

# Unicorn Company Search

원티드 제공 국민연금 데이터 분석을 통한 유니콘 기업 찾기

안태운



# CONTENT

01

프로젝트 개요

02

데이터 전처리

03

데이터 분석

04

가설 수립

05

결론

- 원티드 국민연금 데이터의 기업데이터를 활용해 유니콘 기업으로 보이는 기업 선별
- 비상장 기업은 기업 가치 평가의 절대적인 기준이 존재하지 않으므로 주어진 데이터를 활용해 기업의 가치를 논리적으로 평가하는 것이 사실상 본 프로젝트의 목표라고 할 수 있음
- 분석데이터 : company\_nps\_data.csv(원티드 제공 국민연금 데이터, 회사명은 ID형태로 처리)

```
df = pd.read_csv('./company_nps_data.csv', encoding='cp949')  
print(df.shape)  
df.head()
```

(99801, 6)

	회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
0	233757	41688077	2018	7	8912380	36
1	123464	1760695	2018	3	942380	5
2	342124	3221341	2018	7	781180	4
3	386420	4815584	2015	11	3795900	14
4	129243	9799736	2018	10	40724680	151

➤ 원티드 제공 국민연금 데이터(company\_nps\_data.csv) 로딩

#기본정보

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
```

- 데이터 정보 확인
- 6개 컬럼으로 구성
- 정수형으로 구성된 컬럼들

```
# 결측치 확인
```

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

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

- 결측치는 확인되지 않음
- 매출이 음수인 기업등 이상치라고 의심되는 데이터는 있지만 확신할 수 없어 처리하지 않기로 판단

#기본정보

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
```

- 데이터 정보 확인
- 6개 컬럼으로 구성
- 정수형으로 구성된 컬럼들

```
uni = pd.read_csv('./unicorn.csv', encoding='cp949')
print(uni.shape)
uni.head()
```

(550, 6)

	회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
0	지피클럽	22474017	2015	11	733104.00	8
1	지피클럽	22474017	2015	12	745068.00	7
2	지피클럽	48329929	2016	1	551118.00	7
3	지피클럽	48329929	2016	2	488133.00	7
4	지피클럽	48329929	2016	3	1073632.00	7

- 실제 유니콘 기업의 데이터를 통해 유니콘기업을 유추하는 로직 수립
- 실제 유니콘 기업의 데이터 확보 후 unicorn.csv 파일로 제작



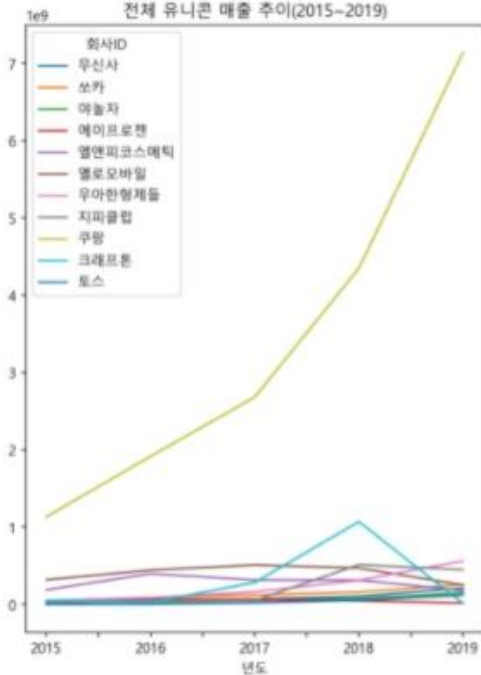
```
uni = pd.read_csv('./unicorn.csv', encoding='cp949')
print(uni.shape)
uni.head()
```

(550, 6)

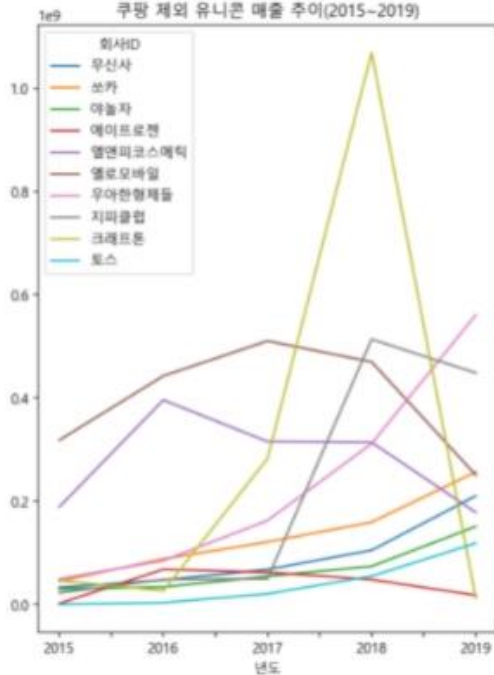
	회사ID	연매출액	년도	월	월별_연금보험료	월별_직원수
0	지피클럽	22474017	2015	11	733104.00	8
1	지피클럽	22474017	2015	12	745068.00	7
2	지피클럽	48329929	2016	1	551118.00	7
3	지피클럽	48329929	2016	2	488133.00	7
4	지피클럽	48329929	2016	3	1073632.00	7

- 실제 유니콘 기업의 데이터를 통해 유니콘기업을 유추하는 로직 수립
- 실제 유니콘 기업의 데이터 확보 후 unicorn.csv 파일로 제작

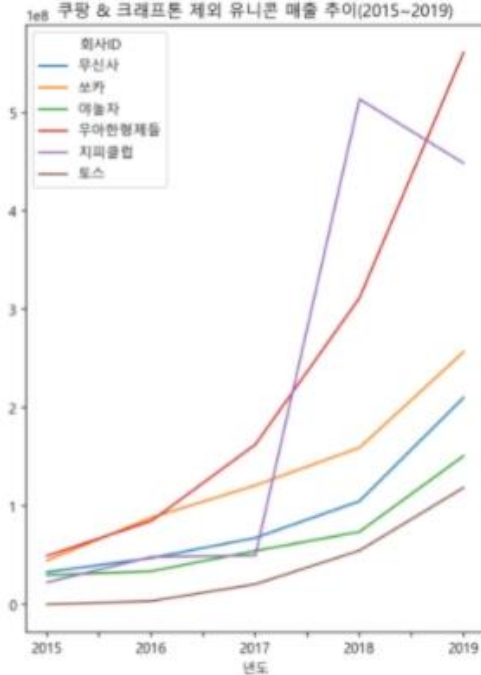
전체 유니콘 매출 추이(2015~2019)



쿠광 제외 유니콘 매출 추이(2015~2019)



쿠광 &amp; 크래프톤 제외 유니콘 매출 추이(2015~2019)



