**12-11 notes**

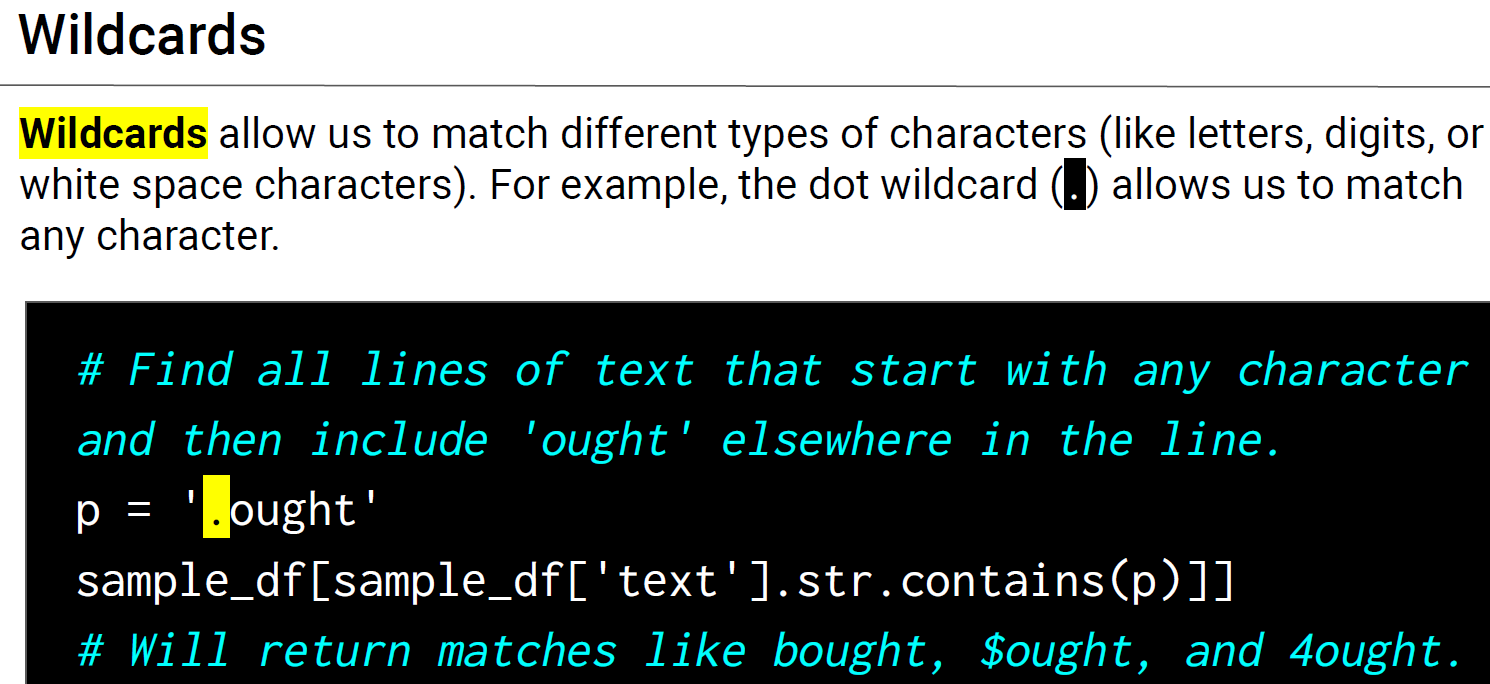
Transformation and Cleaning with Regular Expressions and ETL project work

Goals:

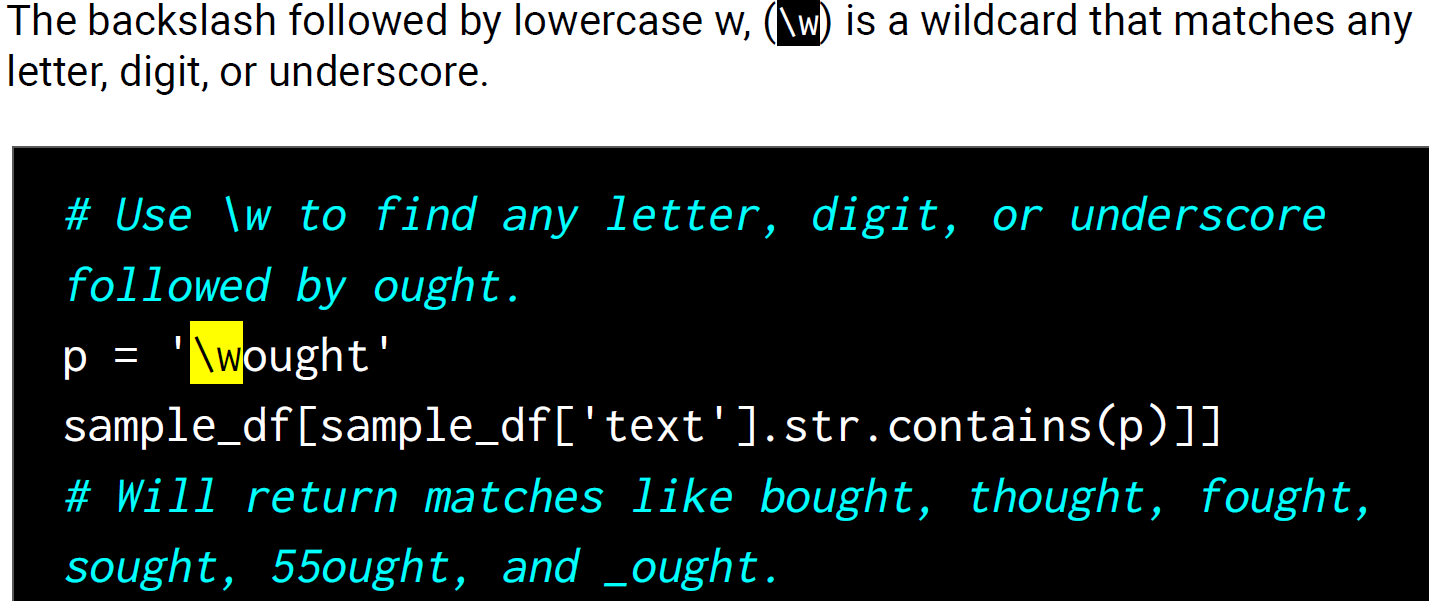
* Use wildcards, sets, and escape characters in regular expressions to extract information.
* Use special characters in regular expressions to extract information.
* Create and use capture groups in regular expressions to extract information.
* Continue working on the ETL mini project.

**Manish Jariwala**  [10:35 AM](https://msuvirtdatapt-wa75790.slack.com/archives/C05JFBCJHBM/p1702222548493819) Sunday 12/10/23 Slack Live Channel

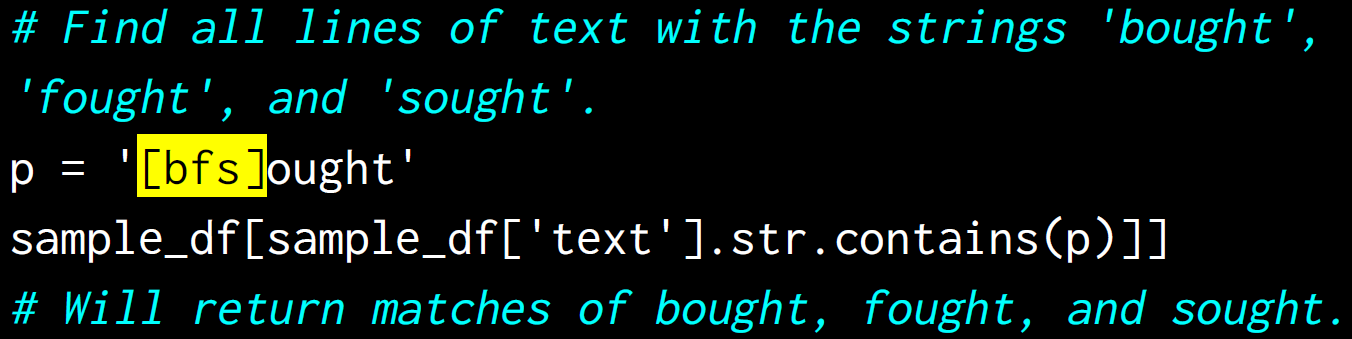
Project 2 - Due Dec 12th.  
Module 12th Challenge - Dec 19thModule 14th Challenge - Jan 4th



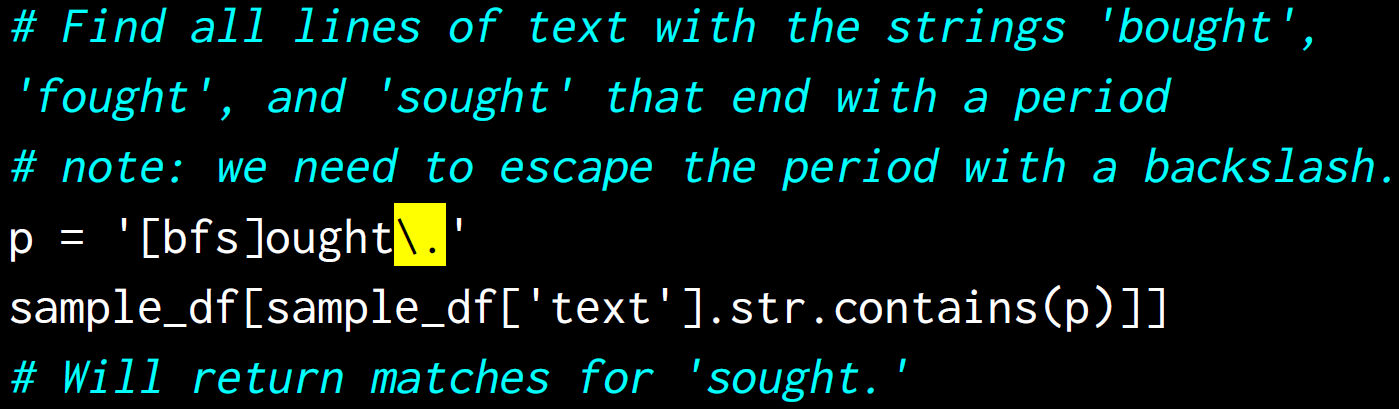
The backslash followed by lowercase w, (\w) is a wildcard that matches any letter, digit, or underscore.



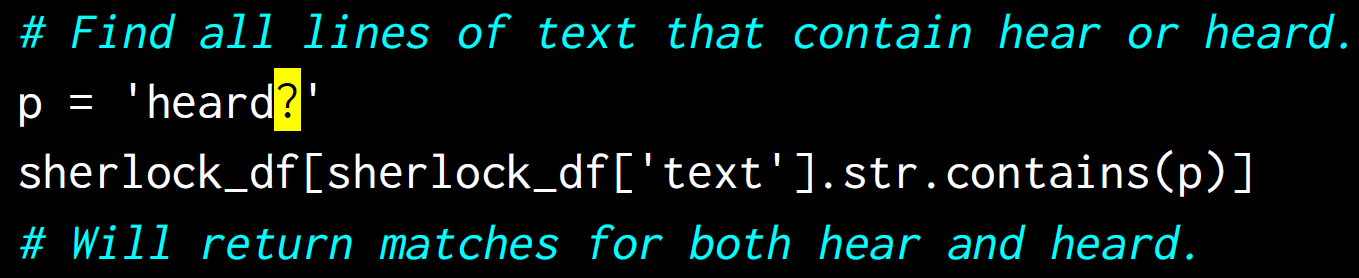
We can use brackets to match specific characters, which is called a set. For example, we can use the [bfs] set in [bfs]ought to match 'bought', 'fought', and 'sought'.



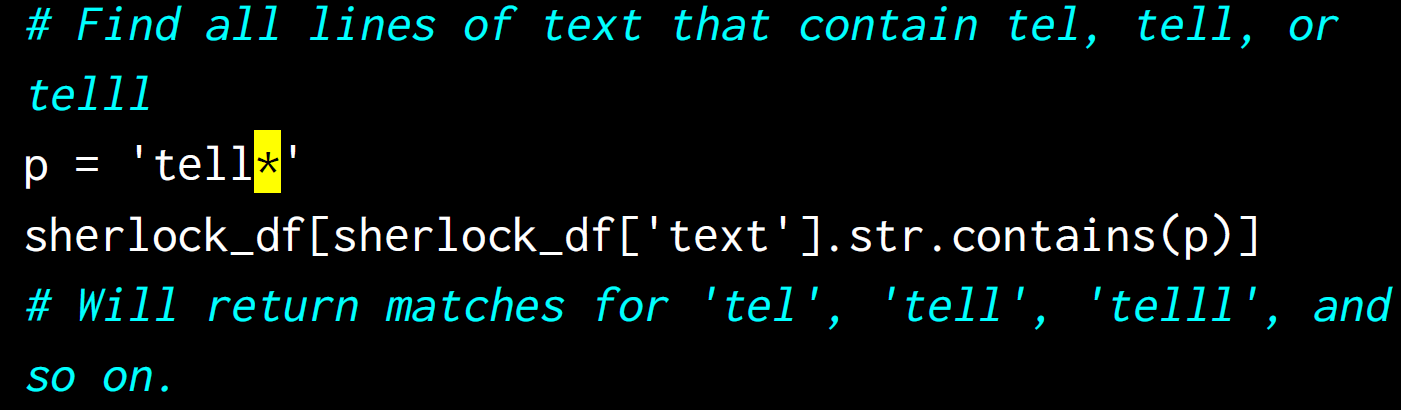
To match a character that also happens to be a regular expression character, we add a backslash (\) as an escape character. For example, to match a period (.) in a sentence, we use a backslash followed by a period (\.).



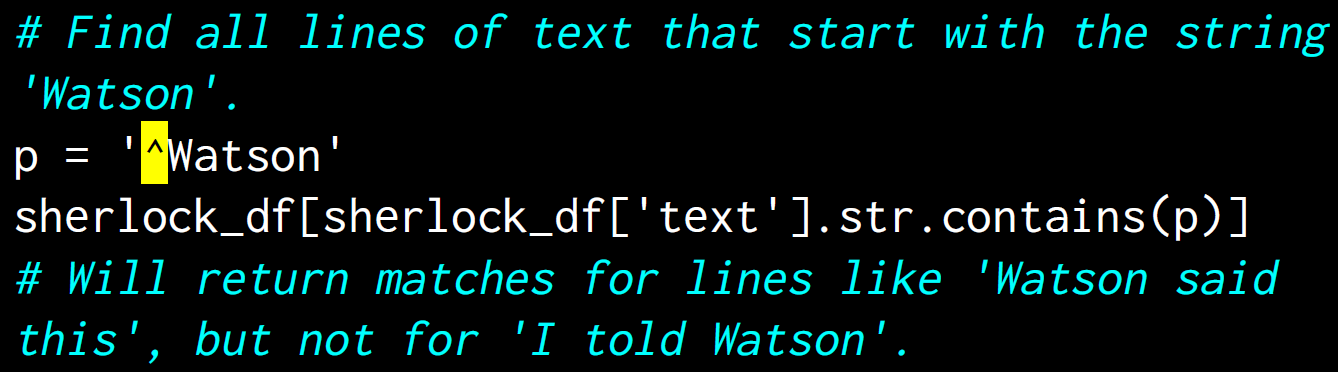
The question mark (?) lets us match either none or one of the preceding characters.



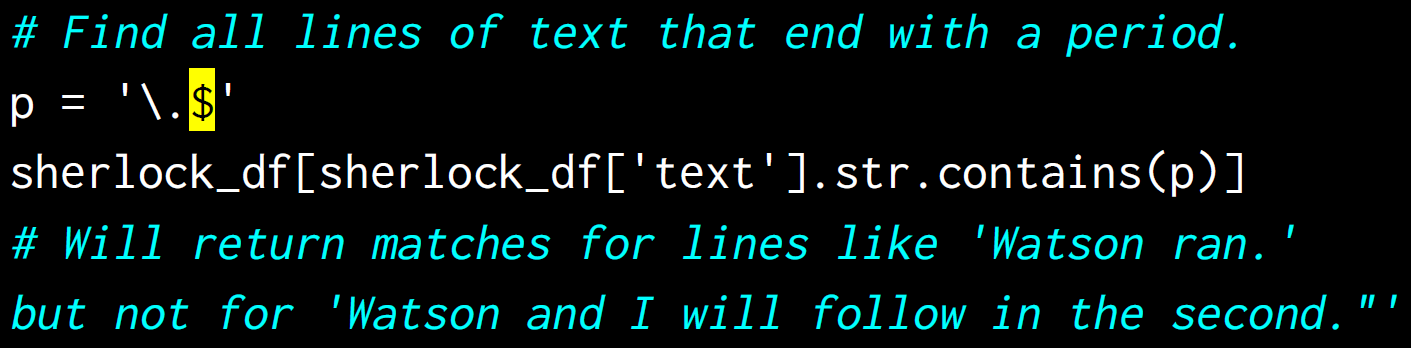
The asterisk (\*) lets us match either none, one, or more than one of the preceding characters.



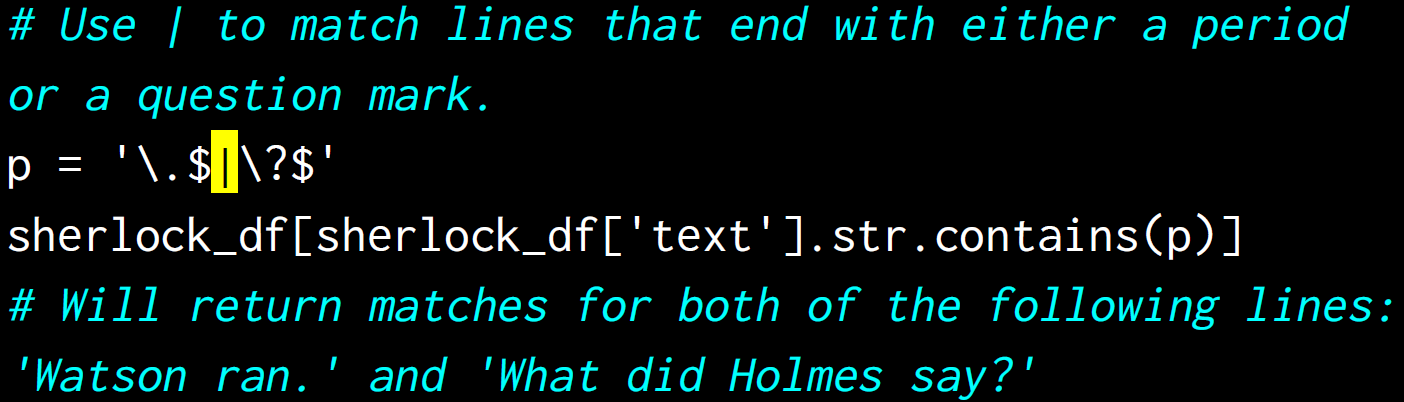
The caret (^) lets us match lines that start with the subsequent expression.

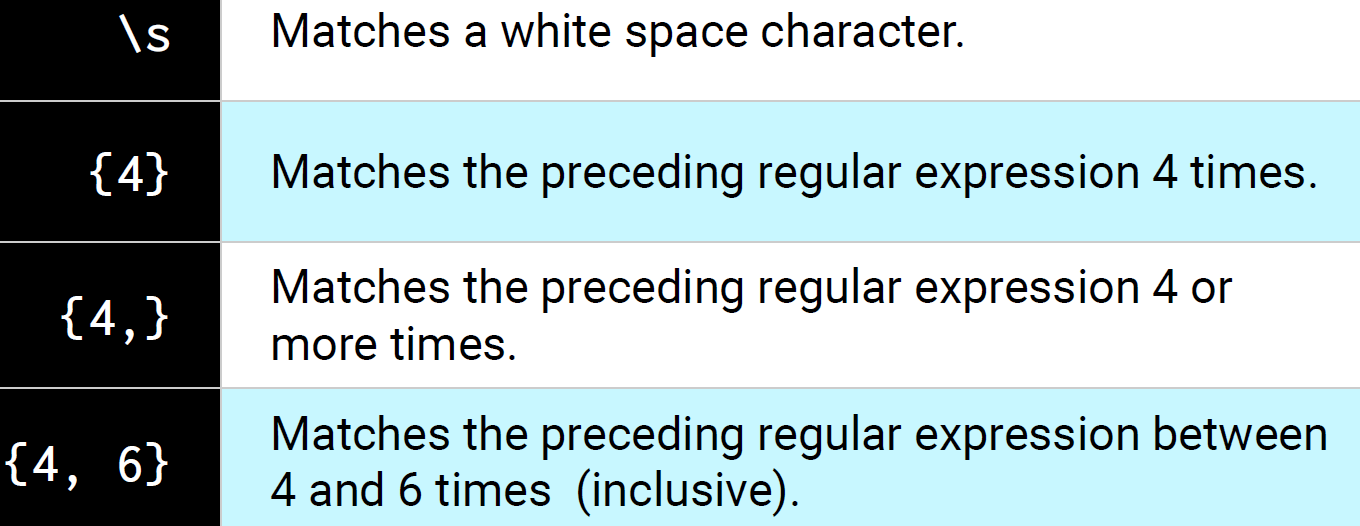


The dollar sign ($) lets us match lines that end with the preceding expression.



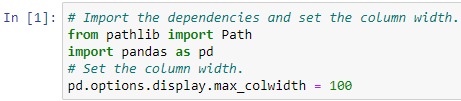
The pipe (|) lets us put a conditional expression in our regex expression to match the term that either precedes or follows it.

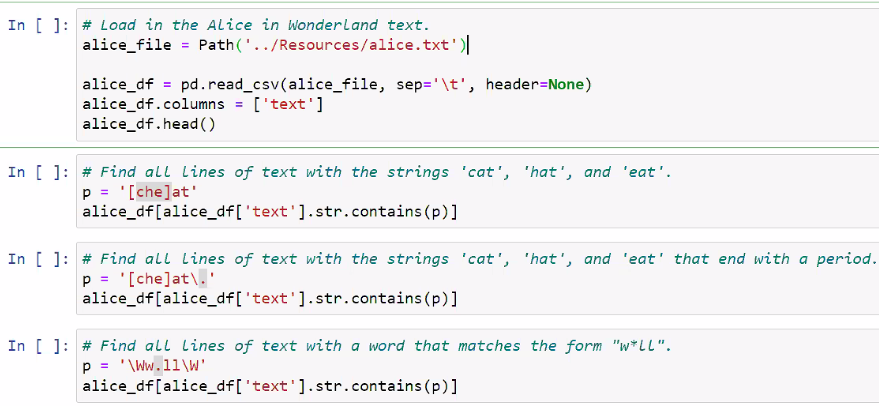




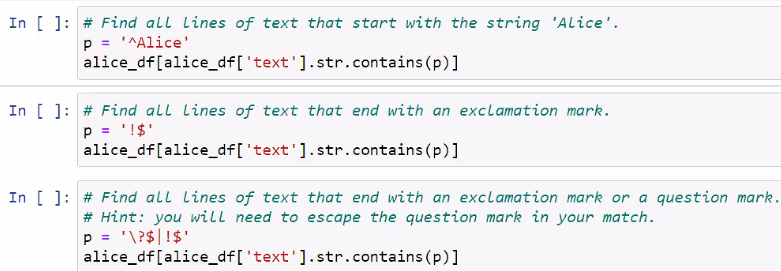
A capture group allows us to further extract information from our regex matches.

**Wildcards Sets Escaping Solution:**

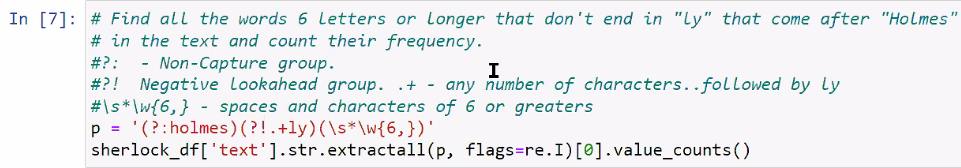




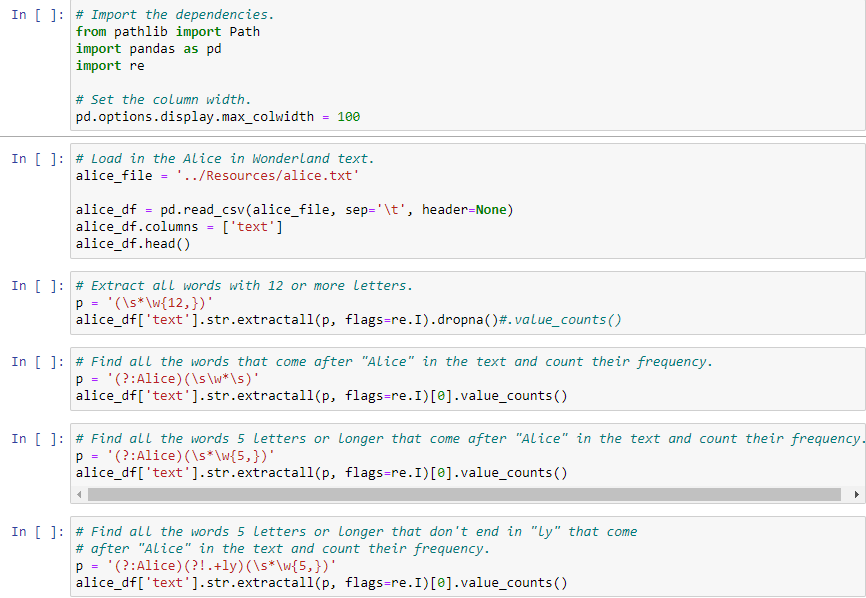
**Regex Special Characters Solution:**



**Instructor Regex Groups Solution:**



**Regex Groups Solution:**



**Transforming and Cleaning IoT Data Solution:**

IoT- Internet of things (measures weather data)

**Useful for project? Contact.csv file portion- IN [8 and 9]**

