BI Tools Analysis

# 1. Introduction

The information in this analysis comes from the website of Gartner, a consulting and research firm, which provides a list of all BI tools.

We will retrieve information on the top 30 out of 200+ business intelligence tools ranked by number of reviews, from most to least, into Excel, such as tool name, number of ratings and review.

More information in the following link: <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>

# 2. Gartner’s Ranking

## 2.1 Importing Gartner’s top 30 by Ratings Number from high to low

Importing the Excel extraction and display the 30 first BI tools by highest ratings number to smallest ones.

# Import Excel sheet  
library(readxl)  
df <- read\_excel("C:/Users/dinho2/Medtronic PLC/Team - Thesis - Thesis/Thesis/4. References/R - BI Tools/BI Tools.xlsx", sheet = "4. Top 30 - Gartner")  
knitr::kable(  
 df,  
 caption = "Gartner Top 30 BI Tools",  
 format = "pipe" # optional, safe default for all outputs  
)

Gartner Top 30 BI Tools

| Gartner’s Ranking | Tool | Review | Ratings Number |
| --- | --- | --- | --- |
| 1 | Tableau (+ CRM Analytics) | 4.4 | 678 |
| 2 | Microsoft Power BI | 4.4 | 613 |
| 3 | Qlik (View + Sense + Cloud Analytics) | 4.4 | 481 |
| 4 | SAP (BusinessObjects + Analytics Cloud) | 4.3 | 269 |
| 5 | Strategy One | 4.6 | 137 |
| 6 | Oracle Analytics (Server + Cloud + Fusion) | 4.1 | 139 |
| 7 | SAS (Enterprise Guide + Visual Analytics + Office Analytics + Viya) | 4.5 | 127 |
| 8 | Spotfire | 4.5 | 123 |
| 9 | Amazon QuickSight | 4.2 | 116 |
| 10 | Sisense Fusion Analytics | 4.0 | 104 |
| 11 | Looker | 4.5 | 88 |
| 12 | IBM Cognos Analytics | 4.3 | 72 |
| 13 | Domo | 4.4 | 67 |
| 14 | Diver Platform | 4.7 | 59 |
| 15 | ThoughtSpot Analytics | 4.5 | 50 |
| 16 | Zoho Analytics | 4.5 | 44 |
| 17 | Pyramid Decision Intelligence Platform | 4.3 | 44 |
| 18 | Birst | 4.4 | 35 |
| 19 | Logi Symphony | 4.4 | 35 |
| 20 | Databricks Data Intelligence Platform | 4.9 | 31 |
| 21 | Tellius | 4.7 | 31 |
| 22 | Yellowfin | 4.6 | 24 |
| 23 | Board | 4.1 | 25 |
| 24 | FineReport | 4.9 | 19 |
| 25 | Alibaba Quick BI | 4.7 | 18 |
| 26 | Sigma | 4.8 | 17 |
| 27 | ibi WebFOCUS | 4.4 | 18 |
| 28 | Minitab Connect | 4.3 | 18 |
| 29 | Incorta | 4.5 | 16 |
| 30 | Alteryx One Platform | 4.9 | 13 |

## 2.2 Transforming and Cleaning Data

Transforming and cleaning such as removing parentheses and trimming data.

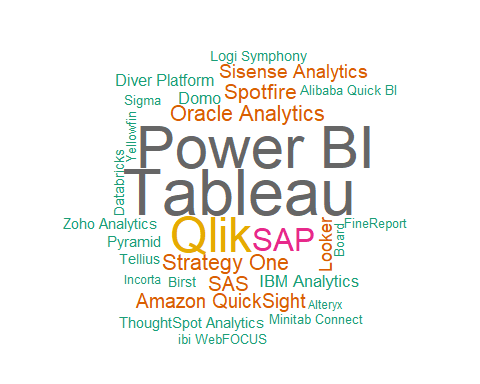
# Remove text in parentheses (and the parentheses themselves)  
df$Tool <- gsub("\\s\*\\(.\*?\\)", "", df$Tool)  
  
