



## CBS Course Bidding Prediction Project

Applied Regression Group Project

Wingate Wellington, Chaitanya Mandalappu, Sri Sundarraaj, Karim Lahoud, Maxim Churkin, Colin Farrell and Victor Lee

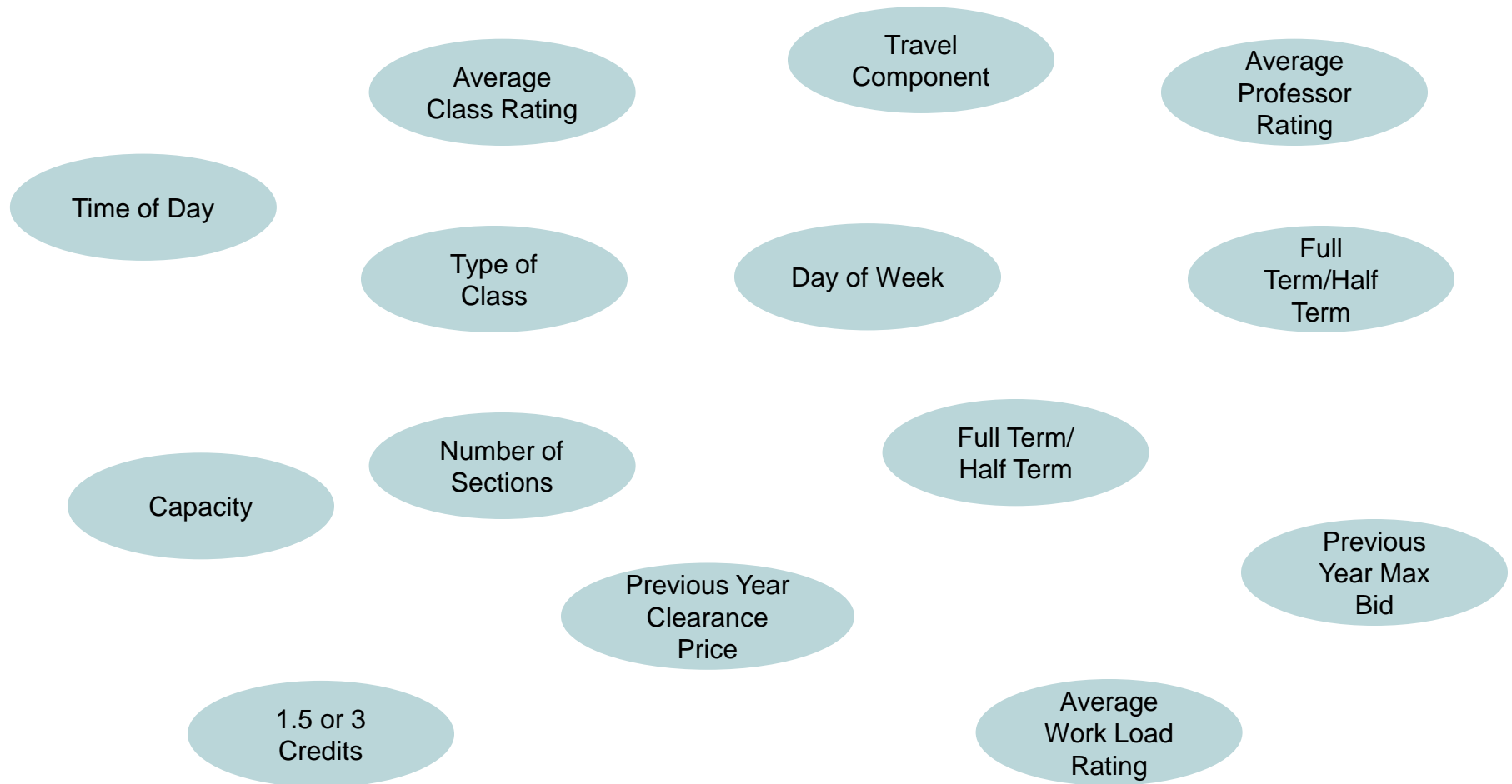
# Table of Contents

---

1. Project Goals
2. Data Findings
3. Initial Statistical Analyses
4. Additional Regression Analyses
5. Insights and Assumptions

