2020-05-06 SQL 문제 발제 – 김지환

데이터는 새로 제공되는 listings\_summary.csv 에서 import 한 테이블을 사용합니다.

0. 뉴욕시의 Airbnb에 올라온 방들에 대한 정보를 데이터베이스에 저장하고자 합니다. 각각의 칼럼에 대한 정보를 반영하여 적절한 조건으로 AIRBNB이름의 테이블을 만드세요.

* HOUSE\_ID: 집 상품을 구분하는 8자리 숫자
* HOST\_ID: 집을 올린 호스트의 아이디. 8자리 숫자
* HOST\_NAME: 집을 올린 호스트의 이름
* NEIGHBOURHOOD: 집이 속해있는 지역 이름
* LATITUDE: 집이 위치한 위도
* LONGITUDE: 집이 위치한 경도
* ROOM\_TYPE: 집의 종류(아파트, 호텔 등)
* PRICE: 가격
* MINIMUM\_NIGHT: 한번 예약 시 최소 숙박해야 하는 일 수
* NUMBER\_OF\_REVIEWS: 해당 상품에 대한 리뷰 개수
* REVIEWS\_PER\_MONTH: 1개월당 평균 리뷰 개수
* CALCULATED\_HOST\_LISTINGS: 해당 호스트가 올린 총 방의 개수
* AVAILABILITY\_365: 365일 중 가능한 일 수

1. 어떤 호스트는 Airbnb 사이트를 자신의 집을 팔기 위한 홍보 사이트로 악용하기도 합니다. 이들은 365일 중 대여 가능한 날을 일부러 적게 걸어놓고, 한번에 최소 예약해야 하는 일 수를 연중 대여 가능한 날 수 보다 더 높게 잡아 방 대여가 사실상 불가능하게끔 합니다. 그렇게 자신의 집을 홍보하여 매매합니다.

사이트를 악용하고 있는 호스트들의 아이디와 이름을 찾아주세요.

2. 지역별, 방 종류 별로 방들의 개수를 누적하여 보고 싶습니다. 지역별, 방 종류 별로 몇 개의 방이 있는지, 부분합은 얼마인지 나타내도록 쿼리를 만들어주세요.

3. 각 집들이 얼마나 오랫동안 Airbnb 사이트에 게시되어 있었는지가 궁금합니다. Months 칼럼을 새로 추가하여 각 집들이 몇 개월 동안 Airbnb 사이트에 게시되어 있었는지 채워 넣으세요. 구하지 못하는 집의 경우 0으로 두세요.

listings\_summary\_2.csv 파일은 원본 파일의 1개월 이후 Airbnb 사이트에 남아있는 방들에 대한 데이터의 일부입니다. 해당 데이터를 다른 테이블로 import 하시고 다음 문제를 풀어주세요.

4. 그 동안 몇 개의 방은 Airbnb 사이트에서 내려갔으며, 새로 올라온 방들은 새로운 데이터에 포함되어 있지 않습니다. 내려간 방들의 id를 찾아주세요.

5. 데이터베이스의 오류로 인해, 새로운 데이터에서 몇몇 집들의 위도와 경도가 다른 집의 위도와 경도로 덮어씌워졌습니다. 이 집들의 id를 출력하고, 원래 데이터를 이용하여 올바른 위도와 경도로 바꾸어 주세요.