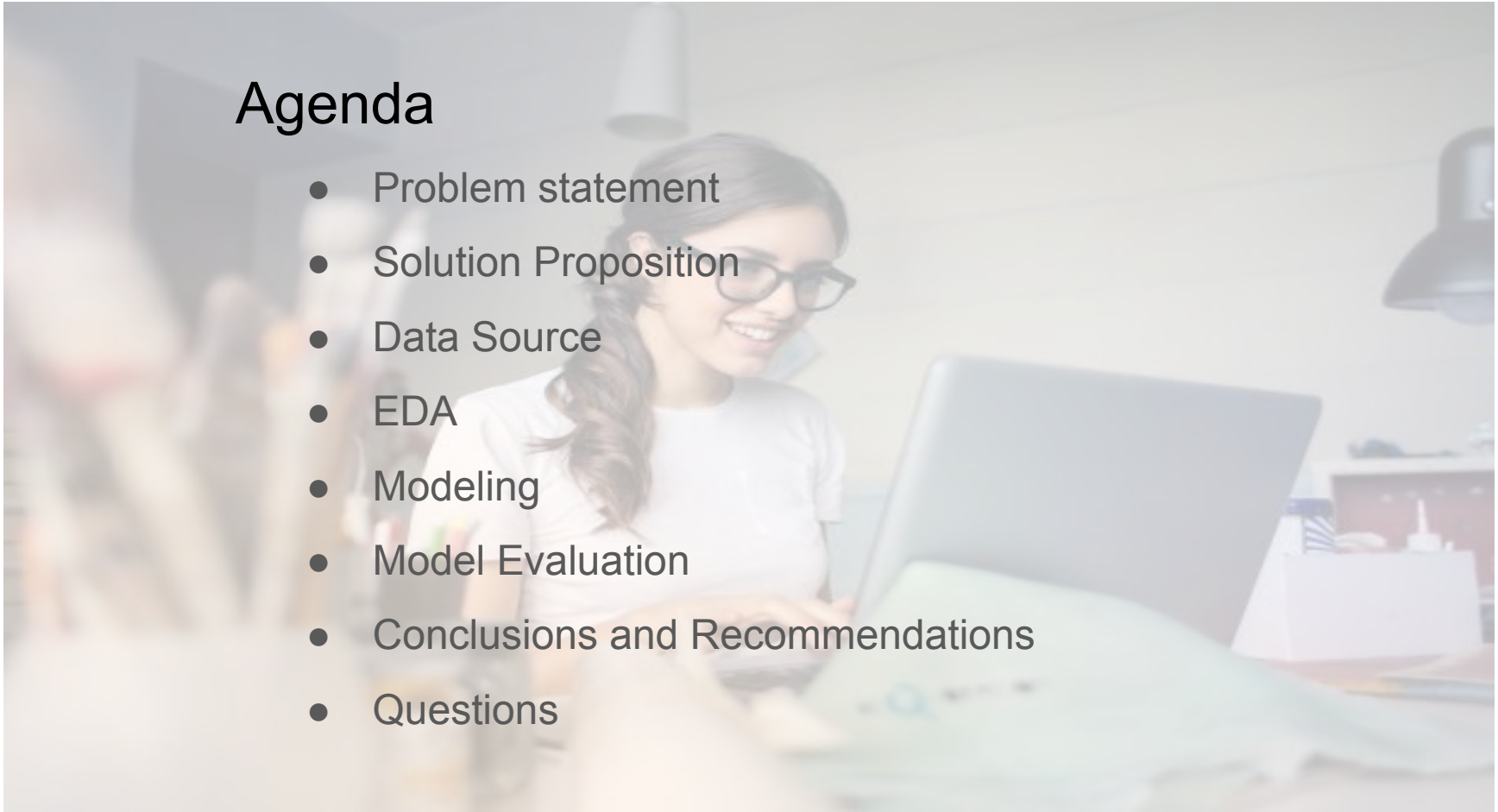
The Best Workplace

By Yunus Herman



Agenda

- Problem statement
- Solution Proposition
- Data Source
- EDA
- Modeling
- Model Evaluation
- Conclusions and Recommendations
- Questions

Problem Statement:

Can we use basic company data (such as size, revenue, and foundation year) and benefits offered to predict employee satisfaction?

Solution Proposition

Using machine learning and basic data on companies and the benefits they provide, I determined whether a company is a good workplace.

