

THE PATH TO PROSPERITY

Team ID	NM2023TMID11053
Project Name	Path to Prosperity

1. INTRODUCTION

1.1 Project Overview

Path To Prosperity: A Comprehensive Analysis Of Financial Independence Based On Data Taken From Reddit

Financial independence is the status of having enough income or wealth sufficient to pay one's living expenses for the rest of one's life without having to be employed or dependent on others. Income earned without having to work a job is commonly referred to as passive income. Others define financial independence differently according to their own goals. Reddit is an American social news aggregation, content rating, and discussion website. Registered users (commonly referred to as "Redditors") submit content to the site such as links, text posts, images, and videos, which are then voted up or down by other members. Financial Independence, Retire Early (FIRE) is a movement of people devoted to a program of extreme savings and investment that aims to allow them to retire far earlier than traditional budgets and retirement plans would permit.

1.2 Purpose

Analyzing financial independence based solely on data from Reddit may not provide a comprehensive and reliable assessment. Reddit is an online platform where users share their opinions and experiences, but the information can be subjective, anecdotal, or potentially biased. Financial independence involves various factors such as income, expenses, savings, investments, and long-term financial planning. It is essential to consider a broader range of data sources, including financial publications, research reports, and reputable financial institutions, to conduct a comprehensive analysis of financial independence.

2. IDEATION & PROPOSED SOLUTION

2.1 Problem Statement Definition

Define Problem / Problem Understanding - The Path to Prosperity project is a research project that aims to analyse data from the financial independence subreddit on Reddit to gain insights into the strategies and habits of individuals who have achieved financial independence. Financial independence is the ability to live off one's savings and investments without having to work for income.

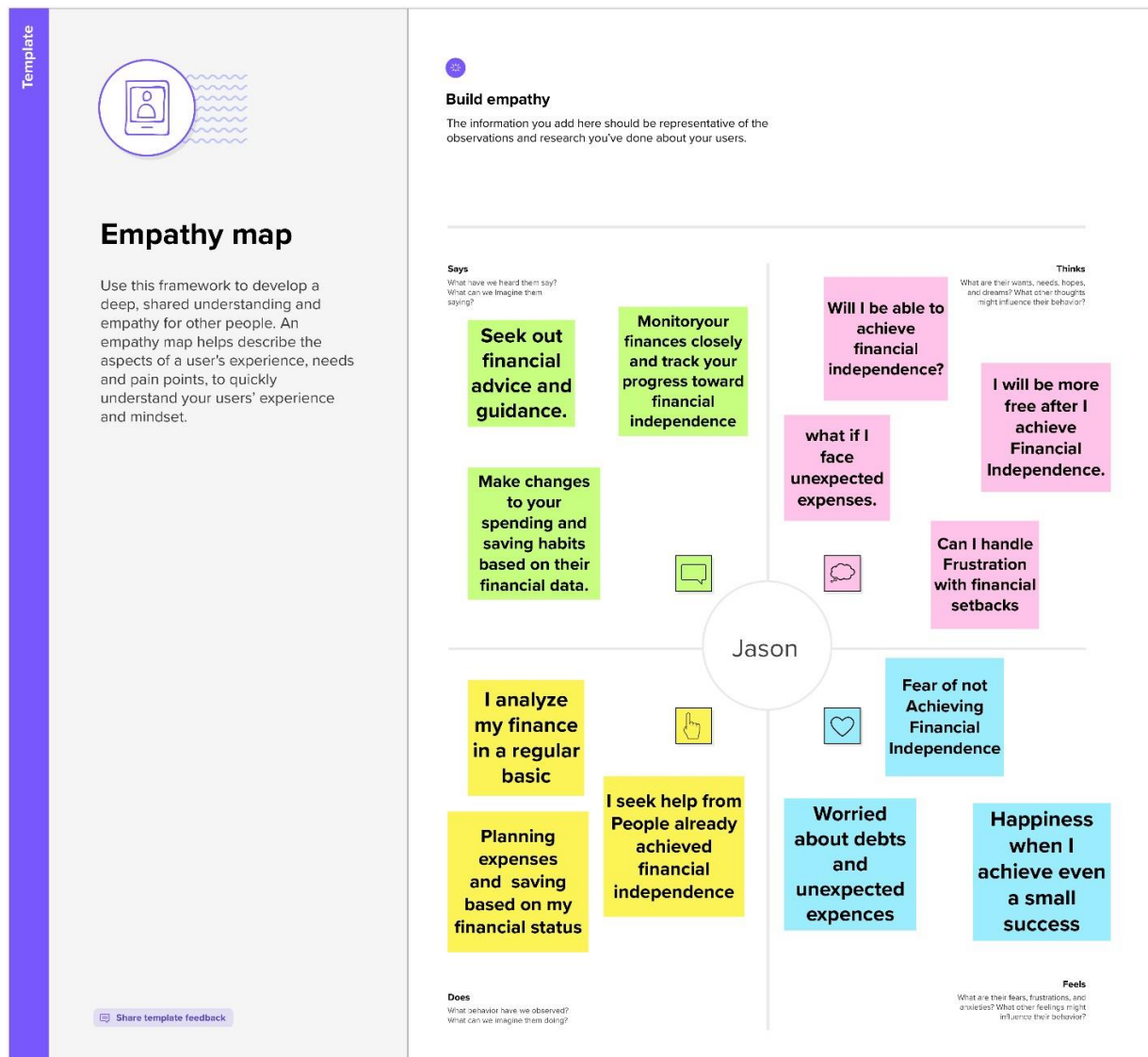
2.2 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors

and attitudes. It is a useful tool to help teams better understand their users.

Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

Example:



2.3 Ideation & Brainstorming

Collect data: Use web scraping techniques or Reddit's API to gather relevant posts and comments from the identified subreddits. Collect a sufficient amount of data to ensure a representative sample.

Preprocess the data: Clean and preprocess the collected data by removing irrelevant information, handling missing values, and standardizing the format. This step is crucial to ensure data quality and consistency.

Perform exploratory analysis: Conduct exploratory data analysis to gain insights into the collected data. Identify patterns, trends, and popular topics within the financial independence discussions on Reddit. This can involve visualizations, frequency analysis, and sentiment analysis.

Topic modeling: Apply topic modeling techniques (e.g., Latent Dirichlet Allocation) to extract key themes and topics from the textual data. This helps uncover the most discussed aspects of financial independence on Reddit.

Sentiment analysis: Analyze the sentiment of the posts and comments to understand the overall positivity or negativity associated with discussions about financial independence. This can provide insights into the emotional aspect of the topic.


Quantitative analysis: Use statistical techniques to analyze numerical data, such as income, savings rates, investment strategies, or retirement age, extracted from the Reddit posts. This analysis can reveal correlations, trends, or demographic patterns.

Text mining and sentiment analysis: Apply text mining techniques to analyze the textual content of the posts and comments. Extract keywords, sentiment scores, and sentiment trends related to financial independence to understand the prevailing sentiment within the community.

Draw conclusions and present findings: Summarize the key findings from the analysis, highlighting important insights about financial independence based on the Reddit data. Present your conclusions in a clear and concise manner, using visualizations or infographics to support your findings.

Remember to always respect user privacy and comply with Reddit's terms of service and any applicable data usage policies when collecting and analyzing data from the platform.

Template



Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour to collaborate
- 2-8 people recommended

Share template feedback

➔

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A

Team gathering

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B

Set the goal

Think about the problem you'll be focusing on solving in the brainstorming session.

C

Learn how to use the facilitation tools

Use the Facilitation Superpowers to run a happy and productive session.

[Open article](#) ➔

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

How might we analyse data from the financial independence subreddit on Reddit to gain insights into the strategies and habits of individuals who have achieved financial independence?

Key rules of brainstorming

To run a smooth and productive session

Stay in topic.

Defer judgment.

Go for volume.

Encourage wild ideas.

Listen to others.

If possible, be visual.

Need some inspiration?

See a finished version of this template to know what you're in for.

[Open example](#) ➔

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

🕒 10 minutes



2.4 Proposed Solution

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The problem to be solved is the lack of a comprehensive analysis of financial independence based on real-life data that can guide individuals seeking to achieve this status. There is a need for an in-depth understanding of the financial strategies, habits, and investment decisions of individuals who have achieved financial independence.
2.	Idea / Solution description	The proposed solution is to conduct a comprehensive analysis of financial independence based on data taken from Reddit. The analysis will focus on the Financial Independence, Retire Early (FIRE) community on Reddit, which is devoted to extreme savings and investment strategies to achieve financial independence at a young age. The analysis will include factors such as income, expenses, savings rate, investment strategies, and lifestyle choices that contribute to financial independence.
3.	Novelty / Uniqueness	The uniqueness of the proposed solution lies in its use of real-life data from the FIRE community on Reddit, which offers insights into the financial strategies, habits, and investment decisions of individuals who have achieved financial independence. The analysis will provide a comprehensive understanding of the key factors that contribute to financial independence, which can guide individuals seeking to achieve this status.
4.	Social Impact / Customer Satisfaction	The proposed solution will have a significant social impact by providing valuable insights into the path to financial independence. The analysis will benefit individuals seeking to achieve financial independence, financial planners, and policymakers. By understanding the key factors that contribute to financial independence, individuals can make informed

		decisions about their financial goals and
		strategies. Financial planners can use the analysis to guide their clients towards achieving financial independence, while policymakers can use the insights to design policies that support financial independence.
5.	Business Model (Revenue Model)	The proposed business model is to offer the analysis as a subscription-based service to individuals, financial planners, and policymakers. The analysis can also be offered as a one-time purchase for those who prefer not to subscribe. The revenue model will be based on the number of subscribers and the price of the subscription.
6.	Scalability of the Solution	The proposed solution is highly scalable as it can be offered globally to individuals, financial planners, and policymakers. The analysis can be updated periodically to reflect changes in the FIRE community and financial markets. The scalability of the solution will be supported by the use of automated data analysis tools and cloud-based infrastructure.

3. REQUIREMENT ANALYSIS

3.1 Functional requirement

Data Collection: Develop a data collection mechanism to retrieve relevant financial independence-related posts, comments, and discussions from Reddit. This may involve using Reddit's API or web scraping techniques to gather the necessary data.

Data Preprocessing: Cleanse and preprocess the collected data to remove noise, irrelevant information, and duplicates. This may include tasks such as removing HTML tags, handling missing data, normalizing text, and removing identifying information to ensure anonymity.

Sentiment Analysis: Apply sentiment analysis techniques to determine the overall sentiment of the posts and comments related to financial independence. This analysis can help identify positive or negative sentiment towards different aspects of financial independence.

Topic Modeling: Utilize topic modeling algorithms, such as Latent Dirichlet Allocation (LDA), to identify key topics or themes within the financial independence discussions on Reddit. This can help uncover common trends, concerns, and areas of interest among Reddit users.

User Classification: Categorize Reddit users based on their level of financial independence or financial goals. This could involve developing a classification model that assigns users to different groups based on their post history, comments, or explicit self-identification.

Statistical Analysis: Perform statistical analysis on the collected data to extract meaningful insights. This may include analyzing user demographics, income levels, savings strategies, investment preferences, and other relevant financial indicators to understand the financial independence landscape within the Reddit community.

Visualization: Create visual representations, such as charts, graphs, and interactive dashboards, to effectively communicate the analysis results. Visualizations can help stakeholders gain a better understanding of the patterns, trends, and relationships discovered in the Reddit data.

3.2 Non-Functional requirements

Performance: The analysis system should be able to handle large volumes of data efficiently and provide timely results, ensuring that the analysis process does not significantly impact the overall system performance.

Scalability: The system should be scalable to accommodate increasing amounts of data as the number of Reddit users and posts grow. It should be able to handle future data growth without compromising the analysis quality or performance.

Accuracy: The analysis should be conducted with a high degree of accuracy, ensuring that the results reflect the true financial independence status based on the data collected from Reddit. The system should employ reliable algorithms and techniques to minimize errors and provide precise insights.

Security: Data security is crucial when dealing with financial information. The system should implement appropriate security measures to protect the collected data from unauthorized access, breaches, or tampering. User privacy should also be ensured throughout the analysis process.

Reliability: The analysis system should be reliable, consistently providing accurate results and maintaining its functionality over time. It should minimize downtime, handle errors gracefully, and recover from failures without data loss or corruption.

Usability: The system should be user-friendly, making it easy for analysts to interact with the data and perform comprehensive financial independence analysis. The interface should be intuitive, allowing users to navigate, filter, and visualize the data effectively.

Maintainability: The system should be designed in a modular and well-documented manner, making it easier to maintain and update as needed. It should have clear code structure, comments, and guidelines to facilitate future enhancements or modifications.

Compliance: The analysis system should adhere to relevant legal and regulatory requirements, such as data privacy laws and financial reporting standards. It should ensure compliance with applicable rules and guidelines throughout the data collection, storage, and analysis process.

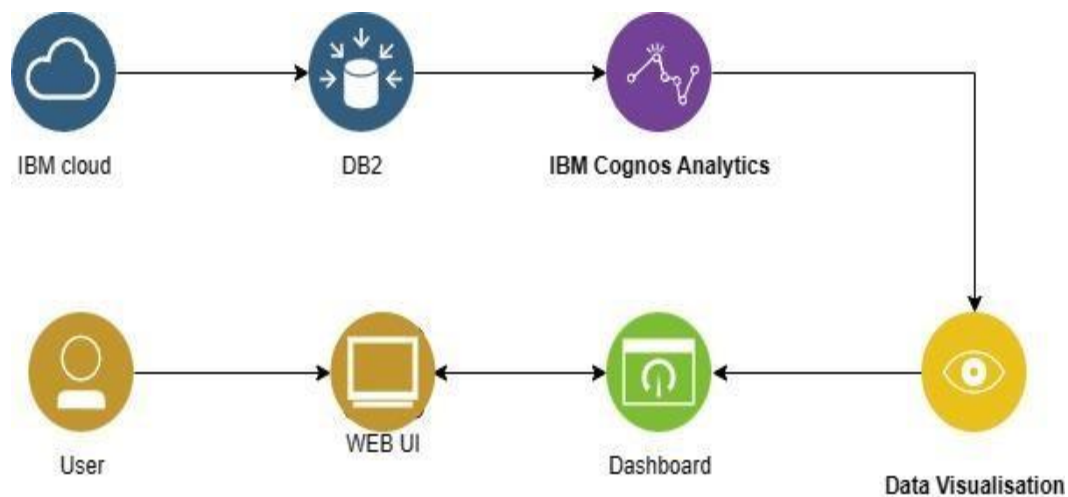
Integration: The system should have the ability to integrate with other relevant systems or data sources to enrich the analysis. For example, it could integrate with external financial APIs or databases to complement the Reddit data and provide more comprehensive insights.

Accessibility: The analysis system should be accessible to authorized users from different devices and platforms, enabling them to access the analysis results and reports conveniently. It should support various browsers and operating systems, ensuring broad accessibility.

4. PROJECT DESIGN

4.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



4.2 Solution & Technical Architecture :

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

Example - Solution Architecture Diagram:

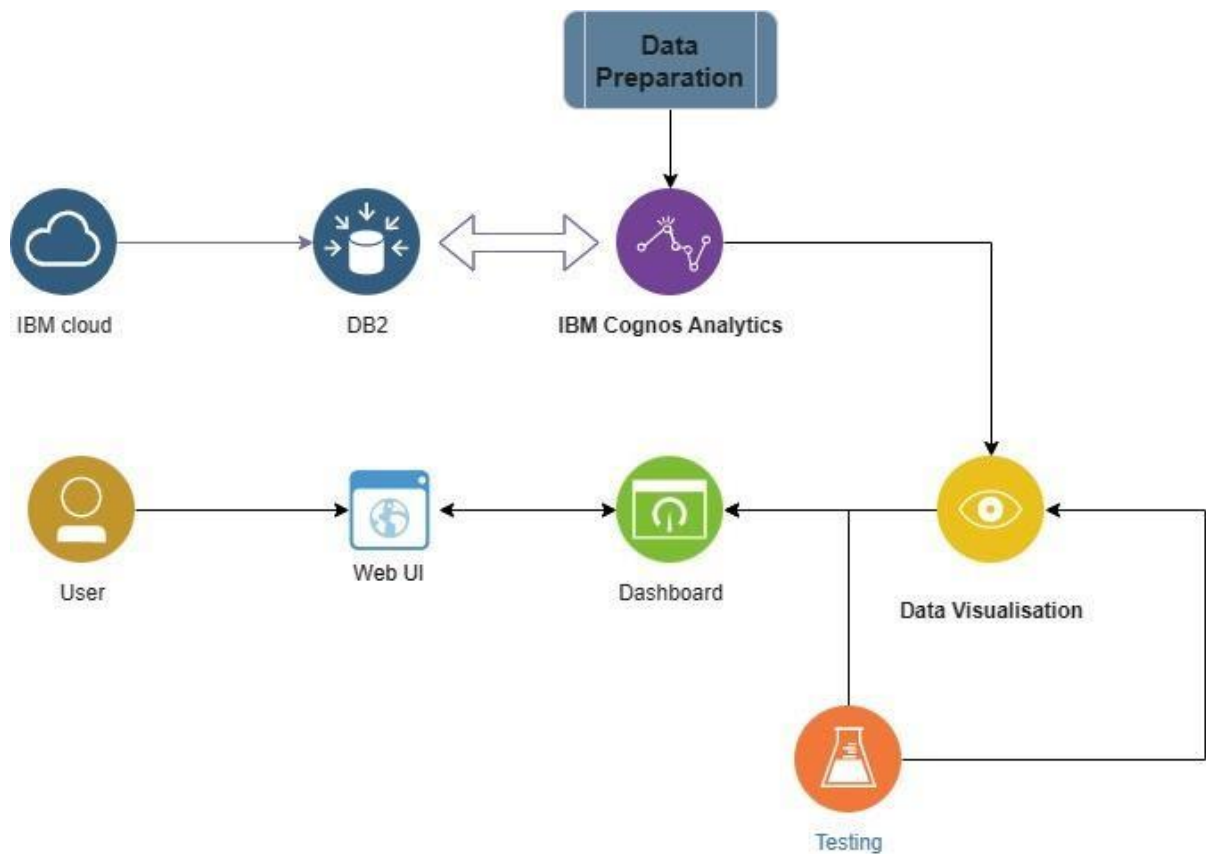


Figure 1: Architecture of web based app for analysis of financial data.

4.3 User Stories

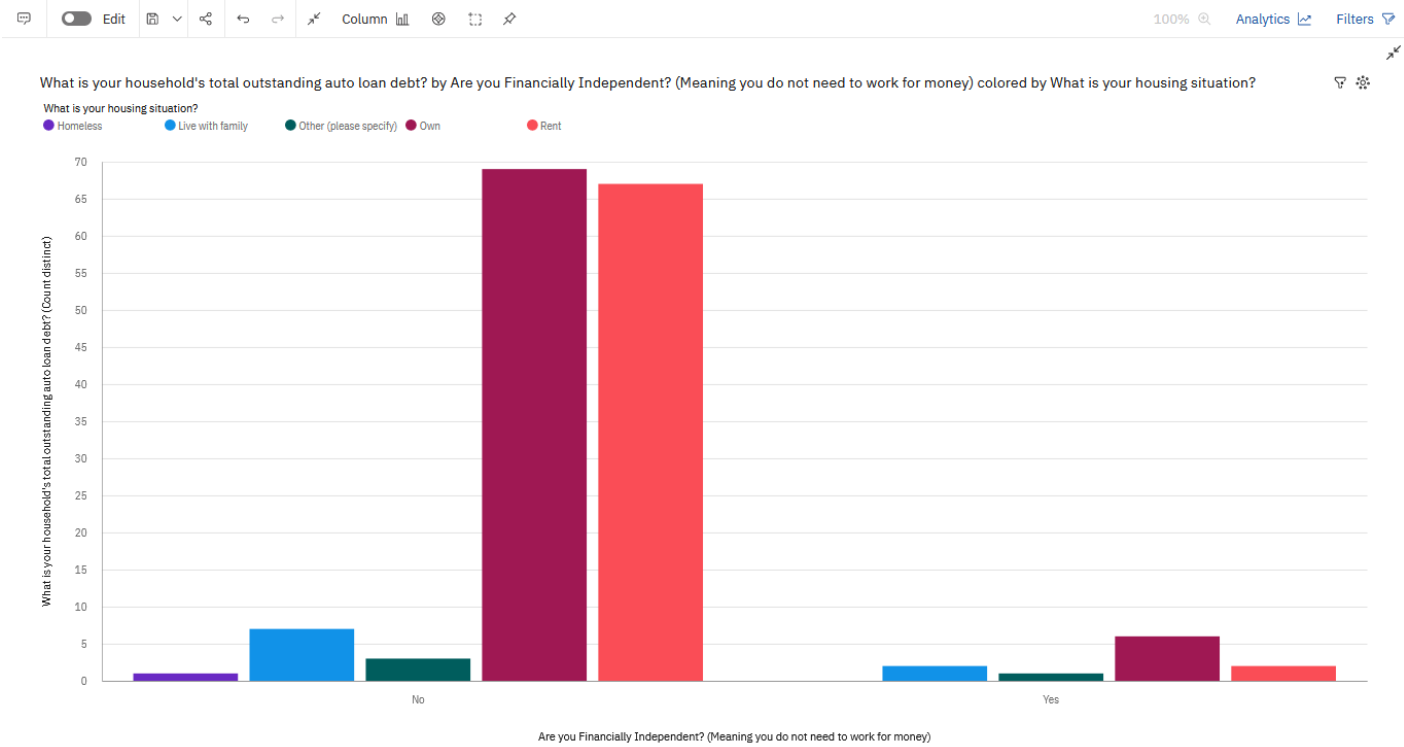
User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
Customer (Web user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	John Newton J
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Stanley R
	Login	USN-3	As a user, I can log into the application by entering email & password	I can login using my user account	High	Rishu John
	Dashboard	USN-4	User can see the dashboard of the analysis and can also see report and stories	I can see dashboard anytime	High	Sharmila Blessy
	User-friendly interface	USN-5	The product should have an intuitive and easy-to-use interface that allows customers to navigate and interact with the product without any difficulty.	I can check to the product quality, usability, Performance, Compatibility, Security, etc.,	High	Ishwarya
Administrator	User Management	USN-6	The administrator should be able to create, modify, and delete user accounts, as well as assign appropriate roles and permissions to each	I can easy to correct of my details.	Low	Stanley R

