

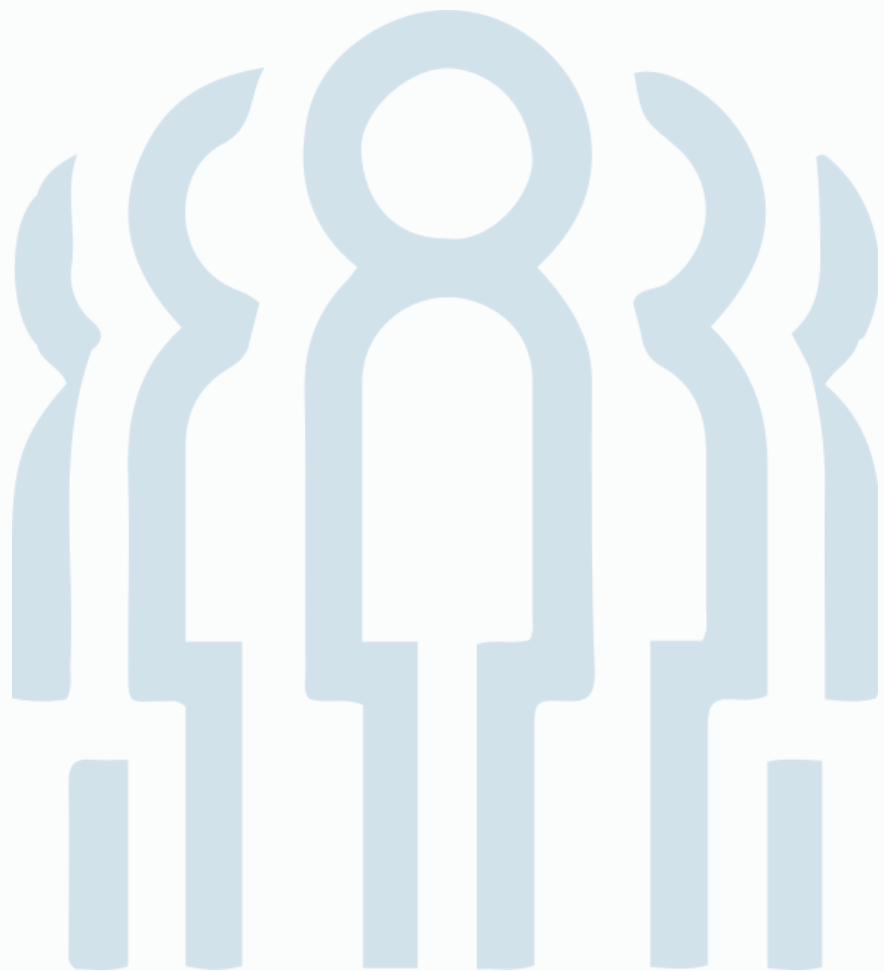


유니콘을 찾아조~

3조

김동섭 김태민 안태윤 이해강

CONTENTS



Chapter **01** 데이터 및 분석 주제 소개

Chapter **02** 분석 방향성 설정

Chapter **03** 분석 내용

Chapter **04** 결과 도출

Chapter 01

테이터 소개 및 분석주제

DATA 소개 및 분석 주제

- 분석 DATA 소개

- 원티드 국민연금 DATA (99801 rows x 6 columns)

	A	B	C	D	E	F	G
1	회사ID	연매출액	년도	월	월별_연금	월별_직원수	
2	233757	41688077	2018	7	8912380	36	
3	123464	1760695	2018	3	942380	5	
4	342124	3221341	2018	7	781180	4	
5	386420	4815584	2015	11	3795900	14	
6	129243	9799736	2018	10	40724680	151	
7	508810	239047	2018	3	528980	5	
8	365169	3650711	2016	4	570320	3	
9	189562	7119908	2016	1	1276900	6	
10	133518	18642904	2018	10	16844840	75	
11	190847	2367117	2015	12	2354660	8	
12	36600	1321988	2018	2	1518840	9	
13	65866	374795	2018	4	832500	5	
14	295938	13589790	2016	12	13187980	66	
15	311326	3578929	2016	2	1931260	8	

- ‘회사ID’ : 기업들을 식별, 5000개의 ID 존재.
- ‘연매출액’ : 해당 기업들의 연매출액을 나타냄.
- ‘년도’ : 2015부터 2019까지의 년도 데이터 존재.
- ‘월’ : 1월부터 12월까지의 월 데이터 존재.
- ‘월별_연금보험료’ : 기업들이 월별로 지불한 연금보험료를 나타냄.
- ‘월별_직원수’ : 기업들의 월별 직원 수 현황을 나타냄.

- 분석 주제

- 위의 데이터를 이용해서 나름의 방식으로 “유니콘 기업” 발굴하기

DATA 소개 및 분석 주제

• DATA 탐색

- 원티드 국민연금 DATA

기본적인 정보

▶ 자료형 및 요약정보 확인

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99801 entries, 0 to 99800
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype  
---  -
0   회사ID          99801 non-null  int64  
1   연매출액        99801 non-null  int64  
2   년도            99801 non-null  int64  
3   월              99801 non-null  int64  
4   월별_연금보험료 99801 non-null  int64  
5   월별_직원수     99801 non-null  int64  
dtypes: int64(6)
memory usage: 4.6 MB
```

```
# 결측치 확인
```

```
df.isnull().sum().to_frame().T
```

▶ NULL 값 확인

회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
0	0	0	0	0	0

- 모든 데이터는 int type
- 당장 눈에 보이는 null은 없음

▶ 컬럼별 통계 정보 확인

```
df.describe()
```

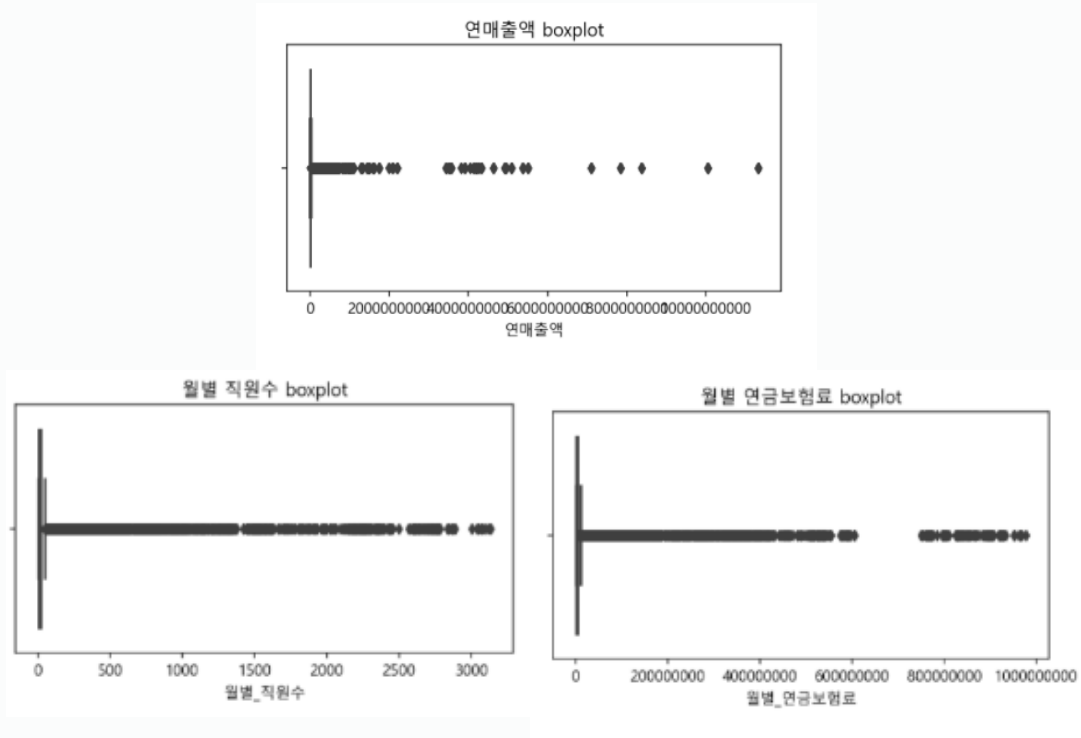
	회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
count	99801.00	99801.00	99801.00	99801.00	99801.00	99801.00
mean	276298.17	30535893.43	2016.97	6.71	9734079.21	37.62
std	133903.15	301664669.02	0.98	3.51	43472560.17	146.95
min	63.00	-7882176.00	2015.00	1.00	0.00	0.00
25%	165530.00	1409760.00	2016.00	4.00	1181160.00	6.00
50%	259149.00	3573585.00	2017.00	7.00	2332760.00	11.00
75%	382656.00	9582321.00	2018.00	10.00	5591860.00	24.00
max	5159723.00	11328779396.00	2019.00	12.00	978067260.00	3133.00

- 월별 연금보험료와 월별 직원수가 0인 회사 존재
- 연매출액이 마이너스인 회사 존재
- 연매출액과 월별 연금보험료, 월별 직원수 모두 최댓값과 제3사분위수값의 차이가 큼

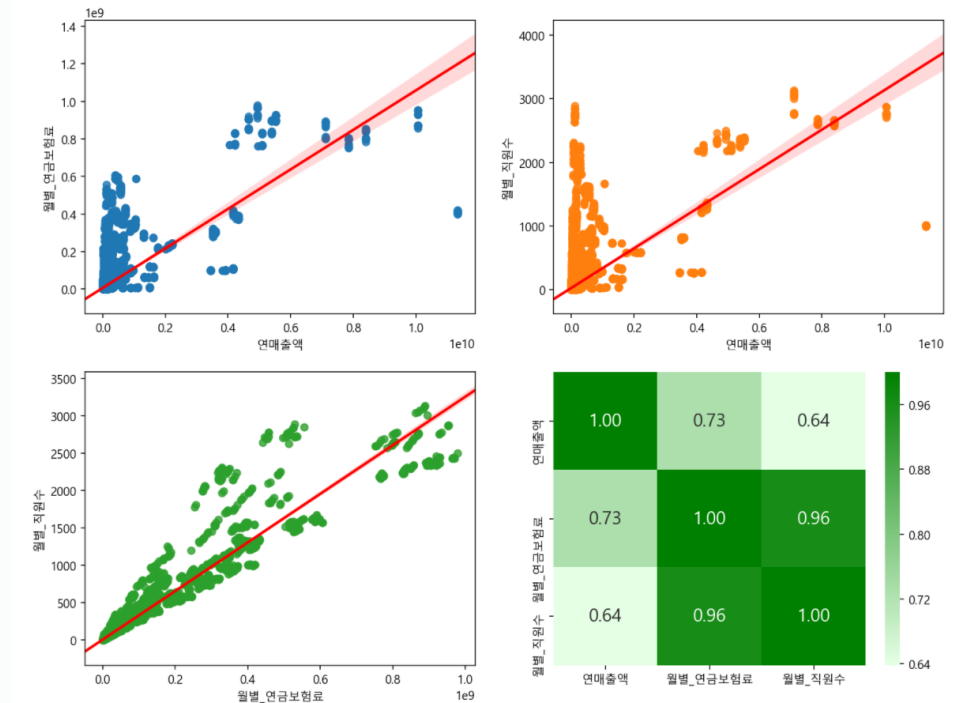
DATA 소개 및 분석 주제

- DATA 탐색

- 원티드 국민연금 DATA



▶ 변수별 이상치 확인(boxplot)



- 연매출액과 월별 연금보험료는 선형적 관계(강한 양의 상관관계: 0.73)
- 연매출액과 월별 직원수는 선형적 관계(양의 상관관계: 0.64)
- 월별 연금보험료와 월별 직원수는 선형적 관계(강한 양의 상관관계: 0.96)

▶ 회귀선 그래프(regplot)와 히트맵(heatmap)을 이용한 상관관계 시각화

Chapter 02

분석방향성 설정

● 분석 주제 알아보기

- 유니콘 기업이란?

정의 : 유니콘 (Unicorn)기업은 **기업가치가 10억 달러(=1조원) 이상**이고 창업한지 **10년 이하인 '비상장 스타트업'**을 말함

유래 : 상장 전 기업가치가 10억 달러 이상 된다는 것이 유니콘처럼 희귀하다는 의미로 명명

- 유니콘 기업 리스트(2020년 4월 기준) :

1. 쿠팡 (2014.05 등극)
2. 크래프톤 (2018.08 등극)
3. 우아한형제들 (2018.12 등극)
4. 위메프 (2019.04 등극)
5. 지피클럽 (2019.06 등극)
6. 아놀자 (2019.02 등극)
7. L&P 코스메틱 (2017.04 등극)
8. 무신사 (2019.11 등극)
9. 비바리퍼블리카 (2018.05 등극)
10. 에이프로젠 (2019.12 등극)
11. 옐로모바일 (2014.11 등극)
12. 쏘카 (2020.05 등극)

매출 순위	기업명	기업가치	매출액		매출 등락률	매출 대비 기업가치	영업이익	
			2019년	2018년			2019년	2018년
1	쿠팡	10.4조	7조 1530억	4조 3545억	64.3	1.5	-7205억	-1조 970억
2	크래프톤	5.8조	1조 874억	1조 1200억	-2.9	5.3	3592억	3002억
3	우아한형제들	3조	5654억	3145억	79.8	5.3	-364억	525억
