```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.impute import SimpleImputer

# a) Read the data with pandas
data = pd.read_csv("/content/CC GENERAL.csv")  # Replace "/content/CC
GENERAL.csv" with your actual dataset file path

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.impute import SimpleImputer

# a) Read the data with pandas
data = pd.read_csv("/content/CC GENERAL.csv")  # Replace "/content/CC
GENERAL.csv" with your actual dataset file path

# Drop non-numeric columns
data_numeric = data.select_dtypes(include=['number'])

# Impute missing values with mean
imputer = SimpleImputer(strategy='mean')
data_imputed = pd.DataFrame(imputer.fit_transform(data_numeric),
columns=data_numeric.columns)

# Check if 'CUST_ID' exists before dropping it
if 'CUST_ID' in data_imputed.columns:
    X = data_imputed.drop(columns=['CUST_ID'])
else:
        X = data_imputed.copy()

        # Split data into features and target variable
        y = data_imputed['BALANCE']  # Target variable

        # Split data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

        # b) Apply suitable model, and train
        model = LinearRegression()
        model.fit(X_train, y_train)

        # c) Print results
        print("Intercept:", model.intercept_)
        print("Coefficients:", model.coef_)
        print("\n")
```

```
        # Predict on the test set
        y_pred = model.predict(X_test)

        # Print evaluation metrics
        print("Mean Squared Error:", mean_squared_error(y_test,
y_pred))
        print("R-squared:", r2_score(y_test, y_pred))


Intercept: 0.0
Coefficients: [ 1.00000000e+00 -1.18642139e-14 -5.21837783e-17
1.24337386e-16
  9.23977245e-17  4.51448259e-16 -7.04412355e-14  8.40792973e-14
  9.26674178e-14 -2.55736530e-14  6.08948405e-16 -5.14156844e-16
  2.11903371e-16 -3.26417778e-16  2.33885095e-16 -3.30579512e-14
 -1.24777860e-14]


Mean Squared Error: 1.3667919996056604e-24
R-squared: 1.0
```