

Estructuras de Datos y Algoritmos

Proyecto Final

Proyecto Análisis de Servicios de Taxis en Chicago



**CHICAGO
DATA PORTAL**



Chicago (USA) está a la vanguardia del registro de información sobre el funcionamiento del servicio de taxis.

<https://data.cityofchicago.org/Transportation/Taxi-Trips-2020/r2u4-wwk3>

<https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew>

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.

Objetivo

- Poner en práctica los conceptos aprendidos en todos los módulos de la clase para poder dar respuesta a problemáticas reales.
- Diseñar e implementar un Tipo Abstracto de Datos (TAD) para modelar un sistema que permita resolver requerimientos (consultas) específicos del sistema de Taxis en Chicago.
- Utilizar las estructuras de datos estudiadas en el curso, teniendo en cuenta las ventajas y desventajas que exhiben y su aplicabilidad a cada uno de los requerimientos.

Fuente de datos – Taxi_Trips-2020.csv

Cada servicio de taxi (carrera) se describe a partir de veinte y tres (23) datos. A continuación se destacan algunos de estos datos a utilizar en el proyecto:

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 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
Company	La empresa de taxis.	Plain Text
Fare	La tarifa del viaje.	Number
Trip Total	Costo total del viaje: tarifa básica, propina, peajes, cargos extra	Number

Fuente de datos – Taxi_Trips-2020.csv

A continuación una muestra de viajes de taxi en el archivo:

Trip ID	Taxi ID	Trip Start Timestamp	Trip End Timestamp	Trip Seconds	Trip Miles	Pickup Community Area	Dropoff Community Area	Fare	Tips	Tolls	Extras	Trip Total	Payment Type	Company
b7466df6f45c05a		1/01/2020 0:00	1/01/2020 0:00	240	0.6		8	8 5.00	0.00	0.00	0.00	5.00	Cash	Choice Taxi Association
b7764b2161590e		1/01/2020 0:00	1/01/2020 0:15	600	1	2E+10	1,7E+10	8 7.25	3.00	0.00	1.00	11.25	Credit	Star North Management
ab5bad07abac49		1/01/2020 0:00	1/01/2020 0:15	383	0.69	2E+10	1,7E+10	8 5.75	0.00	0.00	0.00	5.75	Cash	Flash Cab
ab953ae155ffe17		1/01/2020 0:00	1/01/2020 0:30	1,861	3.16	2E+10	1,7E+10	32 16.00	3.60	0.00	1.50	21.60	Credit	City Service
ac4bf1d5690e6f6c		1/01/2020 0:00	1/01/2020 0:15	960	11.1		76	21 28.75	6.65	0.00	4.00	39.40	Credit	Star North Management
a7c16d64d75005e		1/01/2020 0:00	1/01/2020 0:15	660	2	2E+10	1,7E+10	8 9.00	0.00	0.00	0.00	9.00	Cash	Taxi Affiliation Service
a69e875487ff9e63		1/01/2020 0:00	1/01/2020 0:30	1,324	8.6		8	31 24.00	5.10	0.00	1.00	30.60	Credit	City Service
a350f7a55b5fd5f1		1/01/2020 0:00	1/01/2020 0:15	729	1.5	2E+10	1,7E+10	8 8.25	0.00	0.00	0.00	8.25	Cash	Medallion Leasing
a2ef37ca3fcae486		1/01/2020 0:00	1/01/2020 0:15	1,2	0		6	8 13.50	0.00	0.00	0.00	13.50	Cash	Choice Taxi Association
99f2acf7bf144d30		1/01/2020 0:00	1/01/2020 0:15	789	1.4	2E+10	1,7E+10	32 8.75	3.00	0.00	1.00	13.25	Credit	Sun Taxi
99d8f315a702917f		1/01/2020 0:00	1/01/2020 0:15	360	0		8	8 6.25	3.00	0.00	1.00	10.25	Credit	Taxi Affiliation Service
99757c14fd41bbe		1/01/2020 0:00	1/01/2020 0:15	840	0		8	7 11.75	0.00	0.00	0.00	11.75	Cash	Taxi Affiliation Service
983ff2702d33a620		1/01/2020 0:00	1/01/2020 0:30	1,2	0.2		32	24 15.50	0.00	0.00	1.50	17.00	Cash	Taxi Affiliation Service
92b00ac6d87bec		1/01/2020 0:00	1/01/2020 0:15	868	2.49		1	1 10.75	0.00	0.00	1.00	11.75	Cash	Chicago Carriage Cab
8e07bec2c5732d0		1/01/2020 0:00	1/01/2020 0:00	0	0	2E+10	1,7E+10	8 3.25	0.00	0.00	0.00	3.25	Cash	Taxi Affiliation Service
8ee4b26cce8f048		1/01/2020 0:00	1/01/2020 0:00	60	0		76	76 3.25	0.00	0.00	0.00	3.25	Cash	Blue Ribbon Taxi Association
889552b4f78222c		1/01/2020 0:00	1/01/2020 0:15	720	4.2		32	6 13.75	0.00	0.00	0.00	13.75	Cash	Taxicab Insurance Agency
87097b5d0c58d9		1/01/2020 0:00	1/01/2020 0:30	1,38	2.4	2E+10	1,7E+10	7 14.00	2.32	0.00	1.00	17.82	Credit	Patriot Taxi DBA Peapack
8542e861bca79085		1/01/2020 0:00	1/01/2020 0:15	540	3.5	2E+10	1,7E+10	28 11.50	0.00	0.00	1.00	12.50	Cash	Taxicab Insurance Agency