4	위메프	3조	4653억	4294억	8.3	6.4	-758억	-390억
5	지피클럽	1.5조	4687억	5544억	-15.5	3.2	1227억	2067억
6	아놀자*	1.2조	2450억	1213억	102.0	4.8	-101억	-168억
7	L&P코스메틱	2조	2349억	3207억	-26.8	8.7	-135억	434억
8	무신사	2.2조	2197억	1073억**	104.8	10.0	493억	257억
9	비바리퍼블리카	2.5조	1187억	548억	116.6	20.8	-1154억	-445억
10	에이프로젠	1.2조	279억	533억	-47.7	40.0	-399억	26억
-	옐로모바일	4.6조	-	4699억	-	-	-	-318억

* 금융감독원 연결재무제표 기준 ** 감사 받지 않은 재무제표 기준

<http://chloe-ki.tistory.com>

• 분석 주제 알아보기

- 유니콘 기업이란?

정의 : 유니콘 (Unicorn)기업은 **기업가치가 10억 달러(1조원) 이상**이고 창업한지 **10년 이하인 '비상장 스타트업'**을 말함

유래 : 상장 전 기업가치가 10억 달러 이상 된다는 것이 유니콘처럼 희귀하다는 의미로 명명

- 유니콘 기업 리스트(2020년 4월 기준) :

1. 쿠팡 (2014.05 등극)

2. 크래프톤 (2019.04 등극)

3. 우아한형제들 (2019.04 등극)

4. 위메프 (2019.04 등극)

5. 지피클럽 (2019.06 등극)

6. 아놀자 (2019.02 등극)

7. L&P 코스메틱 (2017.04 등극)

8. 무신사 (2019.11 등극)

9. 비바리퍼블리카 (2014.05 등극)

10. 에이프로젠 (2019.12 등극)

11. 옐로모바일 (2014.11 등극)

12. 쏘카 (2020.10 등극)

“유니콘 기업들의 데이터를 찾아 기준을 잡자!!”

	매출 순위	기업명	기업가치	매출액		매출 동작률	매출 대비 기업가치	영업이익	
				2019년	2018년			2019년	2018년
1. 쿠팡 (2014.05 등극)	1	쿠팡	5.8조	1조 874억	1조 1200억	-2.9	5.3	3592억	3002억
2. 크래프톤 (2019.04 등극)	2	크래프톤	5.8조	1조 874억	1조 1200억	-2.9	5.3	3592억	3002억
3. 우아한형제들 (2019.04 등극)	3	우아한형제들	3조	5654억	3145억	79.8	5.3	-364억	525억
4. 위메프 (2019.04 등극)	4	위메프	3조	4653억	4294억	8.3	6.4	-758억	-390억
5. 지피클럽 (2019.06 등극)	5	지피클럽	1.5조	4687억	5544억	-15.5	3.2	1227억	2067억
6. 아놀자 (2019.02 등극)	6	아놀자*	1.2조	2450억	1213억	102.0	4.8	-101억	-168억
7. L&P 코스메틱 (2017.04 등극)	7	L&P코스메틱	2조	2349억	3207억	-26.8	8.7	-135억	434억
8. 무신사 (2019.11 등극)	8	무신사	2.2조	2197억	1073억**	104.8	10.0	493억	257억
9. 비바리퍼블리카 (2014.05 등극)	9	비바리퍼블리카	2.5조	1187억	548억	116.6	20.8	-1154억	-445억
10. 에이프로젠 (2019.12 등극)	10	에이프로젠	1.2조	279억	533억	-47.7	40.0	-399억	26억
11. 옐로모바일 (2014.11 등극)	-	옐로모바일	4.6조	-	4699억	-	-	-	-318억

* 금융감독원 연결재무제표 기준 ** 감사 받지 않은 재무제표 기준

<https://www.ki.seoul.go.kr>

분석 방향성 설정

- 데이터 수집

- 수집 경로

1. 크레딧잡 (<https://kreditjob.com/>) : 기업별 국민연금, 직원 수 데이터 추출
2. 중소기업현황정보시스템 (<http://sminfo.mss.go.kr/cm/sv/CSV001R0.do>) : 2015-2019 연매출액
3. 다투 (<https://dart.fss.or.kr/>) : 2015-2019 연매출액

- 수집 방법

- 팀원 별 2개 기업 데이터 수집.
- '위메프' 는 결측치가 많아 제외.
- 분석 데이터와 같은 형태로 데이터 수집.
- 11개 기업의 데이터
- 총 550 rows x 6 columns 로 구성

1	회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
2	지피클럽	22474017	2015	11	733104	8
3	지피클럽	22474017	2015	12	745068	7
4	무신사	32930316	2015	11	5828500	61
5	무신사	32930316	2015	12	6015221	66
6	토스	99100	2015	11	29943650.7	27
7	토스	99100	2015	12	38069917.2	28
8	우아한형제들	49500000	2015	11	290972574	216
9	우아한형제들	49500000	2015	12	295823572.2	221
10	야놀자	29899218	2015	11	15376964.85	130
11	야놀자	29899218	2015	12	15973940.39	137
12	엘애피코스메틱	188859253	2015	11	12376873	95
13	엘애피코스메틱	188859253	2015	12	12564176	97
14	크래프톤	46400000	2015	11	35089049	244
15	크래프톤	46400000	2015	12	35005421	246
16	에이프로젠	1823518	2015	11	9353616.75	85
17	에이프로젠	1823518	2015	12	9241173.9	84
18	쏘카	44800000	2015	11	9175651.245	89
19	쏘카	44800000	2015	12	10209293.69	89
20	쿠팡	1133752000	2015	11	728449278	5465
21	쿠팡	1133752000	2015	12	733849435	5538
22	엘로모바일	318200000	2015	11	5392046	35
23	엘로모바일	318200000	2015	12	4943371	35

분석 방향성 설정

- 분석 방법

1. 유니콘 기업 선정에 있어서 가장 중요한 기준은 '기업가치'와 관련된 '영업이익', '주가 관련 지표들'이지만, 분석 데이터에선 연관 데이터들을 살펴볼 수 없다. 따라서, '연매출액', '연금보험료', '직원 수'가 중요한 변수가 된다.
2. 또한, 유니콘 기업 선정 기준에 있어서 **성장률(변수의 추이)** 또한 중요한 변수가 될 수 있다. 따라서, 시계열분석을 할 수 있는 '년도', '월' 변수와 위 3개의 변수들을 활용한 **파생변수들을 생성**할 수 있다.
3. '연매출액', '연금보험료', '직원 수' 및 여러 파생변수를 생성하여 ① 유니콘 기업으로서의 **최소 기준치를 적용**한 뒤, ② 변수 별 **TOP500개의 기업을 추출**하고 ③ 최대한 가능성이 높은(**변수에 가장 많이 겹치는**) 회사를 선정한다.

분석 방향성 설정

- 변수 선정

- 유니콘 데이터 EDA 후, 분석의 기준이 될 Feature 선정

1. 직관적 Feature :

1) '평균 연매출액 순위' (년 기준)

(연매출액이 가장 많은 기업을 추출할 수 있음)

2) '평균 연금보험료 순위'

(연금보험료를 가장 많이 지불하는 기업을 추출할 수 있음)

3) '평균 직원 수 순위'

(가장 직원을 많이 고용한 기업을 추출할 수 있음)

2. 파생 Feature :

4) '연매출액 연평균 성장률' (년 기준)

(해당 기간동안 연매출액의 증가 추세를 파악할 수 있음)

5) '1인당 연금보험료 연평균 성장률' (년 기준)

(직원들의 연금보험료 변화로 임금변화를 알 수 있음)

6) '1인당 매출액 연평균 성장률' (년 기준)

(1인당 기업 연매출에 얼마만큼 기여를 하는지 알 수 있음)

➡ 해당 6개의 변수들을 통해 해당 기업들이 얼마만큼 성장하는지 알 수 있음!!

분석 방향성 설정

- 변수 선정

- 유니콘 데이터 EDA 후, 분석의 기준이 될 Feature 선정

1. 직관적 Feature :

2. 파생 Feature :

1) '평균 연매출액 순위' (년 기준)

(연매출액이 가장 많은 기업을 추출할 수 있음)

4) '연매출액 연평균 성장률' (년 기준)

(해당 기업 5년 연매출액의 증가 추세를 파악할 수 있음)

3) '평균 직원 수 순위'

(가장 직원을 많이 고용한 기업을 추출할 수 있음)

6) '1인당 매출액 연평균 성장률' (년 기준)

(1인당 기업 연매출에 얼마만큼 기여를 하는지 알 수 있음)

**“해당 6개의 변수로 유니콘 기업을 먼저 분석한 후,
기준을 세워, 5000개의 기업 중 유니콘 기업을 찾아낼 수 있다!”**

➡ 해당 6개의 변수들을 통해 해당 기업들이 얼마만큼 성장하는지 알 수 있음!!

Chapter 03

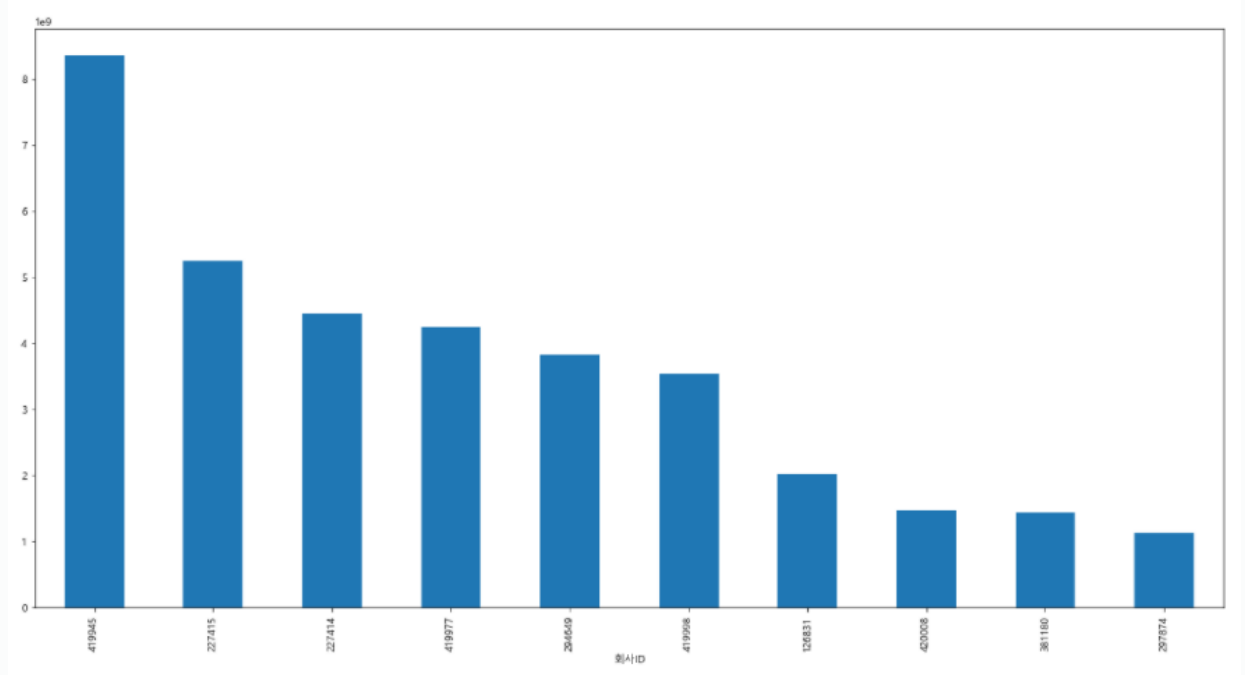
분석내용

분석내용

- '평균 연매출액' 가장 많은 순위 TOP500

- ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 -> 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
- ◎ 유니콘 기업 데이터 기준 등재시점 최소 연매출액인 17857139(천원) 미만의 데이터들 소거

회사 ID	
419945	8.353127e+09
227415	5.241049e+09
227414	4.459102e+09
419977	4.252086e+09
294649	3.833816e+09
419998	3.538393e+09
126831	2.015577e+09
420008	1.466671e+09
381180	1.434090e+09
297874	1.131216e+09



< 평균 연매출액이 가장 높은 기업 500개 중 상위 10개 기업 >

< 상위 10개 기업 시각화 >

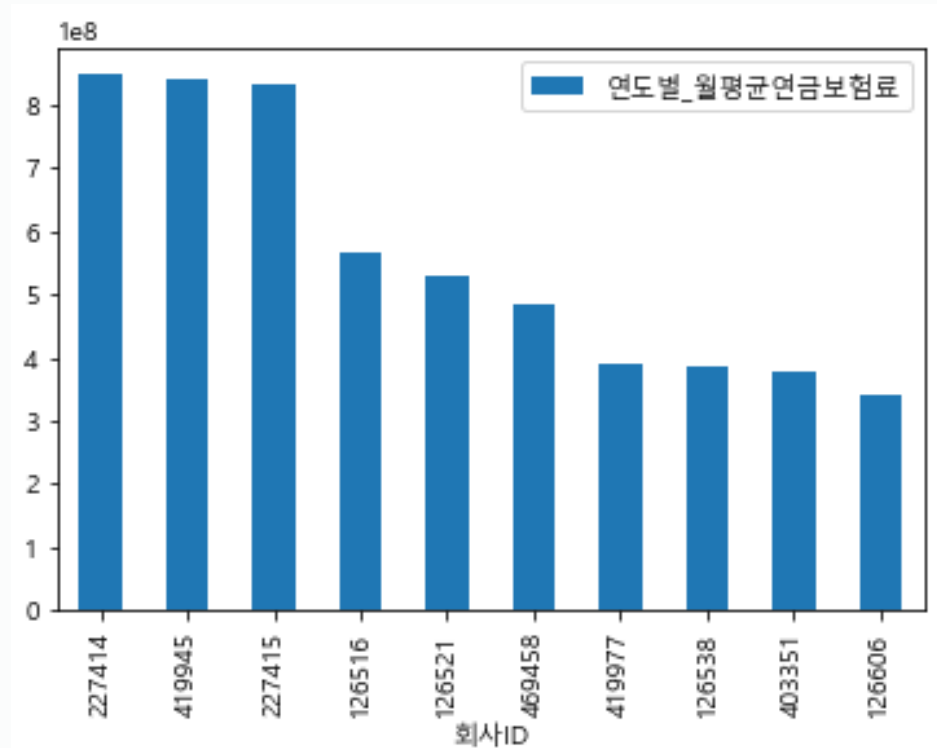
