Digital Attribution at W.M. Winters

Adil Wahab

Project Background

The project uses the "Winters_Attribution" data. The Winters case study uses a unique data set was collected in collaboration with a large online media analytics and optimization platform company. The online media company managed the entire campaign of a U.S.-based retailer. The individual-level data set consists of advertising exposures and user-initiated actions, with users tracked across different advertising channels and media. Note that all observations relate to an order (touchpoints that do not lead to a purchase are absent). The unit of observation is an order-touchpoint, so that the same order is repeated by the number of touches. A brief description of the variables follows: 1. Orderid: actual transaction id from the vendor — if there are multiple touch points, you will see this ID listed on each row 2. Orderdatetime: UTC time of the transaction 3. Saleamount: value of the transaction 4. Newcustomer: Y = yes // N = no 5. Position: touchpoint position 6. Positiondatetime: UTC time of the actual touchpoint 7. Groupname: the group/channel by which the client categorized their marketing activities (e.g. CPM = display advertising, BUZZ AFFILIATE & CJ = affiliates) 8. Networkname: this is the name of the media touchpoint 9. Networkid: this is the id used by C3 10. Brand: this is used for search terms, where N = 'Non-Brand' and Y = 'Brand Search' 11. Positionname: C3 Metrics terminology for the touchpoints. a) "Originator": the first touchpoint b) "Converter": the last touchpoint c) For more than 2 touchpoints: i. "Assist": the penultimate touchpoint ii. "Roster": any touchpoints between "Originator" and "Assist" 12. DaysToConvert: Days between first touch and conversion 13. Touches: Number of touchpoints

EDA

```
head(winters,5)
```

```
##    Orderid   Orderdatetime Saleamount Newcustomer Position Positiondatetime
## 1 11634052 2012-05-01 4:24      341.5           Y        1  2012-05-01 3:49
## 2 11634052 2012-05-01 4:24      341.5           Y        0  2012-05-01 3:47
## 3 11634059 2012-05-01 4:08      339.0           Y        2 2012-04-29 21:01
## 4 11634059 2012-05-01 4:08      339.0           Y        1  2012-04-24 5:29
## 5 11634059 2012-05-01 4:08      339.0           Y        0  2012-04-23 2:46
##             Groupname                          Networkname      Networkid
## 1      BUZZ AFFILIATE                   Buzz CPA Affiliate        buzz23
## 2 SEARCH GOOGLE BRAND G: Medifast Brand Terms > Medifast >        g000793
## 3   PRINT - MAGAZINES                       Medifastok.com medifastok.com
## 4                 CPM                              Armonix        nar7467
## 5   PRINT - MAGAZINES                       Medifastok.com medifastok.com
##   Brand Positionname DaysToConvert
## 1     N    CONVERTER             0
## 2     Y   ORIGINATOR             0
## 3     N    CONVERTER             2
```

```
## 4      N       ASSIST             7
## 5      N   ORIGINATOR             8
```

1. First-touch vs Last-touch

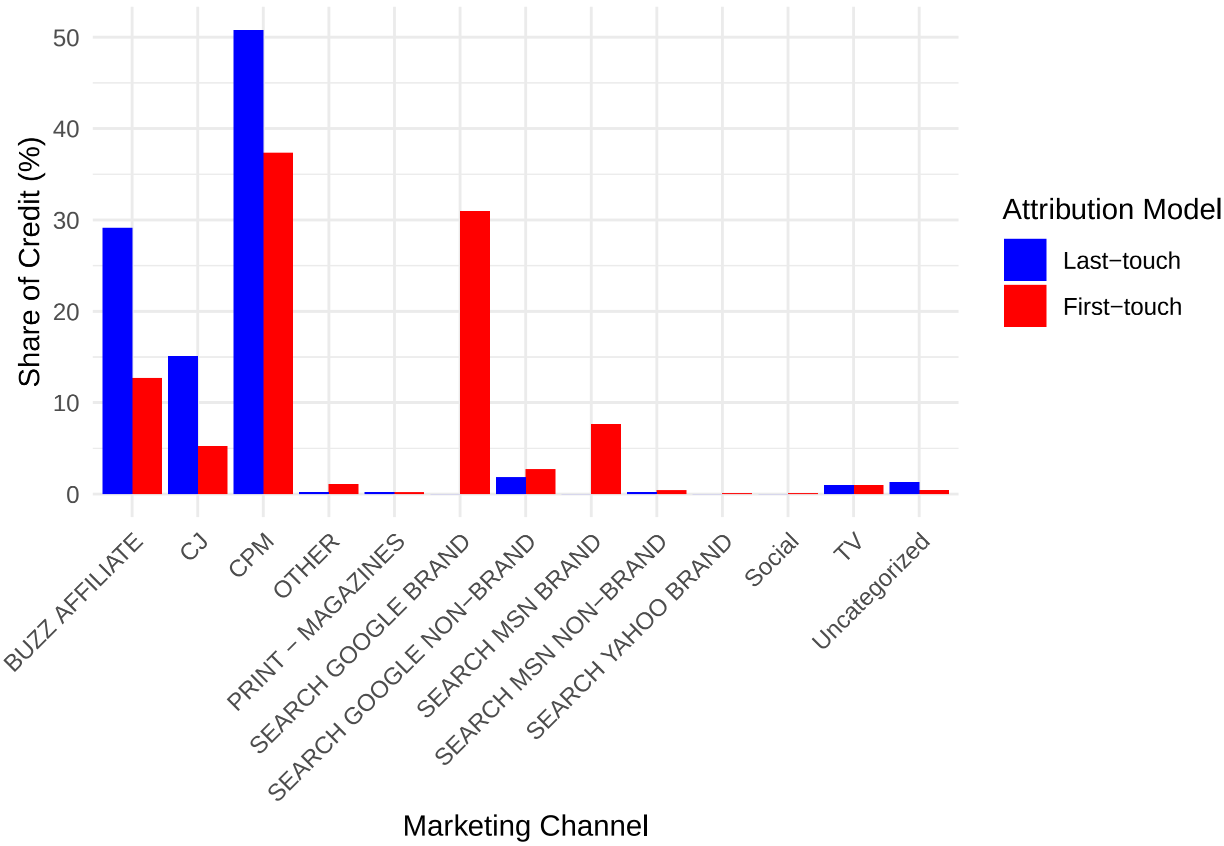
a.

```
# create new columns for originator and converter
winters$Converter = ifelse(winters$Positionname == "CONVERTER", 1, 0)
winters$Originator = ifelse(winters$Positionname == "ORIGINATOR", 1, 0)
n = length(unique(winters$Orderid))
# create table with new columns based on media channels
winters %>% group_by(Groupname) %>%
  summarise(converter = sum(Converter), originator = sum(Originator)) %>%
  mutate(converter_per = round(converter / n,4) * 100,
         originator_per = round(originator/ n,4) * 100) %>%
  select(Groupname, converter, originator, converter_per, originator_per) -> d
# add row of column totals
d[14,] = c("Total", colSums(d[,2:5]))
kable(d)
```

Groupname	converter	originator	converter_per	originator_per
BUZZ AFFILIATE	443	193	29.18	12.71
CJ	229	80	15.09	5.27
CPM	771	567	50.79	37.35
OTHER	4	17	0.26	1.12
PRINT - MAGAZINES	4	3	0.26	0.2
SEARCH GOOGLE BRAND	0	470	0	30.96
SEARCH GOOGLE NON-BRAND	28	41	1.84	2.7
SEARCH MSN BRAND	0	117	0	7.71
SEARCH MSN NON-BRAND	4	6	0.26	0.4
SEARCH YAHOO BRAND	0	1	0	0.07
Social	0	1	0	0.07
TV	15	15	0.99	0.99
Uncategorized	20	7	1.32	0.46
Total	1518	1518	99.99	100.01

b. Share of credit for the first- and last-touch attribution models

```
d %>% select(Groupname, converter_per, originator_per) -> d1
d1 <- d1[1:13,]
dfa <- melt(d1, id.vars="Groupname");

ggplot(dfa) + geom_col(aes(x=Groupname, y=as.numeric(value),
                                       fill = variable), position='dodge') +
  labs(y="Share of Credit (%)", x="Marketing Channel", fill = "Attribution Model") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(labels = c("Last-touch", "First-touch"),values = c("blue", "red"))
```



c. Interpretation

There appears to be a wider distribution spread among the marketing channels for the first-touch attribution model compared to the last-touch model. The last-touch model is dominated by Buzz Affiliate, CJ, and CPM, whereas, the first-touch model's main channels are CPM and branded search on Google.

Allocating marketing budget only on the basis of the last-touch attribution model doesn't account for other prominent marketing channels a customer interacted with. This attribution model can be detrimental for Winters as it will not take budget on search-based ads, particularly branded search on Google, into consideration. Following a last-touch model for marketing budget allocation means a very concentrated investment in Buzz Affiliate, CJ and CPM. This attribution model would miss out on the diversified investment in more marketing channels that a first-touch allocation approach would take.

2. New Customers vs. Old Customers

a. Days to convert

```
winters %>% filter(Originator == 1) -> org
org %>% filter(Newcustomer == "Y") %>% summarise(mean = mean(DaysToConvert)) -> newC
org %>% filter(Newcustomer == "N") %>% summarise(mean = mean(DaysToConvert)) -> oldC

print(paste0("Average number of days for a new customer to convert: ", round(newC,0), " days"))
```

```
## [1] "Average number of days for a new customer to convert: 6 days"
```

```r
print(paste0("Average number of days for an old customer to convert: ", round(oldC,0), " days"))
```

```
## [1] "Average number of days for an old customer to convert: 29 days"
```

b. Average touchpoints

```r
winters %>% group_by(Orderid) %>%
  select(Orderid, Position, Originator, Newcustomer) %>% add_count() %>% filter(Originator == 1, Newcus
winters %>% group_by(Orderid) %>%
  select(Orderid, Position, Originator, Newcustomer) %>% add_count() %>% filter(Originator == 1, Newcus

avg_tp_n = round(mean(tp_n$n),2)
avg_tp_o = round(mean(tp_o$n),2)

print(paste0("Average number of touchpoints for a new customer: ", avg_tp_n, " touches"))
```

```
## [1] "Average number of touchpoints for a new customer: 4.25 touches"
```

```r
print(paste0("Average number of touchpoints for an old customer: ", avg_tp_o, " touches"))
```

```
## [1] "Average number of touchpoints for an old customer: 5.18 touches"
```

c. Average order sales

```r
org %>% filter(Newcustomer == "Y") %>% summarise(mean = mean(Saleamount)) -> O_newC
org %>% filter(Newcustomer == "N") %>% summarise(mean = mean(Saleamount)) -> O_oldC

print(paste0("Average order sales for new customers: $", round(O_newC,2)))
```

```
## [1] "Average order sales for new customers: $271.95"
```

```r
print(paste0("Average order sales for old customers: $", round(O_oldC,2)))
```

```
## [1] "Average order sales for old customers: $207.41"
```

d. Summary of the three variables

New customers have a lower average touchpoint count and they take significantly fewer days to convert. The average sales for a new customer is also higher than an old customer by more than $60.

3. Revenue per marketing channel based on first-touch attribution

a. Average sales and total revenue

```r
winters %>% group_by(Groupname) %>% filter(Positionname == "ORIGINATOR") %>%
  summarise(avg_sales = round(mean(Saleamount),2), revenue = sum(Saleamount)) %>%
  select(Groupname, avg_sales, revenue) -> avg_r
kable(avg_r)
```

Groupname	avg_sales	revenue
BUZZ AFFILIATE	258.05	49803.90
CJ	262.38	20990.32
CPM	242.27	137369.32
OTHER	227.47	3867.04
PRINT - MAGAZINES	324.65	973.96
SEARCH GOOGLE BRAND	250.35	117664.16
SEARCH GOOGLE NON-BRAND	234.58	9617.85
SEARCH MSN BRAND	229.09	26803.53
SEARCH MSN NON-BRAND	274.90	1649.40
SEARCH YAHOO BRAND	258.49	258.49
Social	165.00	165.00
TV	239.25	3588.81
Uncategorized	200.42	1402.91

b. Total incremental gross revenue

```r
# gross margin = 40%, incr_fact_brand = 5%, incr_fact_rem = 10%

winters %>% group_by(Groupname, Brand) %>%
  filter(Positionname == "ORIGINATOR") %>%
  mutate(gross_rev = ifelse(Brand == "Y", Saleamount*0.4*0.05, Saleamount*0.4*0.1)) %>%
  summarise(gross_margin = round(sum(gross_rev),2)) -> b3

b3 %>% filter(Brand == "Y") -> branded
branded[5,] = c("Total","NA", colSums(branded[,3]))

kable(b3)
```

Groupname	Brand	gross_margin
BUZZ AFFILIATE	N	1992.16
CJ	N	839.61
CPM	N	5494.77
OTHER	N	25.21
OTHER	NULL	27.09
OTHER	Y	51.19
PRINT - MAGAZINES	N	38.96
SEARCH GOOGLE BRAND	Y	2353.28
SEARCH GOOGLE NON-BRAND	N	384.71
SEARCH MSN BRAND	Y	536.07
SEARCH MSN NON-BRAND	N	65.98
SEARCH YAHOO BRAND	Y	5.17
Social	N	6.60
TV	NULL	143.55
Uncategorized	N	56.12

```
kable(branded)
```

Groupname	Brand	gross_margin
OTHER	Y	51.19
SEARCH GOOGLE BRAND	Y	2353.28
SEARCH MSN BRAND	Y	536.07
SEARCH YAHOO BRAND	Y	5.17
Total	NA	2945.71

```
print(paste0("The total incremental gross revenue accuring to Winters by originator channel of branded s
```

```
## [1] "The total incremental gross revenue accuring to Winters by originator channel of branded search
```

The total incremental gross revenue accuring to Winters by originator channel of branded search is $2,945.71

c. Advice for the search team

According to the calculation above, the branded search adversiting cost is significantly higher than the incremental gross revenue (by first-touch attribution). The cost is $4,200 while the revenue is only $2,945.71. The search ad team should avoid spending this much on branded search ads as it is exceeding the revenue by a significant amount. However, it is important to not fully neglect branded search ads as we need to keep up with our competitors.

4. Linear/uniform attribution

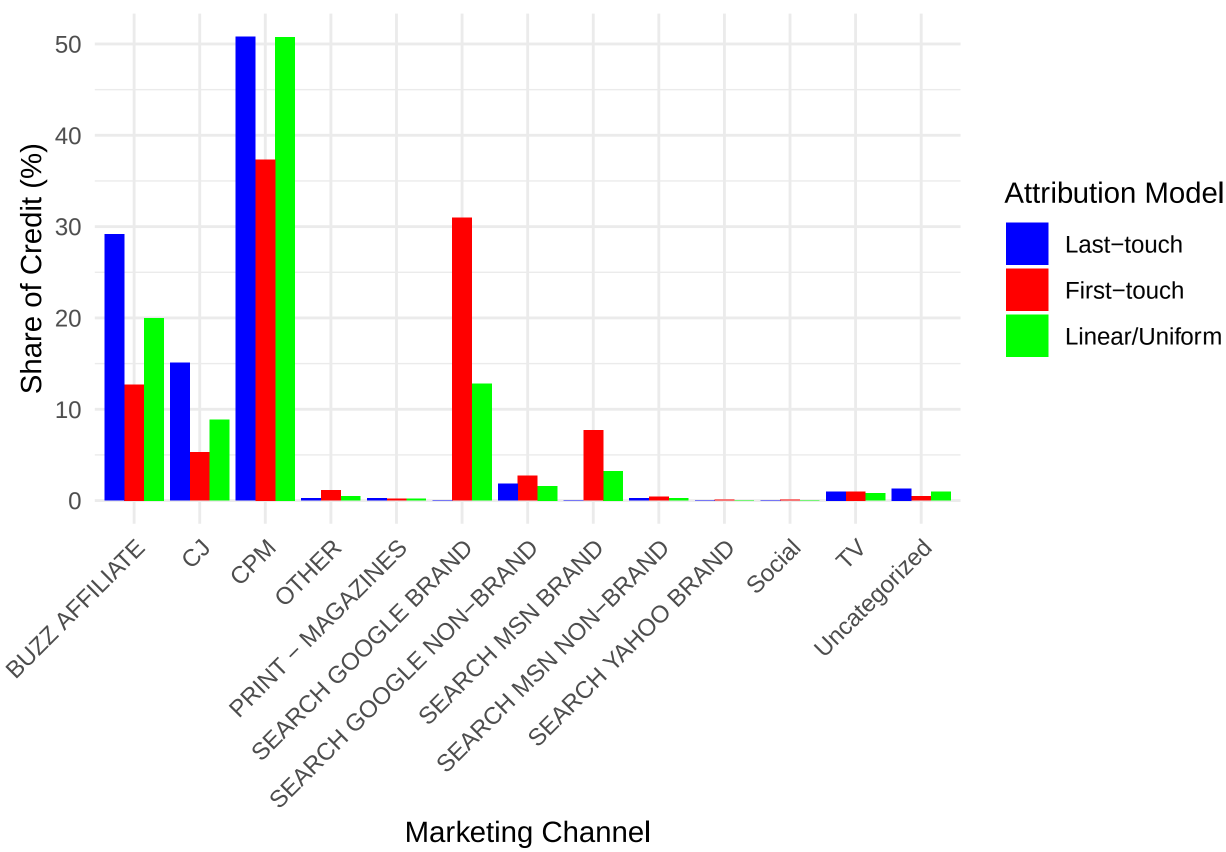
a.

```
winters %>% group_by(Orderid) %>%
  select(Orderid, Position, Originator, Newcustomer, Groupname) %>% add_count() %>%  mutate(LinearAttri

a4 %>% group_by(Groupname) %>%
  summarise(total_share = round(sum(LinearAttributionShare),2)) %>% arrange(desc(total_share)) -> done
done %>% mutate(credit_share = round(as.numeric(total_share)/1518,4)*100) -> k
k[14,] = c("Total", colSums(k[,2:3]))
kable(k)
```

Groupname	total_share	credit_share
CPM	770.76	50.77
BUZZ AFFILIATE	303.48	19.99
SEARCH GOOGLE BRAND	194.07	12.78
CJ	134.68	8.87
SEARCH MSN BRAND	49.12	3.24
SEARCH GOOGLE NON-BRAND	24.09	1.59
Uncategorized	14.58	0.96
TV	12.4	0.82
OTHER	7	0.46
SEARCH MSN NON-BRAND	4.11	0.27
PRINT - MAGAZINES	2.93	0.19
SEARCH YAHOO BRAND	0.43	0.03
Social	0.38	0.03
Total	1518.03	100

b. Plot of all 3 attribution models

```
df2 <- k[1:13,]
df2 %>% select(Groupname, credit_share) -> df2
dfb <- melt(df2, id.vars="Groupname")
comb <- rbind(dfa,dfb)

ggplot(comb) + geom_col(aes(x=Groupname, y=as.numeric(value),
                                        fill = variable), position='dodge') +
  labs(y="Share of Credit (%)", x="Marketing Channel", fill = "Attribution Model") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(labels = c("Last-touch", "First-touch","Linear/Uniform"),values = c("blue", "red","
```



c. Interpretation of the different attribution models

The linear and first-touch models have a more spread out distribution of credit to different marketing channels compared to the last-touch model. However, the linear model appears to share similarities with both of the other models. The last-touch and linear models both give the most credit Buss Affiliate and CPM. On the other hand, the linear and first-touch models both give some credit to branded search ads while the last-touch model gives none. While each one of the three attribution models have their differences, there is no denying of the dominant share of CPM.

5. Intermediate (Roster and Assist) touch points

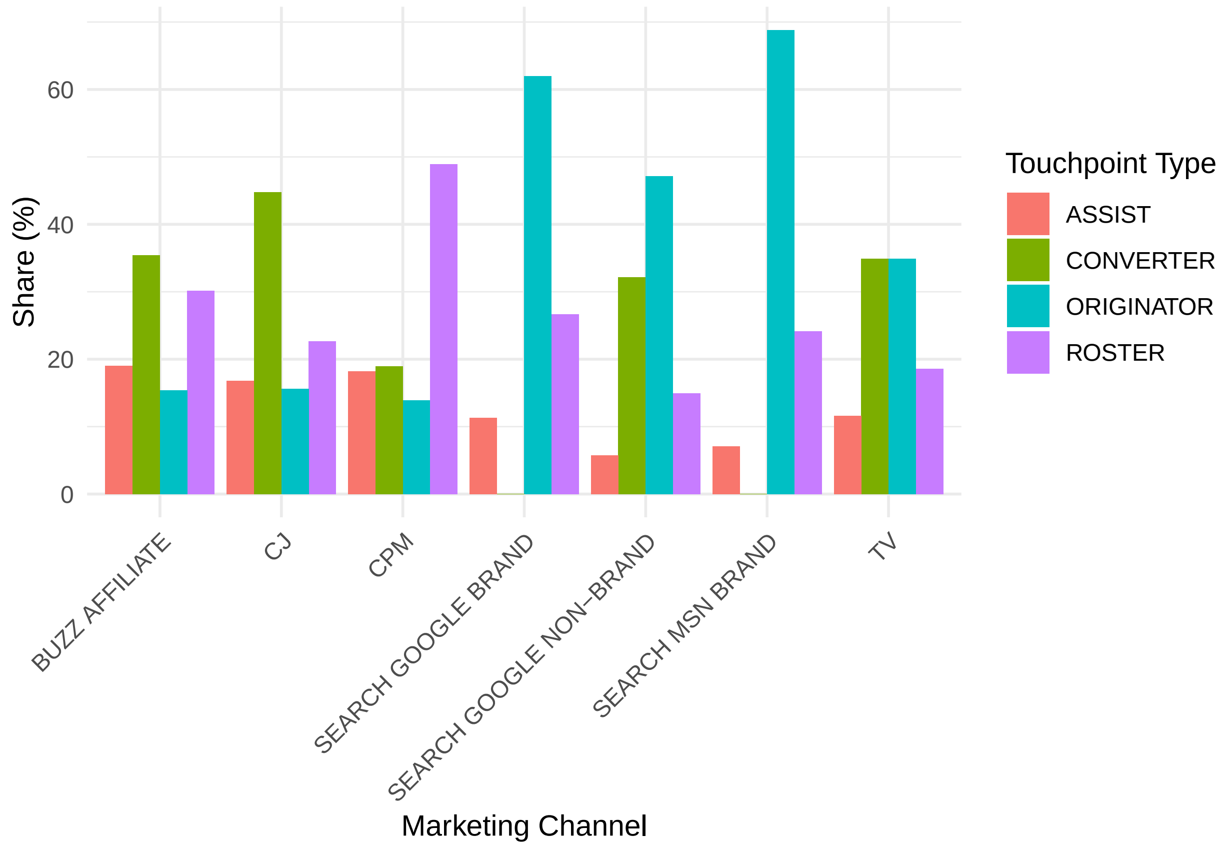
a.

```
prop <- as.data.frame(with(winters, table(Groupname, Positionname)) %>% prop.table(margin=1)*100) %>% fi

prop %>% pivot_wider(names_from = Positionname, values_from = Freq) %>%
  arrange(Groupname) %>% select(Groupname, ORIGINATOR, ROSTER, ASSIST, CONVERTER) %>%
  mutate(Total = (ORIGINATOR+ROSTER+ASSIST+CONVERTER)) -> a5
kable(a5)
```

Groupname	ORIGINATOR	ROSTER	ASSIST	CONVERTER	Total
BUZZ AFFILIATE	15.42766	30.13589	19.024780	35.41167	100
CJ	15.65558	22.70059	16.829746	44.81409	100
CPM	13.92778	48.90690	18.226480	18.93884	100
SEARCH GOOGLE BRAND	62.00528	26.64908	11.345646	0.00000	100
SEARCH GOOGLE NON-BRAND	47.12644	14.94253	5.747126	32.18391	100
SEARCH MSN BRAND	68.82353	24.11765	7.058823	0.00000	100
TV	34.88372	18.60465	11.627907	34.88372	100

b.

```
ggplot(prop) + geom_col(aes(x=Groupname, y=as.numeric(Freq),
                                          fill = Positionname), position='dodge') +
  labs(y="Share (%)", x="Marketing Channel", fill = "Touchpoint Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



c.

CPM, which is display adversiting, has the highest proportion of roster touchpoints relative to all its other touchpoint types, while Google's non-branded search has the lowest proportion of touchpoints dedicated to the roster. CPM and Buzz Affiliate have the most assist touchpoints compared to all other channels. If you look at assist and roster together, CPM and Buzz Affiliate again seems to be getting the most credit from these two types of touchpoints.

The branded search ads receive zero credit under the last-touch model, which may be an issue since branded search ads can play a very crucial role in intent and awareness where the customers search with the sole purpose of making the purchase or gaining the extra confidence to make the purchase. CPM has an extremely high share of roster credit while getting very low credit from last-touch attribution. This may be an issue since display ads play a fundamental role in creating brand awareness that may just be the bridge to get to a roster touchpoint or the difference-maker between an assist and a conversion. From a first-touch perspective CPM, CJ, and Buzz Affiliate all seem to be receiving relatively low credit compared to all the search and TV ads, which may be misleading since these marketing channels can be influential forces in raising brand awareness and exposure.