ANALYSIS

The dataset from Store 4818 contains Jun 16 to Sep 2017 transaction data that have 12 variables. The variables are:

* Business Date
* Ticket Open Time
* Ticket Close Time
* Hop Out Time
* Ticket Number
* Item Price
* Item Quantity
* Product Description
* Total Sales Amount
* Total Discount Amount
* Tender
* Sales Type

Total dataset for store 4818 has about 3MM records. After removing items condiment items from the transaction data, we have 2,6MM records in the dataset. Additionally, we created derived variables from the dataset.

The variables created are:

* Seasonal: Hot, Cold, Cool
* Day of Week: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday
* Daypart: Breakfast, Lunch, Snack, Dinner, Late Night Snack
* Sales Amount: Gross Sales before tax

Table 1: Categorization of variables

|  |  |
| --- | --- |
| Category | Seasonal (Month) |
| Hot | Jun, Jul, Aug |
| Cold | Nov, Dec, Jan |
| Cool | Mar, Apr, May, Sep |

|  |  |
| --- | --- |
| Category | Daypart (Hour) |
| Breakfast | 5am – 11 am |
| Lunch | 11am – 3 pm |
| Snack | 3pm – 5 pm |
| Dinner | 5 pm – 9pm |
| Late Night Snack | 9pm –2am |

The variable Sales Amount is calculated by multiplying Item Price and Item Quantity.

We summarized the variable SALES AMOUNT by TICKET NUMBER to calculate total sales before tax. Once we calculated total sales amount, we aggregated our dataset by BUSINESS DATE and DAY part to get total ticket count, total sales amount and average hop out time. We also created additional two metrics Average Sale Per Ticket and Average Sales Per Service Minute from our aggregated dataset. The aggregated dataset has 616,086 records.

Daily Gross Sales Plot on Figure 1 clearly shows that there is a strong seasonality component in Sales. It appears that sales hike up during summer months and drop during cooler months.

Figure 1: Store 4818 – Daily Gross Sales Trend

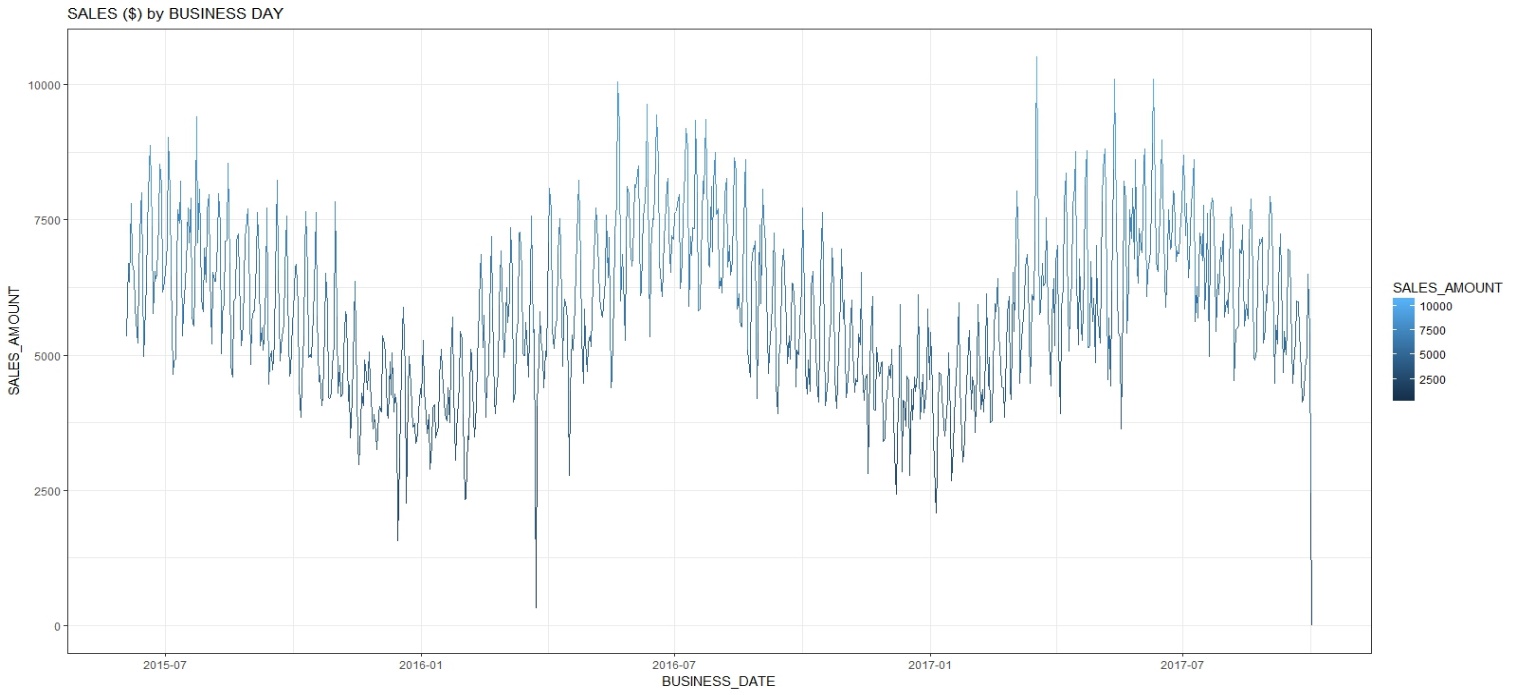
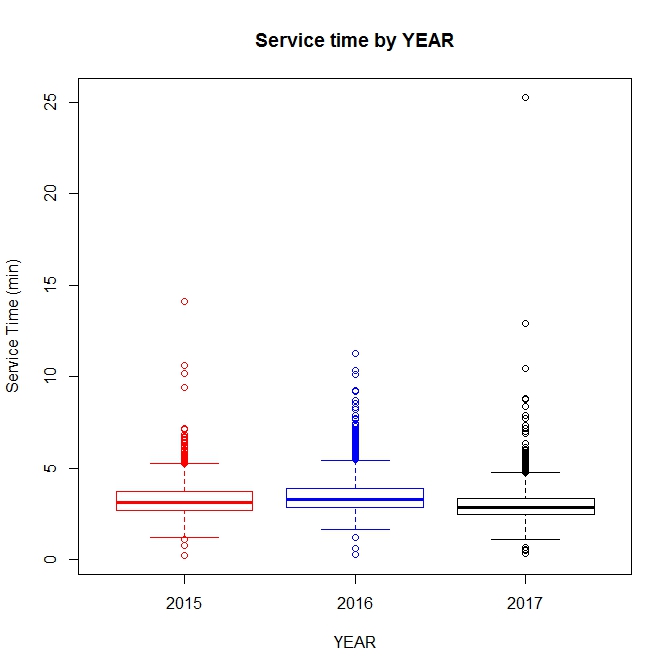


Table 2 shows annualized gross sales before tax.

Table 2: Store 4818 Annualized Gross Sales

|  |  |  |  |
| --- | --- | --- | --- |
| Month | 2015 | 2016 | 2017 |
| Total | 2,054,571 | 2,098,828 | 2,209,860 |

Figure 2: BoxPlot of Service Time( Hop\_Out\_Time)

There are few outliers in the variable service time. It is unusual for QSR to have service time more than 8 minutes. The outlier in 2017 occurred on Sunday during late night shift at the stall indicating it might have happened when the restaurant is not busy.

Figure 3: BoxPlot of Sales ($)

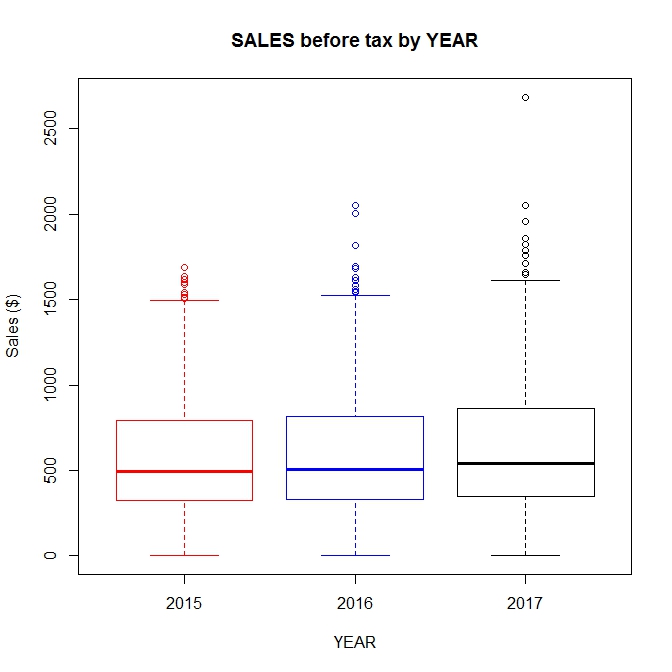
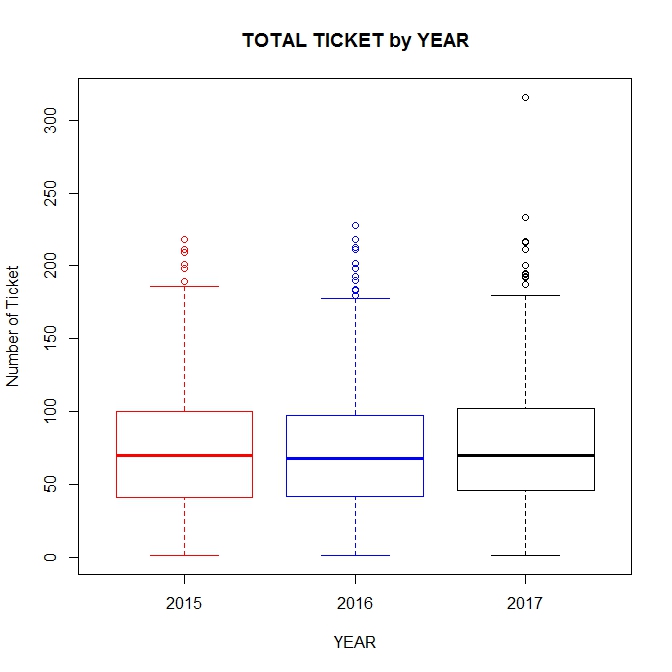


Figure 4: Total Ticket by Year



The outliers in total ticket number and total sales appeared to happened in the same day on Saturday during summer time.

Since the outliers do not appear to be an erroneous entry, we are going to leave it in the dataset.

Table 3: Descriptive Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sales Amount | Hop Out Time | Ticket # | Ave Sales Per Order | Ave\_Sales Per Service MIn |
| Min | 0 | 0.22 | 1 | 0 | 0 |
| 1st Q | 334.9 | 2.67 | 43 | 6.99 | 2.15 |
| Median | 513.5 | 3.12 | 69 | 8.08 | 2.54 |
| Mean | 581.1 | 3.27 | 72.26 | 8.09 | 2.59 |
| 3rd Q | 824.4 | 3.70 | 99 | 9.12 | 2.96 |
| Max | 2686.9 | 25.29 | 316 | 32.28 | 26.80 |
| Sd | 334.17 | 0.94 | 39 | 1.72 | 0.80 |

Figure 5: Scatter plot for Sales amount and Number of Ticket

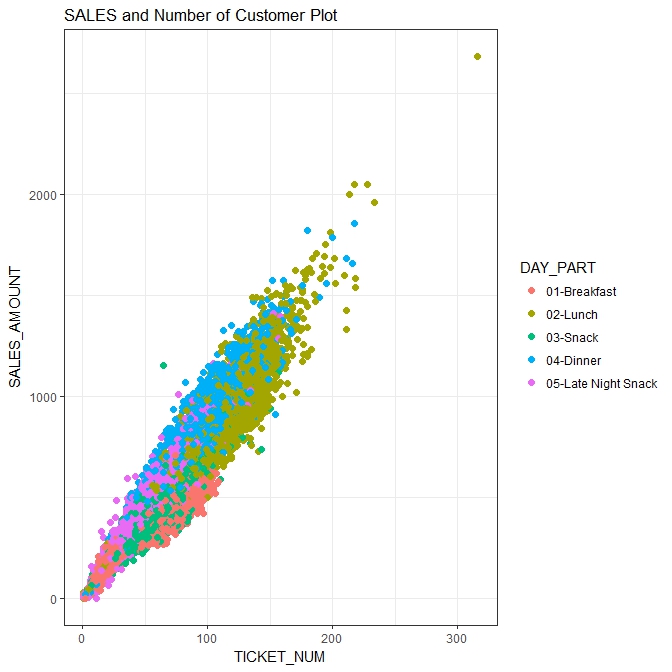
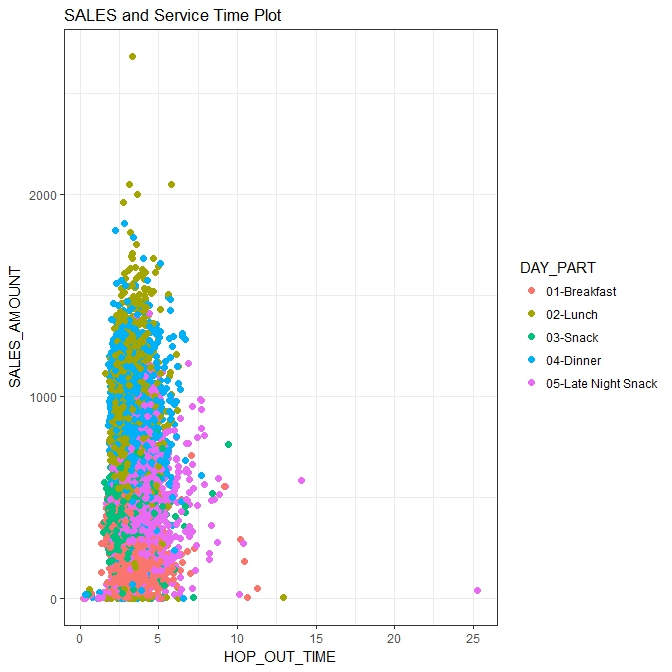


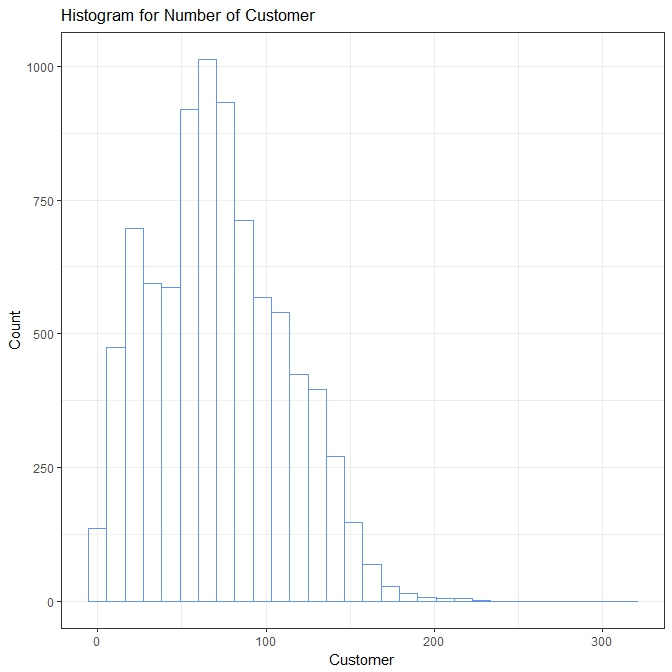
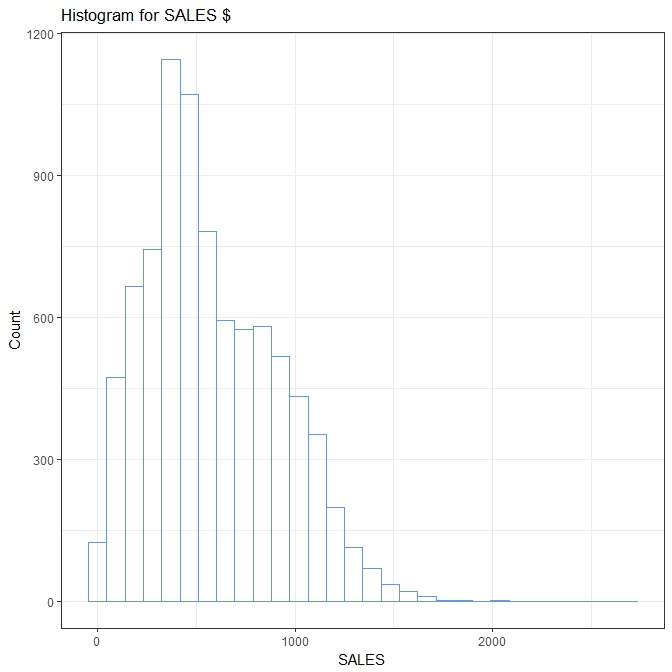
Figure 5 presents a nice view of how the number of tickets drives the total sales. It also shows that the dayparts LUNCH AND DINNER are major contributors in total sales.

Figure 6: Scatter plot for Sales amount and Service Time

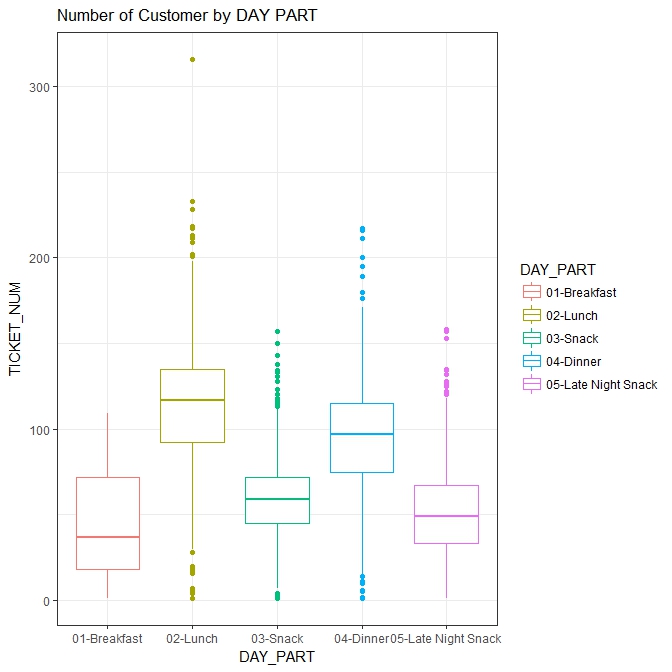


There is no apparent correlation between sales and service time. However, service time is slower during dinner and late-night snack time and goes beyond the industry average.

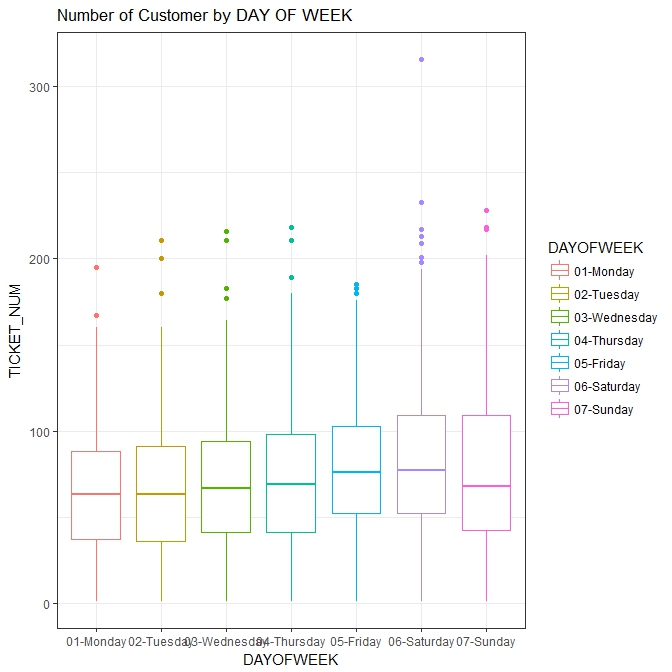
Figure 7: Histogram of Sales Amount and Number of Ticket



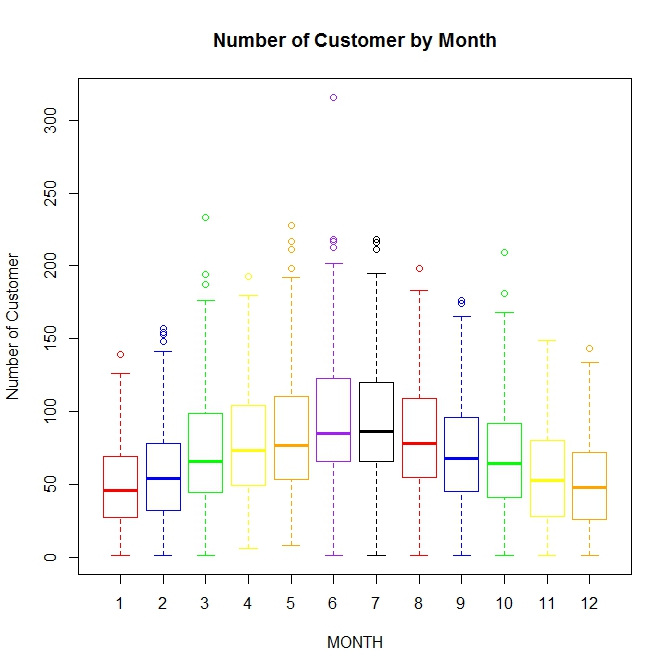
Histograms in Figure 7 show that data distribution of SALES\_AMOUNT and Number of TICKET is positively skewed. The number of Ticket appears pretty close to be normally distributed.

Figure 8: 

As we previously observed, the number of customers that arriving at the store is different by daypart.

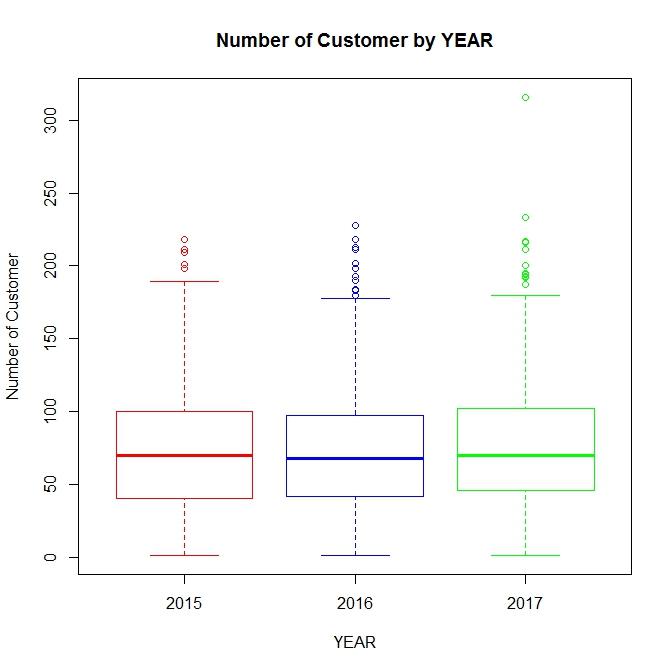
Figure 9: 

While the store has a significant fluctuation in the number of customer by daypart, it does not appear that the number of customer who come to varies significantly by day of week. There is a slight increase in the number of customers on the weekend.

Figure 10: 

However, Figure 10 clearly shows that there is a seasonality component to the number of customers.

Figure 11:



The boxplot for the number of customers shows the store has not seen any growth year over year. While it is not good news for the business, we can also say that the store definitely has a steady customer base that we can leverage to increase sales.

Table 4: Average Sales Per Ticket

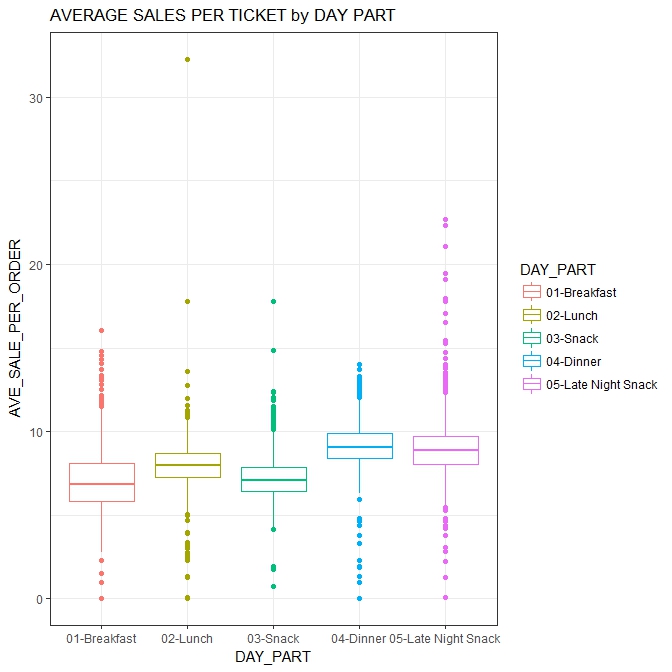
|  |  |  |  |
| --- | --- | --- | --- |
|  | 2015 | 2016 | 2017 |
| Average Sales per Ticket | 7.79 | 8.20 | 8.18 |

Table 5: Average Sales Per Ticket

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| 8.55 | 8.60 | 8.22 | 8.09 | 8.12 | 7.71 | 7.72 | 7.76 | 8.02 | 8.03 | 8.17 | 8.73 |

Average Sales per Ticket fluctuates month by month. However, it appears that the average sales per ticket in summer season is significantly lower than winter season.

Figure 12:

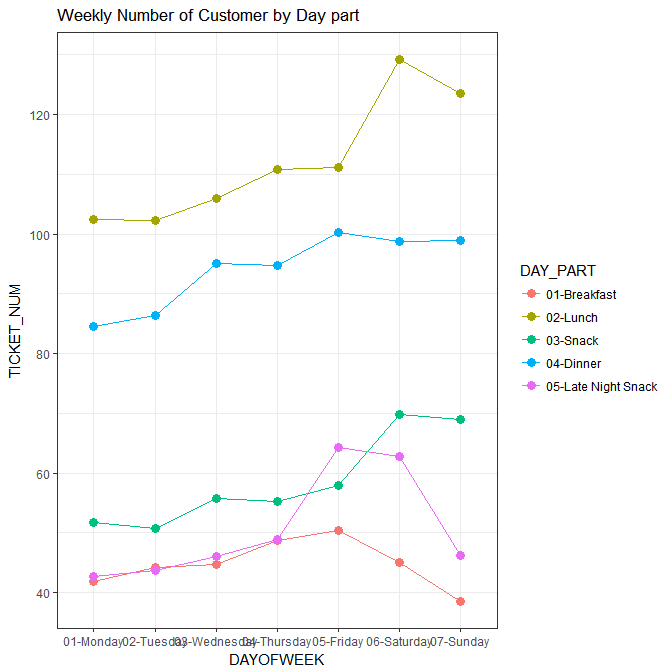


We observed the average service time for dinner and late-night snack were slightly longer than the rest of daypart. But Figure 12 shows that average sales per ticket for dinner and late-night snack time is significantly higher than the rest of daypart.

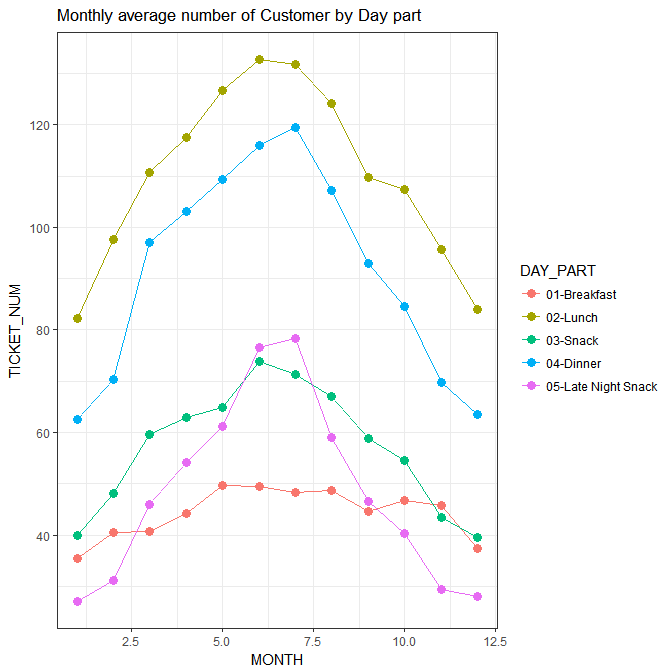
Average Sales per Ticket by Day of Week

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| 7.79 | 7.77 | 7.79 | 7.94 | 8.33 | 8.49 | 8.52 |

It also appears that average sales per ticket on weekend are higher than that of weekday.

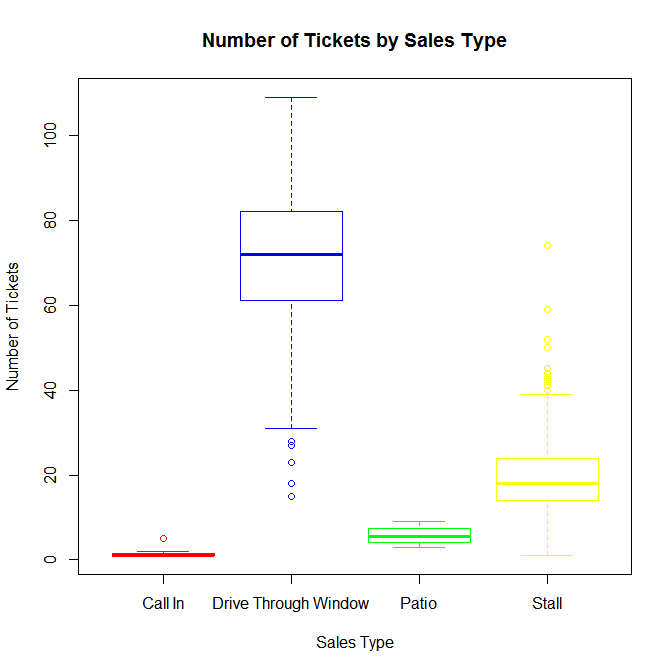


The average number of customers for each daypart is different throughout the day of week. On Monday and Tuesday in all daypart is slower than the rest of the week. Breakfast is the slowest time of a day all week. There are distinct pattern combinations with day of week and day part. The snack time is consistently flat during the weekday but it goes up in the weekend while late night snack goes up on Friday and Saturday and remain flat for the rest of the week.

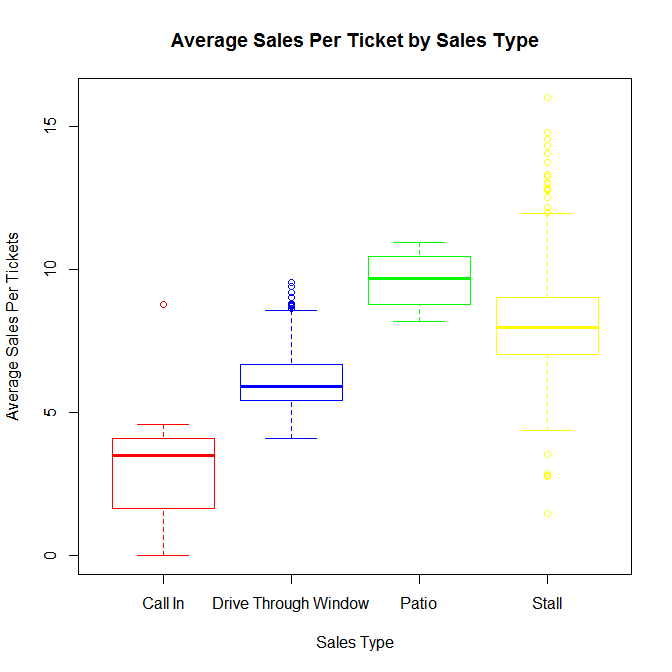


The monthly average number of customer follows seasonal trend for all day part. However, the breakfast is least impacted by seasonality.

BREAKFAST



The biggest number of customers are coming through drive through window for breakfast time



However, the customers who order at patio have highest sales per ticket.

**Models for weekly number of tickets for breakfast**

The Poisson Model 1.

Call:

glm(formula = TICKET\_NUM ~ ., family = "poisson", data = training\_set)

Deviance Residuals:

Min 1Q Median 3Q Max

-7.9324 -2.1152 -0.2003 1.6092 7.7722

Coefficients: (3 not defined because of singularities)

Estimate Std. Error z value Pr(>|z|)

(Intercept) 191.788125 13.557222 14.147 <2e-16 \*\*\*

YEAR -0.093041 0.006723 -13.840 <2e-16 \*\*\*

MONTH -0.001229 0.001402 -0.877 0.38

DAYOFWEEK\_01.Monday 0.129895 0.015344 8.465 <2e-16 \*\*\*

DAYOFWEEK\_02.Tuesday 0.172009 0.015199 11.317 <2e-16 \*\*\*

DAYOFWEEK\_03.Wednesday 0.186118 0.015151 12.284 <2e-16 \*\*\*

DAYOFWEEK\_04.Thursday 0.266530 0.014913 17.873 <2e-16 \*\*\*

DAYOFWEEK\_05.Friday 0.303698 0.014885 20.403 <2e-16 \*\*\*

DAYOFWEEK\_06.Saturday 0.181563 0.015291 11.874 <2e-16 \*\*\*

DAYOFWEEK\_07.Sunday NA NA NA NA

SALES.TYPE\_Call.In -3.974266 0.288864 -13.758 <2e-16 \*\*\*

SALES.TYPE\_Drive.Through.Window 1.336982 0.009524 140.380 <2e-16 \*\*\*

SALES.TYPE\_Patio NA NA NA NA

SALES.TYPE\_Stall NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

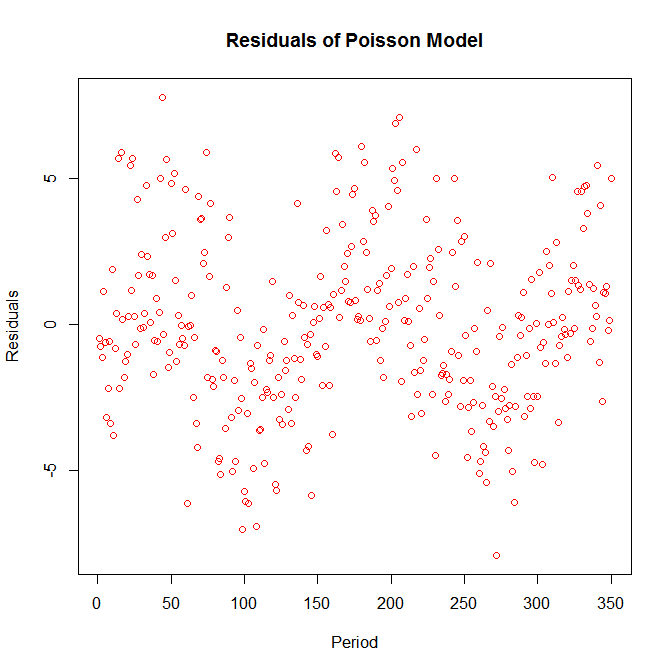
(Dispersion parameter for poisson family taken to be 1)

Null deviance: 30758.0 on 349 degrees of freedom

Residual deviance: 3020.4 on 339 degrees of freedom

AIC: 5415.1

Number of Fisher Scoring iterations: 4



lm(formula = TICKET\_NUM ~ ., data = training\_set)

Residuals:

Min 1Q Median 3Q Max

-136.418 -27.042 -3.398 23.341 137.100

Coefficients: (3 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 36686.4503 8688.2717 4.223 3.11e-05 \*\*\*

YEAR -18.1728 4.3081 -4.218 3.16e-05 \*\*\*

MONTH -0.2482 0.8958 -0.277 0.781853

DAYOFWEEK\_01.Monday 22.2411 9.1614 2.428 0.015715 \*

DAYOFWEEK\_02.Tuesday 29.9169 9.2136 3.247 0.001283 \*\*

DAYOFWEEK\_03.Wednesday 32.7558 9.1615 3.575 0.000401 \*\*\*

DAYOFWEEK\_04.Thursday 49.0677 9.2589 5.300 2.10e-07 \*\*\*

DAYOFWEEK\_05.Friday 56.0198 9.1721 6.108 2.77e-09 \*\*\*

DAYOFWEEK\_06.Saturday 31.7511 9.3063 3.412 0.000723 \*\*\*

DAYOFWEEK\_07.Sunday NA NA NA NA

SALES.TYPE\_Call.In -102.2672 18.2740 -5.596 4.52e-08 \*\*\*

SALES.TYPE\_Drive.Through.Window 228.2643 4.9452 46.159 < 2e-16 \*\*\*

SALES.TYPE\_Patio NA NA NA NA

SALES.TYPE\_Stall NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 45.79 on 339 degrees of freedom

Multiple R-squared: 0.8728, Adjusted R-squared: 0.8691

F-statistic: 232.6 on 10 and 339 DF, p-value: < 2.2e-16



Series: train\_ts[, 4]

ARIMA(1,1,3)

Coefficients:

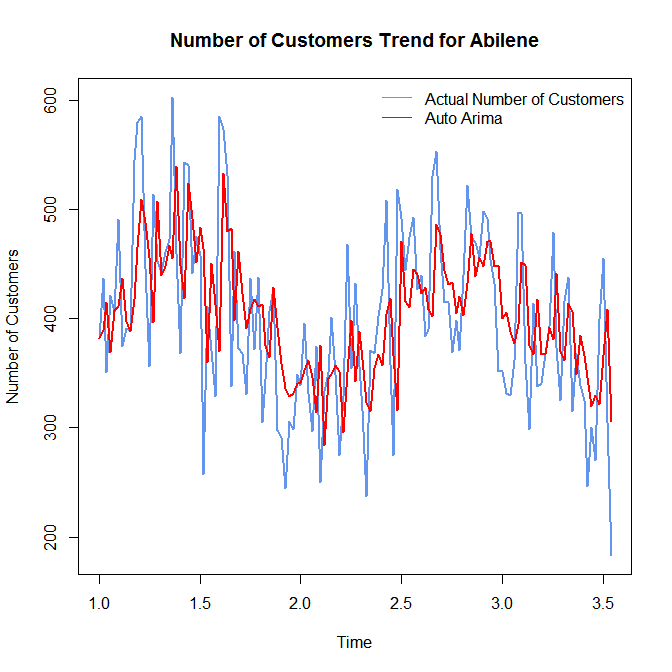
ar1 ma1 ma2 ma3

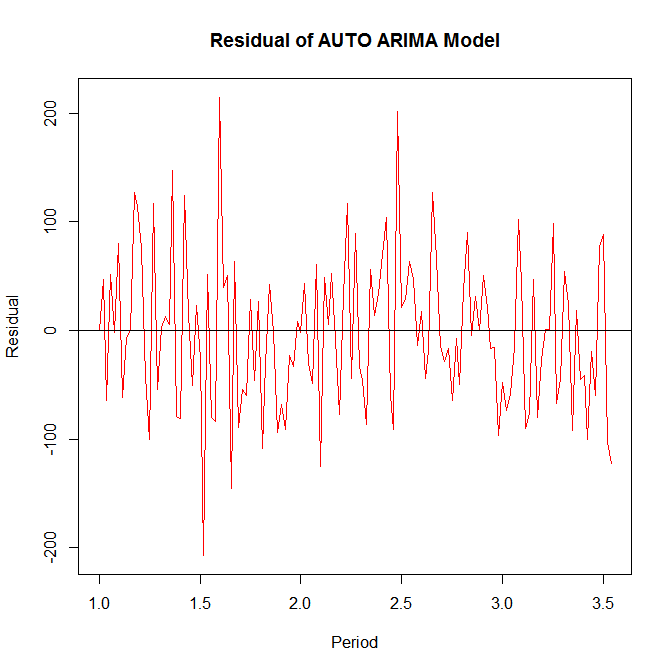
-0.8451 0.3706 -0.6837 -0.3500

s.e. 0.1195 0.1398 0.0849 0.0968

sigma^2 estimated as 5214: log likelihood=-750.69

AIC=1511.39 AICc=1511.86 BIC=1525.8





|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Poisson | Linear Regression | Auto Arima | GBM |
| AIC | 5415 | 3682 | 1511.39 |  |
| BIC | 5457 | 3729 | 1525.8 |  |
| RMSE | 43.69 | 47.98 | 94.23 | 40.32 |
| MAE | 34.02 | 39.24 | 70.41 | 32.12 |

