

Data Analysis Challenge

Problem Statement:

Section 1. Prepare Customers.csv file which contains columns as follows:

1. CustomerID
2. CustomerName
3. City

Using dataset given on below URL.

Dataset Link : <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

Output:

	A	B	C
1	CustomerID	CustomerName	City
2	12520	Claire Gute	Henderson
3			
4			
5			
6			
7			
8			

Hint: Shape of CSV file would be (793, 3)

Constraint:

1. CustomerID should be integer value. Eg. CG-1250 should be transformed to 1250.
Hence data type should be Integer.
2. All columns should have unique values.
3. There should be no Null values in final file.
4. For missing values or null values in dataset's City column
 - >>> if row number is divisible by 2 use Median of City column
 - >>> if row number is divisible by 3 use Mode of City column
 - >>> if row number neither divisible by 2 nor 3 use "Tampa" as City value for respective entry.
5. CustomerID and CustomerName should have all unique values.

Section 2. Prepare Login.csv file which contains columns as follows:

1. LoginID
2. Password
3. LoginType

using Customers.csv generated in last section.

Add 303rd entry as, {MOVIEADMIN, MOVIEADMIN, MOVIEADMIN}

Output:

	A	B	C
1	LoginID	Password	LoginType
2	12520	12520	CUSTOMER
3			
4			
5			
6			
7			
8			

Constraint:

1. There should be no Null value in final file.
2. Datatype for all columns should be String.
3. All values in LoginType must be in uppercase CUSTOMER except 303rd record.