Leadership and Consulting

MCDONALD'S CASE EXPLORATORY DATA ANALYSIS

By: Suyog Mahale

INDEX

THE BUSINESS PROBLEM

- According to the Regional Manager, McDonald's has been losing about 5% customer traffic every year for the past 3 years. The Regional Manager is looking to understand why this is happening.
- To combat the loss of customer traffic, McDonald's has been launching promotions such as the All-Day Breakfast promotion, McMuffin promotion, etc. The Regional Manager wants us to analyze the effectiveness of the recent promotions.

THE DATA

The Regional Manager has provided us with three files of data:

- M395_weekly_sales - Data about the weekly sales of McDonald's for the past three years.
- M395_rest_facts - Details of all the restaurants under the Regional Manager.
- McD Data Variable Descriptions - Describes the variables present in the weekly sales data sheet and restaurant facts data sheet.

The variables in the data which most directly relate to the business problem are:

1. Restaurant data
    a. REST_KEY – Unique identifier for each restaurant
    b. REST_HISP_CONS_MKT – Percentage of trade area that is Hispanic.
    c. REST_AFR_AMR_CONS_MKT - Percentage of trade area that is African American.
    d. REST_ASIAN_CONS_MKT - Percentage of trade area that is Asian.
    e. REST_PLYPL_TYP
    f. REST_TYP – Type of restaurant whether it is in a mall or freestanding.
    g. Ethnic_label – Ethnic population around the area.
    h. Incomeq_label – Income of population in the area.
    i. Urban_label – location of restaurant whether it is urban, rural, etc.
    j. Social_label – Social group of people in the area.
    k. Lstage_label – Life stage of people in the area.
    l. Ppop_09q_label – Ranking of population between 0 – 9 years old.
    m. Pqrowthq_label – Ranking of population growth.

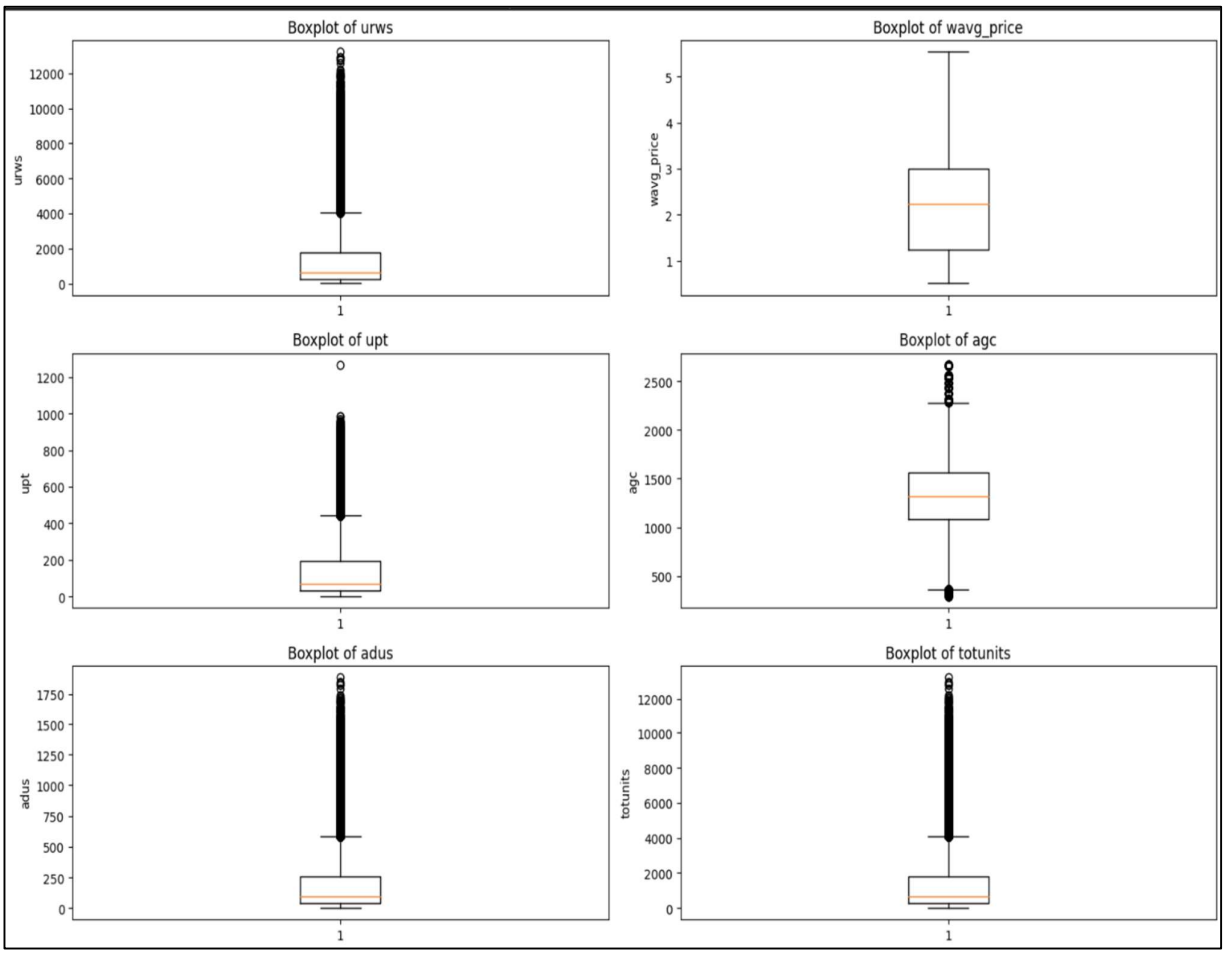
2. Weekly Sales data
    a. REST_KEY – unique identifier for each restaurant.
    b. itemN – Menu item number.
    c. Itemdesc – Menu item name.
    d. wk_ending – Date of week end.

e. Urws – total units sold for the week.
f. wavg_price – Weighted average price of all menu items.
g. Agc – Average transactions per day for a given week.
h. Adus – Average daily units sold.
i. Totunits – Total units sold for the week.

DATA QUALITY CONCERNS

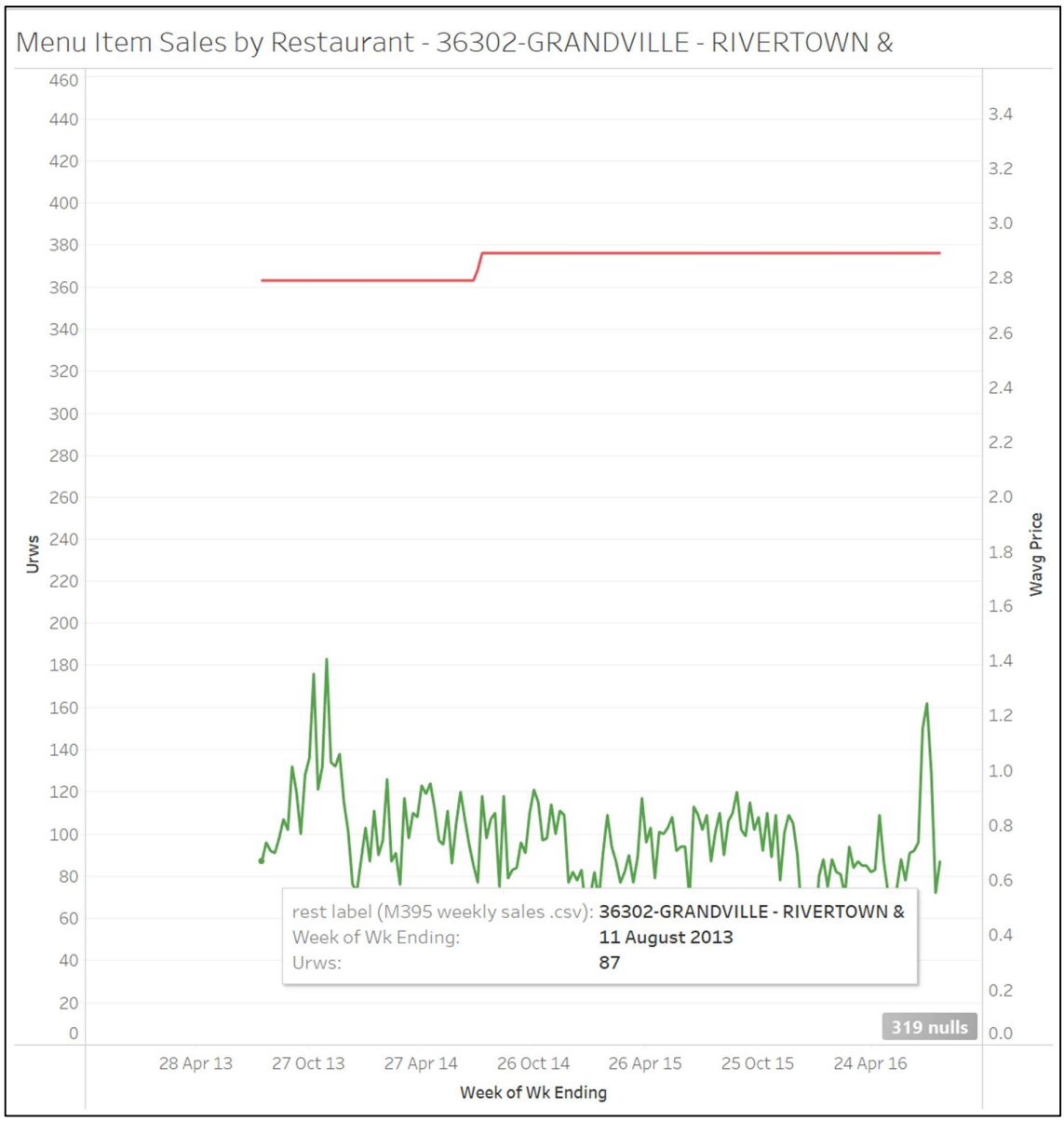
There are multiple quality issues with the data we have been provided by the organisation. Multiple variables have missing values, high number of outliers in many variables, description of variables not given, etc.

1. From the box plots of variables urws, agc, upt, adus, and totunits variables, we can see that there are multiple outliers present in the data. The outliers are the datapoints beyond the whiskers of the given box plots. The whiskers represent the 5th and 95th percentile of the data.



2. There appear to be multiple missing data values for the weekly sales data. The reasons for the missing data could be no restaurant data (restaurant has not opened yet), restaurant closures, or the data may just be missing however we need to deal with the missing data values.

E.g)



We can see that data for the above restaurant begins from 11th August 2013 and it is missing all data points for agc, urws, upt, adus, and totunits before 11th August 2013.

Total Missing/NA Data points:

```
Missing Values in Weekly Sales Data:
REST_KEY          0
rest_label        0
City              0
County            0
latitude          0
longitude         0
owner_label       0
trad_label        0
ItemN             0
itemdesc          0
wk_ending         0
urws           4762
wavg_price     4762
upt            4762
agc            4762
adus           4762
totunits       4762
dtype: int64

Missing Values in Restaurant Facts Data:
REST_KEY                0
rest_label              0
Address                 0
City                    0
Zip                     0
State                   0
County                  0
REST_HISP_CONS_MKT      0
REST_AFR_AMR_CONS_MKT   0
```

```
REST_HISP_CONS_MKT       0
REST_AFR_AMR_CONS_MKT    0
REST_ASIAN_CONS_MKT      0
latitude                 0
longitude                0
REST_PLYPL_TYP           5
REST_DRV_THRU_TYP        0
REST_TYPE                0
coop_label               0
region_label             0
ethnic_label             0
owner_label              0
trad_label               0
subtype_label            0
incomeq_label            2
urban_label              2
social_label             2
lstage_label             2
ppop_09q_label           2
pgrowthq_label           2
dtype: int64
```

3. The urws and totunits variable essentially define the same things within the data i.e, Total weekly unit sales. This makes one of the variables redundant and unnecessary for calculations.

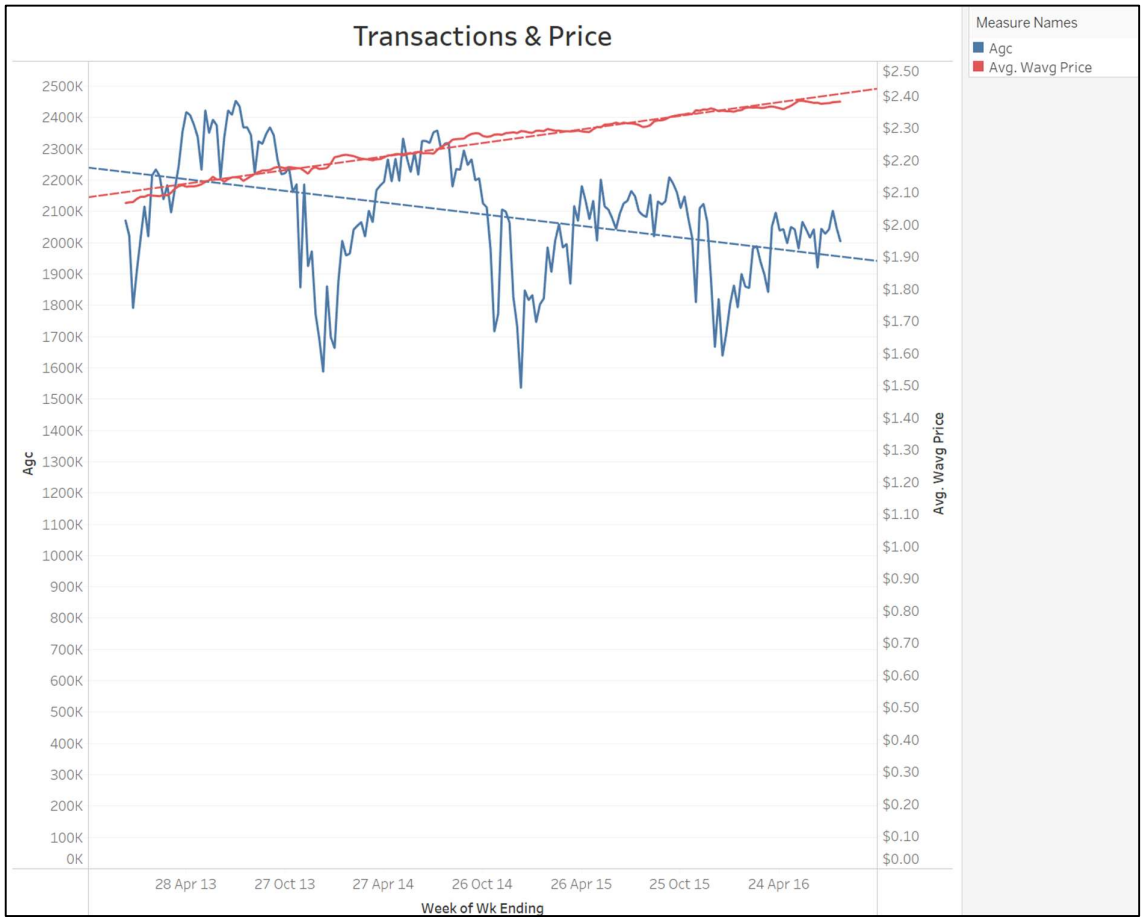
4. The ethnic_label variable in restaurant facts has only two different types of data (HCM, GCM) with HCM only attributed to one restaurant. We can safely discard this variable from calculations as there is not enough data to analyze the variable.

HYPOTHESES

1. The customer traffic at McDonald's has been decreasing year over year.
2. The All-Day breakfast promotions have had a positive impact on sales and performance at McDonald's.

CAN THE DATA CONFIRM OR REJECT OUR HYPOTHESES

1. To confirm the hypothesis that McDonald's traffic has been decreasing year over year, we must analyze the average transactions per day over the past three years.

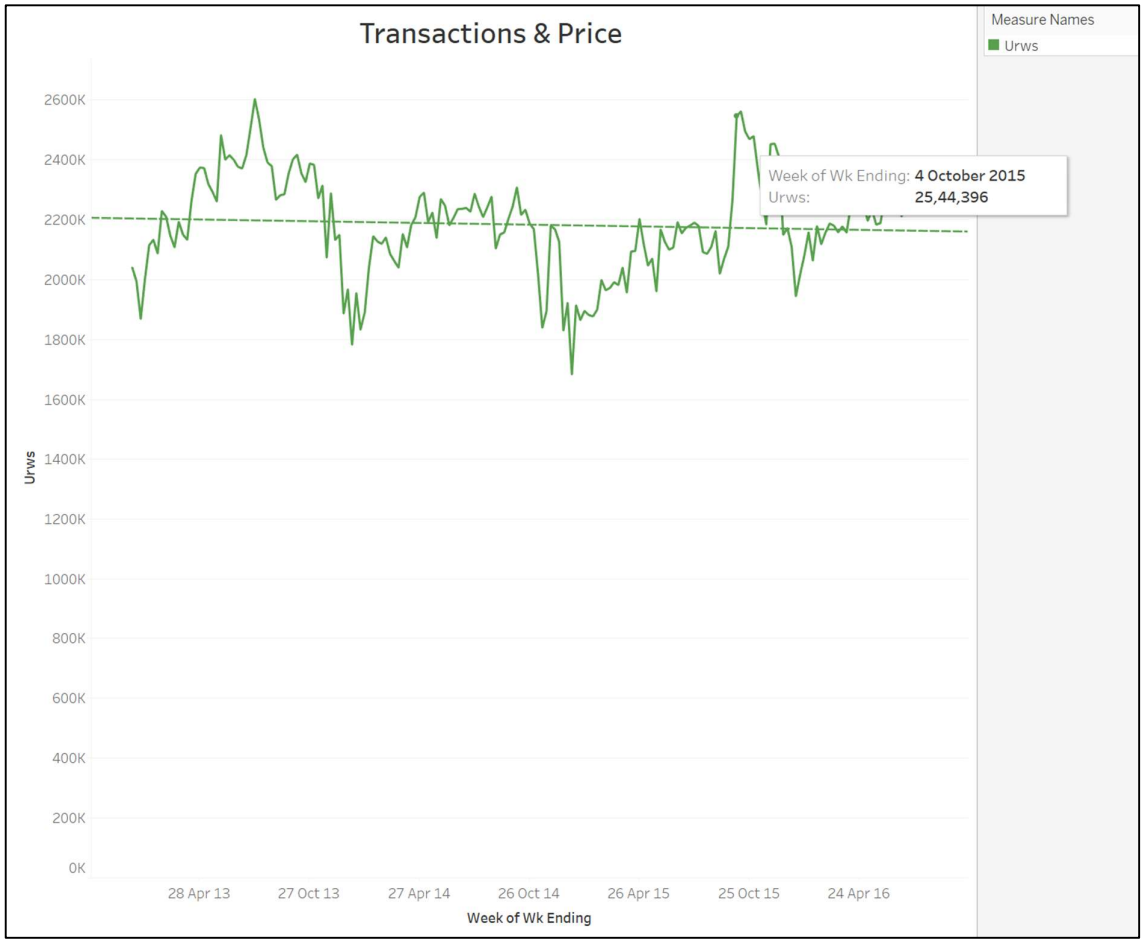


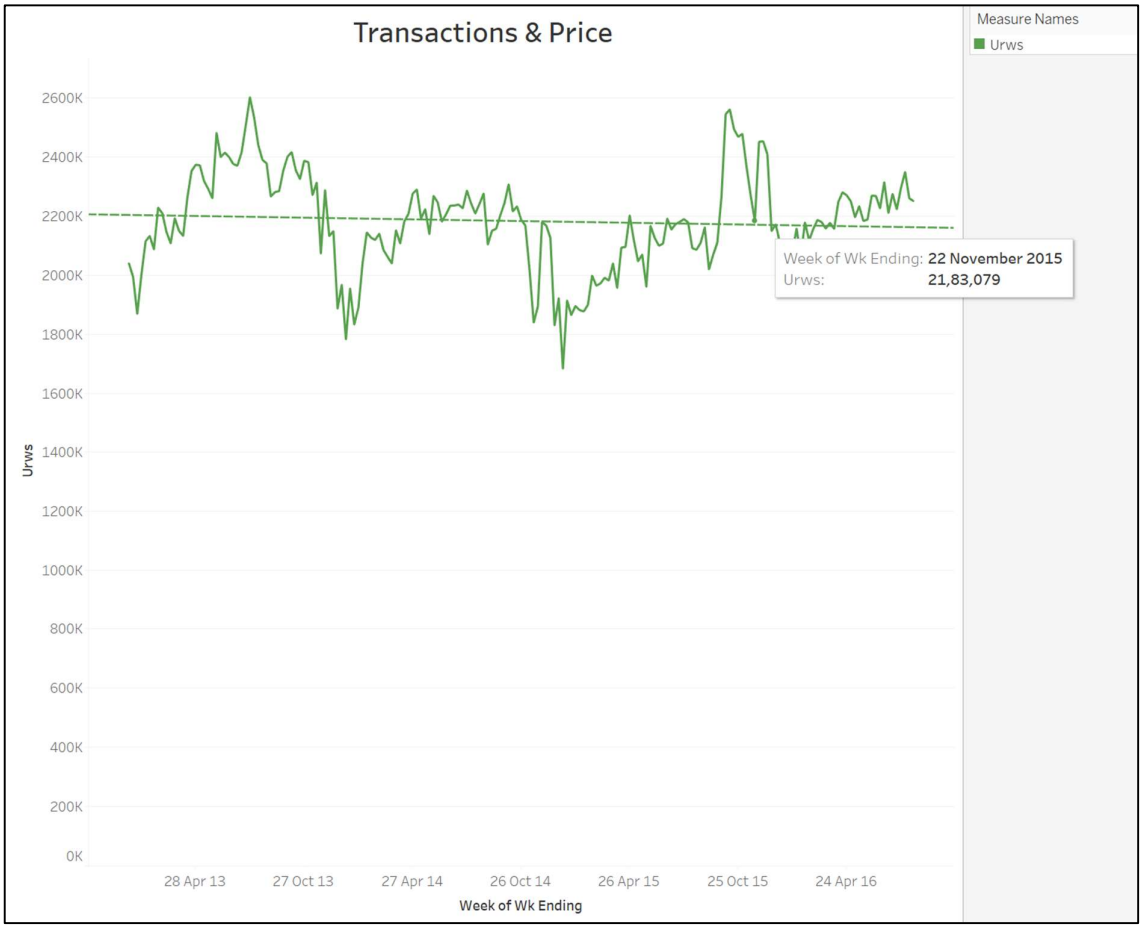
From the above graph, we can see that the Average Daily transactions over a given week (AGC - blue) has been in a downward trend over the past

three years while the average weighted cost of all menu items (AVG WAVG PRICE - red) has been increasing.

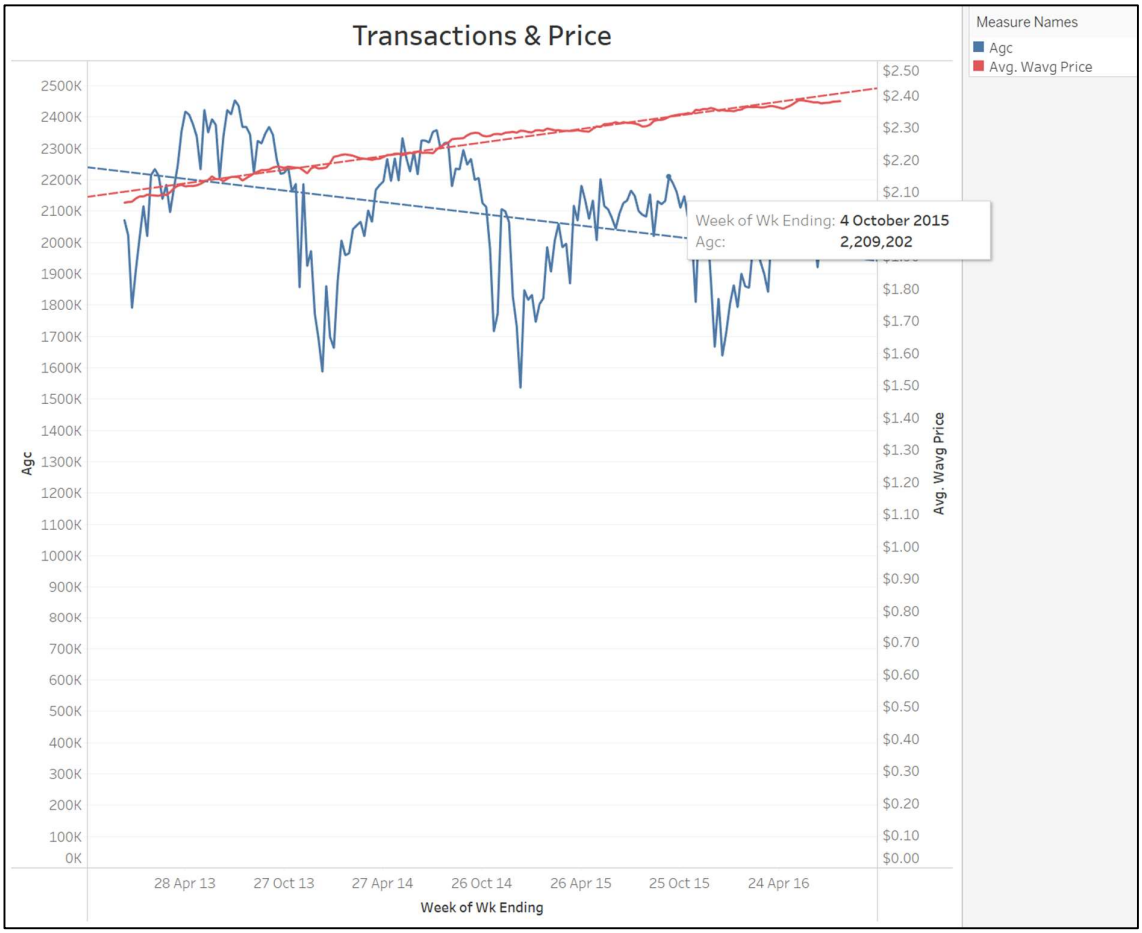
On performing regression analysis of the above graph and calculating the p-value, we find that the p-value is less than the threshold of 0.05, thus confirming our hypothesis that customer traffic has been decreasing year over year.

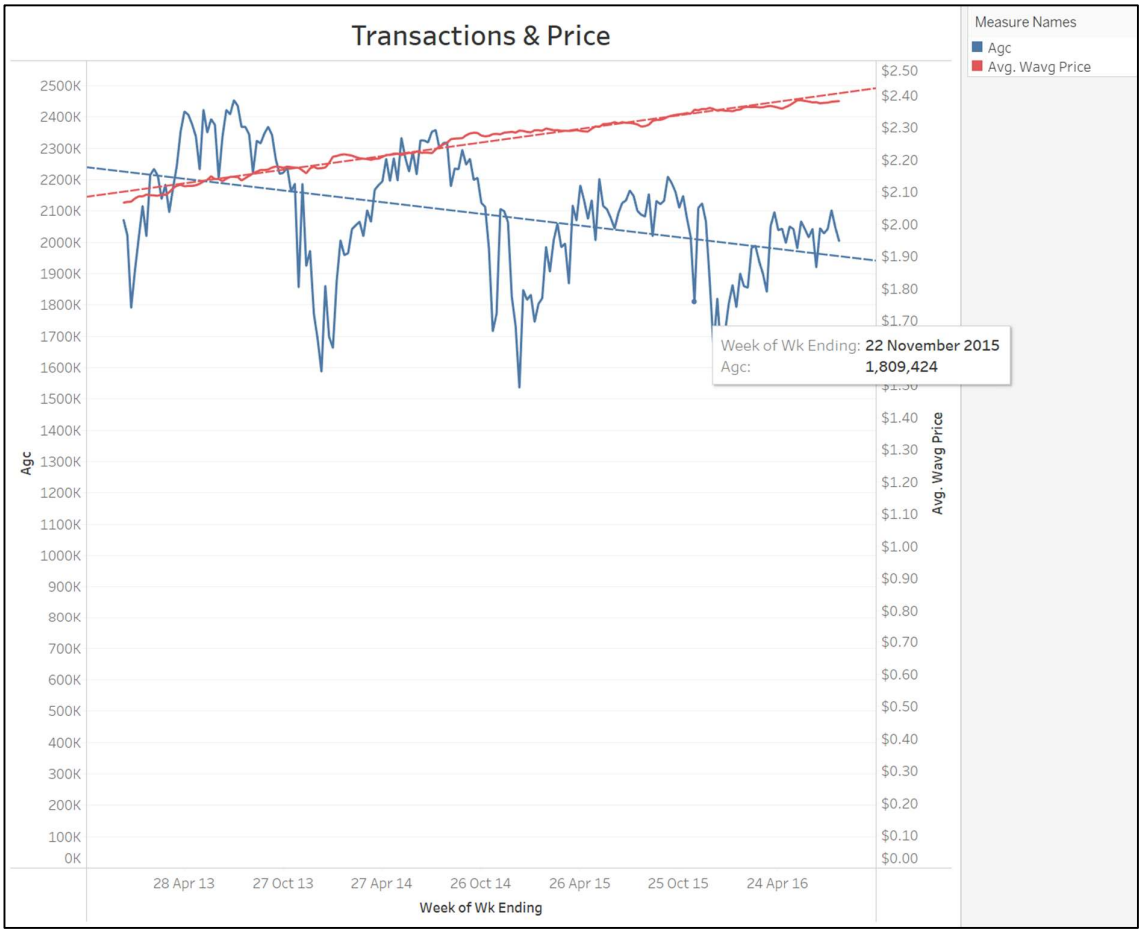
2. To confirm the hypothesis that the All-Day breakfast promotions have had a positive impact on sales and performance at McDonald's we must analyze the total units sold per week after the promotion was launched and also the average transactions per day.





We can see from the above graphs that there has not been a significant impact on the weekly units sold after the All-Day Breakfast promotion was launched. In Fact, the total weekly units sold have seemed to decrease.


Transactions & Price
Measure Names
Agc
Avg. Wavg Price
2500K
2400K
2300K
2200K
2100K
2000K
1900K
1800K
1700K
1600K
1500K
1400K
1300K
1200K
1100K
1000K
900K
800K
700K
600K
500K
400K
300K
200K
100K
0K
Agc
$2.50
$2.40
$2.30
$2.20
$2.10
$1.80
$1.70
$1.60
$1.50
$1.40
$1.30
$1.20
$1.10
$1.00
$0.90
$0.80
$0.70
$0.60
$0.50
$0.40
$0.30
$0.20
$0.10
$0.00
Avg. Wavg Price
Week of Wk Ending: 4 October 2015
Agc: 2,209,202
28 Apr 13
27 Oct 13
27 Apr 14
26 Oct 14
26 Apr 15
25 Oct 15
24 Apr 16
Week of Wk Ending

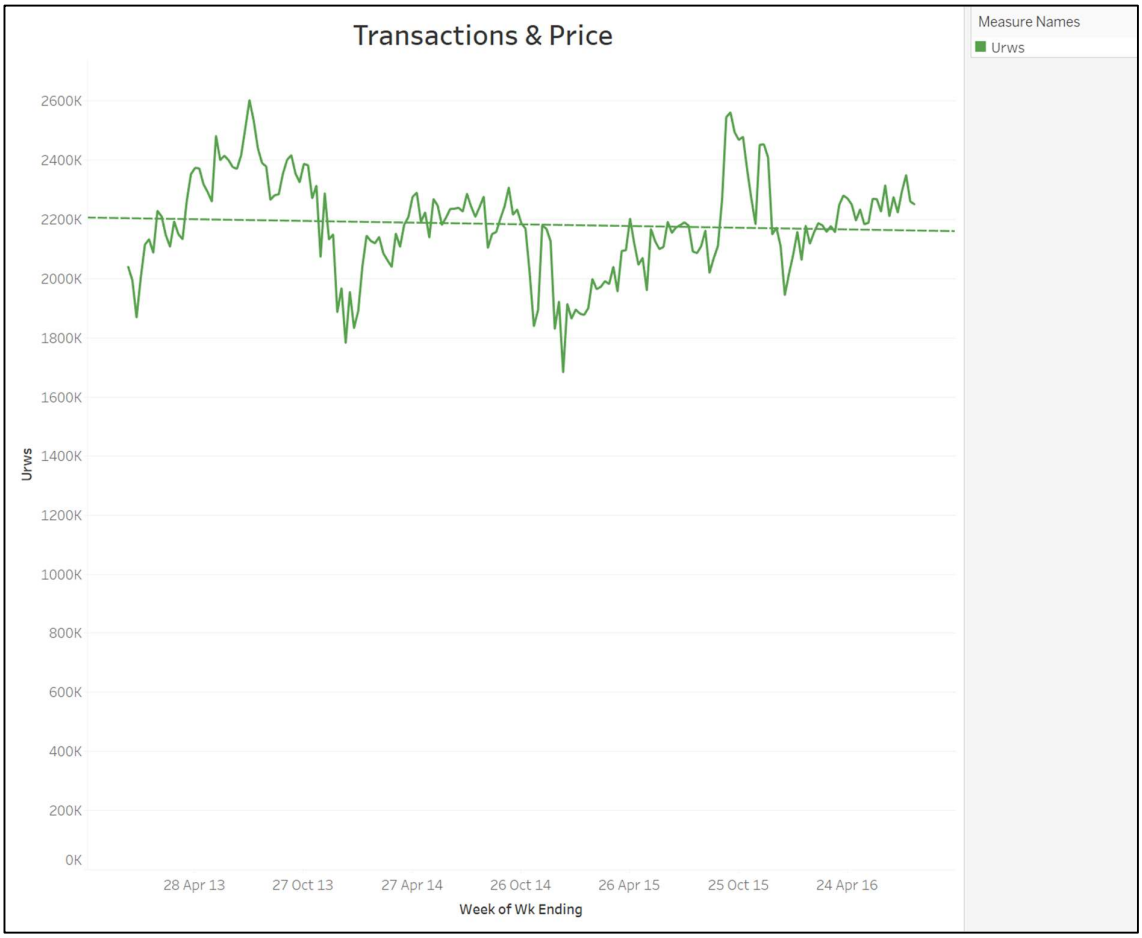


We have a similar case with the average transactions per day. There has not been a significant impact on the average transactions per day after the promotion was launched. Similar to the total units sold, the average transactions per day have decreased.

On Calculating the t-statistic for the above data, we can see that it has a negative value, thus rejecting the hypothesis that the All-Day Breakfast promotion had a positive impact on sales and performance at McDonald's.

SUMMARY

From our analysis of the given data, we can see that customer traffic has been decreasing at McDonald's year over year, however this is not a complete indicator of restaurant performance. While the average transactions have gone down, the weekly units sold remains linear.



We can see from the above graph that the total units sold per week seems to remain linear even with the decreasing customer traffic.

Another indicator which would help us determine performance would be Average Purchase Price of a Transaction at McDonald's. This would give us an idea about the sales.

Regarding the All-Day Breakfast, from the data provided to us we have deduced that the all-day breakfast had no significant impact on sales or performance at McDonald's. We are also missing some crucial data which could help us make a more educated decision such as Resource utilization data (Running costs of breakfast machines, labour costs, raw material cost of breakfast items, etc) and Average price of a transaction. With these extra variables, we can not only determine the effectiveness of promotions but also the performance of McDonald's over the past three years.