

A Study on the Attributes of Companies, and their Effect on the Overall Rating on Glassdoor

Erin M. Swan-Siegel

Northwest Missouri State University, Maryville MO 64468, USA
S201660@nwmissouri.edu

Abstract. Keywords: business · employee satisfaction · company culture · company rating

1 Introduction

Every day, we are bombarded with products and businesses touting the product's utility or the benefits of their particular service through qualitative or quantitative summary statements; declarations such as "9 out of 10 doctors recommend", "4.5 stars with over twenty-thousand reviews!", and "Voted the best in the Midwest". We consumers rely on these ratings and reviews to help us decide on everything from where to eat, what product to buy, where to spend our vacation, and even where to work.

But what special mixture of characteristics contribute to a higher rating, when determined by both past and present employees - the longevity of the company; whether the company answers to shareholders or their cause; the size of the employee base, perhaps? Furthermore, can one predict a company's rating based on key characteristics? Additionally, what business values are present in a job description for that company when the Glassdoor rating is high?

A Glassdoor company profile is created when the first self-reported past or present employee submits a review on the platform. Companies can create a free profile, allowing them to maintain certain aspects of the associated information card.[3] A company's overall rating is a calculated average of all review submissions.

Limitations to this study include the fact that a company's Glassdoor rating is not based on *every* past or present employee - just the subset of self-reported employees who decided to participate. These reviews can also be skewed / affected by employees who are new to the company, leave the company under stressful circumstances, or have recently been denied a promotion. Similarly, company data gathered is initially populated by a self-identifying employee and may or may not have been viewed or updated by an official company representative, the result of which could be incorrect attribute values.[3] In summary, the context and attributes of those contributing to the rating cannot be accounted for or controlled, and must therefore be taken with the common-sense understanding that many consumers do for other location, service, or product review page.

1.1 Associated Literature

Recruiting and retaining top talent is one of the major drives of any business that aims to be successful. Said another way, companies with the ability to keep their employees satisfied have a greater chance of performing well. This satisfaction - like many human experiences - is determined by a multitude of factors including the ability to advance, work-life balance, and the company's overall culture and values.[2],[1]

A company's ability to hit the common criteria for happy employees can be facilitated by higher employee base or increased revenues; for others, it can be a reason for hampering it.

1.2 Goals of this Study

The aim of this study is to predict a company's rating when a data Machine Learning model is provided six key characteristics. Similarly, what key words are unique amongst more highly-rated company's job description - correlating employer ideals with employee satisfaction.

1.3 Methodology

In order to answer the questions posed by this study, company data was scraped from the job platform, Glassdoor. The resulting data frame was further populated with company data scraped by other individuals analyzing Data Science job descriptions on the same platform.

Six key company features - Global Size, Company Industry, Company Headquarters, Company Ownership Type, Year Founded, and Estimated Yearly Revenue - were independent variables identified as those potentially contributing to the company's rating. Records missing one or multiple of the key attributes were removed as one part of the data cleaning process.

[Data was subsequently analyzed (r-values) and a data model was determined. The machine learning training and test sets were. Using SpaCy (NLP), keywords from the company description for the employers with a rating of 3.5 and above were determined.]

2 Data

Data for this study was intended to be gathered solely by a web-scraper, built together with Software Engineer, Daniel Swan. Using code from a Kaggle contributor as a base, the resulting scraper extracts company data from an employer card found on a particular Glassdoor site, clicks into each company's Glassdoor employer page and extracts additional data. Unfortunately for this study, platforms are continuing to build better protections over their systems and their data, and the web scraper built for this study - while it does work as expected - encountered scraping prevention measures.

Web scraping prevention exists for many reasons, including blocking excess traffic that slows down the site for non-bot users and for protecting a company's intellectual property[4]. Options to bypass these security measures are accessible, but required significant re-work of the author's script or a research partnership with the Glassdoor platform itself. With a similar output resource available, company data published to Kaggle in 2020 was used to supplement the web-scraped data.

2.1 Cleaning

Data in the tenth field of the web-scraped output file was parsed into eight additional features using a delimiter. Before the generated and the procured data were joined, the already-cleaned Kaggle set was further processed for this study by removing unnecessary fields, removing the rating from the company name, and using Excel to reduce the records to only those with unique attribute values. Using the approximately 8000 company name values, a V-lookup with the uncleaned Kaggle data allowed for the first respective Job Description to substitute for the "Company Description" attribute. Certain UTF-8 substitutions were identified and reverted to their ASCII counterpart. These steps were completed using MicroSoft Excel to facilitate the desire for permanent alterations, as well as the author's ability to perform the tasks more quickly and precisely.

Following the guidelines and advice of YouTuber Ken Jee[5], in conjuncture with the Visual Studio plugin, Data Wrangler, additional cleaning and processing occurred within Visual Studio, including separating the Company Headquarters state (or country) from the city. Records containing filler values such as "unknown" and "-1" were removed to leave a final data set of approximately 5200 records, with the following set of attributes.

Attribute	Description
Company Name	Name of the company reviewed
Global Size	Number of employees across all of the companys' locations, as reported by current or past employees who leave a review on the platform
Industry	The general industry the company is reported to operate
Description	Job Description data, which often includes company description; this data will be processed through a Natural Language Processor to generate a list of key words which can be associated with a company's rating
Company Rating	The average rating, reported on a scale of 1 to 5, as determined by current and past employees who leave a review on the platform
Headquarters	State (or Country) reported as the reviewed company's main location of administrative operations
Company Ownership Type	Indicates the company's designation as a privately-held, publicly-held, non-profit, educational institute, etc
Year Founded	The year the company is reported to have been founded; this data will be used to calculate the age of the company reviewed
Est Revenue	Estimated yearly revenue of the company, as reported by current or past employees who submit a review on the platform

2.2 Exploration

3 Model

3.1 Type Selection

3.2 Building

3.3 Training and Testing

4 Results

4.1 Summary

4.2 Visualizations

5 Discussion

5.1 Conclusion

5.2 Limitations

5.3 Future Exploration

References

1. Glassdoor: Glassdoor research: Satisfaction drivers remain intact, <https://www.glassdoor.com/research/satisfaction-drivers-remain-intact>
2. Glassdoor: Glassdoor research: Satisfied workers stay, <https://www.glassdoor.com/research/satisfied-workers-stay>
3. Glassdoor: How did my company get on glassdoor?, <https://stage-help.glassdoor.com>
4. GPT, C.: Why websites prevent data scraping, <https://chat.openai.com/>
5. Jee, K.: Data science project from scratch, https://www.youtube.com/@KenJee_ds