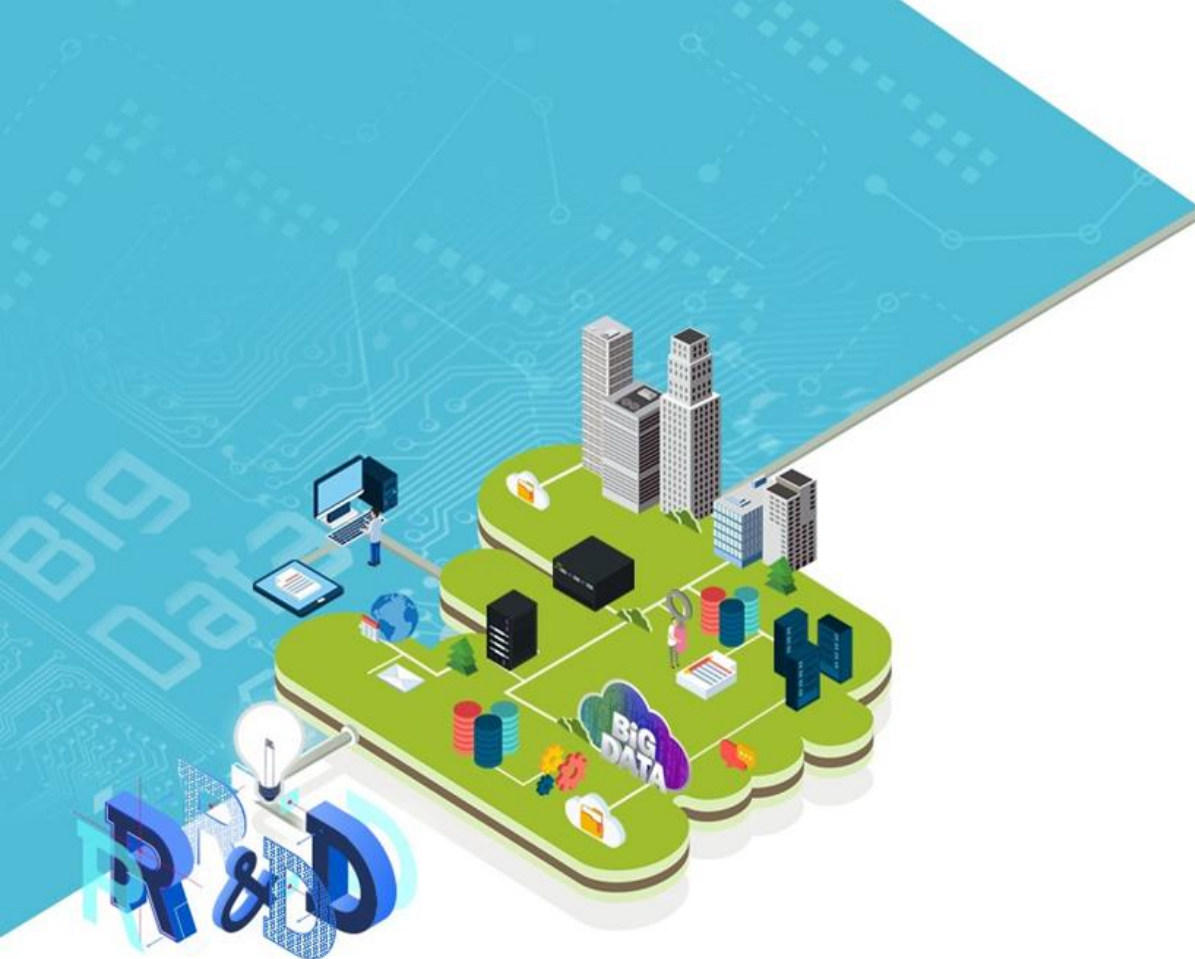


인천국제공항 유동인구 및 교통분담률 분석

로직 메뉴얼





Contents

- I 과제 개요
- II 테이블별 분석 로직
- III 산출 테이블 형태
- IV 데이터 검증

KT 유동인구 데이터를 통한 빅데이터 기반의 인천공항 유동인구와 교통분담률 분석

데이터 수집 및 정제

유동인구 데이터



KT 기지국 데이터



공공데이터



유동인구 및 교통분담률 분석

유동인구

- 인천공항과 영종도의 유동인구를 일자별 생성
- 일정 체류시간 이상의 유동인구만 추출
- 항목별(고객, 일자, 시간대, 구역, 상주지 등) 중복 제거

1. 종합데이터
2. 출발여객데이터
3. 도착여객데이터
4. 인천공항 내 상주자 출퇴근 데이터
5. 방문인구(환송객) 데이터
6. 영종도 OD데이터
7. 여행일수
8. 내부이동 데이터

교통분담률

- 유동인구 데이터 생산 후, 5개 테이블에 한하여 각 고객이 사용한 교통수단(자동차, 버스, 지하철)을 파악
- 2. 출발여객데이터
- 3. 도착여객데이터
- 4. 인천공항 내 상주자 출퇴근 데이터
- 5. 방문인구(환송객) 데이터
- 6. 영종도 OD데이터

집계 및 검증

집계

- 각 항목별 집계
- 보정계수 적용
 - 내국인, 장기외국인 : Market Share 보정 진행
 - 단기외국인: 국적별 출입국통계 보정 진행

검증

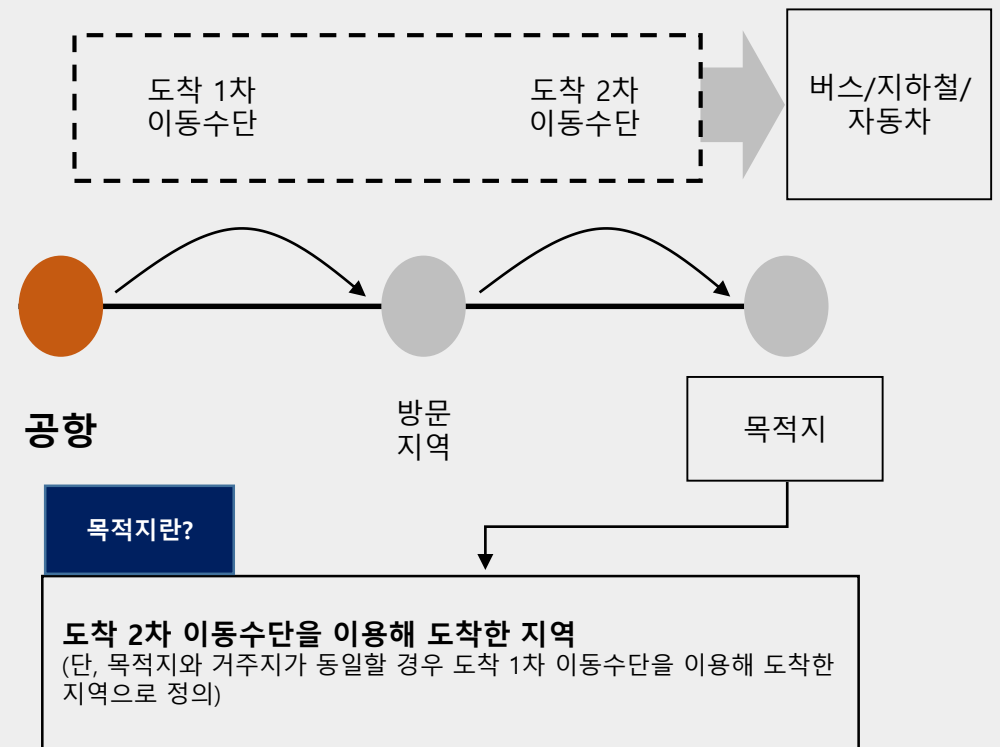
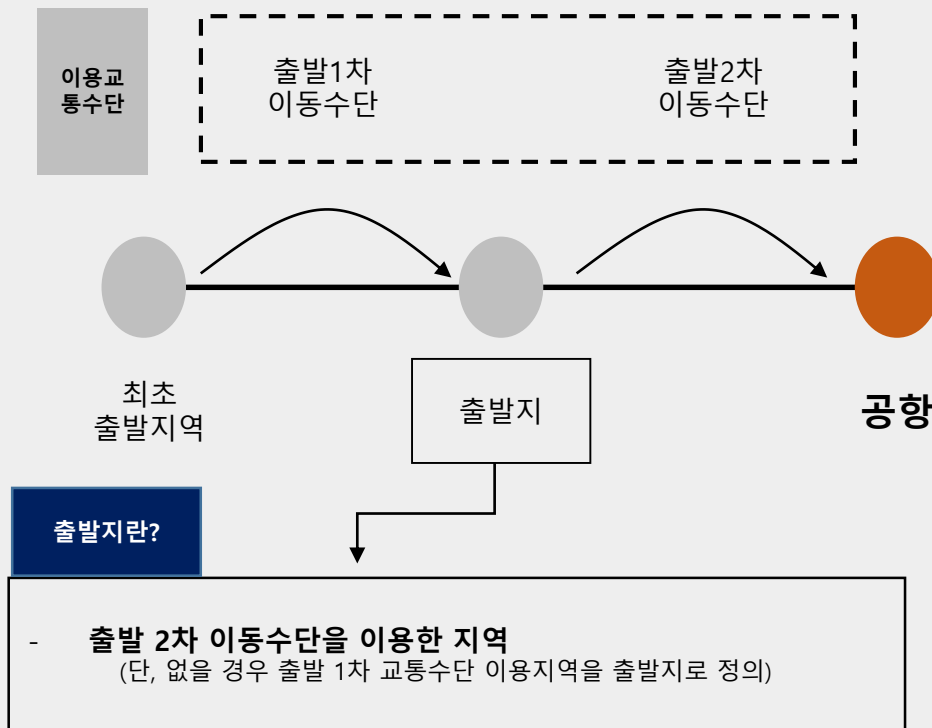
- 검증
 - 종합인구 - (출발여객 + 도착여객 + 상주인구) = 방문인구
 - 일정 기간 동안의 월별/일별 상관성 분석
 - 공항통계 자료 참고
- 검증 후, 수치보정을 위해 체류시간 및 로직 조정

