빅데이터 실기 정리

|  |
| --- |
| * dir(pandas)   'BooleanDtype',  'Categorical',  'CategoricalDtype',  'CategoricalIndex',  'DataFrame', : 데이터프레임 변환  'DateOffset',  'DatetimeIndex',  'DatetimeTZDtype',  'ExcelFile',  'ExcelWriter',  'Flags',  'Float32Dtype',  'Float64Dtype',  'Float64Index',  'Grouper',  'HDFStore',  'Index',  'IndexSlice',  'Int16Dtype',  'Int32Dtype',  'Int64Dtype',  'Int64Index',  'Int8Dtype',  'Interval',  'IntervalDtype',  'IntervalIndex',  'MultiIndex',  'NA',  'NaT',  'NamedAgg',  'Period',  'PeriodDtype',  'PeriodIndex',  'RangeIndex',  'Series',  'SparseDtype',  'StringDtype',  'Timedelta',  'TimedeltaIndex',  'Timestamp',  'UInt16Dtype',  'UInt32Dtype',  'UInt64Dtype',  'UInt64Index',  'UInt8Dtype',  'api',  'array',  'arrays',  'bdate\_range',  'compat',  'concat',  'core',  'crosstab',  'cut',  'date\_range',  'describe\_option',  'errors',  'eval',  'factorize',  'get\_dummies',  'get\_option',  'infer\_freq',  'interval\_range',  'io',  'isna',  'isnull',  'json\_normalize',  'lreshape',  'melt',  'merge',  'merge\_asof',  'merge\_ordered',  'notna',  'notnull',  'offsets',  'option\_context',  'options',  'pandas',  'period\_range',  'pivot',  'pivot\_table',  'plotting',  'qcut',  'read\_clipboard',  'read\_csv', : csv 읽어오기  'read\_excel',  'read\_feather',  'read\_fwf',  'read\_gbq',  'read\_hdf',  'read\_html',  'read\_json',  'read\_orc',  'read\_parquet',  'read\_pickle',  'read\_sas',  'read\_spss',  'read\_sql',  'read\_sql\_query',  'read\_sql\_table',  'read\_stata',  'read\_table',  'read\_xml',  'reset\_option',  'set\_eng\_float\_format',  'set\_option',  'show\_versions',  'test',  'testing',  'timedelta\_range',  'to\_datetime',  'to\_numeric',  'to\_pickle',  'to\_timedelta',  'tseries',  'unique',  'util',  'value\_counts',  'wide\_to\_long' |
| .info() : 정보  .mean() : 평균  .std() : 표준편차  .rank() : 순위 1부터  .sort\_values() : 정렬 (by = [‘기준 컬럼명’], ascending = False) -> 내림차순정렬  .describe() :정보  .count() :  round(값, 수) : 입력한 수 만큼 소수점을 반올림하여 표기  df[df[‘컬럼명’] == “조건”] : 조건에 맞는 행 추출  # \*\*\*\*\*\*\*\*\*첫번째 행부터 80% 훈련 데이터로 추출\*\*\*\*\*\*\*\*\*  idx80 = np.floor(df.shape[0] \* 0.8).astype('int')  df = df[:idx80]  .quantile(.25) : 1사분위 값 추출 |
|  |
|  |
|  |
|  |
|  |
|  |