

2016년 주요국 데이터 주도 비즈니스 현황 및 비즈니스 유형 비교 분석 보고서

주요국 데이터 주도 비즈니스 현황 및 비즈니스 유형 비교·분석 보고서

2016. 9.

주관기관 : 한국데이터진흥원

수행기관 : 주식회사 비석세스

제 출 문

한국데이터진흥원 원장 귀하

본 보고서를 “주요국 데이터 주도 비즈니스 현황 및 비즈니스 유형
비교·분석”에 관한 보고서로 제출합니다.

2016. 9.

수 행 기 관 : 주식회사 비석세스

총괄 책임자 : 정 현 욱 (비석세스 대표이사)

사업 참여자 : 전 진 주 (비석세스 이사)

지 승 원 (비석세스 국문 기자)

요 약 문

- 데이터 주도 비즈니스 유형 등장의 배경은 데이터마이닝(data mining)에 대한 설명부터 시작. 데이터마이닝은 현실에 존재하는 수많은 정보(information)를 유용한 형태의 정보(data)로 가공하는 절차를 의미
- 중국 알리바바 그룹의 마윈 회장은 '정보기술시대(Information Technology)가 저물고 데이터 기술(Data Technology)에 기반을 둔 시장이 열릴 것'이라 강조
- 오래전부터 데이터의 중요성이 강조되어왔음에도 불구하고 국내 데이터 산업은 선진기술 대비 최대 5년에 가까운 기술 격차를 보이는 등 성장 초기 단계에 머물러 있어 관련 시장과 기술 성장을 위한 적극적인 노력이 불가피함
- 이에 국내 데이터 산업 활성화 및 성장을 위해 해외 주요국의 데이터 주도 비즈니스 산업 현황을 조사하여, 글로벌 데이터산업 시대 도래를 대비한 선제적 대응 및 국내 데이터산업 활성화를 위한 전략적 비즈니스 유형 발굴 필요
- 이를 위해 케임브리지 유형을 참고하여 데이터 주도 비즈니스의 6가지 유형을 데이터 출처와 데이터 추출 방법으로 구분하였음
- 조사는 대표적인 국내기업 310개 및 해외기업 379개를 대상으로 이루어짐. 국내 기업은 2010년을 기준으로 신생기업 260개와 전통기업 50개로, 해외기업은 신생기업 293개와 전통기업 86개로 세분하여 분석
- 국내 기업의 경우는 대표적인 유형 5개(A/B/C/D/E)를 선정하였으며, 해외 기업의 유형으로는 B/D/F/G/H 유형을 선정
- 위와 같은 비교 분석을 통해 무료 활용 데이터 측면, 신규 데이터 사업자 판로 측면, 데이터 산업 인력 측면, 데이터 산업 법적 측면에서 문제점과 해결책을 제시

< 목 차 >

요약	i
제1장 서론	1
제1절 연구의 배경 및 필요성	1
1. 데이터 주도 비즈니스에 대한 전망	1
2. 국내 데이터 주도 비즈니스 연구 필요성	2
제2절 연구의 목적과 범위	2
제2장 데이터 주도 비즈니스 유형 선행 연구	4
제1절 데이터 주도 비즈니스 유형 개념	4
1. 데이터 주도 비즈니스 개념	4
2. 데이터 주도 비즈니스 유형 개념 및 프레임워크	4
3. 케임브리지 데이터 주도 비즈니스 유형 프레임워크와 분류기준	6
제2절 데이터 주도 비즈니스 분석 방법론	11
1. 데이터 주도 비즈니스 유형 분석방법	12
2. 데이터 주도 비즈니스 유형 분류기준 변형	13
3. 데이터 주도 비즈니스 유형 분석	16
제3장 국내외 데이터 주도 비즈니스 유형 분석	17
제1절 국내 데이터 기업 현황 분석	17
1. 종합 분석	17
2. 유형별 분석	19
3. 설립시기에 따른 비교 분석	22

제2절 해외 데이터 기업 현황 분석	28
1. 종합 분석	28
2. 유형별 분석	30
3. 설립시기에 따른 비교 분석	33
제3절 국내외 데이터기업 비교 분석	39
1. 종합 분석	39
2. 국내기업 및 해외기업 분류기준별 비교분석	39
 제4장 데이터 주도 비즈니스 활성화 방안	 46
제1절 국내 데이터 주도 비즈니스 문제점	46
제2절 정책 제언	48
 참고문헌	 50

〈 표 목 차 〉

〈표 2-1〉 유럽 디지털 포럼 비즈니스 유형.....	4
〈표 3-1〉 국내기업 종합 분석.....	17
〈표 3-2〉 조사대상 국내기업 리스트.....	18
〈표 3-3〉 국내기업 설립시기에 따른 유형 비중 차이.....	22
〈표 3-4〉 국내기업 설립시기에 따른 산업 비중 차이.....	23
〈표 3-5〉 국내기업 설립시기에 따른 데이터원천별 비중 차이.....	23
〈표 3-6〉 국내기업 설립시기에 따른 데이터추출방법별 비중 차이.....	24
〈표 3-7〉 국내기업 설립시기에 따른 공급형태별 비중 차이.....	25
〈표 3-8〉 국내기업 설립시기에 따른 주요고객별 비중 차이.....	26
〈표 3-9〉 국내기업 설립시기에 따른 수익창출별 비중 차이.....	27
〈표 3-10〉 해외기업 사업유형.....	28
〈표 3-11〉 해외기업 분류기준별 사업유형 현황 종합.....	29
〈표 3-12〉 해외기업 설립시기에 따른 산업 비중 차이.....	34
〈표 3-13〉 해외기업 설립시기에 따른 데이터원천별 비중 차이.....	34
〈표 3-14〉 해외기업 설립시기에 따른 데이터추출방법별 비중 차이.....	35
〈표 3-15〉 해외기업 설립시기에 따른 공급형태별 비중 차이.....	36
〈표 3-16〉 해외기업 설립시기에 따른 주요고객별 비중 차이.....	37
〈표 3-17〉 해외기업 설립시기에 따른 수익창출방법별 비중 차이.....	38
〈표 3-18〉 국내기업 및 해외기업 산업별 비중 차이.....	39
〈표 3-19〉 국내기업 및 해외기업 데이터원천별 비중 차이.....	40
〈표 3-20〉 국내기업 및 해외기업 데이터추출방법별 비중 차이.....	41
〈표 3-21〉 국내기업 및 해외기업 공급형태별 비중 차이.....	42
〈표 3-22〉 국내기업 및 해외기업 주요고객별 비중 차이.....	43
〈표 3-23〉 국내기업 및 해외기업 수익창출 방법별 비중 차이.....	44

〈 그림 목차 〉

〈그림 2-1〉 사업유형별 분류기준.....	4
〈그림 2-2〉 사업유형 분류기준과 DDBM 분류기준	6
〈그림 2-3〉 The data-driven business-model framework.....	8
〈그림 2-4〉 DDBM cluster results and respective medoids.....	9
〈그림 2-5〉 DDBMs.....	11
〈그림 2-6〉 시드 데이터 확보 양식.....	12
〈그림 2-7〉 시드 데이터 필터링.....	13
〈그림 2-8〉 변형된 DDBM 분류기준.....	13
〈그림 2-9〉 분류기준과 세부기준	15
〈그림 3-1〉 국내 기업 사업 유형 분포.....	18
〈그림 3-2〉 국내기업 산업별 비중.....	19
〈그림 3-3〉 국내기업 데이터원천별 비중.....	19
〈그림 3-4〉 국내기업 데이터추출방법별 비중.....	20
〈그림 3-5〉 국내기업 공급형태별 비중.....	20
〈그림 3-6〉 국내기업 주요고객별 비중.....	21
〈그림 3-7〉 국내기업 수익창출별 비중.....	21
〈그림 3-8〉 국내기업 신생기업 비중.....	22
〈그림 3-9〉 국내 신생기업 유형별 비중.....	22
〈그림 3-10〉 국내 전통기업 유형별 비중.....	22
〈그림 3-11〉 국내기업 설립시기에 따른 산업 비중 차이.....	23
〈그림 3-12〉 국내기업 설립시기에 따른 데이터원천별 비중 차이.....	24
〈그림 3-13〉 국내기업 설립시기에 따른 데이터추출방법별 비중 차이.....	25
〈그림 3-14〉 국내기업 설립시기에 따른 공급형태별 비중 차이.....	26
〈그림 3-15〉 국내기업 설립시기에 따른 주요고객별 비중 차이.....	27
〈그림 3-16〉 국내기업 설립시기에 따른 수익창출방법별 비중 차이.....	27
〈그림 3-17〉 해외 기업 사업 유형 검토.....	29
〈그림 3-18〉 해외기업 산업별 비중.....	30
〈그림 3-19〉 해외기업 데이터원천별 비중.....	30
〈그림 3-20〉 해외기업 데이터추출방법별 비중.....	31
〈그림 3-21〉 해외기업 공급형태별 비중.....	31
〈그림 3-22〉 해외기업 주요고객별 비중.....	32

〈그림 3-23〉 해외기업 수익창출 방법별 비중	32
〈그림 3-24〉 해외기업 국가별 비중	33
〈그림 3-25〉 해외기업 신생기업 비중	33
〈그림 3-26〉 해외기업 설립시기에 따른 산업 비중 차이	34
〈그림 3-27〉 해외기업 설립시기에 따른 데이터원천별 비중 차이	35
〈그림 3-28〉 해외기업 설립시기에 따른 데이터추출방법별 비중 차이	36
〈그림 3-29〉 해외기업 설립시기에 따른 공급형태별 비중 차이	36
〈그림 3-30〉 해외기업 수익창출 방법별 비중 차이	37
〈그림 3-31〉 해외기업 설립시기에 따른 수익창출방법별 비중 차이	38
〈그림 3-32〉 국내외 유형 비교 분석	39
〈그림 3-33〉 국내기업 및 해외기업 산업별 비중 차이	40
〈그림 3-34〉 국내기업 및 해외기업 데이터원천별 비중 차이	41
〈그림 3-35〉 국내기업 및 해외기업 데이터추출방법별 비중 차이	42
〈그림 3-36〉 국내기업 및 해외기업 공급형태별 비중 차이	43
〈그림 3-37〉 국내기업 및 해외기업 주요고객별 사업유형 현황	44
〈그림 3-38〉 국내기업 및 해외기업 수익창출 방법별 비중 차이	45

제1장 서론

제1절 연구의 배경 및 필요성

1. 데이터 주도 비즈니스에 대한 전망

데이터 주도 비즈니스 유형 등장의 배경은 데이터마이닝(data mining)에 대한 설명부터 시작한다. 데이터마이닝은 현실에 존재하는 수많은 정보(information)를 유용한 형태의 정보(data)로 가공하는 절차를 의미한다. 태초의 인류가 농업을 시작하고 1년이라는 태양력(太陽曆)을 사용하게 된 배경을 생각해 본다면 데이터마이닝 그 자체의 기원을 따지는 것은 의미가 없다.

다만 현대 산업에서 이러한 기법을 적용하여 ‘Without data, you’re just another person with an opinion’이라는 유명한 명언을 남겼으며, 여러 국제적인 사기업의 고문을 역임한 에드워즈 데밍(Edwards Deming)의 ‘PDCA사이클’을 상기할 필요가 있다. 이는 일본에서 ‘데밍 사이클’로 널리 알려진 통계적 품질관리 기법으로, 1970년대의 오일쇼크로 인한 세계적인 불황속에서도 도요타(Toyota), 소니(SONY), 혼다(Honda) 등의 기업을 통해 일본의 국가적인 성장을 불러왔으며 경영에 있어 데이터의 중요성을 피력하는 기본적인 사례라 할 수 있다. 뿐만 아니라, 2011년 세계경제포럼(World Economic Forum)에서는 ‘Data is the new oil’¹⁾ 라고 언급되어, 데이터의 현대 및 앞으로의 산업에서 위상이 더 높아졌다는 점은 부인할 수 없다.

또한, 중국의 알리바바 그룹의 마윈 회장이 2015년 ‘정보기술시대(Information Technology)가 저물고 데이터 기술(Data Technology)에 기반을 둔 시장이 열릴 것’²⁾이라고 강조한 바 있듯, 데이터 자체가 여러 산업에서의 수단이 아니라 그 자체로 하나의 산업이 되었음을 알 수 있다.

그리고 2016년의 스위스 다보스 포럼의 주제가 4차 산업혁명 이었다는 점을 눈 여겨 볼 필요가 있다. 4차 산업혁명을 한 마디로 정의하기는 어렵다. 다보스 포럼 회장인 클라우스 슈밥의 저서 제4차 산업혁명에 따르면 그 범위와 속도 및 규모가 쉬이 예상할 수

1) data is the new oil wef, 2011; rotella, 2012

2) <http://ar.alibabagroup.com/2015/letter.html> 알리바바 주주 서신 중

있지 않기 때문이다. 다만 이를 유비쿼터스, 모바일, 인터넷 및 클라우드 컴퓨팅으로 특징 지을 수 있으며, 이는 데이터를 기반으로 한다는 점을 고려하였을 때, 향후 데이터의 활용은 기업과 국가의 성장 주도권 확보에 있어서 의무가 되었다는 점은 부인할 수 없다.

2. 국내 데이터 주도 비즈니스 연구 필요성

오래전부터, 그리고 현재에도 데이터의 중요성이 강조되어왔음에도 불구하고 국내 데이터 산업은 선진기술 대비 최대 5년에 가까운 기술 격차를 보이는 등 성장 초기 단계에 머물러 있으며, 이에 관련 시장과 기술을 성장시키기 위한 적극적인 노력이 불가피하다. 빅데이터, 사물인터넷, 클라우드 컴퓨팅은 각각 연평균 성장률 37.9%, 25%, 27.4%의 가파른 상승이 예상되나, 세계 시장에서 국내 신기술 시장 규모는 빅데이터, 사물 인터넷, 클라우드 컴퓨팅이 각각 1.4%, 0.5%, 0.8%로 미미한 수준³⁾이다.

더욱이 한국의 데이터 산업에 속하는 중소기업은 기존 비즈니스 유형에서 벗어나지 못하는 상황에서 새로운 성장의 기회와 창출한 데이터의 가치를 발굴해야할 필요성이 절실하다. 특히 온라인트오프라인(O2O), 사물인터넷(Internet of Things), 빅데이터(Big Data), 인공지능(Artificial Intelligence), 클라우드(Cloud), 모빌리티(Mobility), 소셜미디어(Social Media)가 유례 없는 속도로 성장하고 있는 시기에 데이터 주도 비즈니스는 위기이자 기회라고 할 수 있다.

이런 시대에 한국 데이터 위주의 산업을 활성화 및 성장시키기 위해 해외 주요국의 데이터 주도 비즈니스 산업 현황을 조사하여, 글로벌 데이터산업 시대의 도래를 대비한 선제적 대응 및 국내 데이터산업 활성화를 위한 전략적 비즈니스 유형을 발굴하고자 한다. 또한, 연구결과를 바탕으로 국내의 방대한 공공데이터 등 활용성이 낮은 부분을 해외의 공공데이터 활용 방안과 비교하고 분석하여 한국의 스타트업이 공공데이터를 활용한 서비스 구축 개선 등을 할 수 있기 위해 필요한 정부차원의 정책 및 역할을 고려하고자 한다.

제2절 연구의 목적과 범위

우선 본 연구는 데이터 기반 산업을 기존 DB산업의 범위에 ICT 환경 변화를 반영하고 데이터 생태계관점에서 ‘데이터 산업’으로 새롭게 정의한다. 기존의 DB산업이 데이터를

3) 한국데이터진흥원 (2015), 「 2015 데이터산업 백서 」

저장, 관리, 서비스하는데 필요한 솔루션과 컨설팅 관련 사업, 데이터가 유통 대상인 사업이었다면, 데이터 산업은 최근 IT이슈가 되는 빅데이터, 사물인터넷, 클라우드 컴퓨팅을 통한 데이터 수집, 분석, 활용 등 데이터 관련 요소들을 추려 DB산업 영역과 융합되는 산업이라고 정의한다.

또한, ‘데이터 주도(Data-Driven)’란 어떠한 행동이 직관이나 개인 경험이 아닌 데이터에 의해 일어나게 되는 과정으로 정의한다. ‘데이터 주도 비즈니스(Data-Driven Business)’를 데이터를 축적하고 정보화하는 것에서 벗어나 데이터를 분석하여 의미를 찾아내고 이를 실행에 연결시키는 산업이라고 정의한다. 이를 바탕으로, 본 연구를 통해 국내 데이터산업과 주요 해외 국가의 데이터 산업을 데이터 비즈니스 유형 비교 분석 틀에 따라 비교하고 분석하여, 각각의 장단점을 밝힌다.

그리고 해외에서는 성공적인 산업 군이나 국내 산업에서는 진입이 없는 특정 산업 군을 도출하여 해당 도출된 산업을 구체적으로 분석하고, 성공적인 사례를 통해 국내 데이터 산업에 어떻게 적용할 것인지를 전문가 및 관계자와의 인터뷰를 통하여 구체적인 방향성을 확보하고자 한다. 이를 통해 현재 데이터 관련 산업에 진입 중인 국내 스타트업에게 데이터를 활용한 새로운 사업 기회를 제시하고자 한다.

두 번째로, 공공 데이터 활용에 대한 집중 분석을 실시하고 해외에서 성공적인 공공 데이터 활용에 대한 성공 사례를 분석한다. 그리고 이와 유사한 국내 데이터는 어떻게 활용되고 있는지를 분석하여 공공 데이터의 정책적 활용 방안에 대한 시사점을 도출하고자 한다. 이를 위해 공공 데이터를 활용한 해외 성공 사례를 밝히고 포커스 인터뷰를 진행하여, 같은 산업계의 한국 스타트업의 공공 데이터 활용의 현실을 진단하기로 한다.

제2장 데이터 주도 비즈니스 유형 선행 연구

제1절 데이터 주도 비즈니스 유형 개념

1. 데이터 주도 비즈니스 개념

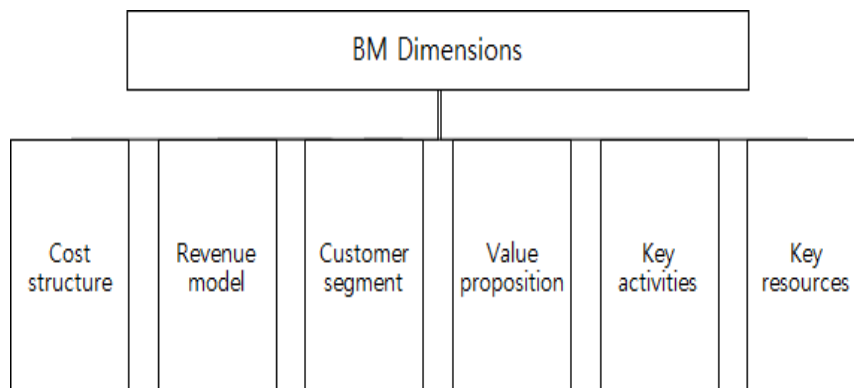
데이터 주도 비즈니스는 데이터를 축적하고 정보화하는 것에서 벗어나 데이터를 분석하여 의미를 찾아내고 이를 실행에 연결시키는 산업이다. 여기서 데이터 주도란 어떠한 행동이 직관이나 개인 경험이 아닌 데이터에 의해 일어나게 되는 과정이다.

데이터 주도 비즈니스는 크게 정부, 민간기업, 대학으로 생태계가 구성되었다. 정부는 정책 및 지원 관련 사업을 담당하며, 민간기업은 실제 데이터 관련 사업을 통해 수익을 창출한다. 대학은 원천 기술 개발 및 인력 교육에 관여하는 역할을 한다.

세 구성원들은 서로가 다양한 형태 및 소스의 데이터를 생성하고 서로로부터의 데이터를 수집 및 저장을 한다. 이는 다시 가공과 분석 등의 과정을 거쳐 최종적으로 활용이 되며, 이러한 과정을 통해 최종 결과물이 소비자에게 제공된다.

서로 상호 연계된 과정의 반복을 통해서 데이터 자체의 규모 및 범위뿐만 아니라, 데이터 주도 비즈니스의 성장 가능성의 규모 및 범위도 점점 증대될 수밖에 없다.

2. 데이터 주도 비즈니스 유형 개념 및 프레임워크



<그림 2-1> 사업유형 분류기준

위의 6가지 분류기준은 현대 경영학에서 활용되는 사업유형을 구분하기 위한 분류기준이다. 이로부터 나아가 데이터 산업의 발전과 함께 데이터 산업 자체에 적용하고 분석하기 위한 유형들이 등장하게 되었는데 대표적으로 케임브리지 유형과 유럽 디지털 포럼 유형이 있다.

유럽 디지털 포럼 유형은 데이터 가치 사슬과 데이터 유형에 따라 비즈니스 유형⁴⁾을 나눴으며 이는 다음의 <표 2-1>로 요약된다.

<표 2-1> 유럽 디지털 포럼 비즈니스 유형

		데이터 가치 사슬		
		데이터 수집	데이터 분석	비즈니스 인텔리전스
데이터 유형	공공 데이터	공공부문의 데이터를 수집, 필터링, 정제	공공부문의 데이터를 분석하여 기업과 공공기관에 유용한 결과를 도출	공공기관을 대상으로 전자정부, 실시간 데이터 관리, 업무상 공공데이터 활용에 대한 컨설팅
	기업 데이터	산재해 있는 데이터소스로부터 기업이 데이터를 목록화하여, 일관성을 유지하고, 사용자 관점의 데이터 세트를 만들 수 있도록 지원	기업을 위해 데이터 분석을 실시하고, 데이터를 가치있는 비즈니스 정보, 비즈니스 통찰로 전환	기업 대상 컨설팅 서비스로, 새로운 비즈니스 기회를 포착하고 비즈니스 절차를 구성해주는 서비스
	개인 데이터	믿을 수 있고 검증된 개인정보 데이터베이스 구축	개인 데이터를 분석하여, 개인 보유 데이터에 대해 이해하도록 지원	개인 데이터를 활용하여 기업이나 공공기관이 데이터에 기반한 의사 결정을 하도록 지원

유럽 디지털 포럼 비즈니스 유형은 데이터 가치사슬과 데이터 유형에 따라 9가지 유형을 제시하여 데이터 가치 사슬과 데이터 유형 구분이 매우 인지적이며, 그 결과 유형에 대한 이해가 용이하다는 장점이 있으나, 실제 비즈니스에서는 데이터를 2가지 이상 이용하는 경우와 데이터 가치 사슬이 2가지 이상 분야에 해당하는 경우가 많다는 단점을 지녔다.

반면 케임브리지 유형은 데이터 주도 비즈니스의 6가지 유형을 데이터 출처와 데이터 추출 방법으로 구분하였고, 이는 스타트업을 대상으로 양적 분석을 한 현실적인 유형이라는 평을 받는다. 다음 내용을 통하여 케임브리지의 6가지 유형을 자세히 살펴본다.

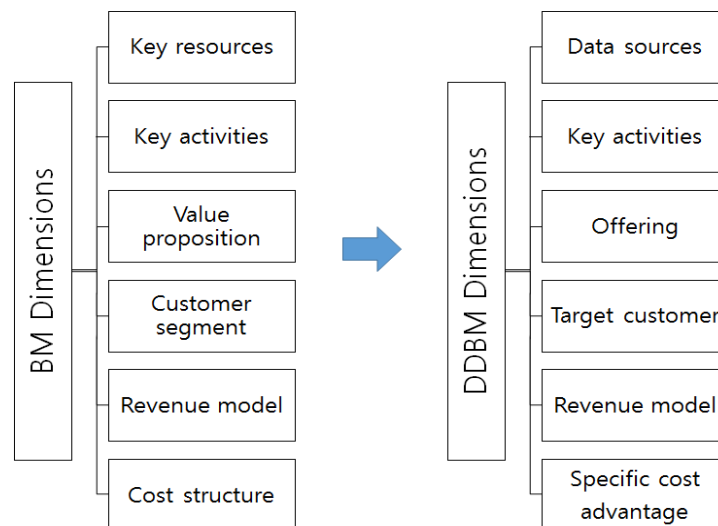
4) 데이터 기술의 시대 - 한국데이터진흥원

3. 케임브리지 데이터 주도 비즈니스 유형 프레임워크와 분류기준

본 연구를 통해 한국의 스타트업이 공공데이터를 활용한 서비스 구축 및 개선 등을 할 수 있는 제도를 마련하고자 한다. 이를 위해 데이터 주도 비즈니스에 대한 고찰이 선행되어야하며 이번 장에서 케임브리지(Cambridge) 대학교의 ‘Big Data for Big Business?, A Taxonomy of Data-driven Business Models used by Start-up Firms’ 라는 연구보고서⁵⁾에서 데이터 주도 비즈니스 유형(Data-driven business models)을 분류한 기준과 분석에 따른 유형들을 살펴본다.

케임브리지 DDBM의 분류체계는 연구보고서에 실린 다음 도표를 참고하여 이해할 수 있다. 이는 2014년 100개의 글로벌 스타트업 회사를 무작위로 추출하여 분석·분류한 연구 결과이다.

아래의 6가지 분류기준은 현대 경영학에서 활용되는 사업유형을 구분하기 위한 분류 기준이며 해당 연구에서는 이를 데이터 산업에 적용하여 하단 <그림 2-2>의 우측과 같이 분류기준을 제시하였다.



<그림 2-2> 사업유형 분류기준과 DDBM 분류기준

5) A Taxonomy of Data-driven Business Models - Cambridge Service Alliance

가. DDBM 분류기준

(1) 데이터 원천(Data sources)

우선적으로, 데이터 출처에 따른 구분이 가능한데, 이는 자체 생산이나 기존의 데이터 등을 활용하는 내부 데이터의 활용 혹은 사용자 입력 데이터, 공공 데이터 등을 활용한 외부 데이터의 활용으로 분류할 수 있다.

(2) 데이터 추출(Key activities)

데이터 추출 방법에 따라서는, 무수히 많은 컴퓨터에 분산 저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술인 크롤링(Crawling), 웹상에서 특정한 주체 데이터의 행동을 추적하는 트래킹(Tracking), 사용자가 직접 생성한 정보를 수집하는 크라우드소싱(Crowdsourcing), 데이터간 통합(Aggregation), 데이터간의 상관관계 등을 통한 분석 및 수집된 정보를 종합적으로 볼 수 있도록 제공하는 기술인 시각화(Visualization) 등이 있다.

(3) 공급(Offering)

제공되는 산출물의 형태에 따라서, 데이터(Data), 정보/지식(Information/knowledge), 데이터나 정보/지식이 간접적으로 활용되어 재화 혹은 용역이 제공되는 형태의 비데이터 형태의 공급(Non-data product/service)으로 분류된다.

(4) 목표 고객(Target customer)

목표 고객에 따른 구분은 일반적인 산업과 동일한, 일반 다수의 소비자를 대상으로 하는 B2C, 기업 고객을 대상으로 하는 B2B 방식이 있다.

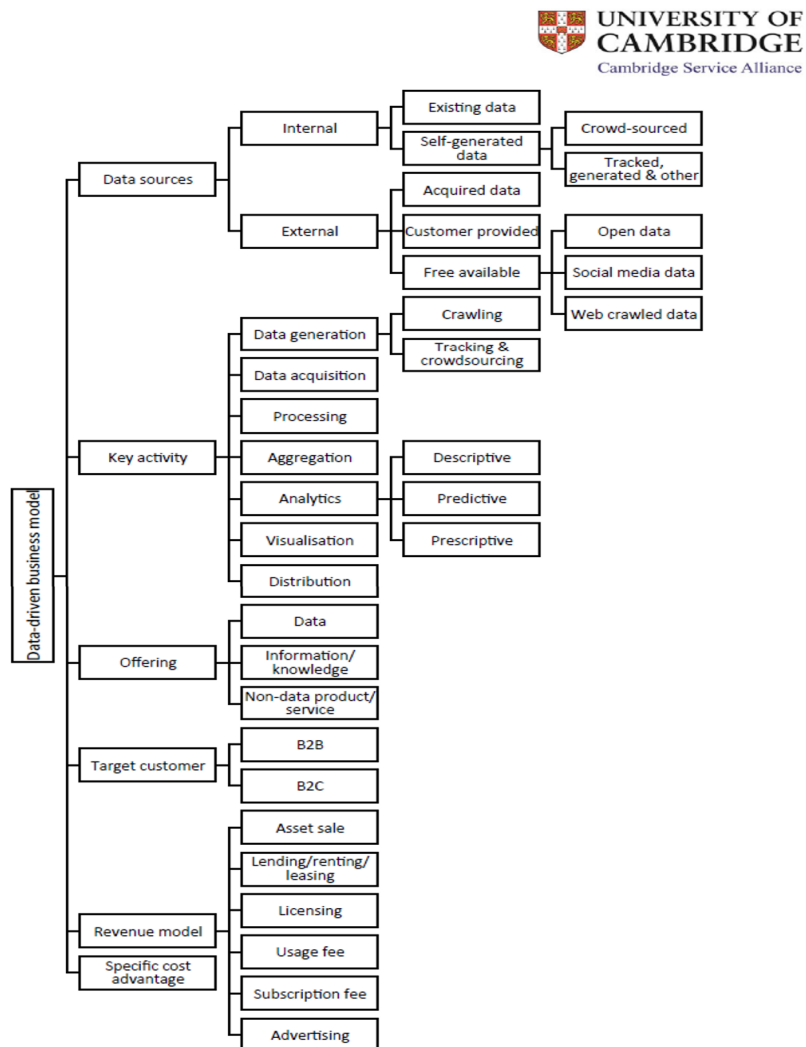
(5) 수익 창출(Revenue model)

수익 창출 방법에 따라서는 등록된 재산권에 대한 대가를 받는 라이선싱(Licensing), 고정 사용료를 수입(Subscription fee), 광고를 통하여 수익(Advertisement)을 창출, 데이터 자체를 판매 혹은 대여하는 등의 방법(Lending/renting/leasing)으로 수익을 창출하는 것으로 나타난다.

(6) 비용측면의 이점(Specific cost advantage)

마지막으로 행동경제학 및 규모의 경제 등으로 인한 특정 비용측면의 이점(Specific cost advantage)을 제시하였는데, 본 연구에서는 이를 변형하여 세분화하여 분석할 계획이다.

이상의 케임브리지의 6가지의 분류기준과 그 세부항목은 다음 <그림 2-3>로 정리할 수 있다.



<그림 2-3> The data-driven business-model framework

나. 케임브리지 DDBM 프레임워크 6가지 유형

해당 보고서에서 표본 기업들에 위의 분류 기준을 적용하여 데이터 원천과 데이터 추출 방법 형태를 조사하고 분석한 결과를 다음과 같이 제시하였다.

DDBM 6가지 유형에 대한 설명을 사례와 함께 제시하면 다음과 같으며 이는 본론의 분석에서도 동일하게 적용되는 기준이다.

DDBM Dimensions				DDBM Cluster Types Percentages					
				A	B	C	D	E	F
				Free data collector and aggregator	Analytics-as-a-service	Data generation and analysis	Free data knowledge discovery	Data-aggregation-as-a-service	Multi-source data mash-up and analysis
Share of companies				18%	29%	17%	15%	6%	15%
Data Source	Internal	Existing data		0%	0%	0%	0%	0%	0%
		Self-generated data	Crowd-sourced	18%	0%	31%	0%	17%	7%
			Tracked, generated and other	6%	11%	88%	0%	0%	21%
		Acquired data		12%	11%	0%	7%	0%	21%
	External	Customer provided		24%	100%	13%	0%	100%	79%
		Free available	Open data	100%	11%	19%	100%	0%	93%
			Social media data	12%	4%	0%	7%	0%	7%
			Web crawled data	71%	11%	13%	50%	0%	57%
	Key Activity	Data generation	Crawling	35%	0%	6%	50%	0%	43%
			Tracking & crowdsourcing	12%	4%	100%	0%	17%	29%
		Data acquisition		24%	21%	6%	21%	17%	29%
		Processing		100%	18%	0%	0%	100%	93%
		Aggregation		12%	82%	69%	86%	0%	93%
		Analytics	Descriptive	0%	50%	13%	21%	0%	14%
			Predictive	0%	11%	6%	7%	0%	7%
			Prescriptive	24%	39%	44%	7%	33%	36%
		Visualisation		100%	36%	19%	29%	83%	50%
		Distribution		35%	0%	6%	50%	0%	43%
Offering	Data			12%	0%	0%	0%	0%	14%
	Information/ knowledge			88%	100%	88%	100%	83%	93%
	Non-data product/service			0%	0%	13%	7%	0%	0%
Target Customer	B2B			71%	96%	63%	86%	83%	86%
	B2C			47%	18%	50%	21%	33%	21%
Revenue Model	Asset Sale			0%	0%	19%	0%	0%	0%
	Lending/renting/leasing			6%	0%	0%	0%	0%	0%
	Licensing			0%	0%	0%	0%	0%	0%
	Usage fee			12%	14%	13%	21%	0%	0%
	Subscription fee			47%	46%	44%	64%	33%	36%
	Advertising			12%	0%	6%	7%	0%	7%

<그림 2-4> DDBM cluster results and respective medoids

(1) 무료 데이터 수집 및 통합 유형(Free data collector and aggregator - Type A)

트위터(Twitter), 페이스북(Facebook), 인스타그램(Instagram) 등 공개된 무료 API를 활용하여, 고객이 원하는 방식으로 제공하는 형식으로 실시간 API 분석을 통한 트렌드 분석 및 조사, 실시간으로 소셜 네트워크의 콘텐츠 데이터를 추출하여 사용자가 재가공할 수 있도록 원소스 제공을 한다. 그립(Gnip)과 같은 서비스가 대표적이며, 본 보고서에서는 이러한 유형을 A유형이라 칭하겠다.

(2) 서비스 분석 유형(Analytics-as-a-service - Type B)

센디파이(sendify)와 같이 콜센터에서 전화가 올 때 해당 전화번호의 주인이 어떤 경력을 가졌고, 어떤 성향이고, 어떤 통화 이력이 있는지 바로 알려주는 서비스를 뜻하는 것으로, 사용자가 생성한 정보를 통해서 사용자에게 기존 데이터를 기반으로 분석 정보를 제공하는 형태이며 이를 B유형이라 칭하겠다.

(3) 데이터 생성 및 분석 유형(Data generation and analysis - Type C)

사용자의 서비스에 특정 코드를 심어서 정보를 수집하는 서비스로, 믹스패널(mixpanel)과 같이 사이트 분석 서비스가 있으며, 사용자의 웹사이트에 믹스패널 코드를 심어서 사용자의 정보를 수집하여 자체 분석된 정보를 사용자에게 제공하는 서비스이다. 이러한 유형의 유형을 C유형이라 칭하겠다.

(4) 무료 데이터 정보 디스커버리 유형(Free data knowledge discovery - Type D)

특정 데이터를 분석하여 관련 정보를 제공하는 서비스로, 길드(guild)와 같이 개발자가 만든 코드를 자동으로 분석해서 해당 개발자의 실력을 평가한 정보를 제공하는 서비스이며 D유형이라 칭하겠다.

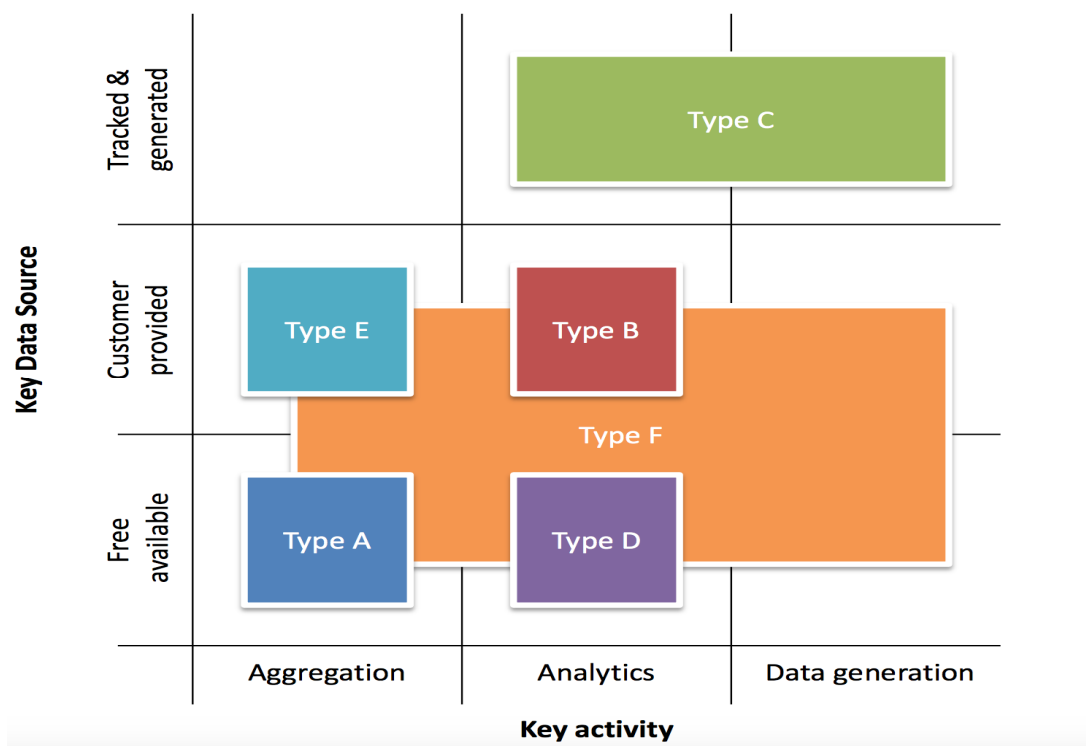
(5) 데이터 통합 유형(Data-aggregation-as-a-service - Type E)

사용자가 내부에서 생성한 다양한 정보를 분석하여 사용자에게 정제된 정보를 제공하는 서비스로 알마(ALMA)와 같이 교내에서 학생들이 생성한 다양한 온라인 정보를 수집 분석하여 학생의 학습 상태, 성장 상태 등 정보를 제공하는 서비스이다. 이와 같은 유형을 E유형이라 칭하겠다.

