|  |  |
| --- | --- |
| **Date** | **17-10-2023** |
| **Team ID** | **3869** |
| **Project Name** | **Future Sales Prediction** |

## 

## Importing Dependencies

import numpy as np  
import pandas as pd  
import plotly as px  
import seaborn as sns  
import matplotlib.pyplot as plt  
from sklearn.preprocessing import MinMaxScaler

# Load Dataset

data = pd.read\_csv('Sales.csv')  
data.head()

TV Radio Newspaper Sales  
0 230.1 37.8 69.2 22.1  
1 44.5 39.3 45.1 10.4  
2 17.2 45.9 69.3 12.0  
3 151.5 41.3 58.5 16.5  
4 180.8 10.8 58.4 17.9

data.shape

(200, 4)

data.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 200 entries, 0 to 199  
Data columns (total 4 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 TV 200 non-null float64  
 1 Radio 200 non-null float64  
 2 Newspaper 200 non-null float64  
 3 Sales 200 non-null float64  
dtypes: float64(4)  
memory usage: 6.4 KB

data.describe().T

count mean std min 25% 50% 75% max  
TV 200.0 147.0425 85.854236 0.7 74.375 149.75 218.825 296.4  
Radio 200.0 23.2640 14.846809 0.0 9.975 22.90 36.525 49.6  
Newspaper 200.0 30.5540 21.778621 0.3 12.750 25.75 45.100 114.0  
Sales 200.0 15.1305 5.283892 1.6 11.000 16.00 19.050 27.0

data.columns

Index(['TV', 'Radio', 'Newspaper', 'Sales'], dtype='object')

# Visualisation of Data

sns.set(style="white")  
sns.jointplot(data=data, x='Sales', y='TV', kind='reg', height=7, color='skyblue', line\_kws={'color':'darkblue'})  
plt.xlabel('Sales', fontsize=14)  
plt.ylabel('TV Advertising', fontsize=14)  
plt.suptitle('Relationship Between Sales and TV Advertising', y=1.02, fontsize=16)

plt.show()

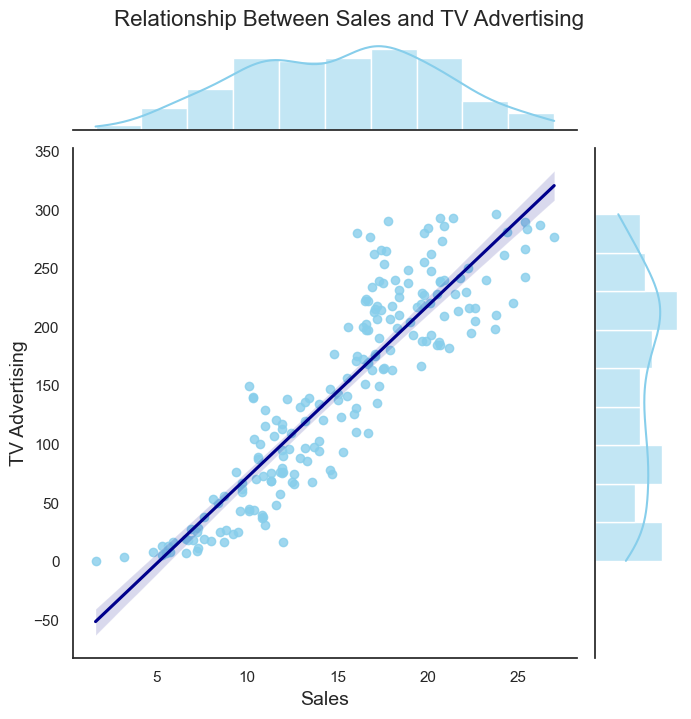


Fig.1 Relationship Between Sales and TV Advertising

sns.set(style="white")  
sns.jointplot(data=data, x='Sales', y='Newspaper', kind='reg', height=7, color='skyblue', line\_kws={'color':'darkgreen'})  
plt.xlabel('Sales', fontsize=14)  
plt.ylabel('Newspaper Advertising', fontsize=14)  
plt.suptitle('Relationship Between Sales and Newspaper Advertising', y=1.02, fontsize=16)  
plt.show()

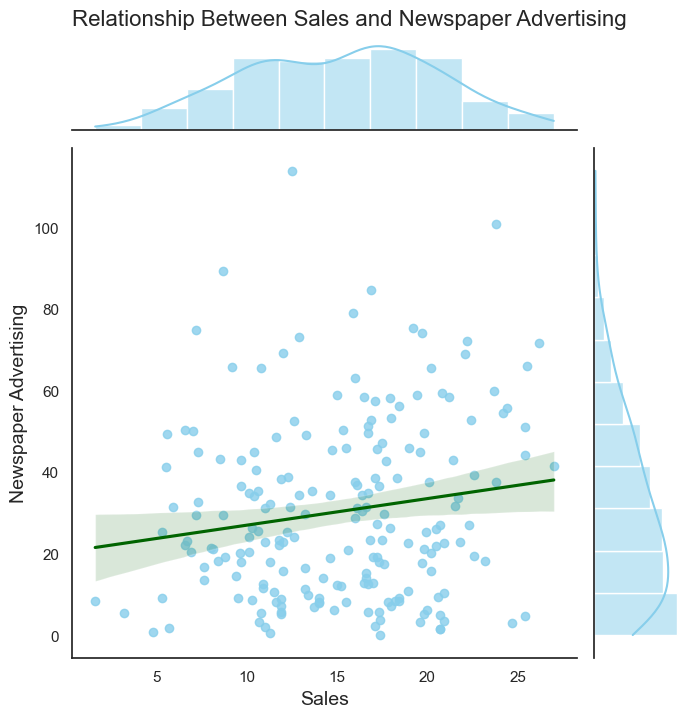


Fig 2. Relationship Between Sales and Newspaper Advertising

sns.set(style="white")  
sns.jointplot(data=data, x='Sales', y='Radio', kind='reg', height=7, color='skyblue', line\_kws={'color':'darkorange'})  
plt.xlabel('Sales', fontsize=14)  
plt.ylabel('Radio Advertising', fontsize=14)  
plt.suptitle('Relationship Between Sales and Radio Advertising', y=1.02, fontsize=16)  
plt.show()

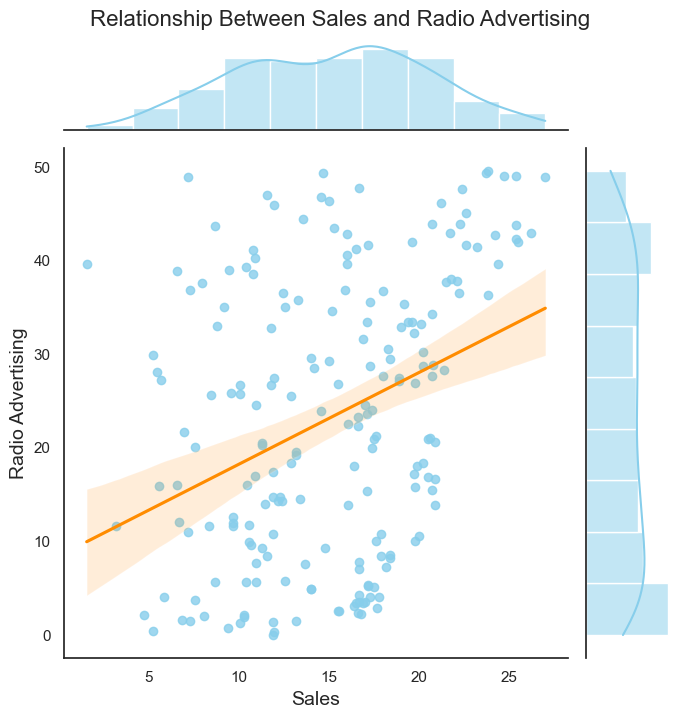


Fig 3. Relationship between Sales and Radio Advertising

correlation = data.corr()  
sales\_correlation = correlation["Sales"].sort\_values(ascending=False)  
styled\_sales\_correlation = sales\_correlation.apply(lambda x: f'{x:.2f}')  
styled\_sales\_correlation = styled\_sales\_correlation.reset\_index()  
styled\_sales\_correlation.columns = ["Feature", "Correlation with Sales"]  
styled\_sales\_correlation.style.background\_gradient(cmap='coolwarm', axis=0)

<pandas.io.formats.style.Styler at 0x1b1017ba250>

plt.figure(figsize=(12,8))  
sns.pairplot(data)

<seaborn.axisgrid.PairGrid at 0x1b17da6b410>

<Figure size 1200x800 with 0 Axes>

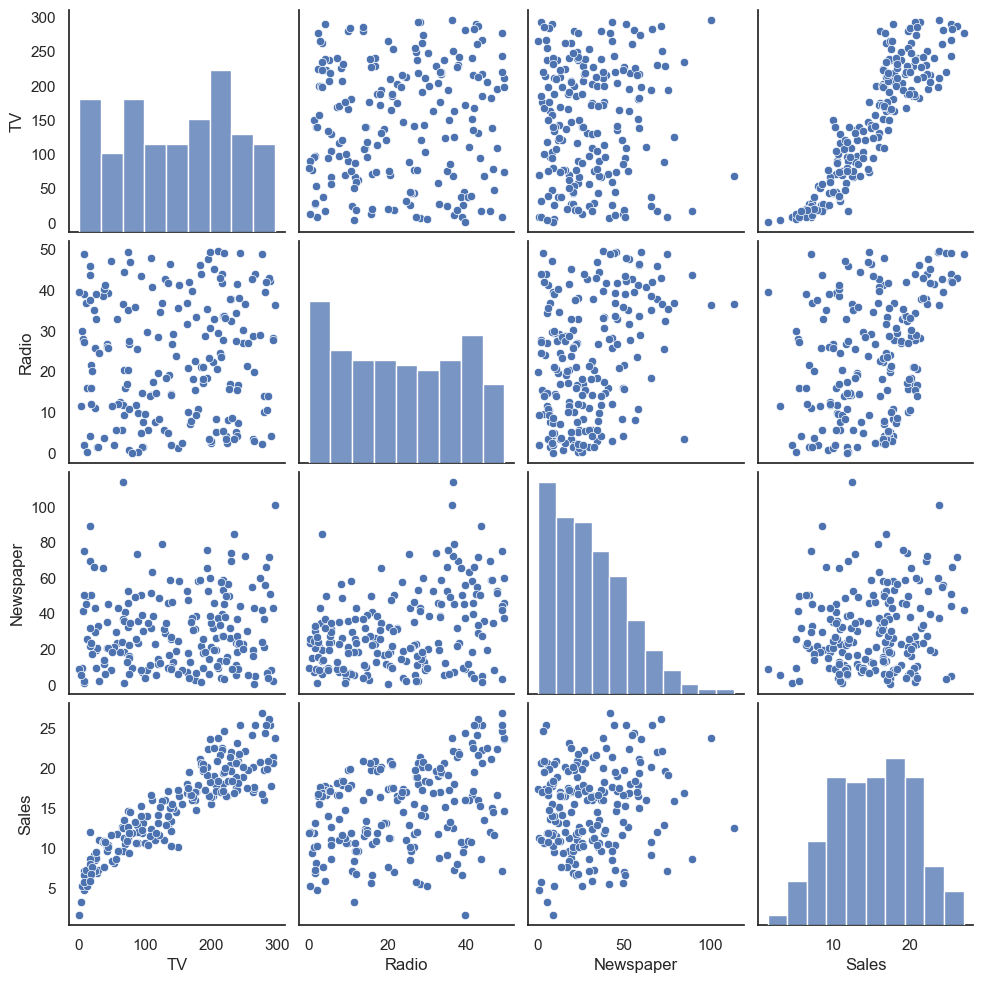


Fig 4. Pair Plot of TV, Radio, Newspaper, and Sales

# Data Preprocessing

data.hist(figsize=(10,8))

array([[<Axes: title={'center': 'TV'}>,  
 <Axes: title={'center': 'Radio'}>],  
 [<Axes: title={'center': 'Newspaper'}>,  
 <Axes: title={'center': 'Sales'}>]], dtype=object)

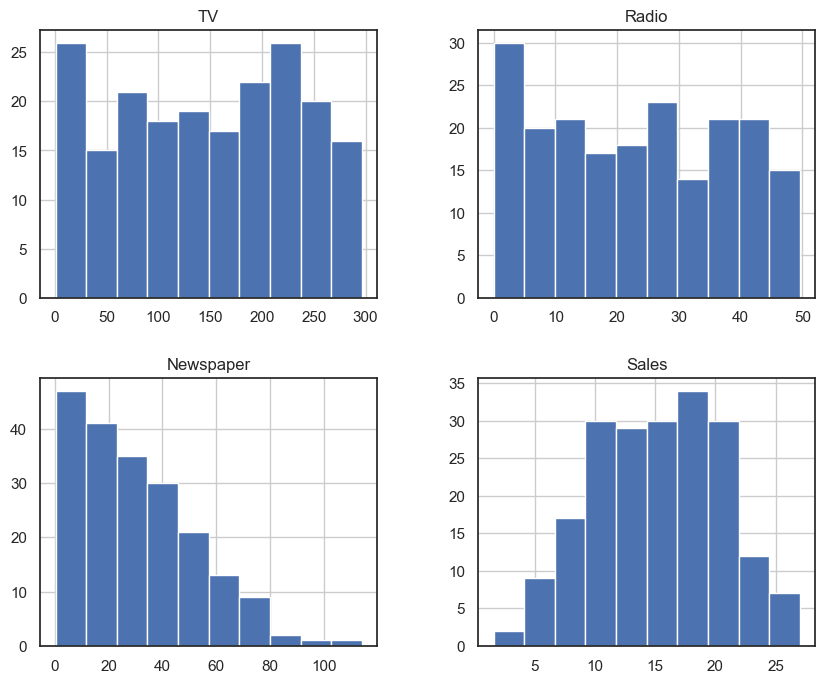


Fig 5. Histograms of TV, Radio, Newspaper, and Sales

# Data normalization

scaler = MinMaxScaler()  
columns\_to\_normalize = ['TV', 'Radio', 'Newspaper']  
data[columns\_to\_normalize] = scaler.fit\_transform(data[columns\_to\_normalize])  
data.head()

TV Radio Newspaper Sales  
0 0.775786 0.762097 0.605981 22.1  
1 0.148123 0.792339 0.394019 10.4  
2 0.055800 0.925403 0.606860 12.0  
3 0.509976 0.832661 0.511873 16.5  
4 0.609063 0.217742 0.510994 17.9

sns.heatmap(data.corr(),annot=True)

<Axes: >

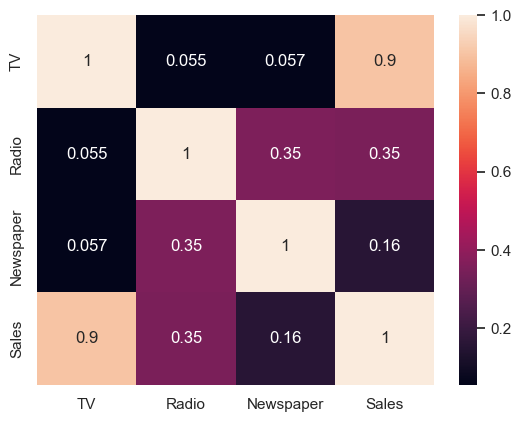


Fig 6. Correlation Heatmap