

Customer Persona Generator

Chi-Yuan (Tina) Peng

2022 . 12 . 15



Table of Contents

1

Introduction

customer segmentation
and persona

2

Objectives

my plan and thoughts

3

Preprocessing

data cleaning and
preprocessing

4

Modeling

standard scaler,
KMeans, PCA

5

Segment Analysis

segment reports
analysis

6

Persona Generator

how I make the
persona generator





Introduction

1

What is **customer segmentation**?

- a process of analyzing customers' buying behavior and clustering customers into groups.

What is **persona**?

- a character that represents the behavior and features of a larger group of customers.

2

Objectives

Collect Data

Data Preprocessing

Modeling



DONE !

Make a Report

Segment Analysis



Segmentation



Preprocessing

3

Data collection

Data cleaning and preprocessing

Collect Data

from Kaggle (customer segmentation data)

Customer Info

- Annual Income
- Year Birth
- Marital Status
- Number of Kids
- Education Level

Buying Behavior

- Amount Spend on Wines
- Amount Spend on Fruits
- Amount Spend on Meats
- Amount Spend on Seafoods
- Amount Spend on Sweets
- Amount Spend on Gold Products
- Amount Spend on Deal Products

Data Preprocessing

New Columns

- Year Birth → Age
- Marital Status → Not Live Alone
- Number of Kids & Not Live Alone → Number of Family Members
- Marital Status & Number of Kids → Single Parent
- Amount Spent on Products → Total Spent
- Amount Spent on XX Products / Total Spent → Pr. Spent on XX Products

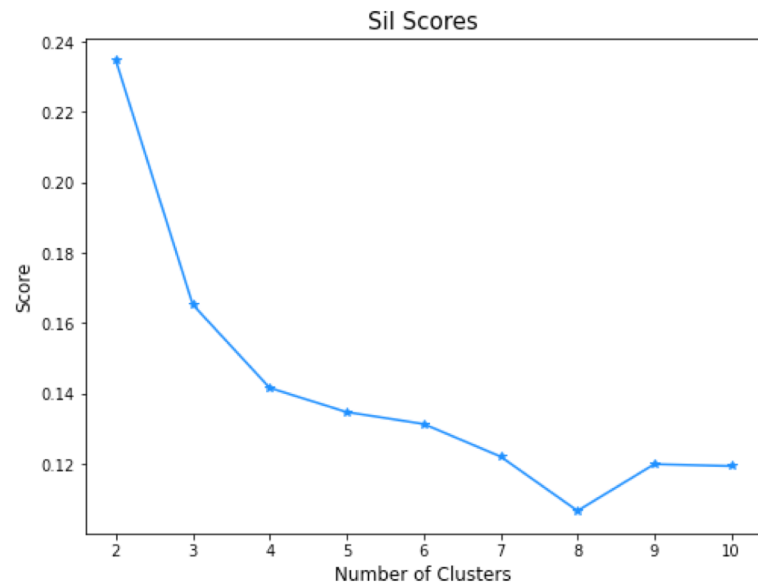
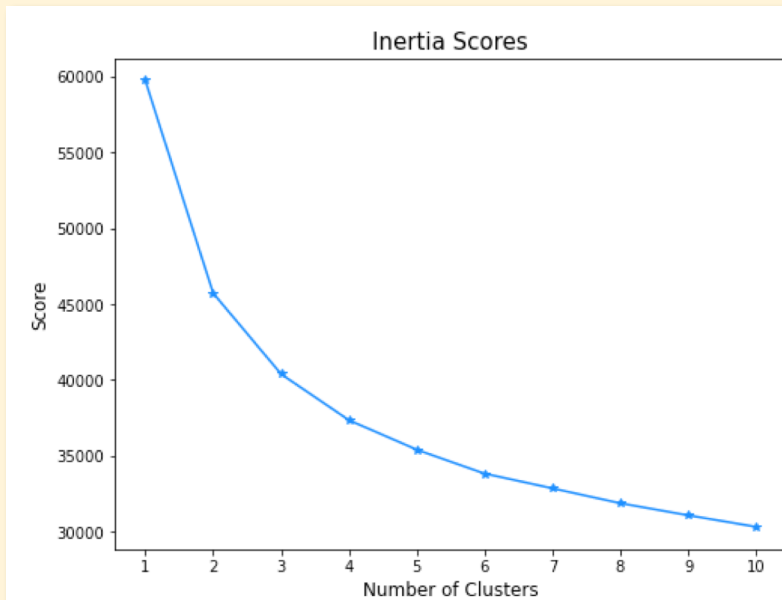


Modeling

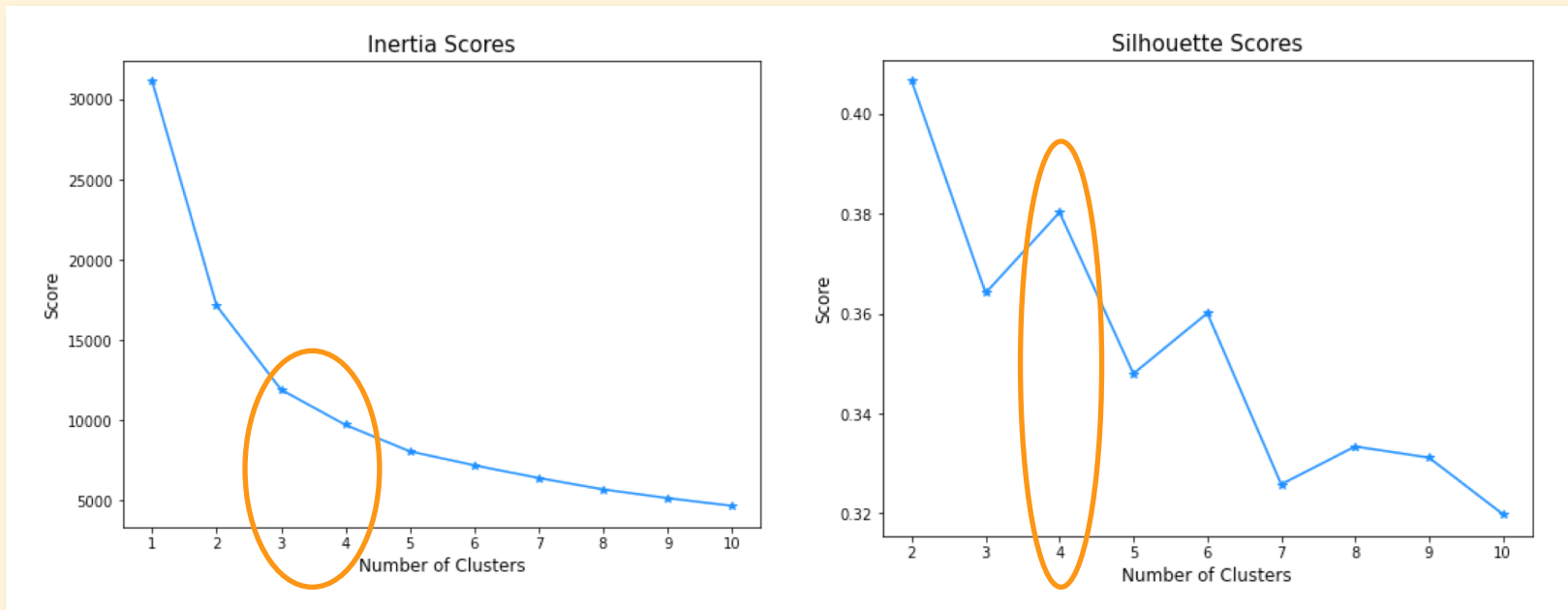
4

Unsupervised learning: **KMeans**

KMeans



PCA + KMeans





Segment Analysis

5

Traditional Segment Analysis

Group A – 29%

- 🛒 **All** of them are **married**
- 🛒 **Largest family** size
- 🛒 Prefer **wine** and **on-sale** products

Group B – 18%

- 🛒 **Single parents**
- 🛒 Most of them has at least one kid
- 🛒 Prefer **wine** and **on-sale** products



Group C – 30%




- 🛒 **Richest** group with highest spending
- 🛒 Most of them has **no kids**
- 🛒 Prefer **meat** products

Group D – 23%

- 🛒 **Lowest** income and spending
- 🛒 **Youngest** group
- 🛒 Prefer **seafood** and **gold** products

Automatic Segment Analysis

Group A – 29%

-  **All** of them are **married**
-  **Largest family** size
-  Prefer **wine** and **on-sale** products

Group A








- Group size: 29%
- Annual income: \$41,000 ~ 59,000
- Spending amount: \$92 ~ 642
- Age: 48 ~ 65
- Marital status: 99% are married
- Family size: 2 ~ 3
- Product preference: **Wine, On-sale**










Persona Generator

6

Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

To randomly generate a name ...

Step 1:

web scraping the **first name** from SSA .



Top Names Over the Last 100 Years

Popular Baby names
Top names by decade
Top 5 names in each year
Popular Names by State

