In [120]:

```
import numpy as np
import re
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder, StandardScaler, OneHotEncoder
from sklearn.metrics import confusion_matrix, accuracy_score, mean_squared_error, r2_scor
e
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.cluster import KMeans
import plotly.express as px
import scipy.stats as stats
import scipy.cluster.hierarchy as ch
import calendar
import datetime
```

In [129]:

```
data = pd.read_csv(r'C:\Users\Mannahil Miftah\Downloads\retail_services.csv')
data.head(10)
```

Out[129]:

	time.index	time.month	time.month name	time.period	time.year	data.inventories.all department stores	data.inventories.all other home furnishings stores	data.inventories.all other merchandise stores	data
0	1	1	Jan	Jan1992	1992	0	0	0	
1	2	2	Feb	Feb1992	1992	0	0	0	
2	3	3	Mar	Mar1992	1992	0	0	0	
3	4	4	Apr	Apr1992	1992	0	0	0	
4	5	5	May	May1992	1992	0	0	0	
5	6	6	Jun	Jun1992	1992	0	0	0	
6	7	7	Jul	Jul1992	1992	0	0	0	
7	8	8	Aug	Aug1992	1992	0	0	0	
8	9	9	Sep	Sep1992	1992	0	0	0	
9	10	10	Oct	Oct1992	1992	0	0	0	

10 rows × 197 columns

In [130]:

```
print(list(data.keys()))
```

```
['time.index', 'time.month', 'time.month name', 'time.period', 'time.year', 'data.invento
ries.all department stores', 'data.inventories.all other home furnishings stores', 'data.
inventories.all other merchandise stores', 'data.inventories.appliances and other electro
nics stores', 'data.inventories.auto and other motor vehicles', 'data.inventories.automob
ile dealers', 'data.inventories.automotive parts and tire stores', 'data.inventories.beer
, wine, and liquor stores', 'data.inventories.book stores', 'data.inventories.building ma
terials and garden supplies dealers', 'data.inventories.building supplies dealers', 'data
.inventories.clothing stores', 'data.inventories.computer and software stores', 'data.inv
entories.discount department stores', 'data.inventories.drinking places', 'data.inventori
es.electronic shopping and mail-order houses', 'data.inventories.electronics and applianc
e stores', 'data.inventories.family clothing stores', 'data.inventories.floor covering st
```

ores', 'data.inventories.food and beverage stores', 'data.inventories.food services and d
rinking places', 'data.inventories.fuel dealers', 'data.inventories.full service restaura
nts', 'data.inventories.furniture and home furnishings stores', 'data.inventories.furnitu
re stores', 'data.inventories.furniture, home furn, electronics, and appliance stores', '
data.inventories.gafo', 'data.inventories.gasoline stations', 'data.inventories.general m
erchandise stores', 'data.inventories.gift, novelty, and souvenir stores', 'data.inventor
ies.grocery stores', 'data.inventories.hardware stores', 'data.inventories.health and per
sonal care stores', 'data.inventories.hobby, toy, and game stores', 'data.inventories.hom
e furnishings stores', 'data.inventories.household appliance stores', 'data.inventories.j
ewelry stores', 'data.inventories.limited service eating places', "data.inventories.men's
clothing stores", 'data.inventories.miscellaneous store retailers', 'data.inventories.mot
or vehicle and parts dealers', 'data.inventories.new car dealers', 'data.inventories.non-
discount department stores', 'data.inventories.non-leased department stores', 'data.inven
tories.nonstore retailers', 'data.inventories.office supplies and stationery stores', 'da
ta.inventories.office supplies, stationery, and gift stores', 'data.inventories.other clo
thing stores', 'data.inventories.other general merchandise stores', 'data.inventories.pai
nt and wallpaper stores', 'data.inventories.pharmacies and drug stores', 'data.inventorie
s.radio, TV, and electronics stores', 'data.inventories.retail trade', 'data.inventories.
retail trade and food services', 'data.inventories.retail trade and food services, ex aut
o', 'data.inventories.retail trade, ex auto', 'data.inventories.shoe stores', 'data.inven
tories.sporting goods stores', 'data.inventories.sporting goods, hobby, book, and music s
tores', 'data.inventories.supermarkets and other grocery (except convenience) stores', 'd
ata.inventories.used car dealers', 'data.inventories.used merchandise stores', 'data.inve
ntories.warehouse clubs and superstores', "data.inventories.women's clothing stores", 'da
ta.ratio.all department stores', 'data.ratio.all other home furnishings stores', 'data.ra
tio.all other merchandise stores', 'data.ratio.appliances and other electronics stores',
'data.ratio.auto and other motor vehicles', 'data.ratio.automobile dealers', 'data.ratio.
automotive parts and tire stores', 'data.ratio.beer, wine, and liquor stores', 'data.rati
o.book stores', 'data.ratio.building materials and garden supplies dealers', 'data.ratio.
building supplies dealers', 'data.ratio.clothing stores', 'data.ratio.computer and softwa
re stores', 'data.ratio.discount department stores', 'data.ratio.drinking places', 'data.
ratio.electronic shopping and mail-order houses', 'data.ratio.electronics and appliance s
tores', 'data.ratio.family clothing stores', 'data.ratio.floor covering stores', 'data.ra
tio.food and beverage stores', 'data.ratio.food services and drinking places', 'data.rati
o.fuel dealers', 'data.ratio.full service restaurants', 'data.ratio.furniture and home fu
rnishings stores', 'data.ratio.furniture stores', 'data.ratio.furniture, home furn, elect
ronics, and appliance stores', 'data.ratio.gafo', 'data.ratio.gasoline stations', 'data.r
atio.general merchandise stores', 'data.ratio.gift, novelty, and souvenir stores', 'data.
ratio.grocery stores', 'data.ratio.hardware stores', 'data.ratio.health and personal care
stores', 'data.ratio.hobby, toy, and game stores', 'data.ratio.home furnishings stores',
'data.ratio.household appliance stores', 'data.ratio.jewelry stores', 'data.ratio.limited
service eating places', "data.ratio.men's clothing stores", 'data.ratio.miscellaneous sto
re retailers', 'data.ratio.motor vehicle and parts dealers', 'data.ratio.new car dealers'
, 'data.ratio.non-discount department stores', 'data.ratio.non-leased department stores',
'data.ratio.nonstore retailers', 'data.ratio.office supplies and stationery stores', 'dat
a.ratio.office supplies, stationery, and gift stores', 'data.ratio.other clothing stores'
, 'data.ratio.other general merchandise stores', 'data.ratio.paint and wallpaper stores',
'data.ratio.pharmacies and drug stores', 'data.ratio.radio, TV, and electronics stores',
'data.ratio.retail trade', 'data.ratio.retail trade and food services', 'data.ratio.retai
l trade and food services, ex auto', 'data.ratio.retail trade, ex auto', 'data.ratio.shoe
stores', 'data.ratio.sporting goods stores', 'data.ratio.sporting goods, hobby, book, and
music stores', 'data.ratio.supermarkets and other grocery (except convenience) stores', '
data.ratio.used car dealers', 'data.ratio.used merchandise stores', 'data.ratio.warehouse
clubs and superstores', "data.ratio.women's clothing stores", 'data.sales.all department
stores', 'data.sales.all other home furnishings stores', 'data.sales.all other merchandis
e stores', 'data.sales.appliances and other electronics stores', 'data.sales.auto and oth
er motor vehicles', 'data.sales.automobile dealers', 'data.sales.automotive parts and tir
e stores', 'data.sales.beer, wine, and liquor stores', 'data.sales.book stores', 'data.sa
les.building materials and garden supplies dealers', 'data.sales.building supplies dealer
s', 'data.sales.clothing stores', 'data.sales.computer and software stores', 'data.sales.
discount department stores', 'data.sales.drinking places', 'data.sales.electronic shoppin
g and mail-order houses', 'data.sales.electronics and appliance stores', 'data.sales.fami
ly clothing stores', 'data.sales.floor covering stores', 'data.sales.food and beverage st
ores', 'data.sales.food services and drinking places', 'data.sales.fuel dealers', 'data.s
ales.full service restaurants', 'data.sales.furniture and home furnishings stores', 'data
.sales.furniture stores', 'data.sales.furniture, home furn, electronics, and appliance st
ores', 'data.sales.gafo', 'data.sales.gasoline stations', 'data.sales.general merchandise
stores', 'data.sales.gift, novelty, and souvenir stores', 'data.sales.grocery stores', 'd
ata.sales.hardware stores', 'data.sales.health and personal care stores', 'data.sales.hob
by, toy, and game stores', 'data.sales.home furnishings stores', 'data.sales.household ap
pliance stores', 'data.sales.jewelry stores', 'data.sales.limited service eating places',
"data.sales.men's clothing stores", 'data.sales.miscellaneous store retailers', 'data.sal

```
es.motor vehicle and parts dealers', 'data.sales.new car dealers', 'data.sales.non-discou
nt department stores', 'data.sales.non-leased department stores', 'data.sales.nonstore re
tailers', 'data.sales.office supplies and stationery stores', 'data.sales.office supplies
, stationery, and gift stores', 'data.sales.other clothing stores', 'data.sales.other gen
eral merchandise stores', 'data.sales.paint and wallpaper stores', 'data.sales.pharmacies
and drug stores', 'data.sales.radio, TV, and electronics stores', 'data.sales.retail trad
e', 'data.sales.retail trade and food services', 'data.sales.retail trade and food servic
es, ex auto', 'data.sales.retail trade, ex auto', 'data.sales.shoe stores', 'data.sales.s
porting goods stores', 'data.sales.sporting goods, hobby, book, and music stores', 'data.
sales.supermarkets and other grocery (except convenience) stores', 'data.sales.used car d
ealers', 'data.sales.used merchandise stores', 'data.sales.warehouse clubs and superstore
s', "data.sales.women's clothing stores"]
```

In [131]:

```
data.shape
```

Out[131]:

```
(289, 197)
```

In [132]:

```
data.dtypes
```

Out[132]:

```
time.index                                                                  int64
time.month                                                                  int64
time.month name                                                            object
time.period                                                                object
time.year                                                                   int64
                                                                           ...
data.sales.supermarkets and other grocery (except convenience) stores       int64
data.sales.used car dealers                                                 int64
data.sales.used merchandise stores                                          int64
data.sales.warehouse clubs and superstores                                  int64
data.sales.women's clothing stores                                          int64
Length: 197, dtype: object
```

How has retail economic activity in the United States changed over the past five years?

In [133]:

```
# columns which are of retail

retail_col = [i for i in data.columns
              if i.startswith('data.sales.retail')]
```

In [134]:

```
retail_col
```

Out[134]:

```
['data.sales.retail trade',
 'data.sales.retail trade and food services',
 'data.sales.retail trade and food services, ex auto',
 'data.sales.retail trade, ex auto']
```

In [135]:

```
# copy the data of retail columns only

last_5_years = data[['time.year'] + retail_col].copy()
last_5_years.head(10)
```

Out[135]:

time.year	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto

	time.year	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto
0	1992	130883	146378	115563	100972
1	1992	131244	147079	115862	100027
2	1992	142488	159336	124200	107352
3	1992	147175	163669	127587	111093
4	1992	152420	170068	133608	115960
5	1992	151849	168663	130274	113460
6	1992	152586	169890	132076	114772
7	1992	152476	170364	134928	117040
8	1992	148158	164617	128734	112275
9	1992	155987	173655	136917	119249

In [136]:

```
last_5_years = last_5_years[last_5_years['time.year'] >= 2012]
last_5_years.head(5)
```

Out[136]:

	time.year	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto
240	2012	315626	355332	293066	253360
241	2012	331777	372782	302230	261225
242	2012	368818	414545	333112	287385
243	2012	349455	393289	319594	275760
244	2012	373374	418940	338945	293379

In [137]:

```
year = last_5_years.groupby('time.year')[retail_col].sum()
year
```

Out[137]:

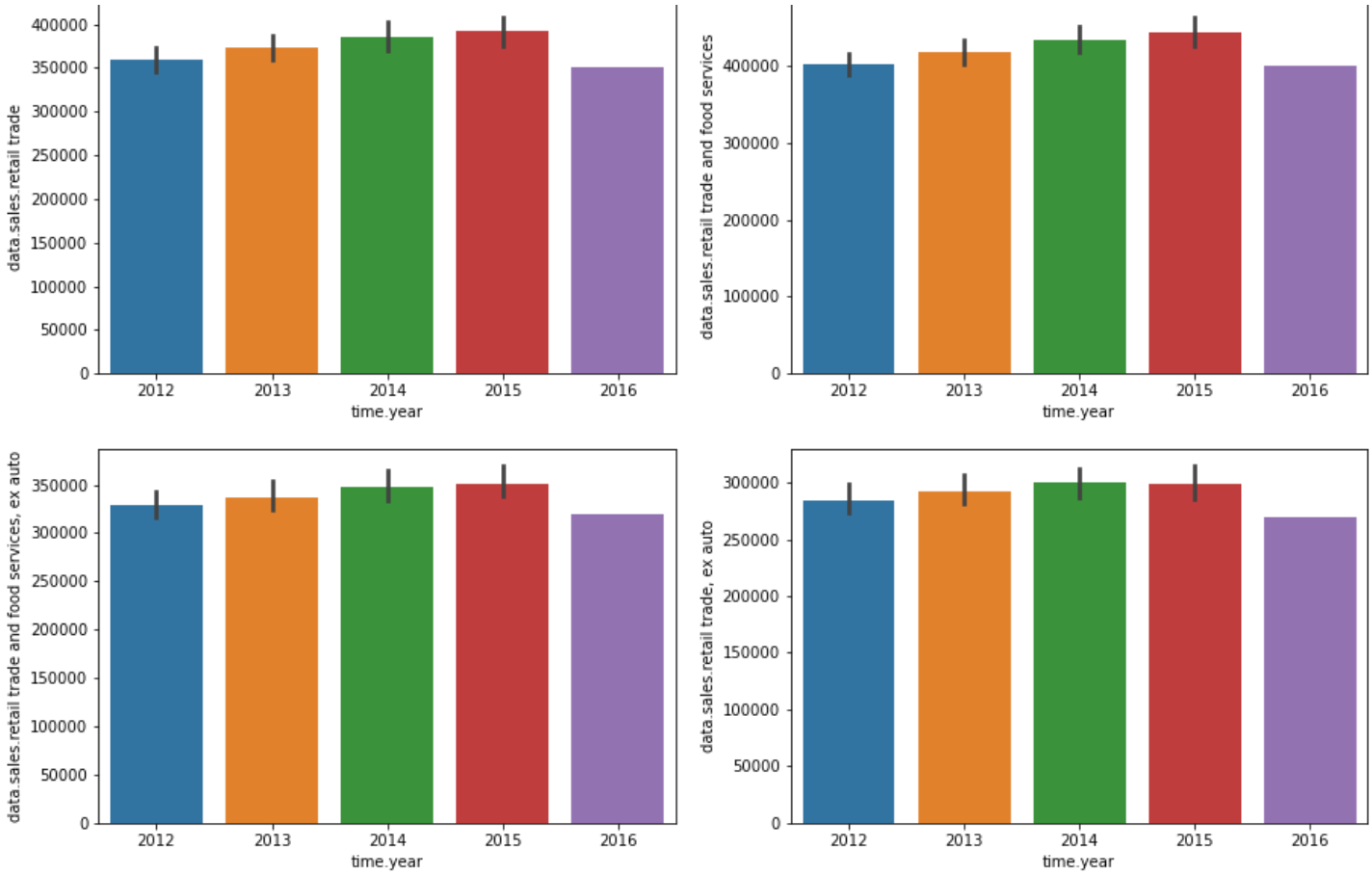
time.year	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto
2012	4306237	4831131	3943141	3418247
2013	4469022	5011740	4049759	3507041
2014	4632289	5208443	4174533	3598379
2015	4699327	5321997	4215309	3592639
2016	350025	400249	319532	269308

