Bikeshare Analysis with R

Kristian Murimi

2022-03-29

Install Packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Import data

```
tripdata <- read.csv('trip_data.csv')
```

Clean up data

There is an unwanted column in the table.

```
trip_data <- tripdata[-c(1)]
head(trip_data)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2 A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3 BD0F91DFF741C66D  classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4 CBB80ED419105406  classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5 DDC963BFDDA51EEA  classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6 A39C6F6CC0586C0B  classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
##              start_station_name start_station_id              end_station_name
## 1      Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 2      Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 3 Sheffield Ave & Fullerton Ave     TA1306000016 Greenview Ave & Fullerton Ave
## 4      Clark St & Bryn Mawr Ave     KA1504000151     Paulina St & Montrose Ave
## 5   Michigan Ave & Jackson Blvd     TA1309000002        State St & Randolph St
## 6         Wood St & Chicago Ave              637       Honore St & Division St
##   end_station_id member_casual trip_duration trip_day time_of_trip
## 1         RP-007        casual          2.95 Thursday      Morning
## 2         RP-007        casual          4.35   Monday      Morning
## 3   TA1307000001        member          4.35  Tuesday        Night
## 4   TA1309000021        casual         14.93  Tuesday        Night
## 5   TA1305000029        member          6.03 Thursday        Night
## 6   TA1305000034        member          3.37  Tuesday    Afternoon
```

The unwanted column has been successfully removed

Summary data

How many members and casual riders use the service throughout the year?

```
user_distribution <- trip_data%>%
  group_by(member_casual)%>%
  summarise(n())
names(user_distribution) <- c("user_type", "user_count")
```

How does bike preference differ between members and casual riders?

```
bike_preference <- trip_data%>%
  group_by(rideable_type, member_casual)%>%
  summarise(n())
```

```
## `summarise()` has grouped output by 'rideable_type'. You can override using the
## `.groups` argument.
```

```
names(bike_preference) <- c("bike_type", "user_type", "user_count")
```

What is the daily rider distribution?

```
weekly_trips <- trip_data%>%
  group_by(trip_day, member_casual)%>%
  summarise(n())
```

```
## `summarise()` has grouped output by 'trip_day'. You can override using the
## `.groups` argument.
```

```
names(weekly_trips) <- c("trip_day", "user_type", "user_count")

weekly_trips$trip_day <- ordered(weekly_trips$trip_day, levels = c("Monday", "Tuesday", "Wednesday", "Th
                                                                   "Friday", "Saturday", "Sunday"))
```

What time of day do users prefer to ride bikes?

```
trip_time <- trip_data%>%
  group_by(time_of_trip,member_casual)%>%
  summarise(n())
```

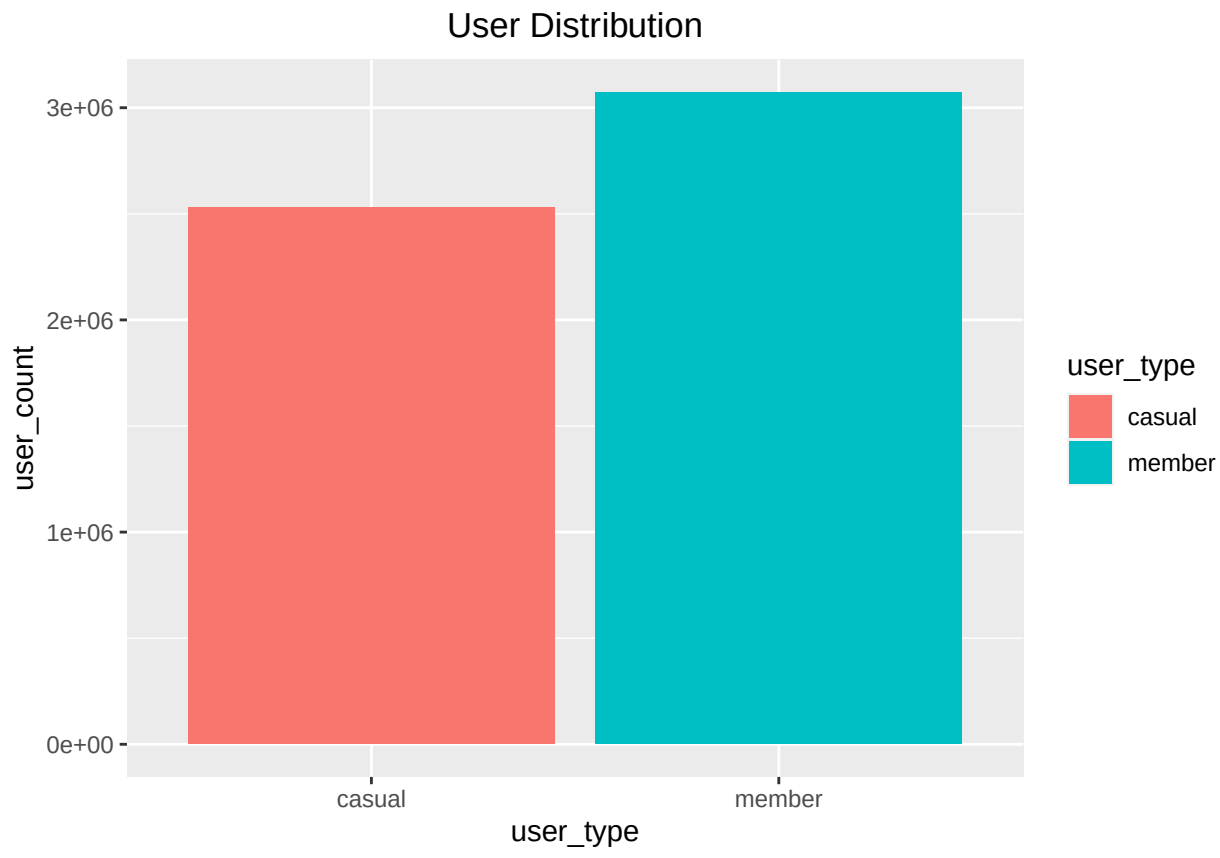
```
## `summarise()` has grouped output by 'time_of_trip'. You can override using the
## `.groups` argument.
```

```
names(trip_time)<- c("time_of_trip","user_type", "user_count")
```

