

# 5-dars (Report)

#drop

#1 ustun bosish.

df.drop(['Size'], axis=1, inplace=True)

df.drop(column=['Size', 'id'])

df.isnull().sum() → tablab uchibaynligini:

df['Size'].fillna(df['Size'].mode()[0], inplace=True)

df.unique() → classlar soni

## • Handling missing values.

- Mean
- Mode
- Median → Basic
- Fixed ( $\rightarrow$  Advanced)
- Drop

## • Encoding

## • Scaling

Encoding (object ga etlibor berishi)  
 Categorical Object  $\Rightarrow$  Numerical  
 int or float

## • Categorical → Numerical

### Turlari

- One-Hot VR
  - Label-Encoding VR
  - Frequency encoding
  - Target encoding
  - Ordinal encoding
- $\circ \Rightarrow df.unique.$



| Original Color | One-hot encoding |        |       | dummies. |
|----------------|------------------|--------|-------|----------|
|                | Red              | Yellow | Green |          |
| Red            | 1                | 0      | 0     |          |
| Red            | 1                | 0      | 0     |          |
| Yellow         | 0                | 1      | 0     |          |
| Green          | 0                | 0      | 1     |          |
| Yellow         | 0                | 1      | 0     |          |

Sumiyesturlas.

Dummies variablei  
 (dum)  
 (ustun)  
 correct

dummies = pd.get\_dummies(df['Size'], prefix='col', dtype=int)  
 dummies.

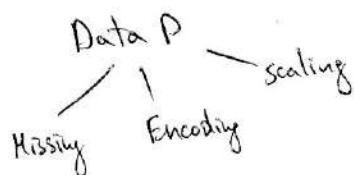
Dummies ni Data setga qo'shishi: 2) df = pd.concat([df.drop(columns=['size'])])

df = pd.concat([df.drop(columns=['size'])], axis=1)

df = pd.concat([df.drop(columns=['size']), dummies], axis=1)

df = pd.concat([df\_1, df\_2])

#dropni qator bosishga tablib yuborish.  
 df.dropna(inplace=True)



Label encoding.

- Label

| Original Data |        | Label Encode Data. |        |
|---------------|--------|--------------------|--------|
| Team          | Points | Team               | Points |
| A             | 25     | 0                  | 25     |
| A             | 12     | 0                  | 12     |
| B             | 15     | 1                  | 15     |
| B             | 14     | 1                  | 14     |
| B             | 19     | 1                  | 19     |
| B             | 23     | 1                  | 23     |
| C             | 25     | 2                  | 25     |
| C             | 28     | 2                  | 28     |

# Label Encoding

```
from sklearn.preprocessing import LabelEncoder.
```

```
encoder = LabelEncoder()
```

```
encoder
```

```
df['player_name'] = encoder.fit_transform(df['player_name'])
```

transform →  
fit-transform → iddiaları organizib transform qiledi.  
fit → organizib berdi.

Encoding:

- Datani bir turdan 2-ga östrəkiz.

- 1- One-hot

- 2- Label

Threshold chegara.

For Loop.