

Text Mining, Tech Giant Reviews using Natural Language Processing (NLP)

By Saaket Varma Y2019

Natural Language processing or Texting Mining	<p>Introduction: More than 70% of potentially usable data exists in unstructured form. This data is stored in text format. This is a form of unstructured data and companies are looking for ways to analyze this data to acquire business intelligence.</p> <p>Text mining or Natural language processing provides us with techniques to gain actionable insights from these data.</p>
Problem Statement	<p>Human Resources Analytics: The HR department of Amazon wants to develop strategies for hiring appropriate talent and prospective employees who are more suited to their work culture. They want to understand what distinguishes Amazon employees with Google employees.</p> <p>The Data Scientists from Springboard have been asked to explore online reviews and provide insights.</p>
Datasets	<p>Amazon and Google Reviews. (500+ and 500-ve reviews) Several employees from both companies submit online reviews. These are categorized into positive and negative reviews. Four data sets are available for analysis. Resource: Datacamp (Ted Kwartler)</p>
Questions Addressed	<ul style="list-style-type: none"> • What are employees saying about Amazon and Google? • Which company has better work-life balance? • Which is company is perceived better in compensating according to the online reviews? • Can HR use the analyses in strategizing their talent hunt?
Methods	<p>Steps involved in Analyzing the Text Data (NLTK library)</p> <ol style="list-style-type: none"> 1) Organization & Text processing (raw text -> sentence segmentation -> tokenization -> corpus -> matrix -> word count) 2) Feature extraction and Analysis (TF, DTM, TF-ID, TF-IDF, DM -> Top terms & Word associations) 3) Machine Learning (clustering, classification, association rules, predictive modeling) 4) Data visualization: Dendrograms, Word cloud. Pyramid plots.
Significance	<p>NLP can provide key business intelligence insights, which are not just limited to human resource management but can also help other departments including marketing to improve sales, accounts and finances, purchasing, customer service support, distribution, research, and development.</p>

A case study in HR analytics

1. Problem definition

Which company has better work life balance? Which has better perceived pay according to online reviews?

2. Unorganized state



3. Organization

4. Feature Extraction

5. Analysis

6. Organized state

Insight, recommendation, analytical output



What are employees saying about Amazon?

Top 15 bigrams (two words)

Amazon Pro		Frequency	Amazon con		Frequency
→	good pay	25	long hours		29
→	great benefits	24	worklife balance		21
	smart people	20	work life		21
→	place work	17	life balance		20
	good benefits	16	not enough		9
	fast paced	16	peak season		8
	great place	12	management not		8
	learn lot	12	not good		8
	work environment	11	high turnover		7
	great pay	11	work environment		6
	good work	10	hard work		6
	pay good	10	hours long		6
	pay great	10	many people		6
	people work	10	work hours		6
	pay benefits	9	not really		6

What are employees saying about Google?

Top 15 bigrams

(two words)

Google Pro

Frequency

smart people	42
free food	41
place work	26
great benefits	22
great perks	20
great work	18
people great	16
work environment	16
great people	16
great place	15
great culture	14
worklife balance	12
people work	12
work life	12
work great	11

Google con

Frequency

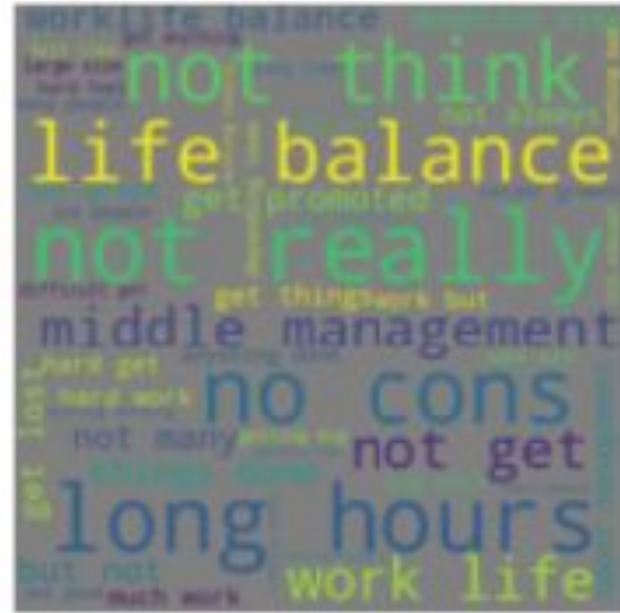
not really	11
no cons	10
long hours	10
not think	9
not get	8
middle management	8
work life	8
life balance	8
worklife balance	8
get promoted	7
not many	7
but not	7
things done	7
get things	6
get lost	6

