

**ANL252**

**Python for Data Analytics**

# **End-of-Course Assessment**

**July 2021 Presentation**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| **Lee Min Kang Bryan** | **E1771573** |

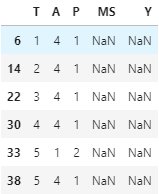
**Tutorial Group: ­­­­­­­­­­­­ T 09**

**Instructor’s Name: Mr. Munish Kumar**

**Submission Date: 13/09/2021**

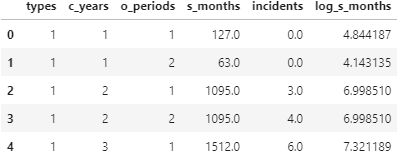
**1ai)**

****



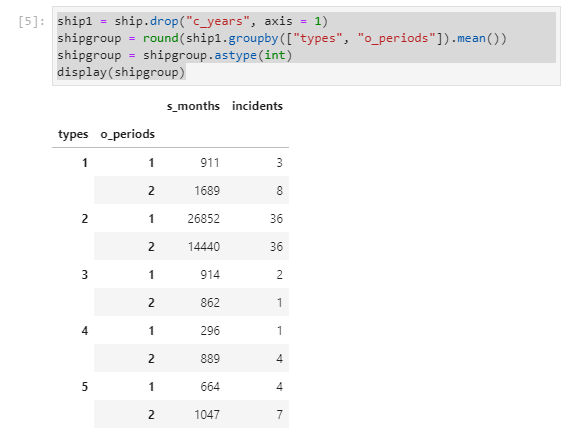
**1aii)**

****



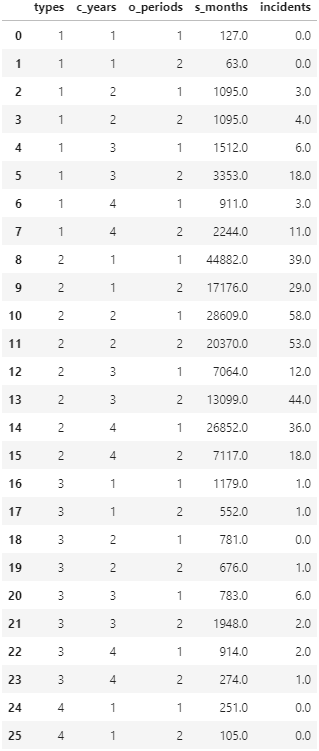
**1aiii)**

****



**1aiv)**

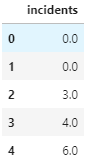
****





**1av)**

****



**1bi)**

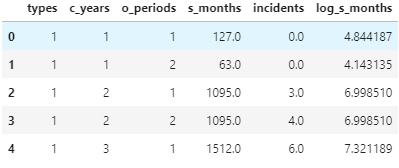
****

**1bii)**

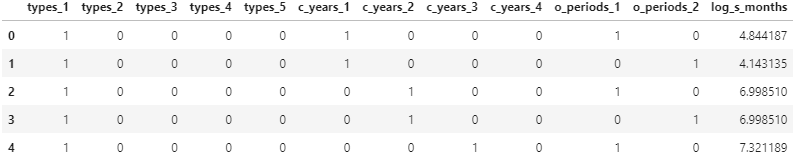
****

**1biii)**

****



X.head()



**1c)** It is not sensible to split the DataFrame into training and testing dataset because with only 40 rows, the DataFrame is considered small and would not be ideal for splitting. There would be insufficient data in the training dataset for the model to learn from and insufficient data in the testing dataset to evaluate the models’ performance. For smaller datasets, a more suitable alternative model evaluation procedure would be the k-fold cross-validation as opposed to the train-test split.

**1d)**

****

**2a)** The Poisson Regression model is a type of Generalised Linear Model (GLM) in which the target variable is a count. The Poisson Regression model is used to predict/estimate a target variable with a Poisson distribution based on multiple independent variables. In python, the Poisson Regression model has a few parameters which can be set by the user to build the model. The parameters and their explanation are as per the table below.

|  |  |
| --- | --- |
| **Parameters** | **Explanation** |
| alpha | To specify if a penalty should be added to the estimator. |
| fit\_intercept | To specify whether to add a bias/intercept to the estimator |
| max\_iter | To set the maximum number of repetitions |
| tol | To set the criteria for stopping the repetitions |
| warm\_start | To reuse the previous fit result as a configuration for coef\_ and intercept\_ |
| verbose | To determine the level of detailed logging output |

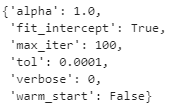
Python uses the fit and predict functions to construct the Poisson Regression model. The fit function is used to fit the Poisson Regression model on a DataFrame. The two parameters of the fit function are X and Y where X contains the independent variables while Y is the target variable. The predict function is used to call out the predicted values of Y based on X. Having said that, the parameter for the predict function is X, the matrix of the independent variables. Performing the predict function would return the predicted values of Y for each row of data. If there are 40 rows of data, there will be 40 predicted values of Y.

**2b)**

Fit the Poisson regression model and get the parameters

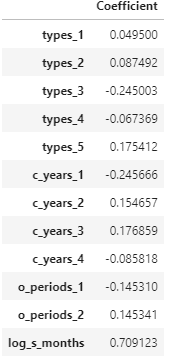
****

The parameters of the estimated model:

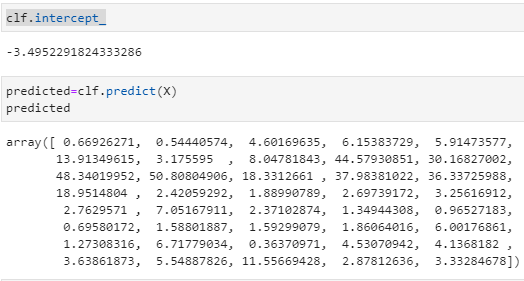


Generate coefficient and present it in a table

****

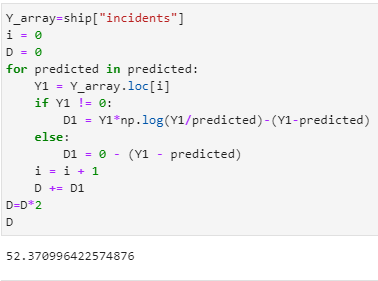


The intercept and predicted values



**2c)**

****



**Appendix**



Full screenshot of Code:

