

한중일 vs 미국 비교 분석

Kaggle 유저 설문을 통한
IT산업 발전도 비교



Our Team

김형준



PPT
데이터 분석



강종혁



데이터 분석
시각화



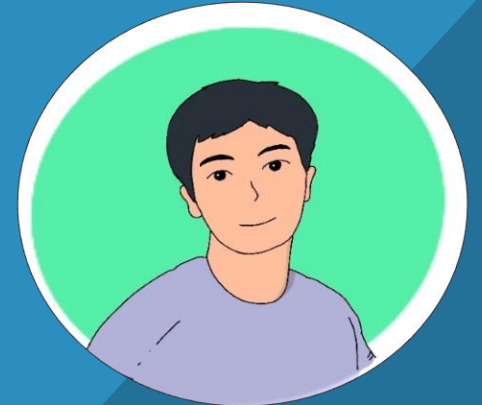
윤정호



데이터 전처리
시각화



김한빈



PPT
데이터 분석



목차

1 서론

- Kaggle?
- 주제 선정 배경
- 진행 순서

2 데이터 분석

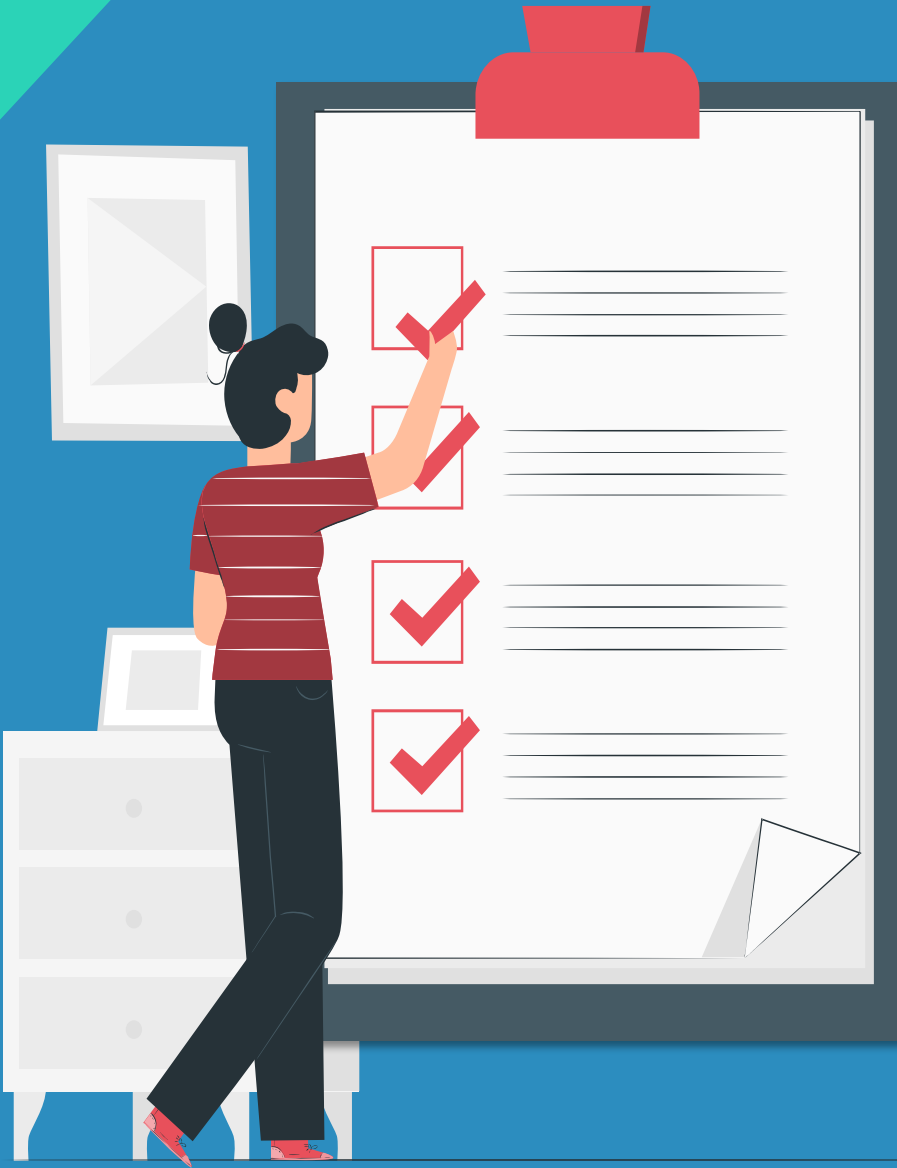
- 사용 데이터 세팅
- 데이터 기초 검증
- 주요 분석
- 결과

3 추가 분석

- 미국
- 중국
- 일본
- 한국
- 결과

4 결론

- 의의 및 한계점
- Kaggle 결과
- 레퍼런스



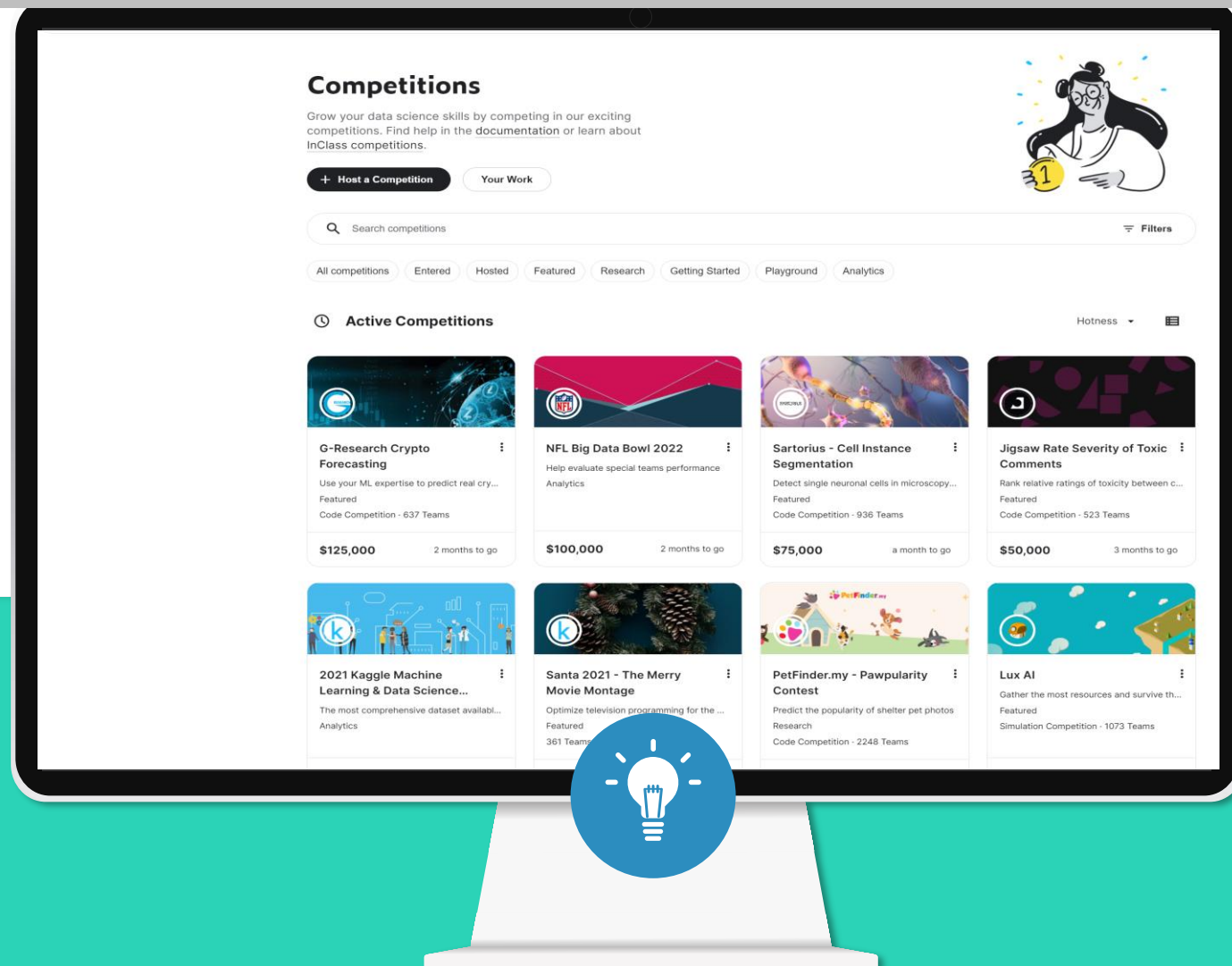
캐글(Kaggle)은 2010년 설립된 예측모델 및 분석 대회 플랫폼

5,000,000

전세계 등록 이용자

1st

수 많은 데이터와
예측모델 학습 가능



미국은 현재 빅데이터 산업 분야의 선두를 달리고 있다.



Kaggler들의 설문 조사를 바탕으로 만들어진 데이터를 분석함으로써 국가들 간의 격차 발전 수준 차이를 알 수 있다.



한국, 중국, 일본은 빅데이터 산업 분야를 국가산업으로 선정하며 발전의 박차를 가하고 있다.



해당 데이터의 특정 요소를 분석함으로써 간접적인 지표로 삼을 수 있다.



사용 데이터 세팅

2021 전체 유저 응답 데이터

수집



한국, 미국, 중국 일본 응답 추출



데이터 전처리

Q1. 연령대

Q3. 국가명

Q4. 유저 학위

Q5. 유저 직업

Q6. 프로그래밍 경력

Q25. 유저 급여수준

데이터 기초 검정

기초 검정

각 국가별
데이터
프레임

주요 항목 별 데이터 비율 추출



결과물 각 국가별 그룹화

주요 분석

핵심 요소 그룹화

연령-급여

학위-급여

경력-급여



각 요인 별 영향 분석

결과

2021 Kaggle Machine Learning & Data Science Survey

본 대회는 전세계 kaggler들의 응답을 기반으로 작성된 데이터로, kaggler들의 많은 정보를 담고 있다. 해당 데이터 중 보다 정확한 분석을 위해 한국, 미국, 중국, 일본의 데이터만을 수집하여 빅데이터 산업 분야와 연결성이 높은 문항만을 정리해보았다.



