

CLEANJO

PROBLEM STATEMENT

- **Unstructured Data**
- **Data Stored**
- **Uncleaned Data**
- **Missing Data Values**
- **Errors in Data**

EXISTING SOLUTION

Few software/websites exist that are used for inputting the data or arranging the data, google sheets, microsoft excel and tableau like applications are used nowadays.

The features they provide-

1. Can input the data or edit the data
2. You can use formulas for the calculations that is time consuming
3. They can present the data into visual format like pie charts and graph
4. It provide a feature of highlighting the data that needed to be separated

LIMITATIONS OF EXISTING PROJECTS

- As there are multiple applications which provide us the facility to enter the data or create a database but they don't provide us the facility to calculate all the data in accordance with the statistics or graphical representation of the data that will help us to understand the growth or loss of the company in a better way .
- This kind of software doesn't Monitor Data Entries.
- These are not user friendly as it can't be reliable for the large database.
- Mostly the google sheet or app like that is not helpful to get the correct data as if we have to calculate every data of a single field differently then you have to calculate the data manually that will be a difficult task to do .
- This kind of application will not work if we have to merge the 2 databases and sort them in accordance to a particular field

PROPOSED SOLUTIONS

The software will be reliable for the large amount of the data.

- This software can be accessed in offline mode.
- The data cleaning function will be added.
- The charts and graph will be linked with the data.
- This will be doing the data merging .
- This software will include the future predictions.
- *Data Entry Monitoring* feature will be added.
- Data Filtering will be added.

ADVANTAGES OF PROPOSED SOLUTION

The software will be reliable for the large amount of the data :

- The softwares like excel or google sheet is not reliable for the large of the data as companies most often doesn't prefer to use google sheets for the matter of handling the data in excel as well the data can't be handled of the large amount so that feature will be provided by our software that the large data could be handled in the way.

This software can be accessed in offline mode :

- Google sheets like applications work in online mode ,users cannot update ,create or view the spreadsheet in offline mode. That is a big issue if the internet connection is weak some day or any other issue related to that then that will be difficult for the person to work so this feature will be included so that it will work in offline mode .

The data cleaning function will be added:

- As the data entered in the database that are rough need to be adjusted by the analyst that include removing the unwanted or repetitive entries in the database or cleaning the NULL or zero entries and sorting the data as we want.

The charts and graph will be linked with the data:

- The charts or any graph that are represented of any particular field accordance the final data that is cleaning will not be updated when the entry of the database changes that will be creating a mess out of it as will be having multiple spreadsheet each having different data so this software will link the data to the graph so that any change occur in the spreadsheet will automatically update the graph or charts.

This will be doing the data integration:

- The term data integration clearly means merging of the data or application will be using features of the data integration so that the merging of any 2 databases will be easy or that can be separated as per the need of the user.

This will be calculating the values automatically with certain operations:

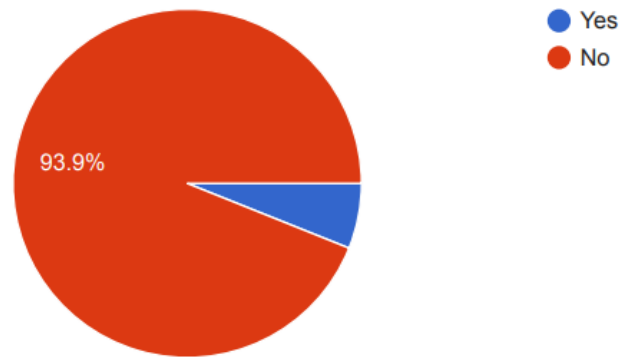
- As in excel we need formulas or query to calculate any kind of data as per the need but these formula or query doesn't work all the time or these formulas or it is difficult to learn these kind of the queries but **our application will** be including **a feature of calculation in other ways**

Google Form Responses

Are you a Data Analyst?

 Copy

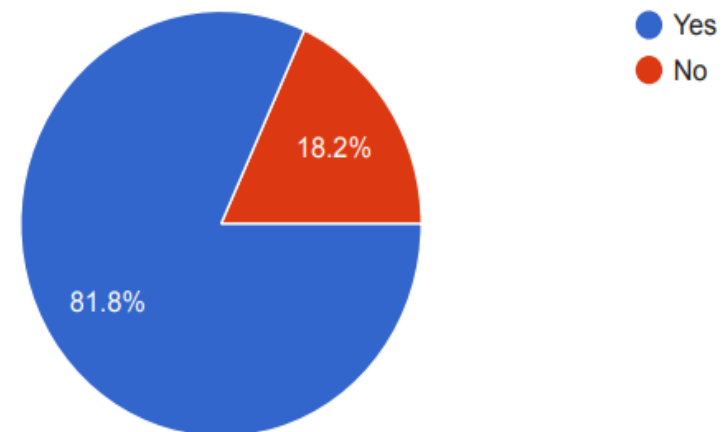
33 responses



Would there be theme section for interface?

 Copy

33 responses

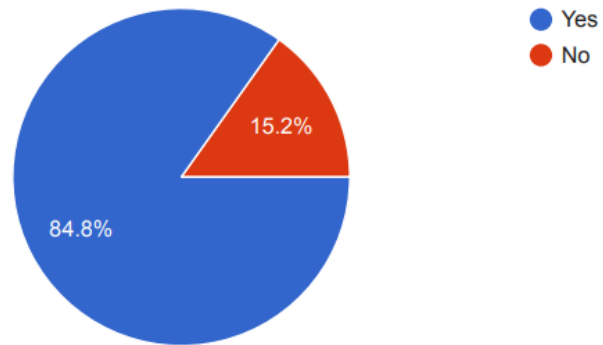


Should Data Monitoring be added in the Software?

 Copy

Data Monitoring :- All activities like modification in data are recorded which data is changed,when changed and who changed the data.

33 responses

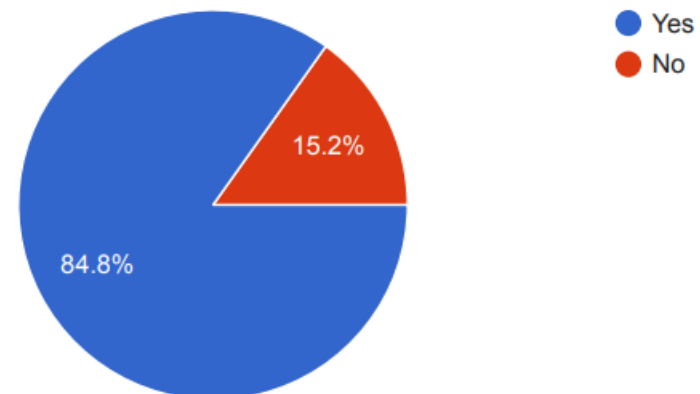


Should efficiency be calculated?

 Copy

Calculating efficiency of employees according to there performance.

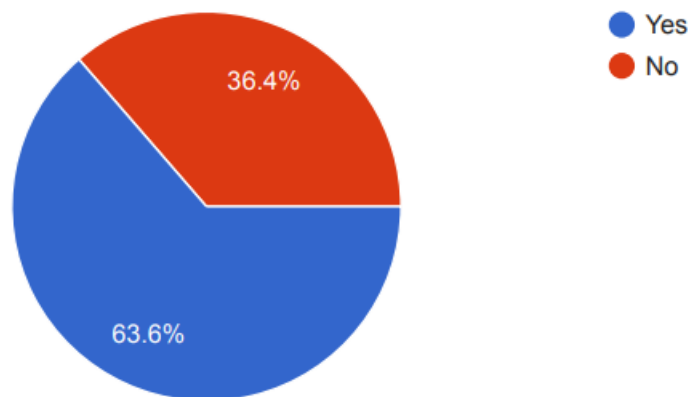
33 responses



Should there be the feature of Data Merging?

 Copy

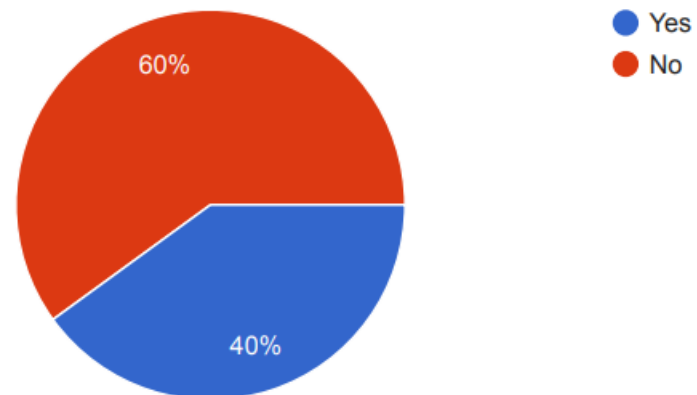
33 responses



Would you want Data mining feature to be added?

 Copy

5 responses

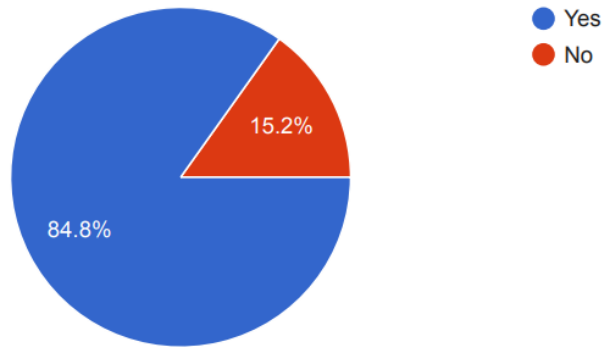


Should Data Monitoring be added in the Software?



Data Monitoring :- All activities like modification in data are recorded which data is changed,when changed and who changed the data.

33 responses

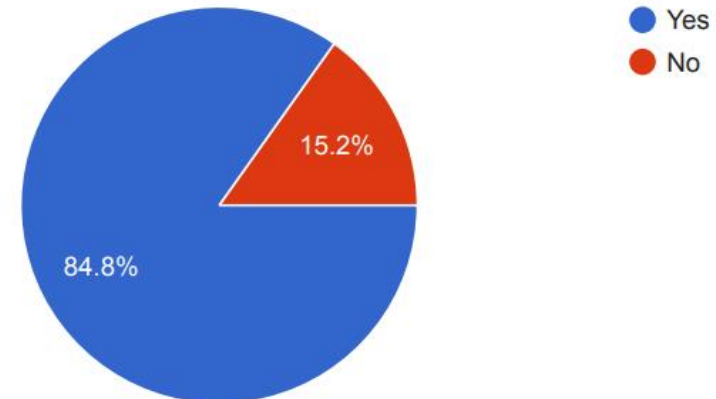


Should efficiency be calculated?



Calculating efficiency of employees according to there performance.

33 responses

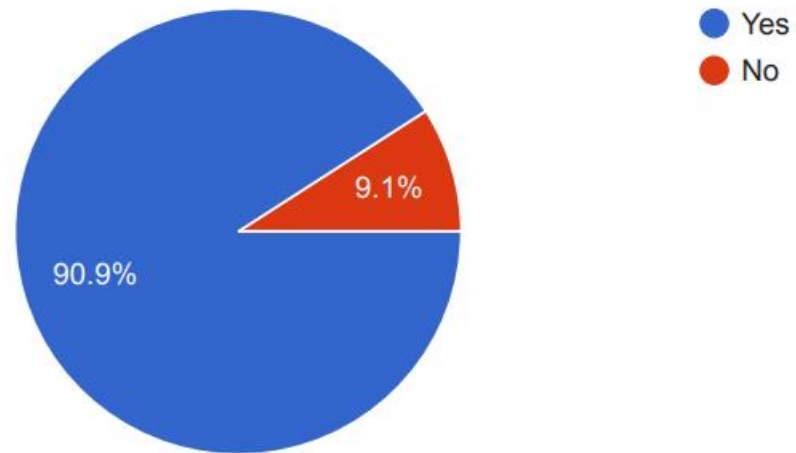


•

Would there be Graphical Representations of Data like pie chart, histogram, Box Plot etc.?

 Copy

33 responses



SOFTWARE REQUIREMENT SPECIFICATION

Hardware Requirements

- At least 4GB of RAM

The operating system takes almost 3GB of the RAM itself, so we need at least 1 GB of extra RAM to run our software.

- 4GB of GPU

GPU memory will be required in graphics-related work like graphical representation

- The processor of 64 bit

This enables the processor to handle far more data and instructions per unit of time so our software will be faster.

Software Requirements

- **HTML/CSS**

This language is used for describing the structure of software components and designing our UI.

- **JAVASCRIPT**

This language will be used to add interactive elements that engage users.

- **DBMS**

This will help to manage the data as it allows users to create, read, update, and delete data in the database.

Functional Requirement

Signup Section

To create a new user account

Login

To access the software

Logout

To give access to a new user

Structured data

Only structured data will be allowed to be input so that software can read it perfectly.

Data Cleaning

In Data Cleaning we will be removing NULL entries, and filling empty entries by predicting the value with the help of the relation between attributes. Mean, Median, Mode, and some Algorithms will help in predicting the accurate values.

Data Filtering

You can extract the particular data in the database using the filter

Data Merging

Multiple Databases can be merged into the single one only if the columns in both the databases are common. So we will provide an interface before merging to select the common columns to merge the data.

Statistics Calculation

After the cleaning of the data statistical values of the data will be calculated for the graphical representation i.e Mean, Median, and Mode.

Graphical Representation

Presenting the data in a diagrammatic form such as Pie charts, histograms, Line charts, etc.

Non- functional requirements

Performance

The application should be able to handle a certain number of concurrent users and processes and a certain amount of data

Security

- The application will have 2 level of verification for the protection of the data.
- Mobile Number Verification by sending OTP via SMS
- Email Verification by sending OTP to the registered email

Usability

The application will have simple features and a user-friendly interface that can be used easily

User Requirements

User ID / Password

Data Cleaning

The application will clean the provided data.

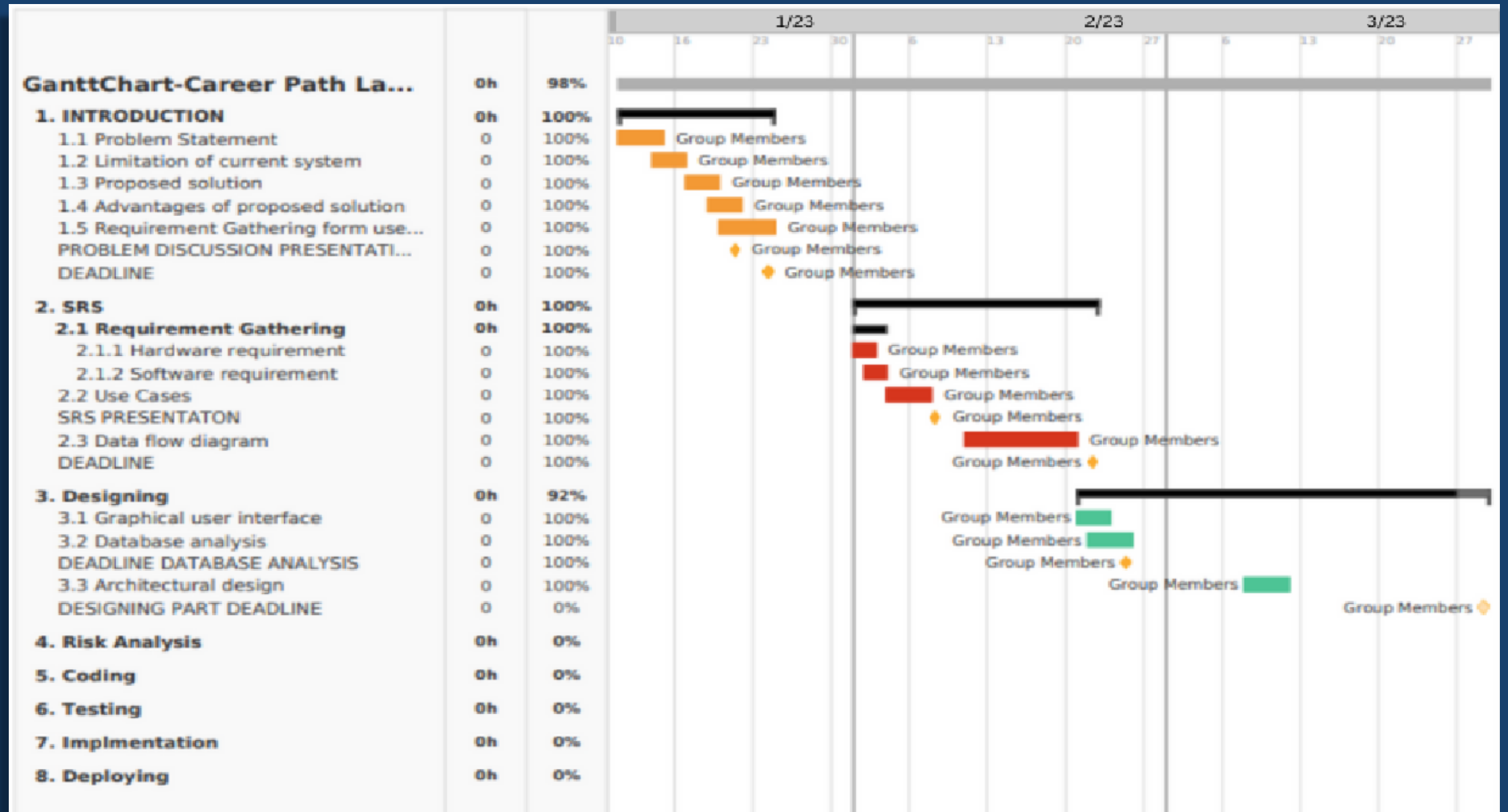
Statistics Calculation

The statistical data will be calculated after the cleaning

Graphical Representation

The database will be presented in the form of graphs for an easy and better understanding of the data

Gantt chart



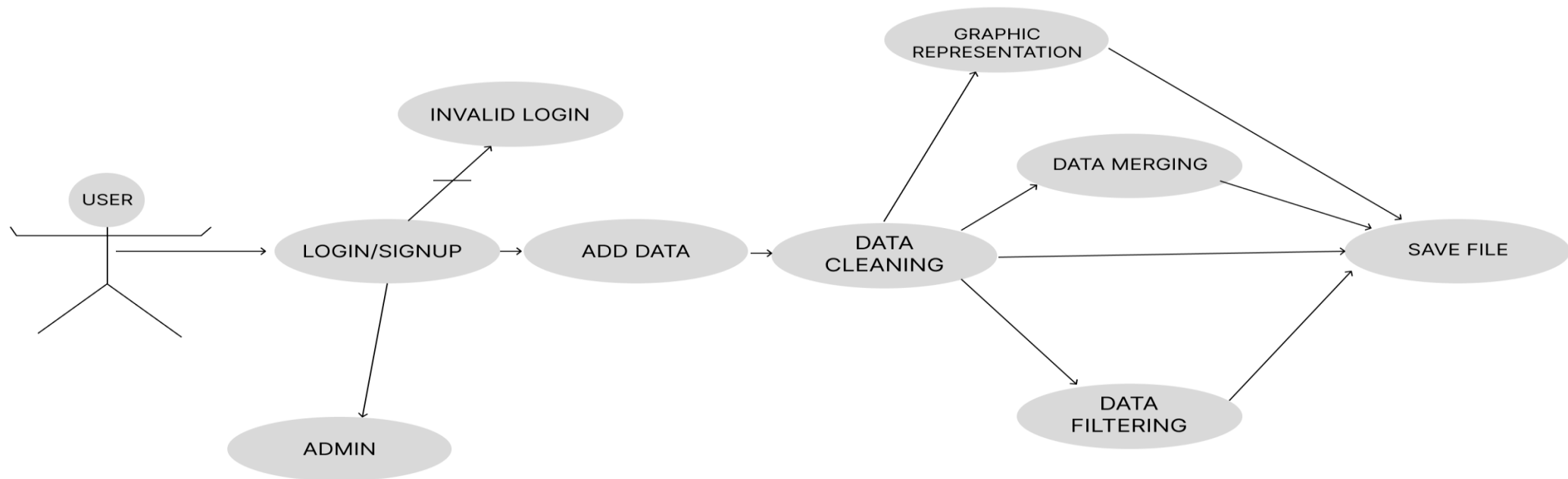
Functional Point Estimation

INFORMATION ON DOMAIN VALUE	COUNT	LOW	AVG	HIGH	COUNT WEIGHT
External input	5	3	4	6	20
External output	4	4	5	7	20
External Interface Files	0	3	4	6	0
Internal Logical Files	3	7	10	15	45
External Inquiry	10	5	7	10	70
Total					155

$FP = \text{count total} * [0.65 + 0.01 * fI]$ $FP = 155 * [0.65 + 0.01 * 51]$ $FP = 155 * 1.16$ $FP = 179.8$

Use Cases

Use Case Diagram



SIGNUP

Brief Discription

It allows users to access the software by signing up with double verification of email and mobile no.

Primary Actor

User

Basic flow

Signup Page ->Name -> Enter Email id -> Enter mobile no.->Date of birth-> Create password -> Confirm password->verify mobile no. with otp -> verify email with otp->User logged in Successfully

Alternative flow

a) Signup Page ->Name-> Enter Email id ->Invalid Email ->User not registered

b) Signup Page ->Name-> Enter Email id ->Email already exist ->User not registered

c) Signup Page ->Name-> Enter Email id ->Enter Mobile no.->Mobile no. already exist ->User not registered

d) Signup Page ->Name -> Enter Email id -> Enter mobile no.->Date of birth-> Create password -> Confirm password->verify mobile no. with otp -> wrong otp entered -> re-enter otp

e) Signup Page ->Name -> Enter Email id -> Enter mobile no.->Date of birth-> Create password -> Confirm password -> verify mobile no. with otp -> mobile number verified-> verify email with otp-> wrong otp entered -> re-enter otp

LOGIN SECTION

Brief Description

Login using the registered email and password to use the software.

Primary Actor

User

Basic Flow

Login Page -> Enter email-> Enter password -> Logged in Successfully

Alternative Flow

Login Page -> Enter email -> Email does not Exist -> Login not Successful

Login Page -> Enter email -> Enter password -> Forget password -> Login not Successful

Login Page -> Enter email -> Forgot password -> Enter mobile no.-> otp -> reset password -> Relogin

STRUCTURED DATA

Brief description

The data which is organised and formatted in a specific way to make it easily readable or understandable

Primary Actor

User

Basic Flow

a) Home page -> add data -> Data added Successfully

Alternative Flow

Home page -> add data -> Data entered is not structured

DATA CLEANING

Brief description

Data cleaning is the process of fixing or removing NULL values, incorrectly formatted, duplicate, or incomplete data within a dataset.

