Exam PA October 11 Project Statement

IMPORTANT NOTICE – THIS IS THE OCTOBER 11, 2022, PROJECT STATEMENT. IF TODAY IS NOT OCTOBER 11, 2022, SEE YOUR TEST CENTER ADMINISTRATOR IMMEDIATELY.

General Information for Candidates

This examination has 12 tasks numbered 1 through 12 with a total of 100 points. The points for each task are indicated at the beginning of the task, and the points for subtasks are shown with each subtask.

Each task pertains to the business problem (and related data files) and data dictionary described below. Additional information on the business problem may be included in specific tasks—where additional information is provided, including variations in the target variable, it applies only to that task and not to other tasks. An .Rmd file accompanies this exam and provides useful R code for importing the data and, for some tasks, additional analysis and modeling. There are five datasets used in this exam. They are all subsets of a larger dataset that is not given to candidates. The .Rmd file has a chunk for each task. Each chunk starts by reading in one or more data files into one or more dataframes that will be used in the task. This ensures a common starting point for candidates for each task and allows them to be answered in any order. When the datafile is read, the variables it contains are assigned a type (e.g., “numerical,” ”factor”). The code that assigns variable types is easily changed (e.g., if month is read in as “numeric” but you want to treat it as a factor).

The responses to each specific subtask should be written after the subtask and the answer label, which is typically ANSWER, in this Word document. Each subtask will be graded individually, so be sure any work that addresses a given subtask is done in the space provided for that subtask. Some subtasks have multiple labels for answers where multiple items are asked for—each answer label should have an answer after it. Where code, tables, or graphs from your own work in R is required, it should be copied and pasted into this Word document.

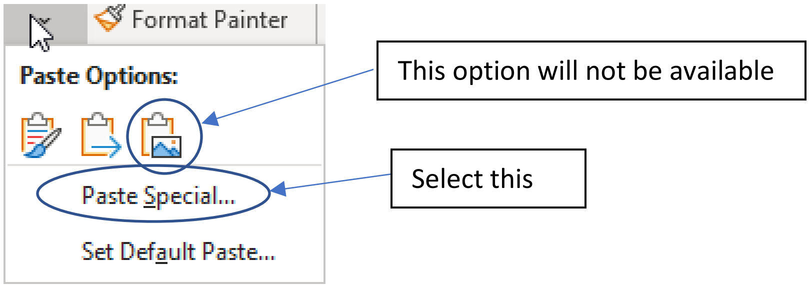
Each task will be graded on the quality of your thought process (as documented in your submission), conclusions, and quality of the presentation. The answer should be confined to the question as set. No response to any task needs to be written as a formal report. Unless a subtask specifies otherwise, the audience for the responses is the examination grading team and technical language can be used. When “for a general audience” is specified, write for an audience **not** familiar with analytics acronyms (e.g., RMSE, GLM, etc.) or analytics concepts (e.g., log link, binarization).

Prior to uploading your Word file, it should be saved and renamed with your five-digit candidate number in the file name. If any part of your exam was answered in French, also include “French” in the file name. Please keep the exam date as part of the file name.

It is not required to upload your .Rmd file or other files used in determining your responses, as needed items from work in R will be copied over to the Word file as specified in the subtasks.

The Word file that contains your answers must be uploaded before the five-minute upload period time expires.

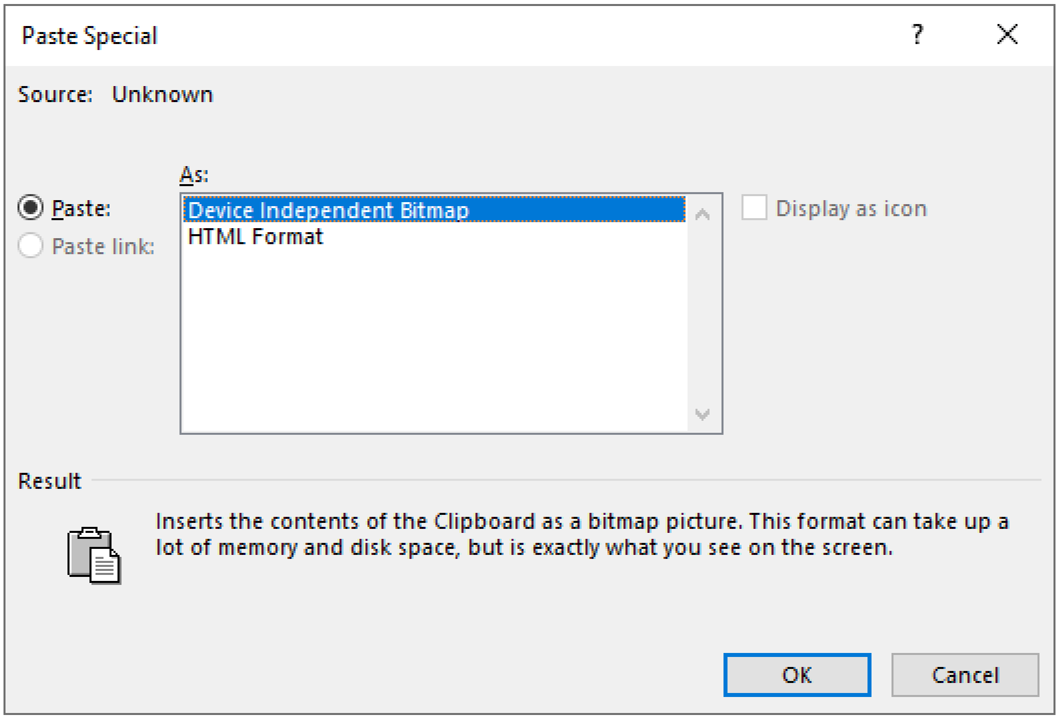
IMPORTANT NOTE: When pasting a picture from RStudio to Word, there is only one approach that will work. After right clicking on the image in RStudio and selecting “copy” the following steps need to be taken in Word. On the Home menu, click on the down arrow under “Paste” and then select “Paste Special …” From the list of options, select “Device Independent Bitmap.” The following images indicate these steps.



This option will not be available

Select this

From this dialog box, make the indicated selection.


Paste Special

Source: Unknown

As:

● Paste: Device Independent Bitmap

○ Paste link:

HTML Format

□ Display as icon

Result

Inserts the contents of the Clipboard as a bitmap picture. This format can take up a lot of memory and disk space, but is exactly what you see on the screen.

Business Problem

Your boss recently started a consulting firm, PA Consultants, specializing in predictive analytics. You and your assistant are the only other employees. Your boss informs you that a local politician from Baton Rouge, Louisiana, USA has hired your firm.

Baton Rouge, a city of about 230,000 residents, is the capital of the state of Louisiana, USA.

The client is about to launch a campaign with the mottos, "Clean up Baton Rouge" and "Treat all Neighborhoods Equally – including yours!" The client wants to improve garbage and waste collection. In particular, the client cares about shortening resolution times and ensuring equitable resolution times throughout the city.

The client wants your ideas and inputs on the following.

