

Identifying Product Preferences for The Cupping Room's New Café

Diane Zhu

1. Introduction

a. Background

The Cupping Room is a modern café located in Canberra, Australia, providing high-quality coffee experiences to customers for the past 10 years. The Cupping Room is designed by the team from Ona Coffee, which is Canberra's largest specialty coffee roaster, to share passion and knowledge about coffee with customers. The Cupping Room positions itself as a diverse café community and tries to open new café in Canberra to serve more customers.



Figure 1. The Cupping Room Dining Area

b. Problem Identification

The purpose of this report is to help the Cupping Room to find the best location for a new storefront and design the new menu based on historical customer orders.

2. Data Pre-processing

a. Data Collection

The location data of café in Canberra is generated by Foursquare which listed all the café around The Cupping Room within 500 kilometres as shown in Figure 2. Since the transactional data for The Cupping Room is confidential, for the purpose of this project, a fictional dataset from Kaggle will be used to represent a placeholder of the model development.

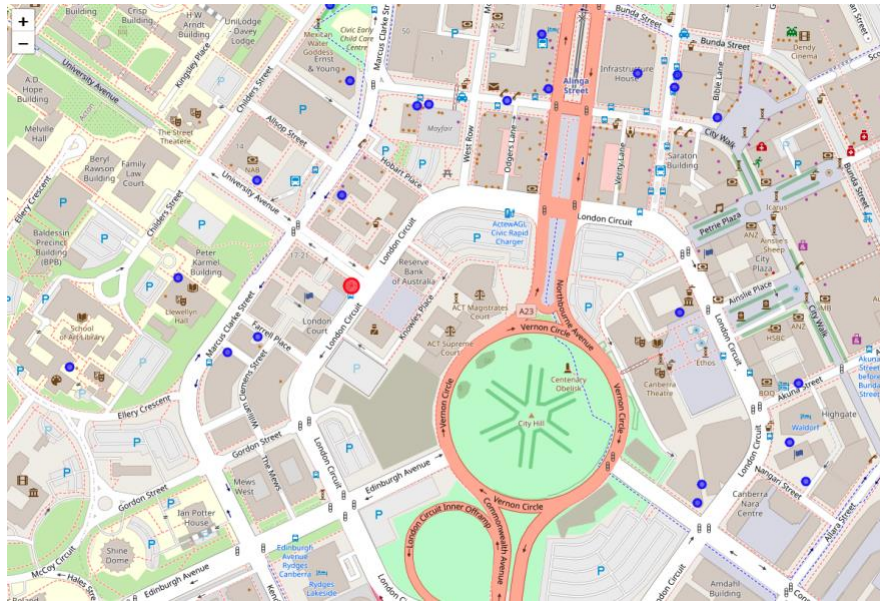


Figure 2. Existing Café around The Cupping Room

b. Data Preparation

As the fictional dataset from Kaggle is clean as shown in the figure below, the only thing required here is parameters selection. Customer name and delivery location were not relevant to the predictive model, hence, will be dropped from the dataset.

	customer	drink	category	delivery_location	time	Gender	Age	Annual_Income
0	Mr Dryden	Latte	Coffee	I9	2019-08-15 07:36:47	Male	19.0	15000.0
1	Jane Olivero	Flat White	Coffee	H Block Admin	2019-08-15 07:36:20	Male	21.0	15000.0
2	Alisha Watson	Ever Green	Smoothie	I19 (next to Drama office)	2019-08-15 07:33:56	Female	20.0	16000.0
3	Elena	Berry Bliss	Smoothie	I18A	2019-08-15 07:06:30	Female	23.0	16000.0
4	Tetias Goh	Chocolate Fix	Smoothie	Music Office	2019-08-15 06:50:58	Female	31.0	17000.0
5	Maree Bowen	Green Tea	Tea Selection	I7	2019-08-15 06:49:31	Female	22.0	17000.0
6	Trina	Latte	Coffee	Hive 3 or G1	2019-08-15 06:38:14	Female	35.0	18000.0
7	Kathy Dyer	Chai	Tea Selection	I19	2019-08-15 06:36:16	Female	23.0	18000.0
8	Chelsey Eunson	Latte	Coffee	I20	2019-08-15 06:34:35	Male	64.0	19000.0
9	Sue Patterson	Flat White	Coffee	Music Office	2019-08-15 06:33:43	Female	30.0	19000.0
10	Steve Roberts	Flat White	Coffee	K1/7	2019-08-15 06:33:19	Male	67.0	19000.0
11	Ruan	Flat White	Coffee	I block staff room	2019-08-15 06:28:50	Female	35.0	19000.0
12	Mrs Anderson-Morse	Flat White	Coffee	K2	2019-08-15 06:26:30	Female	58.0	20000.0
13	Mr Hedstrom	Flat White	Coffee	K11	2019-08-15 06:16:53	Female	24.0	20000.0
14	Rod Frohmuller	Flat White	Coffee	K block staffroom	2019-08-15 06:06:22	Male	37.0	20000.0
15	Mrs T	Chai	Tea Selection	K8	2019-08-15 06:02:26	Male	22.0	20000.0
16	Clare Hannaway	Berry Bliss	Smoothie	Prac Room 1 music hallway across from music of...	2019-08-15 05:56:40	Female	35.0	21000.0
17	Petalia Humphreys	Cappuccino	Coffee	K11	2019-08-15 05:55:23	Male	20.0	21000.0
18	Rob Kelly	Flat White	Coffee	K7 Period 2	2019-08-15 05:52:29	Male	52.0	23000.0

Figure 3. Original Dataset

c. Hyperparameter Tunning

Due to the limitation of the dataset, age, gender, and annual income are chosen to predict the preference of The Cupping Room new menu. As the purpose of the model is to find preference among customers, linear regression and logistic regression are not appropriate.

3. Exploratory Data Analysis

a. Customer Distribution

Based on Figure 4 as shown below, the annual income for existing customers for The Cupping Room is between 15,000 to 45,000, which suggests that The Cupping Room's main target market is focusing on low to medium income group of people.

Also, according to Figure 5, The Cupping Room attracts more female customers than male customers, with 59.7% and 40.3% respectively. Therefore, when selecting new location for a new café, The Cupping Room should aim for places with potentials of more female customers.

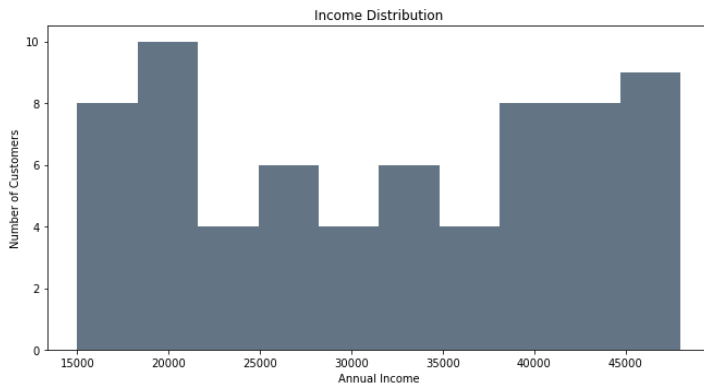


Figure 4. Customer Annual Income Distribution

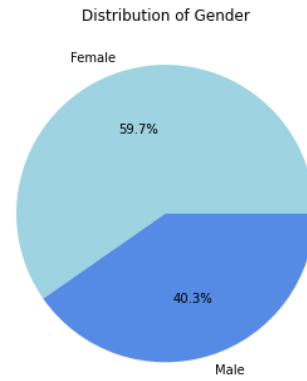


Figure 5. Distribution of Gender

b. Order Selection

After generating the distribution for drinks order of The Cupping Room as shown in the Figure 6, it can be inferred that flat white is the most popular drink in The Cupping Room, and green tea is the least popular drink. Based on this visualisation, The Cupping Room can determine the direction of new drink menu development. In addition, according to Figure 7, customers of The Cupping Room prefer coffee and smoothie than tea selections.

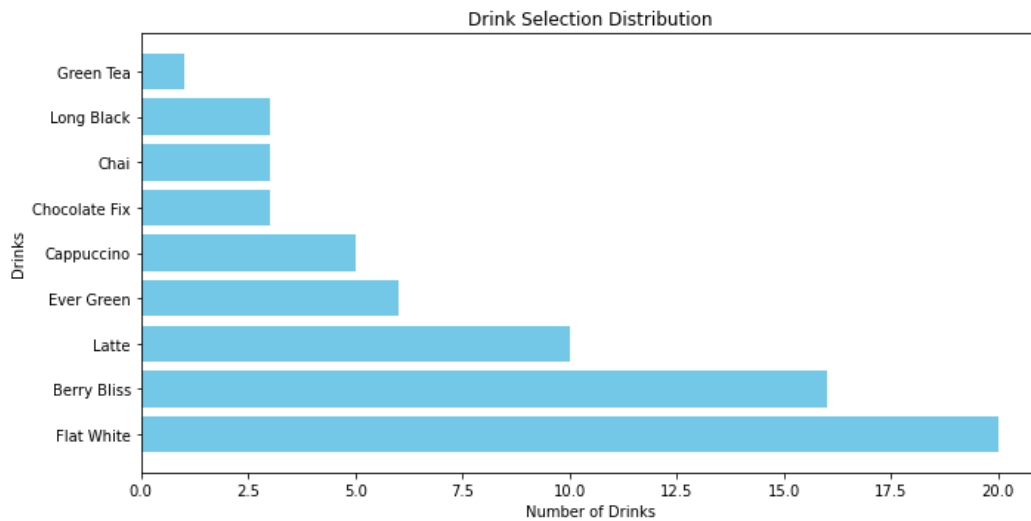


Figure 6. Drink Selection Distribution

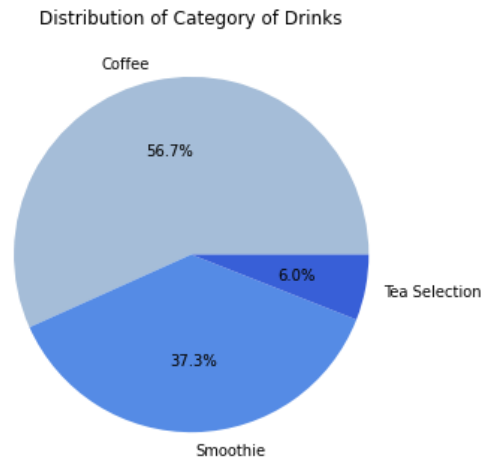


Figure 7. Distribution of Drink Category

4. KNN Model Prediction

To determine `n_neighbors` for the KNN model, I used `GridSearchCV` to repeat the process of fitting the train set data into the KNN model and test the performance using `n_neighbors` values from 1 to 50. And as the training set data will be changing all the time, I set the variable of `best_n_neighbors` to automatically fit each run. Then, I created the KNN model using the train dataset with `n_neighbors` equal to the result from `GridSearchCV`, and plot the result using three 2-D graphs to represent the distances for all features to compare the actual test values and the predicted values. As shown in Figure 8, the prediction is not that accurate for Tea Selection due to the lack of enough sample of data.

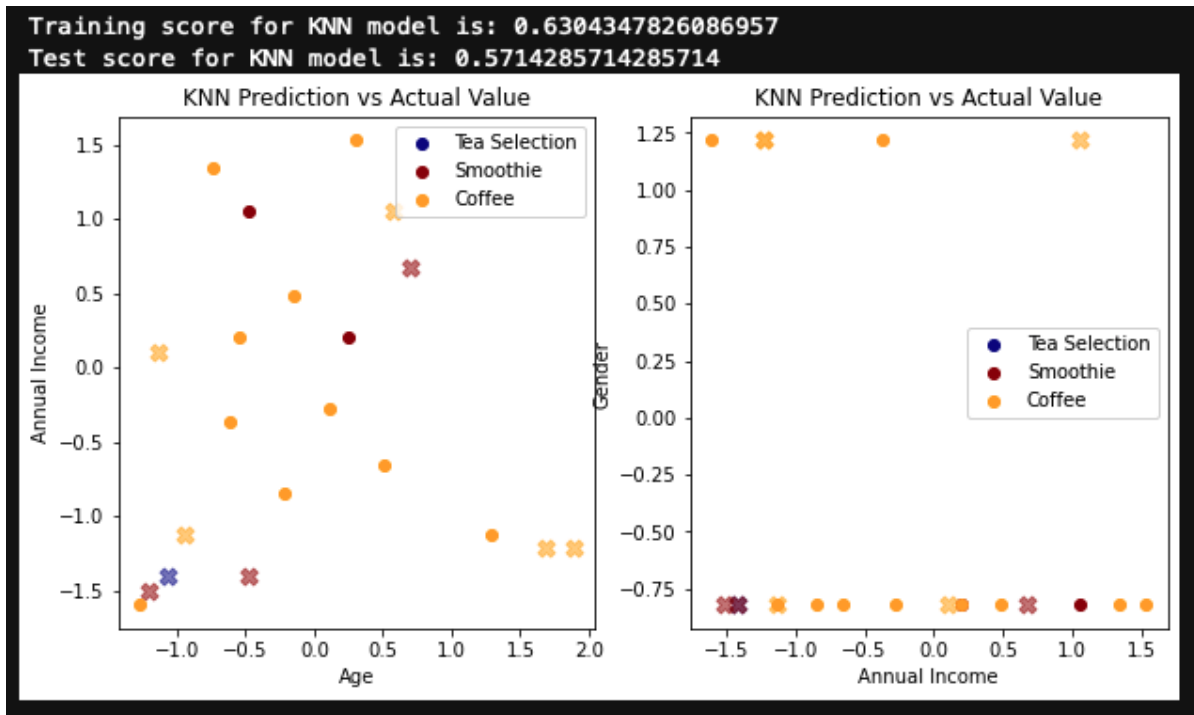


Figure 8. KNN Prediction vs Actual Value

The strengths of this KNN model are its high scores and accuracy for most classes, which can be used to easily predict the class. However, one major pitfall for this model is that it cannot work well with classes with little training data. To be more specific, there are originally very little objects with Tea Selection in the training dataset, which makes the prediction for this class not as accurate as Coffee and Smoothie classes which have many objects.

5. Conclusion and Recommendation

Although the fictional dataset has its limitations when implementing the KNN model, it is still possible to find out the pattern of customer behaviours for The Cupping Room. First, it is highly recommended to develop more coffee products for the new café, and tea products may not be as favourable as coffee or smoothie. Second, customers with a higher annual income may prefer

coffee over smoothie and tea. Therefore, when selecting new location for the café, The Cupping Room should take the surrounding offices or business areas into consideration.