```python
from google.colab import files
uploaded = files.upload()
```

Choose files   No file chosen   Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```python
import shutil

# Make a copy of the CSV file with a new name
shutil.copy('Bank_Marketing_Inspection_test.csv', 'Bank_Marketing_Inspection_project.csv')
```

    'Bank_Marketing_Inspection_project.csv'

```python
import pandas as pd

# Load the new CSV file
df = pd.read_csv('Bank_Marketing_Inspection_project.csv')

# Display the first few rows
df.head()
```

	age	job	marital	education	default	housing	loan	contact	month	day_of_week	...	campaign	pdays	previous	poutcome
0	29	admin.	married	tertiary	no	yes	no	cellular	may	mon	...	1	999	0	nonexisten
1	35	technician	single	secondary	no	no	yes	telephone	jun	tue	...	2	3	1	success

2 rows × 21 columns

```python
import seaborn as sns
import matplotlib.pyplot as plt

# Select only numerical columns
numeric_df = df.select_dtypes(include=['int64', 'float64'])

# Compute the correlation matrix
correlation_matrix = numeric_df.corr()

# Plot the heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix for Numerical Features')
plt.show()
```
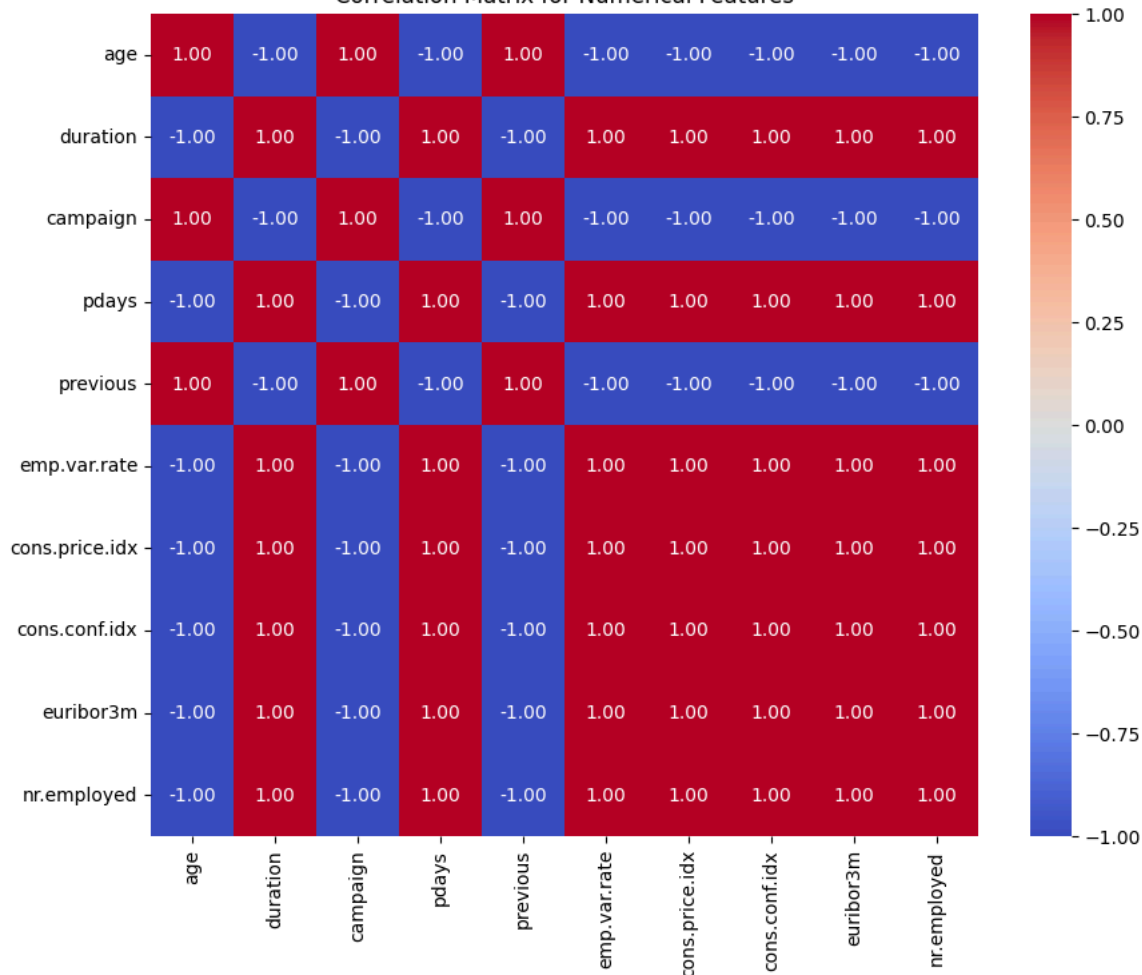
Correlation Matrix for Numerical Features

	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
age	1.00	-1.00	1.00	-1.00	1.00	-1.00	-1.00	-1.00	-1.00	-1.00
duration	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
campaign	1.00	-1.00	1.00	-1.00	1.00	-1.00	-1.00	-1.00	-1.00	-1.00
pdays	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
previous	1.00	-1.00	1.00	-1.00	1.00	-1.00	-1.00	-1.00	-1.00	-1.00
emp.var.rate	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
cons.price.idx	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
cons.conf.idx	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
euribor3m	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00
nr.employed	-1.00	1.00	-1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00

```python
# Check for nulls
print("Missing Values:\n", df.isnull().sum())

# Check for duplicates
print("\nDuplicate Rows:", df.duplicated().sum())

# Data types
print("\nData Types:\n", df.dtypes)
```

```
Missing Values:
 age             0
job             0
marital         0
education       0
default         0
housing         0
loan            0
contact         0
month           0
day_of_week     0
duration        0
campaign        0
pdays           0
previous        0
poutcome        0
emp.var.rate    0
cons.price.idx  0
cons.conf.idx   0
euribor3m       0
nr.employed     0
y               0
dtype: int64

Duplicate Rows: 0

Data Types:
 age              int64
job             object
marital         object
education       object
default         object
housing         object
```

```
loan              object
contact           object
month             object
day_of_week       object
duration           int64
campaign           int64
pdays              int64
previous           int64
poutcome          object
emp.var.rate     float64
cons.price.idx   float64
cons.conf.idx    float64
euribor3m        float64
nr.employed        int64
y                 object
dtype: object
```

```python
sns.countplot(x='job', hue='y', data=df)
plt.title('Subscription Rate by Job Type')
plt.xticks(rotation=45)
plt.show()
```
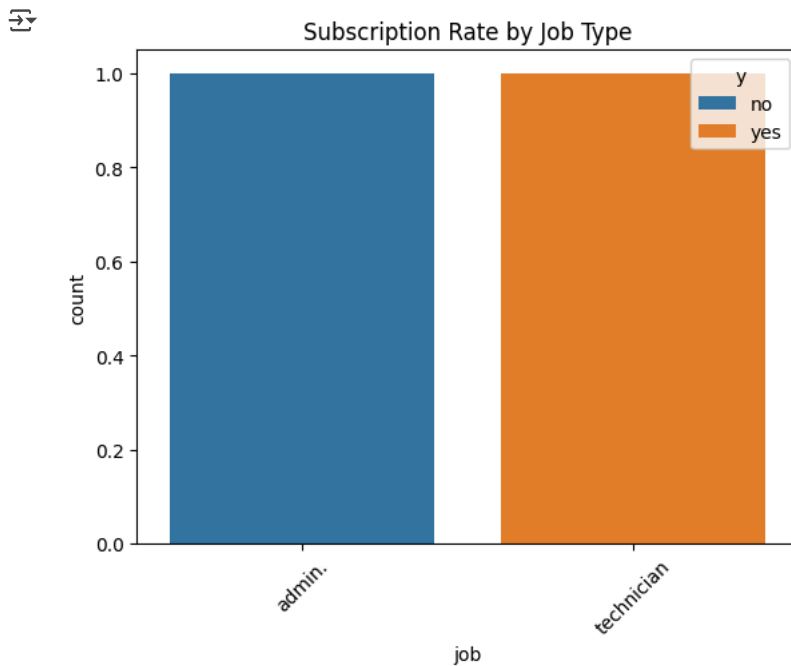


```python
sns.countplot(x='education', hue='y', data=df)
plt.title('Subscription Rate by Education')
plt.xticks(rotation=45)
plt.show()
```

```
sns.histplot(data=df, x='age', bins=20, kde=True)
plt.title('Age Distribution of Clients')
plt.show()
```


Age Distribution of Clients