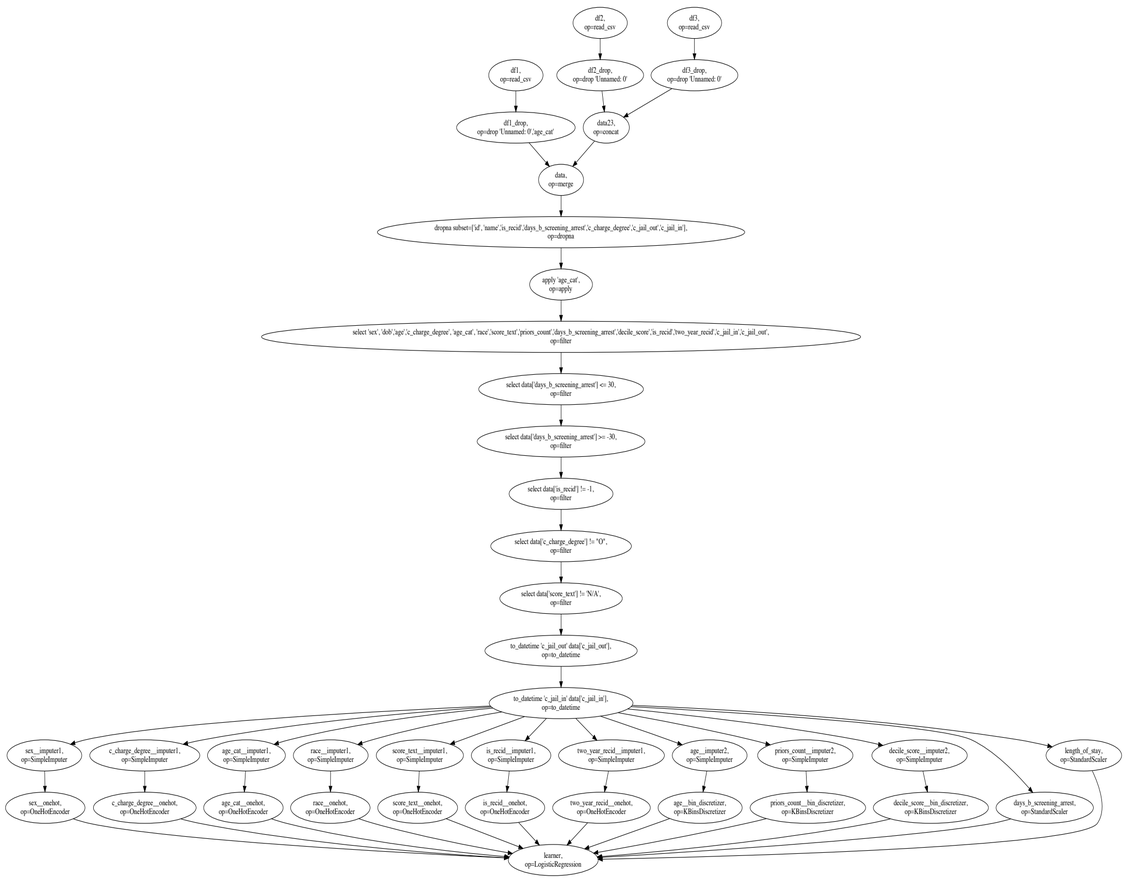
```python
@tracer(cat_col = ['race'], numerical_col = ['age'])
def compas_pipeline(f1_path = '../data/compass/demographic.csv',f2_path = '../data/compass/jailrecord1.csv',f3_path = '../data/compass/jailrecord2.csv'):
    #read csv files
    df1 = pd.read_csv(f1_path)
    df2 = pd.read_csv(f2_path)
    df3 = pd.read_csv(f3_path)

    #drop columns inplace
    df1.drop(columns=['Unnamed: 0','age_cat'],inplace=True)
    df2.drop(columns=['Unnamed: 0'],inplace=True)
    df3.drop(columns=['Unnamed: 0'],inplace=True)

    #JOIN dataframes column-wise and row-wise
    data23 = pd.concat([df2,df3],ignore_index=True)
    data = df1.merge(data23, on=['id','name'])

    #drop rows that miss a few important features
    data = data.dropna(subset=['id', 'name','is_recid','days_b_screening_arrest','c_charge_degree','c_jail_out','c_jail_in'])

    #generate a new column conditioned on existed column
    data['age_cat'] = data.apply(lambda row:'<25' if row['age'] < 25 else '>45' if row['age']>45 else '25-45', axis=1)

    #PROJECTION
    data = data[['sex', 'dob','age','c_charge_degree', 'age_cat', 'race','score_text','priors_count','days_b_screening_arrest',
                 'decile_score','is_recid','two_year_recid','c_jail_in','c_jail_out']]

    #SELECT based on some conditions
    data = data.loc[(data['days_b_screening_arrest'] <= 30)]
    data = data.loc[(data['days_b_screening_arrest'] >= -30)]
    data = data.loc[(data['is_recid'] != -1)]
    data = data.loc[(data['c_charge_degree'] != "O")]
    data = data.loc[(data['score_text'] != 'N/A')]
    # create a new feature
    data['c_jail_out'] = pd.to_datetime(data['c_jail_out'])
    data['c_jail_in'] = pd.to_datetime(data['c_jail_in'])
#     data['length_of_stay'] = data['c_jail_out'] - data['c_jail_in']
    #specify categorical and numeric features
    categorical = ['sex', 'c_charge_degree', 'age_cat', 'race', 'score_text', 'is_recid',
           'two_year_recid']
    numeric1 = ['age','priors_count', 'decile_score']
    numeric2 = ['days_b_screening_arrest','length_of_stay']

    #sklearn pipeline
    impute1_and_onehot = Pipeline([('imputer1', SimpleImputer(strategy='most_frequent')),
                                   ('onehot', OneHotEncoder(handle_unknown='ignore'))])
    impute2_and_bin = Pipeline([('imputer2', SimpleImputer(strategy='mean')),
                                ('bin_discretizer', KBinsDiscretizer(n_bins=4, encode='ordinal', strategy='uniform'))])
    featurizer = ColumnTransformer(transformers=[
            ('impute1_and_onehot', impute1_and_onehot, categorical),
            ('impute2_and_bin', impute2_and_bin, numeric1),
            ('std_scaler', StandardScaler(), numeric2),
        ])

    pipeline = Pipeline([
        ('features', featurizer),
        ('learner', LogisticRegression())
    ])
    return pipeline
```




```
######################## Start Pandas Opeation ########################

------------------------------------------------------------
Inpected df1 = pd.read_csv(f1_path)
------------------------------------------------------------

------------------------------------------------------------
Inpected df2 = pd.read_csv(f2_path)
------------------------------------------------------------

------------------------------------------------------------
Inpected df3 = pd.read_csv(f3_path)
------------------------------------------------------------

------------------------------------------------------------
Inpected df1.drop(columns=['Unnamed: 0','age_cat'],inplace=True)
------------------------------------------------------------

------------------------------------------------------------
Inpected df2.drop(columns=['Unnamed: 0'],inplace=True)
------------------------------------------------------------

------------------------------------------------------------
Inpected df3.drop(columns=['Unnamed: 0'],inplace=True)
------------------------------------------------------------

------------------------------------------------------------
Inpected data23 = pd.concat([df2,df3],ignore_index=True)
------------------------------------------------------------

------------------------------------------------------------
Inpected data = df1.merge(data23, on=['id','name'])
------------------------------------------------------------

**********
Changes in numerical features!
```

