

필수 함수

결측	결측치 개수?	<code>df.isna().sum().sum()</code>
	age의 결측치를 age의 평균으로 대체하여라	<code>df.age.fillna(df.age.mean())</code>
	결측이 있는 행은 제외하고 분석하여라	<code>df=df.dropna()</code>
필터	Segmentation 값이 A 또는 D인 데이터만 사용	<code>df=df[df.Segmentation.isin(['A','D'])]</code>
	previous_owner가 1이 아니고 engine_power가 51인 데이터만 사용	<code>df = df[(df.previous_owners!=1)&(df.engine_power==51)]</code>
	한강조망 여부(0,1)에 따른 가격 평균 구하기	<code>df.groupby('waterfront')['price'].mean()</code>
상관계수	상관계수를 구하시오	<code>df.corr()</code>
파생변수	df.Sex의 F를 1로 M을 0으로 변환하여 Sex_cd 변수를 생성하시오.	<code>df['Sex_cd']=df.Sex.replace({'M':0,'F':1})</code>
	Chance_of_Admit 확률이 0.5를 초과하면 합격으로, 이하이면 불합격으로 구분 후 chance_of_Admit 변수를 만드시오	<code>df['Chance_of_Admit']=df.Chance_of_Admit.map(lambda x: '합격' if x>0.5 else '불합격')</code>
	Na_K_gr 컬럼을 만들고 Na_to_k 값이 10이하는 'Lv1', 20이하는 'Lv2', 30이하는 'Lv3', 30초과는 'Lv4'로 변환하시오.	<pre>def get_Na_K_gr(x): if x<=10: return 'Lv1' elif x<=20: return 'Lv2' elif x<=30: return 'Lv3' else:</pre>

		<pre> return 'Lv4' df['Na_K_gr']=df['Na_to_K'].map(get_Na_K_gr) </pre>
머신러닝		
	모델 생성 DecisionTreeClassifier 를 작성하시오	<pre> model_DT = DecisionTreeClassifier() </pre>
	train test 분할 하지 않을 때 모델 학습하는 코드	<pre> model_DT.fit(X, y) </pre>
	train test 분할하는 코드	<pre> X_train, X_test, y_train, y_test =train_test_split(X, y) </pre>
	모델 생성 DecisionTreeClassifier 을 작성하시오	<pre> model=DecisionTreeClassifier() </pre>
	train test를 분할한 데이터로 학습하는 코드는?	<pre> model.fit(X_train, y_train) </pre>
	모델이 예측한 값을 y_pred로 저장하여라	<pre> y_pred = model.predict(X_test) </pre>
	예측한 값의 정확도를 구하여라	<pre> metrics.accuracy_score(y_test,y_pred) </pre>
통계	성별이 세분화(Segmentation)에 영향을 미치는지 독립성 검정을 수행 import scipy.stats as stats	<pre> cross=pd.crosstab(df.Gender, df.Segmentation) chi2_contingency(cross) </pre>
	성별에 따라 forehead_ratio평균에 차이가 있는지! -> 독립 이표본 t검정! from scipy.stats import ttest_ind	<pre> ttest_ind(Male,Female, equal_var= True) </pre>