1. Suppose that you are the marketing director of a retail company. Your company creates, prints, and distributes weekly price promotion leaflets to attract new and existing shoppers. To encourage people to read the leaflets, they contain a crossword puzzle every week. There is a box in every store where the solution can be placed, and you reward 30 randomly selected respondents with a HUF 15,000 voucher. On average, 9,320 people solve the crossword and place the solution to the box, with a standard deviation of 1,795.

You would like to increase the attractiveness of the crossword puzzle and decide to provide a HUF 20,000 voucher to the selected 30 people. The first 60 weeks of data show that on average 9,820 people solved the crossword.

Execute a hypothesis test in Excel on the increase of the number of people solving the crossword. (5 points)

Calculate the p-value in Excel. (5 points)

Evaluate (on 5% significance level) whether there is enough statistical evidence that the number of people solving the crossword has increased. (10 points)

Please also upload the Excel file containing your answers.

1. You would like to understand how a government environmental policy is welcomed by different people. You executed a survey of 399 people regarding this new policy. Results are in the Exam\_2021-22-2\_task2\_vA.xlsx Excel file.

Create a crosstab and an associated column chart for Age versus Opinion in Excel. Pay attention to the logical ordering of the opinions (Strongly agree – Strongly disagree) and that the table and the chart is following it. (5 points)

Express the counts as percentages (using a new crosstab) so that for either age groups, the percentages add up to 100%. Add a new Excel column chart and also explain the results verbally. (5 points)

Discuss your findings. Specifically, do the different age groups tend to differ in their opinions about the environmental policy? (5 points)

Please also upload the Excel file containing your answers.

1. Assume that you are the manager of a market research company that was tasked to survey the beverage offerings of bars located in Bulgaria. The requirement is to visit 350 bars. You received a dataset containing all the Bulgarian bars (total of 53,462) and their location data (full address).

Calculate the sampling proportion. (5 points)

Suggest the best sampling strategy. Take care of the adequate geographical coverage of the sample. (10 points)

1. A small computer chip manufacturer wants to forecast monthly operating costs as a function of the number of units produced during a month. The company has collected the 16 months of data in the file Exam\_2021-22-2\_task4\_vA.xlsx.

Determine an equation that can be used to predict monthly production costs from units produced. Is the model and the estimated parameters significant at 5% level? Explain your answer. (10 points)

Are there any outliers? (5 points)

How could the regression line obtained in the previous point be used to determine whether the company was efficient or inefficient during any particular month? (10 points)

Please also upload the Excel file containing your answers.

1. A pharmaceutical company has 16% attrition that is exceptionally high compared to its main competitors. The new HR director would like to take actions to reduce this attrition rate. You were asked to help his job by analysing corporate HR data (Exam\_2021-22-2\_task5\_vA.csv dataset) using advanced analytics. The task is to identify the most important factors leading to attrition. The HR department can use this information to create an action plan.

Execute a classification analysis in Orange and answer the following questions:

Identify the five most important factors related to attrition using Information Gain. (5 points)

Based on the results of the Naïve Bayes method, explain how these five factors affect attrition. (10 points)

Evaluate the results based on the percent correctly predicted (there is no need to separate a train and a test dataset). (10 points)

Please also upload the Orange Workflow file (.ows) you created.