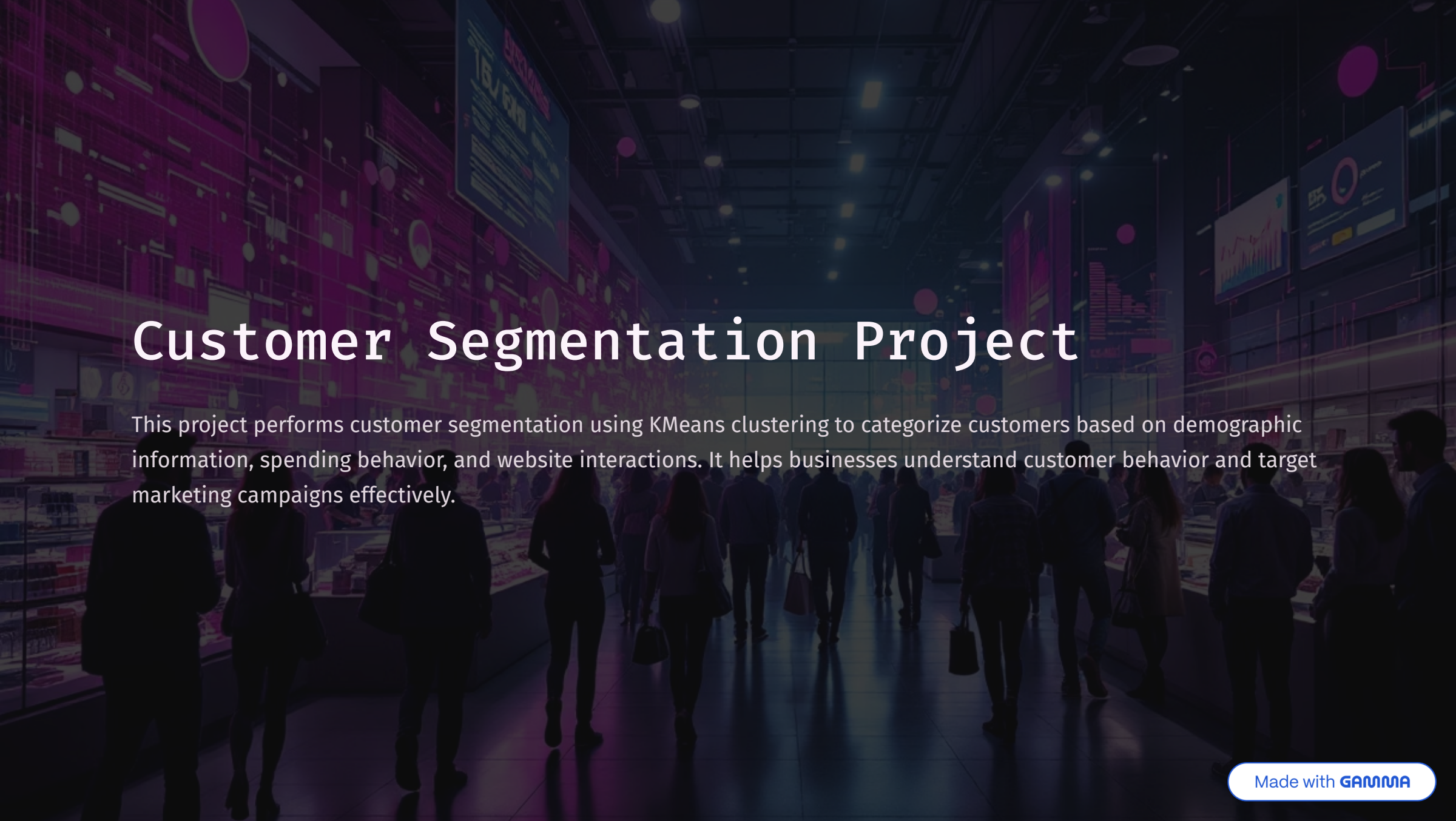


Customer Segmentation Project

This project performs customer segmentation using KMeans clustering to categorize customers based on demographic information, spending behavior, and website interactions. It helps businesses understand customer behavior and target marketing campaigns effectively.

