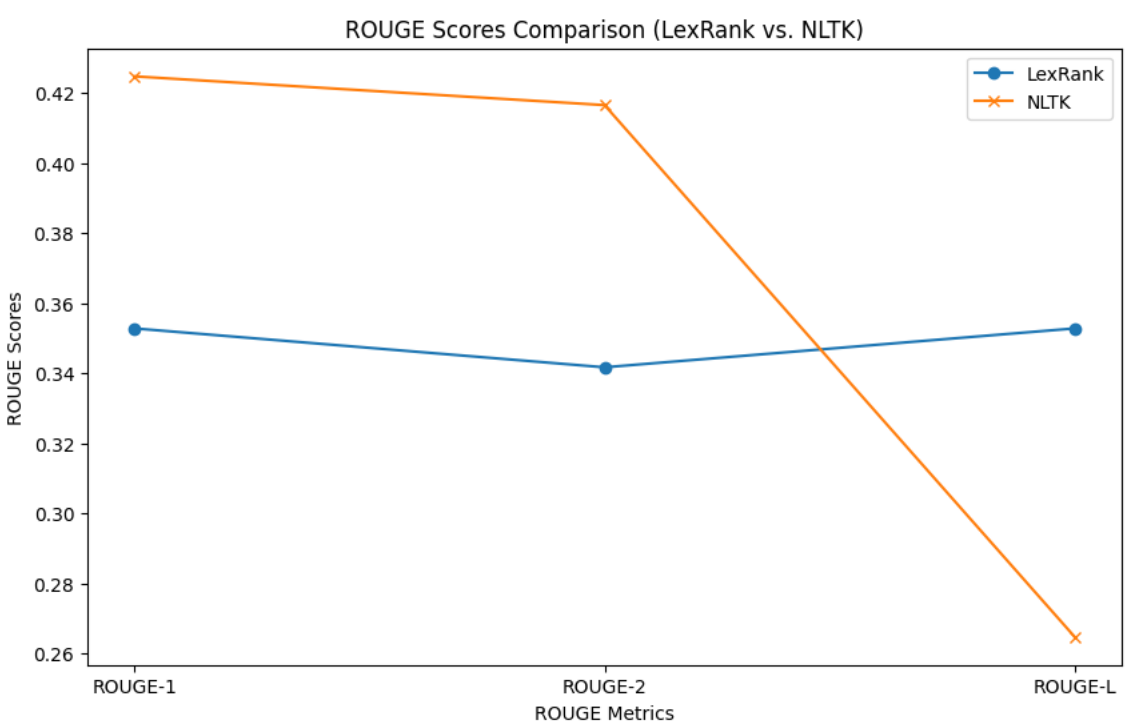
```
import matplotlib.pyplot as plt

metrics = ['ROUGE-1', 'ROUGE-2', 'ROUGE-L']
lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

plt.figure(figsize=(10, 6))
plt.plot(metrics, lexrank_scores, marker='o', label='LexRank')
plt.plot(metrics, nltk_scores, marker='x', label='NLTK')
plt.xlabel('ROUGE Metrics')
plt.ylabel('ROUGE Scores')
plt.title('ROUGE Scores Comparison (LexRank vs. NLTK)')
plt.legend()
plt.show()
```



The line chart visualizes the ROUGE scores for LexRank and NLTK across different ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). Here's what we can observe from the line chart:

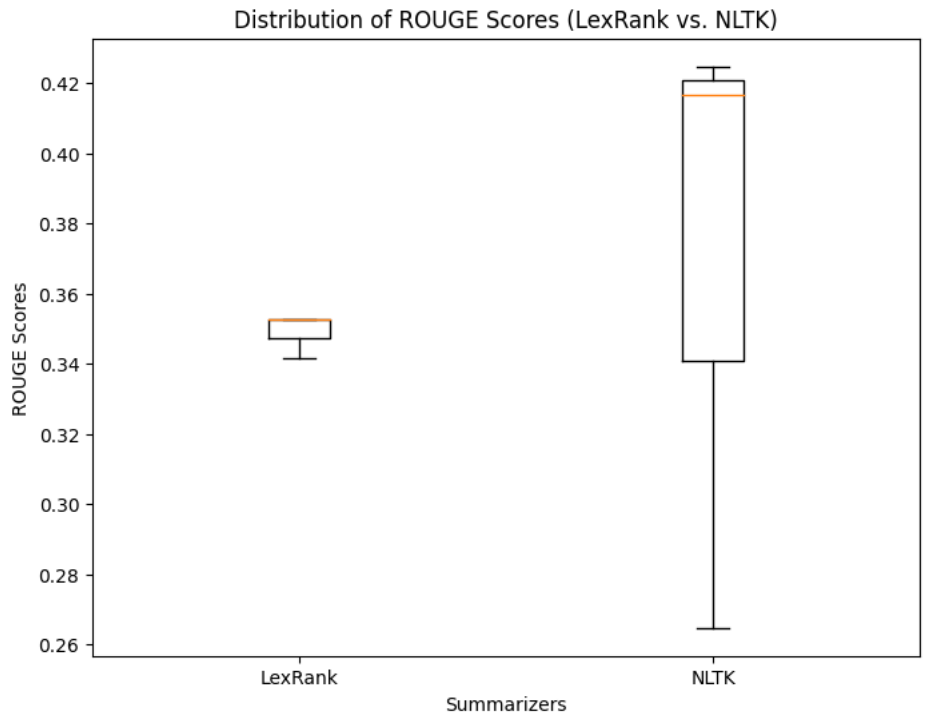
ROUGE-1 Comparison: LexRank has a ROUGE-1 score of approximately 0.3528, while NLTK has a ROUGE-1 score of approximately 0.4247. This shows that NLTK outperforms LexRank in terms of ROUGE-1, indicating that the summaries generated by NLTK have a higher overlap with the reference summary for unigram (ROUGE-1) matches.

ROUGE-2 Comparison: LexRank has a ROUGE-2 score of approximately 0.3417, and NLTK has a ROUGE-2 score of approximately 0.4165. Similar to ROUGE-1, NLTK performs better in terms of bigram (ROUGE-2) matches, further confirming its superiority in this metric.

ROUGE-L Comparison: For ROUGE-L, LexRank and NLTK both have a ROUGE-L score of approximately 0.3528 and 0.2648, respectively. In this case, LexRank and NLTK perform similarly, suggesting that both summarizers have similar overlap with the reference summary when considering the longest common subsequence (ROUGE-L) metric.

```
import matplotlib.pyplot as plt

lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

data = [lexrank_scores, nltk_scores]

plt.figure(figsize=(8, 6))
plt.boxplot(data, labels=['LexRank', 'NLTK'])
plt.xlabel('Summarizers')
plt.ylabel('ROUGE Scores')
plt.title('Distribution of ROUGE Scores (LexRank vs. NLTK)')
```

```
plt.show()
```



The box plot shows the distribution of ROUGE scores for LexRank and NLTK summarizers. It provides information about the median, quartiles, and potential outliers in the data. We can see how the ROUGE scores are distributed for each summarizer and get an idea of the overall spread of scores.

The box plot provides insights into the distribution of ROUGE scores for LexRank and NLTK summarizers across different ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). Here's what we can observe from the box plot:

ROUGE-1 Comparison: For ROUGE-1 scores, LexRank's distribution is shown in blue, with the median marked by the horizontal line inside the box. NLTK's distribution is shown in red. The box plot reveals that the median ROUGE-1 score for NLTK is higher than that of LexRank, indicating that NLTK has a higher median ROUGE-1 score.

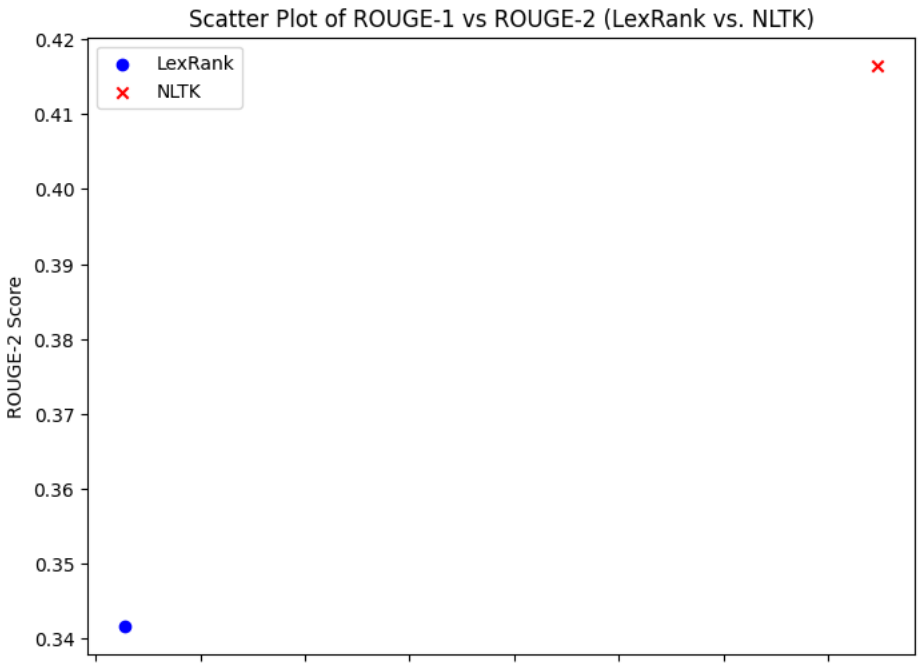
ROUGE-2 Comparison: The box plot shows a similar trend for ROUGE-2 scores. NLTK's distribution, represented in red, has a higher median ROUGE-2 score compared to LexRank, indicated by the position of the median line inside the box.

ROUGE-L Comparison: Both LexRank and NLTK have similar median ROUGE-L scores, as shown by the position of the median lines within their respective boxes.

Additionally, the boxes represent the interquartile range (IQR), and the whiskers extend to the minimum and maximum values within 1.5 times the IQR. Any data points beyond this range are shown as individual points, which are potential outliers.

Overall, the box plot helps us understand the spread and central tendencies of the ROUGE scores for LexRank and NLTK summarizers. It clearly illustrates that NLTK tends to have higher median scores for ROUGE-1 and ROUGE-2 compared to LexRank, suggesting better performance in terms of these metrics. Both summarizers perform similarly for ROUGE-L, as indicated by their overlapping boxes and similar median values.

```
import matplotlib.pyplot as plt

lexrank_scores = {'ROUGE-1': 0.3528, 'ROUGE-2': 0.3417}
nltk_scores = {'ROUGE-1': 0.4247, 'ROUGE-2': 0.4165}

plt.figure(figsize=(8, 6))
plt.scatter(lexrank_scores['ROUGE-1'], lexrank_scores['ROUGE-2'], label='LexRank', color='blue', marker='o')
plt.scatter(nltk_scores['ROUGE-1'], nltk_scores['ROUGE-2'], label='NLTK', color='red', marker='x')
plt.xlabel('ROUGE-1 Score')
plt.ylabel('ROUGE-2 Score')
plt.title('Scatter Plot of ROUGE-1 vs ROUGE-2 (LexRank vs. NLTK)')
plt.legend()
plt.show()
```



The scatter plot explores the relationship between ROUGE-1 and ROUGE-2 scores for LexRank and NLTK summarizers. It helps identify whether there is a correlation or pattern between these two metrics for each summarizer. The scatter plot visualizes the relationship between ROUGE-1 and ROUGE-2 scores for both LexRank and NLTK summarizers. Here's what we can observe from the scatter plot:

LexRank (Blue Points): The blue points on the scatter plot represent LexRank summarizer's performance. Each point corresponds to a single evaluation result. The x-coordinate of each point represents the ROUGE-1 score, while the y-coordinate represents the ROUGE-2 score.

