

Applied Data Science
Capstone Project
By Óscar Luis Rojas
Fernández

“El Tramito a Granel.”



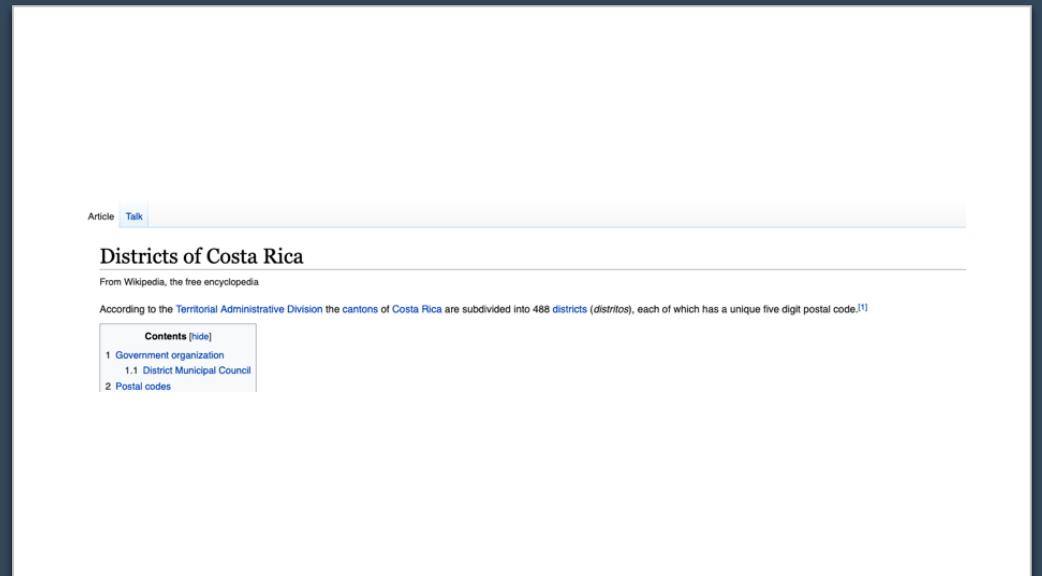
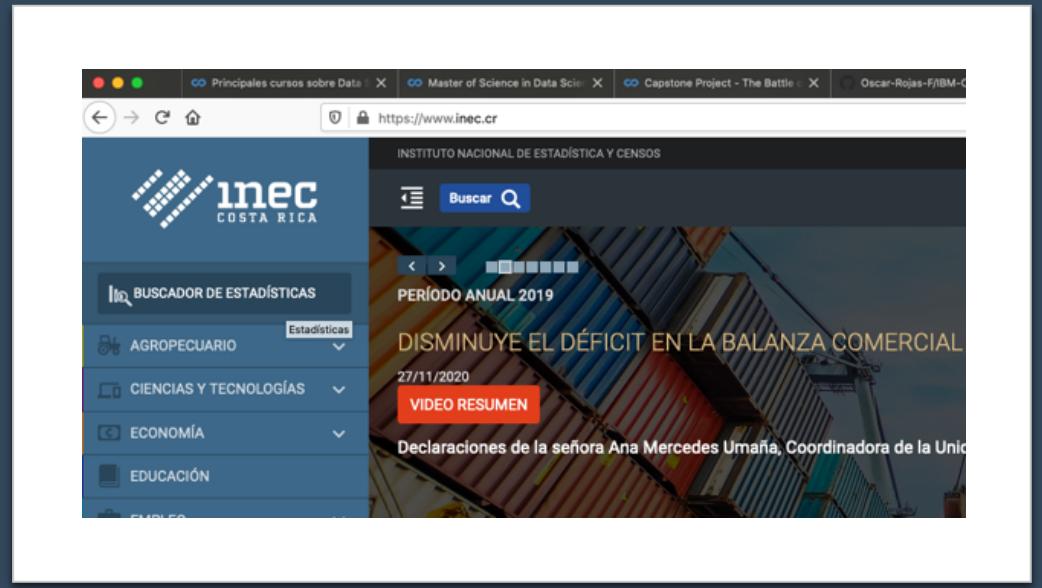
Bringing the experience of the local markets to the suburbs

- El “Tramito a Granel” opened in 2018.
- It takes the experience of the central market’s “tramo” to the suburbs.
- This year they opened in a district called “Escazú”
 - Good results as well.
- Potential new locations could arise from similar districts.
- A clustering exercise will help to list similar districts.



