

# How to add a new dataset to CFTree

1. You need to run the `.ipynb` file under CFTree folder to load a new dataset and generate the counterfactuals.
  - a. after running the ipynb file, we get a pickle file `dice_exp_{datasetName}.pkl`: for example, `dice_exp_LIDC.pkl`
  - b. copy this file into data folder of fastAPIbackend
2. You need to add it to DATASETS config in `main.py` under fastAPIbackend folder:

```
DATASETS = {  
    "income":  
        "df_cfs": "data/dice_exp_income.pkl",  
        "columns_to_drop": ["income", "Original_Index"],  
        "ordinal_attributes": {  
            "education": ["School", "HS-grad", "Some-college", "Bachelors", "M  
        },  
        "target_column_name": "income",  
    }
```

3. You need to add a dataset target name (for example: "income") to NavBar.js:

```
const datasetTargetNames = {  
    income: "income"  
};
```

4. The user needs to add a dataset in the `cfRowData.js` with the format as the following example. For the "originalDataset", the user needs to copy the

originalDataset in {datasetKey}\_x\_test\_first\_20\_with\_target.json to the cfRowData.js.

```
income: {
  name: "income dataset",
  target_column: 'income',
  originalDataset:[
    {
      "age": 29,
      "workclass": "Private",
      "education": "HS-grad",
      "marital_status": "Married",
      "occupation": "Blue-Collar",
      "race": "White",
      "gender": "Female",
      "hours_per_week": 38,
      "target": 0
    },
    ...
  ]
}
```