Word Clouds of Amazon Reviews



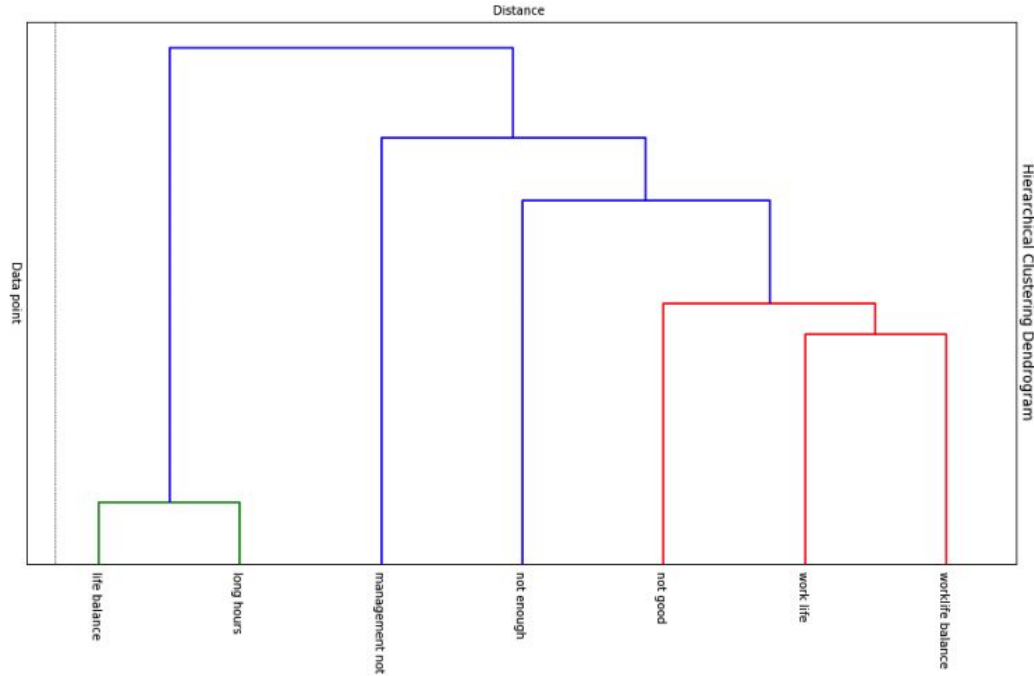
“[Work life balance](#)” bigram in Amazon con reviews appears 21 times.

Word Clouds of Google Reviews



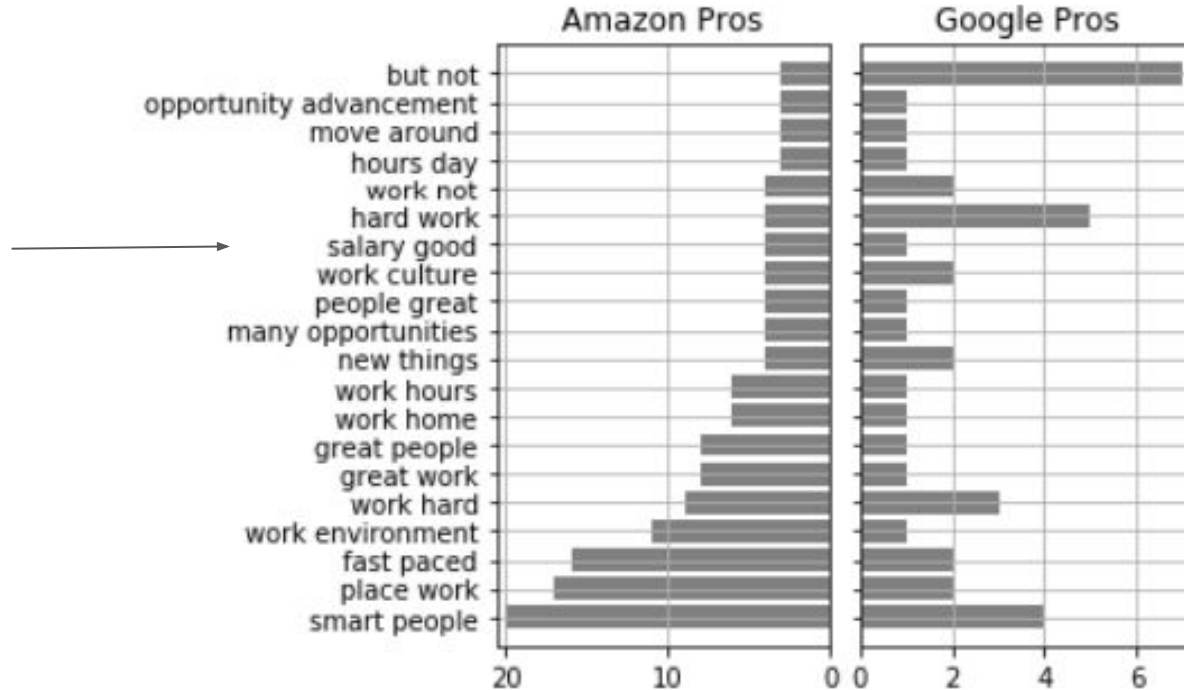
“[Work life balance](#)” bigram appears in Google in pros and cons, 12 and 18 times respectively.

Work Life Balance: Word associations through clustering



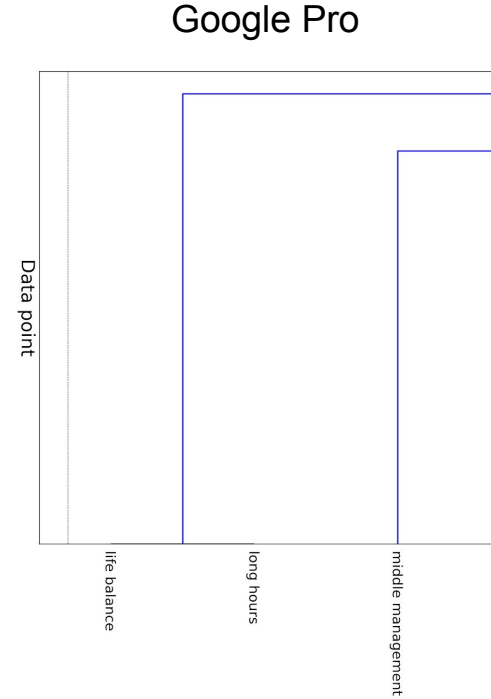
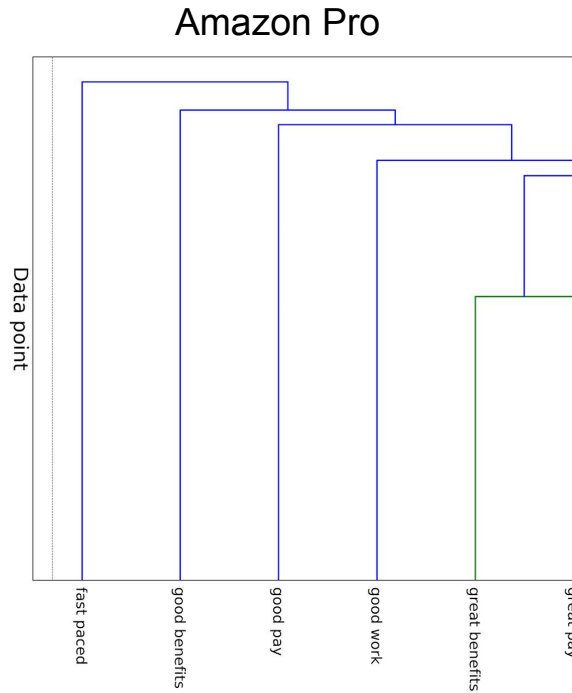
“**Work life balance**” bigram is associated with working long hours and probably apathy shown by the management could be a problem.

Which company is perceived better in compensating according to the online reviews?



More employees in amazon are talking positive about the “good pay” and “benefits” as compared to google.

Can HR use the analyses in strategizing their talent hunt?



Employees like “fast paced” work at Amazon, whereas at Google employees like the work-life balance and love their work.

Sentiment Analysis:

- Through Machine learning approach we also effectively predicted if the review is a pro or a con. Shown in table below are top and lowest bigrams predicted as a pro review.

Pros	P(pros word)
	pro googl 0.94
multipl	opportun 0.94
	take easi 0.94
	everyth veri 0.93
	but long 0.92
	amazon look 0.92
cultur	transpar 0.92
cowork	upward 0.92
	well equal 0.92
	good give 0.92
Cons	P(pros word)
	night work 0.16
	day hr 0.15
	big someth 0.15
balanc	softwar 0.15
	area live 0.15
process	driven 0.15
	due larg 0.14
burnout	real 0.14
	hour manag 0.12
great	worker 0.09

- Application:** We can use this NLP pipeline for any new analyses:
 - Online product reviews
 - Market analysis for startups
 - Analyzing clinical reports

Time to reach a conclusion!

Problem definition

Which company has better work life balance? Which has better perceived pay according to online reviews?

Unorganized state



Organization

Feature Extraction

Analysis

Organized state

Insight, recommendation, analytical output

Conclusions:

I was able to process unstructured data and provide the following key insights from 2000 reviews on Glassdoor

- **Work-life balance rating:** Google has better work life balance ratings than Amazon. Mid-management level needs an evaluation.
- **Compensation:** Amazon seems to provide better compensation than Google.
- **Recommendation for HR:**
 - We see that Amazon employees like fast paced environment. This could be added to their job posting as an essential part of their work culture and set priority for applicants who like the same.
 - Amazon employees like the good pay. This can be used by HR to attract more talent.
 - Google employees prefer work life balance. This can be used to attract more talented employees.
- This NLP pipeline can be used to gain business Intelligence in areas of sales, marketing & commercialization.