

**Reporte Final**

**O21-LDS1021-1 MINERÍA DE DATOS**

**Profesor: Juan González Calleros**

**22 de noviembre de 2021**

**Integrantes del equipo:**

Ma. Paula Ibáñez

ID: 159891

Tao Izzo Elvira

ID:165254

Sabrina Savino

ID: 430551

Miguel Ángel Ibarra Juárez

ID: 160162

Jorge Roberto Polanco Domínguez

ID: 167325

Reporte Final

Tabla de contenido

[Resumen 2](#_Toc88457479)

[Sección 1: Introducción 3](#_Toc88457480)

[Selección de la base de datos 3](#_Toc88457481)

[Preámbulos 3](#_Toc88457482)

[Sección 2: Investigación relevante 4](#_Toc88457483)

[Planteamiento del problema 4](#_Toc88457484)

[Propósito de la investigación 4](#_Toc88457485)

[Estado de la cuestión 4](#_Toc88457486)

[Sección 3: Metodología 5](#_Toc88457487)

[Exploración de datos: 6](#_Toc88457488)

[Transformación y limpieza: 6](#_Toc88457489)

[Sección 4: Resultados 7](#_Toc88457490)

[Sección 5: Discusión 14](#_Toc88457491)

[Sección 6: Conclusión 14](#_Toc88457492)

[Referencias 15](#_Toc88457493)

# Resumen

En este documento se muestra el análisis de una base de datos conformada por los nombres y las coordenadas de las escuelas en la Ciudad de México. Los datos incluyen a distintos grados escolares, entre ellos kínder, primaria, secundaria e incluso escuelas especializadas en lactancia o bien jóvenes con cierto grado de discapacidad. El fin del trabajo es facilitar la decisión sobre que escuela elegir a través de la ubicación de cada una.

# Sección 1: Introducción

## Selección de la base de datos

Se seleccionó esta base con el fin de ayudar a los padres de familia a tomar una decisión basada en los datos de donde se encuentran ubicadas las escuelas al grado que deseen para tomar en cuenta la ubicación dependida del trayecto o cercanía a sus puntos habituales.

## Preámbulos

La escuela es el objetivo principal de la educación; nos da habilidades para convivir en familia y comunidad. La escuela se encarga de formar y preparar a sus estudiantes para desenvolverse y aportar a la sociedad que los rodee. Es muy importante saber seleccionar la escuela que mejor se adapte a nuestras necesidades, pues en esta se desarrollan habilidades sociales que permiten una mejor comunicación y comprensión. Muchas veces en estos lugares se definen los proyectos de vida y aspiraciones de cada individuo, se forjan las herramientas para afrontar las diversas situaciones que nos podemos llegar a encontrar a lo largo de nuestra vida.

Uno de los factores que afectan a la educación de los jóvenes es la ubicación e infraestructura de su escuela, está comprobado que tener escuelas en buen estado es determinante para lograr que los alumnos obtengan los resultados académicos esperados. Según el BANCO DE DESARROLLO DE AMÉRICA LATINA (2021), varios estudios han encontrado que las condiciones físicas de los edificios escolares afectan positivamente las tasas de finalización, culminación del ciclo y el incremento de matrícula.

Buscar, conocer y comparar distintas escuelas ayuda a tomar una decisión más acertada sobre el futuro de los individuos que asistan a dichas escuelas. Es importante tomar en cuenta la ubicación en la que se encuentran las instituciones para establecer una ruta que optimice los tiempos de traslado, costos y recursos necesarios para lograr terminar los cursos de manera satisfactoria.

# Sección 2: Investigación relevante

## Planteamiento del problema

El principal problema radica en que la Ciudad de México es una de las principales ciudades en México, siendo la más grande y poblada. Como consecuencia, muchos habitantes del país deciden trasladarse a la CDMX junto con sus familias en búsqueda de mejores oportunidades laborales. Al ser una metrópoli de gran tamaño muchas veces los padres o tutores inscriben a sus hijos en la escuela más cercana y no en la que es óptima para sus necesidades (escolaridad, presupuesto, especificaciones, etc.).

## Propósito de la investigación

Se planea desarrollar una página/dashboard para apoyar a la toma de decisiones de los padres de familia de manera informada y lógica. Con el mapa desarrollado y los datos de las bases, se muestran las escuelas tanto públicas como privadas cercanas al área seleccionada.

## Estado de la cuestión

Comúnmente la población conoce las principales escuelas o grados (preescolar, primaria, etc.), pero los grados escolares van más allá de estos, por eso decidimos incluir en nuestro mapa escuelas especializadas en lactancia y Centros de Atención Múltiple (CAM). Un CAM de acuerdo con el Gobierno del Estado de México “es un servicio escolarizado que atiende a alumnas(os) desde los 45 días de nacidos hasta los 23 años con discapacidad, con requerimientos de adecuaciones curriculares altamente significativas, así como de apoyos generalizados y permanentes, que las Escuelas de Educación Regular no pueden proporcionar”.

Actualmente existen diferentes plataformas en internet que brindan información como nuestro estudio. Estas plataformas se encargan de proporcionar al usuario listas de escuela dependiendo de las características que el usuario busca, como lo son; ubicación, idioma, nivel educativo, turno, nivel en la que ese encuentra rankeada la escuela.