과제 개요

교통수단 분담률 분석을 위한 OD 출발지 도착지 개념 정의

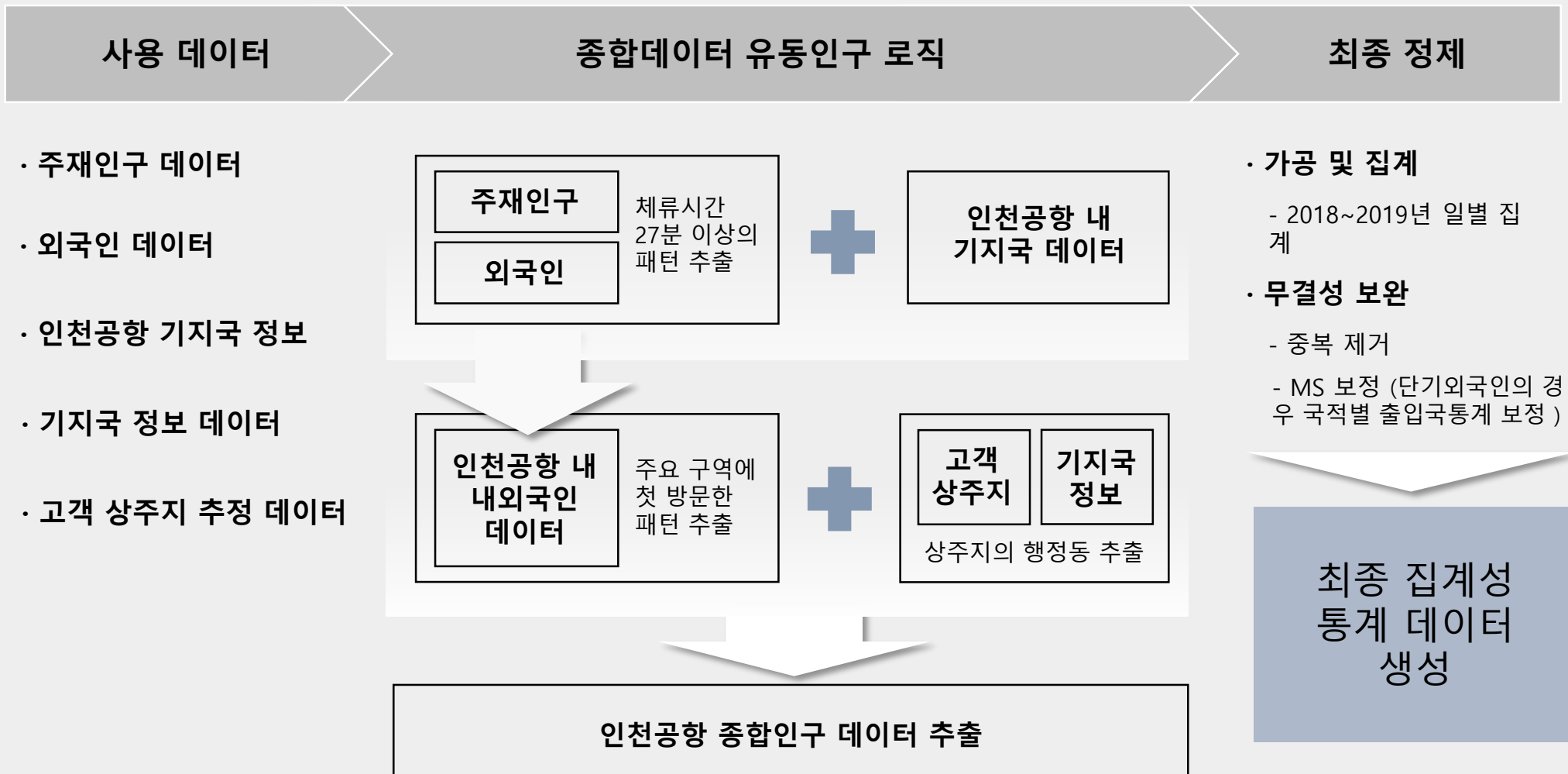
여객 1차/2차 출발/도착지 개념

모든 여객은 반드시 공항 방문 전/후 체류지에 들렸다고 가정합니다.



테이블별 분석 로직 - 1. 인천공항 유동인구 종합 데이터

인천공항 내 7개 지역에 대한 유동인구 종합 데이터의 전체 흐름도 및 유동인구 분석 로직



테이블별 분석 로직 - 2. 인천공항 출발 여객 데이터

인천공항에서 비행기를 타고 출발하는 인원에 대한 데이터의 전체 흐름도 및 유동인구 분석 로직

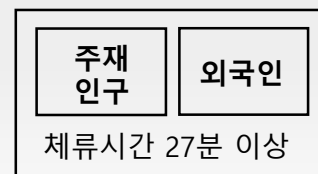
사용 데이터

출발 여객 인구

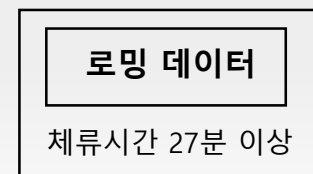
국내선

국제선

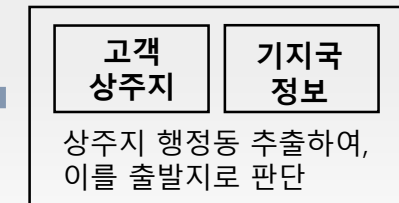
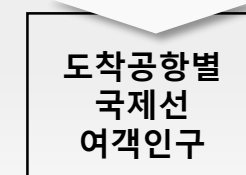
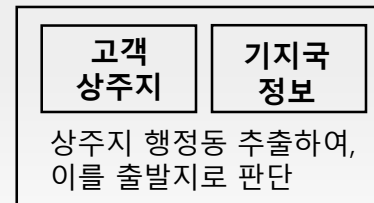
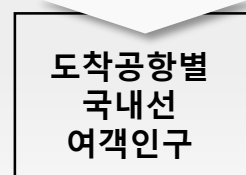
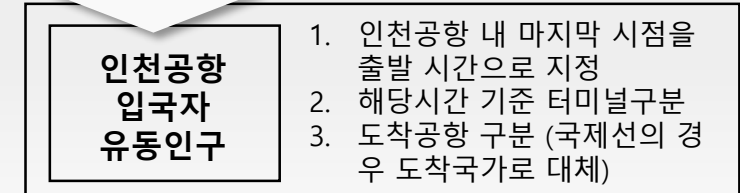
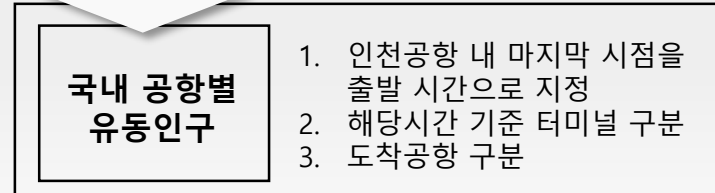
- 주재인구 데이터
- 외국인 데이터
- 로밍데이터
- 고객 상주지 추정 데이터
- 국내공항 기지국 정보
- 기지국 정보 데이터
- 공공데이터
 - 버스정차대 / 지하철역 좌표 데이터
 - 버스 평균속도 정보



