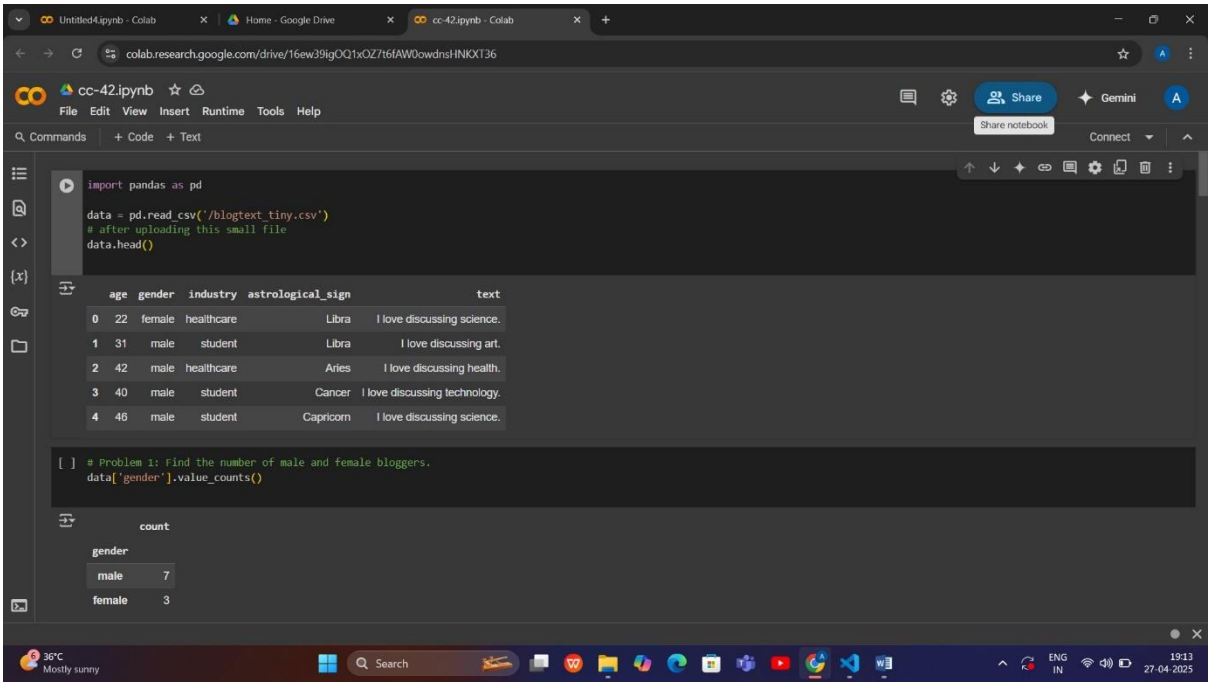
NAME – ASHWINI BASAVARAJ SIMPAGER

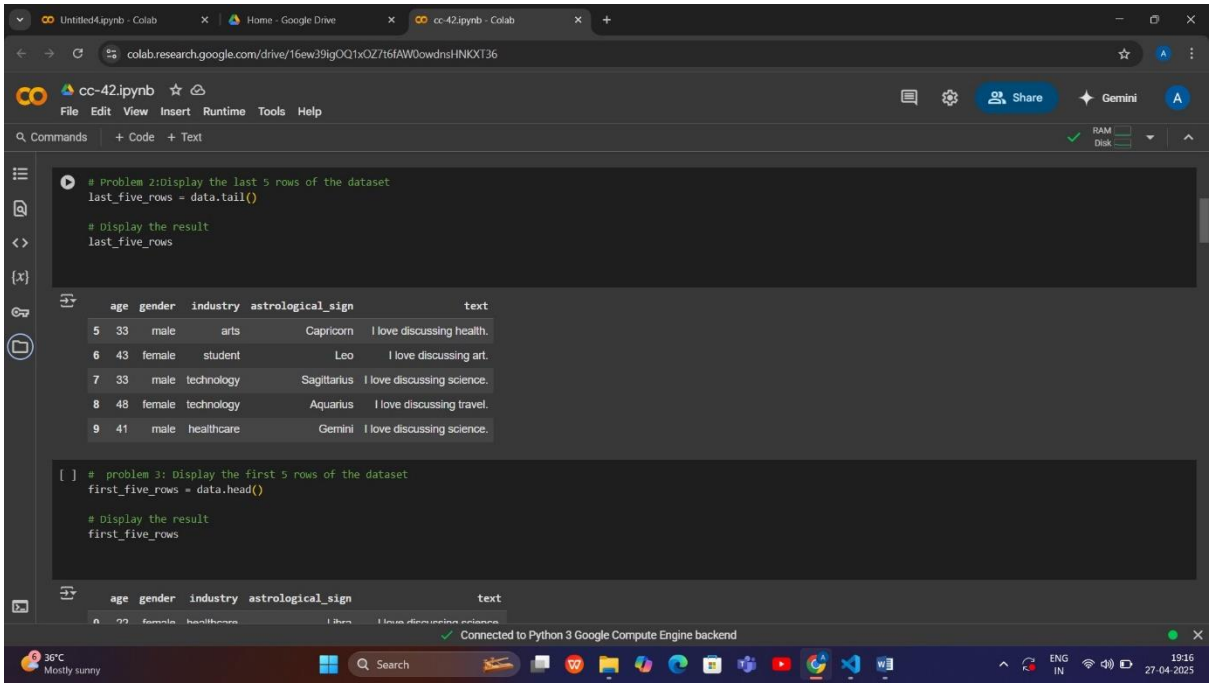
ROLL NO – CC-42

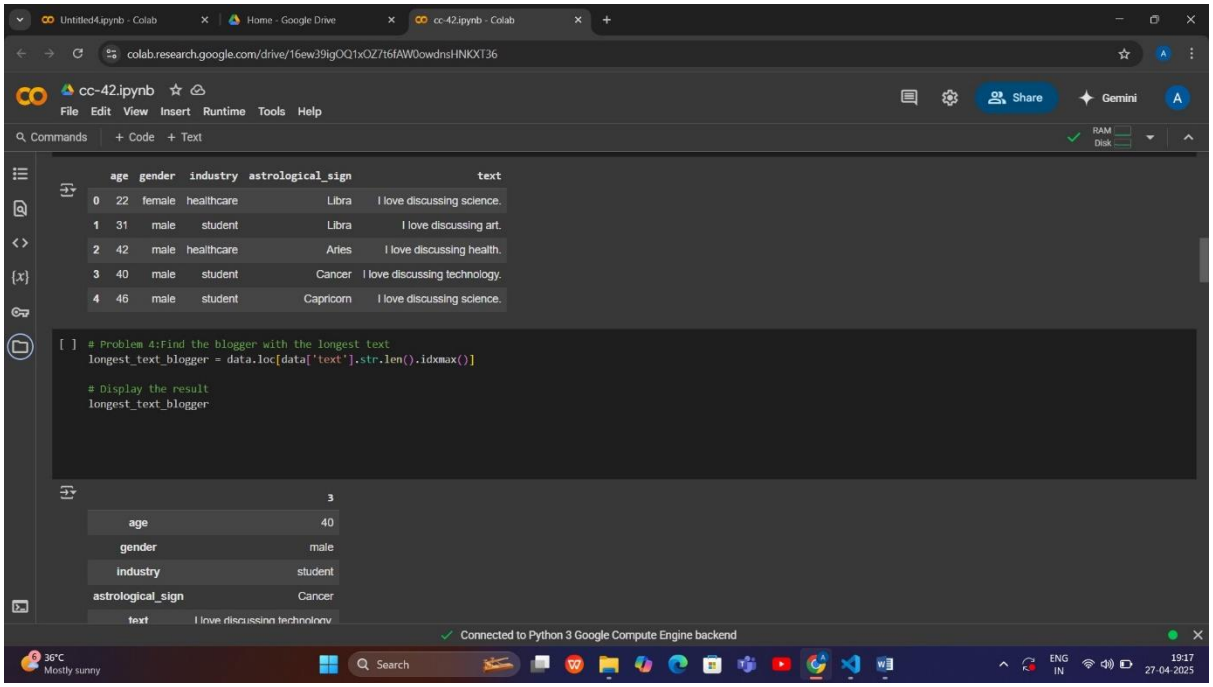
PRN – 202401050071

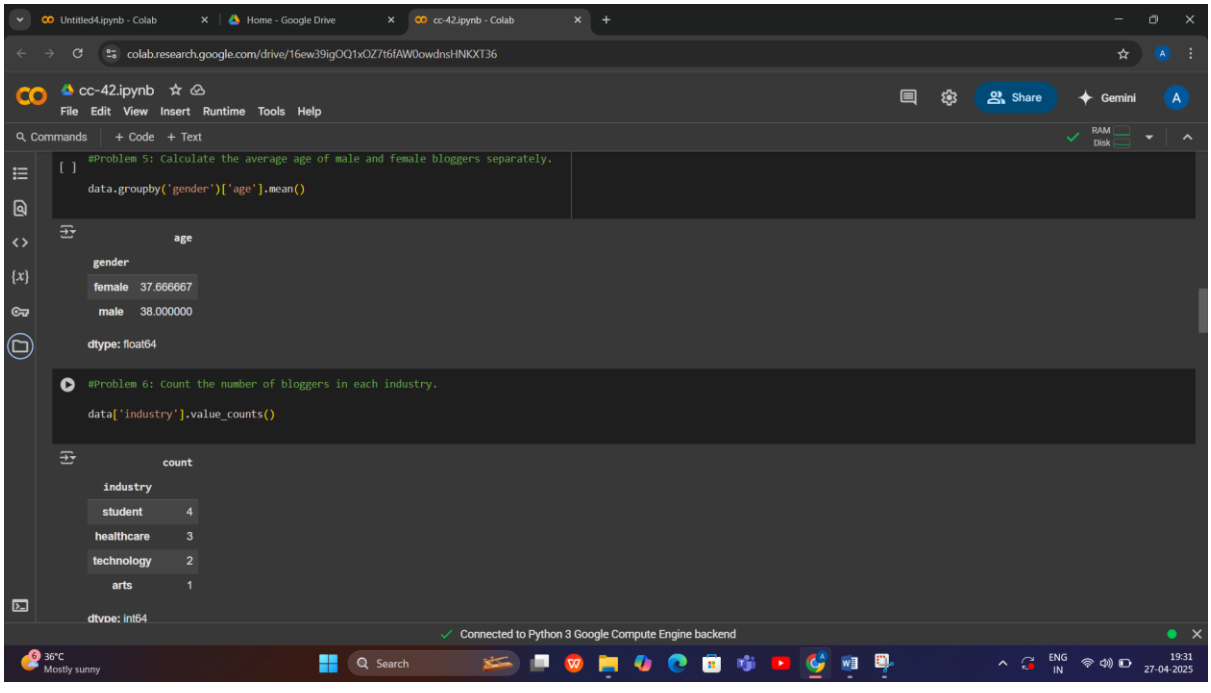
SUB – EDS

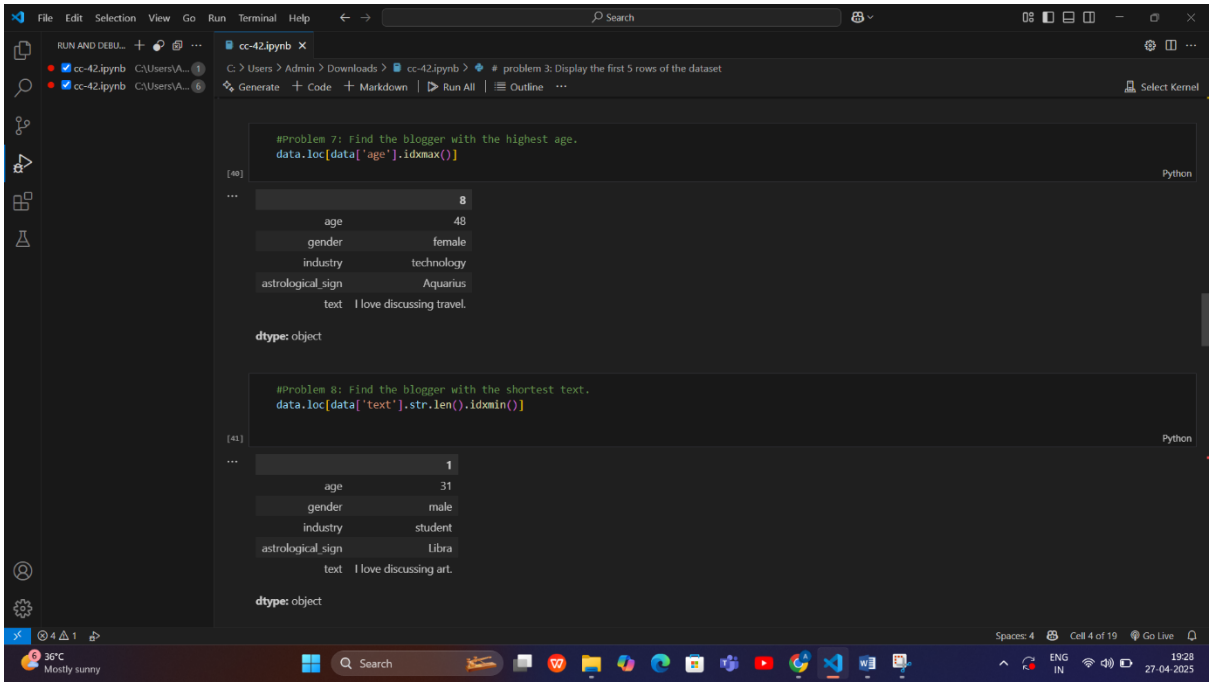
20 PROBLEM STATEMENT ON PAND

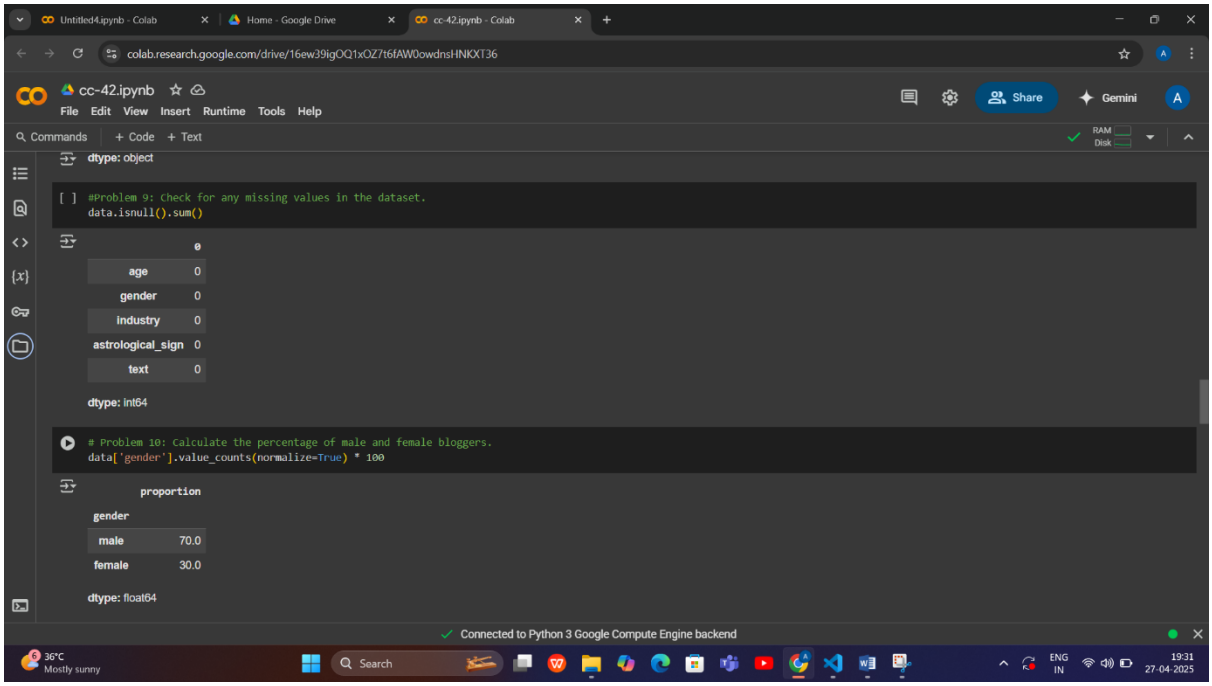


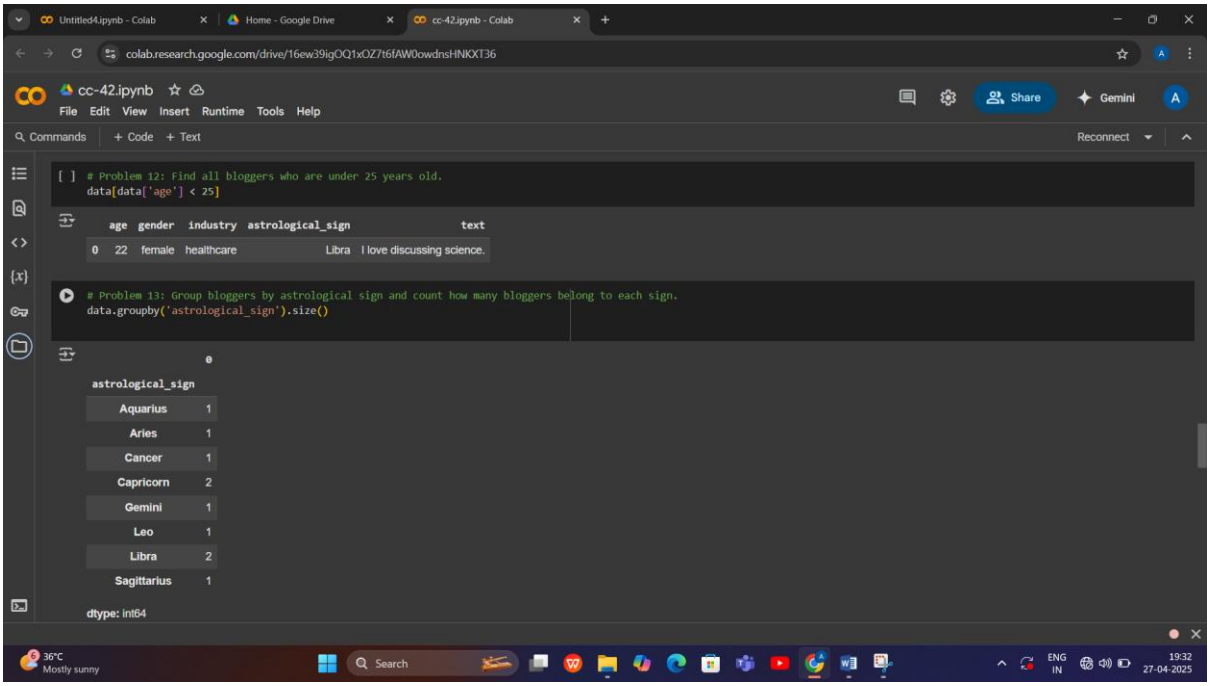


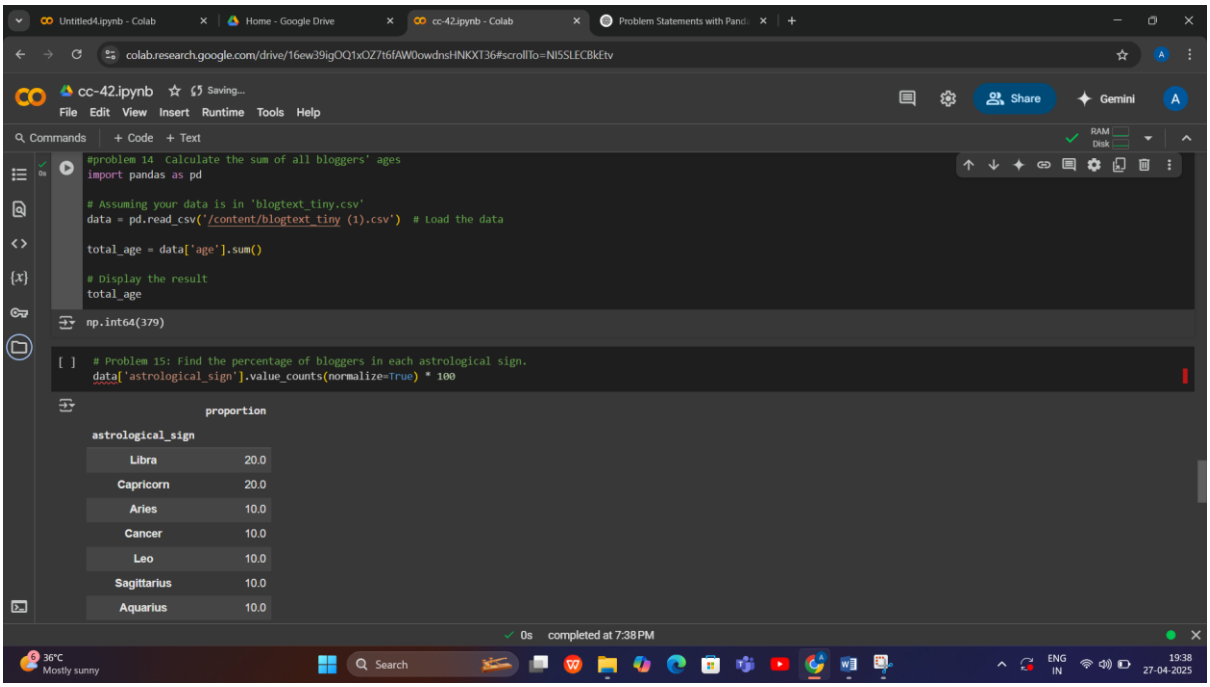

Untitled4.ipynb - Colab
Home - Google Drive
cc-42.ipynb - Colab
colab.research.google.com/drive/16ew39igOQ1xOZ7t6fAW0owdnsHNKXT36
cc-42.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
age gender industry astrological_sign text
0 22 female healthcare Libra I love discussing science.
1 31 male student Libra I love discussing art.
2 42 male healthcare Aries I love discussing health.
3 40 male student Cancer I love discussing technology.
4 46 male student Capricorn I love discussing science.
[ ] # Problem 4:Find the blogger with the longest text
longest_text_blogger = data.loc[data['text'].str.len().idxmax()]
# Display the result
longest_text_blogger
3
age 40
gender male
industry student
astrological_sign Cancer
text I love discussing technology.
Connected to Python 3 Google Compute Engine backend
36°C
Mostly sunny
Search
ENG
IN
19:17
27-04-2025

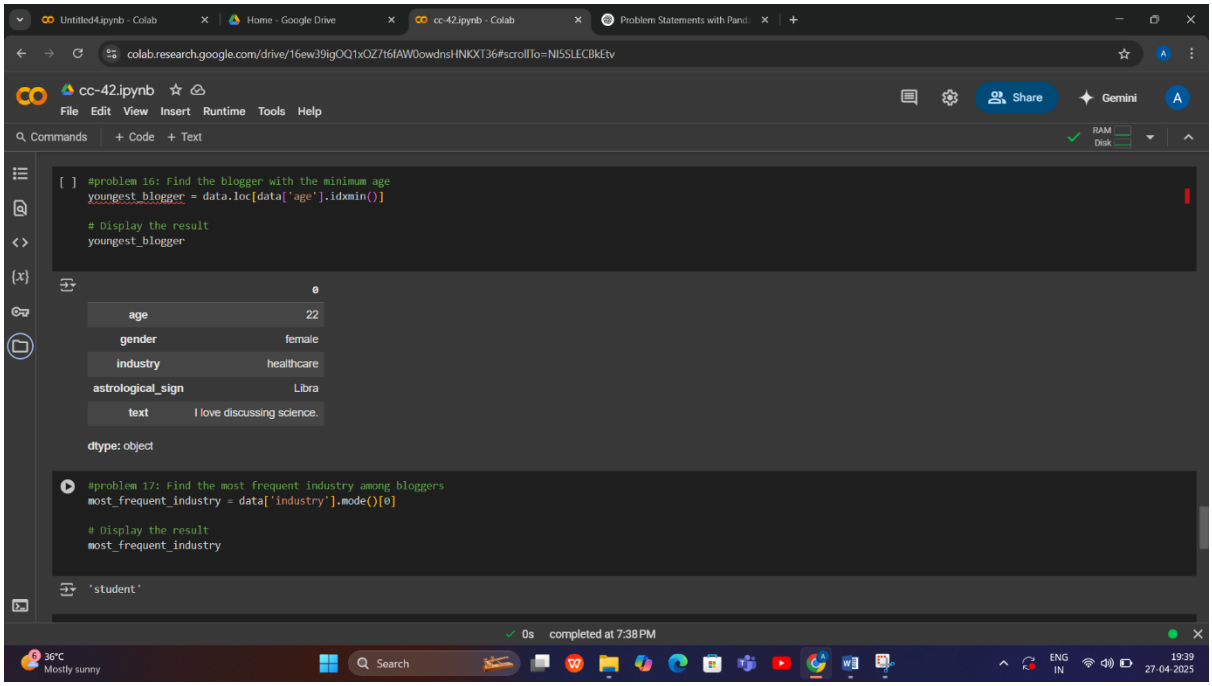

Untitled4.ipynb - Colab
Home - Google Drive
cc-42.ipynb - Colab
colab.research.google.com/drive/16ew39igOQ1xOZ7t6fAW0owdnsHNKXT36
cc-42.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
[ ] #Problem 5: Calculate the average age of male and female bloggers separately.
data.groupby('gender')['age'].mean()
age
gender
female 37.666667
male 38.000000
dtype: float64
#Problem 6: Count the number of bloggers in each industry.
data['industry'].value_counts()
count
industry
student 4
healthcare 3
technology 2
arts 1
dtype: int64
Connected to Python 3 Google Compute Engine backend
36°C
Mostly sunny
Search
ENG
IN
19:31
27-04-2025

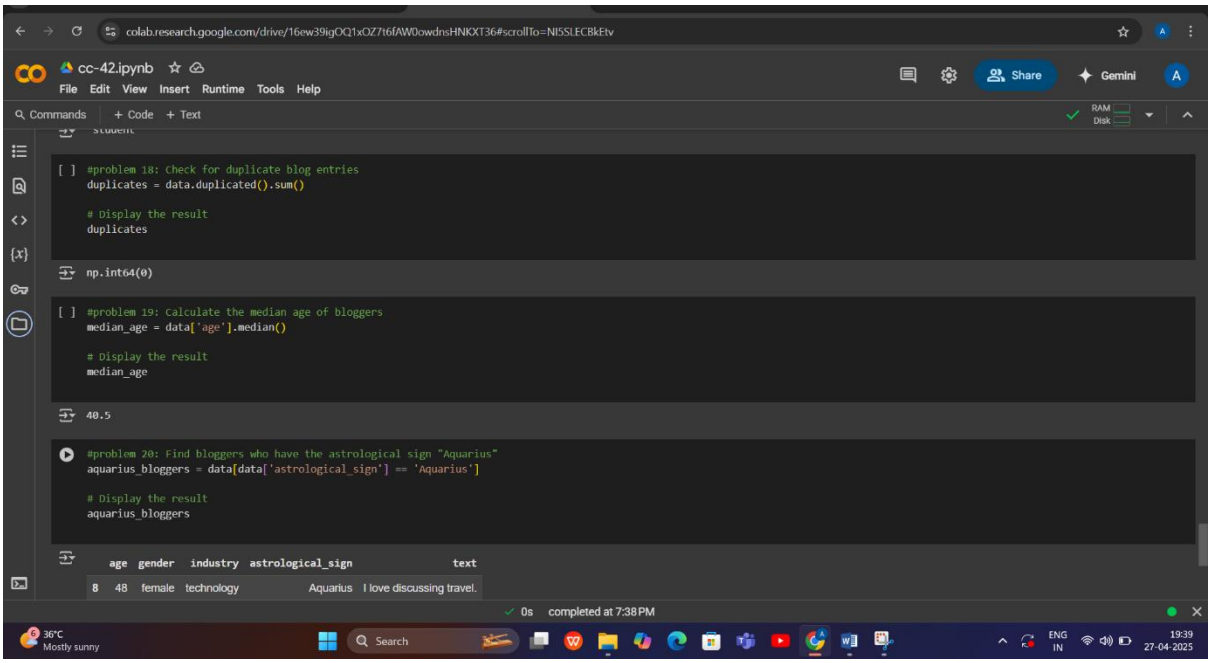

#Problem 7: Find the blogger with the highest age.
data.loc[data['age'].idxmax()]
8
age 48
gender female
industry technology
astrological_sign Aquarius
text I love discussing travel.
dtype: object
#Problem 8: Find the blogger with the shortest text.
data.loc[data['text'].str.len().idxmin()]
1
age 31
gender male
industry student
astrological_sign Libra
text I love discussing art.
dtype: object


dtype: object
#Problem 9: Check for any missing values in the dataset.
data.isnull().sum()
0
age 0
gender 0
industry 0
astrological_sign 0
text 0
dtype: int64
Problem 10: Calculate the percentage of male and female bloggers.
data['gender'].value_counts(normalize=True) * 100
proportion
gender
male 70.0
female 30.0
dtype: float64
Connected to Python 3 Google Compute Engine backend


Untitled4.ipynb - Colab
Home - Google Drive
cc-42.ipynb - Colab
colab.research.google.com/drive/16ew39igOQ1xOZ7t6fAW0owdnsHNKXT36
cc-42.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Reconnect
[ ] # Problem 12: Find all bloggers who are under 25 years old.
data[data['age'] < 25]
age gender industry astrological_sign text
0 22 female healthcare Libra I love discussing science.
# Problem 13: Group bloggers by astrological sign and count how many bloggers belong to each sign.
data.groupby('astrological_sign').size()
0
astrological_sign
Aquarius 1
Aries 1
Cancer 1
Capricorn 2
Gemini 1
Leo 1
Libra 2
Sagittarius 1
dtype: int64
36°C
Mostly sunny
Search
ENG
IN
19:32
27-04-2025


Untitled4.ipynb - Colab
Home - Google Drive
cc-42.ipynb - Colab
Problem Statements with Pandi
colab.research.google.com/drive/16ew39igOQ1xOZ7t6fAW0owdnsHNKXT36#scrollTo=NI5SLECBkEtv
cc-42.ipynb
Saving...
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
RAM
Disk
#problem 14 Calculate the sum of all bloggers' ages
import pandas as pd
# Assuming your data is in 'blogtext_tiny.csv'
data = pd.read_csv('/content/blogtext_tiny (1).csv') # Load the data
total_age = data['age'].sum()
# Display the result
total_age
np.int64(379)
[ ] # Problem 15: Find the percentage of bloggers in each astrological sign.
data['astrological_sign'].value_counts(normalize=True) * 100
proportion
astrological_sign
Libra 20.0
Capricorn 20.0
Aries 10.0
Cancer 10.0
Leo 10.0
Sagittarius 10.0
Aquarius 10.0
0s completed at 7:38 PM
36°C
Mostly sunny
Search
ENG
IN
19:38
27-04-2025


cc-42.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
#problem 16: Find the blogger with the minimum age
youngest_blogger = data.loc[data['age'].idxmin()]
# Display the result
youngest_blogger
0
age 22
gender female
industry healthcare
astrological_sign Libra
text I love discussing science.
dtype: object
#problem 17: Find the most frequent industry among bloggers
most_frequent_industry = data['industry'].mode()[0]
# Display the result
most_frequent_industry
'student'
0s completed at 7:38 PM
36°C Mostly sunny
Search
ENG IN
19:39 27-04-2025


cc-42.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
#problem 18: Check for duplicate blog entries
duplicates = data.duplicated().sum()
# Display the result
duplicates
np.int64(0)
#problem 19: Calculate the median age of bloggers
median_age = data['age'].median()
# Display the result
median_age
40.5
#problem 20: Find bloggers who have the astrological sign "Aquarius"
aquarius_bloggers = data[data['astrological_sign'] == 'Aquarius']
# Display the result
aquarius_bloggers
age gender industry astrological_sign text
8 48 female technology Aquarius I love discussing travel.
0s completed at 7:38 PM
36°C Mostly sunny
Search
ENG IN
19:39 27-04-2025