In [138]:

```
fig, axes = plt.subplots(2,2, figsize=(15, 10))
sb.barplot(data = last_5_years, x = 'time.year', y = 'data.sales.retail trade', ax = axe
s[0,0])
sb.barplot(data = last_5_years, x = 'time.year', y = 'data.sales.retail trade and food s
ervices', ax = axes[0,1])
sb.barplot(data = last_5_years, x = 'time.year', y = 'data.sales.retail trade and food s
ervices, ex auto', ax = axes[1,0])
sb.barplot(data = last_5_years, x = 'time.year', y = 'data.sales.retail trade, ex auto',
ax = axes[1,1])
```

Out[138]:

```
<AxesSubplot:xlabel='time.year', ylabel='data.sales.retail trade, ex auto'>
```



What are the key differences between the Advance Monthly Retail Trade Survey (MARTS) and the Annual Retail Trade Survey (ARTS)?

MARTS is conducted on monthly basis whereas ARTS is conducted is on yearly basis.

MARTS is for short-term analysis whereas ARTS is for long-term analysis.

In [139]:

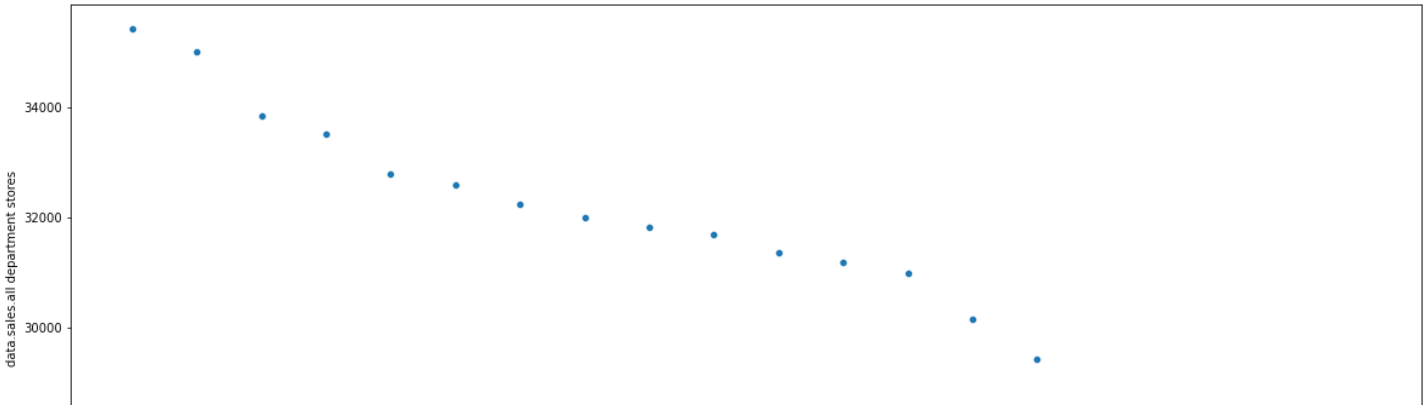
```
marts = pd.DataFrame(data.groupby('time.period')['data.sales.all department stores'].sum(
)).reset_index()
arts = pd.DataFrame(data.groupby('time.year')['data.sales.all department stores'].sum()).
reset_index()
marts = marts.sort_values(by = 'data.sales.all department stores', ascending = False)
arts = arts.sort_values(by = 'data.sales.all department stores', ascending = False)
```

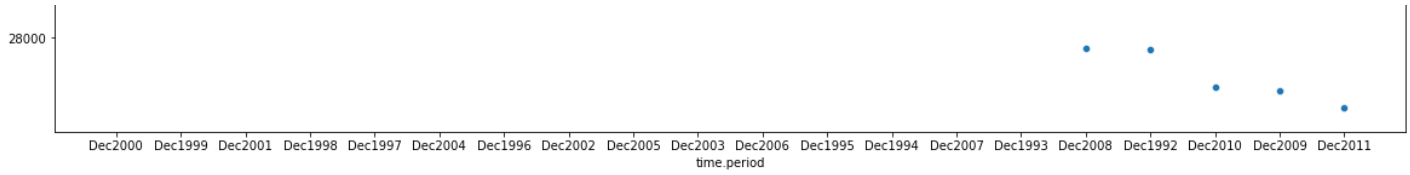
In [140]:

```
plt.figure(figsize=[20,8])
sb.scatterplot(data = marts.head(20), x = 'time.period', y = 'data.sales.all department
stores')
```

Out[140]:

```
<AxesSubplot:xlabel='time.period', ylabel='data.sales.all department stores'>
```



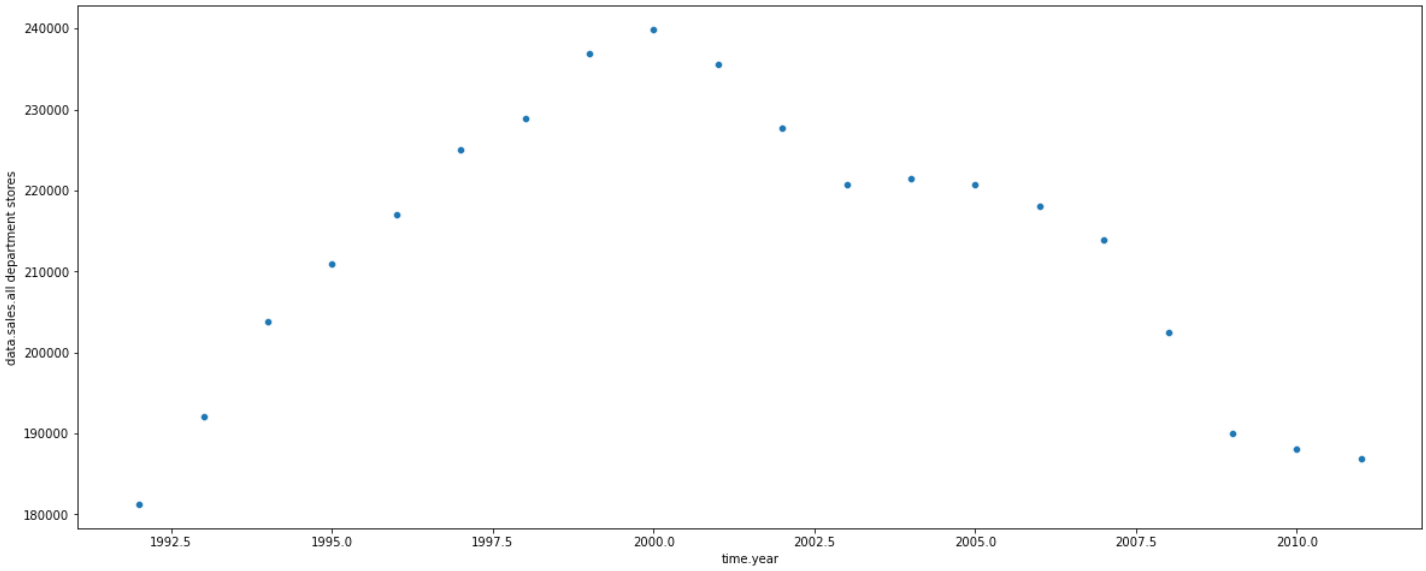


In [141]:

```
plt.figure(figsize=[20,8])
sb.scatterplot(data = arts.head(20), x = 'time.year', y = 'data.sales.all department sto
res')
```

Out[141]:

```
<AxesSubplot:xlabel='time.year', ylabel='data.sales.all department stores'>
```



Can we identify any seasonal patterns or trends in monthly retail sales data?