Diseño de la solución

Para resolver el proyecto, vamos a seguir algunos pasos que nos ayudarán en la solución del problema.

- 1- Identificar los elementos relevantes del problema y sus relaciones
- 2- Identificar los TADs del problema
- 3- Seleccionar las estructuras de datos de implementan los TADs seleccionados
- 4- Plantear algoritmos para la solución de los requerimientos y estimar la complejidad asociada a cada requerimiento
- 5- Implementar una solución para validar la propuesta realizada.

1- Identificación de elementos relevantes

Req. 1 – Reporte general: total de taxis, total de compañías, top M de compañías ordenada por la cantidad de taxis afiliados, top N de compañías que más servicios prestaron

Req. 2 – Sistema de puntos y premios basado en una función alfa diaria:

- Identificar los N taxis con más puntos en una fecha.
- Identificar los M taxis con más puntos en un rango de fechas

Req. 3 – Proponer el mejor horario (hora:minutos) para desplazarse en el menor tiempo entre dos “Community Area” en un rango horario definido. Obtener la hora inicial del viaje, la ruta (secuencia de “community areas”) y el tiempo estimado (segundos) del viaje.

1- Identificación de elementos relevantes

Ayuda: Identifique sustantivos en el enunciado del problema

Req. 1 – Reporte general: total de **taxis**, total de **compañías**, top M de **compañías** ordenada por la cantidad de **taxis** afiliados, top N de **compañías** que más **servicios** prestaron

Req. 2 – Sistema de **puntos** basado en una función alfa diaria:

- Identificar los N **taxis** con más **puntos** en una **fecha**.
- Identificar los M **taxis** con más **puntos** en un rango de **fechas**

Req. 3 – Proponer el mejor **horario** (hora:minutos) para desplazarse en el menor **tiempo** entre dos “**Community Area**” en un **rango horario** definido. Obtener la **hora inicial** del **viaje**, la **ruta** (secuencia de “community areas”) y el **tiempo** estimado (segundos) del viaje.

1- Identificación de elementos relevantes

Ayuda: Identifique sustantivos en el enunciado del problema

Taxi

Compañía

Servicio

Fecha

Ruta

Viaje

Puntos

Horario

Community Area

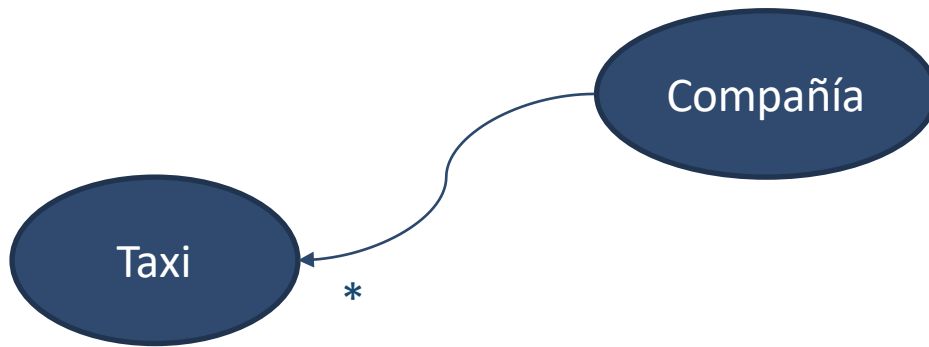
Rango Horario

Tiempo

1- Identificación de elementos relevantes

Cómo se relacionan estos elementos? Trate de usar un dibujo para expresar las relaciones encontradas

Taxi Compañía Servicio Viaje Ruta Puntos Horario Fecha Rango Horario
Community Area Tiempo



1- Identificación de elementos relevantes

Cómo se relacionan estos elementos? Trate de usar un dibujo para expresar las relaciones encontradas

Taxi Compañía Servicio Viaje Ruta Puntos Horario Fecha Rango Horario
Community Area Tiempo

2- Identificación de TADs

Trate de asociar un TAD a los elementos relevantes identificados en el punto anterior. Por ejemplo, hemos identificado el concepto Compañía. Cómo podemos guardar las compañías? Recuerde que podemos utilizar varios TADs para referenciar la misma información.

- Una lista de compañías?
 - Ordenada?
- Un Map de compañías?
 - Cuál sería la pareja <llave, valor> ?
- Un Map Ordenado?
 - Cuál sería la pareja <llave, valor> ?
- Un grafo?
 - Dirigido? Con pesos?

2- Identificación de TADs

Llene las primeras columnas: Entidad/Concepto, TADs y Propósito

Entidad/Concepto c	TAD	Propósito	Estructuras de Datos

3- Identificación de Estructuras de Datos

Ahora identifique las Estructuras de Datos seleccionadas para los TADs

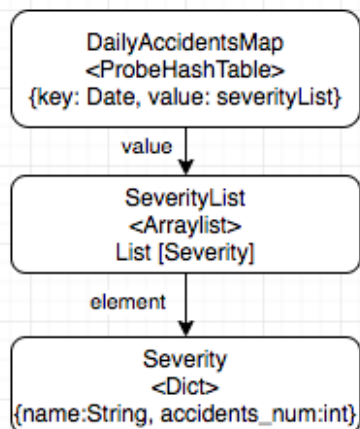
Entidad/Concepto	TAD	Propósito	Estructuras de Datos

4- Plantear algoritmos de solución para los requerimientos

Para cada uno de los requerimientos, se debe explicar cómo se usarán los TADs y Estructuras de Datos

Requerimiento (Ejemplo US Accidents) - Obtener la cantidad de accidentes por severidad para una fecha dada

Entidad/Concepto	TAD	Estructura de Datos	Función	Complejidad Temporal/Espacial
Accidentes	Map	Linear Probing	getNumberAccidentsOnDate(map, date)	$O(1) / O(N)$



Requerimiento 1: Reporte de Información Compañías y Taxis

Reporte que incluye:

- 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.

Entidad/Concepto	TAD	Estructura de Datos	Función	Complejidad Temporal/Espacial

Requerimiento 2: Sistema de Puntos y Premios a Taxis

- 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 1: La función alfa de puntos para un taxi se define como la división del total de millas recorridas entre el total de dinero recibido, esto multiplicado por el total de servicios prestados.

Entidad/Concepto	TAD	Estructura de Datos	Función	Complejidad Temporal/Espacial

Requerimiento 3: Consulta del Mejor Horario en Taxi entre 2 “community areas”

Proponer el mejor horario (hora:minutos) para desplazarse entre dos “Community Area” en un rango horario definido. Obtener la hora inicial del viaje, la ruta (secuencia de “community areas”) y el tiempo estimado (segundos) del viaje.

Entidad/Concepto	TAD	Estructura de Datos	Función	Complejidad Temporal/Espacial