Wanna make better decisions while SHOPPING ONLINE?

Lemme help you

in automating the approach of going through the product reviews using the magic of Data Analytics!

Ready to make the smart choice?
Let’s go!

Introducing

my new project - an R code which calculates the sentiment of all the reviews of a product!

Steps given from next slide onwards.

STEP 1 - Data collection

Product

Reviews

Ratings

using rvest lib.

Scrape

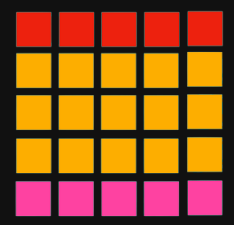
from the website

Store

in a data frame

STEP 2 - Data cleaning

Stored Data



(inside a data frame)

Clean

(using stringr lib. & gsub func.)

New line characters

- \r\n
- \n
- \r

Space characters

- multiple
- leading
- trailing

Non ASCII characters

- emojis
- symbols
- special char

STEP 3 - Data Analysis

Single Review

(has multiple) **Sentences** → $S_1 + S_2 + S_3 + \dots + S_n$ (split using tidyverse lib.)

Sentiment Score (of every sentence) → SS_1 SS_2 SS_3 … SS_n (find using sentimentr lib.)

Average Sentiment score

Store

in the data frame

STEP 4 - Data Visualisation

Final data

(in cleaned format)

Summary plots

- **Bar graph** (shows the frequency of ratings)
- **Histogram** (shows the distribution of sentiments)

(using ggplot2 lib.)

Relationship plots

- **Scatterplot** (shows the relationship b/w ratings & sentiments)
- **Boxplot** (shows the distribution of ratings over the sentiments)

(using ggplot2 lib.)

Text visualisation

- **Word clouds** (shows the collection of words which occur frequently in positive & negative)

(using tm & wordcloud lib.)

Conclusion

Using these insights we can figure out whether the sentiment of a product's review is positive, negative or neutral!

PS - Want to know the sentiment of a specific feature in a product?

Just enter any keyword, & the code will give a sentiment score for that feature.

How cool is that?

Do check out the project on GitHub :)

Thanks for reading, hope you liked it :)

If you found it insightful, then **Follow me for more Data analytics content!**