LEVERAGING PREDICTIVE ANALYTICS TO DRIVE STUDENT ACCESS & SUCCESS

Presented by,

Harshitha Akula

TEXAS A&M UNIVERSITY

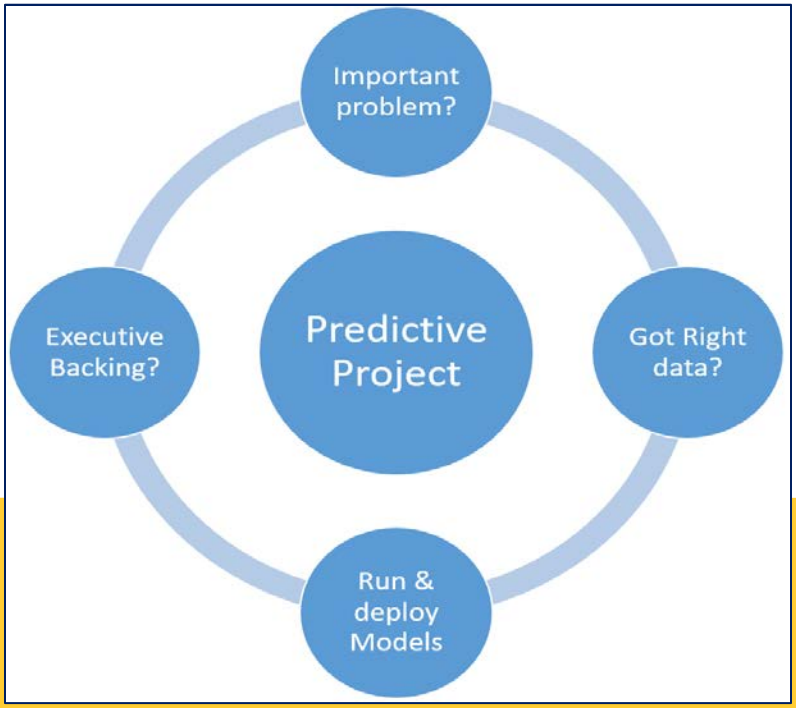
COMMERCE

PREDICTIVE ANALYTICS (BACKGROUND)

- University had an existing external predictive analytics provider
- Report users wanted to better predict student success
- Administrators wanted to better allocate staff time & resources
- IER wanted to build predictive models in-house

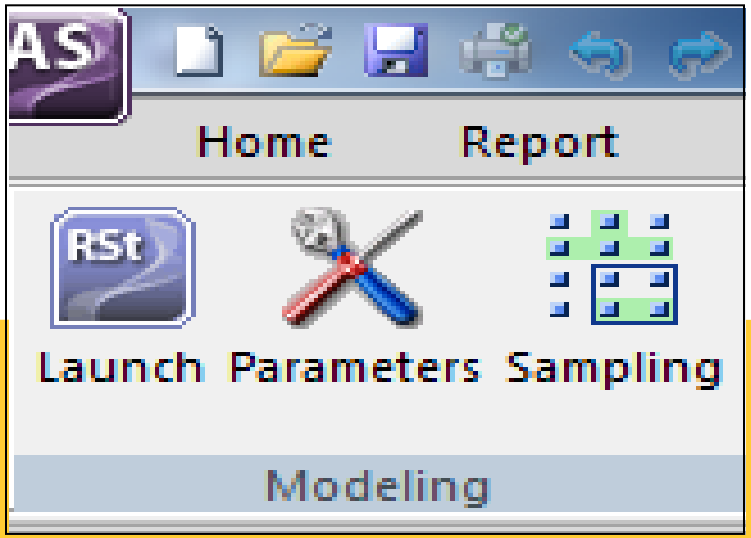
LOCALLY DEVELOPED

- No dependency on external providers
- Flexible & Cost efficient solution
- Leveraged existing university reporting software
- Partnership between IER & ICBE



WEBFOCUS R-STAT

- Add –on to existing university reporting software
- Leverages the power of R
- Graphical User Interface (GUI)
- Ease of predictive model deployment to existing user reports
- Better use of staff time & resources
- Limitless applications



INTERNAL USER EXAMPLE

- Enrollment Management Division:
  - VP of Enrollment Management
  - Admissions Director
  - Financial Aid
  - Registrar office
  - Admissions Recruiters

BUSINESS OBJECTIVE

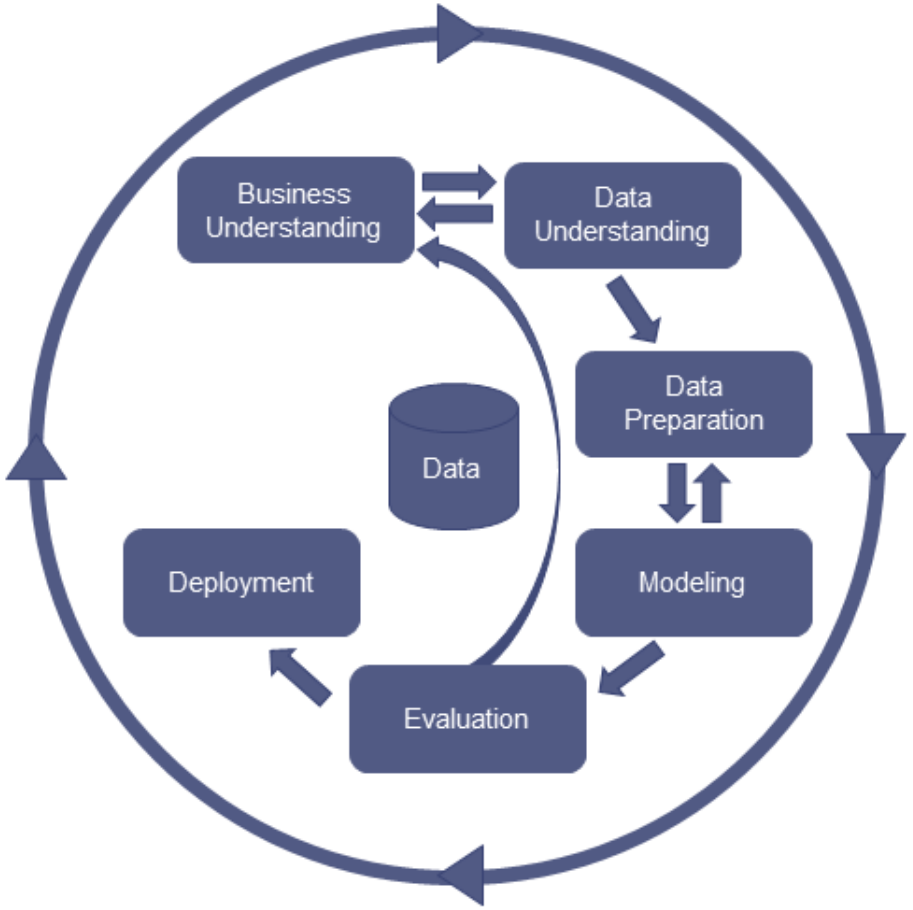
Increase freshmen enrollment for fall 2018

IER Solution:

- Create enrollment probabilities for all admitted freshmen students
- Deploy enrollment model to existing Enrollment Management reports
- Rank students by enrollment probabilities to focus on top enrollment targets

CRISP-DM Model

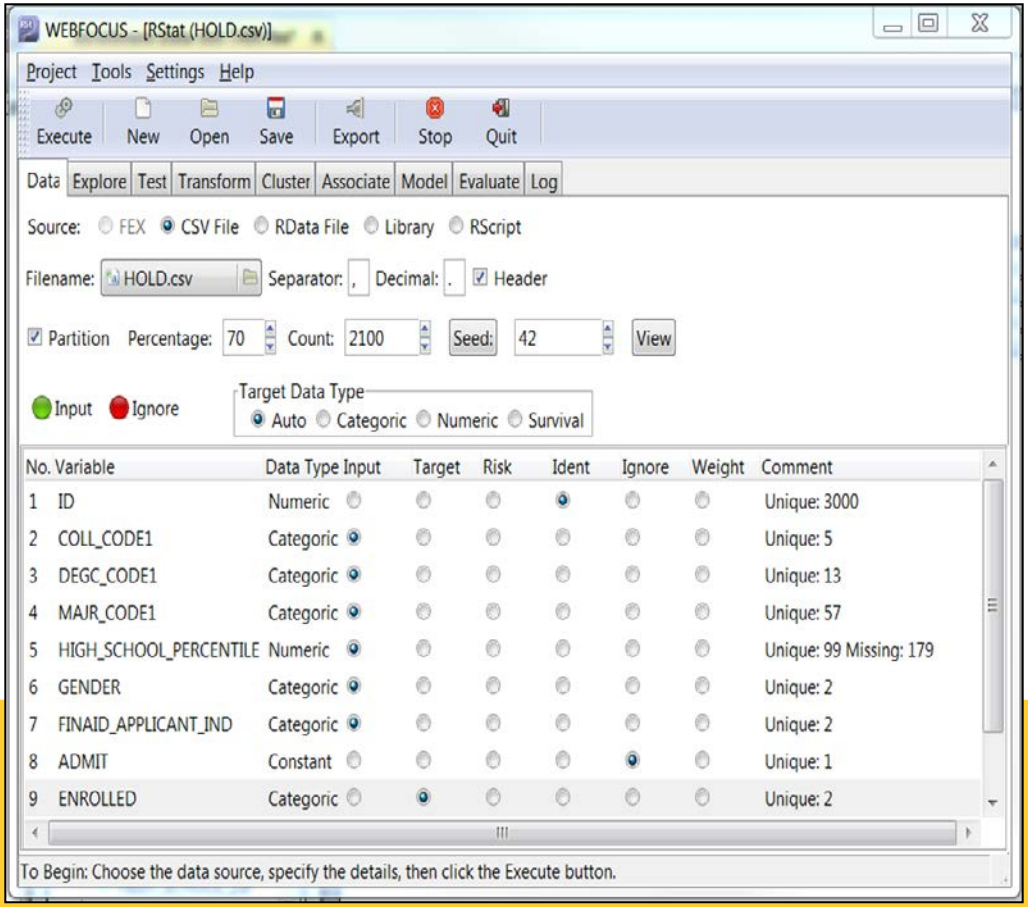
The model splits a data mining project into six phases and it allows for needing to go back and forth between different stages.



