*A project report on*

# PREDICTIVE ANALYSIS OF DEPRESSION USING INTERNET BEHAVIOUR PATTERNS

*Submitted in partial fulfillment for the award of the degree of*

**MCA**

*by*

**SWARUP DAS (17MCA0008)**

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

**SITE**

April,2019

# PREDICTIVE ANALYSIS OF DEPRESSION USING INTERNET BEHAVIOUR PATTERNS

*Submitted in partial fulfillment for the award of the degree of*

**MCA**

*by*

**SWARUP DAS (17MCA0008)**

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

**SITE**

April,2019

## DECLARATION

I hereby declare that the thesis entitled “PREDICTIVE ANALYSIS OF DEPRESSION USING INTERNET BEHAVIOUR PATTERNS” submitted by me, for the award of the degree of Specify the name of the degree VIT  is a record of bonafide work carried out by me under the supervision of Gide Name

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Vellore

Date: Signature of the Candidate

# CERTIFICATE

This is to certify that the thesis entitled “PREDICTIVE ANALYSIS OF DEPRESSION USING INTERNET BEHAVIOUR PATTERNS” submitted by SWARUP DAS (17MCA0008) SITE VIT, for the award of the degree of MCA is a record of bonafide work carried out by him/her under my supervision.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or

any other institute or university. The Project report fulfils the requirements and regulations of

VIT and in my opinion meets the necessary standards for submission.

**Signature of the Guide** **Signature of the Hod**

**Internal Examiner** **External Examiner**

# ABSTRACT

The overall number of deaths by suicide per year is around 800,000 people. Out of that almost 230,000 are from India. With a rising interest in social media alongside the overall difficulty in monitoring of such internet usage has led to a number of reasons that may affect a teenager's mental health. This may continue to a larger age group as the exposure is a basic human right but the effect it has on social wellbeing is rarely accounted for. This project aims to hopefully provide a platform that can start off attempts and research into the coping factor for those that are displaying signs of depression on internet analysis, be it through the usage of text, audio visual media, memes or posting.

# ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Prof. Deepikaa S, Professor, SITE, Vellore Institute of Technology, for her constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Compter Science and Applications

I would like to express my gratitude to Dr. G Vishwanathan, Dr. Sekar Viswanathan, Dr. S. Narayanan, and Dr. Balakrushna Tripathy, SITE, for providing with an environment to work in and for his inspiration during the tenure of the course.

In jubilant mood, I express ingeniously my whole-hearted thanks to Dr. Ramkumar T, Head of Department, all teaching staff and members working as limbs of our university for their not-self-centered enthusiasm coupled with timely encouragements showered on me with zeal, which prompted the acquirement of the requisite knowledge to finalize my course study successfully. I would like to thank my parents for their support.

It is indeed a pleasure to thank my friends who persuaded and encouraged me to take up and complete this task. At last but not least, I express my gratitude and appreciation to all those who have helped me directly or indirectly toward the successful completion of this project.

Place: Vellore

Date: Swarup Das (17MCA0008)

# CONTENTS

**CHAPTER 1**

**LIST OF FIGURES**

**INTRODUCTION**

**CHAPTER 2**

**BACKGROUND AND WORKINGS**

**CHAPTER 3**

**RESULTS**

## CHAPTER 3

## CONLUSION

# Introduction

The prject aims to create a Naïve Bayesian classifier from a given dataset that we have arranged from a Reddit survey. The data has been put through various layers of preprocessing and filtering and then finally the data will be put through an algorithm that determines the Naïve Bayesian classifier values based on the relative frequencies. This helps determine the probability of a student on the verge of suicide. Additionally the results from a testing dataset is used to determine the efficiency of the algorithm. The original dataset has also been put through Orange application to see some data charts about the information. This has all been presented in the results and conclusion phase.

# Literature Survey

Despite the fact that suicide is a profoundly close to home and an individual demonstration, self-destructive conduct is dictated by various individual and social elements. As far back as Esquirol composed that "Every one of the individuals who submitted suicide are crazy" and Durkheim recommended that suicide was a result of social/societal circumstances, the discussion of individual weakness versus social stressors in the causation of suicide has isolated our contemplations on suicide. Suicide is best comprehended as a multidimensional, multifactorial disquietude. Suicide is seen as a social issue in our nation and consequently, mental confusion is given equivalent calculated status with family clashes, social maladjustment and so on.

**[1]** In spite of the extent of the test presented in self harm, what happens is that have a moderately inadequate comprehension in decisively offers ascend in a self harm hazard. To counteract such levels of suicidal self harm, we require a superior comprehension of the fundamental marvels identifying with both the impending danger of suicide (or intense self-destructive hazard) and the long haul dangers. For the two cases, information is incredibly meager, never continuously, and subject to some predisposition. Hardly any target estimates exist to gauge results, also, those that do exist will in general have poor worldly goals (estimated in weeks or months) and are work serious. Streamlining mediation viability or approach level methodologies is troublesome without such information. **[2]** Language, specifically, has turned out to be an intense focal point for the examination of psychological well-being, as proven by extensive usage of the Linguistic Inquiry Word Count. The current techniques for evaluating the occasions encompassing self-destructive emergency bringing about a suicide endeavor are vigorously defenseless to review inclination.

**[3]** In perspective of the vital job of online portrayals may major role in reaching out to helpless people to endeavor suicide, impressive consideration has been given to the conceivable prevention based impacts of proper revealing of hearing about these news on various online channels. Without a doubt, considers have recognized a decline in suicides following the usage of media. **[4]** Generally, the increments in emotional well-being issues showed up crosswise over gatherings paying little mind to race/ethnicity, SES, locale, and age/review. Around 2010, teenagers invested more energy in internet based life and electronic gadgets, exercises decidedly associated with burdensome side effects and suicide-related results. Over that years, young people invested less energy in non-screen exercises, for example, personal platformer collaboration, print media, sports/exercise, and going to religious administrations, exercises contrarily corresponded with burdensome manifestations.

# NAÏVE BAYESIAN CLASSIFIER

There was extensive studies done under this classifier. Under a different name at that time it was introduced as a mainstream algorithm around the 1960s. **[5]** With fitting pre-preparing, it is aggressive in this space with further developed strategies including bolster vector machines **[6]**

The Naive Bayesian classifier relies upon Bayes' theory with the self-rule doubts between markers. A Naive Bayesian model is definitely not hard to work, with no befuddled iterative parameter estimation which makes it particularly important for immense datasets. Notwithstanding its straightforwardness, the Naive Bayesian classifier consistently does incredibly well and is commonly used in light of the way that it regularly outmaneuvers dynamically propelled request systems.

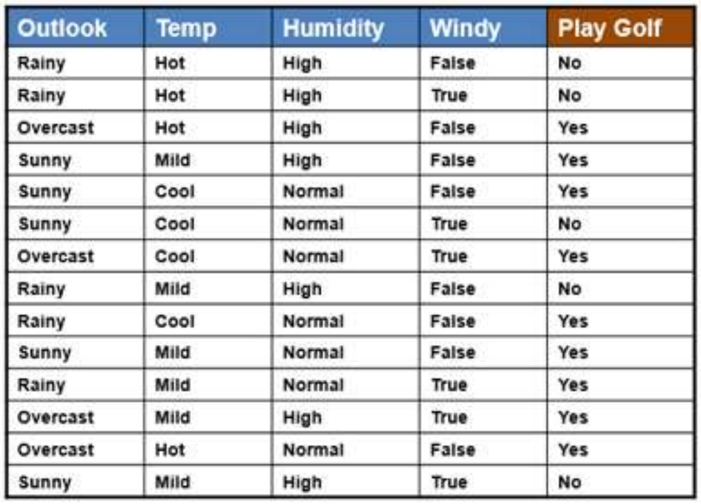
## Algorithm

We can calculate posterior probability using Bayesian method, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. It based on the fact that the predictor ($x$) on a given class ($c$) is independent in comparison to the values of other predictors. The term for this assumption is called class conditional independence.

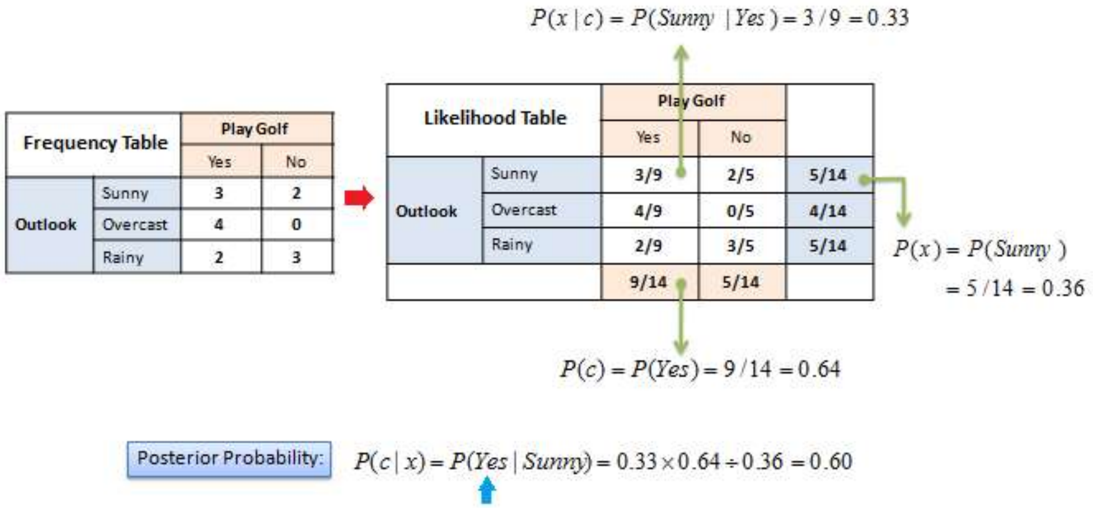
$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

An example of the classifier is given below



| Outlook | Temp | Humidity | Windy | Play Golf |
|---|---|---|---|---|
| Rainy | Hot | High | False | No |
| Rainy | Hot | High | True | No |
| Overcast | Hot | High | False | Yes |
| Sunny | Mild | High | False | Yes |
| Sunny | Cool | Normal | False | Yes |
| Sunny | Cool | Normal | True | No |
| Overcast | Cool | Normal | True | Yes |
| Rainy | Mild | High | False | No |
| Rainy | Cool | Normal | False | Yes |
| Sunny | Mild | Normal | False | Yes |
| Rainy | Mild | Normal | True | Yes |
| Overcast | Mild | High | True | Yes |
| Overcast | Hot | Normal | False | Yes |
| Sunny | Mild | High | True | No |



$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$

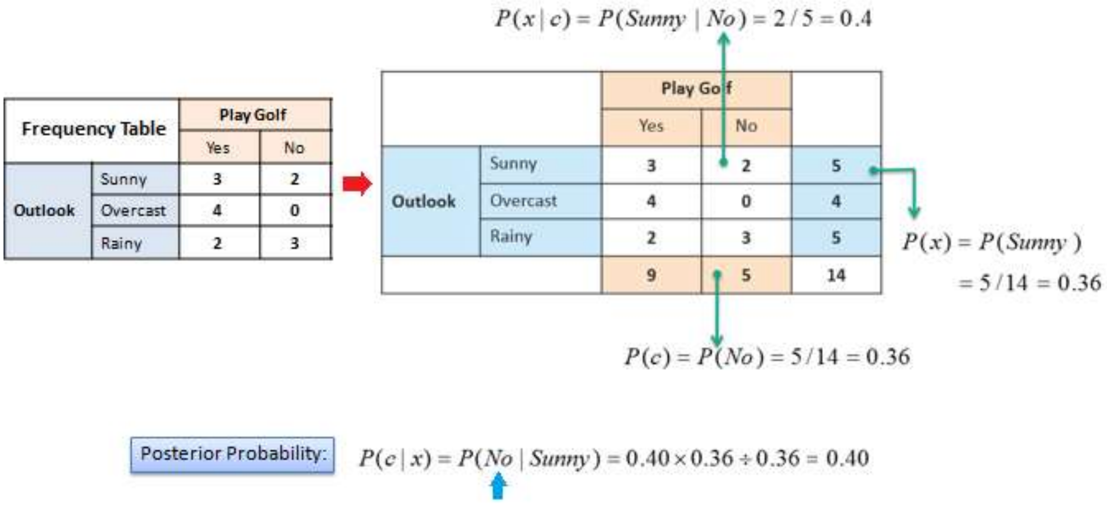
| Frequency Table | | Play Golf | |
|---|---|---|---|
| | | Yes | No |
| Outlook | Sunny | 3 | 2 |
| | Overcast | 4 | 0 |
| | Rainy | 2 | 3 |

| Likelihood Table | | Play Golf | | |
|---|---|---|---|---|
| | | Yes | No | |
| Outlook | Sunny | 3/9 | 2/5 | 5/14 |
| | Overcast | 4/9 | 0/5 | 4/14 |
| | Rainy | 2/9 | 3/5 | 5/14 |
| | | 9/14 | 5/14 | |

$P(x) = P(Sunny)$
$= 5/14 = 0.36$

$P(c) = P(Yes) = 9/14 = 0.64$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$



$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$

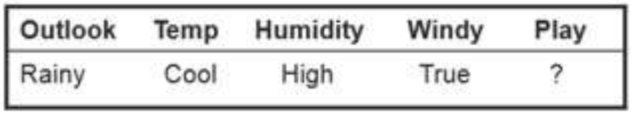
| Frequency Table | | Play Golf | |
|---|---|---|---|
| | | Yes | No |
| Outlook | Sunny | 3 | 2 |
| | Overcast | 4 | 0 |
| | Rainy | 2 | 3 |

| | | Play Golf | | |
|---|---|---|---|---|
| | | Yes | No | |
| Outlook | Sunny | 3 | 2 | 5 |
| | Overcast | 4 | 0 | 4 |
| | Rainy | 2 | 3 | 5 |
| | | 9 | 5 | 14 |

$P(x) = P(Sunny)$
$= 5/14 = 0.36$

$P(c) = P(No) = 5/14 = 0.36$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$



| Outlook | Temp | Humidity | Windy | Play |
|---|---|---|---|---|
| Rainy | Cool | High | True | ? |

$$P(Yes \mid X) = P(Rainy \mid Yes) \times P(Cool \mid Yes) \times P(High \mid Yes) \times P(True \mid Yes) \times P(Yes)$$

$$P(Yes \mid X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = \boxed{0.00529} \longrightarrow 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

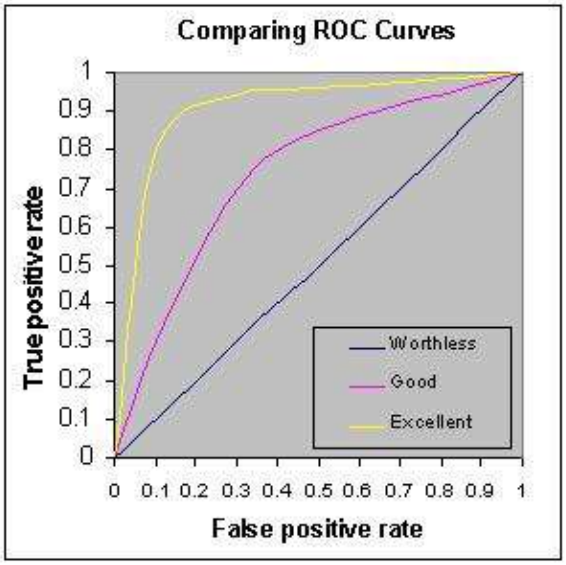
$$P(No \mid X) = P(Rainy \mid No) \times P(Cool \mid No) \times P(High \mid No) \times P(True \mid No) \times P(No)$$

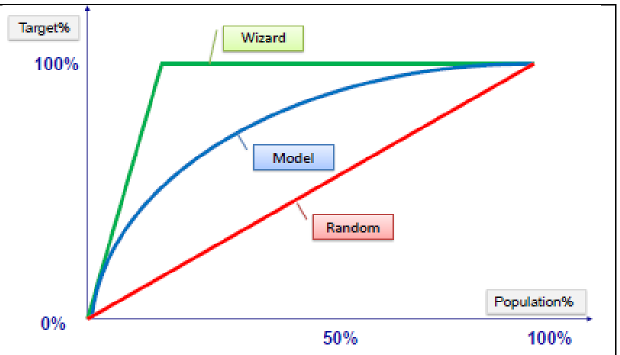
$$P(No \mid X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = \boxed{0.02057} \longrightarrow 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

## Graphs Obtained from Naïve Bayesian

We utilize the ROC and Lift charts available in R to display the results and the efficiency of the code.

ROC Chart: A receiver operating characteristic, or ROC bend, is a graphical plot that delineates the indicative capacity of a paired classifier framework as its separation edge is shifted. The ROC bend is made by plotting the genuine positive rate against the bogus positive rate at different limit settings.



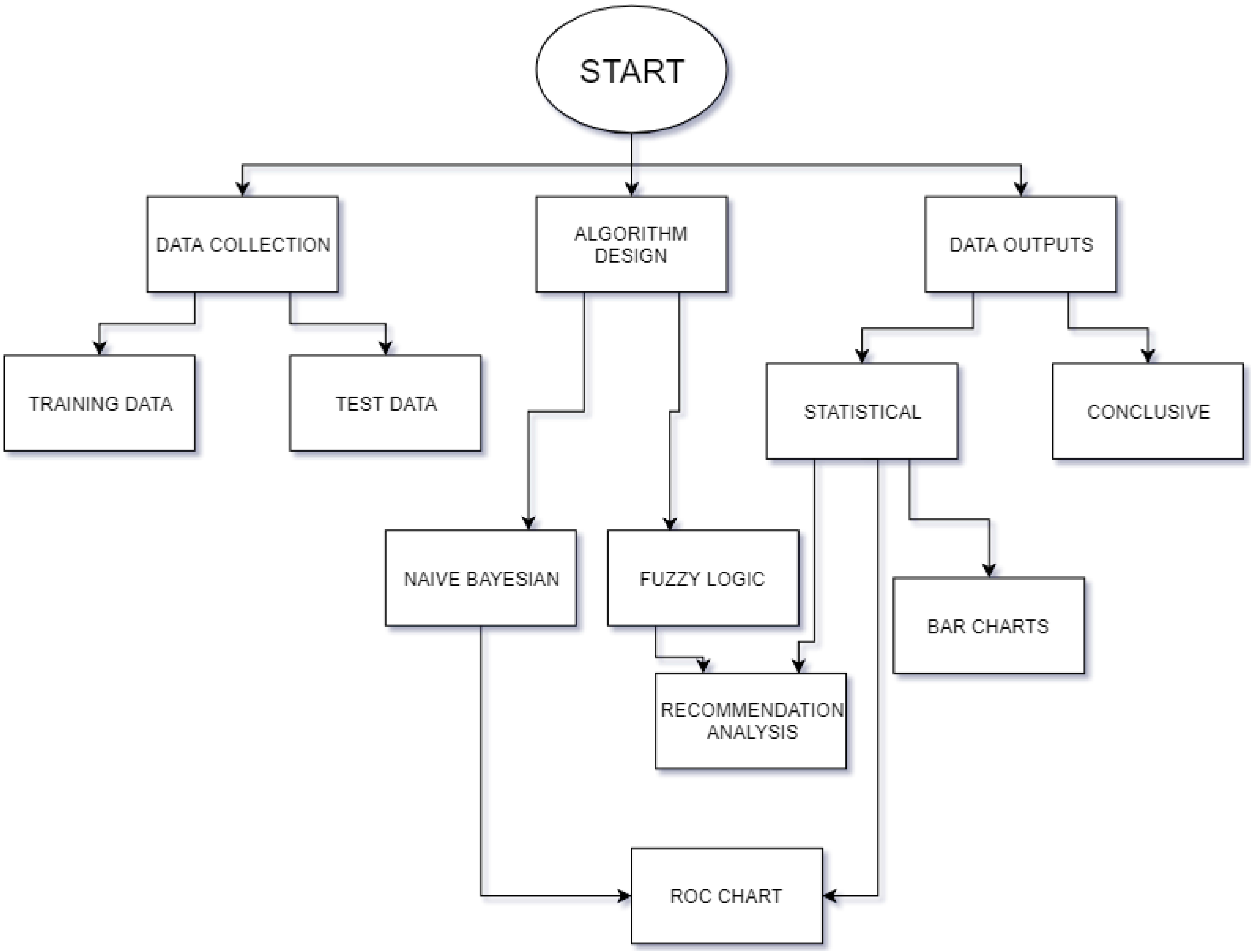


Lift Chart: Lift is a proportion of the effectiveness of a predicitve model determined as the proportion between the outcomes acquired with and without the prescient model.

# Work Plan and System Design

The algorithms utilized in the process will be the traditional Naïve Bayesian Classifier, i.e probability based design and the second process will involve Fuzzy Logic values to determine and classify users into categorized regions of mental health states. This data is then plotted to determine from them various conclusions and references.

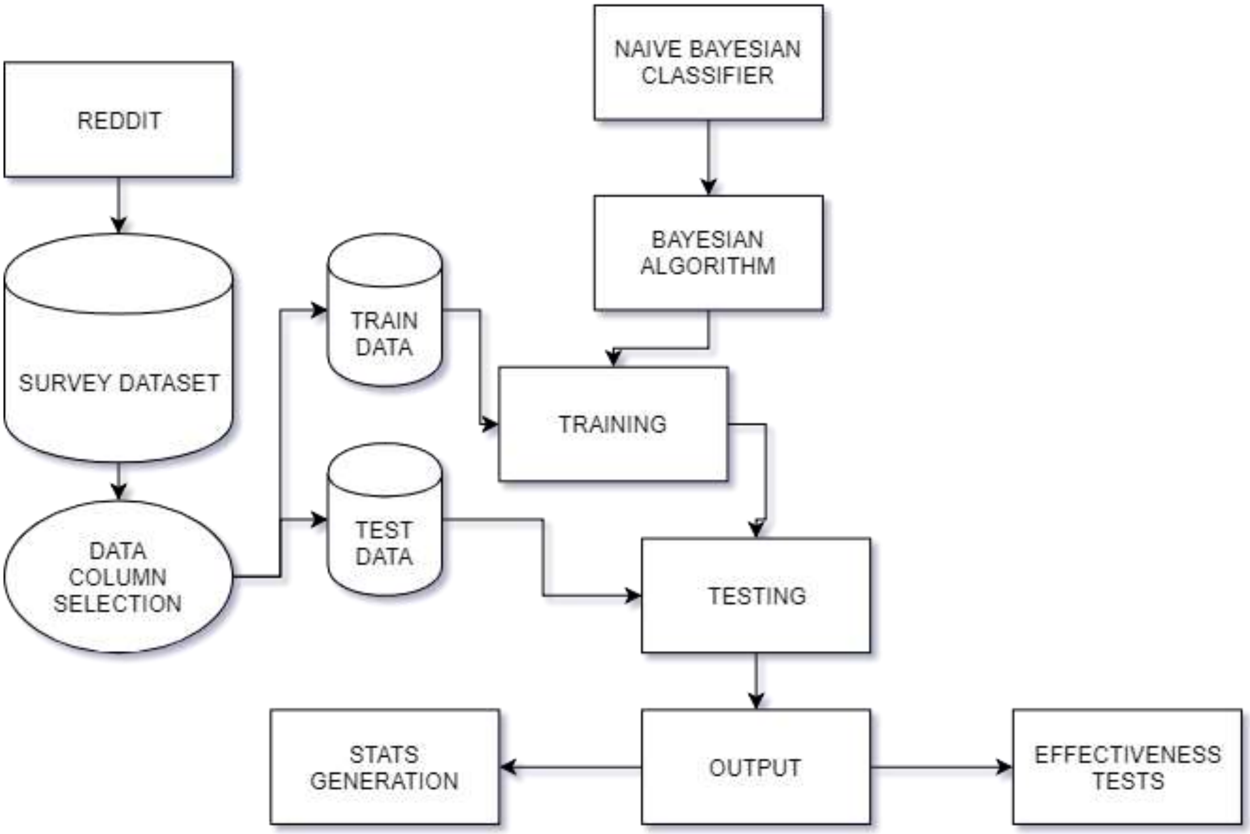
WORKFLOW DESIGN:



DATABASE SCHEMA:

| gender | sexuality | race | bodyweight | social_fear | depressed | attempt_suicide |
|---|---|---|---|---|---|---|

NAÏVE BAYESIAN CLASSIFIER:



# Methodology

The entirety of the project in done and coded in R. The software is useful and intuitive with simple libraries that simplify the entire data processing work. The main data was inserted into excel sheets and exported in a csv format. The csv was then imported and used into databases that store and save the individual values in variables in the software.

The data went through a preprocessing stage where all the information was removed of outliers. This was specially applicable to factors where the members of each column were filled with garbage data.

Even in the original dataset, the number of columns i.e factors was reduced and boiled down to factors that are entirely essential for a young member of society and more prone to suicide. The original dataset of the website included the following information:

Time, gender, sexuality, age, income, race, bodyweight, virgin, prostitution_legal, pay_for_sex, friends, social_fear, depressed, what_help_from_others, attempt_suicide, employment, job_title, edu_level, improve_yourself_how

The rest of the data preprocessing was done in the columns themselves where garbage information entered led to the entire column being removed. After these, the final dataset had only the essential columns and seriously filled survey data.

Orange is a part based visual programming bundle for information perception, AI, information mining, and information investigation.

Orange parts are called gadgets and they extend from basic information perception, subset choice, and preprocessing, to exact assessment of learning calculations and prescient demonstrating.

Visual writing computer programs is actualized through an interface in which work processes are made by connecting predefined or client structured gadgets, while propelled clients can utilize Orange as a Python library for information control and gadget adjustment.

Orange is an open-source programming bundle discharged under GPL. Forms up to 3.0 incorporate center parts in C++ with wrappers in Python are accessible on GitHub. From variant 3.0 onwards, Orange uses basic Python open-source libraries for logical registering, for example, numpy, scipy and scikit-learn, while its graphical UI works inside the cross-stage Qt system. Orange3 has its own different github.

The default establishment incorporates various AI, preprocessing and information perception calculations in 6 gadget sets (information, envision, group, relapse, assess and unsupervised). Extra functionalities are accessible as additional items (bioinformatics, information combination and content mining).

Orange is bolstered on macOS, Windows and Linux and can likewise be introduced from the Python Package Index vault (Pip introduce Orange3).

As of May 2018 the steady form is 3.13 and keeps running with Python 3, while the heritage variant 2.7 that keeps running with Python 2.7 is as yet accessible.

Orange comprises of a canvas interface onto which the client places gadgets and makes an information investigation work process. Gadgets offer essential functionalities, for example, perusing the information, demonstrating an information table, choosing highlights, preparing indicators, contrasting learning calculations, envisioning information components, and so on. The client can intelligently investigate perceptions or feed the chose subset into different gadgets.

Characterization Tree gadget in Orange 3.0

Canvas: graphical front-end for information investigation

Gadgets:

Information: gadgets for information input, information sifting, inspecting, attribution, highlight control and highlight determination

Picture: gadgets for normal representation (box plot, histograms, disperse plot) and multivariate perception (mosaic showcase, sifter chart).

Order: a lot of directed AI calculations for arrangement

Relapse: a lot of managed AI calculations for relapse

Assess: cross-approval, testing based systems, unwavering quality estimation and scoring of expectation strategies

Unsupervised: unsupervised learning calculations for grouping (k-implies, various leveled bunching) and information projection methods (multidimensional scaling, chief segment examination, correspondence investigation).

Additional items:

Partner: gadgets for mining regular itemsets and affiliation rule learning

Bioinformatics: gadgets for quality set investigation, advancement, and access to pathway libraries

Information combination: gadgets for melding distinctive informational indexes, aggregate lattice factorization, and investigation of dormant variables

Instructive: gadgets for encouraging AI ideas, for example, k-implies bunching, polynomial relapse, stochastic angle drop, ...

Geo: gadgets for working with geospatial information

Picture examination: gadgets for working with pictures and ImageNet embeddings

System: gadgets for chart and system examination

Content mining: gadgets for common language handling and content mining

Time arrangement: gadgets for time arrangement investigation and displaying

The program gives a stage to test determination, suggestion frameworks, and prescient demonstrating and is utilized in biomedicine, bioinformatics, genomic research, and instructing. In science, it is utilized as a stage for testing new AI calculations and for executing new methods in hereditary qualities and bioinformatics. In training, it was utilized for encouraging AI and information mining techniques to understudies of science, biomedicine, and informatics.

Scripting dialects have as of late ascended in prominence in all fields of software engineering. Inside the setting of explorative information investigation, they offer favorable circumstances like intuitiveness and quick prototyping by sticking together existing parts or adjusting them for new assignments. Python is a scripting language with clear and basic language structure, which additionally made it famous in training. Its generally moderate execution can be dodged by utilizing libraries that execute the computationally concentrated undertakings in lowlevel dialects.

Python offers countless libraries. Many are identified with AI, counting a few general bundles like scikit-learn (Pedregosa et al., 2011), PyBrain (Schaul et al.,

2010) and mlpy (Albanese et al., 2012). Orange was considered in late 1990s and is among the most seasoned of such instruments. It centers around effortlessness, intuitiveness through scripting, and part based plan.

Orange library is a progressively sorted out tool stash of information mining segments. The low-level methodology at the base of the pecking order, similar to information sifting, likelihood evaluation and highlight scoring, are amassed into larger amount calculations, for example, characterization tree learning. This permits designers to effortlessly include new usefulness at any dimension and breaker it with the current code.

The library is intended to improve the gathering of information investigation work processes and making of information mining comes closer from a mix of existing segments. Other than more extensive scope of highlights,

Orange contrasts from most other Python-based AI libraries by its development (more than 15

long periods of dynamic improvement and use), a substantial client network bolstered through a functioning discussion,

what's more, broad documentation that incorporates instructional exercises, scripting models, informational index vault, and documentation for designers. Orange scripting library is additionally an establishment for its visual programming

stage with graphical UI parts for intuitive information representation.

The two noteworthy bundles that are like Orange are still effectively created are scikitlearn (Pedregosa et al., 2011) and mlpy (Albanese et al., 2012). Both are all the more firmly incorporated with numpy and at present better mix into Python's numerical processing living space. Orange was on the other hand roused by established AI that centers around emblematic techniques. Instead of

supporting just numerical clusters, Orange information structures consolidate emblematic, string and numerical characteristics and meta information data. Client can for example allude to factors and qualities by their names. Factors store mapping capacities, a component which for example enables classifiers to characterize changes on preparing information that are then naturally connected when making expectations.

These highlights additionally make Orange increasingly appropriate for intuitive, explorative information examination.

Orange's center is a gathering of almost 200 C++ classes that spread the fundamental information structures and greater part of preprocessing and displaying calculations. The C++ part is independent, with no

calls to Python that would initiate pointless overhead. The center incorporates a few open source

libraries, including LIBSVM (Chang and Lin, 2011), LIBLINEAR (Fan et al., 2008), Earth (see http://www.milbo.users.sonic.net/earth), QHull (Barber et al., 1996) and a subset of BLAS

(Blackford et al., 2002). The Python layer additionally utilizes prevalent Python libraries numpy for direct variable based math, networkx (Hagberg et al., 2008) for working with systems and matplotlib (Hunter, 2007) for essential perception.

The upper layer of Orange is written in Python and incorporates methods that are not time-basic.

This is additionally the spot at which clients outside the center improvement bunch most effectively add to the venture.

Robotized testing of the framework depends on more than 1,500 relapse tests that are for the most part dependent on

code bits from broad documentation. A piece of the code is additionally secured with stricter unit tests.

Orange is free programming discharged under GPL. The code is facilitated on Bitbucket vault (https:// bitbucket.org/biolab/orange). Orange keeps running on Windows, Mac OS X and Linux, and can likewise

be introduced from the Python Package Index vault (pip introduce Orange). Parallel installer

for Windows and application group for Mac OS X are accessible on task's site (http: /orange.biolab.si).

Orange right now keeps running on Python 2.6 and 2.7. An adaptation for Python 3 and higher is under improvement. There, we will change to numpy-based information structures and scrap the C++ center in support of utilizing schedules from numpy and scipy (Jones et al., 2001– ), scikit-learn (Pedregosa et al., 2011) and comparable libraries that did not exist when Orange was first considered. In spite of arranged

changes in the center, we will keep up in reverse similarity. For existing clients, the progressions of the

Python interface will be minor.

R is a programming language and free programming condition for measurable registering and designs bolstered by the R Foundation for Statistical Computing. The R language is generally utilized among analysts and information excavators for creating factual programming and information investigation. Surveys, information mining studies, and investigations of insightful writing databases show considerable increments in notoriety as of late. as of March 2019, R positions fourteenth in the TIOBE record, a proportion of fame of programming dialects.

A GNU bundle, source code for the R programming condition is composed fundamentally in C, Fortran and R itself, and is unreservedly accessible under the GNU General Public License. Pre-aggregated double forms are accommodated different working frameworks. Despite the fact that R has an order line interface, there are a few graphical UIs, for example, RStudio, a coordinated improvement condition.

R is an execution of the S programming language joined with lexical perusing semantics, roused by Scheme. S was made by John Chambers in 1976, while at Bell Labs. There are some critical contrasts, however a significant part of the code composed for S runs unaltered.

R was made by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is at present created by the R Development Core Team (of which Chambers is a part). R is named mostly after the primary names of the initial two R creators and halfway as a play on the name of S. The task was imagined in 1992, with an underlying rendition discharged in 1995 and a steady beta form in 2000.

R and its libraries actualize a wide assortment of measurable and graphical strategies, including direct and nonlinear displaying, established factual tests, time-arrangement examination, order, grouping, and others. R is effectively extensible through capacities and expansions, and the R people group is noted for its dynamic commitments as far as bundles. A large number of R's standard capacities are written in R itself, which makes it simple for clients to pursue the algorithmic decisions made. For computationally concentrated assignments, C, C++, and Fortran code can be connected and called at run time. Propelled clients can compose C, C++, Java, .NET or Python code to control R protests straightforwardly. R is profoundly extensible using client submitted bundles for explicit capacities or explicit regions of study. Because of its S legacy, R has more grounded item situated programming offices than most factual registering dialects. Expanding R is likewise facilitated by its lexical checking rules.

Another quality of R is static designs, which can deliver production quality charts, including numerical images. Dynamic and intelligent designs are accessible through extra bundles.

R has Rd, its own LaTeX-like documentation design, which is utilized to supply complete documentation, both online in various organizations and in printed copy.

Like other comparative dialects, for example, APL and MATLAB, R bolsters framework number juggling. R's information structures incorporate vectors, networks, exhibits, information outlines (like tables in a social database) and records. Exhibits are put away in segment real request. R's extensible item framework incorporates objects for (among others): relapse models, time-arrangement and geo-spatial directions. The scalar information type was never an information structure of R. Rather, a scalar is spoken to as a vector with length one.

Numerous highlights of R get from Scheme. R utilizes S-articulations to speak to the two information and code. Capacities are five star and can be controlled similarly as information objects, encouraging meta-programming, and permit different dispatch. Factors in R are lexically perused and powerfully composed.

R underpins procedural programming with capacities and, for certain capacities, object-arranged programming with nonexclusive capacities. A conventional capacity acts distinctively relying upon the classes of contentions go to it. As it were, the nonexclusive capacity dispatches the capacity (strategy) explicit to that class of article. For instance, R has a nonexclusive print work that can print pretty much every class of article in R with a basic print(objectname) linguistic structure.

Albeit utilized for the most part by analysts and different professionals requiring a domain for factual calculation and programming advancement, R can likewise work as a general network computation tool kit – with execution benchmarks practically identical to GNU Octave or MATLAB.

The capacities of R are stretched out through client made bundles, which permit particular factual strategies, graphical gadgets, import/send out abilities, revealing instruments (knitr, Sweave), and so forth. These bundles are grown basically in R, and some of the time in Java, C, C++, and Fortran.[citation needed] The R bundling framework is additionally utilized by analysts to make compendia to arrange look into information, code and report records in an efficient manner for sharing and open filing.

A center arrangement of bundles is incorporated with the establishment of R, with in excess of 15,000 extra bundles (as of September 2018) accessible at the Comprehensive R Archive Network (CRAN), Bioconductor, Omegahat, GitHub, and different vaults.

The "Undertaking Views" page (subject rundown) on the CRAN site records a wide scope of errands (in fields, for example, Finance, Genetics, High Performance Computing, Machine Learning, Medical Imaging, Social Sciences and Spatial Statistics) to which R has been connected and for which bundles are accessible. R has likewise been recognized by the FDA as reasonable for translating information from clinical research.

Other R bundle assets incorporate Crantastic, a network site for rating and assessing all CRAN bundles, and R-Forge, a focal stage for the community oriented improvement of R bundles, R-related programming, and activities. R-Forge likewise has numerous unpublished beta bundles, and advancement variants of CRAN bundles.

The Bioconductor venture gives R bundles to the examination of genomic information. This incorporates object-situated information dealing with and examination instruments for information from Affymetrix, cDNA microarray, and cutting edge high-throughput sequencing techniques.

The most specific coordinated improvement condition (IDE) for R is RStudio. A comparable advancement interface is R Tools for Visual Studio. Some conventional IDEs like Eclipse, likewise offer highlights to work with R.

Graphical UIs with even more a point-and-snap approach incorporate Rattle GUI, R Commander, and RKWard.

A portion of the more typical editors with fluctuating dimensions of help for R incorporate (Emacs Speaks Statistics), Vim (Nvim-R module), Neovim (Nvim-R module), Kate, LyX, Notepad++, Visual Studio Code, WinEdt, and Tinn-R.

R usefulness is available from a few scripting dialects, for example, Python, Perl, Ruby, F#, and Julia. Interfaces to other, abnormal state programming dialects, similar to Java and .NET C# are accessible also.

The fundamental R usage is written in R, C, and Fortran, and there are a few different executions gone for improving pace or expanding extensibility. A firmly related usage is pqR (truly speedy R) by Radford M. Neal with improved memory the executives and backing for programmed multithreading. Renjin and FastR are Java usage of R for use in a Java Virtual Machine. CXXR, rho, and Riposte are executions of R in C++. Renjin, Riposte, and pqR endeavor to improve execution by utilizing different processor centers and some type of conceded assessment. The greater part of these elective executions are exploratory and fragmented, with generally couple of clients, contrasted with the primary usage kept up by the R Development Core Team.

TIBCO assembled a runtime motor called TERR, which is a piece of Spotfire.

Microsoft R Open is a completely perfect R appropriation with changes for multi-strung calculations.

Despite the fact that R is an open-source venture upheld by the network creating it, a few organizations endeavor to give business support or potentially augmentations for their clients. This segment gives a few instances of such organizations.

In 2007, Richard Schultz, Martin Schultz, Steve Weston and Kirk Mettler established Revolution Analytics to give business backing to Revolution R, their dispersion of R, which likewise incorporates segments created by the organization. Major extra parts include: ParallelR, the R Productivity Environment IDE, RevoScaleR (for enormous information examination), RevoDeployR, web administrations system, and the capacity for perusing and composing information in the SAS document design. Upheaval Analytics likewise offer a dispersion of R intended to follow built up IQ/OQ/PQ criteria which empowers customers in the pharmaceutical division to approve their establishment of REvolution R. In 2015, Microsoft Corporation finished the obtaining of Revolution Analytics. also, has since coordinated the R programming language into SQL Server 2016, SQL Server 2017, Power BI, Azure SQL Database, Azure Cortana Intelligence, Microsoft R Server and Visual Studio 2017.

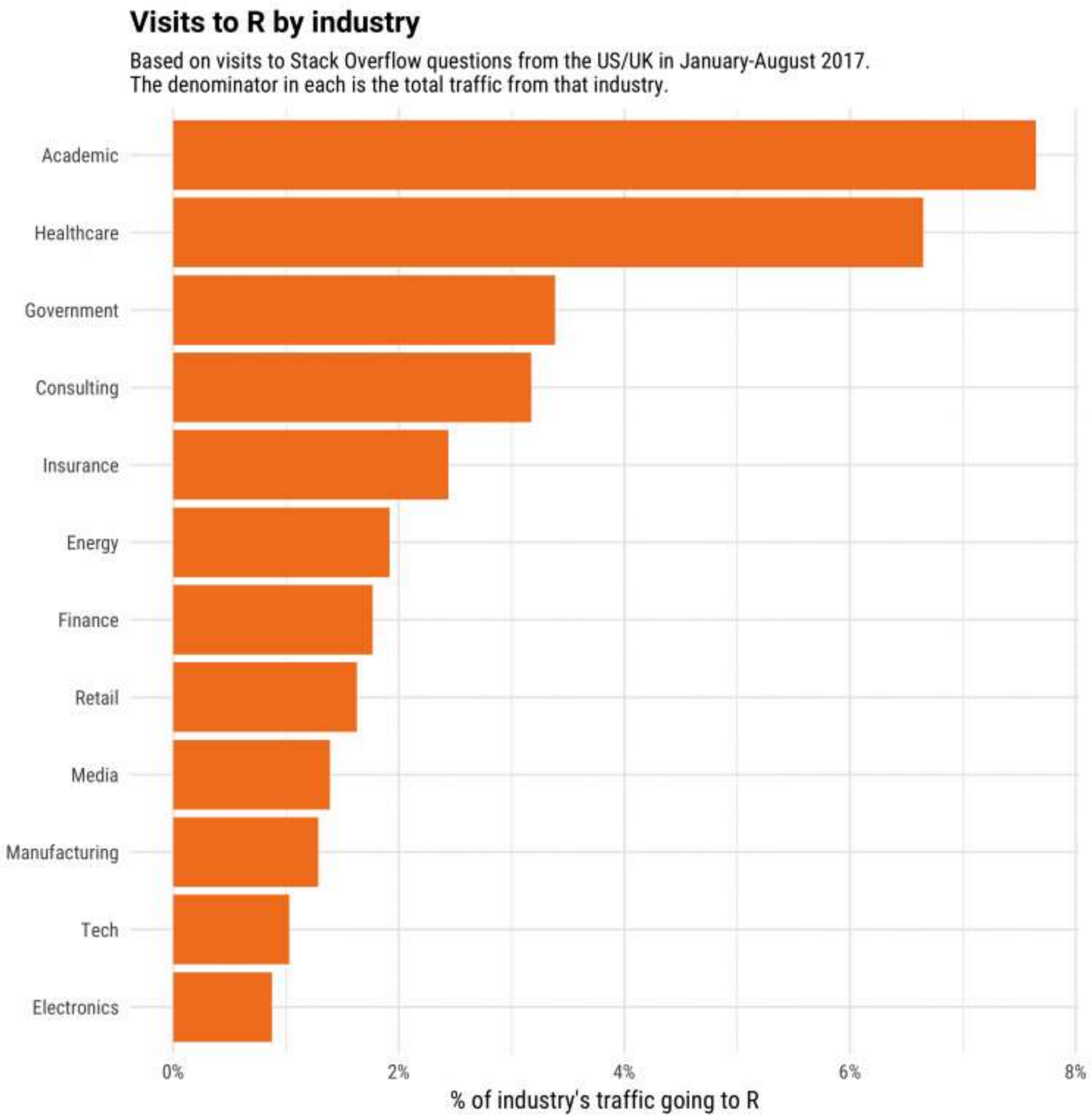
In October 2011, Oracle declared the Big Data Appliance, which incorporates R, Apache Hadoop, Oracle Linux, and a NoSQL database with Exadata equipment. Starting at 2012, Oracle R Enterprise ended up one of two segments of the "Prophet Advanced Analytics Option" (nearby Oracle Data Mining).

IBM offers support for in-Hadoop execution of R, and gives a programming model to hugely parallel in-database examination in R.

Other significant business programming frameworks supporting associations with or coordination with R include: JMP, Mathematica, MATLAB, Microsoft Power BI,Pentaho, Spotfire, SPSS, Statistica, Platform Symphony, SAS, Tableau Software, Esri ArcGIS, Dundas and Statgraphics.

Tibco offers a runtime-rendition R as a piece of Spotfire.

Mango offers an approval bundle for R, ValidR, to make it consistent with medication endorsement organizations, similar to FDA. These offices take into consideration the utilization of any factual programming in entries, if just the product is approved, either by the merchant or support itself.



Visits to R by industry
Based on visits to Stack Overflow questions from the US/UK in January-August 2017.
The denominator in each is the total traffic from that industry.
Academic
Healthcare
Government
Consulting
Insurance
Energy
Finance
Retail
Media
Manufacturing
Tech
Electronics
0%
2%
4%
6%
8%
% of industry's traffic going to R

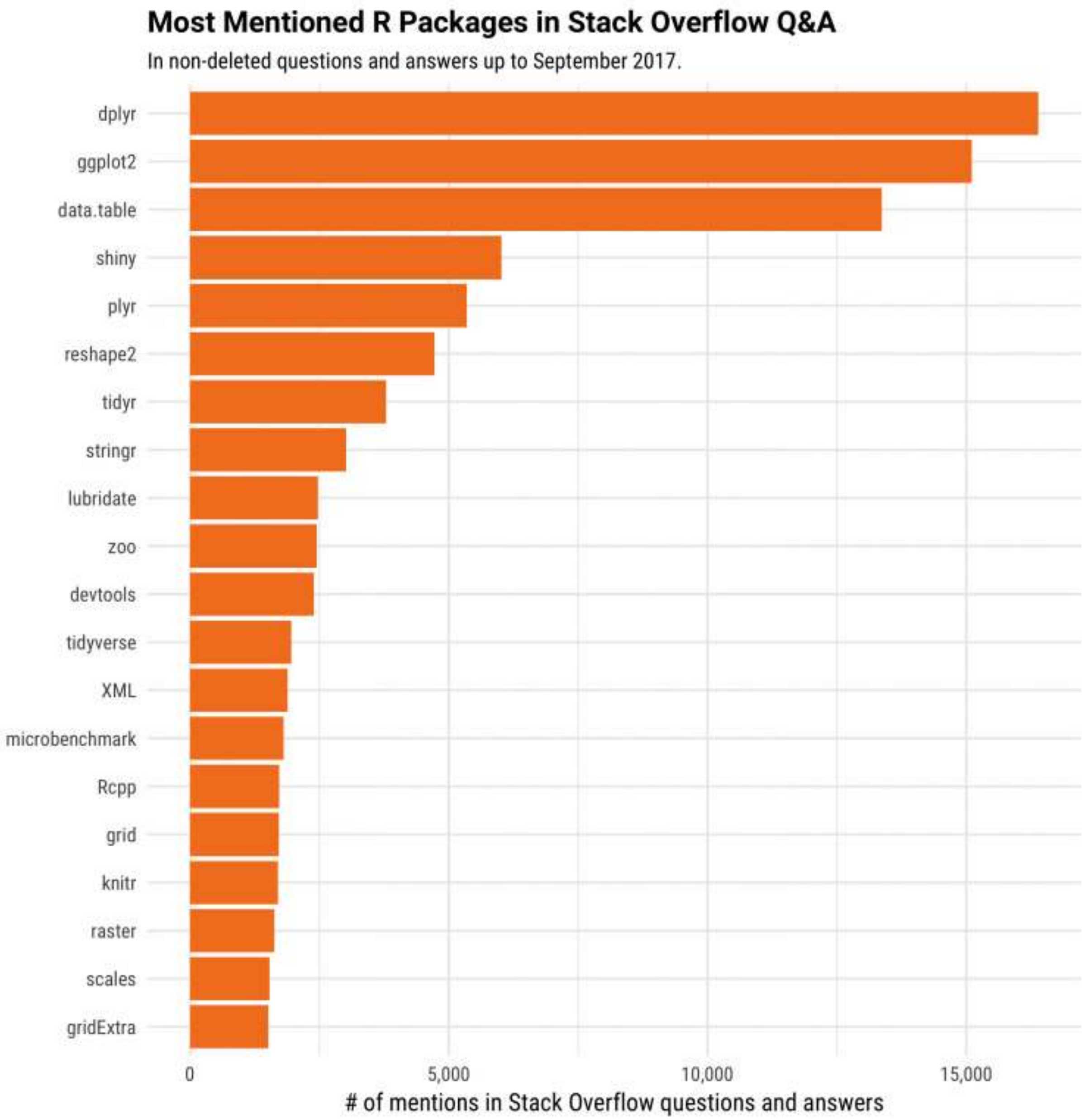
Information mining is the procedure to find fascinating learning from a lot of information [Han and Kanber, 2000]. It is an interdisciplinary field with commitments from numerous zones, for example, insights, AI, data recovery, design acknowledgment and bioinformatics. Information mining is broadly utilized in numerous areas, for example, retail, fund, media transmission and online life.

The primary procedures for information mining incorporate arrangement and forecast, bunching, anomaly

recognition, affiliation rules, arrangement examination, time arrangement investigation and content mining, and furthermore some new methods, for example, informal community investigation and supposition investigation. Point by point presentation of information mining systems can be found in course books on information mining [Han and Kamber, 2000,Handet al., 2001, Witten and Frank, 2005]. In certifiable applications, an information mining procedure can be broken into six noteworthy stages: business understanding, information understanding, information arrangement, demonstrating, assessment and sending, as characterized by the CRISP-DM (Cross Industry Standard Process for Data Mining)

R[R Core Team, 2015b] is a free programming condition for factual figuring and designs.

It gives a wide assortment of factual and graphical systems. R can be effectively reached out with 7324 bundles accessible on CRAN3 (as of October 20, 2015). What's more, there are numerous bundles given on different sites, for example, Bioconductor4, and furthermore a ton of bundles being worked on at R-Forge5 and GitHub6. More insights concerning R are accessible in An Introduction to R 7[Venableset al., 2015] and R Language Definition 8 [R Core Team, 2015d] at the CRAN site. R is generally utilized in both scholarly world and industry



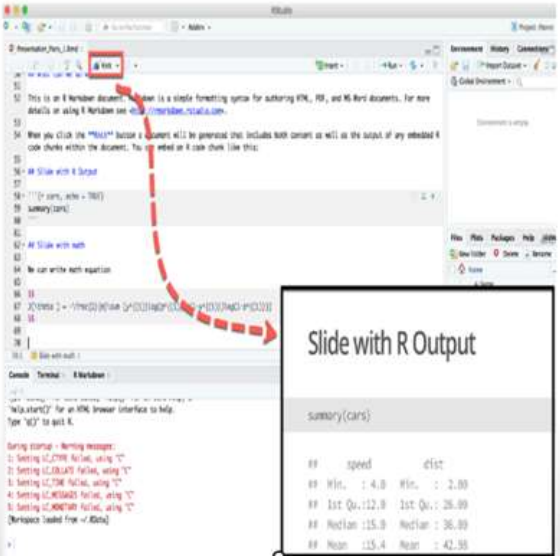
Most Mentioned R Packages in Stack Overflow Q&A

In non-deleted questions and answers up to September 2017.

# What can we do with R

It is really easy to use markdown to create report, papers, book and presentation

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.



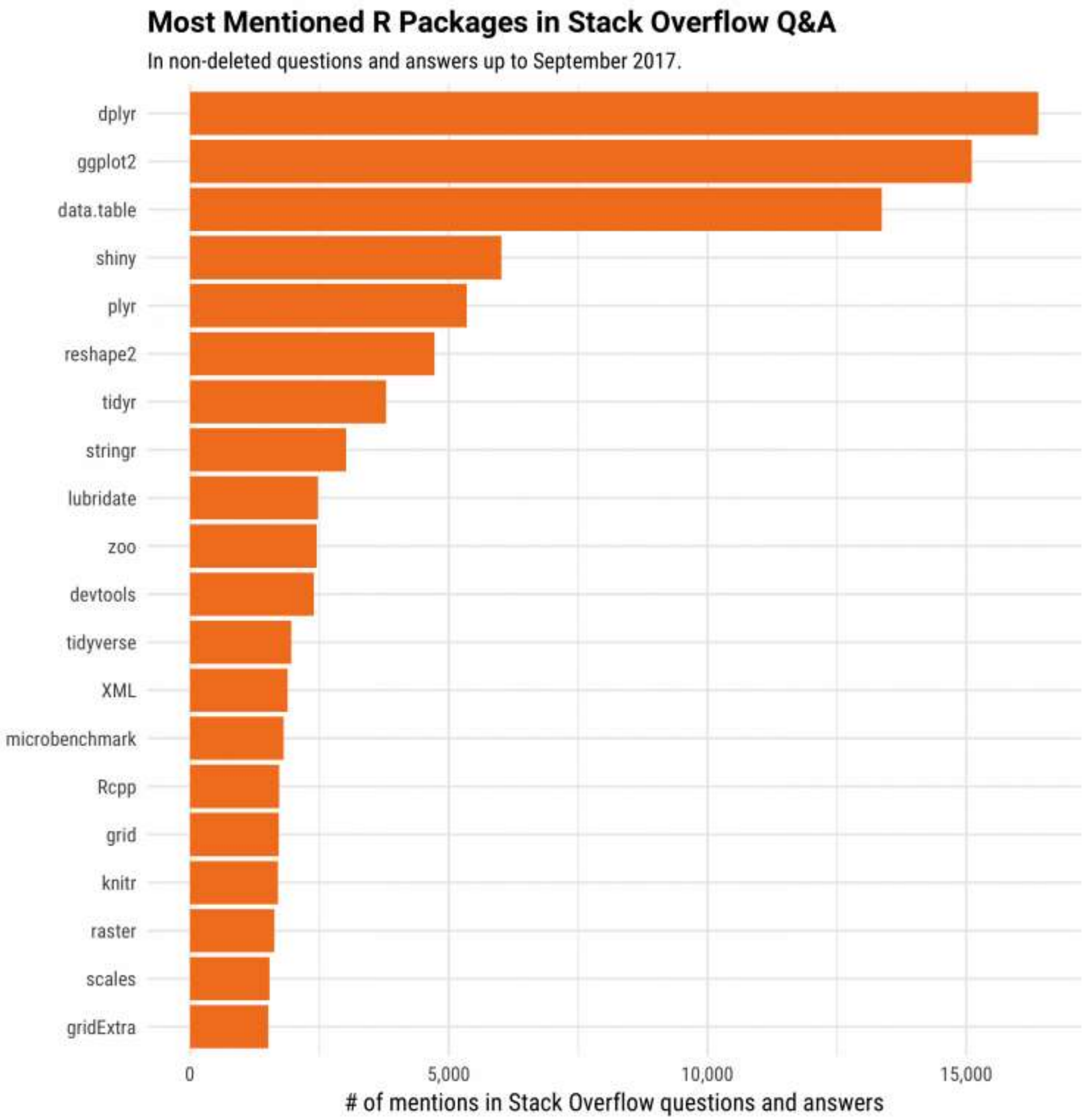
~~~
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more
details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R
code chunks within the document. You can embed an R code chunk like this:

## Slide with R Output

```{r cars, echo = TRUE}
summary(cars)
```

## Slide with math

We can write math equation

$$
[illegible]
$$
~~~

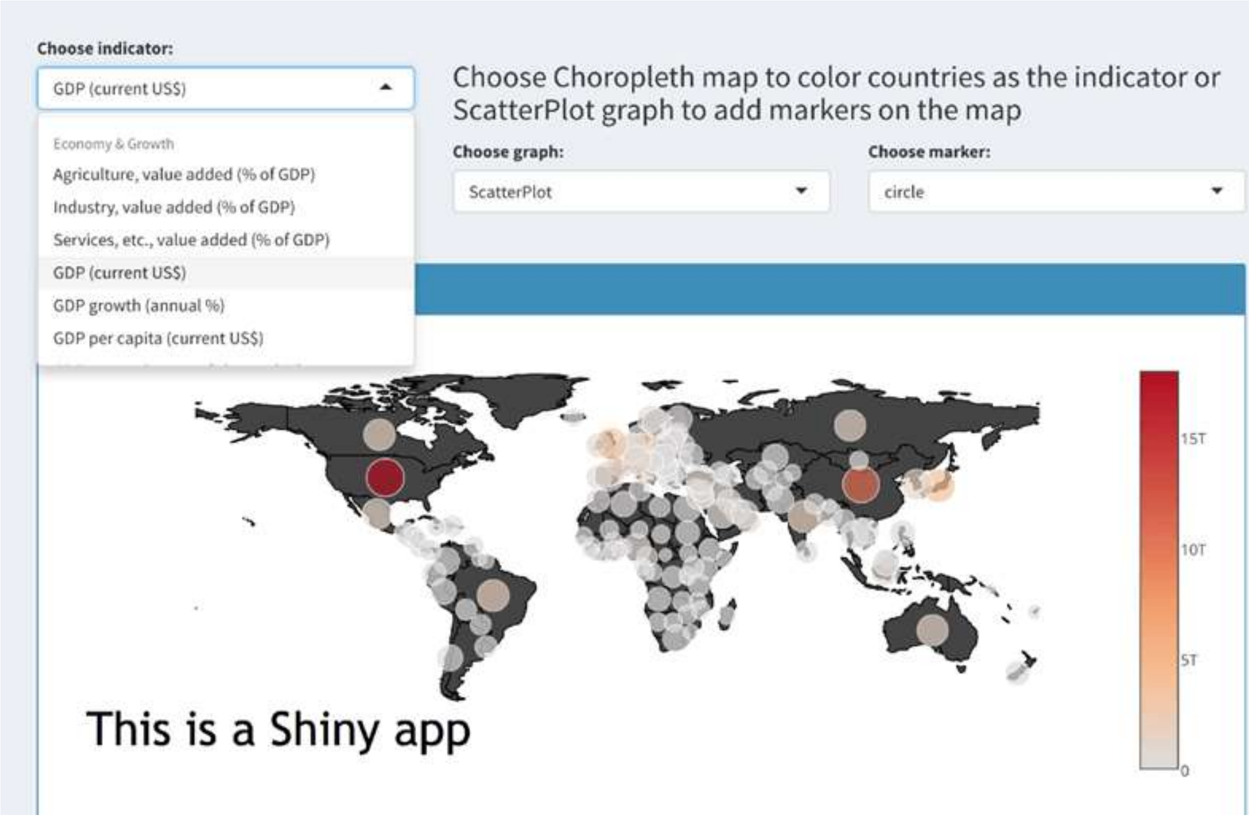
```
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
[Workspace loaded from ~/.RData]
```

Slide with R Output

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
```


Most Mentioned R Packages in Stack Overflow Q&A
In non-deleted questions and answers up to September 2017.
dplyr
ggplot2
data.table
shiny
plyr
reshape2
tidyr
stringr
lubridate
zoo
devtools
tidyverse
XML
microbenchmark
Rcpp
grid
knitr
raster
scales
gridExtra
0
5,000
10,000
15,000
# of mentions in Stack Overflow questions and answers



Choose indicator: GDP (current US$)

Economy & Growth

Agriculture, value added (% of GDP)

Industry, value added (% of GDP)

Services, etc., value added (% of GDP)

GDP (current US$)

GDP growth (annual %)

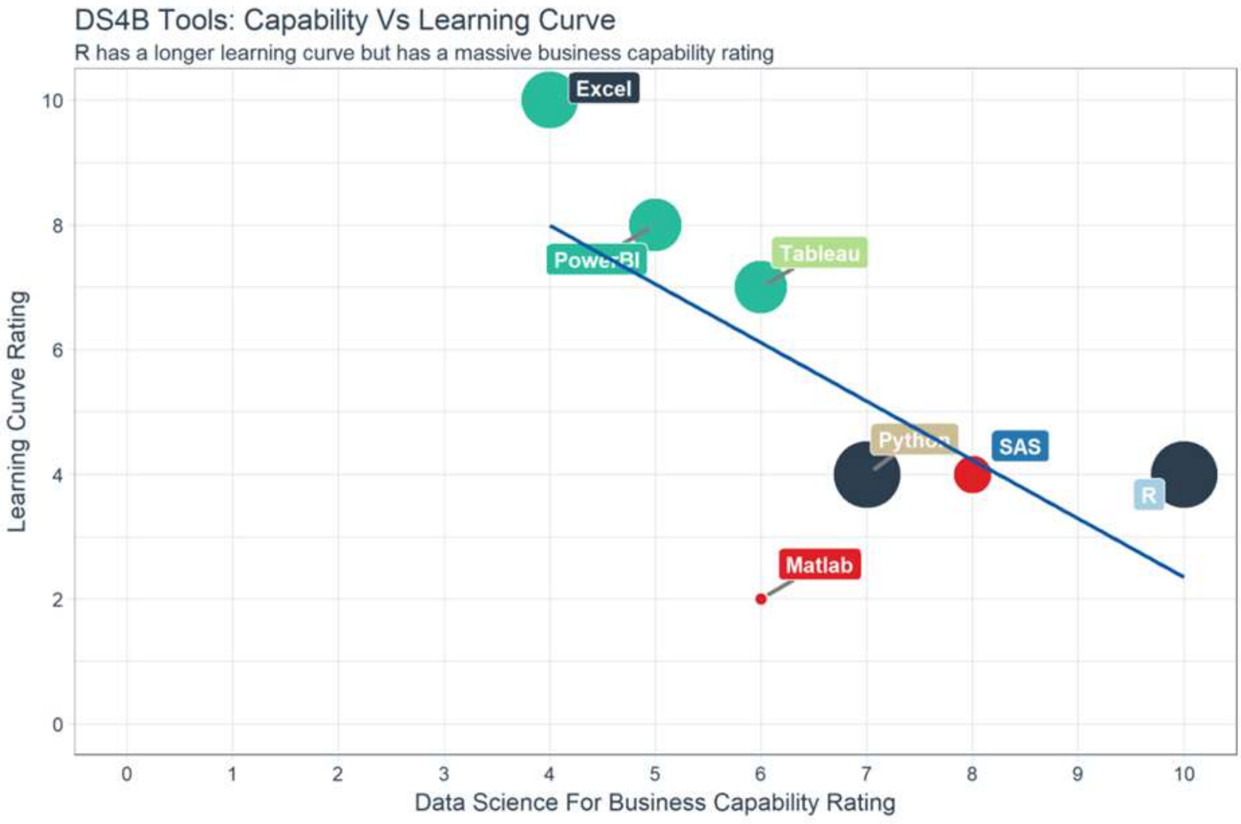
GDP per capita (current US$)

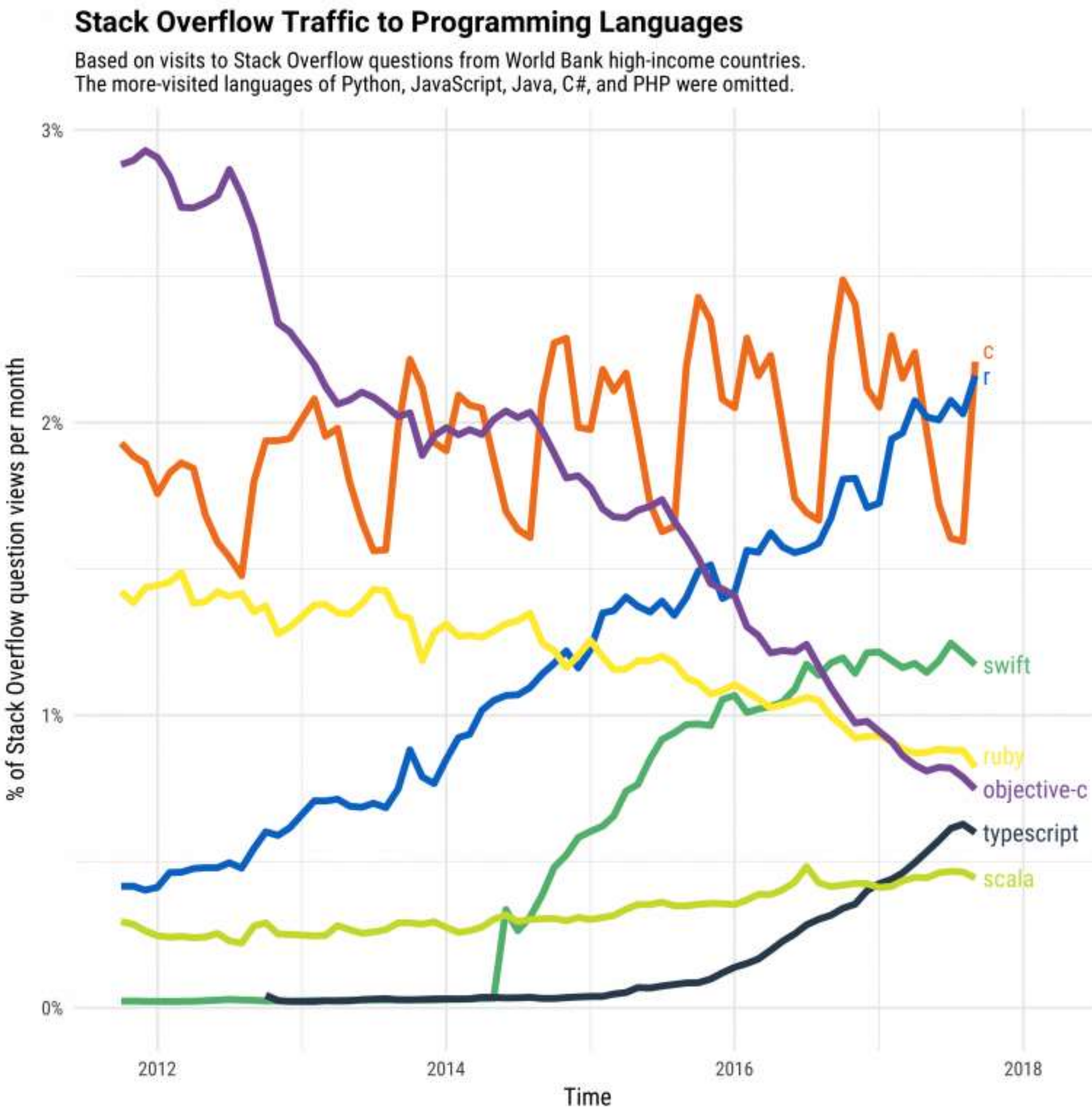
Choose Choropleth map to color countries as the indicator or ScatterPlot graph to add markers on the map

Choose graph: ScatterPlot

Choose marker: circle

This is a Shiny app





Stack Overflow Traffic to Programming Languages

Based on visits to Stack Overflow questions from World Bank high-income countries.
The more-visited languages of Python, JavaScript, Java, C#, and PHP were omitted.

Information researcher can utilize two great instruments: R and Python. You might not have sufficient energy to learn them both, particularly in the event that you begin to learn information science. Learning factual displaying and calculation is undeniably more vital than to gain proficiency with a programming language. A programming language is a device to figure and convey your disclosure. The most vital assignment in information science is the manner in which you manage the information: import, clean, prep, include building, highlight determination. This ought to be your essential core interest. In the event that you are endeavoring to learn R and Python in the meantime without a strong foundation in insights, its plain moronic. Information researcher are not software engineers. Their main responsibility is to comprehend the information, control it and uncover the best methodology. On the off chance that you are pondering which language to learn, how about we see which language is the most proper for you.

The important crowd for information science is business proficient. In the business, one major ramifications is correspondence. There are numerous approaches to convey: report, web application, dashboard. You need an instrument that does this together.

A long time back, R was a troublesome language to ace. The language was befuddling and not as organized as the other programming devices. To conquer this serious issue, Hadley Wickham built up a gathering of bundles called tidyverse. The standard of the diversion changed generally advantageous. Information control become minor and natural. Making a diagram was not all that troublesome any longer.

The best calculations for AI can be actualized with R. Bundles like Keras and TensorFlow permit to make top of the line AI method. R additionally has a bundle to perform Xgboost, one the best calculation for Kaggle rivalry.

R can speak with the other language. It is conceivable to call Python, Java, C++ in R. The universe of enormous information is additionally open to R. You can associate R with various databases like Spark or Hadoop.

At last, R has advanced and permitted parallelizing task to accelerate the calculation. Truth be told, R was condemned for utilizing just a single CPU at any given moment. The parallel bundle lets you to perform undertakings in various centers of the machine.

## History of Reddit

Reddit is a social news accumulation, web content rating, and dialog site. Enlisted individuals submit substance to the site, for example, joins, content posts, and pictures, which are then casted a ballot up or somewhere near different individuals. Posts are composed by subject into client made sheets called "subreddits", which spread an assortment of points including news, science, motion pictures, computer games, music, books, wellness, sustenance, and picture sharing. Entries with additional up-cast a ballot show up towards the highest point of their subreddit and, on the off chance that they get enough votes, at last on the site's first page. Regardless of strict standards denying badgering, Reddit's overseers spend significant assets on directing the site.

Reddit is a site involving client produced content—including photographs, recordings, connections, and content based posts—and talks of this substance in what is basically an announcement board framework. The name "Reddit" is a figure of speech with the expression "read it", i.e., "I read it on Reddit." As of 2018, there are around 330 million Reddit clients, called "redditors". The site's substance is partitioned into classifications or networks referred to on location as "subreddits", of which there are in excess of 138,000 dynamic networks.

As a system of networks, Reddit's center substance comprises of posts from its clients. Clients can remark on others' presents on proceed with the discussion. A key element to Reddit is that clients can cast positive or negative votes, called upvotes and downvotes, for each post and remark on the site. The quantity of upvotes or downvotes decides the posts' perceivability on the site, so the most prevalent substance is shown to the a great many people. Clients can likewise procure "karma" for their posts and remarks, which mirrors the client's remaining inside the network and their commitments to Reddit.

The most famous posts from the site's various subreddits are obvious on the first page to the individuals who peruse the site without a record. As a matter of course for those clients, the first page will show the subreddit r/well known, highlighting top-positioned posts over all of Reddit, barring not-ok for the office networks and others that are most regularly sifted through by clients (regardless of whether they are ok for the office). The subreddit r/all does not channel subjects. Enlisted clients who buy in to subreddits see the top substance from the subreddits to which they buy in on their own front pages.

First page rank—for both the general first page and for individual subreddits—is dictated by a mix of components, including the age of the accommodation, positive ("upvoted") to negative ("downvoted") criticism proportion, and the all out vote-check.

There are around 330 million Reddit clients, called "redditors". Enlisting a record with Reddit is free and does not require an email address. Notwithstanding remarking and casting a ballot, enrolled clients can likewise make their own subreddit on a subject based on their personal preference. In Reddit style, usernames start with "u/". For instance, essential redditors incorporate u/Poem_for_your_sprog, who reacts to messages crosswise over Reddit in section, and u/Shitty_Watercolour, who presents works of art accordingly on posts.

Subreddits are regulated by mediators, Reddit clients who win the title by making a subreddit or being advanced by a present arbitrator. These arbitrators are volunteers who deal with their networks, set and authorize network explicit guidelines, evacuate posts and remarks that abuse these tenets, and for the most part work to keep discourses in their subreddit on theme. Administrators, on the other hand, are paid to work for Reddit.

Discourses on Reddit are composed into client made regions of intrigue called "subreddits". There are around 138,000 dynamic subreddits among an aggregate of 1.2 million, starting at July 2018. Subreddit names start with "r/". For example, r/science is a network dedicated to talking about logical themes and r/TV is a network committed to examining TV appears. In the interim, r/well known highlights top-positioned posts over all of Reddit, barring not-ok for the office networks and others that are most ordinarily sifted through by clients (regardless of whether they are ok for the office). The subreddit r/all does not channel subjects.

In a 2014 meeting with Memeburn, Erik Martin, at that point general administrator of Reddit, commented that their "approach is to give the network arbitrators or keepers however much control as could be expected so they can shape and develop the sort of networks they need". Subreddits frequently use themed variations of Reddit's outsider mascot, Snoo, in the visual styling of their networks.

Reddit Premium (earlier Reddit Gold) is an exceptional enrollment that enables clients to see the site advertisement free. Clients may likewise be skilled coins if another client especially esteemed the remark or post, by and large because of comical or great substance. Reddit Premium opens a few highlights not available to normal clients, for example, remark featuring, select subreddits, and a customized Snoo (known as a "snoovatar"). Reddit Gold was renamed Reddit Premium in 2018. Notwithstanding gold coins, clients can blessing silver and platinum coins to different clients as remunerations for quality substance.

On the site, redditors recognize their "cake day" when a year, on the commemoration of the day their record was made. Cake day includes a symbol of a little cut of cake by the client's name for 24 hours.

In 2017, Reddit built up its own constant visit programming for the site. While some settled subreddits host utilized third-get-together programming to visit about their networks, the organization constructed talk works that it expectations will turn into a vital piece of Reddit. Singular talk rooms were taken off in 2017 and network visit spaces for individuals from a given subreddit were taken off in 2018.

The thought and beginning improvement of Reddit began with then school flat mates Steve Huffman and Alexis Ohanian in 2005. Huffman and Ohanian went to an address by developer business visionary Paul Graham in Boston, Massachusetts, amid their spring break from University of Virginia. In the wake of talking with Huffman and Ohanian following the address, Graham welcomed the two to apply to his startup hatchery Y Combinator. Their underlying thought, My Mobile Menu, was fruitless, and was proposed to enable clients to arrange sustenance by SMS content informing. Amid a meeting to generate new ideas to pitch another startup, the thought was made for what Graham called the "first page of the Internet". For this thought, Huffman and Ohanian were acknowledged in Y Combinator's top of the line. Bolstered by the financing from Y Combinator, Huffman coded the site in Lisp and together with Ohanian propelled Reddit in June 2005.

The group extended to incorporate Christopher Slowe in November 2005. Between November 2005 and January 2006, Reddit converged with Aaron Swartz's organization Infogami, and Swartz turned into an equivalent proprietor of the subsequent parent organization, Not A Bug. Huffman and Ohanian sold Reddit to Condé Nast Publications, proprietor of Wired, on October 31, 2006, for an announced $10 million to $20 million and the group moved to San Francisco. In January 2007, Swartz was terminated for undisclosed reasons.

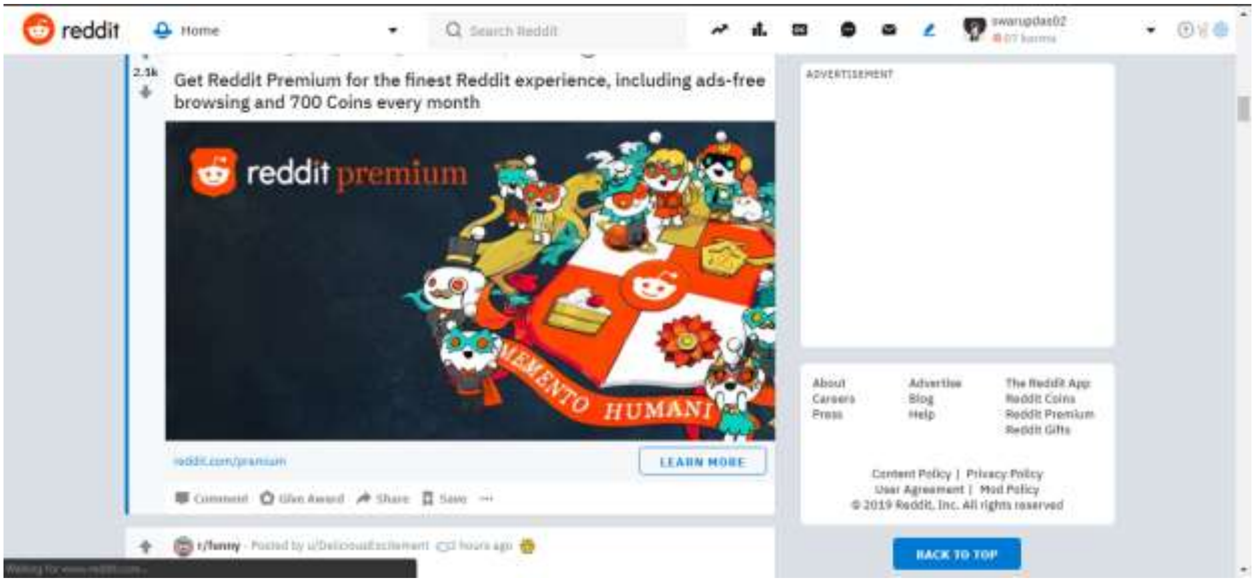
Huffman and Ohanian left Reddit in 2009. Huffman went on to help establish Hipmunk with Adam Goldstein, and later selected Ohanian and Slowe to his new organization. After Huffman and Ohanian left Reddit, Erik Martin, who joined the organization as a network director in 2008 and later wound up general supervisor is 2011, assumed a job in Reddit's development. VentureBeat noticed that Martin was "in charge of propping the site up" under Condé Nast's proprietorship. Martin encouraged the buy of Reddit Gifts and drove philanthropy activities.

Reddit propelled two diverse methods for promoting on the site in 2009. The organization propelled supported substance and a self-serve advertisements stage that year. Reddit propelled its Reddit Gold advantages program in July 2010, which offered new highlights to editors and made another income stream for the business that did not depend on flag advertisements. On September 6, 2011, Reddit turned out to be operationally autonomous of Condé Nast, working as a different auxiliary of its parent organization, Advance Publications. Reddit and different sites took an interest in a 12-hour sitewide power outage on January 18, 2012, in challenge of the Stop Online Piracy Act. In May 2012, Reddit joined the Internet Defense League, a gathering framed to compose future dissents.

Yishan Wong joined Reddit as CEO in 2012. Wong left Reddit in 2014, after over two years at the organization, refering to differences about his proposition to move the organization's workplaces from San Francisco to adjacent Daly City, yet in addition the "unpleasant and depleting" nature of the position. Ohanian acknowledged Wong for driving the organization as its client base developed from 35 million to 174 million. Wong directed the organization as it brought $50 million up in subsidizing and spun off as an autonomous organization. Likewise amid this time, Reddit started tolerating the advanced cash Bitcoin for its Reddit Gold membership administration through an organization with bitcoin installment processor Coinbase in February 2013. Ellen Pao supplanted Wong as between time CEO in 2014 and surrendered in 2015 in the midst of a client revolt over the terminating of a well known Reddit worker. Amid her residency, Reddit started an enemy of provocation approach, prohibited automatic sexualization, and restricted a few discussions that concentrated on intolerant substance or badgering of people.

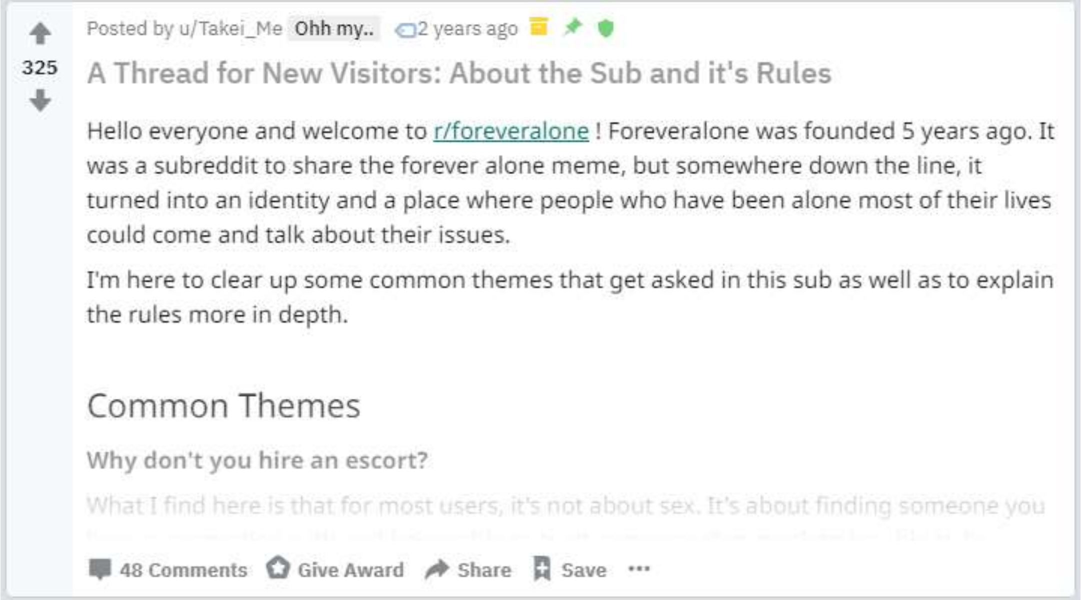
Following five years from the organization, Ohanian and Huffman came back to influential positions at Reddit: Ohanian turned into the full-time official director in November 2014 after Wong's abdication, while Pao's flight on July 10, 2015 prompted Huffman's arrival as the organization's CEO. After Huffman rejoined Reddit as CEO, he propelled Reddit's iOS and Android applications, fixed Reddit's versatile site, and made A/B testing foundation. The organization propelled a noteworthy overhaul of its site in April 2018. Huffman said new clients were killed from Reddit in light of the fact that it had resembled a "tragic Craigslist". Reddit additionally founded a few innovative upgrades, for example, another apparatus that enables clients to shroud posts, remarks, and private messages from chose redditors trying to check online provocation, and new substance rules. These new substance rules were gone for restricting substance affecting viciousness and isolating hostile material. Slowe, the organization's first representative, rejoined Reddit in 2017 as boss innovation officer. Reddit's biggest round of financing came in 2017, when the organization raised $200 million and was esteemed at $1.8 billion. The subsidizing upheld Reddit's site update and video endeavors.

Sample of information on Reddit:



Get Reddit Premium for the finest Reddit experience, including ads-free browsing and 700 Coins every month

© 2019 Reddit, Inc. All rights reserved



Posted by u/Takei_Me Ohh my.. 2 years ago

A Thread for New Visitors: About the Sub and it's Rules

Hello everyone and welcome to r/foreveralone ! Foreveralone was founded 5 years ago. It was a subreddit to share the forever alone meme, but somewhere down the line, it turned into an identity and a place where people who have been alone most of their lives could come and talk about their issues.

I'm here to clear up some common themes that get asked in this sub as well as to explain the rules more in depth.

Common Themes

Why don't you hire an escort?

What I find here is that for most users, it's not about sex. It's about finding someone you

## Implementation

NAÏVE BAYESIAN CODE IN R:

```
#Code Created by Swarup Das 17MCA0008

#Last Edit Date: 21/01/2019

#importing sqldf libraries

#install.packages("sqldf")

library(sqldf)



#importing data to be read

initial<-read.csv("suicide_train.csv")



#sql queries to store the trained values

gender_male_suicide_yes<-sqldf('select count(attempt_suicide) from "initial" where gender="Male" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where gender="Male"')

gender_Male_suicide_no= sqldf('select count(attempt_suicide) from "initial" where gender="Male" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where gender="Male"')

gender_Female_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where gender="Female" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where gender="Female"')

gender_Female_suicide_no= sqldf('select count(attempt_suicide) from "initial" where gender="Female" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where gender="Female"')



sexuality_Straight_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where sexuality="Straight" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where sexuality="Straight"')

sexuality_Straight_suicide_no= sqldf('select count(attempt_suicide) from "initial" where sexuality="Straight" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where sexuality="Straight"')
```

```
sexuality_BisexuaI_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="BisexuaI" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sexuality="BisexuaI"')

sexuality_BisexuaI_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="BisexuaI" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sexuality="BisexuaI"')

sexuality_Gay/Lesbian_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="Gay/Lesbian" and attempt_suicide="Yes"')

sexuality_Gay/Lesbian_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="Gay/Lesbian" and attempt_suicide="No"')

sexuality_Straight_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="Straight" and attempt_suicide="Yes"')

sexuality_Straight_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sexuality="Straight" and attempt_suicide="No"')


race_Asian_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where race="Asian"
and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Asian"')

race_Asian_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where race="Asian"
and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Asian"')

race_BIack_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where race="BIack"
and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="BIack"')

race_BIack_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where race="BIack"
and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="BIack"')

race_caucasian_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="caucasian" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="caucasian"')

race_caucasian_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="caucasian" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="caucasian"')
```

```
race_half_Arab_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="half Arab" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="Arab"')

race_half_Arab_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="half Arab" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where race="Arab"')

race_helicopterkin_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="helicopterkin" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="helicopterkin"')

race_helicopterkin_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="helicopterkin" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where race="helicopterkin"')

race_Hispanic_of_any_race_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Hispanic (of any race)" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="Hispanic (of any race)"')

race_Hispanic_of_any_race_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="Hispanic (of any race)" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where race="Hispanic (of any race)"')

race_Indian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Indian" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="Indian"')

race_Indian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="Indian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where race="Indian"')

race_Middle_Eastern_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Middle Eastern" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="Middle Eastern"')

race_Middle_Eastern_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="Middle Eastern" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where race="Middle Eastern"')

race_Mixed_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Mixed" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where race="Mixed"')
```

```
race_Mixed_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Mixed" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI"
where race="Mixed"')

race_Mixed_race_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Mixed race" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="Mixed race"')

race_Mixed_race_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Mixed race" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="Mixed race"')

race_Mixed_white_asian_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI"
where race="Mixed white/asian" and attempt_suicide="Yes"')/sqIdf('seIect
count(attempt_suicide) from "initiaI" where race="Mixed white/asian"')

race_Mixed_white_asian_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Mixed white/asian" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide)
from "initiaI" where race="Mixed white/asian"')

race_Native_American_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Native American" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="Native American"')

race_Native_American_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Native American" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="Native American"')

race_Pakistani_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Pakistani" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where race="Pakistani"')

race_Pakistani_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Pakistani" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI"
where race="Pakistani"')

race_Turkish_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Turkish" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from "initiaI"
where race="Turkish"')

race_Turkish_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
race="Turkish" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI"
where race="Turkish"')
```

```
race_white_and_asian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="white and asian" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="white and asian"')

race_white_and_asian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="white and asian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="white and asian"')

race_White_non_Hispanic_suicide_yes= sqldf('select count(attempt_suicide) from "initial"
where race="White non-Hispanic" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where race="White non-Hispanic"')

race_White_non_Hispanic_suicide_no= sqldf('select count(attempt_suicide) from "initial"
where race="White non-Hispanic" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where race="White non-Hispanic"')



bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Normal weight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Normal weight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Normal weight" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Normal weight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Obese" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide)
from "initial" where bodyweight="Obese"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Obese" and attempt_suicide="No"')/sqldf('select count(attempt_suicide)
from "initial" where bodyweight="Obese"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Overweight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Overweight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Overweight" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Overweight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Underweight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Underweight"')
```

```
bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Underweight" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Underweight"')



social_fear_yes_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
social_fear="Yes" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where social_fear="Yes"')

social_fear_no_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
social_fear="No" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where social_fear="No"')

social_fear_no_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
social_fear="No" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where social_fear="No"')

social_fear_yes_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
social_fear="Yes" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where social_fear="Yes"')



depressed_yes_suicide_yes=sqldf('select count(attempt_suicide) from "initial" where
depressed="Yes" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where depressed="Yes"')

depressed_yes_suicide_no=sqldf('select count(attempt_suicide) from "initial" where
depressed="Yes" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where depressed="Yes"')

depressed_no_suicide_yes=sqldf('select count(attempt_suicide) from "initial" where
depressed="No" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where depressed="No"')

depressed_no_suicide_no=sqldf('select count(attempt_suicide) from "initial" where
depressed="No" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where depressed="No"')



suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial"')

suicide_no= sqldf('select count(attempt_suicide) from "initial" where
attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"')
```

NAÏVE BAYESIAN GRAPH GENERATION:

```
#storing suicide test and train data

train <- read.csv("suicide_train.csv")

test <- read.csv("suicide_test.csv")



#Rows and Cols

dim(train)

dim(test)



#Columns name

colnames(train)

colnames(test)



#Show

head(train)

head(test)



#importing necessary libraries

library(caret)

library(e1071)

library(AUC)



#training the suicide stats previously shown

model.Bayes <- naiveBayes(attempt_suicide~., data = train)

model.Bayes



#testing the current statistical data
```

```
pc <-NULL

pc <- predict(model.Bayes, test, type = "class")

summary(pc)

xtab <- table(pc, test$attempt_suicide)

caret::confusionMatrix(xtab, positive = "Yes")


#lift chart

pb <-NULL

pb <- predict(model.Bayes, test, type = "raw")

pb <- as.data.frame(pb)

pred.Bayes <- data.frame(test$attempt_suicide,pb$Yes)

colnames(pred.Bayes) <- c("target","score")

lift.Bayes <- lift(target ~ score, data = pred.Bayes, cuts=10, class="Yes")

xyplot(lift.Bayes, main="Bayesian Classifier - Lift Chart", type=c("l","g"), lwd=2

       , scales=list(x=list(alternating=FALSE,tick.number = 10)

                  ,y=list(alternating=FALSE,tick.number = 10)))



#roc chart

labels <- as.factor(ifelse(pred.Bayes$target=="Yes", 1, 0))

predictions <- pred.Bayes$score

auc(roc(predictions, labels), min = 0, max = 1)

plot(roc(predictions, labels), min=0, max=1, type="l", main="Bayesian Classifier - ROC Chart")
```

WEIGHTED NAÏVE BAYESIAN CODE IN R:

```
#Code Created by Swarup Das 17MCA0008
#Last Edit Date: 21/01/2019
#importing sqldf libraries
#install.packages("sqldf")
library(sqldf)


#importing data to be read
initial<-read.csv("suicide_train.csv")


#sql queries to store the trained values
gender_male_suicide_yes<-sqldf('select count(attempt_suicide) from "initial" where
gender="Male" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial"
where gender="Male"')
gender_Male_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
gender="Male" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where gender="Male"')
gender_Female_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
gender="Female" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where gender="Female"')
gender_Female_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
gender="Female" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where gender="Female"')


sexuality_Straight_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Straight" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where sexuality="Straight"')
sexuality_Straight_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Straight" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where sexuality="Straight"')
```

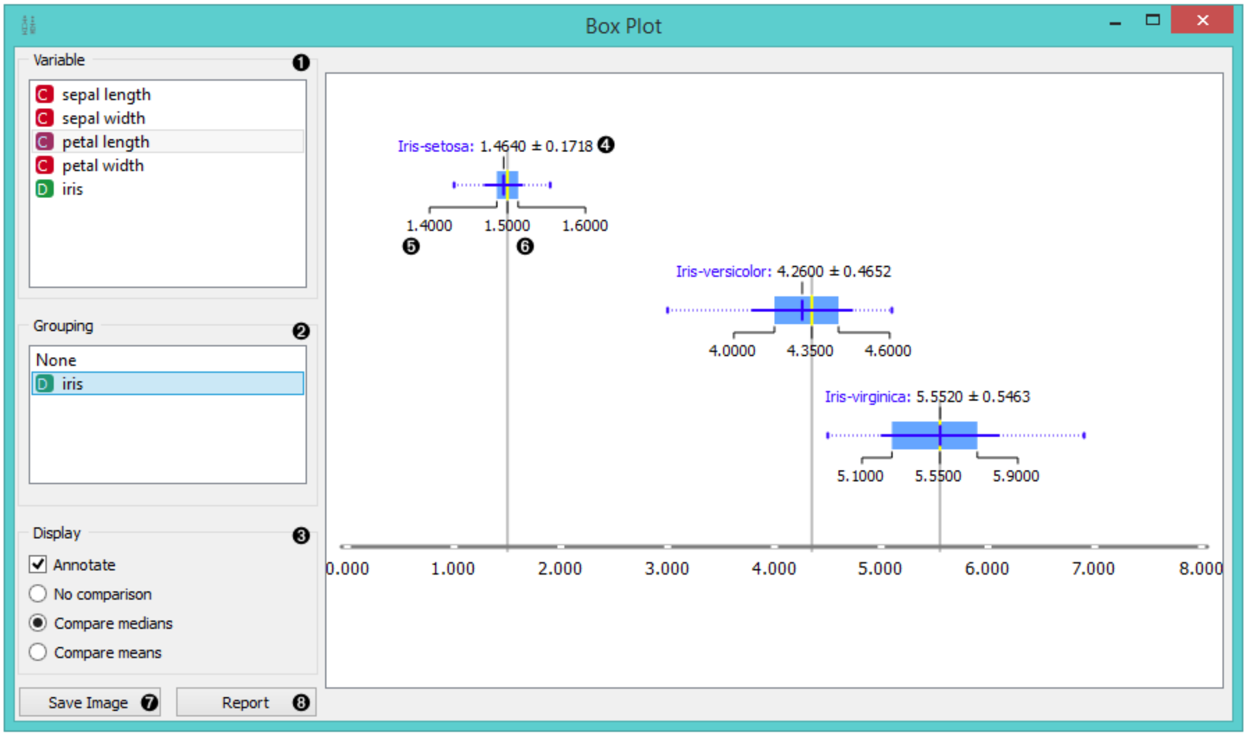
```
sexuality_Bisexual_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Bisexual" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where sexuality="Bisexual"')

sexuality_Bisexual_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Bisexual" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where sexuality="Bisexual"')

sexuality_Gay/Lesbian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Gay/Lesbian" and attempt_suicide="Yes"')

sexuality_Gay/Lesbian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Gay/Lesbian" and attempt_suicide="No"')

sexuality_Straight_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Straight" and attempt_suicide="Yes"')

sexuality_Straight_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
sexuality="Straight" and attempt_suicide="No"')


race_Asian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Asian"
and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where
race="Asian"')

race_Asian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="Asian"
and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where
race="Asian"')

race_Black_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where race="Black"
and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial" where
race="Black"')

race_Black_suicide_no= sqldf('select count(attempt_suicide) from "initial" where race="Black"
and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial" where
race="Black"')

race_caucasian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="caucasian" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="caucasian"')

race_caucasian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="caucasian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="caucasian"')
```

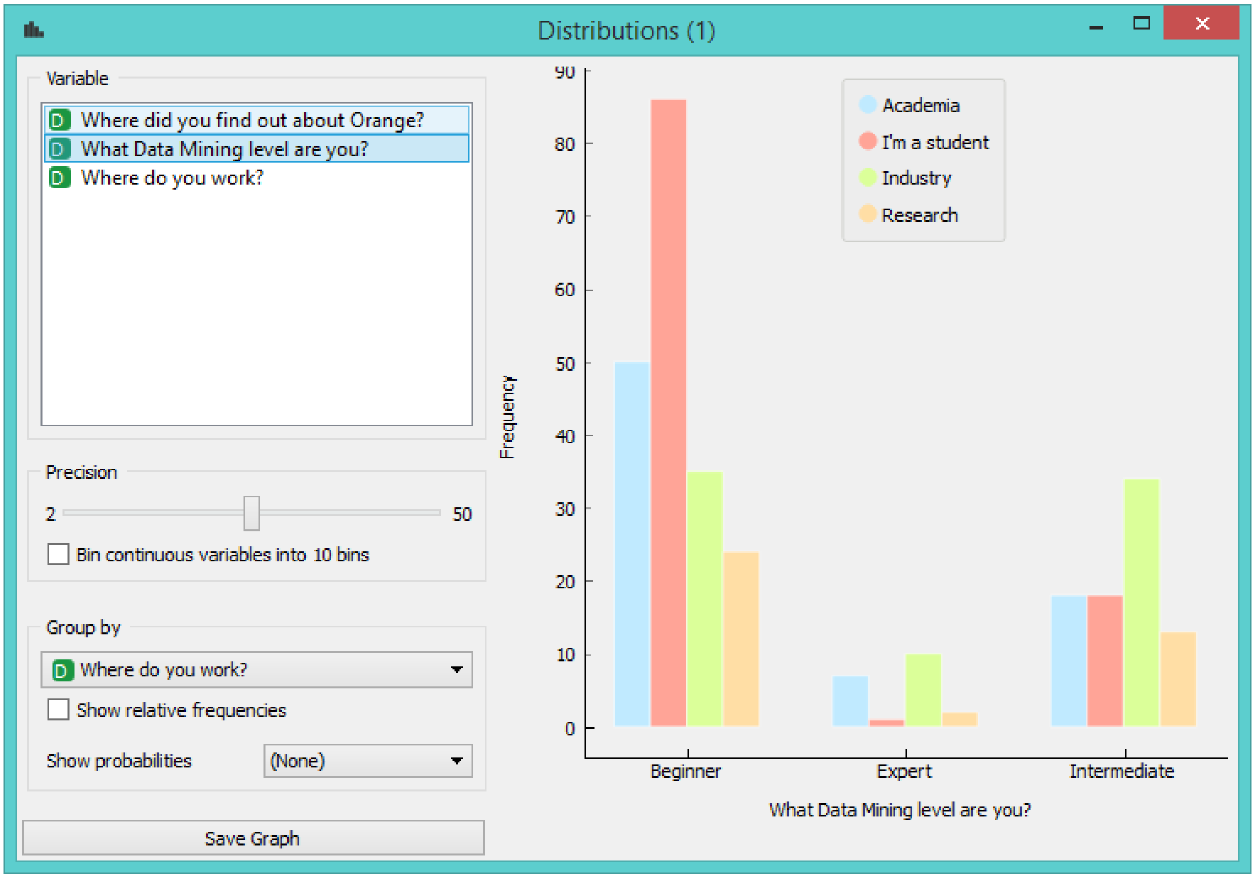
```
race_half_Arab_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="half Arab" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="Arab"')

