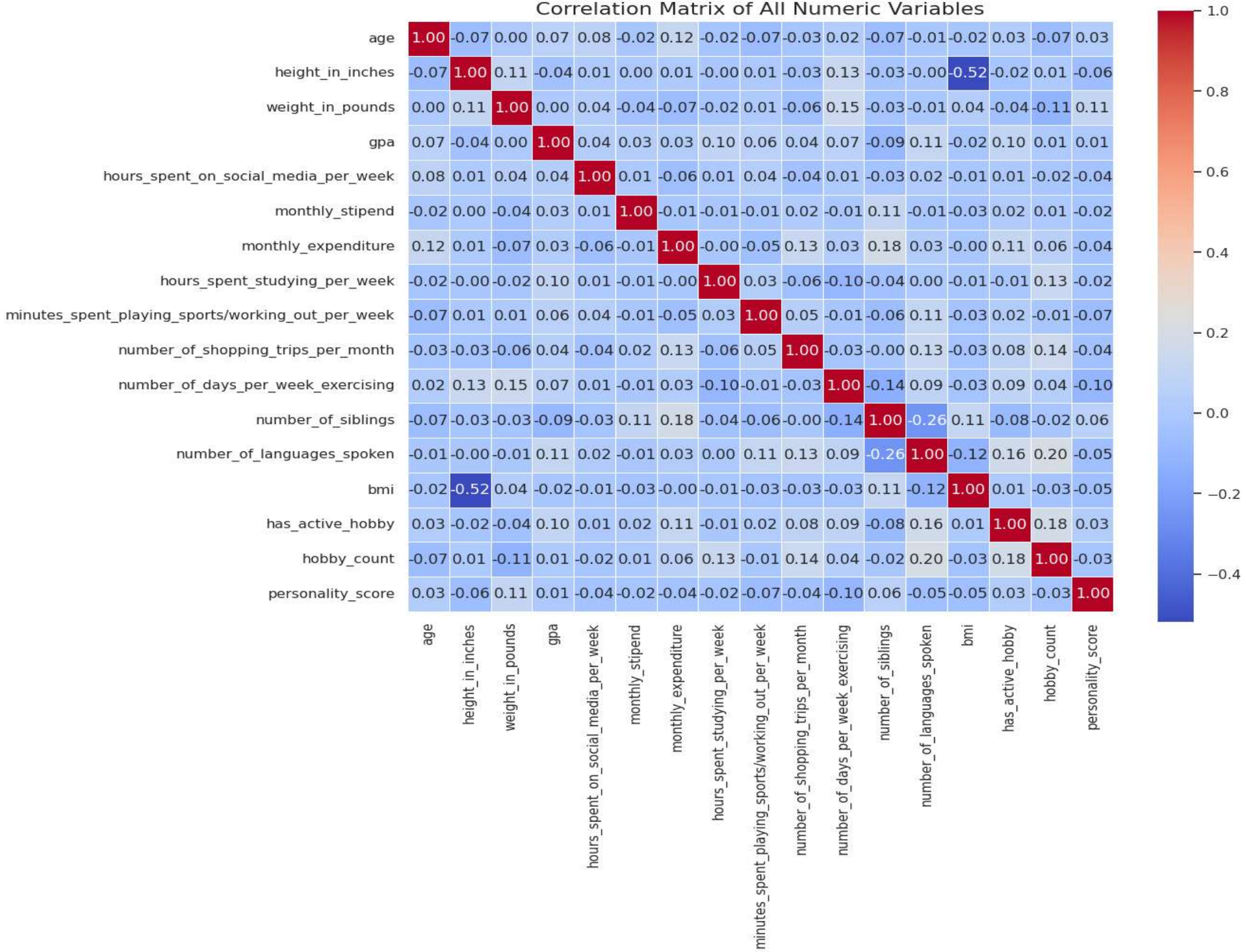
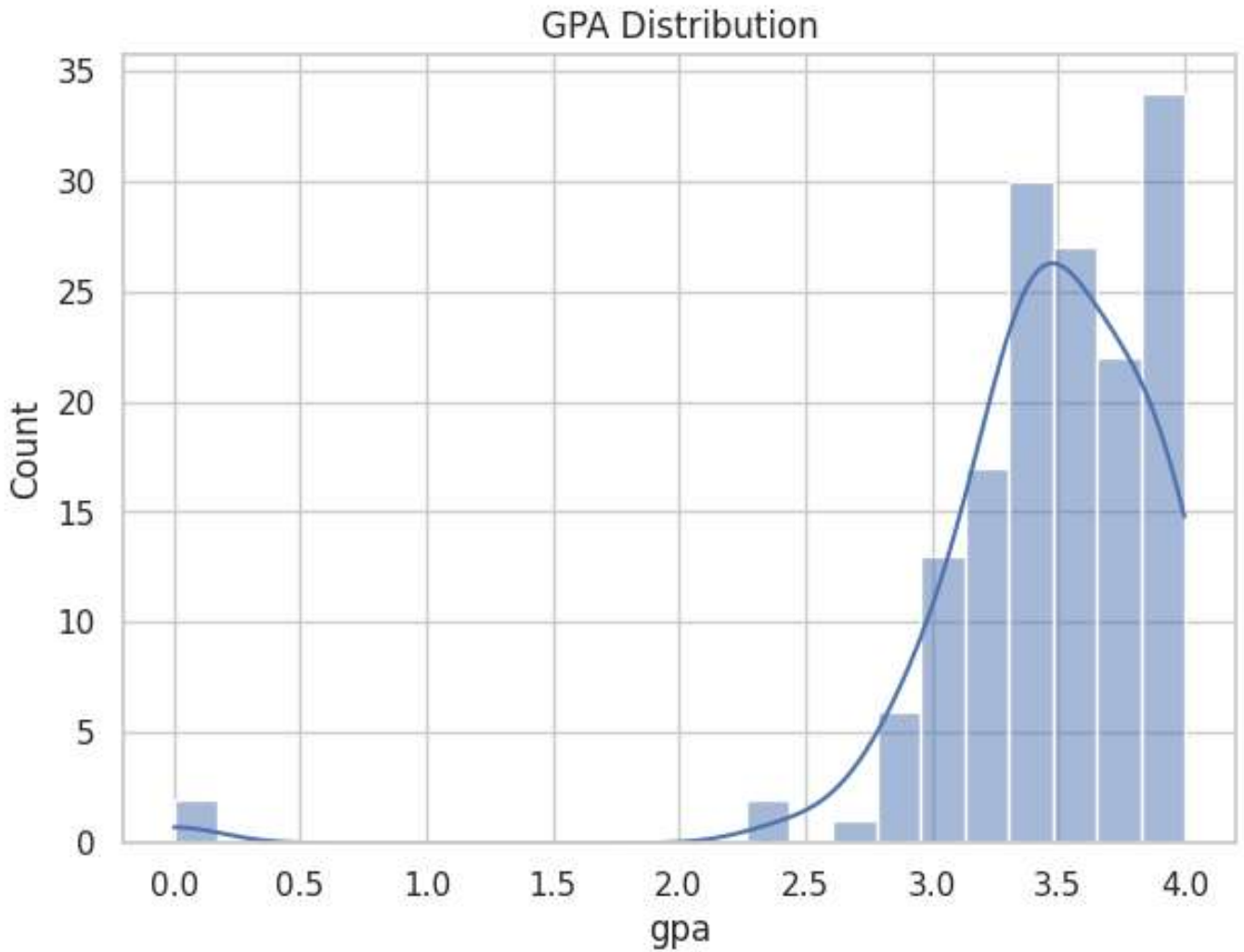
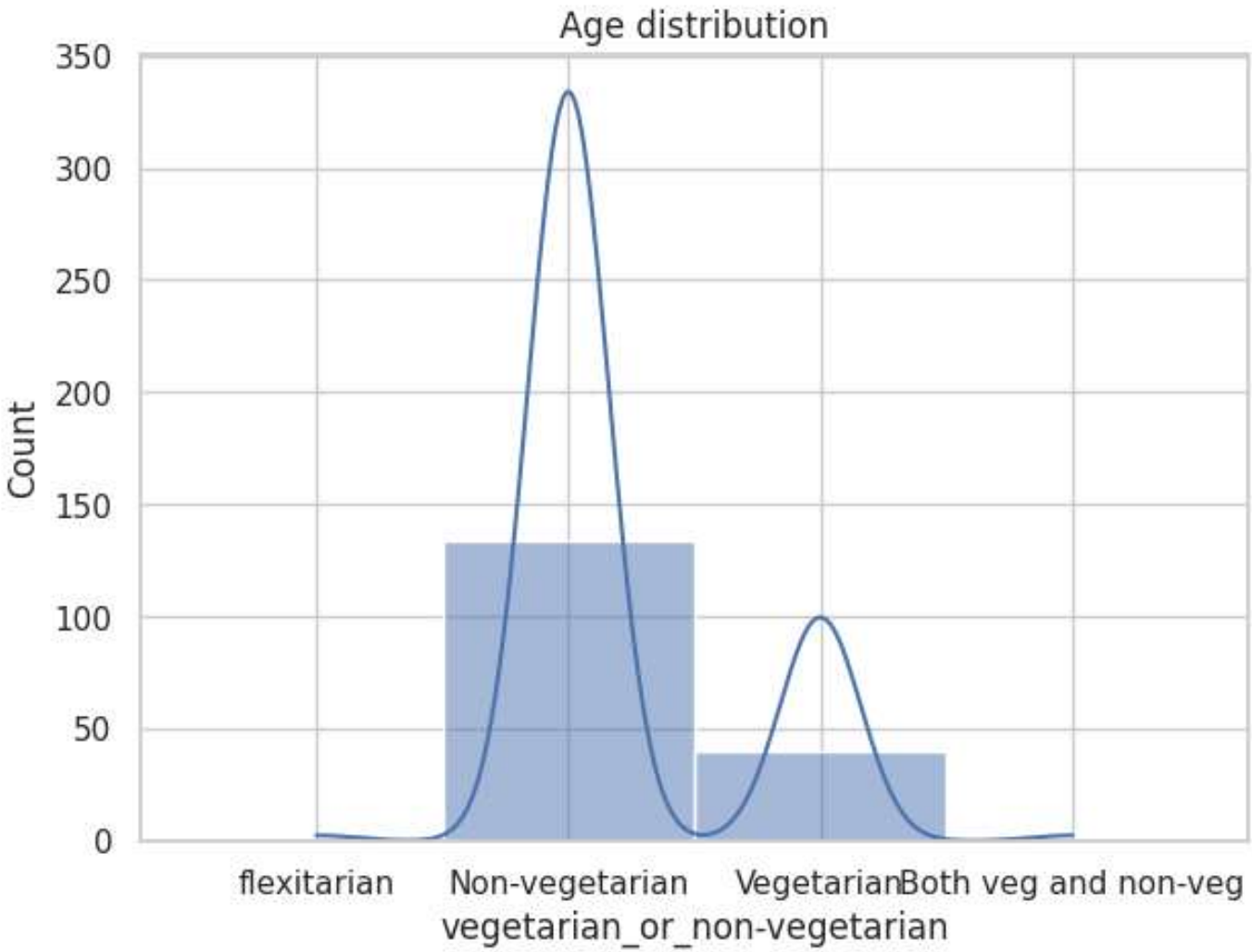
Data Science Project Presentation

