

/* elice */

프로젝트로 배우는 데이터 분석

데이터 분석 시작하기



엘리스

데이터 분석 프로세스

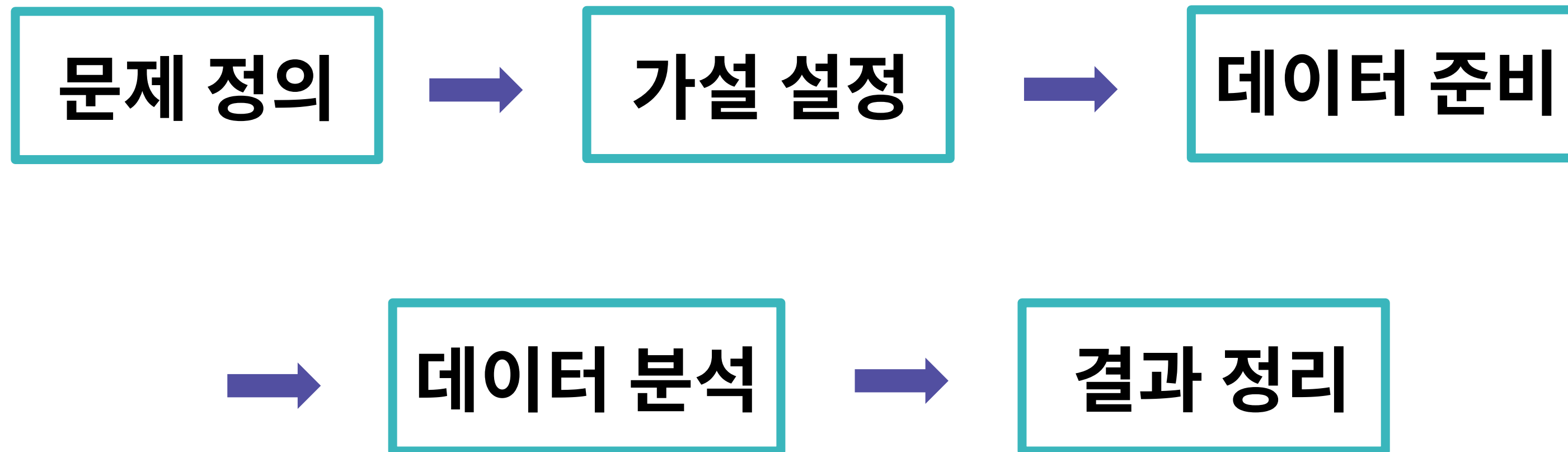
데이터 분석



데이터를 활용하여

원하는 정보를 얻어내기 위한 일련의 과정

데이터 분석 과정



문제 정의

현재 풀고자 하는 **문제**가 무엇인지를 **명확히** 정의

여행을 위해 가장 저렴한 가격으로 항공편을 예약하고자 한다.

→ 언제 비행기 표를 예약해야 할까?

가설 설정

문제를 **해결**하기 위한 데이터 분석의 토대인 가설 설정

가설 1

비수기에 구매하면 가장 저렴 할 것이다.

가설 2

새벽 시간에 구매하면 가장 저렴 할 것이다.

문제와의 **관련성** 고려

데이터 준비

풀고 싶은 문제에 대한 정보를 담고 있는 데이터셋을 선정

다양한 시간대 별 항공권 가격 데이터 수집 및 처리

데이터 수집 및 전처리 과정

데이터 수집

GMP 13:30	→ 01시간 10분	CJU 14:40	 할인석	성인 편도 33,900원
GMP 13:30	→ 01시간 10분	CJU 14:40	 일반석	성인 편도 111,500원
GMP 13:40	→ 01시간 25분	CJU 15:05	 특가석	성인 편도 27,000원
GMP 13:40	→ 01시간 25분	CJU 15:05	 특가석	성인 편도 32,000원

풀고 싶은 **문제에 대한 정보**를 담고 있는 데이터셋

데이터 전처리

데이터 정제(Data Cleaning)

빠진 부분, 중복, 이상값 제거, 형태 변환 등의

초기 데이터(Raw Data)전처리 시행

데이터 분석

본격적인 데이터 분석

데이터 분석 프로젝트의 성공여부는
얼마나 데이터를 이해하고 있느냐에 좌우됨

탐색적 데이터 분석(EDA)

탐색적 데이터 분석(EDA)

Exploratory Data Analysis

데이터의 특징을 찾고, 숨겨진 패턴을 발견하는 과정

결과 정리

분석 과정에서 알아낸 **인사이트(Insight)** 정리

인사이트(In + sight)

사물의 이면을 들여다보는 것

의미 있는 데이터 분석

분석에서 그치는 것이 아닌

문제 해결을 위한 의미 있는 데이터 분석

그러기 위해서는,

명확한 목표 설정과

목표에 맞는 흐름에 따른 데이터 분석 진행

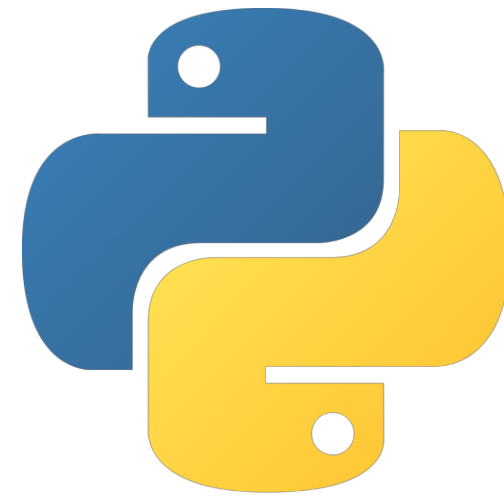
데이터 분석 도구

다양한 도구



데이터 분석을 위한 다양한 도구 존재

파이썬

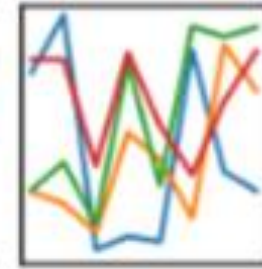


범용적이며 데이터 분석을 위한 라이브러리 지원

파이썬 라이브러리

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib

효과적인 데이터 처리 및 시각화 가능

/* elice */

[실습1]

월드컵 데이터 분석하기





`/*elice*/`

contact@elice.io