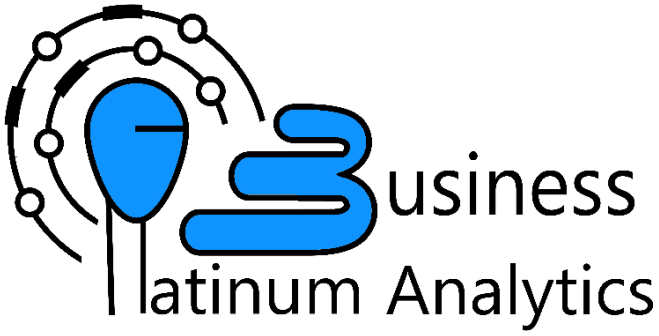
Department of Computer Science

University of Gujrat

PLATINUM BUSINESS ANALYTICS



Session: BSCS Spring 2016-2020

Project Advisor:

Mr. Waqas Haider Khan Bangyal

Submitted By:

Zeeshan Ahmad	***16201519-019***
Muhammad Zaryab	***16201519-153***
Muhammad Shaheryar Khan	***16201519-161***

Department of Computer Science

University of Gujrat

STATEMENT OF SUBMISSION

This is certify that **Zeeshan Ahmad**, **Muhammad Zaryab** and **Muhammad Shaheryar Khan**, Roll No. **16201519-019**, **16201519-153** and **16201519-161** has successfully completed the final year project named as **"Platinum Business Analytics"** at the Department of Computer Science, University of Gujrat, to fulfill the requirement of the degree of **BS in Computer Science**.

Project Management Office
Department of Computer Science,
Faculty of Computer science & IT,
University of Gujrat, Punjab,
Pakistan.

Project Supervisor
Department of Computer Science,
Faculty of Computer Science & IT,
University of Gujrat, Punjab,
Pakistan.

Chairperson
Department of Computer Science,
Faculty of Computer Science & IT,
University of Gujrat, Punjab,
Pakistan.

ACKNOWLEDGEMENT

We truly acknowledge the cooperation and help make by **Mr. Waqas Haider Khan Bangyal** supervisor of our project. He has been a constant source of guidance throughout the course of this project. We would also like to thank **Mr. Muhammad Sami Ullah** our PMO for his help and guidance throughout this project. We are also thankful to our friends and families whose silent support led us to complete our project.

Zeeshan Ahmad: _____

Muhammad Zaryab: _____

Muhammad Shaheryar Khan: _____

Date: August 04, 2020

Abstract

"Platinum Business Analytics" is based on a web application and sentiment analysis of a movie or product. Sentiment analysis is a process of analyzing the text and more specifically reviews of products or about some kind of topic. As businesses spend a lot of money on analyzing their consumers. They monitor the response or reviews of the consumers about a product to make future business decisions. So they require a user friendly visualizations and categorized reviews to understand the consumer response. Our project will provide the interface where user can register and get the most relevant results. The main goal of this project is to improve the accuracy of the sentiment analysis with a graded scale. The scale divides sentiments into five categories: **Negative**, **Partially Negative**, **Neutral**, **Partially Positive** and **Positive**. We will use a data-set from **"Stanford Sentiment Tree Bank".**

TABLE OF CONTENTS

Chapter 1

Project Feasibility Report

1.1. Introduction

The proposed System is based on a web application and sentiment analysis of a movie or product. Sentiment analysis is a process of analyzing the text and more specifically reviews of products or about some kind of topic. Currently, sentiment analysis is a topic of great interest and development since it has many practical applications. Since publicly and privately available information over Internet is constantly growing, a large number of texts expressing opinions are available in review sites, forums, blogs, and social media.
With the help of sentiment analysis systems, this unstructured information could be automatically transformed into structured data of public opinions about products, services, brands, politics, or any topic that people can express opinions about. This data can be very useful for commercial applications like marketing analysis, public relations, product reviews, net promoter scoring, product feedback, and customer service. We will use a data-set from "Stanford Sentiment Tree Bank".

1.2. Project Feasibility Report

Our project is focused on the sentiment analysis of a topic or product. User can select a specific category (like negative, partially negative, neutral, partially positive and positive) and get the most relevant results. Businesses spend a lot of money on analyzing their consumers. Everyone wants to win the race. They monitor the reviews or response of the consumers about a product to make future business decisions. They analyze the reasons why a certain product is getting popularity or disliked so that they can make decisions about those products. They require a user-friendly visualizations and categorized reviews to understand the consumer response.

The goal of this project is to provide sentiment analysis and data analytics of a certain product or a topic. The goal is to provide user with a user-friendly interface and the more accurate results than ever before using state of the art AI techniques. The proposed help users to better understand their business. Our project is feasible according to the requirements of users. In order to show the feasibility of software development system of our project we will discuss all the feasibilities one by one.

1.2.1. Technical Feasibility

Our purposed system perform sentiment analysis that will run on web application. The technologies that we are using are all feasible for our project to perform analysis on the on the reviews using state-of-the-art techniques. The size of our project is large and is useful for all users. We are using Mongo DB database. Since we have thousands of reviews of a movie or a product so Mongo DB is best approach to hold data.

For gathering reviews from the sites like (Rotten Tomatoes, IMDB), we are using Scrapy. This will provide us data (reviews) in the form of JSON. For web application we are using Python, CSS, and Java script, Flask and these are all feasible technologies to provide required results. It is possible to develop and implement the proposed system with all Functionality. All the group members possess the required technical expertise.

1.2.2. Operational Feasibility

Our main operation to create a model that can accurately classify sentiments into five categories. All the modules of our project related to designing and development are feasible to perform this task. Platinum Business Analytics is more efficient and helpful as compare to other analyzers. It satisfies the requirements identified in the requirements analysis phase of system development.

1.2.3. Economic Feasibility

Total cost estimated for this project is 3000$. So, it can be an economically feasible project.

1.2.4. Schedule Feasibility

The **"Platinum Business Analytics"** project can be completed within its scheduled time limits, by a planned due date. Its schedule feasibility will be appraised as high, as this project goes along the planned scheduled according to Gantt chart.

S#	Task Name	Duration	Start Date	End Date
1	Feasibility	10 days	3 September, 2019	16 September, 2019
2	Software Requirement	14 days	17 September,2019	05 October, 2019
3	Web Scraping	15 days	06 October, 2019	26 October, 2019
4	Machine Learning Techniques	18 days	27 October, 2019	20 November, 2019
5	Prototype	17 days	21 November, 2019	15 December, 2019
6	Deep Learning Techniques	23 days	16 December, 2019	15 January, 2020
7	Ensemble Learning Techniques	17 days	16 January, 2020	08 February, 2020
8	Data Analytics	14 days	09 February, 2020	27 February, 2020
9	Web Interface	19 days	28 February, 2020	25 March, 2020
10	Integrating Modules	14 days	26 March, 2020	14 April, 2020
11	Testing	23 days	15 April, 2020	15 May, 2020

Table 1.1: Schedule Feasibility

1.2.5. Specification Feasibility

This system contains the functional requirements as well as non-functional requirements and these are clear and defined. It is assumed that the user has a working internet connection and familiar with Internet and World Wide Web navigational tools.

1.2.6. Information Feasibility

All the information regarding the project will be on hand and made assessed according to the planned activities. The output of each step is tested according to the expected result to ensure its reliability.

1.2.7. Motivational Feasibility

Platinum Business Analytics provides the best sentiment analyses of reviews of a movie or product .User can select a specific category and get the most relevant results. User will be provided with the interface where he can register. User can create a project and can visit the project any time by logging into his account.
User will be able to see the graphs and reviews. There will be a section for analytics that will provide the visualizations like bar graphs and pie charts to better understand the data.

1.2.8. Legal & Ethical Feasibility

Project undertaken meet all legal and ethical requirements. We are developing the application that is both professional and ethical. All the terms and condition are applied. All the data will be kept confidential.

1.3. Project Scope

The project scope is focused on the sentiment analysis of a topic or product. The main goal of this project is to improve the accuracy of the sentiment analysis with a graded scale. The scale divides sentiments into five categories.

- Negative
- Partially negative
- Neutral
- Partially positive
- Positive

Our aim is to provide user a sentiment analysis of a product or topic. We provide user interface where he can register the account and create a projects. Any user can visit our system and check the sentiment analysis on a topic or product. Also user can register for new account for the future use where he can make a project for the topic or product. User will be able to see the graphs and reviews and can visit to that website to which that review belongs. There will be a section for analytics that will provide the visualizations like bar graphs and pie charts to better understand the data. This will ultimately lead to authentic analysis of reviews related to a movie, product or topic. User can visit the dashboard and check any of his projects any time. User can filter the sentiments according to scale (positive, partially negative etc.). User can see the visualizations in the analytics tab and can also export results to a pdf to use later.

1.4. Project Costing

Our Project is between Organic and embedded types so we use Semi-detached type to measure cost.

1.4.1. Project Cost Estimation by Function Point Analysis

Domain Value	External	Complex (Weighting factor)	Total= E*C
External input	20	6	120
External output	18	7	126
External inquiry	10	6	60
Internal logical files	3	15	45
External interface files	10	10	100

Table 1.2: Function Point Analysis table

Total = 451
Value adjustment factor = 30
FP est. = 451 * [0.65 + (0.01 * 30)] = 428.45
Cost / FP = 20,000/428.45= 46.67
Total Project Cost = 451* 46.67 = 21,052.32
Total Estimated Effort = 451/428.45=1.05

1.4.2. Project Cost Estimation by Using COCOMO'81

Our Project is between Organic and embedded types so we use **Semi-detached type to measure cost.**

Basic COCOMO

Type	**Effort**	**Schedule**
Semi-Detached	**PM**= 3(6)1.12=20.16	**TD**= 2.5(20.16)0.35=7.64

PM = person-month (effort) KLOC= lines of code, in thousands
TD = number of months estimated for software development (duration)
People Required = 20.16/7.64 =3

1.5. Task Dependency Table:

Task	Description	Duration in weeks	Dependencies
T1	Detail system study	1 week	None
T2	Proposal making	3 weeks	T1
T3	Software requirement specifications	2 weeks	T1, T2
T4	Analysis	2 weeks	T1, T2
T5	Application interface designing	3 weeks	T1, T3
T6	Application development	8 weeks	T5

T7	Testing	2 weeks	T5, T6
T8	Improvements	2 weeks	T5, T6

Table 1.3: Task Dependency

1.6. CPM - Critical Path Method

Activity	Task	Immediate Predecessor	Duration
T1	Detail system study	None	1
T2	Proposal making	T1	3
T3	Software requirement specifications	T2	2
T4	Analysis	T2 , T3	2
T5	Application interface designing	T3, T4	3
T6	Application development	T3, T5	8
T7	Testing	T5,T6	2
T8	Improvement	T6, T7	2

Table 1.4: Sequence of Activities

Network Diagram:

A network diagram is a schematic depicting the nodes and connections amongst nodes in a network or, more generally, any telecommunications network. Once the activities and their sequencing have been defined, the CPM diagram can be drawn. CPM originally was developed as an activity on node (AON) network, but some project planners prefer to specify the activities on the arcs. Here is the network diagram for Platinum Business Analytics.

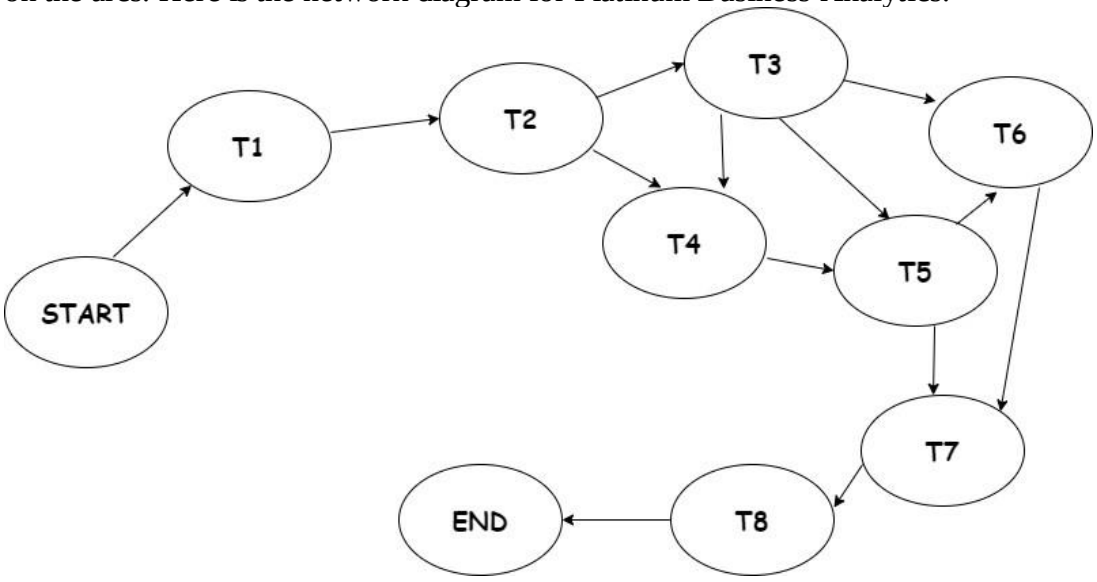


Figure 1.1: Network Diagram

Estimate Activity Completion Time:

Activity	Duration	ES	EF	LS	LF	TS	FS
T1	1	0	1	0	1	0	1
T2	3	1	4	1	2	2	0
T3	2	4	6	2	4	2	2
T4	2	4	6	4	6	0	6
T5	3	6	9	6	9	0	9
T6	8	9	17	9	17	0	17
T7	2	17	19	17	19	0	19
T8	2	19	21	19	21	0	21

Table 1.5: CPM

The critical path is:

T1, T4, T5, T6, T7,T8

1.7. Gantt Chart

This Gantt chart would identify major milestones with their achievement criteria.

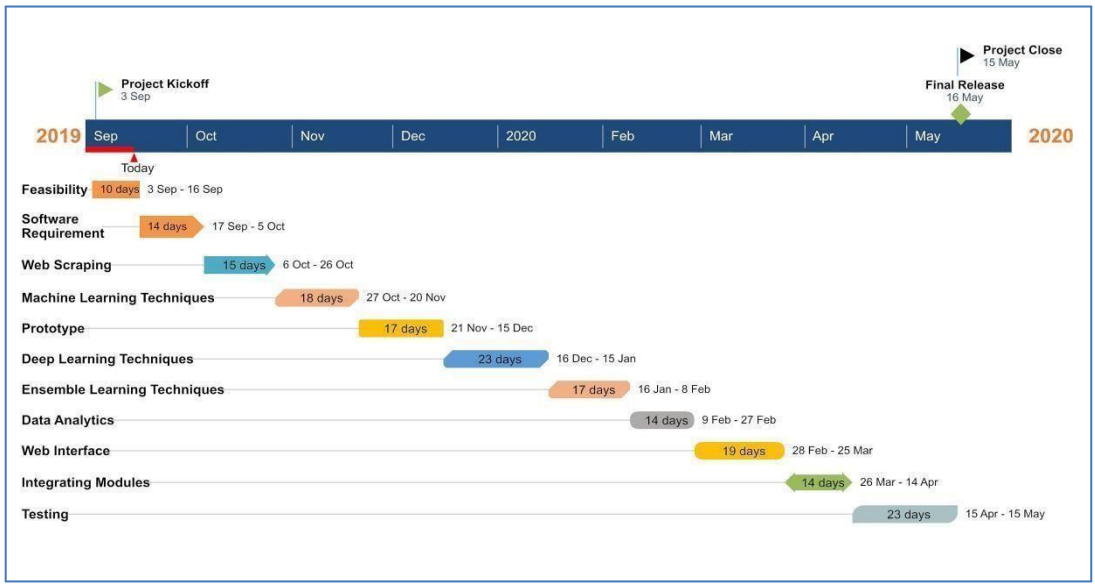


Figure 1.2: **Gantt chart**

1.8. Introduction to Team Member and Their Skill Set

Members Name	Skill Set
Zeeshan Ahmad	Python Expertise, Scraping
Muhammad Zaryab	Back-end development, Rest API Testing
Muhammad Shaheryar Khan	Documentation, Web pages, Testing

Table 1.6: Team members skill set

1.9. Task and Member Assignment Table

Task Durations and Dependencies

Task	Duration	Dependencies	Members
T1	1 week	-	M1
T2	3 weeks	T1	M1
T3	2 weeks	T2	M1,M2,M3
T4	2 weeks	T1, T2	M1,M2,M3
T5	3 weeks	T1, T4	M1
T6	8 weeks	T5	M2,M3
T7	2 weeks	T5 , T6	M1,M2,M3
T8	2 weeks	T6, T7	M1,M2,M3

Table 1.7: Task and Member Assignment

Allocation of People to Activities

Task	Member
T1	Zeeshan Ahmad
T2	Zeeshan Ahmad, Muhammad Zaryab, Shaheryar Khan
T3	Zeeshan Ahmad, Muhammad Zaryab, Shaheryar Khan
T4	Zeeshan Ahmad
T5	Muhammad Zaryab
T6	Zeeshan Ahmad, Muhammad Zaryab, Shaheryar Khan
T7	Zeeshan Ahmad, Muhammad Zaryab, Shaheryar Khan
T8	Zeeshan Ahmad

Table 1.8: Allocation of People to Activities

1.10. Tools and Technology with Reasoning

Tools	
Jupyter Notebook	For training & testing models
Pycharm	For back-end development in flask
Visual Studio Code	For simple HTML & CSS forms
MongoDB Compass Community	A GUI Mongo DB interface
Draw.IO	For creating simple activity and class diagrams
Adobe Illustrator	For designing diagrams for documentation
Technologies	
Scrapy	For scraping data from websites
JSON	For storing scraped data from websites
Mongo DB	For storing scraped and user data

Pandas & Numpy	For performing operations on data during training algorithms
NLTK	It is used for text preprocessing
Scikit-learn	For using machine learning models
Gensim	For creating word embeddings from text
Matplotlib	For visualization of data
Seaborn	For more advance visualization of data
Tensorflow	For creating neural network models
Keras	For creating intial neural network models
Flask	For back-end development
HTML & CSS	For creating simple web-pages and designing
Javascript	For interactive web pages

Table 1.9: Tools and Technology

1.11. Vision Document

For businesses who wish to analyze their reviews of product, the Platinum Business Analytics is a web application that will allow a businesses to monitor the reviews or response of the consumers about a products to make future business decisions. Not only products but also reviews of a movie or a personality will also be monitored and user can see the reviews related to a movie or personality as well. In this regard, a detail analyses tool is required where people can login and analyze the reviews of a product or movie. By using our tool any person can analyze comments. By using our tool any user can User can create a project and can visit the project any time by logging into his account.
Graph will also be available to the user that shows how many people like or dislike the product. Our vision is to make a web application to improve the process of analysis. This project will fruitful for celebrities, brands, movies who are very conscious to know about their image. This is also useful for politicians who are interested to know about public opinion and views about themselves. This will help companies to better understand their business.

1.12. Risk List

The possibility of suffering harm or loss in terms of danger is called risk. Regarding the importance of risks, a list is to be maintained.

- Scraping relevant data
- Noisy data
- Improper time management
- Deadline fulfilment
- Internet failure
- Accuracy of model

1.13. Product Features/ Product Decomposition

The functional requirements capture the intended behavior of the system. This behavior may be expressed as services, tasks or functions the system is required to perform.

- Accuracy of model
- User Login
- A classifier with maximum accuracy
- Functional website for end users
- Fast data processing
- Visualization of insight of data
- Easy to use
- Charming look
- Project will be completed and delivered on time.
- Project will meet the appropriate quality targets.
- Project will meet the agreed scope.
- Project deliverables will be fit for purpose.
- Project team satisfaction target will be achieved.

Chapter 2

System Requirement Specification

2.1. Introduction

The proposed System is a web application. Our project is focused on the sentiment analysis of a topic, movie or product. Sentiment analysis is the process of determining whether a particular piece of writing is positive, negative or neutral. User can select a specific category (like negative, partially negative, neutral, partially positive and positive) and get the most relevant results. As businesses spend a lot of money on analyzing their consumers. Everyone wants to win the race. They monitor the reviews or response of the consumers about a product to make future business decisions. They analyze the reasons why a certain product is getting popularity or disliked so that they can make decisions about those products. They require a user-friendly visualizations and categorized reviews to understand the consumer response.

The goal of this project is to provide sentiment analysis and data analytics of a certain product or a topic. The goal is to provide user with a user-friendly interface and the more accurate results than ever before using state of the art AI techniques. The proposed help users to better understand their business.

2.2. Existing Business

There are many existing websites for the purpose of classifying whether a particular piece of text is positive, negative or neutral. Some of them are:

1. Brand24
2. Parallel dots
3. Repustate

These are discussed here below.

- **Brand24:**

It is also an online monitoring tool which shows us the sentimental analysis into three basic categories positive, negative and neutral. Only you have to give it the keyword of your brand, company or specific keyword you want to know the analysis about. In this application you have the option to create a team accounts. Show line plots of reviews polarity against time. We can get email alerts and it provides facility to export result to pdf. We can create multiple projects.

- **Parallel dots:**

It is an online web based application which shows us the sentimental analysis into three basic categories positive, negative and neutral. It supports 14 languages. It does not have any special module.

- **Repustate:**

It provides fast, reliable & accurate sentiment analysis in 23 languages. It performs sentiment analysis and extract semantic insights from social media, news, surveys, blogs, forums or any of your company data. It have API'S in many languages. Show sentiments by gender and demographics.

2.2.1. Business Organizational Chart

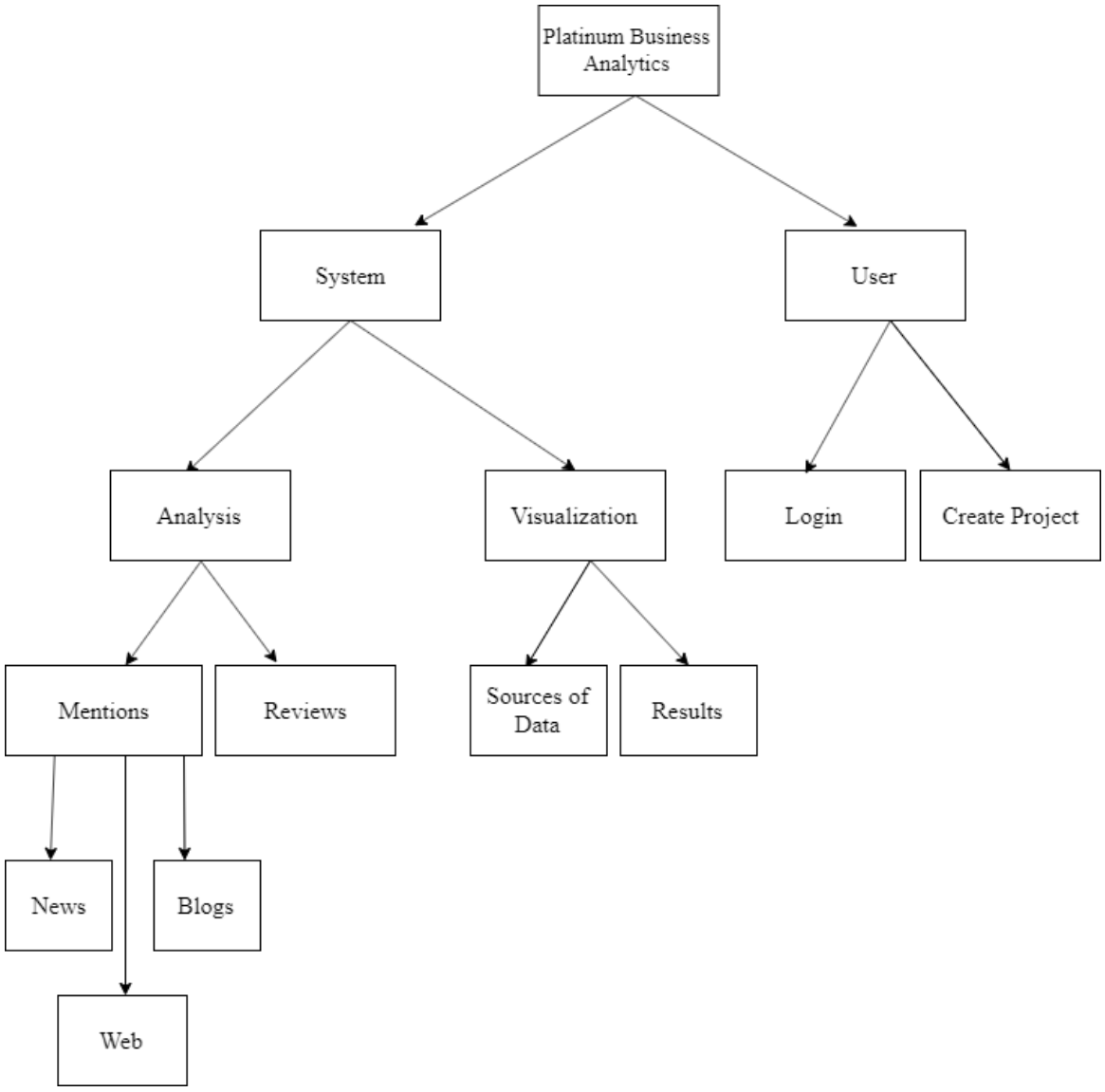


Figure 2.1: Organizational chart

2.3. Scope

This system will help to manage the reviews related to a topic or product. In this Platinum Business Analytics, user will be provided with the interface where he can register. User can create a project and can visit the project any time by logging into his account. User will be able to see the graphs and reviews and can visit to that website to which that review belongs. There will be a section for analytics that will provide the visualizations like bar graphs and pie charts to better understand the data. This will ultimately lead to authentic analysis of reviews related to a movie, product or topic. User can visit the dashboard and check any of his projects any time. User can filter the sentiments according to scale (positive, partially negative etc). User can see the visualizations in the analytics tab and can also export results to a pdf to use later.

2.4. Summary of the Requirements

Our project is mainly focused on:

- Perform sentiment analysis on user's provided text.
- Provide facility to perform sentiment analysis using best state of art algorithms.
- Perform sentiment analysis on user's reviews of a movie or product.
- Provide facility to create a project and visit the project any time by logging In to account.
- See the visualizations in the analytics tab and can also export results to a pdf to use later.
- Filter the sentiments according to scale (positive, partially positive, neutral, negative, partially negative etc).
- Provide facility to filter reviews of their product and see the source of reviews.

2.5. Identify External Entities

The identification of the external entities will be based on the information contained in your Abstract. This identification is done after two phases. We will map the "Platinum Business Analytics" project to make things more comprehensible.

The Identification of External Interfaces is done in two phases.

2.5.1. Over Specify Entities from Abstract

Based on abstract entities are:

- Login
- User
- Fine Grained Sentiment Analysis
- Visualization
- Google Search Results(IMDB, Rotten Tomatoes)
- Reports

2.5.2. Perform Refinement

After over specifying the entities, the following entities are more related to our project are:

- User
- Fine Grained Sentiment Analysis
- Visualization
- Reports

2.6. Capture ‘‘Shall’’ Requirements

Para #	Initial Requirements
1.0	The user shall be able to perform sentiment analysis on user given text.
1.0	The user shall register into our application.
1.0	The user shall login into our application.
1.0	The user shall create project of specific category (movie, product, general).
1.0	The user shall see the scraped data results.
2.0	The user shall see the sources of data scraped.
2.0	The user shall perform sentiment analysis on scraped results.
2.0	The user shall see visualization of scraped data results.
2.0	The user shall see the visualization of sentiment analysis results.
3.0	The user shall perform filter according to the results category.
3.0	The user shall perform filter according to the results fetched category.

Table 2.1: Initial Requirements

2.7. Allocate Requirements

Para#	Initial Requirement	Use-Case Name
1.0	The user shall be able to perform sentiment analysis on user given Text.	UC_sentiment_analysis
1.0	The user shall register into our application.	UC_register
1.0	The user shall login into our application.	UC_login
1.0	The user shall create project of specific category (movie, product, general).	UC_create_project
1.0	The user shall see the scraped data results.	UC_scraped_data
2.0	The user shall see the sources of data scraped.	UC_scraped_data_sources
2.0	The user shall perform sentiment analysis on scraped results.	UC_analysis_scraped_results
2.0	The user shall see visualization of scraped data results.	UC_visualization_of_scraped_results
2.0	The user shall see the visualization of sentiment analysis results.	UC_visualization_of_analysis_results
3.0	The user shall perform filter according to the results category.	UC_result_category_filter
3.0	The user shall perform filter according to the results fetched category.	UC_fetched_category_filter

Table 2.2: Allocate Requirements

2.8. Prioritize Requirements

	Rank	Initial Requirements	Use Case ID	Use Case Name
1.0	High	The user shall be able to perform sentiment analysis on user given Text.	UC_03	UC_sentiment_analysis
1.0	High	The user shall register into our application.	UC_01	UC_register
1.0	High	The user shall login into our application.	UC_02	UC_login
1.0	High	The user shall create project of specific category (movie, product, general)	UC_04	UC_create_project
1.0	High	The user shall see the scraped data results.	UC_05	UC_scraped_data
2.0	High	The user shall perform sentiment analysis on scraped results.	UC_06	UC_analysis_scraped_results
2.0	High	The user shall see the visualization of sentiment analysis results.	UC_08	UC_visualization_of_analysis_results
2.0	Medium	The user shall see the sources of data scraped.	UC_07	UC_scraped_data_sources
2.0	Medium	The user shall see visualization of scraped data results.	UC_09	UC_visualization_of_scraped_results
3.0	Medium	The user shall perform filter according to the results category.	UC_11	UC_result_category_filter
3.0	Medium	The user shall perform filter according to the results fetched category.	UC_10	UC_fetched_category_filter

Table 2.3: Prioritize Requirements

2.9. Requirement Traceability Matrix

	System Specification Text	Build	Use Case Name	Category
1.0	The user shall be able to perform sentiment analysis on user given Text.	B1	UC_sentiment_analysis	Business
1.0	The user shall register into our application.	B1	UC_register	Business
1.0	The user shall login into our application.	B1	UC_login	Business
1.0	The user shall create project of specific category (movie, product, general).	B1	UC_create_project	Business
1.0	The user shall see the scraped data results.	B1	UC_scraped_data	Business
2.0	The user shall see the sources of data scraped.	B1	UC_scraped_data_sources	Business
2.0	The user shall perform sentiment analysis on scraped results.	B1	UC_analysis_scraped_results	Business
2.0	The user shall see visualization of scraped data results.	B1	UC_visualization_of_scraped_results	Business
2.0	The user shall see the visualization of sentiment analysis results.	B1	UC_visualization_of_analysis_results	Business
3.0	The user shall perform filter according to the results category.	B1	UC_result_category_filter	Business
3.0	The user shall perform filter according to the results fetched category.	B1	UC_fetched_category_filter	Business

Table 2.4: Requirement Traceability Matrix Table

2.10. High Level Use Case Diagram

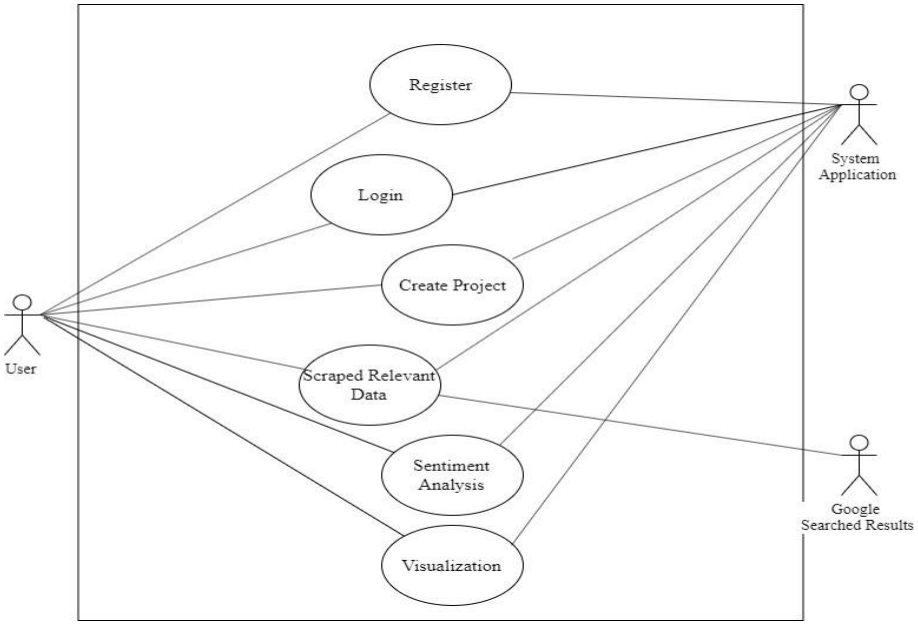


Figure 2.3: High Level Use Case Diagram

2.11. Analysis Level Use Case Diagram

Figure 2.4: Analysis level Use Case

2.12. Use Case Descriptions

UC_ register	
Use Case Name	UC_register
Use Case ID	UC_01
Actors	User
Description	This Use case is used by user to register to our web application.
Basic Flow	**User Action:** • User clicks on Sign Up. • User enters his valid email address. • User enters his password. • Clicks on Register. **System Response:** • Gets Registered successfully. • Redirected to create project.
Pre-Conditions	User must have a valid email address and password.
Post-Conditions	User should get successfully registered and should be redirected to create project.

UC_ login	
Use Case Name	UC_login
Use Case ID	UC_02
Actors	User
Description	This Use Case is used by user to login to our web application whenever he wants.
Basic Flow	**User Action:** • User enters his username. • User enters his password. • Clicks on Login. **System Response:** • Login successfully. • Redirected to create project.
Pre-Conditions	User must be registered.
Post-Conditions	User should be redirected to dashboard.

UC_ sentiment_analysis	
Use Case Name	UC_sentiment_analysis
Use Case ID	UC_03
Actors	User
Description	This Use Case is used by user to perform sentiment analysis on any given sentence.
Basic Flow	**User Action:** • User clicks on Demo. • Enters the text in input field. • Clicks on Analysis button. **System Response:** • Shows the sentiment result.
Pre-Conditions	User must have opened our application.
Post-Conditions	User should have shown sentiment result.

UC_create_project	
Use Case Name	UC_create_project
Use Case ID	UC_04
Actors	User
Description	The user shall create project of specific category (movie, product, general).
Basic Flow	**User Action:** • User clicks on Add New Project. • User selects the Project category (movie, product or general) from drop down list. • User then enters project name in another text field. • Clicks on Next. **System Response:** • Project created successfully. • Redirected to next page of project containing (Mentions & Reviews, Analysis) tabs.
Pre-Conditions	User must be registered and logged in successfully.
Post-Conditions	User should be redirected to next page showing created projects.

UC_scraped_data	
Use Case Name	UC_scraped_data
Use Case ID	UC_05
Actors	User
Description	This Use Case is used by user to see the scraped data results.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Mentions tab. **System Response:** • Shows scraped data results from(All, News , Forums, Blogs, Web).
Pre-Conditions	User must be logged in successfully and has created a project.
Post-Conditions	User should have shown scraped results.

UC_scraped_data_sources	
Use Case Name	UC_scraped_data_sources
Use Case ID	UC_06
Actors	User
Description	This Use Case is used by the user to see the sources of data scraped.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Sources tab. **System Response:** • Shows the names of sources i.e. websites and the no of mentions in that particular website related to our search.
Pre-Conditions	User must be logged in and has created project.
Post-Conditions	User should see sources of data scraped.

UC_analysis_scraped_results	
Use Case Name	UC_analysis_scraped_results
Use Case ID	UC_07
Actors	User
Description	This Use Case is used by user to be able to perform the sentiment analysis on the scraped results.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Analysis tab. **System Response:** • Shows the time series graph indicating one line for each output class i.e. 5 classes (positive, partially positive, neutral, partially negative, negative) for each day. • Shows through composite bar chart sentiment analysis result. One bar for each class.
Pre-Conditions	User must be logged in successfully and has created a project.
Post-Conditions	User should have shown sentiment analysis on scraped results.

UC_visualization_of_scraped_results	
Use Case Name	UC_visualization_of_scraped_results
Use Case ID	UC_08
Actors	User
Description	This Use Case is used by user to see the visualization of the scraped data results.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Analysis tab. **System Response:** • Shows the pie chart indicating the visualization of data scraped from different sources like news, blogs, web, forums)
Pre-Conditions	User must be logged in successfully and has created a project.
Post-Conditions	User should see the visualization of scraped data results.

UC_visualization_of_analysis_results	
Use Case Name	UC_visualization_of_analysis_results
Use Case ID	UC_09
Actors	User
Description	This Use Case is used by user to see the visualization of sentiment analysis results.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Analysis tab. **System Response:** • Shows the time series graph indicating one line for each output class i.e. 5 classes(positive, partially positive, neutral, partially negative, negative) for each day. • Shows through composite bar chart sentiment analysis result. One bar for each class.
Pre-Conditions	User must be logged in successfully and has created project.
Post-Conditions	User should see the visualization of sentiment analysis results.

UC_result_category_filter	
Use Case Name	UC_result_category_filter
Use Case ID	UC_10
Actors	User
Description	This Use Case is used by user to perform filter according to the result category.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Mentions tab. **System Response:** • Shows a slider on the right side by moving it forward or backward filter can be performed according to the result category (positive, partially positive, neutral, partially negative, negative) and only those links and reviews will be displayed.
Pre-Conditions	User must be logged in successfully and has created a project.
Post-Conditions	User should see only those links or reviews that he/she wants.

UC_fetched_category_filter	
Use Case Name	UC_fetched_category_filter
Use Case ID	UC_11
Actors	User
Description	This Use Case is used by user to perform filter according to results fetched category.
Basic Flow	**User Action:** • User clicks on created project. • A new page will be opened in which user clicks on the Mentions tab. **System Response:** • Shows the options like (All, News, Forums, Blogs and Web). • Shows only those links/reviews that user selects from the above mentioned options.
Pre-Conditions	User must be logged in successfully and has created project.
Post-Conditions	User should see links according to result fetched category.

Chapter 3

Design Document

3.1. Introduction

Third deliverable is all about the software design. In the previous deliverable, analysis of the system is completed. So, we understand the current situation of the problem domain. Now we are ready to strive for a solution for the problem domain by using object-oriented approach. Following artifacts must be included in the 3rd deliverable.

a) Domain Model
b) System Sequence Diagram
c) Sequence Diagram
d) Collaboration Diagram
e) Operation Contracts
f) Design Class Diagram
g) State Transition Diagram
h) Data Model

Now we discuss these artifacts one by one in the following section.

3.2. Domain Model Diagram

Domain Model diagram is the UML diagram which is used to depict behaviors and entities of a system with respect to particular domain.

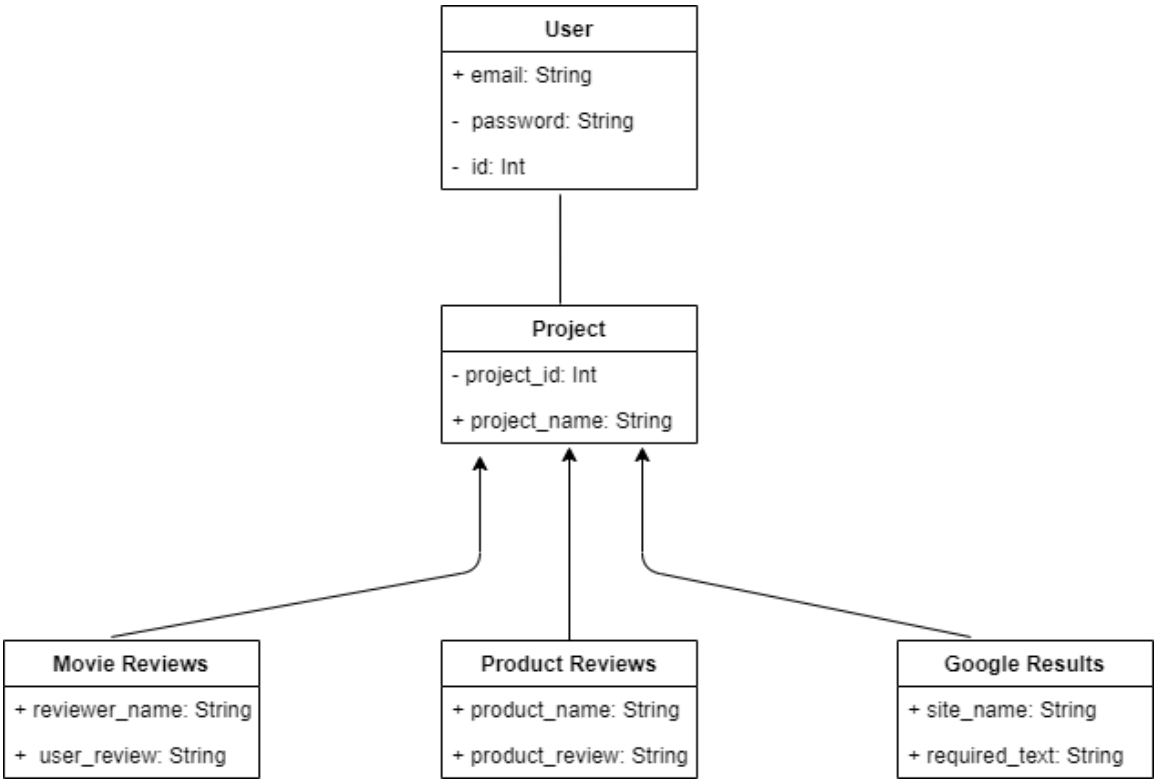


Figure 3.1: Domain Model

3.3. System Sequence Diagram

The UML system sequence diagram (SSD) illustrates events sequentially input from an external source to the system. The SSD will define the system events and operations. System sequence diagrams are a timeline drawing of an expanded use case. Events are related by time with the top events occurring first. System events are the important items. These are events that cause a system response. Use case text may be placed on the left side of the system sequence diagram if desired. If this is done it is best if the use case information lines up with the events in the system sequence diagram.

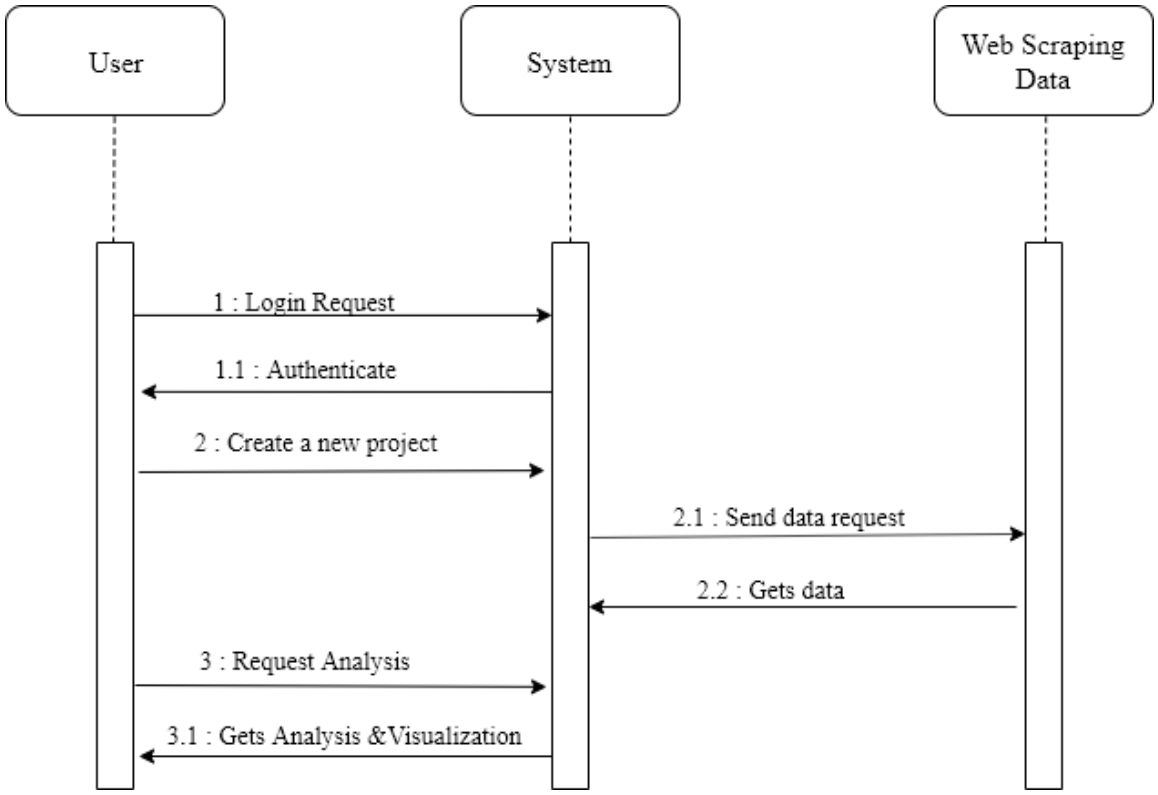


Figure 3.2: System sequence

3.4. Sequence Diagram

Sequence Diagram used to show the interaction of actors with the different behaviors or objects of the system in relevance to the time. A Sequence diagram depicts the sequence of actions that occur in a system. The invocation of methods in each object, and the order in which the invocation occurs is captured in a Sequence diagram.

