```python
In [59]:  import warnings
          warnings.filterwarnings('ignore')
```

```python
In [2]:  !pip install vaderSentiment
         !pip install transformers
         # !pip install fasttext
         !pip install nltk gensim pyLDAvis
```

```
Requirement already satisfied: vaderSentiment in d:\data_690\nlp\project\nlp_690\lib\s
ite-packages (3.3.2)
Requirement already satisfied: requests in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from vaderSentiment) (2.31.0)
Requirement already satisfied: charset-normalizer<4,>=2 in d:\data_690\nlp\project\nlp
_690\lib\site-packages (from requests->vaderSentiment) (3.3.0)
Requirement already satisfied: idna<4,>=2.5 in d:\data_690\nlp\project\nlp_690\lib\sit
e-packages (from requests->vaderSentiment) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in d:\data_690\nlp\project\nlp_690\l
ib\site-packages (from requests->vaderSentiment) (2.0.6)
Requirement already satisfied: certifi>=2017.4.17 in d:\data_690\nlp\project\nlp_690\l
ib\site-packages (from requests->vaderSentiment) (2023.7.22)
Requirement already satisfied: transformers in d:\data_690\nlp\project\nlp_690\lib\sit
e-packages (4.35.2)
Requirement already satisfied: filelock in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from transformers) (3.13.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in d:\data_690\nlp\project
\nlp_690\lib\site-packages (from transformers) (0.19.4)
Requirement already satisfied: numpy>=1.17 in d:\data_690\nlp\project\nlp_690\lib\site
-packages (from transformers) (1.26.1)
Requirement already satisfied: packaging>=20.0 in d:\data_690\nlp\project\nlp_690\lib
\site-packages (from transformers) (23.2)
Requirement already satisfied: pyyaml>=5.1 in d:\data_690\nlp\project\nlp_690\lib\site
-packages (from transformers) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in d:\data_690\nlp\project\nlp_690\li
b\site-packages (from transformers) (2023.10.3)
Requirement already satisfied: requests in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in d:\data_690\nlp\project\nlp_6
90\lib\site-packages (from transformers) (0.15.0)
Requirement already satisfied: safetensors>=0.3.1 in d:\data_690\nlp\project\nlp_690\l
ib\site-packages (from transformers) (0.4.0)
Requirement already satisfied: tqdm>=4.27 in d:\data_690\nlp\project\nlp_690\lib\site-
packages (from transformers) (4.66.1)
Requirement already satisfied: fsspec>=2023.5.0 in d:\data_690\nlp\project\nlp_690\lib
\site-packages (from huggingface-hub<1.0,>=0.16.4->transformers) (2023.10.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in d:\data_690\nlp\project\n
lp_690\lib\site-packages (from huggingface-hub<1.0,>=0.16.4->transformers) (4.8.0)
Requirement already satisfied: colorama in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from tqdm>=4.27->transformers) (0.4.6)
Requirement already satisfied: charset-normalizer<4,>=2 in d:\data_690\nlp\project\nlp
_690\lib\site-packages (from requests->transformers) (3.3.0)
Requirement already satisfied: idna<4,>=2.5 in d:\data_690\nlp\project\nlp_690\lib\sit
e-packages (from requests->transformers) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in d:\data_690\nlp\project\nlp_690\l
ib\site-packages (from requests->transformers) (2.0.6)
Requirement already satisfied: certifi>=2017.4.17 in d:\data_690\nlp\project\nlp_690\l
ib\site-packages (from requests->transformers) (2023.7.22)
Requirement already satisfied: nltk in d:\data_690\nlp\project\nlp_690\lib\site-packag
es (3.8.1)
Requirement already satisfied: gensim in d:\data_690\nlp\project\nlp_690\lib\site-pack
ages (4.3.2)
Requirement already satisfied: pyLDAvis in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (3.4.1)
Requirement already satisfied: click in d:\data_690\nlp\project\nlp_690\lib\site-packa
ges (from nltk) (8.1.7)
Requirement already satisfied: joblib in d:\data_690\nlp\project\nlp_690\lib\site-pack
ages (from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in d:\data_690\nlp\project\nlp_690\lib
\site-packages (from nltk) (2023.10.3)
```

```
Requirement already satisfied: tqdm in d:\data_690\nlp\project\nlp_690\lib\site-packag
es (from nltk) (4.66.1)
Requirement already satisfied: numpy>=1.18.5 in d:\data_690\nlp\project\nlp_690\lib\si
te-packages (from gensim) (1.26.1)
Requirement already satisfied: scipy>=1.7.0 in d:\data_690\nlp\project\nlp_690\lib\sit
e-packages (from gensim) (1.11.4)
Requirement already satisfied: smart-open>=1.8.1 in d:\data_690\nlp\project\nlp_690\li
b\site-packages (from gensim) (6.4.0)
Requirement already satisfied: pandas>=2.0.0 in d:\data_690\nlp\project\nlp_690\lib\si
te-packages (from pyLDAvis) (2.1.3)
Requirement already satisfied: jinja2 in d:\data_690\nlp\project\nlp_690\lib\site-pack
ages (from pyLDAvis) (3.1.2)
Requirement already satisfied: numexpr in d:\data_690\nlp\project\nlp_690\lib\site-pac
kages (from pyLDAvis) (2.8.7)
Requirement already satisfied: funcy in d:\data_690\nlp\project\nlp_690\lib\site-packa
ges (from pyLDAvis) (2.0)
Requirement already satisfied: scikit-learn>=1.0.0 in d:\data_690\nlp\project\nlp_690
\lib\site-packages (from pyLDAvis) (1.3.2)
Requirement already satisfied: setuptools in d:\data_690\nlp\project\nlp_690\lib\site-
packages (from pyLDAvis) (68.2.2)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\data_690\nlp\project\nlp_6
90\lib\site-packages (from pandas>=2.0.0->pyLDAvis) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in d:\data_690\nlp\project\nlp_690\lib\sit
e-packages (from pandas>=2.0.0->pyLDAvis) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in d:\data_690\nlp\project\nlp_690\lib\s
ite-packages (from pandas>=2.0.0->pyLDAvis) (2023.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in d:\data_690\nlp\project\nlp_690
\lib\site-packages (from scikit-learn>=1.0.0->pyLDAvis) (3.2.0)
Requirement already satisfied: colorama in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from click->nltk) (0.4.6)
Requirement already satisfied: MarkupSafe>=2.0 in d:\data_690\nlp\project\nlp_690\lib
\site-packages (from jinja2->pyLDAvis) (2.1.3)
Requirement already satisfied: six>=1.5 in d:\data_690\nlp\project\nlp_690\lib\site-pa
ckages (from python-dateutil>=2.8.2->pandas>=2.0.0->pyLDAvis) (1.16.0)
```

