A
Summer Internship Report
On
"Vachanamrut Visualization"

(CE346 – Summer Internship - I)

Prepared by

Bhishm Daslaniya (17CE023)

Under the Supervision of

Prof. Aayushi Chaudhari

Submitted to

Charotar University of Science & Technology (CHARUSAT)
for the Partial Fulfillment of the Requirements for the
Degree of Bachelor of Technology (B.Tech.)
for Semester 5

Submitted at

Accredited with Grade A by NAAC
Accredited with Grade A by KCG

U & P U. PATEL DEPARTMENT OF COMPUTER ENGINEERING
(NBA Accredited)
Chandubhai S. Patel Institute of Technology (CSPIT)
Faculty of Technology & Engineering (FTE), CHARUSAT
At: Changa, Dist: Anand, Pin: 388421.

July, 2019

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Accredited with Grade A by NAAC
Accredited with Grade A by KCG

CERTIFICATE

This is to certify that the report entitled "**Vachanamrut Visualization**" is a bonafied work carried out by **Bhishm Daslaniya (17CE023)** under the guidance and supervision of **Prof. Aayushi Chaudhari** & **Mr. Ravi Patel** for the subject **Summer Internship – I (CE346)** of 5th Semester of Bachelor of Technology in **Computer Engineering** at Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology & Engineering (FTE) – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred by the examiner(s).

Under the supervision of,

Prof. Aayushi Chaudhari
Assistant Professor
U & P U. Patel Dept. of Computer Engineering
CSPIT, FTE, CHARUSAT, Changa, Gujarat

Mr. Ravi Patel
Chief Technical Officer (CTO)
Milestone Technocrat Pvt. Ltd.

Dr. (Prof.) Amit Ganatra
Head - U & P U. Patel Department of Computer Engineering,
Dean - Faculty of Technology & Engineering (FTE),
CHARUSAT, Changa, Gujarat.

Chandubhai S. Patel Institute of Technology (CSPIT)

Faculty of Technology & Engineering (FTE), CHARUSAT

At: Changa, Ta. Petlad, Dist. Anand, Pin: 388421. Gujarat.

Table of Contents

Abstract

In today's era data has become more and more important but it is also important how we see that data means how we visualize that data in our mind. Data visualization is a quick, easy way to convey concepts in a universal manner. So most data analysts use data visualization techniques to represent data and convey the concept of data. The project entitled as **"Vachanamrut Visualization"** which is based on extracting important information from one of holy Swaminarayan scripture and visualize it by different visualization methods in R tool. Vachanamrut contains discourses of Bhagwan Swaminarayan and his conversation with saints and devotees. Vachanamrut provides spiritual guidance on how to remain happy in this life and beyond, how to attain complete progress in life, how to overcome even the most difficult situations. Vachanamrut is a unique work of prose in the Gujarati language. In Vachanamrut Bhagwan Swaminarayan imparted his wisdom in a clear and unique manner in a question and answer dialogue. Data in Vachanamrut is in a semi-structured format. In this project work we extract some important information from Vachanamrut like place where it disclosed, date, questions asked by different saints, length of particular Vachanamruts, important words used by Bhagwan Swaminarayan, Gap between two Vachanamruts, number of Vachanamrut in different seasons, months, years, etc. and visualize it by different charts. Complete visualization has been done using C#, SQLite, R, R Shiny.

Acknowledgement

First I express our heartiest thanks and gratefulness to Almighty God for His divine blessing makes me possible to complete the internship successfully in Milestone Technocrat Pvt. Ltd. I would like to express our deepest appreciation to all those who provided us the possibility to complete this internship and project. A special gratitude I give to our project guide **Mr. Ravi M. Patel (CTO of Milestone Technocrat)**, whose guidance helped us to complete our project.

I take this opportunity to present our votes of thanks to all guideposts who really acted as lightening pillars to enlighten our way throughout this project that has led to the successful and satisfactory completion of this internship.

I am really grateful and wish our profound our indebtedness to our counselor **Prof. Aayushi Chaudhari** her infinite patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this internship project documentation. Without this valuable support and guidance, this internship could not elevate up to this level of development from my point of view. I would like to express our heartiest gratitude to **Dr. (Prof.) Amit Ganatra,** HOD - U & PU Patel Department of Computer Engineering, Dean - Faculty of Technology & Engineering for giving this opportunity for this summer internship and also to other faculty members of the Computer Engineering Department of CHARUSAT University.

I would like to thank our entire course mate in CHARUSAT University, who took part of this discusses while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

Description of Company

Company name : Milestone Technocrat Pvt. Ltd

Location & Spread of Company: Multinational company located in 3 different countries of India , Canada and USA .

→ **India :** 501 samedh complex , Next to associate petrol pump Near, Panchavati Circle, Chimanlal Girdharlal Road, Ahmedabad, Gujarat – 380006

→ **USA California :** 909 S Central Ave, Compton, CA 90220, USA

→ **Canada Toronto :** Apt 217, 1901 Martin Grove Rao Etobicoke, Buzz 1022, Toronto, ON M9V 3T1, Canada

Number of Employees : 25

My group : Total 3 – All are in data Visualization field

Main Functions of Company : Milestone Technocrat offer extended technology and domain solutions in design, development, deployment, and maintenance of swift applications from bottom to top and it also offer solutions for business performance optimization and increased innovation.

Milestone Technocrat expertise's in following services :

- IT Consulting
- Cloud Integration
- Embedded Services
- Enterprise Mobility
- Web Application Development
- Mobile Application Development
- Big Data Development
- E-commerce Solution
- Digital Marketing & SEO

Scalable Technology :

At Milestone Technocrat they service the world economy right from the average small and medium enterprises to large conglomerates. They offer quality and innovative services in; Mobile, Web, Embedded services, open source frameworks and IT consulting.

Values of Company : 6 pillars of success for sustainable growth, continuous profitability and total client satisfaction

Integrity, Excellence, Respect, Learn, Experience, Deliver

Working Days : Monday to Friday (All week) and Saturday (Odd Week)

Working hours : 9:15 AM to 7:00 PM

List of Figures

List of Tables

Chapter 1: Introduction

1.1. What is data visualization?

Data visualization is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data.

1.2. Why data visualization?

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data are easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

Data visualization can also:

- Identify areas that need attention or improvement.
- Clarify which factors influence customer behavior.
- Help you understand which products to place where.
- Predict sales volumes.

1.3. Project overview

In today's era data visualization is one of the most common methods to represent data in any field. We visualize some important data which is extracted from "The Vachanamrut" one of holy Swaminarayan scripture by charts and different visualization techniques and convey some concepts of Vachanamrut by using technology like R, R Shiny, C#, SQLite.

1.4. Purpose

Our main purpose is to extract some important information from Vachanamrut and visualize it by different visualization a technique so anyone can easily understand that data properly with its original meaning.

1.5. Technology and literature review

For this project we have used R programming language for data visualization. R possesses an extensive catalog of statistical and graphical methods. It includes machine learning algorithm, linear regression, time series, statistical inference to name a few. Most of the R libraries are written in R, but for heavy computational task, C, C++ and Fortran codes are preferred.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

R is not only entrusted by academic, but many large companies also use R programming language, including Uber, Google, Airbnb, Facebook and so on.

Data analysis with R is done in a series of steps; programming, transforming, discovering, modeling and communicate the results.

- **Program:** R is a clear and accessible programming tool
- **Transform:** R is made up of a collection of libraries designed specifically for data science
- **Discover:** Investigate the data, refine your hypothesis and analyze them
- **Model:** R provides a wide array of tools to capture the right model for your data
- **Communicate:** Integrate codes, graphs, and outputs to a report with R Markdown or build Shiny apps to share with the world.

Chapter 2: System Analysis

2.1. Vachanamrut overview

The Vachanamrut is a sacred scripture containing Bhagwan Swaminarayan's divine discourses. These discourses were compiled by four of his scholarly renunciates – Gopalanand Swami, Muktanand Swami, Nityanand Swami and Shukanand Swami. These paramhansas even showed the manuscripts to Bhagwan Swaminarayan during its compilation, adding further authority to this unique scripture.

The compilers meticulously detailed all major aspects of the assembly setting. They noted the date, location and members of audience congregated before Bhagwan Swaminarayan. In many instances they even mention his style of sitting, the direction he was facing in, his clothing and adornment etc.

2.1.1. Focus

The primary emphasis of the Vachanamrut is ultimate spiritual knowledge, that is, explanation and analysis of the entities of Aksharbrahma and Parabrahma, as well as dharma, gnan, vairagya and bhakti – collectively known as 'Ekantik Dharma'. A variety of other topics are raised in the scripture, ranging from deeply philosophical subjects, such as the nature of God, creation of the cosmos and clarifications of classical Hindu philosophies.

The Vachanamrut provides spiritual guidance on how to remain happy in this life and beyond, how to attain complete progress in life, how to overcome even the most difficult situations.

