Team members:

Consider the following artificial dataset:

first_name	satisfaction	signup_date
Lee	high	01/03/2024
Aaliyah	med	05/30/2024
EMILY	high	02/15/2024
Sam	low	03/10/2024
omar	med	01/22/2024

Question 1
For each column, write the variable type and the most appropriate R storage type.

- `first_name`:

- `satisfaction`:

- `signup_date`:

Question 2
You should have identified that one of the columns would be best stored as a `factor`. Why would it be better stored as a `factor` rather than a `character`?

Question 3
You should have identified that one of the columns would be best stored as a `date`. Why would it be better stored as a `date` rather than a `character`?

Question 4
You should have identified that one of the columns would be best stored as a `character`. Why would it be better stored as a `character` rather than a `factor`?

Question 5

Complete the following R code to change each of the columns to the ideal R storage type.

```r
customers <-
  customers |>
  mutate(
    first_name =

    satisfaction =

    signup_date =
  )
```

Question 6

By default, variables stored as factors will have the levels stored in alphabetical order. Complete the R code below to convert the `satisfaction` variable to a factor with a more meaningful ordering and labels appropriate for a plot.

```r
customers <-
  customers |>
  mutate(_____ = factor(

    levels = c(_____, _____, _____),

    labels = c(_____, _____, _____)
  ))
```

Question 7

Describe what the following code seems to be doing.

```r
customers |>
  mutate(membership_length = ymd("2025-02-23") - signup_date) |>
  arrange(membership_length)
```

```
  first_name satisfaction signup_date membership_length
1    Aaliyah          med  2024-05-30          269 days
2        Sam          low  2024-03-10          350 days
3      EMILY         high  2024-02-15          374 days
4       omar          med  2024-01-22          398 days
5        Lee         high  2024-01-03          417 days
```

Question 8
Make a new column called `membership_month` which stores the month that the customer signed up. (Hint: use the `month()` function on the `signup_date` column.)

Question 9
Which of the columns from the original dataset could be described as a "string"?

Question 9
Describe what the following code seems to be doing.

```
customers |>
  filter(first_name != str_to_title(first_name)) |>
  pull(first_name)
```

```
[1] "EMILY" "omar"
```

Question 10
Oh no! It looks like names were not recorded with consistent capitalization. We want them to be recorded title case, meaning the first letter should be capital, followed by lower case letters. Write R code to store the names in title case. (Hint: use the `str_to_title()` function). Make sure to save the change back into the data frame.