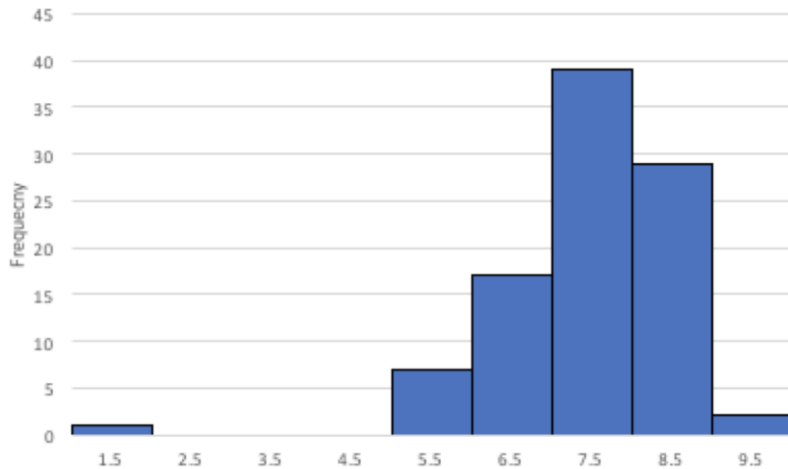
# Project Goal

- *We wanted to see if there were independent variables available that would be able to predict the clearance price of individual classes at CBS.*