Plotting the data

User distribution

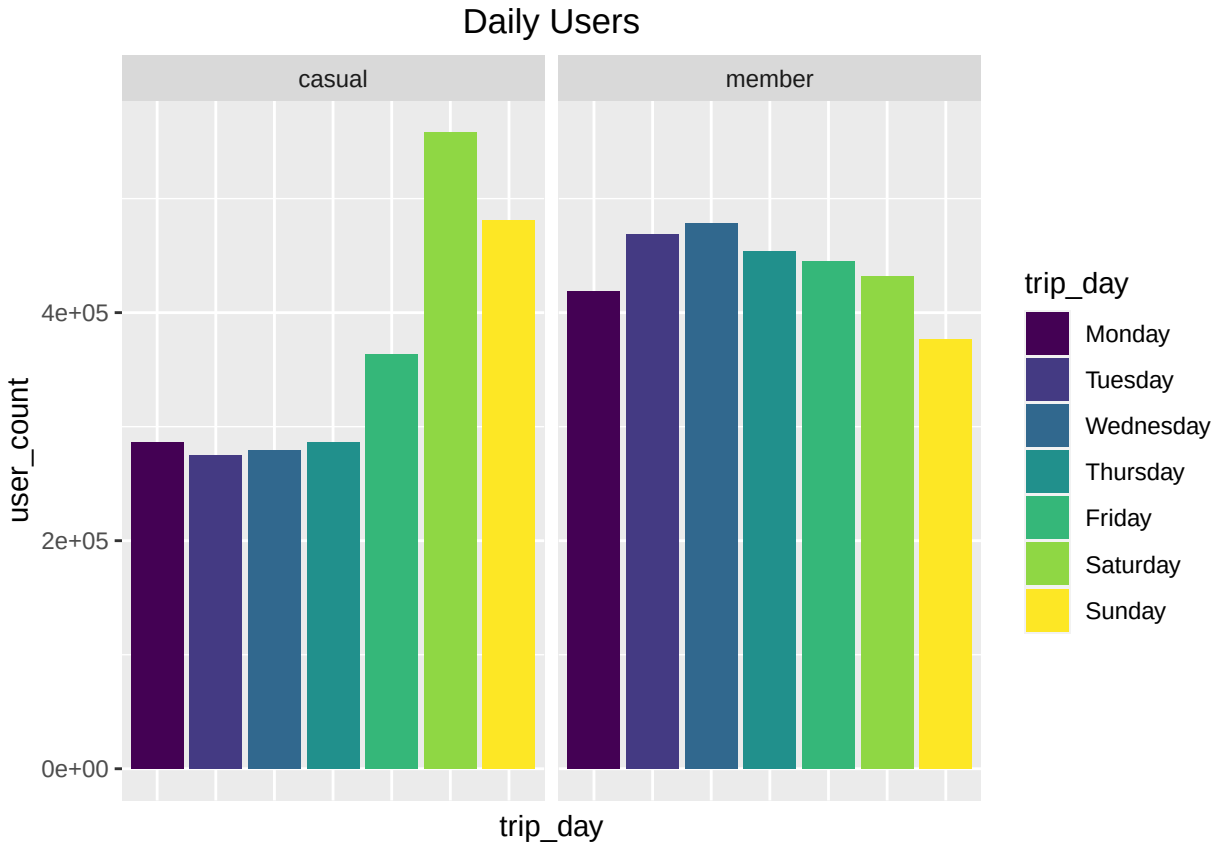
```
ggplot(user_distribution, aes(x =user_type, y = user_count, fill = user_type))+
  geom_bar(stat = "identity")+
  labs(title = "User Distribution")+
  theme(plot.title = element_text(hjust = 0.5))
```



There are more members than casual riders.

Daily user distribution

```
ggplot(weekly_trips, aes(x = trip_day, y = user_count, fill = trip_day))+
  geom_bar(stat = "identity")+
  facet_wrap(~user_type)+
  labs(title = "Daily Users")+
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        plot.title = element_text(hjust = 0.5))
```

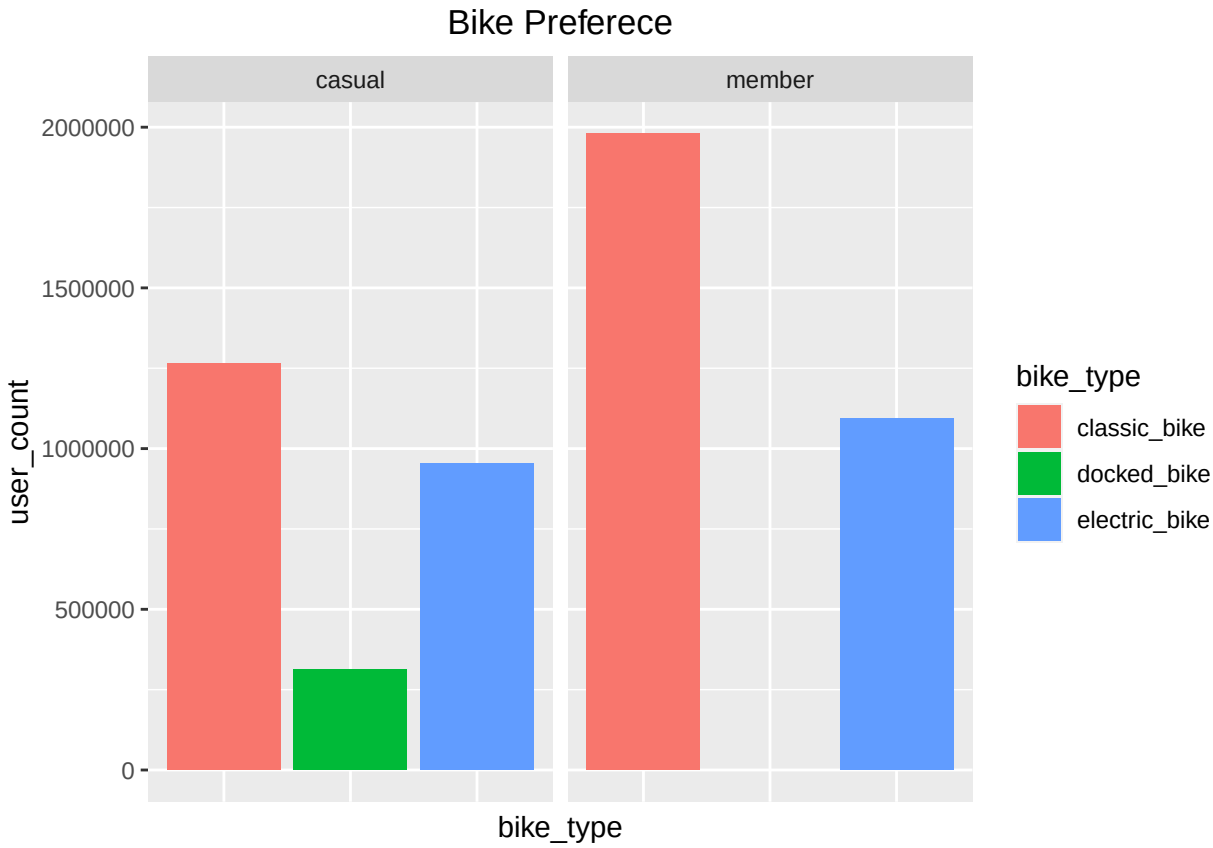


Most of the casual riders take trips on the weekends.This suggests that most casual riders use the bike service for leisure. Members of the ride share service, however, take more trips over the weekdays.This suggests that most of the members use the service for transport.

