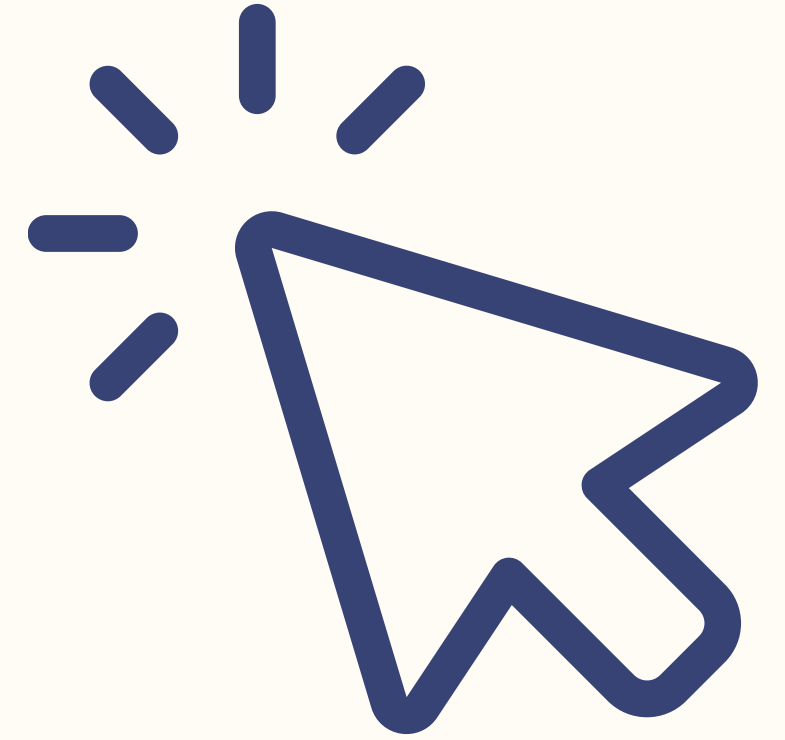


Data Analysis

Portfolio

데이터 분석가 최혜은의 포트폴리오



데이터에서 인사이트를 도출하고, 변화의 방향을 설계합니다.

Contents



- 01 Introduction**
자기소개
- 02 Education**
학력 및 전공소개
- 03 Skill**
역량 및 기술
- 04 Projects**
팀 프로젝트
- 05 Contact**
연락처

Introduction

자기소개



안녕하세요!

기초를 다지고 실전으로 성장하는 분석가 최혜은입니다.

데이터는 과거를 보여주고, 전략은 미래를 만듭니다.

저는 데이터를 분석해 문제를 정의하고, 수치 너머의 맥락을 읽어 전략을 설계하는 데 강점을 지닌 데이터 기반 전략가입니다.

다양한 산업과 주제를 다룬 프로젝트와 공모전에 참여하며, 실제 시장 문제에 분석적 해답을 제시해왔습니다.

통계 기반 분석 → 인사이트 도출 → 실행 전략 수립으로 이어지는 전 과정을 직접 기획하고 리드한 경험이 있습니다.

이러한 경험을 통해 저는 숫자에 설득력을, 전략에 실현 가능성을 담는 훈련을 해왔습니다.

앞으로도 데이터와 전략을 연결해 시장과 조직의 변화를 이끄는 사람이 되고 싶습니다.

이러한 역량을 바탕으로 데이터 분석가로서 회사와 함께 성장하고 싶습니다.

Education

학력 및 자격증

학력

2016년 3월 - 2021년 8월

대진대학교 경영학과 졸업

경제학과 복수전공

2022년 3월 - 2024년 2월

동덕여자대학교 정보통계학과 졸업

학사편입

자격증

- 2023년 6월 **ADsP (데이터분석 준전문가)**
- 2025년 4월 **SQLD (SQL 개발자)**

어학

- 2023년 6월 **토익 수정 예정**
- 2025년 4월 **오픽 수정 예정**



문제 정의 및 분석 설계

- 시장, 소비자, 정책 맥락을 반영하여 문제의 근본 원인을 정의하고, 목적에 맞는 분석 프레임과 데이터 구조 설계
- 단순 통계 분석이 아닌, 의사결정에 필요한 실질적 분석 수행 경험