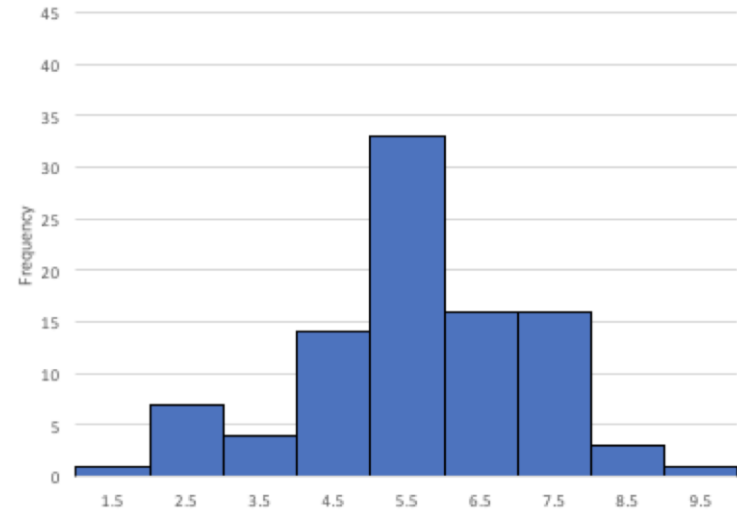


# Histograms

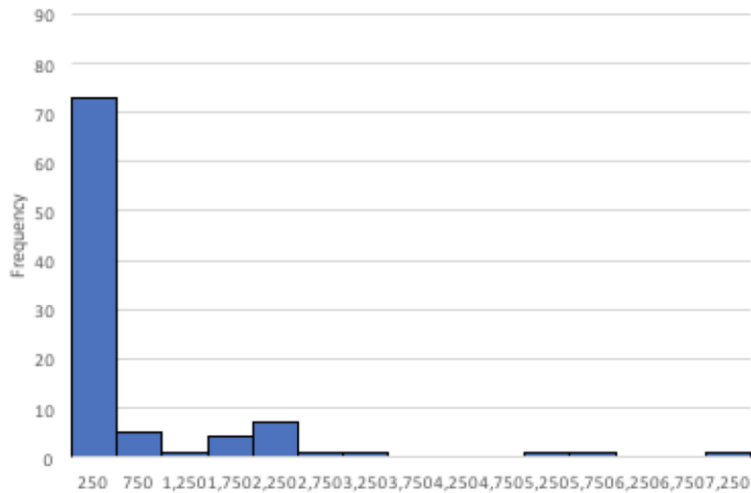
Histogram of Class Ratings



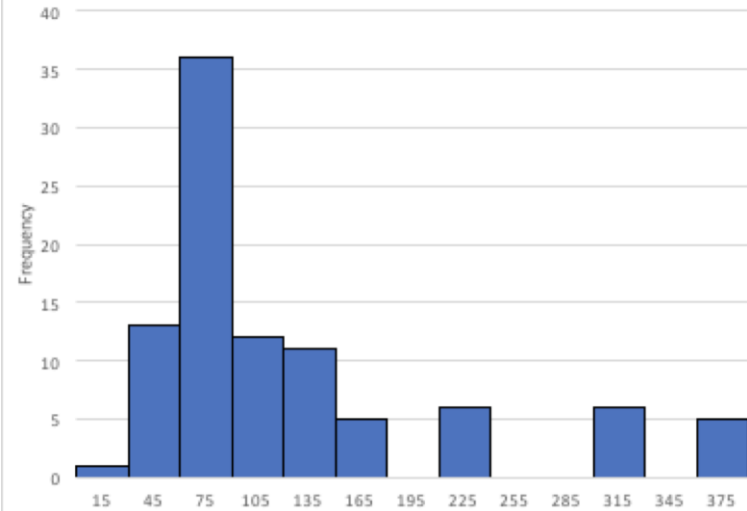
Histogram of Workload Ratings



Histogram of Clearing Prices

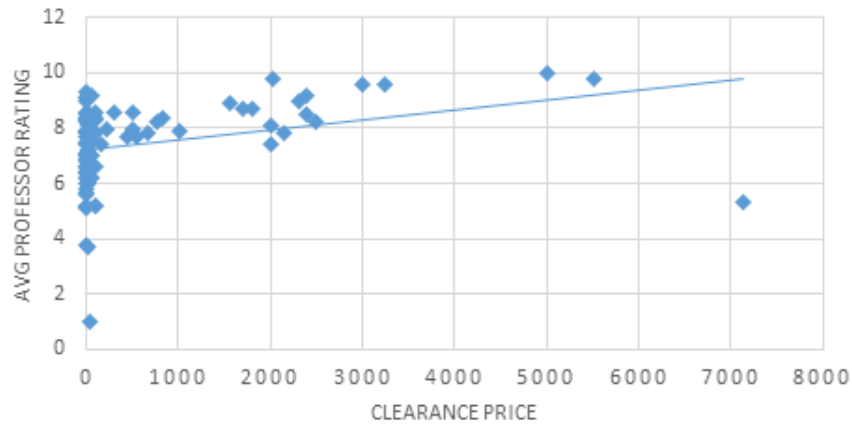


Histogram of Class Capacities

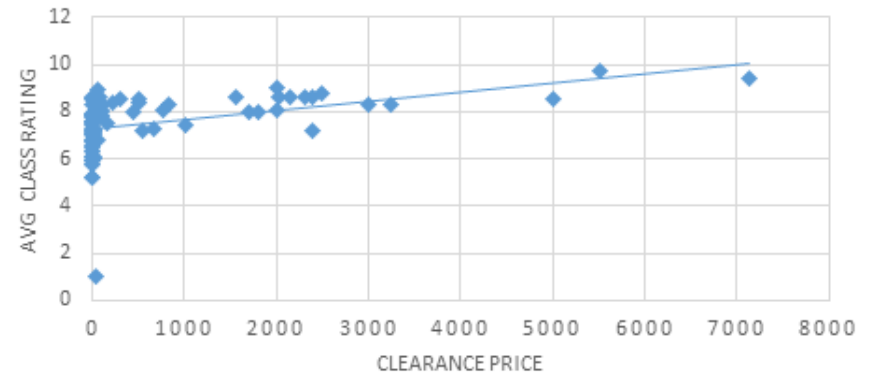


# Scatter Plots

## AVERAGE PROFESSOR RATING



## AVG CLASS RATING VS CLEARANCE PRICE



## LAST YEARS CLEARANCE PRICE VS CURRENT YEAR CLEARANCE PRICE



---

# Half Time MBA Soccer Report

# Halftime Soccer Report

STANFORD GSB		TIES: 1-2-3, 6-0, GOALS FOR:		YALE SOM	
First seed	Stanford	The MBA WINTER CLASSIC AUSTIN, TEXAS 2017		Yale SOM	Second seed
DARDEN FC	Semi-finalist 1	X		MIT SLOAN	Seventh seed
Eighth seed		Sunday, 12:00 Field 1		WHARTON FC2	Third seed
Sunday, 9:00 Field 1		Sponsored by Prodigy Finance		ROSS UM	Sixth seed
HBS A	HBS	Wharton (Winner)			
Fourth seed		Sunday, 10:30 Field 2			
CBS FC	Semi-finalist 3	Wharton			
Fifth seed		Sunday, 10:30 Field 2			
Sunday, 9:00 Field 4		Sunday, 9:00 Field 3			
Time	Field 1	Field 2	Field 3	Field 4	
8:00	Wharton FC 2 4 x 0 Chicago Booth	CBS-FC 2 1 x 2 McCombs Orange	HBS B 0 x 1 Stanford GSB	Wharton FC 1 0 x 2 McCombs White	
9:15	Fuqua United 1 x 1 UCLA Anderson	CBS FC 1 x 0 Haas FC	HBS A 1 x 0 Tuck	Duke Alumni 0 x 1 Darden FC	
10:30	LBS 0 x 1 Yale SOM	IPADE 0 x 2 MIT Sloan	Ross UM 1 x 0 Kellogg	McCombs Orange 0 x 0 Wharton FC 2	
11:45	Chicago Booth 2 x 0 CBS FC	Darden FC 0 x 2 HBS B	Stanford GSB 5 x 1 Wharton FC 1	Haas FC 2 x 0 Fuqua United	
13:00	UCLA Anderson 0 x 1 CBS FC	Yale SOM 1 x 0 Ross UM	Tuck 0 x 4 Duke Alumni	MIT Sloan 0 x 0 LBS	
14:15	Kellogg 2 x 0 IPADE	McCombs White 0 x 7 Stanford GSB	Wharton FC 2 2 x 0 CBS-FC 2	McCombs Orange 1 x 1 Chicago Booth	
15:30	HBS B 1 x 0 Wharton FC 1	Darden FC 6 x 0 Tuck	Fuqua United 0 x 0 CBS FC	Haas FC 1 x 3 UCLA Anderson	
16:45	HBS A 0 x 0 Duke Alumni		LBS 0 x 3 Kellogg	IPADE 1 x 2 Yale SOM	
18:00	Ross UM 3 x 4 MIT Sloan				
GROUP A					
Wharton FC 2	POINTS	NET GOALS	GROUP C	POINTS	NET GOALS