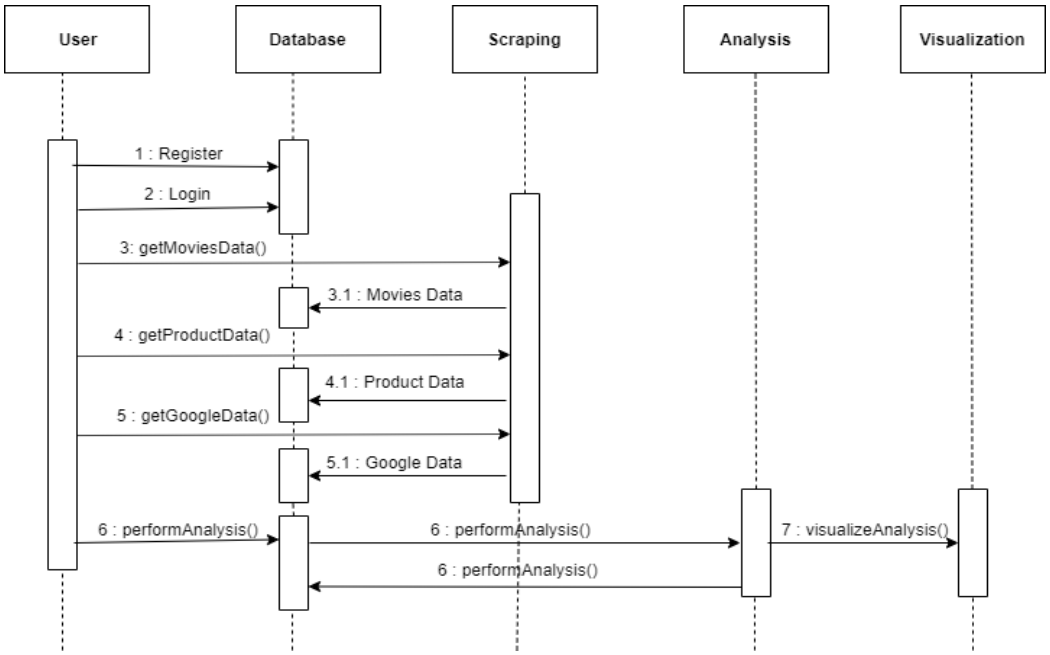


Figure 3.3: Sequence diagram

3.5. Collaboration Diagram

Collaboration diagrams used to show how objects interact to perform the behavior of a particular use case, or a part of a use case. Along with sequence diagrams, collaborations used by designers to define and clarify the roles of the objects that perform a particular flow of events of a use case.

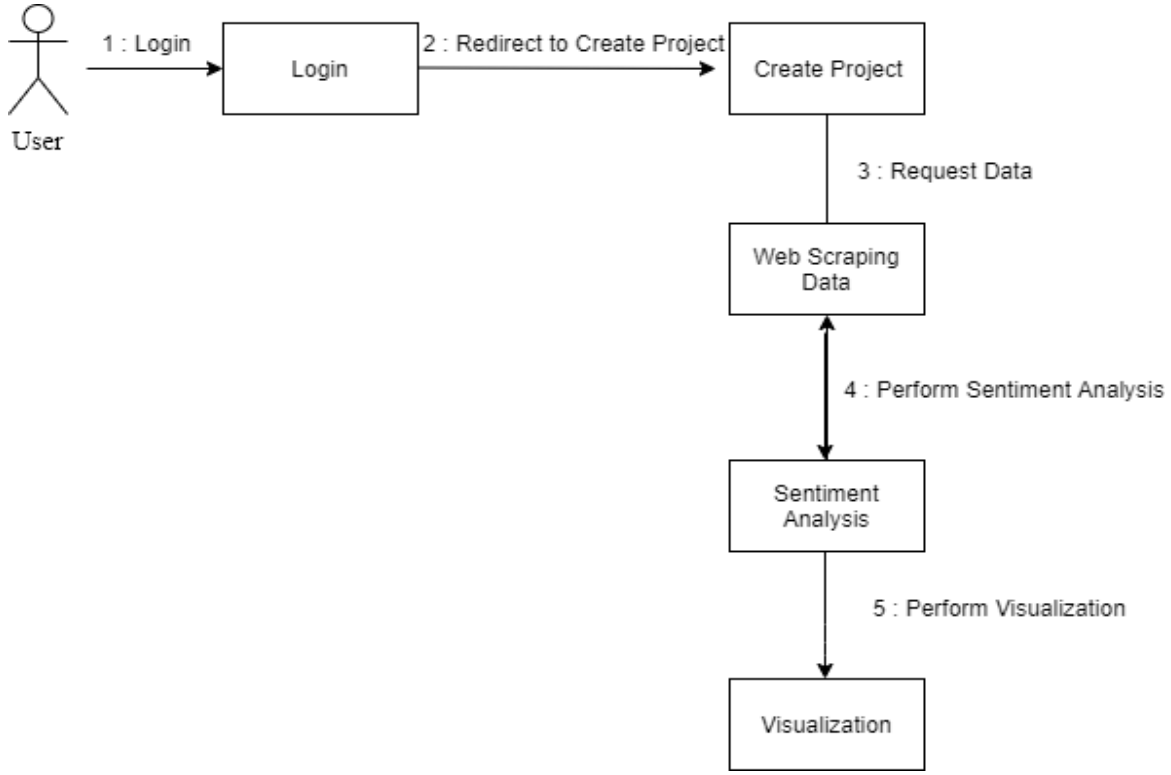


Figure 3.4: Collaboration diagram

3.6. Operation Contracts

Name	Responsibilities	Cross Reference	Exceptions	Pre-condition	Post condition
Login	Logging in Platinum business analytics application	UC_login	Does not login	Must be registered successfully.Get email address and password.	If logged in successful then redirected to create project.
Project creation	Create project of specific category (movie, product, other).	UC_register	Does not create project	Must be a registered user and logged in successfully.	Project created successfully.
Analysis	Perform sentiment analysis on user given text.	UC_sentiment_analysis	No analysis performed	Get input in the form of text from the user	Performed sentiment analysis successfully.
Scraped data results	See the scraped data results just after creating project.	UC_scraped_data	Can not see scraped data.	Must have created project and is logged in.	Sees the scraped data results.
Sources of data scraped	See the sources of data scraped.	UC_scraped_data_sources	Can not see the sources of data scraped	Must have created project and is logged in.	Sees the sources of scraped data.
Analysis on scraped results	Perform sentiment analysis on scraped results.	UC_analysis_scraped_results	No analysis performed	Must have created project and is logged in.	Performed sentiment analysis successfully.
Visualization	See the visualization of sentiment analysis results.	UC_visualization_of_analysis_results	Can not see visualization	Must have created project and is logged in.	Sees the visualization of results

Table 3.1: Trace-ability Matrix

3.7. Design Class Diagram

A class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

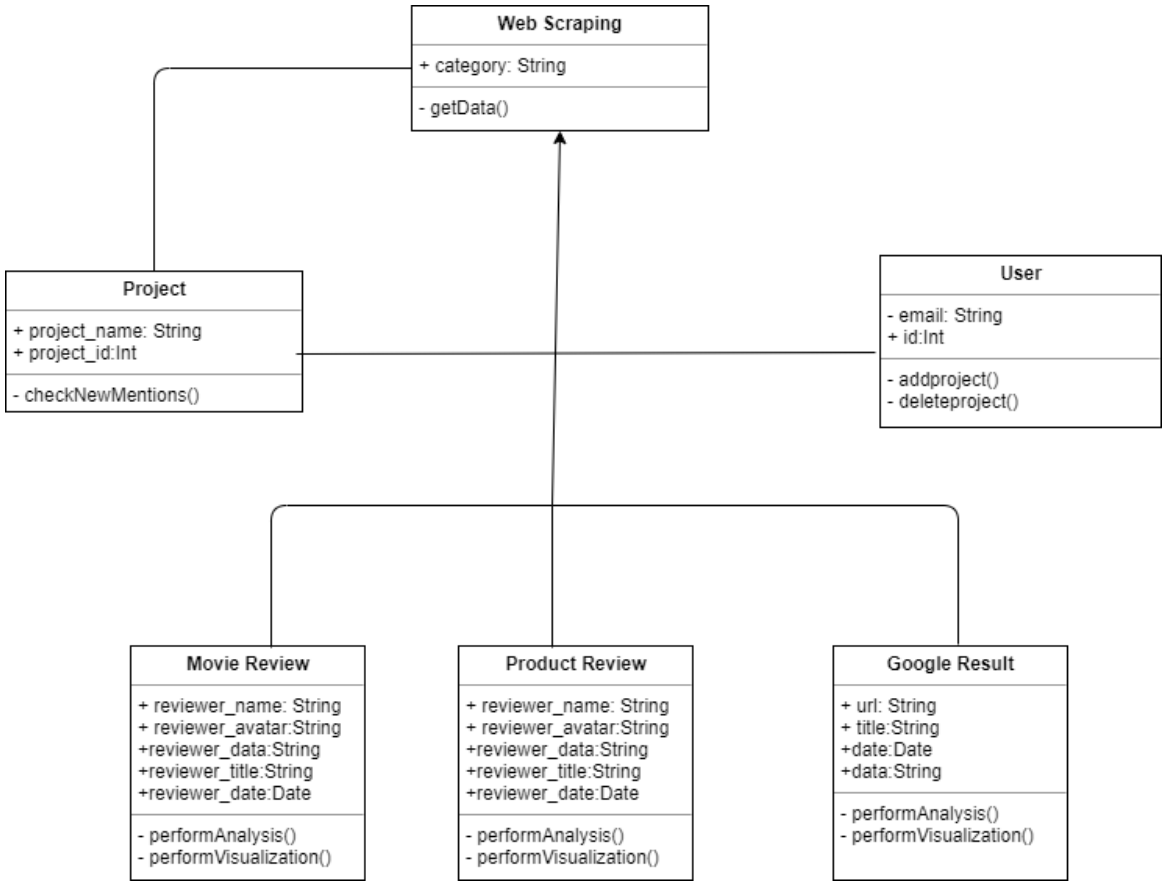


Figure 3.5: Class diagram

The main classes of Platinum Business Analytics are:

- User
- Web Scraping
- Project
- Movie Reviews
- Product Reviews
- Google Results

3.8. State Chart Diagram

State chart diagram is one of the five UML diagrams used to model the dynamic nature of a system. They define different states of an object during its lifetime and these states are changed by events. State chart diagrams are useful to model the reactive systems. In our project, we have drawn State chart diagram using functionality provided by our application and according to the use case given in previous chapter.

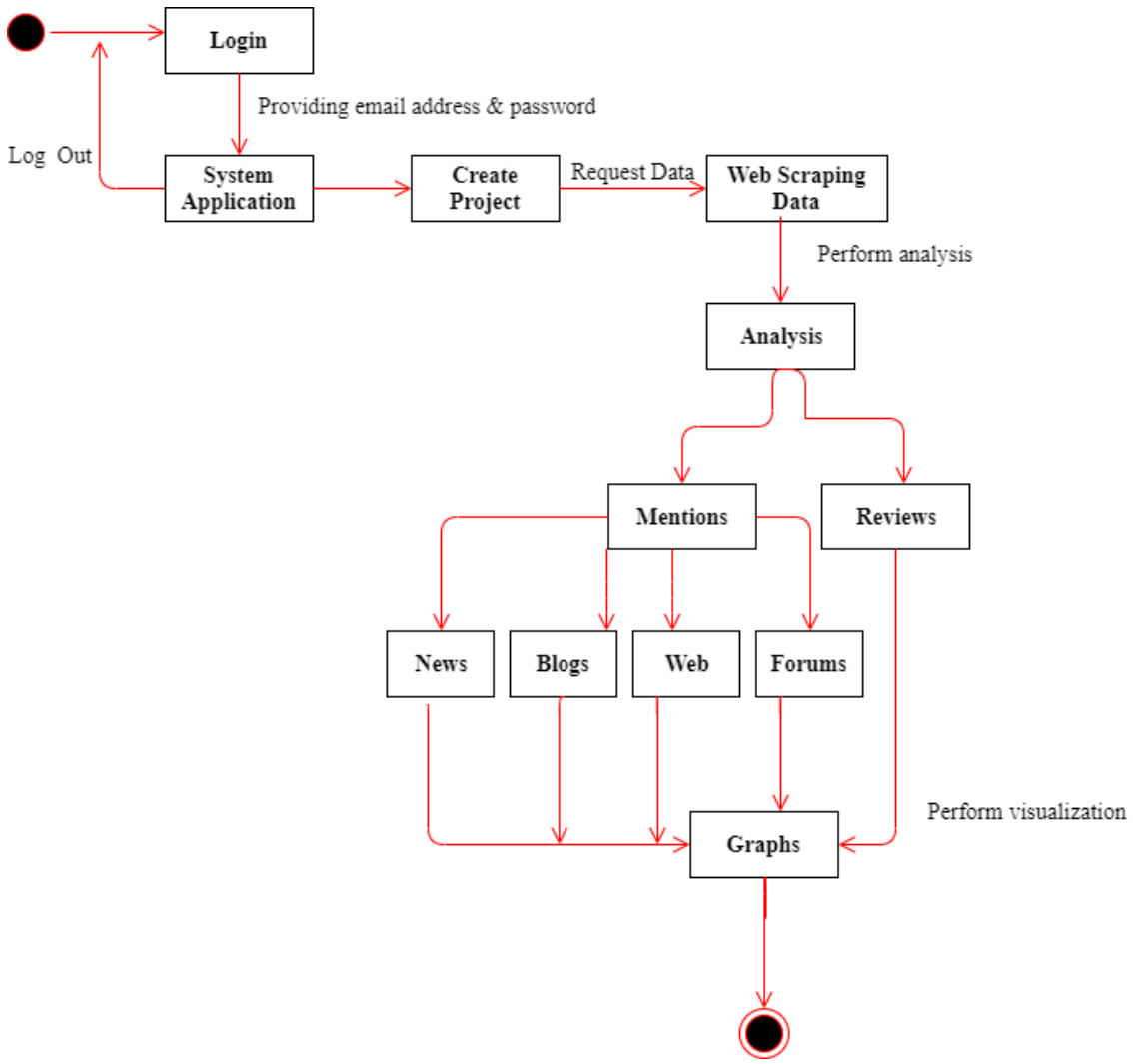


Figure 3.6: State Chart diagram

3.9. Data Model

The entities of data model are:

- User
- Projects
- Data

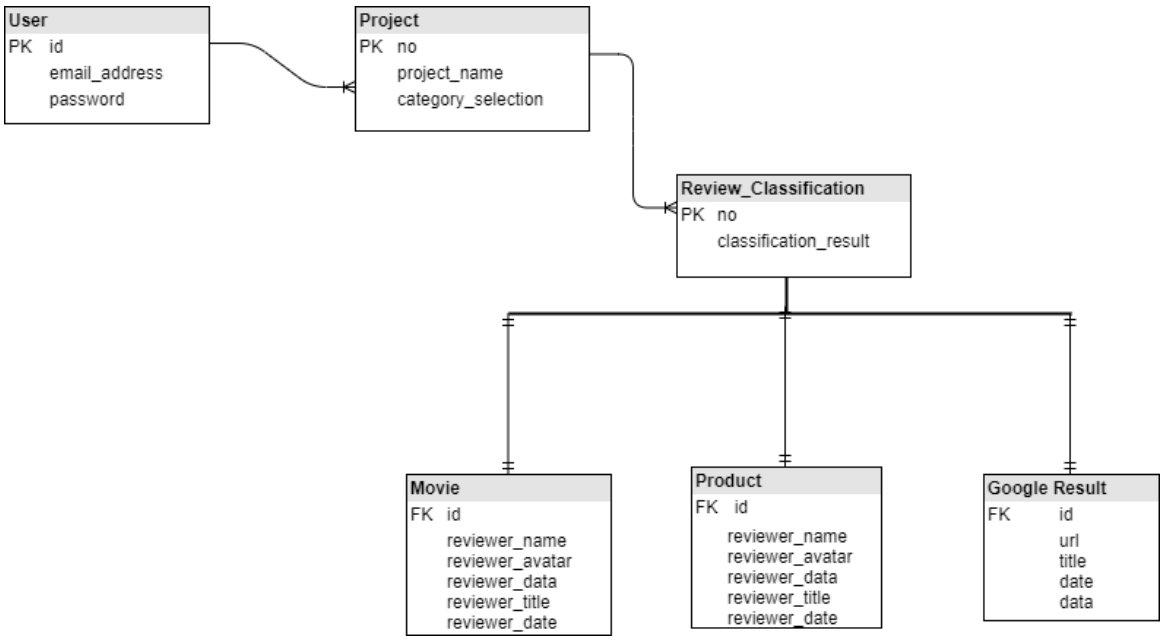


Figure 3.7: Data Model

Chapter 4

User Interface Design

4.1. Introduction

A user interface design consists of three main parts:

Page elements should be visualized on paper before building them in the computer. Just as you draw a site map to plan the site, use cartoons and storyboards to begin blocking out the site's appearance and navigational scheme.

a) Site maps
b) Storyboards
c) Navigational maps

4.2. Site Maps

A site maps main benefit is to give users an overview of the site's areas in a single glance by dedicating an entire page to a visualization of the information architecture. The site map's goal is to give users a single overview of the information space.

The site map for "Platinum Business Analytics" is very simple and compact.

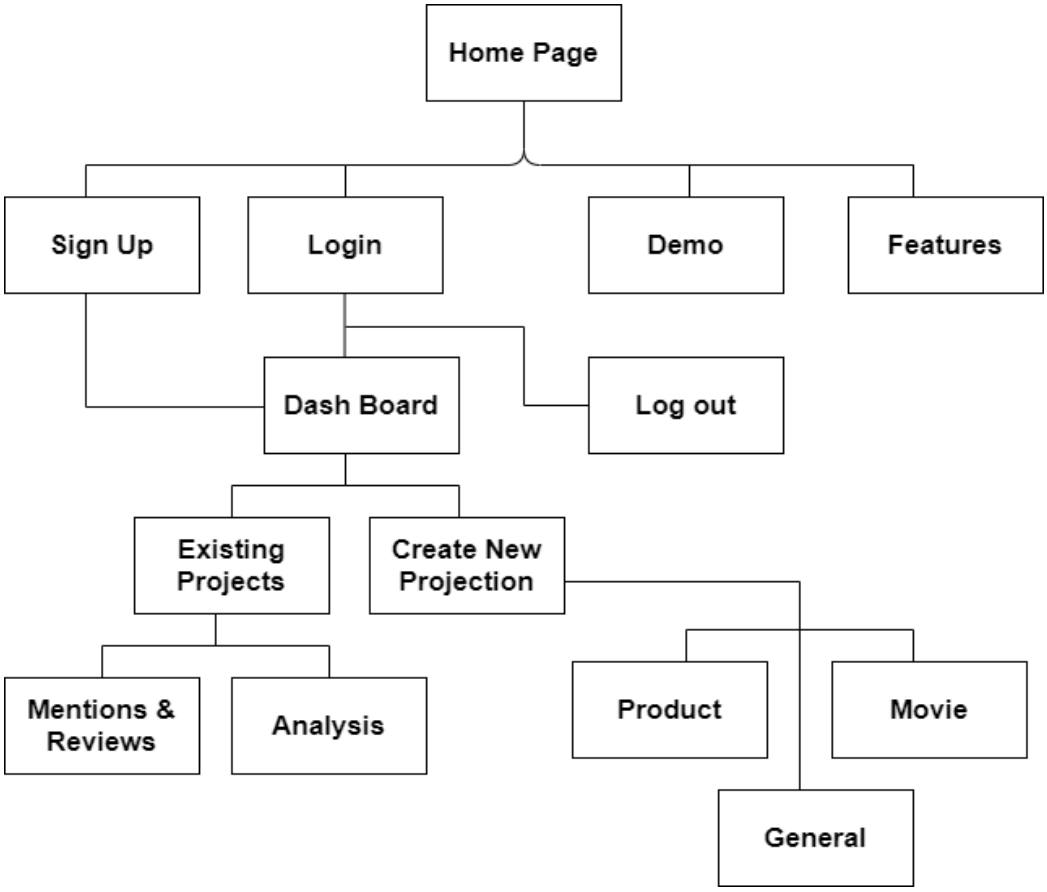


Figure 4.1: Site Map

4.3. Story Boards

A storyboard is a sequence of single images, each of which represents a distinct event or narrative. It is also a visual representation of the script illustrating the interaction between the user and the machine. It can also be imagined as a film in visual-outline form. These are the detailed screens, which pictorially represent the complete view ofthe screens. There would be symbols representing the different elements of the screens and in the end an index that would detail the symbols.

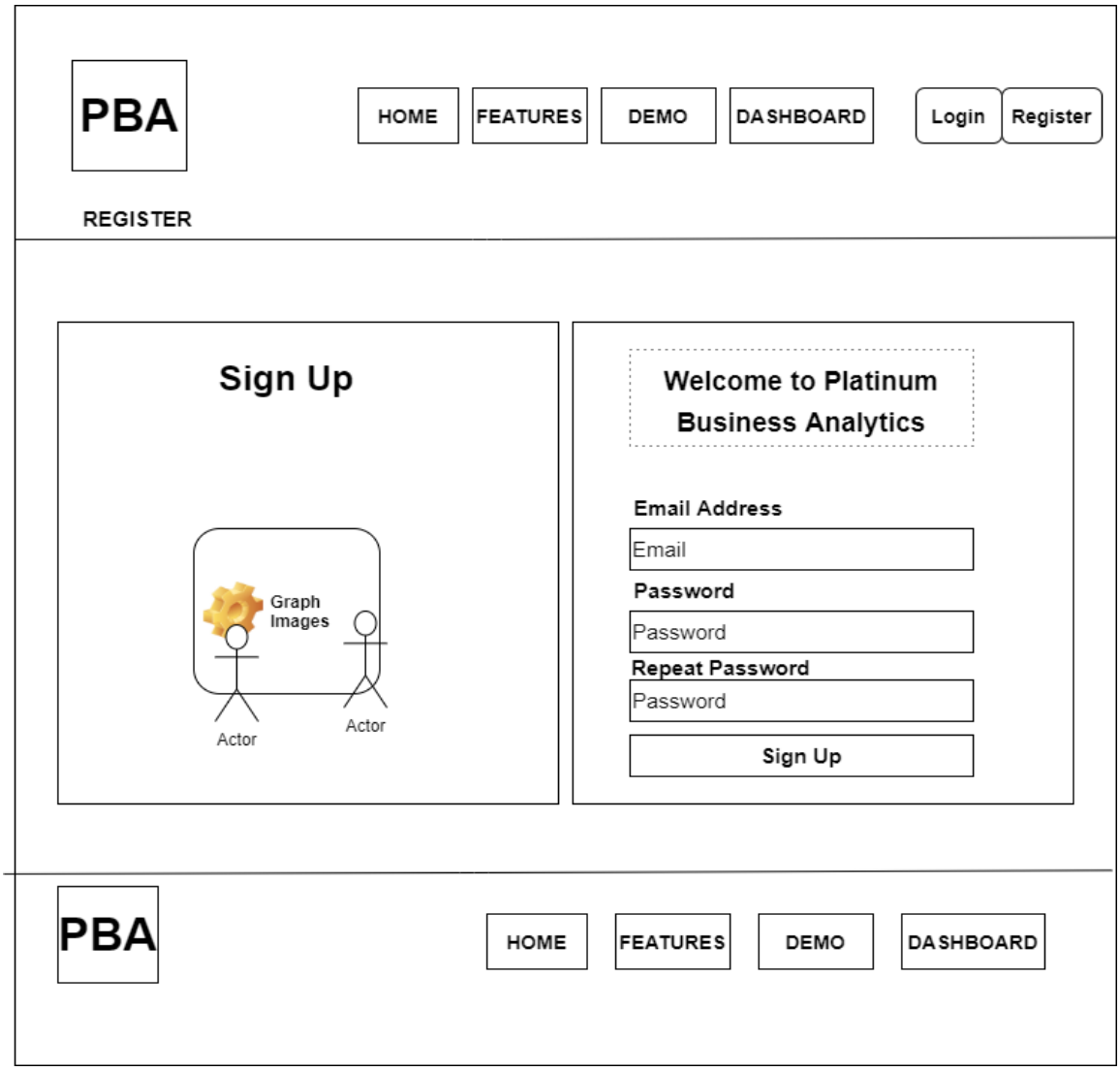


Figure 4.2: Story Board Register Page

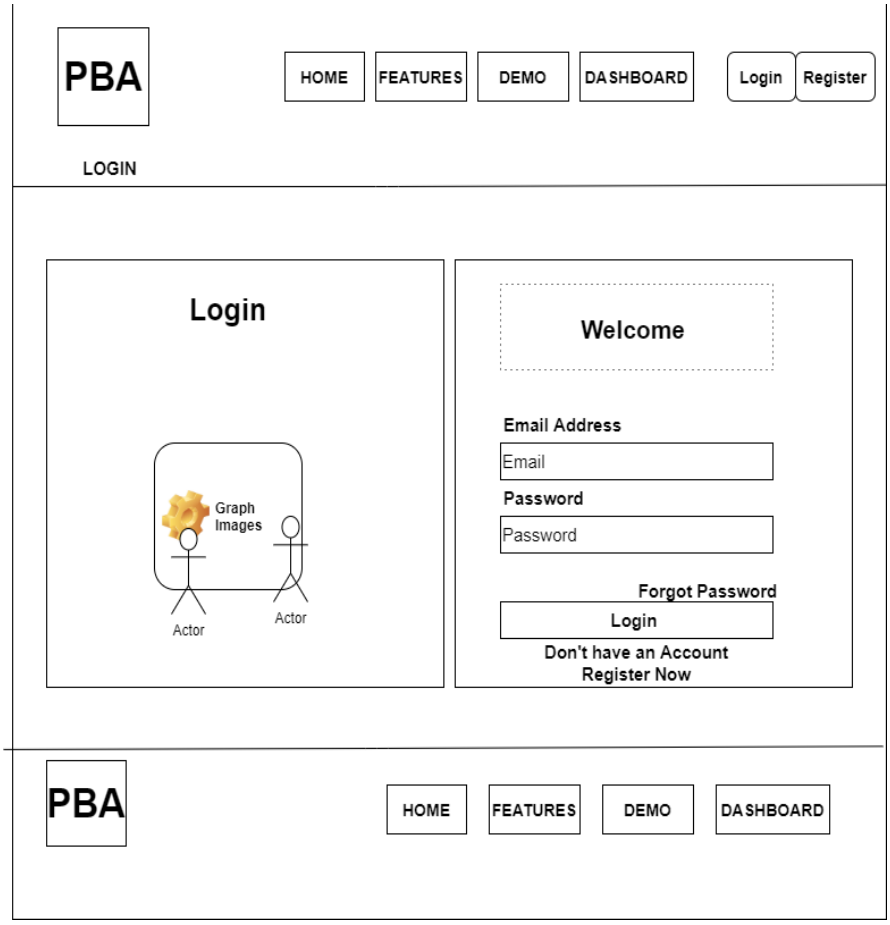


Figure 4.3: Story Board Login Page

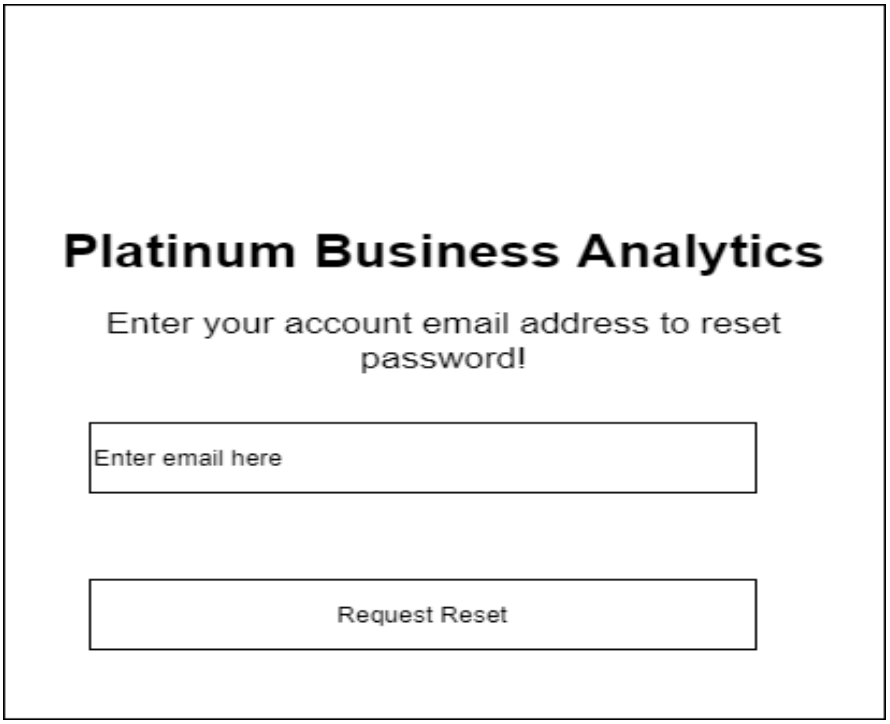


Figure 4.4: Story Board Password Reset

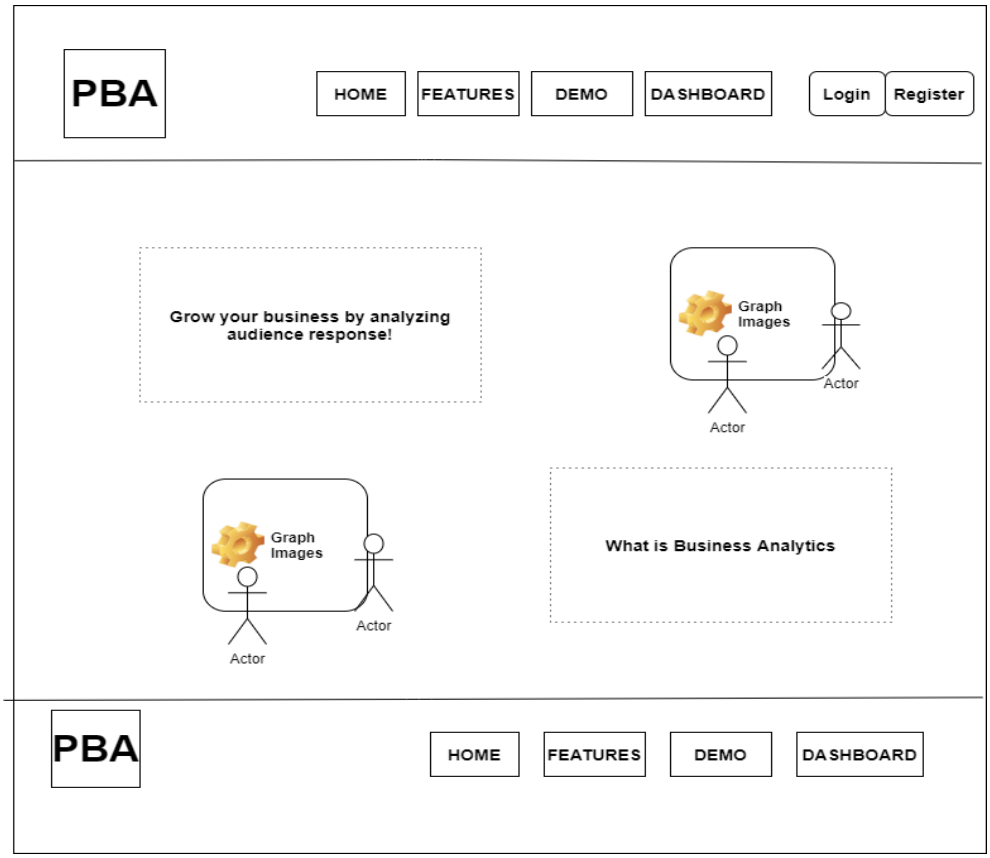


Figure 4.5: Story Board Home Page

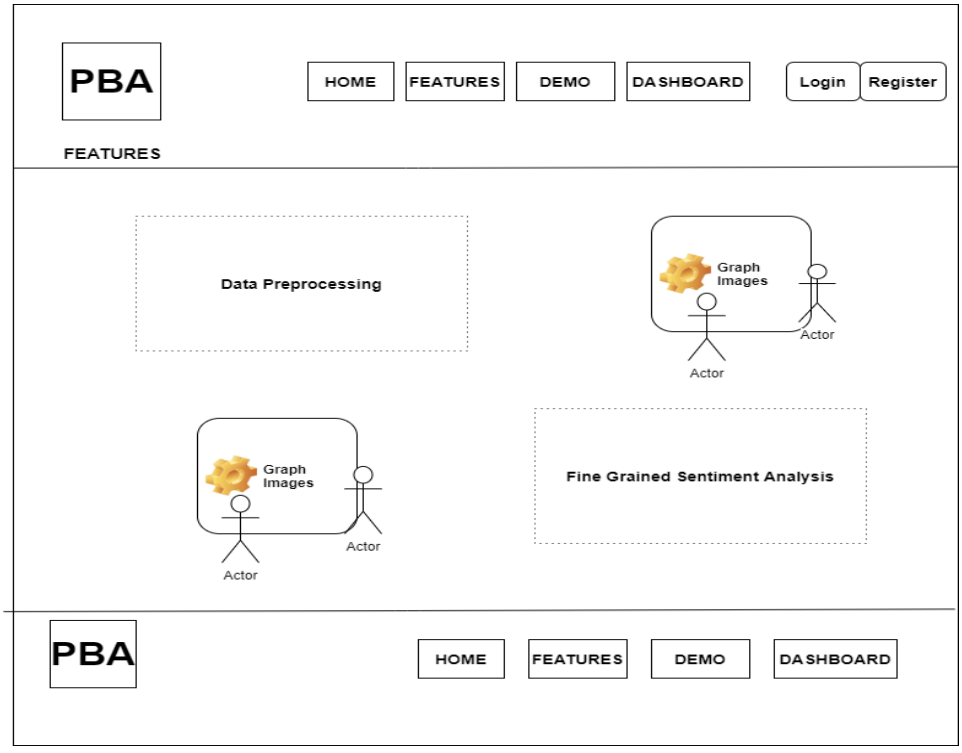


Figure 4.6: Story Board Features Page

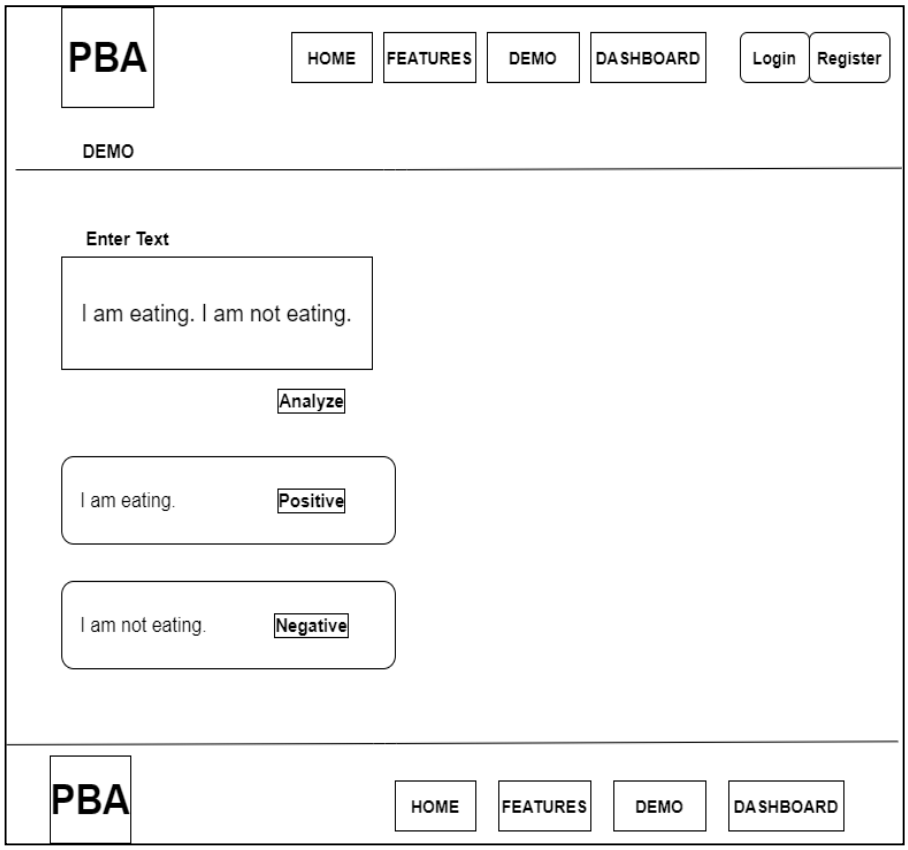


Figure 4.7: Story Board Demo Page

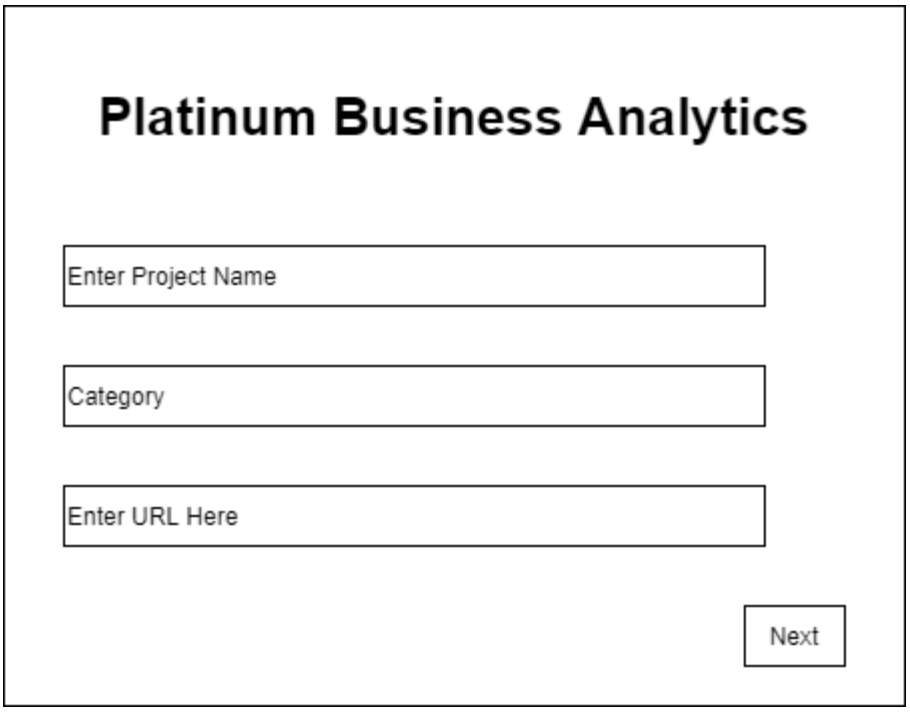


Figure 4.8: Story Board Project Creation Page

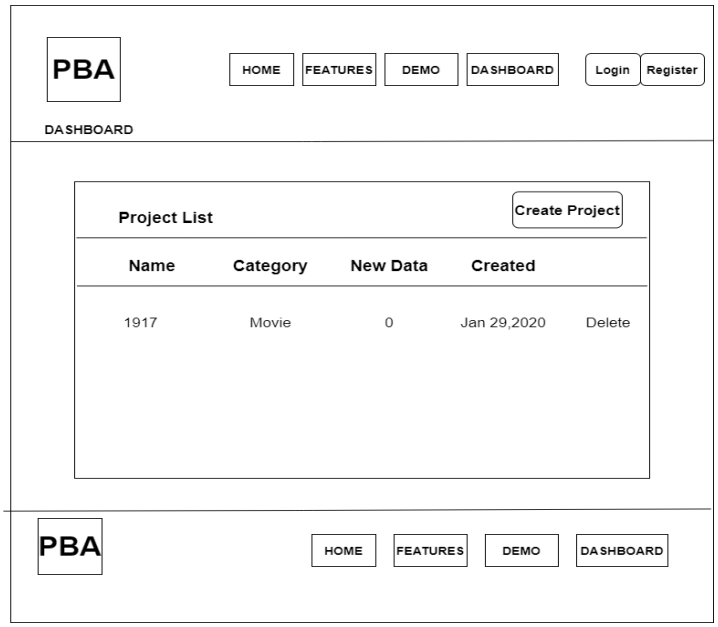


Figure 4.9: Story Board Dashboard Page

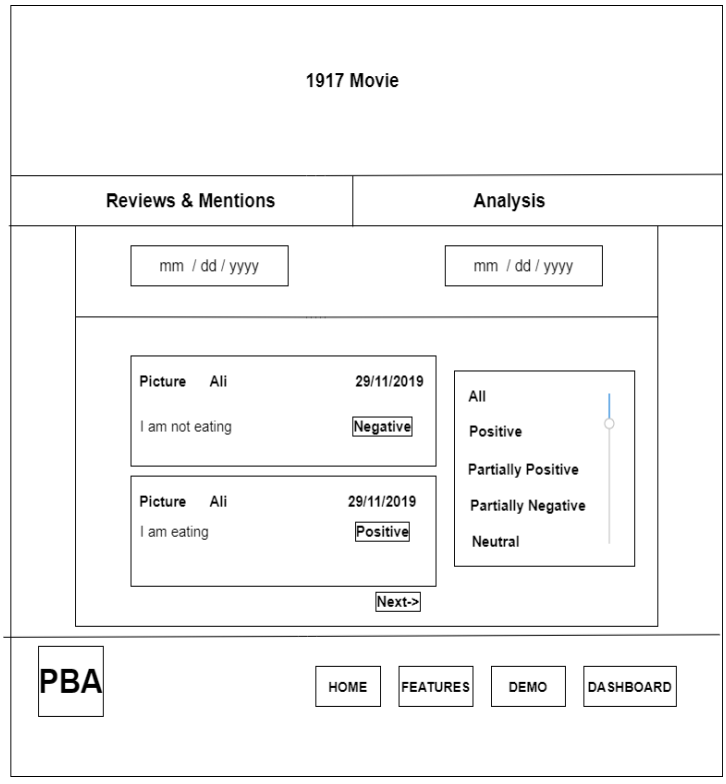


Figure 4.10: Story Board Reviews & Mentions Page

Figure 4.11: Storyboard Analysis Page

4.4. Navigational Maps

The next step is of navigational maps. In these maps, the storyboards are used as an input. The different display buttons or action buttons show the navigation from one screen to the other. In other words when one action button is pressed it would lead to other screens. This path and navigation would be shown.

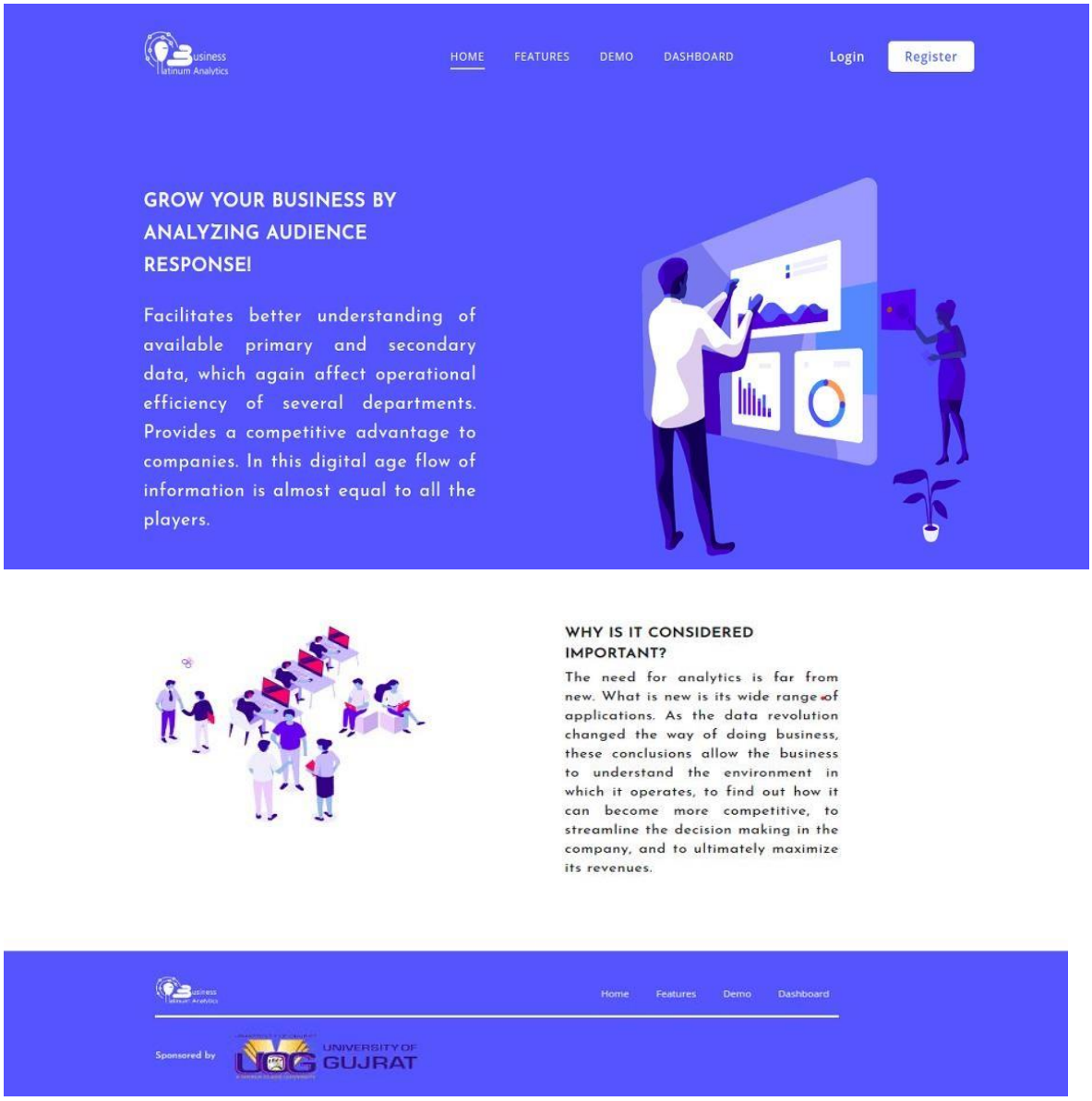


Figure 4.12: Home Page or Landing Page displayed when user enters url in the address bar

