**Curso Ciencia de Datos IBM - Coursera**

**Tarea final – Proyecto Coursera Capstone**

**Título : La Batalla de los barrios (vecindarios) en Montevideo.**

**Tarea Semana 2 ( semana 5 del curso)**

[1- Introducción y descripción del caso. 2](#_Toc89082033)

[2- Datos utilizados. 2](#_Toc89082034)

[2.1- Datos de población 3](#_Toc89082035)

[2.2 - Datos Socioeconómicos 4](#_Toc89082036)

[2.3 - Datos Geográficos y coordenada. 4](#_Toc89082037)

[3- Forma de abordaje y metodología de análisis 5](#_Toc89082038)

[3.1- Primera parte 5](#_Toc89082039)

[3.2- Segunda Parte : 6](#_Toc89082040)

[4- Resultados obtenidos 6](#_Toc89082041)

[4.1 – Resultados preseleccion. 6](#_Toc89082042)

[Análisis Demográfico 7](#_Toc89082043)

[Análisis Económico 8](#_Toc89082044)

[Análisis geográfico: 9](#_Toc89082045)

[4.2 – Resultados con sitios de interés con Foursquare y clustering con K Means. 11](#_Toc89082046)

[Resumen de resultados obtenidos 15](#_Toc89082047)

[5- Debate de Resultados 16](#_Toc89082048)

[6- Conclusión final. 17](#_Toc89082049)

[Fuentes . 19](#_Toc89082050)

# Introducción y descripción del caso.

El caso a analizar refiere a una **empresa multinacional dedicada al rubro gimnasios, centros deportivos y comercialización de artículos deportivos, accesorios y vinculada a una cadena de comida saludable,** que desea conocer la mejor ubicación para instalar una nueva sucursal en la **ciudad de Montevideo, Uruguay**.

Dicha multinacional ya posee ubicaciones en otras principales ciudades de Lationamérica.

La ubicación deberá estar en alguna de las zonas o barrios de Montevideo **con mayor población y mejor nivel socioeconómico, así como estar cercana a espacios abiertos, como zonas costeras, playas y parques.**

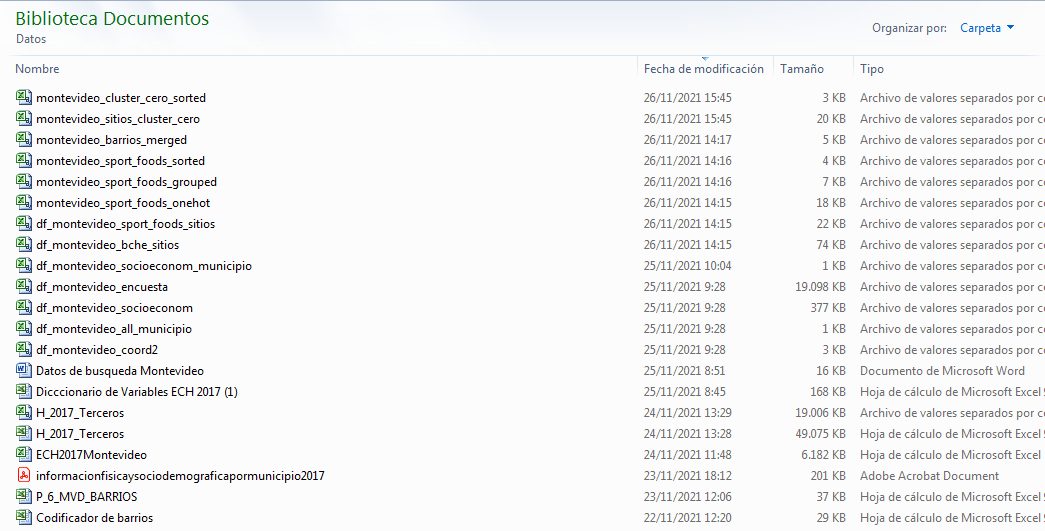
Montevideo cuenta con una franja costera conocida como la Rambla, de más de 15 km de extensión, con playas de arena fina sobre el Rio de la Plata.

