



MS Project End-Semester Progress Report



PersonaGenius - Smart Clustering for Persona Development

Student's Name: Munnazzah Aslam
ERP: 27229

Supervisor's Name: Dr. Tariq Mehmood
(Digital) Signature of Supervisor:
Date: June 15, 2024

Introduction:

Customer segmentation is central to effective marketing strategies, but conventional methods often fall short in capturing the intricate nuances of diverse consumer bases. Our project addresses this issue by developing an innovative framework for customer persona development using clustering techniques. Traditional demographic-focused approaches provide limited insights into consumer behavior, neglecting crucial psychographic traits, behavioral patterns, and pain points necessary for tailored marketing strategies. By leveraging advanced clustering algorithms, our framework aims to overcome these limitations, enabling businesses to create detailed customer personas that offer a comprehensive understanding of their target audience.

Furthermore, we aim to bridge the gap between theoretical insights and practical application by creating a user-friendly website. This platform will empower users to perform clustering analysis on their own data, facilitating the identification of customer segments and providing actionable insights for personalized marketing strategies. Our project combines research insights with user-friendly tools to modernize customer persona development and enhance targeted marketing strategies in the digital era.

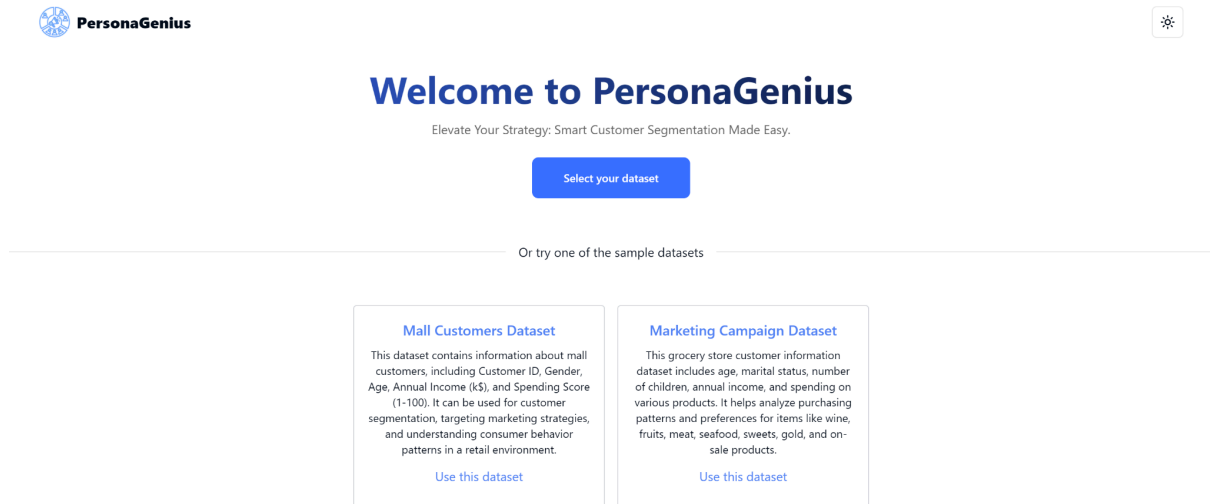
Functionalities of PersonaGenius





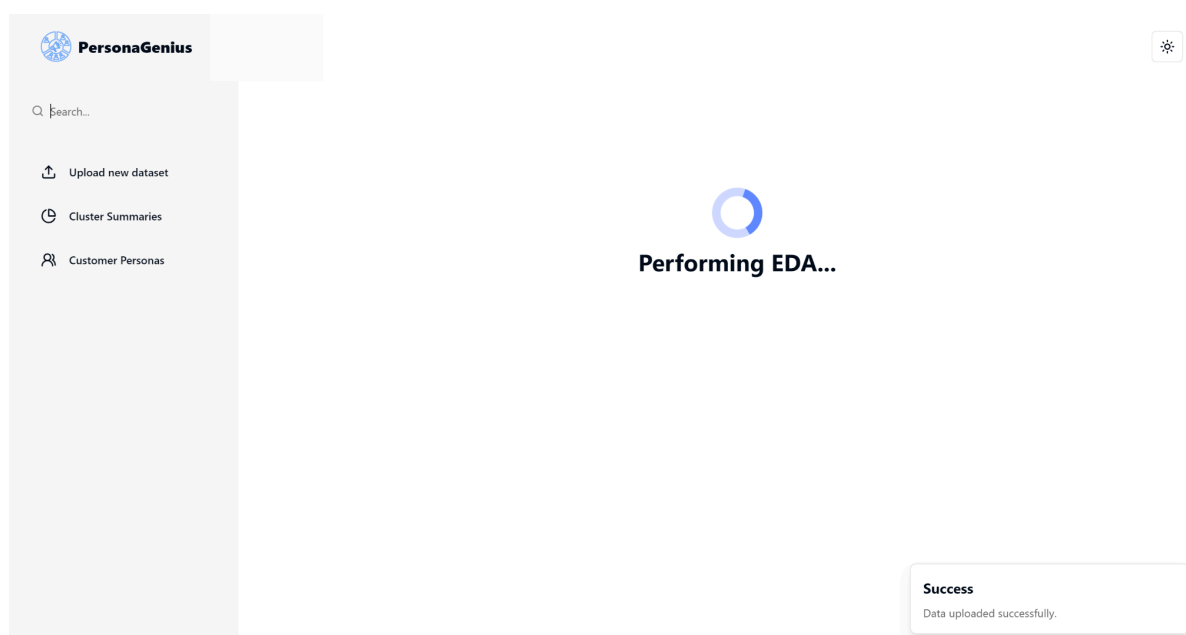
- **Dataset Upload:**

Users can easily upload their own datasets or choose from a selection of sample datasets provided by PersonaGenius.



- **Data Cleaning and Exploratory Data Analysis (EDA):**

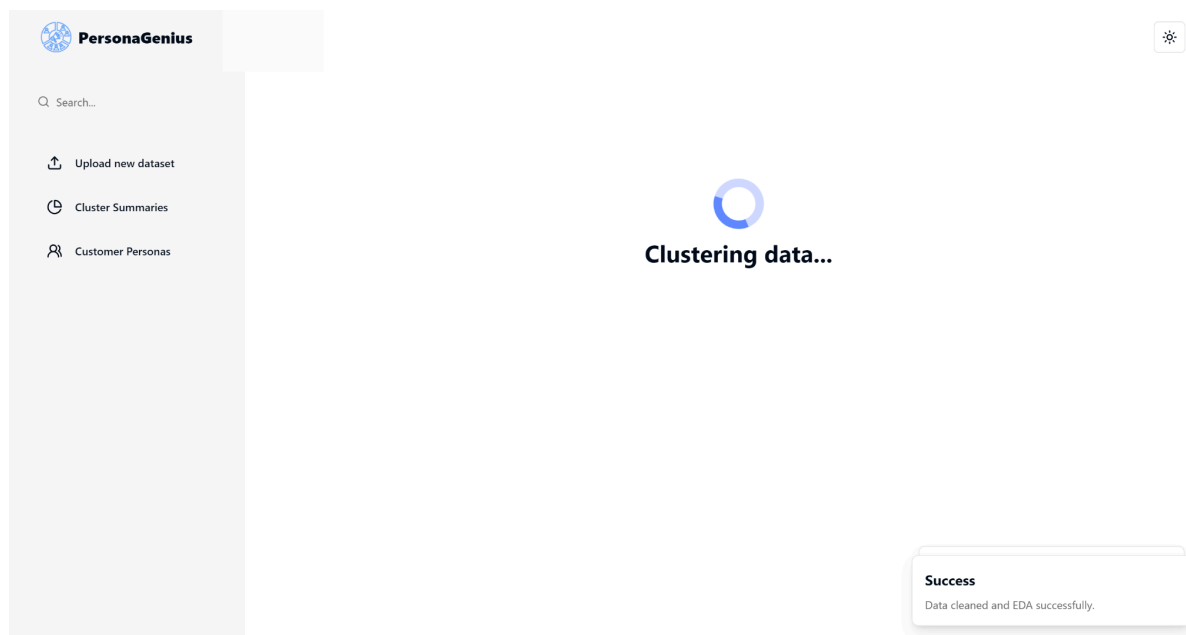
PersonaGenius automatically performs data cleaning tasks to ensure that the uploaded datasets are accurate and free from inconsistencies or errors. The platform then conducts exploratory data analysis to uncover patterns, trends, and relationships within the data. This helps to gain insights into the characteristics of their customer base before proceeding to clustering.





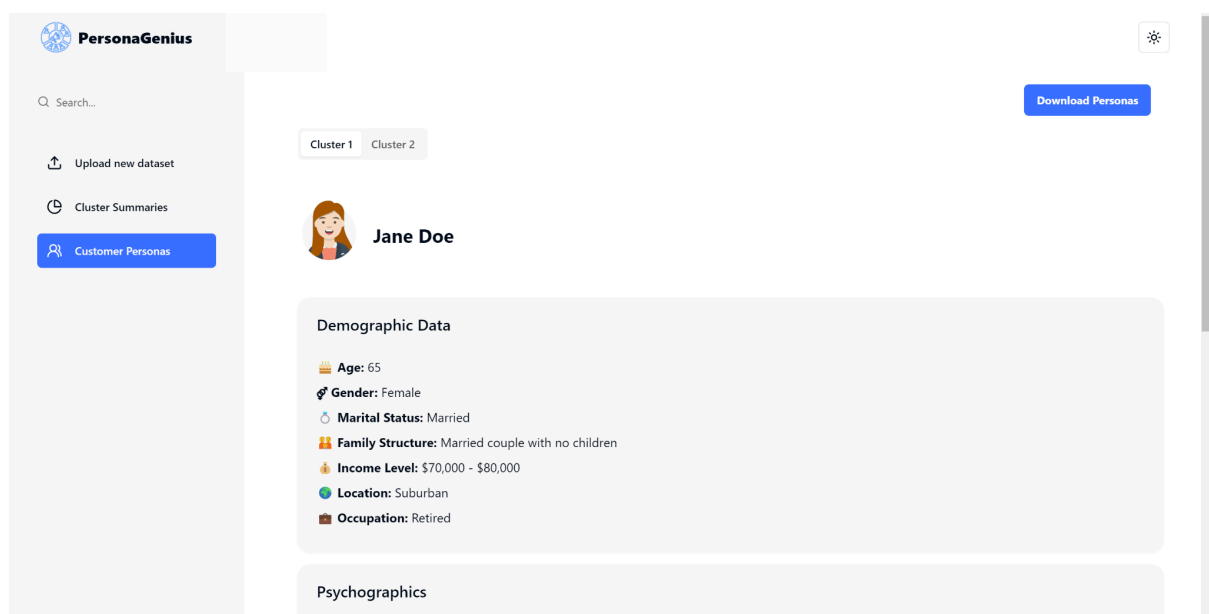
- **Clustering Analysis:**

PersonaGenius utilizes the k-means clustering algorithm to segment the customer data into distinct groups based on similarities in their attributes.



- **Persona Generation:**

After clustering, PersonaGenius generates customer personas by analyzing the clusters and identifying common characteristics, behaviors, and pain points within each segment.





The screenshot shows the PersonaGenius dashboard. On the left is a sidebar with a search bar, an 'Upload new dataset' button, a 'Cluster Summaries' button, and a 'Customer Personas' button. The main content area displays a list of customer personas, each with an icon and a brief description:

- Values and Beliefs:** Appreciates quality and value
- Interests and Hobbies:** Cooking, gardening, traveling
- Lifestyle Choices:** Active and social
- Technology Usage:** Regularly uses internet for shopping and entertainment
- Brand Preferences:** Prefers established brands
- Community Engagement:** Active in local community
- Health and Wellness:** Values healthy living
- Family Dynamics:** Strong family values
- Financial Goals:** Financial security
- Media Consumption:** Watches TV, reads newspapers, uses social media
- Environmental Consciousness:** Environmentally aware
- Cultural Influences:** Traditional values

Below the list, there are two sections: 'Needs and Pain Points' and 'Behavioral Data'. The 'Needs and Pain Points' section lists:

- Needs:** High quality products, convenience, personalized experience
- Pain Points:** Lack of time, finding new products, high prices

The 'Behavioral Data' section lists:

- Behavioral Drivers:** Loyalty, recommendations, special offers

● Summary Generation:

PersonaGenius utilizes Gemini AI to generate summaries from the clustered data.

The screenshot shows the PersonaGenius dashboard with the 'Cluster Summaries' tab selected. The main content area displays two cluster summaries:

Cluster 1

Cluster A is characterized by older customers (born between 1957 and 1976) with higher income levels (between \$63,998 and \$79,823). They are typically married and have graduated from college. They tend to purchase more wine, meat, fish, and gold products. They are also more likely to make web and catalog purchases.

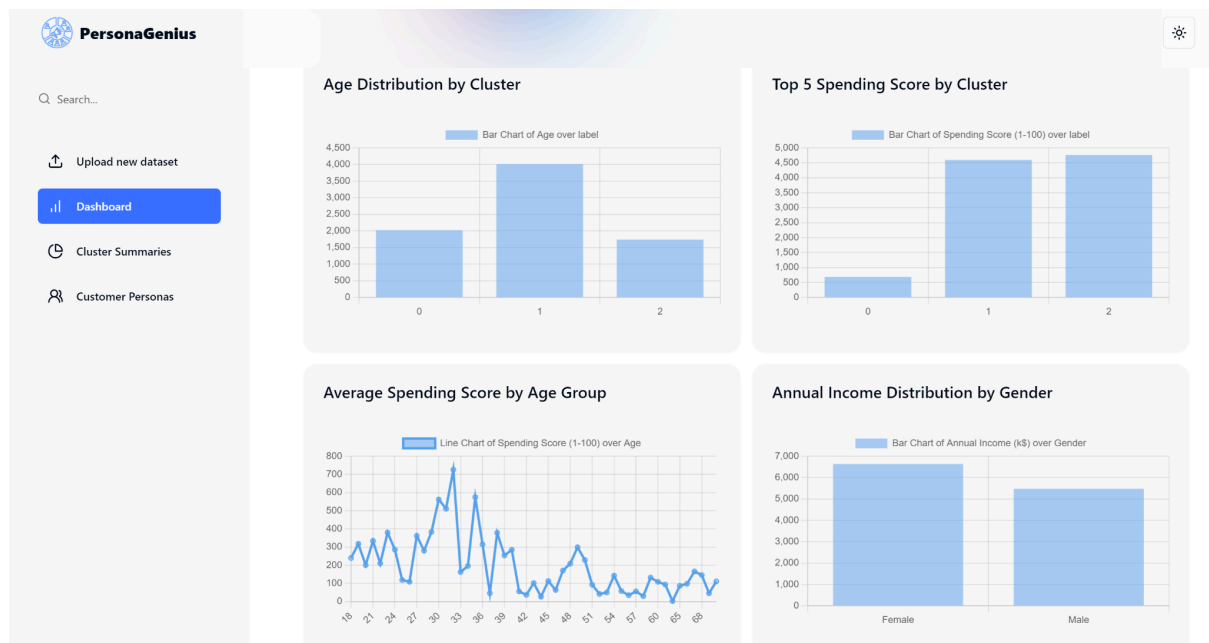
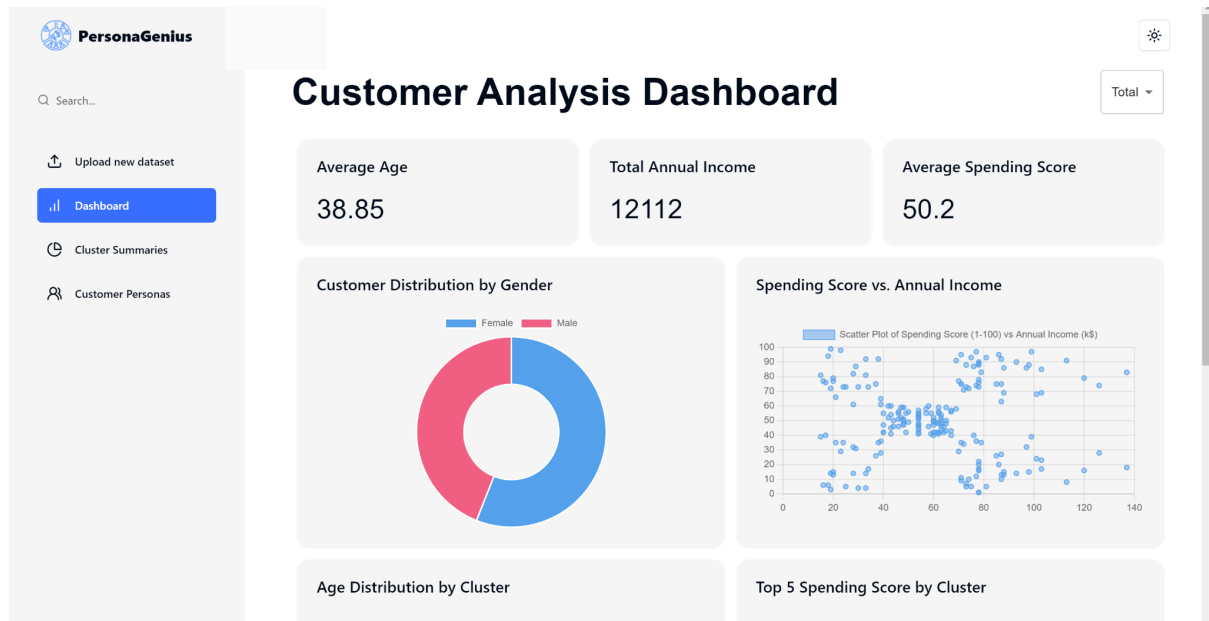
Cluster 2

