**(I) Explore the data:**

Q1. Find the dimensions of the DF. Make the colnames lowercase.

Q2. Rename the column `fire\_size` to `acres\_burned`.

Q3. Convert the data in the `fire\_name` column such that the data starts with a capital letter. (ie `PIGEON LAKE` -> `Pigeon lake`)

Q4. Drop unnecessary rows- having fire size less than 10 acres.

Q5. Drop duplicate rows, if any. If so, keep the first duplicate of each such duplicate set.

Q6. Ensure that the `date` columns have the correct datatype.

Q7. Check for missing values in the `contain\_date` column. Find the number.

a) Check for missing values of `contain\_date` where fires exceed 10000 acres.

b) Check the values of `contain\_date` to see if they are reasonable.

Now explain whether you think the existing data is dependable or not, and whether you should/should not want to handle the missing data in `contain\_date`.

**(II) Prepare the data:**

Q9. Add the `fire\_month` and `days\_burning` columns

Q10. Get the statistical summary of the `days\_burning` column. What trends do you see in the 3rd quartile, mean and Max length of the fires? Is or isn't anything fishy?

Q11. Check the value of `days\_burning` for fires that exceed 100 acres. What do you notice about the acreage and the `days\_burning` for certain outliers?