AdventureWorks Case Study: Assignment Guide

Below is a suggested process for solving the business problem that is based on the programming process. Be aware that there are multiple solutions to this problem and that you do not need to follow the steps below.

Access Data

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

Explore Data

1. Locate the required attributes in their respective tables using the Schema diagram and the Data Dictionary.
2. Explore the following tables and make note of any joins needed using the information from the Requirements document.
    a. customer
    b. person
    c. salesorderheader
    d. salesorderdetail
    e. product

Prepare Data

1. Gather customerid and territoryid from the customer table
2. Gather TotalPurchaseYTD, DateFirstPurchase, MaritalStatus, YearlyIncome, Gender, TotalChildren, NumberChildrenAtHome Education, Occupation, HomeOwnerFlag, NumberCarsOwned, and CommuteDistance from the person table
3. Calculate the customer's age by determinine the number of years between October 31st, 2007 and the customer's birthdate.
4. Determine which customers purchased a bicycle by using the salesorderheader and salesorderdetail tables and the productsubcategoryid.
5. Merge these pieces of information together and flag each customer what purchased a bike with a "1" and a "0" otherwise.
6. Export the final table as a .csv so you can read it into R.

Analyze Data

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

Report Data

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.