

M&WISE 통신 대량 메시지 전송 오류 해결 DT 아이디어 공모전

최준석 김태형 변혜림 서준영 최종현

CONTENTS

01 개요

02 EDA

03 분석

04 고찰



1 - 1

개요

- ❖ 의뢰사: M&WISE
- ❖ 공모전 제공 플랫폼: d-impact
- ❖ 주제: 대량 메시지 전송 시 오류 예측하기
- ❖ 결과물: PPT 25장 이내 (Python, R 등 활용)
- ❖ 혜택: 의뢰사 면접전형 통과 및 장학금 제공
- ❖ 교육: 관계사 담당자 Q&A, Python 중급, R 중급 교육 진행
- ❖ 설정: SQL + R studio 활용 요망



1 - 2

데이터 정보

- ❖ NVECARESENDLOG2017: 2017년 발송로그 1232289건
- ❖ NVECARESENDLOG2018: 2018년 7월까지의 발송로그 1393113건
- ❖ NVECARESENDRESULT2017: 2017년 발송결과 2903건
- ❖ NVECARESENDRESULT2018: 2018년 7월까지의 발송결과 2145건
- ❖ NVECARERECEIPT-2018: 2018년 7월까지의 수신확인 139589건



1 - 2

데이터 정보

- ❖ NVECARERETURNMAIL2012: 2012년 반송메일 6473건
- ❖ NVECARERETURNMAIL2013: 2013년 반송메일 9710건
- ❖ NVECARERETURNMAIL2014: 2014년 반송메일 15121건
- ❖ NVECARERETURNMAIL2015: 2015년 반송메일 4020건
- ❖ NVREJECT2012: 2012년 수신거부 56건
- ❖ NVREJECT2013: 2013년 수신거부 783건
- ❖ NVREJECT2014: 2014년 수신거부 121건
- ❖ NVREJECT2015: 2015년 수신거부 502건



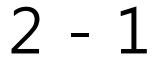
2 - 1

EDA

| | | | | | |
|------------------------------|------------------------------------|----------------|-----------------------------|----------------|---|
| <발송로그> SENDLOG2017, 2018 | | | <발송결과> SENDRESULT2017, 2018 | | |
| ECARE_NO | 이케어번호 (FK) | 시간 순으로 | ECARE_NO | 이케어번호 (FK) | 시간 순으로 |
| RESULT_SEQ | 결과순번 (FK) | 시간 순으로 | RESULT_SEQ | | 시간 순으로 |
| SID | 발송서버ID | MTS1 | | | XE: 발송완료 |
| SEND_DT | 발송일자 | YYYYMMDD | RESULT_STS | 결과상태 | XO: 오류 |
| SEND_TM | 발송시간 | HHMMSS | | | XW: 작업중, END_DT부터 모두 0, TARGET_CNT는 제외 |
| ERROR_CD | 오류코드 (250은 정상, 800이면 ERR_MSG는 701) | | START_DT | 시작일자 | YYYYMMDD |
| SEND_DOMAIN | 발송도메인 | | START_TM | 시작시간 | HHMMSS |
| ERR_MSG | 오류메시지 | | END_DT | 종료일자 | YYYYMMDD |
| | | | END_TM | 종료시간 | HHMMSS |
| <반송메일> RETURNMAIL2012 ~ 2015 | | | TARGET_CNT | 대상자수 | |
| ECARE_NO | 이케어번호 (FK) | 시간 순으로 | SEND_CNT | 발송수 | SUCCESS_CNT ~ INVALID_ADDRESS_CNT의 합과 같음. |
| RESULT_SEQ | 결과순번 (FK) | 시간 순으로 | SUCCESS_CNT | 성공수 | |
| RECEIVE_DT | 반송메일수신일자 | YYYYMMDD | FAIL_CNT | 실패수 | 0 or NULL |
| RECEIVE_TM | 반송메일수신시간 | HHMMSS | UNKNOWN_USER_CNT | UNKNOWN USER 수 | |
| SMTPCODE | SMTP코드 | 400 ~ 699, 888 | UNKNOWN_HOST_CNT | UNKNOWN HOST 수 | |
| | | | SMTP_EXCEPT_CNT | SMTP 예외수 | |
| <수신확인> RECEIPT2018 | | | NO_ROUTE_CNT | NO_ROUTE 수 | |
| ECARE_NO | 이케어번호 (FK) | 6 7 12 13 | REFUSED_CNT | 연결거부수 | |
| RESULT_SEQ | 결과순번 (FK) | 시간 순으로 | ETC_EXCEPT_CNT | 기타예외수 | |
| OPEN_DT | 수신확인일자 | YYYYMMDD | INVALID_ADDRESS_CNT | 무효주소수 | |
| OPEN_TM | 수신확인시간 | HHMMSS | LOG_STS | 로그상태 | DE or RE or LE (로그 Upload완료) or NULL |
| READING_DT | 유효개봉확인일자 | OPEN_DT와 같음 | | | |
| READING_TM | 유효개봉확인시간 | OPEN_TM과 같음 | | | |
| READING_DURATION | 유효개봉시간 | 0 ~ 11 | | | |
| VALID_CNT | 유효개봉수 | 1 | | 주황색은 필요없음 | |