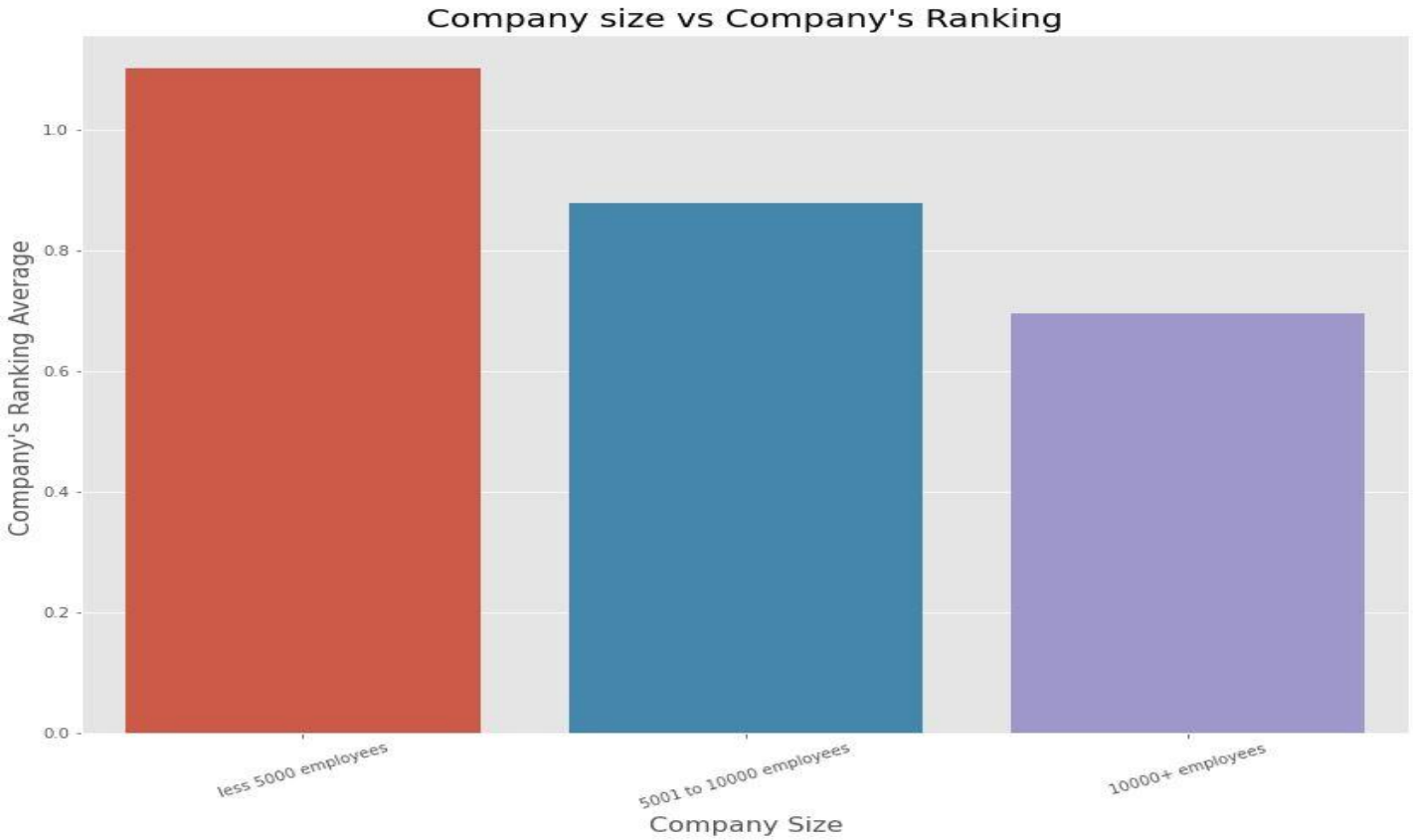
Data

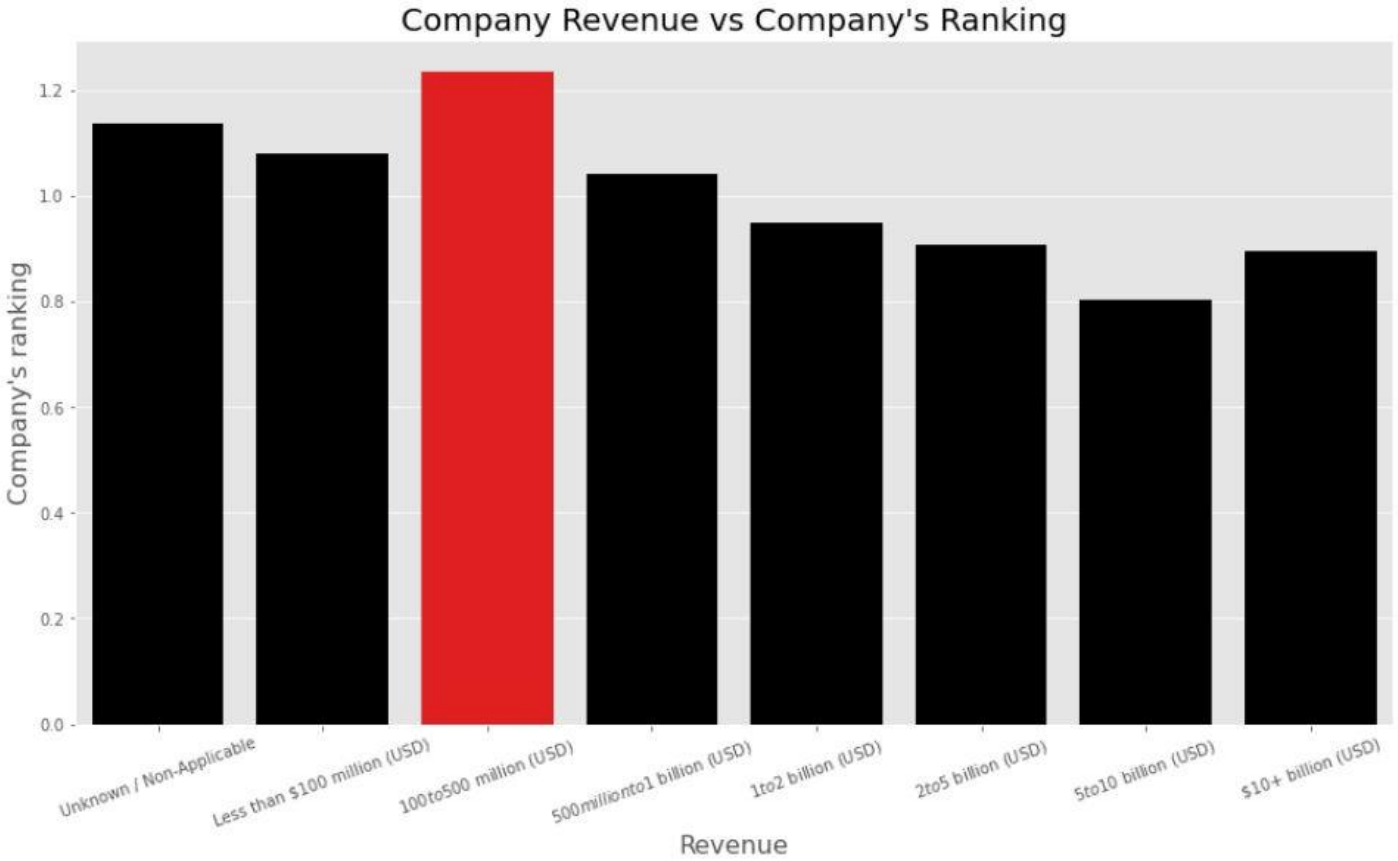
- Web Scraping: **Glassdoor.com**