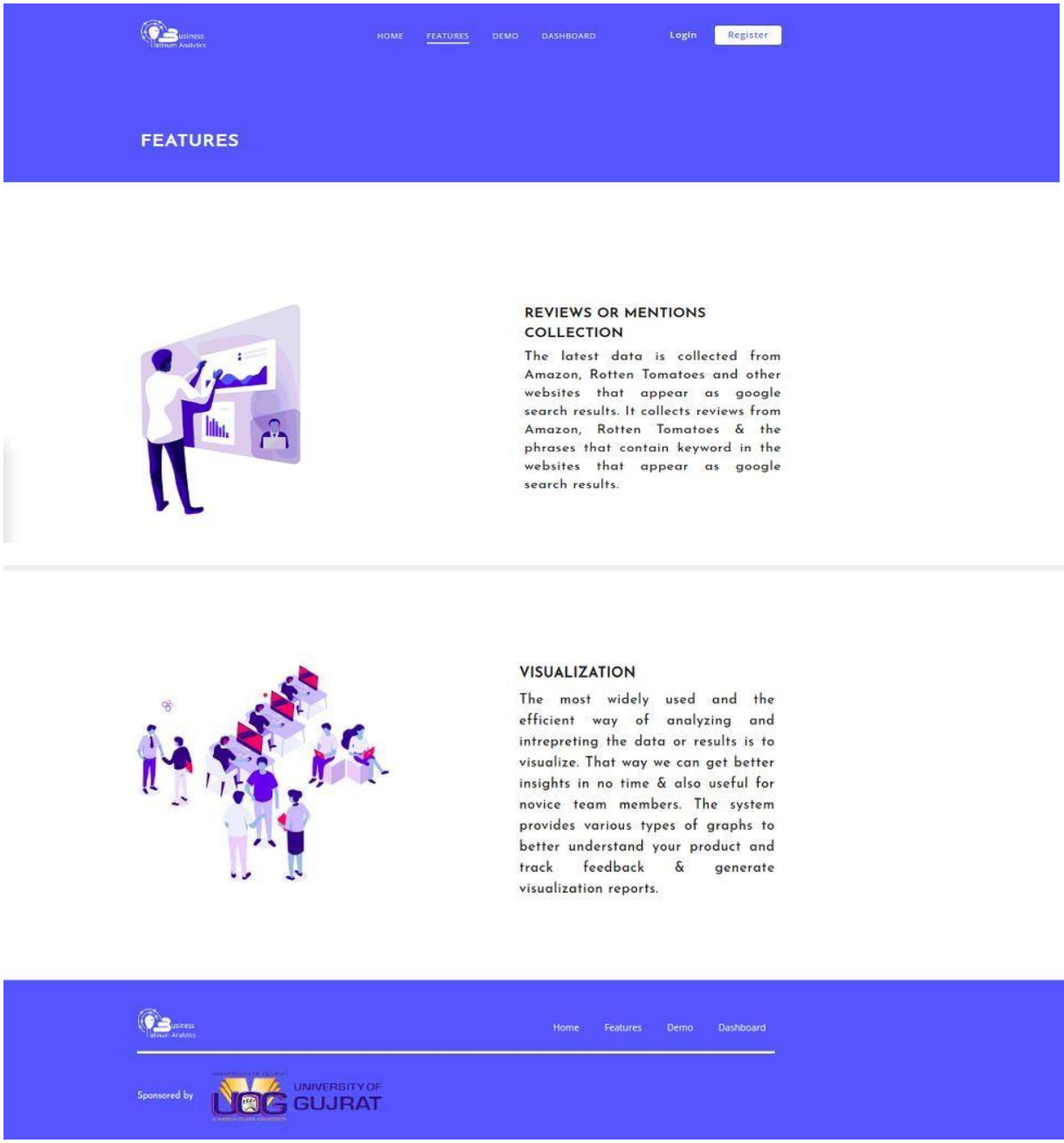


Figure 4.13: Features Page displayed when user clicks on Features tab

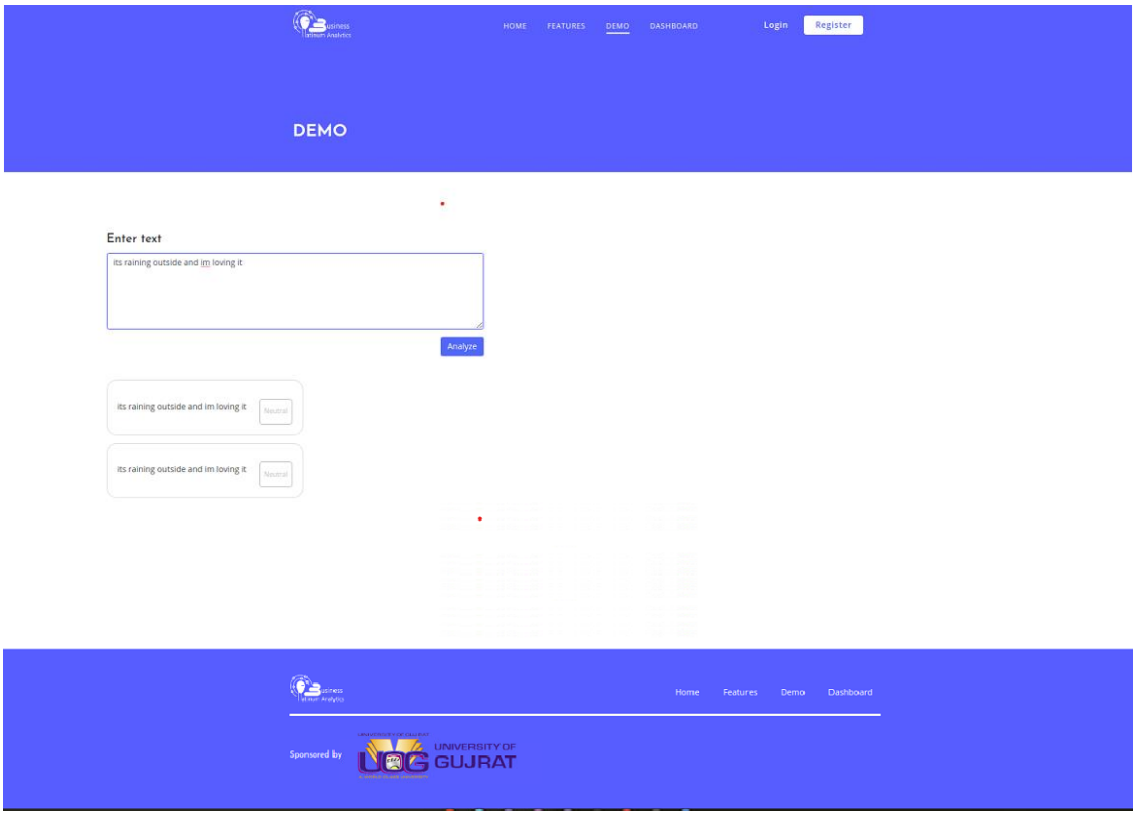


Figure 4.14: Demo Page displayed when user clicks on Demo tab

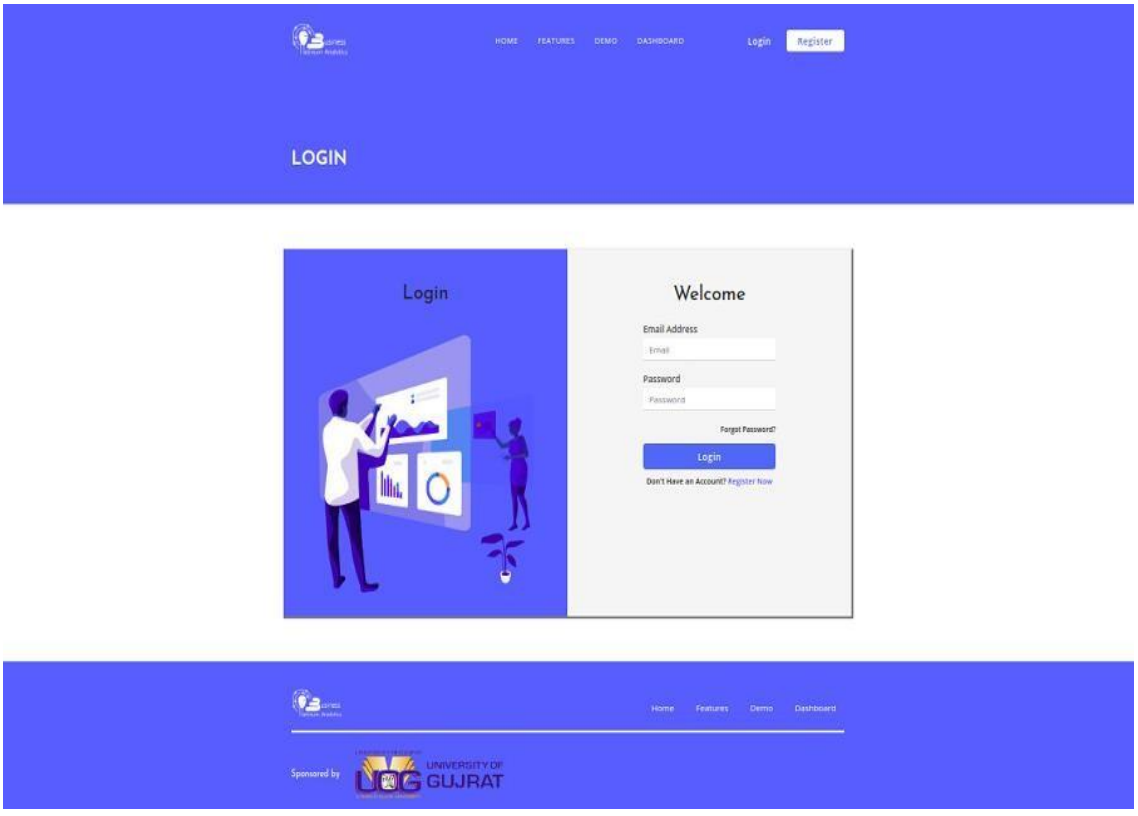


Figure 4.15: Login Page displayed when user clicks on Login button

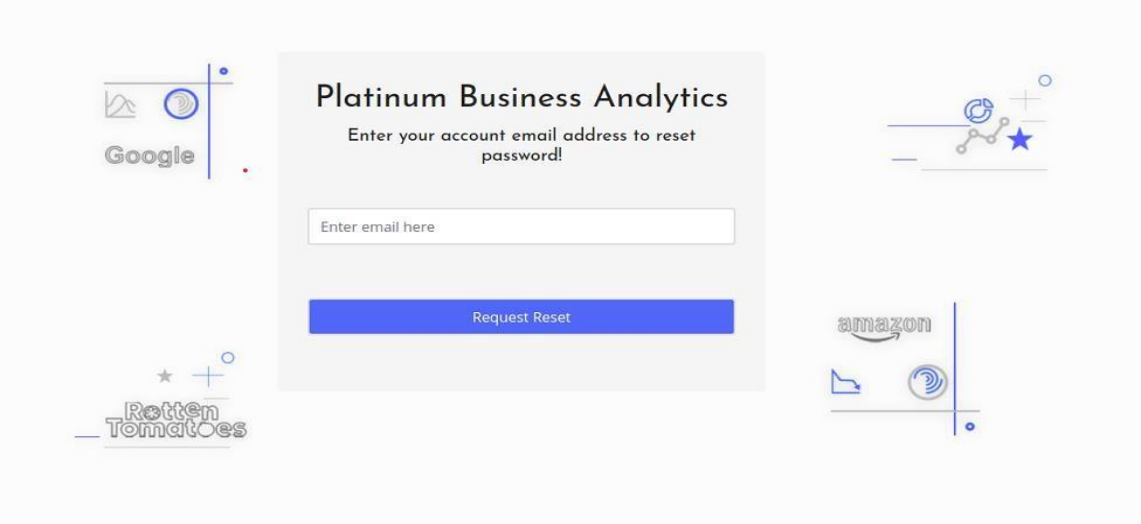


Figure 4.16: Reset Password Page displayed when user clicks on Forgot password in the Login Page

Figure 4.17: Register Page displayed when user clicks on Register button

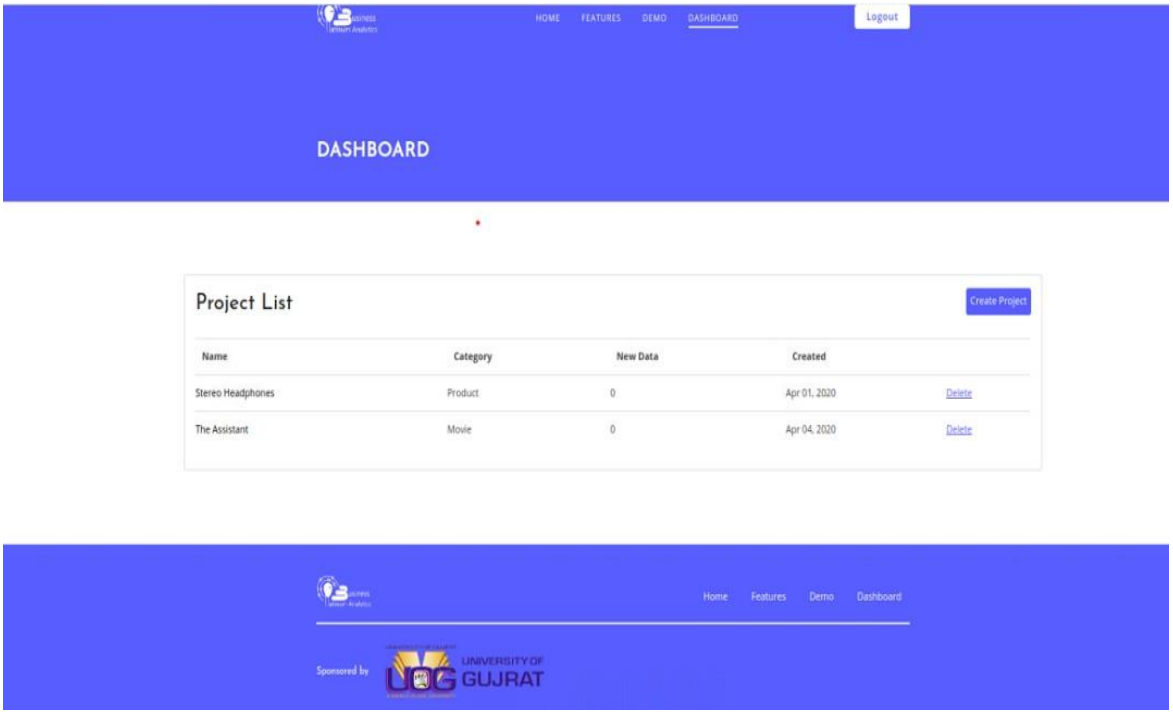


Figure 4.18: Dashboard Page displayed when user successfully login

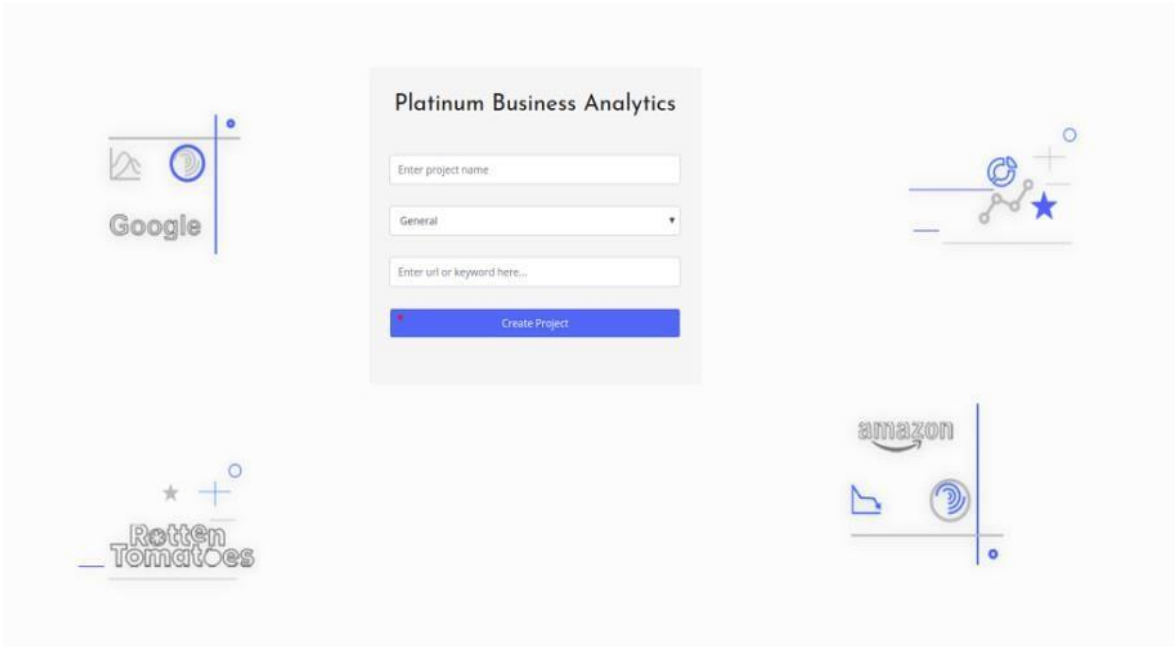


Figure 4.19: Project Creation Page displayed when user clicks on Create Project button in the Dashboard Page

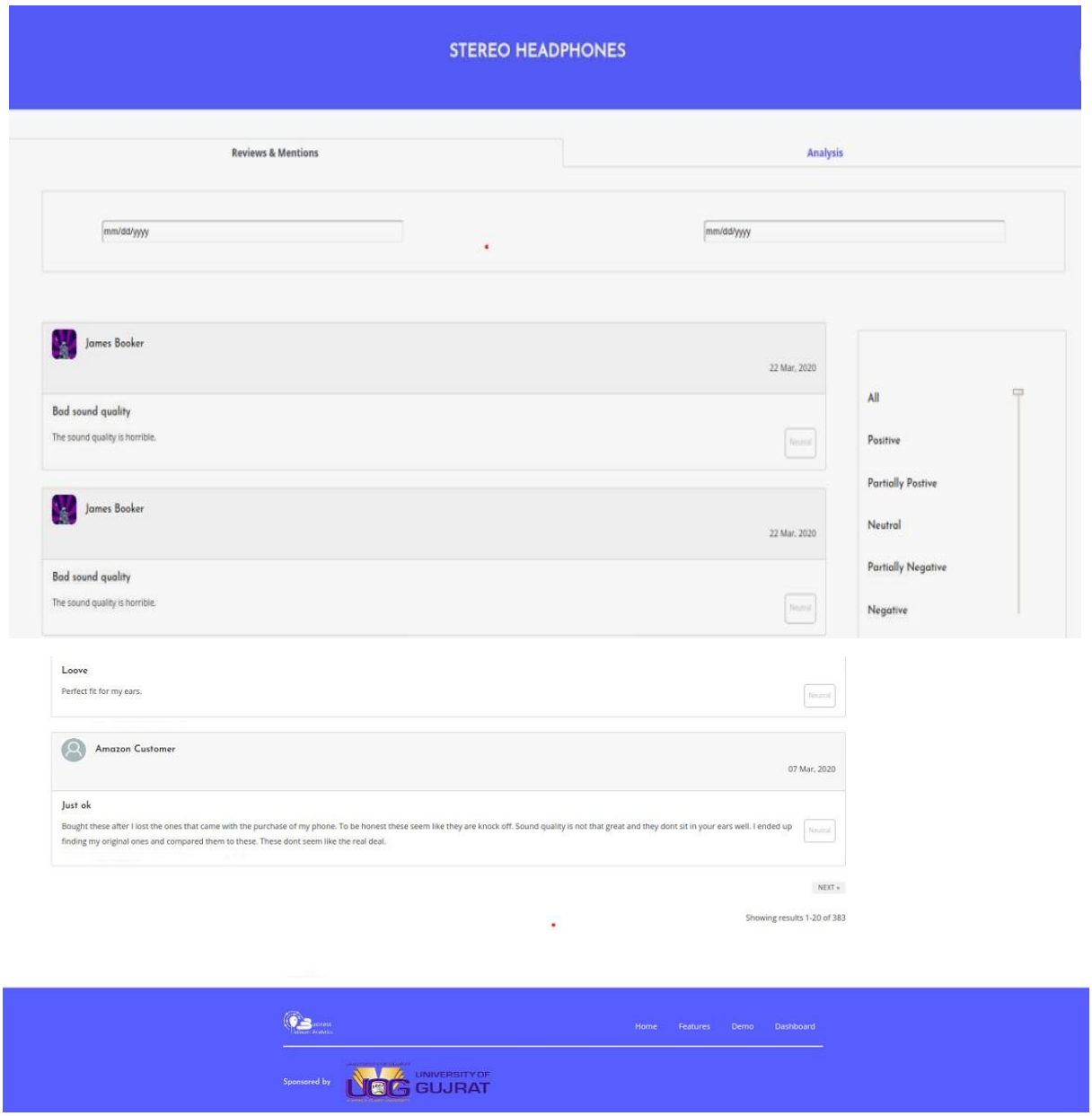


Figure 4.20: Reviews & Mentions Page displayed when user clicks on any of the created projects in Dashboard Page

