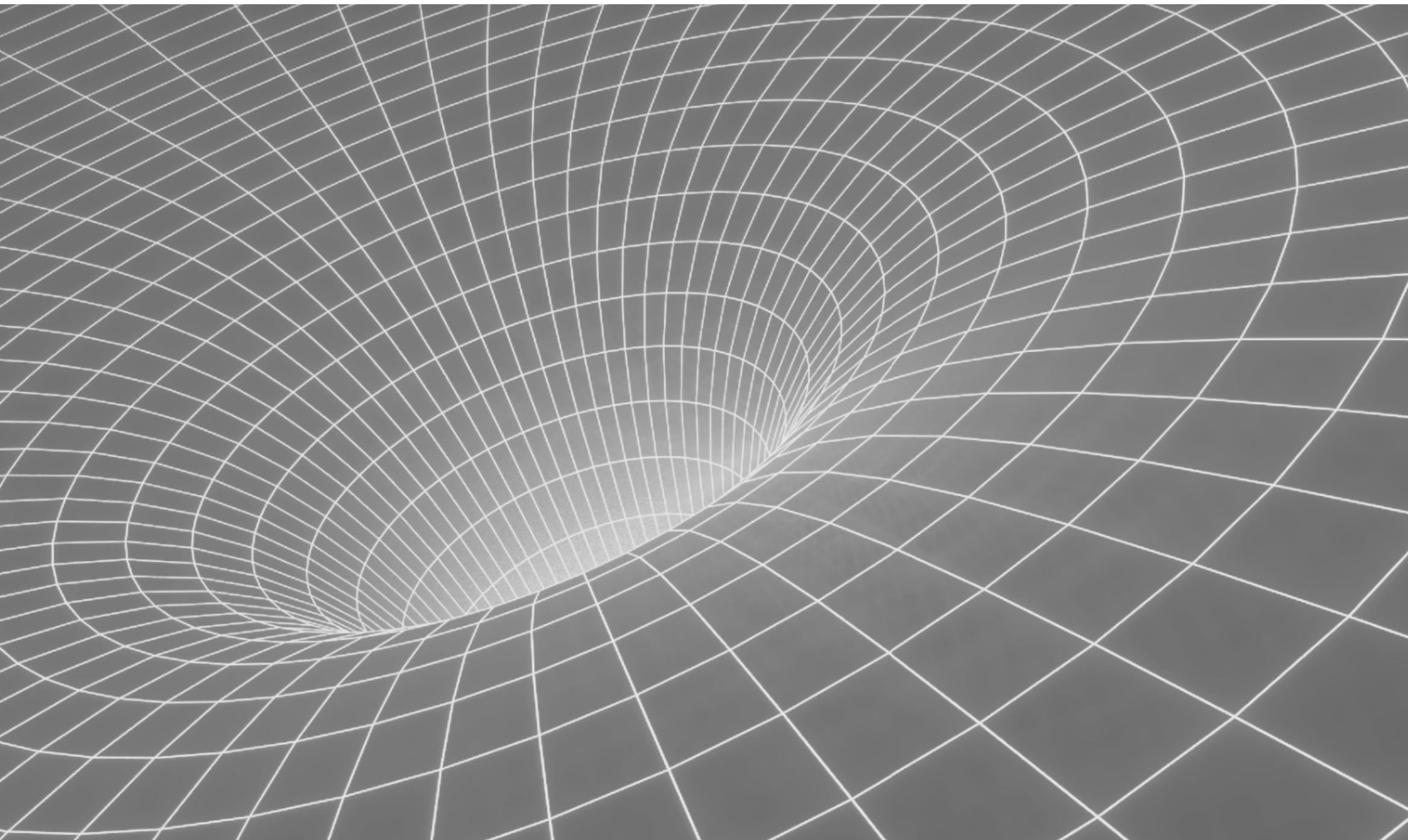


# PORTFOLIO

---



**정찬웅, CHANWOONG  
JEONG**

Developer



jcw7468@naver.com



010 4102 7468

---

<https://chanwoong.dev>

---

<https://github.com/Chanwoong1>

# 정찬웅

CHANWOONG JEONG



## Graduation

2014.02 성사고등학교 졸업

2014.03 동국대학교 입학

2024.02 동국대학교 졸업 예정

## 관심분야

Data Science

Front-end

Data Visualization

## Languages and Tools

Python

Pandas

TypeScript

## 교육

비즈니스 인사이트 도출을 위한 데이터 분석 과정

2021.05 - 2021.11

42SEOUL

2022.04 - 2023.09

## 수행프로젝트

- 2021.08 - 2021.09 | 빅콘테스트 참여  
댐 수위 유입량 예측 모델 개발
- 2021.10 | 스타트업 챌린지 참여  
기업 데이터 활용 인사이트 도출
- 2021.10 - 2021.12 | 빅데이터 인공지능 경진대회  
가스 공급량 예측 모델 개발

## 개인프로젝트

- 2021.09 - 2021.10 | Object Detection  
CCTV 영상을 이용한 객체탐지 모형 향상 대회 참여
- 2021.11 - 2021.12 | Object Detection  
캡슐 내시경 영상을 통한 병변 검출 AI 대회 참여

## About

끊임없이 성장하는 개발자가 되는것이 목표입니다.

제가 구성하고 구현해낸 분석 모형들은 결국 제 노력의 산물일 것이기 때문에,

모형의 성능 또한 제가 공부하고 아는것 만큼 향상될 것입니다.

따라서 배움의 중요성을 느끼며 함께 일하는 사람들의 의견을 존중하며 함께 나아가는것을 소망합니다.

# 댐 유입 수량 예측 모형 도출



물리모델을 통한 댐 유입량 예측의 한계를 보완하기 위하여  
머신러닝을 이용한 모델을 개발하는 프로젝트입니다.

1

작업성격 | Machine Learning

작업역할 | 머신러닝을 이용한 예측 모형 모델링

작업시기 | 2021.08 - 2021.09

# MACHINE LEARNING



# 1 댐 유입 수량 예측 모형 도출

## Problem

### 기후 변화로 인한 예측 정확도 하락

기후 변화로 인해 하락한 예측력은 매년 발생하는 홍수로 인한 피해를 감소시키지 못하고 있습니다.

### 물리모형의 한계

기존 댐 유입량을 파악하는 데 사용했던 물리모형인 COSFIM 모형은 여러가지 한계가 존재합니다.

- 몇 가지 변수만으로 자연 상태의 홍수량을 정확히 예측하는 것의 한계
- 관측 대상의 물리현상이 규명되지 않거나 수식화되지 않을 경우 예측의 한계

## Solution

### 머신러닝을 활용한 모델링

AI를 이용한 댐 유입량 예측 모형은 지속적으로 고안되어 왔으며 다음과 같은 특성들이 물리모형의 한계를 보완합니다.

- 관측된 결과자료와 입력자료 연계 과정 우수
- 복잡한 모의절차 재현 가능
- 물리기반 모형에서의 오차 최소화

### 외부데이터 사용

유입량을 설명하기 위해서는 외부데이터 사용이 불가피합니다.

기후 변화로 인한 유입량을 설명하기 위해 기상데이터를 추가하는 방법 등을 생각해보았습니다.

## Result

### 머신러닝 모형인 LGBM 사용

- LGBM은 부스팅 계열 알고리즘 방식 모델 중 하나입니다. 같은 GBM 계열의 XGBoost의 성능과 비슷함과 동시에 학습시간을 상당히 단축하였습니다.
- LGBM 모델을 사용한 이유는 외부데이터를 사용하면서 카테고리형 변수들이 추가되었기 때문입니다. XGBoost 모형은 카테고리형 변수를 사용하기 위해서 인코딩을 해주어야 하는데 이 방식은 카테고리형 변수가 많을 경우 속도 저하가 발생할 수 있습니다.
- 베이지안 최적화 기법을 이용한 하이퍼파라미터 튜닝으로 모형 성능을 더욱 향상시켰습니다.

## 느낀 점 및 향후 발전 계획

### 외부 데이터 사용을 통한 다양한 데이터 경험

- 외부 데이터 사용을 권장했던 대회로, 기존 제공되는 변수를 제외하고 공급량을 잘 설명할 수 있는 외부 데이터의 변수들을 탐색하고 적용하는 과정을 통해 데이터를 보는 시야가 넓어질 수 있는 기회가 되었습니다.
- 외부데이터 추가에 대해 각자 생각하는 방향이 다름으로 인해서 생겼던 갈등 등을 해결하며 협업으로 인해 생길 수 있는 상황들을 경험해볼 수 있었습니다.

### 머신러닝 분석 경험

- 데이터 분석 과정을 수강하며 경험했던 머신러닝 기법들은 주어진 데이터를 통한 기본적인 사용법 숙지에 불과한 경우가 종종 있었습니다. 직접 데이터를 머신러닝 기법에 적용할 수 있도록 처리하고 분석을 통해 예측하는 모형을 도출하는 과정을 경험하면서 제 역량을 기르기 위해서는 더 많은 경험과 노력이 필요하다는 것을 느꼈습니다.

### Tools



### Auto ML 경험

- 자동으로 머신러닝을 수행할 수 있도록 도와주는 Auto ML 라이브러리 중 Pycaret을 사용해 모형에 적용시켜 보았습니다. 이 과정을 통해 제가 직접 모델링한 모형과 성능을 비교하는 작업을 거쳤습니다. Auto ML 사용은 다양한 데이터에 적용할 수 있어 빠른 분석을 요구할 때 사용할 수 있을 것이라 생각합니다.