El objetivo del caso es determinar cuáles zonas, municipios o barrios son los más aptos para la instalación de la sucursal, y que estén dentro cercanos a la zona de playas y parques. Montevideo cuenta con 62 barrios agrupados en 8 municipios.

Fuente :<https://es.wikipedia.org/wiki/Anexo:Barrios_de_Montevideo>, <https://es.wikipedia.org/wiki/Anexo:Municipios_de_Montevideo>

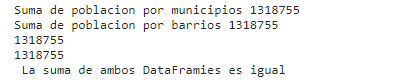
# Datos utilizados.

Detallamos los archivos de datos utilizados que se encuentran en el repositorio de GitHub compartido.

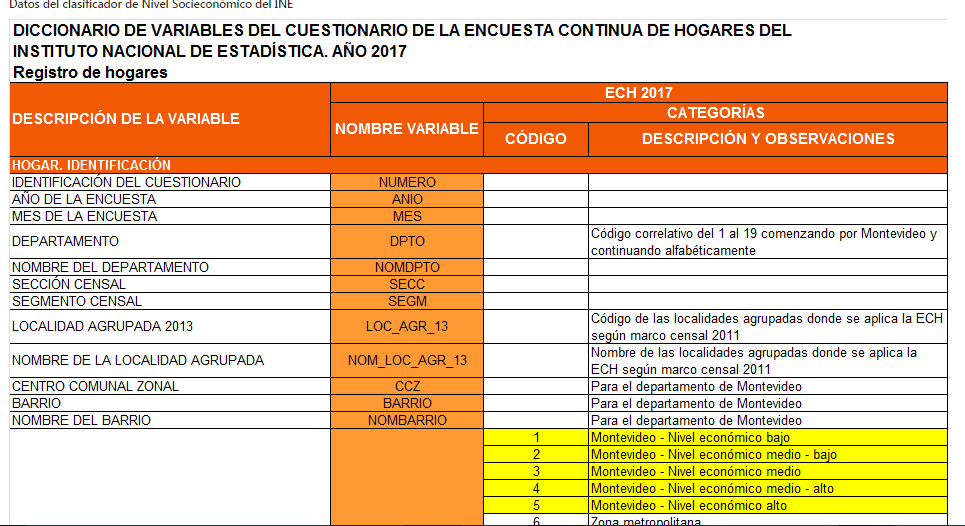
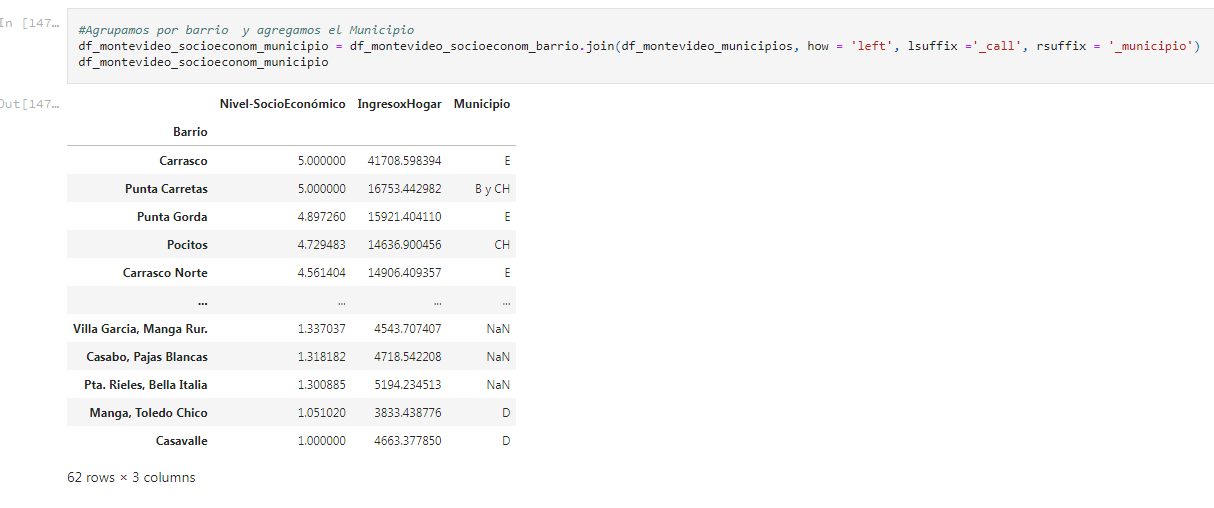


