**Introduction**

Bakeries are a popular type of food service establishment. The smell of freshly baked goods and fantastic coffee is what I call heaven.

My friend Mia loves to bake and just finished her baking school. Opening a bakery presents many unique challenges, soI wants to help her decide on a location where she can open a bakery with a low risk of competition.

**Business problem**

This project aims to analyze and select the best locations in Pune, India, to open a new bakery. This project mainly focuses on the geospatial analysis of Pune City to understand which would be the best place to open a new bakery. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Pune, if a person is looking to open a new bakery, where would you recommend that they open it?

**Data**

To solve the problem, we will need the following data:

• List of neighbourhoods in Pune: It defines this project's scope, which is confined to the city of Pune.

• Latitude and longitude coordinates of those neighbourhoods. This is required to plot the map and also to get the venue data

• Venue data, particularly data related to bakeries. We will use this data to perform clustering on the neighbourhoods.

**Sources of Data and methods to extract the Data**

This Wikipedia page is a list of neighbourhoods in Pune, with 200 neighbourhoods. I have used web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we can get the latitude and longitude coordinates of the neighbourhoods using Python Geocoder package. After that, I have used the Foursquare API to get the venue data for those neighbourhoods.

Foursquare API will provide many categories of the venue data, and we are particularly interested in the Bakery category in order to help us solve the business problem. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).