

Proyecto Ciencia de datos Aplicada IBM

Alejandro Thompson

24/01/2023

Contenido del informe

1. Introducción:
2. Sección de datos:
3. Metodología:
4. Resultados:
5. Debate:
6. Conclusiones:



Introducción

- Problema a resolver:

Como encontrar un apartamento en Manhattan con las siguientes condiciones:

- Apartamento con min 2 dormitorios
- Alquiler mensual no superior a US\$7000/mes
- Situado a poca distancia ($\leq 1,0$ milla, 1,6 km) de una estación de metro en Manhattan
- Lugares y comodidades como en mi residencia actual.

- Público interesado

La metodología, las herramientas y la estrategia utilizadas en este proyecto son pertinentes para una persona o entidad que esté considerando mudarse a una gran ciudad de EE.UU., Europa o Asia, Del mismo modo, puede ser un enfoque útil para explorar la apertura de un nuevo negocio. El uso de datos de FourSquare y de técnicas cartográficas combinadas con el análisis de datos ayudará a resolver las cuestiones clave planteadas. Por último, este proyecto es un buen caso práctico para una persona que esté desarrollando habilidades de Data Science.

Sección de datos

- Geodatos de la residencia actual en Singapur con locales establecidos mediante Foursquare.
- Lista de barrios de Manhattan (MH) con locales agrupados establecidos mediante Foursquare (como en Course Lab).
https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods
- - Lista de estaciones de metro de Manhattan con direcciones y datos geográficos (lat,long):
https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan ,
(<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>)
- - Listado de apartamentos en alquiler en la zona de Manhattan con información sobre ubicación del barrio, dirección, número de camas, superficie, precio mensual del alquiler y complementado con datos geográficos vía Nominatim.
<http://www.rentmanhattan.com/index.cfm?page=search&state=results>
<https://www.nestpick.com/search?city=new-york>
- - Lugar de trabajo en Manhattan (Park Avenue y 53rd St) como referencia
- Fuentes de datos, procesamiento de datos y herramientas utilizadas
- - Los datos y el mapa de Singapur se crearán con Nominatim, Foursquare y Folium mapping.
- - Los barrios de Manhattan se obtuvieron de Wikipedia y se organizaron por barrios con geodatos a través de Nominatim para su mapeo con Folium.
- - La lista de estaciones de metro se obtuvo de Wikipedia, el sitio web de NY Transit y el mapa de Google,
- - La lista de apartamentos en alquiler se consolidó a partir de sitios web inmobiliarios de MH. Los datos de geolocalización (lat, long) se obtuvieron mediante codificación algorítmica y Nominatim.
- - Folium mapa fue la base de la cartografía con diversas características para consolidar todos los datos en un mapa donde se pueden visualizar todos los detalles necesarios para hacer una selección de apartamento.

Metodologia

La estrategia para encontrar la respuesta:

La estrategia se basa en mapear los datos descritos en la sección 2.0, con el fin de facilitar la elección de al menos dos lugares candidatos para alquilar. La información se consolidará en UN MAPA donde se podrán ver los detalles del piso, el conjunto de locales del barrio y la ubicación relativa desde una estación de metro y desde el lugar de trabajo. También se proporcionará un icono de herramienta de medición. Las ventanas emergentes en los elementos del mapa mostrarán el precio del alquiler, la ubicación y el grupo de locales aplicables.

Las herramientas:

Se utilizó el web-scraping de sitios para consolidar la información del marco de datos, que se guardó como archivos csv por comodidad y para simplificar el informe. Geodata se obtuvo mediante la codificación de un programa para utilizar Nominatim para obtener la latitud y longitud de las estaciones de metro y también para cada uno de (144 unidades) los apartamentos en alquiler enumerados.

Geopy_distance y Nominatim se utilizaron para establecer las distancias relativas. Seaborn graphic se utilizó para las estadísticas generales sobre los datos de alquiler.

Los mapas con etiquetas emergentes permiten identificar rápidamente la ubicación, el precio y las características, lo que facilita enormemente la selección.

GeoData Apartamentos en alquiler en Manhattan

```
j: mh_rent=pd.read_csv('MH_rent_latlong.csv')
mh_rent.head()
```

```
j:
```

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5.0	3400	10000	40.799771	-73.966213
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	40.788585	-73.955277
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	40.799771	-73.966213
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	40.730523	-74.001873
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	40.744118	-73.995299

```
j: mh_rent.tail()
```

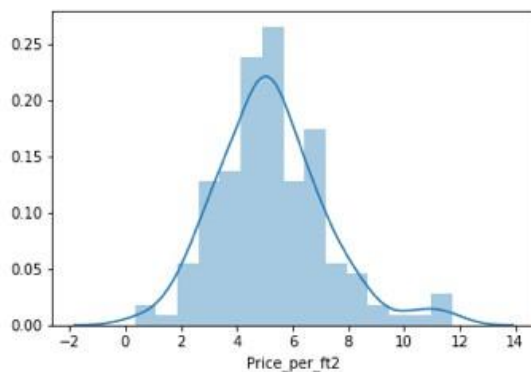
```
j:
```

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	8118	40.758216	-73.965190
142	1930 Broadway	No fee rental in Central Park West	5.06	2.0	1600	8095	40.772474	-73.981901
143	33 West 9th Street	Rental in Greenwich Village	6.67	2.0	1500	10000	40.733691	-73.997323

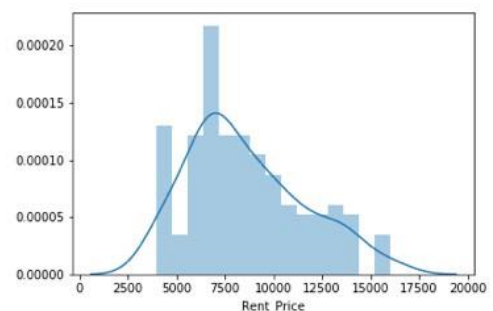
Estadísticas de precios de alquiler de apartamentos MH

Presupuesto US7000/mes es alrededor de la media

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2415fc18>
```

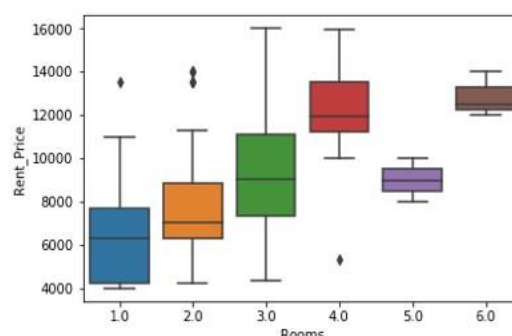


```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```

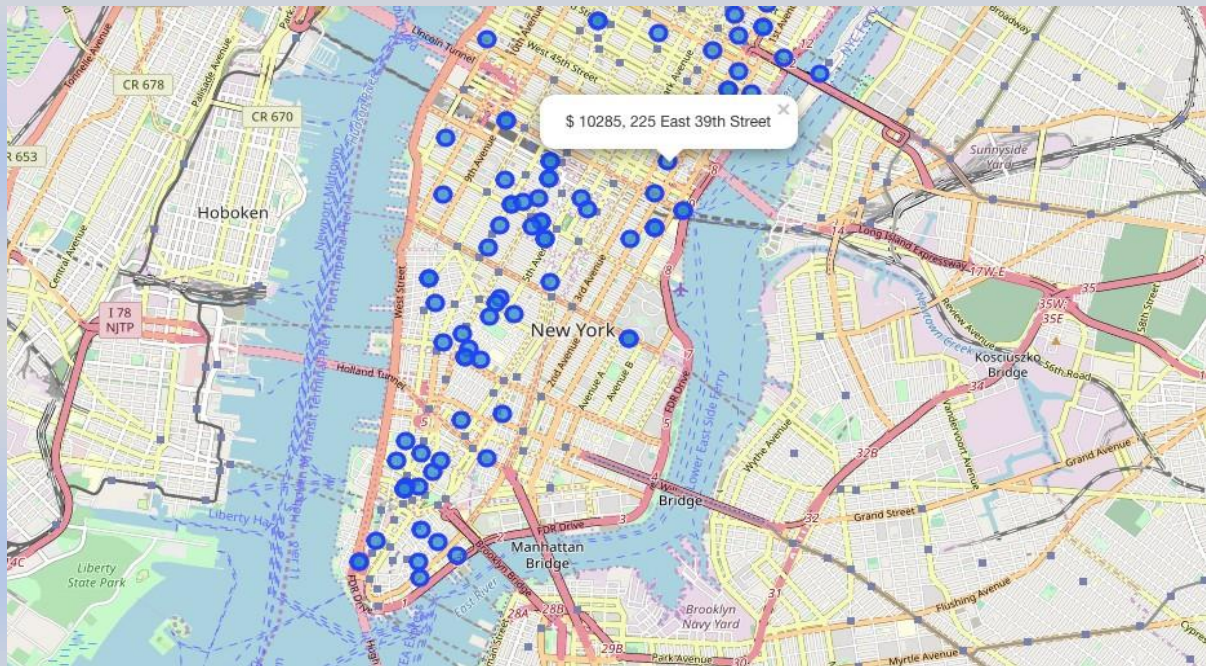


```
sns.boxplot(x='Rooms', y='Rent_Price', data=mh_rent)
```

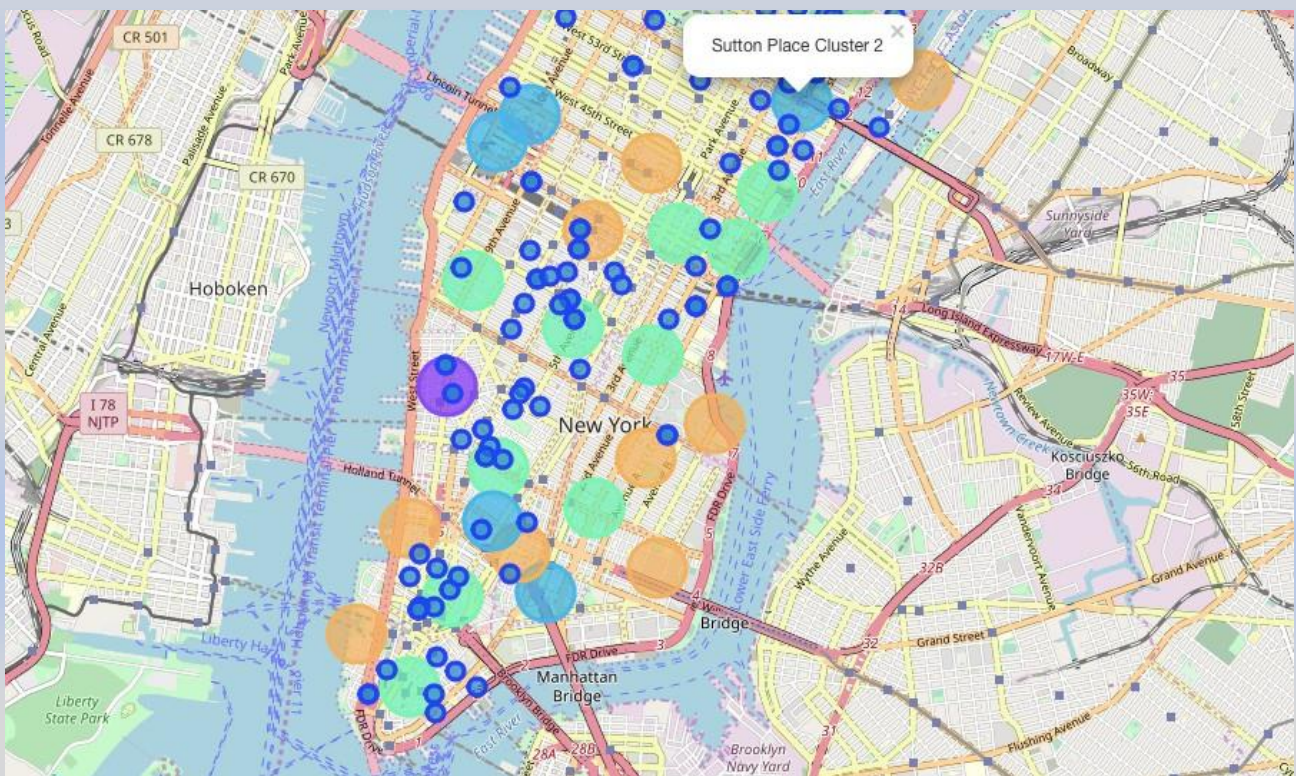
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>
```



Apartamentos en alquiler en MH



MH apartamentos en alquiler con agrupaciones de locales



Sedes del grupo 3

```
## kk is the cluster number to explore
```

```
kk = 3
```

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_m
```

	Neighborhood	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant	Pizza Place	Park	Gym / Fitness Center
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop

Geo datos de las estaciones de metro de Manhattan

click to scroll output; double click to hide

	sub_address	lat	long
0	Dyckman Street Subway Station	170 Nagle Ave, New York, NY 10034, USA	40.861857 -73.924509
1	57 Street Subway Station	New York, NY 10106, USA	40.764250 -73.954525
2	Broad St	New York, NY 10005, USA	40.730862 -73.987156
3	175 Street Station	807 W 177th St, New York, NY 10033, USA	40.847991 -73.939785
4	5 Av and 53 St	New York, NY 10022, USA	40.764250 -73.954525

```
# removing duplicate rows and creating new set mhsubl
```

```
mhsubl=mh.drop_duplicates(subset=['lat', 'long'], keep="last").reset_index(drop=True)
```

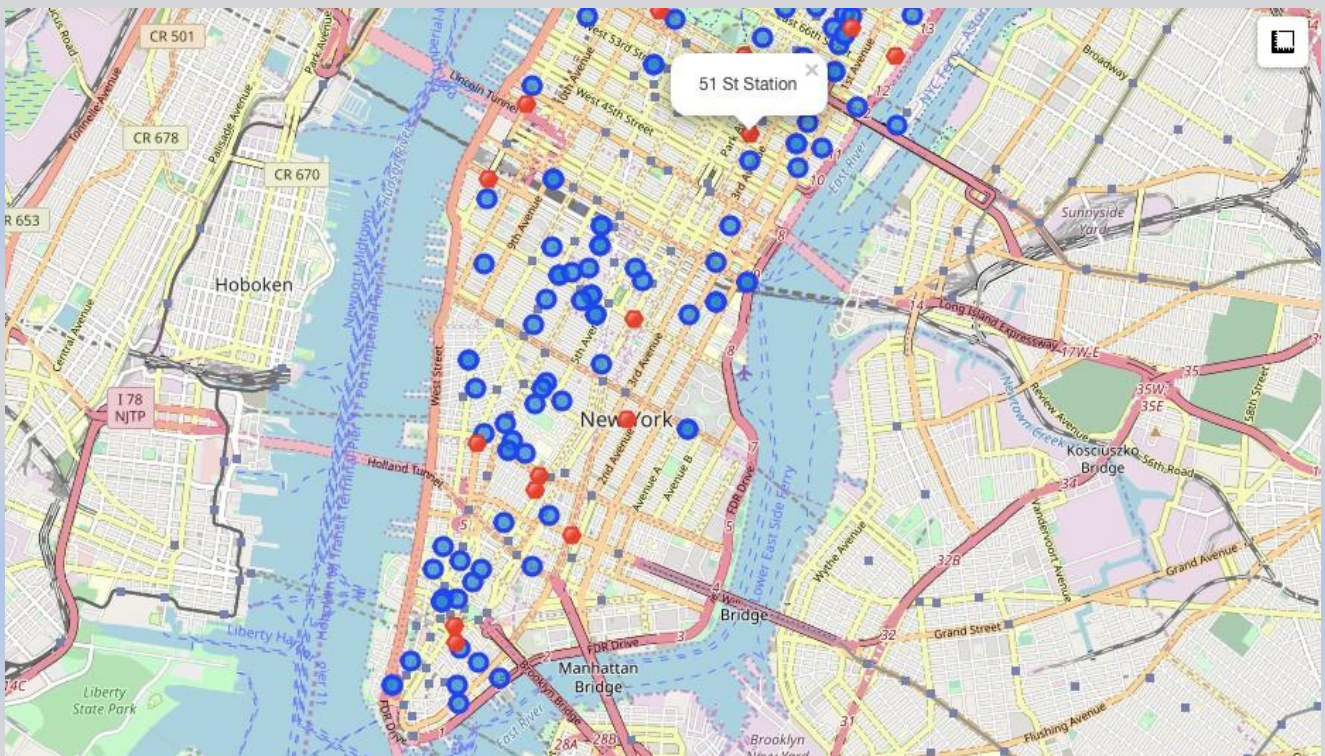
```
mhsubl.shape
```

```
(22, 4)
```

```
: mhsubl.tail()
```

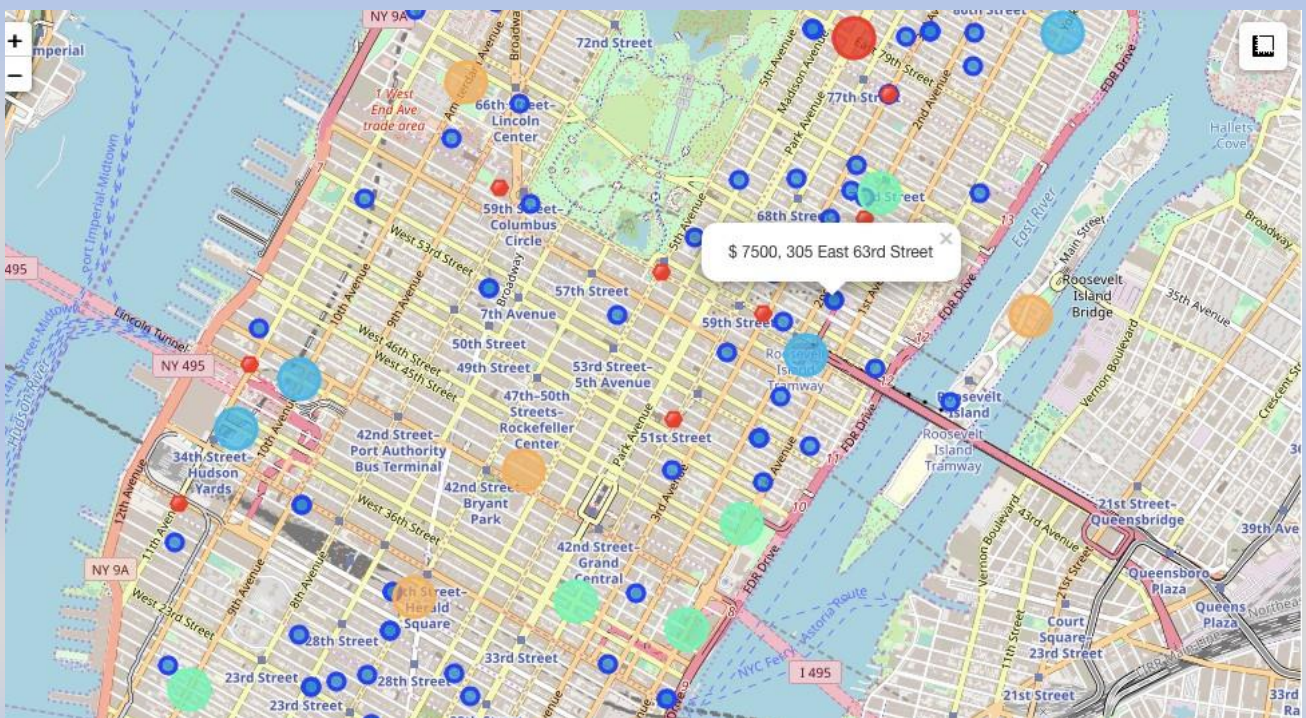
	sub_station	sub_address	lat	long
17	190 Street Subway Station	Bennett Ave, New York, NY 10040, USA	40.858113	-73.932983
18	59 St-Lexington Av Station	E 60th St, New York, NY 10065, USA	40.762259	-73.966271
19	57 Street Station	New York, NY 10019, United States	40.764250	-73.954525
20	14 Street / 8 Av	New York, NY 10014, United States	40.730862	-73.987156
21	MTA New York City	525 11th Ave, New York, NY 10018, USA	40.759809	-73.999282

Pisos en alquiler (azul) y estaciones de metro (rojo)



¡Apartamento seleccionado!

El mapa consolidado ONE muestra toda la información para la decisión:
Apartamentos dirección, precio, barrio, grupo de lugares y la estación de metro cercana.
Puntos azules=apartamentos, Puntos rojos=estación de metro, Burbujas=Conjunto de
locales



Selección de apartamentos

Utilizando el "mapa único" anterior, pude explorar todas las posibilidades ya que las ventanas emergentes proporcionan la información necesaria para tomar una buena decisión.

El precio de alquiler del apartamento 1 es de 7500 dólares, ligeramente por encima del presupuesto de 7000 dólares. El apartamento 1 está situado a 400 metros de la estación de metro de la calle 59 y el lugar de trabajo (Park Ave con la 53) está a otros 600 metros. Puedo caminar al lugar de trabajo y utilizar el metro para otros lugares alrededor. Lugares para este apt son a partir de Cluster 2 y está situado en un buen distrito en el lado este de Manhattan.

Apartamento 2 costo de alquiler es US6935, justo por debajo del presupuesto US7000. Apt 2 se encuentra a 60 metros de la estación de metro en la calle Fulton, pero voy a tener que viajar en metro todos los días para trabajar, posiblemente 40-60 minutos de viaje. Venues for this apt are as of Cluster 3.

Sobre la base de los actuales lugares de Singapur, creo que el tipo de Cluster 2 de los lugares es un parecido más cercano a mi lugar actual. Eso significa que APARTAMENTO 1 es una mejor opción ya que el alquiler mensual extra vale la pena las comodidades que ofrece.

Debate

En general, estoy positivamente impresionado con la organización general, el contenido y los trabajos de laboratorio presentados durante el Curso de Certificación IBM de Coursera

Siento que este proyecto Capstone me presentó una gran oportunidad para practicar y aplicar la Ciencia de Datos y metodologías aprendidas.

Conclusiones

- El mapeo con Folium es una técnica muy potente para consolidar la información y hacer el análisis y la decisión a fondo y con confianza. Recomendaría su uso en situaciones similares.
- Hay que estar al tanto de las nuevas herramientas de DS que siguen apareciendo para su aplicación en diversos campos empresariales.