해보고자 하는 것

- 인사 데이터를 통해, 이 사람이 회사를 나갈것인지 나가지 않을것인지 분석해보기.

- 데이터의 left항목이 해당 데이터의 label

- sales데이터는 one hot encoding이 필요

- salary는 문자를 숫자로 치환이 가능하다고 보았음.

satisfaction_l	last_evaluatic	number_proj	average_mor	time_spend_c	Work_accide	left	promotion_la	sales	salary
0.38	0.53	2	157	3	0	1	0	sales	low
0.8	0.86	5	262	6	0	1	0	sales	medium
0.11	0.88	7	272	4	0	1	0	sales	medium
0.72	0.87	5	223	5	0	1	0	sales	low
0.37	0.52	2	159	3	0	1	0	sales	low
0.41	0.5	2	153	3	0	1	0	sales	low
0.1	0.77	6	247	4	0	1	0	sales	low
0.92	0.85	5	259	5	0	1	0	sales	low
0.89	1	5	224	5	0	1	0	sales	low
0.42	0.53	2	142	3	0	1	0	sales	low
0.45	0.54	2	135	3	0	1	0	sales	low
0.11	0.81	6	305	4	0	1	0	sales	low
0.84	0.92	4	234	5	0	1	0	sales	low
0.41	0.55	2	148	3	0	1	0	sales	low
0.36	0.56	2	137	3	0	1	0	sales	low
0.38	0.54	2	143	3	0	1	0	sales	low
0.45	0.47	2	160	3	0	1	0	sales	low
0.78	0.99	4	255	6	0	1	0	sales	low
0.45	0.51	2	160	3	1	1	1	sales	low
0.76	0.89	5	262	5	0	1	0	sales	low

```
import csv
import pandas as pd
import numpy as np


raw_data = pd.read_csv('HR_comma_sep.csv', header = 0)

hr_list =np.zeros(14999, dtype=np.int)
sales_list =np.zeros(14999, dtype=np.int)
it_list = np.zeros(14999, dtype=np.int)
management_list =np.zeros(14999, dtype=np.int)
marketing_list = np.zeros(14999, dtype=np.int)
product_mng_list = np.zeros(14999, dtype=np.int)
RandD_list = np.zeros(14999, dtype=np.int)
accounting_list = np.zeros(14999, dtype=np.int)
technical_list = np.zeros(14999, dtype=np.int)
support_list = np.zeros(14999, dtype=np.int)


for idx, item in enumerate(raw_data['sales']):
    if item=='hr':
        hr_list[idx]=1
    elif item=='sales':
        sales_list[idx]=1
    elif item=='it':
        it_list[idx]=1
    elif item=='management':
```

```
raw_data['hr'] = hr_list
raw_data['sale'] = sales_list
raw_data['it'] = it_list
raw_data['management'] =management_list
raw_data['marketing'] = marketing_list
raw_data['product_mng'] = product_mng_list
raw_data['RandD'] = RandD_list
raw_data['accounting'] = accounting_list
raw_data['technical'] = technical_list
raw_data['support'] = support_list

raw_data=raw_data.replace(to_replace='low',value=-1)
raw_data=raw_data.replace(to_replace='high',value=1)
raw_data=raw_data.replace(to_replace='medium',value=0)
# print(raw_data)

df = raw_data['average_montly_hours']

df_norm = (df - df.mean()) / (df.max() - df.min())

raw_data['average_montly_hours'] = df_norm

print(raw_data)

raw_data.to_csv('./prep_data.csv')
```

everage_mon	time_spend_c	Work_accider	left	promotion_la	salary	hr	sales	it	management	marketing	product_mng	RandD	accounting	technical	support
-0.2058427	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.28481151	6	0	1	0	0	0	1	0	0	0	0	0	0	0	0
0.33154348	4	0	1	0	0	0	1	0	0	0	0	0	0	0	0
0.10256352	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.1964969	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2245343	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.21471805	4	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.27079282	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.10724142	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2759362	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.3086464	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.48574609	4	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.15397039	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2478988	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2993006	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.2712533	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.191324	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.25210123	6	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.191324	3	1	1	1	-1	0	1	0	0	0	0	0	0	0	0
0.28481151	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.37826946	4	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2525717	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.48107319	4	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2899548	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.2011598	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.19135357	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
0.17733488	5	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.3086464	3	0	1	0	-1	0	1	0	0	0	0	0	0	0	0
-0.3413567	3	0	1	0	-1	0	0	0	0	0	0	0	1	0	0
-0.3226551	3	0	1	0	-1	0	0	0	0	0	0	0	1	0	0
0.43434422	4	0	1	0	-1	0	0	0	0	0	0	0	1	0	0
-0.3133193	3	0	1	0	-1	1	0	0	0	0	0	0	0	0	0
-0.2619175	3	0	1	0	-1	1	0	0	0	0	0	0	0	0	0
-0.2852319	3	0	1	0	-1	1	0	0	0	0	0	0	0	0	0
0.21004516	6	0	1	0	-1	1	0	0	0	0	0	0	0	0	0
0.25210123	4	0	1	0	-1	0	0	0	0	0	0	0	0	1	0