## 2.1- Datos de población

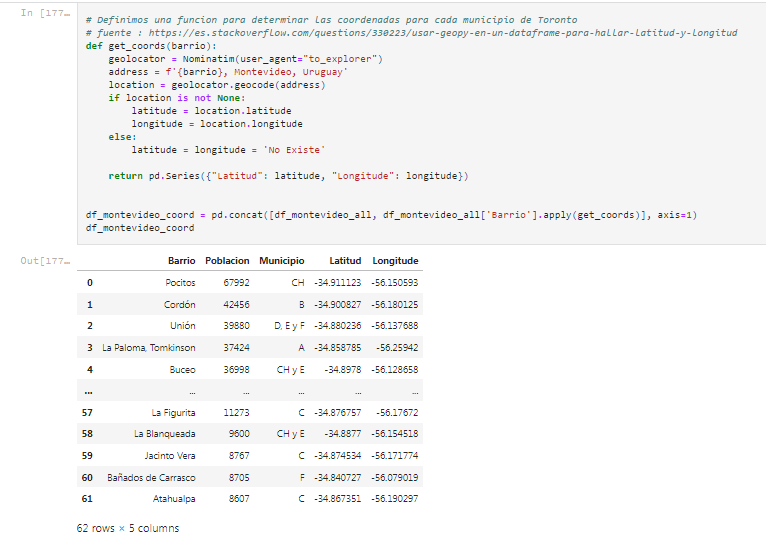




## 2.2 - Datos Socioeconómicos



## 2.3 - Datos Geográficos y coordenada.





# Forma de abordaje y metodología de análisis

Como mencionamos en la introducción o tarea semana 4, vamos a dividir la parte de análisis en dos partes :

3.1- Primera parte vamos a simplificar el marco de datos para el análisis, **seleccionando aquellos municipios que cuentan con las características del público objetivo**.

Por lo cual buscaremos preseleccionar aquellos municipios, y dentro de ellos aquellos barrios o vecindarios , **con mayor nivel socioeconómico**., y por otro lado, desde el punto de vista demográfico preseleccionaremos aquellos municipios, y dentro de ellos aquellos barrios**, con mayor población.** Además, dado que se pretende contar con espacios al aire libre cercanos, asi como cercanía a la costa, preseleccionaremos **aquellos municipios con mayor cantidad de espacios abiertos como plazas y parques** y que visualmente en el mapa se encuentren más cerca de la costa.

Para la preselección socioeconómica tomaremos en cuenta la **Encuesta de Hogares del año 2017, del Instituto Nacional de Estadísticas del Uruguay,** que cuenta con indicadores para Montevideo, y deglosados por barrios, del nivel socioeconómico de los hogares encuestados.

Para la preselección demográfica tomaremos en cuenta los datos del **último Censo Oficial y completo de Población, del año 2011**, realizado por el mismo Instituto antes mencionado, que trae la población de Montevideo desglosada por barrios.

Para la preselección demográfica construiremos una tabla de **coordenadas geoespaciales usando la librerìa Geopy y y el módulo Nominatim de Python** , dado que es el que hemos probado antes que proporciona resultados confiables.

Para asociar los 62 barrios a los 8 municipios usaremos la información proporcionada por el **Gobierno departamental de Montevideo y que está publicada tanto en su sitio web como en Wikipedia,** de esta última importaremos los datos usando métodos panda de Python. Cabe decir que solamente tomaremos los datos del área urbana, no así del área rural del departamento mencionado.

Para esta primera parte buscaremos la información en los sitios web de los organismos mencionados, y ya sea en formato Excel, o en formato .csv, importaremos dichos datos y los depuraremos como se muestra en los **notebooks de jupyter Python, usando métodos pandas y visualización de gráficos con matplotlib.pyplot, así librería geopy-Nomnatim para la búsqueda de coordenadas y librería folium para la visualización de mapas**.

