**Disminuyendo el tiempo de desplazamiento entre casa y trabajo**

**Problema**

En la actualidad Bogotá vive una de las peores congestiones vehiculares de su historia (<https://www.bbc.com/mundo/noticias-39045803>), esto lo podemos constatar según un artículo de la BBC mundo, basados en esta problemática pretendemos realizar un estudio de mercado para buscar el mejor lugar donde establecer un negocio (en este ejercicio instalaremos una cafetería) y a su vez encontrar un lugar de residencia dentro de la localidad seleccionada, con el fin de viabilizar el emprendimiento de un nuevo negocio y disminuir el tiempo de desplazamiento entre estos dos lugares.

**Público objetivo**

Este proyecto está destinado a aquella población emprendedora o que quieran ser independiente, que desee tener una buena ubicación entre el negocio y la vivienda para disminuir los tiempos de desplazamiento y así mejorar en gran medida su calidad de vida.

**Datos**

Según la definición de nuestro problema, los factores que influirán son:

* Concentración de establecimientos comerciales en los barrios.
* Tipos de establecimientos comerciales que no compitan con el nuestro.
* Zonas que se destaquen por su ambiente residencial.
* Distancia entre la ubicación de vivienda y la ubicación comercial.

Se necesitarán de las siguientes fuentes de datos para extraer y generar la información requerida.

* Ubicaciones geo codificadas de los barrios en la ciudad de Bogotá extraídos del **Laboratorio Urbano Bogotá**

<https://bogota-laburbano.opendatasoft.com/explore/dataset/barrios_prueba/table/>

* Número de establecimientos comerciales, tipo y ubicación, que se obtendrá de la **API de Foursquare**.
* Distancia de las zonas residenciales y comerciales que se cotejarán entre las fuentes de datos de **Laboratorio Urbano Bogotá** y la **API de Foursquare**.

**Limpieza de datos y selección de características**

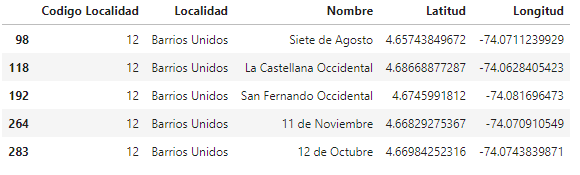
La base de datos del **Laboratorio Urbano Bogotá** está conformada por 3871 filas y 10 columnas descritas de la siguiente manera:

* OBJECTID
* Codigo localidad
* Localidad
* Estado
* Nombre
* Codigo
* SHAPE.AREA
* SHAPE.LEN
* Geo\_shape
* geo\_point\_2d

De las cuales serán usadas:

* Codigo Localidad : Numero con el cual se identifica cada localidad
* Localidad : Nombre de la localidad
* Nombre : Nombre del barrio
* geo\_point\_2d : Ubicación geográfica del barrio

Una vez extraídas estas columnas procedemos a eliminar los espacios sin información en caso de haberlos lo que nos deja 3871 filas y 4 columnas lo que indica una aceptable calidad de los datos, por último es necesario dividir en la columna geo\_point\_2d que contiene en un solo string tanto la latitud como longitud, este procedimiento es necesario para lograr una correcta ingesta de los datos, quedando de un tamaño de 3871 filas por 5 columnas. El último filtro realizado es la elección de la localidad “Barrios Unidos” lo que nos deja una matriz de 66 filas y 5 columnas.



**Metodología**

En este proyecto dirigiremos nuestros esfuerzos a encontrar dos lugares específicos con distintas características dentro de la localidad seleccionada.

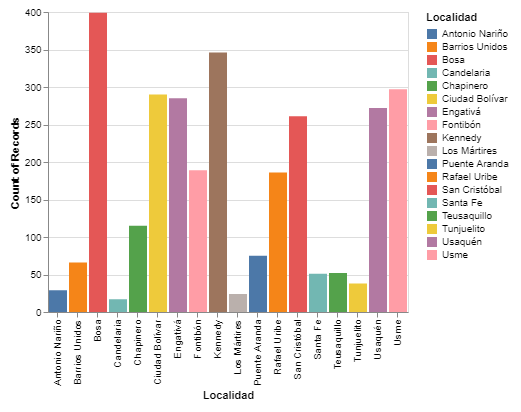
En el primer paso buscaremos dentro de la localidad cuales son los barrios que concentran la mayor parte del comercio, ya que esto nos asegura que en estas zonas tendremos mayor probabilidad de éxito en ventas para nuestra tienda, seguido a este análisis excluiremos aquellos barrios que ya cuenten con negocios de cafeterías puesto que la intención es ingresar en un área que tenga una baja competencia en este campo.

En el segundo paso buscaremos dentro de la localidad los barrios residenciales que identificaremos según las áreas de esparcimiento o descanso, también que tengan una baja densidad de comercio pero lo suficiente para pequeños abastecimientos.

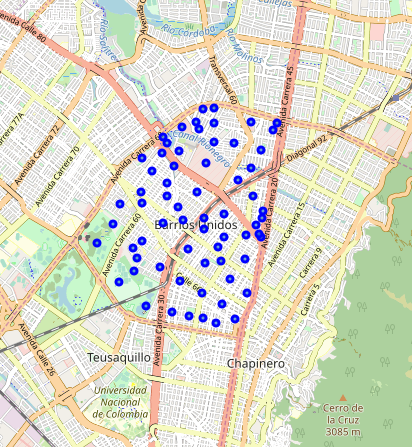
El tercer y último paso será enfocarnos en las áreas más prometedoras tanto de vivienda como comerciales que cumplan con nuestros criterios, cotejaremos las ubicaciones y definiremos cuales son los dos puntos que tengan menor distancia ya que nuestro objetivo es disminuir el tiempo de desplazamiento entre estos dos lugares.

**Exploración y análisis**

Primero fue necesario observar la distribución de los barrios en las localidades para decidir con cuál de ellas se trabajaría, por una limitante técnicas como el bajo número de consultas en la API de Foursquare se decidió trabajar con “Barrios Unidos” que tiene una cantidad de barrios de 66.



Con ayuda de la librería folium marcamos los barrios en la localidad elegida.

****

Esta primera impresión nos muestra claramente los límites de la localidad sobre el territorio y nos da una idea de donde pueden ubicarse las zonas más comerciales, pero para llegar a un concepto claro debemos enriquecer aún más nuestros datos.

Consulta a la API Foursquare

Por medio del uso de la API Foursquare procedemos a consultar los lugares que están en las inmediaciones de nuestros barrios, de esta manera descubrimos cuales son los tipos de establecimiento que se vincula a cada uno de nuestros puntos, para analizar más de fondo nuestros datos es necesario convertirlos a una matriz one hot.

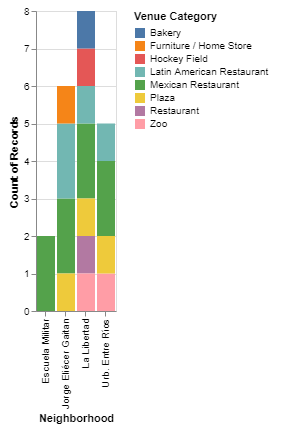
Un rápido conteo de los lugares más populares muestra un alto número de restaurantes, posiblemente un indicador de una riqueza comercial en la zona, también notamos que entre nuestro top 10 aparecen supermercados, jardines y parques, esto es importante para nuestro segundo propósito que es la búsqueda de un lugar de vivienda.

****

Para profundizar más en nuestros datos usamos un modelo de aprendizaje no supervisado de closterizacion llamado Agglomerative Clustering, una vez obtenemos el resultado analizamos cada uno de los grupos resultantes.

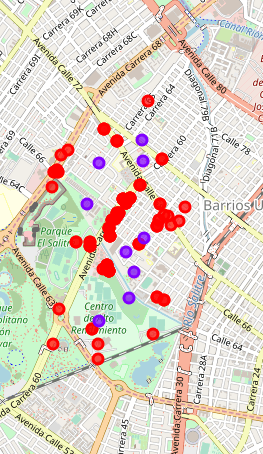
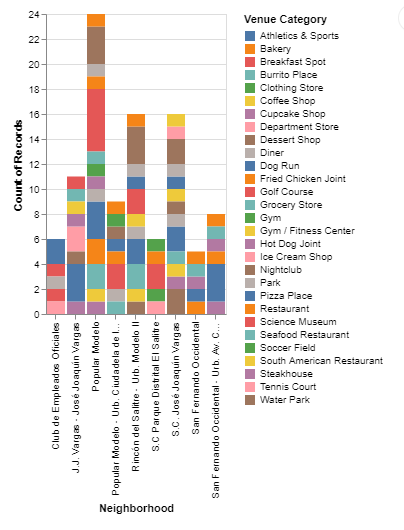
Para facilitar el análisis grafico marcamos en el mapa los barrios de color morado y los establecimientos comerciales en rojo.

