11) Given a dataset with repetitive patterns design a simple physical model for lossless compression Justify how the model reduces data size and provide an example to demonstrate your approach.

1) Physical model: Run length Encoding (RLE)

concept

Instead of storing repeated data values individually, RLE stores:

- The values
- The number of times it repeats (run length).

model representation.

Each sequence is encoded as (value, count).

2) Compression Algorithm

Encoding procedure

- Read the dataset ~~size~~ sequentially.
- count consecutive occurrence of same value.
- Replace repeated values with a pair
- continue until the end of the dataset.

Decoding procedure

- Read each pair.
- Reconstruct the original data by repeating the value count times.

3) Justification of size reduction.

Let:

* Original size = number of elements N.
* Compressed size = number of runs R.

If data contains many repetitions

$R \ll N$.

∴ compression Ratio = $N/(2R)$.

Because each run stores only:

- One value
- One count.

Key insight

- Works best when data long repeated sequence
- Eliminates redundancy by grouping identical values.

4) Example

AAAABBCCCCCAA.

Step 1:

Identify runs..

- A repeated 4 times
- B repeats 2 times
- C repeats 5 times
- A repeats 2 times.

Step 2:

Encod
(A,4), (B,2), (C,5), (A,2).

Step 3

Size compression

Original length =13

compressed length = 4 pairs = 8 values..

5) Code snipped

```
def rle. encode (data):
    encode = []
    count =1
    for i in range (1, len(data)):
        if data [i] = = data [1-1]:
            count+=1:

    else:
            encode. append ((data [i-1], count))
            count =1
    encode .append ((data [-1], count))
    return encode
def rle_decode(encoded):
    decode= []
    for value, count in encoded:
        decoded. extend ([value]* count)
return decoded..
```

6) Advantages

- Simple to implement
- Lossless
- Effective for repetitive datasets.

7) Limitations

- Not efficent for non-repetitive data.
- May increase size, if data has non repetition..

8) Conclusion.

The Run length encoding model compresses data by replacing repeated sequence with (value, count) pairs. ~~ff~~ It reduces data size by eliminating redundancy, making it ~~hig~~ highly effective for datasets with repetitive patterns. It follows ~~Lossles~~ Lossless compression.