

**Technical Documentation (SAS 9.4)**

**By:**

[**Krishnakanth Panakarajupally**](http://www.linkedin.com/pub/krishnakanth-panakarajupally/62/93a/470/)

[**krishnakanth.590@gmail.com**](mailto:krishnakanth.590@gmail.com)

* **Creating library “Zappos,” to store datasets.**

LIBNAME Zappos 'C:\sas\myfolders\Zappos\' ;

* **Defining input data file name and location variable**

FILENAME Inp\_Data 'C:\Analytics\_Challenge\_Data\_2.csv' ;

* **Importing dataset:**

**DATA** Zappos.Dataset\_Crude;

INFILE Inp\_Data DELIMITER=',' FIRSTOBS = **2** MISSOVER DSD;

INFORMAT day mmddyy10. platform $15.;

INPUT day $ site $ new\_customer platform $ visits distinct\_sessions orders gross\_sales bounces

add\_to\_cart product\_page\_views search\_page\_views;

**RUN**;

* **Creating new data set named “Dataset\_Intermediate”, with metrics: conversion\_rate, bounce\_rate, add\_to\_cart\_rate:**

**DATA** Zappos.Dataset\_Intermediate;

SET Zappos.Dataset\_Crude;

Month = MONTH(day); \* Extracting month from date ;

day\_of\_week = WEEKDAY(day); \* Extracting day of the week from from date ;

conversion\_rate = orders / visits;

bounce\_rate = bounces / visits;

add\_to\_cart\_rate = add\_to\_cart / visits;

\* Condition to eliminate observations with missing values ;

WHERE day IS NOT NULL

AND site IS NOT NULL

AND new\_customer ^= **.**

AND new\_customer ^= **.**

AND platform IS NOT NULL

AND visits ^= **.**

AND distinct\_sessions ^= **.**

AND orders ^= **.**

AND gross\_sales ^= **.**

AND bounces ^= **.**

AND add\_to\_cart ^= **.**

AND product\_page\_views ^= **.**

AND search\_page\_views ^= **.**;

**RUN**;

* **Creating new dataset named Dataset\_NomissingValues, by eliminating records with null metric values.**

**DATA** Zappos.Dataset\_NomissingValues;

SET Zappos.Dataset\_Intermediate;

WHERE search\_page\_views ^= **.**

AND conversion\_rate ^= **.**

AND bounce\_rate ^= **.**

AND add\_to\_cart\_rate ^= **.**;

LABEL

day = 'Day'

site = 'Site'

new\_customer = 'New\_Customer?'

platform = 'Platform'

visits = 'Visits'

distinct\_sessions = 'Distinct Sessions'

orders = 'Orders'

gross\_sales = 'Gross Sales'

bounces = 'Bounces'

add\_to\_cart = 'Add to cart'

product\_page\_views = 'Prodcut page views'

search\_page\_views = 'Search page views'

conversion\_rate = 'Conversion rate'

bounce\_rate = 'Bounce rate'

add\_to\_cart\_rate = 'Add to cart rate';

**RUN**;

**Note:** SAS stores the date values as the number of days between January 1, 1960 and the date entered. Without applying format, the dates will be displayed as numerical values. In similar way, months and days of week are also stored as numerical values. Hence we create formats to format those values.

* **Creating SAS formats to format the values of Month, Day\_of\_week and New\_customer variables:**

**PROC** **FORMAT**;

VALUE MONTH\_FMT **1** = 'January'

**2** = 'February'

**3** = 'March'

**4** = 'April'

**5** = 'May'

**6** = 'June'

**7** = 'July'

**8** = 'August'

**9** = 'September'

**10** = 'October'

**11** = 'November'

**12** = 'December';

VALUE WEEK\_FMT **1** = 'Sunday'

**2** = 'Monday'

**3** = 'Tuesday'

**4** = 'Wednesday'

**5** = 'Thursday'

**6** = 'Friday'

**7** = 'Saturday';

VALUE CUST\_FMT **1** = 'Yes'

**0** = 'No';

VALUE WEEK\_FMT\_SHORT **1** = 'S'

**2** = 'M'

**3** = 'T'

**4** = 'W'

**5** = 'TH'

**6** = 'F'

**7** = 'SA';

**RUN**;

* **Applying the above created formats and creating final dataset for analysis:**

**DATA** Zappos.Dataset\_Final;

SET Zappos.Dataset\_NoMissingValues;

\* Applying formats ;

FORMAT day mmddyy10. Month MONTH\_FMT. day\_of\_week WEEK\_FMT.;

**RUN**;

* **Creating frequency distribution table:**

**PROC** **FREQ** DATA = Zappos.Dataset\_Final ;

TITLE "Frequency distribution table";

TABLES site platform month day\_of\_week / NOCUM ;

**RUN**;

* **Creating correlation matrix:**

**PROC** **CORR** DATA = Zappos.Dataset\_Final;

TITLE "Correlation matrix";

VAR visits orders gross\_sales distinct\_sessions bounces

add\_to\_cart product\_page\_views search\_page\_views conversion\_rate bounce\_rate add\_to\_cart\_rate;

**RUN;**

* **Creating bar plots:**

\* Distribution of Orders by month (Percentage of orders);

PATTERN1 COLOR=VIBG VALUE = S;

PATTERN2 COLOR=STRO VALUE = S;

axis1 label=(a=**90** f="Arial/Bold" "Total number of orders") minor=(n=**5**);

axis2 label= (f="Arial/Bold" "Month of the year");

axis3 label= (f="Arial/Bold" "New Customer?");

**PROC** **GCHART** DATA = Zappos.Dataset\_Final;

TITLE "Distribution of Orders by Month (Percentage of orders)";

VBAR Month / SUBGROUP = new\_customer

SUMVAR = Orders

TYPE = SUM DISCRETE

OUTSIDE = PERCENTSUM

INSIDE = PERCENTSUM

raxis=axis1

maxis =axis2

gaxis = axis3;

FORMAT new\_customer CUST\_FMT.;

**RUN**;

\* Distribution of Orders by platform ;

PATTERN1 COLOR=VIBG VALUE = S;

PATTERN2 COLOR=STRO VALUE = S;

axis1 label=( f="Arial/Bold" "Total number of orders") minor=(n=**5**);

axis2 label= (a=**90** f="Arial/Bold" "Platform");

**PROC** **GCHART** DATA = Zappos.Dataset\_Final;

TITLE "Distribution of Orders by Platform";

HBAR Platform /

SUBGROUP = new\_customer

SUMVAR = orders

OUTSIDE = SUM PERCENTSUM

SUMLABEL = 'Orders'

PCTSUMLABEL = 'Percent'

raxis=axis1

maxis =axis2;

FORMAT new\_customer CUST\_FMT.;

**RUN**;

\* Distribution of Orders by site ;

PATTERN1 COLOR=VIBG VALUE = S;

PATTERN2 COLOR=STRO VALUE = S;

axis1 label=(a=**90** f="Arial/Bold" "Total number of orders") minor=(n=**5**);

axis2 label= (f="Arial/Bold" "Platform");

axis3 label= (f="Arial/Bold" "New Customer?");

**PROC** **GCHART** DATA = Zappos.Dataset\_Final;

TITLE "Distribution of Orders by Site";

VBAR site /

SUBGROUP = new\_customer

SUMVAR = orders

OUTSIDE = PERCENTSUM

raxis=axis1

maxis =axis2

gaxis = axis3;

FORMAT new\_customer CUST\_FMT.;

**RUN**;

\* Distribution of Orders by day\_of\_week ;

PATTERN1 COLOR=VIBG VALUE = S;

PATTERN2 COLOR=STRO VALUE = S;

axis1 label=(a=**90** f="Arial/Bold" "Total number of orders") minor=(n=**5**);

axis2 label= (f="Arial/Bold" "Day of the week");

**PROC** **GCHART** DATA = Zappos.Dataset\_Final;

TITLE "Distribution of Orders by day\_of\_week";

VBAR day\_of\_week / DISCRETE

SUBGROUP = new\_customer

SUMVAR = orders

TYPE = SUM

OUTSIDE = PERCENTSUM

INSIDE = SUM

raxis=axis1

maxis =axis2;

FORMAT new\_customer CUST\_FMT.;

**RUN**;

\* Distribution of Orders by day\_of\_week, group by month ;

PATTERN1 COLOR=VIBG VALUE = S;

PATTERN2 COLOR=STRO VALUE = S;

axis1 label=(a=**90** f="Arial/Bold" "Total number of orders") minor=(n=**5**);

axis2 label= (f="Arial/Bold" "Day of the week");

axis2 label= (f="Arial/Bold" "Day of the week");

**PROC** **GCHART** DATA = Zappos.Dataset\_Final;

TITLE "Distribution of Orders by day\_of\_week";

VBAR day\_of\_week / GROUP = month DISCRETE

SUBGROUP = new\_customer

SUMVAR = orders

TYPE = SUM

OUTSIDE = PERCENTSUM

INSIDE = SUM

raxis=axis1

maxis =axis2;

FORMAT new\_customer CUST\_FMT. day\_of\_week Week\_fmt\_short.;

**RUN**;

* **Creating Scatterplots:**

SYMBOL VALUE = DOT COLOR = STRO;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Visits";

PLOT orders\*visits;

**RUN**;

SYMBOL VALUE = DOT COLOR = RED;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Distinct Sessions";

PLOT orders\*distinct\_sessions;

**RUN**;

SYMBOL VALUE = DOT COLOR = VIBG;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Bounces";

PLOT orders\*bounces;

**RUN**;

SYMBOL VALUE = DOT COLOR = OliveDrab;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Add to cart";

PLOT orders\*add\_to\_cart;

**RUN**;

SYMBOL VALUE = DOT COLOR = DarkGoldenrod;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Product page views";

PLOT orders\*product\_page\_views;

**RUN**;

SYMBOL VALUE = DOT COLOR = DarkViolet;

**PROC** **GPLOT** DATA = Zappos.Dataset\_Final;

TITLE "Orders Vs Search Page Views";

PLOT orders\*search\_page\_views;

**RUN**;

* **Analysis of Variance (ANOVA):** Checking if there is significant association between quantitative variable: Orders and Categorical variables (Site, Platform, Month, Day of week)

**PROC** **ANOVA** DATA = Zappos.Dataset\_Final;

TITLE "Orders vs Site";

CLASS site;

MODEL orders = site;

MEANS site / snk;

**RUN**;

**QUIT**;

**PROC** **ANOVA** DATA = Zappos.Dataset\_Final;

TITLE "Orders vs Platform";

CLASS platform;

MODEL orders = platform;

MEANS platform / snk;

**RUN**;

**QUIT**;

**PROC** **ANOVA** DATA = Zappos.Dataset\_Final;

TITLE "Orders vs Month";

CLASS month;

MODEL orders = month;

MEANS month / snk;

**RUN**;

**QUIT**;

**PROC** **ANOVA** DATA = Zappos.Dataset\_Final;

TITLE "Orders vs Day of Week";

CLASS day\_of\_week;

MODEL orders = day\_of\_week;

MEANS day\_of\_week / snk;

**RUN**;

**QUIT**;

**Note:** In order to use categorical variables in regression model, they must be converted into dummy variables. Reference coding is used for dummy variables. If there are N number of values for a categorical variable, N-1 number of variables will be created.

* **Creating dummy variables:**

**DATA** Zappos.Dataset\_final\_dummies;

SET Zappos.Dataset\_final;

\* Dummy variables for site ;

IF site = 'Acme' THEN

DO

site\_botly = **0** ;

site\_pinnacle = **0** ;

site\_sortly = **0** ;

site\_tabular = **0** ;

site\_widgetry = **0**;

END;

ELSE IF site = 'Botly' THEN

DO

site\_botly = **1** ;

site\_pinnacle = **0** ;

site\_sortly = **0** ;

site\_tabular = **0** ;

site\_widgetry = **0**;

END;

ELSE IF site = 'Pinnacle' THEN

DO

site\_botly = **0** ;

site\_pinnacle = **1** ;

site\_sortly = **0** ;

site\_tabular = **0** ;

site\_widgetry = **0**;

END;

ELSE IF site = 'Sortly' THEN

DO

site\_botly = **0** ;

site\_pinnacle = **0** ;

site\_sortly = **1** ;

site\_tabular = **0** ;

site\_widgetry = **0**;

END;

ELSE IF site = 'Tabular' THEN

DO

site\_botly = **0** ;

site\_pinnacle = **0** ;

site\_sortly = **0** ;

site\_tabular = **1** ;

site\_widgetry = **0**;

END;

ELSE IF site = 'Widgetry' THEN

DO

site\_botly = **0** ;

site\_pinnacle = **0** ;

site\_sortly = **0** ;

site\_tabular = **0** ;

site\_widgetry = **1**;

END;

\* Dummy variables for Platform ;

IF platform = 'Android' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'BlackBerry' THEN

DO

platform\_blackberry = **1** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'ChromeOS' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **1**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'iOS' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **1**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'iPad' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **1**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'iPhone' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **1**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'Linux' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **1**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'Macintosh' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **1**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'MacOSX' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **1**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'Other' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **1**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'Unknown' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **1**;

platform\_windows = **0**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'Windows' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **1**;

platform\_windowsphone = **0**;

END;

ELSE IF platform = 'WindowsPhone' THEN

DO

platform\_blackberry = **0** ;

platform\_chromeos = **0**;

platform\_ios = **0**;

platform\_ipad = **0**;

platform\_iphone = **0**;

platform\_linux = **0**;

platform\_macintosh = **0**;

platform\_macosx = **0**;

platform\_other = **0**;

platform\_unknown = **0**;

platform\_windows = **0**;

platform\_windowsphone = **1**;

END;

\* Dummy variables for Month ;

IF month = **1** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **2** THEN

DO

month\_february = **1**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **3** THEN

DO

month\_february = **0**;

month\_march = **1**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **4** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **1**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **5** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **1**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **6** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **1**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **7** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **1**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **8** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **1**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **9** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **1**;

month\_october= **0**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **10** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **1**;

month\_november= **0**;

month\_december= **0**;

END;

IF month = **11** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **1**;

month\_december= **0**;

END;

IF month = **12** THEN

DO

month\_february = **0**;

month\_march = **0**;

month\_april = **0**;

month\_may= **0**;

month\_june= **0**;

month\_july= **0**;

month\_august= **0**;

month\_september= **0**;

month\_october= **0**;

month\_november= **0**;

month\_december= **1**;

END;

**RUN**;

* **Generating regression model:**

SYMBOL VALUE = DOT COLOR = RED;

**PROC** **REG** DATA = Zappos.Dataset\_final\_dummies OUTEST = Zappos.Estimates PLOTS(MAXPOINTS = **11000**);

Predicted\_Sales: MODEL gross\_sales = new\_customer visits bounce\_rate add\_to\_cart\_rate

site\_botly

site\_pinnacle

site\_sortly

site\_tabular

site\_widgetry

platform\_blackberry

platform\_ios

platform\_ipad

platform\_iphone

platform\_linux

platform\_chromeos

platform\_macintosh

platform\_macosx

platform\_other

platform\_unknown

platform\_windows

platform\_windowsphone

month\_february

month\_march

month\_april

month\_may

month\_june

month\_july

month\_august

month\_september

month\_october

month\_november

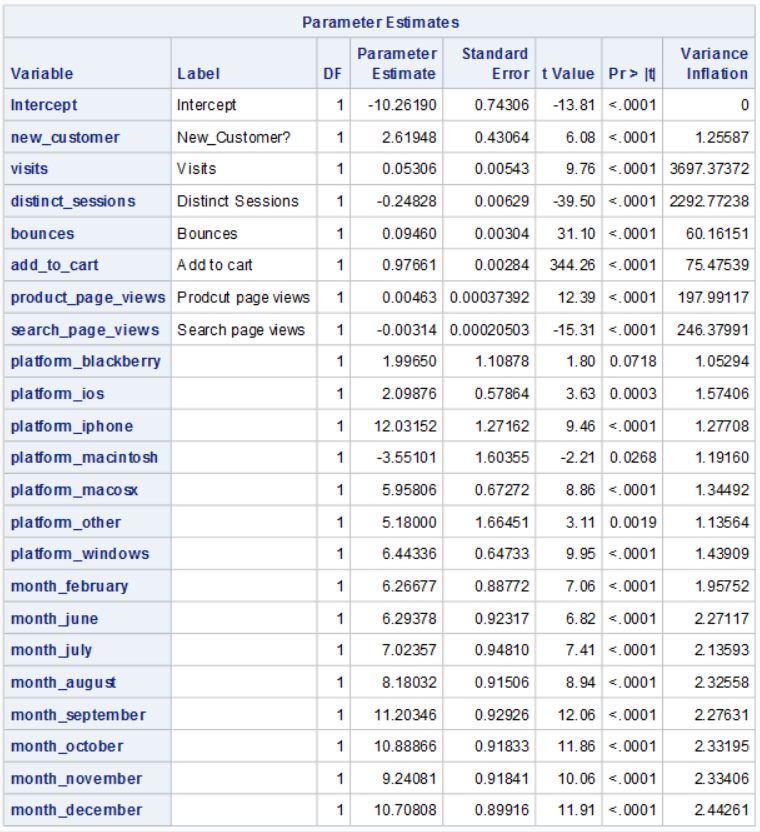
month\_december / SELECTION = stepwise VIF;

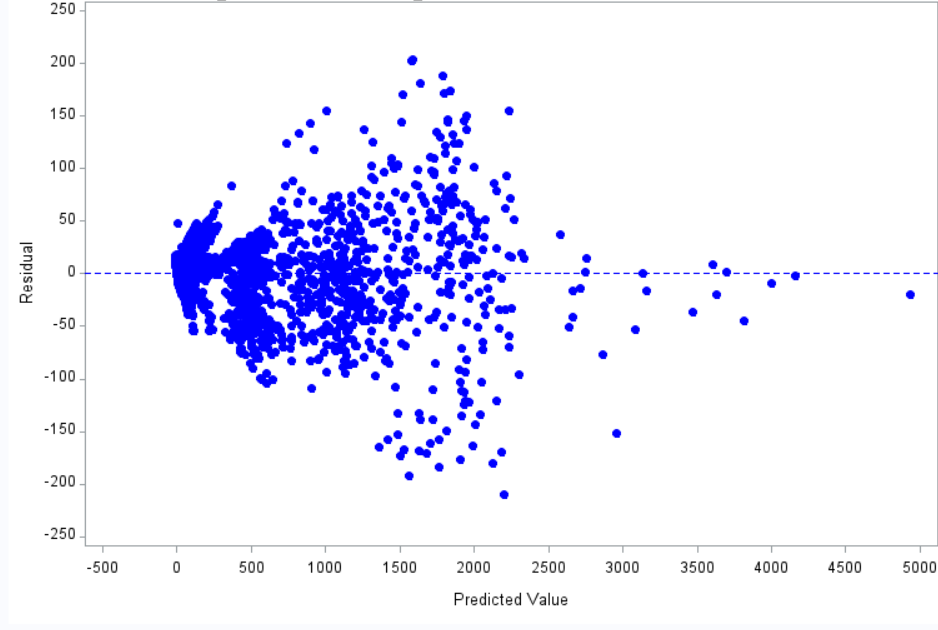
PLOT RESIDUAL.\*PREDICTED.;

**RUN**;

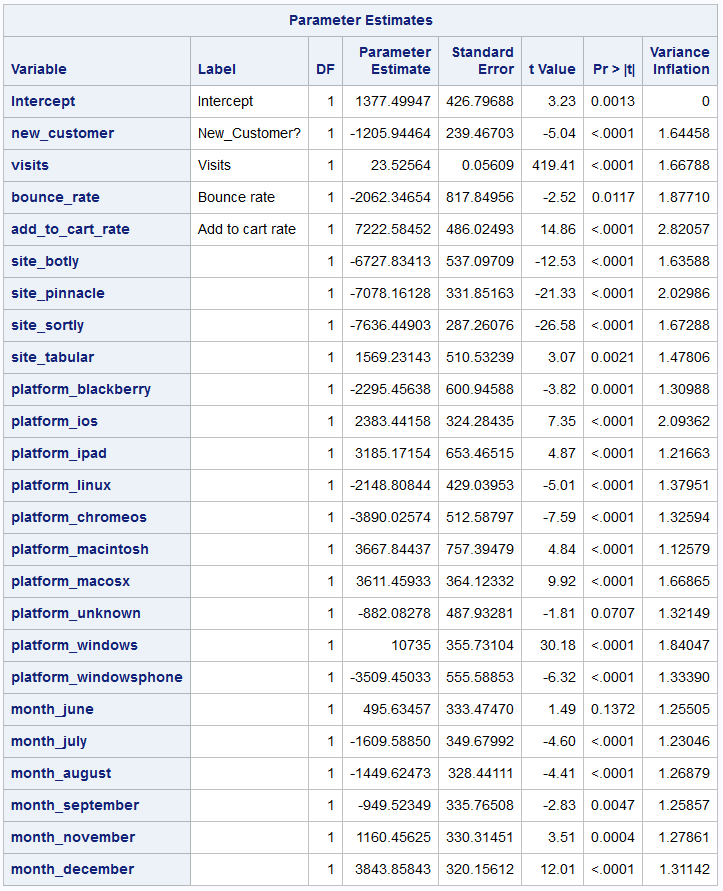
**QUIT**;

**Note:**

* The regression coefficients are saved in a dataset named “Estimates”, so that we can use the coefficients to score the model.
* Several regression models were created by using several combinations of independent variables and the best model was presented in the presentation.
* Acceptable Variance Inflation Factor (VIF) for regression coefficients is assumed as 10.
* Below is the screenshot of the regression coefficients which was generated by including all the variables in the model:
  + 
* In the residual plot, a pattern can be observed, which shows that the prediction error increases as the predicted value increases.



* Since there is high multicollinearity between independent variables, the Variance Inflation Factor (VIF) of regression coefficients is exceeding the acceptable range. Hence, some of the variables were eliminated to make the model stable.
* Below is the screenshot of regression coefficients after eliminating redundant independent variables:

****

* **Scoring the regression model using PROC SCORE:**

**PROC** **SCORE** DATA = Zappos.Dataset\_final\_dummies SCORE = Zappos.Estimates TYPE = PARMS OUT = Zappos.Scored\_Data;

VAR new\_customer visits bounce\_rate add\_to\_cart\_rate

site\_botly

site\_pinnacle

site\_sortly

site\_tabular

platform\_blackberry

platform\_ios

platform\_ipad

platform\_linux

platform\_chromeos

platform\_macintosh

platform\_macosx

platform\_unknown

platform\_windows

platform\_windowsphone

month\_june

month\_july

month\_august

month\_september

month\_november

month\_december;

ID gross\_sales;

**RUN**;

**QUIT**;

* **Creating a dataset with predicted values and residuals:**

**DATA** Zappos.Residuals;

SET Zappos.scored\_data;

residual = gross\_sales-Predicted\_sales;

**RUN**;

* **Calculating the sum of residuals and drawing histogram of residuals.**

**PROC** **UNIVARIATE** DATA = Zappos.Residuals;

VAR residual;

HISTOGRAM residual / NORMAL;

**RUN**;