Overview

This project performs **customer segmentation** using **KMeans clustering** to categorize customers based on demographic information, spending behavior, and website interactions. It helps businesses understand customer behavior and target marketing campaigns effectively.

Demographic Analysis

Categorize customers based on age, income, and family composition to understand different customer profiles.

Spending Behavior

Analyze total spending patterns and purchase preferences across different customer segments.

Website Interactions

Track web visits, store purchases, and campaign responses to measure customer engagement.

Notebook Highlights (analysis_model.ipynb)

The analysis notebook contains comprehensive data processing and machine learning workflows for customer segmentation.

01

Data Cleaning & Preprocessing

Handled missing values, created features like `age`, `total_children`, `total_spending`, and `customer_since`.

02

Exploratory Data Analysis (EDA)

Visualized age, income, total spending, and campaign acceptance patterns.

03

Feature Engineering

Aggregated campaign responses into a single `accepted` column and created `age_groups`.

04

Clustering

Scaled features and applied **KMeans clustering** to segment customers into 6 clusters.

05

PCA Visualization

Reduced dimensions to 2D for cluster visualization.

06

Model Saving

Saved `KMeans` and `StandardScaler` models using `joblib` for future predictions.

Data Cleaning & Preprocessing

Feature Creation

Handled missing values, created features like `age`, `total_children`, `total_spending`, and `customer_since`.

- Calculated customer age from birth year
- Combined kids and teens into total_children
- Aggregated spending across all categories
- Computed customer tenure from enrollment date

Exploratory Data Analysis

Visualized age, income, total spending, and campaign acceptance patterns.

Age Distribution

Analyzed customer age ranges to identify primary demographic segments and generational patterns.

Income Analysis

Examined income levels across customer base to understand purchasing power distribution.

Spending Patterns

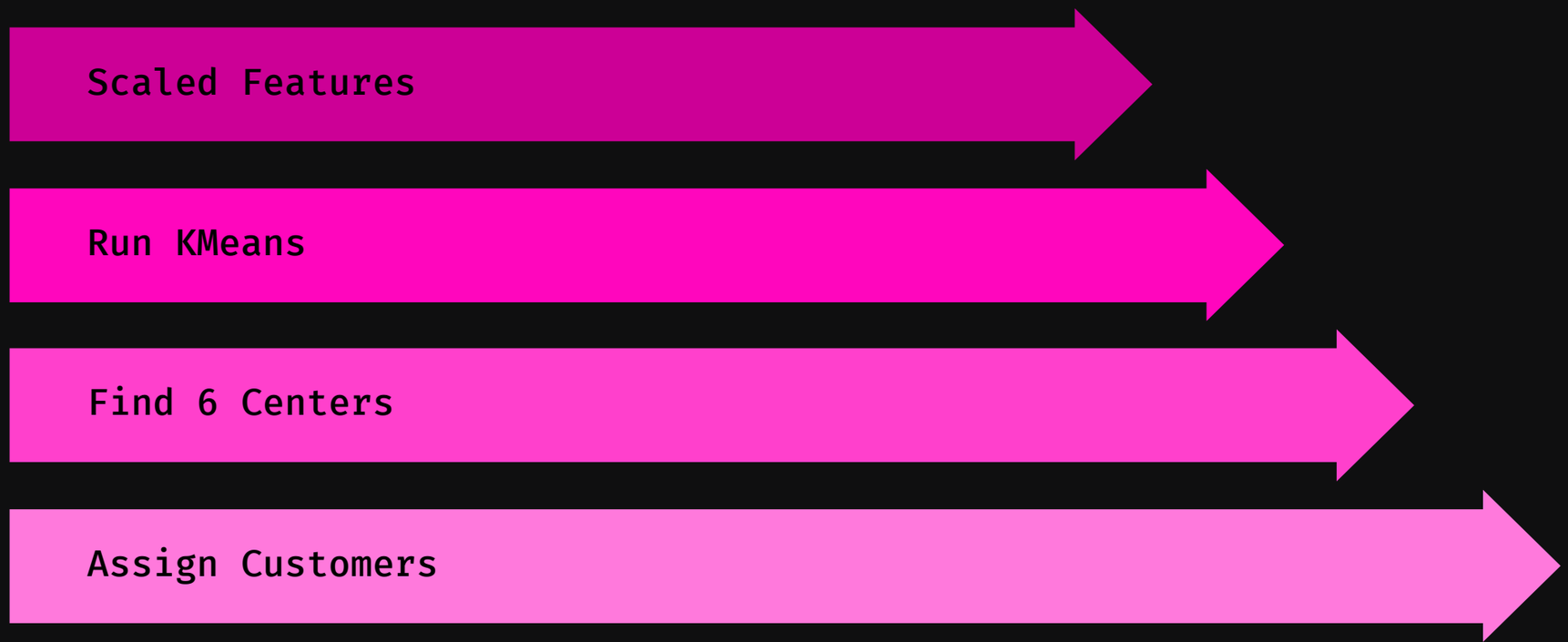
Tracked total spending behavior to identify high-value and low-value customer groups.