In [142]:

```
group = pd.DataFrame(data.groupby('time.period')['data.sales.all department stores'].sum(
))
group = group.sort_values(by = 'data.sales.all department stores', ascending = False).re
set_index()
group.head(10)
```

Out[142]:

	time.period	data.sales.all department stores
0	Dec2000	35437
1	Dec1999	35002
2	Dec2001	33838
3	Dec1998	33509
4	Dec1997	32783
5	Dec2004	32593
6	Dec1996	32246
7	Dec2002	31987
8	Dec2005	31821
9	Dec2003	31697

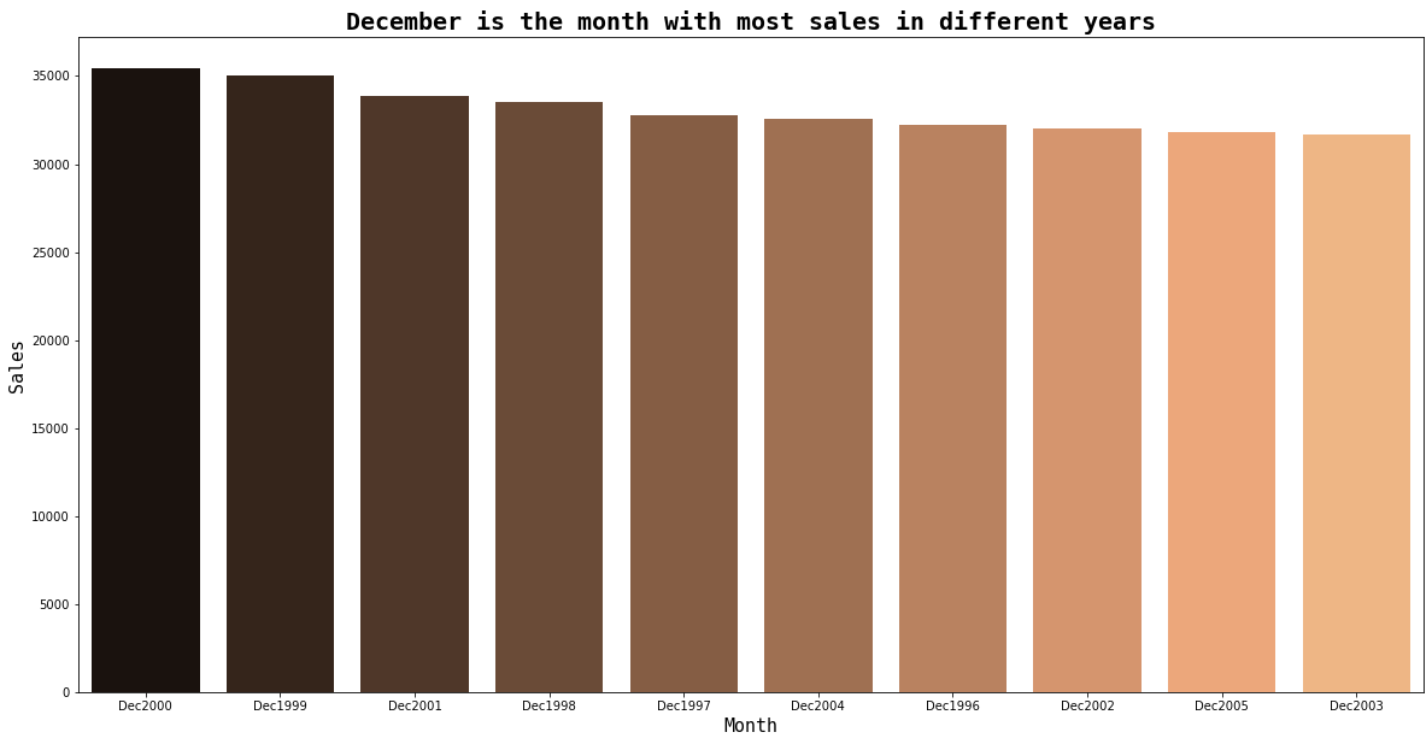
In [143]:

```
plt.figure(figsize=[20,10])
sb.barplot(data = group.head(10), x = 'time.period', y = 'data.sales.all department stor
```

```
es', palette = 'copper')
plt.title('December is the month with most sales in different years', fontdict={'fontname
': 'Monospace', 'fontsize': 20, 'fontweight': 'bold'})
plt.xlabel('Month', fontdict={'fontname':'Monospace', 'fontsize': 15,})
plt.ylabel('Sales', fontdict={'fontname':'Monospace', 'fontsize': 15,})
```

Out[143]:

```
Text(0, 0.5, 'Sales')
```



In [144]:

```
group = group.sort_values(by = 'data.sales.all department stores', ascending = True)
group.head(10)
```

Out[144]:

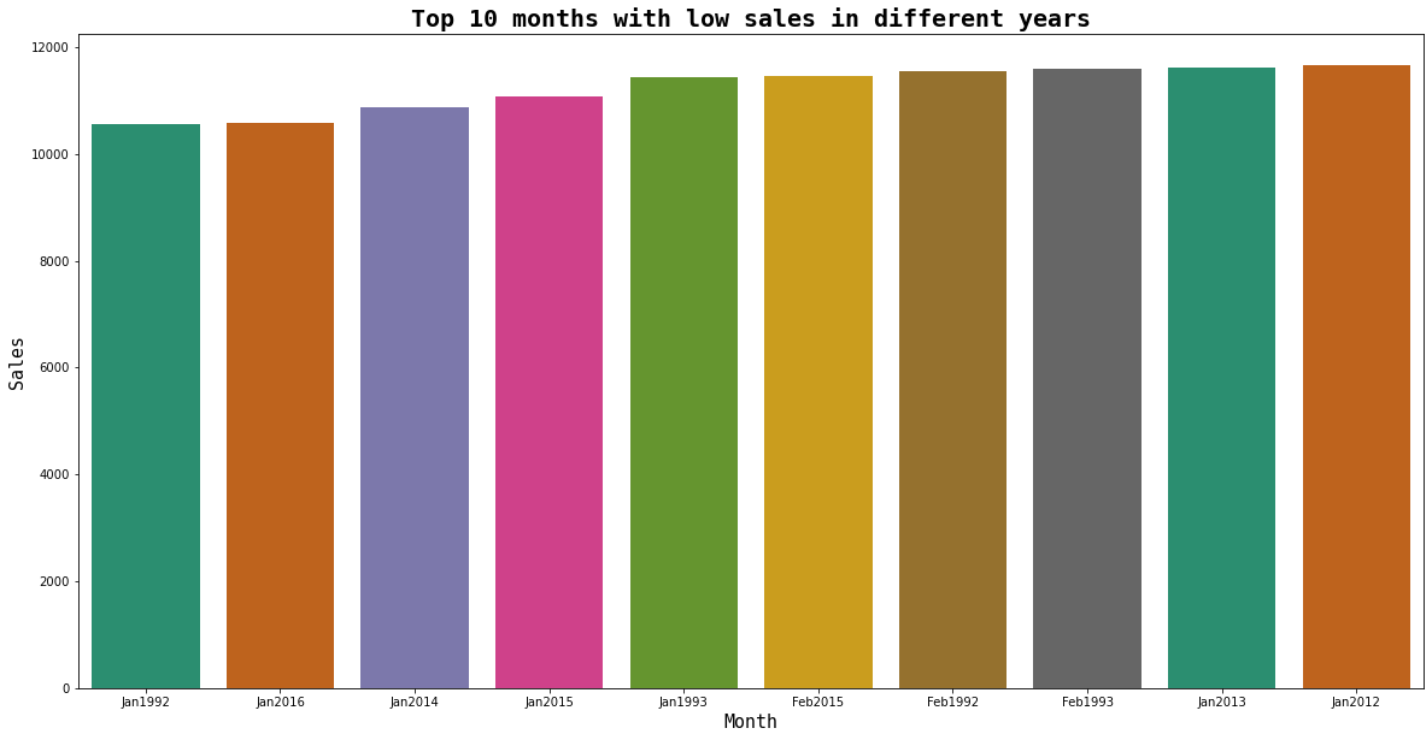
	time.period	data.sales.all department stores
288	Jan1992	10560
287	Jan2016	10571
286	Jan2014	10879
285	Jan2015	11083
284	Jan1993	11429
283	Feb2015	11466
282	Feb1992	11549
281	Feb1993	11595
280	Jan2013	11621
279	Jan2012	11659

In [145]:

```
plt.figure(figsize=[20,10])
sb.barplot(data = group.head(10), x = 'time.period', y = 'data.sales.all department stor
es', palette = 'Dark2')
plt.title('Top 10 months with low sales in different years', fontdict={'fontname': 'Monos
pace', 'fontsize': 20, 'fontweight': 'bold'})
plt.xlabel('Month', fontdict={'fontname':'Monospace', 'fontsize': 15,})
plt.ylabel('Sales', fontdict={'fontname':'Monospace', 'fontsize': 15,})
```

Out[145]:

Text(0, 0.5, 'Sales')



In [146]:

```
month_retail = pd.DataFrame(data.groupby('time.month name')[retail_col].sum())
month_retail = month_retail.reset_index()
month_retail
```

Out[146]:

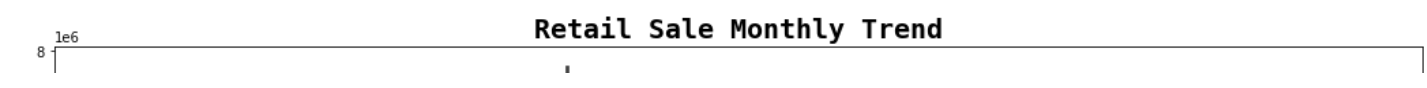
	time.month name	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto
0	Apr	6481734	7230380	5638605	4889959
1	Aug	6854640	7641463	5958068	5171245
2	Dec	7880138	8656280	7177051	6400909
3	Feb	5797595	6477927	5071056	4390724
4	Jan	6164222	6896777	5479533	4746978
5	Jul	6691865	7472735	5815511	5034641
6	Jun	6688287	7453211	5802005	5037081
7	Mar	6578999	7340909	5680510	4918600
8	May	6852754	7639097	5967901	5181558
9	Nov	6710011	7439880	6027567	5297698
10	Oct	6595930	7365613	5841473	5071790
11	Sep	6391135	7131338	5616310	4876107

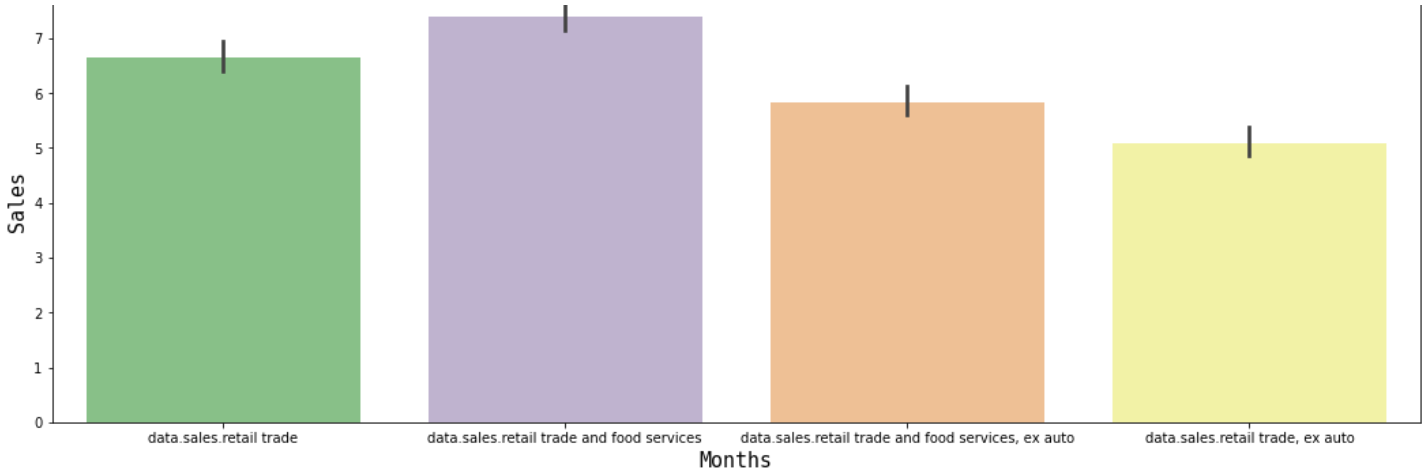
In [147]:

```
plt.figure(figsize=[18,6])
sb.barplot(data = month_retail, palette = 'Accent')
plt.title('Retail Sale Monthly Trend', fontdict={'fontname': 'Monospace', 'fontsize': 20
, 'fontweight': 'bold'})
plt.xlabel('Months', fontdict={'fontname':'Monospace', 'fontsize': 15,})
plt.ylabel('Sales', fontdict={'fontname':'Monospace', 'fontsize': 15,})
```

Out[147]:

Text(0, 0.5, 'Sales')





How does e-commerce activity compare to traditional retail sales on a quarterly basis?

In [173]:

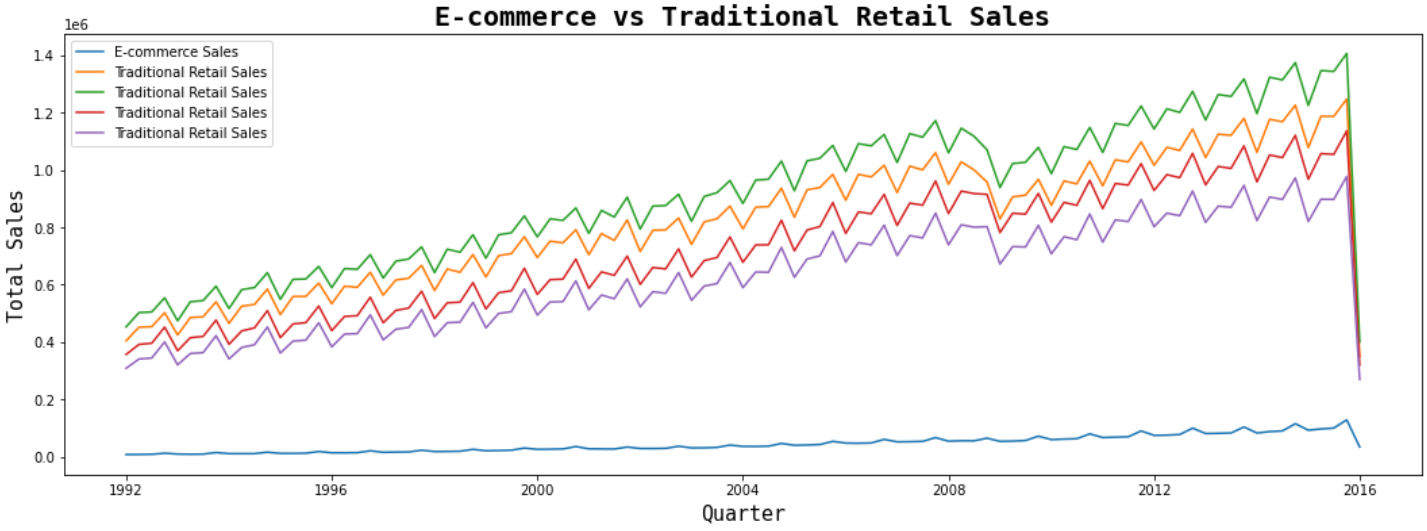
```
data1 = data.copy()
data1 ['time.period'] = pd.to_datetime(data['time.period'])
data1['Quarter'] = data1['time.period'].dt.to_period('Q')
```

In [177]:

```
ecommerce_sales = data1.groupby('Quarter')['data.sales.electronic shopping and mail-order houses'].sum()
traditional_sales = data1.groupby('Quarter')[retail_col].sum()
ecommerce_sales.index = ecommerce_sales.index.to_timestamp()
traditional_sales.index = traditional_sales.index.to_timestamp()
```

In [201]:

```
plt.figure(figsize=[18,6])
plt.plot(ecommerce_sales.index, ecommerce_sales, label='E-commerce Sales')
plt.plot(traditional_sales.index, traditional_sales, label='Traditional Retail Sales')
plt.title('E-commerce vs Traditional Retail Sales', fontdict={'fontname': 'Monospace', 'fontsize': 20, 'fontweight': 'bold'})
plt.xlabel('Quarter', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.ylabel('Total Sales', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.legend()
plt.show()
```



Are there any specific retail sectors that have shown significant growth or decline in recent years?

In [186]:

```
data_copy = data.copy()
data_copy['time.index'] = pd.to_datetime(data['time.index'])
```

In [200]:

```
grouping = data_copy.groupby('time.index')[retail_col].sum()
grouping = grouping.reset_index()
grouping.head(2)
```

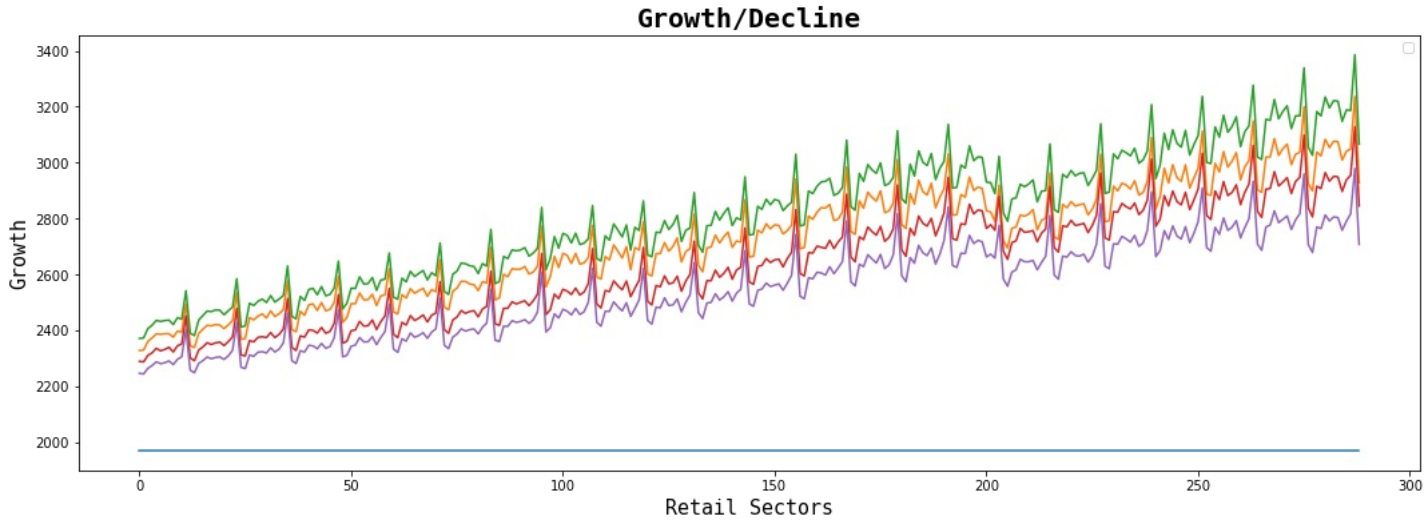
Out[200]:

	time.index	data.sales.retail trade	data.sales.retail trade and food services	data.sales.retail trade and food services, ex auto	data.sales.retail trade, ex auto
0	1970-01-01 00:00:00.000000001	130683	146376	116565	100872
1	1970-01-01 00:00:00.000000002	131244	147079	115862	100027

In [202]:

```
plt.figure(figsize=[18,6])
plt.plot(grouping)
plt.title('Growth/Decline', fontdict={'fontname': 'Monospace', 'fontsize': 20, 'fontweig
ht': 'bold'})
plt.xlabel('Retail Sectors', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.ylabel('Growth', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.show()
```

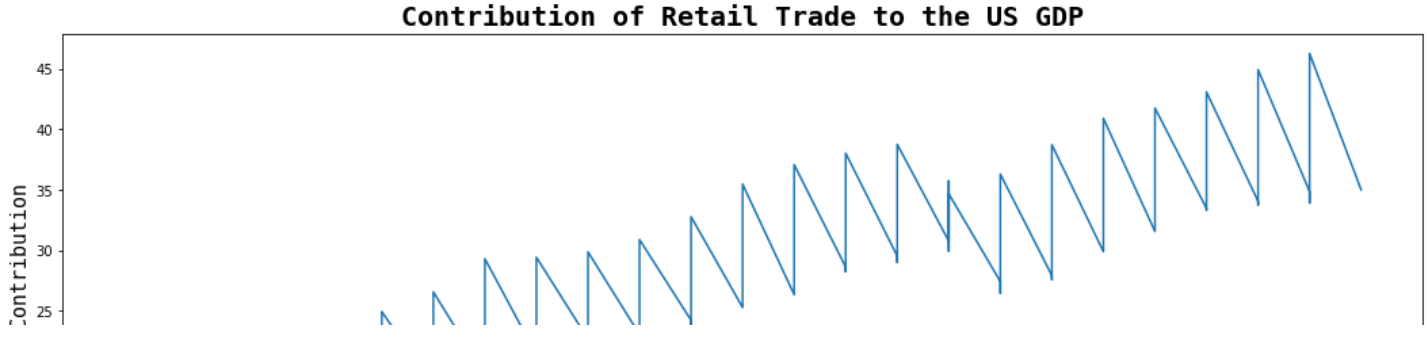
```
No handles with labels found to put in legend.
```

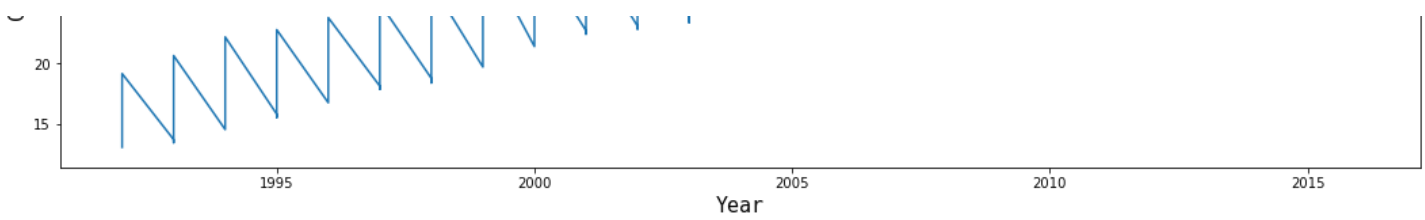


What is the overall contribution of retail trade to the United States' GDP?

In [205]:

```
data_copy['gdp'] = (data_copy['data.sales.retail trade'] / 1000000) * 100
plt.figure(figsize=[18,6])
plt.plot(data_copy['time.year'], data_copy['gdp'])
plt.title('Contribution of Retail Trade to the US GDP', fontdict={'fontname': 'Monospace'
, 'fontsize': 20, 'fontweight': 'bold'})
plt.xlabel('Year', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.ylabel('Contribution', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.show()
```





How do retail operating expenses vary across different types of businesses?

In [225]:

```
businesses = [i for i in data.columns
              if i.startswith('data.inventories')]

diff = pd.DataFrame(data_copy.groupby('time.year')[businesses].sum())
diff.head(2)
```

Out[225]:

time.year	data.inventories.all department stores	data.inventories.all other home furnishings stores	data.inventories.all other merchandise stores	data.inventories.appliances and other electronics stores	data.inventories.auto and other motor vehicles	data.inv
1992	0	0	0	0	0	
1993	0	0	0	0	0	

2 rows × 64 columns

In [226]:

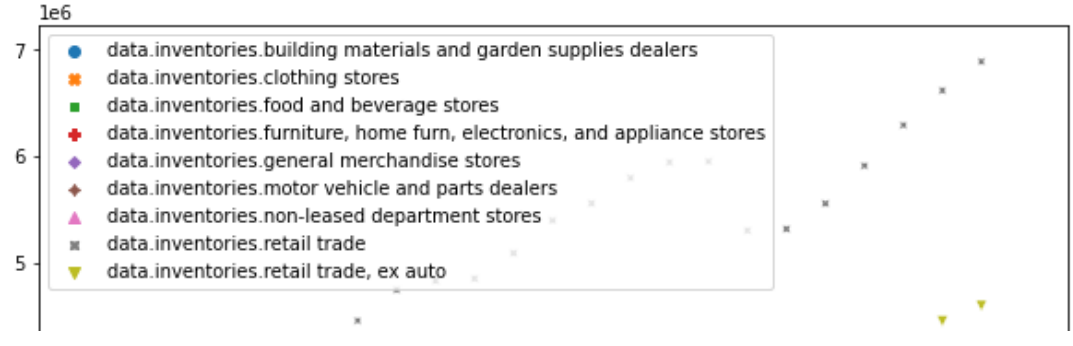
```
zero_cols = [ col for col, is_zero in ((diff == 0).sum() == diff.shape[0]).items() if is
_zero ]
diff.drop(zero_cols, axis=1, inplace=True)
diff.head(2)
```

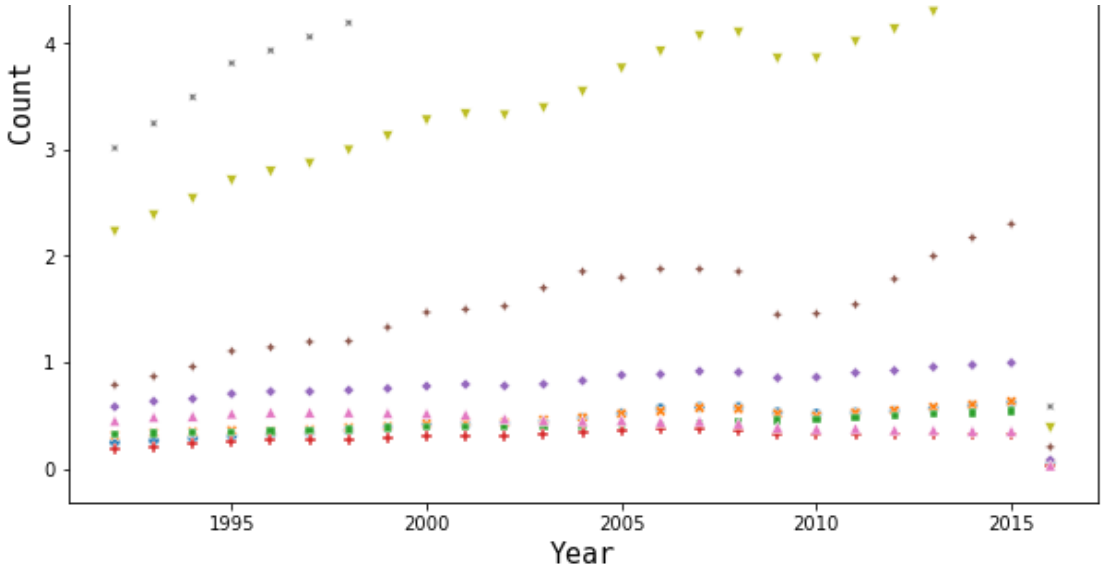
Out[226]:

time.year	data.inventories.building materials and garden supplies dealers	data.inventories.clothing stores	data.inventories.food and beverage stores	data.inventories.furniture, home furn, electronics, and appliance stores	data.inventories.g merchandise s
1992	246458	313147	326498	180985	5
1993	263226	335570	330438	202594	6

In [233]:

```
plt.figure(figsize=[10,8])
sb.scatterplot(data = diff)
plt.xlabel('Year', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.ylabel('Count', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.show()
```





Is there a relationship between e-commerce sales and brick-and-mortar retail sales?

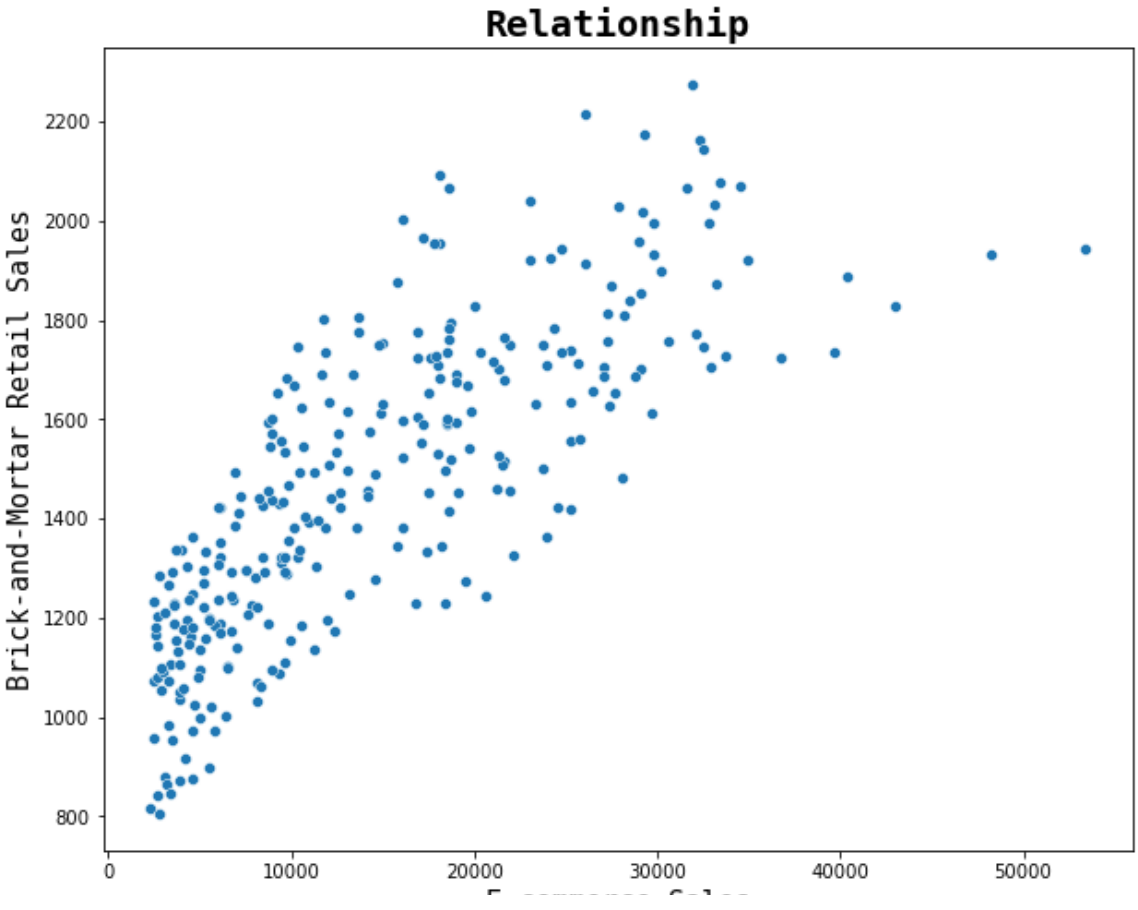
In [234]:

```
corr = data['data.sales.electronic shopping and mail-order houses'].corr(data['data.sales
.hardware stores'])
print("Correlation is", corr)
```

```
Correlation is 0.791367000392785
```

In [236]:

```
plt.figure(figsize=[10,8])
sb.scatterplot(data = data, x = 'data.sales.electronic shopping and mail-order houses',
y = 'data.sales.hardware stores')
plt.title('Relationship', fontdict={'fontname': 'Monospace', 'fontsize': 20, 'fontweight
': 'bold'})
plt.xlabel('E-commerce Sales', fontdict={'fontname': 'Monospace', 'fontsize': 15})
plt.ylabel('Brick-and-Mortar Retail Sales', fontdict={'fontname': 'Monospace', 'fontsize
': 15})
plt.show()
```



E-commerce Sales

How accurate are the monthly estimates compared to the annual survey data?

In [237]:

```
# Monthly survey is performed 12 times a year whereas annual survey is performed only onc
e at then end of year.
# By performing monthly survey we can clear insights about the data and the changes which
might occur whereas in
# annual survey data can be unreliable as changes might happen throughout the year & it c
an be complex as well.
```

Can we identify any correlations between retail sales and macroeconomic indicators such as unemployment rates or consumer sentiment?

Can we identify any discrepancies between the retail data collected by MARTS, MRTS, ARTS, and the Economic Census of Retail Trade?

How do the retail sales patterns differ between rural and urban areas?

Are there any geographical variations in retail sales trends across different states or regions?

What is the market share of e-commerce sales compared to traditional retail sales?

How have retail operating expenses changed over the past decade?

What insights can we gain from comparing the detailed business operating expenses collected in the Business Expenses Supplement with other retail data sources?

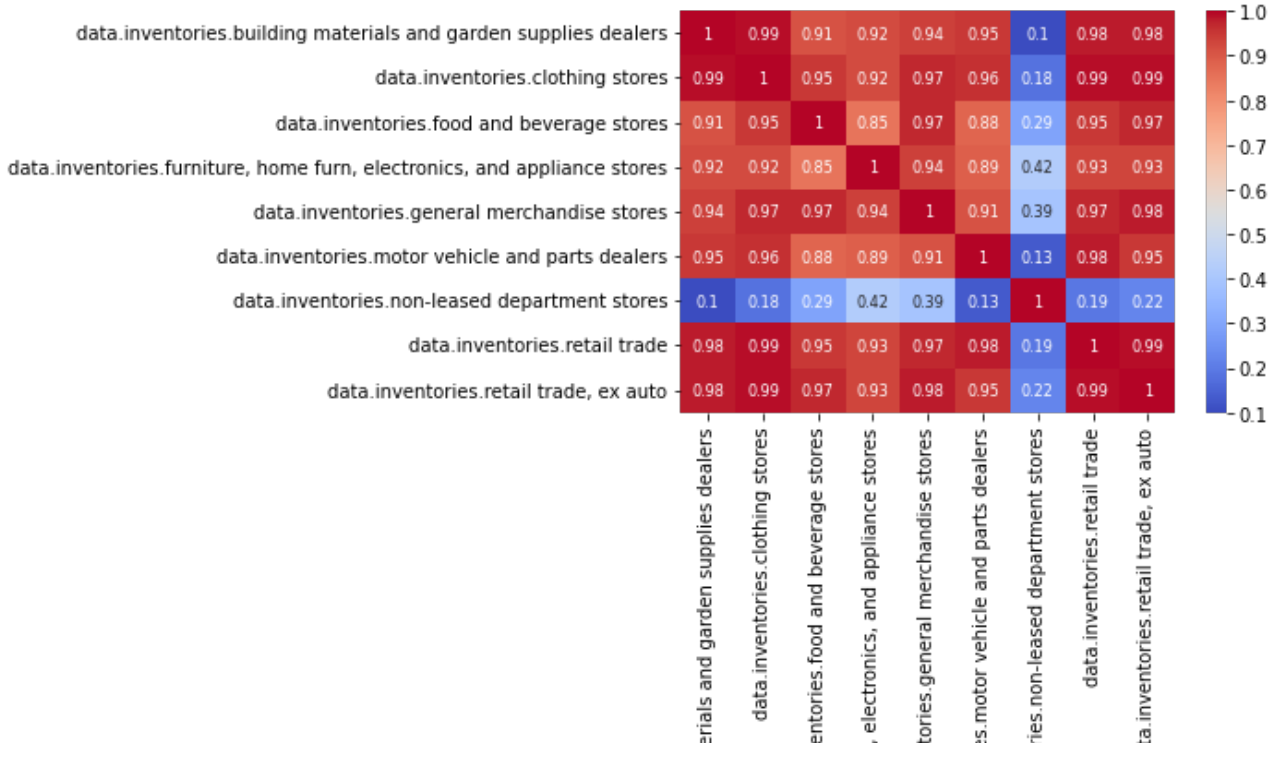
In [238]:

```
# Data is not provided for the above tasks
```

What are the main factors influencing fluctuations in monthly retail sales data?

In [241]:

```
sb.heatmap(data=diff.corr().round(2), cmap='coolwarm', annot=True, annot_kws={"size":8})
plt.show()
```



Can we predict future retail sales based on historical data and other relevant factors?

In [242]:

```
# Yes, by applying ML algorithms
```

Can we identify any outliers or anomalies in the retail sales data that require further investigation?

In [255]:

```
def cal_outliers(data):
    q1 = data.quantile(0.25)
    q3 = data.quantile(0.75)
    iqr = q3 - q1
    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
    result = data[(data < lower_bound) | (data > upper_bound)]
    return result
data1 = diff.keys()
result = cal_outliers(diff[data1])
result
```

Out[255]:

time.year	data.inventories.building materials and garden supplies dealers	data.inventories.clothing stores	data.inventories.food and beverage stores	data.inventories.furniture, home furn, electronics, and appliance stores	data.inventories.g merchandise s
1992	NaN	NaN	NaN	180985.0	
1993	NaN	NaN	NaN	NaN	
1994	NaN	NaN	NaN	NaN	
1995	NaN	NaN	NaN	NaN	
1996	NaN	NaN	NaN	NaN	
1997	NaN	NaN	NaN	NaN	
1998	NaN	NaN	NaN	NaN	
1999	NaN	NaN	NaN	NaN	
2000	NaN	NaN	NaN	NaN	
2001	NaN	NaN	NaN	NaN	
2002	NaN	NaN	NaN	NaN	
2003	NaN	NaN	NaN	NaN	
2004	NaN	NaN	NaN	NaN	
2005	NaN	NaN	NaN	NaN	
2006	NaN	NaN	NaN	NaN	
2007	NaN	NaN	NaN	NaN	
2008	NaN	NaN	NaN	NaN	

time.year	data.inventories.building materials and garden supplies dealers	data.inventories.clothing stores	data.inventories.food and beverage stores	data.inventories.furniture, home furn, electronics, and appliance stores	data.inventories.g merchandise s
2009	NaN	NaN	NaN	NaN	
2010	NaN	NaN	NaN	NaN	
2011	NaN	NaN	NaN	NaN	
2012	NaN	NaN	NaN	NaN	
2013	NaN	NaN	NaN	NaN	
2014	NaN	NaN	NaN	NaN	
2015	NaN	NaN	NaN	NaN	
2016	NaN	51397.0	45124.0	26479.0	79

What are the most significant challenges in reconciling the monthly and annual data for retail economic activity?

In [256]:

```
# Direct Comparison Monthly and annual data as timeframes vary.
# Consistency between monthly data and annual data can be complex as in monthly survey da
ta is revised at the end of each month.
```

In [ ]:

```

```