



Empowering Business with Data Analytics for Actionable Insights (1)



매출액은 증가했지만 영업이익이 증가하지 않는 원인 분석 및 해결방안을 제안 프로젝트

I. 문제정의

1. 프로젝트 목표
2. 프로젝트 전개 요약

II. 데이터 수집 및 전처리

1. 데이터 불러오기
2. 데이터 파악, 전처리
 - 2-1. 지역별 업종기준 기업 통계 데이터
 - 2-2. 주가 데이터
 - 2-3. 재무 데이터
 - 2-4. 사업보고서
 - 2-5. 네이버 뉴스기사 본문

III. 분석

1. 기업의 재무 현황 파악
2. 영업이익 관계 변수 파악
3. 비용 구조 분석
 - 1) 절대적 원인 분석
 - 2) 상대적 원인분석
4. 키워드 분석을 통한 기업의 동향 파악
5. Task Use Idea

IV. 결과 해석

1. 분석 결과의 의미
2. 결과에 대한 해석 및 활용 방안

VI. 프로젝트 요약

1. 프로젝트 과정 요약
2. 프로젝트를 통해 얻은 인사이트, 피드백
3. 사용한 도구 및 라이브러리, 참고자료 모음
4. 전체 과정의 코드

I. 문제정의

1. 프로젝트 목표

- A사는 지난 3개년 동안 매출은 매해 10% 이상 성장해 왔으나 영업이익은 증가하지 않거나 오히려 감소해왔습니다.
- 데이터 분석을 활용하여 이에 대한 원인을 파악합니다.
- 원인 분석을 위한 접근 방법과 활용할 수 있는 분석 기법 및 도구가 무엇인지 제안합니다.
- 분석을 수행하기 위해 어떠한 데이터가 필요한지 정의합니다.
- A사가 속한 산업은 자유롭게 가정 후 주어진 문제를 해결합니다.

2. 프로젝트 전개 요약

⇒ 접근방법을 보다 이해하기 쉽게 설명하기 위해서 **삼성카드**의 실제 재무 데이터를 이용했습니다.

1. 기업의 재무 현황 파악

- 필요 데이터 : 기업의 재무 데이터(dart_fss이용)

2. 영업이익 관계 변수 파악

- 필요 데이터 : 재무 데이터, 재무 관련 변수가 있는 데이

3. 산업군 분석

- 필요 데이터 : 지역별 업종기준 기업 통계 데이터

4. 키워드 분석을 통한 기업의 동향 파악

- 필요 데이터 : 네이버 뉴스 오픈 API를 활용해서 사업체명을 검색했을 나오는 뉴스의 본문 크롤링(검색 키워드 : 기업명 + 영업이익)

5. Task Use Idea

- 검색 키워드 모델 제작
- 매출&영업이익 분석 단어사전 제작

II. 데이터 수집 및 전처리

1. 데이터 불러오기

Data List

Aa 이름	≡ 수집방법	📎 데이터
<u>지역별 업종기준 기업 통계 데이터</u>	통계청	<u>region_category_preprocess.csv</u>
<u>주가 데이터</u>	pykrx라이브러리 이용	<u>stock_preprocess.csv</u>
<u>재무 데이터</u>	dart_fss라이브러리 이용	<u>financial_data_preprocess.csv</u>
<u>삼성카드 사업보고서</u>	dart 홈페이지 다운로드	[삼성카드]사업보고서 (2023.03.08).pdf
<u>네이버 뉴스기사 본문</u>	크롤링(오픈API활용)	

2. 데이터 파악, 전처리

2-1. 지역별 업종기준 기업 통계 데이터

- 데이터 칼럼과 자료형

```
Index(['기준년월', '시도명', '시군구명', '동', '업종대분류코드', '업종중분류코드',
      '업종대분류명', '업종중분류명',
      '기업구분코드', '기업구분', '업력구간코드', '업력구간', '총기업수', '매출총액',
      '매출평균액', '매출중위액',
      '영업이익총액', '영업이익평균액', '영업이익중위액', '총종업원수', '종업원평균수',
      '평균임금_원', '벤처기업수'],
      dtype='object')
```

- 결측치 처리

수치형 변수는 각 변수의 중위값으로 결측값 처리, 범주형 변수는 'unknown'으로 결측값 처리

- 이상치 처리

IQR기준을 이용해서 수치형 변수의 이상치 처리

```
def get_outlier(df=None, column = None, weight=1.5):
    quantile_25 = np.percentile(df[column].values,25)
    quantile_75 = np.percentile(df[column].values,75)
    iqr = quantile_75 - quantile_25
    iqr_weight = iqr * weight

    lowest_val = quantile_25 - iqr_weight
    highest_val = quantile_75 + iqr_weight

    outlier_idx = df[column][ (df[column] < lowest_val) | (df[column] > highest_val) ].index
    return outlier_idx
```

‘기준년월’ 칼럼의 타입을 날짜형 타입으로 변경

```
region_category['기준년월'] = pd.to_datetime(region_category['기준년월'])
```

