

Practice 02

[사용할 데이터 파일]

“월별수출입통계.csv” 파일은 대한민국의 각 국가별 수출입 실적을 2000년도부터 2021년 7월까지 월별로 정리한 데이터이다. 수출/수입 금액 및 무역수지의 단위는 1,000 USD이다.

“ISO국가코드.csv”는 국가 코드를 국가 이름과 함께 기록해놓은 테이블로서 merge(join)시에 사용할 수 있습니다.

“continent_table.csv”는 각 대륙별 국가 목록을 포함한 데이터 테이블입니다.

question 1

수출입 통계 중 수치형 변수들은 쉼표, 마이너스 기호를 포함하고 있어 data loading 시 character type이나 factor로 인식이 된다. 이를 적절하게 전처리하여 수치형 값으로 변환하여야. (방법 설명 필요)

question 2

월별수출입통계 데이터는 월별로 각 국가에 대한 대한민국의 수출/수입 실적을 정리한 자료이다. 월별로 전체 국가에 대해 합산하면, 대한민국 전체 수출량 및 수입량을 월별로 계산할 수 있다. 월별 대한민국의 전체 수출량 및 수입량을 계산하여보고, 월별 수출/수입의 변화가 어떻게 나타나는지 확인해보자.

(예를 들어 여름휴가철에는 수입이 늘어나고 겨울에는 수출이 증가하는 추세가 있다 등)

이러한 추세는 연도별로도 달라지는가? 주어진 데이터를 분석해보고 insight를 얻어보자.

question 3

2020년 초부터 코로나 바이러스 팬데믹이 진행되면서 대한민국 수출입에 영향을 주었을 거라고 생각할 수 있다. 어떠한 영향을 주었을 지 추측해보고 데이터를 통해서 확인해보자.

question 4

국가 코드를 사용하여 월별수출입통계 데이터에 대륙 정보(continent)를 추가하자. 국가명의 표기 방식의 차이로 mismatch가 발생하는 국가가 있는지 확인하고 이를 수정하여 최대한 누락되는 국가가 없도록 integration을 수행하자. (과정 설명)

question 5

우리나라와 가장 많은 무역을 하는 국가는 어디인지 어느 대륙에 주로 분포하는지 확인하여보자.

수출을 많이 하는 국가와 수입을 많이 하는 국가로 나누어서 확인하여보자.

대륙별 수출입 통계를 계산하여 비교하여보자.

question 6

2000년대 들어서 우리나라와 무역규모 (수출 또는 수입)이 가장 크게 상승한 국가는 어디인가?

반대로 무역 규모가 감소하거나 거의 증가하지 않은 국가도 있는지 확인하여보자.

question 7

2000년대 이후 집권한 정부마다 다른 외교 정책을 펼쳤을 것이라 생각이 되는데 이는 무역에도 영향을 주었을 수 있다.

각 정부마다 특이하게 무역량에 변화가 나타난 국가가 있는지 확인하여 보자. (급격하게 증가하거나 증가폭이 줄고 감소하거나 등)

각 정부의 집권 기간은 인터넷 검색으로 확인.

(전체 국가에 대해서 하지 않아도 되고, 미중일러 등 주요국가에 대해서만 수행하면 됨)

데이터 출처:

관세청 수출입무역통계: <https://unipass.customs.go.kr/>