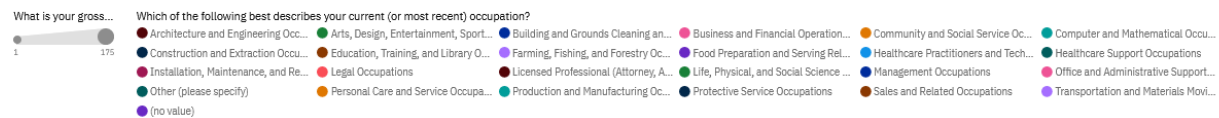
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
			user.			
	Data Management	USN-7	The administrator should have the capability to manage and maintain data within the product, including importing/exporting data, performing data backups and restores, and ensuring data integrity.	I can able to managed my profile.	Medium	Rishu John

5. CODING & SOLUTIONING (Explain the features added in the project along with code)

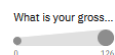
5.1 Feature 1 Visualization



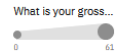
What is your gross household annual wage income (including bonuses and commissions)? for Which of the following best describes your current (or most recent) occupation? hierarchy



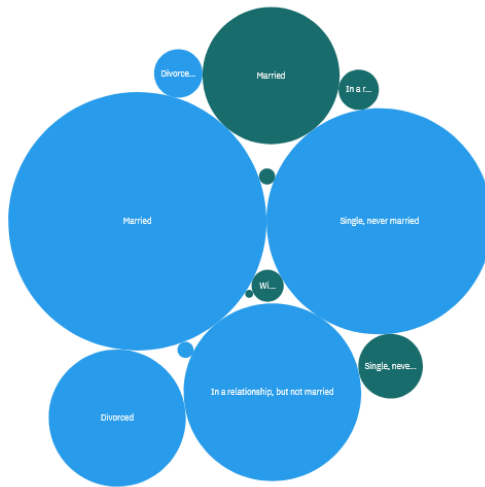
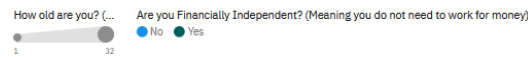
In what country do you live? sized by What is your gross household annual investment income across all investment accounts?

