"An exploratory analysis of quick service restaurants using tidyverse tools in R"

AUTHORS	Khalid M. Dubas (iD) https://orcid.org/0000-0001-5271-4484

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
17	8	17

Khalid M. Dubas (USA)

An exploratory analysis of quick service restaurants using tidyverse tools in R

Abstract

This study presents an exploratory analysis of Quick Service Restaurants (QSR) industry in the US during 2015 by using publicly available data and open-source R software. The data analysis framework and tools utilized here were developed primarily by Hadley Wickham and are included in his tidyverse package in R. This data analysis framework consists of data import, data wrangeling, data exploration (tranformation, visualization, and modeling), and communication of results (Wickham & Grolemund, 2016). These steps are illustrated by exploring the relationship between sales, customer satisfaction, and other characteristics of 65 QSR restaurants in the US. In order to facilitate reproduction and replication of this study, the dataset as well as the R code are included in this study.

Introduction

It is important for a scientific study to be reproducible by the reader of the study. This ensures reliability of the study's findings and helps a discipline grow through replication of the study for different datasets and different settings thus accumulating independently verified scientific knowledge. For this purpose, the data as well as the code for a study should be publicly available. This favors the use of open source software for analysis since a proprietary software would limit access to the reader due to licensing requirements.

While the base R software (The R Foundation, n.d.) is updated a few times a year, there were over 10000 user-contributed packges at the Comprehensive R Archieve Network (CRAN) at the time of writing of this document. These user-contributed packages are frequently updated and new packages are added to CRAN at an exponential rate. Most new developments in R occur through these user-coontributed packages. One recent major addition to R is the tidyverse set of packages (Wickham, 2016; Wickham & Grolemund, 2017). The tidyverse is a set of packages that share a common design philosophy of Hadley Wickham and provide a consistent framework for data science with R. There are four basic principles to a tidy API (Wickham, 2017).

1. Reuse existing data structures.
2. Compose simple functions with the pipe.
3. Embrace functional programming, and
4. Design for humans.

The tidyverse significantly extends the capabilities of R for data science and library (tidyverse) loads the following core packages ("Tidyverse packages", n.d.):

♦ readr for data importing;
♦ tibble for a modern way of data representation;
♦ tidyr for data tidying;
♦ dplyr for data manipulation;
♦ ggplot2 for data visualization; and
♦ purrr for functional programming.

Tidyverse's capabilities are extended by other tools and specialized packages. Here are the main tools and packages as they relate to various stages of data science ("tidyverse", n.d.):

♦ Workflow – scripts, projects, etc.
♦ Import – readr, readxl, heaven, etc.
♦ Wrangle – tibble, tidyr, dplyr, stringr, forcats, lubridate, hms, blob, etc.
♦ Visualize – ggplot2.
♦ Program – purrr, rlang, magrittr, glue, etc.
♦ Model – modelr, purrr, broom, etc.
♦ Communicate – RMarkdown, ggplot2, etc.

Fig. 1 shows some extended tidyverse packages.

While the best place to learn about tidyverse is Wickham & Grolemund (2017), other resources are also available to enhance one's knowledge about tidyverse (Held, 2017; Levy, 2016; Wickham, 2016; R Views, 2017; and "The tidyverse: dplyr, ggplot2, and friends", n.d)

Purpose and scope. The purpose of this study is to illustrate the use of tidyverse tools in R for data science. The complete data analysis framework is presented here by using publicly available data on 65 Quick Service Restaurants

---

Khalid M. Dubas, Dr., Professor of Marketing, University of Mount Olive, USA.

(QSR) in the US in 2015. This study, while illustrative and exploratory in nature, also evaluates the characteristics of different segments of QSR restaurants in the US and explores the factors that influence restaurant sales.

Fig. 1. The tidyverse packages

Reserch questions. This study addresses these research questions:

1. What are the main segments of the U.S. QSR industry?
2. Who are the leaders in each QSR segment?
3. What factors influence the Sales of a QSR restaurant?
4. What factors influence the Sales of a QSR restaurant within each segment?

1. Data import

The tidyverse package was installed before importing data. The data file, qsr65, was imported using R and RStudio. This data file contains information about 65 Quick Service Restaurants (QSR) in the U.S. in 2015. All data, except American Customer Satisfaction Index (ACSI) for QSR were obtained from QSR, Inc. (2016). The data on 2015 ACSI for QSR were obtained from ACSI LLC (2017).

The data file includes the following variables:

♦ company – An American QSR;
♦ rank – The rank of Annual Sales in 2015;
♦ sales – 2015 U.S. System-wide Sales ($ millions);
♦ acsi – American Customer Satisfaction Index for a QSR in 2015. (0 – 100);
♦ segments – QSR segments in 2015. There were burger, chicken, sandwich, pizza, seafood, ethnic, and snack;
♦ seg_rank – The rank of a QSR in its segment;
♦ unit_avg – 2015 U.S. Average Sales per Unit (thousands);
♦ fr_units – Number of Franchised Units in 2015;
♦ co_units – Number of Company Units in 2015;
♦ change – Total Change in Units from 2014.

The data table was produced by kable function in the knitr package.

1.1. Research question #1. What are the main segments of the US QSR industry?

Answer #1. The first research question is answered by the data. There were seven QSR segments: burger, chicken, sandwich, pizza, seafood, ethnic, and snack.

1.2. Research question #2. Who are the leaders in each QSR segment?

Answer #2. The second research question is answered by the data. The leaders in each QSR segment were listed by seg_rank: McDonald's (burger), Chick-fil-A (chicken), Subway (sandwich), Pizza Hut (pizza), Captain D's (seafood), Taco Bell (ethnic), and Starbucks (snack).

Research questions #3 and #4 are explored throughout the rest of this study.

2. Data wrangling

After importing the data file, qsr65, the measurement scales of variables were examined. It was noticed that two variables, sales and unit_avg, were measured on a character scale since they represented sales in US dollars and included commas to facilitate reading of large numbers. Since these variables should have been measured on a numeric scale, they were converted to numeric scales. Their median values were calculated to check that they were properly transformend from a character scale to a numeric scale. The median sales were $776 million and the median unit_avg sales were $1,123 thousands.

Table 1. Quick service restaurants (USA, 2015)

Company	Rank	Sales	Asci	Segments	Seg_rank	Unit_avg	Fr_u	Co-u	Change
Arby's	16	$3,540.00	74	Sandwich	3	$1,073.00	2210	1004	-12
Auntie Anne's	49	$529.20	NA	Snack	7	$500.10	1238	14	59
Baskin-Robbins	46	$584.00	NA	Snack	6	$233.30	2493	10	19
Bojangles'	28	$1,153.00	NA	Chicken	5	$1,838.70	378	281	40
Boston Market	44	$617.90	NA	Chicken	9	$1,364.00	5	453	2
Burger King	4	$9,530.00	72	Burger	2	$1,337.40	7074	52	-16
Captain D's	48	$531.70	NA	Seafood	1	$1,024.00	238	272	-1
Carl's Jr./Hardee's	15	$3,661.30	NA	Burger	5	$1,248.70	2673	285	45
Charleys Philly Steaks	65	$334.10	NA	Sandwich	NA	$692.00	503	48	17
Checkers/Rally's	33	$776.10	NA	Burger	11	$966.50	505	324	25
Chick-fil-A	8	$6,825.90	86	Chicken	1	$3,977.30	1646	337	96
Chipotle	12	$4,501.20	83	Ethnic	2	$2,424.00	0	1971	216
Church's Chicken	31	$838.00	NA	Chicken	6	$729.00	881	250	-37
CiCi's Pizza	54	$440.10	NA	Pizza	NA	$1,002.00	405	41	-6
Cold Stone Creamery	53	$469.80	NA	Snack	NA	$507.40	916	10	19
Corner Bakery Cafe	59	$369.00	NA	Snack	NA	$2,215.80	83	114	13
Culver's	27	$1,196.00	NA	Burger	9	$2,183.80	552	7	31
Dairy Queen	17	$3,532.00	NA	Snack	3	$783.00	4509	2	-1
Del Taco	37	$700.20	NA	Ethnic	5	$1,376.60	247	297	-3
Dickey's Barbecue Pit	62	$358.20	NA	Sandwich	NA	$762.10	499	9	91
Domino's Pizza	11	$4,800.00	75	Pizza	2	$900.00	4816	384	133
Dunkin' Donuts	7	$7,620.00	78	Snack	2	$903.80	8392	39	349
Einstein Bros. Bagels	57	$402.00	NA	Snack	NA	$670.00	285	315	-158
El Pollo Loco	35	$753.40	NA	Chicken	8	$1,798.00	247	186	18
Firehouse Subs	41	$648.50	NA	Sandwich	6	$726.00	913	32	95
Five Guys	26	$1,319.50	NA	Burger	8	$1,123.00	813	402	52
Hungry Howie's Pizza	64	$342.50	NA	Pizza	NA	$638.00	531	18	-4
In-N-Out Burger	45	$613.20	NA	Burger	13	$1,959.20	0	313	12
Jack in the Box	19	$3,395.50	72	Burger	6	$1,510.00	1836	413	-1
Jamba Juice	50	$524.50	NA	Snack	8	$641.20	748	70	12
Jason's Deli	42	$643.60	NA	Sandwich	7	$2,664.00	106	154	7
Jersey Mike's Subs	40	$675.00	NA	Sandwich	5	$645.30	1006	40	189
Jimmy John's Sandwiches	23	$1,978.60	NA	Sandwich	4	$877.10	2354	51	296
KFC	14	$4,328.30	73	Chicken	2	$1,000.00	4064	206	-121
Krispy Kreme	36	$708.10	NA	Snack	4	$1,835.60	181	116	19
Krystal	55	$420.90	NA	Burger	NA	$1,169.90	125	227	0
Little Caesars	18	$3,453.20	74	Pizza	3	$815.00	3626	611	212
Long John Silver's	52	$512.00	NA	Seafood	NA	$628.00	696	41	-78
Marco's Pizza	56	$417.60	NA	Pizza	NA	$683.50	667	0	99
McAlister's Deli	47	$547.70	NA	Sandwich	8	$1,634.50	318	43	24
McDonald's	1	$35,800.00	67	Burger	1	$2,500.00	12899	1360	-91
Moe's Southwest Grill	43	$638.10	NA	Ethnic	6	$1,163.00	633	5	57
Noodles & Company	51	$517.90	NA	Ethnic	NA	$1,105.00	70	422	53
Panda Express	22	$2,550.10	NA	Ethnic	3	$1,487.00	85	1705	61
Panera Bread	10	$4,900.00	80	Sandwich	2	$2,500.00	1071	901	92

Table 1.(cont.) Quick service restaurants (USA, 2015)

Company	Rank	Sales	Asci	Segments	Seg_rank	Unit_avg	Fr_u	Co-u	Change
Papa John's	21	$2,882.90	78	Pizza	4	$892.80	2681	707	48
Papa Murphy's	30	$880.40	NA	Pizza	5	$623.50	1369	127	35
Pizza Hut	9	$5,703.80	78	Pizza	1	$725.00	7311	511	-109
Pollo Tropical	60	$366.70	NA	Chicken	NA	$2,585.00	35	155	29
Popeyes Louisiana Kitchen	20	$3,100.00	NA	Chicken	3	$1,458.50	2469	70	160
Potbelly Sandwich Shop	58	$387.40	NA	Sandwich	NA	$1,060.20	24	372	44
Qdoba Mexican Eats	34	$753.70	NA	Ethnic	4	$1,169.00	339	322	23
Quiznos	63	$354.10	NA	Sandwich	NA	$370.00	953	4	-243
Sonic	13	$4,390.00	NA	Burger	4	$1,200.00	3139	387	41
Starbucks	2	$13,300.00	4	Snack	1	$1,062.20	4962	7559	303
Steak n Shake	29	$1,032.40	NA	Burger	10	$1,866.90	144	417	16
Subway	3	$11,500.00	77	Sandwich	1	$424.30	27103	0	145
Taco Bell	5	$8,820.20	72	Ethnic	1	$1,460.00	5227	894	170
Taco John's	61	$358.80	NA	Ethnic	NA	$931.30	378	10	-5
Tim Hortons	38	$695.70	NA	Snack	5	$1,143.00	881	3	0
Wendy's	6	$8,811.90	73	Burger	3	$1,540.00	5090	632	-58
Whataburger	24	$1,747.00	NA	Burger	7	$2,530.00	119	672	17
White Castle	39	$691.30	NA	Burger	12	$1,408.70	0	390	10
Wingstop	32	$821.20	NA	Chicken	7	$1,130.00	826	19	133
Zaxby's	25	$1,576.40	NA	Chicken	4	$2,174.30	602	123	65

3. Exploratory data analysis

3.1. Summarization. The mosaic package provided summary information about sales, unit_avg, and acsi variables.

Table 2. Quick service restaurants sales by segment

Segments	Min	Q1	Median	Q3	Max	Mean	Sd	N	Missing
Burger	420.9	840.175	1533.25	4207.825	35800.0	5241.793	9275.984	14	0
Chicken	366.7	770.350	995.50	2719.100	6825.9	2038.080	2099.585	10	0
Ethnic	358.8	608.050	726.95	3037.875	8820.2	2355.025	2977.369	8	0
Pizza	342.5	434.475	1881.65	3789.900	5703.8	2365.062	2148.799	8	0
Sandwich	334.1	380.100	646.05	2368.950	11500.0	2155.600	3289.831	12	0
Seafood	512.0	516.925	521.85	526.775	531.7	521.850	13.930	2	0
Snack	369.0	497.150	584.00	2120.050	13300.0	2612.209	4180.544	11	0

Table 3. Quick service restaurants average unit sales by segment

Segments	Min	Q1	Median	Q3	Max	Mean	Sd	N	Missing
Burger	966.5	1212.175	1459.35	1936.125	2530.0	1610.2929	513.5849	14	0
Chicken	729.0	1188.500	1628.25	2090.400	3977.3	1805.4800	945.2114	10	0
Ethnic	931.3	1148.500	1272.80	1466.750	2424.0	1389.4875	459.1363	8	0
Pizza	623.5	672.125	770.00	894.600	1002.0	784.9750	138.5787	8	0
Sandwich	370.0	680.325	819.60	1213.375	2664.0	1119.0417	760.1538	12	0
Seafood	628.0	727.000	826.00	925.000	1024.0	826.0000	280.0143	2	0
Snack	233.3	574.300	783.00	1102.600	2215.8	954.1273	595.9809	11	0

Table 4. Quick service restaurants ACSI index by segment

Segments	Min	Q1	Median	Q3	Max	Mean	Sd	N	Missing
Burger	67	70.75	72.0	72.25	73	71.00	2.708013	4	10
Chicken	73	76.25	79.5	82.75	86	79.50	9.192388	2	8
Ethnic	72	74.75	77.5	80.25	83	77.50	7.778175	2	6
Pizza	74	74.75	76.5	78.00	78	76.25	2.061553	4	4
Sandwich	74	75.50	77.0	78.50	80	77.00	3.000000	3	9
Seafood	NA	NA	NA	NA	NA	NaN	NA	0	2
Snack	74	75.00	76.0	77.00	78	76.00	2.828427	2	9

3.2. Visualization. Dot plots for sales, unit_avg, and acsi are displayed below. For each of these plots, the mean value is displayed in other color. The cowplot package was utilizied to display multiple ggplot plots together. Multiple plots can also be displayed together by using a function written by Chang (n.d.).
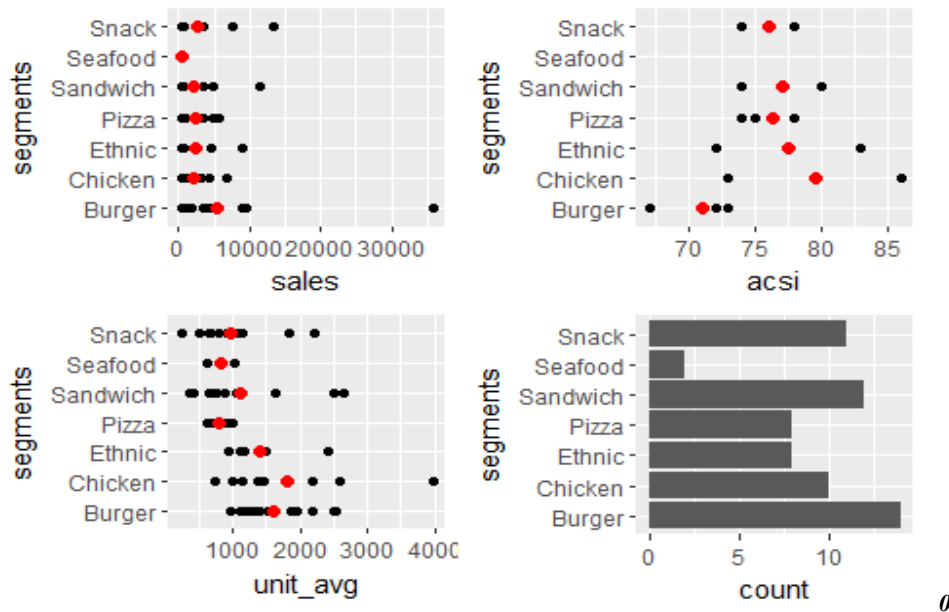

Fig. 2. QSR characteristics by segment

While the leaders in each segment are dispayed in the above figures, two restaurants stand out. McDonald's has the highest level of sales but a low level of ACSI. Chick-fil-A, on the other hand, had a high level of sales per unit and also higher level of ACSI than McDonald's. The number of restaurants by segment, from most to least were burger, sandwich, snack, chicken, ethnic, pizza, and seafood restaurants.

