

## Exam PA April 2026 Project Statement

### General Information for Candidates

This examination has 10 tasks numbered 1 through 10 with a total of 70 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 described below. **We recommend that you read the business problem and data dictionary to learn additional context about each task.** 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. You may use Excel for calculation for any of the tasks, but all answers must be submitted in the Word document. *If you upload the Excel document, it will not be looked at by the graders.* Neither R nor RStudio are available.

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.

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.

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.

## Business Problem

*You work for an actuarial consulting firm that applies predictive modeling techniques to business problems in Los Angeles County, U.S. Los Angeles County has a densely populated inner city with a low-income population and less populous suburbs, which generally have higher-income populations.*

*You are currently working on projects to bring products and services to people in Los Angeles County who do not own cars. To support your work, you are using a dataset with census, survey, and geospatial data for each census tract in Los Angeles County.<sup>1</sup>*

***We recommend that you review the data dictionary to see additional information about each variable.***

---

<sup>1</sup> County of Los Angeles Open Data

## Data Dictionary

Variable	Data Type / Range / Example	Description
tract	Character Example: 0637113231	Unique identifier for census tract. Each tract is a subdivision of Los Angeles County.
no_vehicle	Numeric Range: 0 - 1,268	Number of households surveyed which have no vehicle.
no_vehicle_universe	Numeric Range: 29 - 5,202	Total number of households surveyed.
no_vehicle_pct	Numeric Range: 0 - 82.1	Percentage of households with no vehicle: $100 * \text{no\_vehicle} / \text{no\_vehicle\_universe}$
sup_dist	Character Examples: District 1, District 2	Supervisorial District. Each district has a representative on the Los Angeles County Board of Supervisors
csa	Character Examples: Los Angeles - Hollywood, City of Inglewood	Combined Statistical Area (CSA). A CSA is a population center and adjacent communities.
spa	Character Examples: SPA 2 - San Fernando, SPA 6 - South	Service Planning Area. SPAs are used in providing health services to meet the needs of each area.
Shape__Area	Numeric Range: 48,365 - 16,086,908,851	Land area of the census tract in square feet.
Shape__Length	Numeric Range: 2,815 - 915,243	Perimeter of the census tract in feet.
percent_hs	Numeric Range: 30.1 - 100	Percentage of people in the tract who have graduated high school.
percent_bachelors	Numeric Range: 1.3 - 91.7	Percentage of people in the tract who have a bachelor's degree.
disabled_pct	Numeric Range: 1.1 - 56	Percentage of people in the tract who are disabled.
income_percap	Numeric Range: 9,284 - 214,838	Income per capita for the tract.
med_hh_income	Numeric Range: 7,000 - 250,001	Median Household Income in the census tract. Values greater than \$250,000 are entered as \$250,001

ami_category	Character Examples: Low Income, Moderate Income	Median Household Income category.
--------------	---	-----------------------------------

Task 1 (7 points)

Your actuarial consulting firm has been approached by an economic development organization that is looking to invest in various parts of Los Angeles. They are considering options that enhance the walkability of local communities and are looking for recommendations on potential areas in which to invest.

- (a) (1 point) List four common characteristics of a predictive modeling problem.

**ANSWER:**

---

Your analyst has developed the following table in support of the client request:

District	Count of Census Tracts	No Vehicle Count	No Vehicle Percentage	Disabled Count	Disabled Percentage
District 1	497	68,534	11.1%	214,398	11.2%
District 2	506	71,263	10.8%	223,645	11.4%
District 3	557	63,142	8.3%	211,960	10.5%
District 4	487	44,878	6.8%	210,634	10.4%
District 5	448	39,990	6.1%	203,690	11.0%

- (b) (2 points) Identify the area of predictive analytics (descriptive, predictive, or prescriptive analytics) that the table represents and explain whether it addresses the client's needs. Do not directly discuss the data presented in the table.

**ANSWER:**

---

The client has also provided you with survey information from residents of various neighborhoods being considered for potential development opportunities. Included in this survey information are additional variables capturing job codes, home values, marital status, and religious status.

They have recommended that you build models that include all this additional information.

- (c) (2 points) Critique your client's recommendation of including all the additional variables in your model.

**ANSWER:**

The survey includes a comment box where residents can enter suggestions for future projects.

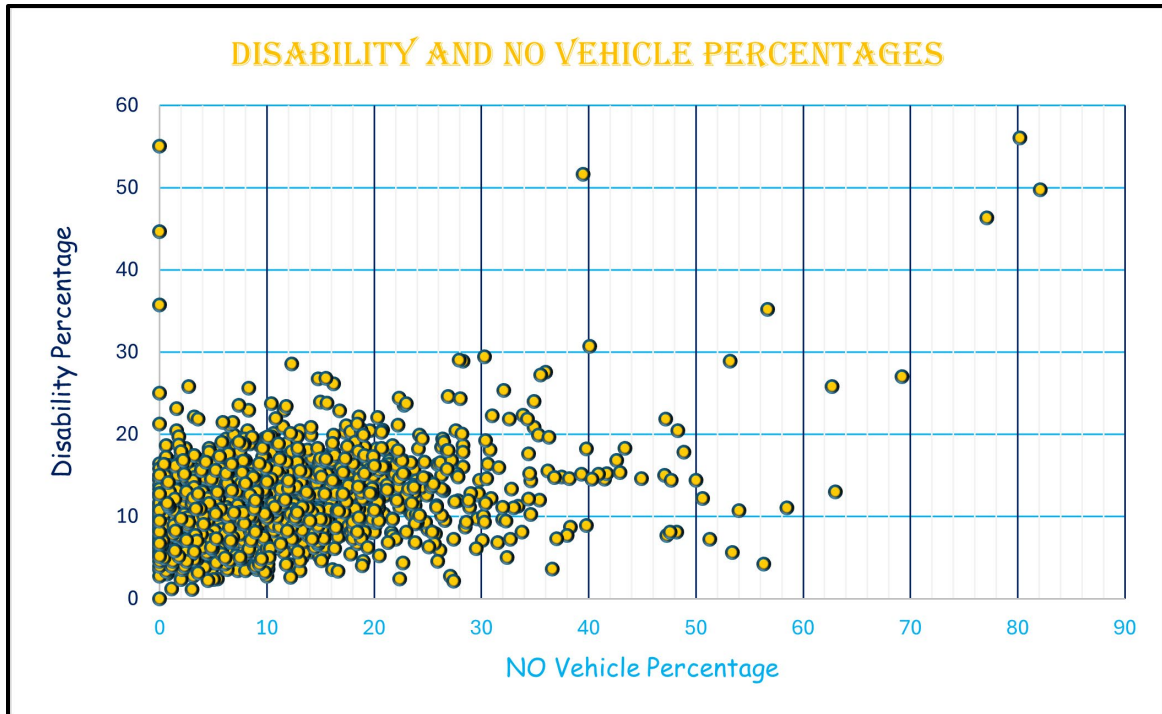
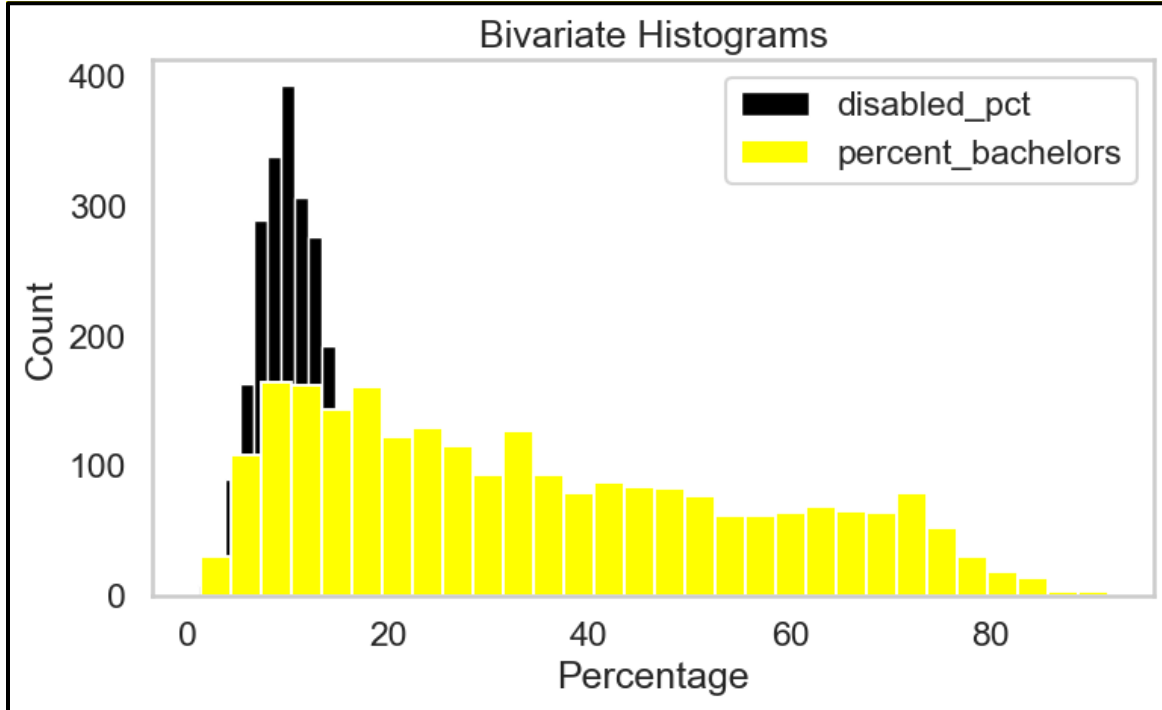
**(d)** (2 points) Identify what kind of data these comments represent and state the pros and cons of including this in your analysis.

**ANSWER:**

---

Task 2 (10 points)

You have asked your assistant to develop bivariate or multi-variate visualizations connecting the rate of vehicle ownership with either education level or disability. They have provided the following graphs as part of this assignment:



(a) (3 points) Critique each graph with regard to data visualization design guidelines, identifying at least two good and at least two bad elements in each graph.

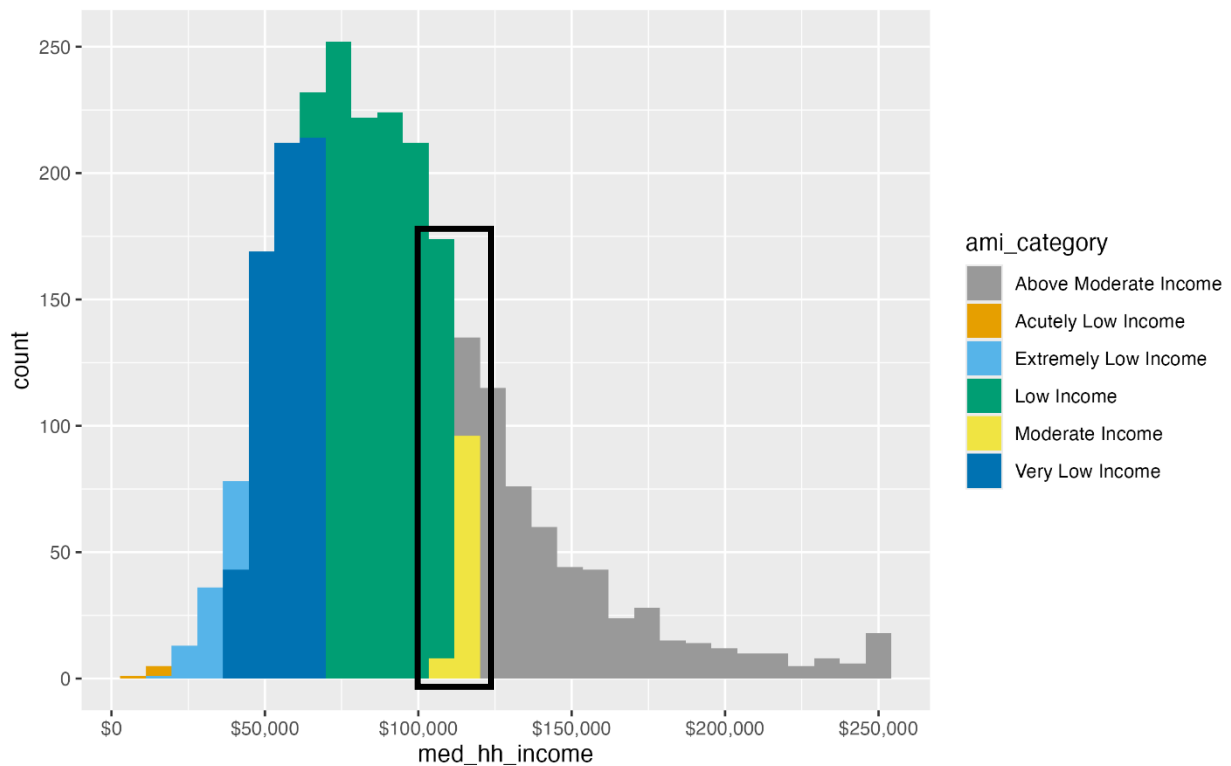
ANSWER:

	Good Elements	Bad Elements
Graph 1		
Graph 2		

(b) (2 points) Critique the two graphs with regard to the analysis you requested.

