

Introduction of the Problem

Our project seeks to comprehensively analyze the Big 4 firms — Deloitte, PwC, KPMG, and EY — delving into employee ratings, company culture, benefits, and overall sentiment, recognizing the widespread interest in prestigious organizations like the Big 4 among job seekers. We aim to address the need for differentiation between these firms and demystify their distinctions, offering valuable guidance to students and professionals who are considering career opportunities within these organizations. Through thorough analysis of employee feedback and other relevant factors, our dataset, focused specifically on the Big 4, illuminates their unique attributes, striving to offer clarity amidst the complexities of the Big 4 landscape and better inform individuals' career decisions.

Analytics approach

Our comprehensive analytics approach is applied to analyze the Big 4 firms, utilizing various techniques to extract valuable insights from employee feedback. Within our dataset, we specifically focus on categories such as the following ratings: 'work-life balance', 'culture values', 'diversity and inclusion', 'career opportunities', 'compensation and benefits', 'senior management', 'overall rating', and other significant columns such as: 'firms', a review of positive facts in 'pros', a review of negative facts in 'cons', and a general comment in 'headline'. This allows us to perform a comprehensive analysis of employee sentiments, company culture, and other pertinent factors. By employing sentiment analysis, regression analyses, topic modeling, and personalized recommendation systems, our aim is to acquire an in-depth understanding of the nuances within each firm and deliver practical insights to individuals exploring career opportunities.

Sentiment Analysis using VADER

During the initial phase of our project, we began on a detailed examination of the sentiments expressed within employee reviews of the Big 4 firms. Recognizing the value of consolidating feedback, we merged the 'pros' and 'cons' columns from the dataset into a unified column. By merging these columns, we created a more cohesive dataset that facilitated easier comparison and interpretation of the sentiment analysis results. This method enabled us to execute a comprehensive assessment of employee perspectives, uncovering broader trends and patterns within their feedback. Leveraging the VADER lexicon for sentiment analysis, we computed compound sentiment scores for each review within the consolidated column text. By aggregating these scores on a per-firm basis, we gained insights into the prevailing sentiments towards each Big 4 company. As a result, the merged column sentiment trended towards positivity, scoring around 0.45, with Deloitte emerging as the frontrunner. Refer to **Table 1** in the Appendix for details.

In parallel, we conducted regression analyses to explore the relationships between various factors. One analysis examined how ratings across various categories (e.g., work-life balance, cultural values) affect headline sentiment. It used headline sentiment as the dependent variable and ratings as the independent variables, shedding light on their impact on the perceptions conveyed in review headlines. Another analysis sought to uncover these same ratings influence the overall rating to support the insights gathered by the previous regression. For further details, please consult **Table 2** and **3** located in the Appendix. Additionally, we delved into the sentiments conveyed through the headlines of employee reviews. Similar preprocessing and sentiment analysis techniques were applied to the 'headline' column, allowing us to gauge the overall sentiment expressed in review headlines for each firm. Refer to **Table 4** in the Appendix for details.

As a result, our analysis was focused on comparing the sentiment derived from the combined pros and cons with that of headlines alone. On average, we discovered that the sentiment from pros and cons was generally more positive than the sentiment expressed in the headlines. Then, conducting regression

analysis, we determined that culture values and career opportunities were identified as the most influential factors impacting sentiment and overall scores. These findings suggest that companies should prioritize these areas to enhance both sentiment and overall ratings.

Identifying Dominant Topics using LDA

During our secondary analysis, we employed Latent Dirichlet Allocation (LDA) to unveil prevalent themes within the 'pros' and 'cons' sections of employee feedback across the four firms. Each firm's dataset underwent individual LDA training, configuring the model to identify five distinct topics, each comprised of four representative words. Our investigation uncovered unique trends across the firms, with some highlighting positive aspects such as supportive work environments and opportunities for professional advancement, while others voiced concerns regarding salary levels and the intensity of workloads. This comprehensive approach provided valuable insights into the diverse experiences of employees within the Big 4 firms, offering insights into both their strengths and areas for enhancement. For detailed findings, please refer to **Table 5** in the Appendix.

Personalized Recommendation System with Cosine Similarity Analysis

In the final phase of our analysis, we aimed to provide personalized recommendations for individuals seeking employment at Big 4 firms based on their specified criteria of what they are looking for in a working place. To achieve this, we employed a matching algorithm that combines user-defined preferences with the pros mentioned in employee reviews. To enhance comparability and maintain consistency in our analysis, we curated the dataset by selecting only 1,000 reviews from each firm and filtering out reviews with fewer than 40 characters. This approach allowed us to focus on meaningful insights and improve the accuracy of our recommendations. We then utilized a bag-of-words representation and pairwise cosine similarity scores to quantify the similarity between the user-provided criteria and the pros extracted from the reviews. By sorting the results based on similarity scores and selecting the firm that appeared the most in the top 100 matches, we can provide users with a tailored recommendation. For instance, if a user inputs criteria like "Good salary, friendly culture, work-life balance, opportunities, benefits, and training", our system will suggest Deloitte as the most suitable option, considering both their preferences and our analysis.

Project Impact

While previous analyses have examined employee ratings and company culture, our approach is distinct in its focus on the Big 4 firms, aiming to unveil nuanced insights that set these organizations apart. While certain analyses may have touched upon similar themes, our project stands out by integrating a range of techniques, including sentiment analysis, topic modeling, and personalized recommendation systems. Our analysis holds significant potential impact such as guiding strategic decisions within Big 4 firms, improving employee satisfaction, fostering better company culture, and attracting top talent. Economically, optimizing satisfaction can enhance productivity and reduce turnover costs, while societally, it promotes transparent and equitable workplaces. Given the prominence of Big 4 firms, our work addresses the challenge of distinguishing between them. By providing comprehensive analyses of employee ratings, culture, and benefits, we empower individuals with insights for informed career decisions. Our aim is to offer clarity and guidance in navigating careers within the competitive Big 4 landscape, contributing to professional development and success.

Project Reflection and Further Enhancement

Upon reflecting on our project, it is important to highlight that our analysis goes beyond the scope of big 4 firms and explores more than just compatibility between companies and employees. Considering the current high rates of unemployment, there's an intriguing opportunity for companies to match job seekers based on their resumes. For instance, imagine a scenario where a tech company seeking innovative talent could utilize our comprehensive resume data to identify individuals with the exact skills and experiences

they need. By leveraging our analysis, these companies could efficiently match with potential candidates who possess the precise qualifications, expectations and attributes required for their specific roles.

Appendix

Table 1: Overall Sentiment on Pros/Cons Column

Firm	Sentiment Score
Deloitte	0.4739
PwC	0.4565
KPMG	0.4320
EY	0.4318

Table 2: Scores Impact on Headline Sentiment

Category	Weight
Work Life Balance	0.0999
Culture Values	0.1762
Diversity Inclusion	0.0038
Career opportunity	0.1582
Compensation & Benefits	0.0516
Senior Management Rating	0.0770

Table 3: Scores Impact on Overall Rating

Category	Weight
Work Life Balance	0.1546
Culture Values	0.3088
Diversity & Inclusion	0.0038
Career Opportunity	0.2980
Compensation & benefits	0.1667
Senior Management Rating	0.2045

Table 4: Overall Sentiment on Headline Column

Firm	Sentiment Score
Deloitte	0.2721
PwC	0.2736
KPMG	0.2568
EY	0.2538

Table 5: Dominant Topics

Big 4 Firm	Pros +	Cons -
Deloitte	Positive work environment, friendly culture, career development, learning opportunities	Salary, career growth, long working hours, project management
PwC	Diverse project offerings, client exposure, pay, salary, career growth	Demanding work hours, workload intensity, dissatisfaction with compensation, teamwork
KPMG	Client service excellence, positive workplace culture, diverse opportunities, growth, work-life balance	Intensity of work during busy seasons, career progression, compensation, work-life balance, achieving equilibrium
EY	Exposure, pay, benefits, brand reputation, friendly workplace atmosphere	High working hours, workplace culture, low pay