

**AI/ML Programming**

**MCA-475**

**Assignment – 04**

***BY***

**HIMANSHU HEDA (24225013)**

**SUBMITTED TO**

**Dr. Manjula Shannhog**

**SCHOOL OF SCIENCES**

**2025-26**

**Importing Libraries**

import pandas as pd

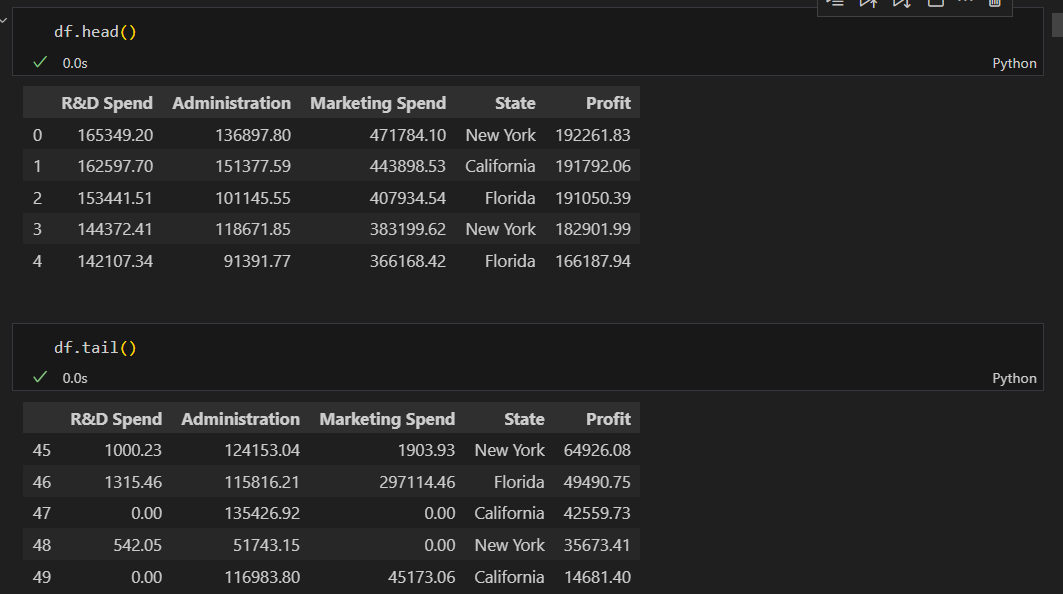
import numpy as np

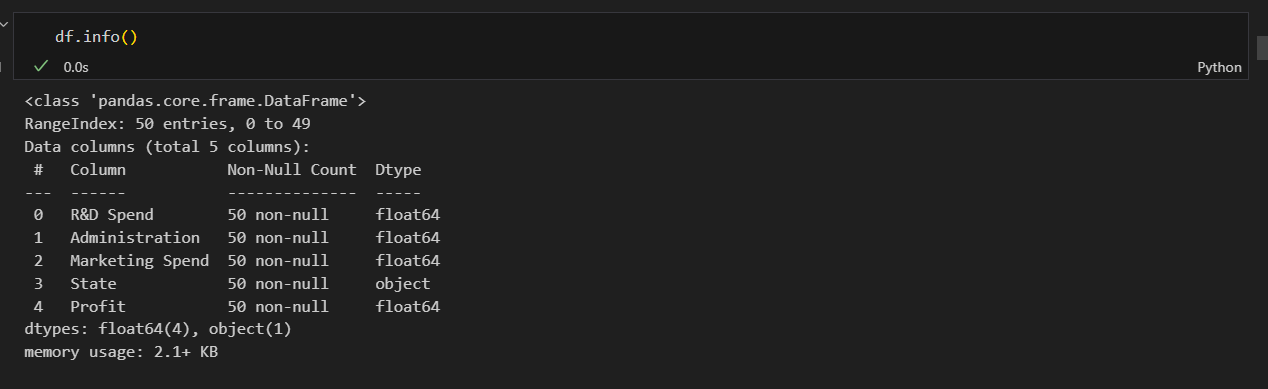
import matplotlib.pyplot as plt

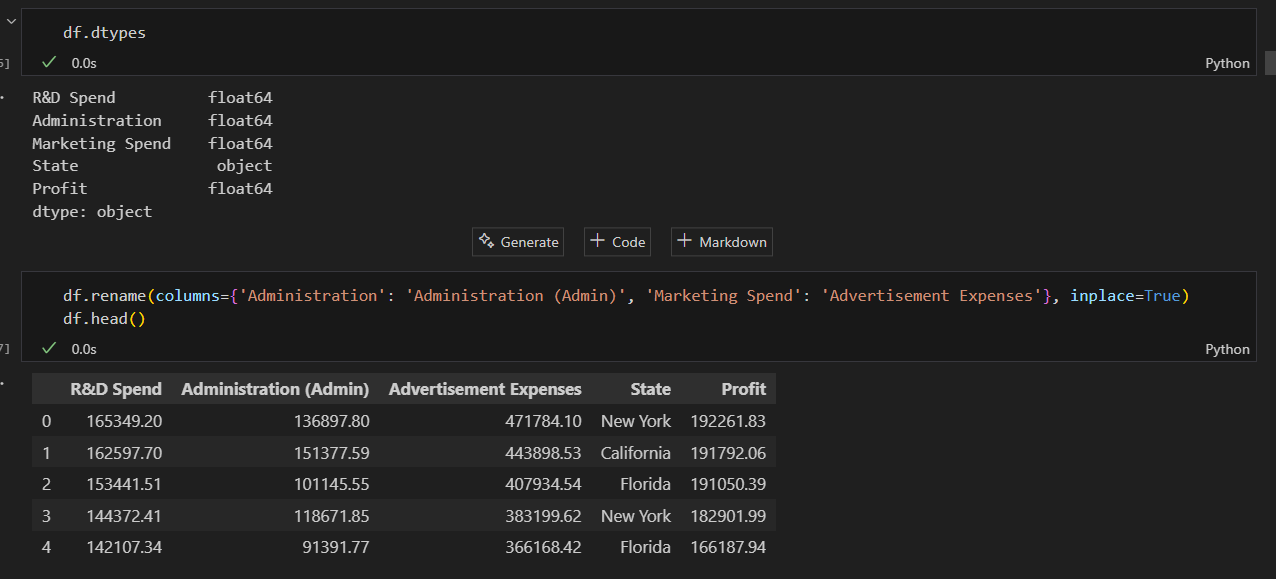
import seaborn as sns

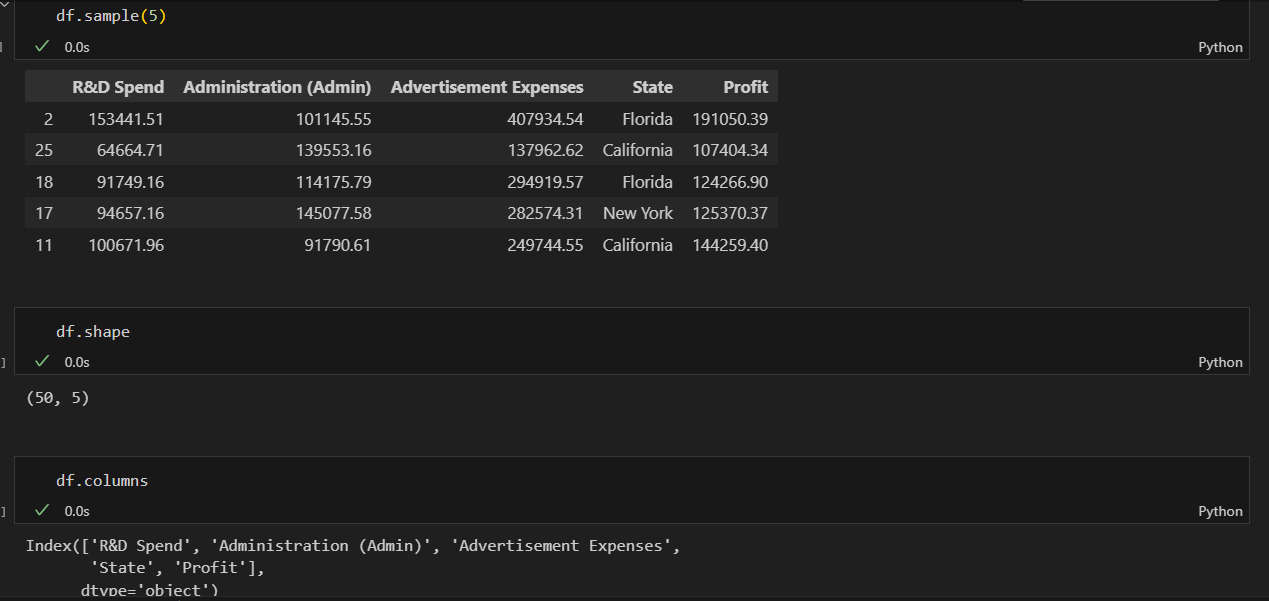
from datetime import date

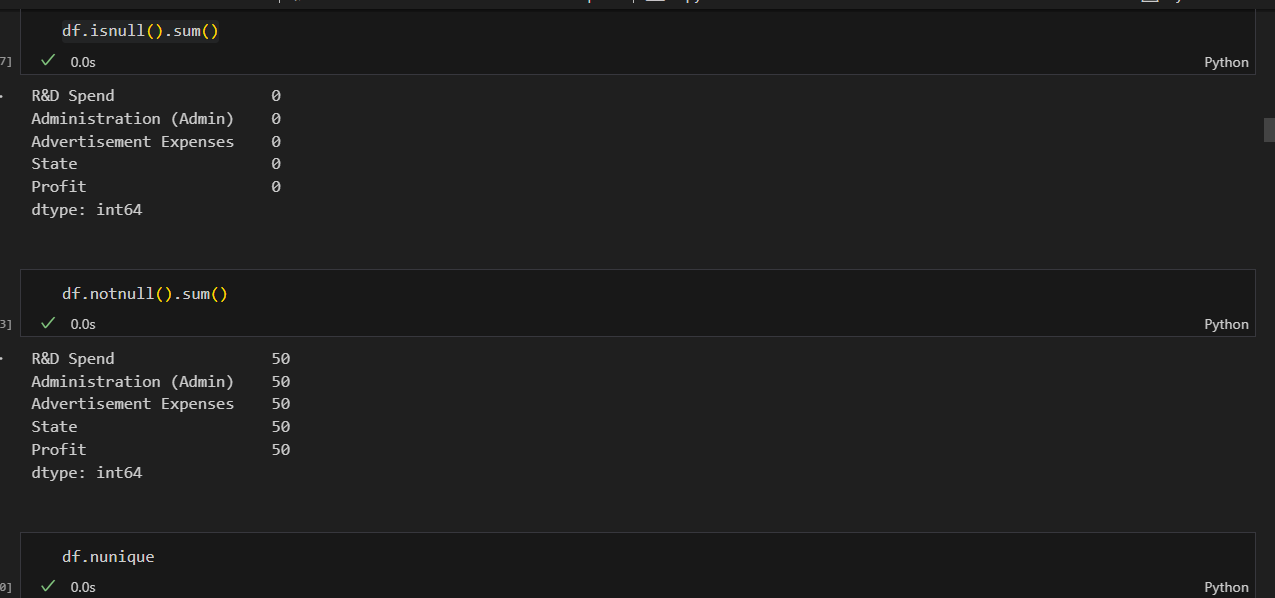
df = pd.read\_csv('./Dataset/50\_Startups.csv')

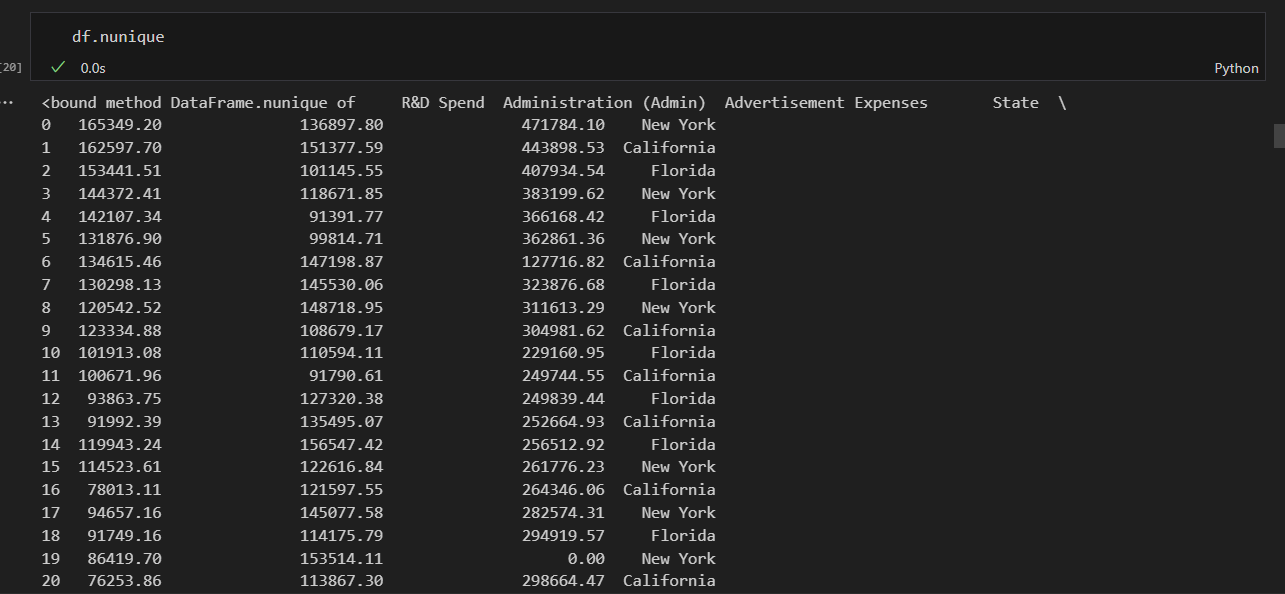
****

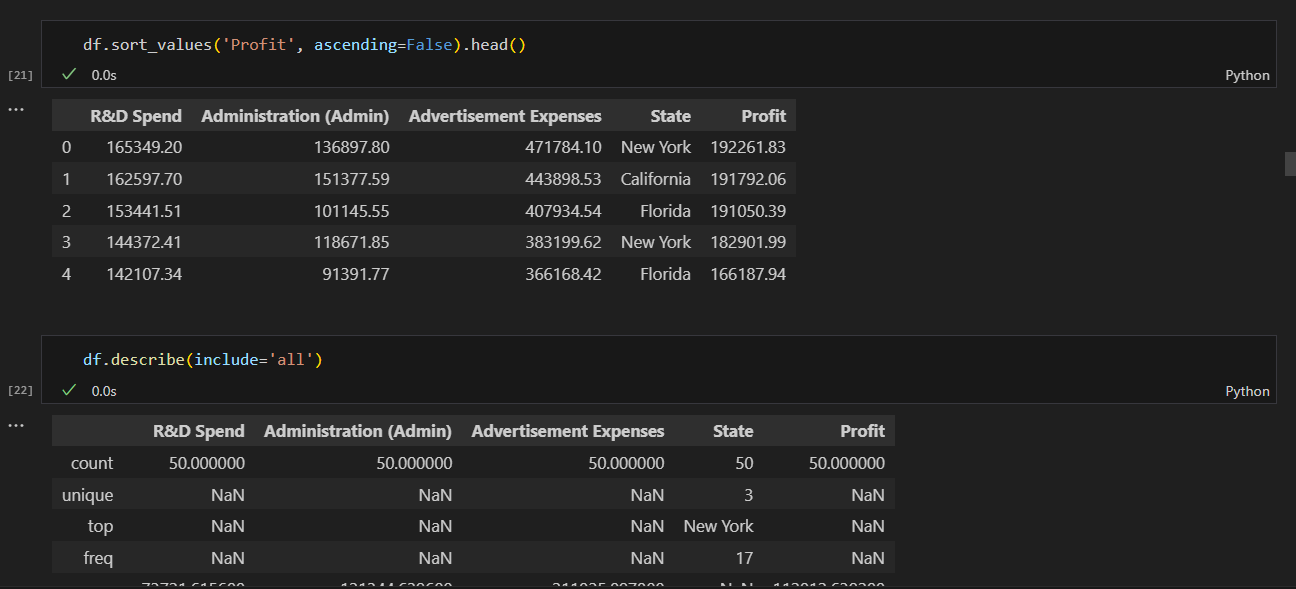
****

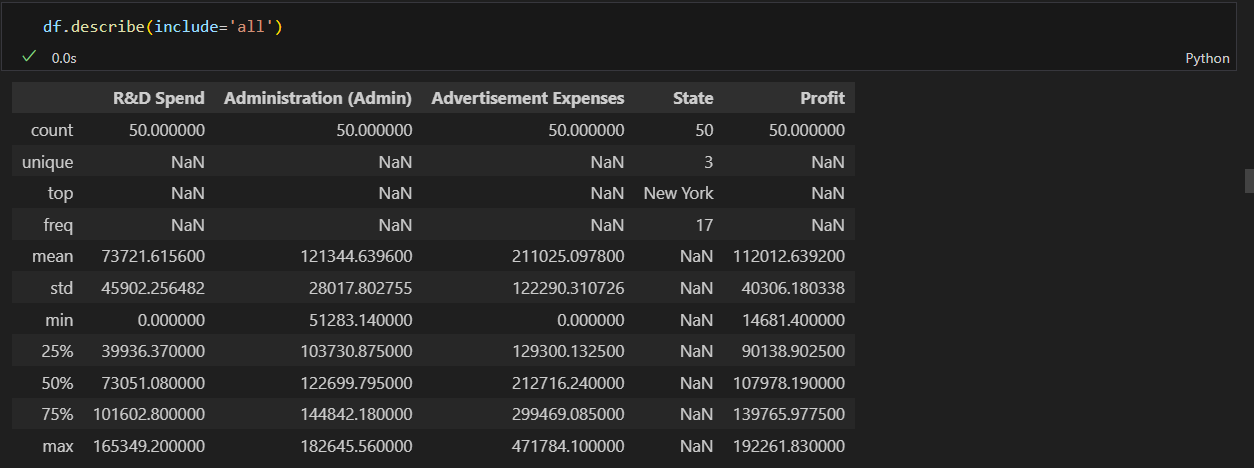
****

****

****

****

****

****

**Graphical Presentation :**

# 1. Histogram of Profit

plt.figure(figsize=(8,5))

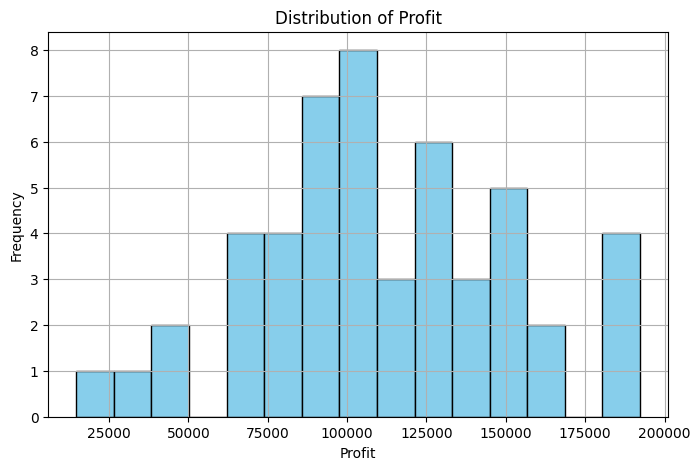
df['Profit'].hist(bins=15, color='skyblue', edgecolor='black')

plt.title('Distribution of Profit')

plt.xlabel('Profit')

plt.ylabel('Frequency')

plt.show()

****

# 2. Boxplot of Advertisement Expenses by State

plt.figure(figsize=(8,5))

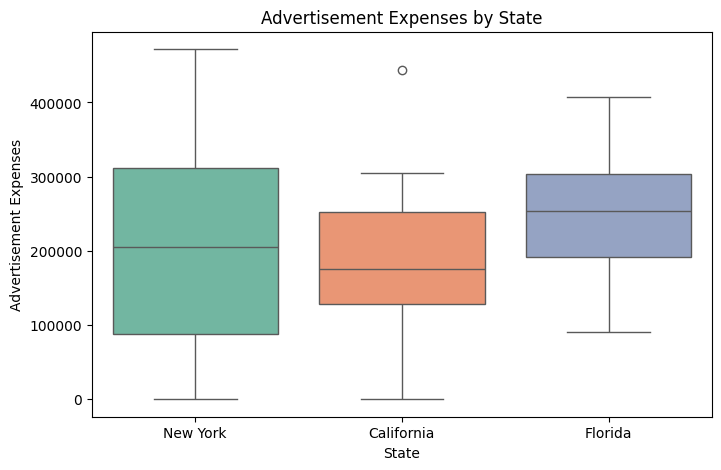
sns.boxplot(data=df, x='State', y='Advertisement Expenses', palette='Set2')

plt.title('Advertisement Expenses by State')

plt.xlabel('State')

plt.ylabel('Advertisement Expenses')

plt.show()

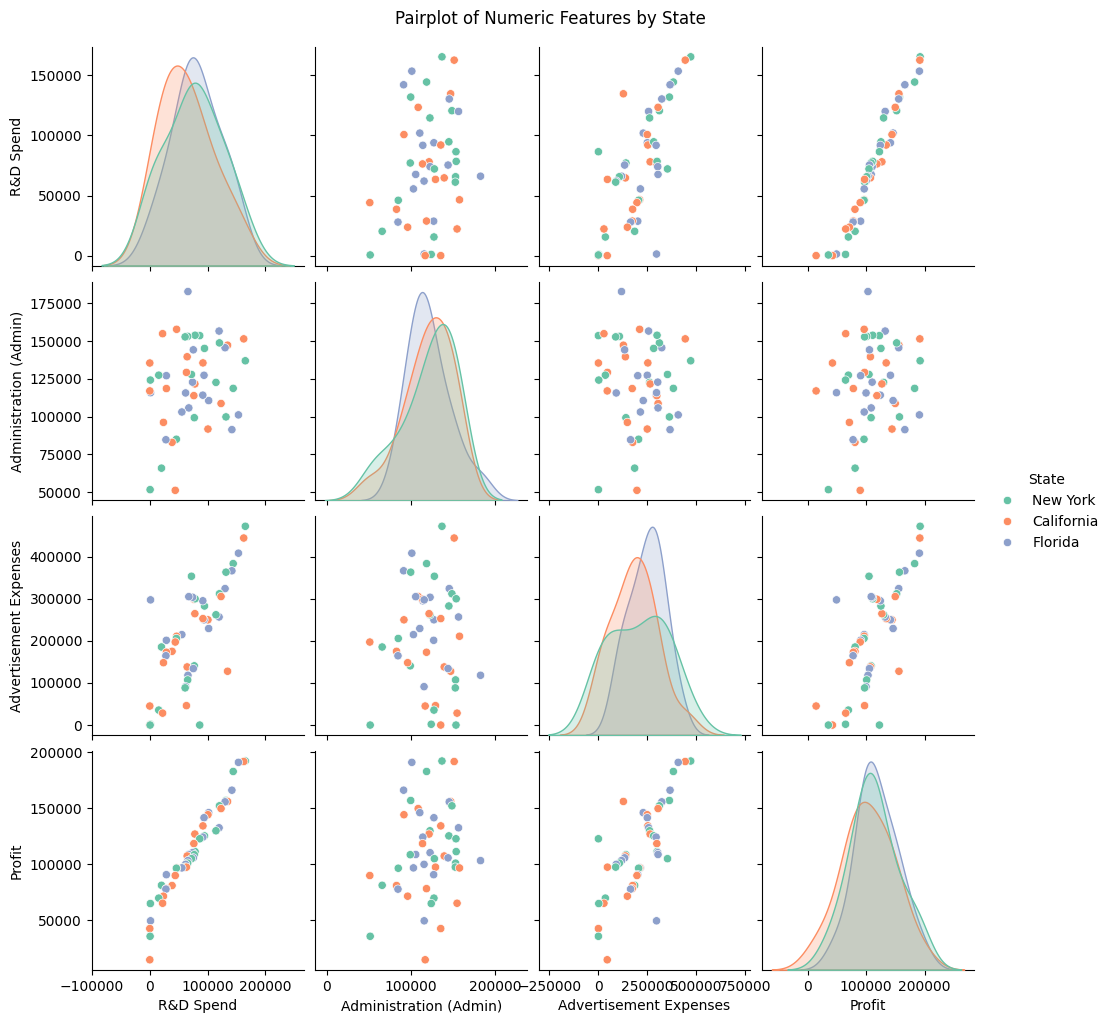
****

# 3. Pairplot of numeric features

sns.pairplot(df, vars=['R&D Spend', 'Administration (Admin)', 'Advertisement Expenses', 'Profit'], hue='State', palette='Set2')

plt.suptitle('Pairplot of Numeric Features by State', y=1.02)

plt.show()

****

# 4. Barplot of average Profit by State

plt.figure(figsize=(7,5))

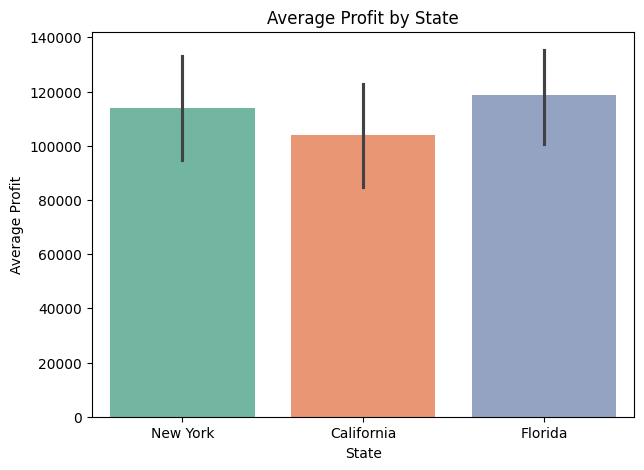
sns.barplot(data=df, x='State', y='Profit', estimator=np.mean, palette='Set2')

plt.title('Average Profit by State')

plt.xlabel('State')

plt.ylabel('Average Profit')

plt.show()

****

# 5. Scatterplot of R&D Spend vs Profit

plt.figure(figsize=(8,5))

sns.scatterplot(data=df, x='R&D Spend', y='Profit', hue='State', palette='Set2', s=80)