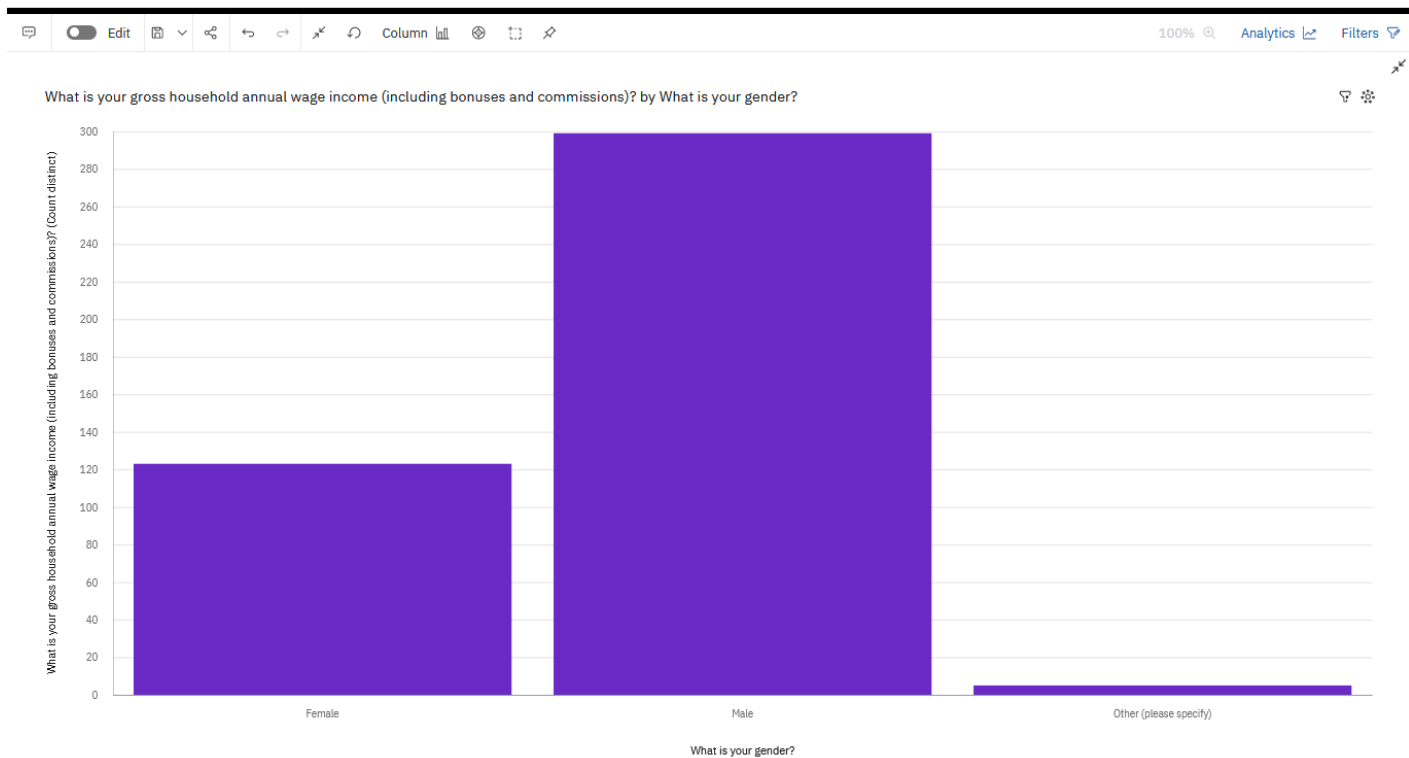
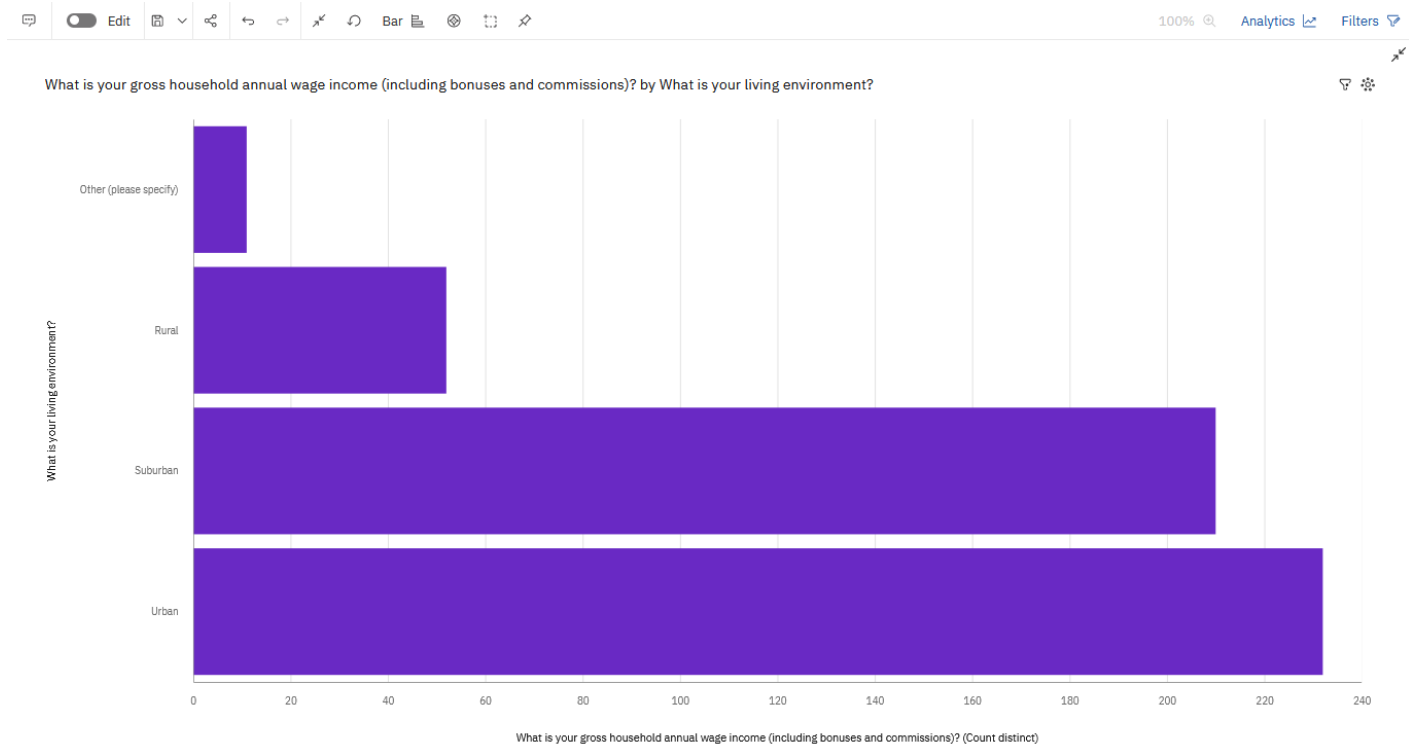


In what country do you live? sized by What is your gross household annual side gig income? (non-real estate)