CBS-FC 2	●●●●●●●●	+6	Fuqua United	●●●●●●●●	-2
McCombs Orange	●●●●●●●●	-6	CBS FC	●●●●●●●●	+2
Chicago Booth	●●●●●●●●	+1	Haas FC	●●●●●●●●	-1
GROUP B	POINTS	NET GOALS	UCLA Anderson	●●●●●●●●	+1
HBS B	●●●●●●●●	-1	GROUP D	POINTS	NET GOALS
Wharton FC 1	●●●●●●●●	-7	HBS A	●●●●●●●●	+3
McCombs White	●●●●●●●●	-7	Duke Alumni	●●●●●●●●	+3
Stanford GSB	●●●●●●●●	+15	Darden FC	●●●●●●●●	+5
GROUP E	POINTS	NET GOALS	Tuck	●●●●●●●●	-11
LBS	●●●●●●●●	-4	GROUP F	POINTS	NET GOALS
IPADE	●●●●●●●●	-5	Yale SOM	●●●●●●●●	+3
Ross UM	●●●●●●●●	-2	MIT Sloan	●●●●●●●●	+4
			Kellogg	●●●●●●●●	+4

THE 20th TEXAS WINTER CLASSIC - FEBRUARY 25th and 26th, 2017 - AUSTIN, TX



# Halftime Soccer Report





---

# Initial Analysis

# Process and Procedure

*We collected our data from BOSS Fall Semester 2015 and 2016 to see if we could predict the future Clearance Price for CBS classes, while also using data from CBS Course Review*

## Process & Procedures

### Regression in Minitab

- Ran regression in Minitab with all independent variables, it had an  $R^2$  of 76.02% and an Adj- $R^2$  of 63.64%.
- The vast majority of independent variables had high P-Values, rendering the regression statistically insignificant.

### Correlation Matrix

- Ran a Correlation Matrix with all variables to remove the highly correlated ones.
- Removed “Total Capacity” as it was highly correlated with “Capacity for Section”
- Removed 1.5 Credits, highly negatively correlated with 3.0 Credits

### Best Subsets in Minitab

- Ran Best Subsets again with remaining variables and it was successful
- The best model had 5 variables and resulted in 64.9 Adj- $R^2$
- Variables included: Avg. Class Rating, “Y Bidding Price, A-Term, Management, Business

### Final Variables

- In the Regression using the 5 variables, resulting in some variables with high p-values.
- In analyzing the various best subsets and individual regression models, we believe a 4 variable regression model with LY Bidding Price, Business Dept. Class, Management Dept. Class and Avg. Class Rating is the best predictor of future clearance price. Adj.  $R^2$  = 64.7%