Clúster 0

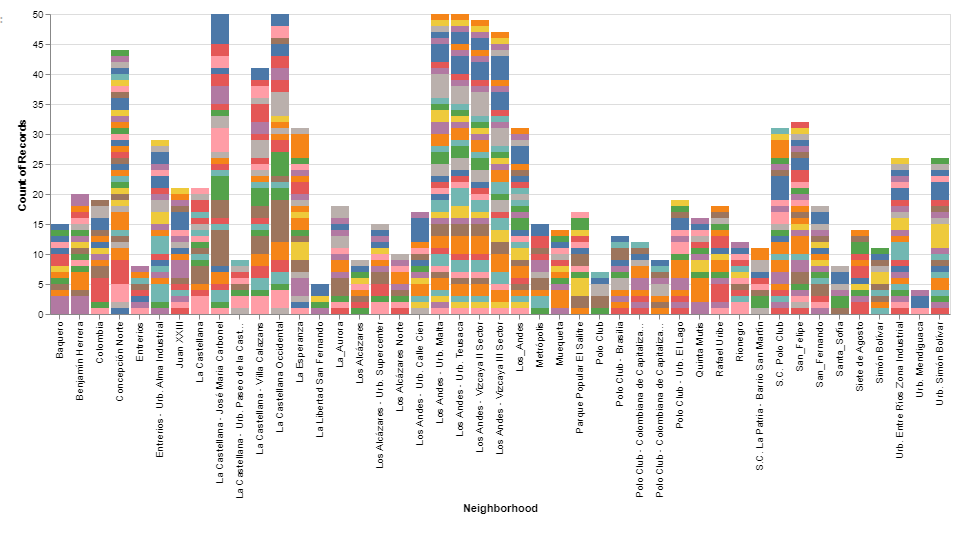
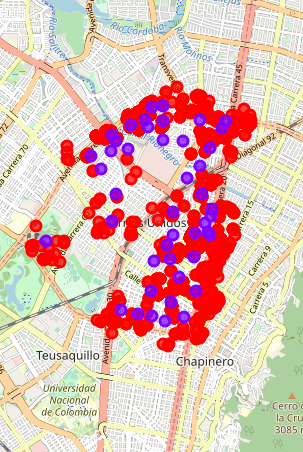
En el grupo 0 vemos que los lugares cercanos a los vecindarios son en su mayoría areas residenciales, el bajo volumen de comercio nos indica una buena opción para hallar nuestra vivienda.

Clúster 1

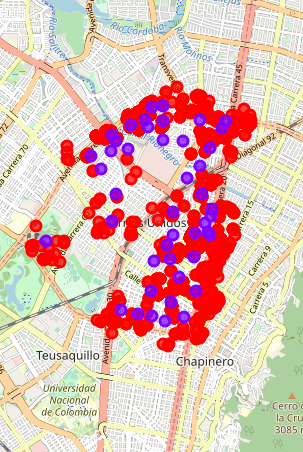
 

El clúster 1 aunque presenta un alto numero de locales comerciales destaca tambien los lugares de esparcimiento social como son las canchas de tennis, parques acuaticos, gyms y canchas de futbol, perfilandose como barrios de vivienda.

Clúster 2

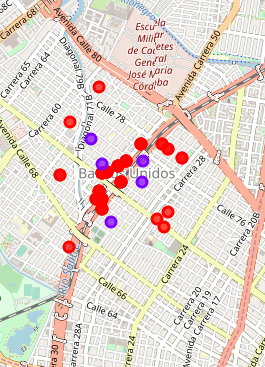
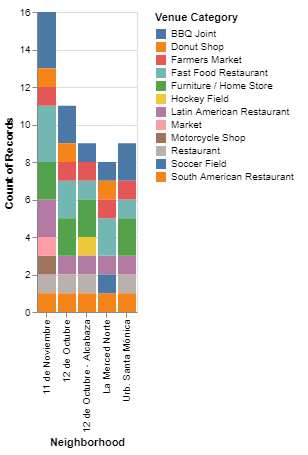
 





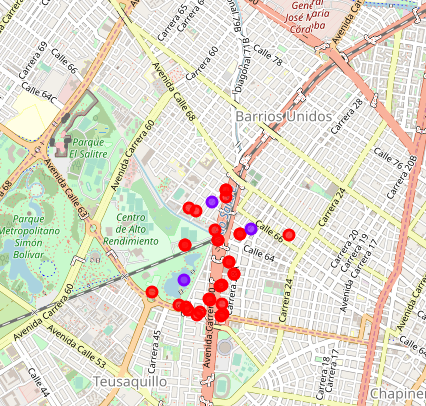
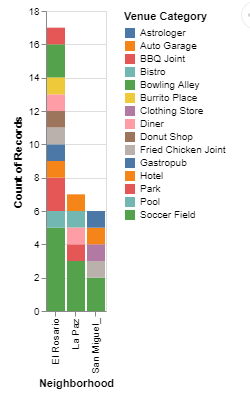
El clúster 2 se destaca por ser el que concentra la mayor cantidad de locales comerciales, nos muestra también que dichos establecimientos bordean los límites de la localidad, coincidiendo con las avenidas principales, estas características nos aseguran un alto flujo de personas y aumentan nuestra probabilidad de tener un alto número de ventas en los negocios.

Clúster 3

Este clúster es una pequeña zona comercial que se gesta alrededor de la avenida que atraviesa la localidad de sur a norte, al igual que el clúster 2 el alto tráfico de personas por estas vías garantizan mayor demanda de los productos y por ende mayores ventas.

Clúster 4

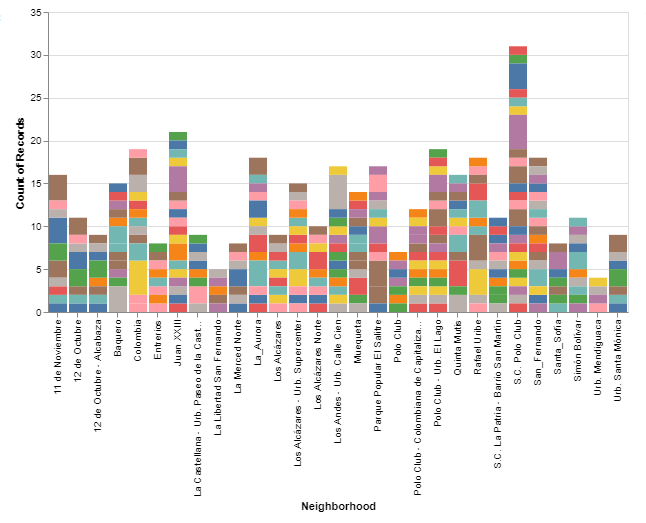
 

Los lugares que conforman este grupo se encuentran boleras, canchas de futbol, parques, piscinas y garajes, esto nos indica que puede ser una zona de bodegas, por lo que descartaremos este clúster para nuestra búsqueda.

Del análisis realizado se deduce:

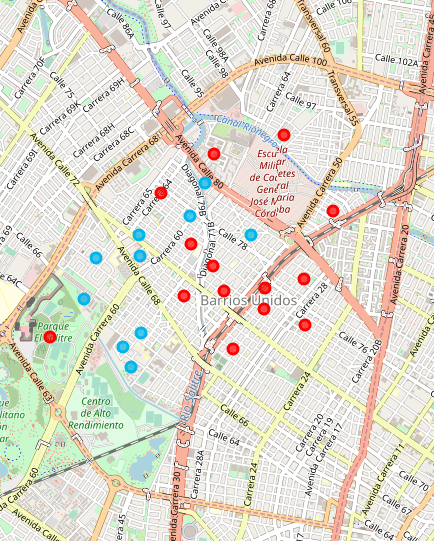
* Los clúster 0 y 1 se utilizarán para encontrar nuestro lugar de vivienda
* Los clúster 2 y 3 se utilizarán para encontrar el mejor lugar para ubicar nuestro negocio.

Decidido esto filtramos de los clúster 2 y 3 partiendo del criterio de eliminar aquellos barrios que ya cuenten con establecimientos de café, pasamos de tener 50 posibles barrios a 29 barrios potenciales donde instalar nuestra cafetería.



Para los grupos de viviendas procedimos a unir los clúster 0 y 1 y descartamos aquellos lugares que no son barrios como, el Club de Empleados Oficiales, Escuela Militar, S.C Parque Distrital El Salitre, ya que estos no entrarían en esta categoría, este proceso nos deja un total de 10 barrios con los cuales trabajar para encontrar nuestra vivienda.

Finalmente se han calculado las distancias entre los barrios de los clústeres con alto comercio y aquellos catalogados como opción de vivienda, se filtraron las distancias entre estos dos tipos de lugares, tomando como criterio que una persona recorre en promedio 5 Km a pie en un tiempo de 60 minutos, se decidió hallar los lugares que se encontraran a menos 15 minutos de recorrido lo que nos da un aproximado de un kilómetro entre ellos, lo que nos deja 24 lugares repartidos en 10 barrios de vivienda y 14 barrios comerciales, de esta manera concentramos nuestros esfuerzos en localizar aquellos 2 puntos que reduzcan nuestro tiempo de trayecto entre nuestro hogar y nuestro lugar de trabajo.



El resultado de este último filtro nos muestra en el mapa de color azul aquellos lugares que consideramos aptos para vivienda, y en rojo aquellos lugares que tienen un alto volumen comercial, la mayoría de ellos con distancias no mayores a 15 minutos de distancia recorridos a pie, aunque entre algunos puntos extremos de los grupos puede llegar a ser mayor el recorrido.

Es importante destacar que la “última milla” de este resultado se basaría en acudir a los puntos que determinamos como nuestra mejor opción, gracias a la tecnología podemos hacer una visita virtual a dos de estos lugares que en mi criterio son buenas opciones.

Punto de vivienda

<https://www.google.com/maps/@4.674412,-74.0721533,3a,75y,237.47h,92.78t/data=!3m6!1e1!3m4!1syPCgfN_T7JHALjeu3y1oDA!2e0!7i13312!8i6656>

Punto de comercio

<https://www.google.com/maps/@4.6703408,-74.0729676,3a,75y,32.43h,102.34t/data=!3m6!1e1!3m4!1sXSXGtGGfZtANo1vqH3jt-w!2e0!7i13312!8i6656>

**Resultados y discusión**

Durante el proceso se implementó un algoritmo de closterizacion encontrando 5 grupos con diferentes tendencias en cuanto a la densidad de lugares y los tipos de comercio, el número de grupos es discutible ya que después de la segmentación fue necesario unir los clúster 2 y 3 para tener un total de 1045 establecimientos, en estos grupos se encontraban la mayor concentración de locales comerciales. lo que indicaba un flujo de personas alto, lo que en teoría puede mejorar la demanda de nuestro producto, abarcamos un radio de 500 metros para encontrar la mayor cantidad de negocios objetivos, en este caso las cafeterías y así asegurarnos de tener los lugares que posiblemente estén más aislados dentro de los grupos, también dejamos un máximo de 50 lugares por punto, estos parámetros pueden ser ajustados para mejorar la precisión de nuestra búsqueda o en aquellos casos donde la densidad poblacional no sea tan alta como en ciudades pequeñas o pueblos.

Los barrios con menor cantidad de establecimientos comerciales presentaban características como el estar cerca de parques, cines, plazas, teatros, colegios y centros comerciales, por supuesto este escenario se presentaba como una buena opción donde ubicar nuestra vivienda, aquellos clústeres con esta tendencia fueron el 0 y el 1, al tener que ser unidos vuelve a replantear la idea de usar un menor número de grupos en el modelo de scikit-learn.

El propósito de este análisis fue encontrar una forma práctica de disminuir el tiempo que dedicamos en el tráfico de la ciudad o en el mejor de los casos evitarlo, razón por la cual se hizo el filtro de las distancias entre los puntos, debido a que en Bogotá el tráfico tiene una alta congestión se optó porque el recorrido fuese a pie, de ahí a que la distancia elegida fuese tan baja.

**Conclusión**

Pese que el objetivo de este análisis constaba simplemente en intentar disminuir o en el mejor de los casos evitar el tráfico durante el desplazamiento entre la vivienda y el trabajo en una de las ciudades consideradas con un alto déficit de movilidad se debieron hacer varios análisis que indirectamente nos fueran llevando al objetivo planteado.

La problemática puede ser abordada de muchas otras maneras como por ejemplo aquellas personas que ya tengan un lugar de trabajo establecido y solo deseen encontrar una mejor opción de vivienda, también aquellas personas que se encuentren en un estado de vulnerabilidad como personas discapacitadas, adultos mayores o madres cabeza de hogar quienes por su situación son quienes más deben emprender debido a que las sociedades tienden a segregarlas únicamente por su condición, esta aplicación puede llegar mejorar sus condiciones de vida disminuyéndoles el tiempo de desplazamiento y a su vez asegurarles el mejor lugar para desarrollar su actividad comercial.

Pese a que la información recabada gracias a la API de Foursquare fue suficiente en este caso, aún se puede enriquecer más al hacer uso de datos como el valor de los arriendos en las diferentes localidades ya que los diferentes estratos sociales que se distribuyen sobre el territorio determinan en gran medida los costos de vida en cada sector, finalmente la persona interesada tomará la decisión basada en sus gustos personales, las características de los barrios, dinámica social y económica, esta aplicación le brindará un panorama mucho más claro acerca de esta cuestión.