**Polars 활용 레시피 100제**

**Polars 데이터 분석 레시피 100제 (초판)**

**발 행**｜2025년 04월 01일

**저 자**｜Evan

**펴낸이**｜한건희

**펴낸곳**｜주식회사 부크크

**출판사등록**｜2014.07.15(제2014-16호)

**주 소**｜서울특별시 금천구 가산디지털1로 119 SK트윈타워 A동 305호

**전 화**｜1670-8316

**이메일**｜info@bookk.co.kr

**ISBN**｜979-11-410-0000-0

**www.bookk.co.kr**

**ⓒ Evan 2025**

본 책은 저작자의 지적 재산으로서 무단 전재와 복제를 금합니다.

Polars 데이터 분석 레시피 100제

(초판)

Evan 지음

Table of Contents

[Chapter 1. 책을 출간하면서 6](#_Toc190514698)

## [1. 책을 집필하면서 느끼는 소회 6](#_Toc190514699)

## [2. 무엇을 배울 수 있는가 6](#_Toc190514700)

## [3. 저자소개 7](#_Toc190514701)

## [4. 추천사 8](#_Toc190514702)

[Chapter 2. 개발환경설정 9](#_Toc190514703)

## [1. Github 회원가입 9](#_Toc190514704)

[Chapter 3. Python 기본문법 10](#_Toc190514705)

## [1. Python Basic Tutorial 10](#_Toc190514706)

## [A. Python 프로그래밍 시작 10](#_Toc190514707)

취업 준비생 분들에게 이 책을 바칩니다.

IT/빅데이터로 전업하고자 하는 많은 사람이 있습니다. 이 글의 저자, Sara, Evan도 전공자는 아니었습니다. 열심히 독학을 하면서, 이 분야로 진입하기 위해 매우 열심히 공부를 하며, 취업 준비를 하고 있습니다.

Evan은 IT/빅데이터와 관련된 첫 직장을 구하기 위해 100여군데의 이력서를 썼고, 같은 회사에 3년 내내 지원하고 떨어지기기를 반복한적도 있습니다. 정규직도 아니고 계약직에. 그래서, Evan은 취업 준비생의 고뇌와 고충을 매우 잘 이해하고 있습니다. 2011년 필리핀에서 만났던 그들도 생존의 위협을 느꼈듯이, 현재 취업 준비생들도 생존의 위협을 느끼고 있음을. Evan 역시 생존의 위협속에서 매일 하루 생존하기 위해 열심히 살고 있습니다.

Sara의 블로그 타이틀은 Grit 입니다. Grit의 핵심은 열정과 끈기, 단 3-6개월이 아닌 3-6년간의 노력이라고 합니다. 이 책이 취업 준비생 분들의 직업의 성공, 인생의 성공에 하나의 디딤돌이 되는 좋은 입문서가 되기를 기원합니다.

'취업 준비생 분들에게 이 책을 바칩니다

# Chapter 1. 책을 출간하면서

## 책을 집필하면서 느끼는 소회

## 무엇을 배울 수 있는가

## 저자소개

Evan

INTJ형 사람으로 가끔 INFJ도 나올 때도 있다. 혼자 있는 걸 좋아하고, 주목받는 걸 싫어하여 Evan으로 출판을 하고 있다. 학점은행 경영학사(2005-2009)를 졸업하였고, 선교와 NGO에 관심이 많아서 ASIA PACIFIC NAZARENE THEOLOGICAL SEMINARY 신학대학에서 석사학위(2009-2013)를 받았고, 한동대학교에서 국제개발협력대학원에서 석사학위(2014-2021)를 받았다. 그 후, 삶의 진로가 여러 번 바뀐 끝에 IT에 관심이 생겼고, 2016년 빅데이터 열풍과 함께 R, Python, SQL을 독학한 뒤, 마케팅, SI, 핀테크 업체에 머물며 3-4년간 경력을 쌓았다. 우연찮은 기회에 국방부와 육군본부에서 강의를 하면서, 강의 경력을 쌓기 시작하였고, 2020년 한동대학교에서 석사학위를 받기 위해 프리랜서로 전향한 뒤, 관련 분야에서 강의를 해오고 있다. 부족한 강의를 들은 친구들이 포트폴리오를 만들어서 인턴, 경력을 거쳐서 유수 좋은 기업에 이직하는 것을 지켜보며 삶의 만족을 느끼고 있으며, 앞으로 국내 기업 외에 글로벌 회사에도 취업하는 것을 돕고 싶은 비전이 있다.

최근에는 멀티캠퍼스, 한국표준협회, 성신여대 등 다양한 기관에서 강의를 진행하고 있으며, ADsP 자격증 강의, 태블로 강의, 단기 속성 강의로 대학교에서도 강의를 진행하면서 점차 강의 역량을 계속 키우고 있다. 배우는 것을 좋아하는 Evan은 전문적인 지식을 습득하고자 2023년부터 국민대학교 비즈니스 IT전문대학원에 박사과정에 재학중이며 박사과정 중에 배운 좋은 내용도 잘 소화해서 취업 준비를 하는 수강생들에게 전하고자 한다.

저서로는 파이썬으로 캐글 뽀개기 파이썬으로 시작해서 포트폴리오로 취업까지(비제이퍼블릭, 2021). 필리핀 스타트업의 기업가적 지향성과 기업성과에 관한 연구 : 사회적 자본의 매개 효과 (한국벤처창업학회, 2020). 학위논문은 The Mediating Effects of Social Capital in the Relationship between Entrepreneurial Orientation and Startup Performance : The cases from the Philippines (한동대학교 국제개발협력대학원, 2021) 등이 있다.

블로그 : <https://dschloe.github.io/>

## 추천사

# **Chapter 2. 개발환경설정**

## Google Colab

본 교재의 독자는 기본적으로 Pandas를 활용할 수 있다고 가정하므로, 별도의 Python 개발환경 설정에 대한 상세한 설명은 생략한다. 별도의 Python 개발환경 설정 설명은 생략한다.[[1]](#footnote-1) 다만, 원활한 실습을 위해 Google Colab에서도 실행할 수 있도록 코드를 구성하였다. Google Colab은 브라우저 기반의 Jupyter Notebook 환경으로, 별도의 설치 없이 Python을 실행할 수 있는 장점이 있다. 또한, 실행 환경이 클라우드에 존재하므로 로컬 컴퓨터의 환경 설정과 무관하게 일관된 실행 결과를 얻을 수 있다.

**Colab에서 패키지 설치 및 업그레이드**

본 교재에서 활용하는 주요 패키지는 **Pandas**와 함께 **Polars**이다. Polars는 Pandas보다 속도가 빠르고 메모리 사용량이 적어 대량의 데이터를 처리하는 데 유리하다. Colab에서 최신 버전을 사용하려면 아래 명령어를 실행한 후 세션을 다시 시작해야 한다.Google Colab이나Jupyter Notebook에서 라이브러리를 설치 할 때는 ‘!’를 앞에 명시적으로 붙여줘야 한다.

|  |
| --- |
| !pip install polars==1.22 --quiet |

### 위 명령어를 실행한 후, [런타임] → [세션 다시 시작] 을 선택하면 변경된 버전이 적용된다.

**패키지 버전 확인**

설치된 패키지의 버전을 확인하려면 다음 명령어를 사용하면 된다. 2025년 2월 기준, Google Colab의 기본 Polars 버전은 **1.9.0**이다. 최신 버전의 변경 사항이 필요할 경우 공식 문서를 참고하면 된다.

|  |
| --- |
| import polars as pl  pl.\_\_version\_\_  [결과]  ‘1.22.0’ |

## 주요 라이브러리

본 교재에서 활용한 주요 라이브러리는 다음과 같다.

