

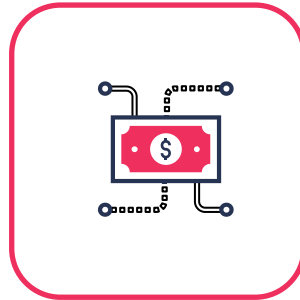
일본과 중국

캐글러의 최신 트렌드

김태용
전우정
정승원
김재원
권도근

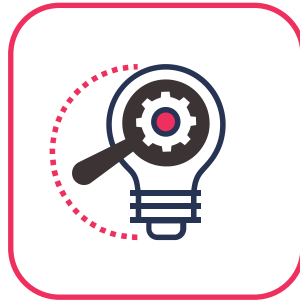


1. 선정 배경



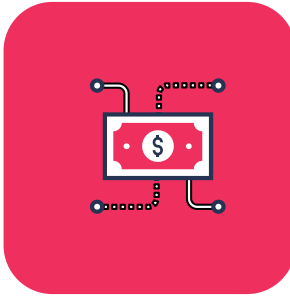
2. 데이터 분석

3. 데이터 결과



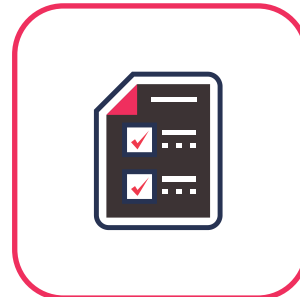
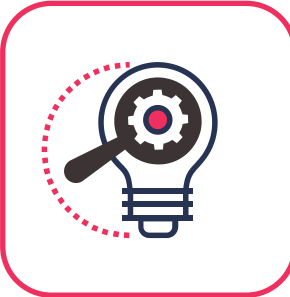
4. 미래 예측

1. 선정 배경



2. 데이터 분석

3. 데이터 결과



4. 미래 예측



1. 선정 배경



동아시아의 가장 큰 영향력



비슷한 캐글러의 분포

일본

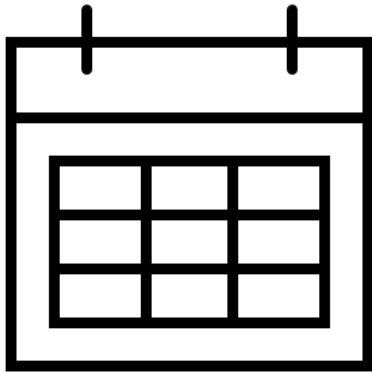
921명

중국

814명

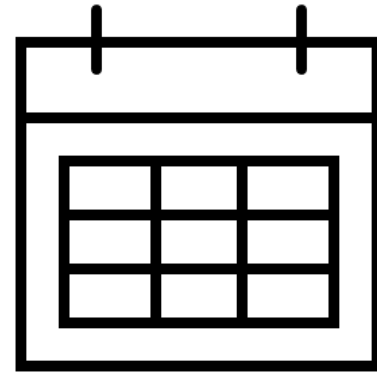


2019

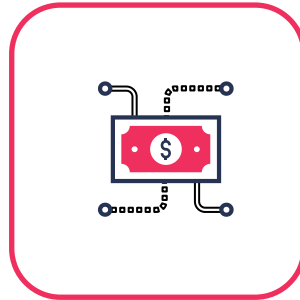


VS

2021

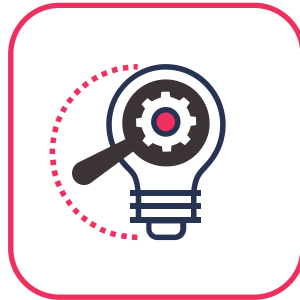


1. 선정 배경



2. 데이터 분석

3. 데이터 결과



4. 미래 예측



The number of

Kaggler

