Scientist)https://doi.org/10.1016/S0386-1112(14)60119-9

[3] H.M. Hacker, E.A. SuchmanSociological approach to accident research author(s), SocProbl, 10 (4) (1963'A), pp. 383-389

[4] World Health Day: 2004 road safety. [Last cited on 2011 Sept 16].

www.who.int/world-health day/previous/2004/en/index.html .

[5] (2000) Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 894: 1-253.

[6] Weight control information network understanding adult obesity 2004.

[7]Swidan A (1997) Nutritional assessment of females 10-60 years old. House hold survey, Baghdad, general federation of Iraqi women.

[8] WORLD HEALTH ORGANIZATION-THE GLOBAL HEALTH OBSERVATORY

[9] S. H. Preston, P. Heuveline, and M. Guillot, Demography: Measuring and Modeling Population Processes, 1 st Edition (Blackwell, Malden-MA, 2001).

[10] O. Ahmad, A. D. Lopez, and M. Inoue, Bull. World Health Organ. 10 (78), 1175 (2000).

[11]. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018) pp. 27-29

APPENDIX

	Age-group	Male	Female	Total
2	Less than 18(years)	8652	2516	11168
3	18-25(years)	29078	4128	33206
4	25-35(years)	34194	4829	39023
5	35-45(years)	27967	4542	32509
6	45-60(years)	19140	3472	22612
7	60 and Above(years)	7201	1803	9004
8	Age not known(years)	3087	504	3591

	S.No.	States	Road accident	Total deaths (per thousand)			Road accident (per thousand)
2	1	Andhra	21,992	5.3	1.61E-05	0.016095	0.016
3	2	Arunachal	237	6.7	1.73E-07	0.000173	0
4	3	Assam	8,350	6	6.11E-06	0.006111	0.006
5	4	Bihar	10,007	6.4	7.32E-06	0.007324	0.007
6	5	Chhattisgarh	13,899	5.8	1.02E-05	0.010172	0.01
7	6	Goa	3,440	4.3	2.52E-06	0.002518	0.003
8	7	Gujarat	17,046	8	1.25E-05	0.012475	0.012
9	8	Haryana	10,944	3.8	8.01E-06	0.008009	0.008
10	9	Himachal Pradesh	2,873	4.5	2.10E-06	0.002103	0.002
11	10	Jammu & Kashmir	5,796	3.3	4.24E-06	0.004242	0.004
12	11	Jharkhand	5,217	5.9	3.82E-06	0.003818	0.004
13	12	Karnataka	40,658	5.9	2.98E-05	0.029755	0.03
14	13	Kerala	41,111	5.9	3.01E-05	0.030087	0.03
15	14	Madhya Pradesh	50,669	6.9	3.71E-05	0.037082	0.037
16	15	Maharashtra	32,925	4.9	2.41E-05	0.024096	0.024
17	16	Manipur	672	5.4	4.92E-07	0.000492	0
18	17	Meghalaya	482	6.3	3.53E-07	0.000353	0.35
19	18	Mizoram	62	6.9	4.54E-08	4.54E-05	0
20	19	Nagaland	358	5.6	2.62E-07	0.000262	0
21	20	Odisha	11,064	6.7	8.10E-06	0.008097	0.008
22	21	Punjab	6,348	5.5	4.65E-06	0.004646	0.005
23	22	Rajasthan	23,480	4.5	1.72E-05	0.017184	0.017
24	23	Sikkim	162	5.8	1.19E-07	0.000119	0
25	24	Tamil Nadu	57,228	4.1	4.19E-05	0.041882	0.042