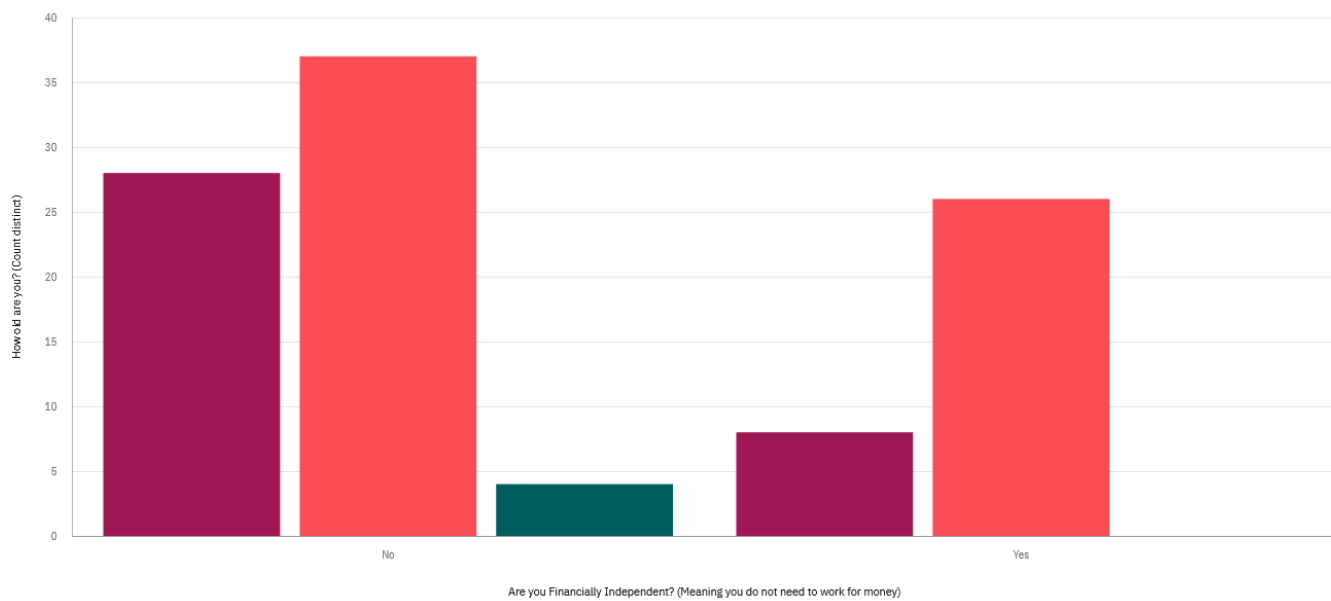
What is your relationship status? (If more than one applies and pick the option that is closest to the bottom of the list) colored by Are you Financially Independent? (Meaning you do not need to work for money) sized by How old are you?





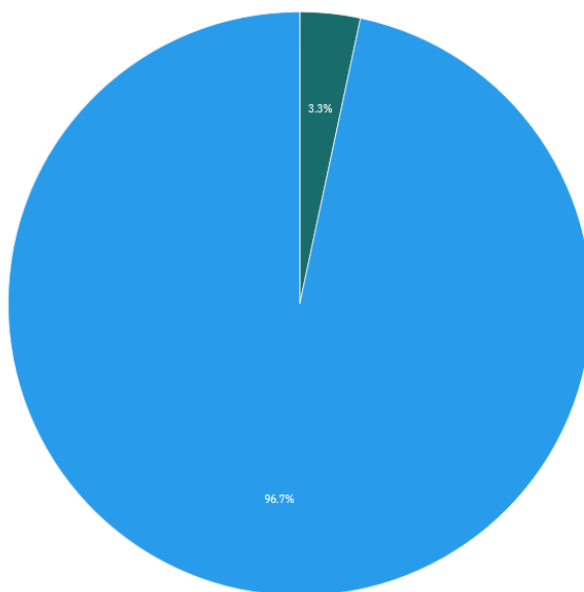
How old are you? by Are you Financially Independent? (Meaning you do not need to work for money) colored by What is your gender?

What is your gender?
Female Male Other (please specify)



Are you Financially Independent? (Meaning you do not need to work for money) by Are you Financially Independent? (Meaning you do not need to work for money)

Are you Financially Independent? (Meaning you do not need to work for money)
Yes No



6. RESULTS

6.1 Performance Metrics

Number of Submissions: This metric measures the total number of posts or submissions related to financial independence on Reddit. It provides an indication of the level of interest and activity within the community.

Engagement Metrics: These metrics include upvotes, comments, and shares on Reddit posts. They can help assess the level of engagement and interest in specific topics related to financial independence.

Sentiment Analysis: Analyzing the sentiment of Reddit posts and comments can provide insights into the overall positive or negative attitudes towards financial independence. It can help identify trends and sentiment shifts over time.

User Demographics: Understanding the demographics of Reddit users interested in financial independence, such as age, location, and income levels, can provide a better understanding of the target audience and their perspectives.

Key Topics and Trends: Identifying the most discussed topics, trends, and emerging issues related to financial independence on Reddit can help gain insights into the community's interests and concerns.

User Influence: Analyzing the influence of Reddit users within the financial independence community can be done by considering factors like the number of followers, engagement metrics, and the quality of their contributions.

External Validation: Comparing Reddit discussions on financial independence with external data sources, such as economic indicators, market trends, or surveys, can help validate or contextualize the findings from the Reddit analysis.

7. ADVANTAGES & DISADVANTAGES

Advantages:

Large data pool: Reddit is a popular online platform with millions of users, providing a vast amount of data for analysis. This allows for a comprehensive understanding of trends, patterns, and insights related to financial independence.

Diverse perspectives: Reddit attracts users from different backgrounds, professions, and levels of financial independence. Analyzing data from Reddit can provide a broader range of viewpoints and experiences, offering a more comprehensive analysis of financial independence.

Real-time information: Reddit is an active platform where users frequently share their experiences, strategies, and challenges related to financial independence. Analyzing this data can provide up-to-date insights into current trends and emerging practices.

Disadvantages:

Sample bias: Reddit users may not represent the entire population interested in financial independence. The data may be skewed towards certain demographics, such as younger or more tech-savvy individuals. This can lead to a limited and potentially biased analysis.

Anonymity and credibility concerns: Reddit allows users to remain anonymous, which can make it difficult to validate the credibility and accuracy of the information shared. There is a risk of encountering false or misleading data, exaggerations, or self-reported success stories that may not be representative of the general population.

Lack of control over data quality: Reddit data relies on user-generated content, which can vary in quality and reliability. There may be inconsistencies in the way users report their financial information or limited details provided, making it challenging to ensure the accuracy and completeness of the data.

