

First Line of a Multi-Line Title
Second Line of a Multi-Line Title

Stephen Turner*
Second Author†

julio 21, 2023

Executive Summary

Lorem ipsum dolor sit amet, semper suscipit sea at. Dico scriptorem nec at, ex qui virtute dolores oportere. Duis tantas ponderum ut has. Est saepe mandamus salutatus et, id sed semper detracto moderatius, ei sit aperiam voluptua. Per esse justo fierent eu, duo quando tempor ut. At elit doming possim vim, ut dolorem appetere nec.

Contents

Preguntar	2
Instrucción de la tarea empresarial	2
Preparar	2
Descripción de las fuentes de los datos utilizadas	2
Procesar	3
Limpieza de los datos y su manipulacion de los datos	3
Analizar y Visualización	4
Resumen del análisis	4

*<____@____.com>
†<____@____.com>

¿En qué se diferencian los socios anuales y los ciclistas ocasionales con respecto al uso de las bicicletas de Cyclistic?

Preguntar

Instrucción de la tarea empresarial

- Tarea Empresarial
 1. *Problema a resolver.*
 - Maximizar el número de miembros anuales en Cyclistic. Actualmente, Los miembros anuales son más rentables que los ciclistas ocasionales, por lo que es crucial convertir a los miembros ocasionales a miembros anuales y así impulsar el crecimiento y el éxito futuro de la empresa.
 2. *Decisiones Empresariales*
 - Con la ayuda del equipo de análisis de datos, y el conocimiento que tenemos, nos permitirá realizar un análisis exhaustivo del uso de las modalidades de las bicicletas de Cyclistic. Al comprender las diferencias en términos de patrones de uso, duración de los viajes, frecuencia de uso, horas de mayor demanda y de menor demanda, y la preferencia de los usuarios, y así podremos identificar oportunidades claves para diseñar una estrategia de marketing efectiva que convierta los ciclistas ocasionales en miembros anuales.

Al terminar el análisis, podremos proporcionar recomendaciones basadas en el análisis, respaldadas por visualizaciones profesionales de los datos, podemos ayudar al equipo de Marketing (Dir. Lily Moreno) a tomar decisiones informadas y desarrollar tácticas de marketing específicas para atraer y retener a los ciclistas ocasionales, gracias a los análisis y recomendaciones ya que estarán respaldadas por datos concretos, lo que aumentará la confianza del equipo ejecutivo de Cyclistic para aprobar y llevar a cabo las estrategias recomendadas.

Preparar

Descripción de las fuentes de los datos utilizadas

Para este análisis, los datos utilizados son datos históricos de los viajes realizados por meses, son proporcionados por la compañía Motivate. Los datos se encuentran en un archivo descargable que contiene los datos recolectados por mes. La fuente de los datos es confiable, ya que se proporciona por una empresa ficticia para fines prácticos para la realización de este caso. Es importante tener en cuenta que los datos no contienen información real de los usuarios.

En términos de autorización, privacidad, seguridad y accesibilidad, se ha asegurado que los datos se utilicen de acuerdo con las políticas y regulaciones de protección de datos aplicables. No se utilizará información de identificación personal ni se realizará ningún intento de conectar los datos con información personal.

Para verificar la integridad de los datos, se pueden realizar diferentes acciones, como comprobar que los archivos descargados estén completos y sin errores, revisar la consistencia de los campos y valores, y comparar los datos con información adicional o registros previos para detectar discrepancias o anomalías.

El análisis de los datos históricos de los viajes de Cyclistic ayudará a identificar tendencias y patrones en el uso de las bicicletas. Esto permitirá comprender mejor los diferentes tipos de clientes y su comportamiento, lo que puede ser útil para tomar decisiones de negocio informadas y optimizar los servicios ofrecidos por Cyclistic. Los datos cuentan con varias celdas vacías, por lo tanto, representa un gran riesgo para tomar cualquier decisión basada en los datos por lo cual es necesario filtrar los datos y ordenar de tal manera en disminuir los sesgos y los errores que pueda tener al momento de realizar el análisis de los datos.

Procesar

Liempieza de los datos y su manipulacion de los datos

Primeramente cargaremos los datos para poder visualizar los datos y comprobar si existe algún tipo de error (anteriormente se pudo observar atrevas de Microsof Excel algunos errores), los datos corresponde a los últimos 12 meses por lo cual se utilizaran los datos correspondiente a los meses que van desde Julio del 2022 a Junio del 2023.

```
## [1] "E:\\Cyclistic/202207-divvy-tripdata.csv"
## [2] "E:\\Cyclistic/202208-divvy-tripdata.csv"
## [3] "E:\\Cyclistic/202209-divvy-publictripdata.csv"
## [4] "E:\\Cyclistic/202210-divvy-tripdata.csv"
## [5] "E:\\Cyclistic/202211-divvy-tripdata.csv"
## [6] "E:\\Cyclistic/202212-divvy-tripdata.csv"
## [7] "E:\\Cyclistic/202301-divvy-tripdata.csv"
## [8] "E:\\Cyclistic/202302-divvy-tripdata.csv"
## [9] "E:\\Cyclistic/202303-divvy-tripdata.csv"
## [10] "E:\\Cyclistic/202304-divvy-tripdata.csv"
## [11] "E:\\Cyclistic/202305-divvy-tripdata.csv"
## [12] "E:\\Cyclistic/202306-divvy-tripdata.csv"
```

Para limpiar los datos es necesario utilizar varias funciones entre ellas **lapply** (permite aplicar una función sobre la lista para poder leer los archivos .csv) y el comando **read_csv** para poder visualizar los datos en un “marco de datos”, se le asignara una nueva columna llamada “month” y se le asigna al nuevo archivo

Se utiliza la función `stringr::str_replace` el cual nos ayudara a simplificar el nombre de la columna “month” y que solo se muestre la fecha del mes. Calculamos la diferencias de tiempo entre el tiempo de inicio del viaje y el final del viaje en bicicleta y se almacenara en una nueva columna, posteriormente este dato que obtenemos se transformara a un dígito y se formateara de nuevo pero en formato de horas:minutos:segundos

Se combinara todos los registros en un solo archivo “marco datos” verticalmente y se le asignara a la variable de datos_limpios, observaremos todos los nombres de las columnas.

