

Chapter 00
Orientation

오리엔테이션

D.COM
DATA-SCIENCE

이번 챕터에서는!

- 1. 데이터 사이언스란 무엇인지**
- 2. 관련하여 무엇을 배워야할지**
- 3. 스터디는 어떻게 진행되는지**

알아봅니다!

1. 데이터 사이언스!

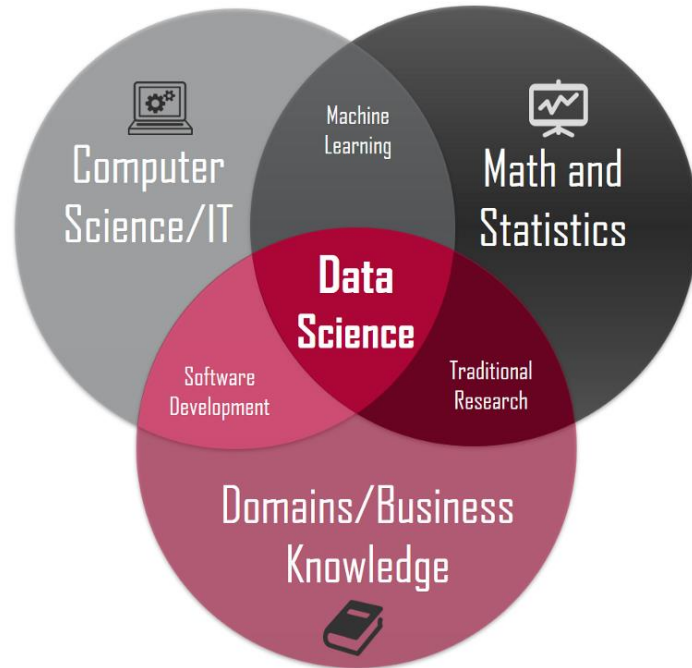
스터디를 시작하기에 앞서
데이터 사이언스란 무엇일까요?

1. 데이터 사이언스!

아래 글을 읽어봅시다.
“데이터 사이언스, 내가 해도 괜찮을까?”

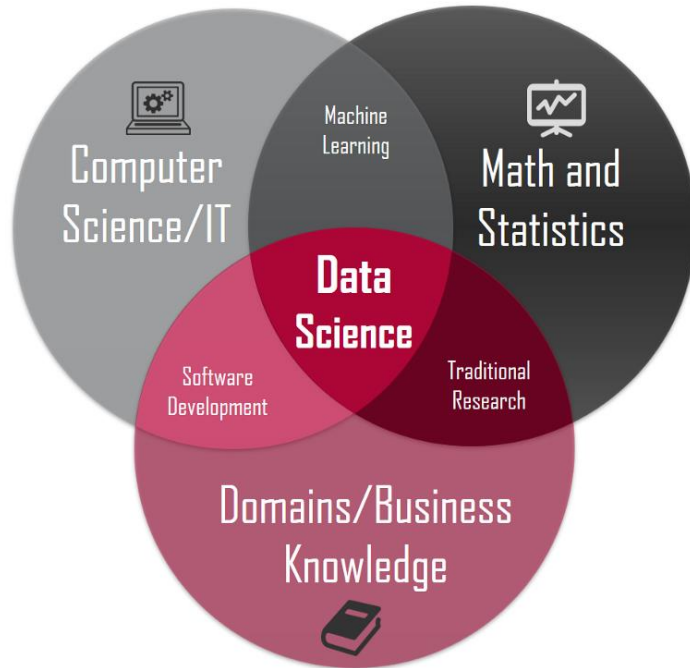
<https://data-artist.tistory.com/2>

1. 데이터 사이언스!



데이터 사이언스는
Drew Conway의 정의에 따르면
수학/통계학 + 컴퓨터과학 + 도메인 지식을
합친 분야라고 합니다.

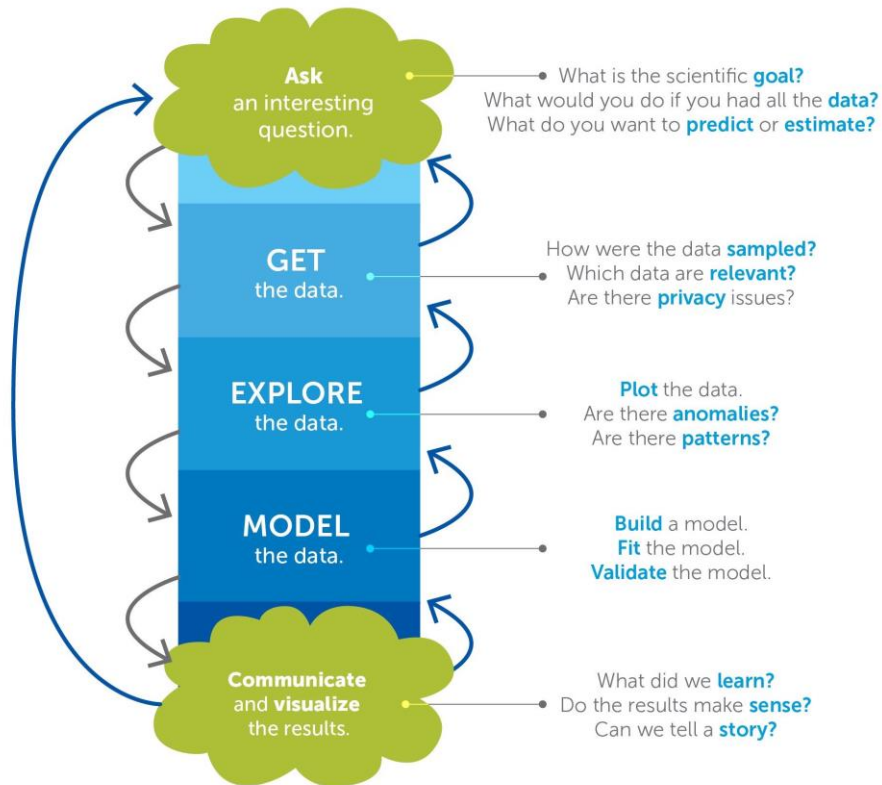
1. 데이터 사이언스!



간단하게, 데이터 사이언스는
앞의 세 요소를 바탕으로
데이터를 통해 내가 원하는 작업을 수행하는
학문입니다.

1. 데이터 사이언스!

The Data Science Process



 Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

데이터 과학은 크게 **5단계**로 이루어져 있습니다.

1. **ASK** (질문하기)
2. **GET** (데이터 수집하기)
3. **EXPLORE** (데이터 탐색하기)
4. **MODEL** (모델링하기)
5. **COMMUNICATE and VISUALIZE** (활용하기)

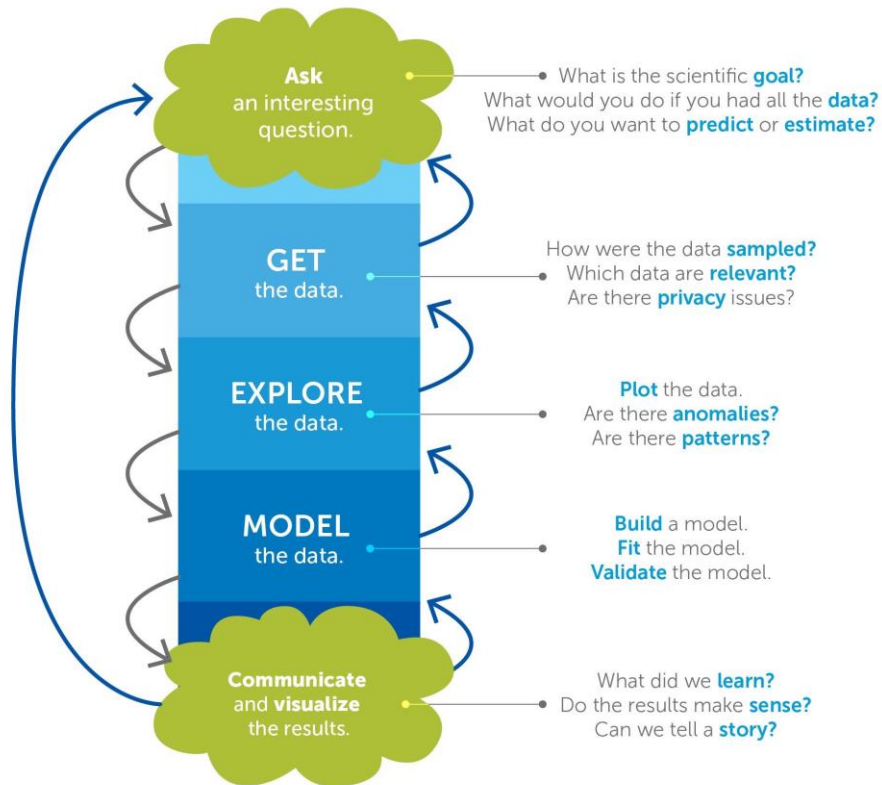
1. 데이터 사이언스!

데이터 과학 5단계를 자세하게 설명한 글입니다.
각 단계에서 어떤 일을 하는지 주목하며 읽어봅시다!

<https://brunch.co.kr/@data/10>

1. 데이터 사이언스!

The Data Science Process



 Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

이번 스터디에서는 **아래 3 단계**에 대해 다룹니다.

1. **ASK** (질문하기)
2. **GET** (데이터 수집하기)
3. **EXPLORE** (데이터 탐색하기)
4. **MODEL** (모델링하기)
5. **COMMUNICATE and VISUALIZE** (활용하기)

1. 데이터 사이언스!

Excel, R, Spark 등 수 많은 데이터 분석 언어가 있습니다.
이번 스터디에서는 “Python”을 사용합니다.

1. 데이터 사이언스!

R VS Python

데이터 분석 툴의 양대 산맥입니다.
두 언어의 장단점은 어떻게 될까요?

<https://media.fastcampus.co.kr/knowledge/dataanalysis-python-r/>

1. 데이터 사이언스!

스터디 실습 환경

본 스터디에서 실습과 과제는
ipynb 확장자로 배부됩니다.
따라서 **jupyter notebook**이나 **colab**을
사용을 추천합니다.

colab에 대한 설명은 아래 링크를 참고해주세요!

<https://theorydb.github.io/dev/2019/08/23/dev-ml-colab/>

1. 데이터 사이언스!

스터디 방식

스터디는 3단계로 이루어집니다!

CONCEPT

매주

PPT를 통한 개념 설명

PRACTICE

매주

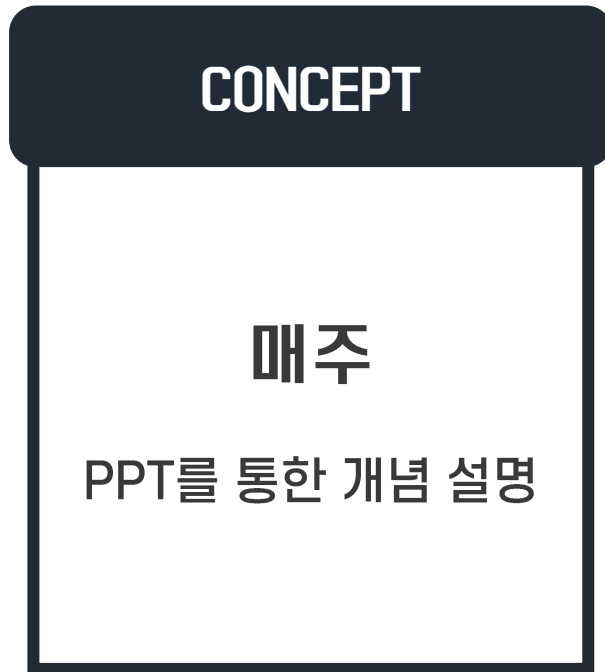
연습문제를 풀어
구글 드라이브에 제출

PROJECT

각 단원마다

케글 프로젝트 진행

1. 데이터 사이언스!



각 단원마다
Concept PPT가 배부됩니다.

이번 데이터 사이언스 스터디의
Concept PPT에는 학습 진행에
필요한 로드맵과 개념 설명이
담겨있습니다.

각 모듈의 구체적인 사용법 등은 다루지 않습니다.

즉, 다음 단계를 해결하기 위해서는
적극적인 구글링이 필요합니다.

1. 데이터 사이언스!

PRACTICE

매 주

연습문제를 풀어
구글 드라이브에 제출

각 단원마다
실습과 연습문제가 ipynb 파일로 배부됩니다.
연습 문제는 모듈 사용법부터 미니프로젝트까지
다양한 유형으로 출제됩니다.

완성된 ipynb 파일을
구글 팀 드라이브에
제출해 주시면 됩니다.

제출한 과제들은 스터디원 모두에게 공개되어 있습니다.
제출 후 다른 사람의 코드와
자신의 코드를 비교해 보아도 좋습니다!

1. 데이터 사이언스!

PROJECT

각 단원마다

캐글 프로젝트 진행

각 단원마다
익힌 내용들을 활용하여 해결할 수 있는
Kaggle Project가 안내됩니다.

본 프로젝트 완성한 후 코드를
구글 드라이브에 업로드 해주시면 됩니다.

Kaggle에 대한 자세한 소개는
아래 링크를 참고해주세요.

<https://m.blog.naver.com/PostView.nhn?blogId=ckdgus1433&logNo=221443513440&proxyReferer=https:%2F%2Fwww.google.com%2F>

1. 데이터 사이언스!

스터디 커리큘럼

본 스터디의 커리큘럼은 아래 링크를 참고해주세요.

<https://github.com/Dcom-KHU/2020-DataScience-Study>

1. 데이터 사이언스!

스터디 QNA

D.COM 슬랙에서
#datascience 채널을 자유롭게
이용해 주시면 됩니다~!