





아마존 기업 리뷰 분석 프로젝트

데이터 소스 탐색

indeed 검색 기업리뷰



Come build the future with us



Amazon.com
3.5 ★★★★★ 73,566건의 리뷰

주요 정보 회사 소개 73,566 리뷰 급여정보

1.0
☆☆☆☆

Amazon warehouses are meatgrinders
Tier_1 (전 직원) - Washington State - 2021년 8월 3일

Worked 5 years at Amazon FCs over 3 states. Did I see any p33 bottles? No, actually. Did managers come talk to you if they saw you were performing 5% too slow? Yeah they would be totally in their rights as a manager for doing that. Terrible work life balance. 10 1/2 hour shifts, 6 days a week during the holidays.

이 리뷰가 도움이 되셨습니까?

3.0
★★★★☆

Decent Work
Sortation Associate (전 직원) - Hebron, KY - 2021년 8월 3일

I worked here at a new station, so we had state of the art equipment. Pay was pretty decent, usually had 20-24 hours per week as I was PT. But it was a dead end job, little room for advancement. It worked for the time.

✓ **장점**
Amazon Discount, Decent pay

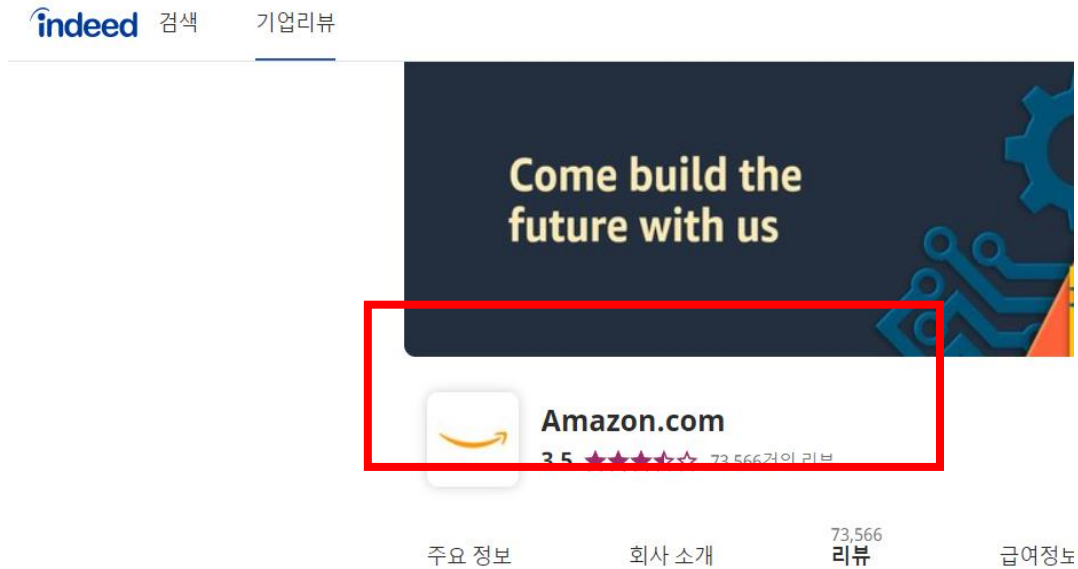
✗ **단점**
Short breaks, little oversight.

이 리뷰가 도움이 되셨습니까?

- 웹사이트: indeed.com

- 아마존에서 일했던 근로자들의 리뷰들이 존재

분석 목적



- 아마존은 세계적으로 성공한 기업임에 불구하고 근로자들의 평가가 5점 만점에 3.5점에 불과한 원인 탐색
- 원인 파악과 동시에 평가를 높이기 위한 방안 제시

사용한 데이터

3.0
★★★★☆

Decent Work

Sortation Associate (전 직원) - Hebron, KY - 2021년 8월 3일

I worked here at a new station, so we had state of the art equipment. Pay was pretty decent, usually had 20-24 hours per week as I was PT. But it was a dead end job, little room for advancement. It worked for the time.

✓ 장점

Amazon Discount, Decent pay

✗ 단점

Short breaks, little oversight.

이 리뷰가 도움이 되셨습니까?

예

아니요

신고하기 공유하기

- 리뷰들을 구분할 리뷰 제목

- 리뷰

- 단점

데이터 전처리 방법

Tfidf Bag of words - Reviews

```
] : def my_tokenizer(r):  
    r1 = word_tokenize(r)  
    r2 = [w.lower() for w in r1 if w.isalpha()]  
    r3 = [w for w in r2 if not w in all_stopwords]  
    r4 = [w for w in r3 if not w.isdigit()]  
    r5 = [lem.lemmatize(w, get_pos(w)) for w in r4]  
    r6 = [w for w in r5 if not w in all_stopwords]  
    return r6
```

all_stopwords

```
{'a',  
 'about',  
 'above',  
 'accomodation',  
 'accrues',  
 'actua',  
 'actually',  
 'advancementtuition',
```

- 특별한 의미가 없는 단어, 오타자, 회사가 위치한 지명 등 분석에 불필요한 단어들은 불용어로 처리

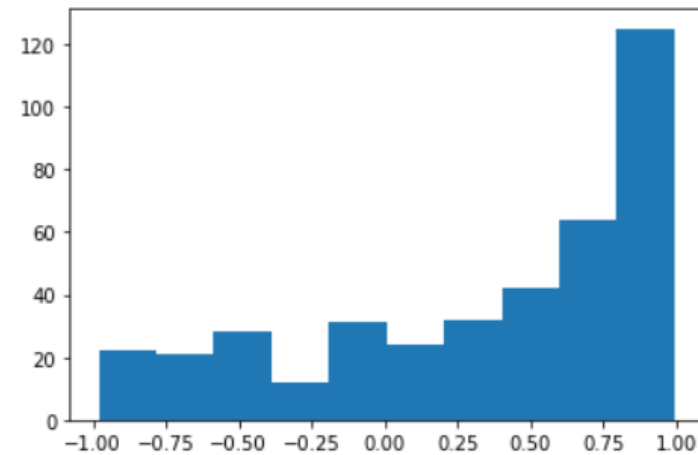
Sentiment Analysis

```
In [48]: employee_review.describe()
```

Out[48]:

Sentiment	
count	401.000000
mean	0.342596
std	0.582210
min	-0.981800
25%	-0.067600
50%	0.542300
75%	0.839600
max	0.994300

```
In [49]: plt.hist(sentiment)  
plt.show()
```



- 리뷰 감정 분석 결과 중위값인 점수가 0.54이고 400개의 리뷰 중 약 120개는 굉장히 긍정적인 리뷰인 것을 알 수 있음.
- 아마존의 평점이 3.5인 이유는 대부분의 사람들은 만족하고 있음에도 불구하고 불만을 느끼는 소수의 인원이 별점을 매우 낮게 줘서 평균 별점이 낮아진 것으로 판단.

해결 방안 모색: Cluster Analysis (리뷰)

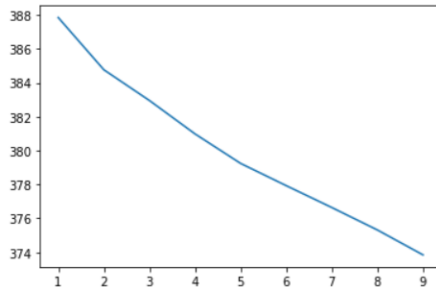
```
In [29]: from sklearn.cluster import KMeans
```

```
In [30]: distance1 = []
```

```
K = range(1,10)
for x in K:
    km_n = KMeans(n_clusters=x, random_state=0)
    km_fit = km_n.fit(tfidf_doc1)
    num = km_fit.inertia_
    distance1.append(num)
```

```
In [31]: import matplotlib.pyplot as plt
```

```
In [32]: plt.plot(K, distance1)
plt.show()
```



```
In [35]: oc1 = km1.cluster_centers_.argsort()[:, :-1]

for i in range(0,3):
    top_words = [tfidf_terms[x] for x in oc1[i, :5]]
    print(top_words)

['time', 'hour', 'break', 'long', 'care']
['people', 'manager', 'give', 'like', 'come']
['good', 'pay', 'great', 'place', 'hour']
```

- 리뷰를 대상으로 클러스터 분석을 진행하기 위해 엘보 기법을 사용했을 때 급격히 감소하는 부분이 없어 클러스터 개수를 3개로 잡고 분석을 진행함.
- 첫 번째 클러스터는 long, time, hour, break 등이 대표적인 단어인 것으로 보아 노동 시간에 관한 클러스터인 것으로 추정. 두 번째는 people, manager로 보아 인간 관계나 HR 관련 클러스터로 추정. 세 번째는 good, great, pay 등을 보아 만족감과 관련된 클러스터로 추정.

