

ANÁLISIS DE DATOS

Examen Parcial

NYC CITI BIKE TRIP

La ciudad de NYC le ha contratado para trabajar en obtener estadísticas y generar tableros de información de los viajes en bicicleta en la ciudad. La principal función de su nuevo contrato es poder generar un sistema de consulta para la toma de decisiones gerenciales y que permita el mejorar el negocio de manera general. La ciudad provee bicicletas en distintos lugares, en alquiler bajo una tarifa determinada (sin importar la ruta a tomar), el usuario de la bicicleta toma la bicicleta en un punto determinado de la ciudad y la puede dejar en otro punto determinado.



El funcionamiento del negocio

Para poder contratar el servicio, debe de descargarse un app y efectuar un registro con los datos básicos del cliente. Existen tres tipos de planes distintos que se pueden contratar:

1. Plan anual: Pagando por anticipado \$17 por mes para un total de \$205.
2. Single ride: se puede contratar el uso de la bicicleta para un viaje único por \$4.49 por 30 minutos. Por cada minuto adicional se cobra \$0.26.
3. Day pass: alquiler de la bicicleta por \$19 por 24 horas. Perfecto y recomendado para turistas.

El proceso del negocio



El set de datos

El dataset que se le entregara para poder efectuar el cubo dimensional y la visualización de datos, tiene almacenados todos los viajes efectuados durante los últimos 5 años y contiene las siguientes columnas:

Start time: Fecha y hora del inicio del viaje.

Stop time: Fecha y hora del fin del viaje.

Trip duration: Duración del viaje (en segundos).

Start Station id: Identificación única de la estación en la cual inicio el viaje.

Start Station Latitude: Latitud de la estación en la cual inicio el viaje.

Start Station Longitude: Longitud de la estación en la cual inicio el viaje.

Stop Station id: Identificación única de la estación en la cual finalizo el viaje.

Stop Station Latitude: Latitud de la estación en la cual finalizo el viaje.

Stop Station Longitude: Longitud de la estación en la cual finalizo el viaje.

Bike id: identificación de la bicicleta utilizada en el viaje.

User Type: tipo de usuario (de los tres identificados anteriormente).

Birth Year: Año de nacimiento del usuario que utilizo el servicio.

Gender: Genero del usuario que utilizo el servicio.