3.3. Box plots and violin plots

The box plots and violin plots for sales and unit_avg sales are displayed below. The violin plot is a compact version of the density plot that is displayed as a boxplot (Wickhham, 2016).
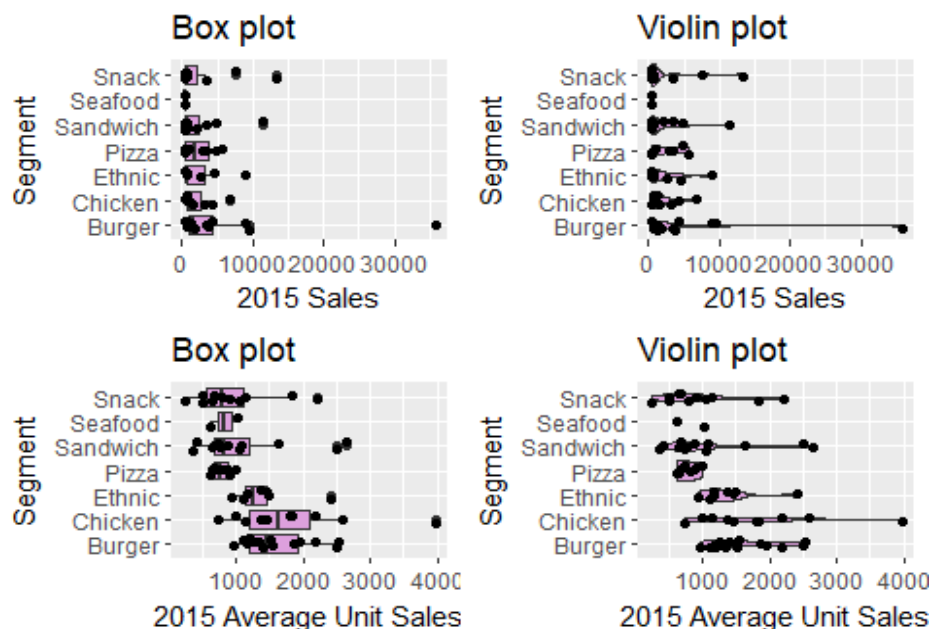

Fig. 3. QSR boxplots and violin plots by segment

3.4. Company sales and average sales per unit versus ACSI

The relationships between company sales (sale) and ACSI (acsi) and also between Average Sales per Unit (unit_avg) and ACSI (acsi) are displayed next. These relationships explore overall patterns at the industry level and address research question #3.
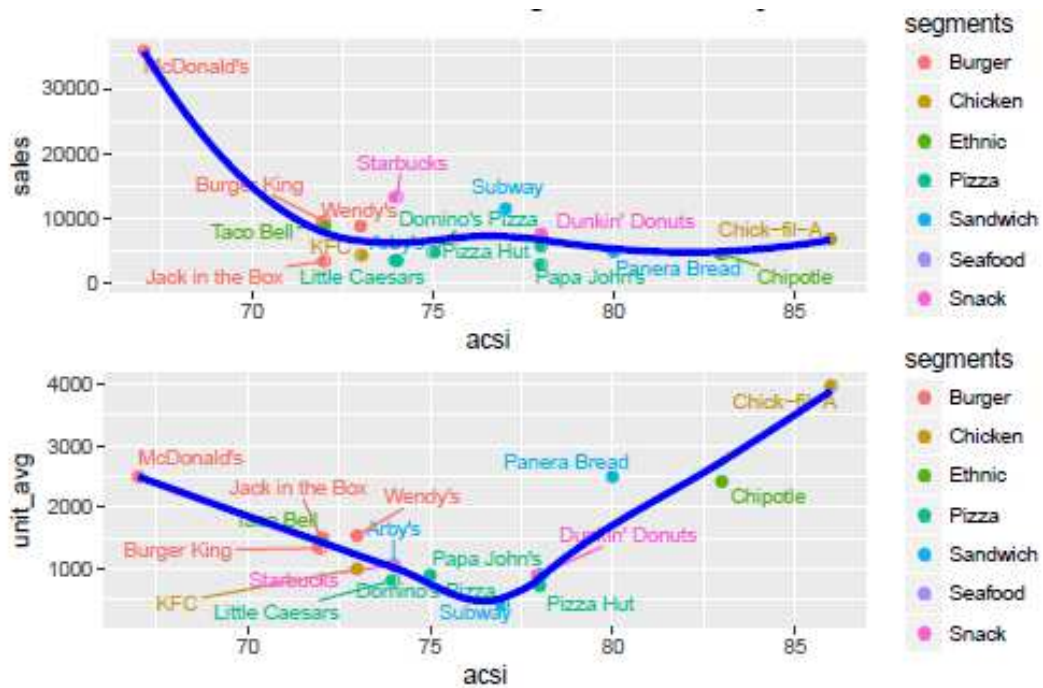
Fig. 4. QSR sales by ACSI and unit average sales by ACSI

3.5. Best-in-segment sales and within-segment relationships

The data for ACSI were not available for all companies so the mean value was substituted for missing acsi values.

The companies with the best-in-segment sales were identified in a plot of the relationship between Sales and ACSI. Next, linear relationships among companies were displayed within each segment. These relationships address research question #4 at the segment level.
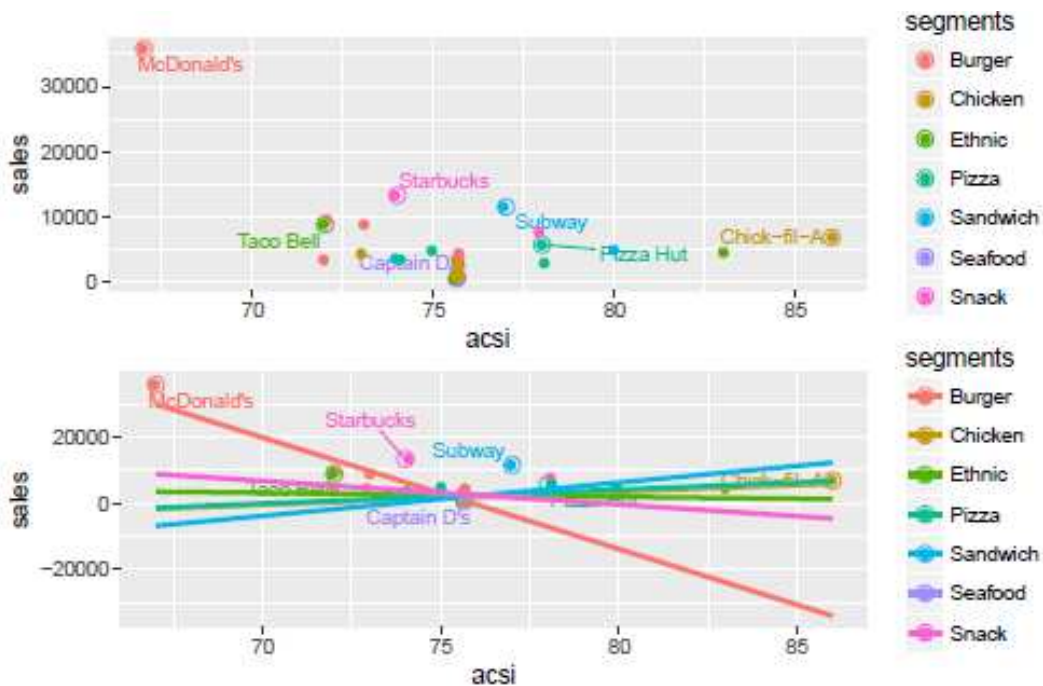

Fig. 5. QSR best-in-segment sales and by segment relationships

3.6. Best-in-segment average sales per unit and within-segment relationships

Next, the best companies in each segment were identified for their Average Sales per Unit (unit_avg). The data for ACSI were not available for all companies so the mean value was

substituted for missing acsi values.

The Average Sales per Unit were plotted against ACSI. For each set of plots, the first plot displays only the companies while the second plot displays linear relationships among companies within segments.
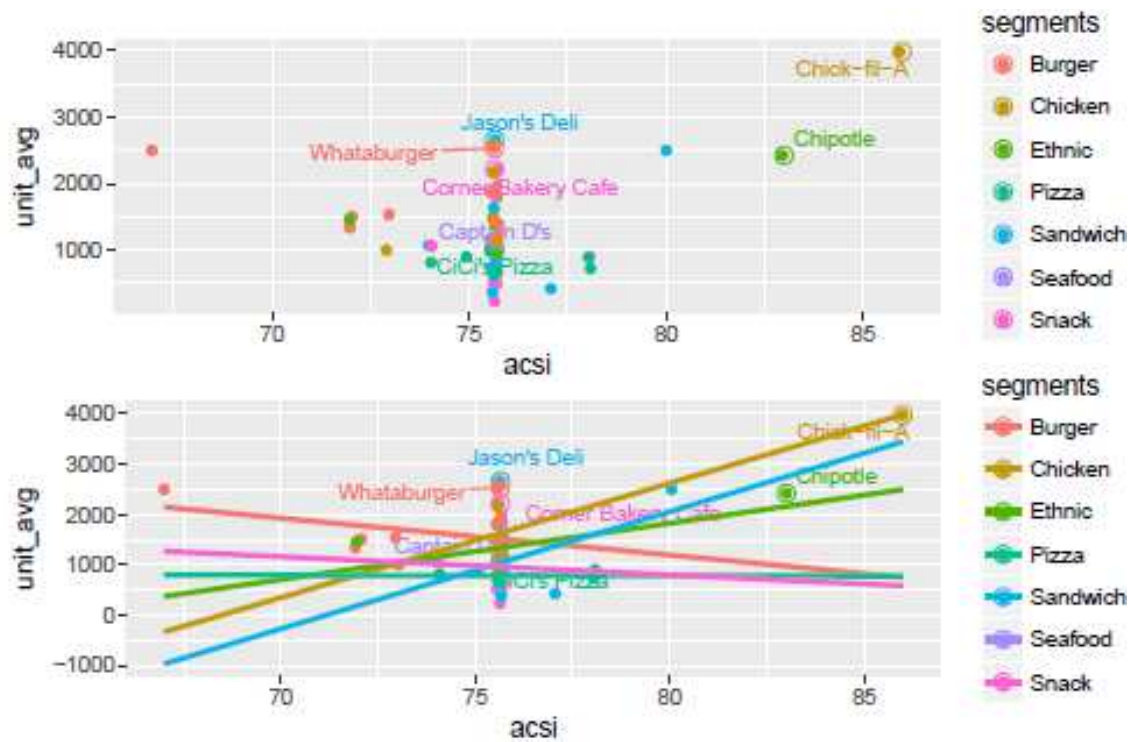
Fig. 6. QSR best-in-segment unit average sales and by segment relationships

4. Exploratory data analysis

4.1. Transformation. Table 5 provides summary statistics, Table 6 presents mean values by segment, and Table 7 shows the outlier restaurant which is McDonalds.

The correlations and histograms show a large,

positive, and significant relationship between the number of units of a restaurant (units) and its Sales (sales). However, there is no statistically significant relationship betweeen ACSI and Sales of a restaurant. The histograms show skewed distributions for most variables that are largely corrected by log transformation of these variables.

Table 5. Descriptive Statistics

|	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Median	776.10	75.65	1123.00	884.00	6.65	4.33	7.02	6.78
Mean	2879.57	75.65	1285.70	2529.20	7.18	4.33	7.02	7.11
SE.mean	630.58	0.28	86.74	518.42	0.14	0.00	0.07	0.14
Cl.mean.0.95	1259.73	0.57	173.29	1035.66	0.29	0.01	0.13	0.28
Var	25846197.69	5.25	489069.37	17469238.01	1.35	0.00	0.29	1.28
Std. Dev	5083.92	2.29	699.33	4179.62	1.16	0.03	0.54	1.13
Coef.var	1.77	0.03	0.54	1.65	0.16	0.01	0.08	0.16
Skewness	4.44	1.01	1.25	3.74	0.73	0.56	-0.18	0.64
Skew.2SE	7.47	1.70	2.10	6.29	1.23	0.94	-0.30	1.07
Kurtosis	24.60	8.79	1.83	17.01	-0.57	8.42	0.03	-0.40
Kurt.2SE	20.98	7.50	1.56	14.51	-0.49	7.18	0.02	-0.34
Normtest.W	0.50	0.63	0.90	0.55	0.90	0.63	0.99	0.95
Normtest.p	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.01
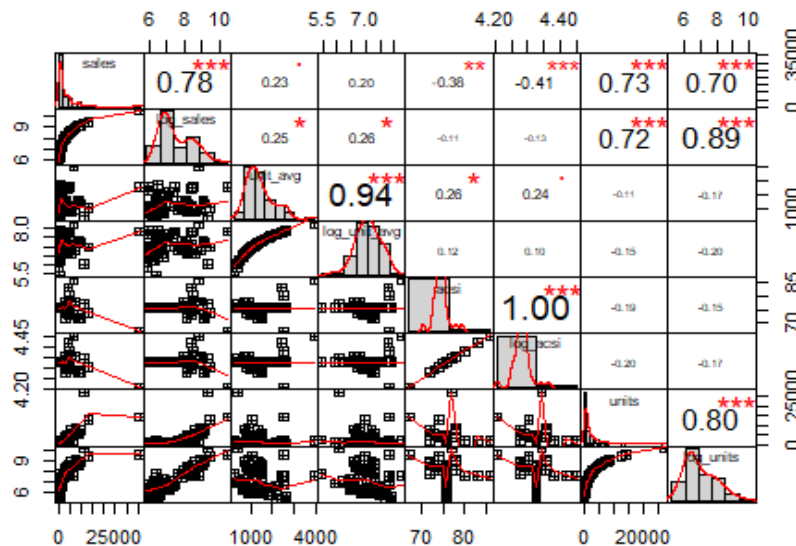
Fig. 7. Scatterplots, histograms, and correlations

Table 6. QSR Mean Values by Segment: Sales ($ millions), Average Sales per Unit (thousands), and ACSI (0 – 100)

Segments	N	Sales	Unit_avg	Acsi
Burger	14	5241.793	1610.2929	74.31933
Chicken	10	2038.080	1805.4800	76.41765
Ethnic	8	2355.025	1389.4875	76.11029
Pizza	8	2365.062	784.9750	75.94853
Sandwich	12	2155.600	1119.0417	75.98529
Seafood	2	521.850	826.0000	75.64706
Snack	11	2612.209	954.1273	75.71123

Table 7. QSR (Sales > 1500): Sales (millions), Average Sales per Unit (thousands), and ACSI (0 – 100)

n	Sales	Unit_avg	Acsi
1	35800	2500	67

The 2015 U.S. Systemwide Sales (millions) are reported by QSR as follows:

1. McDonald's – $35,800;
2. Starbucks – $13,300;
3. Subway – $11,500;
4. Burger King – $9,530.

Here McDonald's stands out for the largest sales of $35,800 million, far above the QSR industry average sales of $2,365 millions. McDonald's had 2015 ACSI of 67 which is far below the QSR industry average. These figures indicate that McDonald's is an outlier (Table 7). Since OLS regression is sensitive to outliers, McDonald's was removed from the data set before OLS regression analyses were performed.

4.2. Modeling. An Ordinary Least-Squares (OLS) regression model was developed for the overall data and also for each segment. Here the response variable was Sales, and the predictor variables were the total number of units (franchisee and company-owned units) and

ACSI. The following relationships among the response variable and the predictors were hypothesized:

♦ The higher the number of units, the higher the sales by a company.
♦ The higher the ACSI, the higher the sales by a company.

4.3. Research question #3. What factors influence the Sales of a QSR restaurant?

4.4. The overall regression model. The research question #3 was explored by running an overall regression model with sales as the response variable and acsi and units as predictors.

4.5. General model and its specifications. This study proposes the following general model to describe QSR restaurant's Sales as a function of its rating on ACSI index and its number of Units (restaurants).

$$Sales = f\,(ACSI, Units)\,. \tag{1}$$

It is postulated that the Sales of a restaurant are positively related to its ACSI and also to its number of Units.

This general model was specified as an additive model and as a multiplicative model for estimation.