	A	B	C	D	E	F	G
1	S.No.	States	Road accident	Total deaths (per thousand)			Road accident (per thousand)
2	1	Andhra	21,992	5.3	1.61E-05	0.01609	0.016
3	2	Arunachal	237	6.7	1.73E-07	0.00017	0
4	3	Assam	8,350	6	6.11E-06	0.00611	0.006
5	4	Bihar	10,007	6.4	7.32E-06	0.00732	0.007
6	5	Chhattisgarh	13,899	5.8	1.02E-05	0.01017	0.01
7	6	Goa	3,440	4.3	2.52E-06	0.00252	0.003
8	7	Gujarat	17,046	8	1.25E-05	0.01247	0.012
9	8	Haryana	10,944	3.8	8.01E-06	0.00801	0.008
10	9	Himachal Pradesh	2,873	4.5	2.10E-06	0.0021	0.002
11	10	Jammu & Kashmir	5,796	3.3	4.24E-06	0.00424	0.004
12	11	Jharkhand	5,217	5.9	3.82E-06	0.00382	0.004
13	12	Karnataka	40,658	5.9	2.98E-05	0.02976	0.03
14	13	Kerala	41,111	5.9	3.01E-05	0.03009	0.03
15	14	Madhya Pradesh	50,669	6.9	3.71E-05	0.03708	0.037
16	15	Maharashtra	32,925	4.9	2.41E-05	0.0241	0.024
17	16	Manipur	672	5.4	4.92E-07	0.00049	0
18	17	Meghalaya	482	6.3	3.53E-07	0.00035	0.35
19	18	Mizoram	62	6.9	4.54E-08	4.54E-05	0
20	19	Nagaland	358	5.6	2.62E-07	0.00026	0
21	20	Odisha	11,064	6.7	8.10E-06	0.0081	0.008
22	21	Punjab	6,348	5.5	4.65E-06	0.00465	0.005
23	22	Rajasthan	23,480	4.5	1.72E-05	0.01718	0.017
24	23	Sikkim	162	5.8	1.19E-07	0.00012	0
25	24	Tamil Nadu	57,228	4.1	4.19E-05	0.04188	0.042
26	25	Telangana	21,570	3.5	1.58E-05	0.01579	0.016
27	26	Tripura	655	7.3	4.79E-07	0.00048	0
28	27	Uttarakhand	1,352	6.9	9.89E-07	0.00099	0.001
29	28	Uttar Pradesh	42,572	6.6	3.12E-05	0.03116	0.031
30	29	West Bengal	10,158	5.9	7.43E-06	0.00743	0.007

Years	Total Number of Road Accidents (in numbers)	Total Number of Persons Killed (in numbers)
1970	114100	14500
1980	153200	24000
1990	282600	54100
1994	325864	64463
1995	351999	70781
1996	371204	74665
1997	373671	76977
1998	385018	79919
1999	386456	81966
2000	391449	78911
2001	405637	80888
2002	407497	84674
2003	406726	85998
2004	429910	92618
2005	439255	94968
2006	460920	105749
2007	479216	114444
2008	484704	119860
2009	486384	125660
2010	499628	134513
2011	497686	142485
2012	490383	138258
2013	486476	137572
2014	489400	139671
2015	501423	146133
2016	480652	150785
2017	464910	147913
2018	467044	151417
2019	449002	151113

S.No.	States	2015	2016	2017	2018	2019
1	Andhra Pradesh	8297	8541	8060	7556	7984
2	Arunachal Pradesh	127	149	110	175	127
3	Assam	2397	2572	2783	2966	3208
4	Bihar	5421	4901	5554	6729	7205
5	Chhattisgarh	4082	3908	4136	4592	5003
6	Goa	311	336	328	262	297
7	Gujarat	8119	8136	7289	7996	7390
8	Haryana	4879	5024	5120	5118	5057
9	Himachal Pradesh	1096	1271	1203	1208	1146
10	Jammu & Kashmir	917	958	926	984	996
11	Jharkhand	2893	3027	3256	3542	3801
12	Karnataka	10856	11133	10609	10990	10958
13	Kerala	4196	4287	4131	4303	4440
14	Madhya Pradesh	9314	9646	10177	10706	11249
15	Maharashtra	13212	12935	12264	13261	12788
16	Manipur	139	81	136	134	156
17	Meghalaya	183	150	182	182	179
18	Mizoram	72	70	60	45	48
19	Nagaland	30	46	41	39	26
20	Odisha	4303	4463	4790	5315	5333
21	Punjab	4893	5077	4463	4740	4525
22	Rajasthan	10510	10465	10444	10320	10563
23	Sikkim	70	85	78	85	73
24	Tamil Nadu	15642	17218	16157	12216	10525
25	Telangana	7110	7219	6596	6603	6964
26	Tripura	158	173	161	213	239
27	Uttarakhand	913	962	942	1,047	867
28	Uttar Pradesh	17666	19320	20124	22256	22655

Type_of_road_accident	2015	2016	2017	2018	2019
Fatal accident	131726	136071	134796	137726	137689
Grievous injury accidents	119,668	120848	120971	125311	126759
Minor injury accidents	192634	187642	174400	169920	157215
Non-injury accidents	57395	36091	34743	34087	27339