캐글러의 숫자

1

Libraries or tools

라이브러리와 툴

3

IDE's & machine learning frameworks

통합개발환경과 프레임워크

5

2

Gender & Age

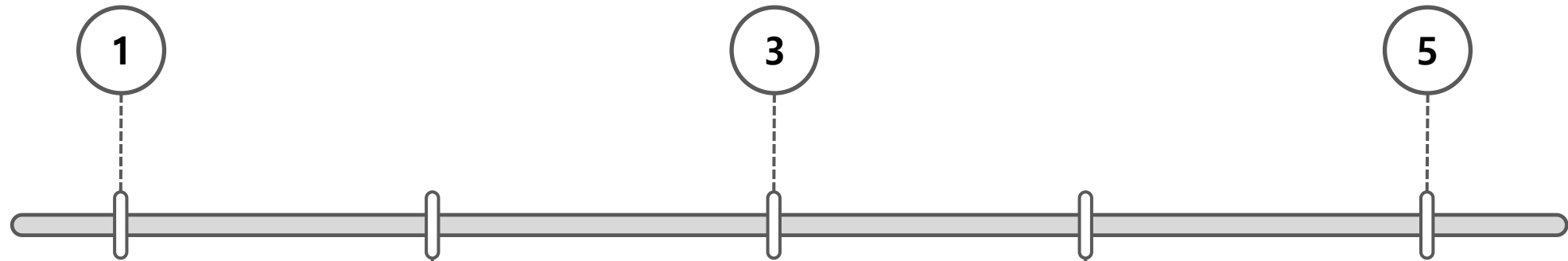
성별과 나이

4

Programming

languages

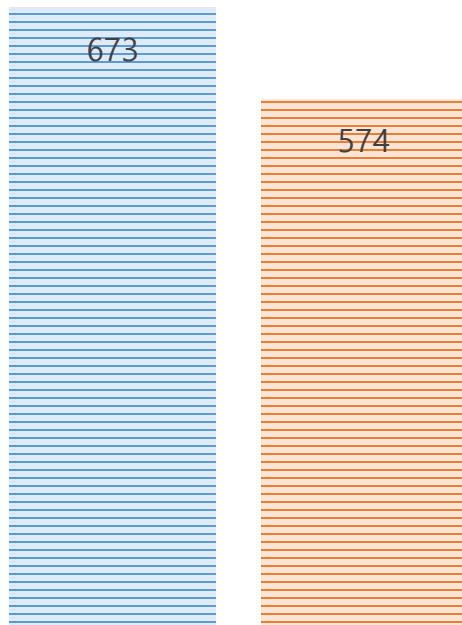
프로그래밍 언어





The number of kaggler

일본 중국



2019



증가율

약 40%

일본 중국



2021



Gender



2019 Japan

남성

90.9%



2021 Japan

남성

