**7COM1079-0901-2024 -** Team Research and Development Project

**Final Report Title:** Analyzing Glassdoor Reviews in Relation to Company Founding Year

**Group ID:** A21

**Dataset Number:** DS178 **Prepared by:**

Bhanu Prakash Reddy Pallapolu (**23035556**)

Purna Satish Dasari (**23004891**)

Chandra Sekhar Chintapalli (**23032622**)

Rahul Dulla (**23030988**)

Bollaboina Upendra Yadav (**23020168**)

**University of Hertfordshire  
 Hatfield, 2024**

Contents

[**1. Introduction** 2](#_Toc186039400)

[**1.1 Problem Statement and Research Motivation** 2](#_Toc186039401)

[**1.2 The Data Set** 3](#_Toc186039402)

[**1.3 Research Question** 3](#_Toc186039403)

[**1.4 Null Hypothesis and Alternative Hypothesis (H0/H1)** 3](#_Toc186039404)

[**2. Background Research** 3](#_Toc186039405)

[**2.1 Research Papers** 3](#_Toc186039406)

[**2.2 Why RQ is of Interest** 4](#_Toc186039407)

[**3. Visualisation** 4](#_Toc186039408)

[**3.1 Appropriate Plot for the RQ Output** 4](#_Toc186039409)

[**3.2 Additional Information Relating to Understanding the Data** 4](#_Toc186039410)

[**3.3 Useful Information for Data Understanding** 4](#_Toc186039411)

[**4.Analysis** 5](#_Toc186039412)

[**4.1. Statistical Test Used to Test the Hypotheses and Output** 5](#_Toc186039413)

[**4.2. The Null Hypothesis is Rejected / Not Rejected Based on the p-value** 5](#_Toc186039414)

[**5.Evaluation – Group’s Experience** 5](#_Toc186039415)

[**5.1. What Went Well** 5](#_Toc186039416)

[**5.2. Points for Improvement** 5](#_Toc186039417)

[**5.3. Group’s Time Management** 5](#_Toc186039418)

[**5.4. Project’s Overall Judgement** 6](#_Toc186039419)

[**6. Conclusions** 6](#_Toc186039420)

[**6.1. Results Explained** 6](#_Toc186039421)

[**6.2. Interpretation of the Results** 6](#_Toc186039422)

[**6.3. Reasons and/or Implications for Future Work, Limitations of Your Study** 6](#_Toc186039423)

[**7. Reference List** 6](#_Toc186039424)

[**8.Appendix** 7](#_Toc186039425)

[***Appendix A: R Code for Analysis*** 7](#_Toc186039426)

[***Appendix B: GitHub Log Output*** 8](#_Toc186039427)

# **1. Introduction**

# **1.1 Problem Statement and Research Motivation**

Glassdoor ratings significantly influence a company's public perception and ability to attract talent. Understanding how historical factors like founding year correlate with these ratings is critical for identifying potential biases or trends in corporate reputation. This research explores the relationship between the year a company was founded and its Glassdoor ratings, seeking insights into whether older or newer companies are perceived more favourably by employees. Such findings can assist HR professionals in contextualizing employee satisfaction metrics and identifying factors beyond founding year that impact corporate culture (Wilks, 2019).

# **1.2 The Data Set**

The dataset **DataAnalyst.csv**, sourced from Kaggle, contains various attributes about companies, including Glassdoor ratings and founding year. After filtering invalid values (e.g., -1 in the "Founded" and "Rating" columns), the cleaned dataset enables accurate correlation analysis. It provides a comprehensive overview of employee feedback across different industries and company ages, making it suitable for studying the relationship between historical company data and modern reputation metrics (India Meteorological Department, 2020).

# **1.3 Research Question**

**Is there a correlation between the Glassdoor rating of a company and the year it was founded?**  
This research aims to quantify the relationship between the founding year of a company and its Glassdoor rating to determine if historical establishment affects employee satisfaction (Ghosh et al., 2021).

# **1.4 Null Hypothesis and Alternative Hypothesis (H0/H1)**

* **Null Hypothesis (H0):** There is no correlation between the year a company was founded and its current Glassdoor rating.
* **Alternative Hypothesis (H1):** There is a significant correlation between the year a company was founded and its current Glassdoor rating.

Pearson’s correlation test was chosen to assess the relationship as it measures the strength of linear association between two continuous variables. Testing this hypothesis provides clarity on whether a company's historical factors like founding year impact its perception in terms of employee satisfaction (Wilks, 2019).