Cluster B is defined by younger customers (born between 1962 and 1978) with lower income levels (between \$28,543.5 and \$49,125). They are also mostly married and have graduated from college. They tend to make more store purchases and visit the website more frequently. They are less likely to purchase expensive products like wine and gold products.



- **Visualizations:**

PersonaGenius produces visualizations such as charts, graphs, and plots to present the clustering results, persona profiles, and summary data in an intuitive and informative manner.





- **Exporting and Sharing:**

Users can export the generated personas, summaries, and visualizations for further analysis or sharing with stakeholders within their organization.

The screenshot displays the PersonaGenius web application interface. On the left is a sidebar with navigation options: 'Search...', 'Upload new dataset', 'Cluster Summaries', and 'Customer Personas' (highlighted in blue). The main content area shows a persona profile for 'Jane Doe' under 'Cluster 1'. The profile includes a 'Download Personas' button in the top right. The profile details are organized into sections: 'Demographic Data' (Age: 65, Gender: Female, Marital Status: Married, Family Structure: Married couple with no children, Income Level: \$70,000 - \$80,000, Location: Suburban, Occupation: Retired) and 'Psychographics' (Values and Beliefs: Appreciates quality and value). Below the main content, a PDF viewer shows a document titled 'customer-persona (3).pdf' with two pages. The right page of the PDF is a detailed view of Jane Doe's persona, including additional psychographic details like 'Interests and Hobbies', 'Lifestyle Choices', 'Technology Usage', 'Brand Preferences', 'Community Engagement', 'Health and Wellness', 'Family Dynamics', 'Financial Goals', 'Media Consumption', 'Environmental Consciousness', and 'Cultural Influences', as well as a 'Needs and Pain Points' section.

PersonaGenius

Search...

Upload new dataset

Cluster Summaries

Customer Personas

Cluster 1 Cluster 2

Jane Doe

Download Personas

Demographic Data

- 👤 Age: 65
- ♀ Gender: Female
- 💍 Marital Status: Married
- 👨‍👩‍👧 Family Structure: Married couple with no children
- 💰 Income Level: \$70,000 - \$80,000
- 📍 Location: Suburban
- 👔 Occupation: Retired

Psychographics

- 💡 Values and Beliefs: Appreciates quality and value

customer-persona (3).pdf

1 / 2 100% + -

Jane Doe

Demographic Data

- 👤 Age: 65
- ♀ Gender: Female
- 💍 Marital Status: Married
- 👨‍👩‍👧 Family Structure: Married couple with no children
- 💰 Income Level: \$70,000 - \$80,000
- 📍 Location: Suburban
- 👔 Occupation: Retired

