

# BIGDATA HACKATHON 문제 구성

- 유형1: 보험상품 정보 분석을 통한 보험상품세트 구성 및 판매대상, 판매기법 등 아이디어 제시.
- 유형2: 우체국의 우체국 택배 배송정보 분석을 통한 예측 및 아이디어 제시

# 유형1 문제 설명

- 보험상품 정보 분석을 통한 보험상품세트 구성 및 판매대상, 판매기법 등 아이디어 제시.
  - "우체국 보험 통계연보"와 공공데이터들(<https://www.data.go.kr>)을 이용하여 자유롭게 아이디어를 제시
  - 예시: 보험 업계에서는 보험 계약자들이 차년도에 계약을 갱신할지, 해약할 지는 매우 중요한 문제
    - 보험 계약자들의 해약률 등은 국가 경제지표와 상관성이 큰 것으로 알려져 있음

## 유형2 문제 설명

- 우체국의 우체국 택배 배송정보 분석을 통한 예측 및 아이디어 제시
- 세부 주제: 제공된 택배 배송 데이터를 분석하여, **택배 배송 수량을 예측**
- 택배 수량 예측은 택배 서비스 운영에 매우 중요한 문제
  - 택배 수량은 **다양한 외부 요인에 의해 변화**
  - 수량이 예측 가능하면 집배원 등을 적재적소에 미리 배치 가능
  - 결과적으로 배송 비용 절감 그리고 배송 시간 단축을 할 수 있음

# 제공 데이터

- 우체국 배송정보: 데이터 컬럼 정보
  - 접수일자
  - 접수국
  - 보내는사람 주소 (상세주소는 제외)
  - 받는 사람 주소 (상세주소는 제외)
  - 물량
- CSV 포맷으로 파일 제공
  - 각각의 컬럼은 콤마(,)로 구분됨
- 1년치 정보를 제공
  - 하반기 공휴일 전후 3일간 물량 정보는 제공되지 않음

# 문제

- 문제 목표
  - 하반기 공휴일(포함한) 전후 3일의 받는 사람 주소별 택배 물량 예측
- 예측 결과 제출 포맷
  - 데이터 컬럼 정보
    - 접수일자
    - 받는 사람 주소 (상세 주소 제외)
    - 수량
  - CSV 포맷의 파일로 제출할 것 (컴마로 구분)
- 외부데이터를 자유롭게 활용
  - 공공데이터포털(<https://www.data.go.kr/>)을 활용해 미리 외부 데이터 준비
  - 함께 제공될 우편 통계집도 활용 가능

# 예측 결과 점수 산출 방식

- 기본적으로 Mean Squared Error 를 이용할 예정

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)$$

- n: 예측 결과 행 개수
  - $Y_i$  : 실제 수량
  - $\hat{Y}_i$  : 예측 수량
- 
- 자세한 산출 방식은 대회 시작 전 공지할 예정

# 예측 결과 제출

- 웹페이지를 통하여 제출
  - 각 팀당 아이디/비밀번호 제공
  - 제한된 횟수로 제출 가능
- 예측 결과 파일을 제출하면 산출된 예측 결과 점수를 제공

# 발표 심사

- 예측 결과 점수 기준으로 상위권 팀들에 대하여 발표 심사
  - 발표 심사 시간: **21일** 오후 **1시-오후5시**
  - 발표 및 **Q&A**로 구성
- 발표 내용
  - 아이디어
    - 어떤 외부 데이터들을 사용하여 수량을 예측하였는지 설명
    - 그러한 외부 데이터들을 활용한 이유를 설명
  - 기술
    - 어떤 프로그래밍/기술적 방법으로 수량값을 예측하였는지 설명