**The final code**

**Group members:**

**Team 38**

|  |  |  |
| --- | --- | --- |
|  | Name | ID |
| 1 | **Mohamed Yousri Ibraheem Abdallah Ibraheem** | **20221509866** |
| 2 | **Abdulrahman Salah Anwar Abdo** | **20221458503** |
| 3 | **Abu-Bakr Mohamed Mahmoud Ragab** | **20221458962** |

In [43]:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

data = pd.read\_csv("Data.csv")

ploted\_data = pd.read\_csv("Data.csv")

ploted\_data = ploted\_data.drop(['Store'], axis=1)

In [44]:

data['year']=pd.to\_datetime(data['Date'],format="%d-%m-%Y").dt.year

data['month']=pd.to\_datetime(data['Date'],format="%d-%m-%Y").dt.month

data['day']=pd.to\_datetime(data['Date'],format="%d-%m-%Y").dt.day

data['quarter']=pd.to\_datetime(data['Date'],format="%d-%m-%Y").dt.quarter

data['semester']=np.where(data['quarter'].isin([1,2]),1,2)

data

Out[44]:

|  | **Store** | **Date** | **Weekly\_Sales** | **Holiday\_Flag** | **Temperature** | **Fuel\_Price** | **CPI** | **Unemployment** | **year** | **month** | **day** | **quarter** | **semester** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | 05-02-2010 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010 | 2 | 5 | 1 | 1 |
| **1** | 1 | 12-02-2010 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010 | 2 | 12 | 1 | 1 |
| **2** | 1 | 19-02-2010 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010 | 2 | 19 | 1 | 1 |
| **3** | 1 | 26-02-2010 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010 | 2 | 26 | 1 | 1 |
| **4** | 1 | 05-03-2010 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010 | 3 | 5 | 1 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 28-09-2012 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012 | 9 | 28 | 3 | 2 |
| **6431** | 45 | 05-10-2012 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012 | 10 | 5 | 4 | 2 |
| **6432** | 45 | 12-10-2012 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012 | 10 | 12 | 4 | 2 |
| **6433** | 45 | 19-10-2012 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012 | 10 | 19 | 4 | 2 |
| **6434** | 45 | 26-10-2012 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012 | 10 | 26 | 4 | 2 |

6435 rows × 13 columns

In [45]:

df = data.groupby('Store')['Weekly\_Sales'].sum().to\_frame().reset\_index()

Max\_Sales = df[(df['Weekly\_Sales']==max(df['Weekly\_Sales']))]

Max\_Sales

Out[45]:

|  | **Store** | **Weekly\_Sales** |
| --- | --- | --- |
| **19** | 20 | 3.013978e+08 |

In [59]:

df = data.groupby('Store')['Weekly\_Sales'].std().to\_frame().reset\_index()

standard\_deviation = df[(df['Weekly\_Sales']==max(df['Weekly\_Sales']))]

standard\_deviation

Out[59]:

|  | **Store** | **Weekly\_Sales** |
| --- | --- | --- |
| **13** | 14 | 317569.949476 |

In [47]:

non\_Holidays = data[(data['Holiday\_Flag']==0)]

Mean = non\_Holidays.mean()

Mean\_NH = Mean['Weekly\_Sales']

Mean\_NH

C:\Users\Dell\AppData\Local\Temp\ipykernel\_16360\550233871.py:2: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric\_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

Mean = non\_Holidays.mean()

Out[47]:

1041256.3802088564

In [48]:

Holidays = data[(data['Holiday\_Flag']==1)]

Holidays\_Bigger\_Than\_Mean =Holidays[(Holidays['Weekly\_Sales']>Mean\_NH)]

Holidays\_Bigger\_Than\_Mean

Out[48]:

|  | **Store** | **Date** | **Weekly\_Sales** | **Holiday\_Flag** | **Temperature** | **Fuel\_Price** | **CPI** | **Unemployment** | **year** | **month** | **day** | **quarter** | **semester** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 12-02-2010 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010 | 2 | 12 | 1 | 1 |
| **31** | 1 | 10-09-2010 | 1507460.69 | 1 | 78.69 | 2.565 | 211.495190 | 7.787 | 2010 | 9 | 10 | 3 | 2 |
| **42** | 1 | 26-11-2010 | 1955624.11 | 1 | 64.52 | 2.735 | 211.748433 | 7.838 | 2010 | 11 | 26 | 4 | 2 |
| **47** | 1 | 31-12-2010 | 1367320.01 | 1 | 48.43 | 2.943 | 211.404932 | 7.838 | 2010 | 12 | 31 | 4 | 2 |
| **53** | 1 | 11-02-2011 | 1649614.93 | 1 | 36.39 | 3.022 | 212.936705 | 7.742 | 2011 | 2 | 11 | 1 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **5819** | 41 | 30-12-2011 | 1264014.16 | 1 | 34.12 | 3.119 | 196.358610 | 6.759 | 2011 | 12 | 30 | 4 | 2 |
| **5825** | 41 | 10-02-2012 | 1238844.56 | 1 | 22.00 | 3.103 | 196.919506 | 6.589 | 2012 | 2 | 10 | 1 | 1 |
| **5855** | 41 | 07-09-2012 | 1392143.82 | 1 | 67.41 | 3.596 | 198.095048 | 6.432 | 2012 | 9 | 7 | 3 | 2 |
| **6334** | 45 | 26-11-2010 | 1182500.16 | 1 | 46.15 | 3.039 | 182.783277 | 8.724 | 2010 | 11 | 26 | 4 | 2 |
| **6386** | 45 | 25-11-2011 | 1170672.94 | 1 | 48.71 | 3.492 | 188.350400 | 8.523 | 2011 | 11 | 25 | 4 | 2 |

220 rows × 13 columns

In [49]:

profit\_Of\_each\_Month =data.groupby(['year','month'])['Weekly\_Sales'].sum().to\_frame().reset\_index()

profit\_Of\_each\_Month

Out[49]:

|  | **year** | **month** | **Weekly\_Sales** |
| --- | --- | --- | --- |
| **0** | 2010 | 2 | 1.903330e+08 |
| **1** | 2010 | 3 | 1.819198e+08 |
| **2** | 2010 | 4 | 2.314124e+08 |
| **3** | 2010 | 5 | 1.867109e+08 |
| **4** | 2010 | 6 | 1.922462e+08 |
| **5** | 2010 | 7 | 2.325801e+08 |
| **6** | 2010 | 8 | 1.876401e+08 |
| **7** | 2010 | 9 | 1.772679e+08 |
| **8** | 2010 | 10 | 2.171618e+08 |
| **9** | 2010 | 11 | 2.028534e+08 |
| **10** | 2010 | 12 | 2.887605e+08 |
| **11** | 2011 | 1 | 1.637040e+08 |
| **12** | 2011 | 2 | 1.863313e+08 |
| **13** | 2011 | 3 | 1.793564e+08 |
| **14** | 2011 | 4 | 2.265265e+08 |
| **15** | 2011 | 5 | 1.816482e+08 |
| **16** | 2011 | 6 | 1.897734e+08 |
| **17** | 2011 | 7 | 2.299114e+08 |
| **18** | 2011 | 8 | 1.885993e+08 |
| **19** | 2011 | 9 | 2.208477e+08 |
| **20** | 2011 | 10 | 1.832613e+08 |
| **21** | 2011 | 11 | 2.101624e+08 |
| **22** | 2011 | 12 | 2.880781e+08 |
| **23** | 2012 | 1 | 1.688945e+08 |
| **24** | 2012 | 2 | 1.920636e+08 |
| **25** | 2012 | 3 | 2.315097e+08 |
| **26** | 2012 | 4 | 1.889209e+08 |
| **27** | 2012 | 5 | 1.887665e+08 |
| **28** | 2012 | 6 | 2.406103e+08 |
| **29** | 2012 | 7 | 1.875095e+08 |
| **30** | 2012 | 8 | 2.368508e+08 |
| **31** | 2012 | 9 | 1.806455e+08 |
| **32** | 2012 | 10 | 1.843617e+08 |

In [50]:

profit\_Of\_each\_semester =data.groupby(['year','semester'])['Weekly\_Sales'].sum().to\_frame().reset\_index()

profit\_Of\_each\_semester

Out[50]:

|  | **year** | **semester** | **Weekly\_Sales** |
| --- | --- | --- | --- |
| **0** | 2010 | 1 | 9.826223e+08 |
| **1** | 2010 | 2 | 1.306264e+09 |
| **2** | 2011 | 1 | 1.127340e+09 |
| **3** | 2011 | 2 | 1.320860e+09 |
| **4** | 2012 | 1 | 1.210765e+09 |
| **5** | 2012 | 2 | 7.893674e+08 |

In [51]:

Weekly\_sales\_sum =data.groupby('Store')["Weekly\_Sales"].sum().reset\_index()

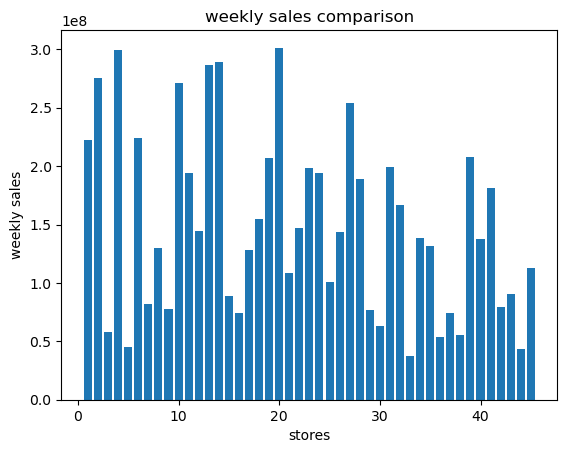
plt.bar(Weekly\_sales\_sum["Store"],Weekly\_sales\_sum["Weekly\_Sales"])

plt.xlabel("stores")

plt.ylabel("weekly sales")

plt.title("weekly sales comparison")

plt.show()

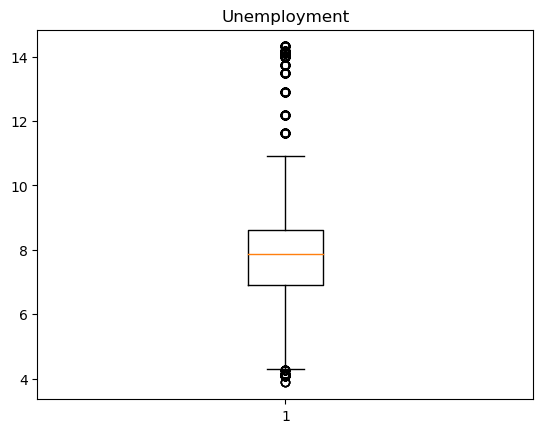


In [52]:

plt.boxplot(data["Unemployment"])

plt.title("Unemployment")

plt.show()

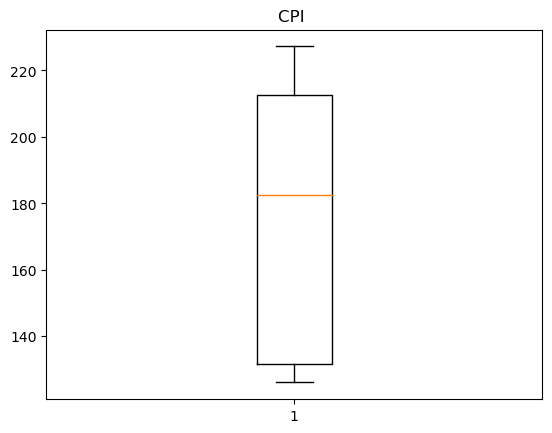


In [53]:

plt.boxplot(data["CPI"])

plt.title("CPI ")

plt.show()



In [54]:

year\_view\_of\_sales=data.groupby(["year"])["Weekly\_Sales"].sum().reset\_index()

year\_view\_of\_sales

Out[54]:

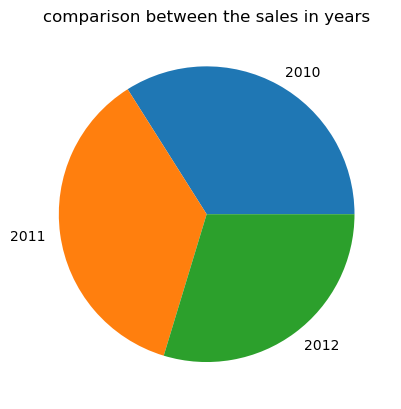
|  | **year** | **Weekly\_Sales** |
| --- | --- | --- |
| **0** | 2010 | 2.288886e+09 |
| **1** | 2011 | 2.448200e+09 |
| **2** | 2012 | 2.000133e+09 |

In [55]:

plt.pie(year\_view\_of\_sales["Weekly\_Sales"],labels=year\_view\_of\_sales["year"])

plt.title("comparison between the sales in years")

plt.show()



In [56]:

sales\_and\_fuel=data.groupby("Store").agg({

'Weekly\_Sales': lambda sales : sales.sum(),

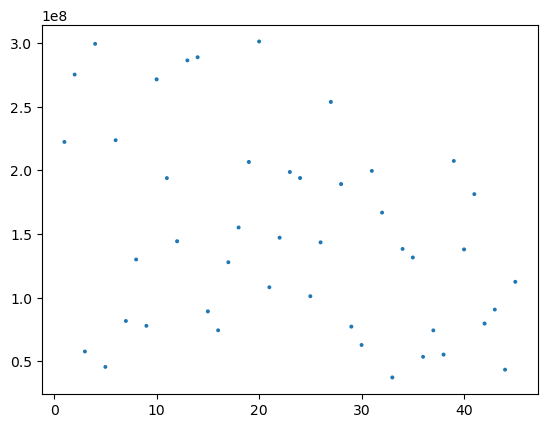
'Fuel\_Price' : lambda fuel : fuel.mean()

}).reset\_index()

plt.scatter(sales\_and\_fuel["Store"],sales\_and\_fuel["Weekly\_Sales"],s=sales\_and\_fuel["Fuel\_Price"],alpha=1)

plt.show()

# fuel price has no effect in the weekly sales



In [57]:

sales\_and\_temp=data.groupby("Store").agg({

'Weekly\_Sales': lambda sales : sales.sum(),

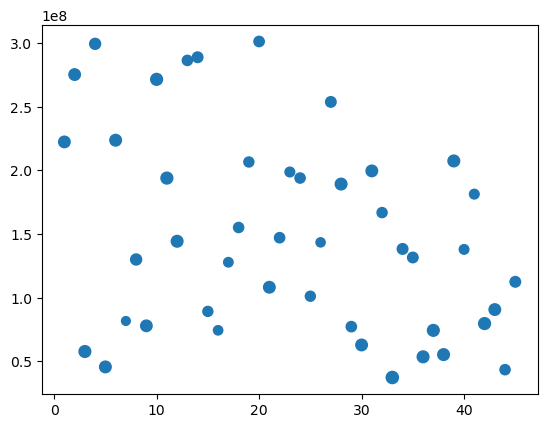
'Temperature' : lambda temp : temp.mean()

}).reset\_index()

plt.scatter(sales\_and\_temp["Store"],sales\_and\_temp["Weekly\_Sales"],s=sales\_and\_temp["Temperature"],alpha=1)

plt.show()

# temperature has no effect in the weekly sales



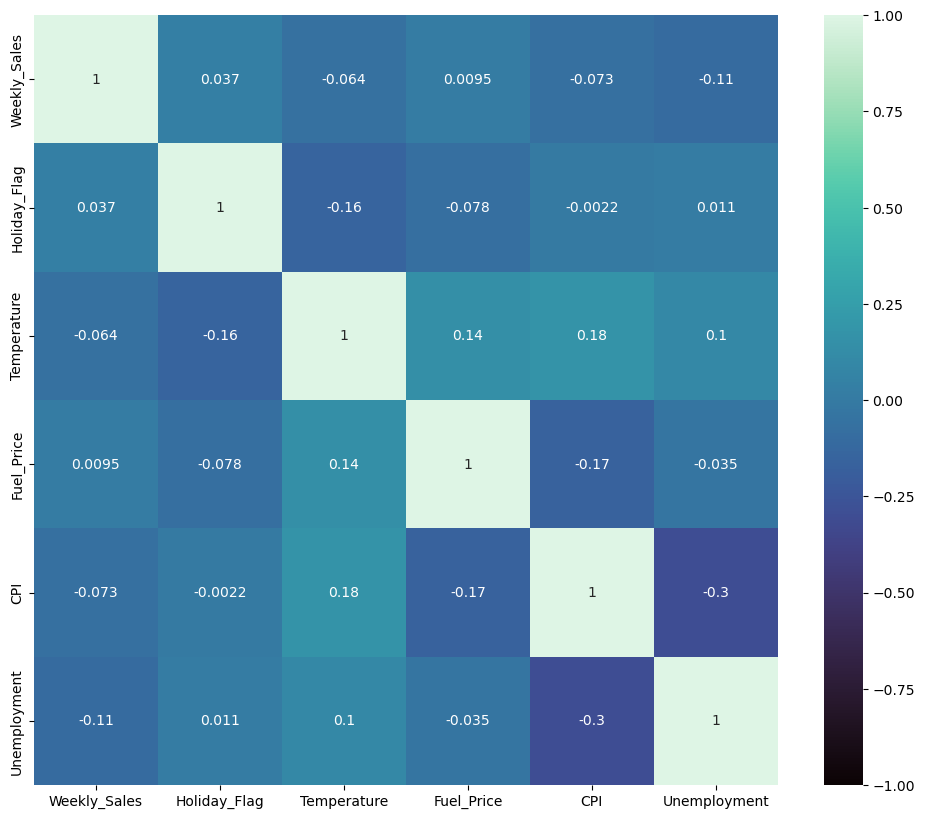
In [58]:

corr = ploted\_data.corr()

plt.figure(figsize=(12, 10))

sns.heatmap(corr, annot=True, vmin=-1.0, cmap='mako')

plt.show()



In [ ]: