**Campaign Data**: Use Python to answer the questions below.

* Find out the top 5 months in terms of impressions.

Ans:

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_excel('Campaign data.xlsx', sheet\_name='Sheet1', engine='openpyxl')

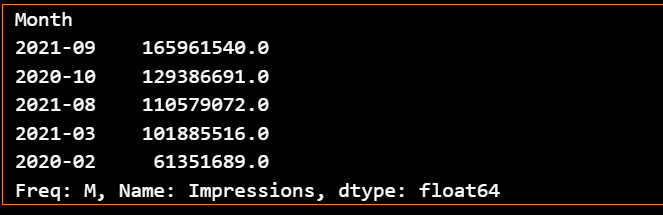
pivot\_df = df.pivot\_table(index=['Date', 'Data source', 'Campaign', 'Media Type', 'City'], columns='Metric', values='Value', aggfunc='sum').reset\_index()

pivot\_df['Date'] = pd.to\_datetime(pivot\_df['Date'])

pivot\_df['Month'] = pivot\_df['Date'].dt.to\_period('M')

top\_5\_months = pivot\_df.groupby('Month')['Impressions'].sum().nlargest(5)

top\_5\_months



* Which model has maximum impressions?

Ans: By model, we have considered campaigns.

max\_impressions\_model = pivot\_df.groupby('Campaign')['Impressions'].sum().idxmax()

max\_impressions\_model

A black and white sign with white text

AI-generated content may be incorrect.

* Find out the top 3 campaigns in terms of impressions for every message type

Ans:

top\_3\_campaigns\_per\_media\_type = pivot\_df.groupby(['Media Type', 'Campaign'])['Impressions'].sum().reset\_index()

top\_3\_campaigns\_per\_media\_type = top\_3\_campaigns\_per\_media\_type.groupby('Media Type').apply(lambda x: x.nlargest(3, 'Impressions')).reset\_index(drop=True)

top\_3\_campaigns\_per\_media\_type

A screenshot of a phone

AI-generated content may be incorrect.

* Does the campaign with maximum impressions also have maximum clicks?

Ans:

max\_impressions\_campaign = pivot\_df.groupby('Campaign')['Impressions'].sum().idxmax()

max\_clicks\_campaign = pivot\_df.groupby('Campaign')['Clicks'].sum().idxmax()

max\_impressions\_equals\_max\_clicks = max\_impressions\_campaign == max\_clicks\_campaign

max\_impressions\_equals\_max\_clicks

A black and orange line

AI-generated content may be incorrect.

* Find out the top 3 campaigns in terms of CPM (CPM stands for cost per mile which means cost per 1000 impressions. CPM = total media cost/total impressions \* 1000)

Ans:

pivot\_df['CPM'] = pivot\_df['Media Cost'] / pivot\_df['Impressions'] \* 1000

top\_3\_campaigns\_by\_cpm = pivot\_df.groupby('Campaign')['CPM'].mean().nlargest(3)

top\_3\_campaigns\_by\_cpm

A screenshot of a computer screen

AI-generated content may be incorrect.

* What is the monthly average media cost for different media types?

Ans:

monthly\_avg\_media\_cost = pivot\_df.groupby(['Month', 'Media Type'])['Media Cost'].mean().unstack()

monthly\_avg\_media\_cost

A screenshot of a computer

AI-generated content may be incorrect.

* Which media type is more expensive in terms of CPM?

Ans:

media\_type\_cpm = pivot\_df.groupby('Media Type')['CPM'].mean().idxmax()

media\_type\_cpm



* When would you use a Pie chart, a Bar chart and a Line Chart? Use the given dataset and provide an example for each type of chart.

Ans:

**Pie chart** is used when wewant to show **proportions or percentages** **of a whole.**

**Bar chart** is used when we want to **compare quantities** **across different categories.**

**Line Chart** is used when we want to show **trends over time**.

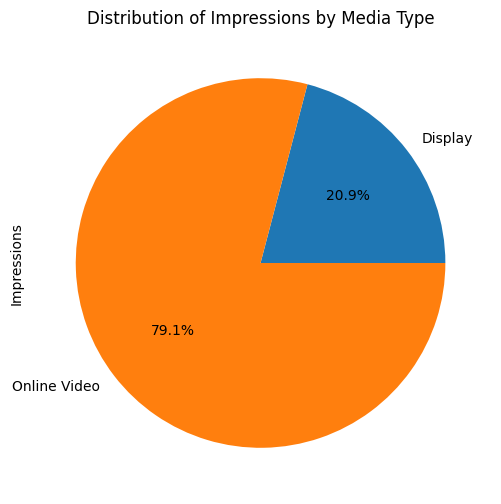
impressions\_by\_media\_type = pivot\_df.groupby('Media Type')['Impressions'].sum()

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

impressions\_by\_media\_type.plot.pie(autopct='%1.1f%%')

plt.title('Distribution of Impressions by Media Type')

plt.show()



plt.figure(figsize=(10, 6))

top\_5\_months.plot(kind='bar')

plt.title('Top 5 Months by Impressions')

plt.xlabel('Month')

plt.ylabel('Impressions')

plt.show()

A graph of blue bars

AI-generated content may be incorrect.

plt.figure(figsize=(12, 8))

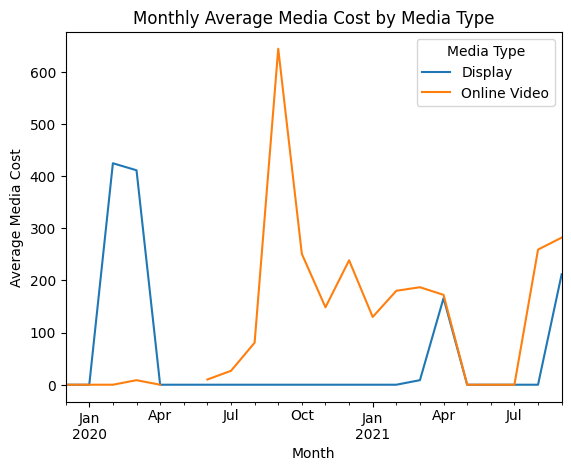
monthly\_avg\_media\_cost.plot(kind='line')

plt.title('Monthly Average Media Cost by Media Type')

plt.xlabel('Month')

plt.ylabel('Average Media Cost')

plt.show()



* **Modelling Data:**
* Analyze Sales of Altima using a Linear Regression model.

Ans: Here we are using a multiple linear regression model to analyze the sales of Altima using Least Squares Method. We are trying to predict it based on other predictor variables.

* Which is your dependent variable?

Ans: Here our dependent variable is sales.

* Which variables are significant in the model?

Ans: Gas\_price(p-value=0.000), discount(p-value =0.000), youtube(p-value =0.029) and print(p-value =0.019).

* How did you evaluate them?

Ans:

import pandas as pd

import statsmodels.api as sm

from sklearn.metrics import mean\_absolute\_percentage\_error

from statsmodels.stats.outliers\_influence import variance\_inflation\_factor

# Load the dataset

df = pd.read\_excel('Modeling dataset.xlsx', sheet\_name='modeling dataset', engine='openpyxl')

# Define the dependent variable (Sales) and independent variables

df['week index']=pd.to\_datetime(df['month']).dt.month

df["Seasonality"] = df.groupby(['week index'])["sales"].transform('mean')/df['sales'].mean()

X = df[['gas\_price', 'discount', 'TV', 'YouTube', 'Radio', 'Print',"Seasonality"]]

y = df['sales']

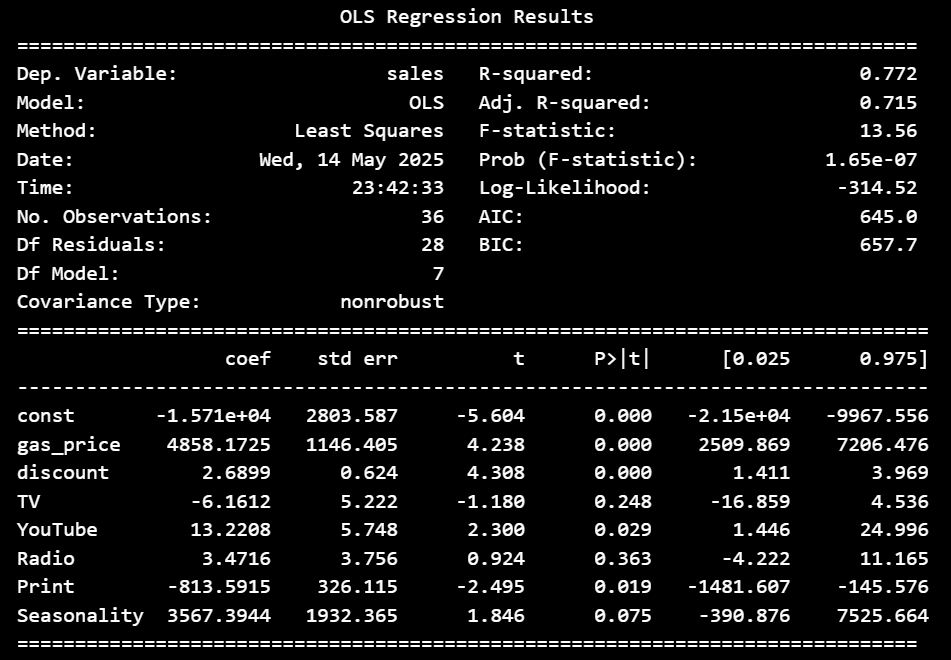
X = sm.add\_constant(X)

# Fit the linear regression model

model = sm.OLS(y, X).fit()

model\_summary = model.summary()

print(model\_summary)



From this table, we can observe the p values(P>|t|) of t-test, which is the test of significance of the beta coefficients. If p-value<0.05, we reject the null hypothesis that beta coefficient=0,that means the variable is significant in predicting sales.

* What is the R2 and Adjusted R2 of the model?

Ans:

r\_squared = model.rsquared

adj\_r\_squared = model.rsquared\_adj

r\_squared, adj\_r\_squared



* Do you observe multicollinearity in the independent variables?

Ans:

# Calculating VIF to check for multicollinearity

vif\_data = pd.DataFrame()

vif\_data['feature'] = X.columns

vif\_data['VIF'] = [variance\_inflation\_factor(X.values, i) for i in range(len(X.columns))]

vif\_data



All the VIF values of the predictor variables are below 5, indicating **no serious multicollinearity**.

* Calculate the MAPE of the model.

Ans:

y\_pred = model.predict(X)

mape = mean\_absolute\_percentage\_error(y, y\_pred)

mape