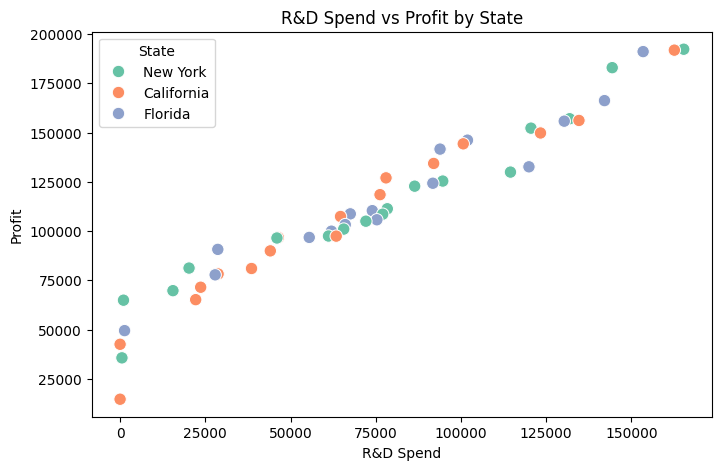
plt.title('R&D Spend vs Profit by State')

plt.xlabel('R&D Spend')

plt.ylabel('Profit')

plt.legend(title='State')

plt.show()

****

# 6. Correlation heatmap

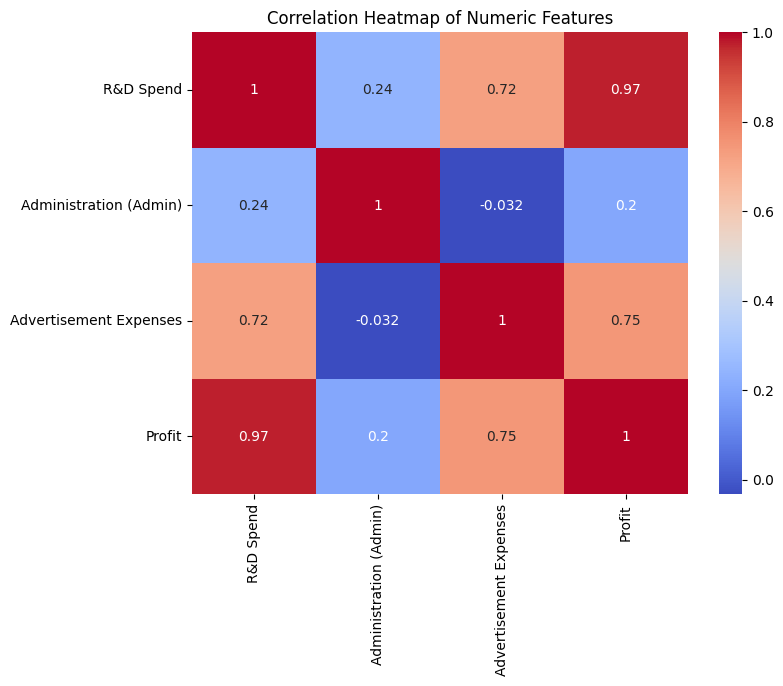
plt.figure(figsize=(8,6))

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

sns.heatmap(corr, annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap of Numeric Features')

plt.show()

****

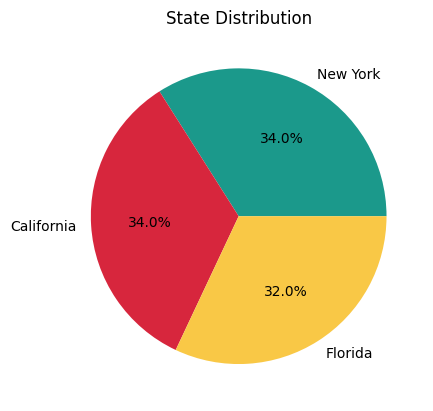
# 7. Pie chart of State distribution

df['State'].value\_counts().plot(kind='pie', autopct='%1.1f%%', colors=['#1b998b','#d7263d','#f9c846'])

plt.title('State Distribution')

plt.ylabel('')

plt.show()

****

# 8. Boxplot of Profit by Advertisement Expenses quartiles

df['Ad\_Quartile'] = pd.qcut(df['Advertisement Expenses'], 4, labels=['Q1','Q2','Q3','Q4'])

plt.figure(figsize=(8,5))

sns.boxplot(data=df, x='Ad\_Quartile', y='Profit', palette='Set2')

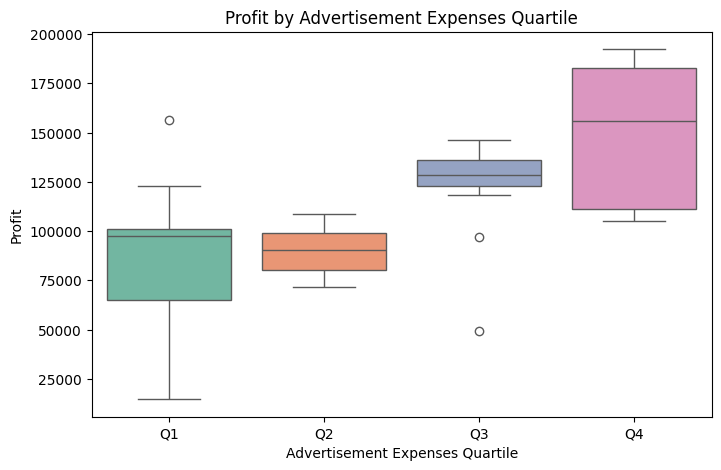
plt.title('Profit by Advertisement Expenses Quartile')

plt.xlabel('Advertisement Expenses Quartile')

plt.ylabel('Profit')

plt.show()

df.drop('Ad\_Quartile', axis=1, inplace=True)

****

# 9. Scatterplot of Administration (Admin) vs Profit

plt.figure(figsize=(8,5))

sns.scatterplot(data=df, x='Administration (Admin)', y='Profit', hue='State', palette='Set2', s=80)

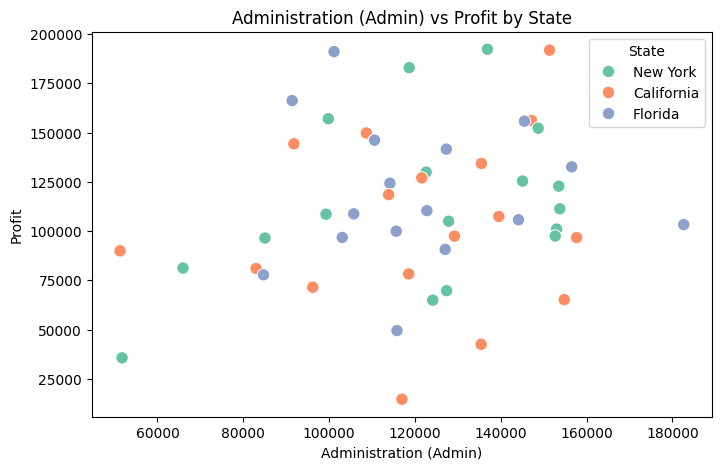
plt.title('Administration (Admin) vs Profit by State')

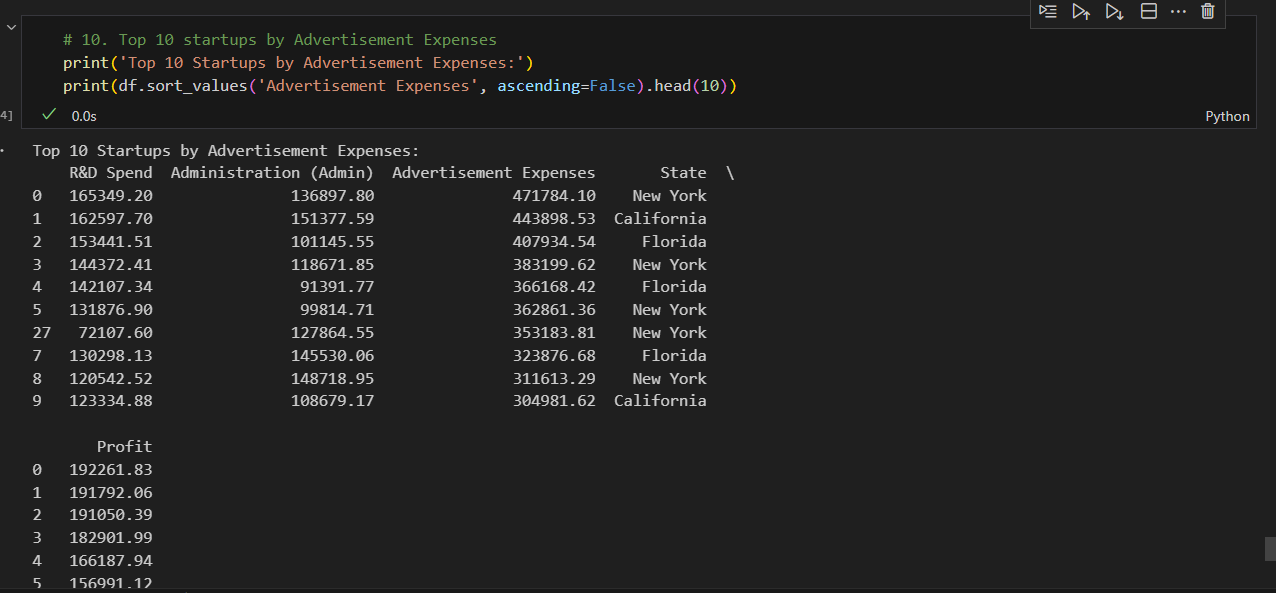
plt.xlabel('Administration (Admin)')

plt.ylabel('Profit')

plt.legend(title='State')

plt.show()

****

****

****