- ❖ 반송메일 데이터는 옛날 데이터이면서 담고 있는 정보가 오류코드 뿐이라 분석에서 제외. 시기 별로 발생하는 오류의 유형이 다를 수 있으나 일관성이 없어서 제외.
- ❖ 수신확인 데이터는 결국 정상 발송된 메시지에 대한 데이터이므로 오류 예측에는 필요 없다고 판단하여 분석에서 제외.
- ❖ 발송로그와 발송결과 중 단일값 등의 무의미한 데이터를 제외하고 분석을 하여 오류를 예측하고자 함.
- ❖ 원본 데이터 형식인 sql 파일은 Python이나 R로 분석하려면 별도의 db 연결 작업을 거쳐야 하므로 csv 파일로 변환하여 분석.



EDA

| | ECARE | NC_RESULT | SI | SEND_DT | SEND_TM | ERROR_CD | SEND_DOF | ERR_MSG | RESULT_S | START_DT | START_TM | END_DT | END_TM | TARGET_C | SEND_CN | SUCCESS | UNKNOWN | UNKNOWN | SMTP_EXC | NO_ROUT | REFUSED | ETC_EXCEP | INVALID | _LOG_STS |
|--|-------|-----------|----------|---------|---------|----------|--------------|---------|----------|----------|----------|----------|--------|----------|---------|---------|---------|---------|----------|---------|---------|-----------|---------|----------|
| | 1 | 1.48E+15 | 20170102 | 181101 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181102 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181112 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181112 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181102 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181139 | 421 | YAHOO.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181102 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181125 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181125 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181101 | 550 | HANA.YOU | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181105 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181110 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181110 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181101 | 550 | NATE.COM | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181117 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181108 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181112 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181220 | 554 | YAHOO.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NATE.COM | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181126 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181106 | 550 | LYCOS.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181107 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181113 | 602 | YAAHU.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181147 | 421 | YAHOO.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181204 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181113 | 550 | LYCOS.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181113 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181213 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181114 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181117 | 550 | HANMAIL | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |
| | 1 | 1.48E+15 | 20170102 | 181113 | 550 | NAVER.CO | [xxx.xxx.xxx | XE | | 20170102 | 181050 | 20170102 | 181230 | 915 | 915 | 866 | 44 | 0 | 1 | 0 | 0 | 4 | 0 | RE |



- ❖ 발송로그 데이터는 각 log마다 발생하는 개행이 제대로 되지 않아 R-studio에서 데이터를 제대로 읽지 못하므로 수정 후 작업.
- ❖ 각 log들의 특성 파악을 위해 발송로그와 발송결과를 'ECARE_NO'와 'RESULT_SEQ'를 key로 하여 inner join한 후 정상발송 (오류코드 250) log를 제거했다. 각 오류코드들의 결과는 아래와 같은데 오류코드 888은 기타 오류들의 집합이므로 결과가 여러가지.

