

# Pizza

## Group members:

Jerrell U2220449D

Melvin U2220256B

Winfred U2221912F





1

# Our Motivation



# Our goal

We are business owners who  
want to open a new pizza  
franchise in the US

## Factors affecting success

1. Location
2. Pricing
3. Competition

The background is a light cream color decorated with various pizza-related illustrations. At the top left, there's a small pizza slice and a mushroom. At the top center, a red heart-shaped cookie cutter and a single olive. At the top right, a whole pizza with toppings and another olive. On the left side, a large pizza server with a green blade and a black handle. On the right side, a whole pizza with toppings. At the bottom left, a mushroom and two olives. At the bottom center, a pizza slice with heart-shaped toppings. At the bottom right, a pizza cutter with a green wheel and a black handle, a mushroom, and an olive. A red heart-shaped cookie cutter is also at the bottom right.

## **Dataset used :**

1. Pizza Restaurants US by Chirag Desai from Kaggle.
2. List of United States cities by population from wikipedia.



2

**Setting the  
stage**

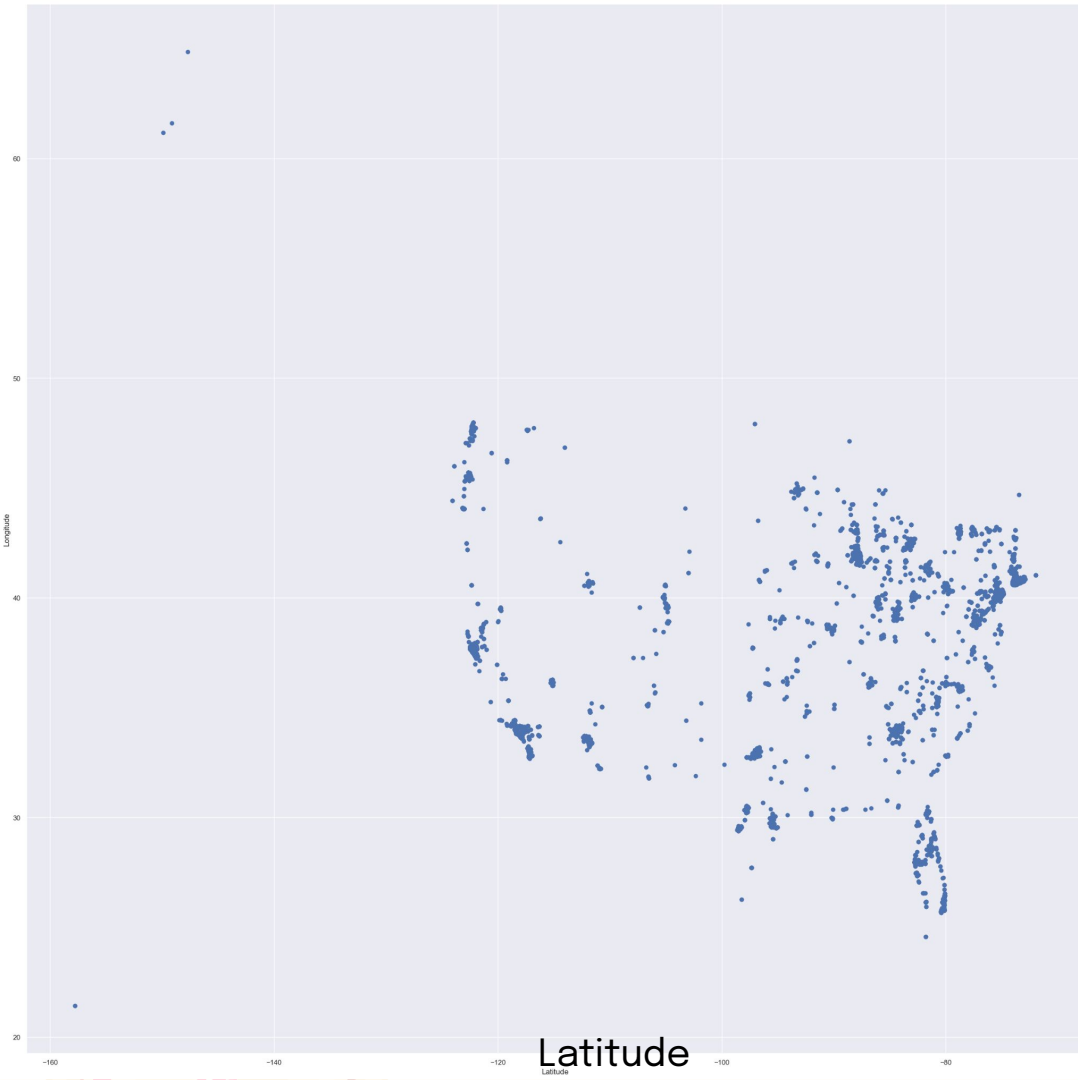
The slide features several decorative illustrations of pizzas and pizza-related items. At the top left, a small mushroom is shown. To its right is a large pizza with various toppings including pepperoni, mushrooms, and olives. Further right is a small pizza with a single slice being pulled out, revealing a thick crust. On the top right is another large pizza with a variety of toppings. In the bottom left corner, a portion of a pizza is visible. At the bottom center, there is a large, empty, light-brown oval shape. In the bottom right corner, there is a small pizza with a thick crust and a single slice being pulled out, revealing a thick crust. The background is a solid light beige color.

# Exploratory Data Analysis

- Scatter plot of pizza shops and cities by location
- Bar plot of cities by population

