

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
2- Datos utilizados.....	2
2.1- Datos de población	3
2.2 - Datos Socioeconómicos.....	4
2.3 - Datos Geográficos y coordenada.....	4
3- Forma de abordaje y metodología de análisis	5
3.1- Primera parte	5
3.2- Segunda Parte :	6
4- Resultados obtenidos.....	6
4.1 – Resultados preseleccion.	6
Análisis Demográfico.....	7
Análisis Económico.....	8
Análisis geográfico:	9
4.2 – Resultados con sitios de interés con Foursquare y clustering con K Means.	11
Resumen de resultados obtenidos	15
5- Debate de Resultados	16
6- Conclusión final.....	17
Fuentes	19

1- 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

2- Datos utilizados.

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

Biblioteca Documentos				Organizar por: Carpeta ▼
Datos				
Nombre	Fecha de modificación	Tamaño	Tipo	
 montevideo_cluster_cero_sorted	26/11/2021 15:45	3 KB	Archivo de valores separados por ci	
 montevideo_sitios_cluster_cero	26/11/2021 15:45	20 KB	Archivo de valores separados por ci	
 montevideo_barrios_merged	26/11/2021 14:17	5 KB	Archivo de valores separados por ci	
 montevideo_sport_foods_sorted	26/11/2021 14:16	4 KB	Archivo de valores separados por ci	
 montevideo_sport_foods_grouped	26/11/2021 14:16	7 KB	Archivo de valores separados por ci	
 montevideo_sport_foods_onehot	26/11/2021 14:15	18 KB	Archivo de valores separados por ci	
 df_montevideo_sport_foods_sitios	26/11/2021 14:15	22 KB	Archivo de valores separados por ci	
 df_montevideo_bche_sitios	26/11/2021 14:15	74 KB	Archivo de valores separados por ci	
 df_montevideo_socioeconom_municipio	25/11/2021 10:04	1 KB	Archivo de valores separados por ci	
 df_montevideo_encuesta	25/11/2021 9:28	19.098 KB	Archivo de valores separados por ci	
 df_montevideo_socioeconom	25/11/2021 9:28	377 KB	Archivo de valores separados por ci	
 df_montevideo_all_municipio	25/11/2021 9:28	1 KB	Archivo de valores separados por ci	
 df_montevideo_coord2	25/11/2021 9:28	3 KB	Archivo de valores separados por ci	
 Datos de busqueda Montevideo	25/11/2021 8:51	16 KB	Documento de Microsoft Word	
 Diccionario de Variables ECH 2017 (1)	25/11/2021 8:45	168 KB	Hoja de cálculo de Microsoft Excel!	
 H_2017_Terceros	24/11/2021 13:29	19.006 KB	Archivo de valores separados por ci	
 H_2017_Terceros	24/11/2021 13:28	49.075 KB	Hoja de cálculo de Microsoft Excel!	
 ECH2017Montevideo	24/11/2021 11:48	6.182 KB	Hoja de cálculo de Microsoft Excel	
 informacionfisicaysociodemograficapormunicipio2017	23/11/2021 18:12	201 KB	Adobe Acrobat Document	
 P_6_MVD_BARRIOS	23/11/2021 12:06	37 KB	Hoja de cálculo de Microsoft Excel!	
 Codificador de barrios	22/11/2021 12:20	29 KB	Hoja de cálculo de Microsoft Excel!	

2.1- Datos de población

```
In [82]: # Agregamos Los datos de poblacion
df_montevideo_all = df_montevideo.merge(df_montevideo_municipios, how = 'left')
df_montevideo_all
```

```
Out[82]:
```

	Barrio	Poblacion	Municipio
0	Pocitos	67992	CH
1	Cordón	42456	B
2	Unión	39880	D, E y F
3	La Paloma, Tomkinson	37424	A
4	Buceo	36998	CH y E
...
57	La Figurita	11273	C
58	La Blanqueada	9600	CH y E
59	Jacinto Vera	8767	C
60	Bañados de Carrasco	8705	F
61	Atahualpa	8607	C

```
62 rows x 4 columns
Suma de poblacion por municipios 1318755
Suma de poblacion por barrios 1318755
1318755
1318755
La suma de ambos DataFrames es igual
```

2.2 - Datos Socioeconómicos

```
In [147]: #Agrupamos por barrio y agregamos el Municipio
df_montevideo_socioeconom_municipio = df_montevideo_socioeconom_barrio.join(df_montevideo_municipios, how = 'left', lsuffix = '_call', rsuffix = '_municipio')
df_montevideo_socioeconom_municipio
```

```
Out[147]:
```

Barrio	Nivel-SocioEconómico	IngresoxHogar	Municipio
Carrasco	5.000000	41708.598394	E
Punta Carretas	5.000000	16753.442982	B y CH
Punta Gorda	4.897260	15921.404110	E
Pocitos	4.729483	14636.900456	CH
Carrasco Norte	4.561404	14906.409357	E
...
Villa García, Manga Rur.	1.337037	4543.707407	NaN
Casabo, Pajas Blancas	1.318182	4718.542208	NaN
Pta. Rieles, Bella Italia	1.300885	5194.234513	NaN
Manga, Toledo Chico	1.051020	3833.438776	D
Casavalle	1.000000	4663.377850	D

62 rows × 3 columns

Datos del clasificador de Nivel Socioeconómico del INE

DICCIONARIO DE VARIABLES DEL CUESTIONARIO DE LA ENCUESTA CONTINUA DE HOGARES DEL INSTITUTO NACIONAL DE ESTADÍSTICA. AÑO 2017 Registro de hogares