2013-10 - Citi Bike trip data												
tripduration	starttime	stoptime	start station id	start station latitude	start station longitude	end station id	end station latitude	end station longitude	bikeid	usertype	birth year	gender
326	2013-10-01 00:01:08	2013-10-01 00:06:34	239	40.69196566	-73.9813018	366	40.693261	-73.968896	16052	Subscriber	1982	1
729	2013-10-01 00:01:21	2013-10-01 00:13:30	322	40.696192	-73.991218	398	40.69165183	-73.999786	19412	Customer	1984	0
520	2013-10-01 00:01:24	2013-10-01 00:10:04	174	40.7381765	-73.97738662	403	40.72502876	-73.99069656	19645	Subscriber	1984	1
281	2013-10-01 00:01:25	2013-10-01 00:06:06	430	40.7014851	-73.98656928	323	40.69236178	-73.98631746	16994	Subscriber	1985	1
196	2013-10-01 00:01:27	2013-10-01 00:04:43	403	40.72502876	-73.99069656	401	40.72019576	-73.98997825	15690	Subscriber	1986	1
1948	2013-10-01 00:01:48	2013-10-01 00:34:16	369	40.73224119	-74.00026394	307	40.71427487	-73.98990025	19846	Subscriber	1977	1
1327	2013-10-01 00:01:48	2013-10-01 00:23:55	254	40.73532427	-73.99800419	539	40.71534825	-73.96024116	14563	Subscriber	1986	2
1146	2013-10-01 00:01:57	2013-10-01 00:21:03	490	40.751551	-73.993934	438	40.72779126	-73.98564945	16793	Subscriber	1959	1
380	2013-10-01 00:01:58	2013-10-01 00:08:18	468	40.7652654	-73.98192338	385	40.75797322	-73.96603308	16600	Customer	1984	0
682	2013-10-01 00:02:05	2013-10-01 00:13:27	300	40.728145	-73.990214	519	40.75188406	-73.97770164	15204	Subscriber	1992	1
863	2013-10-01 00:02:20	2013-10-01 00:16:43	322	40.696192	-73.991218	398	40.69165183	-73.9999786	19494	Customer	1984	0
664	2013-10-01 00:02:45	2013-10-01 00:13:49	482	40.73935542	-73.99931783	505	40.74901271	-73.98848395	17241	Subscriber	1973	1
324	2013-10-01 00:02:58	2013-10-01 00:08:22	463	40.74206539	-74.00443172	254	40.73532427	-73.99800419	19330	Subscriber	1991	1
537	2013-10-01 00:03:22	2013-10-01 00:12:19	285	40.73454567	-73.99074142	462	40.74691959	-74.00451887	16872	Subscriber	1974	1
400	2013-10-01 00:03:23	2013-10-01 00:10:03	499	40.76915505	-73.98191841	352	40.76340613	-73.97722479	19660	Subscriber	1974	1
777	2013-10-01 00:03:28	2013-10-01 00:16:25	410	40.72066442	-73.98517977	532	40.710451	-73.9600876	16628	Subscriber	1985	1
653	2013-10-01 00:03:31	2013-10-01 00:14:24	504	40.73221853	-73.98165557	531	40.71893904	-73.99266288	20304	Subscriber	1987	2
410	2013-10-01 00:03:50	2013-10-01 00:10:40	307	40.71427487	-73.98990025	291	40.713126	-73.984844	18353	Subscriber	1988	1
626	2013-10-01 00:04:07	2013-10-01 00:14:33	262	40.69178232	-73.97372989	143	40.69239502	-73.99337909	17118	Subscriber	1970	1
900	2013-10-01 00:04:09	2013-10-01 00:19:09	441	40.756014	-73.967416	379	40.749156	-73.9916	17079	Subscriber	1964	1
324	2013-10-01 00:04:12	2013-10-01 00:09:36	351	40.70530954	-74.00612572	308	40.71307916	-73.99851193	15865	Subscriber	1988	1
154	2013-10-01 00:04:15	2013-10-01 00:06:49	483	40.73223272	-73.98889957	280	40.73331967	-73.99510132	15370	Subscriber	1994	1
716	2013-10-01 00:04:30	2013-10-01 00:16:26	285	40.73454567	-73.99074142	251	40.72317958	-73.99480012	17702	Subscriber	1981	1

Preguntas

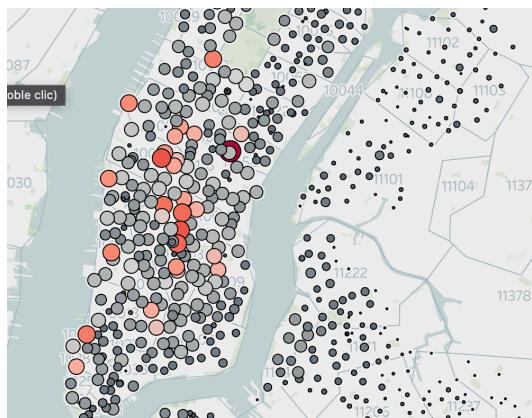
El alcalde de la ciudad de NYC, ha escuchado acerca de los cubos de información, sin embargo el CTO de la alcaldía considera que no es necesario la inversión en un sistema de data warehouse, pues considera que el programa es muy pequeño para que pueda generar un retorno adecuado para sostener la inversión.

El alcalde de NYC, le ha pedido que se le presenten sus recomendaciones al respecto, usted como buen ingeniero en sistemas, sabe que lo mejor para poder generar insights relevantes para el negocio y poder generar de manera adecuada información para la toma de decisiones se necesita generar un cubo de información.

