Customer Research & Insights – Analytics Professional in Training Assignment

Background

TransLink conducts an ongoing survey of its bus, SeaBus and SkyTrain riders to measure how satisfied they are with the transit service that TransLink provides. Interviews are conducted daily, throughout the year, with adults who have used public transit, that is, the bus, SeaBus, or SkyTrain in the past 30 days.

Respondents are asked about their use of transit (frequency, modes used, trip purposes, time of day when they've travelled), followed by a series of questions about their "satisfaction" with the service quality at the transit system level and for each of the transit modes (SeaBus, SkyTrain and Bus). Results are tabulated and reported out on a quarterly basis.

The most recent copies of the report, which also describes the study method, is available on the TransLink website: Corporate Reports | TransLink under the heading of Customer Service Performance Reports.

Assignment

The Vice-President (VP) responsible for overseeing TransLink's customer experience program wants to better understand what factors of the transit service are important to customers, driving their overall satisfaction with the service, as well as their loyalty to the service. The VP comes to you asking if these relationships can be modelled using the Bus, SeaBus, SkyTrain Customer Service Performance study data.

After discussing the Vice-President's request with your manager, you agree to analyze the Bus, SeaBus, SkyTrain dataset and then develop a model to understand the nature and strength of the relationships between the service attributes and customer loyalty. Here is the list of tasks that you and your manager discussed:

Tasks:

1. Prepare a descriptive analysis of the dataset using your statistical software of choice. Produce a summary of the key variables of interest. Identify any issues with the data, how do you plan to address them in order to use them for modelling purposes, and why you decided on that solution.
2. Develop a loyalty metric for modelling purposes. The loyalty literature suggests that there are a number of ways to define "loyalty", both in terms of customers' behaviour (e.g., the number of times a customer uses a service, the number of years a person has been using the service) and attitudes (e.g., intent to repurchase and likelihood to recommend the service). Define a loyalty metric based on the dataset. **Explain the pros and cons of your decision.**
3. Describe how you would model the relationships between the individual service attributes at the transit system level and those measured at the transit mode level as well as how those would be modelled to the customer loyalty measure that you developed in Task #2. Assume that

you are asked to model these relationships using **traditional multivariate analysis techniques (i.e., linear modelling techniques).** Explain the likely strengths and weaknesses of the model and your reason for choosing it. Then estimate the model and report on the results including any strengths and weaknesses of the approach, now that you have generated the empirical model.

4. Similar to Task #3, describe, develop and estimate a model measuring the relationships between the system and mode attributes as well as customer loyalty, except using **a data mining technique (i.e., non-linear or deep learning model)**. Describe your reasons for the approach you have chosen and explain, and if you have time, contrast the results of the model relative to the model you developed in Task #3.
5. Other considerations: identify any other variables/data that you would consider incorporating into the model that could either be collected by modifying the existing survey or through separate datasets.

Deliverables:

- Summarize your finding from Tasks 1 to 5 in a PowerPoint style report. Ideally the presentation should not be longer than 10-15 minutes.
- Submit any supporting materials—e.g., software output, including scripts, etc.

Data and Documents Supplied:

- Data (InternAssignmentData4.csv): The dataset is provided as a *.csv file containing 10,113 interviews with adult customers who have taken transit at least once in the past 30 days, at the time of the interview. A total of 123 variables are included.
- Data Dictionary (InternAssignmentDataDictionary.htm): the data dictionary describes includes the variable names, variable name labels, variable values and the value labels.
- Questionnaire (Bus-SB-ST Questionnaire-Intern): the questionnaire provided is from 2004 and most closely reflects the composition of the dataset.

Tips:

- Review the data dictionary and questionnaire carefully. The latter will inform how best to structure your initial data analysis.
- In order to reduce the burden on the respondent, not all respondents are asked to rate all the services that they said they used. This has an impact on the number of bus routes that they evaluate. For example, they may have taken 3 routes as a part of their most recent trip on the bus, but they are not asked to rate all 3 if they have already rated other modes. Instead they are randomly asked to rate one of the 3 bus routes. Review the instructions in the questionnaire to better understand the data structure.
- Try to complete as much of each task as possible within the time frame.