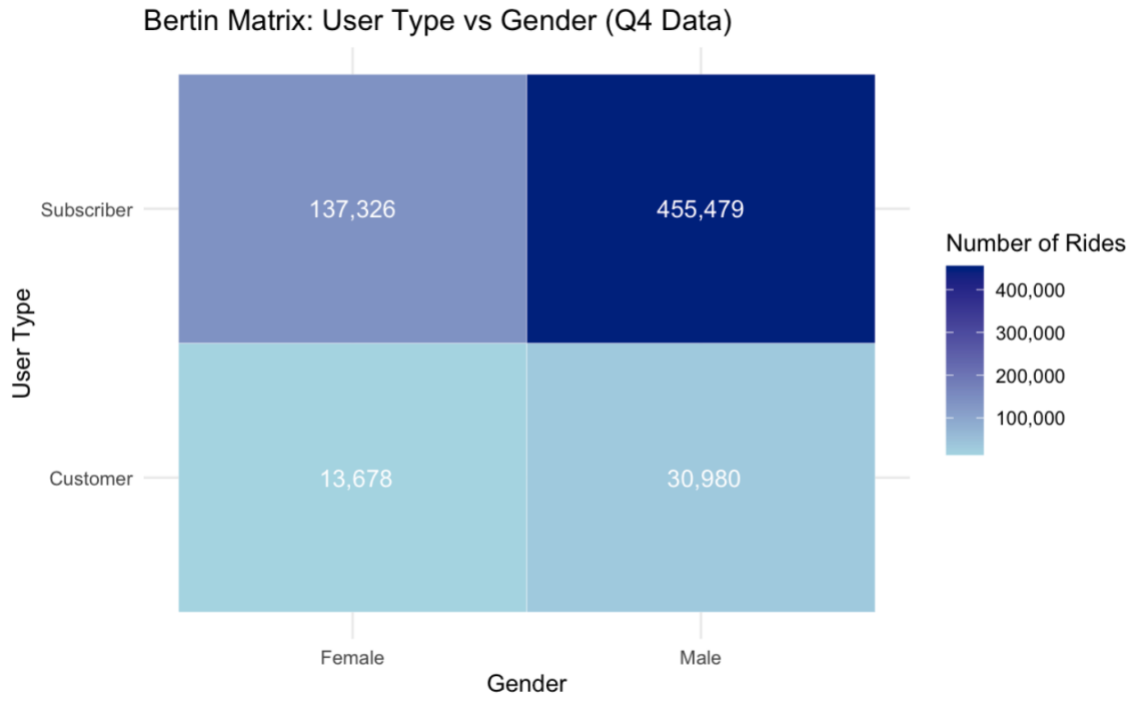
Homework 4 - Question 1

Steps :

I plotted a Bertin Matrix and below are the steps :

Step 1: Loaded the required libraries - tidyverse, ggplot2, and scales to handle the data and create the visualization.
Step 2: Prepared the data - Cleaned the dataset by removing missing values, grouped it by user type and gender, and counted the number of rides in each group.
Step 3: Created the Bertin Matrix – Made the tiles using geom_tile() with white borders, added the ride counts inside each tile with geom_text(), and applied a blue gradient from light to dark based on the count.
Step 4: Fixed number formatting – Added commas to make numbers more readable (like 455,479 instead of 455479) and removed scientific notation.
Step 5: Added labels – I included a clear title, axis labels, and a legend to explain what the chart shows. This way, the Bertin Matrix clearly highlights patterns in ride usage across different user types and genders.



Observations :

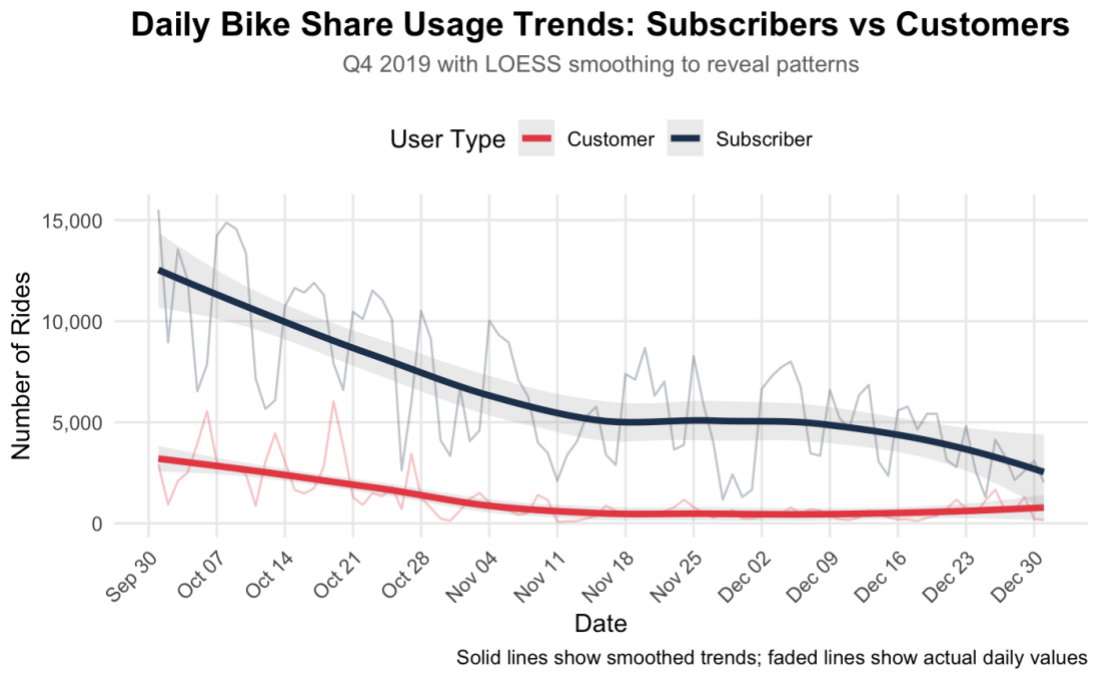
Most regular riders are men, with male subscribers taking 455,479 rides - far more than any other group. In general, men use the bikes about three times more than women, whether they're subscribers or casual users. People with subscriptions (593,805 rides) clearly outnumber one-time customers (44,658 rides), showing that the service mainly attracts regular commuters. Female customers make up the smallest group, with only 13,678 rides, which could be an opportunity to reach more casual female riders. The chart's colors make these patterns easy to see dark blue highlights heavy usage, while light blue shows groups that are less active.

Question 2

Steps:

I created a time series plot to show seasonal trends and differences in user behavior:

Step 1: Loaded libraries – tidyverse, ggplot2, scales, and lubridate to handle and visualize the data.
Step 2: Prepared the data – Filtered out rides with missing start times, converted start_time to just the date, grouped by date and user type, and counted the rides for each day-user type combination.
Step 3: Created the base plot – Set up ggplot() with date on the x-axis and ride counts on the y-axis, coloring lines by user type (Customer vs Subscriber).
Step 4: Added two types of lines – Faded lines (geom_line() with alpha = 0.3) show daily ride counts with all the ups and downs, while bold smoothed lines (geom_smooth() with LOESS) highlight the overall trend.
Step 5: Formatted the plot – I used custom colors (red for Customers, dark blue for Subscribers), formatted the y-axis with commas, set x-axis labels weekly, and added a title, subtitle, and caption to explain the chart.



Observation :

Subscriber rides steadily drop through Q4, going from about 12,000 in early October to around 3,000 by the end of December, showing how colder weather slows down regular use. Casual riders, however, stay consistent, with only 500 - 1,000 rides a day, so they're less affected by the seasons. Throughout the quarter, subscribers take 10 to 15 times more rides than customers, which shows the service is mostly used by regular commuters. You can also see daily ups and downs for subscribers, probably because of weekday versus weekend routines. Overall, winter hits subscribers the hardest, while casual riders keep riding at a steady pace.