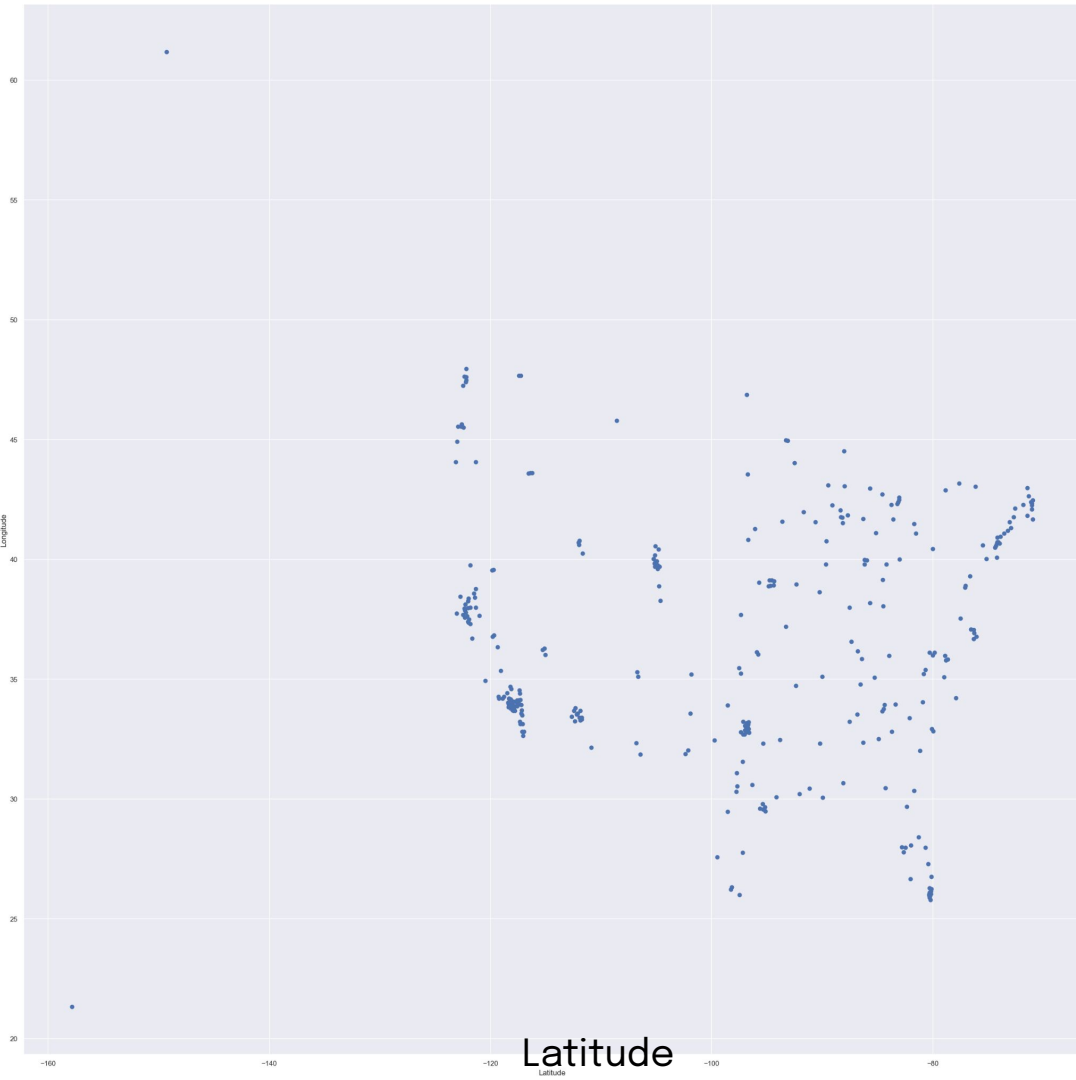
# Pizza shops

Longitude



# Cities

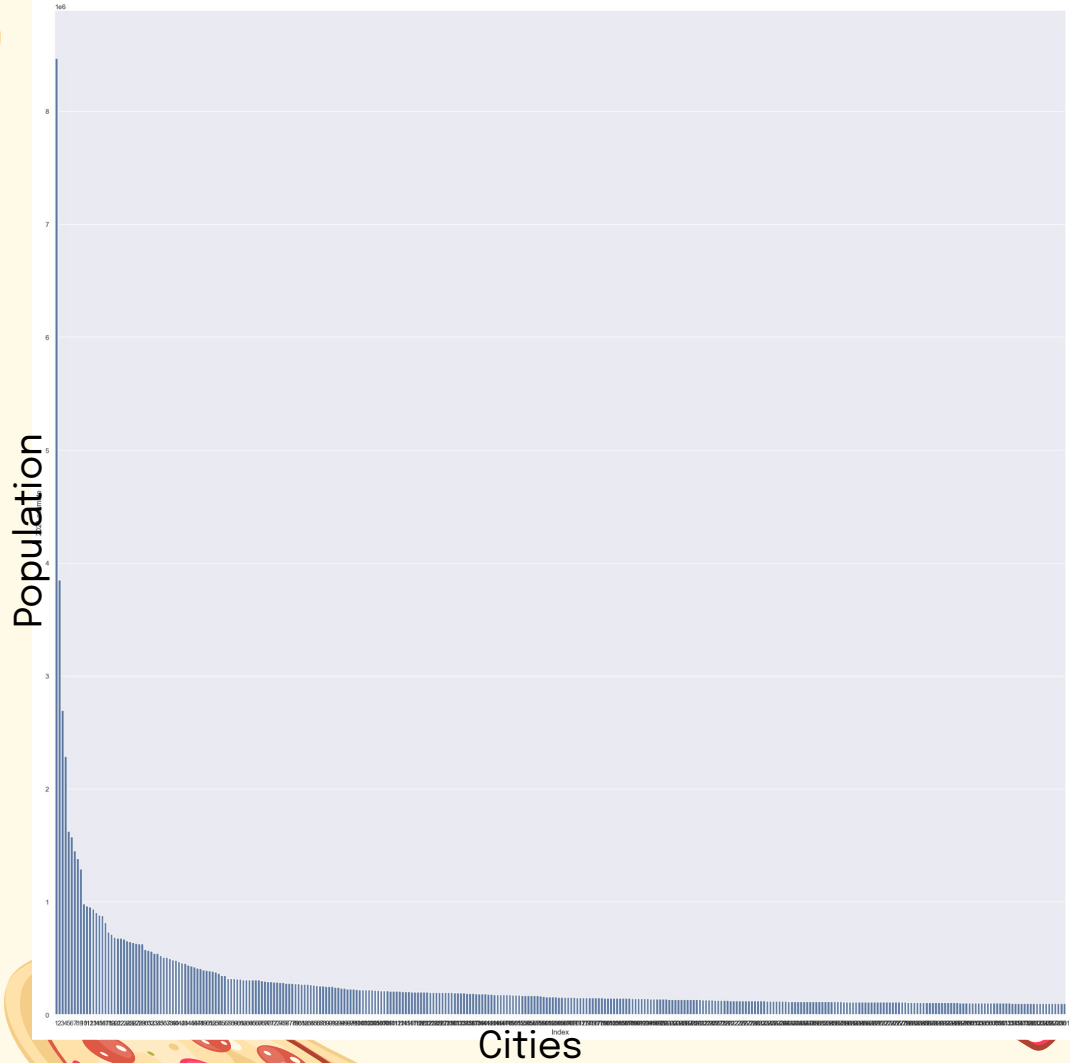
Longitude

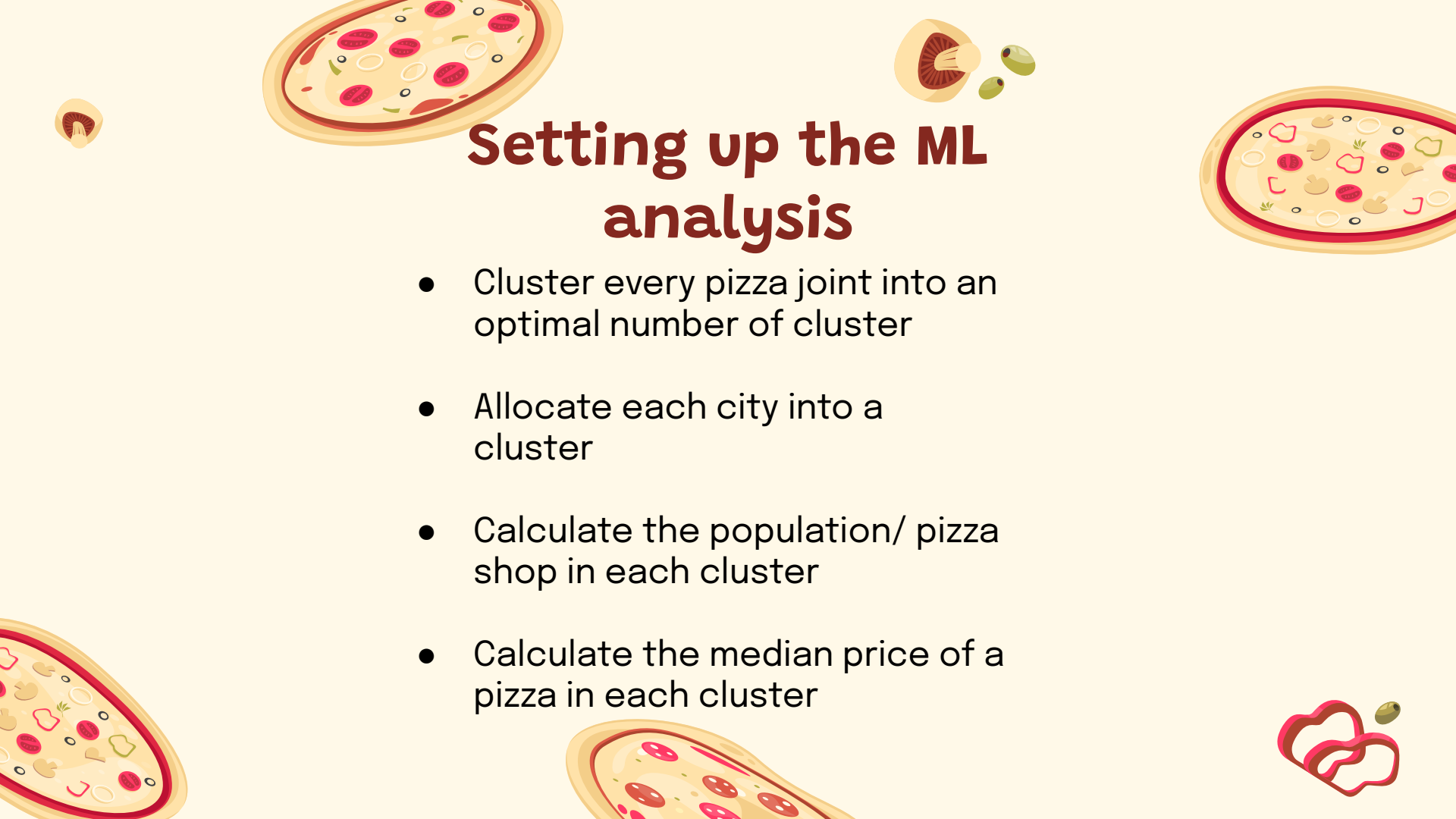


Latitude



# Bar plot for cities



The slide features several decorative illustrations of pizzas and pizza toppings. At the top left is a whole pizza with pepperoni, mushrooms, and olives. To its right is a hand placing a slice of pepperoni onto a pizza. Further right is another whole pizza with various toppings including mushrooms, olives, and heart-shaped toppings. At the bottom left is a partial view of a pizza with heart-shaped toppings. At the bottom center is a partial view of a pizza with pepperoni. At the bottom right is a hand placing a slice of heart-shaped topping onto a pizza. The background is a solid light beige color.

# Setting up the ML analysis

- Cluster every pizza joint into an optimal number of cluster
- Allocate each city into a cluster
- Calculate the population/ pizza shop in each cluster
- Calculate the median price of a pizza in each cluster



# Cleaning & preparation


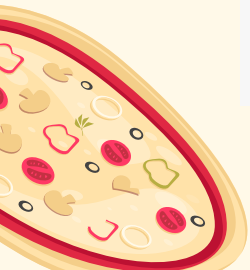
1

## Longitude & Latitude

- Identify US mainland geographical longitude and latitude
- Remove outliers ( such as Hawaii & Alaska )

```
cleaned = pizza[(pizza['latitude'] < 50) & (pizza['longitude'] > -140)]  
long_clean = cleaned["longitude"]  
lat_clean = cleaned["latitude"]
```

```
cities = cities[(cities['Latitude'] < 50) & (cities['Longitude'] > -140)]  
long_clean = cities['Longitude']  
lat_clean = cities["Latitude"]
```





# Cleaning & preparation

## 2

### Price of Pizzas

- Remove the duplicates
- Calculating the median price / shop

```
pizza_clean = cleaned.drop_duplicates(subset=['address'])
```

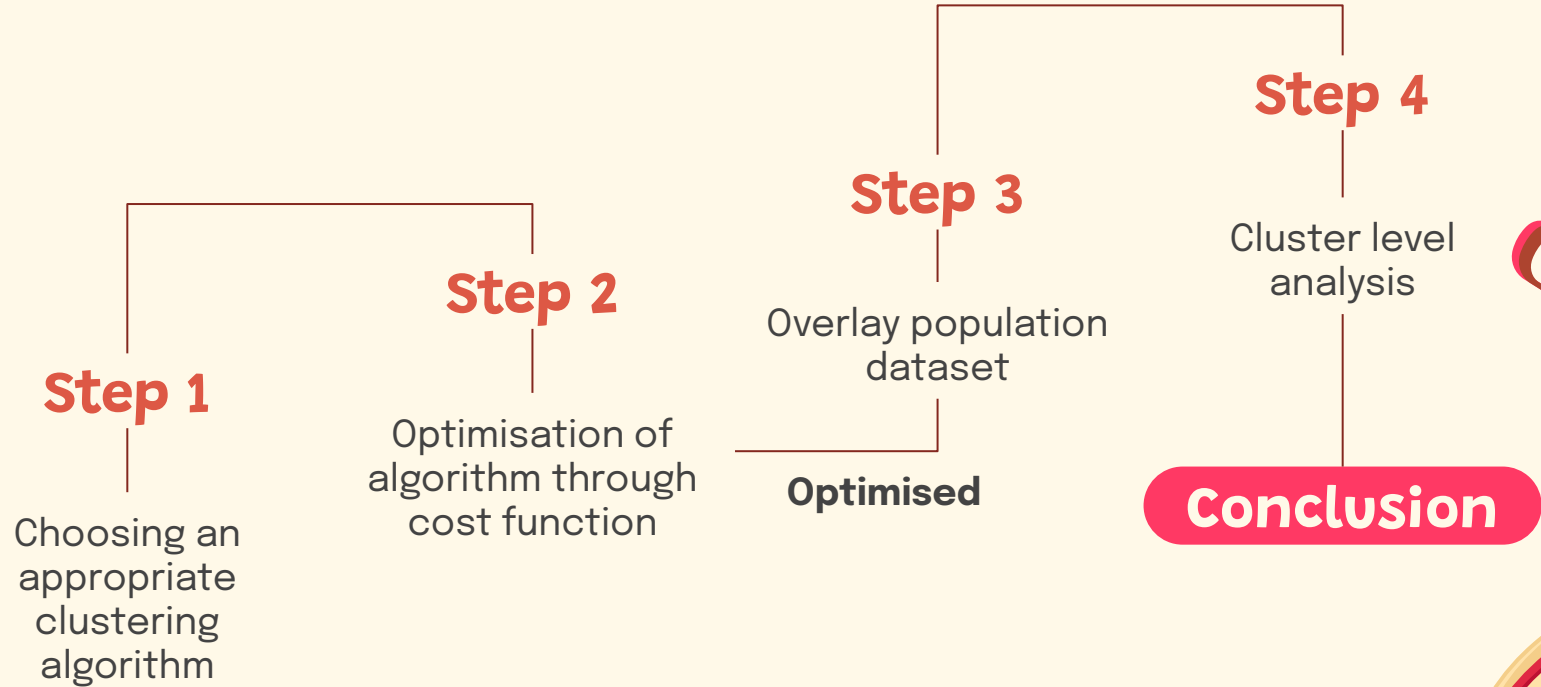
```
grp = cleaned.groupby('address')  
median_price = []  
for add in cleaned.address.unique():  
    median_price.append(grp.get_group(add)["menus.amountMax"].median())  
  
pizza_clean['Median_Price'] = median_price
```

3

# Core Analysis



# Process



# Clustering Algorithms

## K-means

Centroid-based  
algorithm

## Hierarchical

Non-spherical  
clusters

## Fuzzy

Multi-cluster  
assignment

## Density

Density of data  
points and  
handles outliers  
well

## Model

Probabilistic  
models to  
identify clusters

## Spectral

Data with  
complex  
structures

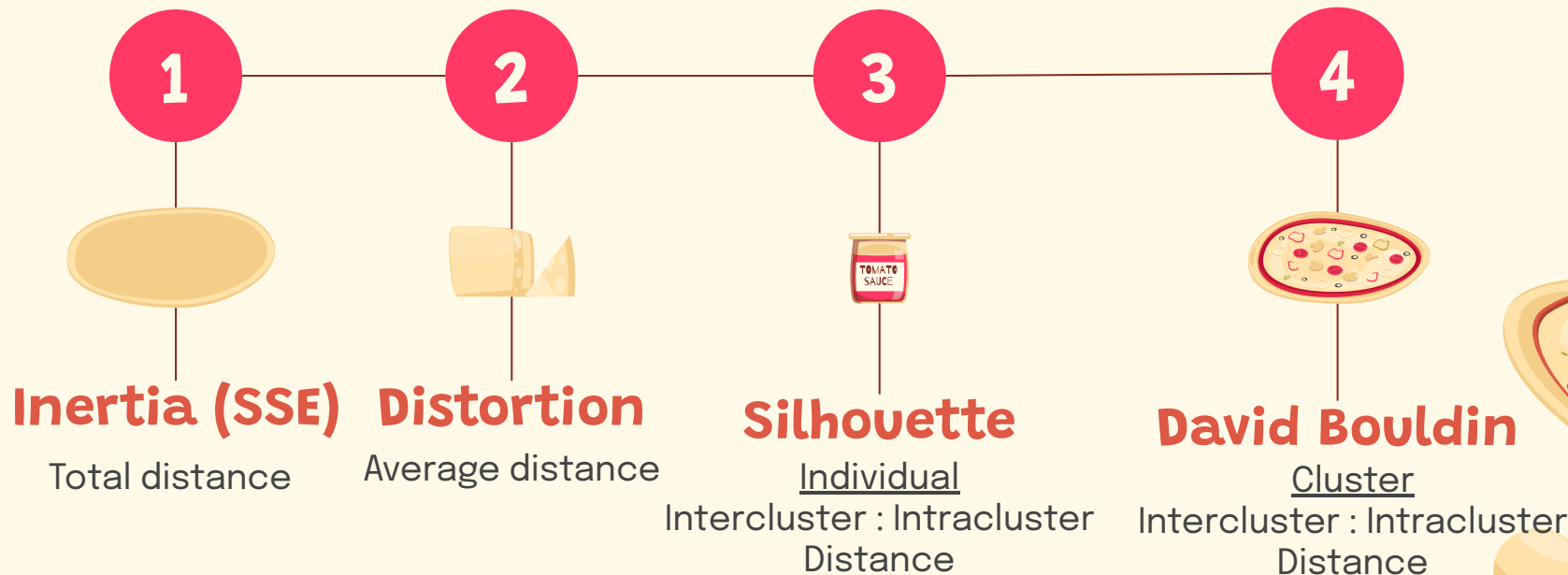
# K-Means

## Benefits over Others

- Simple to implement
- Easily interpreted and analysed
- Centroid-based algorithm allows for overlay of secondary dataset (population)
- Can form clusters quickly and efficiently for large datasets