Filtraremos los datos para eliminar las filas donde la diferencia de sea igual a -1 o 0 y se convertirán las columnas `member_casual` y `rideable_type` en factores, esta conversión nos ayudara para que estas columnas se traten de manera categórica y no por numéricamente, final mente para observar que los cambios han sido efectivos se muestra un resumen con-siso de los datos.

```
## Rows: 4,408,750
## Columns: 19
## $ ride_id          <chr> "954144C2F67B1932", "292E027607D218B6", "5776585258~
## $ rideable_type    <fct> classic_bike, classic_bike, classic_bike, classic_b~
## $ started_at       <dtm> 2022-07-05 08:12:47, 2022-07-26 12:53:38, 2022-07--
## $ ended_at         <dtm> 2022-07-05 08:24:32, 2022-07-26 12:55:31, 2022-07--
## $ start_station_name <chr> "Ashland Ave & Blackhawk St", "Buckingham Fountain ~
## $ start_station_id  <chr> "13224", "15541", "15541", "15541", "TA1307000117",~
## $ end_station_name  <chr> "Kingsbury St & Kinzie St", "Michigan Ave & 8th St"~
## $ end_station_id    <chr> "KA1503000043", "623", "623", "TA1307000164", "TA13~
## $ start_lat         <dbl> 41.90707, 41.86962, 41.86962, 41.86962, 41.89147, 4~
## $ start_lng         <dbl> -87.66725, -87.62398, -87.62398, -87.62398, -87.626~
## $ end_lat           <dbl> 41.88918, 41.87277, 41.87277, 41.79526, 41.93625, 4~
## $ end_lng           <dbl> -87.63851, -87.62398, -87.62398, -87.59647, -87.652~
## $ member_casual     <fct> member, casual, casual, casual, member, member, mem~
## $ month             <chr> "202207", "202207", "202207", "202207", "202207", "~
## $ diff              <drtn> 705 secs, 113 secs, 463 secs, 3509 secs, 1578 secs~
## $ diff_double       <dbl> 705, 113, 463, 3509, 1578, 523, 689, 1853, 333, 687~
```

Table 1: Resumen de los datos

Mes	Estación inicial	Estación final	inicioT	Fin T	Tipo	Duration	Dia
202207	Ashland Ave & Blackhawk St	Kingsbury St & Kinzie St	2022-07-05 08:12:47	2022-07-05 08:24:32	member	00:11:45	Martes
202207	Buckingham Fountain (Temp)	Michigan Ave & 8th St	2022-07-26 12:53:38	2022-07-26 12:55:31	casual	00:01:53	Martes
202207	Buckingham Fountain (Temp)	Michigan Ave & 8th St	2022-07-03 13:58:49	2022-07-03 14:06:32	casual	00:07:43	Domingo
202207	Buckingham Fountain (Temp)	Woodlawn Ave & 55th St	2022-07-31 17:44:21	2022-07-31 18:42:50	casual	00:58:29	Domingo
202207	Wabash Ave & Grand Ave	Sheffield Ave & Wellington Ave	2022-07-13 19:49:06	2022-07-13 20:15:24	member	00:26:18	Miércoles
202207	Desplaines St & Randolph St	Clinton St & Roosevelt Rd	2022-07-01 17:04:35	2022-07-01 17:13:18	member	00:08:43	Viernes
202207	Marquette Ave & 89th St	East End Ave & 87th St	2022-07-18 18:11:01	2022-07-18 18:22:30	member	00:11:29	Lunes
202207	Wabash Ave & Grand Ave	Dearborn Pkwy & Delaware Pl	2022-07-28 20:38:18	2022-07-28 21:09:11	casual	00:30:53	Jueves
202207	Wabash Ave & Grand Ave	Dearborn Pkwy & Delaware Pl	2022-07-10 22:55:59	2022-07-10 23:01:32	member	00:05:33	Domingo
202207	Ashland Ave & Blackhawk St	Orleans St & Merchandise Mart Plaza	2022-07-10 09:35:58	2022-07-10 09:47:25	member	00:11:27	Domingo

```
## $ diff_format      <time> 00:11:45, 00:01:53, 00:07:43, 00:58:29, 00:26:18, ~
## $ day_of_week      <int> 2, 2, 7, 7, 3, 5, 1, 4, 7, 7, 5, 6, 2, 3, 1, 7, 4, ~
## $ dia_semana       <chr> "Martes", "Martes", "Domingo", "Domingo", "Miércoles"
```