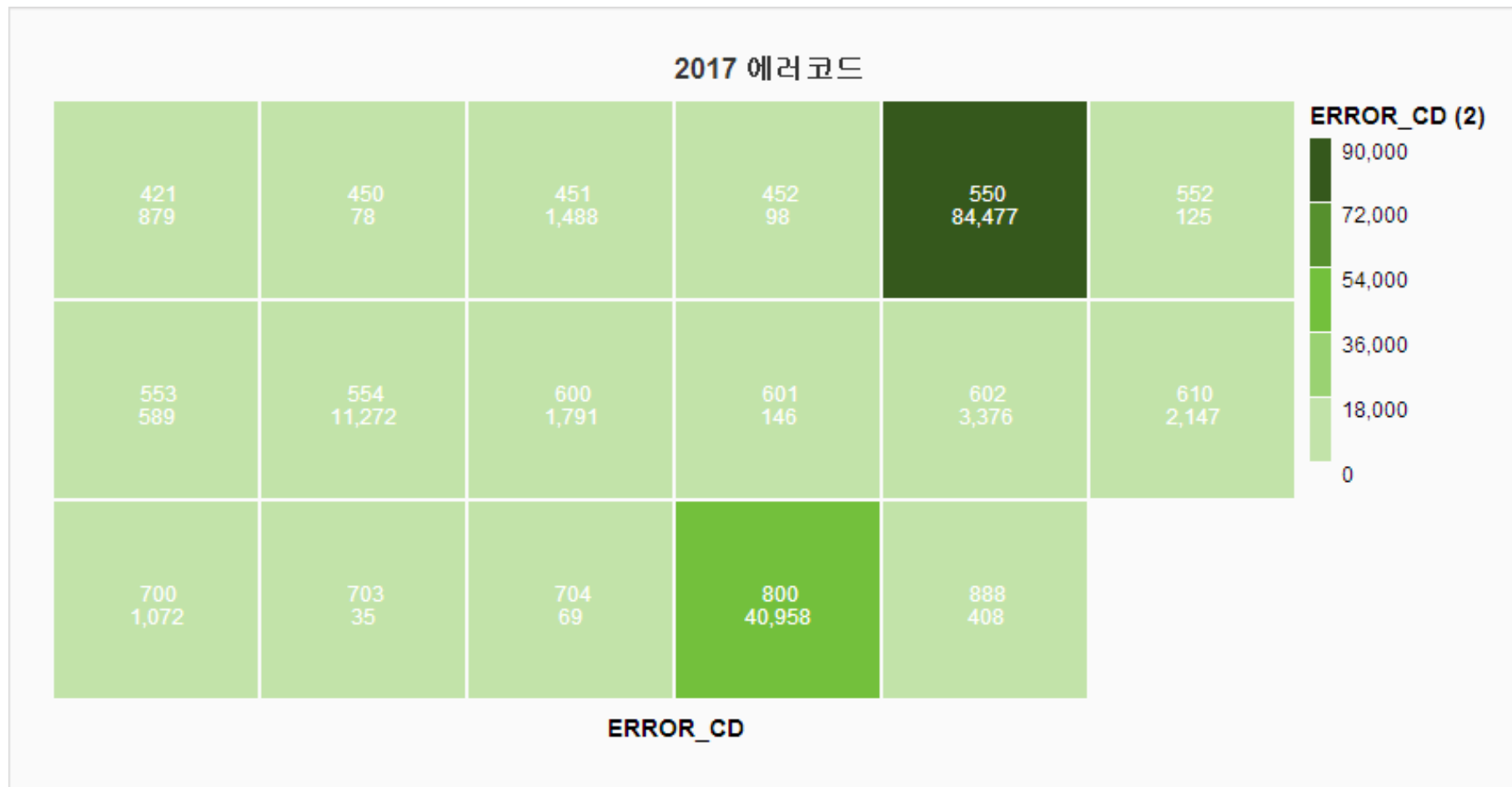
| | | | |
|--------------------|----------------------|----------------------|-------------------------|
| 421 ETC_EXCEPT_CNT | 550 UNKNOWN_USER_CNT | 600 REFUSED_CNT | 700 SMTP_EXCEPT_CNT |
| 450 ETC_EXCEPT_CNT | 552 ETC_EXCEPT_CNT | 601 NO_ROUTE_CNT | 703 SMTP_EXCEPT_CNT |
| 451 ETC_EXCEPT_CNT | 553 UNKNOWN_USER_CNT | 602 SMTP_EXCEPT_CNT | 704 ETC_EXCEPT_CNT |
| 452 REFUSED_CNT | 554 ETC_EXCEPT_CNT | 610 UNKNOWN_HOST_CNT | 800 INVALID_ADDRESS_CNT |