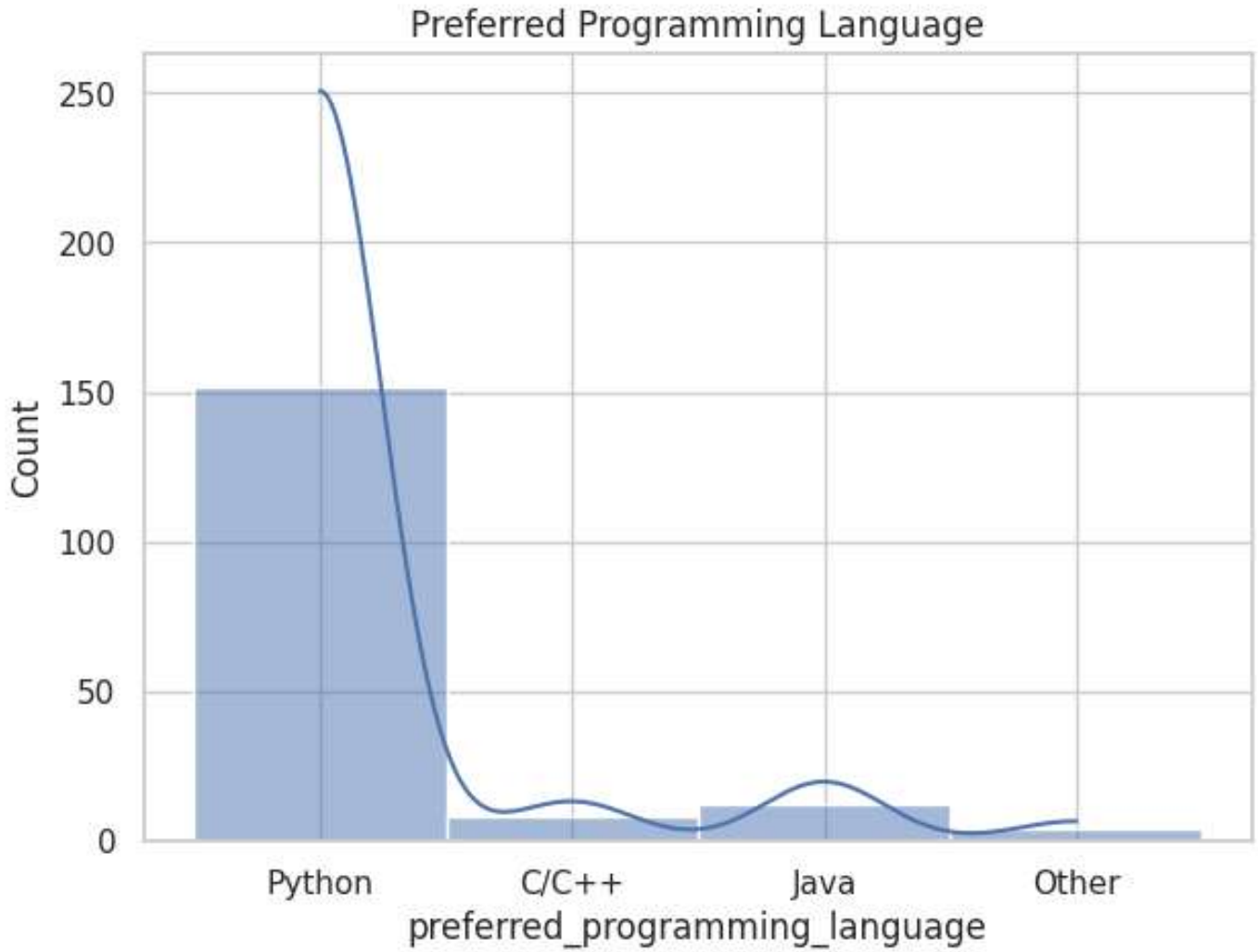
Using Extracted Graphs and Visuals

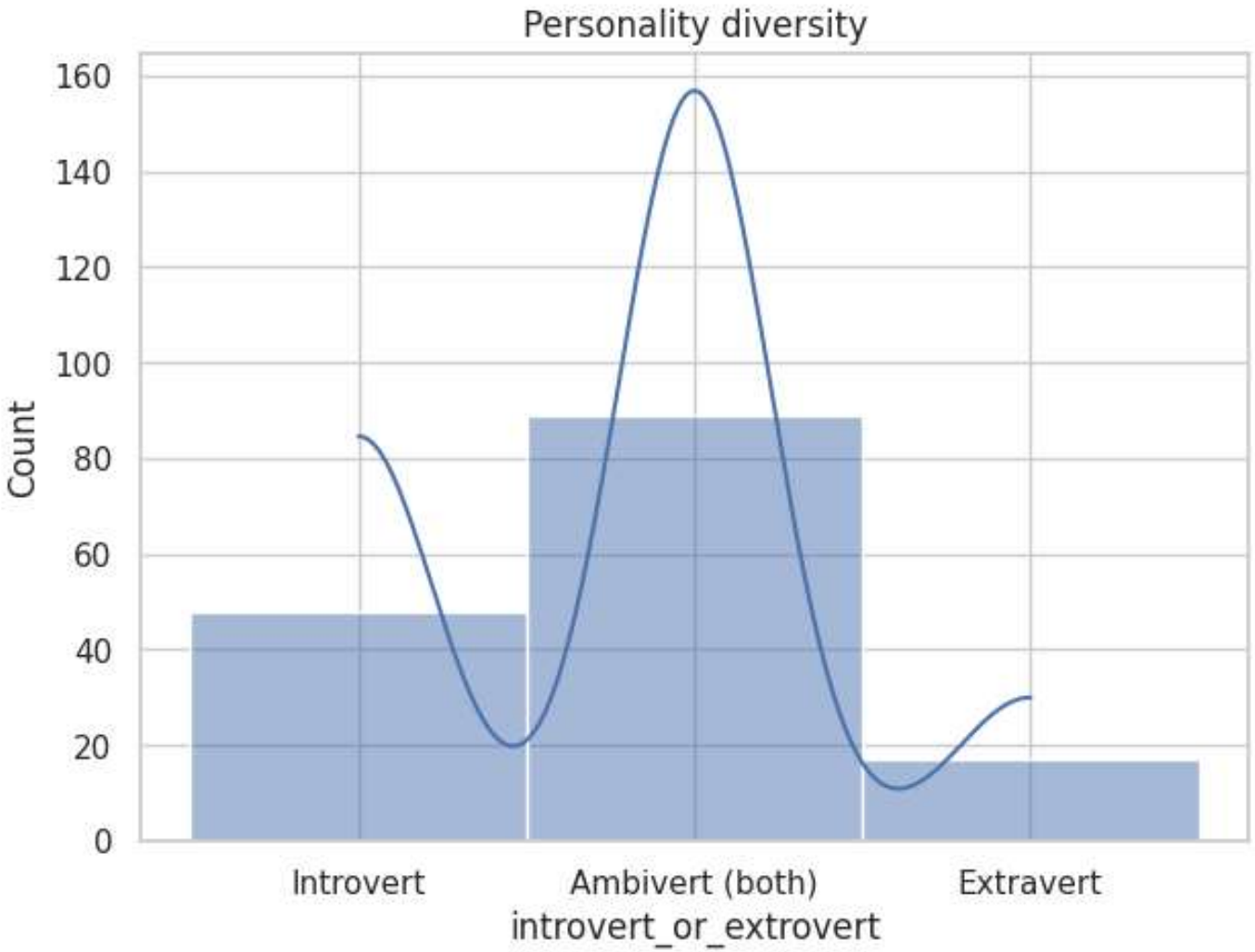

Correlation Matrix of All Numeric Variables
age 1.00 -0.07 0.00 0.07 0.08 -0.02 0.12 -0.02 -0.07 -0.03 0.02 -0.07 -0.01 -0.02 0.03 -0.07 0.03
height_in_inches -0.07 1.00 0.11 -0.04 0.01 0.00 0.01 -0.00 0.01 -0.03 0.13 -0.03 -0.00 -0.52 -0.02 0.01 -0.06
weight_in_pounds 0.00 0.11 1.00 0.00 0.04 -0.04 -0.07 -0.02 0.01 -0.06 0.15 -0.03 -0.01 0.04 -0.04 -0.11 0.11
gpa 0.07 -0.04 0.00 1.00 0.04 0.03 0.03 0.10 0.06 0.04 0.07 -0.09 0.11 -0.02 0.10 0.01 0.01
hours_spent_on_social_media_per_week 0.08 0.01 0.04 0.04 1.00 0.01 -0.06 0.01 0.04 -0.04 0.01 -0.03 0.02 -0.01 0.01 -0.02 -0.04
monthly_stipend -0.02 0.00 -0.04 0.03 0.01 1.00 -0.01 -0.01 -0.01 0.02 -0.01 0.11 -0.01 -0.03 0.02 0.01 -0.02
monthly_expenditure 0.12 0.01 -0.07 0.03 -0.06 -0.01 1.00 -0.00 -0.05 0.13 0.03 0.18 0.03 -0.00 0.11 0.06 -0.04
hours_spent_studying_per_week -0.02 -0.00 -0.02 0.10 0.01 -0.01 -0.00 1.00 0.03 -0.06 -0.10 -0.04 0.00 -0.01 -0.01 0.13 -0.02
minutes_spent_playing_sports/working_out_per_week -0.07 0.01 0.01 0.06 0.04 -0.01 -0.05 0.03 1.00 0.05 -0.01 -0.06 0.11 -0.03 0.02 -0.01 -0.07
number_of_shopping_trips_per_month -0.03 -0.03 -0.06 0.04 -0.04 0.02 0.13 -0.06 0.05 1.00 -0.03 -0.00 0.13 -0.03 0.08 0.14 -0.04
number_of_days_per_week_exercising 0.02 0.13 0.15 0.07 0.01 -0.01 0.03 -0.10 -0.01 -0.03 1.00 -0.14 0.09 -0.03 0.09 0.04 -0.10
