Customer Risk Classifier

By John Park

Introduction

Uses US traffic accident dataset and a data of customers' information (their street address and average miles driven daily) to categorize a customer as a high or low risk client

Assumptions

1. Most customers do not travel outside of their state for a daily commute
2. Circular area determined by the range of a customer's daily commute provides a reasonable region that the customer will dwell in for a large majority of his or her time
3. The likelihood that an accidents will occur in a certain area can be predicted by how often accidents occur within that area for a given amount of time
4. There is a direct correlation between a customer's likelihood of getting into an accident and the regularity of accidents within that area

Hypothesis

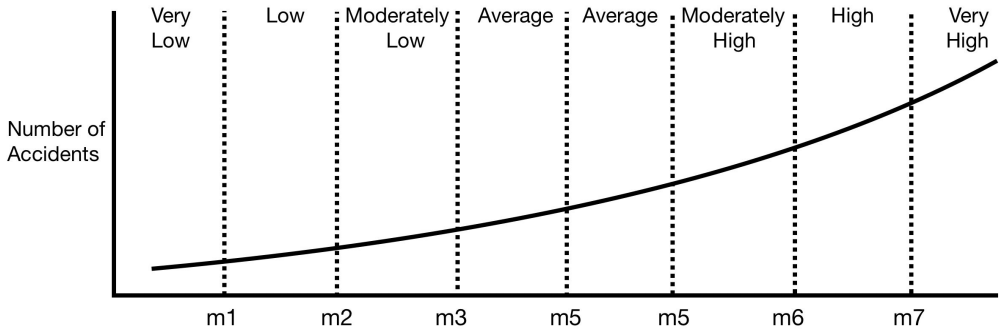
A customer's risk level can be predicted by the regularity of accidents that occur within an area of land determined by the customer's commute radius and his or her location

Procedure

1. Use Google Maps API to convert each customer's home address to its longitudinal and latitudinal coordinates
2. For each accident that occurs within a customer's state, calculate the distance of that accident from the customer's residence by using the customer's previously calculated geographical coordinates and the coordinates of the accident (provided by the US traffic dataset)
3. Count the number of accidents whose distance from the customer's residence fall within the radius of his/her daily commute

Procedure Cont'd

1. After gathering the number of accidents per customer (hereafter described as Accidents Index) sort the data according to the Accidents Index in ascending order and use the median to split the data in two.
2. Repeat the last step 2 more times, further splitting the data into 8 zones (see below) (median was used instead of average since average is sensitive to outliers and especially influential to a small dataset such as a dataset of 30 customers being categorized)



Procedure Cont'd

1. For each customer, determine which zone the customer's Accidents Index falls, and use this to categorize each customer's risk level:

   1 = Very Low 5 = Average
   2 = Low 6 = Moderately High
   3 = Moderately Low 7 = High
   4 = Average 8 = Very High

2. Use Google Sheets API to output the customer classification data to a desired spreadsheet document

Q&A