

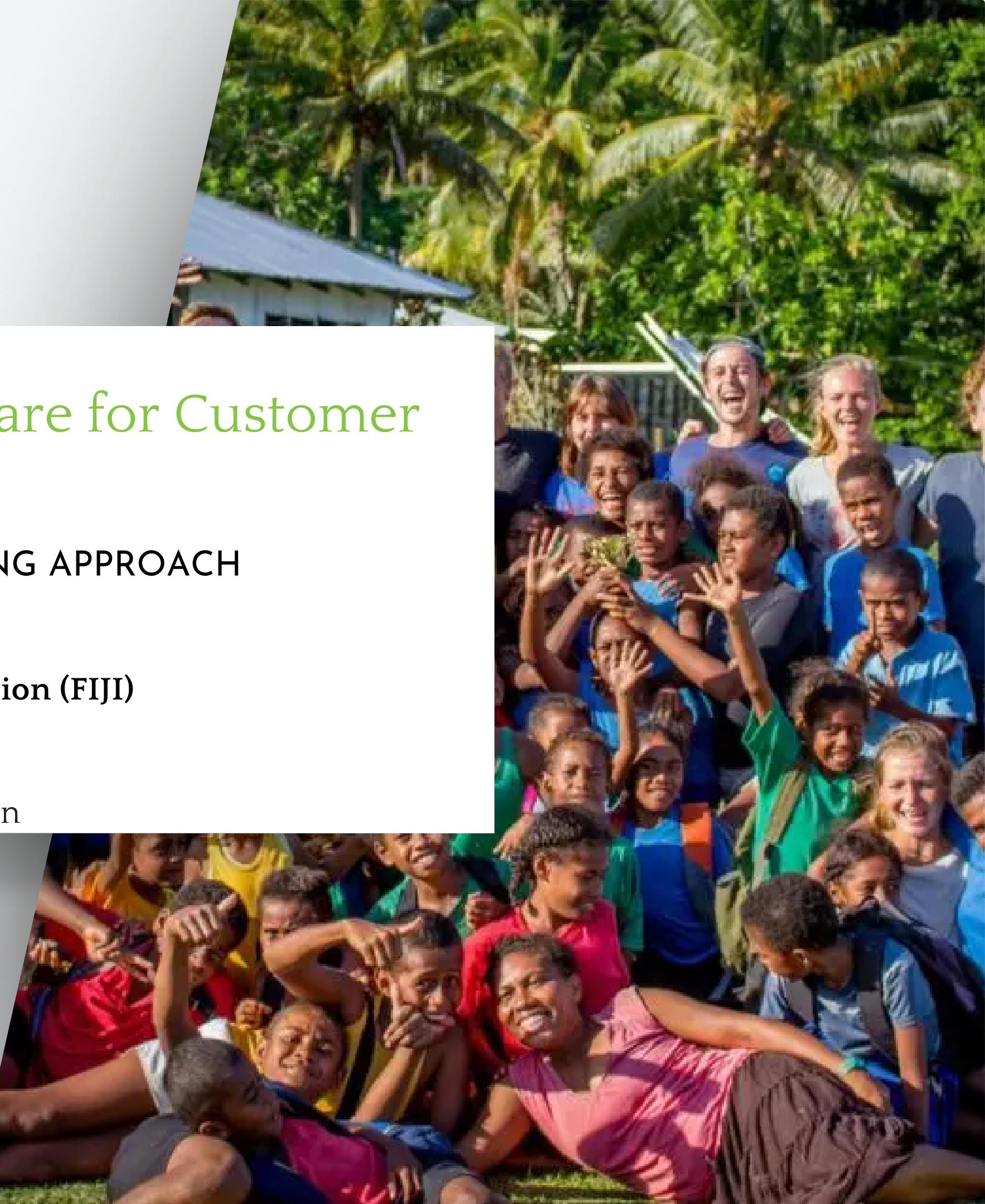


Customer Profiling Software for Customer Behavior Analysis

A MACHINE LEARNING APPROACH

Think Pacific Foundation (FIJI)

by
Sadiq Balogun





CONTENT

01

Introduction

02

Objective

03

Dataset

04

Exploratory Data
Analysis

05

Modeling

06

Model
Deployment

07

Cloud
Deployment

INTRODUCTION



Customer Clustering is a technique that involves dividing a customer base into distinct groups based on shared characteristics



Machine Learning is the development of algorithms that enable computers to learn patterns and predictions or decisions without explicit programming



IMPORTANCE OF CLUSTERING

PERSONALISED MARKETING

Customer profiling allows businesses to tailor marketing strategies to specific customer segments, increasing the effectiveness of campaigns.

RESOURCE OPTIMIZATION

Profiling helps allocate resources efficiently by focusing on high-value customer segments, leading to improved ROI.

RISK MANAGEMENT

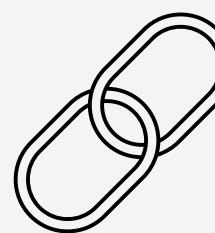
Identifying customer segments prone to churn or non-payment helps businesses take proactive measures to mitigate risks.

OBJECTIVE

To develop a cluster profiling software to identify high-value customer segments for targeted marketing



DATASET



SOURCE

The dataset is an open-source data obtained from Kaggle via the link below.
[Link](#)



OBERVATIONS

Contains 200 observations of a mall's customers



FEATURES

- Gender
- Age
- Annual Income
- Spending Score

UNVEILING THE CUSTOMERS' INSIGHTS





200

CUSTOMERS

56%



44%



18 years

Youngest Customer

70 years

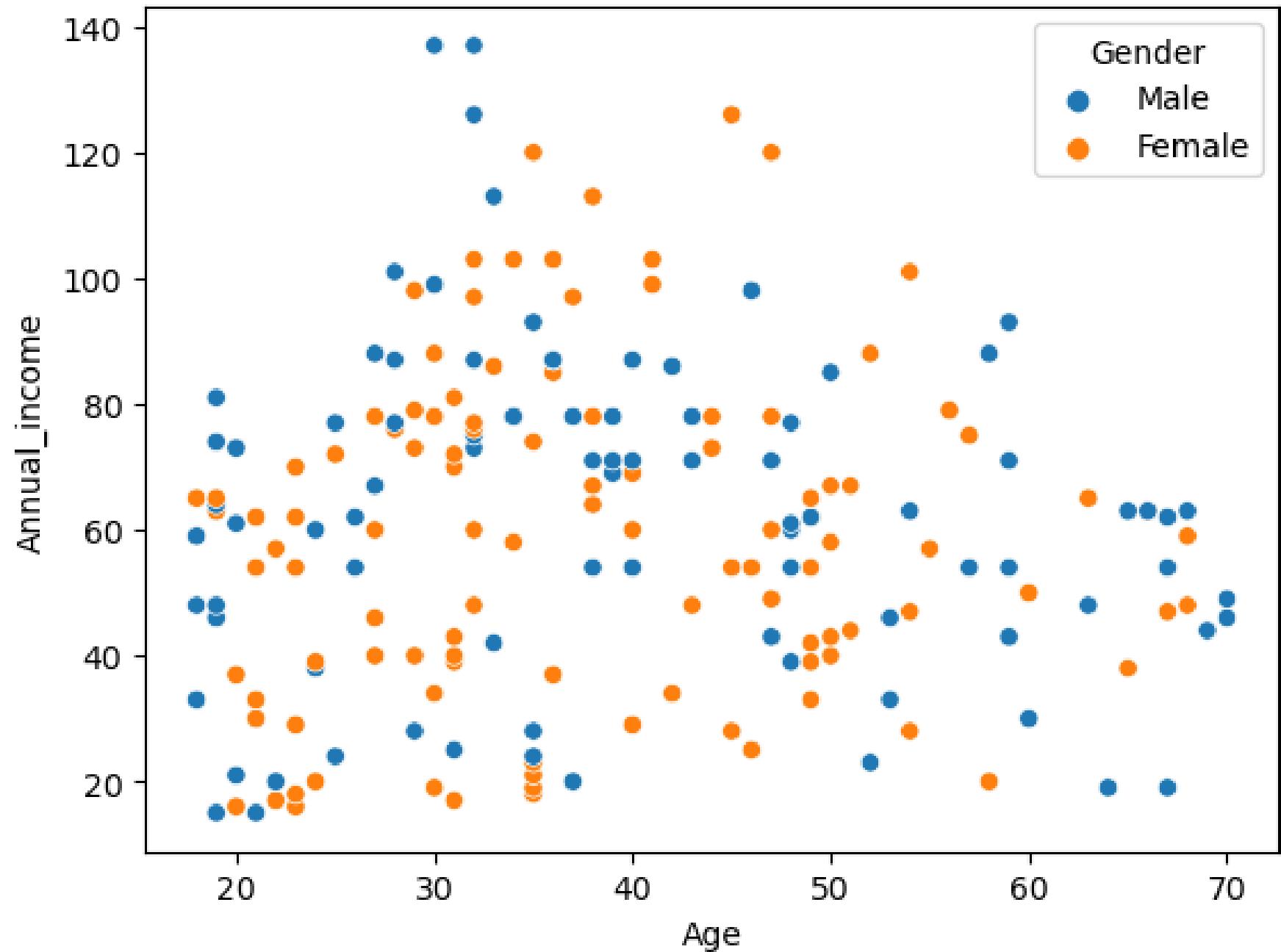
Oldest Customer

STATISTICS

	ANNUAL INCOME (\$)	SPENDING SCORE
Minimum	15,000	1
Average	60,000	50
Maximum	137,000	99

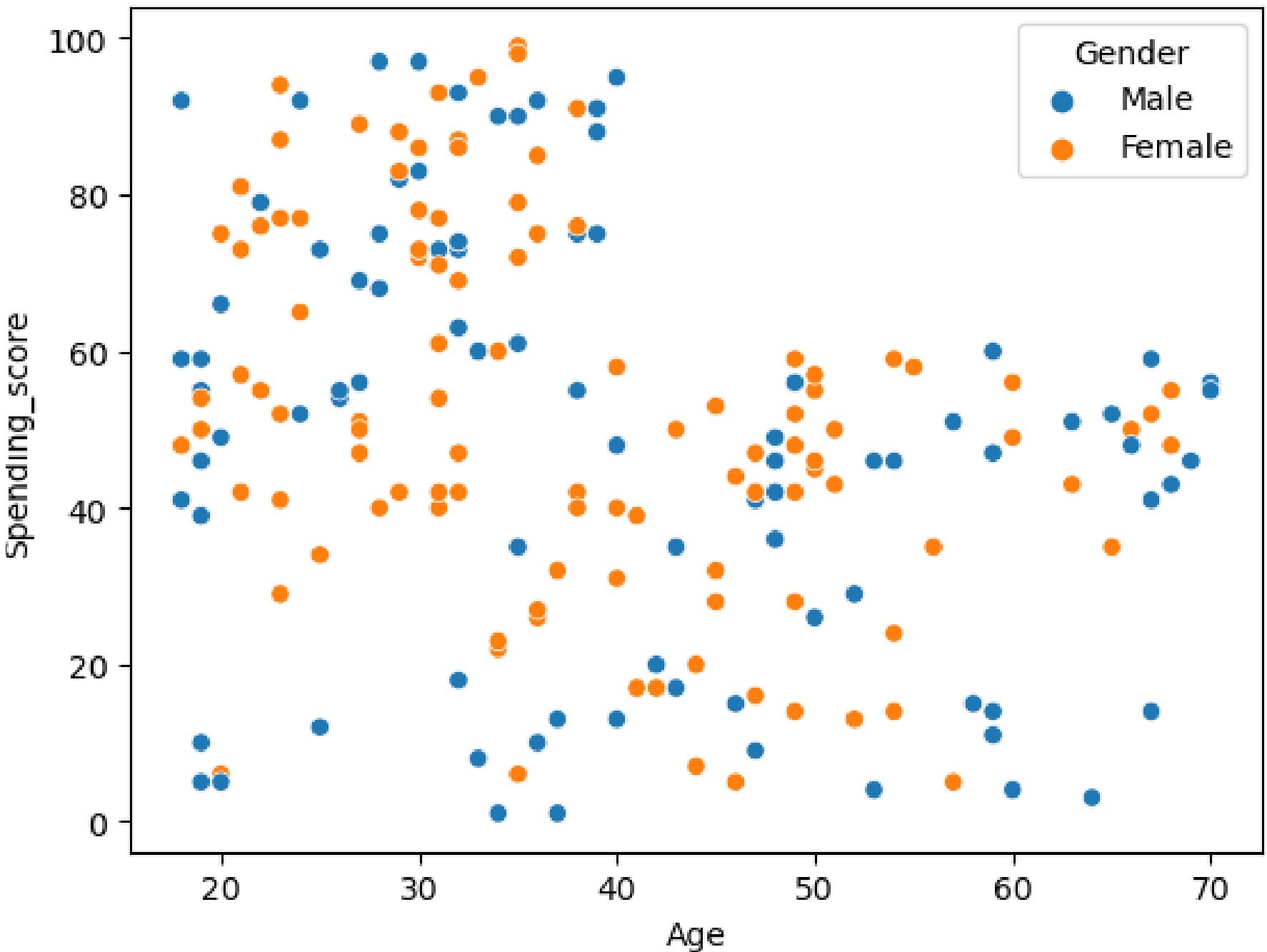
Relationship Between AGE & ANNUAL INCOME

No clear linear relationship
between the features, and there's
diversity in income for every age.



Relationship Between AGE & SPENDING SCORE

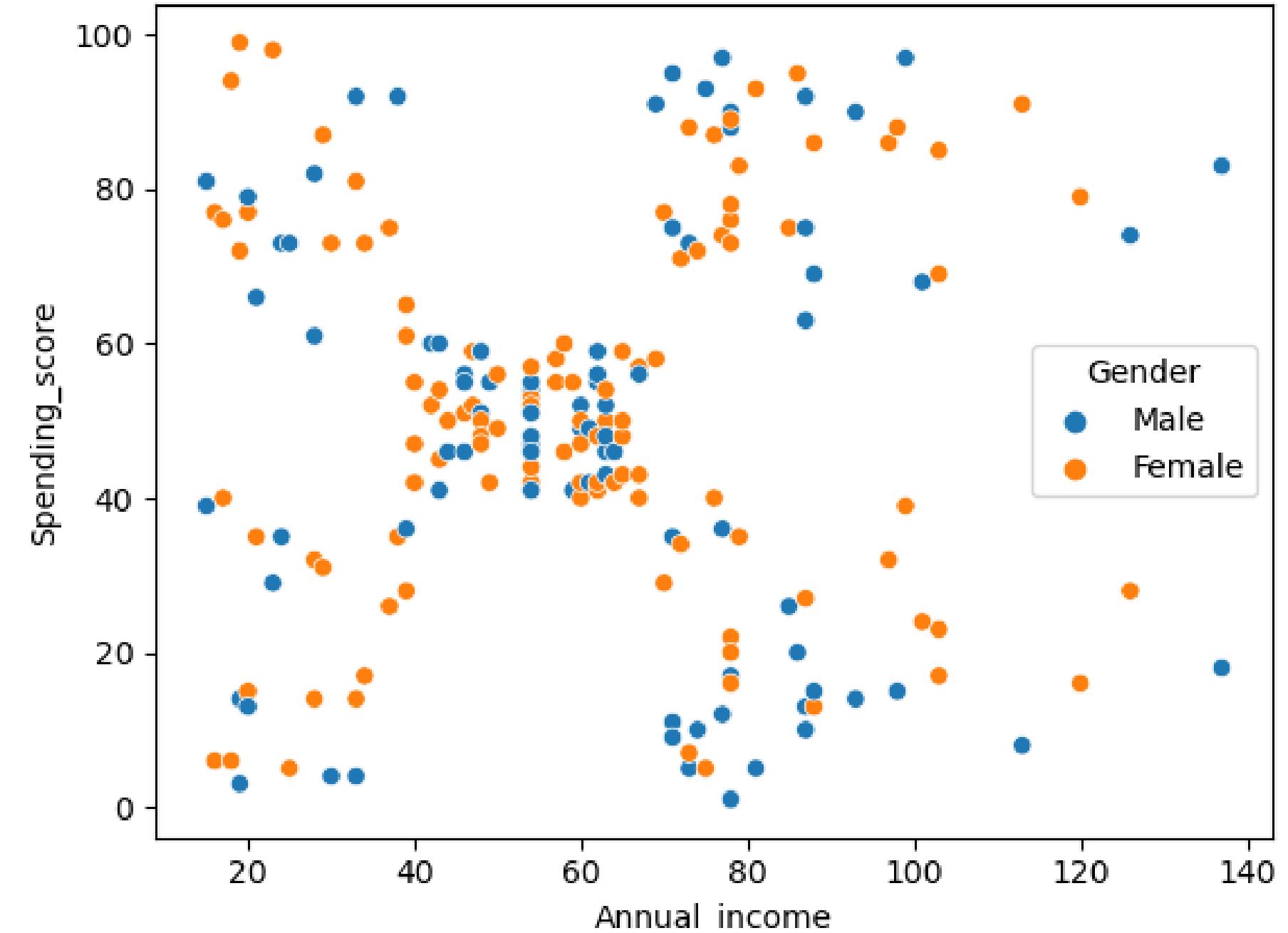
Customers that are older than 40 years have spending scores less than 60.



Relationship Between

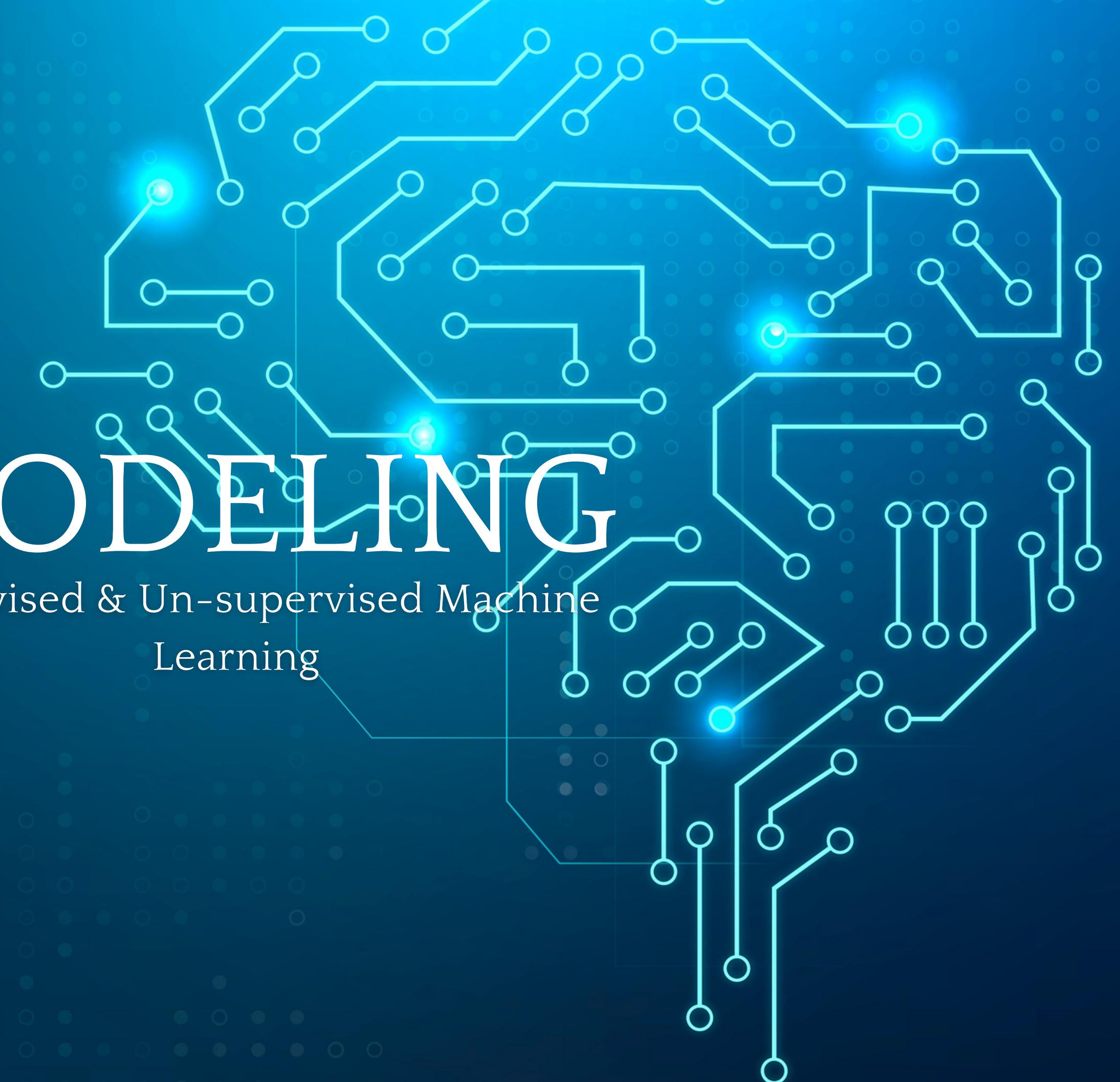
SPENDING SCORE & ANNUAL INCOME

Some customers who have an above-average annual income have a lower-than-average spending score.



MODELING

Supervised & Un-supervised Machine
Learning



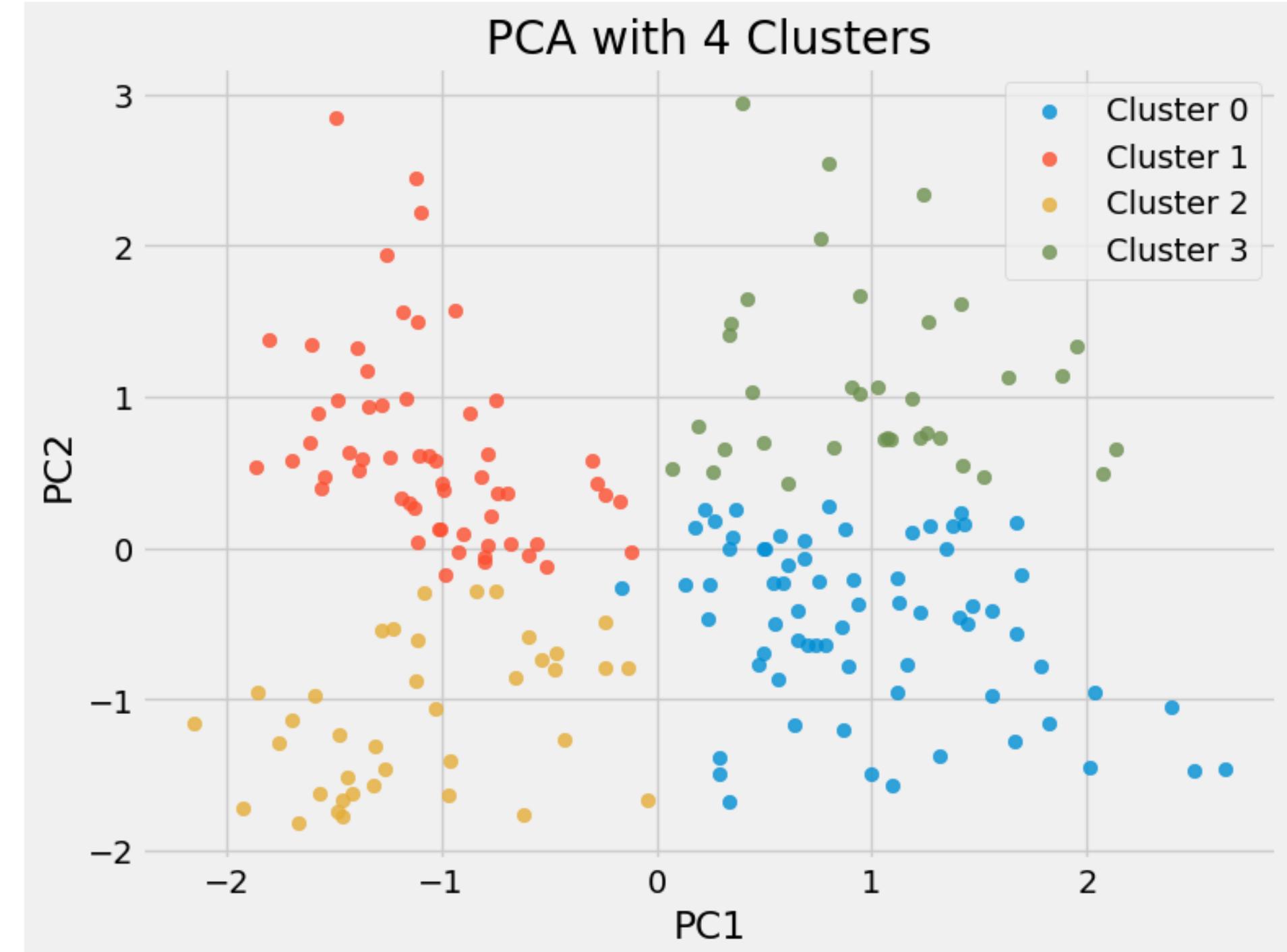
Unsupervised Machine Learning

K-MEANS CLUSTERING

Why K-Means?

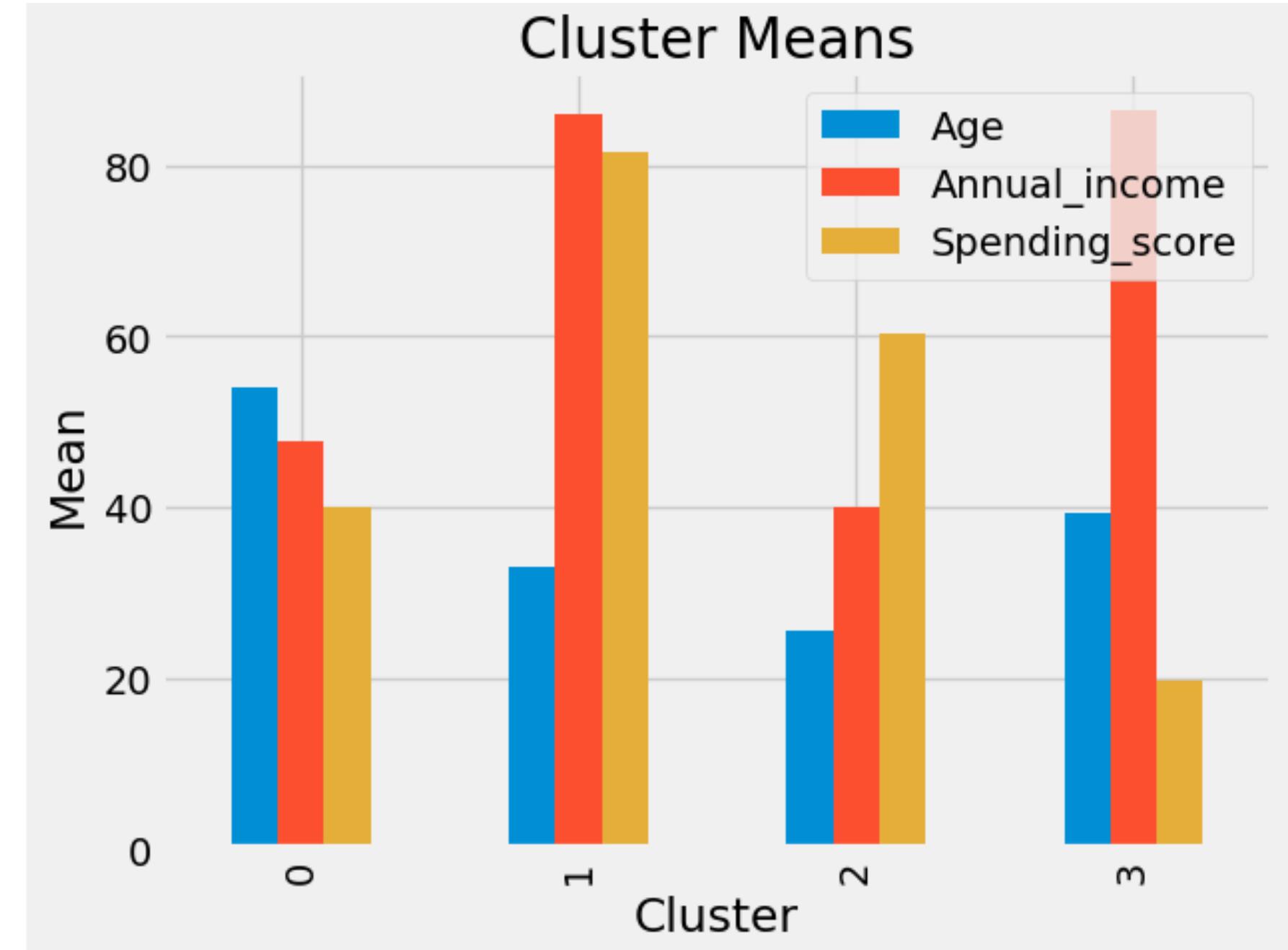
- Efficient in data partitioning
- Computationally efficient and easy to interpret

-
- Successfully categorized the customers into different segments based on their characteristics. (Cluster 0-3)
 - Each colour represents a specific cluster.



UNDERSTANDING THE CLUSTERS

- Cluster 0 = Average spenders
- Cluster 1 = High spenders
- Cluster 2 = High spenders
- Cluster 3 = Low spenders



Supervised Machine Learning

THE ALGORITHMS PERFORMANCE

Two ML algorithms were trained and their performance was tested on new data

	RANDOM FOREST	DECISION TREE
Accuracy	97%	93%

MODEL DEPLOYMENT

- The Random Forest algorithm was deployed as an app using Streamlit
- Tested locally as seen in the screenshot

Customer Segmentation

Select Gender

Male

Age

32

Annual Income

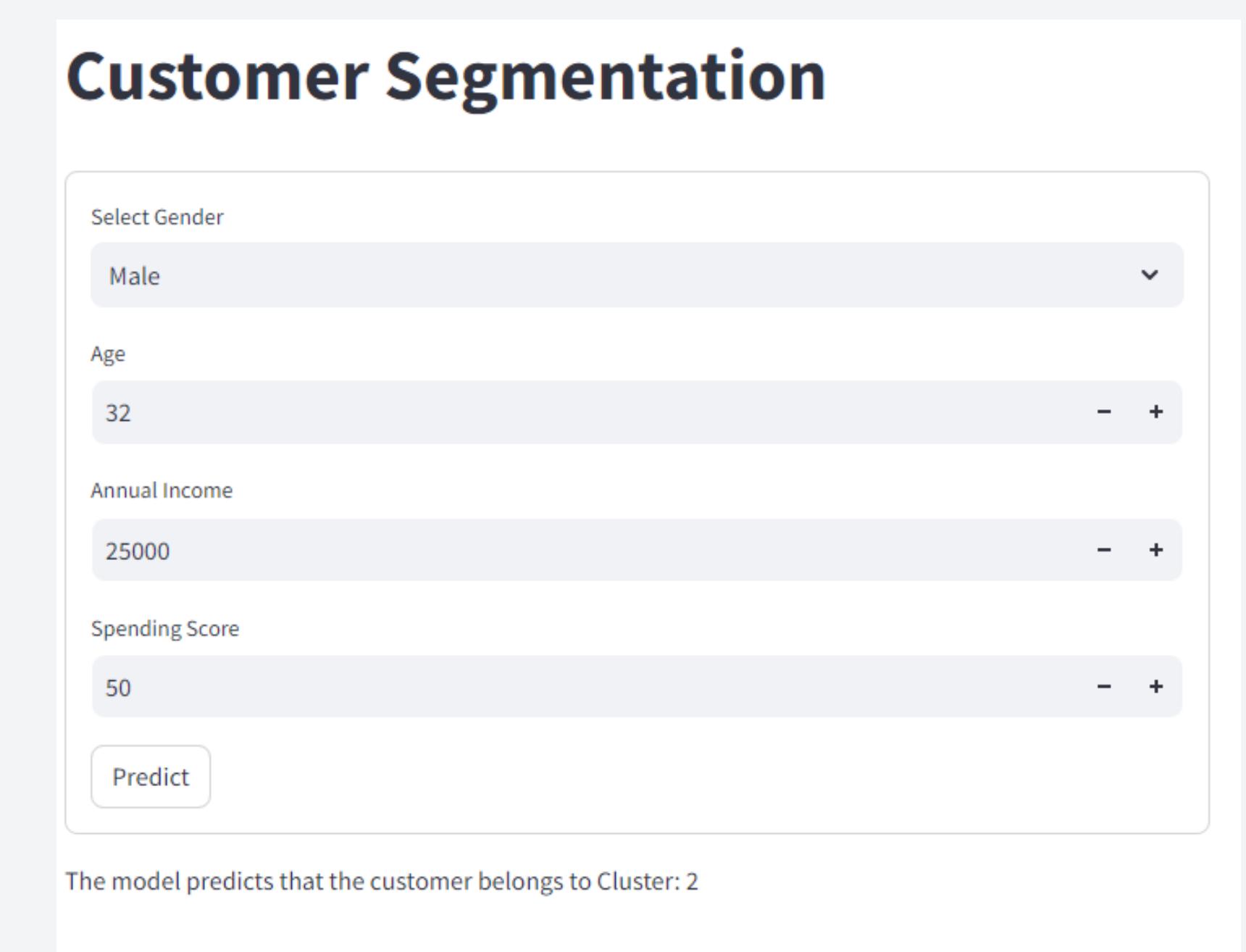
25000

Spending Score

50

Predict

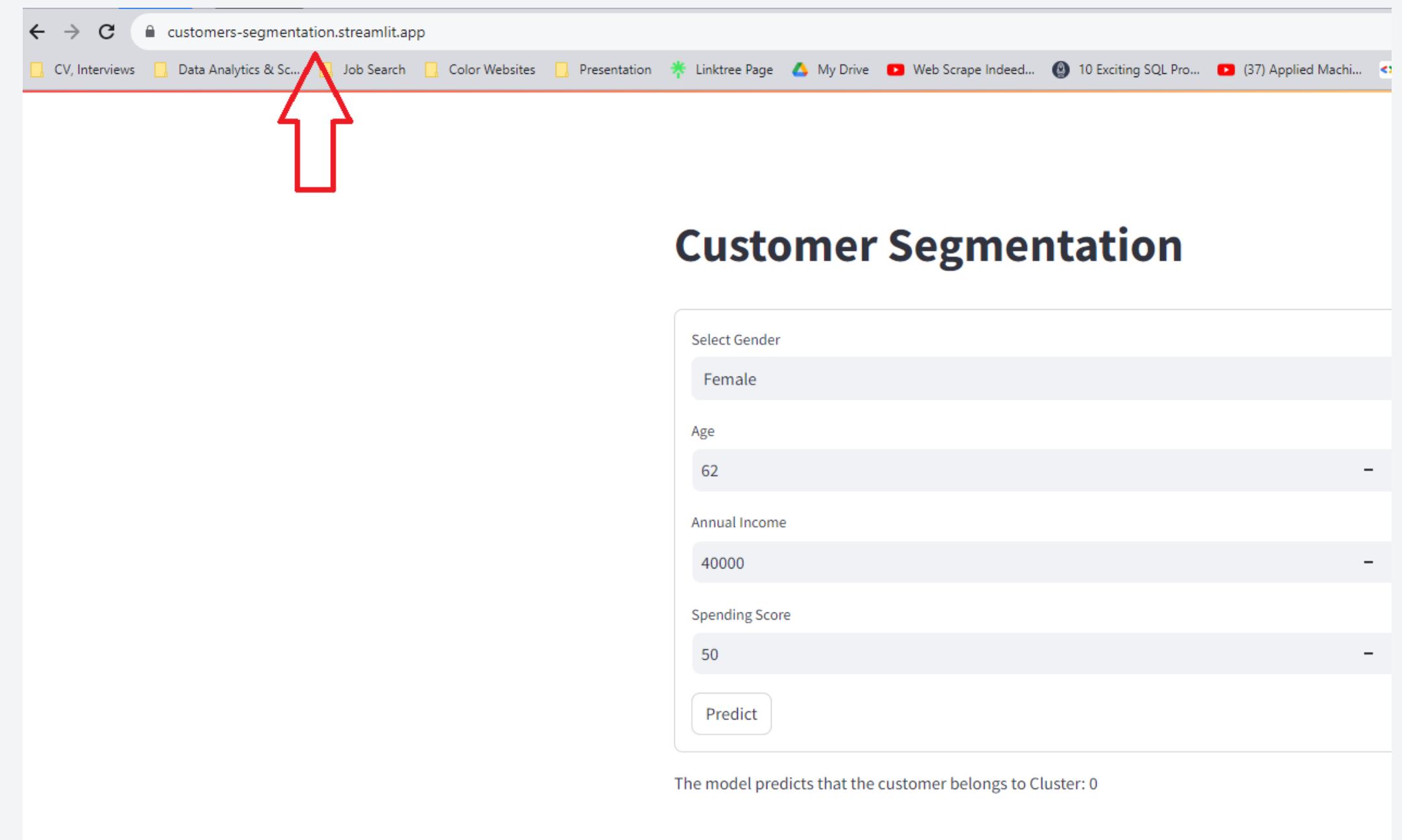
The model predicts that the customer belongs to Cluster: 2



CLOUD DEPLOYMENT

The app has also been deployed in cloud for any user to have access via the below link

[Link](#)





Thank You

Contact



SADIQ BALOGUN

MSc Student at Teesside University

 +44 7867 74996

 linktr.ee/sadiqbalogun

 Middlesbrough, UK

 ballosadiq@gmail.com