데이터 분석 및 시각화 역량

- Python 기반의 Pandas, Scikit-learn을 활용한 데이터 전처리 및 예측 모델링 수행
- 단순 수치 제시가 아닌, 행동 가능한 인사이트 도출 중심의 분석
- Matplotlib, Seaborn, Plotly를 활용한 목적 중심의 맞춤형 시각화 구현



전략 기획 및 인사이트 도출

- 분석 결과를 바탕으로 정책 제안, 마케팅 전략 등 현실 적용 가능한 인사이트 제시
- 정부, 기업, 브랜드 등 이해관계자 중심의 스토리텔링 기반 보고서 작성
- 분석 결과를 ‘숫자 나열’이 아닌, 의사결정에 영향을 줄 수 있는 형태로 구조화

Team Projects

팀 프로젝트

프로젝트 1

경제주체별 환율 분석 및 예측 프로젝트 - 미래 환율을 읽다

경제주체(개인·기업·정부·해외)별 변수로 이벤트 스터디 기반한 환율 변동 원인을 분석하고 예측 모델을 통해 위기 조기 경보 가능성 및 수입단가 예측을 검토한 프로젝트

[프로젝트 목적]

- 환율 변동 요인 구조 파악 및 예측 결과 기반 정책/실무 활용 가능성 제시

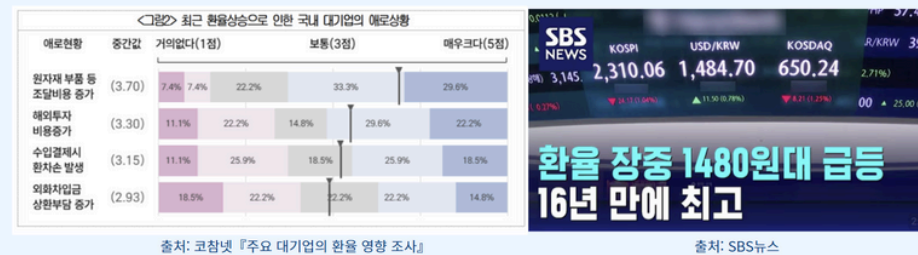
[데이터 정보]

- 한국은행 API (2000~2024, 25개년)
- 변수: 실업률, 원자재 가격, 금리 등

[사용 도구]

- Python (Pandas, Scikit-learn, TensorFlow), Matplotlib, Seaborn, Plotly

01. 배경 및 문제 정의



한국은 수출입 의존도가 높은 만큼, 환율 급등락이 정부의 시장 안정화 대응과 기업의 원가 관리에 직접적인 부담으로 작용한다. 하지만 환율은 복합 변수로, 단순한 분석만으로는 정확한 예측이 어려워 정부와 기업 모두 선제적 대응에 한계를 겪고 있다.

정부 - 기준 금리

가설과 실제 데이터의 차이: 실제 예시 기반 조사

“2017년 3월 미국 기준금리 0.25%p 인상 후 원/달러 환율 하락”

금리 인상 기대의 선반영

- 금리 인상: 이미 시장에 반영
- 투자자: 향후 금리 동결 또는 인하 가능성 주목
- ∴ 금리 인상 → 달러 수요 감소

한국의 견조한 경제 지표

한국: 수출 호조 & 경상수지 흑자 지속
→ 원화에 대한 신뢰 상승
→ 원/달러 환율 하락

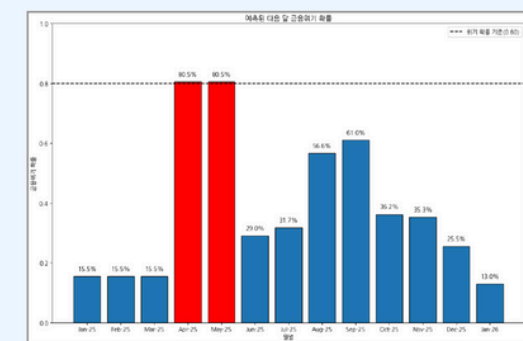
글로벌 리스크 완화

글로벌 금융시장에서의 불확실성 감소
→ 안전자산 선호 현상 약화
→ 달러 수요 감소

05. 머신러닝 성능 비교 (전체 변수)

주요 기능	변수 선택	RMSE	MAE	R2
GRU	L1 정규화 + 단계선택	56.8256	44.4941	0.6192
LSTM	Permutation (top15)	60.4430	48.5842	0.5692
TCN	Permutation (top 3)	50.1332	40.7778	0.7037
Random Forest	Permutation (top10)	159.932331	134.749965	-1.965843
CNN	L1 정규화	85.68	73.50	0.157

06. 활용 방안 제시 [정부]



금융 위기 조기 감지 시스템

거짓정보 최소화를 위하여 IMF 『Predicting Financial Crises』 기준, 리스크 발생 80%인 시점을 경보 구간으로 정의해 예측한 환율을 바탕으로 금융위기 발생 확률을 계산

4월, 5월 금융 위기 조기 감지
⇒ 다음달 5월, 6월 금융 위기 발생할 것이라 예측

미래 환율 예측 프로젝트 - 분석 흐름

1. 문제 정의

- 환율은 한국 경제의 핵심 변수로, 수출입 기업의 원가·수익, 정부의 통화·외환 정책에 직접적 영향
- 복합적인 거시 요인의 영향을 받기 때문에 단일 변수 중심의 예측은 한계가 존재
- 정부 및 기업이 실질적인 대응 전략 수립을 위해 구조적 분석 프레임이 필요함

2. 주체별 변수 정리

- 주요 거시 변수(실업률, 주가, 금리, 외환 등)를 개인·기업·정부·해외 4개 경제주체로 분류
- 환율과 각 주체 간의 구조적 상관관계를 정의하여 주체별 상호작용 기반 분석 프레임 수립
- 변수 중복·불일치 해소를 위한 기준 정비 및 시계열 통합

3. 이벤트 스터디

- 닷컴버블, 글로벌 금융위기, 코로나 팬데믹, 자이언트 스텝 등 4대 주요 이벤트 중심의 환율 변동 분석
- 각 시기마다 지배적인 주체 변수가 상이하며, 환율 반응이 복합적이고 유동적임을 확인
- 정량적 시계열 + 정성적 경제맥락 해석을 병행하여 시기별 반응 구조 파악

4. 모델링

- LSTM, TCN, GRU 등 시계열 딥러닝 모델을 적용하여 1~6개월 예측 성능 비교
- 모델별 성능 지표(RMSE, MAE, R^2)를 기반으로, TCN이 중기 예측에 가장 우수한 성능을 보임
- 변수 중요도 분석을 통해 각 주체별 핵심 영향 요인을 추출하고 해석

5. 정책 및 비즈니스 제안

- 정부: 조기경보 시스템 기반으로 금리·환율 정책의 선제적 대응 체계 마련
- 기업: 수입단가 예측 시뮬레이터를 통해 환헤지 전략 수립 및 조기 발주 등 리스크 관리 강화
- 분석 → 예측 → 활용 흐름을 통합한 정책·비즈니스 연결형 프레임워크로 확장 가능

미래 환율 예측 프로젝트 - 트러블 슈팅 & 기여도

문제 발생 및 해결 사례

시계열 예측 모델의 과적합 발생

- 문제: LSTM, GRU 등 딥러닝 기반 시계열 모델에서 학습 구간은 예측력이 높으나, 테스트 구간에서 성능이 급격히 저하
- 해결: Dropout 조정, EarlyStopping 도입, 예측 구간을 단기/중기로 나누어 성능 비교 수행

거시변수 간 시간 주기 차이 발생

- 문제: 일부 변수는 월간, 일부는 분기별·연간 데이터로 존재하여 시계열 예측의 연속성 저해
- 해결: 선형보간(interpolate)을 통해 월 단위로 보간 정제, resample() 활용하여 단위 맞춤

이벤트 스터디 기간 설정 기준의 불명확성

- 문제: 각 경제 이벤트(닷컴버블, 금융위기, 코로나 등)의 시작·종료 시점을 어떤 기준으로 설정할지 의견 차이 발생
- 해결: IMF, 한국은행 등 주요 기관의 공식 자료에서 정의한 시점을 기준으로 설정하고, 환율의 급변 시점을 보조 지표로 활용하여 해석 병행

가설과 상반된 분석 결과

- 문제: “기준금리가 상승하면 환율도 상승할 것”이라는 가설과 달리 2017년 3월 미국 기준금리 인상 직후에는 원/달러 환율이 하락하는 현상이 관측됨
- 해결: 가설이 모든 시점에 적용되지 않음을 인지하고, 실제 사례 기반으로 “금리 인상 → 환율 상승”이라는 단선적 해석이 항상 유효하지 않음을 정리

기여도

기여도 ●●●●○

- 프로젝트 노션 관리 및 팀 내 정보 공유 체계 구축, 용어 정리 및 분석 가이드 사전 제작
- 정부 주체 변수 분석 담당: 기준금리·부채비율 등의 환율 영향 분석 및 시각화
- TCN 기반 환율 예측 모델링 수행 및 타 모델과의 성능 비교 기여
- 정부 활용 제안 설계: 조기경보 시스템 구조 구상 및 적용 시나리오 작성

참고자료 링크

GitHub

<https://github.com/hyeeun1031/forecasting-exchange-rates>

발표 자료

[https://github.com/hyeeun1031/forecasting-exchange-rates/blob/main/4팀\(데이터톤\)-미래환율을읽다.pdf](https://github.com/hyeeun1031/forecasting-exchange-rates/blob/main/4팀(데이터톤)-미래환율을읽다.pdf)

Team Projects

팀 프로젝트

프로젝트 2

4개국 테러 분석을 통한 공통 대응 전략 - 데이터로 막는 다음 테러

4개국의 테러 데이터를 분석하여 성공 여부를 예측하고 국가별 맞춤 전략과 공통된 대응 방안을 제시한 데이터 기반 테러 분석 프로젝트

[프로젝트 목적]

- 국가별 특성과 데이터 기반 예측 모델을 통해 맞춤형 대응 전략을 제안하고 공통 대응 방안을 도출해 실질적 테러 체계 수립에 기여

[데이터 정보]

- Global Terrorism Database(GTD)를 활용해 1990~2020년까지의 테러 사건 정보를 국가·지역·공격유형·무기유형·타겟 등 다양한 변수로 구성

[사용 도구]

- Python(Pandas, Scikit-learn, XGBoost 등), EDA 및 시각화(Matplotlib, Seaborn), 머신러닝 기반 분류 모델, SMOTE, GridSearchCV

01 문제 정의와 목표

문제 정의

국가마다 발생하는 테러의 양상은 상이
그럼에도 지역/국가별 맥락을 고려하지 않은
일률적인 방식으로 이루어짐

분석 목표

전 세계 테러 데이터를 기반으로
특성이 뚜렷한 4개국을 선정
국가별 테러 성공 예측 모델을 통해 공통점을 종합하여
법지구적 대응 전략을 제시



04 가설(2)

테러 피해자 수 (사망자+부상자) 는
사용된 무기 유형과 밀접한 관련이 있을 것이다.

TRUE

ANOVA

```
from scipy.stats import f_oneway

# 무기 유형별로 그룹 나누기
groups = [group['total_victim'].dropna().values
           for _, group in df_1990.groupby('weaptype1_txt') if len(group) > 10]

# ANOVA 검정 수행
f_stat, p_val = f_oneway(*groups)

print("F-통계량:", round(f_stat, 3))
print("p-value:", round(p_val, 3))

if p_val < 0.05:
    print("→ 무기 유형에 따라 피해 규모의 평균이 통계적으로 유의미하게 다름.")
else:
    print("→ 무기 유형 간 피해 규모 차이는 통계적으로 유의하지 않음.")
```

F-통계량: 117.786
p-value: 0.000
→ 무기 유형에 따라 **피해 규모의 평균**이
통계적으로 유의미하게 다름.

