

마케팅 어널리틱스 스터디

Session1 | Chapter 1&2

채충일



温州肯恩大学
WENZHOUCHEAN UNIVERSITY

2022/01/01 (updated: 2022-01-07)



파이썬언어

파이썬은 범용 프로그래밍 언어. 단순한 구문과 뛰어난 가독성 덕분에 프로그래밍 교육뿐만 아니라 데이터 분석 및 데이터 과학에서 풀 스택 웹 개발에 이르기까지 모든 종류의 프로그래밍 응용 프로그램에서 점점 더 많이 선택되는 언어가 되었음.

언제 파이톤 언어를 쓰는가?

- 다른 곳에서 사용할 수 있는 것보다 더 최신이거나 더 강력한 방법에 대한 액세스가 가능함.
- 분석을 여러 번 실행해야 할때 용이.
- 여러 데이터 세트에 분석을 적용해야 할때
- 새로운 분석 기법을 개발해야 하거나 기존 분석법에 대한 완벽한 제어가 필요할때

교재에서 사용하는 데이터

이 책에서 분석하는 대부분의 데이터세트는 시뮬레이션된 데이터세트

중요 온라인 리소스

이 책의 내용과 관련하여 제공되는 공식적인 리소스는 다음과 같습니다:

- An information website: <https://python-marketing-research.github.io>
- A Github repository: <https://github.com/python-marketing-research/python-marketing-research-1ed>
- The Colab Github browser: <https://colab.sandbox.google.com/github/python-marketing-research/python-marketing-research-1ed>

에러가 발생하는 주요원인

- 파이썬: 들여쓰기, 괄호, 대괄호, 대문자 또는 소문자명령이 길면 조각으로 분해하고 다시 작성
- 패키지 (추가 라이브러리): 패키지는 정기적으로 업데이트됨. 때때로 작동 방식이 변경되거나 잠시 동안 전혀 작동하지 않을 수 있음. 일부는 매우 안정적이고 다른 일부는 자주 변경됨.
- Python 경고 및 오류: Python "경고"는 종종 정보 제공이며 반드시 수정이 필요한 것은 아님. 때때로 패키지가 업데이트되면서 왔다가 사라지기도 하지만 코드와 함께 발생하면 이를 호출합니다. Python에서 "오류"가 표시되면 문제가 발생하여 수정해야 함을 의미.
- 데이터: 책에서 제공된 데이터 세트는 시뮬레이션되고 난수 시퀀스의 영향을 받음. 생성한 데이터가 약간 다르면 처음부터 다시 시도
- 모델: 통계적 추정치가 달라질 수 있는 세 가지 요인이 있음. 데이터의 약간의 차이(이전 항목 참조), 추정치를 약간 다르게 만드는 패키지 변경, 무작위 샘플링을 사용하는 통계 모델. 모델을 실행하고 결과가 매우 유사하지만 약간 다른 경우 이러한 상황 중 하나가 발생했다고 가정할 수 있음
- 출력: 패키지는 보고하는 정보를 변경하는 경우가 있음. 이 책의 출력은 작성 당시의 최신 정보이지만 일부 패키지는 시간이 지남에 따라 내용을 약간 다르게 보고할 것으로 예상할 수 있음.
- 찾을 수 없는 이름: 때때로 패키지는 사용하는 함수 이름이나 결과 구조를 변경. 통계 모델에서 무언가를 추출하려고 할 때 코드 오류가 발생하면 모델의 도움말 파일을 확인

파이선 스크립트의 실행



- 스크립트를 실행
- 인터랙티브 모드에서 실행 (notebook 포함)
- Google Coaboratory

2장은 대부분은 Python의 기초

- Python을 실행하기 위한 여러 환경(CLI, IDE, 스크립트 및 노트북)이 존재함. 노트북(jupyter) 사용을 권장하지만 가장 중요한 것은 자신에게 잘 맞는 환경을 찾는 것
- 컬렉션 인덱싱은 Python, 특히 NumPy에서 매우 강력함.
- 코드의 재현성과 안정성을 향상시키는 함수를 작성하는 것은 Python의 표준 관행임
- 많은 Python 기능은 pip를 사용하여 설치할 수 있는 패키지에 있음
- NumPy 및 pandas는 NumPy 배열 및 pandas의 형태로 Python으로 중요한 데이터 과학 분석을 가능하게 함
- 누락된 값은 명시적으로 처리해야 함
- help() 함수를 사용하여 유용한 정보를 찾을 수 있음
- Python 개체는 pandas를 사용하여 파일을 데이터 프레임으로 읽고 쓰는 기능을 제공함
- 데이터를 CSV로 쉽게 가져오고 내보낼 수도 있음
- 쓸모없는 변수로 인한 혼란과 버그를 피하기 위해 작업 공간을 정기적으로 정리할 필요가 있음

References



Chapman, C. and E. M. Feit (2019). R for marketing research and analytics. 2nd edition: © Springer Nature Switzerland AG 2019. Vol. 67. Springer.

Schwarz, J. S., C. Chapman, and E. M. Feit (2020). Python for marketing research and analytics. Springer.

스터디 참여의 동기배이션

Var1	Freq
주제분야에 대한 학습	4
참여자들과의 인터엑션을 통해 현황 파악	2



Python 속련도

Var1	Freq
1	2
2	1
4	3

R 숙련도

Var1	Freq
1	1
2	1
3	2
5	2

현업에서 주로 쓰는 프로그램이나 툴은?

Var1	Freq
Google Analytics, Adobe Analytics	1
Python	1
R	1
R, SQL, EXCEL	1
엑셀 및 오피스 제품군	2

내 업무는 분석과 주로 밀접하게 관련되어 있다

Var1	Freq
3	2
4	1
5	3

내 업무는 마케팅/소비자분석에 밀접하게 관련되어 있다.

Var1	Freq
3	2
4	1
5	3

마케팅 분석방법론에 대한 나의 숙련도는

Var1	Freq
고급 마케팅 기법활용	3
기초통계	3

범주 양기 반 가능 공 끝 다 루 두
 설도 유 주 관 련 R 것 네 트 워 킹