# Cost Functions / Metrics

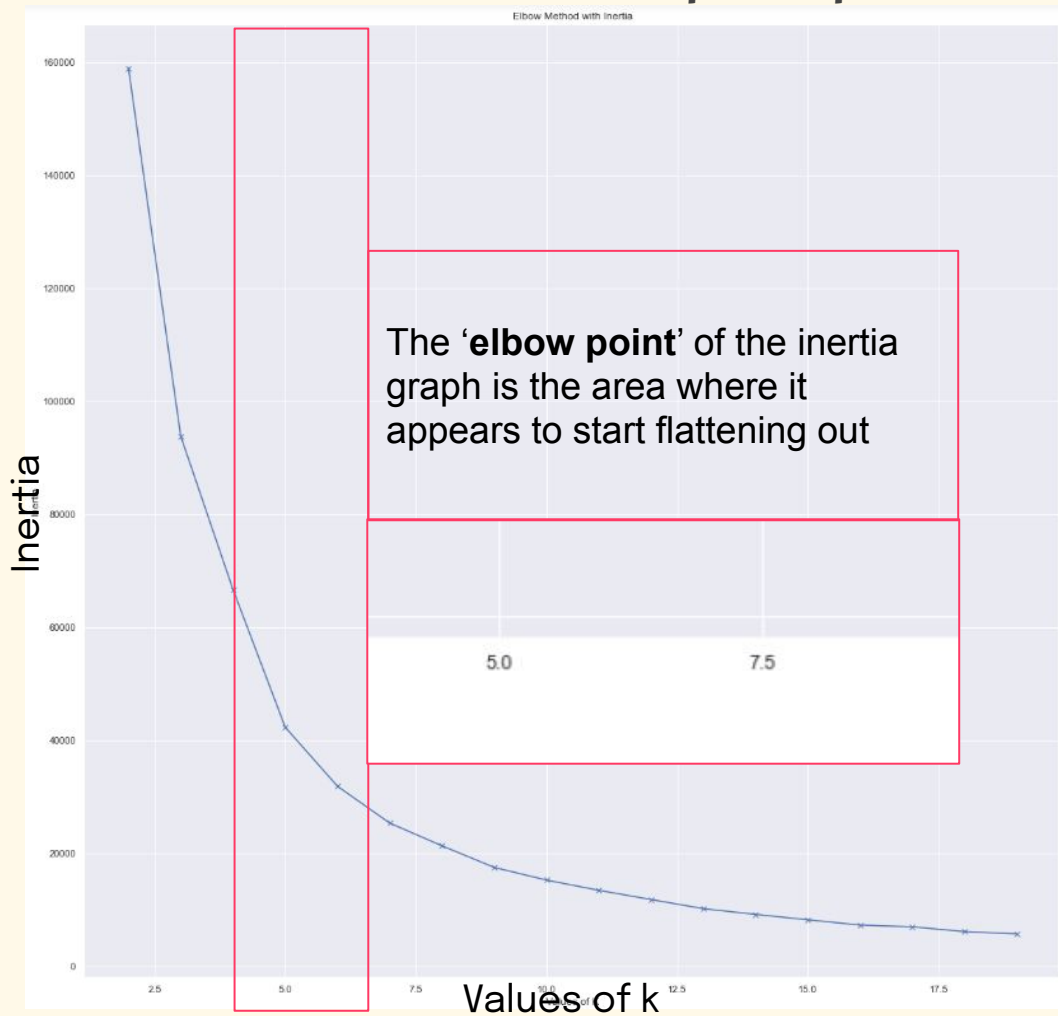


# Inertia

## Description

- **Sums** the distance between all data points and its assigned centroid
- Look for '**elbow point**'

Ideal k [4:8]

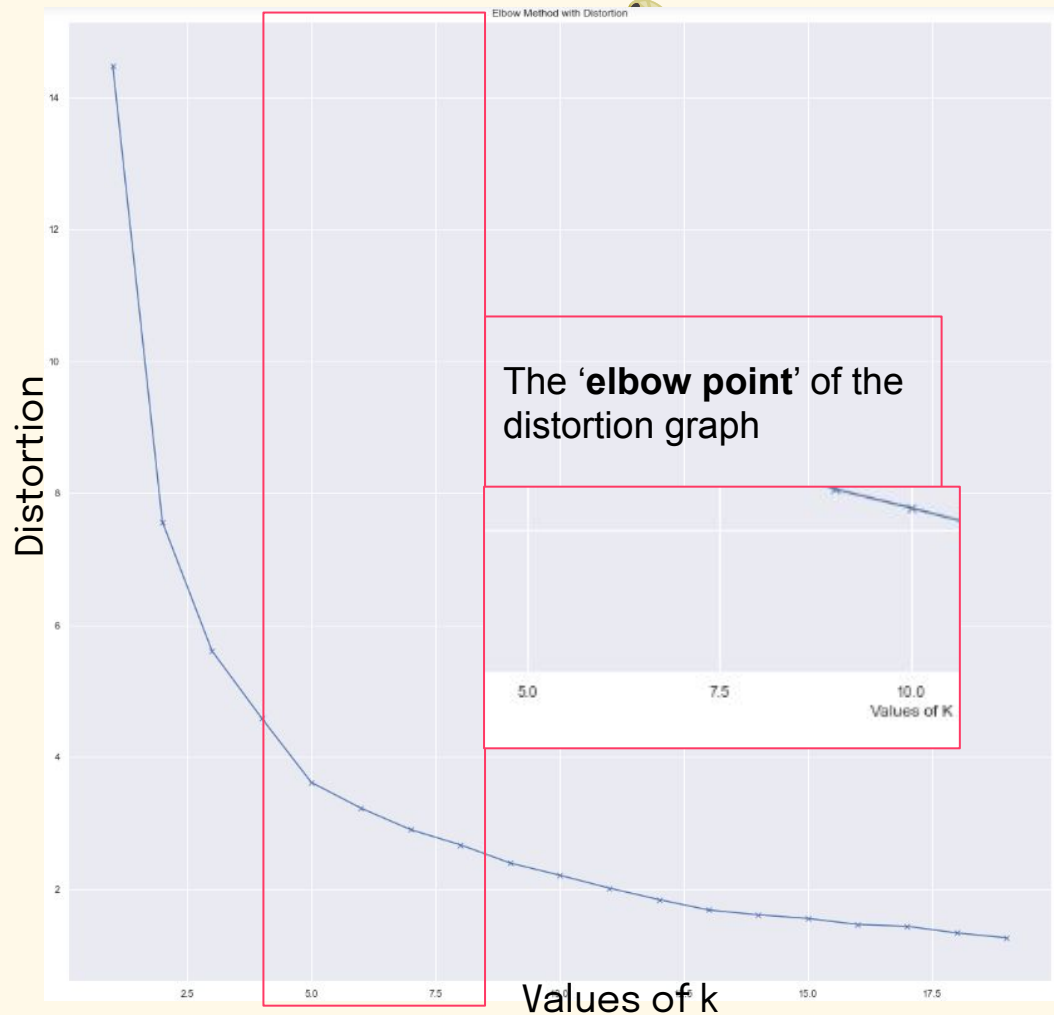


# Distortion

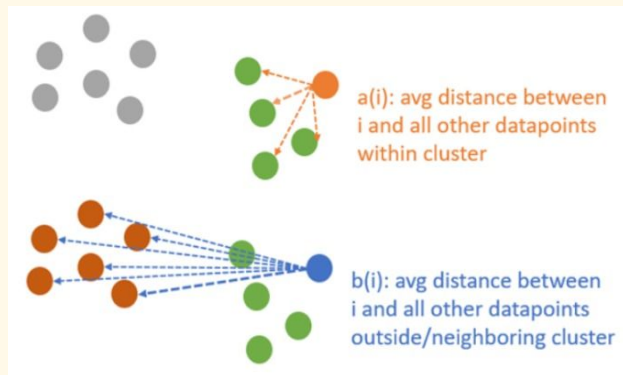
## Description

- **Averages** the distance between all data points and its assigned centroid
- Look for '**elbow point**'

Ideal k [5:10]



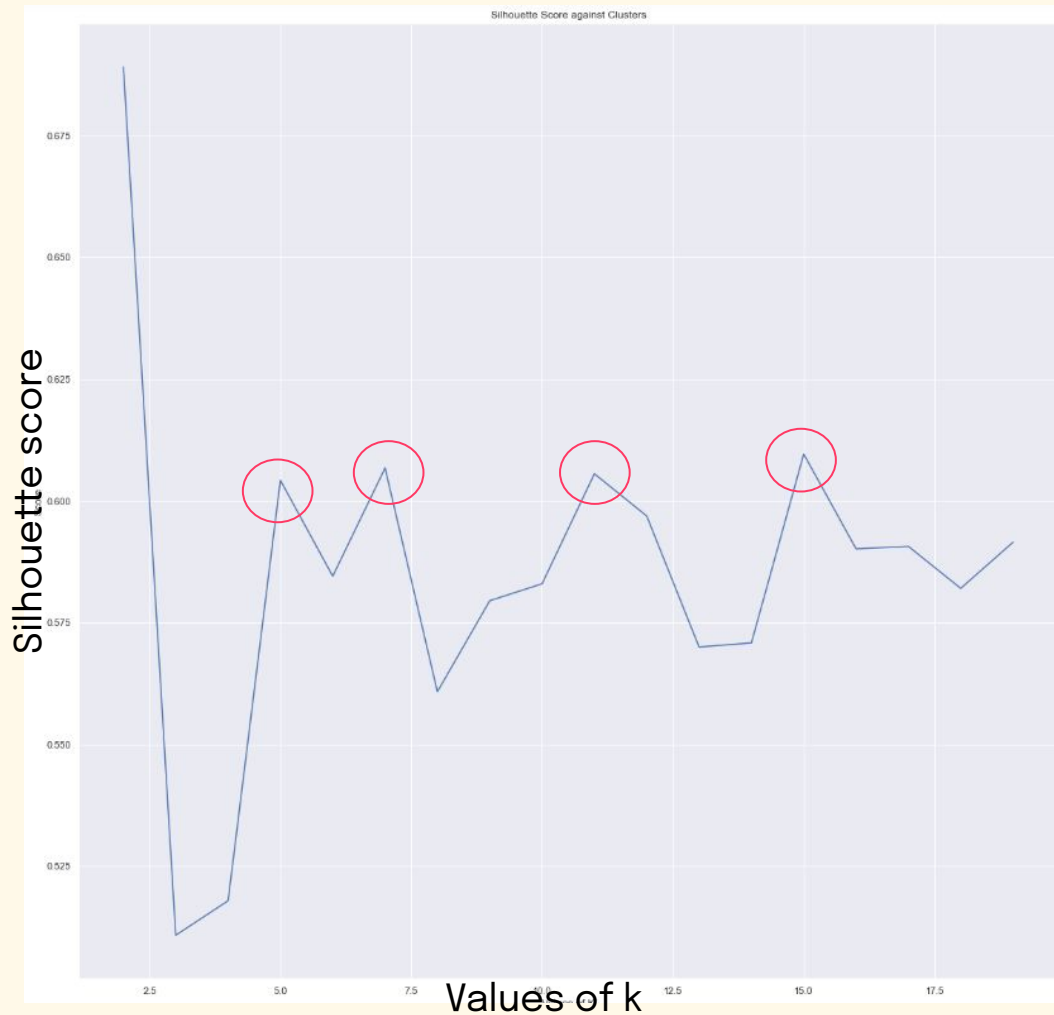
# Silhouette



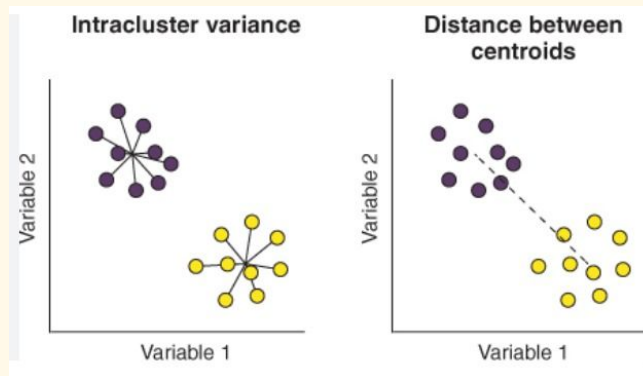
## Description

- Measure of similarity between **data points**
- Compares *intercluster* to *intracluster* distances
- Look for **highest points**

**Ideal k [5,7,11,15]**



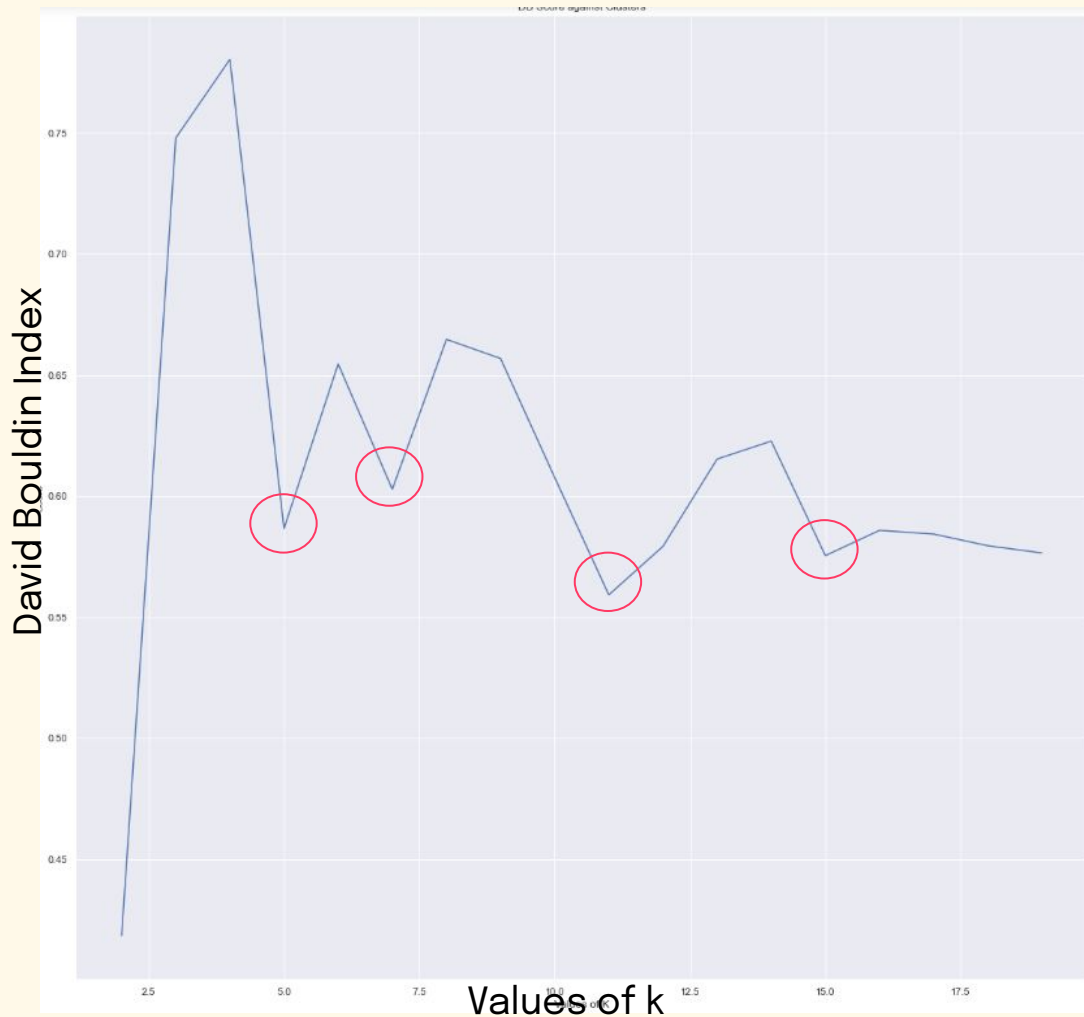
# David Bouldin



## Description

- Measure of similarity between **clusters**
- Compares *intercluster* to *intracluster* distances
- Look for **lowest points**

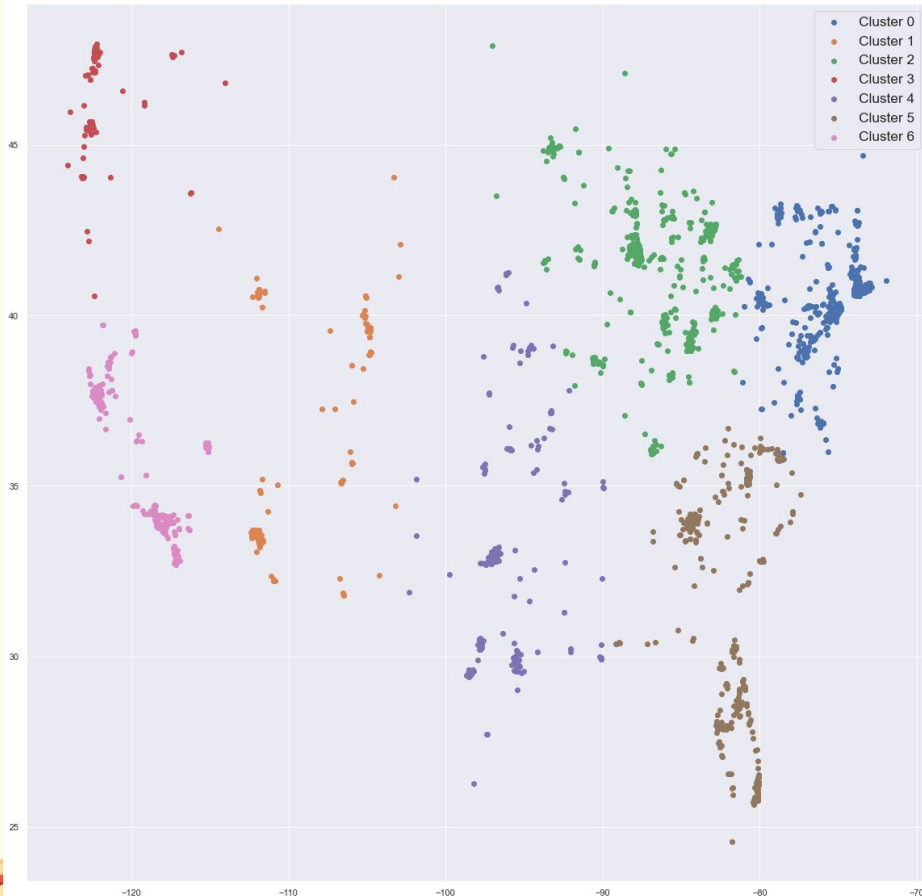
**Ideal k [5,7,11,15]**





# Choosing $k=7$

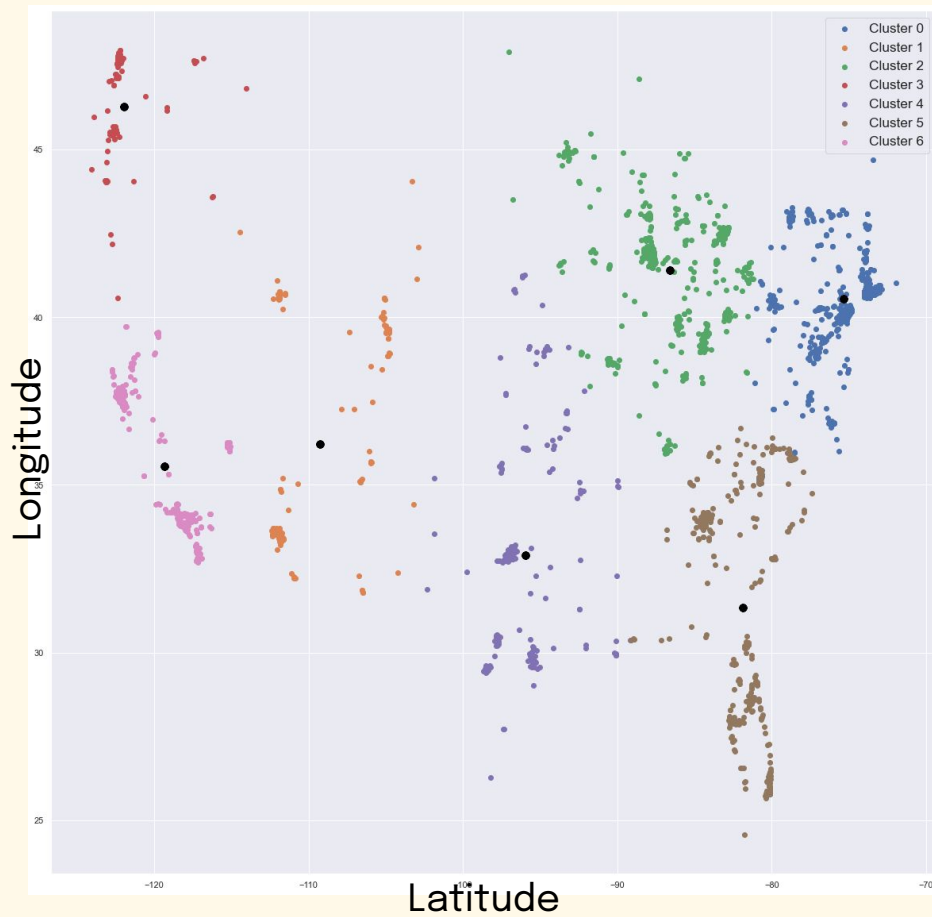
Step #2: Optimisation



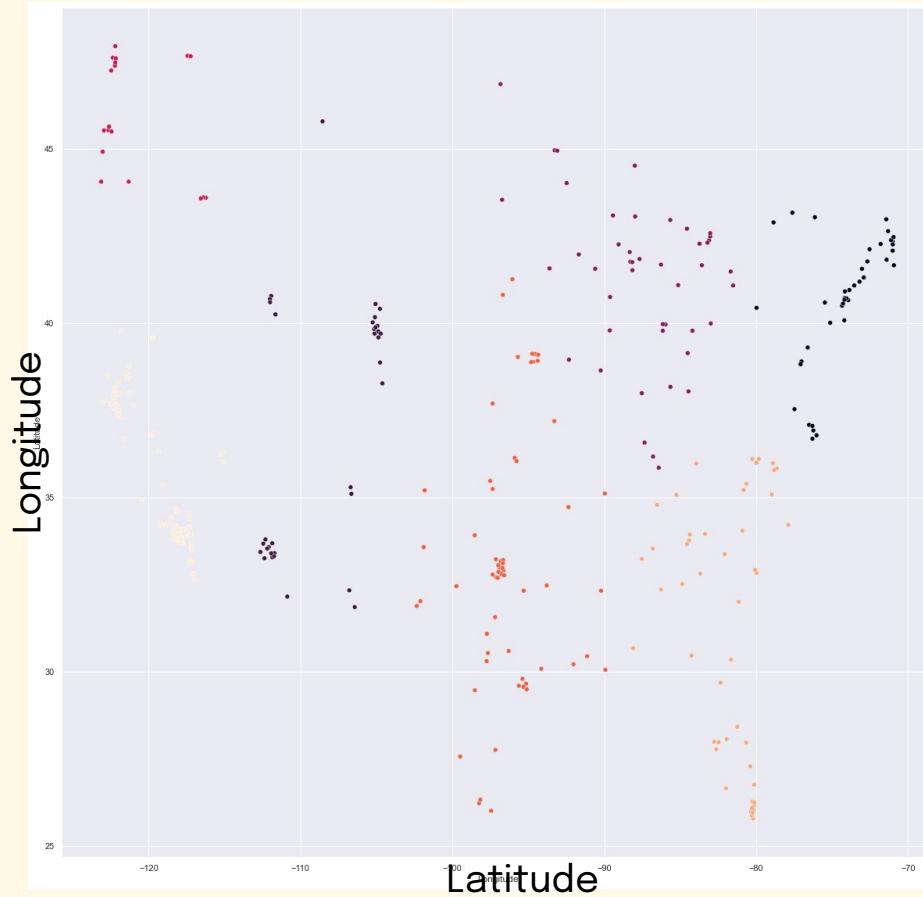
- $k=7$  is optimal as supported by all the other metrics
- $k=5$ ,  $k=11$  and  $k=15$  are good as well, but they are far from the elbow point in the inertia & distortion graph