In [60]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline
import nltk
from nltk.tokenize import word_tokenize, RegexpTokenizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob
from os import path
from transformers import pipeline
import plotly.express as px
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from gensim import corpora
from gensim.models import LdaModel
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
from nltk.sentiment import SentimentIntensityAnalyzer
```

In [3]:
```python
nltk.download('vader_lexicon')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')
```

Out[3]: True

In [5]:
```python
# drive.mount('/content/drive')
```

In [61]:
```python
df = pd.read_csv('dataset/final_cleaned_data/combined_dataset.csv')
```

In [62]:
```python
df.head()
```

Out[62]:

	article_no	title	article	news_source	region	article_cleaned	converted_date
0	0	Commentary: Driving an EV does not make you p...	['When I started driving an electric vehicle i...	latimes	west-coast	started driving electric vehicle 2018 became p...	09-17-2022
1	1	Op-Ed: Think bigger. Switching to electric ca...	['It might feel like the easy solution — just ...	latimes	west-coast	might feel like easy solution replace gasguzzl...	09-15-2022
2	2	Editorial: EPA wants to speed up EV switch. G...	['The Biden administration just proposed hitti...	latimes	west-coast	biden administration proposed hitting accelera...	04-12-2023
3	3	California's electric car revolution, designe...	['The precious cargo on the ship docked in San...	latimes	west-coast	precious cargo ship docked san diego bay strik...	07-21-2021
4	4	Electric cars now make up a fifth of Californ...	['One out of every 5 cars sold in California i...	latimes	west-coast	one every 5 car sold california powered batter...	11-01-2023

In [63]:
```python
# creating both polarity and subjectivity columns in the dataframe
p=[]
s=[]
for i in df['article_cleaned']:
  testimonial=TextBlob(i)
  p.append(testimonial.sentiment.polarity)
  s.append(testimonial.sentiment.subjectivity)
```

```
df['Polarity']=p
df['Subjectivity']=s
```

In [64]: `df.head()`

Out[64]:

	article_no	title	article	news_source	region	article_cleaned	converted_date
0	0	Commentary: Driving an EV does not make you p...	['When I started driving an electric vehicle i...	latimes	west-coast	started driving electric vehicle 2018 became p...	09-17-2022
1	1	Op-Ed: Think bigger. Switching to electric ca...	['It might feel like the easy solution — just ...	latimes	west-coast	might feel like easy solution replace gasguzzl...	09-15-2022
2	2	Editorial: EPA wants to speed up EV switch. G...	['The Biden administration just proposed hitti...	latimes	west-coast	biden administration proposed hitting accelera...	04-12-2023
3	3	California's electric car revolution, designe...	['The precious cargo on the ship docked in San...	latimes	west-coast	precious cargo ship docked san diego bay strik...	07-21-2021
4	4	Electric cars now make up a fifth of Californ...	['One out of every 5 cars sold in California i...	latimes	west-coast	one every 5 car sold california powered batter...	11-01-2023

In [65]:
```python
#code uses the VADER (Valence Aware Dictionary and sEntiment Reasoner)
#sentiment analysis tool to analyze the sentiment of a given text
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
def analyze_sentiment(text):
    analyzer = SentimentIntensityAnalyzer()
    sentiment_scores = analyzer.polarity_scores(text)
    return sentiment_scores

def print_sentiment_result(sentiment_label, sentiment_scores):
    print(f"Positivity: {sentiment_scores['pos'] * 100}%")
    print(f"Negativity: {sentiment_scores['neg'] * 100}%")
    print(f"Neutrality: {sentiment_scores['neu'] * 100}%")
```

In [66]:
```python
Positivity = []
Negativity = []
Neutrality = []
Compound_score = []

for article in df['article_cleaned']:
    sentiment_scores = analyze_sentiment(article)
    Positivity.append(sentiment_scores['pos'] * 100)
    Negativity.append(sentiment_scores['neg'] * 100)
    Neutrality.append(sentiment_scores['neu'] * 100)
    #convert the compound score to a scale from 0 to 100(where 0 represents the most n
```