# Minitab Outputs

Coefficients						
Term	Coef	SE Coef	T-Value	P-Value	VIF	
Constant	-2147	1497	-1.43	0.156		
Total Capacity	9.4	10.9	0.86	0.393	4.98	
Average class rating	90	145	0.62	0.535	4.53	
Average Professor rating	55.4	82.2	0.67	0.503	2.44	
Average Workload rating	14.8	98.7	0.15	0.882	3.62	
Last years bidding price	0.3617	0.0822	4.40	0.000	3.49	
Last years capacity	0.57	4.56	0.12	0.901	26.29	
Capacity for all sections of th	-6.03	6.41	-0.94	0.350	56.65	
Number of sections	204	284	0.72	0.475	41.12	
Travel component	-667	919	-0.73	0.471	8.33	
Day - M	86	387	0.22	0.826	5.86	
Day - T	103	328	0.32	0.754	2.98	
Day - W	15	354	0.04	0.966	4.61	
Day - R	202	309	0.72	0.471	4.19	
Day - F	-2550	1032	-2.47	0.016	3.66	
Day - Sa/Su	151	662	0.23	0.820	2.95	
Day - Block	391	1256	0.31	0.757	22.56	
Full Term	-1755	1212	-1.45	0.153	60.11	
A Term	62	1515	0.04	0.968	63.64	
B Term	192	1562	0.12	0.903	67.62	
9:00 AM	2096	919	2.28	0.026	27.32	
10am to 12.30pm	2265	894	2.53	0.014	17.70	
1pm to 4:30pm	2186	895	2.44	0.017	32.04	
5pm or later	1959	996	1.97	0.054	15.57	
Accounting	-432	848	-0.51	0.612	2.47	
Business	633	514	1.23	0.223	3.01	
Decision, Risk and Operations	-196	398	-0.49	0.624	2.48	
Economics	-92	638	-0.14	0.886	2.07	
Finance	156	301	0.52	0.607	3.43	
Management	383	345	1.11	0.272	3.64	
1.5 credits	-1692	1122	-1.51	0.137	42.71	
3 credits	297	876	0.34	0.736	26.59	
Max Bid 2015	0.0282	0.0655	0.43	0.668	4.23	

Model Summary			
S	R-sq	R-sq(adj)	R-sq(pred)
754.846	76.02%	63.64%	*

# Process and Procedure

*We collected our data from BOSS Fall Semester 2015 and 2016 to see if we could predict the future Clearance Price for CBS classes, while also using data from CBS Course Review*

## Process & Procedures

### Regression in Minitab

- Ran regression in Minitab with all independent variables, it had an  $R^2$  of 76.02% and an Adj- $R^2$  of 63.64%.
- The vast majority of independent variables had high P-Values, rendering the regression statistically insignificant.

### Correlation Matrix

- **Ran a Correlation Matrix with all variables to remove the highly correlated ones.**
- **Removed “Total Capacity” as it was highly correlated with “Capacity for Section”**
- **Removed 1.5 Credits, highly negatively correlated with 3.0 Credits**

### Best Subsets in Minitab

- Ran Best Subsets again with remaining variables and it was successful
- The best model had 5 variables and resulted in 64.9 Adj- $R^2$
- Variables included: Avg. Class Rating, “Y Bidding Price, A-Term, Management, Business

### Final Variables

- In the Regression using the 5 variables, resulting is some variables with high p-values.
- In analyzing the various best subsets and individual regression models, we believe a 4 variable regression model with LY Bidding Price, Business Dept. Class, Management Dept. Class and Avg. Class Rating is the best predictor of future clearance price. Adj.  $R^2$  = 64.7%

# Correlation Matrix

	Clearance Price	Total Capacity	Average class rating	Average Professor rating	Average Workload rating	Last years bidding price	Last years capacity	Capacity for the section	Capacity for all sections	Number of sections	Travel component	Day - M	Day - T	Day - W	Day - R	Day - F	Day - Sa/Su	Day - Block	Full Term	A Term	B Term	9am	10am to 12:30 pm	1pm to 4:30pm	5pm or later	Block Week	Accounting	Business	Decision, Risk and Opera	Economics	Finance	Management	Marketing	1.5 credits	3 credits	Max Bid 2015	
Clearance Price	1.00																																				
Total Capacity	(0.30)	1.00																																			
Average class rating	0.42	(0.16)	1.00																																		
Average Professor rating	0.30	0.05	0.60	1.00																																	
Average Workload rating	0.03	(0.05)	0.42	0.16	1.00																																
Last years bidding price	0.76	(0.40)	0.37	0.22	0.13	1.00																															
Last years capacity	(0.09)	0.18	0.27	0.12	0.25	(0.14)	1.00																														
Capacity for the section	(0.30)	1.00	(0.16)	0.05	(0.05)	(0.40)	0.18	1.00																													
Capacity for all sections	(0.05)	0.17	0.26	0.12	0.20	(0.11)	0.94	0.17	1.00																												
Number of sections	0.08	(0.14)	0.34	0.14	0.18	0.03	0.83	(0.14)	0.92	1.00																											
Travel component	0.36	(0.22)	0.23	0.11	(0.27)	0.23	(0.14)	(0.22)	(0.15)	(0.12)	1.00																										
Day - M	0.02	0.11	(0.01)	(0.07)	(0.11)	(0.07)	0.22	0.11	0.15	0.04	0.06	1.00																									
Day - T	0.11	(0.01)	0.07	(0.00)	0.04	0.10	(0.05)	(0.01)	(0.02)	0.02	(0.13)	(0.40)	1.00																								
Day - W	(0.15)	0.10	0.05	0.13	0.01	(0.13)	0.34	0.10	0.29	0.23	(0.18)	0.06	(0.36)	1.00																							
Day - R	0.04	(0.17)	(0.05)	(0.03)	0.16	0.08	(0.10)	(0.17)	(0.12)	(0.07)	(0.13)	(0.40)	0.24	(0.36)	1.00																						
Day - F	(0.07)	0.01	(0.09)	0.05	(0.05)	(0.05)	(0.09)	0.01	(0.09)	(0.10)	(0.04)	(0.11)	0.10	(0.10)	(0.08)	1.00																					
Day - Sa/Su	(0.06)	0.07	(0.01)	0.05	0.05	(0.07)	(0.06)	0.07	(0.06)	(0.08)	0.38	(0.16)	(0.11)	(0.15)	(0.11)	0.33	1.00																				
Day - Block	0.32	(0.06)	0.26	0.15	(0.16)	0.19	(0.09)	(0.06)	(0.07)	(0.07)	0.80	(0.03)	(0.17)	(0.23)	(0.17)	(0.05)	0.29	1.00																			
Full Term	(0.11)	(0.06)	(0.02)	(0.14)	0.38	(0.03)	0.10	(0.06)	(0.02)	(0.10)	(0.30)	0.11	0.03	0.11	0.24	(0.17)	(0.24)	(0.37)	1.00																		
A Term	0.07	0.10	0.07	0.23	(0.20)	(0.02)	0.02	0.10	0.12	0.17	0.29	(0.14)	0.05	(0.08)	(0.14)	0.10	0.15	0.45	(0.59)	1.00																	
B Term	0.08	(0.02)	(0.06)	(0.04)	(0.31)	0.07	(0.15)	(0.02)	(0.10)	(0.06)	0.08	0.02	(0.08)	(0.03)	(0.14)	0.10	0.15	0.01	(0.59)	(0.27)	1.00																
9am	(0.10)	0.03	(0.12)	0.00	(0.01)	(0.08)	(0.08)	0.03	(0.04)	(0.02)	(0.16)	0.03	0.04	(0.06)	0.16	0.08	(0.01)	(0.19)	0.04	(0.07)	0.04	1.00															