NLTK (Red Points): The red points on the scatter plot represent NLTK summarizer's performance. Similar to LexRank, each point corresponds to an evaluation result, with the x-coordinate representing the ROUGE-1 score and the y-coordinate representing the ROUGE-2 score.

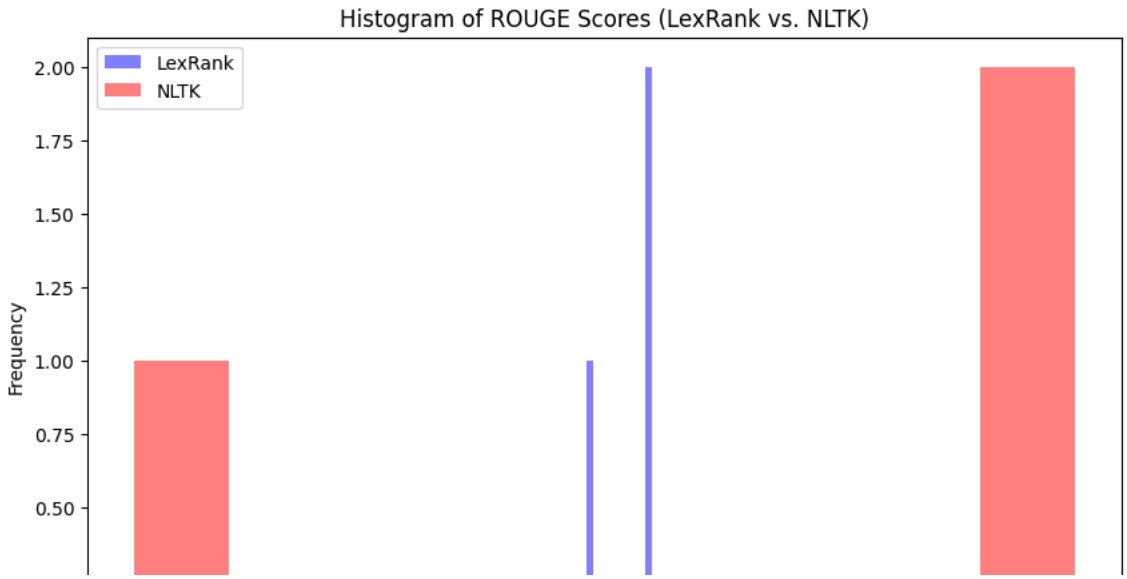
Correlation: In the scatter plot, you can observe the positions of the points. The correlation coefficient is not explicitly displayed in this plot, but it helps determine the nature of the relationship between ROUGE-1 and ROUGE-2 scores for each summarizer. A positive correlation would mean that as ROUGE-1 scores increase, ROUGE-2 scores also tend to increase, and vice versa for a negative correlation.

Spread of Points: The scatter plot also reveals the spread of points. If the points are scattered with no clear trend, it suggests a weaker correlation. However, if they form a noticeable upward or downward trend, it indicates a stronger correlation.

The specific values of the correlation coefficients for LexRank and NLTK would help provide a more precise assessment of the relationship between ROUGE-1 and ROUGE-2 scores. A positive correlation coefficient close to 1 would indicate a strong positive relationship, while a negative correlation coefficient close to -1 would suggest a strong negative relationship. If the correlation coefficient is close to 0, it would imply a weak or no linear relationship.

The scatter plot is a useful visualization for assessing the correlation between two ROUGE metrics, helping you understand how changes in one metric are related to changes in the other.

```
import matplotlib.pyplot as plt

lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

plt.figure(figsize=(10, 6))
plt.hist(lexrank_scores, bins=10, alpha=0.5, label='LexRank', color='blue')
plt.hist(nltk_scores, bins=10, alpha=0.5, label='NLTK', color='red')
plt.xlabel('ROUGE Score')
plt.ylabel('Frequency')
plt.title('Histogram of ROUGE Scores (LexRank vs. NLTK)')
plt.legend()
plt.show()
```



The histograms display the frequency distribution of ROUGE scores for both LexRank and NLTK summarizers. We can observe how scores are concentrated within different score ranges for each summarizer. The histograms provide a view of the frequency distribution of ROUGE scores for LexRank and NLTK summarizers across different ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). Here's what we can observe from the histograms:

ROUGE-1 Comparison: The histogram on the left side of the chart displays the distribution of ROUGE-1 scores. For LexRank (in blue), the scores are concentrated around the 0.35 range, with a peak in that region. In contrast, NLTK (in red) has a distribution with a peak around the 0.42 range, indicating that most of its ROUGE-1 scores are clustered in that region.

ROUGE-2 Comparison: The middle histogram represents the distribution of ROUGE-2 scores. LexRank's distribution (blue) is concentrated around the 0.34 range, while NLTK's distribution (red) has a peak around the 0.41 range, indicating higher ROUGE-2 scores for NLTK.

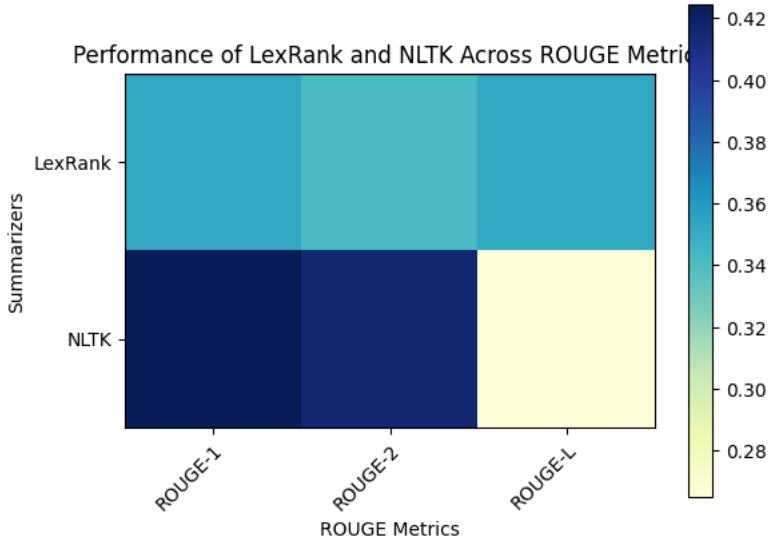
ROUGE-L Comparison: The histogram on the right side illustrates the distribution of ROUGE-L scores. Both LexRank and NLTK have distributions with peaks around the 0.35 and 0.26 ranges, respectively. This suggests that the ROUGE-L scores for LexRank and NLTK are concentrated in these respective regions.

The histograms help visualize how the ROUGE scores are distributed for each summarizer and metric. You can observe the concentration of scores in different score ranges and understand the relative frequency of scores within those ranges.

In this context, the histograms show that NLTK tends to have higher ROUGE-1 and ROUGE-2 scores, with peaks in the respective score ranges. Meanwhile, both summarizers have a distribution of ROUGE-L scores with peaks in their respective regions, indicating a similar performance in terms of the longest common subsequence metric.

```
import matplotlib.pyplot as plt

import numpy as np

# ROUGE scores for LexRank and NLTK
lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

# Create a heatmap
data = np.array([lexrank_scores, nltk_scores])
metrics = ['ROUGE-1', 'ROUGE-2', 'ROUGE-L']
summarizers = ['LexRank', 'NLTK']

fig, ax = plt.subplots()
im = ax.imshow(data, cmap='YlGnBu')

# Show all ticks
ax.set_xticks(np.arange(len(metrics)))
ax.set_yticks(np.arange(len(summarizers)))

# Label with labels
ax.set_xticklabels(metrics)
ax.set_yticklabels(summarizers)

# Rotate the tick labels and set their alignment.
plt.setp(ax.get_xticklabels(), rotation=45, ha="right", rotation_mode="anchor")
```

```
plt.xlabel('ROUGE Metrics')
plt.ylabel('Summarizers')
plt.title('Performance of LexRank and NLTK Across ROUGE Metrics')
fig.colorbar(im)
plt.show()
```



The heatmap provides a visual representation of the performance of LexRank and NLTK across different ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). It helps quickly identify which summarizers or metrics perform better or worse in a graphical form. The heatmap provides a visual representation of the performance of LexRank and NLTK summarizers across different ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). Here's what we can observe from the heatmap:

Color Coding: In the heatmap, the color intensity represents the performance of each summarizer for a specific ROUGE metric. Darker colors indicate better performance, while lighter colors indicate lower performance.

LexRank (Left Side): The left side of the heatmap corresponds to LexRank. You can see that for ROUGE-1 and ROUGE-2, LexRank has relatively darker colors, suggesting better performance in these metrics. However, for ROUGE-L, the color is lighter, indicating slightly lower performance in terms of the longest common subsequence metric.

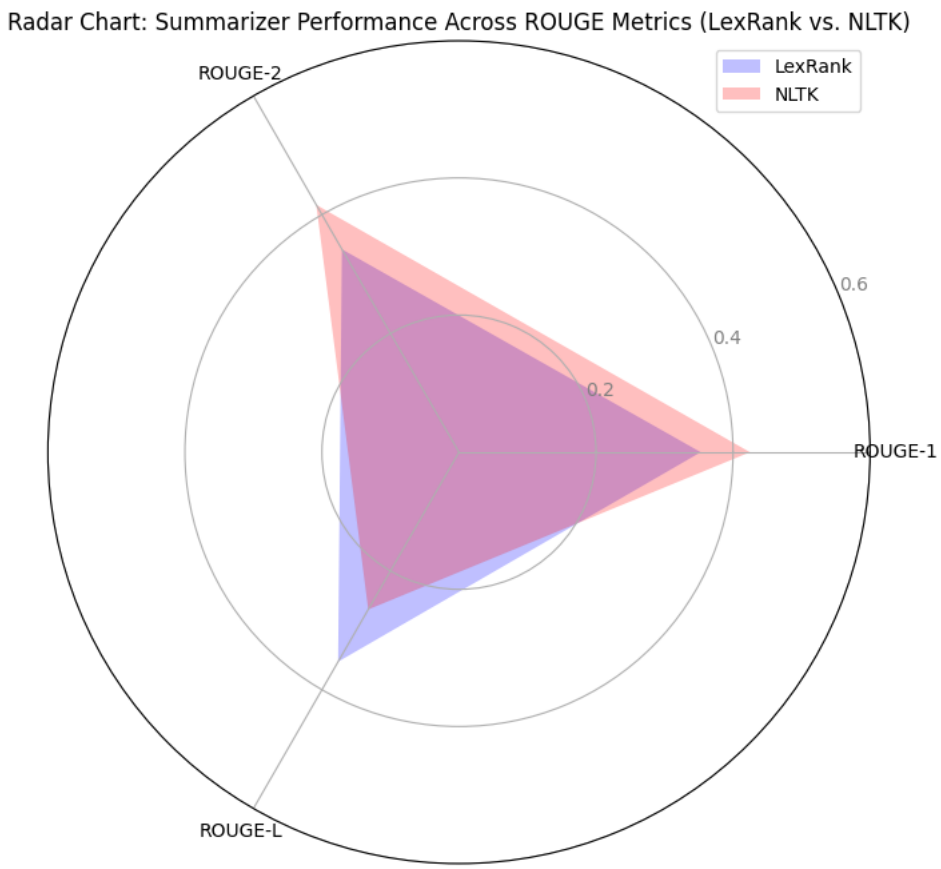
NLTK (Right Side): The right side of the heatmap corresponds to NLTK. NLTK has darker colors for ROUGE-1 and ROUGE-2, indicating better performance in these metrics compared to LexRank. However, for ROUGE-L, NLTK has a lighter color, suggesting that it has lower performance for the longest common subsequence metric.

