

GlassDoor Job Reviews

Python Project Group 2

Andrew Leong
Benson Chen
Lotus Liu
Jay Zhu
Naomi Wong



TABLE OF *CONTENTS*



01

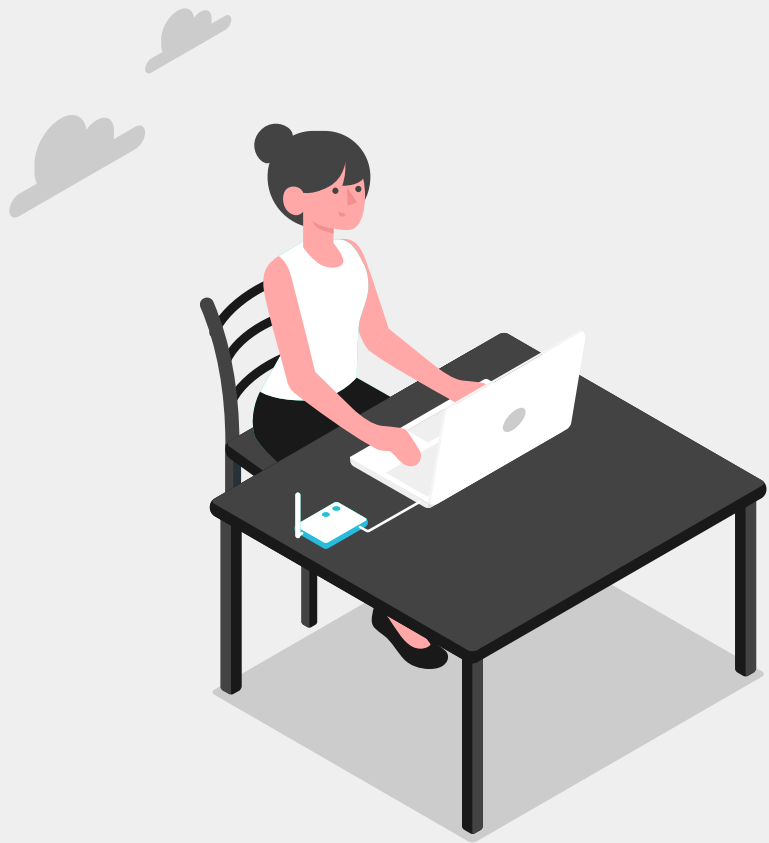
Descriptive
Analytics

02

Interesting
Findings

03

Other
Findings



01

Descriptive *Analytics*



Data *Description*

Our dataset is from Kaggle, which is called 'Glassdoor Job Reviews' dataset.

This dataset contains employee reviews left on the job search platform Glassdoor from 428 companies around the world.

The dataset has 838566 rows and 20 columns. Each row corresponds to an employee review.



Data *Description*

Columns:

1. **column label**: column identifiers (N/A - not relevant to dataset)
2. **firm**: firm that is reviewed
3. **date_review**: date of the review
4. **job title**: job title of the reviewer
5. **current**: current status of the employee (former or current)
6. **location**: job location (Hong Kong, New York, England, London)
7. **overall rating**: overall rating of the firm (scale of 1 to 5)
8. **work_life_balance**: worklife balance rating (scale of 1 to 5)
9. **culture_values**: culture and values rating (scale of 1 to 5)
10. **diversity_inclusion**: diversity and inclusion rating (scale 1 to 5)
11. **career_opp**: career opportunities rating (scale 1 to 5)
12. **comp_benefits**: compensation and benefits rating (scale 1 to 5)
13. **senior_mgmt**: senior management rating (scale 1 to 5)
14. **recommend**: reviewer recommendation (v - Positive, x - Negative, o - No opinion)
15. **ceo_approv**: reviewer approval of the CEO (v - Positive, r - Mild, x - Negative, o - No opinion)
16. **outlook**: reviewer outlook of the firm (v - Positive, r - Mild, x - Negative, o - No opinion)
17. **headline**: headline of the review (string)
18. **pros**: pros (string)
19. **cons**: cons (string)
20. **use**: N/A - all null values

Data *Transformation*



Convert string values to integer

x = -1 (negative), v = 1 (positive), r & o = 0 (neutral and no opinion)



Convert string values to lowercase

DON'T DO IT! -> don't do it!