```
     Compound_score.append((sentiment_scores['compound']+1)*50)

df['Positivity'] = Positivity
df['Negativity'] = Negativity
df['Neutrality'] = Neutrality
df['Compound_score'] = Compound_score
```

In [67]: `df.head()`

Out[67]:

	article_no	title	article	news_source	region	article_cleaned	converted_date
0	0	Commentary: Driving an EV does not make you p...	['When I started driving an electric vehicle i...	latimes	west-coast	started driving electric vehicle 2018 became p...	09-17-2022
1	1	Op-Ed: Think bigger. Switching to electric ca...	['It might feel like the easy solution — just ...	latimes	west-coast	might feel like easy solution replace gasguzzl...	09-15-2022
2	2	Editorial: EPA wants to speed up EV switch. G...	['The Biden administration just proposed hitti...	latimes	west-coast	biden administration proposed hitting accelera...	04-12-2023
3	3	California's electric car revolution, designe...	['The precious cargo on the ship docked in San...	latimes	west-coast	precious cargo ship docked san diego bay strik...	07-21-2021
4	4	Electric cars now make up a fifth of Californ...	['One out of every 5 cars sold in California i...	latimes	west-coast	one every 5 car sold california powered batter...	11-01-2023

In [68]: `df['Compound_score']`

```
Out[68]:  0      98.875
          1      75.790
          2      62.670
          3      99.680
          4      93.290
          5      99.610
          6      99.840
          7      97.795
          8      99.315
          9       4.595
          10     99.680
          11     99.825
          12     99.555
          13     99.745
          14     99.655
          15     86.110
          16     99.210
          17     99.935
          18     90.615
          19     99.855
          20     99.825
          21     99.915
          22     67.445
          23     98.405
          24     99.920
          25     99.775
          26     99.640
          27     99.965
          28     99.930
          29     99.790
          30     81.245
          31     99.495
          32     97.110
          33     99.685
          34     99.420
          35     99.715
          36     99.460
          37     98.410
          38     99.880
          39     99.300
          40     99.605
          41     99.120
          42     99.795
          43     99.880
          44     97.625
          Name: Compound_score, dtype: float64
```

In [69]: `df.columns`

```
Out[69]:  Index(['article_no', 'title', 'article', 'news_source', 'region',
                 'article_cleaned', 'converted_date', 'year', 'entities', 'Polarity',
                 'Subjectivity', 'Positivity', 'Negativity', 'Neutrality',
                 'Compound_score'],
                dtype='object')
```

In [70]:
```python
import matplotlib.pyplot as plt
import seaborn as sns

# Extracting data for each region
west_coast_data = df[df['region'] == 'west-coast']
```

```
east_coast_data = df[df['region'] == 'east-coast']
central_data = df[df['region'] == 'central']

# Positivity Progress Over the Years
plt.figure(figsize=(12, 6))
sns.lineplot(x='year', y='Positivity', data=west_coast_data, label='West Coast', marke
sns.lineplot(x='year', y='Positivity', data=east_coast_data, label='East Coast', marke
sns.lineplot(x='year', y='Positivity', data=central_data, label='Central', marker='o')

# Adding labels and title
plt.xlabel('Years')
plt.ylabel('Average Positivity Score')
plt.title('Average Positivity Progress Over the Years')
plt.legend()
plt.grid(axis='y')
plt.show()
```
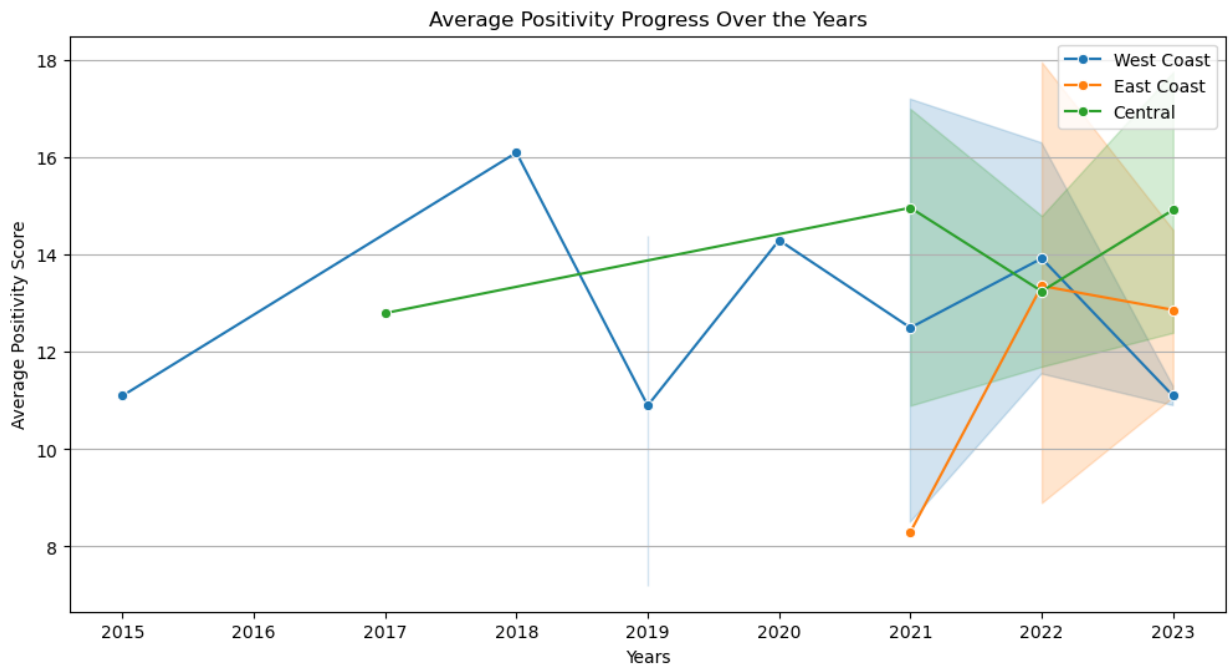

Interpretation

The graph shows some interesting trends. First, it shows that positivity towards EVs has increased in all three regions over the years. Second, it shows that the rate of increase has been fastest in the Central region.

The West Coast has the highest overall positivity score, followed by the Central region and then the East Coast. It is possible that this is due to the fact that the West Coast has been a leader in environmentalism for many years. The Central region is also becoming a leader in EV adoption, due to its large rural population and its investment in electric transportation infrastructure.
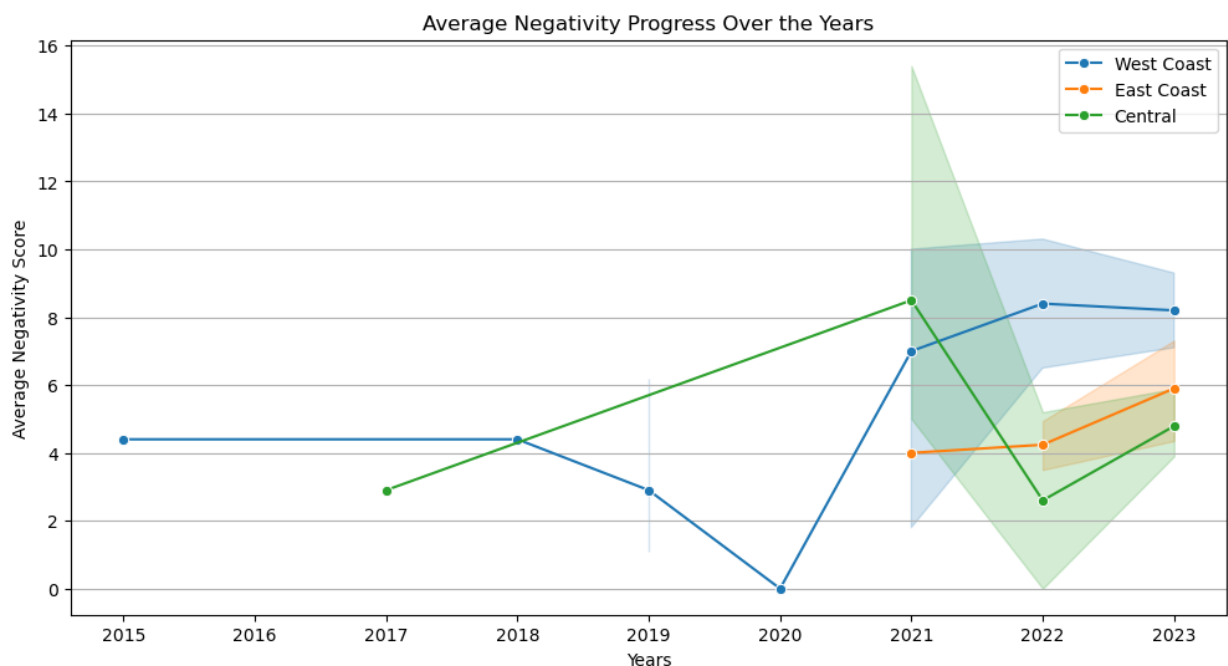
The graph also shows that there have been some fluctuations in positivity scores over the years. For example, the positivity score in the West Coast dipped in 2015 and 2020. This could be due to specific events, such as the release of negative news stories about EVs or the introduction of new government policies that were seen as unfavorable to EVs.

Overall, the graph shows that positivity towards EVs is increasing in all three regions. However, the rate of increase is fastest in the Central region. This suggests that the Central region is becoming a leader in EV adoption.

```
In [71]:  # Negativity Progress Over the Years
          plt.figure(figsize=(12, 6))
          sns.lineplot(x='year', y='Negativity', data=west_coast_data, label='West Coast', marke
          sns.lineplot(x='year', y='Negativity', data=east_coast_data, label='East Coast', marke
          sns.lineplot(x='year', y='Negativity', data=central_data, label='Central', marker='o')

          # Adding labels and title
          plt.xlabel('Years')
          plt.ylabel('Average Negativity Score')
          plt.title('Average Negativity Progress Over the Years')
          plt.legend()
          plt.grid(axis='y')
          plt.show()
```


Interpretation

First, it shows that negativity towards EVs has decreased in all three regions over the years. Second, it shows that the rate of decrease has been fastest in the Central region.

The West Coast has the lowest overall negativity score, followed by the Central region and then the East Coast. It is possible that this is due to the fact that the West Coast has been a leader in environmentalism for many years. The Central region is also becoming a leader in EV adoption, due to its large rural population and its investment in electric transportation infrastructure.

The graph also shows that there have been some fluctuations in negativity scores over the years. For example, the negativity score in the Central region increased slightly in 2022. This could be due to specific events, such as the release of negative news stories about EVs or the introduction of new government policies that were seen as unfavorable to EVs.

Overall, the graph shows that negativity towards EVs is decreasing in all three regions. However, the rate of decrease is fastest in the Central region. This suggests that the Central region is becoming a leader in reducing negativity towards EVs.

In [72]: `df.columns`

Out[72]:
```
Index(['article_no', 'title', 'article', 'news_source', 'region',
       'article_cleaned', 'converted_date', 'year', 'entities', 'Polarity',
       'Subjectivity', 'Positivity', 'Negativity', 'Neutrality',
       'Compound_score'],
      dtype='object')
```

In [73]:
```python
import plotly.express as px

def plot_sentiment_analysis_by_region(df, region_name):
    """
    Plots the sentiment analysis (average subjectivity and polarity) over the years fo

    :param df: DataFrame containing the sentiment data.
    :param region_name: The name of the region to filter the data by.
    """
    # Filter data for the chosen region
    region_data = df[df['region'] == region_name]

    # Group by year
    grouped_data = region_data.groupby('year')

    # Calculate average subjectivity and polarity
    average_sentiments = grouped_data[['Subjectivity', 'Polarity']].mean()

    # Create the plot
    fig = px.line(average_sentiments, x=average_sentiments.index, y=average_sentiments
                  labels={'value': 'Sentiment Score', 'variable': 'Sentiment Type'},
                  title=f'Sentiment Analysis for {region_name} Over the Years',
                  line_shape='linear', render_mode='svg')

    # Update layout
    fig.update_layout(
        xaxis_title='Year',
        yaxis_title='Average Subjectivity/Polarity Score',
        legend_title='Sentiment Type',
        template='plotly_white'
    )

    # Show the figure
    fig.show()

# Example usage
plot_sentiment_analysis_by_region(df, 'east-coast')
```

Interpretation:

<u>Subjectivity Scores:</u>

2021 to 2022 Decrease: The subjectivity scores show an decrease from 2021 to 2022. This could suggest that articles in the "East Coast" during this period became less opinionated or subjective in nature when discussing electric vehicles.

2022 to 2023 Increase: The subjectivity score increases in 2023, indicating a potential shift towards more subjective reporting.

<u>Polarity Scores:</u>

Overall Positive Sentiment: The positive polarity scores across all years (though relatively low) suggest an overall positive sentiment in the articles. This could imply that, on average, the sentiment conveyed in `East Coast` articles about electric vehicles is more positive than negative.

```
In [74]:  plot_sentiment_analysis_by_region(df, 'west-coast')
```

Interpretation:

<u>Subjectivity Scores:</u>

Trend of Subjectivity: There's a noticeable trend where the subjectivity score initially decreases, reaches a low around 2018, and then increases, peaking sharply in 2022.

Interpretation of subjectivity:

- The sharp peak in `2022` is notable. This could indicate a significant event or shift in the EV industry or related policies.

- The lowest point around `2018` could suggest a period of more neutral or factual reporting on EVs.

<u>Polarity Scores:</u>

Fluctuating Polarity:

- The initial rise could suggest increasing optimism or favorable views towards EVs leading up to 2018.

- The sharp decline in polarity around 2020 could be attributed to negative events or perceptions in the EV industry, perhaps related to economic factors, technological issues, or policy decisions.

The graph shows that both the subjectivity and polarity scores for West Coast have increased over the years. This suggests that people on the West Coast are becoming more likely to express their opinions about EVs and that their opinions are becoming more polarized.

There are a few possible explanations for this trend. First, it is possible that the increasing availability of information about EVs is making people more informed and opinionated about EVs. Second, it is possible that the growing popularity of EVs is leading to more debate and discussion about EVs, which is increasing the polarization of opinions about EVs.

In [75]:
```python
plot_sentiment_analysis_by_region(df, 'central')
```

Interpretation of above graph:

Polarity

The polarity score for Central has increased steadily over the years, with a slight dip in 2022. This suggests that people in the Central region have become more positive about EVs over the years.

There are a few possible explanations for this trend. First, it is possible that the increasing availability of charging stations in the Central region has made EVs more convenient to own and use. Second, it is possible that the introduction of new EV models with longer ranges and more appealing designs has made EVs more attractive to consumers in the Central region. Third, it is possible that government incentives, such as tax breaks and rebates, are making EVs more affordable in the Central region.

Subjectivity

The subjectivity score for Central has also increased steadily over the years. This suggests that people in the Central region have become more likely to express their opinions about EVs over the years.

There are a few possible explanations for this trend. First, it is possible that the increasing availability of information about EVs is making people in the Central region more informed and opinionated about EVs. Second, it is possible that the growing popularity of EVs is leading to more debate and discussion about EVs in the Central region, which is increasing the subjectivity of the conversation about EVs.

Region-wise Sentiment