# Find the centroids of the clusters

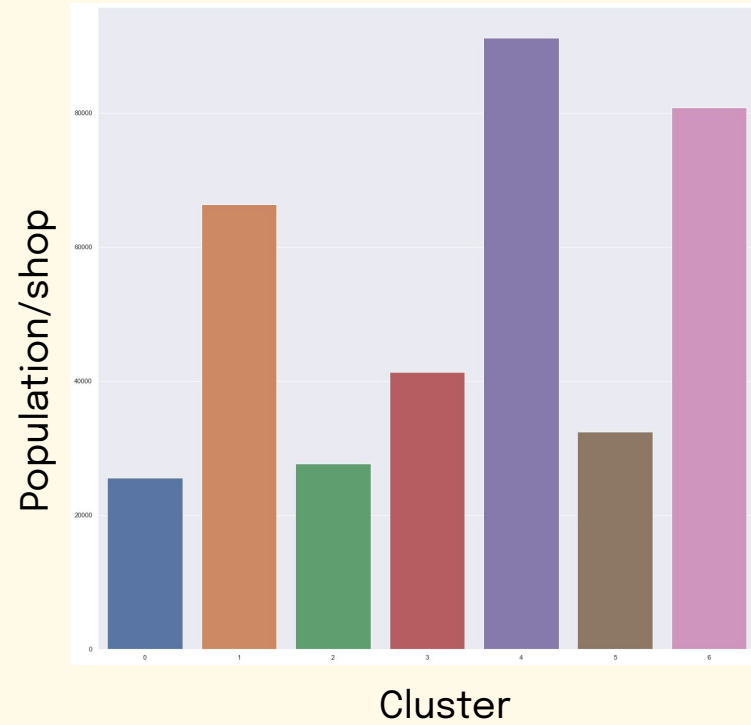
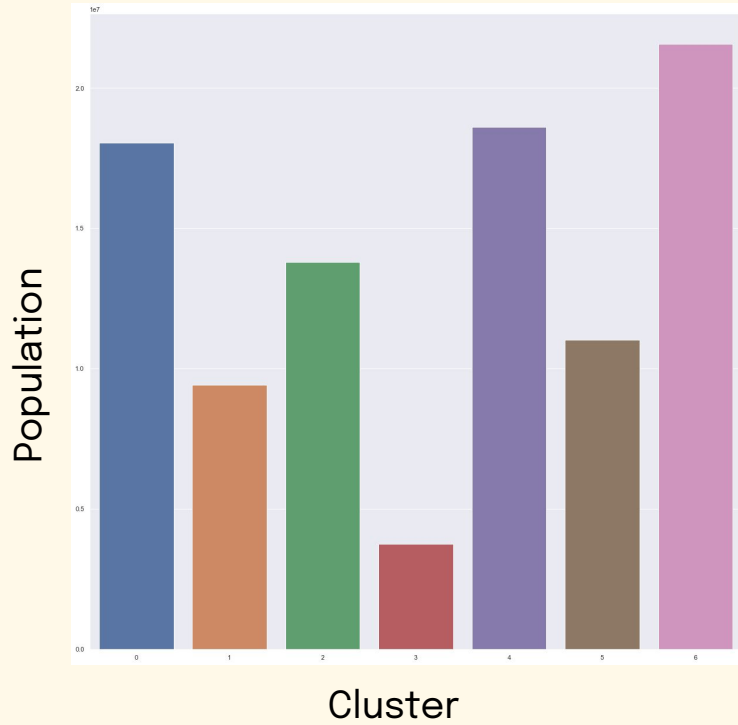


