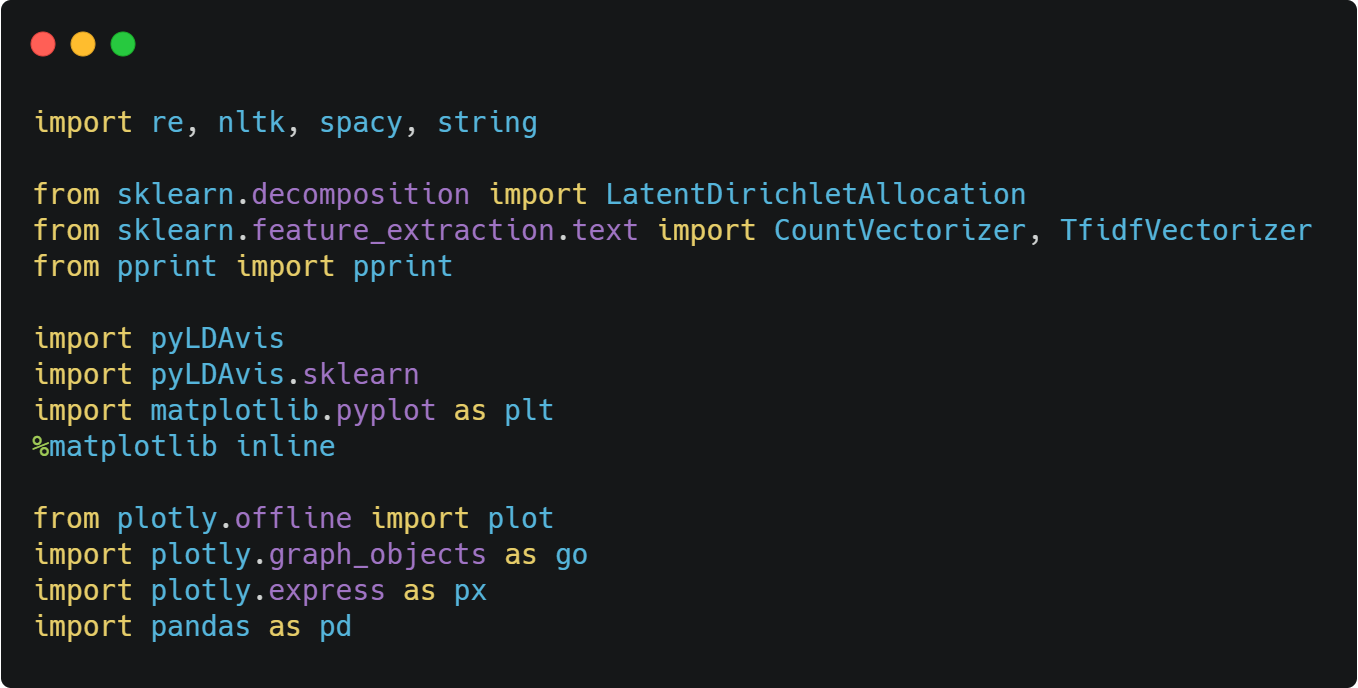
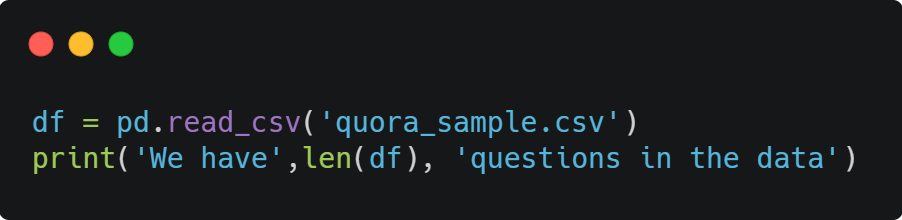
**Project Topic- Quora Topic Modeling using natural language processing and scikit-learn**

Using Carbon to paste images of the source code - <https://carbon.now.sh/>

// *Installing the required packages*



*// Importing the dataset and printing the number of questions in the data*

**

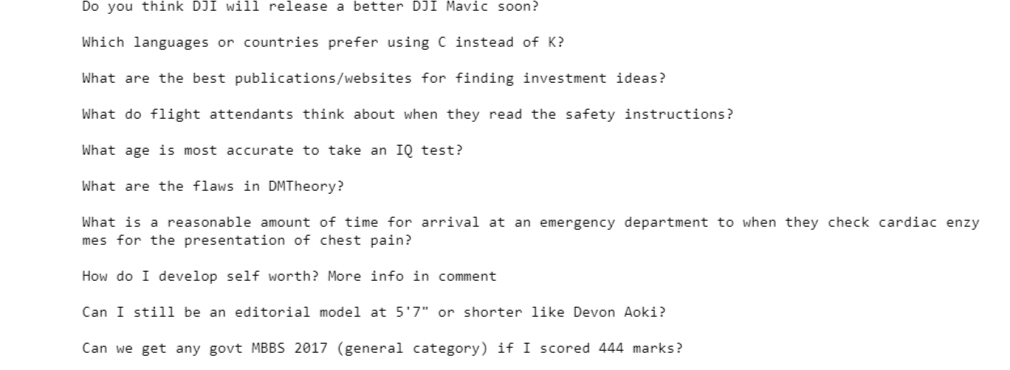
*Output-*

We have 9838 questions in the data

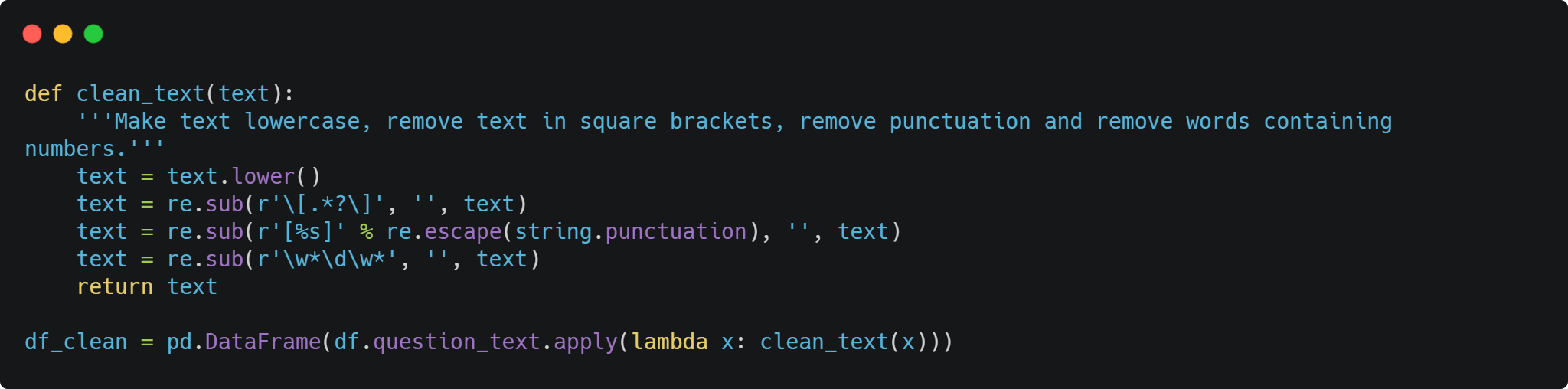
*//Printing some random questions from the data*



*Output*



*//Making text lowercase, removing text in square brackets, punctuation and words containing numbers.*

**

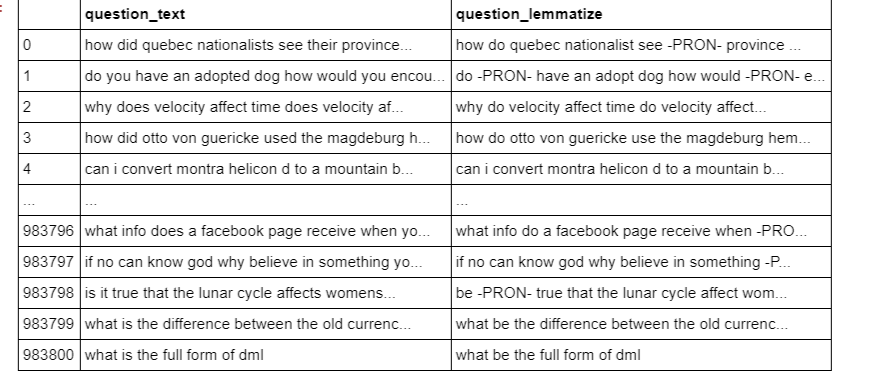
*// Loading the model for English using Spacy library*

**

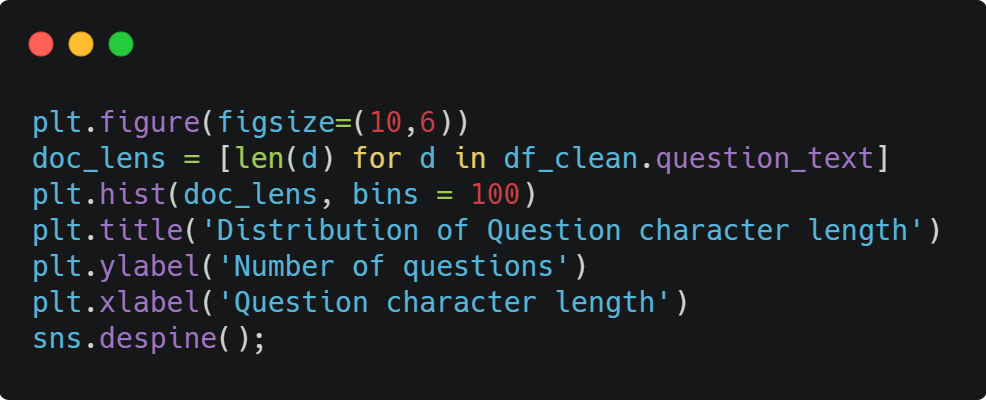
*// implementing lemmatization in python and applying it to df\_clean.*

**

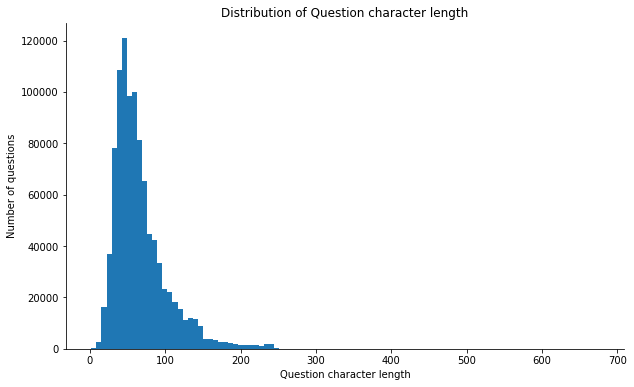
*Output –*

**

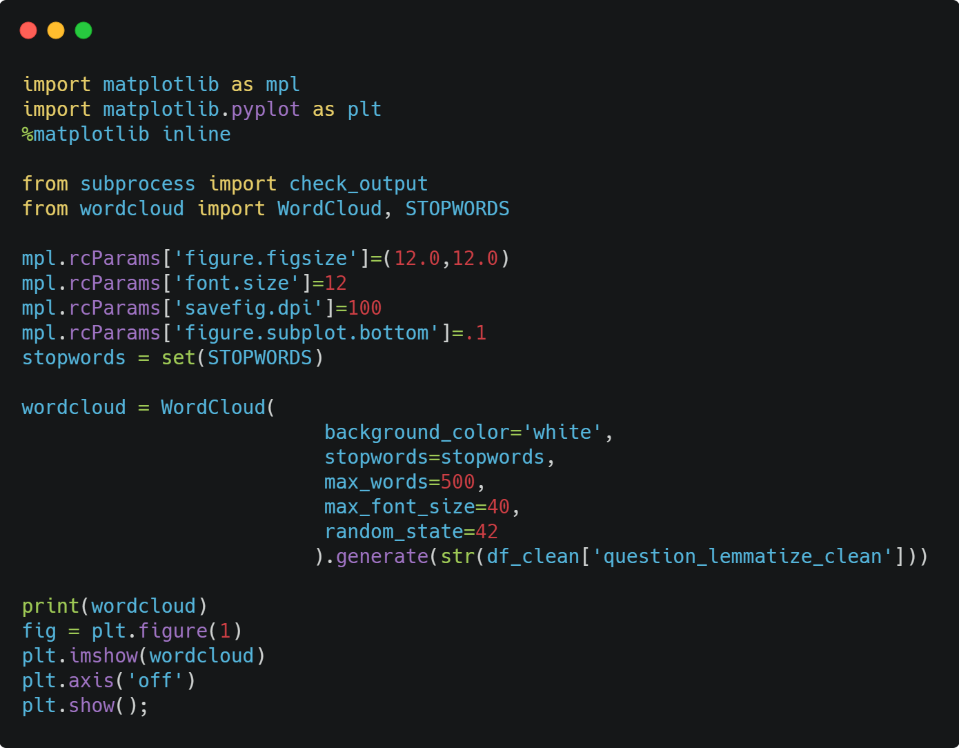
*// Plotting number of questions versus question character length*

**

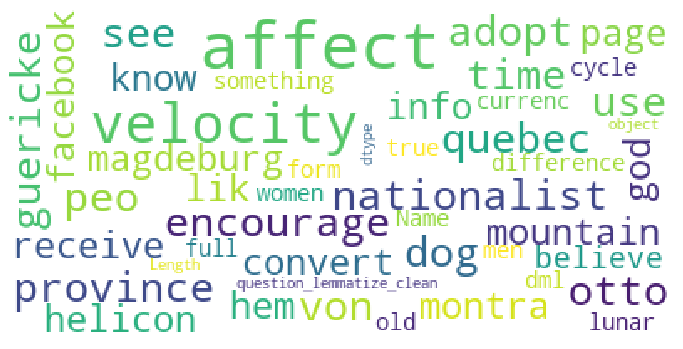
*Output-*

**

*// Using word cloud to print the most prominent words*

**

*Output-*

**

*// replacing -PRON- to a null string to clean the dataset*

**