***Final Case Study***

Libname Horse "Z:\Foundation Exercises\Assignments\Class17 - Final Case Study\Class17 - Final Case Study";

/\* Selecting the required Data by Year and Track id\*/

**data** race;

set horse.race;

date=datepart(race\_date);

year=year(date);

format date date9.;

if year=2005 or year=2006;

if track\_id="AP" or track\_id="FG" or track\_id="CRC";

**proc** **sort**;

by track\_id race\_date race\_number;

**run**;

**data** exotic\_payoff;

set horse.exotic\_payoff;

date=datepart(race\_date);

year=year(date);

format date date9.;

if year=2005 or year=2006;

if track\_id="AP" or track\_id="FG" or track\_id="CRC";

**proc** **sort** nodupkey;

by track\_id race\_date race\_number;

**run**;

/\*Merging Data\*/

**data** horse;

merge race exotic\_payoff;

by track\_id race\_date race\_number;

**run**;

/\* Data Preparation, dropping unnecessary variables\*/

**data** handle;

set horse;

handle=wps\_pool+total\_pool;

if breed\_type="QH" then delete;

if handle=. or handle=0 then delete;

drop division wager\_types wager\_types\_cancelled paid\_to\_others createDT modifyDT rectype

course\_type scheduled\_surface track\_sealed\_indicator

purse\_foreign paid\_to\_others added\_money\_foreign added\_money\_usa payoff\_amount reverts\_money guaranteed\_indicator start\_description abbreviated\_name race\_name program\_selections grade race\_class\_stake\_rule race\_class\_blktype\_rule graded\_stake\_id temporary\_rail\_distance race\_class\_stk\_rule\_wo\_classic carryover chute\_start ace\_class\_bt\_rule\_wo\_classic

number\_of\_rights winning\_numbers number\_of\_tickets\_bet;

**run**;

/\* Creating Dummy Variable \*/

**data** average;

set handle;

day=day(date);

month=month(date);

day\_of\_week=weekday(date);

if day\_of\_week=1 then sunday=1;

else sunday=0;

if day\_of\_week=2 then monday=1;

else monday=0;

if day\_of\_week=3 then tuesday=1;

else tuesday=0;

if day\_of\_week=4 then wednesday=1;

else wednesday=0;

if day\_of\_week=5 then thursday=1;

else thursday=0;

if day\_of\_week=6 then friday=1;

else friday=0;

if day\_of\_week=7 then saturday=1;

else saturday=0;**run**;

**data** restriction;

set average;

if sex\_restriction="B" then Fillies\_Mares=1;

else Fillies\_Mares=0;

if sex\_restriction="F" then Fillies=1;

else Fillies=0;

if age\_restriction="02" then age\_2\_years=1;

else age\_2\_years=0;

if age\_restriction="03" then age\_3\_years=1;

else age\_3\_years=0;

if age\_restriction="04" then age\_4\_years=1;

else age\_4\_years=0;

if age\_restriction="2U" then age\_2U\_years=1;

else age\_2U\_years=0;

if age\_restriction="34" then age\_34\_years=1;

else age\_34\_years=0;

if age\_restriction="35" then age\_35\_years=1;

else age\_35\_years=0;

if age\_restriction="3U" then age\_3U\_years=1;

else age\_3U\_years=0;

if age\_restriction="45" then age\_45\_years=1;

else age\_45\_years=0;

if age\_restriction="4U" then age\_4U\_years=1;

else age\_4U\_years=0;

if age\_restriction="5U" then age\_5U\_years=1;

else age\_5U\_years=0;

**run**;

**data** race\_type;

set restriction;

if race\_type="ALW" then ALLOWANCE=1;

else ALLOWANCE=0;

if race\_type="AOC" then ALLOWANCE\_OPTIONAL\_CLAIMING=1;

else ALLOWANCE\_OPTIONAL\_CLAIMING=0;

if race\_type="CLM" then CLAIMING=1;

else CLAIMING=0;

if race\_type="HCP" then HANDICAP=1;

else HANDICAP=0;

if race\_type="MCL" then MAIDEN\_CLAIMING=1;

else MAIDEN\_CLAIMING=0;

if race\_type="MSW" then MAIDEN\_SPECIAL\_WEIGHT=1;

else MAIDEN\_SPECIAL\_WEIGHT=0;

if race\_type="OCS" then OPTIONAL\_CLAIMING\_STAKES=1;

else OPTIONAL\_CLAIMING\_STAKES=0;

if race\_type="SHP" then STARTER\_HANDICAP=1;

else STARTER\_HANDICAP=0;

if race\_type="STK" then STARTER\_STAKE=1;

else STARTER\_STAKE=0;

if race\_type="STK" then STAKES=1;

else STAKES=0;

if race\_type="STK" then STARTER\_ALLOWANCE=1;

else STARTER\_ALLOWANCE=0;

**run**;

**data** distance\_id;

set race\_type;

if distance\_id=200 then dist\_200=1;

else dist\_200=0;

if distance\_id=450 then dist\_450=1;

else dist\_450=0;

if distance\_id=500 then dist\_500=1;

else dist\_500=0;

if distance\_id=550 then dist\_550=1;

else dist\_550=0;

if distance\_id=600 then dist\_600=1;

else dist\_600=0;

if distance\_id=650 then dist\_650=1;

else dist\_650=0;

if distance\_id=700 then dist\_700=1;

else dist\_700=0;

if distance\_id=750 then dist\_750=1;

else dist\_750=0;

if distance\_id=800 then dist\_800=1;

else dist\_800=0;

if distance\_id=818 then dist\_818=1;

else dist\_818=0;

if distance\_id=832 then dist\_832=1;

else dist\_832=0;

if distance\_id=850 then dist\_850=1;

else dist\_850=0;

if distance\_id=900 then dist\_900=1;

else dist\_900=0;

if distance\_id=950 then dist\_950=1;

else dist\_950=0;

if distance\_id=1000 then dist\_1000=1;

else dist\_1000=0;

if distance\_id>1000 then dist\_1200=1;

else dist\_1200=0;

**run**;

**data** track\_conditions;

set distance\_id;

if track\_condition="FT" then fast=1;

else fast=0;

if track\_condition="GD" then good=1;

else good=0;

if track\_condition="MY" then muddy=1;

else muddy=0;

if track\_condition="SY" then sloppy=1;

else sloppy=0;

if track\_condition="WF" then wet\_fast=1;

else wet\_fast=0;

if track\_condition="FM" then Firm=1;

else Firm=0;

if track\_condition="SF" then soft=1;

else soft=0;

if track\_condition="YL" then yielding=1;

else yielding=0;

**run**;

**data** weather;

set track\_conditions;

if weather="C" then CLear=1;

else CLear=0;

if weather="F" then Foggy=1;

else Foggy=0;

if weather="H" then Hazy=1;

else Hazy=0;

if weather="L" then Cloudy=1;

else CLoudy=0;

if weather="O" then Showery=1;

else Showery=0;

if weather="R" then rainy=1;

else Rainy=0;

**run**;

**data** number\_of\_runners;

set weather;

array race(\*) number\_of\_runners1-number\_of\_runners14;

do i=1 to 14;

race(i)=0;

end;

race(number\_of\_runners)=1;

drop i;

**run**;

**data** race\_number;

set number\_of\_runners;

array race(14) race\_number1-race\_number14;

do i=1 to 14;

race(i)=0;

end;

race(race\_number)=1;

drop i;

**run**;

**data** month;

set race\_number;

array calendar(12) Jan Feb Mar Arp May Jun Jul Aug Sep Oct Nov Dec;

do i=1 to 12;

calendar(i)=0;

end;

calendar(month)=1;

drop i;

**run**;

**data** track;

set month;

if surface="D" then Dirt=1;

else Dirt=0;

if about\_distance\_indicator="A" then about\_distance=1;

else about\_distance=0;

**run**;

**data** groups;

set track;

if purse\_usa<10000 then purse="Below 10k";

if 10000<=purse\_usa<15000 then purse="10k-15k";

if 15000<=purse\_usa<20000 then purse="15k-20k";

if 20000<=purse\_usa<25000 then purse="20k-25k";

if 25000<=purse\_usa<30000 then purse="25k-30k";

if 30000<=purse\_usa<35000 then purse="30k-35k";

if 35000<=purse\_usa<40000 then purse="35k-40k";

if 40000<=purse\_usa<=45000 then purse="40k-45k";

if purse\_usa>45000 then purse="45k above";

if handle<100000 then handle\_category="Below 100k";

if 100000<=handle<150000 then handle\_category="100k-150k";

if 150000<=handle<200000 then handle\_category="150k-200k";

if 200000<=handle<250000 then handle\_category="200k-250k";

if 250000<=handle<300000 then handle\_category="250k-300k";

if 300000<=handle<350000 then handle\_category="300k-350k";

if handle>350000 then handle\_category="350k above";

if 1100 le post\_time le 1159 then clock="11am-12pm";

if 1200 le post\_time le 1259 then clock="12pm-1pm";

if post\_time<=159 then clock="1pm-2pm";

if 200 le post\_time le 259 then clock="2pm-3pm";

if 300 le post\_time le 359 then clock="3pm-4pm";

if 400 le post\_time le 459 then clock="4pm-5pm";

if 500 le post\_time le 559 then clock="5pm-6pm";

if 600 le post\_time le 659 then clock="6pm-7pm";

if 700 le post\_time le 759 then clock="7pm-8pm";

**run**;

**data** date;

set groups;

if date='06feb2005'd or date='05feb2006'd then Superbowl=1;

else Superbowl=0;

if date='17mar2005'd or date='17mar2006'd then StPatrick=1;

else StPatrick=0;

if date='27mar2005'd or date='16apr2006'd then Easter=1;

else Easter=0;

if date='25mar2005'd or date='14apr2006'd then Good\_friday=1;

else good\_friday=0;

if date='05may2005'd or date='05may2006'd then Cinco\_deMayo=1;

else Cinco\_deMayo=0;

if date='30may2005'd or date='29may2006'd then Memorial\_day=1;

else Memorial\_day=0;

if date='04Jul2005'd or date='04Jul2006'd then Independence\_day=1;

else Independence\_day=0;

if date='05Sep2005'd or date='04sep2006'd then Labor\_day=1;

else Labor\_day=0;

if date='11nov2005'd or date='11nov2006'd then Vetern\_day=1;

else Vetern\_day=0;

if date='24Nov2005'd or date='23Nov2006'd then Thanksgiving=1;

else Thanksgiving=0;

if date='25dec2005'd or date='25dec2006'd then Christmas=1;

else Christmas=0;

if date='26dec2005'd or date='26dec2006'd then Boxing\_day=1;

else Boxing\_day=0;

if date='31dec2005'd or date='31dec2006'd then New\_year=1;

else New\_year=0;

if Superbowl=1 or StPatrick=1 or Easter=1 or Good\_friday=1 or Cinco\_deMayo=1 or Memorial\_day=1 or

Independence\_day=1 or Labor\_day=1 or Vetern\_day=1 or Thanksgiving=1 or Christmas=1 or Boxing\_day=1 or

New\_year=1 then holiday=1;

else holiday=0;

**run**;

**data** time;

set date;

if clock="11am-12pm" then time\_11am=1;

else time\_11am=0;

if clock="12pm-1pm" then time\_12pm=1;

else time\_12pm=0;

if clock="1pm-2pm" then time\_1pm=1;

else time\_1pm=0;

if clock="2pm-3pm" then time\_2pm=1;

else time\_2pm=0;

if clock="3pm-4pm" then time\_3pm=1;

else time\_3pm=0;

if clock="4pm-5pm" then time\_4pm=1;

else time\_4pm=0;

if clock="5pm-6pm" then time\_5pm=1;

else time\_5pm=0;

if clock="6pm-7pm" then time\_6pm=1;

else time\_6pm=0;

if clock="7pm-8pm" then time\_7pm=1;

else time\_7pm=0;

**run**;

/\* Exporting Data to draw insights using Excel \*/

**proc** **export** data=time

outfile="Y:\Documents\Projects\Class17 - Final Case Study\Assignment\handle.csv" replace;

**run**;

/\* Splitting the data by Track ID\*/

**data** AP CRC FG;

set time;

if track\_id="AP" then output AP;

if track\_id="CRC" then output CRC;

if track\_id="FG" then output FG;

**run**;

/\* Running Regression Model\*/

**proc** **reg** data=AP;

model handle= holiday friday saturday

STARTER\_STAKE

race\_number3 race\_number4 race\_number5 race\_number7

race\_number8 race\_number9 race\_number11

Dirt about\_distance

Jul Aug Sep

dist\_500 dist\_850 dist\_900 dist\_950 dist\_1000

number\_of\_runners4 number\_of\_runners7 number\_of\_runners8 number\_of\_runners9 number\_of\_runners10

number\_of\_runners11 number\_of\_runners12

sloppy soft Hazy /vif;

**run**;

**proc** **reg** data=CRC;

model handle= Dirt about\_distance

tuesday friday saturday Fillies

ALLOWANCE ALLOWANCE\_OPTIONAL\_CLAIMING MAIDEN\_CLAIMING STARTER\_STAKE

number\_of\_runners6 number\_of\_runners7 number\_of\_runners8 number\_of\_runners9 number\_of\_runners10

number\_of\_runners11 number\_of\_runners12

dist\_200 dist\_500 dist\_700 dist\_750 dist\_800 dist\_832 dist\_850

race\_number1 race\_number3 race\_number4 race\_number5 race\_number6 race\_number7

race\_number8 race\_number9 race\_number10 race\_number11

Jan Arp May Jul Nov Dec

Cinco\_deMayo Labor\_day Vetern\_day Thanksgiving Boxing\_day New\_year

age\_2\_years age\_3\_years age\_34\_years

fast/vif;

**run**;

**proc** **reg** data=FG;

model handle= saturday time\_3pm time\_5pm

ALLOWANCE ALLOWANCE\_OPTIONAL\_CLAIMING MAIDEN\_SPECIAL\_WEIGHT STARTER\_STAKE

number\_of\_runners6 number\_of\_runners7 number\_of\_runners8 number\_of\_runners9 number\_of\_runners10

number\_of\_runners11 number\_of\_runners12

dist\_600 dist\_800

race\_number1 race\_number3 race\_number4 race\_number5 race\_number6 race\_number7

race\_number8 race\_number9

age\_3\_years age\_4U\_years

yielding CLear Cloudy /vif;

**run**;