90.8%



2019 China

남성

81.7%



2021 China

남성

82.6%

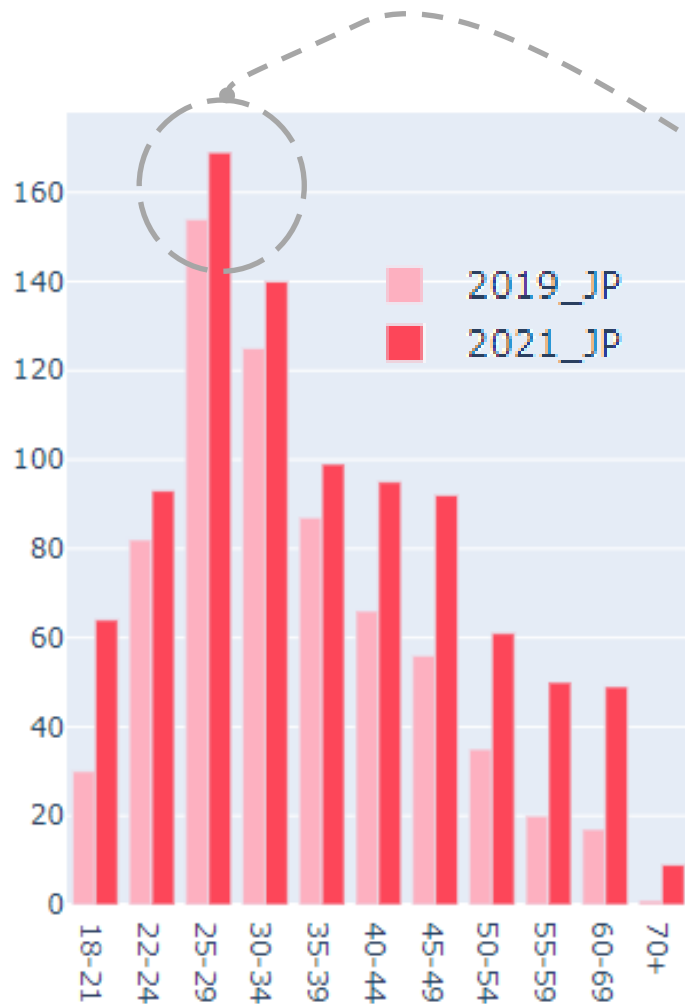


여성

남성



AGE



Japan Age distribution

일본의 나이 분포도



젊은 층의 점유율

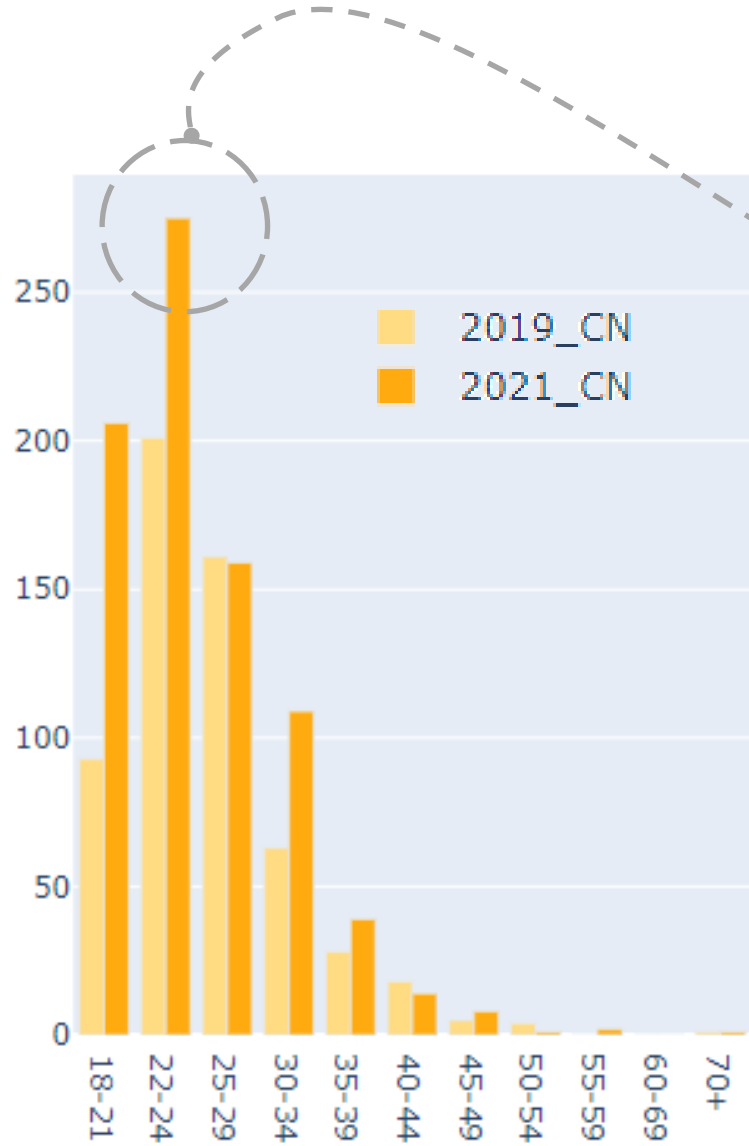


다양한 연령층





AGE



중국의 나이 분포도



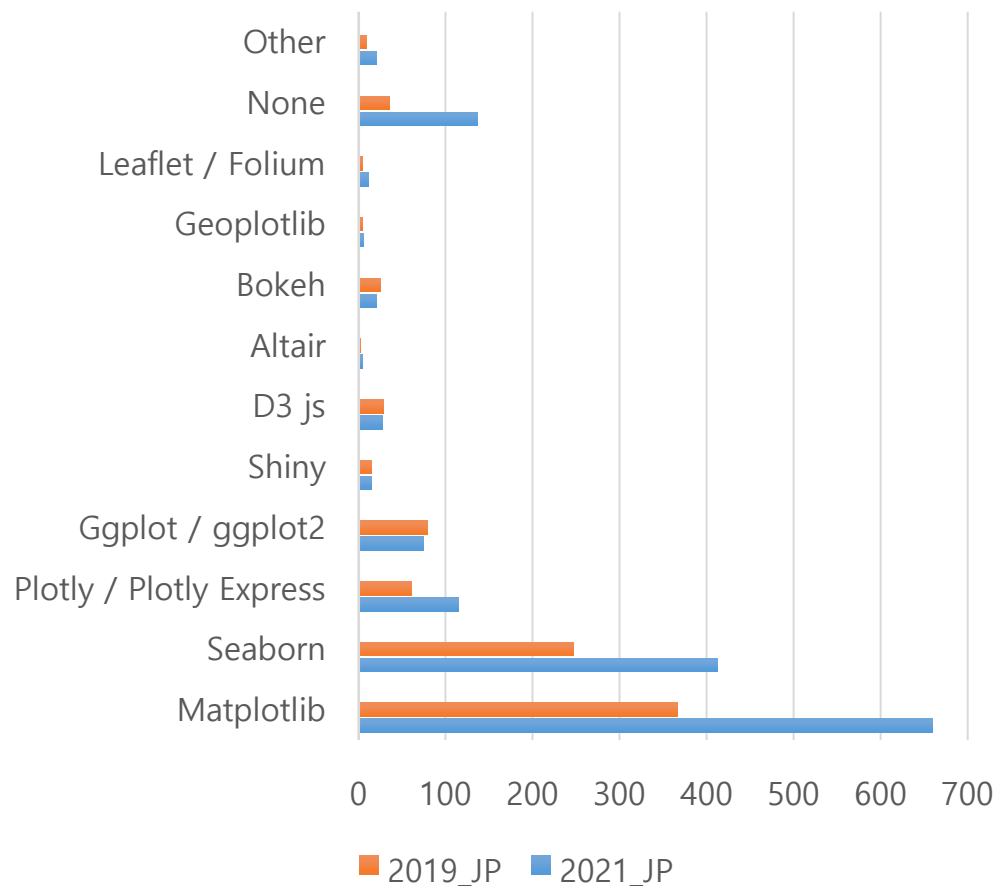
젊은 층의 점유율



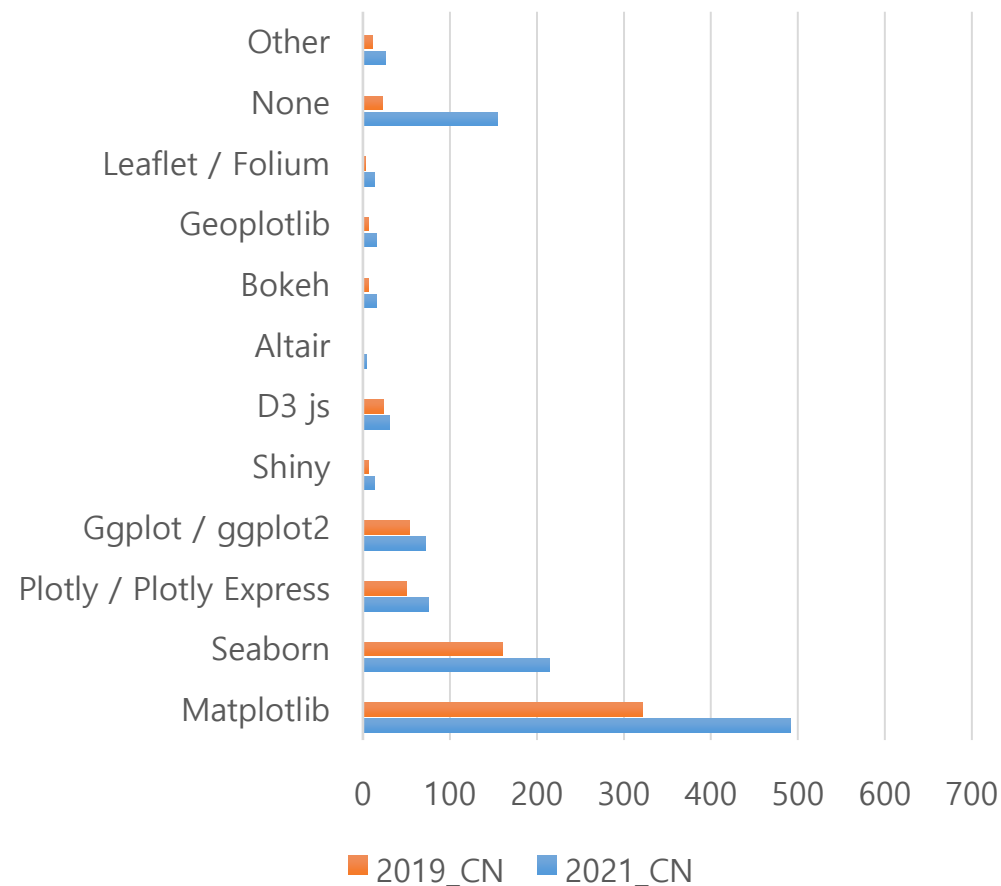
18-34세 **90%**



일본의 라이브러리, 툴 사용 현황

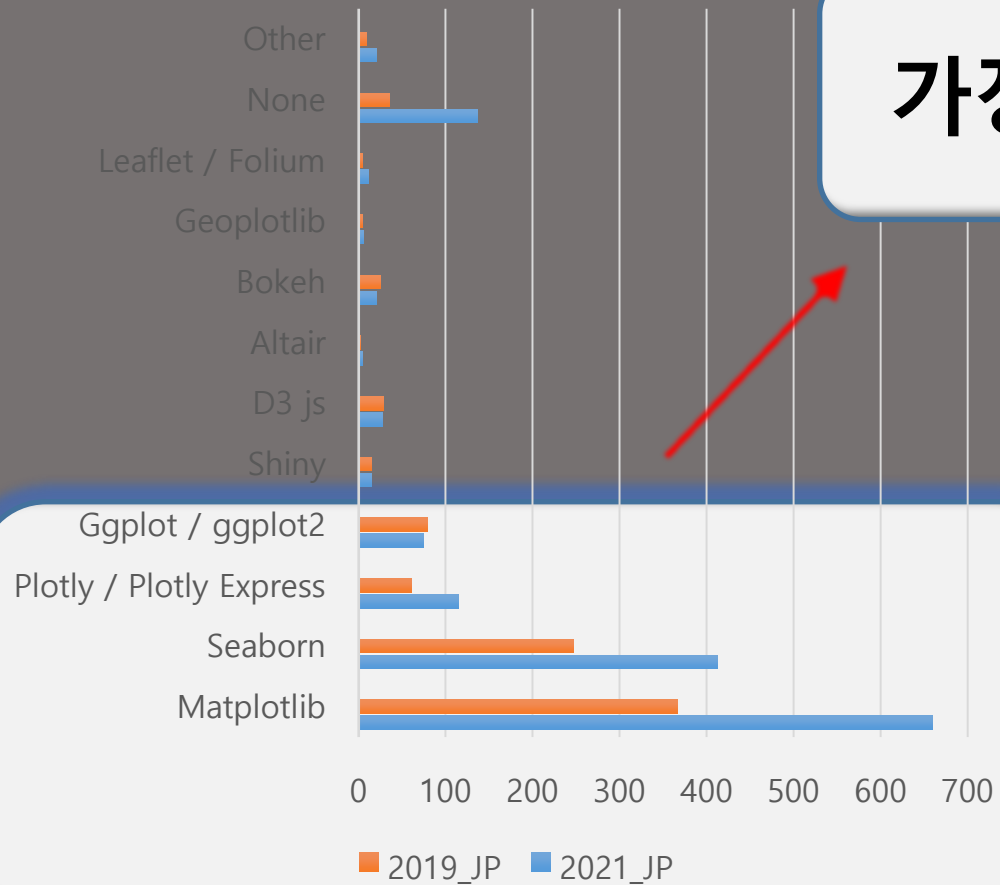


중국의 라이브러리, 툴 사용 현황



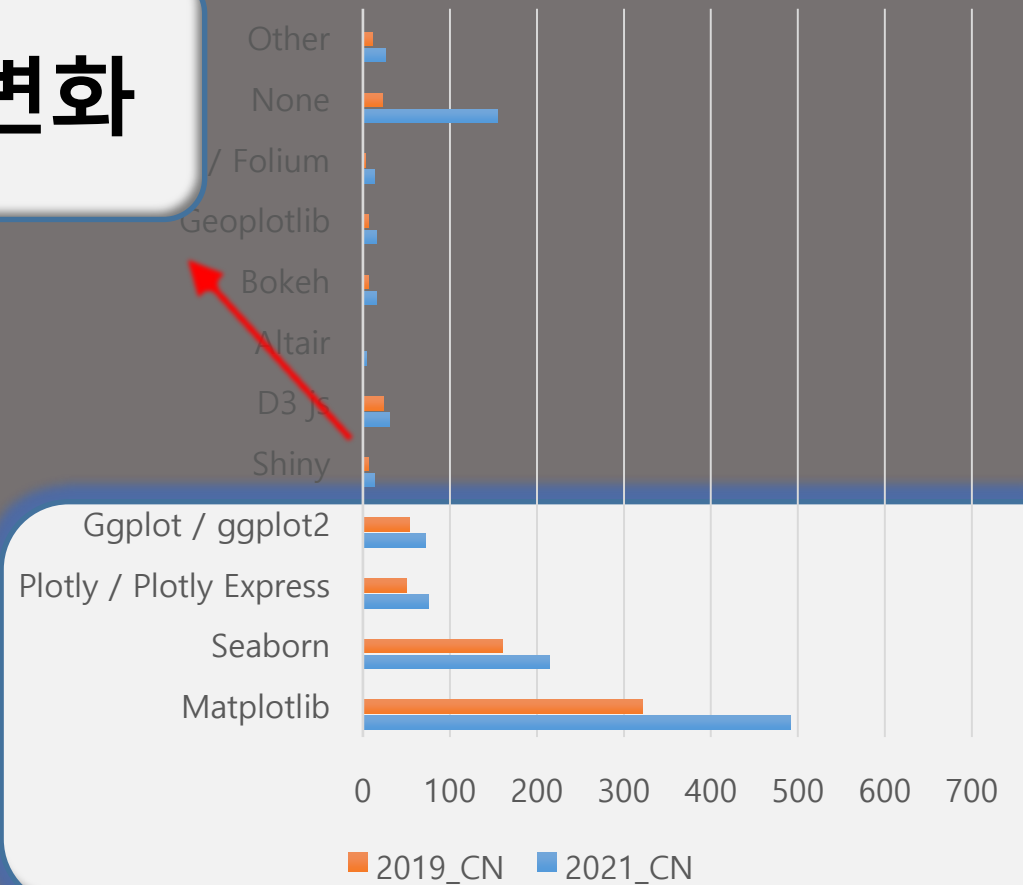


일본의 라이브러리, 툴 사용 현황



가장 큰 변화

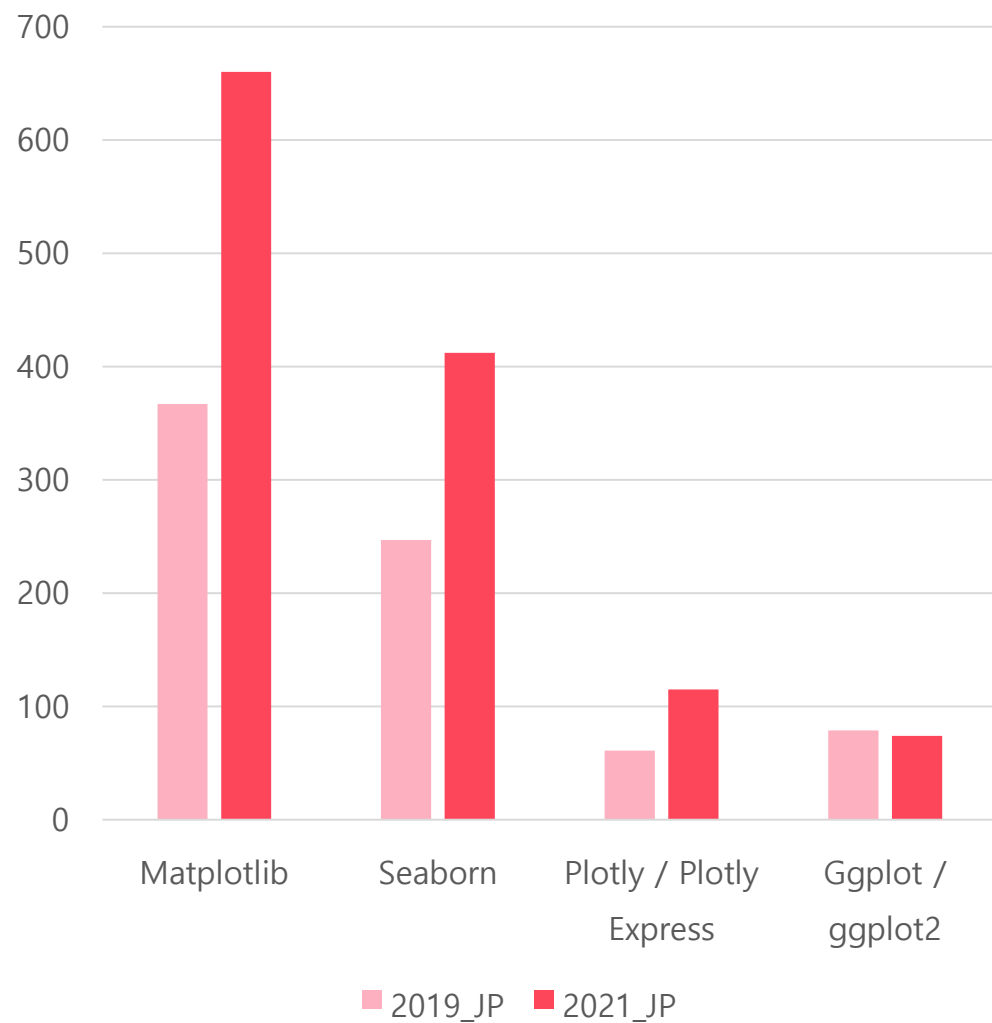
중국의 라이브러리, 툴 사용 현황





Libraries or Tool

일본의 주요 라이브러리, 툴 사용 현황



Matplotlib



Seaborn



Plotly



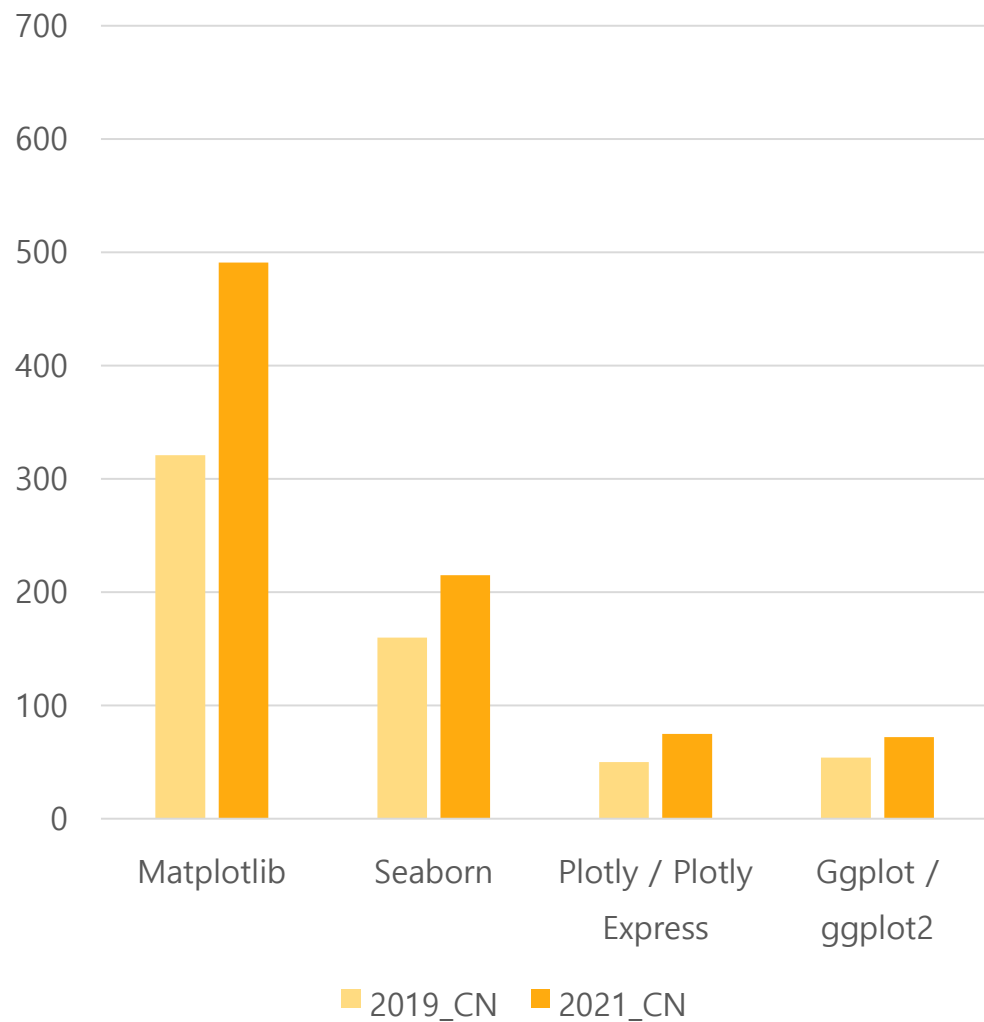
Ggplot





Libraries or Tool