Analizar y Visualización

Resumen del análisis

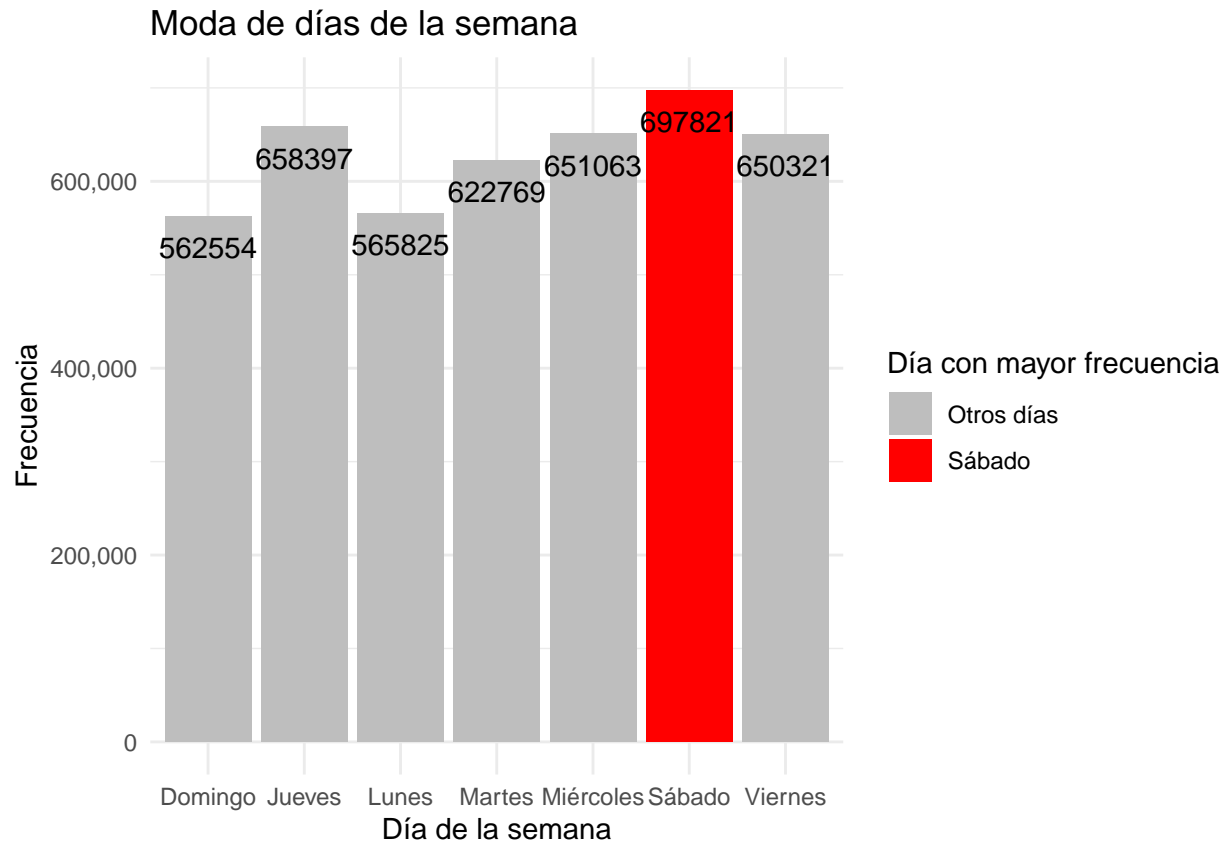
Visualizamos los datos ya limpios y los que tienen una mayor importancia para el análisis y eliminamos datos nulos y así podemos realizar un análisis adecuado.

La tabla muestra los datos limpios y solo se muestra una parte de los datos para tener un contexto de los datos.

Los datos representan varias tendencias en diferentes etapas de uso de las bicicletas, continuación se muestra un resumen de las variables de la diferencias de tiempo, es decir los valores promedios del uso de las bicicletas sin importar el tipo de usuario

Moda	Media	Max
Sábado	16M 3.17107547490832S	22d 5H 55M 27S

Para comprender la diferencia con los demás días de la semana en todo el año se puede observar en la siguiente gráfica



Calculamos la media de las diferencias de la duración del viaje, el máximo tiempo de recorrido, y calculamos la moda del día en que se utiliza las bicicletas por el tipo de usuario, en todo el año y filtramos los datos del tipo de bicicletas mas usadas por mes.

Como podemos observar en las tablas creadas se puede apreciar que en el caso de usuario casual tiene que el mes (202208) con mas viajes con una media de duración de 00:13 y el uso máximo es de 24:05 y el día que mas se usa es Martes con un total de viajes entoldo el año es de 335201

En el caso de usuario con la membresia tiene el mes (202207) con mas viajes con una media de duración de 00:25 y el uso máximo es de 533:55 y el día que mas se usa es Sábado con un total de viajes en todo el año es de 311649

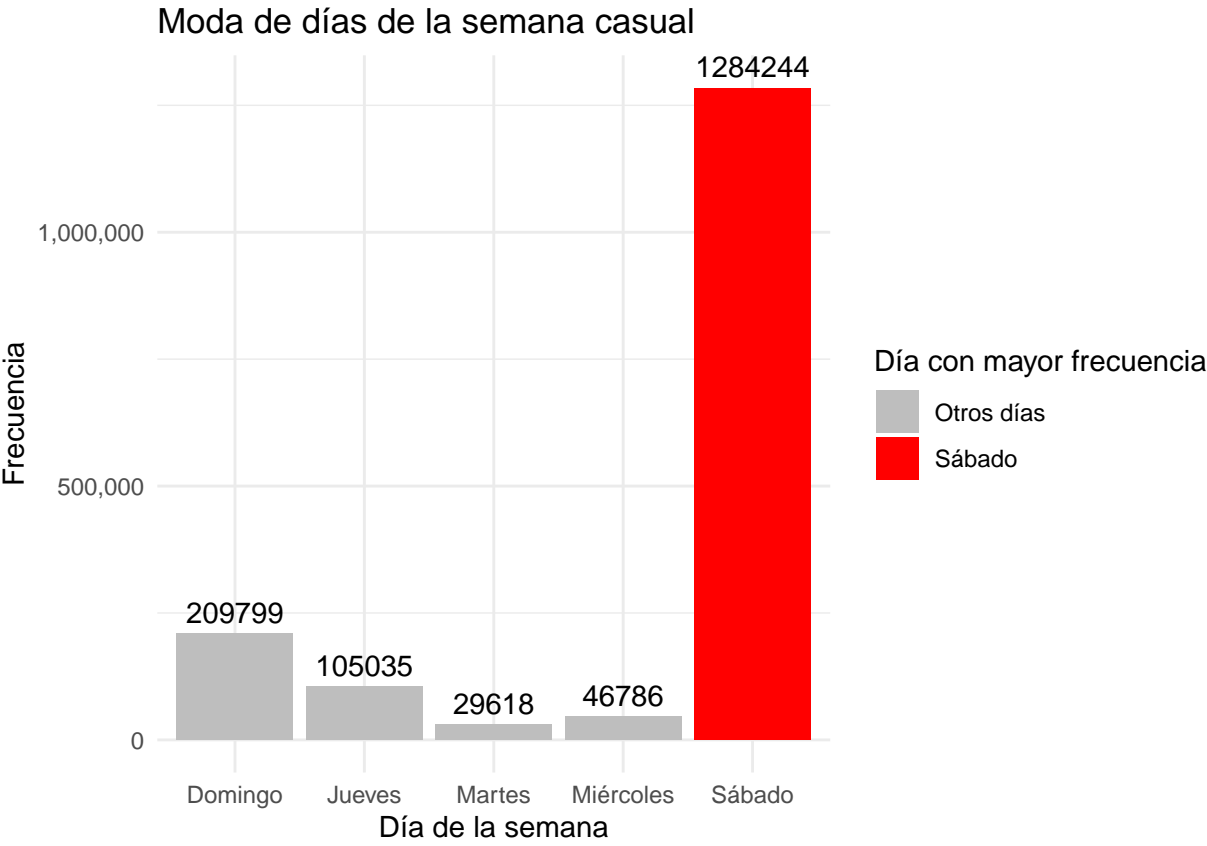
Visualizaremos el día con mas demanda para ambos tipos de usuario.

