Data Preprocessing: Review

- **Handling missing values   1**
    - **Mean**   **Basic**
    - **Mode**
    - **Median**   **Advanced**
    - **Fixed**
    - **Drop**
- **Encoding   2**
    - **One – hot**
    - **Label Encoding**
        Scaling   3

Encoding (continued)

- **One-hot**
- **Label**
- **Frequence**
- **Target**
- **Ordinal**

Frequency encoding

The number of repetitions is replaced by the ratio to the total number of elements.

Height	Sex
173.1	Male
160.4	Female
178.5	Male
155.5	Female
163.7	Female

Frequency Encoding →

Height	Sex
173.1	0.4
160.4	0.6
178.5	0.4
155.5	0.6
163.7	0.6

You can see here Males are two and Females are three and we should encode this Sex column how we do that we take total number of males and divide it to the total number of values and in the Sex column 5 values and we get 0.4 and when everytime we see Male in this case we put 0.4 and like so with Female values.

Target Encoding

Categorical values are converted into numerical values based on the arithmetic mean of their corresponding target values.

Rang	Target
Qizil	1
Ko'k	0
Yashil	1
Qizil	0
Ko'k	1
Yashil	0
Qizil	1

har bir rang uchun target o'rtachasini hisoblaymiz

- **Qizil** → (1 + 0 + 1) / 3 = **0.67**
- **Ko'k** → (0 + 1) / 2 = **0.50**
- **Yashil** → (1 + 0) / 2 = **0.50**

Ordinal Encoding

They are replaced with numbers according to their sequence.

	cost	size	size_endoced
0	50	large	1.0
1	35	small	3.0
2	75	extra large	0.0
3	42	medium	2.0
4	54	large	1.0
5	71	extra large	0.0

Size		Encoding
S	Ordinal encoding	2
XS		1
L		4
M		3
*data with order		

Degree	Degree order
High school	1
Bachelor	2
Master	3
Phd	4

For Loop

A block of code used to automate a process

```
a=[1,2,3]
print(a[0])
print(a[1])
print(a[2])
✓ 0.0s

1
2
3
```

```
for i in [1,2,3]:
    print(i)
✓ 0.0s

1
2
3
```

The first one with three prints is manual

The second one is made easier by writing just one print