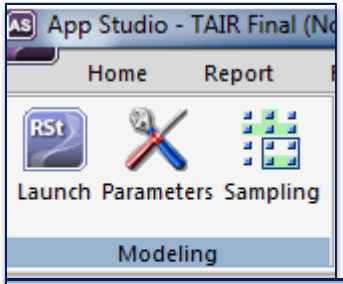
Cross Industry Process for Data Mining

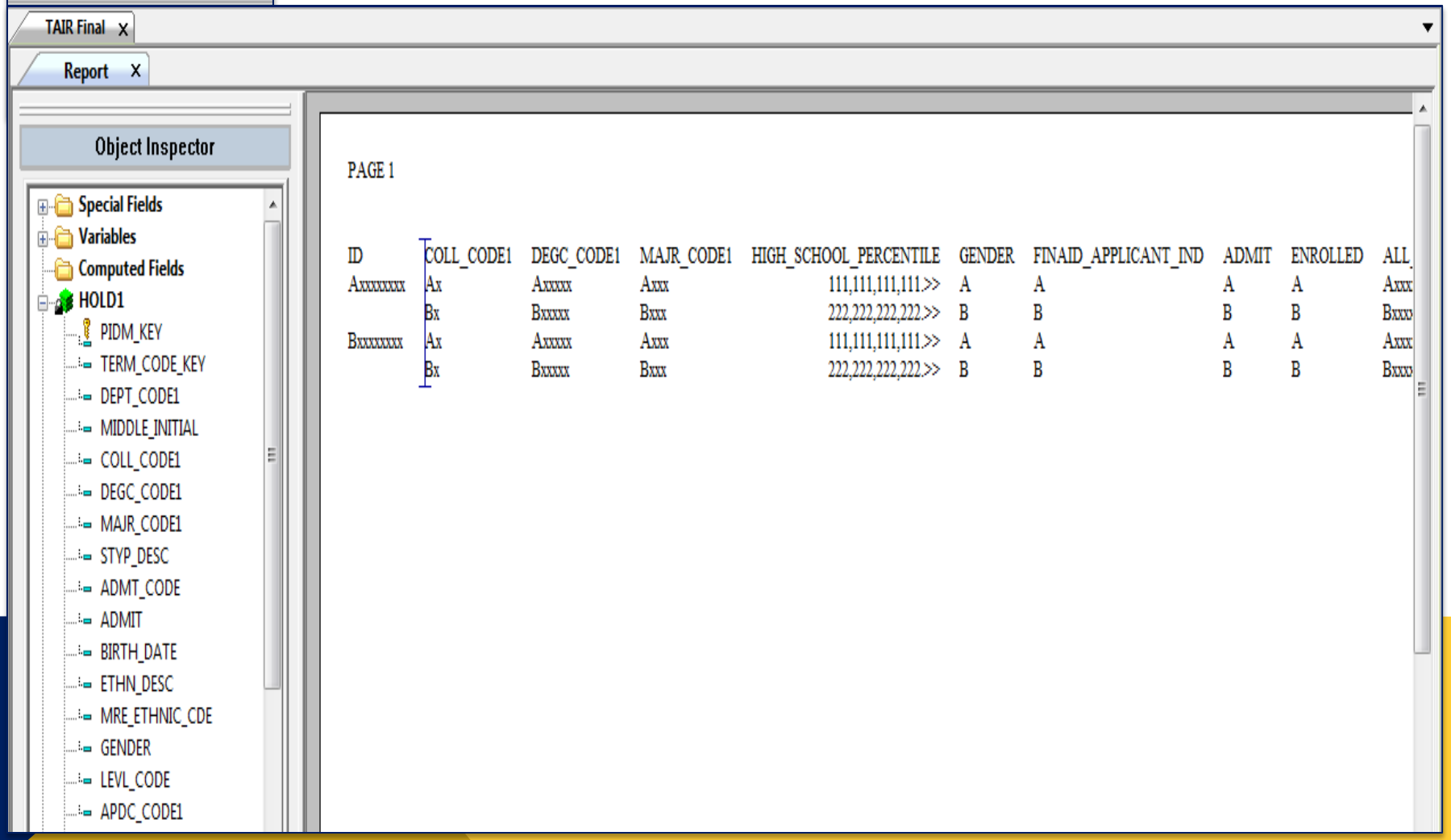
WEBFOCUS R-STAT

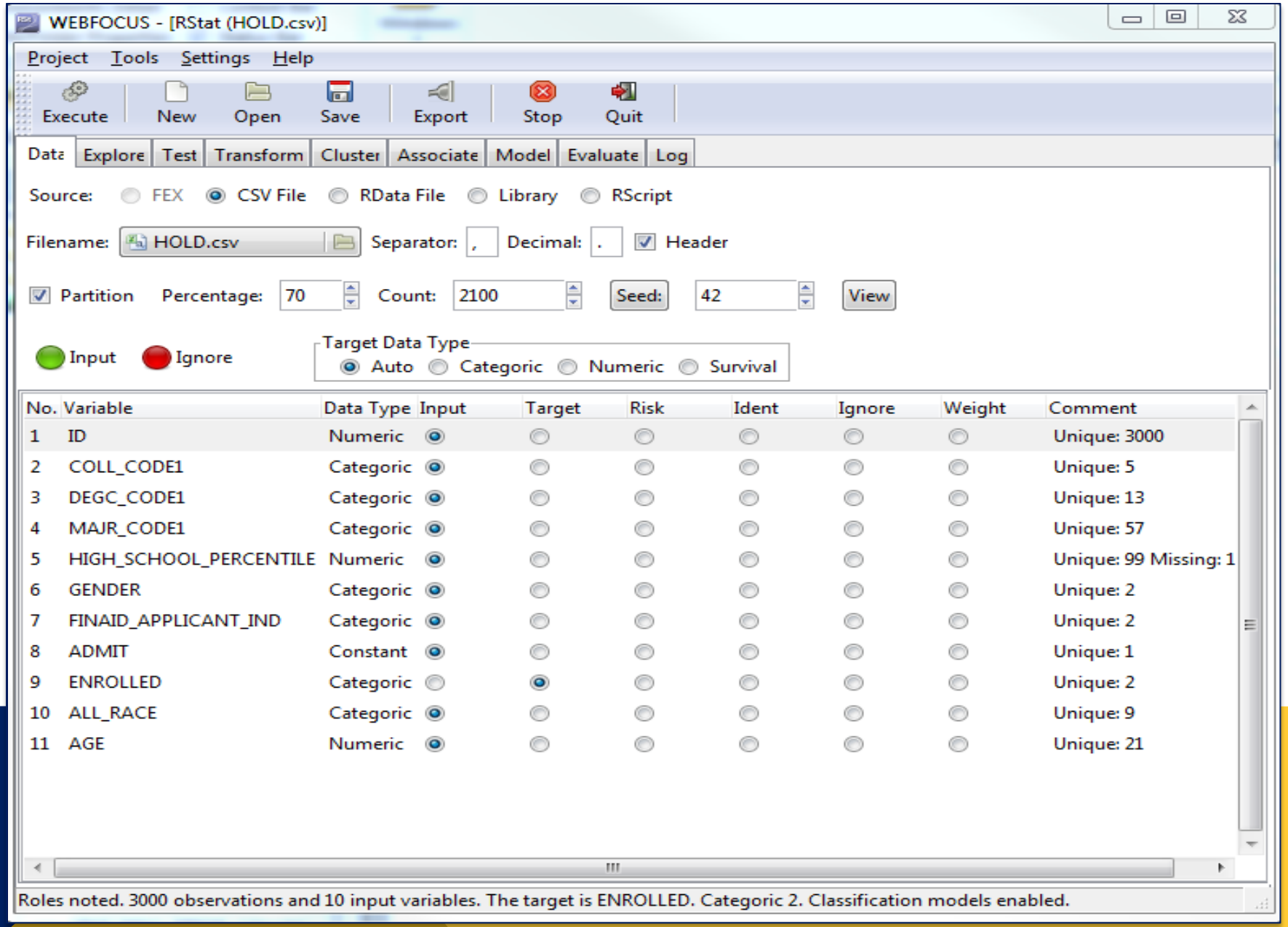
WebFOCUS R-STAT delivers powerful predictive analytics functionality. Business users can leverage a single integrated solution for BI, data modeling, and scoring, so they can make decisions based on accurate, validated future predictions instead of relying on gut instinct alone.

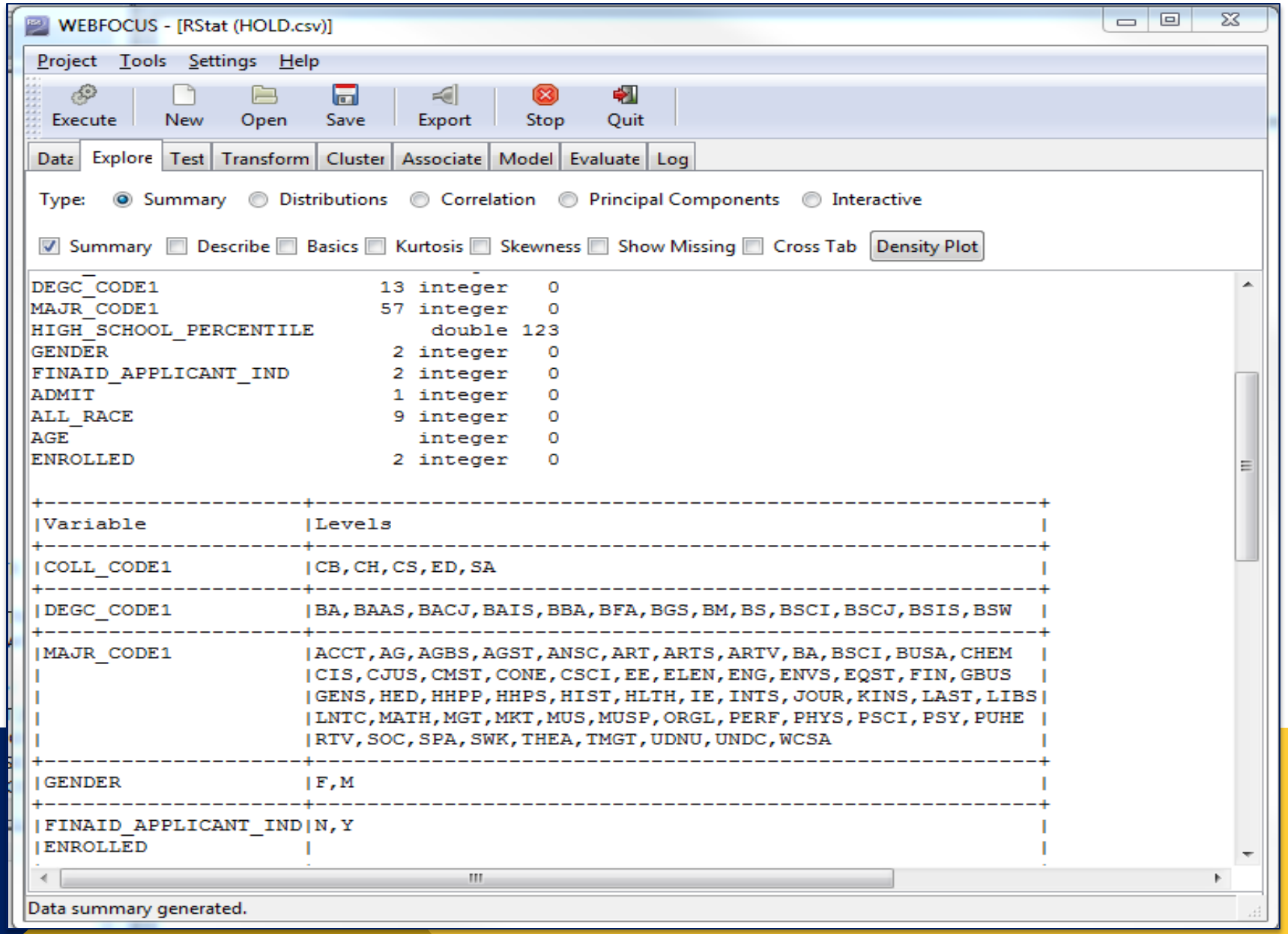


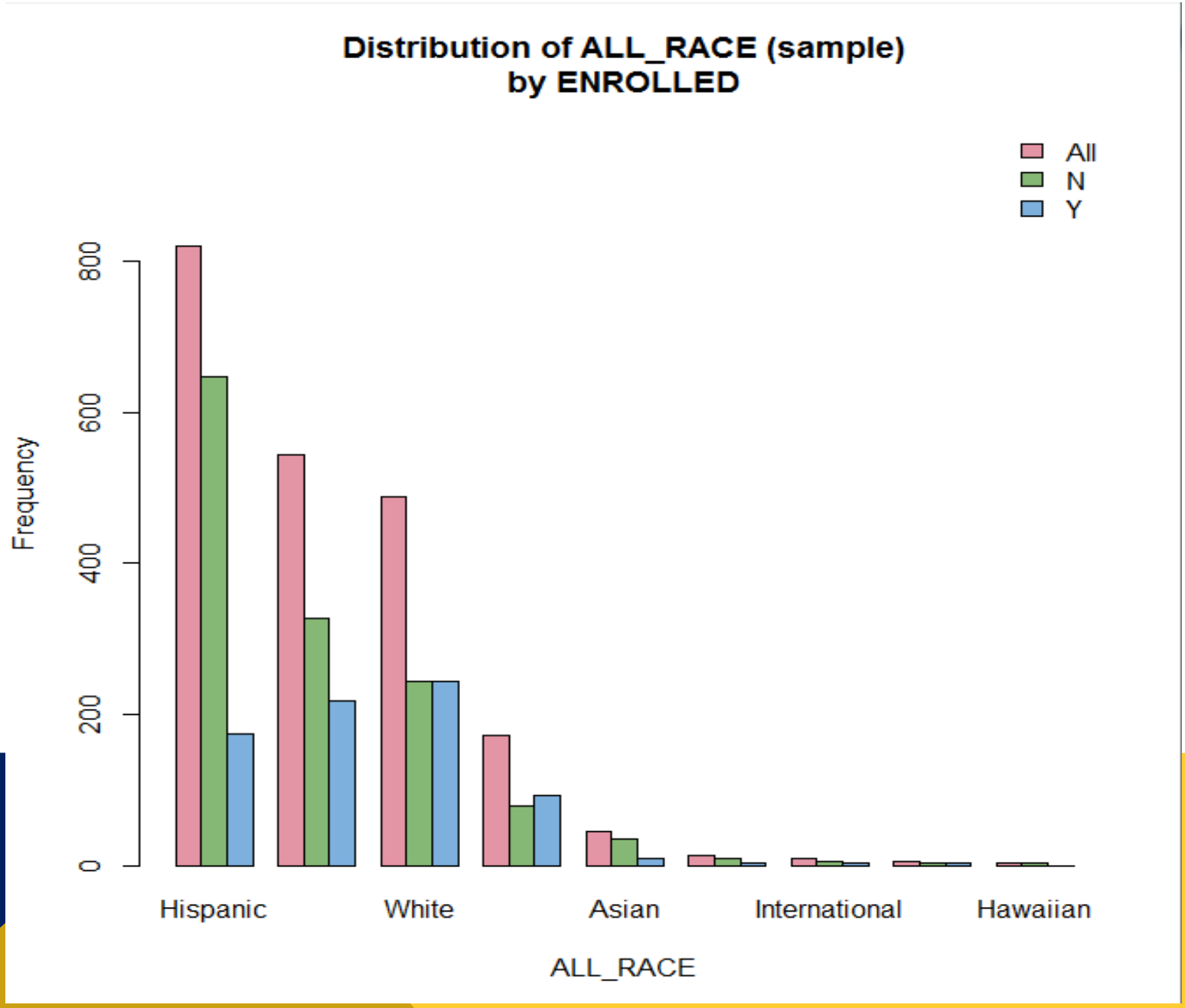
STEPS FOR R-STAT:

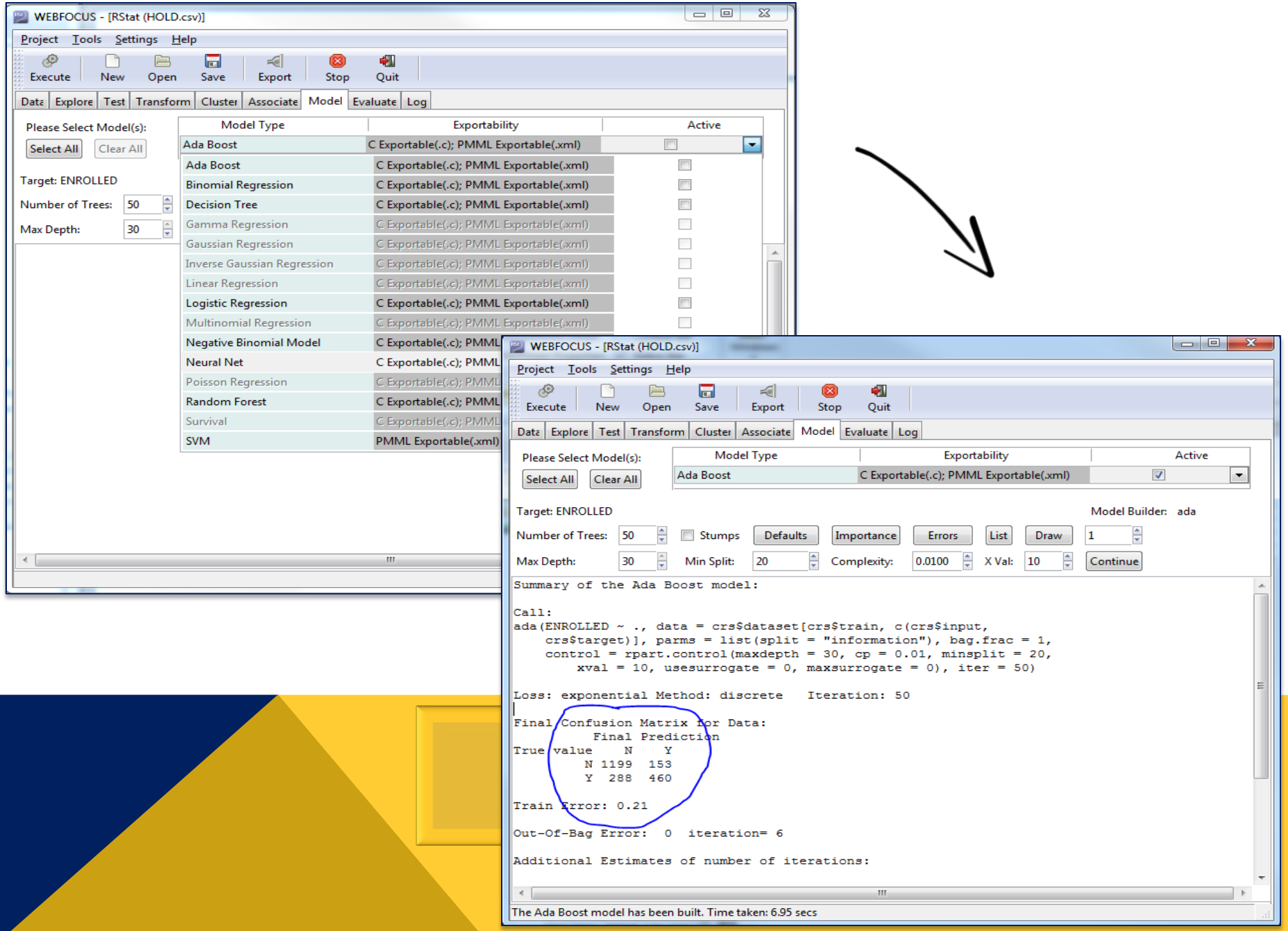


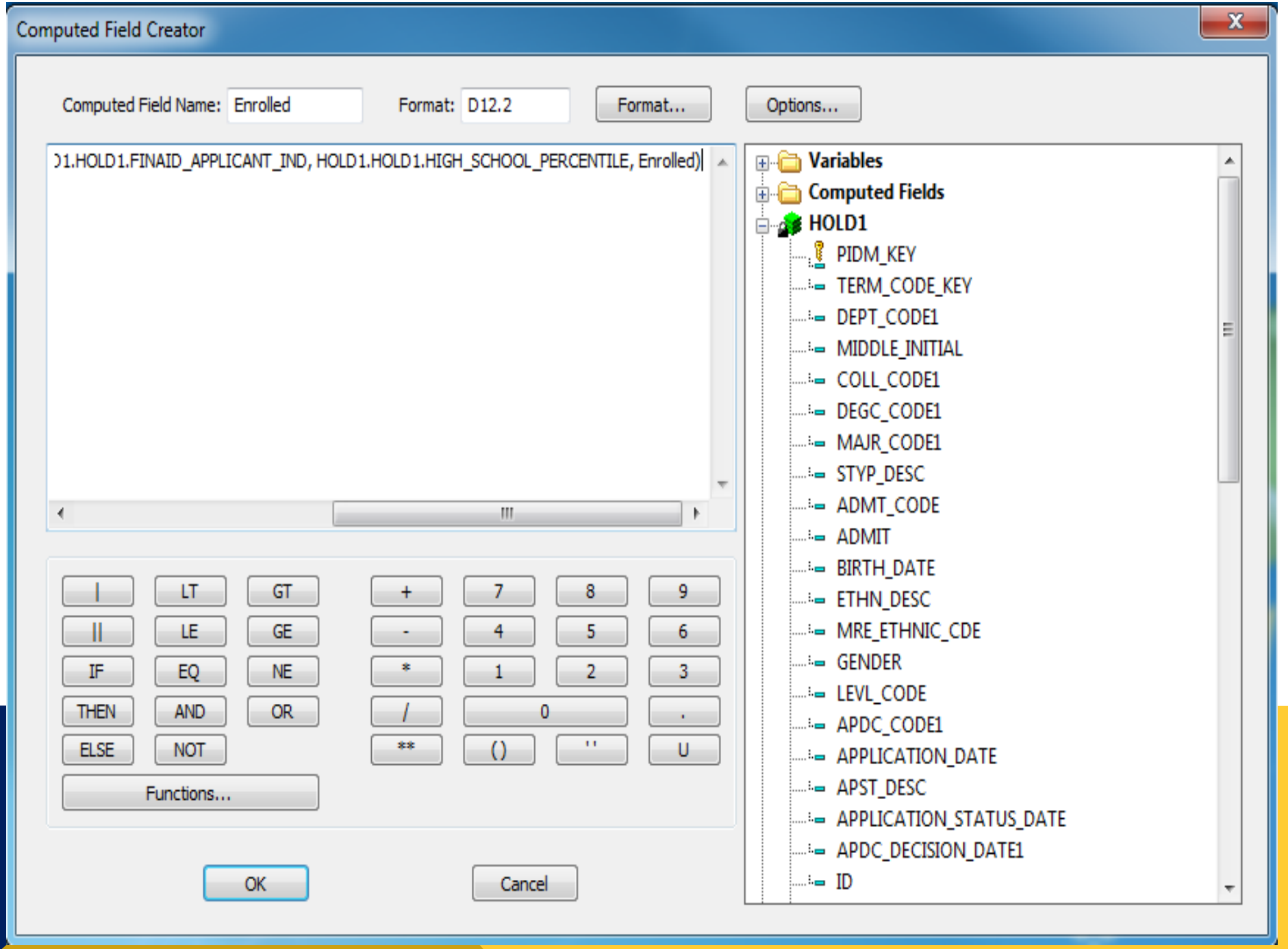


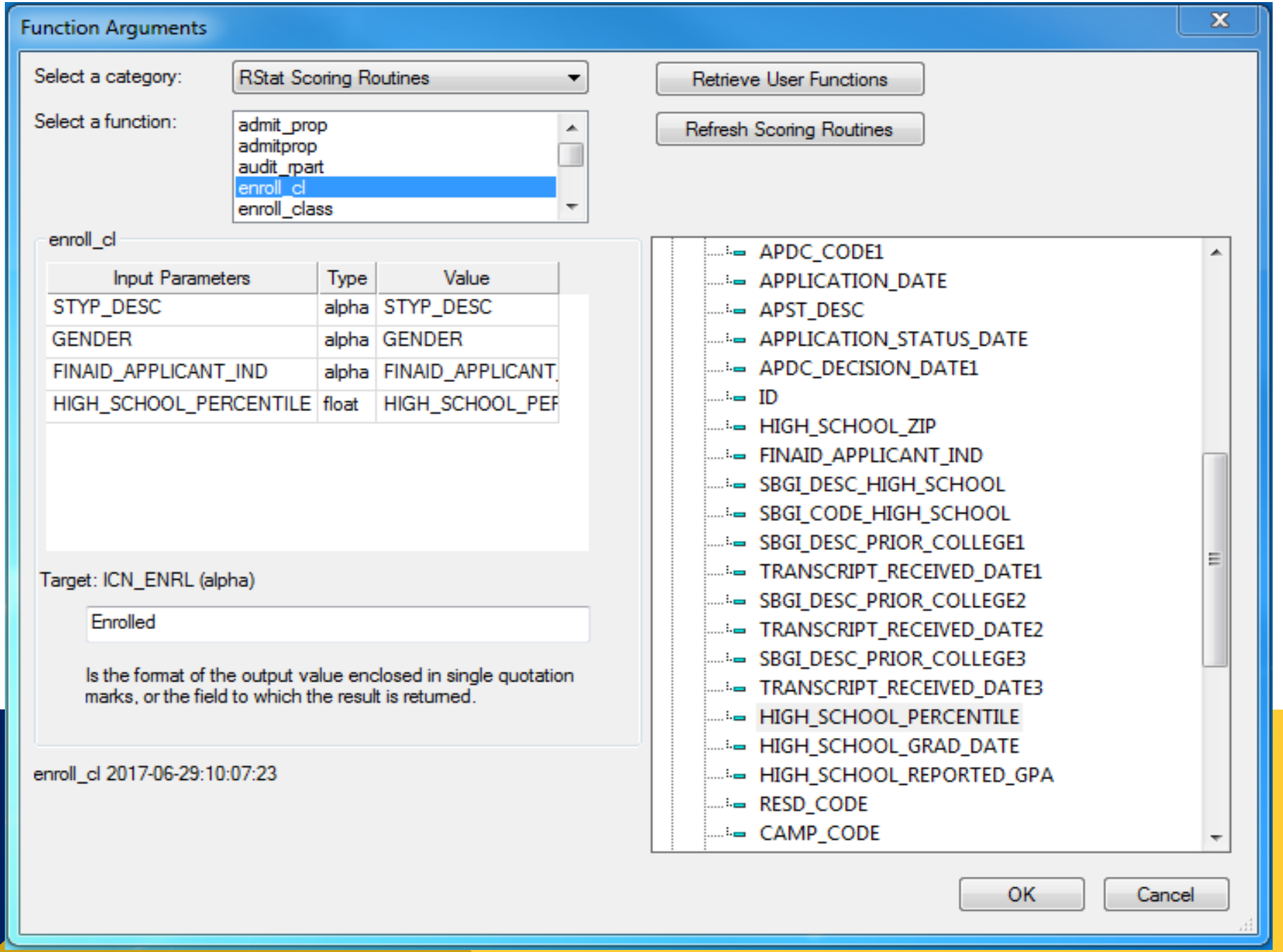

WEBFOCUS - [RStat (HOLD.csv)]
Project Tools Settings Help
Execute New Open Save Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Source: FEX CSV File RData File Library RScript
Filename: HOLD.csv Separator: , Decimal: . Header
Partition Percentage: 70 Count: 2100 Seed: 42 View
Input Ignore
Target Data Type
Auto Categoric Numeric Survival
No. Variable Data Type Input Target Risk Ident Ignore Weight Comment
1 ID Numeric Unique: 3000
2 COLL_CODE1 Categoric Unique: 5
3 DEGC_CODE1 Categoric Unique: 13
4 MAJR_CODE1 Categoric Unique: 57
5 HIGH_SCHOOL_PERCENTILE Numeric Unique: 99 Missing: 1
6 GENDER Categoric Unique: 2
7 FINAID_APPLICANT_IND Categoric Unique: 2
8 ADMIT Constant Unique: 1
9 ENROLLED Categoric Unique: 2
10 ALL_RACE Categoric Unique: 9
11 AGE Numeric Unique: 21
Roles noted. 3000 observations and 10 input variables. The target is ENROLLED. Categoric 2. Classification models enabled.


WEBFOCUS - [RStat (HOLD.csv)]
Project Tools Settings Help
Execute New Open Save Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Type: Summary Distributions Correlation Principal Components Interactive
Summary Describe Basics Kurtosis Skewness Show Missing Cross Tab Density Plot
DEGC_CODE1 13 integer 0
MAJR_CODE1 57 integer 0
HIGH_SCHOOL_PERCENTILE double 123
GENDER 2 integer 0
FINAID_APPLICANT_IND 2 integer 0
ADMIT 1 integer 0
ALL_RACE 9 integer 0
AGE integer 0
ENROLLED 2 integer 0
Variable Levels
COLL_CODE1 CB,CH,CS,ED,SA
DEGC_CODE1 BA,BAAS,BACJ,BAIS,BBA,BFA,BGS,BM,BS,BSCI,BSCJ,BSIS,BSW
MAJR_CODE1 ACCT,AG,AGBS,AGST,ANSC,ART,ARTS,ARTV,BA,BSCI,BUSA,CHEM
CIS,CJUS,CMST,CONE,CSCI,EE,ELEN,ENG,ENVS,EQST,FIN,GBUS
GENS,HED,HHPP,HHPS,HIST,HLTH,IE,INTS,JOUR,KINS,LAST,LIBS
LNTC,MATH,MGT,MKT,MUS,MUSP,ORGL,PERF,PHYS,PSCI,PSY,PUHE
RTV,SOC,SPA,SWK,THEA,TMGT,UDNU,UNDC,WCSA
GENDER F,M
FINAID_APPLICANT_IND N,Y
ENROLLED
Data summary generated.


Distribution of ALL_RACE (sample)
by ENROLLED
All
N
Y
Frequency
0
200
400
600
800
Hispanic
White
Asian
International
Hawaiian
ALL_RACE


WEBFOCUS - [RStat (HOLD.csv)]
Project Tools Settings Help
Execute New Open Save Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Please Select Model(s):
Select All Clear All
Target: ENROLLED
Number of Trees: 50
Max Depth: 30
Model Type
Exportability
Active
Ada Boost
C Exportable(.c); PMML Exportable(.xml)
Binomial Regression
Decision Tree
Gamma Regression
Gaussian Regression
Inverse Gaussian Regression
Linear Regression
Logistic Regression
Multinomial Regression
Negative Binomial Model
Neural Net
Poisson Regression
Random Forest
Survival
SVM
PMML Exportable(.xml)
WEBFOCUS - [RStat (HOLD.csv)]
Project Tools Settings Help
Execute New Open Save Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Please Select Model(s):
Select All Clear All
Model Type
Exportability
Active
Ada Boost
C Exportable(.c); PMML Exportable(.xml)
Target: ENROLLED
Model Builder: ada
Number of Trees: 50
Stumps
Defaults
Importance
Errors
List
Draw
1
Max Depth: 30
Min Split: 20
Complexity: 0.0100
X Val: 10
Continue
Summary of the Ada Boost model:
Call:
ada(ENROLLED ~ ., data = crs$dataset[crs$train, c(crs$input,
crs$target)], parms = list(split = "information"), bag.frac = 1,
control = rpart.control(maxdepth = 30, cp = 0.01, minsplit = 20,
xval = 10, usesurrogate = 0, maxsurrogate = 0), iter = 50)
Loss: exponential Method: discrete Iteration: 50
Final Confusion Matrix for Data:
Final Prediction
True value N Y
N 1199 153
Y 288 460
Train Error: 0.21
Out-Of-Bag Error: 0 iteration= 6
Additional Estimates of number of iterations:
The Ada Boost model has been built. Time taken: 6.95 secs