race_half_Arab_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="half Arab" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where race="Arab"')

race_helicopterkin_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="helicopterkin" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="helicopterkin"')

race_helicopterkin_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="helicopterkin" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="helicopterkin"')

race_Hispanic_of_any_race_suicide_yes= sqldf('select count(attempt_suicide) from "initial"
where race="Hispanic (of any race)" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where race="Hispanic (of any race)"')

race_Hispanic_of_any_race_suicide_no= sqldf('select count(attempt_suicide) from "initial"
where race="Hispanic (of any race)" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where race="Hispanic (of any race)"')

race_Indian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Indian" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial"
where race="Indian"')

race_Indian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Indian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where race="Indian"')

race_Middle_Eastern_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Middle Eastern" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="Middle Eastern"')

race_Middle_Eastern_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Middle Eastern" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="Middle Eastern"')

race_Mixed_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Mixed" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial"
where race="Mixed"')
```

```
race_Mixed_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Mixed" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where race="Mixed"')

race_Mixed_race_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Mixed race" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="Mixed race"')

race_Mixed_race_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Mixed race" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="Mixed race"')

race_Mixed_white_asian_suicide_yes= sqldf('select count(attempt_suicide) from "initial"
where race="Mixed white/asian" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where race="Mixed white/asian"')

race_Mixed_white_asian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Mixed white/asian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide)
from "initial" where race="Mixed white/asian"')

race_Native_American_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Native American" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="Native American"')

race_Native_American_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Native American" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="Native American"')

race_Pakistani_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Pakistani" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="Pakistani"')

race_Pakistani_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Pakistani" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where race="Pakistani"')

race_Turkish_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="Turkish" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from "initial"
where race="Turkish"')

race_Turkish_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="Turkish" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from "initial"
where race="Turkish"')
```

```
race_white_and_asian_suicide_yes= sqldf('select count(attempt_suicide) from "initial" where
race="white and asian" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide) from
"initial" where race="white and asian"')

race_white_and_asian_suicide_no= sqldf('select count(attempt_suicide) from "initial" where
race="white and asian" and attempt_suicide="No"')/sqldf('select count(attempt_suicide) from
"initial" where race="white and asian"')

race_White_non_Hispanic_suicide_yes= sqldf('select count(attempt_suicide) from "initial"
where race="White non-Hispanic" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where race="White non-Hispanic"')

race_White_non_Hispanic_suicide_no= sqldf('select count(attempt_suicide) from "initial"
where race="White non-Hispanic" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where race="White non-Hispanic"')



bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Normal weight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Normal weight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Normal weight" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Normal weight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Obese" and attempt_suicide="Yes"')/sqldf('select count(attempt_suicide)
from "initial" where bodyweight="Obese"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Obese" and attempt_suicide="No"')/sqldf('select count(attempt_suicide)
from "initial" where bodyweight="Obese"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Overweight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Overweight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Overweight" and attempt_suicide="No"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Overweight"')

bodyweight_Normal_weight_suicide_yes=sqldf('select count(attempt_suicide) from "initial"
where bodyweight="Underweight" and attempt_suicide="Yes"')/sqldf('select
count(attempt_suicide) from "initial" where bodyweight="Underweight"')
```

```
bodyweight_NormaI_weight_suicide_yes=sqIdf('seIect count(attempt_suicide) from "initiaI"
where bodyweight="Underweight" and attempt_suicide="No"')/sqIdf('seIect
count(attempt_suicide) from "initiaI" where bodyweight="Underweight"')


social_fear_yes_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sociaI_fear="Yes" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sociaI_fear="Yes"')

sociaI_fear_no_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sociaI_fear="No" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sociaI_fear="No"')

sociaI_fear_no_suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sociaI_fear="No" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sociaI_fear="No"')

sociaI_fear_yes_suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
sociaI_fear="Yes" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where sociaI_fear="Yes"')


depressed_yes_suicide_yes=sqIdf('seIect count(attempt_suicide) from "initiaI" where
depressed="Yes" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where depressed="Yes"')

depressed_yes_suicide_no=sqIdf('seIect count(attempt_suicide) from "initiaI" where
depressed="Yes" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where depressed="Yes"')

depressed_no_suicide_yes=sqIdf('seIect count(attempt_suicide) from "initiaI" where
depressed="No" and attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from
"initiaI" where depressed="No"')

depressed_no_suicide_no=sqIdf('seIect count(attempt_suicide) from "initiaI" where
depressed="No" and attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI"
where depressed="No"')


suicide_yes= sqIdf('seIect count(attempt_suicide) from "initiaI" where
attempt_suicide="Yes"')/sqIdf('seIect count(attempt_suicide) from "initiaI"')

suicide_no= sqIdf('seIect count(attempt_suicide) from "initiaI" where
attempt_suicide="No"')/sqIdf('seIect count(attempt_suicide) from "initiaI"')
```


Box Plot
Variable ❶
C sepal length
C sepal width
C petal length
C petal width
D iris
Grouping ❷
None
D iris
Display ❸
☑ Annotate
○ No comparison
● Compare medians
○ Compare means
Save Image ❼
Report ❽
Iris-setosa: 1.4640 ± 0.1718 ❹
1.4000
1.5000
1.6000
❺
❻
Iris-versicolor: 4.2600 ± 0.4652
4.0000
4.3500
4.6000
Iris-virginica: 5.5520 ± 0.5463
5.1000
5.5500
5.9000
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000
8.000


Distributions (1)
Variable
D Where did you find out about Orange?
D What Data Mining level are you?
D Where do you work?
Precision
2
50
☐ Bin continuous variables into 10 bins
Group by
D Where do you work?
☐ Show relative frequencies
Show probabilities
(None)
Save Graph
Frequency
90
80
70
60
50
40
30
20
10
0
Academia
I'm a student
Industry
Research
Beginner
Expert
Intermediate
What Data Mining level are you?

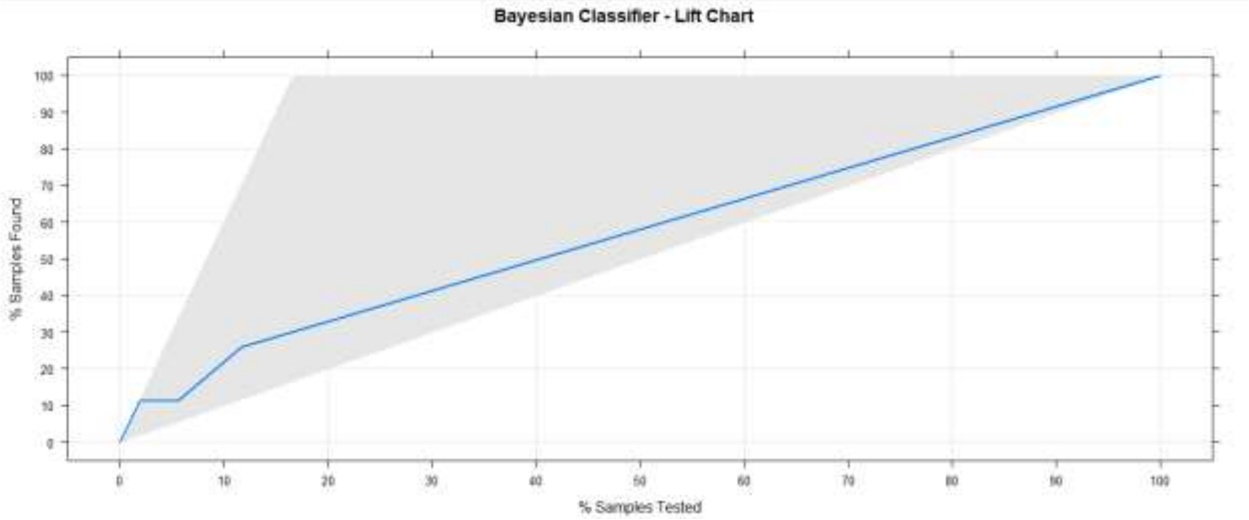
WEIGHTED NAÏVE BAYESIAN GRAPH GENERATION:

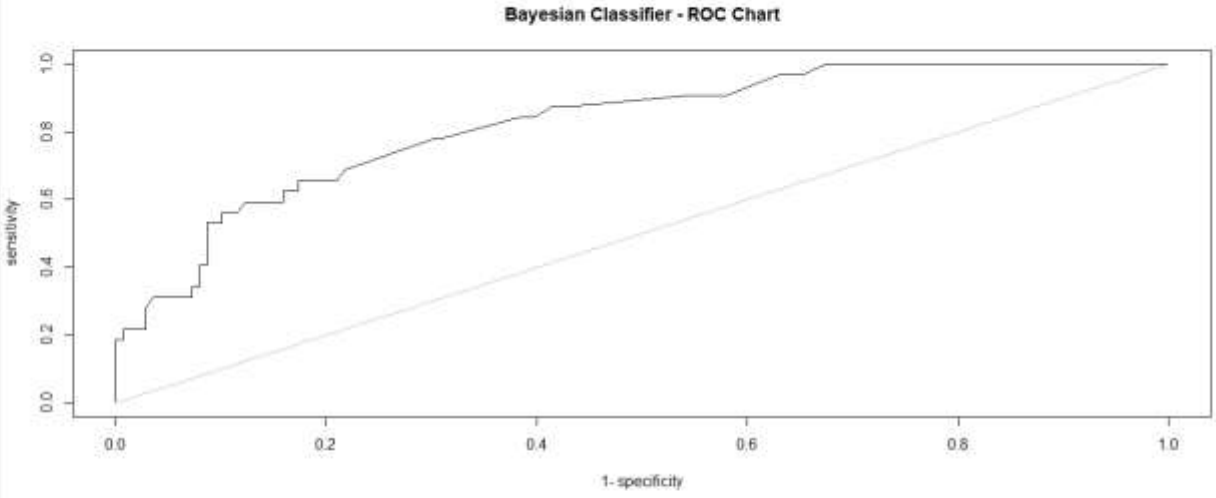
```
#storing suicide test and train data
train <- read.csv("suicide_train.csv")
test <- read.csv("suicide_test.csv")


#Rows and Cols
dim(train)
dim(test)


#Columns name
colnames(train)
colnames(test)


#Show
head(train)
head(test)


#importing necessary libraries
library(caret)
library(e1071)
library(AUC)


#training the suicide stats previously shown
model.Bayes <- naiveBayes(attempt_suicide~., data = train)
model.Bayes
```

```
#testing the current statistical data

pc <-NULL

pc <- predict(model.Bayes, test, type = "class")

summary(pc)

xtab <- table(pc, test$attempt_suicide)

caret::confusionMatrix(xtab, positive = "Yes")


#lift chart

pb <-NULL

pb <- predict(model.Bayes, test, type = "raw")

pb <- as.data.frame(pb)

pred.Bayes <- data.frame(test$attempt_suicide,pb$Yes)

colnames(pred.Bayes) <- c("target","score")

lift.Bayes <- lift(target ~ score, data = pred.Bayes, cuts=10, class="Yes")

xyplot(lift.Bayes, main="Bayesian Classifier - Lift Chart", type=c("l","g"), lwd=2

       , scales=list(x=list(alternating=FALSE,tick.number = 10)

                    ,y=list(alternating=FALSE,tick.number = 10)))



#roc chart

labels <- as.factor(ifelse(pred.Bayes$target=="Yes", 1, 0))

predictions <- pred.Bayes$score

auc(roc(predictions, labels), min = 0, max = 1)