국내 공항별
기지국 데이터



인천공항
기지국 데이터



인천공항 출발 여객 데이터

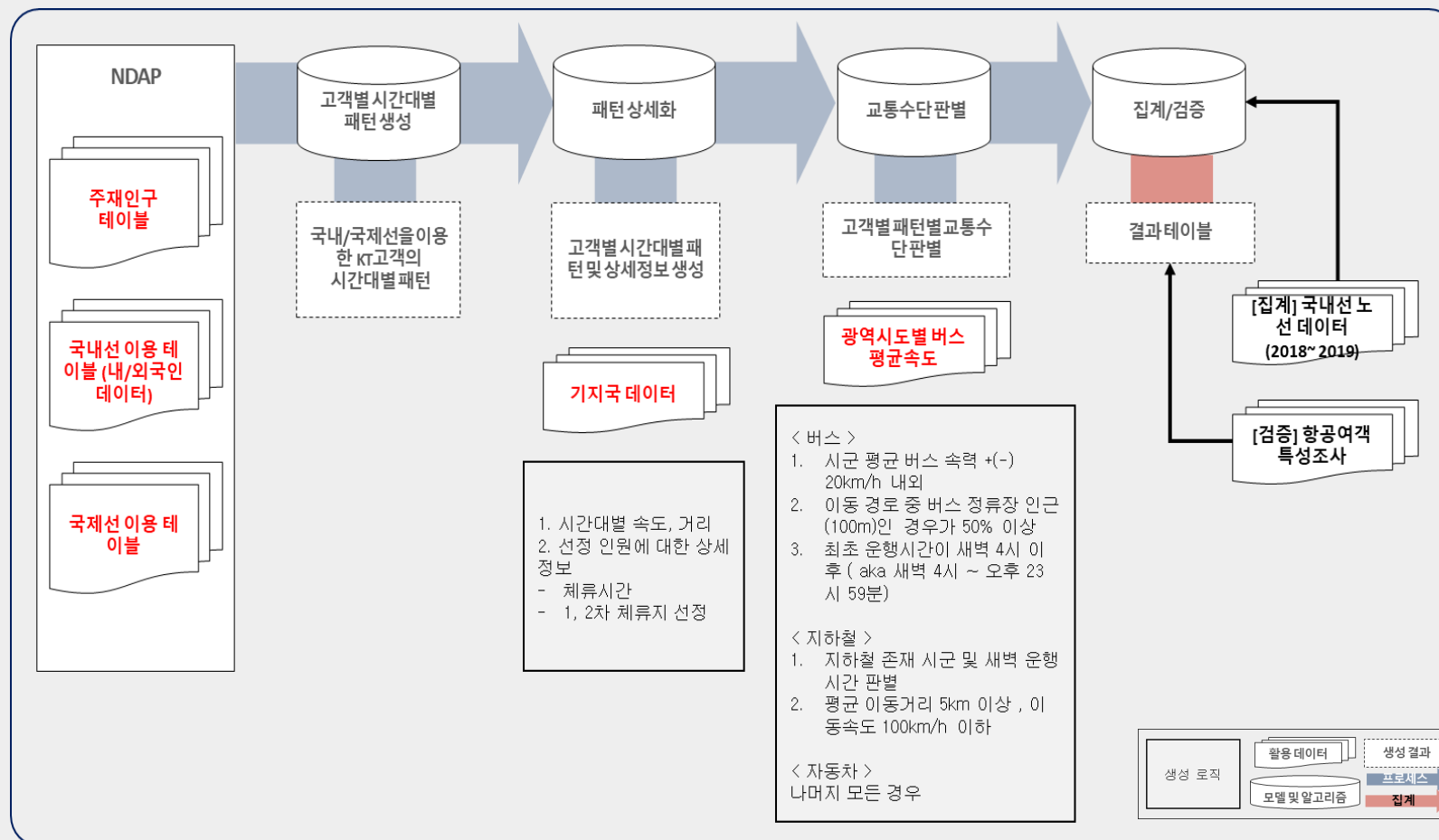
2018년 01~ 2018년 09월은
국내선 이용객만 집계

테이블별 분석 로직 - 2. 인천공항 출발 여객 데이터

인천공항에서 비행기를 타고 출발하는 인원에 대한 데이터의 전체 흐름도 및 유동인구 분석 로직

교통수단 판별

최종 정제



· 가공 및 집계

- 2019년 일별 집계

· 무결성 보완

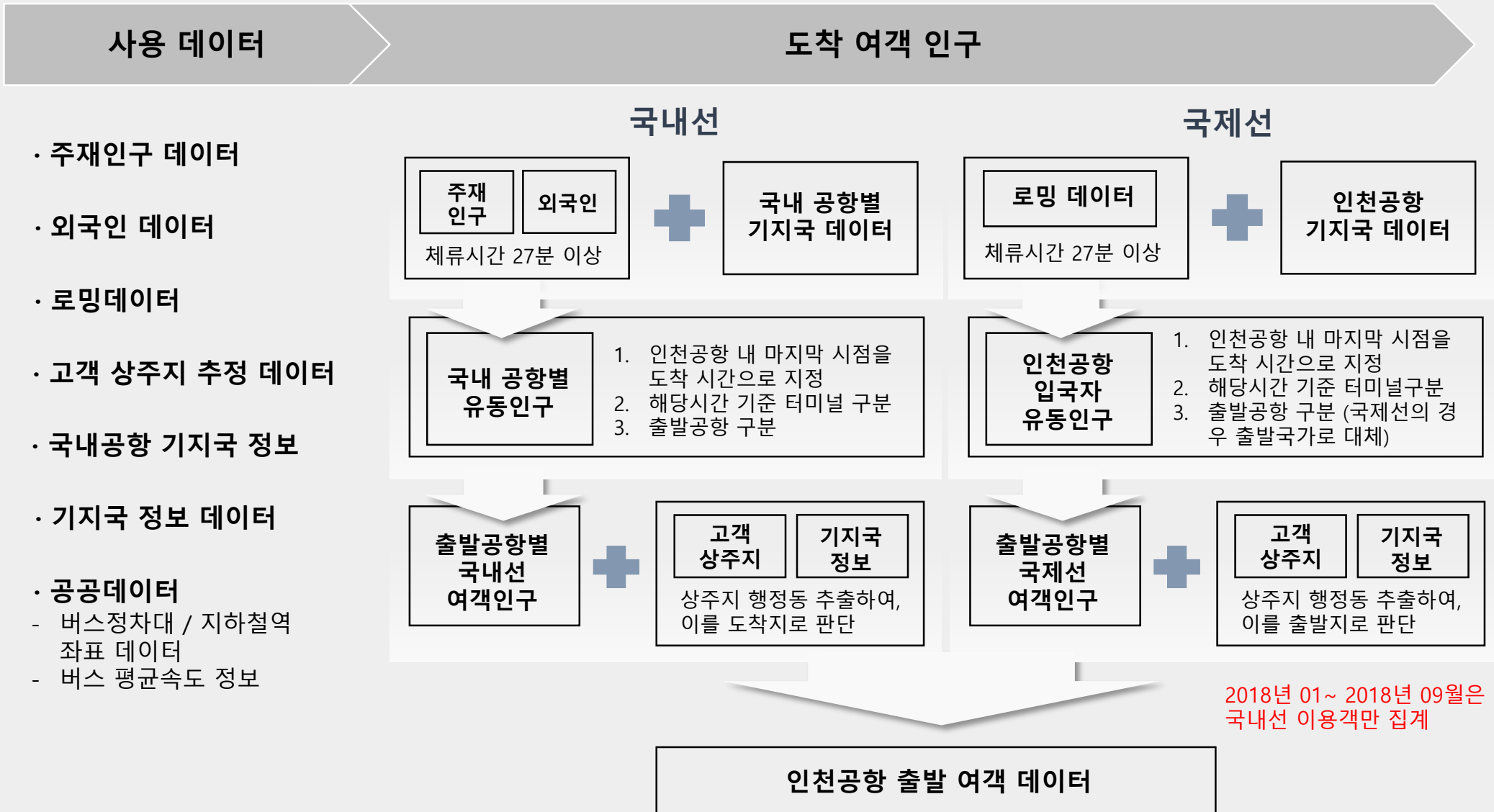
- 중복 제거

- MS 보정 (단기외국인의 경우 국적별 출입국통계 보정)

