

1. 재료 전처리

```
1 df['main'][0:5]

0      ['오리고기 3조각, 짜장라면 1개, 물 300ml']
1      ['새송이버섯', '버터', '어린잎채소', '간장', '맛술', '올리고당']
2      ['새송이버섯', '청양고추', '대파', '버터']
3      ['콩나물', '비엔나소시지', '대파', '양파', '물']
4      ['차돌박이', '깻잎', '팬이버섯', '숙주나물']
Name: main, dtype: object

1 df['main_quantity'][0:5]

0      ['']
1      ['2개', '30g', '', '1T', '1T', '1/2T']
2      ['3동', '1개', '1/2대', '1T']
3      ['1봉', '1봉', '1대', '1/2개', '1컵(종이컵)']
4      ['10장', '5장', '1묵음', '1주먹']
Name: main_quantity, dtype: object
```

레시피에 필요한 재료와 필요한 양이 위와 같이 구분되어 있다. 경우에 따라 구분되어 있지 않은 행도 존재한다. 3번째 행의 경우 재료는 '새송이버섯, 청양고추, 대파, 버터'가 필요하고 용량은 '3동, 1개, 1/2대, 1T'가 필요하다. 이를 db에 넣고 활용하기 좋게 하기 위해 '새송이버섯 3동, 청양고추 1개, 대파 1/2대, 버터 1T'와 같은 형태로 변환해준다. 'main, main_quantity, minor, minor_quantity, source, source_quantity' column을 활용한다.

```
for i in range(0,95709):
    if df['main_quantity'][i]!="['']":
        a=df['main'][i].replace("'",'').replace("
",'').replace(")",'').split(',')
        b=df['main_quantity'][i].replace("'",'').replace("
",'').replace(")",'').split(',')
        df['main_db'][i]=list(map(lambda x, y:x+y, a,b))
    else:
        df['main_db'][i]=df['main'][i].replace("'",'').replace("
",'').replace(")",'').split(',')
```

```
1 df['main_db'][0:5]

0      ['오리고기 3조각, 짜장라면 1개, 물 300ml']
1      ['새송이버섯2개, 버터 30g, 어린잎채소 , 간장 1T, 맛술 1T, 올...']
2      ['새송이버섯3동, 청양고추 1개, 대파 1/2대, 버터 1T']
3      ['콩나물1봉, 비엔나소시지 1봉, 대파 1대, 양파 1/2개, 물 1컵(종이컵)']
4      ['차돌박이10장, 깻잎 5장, 팬이버섯 1묵음, 숙주나물 1주먹']
Name: main_db, dtype: object
```

조리 순서의 경우 다음과 같이 각 순서마다 다른 column에 들어있다.

```
1 df['step0'][0]

'1인분이라 3조각씩 따로 소분해서 올려놓았던 오리고기를 해동시키고나서 물300ml와 함께 밥솥에 넣어서 재가열 9분 해주세요'

1 df['step1'][0]

'고기를 완전히 익히기 위해 재가열 9분 끝난 후 다시 재가열 9분 해주세요'
```

조리 순서에 따른 이미지의 경우도 이와 같이 되어있다. db에 넣고 활용하기 좋게 한 column으로 만든다.

```
for i in range(0,95709):
    step=[]
    for j in range(16,65,2):
        try:
            if math.isnan(df.loc[i][j])==True:
                break
        except:
            step.append(df.loc[i][j])
    df['step'][i]=step
```

```
1 df['step'][0]
```

['1인분이라 3조각씩 따로 소분해서 얼려놓았던 오리고기를 해동시키고나서 물300ml와 함께 밥술에 넣어서 재가열 9분 해주세요',
'고기를 완전히 익히기 위해 재가열 9분 끝난 후 다시 재가열 9분 해주세요',
'재가열이 다 된 후 그대로 짜잔!만 면과 건더기 스프를 넣고 물에 고루 잠기게끔 면을 펼쳐놓고 재가열 9분 해주세요 물 300ml 했을때 면이
에 23만 잠기게 되는데요 그 상태에서 재가열 했을때 면이 꼬들 꼬들하고 딱 적당하게 익는데다가 버릴물도 적어서 좋아요',
'재가열이 다 된 후 남은 물을 따라 버리고 짜잔 스프 넣고 잘 비벼주세요',
'색감을 좋게 하려면 파를 얹게 다져서 위에 올려 드시면 보기에 좋아요']

```
1 df['img'][0]
```

```
[ 'https://recipe1.ezmember.co.kr/cache/recipe/2019/12/30/15e4ab23f52d826aba6607ebb965165e1.jpg',  
  'https://recipe1.ezmember.co.kr/cache/recipe/2019/12/30/3eae5bf3d346248ac074fdaddb6761dc91.jpg',  
  'https://recipe1.ezmember.co.kr/cache/recipe/2019/12/30/c1c2c6155f637f747f373acc6617861d251.jpg',  
  'https://recipe1.ezmember.co.kr/cache/recipe/2019/12/30/89917264100cc3f4855e6f6faa478d821.jpg',  
  'https://recipe1.ezmember.co.kr/cache/recipe/2019/12/30/c2a21733e05784c60daeb910aa94f25f1.jpg']
```

그리고 title, description, step에 쓰인 특수문자들을 제거한다.

```
for i in range(0,95709):
    steps=[]
    for step in df['step'][i]:
        text = re.sub('[-=+,#/\?:^$.@*\'"\*\~&%\`!』\\\'|\(\)\[\]\<|\>\'\'... ]', '',
step)
        steps.append(text)
    df['step'][i]=steps
    if i%100==0:
        print(i)
```