# Step 1: Import libraries

import pandas as pd

import numpy as np

from scipy import stats

import matplotlib.pyplot as plt

import seaborn as sns

from google.colab import files

# Step 2: Upload the CSV file

uploaded = files.upload()  # Prompt user to upload CSV

# Step 3: Load the dataset

df = pd.read\_csv(next(iter(uploaded)))

money = df["money"]

# Step 4: Descriptive statistics

mean = money.mean()

median = money.median()

mode = money.mode().tolist()

std\_dev = money.std()

variance = money.var()

coeff\_var = std\_dev / mean

skewness = money.skew()

kurtosis = money.kurtosis()

z\_scores = stats.zscore(money)

percentiles = np.percentile(money, [25, 50, 75])

quartiles = {

    "Q1 (25%)": percentiles[0],

    "Q2 (Median, 50%)": percentiles[1],

    "Q3 (75%)": percentiles[2]

}

correlation = df.corr(numeric\_only=True)

# Step 5: Print results

print("Mean:", mean)

print("Median:", median)

print("Mode:", mode)

print("Standard Deviation:", std\_dev)

print("Variance:", variance)

print("Coefficient of Variation:", coeff\_var)

print("Skewness:", skewness)

print("Kurtosis:", kurtosis)

print("First 5 Z-scores:\n", z\_scores[:5])

print("Quartiles:", quartiles)

print("\nCorrelation Matrix:\n", correlation)

# Step 6: Optional Visualizations

plt.figure(figsize=(14, 6))

# Histogram

plt.subplot(1, 2, 1)

sns.histplot(money, kde=True, bins=30, color='skyblue')

plt.title('Money Distribution')

# Boxplot

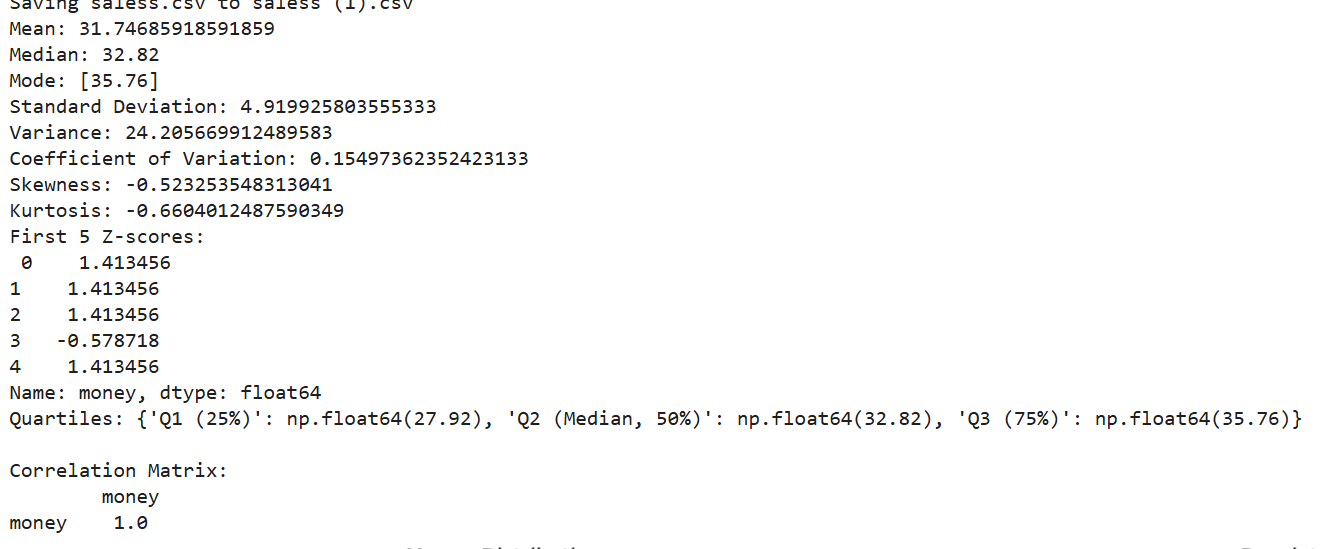
plt.subplot(1, 2, 2)

sns.boxplot(x=money, color='lightgreen')

plt.title('Boxplot of Money')

plt.tight\_layout()

plt.show()



This code performs data analysis and visualization on a dataset (assumed to have a column named money). Here's a brief explanation:

**Steps:**

1. **Import Libraries**: The code imports necessary libraries such as pandas, numpy, scipy, matplotlib, and seaborn for data manipulation, statistical analysis, and plotting. It also imports files from Google Colab to handle file uploads.
2. **Upload CSV**: The files.upload() function prompts the user to upload a CSV file containing the dataset. The dataset is then loaded into a pandas DataFrame.
3. **Descriptive Statistics**:
   * **Mean, Median, Mode**: Basic statistics to measure the central tendency of the money column.
   * **Standard Deviation, Variance**: Measures of spread or dispersion of the money values.
   * **Coefficient of Variation**: A relative measure of variability (standard deviation divided by mean).
   * **Skewness and Kurtosis**: Skewness measures the asymmetry of the data, and kurtosis measures the "tailedness" of the distribution.
   * **Z-scores**: Standardized scores that indicate how many standard deviations each value is from the mean.
   * **Percentiles/Quartiles**: These are values that divide the data into segments (25%, 50%, 75%).
   * **Correlation**: Computes correlations between numeric columns in the dataset.
4. **Print Results**: The calculated statistics (mean, median, mode, etc.) are printed for review.
5. **Optional Visualizations**:
   * **Histogram**: Plots the distribution of the money column, with a Kernel Density Estimate (KDE) curve overlaid to show the data’s probability density.
   * **Boxplot**: Visualizes the spread and outliers of the money values, showing quartiles and potential extreme values.

**Purpose:**

This code helps to:

* Summarize the data with key statistics (central tendency, spread, shape).
* Visualize the distribution and outliers.
* Understand relationships between variables (via correlation).