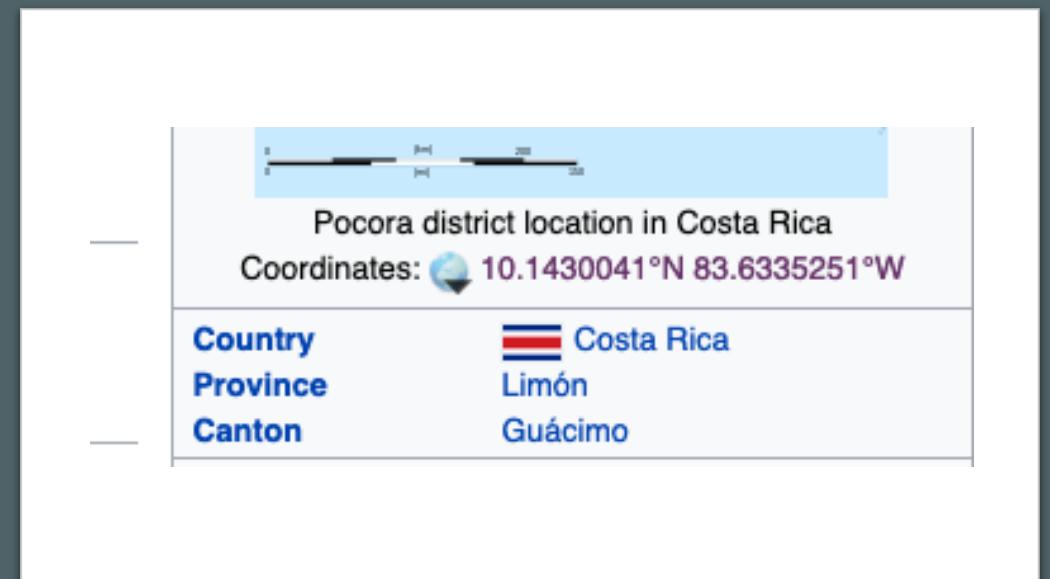
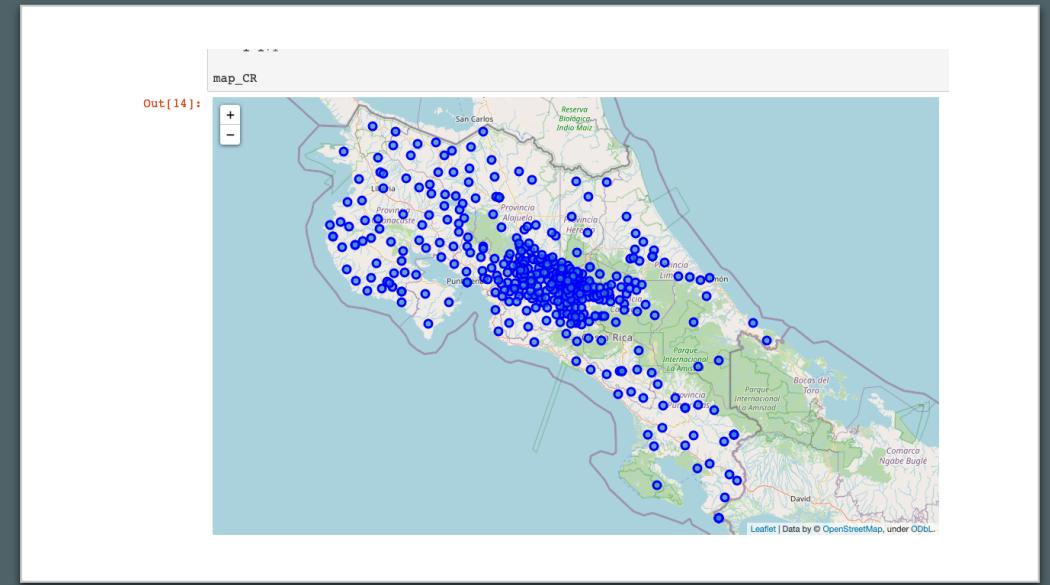
Data sources

