## In this document, type only in the boxes provided (which you may enlarge if necessary) and only use the courier font

(this is what the courier font looks like)

## Dataset Code and Spreadsheet Used

|  |  |
| --- | --- |
| Code (beginning with ds..) | Spreadsheet (csv/xlsx/xls) |
| ds324 | Golf.csv |

## Question 1

What is the dataset about?

(Response limit: Max 30 words but ideally less)

|  |
| --- |
| The dataset compares the driving distances achieved by current and new golf ball models in yards tested by Par Inc. |

## Question 2

Produce the first 5 rows of the dataset. Only include the maximum 5 most interesting columns and rename them if necessary. Do not add any formatting (spaces etc.) unless produced by R.   
(Only use the courier font for this and make sure the columns align. If you rename the columns to make them shorter, underneath the 5 rows list the renamings you have done e.g  
\*acc renamed from “Acceration of the cars”)

|  |
| --- |
| Current New  1 264 277  2 261 269  3 267 263  4 272 266  5 258 262 |

## Question 3

What does each line of the dataset represent?

(Response limit: Max 30 words)

|  |
| --- |
| Each row represents one test result comparing the drive distance of the current and new golf ball models in yards under identical conditions. |

## Question 4

Using basic descriptive statistics, summarize the dataset and highlight some salient features about it on initial viewing.

(Response limit: Max 150 words)

|  |
| --- |
| From the summary, the **current model** ranges from **255 to 289 yards**, with a **mean of 270.3**. The first quartile (**Q1**) is **263.0**, the median (**Q2**) is **270.0**, and the third quartile (**Q3**) is **275.2**. The **new model** ranges from **250 to 289 yards**, with a **mean of 267.5**. The first quartile (**Q1**) is **262.0**, the median (**Q2**) is **265.0**, and the third quartile (**Q3**) is **274.5**. The **new ball’s distances** are slightly lower at all quartiles, suggesting it performs a bit worse overall. Both distributions are approximately normal with **no extreme outliers**, and the results indicate the **new model does not improve on distance performance**. |

**Question 5**

What is the most surprising thing you can see when first looking at the data?

(Response limit: Max 50 words)

|  |
| --- |
| Contrary to expectations, the new golf ball travels **slightly shorter distances** than the current one across most test results. |

## Question 6

Draw a graph using R base graphics showing something interesting about your dataset, and explain it in a single sentence afterwards.

(Put image here with explanation below – max text 30 words)

|  |
| --- |
|  |
| The barplot clearly shows that the current golf ball performs better in most tests, with noticeably more results where it travels farther than the new ball. Only a few tests show equal performance. |

## Question 7

Reproduce the R Code you used to create the graph. Only use graphical techniques taught in the course. Do not use any external libraries. The code you provide should be the same as in the R file you supplied and capable of being copied and pasted to an R prompt to generate a graph

|  |
| --- |
| # Determine which ball goes farther in each test  win <- ifelse(golf$New > golf$Current, "New farther",  ifelse(golf$New < golf$Current, "Current farther", "Equal"))  counts <- table(win)  # Define colours  bar\_colours <- c("steelblue", "lightgreen", "lightcoral")  # Y-axis limit  y\_max <- max(counts) \* 1.2  # Create the barplot  bp <- barplot(counts,  main = "Comparison of Golf Ball Performance",  ylab = "Number of Tests",  xlab = "Result Category",  col = bar\_colours,  border = "grey40",  ylim = c(0, y\_max))  # Add horizontal grid lines  abline(h = seq(0, y\_max, by = 2), col = "lightgray", lty = "dotted", lwd = 0.8)  # Add count labels above each bar  text(x = bp, y = counts + 0.8, labels = counts, cex = 1.1, font = 1, col = "black") |

## Question 8

What interesting conclusions do you think it might be possible to make from a more in depth view of this data, and what real-world implications might it have?

(Response limit: Max 100 words)

|  |
| --- |
| A more detailed examination of the data could identify factors influencing drive distance, such as player consistency or environmental conditions. The findings clearly suggest that the new golf ball performs slightly worse than the current model across most tests. This outcome may prompt a review of the new ball’s design, materials, or manufacturing process to determine why it underperforms. For the company, such insights are valuable in avoiding costly product rollouts that fail to meet expectations and ensuring that future designs genuinely enhance performance and customer satisfaction. |

## Question 9

Provide a short anecdote ideally from your own experience or if not that, relating to something in your country of origin, which has commonalities with some of the issues and concepts visible in the dataset (max 100 words)

|  |
| --- |
| In Sri Lanka, local cricket players often test new ball brands to compare swing and travel distance. Likewise, this dataset checks if a new golf ball design improves or not the driving distance. but, contrary to expectations, it performs slightly worse. |

## Question 10

Include maximum 2 citations (for things cited earlier)

|  |
| --- |
| 1) Kaggle. (2022). *Golf Ball Testing Data Set from Par Inc.* Retrieved from <https://www.kaggle.com/datasets/ipravin/golf-ball-testing-data-set-from-par-inc>  2) University of Hertfordshire. (2025). *Doing Graphics for Assignment 1* (Course Handout). |