Registro de Hogares		ECH 2017	
DESCRIPCIÓN DE LA VARIABLE	NOMBRE VARIABLE	CATEGORÍAS	
		CÓDIGO	DESCRIPCIÓN Y OBSERVACIONES
HOGAR. IDENTIFICACIÓN			
IDENTIFICACIÓN DEL CUESTIONARIO	NUMERO		
AÑO DE LA ENCUESTA	ANIO		
MES DE LA ENCUESTA	MES		
DEPARTAMENTO	DPTO		Código correlativo del 1 al 19 comenzando por Montevideo y continuando alfabéticamente
NOMBRE DEL DEPARTAMENTO	NOMDPTO		
SECCIÓN CENSAL	SECC		
SEGMENTO CENSAL	SEGM		
LOCALIDAD AGRUPADA 2013	LOC_AGR_13		Código de las localidades agrupadas donde se aplica la ECH según marco censal 2011
NOMBRE DE LA LOCALIDAD AGRUPADA	NOM_LOC_AGR_13		Nombre de las localidades agrupadas donde se aplica la ECH según marco censal 2011
CENTRO COMUNAL ZONAL	CCZ		Para el departamento de Montevideo
BARRIO	BARRIO		Para el departamento de Montevideo
NOMBRE DEL BARRIO	NOMBARRIO		Para el departamento de Montevideo
		1	Montevideo - Nivel económico bajo
		2	Montevideo - Nivel económico medio - bajo
		3	Montevideo - Nivel económico medio
		4	Montevideo - Nivel económico medio - alto
		5	Montevideo - Nivel económico alto
		6	Zona metropolitana

2.3 - Datos Geográficos y coordenada.

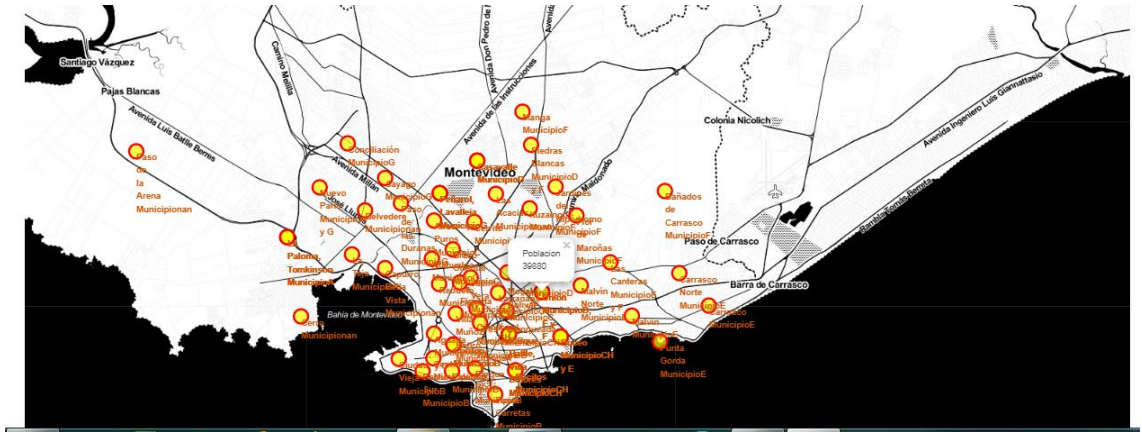
```
In [177]: # Definimos una funcion para determinar las coordenadas para cada municipio de Montevideo
# fuente : https://es.stackoverflow.com/questions/338223/usar-geopy-en-un-dataframe-para-hallar-latitud-y-longitud
def get_coords(barrio):
    geolocator = Nominatim(user_agent="to_explorer")
    address = f'({barrio}), Montevideo, Uruguay'
    location = geolocator.geocode(address)
    if location is not None:
        latitude = location.latitude
        longitude = location.longitude
    else:
        latitude = longitude = 'No Existe'
    return pd.Series({"Latitude": latitude, "Longitude": longitude})

df_montevideo_coord = pd.concat([df_montevideo_all, df_montevideo_all['Barrio'].apply(get_coords)], axis=1)
df_montevideo_coord
```

```
Out[177]:
```

	Barrio	Poblacion	Municipio	Latitud	Longitud
0	Pocitos	67992	CH	-34.911123	-56.150593
1	Cordón	42456	B	-34.900827	-56.180125
2	Unión	39880	D, E y F	-34.880236	-56.137688
3	La Paloma, Tomkinson	37424	A	-34.858785	-56.25942
4	Buceo	36998	CH y E	-34.8978	-56.128658
...
57	La Figurita	11273	C	-34.876757	-56.17672
58	La Blanqueada	9600	CH y E	-34.8877	-56.154518
59	Jacinto Vera	8767	C	-34.874534	-56.171774
60	Bañados de Carrasco	8705	F	-34.840727	-56.079019
61	Atahualpa	8607	C	-34.867351	-56.190297

62 rows × 5 columns



3- 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, así 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 desglosados 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 librería 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-Manhattan, 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, repetiríamos 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.

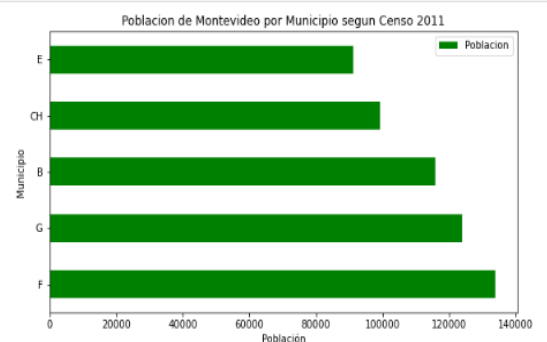
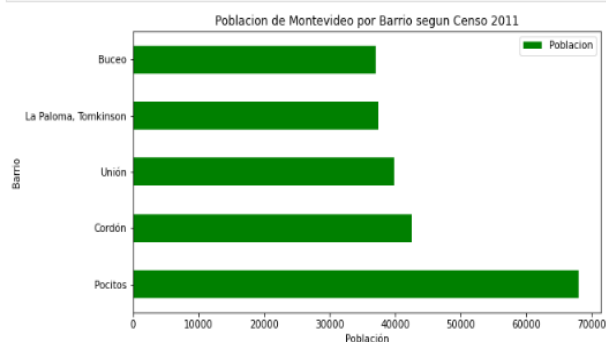
Análisis Demográfico

```
in [84]: # Agrupamos la poblacion por municipio
df_montevideo_all_municipio = df_montevideo_all.groupby('Municipio', as_index=False, sort=True).sum()
df_montevideo_all_municipio.sort_values('Poblacion', ascending=False, inplace=True)
df_montevideo_all_municipio
```