ANSWER:

You want to understand the income-related variables in the dataset better and ask your assistant to investigate the relationship between **med\_hh\_income** and **ami\_category**. They create the histogram below:



(c) (3 points) Interpret the following aspects of the histogram:

- i. The overall relationship between these two variables
- ii. The section of the histogram boxed in black, where there are bars for low, moderate, and above moderate income
- iii. The relatively high density around \$250,000

**ANSWER:**

- i.
- ii.
- iii.

---

**(d)** (2 points) Explain the problems that may arise from using both variables in a GLM, including considerations of model interpretability and performance.

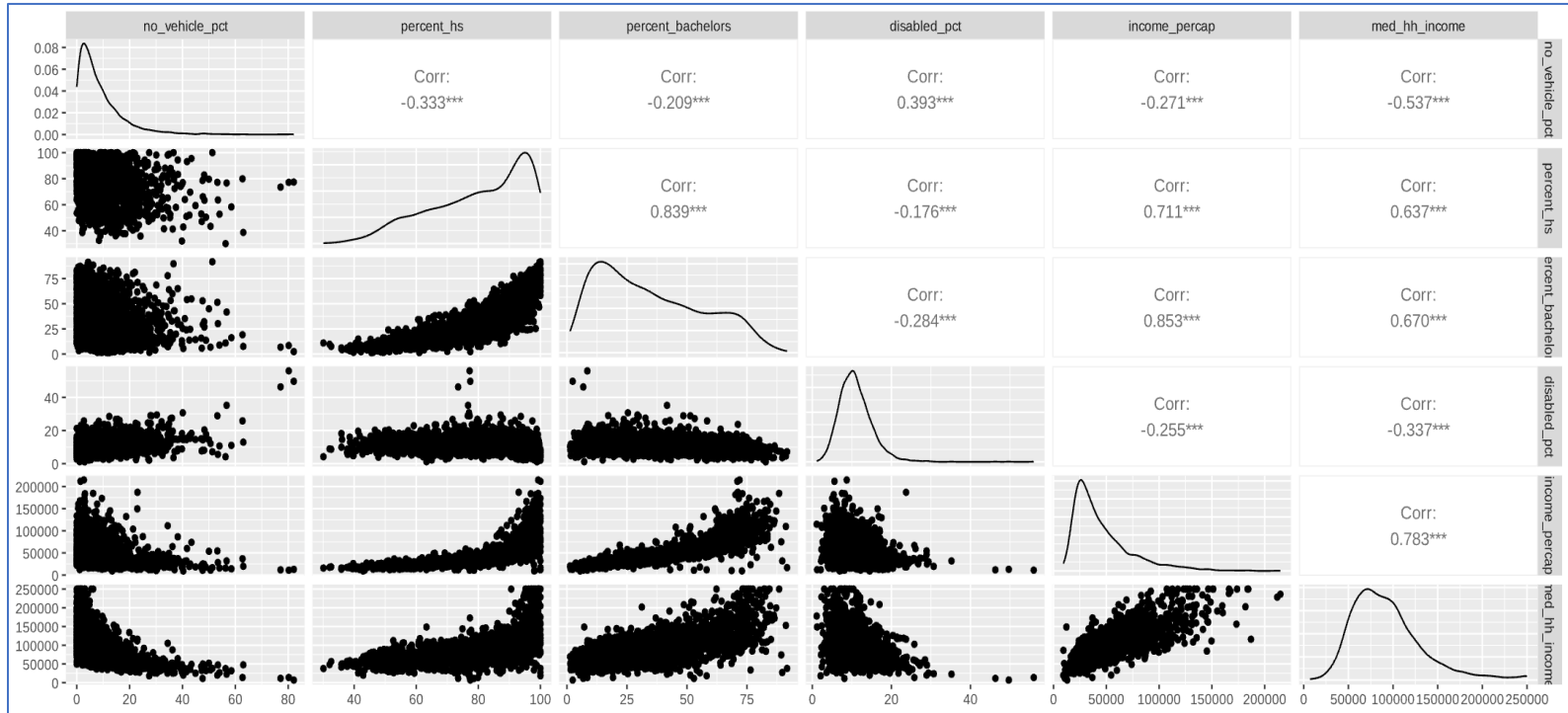
**ANSWER:**

---

Task 3 (6 points)

The Los Angeles County Metropolitan Transportation Authority has asked your firm to cluster census tracts into a small number of neighborhood types to identify transit-vulnerable area clusters. *Transit-vulnerable* is defined as having low socioeconomic status, having poor access to private vehicles, and being more reliant on public transit.

You are provided with a pairs plot along with correlations of the variables selected for this study.



(a) (2 points) Explain the importance of standardizing these six variables before fitting a K-means model.

ANSWER:

---

You fit a 3-cluster K-means model on the standardized variables. The table below shows cluster means on the original scale.

cluster_k3	n_tracts	no_vehicle_pct	percent_hs	percent_bachelors	disabled_pct	income_percap	med_hh_income
<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	890	5.461124	94.71326	59.23528	9.140674	73369.86	127711.58
2	367	23.487738	68.03515	22.07466	15.867847	27583.96	54151.57
3	1196	6.646488	72.57333	21.04933	10.791388	30689.10	80167.38

(b) (2 points) Identify which cluster is most transit-vulnerable. Describe the characteristics of the chosen cluster.

ANSWER:

Your assistant proposes adding **spa**, a categorical variable with 8 levels, to the clustering analysis using the following approach:

- Create seven one-hot variables to replace **spa**: **spa\_1, spa\_2, ..., spa\_7**.
- Standardize all variables.
- Fit K-means using the original standardized variables plus the seven standardized variables **spa\_1, spa\_2, ... spa\_7**.

Assume K-means uses Euclidean distance on the standardized predictors.

(c) (2 points) Critique your assistant's approach of adding the categorical variable, including considerations of how the Euclidean distance calculation will be impacted.

ANSWER:

Task 4 (7 points)

You would like to understand the age distribution of residents across different census tracts. You begin by applying principal components analysis (PCA) to population age data to get more informative variables.

- (a) (2 points) Compare and contrast K-Means Clustering and PCA as unsupervised learning approaches.

**ANSWER:**

---

The data relating to age is split into 14 age buckets, each with a count of the number of individuals. **POP22\_AGE\_0\_4** is the count of people between age 0 and 4, **POP22\_AGE\_5\_9** is the count of people between age 5 and 9, etc. A sample from the data is shown here for reference where each row represents one census tract:

<b>POP22_AGE_0_4</b>	<b>POP22_AGE_5_9</b>	<b>POP22_AGE_10_14</b>	<b>POP22_AGE_15_17</b>	<b>POP22_AGE_18_19</b>	<b>Table continues...</b>
209	251	219	130	90	
206	143	174	97	88	
131	148	130	106	65	
158	199	205	102	68	

You perform PCA on the standardized data and receive the following coefficients for the first three principal components:

Column Name	PC1	PC2	PC3
POP22_AGE_0_4	0.31	-0.17	-0.11
POP22_AGE_5_9	0.30	-0.16	-0.01
POP22_AGE_10_14	0.30	-0.16	0.04
POP22_AGE_15_17	0.29	-0.14	0.08
POP22_AGE_18_19	0.10	-0.27	0.65
POP22_AGE_20_24	0.20	-0.33	0.44
POP22_AGE_25_29	0.27	-0.21	-0.26
POP22_AGE_30_34	0.25	-0.13	-0.38
POP22_AGE_35_44	0.30	-0.05	-0.27
POP22_AGE_45_54	0.32	0.06	-0.05
POP22_AGE_55_64	0.31	0.19	0.04
POP22_AGE_65_74	0.28	0.35	0.10
POP22_AGE_75_84	0.24	0.48	0.16
POP22_AGE_85_100	0.19	0.51	0.17

Your assistant suggests that since all the coefficients in PC1 are positive, we can remove PC1 from our analysis and proceed with PC2, PC3.

(b) (2 points) Critique your assistant's statement.

**ANSWER:**

You are provided with the following information about the **POP22\_AGE\_0\_4** variable:

Mean	158.835
Variance	9518.094

(c) (3 points) Using the PC coefficients above, estimate **POP22\_AGE\_0\_4** for a point with the following weights in the original, unstandardized units:

PC1 Weight	2.194
PC2 Weight	1.234
PC3 Weight	0.099

Show your work.

**ANSWER:**

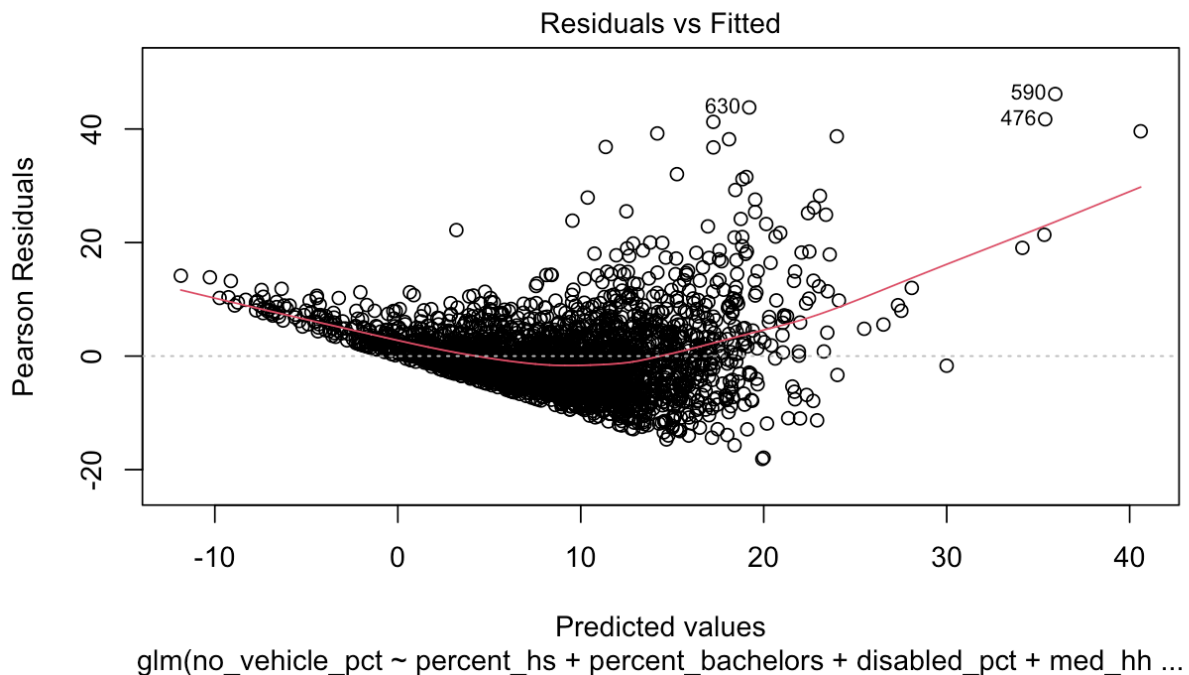


Task 5 (5 points)

Your manager asks you to build a linear model to predict car ownership using demographic and economic variables. You fit the following model:

```
m0 <- glm(  
  no_vehicle_pct ~ percent_hs + percent_bachelors + disabled_pct  
                + med_hh_income,  
  family = gaussian(link = 'identity'),  
  data = mdat,  
  na.action = na.omit  
)
```

A residual plot for this model is shown below:



- (a) (2 points) Critique the choice of the Gaussian family with identity link for modeling **no\_vehicle\_pct**.

ANSWER:

- (b) (2 points) Recommend a more appropriate GLM family and link function. Justify your recommendation based on the plot and characteristics of the response variable.

