

DEV-I PROJECT

REPORT

GitHub Link:

<https://github.com/parthxnarula/045039-ParthNarula.git>

Name- Parth Narula

Roll no.- 045039

Section- H

Introduction:

The **AmbitionBox** Data Analysis project aims to extract, analyze, and derive valuable insights from data collected from **ambitionbox.com**. The website provides information about various companies, including **company ratings, reviews, and other relevant attributes**. This project seeks to provide meaningful conclusions and recommendations based on the collected data.

Project Objectives:

1. Data Collection:

- Retrieve comprehensive data from ambitionbox.com, including company names, ratings, reviews, and company type information.
- Ensure data integrity and completeness during the scraping process.

2. Data Preparation:

- Clean and preprocess the scraped data to eliminate duplicates, inconsistencies, and missing values.
- Format data types for ease of analysis.

3. Exploratory Data Analysis (EDA):

- Conduct an initial exploratory analysis to gain insights into the dataset.
- Explore data distribution, summary statistics, and visualize key trends.

4. Rating Analysis:

- Calculate and analyze the average, median, mode, highest, and lowest company ratings.
- Identify companies with the highest and lowest ratings to understand what sets them apart.

5. Company Type Analysis:

- Categorize companies into various types like Healthcare, Finance, Manufacturing, Consulting and Services and to generate their proportions from the data of the website scrapped.

7. Company Age Analysis:

- Calculate the age of each company based on available data.
- Identify the oldest and youngest companies.

8. Visualization:

- Create insightful visualizations, such as bar charts, scatter plots, and word clouds, to represent findings effectively.
- Use visualizations to communicate key insights and trends.

General Description of Data:

Data was collected using web scraping techniques, leveraging Python libraries such as BeautifulSoup and requests. The following key data points were extracted from the website:

1. **Company Name:** The names of the companies featured on AmbitionBox.com.
2. **Company Type:** The industry or sector to which each company belongs.
3. **Ratings:** Aggregate ratings for each company based on employee reviews. These ratings provide insights into the overall perception of the company.
4. **Total Reviews:** The total number of employee reviews available for each company.
5. **Company Age:** The number of years the company has been in existence since its establishment.
6. **Number of Employees:** The number of employees for each company was also collected during the web scraping process. This data point provides valuable information about the size and scale of each company in the dataset.
7. **Location:** During the web scraping process, location data for each company was collected. This information provides insights into the geographic distribution of companies within the dataset.

	Name	Rating	Reviews	Company Type	No. of Employees	Age of Company	Location
0	TCS	3.8	63.1k Reviews	IT Services & Consulting	1 Lakh+ Employees	55 years old	Mumbai +281 more
1	Accenture	4.1	40.3k Reviews	IT Services & Consulting	1 Lakh+ Employees	34 years old	Dublin +141 more
2	Cognizant	3.9	36.7k Reviews	IT Services & Consulting	1 Lakh+ Employees	29 years old	Teaneck, New Jersey. +111 more
3	Wipro	3.8	33.4k Reviews	IT Services & Consulting	1 Lakh+ Employees	78 years old	Bangalore/Bengaluru +246 more
4	ICICI Bank	4.0	29.7k Reviews	Banking	1 Lakh+ Employees	29 years old	Mumbai +1184 more
5	HDFC Bank	3.9	29.3k Reviews	Banking	1 Lakh+ Employees	29 years old	Mumbai +1422 more
6	Infosys	3.9	27.6k Reviews	IT Services & Consulting	1 Lakh+ Employees	42 years old	Bengaluru/Bangalore +135 more

