

인천시 산업별 종사자 수, 사업체 수, 매출액 분석

팀원 : 황상일, 김경남

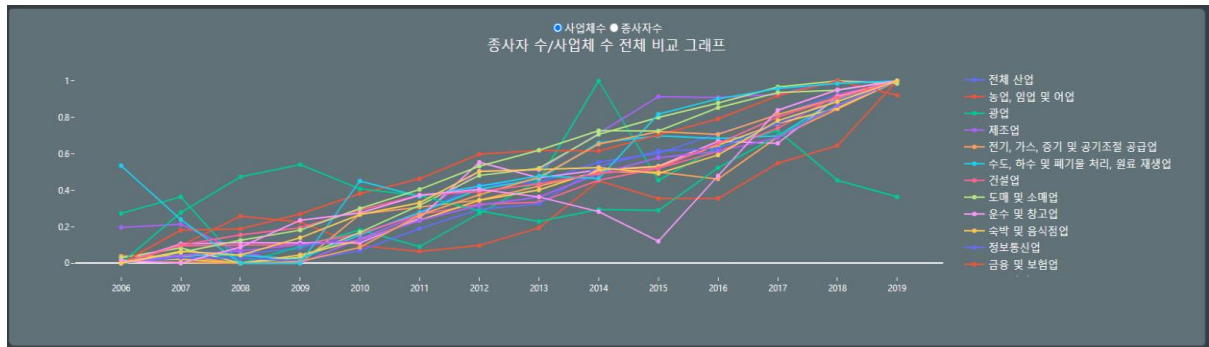
목차

1. 개요	3
2. 분석	
1) 종사자 수/사업체 수 전체 비교 그래프	4
2) 종사자 수/사업체 수 비교 그래프	4
3) 사업체 수/종사자 수 성장률 그래프	5
4) 종사자, 사업체, 매출액 현황	6
5) 히트맵 구현	7
6) 전년도 대비 증가율	8
7) 매출액 회귀 예측	9
4. 결론	9
활용 데이터	10
Python 버전 및 패키지 정보	10

개요

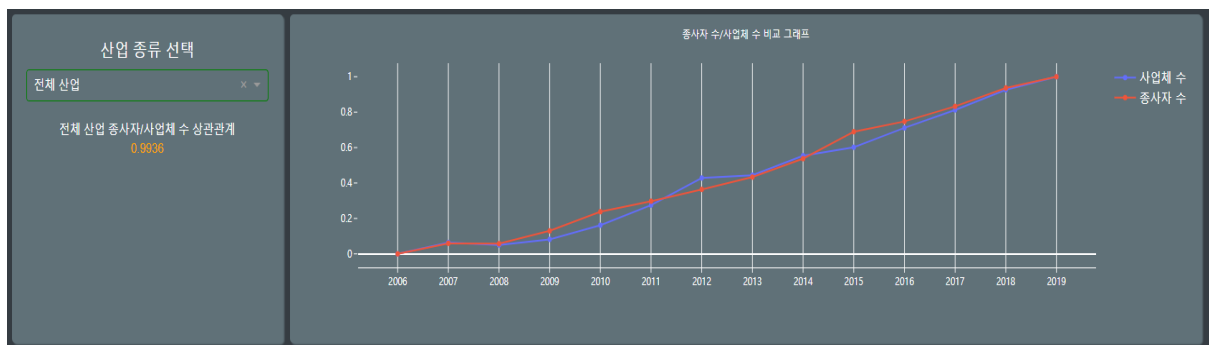
- 인천광역시의 2006 년 ~ 2019 년도의 산업별 종사자, 사업체 수 데이터와 2016~2019 년 종사자 수, 사업체 수, 매출액 데이터를 활용하여 산업별로 변화 추이를 분석해보고, 산업별 트렌드를 파악할 수 있으며, 종사자 수와 사업체 수의 상관분석을 통하여 산업별 동향을 파악할 수 있으며, 매출액 회귀분석을 통한 예상 매출액을 구하여 실제 데이터와 비교해 볼 수 있다.

1) 종사자 수/사업체 수 전체 비교 그래프



- 2006 년부터 2019 년도의 사업체 수, 종사자 수 동향 그래프로 라디오버튼으로 사업체, 종사자로 구분하여 볼 수 있게 구현하였다. 전체적인 흐름을 보기 위해 모든 산업별 데이터를 한 그래프에 넣어 표현하였고, 정규화를 진행하여 산업별로 변화 추이를 동일하게 표현했다.

2) 종사자 수/사업체 수 비교 그래프



- 첫 번째 그래프에서는 전체 산업을 포함한 산업별 데이터를 표현하였지만 여기서는 종사자 수와 사업체 수의 변화 추이를 동일한 산업별로 비교하기 위하여 구현했다. drop down box 를 통하여 산업 종류를 선택하면 해당 산업별로 파란색 선이 '사업체 수', 빨간색 선이 '종사자 수'로 그려지고 drop down box 밑에는 두 종류 간의 상관관계를 작성했다.

- 전체 산업별 상관 계수를 보면 0.9936 으로 매우 강한 상관관계를 보이고 있지만, '농업, 임업, 및 어업' 이나 '광업' 에서는 각각 0.5287, 0.5525 로 약한 상관관계를 보이고 있다.

사업체 수/종사자 수 성장률 그래프



- 2006 년부터 2019 년도의 종사자 수, 사업체 수 변화를 그래프로 나타낸 것이다. 각각 '전체 산업' 을 기준으로 구분했다.

상위 3	하위 3
1. 농업, 임업 및 어업	1. 광업
2. 정보통신업	2. 공공행정, 국방 및 사회보장 행정
3. 보건업 및 사회복지 서비스업	3. 제조업

표 1 사업체 수 성장률 상/하위 top3

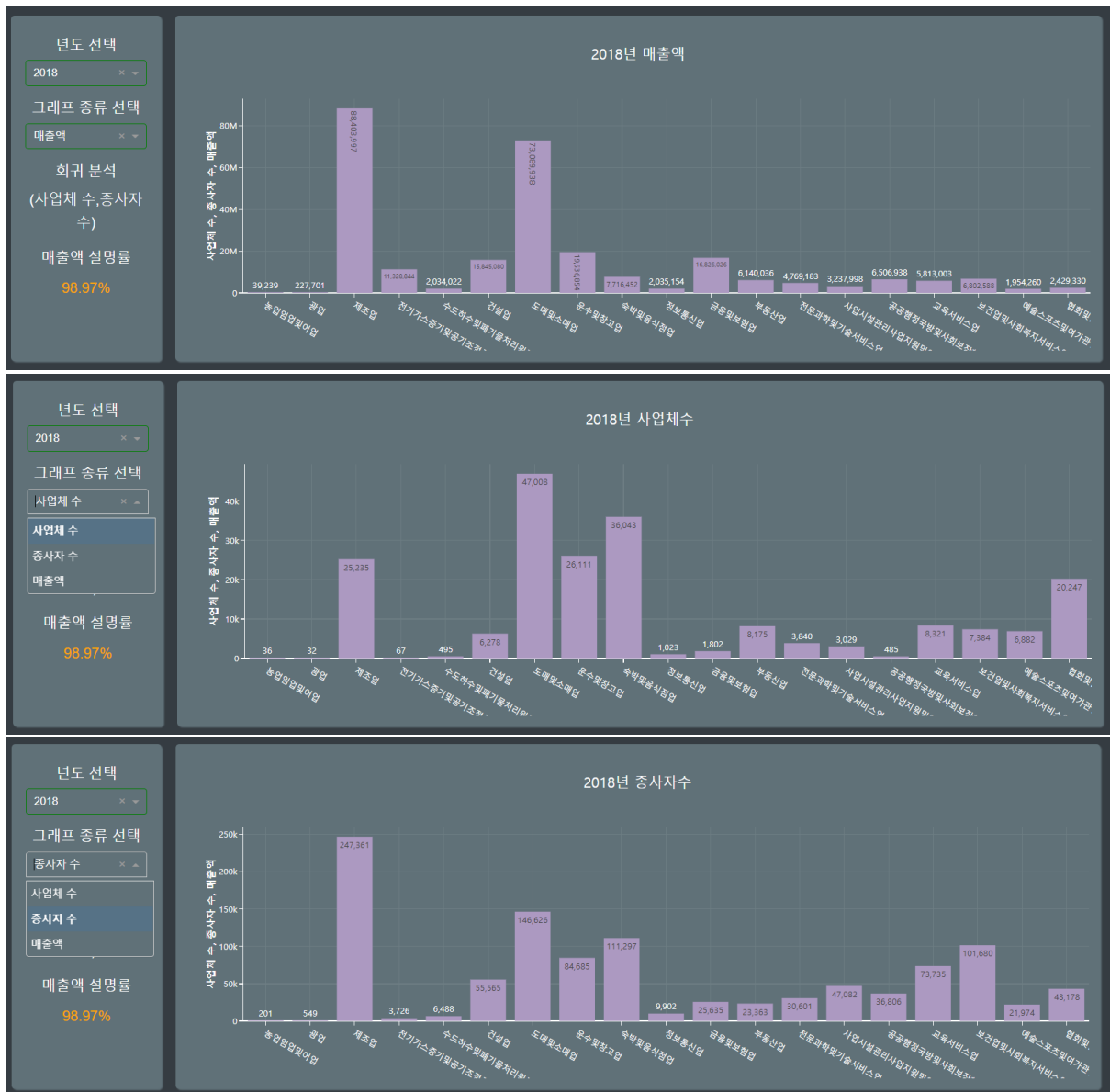
- 상위 3 순위를 보면 성장률 1 위인 '농업, 임업 및 어업' 이 조금 의아한 결과가 나와서 좀더 분석을 해봤다. 사업체 수 데이터는 2006 년 16 개에서 2019 년 47 개로 적은 개체수지만 약 3 배정도 수가 증가한 것을 알 수 있었다. 그러나 사업체 수가 평가 요소가 아니기 때문에 실질적인 평가지표라 할 수 있는 매출액을 보면 2016 년 50,199,000,000 원에서 2019 년 53,457,000,000 원으로 약 6 퍼센트 증가하였지만, 그 전 해인 2018 년도에는 39,239,000,000 으로 2016 년 기준보다 약 22 퍼센트 감소한 것을 알 수 있었다.

