Case Study Hints

I. Accessing Data Hints:

Task

- Open the AdventureWorks.db file in SQLite or DB Browser for SQLite.

Hints

- Download the AdventureWorks.db file from Blackboard
- Open DB Browser for SQLite.
- Open the Database in DB Browser for SQLite.

II. Exploring Data Hints:

Tasks

- Locate the required attributes in their respective tables using the Schema diagram and the Data Dictionary.
- Explore the following tables and make note of any joins needed using the information from the Requirements document.
    i. customer
    ii. person
    iii. salesorderheader
    iv. salesorderdetail
    v. product

Hints

- Find these tables in the schema, and note the available columns.
- Find these tables in the data dictionary and note the columns and their attributes
- Explore these tables in DB Browser for SQLite. Make note of any inconsistencies (e.g., column name mismatches, columns that appear in the database but not in the schema or dictionary).
- Think about how these tables and attributes might help you to determine who might buy a bike, how you could define which customers purchased a bike, and how you might be able to join these tables together if necessary.

III. Preparing Data Hints:

Tasks

- Gather customerid and territoryid from the customer table
- Gather TotalPurchaseYTD, DateFirstPurchase, MaritalStatus, YearlyIncome, Gender, TotalChildren, NumberChildrenAtHome Education, Occupation, HomeOwnerFlag, NumberCarsOwned, and CommuteDistance from the person table
- Calculate the customer's age by determinine the number of years between October 31st, 2007 and the customer's birthdate.
- Determine which customers purchased a bicycle by using the salesorderheader and salesorderdetail tables and the productsubcategoryid.
- Merge these pieces of information together and flag each customer what purchased a bike with a "1" and a "0" otherwise.

- Export the final table as a .csv so you can read it into R.

Hints

- Write a query that selects customerid, territoryid, TotalPurchaseYTD, DateFirstPurchase, MaritalStatus, YearlyIncome, Gender, TotalChildren, NumberChildrenAtHome Education, Occupation, HomeOwnerFlag, NumberCarsOwned, and CommuteDistance from the customer table and the person table joined together
- In the same query calculate age by subtracting the customer's birthdate from October 31st, 2007. *Hint:* Use the substr() function to format the birthdate as yyyy-mm-dd and substract that value from 2007-10-31.
- Write a query that pulls salesorderid, customerid, productid, and productsubcategoryid from salesorderheader and salesorderdetail and product tables where the productsubcategoryid is equal to 1, 2, or 3 (Why?)
- You may want to create tables from these queries and store them in the database instead of doing all of this in one query.
- Write a query that gets all of the predictor variables from your first query, and creates an additional column that uses a CASE statement to flag a bike buying customer using your second query.
- Create a table for your final dataset and use FILE >> EXPORT >> Table(s) as CSV file. . . in DB Browser for SQLite.

IV. Analyzing Data Hints

Tasks

- Read your .csv into R.
- Partition your data into 80% training and 20% testing data sets (set the seed to 12345).
- Fit a logistic regression model.
- Fit a decision tree model.
- Fit a classification method of your choosing by choosing one of the available models listed here.
- Choose the "best" model.

Hints

- You might find the read_csv function in the readr library helpful for reading these data into R.
- Load the caret library
- Use the createDataPartition() function to partition the data into training and test sets.
- Use the train() function in caret using the glm method
- Use the train() function in caret using the rpart method
- Choose one of the available models listed here, and place the keyword in the method argument of the train() function.
- Use the confusionMatrix() function in caret to evaluate each model's fit and select the "best" model as the one with the lowest error rate (highest accuracy rate).

V. Exporting Reports Hints:

Tasks

1. Follow the steps outlines here to create a story in Tableau Public:
    a. Discuss the goal of the overall project
    b. Add a sheet that discusses the data sources and how the variables are measured. Include a table or graphs of summary statistics.
    c. Add a sheet that discusses the different types of analysis that were performed
    d. Add a sheet that evaluates the three models against each other using appropriate tables or graphs. If you're unsure what might constitute a suitable evaluation, check the book or this article.
    e. Showcase your choice of cut off.
    f. Finally, make a final recommendation.

Hints

- The goal is outlined in the Requirements page. Add a sheet that discusses the goal. Text id fine, add a picture of a bike or something to make it fancy. Consider how the 1854 Cholera Outbreak is organized.
- The data are discussed in the data dictionary, schema, and the text on Blackboard under the AdventureWorkds data folder. Add a sheet that outlines pertinent details of the data so that users get a sense of what you have to work with.
- Add a sheet that compares and contrasts the methods you used to fit. Suitable discussions are available in the book and in the like. Use the functionality of tableau to make this information consumable.
- Add a sheet that evaluates the models against each other. Use preattentive attributes to highlight your chosen model. If you're unsure about the types of things that help evaluate classification models, check this article out. If you want to use a ROC curve but don't know how to create one in Tableau, google it! Then create an excel file that has the requisite information formatted as necessary in order to convey the comparison you want.
- Showcase in either a graph or table your choice of cut-off; that is, the probability you indicate to determine which customers will buy a bike and which one's will not. Tell us what the error associated with your cutoff is.
- Make a final recommendation. What proportion of customers would you email to maximize bike buying?