CAGR Analysis

November 24, 2024

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import datetime as dt
```

[2]:
```
df = pd.read_csv('US Superstore.csv', encoding = 'latin-1')
```

[3]:
```
df['Order Date'] = pd.to_datetime(df['Order Date'], errors = 'coerce', dayfirst␣
 ↪= True)
```

[4]:
```
df['Order Year'] = df['Order Date'].dt.year
df['Order Month'] = df['Order Date'].dt.month
```

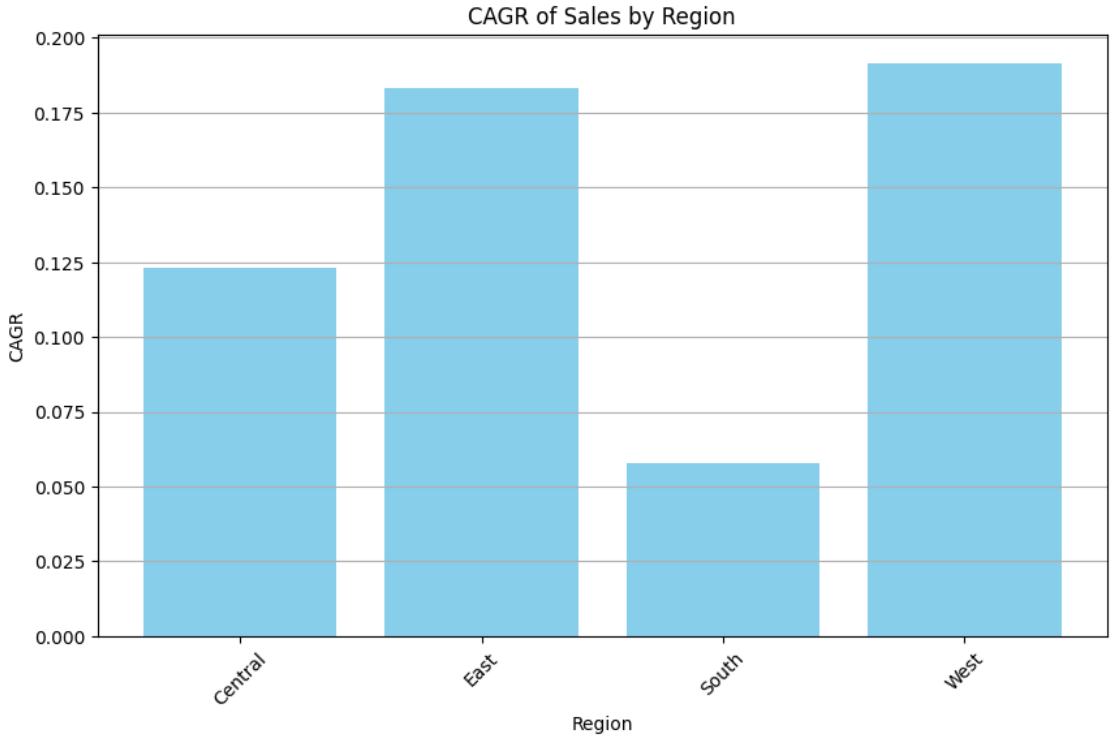
[5]:
```
annual_sales = df.groupby(['Order Year', 'Region'])['Sales'].sum().reset_index()

first_sales = annual_sales.groupby('Region').first().reset_index()
last_sales = annual_sales.groupby('Region').last().reset_index()
sales_summary = pd.merge(first_sales[['Region', 'Sales']],␣
 ↪last_sales[['Region', 'Sales']], on='Region', suffixes=('_start', '_end'))
years_per_region = annual_sales.groupby('Region')['Order Year'].nunique().
 ↪reset_index()
sales_summary = pd.merge(sales_summary, years_per_region[['Region', 'Order␣
 ↪Year']], on='Region')
sales_summary['CAGR'] = ((sales_summary['Sales_end'] /␣
 ↪sales_summary['Sales_start']) ** (1 / (sales_summary['Order Year'] - 1))) - 1
sales_summary[['Region', 'CAGR']]
```

[5]:

	Region	CAGR
0	Central	0.123096
1	East	0.183074
2	South	0.057778
3	West	0.191466

[6]:
```
plt.figure(figsize=(10, 6))
plt.bar(sales_summary['Region'], sales_summary['CAGR'], color='skyblue')
plt.title('CAGR of Sales by Region')
plt.xlabel('Region')
plt.ylabel('CAGR')
plt.xticks(rotation=45)
```

