**Portfolio Project: Oral and PowerPoint Presentation**

Wendell Salas

Colorado State University Global

MIS 581: Capstone: Business Intelligence and Data Analytics

Dr. Jamia Mills

March 12, 2023

**MIS 581 Portfolio Project Code and Github Link**

GitHub Account Link for the MIS 581 Capstone: Portfolio Project

<https://github.com/WJSalas/MIS581Capstone>

SAS Code used for Data Characterization, Exploration, and Summary Statistics

SAS Code for Characterization of Assessor’s Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 12:24 PM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

ods noproctitle;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\*\*\* Analyze categorical variables \*\*\*/

title "Frequencies for Categorical Variables";

**proc freq** data=MYDATA.**'2022\_ASSESSOR\_DATA'n**;

tables State\_Use\_Dscr / plots=(freqplot);

**run**;

/\*\*\* Analyze numeric variables \*\*\*/

title "Descriptive Statistics for Numeric Variables";

**proc means** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** n nmiss min mean median max std;

var MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val Calc\_Taxes;

**run**;

title;

**proc univariate** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** noprint;

histogram MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val Calc\_Taxes;

**run**;

SAS Code for Exploration of Assessor’s Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 11:09 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

options validvarname=any;

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\* Scatter plot matrix macro \*/

**%macro** scatterPlotMatrix(xVars=, title=, groupVar=);

**proc sgscatter** data=MYDATA.**'2022\_ASSESSOR\_DATA'n**;

matrix &xVars / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

diagonal=(histogram normal);

title &title;

**run**;

title;

**%mend** scatterPlotMatrix;

/\* Regression scatter plot macro \*/

**%macro** regressionScatterplot(xVar=, yVar=, title=, groupVar=);

**proc sgscatter** data=MYDATA.**'2022\_ASSESSOR\_DATA'n**;

plot (&yVar)\*(&xVar) / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

reg;

title &title;

**run**;

title;

**%mend** regressionScatterplot;

**%scatterPlotMatrix**(xVars=MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE

Calc\_Assessed\_val Calc\_Taxes,

title="Scatter plot matrix grouped by State\_Use\_Dscr",

groupVar=State\_Use\_Dscr);

**%regressionScatterplot**(xVar=MILL\_LEVY, yVar=Calc\_Taxes,

title="Calc\_Taxes vs MILL\_LEVY grouped by State\_Use\_Dscr",

groupVar=State\_Use\_Dscr);

**%regressionScatterplot**(xVar=TOT\_ACTVAL, yVar=Calc\_Taxes,

title="Calc\_Taxes vs TOT\_ACTVAL grouped by State\_Use\_Dscr",

groupVar=State\_Use\_Dscr);

**%regressionScatterplot**(xVar=Calc\_ASSESS\_RATE, yVar=Calc\_Taxes,

title="Calc\_Taxes vs Calc\_ASSESS\_RATE grouped by State\_Use\_Dscr",

groupVar=State\_Use\_Dscr);

**%regressionScatterplot**(xVar=Calc\_Assessed\_val, yVar=Calc\_Taxes,

title="Calc\_Taxes vs Calc\_Assessed\_val grouped by State\_Use\_Dscr",

groupVar=State\_Use\_Dscr);

**proc sort** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** out=WORK.TempSorted4877;

by State\_Use\_Dscr;

**run**;

**proc boxplot** data=WORK.TempSorted4877;

plot (MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val

Calc\_Taxes)\*State\_Use\_Dscr / boxstyle=schematic;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted4877;

SAS Code for the Summary Statistics of the Assessor’s Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 11:27 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

**proc means** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** chartype mean std min max median n

vardef=df skewness kurtosis q1 q3 qmethod=os;

var MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val Calc\_Taxes;

class State\_Use\_Dscr;

**run**;

**proc univariate** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** vardef=df noprint;

var MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val Calc\_Taxes;

class State\_Use\_Dscr;

histogram MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val Calc\_Taxes /

normal(noprint);

inset mean std min max median n skewness kurtosis q1 q3 / position=ne;