중국의 주요 라이브러리, 툴 사용 현황



Matplotlib



Seaborn



Plotly

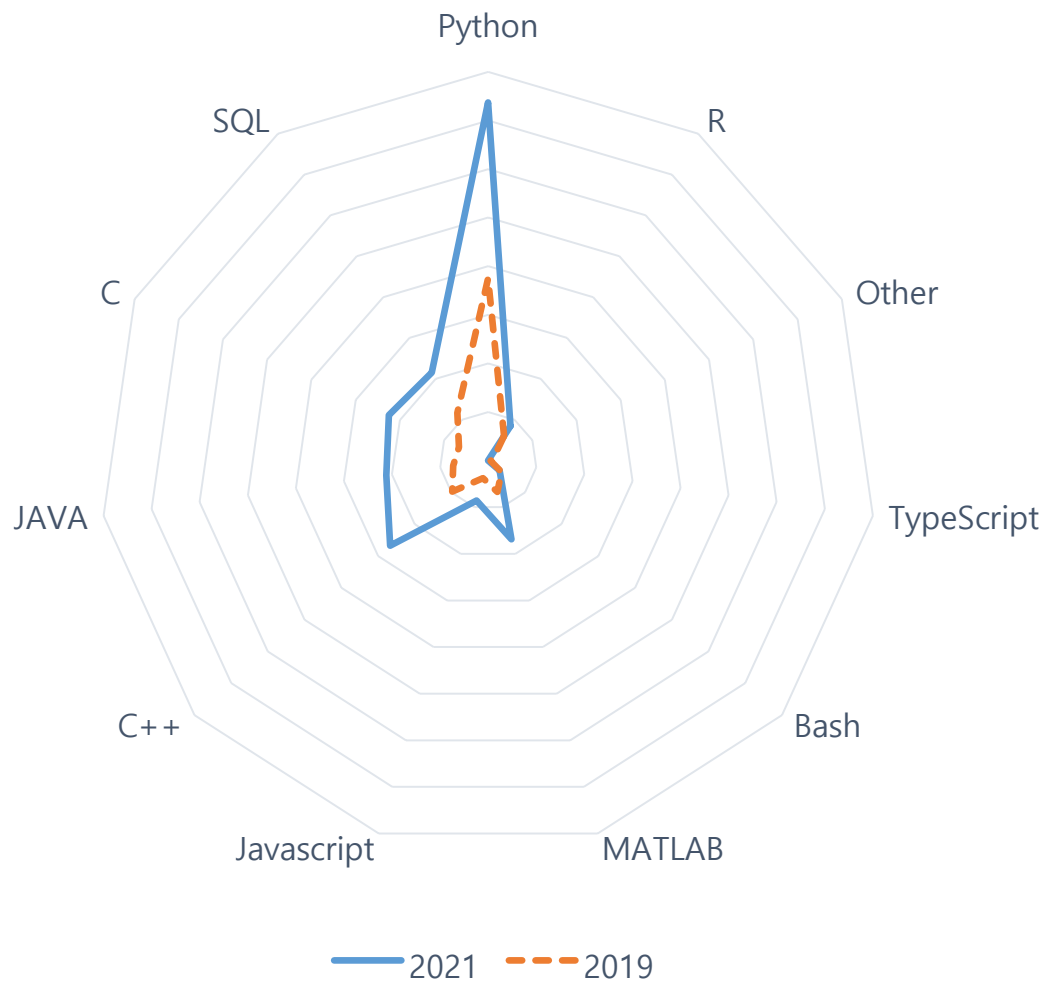


Ggplot

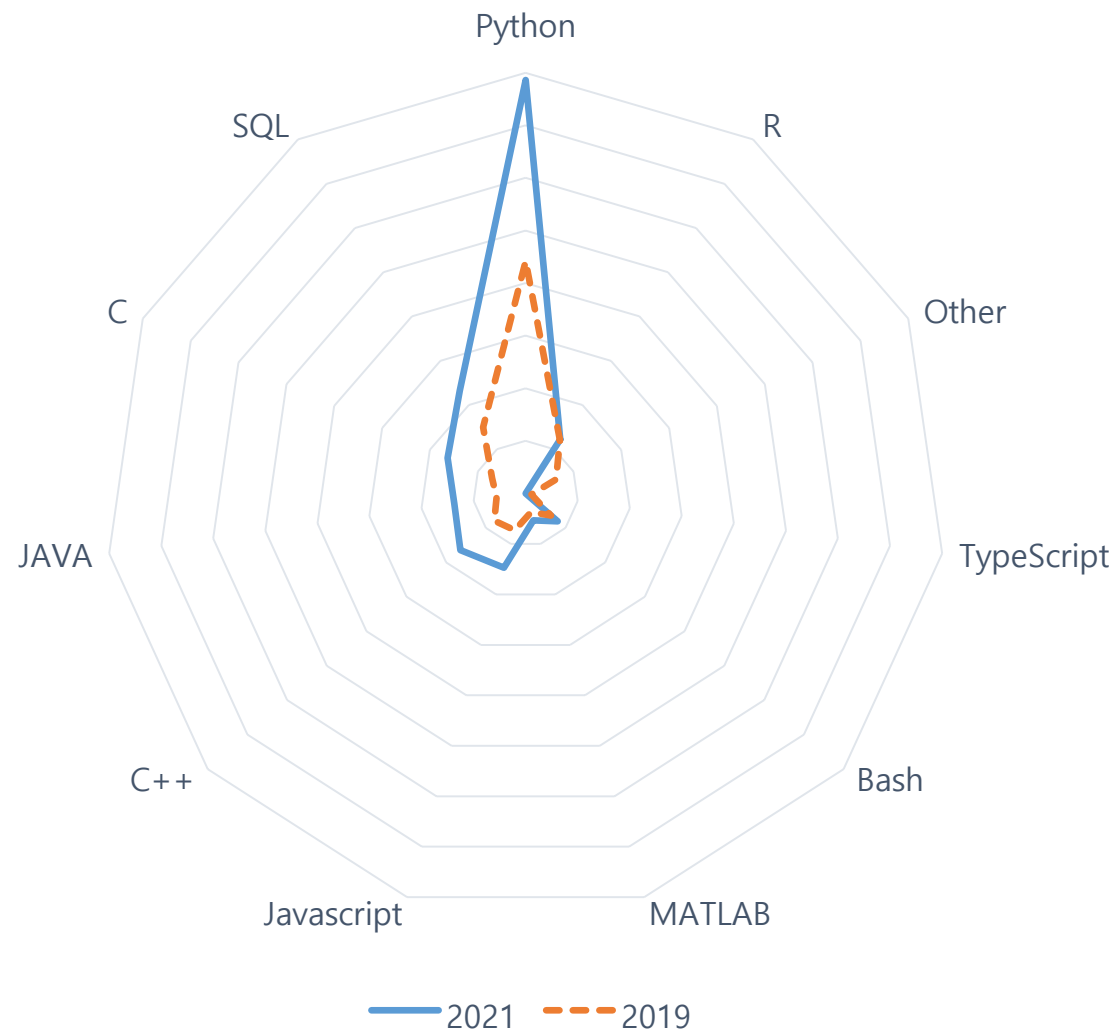




중국의 프로그래밍 언어사용 변화



일본의 프로그래밍 언어사용 변화





통합 개발 환경

2019n2021

2021

Jupyter Notebook

Japan
514
56%China
405
44%

PyCharm

China
494
82%Japan
107
16%

Jupyter

Japan
200
59%China
141
41%

Visual Studio Code (VSCode)

