**Thank you very much for your time ahead.**

**First,** the SQL part is the Tables\_list.xlsx and in the query.sql file in the folder.

The Table\_list file gives the table on what the attached SQL query (query.txt) is gathering and aggregating, while the query.sql has my comments in “comment” labeled close to where I believe it could be improved for the sql performance.

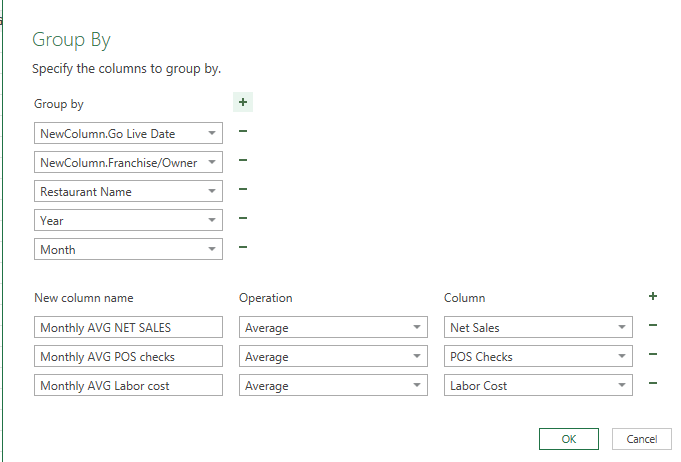
**Second**, the data analysis part is as below.

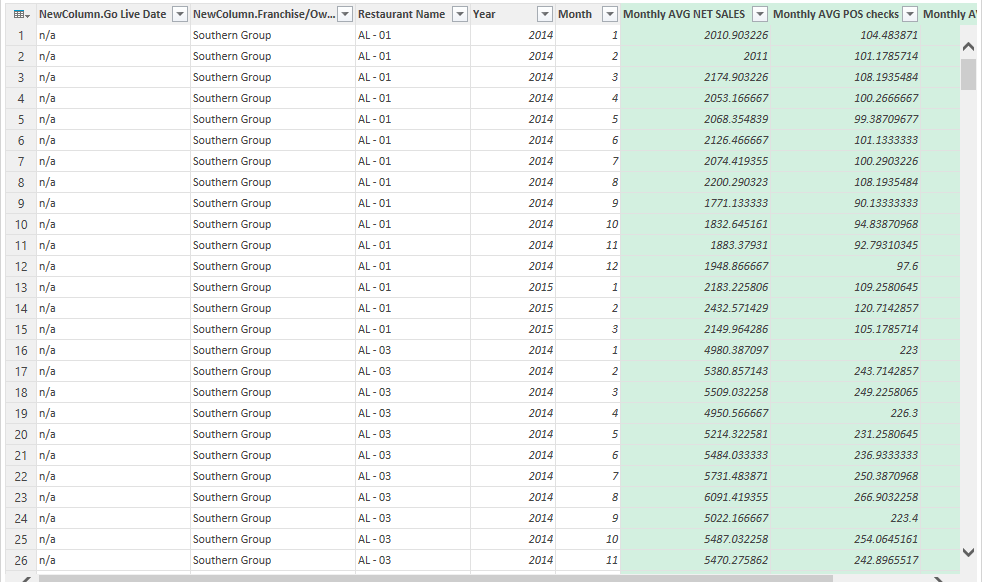
So please install power query if needing to view my querying step if you don’t have it.

Step 1: observe the key data trending pattern on the control group (no device in use)

Using power query in Excel: I applied “restaurant\_data left outer join restaurant\_info”.

Below shows how I group up the data and import it to the new tab

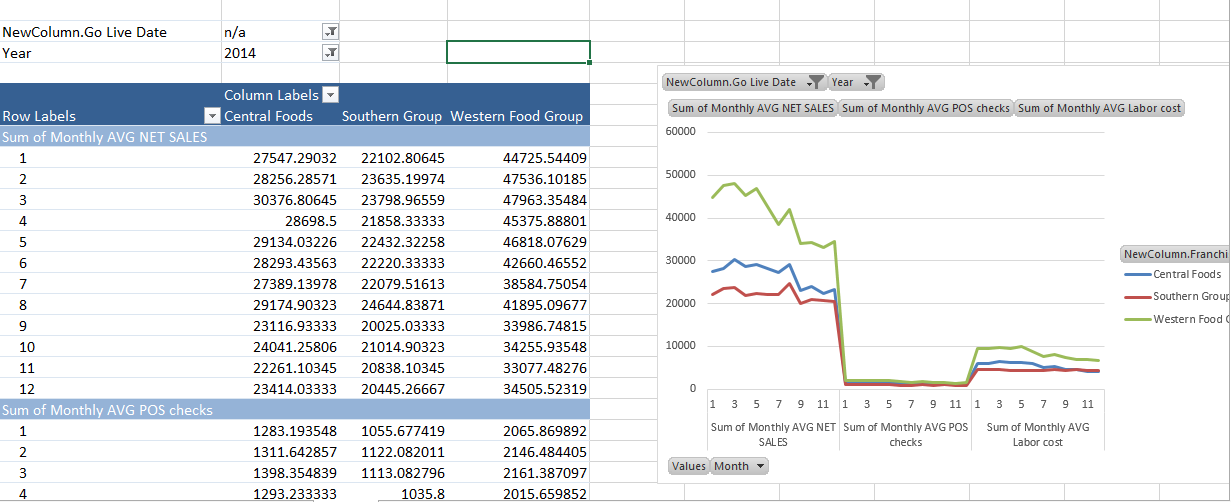




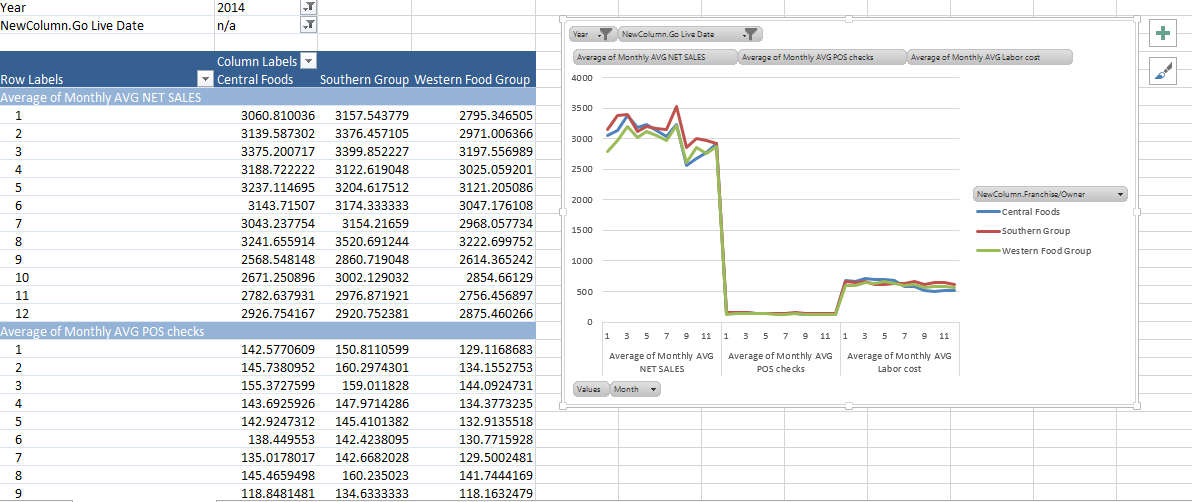
Step 2:

Build a pivot table and shows 2 graphs as below:

The total Net Sales for each month in 2014



The average Net sales for each month in 2014



In the above graphs, we observe that the first 8-months is the best time of the restaurant business as all the data consistently points out. Yet, the total numbers of POS\_check per month has not changed much from month to month during the year. Plus, the labor costs is not too sensitive to sales in each period though it moves along with the restaurants’ business.

Moreover, by comparison, the average Net Sales each month in the Sothern Group shows a higher position in graph whereas the total Net sales each month in the western food Group holds the highest ground.

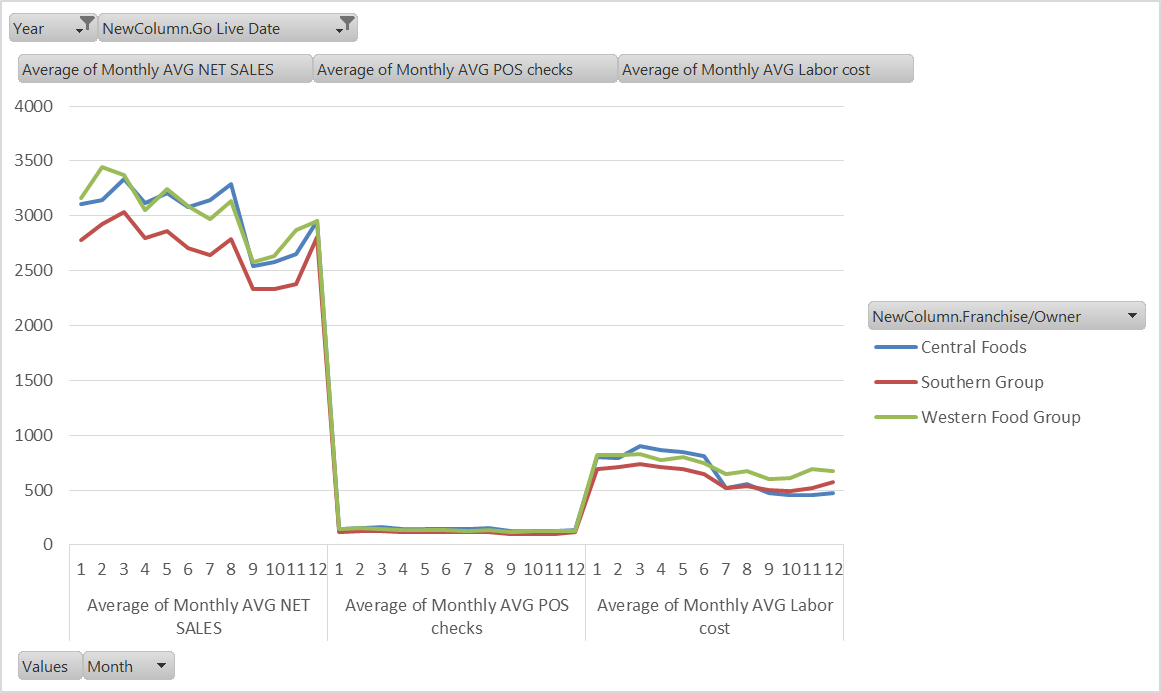
Finally, it is interesting to find that Western food group’s total net sales varied a lot whereas the Southern Group’s average net sales stayed high as time went on in 2014 in the control group.

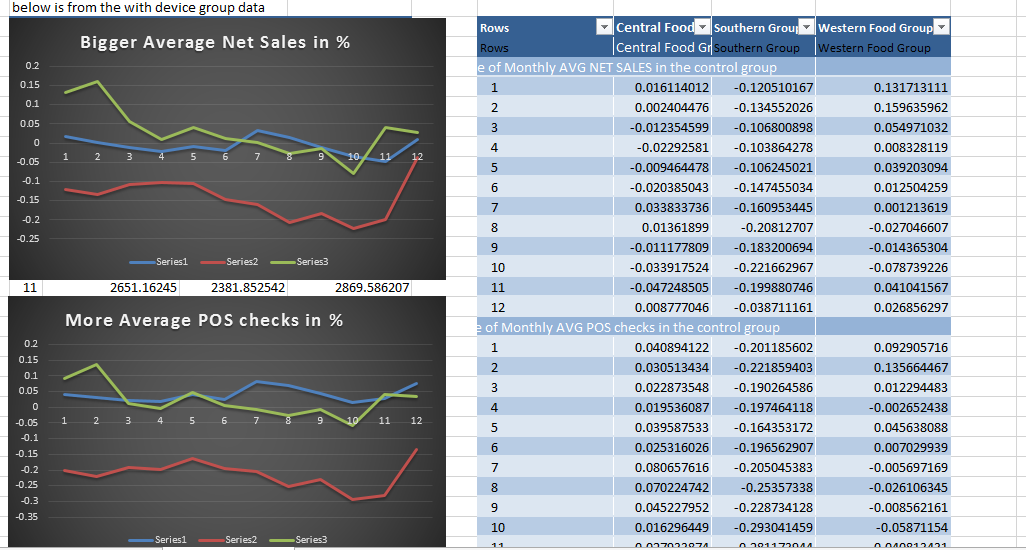
Yet, to compare performance between the control group and the experimental group, it is better to use average per month, since some restaurants may not operate in certain day or lost their data during the month.

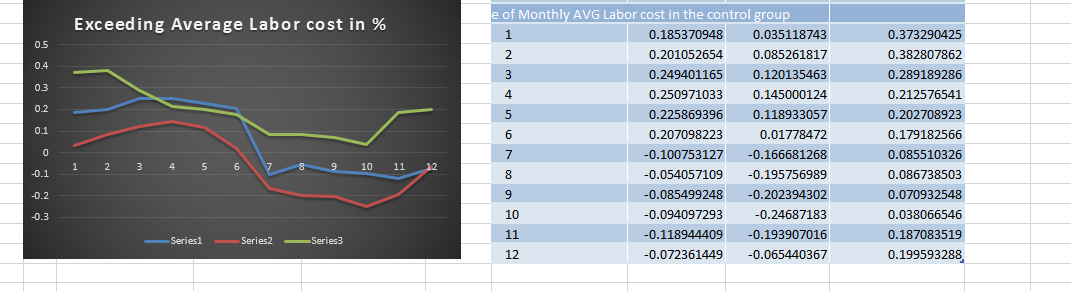
The impact of the introduction of the Presto device vs the control group:

the control group:

the with-presto device group:







The above comparisons show that in terms of the increase in the average figures above, the Southern Food Group has not been doing better though using Presto device while the Western Food group was selling at least by 5% better than the control group. Interestingly, in the Central Group, it seems the restaurants with the presto device was almost like the control group during the year, yet apparently having at least 2% more Pos\_checks than the control group during the year.

Also, very interestingly, within the first 3 months, the Western Food group has witnessed a dramatic better performance both in Net Sales and Pos\_check counts.

Meanwhile, as for the labor costs, both Southern Group and the Central Group got a negative cost comparison since around mid-June, though the average labor costs for Western Food Group stayed higher than the control groups.

**Part 2**

The predictive model for Game Revenue:

**If you want to view my predictive model straight ahead, please go to the final page.**

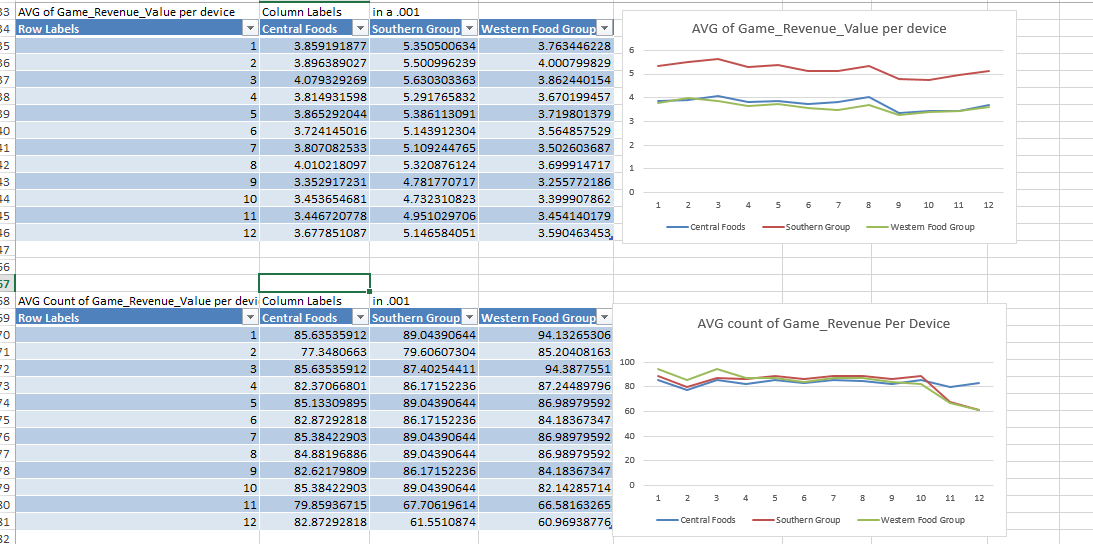
The below step is how I try to observe the data and have a better common sense on what happened there.

Step 1:

**How about each presto device’s revenue amount and service count in each region?**

Below is the graphs show how they perform individually:

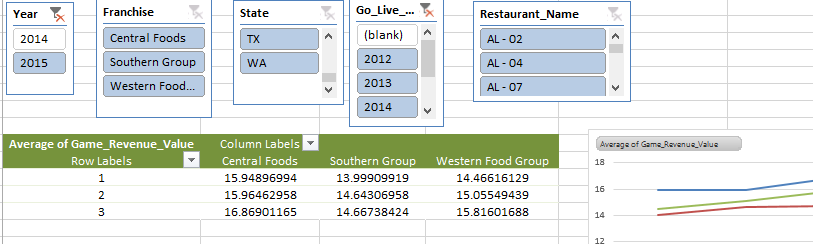
The revenue on each device in its own region is based on .001 in value

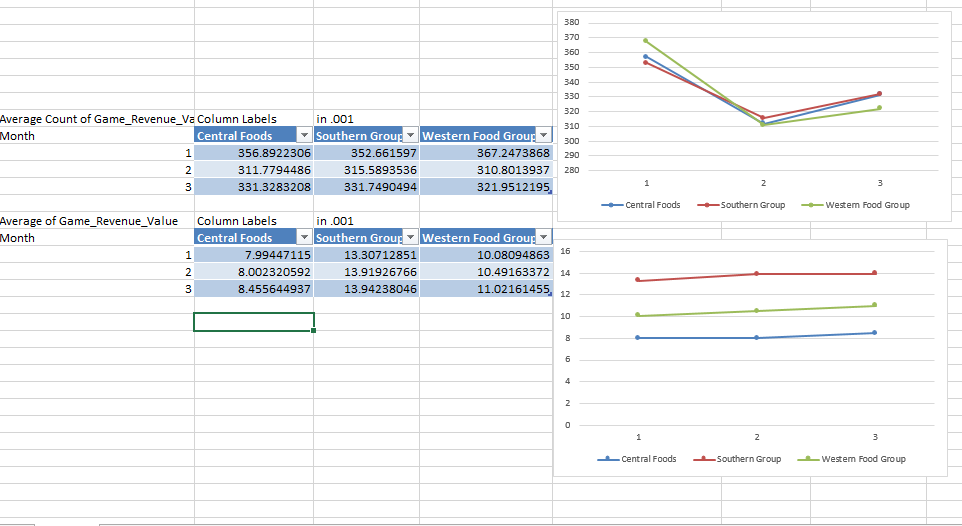


Obviously, the game revenue in return per device has the best result in Southern Group and its business is indeed in need of more presto devices to receive higher overall return. Also, the AVG count of Game Revenue per device each day also shows a declines in Southern Group and Western Food Group. Yet, under this measure, we observe very little differences in each group as for the frequency of people using the device to play games. Also, the variation in the Western food Group in the counts is obviously the largest in a glance.

This unit measure is very sensitive to the growth and the potential of the game’s revenue

Below shows the average count of Game revenue and the average revenue of the Game Revenue in each device in its region,





Here, upon 1st quarter in 2015, we observe a dramatic increase in the average Game revenue and the average count in each device in each region by at least 2 folds compared with that of 2014 after introducing more presto device to the restaurants in 2014.

Thus, we find the “year”, ”Go\_Live\_Year” and “the presto\_device\_install” are very important to our analysis since it affects how we valuate each region’s performance as reflected in the above or in the “Observation\_table” in the Restaurant\_data.xlsx file.

**Step 2:**

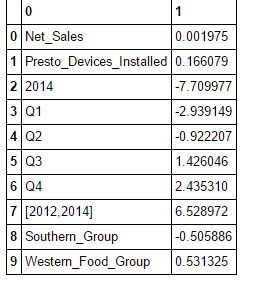
I use ipython notebook for this data model construction.

My predictive model could be viewed at “analysis.ipynb” and “analysis2.ipynb” files, though I decide to use the model call “m6” in “analysis.ipynb” as my final model for this project, since the model incorporates most of the data information and did not over-fit the data (with a high multicollinearity variance inflation) or under-fit as it is what happened in the “analysis” model where is revealed by the scatter plot (the pattern was too obvious).

I use Linear-regression since it is a continuous data set.

My model’s coefficient chart is as below:

X | coefficients



Y is Game\_Revenue

(There is a typo there, [2012,2014] should have been [2012,2014), since it does not include 2014, I am sorry)

The above means the game\_revenue’s future increase was highly leveraged by the investment between 2012 and 2014, and it suggests more revenue growth happening in Quarter 3 and Quarter 4, whereas Presto\_Devices\_installed introduction to the market is very important. Also, the revenue to the Western\_Food\_Group is also vital but not for Southern\_Group(too saturated maybe?).

Thank you very much for your time and patience at reading this.