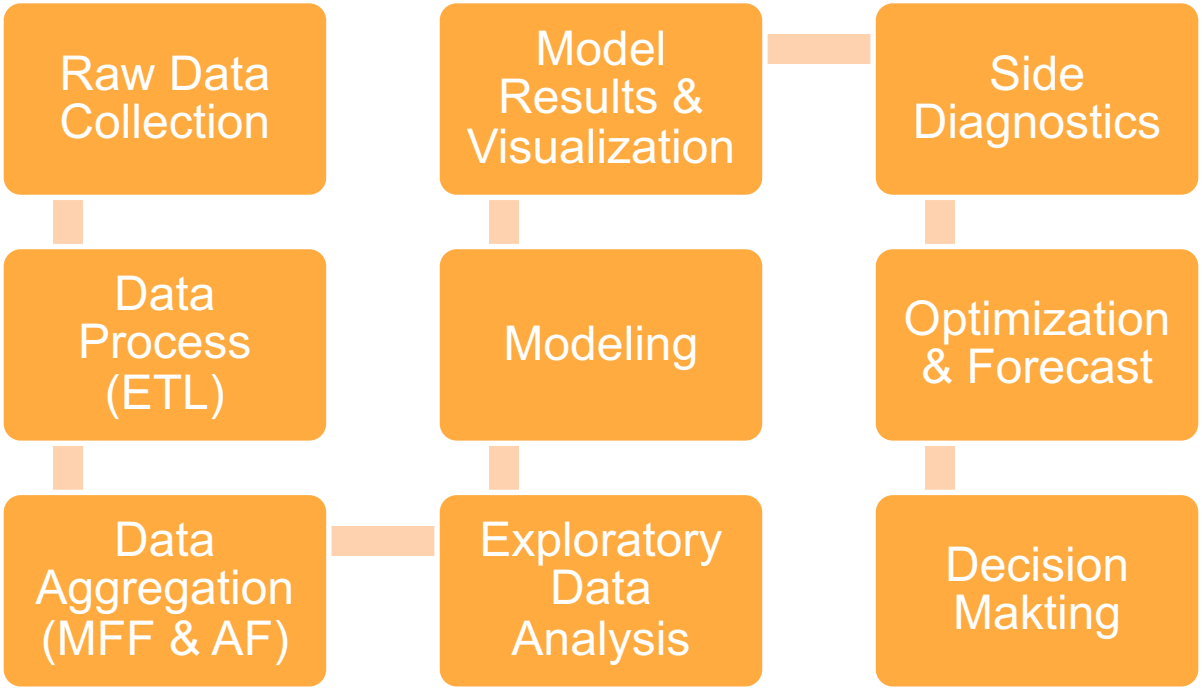
Marketing Mix Modeling

Analyst: Yin(Fien)Xu

Modeling Process

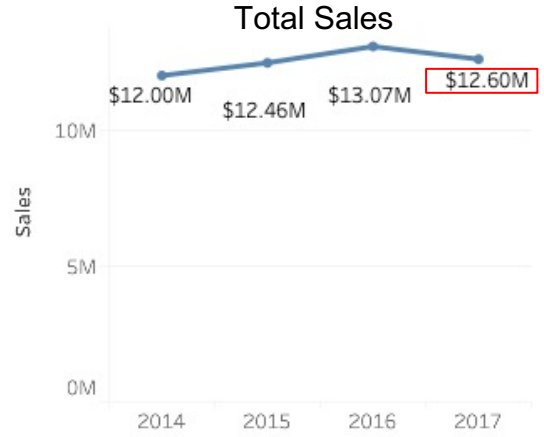


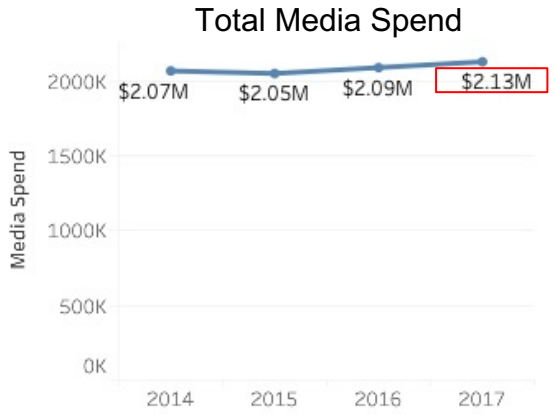
Background

[redacted] is a **leather good fashion design** company, positioning as a middle-class fashion brand targeted white-collar females.

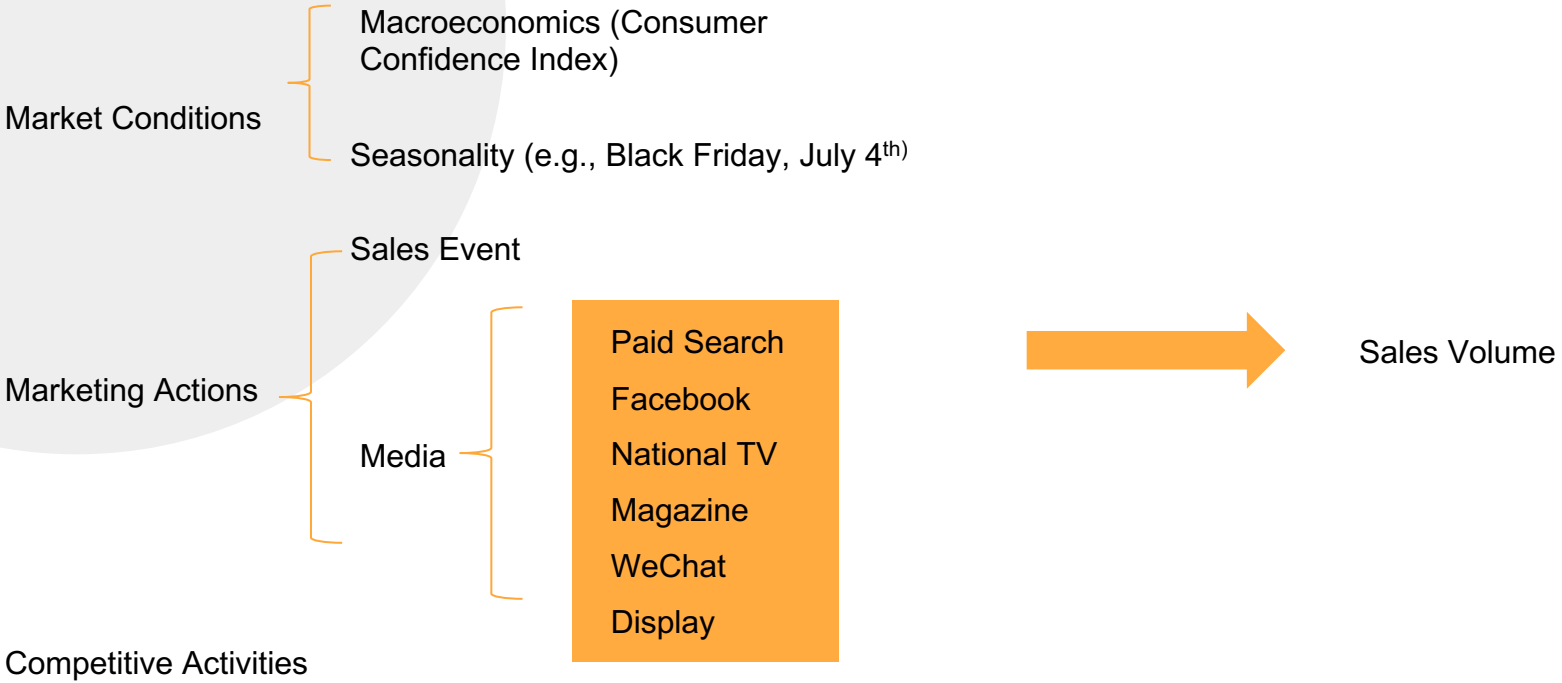
Main Business Questions:

1. What will be the impact from different media channels on sales?
2. What are the ROI for each channel?
3. What will be the optimal spend and mix of marketing investments?
4. What will be the predicted sales in the future?

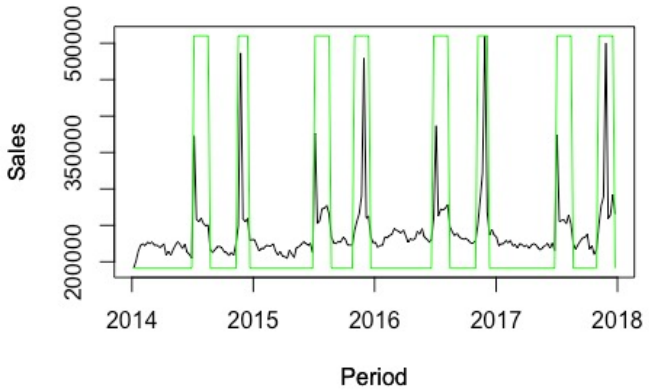




Factors Considered



Exploratory Analysis



Mark the sales spike(July 4th & Black Friday) when building model

```
#install packages
install.packages("ggplot2")
library (ggplot2)

#line chart
plot(data$Period, data$Sales, type='l', xlab = 'Period', ylab = 'Sales')

#add another line
par (new = TRUE) #add another line on the previous line
plot(data$Period, data$Sales.Event, type='l', col='green', xlab = "",ylab="", axes =FALSE)
axis (side=4)

#correlation matrix
correl = cor(data[,c(-1,-2)])
write.csv(correl, file='correl.csv')

#correlation matrix chart
install.packages('corrplot')
library ("corrplot")
corrplot(correl, tl.cex=0.7, tl.col = 'black')

#create some dummy values: to mark the spike of special dates as 1
#first peak --Black Friday
AF$Black_Friday = 0
AF[which(AF$Period=='2014-11-24'),'Black_Friday'] = 1
AF[which(AF$Period=='2015-11-30'),'Black_Friday'] = 1
AF[which(AF$Period=='2016-11-28'),'Black_Friday'] = 1
AF[which(AF$Period=='2017-11-27'),'Black_Friday'] = 1
sum(AF$Black_Friday) # check 4 spikes

#second peak --July 4th
AF$July_4th = 0
AF[which(AF$Period=='2014-07-07'),'July_4th'] = 1
AF[which(AF$Period=='2015-07-06'),'July_4th'] = 1
AF[which(AF$Period=='2016-07-04'),'July_4th'] = 1
AF[which(AF$Period=='2017-07-03'),'July_4th'] = 1
sum(AF$July_4th) # check 4 spikes
```

Modeling

	Decay	Lag	Alpha
National TV	0.95	0	0.6
Magazine	0.9	2	1
Paid Search	0.7	0	1
Display	0.8	0	0.75
Facebook	0.8	0	0.85
WeChat	0.8	1	0.86

Media Variables Transformation

R code:

model = lm (data = AF, Sales ~ CCI + Sales.Event + July_4th + Black_Friday+NationalTV2+PaidSearch1+Wechat2+Magazine2+Display3+Facebook1)

```
Residuals:
   Min     1Q Median     3Q    Max
-25892  -4958   -881   4644  46717

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.502e+05  1.546e+04   9.714  < 2e-16 ***
CCI          5.657e+01  1.286e+02   0.440 0.660541
Sales.Event  3.066e+04  1.741e+03  17.609  < 2e-16 ***
July_4th     1.220e+05  4.440e+03  27.477  < 2e-16 ***
Black_Friday 2.204e+05  5.186e+03  42.496  < 2e-16 ***
NationalTV2  1.518e+03  3.612e+02   4.202 4.01e-05 ***
PaidSearch1  6.413e-02  1.747e-02   3.671 0.000311 ***
Wechat2      1.128e+01  2.598e+00   4.341 2.27e-05 ***
Magazine2    7.818e+02  1.143e+02   6.842 9.61e-11 ***
Display3     8.397e-03  6.832e-03   1.229 0.220489
Facebook1    1.073e-02  1.525e-03   7.042 3.09e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8300 on 197 degrees of freedom
Multiple R-squared:  0.968,     Adjusted R-squared:  0.9664
F-statistic: 596.6 on 10 and 197 DF,  p-value: < 2.2e-16
```

