Coursework Description

Problem 1 [35 Marks]

Aware of your expertise in data analysis, the London Assembly Housing Committee (LAHC) has contacted you to undertake the analysis of data related to house price in London and including deprivation and demographic variables [file London House Prices and Deprivation per LSOA.xlsx]. The data contain variables such as median house price, population size for different age groups, surface area, and various deprivation score rank recorded per Lower Super Output Area (LSOA) in London (4,422 records). You are also provided with another spreadsheet (IoD2019_Domains_of_Deprivation.xlsx) containing the 2019 Index of Multiple Deprivation score, its components, and their definition for all Lower Super Output Area (LSOA) in England (including London). They would like you to answer the questions below. **For each question, indicate what assumptions you have made and justify why you choose a particular approach (graph, hypothesis test, both, etc.). These can be described either in your Excel workbook (using notes) or in your report(provided you don't exceed word count. No more** than 2-3 lines.

a. Median house prices **[12]**

   i. Use suitable summary statistics and chart to analyse the distribution of median house prices. [2]

   ii. Calculate mean, min, max, mode, median, 1st and 2nd quartile to compare the median house price per local authority district. [2]

   iii. Use suitable charts to compare the median house price per local authority district. [2]

   iv. Determine confidence intervals for house prices per local authority district. [2]

   v. Using appropriate techniques, help LAHC find out whether there is a significant difference in prices between: [4]

      1. Camden and neighbouring Westminster
      2. Westminster and Kensington and Chelsea
      3. Lewisham and Camden

b. Age groups **[2]**

i. Calculate the percentage of the population in each age group per local authority district.
ii. Use suitable charts to compare the split between age groups per local authority district.

c. Deprivation variables (note that the values are ranks, 1 being the lowest, e.g., most deprived – these are across all of England, not just London) **[2]**
   i. Use suitable charts to explore the distribution of each variable, excluding the house price.
   ii. Use suitable charts to compare each deprivation variable per local authority district.

d. Relationships between variables. **[2]**
   i. Calculate all correlation coefficients where appropriate.
   ii. Explore the relationship between each of the variable and the house price.

e. Model(s) to forecast house price in terms of the other variables. **[7]**
   i. Develop a model using all variables. Comment on your results. [2]
   ii. Develop an improved model using only significant variables. Comment on your results and potential issues. [2]
   iii. Enhance further the model by including/excluding and transforming variables if and when appropriate. [3]

f. In the report, describe your findings. **[10]**

Problem 2 [17 Marks]

London Coaches (LC) is considering running a coach service between London and Brussels. They may operate the service independently, doing their own sales and marketing. They will also have to compete against the existing service on this route. Alternatively, they could form a partnership with the incumbent operator, leaving the sales and marketing to them, and providing coaches and drivers on demand. The profit per month will depend on the number of customers on the route as shown in the table below. The probabilities for customer demand levels, as judged by L*C*, are also given in the table.

No. passengers	Low	Medium	High
Probability	20%	50%	30%
Profit with independent service	–£100,000	£200,000	£600,000
Profit with franchise service	–£20,000	£180,000	£400,000

Of course, if they choose not to run the service, they will have neither profits nor losses.

a. Using a suitable package such as Precision Tree, represent this problem. Clearly indicate what the final recommendation and expected payoff are. **[3]**

b. In order to make a more informed decision, *LC* could hire a transport consultant from the University of Westminster to give them a forecast on the number of passengers. The forecast costs £6,000 and yield two possible results: a "fairly high" or a "fairly low" passenger number prediction. *LC* believes that these predictions are equally likely. The consultant is good at his job and would not predict a fairly low number of passengers if there will be a high number of passengers or a fairly high number of passengers if there will be a low number of passengers. If he predicts a fairly low number of passengers, then there is a 60% chance that there will be a medium number of passengers and if he predicts a fairly high number of passengers then there is a 40% chance that there will be a medium number of passengers. Find the company's optimal course of action using a decision tree. **[5]**

c. Perform sensitivity analysis on the probability that the consultant's prediction about the number of passengers is "fairly low" or "fairly high" to analyse the robustness of your recommendation to possible changes of this parameter. **[4]**

d. In the report, describe your findings. **[5]**

Problem 3 [30 Marks]

The Lewisham City Council is faced with a severe budget shortage due to unexpected social costs from the Covid-19 pandemic. The council is now reconsidering an earlier proposal submitted by a property developer. This proposal seeks to demolish existing council buildings in the Catford area and replace them with modern residential developments that can attract

higher council tax rates. The developer can build one-bedroom, two-bedroom, three-bedroom or four-bedroom houses. The council must, therefore, determine how many existing buildings to demolish and how many units of the new buildings to put up. The following information is available about the project:

- The Catford area has about 250 units of council buildings that can be demolished. The cost of demolishing a building is budgeted to be £2500. Each building occupies a 0.30-acre lot.
- The table below shows the construction cost and the budgeted lot size required for each type of new house. It also shows the expected annual council tax that could be generated from each type of property.

House Type	**Lot size (acre)**	**Unit cost (£)**	**Council tax (£)**
One-bedroom	0.15	45,000	1,292.63
Two-bedroom	0.30	65,000	1,508.87
Three-bedroom	0.35	120,000	1,723.51
Four-bedroom	0.40	150,000	1,938.95

- Construction policies require that 10% of the available acreage after demolishing be reserved for streets, utility instalments and other emergency facilities. Additionally, at least 10% of the new units should be four-bedroom houses. At least 5% of the new units should be two-bedroom houses. Finally, at most 45% of the new units should be two- and three-bedroom houses.
- The council can access a maximum of £12 million financing facility with a High Street bank.

The council will want to determine the maximum annual council tax that could be generated after this project. Answer the following questions:

a. Formulate a linear programming model to determine the total number of each type of new housing unit to be built and the number of old buildings to be demolished to maximise council tax generated. **[7]**

b. Formulate and solve the LP model using Excel solver. How much tax will the current plan generate? **[8]**

c. Using the sensitivity report, answer the following questions: **[8]**

i. Why are there no three-bedroom houses to be built in the current plan? Advise the council on two ways to achieve positive units of three-bedroom units in the optimal plan.

ii. What will be the effect on the optimal solution and tax if the council tax for a four-bedroom unit is set equal to that for a three-bedroom house?

iii. The council is considering increasing council taxes by 10%. What will be the impact of this policy on the optimal solution and total tax generated?

d. In the report, describe your findings. **[7]**

Problem 4 [18 Marks]

A company sells air conditioners to four regions labelled as East, South, North, and West. The company has gathered historical data of the monthly demand for air conditioners over the past 9 years (since 2013)- see file Data_Pb4.xlsx.

a. Explore the time series for each region. **[6]**

i. Plot the data for each region as time series.

ii. Determine whether these exhibit trend and/or seasonality?

iii. If/Where relevant, calculate seasonal indexes for each region.

iv. State with justification for each region what would be appropriate forecasting methods to forecast monthly demand for the remainder of 2023.

b. Apply two appropriate forecasting methods for each region to forecast monthly demand remainder of 2023. **[7]**

i. For each region, compare performance of the two methods using MAE, RMSE, and MAPE.

ii. Check and comment on forecasting errors for randomness and autocorrelation.

iii. Using the best method, produce monthly forecast for each region remainder of 2023.

c. In the report, describe your findings. **[5]**