

Análisis de datos del uso de las bicicletas de Cyclistic

1. Tarea empresarial *(Preguntar)*

En el presente informe se muestran los resultados obtenidos sobre las diferencias existentes entre los socios anuales y los ciclistas ocasionales, del servicio de bicicletas Cyclistic en Chicago. Para más información consultar la descripción del caso práctico **case_of_study.pdf**.

Como ciclista ocasional se entiende a las personas que toman los servicios de un viaje o el de uso durante un día, de la flota de bicicletas; en cambio, como socio anual se entiende a la gente que contrata la membresía de pago anual para el uso ilimitado durante un año de la flota de bicicletas.

El objetivo de este trabajo es comprender las diferencias existentes entre ambos tipos de usuarios a fin de desarrollar campañas que puedan ayudar a convertir a los clientes ocasionales en socios anuales. Mediante el análisis de los datos pertinentes se pretende comprender la manera más adecuada de aumentar el número de socios anuales provenientes del grupo de clientes ocasionales.

2. Fuentes de datos utilizadas *(Preparar)*

Para el análisis se han tomado los datos correspondientes a los últimos 12 meses (de octubre de 2021 hasta septiembre de 2022). Los datos fueron tomados de el siguiente repositorio público de la empresa, por lo que en un principio son fiables y actuales:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

En la licencia de uso de los datos expedida por Motivate International Inc. se pueden consultar los permisos concedidos y el uso que se puede hacer de los datos. Dicha licencia de uso se puede consultar en el siguiente enlace:

<https://ride.divvybikes.com/data-license-agreement>

La integridad de los datos y su estructura ha sido revisada por medio de RStudio para cerciorarme que todos los archivos (cada archivo representa un mes) son correctos y adecuados para responder las cuestiones planteadas al inicio del presente informe. Todos los archivos tienen las mismas columnas con los nombres iguales. Además, el formato de los datos es correcto para todas las columnas de todos los archivos. Dichas columnas son:

- **ride_id** (identificador de cada fila)
- **rideable_type** (tipo de Bicicleta utilizada)
- **started_at** (fecha y hora de inicio)
- **ended_at** (fecha y hora de finalización del uso)
- **start_station_name** (nombre de la estación de inicio)

- **start_station_id** (*id de la estación inicial*)
- **end_station_name** (*nombre de la estación final*)
- **end_station_id** (*id de la estación final*)
- **start_lat** (*latitud inicial*)
- **start_lng** (*longitud inicial*)
- **end_lat** (*latitud final*)
- **end_lng** (*longitud final*)
- **member_casual** (*cliente con membresía o cliente casual*)

La revisión de los datos ha revelado que en algunos de los archivos (meses) no se ha tomado el dato del nombre e id de la estación de inicio y finalización. En su defecto estos campos han sido cumplimentados con *NA*. En un primer momento esto no parece que vaya a afectar al análisis ya que no son datos muy relevantes para la pregunta que se intenta responder.

Estos pasos de preparación pueden ser consultados más a detalle en el *RMarkdown* presente en la subcarpeta **RMarkdowns** bajo el nombre de **processing_of_data.Rmd** o en la versión en HTML.

3. Limpiezas y manipulaciones de datos (*Procesar*)

En este caso la herramienta seleccionada para el procesamiento y el análisis de los datos es R, puesto que ya tengo experiencia con ella y me permite documentar los procesos por medio de *RMarkdowns* y hacer las visualizaciones. Los paquetes usados en este proceso ha sido *Tidyverse* y *Lubridate*. Todo el procesamiento de los datos se puede encontrar detallado en el documento **processing_of_data.Rm** junto a la preparación.

En el mencionado documento se muestra cómo se garantiza la integridad de los datos y el proceso seguido para limpiarlos. A la hora de limpiarlos ha habido que tener en cuenta que también se recoge el mantenimiento de las bicicletas. Esto se identifica con "HQ QR" por lo que se han podido eliminar dichas filas de forma sencilla. Por último, a fin de facilitar el análisis, se han juntado todos los archivos en un único archivo que contiene los datos correspondientes a un año natural.

4. Resumen del análisis (*Analizar*)

El análisis en detalle se puede revisar en el archivo *.html* o *RMarkdown* **análisis_of_data**.

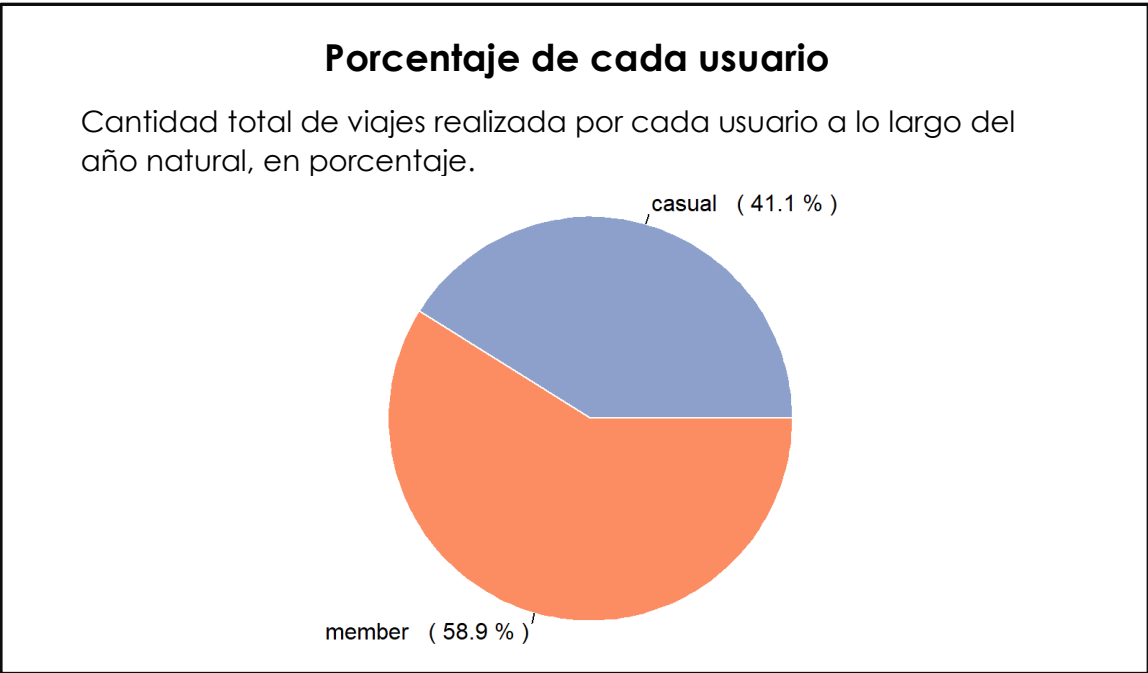
Para el análisis se parte del documento *.csv* ya procesado con anterioridad. Lo primero fue revisar el archivo y cambiar un par de errores relacionados con los días de la semana. Decidí centrarme en la longitud de los viajes y el número de

ellos que realiza cada tipo de usuario cada mes y cada día de la semana. Para ello se generó una tabla para hacerme una idea de las tendencias. Además, calcule que porcentaje del total de usuarios que representa cada tipo de usuario. Por último, para ver más claras las tendencias generé gráficos que resumen los datos obtenidos primeramente en tablas.

A nivel de mes generé una gráfica del número de viajes para comprobar la existencia de estacionalidad. A nivel de día de la semana miré además la longitud media de los viajes que se realizaban. Este último dato presentaba un error estándar estable a lo largo de todos los días de la semana para ambos tipos de usuarios.

5. Visualizaciones y hallazgos claves *(Compartir)*

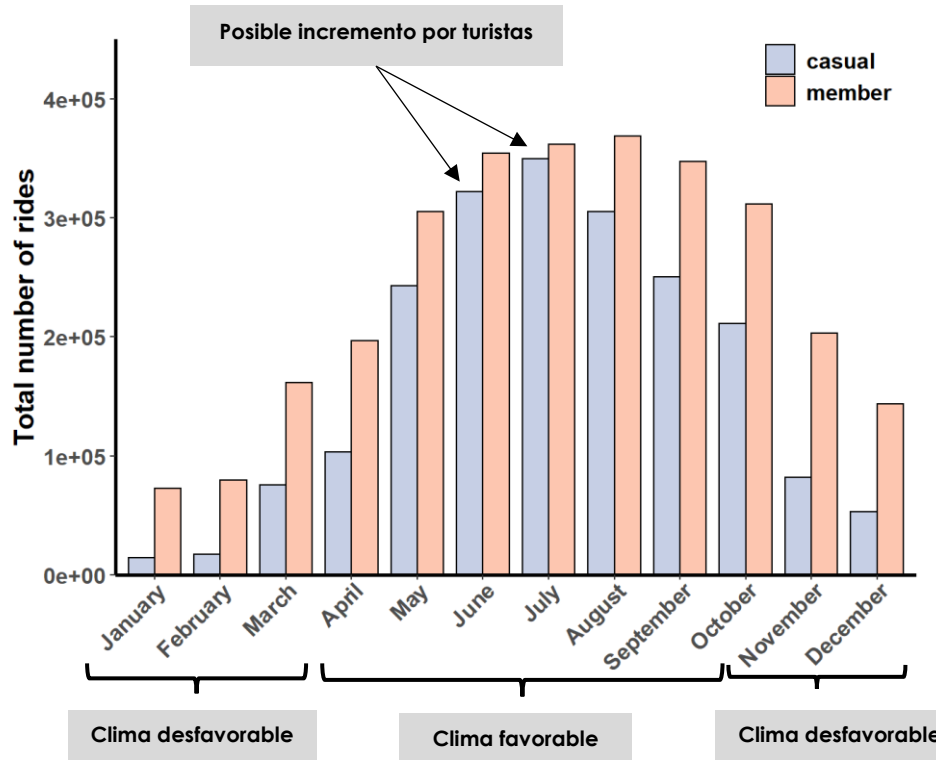
Los usuarios casuales representan un 41,1% de los viajes totales realizados en el año. Esto destaca la importancia de este análisis para aumentar la cantidad de membresías anuales.



Para tener una visión más general de los viajes realizados por cada usuario a lo largo del año, se mira el número de viajes por mes. Se puede observar una distribución desigual a lo largo del año, pero a la vez semejante para ambos tipos. El uso de las bicicletas está más extendido en los meses en el que el clima es más cálido y las precipitaciones son menores. En los meses de invierno se aprecia un pronunciado descenso del uso de las bicicletas. Además, el uso por parte de clientes usuales es mucho más acentuado en verano, lo que puede corresponder a la llegada de turistas. Por último, durante todo el año se ha mantenido un número superior de viajes por parte de los socios.

Uso de las bicicletas a lo largo del año

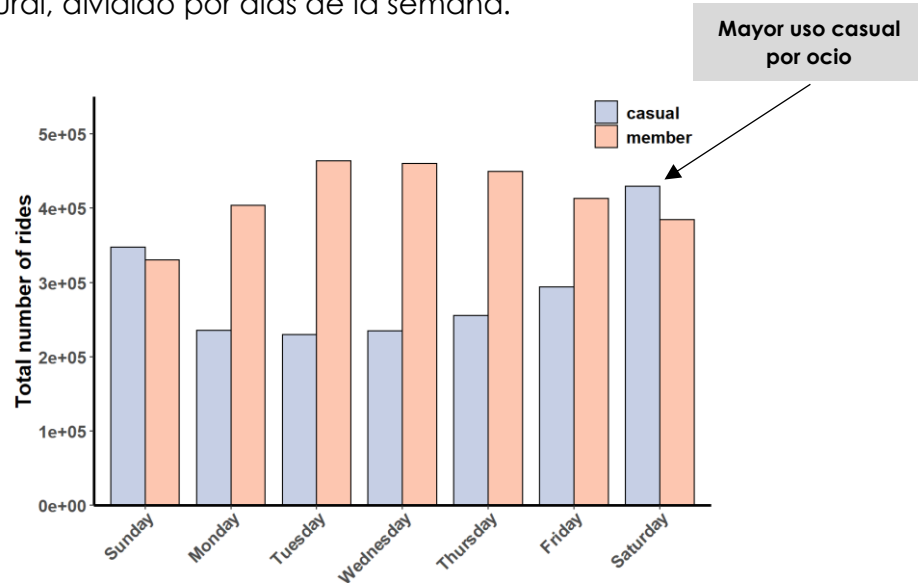
Cantidad total de viajes realizada por cada usuario a lo largo del año natural, mes a mes.



El siguiente gráfico representa los números de viajes totales a lo largo del año para cada día de la semana. El uso de las personas con la membresía es más mantenido a lo largo de la semana lo cual puede estar relacionado con el uso de la bicicleta para ir a trabajar y para ocio los fines de semana. En el caso de los usuarios casuales hay un claro descenso entre semana y, los fines de semana, su uso aumenta hasta superar a la gente con la membresía anual. Está claro que el uso de uso por parte de los usuarios casuales se limita más al ocio. En el caso de las personas con membresía podemos afirmar que posiblemente el uso mayoritario es para ir a trabajar y, los fines de semana como ocio.

