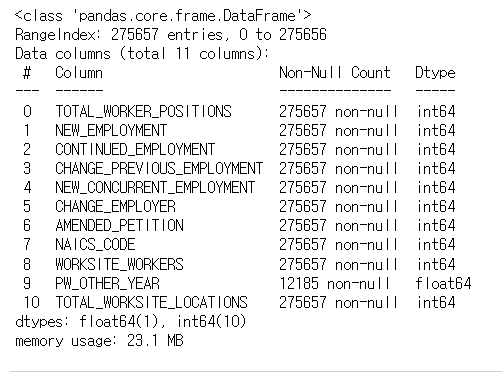
Week 3 Report

이번에는 저번과제에서 불러온 데이터(LCA, PW, PERM)들의 유의미한 분석을 위해 각 데이터들에 대한 EDA를 수행할 것이다. 수행방법은 각 데이터의 numerical data type을 추출하여 각 column간의 Pearson correlation coefficient를 분석하여 이를 기반으로 column간의 유의미한 상관관계를 가지는 column을 추출하여 plot해 볼 것이다.

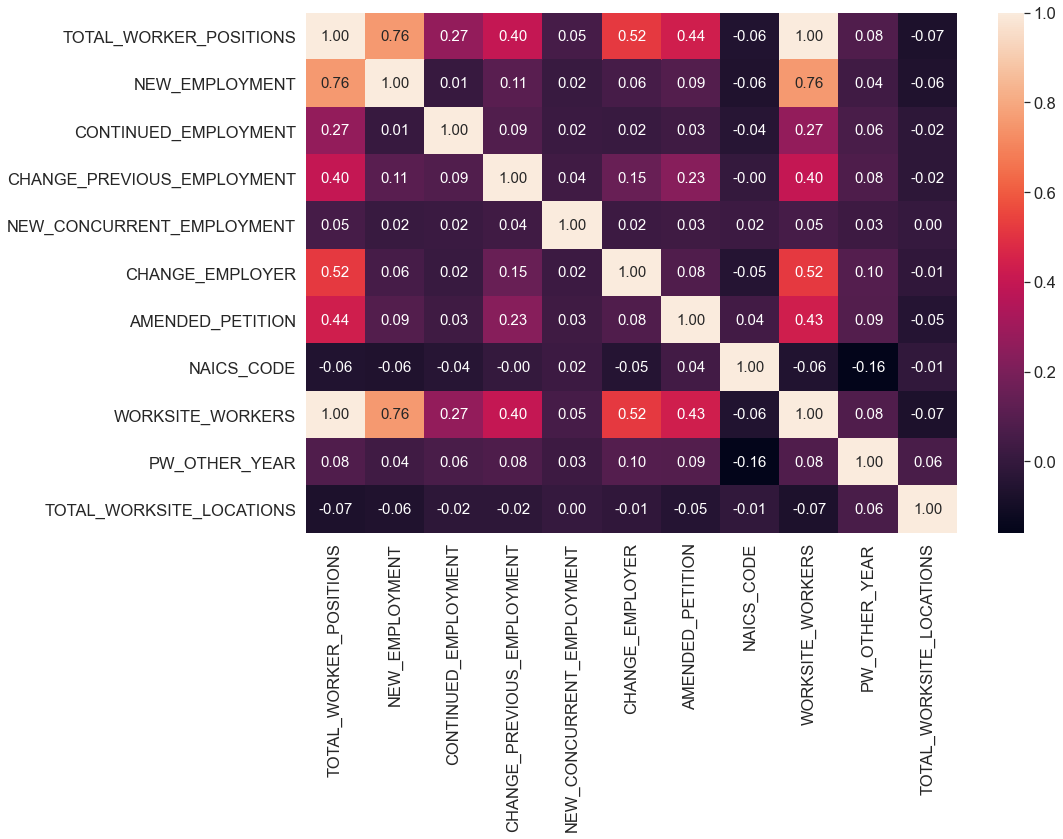
**LCA** **EDA**



<LCA data에서 선별한 numerical data>

LCA의 numerical data type를 추출해 본 결과 아래와 같이 총 10개의 column을 추출할 수 있었다. (phone information에 대한 numerical data 또한 존재했지만 유의미한 결과를 제공하지 않을 것으로 판단하여 제외시켰다.)

이를 기반으로 Pearson correlation coefficient를 분석한 결과 다음과 같았다.



<pearson correlation coefficient of LCA numerical data >

여기서 주목할 점으로 TOTAL\_WORKER\_POSITIONS, NEW\_EMPLOYMENT, WORKERSITE\_WORKERS가서로 강한 양의 상관관계를 보였다. 추출한 columns을 description을 확인한 결과 모두 고용하는 worker의 number와 관련있는 내용이였다.

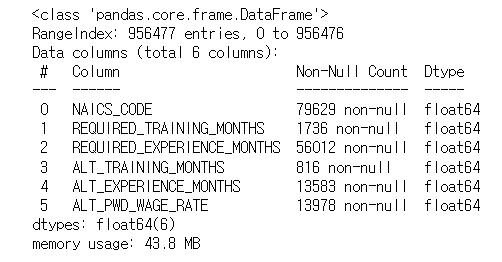
참고

TOTAL\_WORKER\_POSITIONS : Total number of foreign workers requested by the Employer.

NEW\_EMPLOYMENT : Indicates requested worker(s) will begin employment for new employer, as defined by USCIS.

WORKERSITE\_WORKERS : Number of workers placed at the First Worksite location

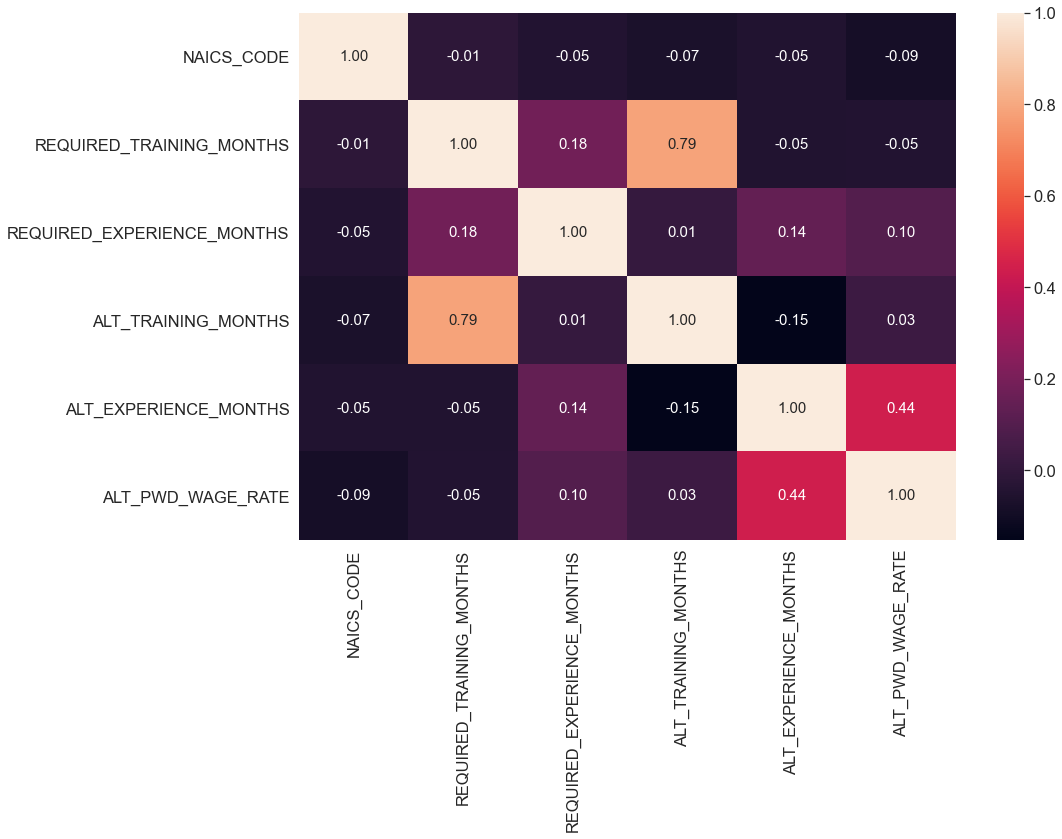
**PW**(**Prevailing Wage**) **EDA**



<PW data에서 선별한 numerical data>

PW의 numerical data type를 추출해 본 결과 위와 같이 총 6개의 column을 추출할 수 있었다. (이중 column만 존재하고 data만 존재하는 PRIMARY\_WORKSITE\_COUNTRY는 표에서 제외했다.)

이를 기반으로 Pearson correlation coefficient를 분석한 결과 다음과 같았다.



<pearson correlation coefficient of PW numerical data >

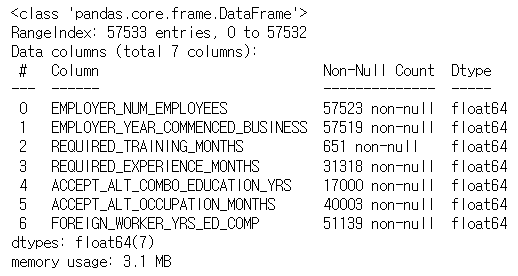
여기서 주목할 점으로 ALT\_TRAINING\_MONTHS와 REQUIERD\_TRAINING\_MONTHS가 강한 양의 상관관계를 보였다. 두 columns 모두 수습기간을 나타내는 정보로 양의 상관관계를 가지는 것을 이해할 수 있다.

참고

ALT\_TRAINING\_MONTHS : If alternative training is accepted for the job opportunity, the number of months training required.

REQUIERD\_TRAINING\_MONTHS : The number of months of training required for the job opportunity, if applicable.

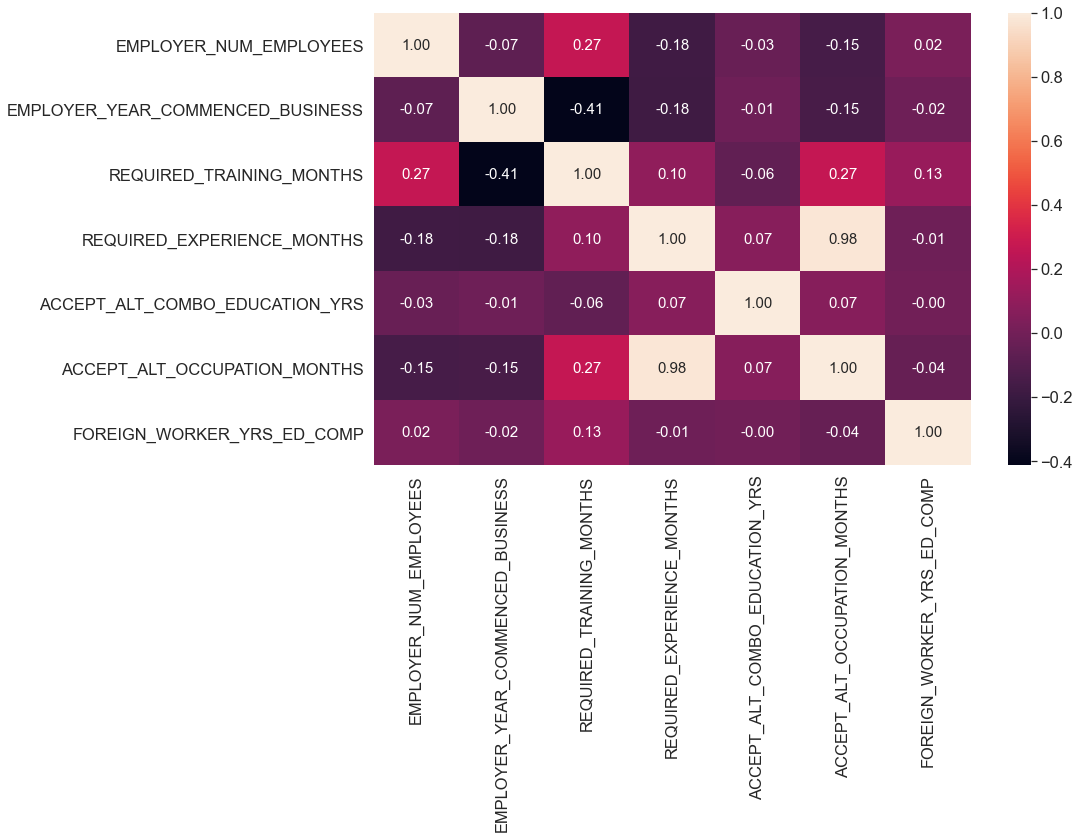
**PERM EDA**



<PERM data에서 선별한 numerical data>

PERM의 numerical data type를 추출해 본 결과 위와 같이 총 7개의 column을 추출할 수 있었다. (이중 phone number에 관한 내용은 LCA와 같은 이유로 표에서 제외했다.)

이를 기반으로 Pearson correlation coefficient를 분석한 결과 다음과 같았다.



<pearson correlation coefficient of PERM numerical data >

여기서 주목할 점으로 ACCEPT\_ALT\_OCCUPATION\_MONTHS와 REQUIERD\_EXPERIENCE\_MONTHS가 매우 강한 양의 상관관계를 보였다. 두 column 모두 experiece에 대한 정보로 양의 상관관계를 가지는 것을 이해할 수 있었다.

참고

ACCEPT\_ALT\_OCCUPATION\_MONTHS : The number of months of experience in the alternate occupation

REQUIERD\_EXPERIENCE\_MONTHS : If experience in the job offered is required, the number of months required

데이터 전처리를 하지 않았기 때문에 전반적으로 유의미한 정보를 추출해지는 못하였다. 특히 가장 주요한 정보인 employ, worker, wage와 같은 정보의 상관관계를 분석하지 못하여 아쉬움이 남는다. 이후 각 column에 해당하는 data를 numerical type으로 casting하여 정보를 분석해 본다면 더욱 유의미한 결과를 도출할 수 있을 것이다.