크루스칼-월리스 H-test

```
from scipy.stats import kruskal

# 동일하게 그룹화
groups = [group['total_victim'].dropna().values
           for _, group in df_1990.groupby('weaptype1_txt') if len(group) > 10]

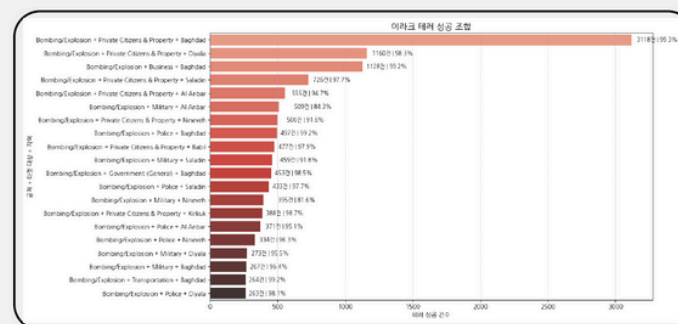
# Kruskal-Wallis H-test
h_stat, p_val = kruskal(*groups)

print("H-통계량:", round(h_stat, 3))
print("p-value:", round(p_val, 3))

if p_val < 0.05:
    print("→ 무기 유형에 따라 피해 규모 분포에 유의미한 차이가 있음.")
else:
    print("→ 피해 규모는 무기 유형 간 유의미한 차이가 없음.")
```

H-통계량: 9465.586
p-value: 0.000
→ 무기 유형에 따라
피해 규모 분포에 유의미한 차이가 있음

05 4개국 분석 : 이라크 - 해석



공격 유형

폭탄/폭발 테러

타겟 대상

민간인 및 민간 단체

지역

바그다드(수도), 다알라

07 결론



소프트 타겟 보호 및 예방 인프라 강화

공공기관/소형공간 보안 대응
매뉴얼 국제 표준화

소프트 타겟 대상
정기 훈련/점검 가이드라인 제정



조기 탐지 및 위험 식별 시스템 구축

위험을 거래 모니터링
(데이터 연계 시스템 국제 협약)

AI 기반 비정상 행동 탐지 시스템



시간간 기반 탄력 보안 전략

지역별 대응 매뉴얼 시스템 구축

핫스팟 시간표를 통한
탄력 보안 인력 운영

테러 예측 프로젝트 - 분석 흐름

1. 문제 정의 및 국가 선정

- 전 세계 테러 사건은 양상과 대응 방식이 국가마다 다르지만, 대응은 일률적이며 지역 맥락이 반영되지 않음
- 이라크, 미국, 터키, 영국 등 테러 발생 특성이 뚜렷한 4개국을 선정하여 비교 분석 및 전략 도출

2. 데이터 전처리 및 변수 구조화

- 테러 발생일, 성공 여부, 무기·공격·타겟 유형 등 핵심 범주형 변수 중심으로 재구조화
- 결측치는 공격 유형/성공 여부 기준으로 그룹 중앙값 보간, 위치 정보는 Nominatim API로 정제

3. EDA 및 통계 검정

- 민간인 대상 공격이 증가하고 있음을 시계열 분석과 카이제곱 검정을 통해 입증
- 무기 유형과 피해 규모 간 상관관계를 ANOVA 및 H-test로 검정
- 난민 수와 테러 수 간의 상관관계, 지역별 무기 유형 차이 등도 통계적으로 유의함을 확인

4. 국가별 머신러닝 모델링 및 성능 비교

- 각 국가별로 테러 성공 여부를 예측하는 분류 모델 구축 (Random Forest, SVM, Voting 등)
- 불균형 데이터 문제 해결을 위해 SMOTE, class_weight 조정, 하이퍼파라미터 튜닝 적용
- 미국은 Stacking, 이라크는 XGBoost, 영국은 SVM이 가장 높은 F1 Score 기록

5. 정책 및 비즈니스 제안

- 국가별 맞춤 대응 전략 도출:
- 예) 미국은 주말/민간 시설 중심 대응 강화, 이라크는 바그다드 중심 위험지역 모니터링 체계
- 공통 대응 전략 제안:
- 소프트 타겟 보호, 위험물 관리, 핫스팟 시간 기반 탄력 경계, AI 기반 비정상 행동 탐지 시스템 구축

테러 예측 프로젝트 - 트러블 슈팅 & 기여도

문제 발생 및 해결 사례

불균형 데이터로 인한 모델 편향

- 문제: 테러 성공 여부를 분류할 때, 일부 국가 (예: 이라크)의 데이터는 성공 사례가 압도적으로 많아 분류 모델이 "성공"에 편향되는 현상 발생
- 해결: SMOTE 오버샘플링을 통해 소수 클래스인 '실패' 데이터를 생성하여 학습 균형을 맞췄고 class_weight 조정 및 여러 모델 성능 비교 (F1 Score 중심)를 통해 최적 모델을 선정함

가설과 통계 결과 간의 괴리

- 문제: '무기 유형이 치명도(사상자 수)에 영향을 줄 것이다'는 가설과 달리, 일부 지역에서는 통계적으로 유의하지 않은 결과 도출
- 해결: 범주 간 분산이 너무 커서 검정력이 약해졌다는 점을 확인하고, 사상자 수를 로그 변환한 후 재분석하여 유의성을 확보함
동시에 지역·국가별로 추가 분할 분석을 병행함

기여도

기여도 ●●●●○

- 이라크 국가 분석 담당으로 주요 변수(지역, 공격 유형, 타겟 등) 기반 이라크 맞춤 정책 및 제안 도출에 기여
- “테러 피해자수는 무기 유형과 관련이 있다“ 등의 가설 하에 추론 통계 분석 수행
- 팀 노션 관리 및 프로젝트 자료 정리를 맡아 협업 체계를 구축하고 전체 흐름을 체계화

참고자료 링크

GitHub

<https://github.com/hyeeun1031/predict-next-terror>

발표 자료

[https://github.com/hyeeun1031/predict-next-terror/blob/main/6조\(파이널\)-데이터로 막는 다음 테러.pdf](https://github.com/hyeeun1031/predict-next-terror/blob/main/6조(파이널)-데이터로 막는 다음 테러.pdf)

Team Projects

팀 프로젝트

프로젝트 3

한국인 행복 지수 개선 시각화 프로젝트 - 한국인의 행복, 데이터로 답을 찾다

한국의 행복지수가 경제 수준에 비해 낮은 원인을 진단하고, 다양한 국가와 비교 분석을
통해 개선 방향과 정책 제안을 도출한 프로젝트

[프로젝트 목적]

- 한국인의 낮은 행복지수를 설명하는 핵심 요인을 추출하고, 비교 국가 사례를 통해
실질적인 정책 제안으로 연결

[데이터 정보]

- World Happiness Report (2018~2022, 157개국), 의료비, 비만율, 알코올 소비량,
언론/정부 신뢰도, 사회 지원 지표 등 외부 데이터 추가 수집

[사용 도구]

- Python (pandas, seaborn, matplotlib), 통계 분석 (상관계수, 카이제곱검정 등)

목표

- 한국 행복지수를 높이기 위해 행복지수 상위 / 하위 20개국 다각적 비교
- 한국 행복 지수 상승 방안을 위해 정책 / 방향 제안



G12 선진국 행복 지수 59위

여러 분석에 따르면, 한국은 G12에 포함될 정도로 국제적 영향
력이 커지며 선진국 반열에 올라서고 있다.
그러나 행복지수는 이러한 경제적-국제적 위상에 비해 상대적으로
낮은 수준을 보이고 있다. 이에 대한 원인을 분석해보았다.

근무시간

- 근무시간이 짧을수록 행복도가 높은 경향
- 한국의 1인당 연간 평균 1936시간 근무
- 이스라엘 & 코스타리카 ↔ 한국

행복 지수 ~ 근무 시간



해결방안

social support

사회적 네트워크와 지원

취약한 저소득층 지원 및 기초생활보장 시스템
기초생활보장 제도를 운영하고 있지만, 저소득
층에 대한 실질적 지원이 부족. 일부 계층은 복지
혜택을 제대로 받지 못하고 있음.

사회적 연대 부족

개인주의적인 성향이 강해 사회적 연대나 공동
체 의식이 상대적으로 부족. 이는 특히 저소득층,
고립된 노인, 청소년에게 더 큰 영향을 미침.

개선책

경제적 지원 및 취약계층 보호
기초생활수급자와 저소득층의 복지 혜택을 확충
하고, 사회적 취약계층에게 실질적이고 지속적
인 지원을 제공하는 제도 강화.

사회적 지원 네트워크

자원봉사 활성화와 사회적 기업 지원을 통해 사
회적 네트워크를 활성화하고, 사회적 지원 시스
템을 지역사회에서 강화.

결론

01. 건강 불평등 해소
의료 접근성 개선 & 예방적 건강 정책 확대
02. 사회적 연대 회복
저소득층의 복지 확대 & 지역 사회 네트워크 강화
03. 개인의 삶의 질 향상
유연한 근무제 도입 & 정신건강 지원 확대
04. 정부 / 언론 신뢰도 향상
부패방지 정책 강화 & 균형 잡힌 미디어 환경 조성
05. 삶의 질 중심 정책 전환
국민 행복을 최우선 가치로 설정

한국인 행복 프로젝트 - 분석 흐름

1. 문제 정의 및 국가 비교 구조 설계

- 한국은 경제력 대비 행복지수가 낮은 현상이 지속
- G12 국가 중 한국은 행복 순위 59위 → 원인 분석 필요
- 상위/하위 국가와 비교하여 구조적 패턴 도출

2. 데이터 정제 및 주요 변수 수집

- 행복지수에 영향을 주는 6대 요인(GDP, 사회지원, 기대수명 등) 정리
- 외부 요인(비만율, 의료비, 언론/정부 신뢰도 등) 병합
- 연도별·국가별 정합성 있는 데이터셋 구축

3. 비교 분석 및 통계 검정

- 변수별 상관관계 분석: GDP(0.80), 기대수명(0.71), 사회지원(0.78)
- 한국은 기대수명·GDP는 높으나 사회지원·정부 신뢰 낮음
- OECD 평균 대비 열위인 요인들을 중심으로 개선 방향 제시

4. 가설 설정 및 사례 분석

- "GDP 높으면 행복하다"는 가설을 미국, 체코, 뉴질랜드 사례로 반례 분석
- 비만율, 알코올 소비, 의료 불평등 등이 기대수명과 행복지수에 영향
- 사회적 네트워크 및 근무시간이 행복도에 주는 영향도 병행 분석

5. 정책 및 비즈니스 제안

- 의료 불평등 해소 및 정신 건강 관리 강화
- 유연근무제 확대, 복지 접근성 개선, 사회적 연대 회복
- 정부/언론 신뢰도 회복 위한 구조적 제도 개편 제안

한국인 행복 프로젝트 - 트러블 슈팅 & 기여도

문제 발생 및 해결 사례

팀원 역량 차이로 인한 진행 속도 불균형

- 문제: 프로젝트 초반, 한 팀원이 데이터 분석 툴 사용 경험이 부족해 분석 프로세스를 따라오지 못하는 상황 발생.
전체 진행 속도가 늦어지고, 역할 분담에 공백이 생길 우려가 있었음.
- 해결: 팀장으로서 매일 진행 상황을 공유하고, 해당 팀원에게 필요한 기술(데이터 전처리, 시각화 코드)을 직접 1:1로 설명하며 피드백 제공.
분석의 핵심 파트는 다른 팀원들이 맡고, 해당 팀원은 데이터 검증·보고서 정리 등 적합한 역할을 수행하도록 조정.
결과적으로 팀 내 부담이 분산되고, 프로젝트 마감 기한을 지킬 수 있었음.

기여도

기여도 ●●●●●

- 팀장으로서 프로젝트 일정 & 역할 조율, 회의 주도
- 노션 기반 프로젝트 기록 체계 구축 및 체계적 관리 진행
- '근무시간'과 '사회적 지원 네트워크' 파트 주요 분석 및 인사이트 도출
- 한국 행복지수 저하 원인을 데이터 기반 해석 및 정책 제언 제시

참고자료 링크

GitHub

<https://github.com/hyeeun1031/korea-happiness-visualization>

발표 자료

[https://github.com/hyeeun1031/korea-happiness-visualization/blob/main/5조 - 한국인의 행복 \(발표자료\).pdf](https://github.com/hyeeun1031/korea-happiness-visualization/blob/main/5조-%20한국인의%20행복%20(발표자료).pdf)

Thank you!

감사합니다!



010-5330-2556

chyeun1031@gmail.com

github.com/hyeun1031
