

Objetivo

- Poner en práctica los conceptos aprendidos en todos los módulos de la clase para poder dar respuesta a problemáticas reales.
- Los estudiantes utilizarán las estructuras de datos estudiadas, teniendo en cuenta las ventajas y desventajas que exhiben y su aplicabilidad a cada uno de los requerimientos.

Contexto

El estado actual de las ciudades, el crecimiento continuo de su población, infraestructura y servicios; hace que las agencias administradoras necesiten contar con información actualizada de los aspectos que afectan la vida de los ciudadanos, tales como: educación, salud, vivienda, transporte, infraestructura, entretenimiento, seguridad, entre muchos otros. Esta información permite a las agencias locales mejorar el servicio a los ciudadanos y tomar mejores decisiones que afecten sus vidas a largo plazo.

El transporte es una problemática muy importante para la ciudad. En particular, el transporte público debe ofrecer soluciones eficientes para transportar un alto volumen de la población, con buena calidad, para así aumentar la calidad de vida de sus ciudadanos. Además, un buen servicio de transporte público puede incentivar la reducción del transporte privado. Por esta razón, el **servicio de taxis** se transforma y mejora prestando el servicio a través de plataformas tecnológicas. Entre las empresas que vienen impulsando este cambio mundial están entidades como Uber, Lyft y Hailo.

En este proyecto, vamos a construir una aplicación que le permita realizar a los administradores y autoridades de la ciudad de Chicago un conjunto de consultas importantes sobre el servicio de taxis para soportar la toma de decisiones acerca de este servicio.

Chicago (USA) está a la vanguardia del registro de información sobre el funcionamiento del servicio de taxis. (URL oficial *Chicago Data Portal* <https://data.cityofchicago.org>). Para analizar este servicio utilizaremos como fuente de información los datos disponibles en el URL: <https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew>.

Como parte de los servicios ofrecidos por el Portal se tiene el API SODA (*Socrata Open Data API*) que permite hacer consultas sobre la información disponible usando como punto de consulta la URL <https://data.cityofchicago.org/resource/wrvz-psew.json> (detalles sobre el uso del API en URL <https://dev.socrata.com/>).

Suministraremos la información obtenida a través del API para este proyecto, por lo que **NO ES NECESARIO** hacer uso directo de esta fuente de información, pero es importante entender de donde proviene la información y para qué es utilizada.

Las Fuentes de Datos

A continuación, se presenta una descripción de las fuentes de datos que se utilizarán en el proyecto.

Archivos de Servicios:

- *Taxi_Trips-2020-subset-small.csv*
- *Taxi_Trips-2020-subset-medium.csv* y
- *Taxi_Trips-2020-subset-large.csv*

Cada archivo de datos contiene el detalle de un subconjunto de servicios de taxi (carreras), de distancias y de tiempo diferentes. Cada servicio de taxi (carrera) se describe a partir de las siguientes veinte y tres (23) columnas de datos:

| Nombre de la Columna | Descripción | Tipo de Dato |
|----------------------------------|--|--------------|
| Trip ID | Identificador único del viaje. | Plain Text |
| Taxi ID | Identificador único del taxi. | Plain Text |
| Trip Start Timestamp | Cuando comenzó el viaje, redondeado a los 15 minutos más cercanos. | Date & Time |
| Trip End Timestamp | Cuando finalizó el viaje, redondeado a los 15 minutos más cercanos. | Date & Time |
| Trip Seconds | Tiempo del viaje en segundos. | Number |
| Trip Miles | Distancia del viaje en millas. | Number |
| Pickup Census Tract | La zona del censo donde comenzó el viaje. Por motivos de privacidad, esta zona del censo no se muestra para algunos viajes. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Plain Text |
| Dropoff Census Tract | La zona del censo donde terminó el viaje. Por motivos de privacidad, esta zona del censo no se muestra para algunos viajes. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Plain Text |
| Pickup Community Area | El Área Comunitaria donde comenzó el viaje. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Dropoff Community Area | El Área Comunitaria donde finalizó el viaje. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Fare | La tarifa del viaje. | Number |
| Tips | La propina para el viaje. Por lo general, las propinas en efectivo no se registrarán. | Number |
| Tolls | Los peajes del viaje. | Number |
| Extras | Cargos extra por el viaje. | Number |
| Trip Total | Costo total del viaje, el total de las columnas anteriores. | Number |
| Payment Type | Tipo de pago del viaje. | Plain Text |
| Company | La empresa de taxis. | Plain Text |
| Pickup Centroid Latitude | La latitud del centro del tramo censal de recogida, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Pickup Centroid Longitude | La longitud del centro del tramo censal de recogida, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Pickup Centroid Location | La ubicación del centro del tramo censal de recogida, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Point |

| Nombre de la Columna | Descripción | Tipo de Dato |
|-----------------------------------|--|--------------|
| Dropoff Centroid Latitude | La latitud del centro del tramo censal de bajada, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Dropoff Centroid Longitude | La longitud del centro del tramo censal de bajada, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Number |
| Dropoff Centroid Location | La ubicación del centro del tramo censal de bajada, o del área comunitaria si el tramo censal se ha ocultado por motivos de privacidad. Esta columna a menudo estará en blanco para ubicaciones fuera de Chicago. | Point |

Carga de Información

Para responder a los siguientes requerimientos cada grupo deberá cargar una única vez la información de uno de los archivos a utilizar (*Taxi_Trips-2020-subset- $\langle\langle size \rangle\rangle$.csv*). El usuario debe tener la posibilidad de escoger entre las opciones “small”, “medium” o “large”. Esta funcionalidad se debe reflejar en el menú de opciones como una carga de datos.

Requerimientos

Parte A

Primero, las autoridades de tránsito con el fin de reducir las quejas, optimizar el servicio prestado y mejorar los estándares de las compañías se desea crear un reporte que permita conocer la siguiente información:

- El número total de taxis en los servicios reportados.
- El número total de compañías que tienen al menos un taxi inscrito.
- El top M de compañías ordenada por la cantidad de taxis afiliados. Se debe reportar para cada compañía su nombre y el número de taxis afiliados. M es un dato de entrada.
- El top N de compañías que más servicios prestaron, el ranking se ordenara por el número de servicios de mayor a menor. Se debe reportar para cada compañía su nombre y el número de servicios de prestados. N es un dato de entrada.

Nota: Los taxis que no registran una compañía se deben agrupar bajo el nombre de la compañía “*Independent Owner*”.

Parte B

Segundo, como parte del plan de mejora del servicio de taxis, Chicago estableció un sistema de puntos y premios. Para calcular los puntos asignados a un taxi se calcula una función alfa diaria; esta función se define como la división del total de millas recorridas entre el total de dinero recibido, esto multiplicado por el total de servicios prestados. Para ello se desea construir un programa que permita:

- Identificar los N taxis con más puntos para en una fecha determinada. N y la *fecha* son datos de entrada.

- Identificar los M taxis con más puntos para un rango entre dos fechas determinadas (fecha inicial y fecha final). M y el *rango de fechas* son datos de entrada.

Nota: Se deben tomar en cuenta los servicios para los cuales se tiene una distancia mayor a 0.0 y se paga un valor mayor a U\$ 0.0.

Parte C

Finalmente, con miras a mejorar el servicio para los usuarios finales se propone la implementación de un servicio de consulta del mejor horario (hora:minutos) para desplazarse entre dos “Community Area”. Para tal fin se solicita resolver el siguiente problema: dados el área de origen, el área de destino y el rango de tiempo en el día en el que se quiere iniciar el viaje; generar un reporte que incluya: el mejor horario (hora:minutos) de inicio de viaje (en el área de origen) en el rango de tiempo dado para tener la menor duración posible, la ruta (secuencia de “community areas”) del viaje y el tiempo estimado (segundos) del viaje.

Tenga en cuenta que, si en los datos hay múltiples viajes iniciando en la misma franja horaria de una “community area” origen a una “community area” destino, la duración de un viaje entre el origen y el destino será la duración promedio de estos viajes.

Notas:

- Si existen viajes que relacionen una “Community Area” consigo misma, esos viajes no deben ser tenidos en cuenta.
- Utilice rangos de tiempo de 15 minutos para poder caracterizar el inicio de los viajes de cada “community area”. Por ejemplo, el inicio de los viajes desde el “community area” X se pueden caracterizar como X-00:00, X-00:15, X-00:30, X-00:45, X-01:00, X-01:15, X-01:30, X-01:45, X-02:00, ..., X-23:00, X-23:15, X-23:30, X-23:45

Ejemplo de consulta:
Dar el mejor horario (hora:minutos) del día para iniciar un viaje desde la “community area” 28 a la “community area” 56 en el rango de hora 11:30 a 12:30. Ambos límites del rango son también aceptables como inicio del viaje.

Entrega (Dic. 13 - 11:59 p.m.)

Como parte de la entrega deben entregar el código fuente y un documento en formato PDF que responda las siguientes preguntas por cada requerimiento:

- ¿Qué TAD utilizaron en la solución del requerimiento?
- ¿Por qué eligieron esa estructura de datos?
- ¿Cuál es la complejidad estimada del algoritmo implementado?