Create dummy variables

Work Status: Current (0), Former (1)

Work Experience: less than 1 year & more than 5 years

Data *Cleaning*



Remove columns with 50% NaNs and irrelevant columns (column_label, use, diversity_inclusion)



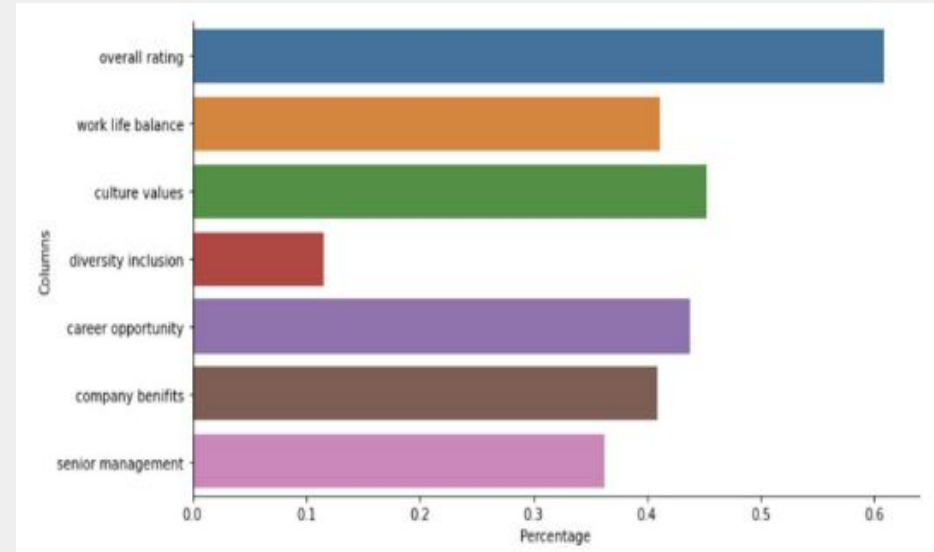
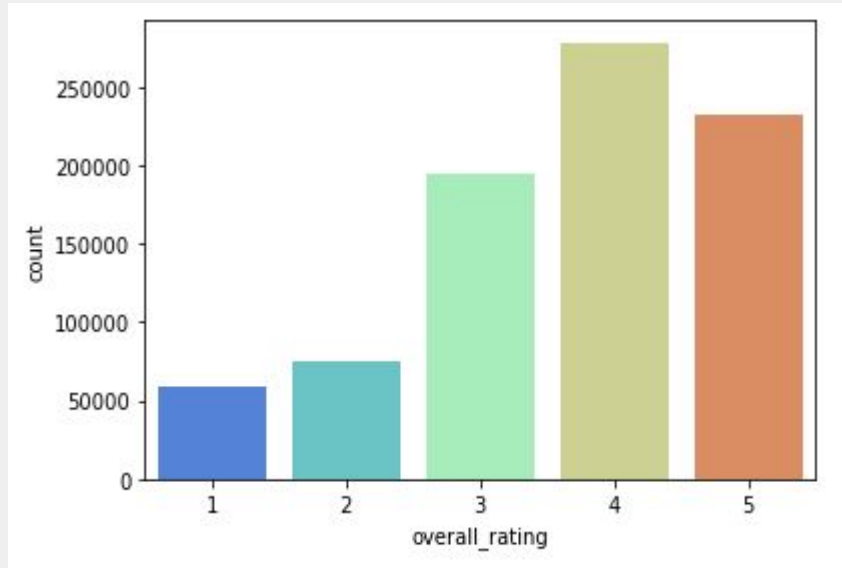
Replace NaNs with ‘ ‘ for string type columns (cons, headline, location)



Drop rows with NaNs for integer type columns



Data *Visualization*





02

Interesting *Findings*



“What affects the overall rating among Big Four accounting firms?”

Question 7

Data *Description*



102,837 reviews among 4 accounting firms

After data cleaning and creating the subset.



The average overall rating is 3.73

Deloitte has the highest rating (3.78) and KPMG has the lowest (3.64)