10am to 12:30pm	0.06	0.05	0.14	0.07	0.18	0.02	0.30	0.05	0.24	0.18	0.01	0.14	(0.08)	0.25	(0.08)	(0.06)	0.05	(0.04)	0.09	(0.15)	0.06	(0.26)	1.00														
1pm to 4:30pm	0.02	(0.18)	(0.10)	(0.22)	(0.07)	0.00	(0.04)	(0.18)	(0.04)	(0.00)	(0.04)	(0.06)	0.05	0.03	0.11	0.03	(0.17)	(0.12)	0.10	(0.05)	(0.05)	(0.49)	(0.35)	1.00													
5pm or later	0.06	0.08	0.01	0.18	(0.15)	0.13	(0.19)	0.08	(0.20)	(0.20)	(0.09)	0.09	0.08	(0.09)	(0.18)	(0.05)	(0.07)	(0.11)	(0.05)	(0.01)	0.08	(0.20)	(0.15)	(0.28)	1.00												
Block Week	(0.00)	0.13	0.18	0.14	0.02	(0.03)	(0.00)	0.13	0.03	0.00	0.47	(0.20)	(0.13)	(0.18)	(0.13)	(0.04)	0.38	0.80	(0.30)	0.50	(0.13)	(0.16)	(0.11)	(0.21)	(0.09)	1.00											
Accounting	(0.07)	(0.02)	0.13	0.05	0.01	(0.05)	0.18	(0.02)	(0.01)	(0.01)	(0.04)	0.19	(0.08)	0.21	(0.08)	(0.02)	(0.03)	(0.05)	0.13	(0.08)	(0.08)	0.08	0.14	(0.12)	(0.05)	(0.04)	1.00										
Business	0.35	(0.16)	(0.09)	(0.10)	(0.33)	0.22	(0.21)	(0.16)	(0.21)	(0.18)	0.59	0.11	(0.05)	(0.20)	0.05	(0.04)	(0.06)	0.46	(0.16)	0.05	0.15	(0.08)	(0.01)	0.02	0.03	0.09	(0.04)	1.00									
Decision, Risk and O	(0.15)	0.14	(0.13)	(0.10)	0.03	(0.07)	(0.11)	0.14	(0.10)	(0.14)	(0.09)	(0.20)	(0.01)	0.13	(0.09)	(0.05)	0.10	(0.11)	(0.26)	(0.09)	0.41	0.03	0.04	0.07	(0.12)	(0.09)	(0.05)	(0.10)	1.00								
Economics	(0.05)	0.12	0.04	(0.00)	(0.12)	(0.06)	(0.02)	0.12	(0.01)	(0.05)	(0.05)	0.23	(0.09)	0.26	(0.09)	(0.03)	(0.04)	(0.06)	0.16	(0.09)	(0.09)	(0.11)	0.09	0.10	(0.06)	(0.05)	(0.03)	(0.05)	(0.06)	1.00							
Finance	0.03	0.02	0.20	0.08	0.53	0.12	0.09	0.02	0.01	(0.07)	(0.19)	0.07	0.00	0.01	0.17	(0.11)	(0.15)	(0.16)	0.37	(0.21)	(0.27)	(0.08)	0.11	(0.05)	0.11	(0.10)	(0.11)	(0.21)	(0.25)	(0.13)	1.00						
Management	0.09	(0.18)	0.20	0.15	(0.03)	(0.05)	0.27	(0.18)	0.38	0.53	0.06	(0.19)	0.07	(0.08)	0.01	0.26	0.25	0.07	(0.20)	0.25	0.01	0.05	(0.04)	0.04	(0.19)	0.16	(0.08)	(0.16)	(0.19)	(0.10)	(0.41)	1.00					
Marketing	(0.20)	0.13	(0.38)	(0.14)	(0.37)	(0.15)	(0.25)	0.13	(0.21)	(0.24)	(0.12)	0.03	0.03	(0.09)	(0.11)	(0.07)	(0.10)	(0.06)	(0.04)	0.10	(0.04)	0.10	(0.20)	(0.04)	0.20	(0.01)	(0.07)	(0.13)	(0.16)	(0.08)	(0.34)	(0.26)	1.00				
1.5 credits	(0.15)	0.08	(0.16)	(0.01)	(0.34)	(0.12)	(0.11)	0.08	(0.01)	0.07	(0.16)	(0.01)	(0.04)	0.11	(0.21)	(0.09)	(0.02)	(0.20)	(0.72)	0.25	0.65	0.05	(0.02)	(0.04)	0.16	(0.16)	(0.09)	(0.09)	0.39	(0.11)	(0.26)	0.03	0.13	1.00			
3 credits	0.16	(0.09)	0.17	0.06	0.32	0.12	0.13	(0.09)	0.02	(0.05)	0.17	0.03	(0.01)	(0.09)	0.22	0.09	0.02	0.21	0.70	(0.23)	(0.63)	(0.03)	0.03	0.01	(0.15)	0.17	0.09	0.09	(0.38)	0.12	0.23	(0.01)	(0.12)	(0.97)	1.00		
Max Bid 2015	0.70	(0.27)	0.46	0.31	0.14	0.78	(0.14)	(0.27)	(0.09)	0.04	0.30	(0.12)	0.13	(0.27)	0.09	(0.07)	0.02	0.32	(0.10)	0.09	0.02	(0.12)	0.01	(0.08)	0.18	0.12	(0.06)	0.18	(0.13)	(0.10)	0.06	0.06	(0.10)	(0.16)	0.17	1.00	

# Process and Procedure

*We collected our data from BOSS Fall Semester 2015 and 2016 to see if we could predict the future Clearance Price for CBS classes, while also using data from CBS Course Review*

## Process & Procedures

### Regression in Minitab

- Ran regression in Minitab with all independent variables, it had an  $R^2$  of 76.02% and an Adj- $R^2$  of 63.64%.
- The vast majority of independent variables had high P-Values, rendering the regression statistically insignificant.

### Correlation Matrix

- Ran a Correlation Matrix with all variables to remove the highly correlated ones.
- Removed “Total Capacity” as it was highly correlated with “Capacity for Section”
- Removed 1.5 Credits, highly negatively correlated with 3.0 Credits

### Best Subsets in Minitab

- **Ran Best Subsets again with remaining variables and it was successful**
- **The best model had 5 variables and resulted in 64.9 Adj- $R^2$**
- **Variables included: Avg. Class Rating, “Y Bidding Price, A-Term, Management, Business**

### Final Variables

- In the Regression using the 5 variables, resulting in some variables with high p-values.
- In analyzing the various best subsets and individual regression models, we believe a 4 variable regression model with LY Bidding Price, Business Dept. Class, Management Dept. Class and Avg. Class Rating is the best predictor of future clearance price. Adj.  $R^2 = 64.7\%$

## Best Subsets in MiniTab

### Best Subsets Regression: Clearance Pr versus Last years b, Average clas, ...

Response is Clearance Price

[illegible]



# Process and Procedure

*We collected our data from BOSS Fall Semester 2015 and 2016 to see if we could predict the future Clearance Price for CBS classes, while also using data from CBS Course Review*

## Process & Procedures

### Regression in Minitab

- Ran regression in Minitab with all independent variables, it had an  $R^2$  of 76.02% and an Adj- $R^2$  of 63.64%.
- The vast majority of independent variables had high P-Values, rendering the regression statistically insignificant.

### Correlation Matrix

- Ran a Correlation Matrix with all variables to remove the highly correlated ones.
- Removed “Total Capacity” as it was highly correlated with “Capacity for Section”
- Removed 1.5 Credits, highly negatively correlated with 3.0 Credits

### Best Subsets in Minitab

- Ran Best Subsets again with remaining variables and it was successful
- The best model had 5 variables and resulted in 64.9 Adj- $R^2$
- Variables included: Avg. Class Rating, “Y Bidding Price, A-Term, Management, Business

### Final Variables

- **In the Regression using the 5 variables, resulting is some variables with high p-values.**
- **In analyzing the various best subsets and individual regression models, we believe a 4 variable regression model with LY Bidding Price, Business Dept. Class, Management Dept. Class and Avg. Class Rating is the best predictor of future clearance price. Adj.  $R^2$  = 64.7%**

# Regression Outputs

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0.81					
R Square	0.66					
Adjusted R Square	0.64687					
Standard Error	743.93					
Observations	95.00					
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	4.00	97,508,092.25	24,377,023.06	44.05	0.00	
Residual	90.00	49,808,671.37	553,429.68			
Total	94.00	147,316,763.62				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	(1,360.13)	555.84	(2.45)	0.02	(2,464.42)	(255.85)
Average class rating	196.78	75.22	2.62	0.01	47.35	346.22
Last years bidding price	0.46	0.05	9.38	0.00	0.36	0.56
Business	1,132.23	307.23	3.69	0.00	521.87	1,742.59
Management	345.29	184.97	1.87	0.07	(22.19)	712.77

---

# Additional Analysis

# Logistic Regression

## Deviance Table

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	9	60.671	6.7413	60.67	0.000
Average class rating	1	3.505	3.5051	3.51	0.061
Average Professor rating	1	1.635	1.6346	1.63	0.201
Average Workload rating	1	0.555	0.5549	0.55	0.456
Last years bidding price	1	20.257	20.2567	20.26	0.000
Capacity for all sections of th	1	1.644	1.6436	1.64	0.200
Day - F	1	0.230	0.2300	0.23	0.632
9:00 AM	1	0.811	0.8106	0.81	0.368
Business	1	1.032	1.0323	1.03	0.310
Management	1	5.337	5.3366	5.34	0.021
Error	85	56.228	0.6615		
Total	94	116.899			

## Model Summary

Deviance	Deviance	
R-Sq	R-Sq(adj)	AIC
51.90%	44.20%	76.23



# Insights and Assumptions

## Assumptions

Only used the data for Round 1 results

Only focused on Fall bidding, due to the Spring 1A bidding round

Did not include EMBA classes or Half Block Week classes

Excluded new classes due to incomplete dataset

## Insights

LY Bidding Price alone is able to explain 58% of the clearing price this year

The conventional wisdom that Avg. Professor Rating matters is not supported by our model

While our model has an Adj-R<sup>2</sup> of 64.7%, it will not instill confidence amongst CBS students as R<sup>2</sup> isn't closer to 100%

Average Class Rating's coefficient is 197. This can be intuitively explained that with each 1 point increase in class rating, the bid price increases by 197.