(6) 멀티 소스 데이터 통합·분석 유형 (Multi-source data mash-up and analysis - Type F)

사용자가 생산한 데이터베이스와 외부의 다양한 무료 데이터를 통합하여 분석 정보를 제공하는 서비스로 위러브로이(welovroi)와 같이, 사용자가 입력한 사이트 정보와 서비스 자체가 다른 공개 데이터를 활용한 정보를 통합하여 회사의 시장 위치 등 관련 정보를 통합적으로 제공해주는 서비스이다.

위에 서술된 유형 6가지를 평면에 나타내 보면 <그림 2-5>와 같다.



<그림 2-5> DDBMs

제2절 데이터 주도 비즈니스 분석 방법론

조사는 대표적인 국내기업 310개 및 해외기업 379개를 대상으로 이루어졌다. 국내기업은 다시 2010년을 기준으로 신생기업 260개와 전통기업 50개로, 해외기업은 신생기업 293개와 전통기업 86개로 세분하여 분석을 하였다. (* 2010년 “스타트업”이라는 개념이 국내에 전파되기 시작한 시기를 기점으로 그 후 설립 기업을 신생 기업으로 정의)

해당 샘플 선정은 데이터 비즈니스 관련 사업체 국내 500개와 해외 500개에서 게임 등 사행성 산업군, 실제 서비스가 작동이 멈추거나 서비스를 변경한 사업체 등을 제외한 국내 기업 310개 및 해외 기업 379개를 선정하였다.

1. 데이터 주도 비즈니스 유형 분석방법

가. 기본적인 시드 데이터 확보 양식

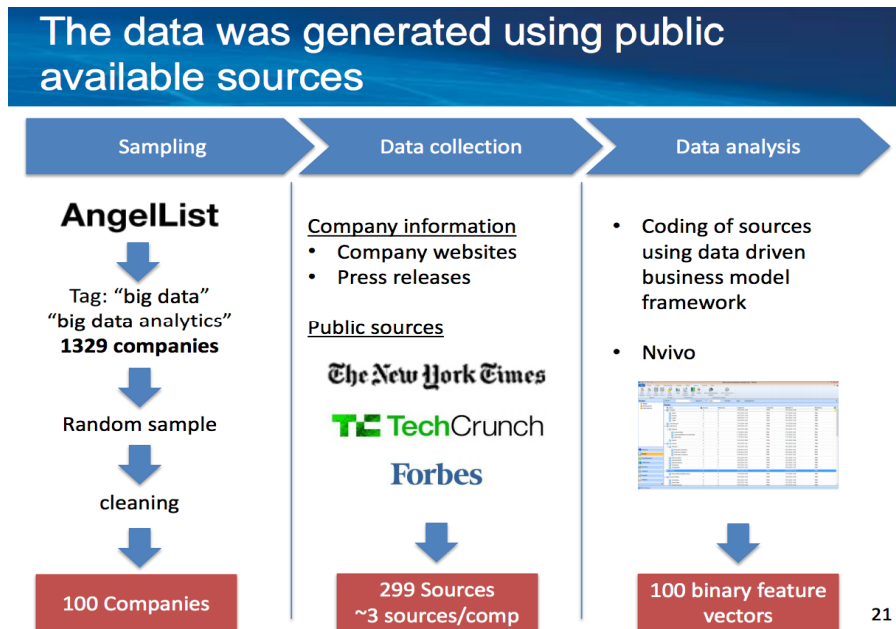
회사명, 사이트 주소, 간략한 회사 소개, 비즈니스 유형, 창업년도, 창업자 정보, 창업자 이메일 주소, 도시, 국가, 총 투자 규모, 투자자, 사업성과 등의 데이터를 다음과 같이 취합하였다.

SNo.	Company Name	Domain Name	Overview	Business Model	Year
1	Ellevest	ellevest.com	Automated investment platform for women.	Stealth	2015
3	WeInvest	weinvest.net	Web based social platform for investment needs	Wealth Management - Retail - Investment Firm	2014
4	GetSquareFeet	getsquarefeet.com	Data-driven investment platform for real estate investors	Private Market Investing - Retail - Real Estate	2014
5	Stockal	stockal.com	Search Engine for stock investors with social media insights	Data Platform - Public Market Investing - Crowdsourced	2014
6	Talkoot	talkoot.in	Social platform for traders and investors	Public Market Investing - Retail - Social Trading Platform - Investor Network	2015
7	Flatons	flatons.in	Real estate investment discovery and investment analytics platform	Private Market Investing - Retail - Real Estate	2015
8	Tavaga	tavaga.com	Social trading platform for users	Stealth	2015
9	PropertyShare	propertyshare.in	Online platform for fractional ownership of residential apartments	Private Market Investing - Retail - Real Estate	2015
10	Wealth Whispers	wealthwhispers.in	Personal financial advisory solution	Wealth Management - Retail - Investment Firm	2015
11	Inves	invesapp.com	Mobile first trading platform	Public Market Investing - Retail - Electronic Trading Platform - Securities - Flat Fee	2015
12	Tauro Wealth	taurowealth.com	Online platform providing investment ideas on thematic investing	Stealth	2015
13	Smallcase	smallcase.com	Platform for thematic investing in public markets	Public Market Investing - Retail - Electronic Trading Platform - Securities - Flat Fee - Thematic	2015
14	TradeWithMe	tradewithme.me	Mentored Trading and Training services portal	Public Market Investing - Retail - Algorithmic Trading Platform	2015
15	Tipbaba	tipbaba.in	Platform to connect to follow top stock performers	Data Platform - Public Market Investing - Analytics	2015
16	Asset Vantage	assetvantage.com	Cloud-based investment management platform	Wealth Management - Enablers - Portfolio Tracking	2013
17	Kuvera	kuvera.in	Online wealth management portal	Stealth	2015

<그림 2-6> 시드 데이터 확보 양식

나. 확보된 시드 데이터 중 성장 중인 기업 선별 후 내부 연구원 개별 조사

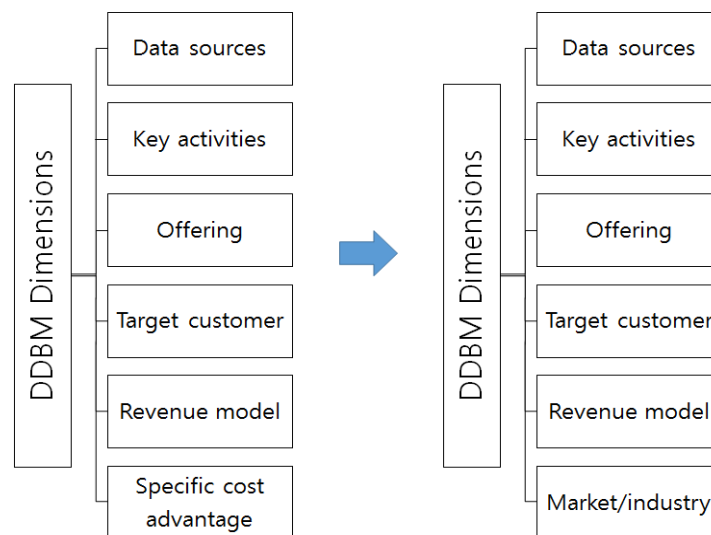
데이터 출처, 데이터 추출 방법, 고객 구분에 따른 방법 등 앞의 데이터 분야 산업 비교 방법에 나열된 기준점으로 세부 분석하였다.



21

<그림 2-7> 시드 데이터 필터링

2. 데이터 주도 비즈니스 유형 분류기준 변형



<그림 2-8> 변형된 DDBM 분류기준

앞서 케임브리지 연구보고서에 나타난 DDBM 분류기준 6가지 항목에서 기존의 특정 비용 측면의 이점을 제외하고, 산업별 분석대상과 그에 대한 주요 세부 서비스 13개 항목을 추가하여 조사 및 분석을 하였으며 13개 항목⁶⁾에 대한 세부 내용은 다음과 같다.

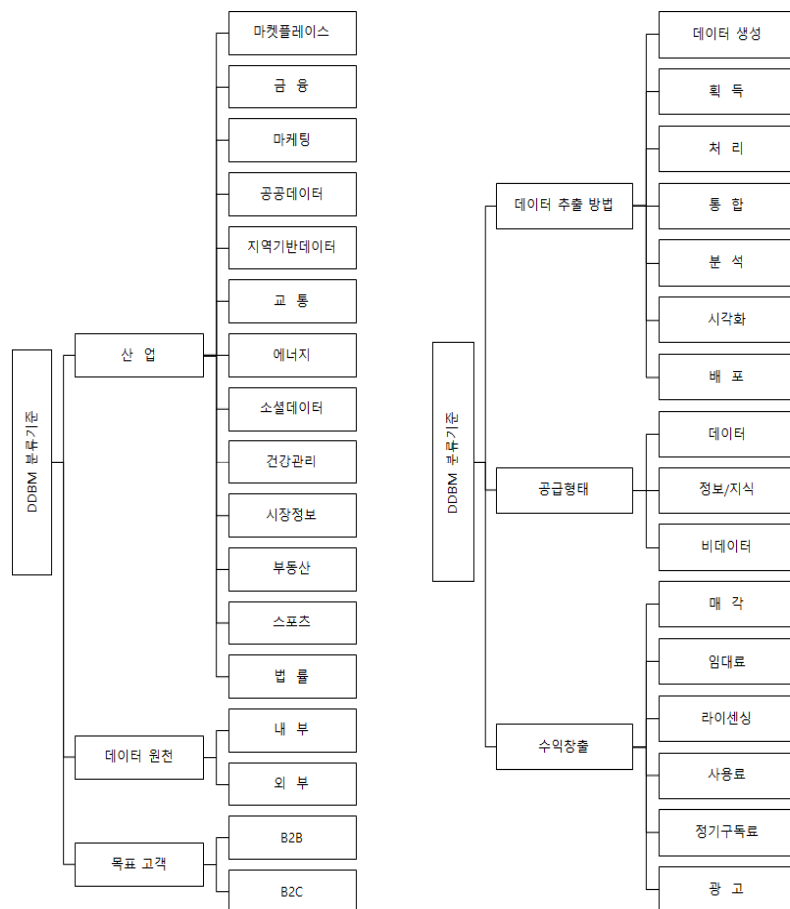
6) 스타트업 리서치 및 데이터 플랫폼 Tracxn 의 “데이터 분야” 대표 산업 분류표

- 마켓플레이스(Marketplace) 산업 분야
- 금융(Finance) 산업 분야
- 마케팅(Marketing) 분야
- 공공데이터(Public data) 분야
- 지역기반데이터(Geospatial) 분야
- 교통데이터 분야
- 에너지데이터 분야
- 소셜데이터 분야
- 건강관리데이터 분야
- 시장정보데이터 분야
- 부동산데이터 분야
- 스포츠데이터 분야
- 법률데이터 분야

- (1) **마켓 플레이스 산업 분야** : 개발자와 기업을 연결해 주는 위시켓 서비스와 같이 사용자 간 가치 창출을 발생시키는 산업 분야
- (2) **금융 산업 분야** : 한국의 모든 금융 상품을 분석해서 사용자에게 최적화된 상품을 추천해주는 핀다 서비스와 같이 금융 분야에서 가치를 발생시키는 산업 분야
- (3) **마케팅 분야** : 소셜 등 온라인에 언급된 문장을 분석해서 기업의 브랜딩 효과를 측정하는 스위즐 랩스와 같이 기업의 마케팅에 분야에서 가치를 발생시키는 산업 분야
- (4) **공공데이터 분야** : 산림청과 국토지리정보원의 공공 데이터를 활용해서 등산 코스를 보여주는 루가 등산 서비스와 같이 공공 기관에서 제공하는 데이터를 활용해서 가치를 발생시키는 산업 분야
- (5) **지역기반 데이터 분야** : 회사 자체적으로 데이터를 수집하는 방식으로 특정 지역 기반으로 서비스를 제공하는 분야, 예를 들어서 데이트팝과 같이 서울 특정 지역의 데이트 코스를 추천하는 서비스
- (6) **교통데이터 분야** : 플러스 서비스와 같은 목적지를 이동하는 사용자의 이동 경로를 파악해서 근거리의 동승자 차량을 연결해주는 서비스로 교통 분야 또는 교통의 핵심인 차량 분야에서 새로운 가치를 창출하고자 하는 서비스 산업
- (7) **에너지 데이터 분야** : 에너지톡 서비스와 같이 실시간으로 사용자에게 댁내의 전기 사용료를 모바일을 통해서 데이터를 제공하는 분야로 전기, 수력, 풍력 등 에너지 분야에 새로운 가치를 창출하고자 하는 서비스 산업
- (8) **소셜 데이터 분야** : 소셜 댓글 서비스인 라이브리와 같이 온라인 소셜 활동 분야에서 사용자에게 편리성을 제공하고, 해당 활동에 기반한 데이터를 활용하는 산업 분야
- (9) **건강 관리 데이터 분야** : 타액을 통한 유전자 분석을 통해 맞춤형 건강관리 방법을 제공하는 제노플랜 서비스와 같이 사람의 건강, 질병 등 생명과 연관된 분야

- (10) **시장 정보 데이터 분야** : 특정 시장의 정보를 활용하는 서비스로 취업 분야의 맞춤형 자기 소개서 샘플을 제공하는 자소서 서비스와 같이 틈새 시장 정보를 제공하는 산업 분야
- (11) **부동산 데이터 분야** : 사용자의 여건에 적합한 빌딩을 추천해 주는 서비스인 애프터 그로스와 같이 부동산 정보를 활용하여 데이터를 제공하는 산업 분야
- (12) **스포츠 데이터 분야** : 사용자의 특성에 맞는 스포츠 종목을 추천해주는 플레이콧 서비스와 같이 스포츠 분야에서 가치를 창출하고자 하는 산업 분야
- (13) **법률데이터 분야** : 사용자의 상황에 맞는 변호사를 연결하고, 유사 사례를 쉽게 검색할 수 있게 하는 로톡과 같은 서비스 분야로, 법률 분야에서 시간, 비용적인 측면에서 사용자에게 가치를 창출하는 산업

본 연구에서 사용된 분류기준과 그 세부기준을 도식화 하면 다음 <그림 2-9>와 같다.



<그림 2-9> 분류기준과 세부기준

3. 데이터 주도 비즈니스 유형 분석

가. 산업 단위로 분석된 정보간 비교

마켓플레이스, 금융, 마케팅, 공공데이터, 지역기반데이터, 교통, 소셜데이터 등 산업 분야별로 국내 기업 분석 정보와 해외 기업 분석 정보를 비교하여, 해외기업 분석 결과 대비 국내 기업 분석 결과가 적은 비즈니스 유형을 발굴한다.

비즈니스 유형별 및 항목별 해외 상황 대비 부족한 기업 산출을 위한 판단 기준은 국내 조사대상 기업 중 각 유형별 항목 비중과 조사대상 해외기업 중 각 유형별 항목 비중의 차이를 비교하였다.

제3장 국내외 데이터 주도 비즈니스 유형 분석

제1절 국내 데이터 기업 현황 분석

1. 종합 분석

앞의 방법론을 적용하여 분석한 결과는 다음과 같다.

<표 3-1> 국내기업 종합 분석

	데이터 배포		데이터 분석		데이터 생성		데이터 시각화		데이터 통합/획득/처리		Total 대상 수	Total %
Row Labels	대상 수	%	대상 수	%	대상 수	%	대상 수	%	대상 수	%		
기존데이터	40	12.90%		0.00%		0.00%		0.00%		0.00%	40	12.90%
무료활용	15	4.84%	9	2.90%	11	3.55%		0.00%	7	2.26%	42	13.55%
이용자제공	96	30.97%	47	15.16%		0.00%	8	2.58%	6	1.94%	157	50.65%
자가생성	38	12.26%	5	1.61%	7	2.26%		0.00%	1	0.32%	51	16.45%
획득데이터	13	4.19%	3	0.97%	2	0.65%		0.00%	2	0.65%	20	6.45%
Grand Total	202	65.16%	64	20.65%	20	6.45%	8	2.58%	16	5.16%	310	100.00%

앞서 서론에서 언급된 케임브리지 연구에 활용된 6개의 범주를 활용하여 분석하였으며, 대표적인 유형 5개를 선정했다.

- A 유형 (이용자 제공 및 데이터 배포) : 96개 기업 (30.97%)
: 위시켓(아웃소싱플랫폼), 스타일쉐어(모바일패션플랫폼), 헤이딜러(내차 판매 견적앱)
- B 유형 (이용자 제공 및 데이터 분석) : 47개 기업 (15.16%)
: 자비스(경영지원 서비스), 와탭(기업 서버 모니터링 서비스), 플라이북(도서 추천 앱)
- C 유형 (기존재 데이터 및 데이터 배포) : 40개 기업 (12.90%)
: 원모먼트(꽃배달), 브레인나우(두뇌디자인 프로그램), XtreamerTV(맞춤형 TV 채널)
- D 유형 (자가 생성 및 데이터 배포) : 38개 기업 (12.26%)
: 차이나탄 (중국어 교육 서비스), 마켓컬리(신선식품 배송), 캠퍼스텐(대학생 잡지)
- E 유형 (무료 활용 및 데이터 배포) : 15개 기업 (4.84%)
: 굿닥(병원 검색 앱), 코인원(비트코인 거래소), 인포케어(법원경매 서비스)

이를 통해서, 국내 데이터 주도 비즈니스의 경우는 데이터를 배포 제공하는 서비스가 65%로 다수를 이루고 있으며, 데이터 원천은 이용자 제공이 45% 이상을 차지함을 알 수 있다. (* “배포” 제공 방법은 원천 데이터를 온라인 화면에 보여주거나 또는 API 형식으로 데이터를 사용자에게 제공하는 방식)

기존재 데이터	Type C				
무료활용	Type E				
이용자 제공	Type A	Type B			
자가 생성	Type D				
획득 데이터					
	데이터 배포	데이터 분석	데이터 생성	데이터 시각화	데이터 통합/획득/처리

〈그림 3-1 국내 기업 사업 유형 분포〉

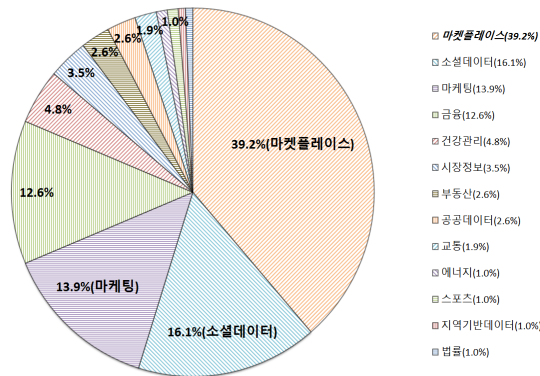
다음 <표 4-3>는 본 연구를 통해 조사한 국내기업 30개를 유형별로 나열한 것이다.

<표 3-2> 조사대상 국내기업 리스트

회사명	산업	데이터 원천	데이터 추출방법	공급	주요 고객	수익구조	설립 시기
미디어웍	마케팅	이용자제공	예측최적화	데이터	B2B	이용료	전통
웍스	공공데이터	공개데이터	크롤링	데이터	B2C	정기이용료	전통
유학닷컴	마켓플레이스	크롤링	예측최적화	정보/지식	B2C	이용료	전통
이데일리	공공데이터	획득데이터	배포	정보/지식	B2C	정기이용료	전통
에듀스파	마켓플레이스	트래킹	트래킹& 클라우드소싱	데이터	B2C	정기이용료	전통
지디에스케이	마케팅	이용자제공	예측최적화	데이터	B2B	정기이용료	전통
인포케어옥션	부동산	공개데이터	배포	데이터	B2B	정기이용료	전통
유웨이	시장정보	이용자제공	예측최적화	데이터	B2C	이용료	전통
중앙교육							
비엔투스	소셜데이터	공개데이터	트래킹& 클라우드소싱	데이터	B2C	광고	신생
LawManager	법률	공개데이터	예측최적화	비데이터	B2C	정기이용료	신생
Goodoc	건강관리	공개데이터	배포	정보/지식	B2C	광고	신생
NIRSIT	건강관리	트래킹	배포	비데이터	B2B	정기이용료	신생
씽크풀	금융	트래킹	배포	데이터	B2C	광고	신생
Brandi	소셜데이터	크롤링	배포	정보/지식	B2C	광고	신생
Meesiq	마켓플레이스	획득데이터	배포	비데이터	B2C	정기이용료	신생
loplat	마케팅	획득데이터	배포	데이터	B2B	정기이용료	신생
CAMSCON	소셜데이터	소셜데이터	배포	정보/지식	B2C	광고	신생
REDTABLE	소셜데이터	소셜데이터	통합	정보/지식	B2C	라이선싱	신생
Anchoreer	시장정보	크롤링	통합	데이터	B2C	광고	신생
AGRISYS	공공데이터	이용자제공	현황	데이터	B2B	라이선싱	신생
Funded	금융	이용자제공	예측	비데이터	B2C	이용료	신생
Gen Port	금융	트래킹	예측	데이터	B2C	라이선싱	신생
inContents	마케팅	이용자제공	배포	데이터	B2B	라이선싱	신생
HELP ME	법률	이용자제공	예측최적화	비데이터	B2C	정기이용료	신생
LOBIG Beta	부동산	공개데이터	현황	데이터	B2B	라이선싱	신생
세틀뱅크	금융	이용자제공	배포	데이터	B2B	이용료	전통
스타뱅크	금융	이용자제공	통합	비데이터	B2B	이용료	전통
진학사	시장정보	이용자제공	배포	데이터	B2C	정기이용료	전통
예람	마켓플레이스	이용자제공	배포	데이터	B2B	정기이용료	전통
올하우	마케팅	이용자제공	시각화	비데이터	B2B	이용료	전통

2. 유형별 분석

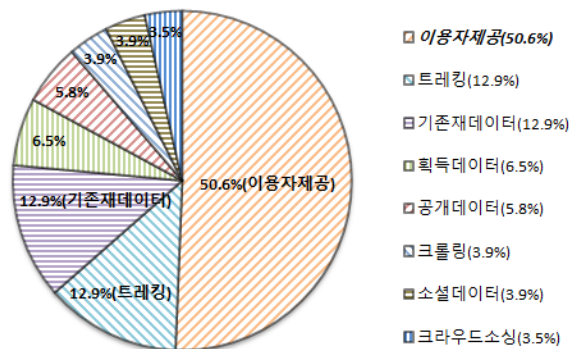
가. 산업별 유형



<그림 3-2> 국내기업 산업별 비중

- 마켓플레이스 분야가 122개사(39%)로 가장 많은 비중을 차지
- 다음으로 소셜데이터 50개사(16.1%), 마케팅이 43개사(13.9%)로 나타남
 - ▶ 마켓 플레이스의 사업 모델은 사용자간을 연결하는 서비스로, 아마존, 우버, 에어비엔비와 같이 급성장하는 기업에 두드러지는 모델로, 정보와 지역의 격차를 해소하고자 하는 증가하는 산업 모델이다.

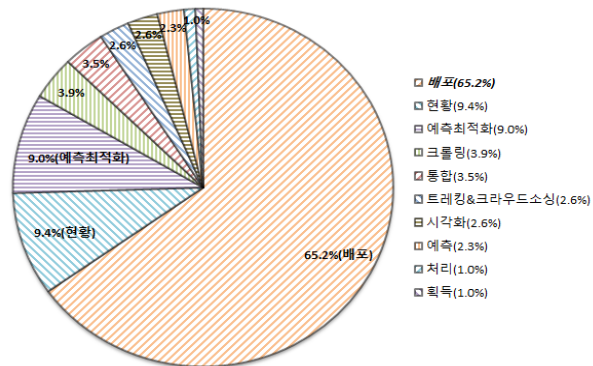
나. 데이터원천별 사업유형



<그림 3-3> 국내기업 데이터원천별 비중

- 50.6%에 해당하는 157개 기업이 이용자(사용자)로부터 데이터를 받음
- 기존 데이터/트래킹을 통한 데이터 축적 기업은 40개사로 12.9% 차지
 - ▶ 국내 기업의 경우, 공개 데이터와 획득 데이터를 활용한 데이터 구축 비율이 12.3%로 이용자 제공에 대비해서 저조한 편이다.

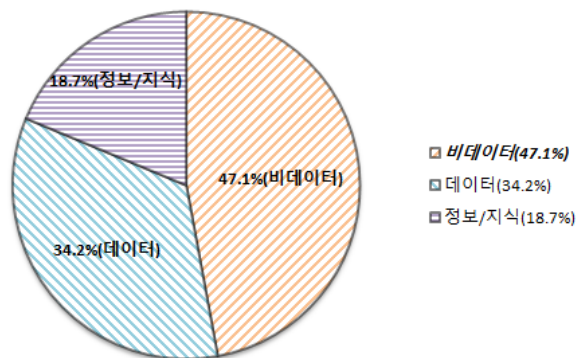
다. 데이터추출방법별 사업유형



<그림 3-4> 국내기업 데이터추출방법별 비중

- “배포” 형식이 202개사(65.2%)로 가장 많음
- 현황 분석이 29개사(9.4%), 예측 최적화가 28개사(9.0%)로 나타남
 - ▶ 배포 형식은 데이터를 나열하는 일반적인 방식으로 온라인 사이트 다수가 해당 전략을 취하고 있으며, 반면 시각화와 예측의 경우는 상대적으로 저조한 편이다.

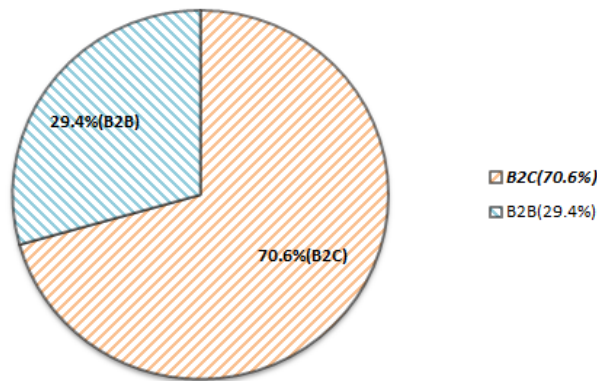
라. 공급형태별 사업유형



<그림 3-5> 국내기업 공급형태별 비중

- 비데이터가 146개사(47.1%), 데이터가 106개사(34.2%)
- 정보/지식 형태가 58개사(18.7%)로 나타남
 - ▶ 사용자가 남긴 후기, 설문, 댓글 등 정해지지 않은 데이터를 사용자로 부터 받은 후 활용하는 경우가 다수를 차지

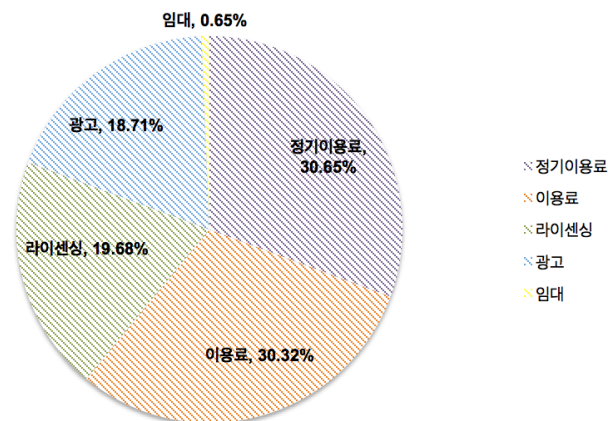
마. 주요고객별 사업유형



<그림 3-6> 국내기업 주요고객별 비중

- 70.6%에 해당하는 219개 기업이 B2C 모델을 가짐
- 기업 대상의 판매를 하는 B2B 모델의 경우는 29.4% 분포됨
 - ▶ 상대적으로 판매 장벽이 낮은 B2C 시장에 접근이 많은 편이다.

바. 수익창출 방법별 사업 유형



<그림 3-7> 국내기업 수익창출 방법별 비중

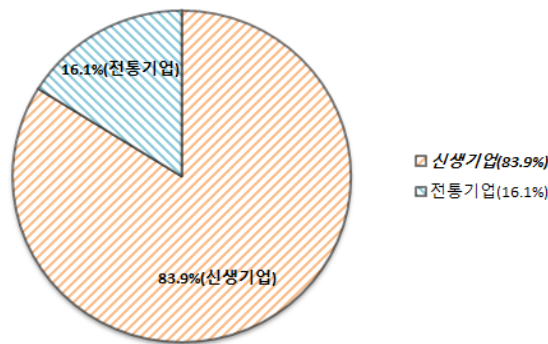
- 정기이용료를 사용하는 수익 방식이 95개사로 30.7%로 이룸
- 사용량 당 청구되는 이용료 청구 방식이 94개로 30.3% 이룸
 - ▶ B2C 사업 모델이 큰 만큼 라이선싱 보다는 정기이용료 또는 이용료 당 산출되는 수익 방식이 60%를 차지한다.

3. 설립시기에 따른 비교 분석

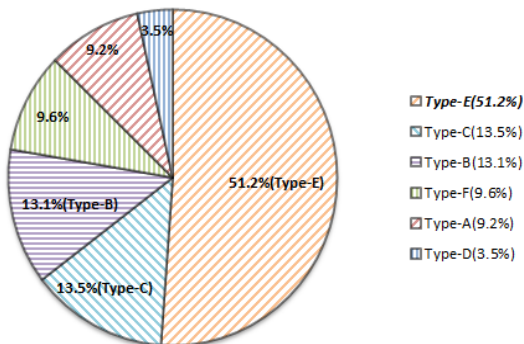
앞선 국내기업 분석을 전통기업과 신생기업으로 나눠 분석하면 다음과 같은 결과가 나온다. 여기서 전통기업과 신생기업은 2010년을 기준으로 구분하였다.

<표 3-3> 국내기업 설립시기에 따른 유형 비중 차이

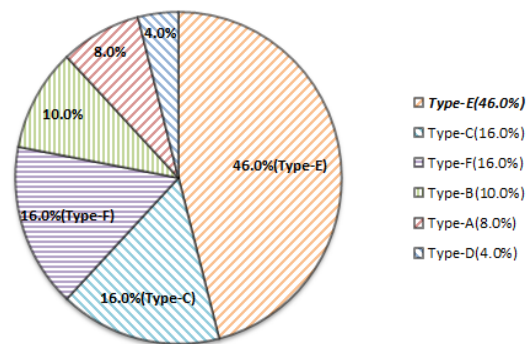
	Type-A	Type-B	Type-C	Type-D	Type-E	Type-F	계
신생기업	24(9.2%)	34(13.1%)	35(13.5%)	9(3.5%)	133(51.2%)	25(9.6%)	260(83.9%)
전통기업	4(8.0%)	5(10.0%)	8(16.0%)	2(4.0%)	23(46.0%)	8(16.0%)	50(16.1%)



<그림 3-8> 국내기업 신생기업 비중



<그림 3-9> 국내 신생기업 유형별 비중



<그림 3-10> 국내 전통기업 유형별 비중

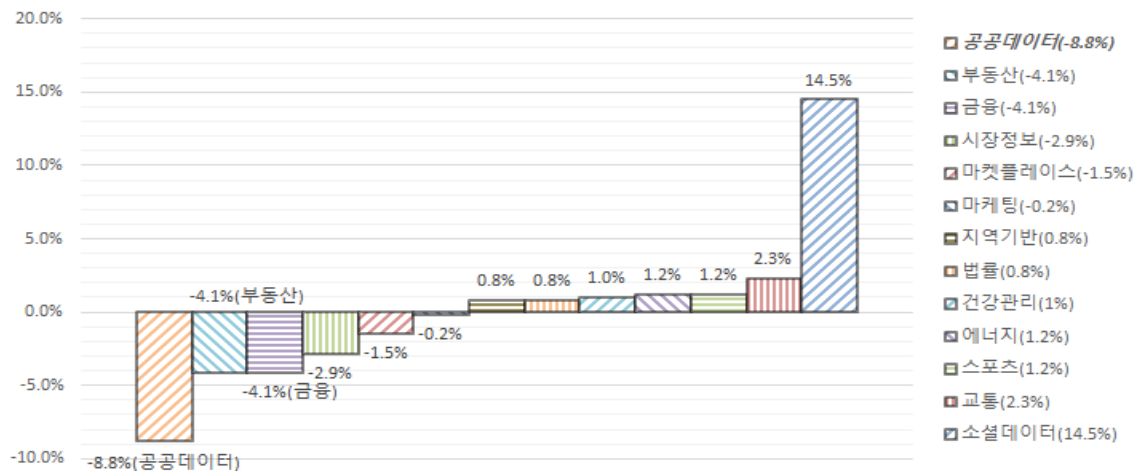
전통기업은 50개사(16.1%)로 나타났으며, 신생기업은 260개사(83.9%)로 나타났다. 구분에 상관없이 E유형의 유형이 가장 많이 활용되었다.

가. 산업별 사업유형

<표 3-4> 국내기업 설립시기에 따른 산업 비중 차이

마켓플레이스	소셜데이터	마케팅	금융	건강관리	시장정보	교통
-1.5%	14.5%	-0.2%	-4.1%	1%	-2.9%	2.3%
부동산	에너지	공공데이터	스포츠	지역기반	법률	
-4.1%	1.2%	-8.8%	1.2%	0.8%	0.8%	

- 신생 기업의 소셜데이터의 분야가 전통 기업 대비해서 14.5% 증가



<그림 3-11> 국내기업 설립시기에 따른 산업 비중 차이

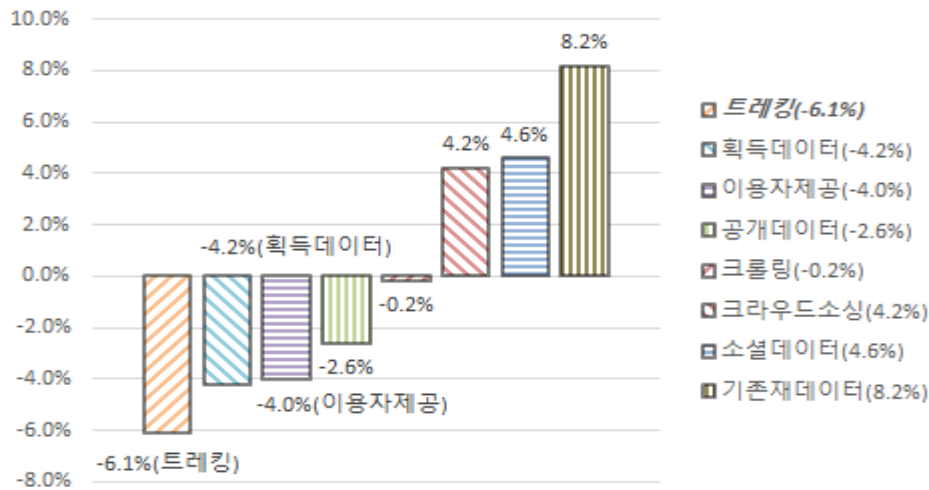
전통기업은 시장에서 금융 분야의 비중(16%)이 신생기업(11.9%)에 비하여 약 4% 높게 나타났다. 신생기업에서는 소셜데이터 부문이 전통기업의 비중(4%)에서 신생기업(18.5%)로 14% 증가하여 2010년 이후 시장에서 상대적으로 큰 성장이 이루어졌음을 알 수 있다.

나. 데이터원천과 사업유형

<표 3-5> 국내기업 설립시기에 따른 데이터원천별 비중 차이

이용자제공	기존재데이터	트래킹	획득데이터
-4.0%	8.2%	-6.1%	-4.2%
공개데이터	소셜데이터	클라우드소싱	크롤링
-2.6%	4.6%	4.2%	-0.2%

● 전통 기업 대비해서 기존재데이터 활용이 8.2% 증가함



<그림 3-12> 국내기업 설립시기에 따른 데이터원천별 비중 차이

전통기업의 주된 데이터 원천 중 획득한 데이터(10%)가 신생기업에서 기존재데이터(14.2%)에게 그 상대적 비중을 내주었다. 이는 시간의 흐름에 따라 더 많은 데이터가 누적되어 존재하게 되는 현실이 반영되었다고 보인다.

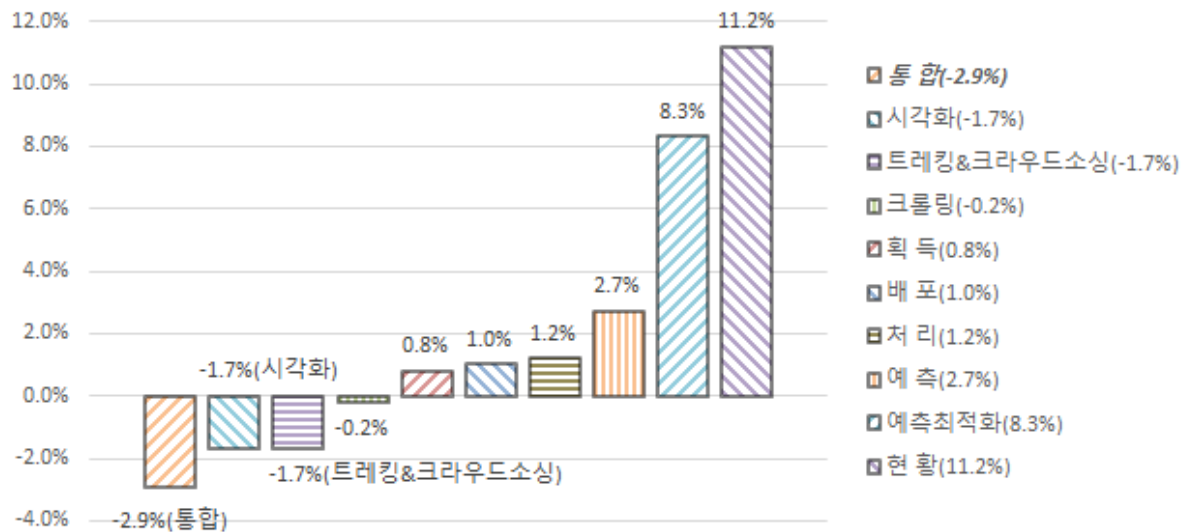
이는 전통기업에서 공개된 자료를 취득하는 업체가 1개사에서 신생기업 7개사로 7배 상승한 것에서도 알 수 있는데, A유형은 무료 데이터 수집 및 통합 유형으로 주로 무료 API를 통한 비즈니스 유형으로, 주요 데이터 원천이 공개정보이다.

다. 데이터활용 방법별 사업유형

<표 3-6> 국내기업 설립시기에 따른 데이터추출방법별 비중 차이

배 포	현 황	예측최적화	크롤링	통 합
1.0%	11.2%	8.3%	-0.2%	-2.9%
예 측	시각화	트래킹&클라우드소싱	처 리	획 득
2.7%	-1.7%	-1.7%	1.2%	0.8%

● 전통 기업 대비해서 현황 분석 방식이 11.2% 증가함



<그림 3-13> 국내기업 설립시기에 따른 데이터추출방법별 비중 차이

전통기업에서의 자료추출 방법 중 통합의 비중(6%)이 신생기업에서는 절반 수준인 3%로 떨어졌으며, 전통기업에서는 나타나지 않던 현황 분석이 신생기업에서는 11.2%의 비중을 차지하게 되었다.

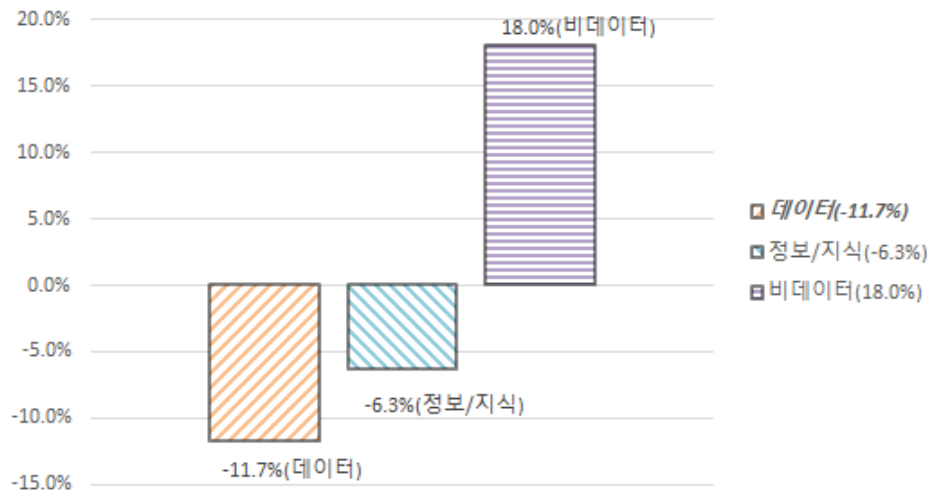
조사대상 전통기업들이 6가지의 단순한 자료추출 방법을 사용했었으나, 신생기업에서는 데이터산업의 성장과 함께 기존 전통기업들의 두 배에 가까운 10가지의 자료추출 방법을 사용하여 자료추출 방법이 과거에 비해 보다 다양하게 발전하였음을 알 수 있다.

라. 공급형태별 사업유형

<표 3-7> 국내기업 설립시기에 따른 공급형태별 비중 차이

비데이터	데이터	정보/지식
18.0%	-11.7%	-6.3%

- 전통 기업 대비해서 신생 기업의 비데이터 활용이 18% 증가함



<그림 3-14> 국내기업 설립시기에 따른 공급형태별 비중 차이

전통기업에서는 데이터 자체의 공급을 위한 회사의 비중(44%)이 가장 높았으나, 신생 기업에서는 데이터를 활용한 용역과 재화의 공급이 50%로 주를 이루었다.

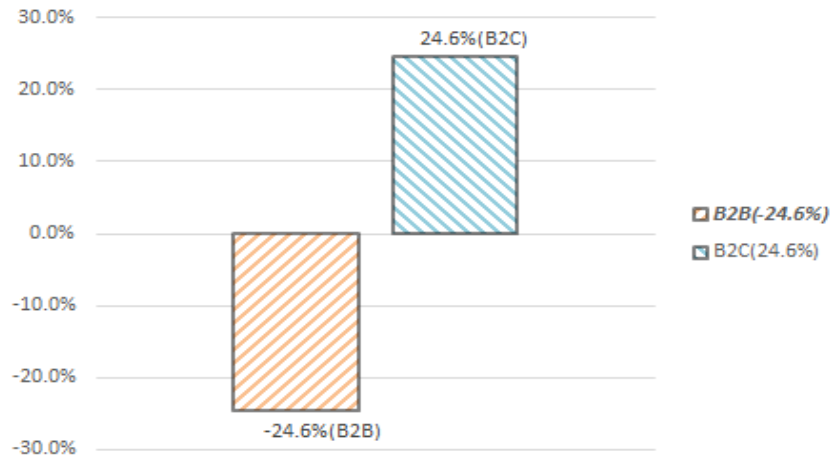
이로부터 국내 데이터 산업에서는 전통기업 50개에서 신생기업 260개로 외형적으로 성장한 듯 보이긴 하였으나, 데이터를 활용은 하되 실질적으로 데이터 산업 자체의 성장이 아닌, 데이터를 활용한 재화나 용역의 공급이 이루어져 전통적인 산업방식에서 크게 벗어나지 못하였다고 할 수 있을 것이다.

마. 주요고객과 사업유형

<표 3-8> 국내기업 설립시기에 따른 주요고객별 비중 차이

B2C	B2B
24.6%	-24.6%

- 전통 기업 대비해서 신생 기업의 B2C 고객 모델이 24.6% 증가함



<그림 3-15> 국내기업 설립시기에 따른 주요고객별 비중 차이

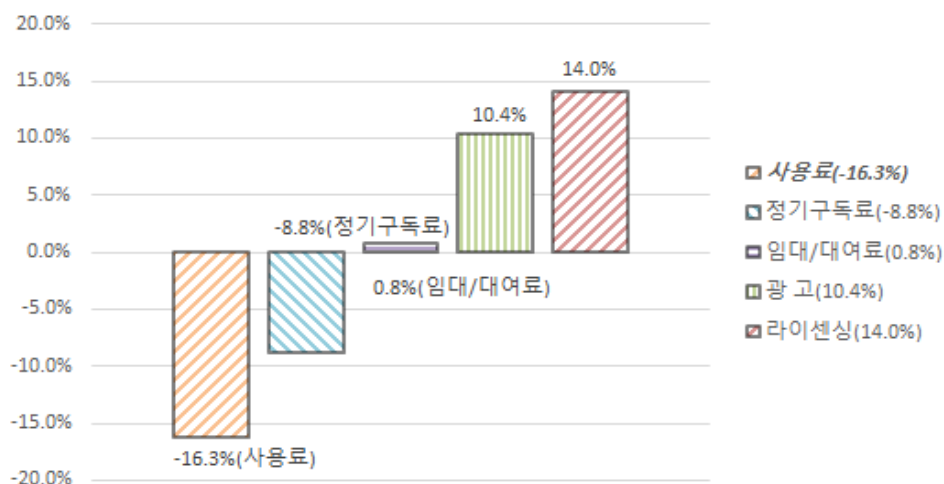
전통기업에서는 주요 수요자가 특정되지 않았으나, 신생기업에서는 주된 수요자가 소비자로나타나, 개별 주체의 데이터 수요 및 활용이 증가하였을 것이라 추측할 수 있다. 신생기업에서 일반 소비자에 편중된 주요고객 현황은 앞으로의 데이터 산업에서의 규모측면에서 성장 가능성을 보여준다고 할 수 있을 것이다.

바. 수익창출 방법별 사업유형

<표 3-9> 국내기업 설립시기에 따른 수익창출방법별 비중 차이

정기이용료	사용료	라이선싱	광 고	임대/대여료
-8.8%	-16.3%	14.0%	10.4%	0.8%

● 전통 기업 대비해서 라이선싱 수익 창출 모델이 14% 증가함



<그림 3-16> 국내기업 설립시기에 따른 수익창출방법별 비중 차이

전통기업에서는 주요 수익 창출 방법에 광고가 포함되었으나, 신생기업에서는 라이선싱이 전통기업의 광고를 대체하였다. 또한 정기이용료를 통하여 수익을 창출하는 기업 중 전통기업은 유형 C를 활용하나, 신생기업은 유형 E를 활용하는 것으로 나타났다.

제2절 해외 데이터 기업 현황 분석

1. 종합 분석

동일한 방법론을 적용하여 분석한 결과는 다음과 같다. 해외기업 379개 업체를 분석하였다.

〈표 3-10〉 해외기업 사업유형

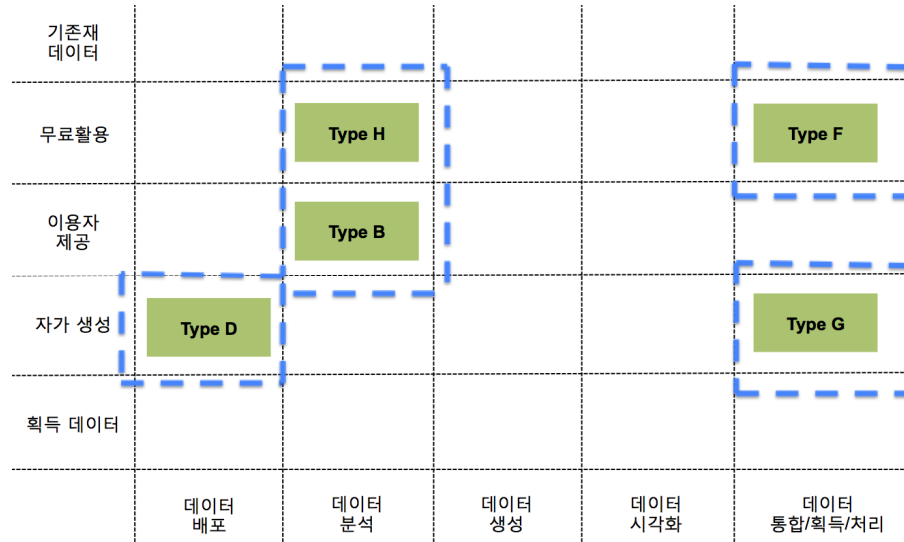
imn Lab											
데이터 배포		데이터 분석		데이터 생성		데이터 시각화		데이터 통합/획득/처리		Total 개수	Total %
Row Lab	개수	%	개수	%	개수	%	개수	%	개수	%	
기존데이터	9	2.37%		0.00%		0.00%		0.00%		0.00%	9 2.37%
무료활용	3	0.79%	28	7.39%	10	2.64%	15	3.96%	149	39.31%	205 54.09%
이용자제공	3	0.79%	35	9.23%		0.00%	3	0.79%	8	2.11%	49 12.93%
자가생성	39	10.29%	26	6.86%	3	0.79%	14	3.69%	30	7.92%	112 29.55%
획득데이터		0.00%	3	0.79%		0.00%		0.00%	1	0.26%	4 1.06%
Grand Total	54	14.25%	92	24.27%	13	3.43%	32	8.44%	188	49.60%	379 100.00%

앞서 서론에서 언급된 케임브리지 연구에 활용된 6개의 범주를 활용하여 분석하였으며, 대표적인 유형 5개를 선정했다. (국내와 동일한 모델의 경우는 A부터 E 유형에서 반영)

- F 유형 (무료 활용 및 데이터 통합/획득/처리) : 149개 기업 (39.31%)
: 스타트업게놈(창업 통합 분석), 쿼들(금융 분석 정보 처리), 서치맨(앱 정보 처리)
- B 유형 (이용자 제공 및 데이터 분석) : 35개 기업 (9.23%)
: 스타트업 랭킹(스타트업 순위), 시티즌미(개인성향 분석), 아이레(신용 평가서비스)
- G 유형 (자가 생성 및 데이터 통합/획득/처리) : 30개 기업 (7.92%)
: 크레딧쿠도스(개인 신용평가), 리빌모바일(앱 사용 분석), 보모다(사용자 행동 분석)
- D 유형 (자가 생성 및 데이터 배포) : 39개 기업 (10.29%)
: 아웃브레인(컨텐츠 추천), 세일쓰루(고객 관리 서비스), 악셀 스프링거(미디어 그룹)
- H 유형 (무료 활용 및 데이터 분석) : 28개 기업 (7.39%)
: 트라피니티(큐레이션 금융 분석), 앱애니(전체 앱 순위 분석), 부루(주식 분석)

국내 데이터 주도 비즈니스의 경우는 데이터를 통합/획득/처리 제공하는 서비스가 39%의 비중을 차지하고 있으며, 데이터 원천은 무료 활용일 54% 이상을 차지함을 알 수 있다.

<그림 3-17> 해외 기업 사업 유형 검토



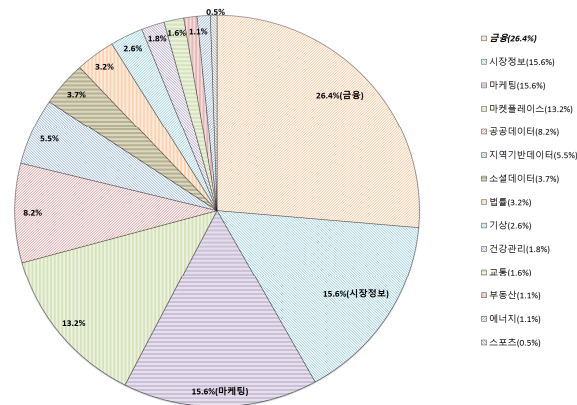
다음 <표 4-25>는 본 연구를 통해 조사한 해외기업 30개를 유형별로 나열한 것이다.

<표 3-11> 해외기업 분류기준별 사업유형 현황 종합

회사명	산 업	데이터 원천	데이터 추출방법	공 급	주요 고객	수익구조	설립 시기
Sportradar	스포츠	크롤링	획득	데이터	B2B	정기이용료	신생
Indix	시장정보	크롤링	획득	데이터	B2B	정기이용료	신생
Snolr	금융	공개데이터	처리	비데이터	B2C	광고	신생
HealthiQ	건강관리	크롤링	획득	데이터	B2B	정기이용료	신생
Enigma	공개데이터	크롤링	획득	데이터	B2B	정기이용료	신생
DueDil	마케팅	이용자제공	획득	데이터	B2B	정기이용료	신생
Datacoup	마켓플레이스	이용자제공	획득	데이터	B2B	광고	신생
Sickweather	건강관리	클라우드소싱	획득	데이터	B2B	광고	신생
Poshly	시장정보	이용자제공	처리	정보/지식	B2B	광고	신생
UberRetail	지역기반	클라우드소싱	시각화	데이터	B2B	광고	신생
Last10k	금융	공개데이터	현황	정보/지식	B2B	광고	신생
Trendlyne	금융	공개데이터	현황	정보/지식	B2B	광고	신생
Owler	금융	클라우드소싱	획득	데이터	B2B	광고	신생
Google	소셜데이터	크롤링	크롤링	데이터	B2C	광고	전통
Baidu	소셜데이터	크롤링	크롤링	데이터	B2C	광고	전통
Yahoo! Inc.	소셜데이터	크롤링	크롤링	데이터	B2C	광고	전통
AOL Inc.	소셜데이터	이용자제공	배포	데이터	B2C	광고	전통
IBM	마케팅	기존재데이터	배포	비데이터	B2B	라이센싱	전통
Benesse	마켓플레이스	트래킹	배포	정보/지식	B2C	이용료	전통
Oracle	마켓플레이스	기존재데이터	배포	데이터	B2B	라이센싱	전통
CoreLogic	금융	이용자제공	예측최적화	데이터	B2B	이용료	전통
MSCI Inc.	금융	이용자제공	예측	데이터	B2B	이용료	전통
Ipsos	마케팅	이용자제공	현황	데이터	B2B	이용료	전통
WesternGeco	지역기반	획득데이터	예측	데이터	B2B	라이센싱	전통
UBM plc	마케팅	이용자제공	시각화	비데이터	B2B	이용료	전통
LexisNexis	법률	이용자제공	현황	데이터	B2B	이용료	전통
BNA	법률	이용자제공	현황	데이터	B2B	이용료	전통
FinGenius	금융	트래킹	예측	정보/지식	B2B	이용료	신생
ARRIA	마케팅	크롤링	예측	데이터	B2B	이용료	신생
Summly	마케팅	크롤링	배포	정보/지식	B2C	이용료	신생

2. 유형별 분석

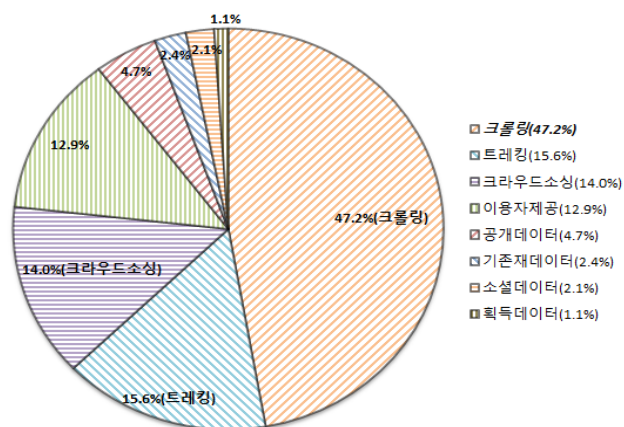
가. 산업별 사업유형



<그림 3-18> 해외기업 산업별 비중

우측 마지막 열을 보면 금융 분야가 100개사(26.4%)로 가장 많은 비중을 차지하였으며 다음으로 마케팅과 시장정보가 각 59개사(15.6%)로 나타났다.