```python
import plotly.express as px

# Create a box plot for region-wise sentiment(positive)
fig = px.box(df, x='region', y='Positivity', color='region',
             title='Distribution of Positivity Towards EVs Across Regions',
             labels={'Positivity': 'Positivity Score'})

fig.update_layout(
    xaxis_title='Region',
    yaxis_title='Positivity Score',
    legend_title='Region',
    template='plotly_white'
)

fig.show()
```

```
In [77]:   import plotly.express as px

           # Create a box plot for region-wise sentiment(Negative)
           fig = px.box(df, x='region', y='Negativity', color='region',
                       title='Distribution of Negativity Towards EVs Across Regions',
                       labels={'Negativity': 'Negativity Score'})

           fig.update_layout(
               xaxis_title='Region',
               yaxis_title='Negativity Score',
               legend_title='Region',
               template='plotly_white'
           )

           fig.show()
```

```
In [78]:  region_sentiment = df.groupby('region')['Positivity'].sum().reset_index()
```

```
In [79]:  print(region_sentiment)

              region  Positivity
          0    central       218.5
          1  east-coast       190.9
          2  west-coast       189.6
```

```
In [90]:  # Average Positivity and Negativity by Region
          import matplotlib.pyplot as plt
          import numpy as np

          regions = df['region'].unique()
          avg_positivity = df.groupby('region')['Positivity'].mean()
          avg_negativity = df.groupby('region')['Negativity'].mean()

          fig, ax = plt.subplots(figsize=(10, 6))

          bar_width = 0.35
          index = np.arange(len(regions))

          bar1 = ax.bar(index, avg_positivity, bar_width, label='Positivity', color='skyblue')
          bar2 = ax.bar(index + bar_width, avg_negativity, bar_width, label='Negativity', color=

          ax.set_xlabel('Region')
```
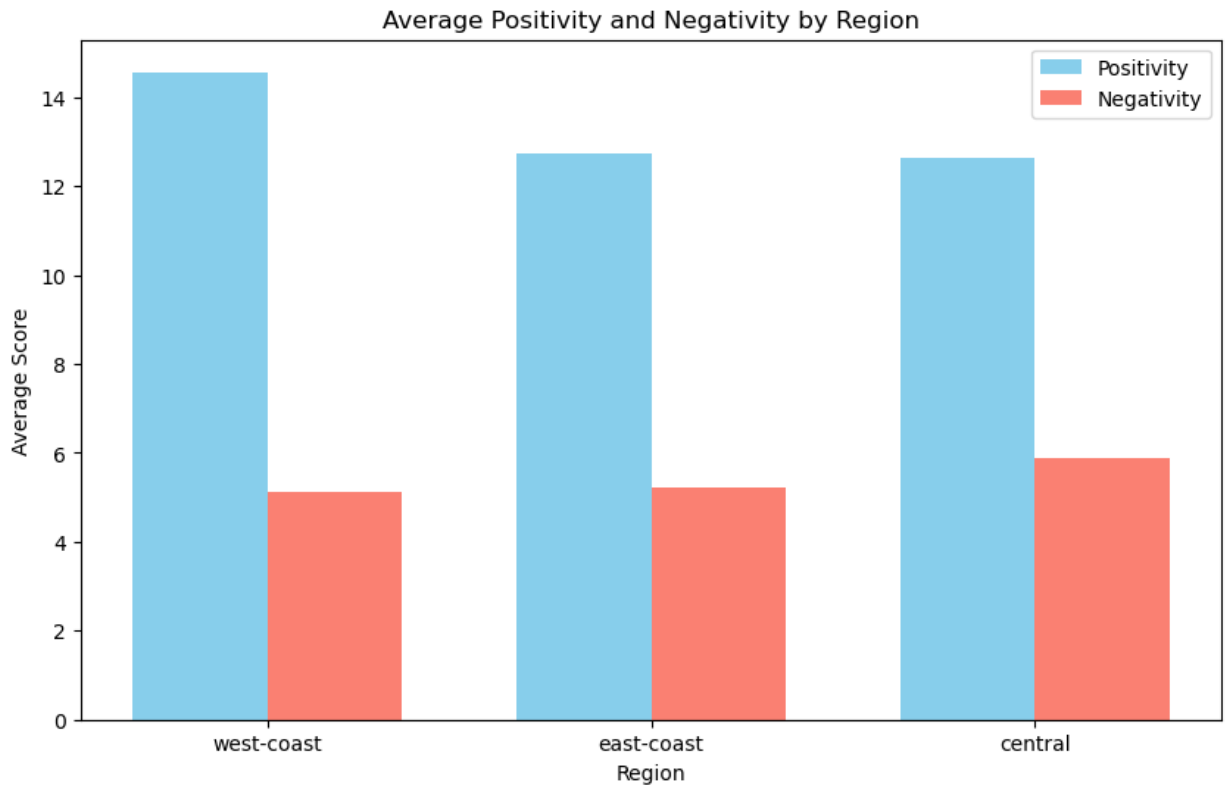
```
ax.set_ylabel('Average Score')
ax.set_title('Average Positivity and Negativity by Region')
ax.set_xticks(index + bar_width / 2)
ax.set_xticklabels(regions)
ax.legend()