Psychographics

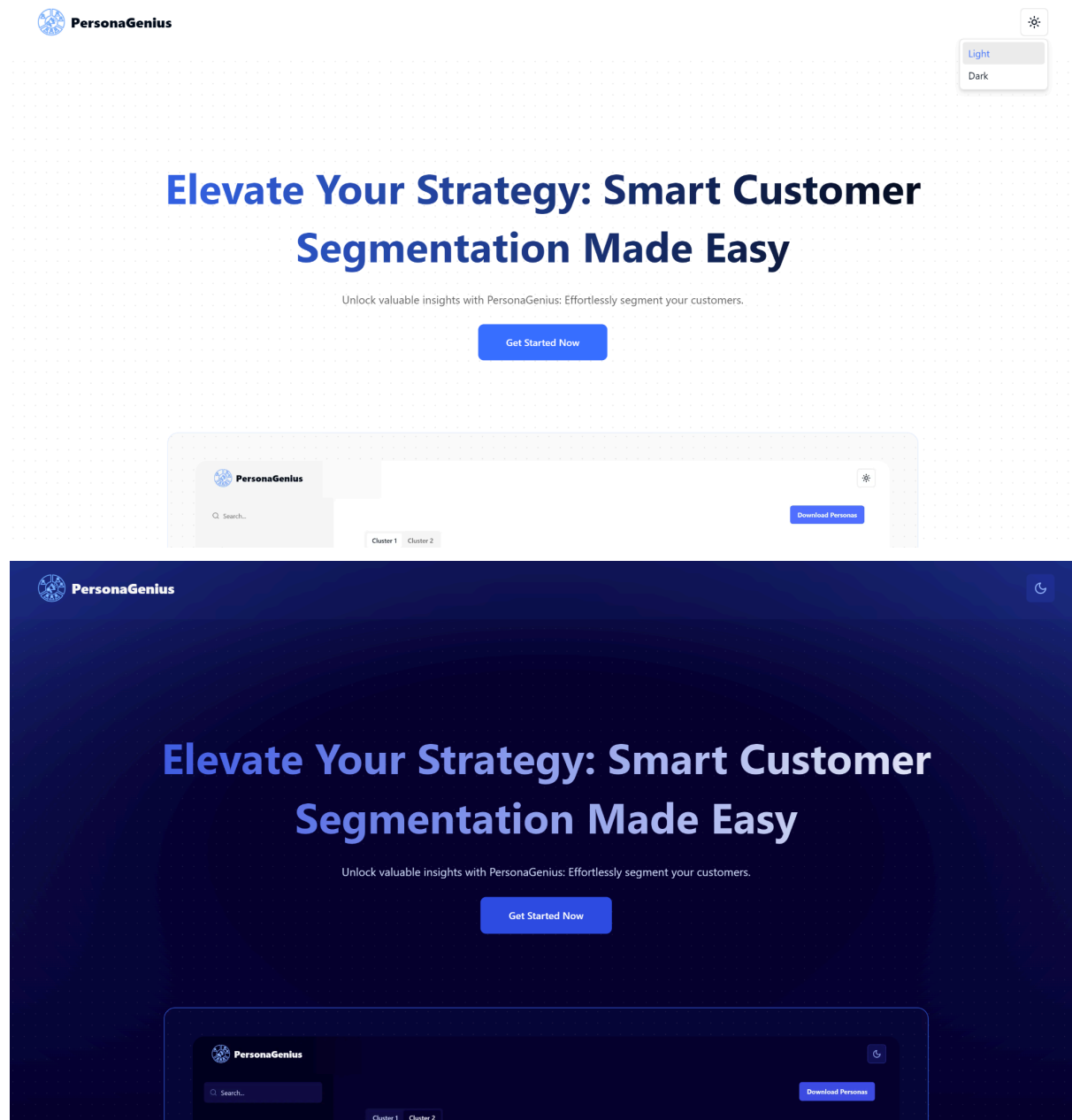
- 💡 Values and Beliefs: Appreciates quality and value
- 🎮 Interests and Hobbies: Cooking, gardening, traveling
- 👤 Lifestyle Choices: Active and social
- 💻 Technology Usage: Regularly uses internet for shopping and entertainment
- 🏷 Brand Preferences: Prefers established brands
- 👥 Community Engagement: Active in local community
- 👤 Health and Wellness: Values healthy living
- 👨‍👩‍👧 Family Dynamics: Strong family values
- 💰 Financial Goals: Financial security
- 📺 Media Consumption: Watches TV, reads newspapers, uses social media
- 🌱 Environmental Consciousness: Environmentally aware
- 🌍 Cultural Influences: Traditional values