Decision *Tree*



Create “high-rating” column

1 : overall rating ≥ 4
0 : overall rating < 4



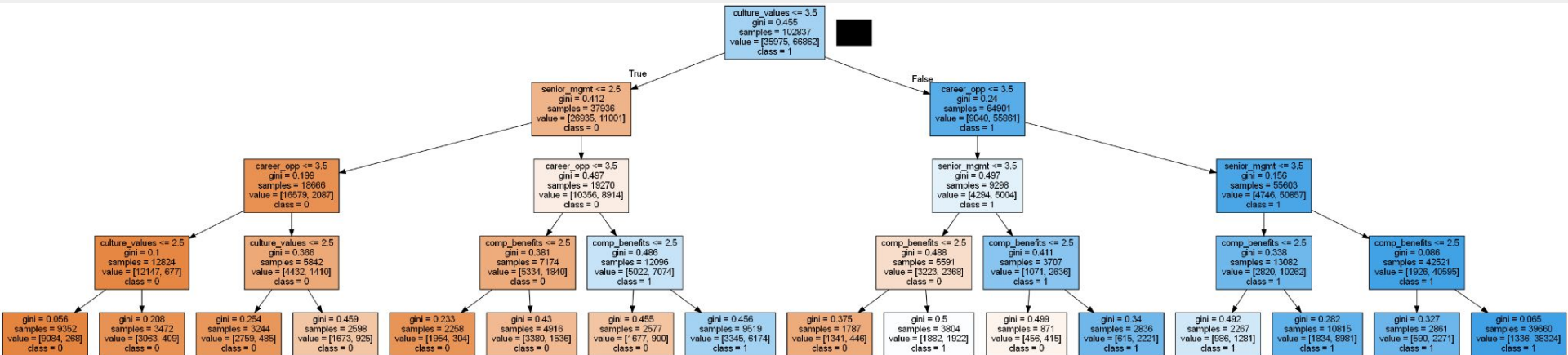
Drop unnecessary columns and create dummy variables



Run decision tree that classifies high-rating firms and low-rating firms



Decision Tree





Tree *Interpretation*

1. When the company has "high culture values" and "high career opportunities", no matter the company benefits are high or low, it is likely that the company will have a high overall rating.
2. When the company has "high culture values", even if their career opportunities are low, the company will have a high overall rating once their "senior management" and "company benefits" are high.
3. Based on decision tree, the most important factor that affects rating is **culture values**.



Clustering

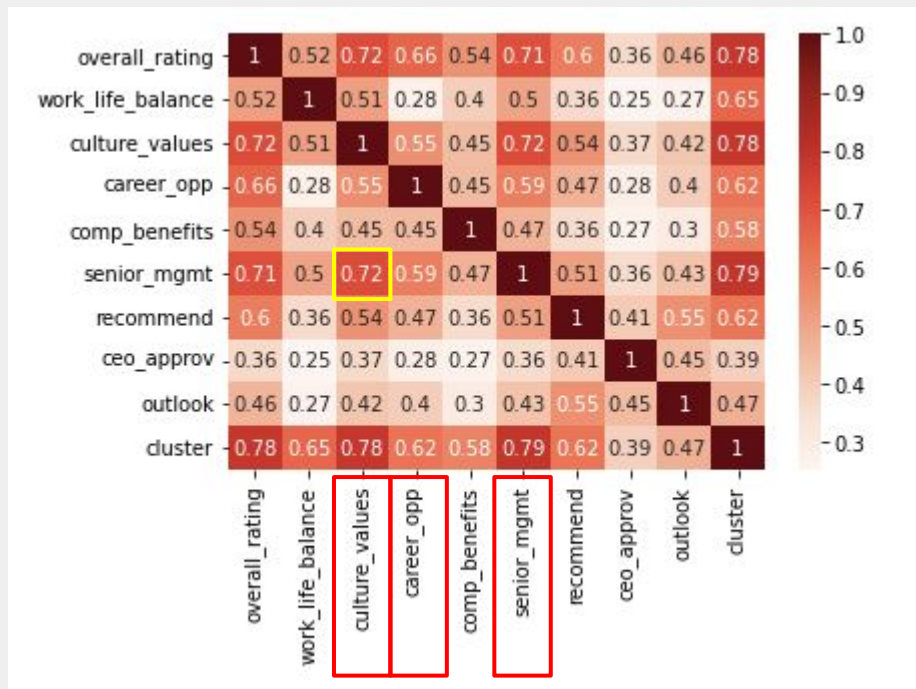
```
df_clus.groupby('cluster').mean()
```

	overall_rating	work_life_balance	culture_values	career_opp	comp_benefits	senior_mgmt
cluster						
0	2.189345	1.693097	1.960166	2.630647	2.292114	1.683651
1	3.655898	2.414008	3.618679	3.916342	3.102234	3.179986
2	4.537759	3.855949	4.635378	4.548547	4.034943	4.330255

1. The clustering basically classified the data into 3 parts. In cluster 1, the overall rating is close to the average rating (3.73) among 4 accounting firms.
2. Accounting firms that have a average overall rating often have a **better career opportunities and culture values**. However, the work-life balance is still low.



Heatmap



1. Culture values, career opportunities and senior management are highly positive correlated with overall rating.
2. Culture value is also highly correlated with senior management.

So it is possible that companies would have a high culture value if they also have good senior management.



Conclusion 7

What kind of companies have high-ratings?

1. Based on decision tree, clustering, and heatmap, we can give suggestions to job seekers that the accounting firms have a high overall rating often have a **high culture value**.
2. We will suggest the accounting firms that they could improve their overall rating by improving their culture values. For example, establishing a committee that reviews/implements the everyday culture and values of the firm.



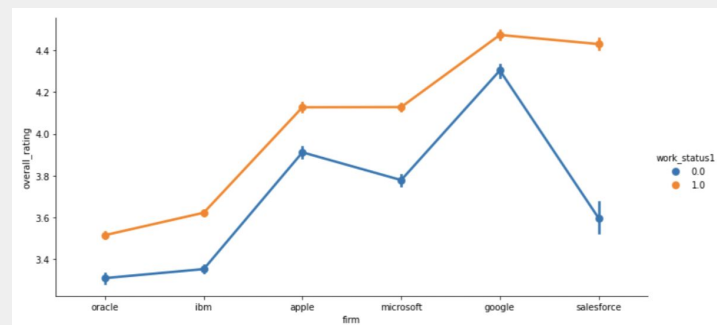
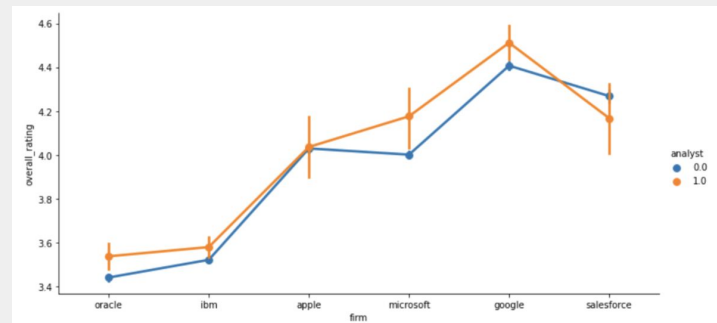
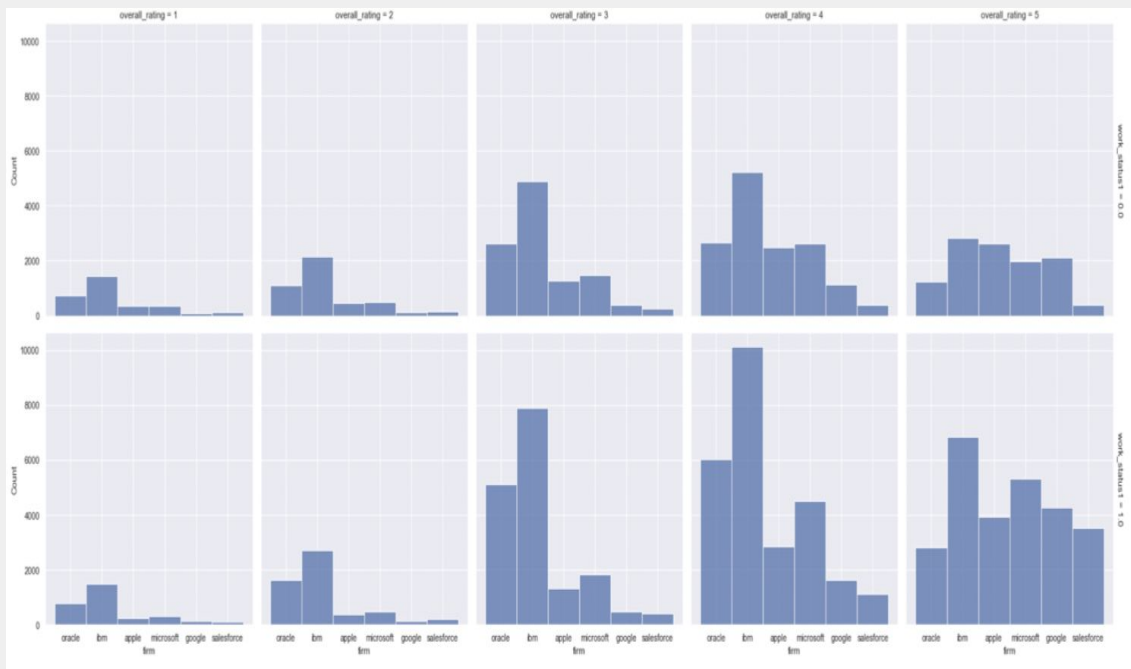
“What are the best and the worst big tech firms to work for?”

Question 2

Data Analysis

122,959 reviews among 5 big tech firms

	overall_rating	work_life_balance	culture_values	career_opp	comp_benefits	senior_mgmt	recommend	ceo_approv	outlook
firm									
oracle	3.538153	3.763855	3.526104	3.254618	2.990361	3.062651	0.346185	0.330120	0.334940
ibm	3.580225	3.721798	3.637753	3.447640	3.034607	3.087191	0.350112	0.247640	0.255281
apple	4.036697	3.486239	4.073394	3.674312	4.059633	3.596330	0.655963	0.651376	0.610092
salesforce	4.167702	4.000000	4.366460	3.981366	4.360248	3.832298	0.639752	0.720497	0.664596
microsoft	4.176471	3.887701	4.021390	3.914439	4.048128	3.775401	0.652406	0.593583	0.636364
google	4.511873	4.168865	4.432718	4.234828	4.358839	4.100264	0.770449	0.593668	0.696570



Google Pros and Cons

pros	cons
782515 the job is very interesting and educational	nothing complicated or problematic, that is all...
782577 great salary and employees. one of the best an...	no cons for such a great company.
782604 it is very free and comfortable. i love the ti...	competition is too intense and work stress is ...
782612 smart people, great perks, interesting projects	big company with lots of people
782631 great compensation for hr positions.	not great at prioritizing work/life balance
782674 work life balance is really good	stress about promotion is real
782726 as a large company with plenty of the resource...	about half of the people i've worked with here...
782737 good team and work culture	management is not good at hiring
782775 amazing culture, people & work environment	no cons whatsoever working here
782782 every thing is cool there.	there is no cons to work there,
782824 love it here. it's amazing	nothing. management is great and awesome
782826 love it here. it's amazing	nothing. management is great and awesome
782830 everyone i worked with was really on top of th...	sometimes it was tough to feel inspired when t...
782879 helpful benefits for growing family	limited career growth over time
782949 management is nice. there are very few bad app...	the job is a little boring. a lot of what made...
782997 its google lmao its perfect	none cons idk none cons
783141 love it love it love it	none at all at all
783155 great benefits, great pay, stock is doing amaz...	if you're stuck with a manager who isn't inves...
783156 great benefits, great pay, stock is doing amaz...	if you're stuck with a manager who isn't inves...
783181 benefits and perks are industry-leading, and w...	very large and ever-growing company creates a ...
783191 challenging work with lots of impact/vnbenefit...	size can sometimes be intimidating and make th...
783264 money, people, benefits, dogs in office	bureaucracy, processes, feeling like a cog in a...
783339 you'll meet some of the smartest people you've...	huge company now, standard enterprise challeng...
783364 perks, good people, decent pay	large company, so comes with politics. senior ...
783391 large scope and lots of learning opportunities...	a bit of a cult
783422 a ton of perks, eg gym, food, a lot of fun work	location of the office in new york is not the ...
783436 full time, high permission, flexible	little amount of work, no
783449 excellent meals. generally a great work enviro...	lots of talent there, but not always clear wha...
783450 knowledge & growth with company and very diverse	can be more organized at times
783458 the base salary was good and they are very acc...	extremely hard to commute with low work offices
783530 excellent compensation and benefits, good work...	very large company, hard to navigate and have ...
783648 great work life balance and absurd pay for the...	none that i can think of at the moment
783722 good work environment and appreciable coworkers	eh not many i can think of
783813 great work life balance and great place to work	none that i can think of