Bike preference

```
ggplot(bike_preference, aes(x=bike_type, y = user_count, fill = bike_type))+
  geom_bar(stat = "identity")+
  facet_wrap(~user_type)+
  labs(title = "Bike Preferece")+
```

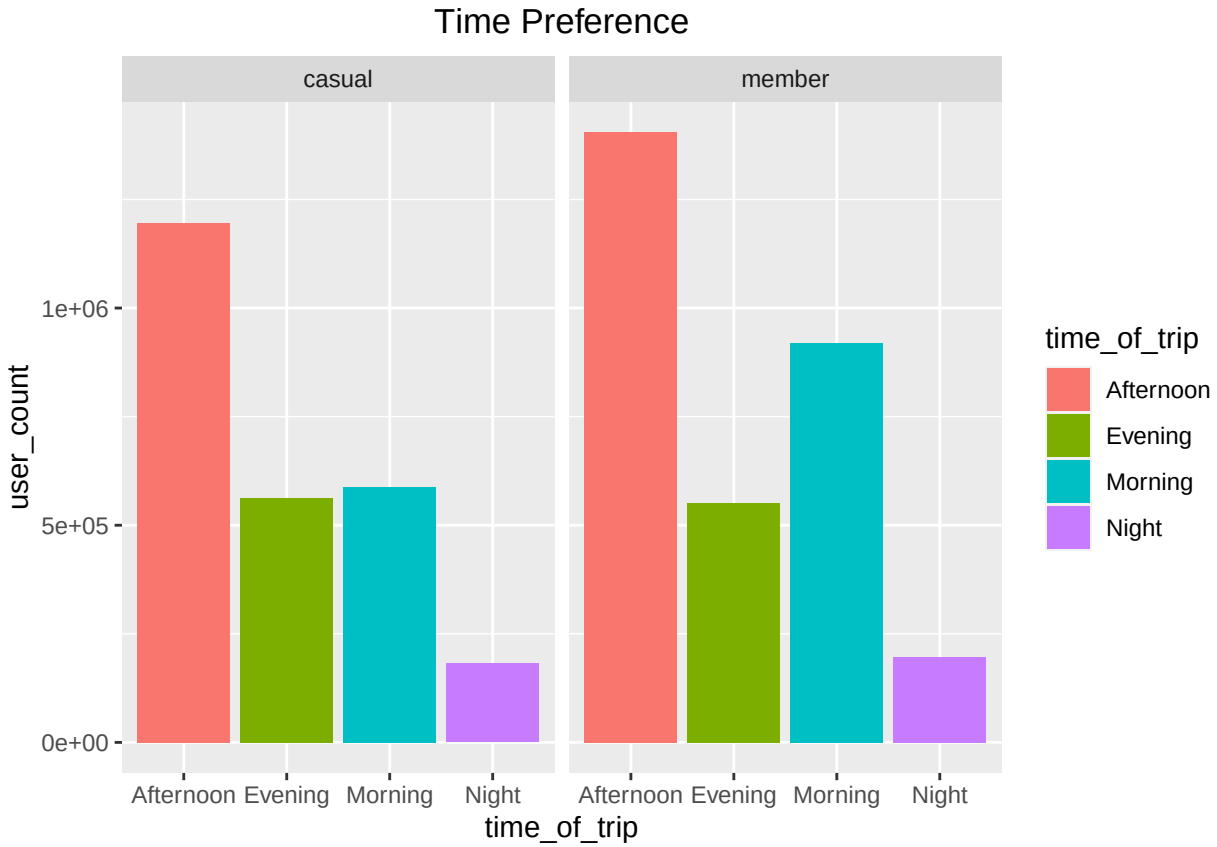
```
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        plot.title = element_text(hjust = 0.5))
```



This shows that the most used bike type is the classic bike, which is the preferred type by both members and casual riders. The least common bike type is the docked bike which is only used by a small number of casual riders and not at all by members.

Time of trip

```
ggplot(trip_time, aes(x= time_of_trip, y = user_count, fill = time_of_trip))+
  geom_bar(stat ="identity")+
  facet_wrap(~user_type)+
  labs(title = "Time Preference")+
  theme(plot.title = element_text(hjust = 0.5))
```



Most of the casual riders use the service in the afternoon suggesting that they use the service mostly for leisure.The member riders mostly ride in the afternoons and morning which may indicate they are using the service to get to and from work.