4.6. Model specification #1: additive model. A linear and additive relationship among the predictor variables is specified as follows:

$$Sales = \beta_0 + \beta_1 \times ACSI + \beta_2 \times Units + \varepsilon\,. \tag{2}$$

4.7. Model specification #2: multiplicative model. A multiplicative model can be transformed to a linear and additive model by taking logs on both sides of the equation.

$$log(Sales) = \gamma_0 + \gamma_1 \times log(ACSI) + \gamma_2 \times log(Units) + \psi \tag{3}$$

The model specifications in (2) and (3) are linear in parameters that can be estimated by OLS regressions.

The estimated additive model is given by:

$$Sales = -3.676 + 61 \times ACSI + 0.62 \times Units. \qquad (4)$$

The standardized model is given below:

#>   acsi units

#> "0.042" "0.821"

The standardized (beta) regression coefficients indicate that the influence of the number of restaurants (units) on Sales (sales) is about 20 times greater than that of ACSI on Sales.

The estimated multiplicative model (converted to the log-log scale) is given below:

$$log(Sales) = -17 + 4 \times log(ACSI) + 0.89 \times log(Units) \qquad (5)$$

Table 8 presents OLS regression and log-log OLS regression results. The OLS regression model has adjusted R-squared of 66.30% while it is 78.60% for the log-log OLS regression. Although ACSI is positive in both models, it has a large standard error for OLS regression making it unreliable. The log-log OLS

regresion model was selected for futher analyses.

Table 8. OLS and log-log OLS models: 2015 dataset

	Dependent variable:	
	Sales	Log(Sales)
ACSI	60.607 (106.705)	
Units	0.618*** (0.055)	
Log(ACSI)		4.048 (2.457)
Log(Units)		0.887*** (0.058)
Constant	−3,676.455 (8,093.578)	-16.653 (10.657)
Observations	64	64
R²	0.674	0.793
Adjusted R²	0.663	0.786
Residual Std. Error (df = 61)	1,719.808	0.507
F Statistic (df = 2; 61)	63.063***	116.603***

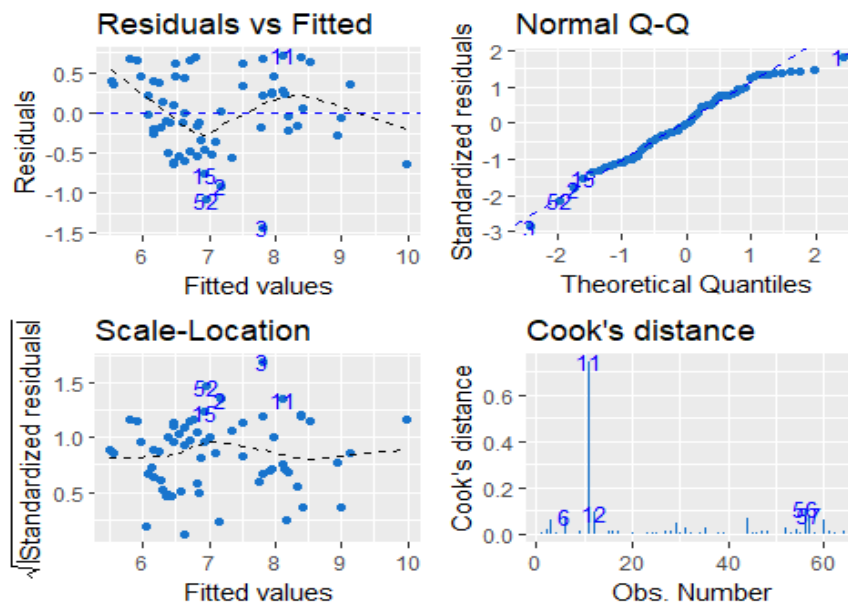Note: *p < 0.1; **p < 0.05; ***p < 0.01..

Fig. 8. Diagnostic plots (log-log OLS model)

The global test of the assumptions of the regression model was performed by the gvlma package. The results (Table 9) showed no violations of the assumptions of the regression model. This was also confirmed by four diagnostic plots, namely, Residuals vs Fitted, Normal Q-Q, Scale-Location, and Cook's Distance plots.

Table 9. Global test of log-log OLS regression model's assumptions

	Value	P-value	Decision
Global stat	4.01	0.40	Assumptions acceptable.
Skewness	2.07	0.15	Assumptions acceptable.
Kurtosis	0.29	0.59	Assumptions acceptable.
Link function	0.95	0.33	Assumptions acceptable.
Heteroscedasticity	0.71	0.40	Assumptions acceptable.

Although Chick-fil-A (number 11) stood out, no restaurant had significant influence on regression coefficients as measured by Cook's distance larger than 4. Recall that McDonald's was already removed from this analysis.

Answer #3. The overall model of the relationship between Sales and ACSI Units was statistically significant with an adjusted R squared value of 0.79 for the log-log model. There was a positive relationship between Sales and ACSI and also between Sales and the number of units of a restaurant. However, the relationship between Sales and ACSI was not significant.

Next, research question #4 was explored by running this regression model at the segment level.

Table 10. Quick service restaurants (segment = sandwich)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Arby's	3540.0	74.00	1073.0	3214	8.17	4.30	6.98	8.08
Charleys Philly Steaks	334.1	75.65	692.0	551	5.81	4.33	6.54	6.31
Dickey's Barbecue Pit	358.2	75.65	762.1	508	5.88	4.33	6.64	6.23
Firehouse Subs	648.5	75.65	726.0	945	6.47	4.33	6.59	6.85
Jason's Deli	643.6	75.65	2664.0	260	6.47	4.33	7.89	5.56
Jersey Mike's Subs	675.0	75.65	645.3	1046	6.51	4.33	6.47	6.95
Jimmy John's Sandwiches	1978.6	75.65	877.1	2405	7.59	4.33	6.78	7.79
McAlister's Deli	547.7	75.65	1634.5	361	6.31	4.33	7.40	5.89
Panera Bread	4900.0	80.00	2500.0	1972	8.50	4.38	7.82	7.59
Potbelly Sandwich Shop	387.4	75.65	1060.2	396	5.96	4.33	6.97	5.98
Quiznos	354.1	75.65	370.0	957	5.87	4.33	5.91	6.86
Subway	11500.0	77.00	424.3	27103	9.35	4.34	6.05	10.21

Table 11. Quick service restaurants (segment = chicken)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Bojangles'	1153.0	75.65	1838.7	659	7.05	4.33	7.52	6.49
Boston Market	617.9	75.65	1364.0	458	6.43	4.33	7.22	6.13
Chick-fil-A	6825.9	86.00	3977.3	1983	8.83	4.45	8.29	7.59
Church's Chicken	838.0	75.65	729.0	1131	6.73	4.33	6.59	7.03
El Pollo Loco	753.4	75.65	1798.0	433	6.62	4.33	7.49	6.07
KFC	4328.3	73.00	1000.0	4270	8.37	4.29	6.91	8.36
Pollo Tropical	366.7	75.65	2585.0	190	5.90	4.33	7.86	5.25
Popeyes Louisiana Kitchen	3100.0	75.65	1458.5	2539	8.04	4.33	7.29	7.84
Wingstop	821.2	75.65	1130.0	845	6.71	4.33	7.03	6.74
Zaxby's	1576.4	75.65	2174.3	725	7.36	4.33	7.68	6.59

4.8. Nested data by segment. The data were nested by segment to facilitate modeling by segments. Tables 10-15 present segment level data.

4.9. Research question #4. What factors influence the Sales of a QSR restaurant within each segment? This question was explored next at the segment level.