Japan
414
56%China
329
44%

Visual Studio

China
154
53%Japan
136
47%

RStudio

Japan
92
56%China
71
44%

Spyder

China
74
51%Japan
70
49%

Vim / Emacs

Japan
99
69%China
45
31%

Sublime Text

China
75
68%

MATLAB

China
158
82%

Notepad++

China
90
70%Japan
29
24%

None

2019

Jupyter

China
372
56%

Japan

293
44%

PyCharm

Japan
240
71%China
98
29%

VS / VSCode

China
200
59%Japan
137
41%

RStudio

China
96
64%Japan
54
36%

Sublime Text

Japan
86
64%China
48
32%

Spyder

Japan
78
63%China
45
37%

Notepad++

Japan
86
81%

China

Vim / Emacs

China
111
60%Japan
73
40%

MATLAB

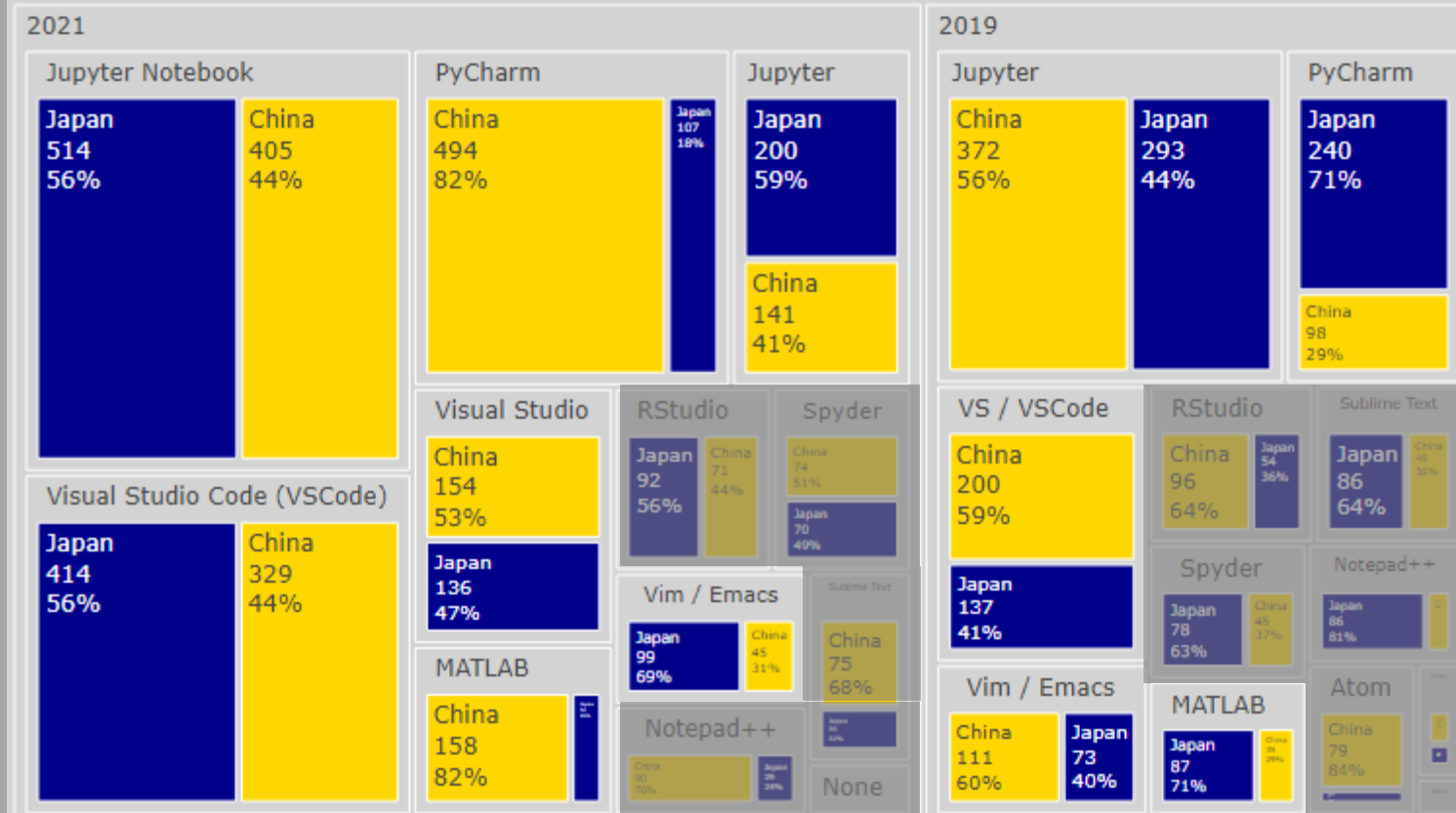
Japan
87
71%China
88
26%

Atom

China
79
84%



2019n2021

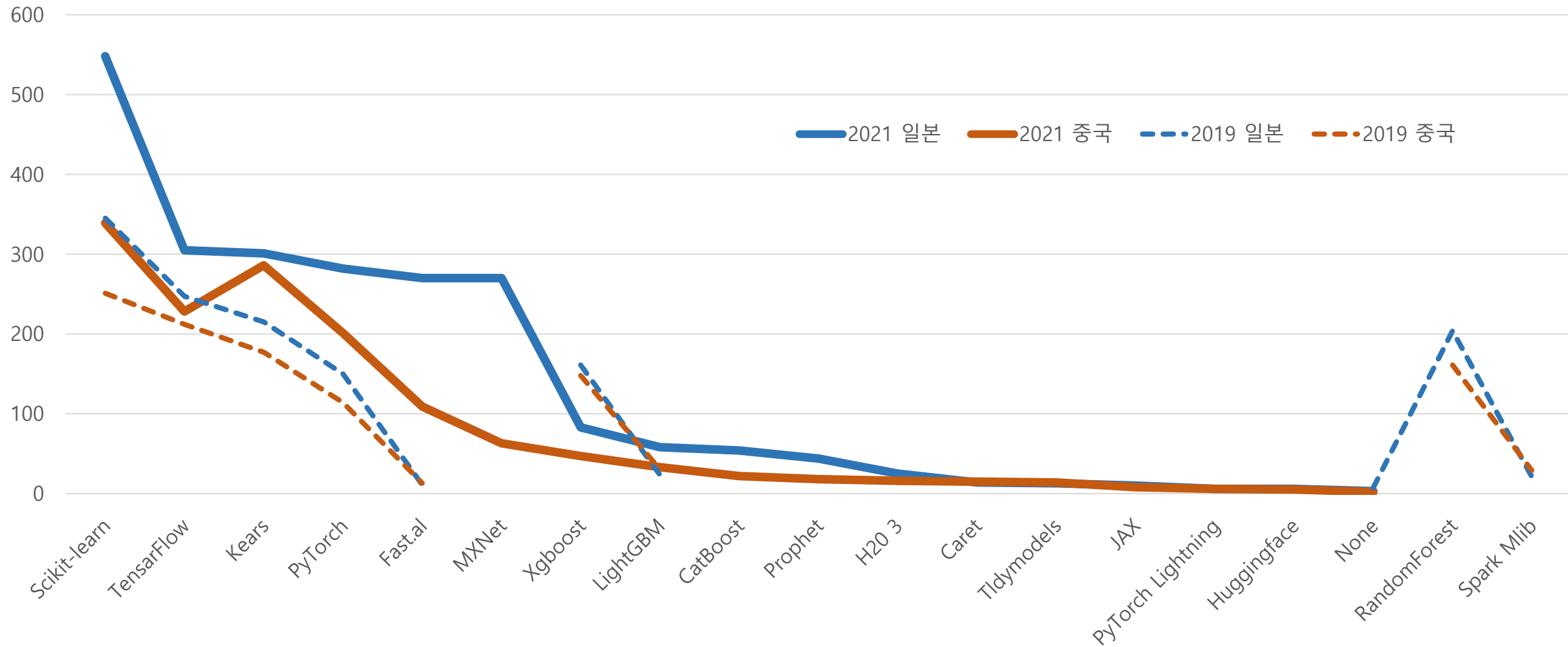


주 사용 개발 환경

- Jupyter
- Visual
- PyCharm
- MATLAB
- Vim/Emacs

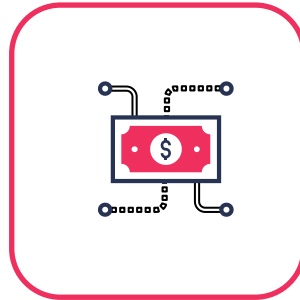


머신러닝 프레임워크





1. 선정 배경



2. 데이터 분석

3. 데이터 결과



4. 미래 예측

일본의 데이터 분석 결과



연령층, 성별

- 일본의 캐글러 수가 약 37% 증가
- 모든 연령층에서 캐글러의 수가 증가
- 고른 나이별 캐글러 분포
- 매우 높은 남성비 (약 90%)

사용 언어, 라이브러리, 툴

- 주요 사용 라이브러리(상위 4개) 중 ggplot 제외 모두 증가
- 라이브러리 및 툴을 사용하지 않는 캐글러 증가
- 모든 언어 항목에서의 증가세
- JAVA의 가장 큰 증가율(144%)
- Python의 가장 큰 증가값(344명)

일본의 데이터 분석 결과



사용 IDE

- Jupyter Notebook와 Visual Studio의 사용자 급증
- PyCharm, Jupyter의 사용자 감소
- 감소세가 보이던 IDE의 사용자와 신규 유입 사용자가 Jupyter Notebook과 Visual Studio로 이동한 것으로 추측

사용 머신러닝 프레임워크

- 프레임워크 항목이 늘어남
- Scikit-learn사용자의 비약적 증가
- Fast.ai, MXNet의 높은 점유율 증가
- Xgboost와 Caret의 점유율이 급감

중국의 데이터 분석 결과



연령층, 성별

- 중국의 캐글러 수 약 42% 증가
- 전체적인 증가세 속에 감소세를 띄는 연령층도 있음(25-29, 40-44)
- 젊은층(18-34)의 매우 높은 점유율(90%, 92%)
- 높은 남성비 (약 82%)

사용 언어, 라이브러리, 툴

- 주요 사용 라이브러리(상위 4개) 모두 증가세
- 라이브러리를 사용하지않는 캐글러 급증 (약 600% 증가)
- 모든 언어 항목에서의 증가세