상위 3	하위 3
1. 보건업 및 사회복지 서비스업	1. 광업
2. 사업시설 관리, 사업지원 및 임대 서비스업	2. 금융 및 보험업
3. 전문, 과학 및 기술 서비스업	3. 제조업

표 2 종사자 수 성장률 상/하위 top3

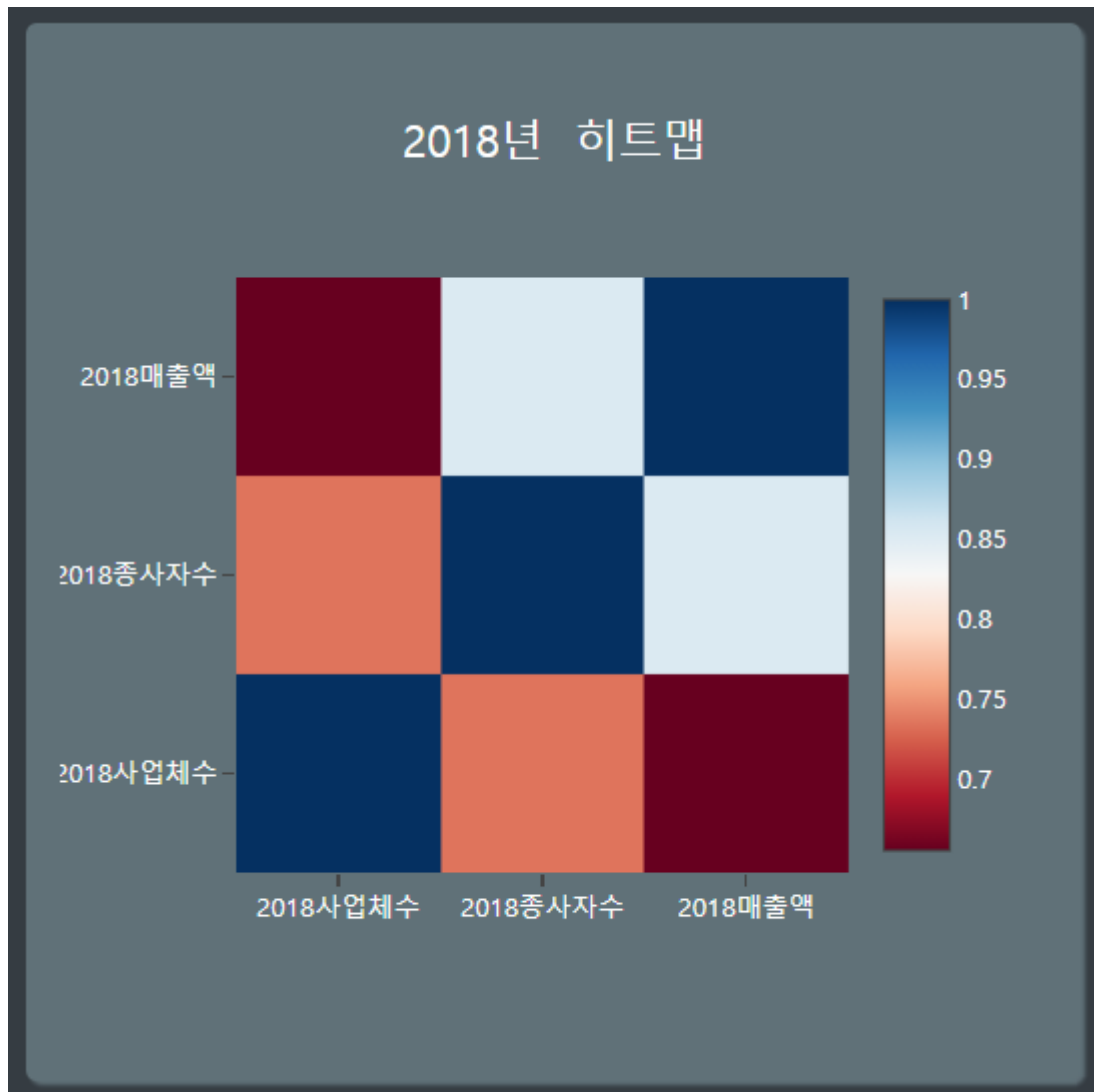
- 상위 3 종류를 보면 대체적으로 서비스업 직종에 종사하시는 분들이 늘어나고 있다는 것을 알 수 있고, 과거 물질적인 생산업에서 보이지 않는 서비스 업종으로 변화가 이뤄지는 것을 알 수 있다.

4) 종사자, 사업체, 매출액 현황



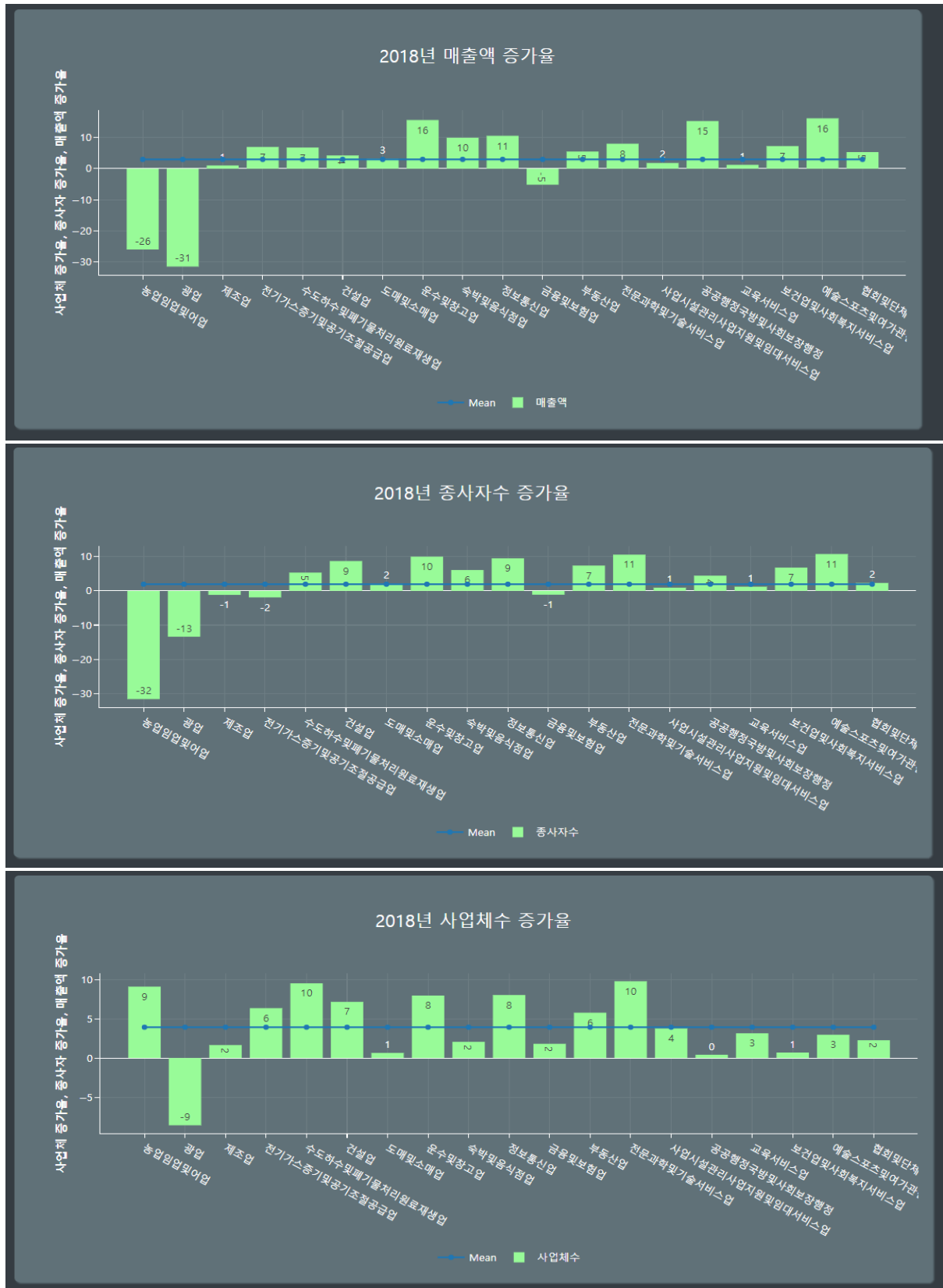
- 왼쪽에 Drop down box 에서 년도와 사업체 수, 종사자 수, 매출액을 선택하여 해당 년도의 산업별로 정보를 확인할 수 있다. 2018 년 매출액 그래프에서는 다른 산업들에 비해 제조업과 도매 및 소매업의 매출액이 다른 산업들에 비해 월등히 높은 걸 확인할 수 있다. 2018 년 사업체 수 그래프와 종사자 수 그래프에서도 제조업과 도매 및 소매업이 높은 걸 확인할 수 있다. 이것로 보아 종사자 수와 사업체 수가 어느정도 상관관계가 있어 보인다. 그래서 회귀분석 모델을 만들어보니 사업체 수와 종사자 수로 매출액을 70% ~ 98%정도 설명할 수 있다고 나온다. 학습데이터와 설명변수의 수가 적다 보니 설명력의 범주가 큰 것 같다. 하지만 설명력은 평균적으로 높아서 회귀 예측 모델을 구성하기에 적합해 보인다.

5) 히트맵 구현



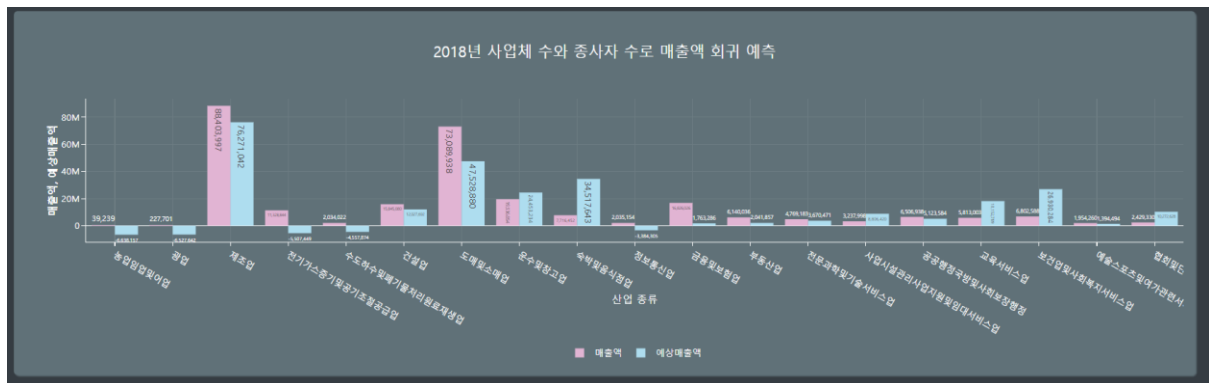
- Heatmap 을 보면 매출액은 사업체 수와 종사자 수와 꽤 강한 상관관계를 보인다. 회귀분석의 설명력도 범주가 크지만 평균적으로 설명력이 높기 때문에 Heatmap 과 같이 보면 매출액을 예측하기에 사업체 수와 종사자 수는 알맞은 설명변수로 보인다. 하지만 종사자 수와 사업체 수는 서로 상관관계가 높기 때문에 다중공선성의 문제가 있다. 적은 변수와 다중공선성의 문제로 모델의 정확성에 영향을 미칠 것으로 예상된다.

6) 전년도 대비 증가율



- 전년도 대비 사업체 수, 종사자 수, 매출액을 Drop down box 을 이용하여 확인해볼 수 있다. 앞에서 사업체 수와 종사자 수가 다중공선성의 문제가 있을 것이라고 예상했지만 위의 그래프를 확인해보면 상관관계가 그렇게 높지 않아 보이는 걸 확인할 수 있다. 그렇기 때문에 매출액 회귀 예측 모델을 구성할 때 사업체 수도 포함해도 될 것이라 생각된다.

7) 매출액 회귀 예측



- 반응 변수로 매출액을 주고 설명 변수로 사업체 수와 종사자 수를 가지고 회귀 예측한 결과이다. 학습 데이터와 설명 변수가 아직은 적어서 아쉬운 예측력임을 확인할 수 있다.

3. 결론

- 산업의 트렌드는 매출액과 매출액 증가율에 의미가 있다고 생각한다. 더 많은 학습데이터와 유의한 설명변수를 추가하면 산업의 트렌드(매출액)를 예측할 때 좋은 성능을 보일 것 같다. 당장의 결과로만 봤을 때는 종사자 수와 사업체 수는 매출액 예측에 큰 영향을 끼치는 것으로 보이기 때문에 가까운 년도를 예측할 때 구인사이트에서 산업별로 종사자와 사업체가 얼마나 증가하는지 종사자와 사업체가 얼마나 줄어드는지 확인하고 데이터로 입력하면 앞으로의 매출액이나 트렌드를 예측함에 있어 영향력이 좋을 것이라 생각된다. 트렌드를 예측해서 기업의 중요한 의사결정이나 투자분야, 기업의 사업추진방향 ,주식시장 동향파악 등에 이용하면 좋을 것이라 생각된다.

활용 데이터

- 인천광역시·산업·사업체구분별_사업체수_종사자수
- 인천광역시·산업별_사업체수_종사자수_및_매출액

출처 - 국가통계포털(<https://kosis.kr/index/index.do>)

Python 버전 및 패키지 정보

- python 3.7.8

package

- Brotli	1.0.9	- pyparsing	2.4.7
- click	8.0.1	- python-dateutil	2.8.2
- colorama	0.4.4	- pytz	2021.1
- cyclер	0.10.0	- scikit-learn	0.24.2
- dash	1.21.0	- scipy	1.7.1
- dash-core-components	1.17.1	- seaborn	0.11.2
- dash-html-components	1.1.4	- setuptools	57.4.0
- dash-table	4.12.0	- six	1.16.0
- Flask	2.0.1	- sklearn	0.0
- Flask-Compress	1.10.1	- tenacity	8.0.1
- future	0.18.2	- threadpoolctl	2.2.0
- importlib-metadata	4.7.0	- typing-extensions	3.10.0.0
- itsdangerous	2.0.1	- Werkzeug	2.0.1
- Jinja2	3.0.1	- zipp	3.5.0
- joblib	1.0.1		
- kiwisolver	1.3.1		
- MarkupSafe	2.0.1		
- matplotlib	3.4.3		
- numpy	1.21.2		
- pandas	1.3.2		
- Pillow	8.3.1		
- pip	21.2.4		
- plotly	5.2.2		