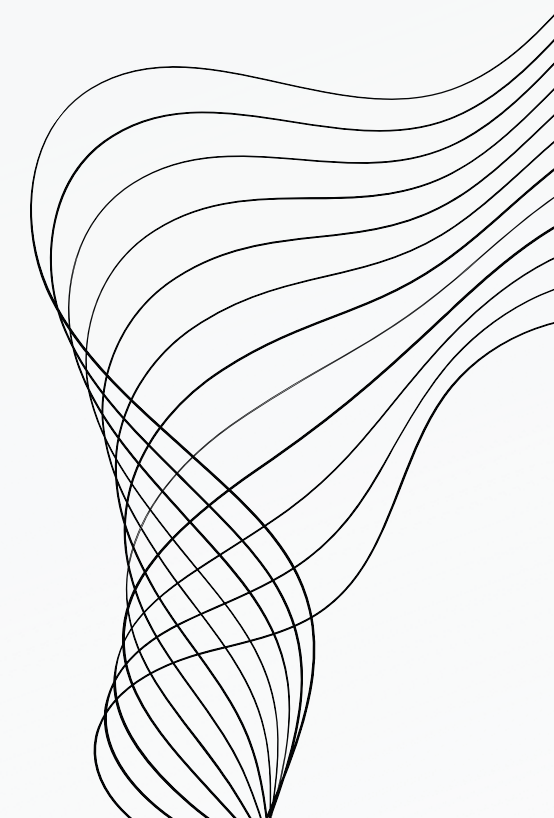


**CUSTOMER
SEGMENTATION
PROJECT**





CONTENT



01

INTRODUCTION

02

DATA PREPROCESSING

03

EDA

04

UNIVARIATE & BIVARIATE ANALYSIS

05

CLUSTERING

06

INTERPRETATION

07

LIMITATION AND FUTURE SCOPE

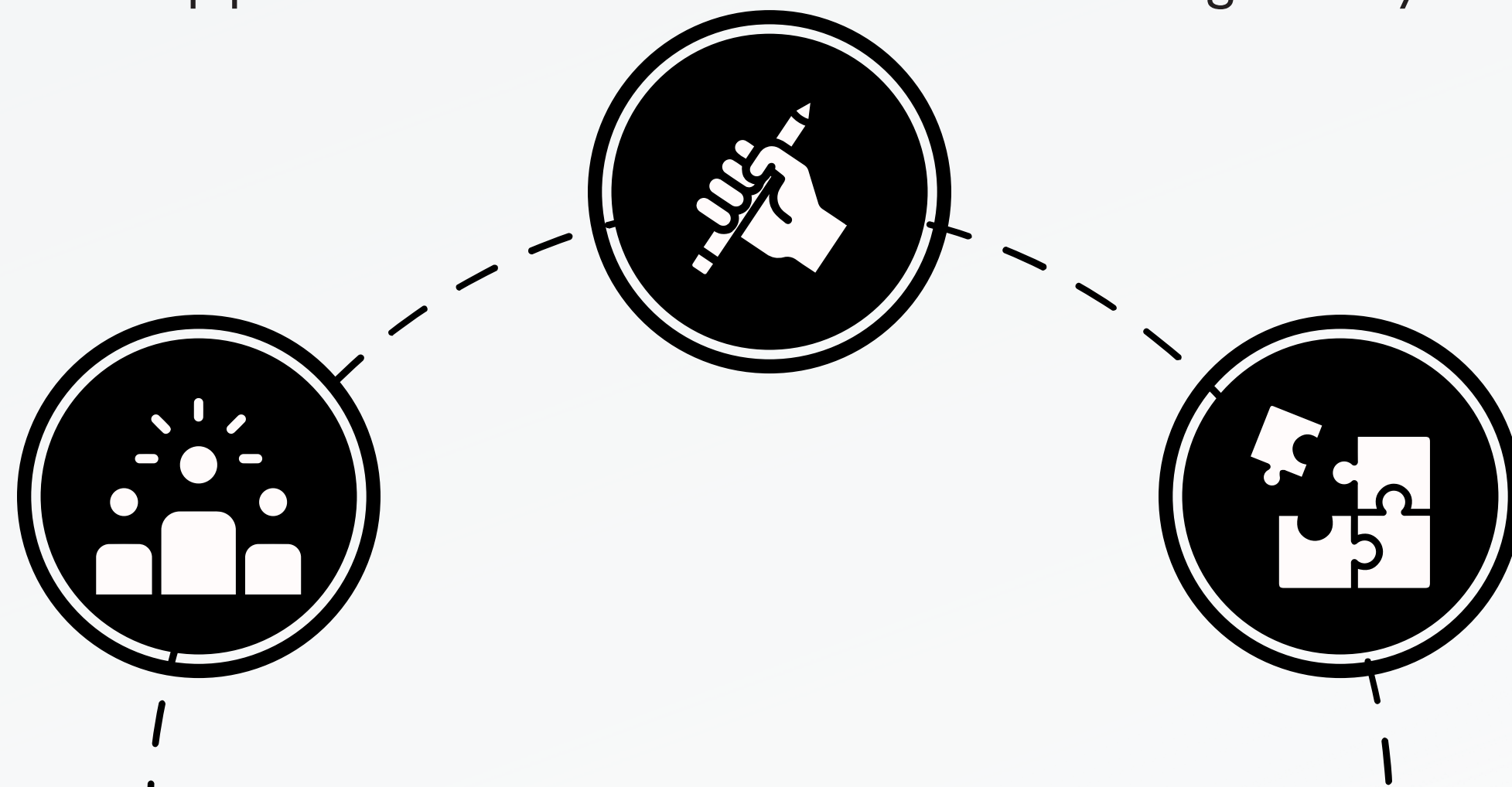
INTRODUCTION

The goal of this project was to perform a customer segmentation analysis using clustering techniques to gain insights into customer behavior and preferences. The dataset used for the analysis contained information about customers, including their gender, age, annual income, and spending score.



DATA PREPROCESSING

The dataset was initially explored to understand its structure and identify any missing values or inconsistencies. The gender column was encoded using a binary encoding scheme, where male was represented as 0 and female as 1. This encoding was applied to facilitate the clustering analysis.