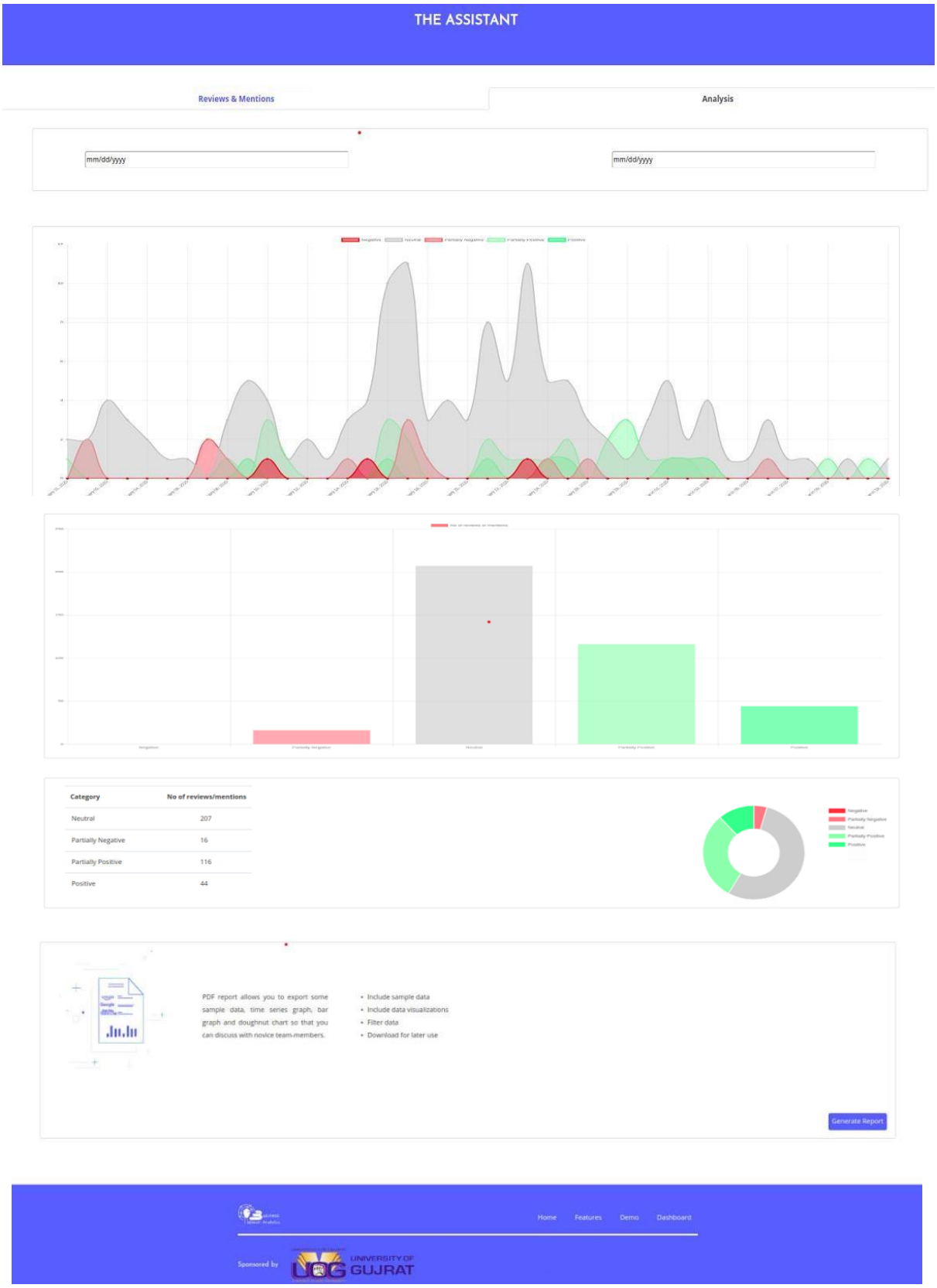


Figure 4.21: Analysis Page displayed when user clicks on Analysis tab from Reviews & Mentions Page

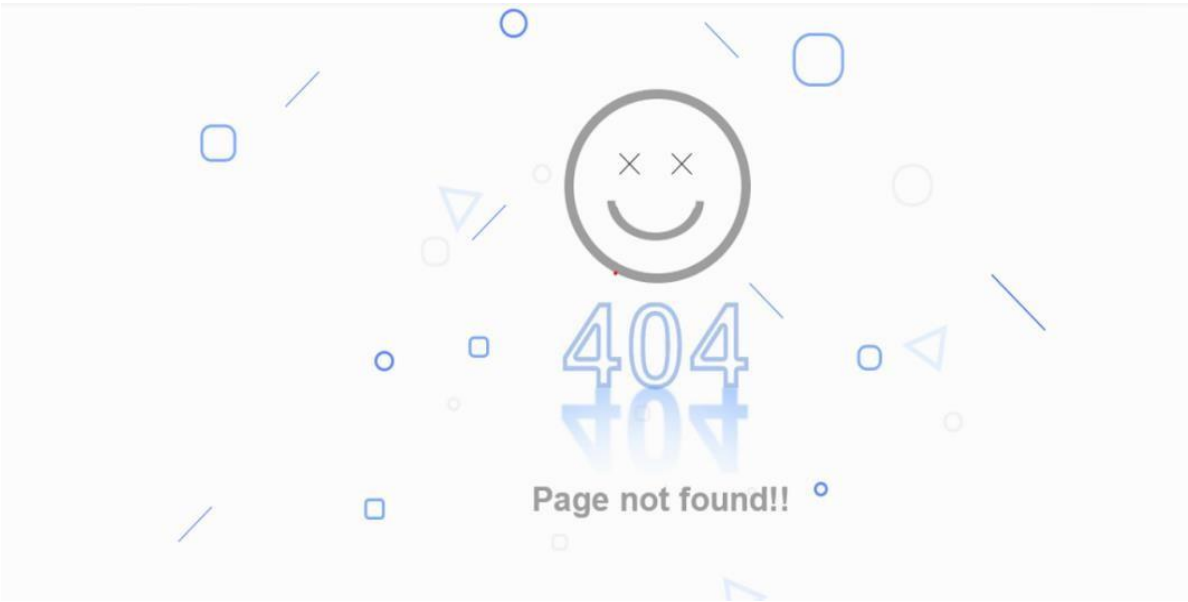


Figure 4.22: In case if wrong URL is entered by user then this page is displayed

Figure 4.23: when user unsubscribe email notifications from Gmail then this page is displayed

Chapter 5

Software Testing

5.1. Introduction

From the start of software engineering different software testing schemes and testing templates has invented. These testing templates are used to test both UI and unit testing. Every credible software house developed their testing documentation for the purpose of developing efficient systems.
To provide a common set of standardized documents the IEEE developed the 829 Standard for Software Test Documentation for any type of software testing, including User Acceptance Testing.
In our testing module, we are going to test different features and modules of the system using documentation template supported by IEEE829.

5.2. Test Plan

5.2.1. Purpose

The purpose of this test plan is to test real-time web application that will give knowledge about audience response related to a product or movie. The purpose of this document presents a detailed explanation of the objectives, features, user interface and application of "Platinum Business Analytics". It will also describe how the system will perform and under which it must operate.

5.2.2. Outline

A test plan shall have the following structure:

a. Test plan identifier
b. Introduction
c. Test items
d. Features to be tested
e. Features not to be tested
f. Approach
g. Item pass/fail criteria
h. Suspension criteria and resumption requirements
i. Test deliverables
j. Testing tasks
k. Environmental needs
l. Responsibilities
m. Staffing and training needs
n. Schedule
o. Risks and contingencies
p. Approvals

5.2.2.1. Test Plan Identifier

The identifier for this test plan: **PBA_TP**
The abbreviation for the identifier is as:

PBA	Platinum Business Analytics
TP	Test Plan

5.2.2.2. Introduction

The test plan is for testing the whole system including modules and their features. The testing of application features will be conducted manually within 8 to 10 days. All these testing procedures needs to be performed in time in order to reduce the complexities. In multilevel test plans, each lower-level plan must reference the next higher-level plan.

5.2.2.3. Test Items

Following are the testing items which we are going to test in this plan:

- User Sign up
- User Login
- Demo Sentiment analysis
- Project creation analysis
- Created Projects analysis

All these are the generalized items which will in fact contain different tests to remove anomalies and defects from the Project.

5.2.2.4. Features to be tested

All features will be tested to check whether they are functioning properly or regularly.

- Validations on user registration
- Validations on user login
- Forget password
- Sentiment Analysis on user given text
- Validations on project creation
- Email notification test
- Unsubscribe email notifications
- Reviews analysis
- Graphs analysis
- Pdf generated report

5.2.2.5. Features not to be tested

As all the testing is going to be check on every feature, so it will make sure that nothing is left without testing.

5.2.2.6. Approach

We are going to use the following testing technique provided in Python to test our project.

- **Unit Testing**

Unit testing is a testing software where we test individual methods, functions and classes. Unit test covers almost 70 percent of test coverage. In unit wise testing we test all features of our application, tasks are divided each member test the application and submit results to group leader.

5.2.2.7. Item Pass/Fail Criteria

All core functionality of the application and its features should function as expected result. All testing tasks will be performed within the ratio of 100% in which passing criteria is 75%. If the test results are less than 75%, it will be considered fail.

- If system do not work properly it will be considered as fail case.
- If expected page of website won't appear then it will be considered as fail case.
- If website does not scrap relevant data then it will be considered as a fail case.
- If the model does not accurately classify sentiments into five categories then it will be considered a fail case.

Each member of a group will test each feature and mark each case as Pass/ Fail. Each member will note the actual result and all relevant data or details. Once all tests will complete, the test leader will review the test report and check it is fail or pass.

If all the written tests are passed successfully then they are termed to be a successful test.

5.2.2.8. Suspension Criteria and Resumption Requirements

Testing work is performed by the individual to ensure that all the functional activities are working well in the system. If modules are not working well resumed the test procedure and check the results. In case of failure, repeat the cycle until the required results are obtained.

If number of errors will be more than test passed, this will be considered as failed test and testing will be suspended. Once all the issues will be address, testing will be resumed again.

5.2.2.9. Test Deliverables

After completion the result will be saved and the tester will circulate the result to the group members and high authorities. The following are the test deliverable:

a) Test plan document

b) Test design specifications document

c) Test case specifications document

d) Test procedure specifications document

e) Defective system

f) Non-defective system

5.2.2.10. Testing Tasks

Following tasks should be performed before testing:

a) Latest version of Pycharm professional addition installed.
b) Add dependencies in the Pycharm needed for test execution.
c) Design and plan test in order to write and execute them.

5.2.2.11. Environmental Needs

We need following environments for the execution of unit tests:

- Access of the internet.
- Pycharm.
- Latest web page supporting browser.
- Latest operating system should be installed.
- Computer system with minimum of 8G RAM.

5.2.2.12. Responsibilities

The main person which is responsible for the testing execution and management is called test manager. In our documentation and project the testing is performed by Muhammad Shaheryar Khan along with the other two members (Muhammad Zaryab and Zeeshan Ahmad).

5.2.2.13. Staffing and Training Needs

a) To apply systematic testing methods test team, require a complete understanding of software testing.

b) To apply test and to understand them a person requires complete knowledge of testing components.

c) Training for the team members can be organized to get understanding of testing framework.

5.2.2.14. Schedule

As we have mentioned in our proposal earlier that we will perform testing of project within 23 days. Following is the testing schedule which will be used:

Activities	Responsible	Time Required
Test Planning	Zeeshan Ahmad	5 days
Test Specification	Zeeshan Ahmad	4 days
Use case Testing	Muhammad Shaheryar Khan	7 days
User Interface Testing	Muhammad Zaryab	7 days

Table 5.1: Testing Schedule

5.2.2.15. Risks and Contingencies

It is assumed that the tester is familiar with World Wide Web navigational tools. If the tester doesn't have the basic understanding of application, testing could be delayed or not conducted properly.

a) **Time Management:** as it requires extensive time for the multiple types of testing for different levels, so this is a crucial part of this project.

b) **Lack of internet:** poor speed of internet may cause application to not work properly and the required results cannot be obtained or altered.

c) **Less interactive:** The system would be rejected by the end users if they found it less interactive and difficult to use.

5.2.2.16. Approvals

Test managers and other team members need to approve the overall test strategy.

a) Zeeshan Ahmad _____

b) Muhammad Zaryab _____

c) Muhammad Shaheryar Khan _____

5.3. Test Design Specification

5.3.1. Purpose

Test design specification deals with the designing and predicting about the features which we are going to test. In this phase we also prioritize the test cases in their order of importance.

This is quite important document and is missed by some organization. This is the biggest problem that people try to execute and develop test without first designing them.

5.3.2. Outline

Following is the content which we are going to describe in test design specification.

a) Test design specification identifier
b) Feature to be tested
c) Approach Refinement
d) Testing identification
e) Feature pass fail criteria

5.3.2.1. Test design Specification Identifier

Test design specification identifier is **PBA_DSI.** The abbreviation for the identifier is as:

PBA Platinum Business Analytics

DSI Design Specification Identifier

5.3.2.2. Features to be tested

Following is the list of all the features which will be used.

- Validations on user registration
- Validations on user login
- Forget password
- Sentiment Analysis on user given text
- Validations on project creation
- Email notification test
- Unsubscribe email notifications
- Reviews analysis
- Graphs analysis
- Pdf generated report

5.3.2.3. Approach Refinement

We will be using the same strategy and approach as mentioned in the test plan phase.

5.3.2.4. Testing Identification

In this phase we make a list of testing features which can be tested and can be taken as a test case in the future step of testing.
In our testing strategies we are taking all the test items as test cases.

5.3.2.4. Feature Pass/Fail Criteria

Testing tasks will be performed within the ratio of 100% in which passing criteria is 75%. If the test results are less than 75%, it will be considered fail. If all the written tests are passed successfully then they are termed to be a successful test. Each member of a group will test each feature and mark each case as Pass/ Fail. Each member will note the actual result and all relevant data or details. Once all tests will complete, the test manager or leader will review the test report and check it is fail or pass.

5.4. Test Case Specification

5.4.1. Purpose

To define a test case identified by a test design specification.

5.4.2. Outline

Following are the contents of test case specification:

a) Test case specification identifier
b) Test items
c) Input specification
d) Output specification

e) Environment needs
f) Special procedural requirements
g) Inter case dependencies

5.4.2.1. Test Case Specification Identifier

Following are test cases identifier that will be used in our test:

Serial No	Test Feature	Test Case Identifier
1	Validations on user registration	TC1
2	Validations on user login	TC2
3	Forget password	TC3
4	Sentiment Analysis on user given text	TC4
5	Validations on project creation	TC5
6	Email notification test	TC6
7	Unsubscribe email notifications	TC7
8	Reviews analysis	TC8
9	Graphs analysis	TC9
10	Pdf generated report	TC10

Table 5.2: Test feature table

5.4.2.2. Test Items

Following is the list of all the features which will be tested:

- Validations on user registration
- Validations on user login
- Forget password
- Sentiment Analysis on user given text
- Validations on project creation
- Email notification test
- Unsubscribe email notifications
- Reviews analysis
- Graphs analysis
- Pdf generated report

5.4.2.3. Input Specifications

Here we will specify the input for every test case:

Serial Number	Test Case Input	Test case identifier
1	1- Leave all fields blank and click Sign Up button. 2- Enter such an email address that does not exists. 3- Password and Repeat password are not same. 4- User is already registered.	TC1

	5- Enter valid email address, password and repeat password.	
2	1- Leave all fields blank and click Login button. 2- Enter invalid email address or password. 3- Enter valid email address and password.	TC2
3	1- Enter a valid email address. 2- Enter an invalid email address.	TC3
4	1- Leave the text box blank. 2- Enter some text.	TC4
5	1- Leave all fields blank and click Create Project button. 2- Repeat a project name within same category that is already created. 3- Enter unique project name, choose a category and enter valid url. 4- Enter an invalid url.	TC5
6	User creates an account.	TC6
7	Click unsubscribe in notification email.	TC7
8	Click on the created project.	TC8
9	Click on the created project and switch to Analysis tab.	TC9
10	Click on Generate Report button.	TC10

Table 5.3: Test case input table

5.4.2.4. Output Specifications

Here we will specify the output for every case:

Serial Number	Test Case Output	Test Case Identifier
1	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Email address is invalid" message will be displayed. 3- "Repeat password must be same" message will be displayed. 4- "Account with same name already exist", message will be displayed. 5- User will be successfully registered and redirected to Dashboard tab.	TC1
2	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Incorrect email or password" message will be displayed. 3- User will be successfully logged in and switched to Dashboard page.	TC2
3	1- A password reset email will be sent to the user. 2- "Invalid email address" message will be displayed.	TC3
4	1- "Please enter some text" message will be displayed. 2- Sentiment for each phrase and whole text will be displayed.	TC4

5	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Project with this name already exists" message will be displayed. 3- The project will be created and user will be redirected to Reviews & Mentions tab. 4- "Invalid URL" message will be displayed.	TC5
6	User will receive a welcome email on his provided email address.	TC6
7	User will be redirected to Unsubscribe page.	TC7
8	Reviews and Mentions will be displayed.	TC8
9	Time series, Bar chart and pie chart will be displayed.	TC9
10	A downloadable PDF report will be generated in new tab.	TC10

Table 5.4: Test case output table

5.4.2.5. Environment Needs

5.4.2.5.1. Hardware

Following test cases need hardware for their successful execution:

Hardware Required	Test Case Identifier
System with min 8GB RAM and i5 processor	TC1 - TC10

Table 5.5: Hardware Requirements

5.4.2.5.2. Software

Following are the software required by some use cases for their successful execution.

a) Pycharm Professional Edition
b) Windows or Linux
c) Browser

5.4.2.6. Special Procedural Requirements

There are no other special procedural requirements for the execution of these tests cases.

5.4.2.7. Inter Case Dependencies

Serial Number	Test Case Input	Test case identifier	Test Case Dependency
1	1- Leave all fields blank and click Sign Up button. 2- Enter such an email address that does not exists. 3- Password and Repeat password are not same. 4- User is already registered. 5- Enter valid email address, password and repeat password.	TC1	-
2	1- Leave all fields blank and click Login button. 2- Enter invalid email address or password.	TC2	TC1

	3- Enter valid email address and password.		
3	1- Enter a valid email address. 2- Enter an invalid email address.	TC3	TC6, TC3
4	1- Leave the text box blank. 2- Enter some text.	TC4	-
5	1- Leave all fields blank and click Create Project button. 2- Repeat a project name within same category that is already created. 3- Enter unique project name, choose a category and enter valid url. 4- Enter an invalid url.	TC5	TC1,TC2
6	User creates an account.	TC6	TC1
7	Click unsubscribe in notification email.	TC7	TC1
8	Click on the created project.	TC8	TC1,TC2,TC5
9	Click on the created project and switch to Analysis tab.	TC9	TC1,TC2,TC5
10	Click on Generate Report button.	TC10	TC1,TC2,TC9

Table 5.6: Inter Case Dependencies

5.5. Test Procedure Specification

5.5.1. Purpose

To specify the steps for executing a set of test cases or, more generally, the steps used to analyze a software item in order to evaluate a set of features.

5.5.2. Outline

Following are the contents of test procedure specification:

a) Test procedure specification identifier
b) Purpose
c) Special requirements
d) Procedure steps

5.5.2.1. Test Procedure Specification Identifier

The identifier for the test procedure specification is **PBA_TPSI.** The abbreviation for identifier is as:

PBA Platinum Business Analytics

TPSI Test Procedure Specification Identifier

5.5.2.2. Purpose

This module describes the testing procedure for the entire testing of system. All the tests will be performed manually with in the specified time constraint. It specifies how to execute the testing procedures of a monitoring system. This procedure executes the following test cases: **TC_001- 012.**

5.5.2.3. Special Requirements

To proceed with the testing in web application one should have knowledge of following techniques:

a) Unit testing
b) Pycharm testing frameworks
c) Web Browser

5.5.2.4. Procedural Steps

Following are the procedural steps for the successful execution of tests.

5.5.2.4.1. Log

Logging in the Python can be performed using following techniques:

a) Manual Logging
b) Debug Logging

We will be using both kind of logging depending upon the scenario.

5.5.2.4.2. Set Up

1- Sentiment analysis model is loading successfully.
2- Maintain a fast internet connection.
3- Scraping scripts are working properly.
4- Database connection is successful.
5- User must be registered.
6- Execute the test.

5.5.2.4.3. Start

Write all test cases in pytest and flask-testing in **tests.py** file and run it in terminal using command **python tests.py**.

5.5.2.4.4. Proceed

We have to execute test cases in the order of their dependency. Test case with no dependencies or solved dependency first.

5.5.2.4.5. Measure

A test either will be successful or failed with an error message.

5.5.2.4.6. Shut down

To shut down the testing we can stop it using **Ctrl + C** command in terminal.

5.5.2.4.7. Restart

Clear terminal log and run **tests.py** file in terminal using command **python tests.py.**

5.5.2.4.8. Stop

A test script will stop implicitly after execution or can be terminated explicitly using **Ctrl + C** command.

5.5.2.4.9. Wrap Up

We can run **tests.py** file multiple times using different parameters and note down the log generated by tests scripts.

5.5.2.4.10. Contingencies

A report of contingencies will be prepared based on test log generated by tests scripts.

5.6. Test Item Transmittal Report

5.6.1. Purpose

To identify the test items being transmitted for testing. It includes the person responsible for each item, its physical location, and its status. Any variations form the current item requirements and designs are noted in this report.

5.6.2. Outline

Following are the contents of test case specification:

a) Transmittal report identifier
b) Transmittal items
c) Location
d) Status
e) Approvals

5.6.2.1. Transmittal Report Identifier

The identifier for the test transmittal report is **PBA_TITI**

The abbreviation for identifier is as:

PBA Platinum Business Analytics

TRI Test Item Transmittal Identifier

5.6.2.2. Transmitted Items

All the items which are mentioned in the Test Plan are part of the test transmittal report. We should also provide all those tests which are failed and again revised and developed for refactoring for further testing.

As we are a small team of three members working in this project so, Zeeshan Ahmad and Shaheryar Khan are responsible for test planning and specification development as indicated in Testing Responsibilities. Whereas these test case specifications then will be transmitted to Muhammad Zaryab for successful running of the test.

5.6.2.3. Location

All the test documents including most important test case specification will be placed in a PDF file and will be transmitted to test manager and test executor by email.
After receiving email tester will be able to convert test case specifications into actual written test cases.

5.6.2.4. Status

At this phase all the test cases which are provided are not tested prior so their status in unresolved or so we can say false. After the testing will be done then if there will be any problem in code, a request will be made to the developers to again correct and refactor the code.

5.6.2.5. Approvals

a) Zeeshan Ahmad _____

b) Muhammad Zaryab _____

c) Muhammad Shaheryar Khan _____

5.7. Test Log

5.7.1. Purpose

To provide a chronological record of relevant details about the execution of tests.

5.7.2. Outline

Following will be the content of test log.

a) Test log identifier
b) Description
c) Activity and event entries

5.7.2.1. Test Log Identifier

The identifier for the test log identifier is **PBA_TLI.** The abbreviation for identifier is as:

PBA Platinum Business Analytics

TPSI Test Log Identifier

5.7.2.2. Description

Here are all the test items which are planned to be tested and also are transmitted to the test manager.

Serial No	Test Feature	Test Case Identifier
1	Validations on User Registration	TC1
2	Validations on User Login	TC2
3	Forget password	TC3
4	Sentiment Analysis on user given text	TC4
5	Validations on Project creation	TC5
6	Email notification test	TC6
7	Unsubscribe email notification	TC7
8	Review Analysis	TC8
9	Graphs Analysis	TC9
10	PDF generated report	TC10

Table 5.7: Test Feature Table

Software Requirements
Software requirements which are included are as:

a) Python 3.x
b) Browser

Hardware Requirements
Hardware requirements which are included as:

a) Corei5 CPU
b) 8GB RAM
c) 180 GB SSD
d) GPU recommended

5.7.2.3. Activity and Event Entries

For each event, including the beginning and end of activities, record the occurrence date and time along with the identity of the author. The information in 5.4.2.3 and 5.4.2.4 should be considered.