- c 의 가장 큰 증가율(240%)
- Python의 가장 큰 증가값(362명)



중국의 데이터 분석 결과



사용 IDE

- Jupyter Notebook과 PyCharm, MATLAB 의 비약적 증가
- Jupyter, Rstudio, Vim의 사용자 감소
- PyCharm의 가장 높은 점유율 (26.4%)

사용 머신러닝 프레임워크

- Scikit-learn의 가장 높은 점유율
- 21년 상위 4개의 프레임워크가 75의 점유율을 가짐
- Fast.ai와 MXNet의 점유율이 큰폭으로 증가
- Xgboost와 Caret의 점유율이 각각 -68%와 -84로 급감

공통점

- 캐글러의 수 **증가**, 젊은층의 가장 높은 **점유율**
- 대부분의 연령층에서 캐글러의 수가 **증가**
- 각각 일본 90% 중국 82%로 **남성**
- 모든 언어 항목에서 **증가**
- Python사용자의 가장 **높은 점유율**
- Jupyter Notebook의 사용자 **급증**
- Scikit-learn의 가장 **높은 점유율**
- Fast.ai와 MXNet의 점유율이 **증가**
- 반면, Xgboost와 Caret의 점유율은 크게 **하락**

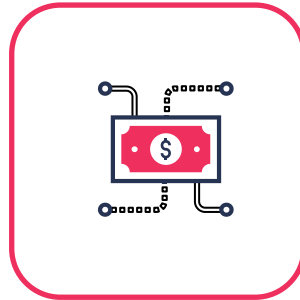


차이점

- 2021년 기준, 중국(92%)의 젊은층 비율이 일본(50%)과 비교해 **매우 높음**
- 일본은 모든 연령층에서 캐글러의 수가 증가했으나, 중국은 **그 수가 감소한 연령층 존재**
- 일본에서는 JAVA(+144%)가, 중국에서는 C(+240%)의 사용자 **증가율이 높았음**
- 일본에서는 VS, 중국에서는 PyCharm과 MATLAB 사용자의 **증가율이 높았음**
- 일본에서는 Scikit-learn의 점유율과 증가율이 **매우 높음**
- 일본의 Fast.ai 점유율 증가율이 중국의 **3배**에 달함

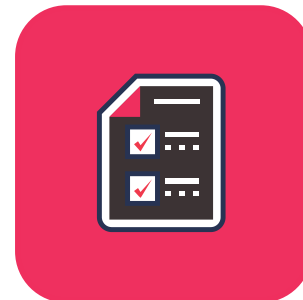
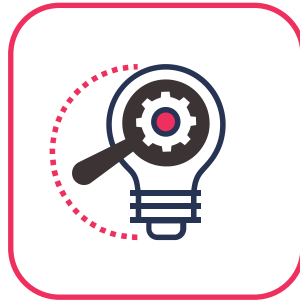


1. 선정 배경



2. 데이터 분석

3. 데이터 결과



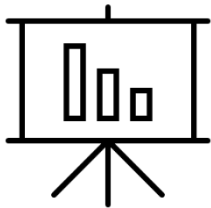
4. 미래 예측



데이터로 본 결과



- 젊은층의 신규 유입



- 가장 높은 점유율 항목



- 두 나라의 트렌드



향후 미래 예측

캐글 시장 가속화

기본 필수 프로그램화
IT산업 유행화 가능성

인접 국가 동아시아에 영향



<https://plotly.com/python/bar-charts/>



<https://plotly.com/python/reference/bar/>



<https://plotly.com/python/pie-charts/>



<https://plotly.com/python/reference/pie/>



[\[Pandas 기초\] 데이터프레임 합치기\(merge, join, concat\) - yg's blog \(yganalyst.github.io\)](#)



[\[Python\] 데이터프레임 열 이름/컬럼명 변경 :: rename \(tistory.com\)](#)



[pandas.DataFrame.insert — pandas 1.3.4 documentation \(pydata.org\)](#)



마치며

Thank
you!