7114056078 HW2 — Bank Marketing Regression

本筆記本將進行資料載入與基礎視覺化。

- 請將 Kaggle 下載的資料檔命名為 `bank.csv` 並放在同資料夾。
- 本專案支援自動偵測分隔符（`,` 或 `;`）以避免載入失敗。

In [1]:
```
# Environment setup
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

sns.set(style='whitegrid', palette='muted')
plt.rcParams['figure.figsize'] = (10, 6)
```

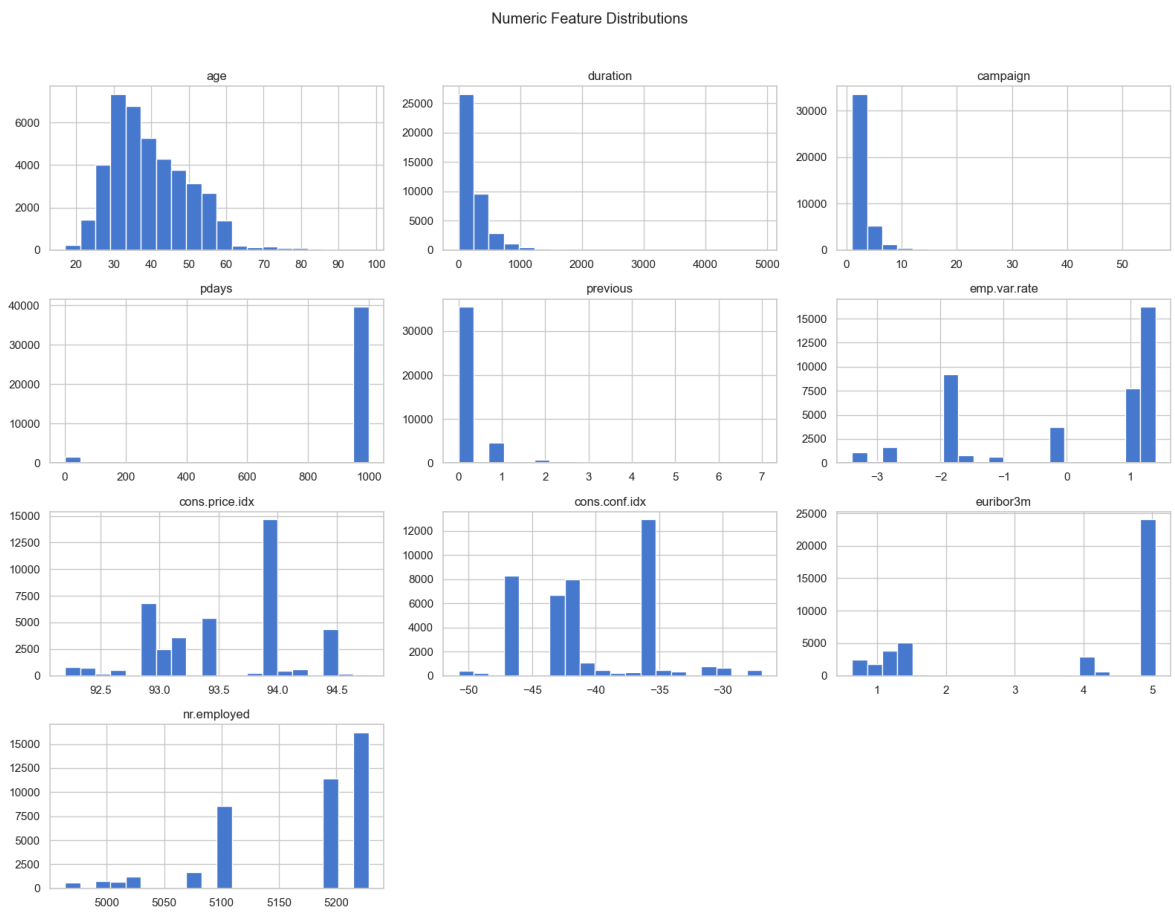
In [2]:
```
# Robust loader: auto-detect delimiter and handle common issues
from pathlib import Path

data_path = Path('bank.csv')
if not data_path.exists():
    raise FileNotFoundError("找不到 bank.csv，請將 Kaggle 檔案放到此資料夾並命名為

# Try pandas' python engine sniffing first, then explicit fallbacks
try:
    df = pd.read_csv(data_path, sep=None, engine='python')
except Exception:
    try:
        df = pd.read_csv(data_path, sep=';')
    except Exception:
        df = pd.read_csv(data_path)

# If only 1 column is read (common when delimiter mismatches), retry with ';'
if df.shape[1] == 1:
    try:
        df = pd.read_csv(data_path, sep=';')
    except Exception:
        pass

display(df.head())
print(f'Shape: {df.shape}')
```

	age	job	marital	education	default	housing	loan	contact	month
0	56	housemaid	married	basic.4y	no	no	no	telephone	may
1	57	services	married	high.school	unknown	no	no	telephone	may
2	37	services	married	high.school	no	yes	no	telephone	may
3	40	admin.	married	basic.6y	no	no	no	telephone	may
4	56	services	married	high.school	no	no	yes	telephone	may

5 rows × 21 columns

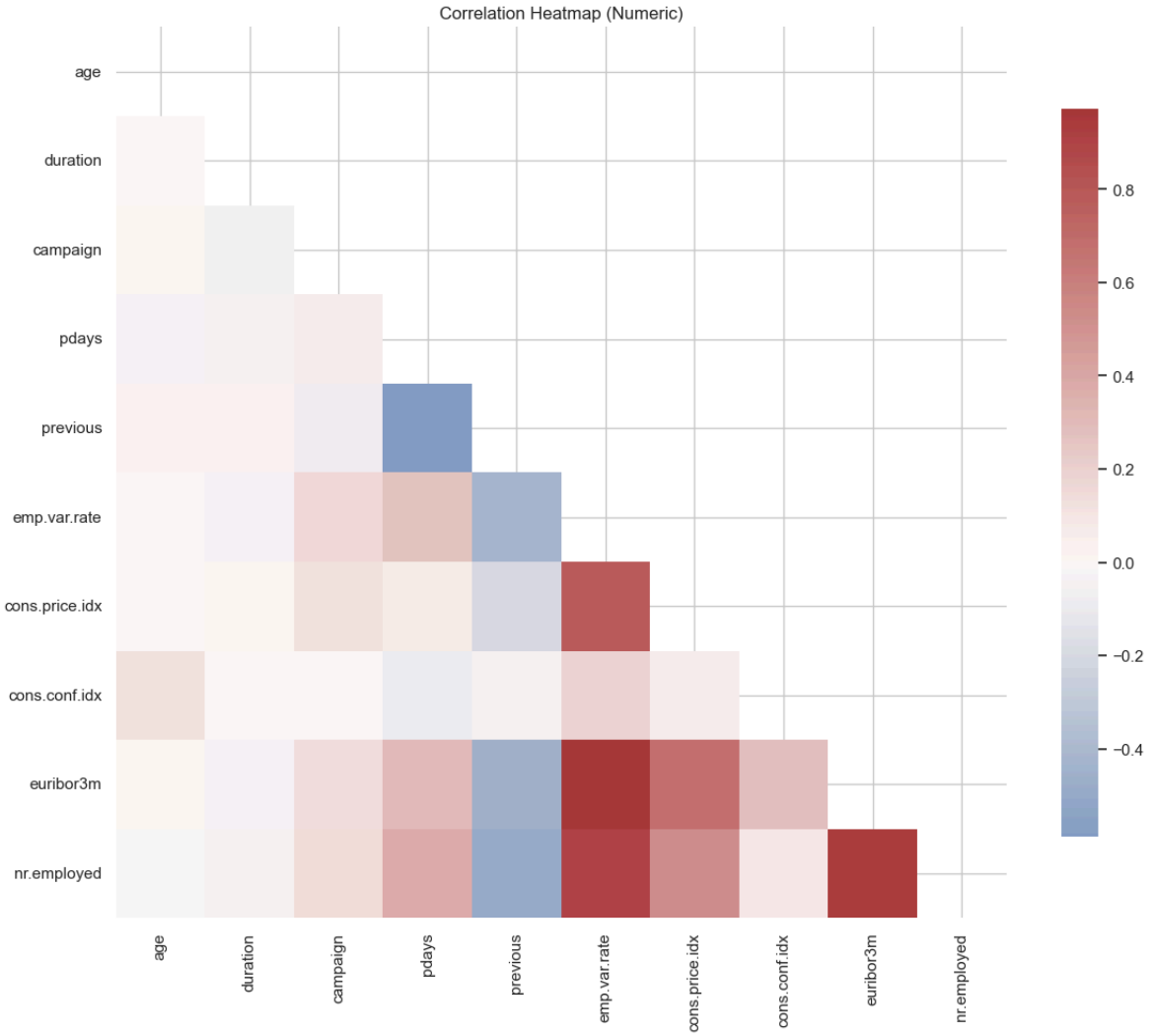
```
Shape: (41188, 21)
```

In [3]:
```python
# Numeric distributions
numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()
if numeric_cols:
    ax = df[numeric_cols].hist(bins=20, figsize=(16, 12))
    plt.suptitle('Numeric Feature Distributions', y=1.02)
    plt.tight_layout()
else:
    print('沒有偵測到數值欄位，請確認資料是否正確載入（分隔符可能錯誤）。')
```



In [4]:
```python
# Correlation heatmap for numeric columns
if len(numeric_cols) >= 2:
    corr = df[numeric_cols].corr(numeric_only=True)
    mask = np.triu(np.ones_like(corr, dtype=bool))
    plt.figure(figsize=(12, 10))
    sns.heatmap(corr, mask=mask, cmap='vlag', center=0, annot=False, square=True
    plt.title('Correlation Heatmap (Numeric)')
```

```
    plt.tight_layout()
else:
    print('數值欄位不足以計算相關係數。')
```



In [5]:
```
# Categorical count plots (optionally against target if present)
cat_cols = df.select_dtypes(include=['object', 'category', 'bool']).columns.toli
target = 'y' if 'y' in df.columns else ('deposit' if 'deposit' in df.columns els
# Avoid plotting the target itself as feature
features = [c for c in cat_cols if c != target]

max_feats = 6  # avoid too many plots
plot_cols = features[:max_feats]
if plot_cols:
    n = len(plot_cols)
    rows = int(np.ceil(n / 2))
    fig, axes = plt.subplots(rows, 2, figsize=(16, 5*rows))
    axes = axes.flatten() if n > 1 else [axes]
    for i, col in enumerate(plot_cols):
        order = df[col].value_counts().index
        if target:
            sns.countplot(data=df, x=col, hue=target, order=order, ax=axes[i])
        else:
            sns.countplot(data=df, x=col, order=order, ax=axes[i])
        axes[i].set_title(f'{col} count')
        axes[i].tick_params(axis='x', rotation=30)
    for j in range(i+1, len(axes)):
        fig.delaxes(axes[j])
    plt.tight_layout()
```

```
else:
    print('沒有可視覺化的類別欄位。')
```