plot(roc(predictions, labels), min=0, max=1, type="l", main="Bayesian Classifier - ROC Chart")
```

# Program Output:

R CODE OUTPUT:




Bayesian Classifier - ROC Chart
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1- specificity
0.0
0.2
0.4
0.6
0.8
1.0

## Orange Output:

DATA ANALYTICS USING ORANGE:

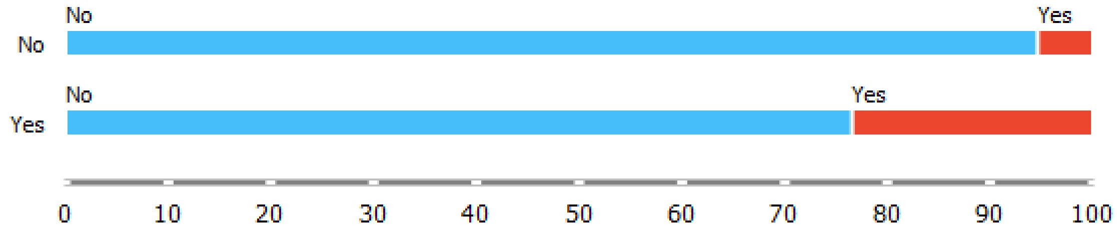


*Figure 1: Depression vs Suicide frequency*

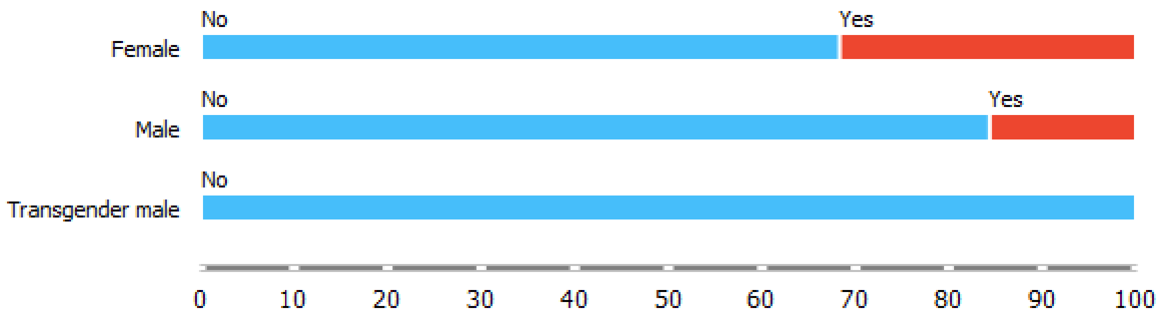


*Figure 2 Depression vs Gender type*

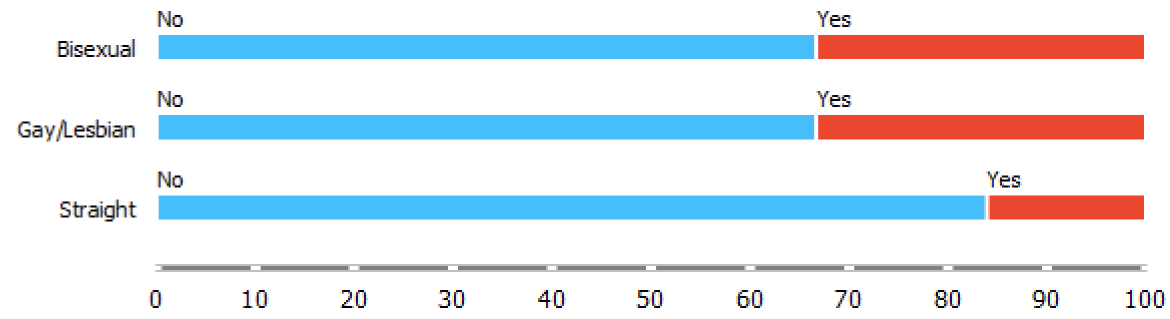


*Figure 3 Suicide Attempt vs Sexual Orientation*

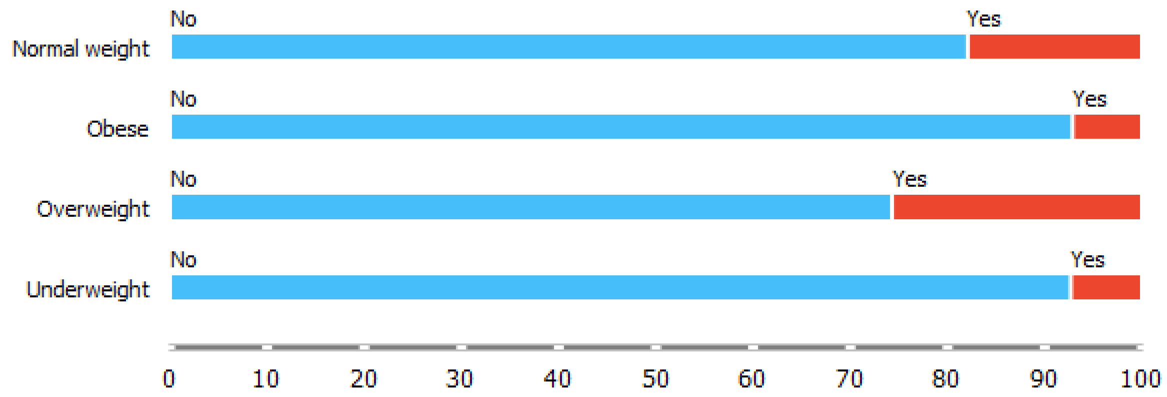


*Figure 4 Suicide Attempt vs Body Weight*

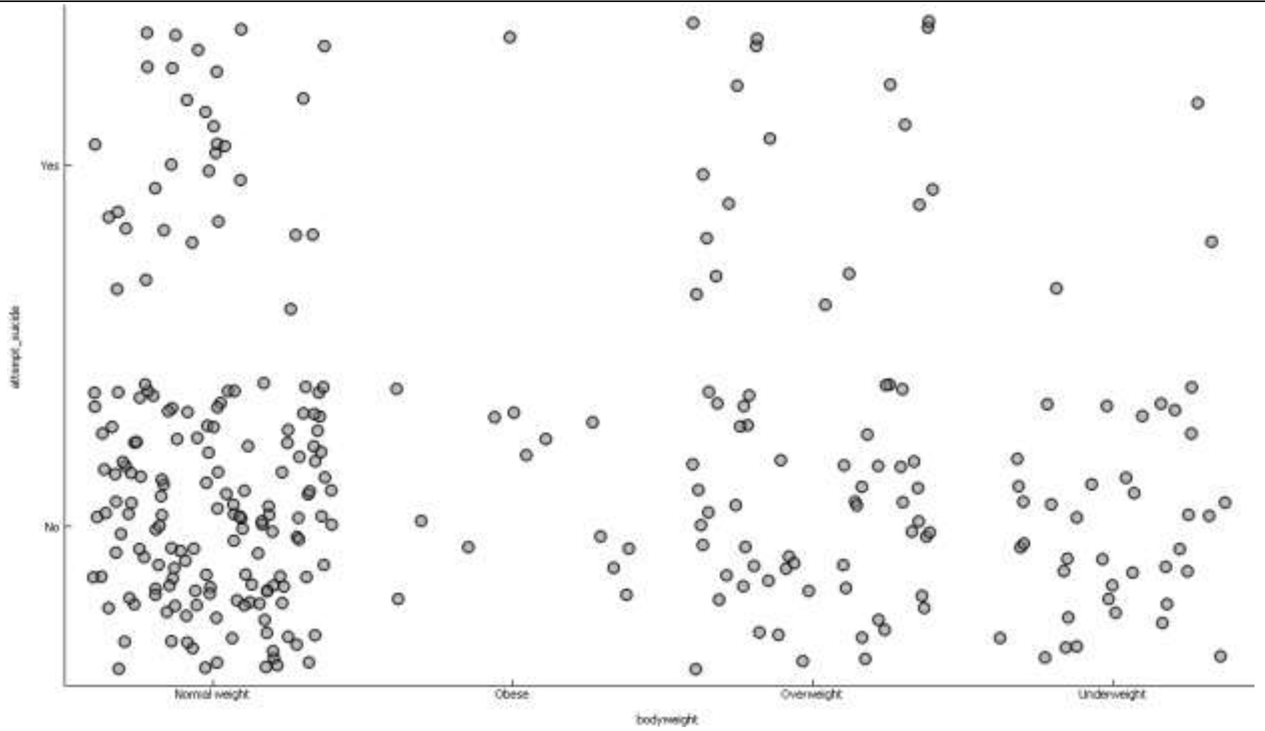


*Figure 5 Bodyweight vs Suicide Attempt Scatter Plot*

# CONCLUSION

## PROGRAM OUTPUT:

```
Confusion Matrix and Statistics


pc     No Yes
  No  134  24
  Yes   0   3

               Accuracy : 0.8509
                 95% CI : (0.7864, 0.9021)
    No Information Rate : 0.8323
    P-Value [Acc > NIR] : 0.305

                  Kappa : 0.1722
 Mcnemar's Test P-Value : 2.668e-06

            Sensitivity : 0.11111
            Specificity : 1.00000
         Pos Pred Value : 1.00000
         Neg Pred Value : 0.84810
             Prevalence : 0.16770
         Detection Rate : 0.01863
   Detection Prevalence : 0.01863
      Balanced Accuracy : 0.55556

       'Positive' Class : Yes
```

## SCOPE OF THE PROJECT

The project was constructed with an intent to make the entire process of understanding teenage emotion a lot better than what it is understood to be now. Especially in the teenage years, it can be much more of a sentimental value when they feel that they are not understood or that the age gap is preventing people from understanding them. It is regularly contended that the most satisfying phase of an individual's life is amid pre-adulthood, while others trust that adulthood, in spite of things like a vocation, family and cash concerns, is better. This exposition concurs with the previous, as opposed to the later view. It will initially talk about how adults are considerably less satisfied than youngsters as a result of the weights they are under and afterward examine how mollified most adolescents are, before arriving at the resolution that the ages of thirteen to eighteen truly are the greatest long periods of our lives.

After achieving development individuals are relied upon to battle for themselves and this frequently prompts misery. This is on the grounds that most grown-ups have lease and bills to pay, just as an accomplice and wards to take care of, which as a rule, prompts them carrying out a responsibility they loath in return for cash. An ongoing report by Queen's University, Belfast found that 79% of individuals would find employment elsewhere on the off chance that they didn't have a home loan and 64% of individuals expressed that their activity made them discouraged sooner or later.

Then again, youngsters are free from these stresses since they are frequently upheld monetarily and free from any genuine inconveniences. Most live with their folks who pay for every one of their needs and the main thing they need to concentrate on is considering. Research completed by Cambridge University found that just 29% of 15-multi year old understudies said they were 'glad', yet when addressed 10 years after the fact 84% said that they were 'a lot more joyful' when they were 16.

All in all, adolescent truly is squandered on the youthful on the grounds that more established individuals are under considerably more strain with regards to cash and individuals depending on them.

Individuals frequently wonder whether immaturity is superior to adulthood. Some think so while others oppose this idea. As I would like to think, each phase in life has its advantages and disadvantages.

Puberty is absolutely a pleasant timeframe. Adolescents have couple of things to stress over. They are not required to acquire cash or bolster a family. Thus, a significant number of them have a lot of leisure time that they can use to take part in exercises they appreciate. Indeed, even the criminal law is thoughtful to adolescents. Offenses submitted by teenagers pull in less serious disciplines than offenses submitted by grown-ups. On the other side, young is likewise a time of incredible passionate unrest. Youngsters are neither children nor grown-ups and have a solid requirement for opportunity. They don't care for it when their folks or educators endeavor to influence them to carry on.

## CONCLUSION

Conversely, grown-ups are increasingly develop and capable. A large number of them have a vocation to meet their costs. They have less extra time than teenagers, yet the greater part of them just couldn't care less. They are increasingly centered around their vocations and need to accomplish something throughout everyday life. Since their family and the law treat them as develop people, they appreciate a more noteworthy dimension of opportunity. For instance, they can choose what they need to do with their life. They can pick a vocation of their preferring or they can decide not to work by any means. While teenagers are minors grown-ups are majors fit for going into contracts enforceable by the law.

Subsequent to taking a gander at the two sides of the circumstance, it isn't difficult to see that pre-adulthood and adulthood have their upsides and drawbacks. As I would see it, it is in this way wrong to guarantee that one phase of life is preferable or more joyful over another stage.

There are various reasons why a young person may wind up discouraged. For instance, youngsters can create sentiments of uselessness and deficiency over their evaluations. School execution, societal position with friends, sexual introduction, or family life can each majorly affect how a teenager feels. Once in a while, adolescent despondency may result from natural pressure. However, whatever the reason, when companions or family - or things that the teenager generally appreciates - don't improve his or her pity or feeling of separation, there's a decent shot that the person in question has high schooler sadness.

Regularly, kids with youngster gloom will have a discernible change in their reasoning and conduct. They may have no inspiration and even turned out to be pulled back, shutting their room entry way after school and remaining in their space for a considerable length of time.

Children with youngster discouragement may rest too much, have an adjustment in dietary patterns, and may even show criminal practices, for example, DUI or shoplifting.\

Misery, which ordinarily begins between the ages of 15 and 30, here and there can keep running in families. Truth be told, teenager discouragement might be increasingly normal among young people who have a family ancestry of dejection.

There aren't a particular restorative tests that can recognize misery. Social insurance experts decide whether a high schooler has wretchedness by leading meetings and mental tests with the adolescent and his or her relatives, instructors, and friends.

The seriousness of the high schooler melancholy and the danger of suicide are resolved dependent on the appraisal of these meetings. Treatment proposals are likewise made dependent on the information gathered from the meetings.

The specialist will likewise search for indications of conceivably existing together mental clutters, for example, nervousness or substance misuse or screen for complex types of melancholy, for example, bipolar confusion (hyper burdensome disease) or psychosis. . The specialist will likewise survey the adolescent for dangers of self-destructive or maniacal highlights. Frequencies of endeavored suicide and self-mutilation is higher in females than guys while finished suicide is higher in guys. A standout amongst the most powerless gatherings for finished suicide is the 18-24 age gathering.

There are an assortment of strategies used to treat melancholy, including drugs and psychotherapy. Family treatment might be useful if family struggle is adding to a teenager's dejection. The teenager will likewise require support from family or educators to help with any school or friend issues. Once in a while, hospitalization in a mental unit might be required for youngsters with extreme discouragement.

Your emotional wellness care supplier will decide the best course of treatment for your high schooler.

The FDA cautions that upper drugs can, once in a while, increment the danger of self-destructive reasoning and conduct in kids and young people with dejection and other mental disarranges. Util ization of antidepressants in more youthful patients, consequently, requires particularly close checking and follow-up by the treating specialist. On the off chance that you have questions or concerns, talk about them with your human services supplier.

Adolescent suicide is a significant issue. Juvenile suicide is the second driving reason for death, following mishaps, among youth and youthful grown-ups in the U.S. It is evaluated that 500,000 adolescents endeavor suicide consistently with 5,000 succeeding. These are plague numbers.

Family challenges, the passing of a friend or family member, or saw disappointments at school or seeing someone would all be able to prompt negative sentiments and despondency. Also, youngster gloom frequently influences issues to appear to be overpowering and the related agony agonizing. Suicide is a demonstration of distress and teenager sorrow is regularly the main driver.

Puberty is a defenseless formative phase of life, ready with state of mind, uneasiness, thought and psychosocial issue. Furthermore, Indian youth is loaded tests, desires, peer weight, savagery and approaching social strains. Is anyone shocked that youths fall prey to despondency at this naive age?

As indicated by a cross-sectional investigation distributed in Indian Journal of Psychological Medicine, emotional well-being education in young people at the pre-college arrange is wretchedly low. Just 29 percent of those reviewed could recognize sadness as an ailment that requires proficient intercession. A negligible 1.31 percent could recognize schizophrenia or psychosis. This demonstrates emotional wellness is a dismissed and fringe to frames of mind and interests throughout everyday life. The investigation specifies that young people favored contacting relatives, particularly moms, instead of look for expert help.

In India, around 200 million are assessed to experience the ill effects of wretchedness sooner or later in their lives. Up until now, Indian safety net providers don't cover mental conditions in their approaches despite the fact that an ongoing round issued by Insurance Regulatory and Development Authority of India (IRDAI) under the Mental Health Act, 2017, guides them to do as such. Truth be told, mental prosperity is a pre-imperative for acquiring medical coverage.

With expanding consumption on wellbeing conditions, protection has turned out to be compulsory on the off chance that one is to get quality treatment and patient consideration.

One can tell if an adolescent is a casualty of discouragement in the event that the individual is always dismal and indicates loss of enthusiasm for exercises. The young person will normally fail to meet expectations at school or college, be cumbersome in social circumstances, feel baffled or even irate over little issues, be peevish and effectively irritated, feel sad, vacant, useless, be focused on past disappointments and enjoy misrepresented self-fault and self-analysis. They may express extraordinary affectability toward dismissal or misfortune and look for over the top consolation. Thinking and recalling things would wind up troublesome. A discouraged young person may feel directionless and futureless and enjoy considerations of passing on, torment and suicide.

Gloom can show in physical issues also. A consistent sentiment of tiredness, sleep deprivation, change of craving, eagerness or an overall condition of fomentation, moderate points of view, disc ourse or development, visit grumblings of body hurts or cerebral pains, social detachment, absence of enthusiasm for individual cleanliness and appearance and a propensity toward self-hurt are run of the mill conduct changes that can be seen as side effects of melancholy.

Guardians must comprehend that sorrow in their young youths is certifiably not an indication of shortcoming or an imperfection in their character. Wretchedness among young people is frequently overlooked on the grounds that individuals ascribe their tension to seething hormones and ideas like age hole. In any case, this is a genuine ailment that can influence learning objectives, physical movement and gainful potential, long into adulthood. Consequently, looking for assistance from a specialist, an attendant or an otherworldly pioneer is constantly fitting over leaving your tyke to adapt without anyone else.

# References

[1] Exploratory Analysis of Social Media Prior to a Suicide Attempt (Glen Coppersmith, Kim Ngo, Ryan Leary, Anthony Wood 2016)

[2] Linguistic Inquiry Word Count (Tausczik and Pennebaker, 2010; Pennebaker et al., 2007; Pennebaker et al., 2001)

[3] Media Contagion and Suicide Among The Young (Madelyn Gould Columbia University, Patrick Jamieson Daniel Romer University Of Pennsylvania 2003)

[4] Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time (Jean M. Twenge, Thomas E. Joiner, Megan L. Rogers, and Gabrielle N. Martin San Diego State University and Florida State University, 2017)

[5] Maron, M. E. (1961). "Automatic Indexing: An Experimental Inquiry" Journal of the ACM

[6] ). Tackling the poor assumptions of Naive Bayes classifiers Rennie, J.; Shih, L.; Teevan, J.; Karger, D. (2003)