EDA- EXPLORATORY DATA ANALYSIS

Several exploratory data analysis techniques were employed to understand the characteristics and distribution of the data.

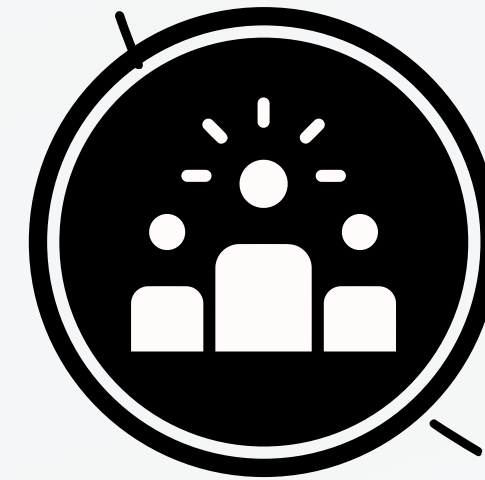
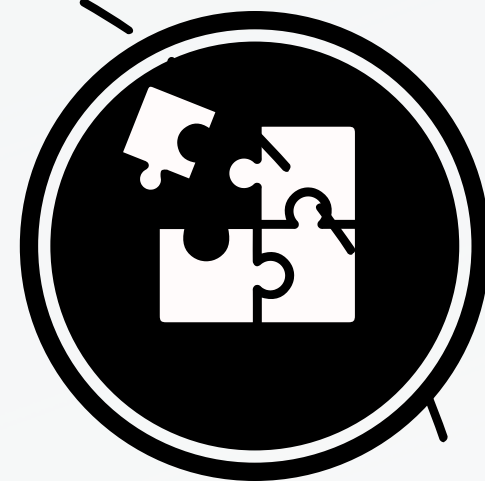
Descriptive statistics, such as mean, mode, minimum, and maximum, were calculated for the different features. Visualizations, including histograms, box plots, and scatter plots, were generated to examine the relationships between variables.



UNIVARIATE & BIVARIATE

Bivariate Analysis:

Bivariate analysis involved examining the relationships between pairs of variables. Pair plots provided a comprehensive view of the interactions and correlations between different variables, enabling the identification of patterns and trends. Heatmaps further emphasized the strength and direction of the correlations between variables, highlighting any significant associations.