Needs and Pain Points

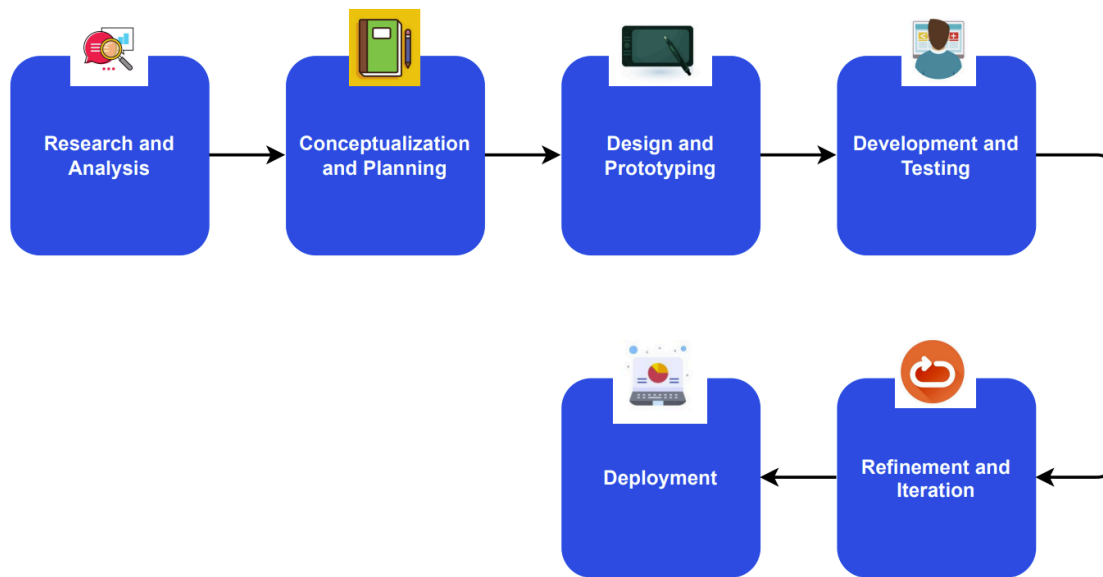


- **Customizable User Experience:**

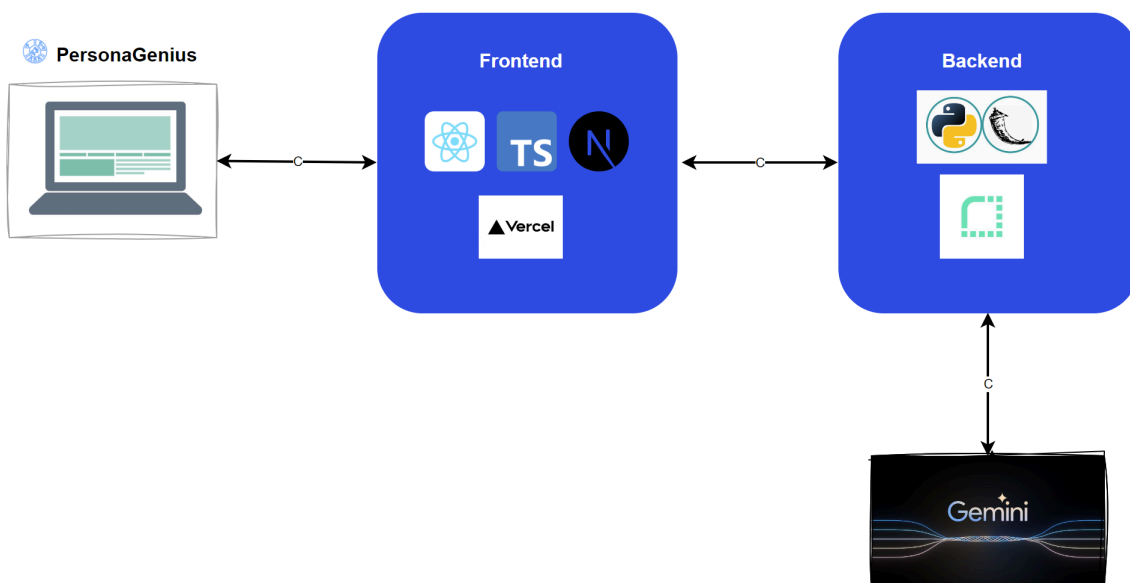
PersonaGenius lets you personalize your workspace with custom themes. Choose between a dark theme or a light theme to match your style or environment. Plus, our platform adapts seamlessly to your device, whether it's a desktop, tablet, or phone. So, you can access PersonaGenius anytime, anywhere, and work however you like.