Each page sheds new light, gives new inspiration, provides spiritual strength and gives constant encouragement. It exposes life's root problems and offers lasting solutions.

2.1.2. Style

The Vachanamrut is a unique work of prose in the Gujarati language. Despite its complicated spiritual subject-matter, Bhagwan Swaminarayan explains complex concepts through simple reasoning, scriptural allegories, proverbs, analogies and

simple everyday examples. He imparted his wisdom in a clear and unique manner in a question and answer dialogue.

2.2. Available resources

Resources available for Vachanamrut is,

- www.anirdesh.com (Vachanamrut is also available on many other sites also but this site gives us better data to process.)
- Vachanamrut Study App (By BAPS)
- It is also available in the form of eBook

2.3. Format

In this project we use Vachanamruts download from www.anirdesh.com .

- It is available in Gujarati and English languages (also available in other languages but here we are not using any other language format)
- For Gujarati it uses Unicode standards

2.3.1. Unicode standard

The Unicode Standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world. In addition, it supports classical and historical texts of many written languages.

The Unicode Standard provides a unique number for every character, no matter what platform, device, application or language. It has been adopted by all modern software providers and now allows data to be transported through many different platforms, devices and applications without corruption. Support of Unicode forms the foundation for the representation of languages and symbols in all major operating systems, search engines, browsers, laptops, and smart phones—plus the Internet and World Wide Web (URLs, HTML, XML, CSS, JSON, etc.). Supporting Unicode is the best way to implement ISO/IEC 10646.

2.4. Classification

We can not classified Vachanamrut directly because it is in unstructured format but by analyze Vachanamrut we conclude that it is like semi-structured format because its divided in two parts. First paragraph always includes all small details like date, location, etc. After that all paragraphs contains core Vachanamrut part.

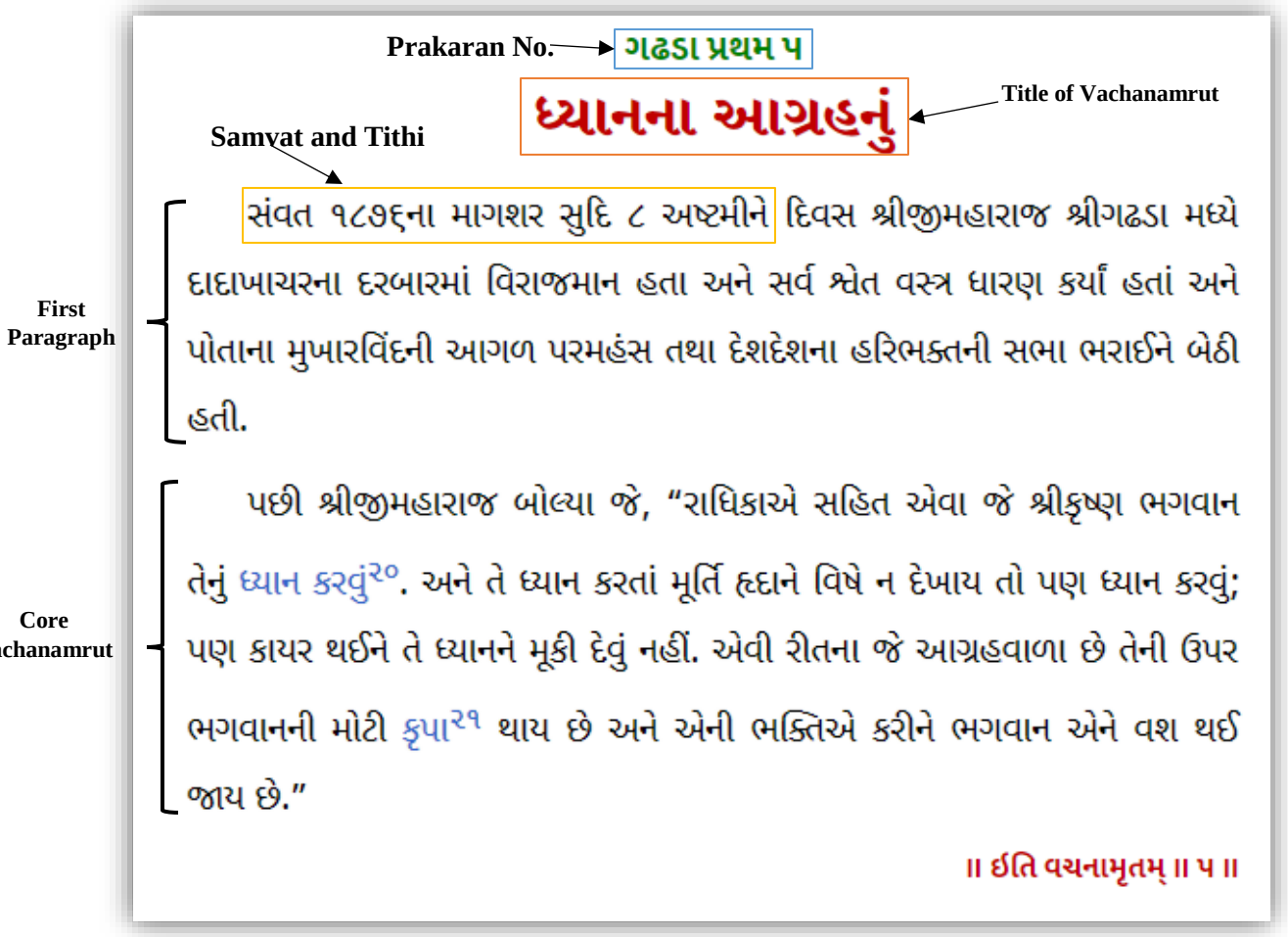


Figure 2.4.1 Semi-structured format of Vachanamrut

Chapter 3: System Design

3.1. Data scraping

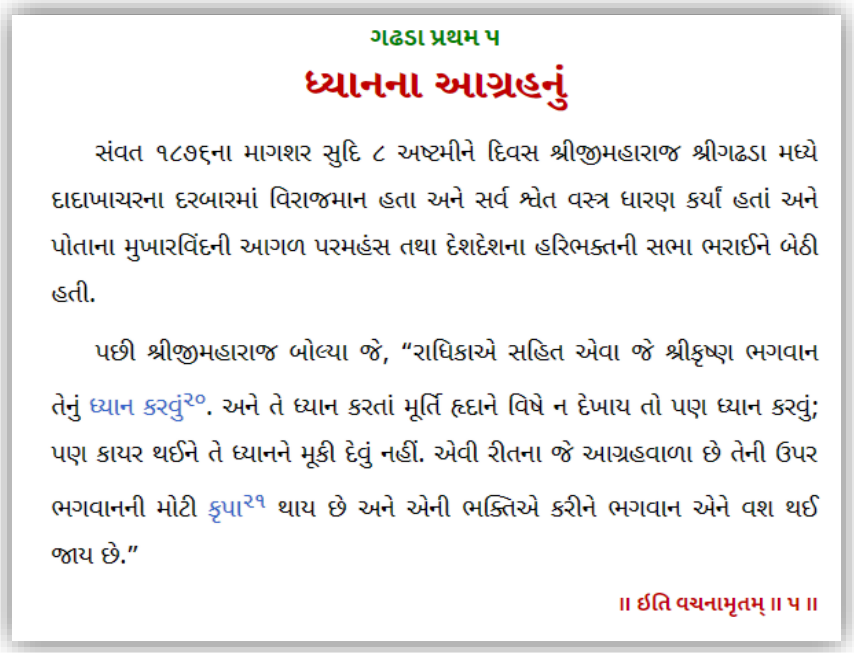
3.1.1. Definition

Data scraping is commonly defined as a process of extracting useful information from sources such as web pages or documents. Typically, source data is unstructured. Data scraping provides results for a variety of uses and automates aspects of data aggregation.

3.1.2. Vachanamrut data scraping

We extracted Vachanamrut from www.anirdesh.com. The website contains entire Vachanamrut granth in Unicode. A single vachanamrut can be fetched with a URL as follows: https://anirdesh.com/vachanamrut/index.php?format=gu&vachno=1

The HTML retrieved from the above URL contains a vachanamrut as well as other elements such as site header, footer, navigation, tooltips etc. We downloaded it with html extension. A C# program was written to download such HTML pages for all vachanamruts and then extract only the vachanamrut text out of it (filtering out unwanted elements such as menu, navigation etc.)



ગઢડા પ્રથમ ૫

ધ્યાનના આગ્રહનું

સંવત ૧૮૭૬ના માગશર સુદિ ૮ અષ્ટમીને દિવસ શ્રીજીમહારાજ શ્રીગઢડા મધ્યે દાદાખાચરના દરબારમાં વિરાજમાન હતા અને સર્વ શ્વેત વસ્ત્ર ધારણ કર્યાં હતાં અને પોતાના મુખારવિંદની આગળ પરમહંસ તથા દેશદેશના હરિભક્તની સભા ભરાઈને બેઠી હતી.

