

ANALYSIS OF VENUES CATEGORIES IN THE NEIGHBORHOODS OF MADRID

DATA

Title	Analysis of venues categories in the neighborhoods of Madrid		
Reference	N/A		
Revision	0		
Location	Madrid (Spain)		
Date	April 3, 2020		
Petitioner	N/A		
Author	Name	Daniel Sánchez Valenciano	
	ID	N/A	
	Qualification	BS in Telecommunications Engineering	
	Official Member's N.	N/A	

SUMMARY

To be completed in week 2.

CONTENTS

1. Introduction/Business problem.....	1
1.1. Business problem.....	1
1.2. Stakeholders.....	1
1.3. Data.....	1

1. INTRODUCTION/BUSINESS PROBLEM

1.1. Business problem

How are venues distributed among the neighborhoods in Madrid? Where are the neighborhoods in which a certain venue category is specially usual? Which is the most common venue category in each neighborhood? Is there any "outlier" neighborhood in Madrid? Could this information be leveraged in any way? **All these questions define the business problem** around which this data science project revolves.

1.2. Stakeholders

Madrid is a city brimming with culture, leisure, catering and hostelry, and so it attracts many entrepreneurs and investors from these fields. **These are the stakeholders in this business problem.** To boost the chances of success of their enterprises and investments, many of them rely on market researches which try to answer questions such as:

- If a business of a specific type is to be put into operation, where are the neighborhoods in Madrid which should be targeted? How are they distributed in the city? In the suburbs, in the down town, in the old city?
- If a the stakeholder plans to put a business of any type into operation in a specific neighborhood of Madrid, what business types should they consider? Is this neighborhood saturated with a particular type of business? Does this neighborhood lack any type of business? What is the market niche?
- Is there any "special" neighborhood in Madrid regarding its venues? Does it offer any special market opportunity to the entrepreneurs and investors?
- Finally, considering that venues open, close or simply change overnight, would it be possible to have this information dynamically updated?

This data science project aims to answer all of these questions for the stakeholders.

1.3. Data

The following data is used:

- Madrid boroughs and neighborhoods data. They are freely available on Madrid City Council Open Data website ¹. They come in the form of geospatial vector data files of Madrid boroughs and neighborhoods. Once downloaded, these files are loaded in QGIS, a free and open-source cross-platform desktop geographic information system, in order to be processed. See Illustration 1.

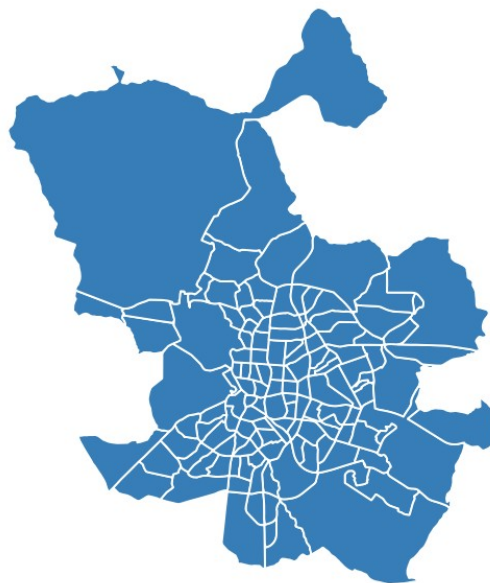


Illustration 1: Madrid boroughs and neighborhoods geospatial vector data files loaded in QGIS

The files are processed leveraging QGIS functionalities so that the corresponding attributes table contains the following information for each neighborhood of Madrid.

- Borough name.
- Neighborhood name.
- Area.
- Perimeter.
- X coordinate (ETRS89/UTM zone 30) of its centroid.
- Y coordinate (ETRS89/UTM zone 30) of its centroid.

See Illustration 2.

¹<https://datos.madrid.es/egob/catalogo/200078-10-districtos-barrios.zip>

	NOMDIS	NOMBRE	area	perimeter	xcoord	ycoord
1	Arganzuela	Legazpi	1414470,6...	5141,6408...	441682,93...	4471151,4...
2	Arganzuela	Chopera	567787,01...	3203,4086...	440672,25...	4471808,8...
3	Arganzuela	Delicias	1054678,7...	4818,0027...	441453,46...	4472027,9...
4	Arganzuela	Acacias	1073437,8...	3950,3269...	439982,85...	4472513,4...
5	Arganzuela	Palos de M...	648442,55...	3647,7465...	441057,64...	4472774,2...
6	Arganzuela	Atocha	735680,59...	4957,6874...	442131,23...	4472353,0...
7	Arganzuela	Imperial	967678,64...	4557,9379...	439045,61...	4473089,6...
8	Barajas	Alameda d...	1970334,8...	6044,9403...	449852,67...	4478600,0...
9	Barajas	Corralejos	4682537,7...	9726,9510...	448524,15...	4479459,2...
10	Barajas	Casco Hist...	549393,99...	3388,6462...	450931,36...	4480530,1...
11	Barajas	Timón	5094465,9...	11840,044...	448908,39...	4480939,2...
12	Barajas	Aeropuerto	29626079,...	28744,094...	452241,98...	4481026,8...

Illustration 2: Madrid boroughs and neighborhoods geospatial vector data files final attributes table

Once the geospatial vector data files are processed, the final attributes table is exported as a comma separated values file (“madrid_neighborhoods.csv”). Then, it is loaded into the Watson Studio project as an asset. See Illustration 3. Finally, this file is read from a Jupyter Notebook using methods from `project_lib` and `pandas` libraries.

IBM Watson Studio				
My projects / Applied Data Science Capstone				
Data assets				
0 assets selected.				
<input type="checkbox"/>	Name	Type	Created by	Last modified
<input type="checkbox"/>	csv Madrid_neighborhoods.csv	Data Asset	Daniel Sánchez Valenciano	Mar 15, 2020, 11:18 PM

Illustration 3. “madrid_neighborhoods.csv” file loaded into the Watson Studio project as an asset

- Foursquare location data.

Foursquare API is leveraged to explore venues in each neighborhood in Madrid. See Illustration 4.

	Name	Category	Latitude	Longitude
0	La Gelateria di Angelo	Ice Cream Shop	40.397951	-3.707739
1	Parque de la Arganzuela	Park	40.398330	-3.708686
2	Restaurante Peruano Mis Tradiciones	Peruvian Restaurant	40.399816	-3.711022
3	sushi raku	Sushi Restaurant	40.404623	-3.708216
4	Le Crust Pizza Bar	Pizza Place	40.400922	-3.709890

Illustration 4. Example of venues data from “Acacias” neighborhood retrieved from Foursquare