# Trim leading/trailing whitespace  
df$Tool <- trimws(df$Tool)

# Shortening names  
df$Tool <- gsub("Microsoft Power BI", "Power BI", df$Tool)  
df$Tool <- gsub("SAP BusinessObjects BI Suite", "SAP Analytics", df$Tool)  
#df$Tool <- gsub("Oracle Analytics", "Oracle", df$Tool)  
#df$Tool <- gsub("Amazon QuickSight", "Amazon", df$Tool)  
df$Tool <- gsub("Sisense Fusion Analytics", "Sisense Analytics", df$Tool)  
df$Tool <- gsub("IBM Cognos Analytics", "IBM Analytics", df$Tool)  
#df$Tool <- gsub("Diver Platform", "Diver", df$Tool)  
#df$Tool <- gsub("ThoughtSpot Analytics", "ThoughtSpot", df$Tool)  
#df$Tool <- gsub("Zoho Analytics", "Zoho", df$Tool)  
df$Tool <- gsub("Pyramid Decision Intelligence Platform", "Pyramid", df$Tool)  
df$Tool <- gsub("Databricks Data Intelligence Platform", "Databricks", df$Tool)  
#df$Tool <- gsub("Alibaba Quick BI", "Alibaba", df$Tool)  
df$Tool <- gsub("Alteryx One Platform", "Alteryx", df$Tool)

## 2.3 Representation of Gartner’s Ranking

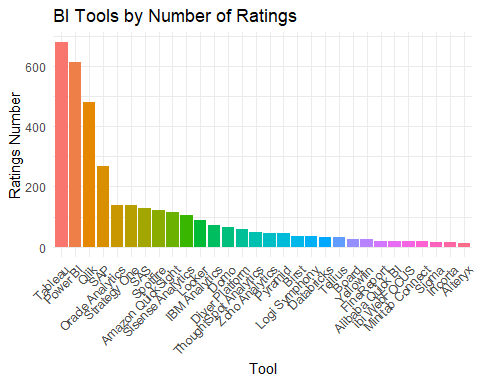
Visualizing the data as an exploratory data analysis.

library(wordcloud)  
library(RColorBrewer)  
set.seed(1234) # You can use any number, but keep it constant  
wordcloud(  
 words = df$Tool,  
 freq = df[["Ratings Number"]],  
 min.freq = 1,  
 max.words = 100,  
 random.order = FALSE,  
 rot.per = 0.1,  
 colors = brewer.pal(8, "Dark2"),  
 scale = c(4, 0.7)  
)



In this wordcloud visualization, we notice two principal business intelligence tools: Tableau and Power BI.

library(ggplot2)  
# Reorder tools by descending Ratings Number  
df$Tool <- factor(df$Tool, levels = df$Tool[order(-df[["Ratings Number"]])])  
  
# Create vertical bar chart  
ggplot(df, aes(x = Tool, y = `Ratings Number`, fill = Tool)) +  
 geom\_col(show.legend = FALSE) +  
 labs(  
 title = "BI Tools by Number of Ratings",  
 x = "Tool",  
 y = "Ratings Number"  
 ) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1, vjust = 1))



In this bar chart, we visualize quantitatively that Tableau and Power BI are far from the other business intelligence tools with more than 600 number of reviews.

# 3. New Ranking

As the Gartner’s ranking from largest to smallest has a bias of popularity, it has been decided to create a score system to better rank these business intelligence tools.

## 3.1 Scoring System

Creating multiple scoring systems to better visualize how the final ranking would be, mainly using squared, root square, and logarithm.

# Core scoring formulas  
df$"Scoring1" <- round(df$Review \* df[["Ratings Number"]], 0) # Review × Ratings  
df$"Scoring2" <- round(df$Review \* sqrt(df[["Ratings Number"]]), 0) # Review × √Ratings  
df$"Scoring3" <- round(df$Review \* log(df[["Ratings Number"]]), 1) # Review × log(Ratings)  
df$"Scoring4" <- round(df$Review^2 \* df[["Ratings Number"]], 0) # Review² × Ratings  
df$"Scoring5" <- round(df$Review^2 \* sqrt(df[["Ratings Number"]]), 0) # Review² × √Ratings  
df$"Scoring6" <- round(df$Review^2 \* log(df[["Ratings Number"]]), 0) # Review² × log(Ratings)  
  
knitr::kable(  
 df,  
 caption = "BI Tools Scoring Table",  
 col.names = c(  
 "Gartner's Ranking", "Tool", "Ratings Number", "Review",  
 "Scoring1", "Scoring2", "Scoring3",  
 "Scoring4", "Scoring5", "Scoring6"  
 )  
)