Out[84]:

	Municipio	Barrio	Poblacion
16	F	Villa García, Manga RuralMangaMaroñas, Parque ...	133776
17	G	Peñarol, LavallejaColón Centro y NoroesteConci...	123867
3	B	CordónCentroPalermoParque RodóBarrio SurCiudad...	115706
9	CH	PocitosParque Batlle, Villa Dolores	99145
14	E	MalvínMalvín NorteCarrascoPunta GordaCarrasco ...	91104
0	A	La Paloma, TomkinsonCasabó, Pajas BlancasLa Teja	88581
6	C	Brazo OrientalReductoLa FiguralJacinto VeraAt...	59555
11	D	CasavalleManga, Toledo Chico	59390
10	CH y E	BuceoLa Blanqueada	46598
12	D y F	Villa EspañolaPiedras Blancas	46041
2	A y G	Nuevo ParísLezica, Melilla	45903
5	B y CH	Punta CarretasTres Cruces	40107
13	D, E y F	Unión	39880
8	C y D	Aires PurosMercado Modelo, Bolívar	31305
4	B y C	AguadaLa Comercial	30001
15	E y F	Las Canteras	22974
1	A y C	Prado, Nueva Savona	20199
7	C y CH	Larrañaga	17956

```
In [86]: # Graficamos los top 5 para barrios y municipios
fig = plt.figure()
ax0 = fig.add_subplot(1,2,1)
ax1 = fig.add_subplot(1,2,2)
#Top 5 barrios
df_montevideo.set_index('Barrio').head(5).plot(kind='barh', figsize=(20,5), color='green', ax=ax0)
ax0.set_xlabel('Población')
ax0.set_ylabel('Barrio')
ax0.set_title('Poblacion de Montevideo por Barrio segun Censo 2011')
#Top 5 municipios
df_montevideo_all_municipio.set_index('Municipio').head(5).plot(kind='barh', figsize=(20,5), color='green', ax=ax1)
ax1.set_xlabel('Población')
ax1.set_ylabel('Municipio')
ax1.set_title('Poblacion de Montevideo por Municipio segun Censo 2011')
plt.show()
```



Análisis Económico

```
In [149_
#Agrupamos por municipio y calculamos la media de Nivel socioeconómico e Ingresos por hogar, ordenado de mayor a menor
df_montevideo_socioeconom_municipio = df_montevideo_socioeconom_municipio.groupby('Municipio', as_index = False, sort = True).mean()
df_montevideo_socioeconom_municipio.sort_values('Nivel-SocioEconómico', ascending=False, inplace = True)
df_montevideo_socioeconom_municipio.set_index('Municipio', inplace = True)
df_montevideo_socioeconom_municipio
```

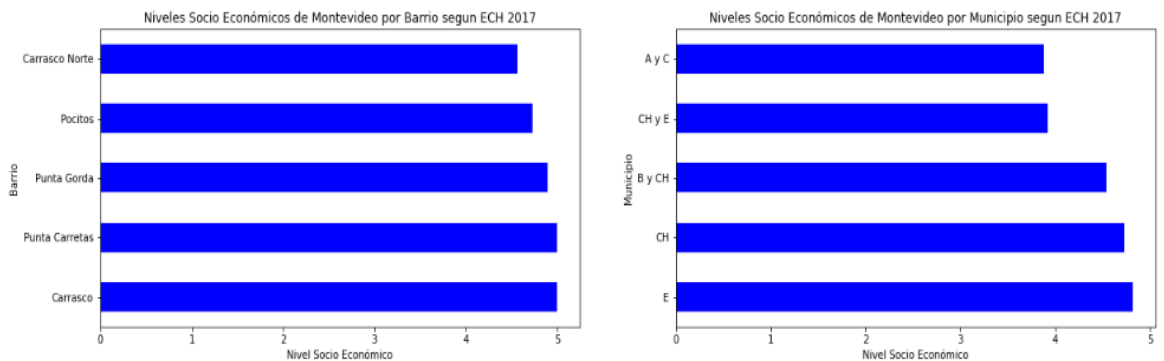
Out[149_

	Nivel-SocioEconómico	IngresosHogar
Municipio		
E	4.819555	24178.803953
CH	4.729483	14636.900456
B y CH	4.543561	12299.547249
CH y E	3.926684	7442.671847
A y C	3.883621	6193.896552
B	3.521299	5832.318959
C	3.426126	5680.093443
C y D	3.357143	7104.066964
B y C	3.314775	5328.958118
G	3.294286	5932.247581
E y F	2.802867	4516.724014
A y G	2.337580	4414.719745
A	2.062645	4764.164793
D y F	1.535971	4933.381295
F	1.426573	4561.744755
D	1.025510	4248.408313

```
In [152_
# Graficamos los top 5 para barrios y municipios
fig2 = plt.figure()
ax0 = fig2.add_subplot(1,2,1)
ax1 = fig2.add_subplot(1,2,2)
#Top 5 barrios
df_montevideo_socioeconom_barrio['Nivel-SocioEconómico'].head(5).plot(kind='barh', figsize=(20,5), color = 'blue', ax=ax0)
ax0.set_xlabel('Nivel Socio Económico')
ax0.set_ylabel('Barrio')
ax0.set_title('Niveles Socio Económicos de Montevideo por Barrio segun ECH 2017')

#Top 5 municipios
df_montevideo_socioeconom_municipio['Nivel-SocioEconómico'].head(5).plot(kind='barh', figsize=(20,5), color = 'blue', ax=ax1)
ax1.set_xlabel('Nivel Socio Económico')
ax1.set_ylabel('Municipio')
ax1.set_title('Niveles Socio Económicos de Montevideo por Municipio segun ECH 2017')

plt.show()
```