Table 2: Resumen de los datos tipo Casual

month	moda_dia_semana	media_diff_hh_mm	max_diff_hh_mm	total_viajes
202207	Sábado	16M	22d 5H 55M 27S	311649
		3.17107547490832S		
202208	Sábado	16M	22d 5H 55M 27S	270074
		3.17107547490832S		
202209	Sábado	16M	22d 5H 55M 27S	220905
		3.17107547490832S		
202210	Sábado	16M	22d 5H 55M 27S	151312
		3.17107547490832S		
202211	Jueves	16M	22d 5H 55M 27S	73533
		3.17107547490832S		
202212	Jueves	16M	22d 5H 55M 27S	31502
		3.17107547490832S		
202301	Martes	16M	22d 5H 55M 27S	29618
		3.17107547490832S		
202302	Domingo	16M	22d 5H 55M 27S	32774
		3.17107547490832S		
202303	Miércoles	16M	22d 5H 55M 27S	46786
		3.17107547490832S		
202304	Sábado	16M	22d 5H 55M 27S	110526
		3.17107547490832S		
202305	Domingo	16M	22d 5H 55M 27S	177025
		3.17107547490832S		
202306	Sábado	16M	22d 5H 55M 27S	219778
		3.17107547490832S		

Table 3: Resumen de los datos tipo Member

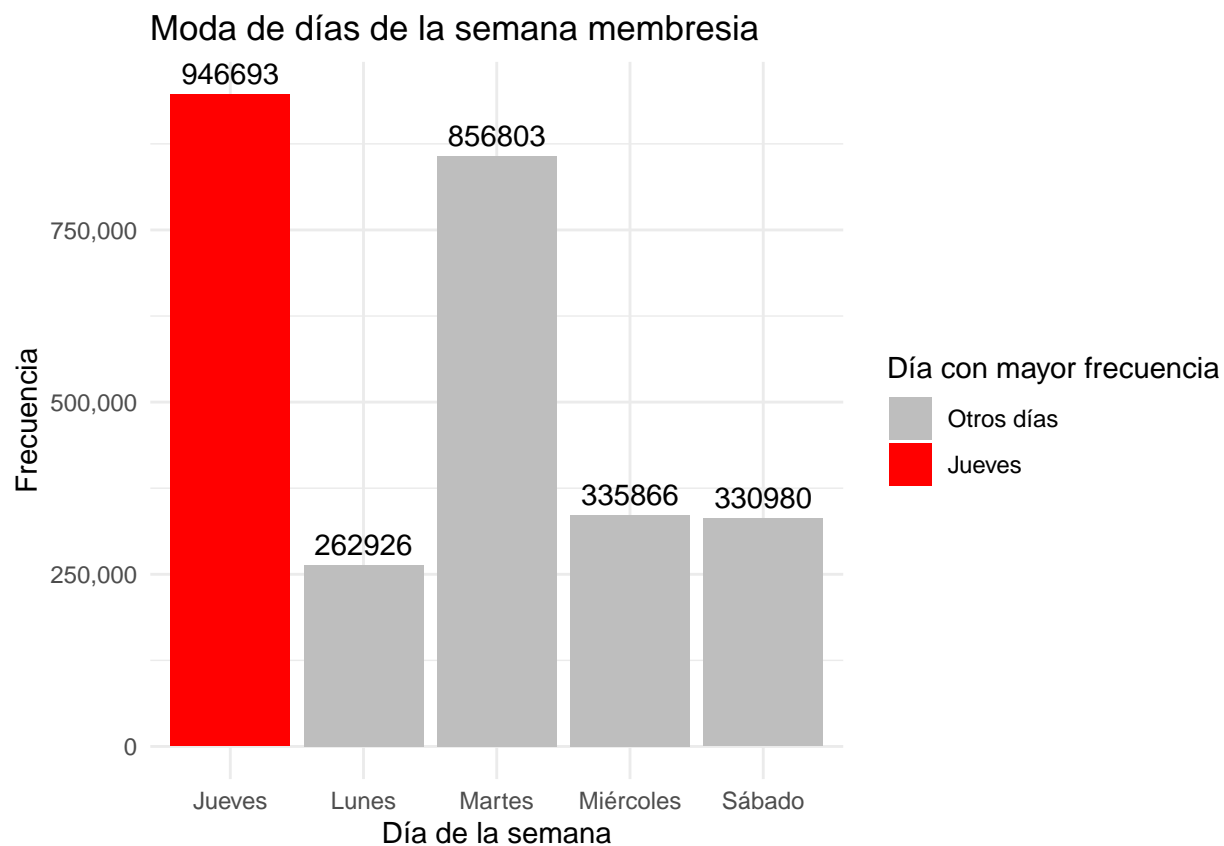
month	moda_dia_semana	media_diff_hh_mm	max_diff_hh_mm	total_viajes
202207	Sábado	16M	22d 5H 55M 27S	330980
		3.17107547490832S		
202208	Martes	16M	22d 5H 55M 27S	335201
		3.17107547490832S		
202209	Jueves	16M	22d 5H 55M 27S	314214
		3.17107547490832S		
202210	Lunes	16M	22d 5H 55M 27S	262926
		3.17107547490832S		
202211	Miércoles	16M	22d 5H 55M 27S	182219
		3.17107547490832S		
202212	Jueves	16M	22d 5H 55M 27S	103891
		3.17107547490832S		
202301	Martes	16M	22d 5H 55M 27S	118662
		3.17107547490832S		
202302	Martes	16M	22d 5H 55M 27S	116778
		3.17107547490832S		
202303	Miércoles	16M	22d 5H 55M 27S	153647
		3.17107547490832S		
202304	Jueves	16M	22d 5H 55M 27S	213647
		3.17107547490832S		
202305	Martes	16M	22d 5H 55M 27S	286162
		3.17107547490832S		
202306	Jueves	16M	22d 5H 55M 27S	314941
		3.17107547490832S		



En este gráfico se puede observar que el tipo de usuario casual lo ocupa el servicio de bicicletas los días sábados y los domingo.

