1. Pima Indian Diabetes 데이터 세트 diabetes.csv 파일을 불러오고 데이터의 평균, 표준편차, 4분위수를 체크하세요. 데이터프레임의 첫 다섯 행도 확인하세요.

2. 'Glucose', 'BloodPressure','SkinThickness','Insulin','BMI' 변수에서 0인 값들을 평균 값으로 대체하세요.

3. Pregnancies 별로 당뇨병에 걸릴 확률을 나타낸 그래프를 그리세요.

4. 데이터셋을 피처 데이터셋, 결과 데이터셋으로 나눠주세요.

5. MinMaxScaler 을 이용해 피처 데이트셋에 스케일링을 적용하고, train/test 데이터셋을 train\_test\_split을 이용하여7:3으로 나눠주세요.

6. Logistic Regression을 이용하여 학습, 예측 한 뒤 accuracy\_score 을 구하세요.

7. 해당 모델로 교차검증을 하고 평균 정확도를 구하세요.

8. 임곗값(threshold)을 0.52로 설정한 binarizer 을 생성하세요. 그리고 위에서 구한 logistic regression 모델의predict\_proba() 예측 확률 array에서 positive(1)에 해당하는 컬럼을 binarize 하세요.

9. 8번에서 구한 예측 결과를 토대로 confusion matrix, accuracy, precision, recall, f1, roc 값을 구하세요.