Análisis geográfico:

```
In [177]: # Definimos una funcion para determinar Las coordenadas para cada municipio de Montevideo
# fuente : https://es.stackoverflow.com/questions/330223/usar-geopy-en-un-dataframe-para-hallar-latitud-y-longitud
def get_coords(barrio):
    geolocator = Nominatim(user_agent="to_explorer")
    address = f'{barrio}, Montevideo, Uruguay'
    location = geolocator.geocode(address)
    if location is not None:
        latitude = location.latitude
        longitude = location.longitude
    else:
        latitude = longitude = 'No Existe'

    return pd.Series({"Latitud": latitude, "Longitud": longitude})

df_montevideo_coord = pd.concat([df_montevideo_all, df_montevideo_all['Barrio'].apply(get_coords)], axis=1)
df_montevideo_coord
```

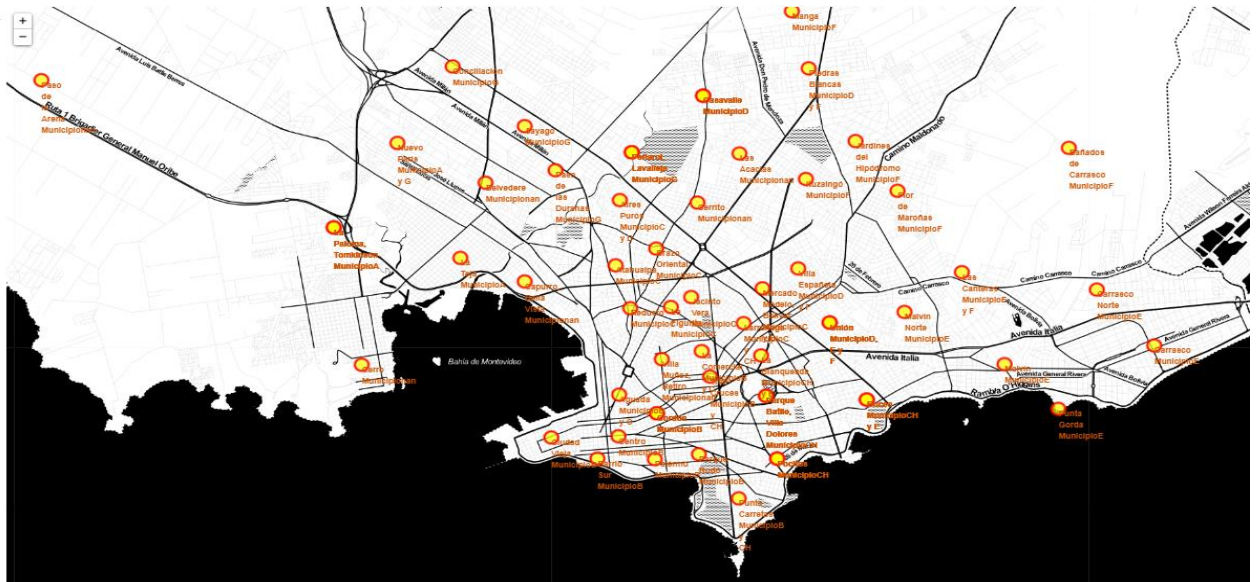
```
Out[177]:
```

	Barrio	Poblacion	Municipio	Latitud	Longitud
0	Pocitos	67992	CH	-34.911123	-56.150593
1	Cordón	42456	B	-34.900827	-56.180125
2	Unión	39880	D, E y F	-34.880236	-56.137688
3	La Paloma, Tomkinson	37424	A	-34.858785	-56.25942
4	Buceo	36998	CH y E	-34.8978	-56.128658
...
57	La Figurita	11273	C	-34.876757	-56.17672
58	La Blanqueada	9600	CH y E	-34.8877	-56.154518
59	Jacinto Vera	8767	C	-34.874534	-56.171774
60	Bañados de Carrasco	8705	F	-34.840727	-56.079019
61	Atahualpa	8607	C	-34.867351	-56.190297

62 rows × 5 columns

```
In [182]: # Añadimos marcadores de los municipios y vecindarios
for lat, lng, municipio, barrio, poblacion in zip(df_montevideo_coord['Latitud'], df_montevideo_coord['Longitud'],
                                                df_montevideo_coord['Municipio'], df_montevideo_coord['Barrio'], df_montevideo_coord['Poblacion']):
    label = 'Poblacion {}'.format(poblacion)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=7,
        popup=label,
        color='red',
        fill=True,
        fill_color='yellow',
        fill_opacity=0.7,
        parse_html=False).add_to(map_montevideo)
    folium.map.Marker(
        [lat, lng],
        icon=DivIcon(
            icon_size=(20, 20),
            icon_anchor=(0, 0),
            html='<div style="font-size: 6; color: #d35400;"><b>{}</b></div>' % '{} {}'.format(barrio, municipio))).add_to(map_montevideo)

map_montevideo
```



Si bien del análisis demográfico se muestra que los municipios más poblados son el F y G, los mismos se encuentran lejos 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 preselección. Esto lo podemos ver más claro en el siguiente mapa extraído 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 :

```
# De acuerdo a Lo requerido, se selecciona Los barrios de Los municipios B, CH y E
# Se unen los dataframes y se eliminan duplicados, reseteando el índice.
# Se descartan municipios que no fueron asignados, asignando a uno fuera de la selección ("A")
df_montevideo_coord2['Municipio'] = df_montevideo_coord2['Municipio'].fillna('A')
df_montevideo_b = df_montevideo_coord2[df_montevideo_coord2['Municipio'].str.contains('B')]
df_montevideo_ch = df_montevideo_coord2[df_montevideo_coord2['Municipio'].str.contains('CH')]
df_montevideo_e = df_montevideo_coord2[df_montevideo_coord2['Municipio'].str.contains('E')]
df_montevideo_bche = pd.concat([df_montevideo_b, df_montevideo_ch, df_montevideo_e])
df_montevideo_bche.drop_duplicates(['Barrio'], inplace = True)
df_montevideo_bche.reset_index(drop = True)
```

