Briefly

Your program will read data from a .txt file that lists first names, sex (F or M), and a count of how many babies had that name in 1996.

Once read into one or more lists, your program will repeatedly ask the user what name to search for, and will show all records with a matching name.

Once you have the basics working well with a small file, you will modify the program to read in the complete file, a little over 26,000 lines of data.

Next, extend your program to offer at least two ways to search the data

a) a complete match on name (you already have this)

   Example

       Search for Jesse

       Result Jesse   F 2686

               Jesse   M 1975

b) Another kind of search

     Top 50   Show 50 most popular girls names

              and 50 most popular boys names

or

c) Partial name match

       Name starts with Jes

       Result   Jessica F 11407

               Jesse   F 2686

               Jesuit F 7

               Jesse M 1975

               etc.

or

d) Name contains ana

       Result Briana F 12407

              Rosana F 512

              Roxana F 419

              Anabelle F 57

              Rexana M 15

or

e) Name ends with olly

       Result Holly F 1256

              Molly F 1156

              Dolly F 897

              Jolly M 567

              Zolly M 15

For full credit, you must provide two search options, two of those listed above, or some other variants that you have figured out (all names used by both girls and boys, perhaps).

Details

Explore the data

       What kind of sequence, if any?

       Does the data need any cleaning?

How to store the data

How can you store data to make your searching easier?

       Should you store names as

              Jordan   Capitalized

              jordan   lower case

              JORDAN   UPPER case

       How can you store 3 items of data:

              Name

              Sex   F or M

              Count   Number of babies with

                     this name

       A list of names, with each item a sublist of 3 items [name, sex, count]

       Parallel lists

              Names

              Sex

              Count

       Parallel lists, another variation

              Girl_Names

              Girl_Counts

              Boy_Names

              Boy_Counts

When you get search results, could it help to store them in a list, or set of parallel lists? Or should you just print the results as you get them?

If you provide an option to save your search results, make sure to provide something easy to read back in, such as a comma separated format

Jordan,M, 8765

or possibly, a fixed length format

Jordan   M   8765

```python
# Experiments with names
name = 'Jordan'
name3 = 'Anne'
look_for = 'An'
if look_for == name3:
    print("Found An")
else:
    print("Did not find An")
# end if
# Starts with?
if name3.startswith(look_for):
    print(name3, "starts with", look_for)
else:
    print(name3, "does not start with", look_for)
# end if
# Ends with?
if name.endswith(look_for):
    print(name, "ends with", look_for)
else:
    print(name, 'does not end with', look_for)
# end if
# Contained within?
if look_for in name:
    print(look_for, "is in", name)
else:
    print(look_for, "is not within", name)
# end if
# Now, go back and make name etc and look_for all lower-case...try the same tests again...
```

One way to read the data from a file:

```python
def read_names(file_name):
    ''' (str) -> list
        Reads records from text file
        Cleans each record, creating a line_list for each
        Appends each line_list to the name_list
        Returns name_list to caller
    '''
    # ------------------------------------------------------------
    # 'with' automatically 'closes' file at end of indented block
    #  process lines of data from file
    # ------------------------------------------------------------
    with open(file_name, "r") as my_file:

        name_list = []
        for line in my_file:

            line = line.strip() # strip newline char.

            # Split the line into a line_list, splitting on commas
            line_list = line.split(",")

            name = line_list[0]
            name = name.lower()        # standardize on lower case
            line_list[0] = name        # use cleaned-up name
            sex = line_list[1]
            count = int(line_list[2])
            line_list[2] = count       # put integer value into list

            name_list.append(line_list)
            #                    ||||T|||| append this little list
        #end for
    # end 'with'
    return name_list
# end def

…
name_list = read_names("yob1996short.txt") # Call the read_names function
```

```
US Baby name data from 1996

---------------------------------------------
26419 records read from baby name file

What kind of search?
F Full match
B Begins with
C Contains
E Ends with
R Range of names
T Top 50 names
S Sort results
X Exit: F

Type a name Jordan
7.763 milliseconds
Jordan .......... F       6,295
---------------------------
Jordan .......... M      13,848

distinct names               2
total names             20,143
```

Full name match

Note that any name could have both girls and boys with that name.