Primary Actor

User

Basic Flow

Home page -> Add data -> Check the data in the attributes which have null entered is active more than 70% -> Data cleaned

Alternative Flow

Home page -> Add data -> Data is not in structured format

DATA FILTERING

Brief description

Data filtering is the process of examining a dataset to exclude, rearrange, sort, or apportion data according to certain criteria.

Primary Actor

User

Basic Flow

Home page -> Add data -> Data cleaned -> Add filter to extract the data

Alternative Flow

Home page -> Add data -> Data cleaning -> Add filter to extract the data -> Invalid filter applied

DATA MERGING

Brief description

Data merging is the process of combining two or more similar records into a single one.

Primary Actor

User

Basic Flow

Home page-> Add data -> Data cleaned->Find the common attribute -> Merge dataset

Alternative Flow

a) Home page->Add data->Data cleaned-> Find the common attribute -> No common attribute found

•

STATISTICS CALCULATION

Brief description

Statistical methods are developed to analyse the large volumes of data and their properties.

Primary Actor

User

Basic Flow

Home page->Add data -> Data cleaned -> Statistics Calculation

Alternative Flow

Homepage -> Add data -> No Statistical field available

GRAPHICAL REPRESENTATION

Brief description

Graphical Representation is a way of analysing numerical data. It is easy to understand and it is one of the most important learning strategies.

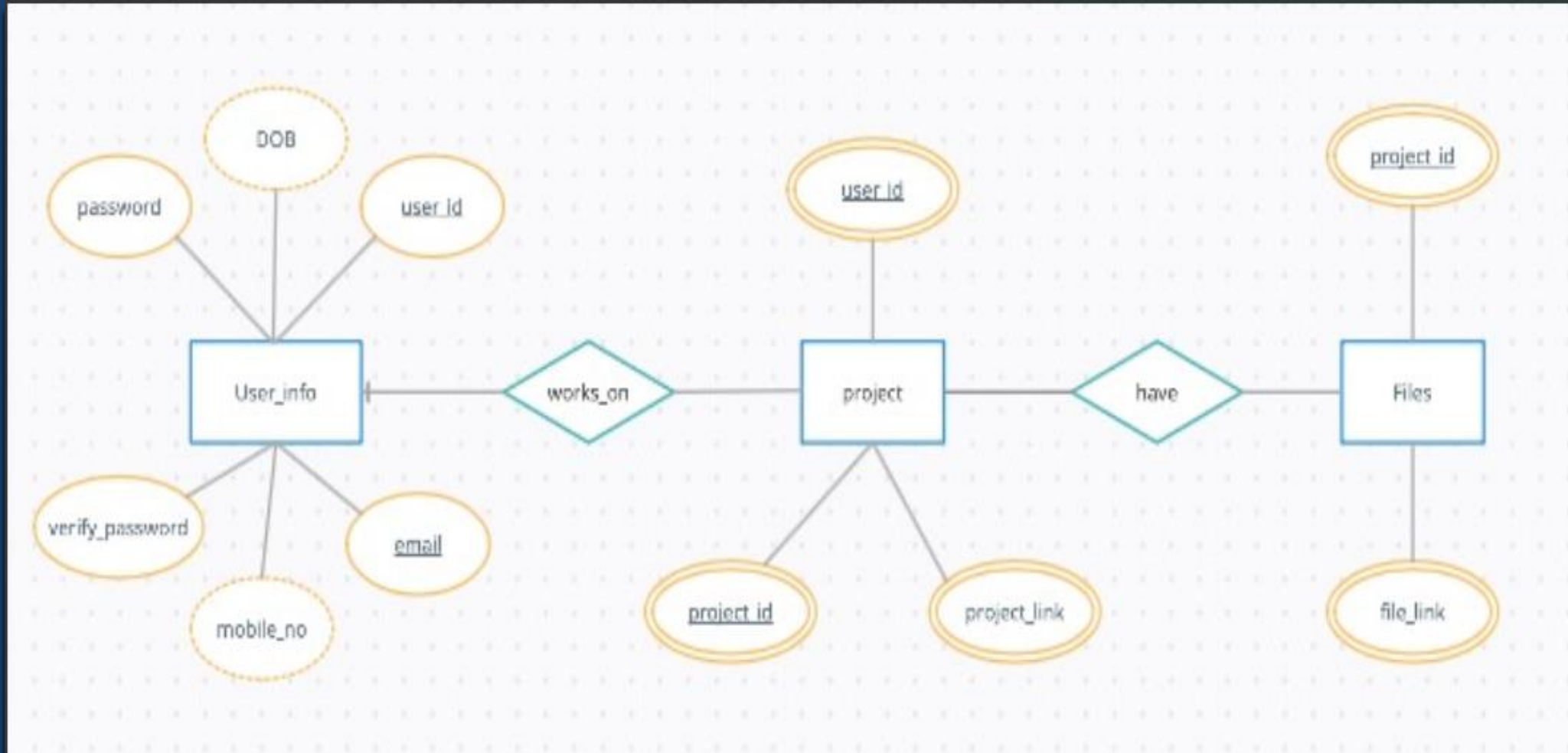
Primary Actor

User

Basic Flow

Home page->Add data ->Data Cleaned-> Statistics data calculated-> Graphical representation

