In [11]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

In [12]:
```
sns.set(style="whitegrid", font_scale=1.1)

df = pd.read_excel("MMM_data.xlsx")
df["DATE"] = pd.to_datetime(df["DATE"], errors="coerce")
df = df.sort_values("DATE").set_index("DATE")

print("Shape:", df.shape)
print("Date range:", df.index.min(), "→", df.index.max())

data_dict = pd.DataFrame([
    ["DATE", "Index", "Calendar date of observation", "Day"],
    ["SALES ($)", "Target", "Total sales revenue", "USD"],
    ["DEMAND ", "Secondary target", "Units sold (quantity of product sold)",
    ["Advertising Expenses (SMS)", "Driver", "Ad spend in SMS channel", "USD
    ["Advertising Expenses (Newspaper)", "Driver", "Ad spend in Newspaper ch
    ["Advertising Expenses (Radio)", "Driver", "Ad spend in Radio channel",
    ["Advertising Expenses (TV)", "Driver", "Ad spend in TV channel", "USD"]
    ["Advertising Expenses (Internet)", "Driver", "Ad spend in Internet char
    ["GRP (SMS)", "Driver", "Gross Rating Points for SMS ads", "GRP points"]
    ["GRP (Newspaper)", "Driver", "Gross Rating Points for Newspaper ads", "
    ["GRP (Radio)", "Driver", "Gross Rating Points for Radio ads", "GRP poir
    ["GRP (TV)", "Driver", "Gross Rating Points for TV ads", "GRP points"],
    ["GRP (Internet)", "Driver", "Gross Rating Points for Internet ads", "GR
    ["CPI", "Control (macro)", "Consumer Price Index (inflation indicator)",
    ["CCI", "Control (macro)", "Consumer Confidence Index", "Index"],
    ["PPI", "Control (macro)", "Producer Price Index", "Index"],
    ["Unit Price ($)", "Control", "Average selling price per unit", "USD)"],
    ["POS/ Supply Data", "Control", "Supply-side data", "Units"]
], columns=["Variable", "Role", "Description", "Unit"])

display(data_dict)
```

```
Shape: (2613, 18)
Date range: 2010-01-01 00:00:00 → 2017-02-25 00:00:00
```

	Variable	Role	Description	Unit
0	DATE	Index	Calendar date of observation	Day
1	SALES ($)	Target	Total sales revenue	USD
2	DEMAND	Secondary target	Units sold (quantity of product sold)	Units
3	Advertising Expenses (SMS)	Driver	Ad spend in SMS channel	USD
4	Advertising Expenses (Newspaper)	Driver	Ad spend in Newspaper channel	USD
5	Advertising Expenses (Radio)	Driver	Ad spend in Radio channel	USD
6	Advertising Expenses (TV)	Driver	Ad spend in TV channel	USD
7	Advertising Expenses (Internet)	Driver	Ad spend in Internet channel	USD
8	GRP (SMS)	Driver	Gross Rating Points for SMS ads	GRP points
9	GRP (Newspaper)	Driver	Gross Rating Points for Newspaper ads	GRP points
10	GRP (Radio)	Driver	Gross Rating Points for Radio ads	GRP points
11	GRP (TV)	Driver	Gross Rating Points for TV ads	GRP points
12	GRP (Internet)	Driver	Gross Rating Points for Internet ads	GRP points
13	CPI	Control (macro)	Consumer Price Index (inflation indicator)	Index
14	CCI	Control (macro)	Consumer Confidence Index	Index
15	PPI	Control (macro)	Producer Price Index	Index
16	Unit Price ($)	Control	Average selling price per unit	USD)
17	POS/ Supply Data	Control	Supply-side data	Units

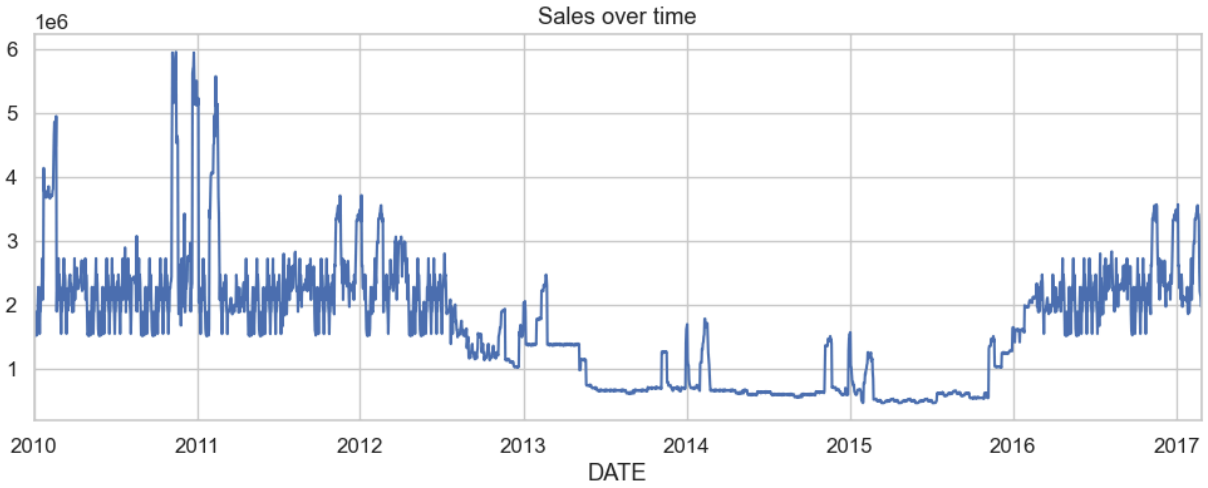
In [13]:
```
print("\nMissing values:")
print(df.isna().sum())
print("\nStatistics:")
print(df.describe().T)
```

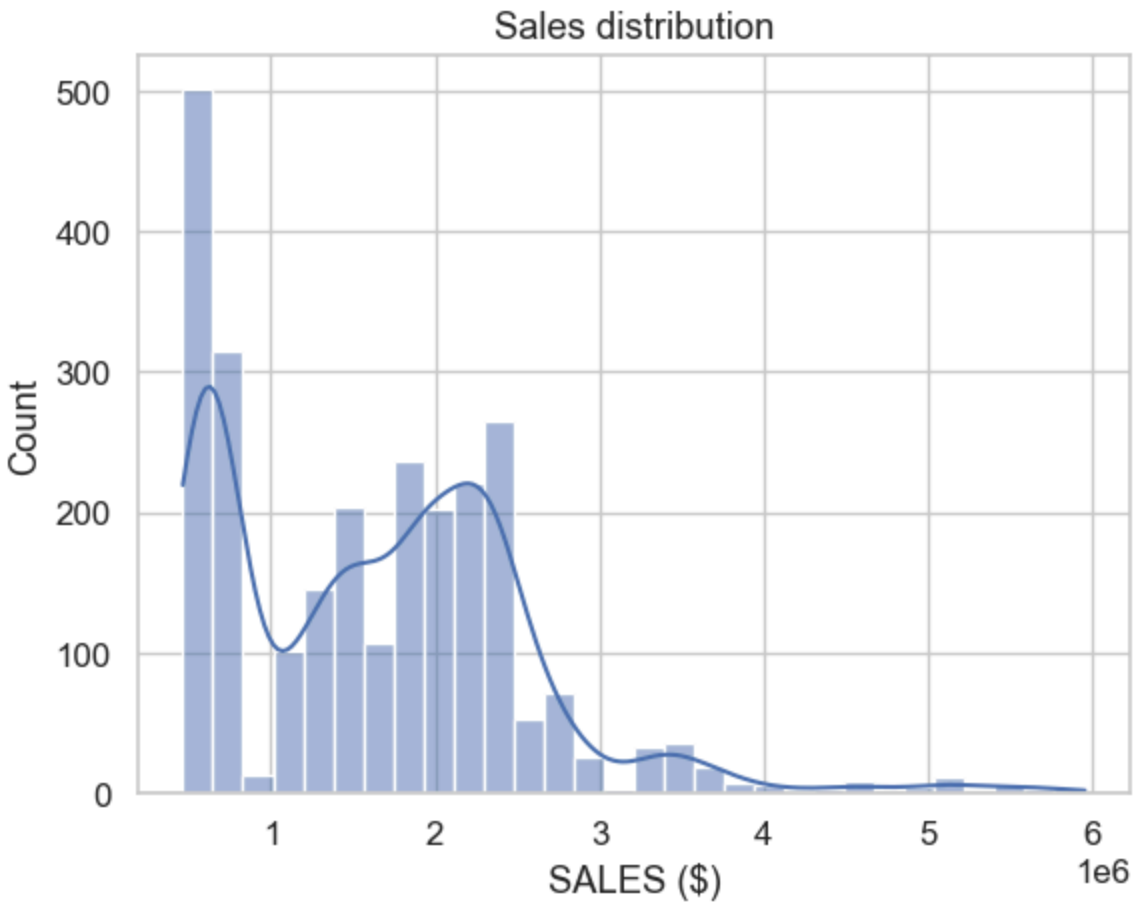
```
Missing values:
TV Manufacturing Brand                0
DEMAND                                0
Consumer Price Index (CPI)            0
Consumer Confidence Index(CCI)        0
Producer Price Index (PPI)            0
Unit Price ($)                        0
POS/ Supply Data                      0
SALES ($)                             0
Advertising Expenses (SMS)            0
Advertising Expenses(Newspaper ads)   0
Advertising Expenses(Radio)           0
Advertising Expenses(TV)              0
Advertising Expenses(Internet)        0
GRP (NewPaper ads)                    0
GRP(SMS)                              0
GRP(Radio                             0
GRP(Internet)                         0
GRP(TV)                               0
dtype: int64

Statistics:
                                    count          mean            std  \
DEMAND                             2613.0  5.021434e+03    2681.197808
Consumer Price Index (CPI)         2613.0  1.026100e+02       1.381355
Consumer Confidence Index(CCI)     2613.0  1.031523e+02       3.167221
Producer Price Index (PPI)         2613.0  1.022382e+02       2.037774
Unit Price ($)                     2613.0  3.632751e+02      26.372958
POS/ Supply Data                   2613.0  4.522971e+03    2603.996535
SALES ($)                          2613.0  1.641507e+06  941667.283659
Advertising Expenses (SMS)         2613.0  6.038850e+01      13.557190
Advertising Expenses(Newspaper ads)2613.0  1.265118e+01       1.117926
Advertising Expenses(Radio)        2613.0  8.807464e+01      12.569956
Advertising Expenses(TV)           2613.0  1.324501e+03     123.677327
Advertising Expenses(Internet)     2613.0  3.079184e+03    1520.891014
GRP (NewPaper ads)                 2613.0  5.052969e+02     488.147782
GRP(SMS)                           2613.0  3.061820e+01      31.570754
GRP(Radio                          2613.0  1.394373e+02     146.042432
GRP(Internet)                      2613.0  2.862287e+02     138.406620
GRP(TV)                            2613.0  1.146114e+03     822.991163

                                          min            25%  \
DEMAND                           1610.000000    2436.000000
Consumer Price Index (CPI)        101.300000     101.400000
Consumer Confidence Index(CCI)     96.300000     102.700000
Producer Price Index (PPI)         99.500000     100.400000
Unit Price ($)                    282.140000     361.600000
POS/ Supply Data                 1510.000000    1776.000000
SALES ($)                      462709.600000  671767.900000
Advertising Expenses (SMS)         37.916700      47.554100
Advertising Expenses(Newspaper ads)10.027128      11.885714
Advertising Expenses(Radio)        62.968800      78.098100
Advertising Expenses(TV)         1067.155700    1251.250000
Advertising Expenses(Internet)      0.000000    2226.429000
GRP (NewPaper ads)                  5.659000     114.957000
GRP(SMS)                            0.000000       0.000000
```

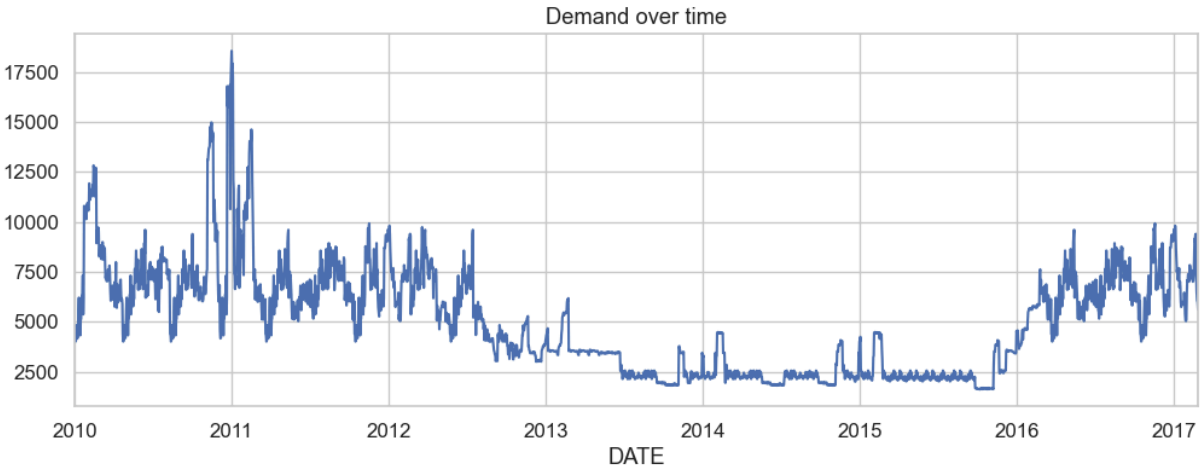
```
GRP(Radio                                66.863600      95.136400
GRP(Internet)                           191.181800     234.954500
GRP(TV)                                 697.636400     849.545500

                                              50%           75%           ma
x
DEMAND                                4.636000e+03  6.834000e+03  1.856500e+0
4
Consumer Price Index (CPI)            1.023000e+02  1.034000e+02  1.065000e+0
2
Consumer Confidence Index(CCI)        1.036000e+02  1.046000e+02  1.079000e+0
2
Producer Price Index (PPI)            1.027000e+02  1.035000e+02  1.072000e+0
2
Unit Price ($)                        3.616200e+02  3.616200e+02  4.001000e+0
2
POS/ Supply Data                      4.412000e+03  6.266000e+03  1.648200e+0
4
SALES ($)                             1.605095e+06  2.267206e+06  5.960221e+0
6
Advertising Expenses (SMS)            6.126790e+01  7.101220e+01  8.972830e+0
1
Advertising Expenses(Newspaper ads)   1.318658e+01  1.343748e+01  1.410419e+0
1
Advertising Expenses(Radio)           8.412920e+01  9.863900e+01  1.184677e+0
2
Advertising Expenses(TV)              1.380696e+03  1.416171e+03  1.479457e+0
3
Advertising Expenses(Internet)        3.302667e+03  4.237095e+03  6.354571e+0
3
GRP (NewPaper ads)                    2.215280e+02  8.543100e+02  1.791183e+0
3
GRP(SMS)                              2.611960e+01  4.678200e+01  1.449951e+0
2
GRP(Radio                             1.096364e+02  1.260455e+02  1.169409e+0
3
GRP(Internet)                         2.614545e+02  2.913182e+02  1.540429e+0
3
GRP(TV)                               9.285455e+02  1.114864e+03  7.307318e+0
3
```

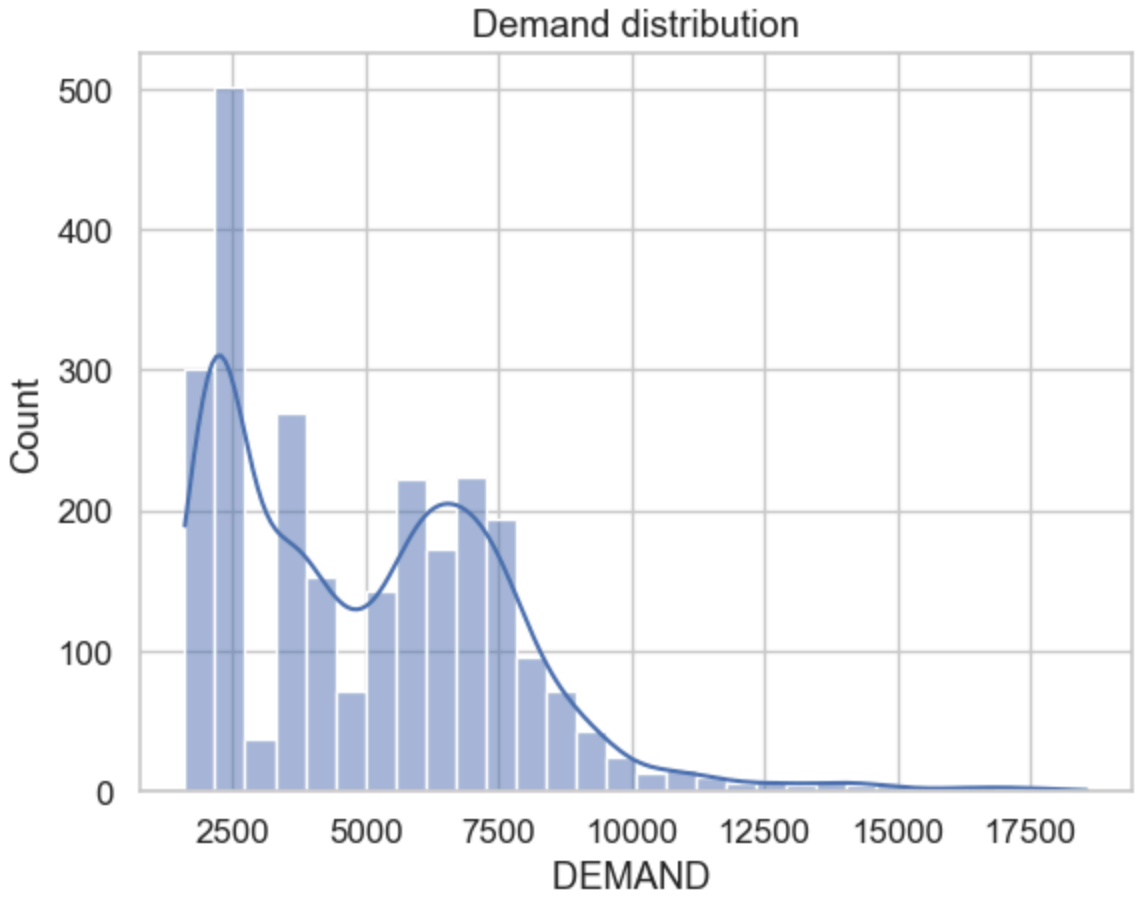
In [14]:
```
# Visualizations
plt.figure(figsize=(12,4))
df["SALES ($)"].plot()
plt.title("Sales over time")
plt.show()
sns.histplot(df["SALES ($)"].dropna(), bins=30, kde=True)
plt.title("Sales distribution")
plt.show()

plt.figure(figsize=(12,4))
df["DEMAND "].plot()
plt.title("Demand over time")
plt.show()
sns.histplot(df["DEMAND "].dropna(), bins=30, kde=True)
```

```
plt.title("Demand distribution")
plt.show()
```



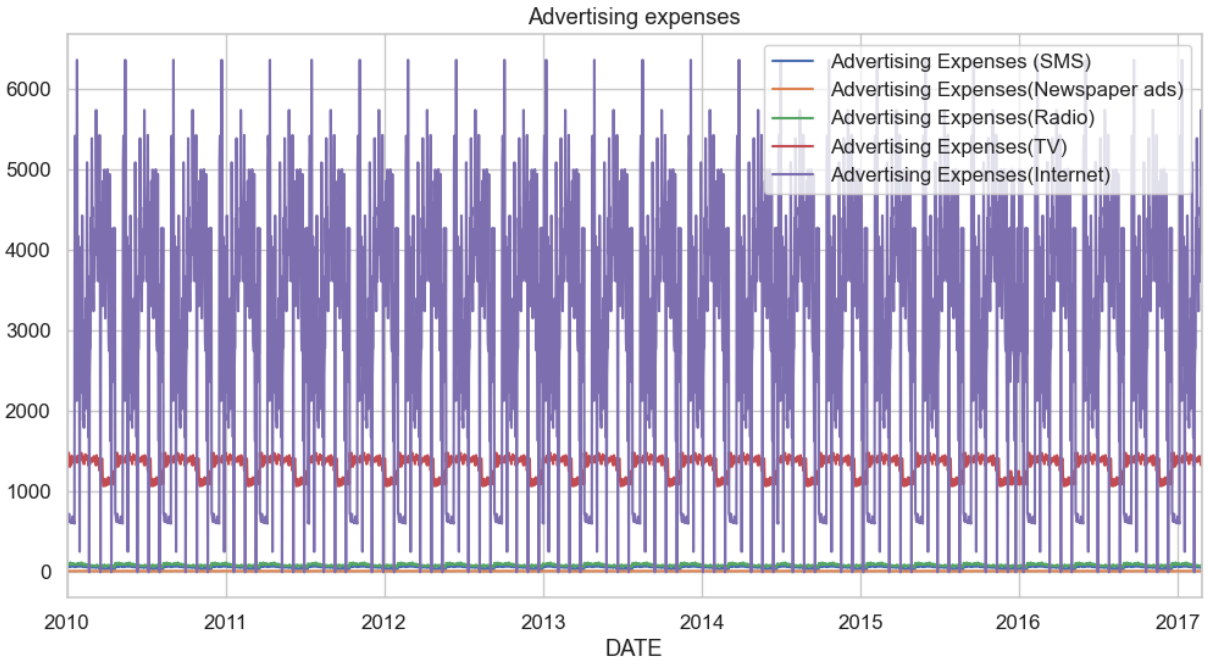


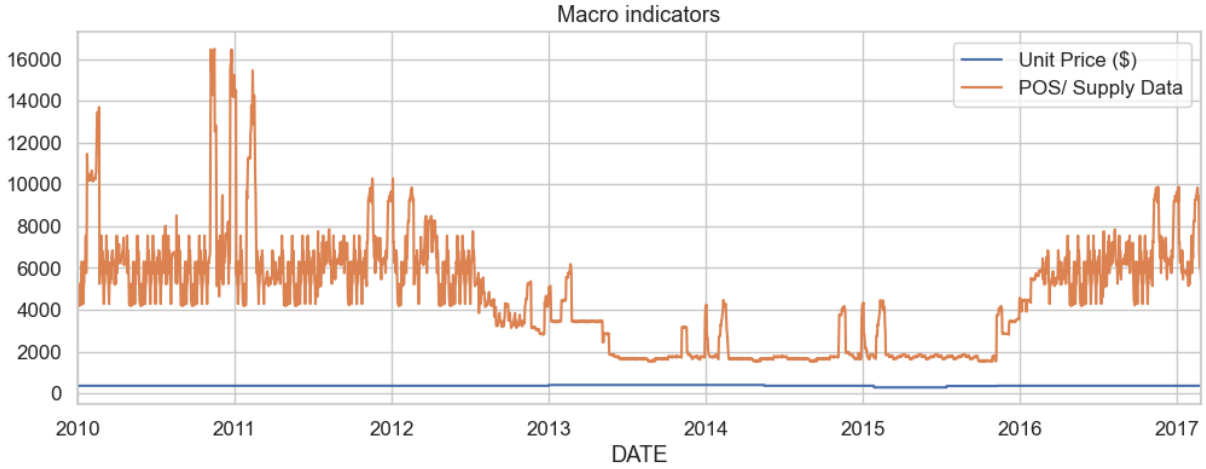




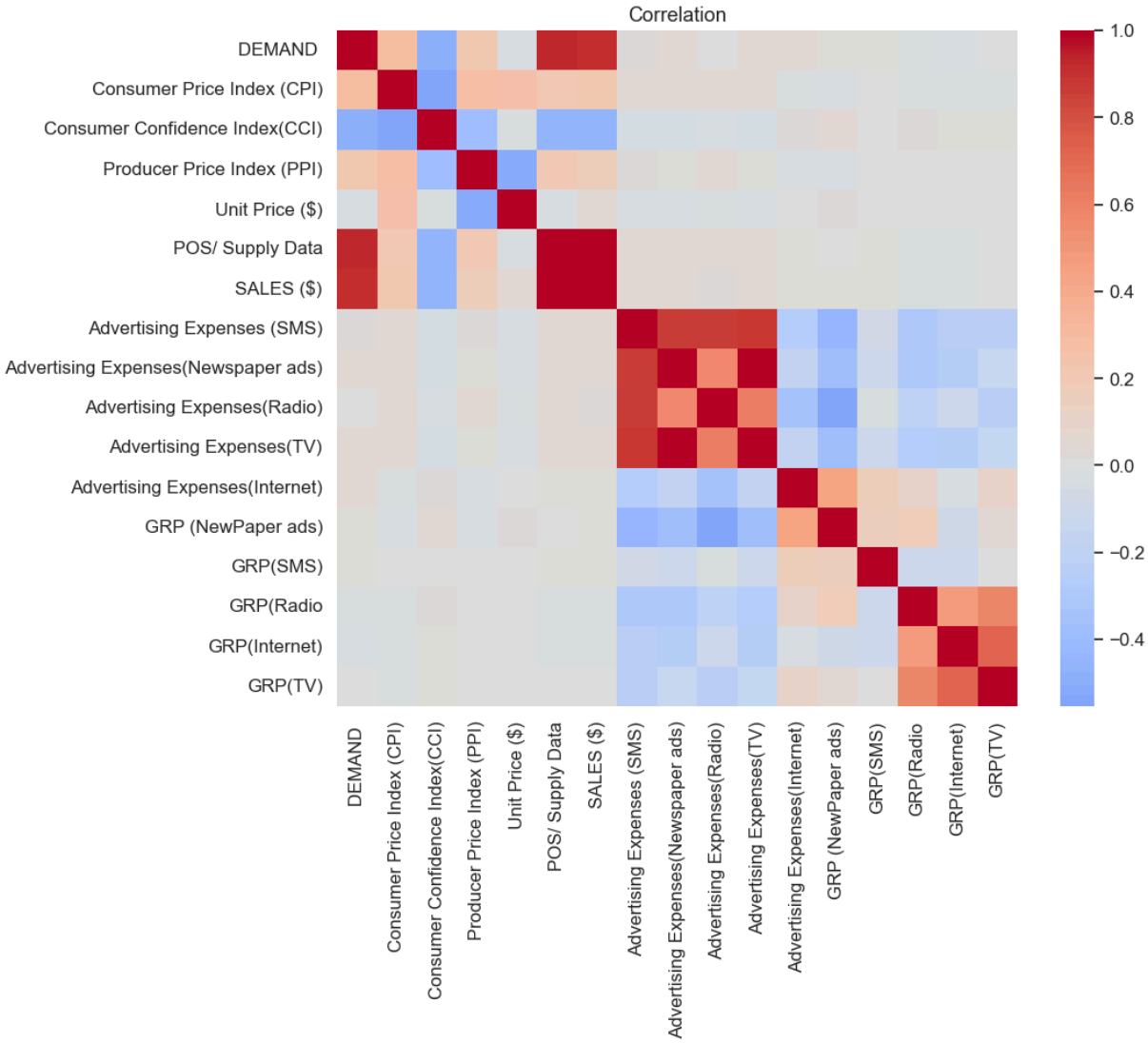
In [15]:
```
# Visualize drivers and controls
ad = [c for c in df.columns if "Expenses" in c] #
grp = [c for c in df.columns if "GRP" in c]
macro = [c for c in df.columns if c in ["CPI","CCI","PPI","Unit Price ($)","

df[ad].plot(figsize=(12,6))
plt.title("Advertising expenses")
plt.show()

df[grp].plot(figsize=(12,6))
plt.title("GRPs")
plt.show()
```

```
df[macro].plot(figsize=(12,4))
plt.title("Macro indicators")
plt.show()
```



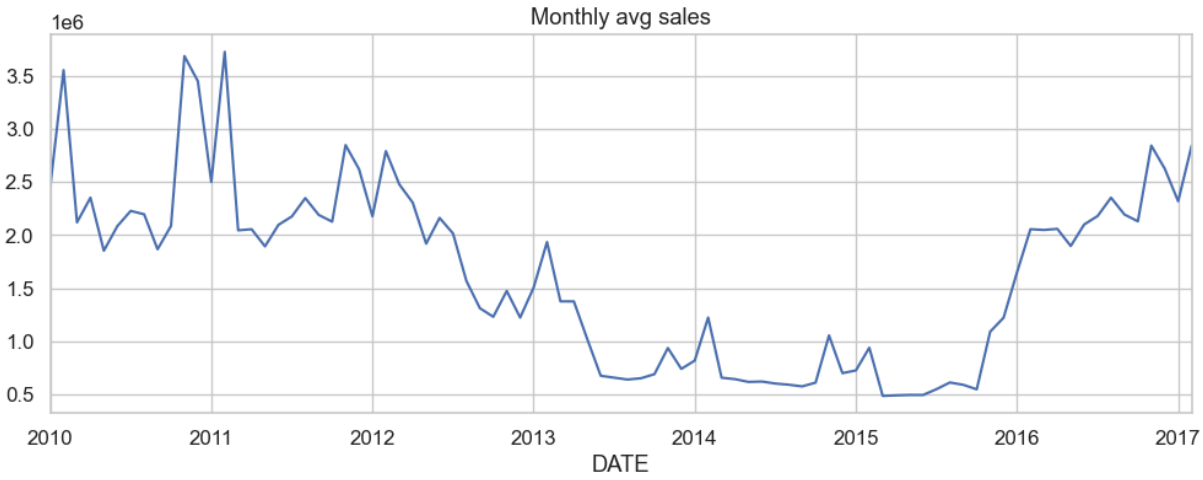


In [16]:
```
plt.figure(figsize=(10,8))
numeric_df = df.select_dtypes(include=[np.number])
sns.heatmap(numeric_df.corr(), cmap="coolwarm", center=0)
plt.title("Correlation")
plt.show()
```

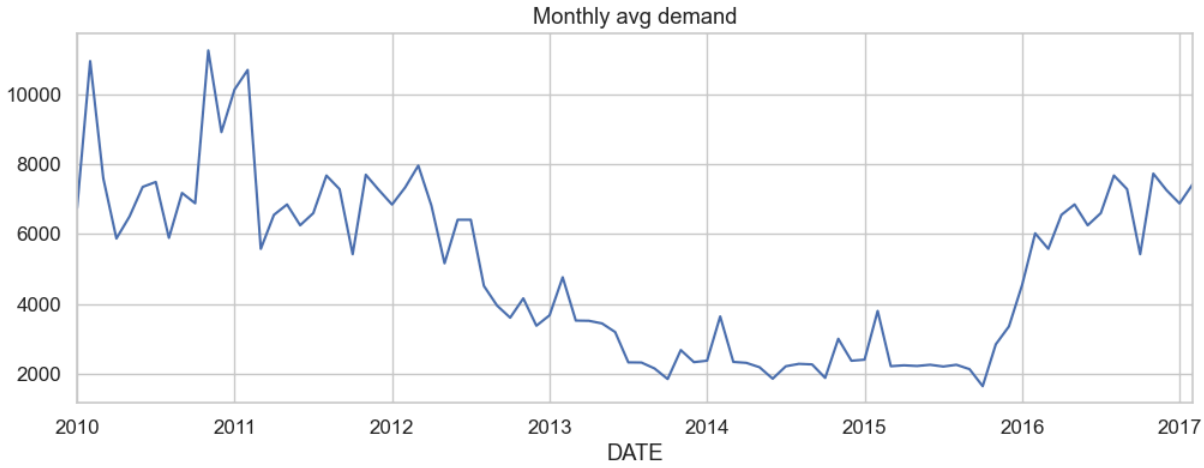


In [17]:
```
plt.figure(figsize=(12,4))
df["SALES ($)"].resample("M").mean().plot()
plt.title("Monthly avg sales")
plt.show()

plt.figure(figsize=(12,4))
df["DEMAND "].resample("M").mean().plot()
plt.title("Monthly avg demand")
plt.show()
```

```
/var/folders/k8/kp1njk3j6ms3bxl4lxwr6xth0000gn/T/ipykernel_25525/2063120492.
py:2: FutureWarning: 'M' is deprecated and will be removed in a future versi
on, please use 'ME' instead.
  df["SALES ($)"].resample("M").mean().plot()
```




```
/var/folders/k8/kp1njk3j6ms3bxl4lxwr6xth0000gn/T/ipykernel_25525/2063120492.
py:7: FutureWarning: 'M' is deprecated and will be removed in a future versi
on, please use 'ME' instead.
  df["DEMAND "].resample("M").mean().plot()
```



In [18]:
```
X = numeric_df.drop(columns=["SALES ($)", "DEMAND "]).fillna(0)
vif = pd.DataFrame({
    "Variable": X.columns,
    "VIF": [variance_inflation_factor(X.values, i) for i in range(X.shape[1]
})
print("\nVariance Inflation Factors:")
print(vif.sort_values("VIF", ascending=False))
```

Variance Inflation Factors:

	Variable	VIF
6	Advertising Expenses(Newspaper ads)	6443.287977
0	Consumer Price Index (CPI)	5947.015317
8	Advertising Expenses(TV)	5880.234554
2	Producer Price Index (PPI)	4099.230218
1	Consumer Confidence Index(CCI)	869.982017
7	Advertising Expenses(Radio)	389.713010
5	Advertising Expenses (SMS)	364.742946
3	Unit Price ($)	344.530661
13	GRP(Internet)	13.497084
14	GRP(TV)	8.400486
9	Advertising Expenses(Internet)	6.683588
4	POS/ Supply Data	5.064756
10	GRP (NewPaper ads)	3.665240
12	GRP(Radio	3.266578
11	GRP(SMS)	2.203121