Trainer: Mahnoor Salman

Instructions:

- Submit Python Notebook file only.

Q#1: Write a regular expression function for the following. By "word", we mean an alphabetic string separated from other words by whitespace, any relevant punctuation, line breaks, and so forth.

1. the set of all strings with two consecutive repeated words (e.g., "Humbert Humbert" and "the the" but not "the bug" or "the big bug");
2. all strings that start at the beginning of the line with an integer and that end at the end of the line with a word;
3. all strings that have both the word grotto and the word raven in them (but not, e.g., words like grottos that merely contain the word grotto);
4. Demonstrate a pattern to check if the email address is valid.
5. Write a function to verify the validity of a phone number, and check if it belongs to the Pakistani Mobile Network.
6. Write a function to remove symbols and non-alphanumeric characters from the string paragraph.
7. Write a function to remove URLs and HTML tags from the text string.
8. Write a function find_acronyms that takes a string of text and returns a list of acronyms found in the text. Assume acronyms are in uppercase letters.
9. Write a function mask_sensitive_info that takes a string of text and masks sensitive information such as phone numbers (in the format XXX-XXX-XXXX).
10. Write a function extract_dates that takes a string of text and returns a list of dates in various formats (e.g., DD-MM-YYYY, MM/DD/YYYY).
11. Write a function extract_currency_amounts that takes a string of text and returns a list of currency amounts (e.g., $100.00, €50).
12. Write a function find_capitalized_words that takes a string of text and returns a list of capitalized words.
13. Write a function find_repeated_words that takes a string of text and returns a list of words that are repeated consecutively.

Reference Material:

- Google Colaboratory as Python IDE
- Regular Expression Documentation
- W3School for Sample RE Implementation

- Build and Test RE Online