

05 데이터 분석 심화 - 탐색적 데이터 분석(EDA)

AI 에이전트 개발

데이터 분석 심화

원티드랩

- [소개](#)
- [데이터 분석 절차 및 고려사항](#)
 - [1. 데이터 불러오기](#)
 - [2. 데이터 정보 확인](#)
 - [3. 결측치 처리](#)
 - [4. 데이터 시각화](#)
 - [7. 결론](#)

소개

EDA(Exploratory Data Analysis)란 결측치 및 이상치를 탐지하여 전처리 방향을 결정하고, 통계 요약 및 시각화를 통해 데이터의 분포, 변수 간 관계 등 핵심 특성을 파악하는 과정이다.

데이터 분석 절차 및 고려사항

1. 데이터 불러오기

- 데이터 불러오는 함수: `pd.read_csv()`, `pd.read_excel()`, `pd.read_spss()`
- 만날 수 있는 예외: `FileNotFoundException`, `UnicodeDecodeError`, `ModuleNotFoundError`

2. 데이터 정보 확인

- 데이터 속성: `data.columns`, `data.shape`
- 데이터 전체 정보: `data.info()`
- 데이터 형 변환: `data.astype()`, `pd.to_datetime()`
- 범주형 변수 카테고리 조회: `data.unique()`, `data.describe()`
- 연속형 변수 요약: `data.describe()`

3. 결측치 처리

- 결측치 개수 파악: `data.isna()`, `data.isnull()`
- 결측치 처리: `data.fillna()`, `data.dropna(subset=[])`

4. 데이터 시각화

1) 변수 분포 파악

- 범주형/순서형/이산형: 막대그래프
- 연속형: 히스토그램, 상자그림
- 시계열: 선그래프
- 전처리가 필요한가? ex. 이상치 제거, 로그 변환 등

2) 변수 관계 파악

- `data.groupby()`, `pd.pivot_table()`, `pd.crosstab()`
- 상관관계: `data.corr()`, `sns.pairplot()`, `sns.heatmap()`

7. 결론

- 어떤 전처리를 진행했나요?
- 데이터의 특징이 무엇인가요?