7	Capgemini	3.9	25.6k Reviews	IT Services & Consulting	1 Lakh+ Employees	56 years old	Paris +87 more
8	Tech Mahindra	3.7	24.1k Reviews	IT Services & Consulting	1 Lakh+ Employees	37 years old	Pune +235 more
9	HCLTech	3.7	23.2k Reviews	IT Services & Consulting	1 Lakh+ Employees	32 years old	Noida +153 more
10	Genpact	4.0	23.1k Reviews	IT Services & Consulting	50k-1 Lakh Employees	26 years old	New York +78 more
11	Axis Bank	3.8	19.6k Reviews	Banking	50k-1 Lakh Employees	30 years old	Mumbai +1221 more
12	Concentrix Corporation	3.9	18.8k Reviews	BPO	50k-1 Lakh Employees	40 years old	Fremont +84 more
13	Amazon	4.2	17.9k Reviews	Internet	1 Lakh+ Employees	29 years old	Seattle +394 more
14	IBM	4.1	17.7k Reviews	Software Product	50k-1 Lakh Employees	112 years old	Armonk +131 more

Analysis Report: Basic Descriptive & Mathematical/Statistical analysis

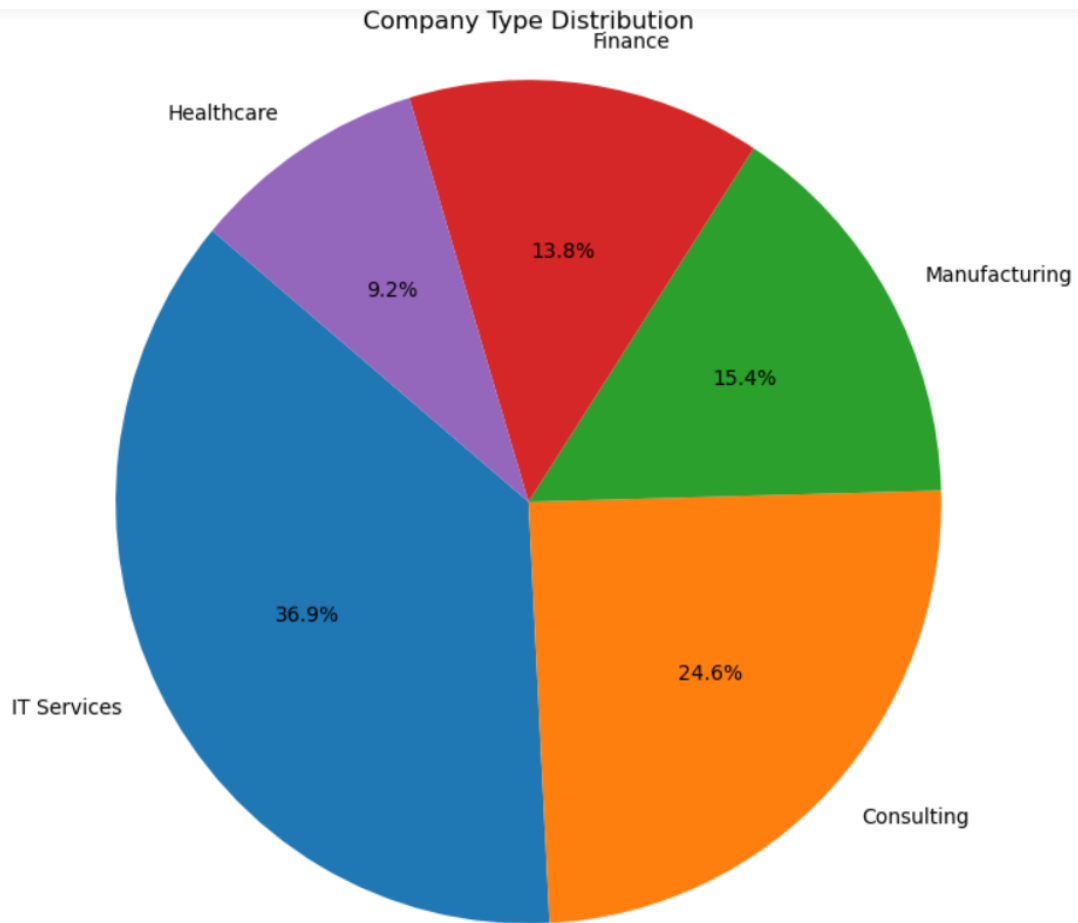
1.) Mathematical/Statistical Analysis

- **Mean for Company Ratings:** The average rating across all companies in the dataset was found to be approximately 4. This suggests a relatively positive sentiment among employees.
