USING ANALYSIS OF VARIANCE(ANOVA) METHOD IN THE IDENTIFICATION OF UNEQUAL MEANS RESULTING FROM SALES INTERACTIONS

KESMEEN TAN; MATTHEW YEE; SIM JING XIANG, Singapore Management University

ABSTRACT

The ability to assess the actual impact of sales representatives' interactions with customers on revenue is pivotal to the success of any pharmaceutical company. Yet current efforts on the company's part fail to take into consideration many factors for it to provide meaningful assessment, which could bring about many detrimental effects - an inaccurate allocation of manpower, or even badly skewed performance reviews. This paper segments the business into different factor groups while performing Analysis of Variance(ANOVA) method to help understand the actual impact of sales interactions on these groups. Factor groups include various sales channels, therapy groups and even financial quarters. Results show that factor groups do indeed have differentiating patterns in sales interaction and sales revenue, making the understanding of their patterns imperative for success. Through our analysis, we have identified groups that benefit from higher interactions, namely neighbourhood clinics and most therapy groups. These groups usually benefit from high level of interactions (≥ 5), with Middle level interactions (<5, ≥ 2) showing no conclusive evidence of improving means.

1 INTRODUCTION

Our sponsor, a global pharmaceutical company, relies heavily on its sales representatives to generate sales for new drugs and maintaining sales relationships with existing customers. Customers include that of government hospitals, private hospitals, polyclinics, to name a few. On a quarterly basis, our sponsor allocates their therapy teams a target number of sales interactions to be made with personnel of these institutions. As our sponsor believes there is a need for full understanding of products before one can make any meaningful headway in sales, it structures its sales teams into different therapy groups, whereby each therapy group focuses on a group of drugs that treat a genre of illnesses (e.g. respiratory diseases, allergies) and people in sales teams are only able to engage in sales related to their allocated therapy group.

The sponsor has long wished to understand the impact of sales interactions on the ability to drive sales. Yet for us to give an accurate picture, there is a need to analyze the different factor groups separately. This paper will focus on identifying these potential factor groups and using Analysis of Variance(ANOVA) individually on these potential factor groups to understand if sales interactions do have an impact on sales revenue for those groups, essentially identifying sections in which the company should place emphasis on manpower interactions and areas where slack could exist for an efficient allocation of resources.

This paper consists seven sections. The first is the introduction. Following that, the Literature Review section discusses the basic method and principles of the Analysis of Variance(ANOVA) technique. Thereafter, the Methodology section will talk about the data used in our analysis, along with the tools used and steps of data preparation. It will also briefly cover our initial data exploration and how we will be using the Analysis of Variance(ANOVA) technique. The Analysis section goes in depth into explaining our results. After that, the Managerial Recommendation section brings our sponsor into context and we will explain what we believe the sponsor should be doing with our results. The Conclusion section then sums up our findings. Lastly, the limitation section brings discussion to limitations in our studies and possible ways to improve it.

2 LITERATURE REVIEW

Analysis of Variance (ANOVA) is a common statistical method used to determine any significant differences between three or more independent groups, usually for one or more independent variables such as dose, time, or age (Lazic, 2008). There are many ways to compare means between groups, but it has been advised that a comparison of means across more than two groups require more accurate methods than just using simple repeating of t-tests (Pandit, 2010). In a typical scenario of experimental biology, animals are categorized into groups that receive a different dose of drug, and tested on response variables such as cognitive performance to determine varying effects (Lazic, 2008). The result, which depicts the outcome (p-value), shows whether there is enough statistical significance (p-value < 0.05) to reject the null hypothesis that different dosage groups have the same performances.

To take a broader perspective on how ANOVA was used in different industries in different ways, this review of literature will involve studying various research papers to understand significant contributions made by researchers and determine any gap that can be complemented by the present study.

In the retail sector, studies have been conducted to determine how shopping trip type can lead to differences in actual, unplanned spending and purchases of promoted items when subjected to in-store promotions or promotional activities (Anic and Radas, 2006). The results which supported that the type of shipping trip is strongly related to purchases made on promotion and different levels of unplanned purchases, enable managers to strategize appropriate promotional and initiatives to maximize purchasing outcomes (Anic and Radas, 2006).

In another paper, Aluregowda (2013) posited that selected strategies from big bazaar retailers such as locating outlets in prime areas and popularizing private-label items can influence the growth of business. Analysis of variance conducted reveals significant differences upon introduction of various strategies and concludes the need for management to carry them out to effectively improve growth (Aluregowda, 2013). Similar studies in the toilet soap industry has been conducted, where Naik (2014) proposed the relationship of different marketing

strategies on customers' interest, so as to curtail the cost of certain ineffective strategies and thus, increasing competitive efficiency.

Though ANOVA has been used widely in marketing strategies to identify and address different target groups, it is not popular in the pharmaceutical industry. Therefore, through our present paper, we hope to analyze the causality relationship of number of interactions on sales revenue in the pharmaceutical industry, characterized by various factors. To discern these factors that play a vital role our analysis, a primary exploratory data analysis will have to be conducted – rationalized by the appreciation that different factors vary in their nature: some factors, like ordered factors, causes an innate variation in results across groups while other factors like unordered factors require a deeper investigation to predict the levels of effect it has on the data. (Singer, 1991).

3 METHODOLOGY

3.1 The Data

With the understanding of our sponsor's motivations, our team sets in motion the data wrangling process – one that encompasses data cleaning, transformation and integration to obtain a consolidated JMP data table used for further analysis.

For this research study, we have obtained a year's worth of data from 2016. It consists of information on invoices, call details, employees and customers – each of which described in the summary table below

File Name	Description	No. of Rows
Call Details	Information on actual interaction between Sales Reps, Sale Targets for a Product Brand	42915
Invoice Details	Information on transactions of product purchases	110372
Employee	Information on employees and their teams - "Therapy Area"	237
HCP	Information on individual doctors	5871
HCO	Information on clinics, organizations	4425

3.2 TOOLS USED

SAS JMP Pro 13 is chosen as our primary tool for data preparation, exploratory and further analysis. It is an analytical software that can perform most statistical analysis on large datasets and generate results with interactive visualizations used by data scientists to manipulate data selection on the go. Furthermore, tutorials and guides are widely available online for us to learn JMP Pro 13's different techniques and functions.

More importantly, its easy-to-use built-in tools enable us to conduct analysis of variance to determine relationship between interactions and sales revenue.

3.3 DATA PREPARATION

Data preparation took us through time-consuming and tedious procedures to obtain high quality data. Though seemingly unrewarding, a set of high quality data allows for more accurate, reliable and consistent analysis of results. Therefore, it is imperative to invest a lot of time and effort on it to avoid getting false conclusions for our hypothesis.

Firstly, using SAS JMP Pro, these tables are scanned for anomalies such as missing values or outliers. We corrected them appropriately, using imputation for missing values and omission for extreme outliers. This step ensures that the data we are using will not give us misleading insights.

Next, to determine the causality relationship between interactions and sales revenue, we need to join Call Details and Invoice Details tables. This leads us to understand variables from both tables to find out 1) which two variables are the same and whether their formats are alike, 2) at which granularity does each row from both tables represents. Upon fully understanding them, we performed aggregations, standardizations of formats and inclusions of HCP and HCO tables to serve as links. Furthermore, we added the dimensionality of employees' teams from Employee table, as it will be useful in describing the relationship as mentioned. Finally, we integrated relevant tables into a consolidated table and loaded it into the JMP server for analysis.
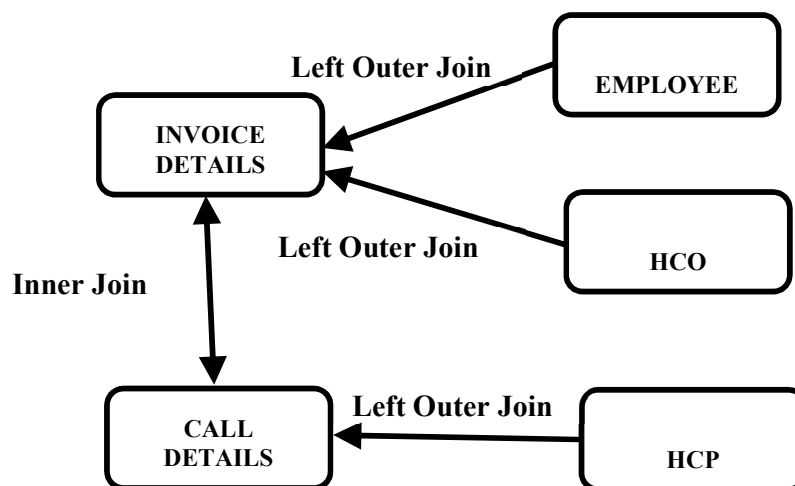

Fig 1: How different data files are joined

The above diagram illustrates how data tables are being integrated, which took a few stages to achieve.

The first stage of data preparation involves cleaning **Invoice Details** and **Call Details** tables.

1. For missing values under "price$" column in **Invoice Details**, we imputed their values to "0" as these rows contain records where free samples given out to customers.

	Sales Qty	Bonus Qty	Sample Qty	Compliance Qty	Amount$	Price$
)	0.00	0.00	12.00	0.00	0.00	•
7	0.00	1.00	0.00	0.00	0.00	•
7	0.00	1.00	0.00	0.00	0.00	•
5	0.00	2.00	0.00	0.00	0.00	•
)	0.00	0.00	5.00	0.00	0.00	•
)	0.00	30.00	0.00	0.00	0.00	•
)	0.00	0.00	40.00	0.00	0.00	•
)	0.00	0.00	6.00	0.00	0.00	•

Fig 2: Some rows where price$ has missing values

Steps Taken:

   i. Press Ctrl-F to show "search data tables" interface
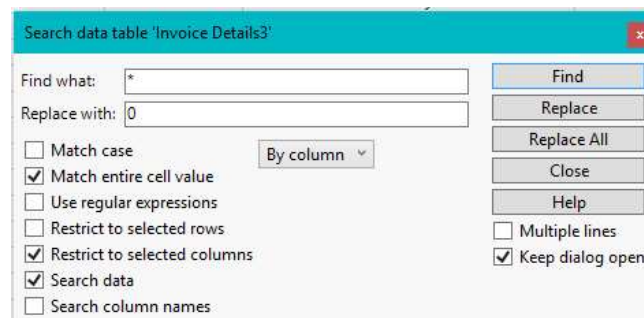
Fig 3: Search and replace function

   ii. Enter the following fields:

      1. Find what: *

      2. Replace with 0

      3. Tick "Match entire cell value"

      4. Tick "Restrict to selected column" of "Price$"

      5. Tick "Search data"

   iii. Click on "Replace All" to apply change

2. For negative "sales qty" and "amount$" values in **Invoice Details**, we did not take any action as they serve as records to void any sales that has been cancelled. Upon aggregation by quarters, no negative values will be present.

3. For 5-digits postal codes in **Invoice Details**, we created a new column to store the converted 'postal code', with data type changed from numerical to categorical and formula function used to add the missing '0'.

> Steps Taken:

> i. Right click on Postal Code's header -> Select "Column Info" -> Change data type to "Character"

> ii. Right click on Postal Code's header -> Select "Insert Columns"

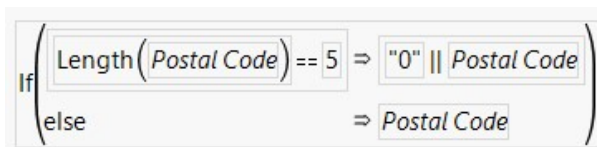> iii. Right click new column -> Select "Formula" -> Enter the formula as shown below

Fig 4: Formula used to reformat postal code

> iv. Click on "OK" to apply change

4. To facilitate understanding quarterly performance of sales revenue in **Invoice Details** and **Call Details**, we added a "year-quarter" column for use during aggregation.

> Steps Taken:

> i. Right click on "Invoice Date" of Invoice Details / "Date" of Call Details

> ii. Select "New Formula Column" -> "Date Time" -> "Year Quarter" to create the new column

5. For "Product Names" differences in **Invoice Details** and **Call Details**, we standardized a common format using the recode function. (No screenshots due to confidentiality clause)

> Steps Taken:

> i. Click on "Product Name" of **Invoice Details** / "Product" of **Call Details**

> ii. Go to "Cols" at the top -> Select "Recode"

> iii. At the "recode" interface, input new values that are the standardized common format

iv. Click on "Done" and "In Place" to replace the old values

The second stage of data preparation involves adding dimensionalities to **Invoice Details** and **Call Details** tables using **HCO**, **HCP** and **Employee** tables.

1. **Invoice Details** and **HCO** are left outer joined to add dimensionality of clinic's "name" needed for further joins, while keeping records in **Invoice Details** intact.

    Steps Taken:

    i. We have identified the matching variables to be "CUSTOMER_CODE" in Invoice Details and "ZP Account" in HCO, both are IDs of clinics

    ii. Go to "Tables" and "Join"

    iii. Enter fields as shown in screenshot below

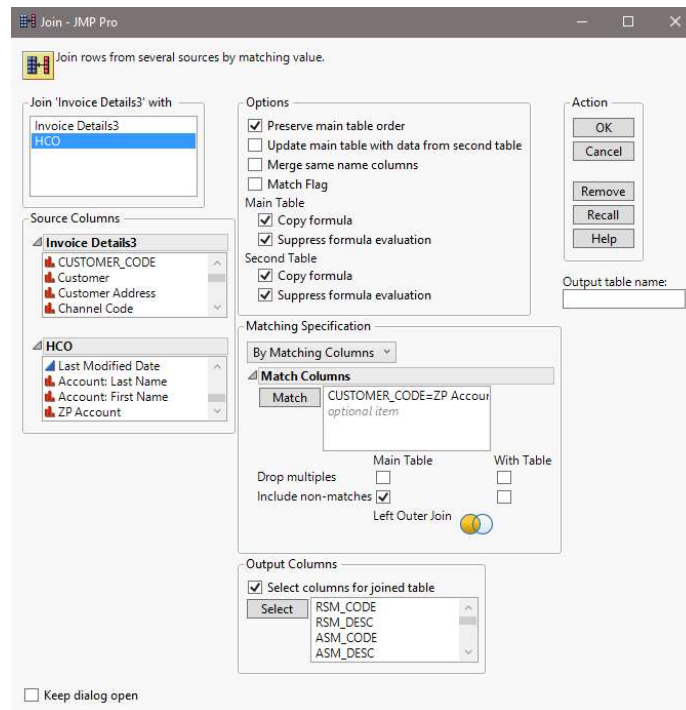*Fig 5: Join function to match Invoice Details and HCO*

    iv. Ensure that Left Outer Join is selected and output columns are populated before clicking "OK" to apply join

2. **Call Details** and **HCP** are left outer joined to add dimensionality of "primary parent" (clinic name of individual doctors) needed for further joins, while keeping records in **Call Details** intact.

Steps Taken:

    i.   We have identified the common variables to be "Account" in Call Details and "Name" in HCP, both are name of doctors

    ii.   Go to "Tables" and "Join"

    iii.   Enter fields as shown in screenshot below
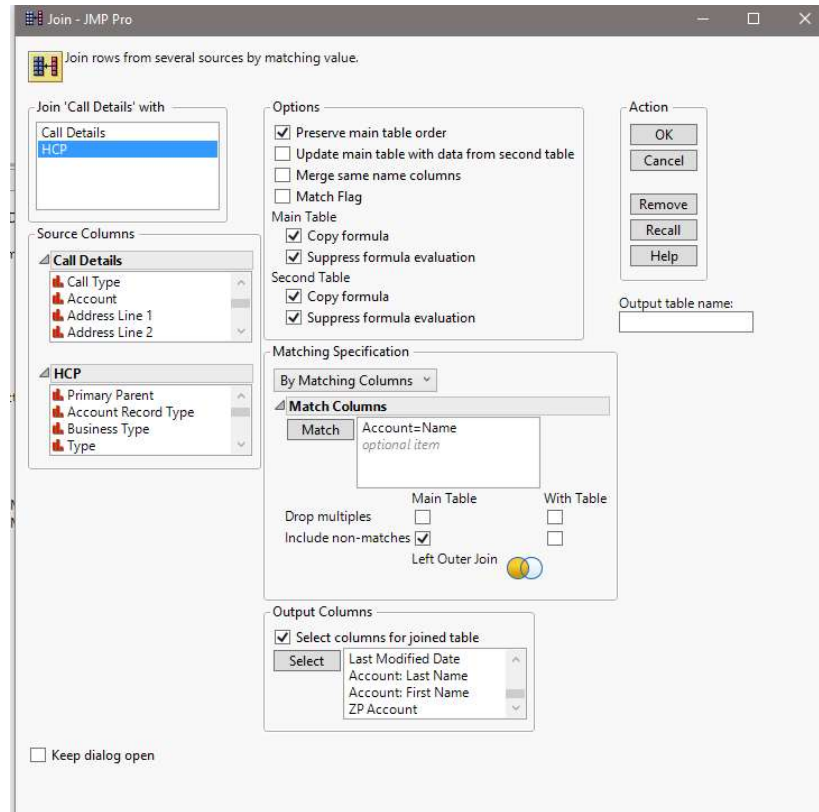
Fig 6: Join function to match Call Details and HCP

    iv.   Ensure that Left Outer Join is selected and output columns are populated before clicking "OK" to apply join

3. **Invoice Details** and **Employee** are left outer joined to add dimensionality of "therapy area" (sales teams) needed for further analysis, while keeping records in **Call Details** intact.

Steps Taken:

    i.   We have identified the matching variables to be "Rep Name" in both Invoice Details and HCP, which are names of sales rep. Additionally,

"year quarter" from both tables are also identified because sales reps may change their "Therapy area" (sales teams) by quarter basis.

ii. We will utilize another function "Update" to perform the same left outer join

iii. Go to "Tables" and "Update"

iv. Enter fields as shown in screenshot below
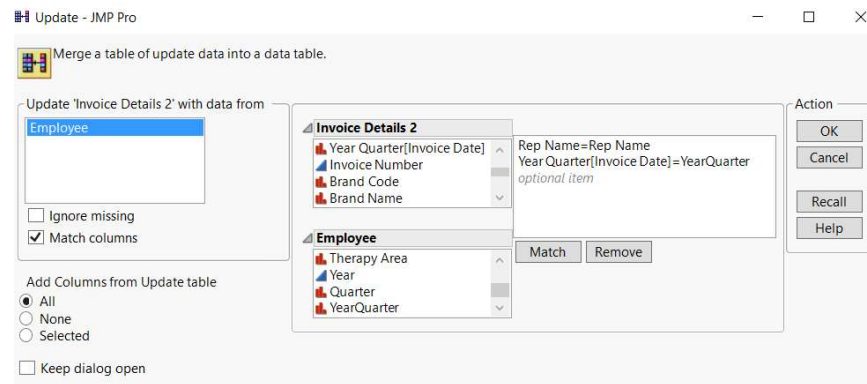
Fig 7: Update function to match Invoice Details and Employee

v. Click "OK" to apply update

The third stage of data preparation involves aggregating **Invoice Details** and **Call Details**

1. **Invoice Details** are aggregated by "year-quarter" to derive additional columns "sum(sales qty)" and "sum(amount$)", addition to variables we are interested at: "channel" (clinic type), "rep name", "product name", "name" (clinic's name) and "therapy area" (sales teams).

   Steps Taken:

   i.   Aggregation will be performed using "Summary" function.

   ii. Go to "Tables" and "Summary"

   iii. Drop fields to Statistics and Group as shown in screenshot below

Fig 8: Summary function to aggregate Invoice Details

   iv. Click "OK" to summarize the table

2. **Call Details** are aggregated by "year-quarter" and "primary parent" to derive additional column of "no. of rows" (interaction count), addition to variables we are interested at: "call: owner name" (sales rep's name) and "product".

      Steps Taken:

        i.   Aggregation will be performed using "Summary" function.

        ii.   Go to "Tables" and "Summary"

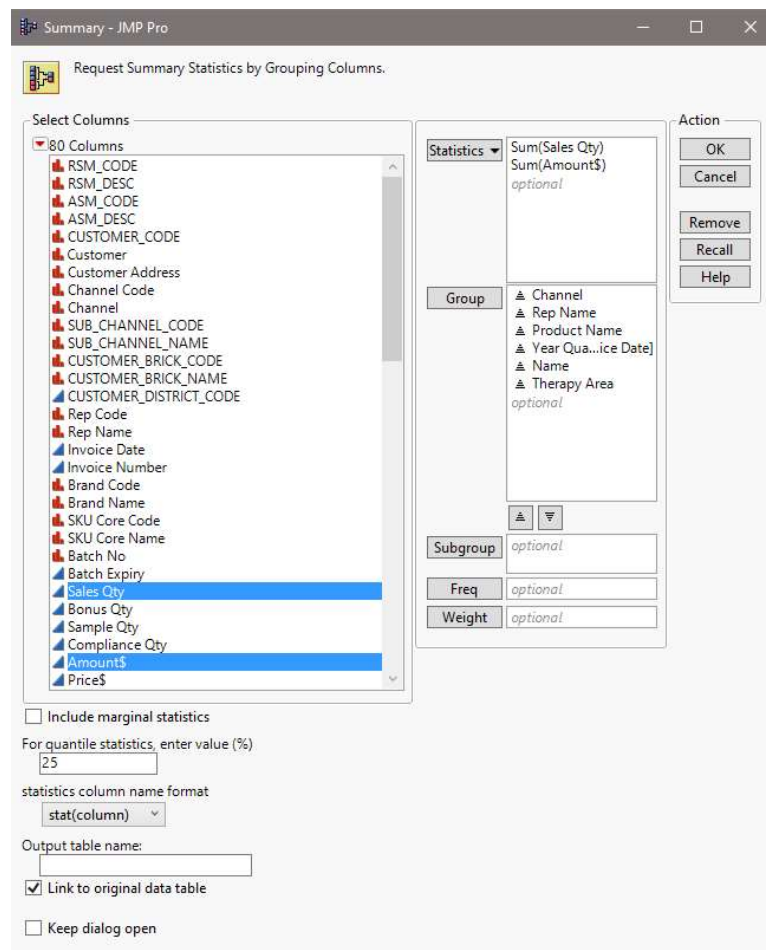   iii.   Drop fields to Statistics and Group as shown in screenshot below
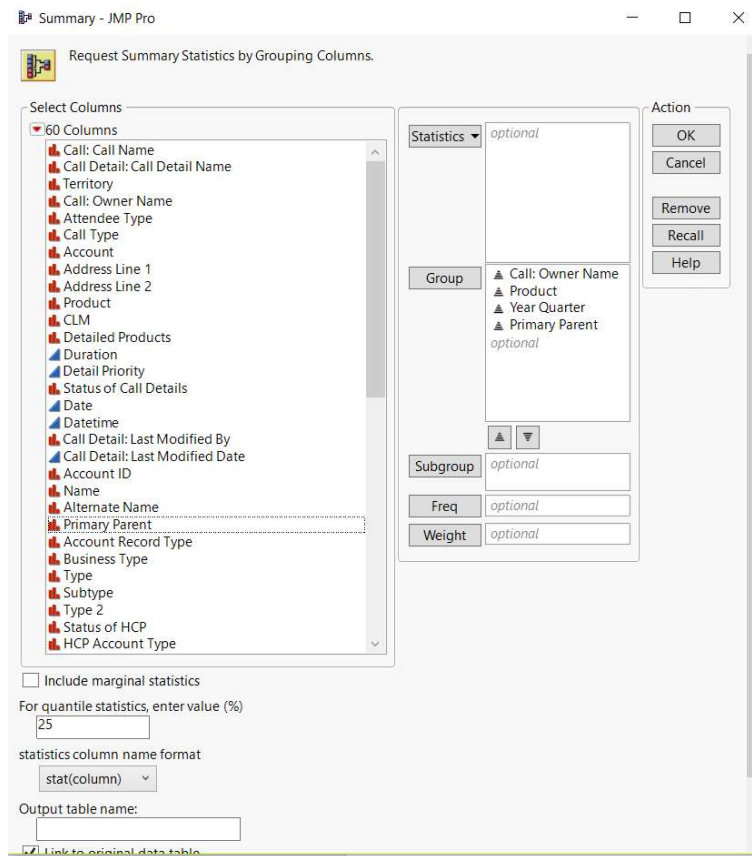
Fig 9: Summary function to aggregate Call Details

        iv.   Click "OK" to summarize the tables

The final stage of data preparation involves joining **Invoice Details** and **Call Details**

1. **Invoice Details** and **Call Details** are inner joined by sales rep's name, clinic's name, product's name and year-quarter. Other variables present in the final table are "channel", "therapy area", "interaction count", "sum(sales qty)" and "sum(amount$).

   Steps Taken:

   i. We have identified the following matching variables from Call Details and Invoice Details

Call Details	Invoice	Description
Call:OwnerName	Rep Name	Name of Sales Rep
PrimaryParent	Name	Name of Clinic
Product	Product Name	Name of Product
YearQuarter	YearQuarter	Year & Quarter

   ii. We will utilize Join function again to perform the inner join

   iii. Go to "Tables" and "Join"

   iv. Enter fields as shown in screenshot below

Fig 10: Join function to match Invoice Details and Call Details

v. Ensure that inner join is selected and output columns are populated before clicking "OK" to finish the join

3.4 INITIAL DATA EXPLORATION AND ANALYSIS
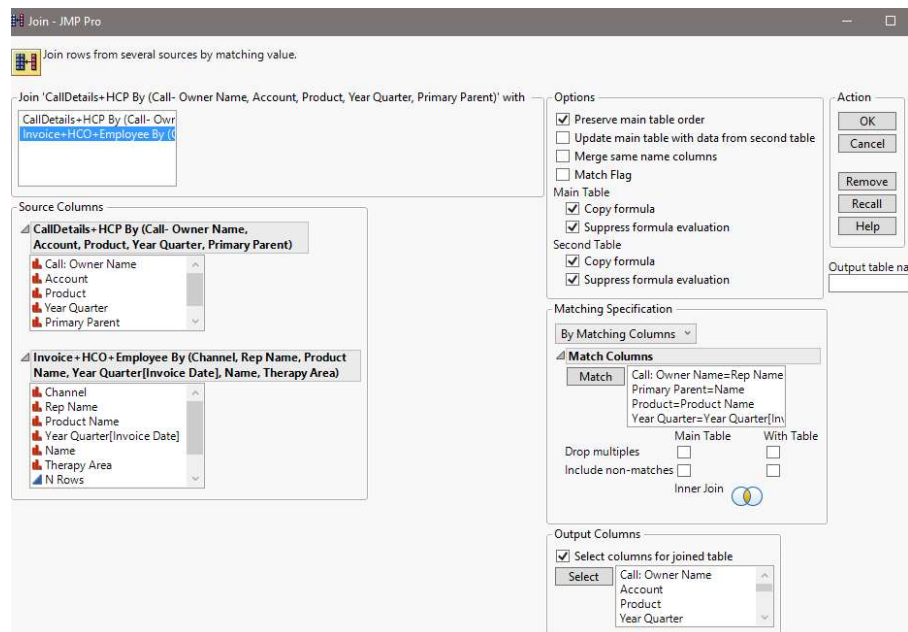
We conducted initial data analysis using Exploratory Data Analysis (EDA) to gain general insights on key determinants that govern the relationship between interaction counts and sales revenue. We hypothesized that such factors could be "channel", "therapy area" (sales team).

"Channel" is the classification for different types of clinics, such as General Practitioners, Restructured & Private Hospitals, Specialists. From our basic understanding, each channel has its own protocols and practices that are likely to affect receptiveness of interactions. For instance, interactions with hospital doctors may not be that impactful as that with GP doctors because hospital doctors get their drugs from a centralized system, while GP doctors have the power to make decisions for their own clinics.

"Therapy area" defines the name of sales teams, such as Uro CNS (Urology), Respi (Respiratory), Paed Vx (Pediatrics Vaccines), Allergy, Al Derm (Dermatology), Ad Vx (Adult Vaccines), and it decides the corresponding product brands to promote. We postulate that different drugs have different demands and established drugs may need small number of interactions to achieve good sales results whereas new types of drugs may need more interactions to achieve the same level of sales results.

To give us a better understanding of the natures of different channels and therapy areas, in this initial data analysis, we will explore how sales revenue differs for different channels across different quarters and how each therapy area performs in terms of sales revenue for different channels.

Exploring sales revenue by channels and quarters allow us to understand significant demand patterns that arise from practices or secondary consumers (patients).

For instance, we will plot a line graph of total sales amount (response) against different quarters (explanatory) by different channels.
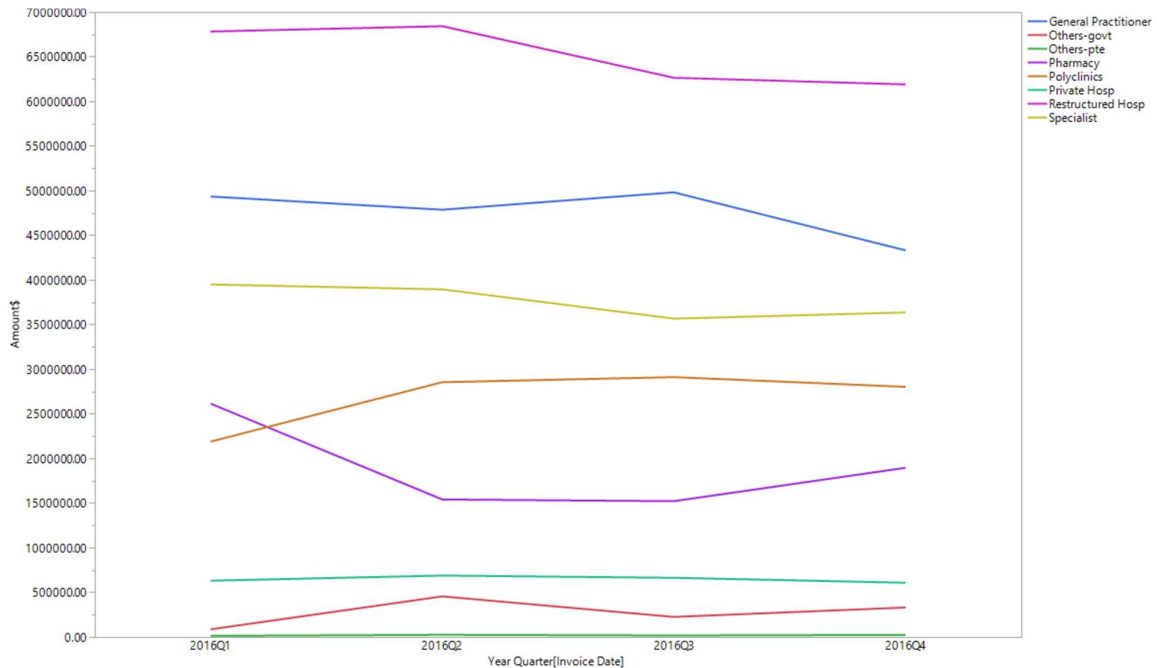
Fig 11: Line graph of Amount($) vs Year Quarter by Sales Channel

A first look at the visualization gives us an understanding that there are indeed intrinsic differences across different channels and quarters.

An observation of trend across quarters is that the highest sales for most channels were made in the first quarter. This is especially prominent for pharmacy, which made more than half of the second quarter sales. To rationalize such trend, we propose two reasons, 1) higher demand from secondary consumers and 2) practice of stockpiling at the start of the year.

Exploring how each therapy area performs in terms of sales revenue for different channels can help to identify which channel each therapy area should focus on.

To visualize, we will plot a mosaic plot of sales revenue by therapy area and channel.
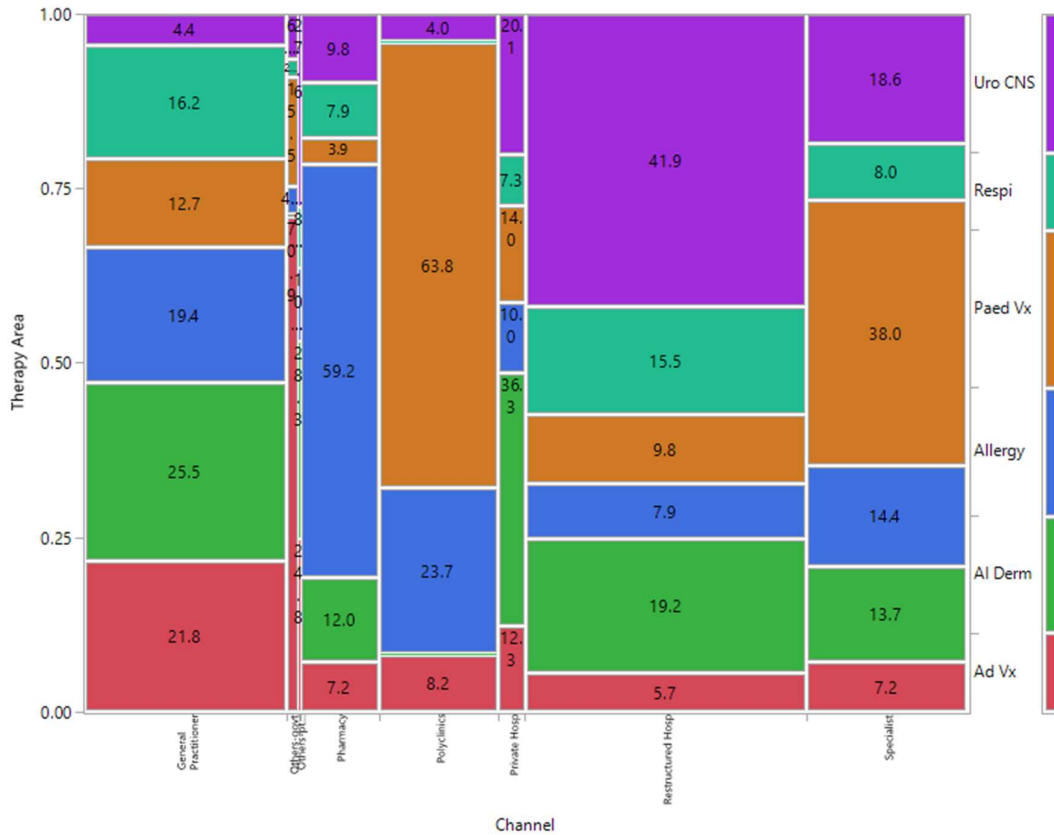
Fig 12: Mosaic Plot of Therapy Area vs Channel, weighted by Sales Amount($)

The use of a mosaic plot is to visualize data from two or more categorical variables in terms of a numerical variable (weight). Referencing Figure 12, x-axis shows the different channels, y-axis shows the different therapy area, area of each box depicts the proportion of sales revenue and labels shows the percentage of each therapy area when compared within a channel.

For instance, we can understand that within a channel, which therapy area generates the highest sales revenue; for Restructured Hospital, Urology team generates the highest sales revenue, for Polyclinics; Paediatrics Vaccine team generates the highest sales revenue.

Secondly, we can also understand how each channel fare against each other in terms of total sales revenue from its relative width; Restructured Hospital has the highest sales revenue, followed by General Practitioner.

This piece of information not only shows inherent variations in terms of sales revenue for therapy area, but also that of therapy area and channel together as it illustrates secondary consumers' (patients) preference over certain channels when it comes to treating different illnesses. This could be important in future work of using two-way ANOVA to understand the effects of multi-factor-group impacts on sales interaction vs sales revenue.

3.5 ACTUAL METHOD: Analysis of Variance (ANOVA) using Fit Y by X

Analysis of Variance is a statistical method used to analyze differences among group means and their variances among and between groups. It is also a form of statistical hypothesis testing to test whether differences between pairs of group means are significant or not.

Prior to using ANOVA, we have attempted using linear regression to generalize the relationship between number of interactions and sales revenue. However, low R-squared values that suggest weak correlation and model not fitting the data were obtained, and these prompted us to carry out similar analysis using nonparametric tests like ANOVA.

The primary step to carry out ANOVA is to discretize our explanatory variable - "interaction count" into bins and as such, converting it from a numerical to categorical variable. The objective of discretization is because we wish to understand whether each of these interaction bins have significant differences between one another when it comes to sales revenue (response).

To define the range of interaction counts for "Low", "Medium" and "High" interaction bins, we consulted our sponsor, who proposed that "Low" is for interaction count less than or equal to 1, "Medium" is for interaction count from 2 to 4 and "High" is for interaction count 5 and above.

The steps taken to discretize interaction counts into bins are as follow:
1. Insert new column right of Interaction Column and name it as Interaction Bin
2. Right click header of Interaction Bin, select "Formula"

3. An interface to formulate the new column is displayed
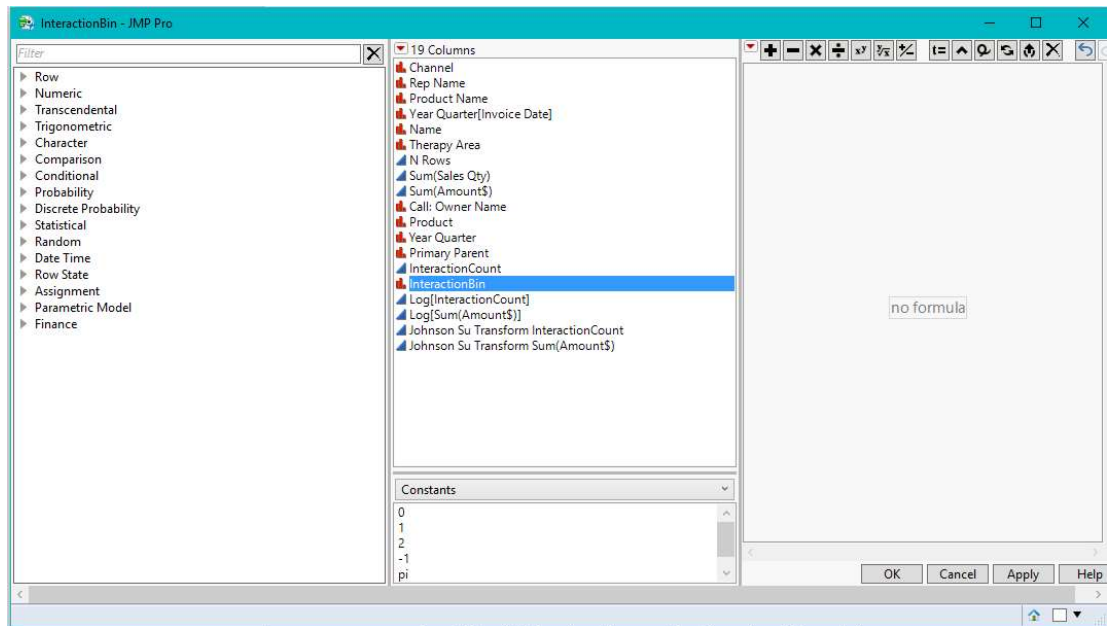
Fig 13: Formula function to discretize interaction counts

4. Using various Conditional and Comparison functions, enter the following formula proposed by our sponsor

Fig 14: Formula derived after consulting sponsor

5. Upon clicking "OK", the new column will be populated with values of "low", "medium" and "high"

The next step to conducting ANOVA would be to use Fit Y by X function. Fit Y by X function can detect whether response or explanatory variables selected are numerical or categorical, and selectively carry out bivariate, oneway, logistic or contingency analysis. In our scenario, our "X, Factor" or explanatory variable is interaction bins (categorical) and "Y, Response" is sales amount (numerical), thus, the analysis conducted would be oneway.

The steps taken to use Fit Y by X function for ANOVA is as follows:

1. Go to "Analyze" and "Fit Y by X"

2. Drop Sum(Amount$) to "Y, Response" and Interaction Bin to "X, Factor"

3. To look into the perspective of individual channels or therapy areas when comparing their means, we will also drop Channel or Therapy Area to "By"

4. Click on "OK" to get one way analysis of Sum(Amount$) by interaction bin for individual channels/ therapy areas

5. To get in-depth details of quantiles for each interaction bin, select the upside-down red arrow and click on "Quantiles"
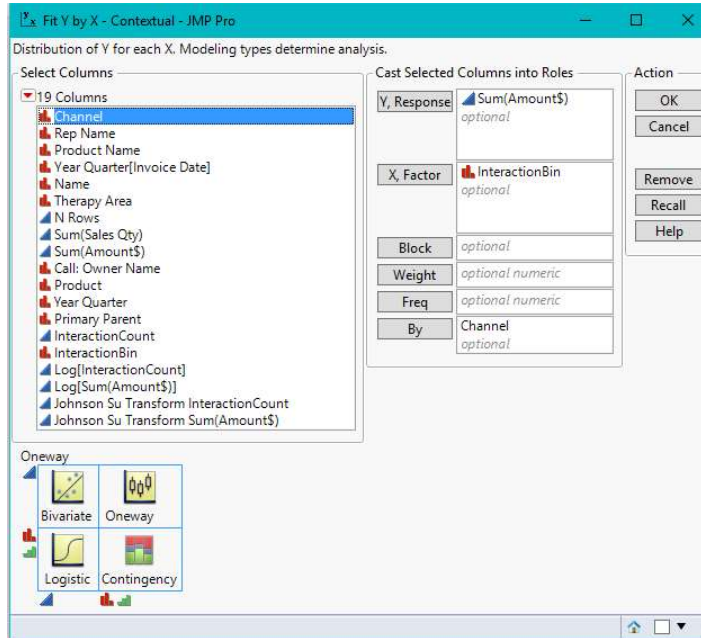
Fig 15: Fit Y by X function

6. Red box plot for each interaction bin will appear
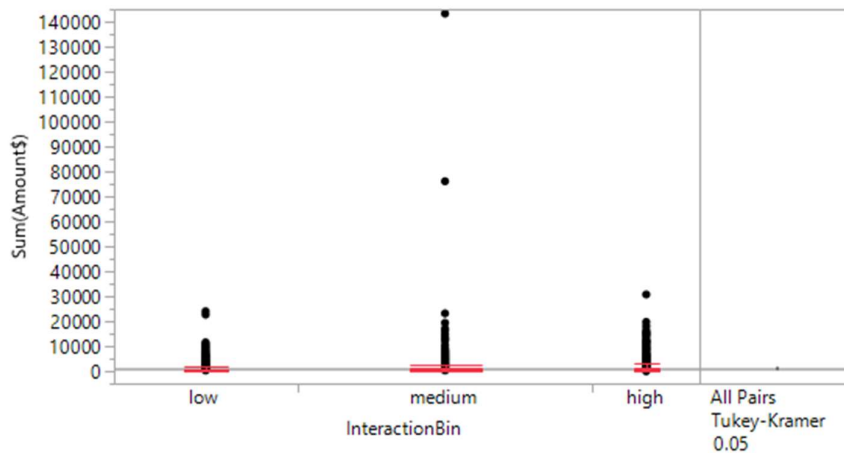
Fig 14: One-way analysis of Sum(Amount$) by Interaction Bin for General Practitioner

7. To conduct Tukey-Kramer HSD test for all pairs of interaction bins, select the upside-down red arrow again, and click on "Compare Means" and "All Pairs, Tukey HSD"

8. A few reports will appear below the graph, but our attention is on the ordered differences report, which calculates p-Value to show whether differences between means of interaction bins are significant or not. Fig 16 below is an instance of the output.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	low	551.1539	83.23091	356.046	746.2623	<.0001*	
high	medium	406.0902	77.17217	225.185	586.9958	<.0001*	
medium	low	145.0637	64.25816	-5.569	295.6965	0.0620	

Fig 16: Ordered Differences Report of Sum(Amount$) by Interaction Bin for General Practitioner

4 ANALYSIS

4.1 BY THERAPY GROUP

Therapy Group 1: Adult Vaccines

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	1710.193	227.4847	1176.71	2243.678	<.0001*	
hi	med	1486.705	200.7101	1016.01	1957.399	<.0001*	
med	low	223.489	181.1401	-201.31	648.289	0.4333	

Fig 17: Ordered Differences Report for Adult Vaccines Therapy Group

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 2: Dermatology

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	11599.79	1128.740	8952.35	14247.23	<.0001*	
hi	med	11128.75	1080.307	8594.91	13662.59	<.0001*	
med	low	471.05	821.523	-1455.82	2397.92	0.8343	

Fig 18: Ordered Differences Report for Dermatology Therapy Group

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 3: Allergy

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	3410.055	900.5882	1298.17	5521.940	0.0005*	
hi	med	2943.662	826.5276	1005.45	4881.873	0.0011*	
med	low	466.394	739.7229	-1268.26	2201.048	0.8033	

Fig 19: Ordered Differences Report for Allergy Therapy Group

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 4: Pediatrics Vaccines

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	5803.216	628.9610	4328.20	7278.232	<.0001*	
hi	med	5151.294	522.3981	3926.18	6376.404	<.0001*	
med	low	651.921	651.1674	-875.17	2179.016	0.5762	

Fig 20: Ordered Differences Report for Pediatrics Vaccines Therapy Group

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 5: Urology

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	35602.88	4716.533	24524.8	46680.97	<.0001*	
hi	med	32173.06	4387.061	21868.8	42477.29	<.0001*	
med	low	3429.82	4349.616	-6786.5	13646.10	0.7102	

Fig 21: Ordered Differences Report for Urology Therapy Group

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.
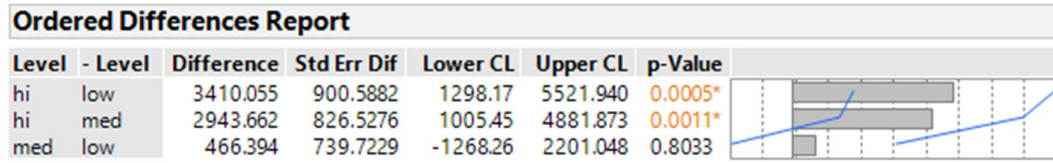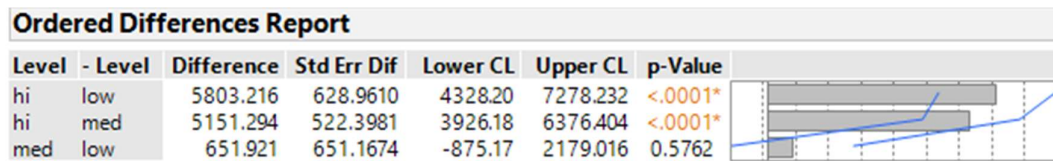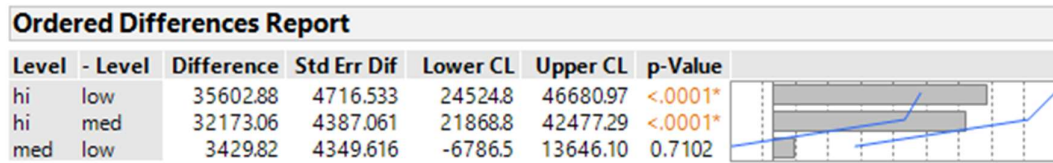
Therapy Group 6: Respiratory

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	med	4272.263	2185.415	-859.59	9404.112	0.1243	
hi	low	2785.862	2500.890	-3086.79	8658.517	0.5057	
low	med	1486.401	2311.274	-3940.99	6913.795	0.7963	

Fig 22: Ordered Differences Report for Respiratory Therapy Group

Results show that across all category pairings (high-medium, high-low, low-medium), the p-value falls in the acceptable range, which means that there is not sufficient evidence for us to reject the null hypothesis that the means are the same.

Discussion for ANALYSIS: BY THERAPY GROUP

Across most therapy groups (5/6), apart from respiratory, it seems that the change from low to high interactions and low to medium interactions bears positive and significant impact on the difference in means, which means that it is always better to perform high level interactions for each therapy group than low and medium level interactions. The change in means reflected is highest on the Urology group. The Respiratory group is an exception, with all p values falling in levels that deny us the chance of rejecting the null hypothesis that the means are the same.

4.2 BY SALES CHANNEL

Sales Channel 1: Pharmacies

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
medium	low	39803.33	76999.01	-153623	233229.7	0.8639	
high	low	37681.13	79231.33	-161353	236715.3	0.8834	
medium	high	2122.20	76999.01	-191304	195548.6	0.9996	

Fig 23: Ordered Differences Report for Pharmacies Sales Channel

Results show that across all category pairings (medium-low, high-low, medium-high), the p-value falls in the acceptable range, which means that there is not sufficient evidence for us to reject the null hypothesis that the means are the same.

Sales Channel 2: Private Hospitals

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	low	7811.272	2326.500	2277.04	13345.50	0.0032*	
medium	low	4956.749	2601.106	-1230.71	11144.21	0.1423	
high	medium	2854.523	2601.106	-3332.93	9041.98	0.5179	

Fig 24: Ordered Differences Report for Private Hospitals Sales Channel

Results show that the p-value between the high-low category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low and high-medium category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different
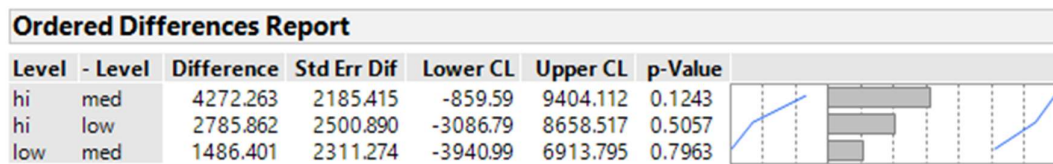
Sales Channel 3: Restructured Hospitals

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	low	23803.26	11008.69	-2076.9	49683.42	0.0788	
high	medium	12593.07	9435.30	-9588.2	34774.38	0.3766	
medium	low	11210.19	13540.61	-20622.2	43042.61	0.6859	

Fig 25: Ordered Differences Report for Restructured Hospitals Sales Channel

Results show that across all category pairings (high-low, high-medium, medium-low), the p-value falls in the acceptable range, which means that there is not sufficient evidence for us to reject the null hypothesis that the means are the same.

Sales Channel 4: Polyclinics

⊿ Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	medium	8766.405	2641.667	2541.80	14991.01	0.0029*	
high	low	6749.466	3558.335	-1635.09	15134.03	0.1415	
low	medium	2016.940	3814.361	-6970.90	11004.78	0.8573	

Fig 26: Ordered Differences Report for Polyclinic Sales Channel

Results show that the p-value between the high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the low-medium and high-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different

Sales Channel 5: Specialists (Neighborhood Clinics)

⊿ Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	low	2863.106	255.7820	2263.302	3462.909	<.0001*	
high	medium	2246.522	214.1281	1744.396	2748.648	<.0001*	
medium	low	616.583	245.4951	40.902	1192.264	0.0324*	

Fig 27: Ordered Differences Report for Specialists (Neighborhood Clinics) Sales Channel

Results show that the p-value between the all pairings(high-low, high-medium, medium-low) is way lower than 0.05 and thus we reject the null hypothesis that the means are the same.

Sales Channel 6: General Practitioner (Neighborhood Clinics)

⊿ Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
high	low	551.1539	83.23091	356.046	746.2623	<.0001*	
high	medium	406.0902	77.17217	225.185	586.9958	<.0001*	
medium	low	145.0637	64.25816	-5.569	295.6965	0.0620	

Fig 28: Ordered Differences Report for General Practitioner (Neighborhood Clinics) Sales Channel

Results show that the p-value between the high-low category and high-medium category is way lower than 0.05 and thus we reject the null hypothesis that the means are the same. For the medium-low category, the p value is higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Discussion for ANALYSIS: BY SALES CHANNEL

A few findings can be seen from these results:

1) Sales interactions increases results in improvement in means for neighbourhood clinics, both Specialists and General Practitioners.
2) Restructured hospitals and Pharmacies have not been proven to be affected by any change in interaction.
3) Different sales channels have extremely different patterns of sales (i.e. affected by interactions in different ways
4) Private hospitals only see a change in revenues when interactions are increased greatly from low to high, but not from a low to medium level.
5) Polyclinics behave the most oddly, with changing of interactions from medium to high having a significant improvement in mean yet there is no conclusive evidence that a change from low to medium interactions result in a change in mean revenues.

4.3 Further Analysis, Quarter Anomalies

For each therapy group, the team has decided to explore deeper by looking at each therapy group by quarterly performance and each sales channel by a quarterly view. We expect quarters to follow the results as in the year's results, which otherwise, could mean that there is a certain action causing an anomaly during the quarter.

Therapy Group 1: Adult Vaccines

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	730.3443	258.2738	123.228	1337.461	0.0135*	
med	low	404.4832	228.4718	-132.579	941.545	0.1806	
hi	med	325.8611	188.2574	-116.670	768.392	0.1947	

Fig 29: Ordered Differences Report for Quarter 4 of Adult Vaccines Therapy Group

Q1, 2 and 3 of the Adult Vaccines Therapy Group follow the same result as in section 4.1. Q4 however, shows an anomaly. Results show that the p-value for the medium-low category now has a p value higher than 0.05, therefore we do not reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 2: Dermatology

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
med	hi	711.0692	1044.848	-1748.92	3171.063	0.7750	
low	hi	586.5623	1031.517	-1842.05	3015.170	0.8369	
med	low	124.5069	251.587	-467.83	716.843	0.8738	

Fig 30: Ordered Differences Report for Quarter 1 of Dermatology Therapy Group

Q2, 3 and 4 of the Dermatology Therapy Group follow the same result as in section 4.1. Q1 however, shows an anomaly. Q1, as in above, have inflated p-values in all categories, which means that we can no longer reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 3: Allergy

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	8813.240	5006.728	-2965.84	20592.32	0.1845	
med	low	4908.673	2780.107	-1631.95	11449.29	0.1826	
hi	med	3904.568	5093.162	-8077.86	15887.00	0.7237	

Fig 31: Ordered Differences Report for Quarter 1 of Allergy Therapy Group

Q1 shows a deviation in the results where p-value for the first two pairings(high-low, medium low) now have inflated p-values, which means that we can no longer reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	low	5330.512	2722.369	-1065.57	11726.60	0.1236	
hi	med	4725.129	1980.577	71.85	9378.41	0.0456*	
med	low	605.382	2500.084	-5268.45	6479.22	0.9682	

Fig 32: Ordered Differences Report for Quarter 2 of Allergy Therapy Group

Q1 shows a deviation in the results where p-value for the first pairing(high-low) now have inflated p-values, which means that we can no longer reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	med	1365.117	1413.494	-1955.16	4685.394	0.5988	
low	med	1186.054	1300.644	-1869.14	4241.246	0.6330	
hi	low	179.064	1551.955	-3466.46	3824.584	0.9927	

Fig 33: Ordered Differences Report for Quarter 4 of Allergy Therapy Group

Q4 shows a deviation in the results where p-value for the first two pairings(high-medium, low-medium) now have inflated p-values, which means that we can no longer reject the null hypothesis that the means are the same as there is not enough evidence to show that the means are different.

Therapy Group 4: Pediatrics Vaccines

There is no deviation from the original yearly results for paediatrics vaccines

Therapy Group 5: Urology

There is no deviation from the original yearly results for paediatrics vaccines

Therapy Group 6: Respiratory

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value	
hi	med	12686.81	5046.549	765.0	24608.66	0.0340*	
hi	low	11684.23	6318.759	-3243.1	26611.51	0.1566	
low	med	1002.58	5377.557	-11701.2	13706.40	0.9810	

Fig 34: Ordered Differences Report for Quarter 3 of Respiratory Therapy Group

Q1, 2 and 4 of the Dermatology Therapy Group follow the same result as in section 4.1. Q3 however, shows an anomaly. We can now reject the null hypothesis that the means are the same for the high-low pairing, therefore we can infer that the means are different with a change in level of interactions.

Discussion for Further Analysis, Quarter Anomalies

Most results follow the original (as in on a yearly view in Section 4.1) with a few exceptions. In the case of adult vaccines, while Q4 shows a higher p-value in the 2nd group, it is not extremely inflated, thus there might still actually be a change in means, making the slight exception negligible. For the case of Dermatology, there needs to be investigation done regarding Q1 as to why there is such a huge deviation in results. Oddly enough, for the allergy group, three quarters differ from the original results, which means that investigation is needed for this area too. Finally, for respiratory, there is oddly enough a category with a p-value of less than 0.05 in the first category and thus more investigation needs to be done in this area.

5 MANAGERIAL RECOMMENDATIONS

<u>Therapy Group</u>

From the above findings, it can be concluded that therapy groups largely follow a certain pattern on interactions vs sales revenue, and it is best for therapy groups to be interacting at a high level to achieve the best results in sales revenue. This means that sales interactions do directly impact the revenue that the company would potentially get and that resources should be invested in this area. It is interesting that the respiratory group follows such a deviation from the rest of the therapy groups, and more can be studied to uncover why there is such an anomaly, possible reasons being a failure to record interactions, or even the nature of drugs; that respiratory drugs do not require much pushing in sales for it to be sellable. What makes it even more peculiar is that the Respiratory group's most profitable channel is the General Practitioner channel, which is an extremely impressionable group and so by right, the number of interactions should have impacted the sales revenue. Therefore, the sponsor should carry an investigation in this area to find out the true meaning of the results – has the respiratory team been skewing their results due to some actions?

Sales Channel

It is unsurprising to see that Specialists and General practitioners are the most impressionable, given that they come from neighbourhood clinics that probably gives them the most autonomy in stocking decisions. Therefore, large emphasis should be placed on interacting with these individuals to drive revenue. It seems that there is no evidence that restructured hospitals and pharmacies are affected by increases in interactions, which means that they probably are clear on what they want and could potentially be hard to persuade, thus less attention can be spent on these areas. Private hospitals are also extremely manpower intensive, which mean that there shouldn't be a wastage of resources by planting medium level interactions, but should instead be allocated full force of interactions to better the business revenue.

Anomalies in quarter: Therapy Group

For the Dermatology group, Q1 could have been so different from the other groups due to a possible introduction of new drug, leading to better results in Q2 onwards, or even a slack in manpower leading to an unfound improvement in mean. For the respiratory group, it is interesting to find a group where there is a conclusive change in means, which could possibly debunk the fact that respiratory drugs are "Sellable" by nature. That then begs the question as to why there is no conclusive change in means when we were analysing it by therapy group. It could possibly be that the respiratory group has not been functioning in a perfect mode, or it could simply be due to a stroke of luck in that full quarter. Once again, there needs to be further investigation done in all these areas for more insights. Lastly, the fact that most results do follow the results in a yearly view show that quarter performances mostly conform to the pattern in a yearly basis, showing high reliability in results.

6 CONCLUSION

In all, we can see that different factor groups (sales channels/therapy groups/quarters) do indeed have differentiating trends in sales interactions vs sales revenues. Neighbourhood Clinics, which include General Practitioners and Specialists are especially impressionable. Most therapy groups benefit from the increase in sales interactions, especially at a High level of interaction. Quarterly results are largely similar to the yearly view, showing high consistency of results across quarters and therefore making our results largely reliable.

7 LIMITATIONS

Given that sales channels and therapy groups are not mutually exclusive groups, it is hard to predict sometimes, if an inferred changed in means in a certain therapy group, could be partially due to the impact coming from the sales channels that they are working with. More work needs to be done in zooming down to specific combination impacts. Another limitation we are unable to pin point are the actual causes of some results like in the case of the respiratory group and more studies need to be done to identify the actual causes.

REFERENCES

Aluregowda, A. (2013). Organized Retail Strategy – A Study at Loyal World Supermarket. IOSR Journal of Business and Management, 7(1), 75-80. doi:10.9790/487x-0717580

Anic, I., & Radas, S. (2006). The Relationships Between Shopping Trip Type, Purchases Made on Promotion, and Unplanned Purchases for a High/Low Hypermarket Retailer. Privredna kretanja i ekonomska politika , 107 , 27-45. Retrieved April 18, 2017.

Lazic, S. E. (2008). Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. BMC Physiology, 8(1), 16. doi:10.1186/1472-6793-8-16

Naik, R. S., Dr. (2014). Marketing Strategies OF Toilet Soap Manufacturing Industry in Karnataka (A Case Study of KS&DL, Bangalore). *International Journal of Business and Administration Research Review,1*(5), april-june, 254-260. Retrieved April 18, 2017.

Pandit, J. J. (2010). The analysis of variance in anaesthetic research: statistics, biography and history ANOVA in anaesthetic research. *Anaesthesia*, *65*(12), 1212-1220. doi:10.1111/j.1365-2044.2010.06542.x

Singer, J. D. (1991) Types of Factors and Their Structural Layouts, in Fundamentals of Exploratory Analysis of Variance (eds D. C. Hoaglin, F. Mosteller and J. W. Tukey), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470316832.ch4