

## Technical Challenge Spot (Data Analyst)

**Tiempo aproximado de desarrollo:** 3-5 horas.

En spot<sup>2</sup> una de las necesidades del día a día es el procesamiento de datos crudos, transformar datos en información y construir una historia con ellos que al contarla, genere insights valiosos para la organización y para la toma de decisiones. Como Data Analyst estarás expuesto a datos de todo tipo, numéricos, categóricos, geoespaciales, series de tiempo, etcétera, a los cuales se tendrá que aplicar un análisis exploratorio, medir la calidad estadística, limpiar, transformar, mezclar y generar valor de ellos.

Para esto, un cliente ha confiado en spot<sup>2</sup> como aliado para su estrategia de expansión, el interés se centra en la ciudad de México, se desea formar un mapa interactivo con información demográfica/poblacional de los ciudadanos y extraer insights significativos para el cliente. La acción a tomar es utilizar los datos públicos que el INEGI pone a nuestra disposición y cruzarlo con un objeto geoespacial que delimite un área geográfica ocupada por un conjunto de manzanas, esta demarcación se llama AGEB. (Ver más en el capítulo 1.5 del [manual del INEGI](#)).

Te compartimos un archivo formato ESRI Shapefile (df\_ageb\_urbana.shp) que contiene un objeto espacial del tipo POLYGON y las claves estándar para identificar la Entidad Federativa (CVE\_ENT), el Municipio (CVE\_MUN), la Localidad (CVE\_LOC), el AGEB (CVE\_AGEB) y un concatenado de estas claves en ese mismo orden. (CVE\_GEO). Este archivo nos es útil para cruzar cualquier dato del INEGI con una demarcación geográfica, en este caso enfocado a un AGEB.

Ejemplo:

CVE_ENT	CVE_MUN	CVE_LOC	CVE_AGEB	CVE_GEO	geometry
09	002	0001	0947	090020001047	POLYGON((-99.18322 19.50387...

Descarga los **Principales resultados por AGEB y manzana urbana** del **Censo de Población y Vivienda 2020** para la Ciudad de México. Descomprime el archivo, estudia el diccionario de datos e importa el conjunto de datos a tu IDE favorito.

En el siguiente link podrás acceder a los datos públicos del INEGI:

[https://www.inegi.org.mx/programas/ccpv/2020/?ps=microdatos#Datos\\_abiertos](https://www.inegi.org.mx/programas/ccpv/2020/?ps=microdatos#Datos_abiertos)

Lo primero que tendrás que hacer es explorar la data del censo poblacional, investigar si hay datos faltantes, tendencias, el tamaño de este data frame ( rows × columns) y cualquier otro aspecto de la data que te parezca relevante, reportalos en tu entrega.

Lo segundo será cruzar esta información con la geometría del archivo (.shp), de esta manera se espera que para cada observación del Censo de Población ligados a un AGEB se tenga una geografía que podamos utilizar en un mapa.

Lo tercero es filtrar este archivo final, como se mencionó, buscamos solo las observaciones ligadas a un AGEB y la siguiente lista de variables:

- La población total.
- La población femenina.
- La población masculina.
- La población de entre 0 y 14 años.
- La población de entre 15 y 64 años.
- La población de 65 y más años.

Una vez filtrada la información y teniendo el cruce con el polígono de cada AGEB, genera un mapa estilo Choropleth dibujando las demarcaciones del AGEB y coloreando con la densidad poblacional total de cada una de ellas.

Por último contesta las siguientes preguntas:

- 1) ¿En qué AGEB/Alcaldía se concentra el mayor número de población? ¿Por qué crees que sea?
- 2) ¿En qué AGEB/Alcaldía se presenta la mayor diferencia entre población de hombres y de mujeres?
- 3) ¿En qué alcaldía recomendarías abrir una guardería dentro de la alcaldía Cuauhtémoc?
- 4) ¿En qué alcaldía recomendarías abrir un asilo de ancianos?
- 5) Realiza un análisis estadístico de las variables, reporta las medidas de tendencia central y de dispersión.
- 6) ¿Qué puedes inferir de los AGEBS con muy poca población?
- 7) ¿Propondrías alguna limpieza o transformación para que la visualización del mapa sea más limpia? ¿Sí, No, por qué?
- 8) Despliega gráficas que ayuden a entender la dinámica poblacional en la Ciudad de México, puede ser a nivel alcaldía y/o AGEB, demuestra tus habilidades para mostrar datos.
- 9) Del catálogo de variables, qué estudio te parecería que generaría mucho valor para la expansión del cliente con un bajo costo técnico.
- 10) ¿Qué otro análisis te parecería interesante desarrollar a futuro con estos datos?

BONUS:

En el mapa Choropleth agrega la ubicación de todas las estaciones del metro y pinta los puntos de naranja.

Los datos de las estaciones los podrás encontrar en:

<https://datos.cdmx.gob.mx/dataset/lineas-y-estaciones-del-metro>

## Requerimientos:

- Tech Stack sugerido: SageMaker, Jupyter, PyCharm, Colab, etc. Usar otras herramientas también es posible, sin embargo en la entrevista debes justificar su uso.

## Resultados esperados:

- Reporte general de los hallazgos, piensa que este se presentará a miembros del equipo sin un background técnico.
- Visualizaciones, mapas y gráficas que ayuden a entender los datos.

## Entregables:

- Entrega código y figuras a través de tu repositorio de GitHub.
  - El README.md debe ser muy claro para poder replicar tu infraestructura.
- Muestra las visualizaciones que creas más interesantes. (Mapas, estadísticas, tablas, cuadros, matrices, diagramas, etc.)
- Presentación con el equipo de data presentando tus hallazgos.

## Consejos:

- La biblioteca de **Plotly**, junto con algunos módulos como **Express**, son muy buenos para desplegar mapas y gráficas interactivas.
- Para la lectura de archivos con información geoespacial se recomienda el uso de **Geopandas**.
- Si decides el mejor estilo de mapas a utilizar, pero si decides usar Plotly, recuerda que este soporta MapBox, si deseas utilizar un estilo que requiera un Mapbox Access token, en el sitio oficial podrás solicitarlo de manera gratuita. ( <https://www.mapbox.com/> )
- No limites tus respuestas a lo solicitado, queremos verte propositivo y encontrando insights que tal vez no sean triviales.
- No hay respuestas incorrectas, queremos entender tu proceso creativo.
- Piensa fuera de la caja.
- ¿Por qué deberíamos adoptar tu solución o qué mejorarías?
- ¿Este sería un código que pondrías en producción?

¿Dudas? Envíame correo a [brian@spot2.mx](mailto:brian@spot2.mx)