- **Mode for Company Ratings:** The most common rating among companies was 3.9.
- **Median for Company Ratings:** The median rating was also 3.9.
- **Mean for Company Age:** The average age of companies in the dataset was approximately 42.1years.
- **Variance for Company Age:** 261.88
- **Standard Deviation for Company Age:** 16.18
- **Correlation Analysis:**
 - Rating vs. Review Count: A positive correlation was observed between company ratings and the number of reviews, suggesting that companies with more reviews tend to have higher ratings.
 - Rating vs. Employee Count: No significant correlation was found between company ratings and the number of employees.
 - Rating vs. Company Age: A weak negative correlation was identified between company ratings and age, indicating that newer companies might have slightly higher ratings.

2.) Other Analysis

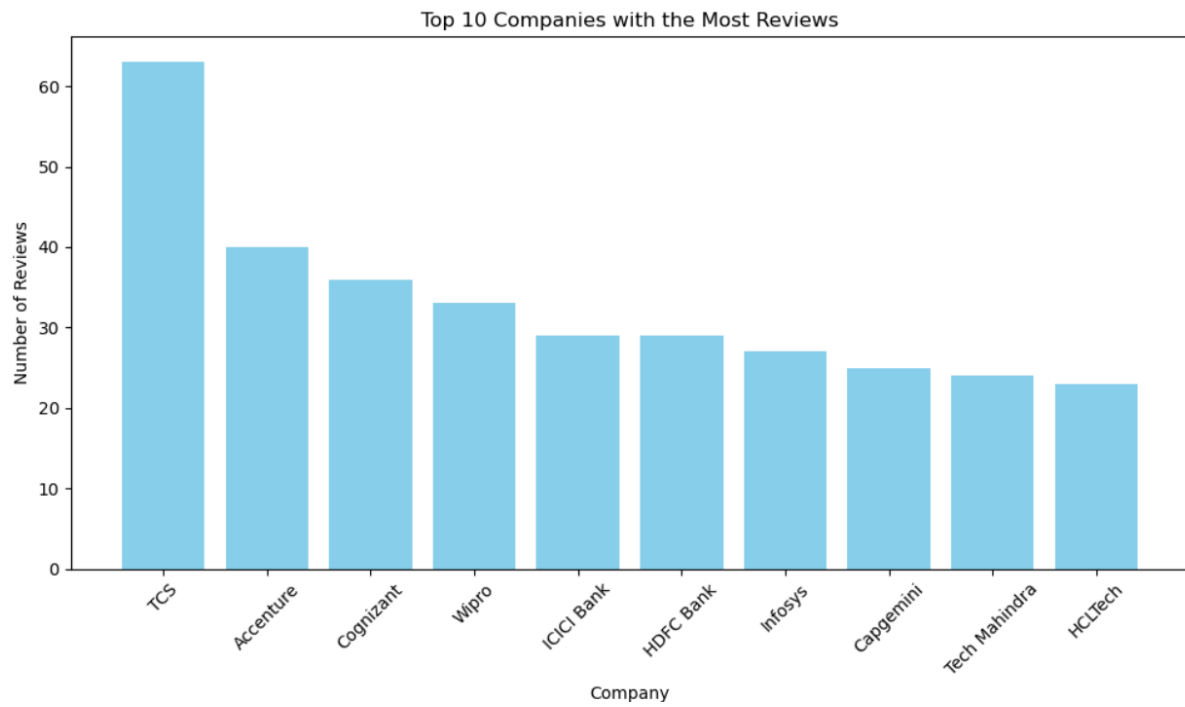
- **Oldest Company:** 78 Years old
- **Youngest Company:**

Cognizant	29 years old
ICICI Bank	29 years old
HDFC Bank	29 years old



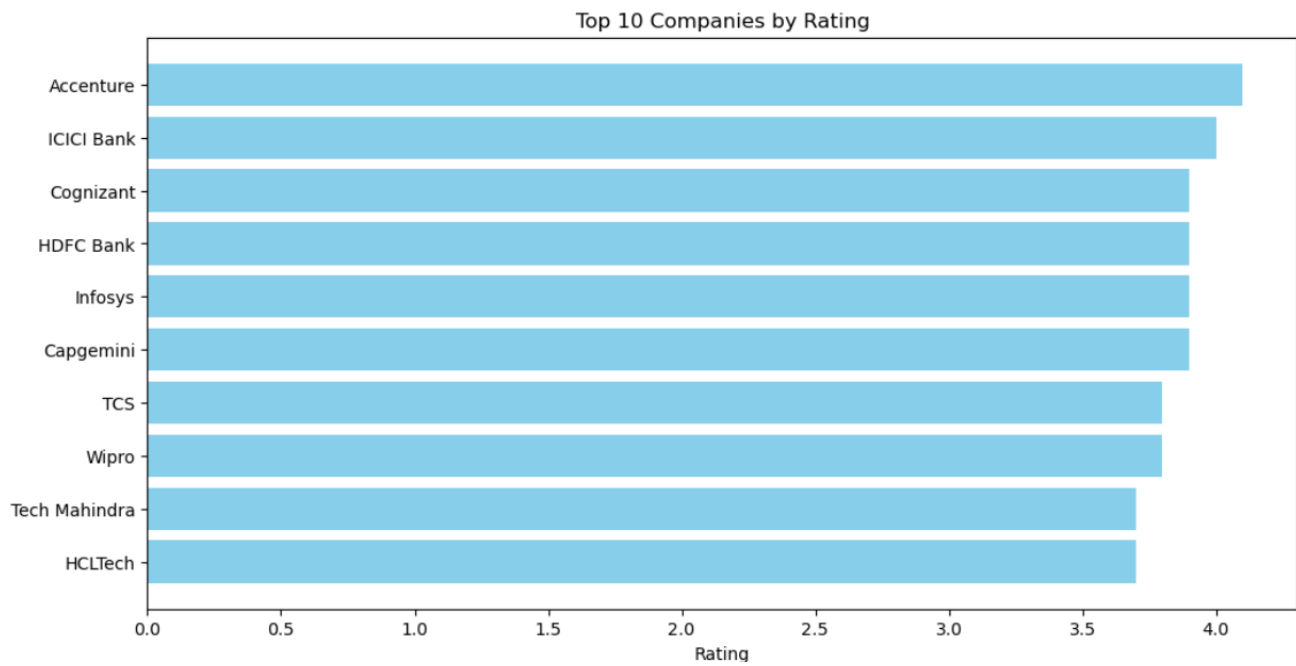
Company Type Distribution (in the form of pie chart pictorial analysis)

- 1.finance:** 13.8%
- 2.Manufacturing:** 15.4%
- 3.Consulting:** 24.6%
- 4.IT services:** 36.9%



Top 10 companies on the basis of reviews:

- 1.TCS
- 2.Accenture
- 3.Cognizant
- 4.Wipro
- 5.ICICI Bank
- 6.Infosys
- 7.Capgemini
- 8.Tech Mahindra
- 9.HCL Tech



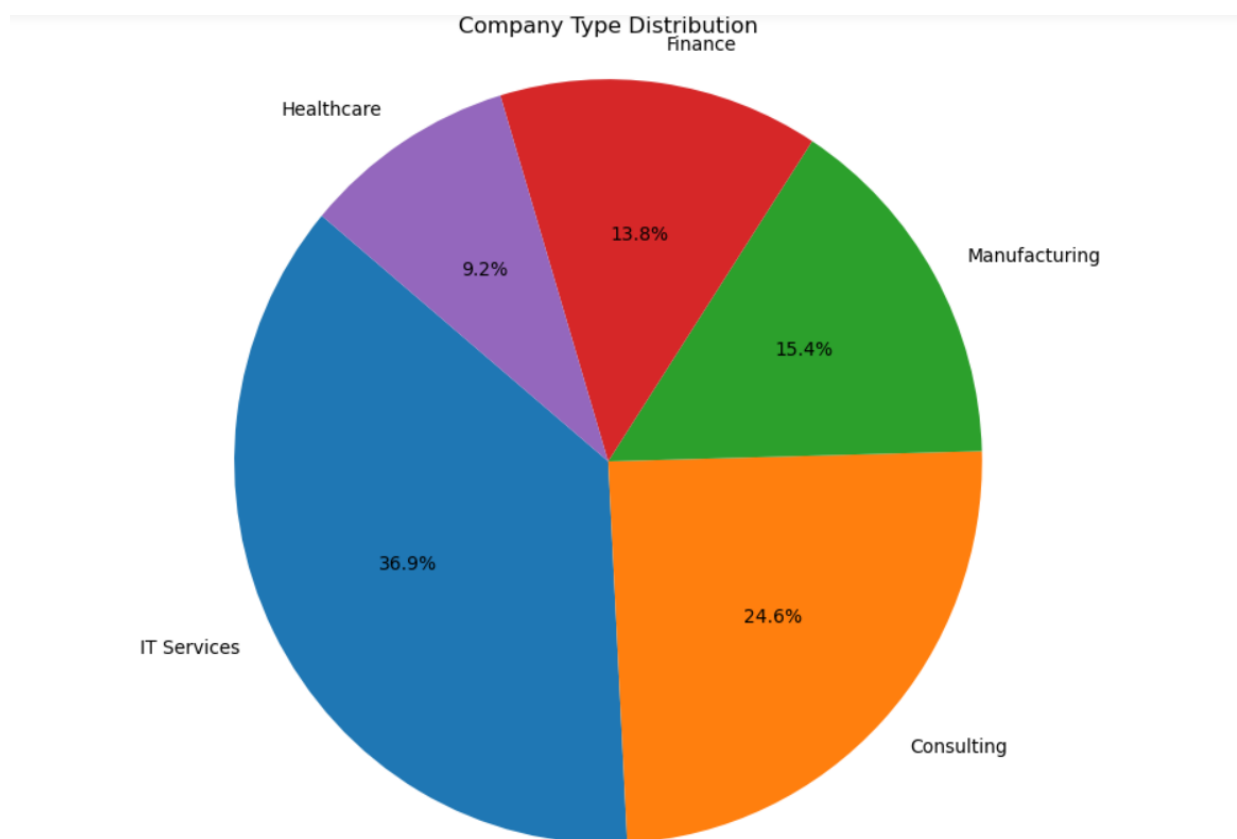
Top 10 companies on the basis of ratings:

- 1.Accenture
- 2.ICICI Bank
- 3.Cognizant
- 4.HDFC Bank
- 5.Infosys
- 6.Capgemini
- 7.TCS
- 8.Wipro
- 9.Tech Mahindra
- 10.HCL Tech

Findings & Inferences

Through Pie Chart Analysis

- **IT and Technology:** The largest segment, accounting for approximately 36.9% of the companies.
- **Consulting:** The second-largest segment, making up around 24.6% of the companies.
- **Manufacturing:** Represented approximately 15.4% of the companies.
- **Finance:** Represented approximately 13.8%
- **Healthcare:** Represented approximately 9.2%



Furthermore key findings

1. **Ratings: (Through Bar graph Analysis above)**

- The average company rating in this dataset is around 3.9, with ratings ranging from 3.2 to 4.7. This suggests that most companies have a moderately positive rating.

2. **Reviews: (Through Bar graph Analysis above)**

- The number of reviews varies widely between companies, with some having as few as 2.2k reviews and others having over 63.1k reviews.
- Higher-rated companies tend to have more reviews, indicating that well-regarded companies attract more feedback.

3. **Company Types:**

- The dataset includes companies from various sectors, including IT Services & Consulting, Banking, Pharma, Automobile, and more.
- IT Services & Consulting and Banking are the dominant sectors, with several companies falling into these categories.

4. **Company Size:**

- A majority of the companies in the dataset have "1 Lakh+ Employees," indicating that they are large organizations.
- Some companies fall into the "50k-1 Lakh Employees" category, while others are in the "10k-50k Employees" or "1k-5k Employees" range.
- A few companies have no employee information available.

5. **Company Age:**

- The age of the companies in the dataset varies significantly, with some being more than a century old and others being relatively young, like "12 years old" or "1 year old."

6. **Locations:**

- Companies are located in different cities and countries, reflecting a global presence.
- Mumbai, Bangalore, and Gurgaon/Gurugram are among the most common locations.

7. **Industry-Specific Observations:**

- IT Services & Consulting companies, such as TCS, Accenture, and Cognizant, dominate the dataset.
- Banking companies, including ICICI Bank, HDFC Bank, and Axis Bank, are well-represented.
- Several companies have no specific information available for employees, age, and location.

8. Highly Rated Companies:

- Amazon and Indian Army stand out as highly-rated organizations with ratings of 4.2 and 4.7, respectively.

9. Diverse Sectors:

- The dataset covers a wide range of sectors, including technology, finance, healthcare, and more, showcasing the diversity of the corporate landscape.

10. Geographic Presence:

- Companies in the dataset are located in various cities globally, indicating their widespread operations.

Managerial Insights | Implications

1. Industry Trends:

The majority of highly-rated companies in this dataset belong to the IT Services & Consulting, Banking, and Pharma industries. Managers in these sectors should consider studying the best practices of these top-rated companies to improve their own operations.

2. Company Size Matters:

Larger companies with over 1 lakh employees tend to dominate the top-rated list. This suggests that scale might contribute to higher employee satisfaction and better reviews. Smaller companies may need to focus on creating a more engaging and positive work environment.

3. Location Impact:

Companies based in major metropolitan areas such as Mumbai, Bangalore, and Gurgaon tend to have a significant presence in the top-rated list. Managers should consider how the location of their offices impacts employee satisfaction and company reputation.

4.Age of Company:

Older companies (with more than 50 years of existence) like TCS and HDFC Bank also feature prominently in the top-rated list. Managers can look at the long-term strategies and stability of such companies as potential sources of inspiration.

5.Company Type Matters:

IT Services & Consulting and Banking companies are leading in the ratings. Managers in other industries should assess what practices from these sectors can be applied to their businesses to improve employee satisfaction.

6. Customer-Centric Approach:

Amazon, a customer-centric internet company, has a high rating. Managers should consider the role of a customer-focused approach in improving employee morale and, ultimately, business success.

7. BPO Industry:

Concentrix Corporation and iEnergizer, both in the BPO industry, have high ratings. This suggests that BPO companies can create a positive work environment, and other BPOs can learn from their practices.

8. EdTech Sector:

BYJU'S, an EdTech company, has a comparatively lower rating. Managers in this sector should focus on improving employee satisfaction, which may lead to better product quality and customer experiences.

9. Global Presence:

Companies with a strong global presence, such as Accenture and IBM, have high ratings. Managers should consider expanding their businesses globally if feasible and beneficial.

10. Employee Engagement:

Companies with high ratings may have effective employee engagement strategies. Managers should invest in initiatives that foster engagement, such as training, career development, and work-life balance programs.

11. Continuous Improvement:

Companies that have consistently maintained high ratings over several years, like TCS and ICICI Bank, likely have effective strategies for sustaining employee satisfaction. Managers should focus on continuous improvement to prevent complacency.

12. Local Presence:

Some companies, such as Reliance Industries and Vodafone Idea, have a strong local presence but did not receive ratings. Managers should assess the reasons behind this and work to improve their image and employee satisfaction at the local level.

13. Employee Feedback:

Encouraging and acting upon employee feedback is crucial. Companies that have a lower rating should actively seek feedback and make necessary improvements to enhance employee satisfaction.

14. Competitive Benchmarking:

Managers can use this data for benchmarking against competitors to identify strengths and weaknesses. This can help in making informed decisions for improving employee satisfaction and overall company performance.

15. Diversity and Inclusion:

Managers should consider the role of diversity and inclusion in their organizations. Companies that prioritize diversity and inclusion tend to have more positive reviews.

16. Brand Reputation:

The reputation of a company, built over the years, plays a significant role in employee satisfaction. Managers should work on enhancing the reputation of their organization both internally and externally.

17. Strategic Partnerships: Companies in this dataset may benefit from strategic partnerships or collaborations with organizations known for their positive work culture and employee satisfaction.