ANSWER:

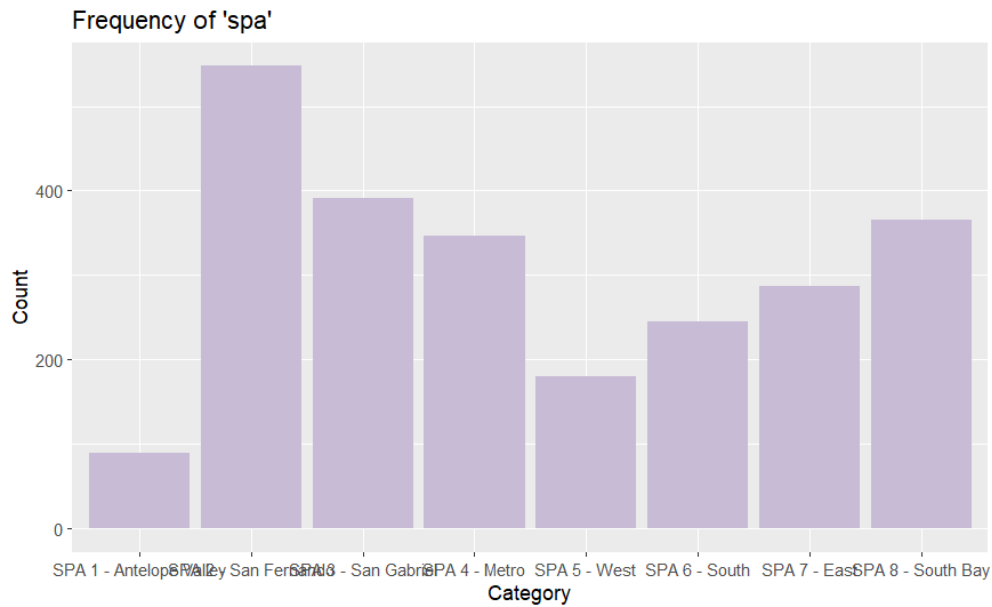
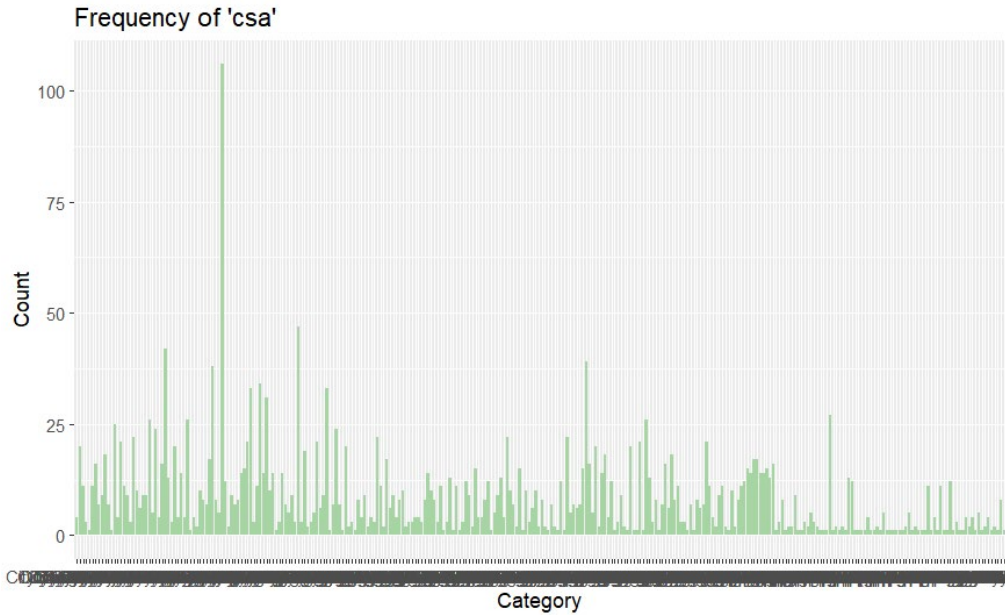
- 
- (c)** (1 point) Explain briefly how interpretation of coefficients changes under your recommended GLM family and link function compared to the GLM with Gaussian family and identity link.

**ANSWER:**

---

Task 6 (7 points)

Your assistant is investigating geographic data to understand how it could be used in models to predict **no\_vehicle\_pct**. They produce the following bar charts for the geographic factor variables **csa** and **spa**:

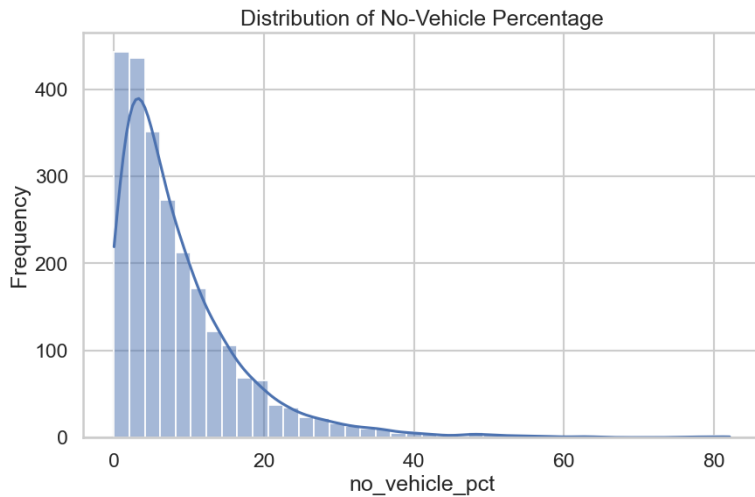


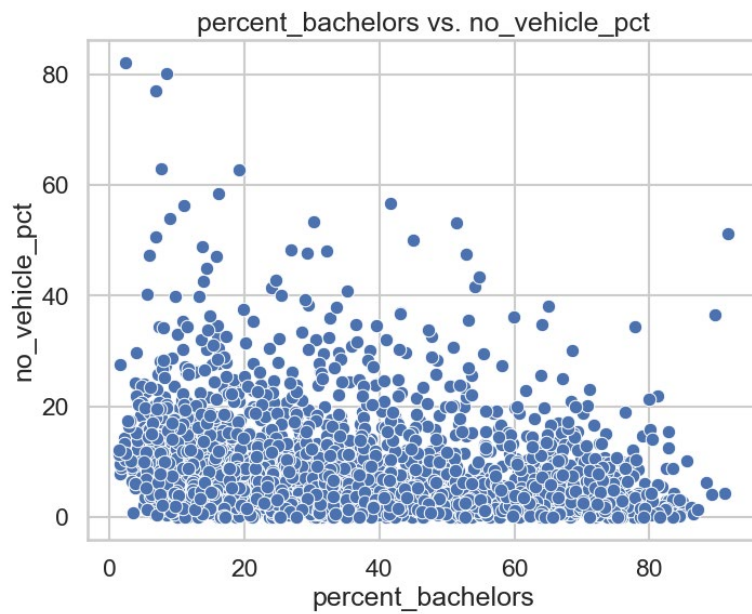
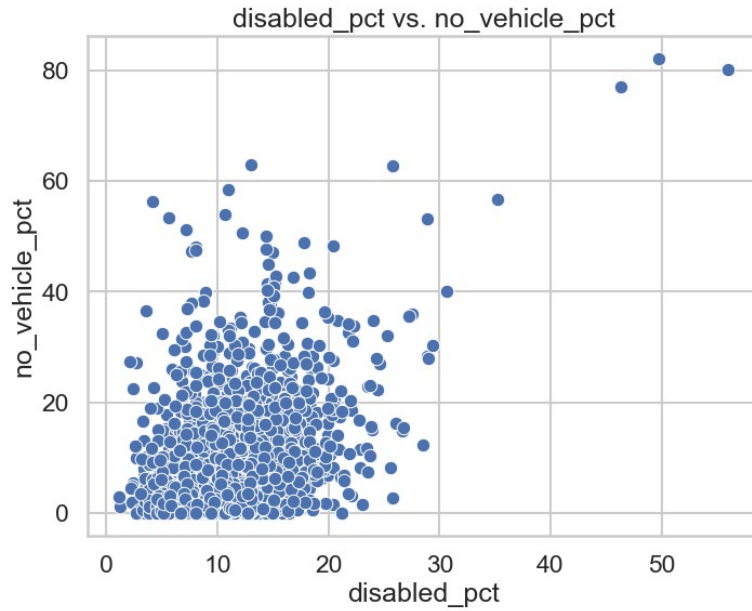
(a) (2 points) Describe one advantage and one disadvantage of using each variable as a predictor in a model.