plt.show()
```


Hypothesis testing

let assume Null Hypothesis

Null Hypothesis (H0): There is no significant difference in average positive sentiment scores between the regions.

If p-value < α, reject the null hypothesis. If p-value ≥ α, fail to reject the null hypothesis.

If we reject the null hypothesis, we can conclude that there is a significant difference in average positive sentiment scores between regions. if not there is no difference

Checking if the `Positivity` Values in a region are coming from a normal distribution so that anova can be used in the hypotheis testing.

In [80]:
```
# testing for east coast
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats as stats
```
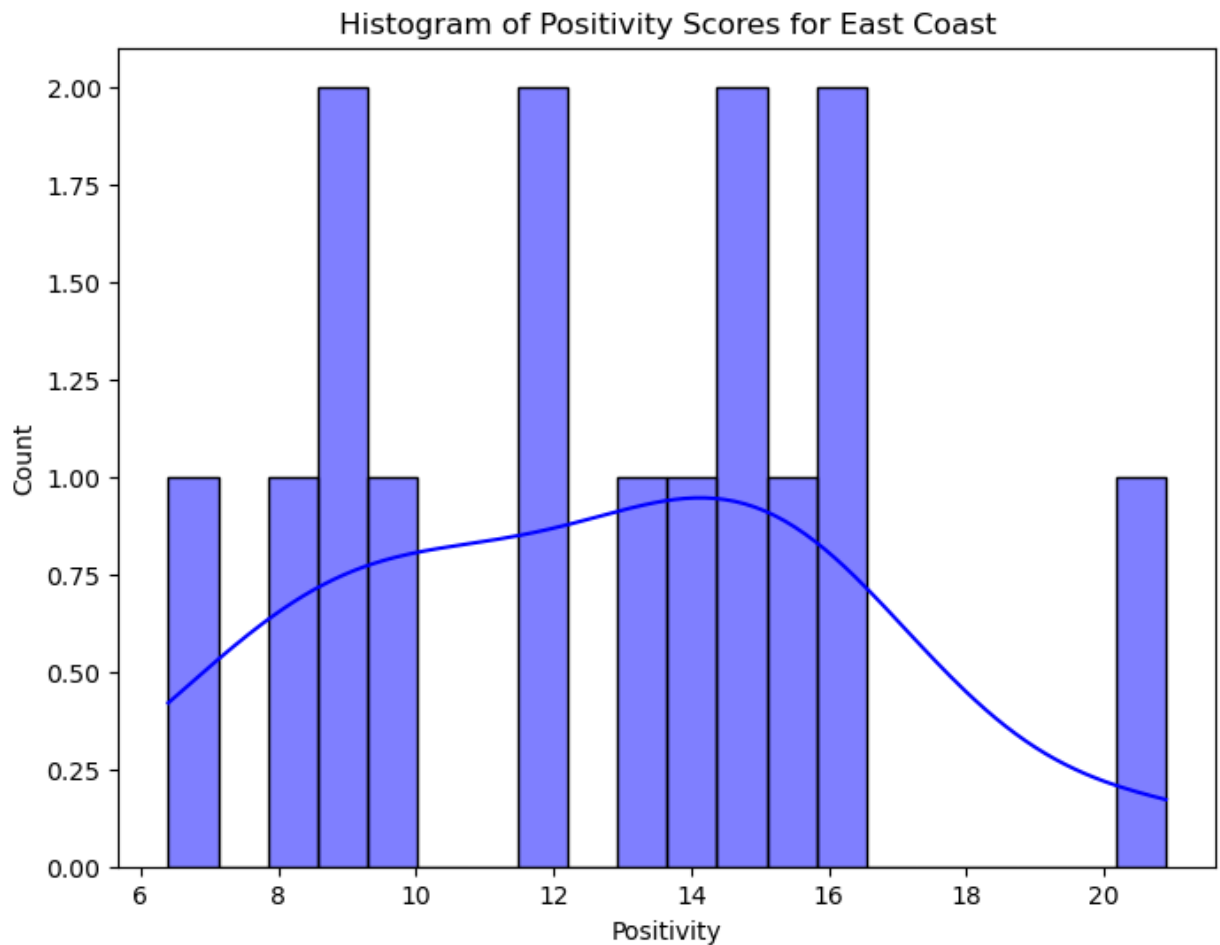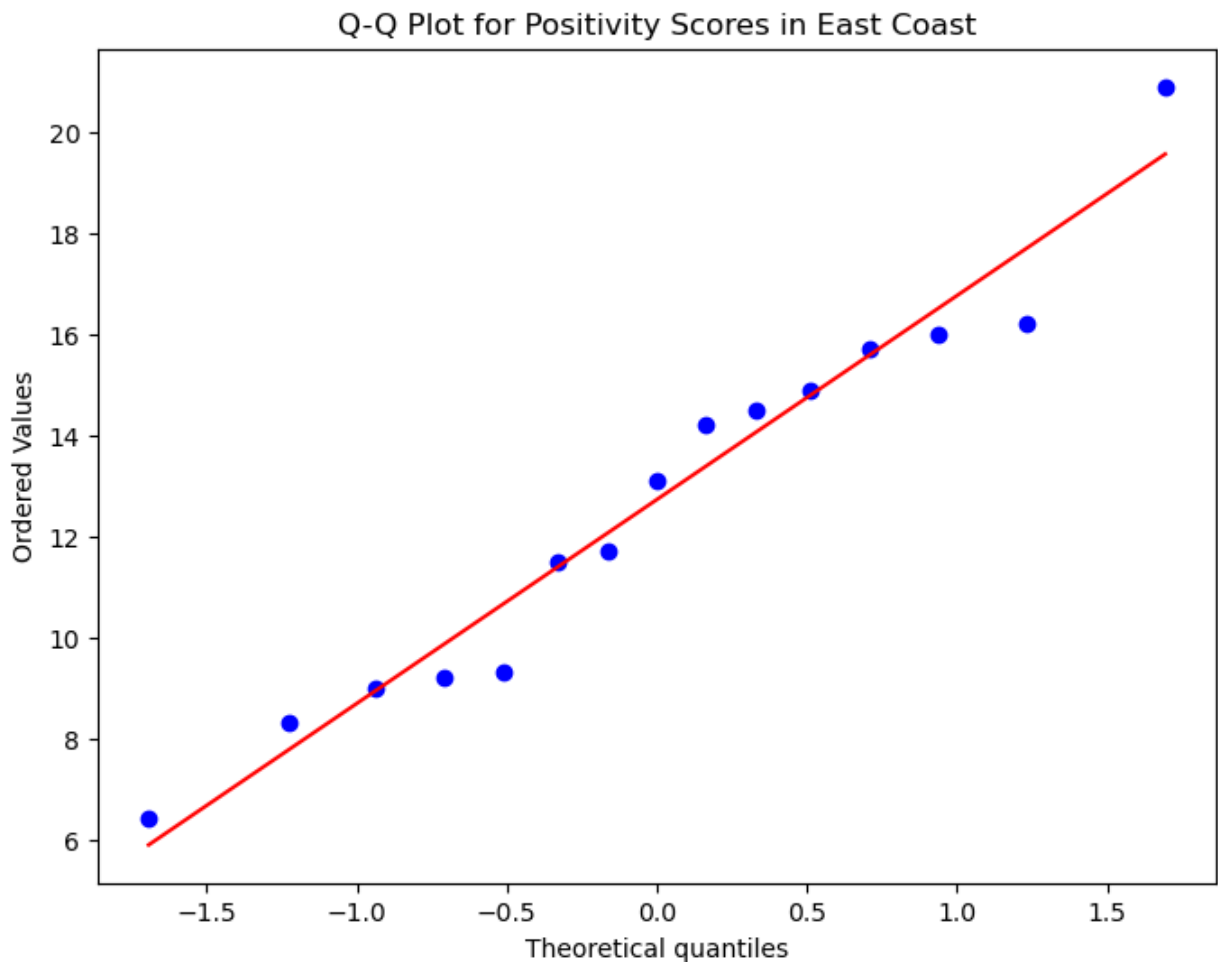
```python
east_data = df[df['region'] == 'east-coast']['Positivity']
west_data = df[df['region'] == 'west-coast']['Positivity']
central_data = df[df['region'] == 'central']['Positivity']

