

```
@tracer(cat_col = ['race', 'occupation', 'education'], numerical_col = ['age', 'hours-per-week'])
def adult_pipeline_easy(f_path = '../pipelines/adult-sample.csv'):

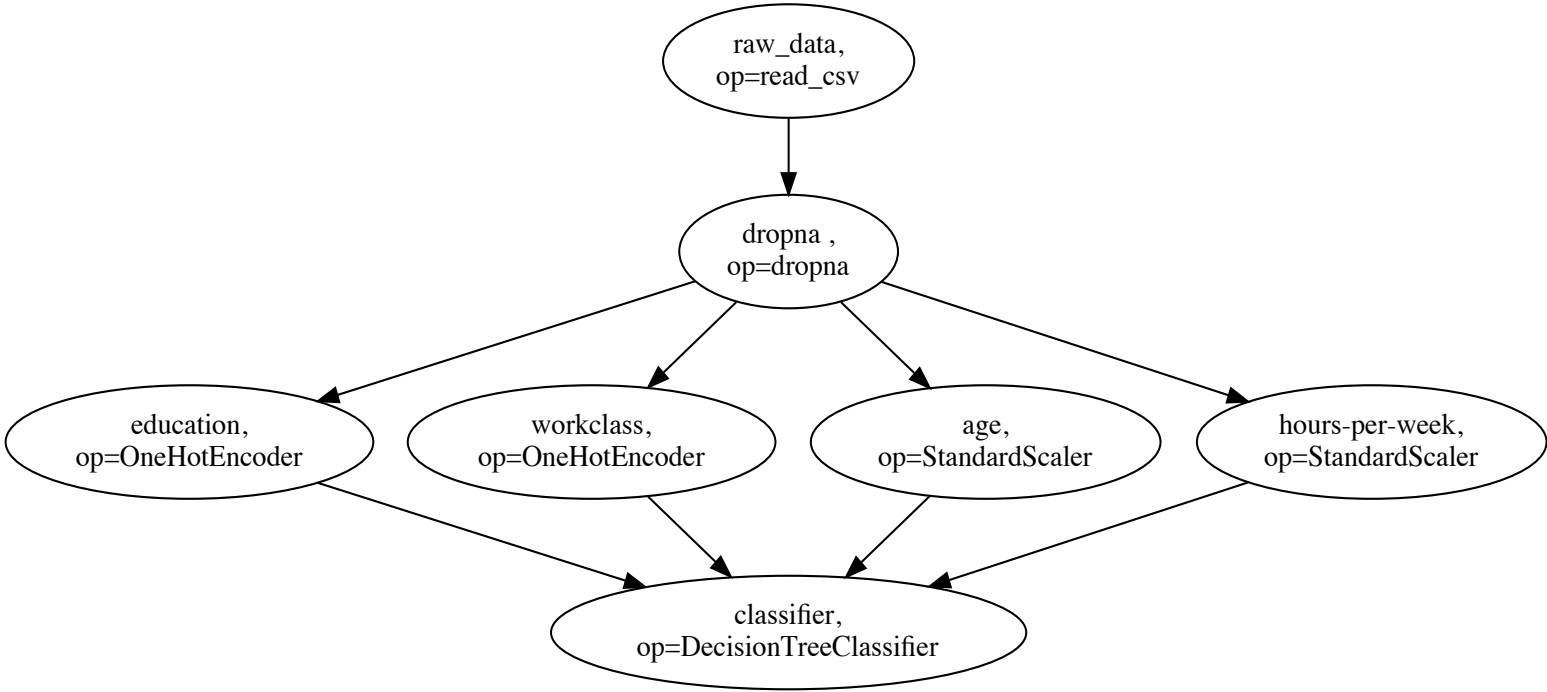
    raw_data = pd.read_csv(f_path, na_values='?')
    data = raw_data.dropna()

    labels = label_binarize(data['income-per-year'], ['>50K', '<=50K'])

    feature_transformation = ColumnTransformer(transformers=[
        ('categorical', OneHotEncoder(handle_unknown='ignore'), ['education', 'workclass']),
        ('numeric', StandardScaler(), ['age', 'hours-per-week'])
    ])

    income_pipeline = Pipeline([
        ('features', feature_transformation),
        ('classifier', DecisionTreeClassifier())
    ])

    return income_pipeline
```



Start Pandas Opeation

Inpected raw_data = pd.read_csv(f_path, na_values='?')

Changes in numerical features!

	count	missing_count	median	mad	range
age	-8.0	0.0	0.0	-0.7413	-19.0
hours-per-week	-8.0	0.0	0.0	-1.4826	0.0

Changes in categorical features!

	missing_count	num_class	class_count	class_percent
race	0.0	0.0	{'White': -6, 'Black': -1, 'Amer-Indian-Eskimo': -1, 'Asian-Pac-Islander': 0, 'Other': 0}	{'White': 0.007, 'Black': -0.0013, 'Amer-Indian-Eskimo': -0.0074, 'Asian-Pac-Islander': 0.0009, 'Other': 0.0009}
occupation	-6.0	0.0	{'Exec-managerial': 0, 'Adm-clerical': 0, 'Craft-repair': -1, 'Sales': 0, 'Prof-specialty': -1, 'Other-service': 0, 'Transport-moving': 0, 'Machine-op-inspct': 0, 'Farming-fishing': 0, 'Protective-serv': 0, 'Handlers-cleaners': 0, 'Tech-support': 0}	{'Exec-managerial': 0.0035, 'Adm-clerical': 0.003, 'Craft-repair': -0.0079, 'Sales': 0.0025, 'Prof-specialty': -0.0083, 'Other-service': 0.0021, 'Transport-moving': 0.0019, 'Machine-op-inspct': 0.0014, 'Farming-fishing': 0.0007, 'Protective-serv': 0.0005, 'Handlers-cleaners': 0.0005, 'Tech-support': 0.0002}
education	0.0	0.0	{'HS-grad': -1, 'Bachelors': 0, 'Some-college': -4, 'Masters': -1, '11th': -2, '7th-8th': 0, 'Assoc-voc': 0, '10th': 0, 'Prof-school': 0, 'Assoc-acdm': 0, '12th': 0, '5th-6th': 0}	{'HS-grad': 0.0152, 'Bachelors': 0.0191, 'Some-college': -0.0235, 'Masters': -0.0057, '11th': -0.0157, '7th-8th': 0.0026, 'Assoc-voc': 0.0026, '10th': 0.0017, 'Prof-school': 0.0009, 'Assoc-acdm': 0.0009, '12th': 0.0009, '5th-6th': 0.0009}

Inpected data = raw_data.dropna()

Start Sklearn Pipeline

Operations OneHotEncoder on education

Changes in categorical features!

	education
missing_count	0
num_class	-10
class_count	{0.0: 90, 1.0: 2}
class_percent	{0.0: 0.9783, 1.0: 0.0217}

Operations StandardScaler on age

Changes in numerical features!

	age
count	0.0000
missing_count	0.0000
median	-36.1059
mad	-12.8706
range	-48.4315

Operations StandardScaler on hours-per-week

Changes in numerical features!

	hours-per-week
count	0.0000
missing_count	0.0000
median	-40.0814
mad	0.0000
range	-63.7616