### 향후 머신러닝 대회 참가

- 이 대회를 통해 머신러닝에 대한 이해를 높일 수 있었습니다. 향후 향상된 이해도를 기반으로 더욱 합리적인 모델링 과정을 수행할 수 있도록 다양한 데이터를 다루는 대회에 참여할 계획입니다.

# 가스 공급량 예측 모형 개발



천연 자원의 가치는 점점 상승하고 있고, 향후 에너지 대란을 대비하기 위해서는 공급량을 미리 예측할 수 있는 모형이 필요합니다.

머신러닝을 이용한 모델을 개발하는 프로젝트입니다.



작업성격 | Machine Learning

작업역할 | 머신러닝을 이용한 예측 모형 모델링

작업시기 | 2021.10 - 2021.12

2

## 2 가스 공급량 예측 모형 개발

### Problem

#### 에너지 대란으로 인한 나비효과 발생

질소 비료를 생산하는 데 주로 사용되는 천연가스와 석탄의 가격 상승은 비료가격 상승의 주된 원인이며, 농산물 가격 상승으로 인한 물가 상승 현상인 애그플레이션을 일으킬 우려가 있습니다.

또한, 많은 산업들이 천연 자원에 의존하고 있기 때문에, 전체적인 산업들에도 영향을 미칠 수 있습니다.

#### 탄소 중립을 향해, 하지만..

탄소 중립을 향하여 정부 및 여러 기업에서 노력하고 있지만 쉽지 않은 것이 현실입니다. 따라서 탄소중립을 향한 발판을 천연 자원 중 가장 환경 친화적이며, 수소 에너지원 추출이 가능한 천연가스로 선정하여 모델 구축 프로젝트를 진행하였습니다.

### Solution

#### 머신러닝을 활용한 모델링

AI를 이용한 가스공급량 예측 모형은 천연가스 역할이 중요해지는 이 시대에 꼭 필요한 모형입니다.

- LGBM은 부스팅 계열 알고리즘 방식 모델 중 하나입니다. 같은 GBM 계열의 XGBoost의 성능과 비슷함과 동시에 학습시간을 상당히 단축하였습니다.
- 베이지안 최적화 기법을 이용한 하이퍼파라미터 튜닝으로 모형 성능을 더욱 향상시켰습니다.

#### 외부데이터 사용

공급량을 설명하기 위해서는 외부데이터 사용이 불가피합니다.

우리나라 아파트의 난방 시스템이 대부분 가스 난방이기 때문에 기온을 고려하였고, 천연가스의 대부분을 수입하기 때문에 수입 물가를 고려하여 여러 지표들을 데이터에 추가하였습니다.

### Result

#### 천연가스 공급량 예측 모델 개발

- 정확한 공급량 예측은 에너지의 안정적인 수급이 가능하게 할 것입니다.
- 단기적 구매는 수입에 의존하는 우리나라 특성 상 무역 손해를 불러올 것입니다. 정확한 예측은 중장기적 구매 계획을 세울 수 있게 도와주어 예산을 효율적으로 사용할 수 있게 도울 것입니다.
- 다른 에너지 자원도 이와 비슷한 모델을 개발 할 수 있을 것입니다. 이는 대부분의 자원을 수입하는 우리나라의 특성 상 필수적인 과업이 될 것입니다.

## 2 가스 공급량 예측 모형 개발

### 느낀 점 및 향후 발전 계획

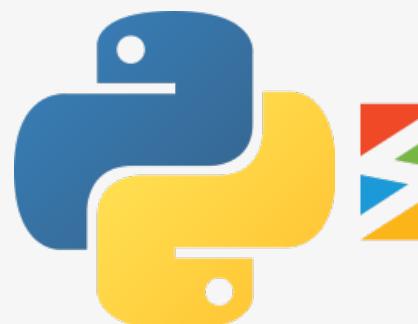
#### 외부 데이터 사용을 통한 다양한 데이터 경험

- 외부 데이터 사용을 권장했던 대회로, 기준 제공되는 변수를 제외하고 공급량을 잘 설명할 수 있는 외부 데이터의 변수들을 탐색하고 적용하는 과정을 통해 데이터를 보는 시야가 넓어질 수 있는 기회가 되었습니다.
- 외부데이터 추가에 대해 각자 생각하는 방향이 다름으로 인해서 생겼던 갈등 등을 해결하며 협업으로 인해 생길 수 있는 상황들을 경험해볼 수 있었습니다.

#### 시계열 분석 경험

- 공급량 예측을 위해 사용한 변수들을 예측하기 위해 시계열 분석을 진행하였습니다. 공급량의 기록이 1시간 단위로 기록되어 있으므로, 다른 변수들 또한 공급량에 맞게 1시간 단위로 처리하여 분석을 진행하였습니다.
- 시계열 분석 모형으로는 ARIMA, PROPHET, LSTM 모형을 구현하여 사용하였습니다.

#### Tools



LightGBM



pandas



#### TensorFlow 경험

- TensorFlow를 사용하여 딥러닝 알고리즘 중 LSTM 모형을 구현, 사용해볼 수 있었습니다.
- TensorFlow를 이용한 알고리즘 구현은 손쉽게 GPU 사용을 가능하게 하여 빠른 연산을 할 수 있도록 도와주었습니다.

#### 댐 공급량 모형 개발 이후 미비점 보완

- 댐 공급량 모형 개발을 하면서 사용하지 못했던 TensorFlow 프레임워크 사용과 천연 가스에 대한 도메인 지식을 더 자세히 숙지하여 전처리 방안에 대해 어떤 방식이 더 옳은 방향으로 갈 수 있는지 고민하였습니다.
- 다양한 주제와 데이터를 통한 문제점 도출이 익숙해 질 수 있도록 앞으로도 다양한 프로젝트에 참여할 계획입니다.

# SKT CCTV 보안영상 이미지 객체 탐지



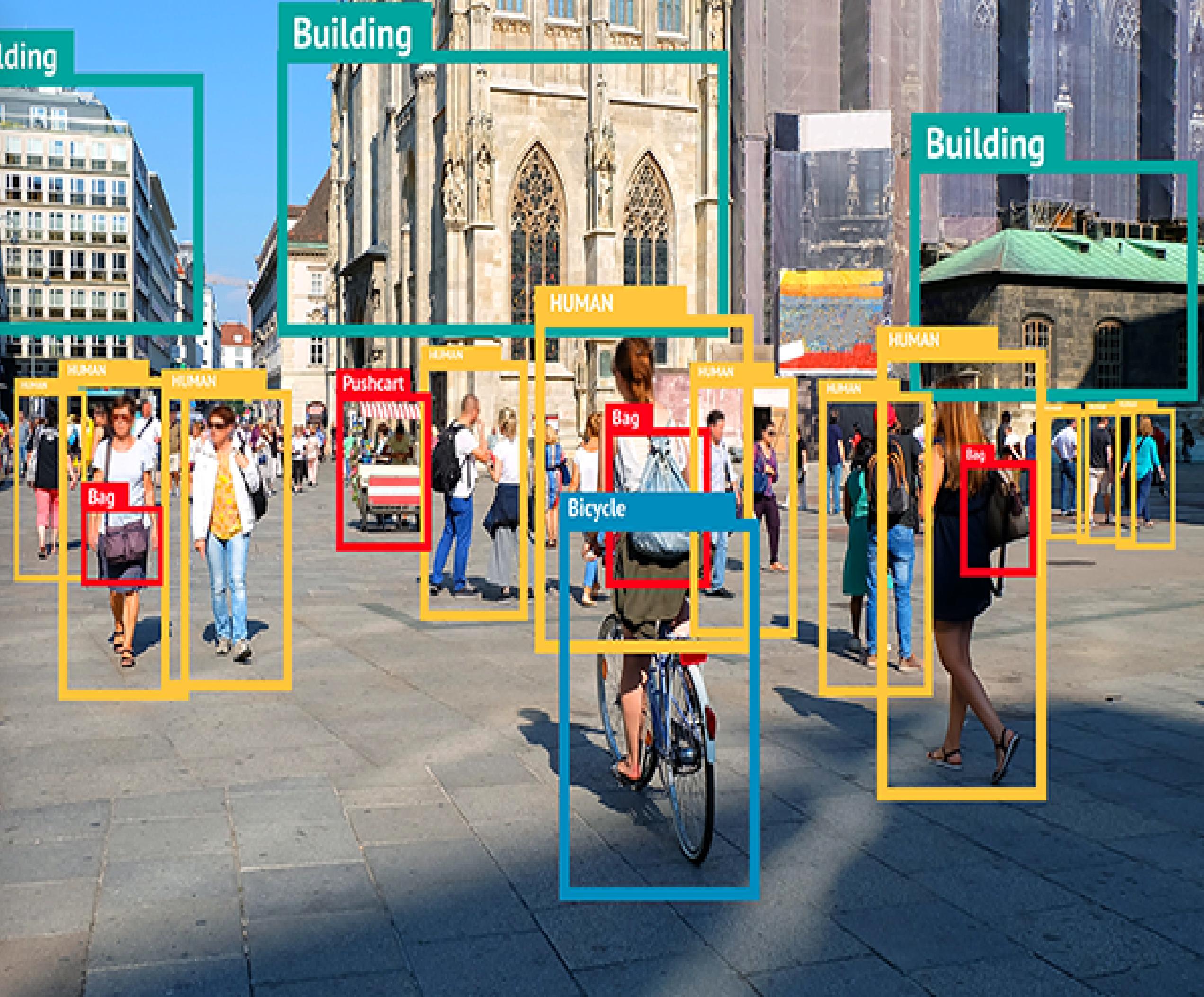
본 프로젝트를 통해 SK텔레콤은 도어가드, 이너가드, 무인 경비, 무인안심존 등의 CCTV 영상 기반 서비스에서 현재 사용 중인 Object Detection 모델의 성능 향상 및 기능 고도화 추진을 가속화할 수 있을 것이라 기대합니다.