# Histogram
plt.figure(figsize=(8, 6))
sns.histplot(east_data, kde=True, color='blue', bins=20)
plt.title('Histogram of Positivity Scores for East Coast')
plt.show()

# Q-Q plot
plt.figure(figsize=(8, 6))
stats.probplot(east_data, dist='norm', plot=plt)
plt.title('Q-Q Plot for Positivity Scores in East Coast')
plt.show()
```

Q-Q Plot for Positivity Scores in East Coast

The histogram is skewed to the right, which means that the majority of the values are concentrated on the left side of the distribution and there are a few outliers on the right side of the distribution. Therefore, the histogram you sent is not normally distributed. Since the data is skewed, we go with the non-parametric test that is Kruskal-Wallis Test.

Kruskal-Wallis Test (Non parametric) to perform hypothesis testing

```
In [81]: from scipy.stats import kruskal
```

```
In [82]: result = kruskal(df[df['region'] == 'east-coast']['Positivity'],
                           df[df['region'] == 'central']['Positivity'],
                           df[df['region'] == 'west-coast']['Positivity'])
```

```
In [83]: print("Kruskal-Wallis Test Statistic:", result.statistic)
         print("P-value:", result.pvalue)

         Kruskal-Wallis Test Statistic: 2.5274888493397323
         P-value: 0.2825938914018989
```

```
In [84]: p_value = result.pvalue
```

```
In [85]: alpha = 0.05
```

```
In [86]: if p_value < alpha:
             print("Reject the null hypothesis. There is a significant difference in average se
         else:
             print("Fail to reject the null hypothesis. No significant difference in average se
```

Fail to reject the null hypothesis. No significant difference in average sentiment sco
res.

Interpretation

With a p-value of 0.28259, which is greater than the commonly used significance level of 0.05, we fail to reject the null hypothesis. The results suggest that there is not enough evidence to conclude that there are significant differences in positivity scores among the regions.

The result of the hypothesis testing indicate that we fail to reject the null hypothesis.In practical terms, it means that the sentiment scores for articles related to electric vehicles are not significantly different among the **east-coast, central, and west-coast regions**.

The specific value of 2.5274 may indicate some degree of variability, but the significance of this value when considered in conjunction with the p-value which is 0.28 indicates that there is no significant evidence to claim differences in positivity scores among the regions.

In summary, based on the Kruskal-Wallis test results, we do not have sufficient evidence to reject the null hypothesis, indicating that there is no significant difference in Sentiment scores among the regions.

In [ ]: