Star Count Class

```
1  ## Author: Ashantha Rosary James
2  import matplotlib.pyplot as plt
3  from pyspark.sql import DataFrame
4  from pyspark.sql.functions import asc
5
6  class StarCountVisualizer:
7      def __init__(self, spark_df: DataFrame):
8          """
9          Initialize the visualizer with a Spark DataFrame.
10
11         Parameters:
12         spark_df (DataFrame): Spark DataFrame with a 'StarCount' column.
13         """
14         self.spark_df = spark_df
15
16     def prepare_data(self):
17         """
18         Group by StarCount, count occurrences, and collect data.
19         """
20         self.star_count_data = (
21             self.spark_df
22             .groupBy("StarCount")
23             .count()
24             .orderBy(asc("StarCount"))
25             .collect()
26         )
```

```python
    def plot_pie_chart(self):
        """
        Create and display a pie chart based on the collected data.
        """
        if not hasattr(self, 'star_count_data'):
            raise ValueError("Data not found. Run 'prepare_data()' first.")

        # Extract StarCount and counts from the collected data
        star_counts = [row['StarCount'] for row in self.star_count_data]
        counts = [row['count'] for row in self.star_count_data]

        # Define labels for the pie chart
        labels = [f'Star Count {x}' for x in star_counts]

        plt.figure(figsize=(8, 8))
        plt.pie(counts, labels=labels, autopct='%1.1f%%',
                colors=plt.cm.Paired(range(len(star_counts))), startangle=140)
        plt.title('Distribution of Star Count')
        plt.show()
```