최종
집계성 통계
데이터
생성

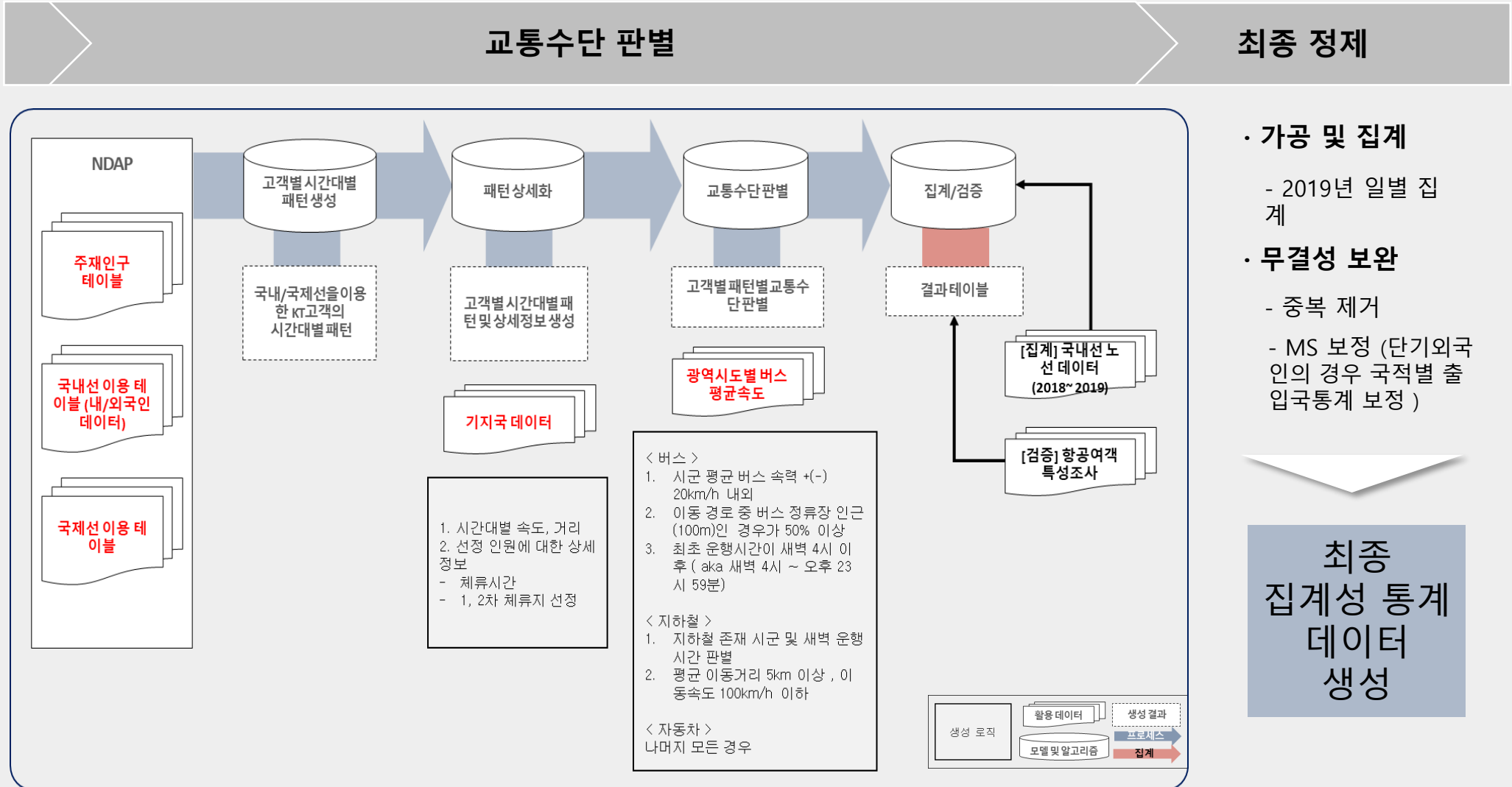
테이블별 분석 로직 - 3. 인천공항 도착 여객 데이터

인천공항에 비행기를 타고 도착하는 인원 에 대한 데이터의 전체 흐름도 및 유동인구 분석 로직



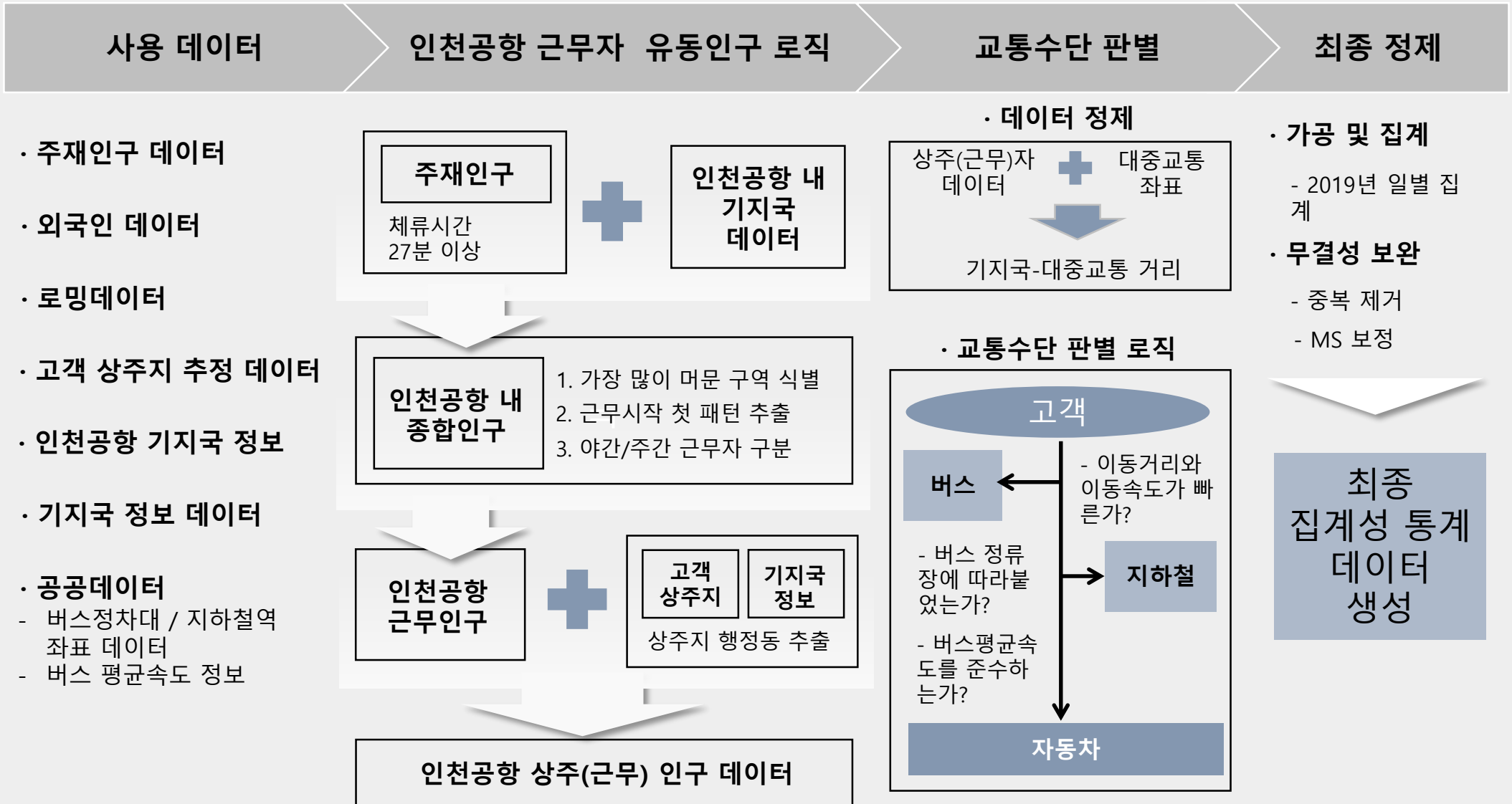
테이블별 분석 로직 - 3. 인천공항 도착 여객 데이터

인천공항에 비행기를 타고 도착하는 인원 에 대한 데이터의 전체 흐름도 및 유동인구 분석 로직



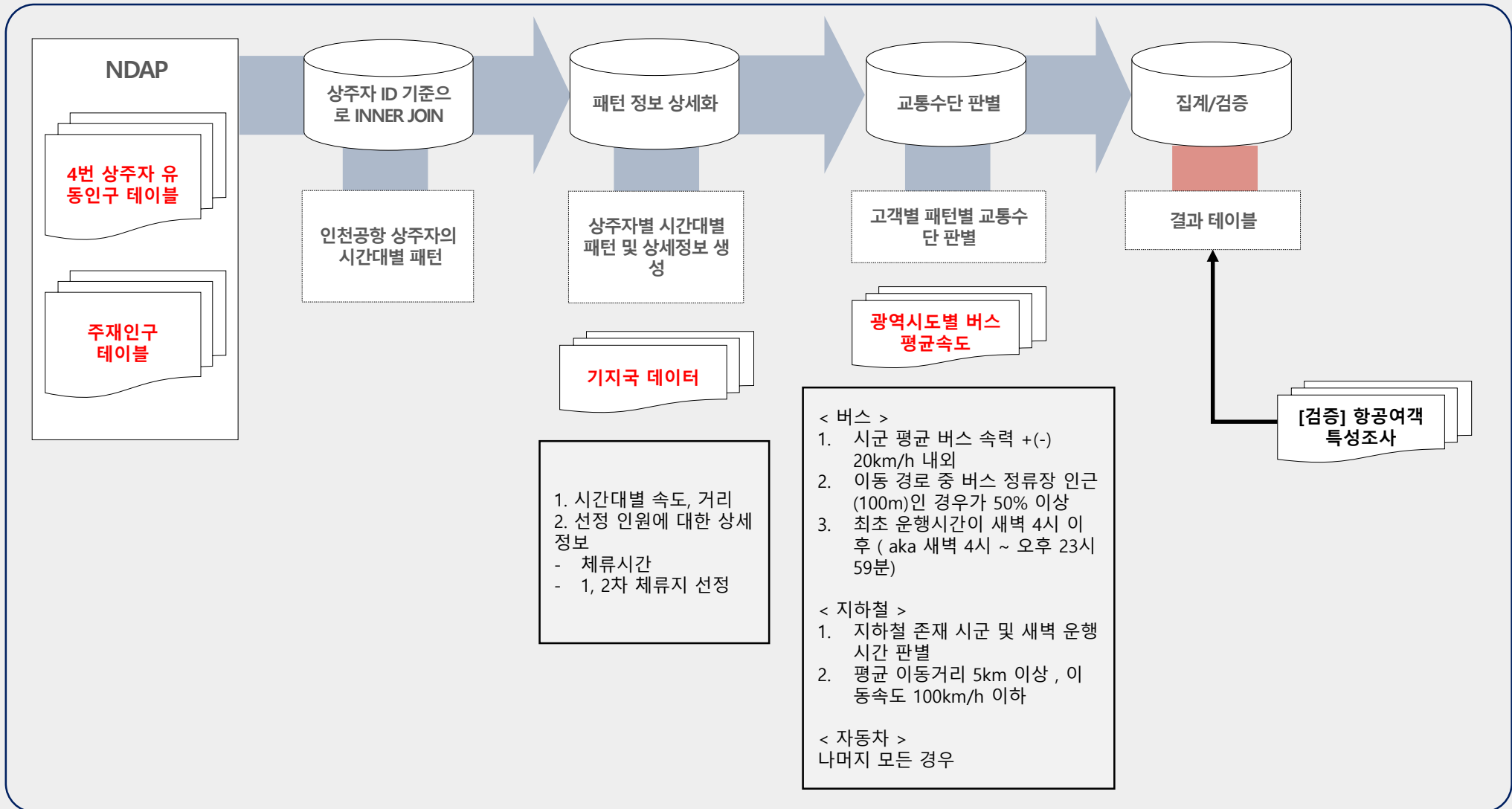
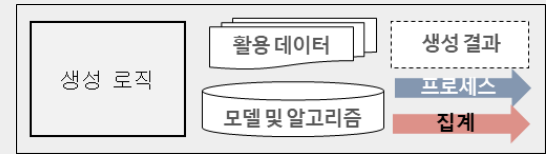
테이블별 분석 로직 - 4. 인천공항 상주자 출퇴근 OD

인천공항 상주자에 대한 출퇴근 OD 데이터의 전체 흐름도 및 유동인구 분석 로직



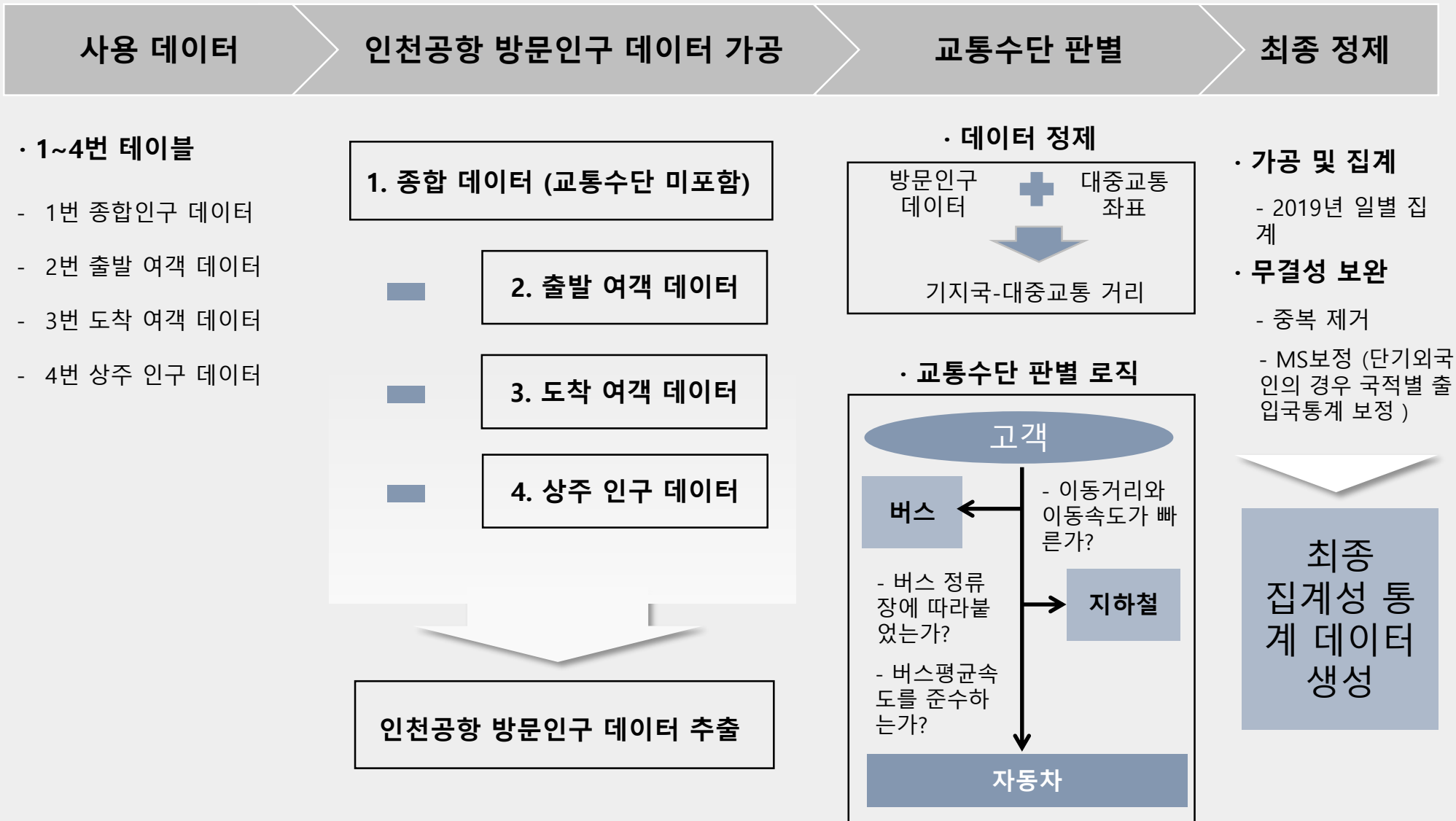
테이블별 분석 로직 - 4. 인천공항 상주자 출퇴근 OD

인천공항 상주자에 대한 출퇴근 OD 데이터의 교통수단 분석 로직



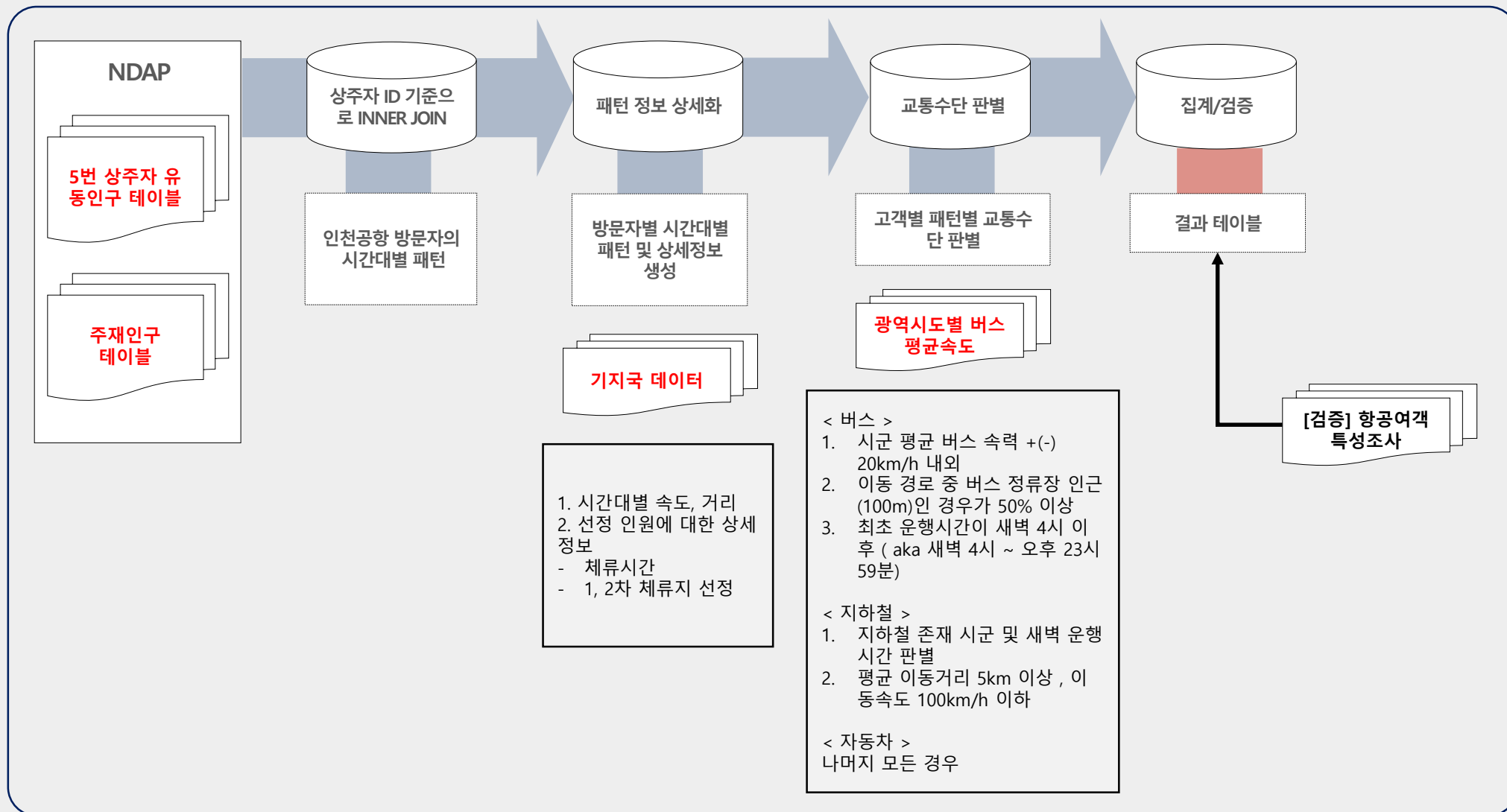
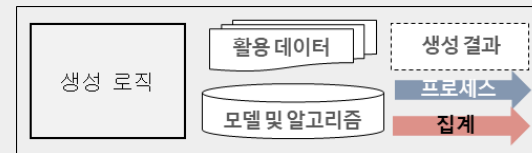
테이블별 분석 로직 - 5. 인천공항 방문인구 데이터