Número total de viajes cada día de la semana

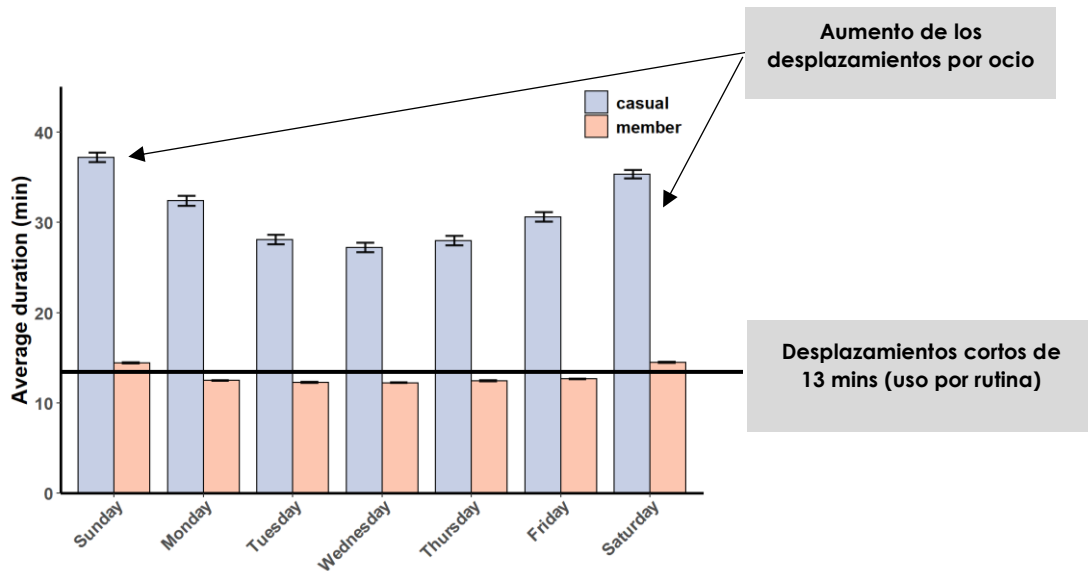
Cantidad total de viajes realizada por cada usuario a lo largo del año natural, dividido por días de la semana.



En lo referente a la duración de los trayectos, destaca el corto uso a lo largo de toda la semana que hacen las personas con membresía. Este confirma un uso más enfocado a desplazamientos rutinarios. Pese a ello esta duración aumenta algunos minutos los fines de semana, por posibles desplazamientos de ocio. En el caso de las personas usuales los desplazamientos son mucho más largos. La media para usuarios casuales es de 32 minutos, frente a los 13 minutos de los miembros anuales. Esto supone un uso superior del 40,6%. Al igual que con los miembros la duración de los trayectos es superior los fines de semana, pero hay mayor variación a lo largo de la semana. Esto responde a un uso más por ocio que por rutina. Al tratarse de datos obtenidos por medio de una media y teniendo en cuenta la gran cantidad de datos usados, he calculado el error estándar. Este error estándar es muy bajo y estable para todas las medidas, lo cual muestra que las conclusiones obtenidas de las tendencias observadas son fiables.

Duración media de los viajes cada día de la semana

Duración media de los viajes en minutos con error estándar por cada usuario a lo largo del año natural, dividido por días de la semana.



Para finalizar podemos concluir que los **clientes con membresía anual** se caracterizan por:

- Realizar actualmente el 58,9% de los viajes totales.
- Hacer mayor uso del servicio entre semana.
- Realizar viajes un 40,6% más cortos al tratarse de desplazamientos rutinarios.

Los **usuarios casuales** se caracterizan por:

- Un aumento pronunciado en los meses de verano (uso turístico).
- Hacer un mayor uso los fines de semana.
- Entre semana se observa una tendencia similar a los miembros, por lo que se deduce que hay gente que usan el servicio para desplazamientos rutinarios.
- Realizar desplazamientos de mayor duración.

Ambos tipos comparten la preferencia de uso en los meses de clima cálido y suave.

6. Recomendaciones basadas en el análisis *(Actuar)*

Ante los hallazgos mostrados en este informe, las tres recomendaciones que se muestran a continuación han sido ideadas por mi solo. En un caso real sería conveniente contar con la opinión de más analistas y de gente más experimentada en el sector, para que las recomendaciones finales sean lo más certeras posible.

- 1) Sería conveniente desarrollar una **nueva membresía anual** más barata y que permita el uso de las bicicletas solo los **viernes, sábado y domingos**. Esto con el fin de contentar a los usuarios que ven el servicio como un **medio de ocio** para los fines de semana.
- 2) Establecer nuevas **membresías** de duración de **6 meses** y de **9 meses**. Con esto se pretende captar a usuarios que no usan el servicio en **meses de invierno** y por ello no contratan la tarifa anual.
- 3) Hacer **comparativas** en la aplicación, a los **usuarios casuales más recurrentes**, de cuanto llevan gastado y el **ahorro** que les supone la contratación de la membresía anual.