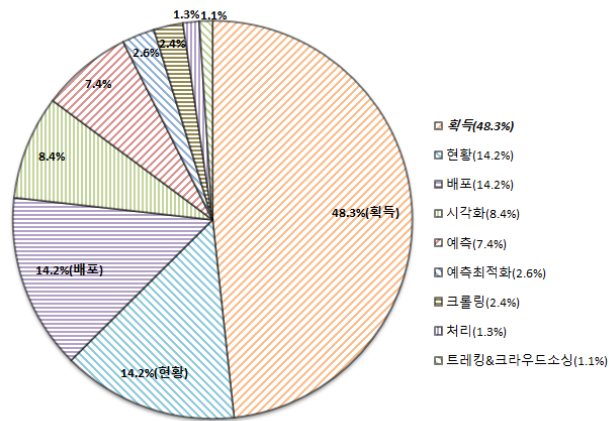
나. 데이터원천별 사업유형



<그림 3-19> 해외기업 데이터원천별 비중

조사대상 기업 중 179개사(47.2%)가 크롤링을 통해 정보를 취득하였으며, 클라우드소싱을 사용하는 기업이 53개사(14.0%), 사용자 정보로 취득하는 형태가 49개사(12.9%)로 나타났다.

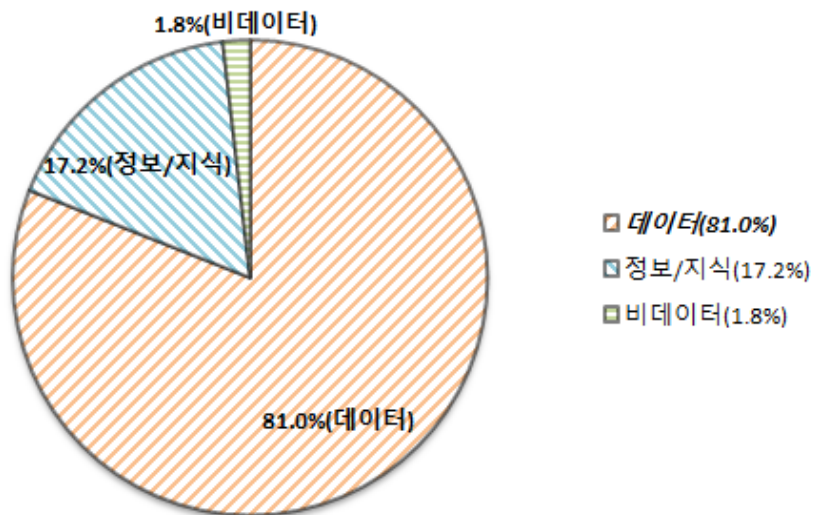
다. 데이터추출방법별 사업유형



<그림 3-20> 해외기업 데이터추출방법별 비중

통합/획득 방식이 183개사(48.3%)로 가장 많았으며 배포식이 54개사(14.2%)로 나타났다.

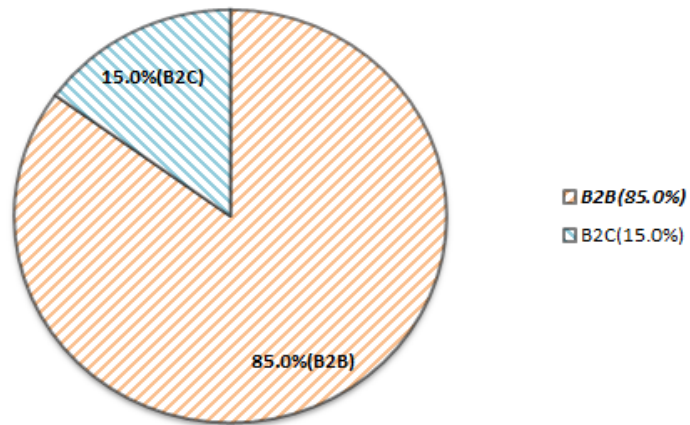
라. 공급형태별 사업유형



<그림 3-21> 해외기업 공급형태별 비중

데이터가 146개사(81.0%), 정보/지식 형태가 106개사(9.5%), 비데이터 형태가 7개사(1.8%)로 나타났다. 데이터 산업내에서 데이터 자체의 공급이 주로 이루어짐을 알 수 있다.

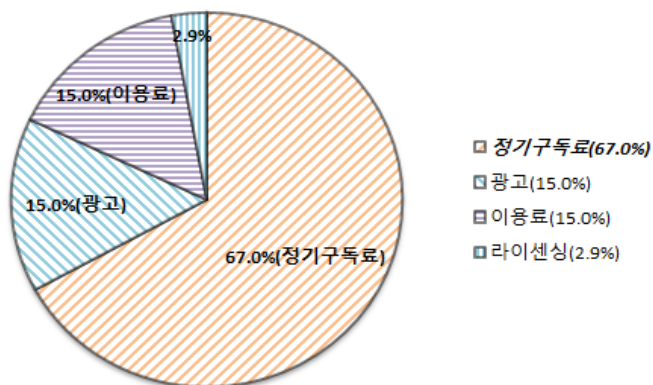
마. 주요고객별 사업유형



<그림 3-22> 해외기업 주요고객별 비중

주된 수요는 321개사(84.7%)로 기업으로 나타났다. 해외에서 기업고객의 데이터에 대한 관심 및 데이터의 중요성에 대한 관점을 보여준다고 할 수 있을 것이다.

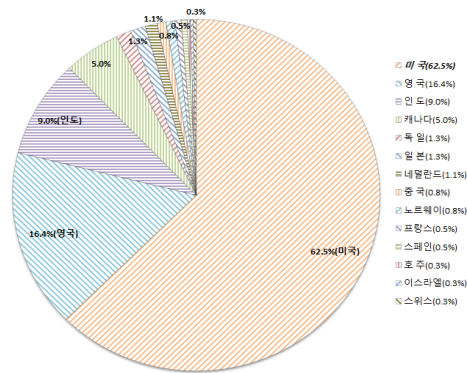
바. 수익창출 방법별 사업유형



<그림 3-23> 해외기업 수익창출 방법별 비중

구독료를 수입하는 방식이 254개사(67.0%)로 가장 많았으나, 사용료 수입과 광고가 57개사 (15.0%)로 나타났다. 이는 해외기업에서 데이터에 대한 정기적이며 꾸준한 수요가 있음을 보여주며, 데이터의 중요성에 대한 관점을 보여주는 것이라 할 수 있을 것이다.

사. 국가별 사업유형

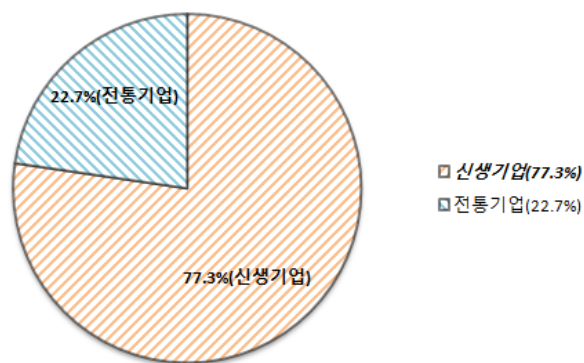


<그림 3-24> 해외기업 국가별 비중

조사대상 해외기업을 국가별로 분류하면 미국이 237개 회사(62.5%)로 전통적인 데이터 산업에서 강국임을 보여준다. 뒤이어 서비스 산업이 주된 산업을 구성하는 영국이 62개사(16.4%), IT산업을 중심으로 성장하는 인도 34개사(9.0%)로 나타났다.

3. 설립시기에 따른 해외기업 분석

앞선 해외기업 분석을 전통기업과 신생기업으로 나눠 분석하면 다음과 같은 결과가 나온다. 여기서 전통기업과 신생기업은 2010년을 기준으로 구분하였다.



<그림 3-25> 해외기업 신생기업 비중

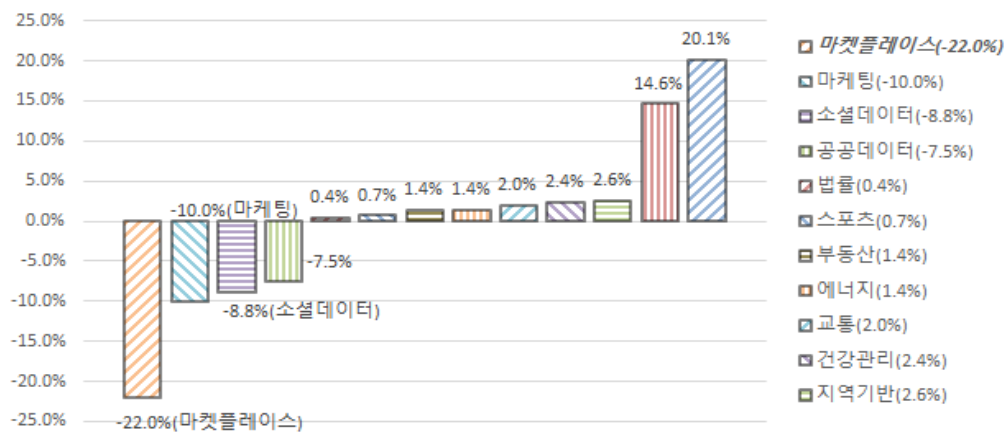
전통기업은 86개사(22.7%)로 나타났으며, 신생기업은 293개사(77.3%)로 나타났다. 86개사의 전통기업에서 293개의 신생기업으로 데이터 산업의 비약적인 성장이 있었음을 볼 수 있다.

가. 산업별 사업 유형

<표 3-12> 해외기업 설립시기에 따른 산업 비중 차이

마켓플레이스	소셜데이터	마케팅	금융	건강관리	시장정보	교통
-22.0%	-8.8%	-10.0%	14.6%	2.4%	20.1%	2.0%
부동산	에너지	공공데이터	스포츠	지역기반	법률	
1.4%	1.4%	-7.5%	0.7%	2.6%	0.4%	

● 전통 기업 대비해서 신생 기업의 시장 정보 모델이 20.1% 증가함



<그림 3-26> 해외기업 설립시기에 따른 산업 비중 차이

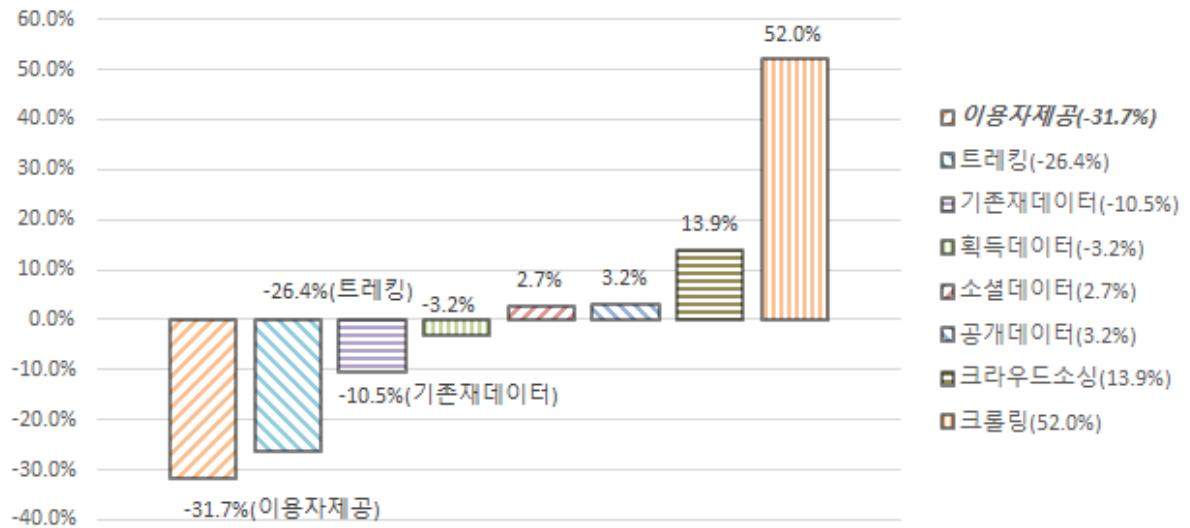
전통기업은 시장에서 마켓플레이스 부문의 비중(30.2%)이 신생기업(8.2%)에 비하여 22% 높게 나타났다. 전통기업에서 마켓플레이스 부문을 대체한 것이 신생기업의 시장정보라고 여겨지며, 시장정보의 주된 고객은 시장에서 생산 및 공급을 담당하는 기업일 것이라 추측할 수 있다.

나. 데이터원천별 사업유형

<표 3-13> 해외기업 설립시기에 따른 데이터원천별 비중 차이

이용자제공	기존재데이터	트래킹	획득데이터
-31.7%	-10.5%	-26.4%	-3.2%
공개데이터	소셜데이터	클라우드소싱	크롤링
3.2%	2.7%	13.9%	52.0%

● 전통 기업 대비해서 신생 기업의 크롤링 데이터 원천이 52% 증가함



<그림 3-27> 해외기업 설립시기에 따른 데이터원천별 비중 차이

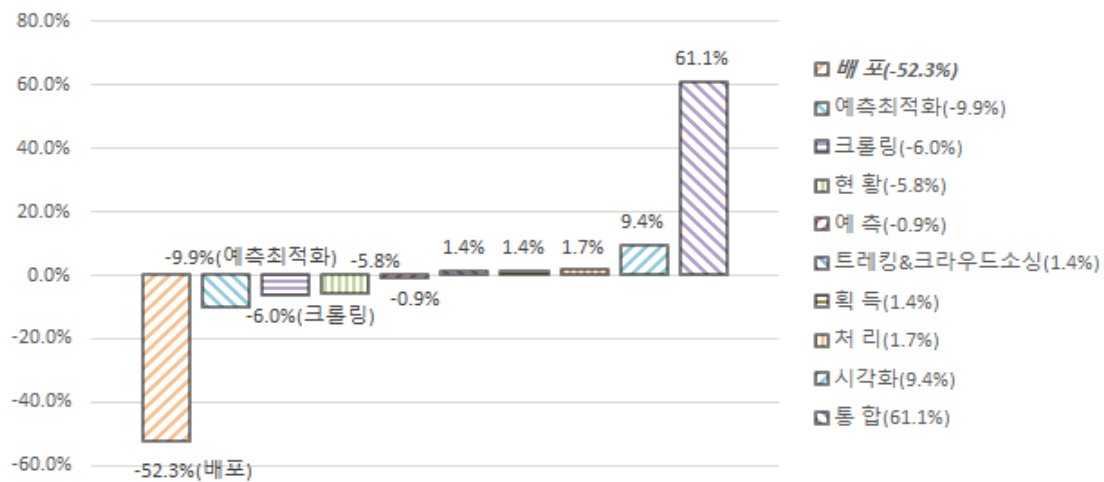
전통기업의 주된 데이터 원천 중 기존재 데이터(10.5%)와 사용자제공정보(37.2%)가 신생 기업의 크롤링(59.0%)과 클라우드소싱(17.4%)에게 그 상대적 비중을 내주었다.

다. 데이터추출 방법별 사업유형

<표 3-14> 해외기업 설립시기에 따른 데이터추출방법별 비중 차이

배 포	현 황	예측최적화	크롤링	통 합
-52.3%	-5.8%	-9.9%	-6.0%	61.1%
예 측	시각화	트래킹&클라우드소싱	처 리	획 득
-0.9%	9.4%	1.4%	1.7%	1.4%

● 전통 기업 대비해서 신생 기업의 통합 데이터 추출 모델이 61% 증가함



<그림 3-28> 해외기업 설립시기에 따른 데이터추출방법별 비중 차이

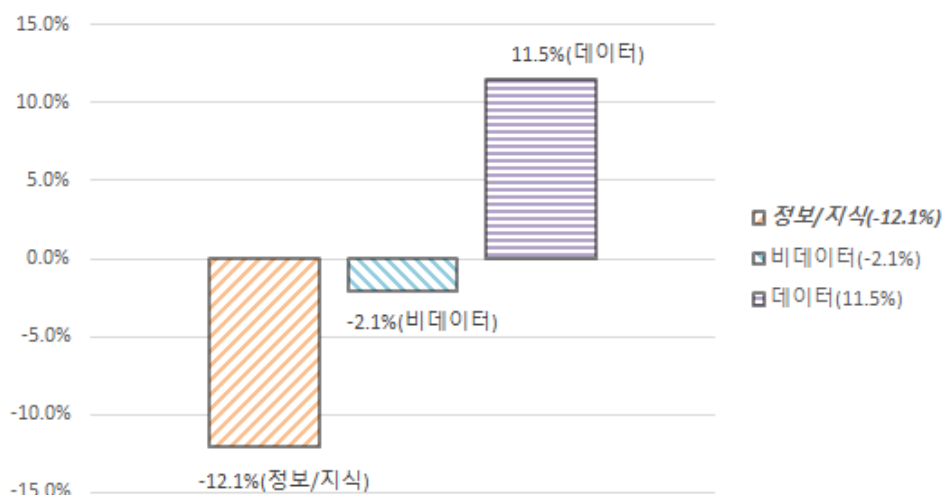
전통기업의 주된 데이터 추출방법 중 배포형(54.7%)과 Prescriptive(11.6%)가 신생기업의 통합형(61.1%)과 시각화(10.6%)에게 그 상대적 비중을 내주었다. 전통기업에서는 6가지의 단순한 데이터 추출방법이 활용되었으나, 신생기업은 10가지 방법의 데이터 추출방법을 활용하여, 과거보다 데이터 산업이 발전되었음을 알 수 있다.

라. 공급형태별 사업유형

<표 3-15> 해외기업 설립시기에 따른 공급형태별 비중 차이

비데이터	데이터	정보/지식
-2.1%	11.5%	-12.1%

● 전통 기업 대비해서 신생 기업의 데이터 공급 방식이 11.5% 증가함



<그림 3-29> 해외기업 설립시기에 따른 공급형태별 비중 차이

전통기업과 신생기업의 구분 없이 데이터 형태, 정보/지식 형태, 비데이터 형태의 공급이 이루어졌다. 비데이터 형태의 공급 비중이 전통기업 3.5%에서 신생기업 1.4%로 감소한 것으로부터 데이터 산업의 성장이 실질적으로 데이터를 위주로 이루어졌음을 알 수 있다.

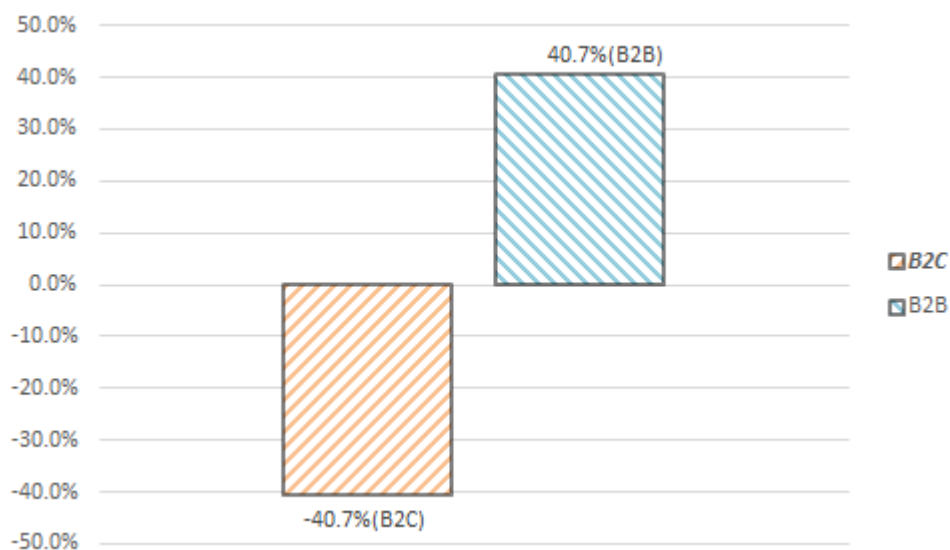
정보/지식 형태의 공급이 이루어지는 회사는 C유형의 사업유형을 주로 보였으며, 데이터 형태의 공급이 이루어지는 회사는 전통기업의 B유형(41.9%)에서 신생기업의 A유형(64.1%)로 가시적인 변동이 있었다고 할 수 있다. 비데이터 형태는 표본이 적어 신빙성이 낮다고 판단할 수 있으나 신생기업과 전통기업에서 모두 특정 유형으로 편중되어 어느 정도 분석의 의미가 있다고 본다.

마. 주요고객별 사업유형

<표 3-16> 해외기업 설립시기에 따른 주요고객별 비중 차이

B2C	B2B
-40.7%	40.7%

● 전통 기업 대비해서 신생 기업의 B2B 고객 모델이 40.7% 증가함



<그림 3-30> 해외기업 수익창출 방법별 비중

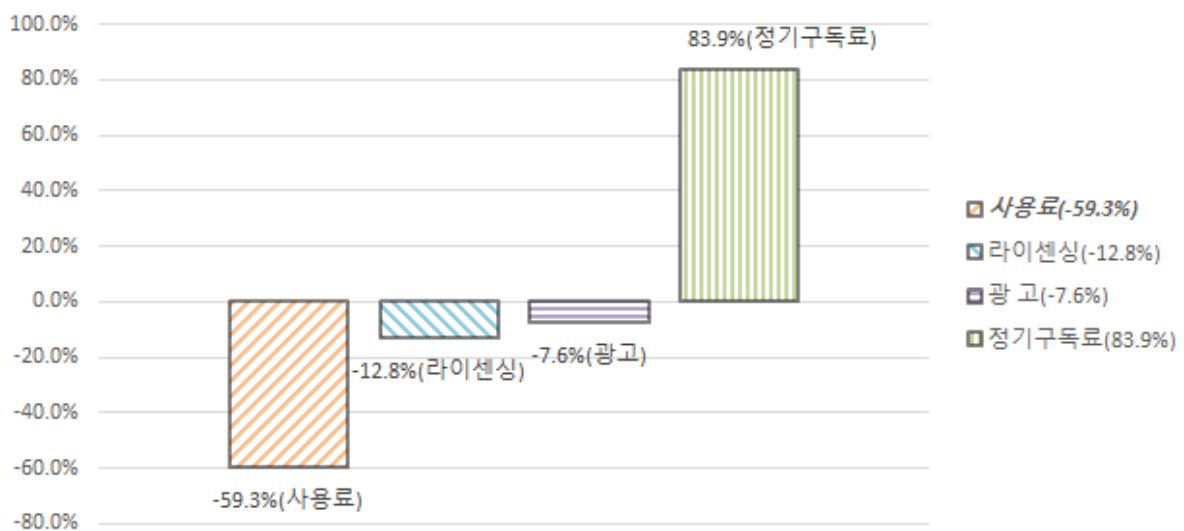
전통기업에서는 주요 고객이 유사한 수준이었으나, 신생기업에서는 주된 고객이 기업 고객으로 나타나, 산업적으로 규모(Volume)있는 시장을 형성하고 있음을 유추할 수 있으며, 데이터 수요 고객인 기업들이 데이터에 대한 중요성을 인지하고 있다는 것을 보여준다.

바. 수익창출 방법별 사업유형

<표 3-17> 해외기업 설립시기에 따른 수익창출방법별 비중 차이

정기이용료	사용료	라이선싱	광 고
83.9%	-59.3%	-12.8%	-7.6%

● 전통 기업 대비해서 신생 기업의 정기이용료 수익 모델이 83.9% 증가함



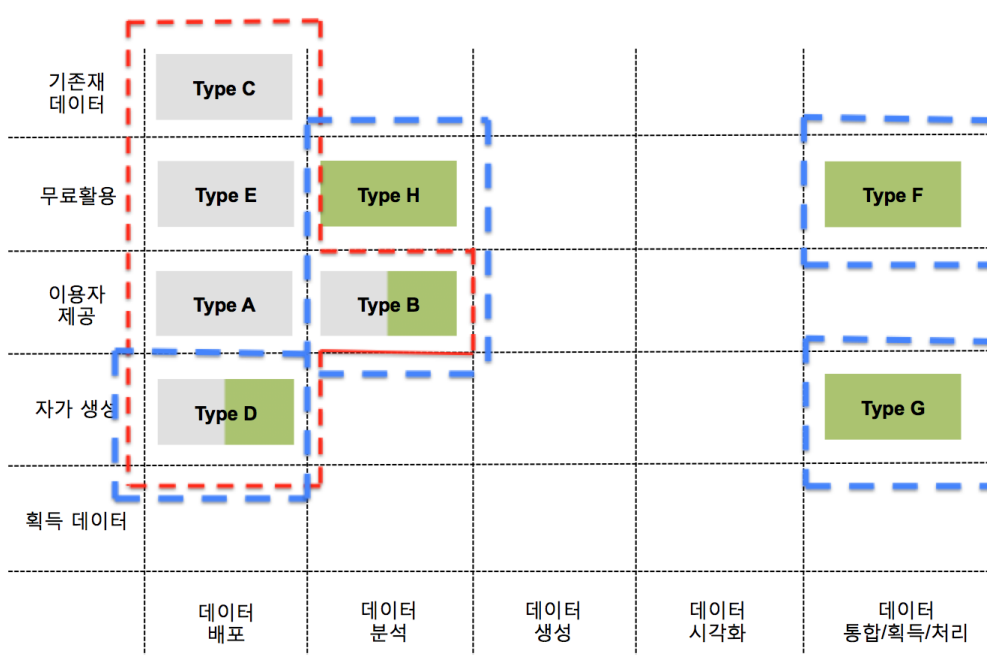
<그림 3-31> 해외기업 설립시기에 따른 수익창출방법별 비중 차이

전통기업에서는 주요 수익 창출 방법이 사용료(59.3%)였으나, 신생기업에서는 정기이용료(84.6%)형태로 전통기업의 사용료를 대체하였다.

이는 데이터산업의 성장에 따라 수요자들이 단발적인 형태의 사용에 머무르지 않고 지속적인 형태로 데이터의 공급을 원하는 형태로 바뀌었다고 볼 수 있을 것이다. 또한 구독료를 통하여 수익을 창출하는 기업 중 전통기업은 C유형 유형을 활용하였으나, 신생기업은 A유형의 유형을 활용하는 것으로 나타났다.

제3절 국내외 데이터기업 비교 분석

1. 종합 분석



(붉은 점선 : 국내 기업 유형, 파란 점선 : 해외 기업 유형)

<그림 3-32> 국내외 유형 비교 분석

국내 기업의 경우는 A 유형부터 E 유형까지 다수가 데이터 배포 방식에 의존도가 높다면 해외 기업의 경우는 데이터 배포 및 통합/획득/처리 방식 비중이 높다.

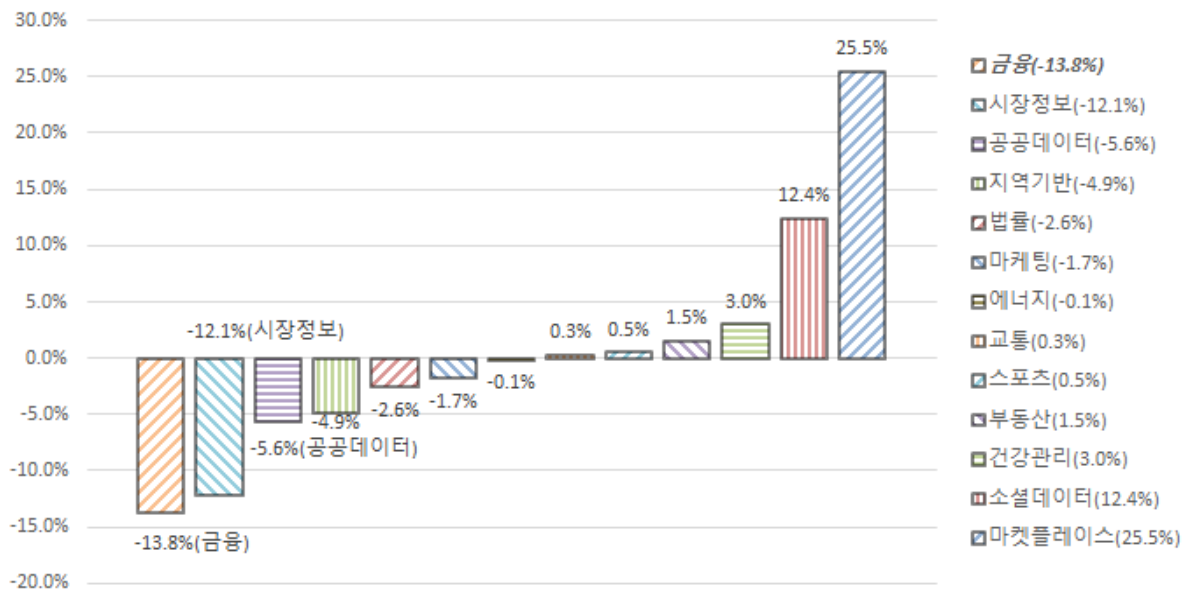
2. 국내기업 및 해외기업 분류기준별 비교분석

가. 산업별 사업유형

<표 3-18> 국내기업 및 해외기업 산업별 비중 차이

마켓플레이스	소셜데이터	마케팅	금융	건강관리	시장정보	교통
25.5%	12.4%	-1.7%	-13.8%	3.0%	-12.1%	0.3%
부동산	에너지	공공데이터	스포츠	지역기반	법률	
1.5%	-0.1%	-5.6%	0.5%	-4.9%	-2.6%	

- 해외 기업 대비해서 국내 기업의 금융 산업 분야가 13.8% 적음



<그림 3-33> 국내기업 및 해외기업 산업별 비중 차이

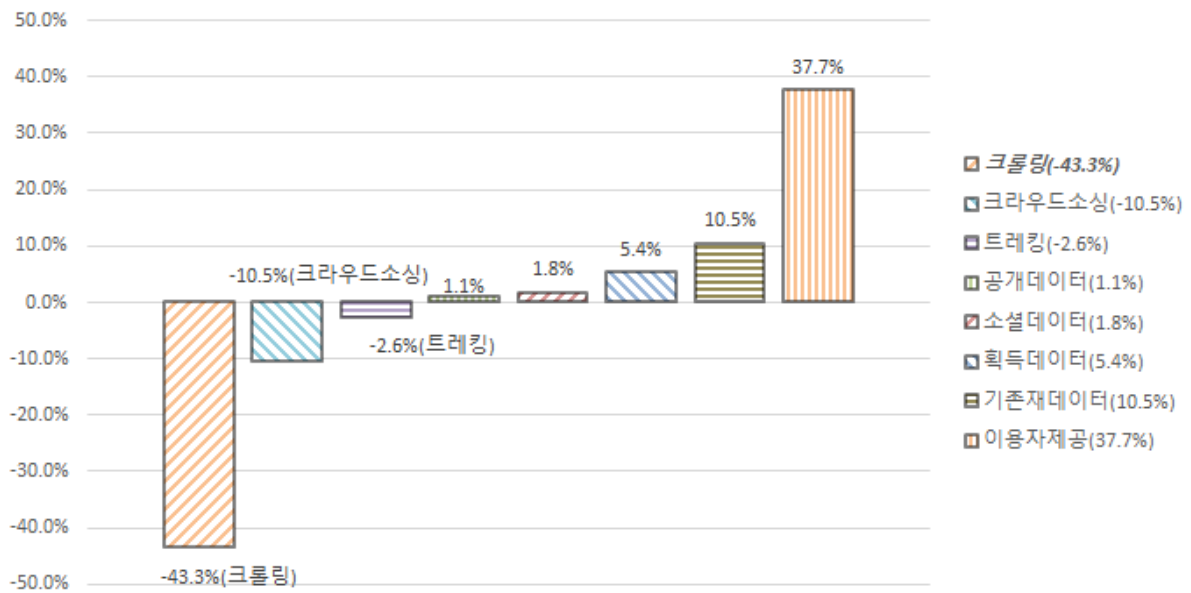
국내기업은 마켓플레이스 산업이 해외기업은 금융 산업이 상당 부분을 차지하는 것으로 나타났다. 국내기업이 해외기업에 비해 대표적으로 금융분야가 13.8%적게 나타났다. 국내기업과 해외기업 모두 마케팅 산업에서 데이터를 상당히 사용하고 해외기업은 시장정보 산업의 비중이 13.2%로 국내기업의 3.5%보다 네 배 가까이 높게 나타난다.

나. 데이터원천별 사업유형

<표 3-19> 국내기업 및 해외기업 데이터원천별 비중 차이

이용자제공	기존제데이터	트래킹	획득데이터
37.7%	10.5%	-2.6%	5.4%
공개데이터	소셜데이터	클라우드소싱	크롤링
1.1%	1.8%	-10.5%	-43.3%

- 해외 기업 대비해서 국내 기업의 경우 크롤링 방식이 43.3% 적음



<그림 3-34> 국내기업 및 해외기업 데이터원천별 비중 차이

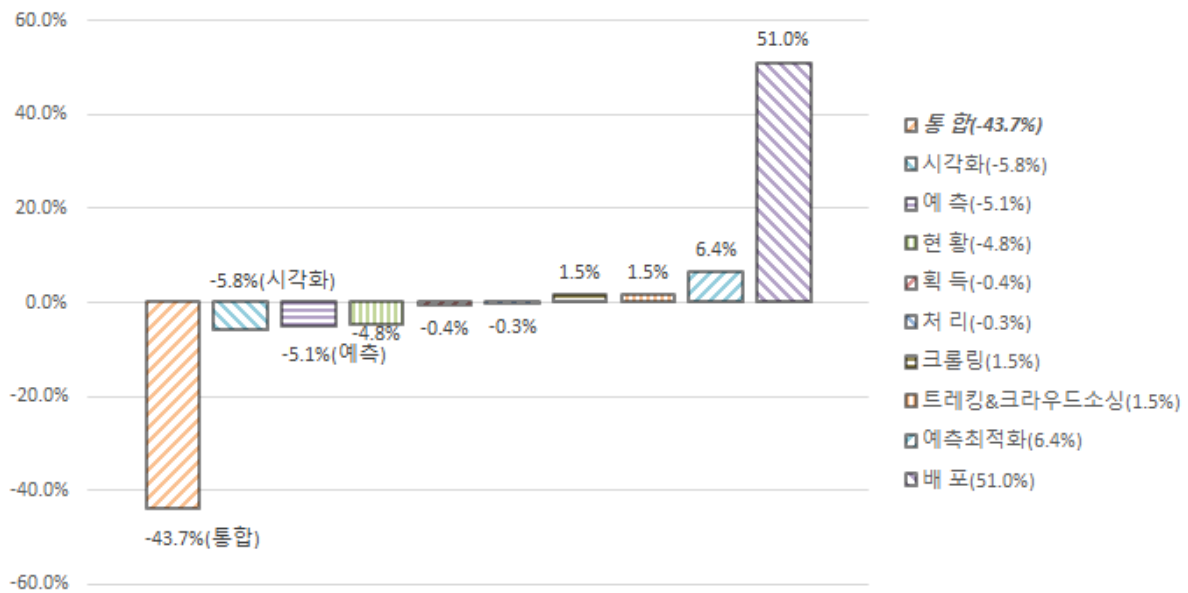
국내기업은 해외기업에 비해 크롤링이 43.3% 적게 나타났다. 국내기업의 주요 정보원천은 사용자정보를 통합을 알 수 있다. 반면 해외기업은 크롤링으로 주로 정보를 취득하는 것으로 나타났다.

다. 데이터추출방법별 사업유형

<표 3-20> 국내기업 및 해외기업 데이터추출방법별 비중 차이

배 포	현 황	예측최적화	크롤링	통 합
51.0%	-4.8%	6.4%	1.5%	-43.7%
예 측	시각화	트래킹&클라우드소싱	처 리	획 득
-5.1%	-5.8%	1.5%	-0.3%	-0.4%

- 해외 기업 대비해서 국내 기업의 경우 데이터 통합 방식이 43.7가 적음



<그림 3-35> 국내기업 및 해외기업 데이터추출방법별 비중 차이

국내기업은 해외기업에 비해서 대표적으로 통합의 사용이 43.7% 적게 나타났다. 국내 기업에서는 주요 데이터추출 방법이 배포형(65.2%)으로 나타났으나, 해외기업에서는 통합 (47.2%)형태로 전통기업의 사용료를 대체하였다.

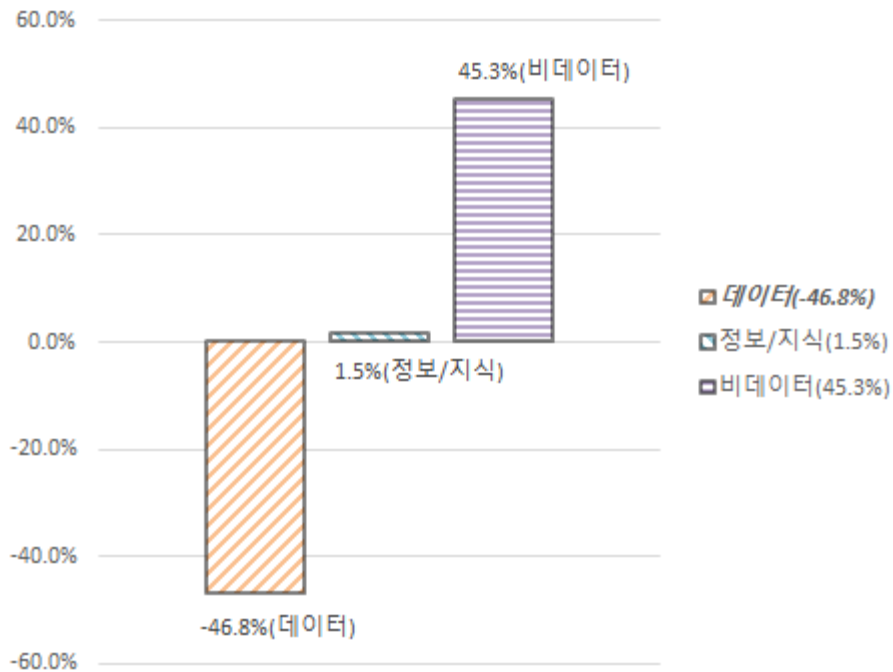
이는 데이터산업의 성장에 따라 수요자들이 지속적인 형태로 공급을 원하는 형태로 바뀌었다고 볼 수 있을 것이다. 또한 구독료를 통하여 수익을 창출하는 기업 중 전통 기업은 C유형 유형을 활용하였으나, 신생기업은 A유형의 유형을 활용하는 것으로 나타났다.

라. 공급형태별 사업유형

<표 3-21> 국내기업 및 해외기업 공급형태별 비중 차이

비데이터	데이터	정보/지식
45.3%	-46.8%	1.5%

- 해외 기업 대비해서 국내 기업의 데이터 공급 비율이 46.8% 가 적음



<그림 3-36> 국내기업 및 해외기업 공급형태별 비중 차이

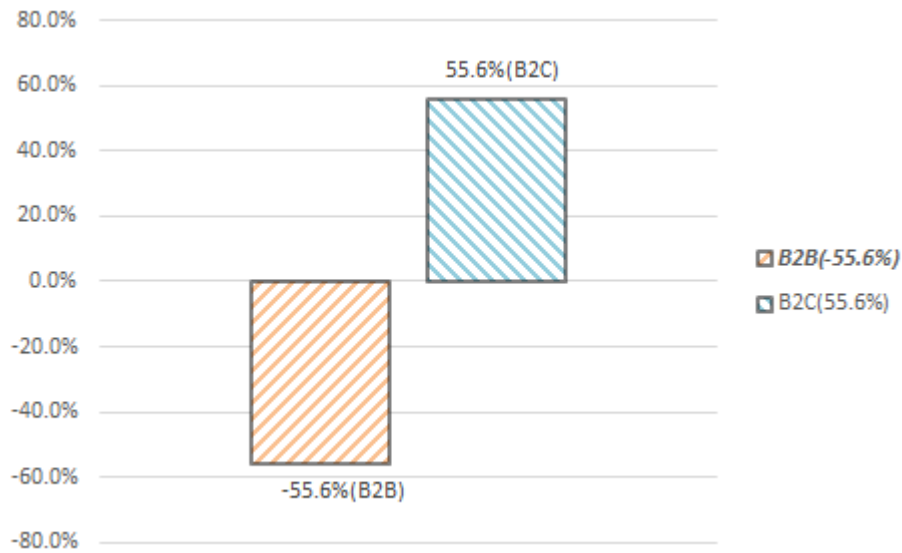
국내기업은 해외기업 대비 대표적으로 데이터 형태의 공급이 46.8% 적게 나타났다. 국내기업은 비데이터 형태의 공급이 상당 부분이나 해외기업은 데이터 형태의 공급이 상당 부분을 차지하며, 데이터 형태의 공급이 해외기업은 81.0% 국내기업은 34.2%로 그 비중이 두 배 이상 차이가 나며 이는 국내의 데이터 산업이 단편적인 수단으로 인식되는 것으로 여겨져 해외의 산업 보다 질적인 측면에서 상당부분 뒤쳐진 것으로 해석할 수 있을 것이다.

마. 주요고객별 사업유형

<표 3-22> 국내기업 및 해외기업 주요고객별 비중 차이

B2C	B2B
55.6%	-55.6%

- 해외 기업 대비해서 국내 기업의 B2B 사업 모델이 55.6% 적음



<그림 3-37> 국내기업 및 해외기업 주요고객별 사업유형 현황

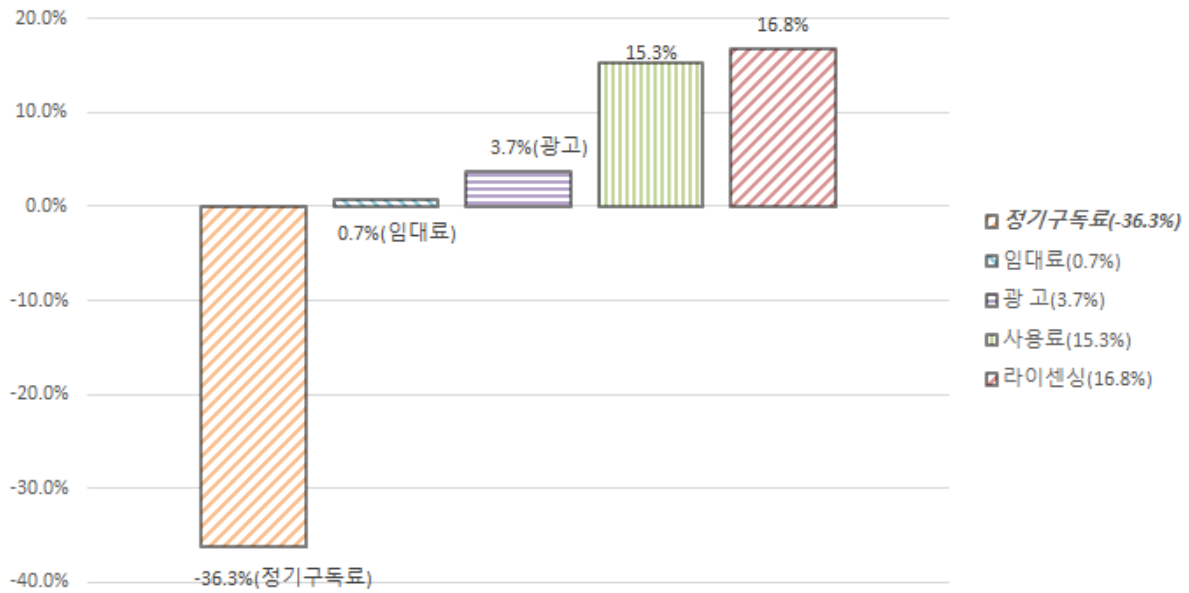
국내기업의 주된 고객은 소비자이다. 반면 해외기업은 주된 고객이 기업으로 그 비중의 차이가 괄목할만하며, 일반적으로 산업과 관계없이 기업고객이 소비자 고객보다 그 거래 규모가 상당히 크다는 점을 고려하면, 이로부터 해외에서 데이터산업 자체가 국내 데이터 산업의 규모보다 상당히 클 것이라는 점을 유추할 수 있다. 반면에, 이는 국내 데이터 산업의 성장가능성이 높다는 것을 의미한다고 볼 수 있다.

바. 수익창출 방법별 사업 유형

<표 3-23> 국내기업 및 해외기업 수익창출 방법별 비중 차이

정기이용료	사용료	라이선싱	광 고	임대료
-36.3%	15.3%	16.8%	3.7%	0.7%

- 해외 기업 대비해서 국내 기업의 정기이용료 분야가 36.3% 적음



<그림 3-38> 국내기업 및 해외기업 수익창출 방법별 비중 차이

국내기업은 해외기업에 비해 정기이용료를 통한 수익창출이 36.3% 적게 나타났다. 국내기업과 해외기업 모두 정기이용료의 형태가 주요 수익창출 형태라고 나타난다.

다만 그 비중이 해외기업은 67.0% 국내기업은 30.7%로 두 배가 넘게 차이가 나는데, 이로부터 데이터를 대하는 관점의 차이를 볼 수 있다. 해외의 데이터 사용형태가 주로 지속적으로 데이터를 활용하는 형태로 데이터의 중요성에 대한 조금 더 성숙한 관점을 보인다고 할 수 있을 것이다.

제4장 데이터 주도 비즈니스 활성화 방안

제1절 국내 데이터 주도 비즈니스 문제

1. 무료 활용 데이터 측면

무료 활용의 경우는 유형적인 면에서도 국내가 한 가지 유형이라면 해외 사례의 경우는 3가지 유형 사례를 가지고 있다.

무료 데이터 활용 사례가 낮은 문제점 중 하나는 국내 무료 활용 데이터는 그 자체의 해상도 매칭이 어렵다. 예를 들어서 우리나라 기업에 대한 공개된 정보를 수집할 경우, 데이터 제공처마다 제공 범위와 단위가 다르게 제공되고 있다.

또한 무료 데이터 자체의 품질 문제다. 무료 제공이지만 데이터 자체의 정확도와 업데이트 주기가 수시로 반영되어야 됨에도, 오래된 정보를 활용도가 낮은 경우가 많다.

더불어 현재 공공 데이터의 경우 큰 범위의 거시 데이터가 다수라 실제 중소기업이 해당 데이터를 활용하기에 필요한 미시(세부) 데이터 제공이 부족한 상황이다.

2. 신규 데이터 사업자 판로 측면

국내 데이터 판매 대상의 경우 국내 시장의 경우 해외 시장 대비해서 B2B 판매 경우가 약 55.6 % 적은 것을 알 수 있다.

B2B 사업의 경우 데이터 사업자 입장에서 판로 개척의 어려움이 존재한다. 재벌 구조 및 대기업의 재계 구조가 굳건한 모습이 한국의 실태이다. 이러한 상황에서 자체의 판매팀 및 영업전문가를 구축하지 않은 대다수의 신규 데이터 사업자의 경우, 실제 좋은 서비스 및 기획안을 가지고 있음에도 적합한 타겟 리스트를 발굴하고, 판매하는 과정의 어려움이 많다는 한계가 존재한다.

국내 모바일 관련 정보를 제공하는 스타트업의 경우, 현재 모바일 정보를 해외 애널리스트 또는 관련 해외 기업에 정보를 유로로 제공하길 원하나 판로 개척에 대한 경험이 없고, 접근 방법에 대한 지식이 없어서 매출을 글로벌로 확대하는 데 어려움을 가지고 있다고 한다.

3. 데이터 산업 인력 측면

데이터 추출 방식의 경우 해외에 비해서 통합 방식이 43.3% 가 부족한 것을 알 수 있다.

데이터를 추출 (크롤링, 스크래핑) 까지의 개발 작업은 일반 중소기업의 시니어 개발자의 경우 기업이 원하는 데이터를 추출 및 표출할 수 있으나, 모아진 데이터 간 통합을 통해 서비스 자체적으로 새로운 정보를 도출하는 기업은 상대적으로 적다.

즉, 시스템 자체적으로 데이터를 분석하고 사용자별 원하는 바를 파악한 후 최적화된 정보 제공, 더 나아가 사용자가 인지하지 못한 정보를 제공하기까지의 기술력을 가진 개발자가 적은 편이다. 대표적으로 머신러닝, 인공지능 기술이 컴퓨터공학 관련 석사, 박사에 포진되어 있는 편이나, 이 또한 다수가 대기업의 연구소 및 국책 연구(R&D)에 투입되고 있는 실정이다.

“현재 부동산 정보 관련 스타트업에 의하면, 현재 국내의 다양한 데이터를 확보한 상태이나 이를 사용자가 원하는 정보로 가공하기 위한 알고리즘 설계 및 시스템 자체에서 계속 진화하는 모델인 머신러닝 및 인공지능 개발자를 확보하기가 현실적으로 어려운 상황이라고 한다.”

4. 데이터 산업 법적 측면

데이터 원천 기준으로 국내기업은 해외기업에 비해 크롤링이 43.3% 적게 나타났다.

이는 데이터의 통합 및 크롤링에 대한 법적 이슈로 인해 통합 및 크롤링 기술이 상대적으로 상대적으로 비활성화 되어 있으며, 사업자 또는 개발자가 공적 목적을 위해서 관련 웹사이트를 해당 소유주의 사전 승인 없이 일방적으로 크롤링 한 경우, 데이터 저작권 등 법적인 이슈로 인해 크롤링을 자제하게 된다.

데이터 수집 방식에 대해 웹사이트 소유주의 데이터의 저작권을 보호와 동시에 데이터가 공적으로 사용될 수 있는 기준이 정립되지 않은 상황이다.

제2절 정책 제언

1. 산업별 데이터의 키 값 통일과 미시적인 데이터 제공 필요

모든 산업별로 데이터의 해상도가 상이하게 존재하고 있는데, 이를 분석해서 산업내의 데이터간 연결 및 통합할 수 있는 개선이 필요하다. 이러한 점이 전제되지 않는다면, 사용자들의 혼동이 있을 수 있으며, 이는 데이터라는 자원이 있음에도 이를 활용하지 못하고 시간이나 비용을 낭비하게 하여 사회적으로 비효율을 가져온다.

이를 방지하기 위한 데이터에 대한 해상도 표준 제시 등 데이터간 연결을 위한 통일된 방법을 마련할 필요가 있다.

또한 현재 공공 데이터의 데이터 제공 수준이 거시적인 큰 범위에서 세부적인 항목별로 제공이 필요하다. 이를 통해 중소기업이 실제 활용 가능한 데이터를 통해서 서비스 구현이 가능할 것으로 판단된다.

“이 부분에 대해 해외 데이터베이스 전문 회사인 크런치베이스 담당자는 초기 세팅 시간이 걸리더라도 중장기적으로 각 지역별, 산업별, 국가별로 어떠한 해상도로 모든 정보를 매칭할 것인지에 대한 내부 가이드라인이 필요하고 이 가이드라인을 다른 데이터 제공자와 사전에 공유하는 것이 필요하다고 언급했다. 또한 데이터 공개 수준 또한 법제도 하에서 세부적인 데이터를 제공했을 때 실제 활용가능한 서비스가 나올 것이라고 덧붙였다.”

2. B2B 국내/해외 판로 개척 지원

B2B 판로의 경우 구매자의 경우 그 목록이 동일한 사업체인 경우가 다수라, 산업별로 주요 또는 잠재 데이터 구매자 리스트 목록을 정리해서 정기적으로 구매자에게 신규 데이터 사업체 등을 소개하는 통합 마케팅이 필요하다.

이러한 중개를 담당하는 회사를 장려하거나, 시장이 열릴 수 있는 오픈 플랫폼을 마련해준다면, 국내 데이터 산업은 수익성 향상을 바탕으로 보다 더 성장할 수 있을 것이다.

또한 B2B 대상이 국내로 한정 짓지 말고 한국 데이터에 관심있는 해외 리서처, 관련 기업 등에게도 제공될 수 있는 판로 개척이 필요하다.

3 머신러닝, 인공지능 개발인력 육성 필요

급변하는 시장 환경에서 기존 개발 교육자에 대한 앞선 기술에 대한 공부が必要하고, 이를 통해 자연적으로 피교육자들의 수준을 향상 시키는 것이 필요하다. 또한 해외의 우수 인공지능 교육자의 교육 영상을 국내 개발자 커뮤니티에 지속 제공하는 것도 병행 필요할 것이다.

“이에 대해 미국 최대 기업 정보 및 채용 정보 회사인 엔젤리스트 담당자는 국내 채용 정보가 급여, 복지, 회사 스톡옵션 등에 대해 우수 개발자에 대해 합당한 대우를 하는 기업 문화가 필요하고, 해당 우수 개발자가 강연 및 블로그를 통해서 자신의 개발 경험을 공유하는 문화가 병행되어야 한다고 언급했다.”

4. 정보 수집에 대한 법적 이슈에 대한 기준 정립 필요

현재 정보 수집 방법에 대한 데이터 주도 비즈니스 사업체라는 명확한 구분이 없으며, 그 시작도 오래되지 않아 관련 법규가 미비하고, 판례 또한 전무한 상태이다.

데이터 산업에 있어 그 근간이 되는 1차 데이터 자체의 수집방법 및 가공된 최종 단계의 데이터에 이르기까지 법적으로 명확한 기준이 제시되어 있지 않아, 데이터 산업 속의 회사들은 보수적이고 소극적인 태도를 보이는 경향이 있다. 이를 개선하기 위한 법적 정비 및 정기적으로 데이터 산업에 대한 개정 내용 등에 대한 교육을 통해 건강한 데이터 수집 및 유통이 이루어지는 문화를 정립할 필요가 있다.

5. 데이터 기반 중소기업에 대한 후속 투자 및 M&A 시장 활성화 필요

국내 데이터 기반 비즈니스 모델의 경우, 투자자가 본엔젤스, DSC 인베스트먼트 등 초기 투자가 주로 이루고 있으면 20억 이상의 후속 투자자는 상대적으로 적어 데이터 기반 중소기업이 서비스를 확장하는 데 어려운 실정이며, 또한 인수, 합병의 경우는 거의 전무한 상황이다.

이를 위해서 데이터 기반 비즈니스 모델이 성장할 수 있는 상장사와의 협업 지원 및 후속 투자 단계의 투자사의 지원을 통해서 국내 기업이 성장할 수 있는 생태계적 관점의 관심과 정책 지원이 필요하다.

참고문헌

Organization for Economic (2015), *Data-Driven InnovationBig Data for Growth and Well-Being*, OECD

Andy Neely (2014), 「*Big Data for Big Business?, A Taxonomy of Data-driven Business Models used by Start-up Firms*」, Cambridge Service Alliance.

Data-Driven Business Models — Cambridge Service Alliance

Tracxn

Crunchbase

AngelList

1. 본 보고서는 미래창조과학부의 출연금으로 수행한 DB산업 육성 사업의 결과입니다.
2. 본 보고서의 내용을 발표할 때에는 반드시 미래창조과학부 DB산업 육성 사업의 결과임을 밝혀야 합니다.
3. 본 보고서 내용과 관련한 문의는 한국데이터진흥원 정책기획실 (02-3708-5362)로 연락해 주시기 바랍니다.