

# 경상대학 빅데이터 분석 연구모임

이 정 우, 김 철 민

한양대학교 지식서비스연구소

# Facilitator

- 이정우, 김철민, 박현지 박사과정



# 참여자가 가져야할 자세

- 파이썬을 배우는걸 목표로 삼지말고  
관심있는 문제를 찾고 그걸 해결하는데 파이썬을 쓰세요.



[Jake VanderPlas](#)

@jakevdp

팔로우

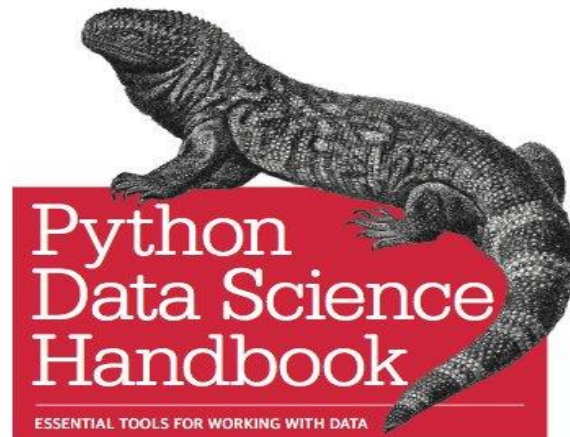
My advice on learning Python:

Don't set out to "learn Python". Choose a problem you're interested in and learn to solve it with Python.

🌐 영어 번역하기

오전 5:24 - 2017년 9월 10일

O'REILLY



powered by  
jupyter

Jake VanderPlas

# 모임 목표

- 데이터 분석 SW Language 습득
  - 데이터 분석용 python language 및 library 활용법 습득
- 실제 문제에 대한 데이터 분석적 접근 (보고서 작성)
  - 학교 안팎의 **문제 발견 및 해결**(Problem Based Learning)
    - 예시1) '경상관 1층 몽랩 카페의 데이터' 분석
    - 예시2) '몰테일(미국 직구) 네이버카페 게시글'에 대한 텍스트 분석
    - 기타 등등 자유 주제

# 모임 기본 운영 방식

- 활동 공간: 경상관 1층 PBL 강의실 107호
- 활동 주기: 주 1회 [매주 수요일 오후 1 ~ 3시]
- 활동 방식: 교재 스터디 모임 & 초청 강연 참석
  - (수업 1시간 전부터: 스터디 자율학습 시간)
  - (10분간) 당일 내용에 대한 쪽지시험 (시험 후 자신이 직접 채점)
  - (1시간 50분간) 개별적으로 할당된 부분에 대한 발표

# 모임 일정 계획

회차	순번	일자	내용
8	1	2017-09-13	신입멤버 오리엔테이션
9	2	2017-09-20	스터디1 (환경설치 및 Jump2Python 2장)
10	3	2017-09-27	특강1 문석현 - (경영자료분석 수업 학생들 다수 참석)
11	4	2017-10-11	특강2 공민서 - (스터디 모임 학생들만 참석)
12	5	2017-10-18	스터디2 (Jump2Python 3장 및 4장)
13	6	2017-10-25	특강3 신윤식 - (스터디 모임 학생들만 참석)
14	7	2017-11-01	스터디3 (Pandas 4장: Numpy 기본)
15	8	2017-11-07*	특강4 이진석 - (스터디 모임 학생들만 참석)
16	9	2017-11-08	스터디4 (Pandas 5장: Pandas 시작하기)
17	10	2017-11-15	특강5 임송이 - (경영자료분석 수업 학생들 다수 참석)
18	11	2017-11-22	스터디5 (Pandas 6장: 데이터 로딩, 저장, 파일 형식)
19	12	2017-11-29	스터디6 (Pandas 7장: 데이터 준비하기)
20	13	2017-12-06	스터디7 (Pandas 8장: 도식화와 시각화)
21	14	2017-12-13	스터디8 (Pandas 9장: 데이터 수집과 그룹 연산)
22	15	2017-12-27	최종결과물 발표 (데이터분석 기획서)

# 교재 분량 할당

## • 최종 멤버 확정 후 공지

회차	일자	Jump2python 책	절	발표자
8	2017-09-20	2장	36~54p	
			54~65p	
			3절(리스트)	
			4절	
			5절, 7절	
12	2017-10-18	3장	1절(if문)	
			2절(while문)	
			3절(for문)	
		4장	1절	
			2절, 3절	

회차	일자	Pandas 책	절	발표자
14	2017-11-01	4장	1절	
			2절~	
16	2017-11-08	5장	1절	
			2절	
			3절, 4절	
			5절	
18	2017-11-22	6장	~236p, 3절	
20	2017-11-29	7장	1절, 2절	
			3절	
			4절, 5절	
21	2017-12-06	8장	1절	
			2절	
			3절, 4절	
21	2017-12-13	9장	1절, 2절	
			3절	
			4절, 5절	