3 - 1

시각화

❖ 2017 에러코드 비율

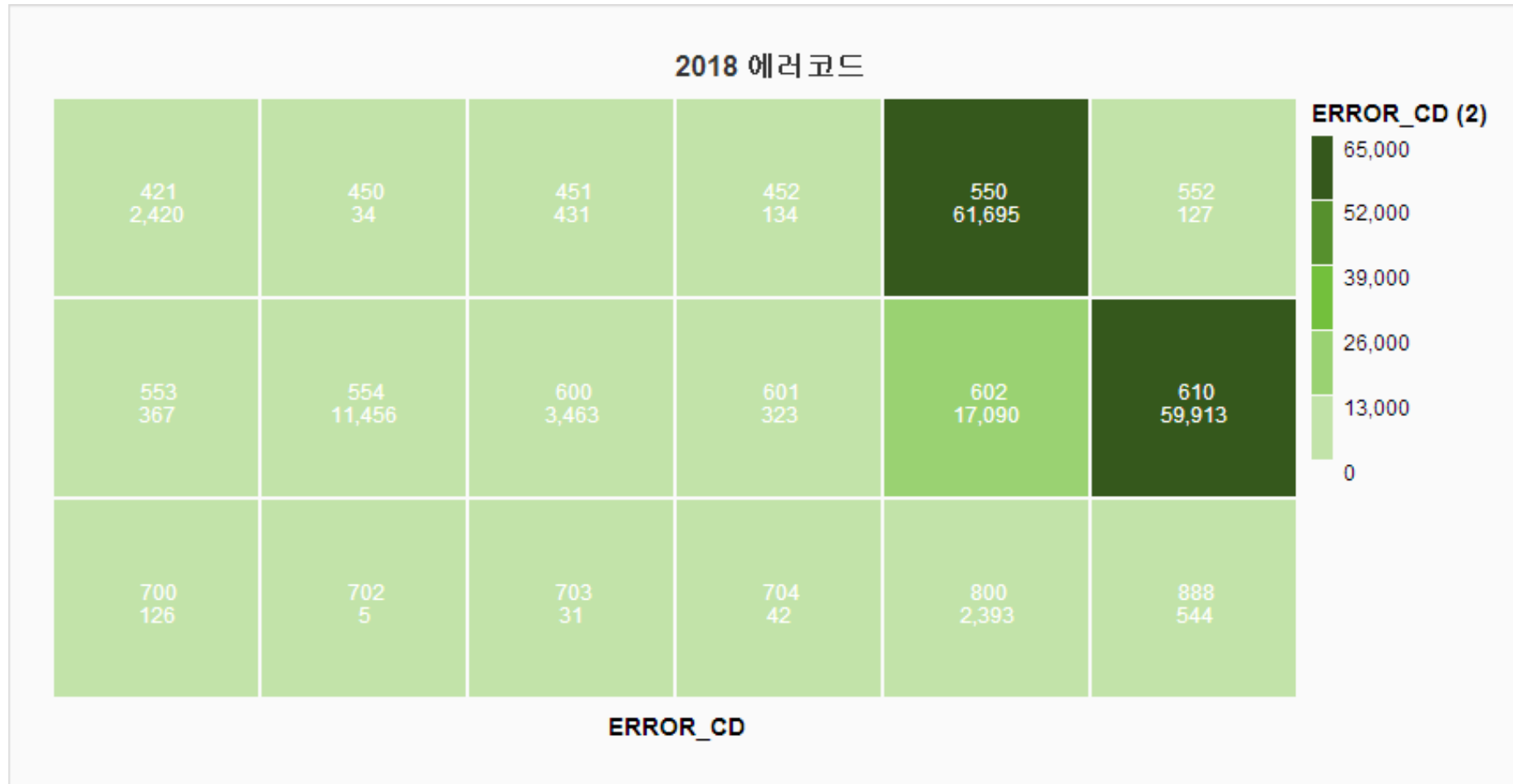




3 - 1

시각화

❖ 2018 에러코드 비율

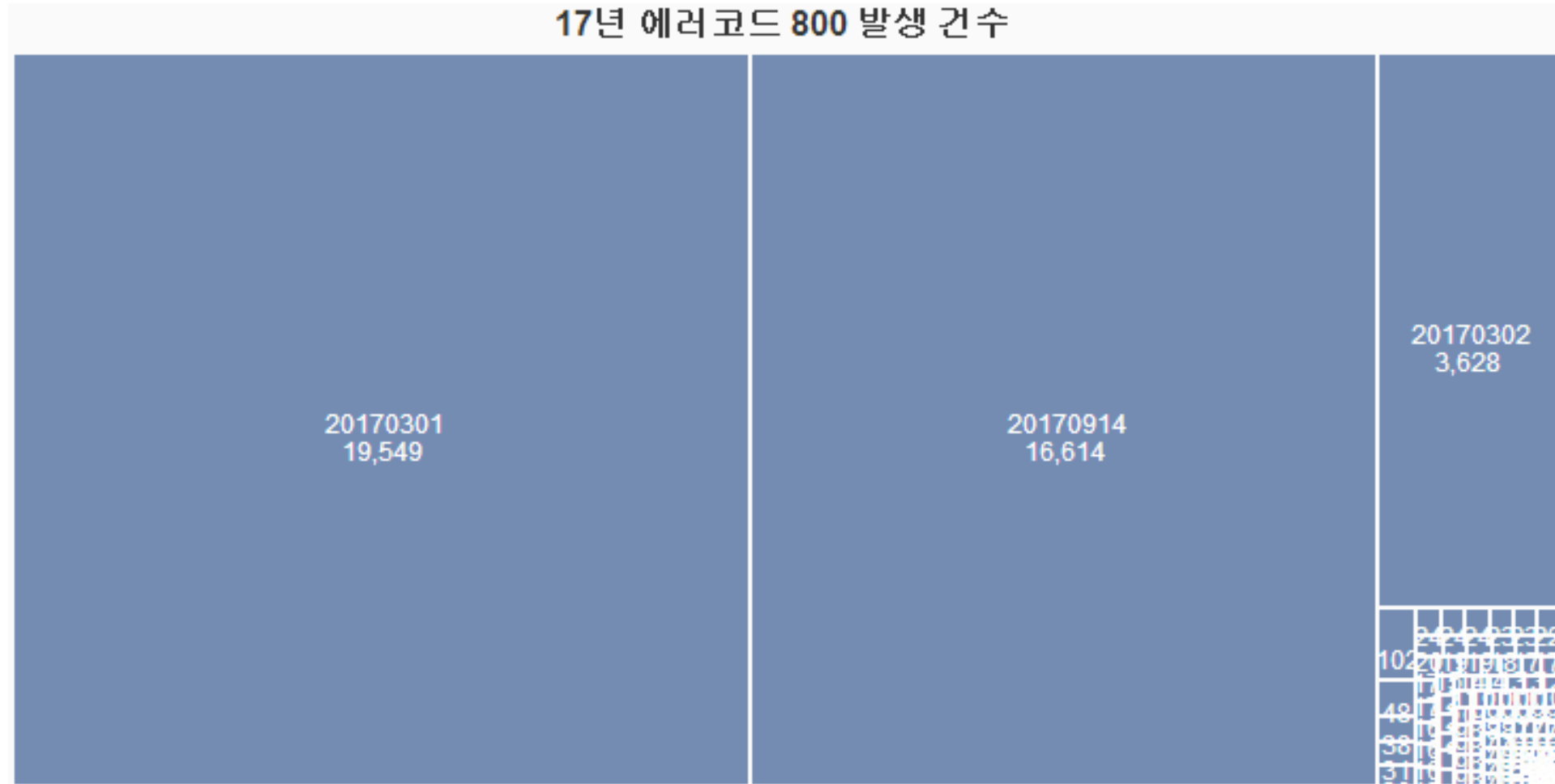




3 - 1

시각화

❖ 2017년 오류코드 800의 일자별 분포 양상





3 - 1

시각화

❖ 2017년 오류코드 800의 일자별 분포 양상





- ❖ 시각화 결과 2017년과 2018년 모두 오류코드 550에 해당하는 log가 제일 많았으나 두번째로 많은 코드는 각각 800, 610으로 다름.
- ❖ 가장 많은 오류인 오류코드 550은 SMTP 표준이며 원인은 'Mailbox unavailable'. 즉, 발송 자체에는 문제가 없지만 mailbox를 찾을 수 없거나 찾아도 접근이 안 되거나, mailbox의 정책에 의해 중간에서 막힌 것.