Campaign Acceptance

Measured response rates to marketing campaigns across different customer segments.

KMeans Clustering Implementation

Scaled features and applied **KMeans clustering** to segment customers into 6 clusters.



Clustering Process

- Features were standardized using StandardScaler
- KMeans algorithm identified 6 distinct customer segments
- Each cluster represents unique behavioral patterns
- Models saved for deployment and future predictions

PCA Visualization

Reduced dimensions to 2D for cluster visualization.

This allows us to visualize high-dimensional customer data in an interpretable two-dimensional space.

Streamlit Appsegmentation.py

The Streamlit application provides an interactive interface for real-time customer segmentation predictions.

1

Interactive Interface

Users can input customer details such as age, income, total spending, web/store purchases, website visits, and recency.

2

Prediction

Inputs are scaled using the saved `StandardScaler`, and clusters are predicted using the trained `KMeans` model.

3

User Feedback

Shows predicted cluster along with a brief description of its characteristics.

4

Ease of Use

Simple web interface for non-technical users to quickly classify customers.

Application Features

User Input Fields

Users can input customer details such as age, income, total spending, web/store purchases, website visits, and recency.

- Age and demographic information
- Income level
- Total spending across categories
- Web and store purchase counts
- Website visit frequency
- Recency of last purchase

Prediction Engine

Inputs are scaled using the saved `StandardScaler`, and clusters are predicted using the trained `KMeans` model.

Shows predicted cluster along with a brief description of its characteristics.

Simple web interface for non-technical users to quickly classify customers.

Impact

Identify High-Value Customers

Enables businesses to **identify high-value customers** and optimize marketing strategies.

Focus resources on segments with highest revenue potential and engagement rates.

Ready-to-Deploy Solution

Provides a **ready-to-deploy solution** for customer segmentation using Python, Pandas, Scikit-learn, and Streamlit.

Complete end-to-end pipeline from data processing to interactive predictions.

Optimize Marketing Campaigns

Target marketing campaigns effectively based on customer behavior patterns and preferences.

Increase conversion rates by tailoring messages to specific customer segments.

Technical Stack & Resources

Technologies Used

- **Python** - Core programming language
- **Pandas** - Data manipulation and analysis
- **Scikit-learn** - Machine learning and clustering
- **Streamlit** - Interactive web application
- **Joblib** - Model serialization
- **KMeans** - Clustering algorithm
- **StandardScaler** - Feature scaling
- **PCA** - Dimensionality reduction

Project Files

- `analysis_model.ipynb` - Jupyter notebook with full analysis
- `segmentation.py` - Streamlit application code
- `customer_segmentation.csv` - Dataset
- `README.md` - Project documentation

Repository Information

Customer segmentation using KMeans clustering to analyze spending behavior, demographics, and web interactions, with a Streamlit app for interactive predictions