Download the dataset (in csv format) from the following link:

https://www.kaggle.com/jahias/microsoft-adventure-works-cycles-customer-data

The Marketing department of Adventure Works Cycles wants to increase sales by targeting specific customers for a mailing campaign. The company's database contains a list of past customers and a list of potential new customers. By investigating the attributes of previous bike buyers, the company hopes to discover patterns that they can then apply to potential customers. They hope to use the discovered patterns to predict which potential customers are most likely to purchase a bike from Adventure Works Cycles.

Part I: Based on Feature Selection, Cleaning, and Preprocessing to Construct an Input from Data Source

(a) Examine the values of each attribute and Select a set of attributes only that would affect to predict future bike buyers to create your input for data mining algorithms. Remove all the unnecessary attributes. (Select features just by analysis).
(b) Create a new Data Frame with the selected attributes only.
(c) Determine a Data value type (Discrete, or Continuous, then Nominal, Ordinal, Interval, Ratio) of each attribute in your selection to identify preprocessing tasks to create input for your data mining.

Part II: Data Preprocessing and Transformation

Depending on the data type of each attribute, transform each object from your preprocessed data.

Use all the data rows (~= 18000 rows) with the selected features as input to apply all the tasks below, do not perform each task on the smaller data set that you got from your random sampling result.

(a) Handling Null values
(b) Normalization
(c) Discretization (Binning) on Continuous attributes or Categorical Attributes with too many different values
(d) Standardization/Normalization
(e) Binarization (One Hot Encoding)

Part III: Calculating Proximity /Correlation Analysis of two features

Make sure each attribute is transformed in a same scale for numeric attributes and Binarization for each nominal attribute, and each discretized numeric attribute to standardization. Make sure to apply a correct similarity measure for nominal (one hot encoding)/binary attributes and numeric attributes respectively.

(a) Calculate Similarity in Simple Matching, Jaccard Similarity, and Cosine Similarity between two following objects of your transformed input data.
(b) Calculate Correlation between two features Commute Distance and Yearly Income