3.2- Segunda Parte : Con los municipios y barrios preseleccionados, en esta **segunda parte** usaremos **la API de Foursquare** sugerida por el curso, para determinar que sitios de interés, tanto deportivo como gastronómico, posee cada barrio. No consideraremos otros sitios de interés comercial, cultural o político. Nos basaremos en la misma metodología vista en el laboratorio de vecindarios de Nueva York-Manhathan, proporcionada por el curso, para el armado y depuración de los dataframes de Python-panda, ( uso de Get.dummies) , de forma de obtener un dataframe por barrio con el top 10 de sitios más comunes detectados.

Luego aplicaremos **clustering usando K-means,** como se vió en el curso. **De los clústeres obtenidos, seleccionaremos aquellos con mayor población y mayor cantidad de sitios vinculados a deportes y espacios libres, así como vinculados al rubro gastronómico**. **Entendemos que, dado que ya hay otras propuestas o sitios vinculados al rubro de negocios que pretendemos desplegar, la población objetivo es más propensa a aceptar una nueva propuesta integral de deportes, recreación y gastronomía como se proyecta.**

Si fuera necesario hacer una subdivisión de alguno de los clústeres seleccionados, repitiríamos los pasos indicados previamente, construyendo los dataframes y usando K-means.

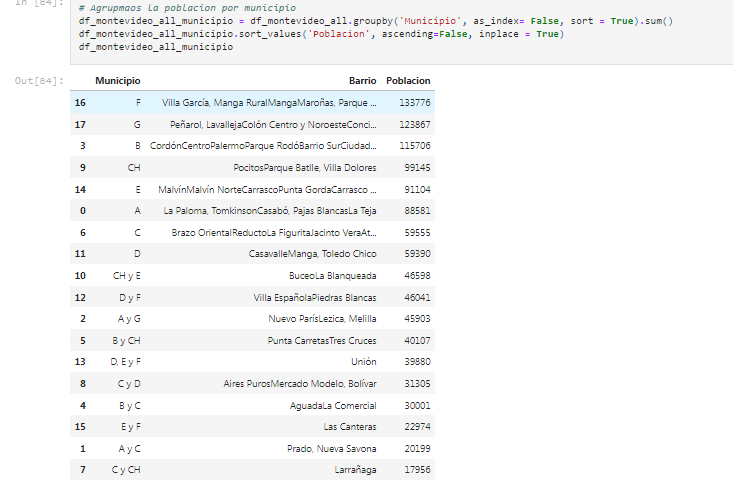
Finalmente, seleccionaríamos aquel barrio que cuente con mayor cantidad de sitios similares al que queremos proyectar, y con mayor población y cercanía a la costa.

# 4- Resultados obtenidos

## 4.1 – Resultados preselección.

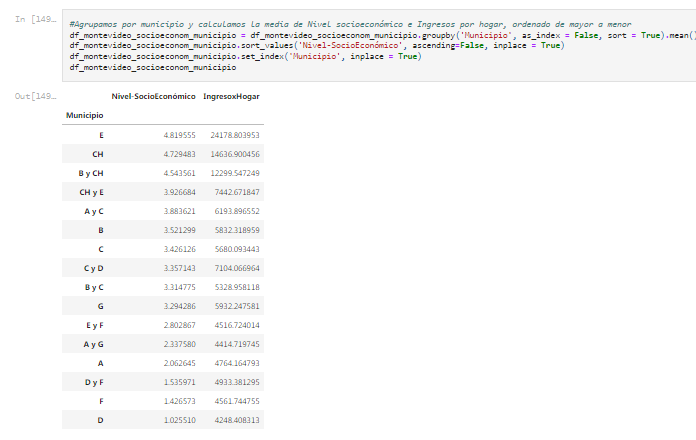
En esta primera parte, luego de descargar, agrupar y filtrar los datos por municipio y barrio, en un análisis demográfico, económico y **geográfico, los municipios seleccionados son B, CH y E,** **dado que presentan mayores niveles de población, mayores indicadores socioeconómicos y mayor cercanía a la franja costera de playas**. Esto se puede apreciar en los gráficos resúmenes generados.

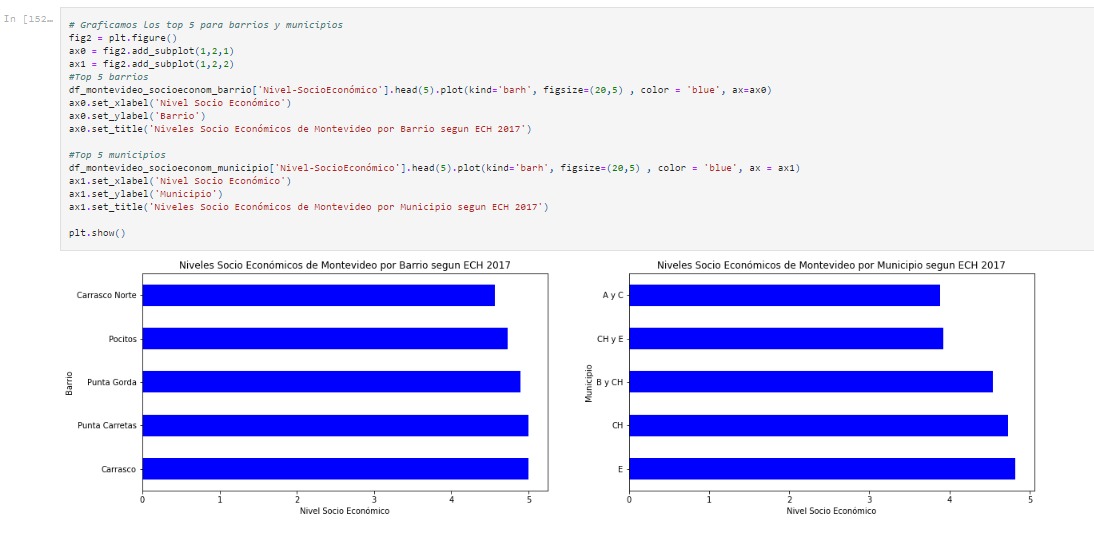
### Análisis Demográfico



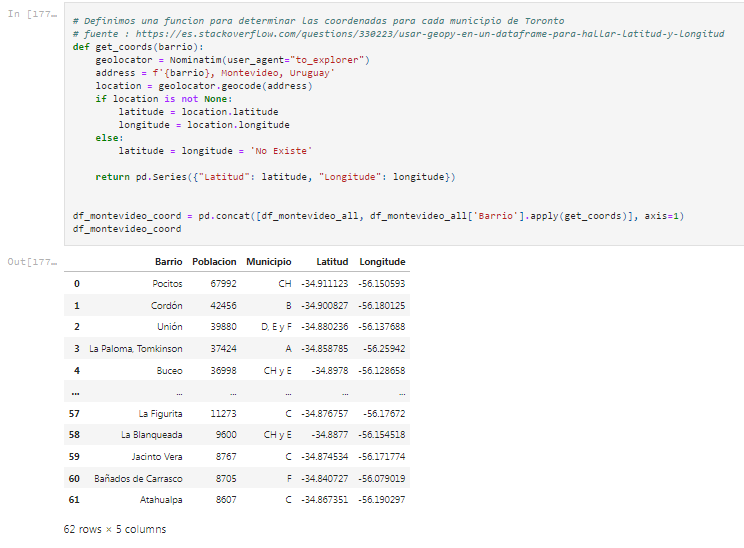


### Análisis Económico





### Análisis geográfico:

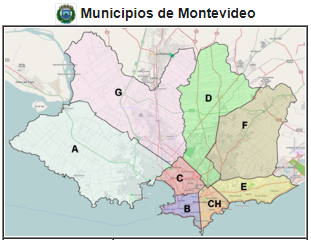






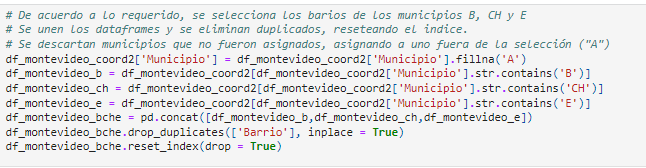
Si bien del análisis demográfico se muestra que los municipios más poblados son el F y G, los mismos se encuentran lejanos de la franja costera, y no presentan indicadores socioeconómicos de la población objetivo, dado que no figuran en el top 5 de municipios con nivel SE superior o igual a 3.