- ❖ 오류코드 800은 SMTP 표준이 아니며 원인은 'Invalid address', 오류 메시지는 모두 '701'이며 이는 'IOException'을 의미. 오류코드 610도 SMTP 표준이 아니며 원인은 'Unknown host', 오류메시지는 모두 '[xxx.xxx.xxx.xxx] 610 UnknownHost'.



- ❖ 2017년에는 3월 1일과 2일, 9월 14일에 오류코드 800인 log가 다량 발생, 2018년 1월 10일에 오류코드 610인 log가 상당량 발생. 두 코드 모두 발송 타겟의 이메일 주소를 잘못 적은 것과 관련.
- ❖ 위의 경우 고객이 이메일 도메인을 잘못 입력한 것이 원인이라면 도메인을 직접 입력이 아닌 선택지 제공 방식으로 입력하게 하는 방법 등으로 해결하면 좋으나 도메인이 너무 많아 완벽한 해결 불가.
- ❖ 그리고 특정 날에 다량의 오류가 발생한 것은 심각한 오류이므로 주어진 데이터 이외에 추가적인 분석이 필요함. 저 날에 시스템 오류나 서버 오류 등이 발생했다면 주어진 데이터로는 오류의 그 결과로만 분석하는 것이므로 해결 불가.



- ❖ 오류코드 800, 610처럼 가시적으로 보이는 오류는 파악하기 비교적 쉽지만 오류코드 550 등의 가시적으로 보이지 않는 오류는 그 원인에 대한 데이터가 있어야 보다 정확한 분석이 가능함. 오류메시지로 원인을 파악하기는 어려움.
- ❖ 랜덤 포레스트 등의 머신 러닝 모델에 주어진 데이터의 일부만 학습시킨 후 남은 데이터를 분석하게 한 정확도는 100%. 즉, 오류의 원인이 아닌 결과를 학습했으므로 새로운 데이터를 분석해도 그대로 결과가 오류코드에 연결되어 100%가 되는 것.
- ❖ 시계열 데이터로 활용해 머신 러닝으로 분석하려면 steady state data이어야 하나 연도별 오류 분포 양상이 너무 달라 포기.

- The End -