

```

# CS 513
# Final Project
# Dataset Clean-Up

# Julia Nelson, Taylor Niedzielski, Sonia Patel, and Noah Suttora

# Clear Environment
rm(list=ls())

# Import Dataset
RawData <- read.csv("/Users/Sonia/Desktop/CS 513/Project/
CAX_Startup_Data.csv", na.string = c("?"))
library(e1071)
library(caret)
View(RawData)

# Remove Columns of DisInterest
Data <- RawData[-c(1,5,6,7,9,11,13,14,15,16,17,24,
                  25,26,28,34,43,46,47,48,49,50,51,
                  52,53,54,56,57,59,61,62,63,64,65,
                  66,67,71,73,74,75,76,79,80,82,83,
                  84,85,86,87,89,90,91,92,93,
                  98,99,100,101,102,103,104,105,
                  106,107,108,109,110,111,112,
                  113,114,115,116)]

# Rename Columns of Interest
colnames(Data) <- c("Status","FoundingYear","Age","FocusFunctions",
                    "EmployeeCount","TeamSizeGrowth","NumSeedInvestors",
                    "NumAngelorVCInvestors","NumCofounders","NumAdvisors",
                    "SeniorLeadershipTeamSize","EmployeeTeamSize",
                    "TopCompanyExp","StartupExp",
                    "SucessfulStartupExp","Big5Partner","ConsultingExp",
                    "ProductorService","DataFocus","ConsumerDataFocus",
                    "DataStructureFocus","SubscriptionBased",
                    "CloudPlatformBased","LocalGlobal","BusinessModel",
                    "CapitalIntensive","CrowdsourcingBased","CrowdfundingBased",
                    "B2CorB2B","GlobalExporsure","HighestEducation",
                    "Fortune100Exp","Fortune500Exp","Fortune1000Exp",
                    "NumFounderRecognition","PricingStrategy",
                    "HyperLocalisation","LongtermFounderRelationship",
                    "GooglePageRank","NumDirectCompetitors","EmployeesPerYear",
                    "LastFundingRoundAmount","RecessionSurvival")

View(Data)

```