Chpt 1. 데이터 1도 몰랐던 멍청한 눈 삽니다.

안녕하세요. 처음 뵙겠습니다.
앞으로 한달간 함께할 제이름은 김덕중입니다. 편하게 DJ라고 불러주세요

대충 이렇게 생겨먹었습니다.jpg

Part 1. 그대의 데이터 눈 “뜸”을, 응원합니다.

새로운 세계에 오신 여러분 환영합니다.
데이터 분석, 머신러닝이라는 거창하고 어려운 말들을 하기 전에
먼저 어떻게 전달해야 여러분께 쉽게 전달할까를 정말 고민했습니다.
고민한 끝에 다다른 결론이 이겁니다

> **만화<슬램덩크>의 가장 중요한 장면에는 대사가 없다.**

말한마디 없이 모든것들이 생생하게 잘 전달되죠.
이렇게 여러분과 함께 하려고 합니다. 누구나 직접 해보면 다 할수있고 재미있는 것들이기 때문에,
어렵고 복잡한 전문용어를 앞세우기보다는 이런 재미있는것을 여러분과 함께 직접 해보고 싶습니다. 그뿐입니다.

아래 요건을 다시한번 잘 봐주시기 바랍니다

- [o] 난 통계도 모르고 프로그래밍도 몰라
- [o] 머신러닝? 프로그래밍? 관심은 있어도 뭐부터 해야될지도 모르겠어
- [o] 머신러닝같이 어려운건 못해도 학업으로, 업무용으로 통계 기초정도는 알고 활용하고 싶어
- [x] 난 프로그래밍도 할줄알고 통계도 조낸 잘함 ~~그만해 미친놈들아...~~

해당하시는 분들만 들으시길 바라며, 아닌분들은 다른 좋은 강의를 찾으시거나, 이 강의 이후에 진행할 차후 강의를 들으세요
자 , 준비가 다 되셨으면 본격적으로 시작해 봅시다!
질문은 언제든 환영이니 언제든 질문해주세요~
새로운 세상에 눈을 뜨고 도전하는 여러분들을 응원합니다

Part 2. 시작했다면, 일단 반은 먹고 들어가고요.

시작이 반이니까 이미 여러분 반 하신겁니다 ~~구라즐~~
일단 예언하나 하면서 시작하겠습니다.

- **예언: 누구나 이부분을 건너뛰지만, 결국은 여길 다시 보게된다**

아닐거 같지만 현실임...
여기가 그만큼 중요하다는 말입니다.
이 강의에서 뿐만이 아니라 무엇인가를 할때는 이말로 시작하는것이 항상 핵심을 꿰뚫게 됩니다.

1. Why: 데이터분석, 그거 도대체 왜 하는겁니까?

가장 중요한 질문입니다만
아쉽게도 답이 정해져 있지 않습니다.
이쯤에서 본인이 본인을 정확히 한번 돌아보실 필요가 잇습니다
내가 지금 데이터 분석을 왜 하려고 하는가? 분석하려면 어떤게 필요한가??어떤 형식의 데이터를 사용할것인가??
이 질문이 중요한 이유는, 이것에 따라 사용하는 분석 방법이 달라지기 때문입니다.
이것을 모르고 뭐 데이터가 어쩌고 알고리즘이 어쩌고 머신 러닝이 어쩌고 떠들어봐야

아이고, 의미없다

앞으로도 이사항을 항상 명심하시고, 계속해서 고민하시면서 머리속에서 가져가시기 바랍니다

---

2. 왜 이제와서 시끄러운겁니까? & 3. 데이터분석과 머신러닝을 하려면 무엇이 가장 중요한가요?

정확한 지적입니다
사실 데이터 분석에서 쓰고있는 방법론이나 알고리즘의 기초들은 적게는 수십년에서 많게는 100년이 넘은 것도 있습니다
근데 왜 이제와서 난리냐면요...

