DV\_FinalProject

**Authors: Rolando Melendez, Hyunji Seo, Catherine Tu**

## This is how our RStudio is configured:

sessionInfo()

## R version 3.2.2 (2015-08-14)  
## Platform: x86\_64-apple-darwin13.4.0 (64-bit)  
## Running under: OS X 10.11.1 (El Capitan)  
##   
## locale:  
## [1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] magrittr\_1.5 tools\_3.2.2 htmltools\_0.2.6 yaml\_2.1.13   
## [5] stringi\_1.0-1 rmarkdown\_0.8.1 knitr\_1.11 stringr\_1.0.0   
## [9] digest\_0.6.8 evaluate\_0.8

## Program file structure

00 Doc - contains .Rmd and .html

01 Data - contains the code to Extract, Transform, and Load our data

03 Tableau Workbooks - this is where the Tableau workbook(s) goes (twbx format)

04 Shiny - contains our workflows and the code for the visualizations along with code used to create Shiny app.

## Explanation of Data:

Hyunji's Intro about banking data add sentece(s) about blended data

## Data summary and subset

#Shows the subset and summary of the data frame  
source("../01 Data/R\_ExtractTransform.R", echo = TRUE)

##   
## > require(dplyr)

## Loading required package: dplyr  
##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

##   
## > setwd("~/Desktop/Desktop/UT/Fall 2015/Data Visualization/DV\_FinalProject/01 Data/CSVs")  
##   
## > file\_path <- "bank-additional.csv"  
##   
## > df <- rename(read.csv(file\_path, stringsAsFactors = FALSE,   
## + sep = ";"), dflt = default)  
##   
## > names(df) <- toupper(names(df))  
##   
## > names(df) <- gsub("\\.+", "\_", names(df))  
##   
## > names(df)  
## [1] "AGE" "JOB" "MARITAL" "EDUCATION"   
## [5] "DFLT" "HOUSING" "LOAN" "CONTACT"   
## [9] "MONTH" "DAY\_OF\_WEEK" "DURATION" "CAMPAIGN"   
## [13] "PDAYS" "PREVIOUS" "POUTCOME" "EMP\_VAR\_RATE"   
## [17] "CONS\_PRICE\_IDX" "CONS\_CONF\_IDX" "EURIBOR3M" "NR\_EMPLOYED"   
## [21] "Y"   
##   
## > measures <- c("AGE", "DURATION", "CAMPAIGN", "PDAYS",   
## + "PREVIOUS", "EMP\_VAR\_RATE", "CONS\_PRICE\_IDX", "CONS\_CONF\_IDX",   
## + "EURIBOR3M", "NR\_ ..." ... [TRUNCATED]   
##   
## > for (n in names(df)) {  
## + df[n] <- data.frame(lapply(df[n], gsub, pattern = "[^ -~]",   
## + replacement = ""))  
## + }  
##   
## > dimensions <- setdiff(names(df), measures)  
##   
## > if (length(measures) > 1 || !is.na(dimensions)) {  
## + for (d in dimensions) {  
## + df[d] <- data.frame(lapply(df[d], gsub, pattern = "[\"']", ..." ... [TRUNCATED]   
##   
## > if (length(measures) > 1 || !is.na(measures)) {  
## + for (m in measures) {  
## + df[m] <- data.frame(lapply(df[m], gsub, pattern = "[^--.0-9]", .... [TRUNCATED]   
##   
## > write.csv(df, paste(gsub(".csv", "", file\_path), ".reformatted.csv",   
## + sep = ""), row.names = FALSE, na = "")  
##   
## > tableName <- gsub(" +", "\_", gsub("[^A-z, 0-9, ]",   
## + "", gsub(".csv", "", file\_path)))  
##   
## > sql <- paste("CREATE TABLE", tableName, "(\n-- Change table\_name to the table name you want.\n")  
##   
## > if (length(measures) > 1 || !is.na(dimensions)) {  
## + for (d in dimensions) {  
## + sql <- paste(sql, paste(d, "varchar2(4000),\n"))  
## + }  
## + .... [TRUNCATED]   
##   
## > if (length(measures) > 1 || !is.na(measures)) {  
## + for (m in measures) {  
## + if (m != tail(measures, n = 1))   
## + sql <- paste(sq .... [TRUNCATED]   
##   
## > sql <- paste(sql, ");")  
##   
## > cat(sql)  
## CREATE TABLE bankadditional (  
## -- Change table\_name to the table name you want.  
## JOB varchar2(4000),  
## MARITAL varchar2(4000),  
## EDUCATION varchar2(4000),  
## DFLT varchar2(4000),  
## HOUSING varchar2(4000),  
## LOAN varchar2(4000),  
## CONTACT varchar2(4000),  
## MONTH varchar2(4000),  
## DAY\_OF\_WEEK varchar2(4000),  
## POUTCOME varchar2(4000),  
## Y varchar2(4000),  
## AGE number(38,4),  
## DURATION number(38,4),  
## CAMPAIGN number(38,4),  
## PDAYS number(38,4),  
## PREVIOUS number(38,4),  
## EMP\_VAR\_RATE number(38,4),  
## CONS\_PRICE\_IDX number(38,4),  
## CONS\_CONF\_IDX number(38,4),  
## EURIBOR3M number(38,4),  
## NR\_EMPLOYED number(38,4)  
## );