➤ 유니콘 기업들의 매출액 추이 확인

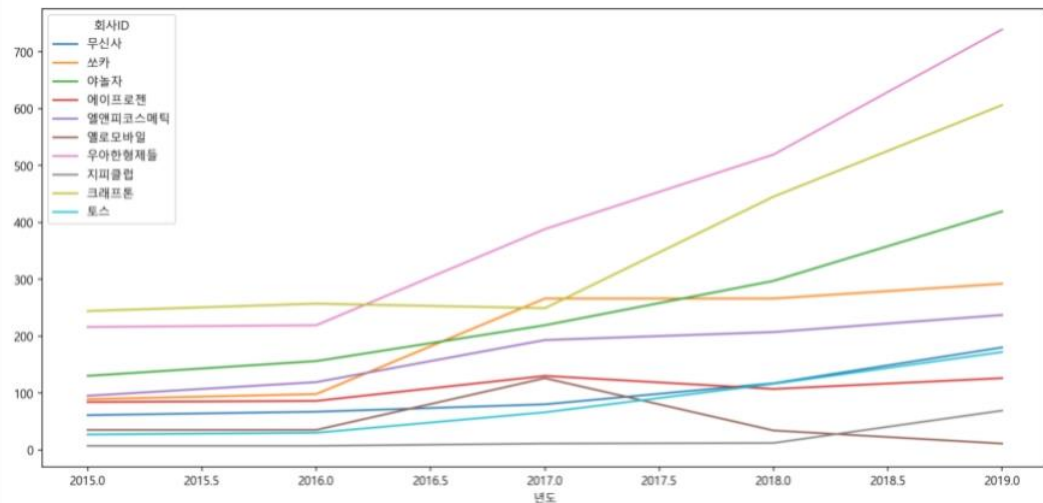
년도	2015	2016	2017	2018	2019	기간수	연평균 성장률(%)
회사ID							
토스	99100	3439965	20591083	54820876	118730969	4	488.33
지피클럽	22474017	48329929	49755792	513728710	448642100	4	111.38
우아한형제들	49500000	84850263	162565237	311535460	561133434	4	83.49
에이프로젠	1823518	68032078	62178381	48100298	17857139	4	76.90
무신사	32930316	47231240	67721259	105012384	210312756	4	58.97

- 유니콘 기업 선정을 위한 연매출액 성장률과 일인당 매출액 성장률의 기준이 필요하다

➤ 유니콘 기업들의 기간별 매출액 성장률 확인

```
# 쿠팡 직원수가 다른 기업에 비해 매우 높으므로 그래프를 그리면 다른 기업들의 추이를 확인하기 어려울. 따라서, 쿠팡 제외하고 시각화
uni.loc[uni['회사ID'] != '쿠팡'].pivot_table(index='회사ID',
                                             columns='년도',
                                             values='월별_직원수',
                                             aggfunc='min').T.plot(kind='line', figsize=(15, 7))
```

<AxesSubplot: xlabel='년도'>



➤ 유니콘 기업들의 직원 수 추이 확인

#기간 내 연금보험료 평균값이 높은 기업 분석

```
uni5 = uni4[['회사ID', '연도별_월평균연금보험료']]
```

```
uni5 = uni5.groupby(['회사ID']).mean()
```

```
uni5.reset_index(inplace=True)
```

```
uni5.sort_values(by='연도별_월평균연금보험료', ascending=False, inplace=True)
```

```
uni5
```

	회사ID	연도별_월평균연금보험료
8	쿠광	802979480.20
6	우아한형제들	762438227.47
10	토스	186789293.03
9	크래프톤	64068470.23
2	야놀자	42779282.34
4	엘애피코스메틱	25250614.65
1	쏘카	24969436.04
0	무신사	15247016.02
3	에이프로젠	14262471.57
5	엘로모바일	8955012.78
7	지피클럽	3732385.48

➤ 유니콘 기업의  
연도별  
월평균연금보  
험료 추이확인



#기간 내 연금보험료 평균값이 높은 기업 분석

```
uni5 = uni4[['회사ID', '연도별_월평균연금보험료']]
```

```
uni5 = uni5.groupby(['회사ID']).mean()
```

```
uni5.reset_index(inplace=True)
```

```
uni5.sort_values(by='연도별_월평균연금보험료', ascending=False, inplace=True)
```

```
uni5
```

	회사ID	연도별_월평균연금보험료
8	쿠광	802979480.20
6	우아한형제들	762438227.47
10	토스	186789293.03
9	크래프톤	64068470.23
2	야놀자	42779282.34
4	엘애피코스메틱	25250614.65
1	쏘카	24969436.04
0	무신사	15247016.02
3	에이프로젠	14262471.57
5	엘로모바일	8955012.78
7	지피클럽	3732385.48

➤ 유니콘 기업의  
연도별  
월평균연금보  
험료 추이확인

년도	2015	2016	2017	2018	2019	기간수	연평균 성장률(%)
회사ID							
토스	1234332.00	1331534.18	1586949.71	1730305.28	1858434.23	4	10.77
무신사	93344.45	97955.25	106840.89	117678.16	133904.69	4	9.44
야놀자	117441.23	108066.54	123614.54	135277.54	153600.66	4	6.94
우아한형제들	1342831.73	1409297.51	1503911.06	1596297.53	1737258.34	4	6.65
크래프톤	143053.02	138837.22	144683.14	157774.69	173417.85	4	4.93

➤ 유니콘 기업의 1인당 연금보험료 성장률 추이 확인

- 유니콘기업들은 다음의 6가지 항목에서 높은 측정값을 보임
  - 1. 직원수 평균
  - 2. 1인당 임금평균 성장률
  - 3. 월평균 임금보험료 평균
  - 4. 연매출액 평균
  - 5. 연매출액 평균성장률
  - 6. 일인당 매출액 평균성장률
- 따라서 위의 기준들에서 높은 값을 보여준다면 유니콘기업이라고 할 수 있음

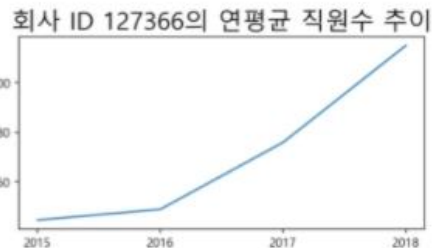
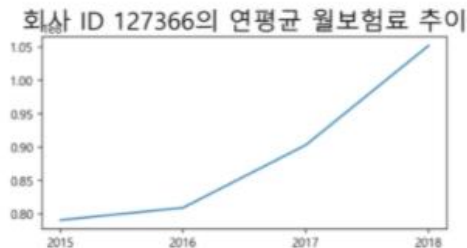
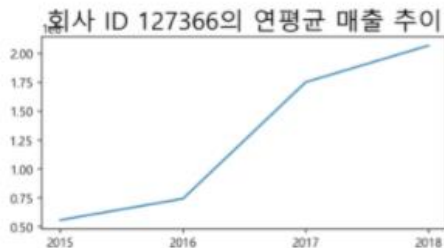
1인당 연금보험료 연평균 성장률	직원수	연매출액 연평균 성장률(%)	일인당 연매출액 연평균 성장률(%)	평균연매출액	월연금보험료
478999	63	177997	417469	419945	227414
515562	95	237193	177997	227415	419945
124581	102	413403	413403	227414	227415
472629	295	432907	246630	419977	126516
146381	414	206055	470208	294649	126521
254150	597	243409	432907	419998	469458
419289	786	446886	359487	126831	419977
513441	917	359487	237193	420008	126538

- 가설에서 정의한 항목별로 높은 값을 가진 기업들  
(항목별 TOP500) 선별

- 유니콘기업의 최소 조건을 만족하면서 6개 항목 TOP500 데이터에서 가장 빈도수가 높은 기업 선별결과 유니콘기업으로 판단되는 기업은 다음과 같음
- 1. 127366
- 2. 138709
- 3. 133493
- 4. 128405
- 5. 135037
- 6. 129527



```
#회사 ID 127366 주요 항목 추이 그래프로 확인
unicorn=df_year_gr[df_year_gr['회사ID']==127366]
plt.figure(figsize=(20,10))
plt.subplot(331)
plt.plot(unicorn['년도'],unicorn['연매출액'])
plt.title("회사 ID 127366의 연평균 매출 추이", fontsize=20)
plt.xticks([2015,2016,2017,2018])
plt.subplot(332)
plt.plot(unicorn['년도'],unicorn['월별_연금보험료'])
plt.title("회사 ID 127366의 연평균 월보험료 추이", fontsize=20)
plt.xticks([2015,2016,2017,2018])
plt.subplot(333)
plt.plot(unicorn['년도'],unicorn['월별_직원수'])
plt.title("회사 ID 127366의 연평균 직원수 추이", fontsize=20)
plt.xticks([2015,2016,2017,2018])
plt.show()
```



➤ 해당기업들의 평가항목 추이 확인