

DISRUPTIVE DATA SUMMER SCHOOL - TEST

Candidato: **LEGITTIMO LUIGI**

1. Descrizione del processo

Link repository: <https://github.com/LuigL/Legittimo-Luigi-Disruptive-Data-Summer-School-Test.git>

Ho scelto il **Project 4**

Per svolgere il progetto ho usato il linguaggio di programmazione R, in particolare ho sviluppato l'analisi sull'ambiente **RStudio**.

Una volta scaricato il file contenente la lista di parole, l'ho caricato nel mio ambiente di lavoro con il nome "Words" usando la funzione `read_excel()`, spiegando dove si trova il file da importare, il range di caselle voluto, il nome della colonna contenente le parole che ho chiamato "Words", il tipo di contenuto presente nelle celle della colonna. Il risultato è un tibble dataframe con 25 osservazioni e una variabile (Words) classificata come "character"; si possono vedere le prime sei osservazioni con la funzione `head()` e la struttura del data frame con `str()`

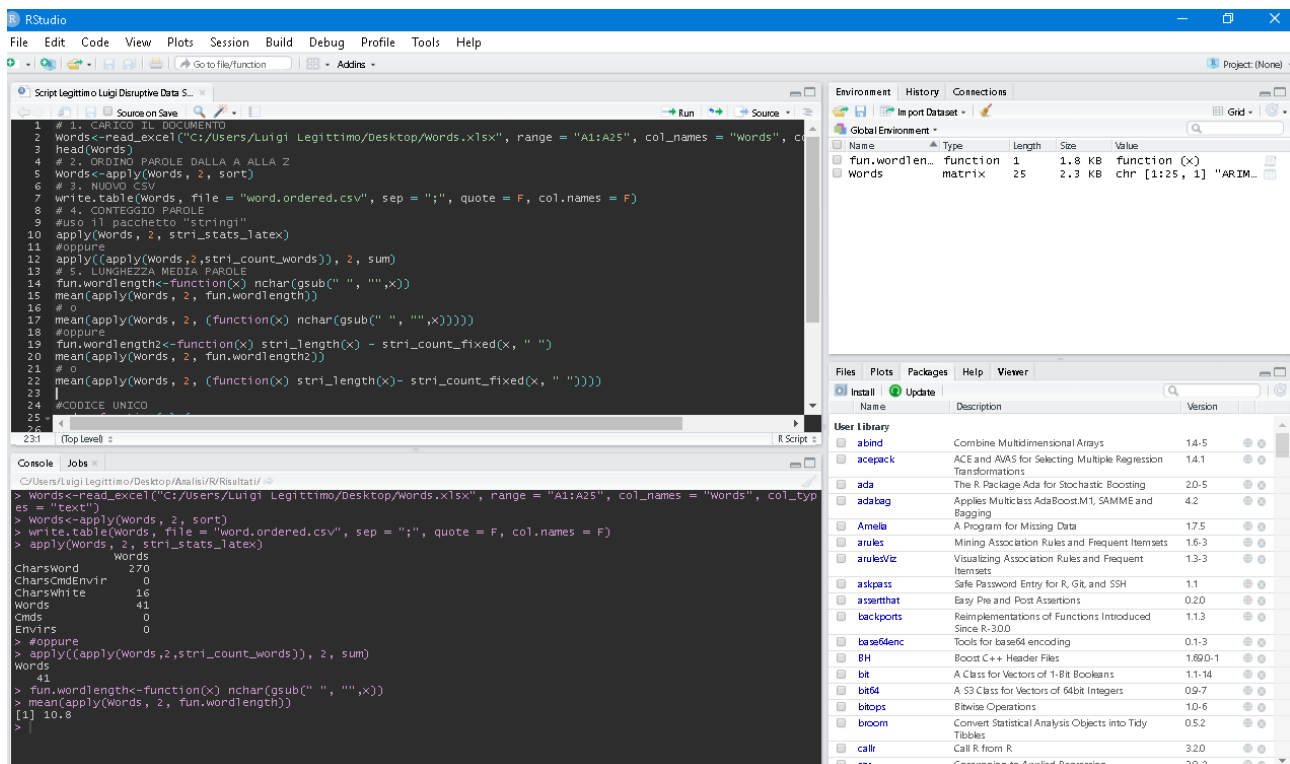
Successivamente, (punto 2) ho ordinato la lista di parole in ordine alfabetico tramite la funzione `sort()` che ordina i dati in modo crescente, essendo di default impostato `sort(x, decreasing = F, ...)`, in seguito ho usato la funzione `apply()` che permette di ripetere delle operazioni su un sottoinsieme di dati, in questo caso all'intera colonna (indicata con il numero "2") del data frame; applicando ad ogni cella la funzione in questione. Al punto 3 ho sovrascritto il data frame ordinato con lo stesso nome (Words). Con la funzione `write.table()` ho salvato la lista ordinata di parole in un file con estensione CSV, senza il nome della colonna e senza le virgolette.

Al punto 4 per conteggiare le parole nella lista, oltre alla già citata funzione `apply()`; mi sono avvalso del pacchetto "*string*" che serve per la manipolazione delle stringhe, tramite la funzione `stri_stats_latex()`, in alternativa si può usare un'altra funzione che conta le singole parole in una cella e poi sommarle.

Per il punto 5, ovvero, la valutazione della lunghezza media delle parole, dato che la funzione `nchar()` o `stri_length()` conteggiavano anche gli spazi vuoti tra una parola e l'altra; ho dovuto creare una funzione apposita chiamata "*fun.wordlength*" che escludesse dal conteggio tali spazi: per `nchar()` ho utilizzato `gsub(" ", "", x)` che sostituisce gli spazi vuoti con una stringa vuota, se utilizzo `stri_length(x)` devo sottrarre a quest'ultimo `stri_count_fixed(x, " ")`. Dopodiché per calcolarmi la lunghezza media delle parole ho utilizzato la funzione `mean()` con oggetto la funzione `apply(Words, 2, fun.wordlength)`.

Oppure, si può creare una funzione che racchiuda tutte le funzioni spiegate sopra e che restituisca in automatico tre output (la lista di parole ordinate, il numero di parole e la loro lunghezza media) oltre a salvare in formato CSV la lista di parole ordinata dalla A alla Z.

Ho creato questa funzione e l'ho chiamata `code()`.



2. Presentazione

Sono Luigi Legittimo, provengo da Casarano (LE), sono nato nel 1994. Ho studiato Analisi Economica (LM-56) alla Sapienza di Roma e sono in procinto di laurearmi. Grazie anche all'offerta didattica del mio corso di laurea magistrale, ho iniziato ad interessarmi prima all'utilizzo di software econometrici partendo dal "classico" Excel, STATA, e poi EView per l'analisi statistica delle serie economiche e successivamente ho iniziato ad avvicinarmi con molto interesse al vastissimo mondo della Data Science, studiando linguaggi di programmazione come R o Python. Ho seguito in merito, sia vari corsi su piattaforme e-learning e sia ho studiato su manuali da autodidatta. Nonostante questo, sono ancora un neofita in questo settore, perché ho ancora tanto da migliorare e soprattutto da imparare. Per tale motivo vorrei partecipare alla vostra Summer School; perché oltre ad avere finalmente la grande possibilità di seguire lezioni con docenti qualificati che toccano vari campi della data science e di svolgere esercitazioni pratiche su quello che si è studiato, avrei l'occasione formativa che mi consentirebbe di addentrarmi sempre di più in questo settore nel quale vorrei lavorare.