# 교재(제공됨)

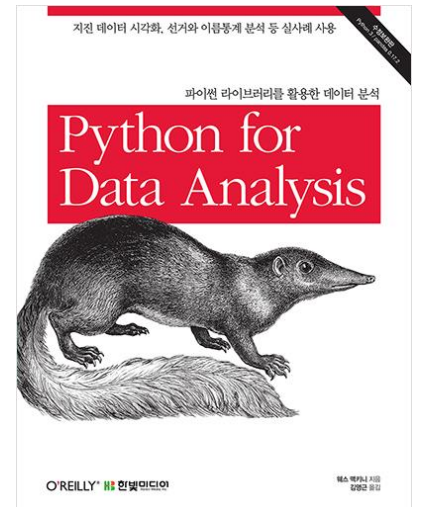
- (기본) 점프 투 파이썬

- 파이썬 기본



- (응용) 파이썬 라이브러리를 활용한 데이터 분석

- 파이썬을 활용한 데이터 분석





# 초청 강연

## • 데이터 분석 분야 전문가 5인 초청

9월 27일 (수) 13~14시 30분(90분) : 문석현 데이터경영연구소 소장

- 주제: 디지털 비즈니스의 열쇠는 데이터에 있다

10월 11일 (수) 13~15시(120분) : 공민서 엔트로피(주) Data Scientist

- 주제: 맛집 지도정보 시각화 및 맛집 추천 알고리즘 구현하기

10월 25일 (수) 13~15시(120분): 신윤식 (주)블루웨일 Service Planning & Analyst

- 주제: 금융 데이터의 로지스틱 회귀분석을 통한 우량채권 특성 발견하기

11월 7일 (화) 13~15시(120분): 이진석 AskDjango 운영자, MS Azure MVP

- 주제: 카카오톡 플러스친구 챗봇 만들기

11월 14일 (수) 13~15시: 임송이 SBS 데이터저널리즘팀

- 주제: 데이터 시각화와 기획하기

수료를 위해선  
4회 이상 참석



# 개인 발표 자료

## • 기본 양식(ppt)을 활용한 PPT 자료로 발표

### if문의 개념

"돈이 있으면 택시를 타고, 돈이 없으면 걸어 간다."

프로그래밍에서 조건을 판단하여 해당 조건에 맞는 상황을 수행하는 데 쓰이는 것이 바로 **if문**이다.

```
>>> money = 1
>>> if money:
...     print("택시를 타고 가라")
... else:
...     print("걸어 가라")
...
택시를 타고 가라
```

### 조건문이란 무엇인가?

앞에서 살펴본 택시 예제를 다음처럼 바꾸려면 어떻게 해야 할까?

"만약 3000원 이상의 돈을 가지고 있으면 택시를 타고 그렇지 않으면 걸어 가라"

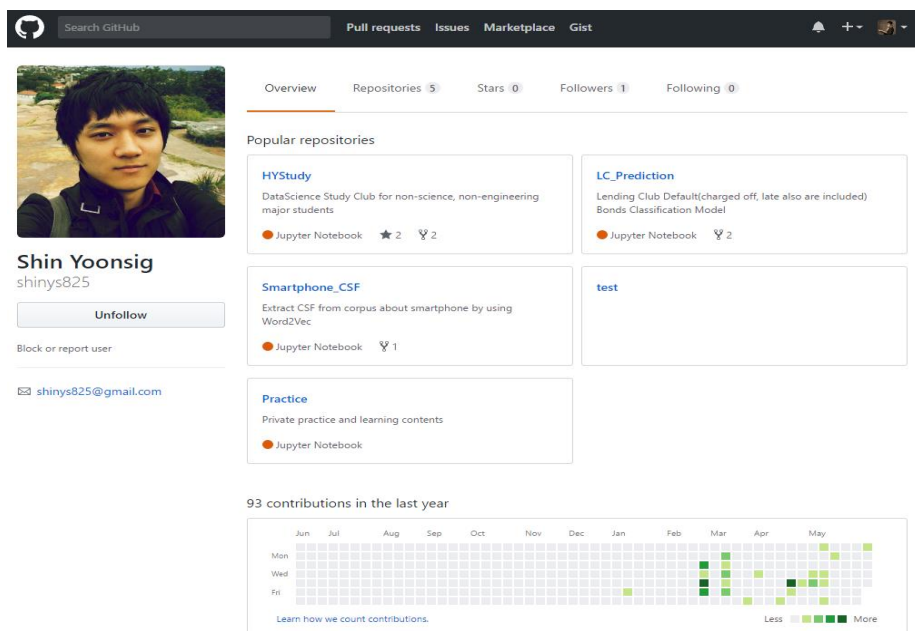
위의 상황은 다음처럼 프로그래밍할 수 있다.

```
>>> money = 2000
>>> if money >= 3000: 거짓
...     print("택시를 타고 가라")
... else:
...     print("걸어가라")
...
걸어가라
>>>
```

# 개인 포트폴리오

## • 데이터 분석 포트폴리오

### ■ 깃허브(Github)



Shin Yoonsig  
shinys825

Unfollow

Block or report user

shinys825@gmail.com

Overview Repositories 5 Stars 0 Followers 1 Following 0

Popular repositories

- HYStudy**  
DataScience Study Club for non-science, non-engineering major students  
Jupyter Notebook ★ 2 🍏 2
- LC\_Prediction**  
Lending Club Default(charged off, late also are included) Bonds Classification Model  
Jupyter Notebook 🍏 2
- Smartphone\_CSF**  
Extract CSF from corpus about smartphone by using Word2Vec  
Jupyter Notebook 🍏 1
- test**
- Practice**  
Private practice and learning contents  
Jupyter Notebook

93 contributions in the last year

Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May

Mon  
Wed  
Fri

Learn how we count contributions.

Less More

[https://github.com/Mosunghoon/HY\\_DATA-STUDY](https://github.com/Mosunghoon/HY_DATA-STUDY)

예시: 코드 및 실행화면 미리보기

#### Scatter plot

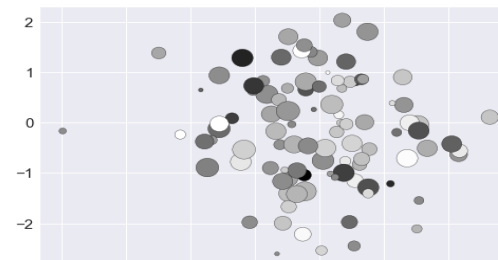
• Doc: [https://matplotlib.org/api/pyplot\\_api.html#matplotlib.pyplot.scatter](https://matplotlib.org/api/pyplot_api.html#matplotlib.pyplot.scatter)

```
In [4]: x = np.random.randh(100)
y = np.random.randh(100)

# make points on coordinate(x, y)
plt.scatter(x, y)
plt.show()
```



```
In [5]: # make points on coordinate(x, y) with style
plt.scatter(x, y,
            s=np.random.randint(10, 500, 100), # size
            c=np.random.randh(100), # color
            edgecolors='black') # edge color
plt.show()
```

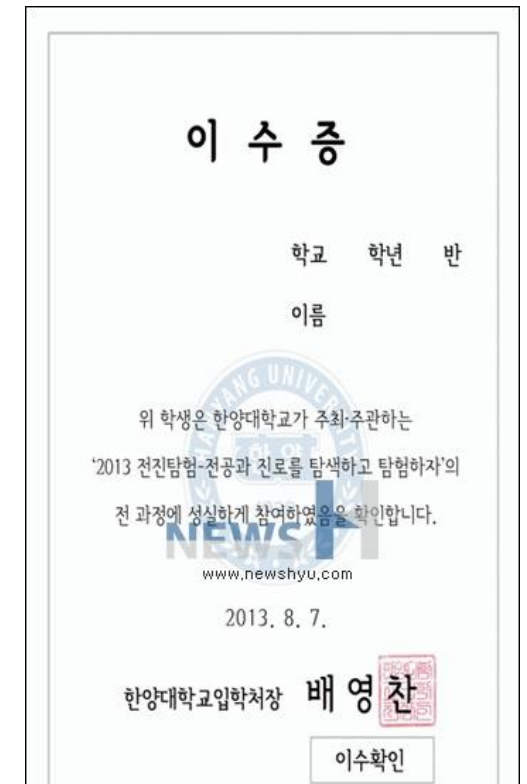


# 데이터분석 보고서

- 12월말까지 최종결과물 수행(아래 내용 반영)
  - 문제 정의
    - 자기가 분석하고 싶은 문제나 내용 ex) 몰테일 네이버 게시판 사용자 의견 분석
  - 데이터 수집
    - 방식: 직접 설문, 웹크롤링, Kaggle 데이터 다운로드 등
  - 문제를 해결할 데이터 분석 기법과 라이브러리 조사
    - 10월 중순 쯤 샘플 제시(ppt)
    - 진도에 따라 실제 분석까지 수행 가능

# 이수증

- 경상대학 학장 명의 이수증 발급
  - 이수증 취득 조건
    - 총 5회 세미나 중 4회 이상 참석
    - 총 8회 스터디 중 6회 이상 참석
    - 최종 결과물(Github 관리 및 데이터분석 보고서)



예시

# 참여 문의

경영학부 + 경영학 다중전공자

- 이정우 연구원 [jungwoolee@hanyang.ac.kr](mailto:jungwoolee@hanyang.ac.kr)
- 김철민 연구원 [friend100n@gmail.com](mailto:friend100n@gmail.com)

혹은...

경상관 529호 [지식서비스연구소](tel:031-400-5038) (031-400-5038)