ANSWER:

	Advantage	Disadvantage
<b>csa</b>		
<b>spa</b>		

After reviewing the geographic variables, you ask your assistant to review the distribution of the response variable, **no\_vehicle\_pct**, and perform univariate analysis on the remaining numeric variables. Your assistant has produced additional statistical summaries and charts to support the analysis:





- (b) (3 points) Describe the distribution of **no\_vehicle\_pct** based on the graphs above and interpret the relationship between this target variable and the other numeric variables.

**ANSWER:**

Your co-worker raises a concern that including all the numeric variables **percent\_hs**, **percent\_bachelors**, **disabled\_pct**, and **med\_hh\_income** would cause collinearity issues in a GLM.

(c) (2 points) Critique the statement from your co-worker.

**ANSWER:**

---

Task 7 (9 points)

You are working with a coffee shop chain in Los Angeles County that plans to open stores with walk-up windows in neighborhoods with low car ownership rates. They provided the following statement:

*To increase profits, we need more high-income customers who don't own cars.*

(a) (2 points) Write down two questions to ask the company to clarify the business issue.

**ANSWER:**

- i. [Question 1]
  - ii. [Question 2]
- 

You fit two Generalized Linear Models (GLMs) to predict the variable **no\_vehicle**.

- Model 1: A Poisson GLM.
- Model 2: A Negative Binomial GLM.

The models use the following variables:

- **income\_10k**: Median income in \$10,000s (10k = 1)
- **disabled\_pct**: Percentage of population with disabilities (10% = 10)
- **sup\_dist**: Supervisory District (Reference = District 1)
- **log(no\_vehicle\_universe)**: Log of the total number of households in the area, used as an offset in both models

Outputs for Model 1 and Model 2 are shown below:

### Model 1 (Poisson)

```
Call:
glm(formula = no_vehicle ~ income_10k + disabled_pct + sup_dist +
  offset(log(no_vehicle_universe)), family = poisson(link = "log"),
  data = train)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.270067  0.0106140 -119.654 < 2e-16 ***
income_10k    -0.1678680 0.0008563 -196.030 < 2e-16 ***
disabled_pct    0.0272979 0.0004028  67.774 < 2e-16 ***
sup_distDistrict 2 -0.0765799 0.0064771 -11.823 < 2e-16 ***
sup_distDistrict 3  0.0323453 0.0067451   4.795 1.62e-06 ***
sup_distDistrict 4 -0.2760543 0.0073312 -37.655 < 2e-16 ***
sup_distDistrict 5 -0.2883799 0.0075855 -38.017 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 157954 on 1718 degrees of freedom
Residual deviance: 78026 on 1712 degrees of freedom
AIC: 88303

Number of Fisher Scoring iterations: 5
```

### Model 2 (Negative Binomial)

```
Call:
glm.nb(formula = no_vehicle ~ income_10k + disabled_pct + sup_dist +
  offset(log(no_vehicle_universe)), data = train, init.theta = 1.727027102,
  link = log)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.690850  0.088867 -19.027 < 2e-16 ***
income_10k    -0.128189  0.005269 -24.331 < 2e-16 ***
disabled_pct    0.030659  0.004563   6.719 1.83e-11 ***
sup_distDistrict 2  0.016376  0.058419   0.280  0.7792
sup_distDistrict 3  0.099816  0.058242   1.714  0.0866 .
sup_distDistrict 4 -0.244548  0.059427  -4.115 3.87e-05 ***
sup_distDistrict 5 -0.345972  0.061548  -5.621 1.90e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.727) family taken to be 1)

Null deviance: 3079.3 on 1718 degrees of freedom
Residual deviance: 1975.4 on 1712 degrees of freedom
AIC: 18750

Number of Fisher Scoring iterations: 1

Theta: 1.7270
Std. Err.: 0.0589

2 x log-likelihood: -18734.3900
```

**(b)** (1 point) Describe two assumptions specific to using a Poisson distribution in a GLM.

**ANSWER:**

---

**(c)** (2 points) Explain why  $\log(\text{no\_vehicle\_universe})$  is included as an offset in the models rather than as a predictor variable.

**ANSWER:**

---

**(d)** (1 point) Calculate the predicted count of households with no vehicle for a district with the following characteristics using Model 2 (Negative Binomial):

- Income: \$50,000 (**income\_10k** = 5)
- Disabled Percentage: 10% (**disabled\_pct** = 10)
- Location: District 3 (**sup\_distDistrict 3**)
- Total Households: 1,000 (**no\_vehicle\_universe** = 1000)

Show your work.

**ANSWER:**

---

Your manager asks you to build Poisson and Gaussian GLMs that include **percent\_bachelors** as a predictor. The GLM specifications and performance on a test dataset are provided below:

Poisson GLM	Call: glm(formula = no_vehicle ~ med_hh_income_10k + percent_bachelors + disabled_pct, family = poisson(link = "log"), data = train, offset = log(no_vehicle_universe))
Gaussian GLM	Call: glm(formula = no_vehicle_pct ~ med_hh_income_10k + percent_bachelors + disabled_pct, family = gaussian(link = "identity"), data = train)

	Residual_Deviance	AIC	MAE	RMSE
Poisson_GLM	72441.84	82761.05	4.006	5.723
Gaussian_GLM	79683.83	11471.60	5.014	7.100

- (e) (2 points) Recommend a model to the client based on performance metrics. Justify your recommendation.