3

작업성격 | Object Detection

작업역할 | 준지도학습을 이용한 객체탐지 모델링

작업시기 | 2021.09 - 2021.10



### 3 SKT CCTV 보안영상 이미지 객체 탐지

#### Problem

##### 보안시장의 성장, 그러나..

보안시장은 사물인터넷 확대, 디지털 경제 가속화와 함께 지속적인 성장세를 보이고 있습니다.

- Computer Vision 기술의 고도화는 보안시장에 서의 보안 기술 성능 고도화에 도움을 줍니다.
- 그러나 보안영상의 경우, 쓰러진 사람을 인식하지 못하는 경우, 악천후와 외부 요인에 의한 성능저하 등이 해결해야 할 과제로 직면하고 있습니다.

##### 객체탐지에서의 지도학습 한계

객체 탐지를 위한 지도학습에서는 이미지 데이터와 그에 맞는 라벨링 데이터를 필요로 합니다.

- 이미지에 맞는 라벨링을 수행하기 위해서는 많은 시간이 소요되며 비용 또한 늘어나게 됩니다.

#### Solution

##### 이미지 데이터 전처리를 통한 성능 향상

이미지 데이터를 전처리 하는 과정을 통해 같은 이미지를 이용하더라도 전처리 방식의 차이가 성능 향상을 이끌어 낼 수 있습니다.

##### 객체탐지모델 구현 성능 향상

모델 성능의 향상은 더 많은 데이터를 효율적으로 처리할 수 있어 시간 단축에도 도움이 될 것입니다.

##### 준지도학습 기반의 객체탐지 모형 구현

라벨링이 되어있는 데이터와 라벨링이 되어있지 않은 데이터를 함께 사용하여 사용 가능한 이미지의 수를 증대시켜 학습에 도움을 주는 모형을 구현하는 것을 목표로 하였습니다.

#### Result

##### 객체탐지 모형인 YOLO 모델 사용

- YOLO 모델은 이미지 전체에 대해서 하나의 신경망이 한 번의 계산만 수행하여 예측하므로 시간 단축에 큰 효과를 발휘합니다.
- YOLO 모델은 예측 시 이미지 전체를 예측합니다. 슬라이딩 윈도를 쓰는 다른 객체 탐지 모델과 달리, 이미지 전체를 예측하여 주변의 정보 또한 학습에 사용할 수 있습니다. 이를 통해 아무 배경이 없는 상태에서 노이즈 발생 시 물체로 인식하는 문제를 해결합니다.

### 3 SKT CCTV 보안영상 이미지 객체 탐지

---

## 느낀 점 및 향후 발전 계획

### 객체 탐지에 대한 기초 이해

- 객체 탐지에 대해 기초 지식 없이 참여한 대회로, 컴퓨터 비전에 대해 알아가고 배워가는 동시에 모델에 적용하여 성능을 향상시킬 수 있었습니다.
- 객체 탐지를 위한 데이터 적용 및 데이터 전처리 과정 또한 객체 탐지에서 중요한 과정으로, Pytorch에 대한 이해가 필수적이었습니다.  
객체탐지와 함께 Pytorch라는 프레임워크를 사용할 수 있게 되었습니다.

### 다양한 데이터 경험

- 데이터 분석 과정을 수강하며 경험했던 대부분의 데이터는 수치형으로 이루어진 정형데이터였습니다. 이미지와 텍스트로 이루어진 비정형 데이터를 다루면서 데이터 처리 방안에 대해 공부할 수 있는 기회가 되었습니다.

### Tools



### 컴퓨터 비전에 대한 관심 증가

- 이번 대회를 진행하면서 이미지 처리를 통한 객체 탐지 알고리즘은 제게 컴퓨터 비전에 대한 새로운 흥미로 다가오게 되었습니다. 이를 계기로 컴퓨터 비전에 대해 공부하며 다른 대회도 참가하여 역량을 발전시킬 계획입니다.

# 스타트업 챌린지 고객은 누구인가



본 프로젝트는 디캠프에서 주관한 스타트업 챌린지 중 '한달어스'라는 스타트업과 협업하였습니다.

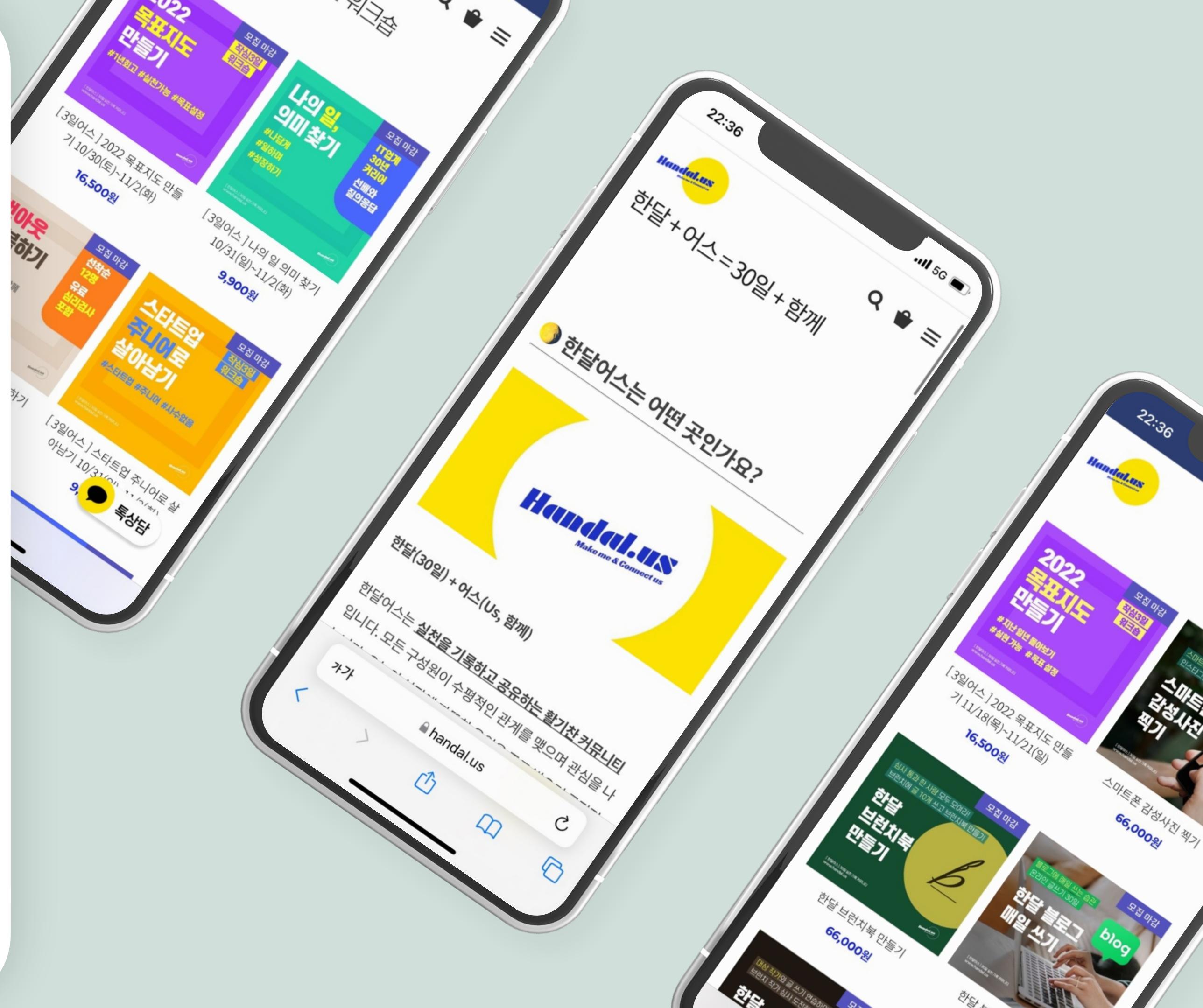
'기업의 고객은 누구인가?'라는 주제로 프로젝트를 진행하였습니다.

작업성격 | Data Analysis

작업역할 | 데이터를 활용한 고객 페르소나 도출

작업시기 | 2021.10

4

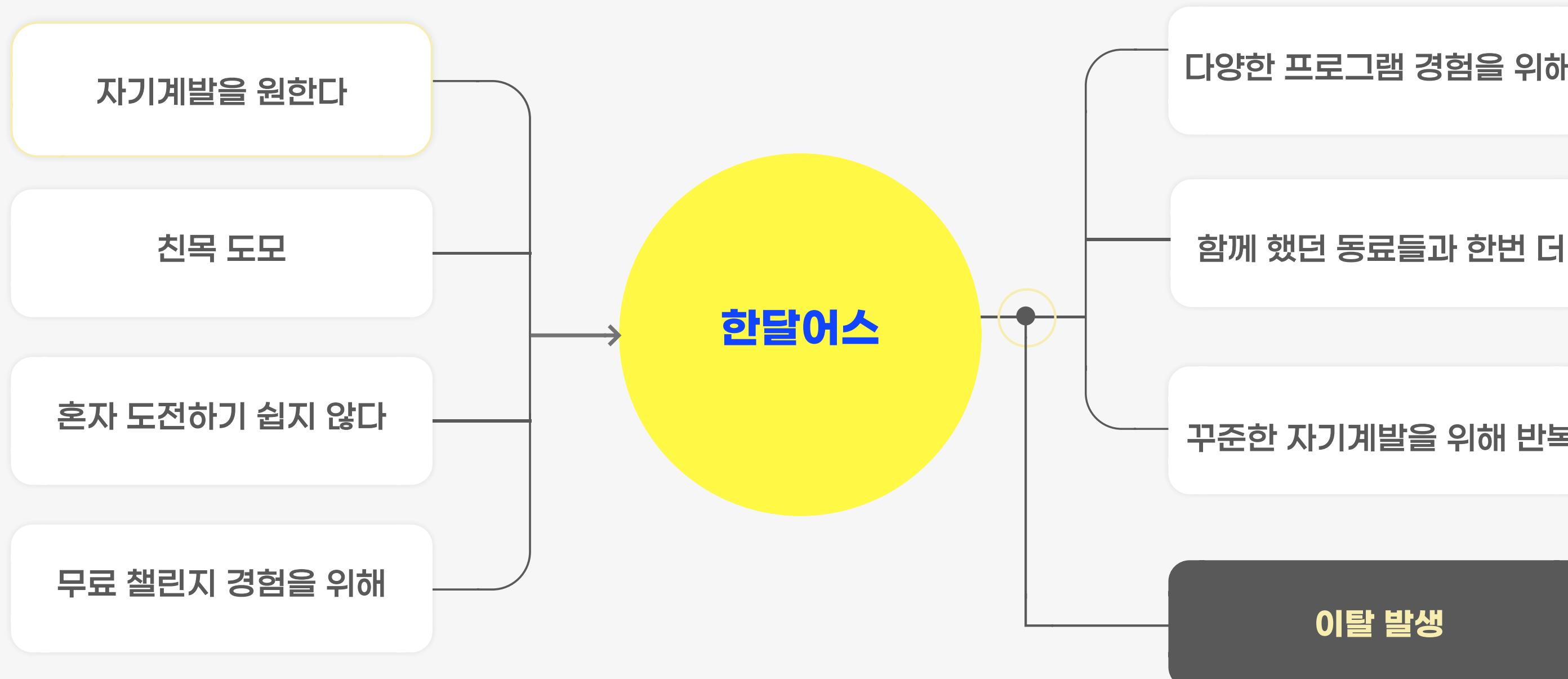


## 4 스타트업 챌린지 - 고객은 누구인가

### Problem

'한달어스'라는 기업은 꾸준한 자기계발을 바라는 2030 고객을 대상으로 한달 챌린지 서비스를 제공중인 스타트업입니다. 한달 챌린지를 할 수 있는 어플 및 상품들이 많고, 경쟁 업체가 많은 교육시장에서 기업이 더욱 성장하기 위해서는 이용 고객 증대와 이용 고객의 특성을 파악한 상품들을 지속적으로 개발 및 서비스 하는것이 중요합니다.

-> 따라서 현재 '한달어스'를 이용하는 고객이 누구인지 특정하기 위해 퍼르소나를 제작하기로 하였습니다.



## 4 스타트업 챌린지 - 고객은 누구인가

### 페르소나 1



이름 : 김영숙

거주지 : 서울 마포구

나이 : 46 세

직업 : 영어 강사

가족 : 남편, 자녀

참여 횟수 : 6번

사용 SNS : 인스타그램, 페이스북, 블로그

**ABOUT :** 영어 강사라는 직업의 경력과 커리어에 대해 고민 중, 한 달어스를 알게 되어 한달 글쓰기로 입문하였다. 글쓰는 것이 처음에는 힘들었지만, 점점 글쓰기에 자신감이 생겼다. 함께 하는 사람들이 좋아 다른 프로그램도 함께 참여하는 중이다.

### Solution

#### 데이터를 통한 고객 특성 파악

- 한달어스를 많이 사용한 고객을 특정하기 위해 여러번 챌린지에 참여한 상위 10%의 고객을 뽑아 성별, 연령, 거주지 등을 정했습니다.
- 또한, 챌린지 첫 날의 SNS 인증 데이터를 참고하여 챌린지에 참여하는 고객의 각 오를 페르소나에 반영하였습니다.

#### 많이 이용하는 프로그램 파악

- 고객이 많이 이용하는 프로그램은 한달 독서 챌린지 프로그램이었습니다.

#### 기대효과

- 이 프로그램을 통한 마케팅으로 더 많은 고객이 유입 될 것으로 기대합니다.
- 또 다른 특성을 지닌 고객 페르소나를 추가할 수 있다면 기업 마케팅의 방향을 설정하는 데 도움이 될 것입니다.

## 4 스타트업 챌린지 - 고객은 누구인가

---

### 느낀 점 및 향후 발전 계획

#### 실제 기업 데이터 경험

- 실제 기업 데이터를 경험하여 기존 프로젝트에서 다뤄보았던 데이터와의 차이를 알 수 있었습니다. 먼저 데이터 수집에 대한 중요성이었습니다. 수집하고자 하는 데이터가 데이터 베이스에 존재해야 했지만, 데이터의 부재로 인한 분석의 한계가 있어 아쉬웠습니다.

#### 데이터 분석 및 활용 방안에 대한 고찰

- 단순히 데이터를 분석하고 끝나는 것이 아닌, 향후 어떻게 사용할 것인지에 대해 고민해볼 수 있는 프로젝트였습니다. 이를 통해 데이터 분석 이전에 미리 향후 활용 계획을 작성할 수 있다면 더 방향에 맞는 분석을 수행할 수 있을것임을 느꼈습니다.

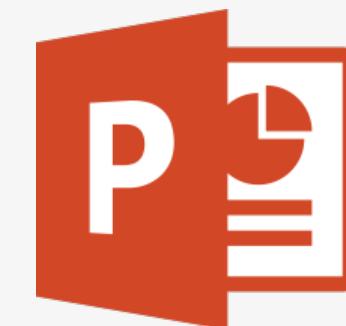
#### 현업에서의 업무 프로세스 체험

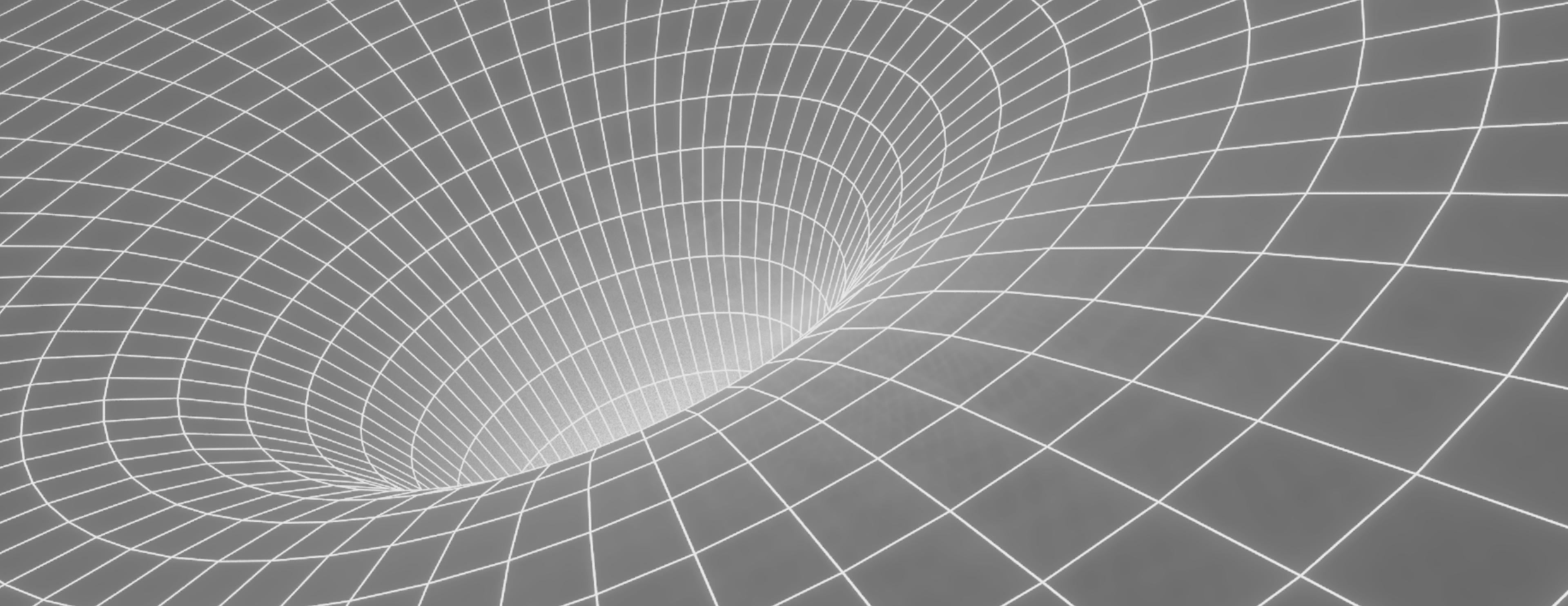
- 문제점 파악 및 분석 주제 설정에 대한 과정들에 대해 체험해 볼 수 있었습니다. 그리고, 위 과정이 제대로 동반하지 않은 프로젝트는 중간에 난항을 겪는 일이 더욱 많아질 수 있기 때문에, 프로젝트에 관계없이 중요한 과정임을 알 수 있었습니다.

#### 기업의 성장을 위한 분석 역량 강화

- 결국, 분석의 목표는 기업의 매출 혹은 규모 성장이 동반되어야 하기 때문에, 데이터 분석의 방향을 결정하는 판단력을 기르기 위해 더 많은 상업적인 데이터를 다루어 보며 경험을 쌓고, 노력해야 함을 느꼈습니다.

Tools





---

★ **THANK YOU**