Table 4: Resumen de los datosr

member_casual	promedio_ride_length
casual	22M 29.8890056712034S
member	12M 6.11458078754129S

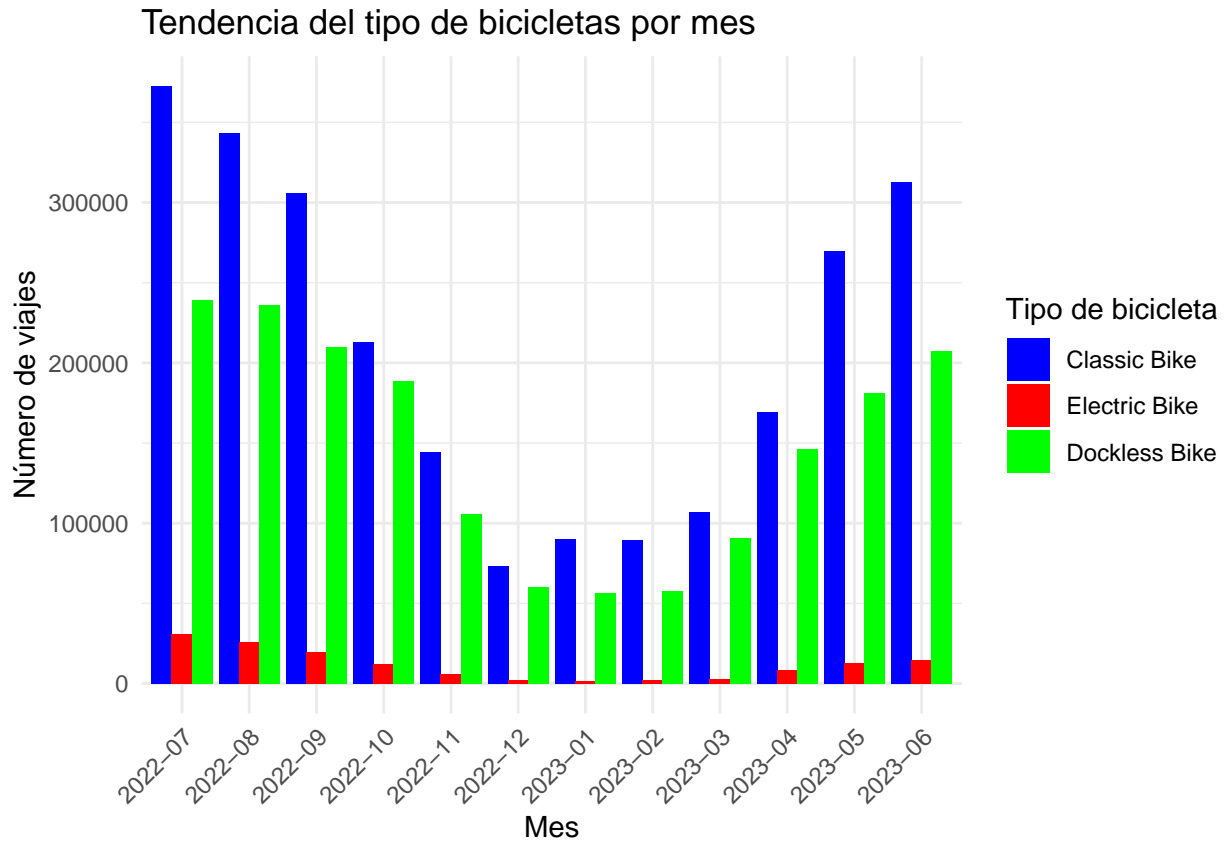


En este gráfico se puede observar que el tipo de usuario membresía lo ocupa el servicio de bicicletas los días jueves y los martes.

Continuación se mostrara varias tablas con los promedios por su categoría y calcularemos el muermo de viajes por usuario y crearemos una grafica para visualizar el tipo de bicicleta mas utilizada.

Table 5: Resumen de los datos

dia_semana	member_casual	promedio_ride_length
Domingo	casual	25M 45.8291169326012S
Domingo	member	13M 25.5021379023167S
Jueves	casual	19M 41.623982824575S
Jueves	member	11M 37.793263223182S
Lunes	casual	22M 30.0616363098509S
Lunes	member	11M 32.2388153648801S
Martes	casual	19M 57.7472282414324S
Martes	member	11M 31.9939186893401S
Miércoles	casual	19M 22.3706985731853S
Miércoles	member	11M 35.9445790586173S
Sábado	casual	25M 33.4341738458456S
Sábado	member	13M 42.2206003484191S
Viernes	casual	21M 42.1370178717257S
Viernes	member	11M 57.1684962981875S



Como podemos observar en este gráfico muestra la tendencia en que los usuarios sin importar de su estatus utilizan las bicicletas, en este caso el uso de las bicicletas mas utilizadas son las clásicas, y las menos utilizadas son las electric bike.

Obtendremos el mes con mas uso

El mes con mas uso es el mes de agosto del 2022, y el usuario con mas uso es el usuario que tiene la membresia.

Crearemos varias gráficas en donde se mostrara los viajes totales por el tipo de usuario, después el gráfico del