#Shows the subset and summary of the data frame  
source("../01 Data/R\_ExtractTransformJob.R", echo = TRUE)

##   
## > require(dplyr)  
##   
## > setwd("~/Desktop/Desktop/UT/Fall 2015/Data Visualization/DV\_FinalProject/01 Data/CSVs")  
##   
## > file\_path <- "job-type.csv"  
##   
## > df <- rename(read.csv(file\_path, stringsAsFactors = FALSE,   
## + sep = ","))  
##   
## > names(df) <- toupper(names(df))  
##   
## > names(df) <- gsub("\\.+", "\_", names(df))  
##   
## > names(df)  
## [1] "JOB\_TYPE" "AVERAGE\_SALARY" "HOURS\_PER\_WEEK"  
##   
## > measures <- c("AVERAGE\_SALARY", "HOURS\_PER\_WEEK")  
##   
## > for (n in names(df)) {  
## + df[n] <- data.frame(lapply(df[n], gsub, pattern = "[^ -~]",   
## + replacement = ""))  
## + }  
##   
## > dimensions <- setdiff(names(df), measures)  
##   
## > if (length(measures) > 1 || !is.na(dimensions)) {  
## + for (d in dimensions) {  
## + df[d] <- data.frame(lapply(df[d], gsub, pattern = "[\"']", ..." ... [TRUNCATED]   
##   
## > if (length(measures) > 1 || !is.na(measures)) {  
## + for (m in measures) {  
## + df[m] <- data.frame(lapply(df[m], gsub, pattern = "[^--.0-9]", .... [TRUNCATED]   
##   
## > write.csv(df, paste(gsub(".csv", "", file\_path), ".reformatted.csv",   
## + sep = ""), row.names = FALSE, na = "")  
##   
## > tableName <- gsub(" +", "\_", gsub("[^A-z, 0-9, ]",   
## + "", gsub(".csv", "", file\_path)))  
##   
## > sql <- paste("CREATE TABLE", tableName, "(\n-- Change table\_name to the table name you want.\n")  
##   
## > if (length(measures) > 1 || !is.na(dimensions)) {  
## + for (d in dimensions) {  
## + sql <- paste(sql, paste(d, "varchar2(4000),\n"))  
## + }  
## + .... [TRUNCATED]   
##   
## > if (length(measures) > 1 || !is.na(measures)) {  
## + for (m in measures) {  
## + if (m != tail(measures, n = 1))   
## + sql <- paste(sq .... [TRUNCATED]   
##   
## > sql <- paste(sql, ");")  
##   
## > cat(sql)  
## CREATE TABLE jobtype (  
## -- Change table\_name to the table name you want.  
## JOB\_TYPE varchar2(4000),  
## AVERAGE\_SALARY number(38,4),  
## HOURS\_PER\_WEEK number(38,4)  
## );

source("../01 Data/R\_Load.R", echo = TRUE)

##   
## > require("jsonlite")

## Loading required package: jsonlite

##   
## > require("RCurl")

## Loading required package: RCurl  
## Loading required package: bitops

##   
## > df <- data.frame(fromJSON(getURL(URLencode("skipper.cs.utexas.edu:5001/rest/native/?query=\"select \* from BNKMKTG\""),   
## + httpheader = c(DB = "j ..." ... [TRUNCATED]   
##   
## > head(df)  
## JOB MARITAL EDUCATION DFLT HOUSING LOAN CONTACT MONTH  
## 1 bluecollar married basic9y no yes no cellular may  
## 2 admin married unknown no yes no telephone apr  
## 3 services married highschool no yes no cellular apr  
## 4 bluecollar divorced basic6y no yes no cellular may  
## 5 admin single universitydegree no yes no cellular jul  
## 6 technician divorced unknown no yes yes cellular oct  
## DAY\_OF\_WEEK POUTCOME Y AGE DURATION CAMPAIGN PDAYS PREVIOUS  
## 1 wed nonexistent no 37 204 2 999 0  
## 2 wed success yes 52 403 1 6 1  
## 3 mon nonexistent no 46 180 1 999 0  
## 4 wed nonexistent no 42 16 2 999 0  
## 5 mon nonexistent no 35 447 3 999 0  
## 6 thu nonexistent no 49 81 1 999 0  
## EMP\_VAR\_RATE CONS\_PRICE\_IDX CONS\_CONF\_IDX EURIBOR3M NR\_EMPLOYED  
## 1 -1.8 92.893 -46.2 1.334 5099.1  
## 2 -1.8 93.749 -34.6 0.654 5008.7  
## 3 -1.8 93.075 -47.1 1.405 5099.1  
## 4 -1.8 92.893 -46.2 1.281 5099.1  
## 5 1.4 93.918 -42.7 4.960 5228.1  
## 6 -3.4 92.431 -26.9 0.754 5017.5  
##   
## > summary(df)  
## JOB MARITAL EDUCATION   
## admin :1012 divorced: 446 universitydegree :1264   
## bluecollar: 884 married :2509 highschool : 921   
## technician: 691 single :1153 basic9y : 574   
## services : 393 unknown : 11 professionalcourse: 535   
## management: 324 basic4y : 429   
## retired : 166 basic6y : 228   
## (Other) : 649 (Other) : 168   
## DFLT HOUSING LOAN CONTACT   
## no :3315 no :1839 no :3349 cellular :2652   
## unknown: 803 unknown: 105 unknown: 105 telephone:1467   
## yes : 1 yes :2175 yes : 665   
##   
##   
##   
##   
## MONTH DAY\_OF\_WEEK POUTCOME Y AGE   
## may :1378 fri:768 failure : 454 no :3668 Min. :18.00   
## jul : 711 mon:855 nonexistent:3523 yes: 451 1st Qu.:32.00   
## aug : 636 thu:860 success : 142 Median :38.00   
## jun : 530 tue:841 Mean :40.11   
## nov : 446 wed:795 3rd Qu.:47.00   
## apr : 215 Max. :88.00   
## (Other): 203   
## DURATION CAMPAIGN PDAYS PREVIOUS   
## Min. : 0.0 Min. : 1.000 Min. : 0.0 Min. :0.0000   
## 1st Qu.: 103.0 1st Qu.: 1.000 1st Qu.:999.0 1st Qu.:0.0000   
## Median : 181.0 Median : 2.000 Median :999.0 Median :0.0000   
## Mean : 256.8 Mean : 2.537 Mean :960.4 Mean :0.1903   
## 3rd Qu.: 317.0 3rd Qu.: 3.000 3rd Qu.:999.0 3rd Qu.:0.0000   
## Max. :3643.0 Max. :35.000 Max. :999.0 Max. :6.0000   
##   
## EMP\_VAR\_RATE CONS\_PRICE\_IDX CONS\_CONF\_IDX EURIBOR3M   
## Min. :-3.40000 Min. :92.20 Min. :-50.8 Min. :0.635   
## 1st Qu.:-1.80000 1st Qu.:93.08 1st Qu.:-42.7 1st Qu.:1.334   
## Median : 1.10000 Median :93.75 Median :-41.8 Median :4.857   
## Mean : 0.08497 Mean :93.58 Mean :-40.5 Mean :3.621   
## 3rd Qu.: 1.40000 3rd Qu.:93.99 3rd Qu.:-36.4 3rd Qu.:4.961   
## Max. : 1.40000 Max. :94.77 Max. :-26.9 Max. :5.045   
##   
## NR\_EMPLOYED   
## Min. :4964   
## 1st Qu.:5099   
## Median :5191   
## Mean :5166   
## 3rd Qu.:5228   
## Max. :5228   
##

## Label Explanations:

AGE - The age of each marketing participant

JOB - Type of job

MARITAL - Marital Status

EDUCATION - Level of education reached by participant

DEFAULT - Has this participant defaulted before

HOUSING - Does this participant have a housing loan (Mortgage)

LOAN - Does this participant have a personal loan

CONTACT - Contact communication type (cellular or telephone)

MONTH - Last contact month of the year

DAY OF WEEK - Last contact day of the week

DURATION - Duration of the last contact to the participant

CAMPAIGN - Number of times this participant was contacted during this campaign

PDAYS - Number of days since participant was last contacted

PREVIOUS - Number of times this participant was contacted before this campaign

POUTCOME - Outcome of previous marketing campaign

EMP VAR RATE - Employment variation rate

CONS PRICE IDX - Consumer price index

CONS CONF IDX - Consumer confidence index

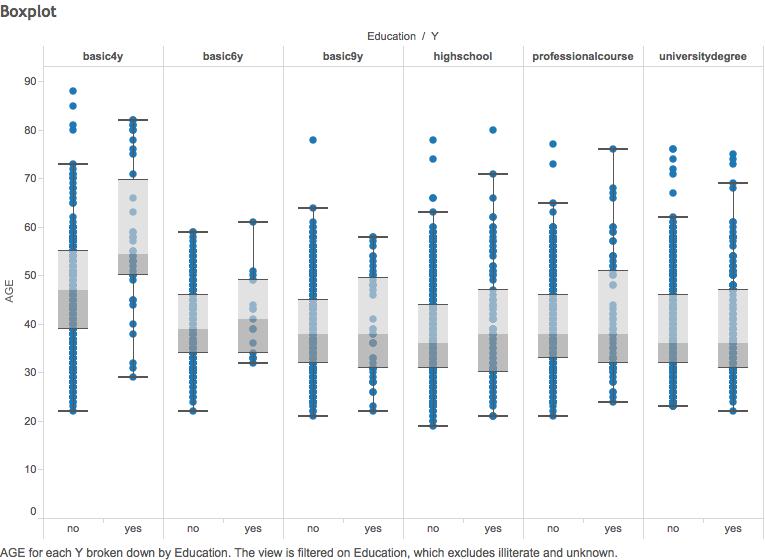
EURIBOR3M - Rate at which banks lend to each other; 3 month loans

NR EMPOLOYED - Number of employees

Y - Outcome, Did the participant subscribe to a term deposit

## Non\_Aggregated Measures Analysis:

#### Boxplot:

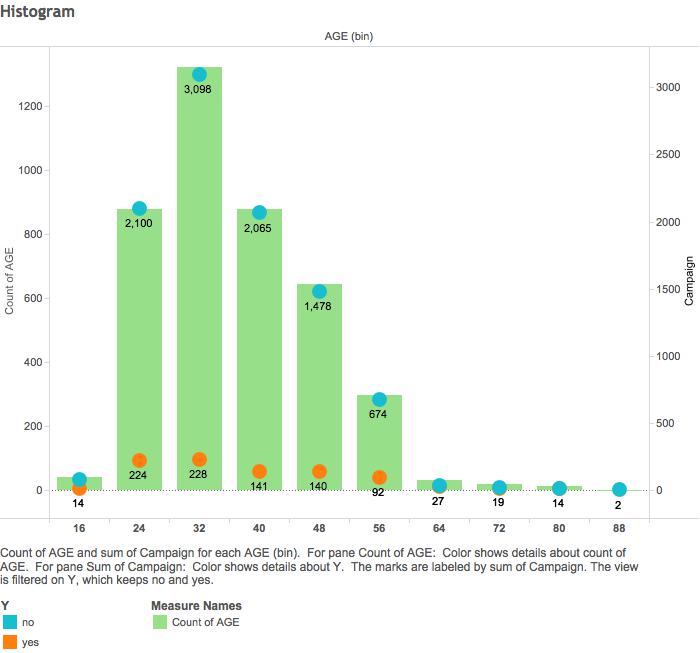


words

## Aggregrated Measures Analysis:

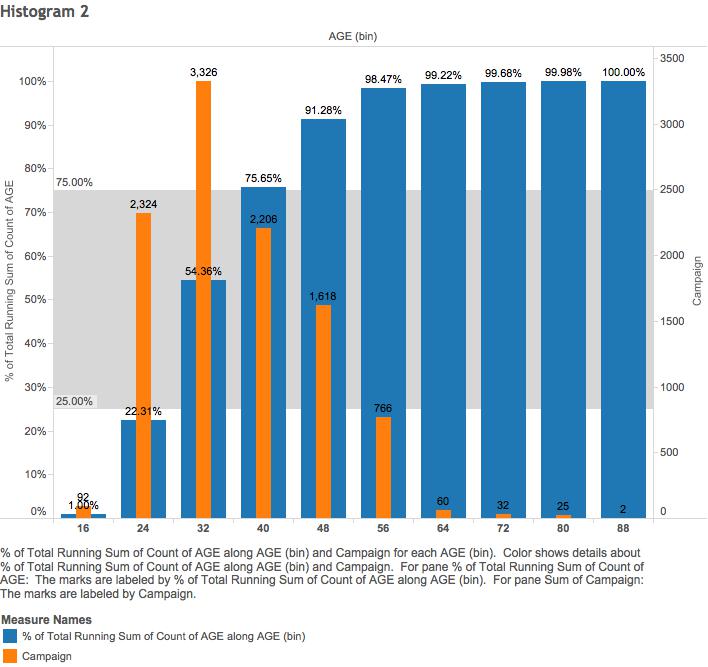
#### Histogram:

##### Legit Title:



words

##### Legit Title No 2:



words

## Scatter Plots:

#### Legit Title 1:

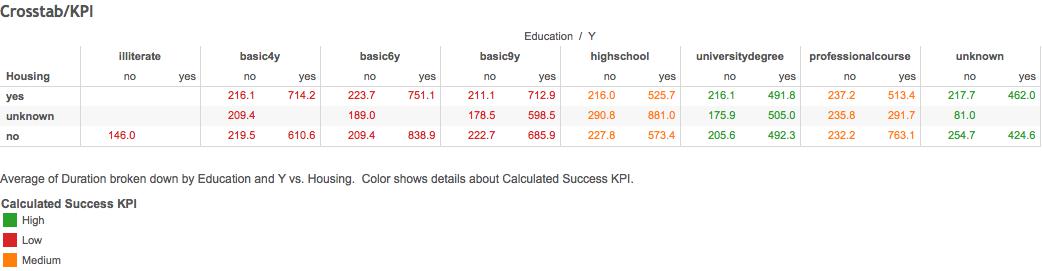
words

#### Legit Title No 2:

words

## Crosstabs:

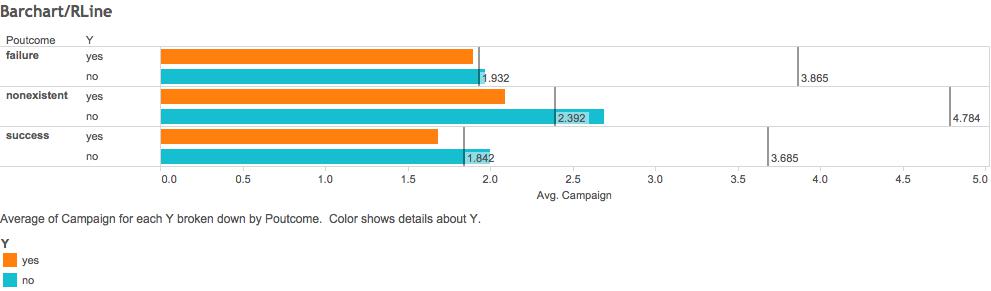
#### Legit Title:



words

## Barcharts:

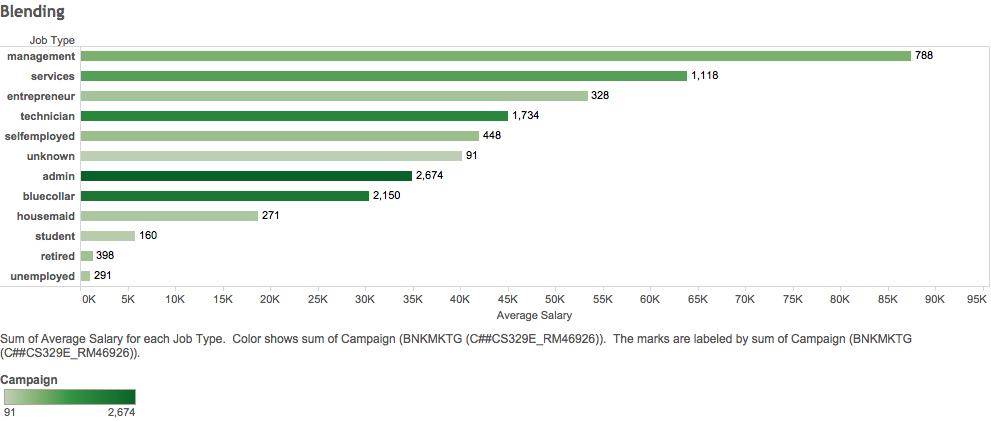
#### Legit Title 1:



words

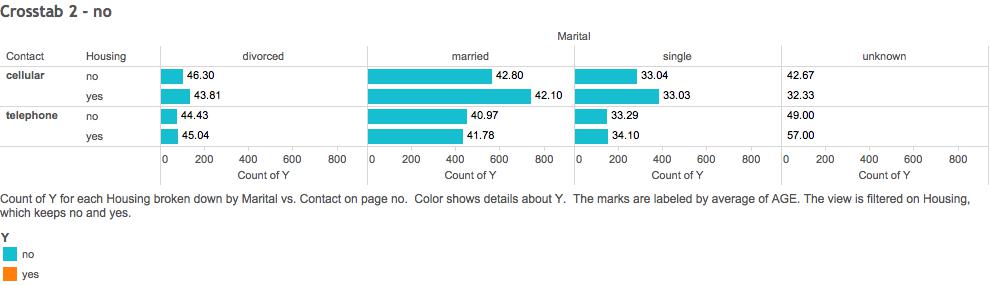
#### Legit Title 2:

*Also demonstrates data blending*



words

#### Legit Title 3:





words