

Applied Data Science Capstone Project

By Óscar Luis Rojas Fernández

“El Tramito a Granel.”

**Problem description**

Tramito a Granel literarily mean “The little section, by measure”.

In many places of Latin America, the central markets were divided in “tramos”, or small sections where the products where taken in bulk and sold by measure (“a granel”) as opposed to prepackaged.

This is particularly useful to local producers that cannot afford prepackage costs, so central markets usually have products that are hard to find otherwise. Also, this way of selling caters to environmentally concious buyers that can bring their own reusable bags or jars.

However, with the expansion of the cities and the modernization of life, the central markets are becoming out of the practical reach of most of the country.

A picture containing text, indoor, shop

Description automatically generatedA picture containing close

Description automatically generated

El Tramito a Granel opened in 2018 in a district called “Tres Ríos”. It takes the experience of the central market’s “tramo” to the suburbs and it has been a great success. This year they opened a second location in a district called “Escazú” with good results as well.

For this capstone project I will play the role of an advisor to the Tramito’s owners and help them find which districts in Costa Rica are similar to the ones where they have succeeded. This will help them make informed decisions on future expansion plans.

**Data sources**

**Graphical user interface

Description automatically generated**

INEC is the national statistics institute in Costa Rica. Among many other things they provide population statistics by province, canton (sort of borough) and district.

This will be the source of the list of districts in Costa Rica as well as population density (which may be used as part of the exercise).

Graphical user interface, text, application, email

Description automatically generated

There is also a wikipedia page (INEC is one of its sources, by the way), so both options are valid. I will explore both and decide for the easiest.

Graphical user interface, application

Description automatically generatedThe geolocation information will be scrapped out of Wikipedia. Each district in Costa Rica has its own page with basic information, which includes its coordinates.

With the districts location, the foursquare API will be used to gather information of each district in terms of the amount and type of venues that it has at a 2km radius.

**Methodology.**

Venue information, and maybe the population density will be used to make a district “footprint” that will serve as the basis for a district clustering exercise.

A visual examination will be made in order to validate the radius for the venue footprint and if the population density could be used as part of the clustering exercise or maybe as a filter to limit the district candidates prior to the clustering exercise.

The result will be k clusters of districts with similar characteristics among each cluster. The clusters that include the two districts where the actual locations are now will be the ones grouping the district candidates for “El Tramito”s eventual expansion.