1. Efectué un diseño dimensional con los datos provistos. Identifique plenamente las dimensiones y la tabla de hechos, describiendo cada uno de los campos que pertenecen a cada entidad (35 puntos)
2. Si usted pudiera enriquecer el diseño agregando datos, campos o otras fuentes de datos, de manera general y sin conocer mucho del negocio, describa que otras cosas le agregaría al diseño. (10 puntos)
3. Dejando aparte el tema de los costos y presupuestos de poder implementar un sistema de data warehouse, y tomando en cuenta los objetivos generales de tener un sistema DW/BI vistos en clase, describa que otros argumentos podría proveer para poder apoyar la implementación de un sistema de DW/BI. (15 puntos)
4. Como puede observar, no en todos los casos el año de nacimiento del cliente se ha registrado correctamente, asumiendo que el se mejora el app por medio de la cual se contratan los servicios de uso de las bicicletas, obligando ahora a actualizar el año de nacimiento de cada usuario, tomando en cuenta el concepto de las slow changing dimensions vistos en clase, cual de los tipos de SCD utilizaría. **JUSTIFIQUE SU RESPUESTA** (20 puntos).
5. El director del programa de NYC Bike, desea conocer cuales son las estaciones de bicicletas con mas demanda, de una manera grafica, El equipo de visualización de datos ha trabajado en tres propuestas para que usted elija, tomando en cuenta lo visto en clase, cual de estas propuestas elegiría usted y porque. **JUSTIFIQUE SU RESPUESTA.** (visualizaciones en la siguiente hoja). (10 puntos)
6. Se le presentan en la hoja no. 5 del presente documento, una serie de visualizaciones de datos del set de datos mencionado. Tomando en cuenta los conceptos vistos en clase de Everyday data Viz y Visual Discovery, clasifique cada uno de las visualizaciones presentadas en alguna de estas dos clasificaciones. Justifique cada elección. (10 puntos).

Pregunta 5

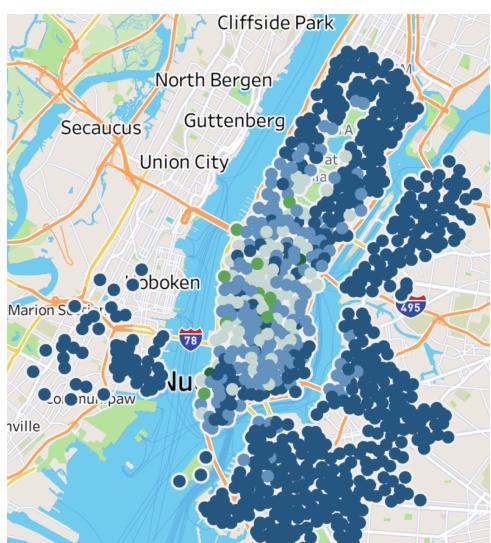
Propuesta 1



Propuesta 2

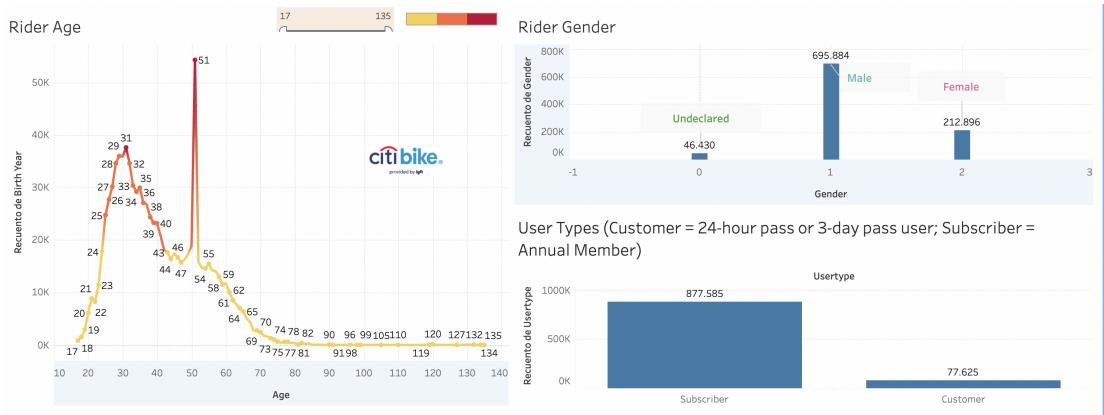


Propuesta 3

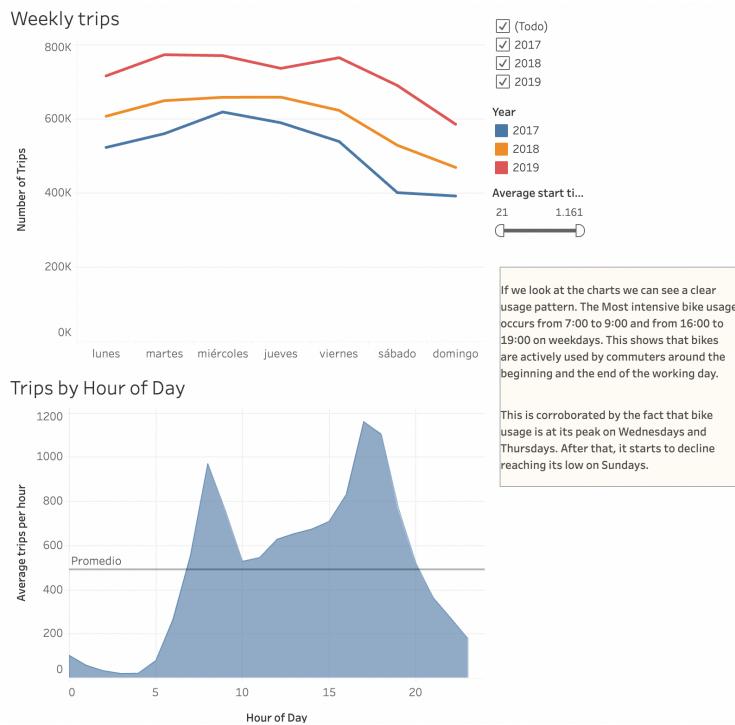


Pregunta 6

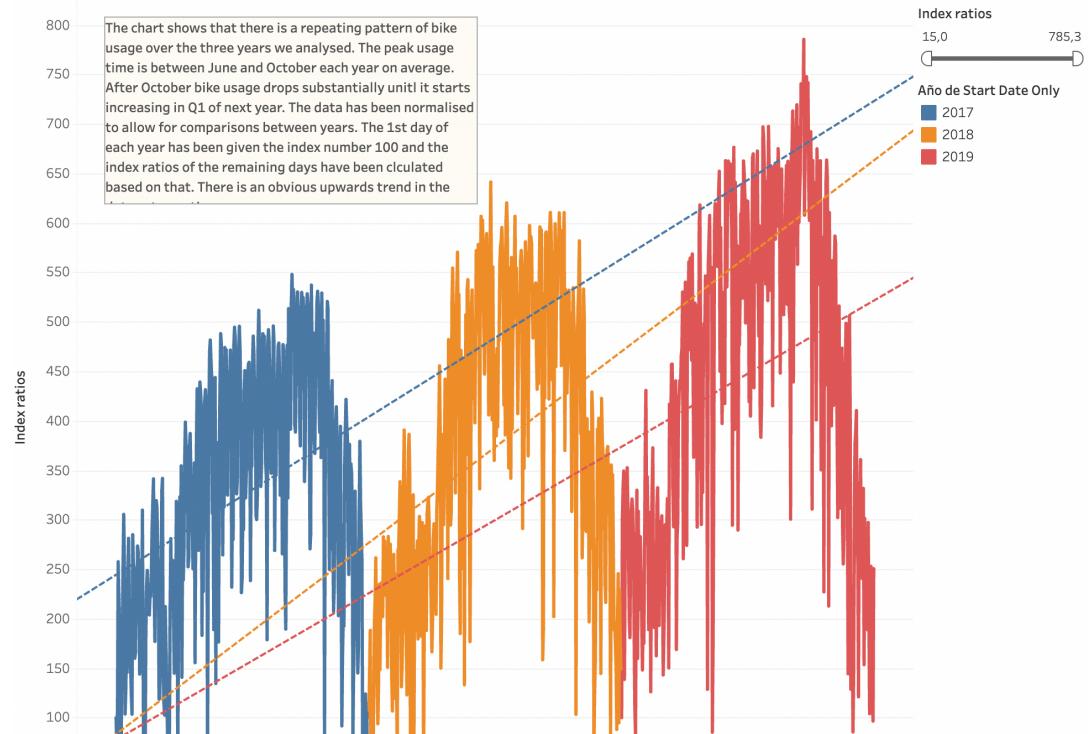
Visualización 1



Visualización 2



Visualización 3



Visualización 4