No real need to draw a line between girls and boys.

```
What kind of search?
F Full match
B Begins with
C Contains
E Ends with
R Range of names
T Top 50 names
S Sort results
X Exit: B

Type a name Jorda
16.605 milliseconds
Jordan .......... F       6,295
Jordana ......... F          57
Jordann ......... F          52
Jordanna ........ F          21
Jordanne ........ F          19
Jordain ......... F          13
Jordane ......... F           5
---------------------------
Jordan .......... M      13,848
Jordanlee ....... M           8
Jordain ......... M           6
Jordano ......... M           6
Jordanchristoph .. M          5
Jordanmichael .... M          5
Jordann ......... M           5
Jordany ......... M           5

distinct names              15
total names             20,350
```

"Begins with" name match

This kind of match usually gets a long list of results.

```
What kind of search?
F Full match
B Begins with
C Contains
E Ends with
R Range of names
T Top 50 names
S Sort results
X Exit: E

Type a name Avid
15.732 milliseconds
David ........... F          62
-------------------------------
David ........... M      22,983
Johndavid ........ M          20
Navid ............ M          15
Javid ........... M          11
Jonathandavid .... M           5

distinct names              6
total names            23,096
```

"Ends with" name match

This kind of match usually gets a long list of results.

Note that you can successfully compare 'avid' with 'David' using endswith()

Using lower case for both your name ('david') and your search term ('avid') will work.

```
What kind of search?
F Full match
B Begins with
C Contains
E Ends with
R Range of names
T Top 50 names
S Sort results
X Exit: T

Display top number of girls names, top number of boys names

How many of the most popular names (same # for girls, boys)? 10
0.018 milliseconds

Emily ............ F      25,146
Jessica .......... F      24,183
Ashley ........... F      23,675
Sarah ............ F      21,016
Samantha ......... F      20,545
Taylor ........... F      19,148
Hannah ........... F      18,590
Alexis ........... F      16,585
Rachel ........... F      16,098
Elizabeth ........ F      15,975
-------------------------------
Michael .......... M      38,361
Matthew .......... M      32,071
Jacob ............ M      31,897
Christopher ...... M      30,895
Joshua ........... M      29,167
Nicholas ......... M      27,714
Tyler ............ M      26,952
Brandon .......... M      25,848
Austin ........... M      25,656
Andrew ........... M      25,221

distinct names             20
total names           494,743
```

Top 50 (or Top "n") names

You have graduated and are starting a family; you want to look at the top 10, or top 50, or top 100 names in popularity.

Our data is listed by most popular girl's name to least popular, and the same for boys, so this is fairly straightforward to program.

5   Read data into list (or lists)

Read the full large file (26,000 names)

5   Clean data as it is read in

Store the data as either capitalized (Jordan) or lower case (jordan)

10   Display loop - Get request, show results, get another...

User selects search option, provides full or partial name to search for

Program displays results of search

10   Full match

5   Actually retrieves the correct names and their data (sex, count)
           Remember that many names get used by both girls and boys.

3   Nicely formatted print   Name   Sex   Count of how many have this name

2   Show total number of names, total number of people with this name somewhat like this

Melissa	F	2000
Melissa	M	15
Names		2
Total with this name	2015

10   Each additional matching method, same details as with Full Match

You need to do one additional matching method for a full 40 point grade.

Begins with

Names that begin with the same letters   Display name, sex, count for each
Searh for Dan
Results   Dan

                 Dana   etc.

Contains

Names that contain the search term   Display name, sex, count for each
Search for Dan
Results   Dan
                 Jordan

                 Jordana etc.

Ends with

Names that end with the search term   Display name, sex, count for each
Search for Dan
Results   Jordan

                 Sheridan

Top 50

The most popular names for girls, boys   Display name, sex, count for each
Search for top 3
Results   Emily
                 Jessica
                 Ashley
                 Michael   etc.