DATA VISUALIZATION

How many reviews are in each sentiment class?

SQL Code:

```sql
SELECT sentiment, COUNT(*)  count_sentiment

FROM dsoaws_deep_learning.reviews

GROUP BY sentiment

ORDER BY sentiment DESC,  count_sentiment
```

Python code for visualization:

```python
import matplotlib.pyplot as plt

chart = df.plot.bar(

        x = "sentiment",

        y = "count_sentiment")

plt.xlabel("sentiment")

plt.show()
```

What is the distribution of review lengths?(number of words)

SQL code:

```sql
SELECT CARDINALITY(SPLIT(review_body, "  "))

FROM dsoaws_deep_learning.reviews
```

I split the review text in the column review body by a space character, which gives me a list of the individual words, and then I calculate the cardinality, which gives me the number of words.

Python code:

```python
Summary = df["num_words"].describe(

        Percentiles = [0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00])

Df["num_words"].plot.hist(
```

Xticks = [0, 16, 32, 64, 128, 256]. Bins = 100,

range = [0,256].axvline(x = summary["100%"], c = "red")

Before I plot the distribution, I want to calculate the percentiles. You can use the described function on the panels data frame that contains the review length, and you specify the percentiles to calculate. To visualize the distribution of review length, I choose a histogram. Histograms represent frequency distributions. They show how often each different value occurs. In this case, I want to group the review length in 100 bins, and I add a red marker to highlight the 100th percentile.