NAME:J.A.MAANSI

TASK-01

CODE:

```python
import pandas as pd

df = pd.read_csv(r"C:\Users\julug\OneDrive\Documents\Mall_customers.csv")

df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")
print("Cleaned Column Names:", df.columns.tolist())

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

if 'gender' in df.columns:
    df['gender'] = df['gender'].astype(str).str.strip().str.lower().replace({'male': 'male', 'm': 'male', 'female': 'female', 'f': 'female'})

if 'country' in df.columns:
    df['country'] = df['country'].astype(str).str.strip().str.title()

if 'join_date' in df.columns:
    df['join_date'] = pd.to_datetime(df['join_date'], errors='coerce', dayfirst=True)

if 'age' in df.columns:
    df['age'] = pd.to_numeric(df['age'], errors='coerce')
    df = df[df['age'].notnull()]

df.to_csv("cleaned_dataset.csv", index=False)

summary = {
    "Removed duplicates": True,
    "Dropped rows with nulls": True,
    "Standardized columns": [col for col in ['gender', 'country'] if col in df.columns],
    "Converted date columns": [col for col in ['join_date'] if col in df.columns],
    "Renamed columns": list(df.columns),
    "Fixed data types": {col: str(df[col].dtype) for col in df.columns}
}

print("\n✅ Cleaning Complete. Summary of Changes:")
for k, v in summary.items():
    print(f"{k}: {v}")
```

OUTPUT:

```
Cleaned Column Names: ['customerid', 'gender', 'age', 'annual_income_(k$)', 'spending_score_(1-100)']

✅ Cleaning Complete. Summary of Changes:
Removed duplicates: True
Dropped rows with nulls: True
Standardized columns: ['gender']
Converted date columns: []
Renamed columns: ['customerid', 'gender', 'age', 'annual_income_(k$)', 'spending_score_(1-100)']
Fixed data types: {'customerid': 'int64', 'gender': 'object', 'age': 'int64', 'annual_income_(k$)': 'int64', 'spending_score_(1-100)': 'int64'}
```