Oracle Pros and Cons

pros	cons
239952 great work environment company culture	not a lot of transparency
239954 great work environment company culture	not a lot of transparency
240002 learning, benefits, environment, pay, and enga...	not much, i was free to work the way i wanted
240135 great flexibility in work timings and autonomy...	salary hike and promotions are rare
240341 i have always enjoyed working at oracle.\nth...	the compensation could be little better but ag...
240345 i have always enjoyed working at oracle.\nth...	the compensation could be little better but ag...
240532 management is great, open territories for sale...	some verticals are harder to sell than others
240534 great compensation \namazing people\ngreat pto...	work life balance could improve \nsome acqui...
240554 great team culture and very corporate / profes...	no cons. it was fantastic and the pay was great
240578 somewhat competitive pay (the stock is also do...	hard to gain momentum (legacy products, legacy...
240600 ceo down focus on client in 1st place/vnpassi...	customer perception is outdated/v\nalumni work...
240632 promoting within, fantastic leadership, great ...	need to be able to react well to changes in th...
240660 benefits seemed good. access to training on or...	oracle tech is not the leading edge and you wi...
241010 lots of resources/v\nbrand name recognition v'r...	constant re-orgs/v\nlittle upward growth witho...
241103 money is good with territory	too many management levels there
241107 money is good with territory	too many management levels there
241509 great brand and health benefits	too much process to get anything answered quickly
241514 flexible work culture, of late it is becoming ...	no salary raise for many years/v\nno clarity o...
241598 a good place to work	very large organization to work
241622 international exposure, job opportunities, goo...	na. na, na, na, na
241626 international exposure, job opportunities, goo...	na. na, na, na, na
241993 great experience and team player	lack of growth in internal management
241997 great experience and team player	lack of growth in internal management
242026 smart people and good reputation	red tape and that's it
242106 good company with ethical values	demanding management kind of less tolerant
242361 lots of great people who are willing to help t...	constant restructuring and reorganization make...
242469 great technology and platform opportunity	employee views does not matter. % decisions d...
242732 unlimited vacation, great benefits, 100% remote	average pay, lots of corporate overhead
242763 all the things related to company	none as of now which i came across.
242767 all the things related to company	none as of now which i came across.
242888 work duties, learning, team, benefits	management does not care about employees
242894 work duties, learning, team, benefits	management does not care about employees

Oracle Pros and Cons

pros	cons
421996 good work environment we can enjoy our work	less salary and no yearly hikes
422098 very flexible in terms of work hour	not really a cons but does not have remote pos...
422182 - high base salary/v\n- exposure to nearly eve...	- constant job cuts and lay offs/v\n- managing...
422185 - high base salary/v\n- exposure to nearly eve...	- constant job cuts and lay offs/v\n- managing...
422204 cutting edge work and teams	compensation and transformation job security
422256 regular paychecks/v\ninteresting clients/v\nma...	salaries are non-negotiable even after a drast...
422450 stability/v\nconsistent stream of work/v\nflex...	upward mobility/v\nan stay at same level for ...
422511 good name - people get impressed when you say ...	- poor management/v\n- managers no skilled/v\n...
422519 good name - people get impressed when you say ...	- poor management/v\n- managers no skilled/v\n...
422568 great work learning and diverse culture	solid commitment to embracing the future, but ...
422642 flexibility, networking, skills-gain, good dev...	the company needs to invest in strengthening L...
422666 thoughtful management and great support. every...	it would be great to see the ibmer weekly news...
422688 -intelligent, focused, legacy tech company. V...	-marketing division group think/v\n-too many L...
422912 great work flexibility, opportunity for growth...	big company, slow to respond, and internal pol...
422917 great work flexibility, opportunity for growth...	big company, slow to respond, and internal pol...
423068 reasonable compensation and good benefits	ok pay, less than competitive bonus
423157 flexibility, cutting edge work, learning	productivity and complexity and result
423242 work / life balance; genuinely good people	incredibly inward looking; overly political; d...
423359 good money if you work hard	long hours, lots of politics
423370 pay, benefits, boss was great	the company was so large you often got lost in it
423673 global footprint/v\niversity/v\nmany smart an...	products are outdated or subpar/v\nbad behavio...
423681 global footprint/v\niversity/v\nmany smart an...	products are outdated or subpar/v\nbad behavio...
423714 good brand name, well known with clients	nobody is on the same page
423718 good work environment, interesting projects	stressful company and sometime unrealistic obj...
423721 \$ and kindness and people	too nice they pay too well
423725 great networking opportunities, experience to ...	unsure of role before starting
423740 if you luck out with a good management chain t...	number driven, executives are not supportive, ...
423748 if you luck out with a good management chain t...	number driven, executives are not supportive, ...
423753 develop admin and programming skills	to much work for a only position
423832 training/v\ngood projects/v\nopportunities to ...	large company, lots of processes
423901 though i don't like the new name much. as long...	i wish they kept the same name though.
423993 great, smart people that ibm is lucky to have...	seems to most frequently lay off people with a...
423998 great, smart people that ibm is lucky to have...	seems to most frequently lay off people with a...

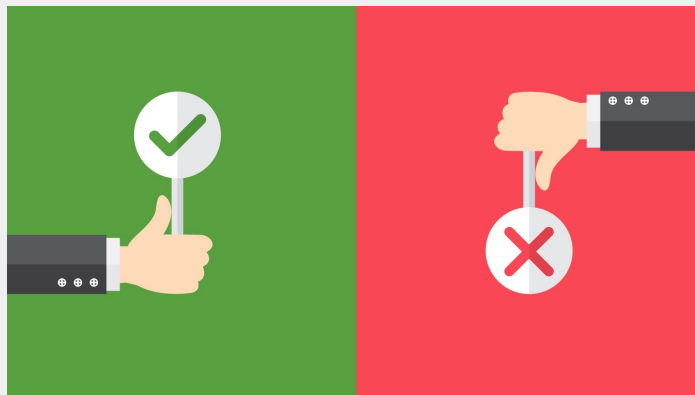


Conclusion 2

Where should we work for?

1. The best place to work at is Google.
The worst place to work at are Oracle and IBM.
2. Oracle and IBM can improve their management system and they should focus on work-life balance. For example, they can work on team-building exercises, more networking events with coworkers, clients, and managerial staff.
3. As large of a company Oracle is, we suggest to break Oracle into more smaller departments to improve efficiencies.





**“What are the most common words
used in pros and cons?”**

Question 3



Data *Analysis*

```
df.pros.head()
```

```
1    people in leadership roles have been there for...
2    great exposure, respectful environment and bea...
3                good team, support and training
4    exposure to artworks, great team of specialists
5    fantastic art, nice enthusiastic people and a ...
Name: pros, dtype: object
```

```
df.cons.head()
```

```
1    pressure to do whatever the leaders want\r\nno...
2                ot during peak seasons and often underpaid
3                high-stress with relatively low compensation
4                work life balance, company culture
5                very low pay and huge workload
Name: cons, dtype: object
```



Data *Analysis*

```
from collections import Counter
counts = Counter(cleaned_pro_words)
counts.most_common(20)
```

```
[('people', 121211),
 ('benefits', 87471),
 ('company', 87420),
 ('environment', 66394),
 ('opportunities', 63194),
 ('culture', 60758),
 ('life', 59937),
 ('balance', 54095),
 ('pay', 48775),
 ('flexible', 45861),
 ('team', 45786),
 ('nice', 41824),
 ('friendly', 38074),
 ('career', 35641),
 ('management', 35001),
 ('training', 34098),
 ('hours', 32385),
 ('job', 32103),
 ('learn', 31543),
 ('time', 31309)]
```

```
from collections import Counter
counts = Counter(cleaned_con_words)
counts.most_common(20)
```

```
[('management', 101792),
 ('hours', 76906),
 ('company', 68939),
 ('people', 62386),
 ('pay', 58788),
 ('time', 53927),
 ('life', 48538),
 ('balance', 41026),
 ('managers', 39352),
 ('employees', 38580),
 ('salary', 38407),
 ('staff', 38084),
 ('low', 36142),
 ('job', 34862),
 ('poor', 33099),
 ('sometimes', 31068),
 ('hard', 30957),
 ('team', 29810),
 ('lack', 29621),
 ('don', 29601)]
```




Data *Analysis*

PROS



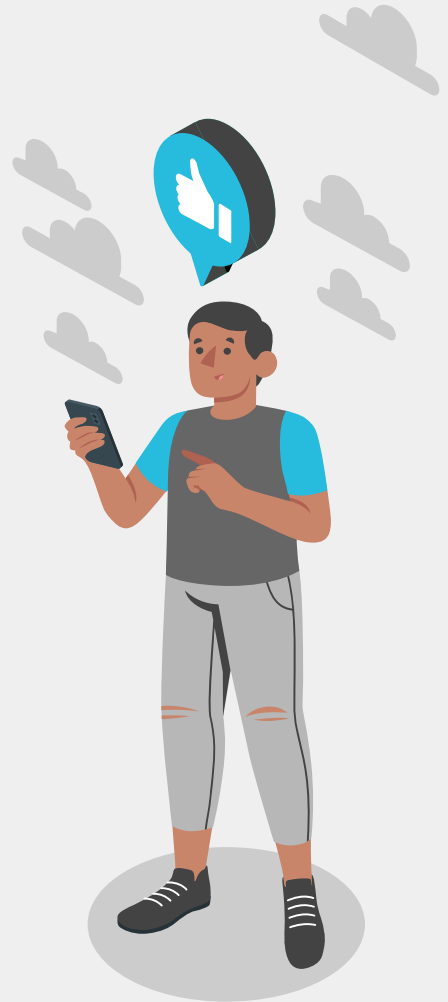
CONS



Conclusion 3

What are the most common words?

1. The most common **positive** phrases are company, good team/people to work with, friendly working environment, opportunity, work life balance, and benefits.
2. The most common **negative** phrases are related to hour, time, company, people, manager and life balance.
3. Firms should put emphasis on recruiting, HR, resource planning as well as providing more learning / career opportunities for their employees.





03

Other *Findings*

```
#find the top 10 overall_rating companies
df.groupby('firm')['overall_rating'].mean().nlargest(10)
```

```
firm
The-Survey-Association    5.000000
Time-Etc                  4.785714
Northern-Gas-and-Power    4.719682
CarShop-UK                4.651163
Michelmores               4.610390
Bain-and-Company          4.559757
Gateley                   4.540984
Kingsley-Healthcare       4.500000
Sport-England             4.500000
Taylor-Wimpey             4.497396
Name: overall_rating, dtype: float64
```


```
In [20]: #find the number of interviewed employees from the top 10 overall_rating companies
len(df[df.firm == 'The-Survey-Association'])
len(df[df.firm == 'Time-Etc'])
len(df[df.firm == 'Northern-Gas-and-Power'])
len(df[df.firm == 'CarShop-UK'])
len(df[df.firm == 'Michelmores'])
len(df[df.firm == 'Bain-and-Company'])
len(df[df.firm == 'Gateley'])
len(df[df.firm == 'Kingsley-Healthcare'])
len(df[df.firm == 'Sport-England'])
len(df[df.firm == 'Taylor-Wimpey'])
```

```
Out[20]: 3
Out[20]: 42
Out[20]: 503
Out[20]: 258
Out[20]: 77
Out[20]: 3623
Out[20]: 61
Out[20]: 122
Out[20]: 24
Out[20]: 384
```

Other Findings

The highest overall_rating company is
The-Survey_Association!

Other Findings



```
In [40]: # find the gap from current employee with former employee
df_difference = df_current_mean - df_former_mean
df_difference
```

```
Out[40]: overall_rating    0.412651
work_life_balance    0.345389
culture_values    0.455617
career_opp    0.399405
comp_benefits    0.232490
senior_mgmt    0.436534
recommend    0.256261
ceo_approv    0.186392
outlook    0.221395
dtype: float64
```

Current employees have better job satisfaction on every aspects than former employees.

Other Findings

```
In [43]: #find the average overall_rating from different working year groups of current employee
df_current_employee[df_current_employee.work_more_than_10years == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.work8_to_10years == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.work5_to_8years == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.work3_to_5years == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.work1_to_3years == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.work0_to_1year == 1.0].overall_rating.mean()
df_current_employee[df_current_employee.current_worker == 1.0].overall_rating.mean()
```

Out[43]: 3.772060396081954

Out[43]: 3.808349808117223

Out[43]: 3.7856380138814734

Out[43]: 3.7578114870997017

Out[43]: 3.7594451470123302

Out[43]: 3.938727042431919

Out[43]: 3.809155302899126

Current employee:

those who works **less than 1 year** has the highest overall_rating,
those who works **3 to 5 years** tends have the lowest overall_rating on average.

Other Findings

```
In [46]: #find the average overall_rating from different working year groups of former employee
df_former_employee[df_former_employee.work_more_than_10years == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.work8_to_10years == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.work5_to_8years == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.work3_to_5years == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.work1_to_3years == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.work0_to_1year == 1.0].overall_rating.mean()
df_former_employee[df_former_employee.former_worker == 1.0].overall_rating.mean()
```

```
Out[46]: 3.3806734992679357
```

```
Out[46]: 3.45955474967261
```

```
Out[46]: 3.4553407647805843
```

```
Out[46]: 3.4342398238723075
```

```
Out[46]: 3.342506867433106
```

```
Out[46]: 3.2252705763906366
```

```
Out[46]: 3.396393783016275
```

Former Employee:

Former employee who works **8 to 10 years** has the highest overall_rating score of the company. Those who works **more than 3 years**(3-5,5-8,8-10,more than 10) have very similar average overall_rating.

The lowest overall-rating comes from those who works **less than 1 year** in the company.



THANKS!

DO YOU HAVE
ANY QUESTIONS?