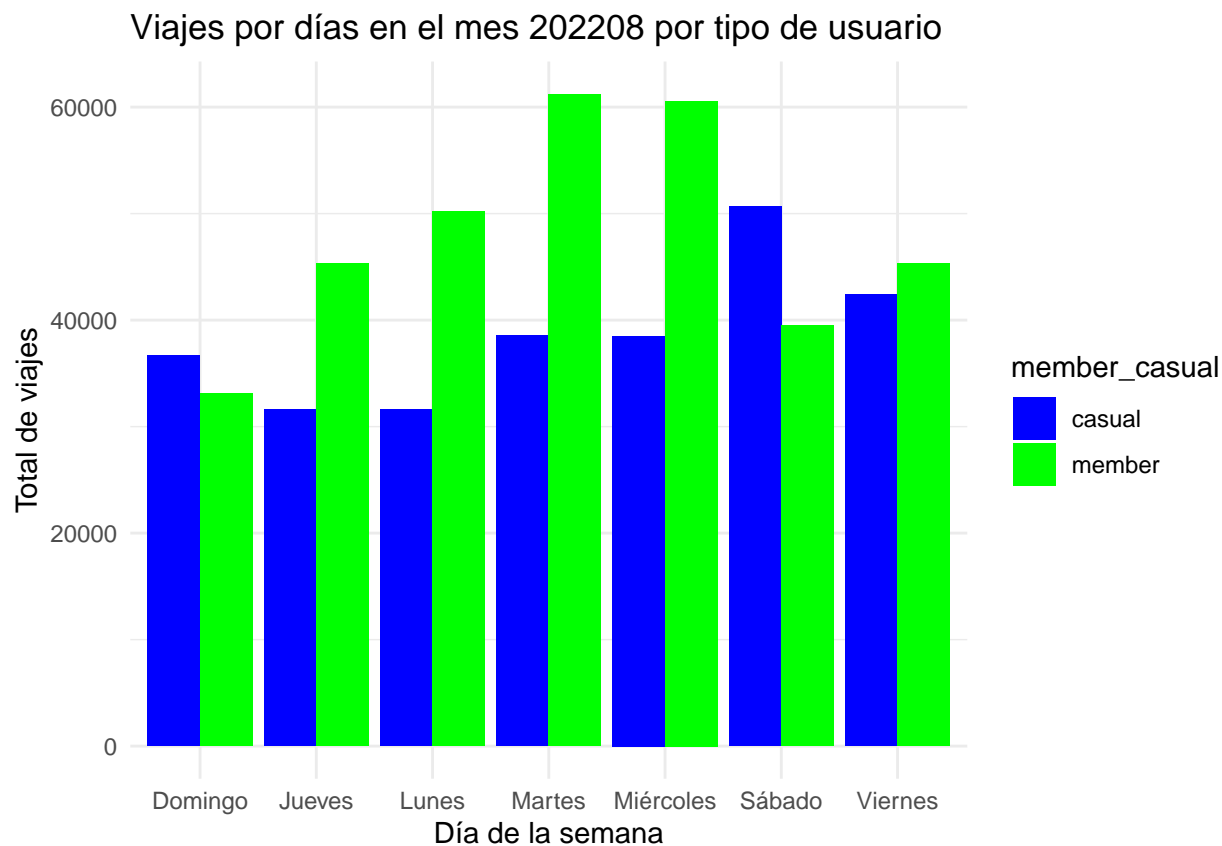
Table 6: Mese con mas uso

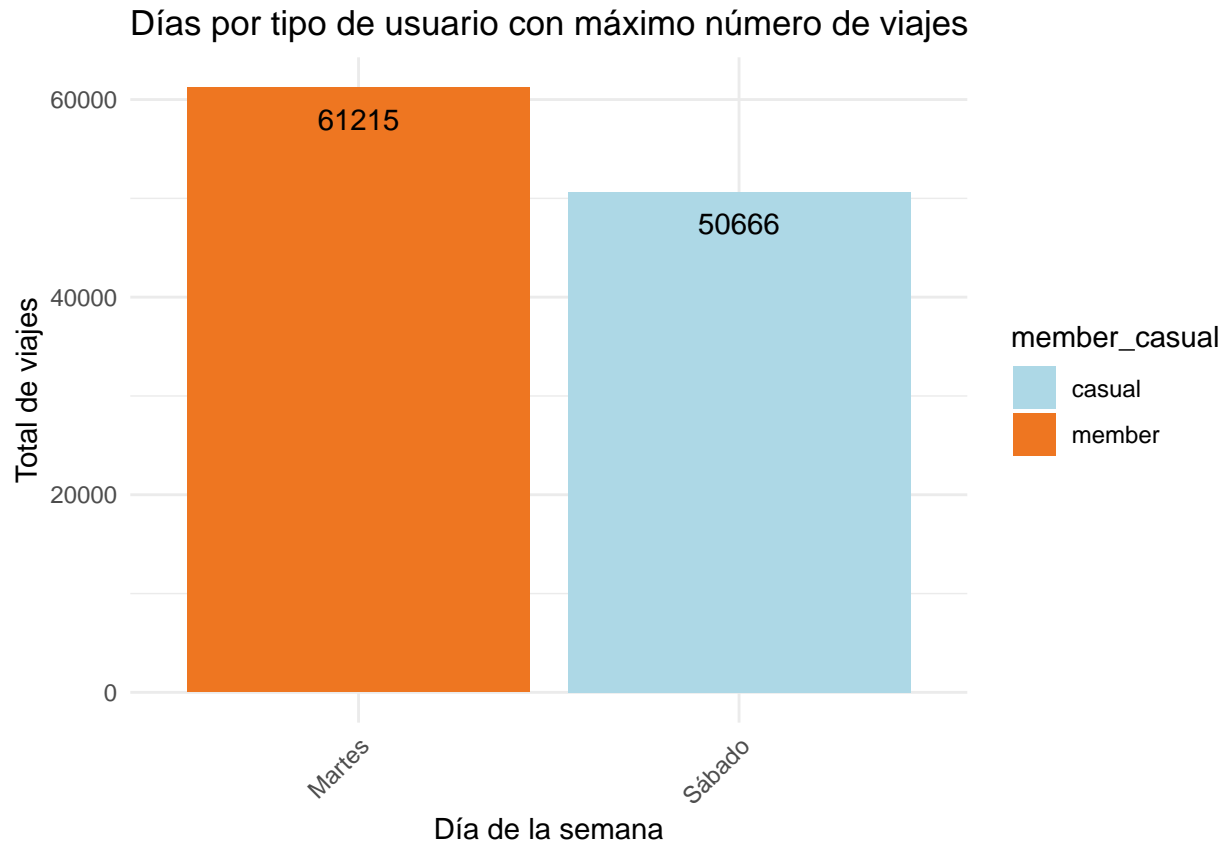
mes_max_viajes	tipo_usuario_max_viajes
202208	member

mes para corroborar si es correcto los datos anteriores y así visualizar el usuario con mas uso, y el ultimo gráfico se especificara el mes con mas uso de tanto el tipo de usuario casual y el de tipo memebresia.

Como podemos observar el tipo de usuario que cuenta con una memebresía utiliza las bicicletas con mas frecuencia, mientras que el tipo de usuario casual, la su ocupación de la bicicletas es baja a comparación del otro tipo de usuario.

