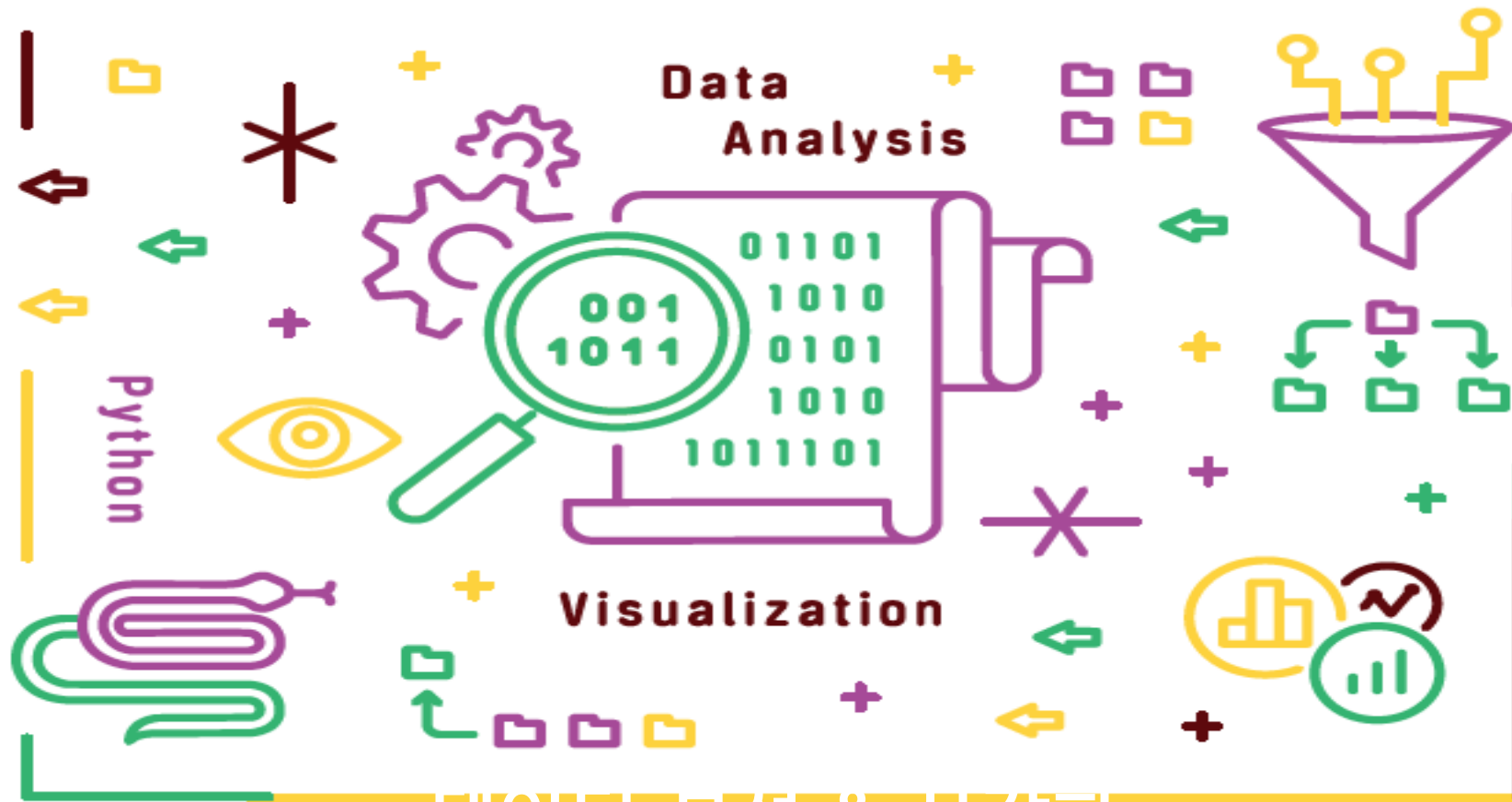


데이터 분석 & 시각화



목차

1. 데이터 분석 이해

- 데이터 분석 이해 및 활용
- 데이터 분석 도구 및 처리과정

2. 데이터 분석 라이브러리

- 배열처리(NumPy)
- 데이터분석 도구(Pandas)
- 데이터 시각화 도구(Matplotlib, Seaborn)

3. 통계 분석과 시각화

- 기술 통계 분석과 시각화
- 상관 분석 + 히트맵 시각화

4. 텍스트 빈도 분석과 시각화

- Wordcloud, pytagcloud 시각화 생성하기
- 영문 분석 및 시각화
- 한글 분석(KoNLPy) 및 시각화

5. 지리정보 분석과 시각화

- Forium 사용 지도 생성하기
- 지리정보 분석 후 맵 생성하기

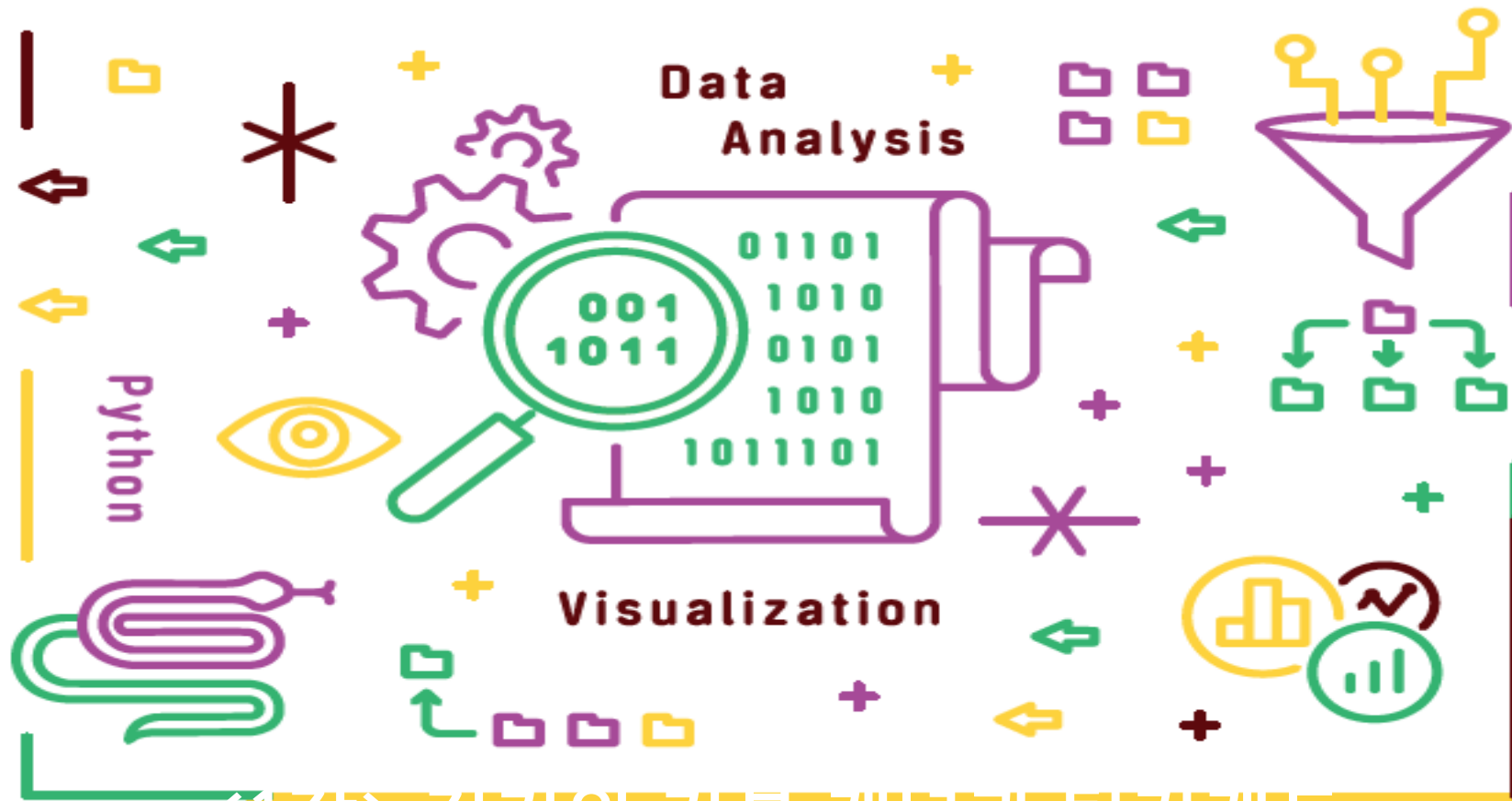
6. 다양한 데이터 분석 및 시각화 방법

- 데이터 수집에 도움되는 사이트
- 데이터 분석 및 시각화를

7. Django 시각화 결과 웹 배포

- Django
- 웹프로그래밍

1. 데이터 분석 개요



학습 목표

- 데이터 분석이 무엇인지 알고 필요성에 대하여 설명한다.
- 데이터 분석 활용 분야에 대하여 설명한다.
- 데이터 분석 도구 종류에는 무엇이 있는지 알고, 데이터 분석 도구로써 파이썬을 설명한다.
- 데이터 분석 처리 과정에 대하여 설명한다.

목차

01 데이터 분석 이해

02 데이터 분석 활용 분야

03 데이터 분석 도구

04 데이터 분석 처리 과정

01

데이터 분석 이해

1. 데이터 분석 이해

■ 데이터 분석 정의

- 데이터(Data)
 - 사실(Fact)을 표현한 것
 - 숫자, 문자, 이미지, 오디오, 동영상 등의 형식으로 나타낸 모든 것을 의미
 - 필요에 따라 정보(Information)가 될 수 있음

데이터 분석(Data Analysis)

문제 해결 또는 인사이트 도출을 위한 목적으로 데이터 수집, 전처리 및 정형화, 분석하는 과정이다. 데이터 분석은 여러 가지 접근 방식이 있으며 다양한 기술을 조합하여 서로 다른 비즈니스, 과학, 사회과학 등 다양한 분야에서 사용된다. 오늘날 비즈니스 분야에서 데이터 분석은 의사 결정을 더욱 과학적이고 효율적으로 운영할 수 있도록 도와주는 중요한 역할을 한다.

1. 데이터 분석 이해

■ 데이터 분석 정의

- 데이터 분석을 위한 데이터 특징
 - 문제 해결또는 인사이트 도출을 위한 데이터
 - 데이터 분석을 하는 목적은 문제 해결이나 개선을 위함
 - 정확한 수치 데이터
 - 데이터 분석을 위해서는 수집한 데이터가 정확한 수치 데이터여야 하고 그 수치 표현 형식도 일관되어야 함
 - 적당한 양의 데이터
 - 데이터 분석을 할 때 적은 양의 데이터를 이용하게 되면 한쪽으로 편향된 결과가 나타나 문제 해결 및 인사이트 도출에 적절하지 않은 정보를 제시할 수 있음

1. 데이터 분석 이해

■ 데이터 분석 필요성

■ 필요성


데이터 분석의 필요성

1. 문제 상황을 정확히 파악한다.
2. 문제 해결 및 인사이트 도출을 위한 목적을 정확히 정의한다.
3. 목적에 따라 효율적인 방안을 제시한다.
4. 만족스러운 결과가 도출될 때까지 실행 과정을 평가하고 개선한다.

1. 데이터 분석 이해

■ 데이터 분석 필요성

- 데이터 분석가
 - 데이터 분석을 전문적으로 하는 사람

| 50 Best Jobs in America for 2022 | | | | |
|--|--------------------|------------------|--------------|---------------------------|
| Best Places to Work Top CEOs Best Jobs Best Cities for Jobs Highest Paying Jobs | | | | |
| 2022 United States | | | | |
|  Discover Glassdoor's Best Jobs in 2022 Using Glassdoor's unique data on jobs, salaries, and companies, we compiled a list of the 50 Best Jobs in America to help people find jobs they'll love. Each job stands out for its earning potential (median salary), job satisfaction, and job openings. Are you considering a new position? Check out this comprehensive list to see what jobs made the list this year, and view open jobs at companies across the country. | | | | |
| Job Title | Median Base Salary | Job Satisfaction | Job Openings | |
| #1 Enterprise Architect | \$144,997 | 4.1/5 | 14,021 | View Jobs |
| #2 Full Stack Engineer | \$101,794 | 4.3/5 | 11,252 | View Jobs |
| #3 Data Scientist | \$120,000 | 4.1/5 | 10,071 | View Jobs |

| | | | | | |
|-----|---------------------|----------|-------|--------|---------------------------|
| #35 | Data Analyst | \$74,224 | 4.0/5 | 13,657 | View Jobs |
| #36 | Business Analyst | \$81,556 | 3.9/5 | 15,238 | View Jobs |
| #37 | QA Engineer | \$87,626 | 4.1/5 | 4,481 | View Jobs |
| #38 | Front End Engineer | \$81,136 | 4.0/5 | 10,044 | View Jobs |
| #39 | HR Business Partner | \$95,431 | 4.0/5 | 4,108 | View Jobs |
| #40 | Project Manager | \$86,000 | 3.8/5 | 42,554 | View Jobs |
| #41 | Compliance Officer | \$80,000 | 3.9/5 | 12,053 | View Jobs |
| #42 | Program Manager | \$81,335 | 3.9/5 | 26,881 | View Jobs |

그림 1-1. 2022년 미국의 베스트 직업 순위

02

데이터 분석 활용 분야

2. 데이터 분석 활용 분야

■ 국외 활용 사례

- 구글, 아마존, DHL, 넷플릭스, 스타벅스 등 국외 기업들이 데이터 분석 기술을 활용

표 1-2. 데이터 분석 국외 활용 사례

| 분류 | 국외 기업 | 활용 내용 |
|----------------------|----------|--|
| 고객 관계 관리 / 경험의 변화 | 아비바생명 | 고객 맞춤형 보험 상품 제공 |
| | 사우스웨스트항공 | 고객 맞춤형 광고 |
| 고객 관계 관리 / 경험의 변화 | 타겟 | 고객 맞춤형 상품 프로모션 |
| | 넷플릭스 | 시청자들의 행동 관찰, 선호도 분석을 통한 맞춤형 콘텐츠 제작 |
| | 스타벅스 | 데이터 기반의 상권 분석, 소비자 정보 기반의 추천 서비스 |
| 내부 프로세스 / 효율성 개선 | DHL | 물류 효율화, 투자 의사결정에 참고 |
| | 자라(Zara) | 효율적인 물류 배송망 운영 |
| | 구글 | 데이터 센터 성능 및 에너지 사용 최적화 |
| | 테슬라 | 데이터 분석을 통해 의사결정을 내리는 데이터 드리븐(Data Driven) 경영 방식 도입 |
| 신규 가치 제안 창출 | 아마존 | 고객 주문 전에 예측 배송 |
| | GE | 산업 인터넷으로 스마트 공장 구현 |
| | 후지쯔 | 농업용 빅데이터 분석 솔루션 제공 |

2. 데이터 분석 활용 분야

■ 국내 활용 사례

- 신한카드, 엔씨소프트, 이마트, SK텔레콤 등 국내 기업들이 데이터 분석 기술을 선택이 아닌 필수적으로 활용

표 1-3. 데이터 분석 국내 활용 사례

| 분류 | 국내 기업 | 활용 내용 |
|----------------------|---------------|---|
| 고객 관계 관리 / 경험의 변화 | 신한카드 | 고객 라이프 스타일 맞춤형 상품 |
| | KB국민카드 | 소비자 정보 기반 카드 이용 및 편의성 서비스 |
| | 엔씨소프트 | 사기 탐지(Fraud Detection) 알고리즘을 통한 불법 행위 모니터링 |
| | 옥션, G마켓, 11번가 | 고객 구매력에 따른 개인화된 상품 추천과 광고 배달 |
| 내부 프로세스 / 효율성 개선 | 삼성SDS | 제조업 생산성 향상 지원 솔루션 |
| | 이마트 | 데이터 분석을 통해 의사결정을 내리는 데이터 드리븐(Data Driven) 경영 방식으로 생산한 자체생산(PB) 상품 |
| 신규 가치 제안 창출 | SK텔레콤 | 상권 분석 및 타겟 마케팅 지원 서비스 |

03

데이터 분석 도구

3. 데이터 분석 도구

■ 데이터 분석 도구를 선택할 때 고려 사항

- 처리 성능
- 데이터 구성
- 분석 기법
- 시각화

3. 데이터 분석 도구

■ 데이터 분석 도구 종류

■ R

- 로버트 젠틀맨(Robert Gentleman)과 로스 이하카(Ross Ihaka)가 만든 오픈 소스 통계 프로그래밍 언어

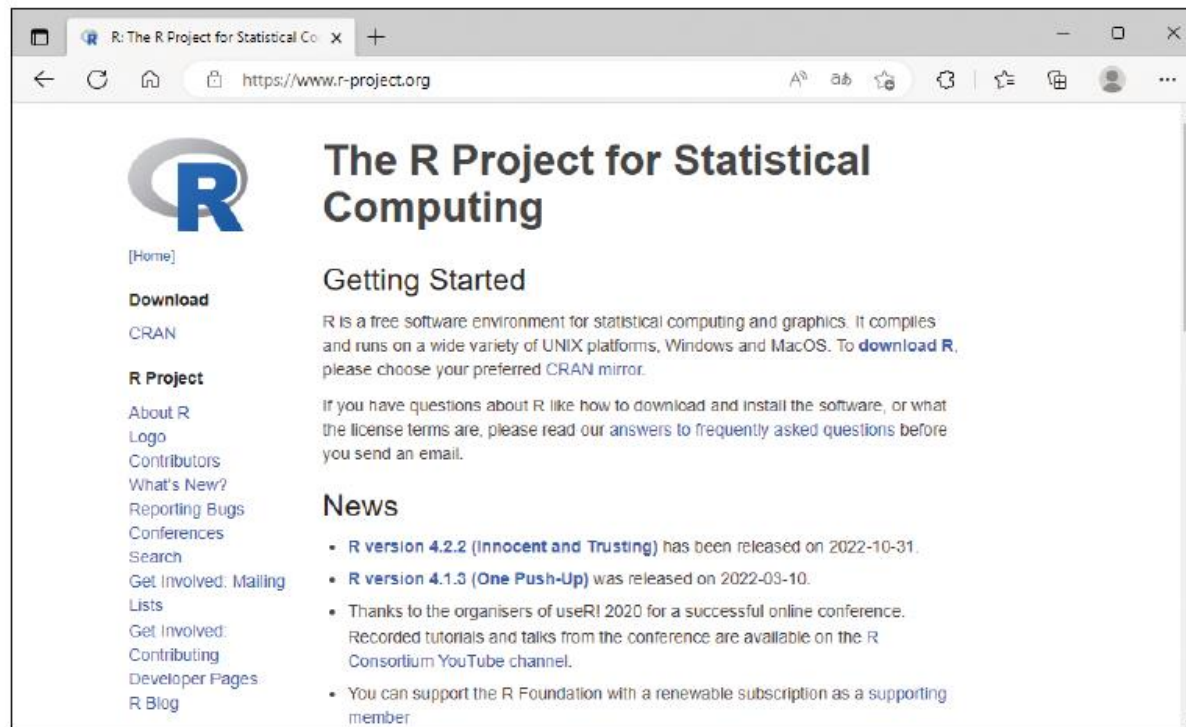


그림 1-2. R 홈페이지

3. 데이터 분석 도구

■ 데이터 분석 도구 종류

■ 파이썬

- 귀도 반 로섬(Guido van Rossum)이 만든 프로그래밍 언어

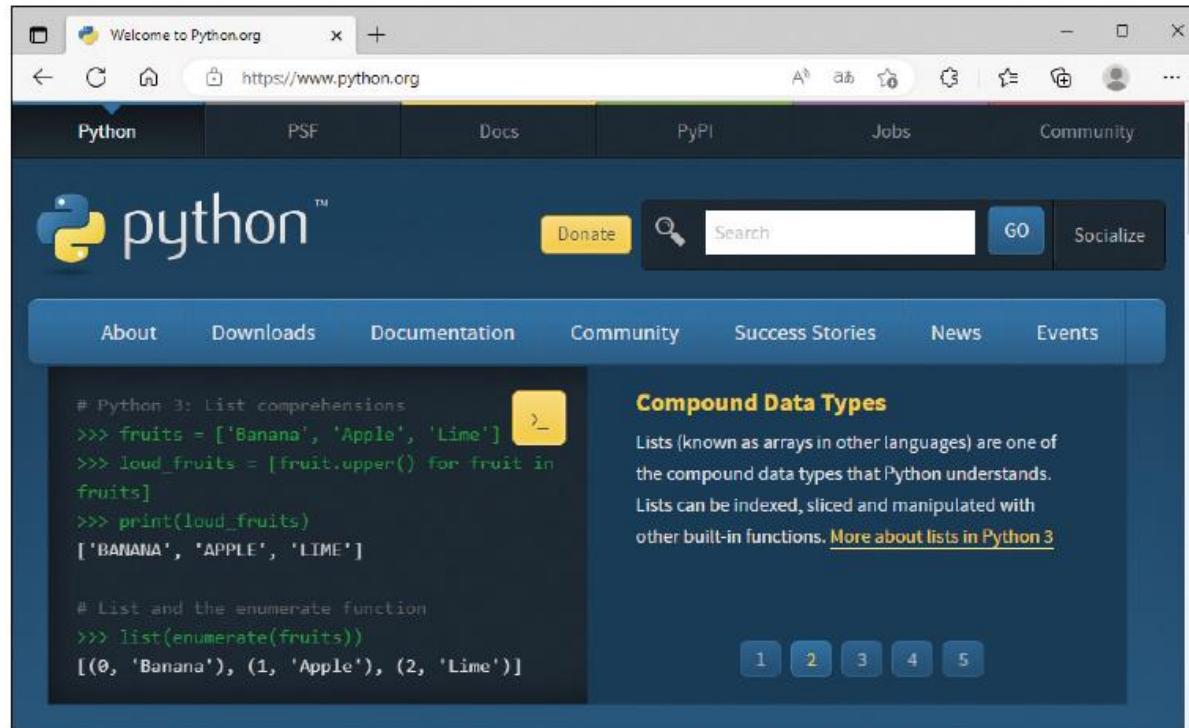


그림 1-3. 파이썬 홈페이지

3. 데이터 분석 도구

■ 데이터 분석 도구 종류

- 파이썬
 - 아나콘다

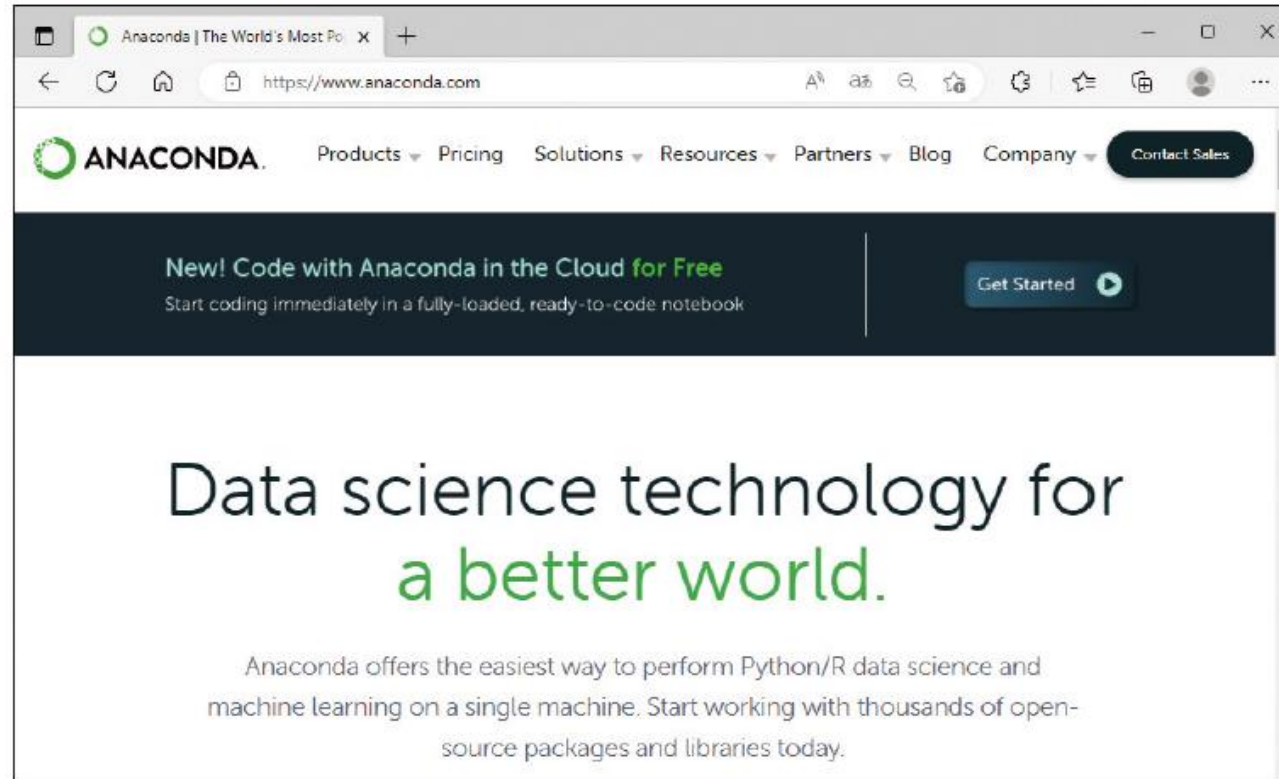


그림 1-4. 아나콘다 홈페이지

3. 데이터 분석 도구

■ 데이터 분석 도구 종류

■ 파이썬

- 구글 코랩

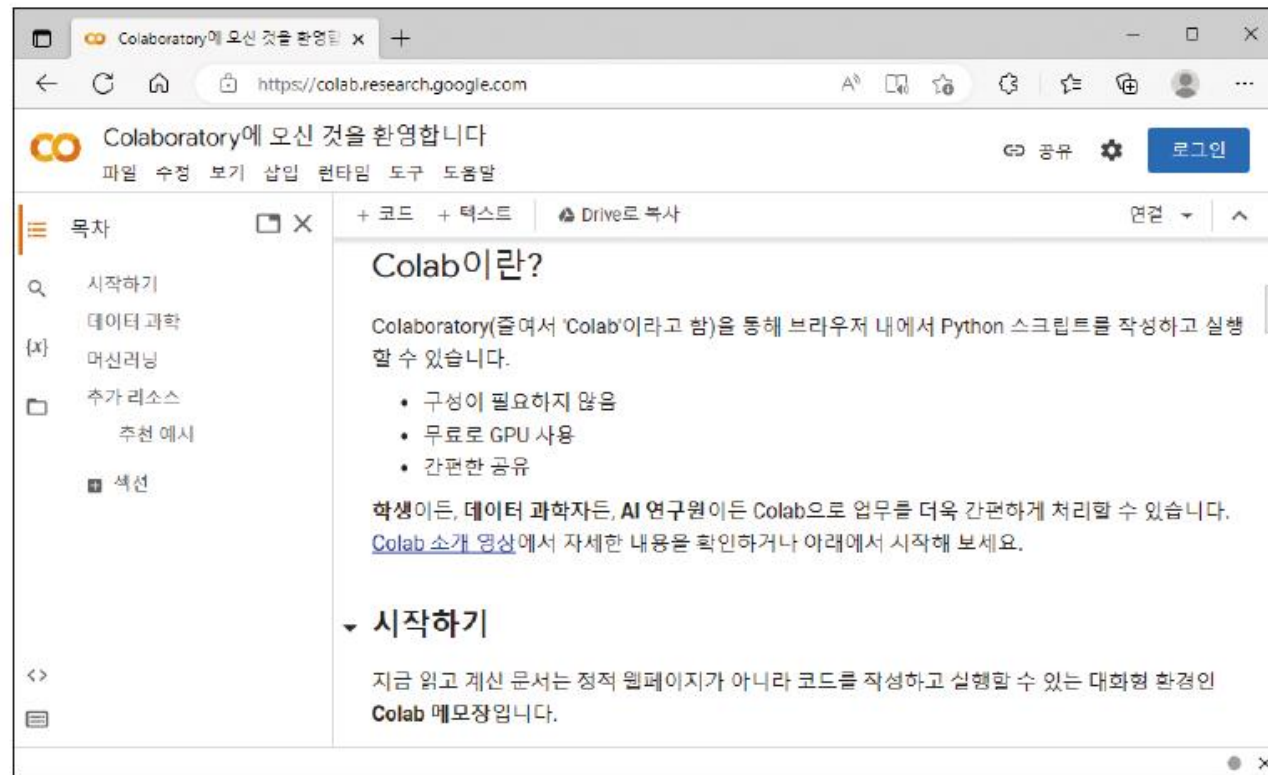


그림 1-5. 구글 코랩 홈페이지

3. 데이터 분석 도구

■ 데이터 분석 도구 종류

■ 엑셀

- 마이크로소프트사에서 만든 계산용 스프레드시트 프로그램

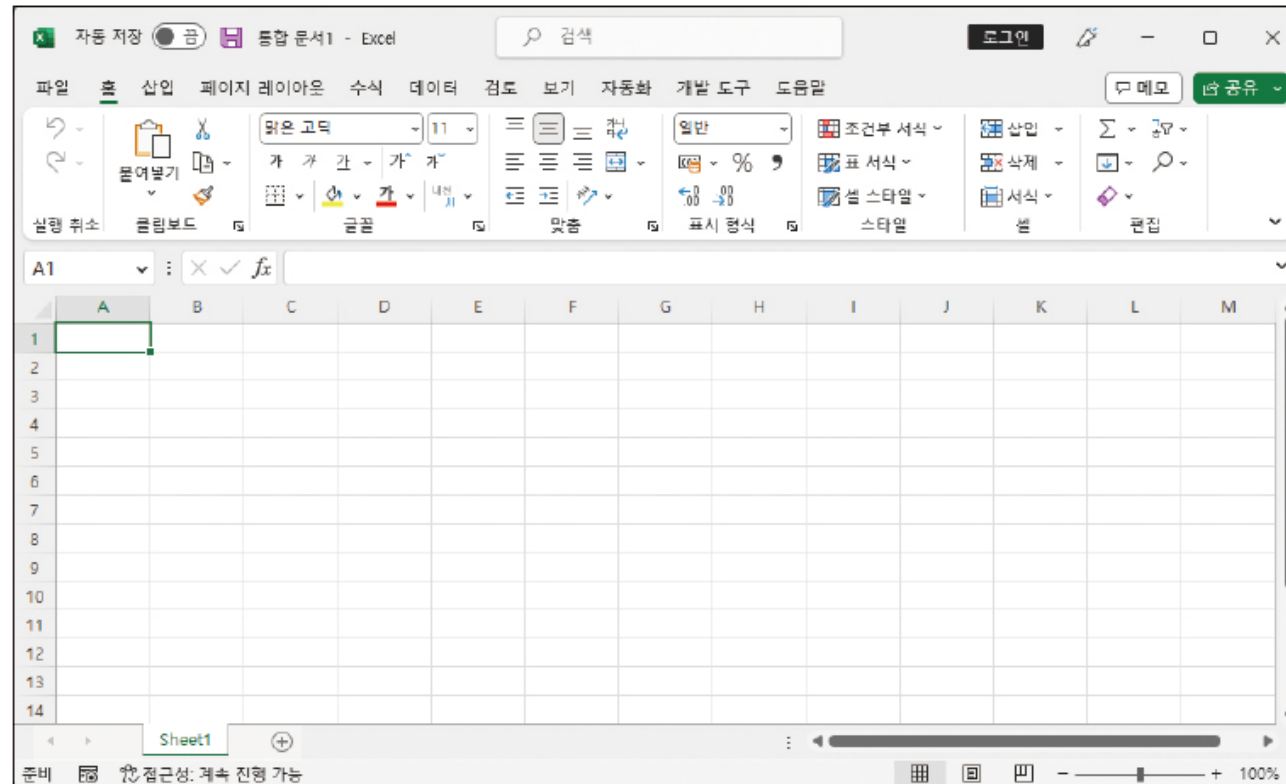


그림 1-6. 엑셀 화면

3. 데이터 분석 도구

■ 데이터 분석 도구로서의 파이썬

- 파이썬이 데이터 분석 도구로 유용한 이유
