Who Dominates Taxi Business?

STAT 432 FA2019 - Team LRHK

Eric Jong Bum Kim (jekim3)
Israel Reyes (ir2)
Jihee Hwang (jhwang55)
Hyun Suk Lee ()

November 15, 2019

Abstract

Coming Soon

Goal

Classification model predicting the company

Description of the dataset that is sufficient for a reader to understand your motivation for using the dataset. Background information on the dataset, including specific citation of its source.

Introduction

Eric

Data Description

The Chicago Taxi Trips data is from the Kaggle datasets.

There are 23 columns of information about each tripes. There are trip logistics information including trip starting and ending timestamps, trip duration and miles, pick up and drop off locations, latitude, and longitude. In additiona, there are finnanical-related information of each trips including, trip fare, tips, number of tolls, extra charges, total trip costs, and payment methods. Finally, third catergories include infomation about taxi and its company. More explanation of each columns can be found in the appendix of the file at the end of the project.

The data originates from the UCI Machine Learning Repository3. The data was accessed via the ucidata package on GitHub.4 The data was collected from four sources:

Hungarian Institute of Cardiology. Budapest (Andras Janosi, M.D.) University Hospital, Zurich, Switzerland (William Steinbrunn, M.D.) University Hospital, Basel, Switzerland (Matthias Pfisterer, M.D.) V.A. Medical Center, Long Beach and Cleveland Clinic Foundation (Robert Detrano, M.D., Ph.D.) Each patient in the data was graded on a scale from 0 to 4, indicating the severity of heart disease. A score of 0 indicates no presence of heart disease. Scores 1 through 4 indicate increasing severity of heart disease. In particular, the scores from 1 to 4 indicate the number of vessels found through angiography5 with greater than 50% diameter narrowing.

Background Informaton On Data

Jihee

Statistical Learning Task

Israel

Data Loading

Hyun Suk

Data Modeling

Hyun Suk