Out[4]:

	Barrio	Poblacion	Municipio	Latitud	Longitude
0	Cordón	42456	B	-34.900827	-56.180125
1	Punta Carretas	24181	B y CH	-34.920202	-56.160065
2	Centro	22120	B	-34.906067	-56.189656
3	Aguada	18557	B y C	-34.896676	-56.189366
4	Tres Cruces	15926	B y CH	-34.892518	-56.167113
5	Palermo	12992	B	-34.911288	-56.180698
6	Parque Rodó	12944	B	-34.910082	-56.169778
7	Barrio Sur	12639	B	-34.911202	-56.194784
8	Ciudad Vieja	12555	B	-34.906351	-56.205979
9	La Comercial	11444	B y C	-34.886749	-56.169093
10	Pocitos	67992	CH	-34.911123	-56.150593
11	Buceo	36998	CH y E	-34.897800	-56.128658
12	Parque Batlle, Villa Dolores	31153	CH	-34.896868	-56.153383
13	Larrañaga	17956	C y CH	-34.880441	-56.158788
14	La Blanqueada	9600	CH y E	-34.887700	-56.154518
15	Unión	39880	D, E y F	-34.880236	-56.137688
16	Malvín	28102	E	-34.889692	-56.094768
17	Las Canteras	22974	E y F	-34.868839	-56.105227
18	Malvín Norte	19916	E	-34.877897	-56.119281
19	Carrasco	15488	E	-34.885577	-56.058150
20	Punta Gorda	14259	E	-34.899841	-56.081571
21	Carrasco Norte	13339	E	-34.872898	-56.072086

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).

```
In [19]: # Filtramos el dataframe por las categorías que tengan que ver con deportes, gym, parques, outdoors, playas, etc. y lugares de comida.
df_montevideo_gym = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Gym')]
df_montevideo_sport = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Sport')]
df_montevideo_outdoors = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Outdoor')]
df_montevideo_parques = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Park')]
df_montevideo_plazas = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Plaza')]
df_montevideo_estadios = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Stadium')]
df_montevideo_playas = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Beach')]
df_montevideo_basket = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Basket')]
df_montevideo_futbol = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Soccer')]
df_montevideo_golf = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Golf')]
df_montevideo_tennis = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Tennis')]
df_montevideo_restaurant = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Restaurant')]
df_montevideo_food = df_montevideo_bche_sitios[df_montevideo_bche_sitios['Local Categoría'].str.contains('Food')]

df_montevideo_sport_foods_sitios = pd.concat([df_montevideo_gym, df_montevideo_sport, df_montevideo_outdoors, df_montevideo_parques,
                                             df_montevideo_plazas, df_montevideo_estadios, df_montevideo_playas, df_montevideo_basket,
                                             df_montevideo_futbol, df_montevideo_tennis, df_montevideo_golf, df_montevideo_food,
                                             df_montevideo_restaurant])

df_montevideo_sport_foods_sitios.drop_duplicates(['Local'], inplace = True)

# Exportamos y mostramos el nuevo dataframe solo con los sitios que nos interesan.
df_montevideo_sport_foods_sitios.to_csv('df_montevideo_sport_foods_sitios.csv', index = False)
df_montevideo_sport_foods_sitios.reset_index(drop=True)
```

```
Out[19]:
```

	Barrio	Barrio Latitud	Barrio Longitud	Local	Local Latitud	Local Longitud	Local Dirección(esquina)	Local Categoría
0	Cordón	-34.900827	-56.180125	Asociación Cristiana de Jóvenes	-34.900806	-56.176413	Colonia 1870 (Eduardo Acevedo)	Gym / Fitness Center
1	Cordón	-34.900827	-56.180125	Club BPS	-34.901414	-56.180063	Mercedes 1723 (Magallanes)	Gym
2	Punta Carretas	-34.920202	-56.160065	Instinto Crossfit	-34.917322	-56.157522	21 De Setiembre 2745 (Luis De La Torre)	Gym / Fitness Center
3	Punta Carretas	-34.920202	-56.160065	Beauty Planet	-34.921482	-56.156163	Blanca Del Tabaré 2990 (Elauri)	Gym
4	Punta Carretas	-34.920202	-56.160065	Via Aqua Spa	-34.924113	-56.157570	Victor Soliño 375	Gym
...
156	Carrasco	-34.885577	-56.058150	Dackel	-34.889256	-56.057562	Gabriel Otero 6438 (Arocena)	German Restaurant
157	Carrasco	-34.885577	-56.058150	Hoy Te Quiero	-34.889672	-56.056654	Arocena (Otero)	Restaurant
158	Carrasco	-34.885577	-56.058150	Tajamar Restó	-34.883163	-56.061224	Lieja 6416 (Divina Comedia)	American Restaurant
159	Punta Gorda	-34.899841	-56.081571	Hemingway	-34.898744	-56.082226	Rambía República de México 5535	Seafood Restaurant
160	Punta Gorda	-34.899841	-56.081571	Chivitos Marcos	-34.896074	-56.080180	Coimbra (Gral. Paz)	Latin American Restaurant

161 rows x 8 columns

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

5. Análisis de clustering por K-Means

5.1 Preparación y armado de los clusters

Ejecutemos k-means para agrupar los barrios en 5 agrupaciones.

```
[33]: # establecer el número de agrupaciones
kclusters = 5

montevideo_grouped_clustering = montevideo_sport_foods_grouped.drop('Barrio', 1)

# ejecutar k-means
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(montevideo_grouped_clustering)

# revisar las etiquetas de las agrupaciones generadas para cada fila del dataframe
kmeans.labels_[0:10]
```

```
[33]: array([2, 0, 4, 0, 4, 0, 0, 0, 0, 3])
```