	count	missing_count	median	mad	range
age	-307.0	0.0	0.0	0.0	0.0

```
**********

**********
Changes in categorical features!
```

	missing_count	num_class	class_count	class_percent
race	0.0	0.0	{'African-American': -159, 'Caucasian': -76, 'Hispanic': -53, 'Other': -17, 'Asian': 0, 'Native American': -2}	{'African-American': -0.0002, 'Caucasian': 0.0041, 'Hispanic': -0.0037, 'Other': -0.0001, 'Asian': 0.0002, 'Native American': -0.0002}

```
**********
------------------------------------------------------------
Inpected data = data.dropna(subset=['id', 'name','is_recid','days_b_screening_arrest','c_charge_degree','c_j
ail_out','c_jail_in'])
------------------------------------------------------------

------------------------------------------------------------
Inpected data = data[['sex', 'dob','age','c_charge_degree', 'age_cat', 'race','score_text','priors_count','d
ays_b_screening_arrest','decile_score','is_recid','two_year_recid','c_jail_in','c_jail_out']]
------------------------------------------------------------

**********
Changes in numerical features!
```

	count	missing_count	median	mad	range
age	-284.0	0.0	0.0	0.0	0.0

```
**********

**********
Changes in categorical features!
```

	missing_count	num_class	class_count	class_percent
race	0.0	0.0	{'African-American': -158, 'Caucasian': -87, 'Hispanic': -27, 'Other': -9, 'Asian': 0, 'Native American': -3}	{'African-American': -0.0019, 'Caucasian': 0.0016, 'Hispanic': -0.0005, 'Other': 0.0009, 'Asian': 0.0002, 'Native American': -0.0004}

```
**********
------------------------------------------------------------
Inpected data = data.loc[(data['days_b_screening_arrest'] <= 30)]
------------------------------------------------------------

**********
Changes in numerical features!
```

	count	missing_count	median	mad	range
age	-451.0	0.0	0.0	0.0	0.0

```
**********

**********
Changes in categorical features!
```

	missing_count	num_class	class_count	class_percent
race	0.0	0.0	{'African-American': -204, 'Caucasian': -188, 'Hispanic': -48, 'Other': -8, 'Asian': -1, 'Native American': -2}	{'African-American': 0.0042, 'Caucasian': -0.0052, 'Hispanic': -0.0016, 'Other': 0.0026, 'Asian': 0.0002, 'Native American': -0.0002}

```
**********
------------------------------------------------------------
Inpected data = data.loc[(data['days_b_screening_arrest'] >= -30)]
------------------------------------------------------------

------------------------------------------------------------
Inpected data = data.loc[(data['is_recid'] != -1)]
------------------------------------------------------------

------------------------------------------------------------
Inpected data = data.loc[(data['c_charge_degree'] != "O")]
------------------------------------------------------------

------------------------------------------------------------
Inpected data = data.loc[(data['score_text'] != 'N/A')]
------------------------------------------------------------


######################## Start Sklearn Pipeline ########################

------------------------------------------------------------
Operations SimpleImputer on race
------------------------------------------------------------

------------------------------------------------------------
Operations OneHotEncoder on race
------------------------------------------------------------

**********
Changes in categorical features!
```

	race
missing_count	0
num_class	-4
class_count	{1.0: 3175, 0.0: 2997}
class_percent	{1.0: 0.5144, 0.0: 0.4856}

```
**********
------------------------------------------------------------
Operations SimpleImputer on age
------------------------------------------------------------

------------------------------------------------------------
Operations KBinsDiscretizer on age
------------------------------------------------------------

**********
Changes in numerical features!
```

	age
count	0.0000
missing_count	0.0000
median	-31.0000
mad	-10.3782
range	-75.0000

```
**********
```