빅데이터 시험 요약 (loc & iloc 포함 완전판)

0. 환경 설정

Python / VS Code / pip install numpy, pandas → pip install numpy pandas
Colab → https://colab.research.google.com
Anaconda / Jupyter → jupyter notebook

1. Numpy

array() / ndarray → np.array([1,2,3]) → array([1,2,3])
사칙연산 → np.array([1,2,3]) + np.array([4,5,6]) → array([5,7,9])
속성: shape / ndim / size / dtype / itemsize
인덱싱/슬라이싱 → a[0], a[:,1], a[a>3]

2. Pandas

Series → pd.Series([10,20,30], index=['a','b','c'])
DataFrame → pd.DataFrame({'name':['A','B'],'score':[90,80]})
head()/info()/describe() → 기본 정보 확인

3. 데이터 처리

결측치(NaN) → df.isnull(), df.fillna(0), df.dropna()
정렬/필터링 → df.sort_values('score'), df[df['score']>80]
.loc (라벨 기반) → df.loc[0,'name'], df.loc[df['score']>=85,['name','score']]
.iloc (정수 기반) → df.iloc[0,0], df.iloc[0:2,0:2]
groupby() → df.groupby('class')['score'].mean()
concat()/merge() → pd.concat([df1,df2]), pd.merge(df1,df2,on='id')

4. 파일 입출력

CSV → pd.read_csv('data.csv'), df.to_csv('out.csv', index=False)
Excel → pd.read_excel('data.xlsx'), df.to_excel('out.xlsx')

5. 통계 함수

mean(), median(), max(), min(), std(), corr()

6. 데이터 타입 변환

astype() → df['score']=df['score'].astype(float)
to_datetime() → df['date']=pd.to_datetime(df['date'])

7. 실습 포인트

Numpy 연산 → a*np.array([1,2,3])*2

Pandas 조작 → df['avg'] = (df['mid'] + df['final'])/2
NaN 처리 → df.fillna(df.mean())
정렬/그룹화/병합 → sort_values(), groupby(), merge()