5.2 Preparación y armado del Dataframe

Generemos un nuevo dataframe que incluya la agrupación así como los 10 sitios más populares de cada barrio.

```
[34]: # añadir etiquetas
montevideo_sport_foods_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

montevideo_barrios_merged = df_montevideo_bche

# juntar manhattan_grouped con manhattan_data
montevideo_barrios_merged = montevideo_barrios_merged.join(montevideo_sport_foods_sorted.set_index('Barrio'), on='Barrio')

montevideo_barrios_merged.head() # revisar las últimas columnas
```

```
In [41]:
```

	Barrio	Poblacion	Municipio	Latitud	Longitud	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Cordón	42456	B	-34.900827	-56.180125	0	South American Restaurant	Basketball Court	Plaza	Gym / Fitness Center	Gym	Soccer Field	American Restaurant	Other Great Outdoors	Park	Pedestrian Plaza
1	Punta Carretas	24181	B y CH	-34.920202	-56.160065	0	Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Italian Restaurant	Gym / Fitness Center	Gym	Mexican Restaurant	Golf Course	South American Restaurant	Modern European Restaurant
2	Centro	22120	B	-34.906067	-56.189656	0	Plaza	Restaurant	Gym / Fitness Center	Food Truck	Chinese Restaurant	Latin American Restaurant	Kebab Restaurant	Italian Restaurant	Spanish Restaurant	Gym
3	Aguada	18557	B y C	-34.896676	-56.189366	2	Food & Drink Shop	American Restaurant	South American Restaurant	Modern European Restaurant	Other Great Outdoors	Park	Pedestrian Plaza	Plaza	Restaurant	Seafood Restaurant
4	Tres Cruces	15926	B y CH	-34.892518	-56.167113	0	Latin American	Food Court	Chinese Restaurant	Plaza	Soccer Field	Southern / Soul Food	Other Great Outdoors	Park	Pedestrian Plaza	Restaurant

```
Out[41]:
```

Podemos verlos en el mapa diferenciados por colores:

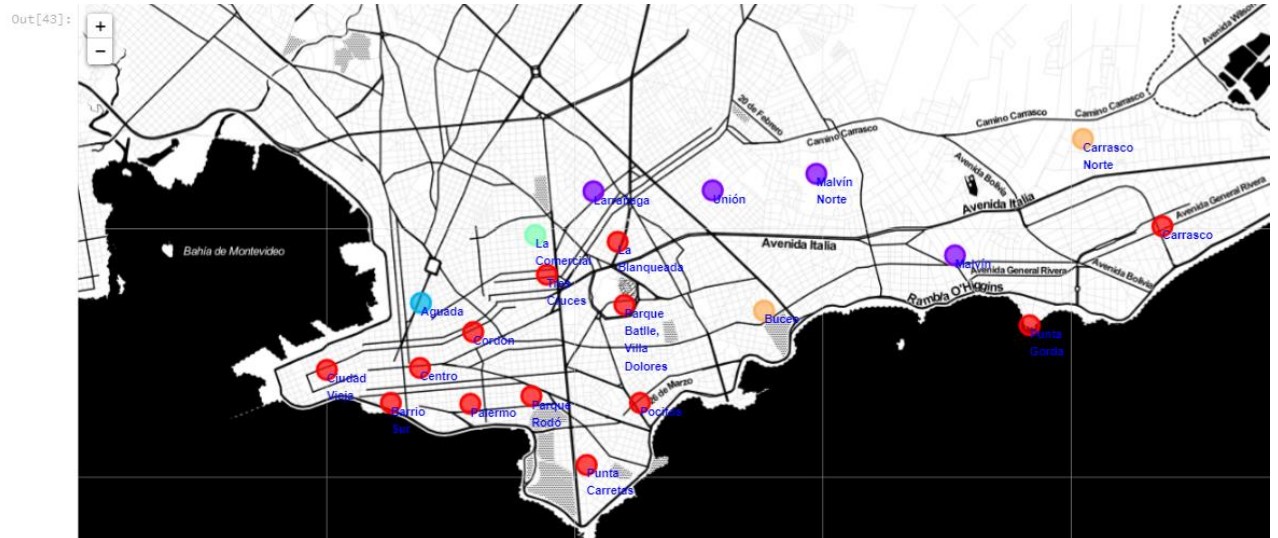
```
In [43]:
```

```
# crear mapa
montevideo_barrios_clusters = folium.Map(location=[latitude, longitude], zoom_start=13, tiles = 'Stamen Toner')

# establecer el esquema de color para las agrupaciones
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# añadir marcadores al mapa
markers_colors = []
for lat, lon, bar, cluster, first in zip(montevideo_barrios_merged['Latitud'], montevideo_barrios_merged['Longitud'],
montevideo_barrios_merged['Barrio'], montevideo_barrios_merged['Cluster Labels'],
montevideo_barrios_merged['1st Most Common Venue']):
    label = folium.Popup(str(bar) + ' Cluster ' + str(cluster) + ' - ' + str(first), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=10,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(montevideo_barrios_clusters)
    folium.map.Marker(
        [lat, lon],
        icon=DivIcon(
            icon_size=(20,20),
            icon_anchor=(0,0),
            html='<div style="font-size: 5; color: blue;"><b>%s</b></div>' % '{}'.format(bar))).add_to(montevideo_barrios_clusters)

montevideo_barrios_clusters
```



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


```
In [44]: montevideo_barrios_merged.loc[montevideo_barrios_merged['Cluster Labels'] == 0, montevideo_barrios_merged.columns[[0]+[1] + list(range(5, 11))]]
```

```
Out[44]:
```

