Data Science for Hackers-Regex Reference Sheet

PYTHON REGEX COMMANDS	
<*Series*>.str.contains(<*regex_pattern*>)	Returns a boolean series if the series element contains the regex_pattern
<*Series*>[<Series>.str.contains(<*regex_pattern*>)]	Returns a series only containing those elements which match the regex_pattern.
<*Series*>**.str.extract**(<*regex_pattern*>)	Returns a new series containing the regex_pattern specified within the parentheses. Works for multiple capture groups.
<*Series*>**.str.replace**(<*regex_pattern*>, 'new string')	Finds the regex_pattern and replaces with the new string.
<*Series*>.**str.match**(<*regex_pattern*>)	returns a boolean if the series element exactly matches the regex_pattern.

Character	Definition	Example
^	Pattern starts the string.	^cat matches any string that begins with cat
$	The pattern has to appear at the end of a string.	cat$ matches any string that ends with cat
.	Matches one of any character.	cat. matches catT and cat2 but not catty
[]	Bracket expression. Matches one of any characters enclosed.	gr[ae]y matches gray or grey
[^]	Negates a bracket expression. Matches one of any characters EXCEPT those enclosed.	1[^02] matches 13 but not 10 or 12
[-]	Range. Matches any characters within the range.	[1-9] matches any single digit EXCEPT 0
?	Preceding item must match one or zero times.	colou?r matches color or colour but not colouur
+	Preceding item must match one or more times.	be+ matches be or bee but not b
*****	Preceding item must match zero or more times.	be* matches b or be or beeeeeeeee
()	Parentheses. Creates a substring or item that metacharacters can be applied to	a(bee)?t matches at or abeet but not abet
{n}	Bound. Specifies exact number of times for the preceding item to match.	[0-9]{3} matches any three digits

{n,}	Bound. Specifies minimum number of times for the preceding item to match.	[0-9]{3,} matches any three or more digits
{n,m}	Bound. Specifies minimum and maximum number of times for the preceding item to match.	[0-9]{3,5} matches any three, four, or five digits
|	Alternation. One of the alternatives has to match.	July (first|1st|1) will match July 1st but not July 2
Special Characters		
\d	A single digit character, or [0-9]	/a\db/i matches a2b but not acb
\D	A single non-digit character, or [^0-9]	/a\Db/i matches aCb but not a2b
\s	A single whitespace character	/a\sb/ matches a b but not ab
\S	A single non-whitespace character	/a\Sb/ matches a2b but not a b
\t	The tab character. (ASCII 9)	/\t/ matches a tab.
\w	A single word character - alphanumeric and underscore, or [0-9a-zA-Z_]	/\w/ matches 1 or _ but not ?
\W	A single non-word character, or [^a-zA-Z0-9_]	/a\Wb/i matches a!b but not a2b

USEFUL REFERENCES:

www.regexone.com - A quick interactive tutorial on basic regex.
www.regexpal.com - An interactive tester for for Java based regex.