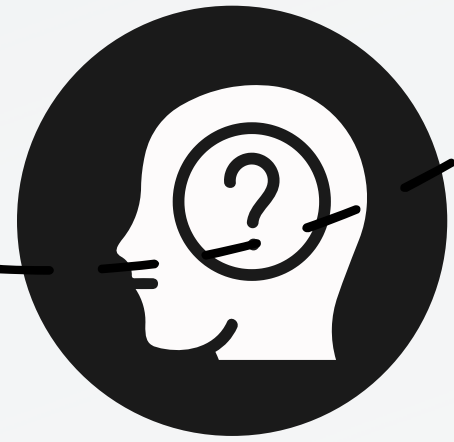
Univariate Analysis:

Univariate analysis was conducted to understand the central tendencies and dispersion of the variables. Descriptive statistics, including mean, mode, maximum, and minimum, were calculated for each variable. These statistics provided an overview of the data distribution and allowed for comparisons between different segments.

CLUSTER ANALYSIS



K-means clustering was employed to segment the customers based on their gender, age, annual income, and spending score. The elbow method was utilized to determine the optimal number of clusters, which involved plotting the within-cluster sum of squares (WCSS) against the number of clusters and selecting the number of clusters at the "elbow" point where the rate of decrease in WCSS diminished. Additionally, silhouette analysis was performed, calculating the average silhouette score for different numbers of clusters to assess the quality of the clustering.



