**Context:** You are a data scientist in the car insurance area of an insurance company. Attached is a dataset “CaseStudyDatasetQ.csv” that contains data related to the historical car accident claims received by your company.

The dataset contains feature vector 𝒙 = (𝑥1, 𝑥2, 𝑥3, 𝑥4, 𝑥5) where:

• 𝑥1 refers to car weight: weight;

• 𝑥2 refers to annually driven distance: distance;

• 𝑥3 refers to age of driver: age;

• 𝑥4 refers to age of car: carage;

• 𝑥5 refers to gender of driver: gender;

In addition to feature vector 𝒙𝑖, the data includes volume: 𝑣𝑖 (exposure) and claims count (Counts) 𝑁𝑖~𝑃𝑜𝑖𝑠𝑠𝑜𝑛(𝑣𝑖𝜆(𝑥𝑖)) for a portfolio of car insurance policies 𝑖 = 1, … , 𝑛. Requirement:

Draft a technical report to your manager outlining an approach to predicting the claims frequency of this portfolio. Consider the following tasks:

1. Perform exploratory data analysis on the dataset including a. Data checks b. summary statistics of the variables c. any relationships between the variables including any visualisations d. any hypothesis based on these relationships

2. Estimate 𝜆(𝒙) assuming linear terms in 𝑥1, … , 𝑥5, quadratic terms in 𝑥1 2,… , 𝑥5 2 and

mixed terms 𝑥1𝑥2, 𝑥1𝑥3, 𝑥1𝑥4, 𝑥2𝑥3, 𝑥2𝑥4, 𝑥3𝑥4. Select the best model.

3. Estimate 𝜆(𝒙) using Poisson regression tree, including in your commentary the optimal size of the tree.

4. Estimate 𝜆(𝒙) using Poisson boosting tree method (with no base model), including in your commentary the number of boosting steps and the size of the tree

5. Compare the models fitted in (2-4) and any other model you choose using 10-fold cross validation error to select the best model

Summarise your results and findings in a technical report of 1-2 pages in length covering: background, methodology and approach for model selection and results.

Include an appendix to hold

1. Any analysis you have conducted that you would like to share outside of the body of the report as Appendix A

2. Your code created for the analysis as Appendix B Submit both your report and any code created.