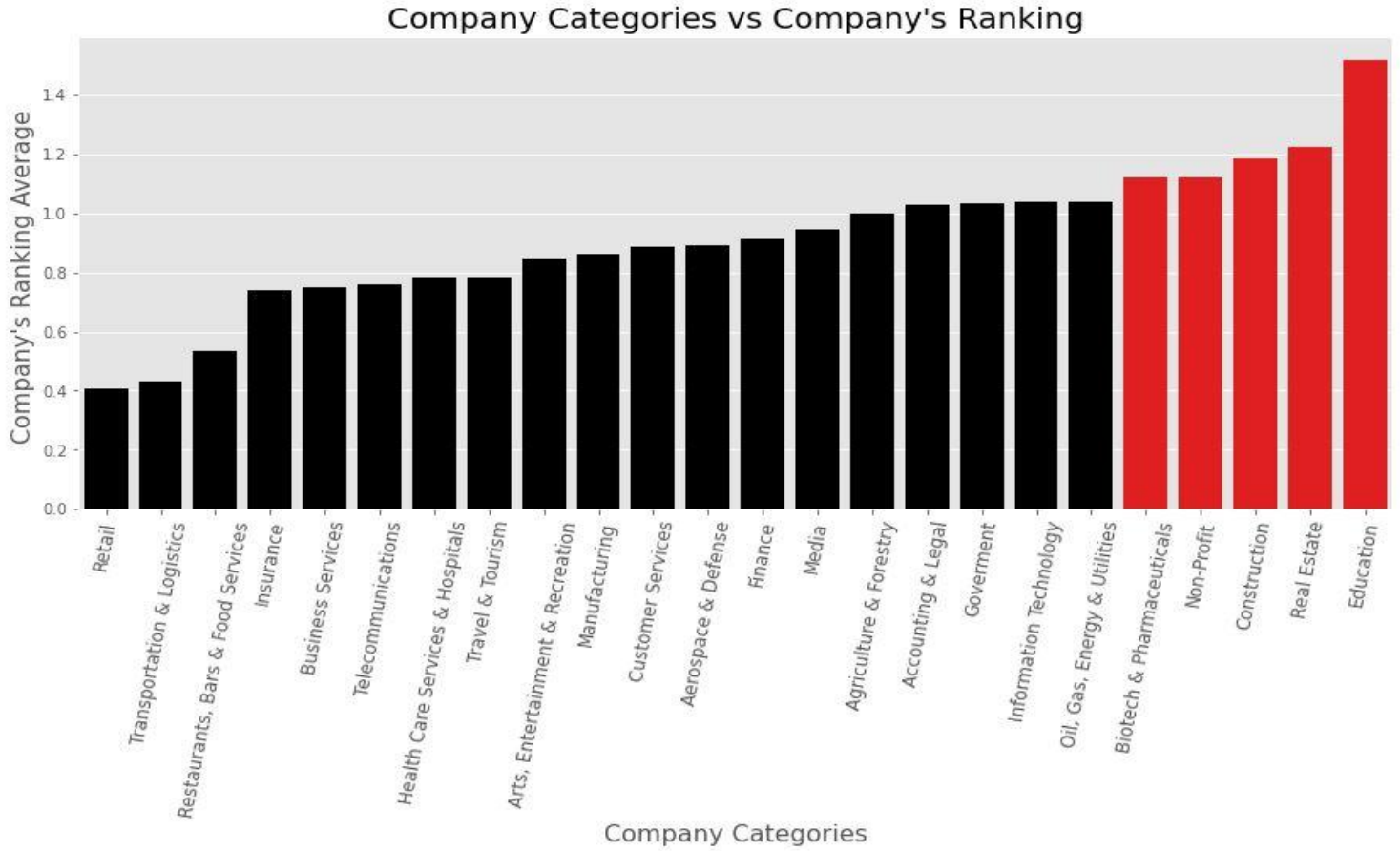
- Result: **2651 Companies** and **80 Features**

  Date of collection: **March 10, 2021**

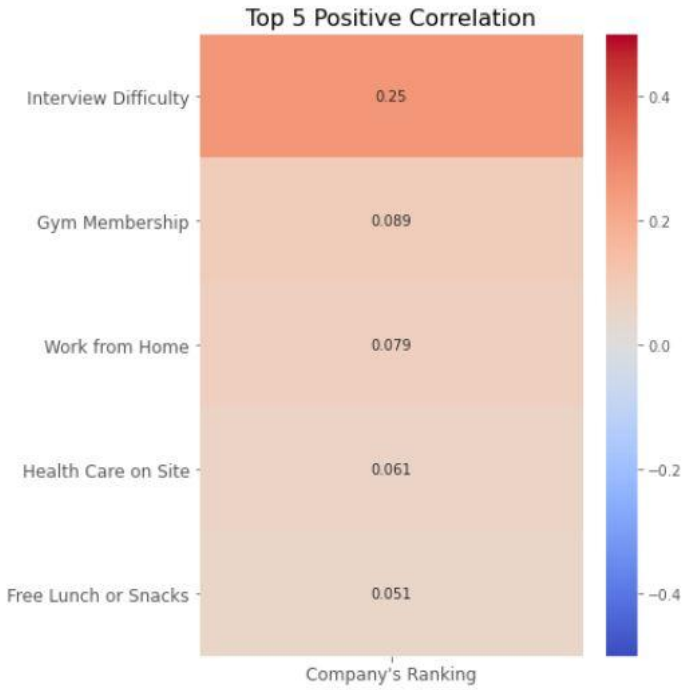
- Convert company ratings 1-5 stars (including fractions) into just 3 categories: above average, average, below average companies

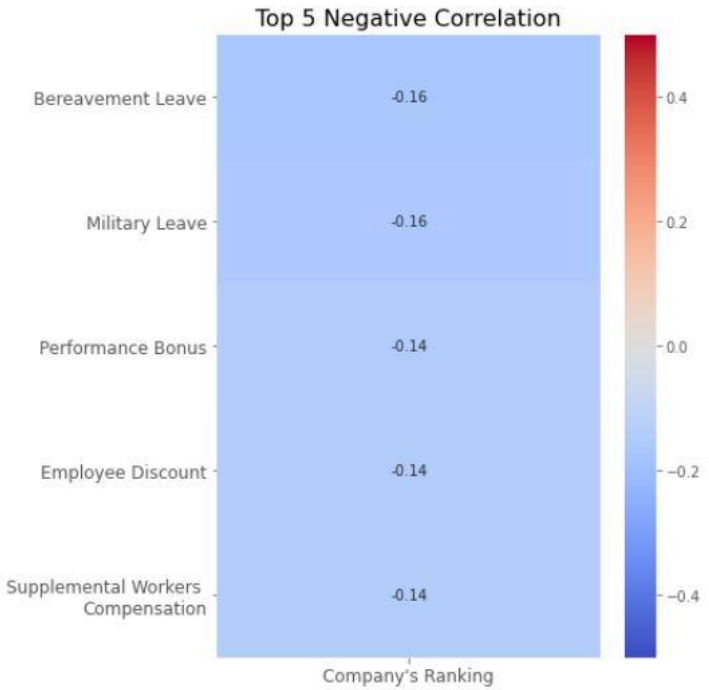

Company size vs Company's Ranking
Company's Ranking Average
1.0
0.8
0.6
0.4
0.2
0.0
less 5000 employees
5001 to 10000 employees
10000+ employees
Company Size


Company Revenue vs Company's Ranking
Company's ranking
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Unknown / Non-Applicable
Less than $100 million (USD)
100to500 million (USD)
500milliontо1 billion (USD)
1to2 billion (USD)
2to5 billion (USD)
5to10 billion (USD)
$10+ billion (USD)
Revenue


Company Categories vs Company's Ranking
Company's Ranking Average
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Retail
Transportation & Logistics
Restaurants, Bars & Food Services
Insurance
Business Services
Telecommunications
Health Care Services & Hospitals
Travel & Tourism
Arts, Entertainment & Recreation
Manufacturing
Customer Services
Aerospace & Defense
Finance
Media
Agriculture & Forestry
Accounting & Legal
Goverment
Information Technology
Oil, Gas, Energy & Utilities
Biotech & Pharmaceuticals
Non-Profit
Construction
Real Estate
Education
Company Categories

Correlation between features and company's ranking





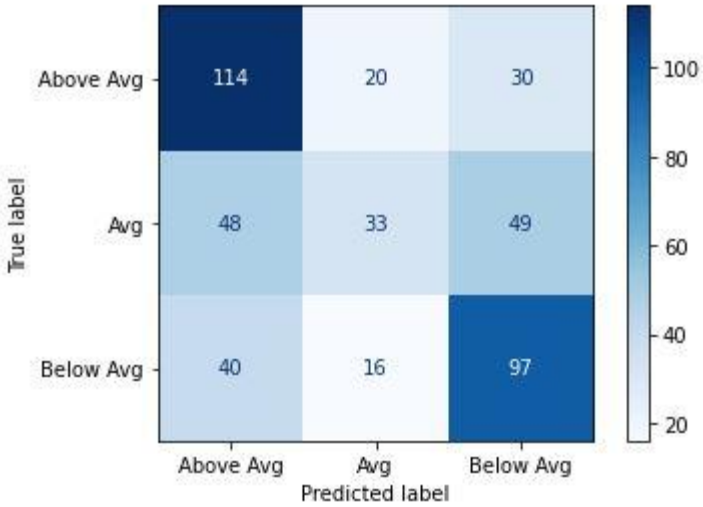
Modeling

Classification:

- SVC
- Random Forest
- ADA Boost
- Logistic Regression
- Decision Tree
- Extra Tree
- K-NN
- **Gradient Boost**

Model Evaluation

- Gradient Boost Classifier is the best model
- This model has an accuracy rate of 55%
- This may seem low but is nevertheless statistically significant, because the baseline is 36%
- The model is not so good at predicting a satisfaction levels of the ‘average’ class but is fairly accurate for ‘the above average’ and ‘below average classes’



Conclusions and recommendations

- Normally the benefits considered most important for employees are healthcare and pension/401k
- Interestingly, the results of my analysis indicate that other features have a much higher predictive value when it comes to employee satisfaction
- To improve the model, results would be more robust if companies are divided into just two classes: ‘above average’ and ‘below average’.
- Find the way to obtain data about other importance features such as salary range
- We should bear in mind that employee satisfaction is a highly subjective notion


Questions?