5.7.2.3.1. Execution Description

While executing the tests whole team was present so that all of the team get through understanding of testing process.
The tests were executed on April 20, 2019 in Python using Flask-testing and pytest with all hardware and software specification already provided.

5.7.2.3.2. Procedure Results

Following is the complete result of our test execution based on the pass/fail criteria already mentioned in planning phase.

Serial Number	Test Case Identifier	Test Revision	Test Case Result	Test Status
1	TC1	1st	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Email address is invalid" message will be displayed. 3- "Repeat password must be same" message will be displayed. 4- "Account with same name already exist", message will be displayed. 5- User will be successfully registered and redirected to Dashboard tab.	Pass
2	TC2	1st	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Incorrect email or password" message will be displayed. 3- User will be successfully logged in and switched to Dashboard page.	Pass
3	TC3	1st	1- A password reset email will be sent to the user.	Pass

			2- "Invalid email address" message will be displayed.	
4	TC4	1st	1- "Please enter some text" message will be displayed. 2- Sentiment for each phrase and whole text will be displayed.	Pass
5	TC5	1st	1- For each blank field a message will be displayed: "Please fill out this field". 2- "Project with this name already exists" message will be displayed. 3- The project will be created and user will be redirected to Reviews & Mentions tab. 4- "Invalid URL" message will be displayed.	Pass
6	TC6	1st	User will receive a welcome email on his provided email address.	Pass
7	TC7	1st	User will be redirected to Unsubscribe page.	Pass
8	TC8	1st	Reviews and Mentions will be displayed.	Pass
9	TC9	1st	Time series, Bar chart and pie chart will be displayed.	Pass
10	TC10	1st	A downloadable PDF report will be generated in new tab.	Pass

Table 5.8: Test Case Result Table

5.7.2.3.3. Environment Information

- Ubuntu 16.04
- RAM 8GB (minimum) (recommended)

5.7.2.3.4. Anomalous Events

Our tests are successful passed at the moment and no anomaly occurred except in one case which will be later investigated. But once developer start writing more code, problems can occur at any time. So, it is necessary to use test driven approach.

5.7.2.3.5. Incident Report Identifiers

Record the identifier of each test incident report, whenever one is generated.

5.8. Test Incident Report

5.8.1. Purpose

To document any event that occurs during the testing process that requires investigation.

5.8.2. Outline

A test incident report shall have the following structure.

a. Test incident report identifier
b. Summary

c. Incident description
d. Impact

5.8.2.1. Test Incident Report Identifier

The identifier for incident report is **PBA_TIRI.**

The abbreviation for identifier is as:

PBA Platinum Business Analytics

TIRI Test Incident Report Identifier

5.8.2.2. Summary

During the execution of test, we have faced an anomaly when one of the test get failed. TC4 which get failed. In this section we will analyze and investigate the reason behind failure of test case.

5.8.2.3. Incident Description

Here is the complete description of the anomalies occurred during the testing phase. Also, a complete solution and refactoring is suggested in the description.

Input	Expected Result	Actual Result	Anomalies
Enter some text.	Sentiment for each phrase and whole text will be displayed.	Sentiment for each phrase and whole text was not displayed.	After investigation we found that the main anomaly was that the system was accepting invalid text.

Date and time	Environment	Attempt to repeat	Tester	Observer
April 20, 2020	Python, flask, Pytorch	After correcting the anomaly, test case is released.	Shaheryar Khan	Zeeshan Ahmad

Table 5.9: Incident Description Table

5.8.2.4. Impact

The incident will have only impact on the test case specification. We will need to rewrite test case specification for the failed test case after sorting out the issue.

5.9. Test Summary Report

5.9.1. Purpose

To summarize the result of the designated testing activities and to provide evaluations based on these results.

5.9.2. Outline

A test summary report shall have the following structure:

a) Test summary report identifier

b) Summary
c) Variance
d) Comprehensive assessment
e) Summary of results
f) Evaluation
g) Summary of activities
h) Approvals

5.9.2.1. Test Summary Report Identifier

Test summary report identifier is as **PBA_TSRI.**

The abbreviation for identifier is as:

PBA Platinum Business Analytics

TSRI Test Summary Report Identifier

5.9.2.2. Summary

All the items described in the test plan are tested and after attestation we figured out that one of them get failed. Except one test case all others test cases passed. We determine passing and failure of test cases according to the criteria mentioned above in test planning phase.
After figuring out that one of the test case is failed we prepared its incidental report so that it can be again refactor by the developer and send again. After refactoring test cases with defects, we again tested it and it passed the testing criteria.

5.9.2.3. Variances

Every test is executed according to the test plan, test specification and test case specification. We follow the testing approach and strategy stated in the test planning phase. Test were executed according to test procedure specification and they are deemed passed and failed according to testing criteria mentioned.

5.9.2.4. Comprehensiveness Assessment

All the features described in the test planning phase are tested and no test case is left out of testing due to any sort of reason. Our test coverage report is more than 80 percent.

5.9.2.5. Summary of Results

All the tests were passed except test with test case id TC4. In TC4 we try to enter some text and detect sentiment for each phrase and whole text to be displayed. But, unfortunately our test did not work as planned. We send the test case back to the test planner and designer. Test planner and designer modify the test case and update transmittal report. After receiving updated transmittal report test manager again execute the revised test case and it get passed.

5.9.2.6. Evaluation

We have covered almost 80 percent code with unit and UI testing. Moreover, we use test driven approach in our code. Out of seventeen test cases only one test case gets failed which was again refactored. So, our team including test planner, test designer and test manager work in great harmony to execute all phases of testing.

5.9.2.7. Summary of Activities

There are three development phases in testing. The first one was planning followed by test specification and designing and test execution. All these test activities and phases are planned and executed according to Gantt chart. Here is the breakdown of all the activities with their time and staff who was responsible.

Testing activities	Person Responsible	Time Elapsed
Test Planning	Zeeshan Ahmad	7 Days
Test Designing	Muhammad Zaryab	8 Days
Test Execution	Muhammad Shaheryar Khan	8 Days

Table 5.9: Activity Summary Table

5.9.2.8. Approvals

We need approval of test planner, test designer and test manager.

a) Zeeshan Ahmad _____

b) Muhammad Zaryab _____

c) Muhammad Shaheryar Khan _____

Chapter 6

User Help Manual

6.1. General information

We are living in an era where there is a lot of competition. The technology has changed the business. Businesses spend a lot of money on analyzing their consumers. Everyone wants to win the race. They monitor the reviews or response of the consumers about a product to make future business decisions. They analyze the reasons why a certain product is getting popularity or disliked so that they can make decisions about those products. They require a user-friendly visualizations and categorized reviews to understand the consumer response.
Sentiment analysis is a process of analyzing the text and more specifically reviews of products or about some kind of topic. People think that there are only three kinds of sentiments positive, negative and neutral. But we can expand it to graded scale where we can analyze and divide the sentiment into furthermore categories like negative, partially negative, neutral and etc. Our purposed system "Platinum Business Analytics" performs sentiment analysis and analyzes the reviews of a movie or product into five categories i.e. Positive, Partially Positive, Neutral, Partially Negative and Negative. In order to provide user more accurate results. User will be able to see the graphs and reviews and can visit to that website to which that review belongs.

6.1.1. System Overview

Platinum Business Analytics is a web application. The proposed system is focused on the sentiment analysis of a topic, movie or product. User can select a specific category and get the most relevant results. The main goal of this project is to improve the accuracy of the sentiment analysis with a graded scale. The scale divides sentiments into five categories: Negative, Partially negative, Neutral, Partially positive and Positive. User will be provided with the interface where he can register. User can create a project and can visit the project any time by logging into his account. There will be a section for analytics that will provide the visualizations like bar graphs and pie charts to better understand the data.

6.2. System Summary

System Summary section provides a general overview of the system. The summary outlines the uses of the system's hardware and software requirements, system's configuration, user access levels and system's behavior in case of any contingencies.

6.2.1. System Configuration

Platinum Business Analytics can run on a computer system with any web browser (Microsoft edge, Google chrome, Firefox, Opera, and Safari) and a minimum of 8GB RAM. A strong connection to internet is necessary for scraping data. We can easily use the website by entering URL in the address bar of browser window.

6.2.2. User Access Levels

Any user who wants to see reviews of a product (AMAZON) or movie (ROTTEN TOMATOES) can use our website. After opening website by entering URL, get registered by providing the credentials and clicking "Register" button from Home Page. After creating account, login in to our website.

6.2.3. Contingencies

In case there is no Internet connection, website will not be displayed.

6.3. Getting started

Getting Started section explains how to reach website and use it.

6.3.1. Home Page

Home Page or Landing Page is displayed when user enters URL in the address bar of the browser window.

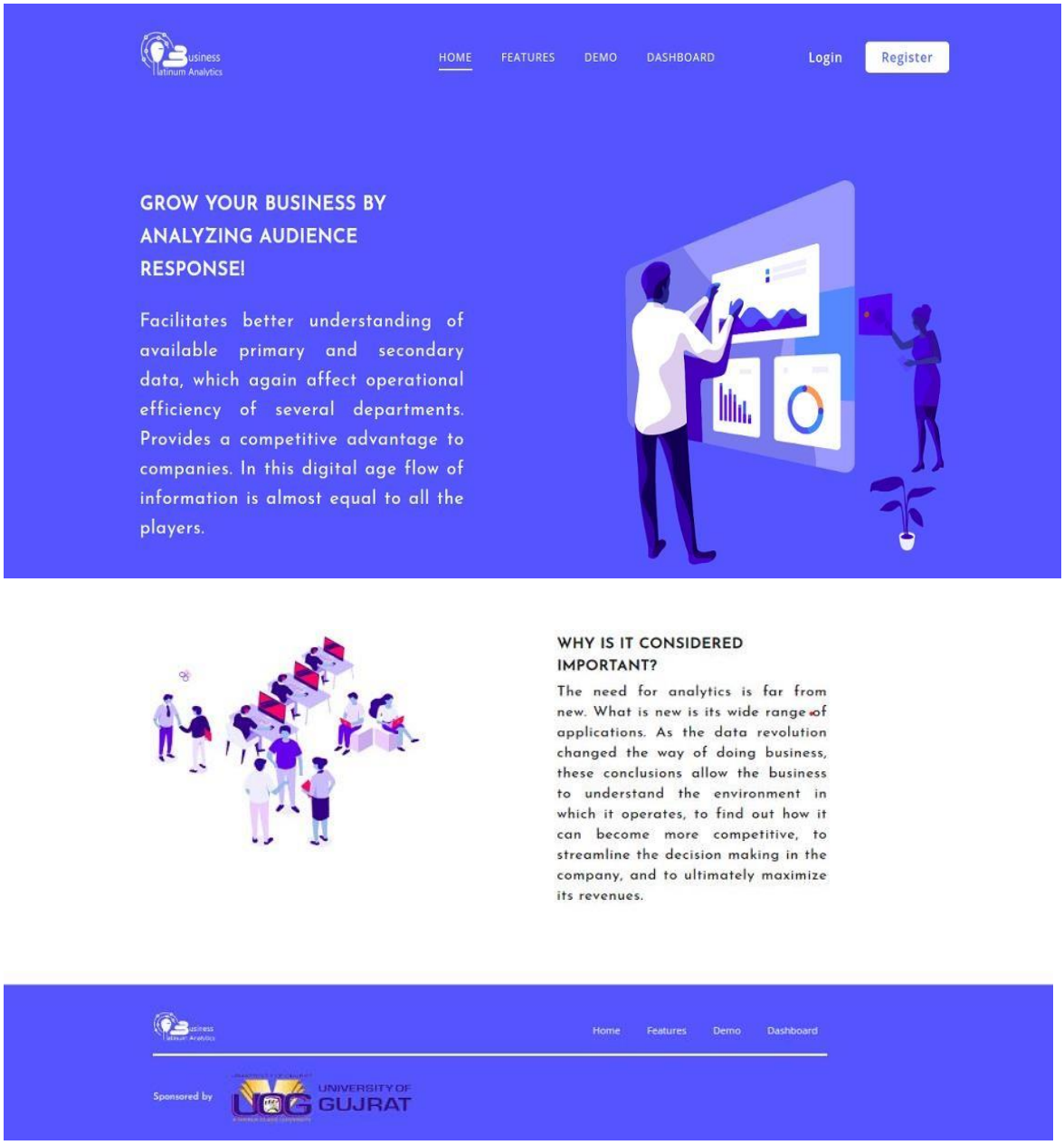


Figure 6.1: Home Page

6.3.2. Features Page

Features Page is displayed when user clicks on Features tab. It shows different features of the website.

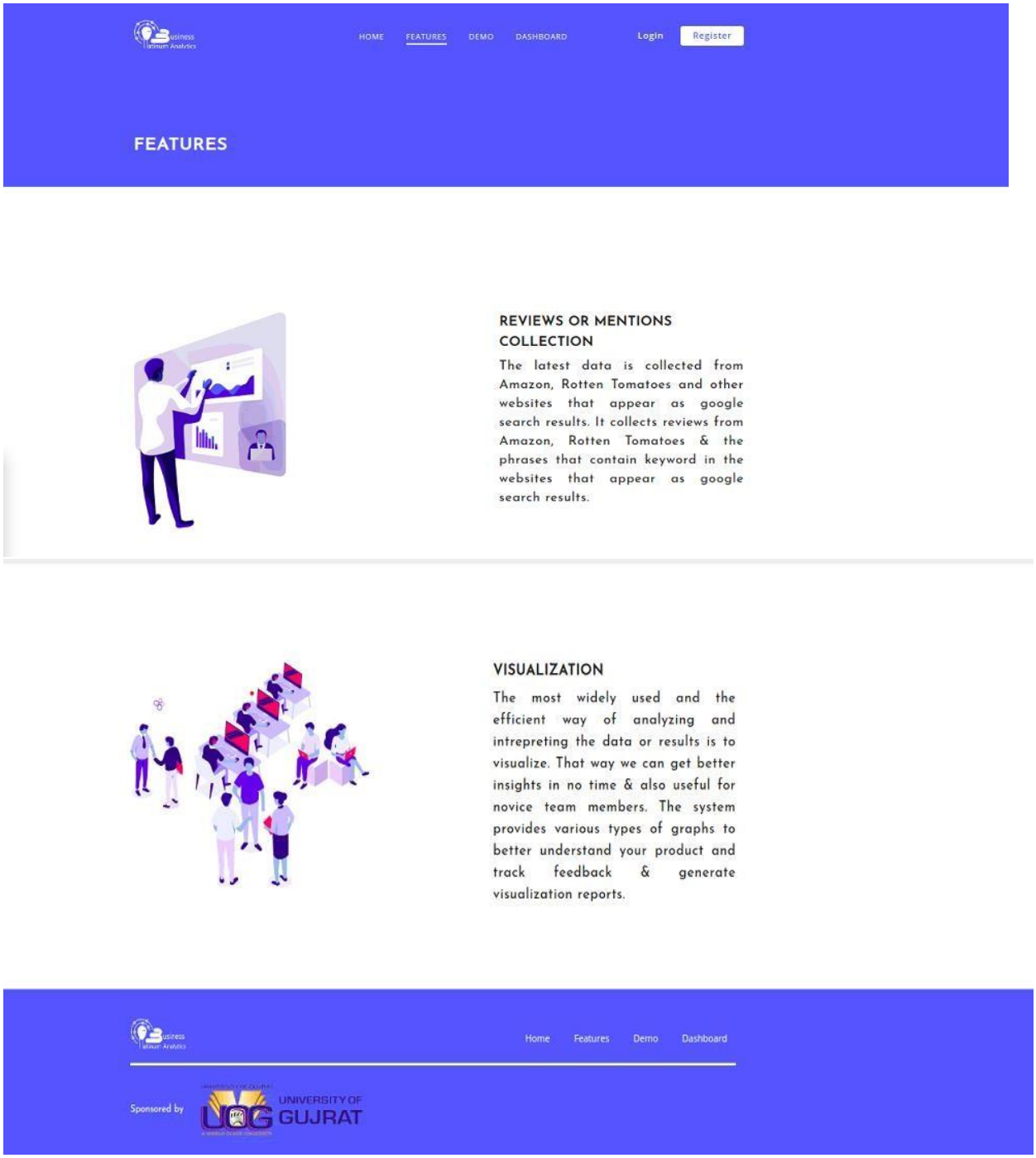


Figure 6.2: Features Page

6.3.3. Demo Page

If you want to check the sentiment of a particular piece of text, then place the text in the provided text box and click on Analyze button.

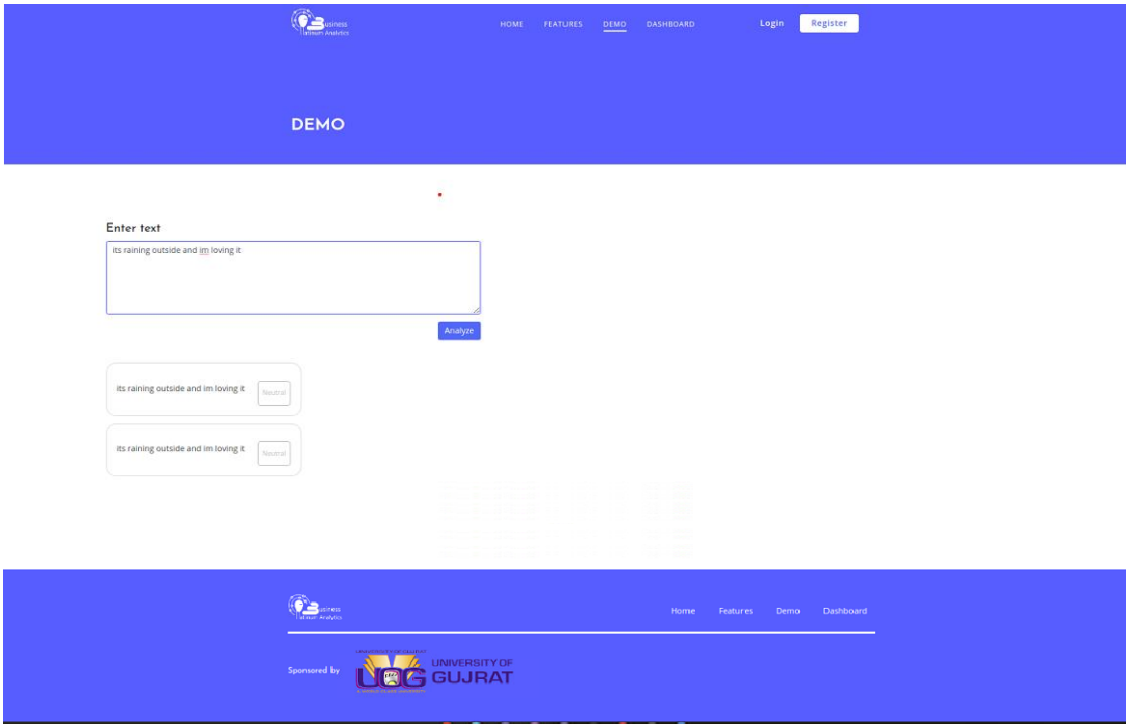


Figure 6.3: Demo Page

6.3.4. Register Page

Register Page is displayed when user clicks on Register button from Home page.

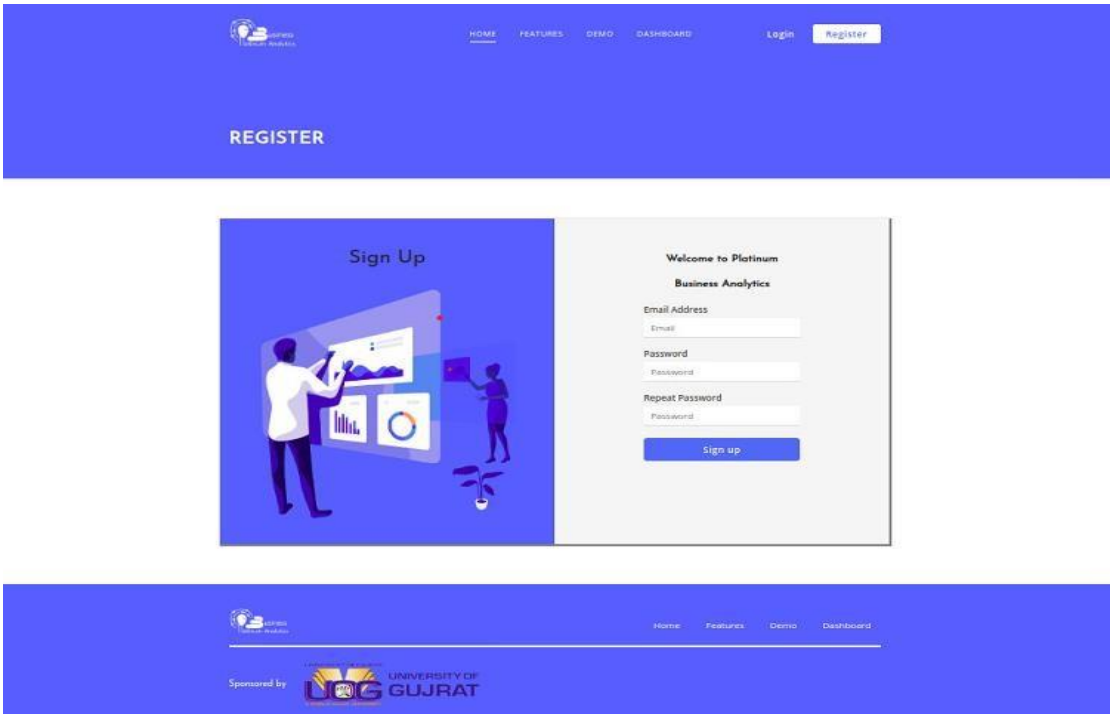


Figure 6.4: Register Page

6.3.5. Login Page

Login Page is displayed when user clicks on Login button from Home page. User can login whenever he/she wants.

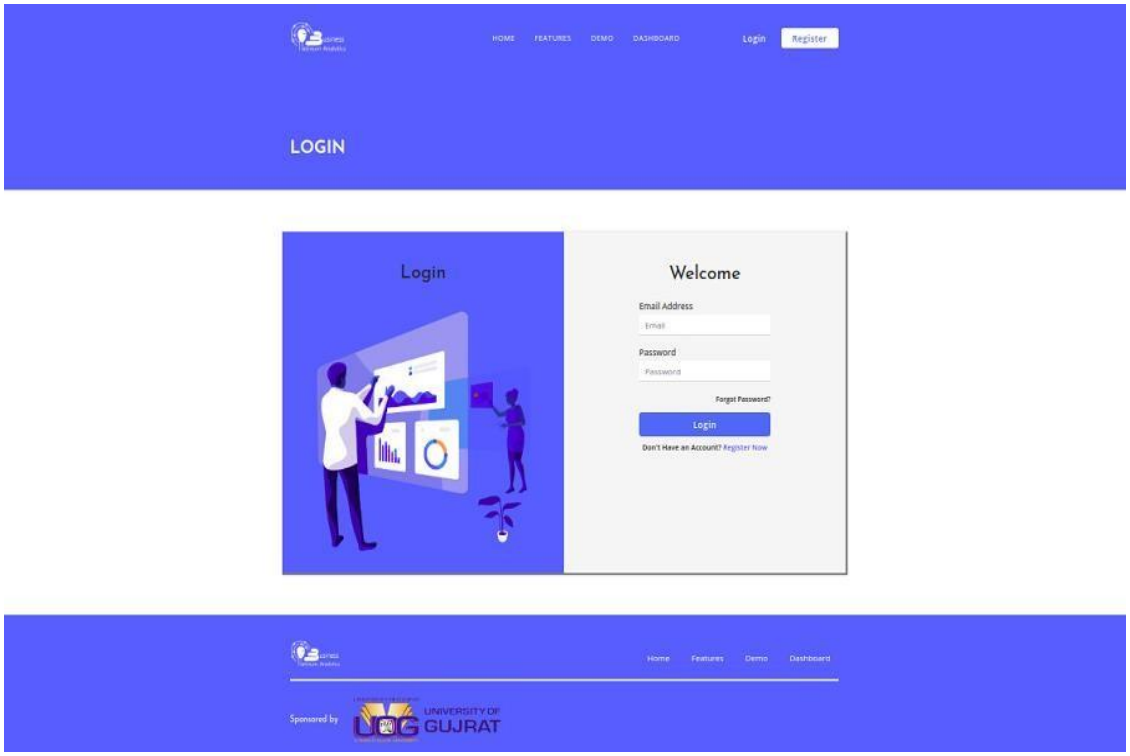


Figure 6.5: Login Page

6.3.6. Reset Password Page

Reset Password Page is displayed when user clicks on Forgot password in the Login Page

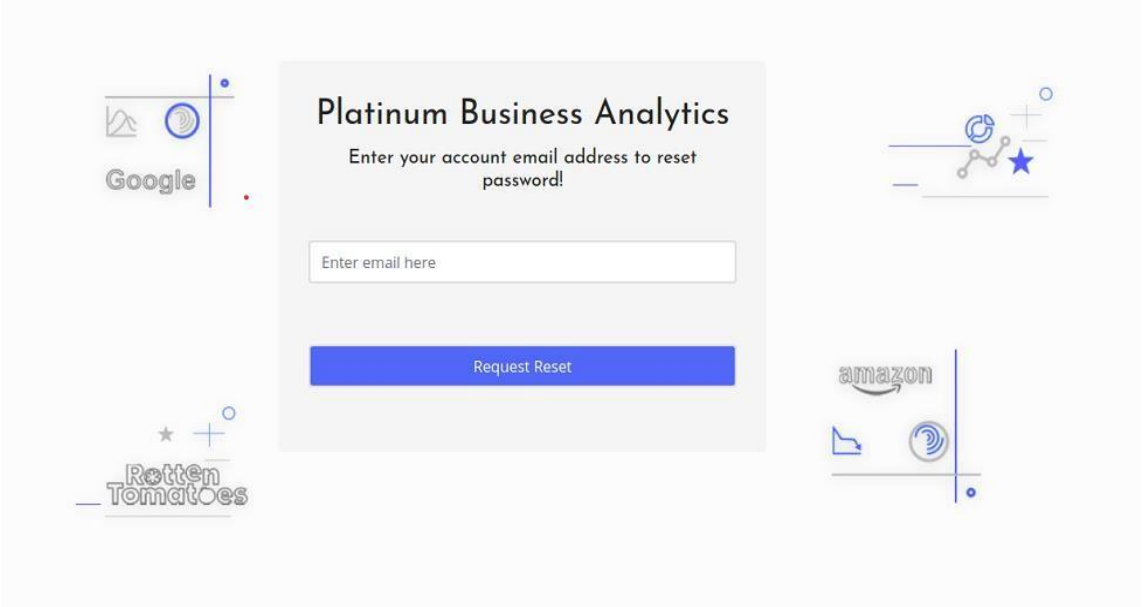


Figure 6.6: Reset Password Page

6.3.7. Dashboard Page

Dashboard Page is displayed when user is successfully logged in to the website. Here a list of projects will be shown that user has already created.

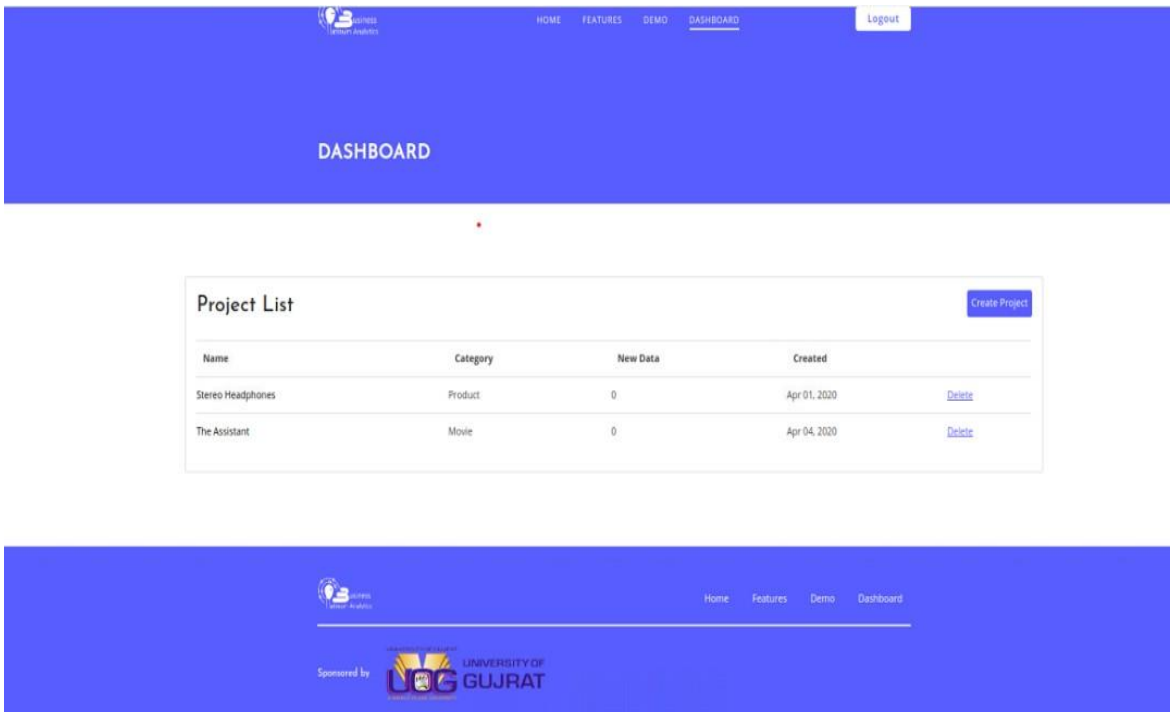


Figure 6.7: Dashboard Page

6.3.8. Project Creation Page

Project Creation Page is displayed when user clicks on Create Project button in the Dashboard Page. Here user can create a project of a specific category (movie, product, general).

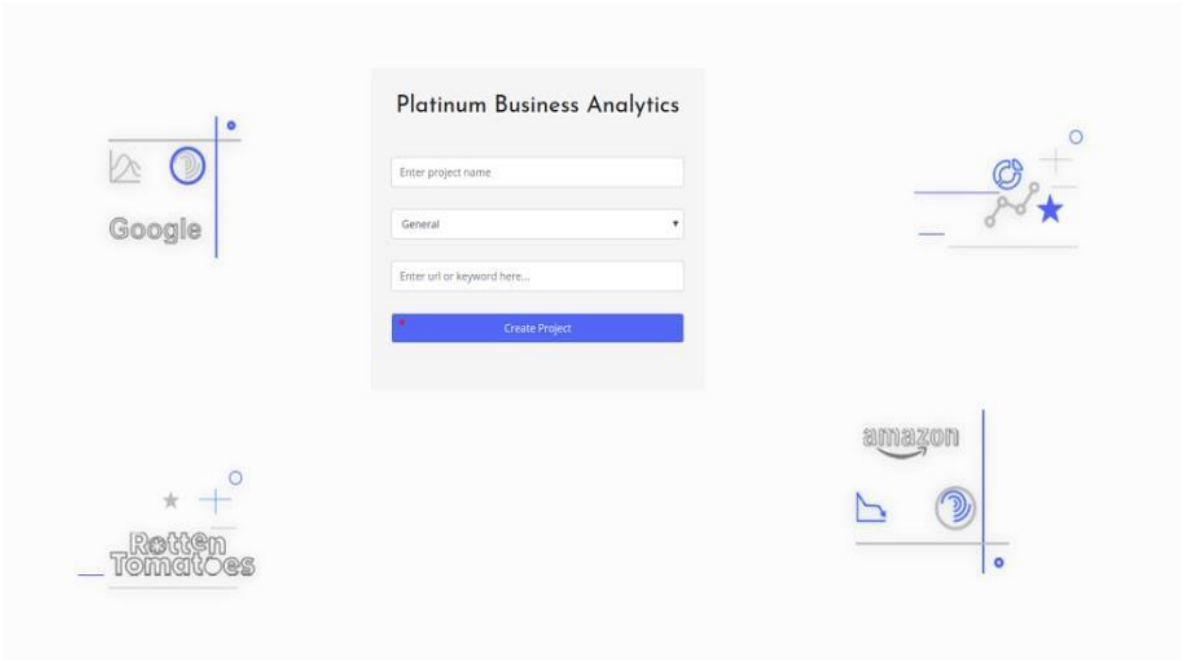


Figure 6.8: Project Creation Page

6.3.9. Reviews & Mentions Page

Reviews & Mentions Page is displayed when user clicks on any of the created projects in Dashboard Page. Here user can see the reviews given by people related to that particular movie or product. We can also filter reviews of a particular class by using slider bar on the right side.

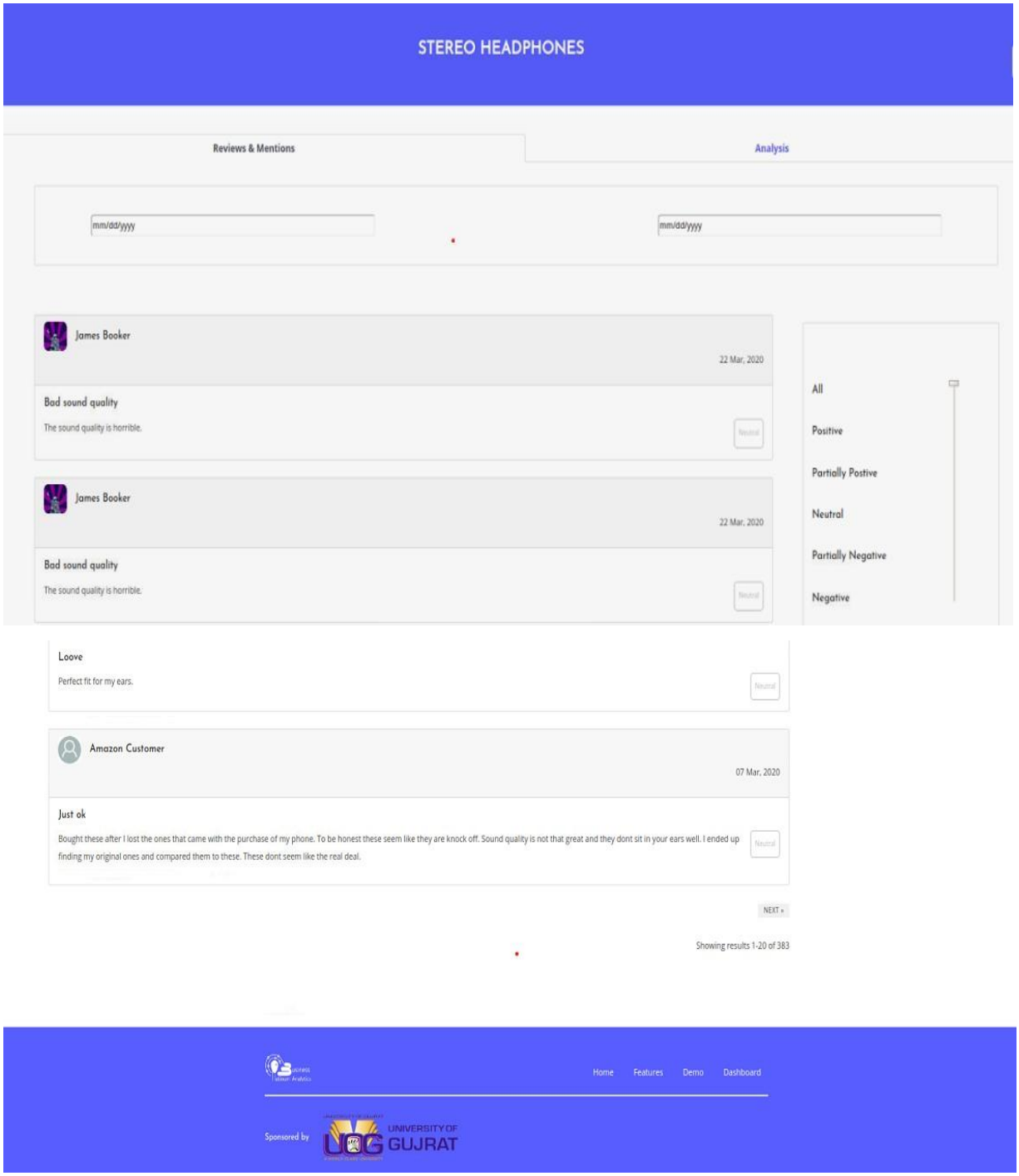


Figure 6.9: Reviews & Mentions Page

6.3.10. Analysis Page

Analysis Page is displayed when user clicks on Analysis tab from Reviews & Mentions Page. Here Time series graph, Bar chart and Pie chart will be shown as well as user can generate a PDF report containing visualization of scraped results i.e. graphs.

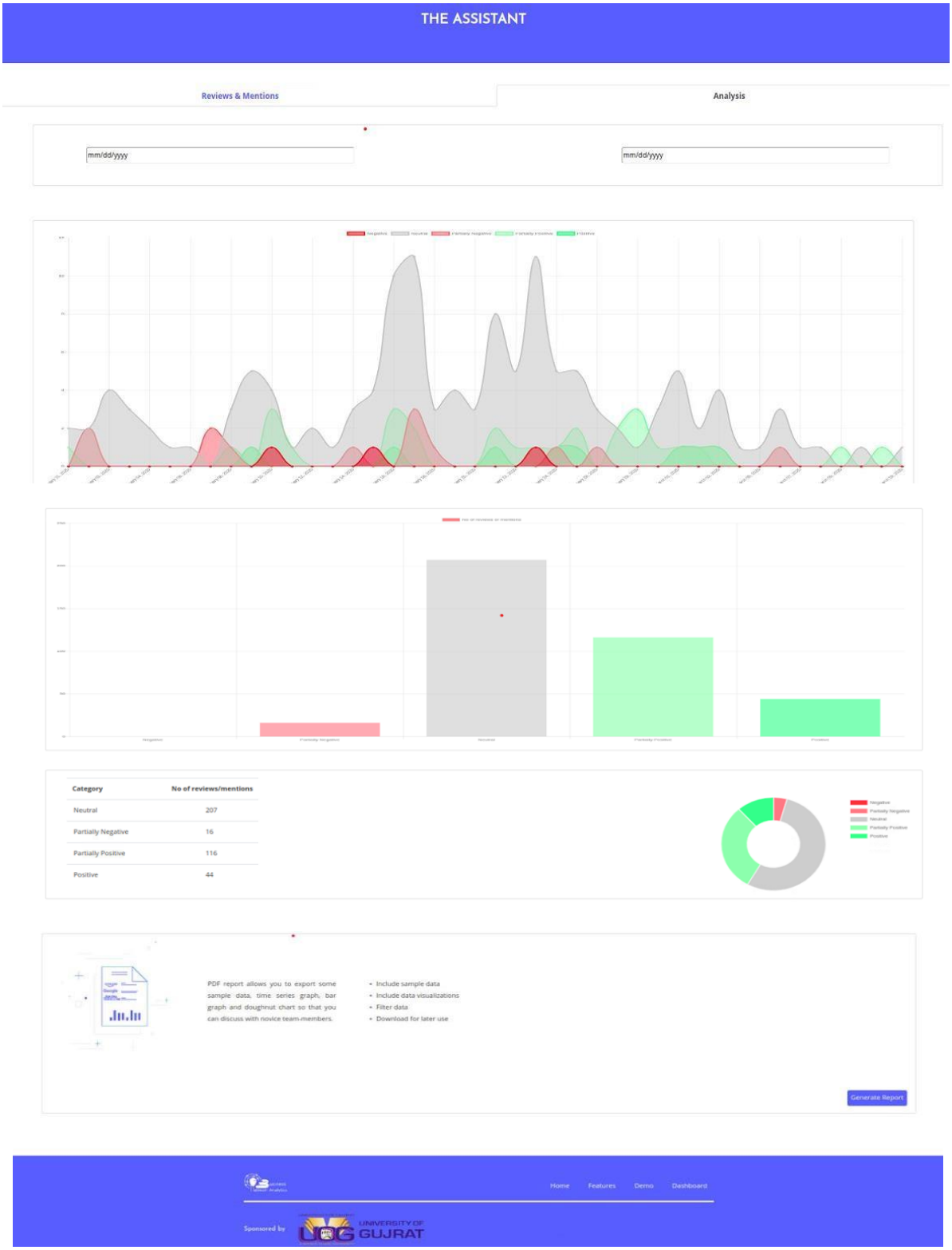


Figure 6.10: Analysis Page

6.3.11. 404 Error Page

In case if wrong URL is entered by user then this page will be displayed.

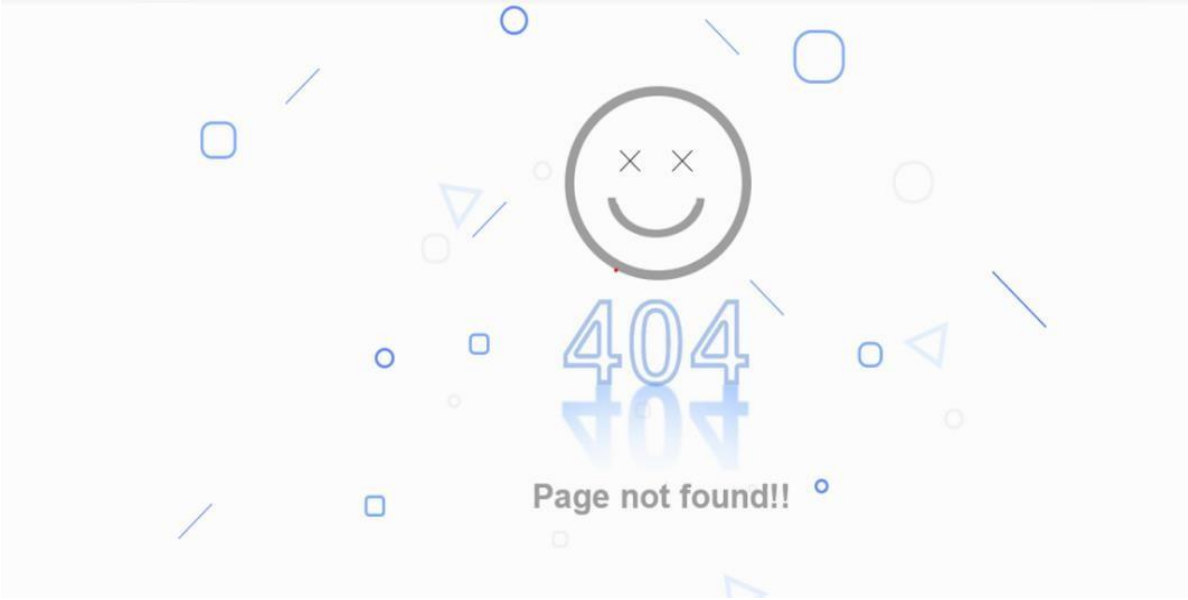


Figure 6.11: 404 Error Page

6.3.12. Unsubscribe Email Notification Page

When user unsubscribe email notifications from Gmail then this page will be displayed.

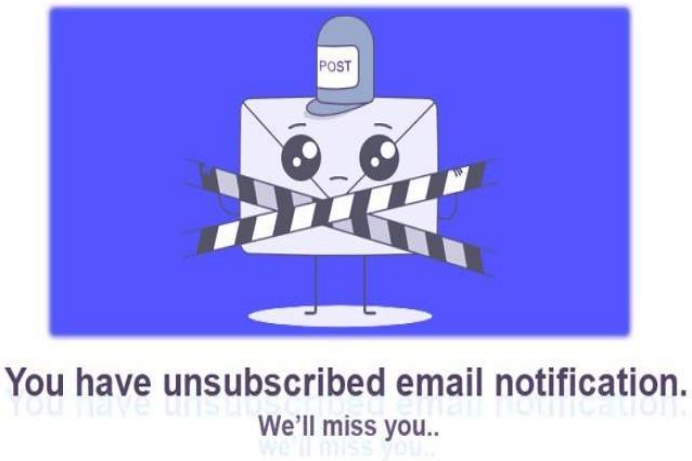


Figure 6.12: Unsubscribe Email Notification Page

Results

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	76.55	49.83	64.73	44.14
Random Forest	73.53	39.10	68.05	34.06
Decision Tree	73.70	47.28	51.10	44.98
SVM	65.67	37.03	50.75	34.64
KNN	74.90	47.69	63.87	41.59
Adaboost	71.94	38.32	54.29	34.10
MLP / BPA	78.93	57.96	64.50	54.09
Naive Bayes	72.95	39.09	50.29	35.31
LSTM	80.20	60.64	68.88	56.90
BERT	83.50	-	-	-

Appendix

- **What is Fine Grained Sentiment Analysis?**
  Fine-grained sentiment analysis refers to the detection of **sentiment**, not on the document or post level, but rather on the sentence, sub-sentence (phrase/clause), or even aspect-based level. In fine grained sentiment analysis more than two classes are used for sentiment classification.

- **What is Pytest?**
  A robust Python testing tool, **Pytest** can be used for all types and levels of software testing. **Pytest** is a software test framework, which means **Pytest** is a command-line tool that automatically finds tests you've written, runs the tests, and reports the results.

- **What is API?**
  An application programming interface (**API**) is a computing interface which defines interactions between multiple software intermediaries. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc.

- **What is Keras?**
  Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

- **What is Gensim?**
  Gensim is an open-source library for unsupervised topic modeling and natural language processing, using modern statistical machine learning. Gensim is implemented in Python and Cython.

- **What is MongoDB?**
  MongoDB is a cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemas. MongoDB is developed by MongoDB Inc. and licensed under the Server Side Public License.

- **What is Flask?**
  Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.