S.No.	States	2015	2017	2019
1	Andhra	24258	25727	21992
2	Arunachal	284	241	237
3	Assam	6959	7170	8350
4	Bihar	9555	8855	10007
5	Chhattisgarh	14446	13563	13899
6	Goa	4338	3917	3440
7	Gujarat	23183	19081	17046
8	Haryana	11174	11258	10944
9	Himachal Pradesh	3010	3114	2873
10	Jammu & Kashmir	5836	5624	5796
11	Jharkhand	5162	5198	5217
12	Karnataka	44011	42542	40658
13	Kerala	39014	38470	41111
14	Madhya Pradesh	54947	53399	50669
15	Maharashtra	63805	35853	32925
16	Manipur	671	578	672
17	Meghalaya	606	675	482
18	Mizoram	70	68	62
19	Nagaland	54	531	358
20	Odisha	10542	10855	11064
21	Punjab	6702	6273	6348
22	Rajasthan	24072	22112	23480
23	Sikkim	219	196	162
24	Tamil Nadu	69059	65562	57228
25	Telangana	21252	22484	21570
26	Tripura	647	503	655
27	Uttarakhand	1523	1603	1352
28	Uttar Pradesh	32385	38783	42572
29	West Bengal	13208	11631	10158

S.No.	Traffic rules violation	Number of accidents(2018)	Persons Killed(2018)	Persons injured(2018)	Number of accidents(2019)	Persons Killed(2019)	Persons injured(2019)
1	Over-speeding	310612	97588	316421	319028	101723	326850
2	Drunken driving/consumpti on of alcoh	12018	4188	9944	12256	5325	10564
3	Driving on wrong side/ Lane indiscipline	24781	8764	24100	24431	9201	24628
4	Jumping red light	4441	1545	4126	4443	1797	4006
5	Use of mobile phone	9039	3707	7878	10522	4945	8144
6	Others*	106150	35625	106949	78322	28122	77169

Obesity survey

ID	What is yc	Select you	Do you try	How ofter	Select the	Which of	Non- vege	Pizza,Burg	How ofter	What	How ofter	Hypothyrc	Are you av	Who ar
1	Male	20- 30 yea	Yes	2-4 times	Fit	5	3	3	Often	Juices;Cof	Sometime	No	Yes	Parents
2	Male	20- 30 yea	Yes	Daily	Fit	1	1	1	Always	Tea ;	Frequentl	No	Yes	None o
3	Female	40 - 60 yea	Maybe	Daily	Healthy	2	5	5	Always	Tea ;Coffe	Frequentl	No	Yes	Grandp
4	Male	20- 30 yea	Yes	Once a we	Obese	4	4	1	Always	Coffee;	Never	No	Yes	None o
5	Female	20- 30 yea	Yes	Once a we	Overweig	5	2	1	Often	Tea ;Coffe	Sometime	Yes	Yes	Grandp
6	Female	20- 30 yea	No	Once a we	Underwei	3	4	3	Always	Juices;Col	Never	No	Yes	None o
7	Male	20- 30 yea	Yes	Once a we	Overweig	5	1	2	Always	Tea ;Coffe	Never	No	No	None o
8	Female	20- 30 yea	Yes	2-4 times	Slim	5	2	3	Always	Tea ;Juice	Never	No	Yes	Grandp
9	Female	20- 30 yea	Yes	Daily	Underwei	5	5	4	Always	Tea ;Coffe	Sometime	No	No	Parents
10	Female	20- 30 yea	Yes	Once a we	Healthy	4	4	1	Always	Green tea	Never	No	Maybe	None o
11	Female	20- 30 yea	Yes	Once a we	Healthy	5	1	3	Always	Juices;	Never	Yes	Yes	Parents
12	Female	20- 30 yea	No	Once a we	Overweig	4	5	2	Often	Tea ;	Sometime	Yes	Yes	None o
13	Male	Less than	No	2-4 times	Fit	3	4	1	Always	Milk;Tea ;	Never	No	Maybe	None o
14	Male	Less than	No	2-4 times	Fit	4	4	5	Always	Coffee;Mi	Sometime	No	No	Grandp
15	Male	20- 30 yea	Yes	2-4 times	Fit	5	5	5	Sometime	Green tea	Never	May be	Maybe	Son or
16	Female	20- 30 yea	Yes	Once a we	Slim	1	1	2	Always	Milk;	Never	No	No	None o
17	Female	Less than	Yes	2-4 times	Underwei	4	1	2	Often	Milk;	Never	No	Yes	Grandp
18	Male	20- 30 yea	Maybe	Once a we	Overweig	1	1	1	Sometime	Tea ;	Never	No	Maybe	Relativ
19	Male	40 - 60 yea	Yes	Once a we	Fit	5	3	1	Always	Tea ;	Never	No	Yes	Relativ
20	Male	20- 30 yea	No	Once a we	Healthy	5	5	1	Always	Milk;	Sometime	No	No	Relativ
21	Female	Less than	Yes	2-4 times	Healthy	4	3	2	Always	Green tea	Never	No	Yes	None o
22	Male	20- 30 yea	Yes	Once a we	Healthy	1	4	5	Always	Milk;	Never	No	Yes	Relativ
23	Female	Less than	Yes	2-4 times	Fit	4	4	2	Always	Milk;	Never	No	Yes	None o
24	Female	20- 30 yea	No	2-4 times	Healthy	4	2	2	Sometime	Coffee;Jui	Never	No	No	None o