- 전처리 완료 후 최종 데이터

	기준 년월	시 도 명	시 군 구 명	동	업 종 대 분 류 코 드	업 종 대 분 류 명	업 종 대 분 류 명	기 업 구 분 코 드	기 업 구 분	...	매출총액	매출평균액	매출증위액	영업이익 총액	영업이익 평균액	영업이익 증위액	총 종 업 원 수	종 업 원 평 균 수	평균임금_원	벤 치 기 업 수
1	2018-12-01	강원	강릉시	교1동	A	A01	농업, 임업 및 어업	99	미분류	...	2,582,448.5	1,732,921.0	1,605,320.5	66,407.0	49,669.0	46,024.0	11.0	7.0	30,185,328.5	1.0
2	2018-12-01	강원	강릉시	교1동	A	A01	농업, 임업 및 어업	99	미분류	...	2,582,448.5	1,732,921.0	1,605,320.5	66,407.0	49,669.0	46,024.0	11.0	7.0	30,185,328.5	1.0
3	2018-12-01	강원	강릉시	교1동	A	A02	농업, 임업 및 어업	4	스기업	...	2,582,448.5	1,732,921.0	1,605,320.5	66,407.0	49,669.0	46,024.0	11.0	7.0	30,185,328.5	1.0

2-2.주가 데이터

- 데이터 칼럼과 자료형

```
Index(['시가', '고가', '저가', '종가', '거래량', '거래대금', '등락률'], dtype='object')
```

#	Column	Non-Null Count	Dtype
0	시가	496 non-null	int32
1	고가	496 non-null	int32
2	저가	496 non-null	int32
3	종가	496 non-null	int32
4	거래량	496 non-null	int32
5	거래대금	496 non-null	int64
6	등락률	496 non-null	float32

dtypes: float32(1), int32(5), int64(1)
memory usage: 19.4 KB

- 결측치 처리

결측치 X

- 이상치 처리

이상치 처리가 필요 X

- 전처리 완료 후 최종 데이터

	시가	고가	저가	종가	거래량	거래대금	등락률
날짜							
2020-01-02	38600	38800	37400	37450	81976	3102019350	-2.98
2020-01-03	37800	38050	37350	37600	87649	3295520750	0.40
2020-01-06	36800	37550	36550	37150	66830	2477696250	-1.20

2-3.재무 데이터

- 데이터 칼럼과 자료형

Index(['Statement of comprehensive income(Unit: KWR)', 'I.영업수익', '1.신용판매수익',
 '(1)가맹점수수료', '(2)신판이자', '(3)연회비', '(4)기타', '2.금융상품수익', '3.할부금융수익',
 '4.리스수익', '(1)금융리스수익', '(2)운용리스수익', '(3)기타리스수익', '5.유가증권평가및처분이익',
 '6.파생상품평가및거래이익', '7.외화평가및거래이익', '8.기타수익', 'III.영업비용', '1.금융비용', '2.판매비',
 '3.일반관리비', '4.유가증권평가및처분손실', '5.파생상품평가및거래손실', '6.외화평가및거래손실', '7.대손상각비',
 '8.기타비용', 'III.영업이익', 'IV.영업외수익', '1.관계기업투자주식평가이익',
 '2.유·무형자산처분이익및손상차손환입', '3.기타영업외수익', 'V.영업외비용', '1.관계기업투자주식처분손실',
 '2.유·무형자산처분손실및손상차손', '3.기부금', '4.기타영업외비용', 'VI.법인세비용차감전순이익',
 'VII.법인세비용', 'VIII.당기순이익', 'IX.기타포괄손익', '1.후속적으로당기손익으로재분류될수있는항목',
 '(1)현금흐름위험회피파생상품평가손익', '(2)당기손익으로재분류될수있는항목의법인세',
 '2.후속적으로당기손익으로재분류될수없는항목', '(1)공정가치측정_기타포괄손익금융자산평가손익',
 '(2)관계기업의기타포괄손익에대한지분', '(3)확정급여제도의재측정요소', '(4)당기손익으로재분류될수없는항목의법인세',
 'X.당기총포괄손익', 'XI.주당이익', '1.기본주당이익', '2.희석주당이익'],

- 결측치 처리

결측치는 중위값으로 대체

- 전처리 완료 후 최종 데이터

	Statement of comprehensive income(Unit: KWR)	I.영업수익	1.신용판매수익	(1)가맹점수수료	(2)신판이자	(3)연회비	
0	2010	3607163509929	1538503900261	8.904105e+11	6.211558e+11	2.380164e+11	6.4
1	2011	3204205551561	1806954019397	8.904105e+11	6.211558e+11	2.380164e+11	6.4
2	2012	3817669563919	1881303141752	8.904105e+11	6.211558e+11	2.380164e+11	6.4

3 rows × 51 columns

2-4. 사업보고서

사업보고서

(제 40 기)

사업연도 2022년 01월 01일 부터
2022년 12월 31일 까지

금융위원회
한국거래소 귀중

2023년 3월 8일

제출대상법인 유형 :
면제사유발생 :

주권상장법인
해당사항 없음

회 사 명 :
대 표 이 사 :
본 점 소 재 지 :

삼성카드 주식회사
김 대 환
서울특별시 중구 세종대로 67
(전 화) 1588-8700

(홈페이지) <http://www.samsungcard.com>

작 성 책 임 자 :

(직 책) 경영지원실장 (성 명) 김 상 규
(전 화) 02-2172-7342

2-5. 네이버 뉴스기사 본문

	title	link	pubdate	content
0	'자산 4배 성장'시킨 정태영 현대카드 부회장, 현대캐피탈사 특별공로금 70억...	https://n.news.naver.com/mnews/article/014/000...	Thu, 06 Apr 2023 17:03:00 +0900	WnWn지난 2003년 현대캐피탈 대표이사로 취임해 18년간 현대캐피탈 이끌어 Wn...
1	삼성카드, 순이익 13% 증가...영업이익 신안 재쳤다	https://n.news.naver.com/mnews/article/374/000...	Thu, 09 Feb 2023 11:21:00 +0900	WnWnWnWnWnWn삼성카드가 지난해 당기순이익 6천223억원을 달성했습니다. 영...
2	1등 카드 신한 따라잡은 삼성카드, 한 뼘잡나[머니워니]	https://n.news.naver.com/mnews/article/016/000...	Fri, 10 Feb 2023 10:18:00 +0900	WnWn영업이익으로 신한카드 제치[에월드경제=홍승희 기자] 삼성카드가 업계 1위 신한...
3	빠르게 움직인 삼성카드 '나홀로 호심작'	https://n.news.naver.com/mnews/article/648/000...	Mon, 13 Feb 2023 11:07:00 +0900	WnWn[워치전망대]지난해 순이익 6223억원...전년비 12.9%↑ "조달시장 선제 대...
4	현대캐피탈 자산 4배 성장 주역... 정태영, 특별공로금 70억 받았다	https://n.news.naver.com/mnews/article/014/000...	Thu, 06 Apr 2023 18:08:00 +0900	WnWn상여금 포함 보수 74억7300만원 WnWnWnWn 정태영 현대카드 부회장(...
5	삼성카드, 3분기 영업이익 1940억원...0.3% 증가에 그쳐	https://n.news.naver.com/mnews/article/374/000...	Wed, 26 Oct 2022 10:58:00 +0900	WnWnWnWnWnWn삼성카드의 올해 3분기 영업이익 증가율이 지난해에 비해 크게 ...
6	삼성카드, 3분기 순익 1405억 원...전년비 0.8% 증가	https://n.news.naver.com/mnews/article/629/000...	Wed, 26 Oct 2022 10:12:00 +0900	WnWn영업이익 1940억 원...0.3% 늘어WnWnWnWn삼성카드는 올해 3분기 당...

III. 분석

1. 기업의 재무 현황 파악

• 사업의 개요

dart의 사업보고서에서 기업의 사업 부문별 주요 취급 업무를 파악합니다.

영업부문	취급업무	주요 내용
카드사업 부문	개인신판	개인고객을 대상으로 한 할부 및 일시불 신용제공 서비스
	법인신판	법인고객을 대상으로 한 신용제공 서비스
	카드대출	단기카드대출, 장기카드대출 등 신용카드회원에 대한 장·단기 자금대여 서비스
할부리스사업 부문	할부사업	전자, 내구재, 신차 및 중고차에 대한 할부제공 서비스
	리스사업	설비, 오토, 플릿리스 및 렌트제공 서비스
	일반대출	고객을 대상으로 한 일반 자금대여 서비스
	팩 토 링	기업으로부터 매출채권 등을 매입하고 매입대금을 지불하는 상품
기타 부문	생활편의서비스	보험대리판매, 온라인 쇼핑물 및 여행알선 등의 서비스 제공

• 영업이익 분야

(단위 : 천원, %)

구 분	제40기		제39기		제38기	
	금 액	구성비	금 액	구성비	금 액	구성비
1. 신용판매수익	2,140,670,778	56.42	1,914,466,807	53.36	1,744,083,977	51.80
2. 금 융 수 익	1,008,587,714	26.58	988,120,074	27.54	935,540,217	27.79
3. 할부금융수익	14,385,992	0.38	14,193,473	0.40	20,704,882	0.61
4. 리 스 수 익	223,787,134	5.90	259,685,318	7.24	278,937,737	8.28
5. 유가증권평가 및 처분이익	-	-	7,764,252	0.22	165,352	0.00
6. 파생상품평가 및 거래이익	144,843,633	3.82	156,804,174	4.37	11,210,833	0.33
7. 외화평가 및 거래이익	32,208,914	0.85	20,408,485	0.56	139,290,940	4.14
8. 기타수익	230,087,697	6.05	226,609,227	6.31	237,127,694	7.05
합 계	3,794,571,862	100.00	3,588,051,810	100.00	3,367,061,632	100.00

→ 삼성카드의 경우 신용판매수익과 금융수익이 영업이익의 80%가량을 차지하는 걸 확인할 수 있습니다. 이렇게 기업의 영업이익의 구성요소를 파악합니다.

• 부문손익에 관한 정보

(당기)

(단위:천원)

구 분	카드사업	할부리스사업	기 타	합 계	조정사항	공시재무제표
영업수익	3,328,963,665	247,586,843	218,110,141	3,794,660,649	(88,787)	3,794,571,862
영업비용	2,012,339,223	226,642,357	140,928,259	2,379,909,839	123,426,832	2,503,336,671
대손전 영업이익	1,316,624,442	20,944,486	77,181,882	1,414,750,810	(123,515,619)	1,291,235,191
대손상각비	442,145,475	157,516	(10)	442,302,981	(14,600)	442,288,381
대손후 영업이익	874,478,967	20,786,970	77,181,892	972,447,829	(123,501,019)	848,946,810
전시간접경비	117,063,492	2,160,592	6,899,366	126,123,450	(126,123,450)	-
영업외손익	-	-	-	-	1,807,598	1,807,598
법인세차감전순이익	757,415,475	18,626,378	70,282,526	846,324,379	4,430,029	850,754,408

→ 부문손익에 관한 정보를 통해서 영업수익에 많은 비율을 차지하는 부문을 확인합니다. 여기서 카드사업이 많은 비율을 차지합니다.

• 취급 업무 및 상품서비스 개요

기업의 취급 업무 및 상품서비스를 확인합니다.