분석내용

- '평균 월별_연금보험료' (기업이 지불한 연금보험료의 평균)가 가장 큰 순위 TOP500

- ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 -> 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
- ◎ 유니콘 기업 데이터 기준 등재시점 최소 월별_연금보험료인 3796700.42(원) 미만의 데이터들 소거

회사ID	
227414	848532802.50
419945	838814933.33
227415	833628100.42
126516	564440444.44
126521	528434259.58
469458	482713946.11
419977	391094838.33
126538	386311959.58
403351	379280735.00
126606	340740658.33

< 평균 연금보험료가 가장 높은 기업 500개 중 상위 10개 기업 >



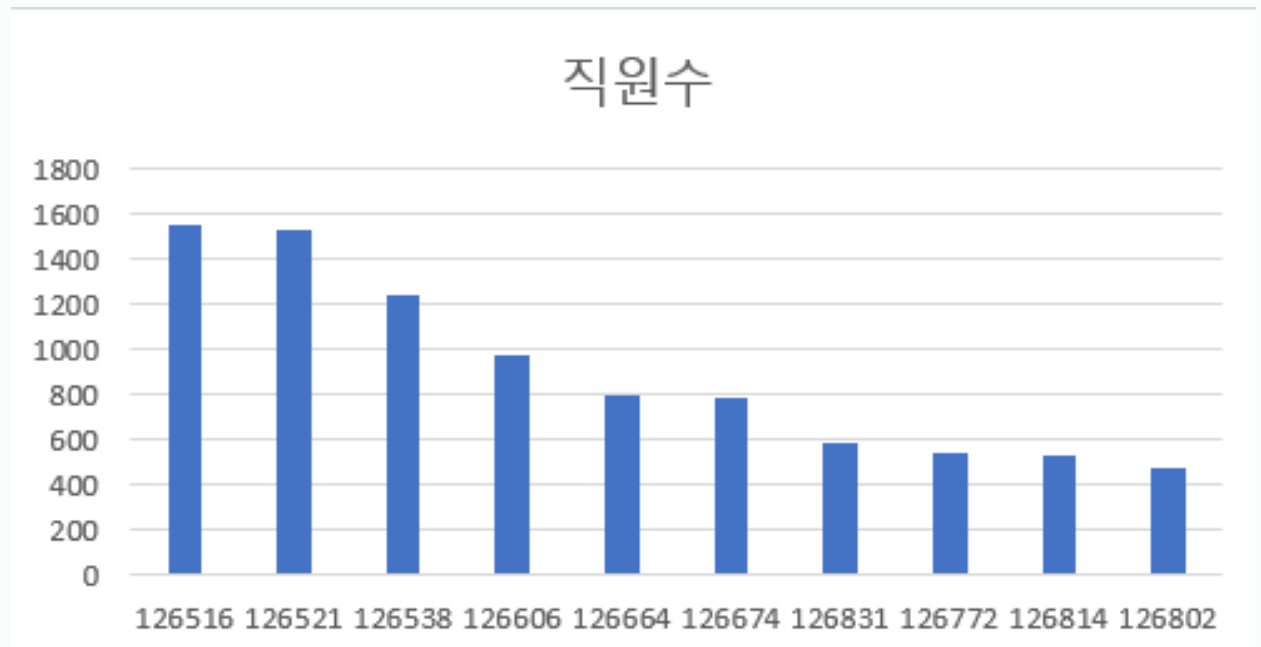
< 상위 10개 기업 시각화 >

분석내용

- '평균 월별_직원 수' 가 가장 많은 순위 TOP500

- ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 -> 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
- ◎ 유니콘 기업 데이터 기준 등재시점 최소 월별_직원 수 35(명) 미만의 데이터들 소거

회사ID	직원수
126516	1555.277778
126521	1532.583333
126538	1236.166667
126606	974.2708333
126664	796.9583333
126674	791.2291667
126831	584.7916667
126772	541.3125
126814	534.3125
126802	473.1875



< 평균 월별 직원 수가 가장 많은 기업 500개 중 상위 10개 기업 >

< 상위 10개 기업 시각화 >

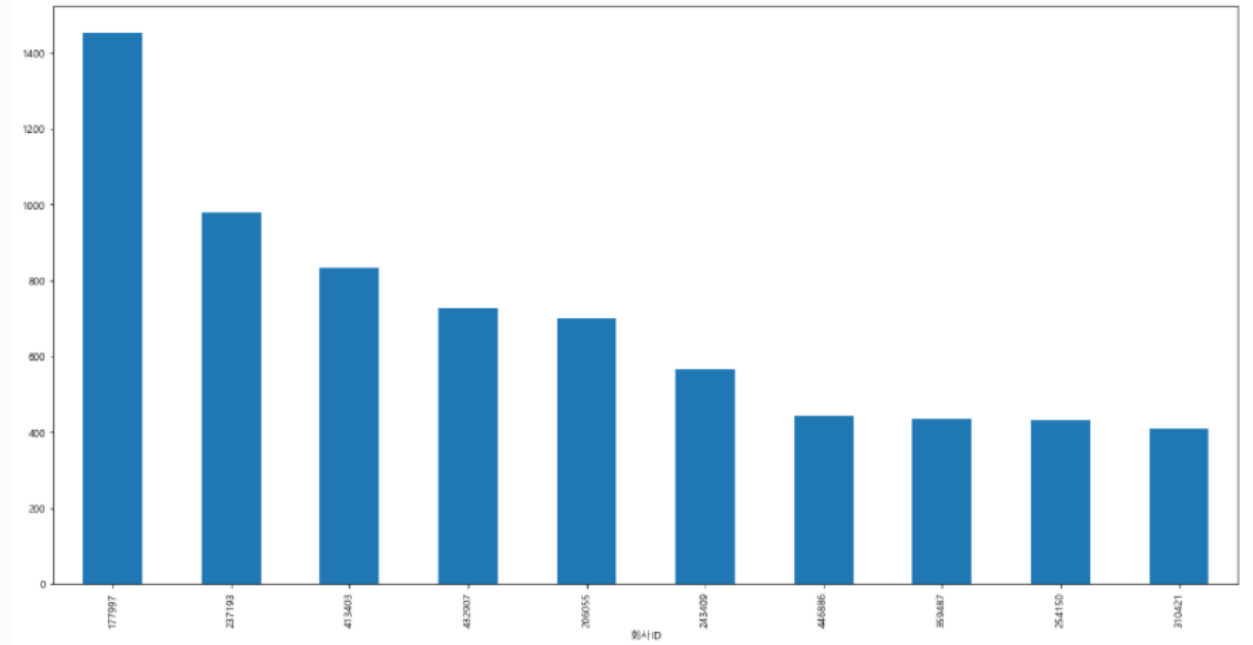
분석내용

- '연매출액 연평균 성장률(%)' 이 가장 큰 순위 TOP500

- '연평균 성장률(%)' : $((\text{종료값}/\text{시작값})^{(1/\text{기간수})}) - 1$

- ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 -> 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
- ◎ 유니콘 기업 데이터 기준 최소 연매출액 99100(천원) 미만의 데이터들 소거
- ◎ 유니콘 기업 데이터 기준 최소 연평균 성장률 -28.54% 미만의 데이터들 소거
- ◎ 유니콘 기업 데이터 기준 성장률 최고치의 최소값인 39.2% 미만의 데이터들 소거 -> 유니콘 기업들은 해당 년도에 한번은 최소 39.2%의 성장을 기록

회사ID	
177997	1451.48
237193	979.16
413403	834.29
432907	727.52
206055	700.00
243409	566.91
446886	443.21
359487	435.92
254150	432.73
310421	408.40