Para tener mas información respecto a este mes se realizo el siguiente gráfico



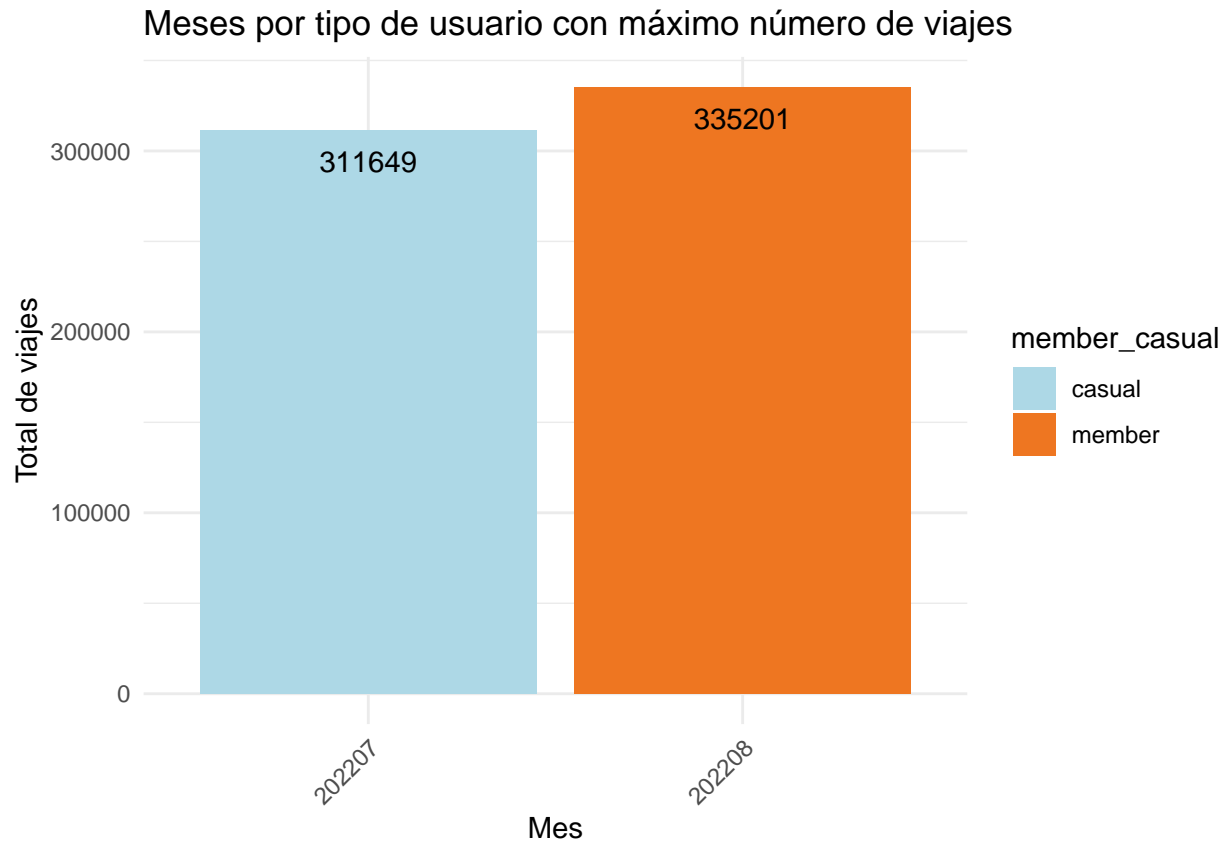


En este gráfico se muestra por días en ese mes y se puede observar que el día con mas demanda fueron dos el día sábado el usuario casual tubo una mayor importación y el día martes la tubo el tipo de usuario con membrsía.

Los dos mesa mas importante fueron el mes de julio y de agosto con una mayor demanda a comparación de los demás meses.

Table 7: Estaciones de inicio con mas demanda

start_station_name	num_usuarios
Streeter Dr & Grand Ave	65891
DuSable Lake Shore Dr & Monroe St	37939
Michigan Ave & Oak St	36035
DuSable Lake Shore Dr & North Blvd	35091



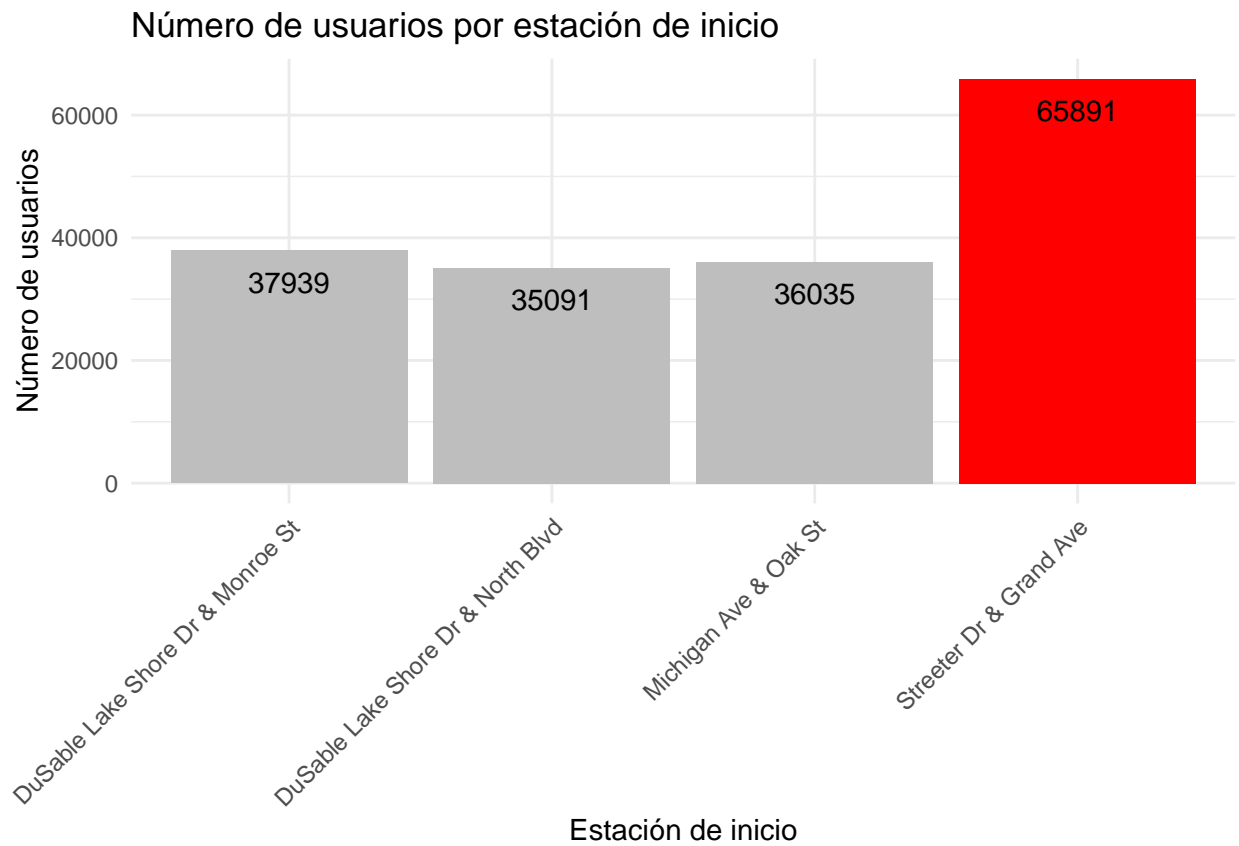
Con los datos de la estaciones podremos observar en que estación los usuarios comenzaron su recorrido y así podremos ver la estaciones con mas demanda y también la estaciones de destino con mas demanda.

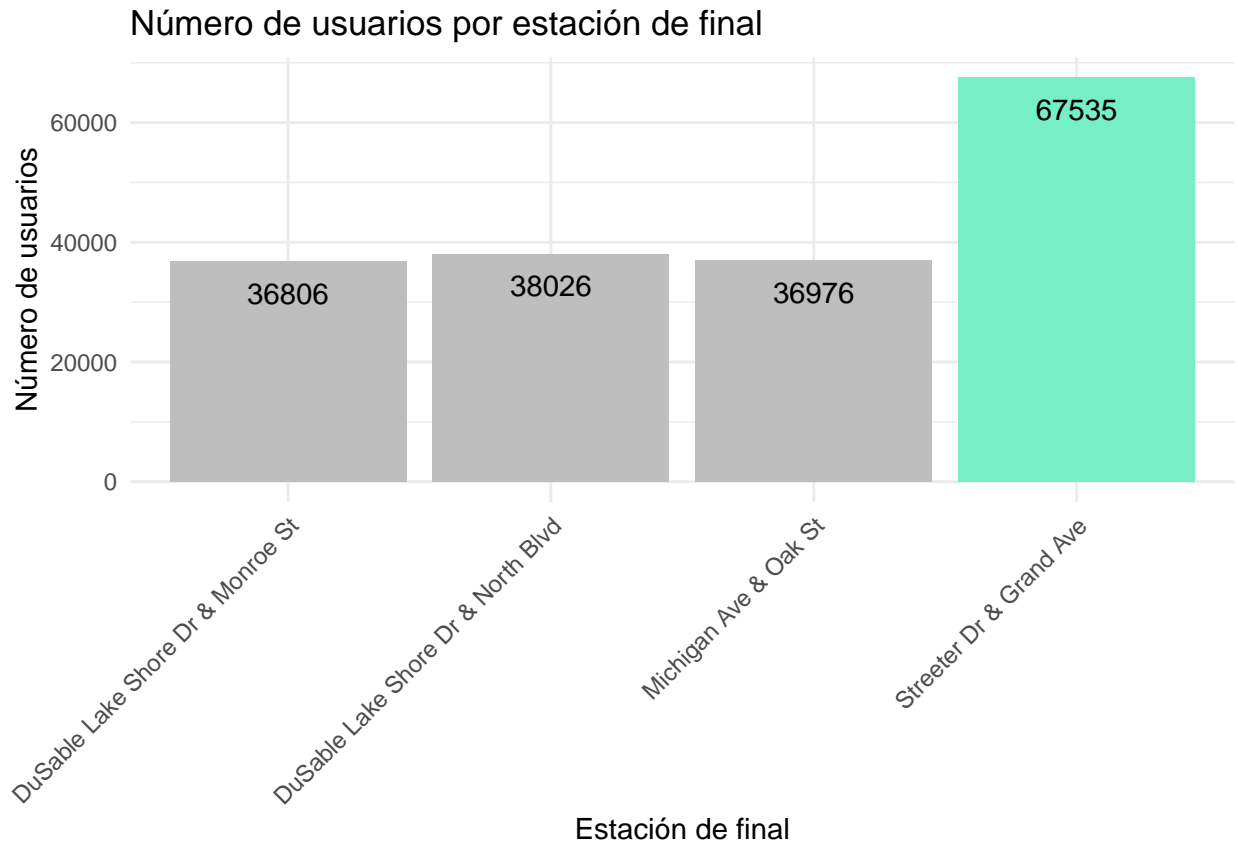
```
## # A tibble: 4 x 2
##   start_station_name      num_usuarios
##   <chr>                  <int>
## 1 Streeter Dr & Grand Ave    65891
## 2 DuSable Lake Shore Dr & Monroe St 37939
## 3 Michigan Ave & Oak St     36035
## 4 DuSable Lake Shore Dr & North Blvd 35091

## # A tibble: 4 x 2
##   end_station_name      num_usuarios
##   <chr>                  <int>
## 1 Streeter Dr & Grand Ave    67535
## 2 DuSable Lake Shore Dr & North Blvd 38026
## 3 Michigan Ave & Oak St     36976
## 4 DuSable Lake Shore Dr & Monroe St 36806
```

Table 8: Estaciones de destino con mas demanda

end_station_name	num_usuarios
Streeter Dr & Grand Ave	67535
DuSable Lake Shore Dr & North Blvd	38026
Michigan Ave & Oak St	36976
DuSable Lake Shore Dr & Monroe St	36806





Calculamos el promedio del tiempo en que los usuarios hicieron un recorrido y para tener una mejor visualización crearemos una gráfica de líneas donde representa el promedio de la duración de viajes por el tipo de usuario