Algunos ejemplos de estas plataformas son:

-Mejora tu escuela.org

-La mejor escuela para ti

-Los mejores colegios.mx

También existen escuelas e instituciones que explican cómo hacer análisis como el nuestro para poder elegir la mejor opción de escuela. Como lo hizo el instituto Justo Sierra, el cual elaboro una página web donde muestra los aspectos a considerar para tomar la decisión más apta para elección de escuela de nivel básico, ellos indican que hay que considerar seis aspectos importantes, los cuales son: ¿Cuál es la metodología de estudio del colegio?, ¿Se encuentran actualizados con herramientas y tecnologías?, ¿Cuentan con un segundo idioma?, ¿Cuáles son los valores de la escuela?, ¿Cuentan con actividades extracurriculares? y ¿La ubicación es cercana?, las cuales se parecen a las variables que trabajaremos, en este proyecto.

La revista digital expansión, realizo un reporte sobre los seis puntos básicos para elegir escuela, en el cual explican que uno de los aspectos a considerar es la distancia, la cual según si explicación es importante para no perder una gran cantidad de tiempo en el traslado y ese tiempo se pueda usar en otras actividades.

# Sección 3: Metodología

Para empezar a realizar este trabajo se tenía que seleccionar una base de datos para trabajar, por lo que se procedió a realizar una búsqueda de una base que nos permitiera realizar los diferentes tipos de análisis vistos en clase, con suficientes variables y datos. Para esto se usaron varias herramientas de internet como: portal de datos abiertos de la CDMX, portal de datos abiertos de Puebla, Kaggle, Google public data sets, UCI machine learning repository, Quandl, etc. Durante la investigación se seleccionó dos bases de datos del portal de datos abiertos de la CDMX, para tener un proyecto con información más completa se añadió una base de población de la CDMX, la cual se tuvo que hacer a mano, obteniendo los datos del INEGI (Instituto Nacional de Estadística y Geografía).

Las bases seleccionadas fueron “Escuelas públicas” con 2384 registros y “Escuelas privadas” con 3697 registros, las cuales contienen ocho y doce variables respectivamente. Las variables de la primera base de datos son: id, nombre, domicilio con nombre, coordenadas, longitud, geopoint y domicilio (ocho variables). Las variables de la segunda base son: id, nivel, nombre, turno, domicilio, colonia, alcaldía, sostenimiento, ubicación, lat, lon y coordenadas (doce variables).

Debido a las diferencias en las variables de las bases, se realizó el análisis de manera independiente, obteniendo más insights de las escuelas privadas que publicas.

Los procesos que se realizaron en ambas bases fueron:

Exploración de datos: se hizo un análisis exploratorio de datos para saber que variables eran de utilidad y cuales no, debido a que varias variables se repetían en otras con ligeras modificaciones se depuraron estas, pero antes de llevar a cabo esto se recataron datos de importancia o para equilibrar los datos.

Transformación y limpieza: En la base de escuelas públicas a falta de la variable “Alcaldía”, la información se logró extraer de la variable “domicilio con nombre”. Al hacer esta extracción se presentaron duplicados por la diferencia de un acento. Si bien todavía se encontraban datos en alcaldía, se hizo una segunda extracción de estas de forma diferente, o bien analizando los pocos datos que hacían falta, para no depurarlos.

De igual manera que se hizo lo anterior con las alcaldías, se hizo el mismo proceso con el grado/nivel de las escuelas, el cual se encontraba en la variable “nombre” junto con el nombre de la escuela o bien era parte del nombre oficial. Hubo datos que en este primer proceso de transformación se tuvieron que depurar a falta del grado en el nombre y otros en los cuales se tuvo que analizar por separado esos datos y hacer la corrección de ellos.

La última transformación que se realizó en esta base fue con los puntos geográficos, es decir la longitud, latitud y geopoint, debido a que en latitud y longitud había datos faltantes, se tenían que eliminar estos datos con toda la información que contenían, pero hubo ocasiones en donde estos datos se encontraban vacíos pero el geopoint los contenía. Lo que se hizo fue completar estos datos con esa información dándonos de esta forma la misma cantidad de datos faltantes en las tres variables

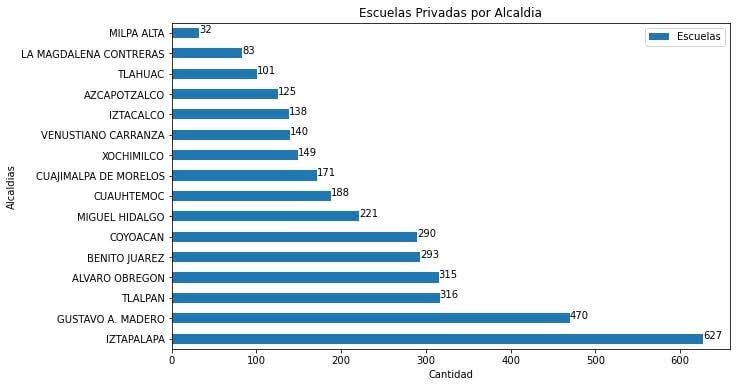
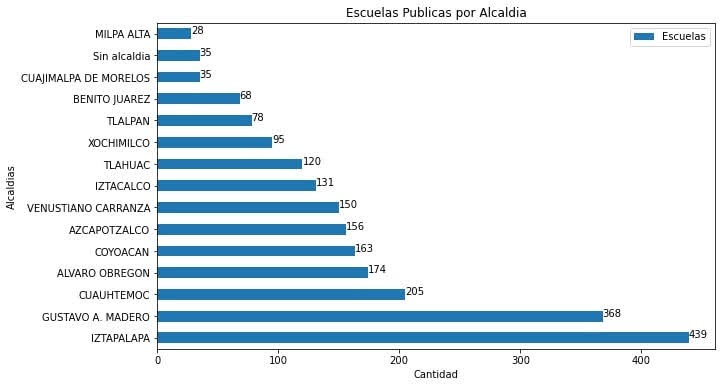
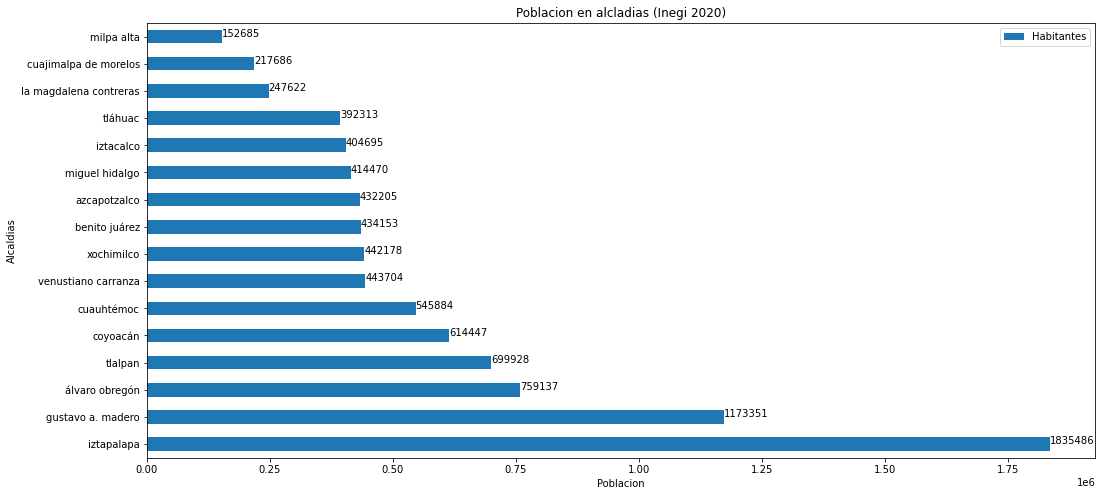
La base de escuelas públicas quedo teniendo las siguientes variables con los datos completos en todas sus columnas: nombre, domicilio con nombre, latitud, longitud, alcaldía y grado. Con respecto a la base de escuelas privadas, que ya contenían las variables que se crearon en la base de escuelas públicas el proceso de limpieza fue más rápido de depuraron los datos faltantes correspondiente a lon, lat y coordenadas, debido a que no contenían la información.

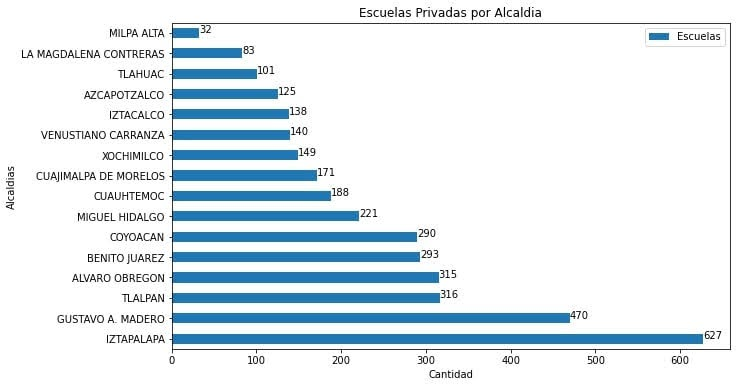
La base de escuelas privadas quedo teniendo las siguientes variables con los datos completos en todas sus columnas: nombre, ubicación, latitud, longitud, alcaldía, nivel, colonia y turno

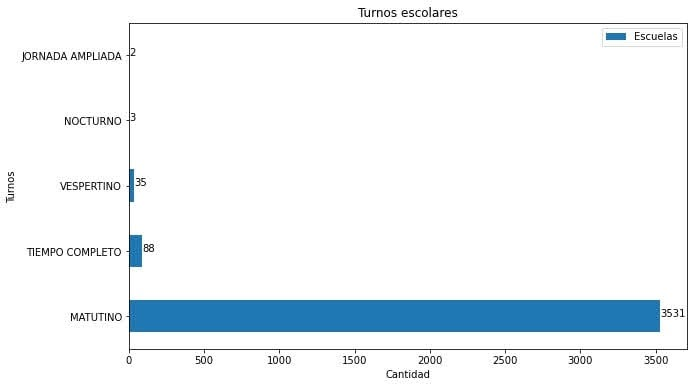
Una vez que las dos bases quedaron limpias y con las variables adecuadas se procedió a sacar estos insights de cada una, así como graficar las escuelas por separado en un mapa interactivo, donde además de graficar las escuelas se hizo una clusterizacion para agrupar las escuelas por la cercanía y mostrar estas según el usuario seleccione el grado/nivel que desea visualizar.

La base de Población se usó para ver qué tipo de relación había entre las escuelas públicas/privadas por alcaldía y la población en esa alcaldía. Por lo anterior se tuvo que llevar a cabo un cruce de datos para tener todo en una sola tabla y a partir de ella ver estas relaciones.

