ESSENTIALS OF DATA SCIENCE

Theory Activity No. 1

Name – Rushikesh Ganesh Shinde

PRN – 202401040040

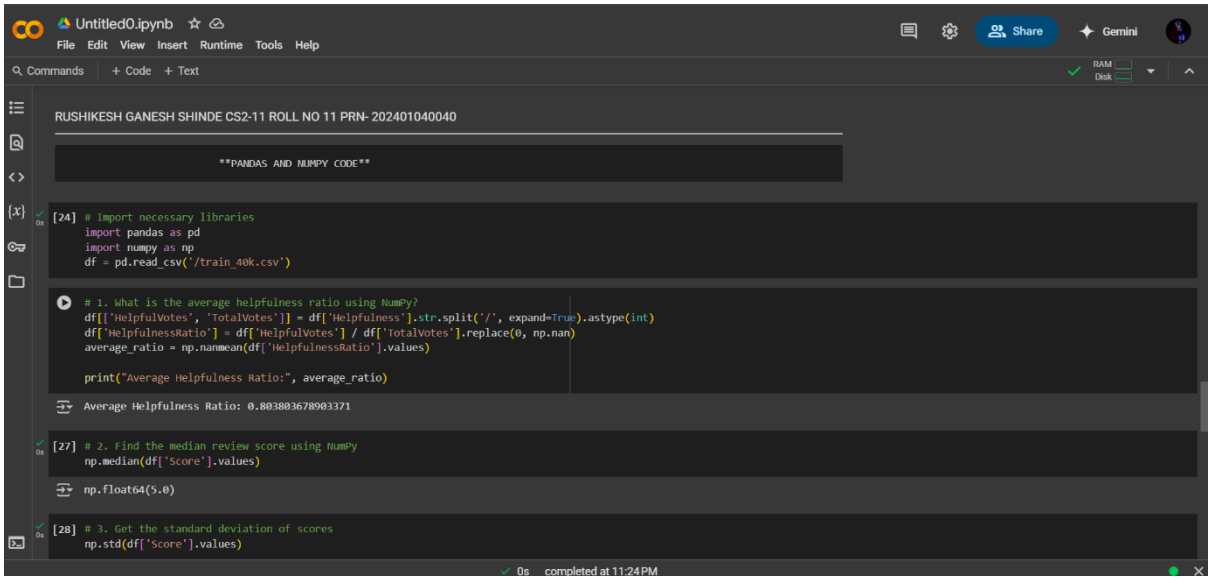
Roll N0 – CS2-11

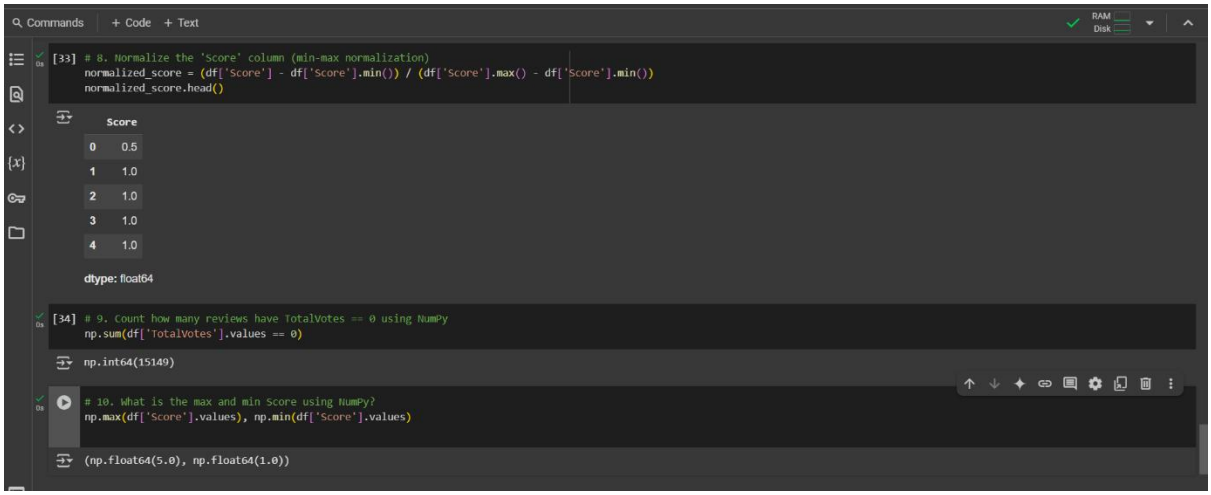
Div – CS2

- 20 problem statements for **Kaggle Text Classification Dataset** using Numpy and Pandas.

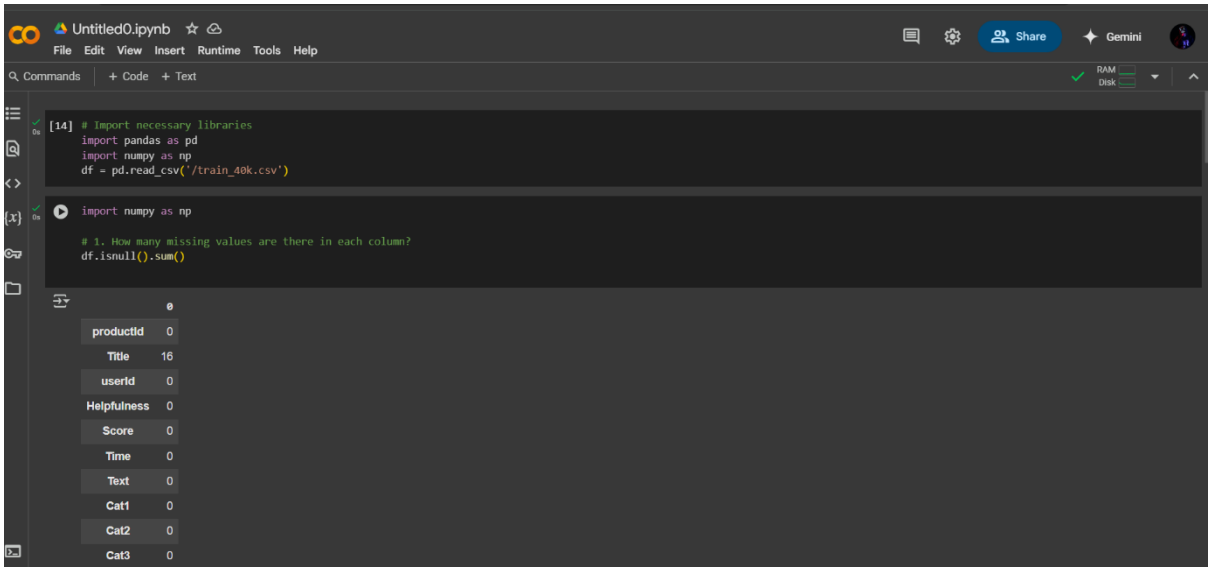
---

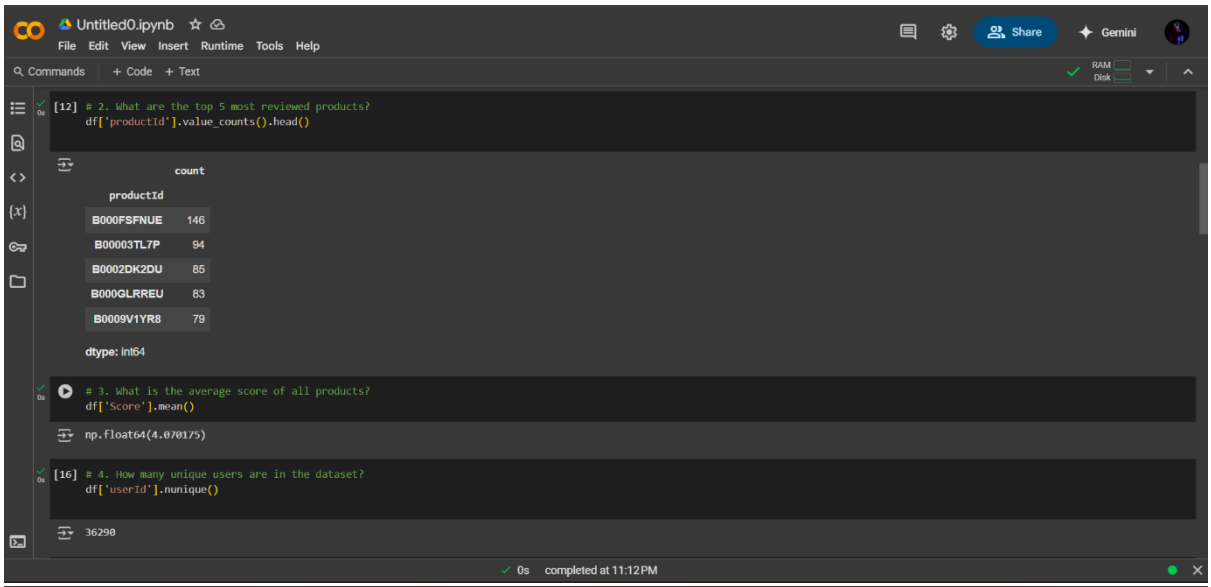
1) **Numpy Question and Answers :-**

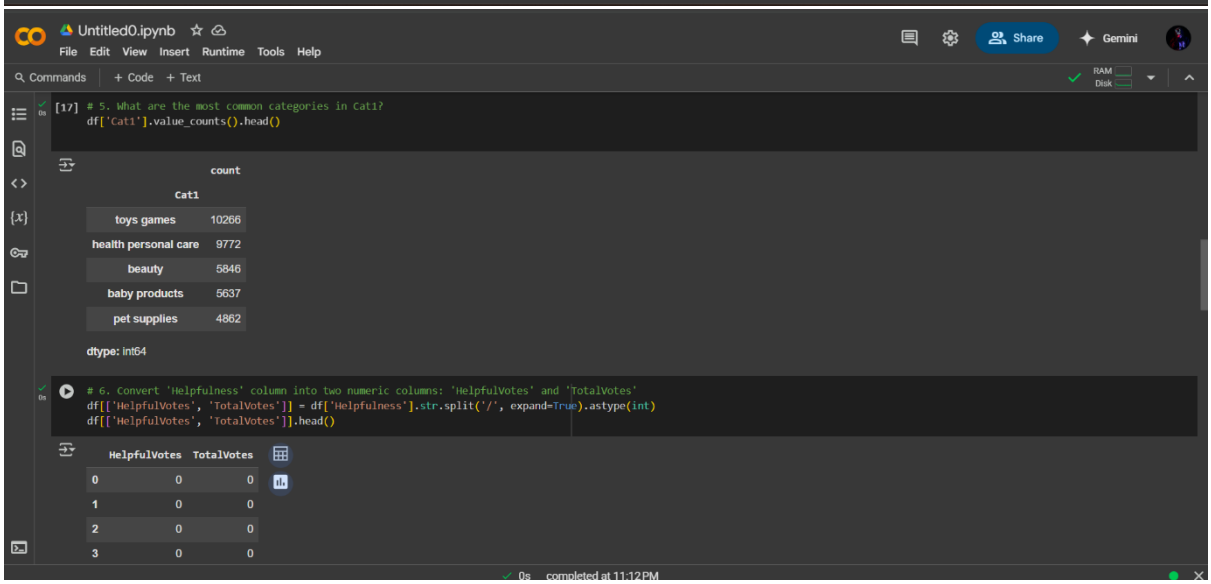


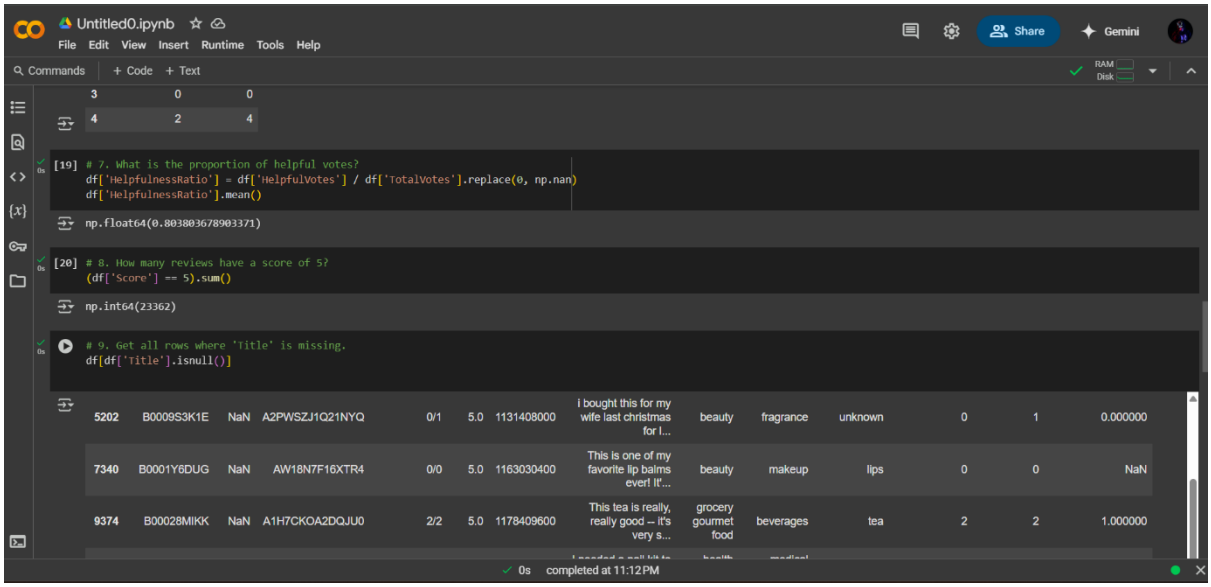


2) Pandas Question and Answers :-








Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
3 0 0
4 2 4
[19] # 7. What is the proportion of helpful votes?
df['HelpfulnessRatio'] = df['HelpfulVotes'] / df['TotalVotes'].replace(0, np.nan)
df['HelpfulnessRatio'].mean()
np.float64(0.803803678903371)
[20] # 8. How many reviews have a score of 5?
(df['Score'] == 5).sum()
np.int64(23362)
# 9. Get all rows where 'Title' is missing.
df[df['Title'].isnull()]
5202 B0009S3K1E NaN A2PWSZJ1Q21NYQ 0/1 5.0 1131408000 i bought this for my wife last christmas for l... beauty fragrance unknown 0 1 0.000000
7340 B0001Y6DUG NaN AW18N7F16XTR4 0/0 5.0 1163030400 This is one of my favorite lip balms ever! It'... beauty makeup lips 0 0 NaN
9374 B00028MIKK NaN A1H7CKOA2DQJU0 2/2 5.0 1178409600 This tea is really, really good -- it's very s... grocery gourmet food beverages tea 2 2 1.000000
0s completed at 11:12PM


Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
12117 B000KGRNMM NaN A2EVLYSFT7IDC6 99/100 5.0 1199232000 take care of my Mother'... personal care supplies equipment daily living aids 99 100 0.990000
13089 B000GUFFQS NaN A1P0IHM5P5WVOZ 1/1 5.0 1204934400 My 93 year old father in law recently fell (lu... health personal care medical supplies equipment beds accessories 1 1 1.000000
14881 B000KGRNMM NaN A3V0HBE74ZUFKB 52/52 5.0 1218672000 I am delighted with my nail care kit. I am in ... health personal care medical supplies equipment daily living aids 52 52 1.000000
18850 B000634EEO NaN A3A6B6RYANZYJN 6/6 1.0 1244937600 We bought this yesterday at Petco. Our betta, ... pet supplies fish aquatic pets aquarium starter kits 6 6 1.000000
19395 B000GUFFQS NaN A1W0PU8IJ06LJJ 3/3 5.0 1249603200 I put this on my grandmother's bed. It was ver... health personal care medical supplies equipment beds accessories 3 3 1.000000
20193 B000GUFFQS NaN A3URJ91T1WXQHE 1/1 5.0 1255392000 My father, who had a stroke several years ago,... health personal care medical supplies equipment beds accessories 1 1 1.000000
24905 B000GUFFQS NaN A3IEU3BT51AUKF 0/0 5.0 1284336000 My mother was issued a hospital bed to prevent... health personal care medical supplies equipment beds accessories 0 0 NaN
25759 B000PC4KX6 NaN A2SBECPKOH229L 1/1 4.0 1288828800 I bought these after they were used on my son ... health personal care medical supplies equipment health monitors 1 1 1.000000
26912 B000GUFFQS NaN A2FXS0OQKM3JQD 0/0 4.0 1294012800 Purchased it for older neighbor and she loves ... health personal care medical supplies equipment beds accessories 0 0 NaN
Handy prepackaged health
0s completed at 11:12PM


Commands
+ Code
+ Text
RAM
Disk
26912 B000GUFFQS NaN A2FXS0OQKM3JQD 0/0 4.0 1294012800 Purchased it for older neighbor and she loves ... health personal care medical supplies equipment beds accessories 0 0 NaN
31253 B000PJJNRC NaN A1JL9EIJJBARM1 1/1 5.0 1312588800 Handy prepackaged alcohol wipes a great bulk p... health personal care health care first aid 1 1 1.000000
32483 B000GUFFQS NaN A2TH88LHARKKXC 2/2 5.0 1317772800 I ordered this bedcane for my husband who is 6... health personal care medical supplies equipment beds accessories 2 2 1.000000
39455 B000GUFFQS NaN A1H5TCWJB5CV77 0/0 5.0 1342224000 I purchased another bedrail that was too low f... health personal care medical supplies equipment beds accessories 0 0 NaN
10. Convert 'Time' column from UNIX timestamp to readable date
df['ReviewDate'] = pd.to_datetime(df['Time'], unit='s', errors='coerce')
df[['Time', 'ReviewDate']].head()
Time ReviewDate
0 -1 1969-12-31 23:59:59
1 860630400 1997-04-10 00:00:00
2 883008000 1997-12-25 00:00:00
3 897696000 1998-06-13 00:00:00
4 911865600 1998-11-24 00:00:00