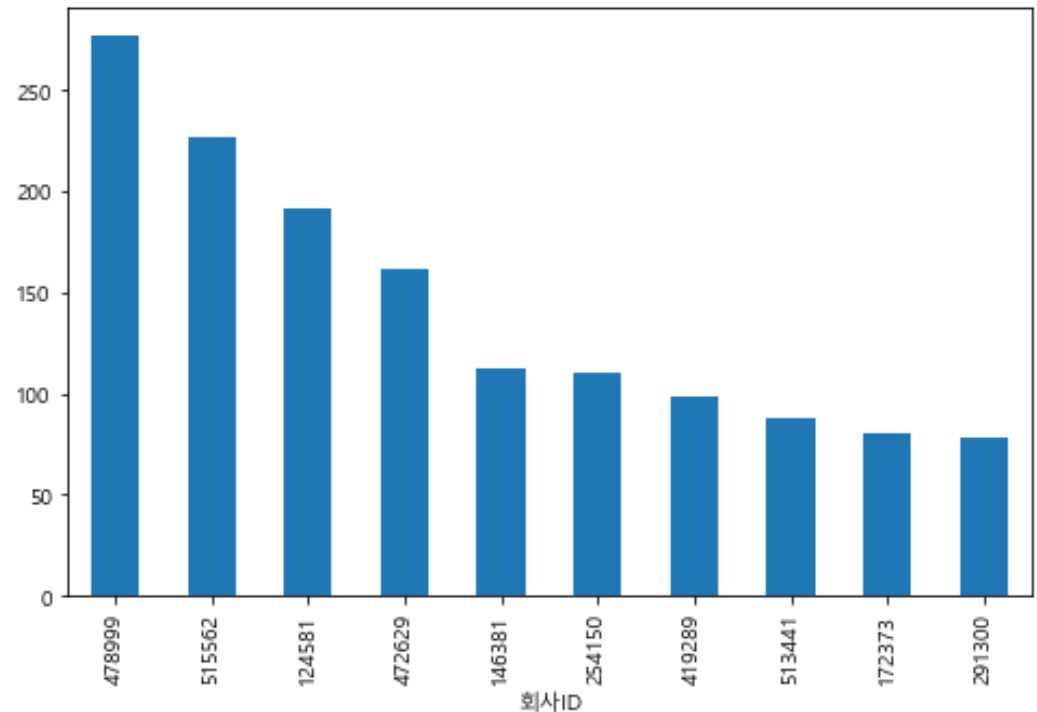
< 연매출액 연평균 성장률(%)이 가장 높은 기업 500개 중 상위 10개 기업 >

< 상위 10개 기업 시각화 >

분석내용

- '1인당 연금보험료 연평균 성장률(%)' (직원들의 연금보험료 변화 → 임금의 변화)이 가장 큰 순위 TOP500
 - '1인당 연금보험료' : (연금보험료) / (직원 수)
 - '연평균 성장률(%)' : $((\text{종료값} / \text{시작값})^{(1/\text{기간수})}) - 1$
 - ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 → 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
 - ◎ 유니콘 기업 데이터 기준 최소 연평균 성장률 1.25% 미만의 데이터들 소거

회사 ID	
478999	276.79
515562	226.54
124581	191.71
472629	161.71
146381	111.88
254150	110.77
419289	98.81
513441	88.02
172373	80.25
291300	77.83



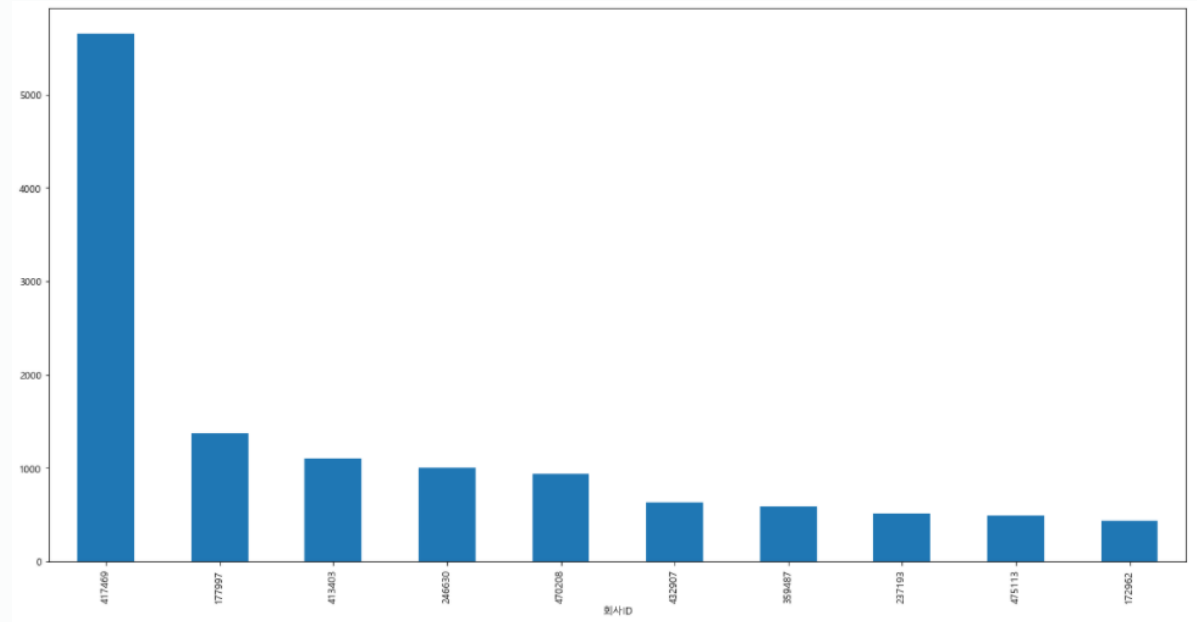
< 1인당 연금보험료 연평균 성장률(%)이 가장 높은 기업 500개 중 상위 10개 기업 >

< 상위 10개 기업 시각화 >

분석내용

- '1인당 연매출액 연평균 성장률(%)' (1인당 회사매출 기여도) 이 가장 큰 순위 TOP500
 - '1인당 연매출액' : (연매출액) / (직원 수)
 - '연평균 성장률(%)' : $((\text{종료값}/\text{시작값})^{(1/\text{기간수})}) - 1$
 - ◎ 기간이 0인 회사ID는 제외. (예를들어, 1년치 데이터만 존재하는 회사들은 소거 -> 회사가 얼마만큼 성장하는지 추이를 알 수 없기 때문)
 - ◎ 유니콘 기업 데이터 기준 최소 1인당 연매출액 3604(천원) 미만의 데이터들 소거
 - ◎ 유니콘 기업 데이터 기준 최소 연평균 성장률 -44.66% 미만의 데이터들 소거
 - ◎ 유니콘 기업 데이터 기준 1인당 연매출액 성장률 최고치의 최소값인 23.2% 미만의 데이터들 소거

회사ID	
417469	5646.83
177997	1377.14
413403	1103.88
246630	1002.49
470208	936.55
432907	635.00
359487	589.48
237193	508.90
475113	493.66
172962	439.04



< 1인당 연매출액 연평균 성장률(%)이 가장 높은 기업 500개 중 상위 10개 기업 >

< 상위 10개 기업 시각화 >

Chapter 04

결과 도출

결과도출

- 분석 결과

- 6개 Feature에 가장 많이 포함된 기업 추출

1. 6개의 Feature 별, 상위 랭크 TOP500(상위 10%)에 있는 '회사ID' 들을 추출.
2. 추출 후, 가장 많이 겹치는 기업 6곳을 선정.
3. 6개의 Feature에서 상위권을 기록했으므로, 직관적인 수치에서도 높은 값을 보이며, 성장률도 높은 수치를 보이므로 유니콘 기업일 가능성이 가장 높다.



1 2 7 3 6 6

1 3 8 7 0 9

1 3 3 4 9 3

1 2 8 4 0 5

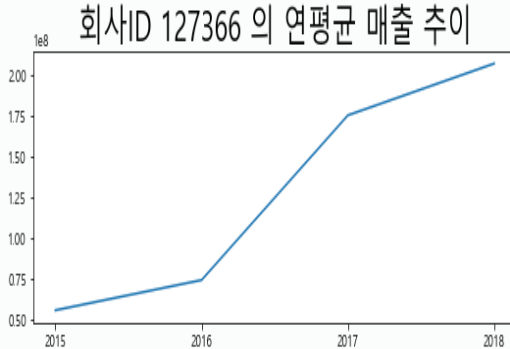
1 3 5 0 3 7

1 2 9 5 2 7

결과도출

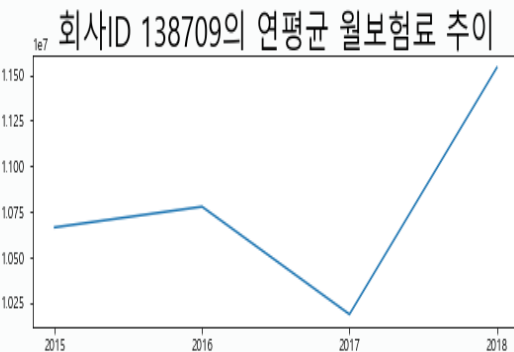
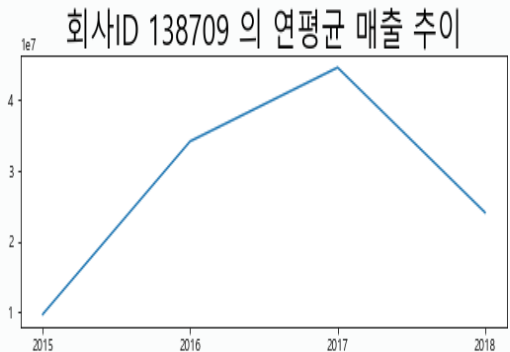
- 분석 결과 검증

- 회사ID 127366



- 매출액, 연금보험료, 직원수에 있어서 꾸준한 성장을 보임.

- 회사ID 138709

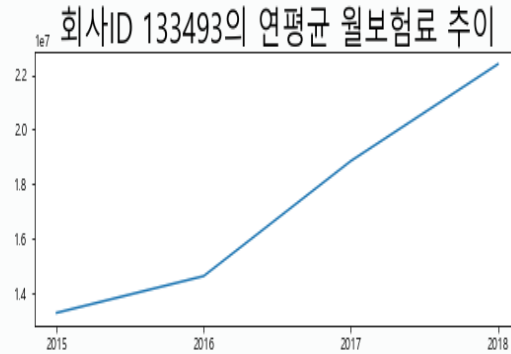
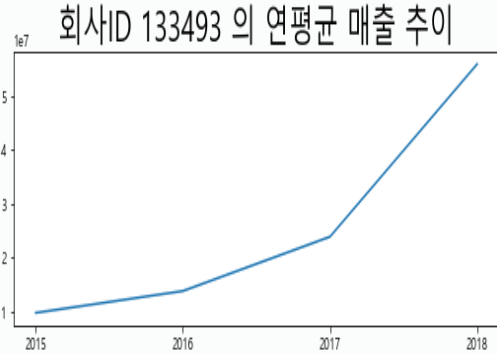


- 연매출이 꾸준히 증가하다 18년에 감소하지만 연금보험료와 직원수는 증가.
- 성장을 위해 도약을 준비하는 단계로도 해석 가능. (유니콘 데이터에서도 관찰 가능)

결과 도출

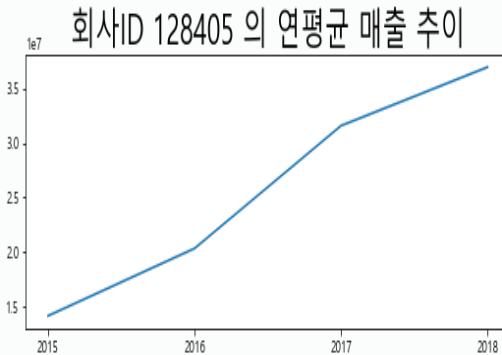
- 분석 결과 검증

- 회사ID 133493



- 매출액, 연금보험료, 직원수에 있어서 꾸준한 성장을 보임.

- 회사ID 128405

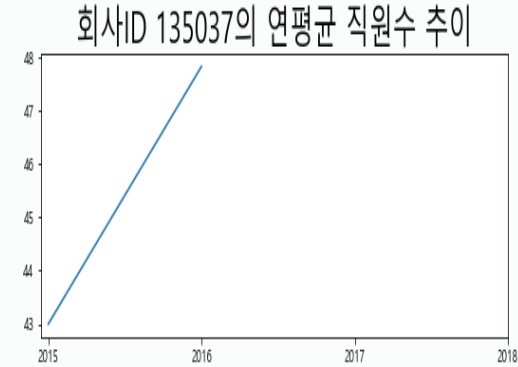


- 연금보험료가 16년에 주춤했지만 이후 꾸준히 성장.
- 매출액, 직원수에 있어서 꾸준한 성장을 보임.

결과도출

● 분석 결과 검증

- 회사ID 135037



- 매출액, 연금보험료, 직원수에 있어서 꾸준한 성장을 보임.
- 단, 데이터가 2년치만 존재.

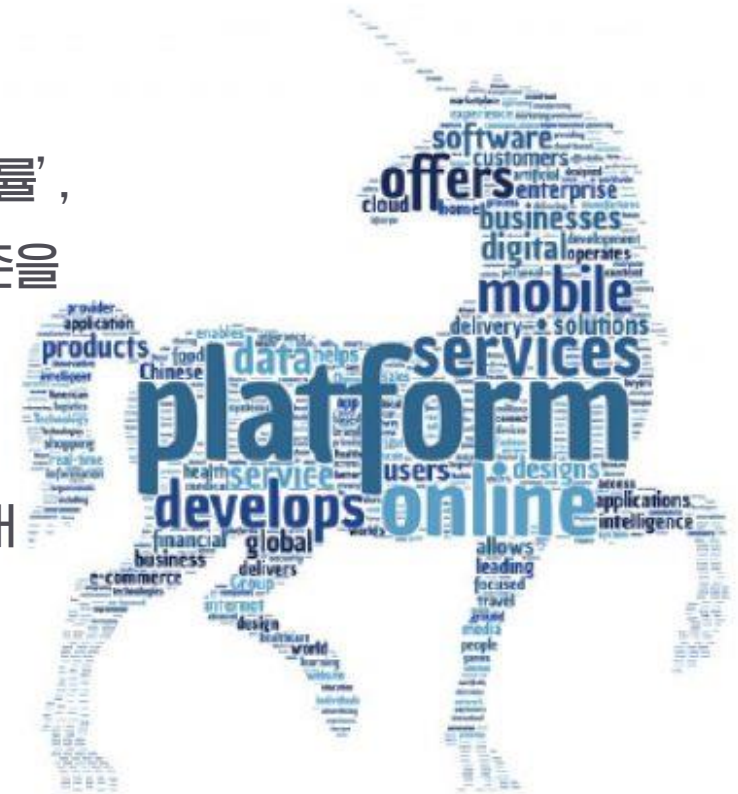
- 회사ID 129527



- 매출액, 연금보험료, 직원수에 있어서 꾸준한 성장을 보임.
- 단, 데이터가 2년치만 존재.

- 분석 결과 검증

1. 해당 6개의 기업들은 '유니콘 기업 데이터' 에서 추출한 기준인 '평균 연매출액 17857139(천원) 이상', '평균 연금보험료 3796700.42(원) 이상', '평균 직원 수 35(명) 이상', '최소 2개년 데이터 존재' 를 모두 만족한다.
2. 또한, 해당 기업들은 '유니콘 기업 데이터' 에서 추출한 '매출액 연평균 성장률', '1인당 연금보험료 연평균 성장률', '1인당 매출액 연평균 성장률' 의 기준을 모두 만족한다.
3. '유니콘 데이터' 의 기준을 모두 만족하고, 뚜렷한 성장률을 보이는 해당 6개 기업이 5000개의 기업 중, 유니콘 기업일 가능성이 가장 높다!



참고 및 출처

Python 분석 코드 – https://github.com/kevinpark2021/team3_wanted_onboarding (깃허브)

그림1 , 게티이미지뱅크 (gettyimagesbank.com)

그림2 , KOCAM (<http://www.kocam.org/?p=18892>)

DATA , 원티드랩 (<https://www.wanted.co.kr/>)

DATA, 크레딧잡 (<https://kreditjob.com/>)

DATA, 중소기업현황정보시스템 (<http://sminfo.mss.go.kr/cm/sv/CSV001R0.do>)

DATA, 다투 (<https://dart.fss.or.kr/>)

참고자료, 유니콘 기업(<https://better-together.tistory.com/191>)

참고자료, 스타트업의 기업 가치 평가(<https://brunch.co.kr/@jyzz21/7>)

감사합니다.