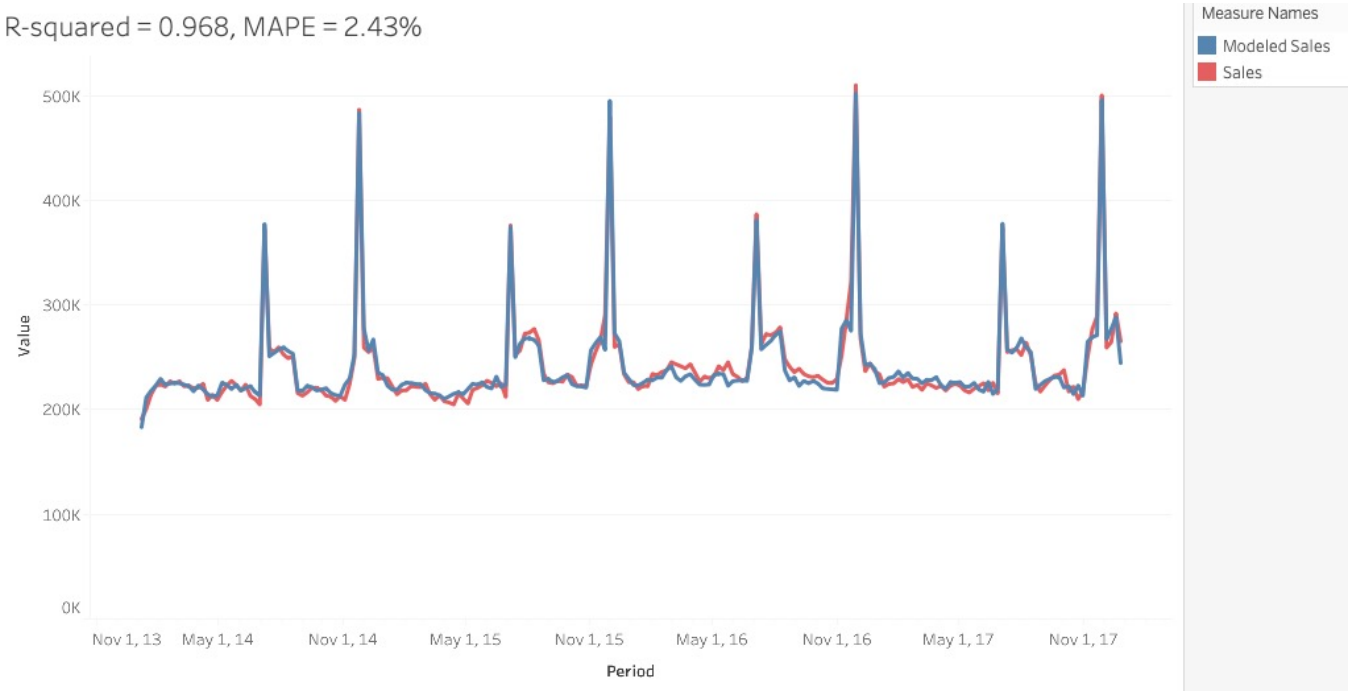
Optimal:

t-value: 2 indicates a positive difference between our sample data and the null hypothesis (prefer > 2);

R-Square: 1.0 means model is prefect;

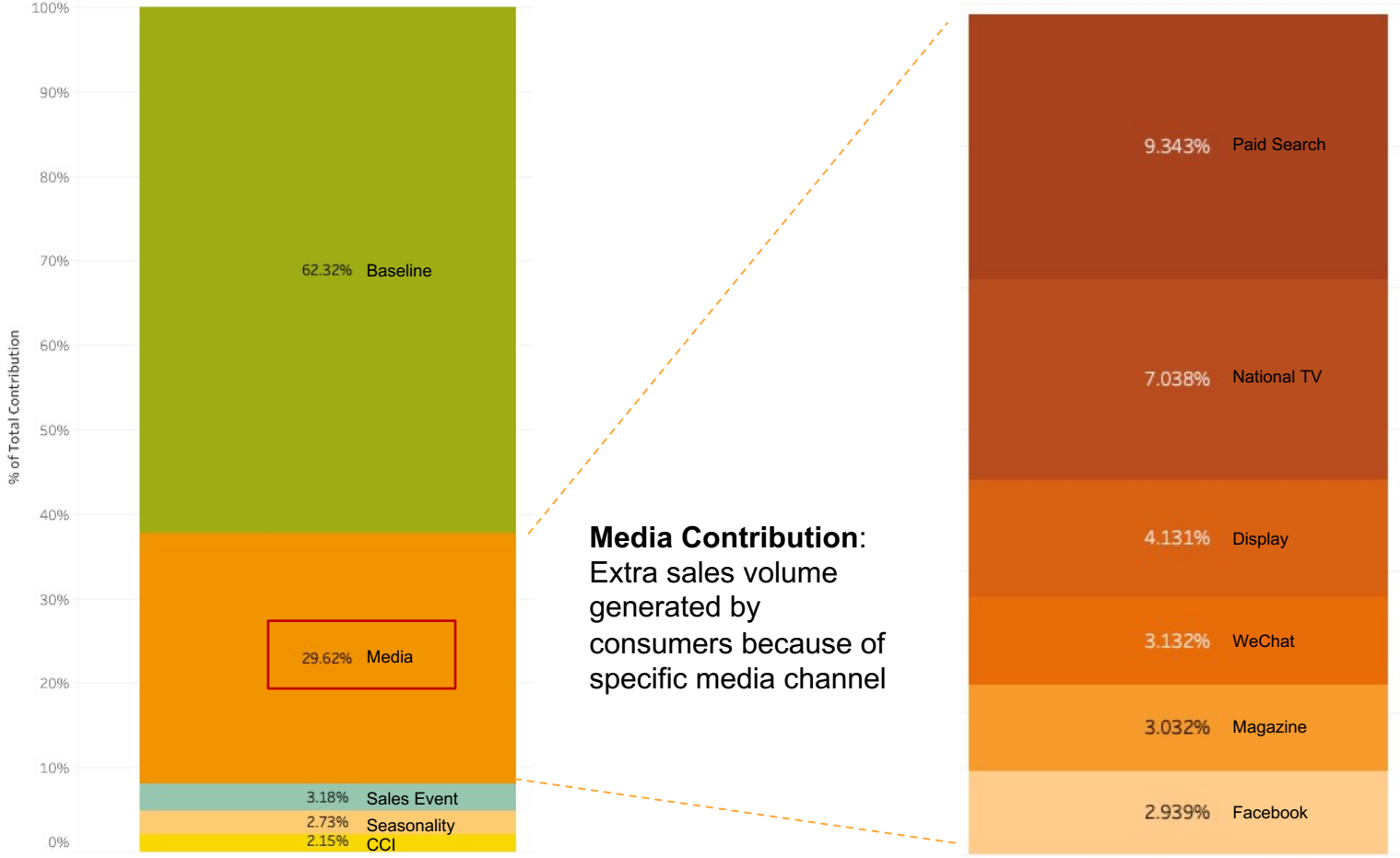
P-value < 0.05: Statistically significant.

Actual Versus Model



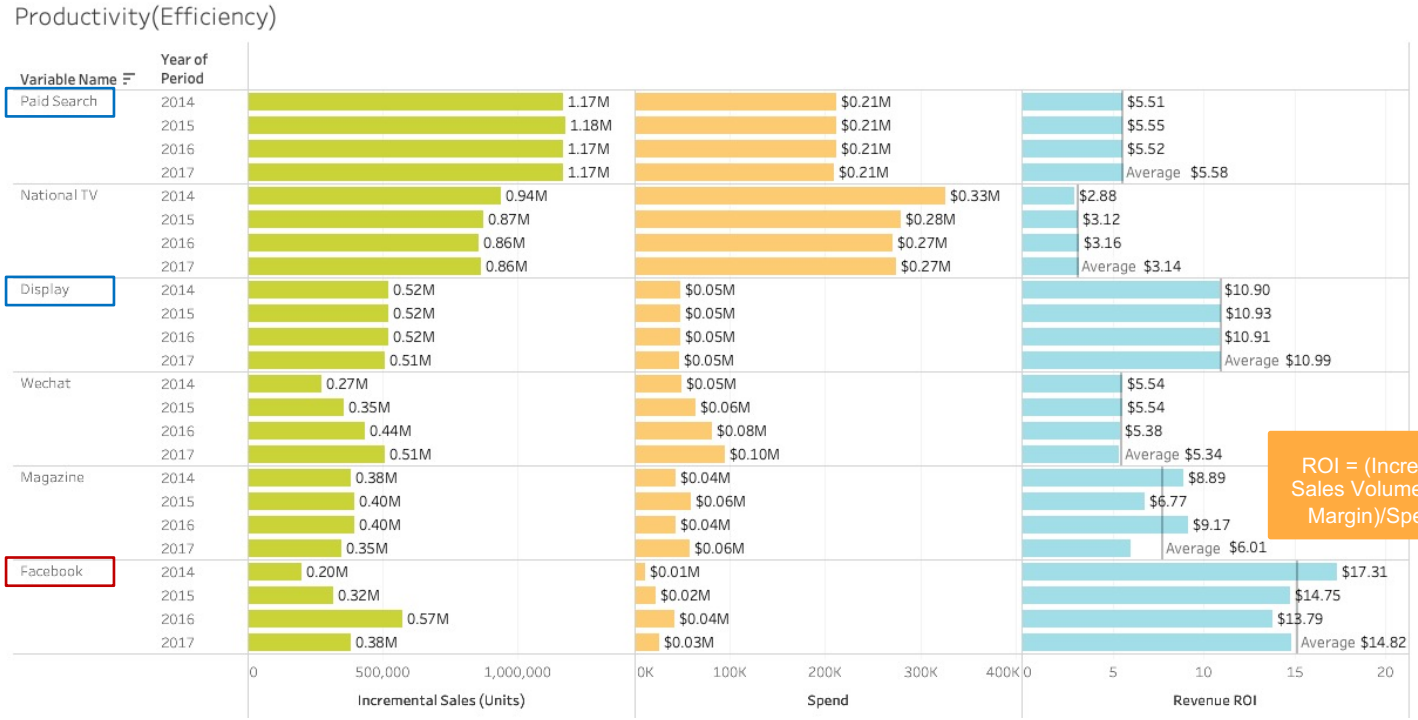
The model matched the actual sales performance

Media Contribution



Media Contribution: Extra sales volume generated by consumers because of specific media channel

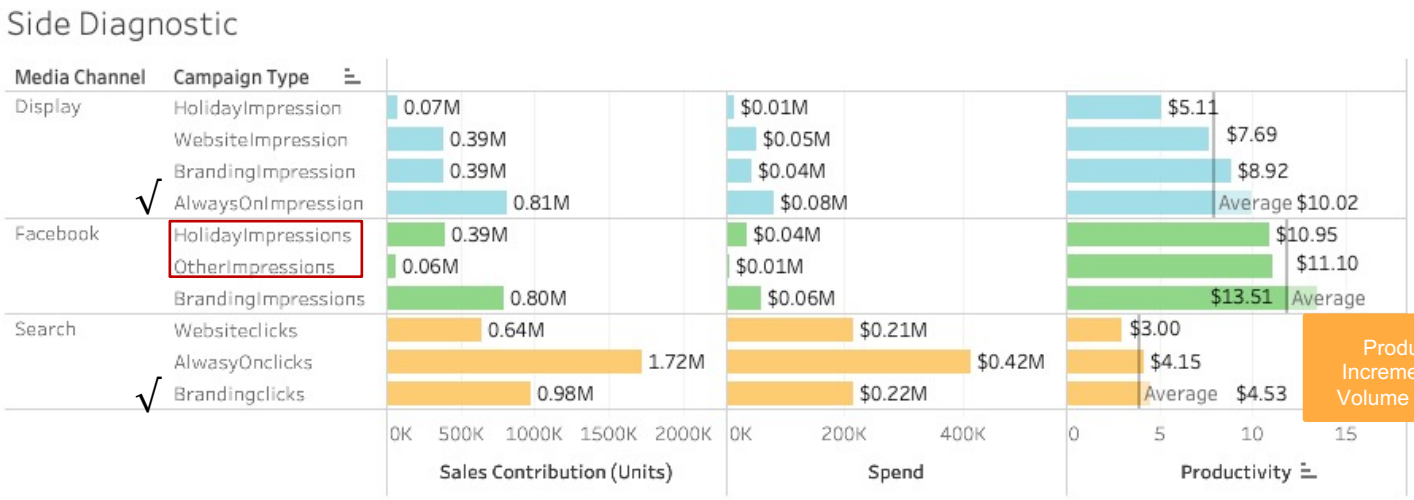
Channels Break Down



ROI = (Incremental Sales Volume * Profit Margin)/Spending

- Facebook channel has the highest ROI (Efficiency)
- Display & Paid Search channels have the least number of year that fall behind the average ROI

Side Diagnostics



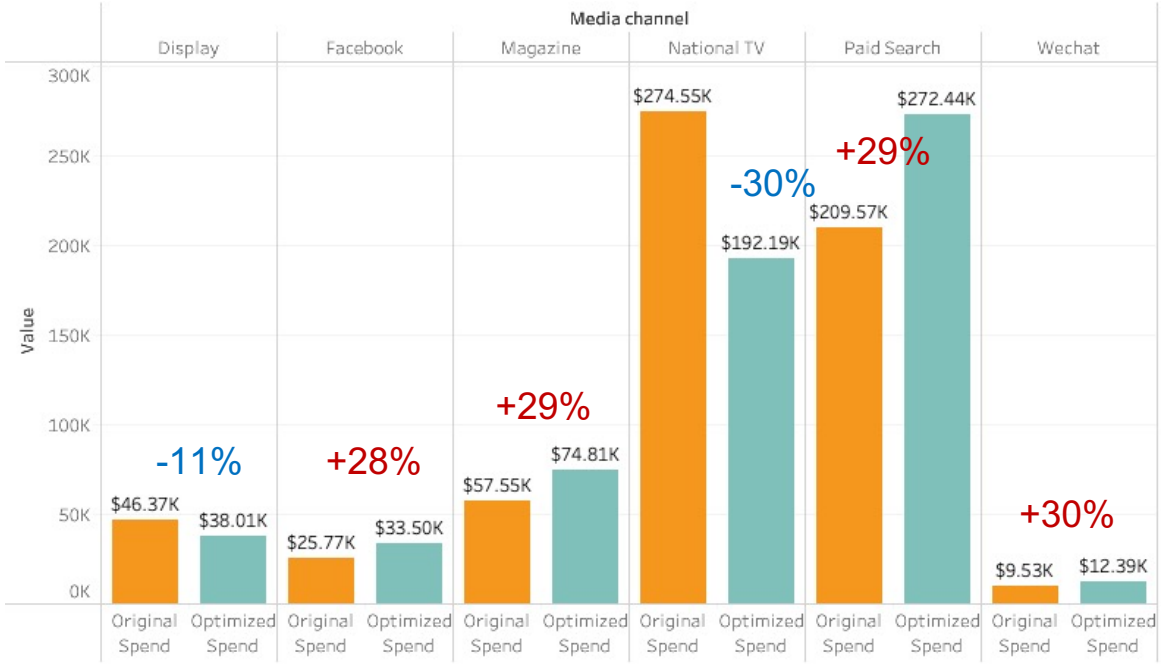
Percentage of ROI over total that BELOW average:

- Facebook: Holiday Impressions & Others
- Display: Holiday & Website Impressions
- Search: Website clicks

Suggestions:

- Pay close attention on Holiday branding campaigns
- Buy branding keywords with SEO strategy
- Engage with customers by creating connective tissue

2018 Budget Optimization



Objective: Total Sales
Variables: Media Spend
Constraint:

1. Total Media spend is unchanged
2. Each optimized media spend is constrained within +/-30% range

Result:

- Total efficiency is increased from 1.01 to 3.45
- Total sales is increased from 777K to 2151K

Summary & Next Step

- The most efficient media channel is Facebook, which triggers more incremental profit gained per dollar spent on marketing activities
- Paid Search contributes to incremental sales volume more than the other channels
- Through optimization, both predicted total sales & media efficiency (ROI) are increased
- Increase budget for all channels, except for National TV and Display channels
- Need to consider modeling DMA level data if available to use more granular data

THANKS

Do you have any questions?
Please feel free to contact me
for detail R & SQL Code

yinxufien@gmail.com

Yin(Fien)Xu

Appendix

SQL Data Cleaning (Sample):

```sql
/*Facebook*/
/*1*/
create table facebook_extracted as
(select * from mmm.mmm_facebook_raw);

/*2*/
create table facebook_transformed as
(
select ap_total_imps as FacebookImpressions, ap_total_clicks as FacebookClicks, round(ap_total_clicks/ap_total_imps,2) as FacebookCTR
from facebook_extracted
where ap_total_imps !=0
);
/*update2*/
select
Period,
sum(ap_total_imps) as FacebookImpressions,
sum(ap_total_clicks) as FacebookClicks,
sum(ap_total_clicks)/sum(ap_total_imps)  as CTR
from mmm.mmm_facebook_raw
group by Period
;

/*update 2*/
CREATE TABLE facebook_transform
(
SELECT b.`week`, SUM(ap_total_imps) AS FacebookImpressions,
       SUM(ap_total_clicks) AS FacebookClicks,
       ROUND(SUM(ap_total_clicks)/SUM(IF(ap_total_imps = 0, 1, ap_total_imps)),3) AS FacebookCTR
FROM mmm_facebook_raw a
LEFT JOIN mmm_date_metadata b ON a.period = b.`Day`
GROUP BY 1
);
```

Appendix

R Code (Sample):

```
#Step7: add campaigns based on spending ~ FB
model7 =  lm (data = AF, Sales ~ CCI + Sales.Event + July_4th + Black_Friday+NationalTV2+PaidSearch1+Wechat2+Magazine2+Display3+Facebook1)
summary(model7)

#check VIF
# In regression, "multicollinearity" refers to predictors that are correlated with other predictors.  Multicollinearity occurs when your model includes multiple factors that are correlated not just to your r
# Variance inflation factor (VIF) is used to detect the severity of multicollinearity in the ordinary least square (OLS) regression analysis.)
#If the VIF is equal to 1 there is no multicollinearity among factors
#If the VIF is greater than 1, the predictors may be moderately correlated.
#A VIF between 5 and 10 indicates high correlation that may be problematic. And if the VIF goes above 10, you can assume that the regression coefficients are poorly estimated due to multicollinearity.
install.packages('car')
library('car')
vif(model7)#all vif of each variables < 3, which variables are moderately corelated-> accetaable

#AVM --actrual vs model
AVM = cbind.data.frame(AF$Period, AF$Sales,model7$fitted.values)
colnames(AVM) = c('Period','Sales','Modeled Sales') #change column name
write.csv(AVM,file = 'AVM1.csv', row.names=F)

#MAPE
MAPE = abs(AVM$Sales-AVM$`Modeled Sales`)/AVM$Sales
mean(MAPE)

#Calculate & Export contribution
model7 =  lm (data = AF, Sales ~ CCI + Sales.Event + July_4th + Black_Friday+NationalTV2+PaidSearch1+Wechat2+Magazine2+Display3+Facebook1, x=TRUE) # x means aggregated the variables to a new column
View(model7$x)
model7$coefficients
contribution =  sweep(model7$x,2,model7$coefficients,"*") #different dimensions multiply
View(contribution)
contribution =  data.frame(contribution)
contribution$Perid = AF$Period
names(contribution) = c(names(model7$coefficients),'Period')

#Transform to long format to better visualize in Tableau
install.packages('reshape')
library(reshape)
contri = melt(contribution, id.vars = "Period")
View(contri)
write.csv(contri, file ='contribution.csv',row.names = F)
```

Appendix

Tableau (Sample):