


Name :- Sumit Balasaheb Jadhav

Batch :- CS6

Roll No :- CS6-61

PRN :- 202401080036

 Overview of Colaboratory Features  Cannot save changes

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text Copy to Drive

Files

Analyze your files with code written by Gemini Upload

..

sample_data

README.md

Reviews.csv.zip

0s [30] print(df['ProductId'].value_counts().head(5))

ProductId	
B007JFMH8M	913
B002QWHDJOU	632
B0026RQTGE	632
B002QW8P89S	632
B002QW8P8H0	632

Name: count, dtype: int64

0s [31] print("Reviews with score 5:", (df['Score'] == 5).sum())

Reviews with score 5: 363122

RAM

Disk

Share Gemini A



Share



Gemini



Commands + Code + Text Copy to Drive

RAM Disk

Files

Analyze your files with code written by Gemini

Upload

sample_data

README.md
Reviews.csv.zip
anscombe.json
california_housin...
california_housin...
mnist_test.csv
mnist_train_sma...

Disk 70.51 GB available

```
[39] df['TextLength'] = df['Text'].str.len()
top5_users_longest = df.groupby('UserId')['TextLength'].mean().sort_values(ascending=False).head(5)
print("Top 5 users with longest average review text:\n", top5_users_longest)
```

Top 5 users with longest average review text:

UserId	
A24HWYHR28JJ7A	12743.5
A1kW22EAVQNM56	12331.0
#oc-R1ZR5L29T4LSAE	11861.0
A3AGQXIC8XC0JG	10327.0
A150QS4ITB3XJ	9919.0

Name: TextLength, dtype: float64


```
[40] love_avg_score = df[df['Text'].str.contains('love', case=False, na=False)]['Score'].mean()
print("Average score for reviews mentioning 'love':", love_avg_score)
```

Average score for reviews mentioning 'love': 4.5218330868478755

```
df['Time'] = pd.to_datetime(df['Time'], unit='s')
reviews_2012 = df[df['Time'].dt.year == 2012]
print("Reviews written in 2012:", reviews_2012.shape[0])
```

Reviews written in 2012: 198659

```
[42] df['WordCount'] = df['Text'].str.split().str.len()
long_reviews = df[df['WordCount'] > 100]
print("Reviews with more than 100 words:", long_reviews.shape[0])
```

 Overview of Colaboratory Features Cannot save changes

File Edit View Insert Runtime Tools Help

Q Commands

+ Code + Text

Copy to Drive

✓

RAM

Disk

↕

↕

Files

Analyze your files with code written by Gemini

Upload

..

sample_data

README.md

Reviews.csv.zip

anscombe.json

california_housin...

california_housin...

mnist_test.csv

mnist_train_sma...

0s

[34] df['Time'] = pd.to_datetime(df['Time'], unit='s') # Convert timestamp

print("Earliest review date:", df['Time'].min())

↕

Earliest review date: 1999-10-08 00:00:00

0s

[35] most_active_user = df['UserId'].value_counts().idxmax()

print("User with most reviews:", most_active_user)

↕

User with most reviews: A30XHLG6DIBRW8

17s

[36] df['WordCount'] = df['Text'].str.split().str.len()

print("Average word count:", df['WordCount'].mean())

↕

Average word count: 80.26402312236347

0s

▶

most_1star_product = df[df['Score'] == 1]['ProductId'].value_counts().idxmax()

print("Product with most 1-star reviews:", most_1star_product)

↕

Product with most 1-star reviews: B000X1Q1G8

3s

[38] bad_count = df['Text'].str.contains('bad', case=False, na=False).sum()

print("Reviews containing 'bad':", bad_count)

↕

Reviews containing 'bad': 29165

Disk

70.51 GB available

```
B00004RAMV      2.111111
B00006G930      2.875000
B00006IDJO      1.000000
B000084EKD      2.000000
...
B009MRNOR8      1.000000
B00907B1I0      1.000000
B009P4KMZA      1.000000
B009UOFTUT      1.000000
B009UOFU20      1.000000
Name: Score, Length: 7657, dtype: float64
```

```
0s print("Earliest review date:", df['Time'].min())
```

```
Earliest review date: 1999-10-08 00:00:00
```

```
0s [46] print("Latest review date:", df['Time'].max())
```

```
Latest review date: 2012-10-26 00:00:00
```

```
3s [44] refund_return_count = df['Text'].str.contains('refund|return', case=False, na=False).sum()
print("Reviews mentioning 'refund' or 'return':", refund_return_count)
```

```
Reviews mentioning 'refund' or 'return': 9742
```

```
0s [25] print("Total reviews:", df.shape[0])
```

```
Total reviews: 568454
```

Overview of Colaboratory Features

File Edit View Insert Runtime Tools Help

Cannot save changes

Share

Gemini

A

Commands

+ Code + Text

Copy to Drive

RAM

Disk

Files

Analyze your files with code written by Gemini

Upload

..

sample_data

README.md

Reviews.csv.zip

anscombe.json

california_housin...

california_housin...

mnist_test.csv

mnist_train_sma...

[24] df.head()

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...

print("Unique products:", df['ProductId'].nunique())

Unique products: 74258


```
[42] df['WordCount'] = df['Text'].str.split().str.len()
long_reviews = df[df['WordCount'] > 100]
print("Reviews with more than 100 words:", long_reviews.shape[0])
```

➡ Reviews with more than 100 words: 136017

```
low_score_products = df.groupby('ProductId')['Score'].mean()
products_below_3 = low_score_products[low_score_products < 3]
print("Products with average score below 3:\n", products_below_3)
```

➡ Products with average score below 3:

ProductId	Score
B00002N8SM	1.789474
B00004RAMV	2.111111
B00006G930	2.875000
B00006IDJO	1.000000
B000084EKD	2.000000
...	...
B009MRNORS	1.000000
B009O7B1I0	1.000000
B009P4KMZA	1.000000
B009UOFTUI	1.000000
B009UOFU20	1.000000

Name: Score, Length: 7657, dtype: float64

```
[45] print("Earliest review date:", df['Time'].min())
```

➡ Earliest review date: 1999-10-08 00:00:00



Share



Gemini



Commands + Code + Text Copy to Drive

RAM Disk

Files

Analyze your files with code written by Gemini Upload



- sample_data
- README.md
- Reviews.csv.zip
- anscombe.json
- california_housin...
- california_housin...
- mnist_test.csv
- mnist_train_sma...

Disk 70.51 GB available

```
excellent_reviews = df[df['Text'].str.contains('excellent', case=False, na=False)]
print(excellent_reviews)
```


	Id	ProductId	UserId	ProfileName	
28	29	B00144C10S	A2F4LZVGFLD10B	DaisyH	
34	35	B001E05QW8	A2EBGOGOWCRU5H	CorbyJames	
40	41	B001E05QW8	AQCY5KRO7489S	Garrett	
43	44	B001E05QW8	A17DW6SUOC70DJ	Mother of 9	
106	107	B004K2IHU0	A1ZKFQLHFZAEH9	S. J. Monson "world citizen"	
...	
568338	568339	B000KOSIP0	A3EDVUU7LRFZEJ	Bob Dobolina	
568389	568390	B001EQ506Y	A2LGOMSYQJ5PZS	N. Hjort	
568402	568403	B001EQ506Y	A1A3H22VZYUKW	msfreixy	
568403	568404	B001EQ506Y	A2891E3BMAKGYN	PCNiles "reader/writer"	
568437	568438	B000NY4SAG	A1RKKPSXF9QIZF	Allen "RcDriver"	
	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	
28	0	0	5	1338854400	
34	9	9	5	1175212800	
40	0	0	5	1303430400	
43	0	0	5	1255392000	
106	2	8	3	1236384000	
...	
568338	1	1	2	1337385600	
568389	2	2	5	1257552000	
568402	1	3	1	1187049600	
568403	0	6	2	1264118400	
568437	4	4	5	1233792000	

Summary \

28 YUMMY!

34 Great Irish oatmeal for those in a hurry!

0s completed at 12:19 AM

 Overview of Colaboratory Features Cannot save changes

File Edit View Insert Runtime Tools Help

Q Commands

+ Code + Text

Copy to Drive

✓

RAM

Disk

Files

Analyze your files with code written by Gemini Upload

..

sample_data

README.md

Reviews.csv.zip

anscombe.json

california_housin...

california_housin...

mnist_test.csv

mnist_train_sma...

[27] print("Unique products:", df['ProductId'].nunique())

Unique products: 74258

[28] print("Average score:", df['Score'].mean())

Average score: 4.183198640523243

df['TextLength'] = df['Text'].str.len()
longest_review = df.loc[df['TextLength'].idxmax()]
print("Longest review:\n", longest_review)

Longest review:

Id	290808
ProductId	B005HG9ESG
UserId	A24HWMYHR28JJ7A
ProfileName	ƒůŽźŸ ωŬ&#...
HelpfulnessNumerator	10
HelpfulnessDenominator	11
Score	3
Time	1348272000
Summary	An okay filtered municipal tap water with slig...
Text	Fuzzy Wuzzy's Summary: *** Somewhat recom...
TextLength	21409

Name: 290807, dtype: object