- INEC (National census and statistics institute)
- Wikipedia



Data preparation

- Data was taken from INEC, loaded and formatted.
- Calls to wikipedia were used to populate geo location information.
- Visual help from maps was used to validate and depurate.



Methodology

- Foursquare api was used to generate a footprint for each district based on venue information.
- K-means was used to cluster the 400+ districts into 20 groups.

Out[35]:

	Distrito_extendido	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Aguabuena,Coto Brus,Puntarenas	Recreation Center	Breakfast Spot	Farm	Drugstore	Electronics Store	Entertainment Service	Event Service	Electronics Store
1	Aguas Claras,Upala,Alajuela	Resort	Volcano	South American Restaurant	Yoga Studio	Falafel Restaurant	Drugstore	Electronics Store	Food Court
2	Aguas Zarcas,San Carlos,Alajuela	Restaurant	Pizza Place	Bar	Department Store	Italian Restaurant	Gym	Pharmacy	Gas Station

The clusters containing actual locations were identified

```
In [38]: print('This is the cluster were the first location is:')
CR_merged.loc[CR_merged['Distrito_extendido'] == 'San Juan,La Unión,Cartago', CR_merged.columns[[0] + list(range(3, CR_merged.shape[1]))]]
```

This is the cluster were the first location is:

Out[38]:

	Distrito_extendido	Cantón	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
247	San Juan,La Unión,Cartago	La Unión	17.0	Sandwich Place	Supermarket	Ice Cream Shop	Gym	Movie Theater	Restaurant	Shop Mall

```
In [39]: print('This is the cluster were the second location is:')
CR_merged.loc[CR_merged['Distrito_extendido'] == 'San Antonio,Escazú,San José', CR_merged.columns[[0] + list(range(3, CR_merged.shape[1]))]]
```

This is the cluster were the second location is:

Out[39]:

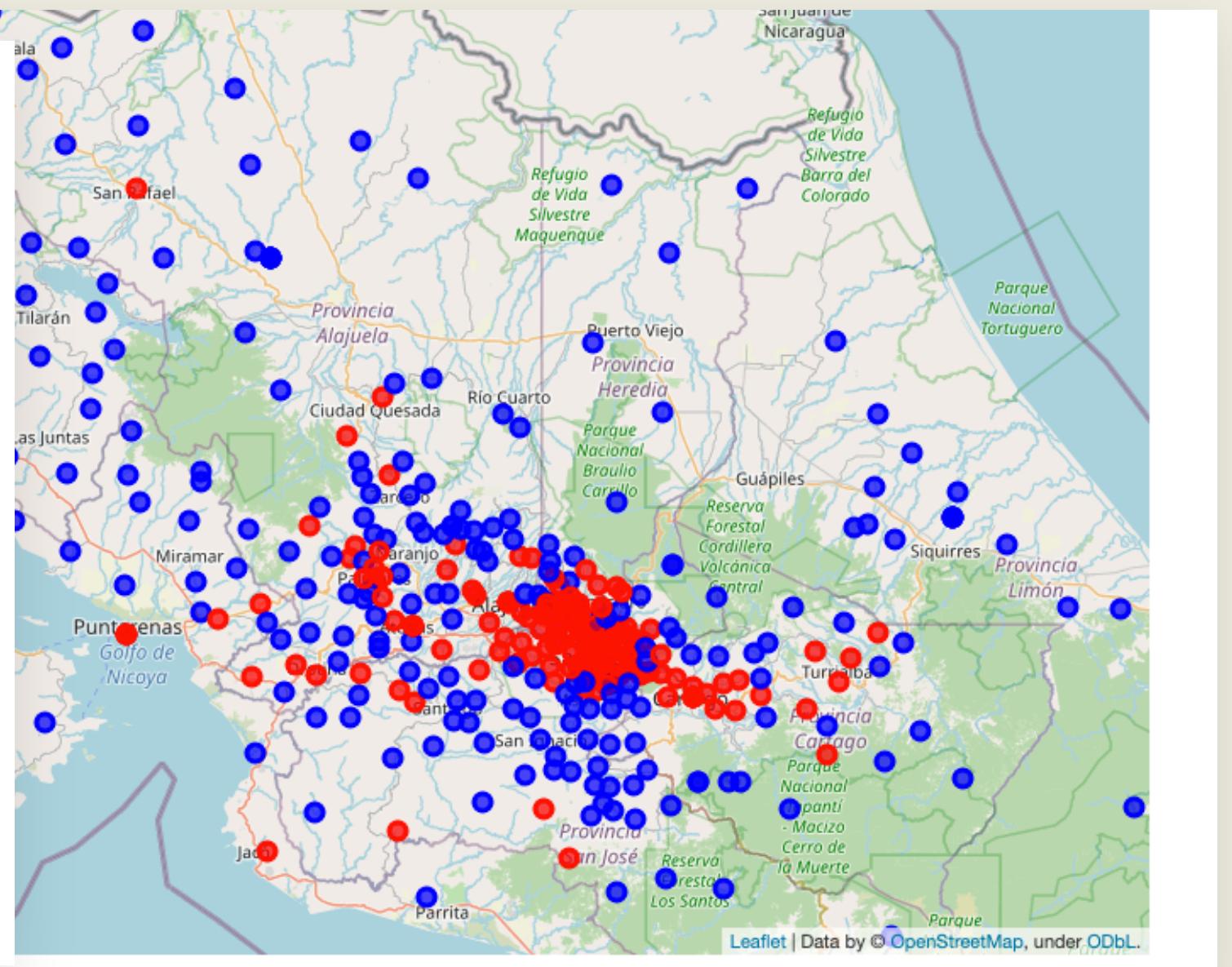
	Distrito_extendido	Cantón	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
12	San Antonio,Escazú,San José	Escazú	1.0	Hotel Bar	Mountain	Forest	Trail	Scenic Lookout	Soccer Stadium	Superma

```
In [40]: # Lets understand a little bit these special clusters.
Best_cluster1=CR_merged.loc[CR_merged['Cluster Labels'] == 17, CR_merged.columns[[0] + list(range(3, CR_merged.shape[1]))]]
Best_cluster1.shape
```

Out[40]: (144, 13)

```
In [41]: Best_cluster2=CR_merged.loc[CR_merged['Cluster Labels'] == 1, CR_merged.columns[[0] + list(range(3, CR_merged.shape[1]))]]
Best_cluster2.shape
```

The results
were mapped
for visual
analysis.



The result was also expressed in a list of “cantones” (burghs) ordered by occurrence of expansion candidate districts.

Cantón	
San José	11
Alajuela	10
Cartago	6
La Unión	6
Desamparados	6
Goicoechea	6
Tibás	5
Santo Domingo	5
Palmares	5
San Rafael	5



Conclusion and future ideas.

- It is possible to use a unsupervised clustering algorithm to create a list of location candidates for an eventual expansion.
- Data and analysis show that “El Tramito” has a favorable outlook for growth and expansion in Costa Rica.
 - We were able to create a list of “cantones”, based on their number of districts with favorable conditions.
- Current data sources are valid, with lots of cleaning and manual correction.
- This exercise could be greatly benefited from the 2021 Costa Rica census information.