પછી શ્રીજીમહારાજ બોલ્યા જે, "રાધિકાએ સહિત એવા જે શ્રીકૃષ્ણ ભગવાન તેનું ધ્યાન કરવું[૨૦]. અને તે ધ્યાન કરતાં મૂર્તિ હૃદાને વિષે ન દેખાય તો પણ ધ્યાન કરવું; પણ કાયર થઈને તે ધ્યાનને મૂકી દેવું નહીં. એવી રીતના જે આગ્રહવાળા છે તેની ઉપર ભગવાનની મોટી કૃપા[૨૧] થાય છે અને એની ભક્તિએ કરીને ભગવાન એને વશ થઈ જાય છે."

॥ ઇતિ વચનામૃતમ્ ॥ ૫ ॥

Figure 3.1.2. Vachanamrut sample

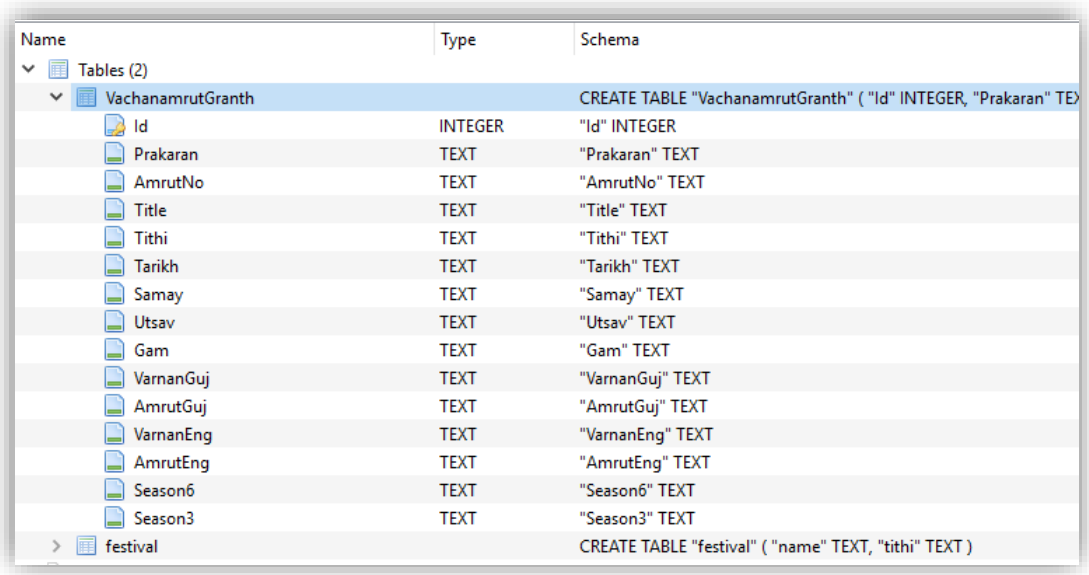
Here we give sample image of Vachanamrut format in which we downloaded it from www.anirdesh.com .

3.2. Column identification and data manipulation

After analyzing some Vachanamruts format we conclude that we can convert all Vachanamruts data in relational database. For that purpose we decide to divide each Vachanamrut in particular columns for Id of Vachanamrut, Title of Vachanamrut, First Paragraph of Vachanamrut (which contains place, date, audience, adornment etc details), Core Vachanamrut part, date of Vachanamrut, etc.

All are stored in separate columns with its own data type.

We create one database from all data files of Vachanamruts with C# script. After creating database with all data columns we created some extra calculated columns which is useful for visualization like Village name, Season, Festivals, etc with SQL queries.



Name	Type	Schema
Tables (2)		
VachanamrutGranth		CREATE TABLE "VachanamrutGranth" ("Id" INTEGER, "Prakaran" TE
Id	INTEGER	"Id" INTEGER
Prakaran	TEXT	"Prakaran" TEXT
AmrutNo	TEXT	"AmrutNo" TEXT
Title	TEXT	"Title" TEXT
Tithi	TEXT	"Tithi" TEXT
Tarikh	TEXT	"Tarikh" TEXT
Samay	TEXT	"Samay" TEXT
Utsav	TEXT	"Utsav" TEXT
Gam	TEXT	"Gam" TEXT
VarnanGuj	TEXT	"VarnanGuj" TEXT
AmrutGuj	TEXT	"AmrutGuj" TEXT
VarnanEng	TEXT	"VarnanEng" TEXT
AmrutEng	TEXT	"AmrutEng" TEXT
Season6	TEXT	"Season6" TEXT
Season3	TEXT	"Season3" TEXT
festival		CREATE TABLE "festival" ("name" TEXT, "tithi" TEXT)

Figure 3.2.1. Database table structure

3.3. GUI Forms

In this project we created simple GUI with R shiny library.

3.3.1. R shiny

Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R. Shiny helps you turn your analyses into interactive web applications without requiring HTML, CSS, or JavaScript knowledge.

3.3.2. GUI Layout

Figure 3.3.2.1. GUI Layout

Chapter 4: Implementation

4.1. Module specification

4.1.1. Important words module

In this module, we use wordcloud for visualizing the most frequently used words from the core Vachanamrut part. For creating wordcloud we create an algorithm that followed by all processes like Text cleaning, stemming, removing stopwords.

4.1.1.1. Wordcloud

A Word Cloud or Tag Cloud is a visual representation of text data in the form of tags, which are typically single words whose importance is visualized by way of their size and color. As unstructured data in the form of text continues to see unprecedented growth, especially within the field of social media, there is an ever-increasing need to analyze the massive amounts of text generated from these systems. A Word Cloud is an excellent option to help visually interpret text and is useful in quickly gaining insight into the most prominent items in a given text, by visualizing the word frequency in the text as a weighted list.

4.1.1.2. Algorithm for creating wordcloud:

→ Fetch table data, stemming file and stopwords file.

→ Fetch AmrutGuj column which contains core Vachanamrut part in one variable

→ Replace all special characters by space

Code Snippet: str_replace_all(Text_guj,'[\r\n.;“”"\',?()!‘’]',' ')

→ Perform stemming (use stemming.txt file)

Example of stemming :

દ્રષ્ટિ <= દ્રષ્ટિએ, દ્રષ્ટિને, દ્રષ્ટિમાં

Means all related words counted in its original word.

→ Remove stopwords (use stopwords.txt file)

→ Count frequency of all words

→ Generate Wordcloud

4.1.2. Vachanamrut varnan module

In this module, all the processes are the same as Important words module only difference is that in this module wordcloud is generated from Varnan Paragraph means only the first paragraph. Stemming file and stopwords file are different for both modules.

4.1.3. Question analysis module

In this module we extract the number of questions asked by per person by analyzing some string pattern in Vachanamrut. For visualization of this module we use simple pie chart and we also use special Pareto chart for visualizing the number of questions asked by a person with its contribution in the total number of questions in the form of a percentage.

In this module we get 90% accuracy in the extraction of the number of questions.

4.1.3.1. Pareto chart

A Pareto chart is a type of chart that contains both bars and a line graph, where individual values are represented in descending order by bars, and the cumulative total is represented by the line.

The purpose of the Pareto chart is to highlight the most important among a (typically large) set of factors.

4.1.4. Length analysis module

In this module, we calculate the length of each Vachanamrut in the form of the number of words. For visualization of that, we use histograms and Scatter plot.

4.1.4.1. Histogram

A histogram is an accurate representation of the distribution of numerical data. It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one.

Notice that, unlike a bar chart, there are no "gaps" between the bars. This is because a histogram represents a continuous data set, and as such, there are no gaps in the data.

4.1.4.2. Scatter plot

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded (color/shape/size), one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

4.1.5. Vachanamrut count module

In this module we count the number of Vachanamruts in particular month, year, season, village, tithi. For visualization of this module, we use bar charts. In this module we calculate the number of Vachanamruts in six different ways :

1. Seasonwise Vachanamruts
2. Villagewise Vachanamruts
3. Monthwise Vachanamruts
4. Tithiwise Vachanamruts
5. Samvatwise Vachanamruts
6. Yearwise Vachanamruts

4.1.6. Gap analysis module

In this module we calculate the gap between two consecutive Vachanamruts. For visualization of this module, we use a Bubble graph, Scatter plot, Histogram. For generating graphs we generate all the dates between the first date of Vachanamrut and the Last date of Vachanamrut. After that, we generate different graphs to visualize gaps between two Vachanamruts. Here the first date is 21/11/1819 and the last date is 25/7/1829 means the total number of days is 3535.

4.2. Coding standards of R

4.2.1. Syntax

- **Line Length:** The maximum line length is 80 characters.

- **Spacing :**

  Place spaces around all binary operators (=, +, -, <-, etc.).

  Exception: Spaces around ='s are optional when passing parameters in a function call.

  Do not place a space before a comma, but always place one after a comma.

```
tab.prior <- table(df[df$days.from.opt < 0, "campaign.id"])
total <- sum(x[, 1])
total <- sum(x[1, ])
```

- **Curly Braces:**

  An opening curly brace should never go on its own line; a closing curly brace should always go on its own line. You may omit curly braces when a block consists of a single statement; however, you must consistently either use or not use curly braces for single statement blocks.

```
if (is.null(ylim)) {
  ylim <- c(0, 0.06)
}
```

- **Assignment:**
  - Use <-, not =, for assignment.

  GOOD: x <- 5

  BAD: x = 5

- **Semicolons:**

  Do not terminate your lines with semicolons or use semicolons to put more than one command on the same line.

4.2.2. Organization

- **Function Definitions and Calls:**

  Function definitions should first list arguments without default values, followed by those with default values.

  In both function definitions and function calls, multiple arguments per line are allowed; line breaks are only allowed between assignments.

```
PredictCTR <- function(query, property, num.days,
                       show.plot = TRUE)
```

- **Function Documentation:**

  Functions should contain a comments section immediately below the function definition line. These comments should consist of a one-sentence description of the function; a list of the function's arguments, denoted by Args:, with a description of each (including the data type); and a description of the return value, denoted by Returns:. The comments should be descriptive enough that a caller can use the function without reading any of the function's code.

Chapter 5: Analysis of Visualization

5.1. Analysis of wordclouds

5.1.1. Important words

Figure 5.1.1.1. Wordcloud of core Vachanamrut

Analysis : From this wordcloud we can analyze that which words are most frequently used by Bhagwan Swaminarayan in Vachanamruts.

5.1.2. Bar graph of Important words

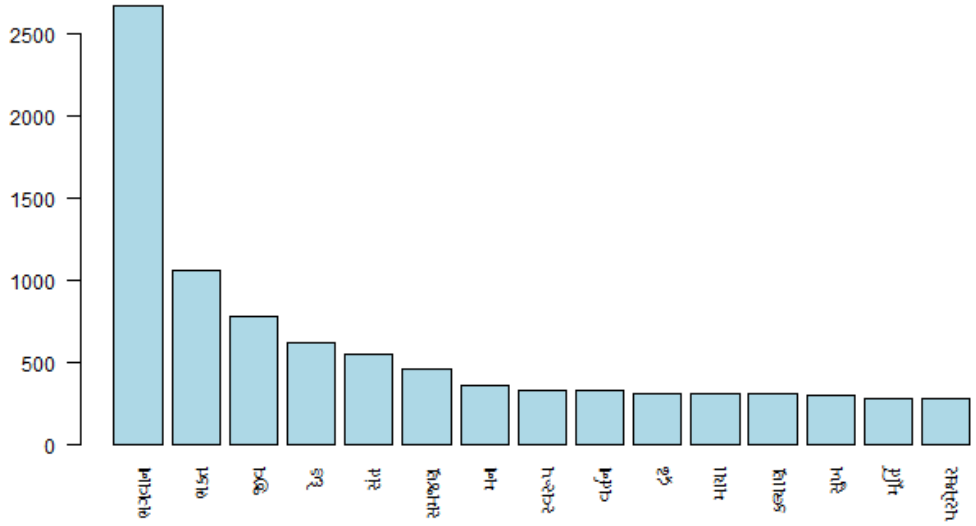


Figure 5.1.2.1. Bar graph of Important words

Table of analysis :

Words	Freq
ભગવાન	2671
ભક્ત	1057
જીવ	775
દેહ	618
સંત	548
સમજણ	459
મન	359
સ્વરૂપ	335
વર્તન	329
દ્રઢ	314
માયા	313
કલ્યાણ	306
સુખ	304
મૂર્તિ	278
પરમેશ્વર	277

Table 5.1.2.1. Table of frequency for Important words

5.1.3. Vachanamrut varnan

Figure 5.1.2.1. Wordcloud of varnan

Analysis : From this wordcloud of Varnan Paragraph, we can conclude that Bhagwan Swaminarayan mostly wears white cloths. Wordcloud gives us an idea of the variety of

cloths, It also shows that place where Bhagwan Swaminarayan mostly stayed, It also gives some idea of tithi related to Vachanamruts.

5.1.4. Bar graph of Vachanamrut varnan

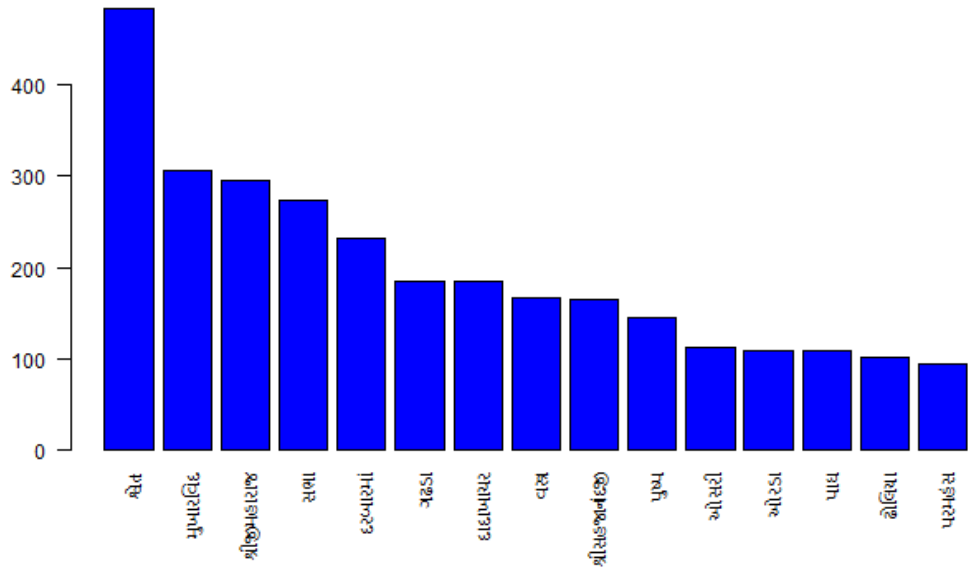


Figure 5.1.4.1. Bar graph for Vachanamrut varnan

Table of analysis:

Words	Freq
શ્વેત	484
મુખારવિંદ	306
શ્રીજીમહારાજ	295
સભા	274
દરબારમાં	231
ગઢડા	185
દાદાખાચર	184
વસ્ત્ર	166
શ્રીસહજાનંદજી	165
પુષ્પ	144
ઓસરી	113
ઓરડા	109
પાઘ	108
ઢોલિયા	101
પરમહંસ	95

Table 5.1.4.1. Table of frequency for Vachanamrut varnan

5.2. Question analysis

5.2.1. Pie chart

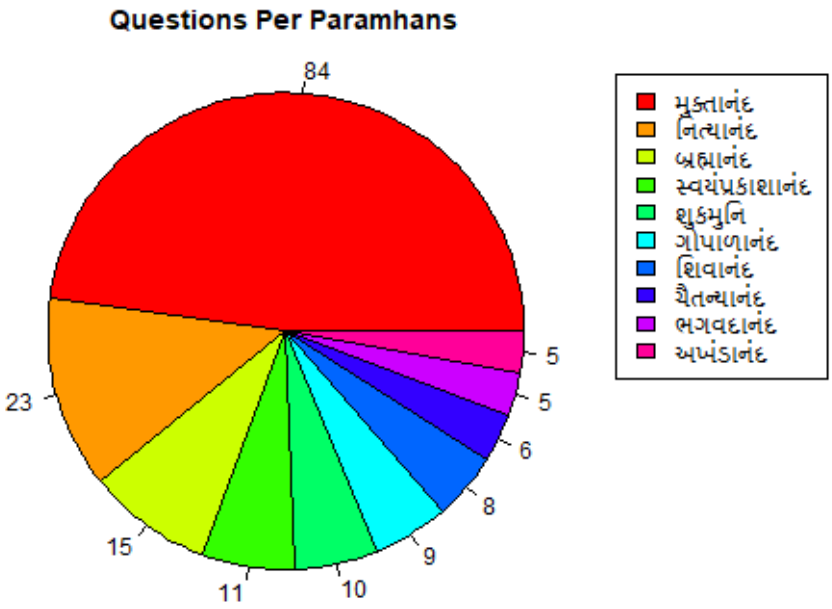


Figure 5.2.1.1. Pie chart of questions per person

Analysis table:

Paramhans_Name	No_Of_Questions
મુક્તાનંદ સ્વામી	84
નિત્યાનંદ સ્વામી	23
બ્રહ્માનંદ સ્વામી	15
સ્વયંપ્રકાશાનંદ સ્વામી	11
શુકમુનિ સ્વામી	10
ગોપાળાનંદ સ્વામી	9
શિવાનંદ સ્વામી	8
ચૈતન્યાનંદ સ્વામી	6
ભગવદાનંદ સ્વામી	5
અખંડાનંદ સ્વામી	5

Table 5.2.1.1. Table for number of questions per Paramhans

From this pie chart we can see the number of questions asked by per person/saint. Here we get 90% of accuracy in extract the number of questions per person.

5.2.2. Pareto chart

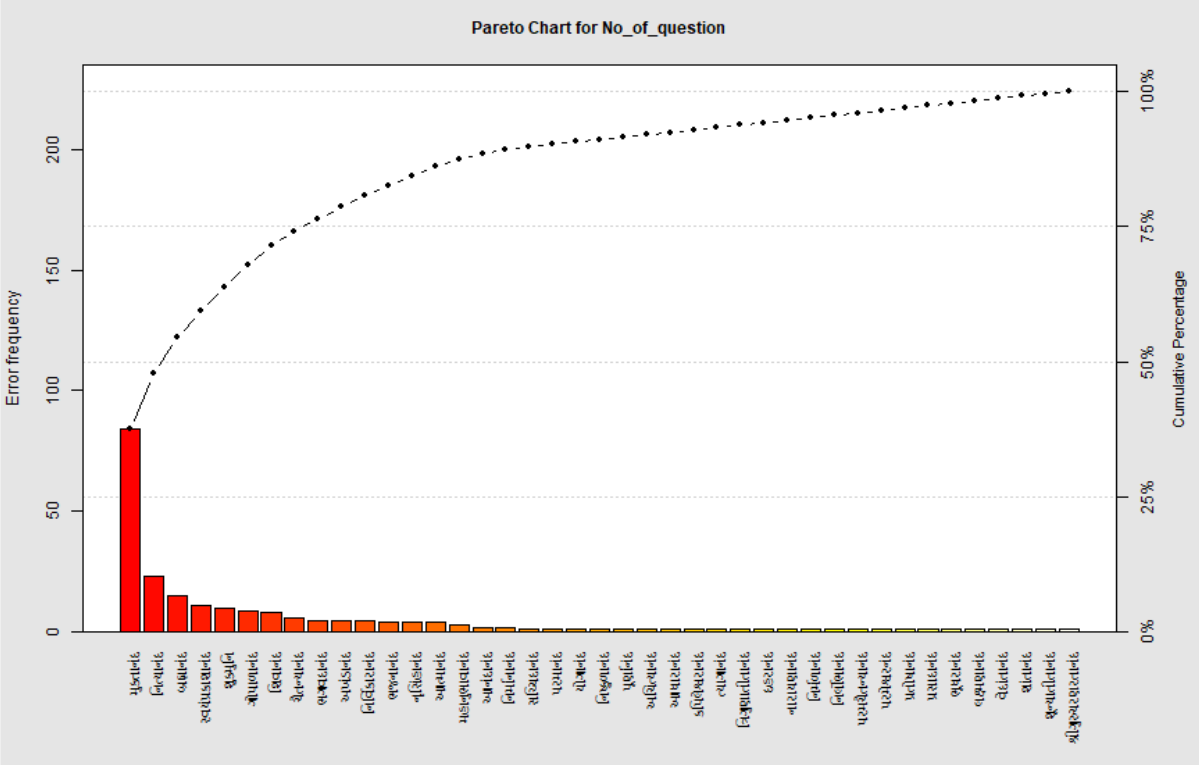


Figure 5.2.2.1. Pareto chart for question analysis

Analysis from pareto chart :

From this Pareto Chart, we can see that 80% of questions of total questions asked by the first 8-9 saints. Here 80-20 ratio type situation is there means 80% of questions asked by 20% of saints and other 20% questions asked by other 80%.of saints.

5.3. Length analysis

5.3.1. Histogram

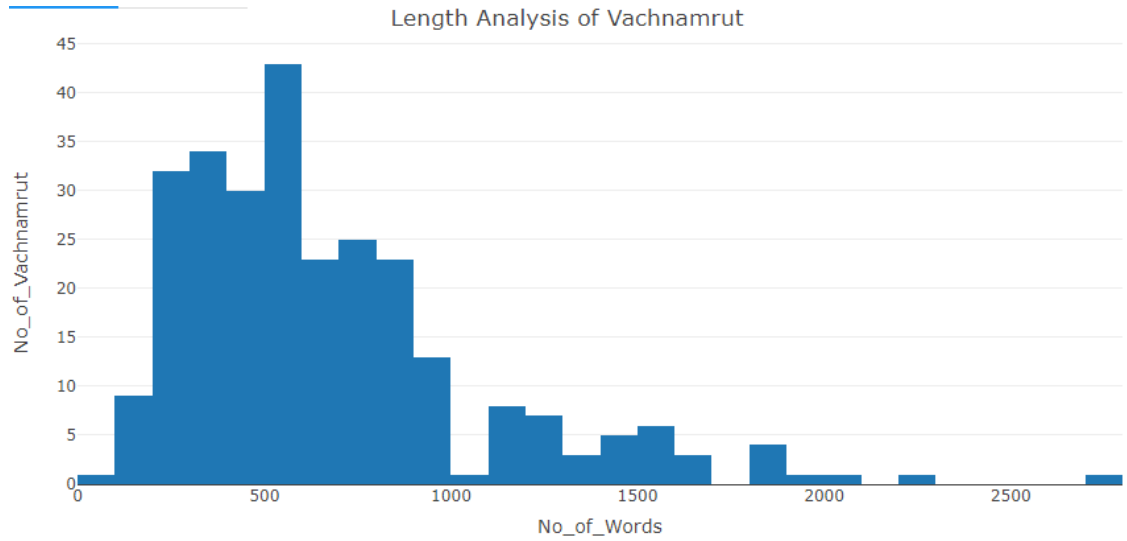


Figure 5.3.1.1. Histogram for length analysis

Analysis : From this histogram, we can say that most of the Vachanamruts are of length 500-1000 words. Rare Vachanamruts are the length of more than 1500 words.

5.3.2. Scatter plot

Figure 5.3.2.1. Scatter plot for length analysis

Analysis : From this Scatter plot, we can say that longest Vachanamrut is ગઢડા પ્રથમ ૭૮ and shortest Vachanamrut is ગઢડા પ્રથમ ૫. We can see increment in length on X-axis.The longest Vachanamrut is of length 2771 words, ગઢડા પ્રથમ ૭૮.The Shortest Vachanamrut is of length 89 words, ગઢડા પ્રથમ ૫.

5.4.Vachanamrut counts

5.4.1. Seasonwise vachanamruts

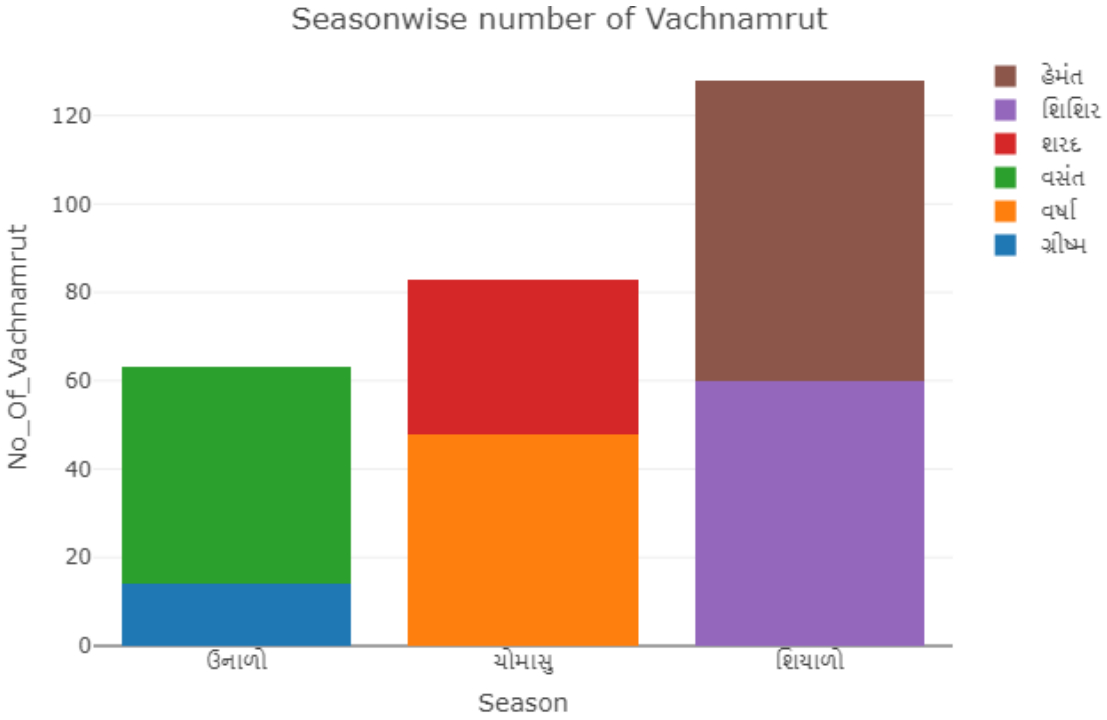


Figure 5.4.1.1. Stacked bar chart for Seasonwise Vachanamruts count

Analysis table:

Gujarati_Season	English_Season	No_Of_Vachnamrut
ગ્રીષ્મ	ઉનાળો	14
વર્ષા	ચોમાસુ	48
વસંત	ઉનાળો	49
શરદ	ચોમાસુ	35
શિશિર	શિયાળો	60
હેમંત	શિયાળો	68

Table 5.4.1.1 Table of seasonwise vachanamruts count

From this staked bar chart and table 5.4.1.1. we can see that more number of Vachanamruts are disclosed in winter. So we can say that winter was most preferred season for assemblies and summer was the least preferred. Even in summer, more hotter "ગ્રીષ્મ" has less vachnamruts than "વસંત" which is pleasant season.

5.4.2. Villagewise vachanamruts

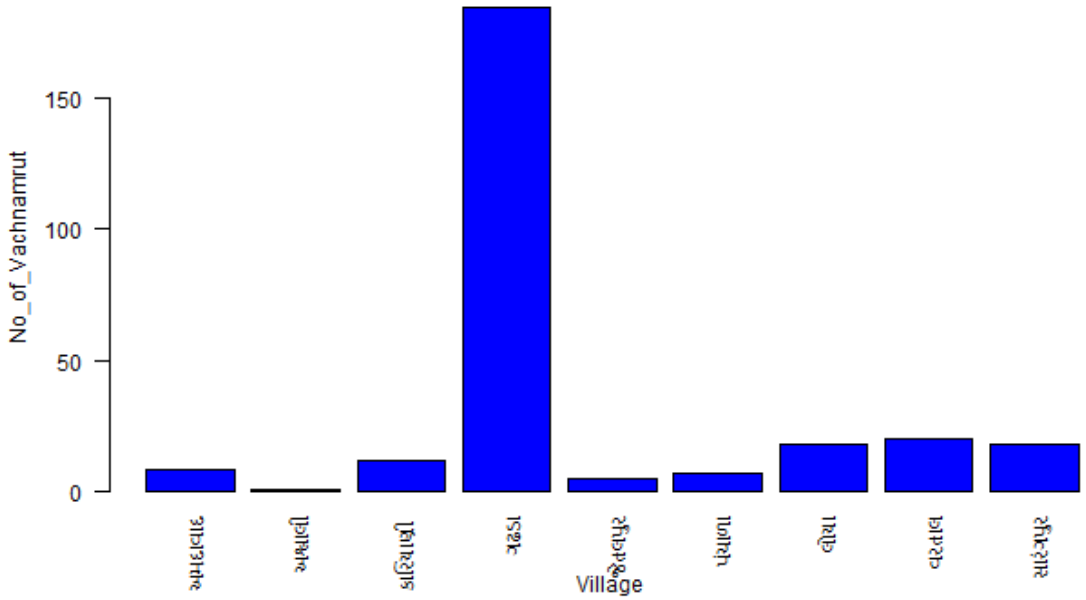


Figure 5.4.2.1. Bar chart for villagewise vachanamruts count

Analysis table:

Gam	No_of_vachnamrut
અમદાવાદ	8
અશ્લાલી	1
કારિયાણી	12
ગઢડા	185
જેતલપુર	5
પંચાળા	7
લોયા	18
વરતાલ	20
સારંગપુર	18

Table 5.4.2.1. Table of villagewise vachnamruts count

From this we can say that Bhagwan Swaminarayan spent more time of his life in ગઢડા because more number of Vachanamruts are in ગઢડા.

5.4.3. Monthwise vachanamruts

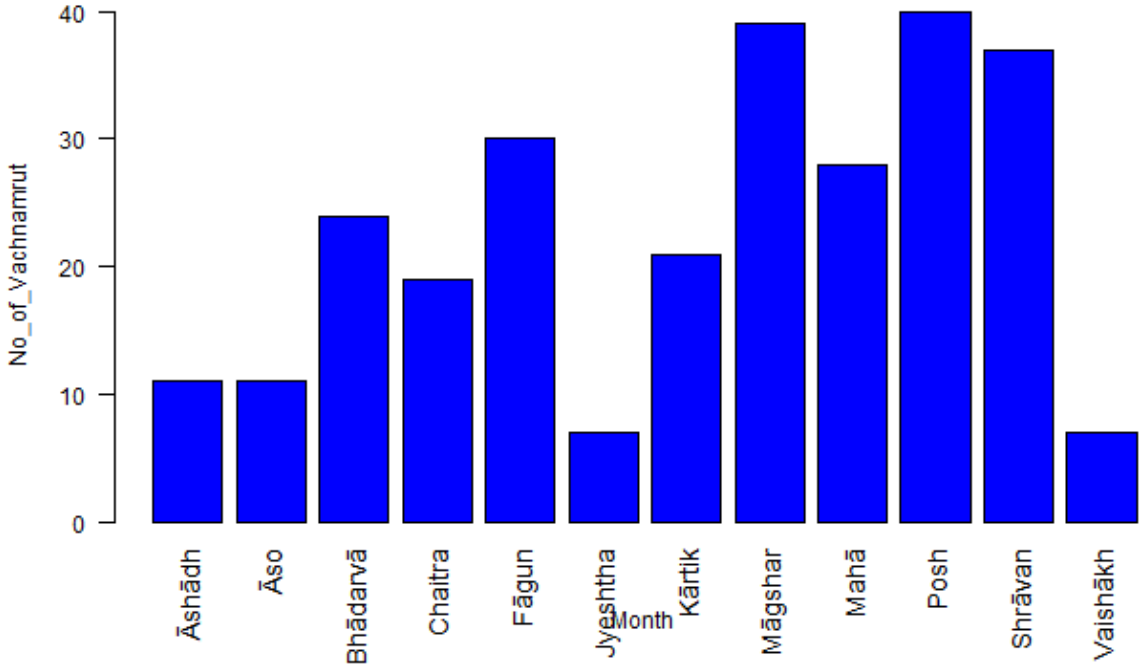


Figure 5.4.3.1. Bar chart for monthwise vachanamruts count

Analysis table:

Tithi	Count
Āshādh	11
Āso	11
Bhādarvā	24
Chaitra	19
Fāgun	30
Jyeshtha	7
Kārtik	21
Māgshar	39
Mahā	28
Posh	40
Shrāvan	37
Vaishākh	7

Table 5.4.3.1. Table of monthwise vachanamruts count

From this graph it is clear that Posh and Magshar was more preffered for assembly.

5.4.4. Tithiwise vachanamruts

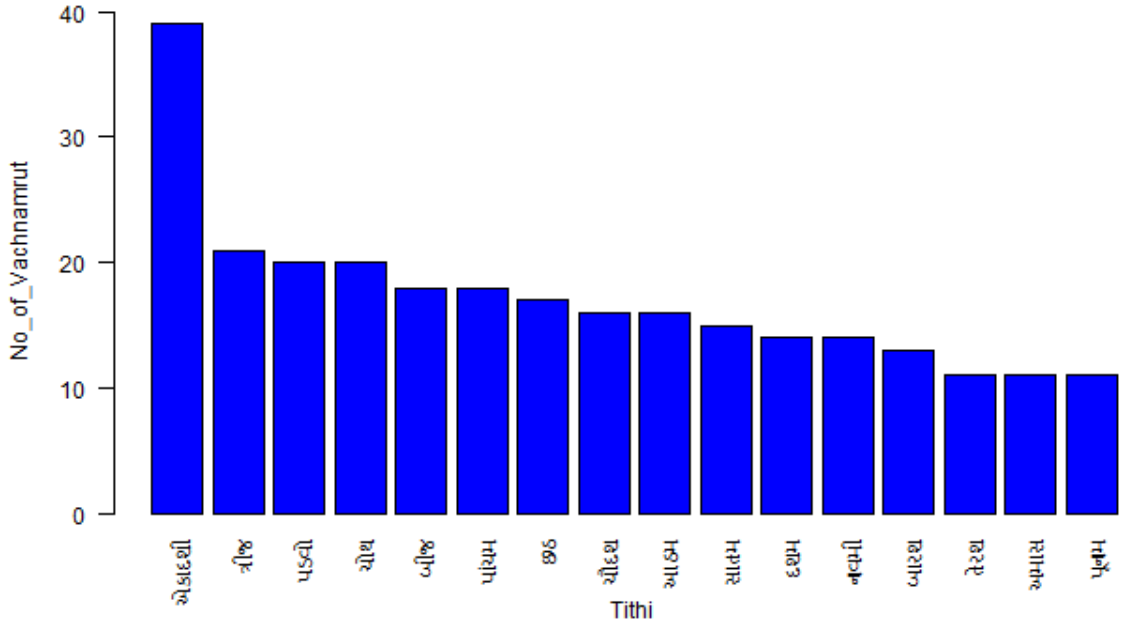


Figure 5.4.4.1. Bar chart for tithiwise vachanamruts count

Analysis table:

G_Tithi	No_of_vachnamrut
એકાદશી	39
ત્રીજ	21
પડવો	20
ચોથ	20
બીજ	18
પાંચમ	18
છઠ	17
ચૌદશ	16
આઠમ	16
સાતમ	15
દશમ	14
નવમી	14
બારશ	13
તેરશ	11
અમાસ	11
પૂનમ	11

Table 5.4.4.1. Table of tithiwise vachnamruts count

From this graph It is clear that Ekadashi was most preferred day of the month. It is also considered auspicious.