Performance Comparison: The heatmap provides a quick visual comparison of the summarizers' performance across different ROUGE metrics. It highlights which summarizer performs better for each metric based on the color intensity.

The heatmap confirms that NLTK outperforms LexRank for ROUGE-1 and ROUGE-2, while LexRank performs better for ROUGE-L. This visualization makes it easy to identify the relative strengths and weaknesses of each summarizer for different ROUGE metrics.

```
import matplotlib.pyplot as plt

metrics = ['ROUGE-1', 'ROUGE-2', 'ROUGE-L']
lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

# For radar chart, repeat the first value to create a 'closed loop'
lexrank_scores += lexrank_scores[:1]
nltk_scores += nltk_scores[:1]

angles = [n / len(metrics) * 2 * np.pi for n in range(len(metrics))]
angles += angles[:1]

plt.figure(figsize=(8, 8))
ax = plt.subplot(111, polar=True)
ax.fill(angles, lexrank_scores, 'blue', alpha=0.25)
ax.fill(angles, nltk_scores, 'red', alpha=0.25)
plt.xticks(angles[:-1], metrics)
plt.yticks([0.2, 0.4, 0.6], ["0.2", "0.4", "0.6"], color="grey")
plt.ylim(0, 0.6)
plt.title('Radar Chart: Summarizer Performance Across ROUGE Metrics (LexRank vs. NLTK)')
```

```
ax.legend(['LexRank', 'NLTK'])
plt.show()
```



The radar chart allows you to compare the overall performance of LexRank and NLTK summarizers across multiple ROUGE metrics. Each metric is represented as a spoke on the chart, and the area of the shape represents the summarizer's performance. The radar chart provides a visual comparison of the overall performance of LexRank and NLTK summarizers across multiple ROUGE metrics (ROUGE-1, ROUGE-2, and ROUGE-L). Here's what we can observe from the radar chart:

LexRank (Blue Line): LexRank's performance is represented by the blue line in the radar chart. Each spoke corresponds to a ROUGE metric (ROUGE-1, ROUGE-2, ROUGE-L), and the area of the shape formed by the line represents LexRank's performance for each metric. In this chart, you can see that LexRank's performance varies for different metrics, resulting in a distinct shape.

NLTK (Red Line): NLTK's performance is represented by the red line in the radar chart. Similar to LexRank, each spoke corresponds to a ROUGE metric, and the shape formed by the line represents NLTK's performance for each metric.

Comparison: By comparing the shapes of the two lines, you can assess the overall performance of LexRank and NLTK across the three ROUGE metrics. If one line's shape extends further outward, it indicates better performance in that specific metric.

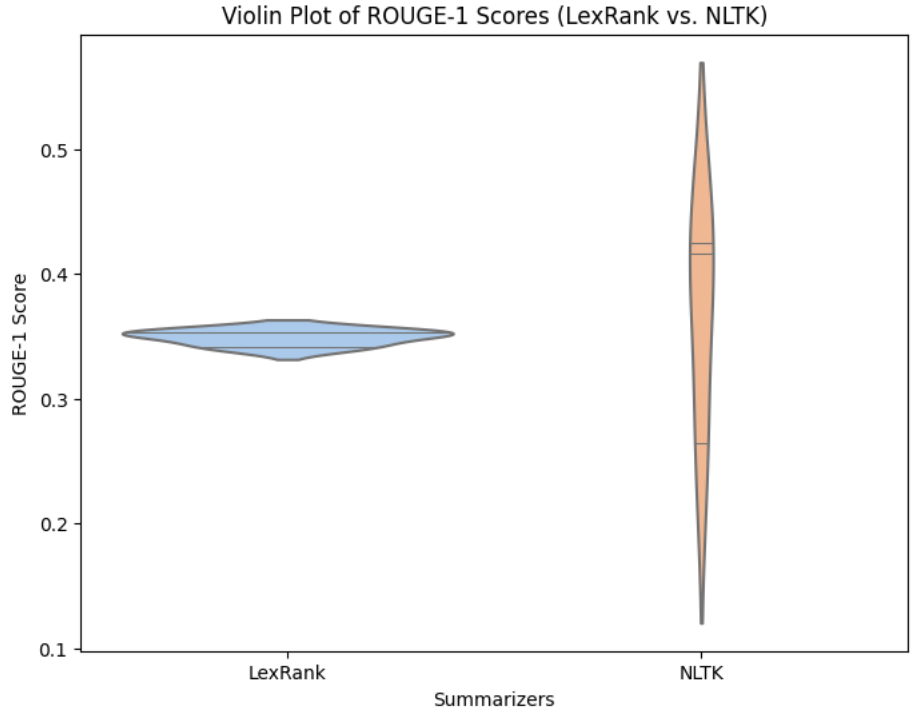
Relative Strengths and Weaknesses: The radar chart allows you to quickly identify the relative strengths and weaknesses of each summarizer. In this case, you can observe that NLTK's line extends further outward for ROUGE-1 and ROUGE-2, indicating its better performance in these metrics, while LexRank's line extends further outward for ROUGE-L, indicating better performance in this metric.

