Paper 1188-2017

Where Does Cleopatra Really Belong? An Analysis of Slot Machine Placement and Performance Using SAS®

Stephen Tam, British Columbia Lottery Corporation

ABSTRACT

In the world of gambling, superstition drives behavior, which can be difficult to explain. Conflicting evidence suggests that slot machines, like BCLC's Cleopatra, perform well regardless of where they are placed on a casino floor. Other evidence disputes this, arguing that performance is driven by their strategic placement (for example, in high-traffic areas). We explore and quantify the location sensitivity of slot machines by leveraging SAS® to develop robust models. We test various methodologies and data import techniques (such as casino CAD floor plans) to unlock some of the nebulous concepts of player behavior, product performance, and superstition. By demystifying location sensitivity, key drivers of performance can be identified to aid in optimizing the placement of slot machines.

INTRODUCTION

For the most part, slot machine placement within a casino is more art than science. Casino operators rely on performance data together with their own experience and intuition to determine placement of slot machines. Several journals have published papers on topics relating to slot machine placement, with many agreeing with casino operators that visibility, traffic volume, as well as physical attributes such as placement within a bank of machines, are all factors which affect performance significantly. The popularity of the game theme is also an important driver of performance.

This paper seeks to address the location sensitivity of every slot machine, using proximities to different structures within a casino as inputs. We differentiate slot machines based on their game theme, game denomination and operating platform. By identifying slot machines which are location insensitive (for example, a slot machine which is so popular that it performs well regardless of where it is located on the casino floor), you could place these machines in less desirable locations, while filling the more desirable locations with machines which are location sensitive, thereby optimizing the placement of slot machines across the whole casino floor.

Working with a BCLC Service Partner, AutoCAD® was used to extract location coordinates of every slot machine and different structures within a casino. With the location data ready to be imported, here are the SAS procedures used:

1. IMPORT procedure: Imports location coordinates for slot machines and physical structures, such as casino entrance, cashier and windows.
2. SQL procedure: Extracts data from a data warehouse and calculates minimum straight line distance measurements to different physical structures for each slot machine.
3. TRANSPOSE procedure: Transposes distance information (long format/stacked) into a series of variables (wide format/unstacked).
4. MEANS procedure: Calculates monthly averages from previous years to normalize for seasonality effects in the model dataset.
5. SORT procedure: Sorts the model dataset based on the BY variable to ensure that the model dataset is arranged and ready for modelling.
6. CORR procedure: Examines correlation between variables to remove variables which are highly correlated from the models.
7. GLMSELECT procedure: Performs variable selection, builds general linear models for each BY group, and generates outputs which can be interpreted for the Business.

DATA PREPARATION

COLLECTING LOCATION DATA

For casinos in British Columbia, each location where a slot machine can be placed is given an “address” known as area-section-location (A-S-L). Using a tool such as AutoCAD, we can extract location coordinates from a casino floor plan for each A-S-L. Similarly, we can also extract these coordinates for any physical structures we are interested in examining. For this particular casino, we included the following physical structures in our dataset:

- Cashier
- Casino Entrance
- Emergency Exit
- Guest Services
- Race Book
- Restaurant Pub
- Self Serve Coffee Station
- Smoking Patio
- Table Game
- Washroom Entrance
- Window

IMPORTING LOCATION DATA

Once we have defined all macro variables at the beginning of the program, we can use the following code to import machine locations into SAS:

```
proc import         out =       machine_locations
                    datafile = "&file_path.&input_name."
                    dbms =      excel
                    replace;
       range =      "&sheet1_name.";
       getnames = yes;
       scantext = yes;
run;
```

	area	section	location	machine_x	machine_y
1	1	1	1	11	9
2	1	1	2	8	9
3	1	1	3	6	9

Table 1. Partial contents of machine_locations

The process is repeated for structure locations:

```
proc import         out =       structure_locations
                    datafile = "&file_path.&input_name."
                    dbms =      excel
                    replace;
       range =      "&sheet2_name.";
       getnames = yes;
       scantext = yes;
run;
```

	structure	structure_x	structure_y
1	Washroom Entrance	51	53
2	Guest Services	27	8
3	Cashier	150	72
4	Cashier	151	67
5	Cashier	151	62
6	Cashier	150	57
7	Restaurant Pub	126	25

Table 2. Partial contents of structure_locations

CALCULATING DISTANCE MEASUREMENTS

As we are looking for all possible combinations between each A-S-L and structure, we use PROC SQL to create the Cartesian Product between these two tables, and calculate the straight line distance between each combination using the Pythagorean equation:

```
proc sql;
     create table machine to structure distance as
     select     put(a.area, 1.)||"-"||put(a.section, 1.)
                ||"-"||put(a.location, 1.) as a_s_l
          ,     b.structure
          ,     b.structure x
          ,     b.structure y
          ,     sqrt((b.structure x - a.machine x) ** 2 +
                (b.structure y - a.machine_y) ** 2) as distance
                format = comma8.2
     from       machine locations a
          ,     structure locations b
     order by   calculated a_s_l
          ,     b.structure
          ,     b.structure x
          ,     b.structure_y
     ;
quit;
```

	a_s_l	structure	structure_x	structure_y	distance
1	1-1-1	Cashier	150	57	147.05
2	1-1-1	Cashier	150	72	152.61
3	1-1-1	Cashier	151	62	149.70
4	1-1-1	Cashier	151	67	151.54
5	1-1-1	Casino Entrance	40	0	30.36
6	1-1-1	Casino Entrance	54	0	43.93
7	1-1-1	Emergency Exit	66	142	143.92
8	1-1-1	Emergency Exit	123	142	173.88
9	1-1-1	Guest Services	27	8	16.03
10	1-1-1	Race Book	79	22	69.23
11	1-1-1	Restaurant Pub	126	25	116.11

Table 3. Partial contents of machine_to_structure_distance

For certain structures, there are multiple locations, such as the four cashiers that are present at this casino. We are interested in the shortest distance to each type of structure by taking the minimum distance from the A-S-L to each type of structure:

```
proc sql;
      create table machine_to_structure_minimum as
      select     a s l
            ,    structure
            ,    min(distance) as shortest distance format = comma8.2
      from       machine_to_structure_distance
      group by   a s l
            ,    structure
      ;
quit;
```

	a_s_l	structure	shortest_distance
1	1-1-1	Cashier	147.05
2	1-1-1	Casino Entrance	30.36
3	1-1-1	Emergency Exit	143.92
4	1-1-1	Guest Services	16.03
5	1-1-1	Race Book	69.23
6	1-1-1	Restaurant Pub	116.11
7	1-1-1	Self Serve Coffee Station	104.01
8	1-1-1	Smoking Patio	113.99
9	1-1-1	Table Game	95.85
10	1-1-1	Washroom Entrance	59.46
11	1-1-1	Window	11.70

Table 4. Partial contents of machine_to_structure_minimum

TRANSPOSING DATA TO BE USED FOR MODELLING

We are using the shortest distance to each structure as independent variables in our general linear models. As such, we need to transpose the data above from a "long" format into a "wide" format using PROC TRANSPOSE:

```
proc transpose  data =     machine to structure minimum
                out =      machine_to_structure_wide
                name =     a_s_l
                ;
     by         a s l;
     id         structure;
     var        shortest_distance;
run;
```

	a_s_l	Cashier	Casino Entrance	Emergency Exit	Guest Services
1	1-1-1	147.05	30.36	143.92	16.03
2	1-1-2	149.89	33.24	145.10	19.03

Table 5. Partial contents of machine_to_structure_wide

NORMALIZING PERFORMANCE DATA

We are now ready to join this dataset with daily performance data for each A-S-L retrieved from our data warehouse. As there is a heavy seasonality pattern in our data, we need to normalize our data using a monthly index, and then aggregate the data at a weekly level to take into account variation in performance by day of week. Here is an example of how we use PROC MEANS to do this:

```
proc means data =      summary_month_normalised
                       noprint
                       nway
                       ;
     class             month
                       ;
     var               meter coin in amt index
                       real net win amt index
                       games_played_cnt_index
                       ;
     where             year in ("2013", "2014", "2015")
                       ;
     output            out =      summary_normalised
                       (drop =     TYPE
                                  _FREQ_
                       )
                       mean =
                       ;
run;
```

MODEL PREPARATION

GROUPING SIMILAR SLOT MACHINES

In the world of slot machines, there are many game themes, and in a lot of cases, there are only one or two slot machines with the same game theme across the entire casino floor. If we built regression models for each game theme, we would run into issues with small sample sizes. On the other hand, if we built models for each game denomination (penny, nickel, and so on) or operating platform, we would have slot machines which are very dissimilar from each other within each group, and the performance of the model would suffer as a result.

Ideally, we want to group slot machines into homogenous groups so that, all other things being equal, the performance of each slot machine within the group would be similar. We worked with our Casino Product team to derive “hybrid groups” based on a combination of game theme, game denomination and operating platform for each slot machine. A new variable, hybrid_group, was generated in our dataset using a CASE statement in PROC SQL.

SORTING DATA PRIOR TO MODELLING

Before we run the dataset through PROC GLMSELECT, we need to sort the data if the BY groups (in this case, hybrid_group) are not already in ascending order. There are others options that you could choose in PROC GLMSELECT if you want to avoid doing this, but sorting the data is the most straightforward way to proceed.

EXAMINING CORRELATION BETWEEN INDEPENDENT VARIABLES

Another thing to watch out for is correlation between independent variables. Collinearity is a problem because the performance of your model could be affected, such as your parameter estimates. We use PROC CORR to examine this:

```
proc corr  data =     location_sensitivity
                      noprob
                      ;
           var        cashier
                      casino entrance
                      emergency exit
                      guest services
                      race book
                      restaurant pub
                      self serve coffee
                      smoking patio
                      table game
                      washroom
                      window
                      ;
run;
```

	cashier	casino_entrance	emergency_exit	guest_services	race_book
cashier	1.00000	-0.38685	0.28157	-0.55179	-0.01527
casino_entrance	-0.38685	1.00000	-0.91696	0.98021	0.89863
emergency_exit	0.28157	-0.91696	1.00000	-0.87584	-0.79258
guest_services	-0.55179	0.98021	-0.87584	1.00000	0.80711
race_book	-0.01527	0.89863	-0.79258	0.80711	1.00000

Display 1. Partial output from PROC CORR

The output from PROC CORR shows that casino_entrance is highly correlated with guest_services at this casino. It makes sense, as guest services is immediately beside the entrance to the casino. Based on this output, we removed several independent variables, and kept the ones below for our model:

- Cashier
- Casino Entrance
- Restaurant Pub
- Self Serve Coffee Station
- Smoking Patio
- Window

MODEL BUILDING AND INTEPRETATION

BUILDING MODELS USING PROC GLMSELECT

We are now ready to investigate whether we can predict the performance of each group of slot machines based on their location parameters. Since there are many combinations of bets for different slot machines, resulting in different amount of coin in (total wager) per spin, we look at the number of games played (spins) as the indicator for performance. PROC GLMSELECT performs stepwise variable selection by default, but other selection methods are also available. The code below shows some of the functionality PROC GLMSELECT provides in order to help you understand how the models are derived:

```
ods graphics on;
proc glmselect  data =     location_sensitivity_sorted
                plots =    all
                ;
     class      denom cd
                special_area
                ;
     model      games played =         denom cd
                                       special_area
                                       cashier
                                       casino entrance
                                       restaurant pub
                                       self serve coffee
                                       smoking_patio
                                       window
                                       /
                                       details =  all
                                       stats =    all
                ;
     by         hybrid_group
                ;
     where      days_in_week = 7
                ;
     output     out =                  location_sensitivity_glmselect
                p =                    yhat
                r =                    resid
                ;
     ods output ParameterEstimates =   location sensitivity estimates
                FitStatistics =        location_sensitivity_fit
                ;
run;
ods graphics off;
```

Here are explanations of some of the functionalities that were used:

- ODS Graphics is used in conjunction with PLOTS = ALL to graphically illustrate some of the modelling steps in this procedure.
- We specify two classification variables (denomination setting, and a binary indicator for whether the slot machine is located in a special area at the casino) using a CLASS statement.
- The MODEL statement indicates the dependent and independent variables, with two options DETAILS = ALL and STATS = ALL added in for you to examine each step of the modelling process.
- Models are built for each hybrid_group, using only complete weeks of data, and ignore weeks where there were fewer than seven days of data.
- The OUTPUT statement generates a replica of the original dataset, with the predicted value and its corresponding residual stored as additional variables in the dataset.
- Additional ODS Tables from the output, specifically model parameter estimates and model fit statistics, are stored as additional datasets.

INTEPRETING THE OUTPUT

The output from PROC GLMSELECT contains a lot of information, especially if you have the DETAILS = ALL option as it would output statistics at every step of the stepwise selection process. There are two key pieces of output that are especially relevant to us.

Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	Model R-Square	Adjusted R-Square
0	Intercept		1	1	0.0000	0.0000
1	casino_entrance		2	2	0.2826	0.2801
2	window		3	3	0.4328	0.4290
3	self_serve_coffee		4	4	0.4809	0.4755
4	smoking_patio		5	5	0.5293	0.5229*

Display 2. Partial Stepwise Selection Summary for hybrid_group = Cleopatra

*** denotes Optimal Value Criterion as determined by PROC GLMSELECT**

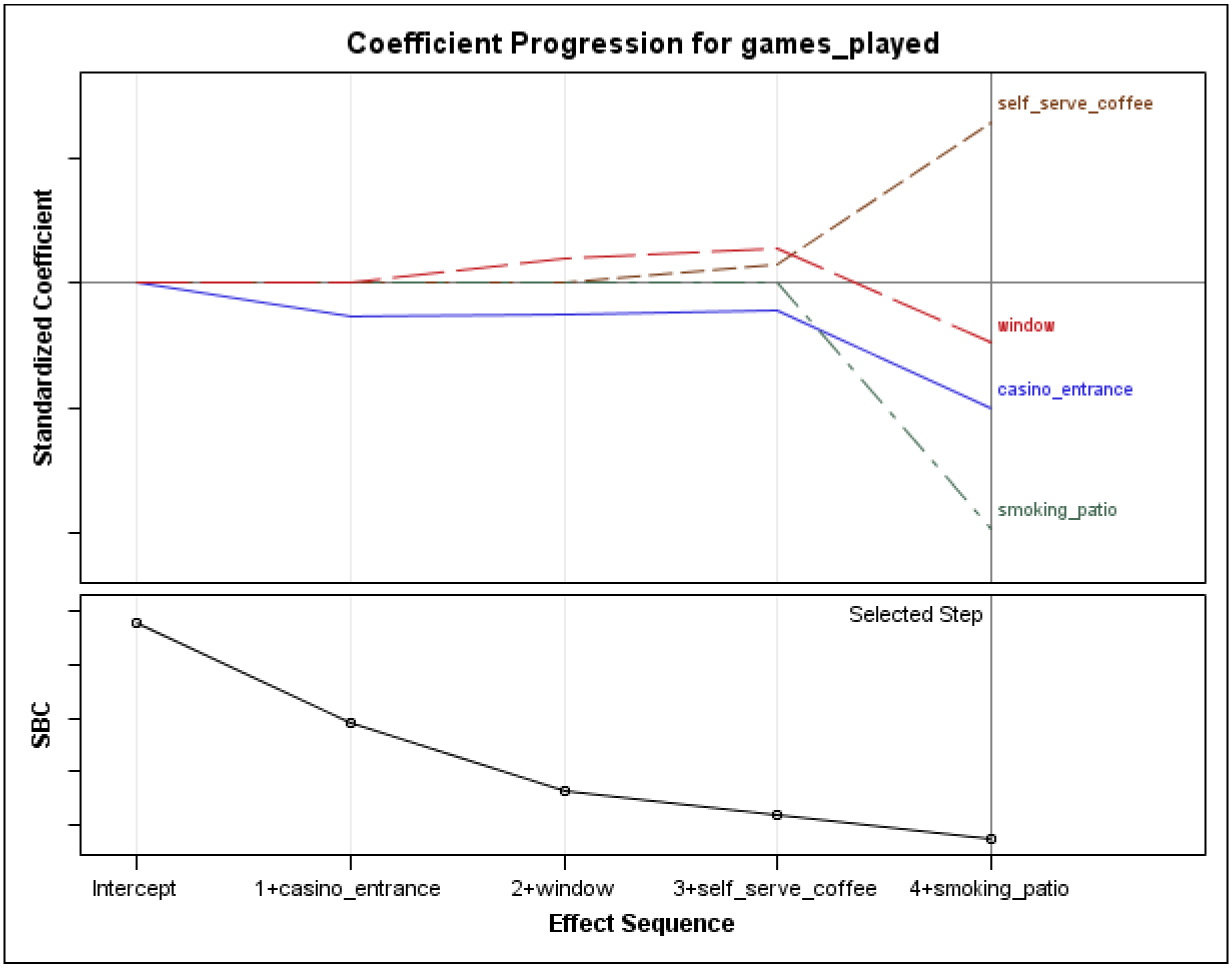
R-squared is a measure of fit for the chosen model, indicating the amount of variation that is explained by the model. This figure is adjusted as you add more independent variables into the model. In this particular example with Cleopatra, PROC GLMSELECT picks an optimum model with four independent variables which explains 52% of the variation in the data. If we had allowed only one independent variable, the model would have picked casino_entrance, explaining 28% of the variation. A model with casino_entrance and window as its independent variables would explain 43% of the variation, and so on.

Since adjusted R-squared explains how much variation is explained by our model, which is based on its location parameters, we use this metric as a proxy for a location sensitivity score, as a low R-squared indicates that the location parameters do not predict the number of games played accurately. This could be due to:

1. This group of slot machines is inherently location insensitive, therefore changing location parameters have very little effect on its performance.
2. Our data at this particular casino suggest there is little evidence of location sensitivity, which could be due to the lack of randomness in the way we placed the slot machines in the first place.

In the case of Cleopatra, a location sensitivity score of 52% is considered moderately location sensitive when we rank the scores for every hybrid group. As a comparison, another popular game theme in British Columbia, has a score of 22%. We interpret this as evidence that this other product is relatively location insensitive compared to Cleopatra. The scores for other hybrid groups can be found in the ODS table FitStatistics.

The location sensitive score on its own tells an incomplete story. By now, we know that Cleopatra is sensitive to certain locations, but how is it sensitive to these locations? In addition to understanding which location factors affect a slot machine's performance, it is also important to understand how each of the location factors affects the overall performance. Parameter estimates, or coefficients, are examined for this purpose.



Display 3. Partial Coefficient Panel for hybrid_group = Cleopatra

A positive coefficient means an increase in the regressor will lead to a positive change in the expected response. In the case of Cleopatra, self_serve_coffee has a positive coefficient, meaning an increase in the distance between the slot machine and the Self Serve Coffee Station is expected to result in an increase in games played, and you might want to put the slot machine as far away from it as possible. On the other hand, you might want to place Cleopatra machines as close to the Smoking Patio, Casino Entrance and Window as possible. ODS table ParameterEstimates contains all the information you would need to make similar interpretations for every hybrid group.

CONCLUSION

PROC GLMSELECT generates a series of ODS tables, which we can use to provide recommendations to the business. In this particular case, adjusted R-squared in FitStatistics determines the location sensitivity score for each hybrid group, and the corresponding coefficients showing the influence of each structure are found in ParameterEstimates.

We presented our findings to our Service Partner, and created a test proposal to validate the findings. Several banks of machines with varying degrees of location sensitivity were identified, and were swapped into different locations based on their parameter estimates. We are currently conducting this test, and results will be analyzed in order to obtain learnings for future models.

As mentioned before, this piece of analysis focused on one casino, and some of these findings might be specific to the site. Generalized models can be built with a regional dataset (for example, all casinos within Greater Vancouver), or even based on data for the whole province of British Columbia.

The approach could also be expanded to incorporate other variables which other studies have investigated before, including location descriptors such as proximity to major aisles, physical attributes such as signage, or hidden attributes like game volatility. This would provide a more comprehensive view of slot machine performance.

REFERENCES

Lucas, A. and Roehl, W. 2002. “Influences on Video Poker Machine Performance.” Journal of Travel & Tourism Marketing, 12:4:75–92

Lucas, A., Dunn, T., Roehl, W. and Wolcott, G. 2004. “Evaluating slot machine performance: A performance-potential model.” International Journal of Hospitality Management, 23:2:103–121

Lucas, A. and Dunn, W. 2005. “Estimating the Effects of Micro-Location Variables and Game Characteristics on Slot Machine Volume: A Performance-Potential Model” Journal of Hospitality & Tourism Research, 29:2:170–193

Thalheimer, R. and Ali, M. 2008. “Table games, slot machines and casino revenue.” Applied Economics, 40:18:2395–2404

Ghaharian, K. 2010. “A Mathematical approach for optimizing the casino slot floor: A linear programming application.” UNLV Theses, Dissertations, Professional Papers, and Capstones. 716.

ACKNOWLEDGMENTS

The author would like to thank BCLC and their casino operators for supporting his participation in SAS Global Forum 2017. He would also like to thank Michael Raithel, the author’s mentor for SGF, for his guidance and support in putting this paper together.

CONTACT INFORMATION

The author would love to receive any feedback on the paper, as well as any tips on better ways to tackle any of the tasks in the paper, especially ways to perform modelling. Please contact the author at:

Stephen Tam
British Columbia Lottery Corporation
stam@bclc.com
http://www.linkedin.com/in/s-tam/