State/Union Territory	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Andaman & Nicobar Islands	19	27	31	34	31	27	25	23	24	24	22	20	16	14	9
Andhra Pradesh	59	57	56	54	52	49	46	43	41	39	39	37	34	32	29
Arunachal Pradesh	38	37	40	37	32	32	31	32	33	32	30	30	36	42	37
Assam	66	68	67	66	64	61	58	55	55	54	49	47	44	44	41
Bihar	61	61	60	58	56	52	48	44	43	42	42	42	38	35	32
Chandigarh	21	19	23	27	28	25	22	20	20	21	23	21	14	14	13
Chhattisgarh	60	63	61	59	57	54	51	48	47	46	43	41	39	38	41
Dadra & Nagar Haveli	48	42	35	34	34	37	38	35	33	31	26	21	17	13	13
Daman & Diu	37	28	28	27	31	24	23	22	22	20	18	18	19	17	16
Delhi	32	35	37	36	35	33	30	28	25	24	20	18	18	16	13
Goa	17	16	15	13	10	11	10	11	10	9	10	9	8	9	7
Gujarat	53	54	53	52	50	48	44	41	38	36	35	33	30	30	28
Haryana	61	60	57	55	54	51	48	44	42	41	36	36	33	30	30
Himachal Pradesh	51	49	50	47	44	45	40	38	36	35	32	28	25	22	19
Jammu and Kashmir	49	50	52	51	49	45	43	41	39	37	34	26	24	23	22
Jharkhand	49	50	49	48	46	44	42	39	38	37	34	32	29	29	30
Karnataka	49	50	48	47	45	41	38	35	32	31	29	28	24	25	23
Kerala	12	14	15	13	12	12	13	12	12	12	12	12	10	10	7
Lakshadweep	30	22	25	24	31	25	25	24	24	24	20	20	19	20	14
Madhya Pradesh	79	76	74	72	70	67	62	59	56	54	52	50	47	47	48
Maharashtra	36	36	35	34	33	31	28	25	25	24	22	21	19	19	19
Manipur	14	13	11	12	14	16	14	11	10	10	11	9	11	12	11
Meghalaya	54	49	53	56	58	59	55	52	49	47	46	42	39	39	33
Mizoram	19	20	25	23	37	36	37	34	35	35	32	32	27	15	5
Nagaland	17	18	20	21	26	26	23	21	18	18	14	12	12	7	4
Odisha	77	75	73	71	69	65	61	57	53	51	49	46	44	41	40
Puducherry	24	28	28	25	25	22	22	19	17	17	14	11	10	11	11
Punjab	45	44	44	43	41	38	34	30	28	26	24	23	21	21	20
Rajasthan	67	68	67	65	63	59	55	52	49	47	46	43	41	38	37
Sikkim	32	30	33	34	33	34	30	26	24	22	19	18	16	12	7
Tamil Nadu	41	37	37	35	31	28	24	22	21	21	20	19	17	16	15
Telangana	.	.	.	.	.	.	.	.	.	.	35	34	31	29	27
Tripura	32	31	36	39	34	31	27	29	28	26	21	20	24	29	27
Uttar Pradesh	72	73	71	69	67	63	61	57	53	50	48	46	43	41	43
Uttarakhand	42	42	43	48	44	41	38	36	34	32	33	34	38	32	31
West Bengal	40	38	38	37	35	33	33	32	32	31	28	26	25	24	22
ALL INDIA	58	58	57	55	53	50	47	44	42	40	39	37	34	33	32