	Barrio	Poblacion	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cordón	42456	0	South American Restaurant	Basketball Court	Plaza	Gym / Fitness Center	Gym
1	Punta Carretas	24181	0	Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Italian Restaurant	Gym / Fitness Center
2	Centro	22120	0	Plaza	Restaurant	Gym / Fitness Center	Food Truck	Chinese Restaurant
4	Tres Cruces	15926	0	Latin American Restaurant	Food Court	Chinese Restaurant	Plaza	Soccer Field
5	Palermo	12992	0	Restaurant	Gym / Fitness Center	Other Great Outdoors	Basketball Court	Basketball Stadium
6	Parque Rodó	12944	0	Restaurant	Gym	American Restaurant	Tapas Restaurant	Italian Restaurant
7	Barrio Sur	12639	0	American Restaurant	Food & Drink Shop	Restaurant	Vegetarian / Vegan Restaurant	Beach
8	Ciudad Vieja	12555	0	Restaurant	Plaza	Vegetarian / Vegan Restaurant	Food & Drink Shop	Pedestrian Plaza
10	Pocitos	67992	0	Beach	Italian Restaurant	Sporting Goods Shop	South American Restaurant	Restaurant
12	Parque Batlle, Villa Dolores	31153	0	Soccer Stadium	Athletics & Sports	Track Stadium	Park	Sports Club
14	La Blanqueada	9600	0	Basketball Court	Athletics & Sports	Latin American Restaurant	American Restaurant	South American Restaurant
18	Carrasco	15488	0	Restaurant	American Restaurant	Tennis Court	Park	Comfort Food Restaurant
19	Punta Gorda	14259	0	Plaza	Latin American Restaurant	Beach	Seafood Restaurant	American Restaurant

```
In [45]: print('La poblacion del cluster 0 es :',montevideo_barrios_merged.loc[montevideo_barrios_merged['Cluster Labels'] == 0, montevideo_barrios_merged.columns[[1]]].sum())
```

La poblacion del cluster 0 es : Poblacion 294305
dtype: int64

Compuesto por los siguientes barrios:

```
In [76]: print(montevideo_cluster_cero['Barrio'])
```

```
0          Cordón
1      Punta Carretas
2          Centro
3      Tres Cruces
4          Palermo
5      Parque Rodó
6      Barrio Sur
7      Ciudad Vieja
8          Pocitos
9  Parque Batlle, Villa Dolores
10     La Blanqueada
11          Carrasco
12     Punta Gorda
Name: Barrio, dtype: object
```

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).

La poblacion del sub-cluster 0 es : Poblacion 172353
dtype: int64

	Barrio	Poblacion	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cordón	42456	0	South American Restaurant	Basketball Court	Plaza	Gym / Fitness Center	Gym
2	Centro	22120	0	Plaza	Restaurant	Gym / Fitness Center	Food Truck	Chinese Restaurant
3	Tres Cruces	15926	0	Latin American Restaurant	Food Court	Chinese Restaurant	Plaza	Soccer Field
8	Pocitos	67992	0	Beach	Italian Restaurant	Sporting Goods Shop	South American Restaurant	Restaurant
10	La Blanqueada	9600	0	Basketball Court	Athletics & Sports	Latin American Restaurant	American Restaurant	South American Restaurant
12	Punta Gorda	14259	0	Plaza	Latin American Restaurant	Beach	Seafood Restaurant	American Restaurant

print('La poblacion del sub-cluster 1 es :',montevideo_cluster_cero_merged.loc[montevideo_cluster_cero_merged['Cluster Labels'] == 1, montevideo_cluster_cero_merged.loc[montevideo_cluster_cero_merged['Cluster Labels'] == 1, montevideo_cluster_cero_merged.columns[0]+1] + list(range(5, 11))

La poblacion del sub-cluster 1 es : Poblacion 31153
dtype: int64

	Barrio	Poblacion	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Parque Batlle, Villa Dolores	31153	1	Soccer Stadium	Athletics & Sports	Track Stadium	Park	Sports Club

print('La poblacion del sub-cluster 2 es :',montevideo_cluster_cero_merged.loc[montevideo_cluster_cero_merged['Cluster Labels'] == 2, montevideo_cluster_cero_merged.loc[montevideo_cluster_cero_merged['Cluster Labels'] == 2, montevideo_cluster_cero_merged.columns[0]+1] + list(range(5, 11))

La poblacion del sub-cluster 2 es : Poblacion 90799
dtype: int64

	Barrio	Poblacion	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Punta Carretas	24181	2	Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Italian Restaurant	Gym / Fitness Center
4	Palermo	12992	2	Restaurant	Gym / Fitness Center	Other Great Outdoors	Basketball Court	Basketball Stadium
5	Parque Rodó	12944	2	Restaurant	Gym	American Restaurant	Tapas Restaurant	Italian Restaurant
6	Barrio Sur	12639	2	American Restaurant	Food & Drink Shop	Restaurant	Vegetarian / Vegan Restaurant	Beach
7	Ciudad Vieja	12555	2	Restaurant	Plaza	Vegetarian / Vegan Restaurant	Food & Drink Shop	Pedestrian Plaza

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

5- Debate de Resultados

El análisis demográfico y económico permiten determinar 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.**

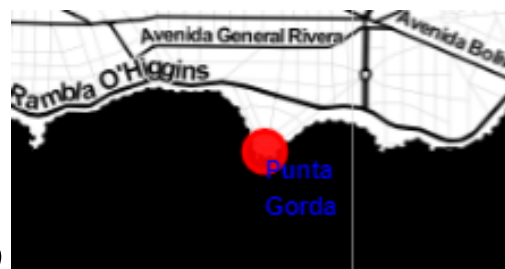
Pero luego, al agrupar los barrios en 5 clústeres en función de los top 10 de sitios más populares, vemos **que no quedaron tan homogéneamente agrupados geográficamente**, dado que algunos clústeres tenían barrios cercanos y lejanos a la zona costera de playas, como por ej. el clúster cero (ver el 1er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

6- 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 indicados 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**.

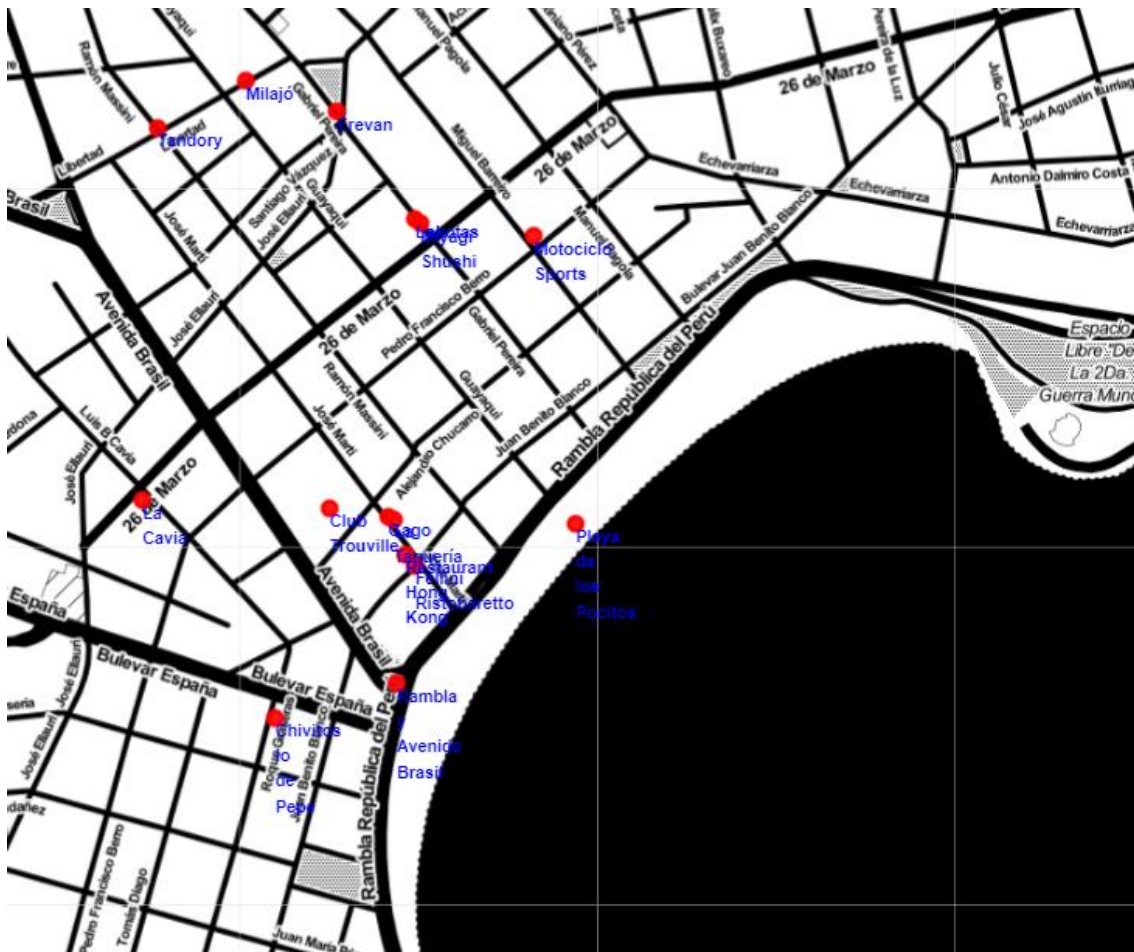
Por lo cual, la recomendación final es instalar el proyectado centro deportivo, comercial y gastronómico en el barrio Pocitos.



n:

	Barrio	Poblacion
0	Cordón	42456
2	Centro	22120
3	Tres Cruces	15926
8	Pocitos	67992
10	La Blanqueada	9600
12	Punta Gorda	14259

Out[121]	Barrio	Barrio Latitud	Barrio Longitud	Local	Local Latitud	Local Longitud	Local Dirección(esquina)	Local Categoría
480	Pocitos	-34.911123	-56.150593	Motociclo Sports	-34.909032	-56.146623	Miguel Barreiro 3243 (Berro)	Sporting Goods Shop
461	Pocitos	-34.911123	-56.150593	Club Trouville	-34.912473	-56.149755	Chucarro 1031	Basketball Stadium
433	Pocitos	-34.911123	-56.150593	Playa de los Pocitos	-34.912655	-56.145973	Rambla República del Perú (Avenida de Brasil)	Beach
448	Pocitos	-34.911123	-56.150593	Rambla y Avenida Brasil	-34.914673	-56.148711	Montevideo Montevideo	Beach
425	Pocitos	-34.911123	-56.150593	Miyagi Shushi	-34.908886	-56.148351	Gabriel Pereira 3138 (26 de marzo)	Japanese Restaurant
426	Pocitos	-34.911123	-56.150593	Fellini Ristobaretto	-34.913196	-56.148436	José Martí 3408 (Benito Blanco)	Italian Restaurant
428	Pocitos	-34.911123	-56.150593	Gago	-34.912575	-56.148849	Jose Martí 3383 (Juan Benito Blanco)	Falafel Restaurant
430	Pocitos	-34.911123	-56.150593	Tandory	-34.907692	-56.152395	Ramon Masini 1130 (Libertad)	Restaurant
436	Pocitos	-34.911123	-56.150593	La Cavia	-34.912359	-56.152623	26 de Marzo (Cavia)	Latin American Restaurant
445	Pocitos	-34.911123	-56.150593	Milajó	-34.907083	-56.151044	Libertad (Gabriel Pereira)	Italian Restaurant
450	Pocitos	-34.911123	-56.150593	Lokotas	-34.908826	-56.148432	Montevideo Montevideo	Empanada Restaurant
451	Pocitos	-34.911123	-56.150593	Chivitos lo de Pepe	-34.915114	-56.150585	Roque Graseras (Bvar. España)	South American Restaurant
452	Pocitos	-34.911123	-56.150593	Erevan	-34.907465	-56.149652	Elauri 1308 (Gabriel Pereira)	Middle Eastern Restaurant
469	Pocitos	-34.911123	-56.150593	La Taquería	-34.912611	-56.148763	José Martí 3373 (Benito Blanco)	Mexican Restaurant
478	Pocitos	-34.911123	-56.150593	Restaurant Hong Kong	-34.913050	-56.148591	José Martí 3397 (Benito Blanco)	Chinese Restaurant



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:Barrios_de_Montevideo) ;
[https://es.wikipedia.org/wiki/Anexo:Municipios de Montevideo](https://es.wikipedia.org/wiki/Anexo:Municipios_de_Montevideo)