```
plt.grid(True, axis='y')
plt.show()
```



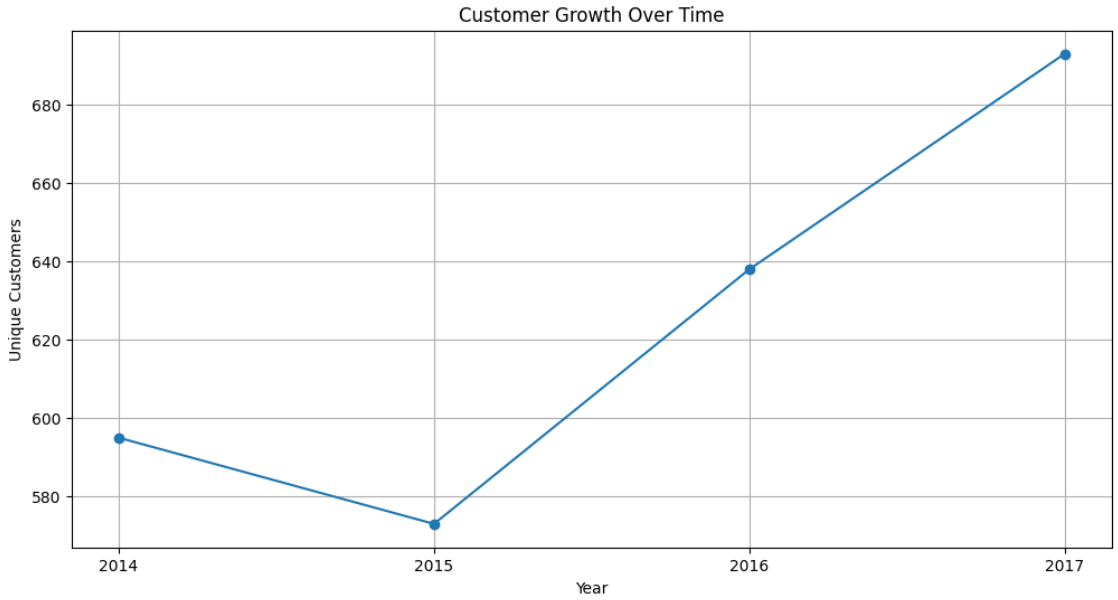
[7]:
```
customerCount = df.groupby('Order Year')['Customer ID'].nunique().reset_index()
customerCount.rename( columns= {'Customer ID' : 'Customer Count'}, inplace =␣
 ↪True)
```

[8]:
```
firstYearCustomers = customerCount['Customer Count'].iloc[0]
lastYearCustomers = customerCount['Customer Count'].iloc[-1]
years = customerCount['Order Year'].iloc[-1] - customerCount['Order Year'].
 ↪iloc[0]

CAGR = (lastYearCustomers / firstYearCustomers) ** (1 / years) -1
print(f'The customer growth over the years is: {CAGR*100:.2f}%')
```

```
The customer growth over the years is: 5.21%
```

[9]:
```
plt.figure(figsize = (12, 6))
plt.plot(customerCount['Order Year'], customerCount['Customer Count'], marker =␣
 ↪'o')
plt.title('Customer Growth Over Time')
plt.xlabel('Year')
plt.ylabel('Unique Customers')
```

```
plt.xticks(df['Order Year'].unique())
plt.grid(True)
plt.show()
```



[ ]: