

task5__030

August 30, 2024

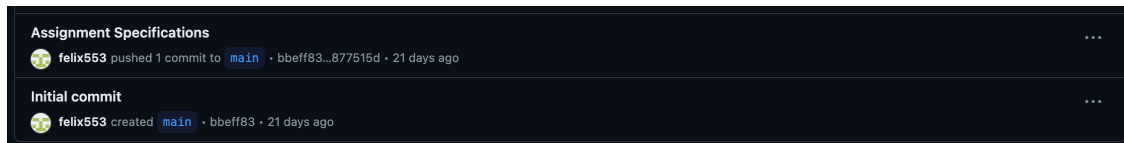
1 TASK 5 - Development History

1.0.1 Date: 09/08/2023

Today, we went through the assignment specification, and started exploring task 1 data.

contribution: Group member 1: - understood the data structure by exploring data via browsing the sample output and input

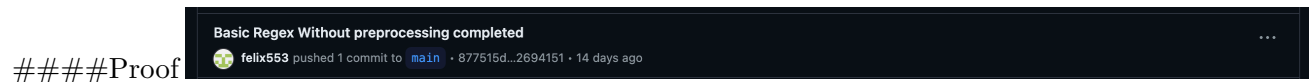
Group member 2: - understood the data structure by exploring data via browsing the sample output and input



proof

1.0.2 Date 16/08/2023







Today Michael xie implemented basic regex to extract the reviews on review text, ratings, time etc.
Group member 1: - Set the structure to the whole extraction process



1.0.3 Date 18/08/2023 - 21/08/2023

These days, Weiliang Huang further complete the extraction on the xml file on if_pic, time etc, and functionized the extracting process. #####Contribution Group member 1: - Set extraction on if_pic, picture dimension and if_response etc.

####Proof


functionise capturing rach pattern and combining them into json	...
 william99987 pushed 1 commit to main · 2113133...3f9fc20 · 9 days ago	
combine data into json format	...
 william99987 pushed 1 commit to main · 378022b...2113133 · 10 days ago	
Fixed if_pic	...
 william99987 pushed 1 commit to main · 2715426...378022b · 11 days ago	
Fixed time and finalized earliest time and latest time	...
 william99987 pushed 1 commit to main · b9ee966...2715426 · 11 days ago	
picture width and height if_pic doen	...
 william99987 pushed 1 commit to main · 6bae8de...b9ee966 · 11 days ago	
time format finished	...
 william99987 pushed 1 commit to main · 2694151...6bae8de · 12 days ago	

1.0.4 Date: 22/08/2023 - 24/08/2023

These two days, we finalized the task1 output

contribution: Group member 1: - write the code to produce the dataframe output of the excel file groupby the dataframe and merge it with the xml output. Group member 2: - Fixing the output format of the combining json file, Format the task_1 notebook to make it look good, align the output with the sample ouput.

Conflict Resolution

 felix553 pushed 1 commit to `main` • 90a0732...05e1cb5 • 8 days ago

excel to json done and merge xml with txt output

 william99987 pushed 1 commit to `main` • 7e73acc...90a0732 • 8 days ago

Implemented language validation and remove duplicate logic

 felix553 pushed 1 commit to `main` • fd31edf...7e73acc • 8 days ago


Fixed review text regex

 felix553 pushed 1 commit to `main` • 492be1f...fd31edf • 8 days ago

change to pandas dataframe of json

 william99987 pushed 1 commit to `main` • 6e6b928...492be1f • 8 days ago


excel to json

 william99987 pushed 1 commit to `main` • b30086d...6e6b928 • 8 days ago


json output foremat done

 william99987 pushed 1 commit to `main` • b0a6a59...b30086d • 8 days ago


Fixed if_pic and if_response bug

 felix553 pushed 1 commit to `main` • 83fed41...b0a6a59 • 8 days ago

CSV output completed


 felix553 pushed 1 commit to `main` • ea8082c...83fed41 • 8 days ago

Remove Emoji completed


 felix553 pushed 1 commit to `main` • 3f9fc20...ea8082c • 8 days ago

proof


Implemented Translation Extraction + Added comments

 felix553 pushed 1 commit to `main` • f056e13...5a8185e • 6 days ago


Implemented new JSON output logic

 felix553 pushed 1 commit to `main` • 3fd2224...f056e13 • 7 days ago

Version 2 of json output

 felix553 pushed 1 commit to `main` • 0a52c91...3fd2224 • 7 days ago

Changed pic dim to list


 felix553 pushed 1 commit to `main` • 582333a...0a52c91 • 7 days ago


##Task 2


1.0.5 Date:26/08/2023 - 28/08/2023

These days, Weiliang Huang finished task2.


contribution: Group member 1: - Weiliang Huang finished generating the countvect and vocab txt.

Countvec file output
 **william99987** pushed 1 commit to **main** · ccb67fc...bffc2c2 · 3 days ago

vocab change sequence
 **william99987** pushed 1 commit to **main** · ae4fb45...ccb67fc · 3 days ago

vocab roughly done
 **william99987** pushed 1 commit to **main** · 64f1d09...ae4fb45 · 3 days ago

add a file
 **william99987** pushed 1 commit to **main** · cdce32e...64f1d09 · 4 days ago







get text dataframe out
 **william99987** pushed 1 commit to **main** · 9cce0c0...cdce32e · 4 days ago

proof

Added some explanation
 **william99987** pushed 1 commit to **main** · 2e07ab6...45a504e · 2 days ago

fix review text none
 **william99987** pushed 1 commit to **main** · 15cef4b...2e07ab6 · 2 days ago

task2
 **william99987** pushed 1 commit to **main** · f24d466...15cef4b · 2 days ago

countvec.txt
 william99987 pushed 1 commit to main • 658649c...5f5e61c • 2 days ago
countvec.txt
 william99987 pushed 1 commit to main • 13b2037...658649c • 2 days ago
task 2 ouput file fixed again
 william99987 pushed 1 commit to main • 2de2da4...13b2037 • 2 days ago
task 2 ouput file fixed
 william99987 pushed 1 commit to main • 6273c90...2de2da4 • 2 days ago
task 2 ouput file
 william99987 pushed 1 commit to main • 99c5aeb...6273c90 • 2 days ago
output vocab using Mget_feature_names_out
 william99987 pushed 1 commit to main • 23937c9...39abb03 • 2 days ago

##Task 3

1.0.6 Date:26/08/2023 - 28/08/2023

These days, Michael Xie finished task3

contribution: Group member 1: - Michael xie used out put from task 1 and 2 to generate data exploration insights and charts.

proof

Fixed rating type bug



felix553 pushed 1 commit to `main` • bffc2c2...f24d466 • 2 days ago

Implemented MISC analysis



felix553 pushed 1 commit to `main` • f38cb17...99c5aeb • 2 days ago

3 Insights completed



felix553 pushed 1 commit to `main` • b04a9ad...f38cb17 • 2 days ago

Completed Initial Findings for review data



felix553 pushed 1 commit to `main` • 45a504e...b04a9ad • 2 days ago

Word insight generated













felix553 pushed 1 commit to `main` • 5f5e61c...23937c9 • 2 days ago

1.0.7 Date: 29/08/2023 - 30/08/2023

These two days, Michael Xie prepared the presentation slides and contents for the presentation. Weiliang Huang write dev history and write methodology to task1.

contribution: Group member 1: - prepared the presentation slides and contents for the presentation

Group member 2: - write dev history and write methodology for task1

task1 explanation	...
 william99987 pushed 1 commit to main · 5a9d5f0...4c9fa5b · 1 hour ago	
Completed Presentation Powerpoint	...
 felix553 pushed 1 commit to main · 404db04...5a9d5f0 · 2 hours ago	
Added some more comments Task3	...
 felix553 pushed 1 commit to main · 893d044...404db04 · 6 hours ago	
Completed final insight	...
 felix553 pushed 1 commit to main · 5c4de3d...893d044 · 13 hours ago	
Added comments for count vec	...
 felix553 pushed 1 commit to main · cef338f...5c4de3d · 14 hours ago	
Added comments for price and category	...
 felix553 pushed 1 commit to main · ae088dd...cef338f · 17 hours ago	
Added Comments for MISC and Location	...
 felix553 pushed 1 commit to main · c06760f...ae088dd · 18 hours ago	
Added comments for Dataset review stage	...
 felix553 pushed 1 commit to main · 2639b9c...c06760f · 20 hours ago	
mehodology	...
 william99987 pushed 1 commit to main · 6313f7a...2639b9c · yesterday	
Comment codes	...
 william99987 pushed 1 commit to main · 39abb03...6313f7a · yesterday	

proof