Computed Field Creator
Computed Field Name: Enrolled
Format: D12.2
Format...
Options...
)1.HOLD1.FINAID_APPLICANT_IND, HOLD1.HOLD1.HIGH_SCHOOL_PERCENTILE, Enrolled)
| LT GT + 7 8 9
|| LE GE - 4 5 6
IF EQ NE * 1 2 3
THEN AND OR / 0 .
ELSE NOT ** ( ) '' U
Functions...
OK
Cancel
Variables
Computed Fields
HOLD1
PIDM_KEY
TERM_CODE_KEY
DEPT_CODE1
MIDDLE_INITIAL
COLL_CODE1
DEGC_CODE1
MAJR_CODE1
STYP_DESC
ADMT_CODE
ADMIT
BIRTH_DATE
ETHN_DESC
MRE_ETHNIC_CDE
GENDER
LEVL_CODE
APDC_CODE1
APPLICATION_DATE
APST_DESC
APPLICATION_STATUS_DATE
APDC_DECISION_DATE1
ID


Function Arguments
Select a category:
RStat Scoring Routines
Retrieve User Functions
Select a function:
admit_prop
admitprop
audit_rpart
enroll_cl
enroll_class
Refresh Scoring Routines
enroll_cl
Input Parameters
Type
Value
STYP_DESC
alpha
STYP_DESC
GENDER
alpha
GENDER
FINAID_APPLICANT_IND
alpha
FINAID_APPLICANT
HIGH_SCHOOL_PERCENTILE
float
HIGH_SCHOOL_PEF
Target: ICN_ENRL (alpha)
Enrolled
Is the format of the output value enclosed in single quotation marks, or the field to which the result is returned.
enroll_cl 2017-06-29:10:07:23
APDC_CODE1
APPLICATION_DATE
APST_DESC
APPLICATION_STATUS_DATE
APDC_DECISION_DATE1
ID
HIGH_SCHOOL_ZIP
FINAID_APPLICANT_IND
SBGI_DESC_HIGH_SCHOOL
SBGI_CODE_HIGH_SCHOOL
SBGI_DESC_PRIOR_COLLEGE1
TRANSCRIPT_RECEIVED_DATE1
SBGI_DESC_PRIOR_COLLEGE2
TRANSCRIPT_RECEIVED_DATE2
SBGI_DESC_PRIOR_COLLEGE3
TRANSCRIPT_RECEIVED_DATE3
HIGH_SCHOOL_PERCENTILE
HIGH_SCHOOL_GRAD_DATE
HIGH_SCHOOL_REPORTED_GPA
RESD_CODE
CAMP_CODE
OK
Cancel

Result:

PAGE 1

ID	PROB	enrolled	COLL_CODE1	DEGC_CODE1	MAJR_CODE1	HIGH_SCHOOL_PERCENTILE	GENDER
Axxxxxxx	111,111,111.11	Axxxxxxxxxx>>	Ax	Axxxx	Axxx	111,111,111,111.>>	A
	222,222,222.22	Bxxxxxxxxxx>>	Bx	Bxxxx	Bxxx	222,222,222,222.>>	B
Bxxxxxxx	111,111,111.11	Axxxxxxxxxx>>	Ax	Axxxx	Axxx	111,111,111,111.>>	A
	222,222,222.22	Bxxxxxxxxxx>>	Bx	Bxxxx	Bxxx	222,222,222,222.>>	B

- In Excel:

ID	PROB	Enrolled	COLL_CODE1	DEGC_CODE1	MAJR_CODE1	HIGH_SCHOOL_PERCENTILE	GENDER	FINAID_APPLICANT_IND	ADMIT	ENROLLED	ALL_RACE	AGE
Student ID 1	.06	N	CS	BS	IE	26.00	M	N	Y	N	Hispanic	19
Student ID 2	.70	Y	CS	BA	MATH	9.00	M	Y	Y	Y	Black	31
Student ID 3	.33	N	ED	BA	PSY	.	F	Y	Y	Y	White	29
Student ID 4	.33	N	ED	BS	UDNU	60.00	F	Y	Y	N	White	29
Student ID 5	.70	Y	CB	BS	BA	12.00	F	Y	Y	N	Multiple Race	26
Student ID 6	.33	N	CS	BS	BSCI	74.00	M	Y	Y	Y	White	24
Student ID 7	.61	Y	SA	BS	ANSC	74.00	M	Y	Y	Y	White	23
Student ID 8	.33	N	CH	BA	LAST	.	M	Y	Y	Y	White	25
Student ID 9	.70	Y	CB	BS	BA	38.00	F	Y	Y	N	Black	22
Student ID 10	.42	N	ED	BS	UDNU	47.00	F	Y	Y	N	Black	24
Student ID 11	.33	N	CS	BS	MATH	67.00	M	Y	Y	N	Black	21
Student ID 12	.06	N	CS	BS	IE	57.00	M	N	Y	Y	White	25
Student ID 13	.33	N	CS	BS	BSCI	99.00	F	Y	Y	Y	Hispanic	18
Student ID 14	.33	N	ED	BS	PSY	.	M	Y	Y	N	Black	21
Student ID 15	.06	N	CH	BA	PHYS	63.00	M	N	Y	N	Hispanic	18

Any Questions ?