BI Tools Scoring Table

| Gartner’s Ranking | Tool | Ratings Number | Review | Scoring1 | Scoring2 | Scoring3 | Scoring4 | Scoring5 | Scoring6 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Tableau | 4.4 | 678 | 2983 | 115 | 28.7 | 13126 | 504 | 126 |
| 2 | Power BI | 4.4 | 613 | 2697 | 109 | 28.2 | 11868 | 479 | 124 |
| 3 | Qlik | 4.4 | 481 | 2116 | 96 | 27.2 | 9312 | 425 | 120 |
| 4 | SAP | 4.3 | 269 | 1157 | 71 | 24.1 | 4974 | 303 | 103 |
| 5 | Strategy One | 4.6 | 137 | 630 | 54 | 22.6 | 2899 | 248 | 104 |
| 6 | Oracle Analytics | 4.1 | 139 | 570 | 48 | 20.2 | 2337 | 198 | 83 |
| 7 | SAS | 4.5 | 127 | 572 | 51 | 21.8 | 2572 | 228 | 98 |
| 8 | Spotfire | 4.5 | 123 | 554 | 50 | 21.7 | 2491 | 225 | 97 |
| 9 | Amazon QuickSight | 4.2 | 116 | 487 | 45 | 20.0 | 2046 | 190 | 84 |
| 10 | Sisense Analytics | 4.0 | 104 | 416 | 41 | 18.6 | 1664 | 163 | 74 |
| 11 | Looker | 4.5 | 88 | 396 | 42 | 20.1 | 1782 | 190 | 91 |
| 12 | IBM Analytics | 4.3 | 72 | 310 | 36 | 18.4 | 1331 | 157 | 79 |
| 13 | Domo | 4.4 | 67 | 295 | 36 | 18.5 | 1297 | 158 | 81 |
| 14 | Diver Platform | 4.7 | 59 | 277 | 36 | 19.2 | 1303 | 170 | 90 |
| 15 | ThoughtSpot Analytics | 4.5 | 50 | 225 | 32 | 17.6 | 1012 | 143 | 79 |
| 16 | Zoho Analytics | 4.5 | 44 | 198 | 30 | 17.0 | 891 | 134 | 77 |
| 17 | Pyramid | 4.3 | 44 | 189 | 29 | 16.3 | 814 | 123 | 70 |
| 18 | Birst | 4.4 | 35 | 154 | 26 | 15.6 | 678 | 115 | 69 |
| 19 | Logi Symphony | 4.4 | 35 | 154 | 26 | 15.6 | 678 | 115 | 69 |
| 20 | Databricks | 4.9 | 31 | 152 | 27 | 16.8 | 744 | 134 | 82 |
| 21 | Tellius | 4.7 | 31 | 146 | 26 | 16.1 | 685 | 123 | 76 |
| 22 | Yellowfin | 4.6 | 24 | 110 | 23 | 14.6 | 508 | 104 | 67 |
| 23 | Board | 4.1 | 25 | 102 | 20 | 13.2 | 420 | 84 | 54 |
| 24 | FineReport | 4.9 | 19 | 93 | 21 | 14.4 | 456 | 105 | 71 |
| 25 | Alibaba Quick BI | 4.7 | 18 | 85 | 20 | 13.6 | 398 | 94 | 64 |
| 26 | Sigma | 4.8 | 17 | 82 | 20 | 13.6 | 392 | 95 | 65 |
| 27 | ibi WebFOCUS | 4.4 | 18 | 79 | 19 | 12.7 | 348 | 82 | 56 |
| 28 | Minitab Connect | 4.3 | 18 | 77 | 18 | 12.4 | 333 | 78 | 53 |
| 29 | Incorta | 4.5 | 16 | 72 | 18 | 12.5 | 324 | 81 | 56 |
| 30 | Alteryx | 4.9 | 13 | 64 | 18 | 12.6 | 312 | 87 | 62 |

At first glance, it is not easy to compare scoring systems with each other, but we note that some systems have a wide dispersion of scores.

## 3.2 Ranking System

We will transform each scoring system into a business intelligence tool ranking. Each system will have its own ranking.

library(dplyr)  
# Rank: Review × Ratings  
df <- df %>%  
 arrange(desc(`Scoring1`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring1` = row\_number())  
  
# Rank: Review × √Ratings  
df <- df %>%  
 arrange(desc(`Scoring2`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring2` = row\_number())  
  
# Rank: Review × log(Ratings)  
df <- df %>%  
 arrange(desc(`Scoring3`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring3` = row\_number())  
  
# Rank: Review² × Ratings  
df <- df %>%  
 arrange(desc(`Scoring4`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring4` = row\_number())  
  
# Rank: Review² × √Ratings  
df <- df %>%  
 arrange(desc(`Scoring5`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring5` = row\_number())  
  
# Rank: Review² × log(Ratings)  
df <- df %>%  
 arrange(desc(`Scoring6`), desc(`Ratings Number`)) %>%  
 mutate(`Rank: Scoring6` = row\_number())

### 3.2.1 Averaging the Scores, Sorting and Indexing

Based on each ranking for each business intelligence tool, an average is calculated from this ranking.

This average will give the average ranking for each of these systems.

Then, this average will be sorted from highest to lowest.

Finally, indexing will be performed based on the sorted average, which will create a temporary ranking.

# Calculate average and standard deviation of ranks  
rank\_columns <- grep("^Rank:", names(df), value = TRUE)  
df$`Rank Average` <- round(rowMeans(df[rank\_columns]), 1)  
df$`Rank Std Dev` <- round(apply(df[rank\_columns], 1, sd), 1)

# Sorting the final ranking by lowest Rank Average  
df <- df[order(df$`Rank Average`, df$`Rank Std Dev`), ]

# Indexing the final ranking using the Rank Average resulting from each scoring system  
df$`Temporary Ranking` <- seq\_len(nrow(df))  
knitr::kable(  
 df[, c("Tool", "Review", "Ratings Number", "Rank Average", "Rank Std Dev", "Temporary Ranking")],  
 caption = "Summary of BI Tool Scores and Temporary Ranking",  
 col.names = c("Tool", "Review", "Ratings Number", "Rank Average", "Rank Std Dev", "Temporary Ranking")  
)

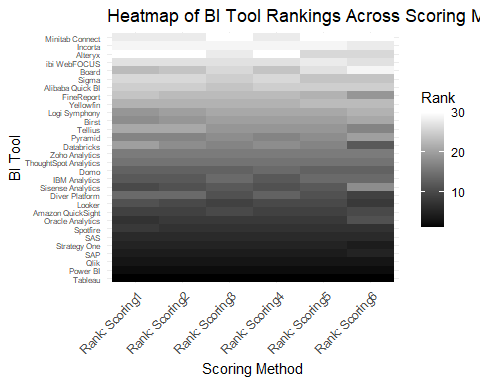
Summary of BI Tool Scores and Temporary Ranking

| Tool | Review | Ratings Number | Rank Average | Rank Std Dev | Temporary Ranking |
| --- | --- | --- | --- | --- | --- |
| Tableau | 4.4 | 678 | 1.0 | 0.0 | 1 |
| Power BI | 4.4 | 613 | 2.0 | 0.0 | 2 |
| Qlik | 4.4 | 481 | 3.0 | 0.0 | 3 |
| SAP | 4.3 | 269 | 4.2 | 0.4 | 4 |
| Strategy One | 4.6 | 137 | 4.8 | 0.4 | 5 |
| SAS | 4.5 | 127 | 6.0 | 0.0 | 6 |
| Spotfire | 4.5 | 123 | 7.2 | 0.4 | 7 |
| Oracle Analytics | 4.1 | 139 | 8.3 | 1.4 | 8 |
| Amazon QuickSight | 4.2 | 116 | 9.3 | 0.5 | 9 |
| Looker | 4.5 | 88 | 9.7 | 1.0 | 10 |
| Diver Platform | 4.7 | 59 | 12.0 | 2.0 | 11 |
| Sisense Analytics | 4.0 | 104 | 12.3 | 2.9 | 12 |
| IBM Analytics | 4.3 | 72 | 13.0 | 1.1 | 13 |
| Domo | 4.4 | 67 | 13.2 | 0.4 | 14 |
| ThoughtSpot Analytics | 4.5 | 50 | 15.0 | 0.0 | 15 |
| Zoho Analytics | 4.5 | 44 | 16.0 | 0.0 | 16 |
| Databricks | 4.9 | 31 | 17.0 | 2.7 | 17 |
| Pyramid | 4.3 | 44 | 17.8 | 1.2 | 18 |
| Tellius | 4.7 | 31 | 19.3 | 1.5 | 19 |
| Birst | 4.4 | 35 | 19.7 | 1.0 | 20 |
| Logi Symphony | 4.4 | 35 | 20.7 | 1.0 | 21 |
| Yellowfin | 4.6 | 24 | 22.3 | 0.5 | 22 |
| FineReport | 4.9 | 19 | 22.3 | 1.8 | 23 |
| Alibaba Quick BI | 4.7 | 18 | 24.8 | 0.4 | 24 |
| Sigma | 4.8 | 17 | 25.2 | 1.0 | 25 |
| Board | 4.1 | 25 | 25.5 | 2.3 | 26 |
| ibi WebFOCUS | 4.4 | 18 | 27.2 | 0.4 | 27 |
| Alteryx | 4.9 | 13 | 28.3 | 2.0 | 28 |
| Incorta | 4.5 | 16 | 28.8 | 0.4 | 29 |
| Minitab Connect | 4.3 | 18 | 29.0 | 1.1 | 30 |

## 3.3 Testing the scoring and ranking systems.

To find the best scoring system, a heatmap will be made in order to better visualize all the scoring systems.

library(reshape2)  
library(ggplot2)  
  
# Step 1: Extract only ranking columns and Tool  
rank\_columns <- grep("^Rank:", names(df), value = TRUE)  
ranking\_matrix <- df[, c("Tool", rank\_columns)]  
  
# Step 2: Melt into long format  
ranking\_long <- melt(ranking\_matrix, id.vars = "Tool", variable.name = "Scoring Method", value.name = "Rank")  
  
# Step 3: Merge with Rank Average to reorder Tool correctly  
ranking\_long <- merge(ranking\_long, df[, c("Tool", "Rank Average")], by = "Tool")  
  
# Step 4: Heatmap  
ggplot(ranking\_long, aes(x = `Scoring Method`, y = reorder(Tool, `Rank Average`))) +  
 geom\_tile(aes(fill = Rank)) +  
 scale\_fill\_gradient(low = "black", high = "white", name = "Rank") +  
 labs(  
 title = "Heatmap of BI Tool Rankings Across Scoring Methods",  
 x = "Scoring Method",  
 y = "BI Tool"  
 ) +  
 theme\_minimal() +  
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1),  
 axis.text.y = element\_text(size = 6)  
 )



Based on this heat map, we can see that the scoring3 ranking closely follows and is very similar to the overall average ranking.

### 3.3.1 Finding the best scoring system

Counting the matches of each scoring system to the temporary ranking.

# Step 1: Compare each score ranking to the final New Ranking  
rank\_columns <- grep("^Rank:", names(df), value = TRUE)  
  
# Step 2: Count how many times each scoring method matches New Ranking  
match\_counts <- sapply(rank\_columns, function(col) {  
 sum(df[[col]] == df$`Temporary Ranking`, na.rm = TRUE)  
})  
  
# Step 3: Convert to data frame  
match\_df <- data.frame(  
 ScoringMethod = rank\_columns,  
 MatchesFinalRank = as.numeric(match\_counts)  
)  
  
# Step 4: Sort descending by match count  
match\_df <- match\_df[order(-match\_df$MatchesFinalRank), ]  
  
# Step 5: View results  
knitr::kable(  
 match\_df,  
 caption = "Counting Match Between Scoring Methods and Final Ranking",  
 col.names = c("Scoring Method", "Exact Matches with Final Rank")  
)

Counting Match Between Scoring Methods and Final Ranking

|  | Scoring Method | Exact Matches with Final Rank |
| --- | --- | --- |
| 3 | Rank: Scoring3 | 26 |
| 5 | Rank: Scoring5 | 21 |
| 4 | Rank: Scoring4 | 20 |
| 2 | Rank: Scoring2 | 16 |
| 1 | Rank: Scoring1 | 12 |
| 6 | Rank: Scoring6 | 9 |

The scoring system 3: “Review x log(Ratings)” is the most relevant one and will be chosen to make the final ranking.

This scoring system shows that using a logarithm for the number of reviews is the right solution for balancing reviews against the number of ratings. It takes into account quantitative and qualitative review data and the popularity of the number of reviews. Using a logarithm compresses this range and rewards popularity without overdoing it.

# 4. New Table with the Final Ranking

In making this final ranking, we omit scoring systems that miscalculate the qualitative data of reviews and the popularity of the number of reviews, and only take into account the scoring system that allows for a more accurate final ranking.

df <- df[, c("Rank: Scoring3", "Gartner's Ranking", "Tool", "Ratings Number", "Review", "Scoring3")]  
  
df %>%  
 arrange(`Rank: Scoring3`) %>%  
 knitr::kable(  
 caption = "Filtered BI Tools Scoring Table",  
 col.names = c("Final Ranking", "Gartner's Ranking", "Tool", "Ratings Number", "Review", "Score")  
 )

Filtered BI Tools Scoring Table

| Final Ranking | Gartner’s Ranking | Tool | Ratings Number | Review | Score |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | Tableau | 678 | 4.4 | 28.7 |
| 2 | 2 | Power BI | 613 | 4.4 | 28.2 |
| 3 | 3 | Qlik | 481 | 4.4 | 27.2 |
| 4 | 4 | SAP | 269 | 4.3 | 24.1 |
| 5 | 5 | Strategy One | 137 | 4.6 | 22.6 |
| 6 | 7 | SAS | 127 | 4.5 | 21.8 |
| 7 | 8 | Spotfire | 123 | 4.5 | 21.7 |
| 8 | 6 | Oracle Analytics | 139 | 4.1 | 20.2 |
| 9 | 11 | Looker | 88 | 4.5 | 20.1 |
| 10 | 9 | Amazon QuickSight | 116 | 4.2 | 20.0 |
| 11 | 14 | Diver Platform | 59 | 4.7 | 19.2 |
| 12 | 10 | Sisense Analytics | 104 | 4.0 | 18.6 |
| 13 | 13 | Domo | 67 | 4.4 | 18.5 |
| 14 | 12 | IBM Analytics | 72 | 4.3 | 18.4 |
| 15 | 15 | ThoughtSpot Analytics | 50 | 4.5 | 17.6 |
| 16 | 16 | Zoho Analytics | 44 | 4.5 | 17.0 |
| 17 | 20 | Databricks | 31 | 4.9 | 16.8 |
| 18 | 17 | Pyramid | 44 | 4.3 | 16.3 |
| 19 | 21 | Tellius | 31 | 4.7 | 16.1 |
| 20 | 18 | Birst | 35 | 4.4 | 15.6 |
| 21 | 19 | Logi Symphony | 35 | 4.4 | 15.6 |
| 22 | 22 | Yellowfin | 24 | 4.6 | 14.6 |
| 23 | 24 | FineReport | 19 | 4.9 | 14.4 |
| 24 | 25 | Alibaba Quick BI | 18 | 4.7 | 13.6 |
| 25 | 26 | Sigma | 17 | 4.8 | 13.6 |
| 26 | 23 | Board | 25 | 4.1 | 13.2 |
| 27 | 27 | ibi WebFOCUS | 18 | 4.4 | 12.7 |
| 28 | 30 | Alteryx | 13 | 4.9 | 12.6 |
| 29 | 29 | Incorta | 16 | 4.5 | 12.5 |
| 30 | 28 | Minitab Connect | 18 | 4.3 | 12.4 |

This new ranking takes into account not only the quality of a business intelligence tool, but also its popularity.

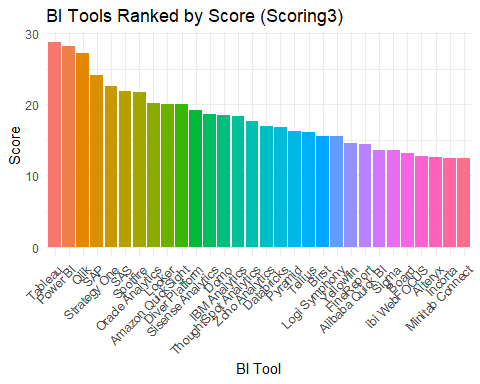
For example:

* Tableau will always score higher than Power BI because they have the same rating, but also because it has more reviews
* Spotfire scores higher and ranks better than Oracle Analytics because it has a better rating, even though it has slightly fewer reviews. This difference in numbers will not make a big difference to Spotfire’s final rating, unless Spotfire is in turmoil or they release a bad update to the tool.
* Oracle Analytics, Looker, and Amazon QuickSight have very similar scores. The difference lies in their popularity and ratings. But it’s a very close call.

# 5. Plotting the Final Ranking

Plotting again the final ranking taking into account the reviews and the logarithm of popularity this time.

library(ggplot2)  
  
# Optional: reorder tool factor based on final rank (smallest = top)  
df$Tool <- factor(df$Tool, levels = df$Tool[order(df$`Rank: Scoring3`)])  
  
# Bar chart  
ggplot(df, aes(x = Tool, y = `Scoring3`, fill = Tool)) +  
 geom\_col(show.legend = FALSE) +  
 labs(  
 title = "BI Tools Ranked by Score (Scoring3)",  
 x = "BI Tool",  
 y = "Score"  
 ) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# 6. Exporting to Excel

library(writexl)  
write\_xlsx(df, "BI\_Tools\_Scoring\_Comparison.xlsx")