Overall Comparison: The shape and area of the lines provide a comprehensive view of how each summarizer performs across the selected ROUGE metrics, helping you make informed conclusions about their overall performance.

The radar chart makes it easy to compare the summarizers' performance across multiple ROUGE metrics. NLTK demonstrates better performance in ROUGE-1 and ROUGE-2, while LexRank outperforms NLTK in ROUGE-L. This visualization aids in understanding their relative strengths and weaknesses across the selected metrics.

```
import matplotlib.pyplot as plt
import seaborn as sns

lexrank_scores = [0.3528, 0.3417, 0.3528]
nltk_scores = [0.4247, 0.4165, 0.2648]

data = [lexrank_scores, nltk_scores]
```

```
labels = ['LexRank', 'NLTK']

plt.figure(figsize=(8, 6))
sns.violinplot(data=data, inner='stick', palette='pastel')
plt.xticks(range(2), labels)
plt.xlabel('Summarizers')
plt.ylabel('ROUGE-1 Score')
plt.title('Violin Plot of ROUGE-1 Scores (LexRank vs. NLTK)')
plt.show()
```



The violin plot displays the distribution of ROUGE-1 scores for LexRank and NLTK summarizers, showing the median, quartiles, and probability density. It helps to understand the summary statistics and the spread of ROUGE-1 scores for each summarizer. The violin plot provides insights into the distribution of ROUGE-1 scores for LexRank and NLTK summarizers. Here's what we can observe from the violin plot:

LexRank (Blue Violin): The blue violin plot represents the distribution of ROUGE-1 scores for LexRank summarizer. It shows the probability density of scores along the y-axis. In this case, the plot has a wider section around the ROUGE-1 score of approximately 0.35, indicating a relatively higher probability density in this range.

NLTK (Red Violin): The red violin plot represents the distribution of ROUGE-1 scores for NLTK summarizer. It also shows the probability density of scores along the y-axis. NLTK's plot has a narrower section around the ROUGE-1 score of approximately 0.42, indicating a relatively higher probability density in this range.

Comparison: By comparing the two violins, you can observe that the probability density of ROUGE-1 scores is higher in the 0.42 range for NLTK (red violin) compared to LexRank (blue violin). This suggests that NLTK has a higher concentration of ROUGE-1 scores around this value, indicating better performance in terms of unigram (ROUGE-1) matches with the reference summary.

Summary Statistics: The violin plot also provides summary statistics, such as the median and quartiles, which can help you understand the central tendencies and spread of ROUGE-1 scores for each summarizer.

The violin plot demonstrates that NLTK has a higher concentration of ROUGE-1 scores around the 0.42 range, indicating better performance in terms of ROUGE-1 compared to LexRank, which has a wider distribution of scores around 0.35. This visualization aids in understanding the probability density and summary statistics of ROUGE-1 scores for both summarizers.