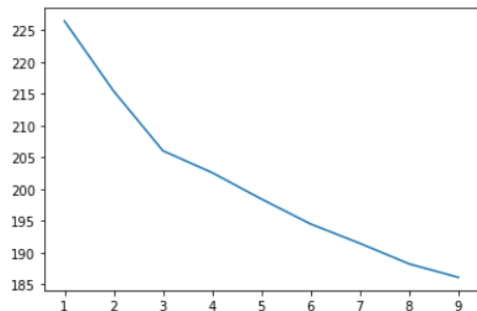
Cluster Analysis (단점)

Cons

In [36]: distance2 = []

```
K = range(1,10)
for x in K:
    km_n = KMeans(n_clusters=x, random_state=0)
    km_fit = km_n.fit(tfidf_doc2)
    num = km_fit.inertia_
    distance2.append(num)
```

In [37]: plt.plot(K, distance2)
plt.show()



In [117]: oc2 = km2.cluster_centers_.argsort()[:, :-1]

```
for i in range(0,3):
    top_words = [tfidf_terms2[x] for x in oc2[i, :6]]
    print(top_words)
```

```
['bad', 'break', 'lack', 'balance', 'boring', 'life']
['pay', 'bad', 'low', 'workload', 'much', 'short']
['break', 'short', 'hour', 'long', 'cleanest', 'boring']
```

- 이번엔 단점을 대상으로 클러스터 분석을 진행. 이번에는 엘보우 기법을 활용하여 클러스터 3개가 적합할 것이라고 판단.
- 클러스터들이 break, life balance, workload, hour 등을 중심으로 형성된 것을 보아 과도한 노동 시간이 불만의 원인인 것을 파악함.

LDA 토픽 모델링

```
In [197]: lda_model2.print_topics(num_words=3)
```

```
Out[197]: [(0, '0.030*"break" + 0.021*"bad" + 0.018*"time"'),  
           (1, '0.050*"hour" + 0.039*"long" + 0.021*"boring"'),  
           (2, '0.068*"break" + 0.068*"short" + 0.019*"hour"')]
```

- LDA 토픽 모델링으로 분석을 진행해도 앞선 클러스터 분석과 비슷한 결과 도출 가능. 즉, 클러스터 분석에서 놓친 숨겨진 토픽이 없다는 걸 알 수 있음.

결론

→ 휴식 시간을 늘리고, 노동 시간을 과도하지 않게 조절하면 불만이 줄어들고 기업 평점도 높아질 것으로 보임.

프로젝트의 한계

1.0 Amazon warehouses are meatgrinders

☆☆☆☆☆

Tier 1 (전 직원) - Washington State - 2021년 8월 3일

Worked 5 years at Amazon FCs over 3 states. Did I see any p33 bottles? No, actually. Did managers come talk to you if they saw you were performing 5% too slow? Yeah they would be totally in their rights as a manager for doing that. Terrible work life balance. 10 1/2 hour shifts, 6 days a week during the holidays.

이 리뷰가 도움이 되셨습니까?

예

아니요

신고하기 공유하기

3.0 Decent Work

☆☆☆☆☆

Sortation Associate (전 직원) - Hebron, KY - 2021년 8월 3일

I worked here at a new station, so we had state of the art equipment. Pay was pretty decent, usually had 20-24 hours per week as I was PT. But it was a dead end job, little room for advancement. It worked for the time.

✓ 장점

Amazon Discount, Decent pay

✗ 단점

Short breaks, little oversight.

이 리뷰가 도움이 되셨습니까?

예

아니요

신고하기 공유하기

- 이번에는 부서, 지역 상관 없이 분석 당시 기준 최근 400개의 아마존 리뷰를 긁어 와서 분석을 진행했지만, 업무별 혹은 지역별로 리뷰를 나누어 분석을 진행할 경우 결과가 다르게 나타날 가능성이 있음.