 - 직관적이고 쉬운 문법
 - 대량의 데이터를 빠르게 처리
 - 다양한 라이브러리 지원

04

데이터 분석 처리 과정

4. 데이터 분석 처리 과정

■ 데이터 분석 처리 과정

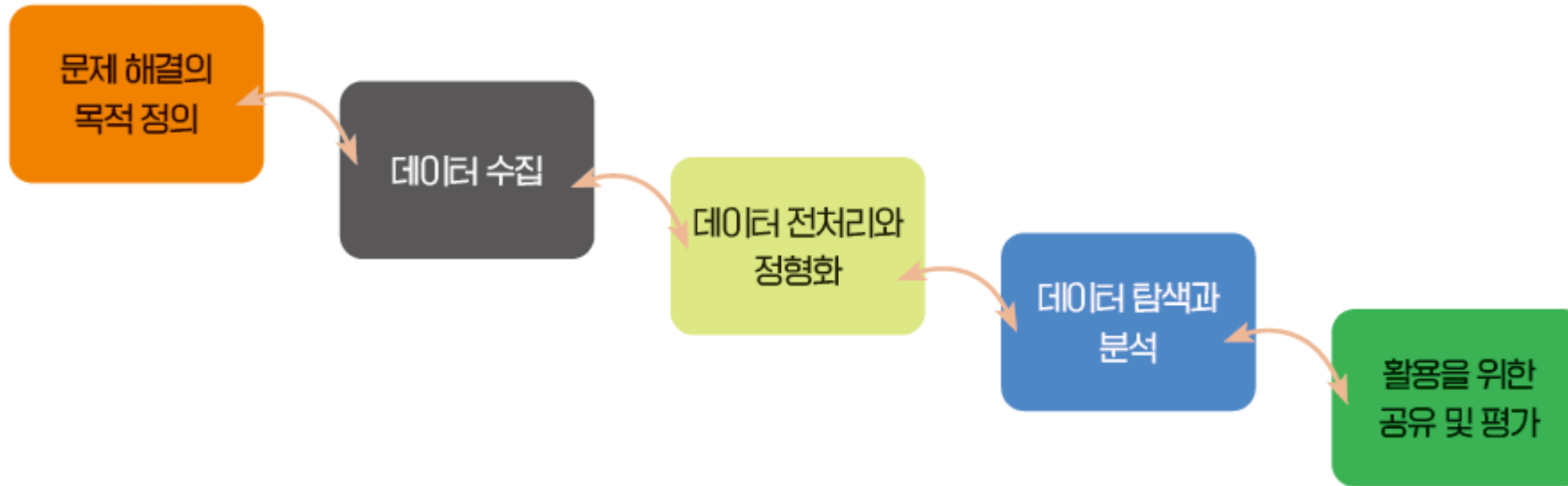


그림 1-7. 데이터 분석 처리 과정

4. 데이터 분석 처리 과정

■ 문제 해결의 목적 정의

- 문제 해결을 위한 데이터 분석의 목적을 정의할 때 고려 사항
 - 문제의 배경
 - 요구 사항 파악
 - 분석 범위

4. 데이터 분석 처리 과정

■ 데이터 수집

- 문제를 해결하기 위해 필요한 준비물을 준비하는 것
- 수집할 데이터의 속성을 찾아내서 정의해야 함
- 공공기관 홈페이지에서 무료로 제공하는 공공 데이터를 다운로드하여 수집

표 1-7. 공공기관 홈페이지의 인터넷 주소

| 기관명 | 인터넷 주소 |
|--------------------------|---|
| ACKIS(문화 예술 지식 정보 시스템) | https://policydb.kcti.re.kr/main.do |
| AI 허브 | https://aihub.or.kr/ |
| Kaggle (세계의 다양한 주제의 데이터) | https://www.kaggle.com/datasets |
| Quandl(경제, 의료, 부동산 등) | https://www.quandl.com |
| SGIS(통계 지리 정보 서비스) | http://sgis.kostat.go.kr/ |
| TAAS(교통 사고 분석 시스템) | http://taas.koroad.or.kr/ |
| Uniprot(염기서열) | https://www.uniprot.org |
| 건축데이터 민간개방 시스템 | https://open.eais.go.kr/ |
| 경기도 공공데이터 개방포털 | https://data.gg.go.kr/portal/mainPage.do |
| 경기버스정보 | http://www.gbis.go.kr/ |

4. 데이터 분석 처리 과정

표 1-7 공공기관의 홈페이지 인터넷 주소

| | |
|-----------------------|---|
| 고속도로 공공데이터 포털(한국도로공사) | http://data.ex.co.kr/ |
| 공공데이터포털 | https://www.data.go.kr/index.do |
| 관광지식정보시스템 | http://www.tour.go.kr/main.asp |
| 국가공간정보포털(국토교통부) | http://www.nsdi.go.kr/ |
| 국가교통DB | https://www.ktdb.go.kr/ |
| 국가통계포털 | http://kosis.kr/ |
| 기미나인(공공데이터 큐레이션 서비스) | https://www.gimi9.com/ |
| 기상자료개방포털 | https://data.kma.go.kr/ |
| 농림축산식품 공공데이터 | https://data.mafra.go.kr/main.do |
| 데이콘(AI 해커톤 플랫폼) | https://dacon.io/ |
| 보건의료빅데이터개방시스템 | http://opendata.hira.or.kr/op/opc/selectOpenDataList.do |

| | |
|-------------------------|---|
| 부동산통계정보시스템 | https://www.reb.or.kr/r-one/main.do |
| 서울 열린 데이터 광장 | http://data.seoul.go.kr |
| 서울특별시 대기환경정보 | https://cleanair.seoul.go.kr/ |
| 영화진흥위원회 | http://www.kofic.or.kr/kofic/business/main/main.do |
| 인공지능 팩토리 | http://aifactory.space/ |
| 국민재난안전포털 | https://www.safekorea.go.kr/idsiSFK/neo/main/main.html |
| 정보공개포털 | https://www.open.go.kr/ |
| 주민등록 인구통계(행정안전부) | https://jumin.mois.go.kr/ |
| 지방행정 인허가 데이터개방 | http://www.localdata.kr/ |
| 철도안전정보종합관리시스템 | https://www.railsafety.or.kr/index.do |
| 티머니 카드&페이(대중교통 통계자료) | https://pay.tmoney.co.kr/ncs/pct/ugd/ReadUgdMainGd.dev |
| 학교알리미(초·중등 교육정보 공시 서비스) | https://www.schoolinfo.go.kr/Main.do |
| 한국소비자원 참가격 | http://www.price.go.kr/tprice/portal/main/main.do |

4. 데이터 분석 처리 과정

■ 데이터 전처리와 정형화

- 데이터 전처리

- 필터링, 데이터 형 변환, 정제 작업 등으로 이루어짐
- 데이터 수집 단계에서 손실된 값을 보정하거나 무의미한 데이터를 제거하는 것

- 데이터 정형화

- 에서는 데이터 분석이 수월하도록 데이터 통합, 데이터 변환, 데이터 축소 등의 가공 작업이 이루어짐
- 데이터 가공의 정형화를 통하여 분석의 효율을 높이는 것

4. 데이터 분석 처리 과정

■ 데이터 탐색과 분석

- 문제 해결 목적에 맞는 데이터를 수집하고 전처리 및 정형화 단계를 거친 이후의 단계
- 미래 상황을 예측할 수 있는 자료로 제공될 수 있도록 다양한 수학적 연산과 추론 방법 이용

표 1-8. 데이터 탐색 과정의 고려 사항

고려 사항1 : 문제 상황의 정확한 표현

고려 사항2 : 일관성 있는 문제 상황 반영

4. 데이터 분석 처리 과정

■ 데이터 탐색과 분석

■ 데이터 일반화

데이터 일반화 과정의 주요 작업

1. 데이터의 대푯값(평균, 최대, 최소 등) 파악
2. 데이터 범위, 분산 및 편향성을 이용한 규칙성과 특이점 발견
3. 조건에 따른 필터링, 그룹화를 이용하여 데이터 선별 및 비교
4. 미래 예측 및 개선 방향 제안

4. 데이터 분석 처리 과정

■ 활용을 위한 공유 및 평가

- 올바른 개선 또는 해결책 제시를 위하여 공유하여 평가

공유 사항

1. 데이터 분석 계기, 문제 해결 목적과 정의된 문제
2. 수집된 데이터의 상세 내용 및 수집 방법
3. 정의된 목적에 따른 분석 방법과 결과
4. 분석 결과를 통해 도출한 과거와 현재 상황
5. 예측 가능한 미래 상황과 대안

Q&A