인천공항을 방문하는 인구에 대한 데이터의 전체 흐름도 및 유동인구 분석 로직



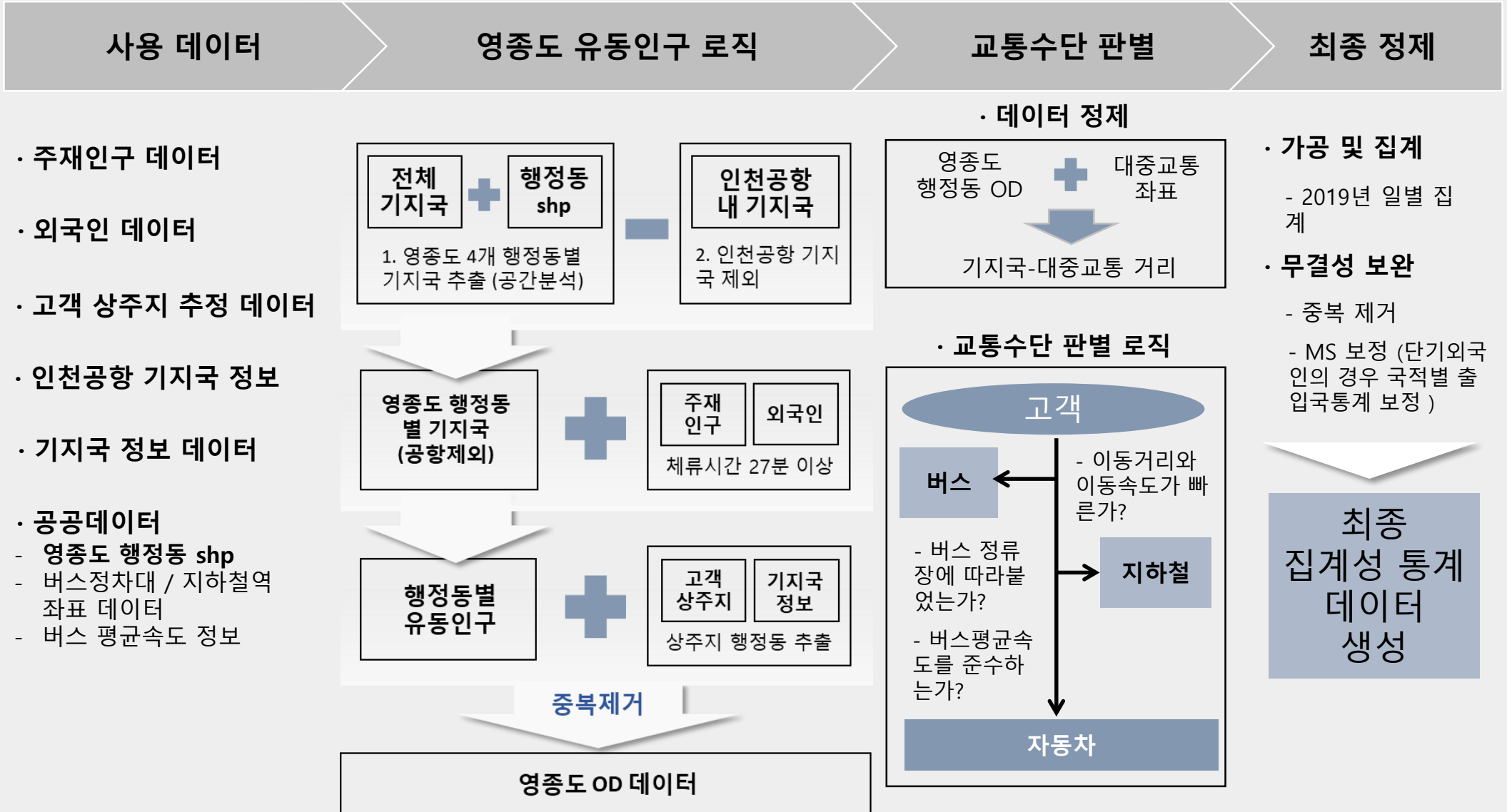
테이블별 분석 로직 - 5. 인천공항 방문인구 데이터

인천공항을 방문하는 인구에 대한 데이터의 교통수단 분석 로직



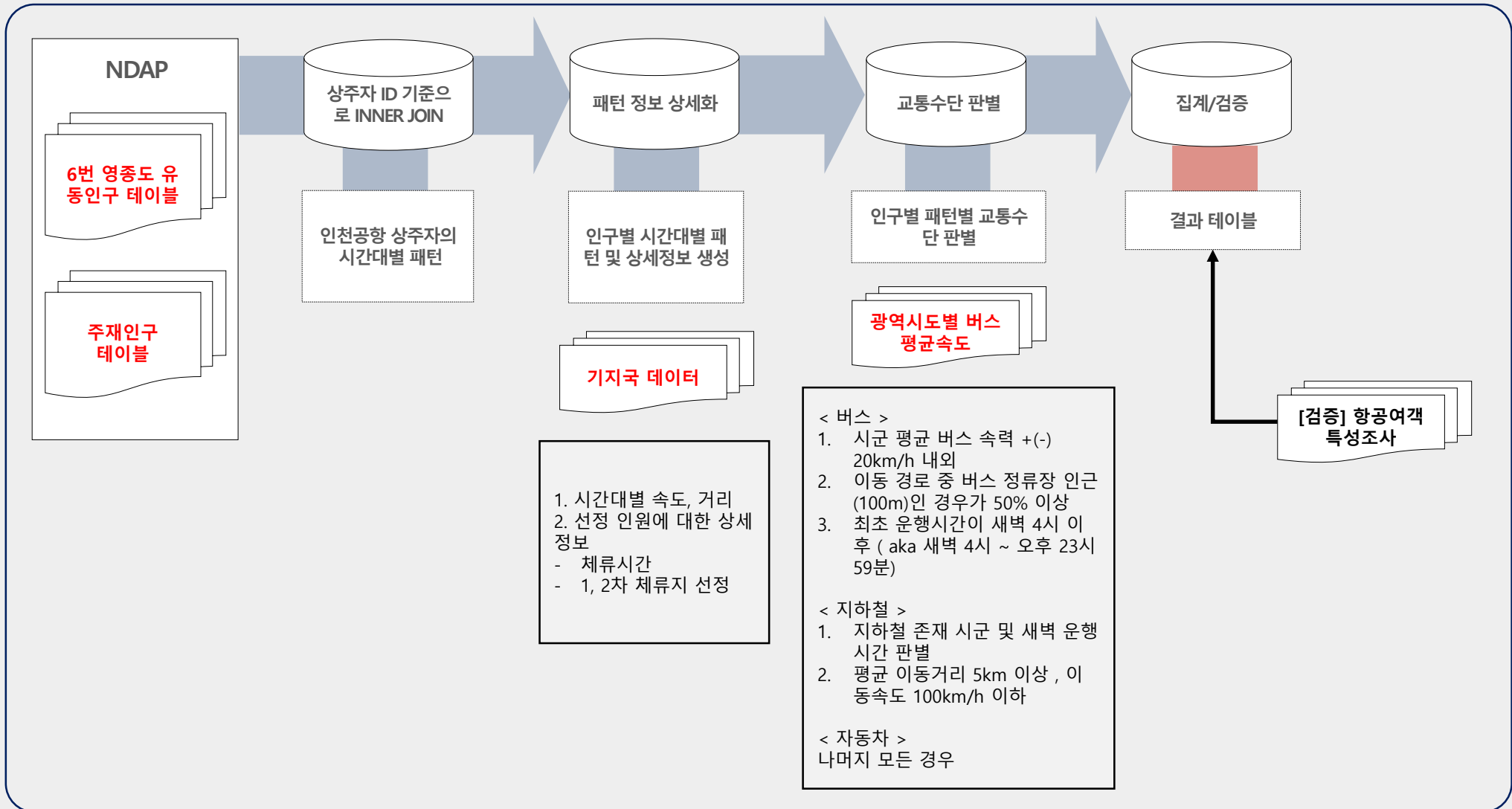
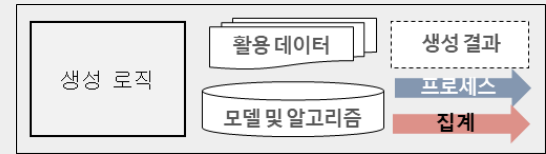
테이블별 분석 로직 - 6. 영종도 OD 데이터

영종도 내 (공항 내부 제외) 유동인구의 OD데이터의 전체 흐름도 및 유동인구 분석 로직



테이블별 분석 로직 - 6. 영종도 OD 데이터

영종도 내 (공항 내부 제외) 유동인구의 OD데이터의 교통수단 분석



테이블별 분석 로직 - 7. 여행일수 데이터

국가별 여행일수별 데이터의 전체 흐름도 및 유동인구 분석 로직

데이터 정의

여행일수 데이터 가공

최종 정제

· 로밍 데이터

· 여행일수 추정

- 해외로밍 서비스 사용자의 여행일수 추정

1. 해외여행 정의

데이터 사용

OR

통화 사용



해외여행
중으로 판단
1day 카운트

2. 개별 패턴 추적 최대기간 필터링
(e.g. 해외여행 기간 6개월이상 제외)

여행일수 데이터 생성

· 정제 및 집계

- 2019년의 국가별 숙박일수 별 집계

· 무결성 보완

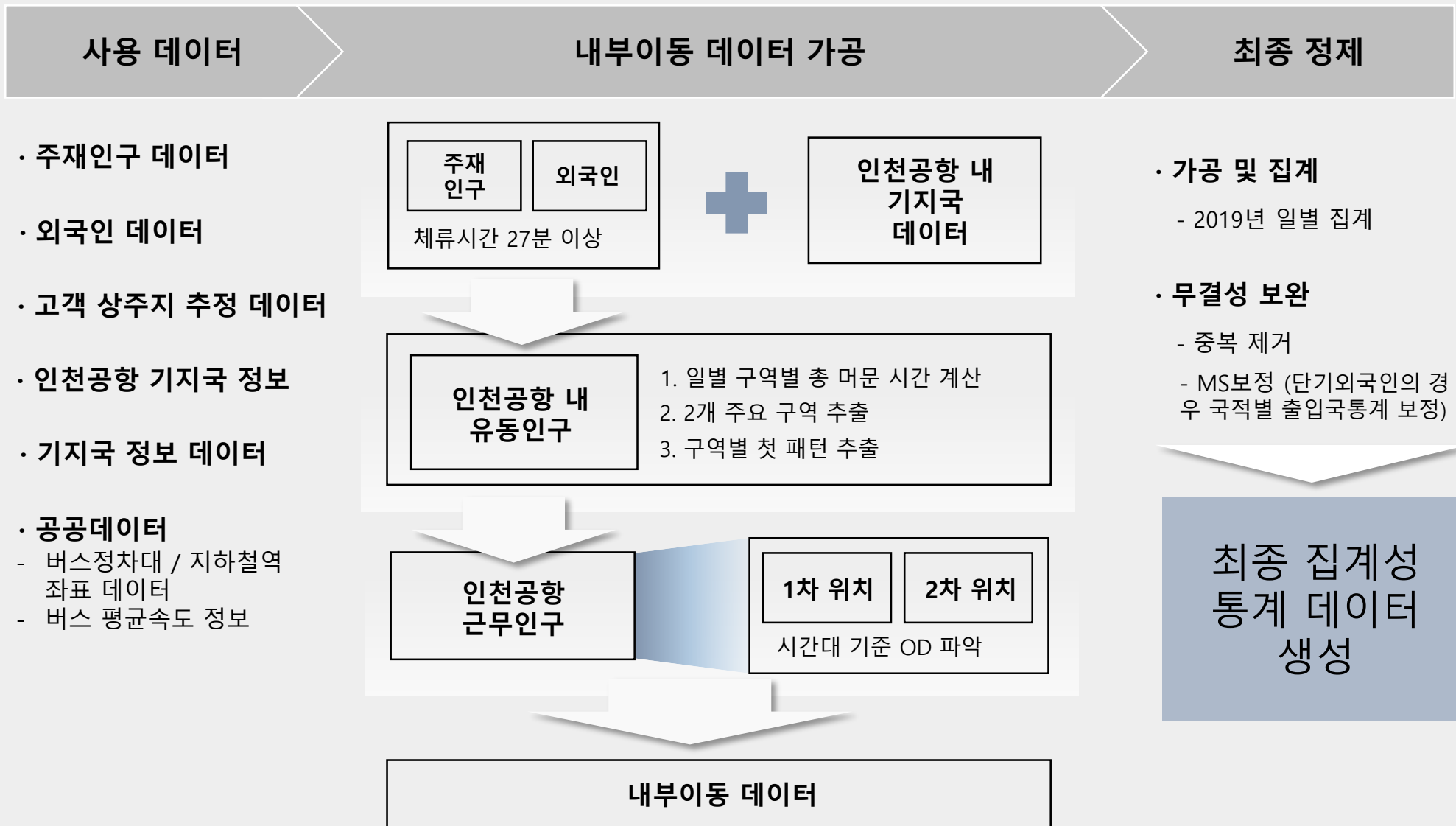
- 중복 제거

- MS보정 (단기외국인의 경우 국적
별 출입국통계 보정)

최종 집계성
통계 데이터
생성

테이블별 분석 로직 - 8. 인천공항 내부이동 데이터

인천공항 내부에서 이동한 정보를 지닌 데이터의 전체 흐름도 및 유동인구 분석 로직



1. 종합 데이터 (2018.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	지역구분 (ICH_SE)	거주지 (ADMI_CD)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	T1	NULL	외국인	NULL	NULL	100.1048
20191202	T2	28110520	내국인	M	A10	7.15
20191203	TF	11305545	외국인	F	A30	46.47
..

2. 출발 여객 데이터 (2018.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	시간대 (TIME_CD)	터미널구분 (ICH_SE)	도착공항 (ARRIVE_CD)	출발지 (ADMI_CD)	교통수단 (TRNS_PRT)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	01	T1	미국	NULL	자동차	외국인	NULL	NULL	100.1048
20191202	10	T2	대구공항	28110520	지하철	내국인	M	A10	7.15
20191203	11	T2	김해공항	11305545	버스	외국인	F	A20	46.47
..

3. 도착 여객 데이터 (2018.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	시간대 (TIME_CD)	터미널구분 (ICH_SE)	출발공항 (DEPARTURE_CD)	도착지 (ADMI_CD)	교통수단 (TRNS_PRT)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	01	T1	미국	NULL	자동차	외국인	NULL	NULL	100.10
20191202	10	T2	대구공항	28110520	지하철	내국인	M	A10	7.15
20191203	11	T2	김해공항	11305545	버스	외국인	F	A20	46.47
..

4. 상주자 출퇴근 OD 데이터 (2019.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	시간대 (ST_TIMEZN_CD)	근무지 (ICH_SE)	거주지 (ADMI_CD)	교통수단 (TRNS_PRT)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	02	T1	NULL	자동차	외국인	NULL	NULL	100.1048
20191202	15	IIAC1	28110520	지하철	내국인	M	A10	7.15
20191203	21	TF	11305545	버스	외국인	F	A30	46.47
..

→ 낮에 근무시, 야간상주지
밤에 근무시 주간상주지 추정

5. 방문인구(환송객) 데이터 (2019.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	시간대 (ST_TIMEZN_CD)	지역구분 (ICH_SE)	거주지 (ADMI_CD)	교통수단 (TRNS_PRT)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	02	T1	NULL	자동차	외국인	NULL	NULL	100.1048
20191202	15	IIAC1	28110520	지하철	내국인	M	A10	7.15
20191203 ..	21	TF ..	11305545 ..	버스	외국인 ..	F ..	A30 ..	46.47 ..

6. 영종도 OD 데이터 (2019.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	시간대 (ST_TIMEZN_CD)	지역구분 (YJD_NM)	거주지 (ADMI_CD)	교통수단 (TRNS_PRT)	내/외국인구분 (IN_FORN_DIV_NM)	성별 (SEX_TYPE_ITG_CD)	연령대 (AGE_CD)	인구수 (CNT)
20191201	02	용유동	NULL	자동차	외국인	NULL	NULL	100.1048
20191202	15	운서동	28110520	지하철	내국인	M	A10	7.15
20191203 ..	21	영종동 ..	11305545 ..	버스	외국인 ..	F ..	A30 ..	46.47 ..

7. 여행일수 데이터 (2019)

정보가 없는 경우, NULL로 반환

국가코드 (CNTRY_CD)	여행일수 (TRVL_DAY)	인구수 (CNT)
202	1	100
202	2	7
202	3	46
...

8. 인천공항 내부이동 데이터 (2019.01.01 ~ 2019.12.31)

정보가 없는 경우, NULL로 반환

일자 (ETL_YMD)	이전위치 (ICH_SE_1)	현재위치 (ICH_SE_2)	내/외국인구분 (IN_FORN_DIV_N M)	성별 (SEX_TYPE_ITG_C D)	연령대 (AGE_CD)	인구수 (CNT)
20191201	T1	IOC	외국인	NULL	NULL	100
20191202	IIAC1	IIAC2	내국인	M	A10	7
20191203	TF	ICLC	외국인	F	A30	46
..

각 집계 테이블 생산 후, 일별/월별 총 유동인구 수를 도출하여 비교 검증

검증 방법

- 일별 유동인구를 도출하여, 이상치 존재여부 및 수치의 일관성 검토
- 통계자료와 비교 검토
 - 여객 데이터와 인천공항 운행 데이터의 일별 상관관계 분석
 - 구역별 유동인구 수 검토 (e.g. 1번 종합데이터의 경우, T1 : T2 = 7 : 3 의 유동인구 비율)
 - 종사자 수 현황 자료(2019년 12월)와 4번 상주자 테이블의 월별 수치 비교
- 상주지 항목을 시군구 단위로 집계 하여, 거주자 수와 비교
- 종합인구 - (출발여객 + 도착여객 + 상주인구) = 방문인구

감사합니다