# Instrucción Clara de la Tarea Empresarial

**Título:** Análisis de Tendencias entre Socios Anuales y Ciclistas Ocasionales

**Objetivo:**Identificar las diferencias en el comportamiento y las características entre los socios anuales y los ciclistas ocasionales. Este análisis permitirá diseñar estrategias de marketing efectivas para convertir a los ciclistas ocasionales en miembros anuales.

**Problema a Resolver:**Determinar las tendencias y patrones que diferencian a los ciclistas ocasionales de los socios anuales, así como los factores que podrían influir en la decisión de un ciclista ocasional para convertirse en socio anual.

**Metodología:**

1. **Limpieza de Datos:** Aplicar técnicas de limpieza de datos para asegurar la integridad y la calidad de los datos disponibles.
2. **Análisis Exploratorio:** Realizar un análisis exploratorio de datos (EDA) para identificar tendencias, patrones y correlaciones entre las variables de interés.
3. **Visualización de Datos:** Crear visualizaciones que resalten las diferencias y similitudes entre los dos grupos de usuarios, facilitando la comprensión de los hallazgos.
4. **Identificación de Factores Clave:** Determinar qué características o comportamientos son más comunes entre los socios anuales en comparación con los ciclistas ocasionales.

**Entregable:**

* Un informe detallado que incluya:
  + Análisis de las tendencias encontradas.
  + Visualizaciones que apoyen los hallazgos.
  + Recomendaciones para estrategias de marketing orientadas a convertir ciclistas ocasionales en miembros anuales.

**Interesados Clave:**

* **Lily Moreno** (Gerente y Directora de Marketing)
* **Equipo Ejecutivo de Cyclist**

# Preparación de los datos para el análisis

1. **Ubicación de los datos**: Los datos están almacenados en archivos CSV, lo que facilita su manipulación y análisis dentro de programas como R o Python.
2. **Organización de los datos**: Al ser un archivo CSV, la información se encuentra separada por comas, lo que permite identificar cada columna con los diferentes atributos del conjunto de datos, como la duración de los viajes, el tipo de usuario, las estaciones de inicio y fin, entre otros.
3. **Sesgo y credibilidad de los datos**:
   * Los datos proporcionados son de primera fuente y provienen directamente del sistema de bicicletas compartidas de Cyclistic.
   * Estos datos incluyen variables como datos nominales, continuos y discretos, lo que los hace confiables para el análisis.
   * No parece haber problemas evidentes de sesgo o manipulación con el fin de obtener ciertos resultados, cumpliendo con los principios de ROCCC (confiables, originales, integrales, actuales y citados).
4. **Autorización, privacidad, seguridad y accesibilidad**:
   * El análisis se realiza cumpliendo con las normativas estipuladas en la licencia bajo la que se obtuvieron los datos.
   * Se maneja la información de manera responsable, asegurando que la privacidad de los usuarios no sea comprometida, ya que no se incluyen datos personales identificables.
5. **Verificación de la integridad de los datos**:
   * Para garantizar que los datos sean íntegros, se sigue una lista de verificación que incluye:
     + Eliminación de duplicados.
     + Corrección de errores tipográficos.
     + Validación de consistencia en las categorías de datos (por ejemplo, tipos de usuarios o estaciones).
6. **Relevancia para el análisis**:
   * Estos datos son cruciales para identificar tendencias en el comportamiento de los usuarios miembros y casuales, lo que permitirá proponer estrategias para aumentar la cantidad de membresías anuales.
7. **Problemas identificados en los datos**:
   * Se ha detectado que en algunos casos no se registra la estación donde se dejan las bicicletas. Esto puede deberse a un posible extravío o a un fallo en el registro. Este detalle debe considerarse al momento de realizar ciertos análisis, como los que involucren rutas o estaciones específicas.

# Procesamiento:

1. **Iteración a través de una lista de archivos CSV**: El script recorre cada archivo CSV en una lista de nombres de archivos, procesándolos uno por uno.
   * Función utilizada: read.csv() para leer cada archivo CSV y convertirlo en un data frame en R.
2. **Limpieza y transformación de los datos**: Se limpian los datos y se aplican transformaciones clave como el cálculo de la duración del viaje y la extracción del día de la semana.
   * Función clean\_names() del paquete janitor para estandarizar los nombres de las columnas, haciéndolos más legibles y consistentes.
   * Función drop\_na() para eliminar filas con valores faltantes (NA), asegurando la calidad de los datos.
   * Función mutate() del paquete dplyr para:
     + Convertir las columnas de tiempo (started\_at y ended\_at) a un formato de fecha-hora con ymd\_hms(), una función del paquete lubridate.
     + Calcular la duración del viaje en segundos con difftime().
     + Extraer el día de la semana con wday(), también de lubridate, lo cual permite obtener el nombre del día (lunes, martes, etc.).
3. **Asignación dinámica de los data frames**: Cada data frame procesado se guarda en una nueva variable con un nombre dinámico basado en una lista de nombres, facilitando la organización de los datos.
   * Función assign() para asignar el nombre dinámico a cada data frame basado en el nombre del archivo que se está procesando.

### **Funciones principales empleadas:**

* read.csv(): Permite leer archivos CSV y convertirlos en data frames.
* clean\_names(): Limpia los nombres de las columnas de manera eficiente.
* drop\_na(): Elimina filas con valores faltantes (NA).
* mutate(): Permite modificar y crear nuevas columnas en un data frame.
* ymd\_hms(): Convierte texto en formato de fecha y hora.
* difftime(): Calcula la diferencia entre dos fechas u horas.
* wday(): Extrae el día de la semana de una fecha.
* assign(): Asigna dinámicamente un nombre a un objeto en R.

Este enfoque me permitió automatizar el procesamiento de grandes volúmenes de datos y realizar análisis descriptivos de forma rápida y precisa.

# **Análisis de datos**

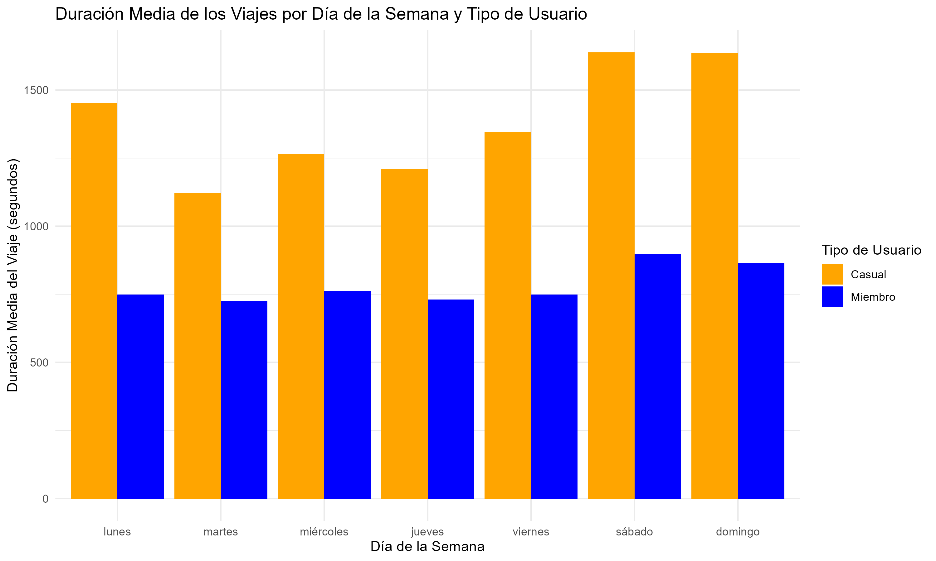
Con base en el análisis de datos, se identifican las siguientes tendencias clave que pueden ser aprovechadas para convertir a los ciclistas ocasionales en miembros anuales de Cyclistic:

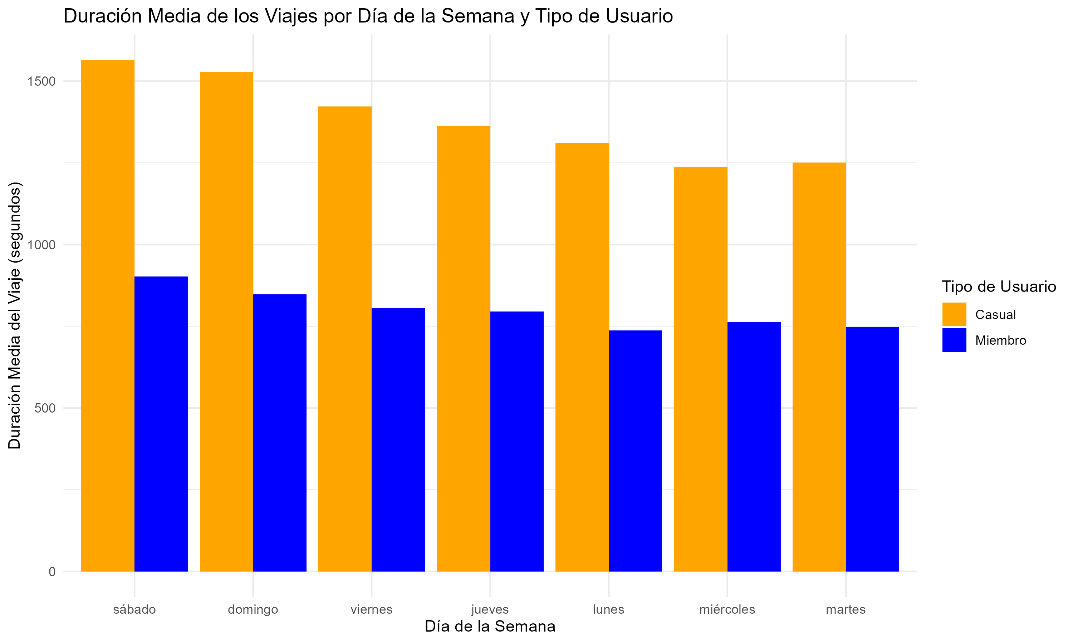
1. **Mayor número de viajes por parte de los miembros**: Los miembros realizan más viajes mensualmente en comparación con los usuarios casuales. Una estrategia efectiva podría ser ofrecer promociones basadas en la cantidad de viajes. Por ejemplo, tras alcanzar un número determinado de viajes en un mes (por ejemplo, 40), los usuarios podrían obtener viajes gratuitos o descuentos en la membresía. Esto incentivaría a los usuarios casuales a utilizar más el servicio y optar por una membresía anual.
2. **Duración promedio de los viajes más corta entre los miembros**: En general, los miembros tienden a realizar viajes más cortos que los usuarios casuales, especialmente los fines de semana. Esto podría indicar que los miembros utilizan las bicicletas para trayectos más cotidianos y rápidos, lo que se podría enfatizar en la comunicación de marketing. Se podría promocionar la membresía anual como una opción ideal para viajes regulares, rápidos y accesibles.
3. **Preferencia por bicicletas eléctricas**: Tanto los miembros como los ciclistas ocasionales muestran una preferencia por las bicicletas eléctricas en lugar de las tradicionales. Esto abre la puerta a ofrecer membresías exclusivas con beneficios adicionales para el uso de bicicletas eléctricas, como descuentos especiales o acceso prioritario, destacando las ventajas de estas bicicletas para viajes más cómodos y eficientes.
4. **Comportamiento de uso diario**: Los usuarios casuales tienden a concentrar sus viajes en picos específicos, mientras que los miembros muestran un patrón de uso más constante a lo largo del mes. Este comportamiento podría ser aprovechado para lanzar campañas dirigidas a usuarios casuales, ofreciendo incentivos en días de menor actividad o tarifas reducidas para viajes regulares entre semana. Esto podría estimular una mayor frecuencia de uso y, a su vez, fomentar la transición a una membresía anual.

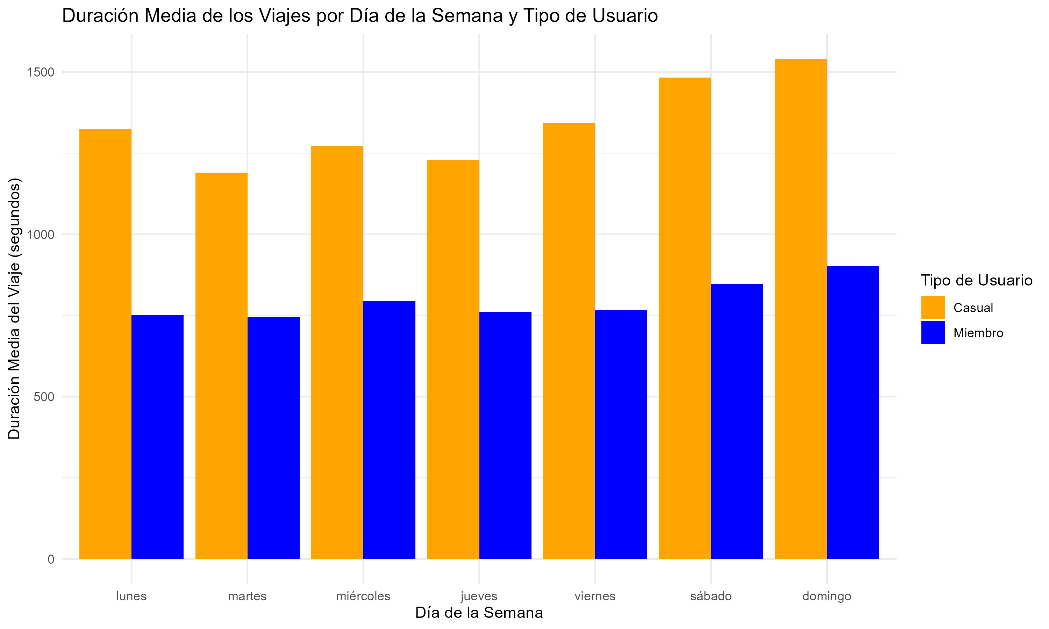
Implementar estas estrategias basadas en las tendencias observadas podría ser clave para maximizar el número de membresías anuales y fortalecer la relación entre Cyclistic y sus usuarios.

# Visualización

Agosto



Julio

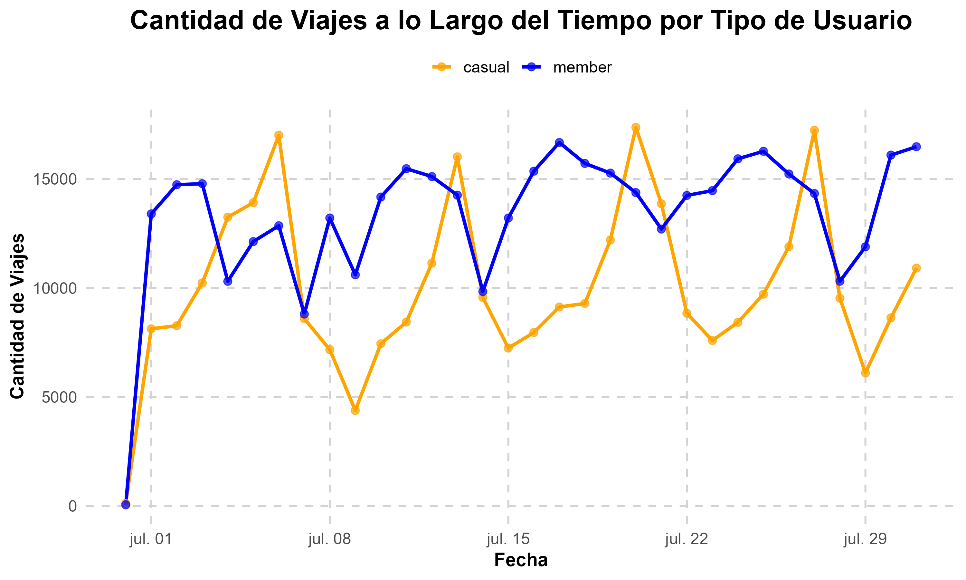
Junio

Agosto

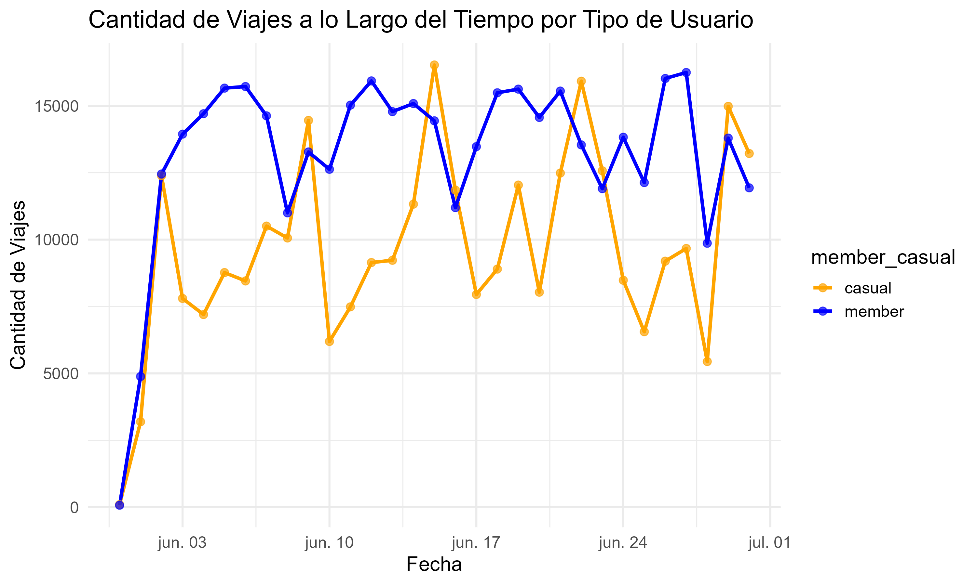
Gráfico

Descripción generada automáticamente

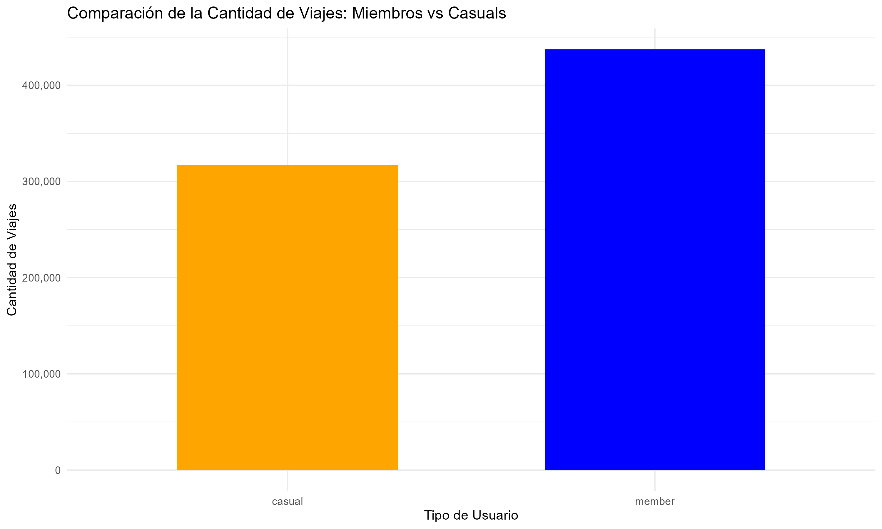
Julio



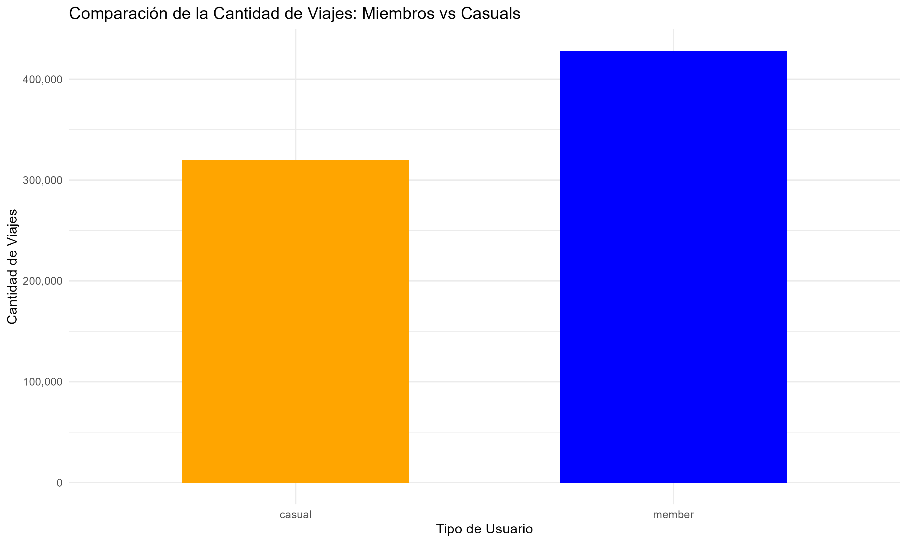
Junio



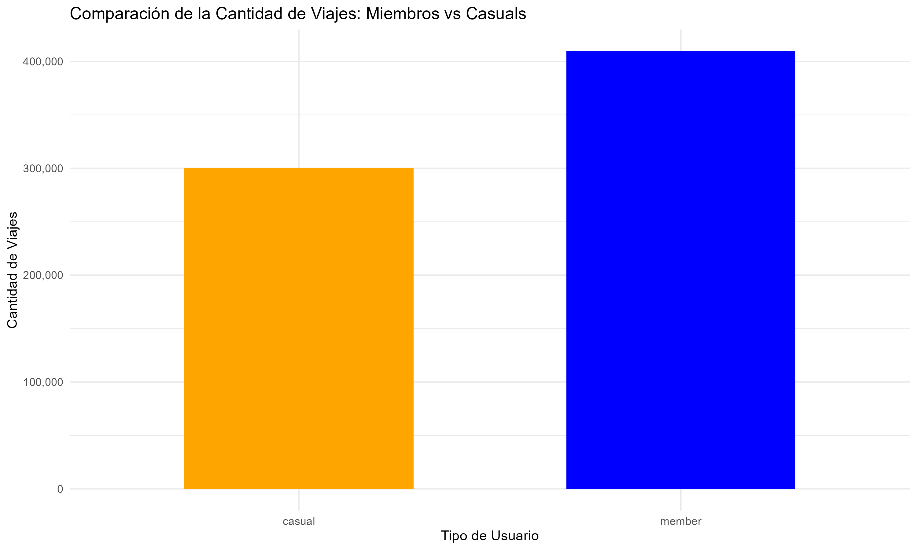
Agosto



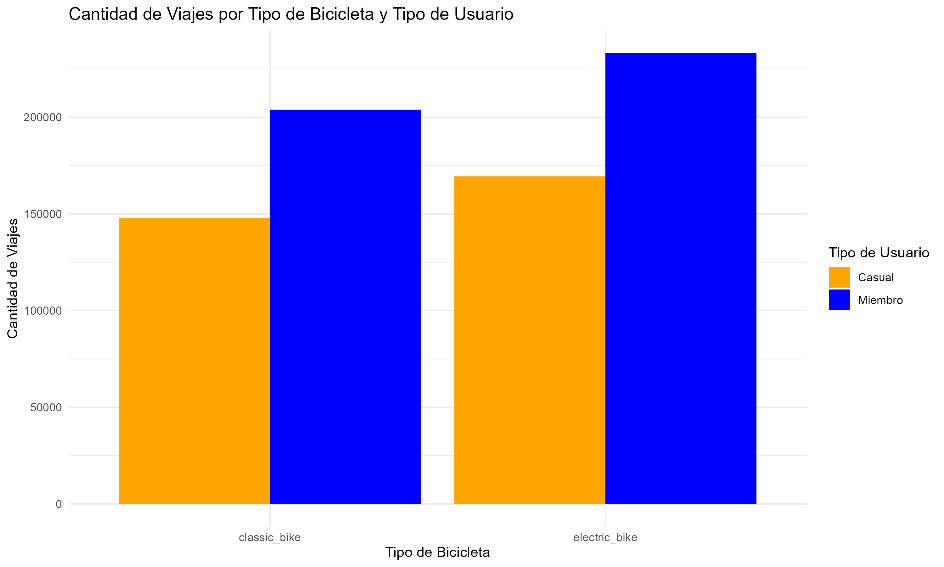
Julio



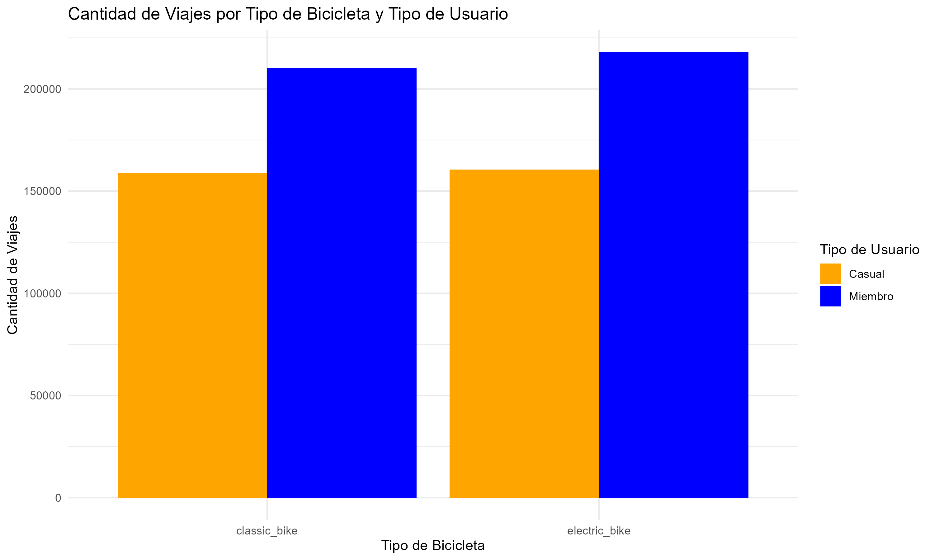
Junio



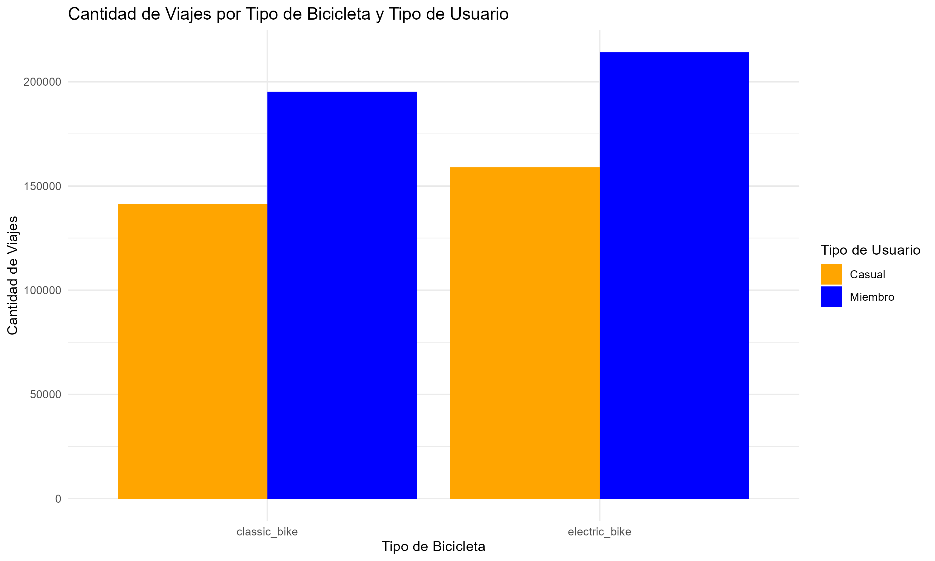
Agosto



Julio



Junio



# Actuar

para Cyclistic, las siguientes estrategias basadas en los datos analizados podrían convertir a los ciclistas ocasionales en miembros anuales:

1. **Promociones por frecuencia de uso**: Incentivar a los ciclistas casuales a alcanzar un número de viajes mensuales con descuentos, viajes gratuitos o una membresía a precio reducido tras completar 40 viajes.
2. **Viajes cortos y frecuentes**: Resaltar los beneficios de la membresía para quienes hacen varios viajes cortos, especialmente los fines de semana, destacando la conveniencia económica.
3. **Beneficios exclusivos para bicicletas eléctricas**: Ofrecer tarifas especiales o descuentos en bicicletas eléctricas para miembros anuales, aprovechando la preferencia de los usuarios por este tipo de bicicleta.
4. **Campañas en días pico**: Aprovechar los picos de uso de los usuarios casuales en fines de semana o eventos, destacando la estabilidad y conveniencia de una membresía para uso constante.

Estas estrategias buscan convertir el uso ocasional en membresías a largo plazo.