

CAPSTONE PROJECT - ABRIR UN NUEVO RESTAURANTE

Applied Data Science Capstone by IBM/Coursera

TABLA DE CONTENIDO

- Introducción
- Datos
- Metodología
- Análisis
- Resultados
- Conclusiones

INTRODUCCIÓN

En el presente proyecto se aplicará la metodología de la Ciencia de Datos, así como las distintas herramientas para el manejo de datos, con la finalidad de resolver la pregunta: **¿Dónde establecer un nuevo restaurante?**, para que sea redituable y tenga éxito.

Cuando un emprendedor decide iniciar un nuevo negocio, uno de los principales problemas con los que se topa, es el decidir **la ubicación para establecer el negocio**, este es un factor determinante en el éxito del mismo, ya que además de la calidad del producto o servicio a ofrecer, es muy importante el lugar donde se establezca, el cual debería ser un lugar visible, con buena afluencia de personas, de fácil acceso, y que tenga poca competencia.

Por tal motivo, es de suma importancia antes de emprender un negocio identificar el lugar en donde establecerlo, y que cumpla con las características definidas anteriormente; para ello necesitaremos saber entre otros datos:

INTRODUCCIÓN

- **El tamaño de la población del lugar**, cuanto mayor sea la población el número de posibles clientes será mayor y la afluencia de gente al negocio será mejor.
- **La densidad de carreteras**, cuanto mayor sea el flujo de carreteras en el lugar elegido, se tendrá un mejor acceso al negocio, además de que se reducirán los tiempos de traslado de los clientes.
- Por último, se deberán **identificar los posibles competidores**, es decir, los restaurantes similares al nuestro por lo que se deberá elegir el lugar en donde existan los menos posibles, o no existan restaurantes en la zona.

En el presente proyecto nos enfocaremos en determinar **la mejor ubicación donde establecer un nuevo restaurante de comida tradicional en México, específicamente en la ciudad de Toluca, Estado de México**, y de esta forma, mejorar las expectativas de éxito para el nuevo restaurante.

DATOS

Para obtener la información requerida, de densidad de población y sus características como la población de hombres, población de mujeres y si son económicamente activos; utilizaremos los datos que proporciona el Instituto Nacional de Estadística Geografía e Informática (INEGI), el cual es un organismo gubernamental que recaba la información de todos los habitantes de la República Mexicana, en sus censos de población y vivienda, en específico utilizaremos las bases de datos del censo realizado en el año 2020.

Además, se utilizará el API de Foursquare, la cual es un API que proporciona la información de lugares existentes en una determinada ubicación, para obtener la información de los restaurantes existentes en el municipio de Toluca.

METODOLOGÍA

Mediante la información proporcionada por el INEGI obtendremos datos como la población total por barrio, el género, la edad, si es económicamente activo, así como nombre del barrio, su latitud y longitud; información que será de utilidad para determinar los índices de población y el índice integral para encontrar el mejor lugar o lugares para establecer el nuevo restaurante.

De esta manera, podremos determinar el número de personas en cada barrio, así como la edad, el género y si se encuentran trabajando, datos importantes para un adecuado análisis y determinar a los posibles clientes.

Por medio del API de Foursquare y las coordenadas proporcionadas por el INEGI, obtendremos la ubicación de los restaurantes ya establecidos en la zona, así como el tipo de restaurantes, sus horarios de operación y hasta los comentarios que tengan para determinar el nivel de competencia, además se podrá visualizar la afluencia de carreteras en el lugar.

Por último, para determinar el mejor barrio o barrios donde establecer el nuevo restaurante, será calculando un índice integral, por lo que para el presente proyecto utilizaremos el siguiente supuesto:

$$\text{Índice integral} = \text{índice de población} \times 0.4 - \text{índice de restaurantes (competidores)} \times 0.1.$$

RECOLECCIÓN DE DATOS

Iniciaremos con la recolección de datos de la base de datos del Censo 2020 del INEGI, solo obtendremos los datos del Estado de México para posteriormente filtrar la información de los barrios de la ciudad de Toluca.

cve_entidad	nom_entidad	cve_municipio	nom_municipio	cve_localidad	nom_localidad	Longitud	Latitud	pob_h	pob_m	pob_total	pop12_econ_act	pobf12_econ_act	pobm12_econ_act	tot_viviendas	tot_viv_hab	tam_localidad	
0	15	México	106	Toluca	1	Toluca de Lerdo	99°39'26.340" W	19°17'36.557" N	104524	119352	223876	117499	54400	63099	75817	66082	11
1	15	México	106	Toluca	43	Cacalomacán	99°42'16.517" W	19°15'11.914" N	6737	7059	13796	6692	2612	4080	3746	3232	7
2	15	México	106	Toluca	44	Calixtlahuaca	99°41'07.710" W	19°20'04.219" N	4597	4799	9396	4851	2077	2774	2589	2278	6
3	15	México	106	Toluca	46	Capultitlán	99°39'46.935" W	19°14'57.062" N	9950	10753	20703	10524	4661	5863	6771	5567	8
4	15	México	106	Toluca	49	Arroyo Vista Hermosa	99°33'03.357" W	19°20'15.353" N	2278	2443	4721	2191	862	1329	1339	1093	5

RECOLECCIÓN DE DATOS

Se filtran las columnas eliminando aquellas que no utilizaremos. Además, se limpia la información eliminando espacios en blanco al inicio y al final de cada columna y se sustituyen los valores numéricos nulos por 0.

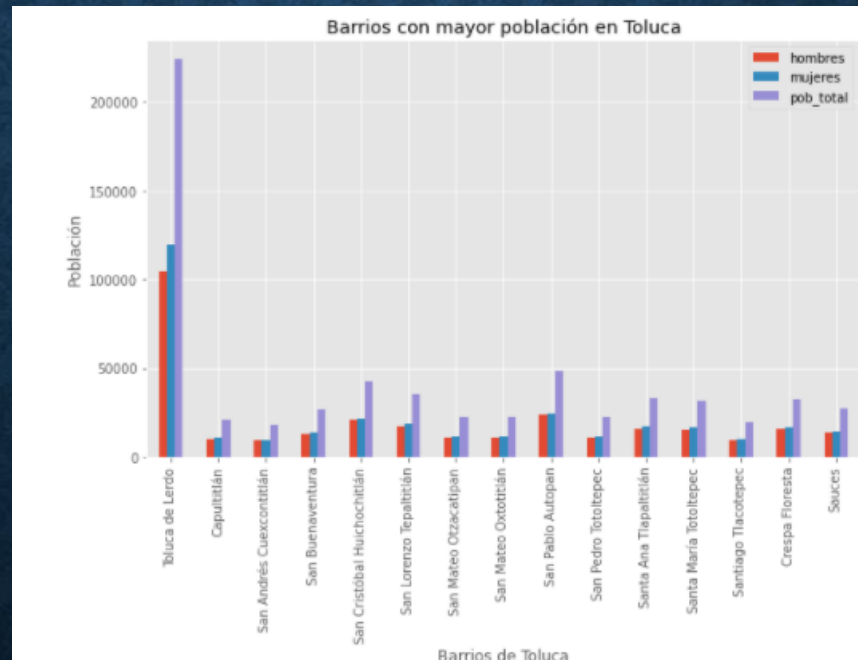
	cve_municipio	nom_municipio	cve_localidad	nom_localidad	Longitud	Latitud	hombres	mujeres	pob_total	pop12_econ_act	muj12_econ_act	hom12_econ_act	tot_viviendas	tot_viv_hab	tam_localidad
0	106	Toluca	1	Toluca de Lerdo	99°39'26.340" W	19°17'36.557" N	104524	119352	223876	117499	54400	63099	75817	66082	11
1	106	Toluca	43	Cacalomacán	99°42'16.517" W	19°15'11.914" N	6737	7059	13796	6692	2612	4080	3746	3232	7
2	106	Toluca	44	Calixtlahuaca	99°41'07.710" W	19°20'04.219" N	4597	4799	9396	4851	2077	2774	2589	2278	6
3	106	Toluca	46	Capultitlán	99°39'46.935" W	19°14'57.062" N	9950	10753	20703	10524	4661	5863	6771	5567	8
4	106	Toluca	49	Arroyo Vista Hermosa	99°33'03.357" W	19°20'15.353" N	2278	2443	4721	2191	862	1329	1339	1093	5

ANÁLISIS

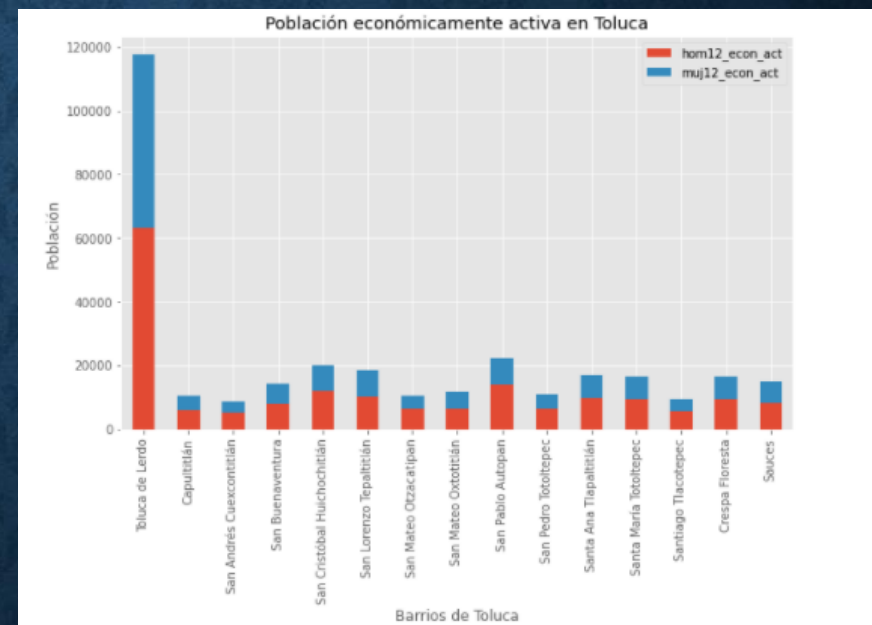
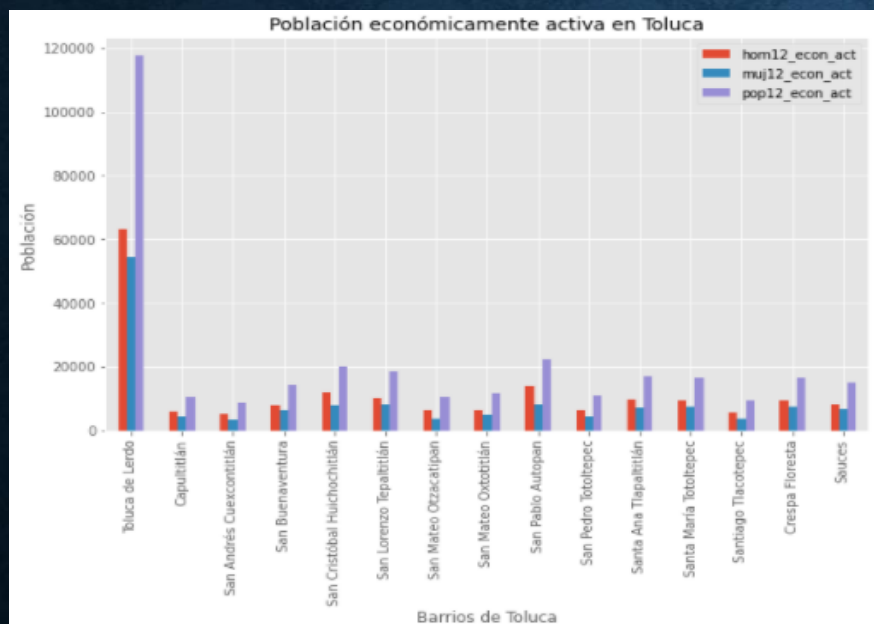
- Mediante las siguientes gráficas se muestra la distribución y características de la población de los barrios de Toluca, para el análisis se descartarán los barrios con población menor a 15,000 habitantes, y solo se tomarán en cuenta los barrios con mayor población, tomando como base el tipo de localidad, de acuerdo a la siguiente tabla:

Tamaño localidad	descripcion
1	1 a 249 habitantes
2	250 a 499 habitantes
3	500 a 999 habitantes
4	1 000 a 2 499 habitantes
5	2 500 a 4 999 habitantes
6	5 000 a 9 999 habitantes
7	10 000 a 14 999 habitantes
8	15 000 a 29 999 habitantes
9	30 000 a 49 999 habitantes
10	50 000 a 99 999 habitantes
11	100 000 a 249 999 habitantes
12	250 000 a 499 999 habitantes
13	500 000 a 999 999 habitantes
14	1 000 000 y más habitantes

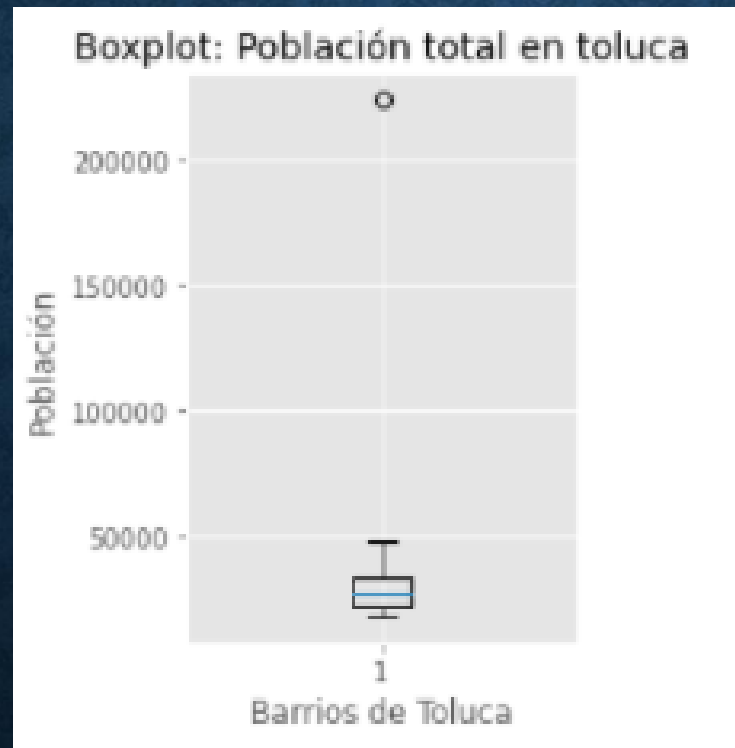
En la siguiente gráfica se observa la población total, así como la población de hombres y mujeres de cada uno de los barrios, sobresaliendo el barrio de Toluca de Lerdo como el barrio con mayor población, otra característica que destaca es que en todos los barrios la población de mujeres es mayor que la población de hombres.



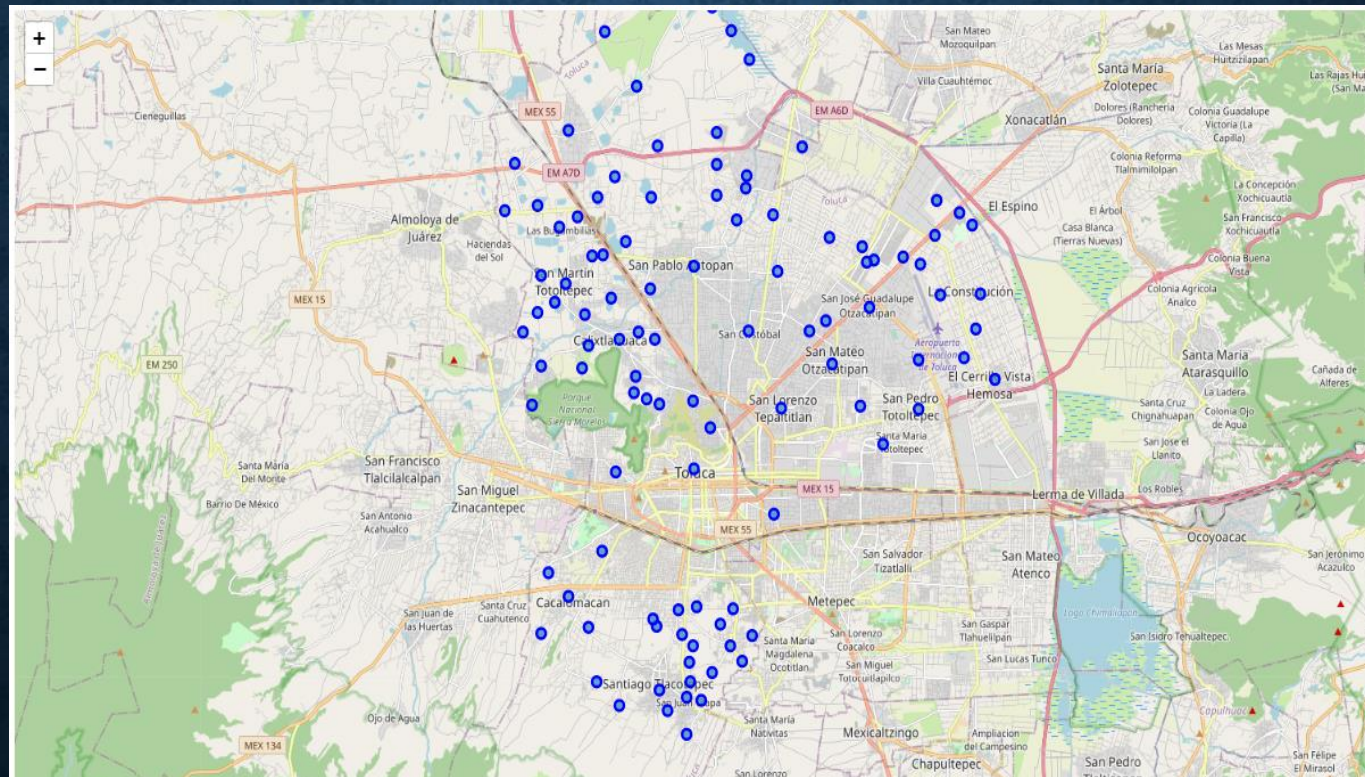
En los siguientes gráficos, se muestra la población mayor a 12 años de edad económicamente activa, y el total por género de cada uno de los barrios.



Mediante el siguiente diagrama de caja podemos observar la dispersión en los datos de población y como el barrio de Toluca de Lerdo, sobresale de los demás barrios por la considerable diferencia de población.



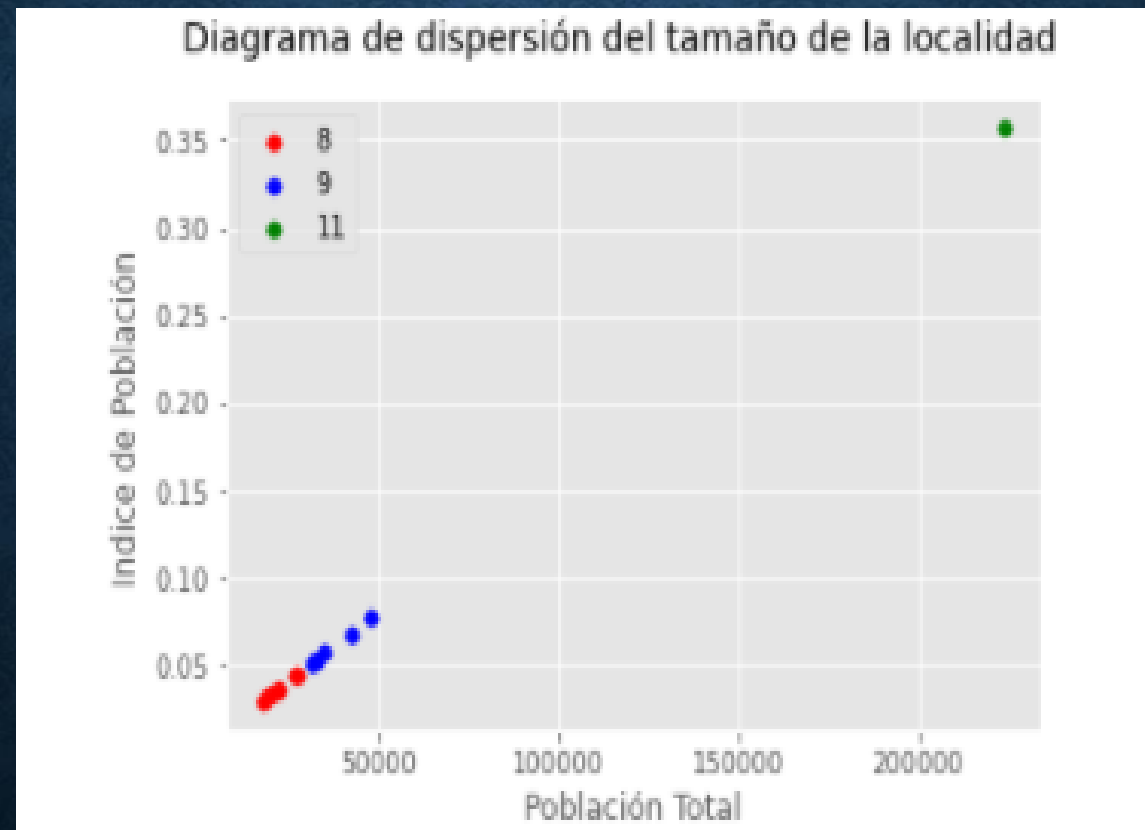
Para una mejor visualización de la distribución de barrios de la ciudad de Toluca se crea el siguiente mapa mediante la librería de Folium, y con las coordenadas geográficas de cada uno de los barrios se dibujan en el mapa de la ciudad de Toluca.



Se crea un índice de población, dividiendo la población de cada barrio entre el total de población de todos los barrios, esto para simplificar el análisis y obtener el índice integral.

	nom_localidad	latitude	longitude	pob_total	tam_localidad	indice_pob
0	Toluca de Lerdo	19.293488	-99.657317	223876	11	0.356959
1	Capultitlán	19.249184	-99.663038	20703	8	0.033010
2	San Andrés Cuexcontitlán	19.355780	-99.625639	18180	8	0.028987
3	San Buenaventura	19.267702	-99.692161	26968	8	0.042999
4	San Cristóbal Huichochitlán	19.336907	-99.636576	42320	9	0.067477
5	San Lorenzo Tepaltitlán	19.312778	-99.624347	35292	9	0.056271
6	San Mateo Otzacatipan	19.326717	-99.605106	22574	8	0.035993
7	San Mateo Oxtotitlán	19.292521	-99.686870	22500	8	0.035875
8	San Pablo Autopan	19.357432	-99.657216	47932	9	0.076425
9	San Pedro Totoltepec	19.312412	-99.572497	22374	8	0.035674
10	Santa Ana Tlapaltitlán	19.279259	-99.627182	33089	9	0.052759
11	Santa María Totoltepec	19.301545	-99.585860	31689	9	0.050526
12	Santiago Tlacotepec	19.223879	-99.670501	19744	8	0.031481
13	Crespa Floresta	19.313538	-99.594541	32307	9	0.051512
14	Sauces	19.358802	-99.592261	27628	8	0.044051

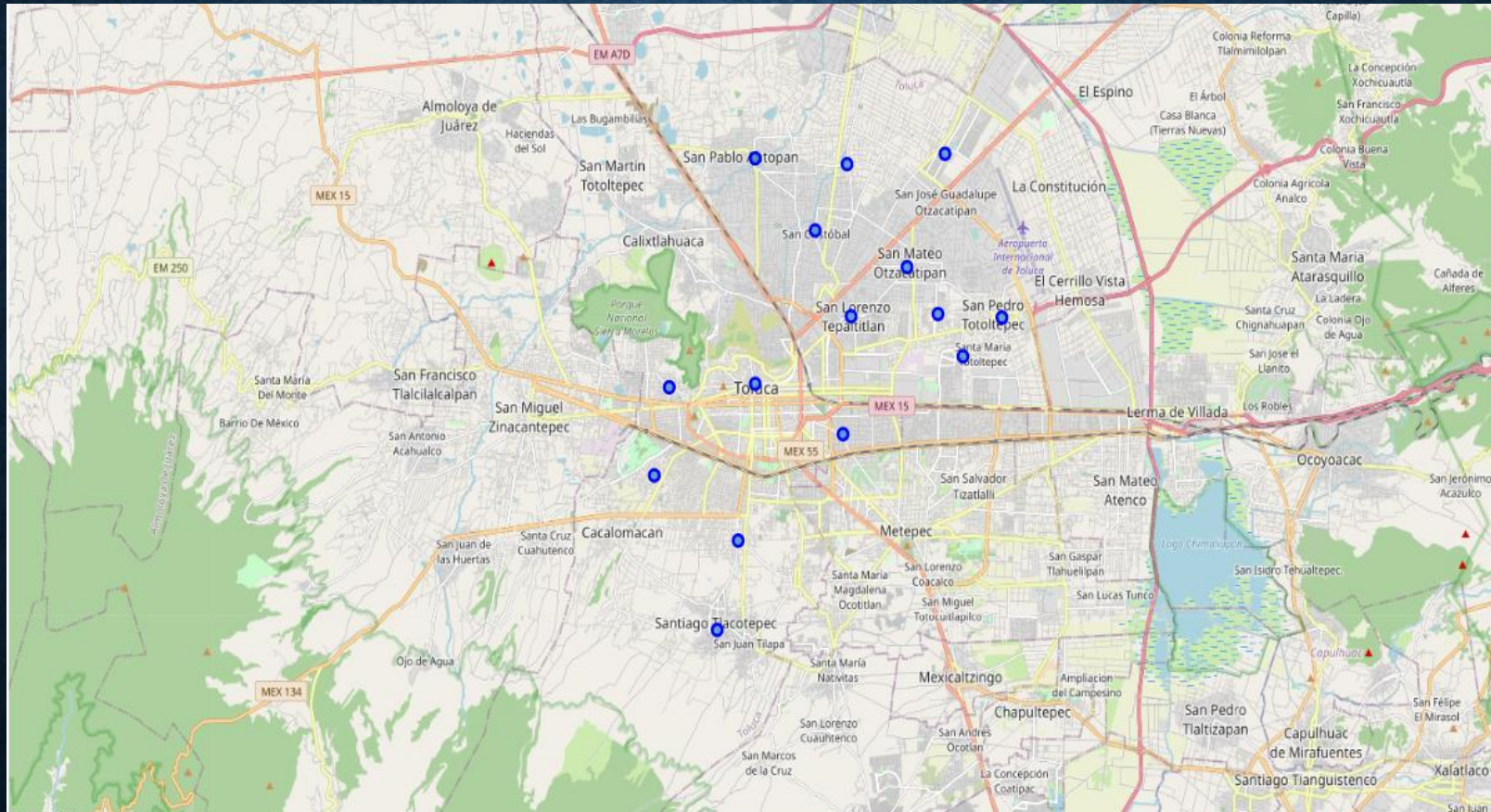
En el siguiente gráfico de dispersión se muestra la dispersión de los barrios de acuerdo al tamaño de la localidad.



Mediante el API de Foursquare obtendremos todos los lugares de comida existentes en cada uno de los 15 barrios seleccionados y los agruparemos por categorías.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	ven_latitude	ven_longitude	ven_category
0	Toluca de Lerdo	19.293488	-99.657317	Tortas El Torito Enano	19.292500	-99.656037	Fast Food Restaurant
1	Toluca de Lerdo	19.293488	-99.657317	Bistró Mecha Centro Histórico	19.290751	-99.654963	Bistro
2	Toluca de Lerdo	19.293488	-99.657317	Los Rancheros Del Sur	19.291789	-99.659255	Mexican Restaurant
3	Toluca de Lerdo	19.293488	-99.657317	Restaurante Biarritz	19.290825	-99.658276	Breakfast Spot
4	Toluca de Lerdo	19.293488	-99.657317	La Botticella	19.294278	-99.654998	Italian Restaurant
5	Toluca de Lerdo	19.293488	-99.657317	Sushi Haikú	19.291083	-99.655110	Japanese Restaurant
6	Toluca de Lerdo	19.293488	-99.657317	Tortería La Barca	19.291296	-99.657832	Mexican Restaurant
7	Toluca de Lerdo	19.293488	-99.657317	Cafe Con Leche Independencia	19.292009	-99.654728	Restaurant
8	Toluca de Lerdo	19.293488	-99.657317	Sushi Haiku	19.291070	-99.655210	Sushi Restaurant
9	Toluca de Lerdo	19.293488	-99.657317	Escamilla	19.293652	-99.653208	Seafood Restaurant
10	Toluca de Lerdo	19.293488	-99.657317	Huaraches "El Zacatecano"	19.290900	-99.659282	Mexican Restaurant

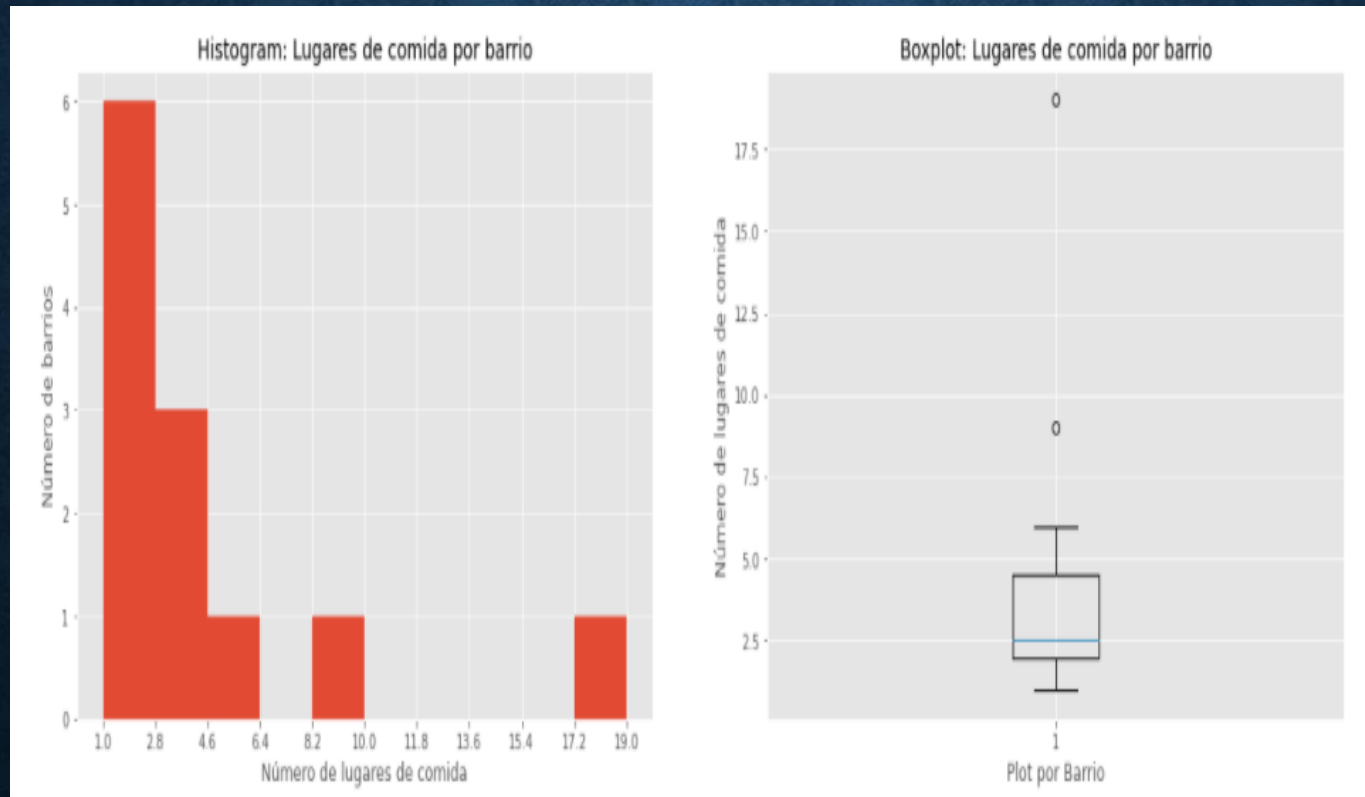
En el siguiente mapa se muestran solo los 15 barrios seleccionados para el análisis, los cuales se identifican con los círculos azules.



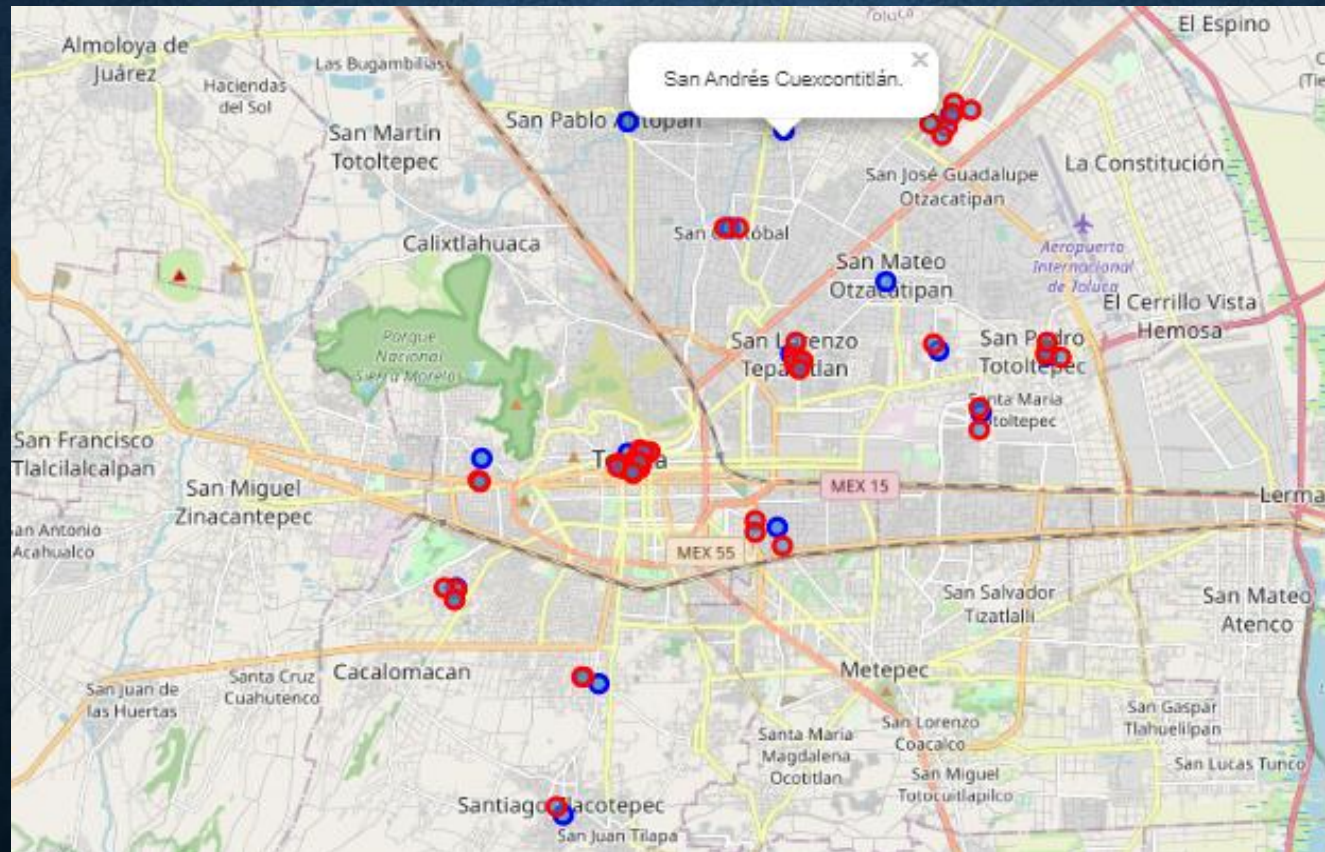
En la siguiente tabla agrupamos por barrio solo los lugares de comida, en donde destaca el barrio de Toluca con 19 lugares de comida, seguido por San Lorenzo Tepaltitlán con 9 y Sauces con 6.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	ven_latitude	ven_longitude	ven_category
Capultitlán	2	2	2	2	2	2
Crespa Floresta	1	1	1	1	1	1
San Buenaventura	4	4	4	4	4	4
San Cristóbal Huichochitlán	2	2	2	2	2	2
San Lorenzo Tepaltitlán	9	9	9	9	9	9
San Mateo Oxtotitlán	2	2	2	2	2	2
San Pedro Totoltepec	4	4	4	4	4	4
Santa Ana Tlapaltitlán	3	3	3	3	3	3
Santa María Totoltepec	2	2	2	2	2	2
Santiago Tlacotepec	1	1	1	1	1	1
Sauces	6	6	6	6	6	6
Toluca de Lerdo	19	19	19	19	19	19

- Generamos un Histograma y un boxplot para observar la distribución de los lugares de comida por barrio, en donde como ya se mencionó sobresalen de los demás barrios los barrios de Toluca de Lerdo y San Lorenzo Tepaltitlán.



Ahora agregamos al mapa de la ciudad de Toluca generado anteriormente, los lugares de comida para observar de una mejor manera su distribución, aquí nos podemos dar cuenta que existen barrios como San Pablo Autopan o San Andrés Cuexcontitlán en donde no existen restaurantes.



Posteriormente se agrega un índice que dependa del número de lugares por barrio, para calcular el índice integral de acuerdo al siguiente supuesto:

- **índice integral = índice de población * 0.4 – índice de lugares * 0.1**

Se calcula en índice y se agrega la columna al dataframe:

	nom_localidad	latitude	longitude	pob_total	tam_localidad	indice_pob	Venue	indice_venue	indice_integral
0	Toluca de Lerdo	19.293488	-99.657317	223876	11	0.356959	19	0.345455	0.108238
1	Capultitlán	19.249184	-99.663038	20703	8	0.033010	2	0.036364	0.009568
2	San Andrés Cuexcontitlán	19.355780	-99.625639	18180	8	0.028987	0	0.000000	0.011595
3	San Buenaventura	19.267702	-99.692161	26968	8	0.042999	4	0.072727	0.009927
4	San Cristóbal Huichochitlán	19.336907	-99.636576	42320	9	0.067477	2	0.036364	0.023354
5	San Lorenzo Tepaltitlán	19.312778	-99.624347	35292	9	0.056271	9	0.163636	0.006145
6	San Mateo Oztacatipan	19.326717	-99.605106	22574	8	0.035993	0	0.000000	0.014397
7	San Mateo Oxtotitlán	19.292521	-99.686870	22500	8	0.035875	2	0.036364	0.010714
8	San Pablo Autopan	19.357432	-99.657216	47932	9	0.076425	0	0.000000	0.030570
9	San Pedro Totoltepec	19.312412	-99.572497	22374	8	0.035674	4	0.072727	0.006997
10	Santa Ana Tlapaltitlán	19.279259	-99.627182	33089	9	0.052759	3	0.054545	0.015649
11	Santa María Totoltepec	19.301545	-99.585860	31689	9	0.050526	2	0.036364	0.016574
12	Santiago Tlacotepec	19.223879	-99.670501	19744	8	0.031481	1	0.018182	0.010774
13	Crespa Floresta	19.313538	-99.594541	32307	9	0.051512	1	0.018182	0.018787
14	Sauces	19.358802	-99.592261	27628	8	0.044051	6	0.109091	0.006711

Por medio de K-means se realiza la agrupación de barrios en 4 categorías, utilizando el índice integral calculado en el paso anterior, se agrega una nueva columna al dataframe con el nombre de Cluster Labels para identificar las categorías.

	nom_localidad	Cluster Labels	latitute	longitute	pob_total	tam_localidad	indice_pob	Venue	indice_venue	indice_integral
0	Toluca de Lerdo	1	19.293488	-99.657317	223876	11	0.356959	19	0.345455	0.108238
1	Capultitlán	0	19.249184	-99.663038	20703	8	0.033010	2	0.036364	0.009568
2	San Andrés Cuexcontitlán	0	19.355780	-99.625639	18180	8	0.028987	0	0.000000	0.011595
3	San Buenaventura	0	19.267702	-99.692161	26968	8	0.042999	4	0.072727	0.009927
4	San Cristóbal Huichochitlán	2	19.336907	-99.636576	42320	9	0.067477	2	0.036364	0.023354
5	San Lorenzo Tepaltitlán	0	19.312778	-99.624347	35292	9	0.056271	9	0.163636	0.006145
6	San Mateo Oztzacatipan	3	19.326717	-99.605106	22574	8	0.035993	0	0.000000	0.014397
7	San Mateo Oxtotitlán	0	19.292521	-99.686870	22500	8	0.035875	2	0.036364	0.010714
8	San Pablo Autopan	2	19.357432	-99.657216	47932	9	0.076425	0	0.000000	0.030570
9	San Pedro Totoltepec	0	19.312412	-99.572497	22374	8	0.035674	4	0.072727	0.008997
10	Santa Ana Tlapaltitlán	3	19.279259	-99.627182	33089	9	0.052759	3	0.054545	0.015649
11	Santa María Totoltepec	3	19.301545	-99.585860	31689	9	0.050526	2	0.036364	0.016574
12	Santiago Tlacotepec	0	19.223879	-99.670501	19744	8	0.031481	1	0.018182	0.010774
13	Crespa Floresta	3	19.313538	-99.594541	32307	9	0.051512	1	0.018182	0.018787
14	Sauces	0	19.358802	-99.592261	27628	8	0.044051	6	0.109091	0.006711

RESULTADOS

Como se mencionó anteriormente, para la elección del mejor lugar (barrio) para establecer un nuevo restaurante de comida mexicana, se realiza en base al barrio con el mejor índice integral, el cual se determinó con la fórmula:

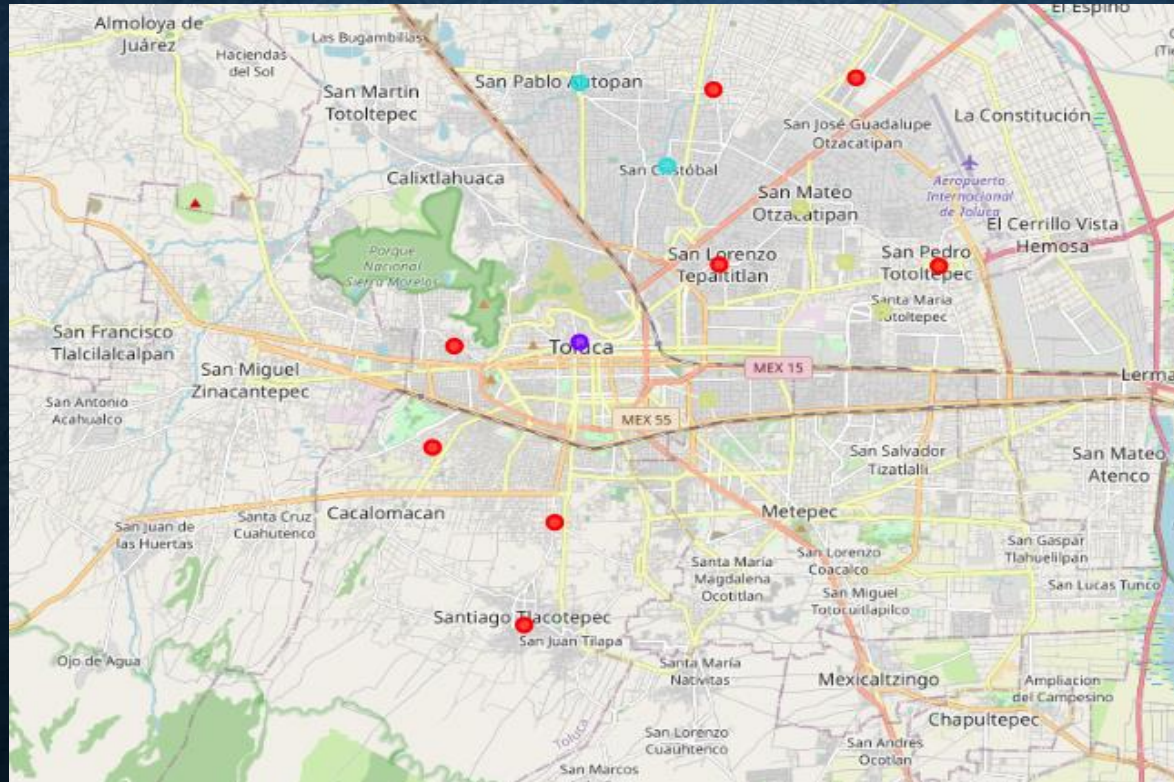
- $\text{índice_población} * 0.4 - \text{índice_lugares} * 0.1$.

De acuerdo a los resultados obtenidos, se observa que los barrios con un mejor índice integral, son aquellos con una población mayor a 15,000 habitantes, y los barrios con un menor número de lugares de comida, a excepción del barrio de Toluca de Lerdo, el cual es un barrio atípico ya que cuenta con el mayor número de habitantes (223,876) y posee 19 lugares de comida.

Para visualizar de una mejor manera los resultados, se agruparon los barrios en 4 grupos, siendo los barrios con un mejor índice integral los que se encuentran dentro de los grupos 1,2 y 3; en ese orden de importancia.

En el siguiente mapa se muestran por color las agrupaciones de barrios, siendo los mejores barrios:

- Morado: Cluster 1 (Toluca de Lerdo)
- Azul: Cluster 2 (San Pablo Autopan, San Cristóbal Huichochitlán)
- Verde Cluster 3 (San Mateo Otzacatipan, Santa Ana Tlapaltitlán, Santa María Totoltepec, Crespa Floresta)



Los mejores barrios para establecer un nuevo restaurante de comida mexicana, son los clasificados en los grupos 1 y 2 con el mejor índice integral.

Por lo que los mejores barrios para establecer un nuevo restaurante de comida mexicana son:

- 1. Toluca de Lerdo,**
- 2. San Pablo Autopan,**
- 3. San Cristóbal Huichochitlán.**

CONCLUSIONES

Con base en el desarrollo del presente proyecto Capstone para determinar el mejor lugar para establecer un nuevo restaurante de comida mexicana en la ciudad de Toluca, he aprendido a utilizar las herramientas que proporciona la Ciencia de Datos para facilitar el ciclo de vida de todo el proyecto.

Herramientas como los dataframes de pandas, las gráficas e histogramas de matplotlib o los mapas de folium, aunado a API's de terceros como Foursquare, la cual es una poderosa herramienta en proyectos donde se deba trabajar con mapas de todo tipo ya que proporciona información valiosa de los lugares de todo el mundo.

Además, por medio de la librería de pandas de Python se explotó la información del INEGI, de donde se obtuvo la población residente en los barrios de la ciudad de Toluca, el género y si es económicamente activa; así como la latitud y longitud de los barrios, información útil al momento de la creación de mapas de folium y de la obtención de lugares de comida a partir del API de Foursquare.

Por lo que, puedo concluir que hoy en día el manejo de herramientas de Ciencia de Datos es indispensable en el desarrollo de cualquier proyecto, para facilitar la recolección, limpieza análisis y visualización de los datos, así como la obtención de resultados y la publicación de los mismos.