

Curso Ciencia de Datos IBM - Coursera

Tarea final – Proyecto Coursera Capstone

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

Tarea Semana 1 (semana 4 del curso)

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:Barrios_de_Montevideo), https://es.wikipedia.org/wiki/Anexo:Municipios_de_Montevideo

2- Forma de abordaje y metodología de análisis

En una primera parte vamos a simplificar el marco de datos para el análisis, seleccionando aquellos municipios que cuentan con las características del público objetivo.

Montevideo posee 8 municipios, y si bien en todos la población está integrada y es heterogénea en cuanto a aspectos socioeconómicos, en algunos municipios predominan los estratos socioeconómicos más altos (de clase media a clase media alta y clase alta) y en otros predominan los de estratos socioeconómico más bajo. Por lo cual buscaremos preseleccionar aquellos municipios, y dentro de ellos aquellos barrios o vecindarios , con mayor nivel socioeconómico.

Por otro lado, hay municipios con mayor densidad demográfica que otros, generalmente los que tienen mayor acceso a servicios, transporte, y cuentan con mayor historia en la ciudad, por lo cual desde el punto de vista demográfico preseleccionaremos aquellos municipios, y dentro de ellos aquellos barrios, con mayor población.

Finalmente, dado que se pretende contar con espacios al aire libre cercanos, así como cercanía a la costa, preseleccionaremos aquellos municipios y dentro de ellos aquellos barrios, 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 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 pandas de Python. Cabe decir que solamente tomaremos los datos del área urbana, no así del área rural del departamento mencionado.

Luego de completada esta primera parte, y preseleccionados los municipios y barrios, 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.

Determinados los sitios de interés, procederemos a aplicar la metodología vista en el curso para los barrios de Manhattan, Nueva York, para el armado y depuración de los dataframes de Python-pandas, y para luego aplicar clustering usando K-means, como se vio en el curso. De los clústeres obtenidos, seleccionaremos aquellos con mayor población (no procederemos a seleccionar por nivel socioeconómico, dado que como señalamos, los municipios son bastante heterogéneos en ese aspecto), 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.

Por lo cual el trabajo propuesto cuenta de dos partes: una de preselección usando dataframes y pandas, y otra de análisis de los municipios preseleccionados, usando Foursquare, y clustering con K-means.

3- 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 búsqueda Montevideo	25/11/2021 8:51	16 KB	Documento de Microsoft Word	
 Dicccionario 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!	

3.1- 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
```

3.2 – 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]:
```

	Nivel-SocioEconómico	IngresoxHogar	Municipio
Barrio			
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 x 4 columns

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

Registro de hogares

DESCRIPCIÓN DE LA VARIABLE		ECH 2017	
		NOMBRE VARIABLE	CATEGORÍAS
			CÓDIGO
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

3.3 Datos Geográficos y coordenada.

```
In [177]: # Definimos una funcion para determinar Las coordenadas para cada municipio de Toronto
# 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'

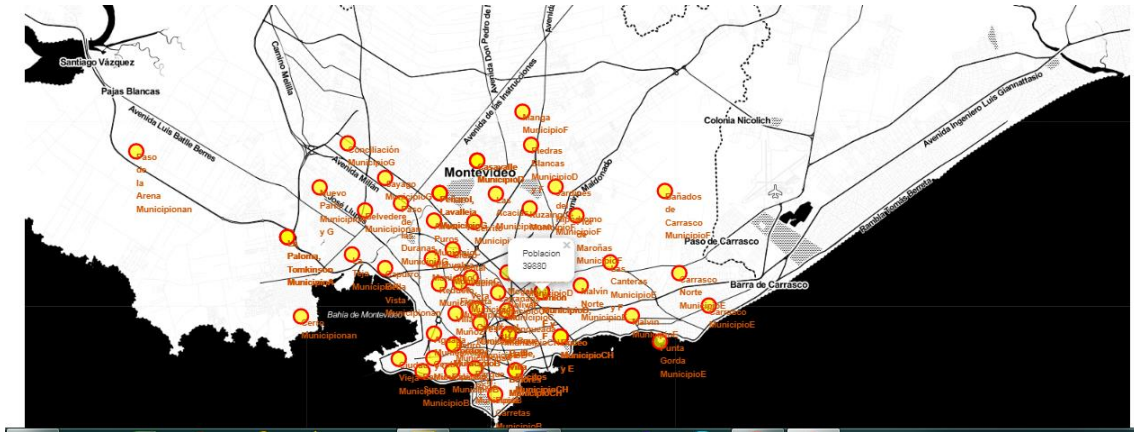
    return pd.Series({"Latitud": latitude, "Longitud": longitude})

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



27-11-2021

HB.