5.4.5. Samvatwise vachanamruts

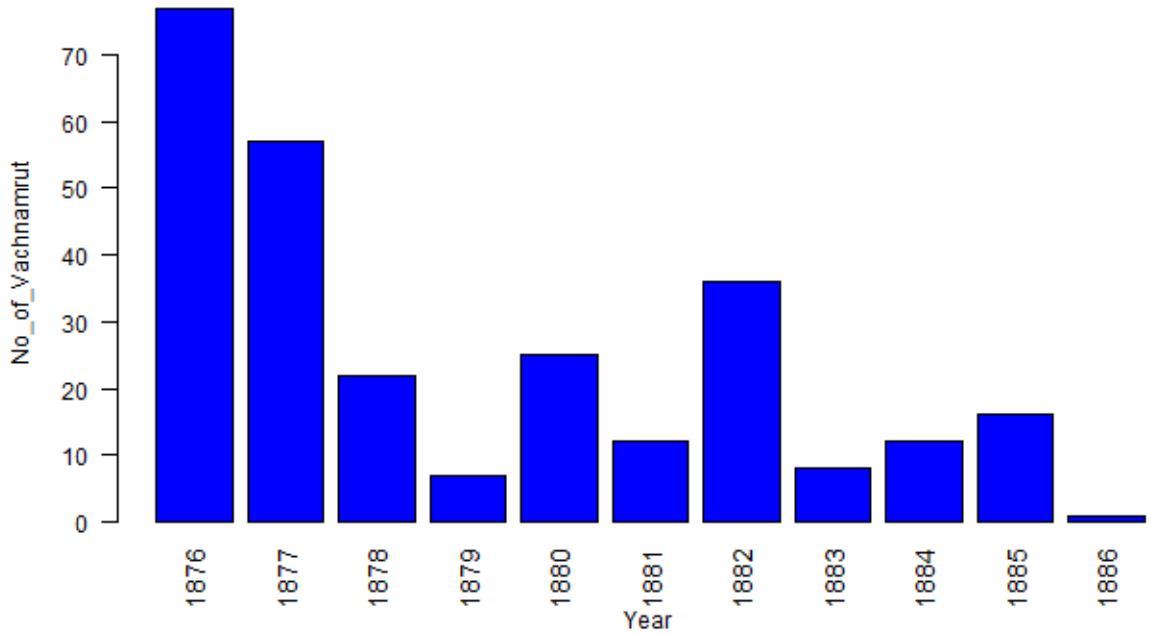


Figure 5.4.5.1. Bar chart for samvatwise vachanamruts count

Analysis table:

G_Year	No_of_vachnamrut
1876	77
1877	57
1878	22
1879	7
1880	25
1881	12
1882	36
1883	8
1884	12
1885	16
1886	1

Table 5.4.5.1. Table of samvatwise vachanamruts count

From this graph, we can see that in the first year the highest number of Vachanamrut is composed after that the number of Vachanamruts per year decreases and in last year only one Vachanamrut is composed.

5.4.6. Yearwise vachanamruts

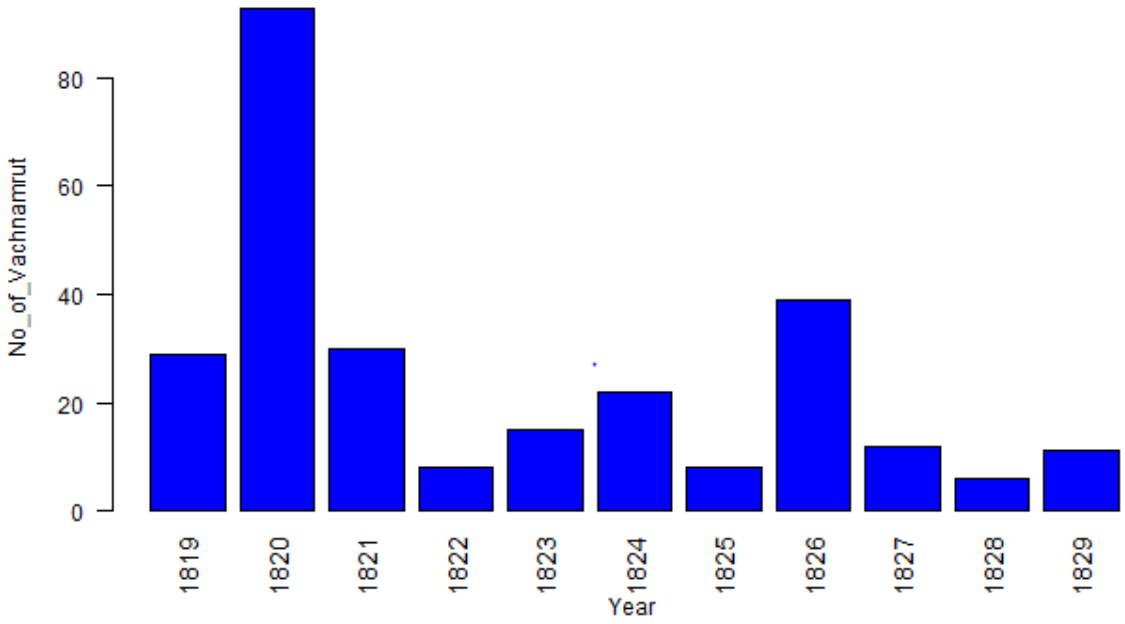


Figure 5.4.6.1. Bar chart for yearwise vachanamruts count

Analysis table:

G_Year	No_of_vachnamrut
1819	29
1820	93
1821	30
1822	8
1823	15
1824	22
1825	8
1826	39
1827	12
1828	6
1829	11

Table 5.4.6.1. Table of yearwise vachnamruts count

From this graph, we can analyze that in English Year the highest number of Vachanamruts was composed in 1820 that is 93 and the lowest number of Vachanamruts is composed in 1828 that is only 6.