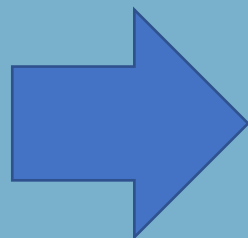
데이터 분석

#사용 데이터 세팅 #데이터 전처리 #시각화 작업 #결과

이용 데이터 설명

2021 Kaggle Machine Learning & Data Science Survey

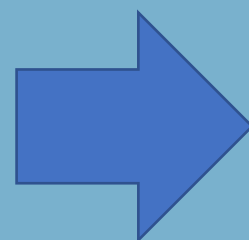
Time from Q1	Q2	Q3	Q4	Q5	Q6	Q7_Part_1	Q7_Part_2	Q7_Part_3	Q7_Part_4	Q7_Part_5	Q7_Part_6	Q7_Part_7
Duration (What is your current duration in years)	What is your gender	In which country are you currently living	What is the highest level of education that you have completed	Select the programming language(s) you use most frequently	For how long have you been using the language(s) you selected	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently	What programming language(s) do you use most frequently
910 50-54	Man	India	Bachelor's	Other	5-10 years	Python	R		SQL	C	C++	Java
784 50-54	Man	Indonesia	Master's	Program/degree	20+ years				SQL	C	C++	Java
924 22-24	Man	Pakistan	Master's	Software Engineering	1-3 years	Python					C++	Java
575 45-49	Man	Mexico	Doctoral/degree	Research	20+ years	Python						
781 45-49	Man	India	Doctoral/degree	Other	< 1 years	Python					C	
1020 25-29	Woman	India	I prefer not to answer	Currently	< 1 years	Python						
141 18-21	Woman	India	Some college	Student	1-3 years						C++	Java
484 30-34	Man	India	Bachelor's	Data Science	5-10 years	Python						JavaScript
1744 22-24	Man	Russia	Bachelor's	Currently	3-5 years	Python			SQL			
655 30-34	Man	Turkey	I prefer not to answer	Other	1-3 years	Python			SQL			
1777 40-44	Man	Australia	Doctoral/degree	Other	1-3 years	Python	R		SQL			
3081 18-21	Woman	India	Master's	Student	< 1 years	Python	R				C++	
1922 18-21	Woman	India	Master's	Student	< 1 years	Python						
852 45-49	Man	Nigeria	Master's	Program/degree	5-10 years	Python			SQL			
838 22-24	Man	Nigeria	Bachelor's	Other	< 1 years	Python						
563 35-39	Man	Greece	Doctoral/degree	Research	10-20 years	Python					C	C++
1315 50-54	Man	Belgium	Bachelor's	Data Analyst	20+ years	Python			SQL			
479 18-21	Man	Pakistan	Bachelor's	Data Science	1-3 years	Python			SQL			
249 22-24	Man	Japan	Master's	Software Engineering	3-5 years	Python			SQL	C		Java
650 30-34	Man	Egypt	Bachelor's	Other	< 1 years							JavaScript
1461 70+	Man	Singapore	Bachelor's	Other	< 1 years	Python						
551 25-29	Woman	Turkey	Bachelor's	Data Science	3-5 years	Python	R		SQL		C++	
258 30-34	Man	Indonesia	Master's	Student	1-3 years		R					
773 35-39	Man	Brazil	Master's	Machine Learning	20+ years	Python			SQL		C++	
630 18-21	Man	India	Bachelor's	Student	1-3 years	Python	R		SQL	C	C++	
2467 25-29	Woman	Poland	Master's	Machine Learning	3-5 years	Python					C++	
7750 45-49	Man	Brazil	Doctoral/degree	Research	< 1 years	Python	R					
607 22-24	Man	China	Master's	Student	3-5 years	Python				C	C++	
525 22-24	Woman	Iran, Islamic Republic of	Bachelor's	Data Science	3-5 years	Python						
501 18-21	Man	India	Bachelor's	Student	< 1 years	Python						Java
415 22-24	Man	India	Bachelor's	Other	1-3 years	Python						
317 30-34	Man	India	Master's	Data Science	5-10 years	Python	R		SQL			
171 22-24	Nonbinary	United States of America	Some college	Data Analyst	< 1 years		R					
744 30-34	Woman	Egypt	Bachelor's	Data Analyst	3-5 years	Python						
668 22-24	Man	India	Bachelor's	Student	1-3 years	Python				C	C++	Java
460 30-34	Woman	Brazil	Bachelor's	Currently	< 1 years	Python	R		SQL			JavaScript
244 25-29	Woman	Egypt	Master's	Currently	< 1 years	Python			SQL			
901 30-34	Man	Brazil	Bachelor's	Student	3-5 years	Python					Java	JavaScript
625 22-24	Man	Japan	No formal education	Software Engineering	3-5 years	Python			SQL		Java	
1239 18-21	Man	India	Bachelor's	Student	< 1 years	Python					C++	
229998 25-29	Man	Brazil	Master's	Data Science	5-10 years	Python			SQL			
312 25-29	Man	Italy	Master's	Other	1-3 years	Python						
976 55-59	Man	United States of America	Master's	Software Engineering	10-20 years	Python			SQL		Java	
1890 30-34	Woman	Viet Nam	Master's	Machine Learning	1-3 years	Python						
357 18-21	Man	United States of America	Some college	Student	1-3 years	Python				C	C++	
660 40-44	Man	Israel	Master's	Data Science	3-5 years	Python					C++	
1524 22-24	Man	China	Some college	Data Analyst	< 1 years	Python			SQL	C	C++	Java
453 18-21	Woman	Egypt	Bachelor's	Student	1-3 years	Python	R					JavaScript
793 25-29	Man	India	Bachelor's	Machine Learning	3-5 years	Python			SQL			Java
557 30-34	Woman	United States of America	Master's	Data Science	5-10 years	Python			SQL			



국가별(한,중,일,미)응답자 분류

Q1	Q2	Q3	Q4	Q5	Q6	Q7_Part_1	Q7_Part_2	Q7_Part_3
22-24	Man	China	Master's	Student	3-5 years	Python		
22-24	Nonbinary	United States of America	Some college	Data Analyst	< 1 years		R	
55-59	Man	United States of America	Master's	Software Engineer	10-20 years	Python		SQL
18-21	Man	United States of America	Some college	Student	1-3 years	Python		
22-24	Man	China	Some college	Data Analyst	< 1 years	Python		SQL
30-34	Woman	United States of America	Master's	Data Scientist	5-10 years	Python		SQL
40-44	Man	United States of America	Bachelor's	Research Scientist	< 1 years			
30-34	prefer not to answer	China	I prefer not to answer	Data Analyst	< 1 years	Python		
50-54	Man	United States of America	Master's	Data Engineer	20+ years	Python		
22-24	Man	China	Master's	Business Analyst	< 1 years	Python		
35-39	Man	United States of America	Professional	Research Scientist	10-20 years	Python	R	SQL
25-29	Man	United States of America	Master's	Data Scientist	1-3 years	Python		SQL
55-59	Man	United States of America	Bachelor's	Software Engineer	20+ years	Python		SQL
30-34	Man	United States of America	Master's	Business Analyst	1-3 years	Python		SQL
55-59	Man	United States of America	Bachelor's	Other	1-3 years	Python		SQL
40-44	Woman	United States of America	Master's	Other	5-10 years	Python		
55-59	Man	United States of America	Master's	Data Scientist	20+ years	Python	R	SQL
70+	Man	United States of America	Master's	Other	3-5 years	Python		
22-24	Man	United States of America	Master's	Student	5-10 years	Python		
70+	Woman	United States of America	Master's	Data Analyst	20+ years	Python		SQL
22-24	Man	United States of America	Bachelor's	currently not employed	1-3 years	Python	R	SQL
60-69	Man	United States of America	Master's	Product Manager	20+ years	Python		SQL
30-34	Man	United States of America	Master's	Data Scientist	10-20 years	Python	R	SQL
30-34	Man	United States of America	Master's	Business Analyst	we never written code			
55-59	Man	United States of America	Master's	Data Scientist	20+ years	Python		SQL
30-34	Man	United States of America	Some college	Other	< 1 years			
30-34	Man	United States of America	Master's	Data Scientist	1-3 years	Python	R	
60-69	Man	United States of America	Master's	Student	10-20 years		R	
22-24	Man	China	Master's	Data Scientist	1-3 years	Python		
25-29	Man	United States of America	Bachelor's	Student	1-3 years	Python		

Q1	Q2	Q3	Q4	Q5	Q6	Q7_Part_1	Q7_Part_2	Q7_Part_3
22-24	Man	China	Master's degree	Student	3-5 years	Python		
22-24	Nonbinary	United States of America	Some college	Data Analyst	< 1 years		R	
55-59	Man	United States of America	Master's degree	Software Engineer	10-20 years	Python		SQL
18-21	Man	United States of America	Some college	Student	1-3 years	Python		
22-24	Man	China	Some college	Data Analyst	< 1 years	Python		SQL
30-34	Woman	United States of America	Master's degree	Data Scientist	5-10 years	Python		SQL
40-44	Man	United States of America	Bachelor's degree	Research Scientist	< 1 years			
30-34	Gender not listed	China	I prefer not to answer	Data Analyst	< 1 years	Python		
50-54	Man	United States of America	Master's degree	Data Engineer	20+ years	Python		
22-24	Man	China	Master's degree	Business Analyst	< 1 years	Python		
35-39	Man	United States of America	Professional degree	Research Scientist	10-20 years	Python	R	SQL
25-29	Man	United States of America	Master's degree	Data Scientist	1-3 years	Python		SQL
55-59	Man	United States of America	Bachelor's degree	Software Engineer	20+ years	Python		SQL
30-34	Man	United States of America	Master's degree	Business Analyst	1-3 years	Python		SQL
55-59	Man	United States of America	Bachelor's degree	Other	1-3 years	Python		SQL
40-44	Woman	United States of America	Master's degree	Other	5-10 years	Python		
55-59	Man	United States of America	Master's degree	Data Scientist	20+ years	Python	R	SQL
70+	Man	United States of America	Master's degree	Other	3-5 years	Python		
22-24	Man	United States of America	Master's degree	Student	5-10 years	Python		
70+	Woman	United States of America	Master's degree	Data Analyst	20+ years	Python		SQL
22-24	Man	United States of America	Bachelor's degree	Currently not employed	1-3 years	Python	R	SQL
60-69	Man	United States of America	Master's degree	Product Manager	20+ years	Python		SQL
30-34	Man	United States of America	Master's degree	Data Scientist	10-20 years	Python	R	SQL
30-34	Man	United States of America	Master's degree	Business Analyst	Have never written code			
55-59	Man	United States of America	Master's degree	Data Scientist	20+ years	Python		SQL
30-34	Man	United States of America	Some college	Other	< 1 years			
30-34	Man	United States of America	Master's degree	Data Scientist	1-3 years	Python	R	
60-69	Man	United States of America	Master's degree	Student	10-20 years		R	
22-24	Man	China	Master's degree	Data Scientist				
25-29	Man	United States of America	Bachelor's degree	Student				

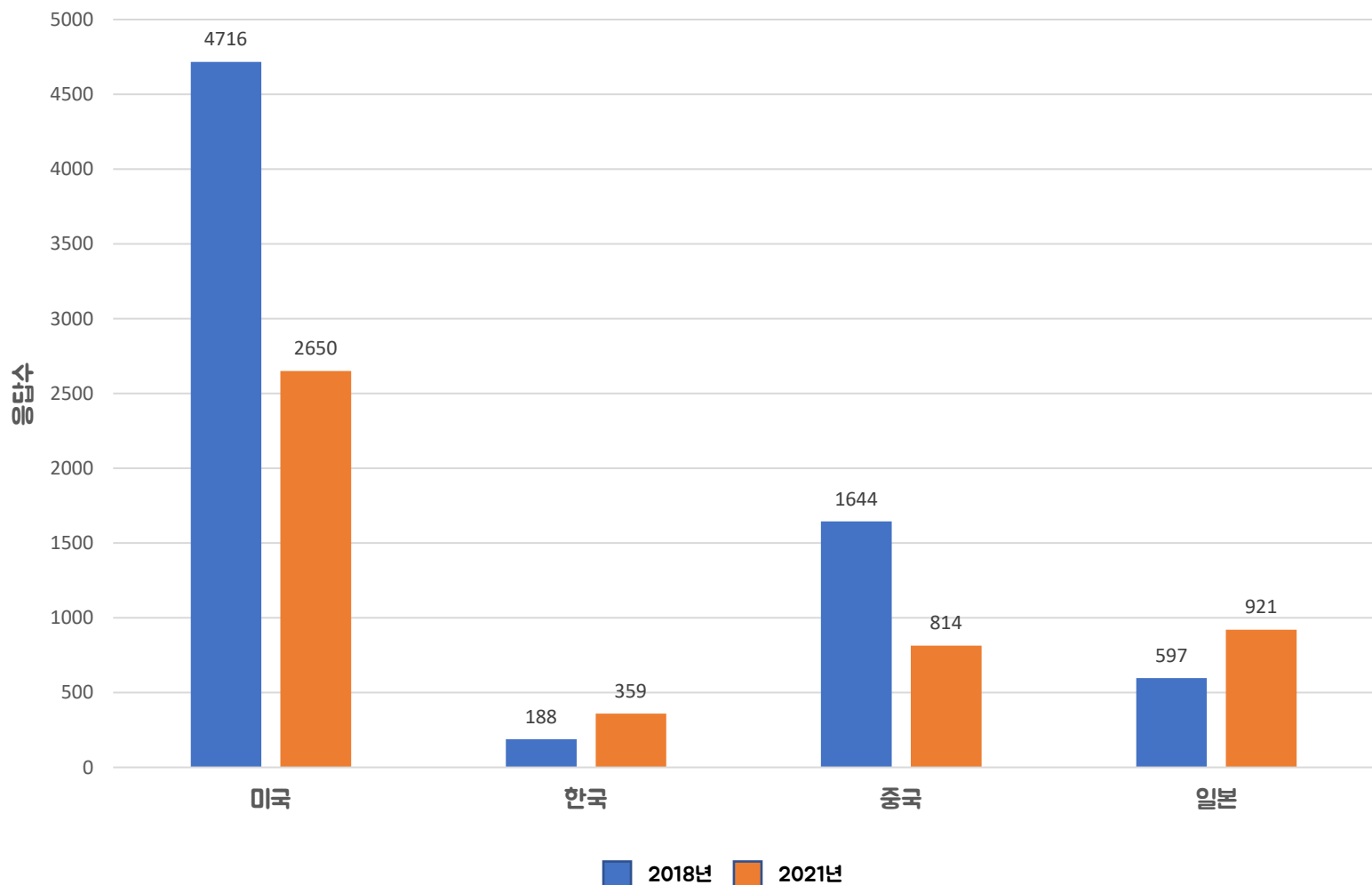


Column_1	나이	Q2	Q3	학위수준	직군	경력	연봉(\$)
2180	45	Man	USA	4	Data Scientist	29	399845
2226	47	Man	USA	5	Software Engineer	21	396529
1338	58	Man	USA	4	Data Scientist	22	396045
2582	47	Man	USA	5	Research Scientist	23	395305
1054	53	Man	USA	5	Product Manager	27	392155
614	66	Man	USA	5	Other	25	391771
441	50	Man	USA	5	Research Scientist	21	391707
2651	42	Man	USA	3	Product Manager	3	387560
2207	26	Man	USA	3	Data Engineer	9	387195
7	52	Man	USA	4	Data Engineer	28	377002
2208	35	Man	USA	5	Other	17	371169
1212	57	Man	USA	5	Other	20	370637
2695	48	Man	USA	4	Data Scientist	24	368608
2214	44	Man	USA	3	Product Manager	15	365514
229	44	Man	USA	3	Other	16	357184
1741	39	Man	USA	4	Machine Learning Engineer	17	354513
2023	57	Man	USA	5	Data Scientist	15	353680
1015	46	Man	USA	3	Software Engineer	29	342504
937	64	Man	USA	5	Research Scientist	24	334620
1063	43	Man	USA	5	Data Scientist	13	333553
1276	35	Man	USA	4	Other	3	330671
1488	35	Man	USA	4	Machine Learning Engineer	7	330440
1079	35	Man	USA	4	Data Scientist	13	329938
814	41	Man	USA	4	Developer	9	328589
1937	69	Man	USA	5	Research Scientist	24	324871
2120	47	Man	USA	4	Other	3	324156
78	57	Man	USA	4	Other	0	323907
2899	47	Man	USA	5	Other	16	322441
2381	44	Man	USA	5	Research Scientist	2	321611
1835	48	Man	USA	4	Product Manager	7	312759
1515	42	Man	USA	5	Software Engineer	4	309262
961	31	Man	USA	3	Software Engineer	10	308962
1265	47	Man	USA	4	Other	29	306282
311	41	Woman	USA	5	Research Scientist	17	300750
1951	42	Man	USA	4	Other	3	299482
1886	35	Man	USA	4	Machine Learning Engineer	6	295653
			USA	4	Data Scientist	4	295521
			USA	5	Machine Learning Engineer	25	295093

- 42개의 질문 중 7개의 질문으로 요약
- 각 나라별 질문 응답에 대한 전처리 작업 진행

2018년/2021년 4개국 총 응답자 수 비교

2018년/2021년 국가별 Kaggle 이용자 수

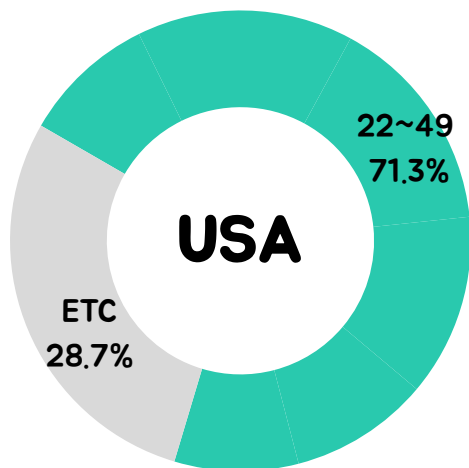


■ 2018년에 비해 2021년 미국, 중국의 경우 이용자 응답 수 가 절반수준 줄어든 것이 확인된다.

■ 반면, 한국과 일본은 유의미한 상승률을 보이고 있다.

■ 빅데이터 분야에 발전에 투자하고 있는 한국, 중국, 일본 중 중국만 줄어든 것은 외교적 요인이 작용 했을 것으로 예상된다.

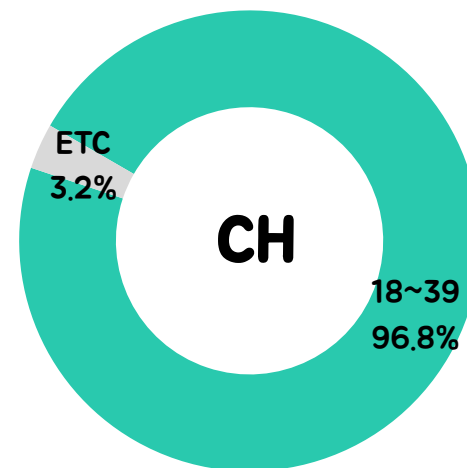
Q1. 주요 연령대 분포



■ 22~49세 비율이 전체 71.3%

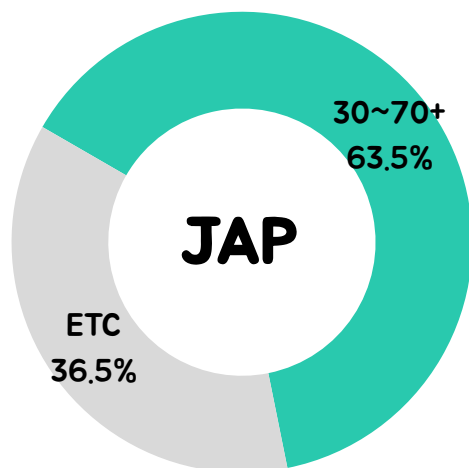
■ 직업활동 가능한 인구 많음

■ 50+ 이상 인구 수 가장 많음



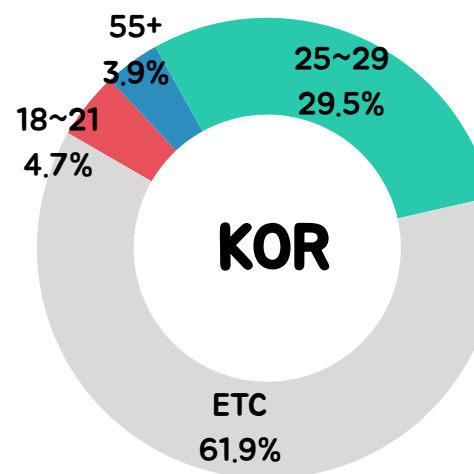
■ 18~39세 비율이 96.8%로 가장 많음.

■ 후에 분석할 자료와 연관성이 크다.



■ 22~49세 비율이 전체 74.3%

■ 단, 30세 이상 비율 63.5%로 미국이 5.6% 앞선다.

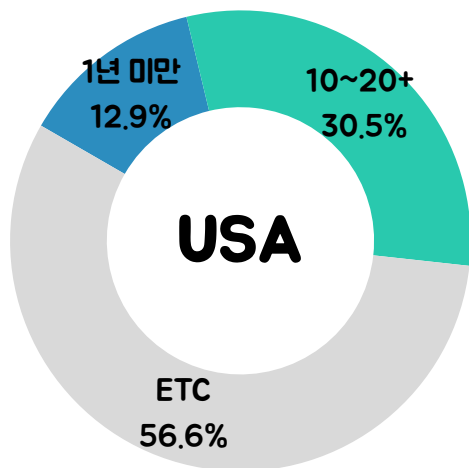


■ 25~29세 비율이 29.5%로 4개국 중 가장 많음

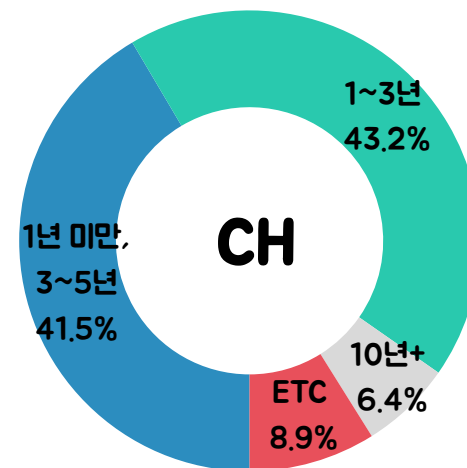
■ 반대로 55+, 21세 이하 비율은 각각 3.9%, 4.7%로 가장 적음

■ 이는 한국의 군입대 연령, 정년퇴임 연령대와 어느정도 연관이 있는 것으로 보인다.

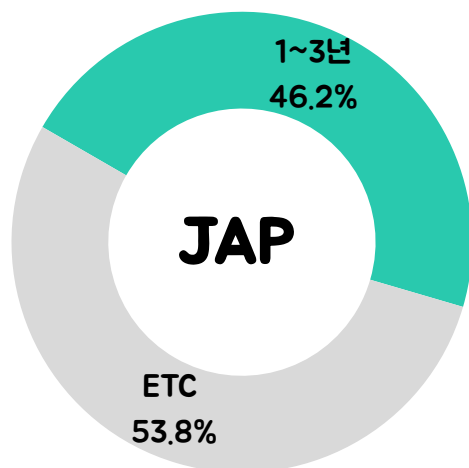
Q6. 프로그래밍 활동 기간 분포 비교



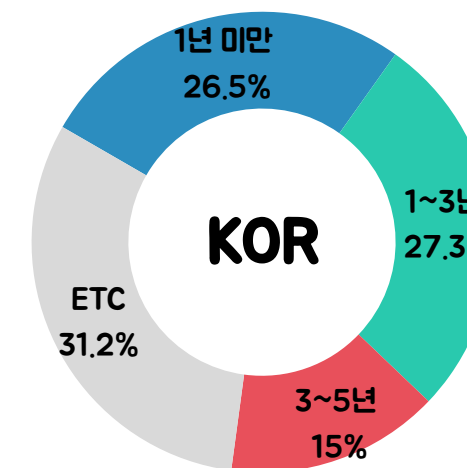
- 경력 1년 미만 비율이 12.9%로 가장 적음
- 10~20+ 년 이상의 경력 비율은 30.5%로 가장 많음
- 연령 분포와 연관성을 보인다.



- 5년 미만의 비율이 84.7%로 가장 높게 나타났다.
- 10년+ 비율은 6.4%로 가장 낮게 나타났다.
- 1~3년 경력이 43.2%로 가장 비율이 컸다.

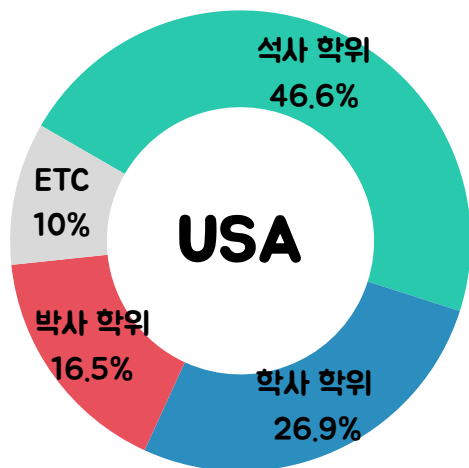


- 연령대와 마찬가지로 비슷한 분포를 보인다.
- 눈에 띄는 차이점은 1~3년 비율이 46.2%로 미국에 비해 14%정도 많았다.
- 연령대에 비해 경력의 수준은 낮은 수준의 비율이 많다.

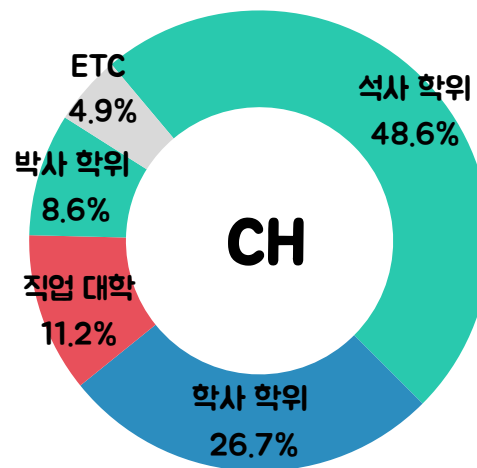


- 5년 미만 응답자 68.8% 중국에 비해 상대적으로 적었다.
- 1~3년 경력이 27.3%로 중국과 큰 차이를 보인다.

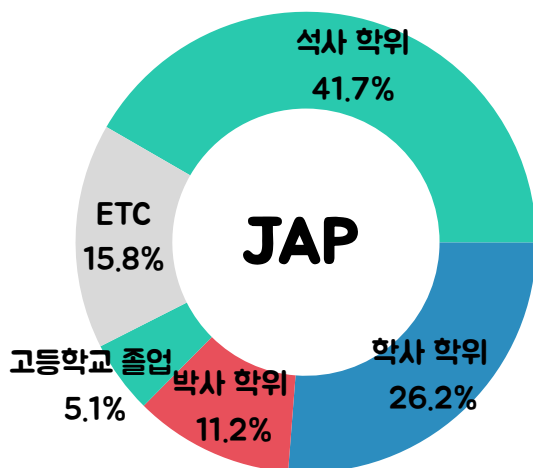
Q4. 학위 수준 분포 비교



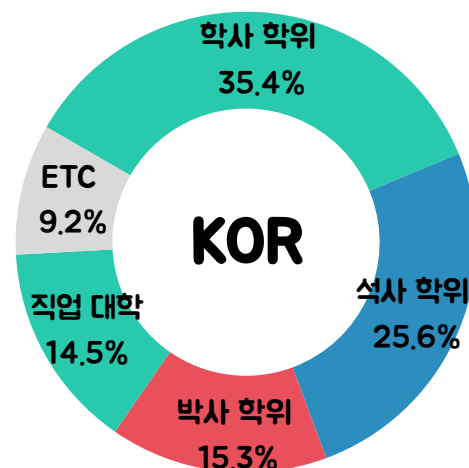
- 박사 학위 취득률이 16.5% 가장 높게 나타났다.



- 석사 학위 비율이 48.6%로 가장 높게 나타났다.
- 직업대학 응답자의 비율이 11.2%로 박사 학위 응답자 8.6%로 직업대학 응답자가 더 많은 유일한 국가이다.

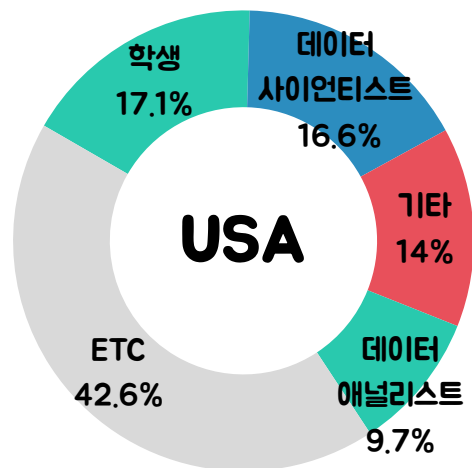


- 미국과 전반적으로 유사한 분포도를 보이고 있다.
- 고등학교 졸업만 한 응답자의 비율이 5.1%로 가장 높았다.



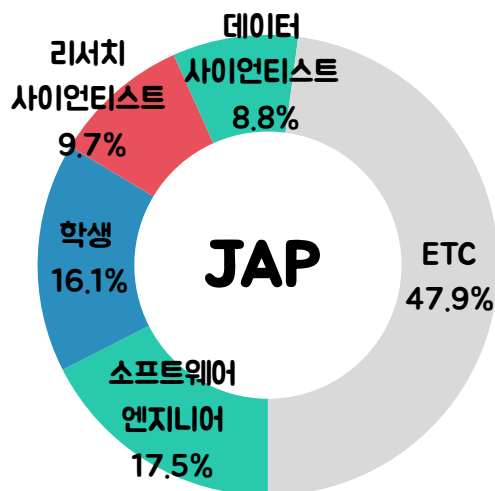
- 4개국 중 학사 학위가 35.4%로 가장 높고, 석사 비율보다 많다.
- 직업대학 응답 비율도 14.5%로 가장 높다.
- 박사 학위 응답 비율 또한 15.3%로 아시아권 3개 국 중 가장 높다.

Q5. 직군 분포 비교



■ 데이터 사이언티스트, 데이터 애널리스트의 비율이 16.6%, 9.7%로 많은 비중을 차지했다.

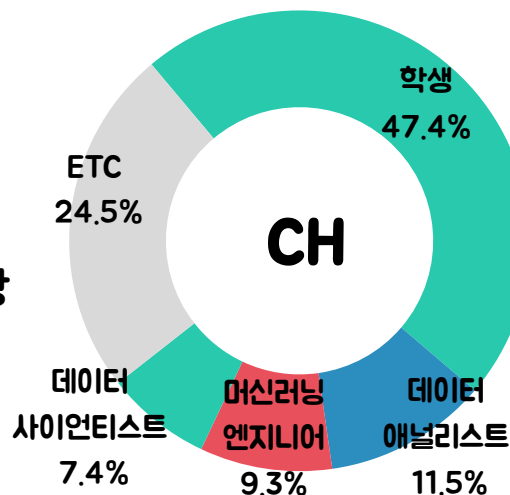
■ 기타 직업군 응답률이 14%로 가장 높게 나타났다.



■ 일본의 경우 미국과 대비되는 모습을 보였다.

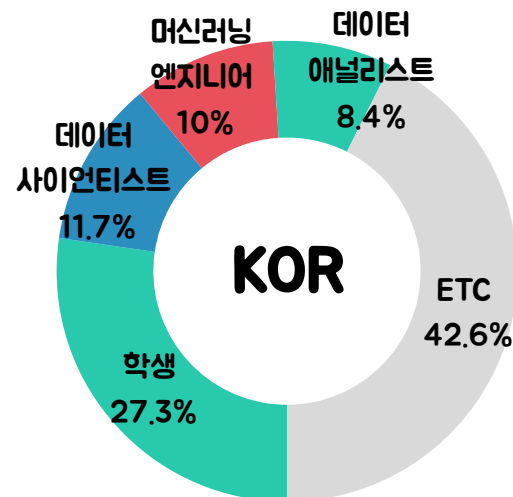
■ 빅데이터 분야 직군보다 소프트웨어 엔지니어 비율이 17.5%로 가장 높게 나타났다.

■ 빅데이터 분야 직군의 응답비율은 가장 낮게 분포되어 있었다.



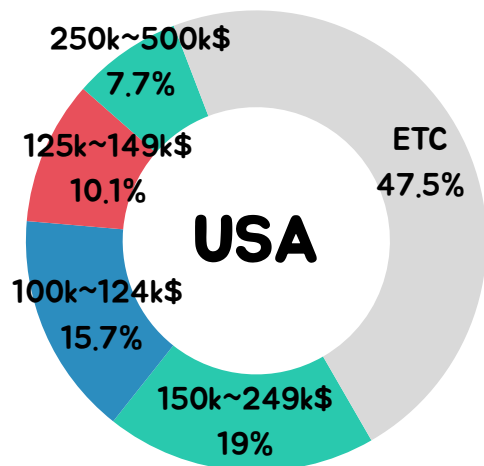
■ 학생 비율이 47.4%로 4개 국 중 가장 높은 비율을 보여준다.

■ 학생을 제외한 상위 3개 비율의 직업군이 모두 최근 각광받는 직업군이 차지하고 있다.

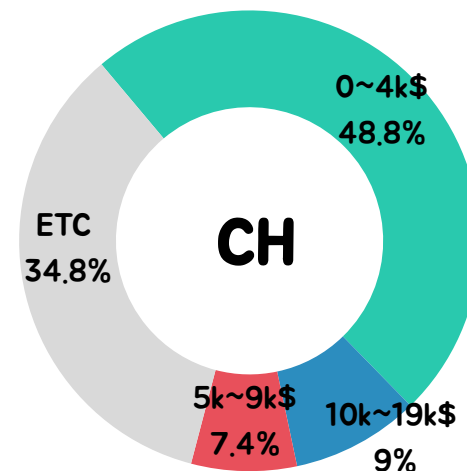


■ 한국의 경우 학생 비율에서만 중국과 차이를 보일 뿐, 큰 차이점은 나타나지 않는다.

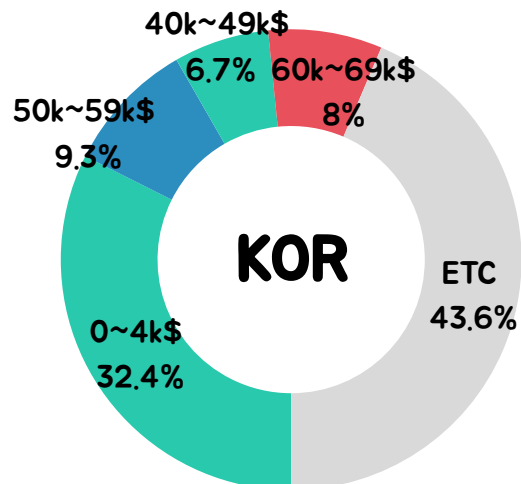
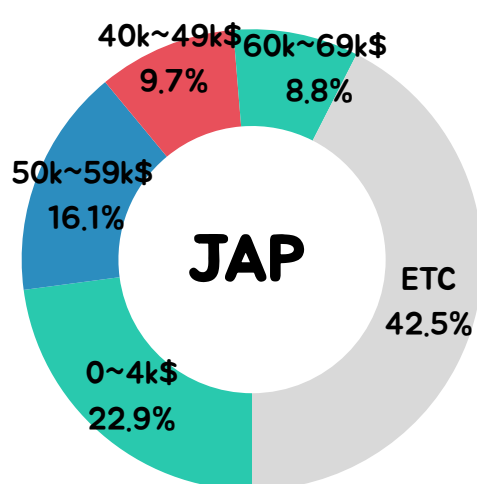
Q25. 급여 수준 분포 비교



- 상위 분포 3개 항목
100k~249k라 응답한 비율이
44.8%로 가장 높다.
- 250k~500k라 응답한 비율도
7.7%로 가장 높다.
- 전반적인 평균 급여 수준이 높은
것을 알 수 있다.



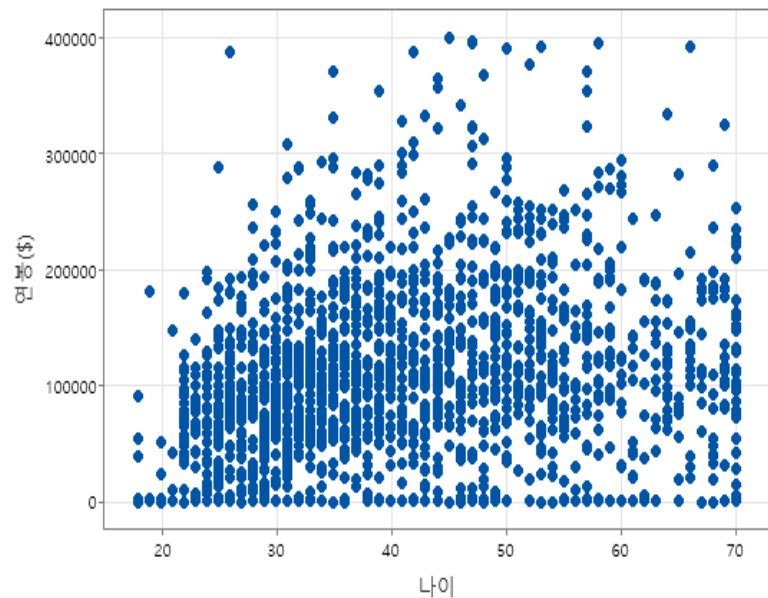
- 0~4k \$의 응답률이 48.8%로
가장 높다.
- 평균적인 급여 수준이 가장 낮다.



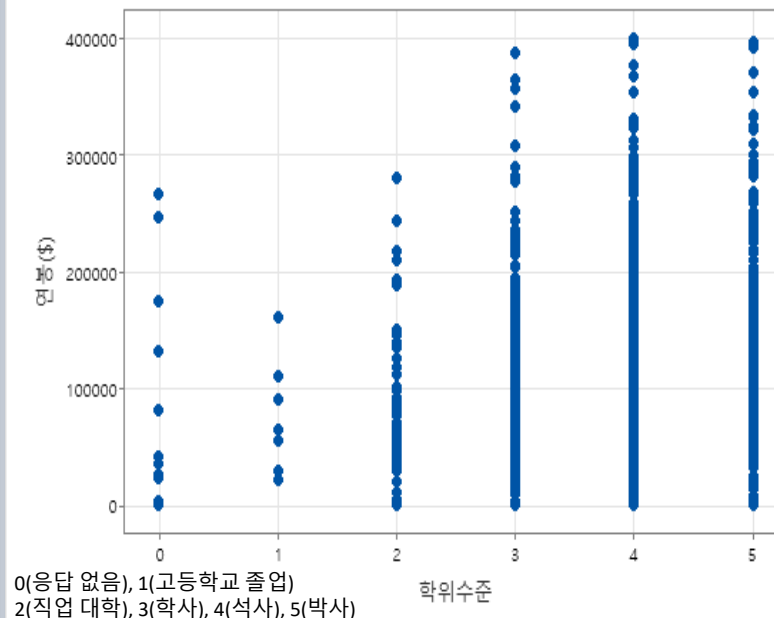
- 일본과 한국은 급여 수준에서 매우 유사한 분포를
나타냈다.
- 0~4k\$ 수준이 가장 많은 비율을 차지했다.
- 그 다음 상위 3개 항목은 40k~69k\$ 부분에
분포되어 있다.

#미국 각 요인 별 영향 분석 결과

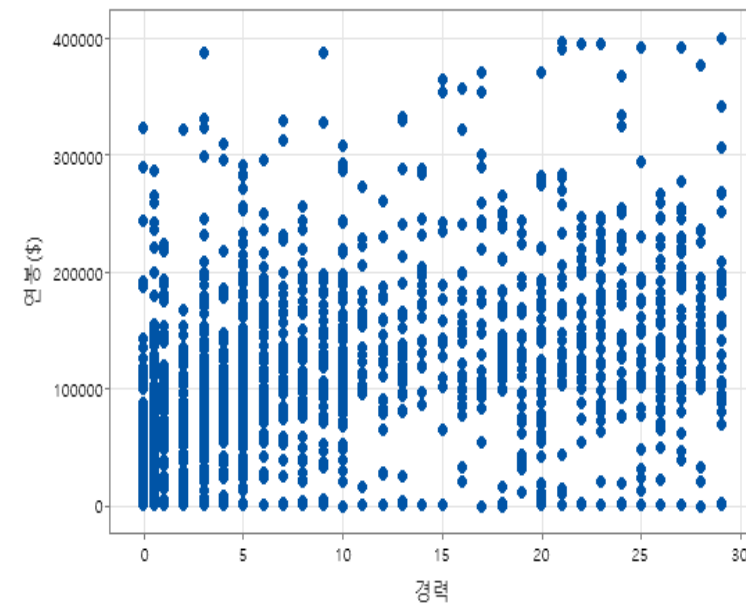
연봉(\$) 대 나이의 산점도



연봉(\$) 대 학위수준의 산점도



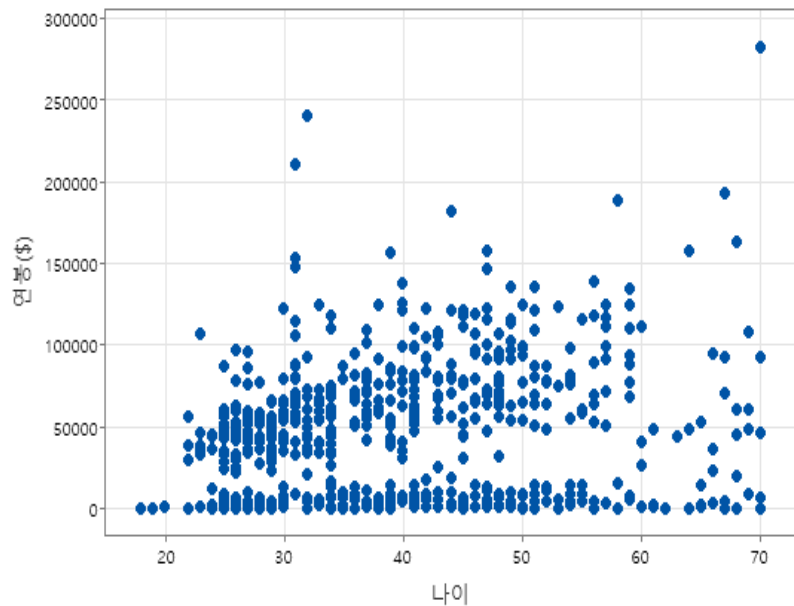
연봉(\$) 대 경력의 산점도



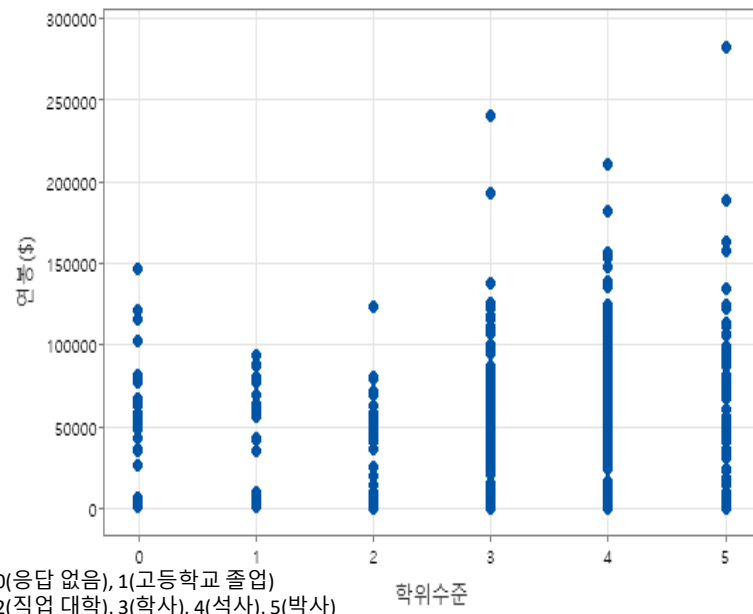
■ 미국의 경우 학위수준, 경력에 따른 급여 수준의 차이가 크게 들어난다.

#일본 각 요인 별 영향 분석 결과

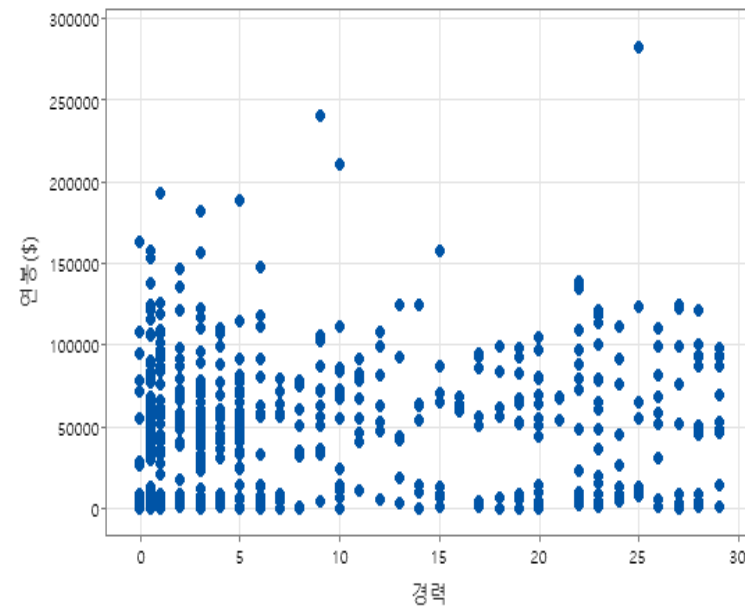
연봉(\$) 대 나이의 산점도



연봉(\$) 대 학위수준의 산점도



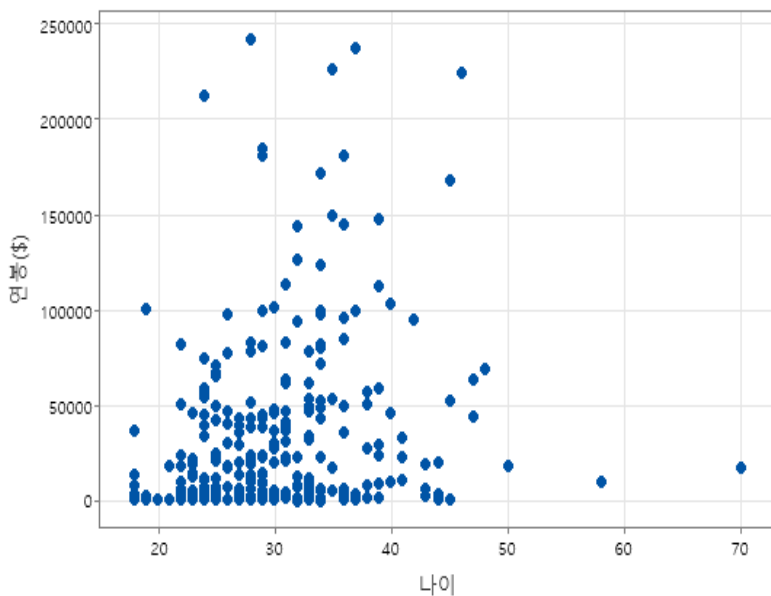
연봉(\$) 대 경력의 산점도



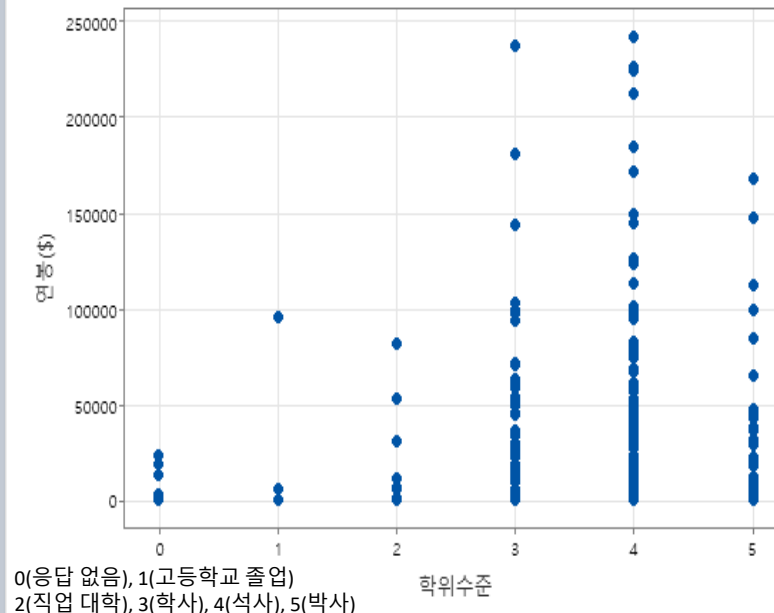
- 일본의 경우 학위수준에 따른 급여의 분포가 크지 않다.
- 연령에 따른 급여 수준의 분포가 유의미하게 증가되는 것이 확인 가능하다.

#중국 각 요인 별 영향 분석 결과

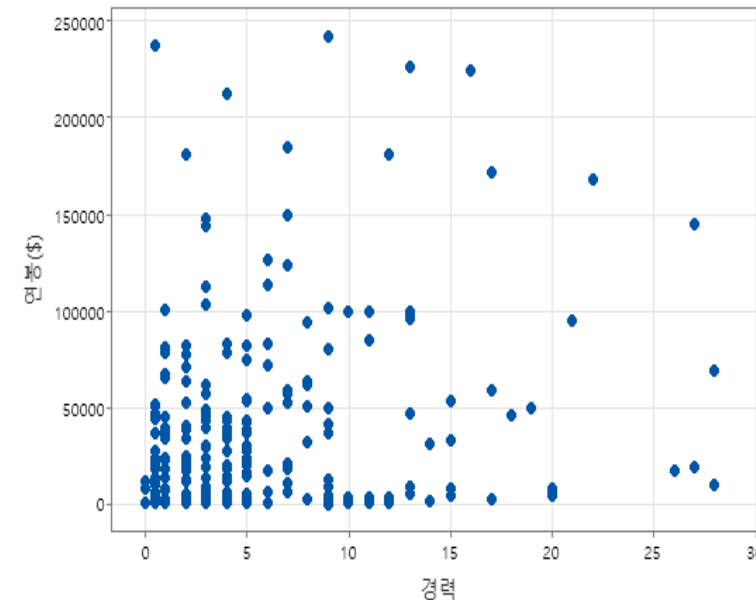
연봉(\$) 대 나이의 산점도



연봉(\$) 대 학위수준의 산점도



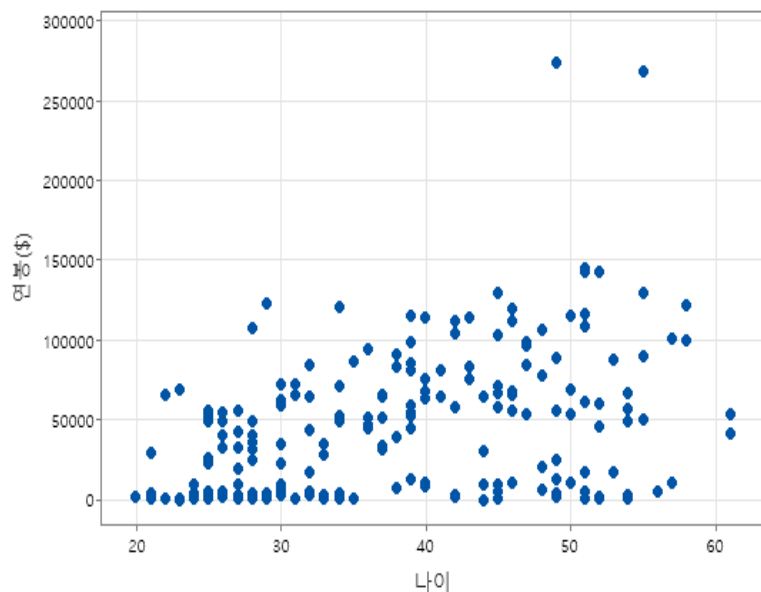
연봉(\$) 대 경력의 산점도



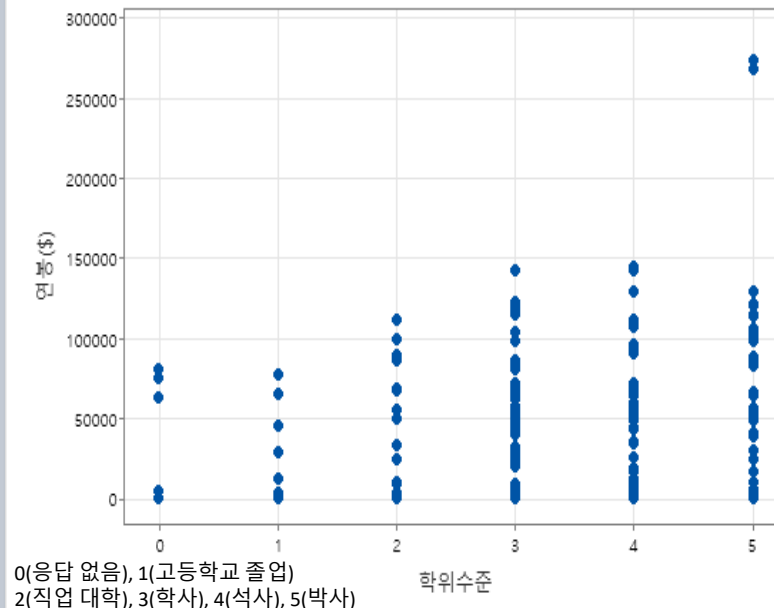
- 학위 수준에 따른 급여차이가 눈에 띄게 나타난다.
- 연령에 따른 급여차이가 미미하나, 영향이 있음을 알 수 있다.

#한국 각 요인 별 영향 분석 결과

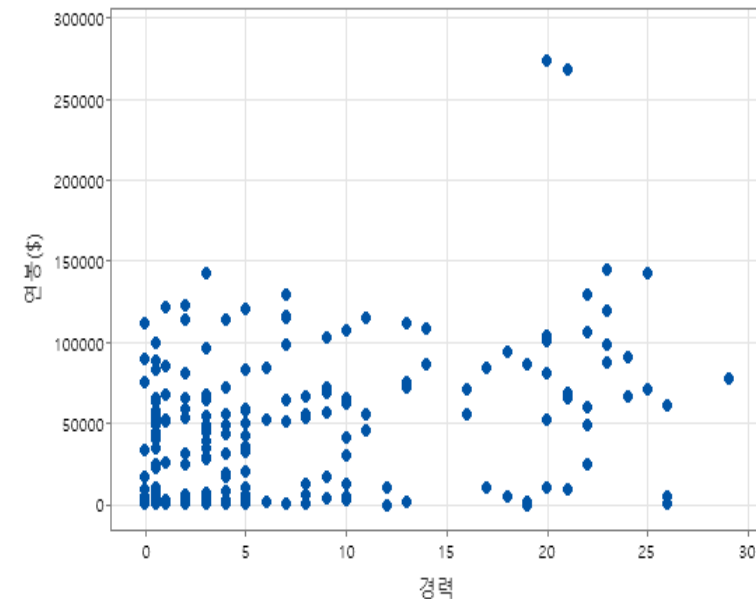
연봉(\$) 대 나이의 산점도



연봉(\$) 대 학위수준의 산점도



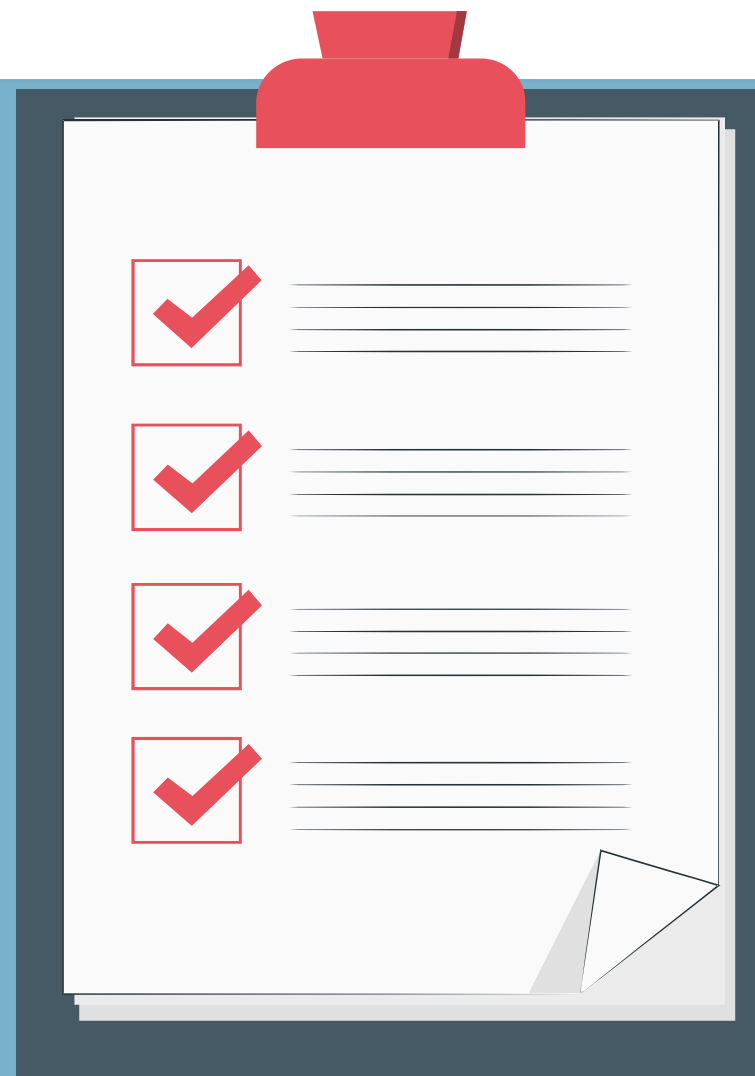
연봉(\$) 대 경력의 산점도



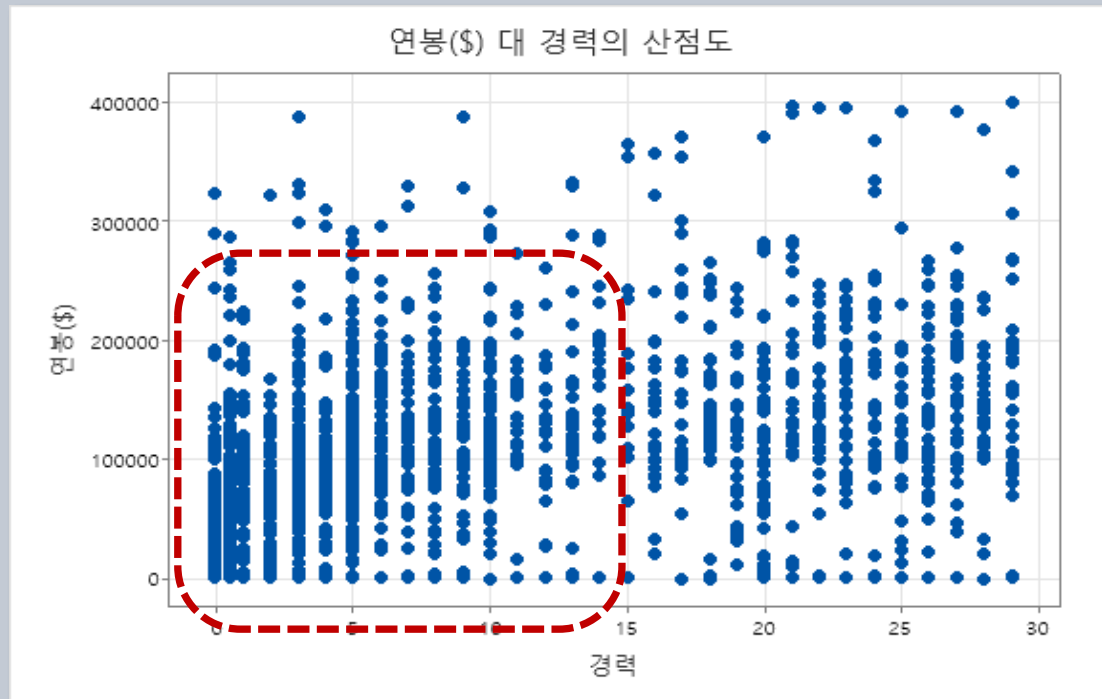
■ 한국의 경우 나이, 학위수준에 따른 급여가 증가됨을 알 수 있다.

■ 타 국가에 비해 아주 높은 수준의 급여를 받는 케이스가 매우 적다.

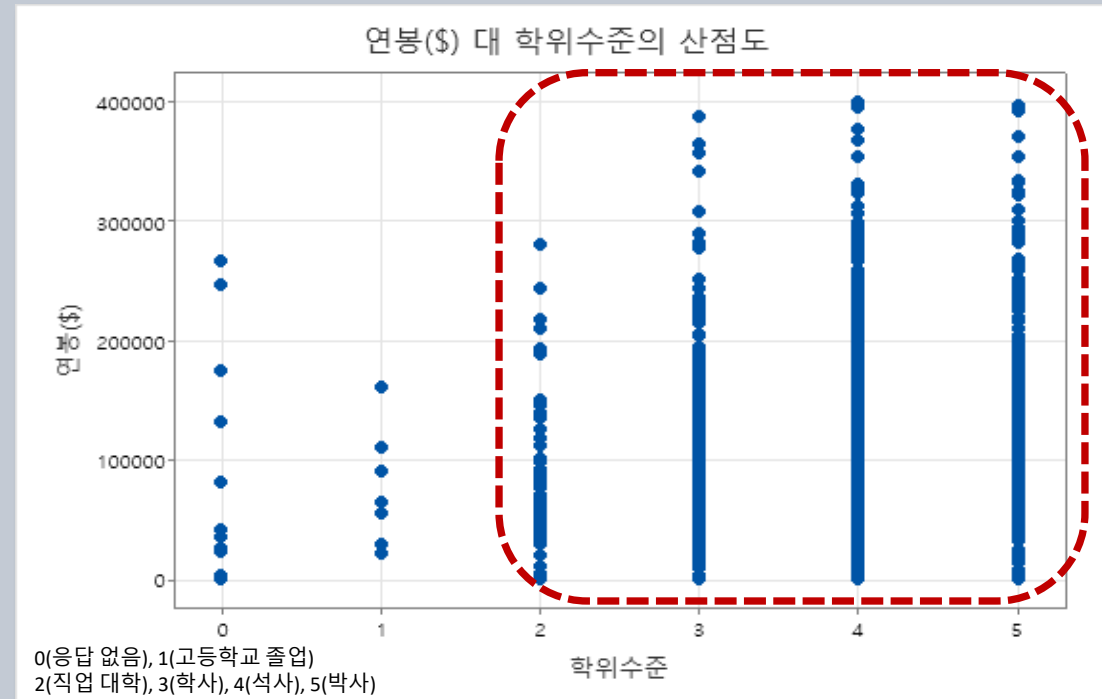
- 기초 검정 결과와 주요 분석 결과를 바탕으로 이러한 결과값에 대한 추가적인 분석과 다른 요소가 작용했는지 확인해본다.



경력, 학위수준에 따른 급여 수준분석 결과

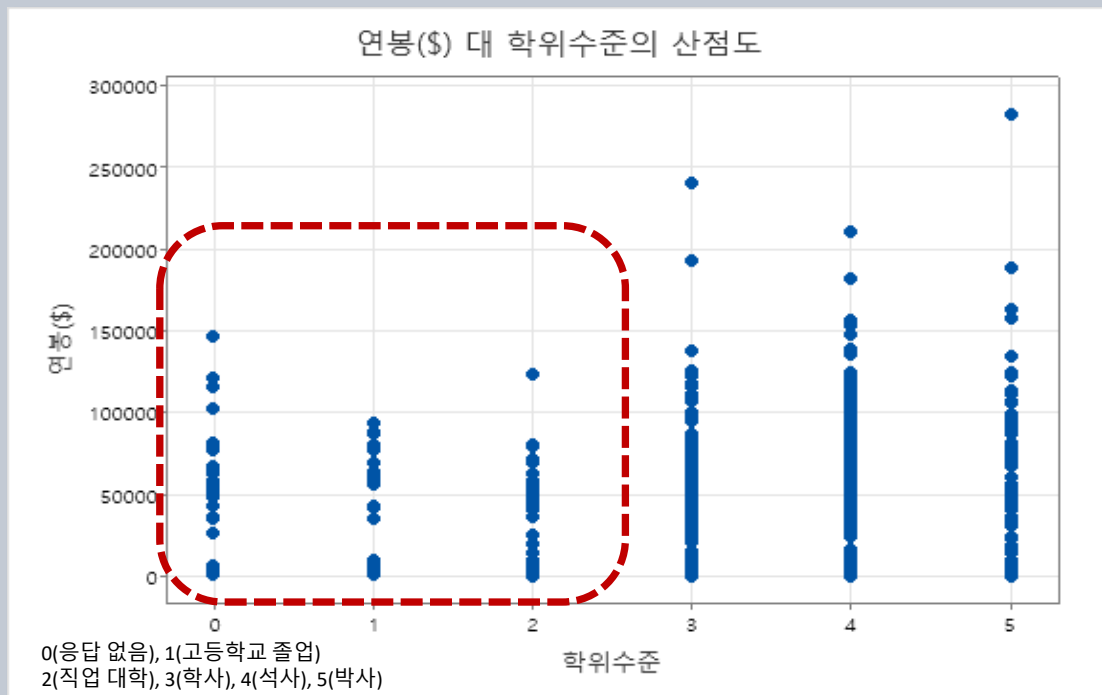


- 0~5의 경력 수준에서 100,000불 이하 분포가 눈에 띄게 많다.
- 5년 이상 경력 수준에서 100,000불 ~ 200,000불 분포가 증가되는 걸 확인 할 수 있다.

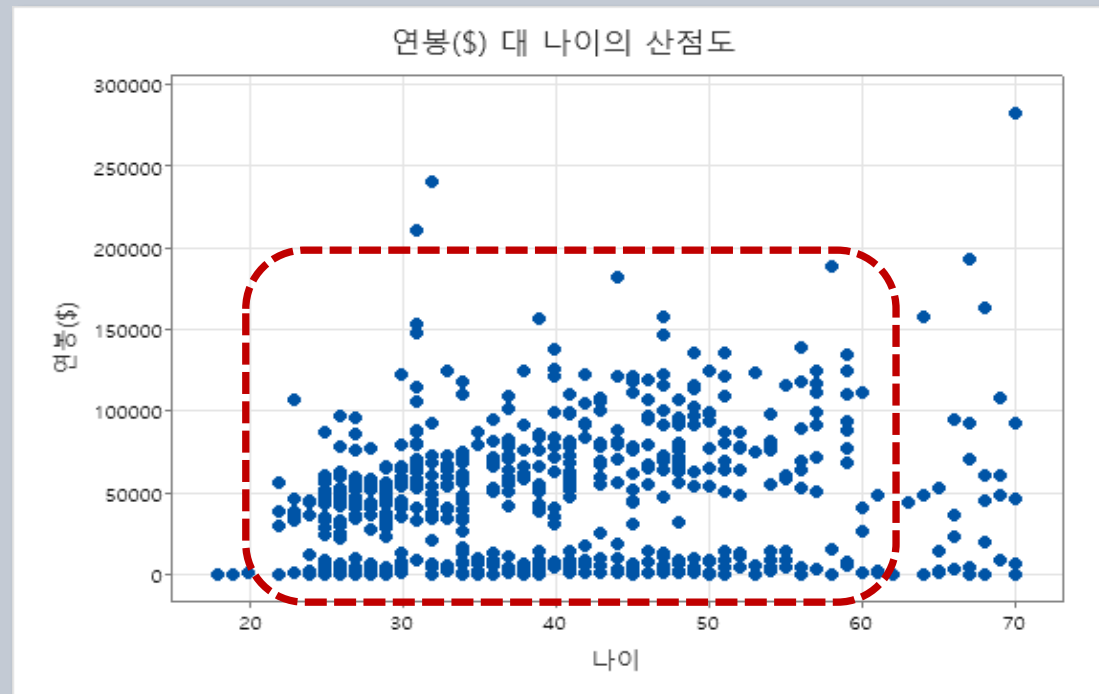


- 100,000불 이상의 연봉의 경우 3(학사) 수준에서 분포가 증가된다.
- 4(석사) 수준에서 평균적으로 가장 높은 수준의 분포도를 나타내고 있다.

연령, 학위수준에 따른 급여 수준분석 결과

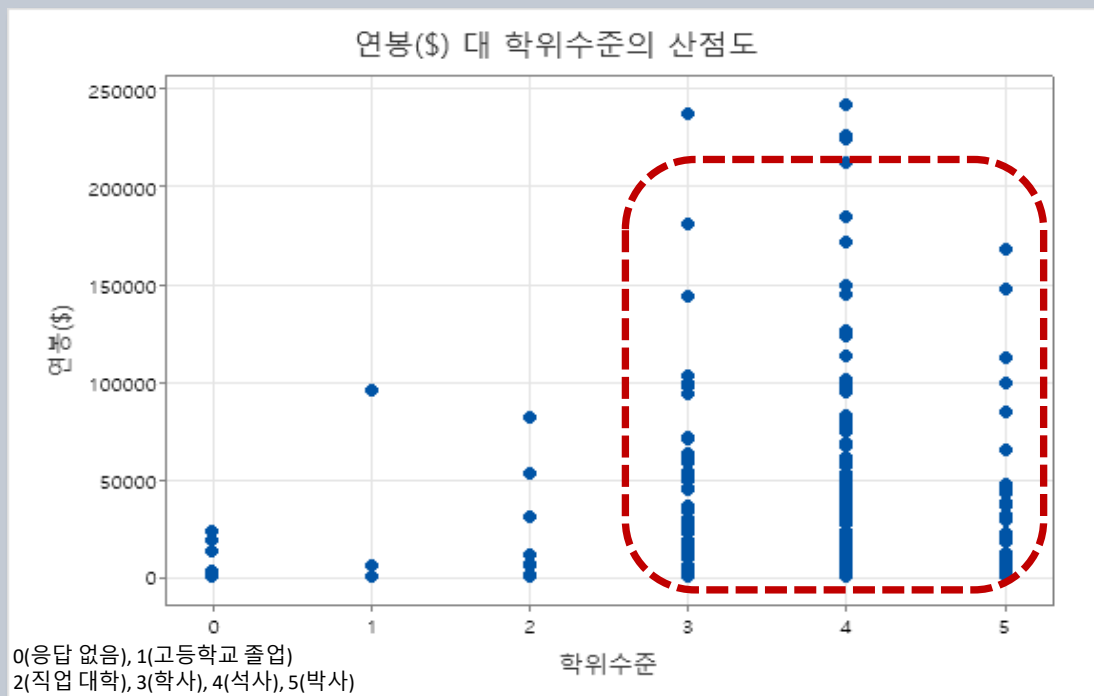


- 0(고졸), 1(직업학교)의 급여분포와 2(학사)의 급여분포가 큰 차이를 보이지 않고 있다.
- 이는 일본의 취업에 대학의 진학 여부가 큰 영향을 주지 않는 문화가 영향을 준 것으로 보인다.

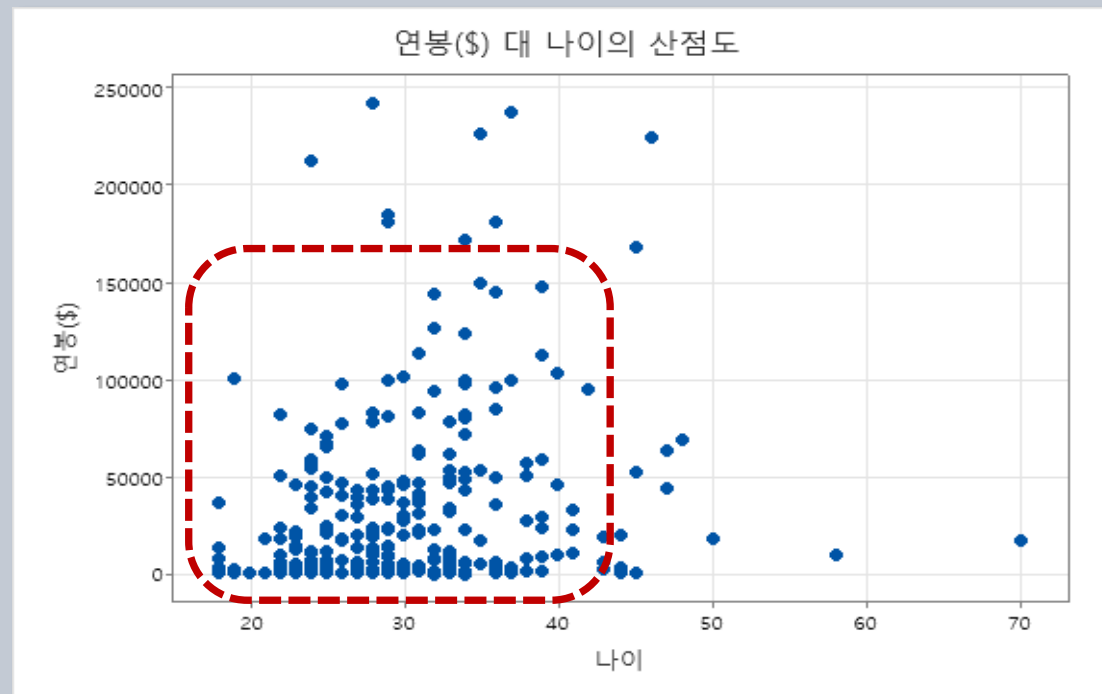


- 100,000불 이상의 연봉의 경우 3(학사) 수준에서 분포가 증가된다.
- 4(석사) 수준에서 평균적으로 가장 높은 수준의 분포도를 나타내고 있다.

연령, 학위수준에 따른 급여 수준분석 결과

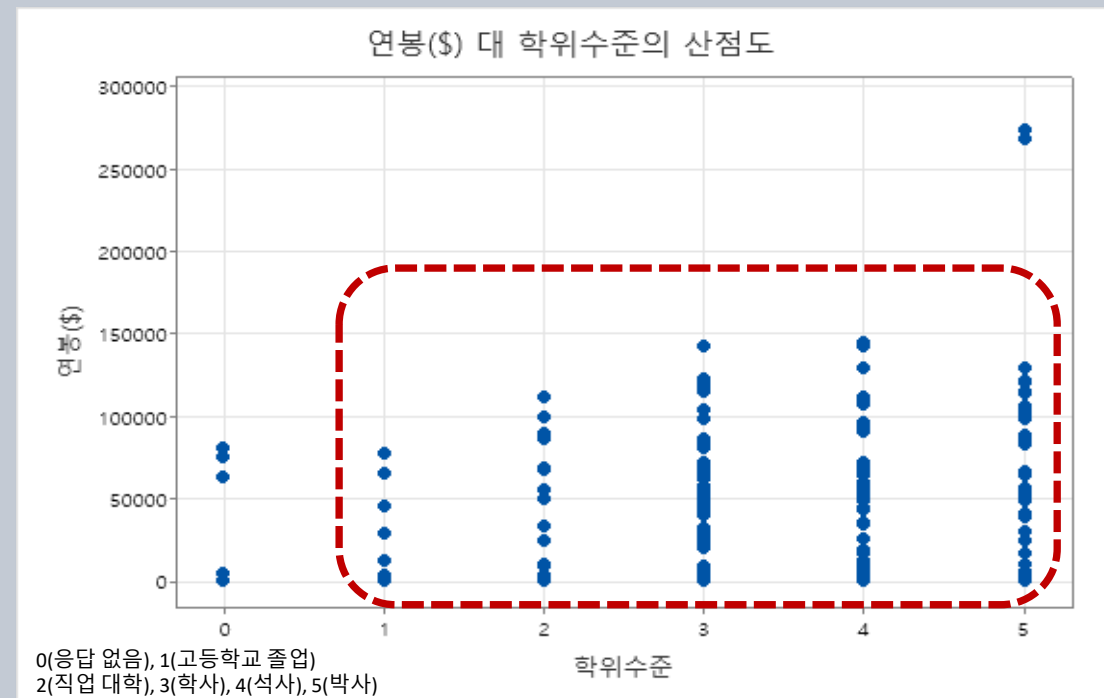
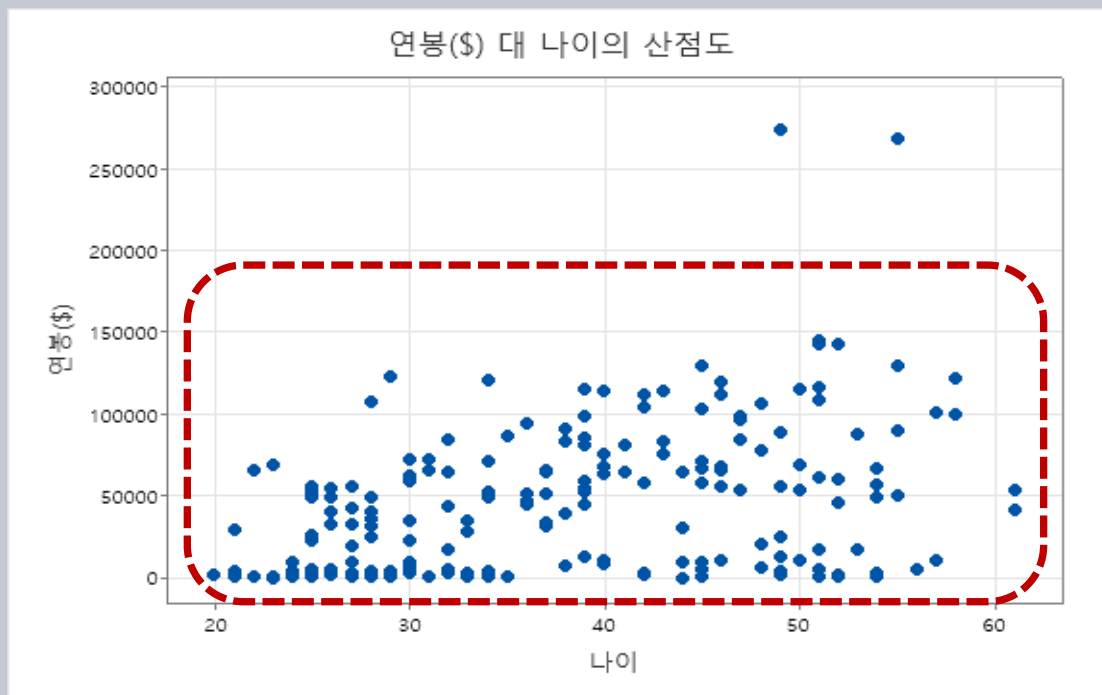


- 중국의 경우 3~5(학사~박사) 수준의 응답 비율이 극단적으로 많았다.
- 평균적인 급여 수준의 경우 4(석사)의 분포가 가장 높게 나타난 것을 확인 할 수 있다.



- 급여를 받는 대부분의 20~40대 비중이 많았다.
- 20~30대까지는 급여수준이 비슷하지만, 30대를 넘어가는 분기에서 급여수준이 상승되는 분포를 보인다.

연령, 학위수준에 따른 급여 수준분석 결과



- 한국의 경우 연령대 분포가 일정하지만, 급여 분포의 경우 0~100,000불 수준에 고르게 분포가 되어 있다.
- 한국의 프로그래밍 경력 비율 수준을 고려 했을 때, 30대 이상에 프로그래밍을 시작하는 사람의 수가 적지 않음을 알 수 있다.
- 학사와 석사의 비중이 가장 많지만, 학위 수준에 따른 급여 수준이 상승하는 분포를 보이고 있다.
- 일본과 비슷한 분포를 보이지만 높은 대학 진학률과 박사학위 취득률이 상대적으로 낮은 것이 변수로 작용한 것으로 보인다.

종합 결과

- 미국의 데이터분야의 초입하는 유저의 수가 다른 국가에 비해 적은 것으로 나타났다. 빅데이터 분야에 진출을 준비하는 학생들이 많은 공부와 경력을 쌓고 높은 평균 급여 수준과 다양한 직군의 다양성이 있는 미국으로 진출 가능성을 확인.
- 경력 5년 미만 응답자 수가 많았던 한국과 중국이지만, 한국의 경우 40,000\$ 이상의 급여를 받는다는 응답의 비율이 중국에 비해 압도적으로 많았다.
- 최고 수준의 급여분포와 한국 일본의 물가를 고려했을 때 최고 수준의 급여 비중이 미국에 비해 많이 부족하다.



의의

- 빅데이터 분야에 연관된 사람들만의 데이터를 분석함으로써 한국과 다른 국가와의 차이점을 알 수 있었다.
- 한국은 미국에 비해 경력 수준, 학사 비율에서 크게 뒤쳐지지만, 젊은 연령대의 유입이 많고, 현재 국가에서 많은 투자를 하고 있기 때문에 빠른 속도로 발전할 것을 기대 할 수 있다.
- 중국의 경우 4개국 중 가장 후발주자의 형태를 보이고 있었다. 하지만 직군의 집중도와 높은 학생 비율을 미루어 보았을 때 순식간에 성장을 해낼 수 있을 것이라 기대할 수 있다.

한계점

- 유저 설문으로 이루어진 데이터의 한계상 누락된 값들이 많고 다양한 분석이 불가능 했다.
- 특히, 중점적으로 다루었던 유저 급여수준 항목에서 설문의 의미가 모호하여 잘못된 항목을 고른 유저의 수가 많아 신뢰도와 연관성이 높지 못했다.
- 한국의 경우 전체 응답수의 1.2% 수준으로 케이스가 타 국가에 비해 절대적으로 부족해 분석의 신뢰도가 높지 않았다.
- 각 요인 별 상관수준을 비교하고 싶었지만, 명목변수로만 이루어져 분석에 한계점이 있었다.



Kaggle 대회 결과 이미지

참고 문헌

- [1] Claus O. Wike, (2020, February, 20) 데이터 시각화 교과서(데이터 분석의 본질을 살리는 그래프와 차트 제작의 기본 원리와 응용)
- [2] R&D Int, 최재경_SK경영경제연구소 수석연구원 빅데이터 분석의 국내외 활용현황과 시사점
- [3] 중앙일보, [지식리뷰] 일본에선 "대학 안 가?" 대신 "꿈이 뭐야?" 묻는다.
- [4] 한국경제, 미국은 '10억명 빅데이터'로 떼돈 버는데...규제 딜레마 빠진 일본 EU
- [5] 이데일리, [미래기술25]크기만 하면 빅데이터? 분석 활용 잘해야 빅데이터!

사용데이터

데이터명	기간	활용 목적	출처
kaggle_survey_2021_Responses	2021년 기준	각 국가별 응답 비율 확인	2021 Kaggle Machine Learning & Data Science Survey
2018_kaggle_ds_and_ml_survey_responses_only	2018년 기준	2021년 데이터와 비교 및 추세 확인	2018 Kaggle Machine Learning & Data Science Survey

분석도구



Thank You

