import warnings

warnings.filterwarnings('ignore')

%matplotlib inline

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

import seaborn as sns

import time

import scipy.stats as stats

from scipy.stats import linregress

shopping\_trends\_path=("shopping\_trends\_updated.csv")

shopping\_trends\_df = pd.read\_csv(shopping\_trends\_path)

shopping\_trends\_df.head()

shopping\_trends\_df.info()

shopping\_trends\_df["Purchase Amount (USD)"].head()

shopping\_trends\_df["Review Rating"].head()

average\_age=shopping\_trends\_df["Age"].mean()

average\_age

# shopping\_trends\_df.groupby("Review Rating")["Age"]

df2= shopping\_trends\_df.groupby(['Purchase Amount (USD)','Age']).size().reset\_index(name='Count')

df2

# shopping\_trends\_df.groupby("Review Rating")["Age"]

df3= shopping\_trends\_df.groupby(['Review Rating', 'Age']).size().reset\_index(name='Count')

df3

average\_purchase\_amount\_by\_age = shopping\_trends\_df.groupby("Age")["Purchase Amount (USD)"].mean().reset\_index()

average\_purchase\_amount\_by\_age

# Merge the average purchase amount back into the original DataFrame

shopping\_trends\_df = pd.merge(shopping\_trends\_df, average\_purchase\_amount\_by\_age, on="Age", how="left", suffixes=('', '\_avg'))

shopping\_trends\_df

def plot\_linear\_regression(df, x\_column, y\_column):

# Extract x and y values

x\_values = df[x\_column]

y\_values = df[y\_column]

# Run regression

slope, intercept, rvalue, pvalue, stderr = linregress(x\_values, y\_values)

# Calculate the regression values

regress\_values = x\_values \* slope + intercept

# Create the equation of the line

line\_eq = f"y = {round(slope, 2)}x + {round(intercept, 2)}"

# Plot values

plt.scatter(x\_values, y\_values)

plt.plot(x\_values, regress\_values, "-r")

plt.annotate(line\_eq, (min(x\_values), min(y\_values)), fontsize=15, color="red")

plt.xlabel(x\_column)

plt.ylabel(y\_column)

plt.title(f"Linear Regression: {x\_column} vs. {y\_column}")

plt.show()

print(f"The r-squared is: {rvalue\*\*2}")

# Plot the linear regression

plot\_linear\_regression(shopping\_trends\_df, "Age", "Purchase Amount (USD)\_avg")

average\_review\_rating\_by\_age = shopping\_trends\_df.groupby("Age")["Review Rating"].mean().reset\_index()

average\_review\_rating\_by\_age

# Merge the average review rating back into the original DataFrame

shopping\_trends\_df = pd.merge(shopping\_trends\_df, average\_review\_rating\_by\_age, on="Age", how="left", suffixes=('', '\_avg'))

shopping\_trends\_df

plot\_linear\_regression(shopping\_trends\_df,"Age","Review Rating\_avg")

average\_purchase\_amount\_by\_age.mean()

average\_purchase\_amount\_by\_age.median()

average\_review\_rating\_by\_age.mean()

average\_review\_rating\_by\_age.median()

average\_purchase\_amount\_by\_age.describe()

average\_review\_rating\_by\_age.describe()

shopping\_trends\_df.groupby("Age",as\_index=False)["Purchase Amount (USD)\_avg"].median().corr()

sns.violinplot(data=shopping\_trends\_df,x="Gender",y="Purchase Amount (USD)")

plt.title("Gender vs Purchase Amount (USD)")

plt.show()

shopping\_trends\_df.groupby("Gender").agg({"Purchase Amount (USD)":["mean","var"]})

shopping\_trends\_df.groupby("Gender").agg({"Review Rating":["mean","var"]})

males= shopping\_trends\_df.loc[shopping\_trends\_df.Gender == "Male", "Purchase Amount (USD)"]

print(males.mean())

print(males.var())

females= shopping\_trends\_df.loc[shopping\_trends\_df.Gender == "Female", "Purchase Amount (USD)"]

print(females.mean())

print(females.var())

stats.ttest\_ind(males, females, equal\_var=True)

males= shopping\_trends\_df.loc[shopping\_trends\_df.Gender == "Male", "Review Rating"]

print(males.mean())

print(males.var())

females= shopping\_trends\_df.loc[shopping\_trends\_df.Gender == "Female", "Review Rating"]

print(females.mean())

print(females.var())

stats.ttest\_ind(males, females, equal\_var=True)

sns.violinplot(data=shopping\_trends\_df,x="Gender",y="Review Rating")

plt.title("Gender vs Review Rating")

plt.show()