STRATEGY N°1

STRATEGY N°2

STRATEGY N°3

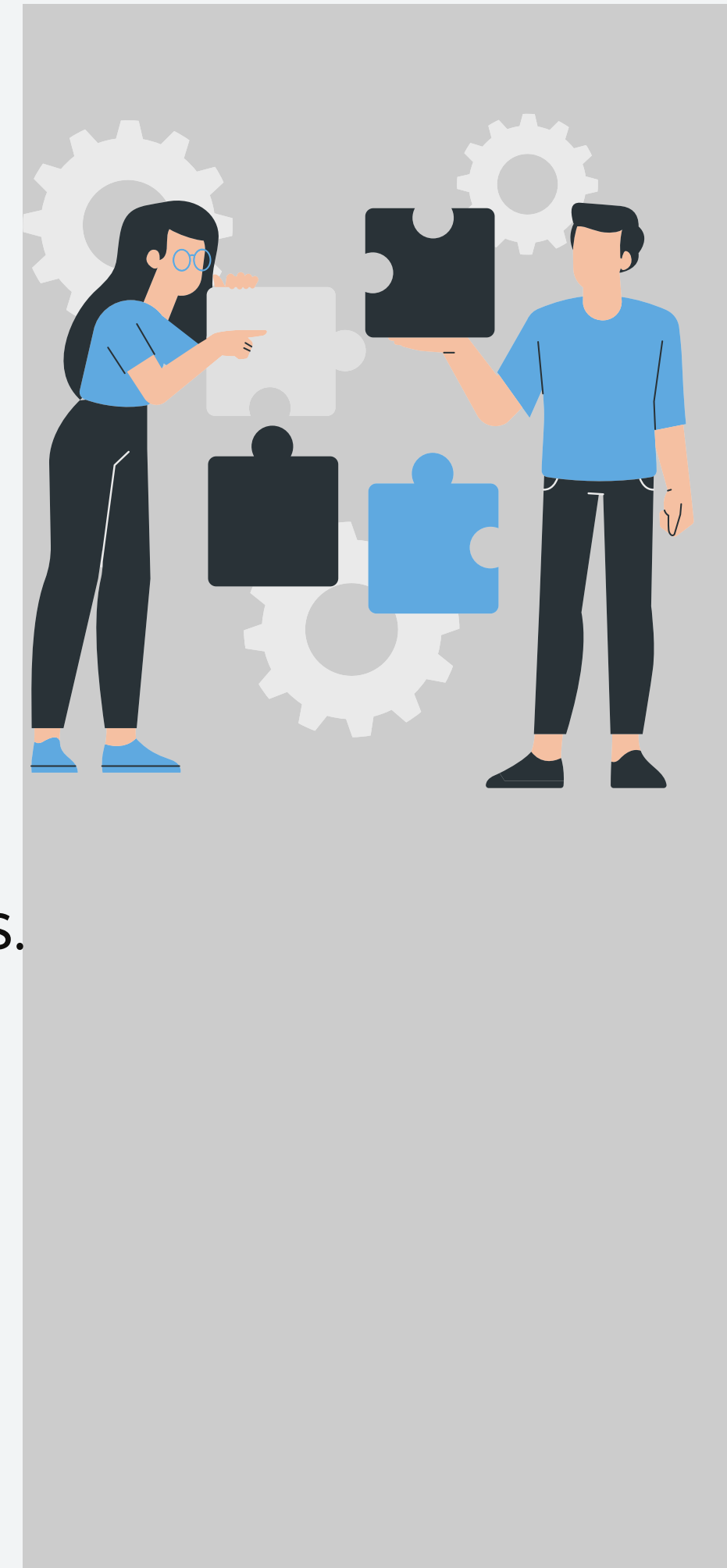


INTERPRETATION

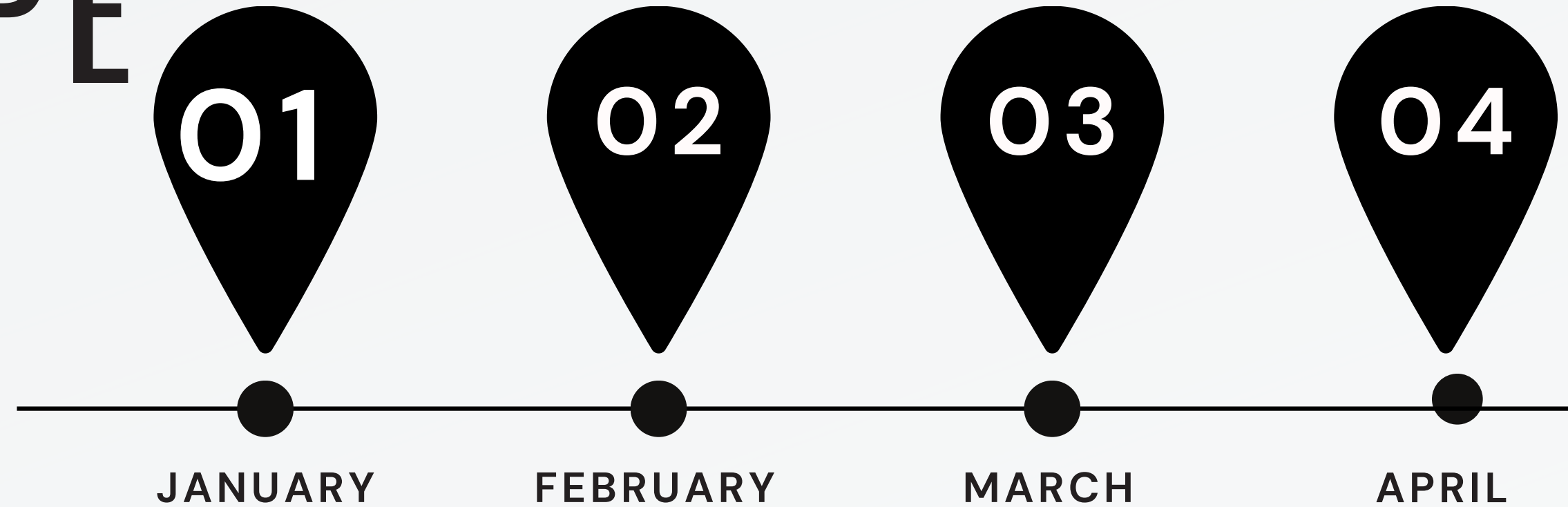


LIMITATION

It is important to acknowledge that clustering analysis is a descriptive technique and does not establish causal relationships. Furthermore, the interpretation of the clusters is subjective and may require validation and refinement based on domain knowledge and further analysis.



FUTURE SCOPE



In future research, incorporating additional variables such as purchase history or online behavior could enhance the clustering analysis and provide a more comprehensive understanding of customer segmentation. Furthermore, exploring advanced clustering algorithms or combining clustering with other machine learning techniques may improve the accuracy and effectiveness of the segmentation.