**ANSWER:**

- (f) (1 point) Recommend a next step to assure the company that the model can be trusted when used on real life data.

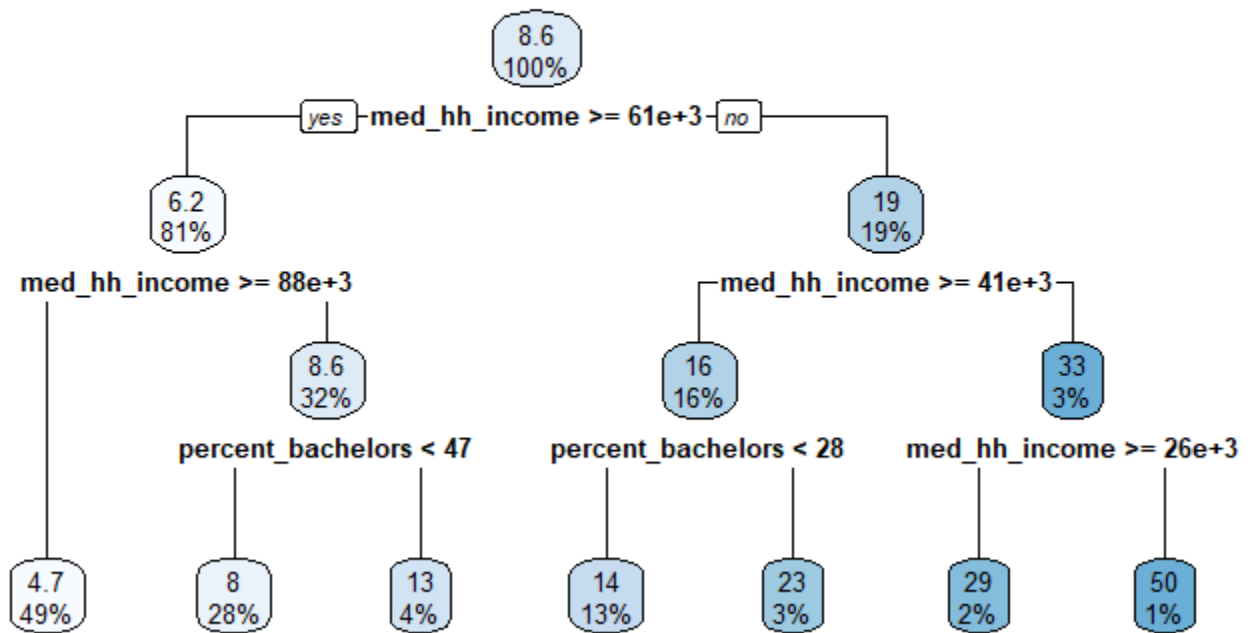
**ANSWER:**

Task 8 (6 points)

Your manager would like you to investigate whether trees can provide insight into factors related to car ownership.

You have fit a regression tree with **no\_vehicle\_pct** as the target variable and **med\_hh\_income**, **disabled\_pct**, **percent\_bachelors**, **percent\_hs**, and **income\_percap** as predictor variables. The results of the tree are depicted below:

**Exhibit 1: Initial Regression Tree**



(a) (2 points) Populate the table below with predictions from the tree:

ANSWER:

		percent_bachelors		
		0 – 28	28 – 47	> 47
med_hh_income	0 – 26,000			
	26,000 – 41,000			
	41,000 – 61,000			
	61,000 – 88,000			
	>88,000			

- (b) (1 point) Interpret how the impact of having a high percentage of bachelors degrees (>47%) differs between high-income and middle-income tracts.

ANSWER:

---

To increase the number of variables included, you decide to decrease the complexity parameter (cp) and now the tree contains more splits.

Regression tree:

```
rpart(formula = no_vehicle_pct ~ ., data = train, method = "anova",
      control = rpart.control(cp = 0.005, minbucket = 10))
```

variables actually used in tree construction:

```
[1] disabled_pct      income_percap      med_hh_income      percent_bachelors  percent_hs
```

Root node error: 127900/1719 = 74.403

n= 1719

	CP	nsplit	rel error	xerror	xstd
1	0.3223761	0	1.00000	1.00120	0.085923
2	0.0958693	1	0.67762	0.70260	0.057056
3	0.0397752	2	0.58175	0.62313	0.043781
4	0.0287452	3	0.54198	0.59855	0.043631
5	0.0246791	4	0.51323	0.60094	0.044015
6	0.0105465	5	0.48856	0.56944	0.042395
7	0.0095581	6	0.47801	0.56918	0.040955
8	0.0093560	7	0.46845	0.56857	0.041441
9	0.0089391	8	0.45909	0.56738	0.041437
10	0.0075548	9	0.45016	0.54276	0.037687
11	0.0074294	10	0.44260	0.54172	0.037679
12	0.0064832	11	0.43517	0.53928	0.037632
13	0.0050000	12	0.42869	0.52380	0.036278

- (c) (2 points) Identify the optimal number of splits according to the 1-Standard-Error Rule.

ANSWER:

- 
- (d) (1 point) Explain why a single decision tree may have poor model performance.

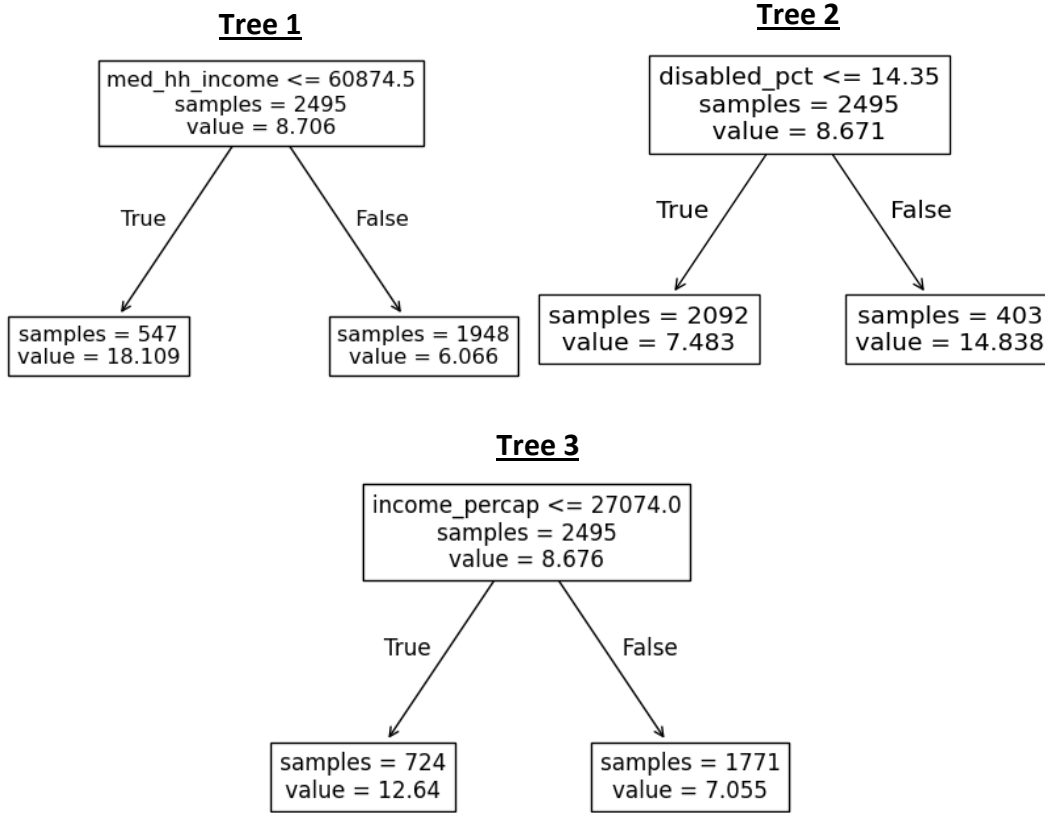
ANSWER:



Task 9 (8 points)

Your manager believes that ensemble tree models are a promising approach to modeling car ownership and asks you to build some ensemble models.

You start by building a simple a bagged tree model, which consists of the three trees below:



You are given three data points and for which trees they were used in the fitting process:

ID	no_vehicle_pct	disabled_pct	income_percap	med_hh_income	Tree 1	Tree 2	Tree 3
1	2.5	9.5	33932	109088	In		
2	15.7	9.4	27037	71938		In	In
3	11.6	8.2	43325	114550			In

- (a) (3 points) Calculate the Out-of-Bag MSE for the bagged model based on these three observations.

**ANSWER:**

You are also interested in considering other ensemble methods. Your assistant suggests boosting as an alternative to bagging.

**(b)** (1 point) Explain how the following hyperparameters are used in a boosted tree model

- i. The number of iterations,  $B$
- ii. The shrinkage parameter,  $\lambda$

**ANSWER:**

---

Based on client's feedback, they are interested in understanding segments where **no\_vehicle\_pct** is greater than 20%. You create a binary variable, **low\_vehicle\_own** as:

- **low\_vehicle\_own** = Yes if **no\_vehicle\_pct** > 20
- **low\_vehicle\_own** = No if **no\_vehicle\_pct** ≤ 20

Your assistant builds a classification decision tree. A confusion matrix from this model is below:

		Predicted	
		No	Yes
Actual	No	362	82
	Yes	9	38

**(c)** (2 points) Calculate accuracy, precision, sensitivity, and specificity from the confusion matrix. Show your work.

**ANSWER:**

- i. Accuracy
- ii. Precision
- iii. Sensitivity
- iv. Specificity

---

Your assistant observes that the variable **low\_vehicle\_own** is unbalanced between the two classes.

**(d)** (1 point) Explain your assistant's observation using the information provided.

**ANSWER:**

---

**(e)** (1 point) Describe two methods for addressing unbalanced data.

**ANSWER:**

---

### Task 10 (5 points)

You plan to have your assistant construct a linear regression model but are concerned that due to the high variable count in the dataset, there is a risk that the model will overfit to the data. You would like to consider modifications that reduce the risk of overfitting while retaining the linear model type.

(a) (2 points) Describe two advantages of regularized regression over Best Subset Selection (BSS).

**ANSWER:**

---

You begin by investigating subsets of the predictors that could be used to calibrate a model. To begin, you ask your assistant to incrementally select counts of variables to use and calibrate a standard linear regression on each of them using all observations in the dataset. The table below summarizes the performance of each model:

Model ID	Variables Used	RSS	AIC	BIC
1	5	156305.9	62.8	63.2
2	10	140944.6	56.8	57.6
3	15	132194.6	53.4	54.7
4	20	114901.0	46.6	48.3
5	25	97713.3	39.9	42.0

You are concerned that the AIC and BIC continue to decrease even though they are supposed to penalize for adding additional variables to the model training.

(b) (1 point) Explain why the RSS, AIC, and BIC continue to decrease as the number of variables used increases, even though AIC and BIC penalize for additional variables.

**ANSWER:**

---

You would like your assistant to try using Lasso regression. You think that performing a Lasso regression using an edge case for the lambda parameter could help better understand its impact.

You provide the following data points for your assistant to use:

ID	no_vehicle_pct	Shape_Area	Shape_Length	percent_hs	percent_bachelors	disabled_pct
1	12.9	12,329,806	14,766	90.1	27.5	11.4
2	0.7	29,088,657	30,028	82.1	33.7	13.1
3	5.4	7,532,651	12,505	79.7	31.4	15.7

4	14.8	3,874,399	8,938	71.7	23.7	26.7
5	21	3,090,538	9,751	78.5	19.1	18.1
6	2.6	27,694,822	28,701	91.8	37.2	16.4
7	14	67,867,979	41,887	88.5	25.2	14.0
8	10.2	12,811,195	16,199	90.9	29.2	15.2
9	4.7	179,333,612	27,119	93.1	44.7	9.6
10	1.3	5,257,057	9,953	77.8	20.9	7.5
11	1.7	139,121,772	63,443	82.8	37.0	14.7
12	2.4	21,327,765	24,166	93.7	35.7	12.1
13	8.5	13,323,876	16,273	82.0	17.8	13.2
14	12.5	21,040,409	18,173	79.2	26.3	12.1

(c) (2 points) Fit a Lasso regression with **no\_vehicle\_pct** as the target variable and a lambda approaching infinity. Provide the resulting model coefficients in the table below. Explain how you arrived at your answer.

**ANSWER:**

*Coefficients:*

Variable	Coefficient
Intercept	
Beta(Shape__Area)	
Beta(Shape__Length)	
Beta(percent_hs)	
Beta(percent_bachelors)	
Beta(disabled_pct)	

*Explanation:*

---