**1. title값에 문제가 있어 보이는 경우, title\_error(초록색)로 표시하였으며, title\_error의 의미는 아래와 같습니다.**

1. title\_error=1

: title에 해당 기업이 아닌 다른 기업의 analyst report가 들어간 경우

|  |
| --- |
| Ex) KB 손해보험 |
| …  - KB 손해보험의 title에 ‘LIG손해보험’이 있는 경우 |

**KB 손해보험의 이전 이름이 LIG 손해보험 이었음 (문제없음)**

1. title\_error=2

: title에 **기업명만** 들어간 경우 **(의미 없는 자료이니 제거할 것)**

|  |
| --- |
| Ex) KB 손해보험 |
| - KB 손해보험의 title에 ‘KB손해보험’만 즉, **기업명만 있는 경우** |
| Ex) LG 화학 |
| - LG화학의 title에 ‘LG화학’만 있는 경우 |
| Ex) 한화솔루션 |
| - 한화솔루션의 title에 ‘한화케미칼’만 있는 경우 (한화솔루션=한화케미칼) |
| Ex) 오리온홀딩스 |
| - 오리온홀딩스의 title에 ‘오리온’만 있는 경우 (한화솔루션=한화케미칼) |

1. title\_error=3

: title에 해당 기업명이 없어 해당 기업의 title인지 확인 필요한 경우

Ex) 예를 들어 ‘KB 손해보험’의 title이 ‘손해보험-1월 손해율 개선을 통해 시장의 우려 해소’ 인 경우, 손해보험이 **KB손해보험인지 LIG손해보험인지 확인이 필요**해 보입니다.   
그 이유는 title\_error=1과 같이 title에 ‘KB손해 보험’이 아닌 ‘LIG손해보험’이 들어간 경우가 분명히 존재하기 때문입니다. **(두 회사는 동일 기업)**

Ex) 또 다른 예로 **기업이 다른 기업**임에도 불구하고(메리츠화재, 삼성화재, 현대해상, KB 손해보험, DB 손해보험) **title이 동일한 경우**가 있었습니다. 년, 월, 일은 각각 2016년, 3월, 2일로 동일합니다.   
(해당 리포트는 특정기업이 아니라 업계 현황에 대한 리포트이므로 제거)

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

**※ 추가적으로 title\_error=3인 경우의 예시는 아래와 같습니다.**

Ex) 메리츠화재, 한화손해보험, 삼성화재, 현대해상의 **title이 동일한 경우** (2016년 11월 30일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, 현대해상, DB손해보험의 **title이 동일한 경우** (2017년 6월 1일)

|  |
| --- |
|  |

Ex) 메리츠화재, 현대해상의 **title이 동일한 경우** (2018년 1월 1일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, 현대해상, 코리안리, DB손해보험의 **title이 동일한 경우** (2018년 1월 2일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, KB손해보험, 코리안리, DB손해보험의 **title이 동일한 경우** (2019년 4월 3일)

|  |
| --- |
|  |

Ex) 삼성화재와 KB손해보험의 **title이 동일한 경우** (2009년 5월 11일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2010년 1월 5일)

|  |
| --- |
|  |

Ex) 삼성화재와 현대해상의 **title이 동일한 경우** (2010년 2월 24일)

|  |
| --- |
|  |

Ex) 메리츠화재, 현대해상, KB손해보험, 코리안리, DB손해보험의 **title이 동일한 경우** (2011년 8월 1일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2014년 10월 20일)

|  |
| --- |
|  |

Ex) 한화손해보험, 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2014년 10월 20일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, KB손해보험의 **title이 동일한 경우** (2014년 12월 1일)

|  |
| --- |
|  |

Ex) 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2015년 12월 1일)

|  |
| --- |
|  |

Ex) 메리츠화재, 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2015년 12월 1일)

|  |
| --- |
|  |

Ex) 메리츠 화재, 한화손해보험, 삼성화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2015년 12월 1일)

|  |
| --- |
|  |

Ex) 메리츠화재, 현대해상, KB손해보험, DB손해보험의 **title이 동일한 경우** (2015년 3월 2일)

|  |
| --- |
|  |

Ex) 우리금융, 신한지주, 하나금융지주의 **title이 동일한 경우** (2009년 1월 6일)

|  |
| --- |
|  |

Ex) ‘세방전지’와 ‘한국아트라스비엑스’의 **title이 동일한 경우** (2009년 7월 9일)

|  |
| --- |
|  |

Ex) ‘삼성전자’와 ‘LG 이노텍’의 **title이 동일한 경우** (2010년 8월 6일)

|  |
| --- |
|  |

Ex) ‘한화솔루션’와 ‘LG 화학’의 **title이 동일한 경우** (2011년 1월 18일)

|  |
| --- |
|  |

**2. 행 전체가 type 측면에서 문제가 있어 보이는 경우, type\_error(노란색)로 표시하였으며, 예시는 아래와 같습니다.**

|  |
| --- |
| Ex) |
| …  - code에 “예상보다…” 라는 문장이 존재  - idx에 “BUY” 존재 |

**(잘못 입력된 것으로 보임. 제거할 것)**

**3. 해당 기업의 title이 잘 들어간 data로 error는 아니지만, 같은 값이 중복된 경우, title\_중복(파란색)으로 표시하였으며, 예시는 아래와 같습니다.**

**(중복된 것을 하나로 만들것)**

|  |
| --- |
| Ex) |
| … |

**4. 위에서 언급한 1번, 2번, 3번을 수작업으로 처리하기에는 한계가 있어서 coding을 통해 1번, 2번, 3번의 경우 중 하나라도 속한다면, ‘duplication=1’로 표시하였습니다.**