Top Names Over the Last 100 Years

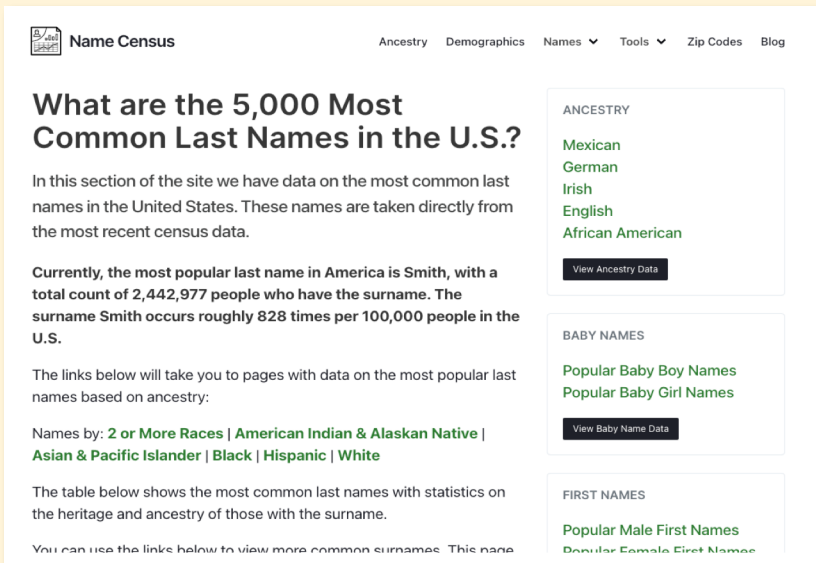
The following table shows the 100 most popular given names for male and female babies born during the last 100 years, 1922-2021. For each rank and sex, the table shows the name and the number of occurrences of that name. These time-tested popular names were taken from a universe that includes 177,238,032 male births and 172,053,794 female births.

Please note that popular names listed below are not necessarily *consistently* popular in every year. For example, the name James, ranked as the most popular male name over the last 100 years, has been ranked as low as number 19. Similarly, the most popular female name in the table, Mary, ranked as low as 133.

Rank	Males		Females	
	Name	Number	Name	Number
1	James	4,663,035	Mary	3,124,584
2	Robert	4,407,377	Patricia	1,555,054
3	John	4,403,862	Jennifer	1,469,031
4	Michael	4,340,931	Linda	1,448,283
5	David	3,564,313	Elizabeth	1,411,916
6	William	3,524,670	Barbara	1,391,959

Step 2:

web scraping the **last name** from Name Census.



Name Census

Ancestry Demographics Names Tools Zip Codes Blog

What are the 5,000 Most Common Last Names in the U.S.?

In this section of the site we have data on the most common last names in the United States. These names are taken directly from the most recent census data.

Currently, the most popular last name in America is Smith, with a total count of 2,442,977 people who have the surname. The surname Smith occurs roughly 828 times per 100,000 people in the U.S.

The links below will take you to pages with data on the most popular last names based on ancestry:

Names by: **2 or More Races** | **American Indian & Alaskan Native** | **Asian & Pacific Islander** | **Black** | **Hispanic** | **White**

The table below shows the most common last names with statistics on the heritage and ancestry of those with the surname.

You can use the links below to view more common surnames. This page

ANCESTRY

- Mexican
- German
- Irish
- English
- African American

[View Ancestry Data](#)

BABY NAMES

- Popular Baby Boy Names
- Popular Baby Girl Names

[View Baby Name Data](#)

FIRST NAMES








- Popular Male First Names
- Popular Female First Names

Customer Persona Draft

Male first name x 100

Female first name x 100

Last name x 100

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			








Customer Persona Draft

Male first name x 100








Female first name x 100

Last name x 100








Jennifer Morris

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Customer Persona Draft

NAME Gender Age Group Size

Photo

Profile

Preferences




















Beer Glass, Chicken Drumstick, Crab, Pizza Slice

The percentage of the group in the whole dataset.

Ex.

Group A's group size is 30% means that 30% of the customers in this whole dataset are labeled group A.

Customer Persona Draft








NAME	Gender	Age	Group Size						
<div>Photo</div> 	<div>Profile (ex. family size, income, spent)</div>								
<div>Product preference</div> <table><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr></table>									
									
									

To randomly generate a persona photo ...








Python package: [python-avatars](#)



Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Customer Persona Draft

NAME	Gender	Age	Group Size
Photo 	Profile (ex. family size, income, spent)		
Product preference			
			
			

Persona Generator

Display through Streamlit



Customer Persona

Group A Group B Group C Group D Summary

NAME	GENDER	AGE	GROUP SIZE
Raymond Chavez	Male	52	29%

CUSTOMER PROFILE



- Annual income: \$50,000
- Spending level: Medium
- Marital status: Married
- Number of family member: 3
- Number of kids: 1

PRODUCT PREFERENCE

Wine

Fruits

Meats

Seafood

Sweets

Gold

On-sale Products



Thank you
Any questions?

Citations



- <https://www.kaggle.com/code/karnikakapoor/customer-segmentation-clustering/data>
- <https://pypi.org/project/python-avatars/>
- https://github.com/ibonn/python_avatars
- <https://streamlit.io/>
- <https://www.ssa.gov/oact/babynames/decades/century.html>
- <https://namecensus.com/last-names/>
- <https://www.gartner.com/en/sales/glossary/customer-segmentation>
- <https://xd.adobe.com/ideas/process/user-research/putting-personas-to-work-in-ux-design/>
- <https://excalidraw.com/>
- <https://slidesgo.com/theme/flat-style-buyer-persona-infographics#search-persona&position-5&results-12&rs=search>