|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Project Activity Log | | | | | | | |
| Learner Name | | Gleb Sokolovskyi | |  | Learner number | 8279 |  |
|  | |  | | |  | |  |
| Centre Name | | Fettes College | |  | Centre Number | 60772 |  |
|  | |  | | |  | |  |
| Unit Name | | Artefact | |  | Unit number | P304 |  |
|  | |  | | |  | |  |
| Teacher Assessor | | Jason Pitt | |  |  | |  |
|  | |  | | |  | |  |
| Proposed project title | | | **How accurately can machine learning predict future sales for a small coffee shop in Edinburgh?** | | | |  |
| Prior to the dates below:   * Identified Different Factors for Possible Effects of sales data * Looked at possibility of doing a rule-based system; Looked at possibility of doing a neural network; Compared the 2 systems – decided to do neural network * Read a book about AI = “The Essence of Artificial Intelligence” by Alison Cawsey * Made a presentation and Draft write up * Decided that the program is going to be done in python | | | | | | | |
| Date | Comments | | | | | | |
| Tue 19/02/19 | * Drew Neural Network for all different Factors and possible outputs * Looked up all the dates for holidays and Fringe and any other dates that were needed for the project | | | | | | |
| Wed 20/02/19 – Sat 23/02/19 | Performing calculations to figure out weights for each input neuron | | | | | | |
| Mon 25/02/19 – Sun 03/03/19 | * Working on write up * Practisiing using the PESC model (Political, Economic, Social, Cultural) | | | | | | |
| Mon 04/03/19  Fri 08/03/19 – Wed 13/03/19 | Working on the “Abstract” of the write-up  Writing the “Writing the Code” section of the write up | | | | | | |
| Fri 22/03/19 – Sun 31/03/19 | I have figured out that I need an economic question for my epq. Some | | | | | | |

|  |  |
| --- | --- |
|  |  |
| 29/04/19 – 05/05/19 | Saving all the excel sheets as a CSV file and transferring to python environment. Didn’t know how to do it so had to search it up. Website used: <https://docs.python.org/2/library/csv.html> |
| 01/07/19 – 11/07/19 | Learnt all about reading data from files. They key is to transfer all the data from excel to a text file. Then use “open()” and “readlines()” functions, specifying the delimitor – which is “,” in my case – and then adding every value to a list. |
| 12/07/19 | Learning about regression lines. Importing “statistics”, “numpy” and “matplotlib.pyplot” to find line of best fit, y-intercept and plot the graph. Use points as (amount sold, temp). |
| 13/07/19  13/07/19 – 22/07/19 | Got all the week average temperature for 2015   * Was solving the problem of trying to get all products units sold as x-coordinates and all temp values for that each week as y-coordinates * Found all rugby matches at <https://www.scottishrugby.org/fixtures-and-results/> * Had to stop plotting axis as a bigger issue accred |
| 23/07/19 – 29/07/19 | Spoke to an economics student – he explained to me that to add accuracy to the program, it should include Price, people’s income in Ediburgh and multiple linear regression. Have done:   * A program which calculates consumption per week for each week for each product and asks the user which product's sales data they would like to see and it outputs how much of that product was sold each week and plots a graph of that data (with date on the x-axis – in chronological order), which helps to spot patterns in data. This was mainly created to see whether having the same price over a year’s time would affect sales in positive manner, however this can be used for any other factor now too. |
| 31/07/19 | * I’ve realized that I’ve been using wrong temperature data. I was doing calculations with average weekly temperature, which takes into account night temperature, which are irrelevant for my calculations, as the café wasn’t working at nights. Therefore I’ve changed my data in the Factors.csv * Also found a way to make my code shorter for both Plotting\_PED.py, Finding\_Pearson\_Correlation.py and any other programs which include extracting info from sales data files by creating a new python file called “Extracting\_Info.py”, where I only extract it once and in any file which needs the variables from Extracting\_Info.py, I can do one simple line of code: “from Extracting\_Info import …”. |
| 01/08/19 | * Realized I made another mistake in Finding\_Pearson\_Correlation.py because I was looking at the r value between sales and time but sales depends on season so it won’t just have strong positive correlation |
| 02/08/19 – 18/08/19 | * Calculating by how much Fringe affects sales (Finding\_Fringe\_per.py) * Working on write up * Model\_table.py – creates a table of all independent variables and their coefficient, standard deviation, t value and lower and upper bounds of 95% confidence interval with the dependant value (sales) |
| 19/08/19 – 03/09/19 | * Working on write up * Realized ANOTHER mistake: Model\_table.py calculates relationship between independent variables and sales, while independent variables overlap. Eg. When it was outputting coefficient between Fringe and sales, it was actually outputting coefficient between (Fringe and holidays) and sales, so it was not accurate. Rookie mistookie. * So…to recover from such failure I wrote Finding\_Fringe\_Matches\_Holidays\_Percentages.py which calculates coefficients accurately this time AND outputs the final Multiple Linear Regression equation for each product. |
| 04/09/19  05/09/2019 | * Finished “Testing.py” * Handed first final draft |
| 15/09/2019 – 05/10/2019 | * Feedback came back. Mainly working on rephrasing everything in the write up to make sure it has proper meaning. The structure is also pretty bad – quite confusing to follow, so have to change that as well. * Bibliography also lacking |
| 10/10/2019 | * Changed a bit in “Testing.py”. Realised the variable   “testing\_results” can be negative so had to change to: |
| 11/10/2019 – 01/11/2019 | * Working on wording of the write up * Adding comments to my code * Changed the layout of the write up 🡨 had a lot of bullet points – now more sentences and paragraphs |

