PROG1000 – Introduction to Programming

Semester assignment 2025

Handed out: 26 March 2025 at Canvas

Handed in: 16 April 2025, 23:59 in Canvas.

(Last update: 26 March 2025)

You will produce two outputs:

- **Report**: A written report of 1500-2000 words, containing analysis and graphs. Upload this report as a .pdf file.
- **Code:** You have to upload the file with the code behind your analysis. This will either be a file created by Jupyter notebook (.ipynb) or a file created by Spyder (.py).
- Please write the group <u>number</u> and the <u>names</u> of all of the group members on the first page of the report

You work as an analyst at Murloc Inc. when your boss, John Snow, asks you to analyse the effect of product pricing vs. product demand (sold units). This willl help Murloc Inc. to set the optimal price for their product, thereby increasing the profitability.

However, the Chief Security Officer, Cercei Lannister, is preoccupied with the safety of business secrets and would like you to make a password generator to protect the running of the analysis code.

1)
    a) The file with the password generator shall be named "password.ipynb" or "password.py".

    b) The user must input a number between 15 and 20 which will decide the number of characters in the password. The password shall contain random characters from letters (a-z, and A-Z), and special characters (! @ # - _ . , : ; ( ) = ? / \ ).

    c) A randomly generated password shall be generated and stored for later use.

The data base administrator, Lyanna Mormont, has downloaded two data sets for you from the data base. These might help you reach some conclusions about the products:

- The file "orders.csv" contains the following variables:
    o Month
    o Product_id
    o Units_sold
    o Sales_price_per_unit
- The Excel file "products.xlsx" contains the following variables:
    o Product_id
    o Product_name
    o Product_cost_per_unit

2)
   a) Import the file "orders.csv".
   b) Import the file "products.xlsx"
   c) Merge the two data sets. Use "product_id" as a common column.
   d) Drop any NaNs.
   e) Create a new variable called "order_cost", which is the product amount multiplied with product cost.

3) Explore the data set.
   a) Present descriptive statistics for the data set.
   b) Create a figure with four sub-plots in two rows and two columns, one for each product. "price" should be on the x-axis and quantity should be on the y-axis. Add the correlation coefficients in the subplot titles.

4)
   a) Create a bar plot to show the revenue (sum of sales) for each product.
   b) Create a plot with four subplots, one for each of the products. The subplots should be a histogram of the sales price for each product.

5)

a) Estimate the following regression model for each of the products:
$$quantity = \alpha + \beta_1 \times price + \epsilon$$
How much of the variation in quantity does the model explain for each product?

b) If the products are priced by their average price. How much of the four products would Murloc Inc. sell, on average?

c) If the product prices were increased by 10% from the average prices, how much of each product would Murloc Inc. sell?

6)

a) Create a function which takes in two arguments (y) and (x), where y is the dependent variable and x is the explanatory variables. The function should run a regression and return the models' adjusted R^2. This function shall only work if the securely generated password from task 1 is entered. If a wrong password is entered, the user shall be notified and may try again.

b) Try the function by passing in quantities and prices for one of the products.

Write a report where you:

Present your investigation and analysis of the four products and their market demand. The report must be between 1500 and 2000 words, and it should provide answers to all the tasks above. Beside the defined tasks, you are of course free to include what you have found out from the data sets that is not a part of the task descriptions.

The pdf report should not include any code. Instead, you should explain and summarize your results in a way that the reader is able to understand the report without any knowledge of Python or your Jupyter notebook.

Remember to not just present the results from your Jupyter notebook, the report should also explain those results. For example, it is not sufficient to simply include graphs, but you must also explain them (e.g. this figure shows..., in this figure we have plotted... etc.).