ER DIAGRAM



DATABASE

Field	Type	Null	Key	Default	Extra
Name	char(50)	NO		NULL	
Email	varchar(50)	NO	PRI	NULL	
Mobile_no	int	NO		NULL	
DOB	date	NO		NULL	
Password	varchar(50)	NO		NULL	
Verify_Password	varchar(50)	NO		NULL	
User_ID	varchar(10)	NO	MUL	NULL	
Project_no	int	YES		NULL	

Field	Type	Null	Key	Default	Extra
user_id	varchar(10)	NO	PRI	NULL	
Project_name	varchar(50)	YES		NULL	
Project_id	varchar(10)	YES	MUL	NULL	
Project_link	varchar(50)	YES		NULL	

Field	Type	Null	Key	Default	Extra
User_id	varchar(10)	NO	PRI	NULL	
File_link	varchar(50)	YES		NULL	

ARCHITECTURAL DESIGN

SIGNUP

Name

Email Address

Mobile Number

Date of Birth

Password

Verify Password

SIGNUP

WELCOME

LOGIN

User I'd

Password

LOGIN

[Reset Password](#)

YOU DON'T HAVE ANY
ACCOUNT!!

SIGNUP

RESET

Sent OTP

Enter OTP

Verify

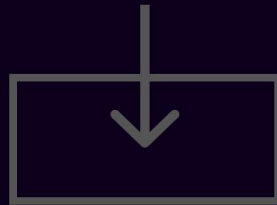
New Password

Verify Password

Change

User's Name

UID XXXXX



UPLOAD

Upload your .xlsx file

LOG OUT

	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						

User's Name

UIDXXXXX



CLEAN DATA

MERGE ANOTHER DATA

FILTERS

GRAPHIC REPRESENTATION

EXPORT FILE



LOG OUT


```
    return "Duplicate values removed";  
}  
}
```

```
on click("Extract file")  
{  
    return ;  
}
```

```
    on click("Calculate Statistics")  
    {  
if(data>70%)  
{  
    return Statistics  
}
```

```
else  
{  
    return "The amount of data is not appropriate for satastical calculation\nFirst fill the values then calculate  
statistics "  
}  
}
```


	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						

User's Name

UID:XXXXXX



FILTERS

SORT- HIGH TO LOW

SORT- LOW TO HIGH

SORT- ALPHABATICAL

SELECT ATTRIBUTE

FILTER



LOG OUT

SELECT COMMON ATTRIBUTES

Attribute Name file2

Attributes of uploaded
file

A

D

F

K

L

O

more...

MERGE

User's Name

UIDXXXXX



CLEAN DATA

MERGE ANOTHER DATA

FILTERS

GRAPHIC REPRESENTATION

EXPORT FILE



LOG OUT


```
    }  
    else  
    {  
  
        select(attribute you want to merge)  
        {  
            if(on_check(Attributes data matched))  
            {  
                return Merged files  
            }  
            else  
            {  
                return "Files are not appropriate to merge "  
            }  
        }  
    }  
    on click("Done")  
    {  
        goto(export file page);  
    }
```

db. insert(Project, file)

database table used are:- Project and file

User's Name
UIDXXXXX



EXPORT FILE

SELECT GRAPH TYPE

BAR GRAPH

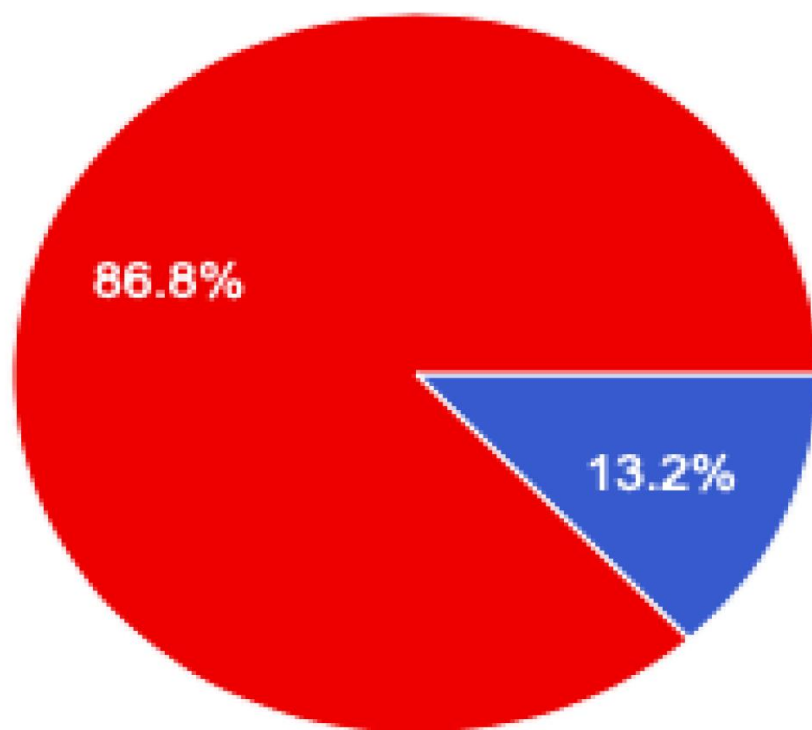
PIE CHART

HISTOGRAM

LINE GRAPH



LOG OUT



User's Name

UIDXXXXX



EXPORT FILE



LOG OUT

TEST CASES

Module: Signup

TEST CASES(INPUT)

Test case 1: Email should contain @ and .

Test case 2: Mobile no. should contain 10 digit

Test case 3: incorrect format of DOB

Test case 4: Password should contain upper case alphabet, numbers, and signs

Test case 5: Verify password should be same as password

Test case 6: Incorrect field entry

Test case 7: Empty field

Test case 8: Click on login page

DESIRED OUTPUT

1. Valid email
2. Valid Mobile Number
3. Invalid DOB
4. Invalid Password
5. Password not same
6. Incorrect field entries
7. Required Field
8. Go to login page

Module: Login

TEST CASES(INPUT)

Test case 1: correct username and correct password

Test case 2: Incorrect username and password

Test case 3: Empty field

Test case 4: click on forget password

Test case 5: click on sign up page

DESIRED OUTPUT

1. Login successfully
2. Invalid entries
3. Please fill in the required fields
4. Go to the verification page
5. Go to the signup page

Module: Verification Page

TEST CASES (INPUT)

Test case 1: Incorrect Mobile OTP

Test case 2: Incorrect email OTP

Test case 3: Correct Mobile & email OTP

DESIRED OUTPUT

1.

The mobile otp you have entered is wrong

2.The Email otp you have entered is wrong

3.Successfully entered

Module : Home Page

TEST CASES(Input)

Test case 1: Structured file inputted

Test case 2: Unstructured file inputted

Test case 3: clicked on logout

Test case 4: clicked on Profile

DESIRED OUTPUT

1. Successfully uploaded file
2. Alerted to upload a structured file only
3. Log out of the user and go to the login page
4. Go to the profile window Module

Module : Data Cleaning

TEST CASES(Input)

Test case 1: Clicked on Clean data

Test case 2: Clicked on merge another data

Test case 3: clicked on data filter

Test case 4: clicked on export file

Test case 5: clicked on profile

DESIRED OUTPUT

1. Data cleaned
2. Merge data window Module
3. Filter data window Module
4. Export the file
5. Go to the profile window Module

Module : Data Merging

TEST CASES(Input)

Test case 1:Clicked on merge another data

Test case 2: clicked on import file

Test case 3: clicked on select the same attribute

Test case 4: clicked on profile

Test case 5: clicked on export data

DESIRED OUTPUT

1. Import file or download file window open
2. The file will be imported from device
3. The same attribute will be selected by the user
4. Go to the profile window Module
5. Export the file

Module : Data Filtering

TEST CASES(Input)

Test case 1:Clicked on apply filters

Test case 2: clicked on sort file

Test case 3: clicked on sorting value

Test case 4: clicked on extract data.

Test case 5: clicked on profile

Test case 6: clicked on export data

DESIRED OUTPUT

1. Choose the option for the filter
2. Ask the user to enter the value for the filtering purpose
3. The file will sorted
4. Ask the user to end the upper or lower value or field for extracting the data
- 5.GO to the profile window
6. Export the file

Module : Graphical Representation

TEST CASES(Input)

Test case 1:Clicked on graphical presentation

Test case 2: Click on the type of graph you want to build

Test case 3: clicked on profile

Test case 4: clicked on export data

DESIRED OUTPUT

1. Ask the user to enter the value for the graphical representation
2. The graph will be displayed
3. Go to the profile window Module
4. Export the file