

# IBM Data Science Professional

*The best place to live for a family with children in San Francisco*

Victoria Gogulya  
April 2021



# Introduction

Thousands of people dream about moving to San Francisco. Of course, a life in a big city on the Pacific coast has many benefits. In fact, the list of reasons to live in San Francisco is rather extensive. It is a city of great opportunities. Many people from around the world move here.

What about moving to San Francisco with children? It is very important to find a place or neighborhood which will be comfortable for kids. With schools, parks, ice cream cafes and so on. Kids should love this place and be happy there.

So, the question is how to find a great place to live in San Francisco with children?

Choosing a neighborhood to live in with children can be a complicated task to do, but with the help of Foursquare we can make it a little bit easier.

## Target audience:

- People interested in moving to San Francisco and looking for a perfect neighborhood for their needs
- Beginner data scientists who may use this research as an example

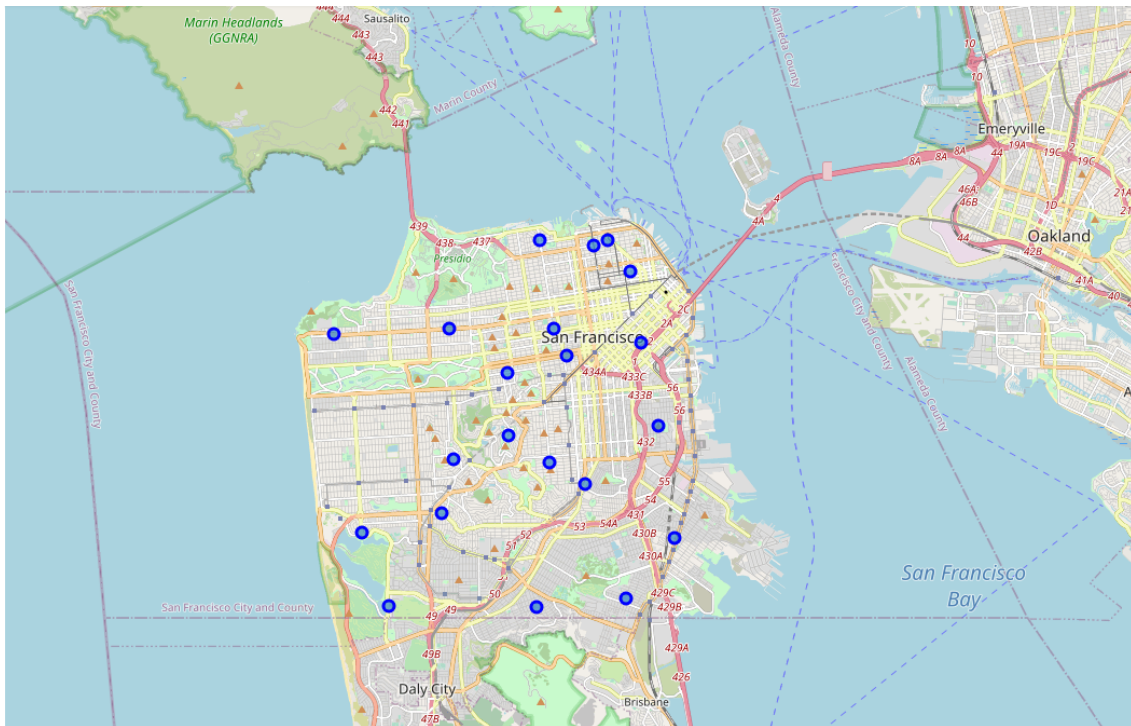
# Data

To find a solution for the question above here we'll need the following data

- List of San Francisco neighborhoods: [SF zip codes and neighborhoods](#)
- Google Geocoding geo coordinates of neighborhoods
- Foursquare venues data for closest schools, parks, kids sores and other attractions for each neighborhood: [foursquare](#)

# Methodology

First of all we need to get neighborhoods of San Francisco and their geo coordinates. List of neighborhoods was taken from the web(link is in the "Data" section of this report). Web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data were used. For extracting geo coordinates, the Geocoder package was used. Show all neighborhoods on map using Folium library.



We get venues using Foursquare API. And additionally we perform search for schools for each neighborhood with radius 500, because general search returns only two types of schools and we need more.

Filter venues. Filtering is based on the knowledge of more appropriate venues for children. Schools, parks, ice cream cafes and so on.

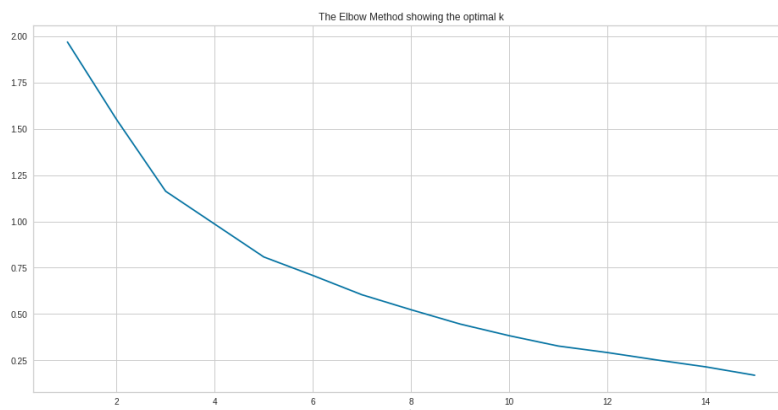
We check if not all neighborhoods have venues for children. We see that all neighborhoods have venues for children.

