**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

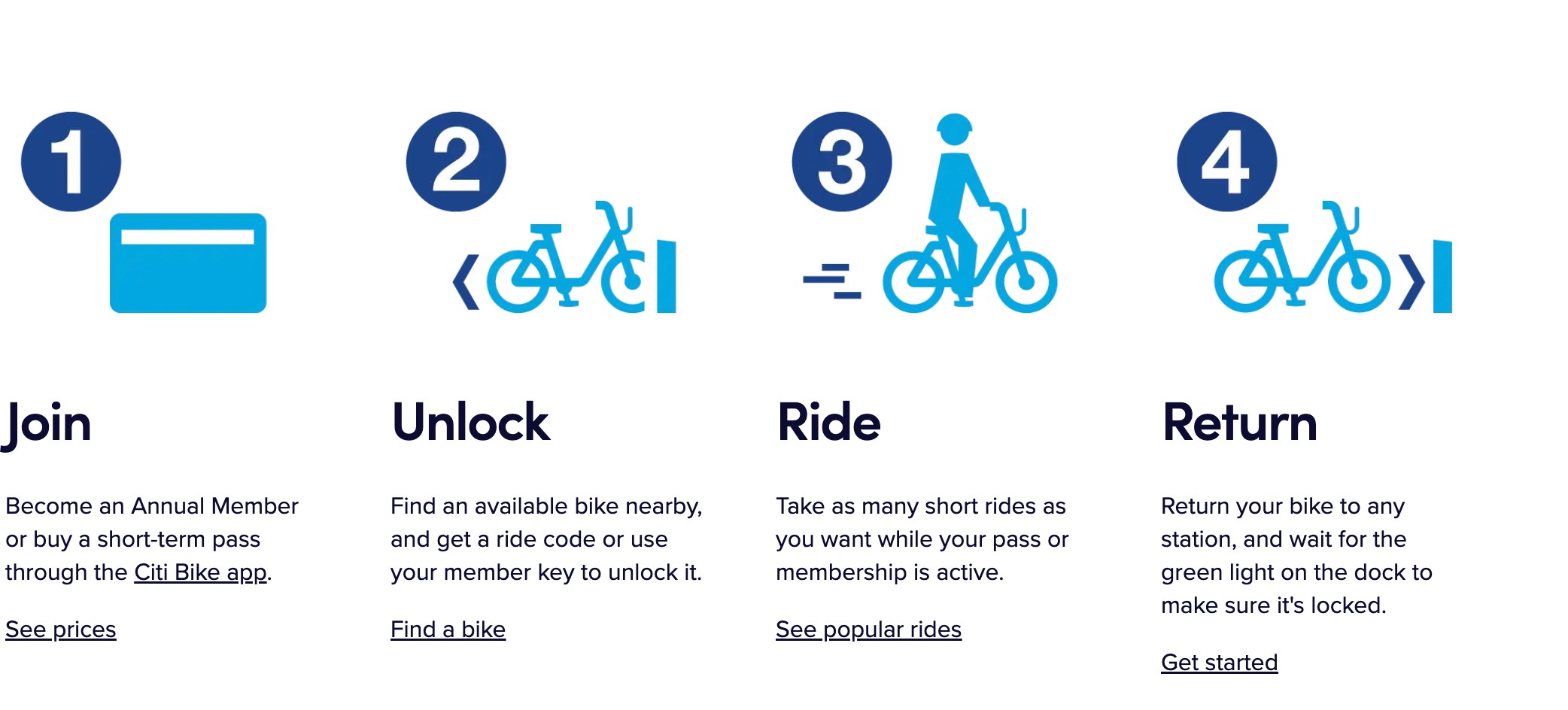
Para poder contratar el servicio, debe de descargarse un app y efectuar un registro con los datos básicos del cliente. Existen tres tipos de planes distintos que se pueden contratar:

1. Plan anual: Pagando por anticipado $17 por mes par aun 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.

1. 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.

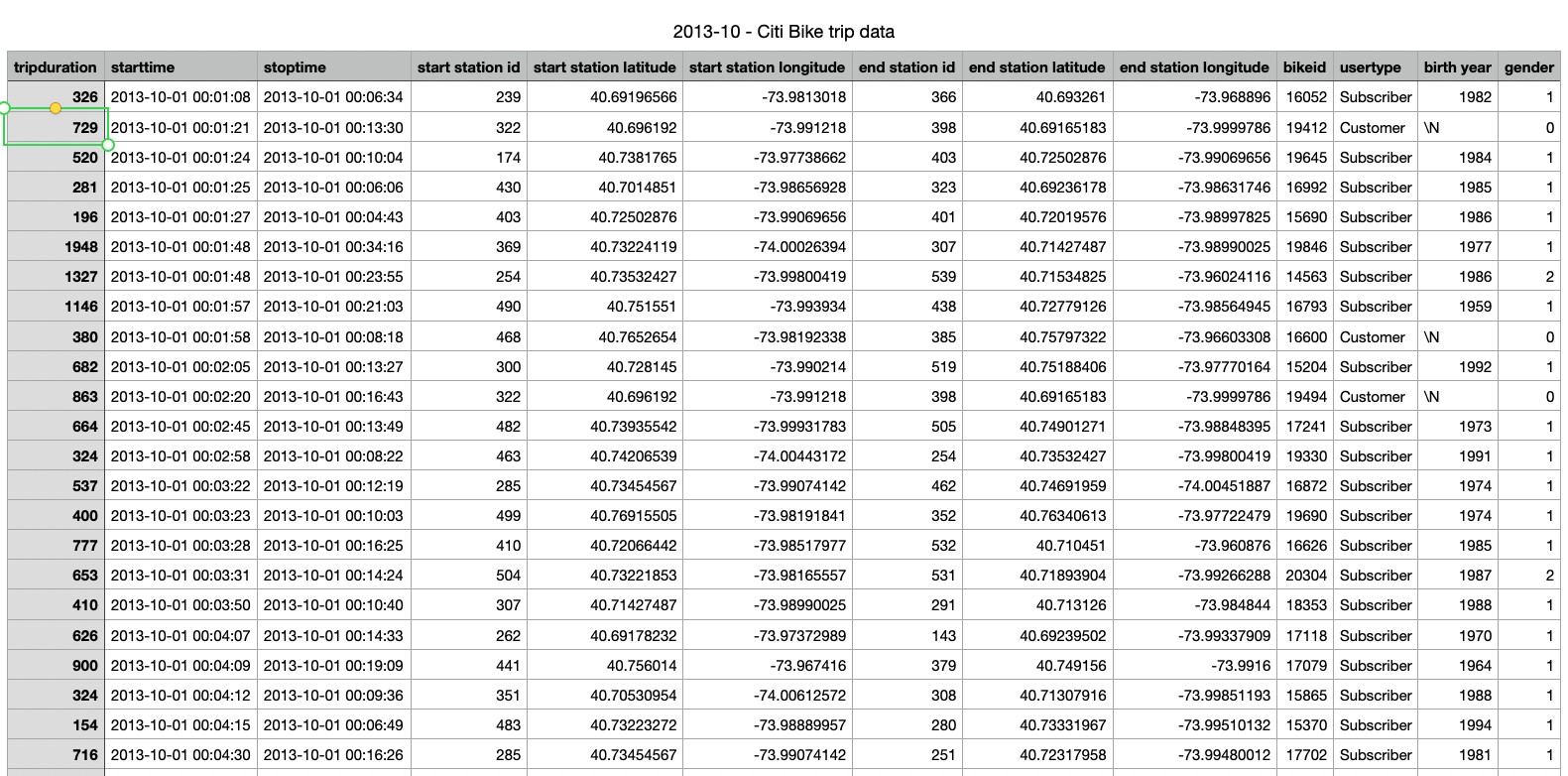
**Stop Station Longitude:** Longitud de la estación en la cual finalizo el viaje.

**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.



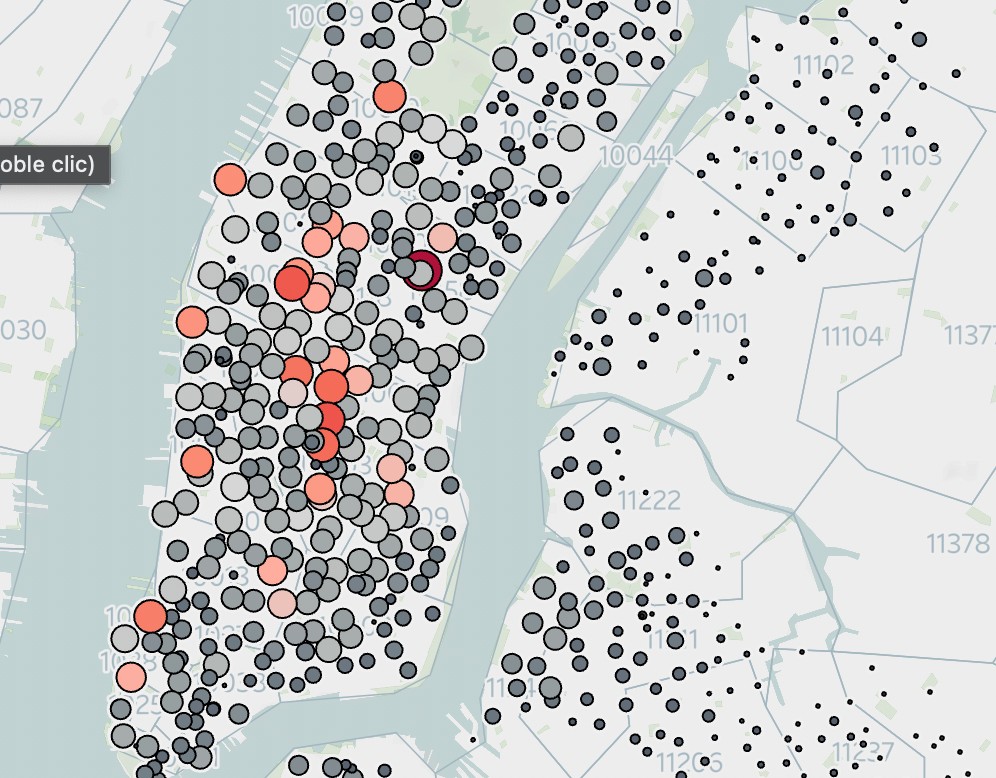
## Preguntas

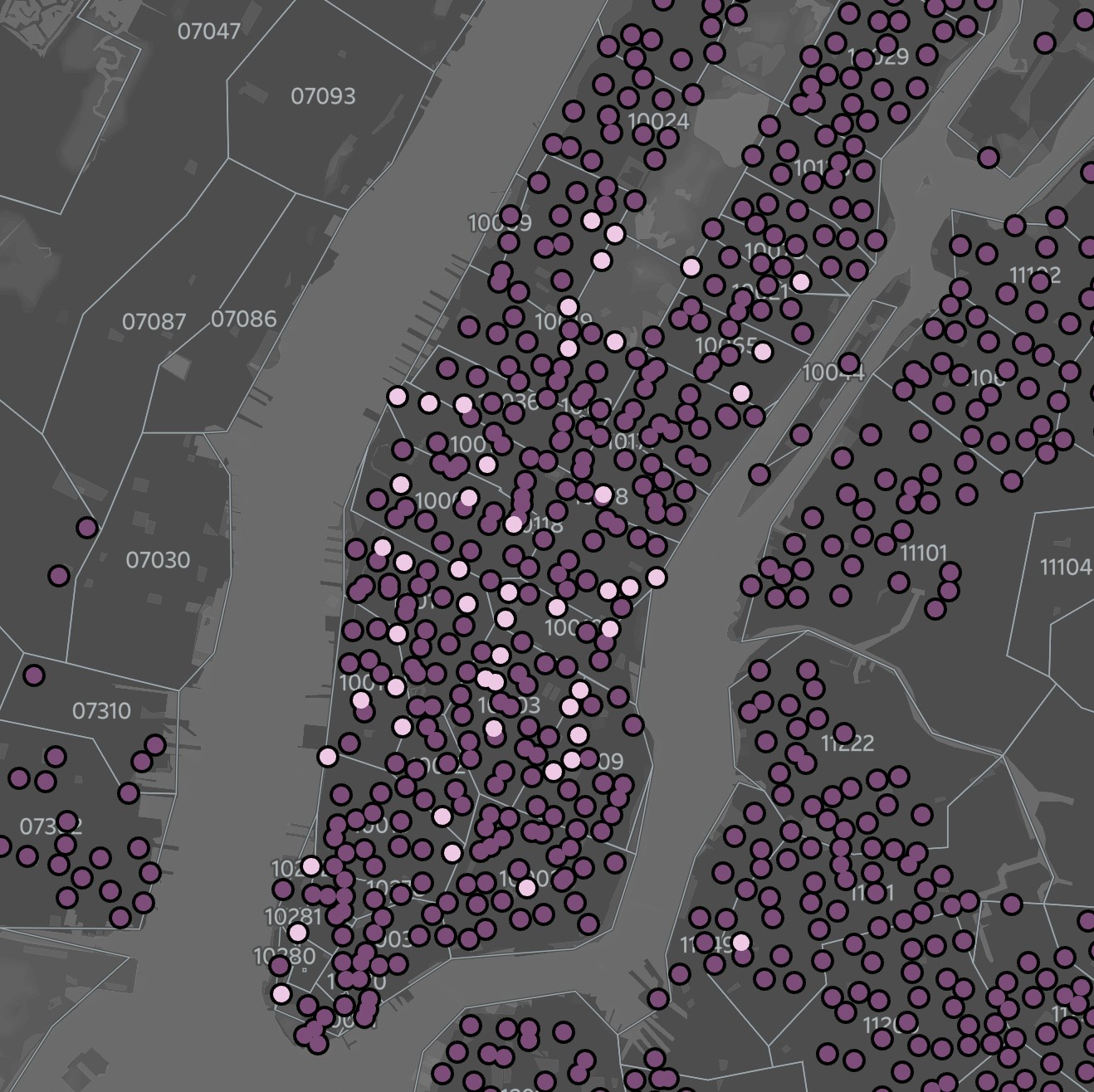
El alcalde de la ciudad de NYC, ha escuchado acerca de los cubos de información, sin embargo el CTO de la alcaldía considera que no es necesario la inversión en un sistema de data warehouse, pues considera que el programa es muy pequeño para que pueda generar un retorno adecuado para sostener la inversión.

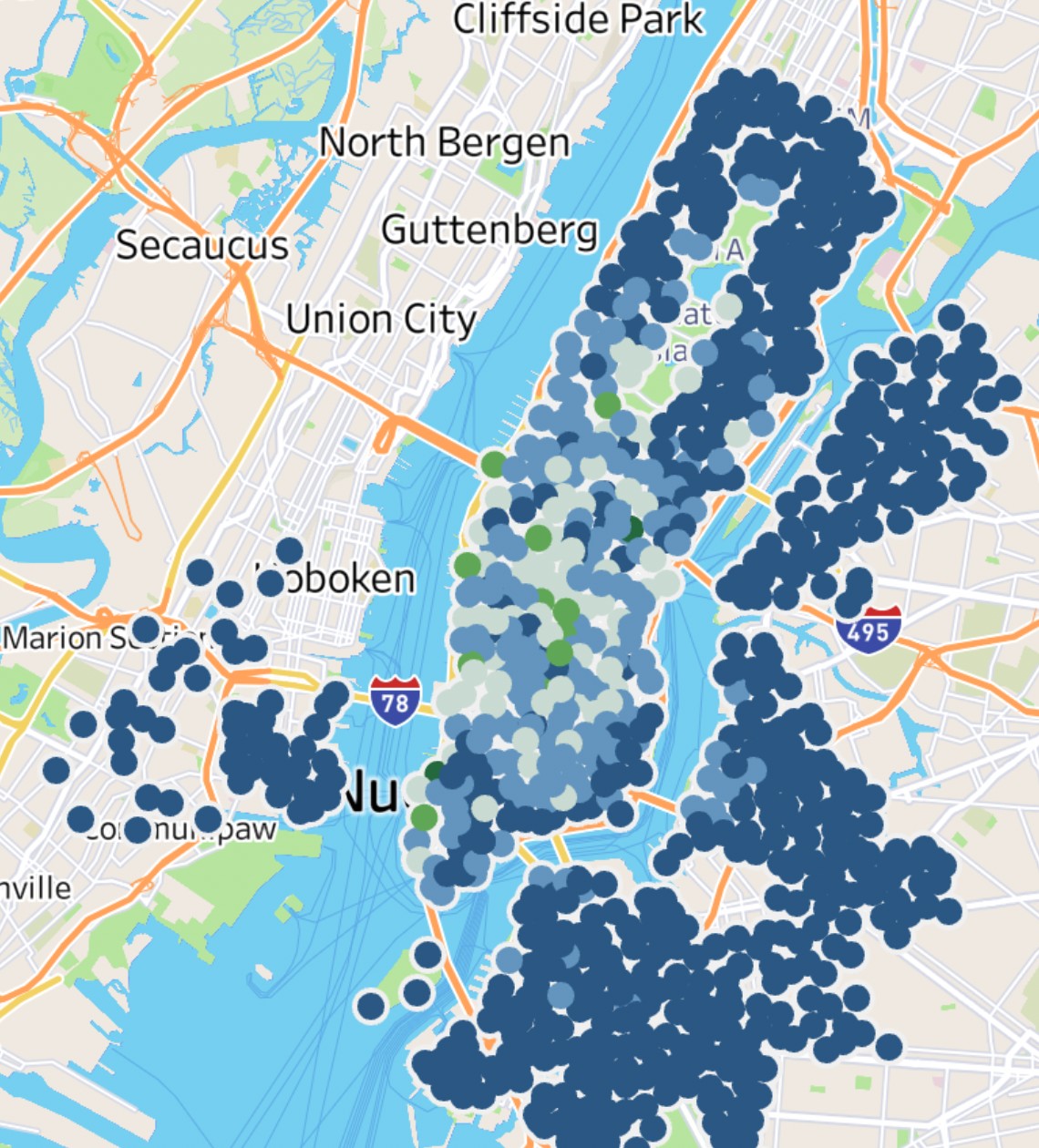
El alcalde de NYC, le ha pedido que se le presenten sus recomendaciones al respecto, usted como buen ingeniero en sistemas, sabe que lo mejor para poder generar insights relevantes para el negocio y poder generar de manera adecuada información para la toma de decisiones se necesita generar un cubo de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

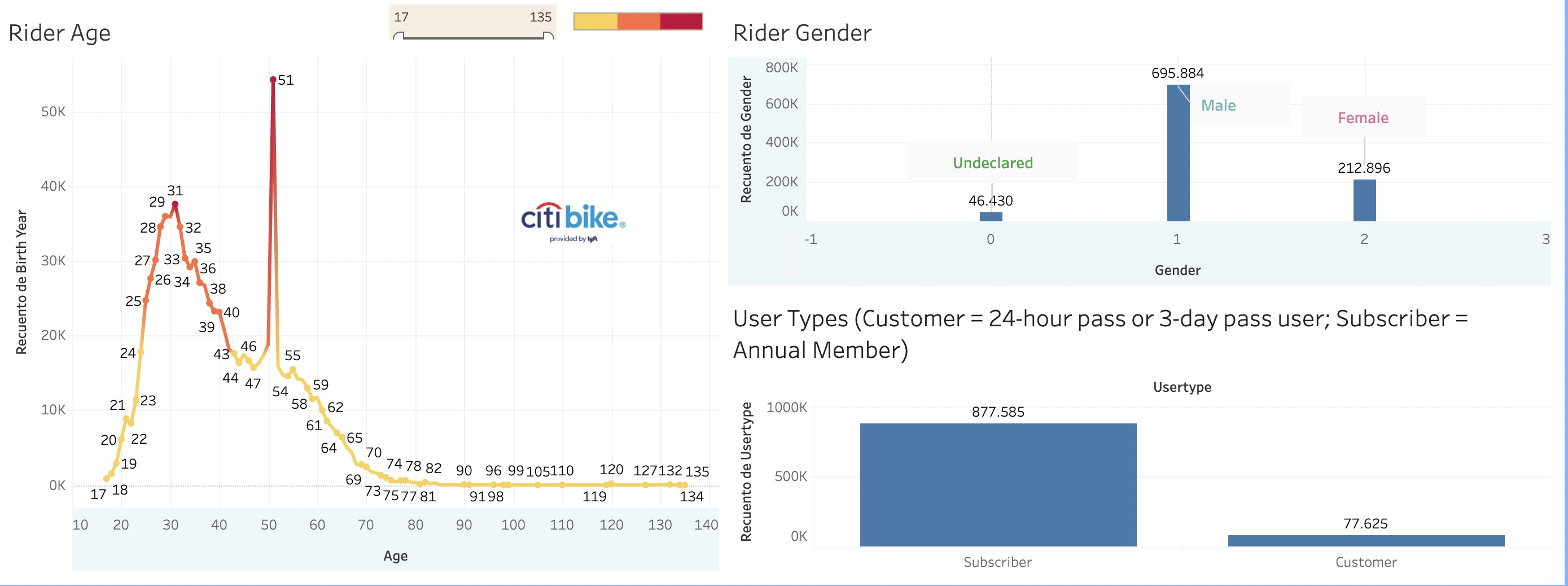
Propuesta 1

Propuesta 2

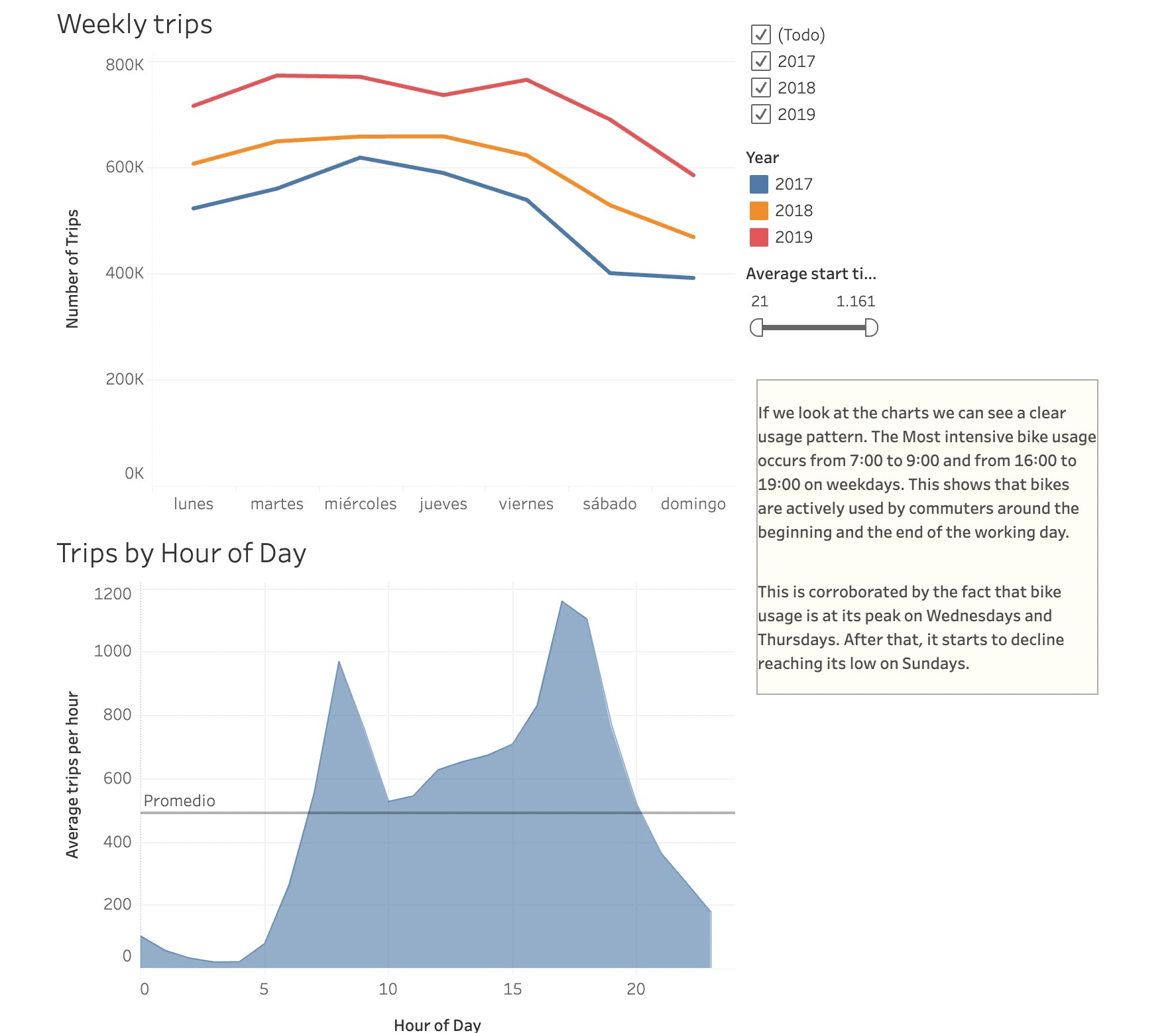
Propuesta 3

# Pregunta 6

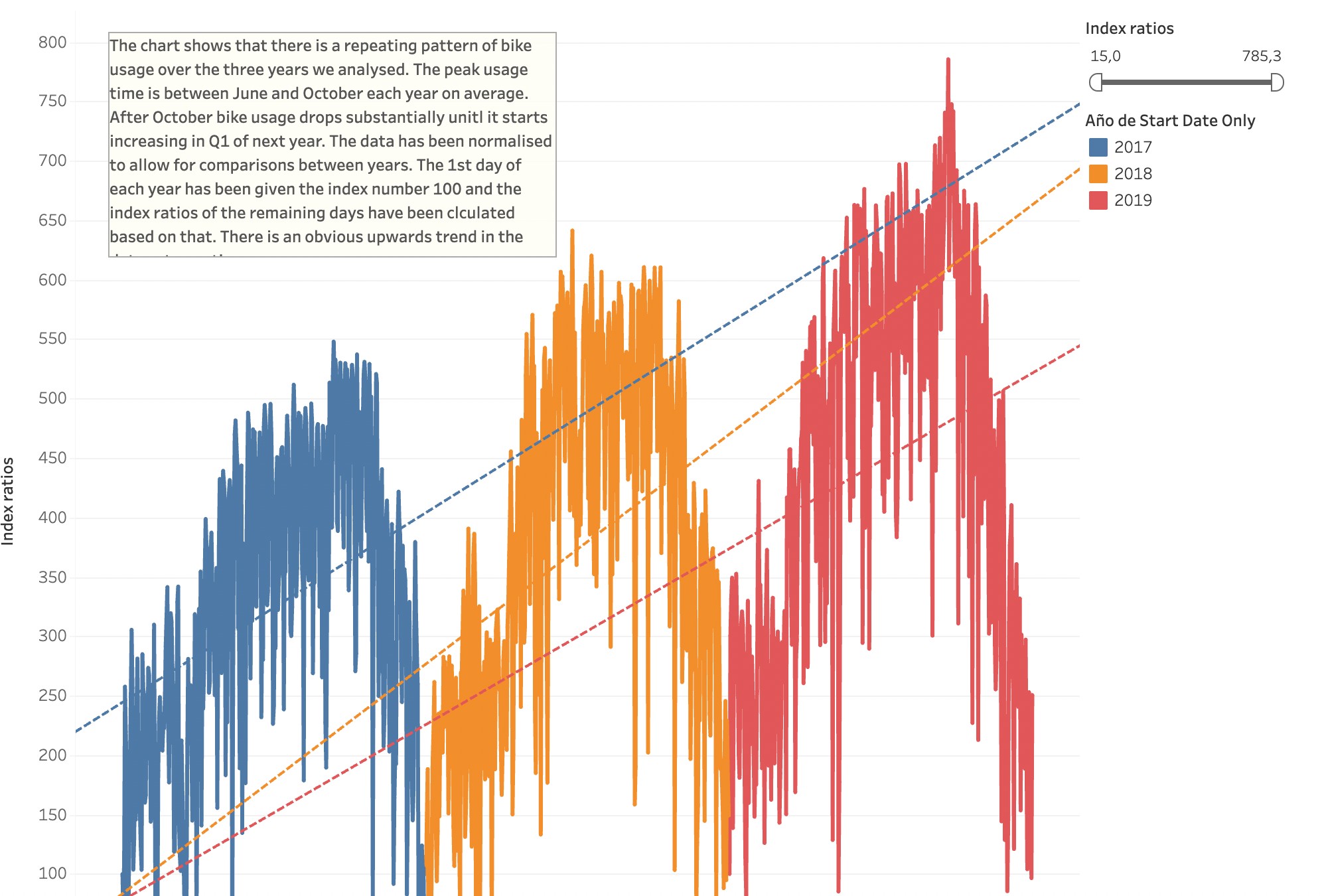
**Visualización 1**



# Visualización 2



**Visualización 3**



# Visualización 4