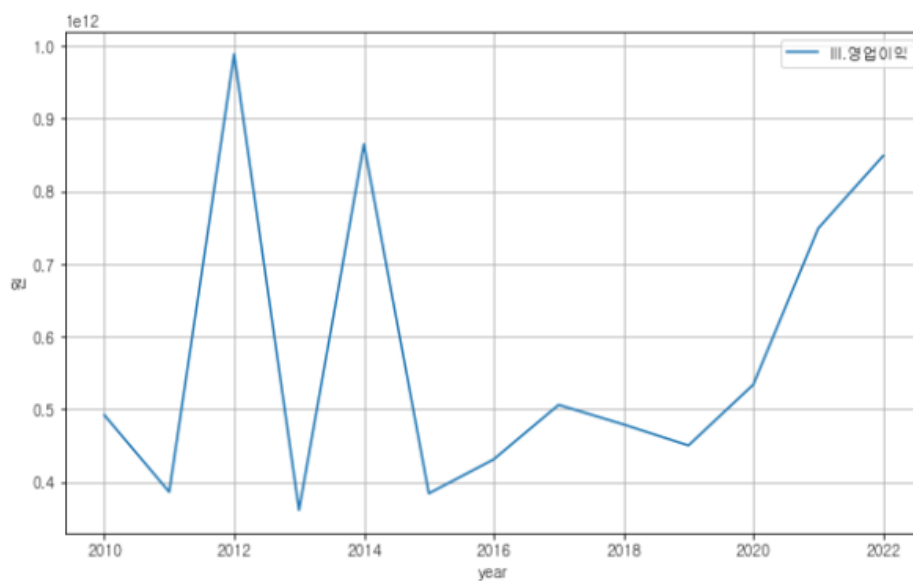
삼성카드의 개요는 다음과 같습니다.

카드사업 : 신용카드, 체크카드, 선불카드, 단기카드 대출, 장기카드대출

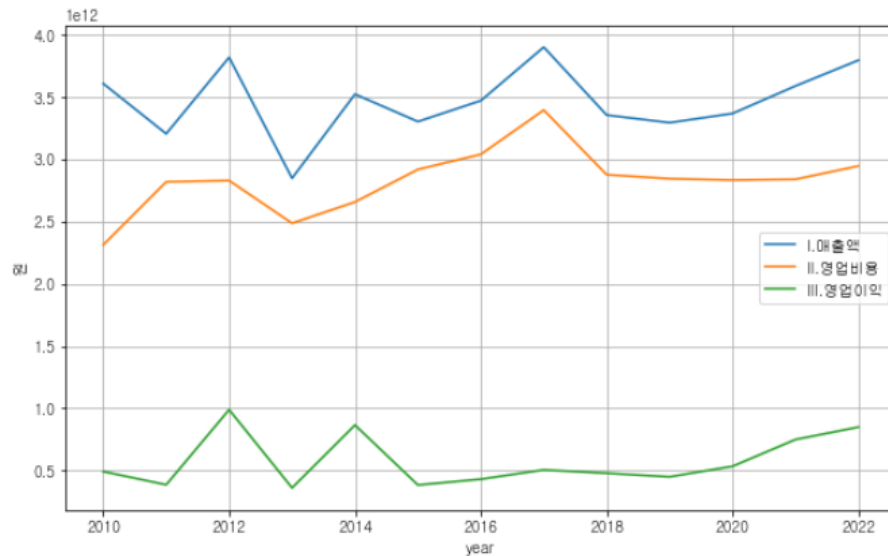
할리스사업 : 할부금융(자동차할부, 일반할부), 일반대출, 팩토링, 설비리스, 설비렌탈, 오토리스, 장기렌터카

기타 부문: 보험대리판매, SMS알림상품, 온라인쇼핑몰, 여행알선, 안심쇼핑플러스

• 영업이익 추세



• 매출액, 영업이익, 영업비용 추세



-매출액, 영업비용, 영업이익 모두 비슷한 추세를 보입니다.

-전반적으로 비슷하게 움직이지만 매출액의 증감폭이 더욱 큼니다.

-2017년에는 매출액과 영업비용은 증감했다 감소하지만 영업이익에는 변화가 없는걸 확인할 수 있습니다. 그러므로 이 시기에는 매출원가와 판매관리비가 증가했을 것으로 확인할 수 있습니다.

• 기업의 손익계산서 항목

손익계산서	
	매출
-	매출원가
	매출총이익
-	판매관리비
	영업이익
+	영업외수익
-	영업외비용
	당기순이익

과 목
매출액
매출원가
매출총이익
영업비용
급여
임차료
영업이익
금융수익
이자수익
금융비용
이자비용
법인세차감전순이익
법인세비용
당기순이익
기타포괄손익
매도가능금융자산평가이익
총포괄이익
주당순이익

☀기업의 매출이 늘었는데 영업비용이 늘어나지 않은 상황에서 무슨 일이 있었을까요?

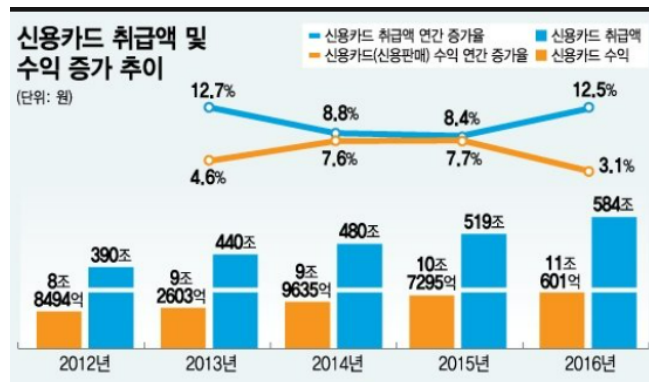
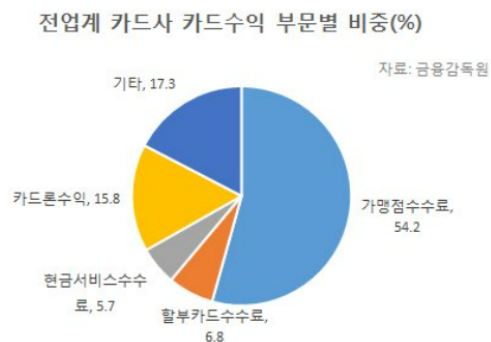


카드사의 경우 매출원가는 큰 영향이 없고 **판매관리비가 증가**해서 매출액은 증가하고 영업이익이 감소하게 됩니다. 해당 문제에서는 이전 시기에 판촉비용을 많이 늘려서 매출액은 늘어났지만 판촉이 효과를 내지 못해서 영업이익이 증가하지 않은 상황을 의미합니다.