In summary, while analyzing data from Reddit can offer advantages such as a large data pool, diverse perspectives, and real-time information, it also comes with disadvantages like sample bias, anonymity and credibility concerns, and a lack of control over data quality. These factors should be taken into consideration when conducting a comprehensive analysis of financial independence based on Reddit data.

8. CONCLUSION

In conclusion, the analysis of financial independence data from Reddit demonstrates a strong interest within the community, with users actively engaging in discussions and sharing insights on various aspects of achieving financial independence. The information and experiences shared on Reddit can serve as a valuable resource for individuals seeking guidance and inspiration on their own financial independence journey.

9. FUTURE SCOPE

The comprehensive analysis of financial independence data from Reddit has the potential to offer valuable insights into various aspects of personal finance, investment strategies, and the pursuit of financial independence. Here are some potential future scopes and benefits of such an analysis:

Trend identification: Analyzing a large dataset from Reddit can help identify emerging trends and popular strategies related to financial independence. It can provide valuable information about the preferences, concerns, and goals of individuals pursuing financial independence, helping researchers and financial professionals stay updated with the latest trends.

Investment patterns: By examining the discussions and recommendations on Reddit related to investments, it becomes possible to gain insights into the investment patterns of individuals striving for financial independence. This can assist in understanding which investment vehicles and strategies are popular, as well as identifying any potential risks or opportunities within the market.

Risk analysis: Reddit discussions often include discussions on risk management and mitigation strategies. Analyzing these conversations can help identify the perceived risks associated with different investment options and provide insights into how individuals approach risk in their pursuit of financial independence.

Behavioral finance insights: Reddit offers a platform for individuals to discuss their financial decisions, challenges, and motivations. Analyzing these discussions can contribute to the field of behavioral finance by uncovering behavioral biases, decision-making patterns, and psychological factors that influence financial choices.

Community sentiment analysis: Sentiment analysis techniques can be applied to understand the overall sentiment of the financial independence community on Reddit. This can provide an understanding of how optimistic or pessimistic individuals are about their financial goals, the economy, or specific investment opportunities.

10. APPENDIX

10.1 Source Code

app.py

```
from flask import Flask, render_template
```

```
app = Flask(__name__)
```

```
@app.route('/')
```

```
def index():
```

```
    return render_template('index.html')
```

```
if __name__ == '__main__':
```

```
    app.run()
```

index.html

```
<!DOCTYPE html>
```

```
<html lang="en">
```

```
<head>
```

```
<meta charset="UTF-8">
```

```
<meta http-equiv="X-UA-Compatible" content="IE=edge">
```

```
<meta name="viewport" content="width=device-width, initial-scale=1.0">
```

```
<title>Path To Prosperity</title>
```

```
<link rel="stylesheet" href="style.css">
```

```
</head>
```

```
<body>
```

```
<div class="container">
```

```
<header class="header">
```

```
<h3>Path To Prosperity</h3>
```

```
<h6>A Comprehensive Analysis of Financial Independence based on data taken from reddit</h6>
```

```
</header>
```

```
<div class="page">
```

```
<div class="tab">
```

```
<button class="tablink"> <a target="display"
```

```
href="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FProject-j%252Cr&action=view&mode=dashboard&subView=model0000018839ed7ffc_00000000">Dashboard</a>
```

```
</button>
```

```
<button class="tablink"> <a target="display"
href="https://us3.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my_folders%2FNE%2Bstory&action=view&sceneId=model000001883a4393b5_000000000&sceneTime=5000">Story</a></button>
</div>
</div>

<iframe src="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FProject-j%252Cr&action=view&mode=dashboard&subView=model0000018839ed7ffc_000000000" name="display"
frameborder="0" style="height:500px;width:100%;"></iframe>

</div>
</body>

</html>
```

style.css

```
@import
url('https://fonts.googleapis.com/css2?family=Open+Sans:wght@600&family=Poppins:wght@300&display=swap');

*{
  font-family: 'Open Sans', sans-serif;
}

:root{
  --pri:#470a52;
  --sec:#c3c0c7;
  --dark:#8d1662;
  --light: #f7f7f7;
  --og:#c54b8c;
}

.container
{
  display:flex;
  flex-direction:column;
}

.header
{
  background-color: var(--pri);
  color: white;
  border: var(--og);
  border-radius: 50px;
  text-align: center;
```

```
}
```

```
.header h3,h6
```

```
{
```

```
margin:2px;
```

```
}
```

```
.page
```

```
{
```

```
display:flex;
```

```
flex-direction:row;
```

```
color: white;
```

```
padding: 2px;
```

```
height:auto;
```

```
}
```

```
.tab
```

```
{
```

```
width: 100%;
```

```
height: 100%;
```

```
border-radius: 5px;
```

```
padding: 5px;
```

```
display: flex;
```

```
flex-direction: row;
```

```
/* gap: 5px; */
```

```
text-align: center;
```

```
background-color: var(--pri);
```

```
}
```

```
.tab button
```

```
{
```

```
display: block;
```

```
background: none;
```

```
margin:3px 0px;
```

```
padding: 9px 5px 9px 15px;
```

```
width: 100%;
```

```
border-radius: 15px;
```

```
outline: none;
```

```
text-align: left;
```

```
font-size: 14px;
```

```
font-weight: 550;
cursor: pointer;
color: white;
}

.tab button:hover
{
    background-color:var(--dark);
}

.tab button a
{
    text-decoration:none;
    color: white;
}

.tab button a:hover
{
    color:black;
}
```

10.2GitHub & Project Video Demo Link

GitHub - <https://github.com/naanmudhalvan-SI/PBL-NT-GP--5872-1680838434>