**run**;

**proc sort** data=MYDATA.**'2022\_ASSESSOR\_DATA'n** out=WORK.TempSorted2236;

by State\_Use\_Dscr;

**run**;

**proc boxplot** data=WORK.TempSorted2236;

plot (MILL\_LEVY TOT\_ACTVAL Calc\_ASSESS\_RATE Calc\_Assessed\_val

Calc\_Taxes)\*State\_Use\_Dscr / boxstyle=schematic;

insetgroup mean stddev min max n q1 q2 q3 / position=top;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted2236;

**run**;

SAS Code for Characterization of Well Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 12:27 PM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

ods noproctitle;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\*\*\* Analyze categorical variables \*\*\*/

title "Frequencies for Categorical Variables";

**proc freq** data=MYDATA.**'2022\_WELL\_DATA'n**;

tables Well\_Stat / plots=(freqplot);

**run**;

/\*\*\* Analyze numeric variables \*\*\*/

title "Descriptive Statistics for Numeric Variables";

**proc means** data=MYDATA.**'2022\_WELL\_DATA'n** n nmiss min mean median max std;

var Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est;

**run**;

title;

**proc univariate** data=MYDATA.**'2022\_WELL\_DATA'n** noprint;

histogram Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est;

**run**;

SAS Code for Exploration of Well Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 11:18 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

options validvarname=any;

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\* Scatter plot matrix macro \*/

**%macro** scatterPlotMatrix(xVars=, title=, groupVar=);

**proc sgscatter** data=MYDATA.**'2022\_WELL\_DATA'n**;

matrix &xVars / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

diagonal=(histogram normal);

title &title;

**run**;

title;

**%mend** scatterPlotMatrix;

/\* Regression scatter plot macro \*/

**%macro** regressionScatterplot(xVar=, yVar=, title=, groupVar=);

**proc sgscatter** data=MYDATA.**'2022\_WELL\_DATA'n**;

plot (&yVar)\*(&xVar) / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

reg;

title &title;

**run**;

title;

**%mend** regressionScatterplot;

**%scatterPlotMatrix**(xVars=Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est,

title="Scatter plot matrix grouped by Well\_Stat", groupVar=Well\_Stat);

**%regressionScatterplot**(xVar=Oil\_Sales, yVar=Tax\_Est,

title="Tax\_Est vs Oil\_Sales grouped by Well\_Stat", groupVar=Well\_Stat);

**%regressionScatterplot**(xVar=Gas\_Sales, yVar=Tax\_Est,

title="Tax\_Est vs Gas\_Sales grouped by Well\_Stat", groupVar=Well\_Stat);

**%regressionScatterplot**(xVar=tot\_mill, yVar=Tax\_Est,

title="Tax\_Est vs tot\_mill grouped by Well\_Stat", groupVar=Well\_Stat);

**%regressionScatterplot**(xVar=asd\_val, yVar=Tax\_Est,

title="Tax\_Est vs asd\_val grouped by Well\_Stat", groupVar=Well\_Stat);

**proc sort** data=MYDATA.**'2022\_WELL\_DATA'n** out=WORK.TempSorted4877;

by Well\_Stat;

**run**;

**proc boxplot** data=WORK.TempSorted4877;

plot (Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est)\*Well\_Stat /

boxstyle=schematic;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted4877;

**run**;

SAS Code for Summary Statistics of Well Data

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/7/23, 11:31 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS04-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-76295-doW

\*

\*/

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

**proc means** data=MYDATA.**'2022\_WELL\_DATA'n** chartype mean std min max median n

vardef=df skewness kurtosis q1 q3 qmethod=os;

var Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est;

class Well\_Stat;

**run**;

**proc univariate** data=MYDATA.**'2022\_WELL\_DATA'n** vardef=df noprint;

var Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est;

class Well\_Stat;

histogram Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est / normal(noprint);

inset mean std min max median n skewness kurtosis q1 q3 / position=ne;

**run**;

**proc sort** data=MYDATA.**'2022\_WELL\_DATA'n** out=WORK.TempSorted2236;

by Well\_Stat;

**run**;

**proc boxplot** data=WORK.TempSorted2236;

plot (Oil\_Sales Gas\_Sales tot\_mill asd\_val Tax\_Est)\*Well\_Stat /

boxstyle=schematic;

insetgroup mean stddev min max n q1 q2 q3 / position=top;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted2236;

**run**;

R Code for Combining Assessor’s and COGCC’s Datasets and Outputting the Third Dataset

MIS 581 Capstone Project

Author: Wendell Salas

First we need to load the required libraries for running our R scripts.

Hide

# Load Library

library(fields)

Next, we will read in the two data files for our project

Hide

## Working Files

Assessor\_Master\_Data <- paste(getwd(), "Working/Assessor\_Master\_Res\_Data\_2022.csv", sep = "/")

Well\_Master\_Data <- paste(getwd(), "Working/Master\_Well\_Data\_Combined\_v3.csv", sep = "/")

## Read in .csv file

Assessor\_Data <- read.csv(Assessor\_Master\_Data, header = TRUE)

Well\_Data <- read.csv(Well\_Master\_Data, header = TRUE)

Then, we will create a Stacked ***for loop*** to calculate the distance from each well in the Well\_Data file to each home in the Assessor\_Data file. We will also incorporate counts for the number of homes within one mile of each well, within 1000 ft of a well, and sum up the total home property taxes for all of the homes within one mile of each well.

Hide

# 1,000 ft = 0.3048 km

# 1 mile = 1.609344 km

## Stacked for loop

# Count the number of rows in each of the dataframes. These will be used to set our for loops.

n1 <- nrow(Well\_Data)

n2 <- nrow(Assessor\_Data)

# Create a new dataframe called df where we will keep track of our outputs from the for loops.

df <- data.frame(Well\_API=character(),

Well\_ID=character(),

Well\_Status=character(),

Oil\_Sales=integer(),

Gas\_Sales=integer(),

Stripper\_flg=character(),

Well\_Tot\_Mill=numeric(),

Well\_Asd\_Val=integer(),

H1000=integer(),

H5280=integer(),

Well\_PTax=numeric(),

Sum\_Home\_Asd\_Val=integer(),

Sum\_Home\_PTax=numeric(),

stringsAsFactors = FALSE)

for (i in 1:n1){

# Initialize the counting veriables for each loop through the Well\_Data dataframe.

tax\_sum <- 0

h\_asd\_sum <- 0

h\_count <- 0

H1K <- 0

for (j in 1:n2){

# Assign each Lat and Long to the correct input

long1 <- Well\_Data$long1[i]

lat1 <- Well\_Data$lat1[i]

long2 <- Assessor\_Data$long2[j]

lat2 <- Assessor\_Data$lat2[j]

# Calculate the distances between the Well and home

dist\_km <- rdist.earth(matrix(c(long1,lat1), ncol=2), matrix(c(long2,lat2), ncol=2), miles = FALSE, R = 6371)

# if-else statement that will be used to tally the number of homes within the two different ranges that we are interested in and exclude anything over 1 mile away from a well.

if (dist\_km > 1.609344){

next

}else if (dist\_km <= 1.609344 & dist\_km > 0.3048){

H1K\_new <- 0

h\_count\_new <- 1

h\_asd <- Assessor\_Data$TOT\_ACTVAL[j]

h\_tax <- Assessor\_Data$Calc\_Taxes[j]

}else if (dist\_km <= 0.3048){

H1K\_new <- 1

h\_count\_new <- 1

h\_asd <- Assessor\_Data$TOT\_ACTVAL[j]

h\_tax <- Assessor\_Data$Calc\_Taxes[j]

}

# Count the number of homes within 1000 ft

H1K <- H1K + H1K\_new

# Count the number of homes within one mile of the well

h\_count <- h\_count + h\_count\_new

# Sum up the total home assessed values

h\_asd\_sum <- h\_asd\_sum + h\_asd

# Summaries the total Property Tax amount for all of the homes within one mile of the well

tax\_sum <- tax\_sum + h\_tax

}

# Create a dataframe from the ouput of the if-else statement and append it to the df

df\_new <- c(Well\_Data$API\_Label[i], Well\_Data$Well\_ID[i], Well\_Data$Well\_Stat[i], Well\_Data$Oil\_Sales[i], Well\_Data$Gas\_Sales[i], Well\_Data$stripper\_flg[i], Well\_Data$tot\_mill[i], Well\_Data$asd\_val[i], H1K, h\_count, Well\_Data$Tax\_Est[i], h\_asd\_sum, tax\_sum)

df <- rbind(df, df\_new)

}

# Rename the columns in the df Output

colnames(df) <- c("Well\_API", "Well\_ID", "Well\_Status", "Oil\_Sales", "Gas\_Sales", "Stripper\_flg", "Well\_Tot\_Mill", "Well\_Asd\_Val", "H1000", "H5280", "Well\_PTax", "Sum\_Home\_Asd\_Val", "Sum\_Home\_PTax")

Now, lets take a look at our new dataframe that contains the output from our **Staked for loop**.

Hide

View(df)

Now we will export the df dataframe that has the output form the for loops, so that we can work with it in another application.

Hide

Output <- paste(getwd(), "Working/Assessor\_Well\_Output\_Data\_v2.csv", sep = "/")

write.csv(df, Output, row.names = TRUE)

SAS Code for Characterizing, Exploring, Summary Statistics of the R Output Dataset

SAS Code for Characterization of the R Output Dataset

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/9/23, 11:25 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS01-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-118825-j

\*

\*/

ods noproctitle;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\*\*\* Analyze numeric variables \*\*\*/

title "Descriptive Statistics for Numeric Variables";

**proc means** data=MYDATA.WELL\_COI\_OUTPUT\_DATA n nmiss min mean median max std;

var Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate;

**run**;

title;

**proc univariate** data=MYDATA.WELL\_COI\_OUTPUT\_DATA noprint;

histogram Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate;

**run**;

SAS Code for the Exploration of the R Output Dataset (First Set of Six Variables)

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/9/23, 11:30 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS01-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-118825-j

\*

\*/

options validvarname=any;

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\* Scatter plot matrix macro \*/

**%macro** scatterPlotMatrix(xVars=, title=, groupVar=);

**proc sgscatter** data=MYDATA.WELL\_COI\_OUTPUT\_DATA;

matrix &xVars / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

diagonal=(histogram normal);

title &title;

**run**;

title;

**%mend** scatterPlotMatrix;

**%scatterPlotMatrix**(xVars=Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000

H5280, title="Scatter plot matrix grouped by Well\_Status",

groupVar=Well\_Status);

**proc sort** data=MYDATA.WELL\_COI\_OUTPUT\_DATA out=WORK.TempSorted4877;

by Well\_Status;

**run**;

**proc boxplot** data=WORK.TempSorted4877;

plot (Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280)\*Well\_Status

/ boxstyle=schematic;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted4877;

**run**;

SAS Code for the Exploration of the R Output Dataset (Second Set of Six Variables)

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/9/23, 11:31 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS01-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-118825-j

\*

\*/

options validvarname=any;

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\* Scatter plot matrix macro \*/

**%macro** scatterPlotMatrix(xVars=, title=, groupVar=);

**proc sgscatter** data=MYDATA.WELL\_COI\_OUTPUT\_DATA;

matrix &xVars / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

diagonal=(histogram normal);

title &title;

**run**;

title;

**%mend** scatterPlotMatrix;

**%scatterPlotMatrix**(xVars=Well\_PTax Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home

Avg\_H1000\_NoWell H1000\_Effect,

title="Scatter plot matrix grouped by Well\_Status", groupVar=Well\_Status);

**proc sort** data=MYDATA.WELL\_COI\_OUTPUT\_DATA out=WORK.TempSorted4877;

by Well\_Status;

**run**;

**proc boxplot** data=WORK.TempSorted4877;

plot (Well\_PTax Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell

H1000\_Effect)\*Well\_Status / boxstyle=schematic;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted4877;

**run**;

SAS Code for the Exploration of the R Output Dataset (Third Set of Six Variables)

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/9/23, 11:34 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS01-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-118825-j

\*

\*/

options validvarname=any;

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

/\* Scatter plot matrix macro \*/

**%macro** scatterPlotMatrix(xVars=, title=, groupVar=);

**proc sgscatter** data=MYDATA.WELL\_COI\_OUTPUT\_DATA;

matrix &xVars / %if(&groupVar ne %str()) %then

%do;

group=&groupVar legend=(sortorder=ascending) %end;

diagonal=(histogram normal);

title &title;

**run**;

title;

**%mend** scatterPlotMatrix;

**%scatterPlotMatrix**(xVars=Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect

Net\_Difference PA\_Candidate,

title="Scatter plot matrix grouped by Well\_Status", groupVar=Well\_Status);

**proc sort** data=MYDATA.WELL\_COI\_OUTPUT\_DATA out=WORK.TempSorted4877;

by Well\_Status;

**run**;

**proc boxplot** data=WORK.TempSorted4877;

plot (Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate)\*Well\_Status / boxstyle=schematic;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted4877;

**run**;

SAS Code for the Summary Statistics of the R Output Dataset

/\*

\*

\* Task code generated by SAS Studio 3.8

\*

\* Generated on '3/9/23, 11:38 AM'

\* Generated by 'u50204747'

\* Generated on server 'ODAWS01-USW2.ODA.SAS.COM'

\* Generated on SAS platform 'Linux LIN X64 3.10.0-1062.9.1.el7.x86\_64'

\* Generated on SAS version '9.04.01M7P08062020'

\* Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Sa

\* Generated on web client 'https://odamid-usw2.oda.sas.com/SASStudio/main?locale=en\_US&zone=GMT-07%253A00&ticket=ST-118825-j

\*

\*/

ods noproctitle;

ods graphics / imagemap=on;

footnote "Report Created on %sysfunc(today(), MMDDYYD10.) at %sysfunc(time(), timeampm.)";

**proc means** data=MYDATA.WELL\_COI\_OUTPUT\_DATA chartype mean std min max median n

vardef=df skewness kurtosis q1 q3 qmethod=os;

var Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate;

class Well\_Status;

**run**;

**proc univariate** data=MYDATA.WELL\_COI\_OUTPUT\_DATA vardef=df noprint;

var Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate;

class Well\_Status;

histogram Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate / normal(noprint);

**run**;

**proc sort** data=MYDATA.WELL\_COI\_OUTPUT\_DATA out=WORK.TempSorted2236;

by Well\_Status;

**run**;

**proc boxplot** data=WORK.TempSorted2236;

plot (Oil\_Sales Gas\_Sales Well\_Tot\_Mill Well\_Asd\_Val H1000 H5280 Well\_PTax

Sum\_Home\_Asd\_Val Sum\_Home\_PTax Avg\_PTax\_Home Avg\_H1000\_NoWell H1000\_Effect

Well\_in\_View Avg\_H5280\_NoWell H5280\_Effect Ptax\_Effect Net\_Difference

PA\_Candidate)\*Well\_Status / boxstyle=schematic;

insetgroup mean stddev min max n q1 q2 q3 / position=top;

**run**;

**proc datasets** library=WORK noprint;

delete TempSorted2236;

**run**;

SAS Code for Normalization of R Output Parameters



SAS Code for the Multiple-Linear Regression Model using the R Output Dataset