number_of_siblings -0.07 -0.03 -0.03 -0.09 -0.03 0.11 0.18 -0.04 -0.06 -0.00 -0.14 1.00 -0.26 0.11 -0.08 -0.02 0.06
number_of_languages_spoken -0.01 -0.00 -0.01 0.11 0.02 -0.01 0.03 0.00 0.11 0.13 0.09 -0.26 1.00 -0.12 0.16 0.20 -0.05
bmi -0.02 -0.52 0.04 -0.02 -0.01 -0.03 -0.00 -0.01 -0.03 -0.03 -0.03 0.11 -0.12 1.00 0.01 -0.03 -0.05
has_active_hobby 0.03 -0.02 -0.04 0.10 0.01 0.02 0.11 -0.01 0.02 0.08 0.09 -0.08 0.16 0.01 1.00 0.18 0.03
hobby_count -0.07 0.01 -0.11 0.01 -0.02 0.01 0.06 0.13 -0.01 0.14 0.04 -0.02 0.20 -0.03 0.18 1.00 -0.03
personality_score 0.03 -0.06 0.11 0.01 -0.04 -0.02 -0.04 -0.02 -0.07 -0.04 -0.10 0.06 -0.05 -0.05 0.03 -0.03 1.00
age
height_in_inches
weight_in_pounds
gpa
hours_spent_on_social_media_per_week
monthly_stipend
monthly_expenditure
hours_spent_studying_per_week
minutes_spent_playing_sports/working_out_per_week
number_of_shopping_trips_per_month
number_of_days_per_week_exercising
number_of_siblings
number_of_languages_spoken
bmi
has_active_hobby
hobby_count
personality_score
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

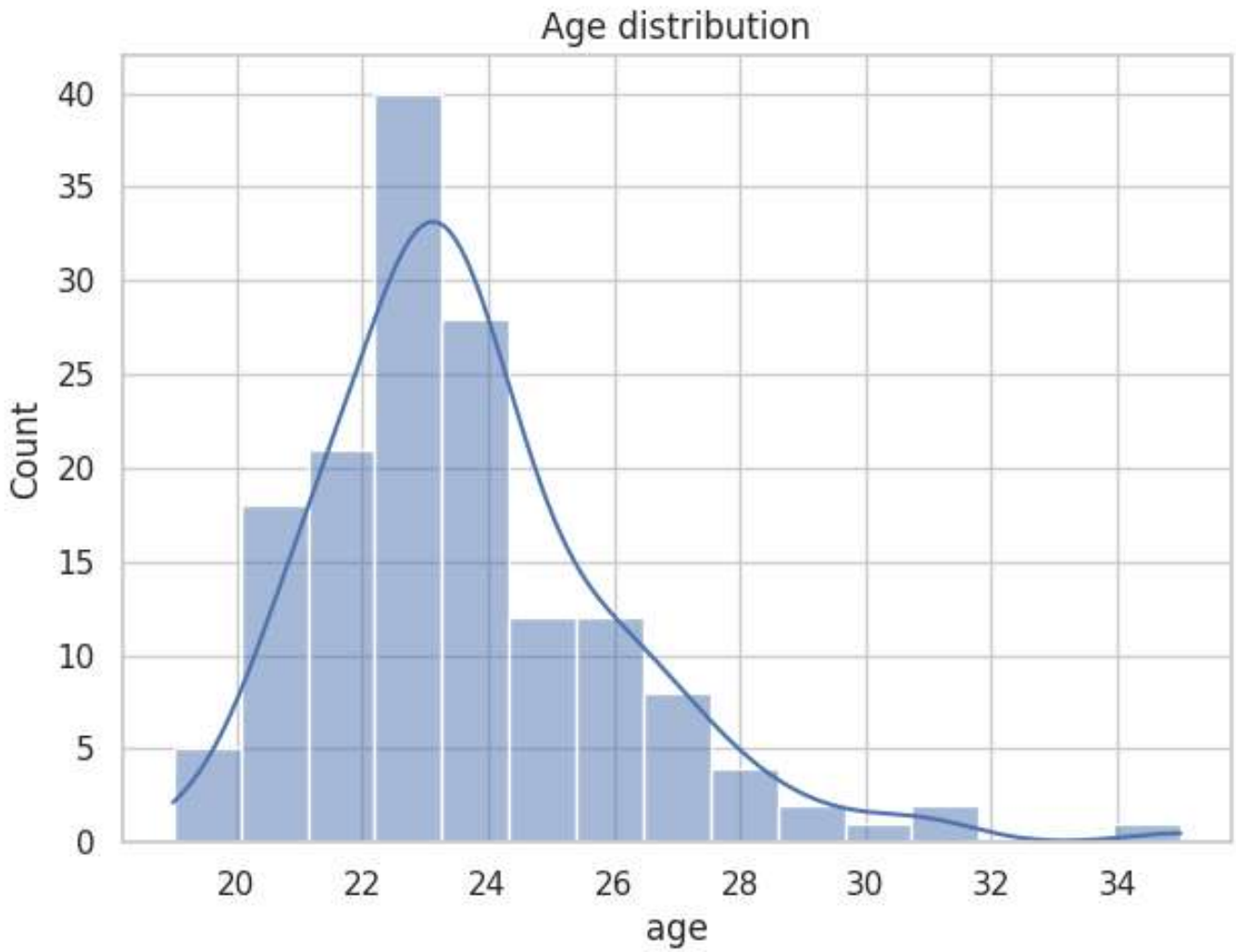

GPA Distribution
Count
35
30
25
20
15
10
5
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
gpa


Age distribution
350
300
250
200
150
100
50
0
Count
flexitarian
Non-vegetarian
Vegetarian
Both veg and non-veg
vegetarian_or_non-vegetarian


iOS vs Android
Count
250
200
150
100
50
0
Android
IOS
mobile_operating_system_android_or_ios

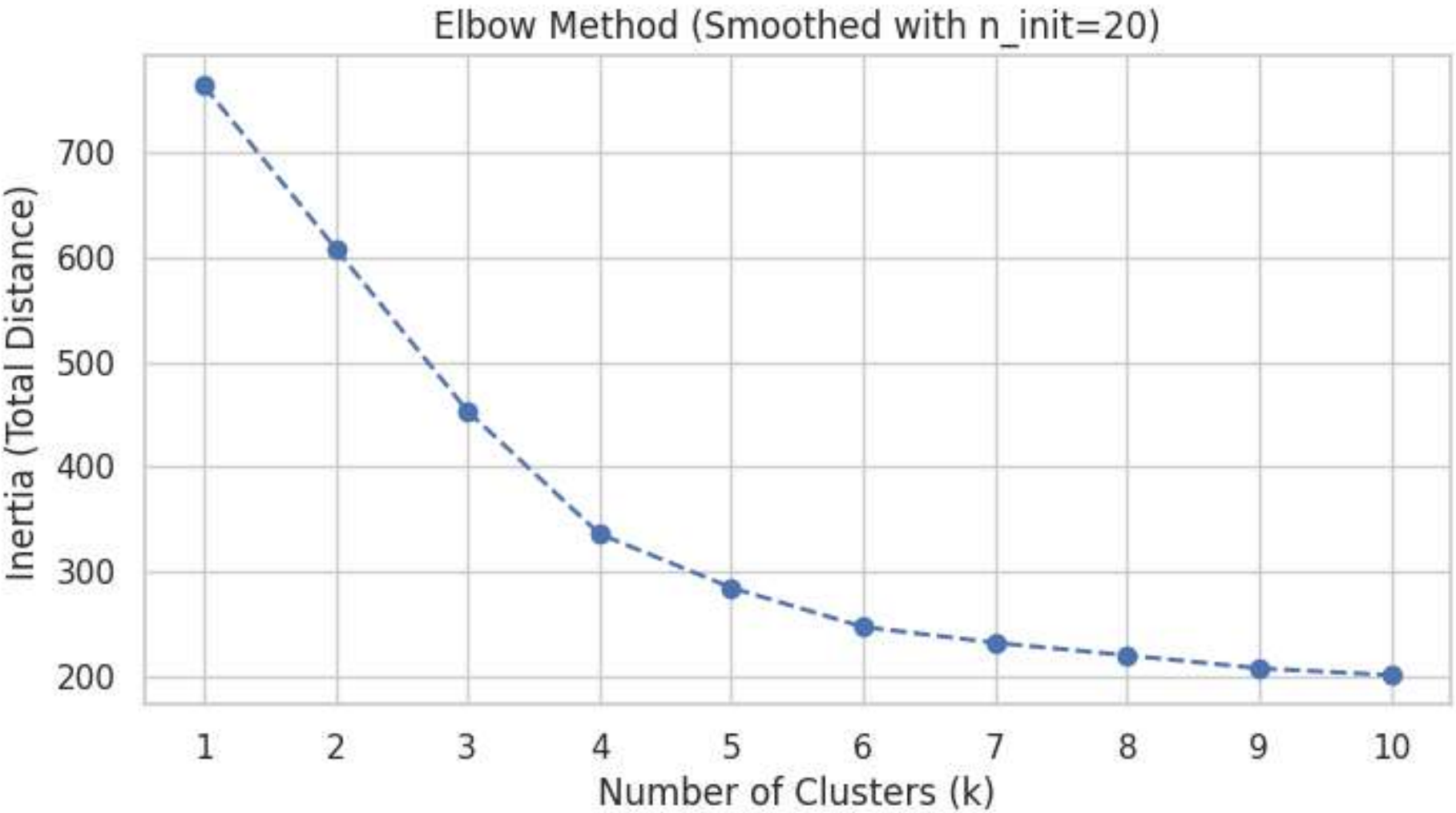

Preferred Programming Language
250
200
150
100
50
0
Count
Python
C/C++
Java
Other
preferred_programming_language

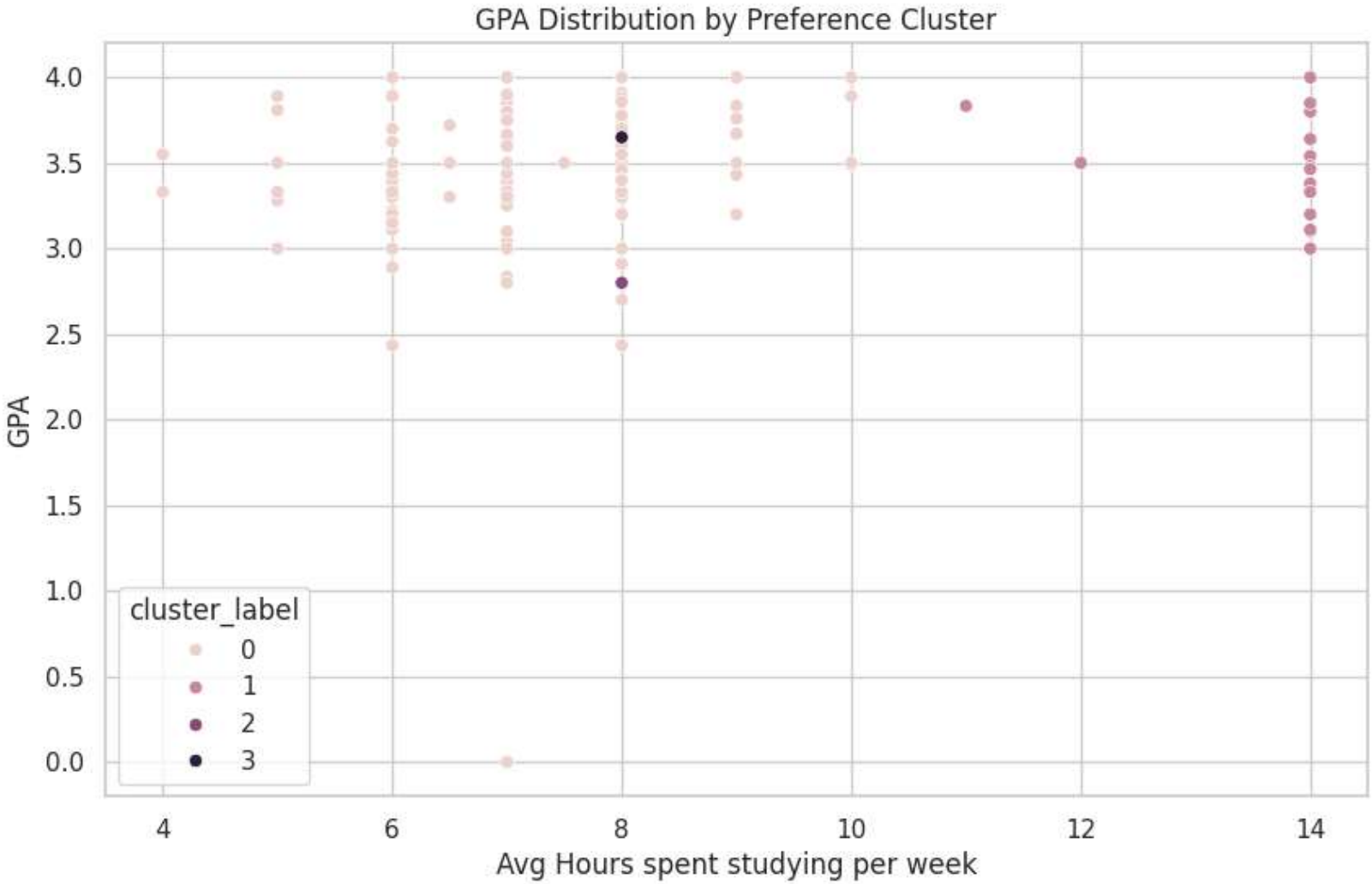

Personality diversity
160
140
120
100
80
60
40
20
0
Count
Introvert
Ambivert (both)
Extravert
introvert_or_extrovert

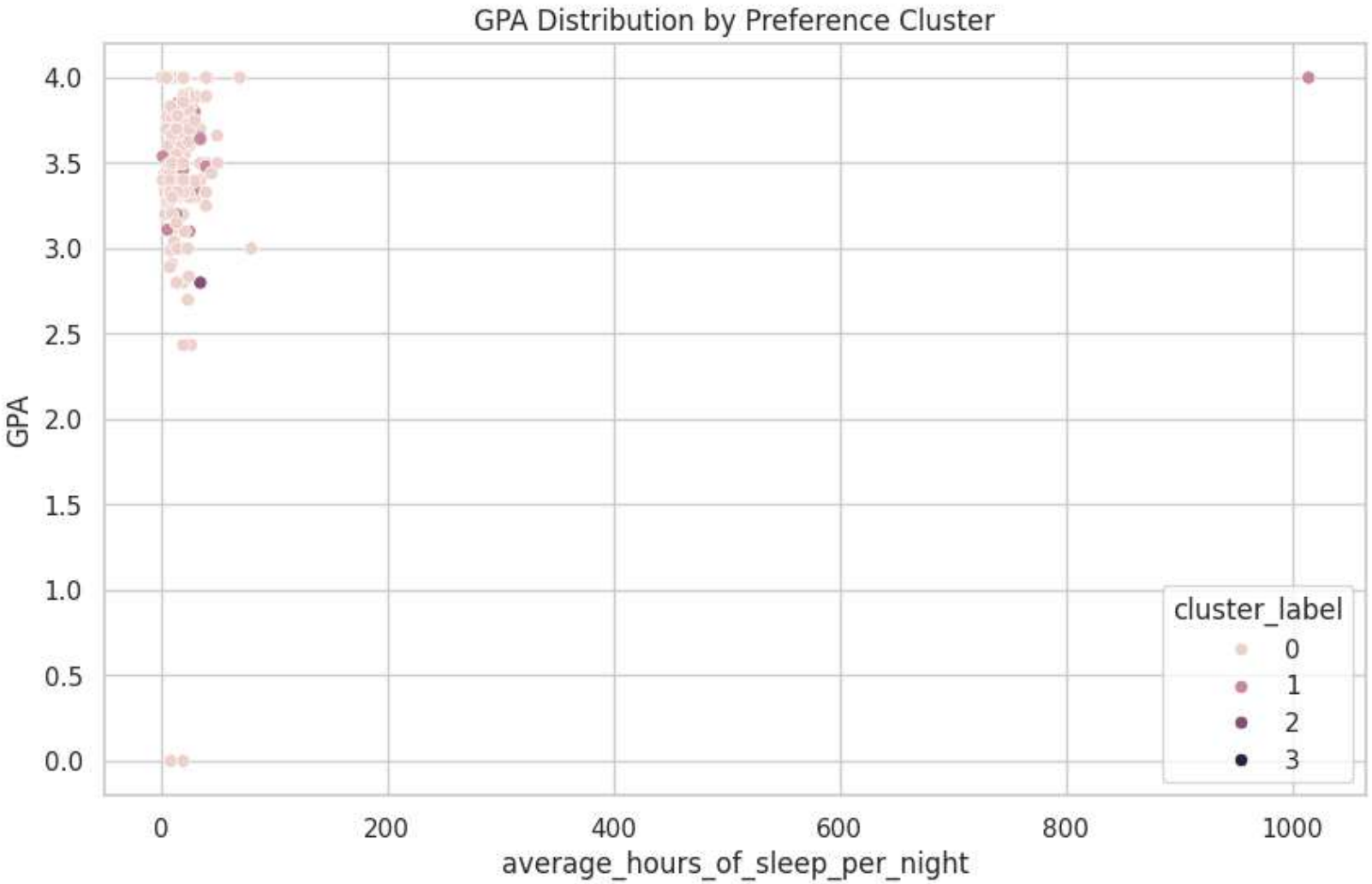

Age distribution
Count
40
35
30
25
20
15
10
5
0
20
22
24
26
28
30
32
34
age

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gpa   R-squared:                       0.010
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     1.574
Date:                Sun, 07 Dec 2025   Prob (F-statistic):              0.211
Time:                        02:40:45   Log-Likelihood:                -117.91
No. Observations:                 154   AIC:                             239.8
Df Residuals:                     152   BIC:                             245.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.8000      0.524      5.346      0.000       1.765       3.835
has_active_hobby     0.6593      0.525      1.255      0.211      -0.379       1.697
==============================================================================
Omnibus:                      161.027   Durbin-Watson:                   1.572
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3741.460
Skew:                          -3.843   Prob(JB):                         0.00
Kurtosis:                      25.891   Cond. No.                         24.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

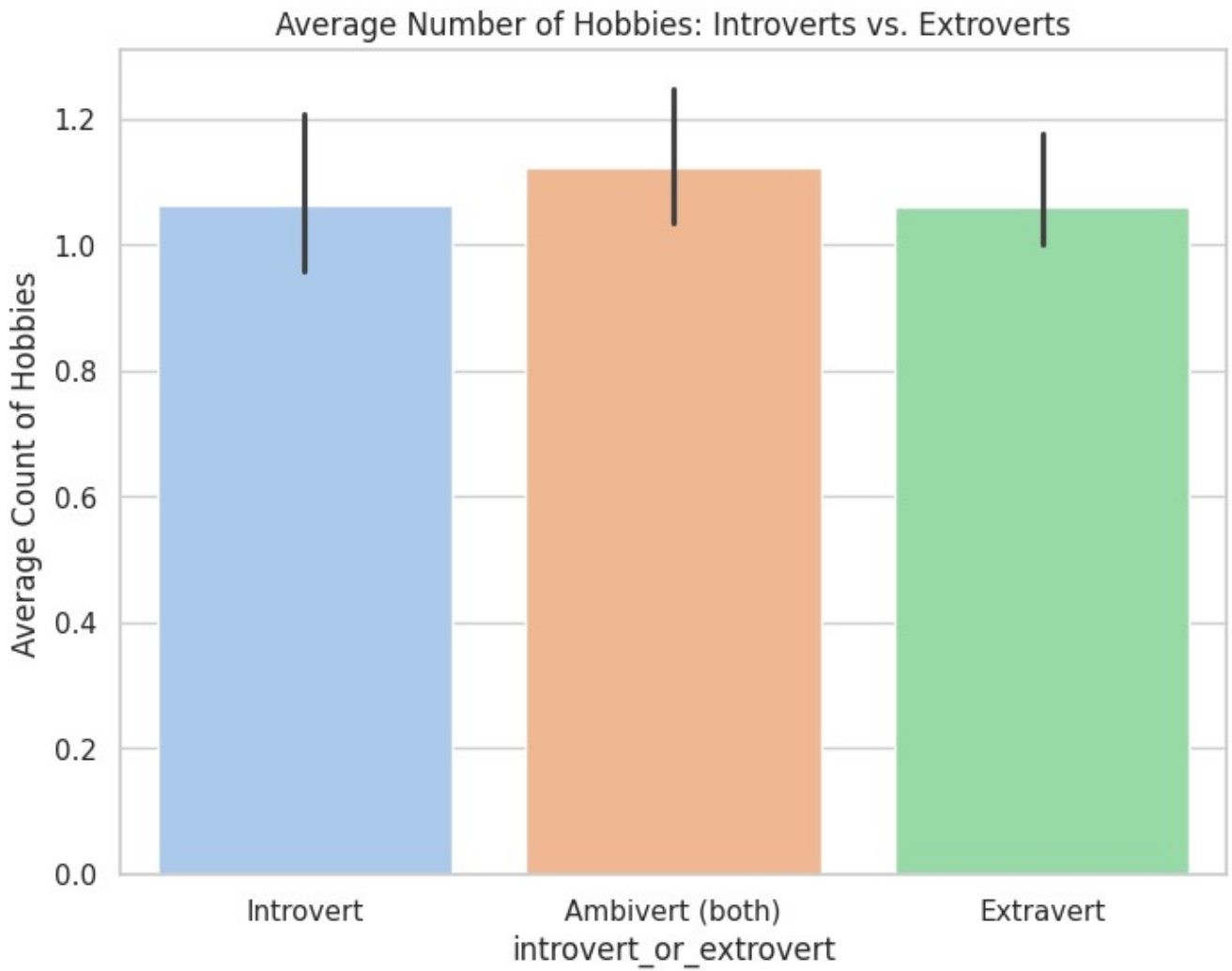
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gpa   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.007
Method:                 Least Squares   F-statistic:                     1.522
Date:                Sun, 07 Dec 2025   Prob (F-statistic):              0.222
Time:                        02:40:45   Log-Likelihood:                -117.17
No. Observations:                 154   AIC:                             240.3
Df Residuals:                     151   BIC:                             249.4
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const                        2.7779      0.523      5.309      0.000       1.744       3.812
hours_spent_studying_per_week     0.0006      0.001      1.211      0.228      -0.000       0.002
personality_score                 0          0        nan        nan           0           0
has_active_hobby             0.6654      0.525      1.268      0.207      -0.371       1.702
==============================================================================
Omnibus:                      162.114   Durbin-Watson:                   1.588
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3834.542
Skew:                          -3.876   Prob(JB):                         0.00
Kurtosis:                      26.184   Cond. No.                          inf
==============================================================================
```

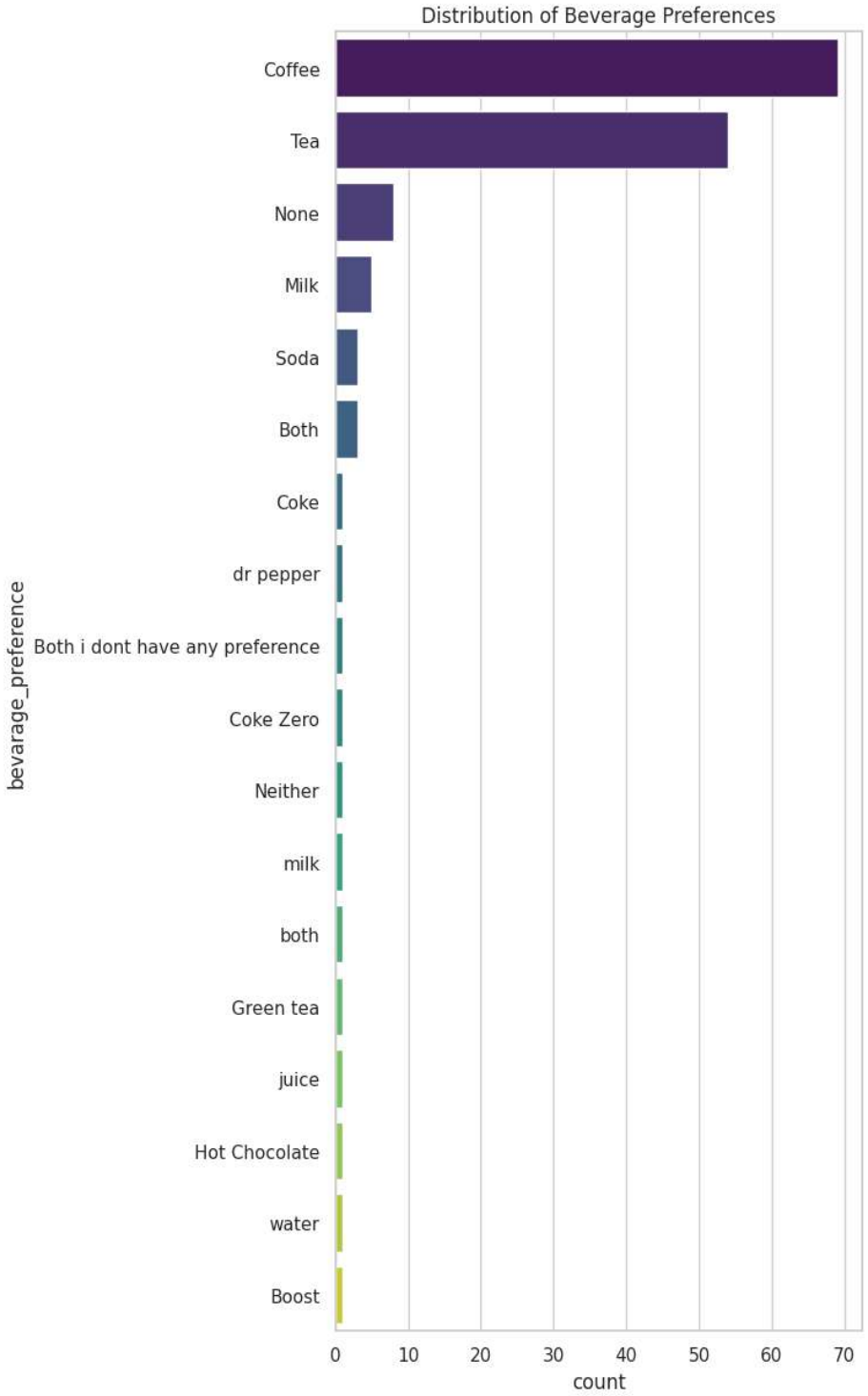

Elbow Method (Smoothed with n_init=20)
Inertia (Total Distance)
700
600
500
400
300
200
1
2
3
4
5
6
7
8
9
10
Number of Clusters (k)

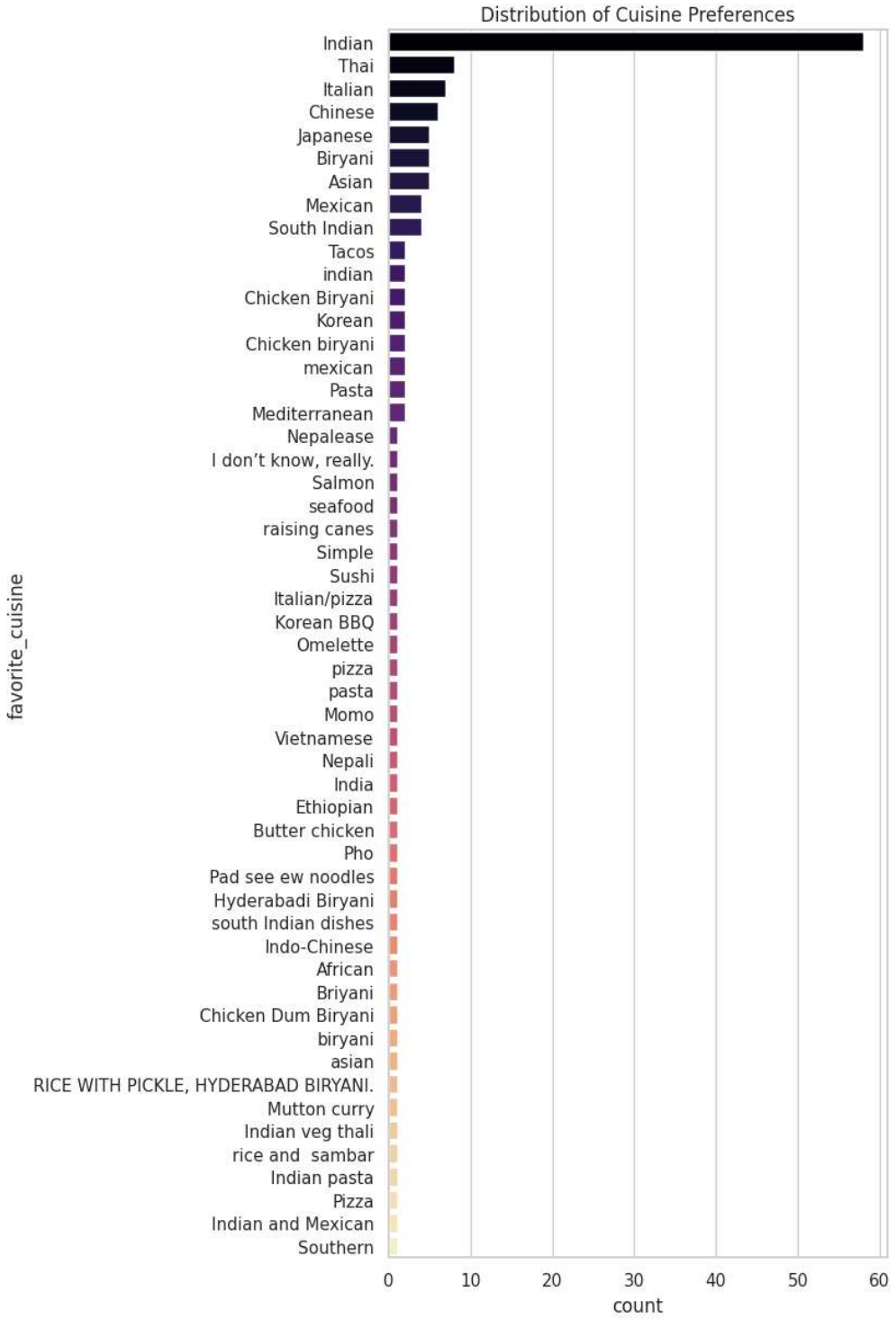

GPA Distribution by Preference Cluster
GPA
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
cluster_label
0
1
2
3
4
6
8
10
12
14
Avg Hours spent studying per week


GPA Distribution by Preference Cluster
GPA
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
200
400
600
800
1000
average_hours_of_sleep_per_night
cluster_label
0
1
2
3

```
Optimization terminated successfully.
         Current function value: 0.346891
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:           is_extrovert   No. Observations:                  154
Model:                          Logit   Df Residuals:                      152
Method:                           MLE   Df Model:                            1
Date:                Sun, 07 Dec 2025   Pseudo R-squ.:                0.001261
Time:                        02:40:47   Log-Likelihood:                -53.421
converged:                       True   LL-Null:                       -53.489
Covariance Type:            nonrobust   LLR p-value:                    0.7134
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -1.8459      0.747     -2.470      0.013      -3.310      -0.381
hobby_count   -0.2233      0.661     -0.338      0.735      -1.518       1.072
==============================================================================

Classification Report:

              precision    recall  f1-score   support

           0       0.89      1.00      0.94        42
           1       0.00      0.00      0.00         5

    accuracy                           0.89        47
   macro avg       0.45      0.50      0.47        47
weighted avg       0.80      0.89      0.84        47
```


Average Number of Hobbies: Introverts vs. Extroverts
Average Count of Hobbies
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Introvert
Ambivert (both)
Extravert
introvert_or_extrovert


Distribution of Beverage Preferences
bevarage_preference
Coffee
Tea
None
Milk
Soda
Both
Coke
dr pepper
Both i dont have any preference
Coke Zero
Neither
milk
both
Green tea
juice
Hot Chocolate
water
Boost
0
10
20
30
40
50
60
70
count


Distribution of Cuisine Preferences
favorite_cuisine
Indian
Thai
Italian
Chinese
Japanese
Biryani
Asian
Mexican
South Indian
Tacos
indian
Chicken Biryani
Korean
Chicken biryani
mexican
Pasta
Mediterranean
Nepalease
I don’t know, really.
Salmon
seafood
raising canes
Simple
Sushi
Italian/pizza
Korean BBQ
Omelette
pizza
pasta
Momo
Vietnamese
Nepali
India
Ethiopian
Butter chicken
Pho
Pad see ew noodles
Hyderabadi Biryani
south Indian dishes
Indo-Chinese
African
Briyani
Chicken Dum Biryani
biryani
asian
RICE WITH PICKLE, HYDERABAD BIRYANI.
Mutton curry
Indian veg thali
rice and sambar
Indian pasta
Pizza
Indian and Mexican
Southern
0
10
20
30
40
50
60
count