파일명 : requirements.txt

|  |
| --- |
| polars==1.22.0  pandas==2.2.3  numpy==2.0.2  matplotlib==3.9.4  seaborn==0.13.2  scikit-learn==1.6.1  graphviz==0.20.3  statsmodels==0.14.4  jupyterlab==4.3.5  pyarrow==19.0.0  deltalake==0.24.0  xlsx2csv==0.8.4  xlsxwriter==3.2.2  openpyxl==3.1.5  xlrd==2.0.1 |

데이터 분석을 수행할 때 사용되는 라이브러리는 매우 다양하다. 본 교재에서는 **데이터 처리, 분석, 시각화, 머신러닝**을 포함한 다양한 작업을 수행할 수 있도록 필요한 패키지를 정리하였다. 특히, 로컬 환경에서도 동일한 코드를 실행할 수 있도록 주요 라이브러리의 버전 정보를 명시하였다. 각 라이브러리에 대한 설명은 다음 표에서 대략적으로 확인 할 수 있다.

|  |  |
| --- | --- |
| 라이브러리 | 주요기능 |
| polars | 대용량 데이터 처리를 위한 고속 데이터프레임 연산 |
| pandas | 전통적인 데이터 분석 및 테이블 형태 데이터 조작 |
| numpy | 벡터 및 행렬 연산을 위한 과학 연산 라이브러리 |
| matplotlib | 기본적인 데이터 시각화 도구 |
| seaborn | 통계적 데이터 시각화 |
| scikit-learn | 머신러닝 모델 구축 및 평가 |
| graphviz | 그래프 및 네트워크 시각화 |
| statsmodels | 통계적 분석 및 회귀 모델링 |
| jupyterlab | Jupyter Notebook의 확장형 개발 환경 |
| pyarrow | Apache Arrow 기반의 고속 데이터 처리 |
| deltalake | Delta Lake 지원 (데이터 레이크 관리) |
| xlsx2csv | XLSX 파일을 CSV로 변환 |
| xlsxwriter | Excel 파일 생성 및 편집 |
| openpyxl | Excel 파일 읽기/쓰기 |
| xlrd | XLS 파일 읽기 지원 |

각 개인 Local에서 개발환경을 설정 할 때는 프로젝트 경로에서 다음과 같이 명령어를 실행하면 된다.

|  |
| --- |
| pip install -r requirements |

**패키지 버전 확인**

주요 라이브러리의 설치 버전을 확인하고 싶다면 다음과 같은 코드를 작성하면 확인할 수 있다.

|  |
| --- |
| import pkg\_resources  # requirements.txt 파일에 명시된 패키지들의 설치 여부와 버전을 확인하는 코드  # 각 패키지에 대해 현재 환경에 설치된 버전을 출력하며, 설치되지 않은 경우 메시지 표시  packages = [  'polars', 'pandas', 'numpy', 'matplotlib', 'seaborn',  'scikit-learn', 'graphviz', 'statsmodels', 'jupyterlab',  'pyarrow', 'deltalake', 'xlsx2csv', 'xlsxwriter', 'openpyxl', 'xlrd'  ]  # 각 패키지를 순회하면서 버전 정보 확인  for package in packages:  try:  # pkg\_resources를 사용하여 설치된 패키지의 버전 정보 조회  version = pkg\_resources.get\_distribution(package).version  print(f"{package}: {version}")  except pkg\_resources.DistributionNotFound:  # 패키지가 설치되지 않은 경우 예외 처리  print(f"{package}: 설치되지 않음")  [결과]  polars: 1.22.0  pandas: 2.2.3  numpy: 2.0.2  matplotlib: 3.9.4  seaborn: 0.13.2  scikit-learn: 1.6.1  graphviz: 0.20.3  statsmodels: 0.14.4  jupyterlab: 4.3.5  pyarrow: 19.0.0  deltalake: 0.24.0  xlsx2csv: 0.8.4  xlsxwriter: 3.2.2  openpyxl: 3.1.5  xlrd: 2.0.1 |

# Chapter 3. Python 기본문법

## 레시피 1

### Python 프로그래밍 시작

프로그래밍의 첫번째 시작은 “Hello World!”를 출력하는 것에서부터 시작한다. Python에서는 다음과 같이 print()함수를 사용한다.

|  |
| --- |
| print("Hello, Streamlit!")  [결과]  Hello, Streamlit! |

Table 1. Python Data Structures

|  |  |
| --- | --- |
| Primitive | Non-Primitive |
| Integer  Float  String  Boolean | Array  List  Tuple  Dictionary  Set  File |

1. 개발환경설정이 궁금하다면 다음 필자의 유투브에 접속해서 확인한다. https://youtu.be/BqkwJ5zCMb0?si=ylqt185WSSPJ-x9q [↑](#footnote-ref-1)