Project Phases:



Design Methodology Diagram:



Components of System Architecture:

- **PersonaGenius (User Interface):** The user interface (UI) serves as the primary point of interaction for users, enabling them to upload datasets, execute clustering, and visualize the generated personas and insights.
- **Frontend:** The frontend handles all client-side operations and presentation, facilitating communication with the backend for data exchange.

Components:

1. React: A JavaScript library for building user interfaces, particularly single-page applications.
 2. Next.js: A React framework that supports server-side rendering and static site generation to improve performance and SEO.
 3. TypeScript: A statically typed superset of JavaScript that enhances code quality and maintainability.
- **Backend:** The backend manages business logic, data processing, clustering, and AI functionalities. It processes uploaded datasets and generates personas and visualizations.

Components:

1. Python: The primary programming language for backend development due to its versatility and extensive library support.
 2. Flask: A lightweight WSGI web application framework used to build the backend services and APIs.
 3. Google Generative AI - Gemini: For generating detailed personas from the clustered data.
- **Deployment Environment**
 1. **Vercel:** The frontend application is deployed on Vercel, providing efficient and reliable hosting with support for serverless functions and static site generation.
 2. **Render:** The backend application is deployed on Render, a platform that ensures scalability and manages infrastructure for web applications and APIs.



Data Flow and Interaction:

1. User Interaction:

Users interact with the PersonaGenius interface to upload datasets and initiate the processes of clustering and persona generation.

2. Frontend Processing:

The frontend captures user inputs and communicates with the backend through APIs.

3. Backend Processing:

- The backend processes the incoming data, performing tasks such as data cleaning and exploratory data analysis using Pandas and Numpy.
- Machine learning algorithms are applied using Torch and Scikit-learn to cluster the data.
- Detailed personas are generated with the help of Google Generative AI - Gemini and Langchain.

4. Data Visualization:

Visualizations are created using Matplotlib and Seaborn and sent back to the frontend for display.

5. Frontend Display:

The frontend displays the generated personas and visualizations, providing a comprehensive view of the clustering results.

Implementation Details:

<u>Category</u>	<u>Details</u>
Frontend	<ul style="list-style-type: none">• React (v18.0)• Next.js (v14.2.3)• TypeScript (v5.3.3)• Radix UI (v1.1.2)
Backend	<ul style="list-style-type: none">• Python (v3.11)• Flask (v3.0.3)• Torch (v2.3.1)• Scikit-learn (v1.5.0)• Google Generative AI - Gemini (v0.6.0)• Langchain (v0.2.3)
Deployment Environment	<ul style="list-style-type: none">• Frontend: Vercel• Backend: Render



Datasets	<ul style="list-style-type: none">● UCI Bank Marketing Dataset● Customer Segmentation Dataset● Marketing Campaign● Clustering Dataset● Sales Dataset
----------	--

Potential Impact of PersonaGenius:

While we haven't conducted a formal demo yet, PersonaGenius holds promise for various industries. By providing detailed insights into customer behavior and preferences, it could revolutionize marketing strategies, leading to more personalized campaigns and improved customer satisfaction. However, before it's ready for sale, we need to conduct thorough validation and testing, and define a clear business model and pricing strategy.

Github Repo Link:

Repository Link: <https://github.com/MunnazzahAslam/PersonaGenius>

Project Live Link: <https://persona-genius.vercel.app>

References:

- [1] <https://github.com/MunnazzahAslam/customer-segmentation>
- [2] <https://github.com/tinacypeng/customer-persona-generator>
- [3] <https://github.com/MunnazzahAslam/DarazDataset>
- [4] https://github.com/rahulsingh2312/innov_2024_k-means-gamble
- [5] https://python.langchain.com/v0.1/docs/integrations/chat/google_generative_ai/
- [6] <https://link.springer.com/article/10.1007/s10618-022-00869-6>