

👉 데이터 분석 프로세스

데이터 분석 프로세스 가이드

1. 비즈니스 문제 정의하기

- a. 분석 배경 (Background)
 - b. 분석 목표 (Objective)
 - c. 분석 방법 (Methodology)
 - d. 분석 전제 및 가정 (Assumption)
 - e. 분석 한계점 (Limitation)
-

2. 데이터셋 추출 및 수집하기

- a. 원본 데이터 (Raw Data)
 - b. 변형 데이터 (Intermediate Data)
 - c. 최종 데이터 (Final Data)
-

3. 데이터 분석 설계 및 모델링하기

- a. 데이터 프로파일링 & 카탈로그
 - b. 분석 지표 설정
 - c. 탐색적 데이터 분석 (EDA)
 - d. 데이터 전처리 & 정제
 - e. 데이터 모델링 & 모델 평가
-

4. 인사이트 및 결론 해석하기

- a. 통계 분석적 해석
 - b. 비즈니스적 해석
-

5. 발표 및 보고하기

- a. 발표 프리젠테이션
 - b. 분석 대시보드
 - c. 인포그래픽
 - d. 코드 & 리포지토리
-

6. 액션 아이템 정하기 & 비즈니스 적용하기

- 액션 아이템
- 비즈니스 적용 범위

👉 비즈니스 문제 정의

👉 비즈니스 문제 정의

? 우리가 분석 문제부터 설정해야 하는 이유

+ 분석 배경

+ 분석 목표

+ 분석 방법

+ 분석 전제 및 가정

+ 분석 한계점

? 문제부터 설정해야 하는 이유

? 우리가 분석 전 문제부터 정의해야 하는 이유

데이터 분석 프로젝트에서 해결해야 할 비즈니스 문제를 구체적으로 진단하는 것이 매우 중요합니다.

데이터 분석 프로젝트에서 비즈니스 문제를 정의하는 것이 중요한 4가지 이유는 다음과 같습니다.

✓ 문제 정의는 곧 실현 가능한 목표 설정으로 이어지기 때문입니다.

비즈니스가 마주한 문제를 제대로 이해하면,

데이터 분석을 통해 해결해야 할 실제 문제와 달성해야 할 목표를 더 명확하게 파악할 수 있습니다

비즈니스에 맞는 올바른 질문을 하고 그에 따라 분석의 목적을 구체화할 수 있습니다.

✓ 문제에 맞추어 분석 범위와 접근 방식을 정할 수 있기 때문입니다.

문제 해결에 필요한 비즈니스 상황, 제약 조건, 이해관계자 요구 사항 등을 고려하여 분석에 적합한 범위와 접근 방식을 설정할 수 있습니다.

사용 가능한 데이터, 리소스 및 시간의 한계를 이해하고 그에 따라 분석을 계획하는 것이 좋습니다.

✓ 실행 가능하고 영향력 있는 분석을 만들 수 있기 때문입니다.

비즈니스에게 꼭 필요한 내용만 추려서 의사 결정에 활용할 수 있습니다.

비즈니스 요구사항과 밀접하게 연계된 분석 결과는 실제 영향력이 있는 문제를 해결하는 데 더 효과적입니다.

✓ 이해관계자와 효과적으로 소통할 수 있기 때문입니다.

비즈니스 배경을 이해하면 이해관계자에게 분석 프로세스와 결과를 명확하게 전달할 수 있습니다. 이해관계자의 요구 사항을 더 잘 이해하고 그에 따라 분석을 맞춤화할 수 있습니다.

결론적으로, 비즈니스 배경에 대한 이해는 데이터 분석 프로젝트의 방향과 효과를 개선하여 비즈니스 목표 달성에 크게 기여할 수 있습니다. 이는 데이터 분석가의 역량과 분석 결과의 가치를 높이는 데 있어 핵심적인 요소입니다.

+ 분석 배경

+ 분석 배경

[Who | What | Where | When | Why | How]

비즈니스 문제를 **W5H1** 프레임워크로 전환해보세요.

1. 비즈니스 문제 정의

무엇이 문제인가요? 누가 문제를 가지고 있나요? (Who)

2. 솔루션의 필요성

비즈니스 솔루션이 필요한 이유는 무엇인가요? (Why)

3. 성공 지표

이해관계자가 프로젝트의 성공 여부를 어떻게 판단하나요? (What)

4. 분석 접근 방식

문제 해결을 위한 분석 접근 방식의 일반적인 특성은 무엇인가요?

Supervised 또는 Unsupervised 접근 방식이 사용되나요?

Supervised 접근 방식이라면 문제는 회귀 문제인가요? 분류 문제인가요?

Supervised 접근 방식이라면 목표 변수는 무엇인가요? (Where)

5. 범위:

무엇을 전달하고 무엇을 범위에서 제외할 것인가?

나중에 무엇이 추가될 수 있는가? (When)

6. 세부 사항:

누가 프로젝트를 실행할 예정인가요? 프로젝트는 언제 완료되나요? 중요한 프로젝트 마일스톤이 있나요? (How)

Our W5H1 Framework

7. **What:** Identify and analyze patterns in machine downtime and factors influencing breakdowns.
8. **When:** Examine the frequency and timing of breakdowns, including seasonal or operational peaks.
9. **Where:** Look at the specific locations, plants, and machines affected by downtimes.

10. **Who:** Involve maintenance and production teams to understand the root causes and resolutions.
11. **Why:** Identify underlying causes of machine failures to enable predictive insights.
12. **How:** Examine maintenance processes and whether preventive or corrective actions could be optimized.

+ 분석 목표

+ 분석 목표

[SMART]

- 데이터 분석으로 해결하고자 하는 구체적인 목표 또는 대답하고 싶은 질문을 정의합니다.
 - **Specific** (구체적인)
 - **Measurable** (측정 가능한)
 - **Attainable** (달성 가능한)
 - **Relevant** (관련 있는)
 - **Time-bounded** (시간 제한적인)
- 분석 목표가 측정 가능한 요소들로 명확하게 측정이 가능한지 확인합니다.
- 구체적인 목표가 달성 가능한지, 그리고 왜 달성하면 좋은지에 대해 논의합니다..
 - 목표를 달성했을 때 혜택 및 효과는 무엇인가요?
 - 해당 목표가 다른 목표 보다 더 중요한 이유는 무엇인가요?

Our Objective & Goal

- The main goal is to develop a predictive model that forecasts machine downtime, identifies critical components prone to breakdown, and quantifies the financial and productivity impact.
- This model aims to provide actionable insights for proactive maintenance, reducing costs, and enhancing efficiency.

+ 분석 방법

+ 분석 방법

- Supervised
 - Classification
 - Regression
 - Unsupervised
 - Clustering
 - Abnomal Detection
 - Dimensional Reduction
-

+ 분석 전제 및 가정

+ 분석 전제 및 가정

- 데이터 분석에서 미리 가정할 수 있는 상황이나 전제 조건을 나열합니다.
 - 데이터의 사용 범위 및 여부, 관련 리소스, 분석 마일스톤 타임라인 등과 같은 외부 요인을 충분히 고려합니다.
-

Our Assumptions

- Downtime patterns are statistically consistent across the machine and plants.
- Parts causing downtime can be pre-stocked, allowing for timely repairs.

+ 분석 한계점

+ 분석 한계점

- 분석 중에 발생할 수 있는 잠재적인 한계점이나 문제점을 파악합니다.
 - 이러한 한계를 해결하거나 완화할 방법을 계획합니다.
-

Our Limitations

- **Missing Data:** Some work orders have missing proportion variables, affecting the analysis accuracy.
- **Unrecorded Equipment Tracking:** Only significant breakdowns make it to the IWC, potentially missing patterns from minor issues.
- **Unidentified Machine age:** The range of factors (machine age, usage patterns) may not fit for model development and impact prediction accuracy.

👉 데이터셋 추출 및 생성

데이터셋

-
- 원본 데이터
 - 변형 데이터
 - 최종 데이터
-



Raw Stage

원본 데이터 (**Data Sources**)

- 내부 및 외부에 있는 관련 데이터 소스에 대해 파악합니다.
- 각 데이터의 구조, 형식, 품질에 대해 이해합니다.



Dev Stage

변형 데이터 (**Data ETL**)

- 필요한 데이터를 수집하기 위한 체계적인 접근 방식과 인프라를 구축합니다.
- 일관되고 신뢰할 수 있는 방식으로 데이터를 관리하고 접근할 수 있을지를 판단합니다.

 Prod Stage

최종 데이터 (**Final Data**)

- Actual Dataset to be used for the analysis
- Variable & Metrics defined and labeled as appropriate

👉 분석 과정 및 모델링

분석 과정 및 모델링

데이터 프로파일링

분포, 상관관계, 이상값 등 데이터의 기본 특성을 분석하세요.
데이터에 대해 더 깊이 이해하고 잠재적인 인사이트를 파악하세요.

데이터 시각화

차트, 그래프, 대시보드와 같은 시각화를 만들어 데이터를 탐색하세요.
데이터 내의 패턴, 추세 및 관계 식별

가설 생성 및 검증

탐색적 분석을 통해 얻은 인사이트를 기반으로 가설을 수립합니다.
추가 분석을 통해 검증할 수 있는 테스트 가능한 가설 개발



데이터 프로파일 & 카타로그



데이터 프로파일 & 카탈로그 (Data Profiling & Catalog)

- 분포, 상관관계, 이상값 등 데이터의 기초 통계 관련 특성을 분석합니다.
- 해당 데이터에 대한 잠재적인 인사이트를 파악합니다
- 독립 변수와 종속 변수가 무엇인지 파악하고 서로 관계에 대해 정의합니다.



분석 지표



분석 지표 (**Metrics**)

- Which key number/ measurable metrics are we using for the analysis?



탐색적 데이터 분석 EDA

탐색적 데이터 분석 (EDA)

데이터 프로파일링

- 분포, 상관관계, 이상값 등 데이터의 기본 특성을 분석하세요.
- 데이터에 대해 더 깊이 이해하고 잠재적인 인사이트를 파악하세요.

데이터 시각화

- 차트, 그래프, 대시보드와 같은 시각화를 만들어 데이터를 탐색하세요.
- 데이터 내의 패턴, 추세 및 관계를 식별하고 시각적으로 구성합니다.

가설 생성

- 탐색적 분석을 통해 얻은 인사이트를 기반으로 가설을 수립합니다.
- 추가 분석을 통해 검증할 수 있는 테스트 가능한 가설을 개발하는 것에 집중합니다.



데이터 전처리 & 클린징

데이터 전처리 (**Data Preprocessing**)

데이터 클리닝 (**Data Cleansing**)



모델링 & 모델 평가

모델링 (Modeling)

피처 엔지니어링

- 모델의 성능을 향상시킬 수 있는 새로운 피처를 식별하고 생성합니다.
- 기존 데이터를 변환하거나 결합하여 보다 의미 있는 변수를 만들어냅니다.

모델 비교 및 파인 튜닝

- 분석 목표에 따라 적절한 통계 또는 머신러닝 모델을 선택합니다.
- 다양한 모델링 기법의 장점과 한계점을 비교하고 선택합니다.

모델 평가 (Model Evaluation)

Model Evaluation

- 선택한 모델의 성능 및 정확도를 평가합니다.
- 적절한 성능 지표와 검증 기법을 사용하여 모델을 검증하는 과정을 거칩니다.

👉 인사이트 및 결론 해석

👉 인사이트 및 결론 (Insight & Key Finding)

인사이트 및 결론

- 데이터 분석 결과를 통해 비즈니스와 이해관계들에게 의미 있는 인사이트를 도출합니다.
- 앞에서 설정한 비즈니스 목표를 달성하기 위한 결과물을 정리하고 요약하고 종합합니다.



분석적 해석



분석적 해석 (**Analytical Interpretation**)



비즈니스적 해석



비즈니스적 해석 (**Business Interpretation**)

👉 발표 및 보고서

👉 발표 및 보고서 (Presenation & Reporting)

- 명확하고 간결한 보고서 양식 또는 프레젠테이션을 준비하여 데이터 분석 결과를 전달하세요.
- 이해관계자의 구체적인 요구와 선호도에 맞게 커뮤니케이션을 맞춤화합니다.



발표 자료

발표 자료



대시보드



인포그래픽

인포그래픽



코드



코드

👉 액션 아이템 & 비즈니스 적용 범위

👉 액션 아이템 & 비즈니스 적용 범위 (Action Item)



액션 아이템

액션 아이템



비즈니스 적용 범위

비즈니스 적용 범위