- *Understanding time trends*
- *Seeing whether different responding departments have different resolution times for similar tasks*
- *Predicting resolution times for any type(s) of complaint.*

Your boss directs you to use a dataset[1] of public data that includes all the service requests from January, 2016 – March, 2022. There are over 300,000 service requests in this time period. Your assistant has prepared five subsets of the public data and has provided the following data dictionary that contains all the variables appearing in the subsets. Note that all variables do not appear in every subset datafile.

[1] *Source: City of Baton Rouge Parish of East Baton Rouge*.

Data Dictionary

Variable Name	**Variable Values**
Time.to.resolution	Days from service request to resolution
quarter	“Q1”, “Q2”, “Q3”, “Q4”; quarter of service request
month	1 to 12, month of service request
year	2016 to 2022, year of service request
year.mo	201601 to 202203, 100*year + month
DEPARTMENT	“GROUNDS”,”BLIGHT”,”SANITATION”
LATITUDE	Latitude of service location, 30.2 to 30.6
LONGITUDE	Longitude of service location, -91.3 to -90.9
area	“N”,”W”,”D”,”LSU”; neighborhood of service location
Latitude_Binned	Latitude range for binned data (geo.grid.csv only)
Longitude_Binned	Longitude range for binned data (geo.grid.csv only)
Ave.time.to.resolution	Average Time.to.resolution for binned data (geo.grid.csv only)
call.count	Number of service requests for binned data (geo.grid.csv only)
TYPEid	An id representing a specific type of service request

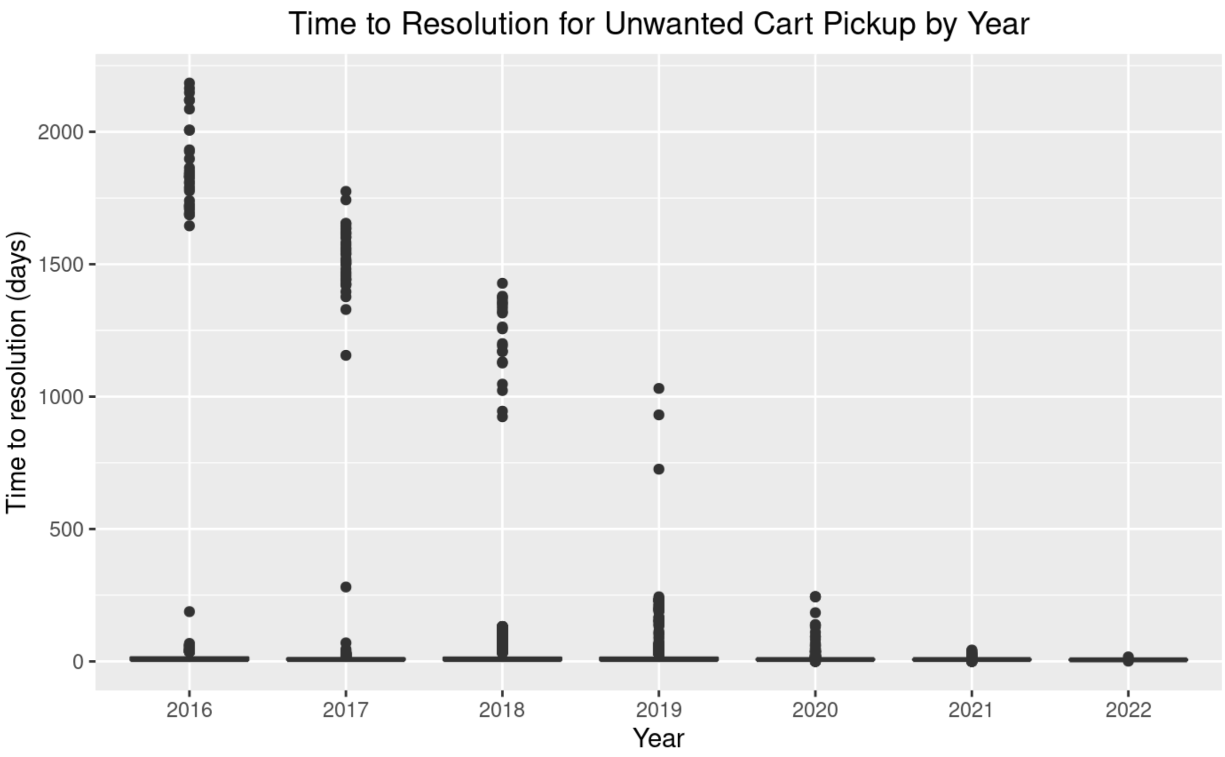
Comments

Requests for service do not appear in the dataset until they are resolved.

Task 1 (*7 points*)

Your boss asks you to review the quality of the data below. The data shows Time to Resolution for calls to pick up unwanted garbage carts. (This data is not found in any of the supplied files.)

(a) (*2 points*) Review the box plot below that your assistant made and describe an issue with the data.



ANSWER:

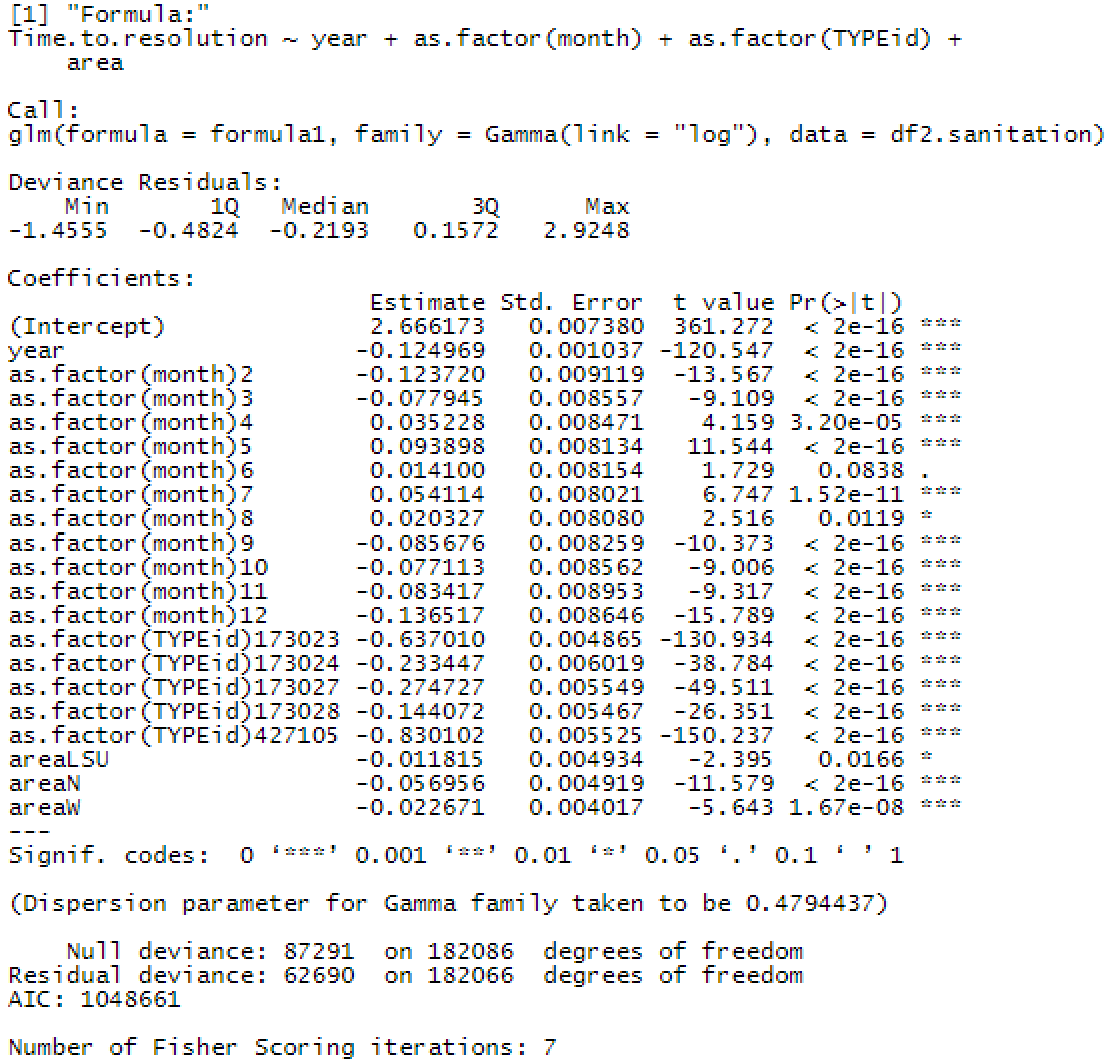
(b) (*1 point*) List three options for handling the data issue.

ANSWER:

(c) (*2 points*) Select and explain which option from part (b) you would recommend.

ANSWER:

(d) (2 *points*) Your assistant produces the following output from a GLM. (Note your assistant redefined year as years since 2016.)



```
[1] "Formula:"
Time.to.resolution ~ year + as.factor(month) + as.factor(TYPEid) +
    area

Call:
glm(formula = formula1, family = Gamma(link = "log"), data = df2.sanitation)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4555  -0.4824  -0.2193   0.1572   2.9248

Coefficients:
                         Estimate Std. Error  t value Pr(>|t|)
(Intercept)              2.666173   0.007380  361.272  < 2e-16 ***
year                    -0.124969   0.001037 -120.547  < 2e-16 ***
as.factor(month)2       -0.123720   0.009119  -13.567  < 2e-16 ***
as.factor(month)3       -0.077945   0.008557   -9.109  < 2e-16 ***
as.factor(month)4        0.035228   0.008471    4.159 3.20e-05 ***
as.factor(month)5        0.093898   0.008134   11.544  < 2e-16 ***
as.factor(month)6        0.014100   0.008154    1.729   0.0838 .
as.factor(month)7        0.054114   0.008021    6.747 1.52e-11 ***
as.factor(month)8        0.020327   0.008080    2.516   0.0119 *
as.factor(month)9       -0.085676   0.008259  -10.373  < 2e-16 ***
as.factor(month)10      -0.077113   0.008562   -9.006  < 2e-16 ***
as.factor(month)11      -0.083417   0.008953   -9.317  < 2e-16 ***
as.factor(month)12      -0.136517   0.008646  -15.789  < 2e-16 ***
as.factor(TYPEid)173023 -0.637010   0.004865 -130.934  < 2e-16 ***
as.factor(TYPEid)173024 -0.233447   0.006019  -38.784  < 2e-16 ***
as.factor(TYPEid)173027 -0.274727   0.005549  -49.511  < 2e-16 ***
as.factor(TYPEid)173028 -0.144072   0.005467  -26.351  < 2e-16 ***
as.factor(TYPEid)427105 -0.830102   0.005525 -150.237  < 2e-16 ***
areaLSU                 -0.011815   0.004934   -2.395   0.0166 *
areaN                   -0.056956   0.004919  -11.579  < 2e-16 ***
areaW                   -0.022671   0.004017   -5.643 1.67e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.4794437)

    Null deviance: 87291  on 182086  degrees of freedom
Residual deviance: 62690  on 182066  degrees of freedom
AIC: 1048661

Number of Fisher Scoring iterations: 7
```

Calculate the residual for the predicted time to resolution using the values in the following table for a single observation. Show both the formula(s) used (with values substituted for variables) and the final value to two decimal places.

TYPEid	month	year	Area	Time to Resolution
173023	2	4	N	5

ANSWER:

Task 2 *(11 points)*

The client is interested in improving the debris collection performance.

(a) (2 points) Create a table showing number of observations by year and month. Paste the R code and the table below.

ANSWER:

(b) (2 *points*) Recommend which time period you will choose to use for your analysis (in terms of years and months). Justify your recommendation.

ANSWER:

Your boss told your assistant to use stratified sampling when separating the chosen dataset into a training dataset and a testing dataset.

(c) (2 *points*) Discuss the benefits of stratified sampling.

ANSWER:

Your assistant has stratified the entire dataset, based on month and year, and divided it into train and test datasets. You need to remove any observations that you decided not to use in (b).

(d) (*2 points*) Remove the observations that you decided in (b) not to use from the train and test datasets. Copy the code to adjust datasets.

ANSWER:

Code to adjust datasets:

Your assistant has prepared glm1 and glm2. Run the .Rmd file to fit the models.

(e) (3 *points*) State the better of the two models, based on RMSE. Copy the code (i.e., the glm command, and any further lines of code) for both of the models that you used to make the choice.

ANSWER:

Model choice (erase one): Gamma GLM or Poisson GLM

RMSE for Gamma GLM:

RMSE for Poisson GLM:

Code to calculate Gamma GLM RMSE:

Code to calculate Poisson GLM RMSE:

Task 3 *(9 points)*

Your boss is interested in providing an update to the client about interesting findings coming out of the exploratory data analysis. Specifically, your boss would like to present on how resolution times vary by year and if there are any differences by department:

(a) *(3 points)* Create a graph to show how resolution times vary by year, split out by department type. Paste the graph and paste code for the graph below. (Refer to the code provided in the RMD file.)

ANSWER:

Graph:

Code:

(b) (*2 points*) Describe any trends seen in the graph.

ANSWER:

You have asked your assistant to use stepwise selection as a possible method to select predictors in a final model.

(c) (*2 points*) Contrast best subset and stepwise selection for selecting predictors.

ANSWER:

Your assistant believes that stepwise selection could lead to a suboptimal model being fit and that best subset selection should always be performed.

(d) (*2 points*) Critique the assistant's assertion that best subset selection should always be used since stepwise selection could lead to a suboptimal choice for the model.

ANSWER:

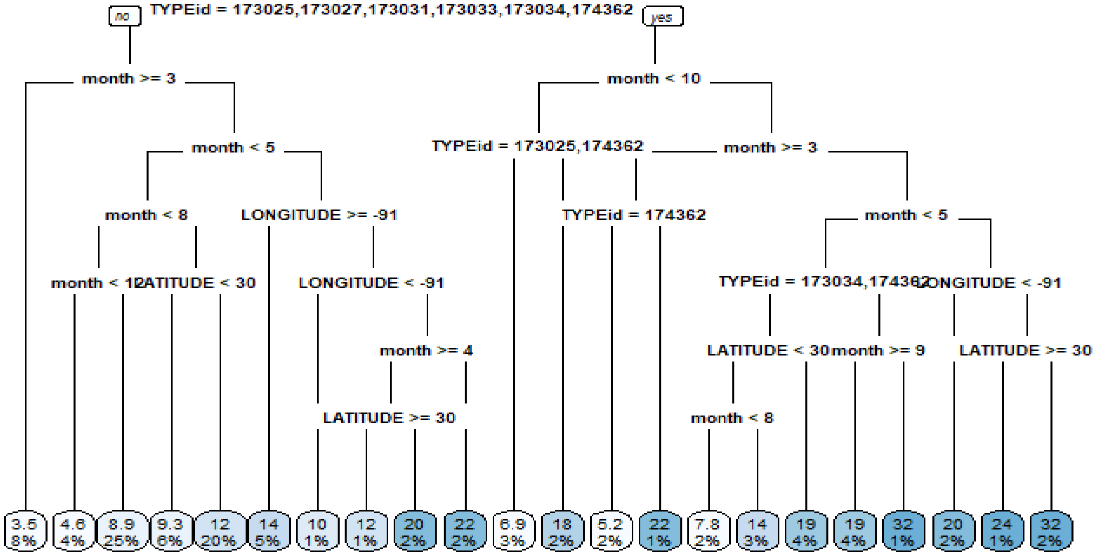
Task 4 *(7 points)*

Your boss wants you to build a tree model to better understand the Time.to.resolution for discarded couches and mattresses. (The data used is not found in any of the supplied files.)

(a) (2 *points*) Describe two ways impurity measures are used in a classification tree.

ANSWER:

Your assistant has built a tree model and noticed that for all values of the *cp* parameter the model never splits on DEPARTMENT but always includes splits on TYPEid.



Your assistant also generated a summary table by DEPARTMENT and TYPEid for you to review.

DEPARTMENT	TYPEid	n()	mean(Time.to.resolution)
BLIGHT	173034	19	22.421053
GROUNDS	173021	3	5
GROUNDS	174362	337	20.175074
SANITATION	173020	1	3
SANITATION	173023	13	8.615385
SANITATION	173024	3012	10.134462
SANITATION	173025	1	18
SANITATION	173026	23	8.347826
SANITATION	173027	686	15.262391
SANITATION	173029	1	5
SANITATION	173031	1	19
SANITATION	173032	9	5.333333
SANITATION	173033	17	16.647059

(b) (*2 points)* Explain why the classification tree does not split on DEPARTMENT.

ANSWER:

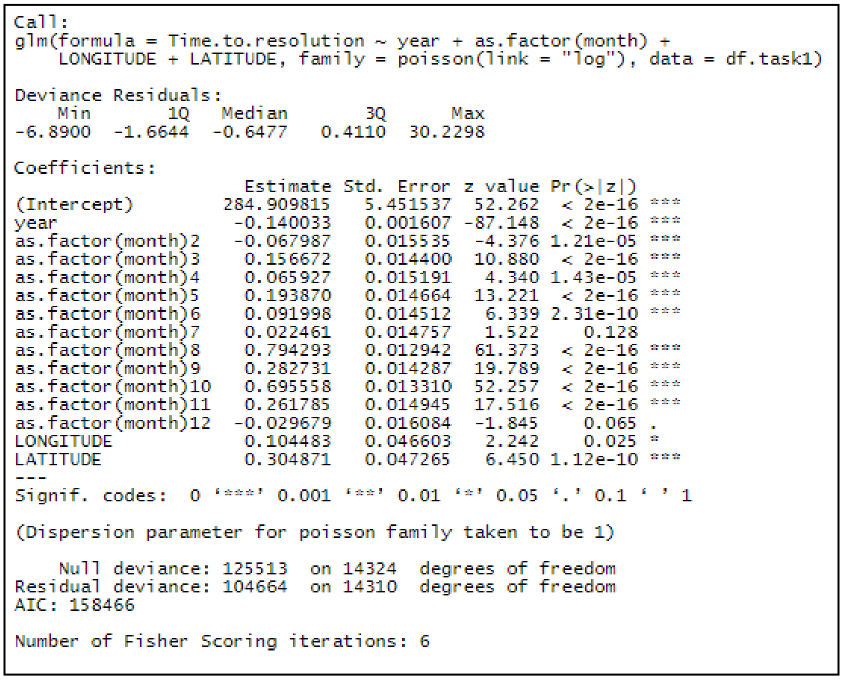
Based on the preliminary findings, your boss suggests you round the values of LONGITUDE and LATITUDE variables to 1 decimal place.

(c) (*3 points*) Explain potential issues with the LONGITUDE and LATITUDE variable before they were rounded and how your boss's suggestion would address these concerns.

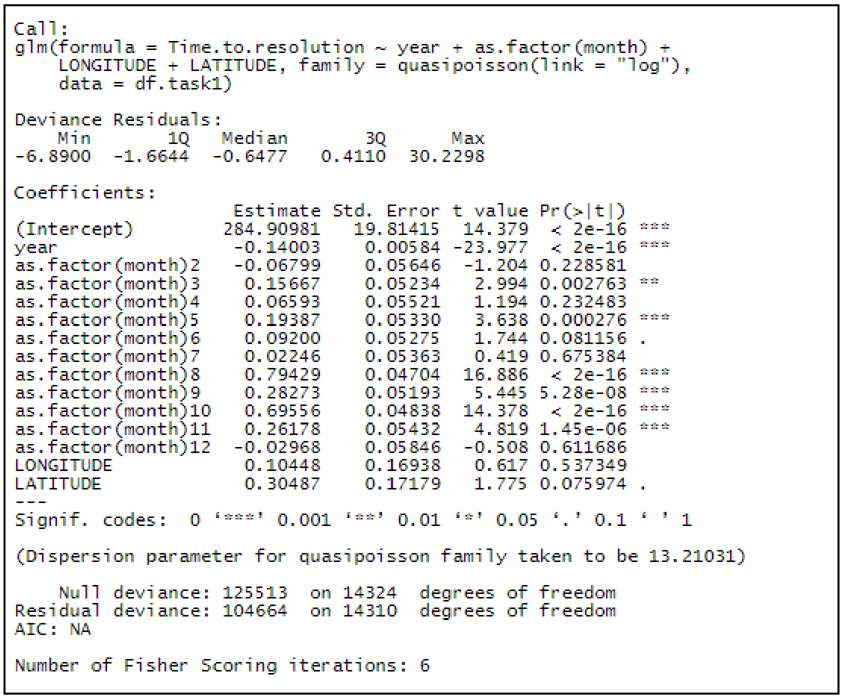
ANSWER:

Task 5 *(5 points)*

Your assistant fit a GLM to predict the resolution time for garbage cart requests from new residents. (The data used is not in any of the supplied files.) The assistant chose to fit two different distributions, a Poisson and a Quasi-Poisson distribution. Refer to output below:



```
Call:
glm(formula = Time.to.resolution ~ year + as.factor(month) + 
    LONGITUDE + LATITUDE, family = poisson(link = "log"), data = df.task1)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-6.8900  -1.6644  -0.6477   0.4110  30.2298  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)        284.909815   5.451537  52.262  < 2e-16 ***
year                -0.140033   0.001607 -87.148  < 2e-16 ***
as.factor(month)2   -0.067987   0.015535  -4.376 1.21e-05 ***
as.factor(month)3    0.156672   0.014400  10.880  < 2e-16 ***
as.factor(month)4    0.065927   0.015191   4.340 1.43e-05 ***
as.factor(month)5    0.193870   0.014664  13.221  < 2e-16 ***
as.factor(month)6    0.091998   0.014512   6.339 2.31e-10 ***
as.factor(month)7    0.022461   0.014757   1.522    0.128    
as.factor(month)8    0.794293   0.012942  61.373  < 2e-16 ***
as.factor(month)9    0.282731   0.014287  19.789  < 2e-16 ***
as.factor(month)10   0.695558   0.013310  52.257  < 2e-16 ***
as.factor(month)11   0.261785   0.014945  17.516  < 2e-16 ***
as.factor(month)12  -0.029679   0.016084  -1.845    0.065 .  
LONGITUDE            0.104483   0.046603   2.242    0.025 *  
LATITUDE             0.304871   0.047265   6.450 1.12e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 125513  on 14324  degrees of freedom
Residual deviance: 104664  on 14310  degrees of freedom
AIC: 158466

Number of Fisher Scoring iterations: 6
```



```
Call:
glm(formula = Time.to.resolution ~ year + as.factor(month) + 
    LONGITUDE + LATITUDE, family = quasipoisson(link = "log"), 
    data = df.task1)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-6.8900  -1.6644  -0.6477   0.4110  30.2298  

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)        284.90981   19.81415  14.379  < 2e-16 ***
year                -0.14003    0.00584 -23.977  < 2e-16 ***
as.factor(month)2   -0.06799    0.05646  -1.204 0.228581    
as.factor(month)3    0.15667    0.05234   2.994 0.002763 ** 
as.factor(month)4    0.06593    0.05521   1.194 0.232483    
as.factor(month)5    0.19387    0.05330   3.638 0.000276 ***
as.factor(month)6    0.09200    0.05275   1.744 0.081156 .  
as.factor(month)7    0.02246    0.05363   0.419 0.675384    
as.factor(month)8    0.79429    0.04704  16.886  < 2e-16 ***
as.factor(month)9    0.28273    0.05193   5.445 5.28e-08 ***
as.factor(month)10   0.69556    0.04838  14.378  < 2e-16 ***
as.factor(month)11   0.26178    0.05432   4.819 1.45e-06 ***
as.factor(month)12  -0.02968    0.05846  -0.508 0.611686    
LONGITUDE            0.10448    0.16938   0.617 0.537349    
LATITUDE             0.30487    0.17179   1.775 0.075974 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 13.21031)

    Null deviance: 125513  on 14324  degrees of freedom
Residual deviance: 104664  on 14310  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6
```

(a) (3 *points*) Assess the two chosen distributions with respect to reasonability in modeling Time.to.resolution as a target variable, using the output provided by the assistant.

ANSWER:

Your boss would like you to consider other distributions for the GLM.

(b) (*2 points*) Recommend two additional distributions along with link functions that are reasonable choices to model Time.to.resolution. Justify your recommendations.

ANSWER:

Task 6 *(10 points)*

The client is interested in improving furniture disposal pickup times. Your assistant prepares a GLM and a decision tree that model Time.to.resolution using LATITUDE and LONGITUDE as predictor variables. (The data used is not in any of the supplied files.)

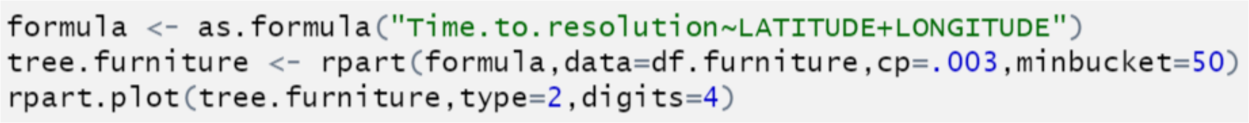
(a) (*2 points*) Contrast using a GLM versus a decision tree given the client's goals and the variables chosen to use in these models.

ANSWER:

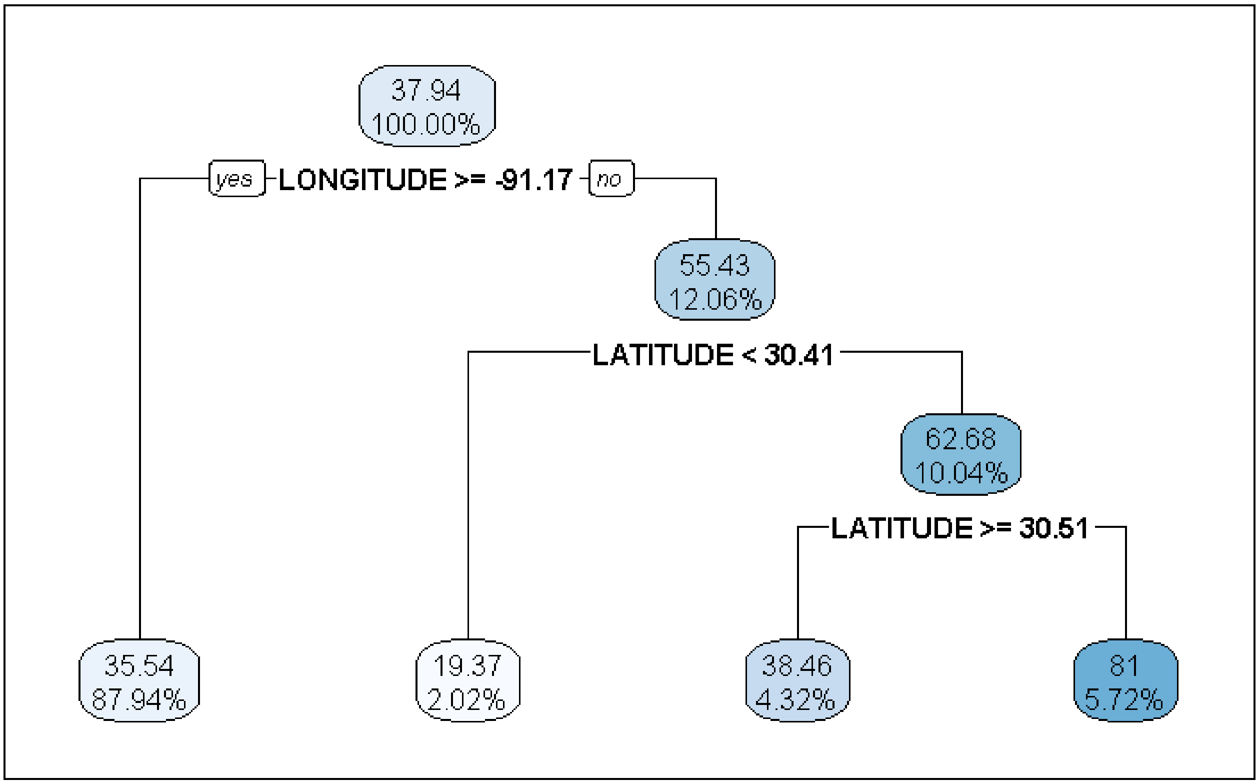
(b) (*2 points*) Recommend either the GLM or decision tree to use and justify your recommendation.

ANSWER:

Your assistant produced a decision tree to predict Time.to.resolution using LATITUDE and LONGITUDE. Your assistant provides you with the following code and output below:



```
formula <- as.formula("Time.to.resolution~LATITUDE+LONGITUDE")
tree.furniture <- rpart(formula,data=df.furniture,cp=.003,minbucket=50)
rpart.plot(tree.furniture,type=2,digits=4)
```



(c) (*3 points*) Interpret a few select components of this plot by filling out the table below:

ANSWER:

Component of Plot	Interpretation
55.43	
12.06%	
Latitude < 30.41	
38.46	
5.72%	

Your assistant wants to recommend that the client includes shortening furniture disposal service request resolution times as part of their campaign.

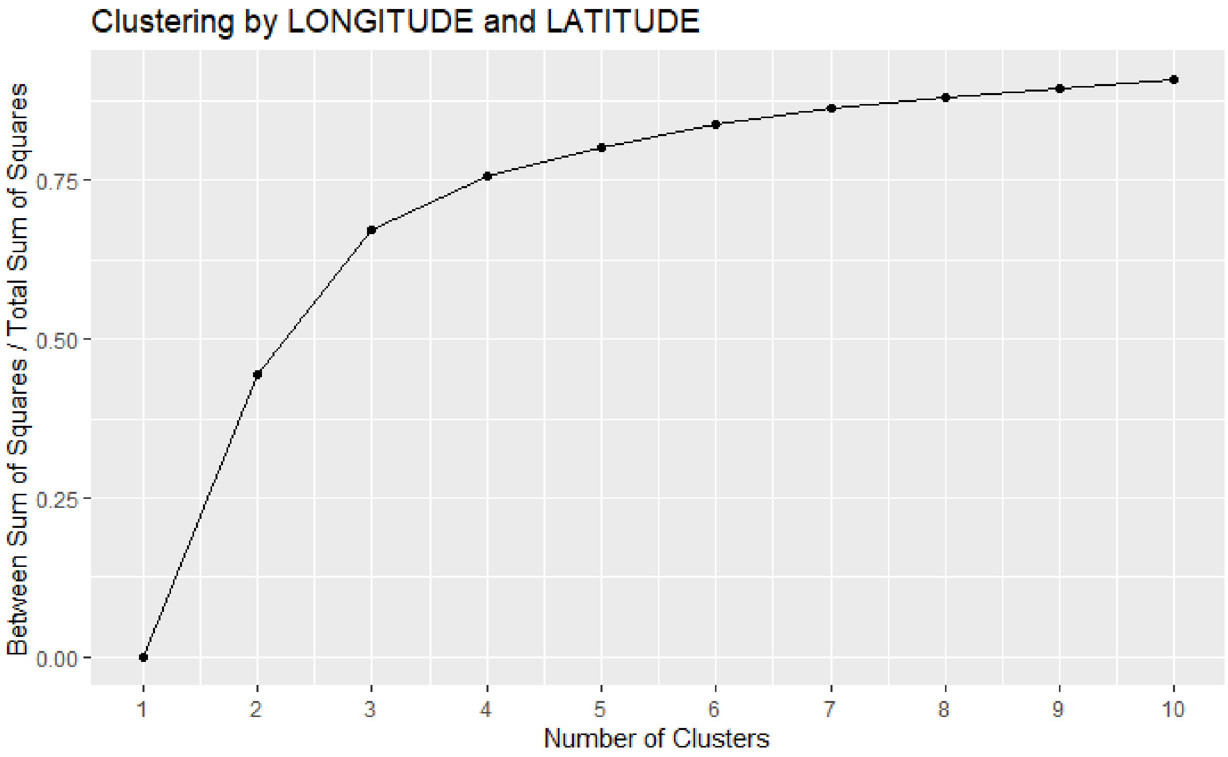
(d) (*3 points*) Critique your assistant's recommendation and consider model efficacy and potential equity concerns.

ANSWER:

Task 7 (9 *points*)

As part of the exploratory data analysis process, your assistant decides to focus on service requests involving mattresses, couches, and sofas. To reduce the number of observations, your assistant also restricts the dataset to include only observations from prior to 2020. (The data used is not in any of the supplied files.)

Using the dataset described above, your assistant decides to use K-means clustering using LATITUDE and LONGITUDE variables and produces the plot shown below.



(a) (*2 points*) Identify the type of plot and explain what it depicts.

ANSWER:

(b) (2 *points*) Recommend the number of clusters to use and justify your recommendation.

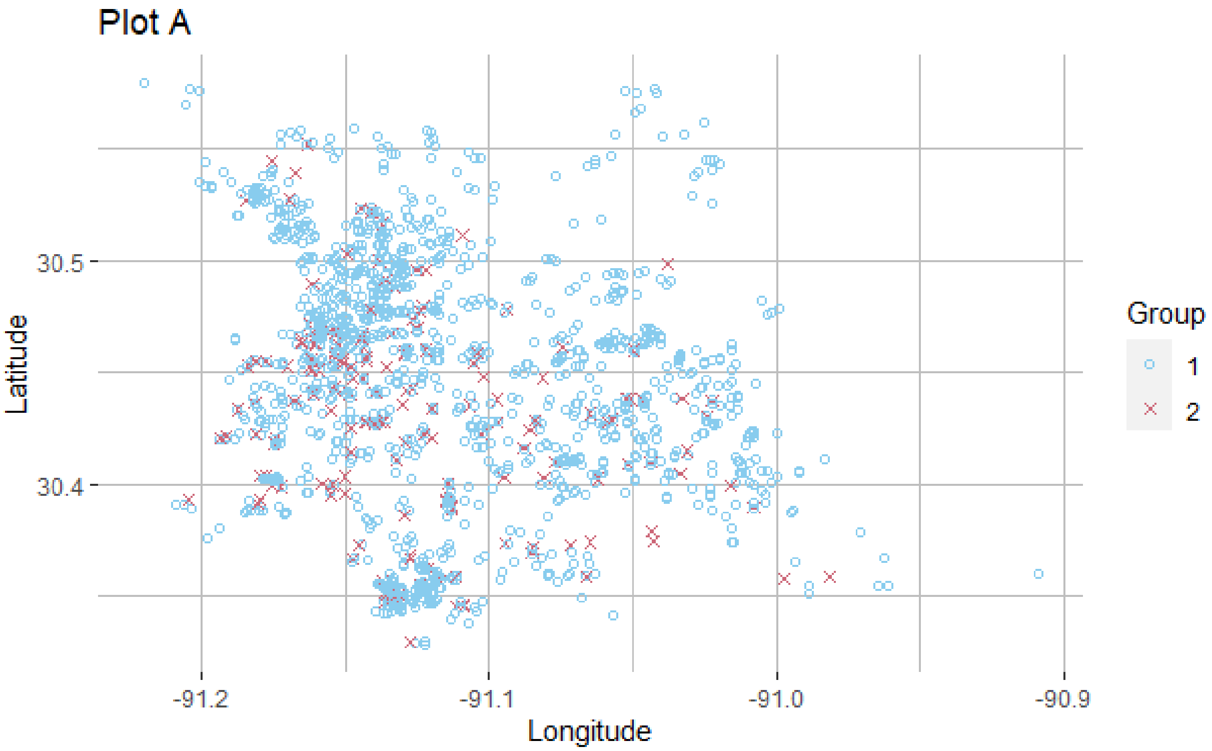
ANSWER:

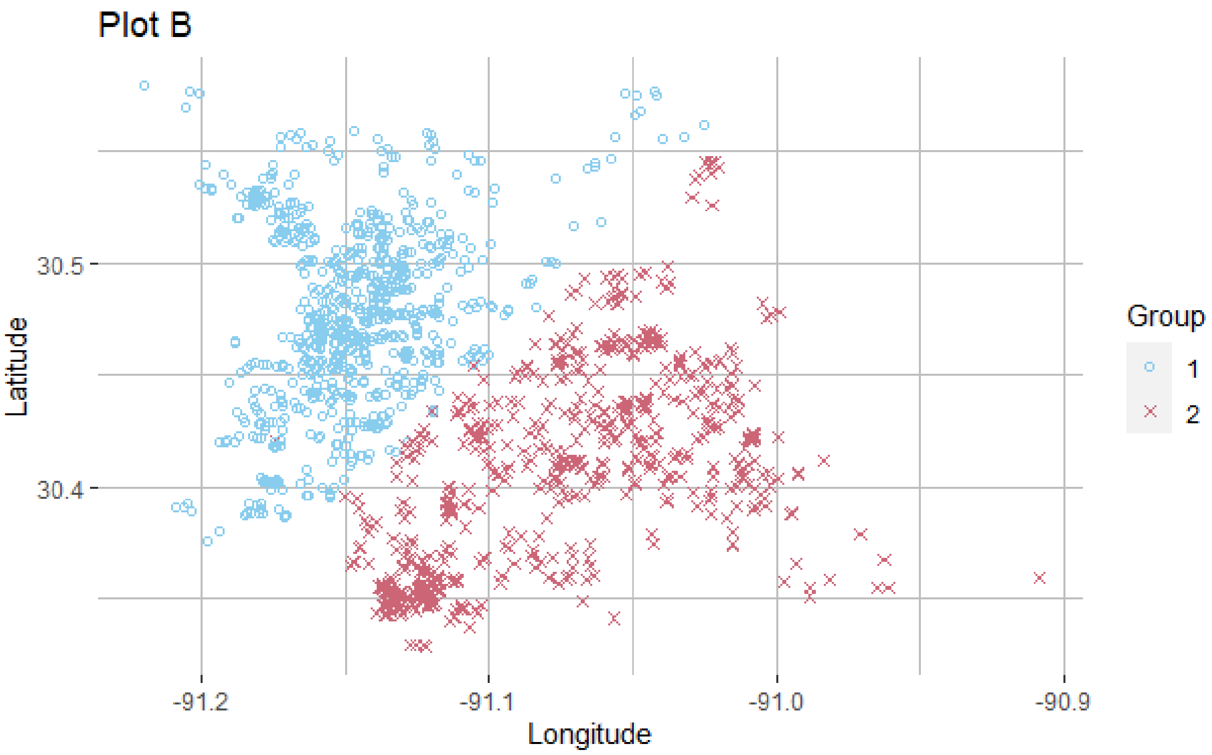
Your assistant is using K-means clustering to create a feature to be used as a predictor variable in a GLM where the target variable is time to resolution. Your assistant wants to use the following three variables as inputs into the K-means clustering algorithm: LATITUDE, LONGITUDE, and Time.to.resolution.

(c) (*2 points*) Critique the recommendation described above.

ANSWER:

Using the variables LATITUDE, LONGITUDE, and Time.to.resolution, your assistant performs K-means clustering with K=2. Your assistant performs the analysis with **and** without variable scaling but forgets to label the output properly. When clustering is done on scaled variables, the plot is made using the unscaled values. Note that the plots below depict only two of the three variables.





(d) (*3 points*) Identify which plot reflects the version where the variables were scaled and discuss how scaling created such large differences in how the data appear in these plots.

ANSWER:

Task 8 *(11 points)*

Your assistant shared with you a decision tree built to model Time.to.resolution for furniture-related complaints. The predictor variables are year, LONGITUDE, and LATITUDE. The resulting tree appears to be overly complex. Your assistant seeks your guidance to help improve this model. (The data used is not in any of the supplied files.)

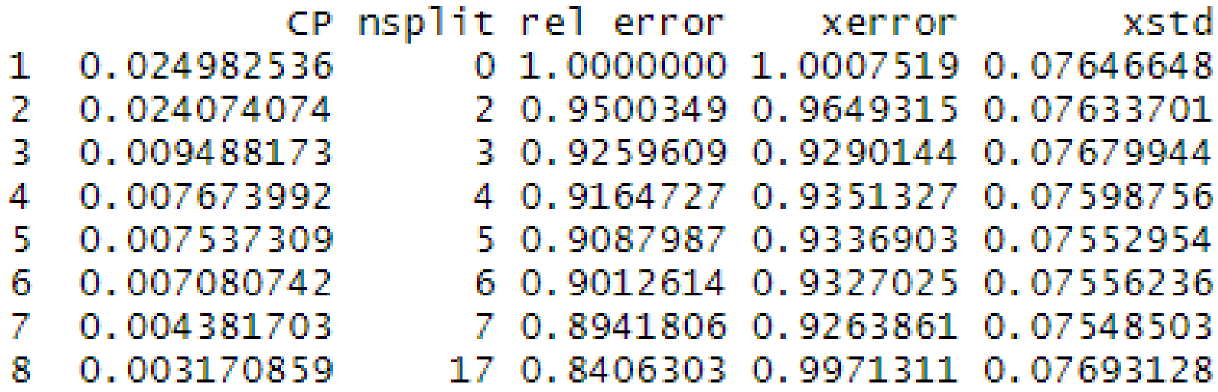
(a) (*3 points*) Describe the cost-complexity pruning algorithm and what purpose it serves.

ANSWER:

(b) (*3 points*) Describe two common approaches for choosing a complexity parameter based on cross-validation results.

ANSWER:

Your assistant shares the results of the complexity parameter below.



```
           CP nsplit rel error    xerror       xstd
1 0.024982536      0 1.0000000 1.0007519 0.07646648
2 0.024074074      2 0.9500349 0.9649315 0.07633701
3 0.009488173      3 0.9259609 0.9290144 0.07679944
4 0.007673992      4 0.9164727 0.9351327 0.07598756
5 0.007537309      5 0.9087987 0.9336903 0.07552954
6 0.007080742      6 0.9012614 0.9327025 0.07556236
7 0.004381703      7 0.8941806 0.9263861 0.07548503
8 0.003170859     17 0.8406303 0.9971311 0.07693128
```

(c) (*2 points*) Apply both methods described in part (b). Recommend the number of splits to use in the pruned tree. Justify your recommendation.

ANSWER:

(d) (*3 points*) Recommend either to prune the overly complex tree with an optimally selected complexity parameter or to build a new tree with that same complexity parameter. Justify your recommendation.

ANSWER:

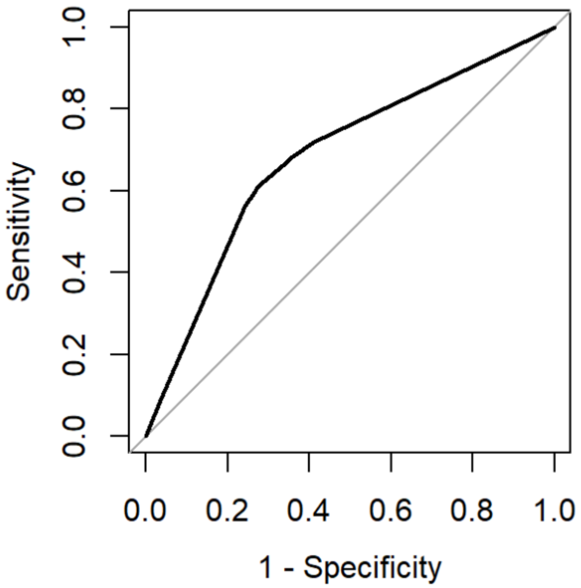
Task 9 *(10 points)*

Your client has a goal to resolve missed pickups service calls to fewer than two days. Your boss wants you to build a model to evaluate this and suggests using AUC as a performance metric. (The data used is not in any of the supplied files.)

(a) *(2 points)* Explain the difference between accuracy and AUC in terms of overall model assessment.

ANSWER:

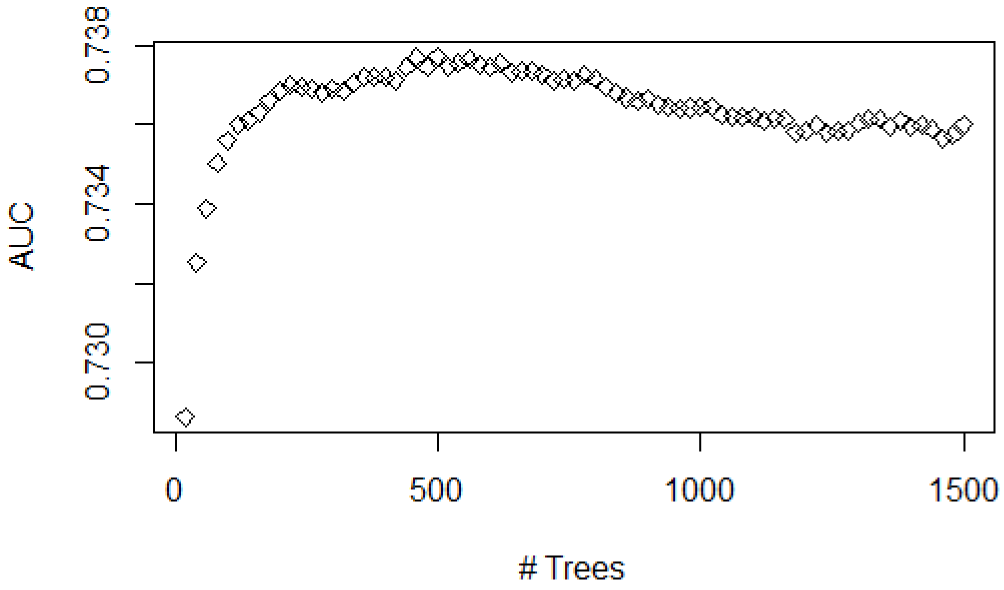
Your assistant built a model and plotted the ROC curve below.



(b) (2 points) Explain why the ROC curve always goes through (0,0) and (1,1).

Your boss suggests a boosted tree can increase model performance by reducing bias, however, setting hyperparameters is critical. You are asked to build a gradient boosting machine (GBM) tree model to assess the performance improvement.

The GBM tree model performance using the test data set is shown below.



(c) (2 *points*) Explain why model performance improves at beginning then deteriorates as the number of trees increases.

ANSWER:

(d) (2 *points*) Describe two hyperparameters you could adjust to improve model performance.

ANSWER:

(e) (*2 points*) Explain the process of how to tune a hyperparameter.

ANSWER:

Task 10 *(6 points)*

The client is interested in estimating the impact of various predictors on Time.to.resolution for two common complaints: “MISSED GARBAGE SERVICE DAY (GENERAL PICK-UP)” and “MISSING GARBAGE CART.” The client is interested in resolution time trends. Another concern is whether resolution times differ for certain areas within the city.

Run the given code and use the output to answer the following.

(a) (*3 points*) Interpret the coefficients for the time variables (year, quarter) for the two models (one for each complaint) using the summary() output. Also describe the trends of resolution times for each of the two complaints.

ANSWER:

(b) (*3 points*) Using the summary() and drop1() output, compare and contrast the significance of the area variables in the two models. Quantify significant differences in resolution times.

ANSWER:

Task 11 *(8 points)*

You are investigating data on calls for damaged carts using Time.to.resolution as the target variable. This dataset includes an additional variable "Service.Request.Id." This variable is set to 1 for the first request and incremented by one at each subsequent request. Your assistant has removed this variable, arguing that it is not of any value for predicting Time.to.resolution, given that is merely a counter that reflects the row of the observation. (The data used is not in any of the supplied files.)

(a) (2 *points*) Critique the assistant's recommendation.

ANSWER:

(b) (*1 point*) Define an interaction effect.

ANSWER:

Many service calls for damaged carts have resolution times over 60 days. You have been asked to look at these in more detail. Your assistant has built an initial model to predict if a damaged cart call will take more than 60 days to service. The predictive variables used are: year, month, DEPARTMENT, LATITUDE, LONGITUDE. Consider interactions among the predictor variables.

(c) (*2 points*) Propose two variables to make an interaction term that may improve model accuracy. Justify your proposal.

ANSWER:

You continue working on a model to predict if a call for a damaged cart will have resolution times over 60 days. A new indicator variable "Over60" has been created to identify records that have a resolution time greater than 60 days.

Your assistant is testing different link functions for predicting Over60. Your assistant notes that some model predictors are highly statistically significant with certain link functions but not with others.

(d) (*3 points*) Explain how changing the link function in the GLM impacts the model fitting and how this can impact predictor significance.

ANSWER:

Task 12 *(7 points)*

Your assistant mentions that using latitude and longitude for each service call would allow the mapping of each call to a zip code. By using publicly available census information, the data by zip code could be combined with information such as average age, predominant race, and average household income.

(a) (*1 point*) Define proxy variable.

ANSWER:

(b) (3 *points*) Evaluate your assistant's recommendation for any potential legal or ethical concerns including whether proxy variables should be used in this project.

ANSWER:

Your assistant states that the values for latitude and longitude are too granular and proposes that the data be grouped for modeling. Your assistant groups the data by splitting the ranges of both latitude and longitude into 20 equally spaced bins and creating factor variables Latitude_Binned and Longitude_Binned. For each combination of Department, year, month, Latitude_Binned and Longitude_Binned the average Time.to.resolution and the total count is stored in variables Ave.Time.to.resolution and call_count.

Using this grouped data, your assistant then models the Ave.Time.to.resolution using two Poisson regression models, Poisson.1 and Poisson.2. The code for these models is provided.

(c) (3 *points*) Assess the differences between the two models, including fitted parameters, coefficient estimates, goodness of fit.

ANSWER: