NYU TANDON SCHOOL OF ENGINEERING

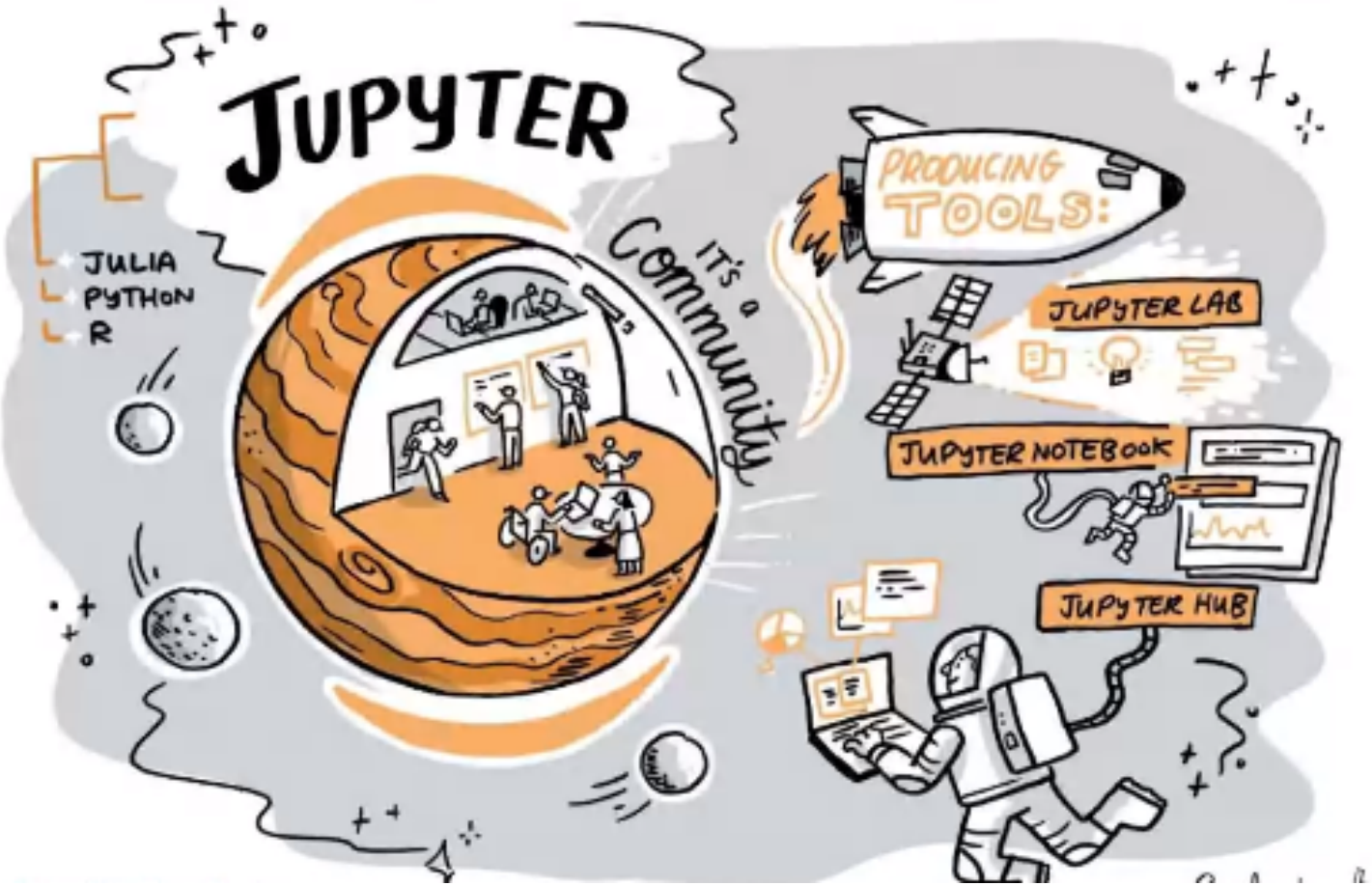
Lab 6
Text

Programming for Business Analytics
MG-GY 8401

NYU TANDON SCHOOL OF ENGINEERING

Agenda

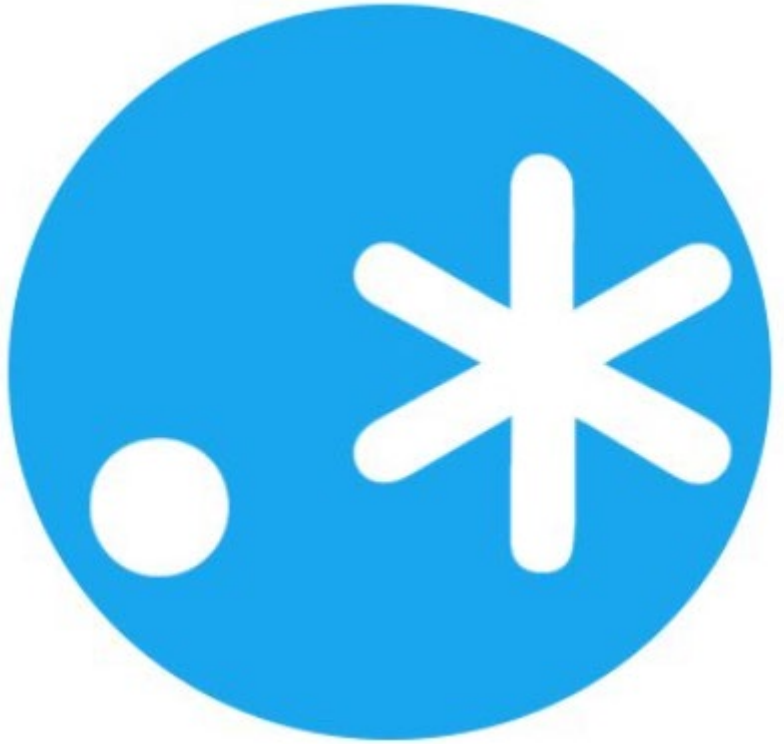
- Text
- Regular Expressions



- Describe the learning objectives.
- Summarize the relevant take-aways.
- Ask about unclear information.

The ability to extract patterns of characters in text leads to many applications

- Access information in digital libraries
- Filter text such as spam emails
- Validate data-entry fields such as date or URL



operation	order	example	matches	does not match
concatenation	3	AABAAB	AABAAB	every other string
or	4	AA|BAAB	AA BAAB	every other string
closure (zero or more)	2	AB*A	AA ABBBBBBA	AB ABABA
parenthesis	1	A(A|B)AAB	AAAAB ABAAB	every other string
		(AB)*A	A ABABABABA	AA ABBA

operation	example	matches	does not match
any character (except newline)	`.U.U.U.`	`CUMULUS` `JUGULUM`	`SUCCUBUS` `TUMULTUOUS`
character class	`[A-Za-z][a-z]*`	`word` `Capitalized`	`camelCase` `4illegal`
at least one	`jo+hn`	`john` `jooooooohn`	`jhn` `jjohn`
zero or one	`joh?n`	`jon` `john`	`any other string`
repeated exactly {a} times	`j[aeiou]{3}hn`	`jaoehn` `jooohn`	`jhn` `jaeiouhn`
repeated from a to b times: {a,b}	`j[ou]{1,2}hn`	`john` `juohn`	`jhn` `jooohn`

Example

regex	matches	does not match
`.*SPB.*`	RASPBERRY CRISPBREAD	SUBSPACE SUBSPECIES
`[0-9]{3}-[0-9]{2}-[0-9]{4}`	231-41-5121 573-57-1821	231415121 57-3571821
`[a-z]+@([a-z]+\.)+(edu|com)`	horse@pizza.com horse@pizza.food.com	frank_99@yahoo.com hug@cs

operation	example	matches	does not match
built-in character classes	`\w+` `\d+`	`fawef` `231231`	`this person` `423 people`
character class negation	`[^a-z]+`	`PEPPERS3982` `17211!↑å`	`porch` `CLAmS`
escape character	`cow\.com`	`cow.com`	`cowscom`

operation	example	matches	does not match
beginning of line	`^ark`	`ark two` `ark o ark`	`dark`
end of line	`ark$`	`dark` `ark o ark`	`ark two`
lazy version of zero or more *?	`5.*?5`	`5005` `55`	`5005005`

5*5 would match this!

Write a regular expression to match the pattern

- XYYYY!

at the end of strings. Here

- X is a digit
- Y is a letter
- ! is the exclamation point