**Description**

The dataset we will be working with is a compilation of employee reviews for Google, Amazon, Apple, Microsoft and Netflix, which were collected by web scraping Glassdoor. Glassdoor is a website where (former) employees can leave anonymous reviews of their employers. For example, reviews often entail employees’ opinions on management, company culture, and company policies. The dataset contains over 67,000 reviews described across 16 variables, which are elaborated on below. It is worth noting that the data type for each variable describes its original form, which might not be the accurate characterisation.

**Variables**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Data Type** | **Description** |
| X | Integer | A count for each review |
| company | Factor | The name of the company |
| location | Factor | The location of where the employee worked in text, described in the format of ‘city, state’. |
| dates | Factor | The date on which the review was created. Written in the format of ‘month date, year’ |
| job-title | Factor | Describes whether the employee is a current or former employee and the job title of the employee, in the format ‘status employee, job title’ The reviewer may elect to choose to keep their title anonymous. |
| summary | Factor | A text summary of the employee’s review |
| pros | Factor | Describes in text the employee’s opinion of the positive aspects of working at the company |
| cons | Factor | Describes in text the employee’s opinion of the negative aspects of working at the company |
| advice-to-mgmt | Factor | A text of the employee’s advice to management for improving their practices |
| overall-ratings | Numerical | A numerical rating for the employee’s overall impression of working at the company. Rated on a scale from 1-5, where 1 is the lowest score and 5 is the highest score. |
| work-balance-stars | Factor | The individual employees’ rating of how good they found the work-life balance to be. Expressed as a rating from 1-5, where 1 exhibits a low work-life balance, and 5 a high (positive) work-life balance. |
| culture-values-starts | Factor | A numerical rating for the employee’s impression of the culture and values of the company. Rated on a scale from 1-5, where 1 is the lowest score and 5 is the highest score. |
| carrer-opportunities-stars | Factor | A numerical rating for the employee’s impression of the career opportunities at the company. Rated on a scale from 1-5, where 1 is the lowest score and 5 is the highest score. |
| comp-benefit-stars | Factor | A numerical rating for the employee’s impression of the compensation and benefits of the company. Rated on a scale from 1-5, where 1 is the lowest score and 5 is the highest score. |
| senior-mangemnet-stars | Factor | A numerical rating for the employee’s impression of the competency of their senior management at the company. Rated on a scale from 1-5, where 1 is the lowest score and 5 is the highest score. |
| helpful-count | Integer | A score of how many readers that have clicked that they found that particular review helpful |
| link | Factor | The URL to the particular review |

**Potential Questions**

* Which company has the happiest employees?
* What are the most prominent positive aspects of working for each company?
* What are the most prominent negative aspects of working for each company?
* Are anonymous (hidden job title) reviews more negative than non-anonymous reviews?
* Does the sentiment of reviews differ between locations?
  + i.e. are people more happy working at Google’s headquarters in Mountain View, than in alternative locations?
* Are we able tell how positive (negative) the review would be solely based on the summary of the review?
* Are there specific types of employees (job roles) which are more satisfied than others?