Por otro lado, si bien los municipios A y c presentan indicadores SE superiores a 3, no se encuentran dentro de la zona costera, por lo cual son descartados también en esta preseleccion. Esto lo podemos ver más claro en el siguiente mapa extraìdo de la página de Municipios de Montevideo de Wikipedia.



## 4.2 – Resultados con sitios de interés con Foursquare y clustering con K Means.

Los barrios incluidos en los municipios preseleccionados son :





Luego, aplicando **la API de Foursquare**, seleccionamos aquellos sitios dentro de esos barrios que se relación con el negocio proyectado ( deportes, sitios al aire libre, gastronomía).





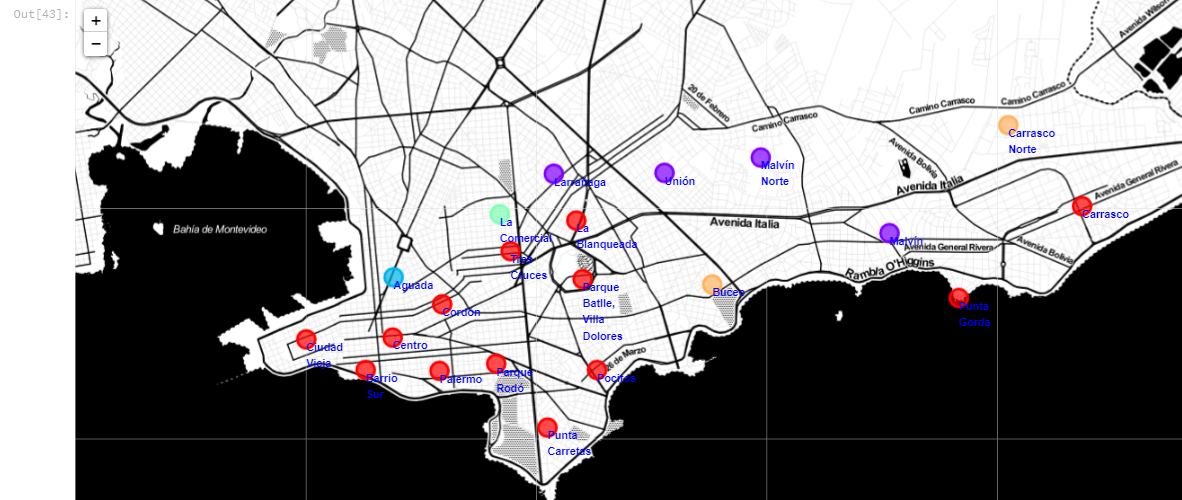
Usando la metodología vista en el laboratorio de los barrios de Manhathan, preparamos el **dataframe y aplicamos K means con 5 clusteres**:





Podemos verlos en el mapa diferenciados por colores:

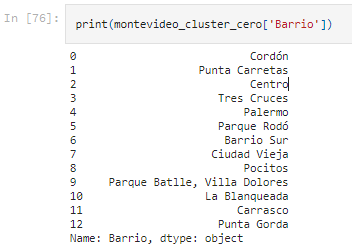




**El más poblado de los 5 clústeres es el clúster cero, en color rojo en el mapa.**



Compuesto por los siguientes barrios:

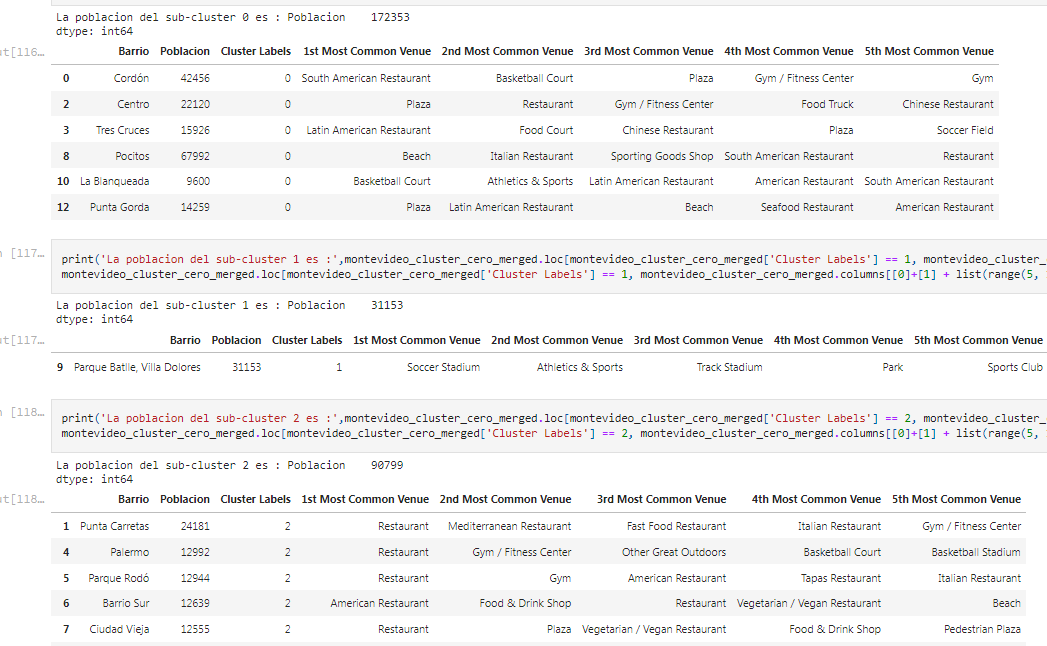


Dado que este clúster aún presenta muchas variedades geográficas y de sitios para hacer una selección definitiva, **volvemos a repetir el proceso de clustering, pero solo para los barrios del clúster cero, generando un nuevo dataframe, y aplicando K means con K igual a 3.**

Podemos visualizarlo en el siguiente mapa, siendo **rojo el sub-clúster cero, azul el sub-clúster 1 y verde el sub-clúster 2.**



De los tres sub-clústeres **el más poblado es el sub-clúster cero, (rojo).**



**Además, podemos hacer un conteo de sitios de interés detectados dentro de cada cluster.**

### Resumen de resultados obtenidos

El **sub-clúster cero** del clúster cero original, es el que tiene mayor población y en el top 5 tiene:

* 2 playas
* 3 gimnasios
* 3 estadios deportivos
* 2 sitios de artículos deportivos
* 4 plazas
* 16 sitios de comida y restaurantes

El **sub-clúster dos** del clúster cero original, es el que tiene sigue en cantidad de población y en el top 5 tiene:

* 1 playas
* 3 gimnasios
* 3 estadios deportivos
* 2 parque o sitios al aire libre
* 0 sitios de artículos deportivos
* 2 plazas
* 19 sitios de comida y restaurantes

El **sub-clúster uno** del clúster cero original, es el de menor población y en el top 5 tiene:

* 0 playas
* 0 gimnasios
* 2 estadios deportivos
* 1 parque o sitios al aire libre
* 1 sitios de artículos deportivos
* 0 plazas o sitios al aire libre
* 0 sitios de comida y restaurantes
* 1 club deportivo

# Debate de Resultados

El análisis demográfico y económico permiten determinan que **tres de los 5 municipios más densamente poblados, son los que a su vez cuentan con mejor nivel socio económico y más cercanos a la zona costera de playas y parques**.( ver pto. 4.1 para el análisis demográfico, económico y geográfico). Como vimos, **estos son los municipios B, CH y E.**

Pero luego, al agrupar los barrios **en 5 clústeres en función de los top 10 de sitios más populares**, vemos **que no quedaron tan homogéneamente agrupados geográficamente**, dado que algunos clústeres tenían barrios cercanos y lejanos a la zona costera de playas, como por ej. el clúster cero ( ver el 1er mapa con los 5 clústeres).

Por lo cual **fue necesario otra agrupación para el clúster cero,** y aquí se puede ver que los **barrios Cordón, Centro, Tres Cruces, Pocitos, la Blanqueada y Punta Gorda (sub-clúster cero o rojo en el segundo mapa ),** son los que cuentan con la **mayor cantidad de sitios dentro de las preferencias** buscadas con Foursquare, si bien **no se encuentran todos geográficamente cerca de la zona costera de playas.**

**Solo dos de los barrios mencionados se encuentran cerca de la zona costera, Pocitos y Punta Gorda. De estos dos elegiremos el más poblado.**

 **VS** 

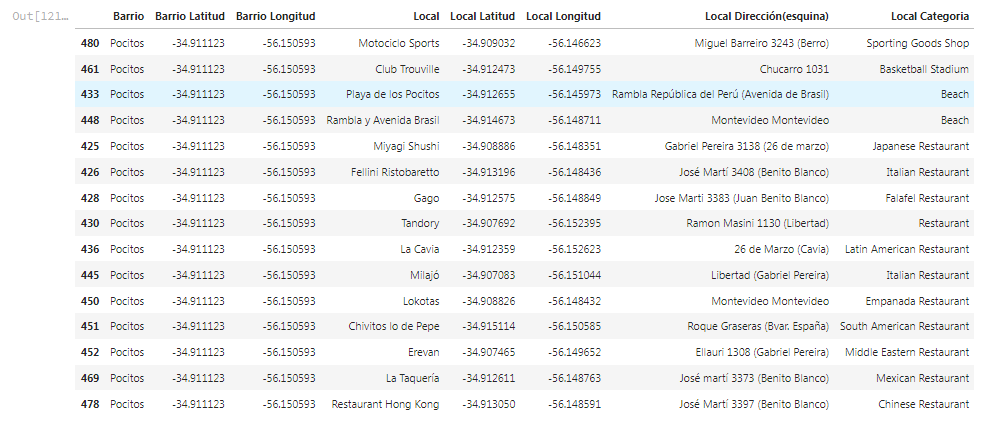
***NOTA : EN LAS JUPYTER NOTEBOOKS PARTE 1 Y 2 ADJUNTAS A ESTE TRABAJO SE PUEDE VER TODOS LOS RESULTADOS DE FORMA COMPLETA.***

# Conclusión final.

**En resumen,**  luego de agrupar primero en **5 clústeres** los 21 barrios que componen los municipios **preseleccionados B, CH y E,** y luego de sub-agrupar **en 3 clústeres** los 13 barrios que componían el clúster cero original, llegamos a los **6 barrios del sub-cluster cero**, que poseen las características preseleccionadas tanto desde el punto de vista demográfico como socioeconómico, y simultáneamente cuentan con la mayor cantidad de sitios acordes a la propuesta proyectada, **lo cual nos indica que la población objetivo propensa a incorporarla se encuentra en alguno de esos 6 barrios indicaos en el punto anterior**. Solo dos de ellos se encuentran **cercanos a la costa, Pocitos y Punta Gorda**, y de los dos, **el más poblado es Pocitos, con casi de 68000 habitantes.**

**Por lo cual, la recomendación final es instalar el proyectado centro deportivo, comercial y gastronómico en el barrio Pocitos.**







# Fuentes .

INE Censo 2011 : <https://www.ine.gub.uy/censos-2011>

<https://www.ine.gub.uy/c/document_library/get_file?uuid=4718ab8a-c64e-439c-ba8e-7d56371b6c89&groupId=10181>

INE ECH 2017 : <https://www.ine.gub.uy/web/guest/encuesta-continua-de-hogares1> ; <https://www.ine.gub.uy/c/document_library/get_file?uuid=75b6cb00-387e-40ee-a694-eb64314e81db&groupId=10181>

Gobierno Departamental : <https://municipios.montevideo.gub.uy/>; <https://es.wikipedia.org/wiki/Anexo:Barrios_de_Montevideo> ; <https://es.wikipedia.org/wiki/Anexo:Municipios_de_Montevideo>