```
#  Install PySpark
!pip install pyspark

#  Import libraries
from pyspark.sql import SparkSession
from pyspark.sql.functions import col

#  Spark Session
spark = SparkSession.builder.appName("Codetech Task 1 - Data Cleaning").getOrCreate()

#  Upload file from local
from google.colab import files
uploaded = files.upload()

# Read the uploaded CSV
df = spark.read.csv("sales_data.csv", header=True, inferSchema=True)

# Show original data
print(" Original Data:")
df.show()

# Check null values
print("\n Missing values per column:")
df.select([col(c).isNull().cast("int").alias(c) for c in df.columns])\
  .groupBy().sum().show()

#  Drop nulls
df_cleaned = df.dropna()

#  Drop duplicates
df_cleaned = df_cleaned.dropDuplicates()

#  Show cleaned data
print("\n Cleaned Data:")
df_cleaned.show()

# Save cleaned data
df_cleaned.write.csv("cleaned_sales_data.csv", header=True, mode="overwrite")

print("\n Task 1 complete. Cleaned data saved as cleaned_sales_data.csv")
```

```
Requirement already satisfied: pyspark in /usr/local/lib/python3.11/dist-packages (3.5.1)
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.11/dist-packages (from pyspark) (0.10.9.7)
```

Choose files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving sales_data.csv to sales_data.csv
 Original Data:
+-------------+-----------+------+
|Customer_Name|   Category|Amount|
+-------------+-----------+------+
|        Alice|Electronics|  1500|
|          Bob|   Clothing|   800|
|      Charlie|Electronics|  1800|
|        Diana|  Groceries|   300|
|          Eve|Electronics|  2000|
|        Frank|   Clothing|  1200|
|        Grace|  Groceries|   400|
|        Alice|Electronics|  1500|
+-------------+-----------+------+


 Missing values per column:
+------------------+-------------+-----------+
|sum(Customer_Name)|sum(Category)|sum(Amount)|
+------------------+-------------+-----------+
|                 0|            0|          0|
+------------------+-------------+-----------+
```