⇒ 기업의 **비용구조**를 초점에 두고 분석을 진행해야 합니다.



이외에도 영업소득 분석, 수익구조 분석, 성장 추세, 총 이익률 분석을 진행합니다.

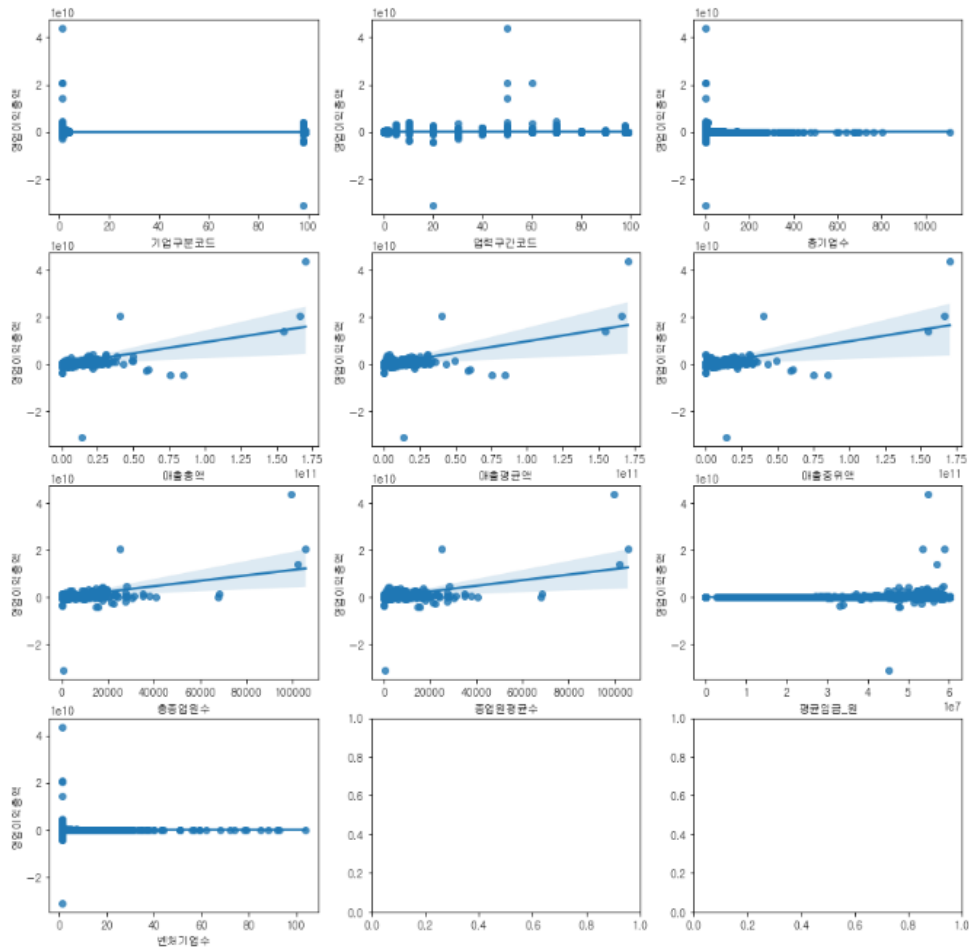


출처 : 머니투데이

2. 영업이익 관계 변수 파악

1) 다른 변수와 상관관계 분석

- 영업이익과 다양한 변수(매출액, 인건비, 광고비 등) 간의 관계를 파악합니다.
- 해당 결과를 바탕으로 어떠한 변수가 영업이익에 큰 영향을 줄지 가정을 수립합니다.



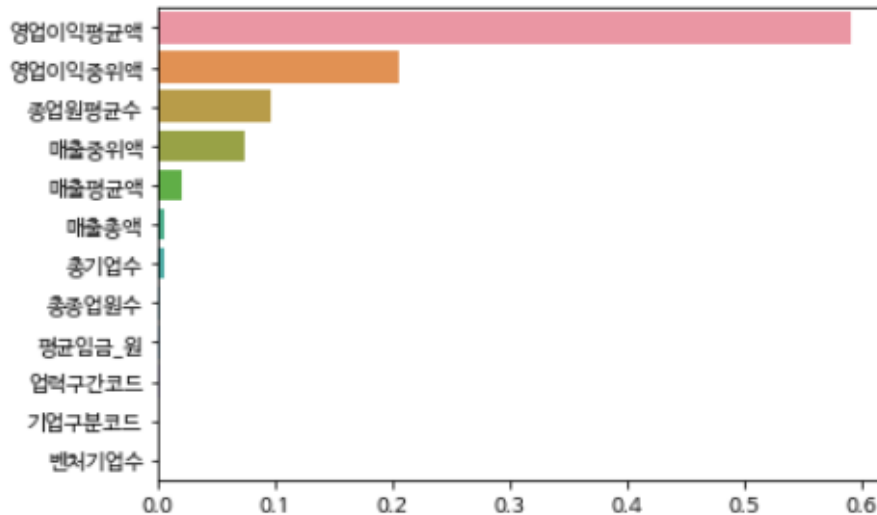
→ 매출총액, 매출평균액, 매출중위액, 총종업원수와 약한 선형관계를 보입니다.

2) 회귀분석을 통한 영업이익의 예측 모델 제작

- tree 계열 모델을 사용해서 영업이익의 예측모델을 제작합니다.

```
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor(random_state=0, n_estimators=100)
rf.fit(X_train, y_train)
y_preds = rf.predict(X_test)
mse = mean_squared_error(y_test, y_preds)
rmse = np.sqrt(mse)
```

- Feature Importance를 통해서 실제 영업이익에 주요한 영향을 주는 변수를 파악합니다.



영업이익 관련 변수를 제외하면 ‘종업원평균수’, ‘매출액’ 관련 변수가 영업이익에 중요한 영향을 주는 변수라는 것을 확인할 수 있습니다.

3.비용 구조 분석

1)절대적 원인 분석

[분석 목표]

- 해당 기업의 내부 구조 분석을 통해서 특정 시기에 비용 구조의 변화가 존재했는지 파악합니다.

[사용 데이터]

- 과거 5년간 해당기업의 영업이익 감소율, 영업이익률 감소율
- 매출 상품 별 인건비 원자재 판관비 금융비용 등의 비중과 증가율

[분석 방법론]

- 각 데이터의 시계열 그래프를 통해서 추세를 분석합니다.

[결과 해석]

- 영업이익/영업이익률의 감소 시점을 통해 문제가 가속화된 시기를 확정할 수 있습니다.
- 인건비, 원자재 비용 등 비용이 특정 시기에 증가했다면 해당 분야의 비용관리 실패가 영업이익 감소의 원인이 되었다는 사실을 파악할 수 있습니다.

2) 상대적 원인분석

[필요 데이터]

- peer그룹의 영업이익 감소율, 영업이익 감소율, 인건비 원자재 판매비 금융비용등의 비중과 증가율
- 거시경제지표 : 금융기관유동성, 대출금리, 콜금리, 국채 수익률, 회사채 수익률, 협의 통화 등

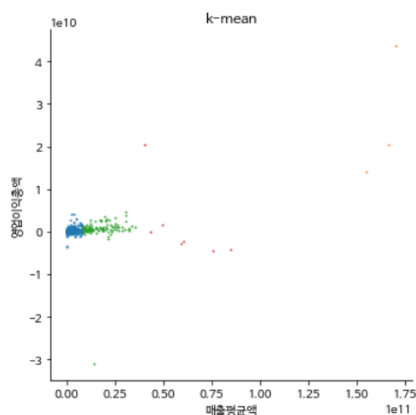
[분석 방법론]

- 특정 시점 별 해당 기업과 peer그룹의 **비용구조 값의 그래프**를 통해서 비교 분석합니다.
- 영업이익과 다양한 **거시경제 지표간의 관계**를 파악합니다.

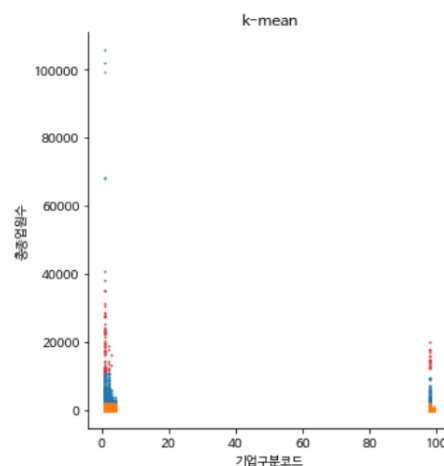


- 클러스터링을 통해서 영업이익과 관련된 특성들을 기준으로 기업끼리 군집화하여 분석합니다.

영업이익총액과 매출평균액 기준



총종업원수와 기업구분코드 기준



[결과 해석]

- peer그룹 대비 해당 기업에서만 비용 증가 및 이익률 감소가 발생한 경우 해당 **기업의 경영 문제**가 영업이익 감소의 원인이 됩니다.
- peer그룹 대비 해당기업에서 비용 증가 및 이익률 감소가 크게 발생한 경우 **경제 사이클에 영향**을 받았고 해당기업의 **위험관리 실패**가 영업이익 감소의 원인이 됩니다.

- 이를 통해서 영업이익 감소의 원인이 기업 내부에 있는지 거시적 경제 사이클에 영향을 받은 것인지 파악할 수 있습니다.
- 어떠한 **거시경제지표** 변수가 영업이익에 큰 영향을 줄지 가정을 수립합니다.

분석 확장 PROCESS

- 클러스터링 기법을 이용해서 동종 산업군에서 ‘영업이익’이 비슷한 기업끼리 군집화합니다.
- 군집별로 각각의 특성을 파악합니다.
- 해당 기업이 속하는 군집에서의 대응 전략을 수립합니다.
- 군집화 방법으로는 단일결합법, 완전결합법, k means등을 활용할 수 있습니다.

4. 키워드 분석을 통한 기업의 동향 파악

- 네이버 뉴스 오픈 API를 활용해서 사업체명을 검색했을 나오는 뉴스의 본문 크롤링(검색 키워드 : 기업명 + 영업이익)했습니다.

```
## 네이버 뉴스 link가 있는 뉴스만 검색하기
titles = []
links = []
pubdates = []
contents = []
article_ids = ['newsct_article', 'articeBody']

p = re.compile('https://n.news.naver.com/.+')
for i, item in enumerate(search_results['items']):
    if p.match(item['link']): ## <link>태그의 문자열이 n.news.naver.com/으로 시작하는 결과만 추출
        title = sub_html_special_char(remove_tag(item['title']))
        link = item['link']
        pubdate = item['pubDate']
        titles.append(title)
        links.append(link)
        pubdates.append(pubdate)

html = urllib.request.urlopen(link)
bs_obj = BeautifulSoup(html, 'html.parser')
for article_id in article_ids:
    print(article_id)
    content = bs_obj.find_all('div', {'id':article_id})
    if len(content) > 0:
        contents.append(content[0].text)
        break
    else:
        continue
#하나의 사이트 접속할때마다 5초의 시간, 연속적으로 해서 block당하는 걸 막기위해서 시차를 줘야함
time.sleep(5)
```

- 기사 본문에서 빈도수가 높은 상위 5개의 키워드 파악했습니다.

1)기사 본문을 하나의 리스트로 제작

```
content = []
for i in news['content']:
    content.append(i)
content
```

2)Tokenization : 문자열을 단어 형태로 분리

```
str = ' '.join(s for s in content)
pos_tag = okt.pos(str)
tokens = okt.morphs(str)
tokens
```

3)stopword 제거

```
with open('stopwords.txt', 'r', encoding='utf-8') as f:
    stopwords = f.readlines()
stopwords=[x.replace('\n','') for x in stopwords] #문단 나누기 제거
result = [x for x in tokens if x not in stopwords]
```

4)특수문자 제거

```
results = []
p = re.compile(r"^[^0-9A-Za-z]+") # 특수문자 제거
for word in result:
    if p.match(word):
        result.remove(word)
    results.extend(result)
results
```

5)vocabulary set 제작

```
[('카드', 33659), ('지난해', 12681),('영업', 11892), ('대비', 10090),
('캐피탈', 8391), ('이익', 7311), ('1년', 7240), ('기간', 7102), ('올해', 6817),
('부회장', 6644), ('비용', 6494)]
```

- 키워드를 바탕으로 기업의 동향을 파악

삼성카드의 경우 **카드, 영업, 캐피탈, 이익, 부회장, 비용** 등 이 영업이익과 관련된 주요 키워드로 나타났습니다.

5. Task Use Idea

📌 수평적 분석(horizontal analysis) : 과거 데이터와 해당 기간의 데이터를 비교 분석하여 다양한 측면에서 문제점을 드러내는 심층적인 분석론



매출액 ↑ & 영업이익 ↓
추세를 보이는 연도 확인

2. 네이버 뉴스 기사 본문 크롤링



해당 연도의 기업명 + 영업이익에 해당하는 검색 키워드 추출

- 영업이익 추세그래프를 확인한 후 매출액이 오르고 영업이익이 동결해있는 연도를 확인합니다.
- 해당 연도의 기업+영업이익과 관련된 검색 키워드를 뽑아냅니다.

3. 연도별 매출 & 영업이익 단어 사전 제작

연도	TOP 10 빈도수 키워드
2010	카드, 영업, 대출, ...
2011	부회장, 캐피탈, 비용 ...
2012	이익, 구조조정, 불황 ...

연도별 빈도수 상위 10개인 키워로 구성된 데이터셋 생성

4. 검색 키워드 모델 확장



동종 산업군의 다른 기업의 검색 키워드를 합쳐서 검색 키워드 모델 제작 및 확장

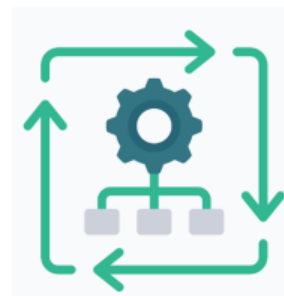
- 공통된 키워드를 파악한 뒤 가장 빈도수가 많이 나온 키워드로 연도별 키워드 단어 사전을 제작합니다.
- 위 프로세스를 동종 산업군의 다른 기업에도 적용하여 단어 사전을 확장합니다. 단어 사전을 제작하는 검색 키워드 모델을 제작합니다.

5. 매출 & 영업이익 분석 단어 사전

연도	TOP10 빈도수 키워드	매출 & 영업이익 패턴
2010	카드, 영업, 대출, ...	매출 상승 영업이익 상승
2011	부회장, 캐피탈, 비용 ...	매출 하락 영업이익 상승
2012	이익, 구조조정, 불황 ...	매출 상승 영업이익 하락

키워드 & 매출액 & 영업이익 패턴 파악

6. 자동화 시스템 구축



검색 키워드 모델 & 단어 사전 구축 자동화

- 어떤 키워드가 나왔을 때 매출액과 영업이익이 변동하는지 패턴을 파악합니다. 앞에서 만들었던 키워드 단어 사전에 패턴 칼럼을 추가 후 매출&영업이익 분석 단어사전을 제작합니다.
- 실시간으로 검색 키워드를 모으도록 자동화 시킨후 데이터베이스로 제작해서 모델의 정확도를 올리며 실제 업무에 적용합니다.

IV. 결과 해석

1. 분석 결과의 의미

- 해당 프로젝트의 목표는 매출액은 증가했지만 영업이익은 동결된 상황을 분석하고 원인을 파악하는 것이었습니다. 실제 삼성카드의 재무 데이터를 활용하여서 접근 방법 구현에 대한 이해를 도왔습니다.
- 기업의 재무 데이터, 사업보고서 등을 활용해서 재무 상태 추세를 파악하였고 매출액과 영업이익 간의 관계를 확인하였습니다.

- 절대적 원인분석과 상대적 원인분석을 통해서 기업의 비용구조 분석을 진행했습니다.
- 이를 바탕으로 실제 업무에 활용할 수 있는 단어사전과 모델을 기획해보았습니다.

2. 결과에 대한 해석 및 활용 방안

- 영업이익과 관련된 변수들의 상관관계 시각화를 통해서 실제 업무를 진행하며 어떤 변수들이 관련이 있는지 대략적으로 파악할 수 있습니다.
- 회귀분석 모델을 활용해서 제작한 예측 모델을 이용해서 중요한 변수들을 파악할 수 있고 새로운 데이터값이 들어왔을 때 영업이익이 얼마가 될 지 예측할 수 있습니다.
- 클러스tring을 통해서 동종 산업군에서 '영업이익'을 기준으로 기업을 군집화 시킬수 있고 해당 군집에서 대응 전략을 수립할 수 있습니다. 경쟁사 분석으로도 이용할 수 있습니다.
- 영업이익 관련 키워드 분석을 통해서 기업의 동향을 파악할 수 있습니다.
- 위에서 사용한 기본 분석들을 바탕으로 **매출&영업이익 분석 단어사전**을 제작해서 실제 업무에 활용할 수 있습니다. 네이버 오픈 API를 이용해서 무료로 데이터를 가져올 수 있으므로 **경제적**으로 사용할 수 있는 단어사전입니다.
- 매출과 영업이익의 관계를 파악할 일이 생겼을 때 네이버에 기업이름+영업이익을 검색한 후 키워드를 뽑아냅니다. 그 후 **매출&영업이익 분석 단어사전**에서 해당 키워드를 찾아서 패턴이 어떤식으로 진행될지 확인할 수 있습니다.

VI. 프로젝트 요약

1. 프로젝트 과정 요약

1. 기업의 재무 현황 파악

- 재무 지표 분석 : 성장 추세, 영업이익 추세, 영업소득 분석 등

2. 영업이익 관계 변수 파악

1) 변수간 상관관계 파악

- 영업이익과 다양한 변수(매출액, 인건비, 광고비 등) 간의 관계를 파악
- 해당 결과를 바탕으로 어떠한 변수가 영업이익에 큰 영향을 줄지 가정 수립

2) 회귀분석을 통한 영업이익 예측 모델 제작

- tree계열 모델 사용

- 변수의 중요도를 파악하여 영업이익에 영향을 미치는 요인을 파악

3. 비용구조 분석

1) 절대적 원인 분석

- 기업 내부 구조 분석을 통해서 진행

2) 상대적 원인 분석

- 국내외 peer그룹과의 비교를 통해서 진행
- 클러스터링 기법을 이용해서 영업이익이 비슷한 기업끼리 군집화하여 비교 분석, 클러스터링 기법을 이용해서 영업이익이 비슷한 기업끼리 군집화하여 비교 분석, 군집별로 각각의 특성을 파악하여 대응 전략을 수립

4. 키워드 분석을 통한 기업의 동향 파악

- 네이버 뉴스 오픈 API를 활용해서 사업체명을 검색했을 나오는 뉴스의 본문 크롤링(검색 키워드 : 기업명 + 영업이익)
- 기사당 빈도수가 높은 상위 10개의 키워드 파악
- 키워드를 바탕으로 기업의 동향을 파악

5. Task Use Idea

- 수평적 분석 이용
- 검색 키워드 모델 제작
- 매출&영업이익 분석 단어사전 제작

2. 프로젝트를 통해 얻은 인사이트, 피드백

1) 새롭게 학습한 내용

- 이번 프로젝트는 기존의 재무 데이터 분석에서 사용하던 **하향식 접근 방식**에서 시작해서 빅데이터 분석의 접근 방식인 **상향식 접근 방식**으로 진행되었습니다.
- 기존의 통계 접근 방식에서 빅데이터 분석 접근 방식을 사용해서 인사이트를 도출할 수 있었습니다.
- 재무재표 구성 요소에 대해서 알아볼 수 있었고 사업보고서 구성요소를 파악해볼 수 있었습니다.
- 유사도 기반 추천시스템의 개념을 알아볼 수 있었습니다.

-

2) 아쉬운 점

기업의 실제 데이터가 없는 상황에서 데이터 분석을 처음 진행하다보니 많은 시행착오를 겪었던 것 같습니다.

분석을 진행하기 전에 '삼성카드'의 데이터를 모의 데이터로 만들어 사용하였습니다. 하지만 정보가 없다보니 실제 데이터와 유사한 분포를 갖고 있는지 검증할 수 없었던게 힘들었던 것 같습니다.

3) 다음 분석에서 보완할 것

데이터의 종류와 방법을 고려해보는 것이 중요하다고 느꼈습니다. 실제로 데이터를 수집하고 분석한다면 이번 프로젝트에서 제안한 방식으로 모델링을 개선할 수 있을 것이라고 기대합니다. 이러한 방식으로 기업의 데이터 분석 역량을 지속적으로 강화할 것입니다.

3. 사용한 도구 및 라이브러리, 참고자료 모음

효과적인-재무-보고서를-분석하기-위한-3단계-과정/

6대-재무제표를-작성하는-데-도움이-되는-데이터-분석/

파이썬으로 공시 데이터 불러오기

[추천시스템] 유사도(Similarity) 구하는 방법

4. 전체 과정의 코드

- 전처리 코드



https://drive.google.com/file/d/1HXBdnlqH3klzHR_UhSd-W8su2MDnB4wU/view?usp=sharing

- 분석 코드



https://drive.google.com/file/d/1HXBdnlqH3klzHR_UhSd-W8su2MDnB4wU/view?usp=sharing