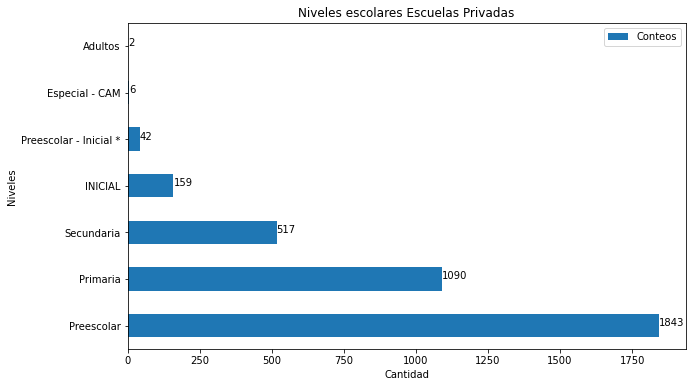
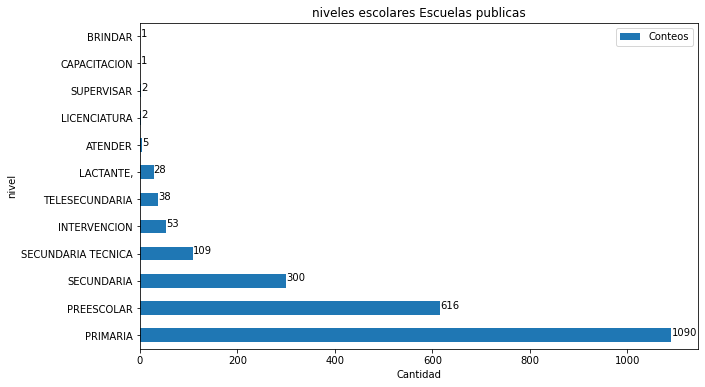
# Sección 4: Resultados

1.- Categorizamos las escuelas públicas y privadas por Alcaldía, y comparando la cantidad de escuelas con la cantidad de habitantes por alcaldía como uno esperaría Iztapalapa y Gustavo A. Madero que son las alcaldías con mayor población también tienen el mayor número de escuelas tanto públicas como privadas. Las Alcaldías con menor población, Milpa Alta y Cuajimalpa Morelos, tienen la menor cantidad de escuelas públicas, Milpa Alta también cuenta con la menor cantidad de escuelas privadas, sin embargo, Cuajimalpa de Morelos siendo la segunda alcaldía con menor población está en la posición número nueve en cuanto a escuelas privadas, lo que nos lleva a entender que tienen un mayor índice de estudiantes en escuelas privadas que en escuelas públicas. 

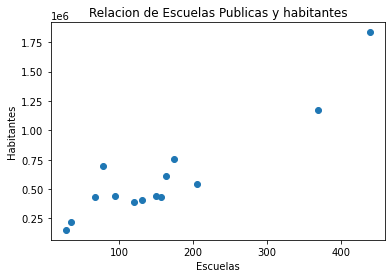
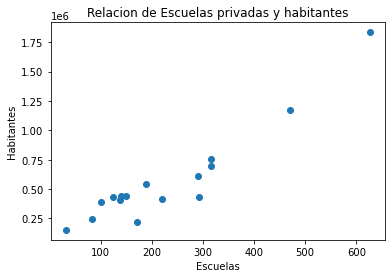
2.- Nuestra base de datos contaba con la columna de “turno”, por lo que podíamos filtrar por categoría para obtener el número de escuelas que laboraban por horario. Nuestra búsqueda concluyo que la gran mayoría de las escuelas laboran en el periodo matutino, un 2.4% de las escuelas laboran en horario de tiempo completo la cual consta en jornadas de 8 horas en lugar de las 4 horas y media de una jornada regular, un mínimo 0.095 en horario vespertino, solamente 3 en horario Nocturno y la menor cantidad de escuelas por turno es por jornada ampliada con solo 2 escuelas laborando bajo ese horario. 



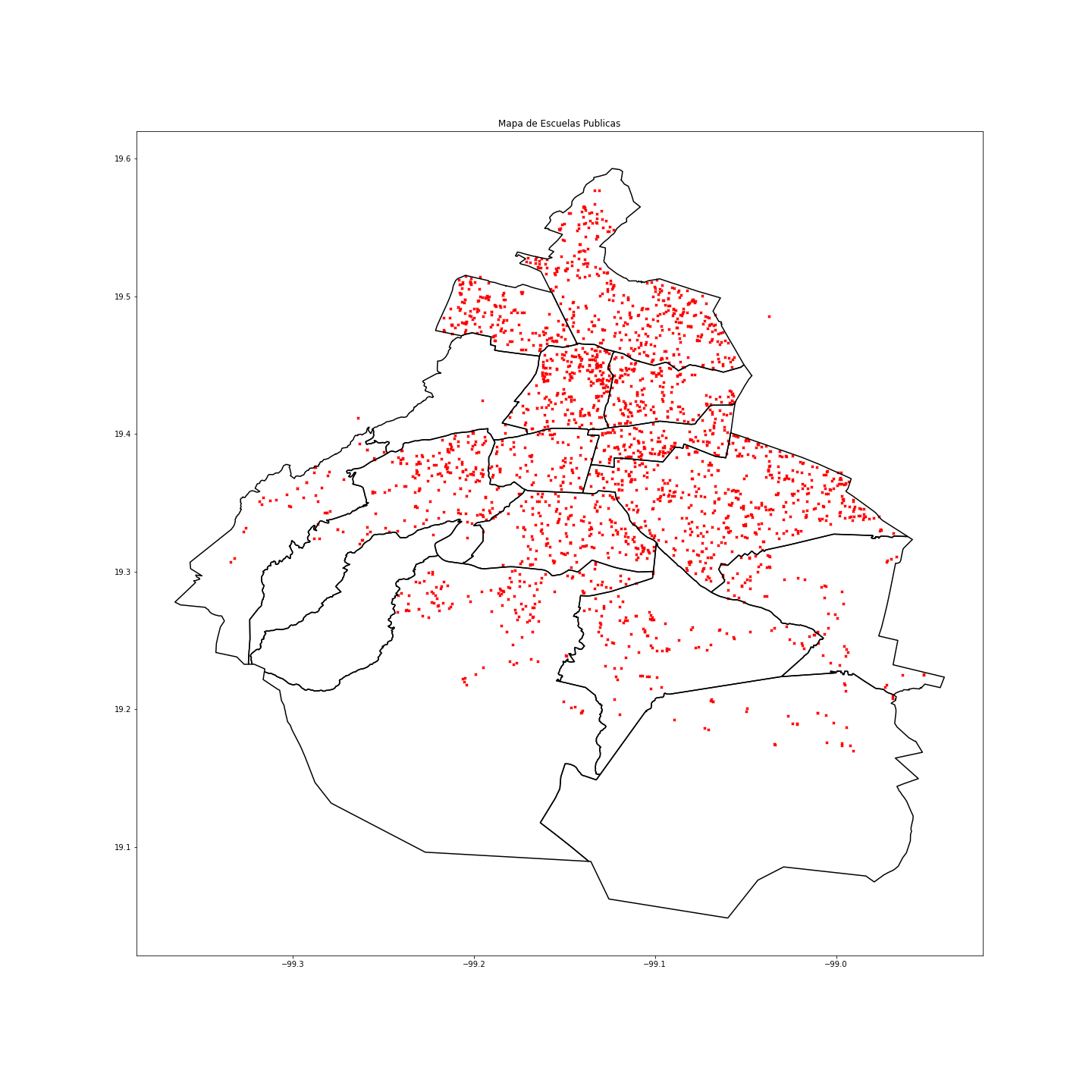
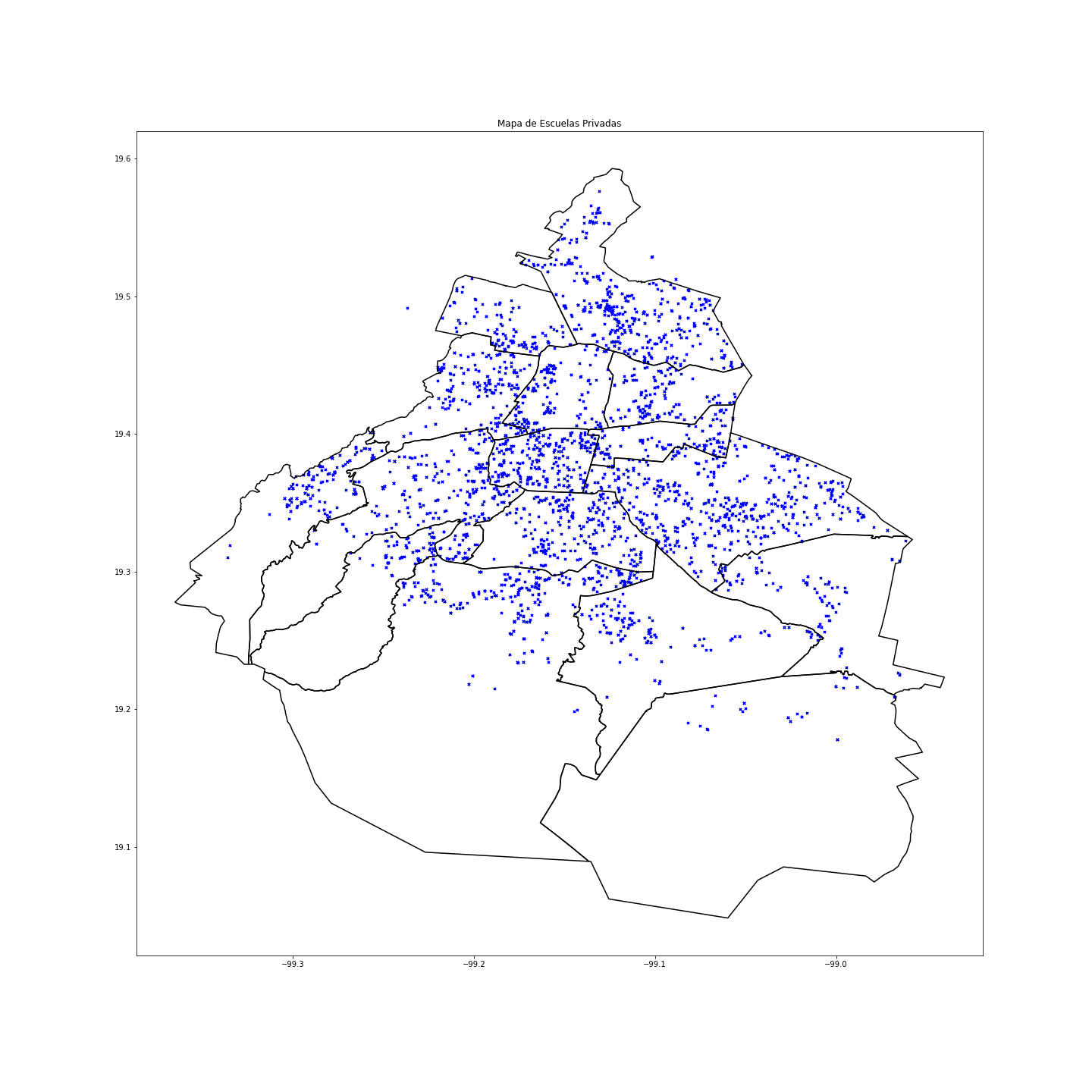
3.- En esta sección desglosamos los diferentes niveles de escolaridad que se presentan en instituciones tanto públicas como privadas, se puede apreciar que en ambos públicos y privados predominan los Preescolares, seguidos de primaria y después secundaria, las categorías subsecuentes tienen un porcentaje poco importante dividido entre ellas, pero si notamos la tendencia es que entre mayor grado de escolaridad haya, menor cantidad de instituciones hay, por lo que se puede inferir una de dos cosas que la atención que se le da a cada individuo es menor con mayor escolaridad o que menor cantidad de estudiantes están continuando con sus estudios al terminar un grado.

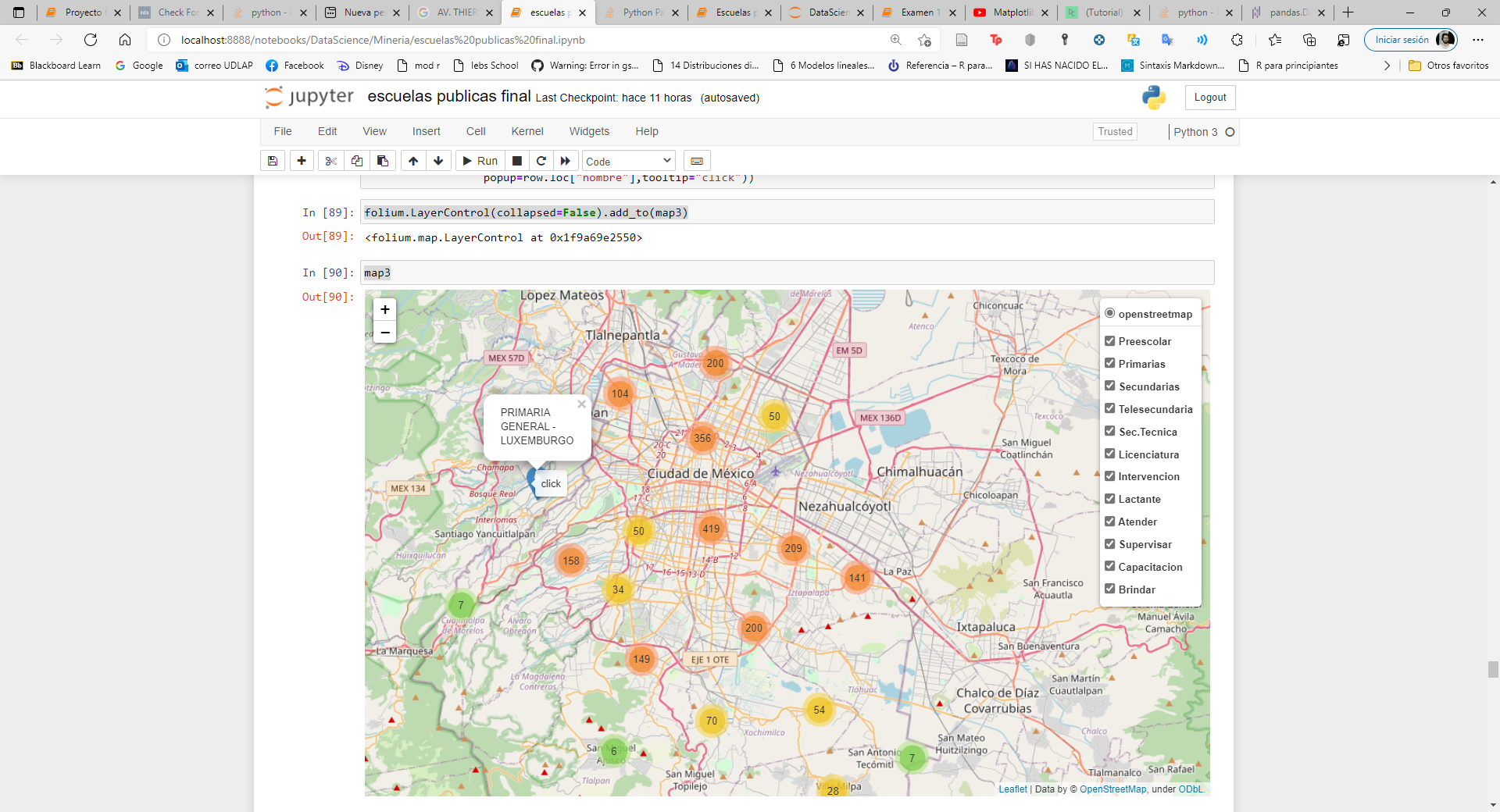


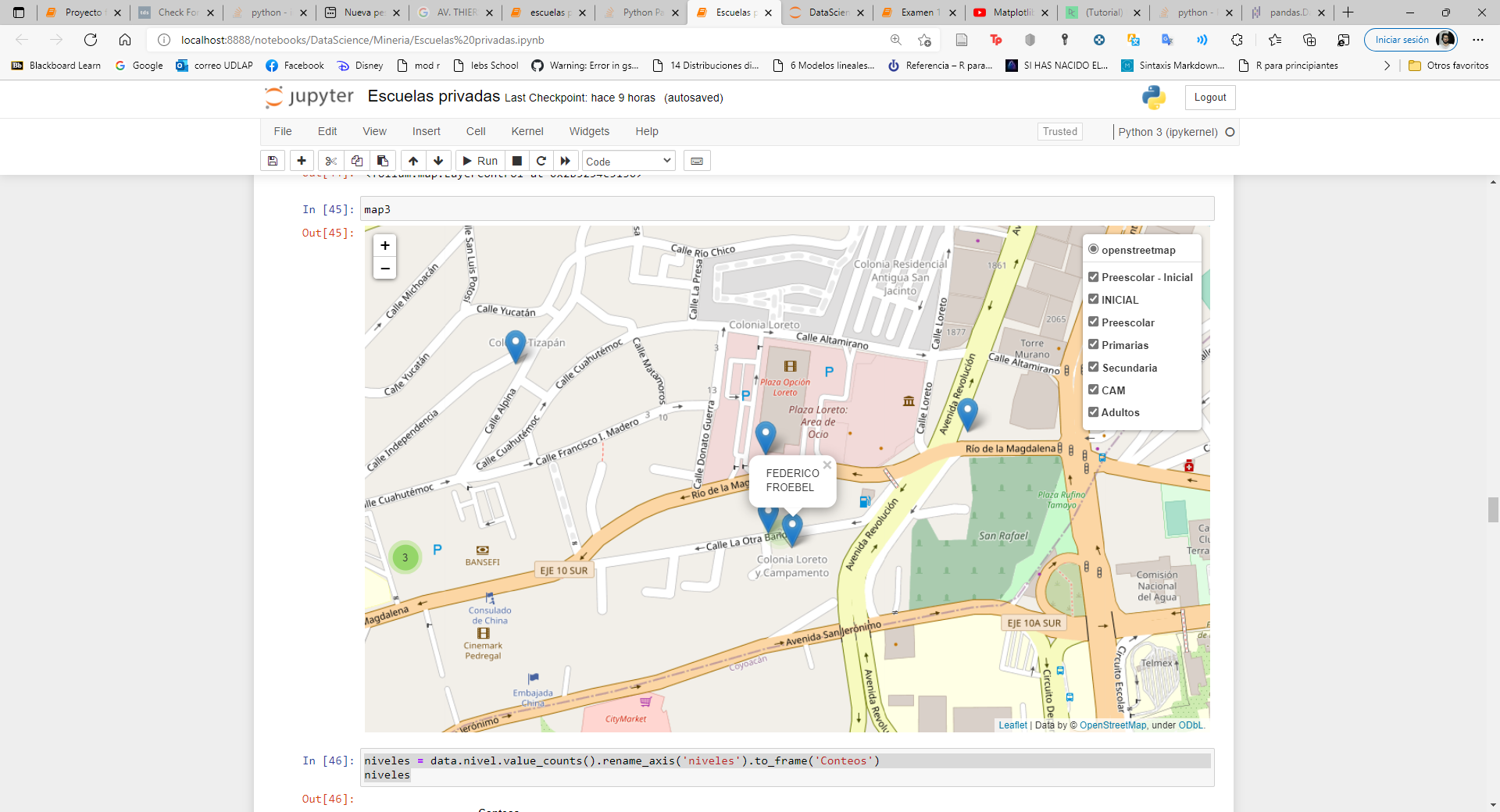
4.- Aquí está la relación entre habitantes de la alcaldía y numero de escuelas, Podemos notar que la gráfica de las escuelas privadas que arroja los datos tiene una similitud a la recta identidad Y=X, por lo que se puede decir que la cantidad de escuelas privadas tiene una correlación alta de 93.82 con el número de habitantes, en el caso de las escuelas públicas sigue teniendo una similitud a la recta identidad con una correlación de 91.93, pero tiene unos valores aislados como sería el caso de Cuajimalpa que consta de la segunda población más pequeña, pero tiene ocupa la novena posición en cuanto a escuelas privadas, además de otros.



5.- Se crearon varios mapas, unos para ver las distribuciones en las alcaldías y otros mapas interactivos para poder visualizar las escuelas por su nivel, las cuales aparecen juntas según su cercanía, permitiendo saber cuántas escuelas hay en cada zona. Si el usuario selecciona algún punto se le mostrará el nombre de la escuela.







6.- Comparamos el ambiente en cuestión de seguridad fuera de la mejor y peor escuela según su calificación de la prueba PLANEA, efectuada en 2018, la que peor calificación tuvo fue el Colegio Humanista de las Américas una escuela privada ubicada en la alcaldía de Iztapalapa, un factor ajeno a la escuela, puede ser la inseguridad ya que según datos de INEGI 74% de la población de Iztapalapa ha reportado haber tenido un conflicto o enfrentamiento en el 2021 y el 44.9% de los hogares conto con al menos 1 víctima de robo y/o extorsión. En comparación la escuela con mejor clasificación de la prueba planea de la Ciudad de México fue el Colegio Libertad que conforme al semáforo delictivo, tiene menor cantidad de enfrentamientos, robos y/o extorsiones. Con lo anterior se concluye que Iztacalco es más seguro que Iztapalapa dando como resultado un ambiente escolar más tranquilo y por ende facilitando el aprendizaje.

El propósito por el cual decidimos realizar este proyecto es facilitar la toma de decisiones de los padres de familia en cuanto a en que colegio pueden inscribir a su hijo, esto debido a que la Ciudad de México está en constante crecimiento y puede llegar a dificultar esta tarea. En el proyecto hicimos varios mapas interactivos que cuentan con filtros para poder encontrar las escuelas más cercanas a ti, hubo un gran problema con nuestro proyecto y es que no se encontraban en la base de dato

# Sección 5: Discusión

El propósito por el cual decidimos realizar este proyecto es facilitar la toma de decisiones de los padres de familia en cuanto a en que colegio pueden inscribir a sus hijos, esto debido a que la Ciudad de México está en constante crecimiento y esto puede llegar a dificultar esta tarea. En el proyecto hicimos varios mapas interactivos que cuentan con filtros para poder encontrar las escuelas más cercanas a la ubicación del usuario. Hubo un gran problema con nuestro proyecto, el cual fue que no se encontraba información en la base de datos sobre preparatorias y universidades, esto ocasionando que solo se pueda facilitar la selección de escuelas con de nivel de kínder a secundaria. Otro problema con el que nos encontramos es que no había un organismo confiable con el cual pudiéramos proporcionar rangos de precio (dentro de las escuelas privadas), nivel de las instalaciones ni una evaluación a los docentes, todas las fuentes que proporcionaban un análisis que incluía estos rubros lo hacía por medio de comentarios de padres de familia, docentes o directores. Esto conlleva a que estén involucrados en la escuela y no puedan mantenerse imparciales. Con lo previamente mencionado, se facilitó la toma de decisiones hasta cierto punto, ya que puedes encontrar las escuelas en diferentes grados de escolaridad cerca de tu ubicación o la ubicación que busques dentro de la Ciudad de México, sin embargo, dada la oportunidad de obtener más información sobre las escuelas podría facilitarse aún más la toma de decisiones.

# Sección 6: Conclusión

Al final de la elaboración del proyecto, brindamos una herramienta más que ayude a los padres o tutores a elegir la mejor opción escolar para sus hijos, desde nivel básico a nivel medio. Esta herramienta ayudara a mejorar la calidad de vida tanto de los estudiantes como de sus tutores o padres, ya que el punto más fuerte del trabajo es poder ver las distancia que existe entre la escuela y la ubicación del usuario, decimos que puede mejorar su calidad de vida, ya que al estar en la CDMX una de las ciudades más pobladas del mundo, el trasladarse dentro de la ciudad puede llevar a una gran cantidad de tiempo debido al tránsito vehicular, así que si la selección de la escuela se toma en cuenta con la distancia, las personas no estarán detenidas mucho tiempo en el tráfico vehicular, el ruido de los demás carros ni soportar los climas extremistas de la CDMX dentro del carro ocasionando que se disminuya su estrés.

# Referencias

C. (2019, 9 agosto). *La importancia de la intervención temprana en educación especial*. El Regional Del Sur. Recuperado 21 de noviembre de 2021, de <https://elregional.com.mx/la-importancia-de-la-intervencion-temprana-en-educacion-especial>

CHILANGO. (s. f.). *Chilango.com*. LA MEJOR ESCUELA PARA TI. Recuperado 20 de noviembre de 2021, de <https://escuelas.chilango.com>

ESCALE. (s. f.). *Mapa de Escuelas*. ESCALE (Estadística de la Calidad Educativa). Recuperado 21 de noviembre de 2021, de <https://www.scribbr.es/detector-de-plagio/generador-apa/new/webpage/>

Gobierno del Estado de México. (s. f.). *Cédula de Registro del Trámite o Servicio*. Inscripción al Centro de Atención Múltiple (CAM). Recuperado 21 de noviembre de 2021, de <http://sistemas2.edomex.gob.mx/TramitesyServicios/Tramite?tram=1058&cont=0>

Los mejores colegios. (2021, 13 mayo). *Guía de Las Mejores Escuelas de México Colegios Primarias Preparatoria*. Los Mejores Colegios de México. Recuperado 21 de noviembre de 2021, de <https://losmejorescolegios.mx>

Lozano, T. S. P. (s. f.). *La importancia de la escuela, el profesor y el trabajo educativo en la atención a la deserción escolar*. SciELO. Recuperado 21 de noviembre de 2021, de <http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1992-82382018000300024>

Mejora tu escuela. (s. f.). *Mejora tu escuela*. Recuperado 18 de noviembre de 2021, de <http://www.mejoratuescuela.org/compara/#?entidad=9&p=1&sort=Sem%C3%A1foro%20de%20Resultados%20Educativos&type_test=planea&schoolStatus=1&niveles=12&control=2>

SEP. (s. f.). *GIS - SEP(DF)*. Sistema de Información Geográfica para Escuelas de Educación Básica. Recuperado 19 de noviembre de 2021, de <https://gis.aefcm.gob.mx/indexs.html>

INEGI. (2020). *División municipal. Distrito Federal*. Cuentame INEGI. <http://www.cuentame.inegi.org.mx/monografias/informacion/df/territorio/div_municipal.aspx?tema=me&e=09>

Justo Sierra. (2020). ¿Cómo elegir entre escuelas de Educación Básica en CDMX? Recuperado 21 de noviembre de 2021, de <https://blog.justo-sierra.edu.mx/como-elegir-escuelas-educacion-basica-cdmx>

Expansión. (2010). Los 6 puntos básicos para elegir escuela. Recuperado 21 de noviembre de 2020, de <https://expansion.mx/actualidad/2010/01/21/como-escoger-escuela>