We use One-hot encoding to transform all categories to the appropriate format.

Then, we use the K-Means clustering algorithm to find out which areas(clusters) are more convenient for living with children.

1. Find the best K value.

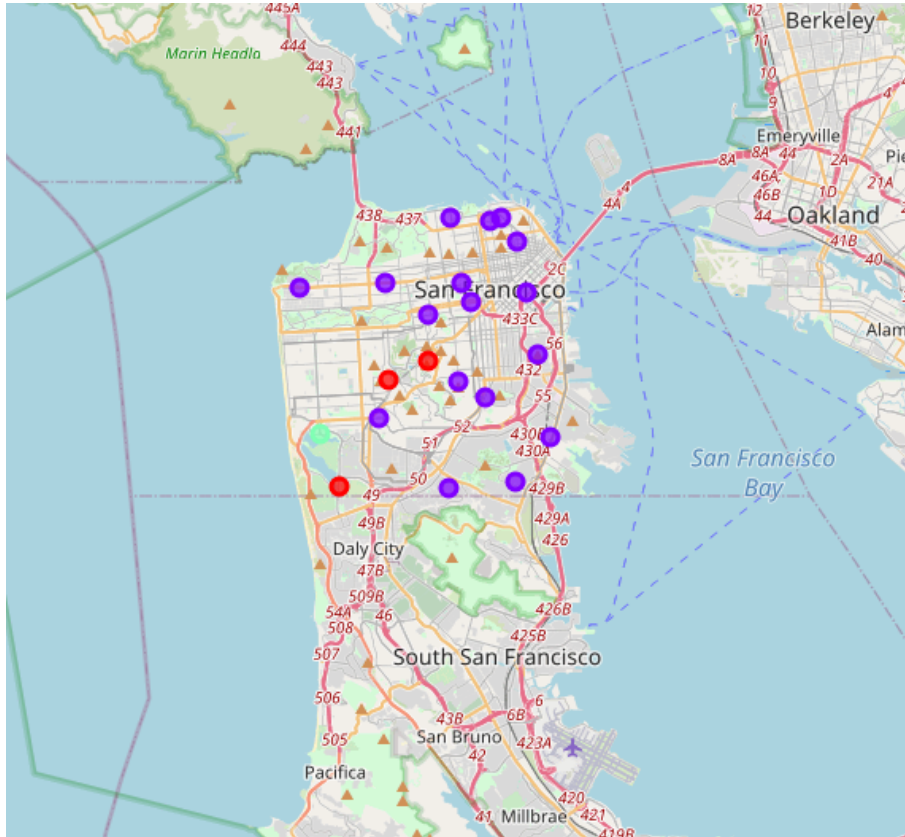
Distortion is low but it seems OK to choose K=3 for our purposes. There is a soft “elbow” here.



2. Run clustering algorithm with  $K=3$

## Results

We get 3 clusters and show them on a map.



The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the specific details of each cluster.

- **Cluster 1: “Welcome with Small children”**. This cluster is better for living with small kids. There is just preschool there, parks, playgrounds. This is a quite small cluster, 3 neighborhoods.
- **Cluster 2: “Welcome with All children”**. This cluster can be used for living with children of any ages. There are many types of schools there, parks, kids stores, ice cream and pizza cafes and so on. A lot of different venues are located here. This is a big cluster, 17 neighborhoods.
- **Cluster 3: “Welcome for a walk with children”**. There are no schools, kids stores and other places which are needed for a life with children. But there is a Zoo, parks and green area. It is nice for a walk!
- This is the smallest cluster, 1 neighborhood.

# Discussion

So, we got 3 clusters in San Francisco.

1. "Welcome with **Small** children"
2. "Welcome with **All** children"
3. "Welcome for a walk with children"

The smallest cluster is the last one - **cluster #3 "Welcome for a walk with children"**. There is only 1 neighborhood there. This is a green area with parks and areas for picnics. There is a Zoo and playgrounds there. But no schools, no kids stores and other venues for children.

Cluster #3 is **green** on the map.

I would recommend this cluster for walking with children.

The biggest cluster is a **cluster #2 "Welcome with All children"**. There are 17 neighborhoods there. There are a lot of nice places to visit with children: ice cream, chocolate and dessert shops. There are many types of schools for different ages. Also there are parks and playgrounds, kids stores and shopping malls.

Cluster #2 is **purple** on the map.

I would recommend this cluster for living with kids of all ages. This cluster has very wide infrastructure. All that you need is here. Recommend!

The middle cluster is a **cluster #1 "Welcome with small children"**. There are 3 neighborhoods there. There are venues for children with small kids: prks, preschool and playgrounds. Other types of schools are absent there.

Cluster #1 is **red** on the map.

I would recommend this cluster for living families with small children.

Each family can choose an area(cluster) which is more comfortable for them to live in San Francisco.

# Conclusion

Here was done a research about finding the best place for living for a family with children. We used data from the internet about neighborhoods of San Francisco city. It was analyzed and filtered. Then, we applied the K-Means clustering algorithm for segmentation and finding the best locations. Two clusters are good for living with children. And one of them is the best because it is suitable for children of any age.