출처: '하용호'님 슬라이드 쉐어

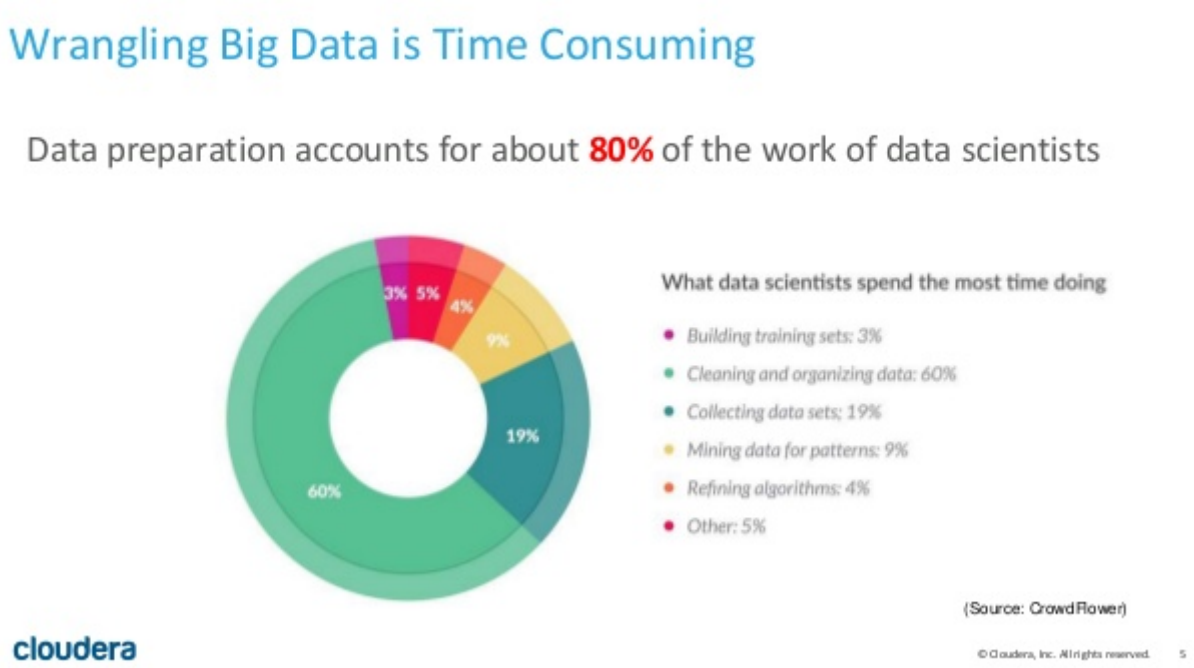
첫째, 데이터가 없었다
: 분석방법 만들어봤자, 그것을 활용할 데이터가 충분히 축적되지 못했었습니다
둘째, 기술력이 따라오지 못했음요
: 대량의 데이터가 있다 손 치더라도 그것을 물리적으로 처리해줄 하드웨어 및 소프트웨어가 부족했습니다. 즉, 인프라부족
셋째, 알고리즘
: 기존에도 분석방법론들이 있었지만, 특히 요즘와서야 딥러닝쪽 알고리즘이 제대로 다시 논의되고 연구되고 있습니다.

사실 위의 얘기는 굉장히 뼈가 있는 말입니다.
미칠듯한 성능의 CPU와 빠른연산의 GPU요? 다나와 ~~ppl아닌~~ 가서 돈으로 바르면 됩니다
알고리즘이요? 예전이야 별로 없었지 요즘들어 연구가 많이 되서, 구글링하면 소스코드까지 다나옵니다.
핵심은 사실....
DATA입니다!! (feat.분석방법론)
많은 사람들이 머신러닝이라는 **'단어'**에 혹해서, 고성능의 하드웨어와 최적화 알고리즘이 머신러닝의 핵심이라고 생각하고있습니다.
하지만 이것은 장담하건데, 틀린 방식입니다

데이터 분석에 가장 중요한것은 데이터입니다!!**

데이터를 분석해서 어떤것을 할것인가에 따라
'어떤 데이터를 어떻게 정의하여 쓸것인가, 처리는 어떻게 할것인가, 어떤 분석방법론을 사용할것인가'가 가장 중요합니다.

머신러닝 하면 어차피 다 되는거 아니냐? 라고 묻는 사람들에게 유명한 한마디 드리겠습니다.
"쓰레기를 넣으면 쓰레기가 나온다"



- 데이터 과학자들의 업무 비중, 보시다시피 필요한 데이터를 수집하고 가공하는데에만 전체 업무시간의 80%가량이 소요됩니다. data분석에서 가장 중요한것은 data입니다.
출처: Crowd Flower

---

4. 아무말 대잔치: 응용으로 기초하기

뭔가 이상하죠? 응용으로 기초 한다니...말이 안됩니다
근데 대부분의 데이터 분석 강의를 듣다보면 이러한 현상이 종종 발생됩니다

- **여러분들 보고 아주 획기적인 새로운 알고리즘을 만들어내라는 얘기는 아닙니다**
  : 하지만, 적어도 지금쓰고있는 데이터 방법론에 대하여 기본적인건 이해하고 분석결과가 안좋은 경우 그 원인을 알아야 합니다

기초를 모르고 응용만 해서는 절대로 문제의 근본을 해결할 수 없습니다.
머신러닝의 다양한 것들을 이해하기 위해서는 몇가지 근본적으로 알아야될 기초들이 있습니다.
그중 하나는, 전통적인 통계방법론, 그중에서도 특히 회귀의 기초는 알아야 합니다.
왜냐면, 사실 머신러닝은 엄청 복잡한 알고리즘으로 구성되어 있을뿐, **회귀분석의 일종**이기 때문입니다.

> **응용으로는 기초할수 없습니다. 하지만, 기초로는 응용할 수 있습니다.**