5.5. Gap analysis

5.5.1. Bubble graph

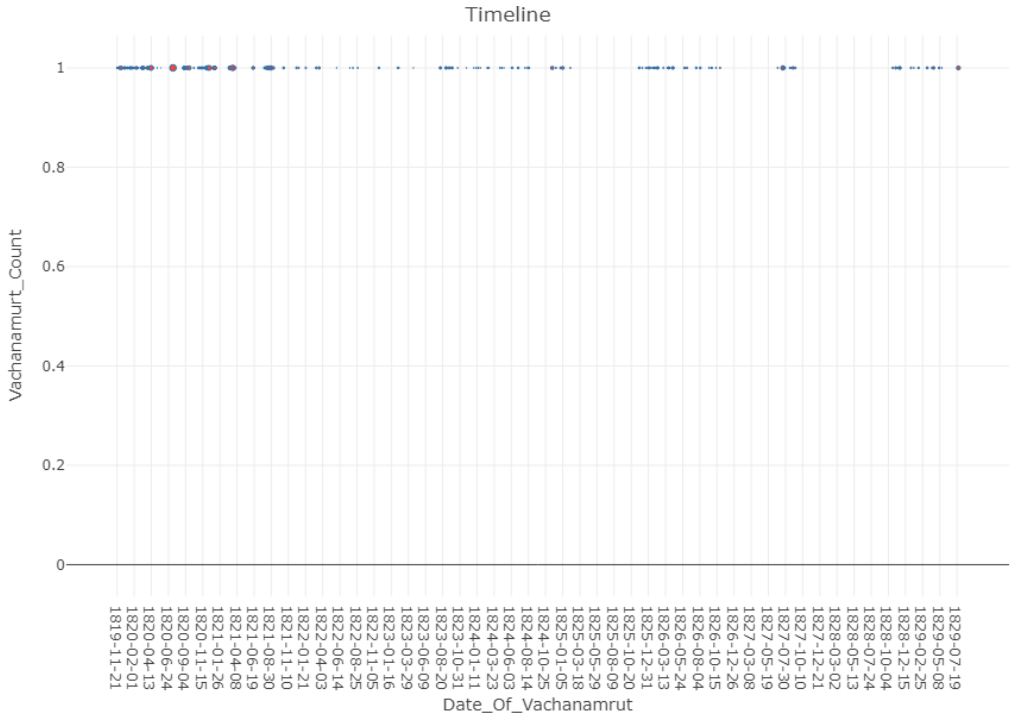


Figure 5.5.1.1. Bubble graph for gap analysis

5.5.2. Scatter plot

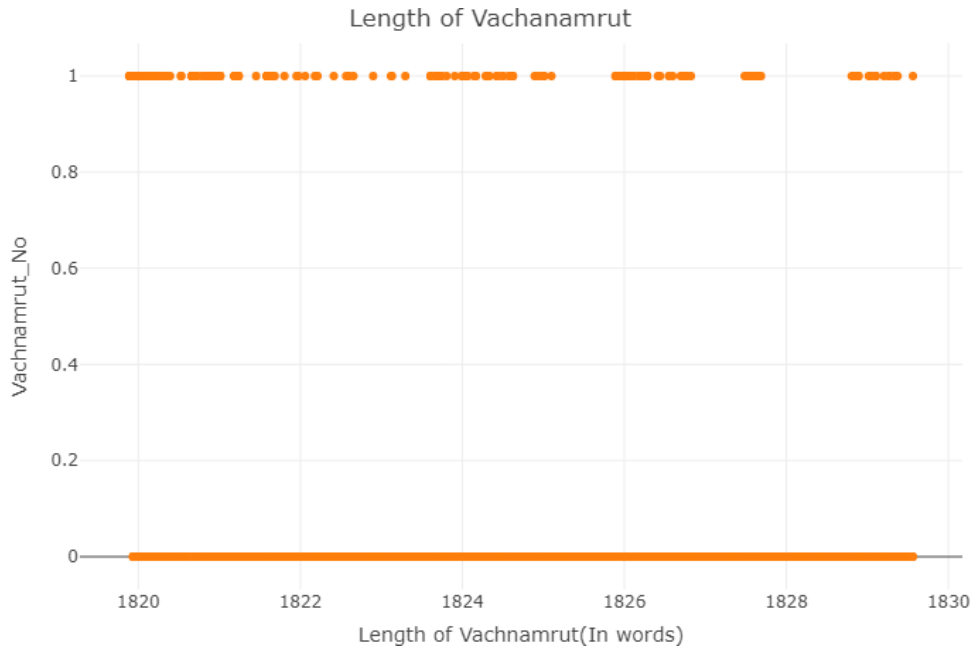


Figure 5.5.2.1. Scatter plot for gap analysis

Analysis from bubble graph and scatter plot :

From this bubble graph and scatter plot we can see the date difference between two Vachanamruts. We can assume that gap is because of some events happened at that time.

One of the event noted by Raymond Brady Williams in Introduction to Swaminarayan Hinduism.

"The third famine came in 1825. In this time of disaster and scarcity, armed bands roamed the countryside killing and looting, and the weak were cowed by violence. Bishop Reginald Heber reported as he traveled through Gujarat in 1825 that no area was more disturbed, so the exercise of authority was more expensive in Gujarat than elsewhere. Local chiefs and then the British maintained large armed forces to quell rebellions. Nevertheless, in no place was there more bloodshed or were the roads more insecure (R. Heber 1846 2:105). One writer summarized the situation in stark terms :

Never had there been such intense and general suffering in India; the native states were disorganized, and society on the verge of dissolution; the people crushed by despots and ruined by exactions; the country overrun by bandits and its resources wasted by enemies; armed forces existed only to plunder, torture and munity; government had ceased to exist; there remained only oppression and misery.(Dodwell 1963 : 367-7)."

So, we can assume that the famine is responsible for only 8 Vachanamruts in 1825. This is only one event we noticed, It may be possible that this type of more events were responsible for gaps.

5.5.3. Histogram

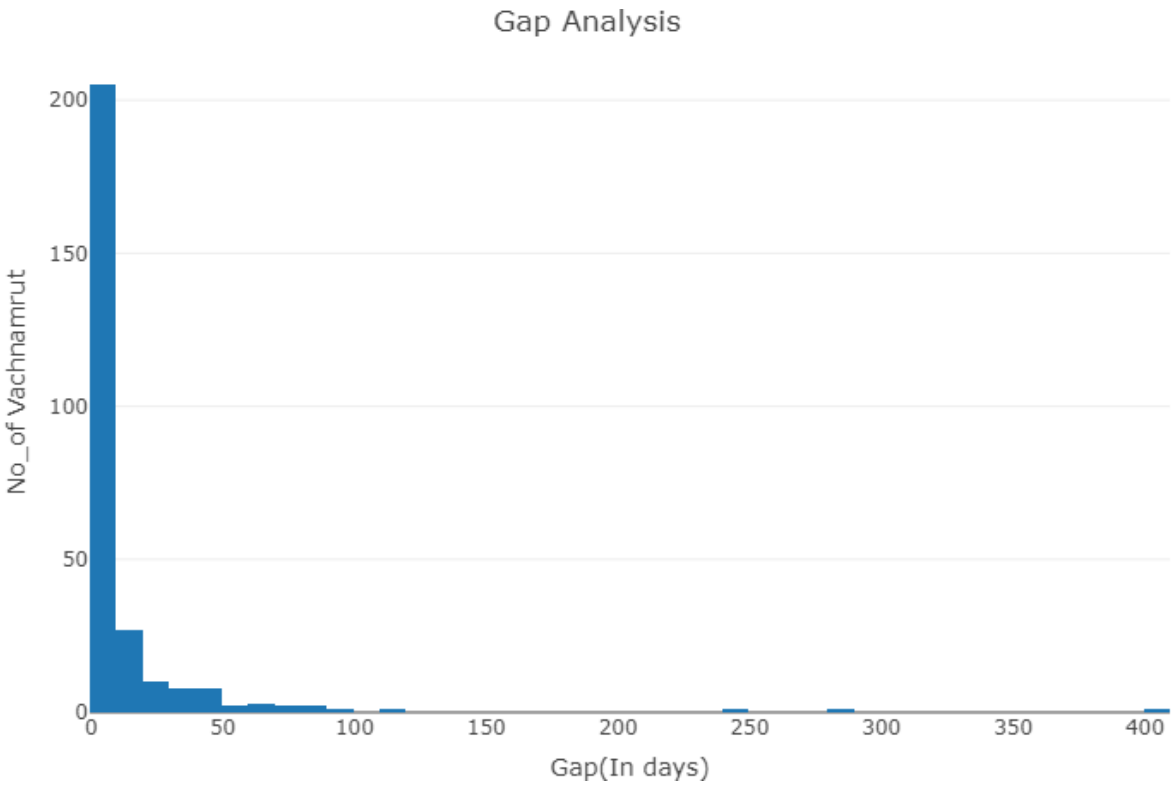


Figure 5.5.3.1. Histogram for gap analysis

Analysis from histogram:

From histogram we can see that there is more than 200 gaps between two Vachanamruts in the range of 0-10 days gap. Very few gaps indicates gap more than 50 days. There is 3 dates on which two Vachanamruts are disclosed on that date. Highest gap is 409 days gap between ગઢડા અંત્ય ૨૨ and ગઢડા અંત્ય ૨૩.

Chapter 6: Future Enhancement and Conclusion

6.1. Future Enhancement

- We are trying to get more accuracy in the question analysis module.
- As we know that in Vachanamrut there are question and answer between Bhagwan Swaminarayan and his devotees and saints. So Vachanamrut gives the solution of every real-life problem. So in future, we are going to add some machine learning and deep learning concepts in this project which analyze the problem of user and gives Vachanamruts which contains the solution of that problem.
- We are also trying to perform more text mining on Vachanamrut and try to extract more important information and visualize it.

6.2. Conclusion

In this project we faced many problems like dealing with the Gujarati language, making GUI in shiny, getting less accuracy, etc. So from this project, I have learnt so many new tools and technologies like R, R Shiny, C# and how to deal with non-English characters with Unicode standards. From this internship, we learnt about new concepts of Data analytics like Data scraping, Data mining, Text Cleaning, Text mining, data visualization, etc. This internship experience gives us technical skills as well as soft skills like communicate with professionals, work together with the team, be punctual, be humane, be attentive etc.

References

Weblinks:

- https://www.datacamp.com/courses/free-introduction-to-r
- https://www.datacamp.com/courses/data-visualization-in-r
- https://www.datacamp.com/tracks/importing-cleaning-data-with-r
- https://www.datacamp.com/courses/intro-to-text-mining-bag-of-words
- http://www.mit.edu/~rnielsen/Working%20with%20Unicode%20in%20R.txt
- https://plot.ly/r/reference/
- https://unicode.org/standard/standard.html
- https://cran.r-project.org/web/packages/readtext/vignettes/readtext_vignette.html
- https://kevinushey.github.io/blog/2018/02/21/string-encoding-and-r/
- https://db.rstudio.com/databases/sqlite/
- https://www.rdocumentation.org/packages/xfun/versions/0.5/topics/read_utf8