Python Program Output (Terminal)

```
(base) divitshetty@Divits-MBP-2 A4 % python3 A4.py
Question 1

Average testscore for all schools: 753.854
Average testscore for schools with a relatively large class size: 748.271
Average testscore for schools with a small class size: 757.608
Difference in averages (large-class minus small-class): -9.337

Question 1(b)

                            OLS Regression Results
==============================================================================
Dep. Variable:              testscore   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     2.893
Date:                Wed, 22 Nov 2023   Prob (F-statistic):             0.0896
Time:                        21:51:02   Log-Likelihood:                -2757.2
No. Observations:                 500   AIC:                             5518.
Df Residuals:                     498   BIC:                             5527.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    757.6077      3.481    217.650      0.000     750.769     764.447
NS            -9.3370      5.490     -1.701      0.090     -20.123       1.449
==============================================================================
Omnibus:                       13.296   Durbin-Watson:                   2.063
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               13.600
Skew:                           0.385   Prob(JB):                      0.00111
Kurtosis:                       3.242   Cond. No.                         2.45
==============================================================================

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

Estimated Slope: -9.337
Estimated Intercept:  757.608
The p-value,  0.090, is greater than 0.05, thus the estimated slope is not statistically significant
The 95% Confidence Interval is: [-20.123, 1.449]

Question 1(c)

S_Bay: t-value = 4.945, p-value = 0.000
NS_NonBay: t-value = 3.243, p-value = 0.001
NS_Bay: t-value = 2.881, p-value = 0.004
```

```
Question 2

Question 2(a)

                            OLS Regression Results
==============================================================================
Dep. Variable:               employed   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.019
Method:                 Least Squares   F-statistic:                     54.22
Date:                Wed, 22 Nov 2023   Prob (F-statistic):           4.83e-24
Time:                        21:51:02   Log-Likelihood:                -1628.7
No. Observations:                5412   AIC:                             3263.
Df Residuals:                    5409   BIC:                             3283.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.3075      0.055      5.619      0.000       0.200       0.415
age            0.0283      0.003     10.293      0.000       0.023       0.034
age2          -0.0003   3.28e-05     -9.971      0.000      -0.000      -0.000
==============================================================================
Omnibus:                     2193.252   Durbin-Watson:                   1.911
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6593.720
Skew:                          -2.215   Prob(JB):                         0.00
Kurtosis:                       6.100   Cond. No.                     2.68e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.68e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

Formal F-test:
F-statistic: 54.224
P-value: 0.000

Predicted Probability for 20-year-old worker: 0.74
Predicted Probability for 40-year-old worker: 0.92
Predicted Probability for 60-year-old worker: 0.83

Question 2(b)

                           Logit Regression Results
==============================================================================
Dep. Variable:               employed   No. Observations:                 5412
Model:                          Logit   Df Residuals:                     5409
Method:                          MLE   Df Model:                            2
Date:                Wed, 22 Nov 2023   Pseudo R-squ.:                0.02349
Time:                        21:51:02   Log-Likelihood:                -1986.4
converged:                       True   LL-Null:                       -2034.2
Covariance Type:            nonrobust   LLR p-value:                 1.775e-21
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -2.4898      0.437     -5.692      0.000      -3.347      -1.632
age            0.2255      0.023      9.885      0.000       0.181       0.270
age2          -0.0026      0.000     -9.518      0.000      -0.003      -0.002
==============================================================================

Likelihood Ratio Test:
Likelihood Ratio Test Statistic:  86.742
P-value:  0.000

Predicted Probability for 20-year-old worker: 0.73
Predicted Probability for 40-year-old worker: 0.91
Predicted Probability for 60-year-old worker: 0.83
```

Python Program Output(Text)

Question 1

Question 1(a)

- Average test score for all schools: **753.854**
- Average test score for schools with a relatively large class size(NS =1): **748.271**
- Average test score for schools with a small class size (NS =0): **757.608**
- Difference in averages (large-class minus small-class): **-9.337**

Question 1(b)

OLS Regression Results

Dep. Variable:	testscore	R-squared:	0.006
Model:	OLS	Adj. R-squared:	0.004
Method:	Least Squares	F-statistic:	2.893
Date:	Wed, 22 Nov 2023	Prob (F-statistic):	0.0896
Time:	18:43:55	Log-Likelihood:	-2757.2
No. Observations:	500	AIC:	5518.
Df Residuals:	498	BIC:	5527.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	757.6077	3.481	217.650	0.000	750.769	764.447
NS	-9.3370	5.490	-1.701	0.090	-20.123	1.449

Omnibus:	13.296	Durbin-Watson:	2.063
Prob(Omnibus):	0.001	Jarque-Bera (JB):	13.600
Skew:	0.385	Prob(JB):	0.00111
Kurtosis:	3.242	Cond. No.	2.45

==========

- Estimated Slope: **-9.337**
    - This is the same as the difference in averages that was calculated in 1(a)
- Estimated Intercept: **757.608**
    - This is the same as the avg test score for schools with a small class size from 1(a)
        - Makes sense because the intercept is the avg test score when NS is held constant and equals 0
- **Estimated slope is not statistically significant at the 5% level** because the p-value (0.090) is greater than 0.05.
- The 95% Confidence Interval is: **[-20.123, 1.449]**

Question 1(c)

- S_Bay: T-value = 4.945, P-value = 0.000
    - High t-value and low p- value. Shows that the effect of small class size (represented by S_Bay) is statistically significant for schools in the Bay Area
- NS_NonBay: T-value = 3.243, P-value = 0.001
    - Relatively high t-value, and low p-value indicates that the effect of large class size (represented by NS_NonBay) is statistically significant for schools not in the Bay Area
- NS_Bay: T-value = 2.881, P-value = 0.004
    - Still a relatively high t-value, and low p-value indicates that the effect of large class size (represented by NS_NonBay) is statistically significant for schools in the Bay Area
- **The t-stats and p-values suggest that the effect of (large) class size has a significant effect for both Bay Area and non-Bay Area schools.**

Question 2

Question 2(a)

OLS Regression Results

==========

Dep. Variable:	employed	R-squared:	0.020
Model:	OLS	Adj. R-squared:	0.019
Method:	Least Squares	F-statistic:	54.22
Date:	Wed, 22 Nov 2023	Prob (F-statistic):	4.83e-24
Time:	21:18:24	Log-Likelihood:	-1628.7
No. Observations:	5412	AIC:	3263.

Df Residuals: 5409 BIC: 3283.
Df Model: 2
Covariance Type: nonrobust

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.3075	0.055	5.619	0.000	0.200	0.415
age	0.0283	0.003	10.293	0.000	0.023	0.034
age2	-0.0003	3.28e-05	-9.971	0.000	-0.000	-0.000

Omnibus: 2193.252 Durbin-Watson: 1.911
Prob(Omnibus): 0.000 Jarque-Bera (JB): 6593.720
Skew: -2.215 Prob(JB): 0.00
Kurtosis: 6.100 Cond. No. 2.68e+04

- The p-value associated with age is 0.000
    - P < 0.05 (significance level of 5%)
        - Reject the null hypothesis
            - **Therefore, age was a statistically significant determinant of employment in April 2009.**
- Formal F-test:
    - **F-statistic: 54.224**
        - F > 2.5 so we can reject the null hypothesis.
    - **P-value: 0.000**
    - The p-value = 0.000, which is less than sig level of 0.05.
        - Therefore, reject the null hypothesis.
            - Provides evidence to suggest that there is a statistically significant effect of age (both ‘age’ and ‘age^2’) on the probability of being employed.
    - **Based on the p-value of the formal F-test, there is evidence of a nonlinear effect of age on the probability of being employed.**
- Predicted Probability for:
    - 20-year-old worker: **0.74**
    - 40-year-old worker: **0.92**
    - 60-year-old worker: **0.83**

Question 2(b)

Logit Regression Results

==========

Dep. Variable:	employed	No. Observations:	5412
Model:	Logit	Df Residuals:	5409
Method:	MLE	Df Model:	2
Date:	Wed, 22 Nov 2023	Pseudo R-squ.:	0.02349
Time:	21:51:02	Log-Likelihood:	-1986.4
converged:	True	LL-Null:	-2034.2
Covariance Type:	nonrobust	LLR p-value:	1.775e-21

==========

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.4898	0.437	-5.692	0.000	-3.347	-1.632
age	0.2255	0.023	9.885	0.000	0.181	0.270
age2	-0.0026	0.000	-9.518	0.000	-0.003	-0.002

==========

- Once again, the p-value associated with age is 0.000
    - $P < 0.05$ (significance level of 5%)
        - Reject the null hypothesis
            - **Therefore, age was a statistically significant determinant of employment in April 2009.**
- Likelihood Ratio Test:
    - **Likelihood Ratio Test Statistic: 86.742**
    - **P-value: 0.000**
    - The p-value = 0.000, which is less than sig level of 0.05.
        - Therefore, reject the null hypothesis.
            - Provides evidence to suggest that there is a statistically significant effect of age (both ‘age’ and ‘age^2’) on the probability of being employed.
    - **Once again, because on the small p-value, there is evidence of a nonlinear effect of age on the probability of being employed.**
- Predicted Probability for:
    - 20-year-old worker: **0.73**
    - 40-year-old worker: **0.91**
    - 60-year-old worker: **0.83**