# Finding the Best Place to Open a New Juku Campus

HIKARU HASHIMOTO

2021/JAN.

# Introduction

2021/1/8

# Juku

Juku is a kind of private fee-paying school in Japan for children to learn subjects such as maths or English after their regular school hours.

Students study hard in jukus offering supplementary lessons to students to pass entrance exams.

# Business Problem

The juku I work for is looking for a place to build a new Campus in Tokyo.

We want a similar area to Shibuya where one of our campus is placed because our business is going well now.

# Data

# Dataset and Preprocessing

Dataset:

4005 zip codes of Tokyo retrieved from Japan Post.

Preprocessing:

1. Obtaining the geographical coordinates based on the zip codes.

-We obtained 1033 geographical coordinates successfully out of 4005 zip codes.

2. Collecting venues information around the geographical coordinates.

# Methodology

1. COSINE SIMILARITY

2. K-MEANS ALGORITHM

# 1. Cosine Similarity

Reducing the size of the dataset by the cosine similarity.

The cosine similarity computes similarity between X and Y as;

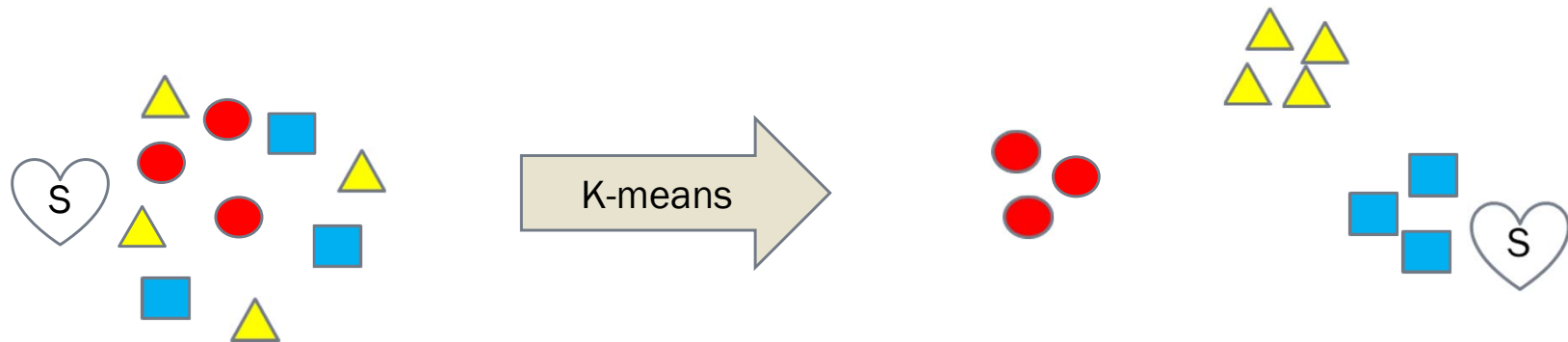$$Cosine\_Similarity(X, Y) = <X, Y> / (||X||*||Y||)$$

We choose top 40 similar areas to our Shibuya campus as candidate areas.

# 2. K-Means Algorithm

Find candidate areas by the k-means algorithm.

The k-means algorithm makes clusters of data based on the features of it.

The cluster which has Shibuya area in it is the target. Areas in the cluster are candidate locations (the cluster of squares in the figure below).

# Result

# Results; Cluster

Three areas fall in the same cluster as Shibuya area where we have a campus.

These three are similar to Shibuya in respect of the cosine similarity and k-means. These two methods consider different aspects of the data. Therefore, these locations are great candidates.

| zip_code | Area | cos_score |
|---|---|---|
| 1500002 | SHIBUYA | 1.000000 |
| 1540024 | SANGENJAYA | 0.801115 |
| 1500011 | HIGASHI | 0.773443 |
| 1000001 | CHIYODA | 0.743232 |

# Results; Location

Blue circles are candidates in the map;
Sangenjaya, Higashi, Chiyoda from left to
light.

Red circles are our campus.

Sangenjaya and Higashi areas seem to be
covered by three of our campuses while
Chiyoda area is not.



2021/1/8

# Conclusion

# Conclusion

In this project,

-We search for candidate areas to build a new juku campus.

-We use the cosine similarity and k-means algorithm.

- Sangenjaya, Higashi, Chiyoda areas are the best three candidates in our data. Chiyoda is the best area because of our current campuses' coverage.