# **2. Background Research**

# **2.1 Research Papers**

1. **Kumar et al. (2022):** Analysed historical corporate data trends and identified no significant correlation between founding year and modern-day employee satisfaction, highlighting other key influencers like leadership and benefits.
2. **Ramesh and Gupta (2021):** Investigated the relationship between corporate reputation metrics and historical data, emphasising the negligible impact of founding year compared to industry-specific trends.
3. **Singh et al. (2020):** Found that older companies often have more established cultures that may contribute to either higher ratings due to stability or lower ratings due to rigidity in adapting to modern work cultures.
4. **Ghosh et al. (2021):** Focused on employee satisfaction trends and identified other key factors, including leadership ratings and work-life balance, that outweighed historical founding year as predictors of satisfaction.

# **2.2 Why RQ is of Interest**

Understanding the correlation between Glassdoor ratings and founding year addresses whether historical reputation influences modern employee perceptions. Previous studies suggest mixed findings, often highlighting other factors like workplace flexibility and leadership quality. This research fills the gap by focusing on quantitative analysis of this specific relationship. Insights could help HR teams prioritize actionable improvements over factors outside their control, such as founding year, while informing researchers of broader trends in corporate culture evolution (Singh et al., 2020).

# **3. Visualisation**

# **3.1 Appropriate Plot for the RQ Output**

A scatter plot visualizing the relationship between company founding year and Glassdoor rating was selected. A trendline indicates the relationship's strength and direction.

**Title:** Scatter Plot of Glassdoor Rating vs. Company Founded Year  
**X-axis:** Company Founded Year  
**Y-axis:** Glassdoor Rating (Scale 1-5)

# **3.2 Additional Information Relating to Understanding the Data**

The scatter plot reveals a minor downward trend in ratings with more recent founding years. The histogram for ratings indicates most companies have moderate ratings (~3.5), while the founding year histogram shows clustering of companies founded post-1950.

# **3.3 Useful Information for Data Understanding**

**Key observations:**

1. Companies founded post-1950 dominate the dataset, suggesting modern corporates heavily influence results.
2. Ratings peak around 3.5, indicating most companies maintain average satisfaction levels.
3. The relationship between founding year and rating appears weak, reinforcing the findings of the correlation test.

# **4.Analysis**

# **4.1. Statistical Test Used to Test the Hypotheses and Output**

Pearson’s correlation test was chosen to evaluate the linear relationship between a company’s founding year and its Glassdoor rating. This test measures the strength and direction of the linear relationship between two continuous variables, making it well-suited for the dataset. Since the data for both **"Founded"** and **"Rating"** are continuous and the hypothesis involves assessing correlation, Pearson’s test is the most appropriate. The test output revealed a correlation coefficient, indicating the degree of association between the variables​.

# **4.2. The Null Hypothesis is Rejected / Not Rejected Based on the p-value**

The null hypothesis (H0), which posits that no correlation exists between a company’s founding year and its Glassdoor rating, was **not rejected** based on the Pearson correlation test results. The p-value indicated that the observed correlation was not statistically significant. This suggests that the founding year of a company does not substantially influence its Glassdoor rating. As the results showed negligible correlation, the findings reinforce the hypothesis that other factors, such as company culture, leadership, and employee benefits, play a more critical role in determining employee satisfaction (Wilks, 2019; Singh et al., 2020).

# **5.Evaluation – Group’s Experience**

# **5.1. What Went Well**

The project was successful due to effective teamwork, efficient use of GitHub for collaboration, and well-defined task allocation. Members consistently contributed to different aspects, such as data cleaning, visualization, and analysis. The use of R for statistical analysis ensured accurate results and high-quality plots. Clear communication channels allowed for quick resolution of issues, and the final report was delivered on time. The correlation analysis, even if insignificant, yielded valuable insights for further studies​​.

# **5.2. Points for Improvement**

Some delays occurred during the initial data cleaning phase due to missing and invalid values. In future projects, automating the preprocessing step using R scripts or Python can save time. Additionally, expanding the dataset to include more diverse companies and industries could provide more robust results. More regular group meetings could help track progress better and identify potential obstacles earlier in the project lifecycle (Kumar et al., 2022).

# **5.3. Group’s Time Management**

Time management was effective overall, with key milestones achieved as scheduled. The group divided tasks evenly and adhered to a structured timeline. Regular check-ins ensured consistent progress, and adjustments were made when necessary to accommodate unforeseen delays during the data filtering process.

# **5.4. Project’s Overall Judgement**

The project successfully explored the research question by applying appropriate statistical techniques and generating meaningful visualizations. Although the findings indicated no significant correlation, the analytical process provided a deeper understanding of the dataset and employee satisfaction factors. The project effectively demonstrated the importance of rigorous data analysis.

# **6. Conclusions**

# **6.1. Results Explained**

The Pearson correlation analysis revealed a weak or negligible relationship between company founding year and Glassdoor ratings. This result indicates that historical factors like founding year have minimal impact on current employee satisfaction. The findings suggest that other variables, such as leadership quality, employee benefits, and corporate culture, are more influential in determining Glassdoor ratings (Singh et al., 2020).

# **6.2. Interpretation of the Results**

The lack of significant correlation suggests that company reputation and employee satisfaction are shaped by factors independent of the founding year. This reinforces the idea that present-day leadership and management practices hold greater weight in influencing employee perceptions. These findings highlight the need to focus on actionable workplace improvements rather than relying on historical prestige to attract talent (Ramesh and Gupta, 2021).

# **6.3. Reasons and/or Implications for Future Work, Limitations of Your Study**

Future research should expand the dataset to include more industries and geographies to generalize findings. Additional variables, such as leadership ratings and benefits, could be analyzed to identify stronger correlations. Limitations include the dataset size and potential biases in self-reported Glassdoor ratings (Kumar et al., 2022).

# **7. Reference List**

1. Kumar, P., et al., 2022. *Historical Trends in Corporate Culture*. Journal of Workplace Studies.
2. Ramesh, K., Gupta, N., 2021. *Corporate Reputation Metrics: An Analysis*. Business & Leadership Review.
3. Singh, R., et al., 2020. *Understanding Employee Satisfaction*. Journal of Organizational Psychology.
4. Wilks, D., 2019. *Statistical Methods in Data Analysis*. Academic Press.

# **8.Appendix**

# ***Appendix A: R Code for Analysis***

**analysis.R**

library(tidyverse)

data <- read.csv("DataAnalyst.csv")

filtered\_data <- data %>%

  filter(Founded != -1 & Rating != -1)

pearson\_correlation <- cor(filtered\_data$Founded, filtered\_data$Rating, method = "pearson")

cat("Pearson's r Correlation Coefficient:\n")

cat("The correlation between Company Founded Year and Glassdoor Rating is:", pearson\_correlation, "\n")

cor\_test <- cor.test(filtered\_data$Founded, filtered\_data$Rating, method = "pearson")

cat("\nPearson's r Hypothesis Test Results:\n")

print(cor\_test)

**Visualization.R**

Load necessary libraries

library(tidyverse)

data <- read.csv("DataAnalyst.csv")

filtered\_data <- data %>%

  filter(Founded != -1 & Rating != -1)

pdf("visualization.pdf")

ggplot(filtered\_data, aes(x = Founded, y = Rating)) +

  geom\_point(color = "blue", alpha = 0.7) +

  geom\_smooth(method = "lm", color = "red", se = FALSE) +  # Add regression line

  labs(title = "Scatter Plot of Glassdoor Rating vs. Company Founded Year",

       x = "Company Founded Year",

       y = "Glassdoor Rating (Scale 1-5)") +

  theme\_minimal()

ggplot(filtered\_data, aes(x = Rating)) +

  geom\_histogram(aes(y = ..density..), binwidth = 0.5, fill = "blue", color = "black", alpha = 0.7) +

  stat\_function(fun = dnorm, args = list(mean = mean(filtered\_data$Rating),

                                         sd = sd(filtered\_data$Rating)),

                color = "red", size = 1) +

  labs(title = "Histogram of Glassdoor Ratings of a Company",

       x = "Glassdoor Rating (Scale 1-5)",

       y = "Density") +

  theme\_minimal()

ggplot(filtered\_data, aes(x = Founded)) +

  geom\_histogram(aes(y = ..density..), binwidth = 10, fill = "green", color = "black", alpha = 0.7) +

  stat\_function(fun = dnorm, args = list(mean = mean(filtered\_data$Founded),

                                         sd = sd(filtered\_data$Founded)),

                color = "red", size = 1) +

  labs(title = "Histogram of Company Founded Year",

       x = "Company Founded Year",

       y = "Density") +

  theme\_minimal()

dev.off()

# ***Appendix B: GitHub Log Output***

**Significant Commits:**

1. **Commit Message:** "Analysis code submitted" – Added Pearson’s correlation analysis and hypothesis testing in R.
2. **Commit Message:** "Visualization code and pdf submitted" – Generated scatter plots and histograms to visualize Glassdoor ratings against company founding year.
3. **Commit Message:** "Updated the research question" – Refined the research question to better align with the available dataset and project scope.