4.10. Regression models by segments. The nested data were input to a function, segment_model, to obtain OLS regression models for each segment. These regression models were then stored in by_segment database and displayed below.

```
#> # A tibble: 7 x 3
#>   segments         data    model
#>   <chr>            <list>  <list>
#> 1 Sandwich <tibble [12 x 9]> <S3: lm>
#> 2   Snack <tibble [11 x 9]> <S3: lm>
```

Table 12. Quick service restaurants (segment = burger)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Burger King	9530.0	72.00	1337.4	7126	9.16	4.28	7.20	8.87
Carl's Jr./Hardee's	3661.3	75.65	1248.7	2958	8.21	4.33	7.13	7.99
Checkers/Rally's	776.1	75.65	966.5	829	6.65	4.33	6.87	6.72
Culver's	1196.0	75.65	2183.8	559	7.09	4.33	7.69	6.33
Five Guys	1319.5	75.65	1123.0	1215	7.19	4.33	7.02	7.10

Table 12. (cont.) Quick service restaurants (segment = burger)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
In-N-Out Burger	613.2	75.65	1959.2	313	6.42	4.33	7.58	5.75
Jack in the Box	3395.5	72.00	1510.0	2249	8.13	4.28	7.32	7.72
Krystal	420.9	75.65	1169.9	352	6.04	4.33	7.06	5.86
Sonic	4390.0	75.65	1200.0	3526	8.39	4.33	7.09	8.17
Steak n Shake	1032.4	75.65	1866.9	561	6.94	4.33	7.53	6.33
Wendy's	8811.9	73.00	1540.0	5722	9.08	4.29	7.34	8.65
Whataburger	1747.0	75.65	2530.0	791	7.47	4.33	7.84	6.67
White Castle	691.3	75.65	1408.7	390	6.54	4.33	7.25	5.97

Table 13. Quick service restaurants (segment = seafood)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Captain D's	531.7	75.65	1024	510	6.28	4.33	6.93	6.23
Long John Silver's	512.0	75.65	628	737	6.24	4.33	6.44	6.60

Table 14. Quick service restaurants (segment = ethnic)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
Chipotle	4501.2	83.00	2424.0	1971	8.41	4.42	7.79	7.59
Del Taco	700.2	75.65	1376.6	544	6.55	4.33	7.23	6.30
Moe's Southwest Grill	638.1	75.65	1163.0	638	6.46	4.33	7.06	6.46
Noodles & Company	517.9	75.65	1105.0	492	6.25	4.33	7.01	6.20
Panda Express	2550.1	75.65	1487.0	1790	7.84	4.33	7.30	7.49
Qdoba Mexican Eats	753.7	75.65	1169.0	661	6.62	4.33	7.06	6.49
Taco Bell	8820.2	72.00	1460.0	6121	9.08	4.28	7.29	8.72
Taco John's	358.8	75.65	931.3	388	5.88	4.33	6.84	5.96

Table 15. Quick service restaurants (segment = pizza)

Company	Sales	Acsi	Unit_avg	Units	Log_sales	Log_acsi	Log_unit_avg	Log_units
CiCi's Pizza	440.1	75.65	1002.0	446	725.0	4.33	6.91	6.10
Domino's Pizza	4800.0	75.00	900.0	5200	8.48	4.32	6.80	8.56
Hungry Howie's Pizza	342.5	75.65	638.0	549	5.84	4.33	6.46	6.31
Little Caesars	3453.2	74.00	815.0	4237	8.15	4.30	6.70	8.35
Marco's Pizza	417.6	75.65	683.5	667	6.03	4.33	6.53	6.50
Papa John's	2882.9	78.00	892.8	3388	7.97	4.36	6.79	8.13
Papa Murphy's	880.4	75.65	623.5	1496	6.78	4.33	6.44	7.31
Pizza Hut	5703.8	78.00	725.0	7822	8.65	4.36	6.59	8.96

Table 16. Quality measures of regression models

Segments	R.squared	Adj.r.squared	Sigma	Statistic	P.value	Df	LogLik	AIC	BIC	Deviance	Df.residual
Sandwich	0.82	0.78	0.56	20.44	0	3	24.72	24.72	26.66	2.83	9
Snack	0.77	0.71	0.69	13.07	0	3	27.47	27.47	29.06	3.78	8
Chicken	0.90	0.87	0.33	32.06	0	3	10.77	10.77	11.98	0.77	7
Burger	0.94	0.93	0.27	81.74	0	3	7.26	7.26	9.51	0.72	10
Seafood	1.00	NaN	NaN	NaN	NaN	2	Inf	-Inf	-Inf	0.00	0
Ethnic	0.99	0.99	0.10	464.53	0	3	8.95	-9.89	-9.58	0.05	5
Pizza	0.97	0.96	0.22	96.90	0	3	2.56	2.89	3.20	0.25	5

```
#> 3  Chicken <tibble [10 x 9]> <S3: lm>
#> 4  Burger  <tibble [13 x 9]> <S3: lm>
#> 5  Seafood <tibble [2 x 9]>  <S3: lm>
#> 6  Ethnic  <tibble [8 x 9]>  <S3: lm>
#> 7  Pizza   <tibble [8 x 9]>  <S3: lm>
```

4.11. Displaying models with the broom package.
The glance option in the broom package in R, facilitates displaying the quality measures of multiple models. The list columns are dropped (.drop = TRUE) to facilitate display of important results.

Table 16 presents the model quality measures for each segment while Table 17 presents the regression coefficients for each segment.

Answer #4. The models depicting the relationship between Sales and ACSI at the segment level were statistically significant and had high adjusted R squared values (71% - 90%) implying good models for prediction. The seafood segment was excluded due to limited dataset of only two restaurants.

The tidy option in the broom package facilitates displaying the regression coefficients of multiple models.

The relationship between Sales and ACSI varied by segment:

- It was positive and statistically significant for chicken and ethnic restaurants.
- It was negative and not statistically significant for sanck, burgers, and pizza restaurants.
- The seafood segment was excluded due to limited dataset of only two restaurants.

Limitations

Like most studies, this study had some limitations as well:

- The main limitation was a small sample of 65 restaurants that were divided into seven segments. Some segments had too few restaurants to develop significant regression models. For example, the seafood segment had only two restaurants so a regression analysis could not be performed for this segment. Therefore, for segments with very few restaurants, the earlier visual models were quite useful.
- In addition, the ACSI score was not available for all resaturants so the mean value for this variable was substitued for the missing values.
- The factors influencing profit or customer satisfaction of a restaurant were not investigated.

Table 17. Coefficients of regression models

Segments	Term	Estimate	Std.Error	Statistic	P.Value
Sandwich	(Intercept)	-62.20	40.98	-1.52	0.16
Sandwich	log(acsi)	14.71	9.52	1.55	0.16
Sandwich	log(units)	0.77	0.14	5.62	0.00
Snack	(Intercept)	4.24	79.07	0.05	0.96
Snack	log(acsi)	-0.78	18.28	-0.04	0.97
Snack	log(units)	0.84	0.17	5.11	0.00
Chicken	(Intercept)	-31.54	11.11	-2.84	0.03
Chicken	log(acsi)	7.61	2.58	2.95	0.02
Chicken	log(units)	0.84	0.12	7.01	0.00
Burger	(Intercept)	21.30	23.27	0.92	0.38
Burger	log(acsi)	-4.61	5.28	-0.87	0.40
Burger	log(units)	0.86	0.10	8.79	0.00
Seafood	(Intercept)	6.92	NaN	NaN	NaN
Seafood	log(units)	-0.10	NaN	NaN	NaN
Ethnic	(Intercept)	-23.02	4.23	-5.45	0.00
Ethnic	log(acsi)	5.01	0.97	5.18	0.00
Ethnic	log(units)	1.23	0.04	30.41	0.00
Pizza	(Intercept)	4.11	20.51	0.20	0.85
Pizza	log(acsi)	-1.10	4.77	-0.23	0.83
Pizza	log(units)	1.05	0.08	13.54	0.00

Conclusion and recommendations for future research

This study utilized the open source R language and environment and the tidyverse package to analyze the role of American Customer Satisfaction on the Sales of Quick Service Restaurants in the US. The answers for the four research questions in this study are summarized below:

1. There were seven QSR segments: burger, chicken, sandwich, pizza, seafood, ethnic, and snack.
2. The leaders in each QSR segment were listed by seg_rank: McDonald's (burger), Chick-fil-A (chicken), Subway (sandwich), Pizza Hut (pizza), Captain D's (seafood), Taco Bell (ethnic), and Starbucks (snack).
3. This study showed a positive relationship between Sales, the number of restaurants, and American Customer Satisfaction (ACSI) with a QSR restaurant. While the overall model was good and statistically significant, the regression coefficient for the number of restaurant units was statistically significant, that for ACSI was not statistically significant.
4. The relationship between Sales and American Customer Satisfaction with a QSR varied by QSR segment. It was statistically significant and positive for chicken and ethnic restaurants and not significant for other segments. The relationship between Sales and the number of restaurant units was positive and statistically significant for all segments.

In addition, McDonald's and Chick-fil-A stood out from other QSR restaurants. McDonald's had the largest sales volume but the lowest customer satisfaction level, while Chick-fil-A had the highest customer satisfaction level and highest average sales per unit. It should also be noted that, unlike other QSR in the industry, Chick-fil-A is closed on Sundays. Both of these restaurants were the leaders in their own segments.

Studies that can be reproduced and replicated by the readers help in accumulating scientific knowledge in a discipline. For this purpose, it is imperative to utilize open source software and make both the code and the data publicly available.

Future researchers should reproduce this study with the R code and data provided here and replicate these findings with data for other years. They should also utilize a larger sample size and larger samples per segment for their model building efforts.

They should also improve upon the linear model used here by exploring non-linear models and by utilizing more comprehensive data about characteristics of various restaurants. They should also replicate these findings for data from other countries. They could also explore factors that influence a company's profit and customer satisfaction and provide guidelines to managers.

References

1. ACSI LLC (2017). American Customer Satisfaction Index Benchmarks By Industry: Limited-Service Restaurants. Retrieved from http://www.theacsi.org/index.php?option=com_content&view=article&id=147&catid=&Itemid=212&i/
2. Chang, Winston. (n.d.). Multiple graphs on one page (ggplot2). In *Cookbook for R*. Retrieved from http://www.cookbook-r.com/Graphs/Multiple_graphs_on_one_page_(ggplot2)/
3. Held, Fabian P. (2017). Introduction to the Tidyverse. Retrieved from http://www.fabianheld.com/r-tidyverse/ (accessed on June 23, 2017).
4. Levy, Michael. (2016). *Tidyverse Tutorial.* Retrieved from https://www.r-bloggers.com/tidyverse-tutorial/ (accessed on October 17, 2016)
5. QSR, Inc. (2016). *The QSR 50: An inside look at the top 50 brands in quick service and fast casual.* Retrieved from https://www.qsrmagazine.com/reports/qsr50-2016-top-50-chart (accessed on August, 2016)
6. R & Data Science. Retrieved from http://www.sthda.com/english/wiki/ggplot2-dot-plot-quick-start-guide-r-software-and-data-visualization
7. R Views (2017). *What is the tidyverse?* Retrieved from https://www.r-bloggers.com/what-is-the-tidyverse/ (accessed on June, 2017).
8. The R Foundation (n.d.). The R Project for Statistical Computing. Retrieved from https://www.r-project.org/
9. The tidyverse: dplyr, ggplot2, and friends. (n.d). Retrieved from https://monashbioinformaticsplatform.github.io/
10. Tibble Data Format in R: Best and Modern Way to Work with Your Data. Retrieved from http://www.sthda.com/english/wiki/tibble-data-format-in-r-best-and-modern-way-to-work-with-your-data
11. Tidyverse (n.d.) Retrieved from http://tidyverse.tidyverse.org/
12. Tidyverse packages (n.d.). Retrieved from https://www.tidyverse.org/packages/
13. Wickham, Hadley, & Grolemund, Garrett. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly. Retrieved from http://r4ds.had.co.nz/
14. Wickham, Hadley. (2016). Tidyverse 1.0.0. Retrieved from https://blog.rstudio.com/2016/09/15/tidyverse-1-0-0/ (accessed on September 5, 2016).
15. Wickham, Hadley. (2017). *The Tidy Tools Manifesto.* Retrieved from https://mran.microsoft.com/web/packages/tidyverse/vignettes/manifesto.html (accessed on January 27, 2017).
16. Wickhham, Hadley. (2016). G*gplot2*. Second edition. Springer.
17. Wilke, C. O. (2016). *Introduction to cowplot.* Retrieved from https://cran.r-project.org/web/packages/

Appendix (R Code)

```r
knitr::opts_chunk$set(echo = FALSE, comment = "#>", collapse = TRUE,  message = FALSE,
                      warning = FALSE, scipen = 5, digits = 3)

# install.packages("tidyverse")
library(tidyverse)
library(modelr)
library(ggrepel)

# read data file
qsr65 <- read_csv("qsr65_v1.csv")

# print data table
library(kableExtra)
knitr::kable(qsr65, format = "latex", digits = 2,
             caption = 'Quick Service Restaurants (USA, 2015)', booktabs = T) %>%
# kable_styling(latex_options = c("striped", "scale_down", font_size = 7))
 kable_styling(bootstrap_options = "striped", font_size = 7)

# Helper function by Wickham & Grolemund (2016, p. 434). Set number of digits to reasonable
# accuracy and make it easier to read numbers by inserting commas.
comma <- function(x) format(x, digits = 2, big.mark = ",")

library(tidyverse)
library(modelr)

# replace $ and commas with blanks "" and change character to numeric
qsr65$sales <- as.numeric(gsub("[\\$,]", "", qsr65$sales))
qsr65$unit_avg <- as.numeric(gsub("[\\$,]", "", qsr65$unit_avg))

library(mosaic)
sales_seg <- favstats(sales ~ segments, data= qsr65)

knitr::kable(sales_seg, caption = 'Quick Service Restaurants Sales by Segment') %>%
kable_styling(bootstrap_options = "striped", font_size = 7)

unit_avg_seg <- favstats(unit_avg ~ segments, data=qsr65)
knitr::kable(unit_avg_seg, caption = 'Quick Service Restaurants Average Unit Sales by Segment')
%>%
kable_styling(bootstrap_options = "striped", font_size = 7)

acsi_seg <- favstats(acsi ~ segments, data=qsr65)
knitr::kable(acsi_seg, caption = 'Quick Service Restaurants ACSI Index by Segment') %>%
kable_styling(bootstrap_options = "striped", font_size = 7)

library(ggplot2)
# theme_set(theme_classic)
library(ggrepel)
library(cowplot)
theme_set(theme_gray()) # switch to default ggplot2 theme

p1<- ggplot(qsr65, aes(segments, sales)) + geom_point(na.rm = TRUE) +
  stat_summary(geom = "point", fun.y = "mean", color = "red", size = 2) + coord_flip()

p2 <- ggplot(qsr65, aes(segments, unit_avg)) + geom_point(na.rm = TRUE) +
  stat_summary(geom = "point", fun.y = "mean", color = "red", size = 2) + coord_flip()

p3 <- ggplot(qsr65, aes(segments, acsi)) + geom_point(na.rm = TRUE) +
  stat_summary(geom = "point", fun.y = "mean", color = "red", size = 2) + coord_flip()

# bar chart
p4 <- ggplot(data = qsr65, mapping = aes(x = segments)) + geom_bar(na.rm = TRUE) + coord_flip()

p <- plot_grid(p1, p3, p2, p4, align = "v")

title <- ggdraw() + draw_label("QSR Characteristics by Segment", fontface='bold')
plot_grid(title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title margins
```

```r
# Boxplots with dotplots with filtered data
p5 <- ggplot(qsr65, aes(segments, sales)) +
  geom_boxplot(fill="plum") + geom_jitter(width = 0.2) +
  coord_flip() + labs(title="Box plot", y="2015 Sales", x="Segment")

p6 <- ggplot(qsr65, aes(segments, sales)) +
  geom_violin(fill="plum") + geom_jitter(width = 0.2) +
  coord_flip() + labs(title="Violin plot", y="2015 Sales", x="Segment")

p7 <- ggplot(qsr65, aes(segments, unit_avg)) +
  geom_boxplot(fill="plum") + geom_jitter(width = 0.2) +
  coord_flip() + labs(title="Box plot", y="2015 Average Unit Sales", x="Segment")

p8 <- ggplot(qsr65, aes(segments, unit_avg)) +
  geom_violin(fill="plum") + geom_jitter(width = 0.2) +
  coord_flip() + labs(title="Violin plot", y="2015 Average Unit Sales", x="Segment")

p <- plot_grid(p5, p6, p7, p8, align = "v")

title <- ggdraw() + draw_label("QSR Boxplots and Violin plots by Segment", fontface='bold')
plot_grid(title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title margins

p9 <- ggplot(qsr65, aes(x = acsi, y=sales, label = company, color=segments)) +
  geom_point(na.rm = TRUE, size = 2, shape = 1) +
  ggrepel::geom_text_repel(aes(label = company), size = 3) +
  geom_jitter(width = 0.1, height = 0.1) +
  geom_smooth(method = 'loess', color = "blue", size = 1.5, na.rm = TRUE, se = FALSE)

p10 <- ggplot(qsr65, aes(x = acsi, y=unit_avg, label = company, color=segments)) +
  geom_point(na.rm = TRUE, size = 2, shape = 1) +
  ggrepel::geom_text_repel(aes(label = company), size = 3) +
  geom_jitter(width = 0.1, height = 0.1) +
  geom_smooth(method = 'loess', color = "blue", size = 1.5, na.rm = TRUE, se = FALSE)

p <- plot_grid(p9, p10, ncol = 1)

title <- ggdraw() + draw_label("QSR Sales and Average Unit Sales by ACSI", fontface='bold')
plot_grid(title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title margins

# replace missing valuse for acsi with its average value
qsr65$acsi[is.na(qsr65$acsi)] <- mean(qsr65$acsi, na.rm = TRUE)

best_sales_in_segment <- qsr65 %>% group_by(segments) %>% filter(row_number(desc(sales)) == 1)

p11 <- ggplot(qsr65, aes(x = acsi, y=sales, color=segments)) +
  geom_point(na.rm = TRUE, size = 3, shape = 1, data = best_sales_in_segment) +
  ggrepel::geom_text_repel(aes(label = company), size = 3, data = best_sales_in_segment) +
  geom_jitter(width = 0.1, height = 0.1)

# Remove confidence intervals and extend regression lines
p12 <- ggplot(qsr65, aes(x = acsi, y=sales, color=segments)) +
  geom_point(na.rm = TRUE, size = 3, shape = 1, data = best_sales_in_segment) +
  ggrepel::geom_text_repel(aes(label = company), size = 3, data = best_sales_in_segment) +
  geom_jitter(width = 0.1, height = 0.1) +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)

p <- plot_grid(p11, p12, ncol = 1)

title <- ggdraw() + draw_label("QSR Best Sales in Each Segment by ACSI", fontface='bold')
plot_grid(title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title margins

best_unit_avg_in_segment <- qsr65 %>%
  group_by(segments) %>%
  filter(row_number(desc(unit_avg)) == 1)

p13 <- ggplot(qsr65, aes(x = acsi, y=unit_avg, color=segments)) +
  geom_point(na.rm = TRUE, size = 3, shape = 1, data = best_unit_avg_in_segment) +
  ggrepel::geom_text_repel(aes(label = company), size = 3, data = best_unit_avg_in_segment) +
  geom_jitter(width = 0.1, height = 0.1)
```

```r
# Remove confidence intervals and extend regression lines
p14 <- ggplot(qsr65, aes(x = acsi, y=unit_avg, color=segments)) +
  geom_point(na.rm = TRUE, size = 3, shape = 1, data = best_unit_avg_in_segment) +
  ggrepel::geom_text_repel(aes(label = company), size = 3, data = best_unit_avg_in_segment) +
  geom_jitter(width = 0.1, height = 0.1) +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)


p <- plot_grid(p13, p14, ncol = 1)


title <- ggdraw() + draw_label("QSR Sales and Average Sales per Unit by Segment",
fontface='bold')
plot_grid(title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title margins


# calculate total number of units for a company
qsr65$units <- qsr65$fr_u + qsr65$co_u


# Transform to natural log scale
qsr65$log_sales <- log(qsr65$sales)
qsr65$log_acsi <- log(qsr65$acsi)
qsr65$log_unit_avg <- log(qsr65$unit_avg)
qsr65$log_units <- log(qsr65$units)


library(PerformanceAnalytics)
chart.Correlation(qsr65[, c("sales", "log_sales", "unit_avg", "log_unit_avg", "acsi",
                            "log_acsi", "units", "log_units")], method="pearson",
                    histogram=TRUE, pch=12, main = "Correlations and Histograms")


# Descriptive statistics of selected variables
library(pastecs)
mytable <- qsr65 %>% select(-company, -rank, -segments, -seg_rank, -fr_u, -co_u, -change) %>%
  stat.desc(qsr65, norm = TRUE) # descriptive statistics


library(knitr)
library(kableExtra)
knitr::kable(mytable, format = "latex", digits = 2,
             caption = "Descriptive Statistics", booktabs = T) %>%
# kable_styling(latex_options = c("striped", "scale_down"))
kable_styling(bootstrap_options = "striped", font_size = 7)


# Summary of full dataset
library(dplyr)
segment_means <- qsr65 %>% group_by(segments) %>%
  summarize(n = n(), sales = mean(sales), unit_avg = mean(unit_avg),
            acsi = mean(acsi, na.re = TRUE))


knitr::kable(segment_means, caption = 'QSR Mean Values by Segment:
             Sales ($ millions), Average Sales per Unit (thousands), and ACSI (0 – 100)')  %>%
kable_styling(bootstrap_options = "striped", font_size = 7)


# summary of outliers
outliers <- qsr65 %>% filter(sales > 15000) %>% summarize(n = n(), sales, unit_avg, acsi)


knitr::kable(outliers, caption = 'QSR (Sales > $1,500):
             Sales ($ millions), Average Sales per Unit (thousands), and ACSI (0 – 100)') %>%
kable_styling(bootstrap_options = "striped", font_size = 7)


# filtered data after removing outliers
qsr65_short <- qsr65 %>%
  filter(sales <= 15000)


# use 64 companies after removing McDonald's as an outlier
qsr65 <- qsr65_short


# OLS model on overall dataset (Additive Model)
fit_qsr <- lm(sales ~ acsi + units, data = qsr65)


# Standardize Regression Model
library(QuantPsyc)
fit_qsr_beta <- lm.beta(fit_qsr)
```

```r
comma(fit_qsr_beta)
#knitr::kable(fit_qsr_beta,  format = "latex", digits = 2,
#            caption = 'Standardized Regression Coefficients for Overall Dataset',
#            booktabs = T)

# OLS model on overall dataset (Multiplicative Model)
fit_qsr_log <- lm(log(sales) ~ log(acsi) + log(units), data = qsr65)

library(stargazer)
stargazer(fit_qsr, fit_qsr_log, title="OLS and Log-Log OLS Models: 2015 Dataset",
          dep.var.labels=c("Sales", "Log(Sales)"),
          covariate.labels=c("ACSI", "Units", "Log(ACSI)", "Log(Units)"), header=FALSE)

library(ggfortify)
ggplot2::autoplot(fit_qsr_log, which = 1:4, colour = 'dodgerblue3', smooth.colour = 'black',
                  smooth.linetype = 'dashed', ad.colour = 'blue', label.size = 3.5,
                  label.n = 5, label.colour = 'blue')

cooksd <- cooks.distance(fit_qsr_log)
# influential row numbers
influential <- as.numeric(names(cooksd)[(cooksd > 4*mean(cooksd, na.rm=T))])
influential

# Global test of model assumptions
library(gvlma)
gvmodel <- gvlma(fit_qsr_log)
global_test <- summary(gvmodel)

library(knitr)
library(kableExtra)
knitr::kable(global_test, digits = 2,
             caption = "Global Test of Log-Log OLS Regression Model's Assumptions",
             booktabs = T) %>% kable_styling(latex_options = c("striped", "scale_down")) %>%
kable_styling(bootstrap_options = "striped", font_size = 7)

# Select variables for analysis
qsr65_b <- qsr65 %>% dplyr::select(-rank, -seg_rank, -fr_u, -co_u, -change)

# nested  data (Wickham p. 401-402)
by_segment <- qsr65_b %>%  group_by(segments) %>% nest()

# sandwich
knitr::kable(by_segment$data[[1]], format = "latex", digits = 2,
             caption = 'Quick Service Restaurants (Segment = sandwich)', booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

# snack
knitr::kable(by_segment$data[[2]], format = "latex", digits = 2,
caption = 'Quick Service Restaurants (Segment = snack)', booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

# chicken
knitr::kable(by_segment$data[[3]], format = "latex", digits = 2,
             caption = 'Quick Service Restaurants (Segment = chicken)', booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

# burger
knitr::kable(by_segment$data[[4]], format = "latex", digits = 2, caption =
             'Quick Service Restaurants (Segment = burger)', booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

# seafood
knitr::kable(by_segment$data[[5]], format = "latex", digits = 2,
             caption = 'Quick Service Restaurants (Segment = seafood)',  booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

# ethnic
knitr::kable(by_segment$data[[6]], format = "latex", digits = 2,
             caption = 'Quick Service Restaurants (Segment = ethnic)', booktabs = T) %>%
```

```r
  kable_styling(latex_options = c("striped", "scale_down"))

# pizza
knitr::kable(by_segment$data[[7]], format = "latex", digits = 2,
caption = 'Quick Service Restaurants (Segment = pizza)', booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

segment_model <- function(df){lm(log(sales) ~ log(acsi) + log(units), data = df)}

# build models by segments
models <- map(by_segment$data, segment_model)

# store this information in by_segment database
by_segment <- by_segment %>% mutate(model = map(data, segment_model))
by_segment

# model quality
glance <- by_segment %>% mutate(glance = map(model, broom::glance)) %>%
  unnest(glance, .drop = TRUE)

knitr::kable(glance, format = "latex", digits = 2, caption = 'Models Quality Measures',
             booktabs = T) %>% kable_styling(latex_options = c("striped", "scale_down"))

# model coefficients
tidy <- by_segment%>% mutate(tidy = map(model, broom::tidy)) %>% unnest(tidy, .drop = TRUE)

knitr::kable(tidy, format = "latex", digits = 2, caption = 'Regression Model Coefficients',
             booktabs = T) %>% kable_styling(latex_options = c("striped", "scale_down"))
```