5-dars (Report)

#drop

#1 ustun bo'yicha.

```
df.drop('Size', axis=1, inplace=True)
df.drop(column=['Size','id'])
df.isnull().sum() → tashlab ketibgan qiymat:
```

#dropni qator bo'yicha tashlab yuborish.

```
df.dropna(inplace=True)
```

Data P
- Missing
- Encoding
- scaling

```
df['Size'].fillna(df['Size'].mode()[0], inplace=True)
df.nunique()  → classlar soni
```

- Handling missing values.
  - Mean
  - Mode
  - Median ⇒ Basic
  - Fixed (⇒ Advanced)
  - Drop
- Encoding
- Scaling

Encoding: (object ga e'tibor berish)

Categorical / Object ⇒ Numerical (int64, float)

- Categorical → Numerical
- Turlari
  - One-Hot ✓
  - Label-Encoding ✓
  - Frequency encoding
  - Target encoding
  - Ordinal encoding

  0 ⇒ df.nunique.

One-hot encoding → dummies.

original Color	Red	Yellow	Green
Red	1	0	0
Red	1	0	0
Yellow	0	1	0
Green	0	0	1
Yellow	0	1	0

Sunniy ustunlar.

Concat: Dummies yaratibolish

```
dummies = pd.get_dummies(df['Size'], prefix='col', dtype=int)
dummies
```

(prefix: ustun)

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

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

dummies

```
df = pd.concat([df.drop(columns=['Size']), dummies], axis=1)
df = pd.concat(df_1, df_2)
```

Label encoding.

- Label

Original Data

Team	Points
A	25
A	12
B	15
B	14
B	19
B	23
C	25
C	29

→

Label Encode Data.

Team	Points
0	25
0	12
1	15
1	14
1	19
1	23
2	25
2	29

Label Encoding

```
from sklearn.preprocessing import LabelEncoder.
    encoder = LabelEncoder()
    encoder
    df['player_name'] = encoder.fit_transform(df['player_name'])
```

↙

transform →

fit_transform → ichidagi qiymatlarini organib transform qiladi.

fit → organib beradi.

Encoding:

- Datani bir turdan 2-ga o'tkazish. threshold chegara.
  1- One-hot
  2- Label

For Loop.