predict whether a person survived or not (0 or 1) Logistic Regression is like Yes or NO

DataFrame's columns as the x axis and the sum of all missing values in each column in the y axis

Sns **.** barplot ( x **=** titanic\_data **.** columns , y **=** titanic\_data **.** isnull() **.**sum() **.**values)

ptl.show()

DATA . isnull( ) . sum() -> it will show us missing values in the out data.

Did more men or females survive? Recall that ***hue*** parameter seaborn gives us access too.

mean\_age = int(DATA ['Age'] .mean())

DATA ['Age'] = DATA ['Age'].apply(lambda age : mean\_age if pd.isnull(age) else age)

It will clean all missing values in the data. If there are too much missing DATA then we have to drop this column

titanic\_data.drop(labels=['Cabin'], axis=1, inplace=True)

titanic\_data.dropna(inplace=True)

handle categorical variables since machine learning algorithms can only understand numbers.

We first need to split our data into training and testing sets. This can be done using sklearn's ***train\_test\_split(X, y, test\_size)*** function.

ALWAYS FIRST SPLIT the DATA

Since we're now dealing with classification, we'll import sklearn's ***classification\_report*** and ***confusion\_matrix*** to evaluate our model. Both of these take the true values and predictions as parameters.