**REGEX fast**

https://docs.python.org/2/howto/regex.html

"Regular expressions (called REs, or regexes, or regex patterns) are essentially a tiny, highly specialized programming language embedded inside Python"

"For a detailed explanation of the computer science underlying regular expressions (deterministic and non-deterministic finite automata), you can refer to almost any textbook on writing compilers."

think rules for variables; compiler accepts or rejects variable names vs. reserved words

think vending machines; combinations of coins that equal a dollar

basically, acceptable strings or legal sentences

note: easy to accept things but hard to reject everything not acceptable

**cheat sheet:**

\d : matches any decimal digit; this is equivalent to the class [0-9]

\D : matches any non-digit character; this is equivalent to the class [^0-9]

\s : matches any whitespace character; this is equivalent to the class [ \t\n\r\f\v]

\S : matches any non-whitespace character; this is equivalent to the class [^ \t\n\r\f\v]

\w : matches any alphanumeric character; this is equivalent to the class [a-zA-Z0-9_]

\W : matches any non-alphanumeric character; this is equivalent to the class [^a-zA-Z0-9_]

^ : matches at the beginning of lines

. : matches anything except a newline character

? : matches previous character zero or one times

* : matches previous character zero or more times

ca*t will match ct, cat, caaat, etc.

+ : matches previous character one or more times

ca+t will match cat, caaat, but not ct

{m,n} : at least m repetitions, and at most n

{0,} is the same as *, {1,} is equivalent to +, and {0,1} is the same as ?

| : the "or" operator.

[^5] will match any character except '5', and [^^] will match any character except '^'

example: [\s,.] is a character class that will match any whitespace character, or ',' or '.'

example: expression for finding doubled words \b(\w+)\s+\1\b

**Exercises:**

load re3.py
open ph.csv

1. find all 800 numbers
2. find all 800 or 900 numbers
3. find all instances of the phone number 412-265-0997, 1-412-265-0997, 4122650997, 412 265 0997