# Assign city to cluster centroids

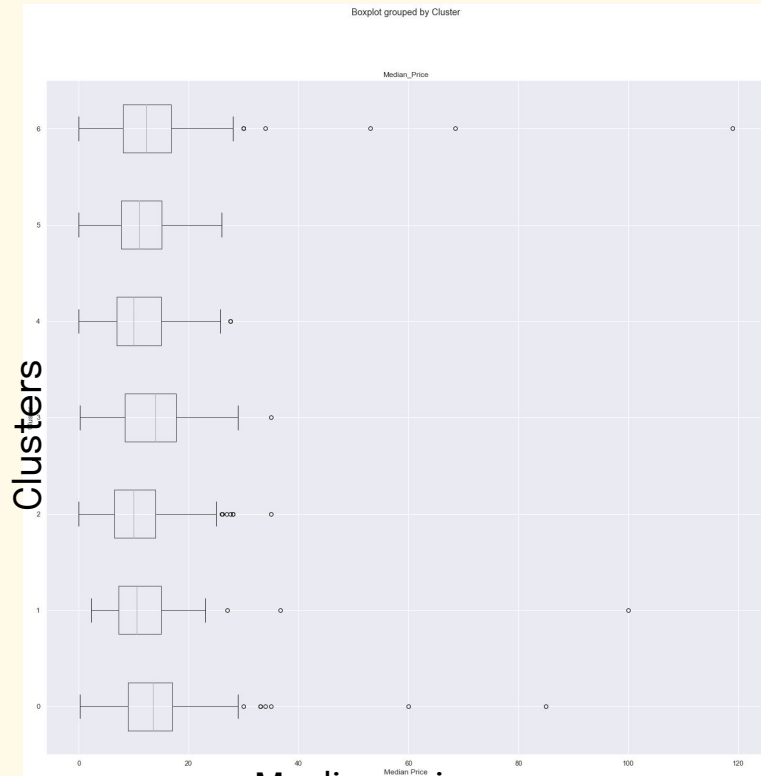




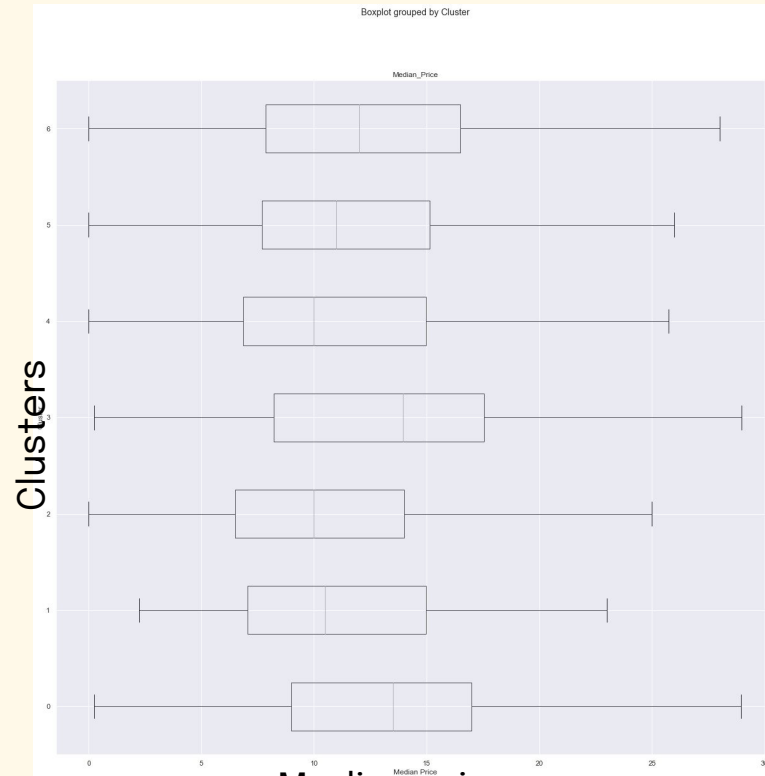
## Step #4: Cluster analysis



# Median prices



Median price  
**Before**

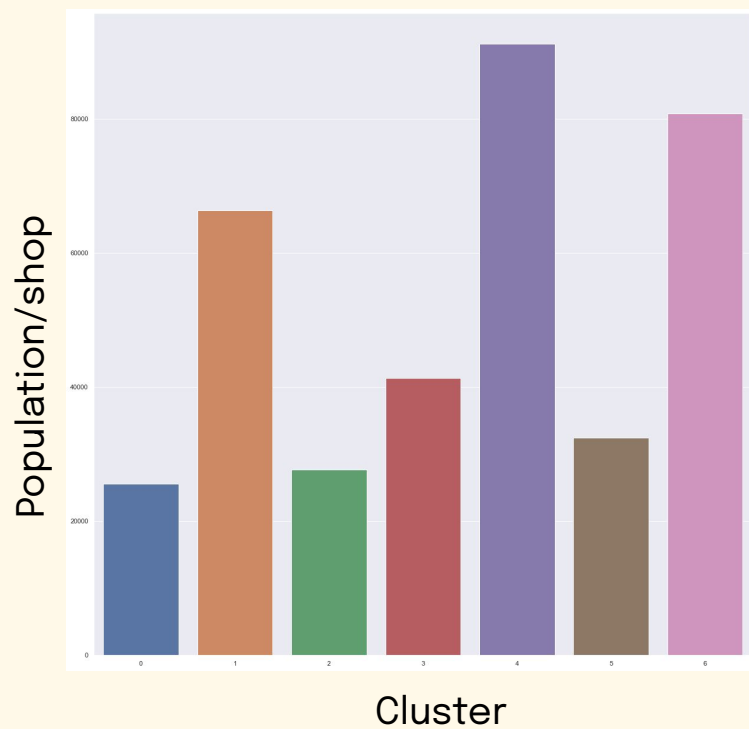
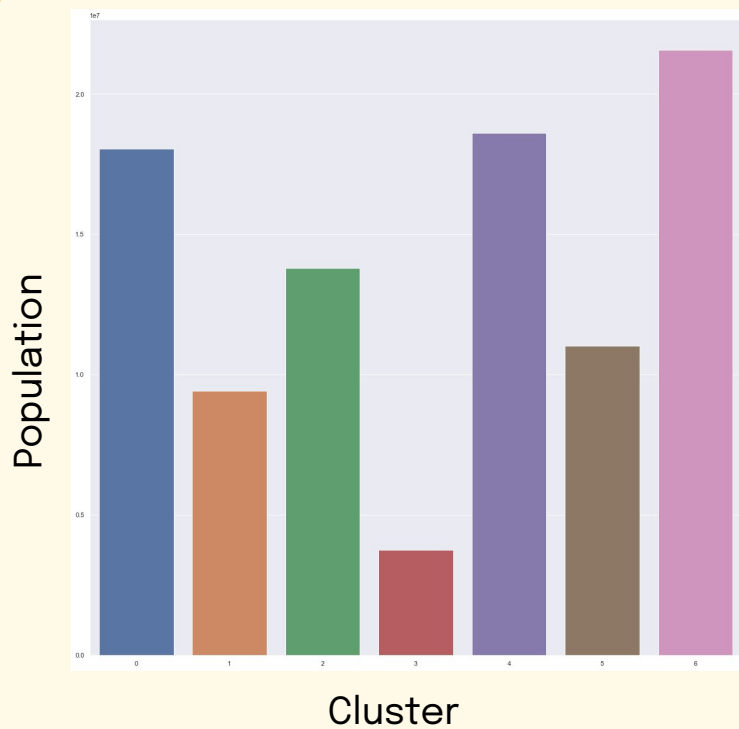


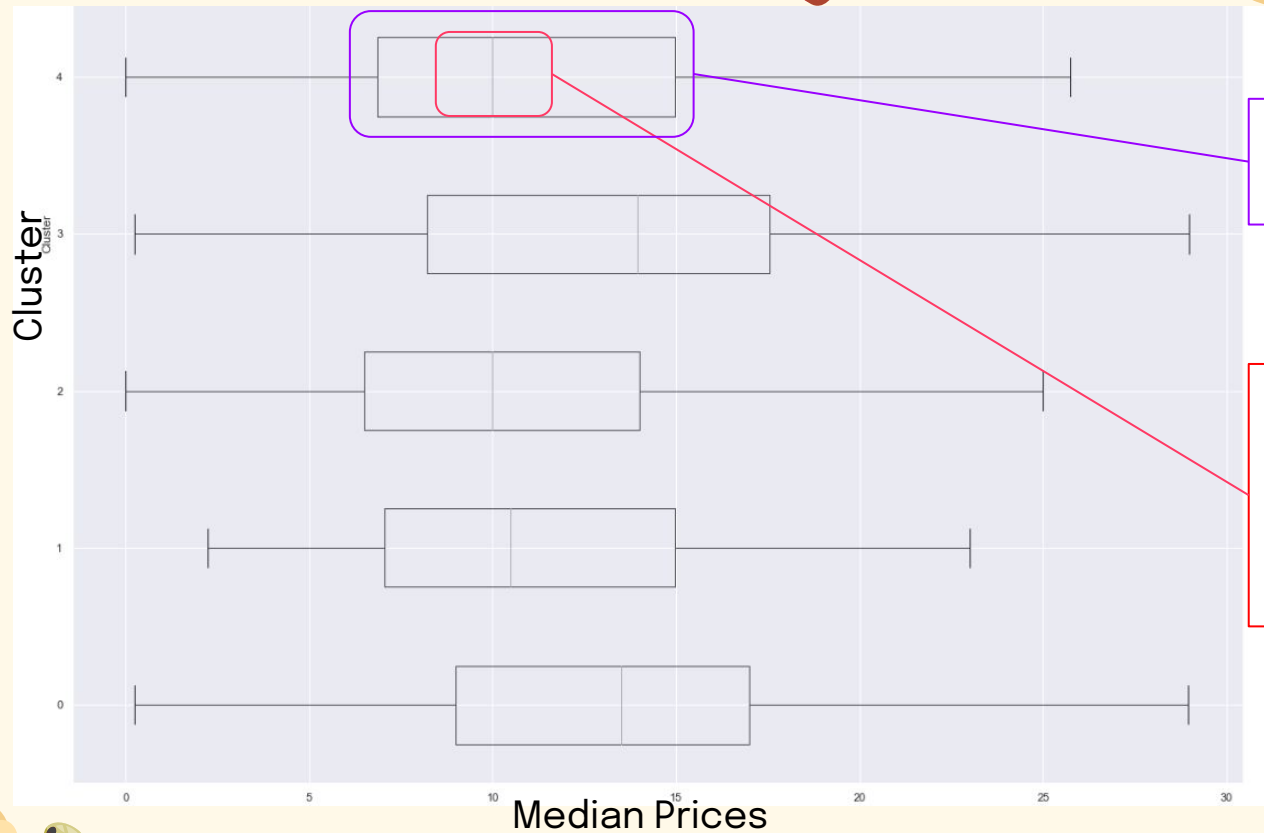
Median price  
**After**



4

# Conclusion





Competitors' menu item prices

Median price of rival shops set ~\$10. To be used as a guideline for menu pricing



# Outcome

- Cluster 4 has the highest pop/shop
- Lowest level of competition with rivals
- It also has 2nd highest population
- \$10 serves as a good median price in this cluster

# What we learned

- Using K-Means clustering
- Using different metrics to optimise the number of clusters

The background is a light cream color. It is decorated with several stylized, flat-design illustrations of pizza-related items. In the top left, there is a partial view of a pizza slice with toppings. In the top center, there is a pink, heart-shaped cookie cutter. In the top right, there is a whole pizza with various toppings like pepperoni, mushrooms, and olives, with a single olive floating nearby. In the bottom left, there is a mushroom and two olives. In the bottom center, there is a partial view of a pizza slice. In the bottom right, there is a mushroom and a pink, heart-shaped cookie cutter. The text "Thank you!" is centered in a bold, dark red font.

# Thank you!