



<a href="#">&lt; Previous</a>	<div><div>📄</div><div>✓</div></div>	<div><div>📝</div><div>✓</div></div>	<div><div>📝</div></div>	<div><div>📄</div></div>	<a href="#">Next &gt;</a>
-------------------------------	-------------------------------------	-------------------------------------	-------------------------	-------------------------	---------------------------

## Case Study 2 Homework: Exercises 5-6

[🔖](#) Bookmark this page

Homework due Jul 13, 2021 22:00 +06

## Exercise 5

2/2 points (graded)

In Exercise 5, we will join all the data summaries for text Hamlet translation.

### Instructions

The previous code for summarizing a particular translation of Hamlet is consolidated into a single function called `summarize_text`. Create a `pandas` dataframe `grouped_data` consisting of the results of `summarize_text` for each translation of Hamlet in `hamlets`.

- Use a `for` loop across the row indices of `hamlets` to assign each translation to a new row.
- Obtain the `ith` row of `hamlets` to variables using the `.iloc` method, and assign the output to variables `language` and `text`.
- Call `summarize_text` using `language` and `text`, and assign the output to `sub_data`.
- Use the `pandas` `.append()` function to append `pandas` dataframes row-wise to `grouped_data`.

The code below defines `summarize_text`:

```
def summarize_text(language, text):
    counted_text = count_words_fast(text)

    data = pd.DataFrame({
        "word": list(counted_text.keys()),
        "count": list(counted_text.values())
    })

    data.loc[data["count"] > 10, "frequency"] = "frequent"
    data.loc[data["count"] <= 10, "frequency"] = "infrequent"
    data.loc[data["count"] == 1, "frequency"] = "unique"

    data["length"] = data["word"].apply(len)

    sub_data = pd.DataFrame({
        "language": language,
        "frequency": ["frequent", "infrequent", "unique"],
        "mean_word_length": data.groupby(by = "frequency")["length"].mean(),
        "num_words": data.groupby(by = "frequency").size()
    })

    return(sub_data)

# write your code here!
```

What is the average word length of the frequent words in the German translation?

4.528053



 Hide Notes

4.528053

How many frequent words are there in the Portugese translation?

261



261

Submit

You have used 1 of 10 attempts

✓ Correct (2/2 points)

## Exercise 6

1/1 point (graded)

In Exercise 6, we will plot our results and look for differences across each translation.

### Instructions

Plot the word statistics of each translation on a single plot. Note that we have already done most of the work for you. Consider whether the word statistics differ by translation.

This code will do most of the plotting work:

```
colors = {"Portuguese": "green", "English": "blue", "German": "red"}
markers = {"frequent": "o", "infrequent": "s", "unique": "^"}
import matplotlib.pyplot as plt
for i in range(grouped_data.shape[0]):
    row = grouped_data.iloc[i]
    plt.plot(row.mean_word_length, row.num_words,
             marker=markers[row.frequency],
             color = colors[row.language],
             markersize = 10
    )

color_legend = []
marker_legend = []
for color in colors:
    color_legend.append(
        plt.plot([], [],
                 color=colors[color],
                 marker="o",
                 label = color, markersize = 10, linestyle="None")
    )
for marker in markers:
    marker_legend.append(
        plt.plot([], [],
                 color="k",
                 marker=markers[marker],
                 label = marker, markersize = 10, linestyle="None")
    )
plt.legend(numpoints=1, loc = "upper left")

plt.xlabel("Mean Word Length")
plt.ylabel("Number of Words")
# write your code to display the plot here!
```

For which word category do the statistics differ most by translation?

☐ frequent

Hide Notes

☐ infrequent

☒ unique



Submit

You have used 1 of 1 attempt

✓ Correct (1/1 point)

◀ Previous

Next ▶

© All Rights Reserved



## edX

[About](#)

[Affiliates](#)

[edX for Business](#)

[Open edX](#)

[Careers](#)

[News](#)

## Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

[Accessibility Policy](#)

[Trademark Policy](#)

[Sitemap](#)

## Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Media Kit](#)

[Donate](#)



Hide Notes

