```
In [4]:  import pandas as pd
         from pytrends.request import TrendReq
         from sklearn.linear_model import Lasso
         from sklearn.preprocessing import StandardScaler
         import matplotlib.pyplot as plt
         import numpy as np
```

1.) Clean the Apple Data to get a quarterly series of EPS.

```
In [7]:  y = pd.read_csv("AAPL_quarterly_financials.csv")
```

```
In [8]:  y.index = y.name
```

```
In [9]:  y = pd.DataFrame(y.loc["BasicEPS", :]).iloc[2:,:]
```

```
In [10]: y.index = pd.to_datetime(y.index)
```

```
In [11]: # CHECK IF NAS ARE NO DIVIDEND PERIOD
         y = y.sort_index().fillna(0.)
```

2.) Come up with 6 search terms you think could nowcast earnings. (Different than the ones I used) Add in 3 terms that that you think will not Nowcast earnings. Pull in the gtrends data

```
In [219…  # Create pytrends object
          pytrends = TrendReq(hl='en-US', tz=360)

          # Set up the keywords and the timeframe
          keywords = ["iPhone", "Reccession","Samsung","Interest Rates", "New Phone
          start_date = '2004-01-01'
          end_date = '2024-01-01'

          # Create an empty DataFrame to store the results
          df = pd.DataFrame()

          # Iterate through keywords and fetch data
          for keyword in keywords:
              pytrends.build_payload([keyword], cat=0, timeframe=f'{start_date} {en
              interest_over_time_df = pytrends.interest_over_time()
              df[keyword] = interest_over_time_df[keyword]
```

```
In [220…  X = df.resample("Q").mean()
```

```
In [221...   # ALIGN DATA
             temp = pd.concat([y, X],axis = 1).dropna()
             y = temp[["BasicEPS"]].copy()
             X = temp.iloc[:,1:].copy()
```
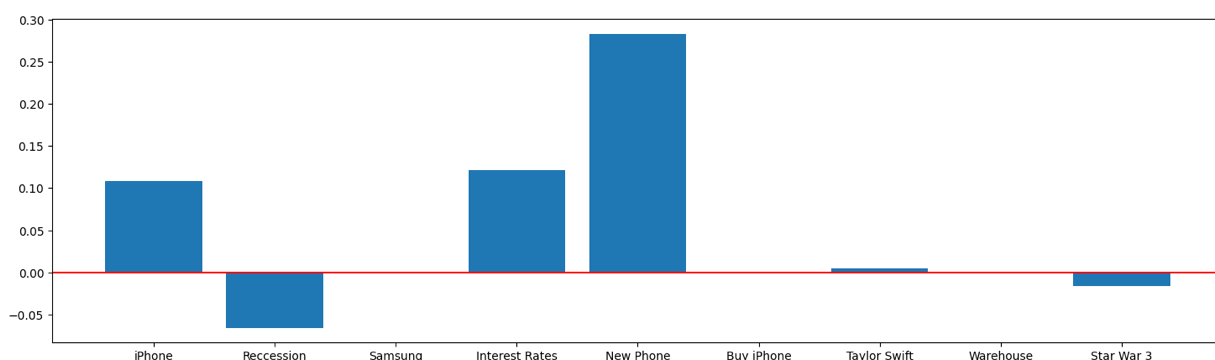
3.) Normalize all the X data

```
In [222...   scaler = StandardScaler()
             X_scale = scaler.fit_transform(X)
```

4.) Run a Lasso with lambda of .5. Plot a bar chart.

```
In [225...   lasso_model = Lasso(alpha = .01)
             lasso_model.fit(X_scale,y)
             coefficients = lasso_model.coef_
```

```
In [226...   plt.figure(figsize = (18,5))
             plt.bar(range(len(coefficients)), coefficients, tick_label = X.columns)
             plt.axhline(0, color = "red")
             plt.show()
```


5.) Do these coefficient magnitudes make sense?

I think these coefficient maginitudes make sense. The three terms we presumed would not significantly impact earnings indeed have coefficients close to zero, while the six terms anticipated to affect earnings possess non-zero coefficients. The results indicate that the term 'New Phone' contributes most significantly to earnings predictions, which is a logical outcome.