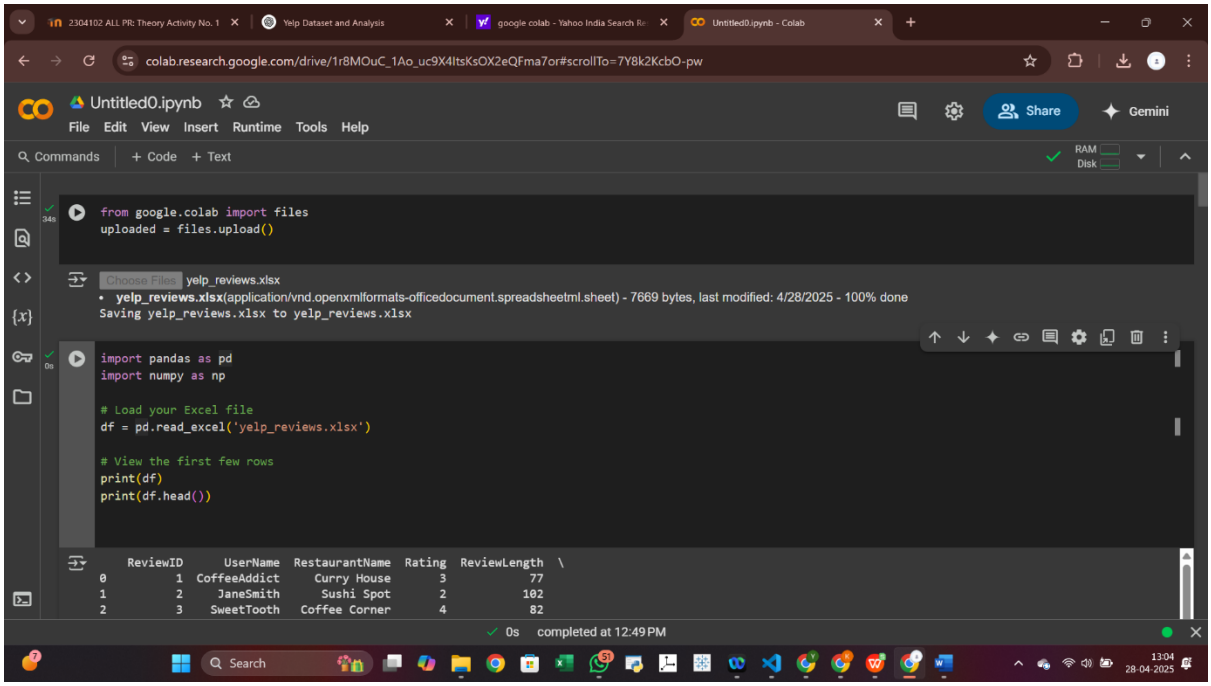
Activity no :- 1

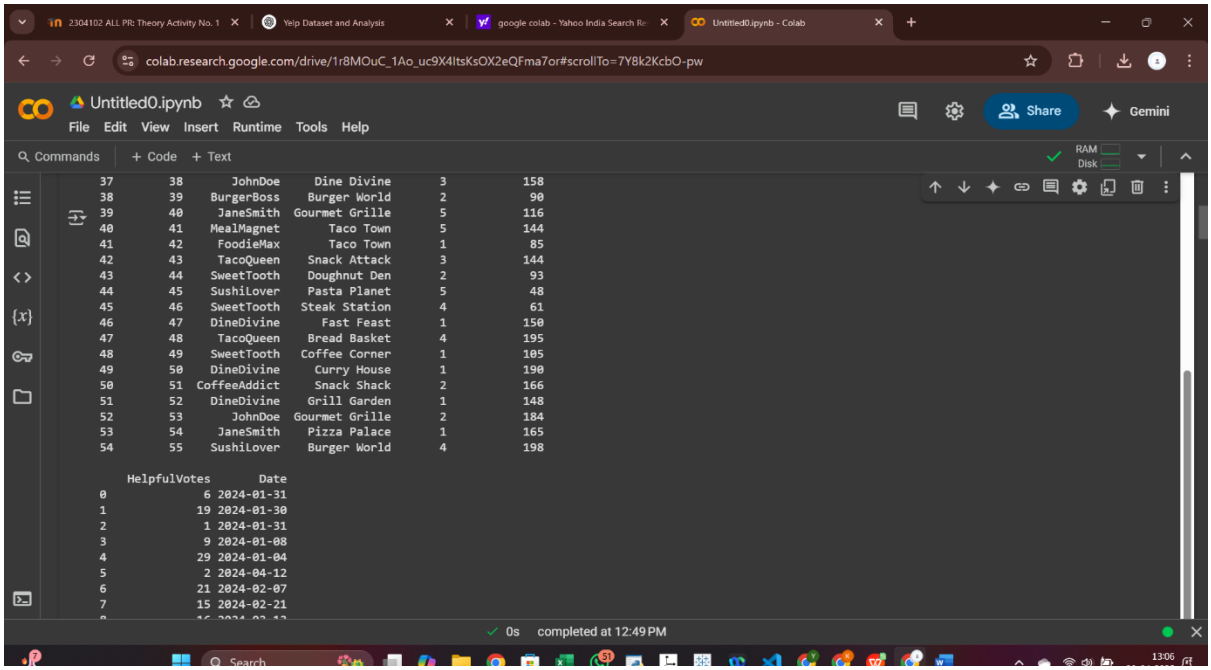
By:-

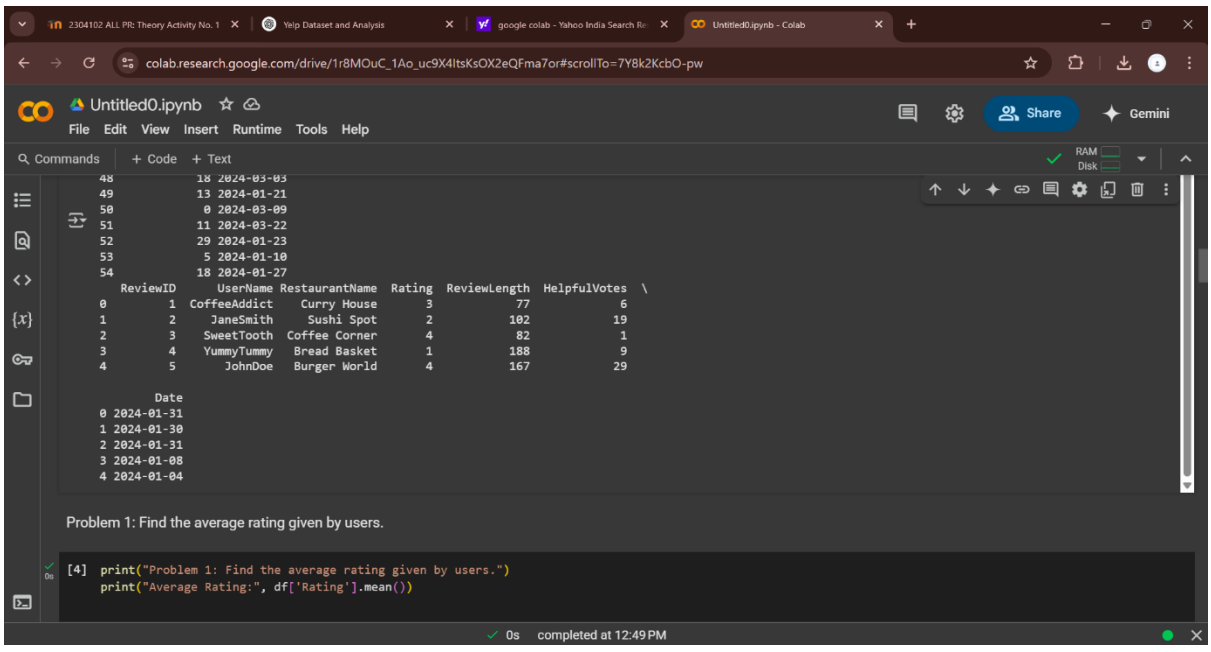
Kishor Kaple

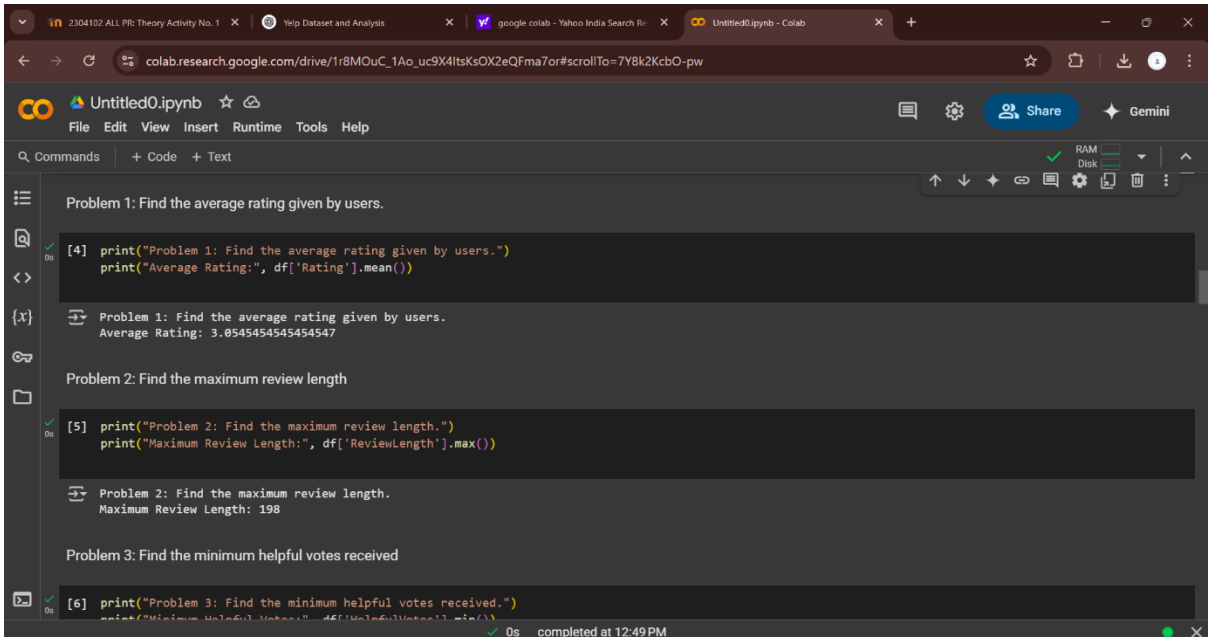
CS7-40

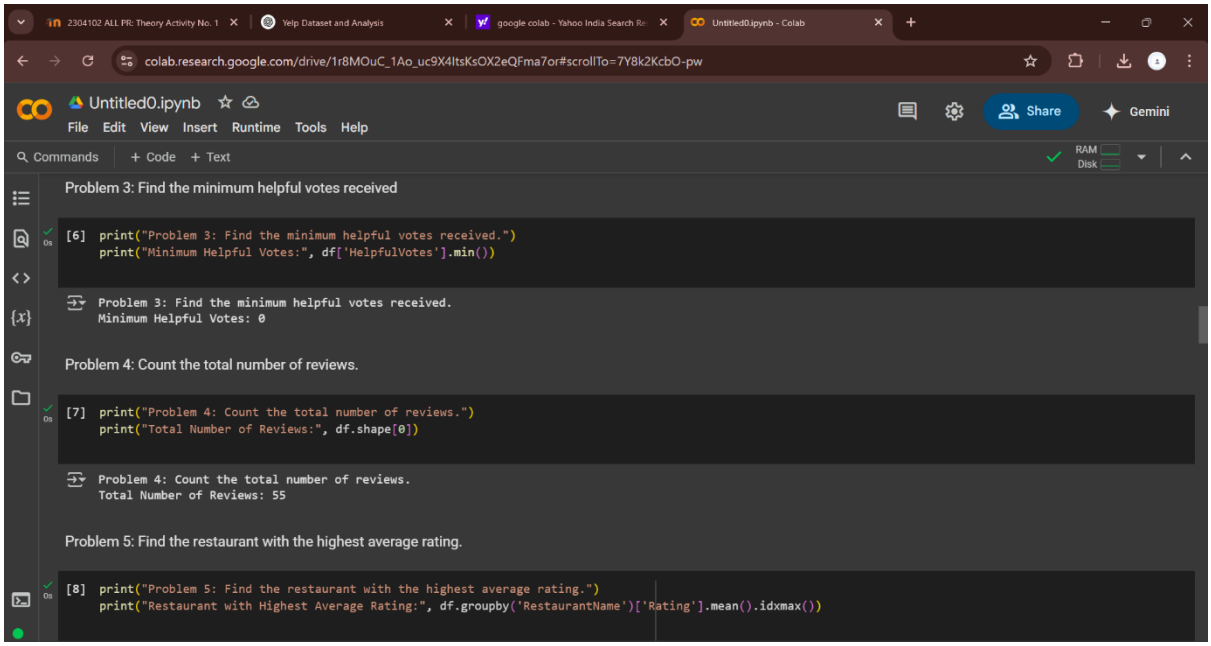
202401110036

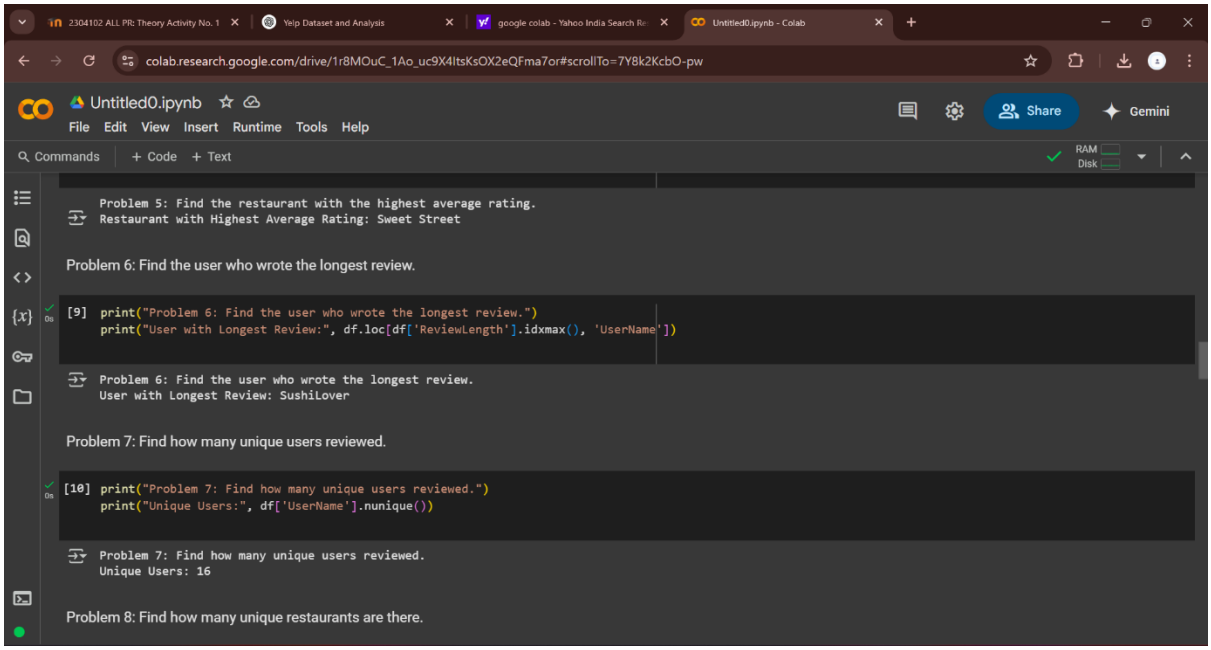

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
from google.colab import files
uploaded = files.upload()
Choose Files
yelp_reviews.xlsx
yelp_reviews.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 7669 bytes, last modified: 4/28/2025 - 100% done
Saving yelp_reviews.xlsx to yelp_reviews.xlsx
import pandas as pd
import numpy as np
# Load your Excel file
df = pd.read_excel('yelp_reviews.xlsx')
# View the first few rows
print(df)
print(df.head())
ReviewID UserName RestaurantName Rating ReviewLength \
0 1 CoffeeAddict Curry House 3 77
1 2 JaneSmith Sushi Spot 2 102
2 3 SweetTooth Coffee Corner 4 82
0s completed at 12:49 PM
Search
13:04
28-04-2025

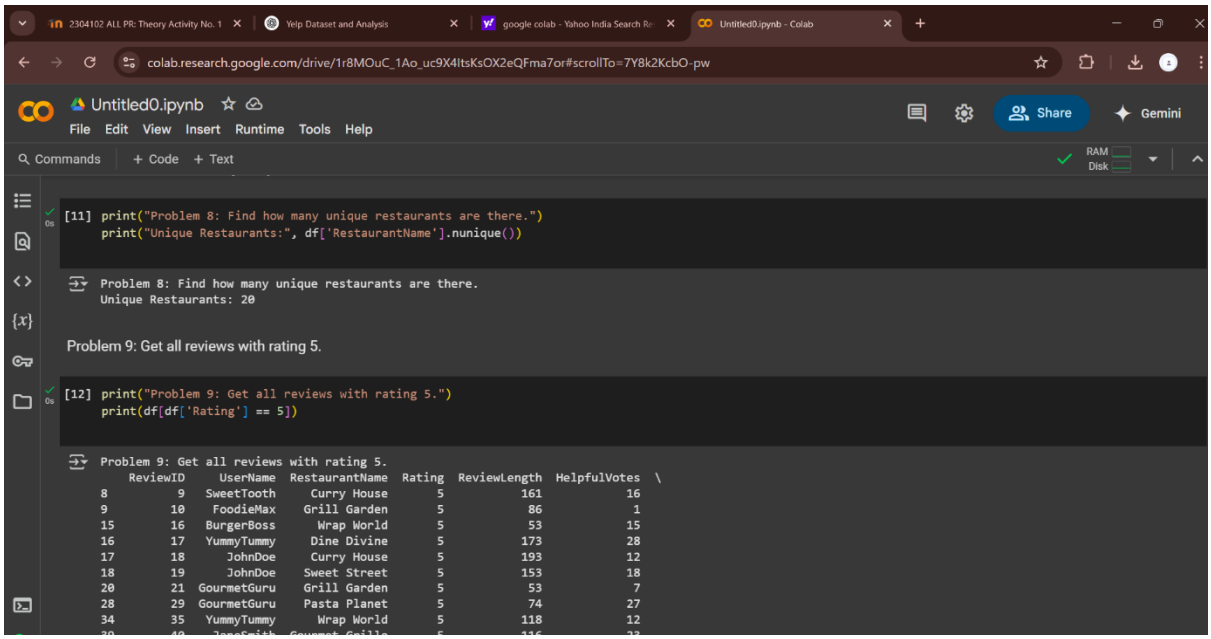

Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
37 38 JohnDoe Dine Divine 3 158
38 39 BurgerBoss Burger World 2 90
39 40 JaneSmith Gourmet Grille 5 116
40 41 MealMagnet Taco Town 5 144
41 42 FoodieMax Taco Town 1 85
42 43 TacoQueen Snack Attack 3 144
43 44 SweetTooth Doughnut Den 2 93
44 45 SushiLover Pasta Planet 5 48
45 46 SweetTooth Steak Station 4 61
46 47 DineDivine Fast Feast 1 150
47 48 TacoQueen Bread Basket 4 195
48 49 SweetTooth Coffee Corner 1 105
49 50 DineDivine Curry House 1 190
50 51 CoffeeAddict Snack Shack 2 166
51 52 DineDivine Grill Garden 1 148
52 53 JohnDoe Gourmet Grille 2 184
53 54 JaneSmith Pizza Palace 1 165
54 55 SushiLover Burger World 4 198
HelpfulVotes Date
0 6 2024-01-31
1 19 2024-01-30
2 1 2024-01-31
3 9 2024-01-08
4 29 2024-01-04
5 2 2024-04-12
6 21 2024-02-07
7 15 2024-02-21
0s completed at 12:49 PM

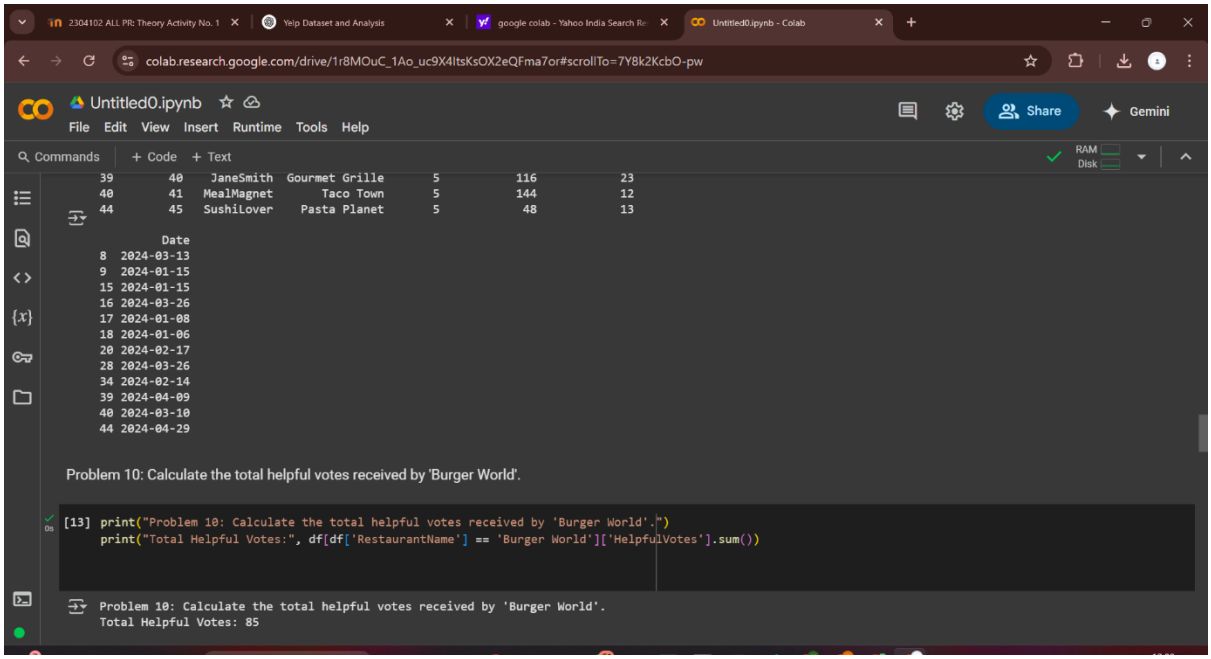

Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
48 18 2024-03-03
49 13 2024-01-21
50 0 2024-03-09
51 11 2024-03-22
52 29 2024-01-23
53 5 2024-01-10
54 18 2024-01-27
ReviewID UserName RestaurantName Rating ReviewLength HelpfulVotes \
0 1 CoffeeAddict Curry House 3 77 6
1 2 JaneSmith Sushi Spot 2 102 19
2 3 SweetTooth Coffee Corner 4 82 1
3 4 YummyTummy Bread Basket 1 188 9
4 5 JohnDoe Burger World 4 167 29
Date
0 2024-01-31
1 2024-01-30
2 2024-01-31
3 2024-01-08
4 2024-01-04
Problem 1: Find the average rating given by users.
[4] print("Problem 1: Find the average rating given by users.")
print("Average Rating:", df['Rating'].mean())
0s completed at 12:49 PM

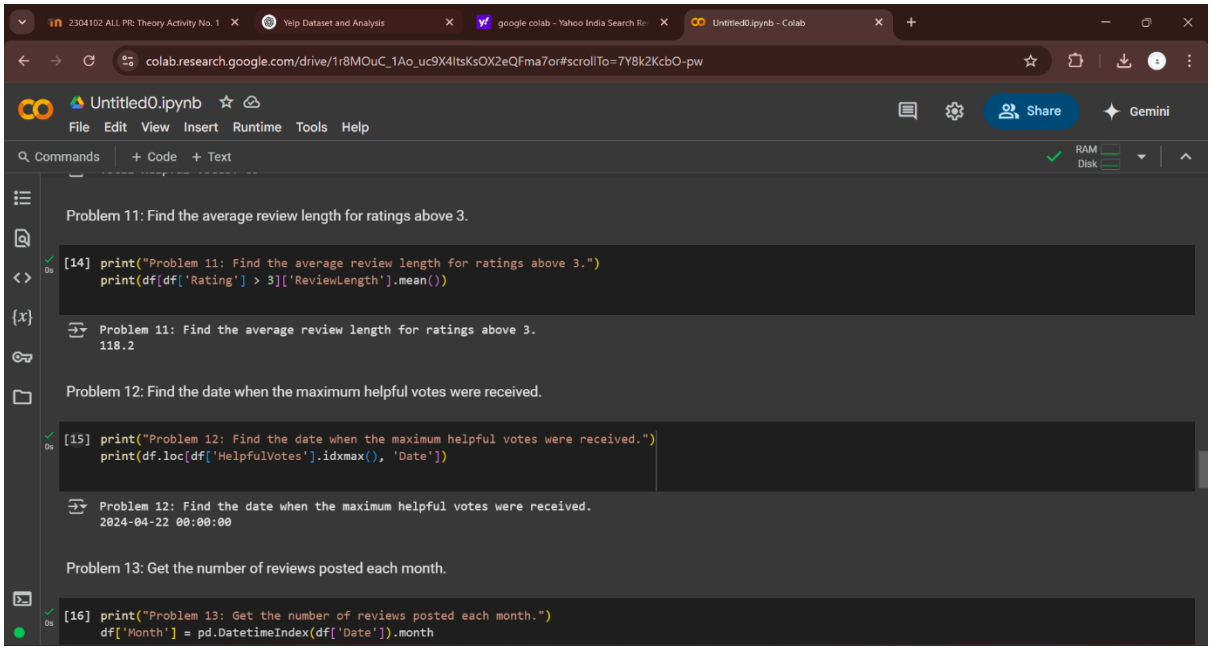

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
Problem 1: Find the average rating given by users.
[4] print("Problem 1: Find the average rating given by users.")
print("Average Rating:", df['Rating'].mean())
Problem 1: Find the average rating given by users.
Average Rating: 3.0545454545454547
Problem 2: Find the maximum review length
[5] print("Problem 2: Find the maximum review length.")
print("Maximum Review Length:", df['ReviewLength'].max())
Problem 2: Find the maximum review length.
Maximum Review Length: 198
Problem 3: Find the minimum helpful votes received
[6] print("Problem 3: Find the minimum helpful votes received.")
0s completed at 12:49 PM

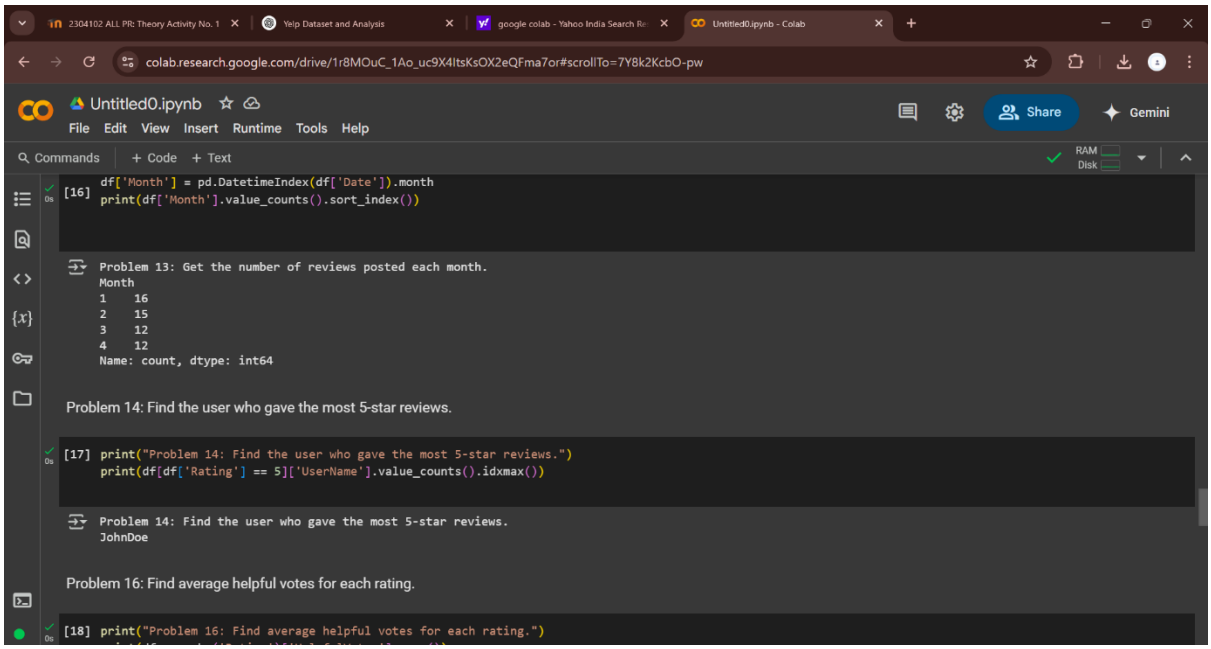

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
Problem 3: Find the minimum helpful votes received
[6] print("Problem 3: Find the minimum helpful votes received.")
print("Minimum Helpful Votes:", df['HelpfulVotes'].min())
Problem 3: Find the minimum helpful votes received.
Minimum Helpful Votes: 0
Problem 4: Count the total number of reviews.
[7] print("Problem 4: Count the total number of reviews.")
print("Total Number of Reviews:", df.shape[0])
Problem 4: Count the total number of reviews.
Total Number of Reviews: 55
Problem 5: Find the restaurant with the highest average rating.
[8] print("Problem 5: Find the restaurant with the highest average rating.")
print("Restaurant with Highest Average Rating:", df.groupby('RestaurantName')['Rating'].mean().idxmax())

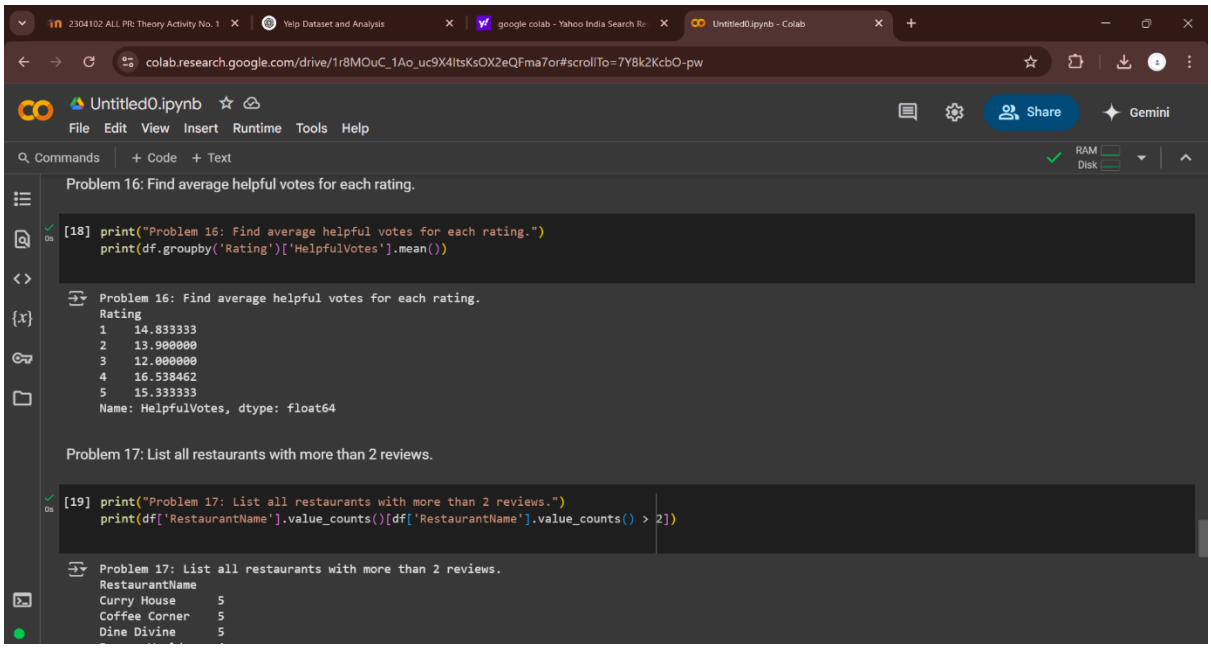

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
Problem 5: Find the restaurant with the highest average rating.
Restaurant with Highest Average Rating: Sweet Street
Problem 6: Find the user who wrote the longest review.
[9] print("Problem 6: Find the user who wrote the longest review.")
print("User with Longest Review:", df.loc[df['ReviewLength'].idxmax(), 'UserName'])
Problem 6: Find the user who wrote the longest review.
User with Longest Review: SushiLover
Problem 7: Find how many unique users reviewed.
[10] print("Problem 7: Find how many unique users reviewed.")
print("Unique Users:", df['UserName'].nunique())
Problem 7: Find how many unique users reviewed.
Unique Users: 16
Problem 8: Find how many unique restaurants are there.

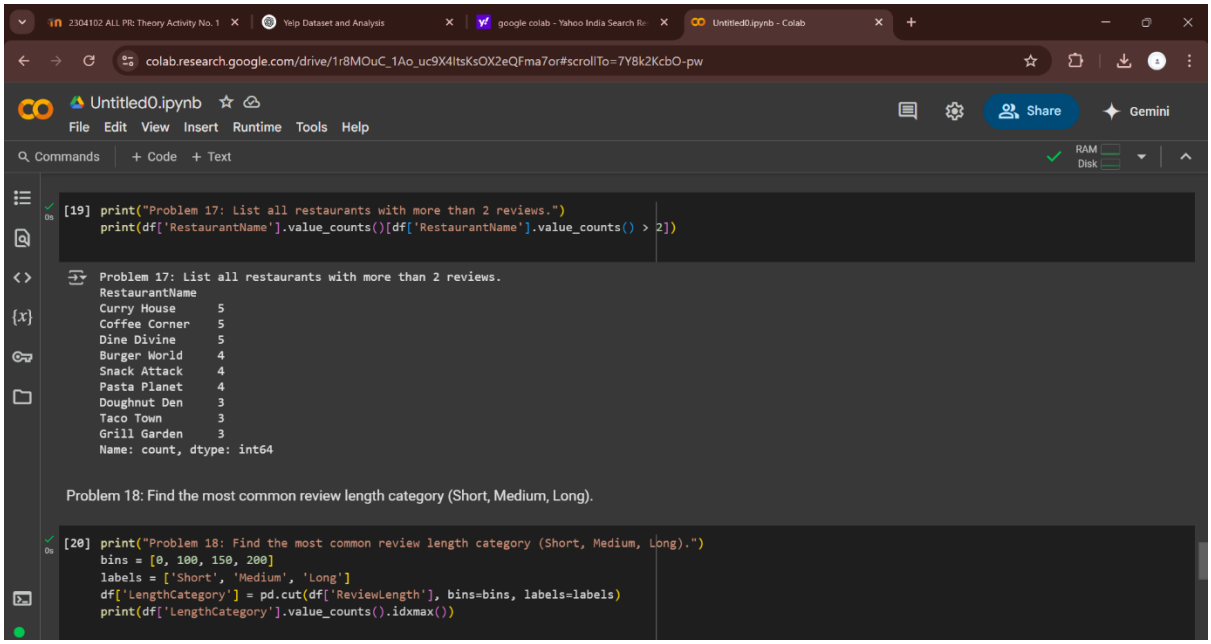

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
[11] print("Problem 8: Find how many unique restaurants are there.")
print("Unique Restaurants:", df['RestaurantName'].nunique())
Problem 8: Find how many unique restaurants are there.
Unique Restaurants: 20
Problem 9: Get all reviews with rating 5.
[12] print("Problem 9: Get all reviews with rating 5.")
print(df[df['Rating'] == 5])
Problem 9: Get all reviews with rating 5.
ReviewID UserName RestaurantName Rating ReviewLength HelpfulVotes \
8 9 SweetTooth Curry House 5 161 16
9 10 FoodieMax Grill Garden 5 86 1
15 16 BurgerBoss Wrap World 5 53 15
16 17 YummyTummy Dine Divine 5 173 28
17 18 JohnDoe Curry House 5 193 12
18 19 JohnDoe Sweet Street 5 153 18
20 21 GourmetGuru Grill Garden 5 53 7
28 29 GourmetGuru Pasta Planet 5 74 27
34 35 YummyTummy Wrap World 5 118 12

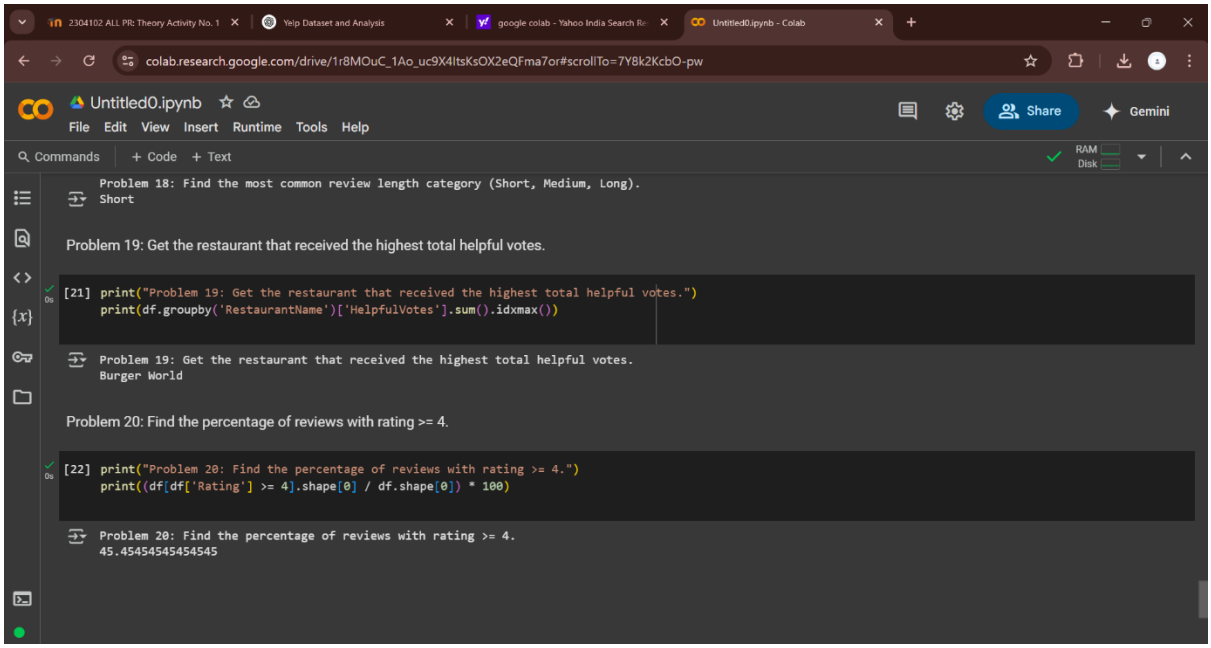

2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
39 40 JaneSmith Gourmet Grille 5 116 23
40 41 MealMagnet Taco Town 5 144 12
44 45 SushiLover Pasta Planet 5 48 13
Date
8 2024-03-13
9 2024-01-15
15 2024-01-15
16 2024-03-26
17 2024-01-08
18 2024-01-06
20 2024-02-17
28 2024-03-26
34 2024-02-14
39 2024-04-09
40 2024-03-10
44 2024-04-29
Problem 10: Calculate the total helpful votes received by 'Burger World'.
[13] print("Problem 10: Calculate the total helpful votes received by 'Burger World'.")
print("Total Helpful Votes:", df[df['RestaurantName'] == 'Burger World']['HelpfulVotes'].sum())
Problem 10: Calculate the total helpful votes received by 'Burger World'.
Total Helpful Votes: 85


2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
Problem 11: Find the average review length for ratings above 3.
[14] print("Problem 11: Find the average review length for ratings above 3.")
print(df[df['Rating'] > 3]['ReviewLength'].mean())
Problem 11: Find the average review length for ratings above 3.
118.2
Problem 12: Find the date when the maximum helpful votes were received.
[15] print("Problem 12: Find the date when the maximum helpful votes were received.")
print(df.loc[df['HelpfulVotes'].idxmax(), 'Date'])
Problem 12: Find the date when the maximum helpful votes were received.
2024-04-22 00:00:00
Problem 13: Get the number of reviews posted each month.
[16] print("Problem 13: Get the number of reviews posted each month.")
df['Month'] = pd.DatetimeIndex(df['Date']).month


2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
[16] df['Month'] = pd.DatetimeIndex(df['Date']).month
print(df['Month'].value_counts().sort_index())
Problem 13: Get the number of reviews posted each month.
Month
1 16
2 15
3 12
4 12
Name: count, dtype: int64
Problem 14: Find the user who gave the most 5-star reviews.
[17] print("Problem 14: Find the user who gave the most 5-star reviews.")
print(df[df['Rating'] == 5]['UserName'].value_counts().idxmax())
Problem 14: Find the user who gave the most 5-star reviews.
JohnDoe
Problem 16: Find average helpful votes for each rating.
[18] print("Problem 16: Find average helpful votes for each rating.")


2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
Problem 16: Find average helpful votes for each rating.
[18] print("Problem 16: Find average helpful votes for each rating.")
print(df.groupby('Rating')['HelpfulVotes'].mean())
Problem 16: Find average helpful votes for each rating.
Rating
1 14.833333
2 13.900000
3 12.000000
4 16.538462
5 15.333333
Name: HelpfulVotes, dtype: float64
Problem 17: List all restaurants with more than 2 reviews.
[19] print("Problem 17: List all restaurants with more than 2 reviews.")
print(df['RestaurantName'].value_counts()[df['RestaurantName'].value_counts() > 2])
Problem 17: List all restaurants with more than 2 reviews.
RestaurantName
Curry House 5
Coffee Corner 5
Dine Divine 5


2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
[19] print("Problem 17: List all restaurants with more than 2 reviews.")
print(df['RestaurantName'].value_counts()[df['RestaurantName'].value_counts() > 2])
Problem 17: List all restaurants with more than 2 reviews.
RestaurantName
Curry House 5
Coffee Corner 5
Dine Divine 5
Burger World 4
Snack Attack 4
Pasta Planet 4
Doughnut Den 3
Taco Town 3
Grill Garden 3
Name: count, dtype: int64
Problem 18: Find the most common review length category (Short, Medium, Long).
[20] print("Problem 18: Find the most common review length category (Short, Medium, Long).")
bins = [0, 100, 150, 200]
labels = ['Short', 'Medium', 'Long']
df['LengthCategory'] = pd.cut(df['ReviewLength'], bins=bins, labels=labels)
print(df['LengthCategory'].value_counts().idxmax())


2304102 ALL PR: Theory Activity No. 1
Yelp Dataset and Analysis
google colab - Yahoo India Search Re
Untitled0.ipynb - Colab
colab.research.google.com/drive/1r8MOuC_1Ao_uc9X4ItsKsOX2eQFma7or#scrollTo=7Y8k2KcbO-pw
Untitled0.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
Problem 18: Find the most common review length category (Short, Medium, Long).
Short
Problem 19: Get the restaurant that received the highest total helpful votes.
[21] print("Problem 19: Get the restaurant that received the highest total helpful votes.")
print(df.groupby('RestaurantName')['HelpfulVotes'].sum().idxmax())
Problem 19: Get the restaurant that received the highest total helpful votes.
Burger World
Problem 20: Find the percentage of reviews with rating >= 4.
[22] print("Problem 20: Find the percentage of reviews with rating >= 4.")
print((df[df['Rating'] >= 4].shape[0] / df.shape[0]) * 100)
Problem 20: Find the percentage of reviews with rating >= 4.
45.45454545454545