



**MCDA 5520**  
**STATISTICS & BUSINESS ANALYTICS**

**Project Case –**  
**Vaycaynation: Driving website traffic through second-screen analytics**

**Team Members**

Names	A#
Bhattacharjee, Juthika	A00442303
Chacko, Jelson	A00446838
Chauhan, Raj Bhupendrakumar	A00447694
Deshmukh, Kishor P.	A00439253
Ganesan Kabaleeswaran, Sivakarhikeyan	A00442434

**Q1: Which of the spots has most contributed to increase in website visits?**

Answer:

The lifts for the 2-minute and the 20-minute mark/window have been calculated using the following formula:

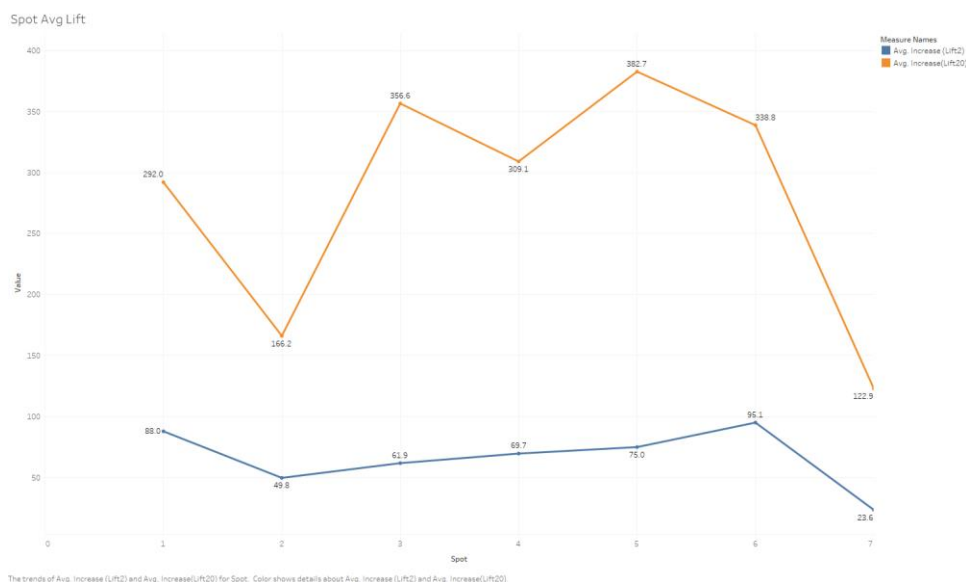
- $\text{Lift2} = \text{AfterVisits2} - \text{BeforeVisits2}$
- $\text{Lift20} = \text{AfterVisits20} - \text{BeforeVisits20}$

The AfterVisits2, BeforeVisits2, AfterVisits20 as well BeforeVisits20 data columns are used from the excel sheet given. Hence, Lift2 and Lift20 are the derived columns/features.

This is needed as the 20-minute window should also be analyzed for Weber to come up with the most contributive spot for the lifts. The average lifts have been calculated per spot for the 2 marks/windows (2-minute and 20-minute).

	Avg Lift2	Avg Lift20	Avg
Spot 1	87.97	292.03	190.00
Spot 2	49.82	166.18	108.00
Spot 3	61.89	356.56	209.22
Spot 4	69.74	309.05	189.39
Spot 5	75.04	382.65	228.85
Spot 6	95.11	338.78	216.94
Spot 7	23.57	122.86	73.21

For the 2-minute mark, it can be concluded that Spot 6 has the highest average lifts, and for the 20-minute mark, Spot 5 accounts to the highest mean lifts.



Hence, to conclude that which spot is the most contributive, we must consider both windows.

Overall, from the above graph, the mean lift in Vaycaynation's website visits is the highest for Spot 5, meaning choosing Spot 5 for airing the commercial will yield company, the highest increase in visits.

**Q2: How important is content (i.e., spots 1-7) in explaining website visits relative to placement and timing factors? In addition to the variables in the data set, this analysis should also include the day of the week and whether an ad aired during prime time (i.e., between 8 and 10 pm, or 20 to 22 h).**

Answer:

This question focuses on the spots, the time slots, and the placement variables for determining the lifts in visits, which clearly indicates that the focus is on the day of the week and the time slot (e.g., 8pm – 10pm is the prime-time slot) for the airing of the commercial, too. These two variables were obtained by modifying the StartSp and EndSp columns in excel.

As we can see, since the effect of timing and placement variables on the lifts in visits is to be analysed, multiple regression technique is employed for this purpose.

Clearly,

The independent variables used are:

- Spot # (Spot 1-7)
- Day of the week (Mon – Sun)
- Prime Time Slot (1 = Prime Time airing slot, 0 = Non-prime Time airing slot)
- Position % (= Position / Total\_Positions) [higher the %, later is the airing done]
- Channel (Vox, Kabel1, RTL)
- Genre (Thriller, Sitcom, Basketball, etc.)

Note: Prime Time Slot is 1 if the airing is done after 8pm. Also, the categorical variables are one-hot encoded, so Spot# is split into 6, Day of the week into 6, Prime Time Slot into 1, Channel into 2, and Genre into 19 one-hot encoded columns.

The sample snapshot of how the data looks like is here:

Spot 5	Spot 6	Monday	Tuesday	Kabel1	VOX	Adventure	Basketbal	Document	News
1	0	1	0	1	0	1	0	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0

The dependent/target variables are the lifts in website visits for 2-min as well as the 20-min window. So, we perform multiple regression for the lifts in 2-min window and then for the lifts in the 20-min window.

R Studio is used to come up with the regression analysis results.

This is the regression analysis result for the lifts in visits for the 2-minute window.

```

Coefficients:
(Intercept)      -51.8049    56.2187   -0.921 0.359344
i..Spot1         36.0957    18.9097    1.909 0.059579 .
Spot2            23.7039    19.7981    1.197 0.234451
Spot3            32.1758    22.2362    1.447 0.151490
Spot4            13.8755    18.6911    0.742 0.459869
Spot5            25.6453    19.3676    1.324 0.188927
Spot6            53.0730    22.4383    2.365 0.020239 *
Monday           -67.5403    27.6254   -2.445 0.016507 *
Tuesday          -50.8394    28.0763   -1.811 0.073631 .
Wednesday        -60.9920    31.7261   -1.922 0.057821 .
Thursday         -65.4596    26.5804   -2.463 0.015760 *
Friday           -68.0373    28.4332   -2.393 0.018868 *
Saturday         -17.6807    27.0702   -0.653 0.515386
PrimeNo          48.1567    21.2423    2.267 0.025868 *
Kabel1           54.7369    22.4704    2.436 0.016892 *
VOX              81.0245    23.5417    3.442 0.000890 ***
PositionPercent  -0.5405    16.4317   -0.033 0.973834
Adventure        73.2783    59.0521    1.241 0.217975
Basketball       87.5196    42.7115    2.049 0.043466 *
Documentry       89.3535    46.2581    1.932 0.056660 .
News            72.1925    46.5832    1.550 0.124831
Romance          186.3628    57.0767    3.265 0.001566 **
Science          69.0602    48.3919    1.427 0.157130
Thriller         110.7826    45.9980    2.408 0.018132 *
Travel           81.0632    52.9463    1.531 0.129386
Cartoon          98.0080    61.7551    1.587 0.116131
Humor            86.6126    53.5588    1.617 0.109466
Magazine         113.8333    53.3467    2.134 0.035670 *
OtherMovies      72.1723    54.0569    1.335 0.185321
ScienceFiction   107.3130    50.6894    2.117 0.037109 *
SitCom           111.0714    49.2524    2.255 0.026634 *
Soap             80.0271    47.2516    1.694 0.093911 .
Soccer           183.9275    53.0699    3.466 0.000823 ***
Politics         115.8183    50.2355    2.306 0.023517 *
Nature           105.0681    61.3506    1.713 0.090351 .
Family           115.8471    53.0668    2.183 0.031726 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.24 on 87 degrees of freedom
Multiple R-squared:  0.6551,    Adjusted R-squared:  0.5164
F-statistic: 4.722 on 35 and 87 DF,  p-value: 2.163e-09

```

Clearly, the R-squared value for the combination of the independent variables is 0.6551. This means 65.51% of the variation in lifts in visits (2-min window) is explained by the independent variables. This also indicates the combined effect of the content, timing and placement columns does affects the increase in visits.

Even for the 20-min window (shown below), it is evident that 52.88% of the variation is brought about by the overall combination of the independent variables.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-105.461	285.104	-0.370	0.7124
i..Spot1	76.360	95.897	0.796	0.4280
Spot2	88.600	100.403	0.882	0.3800
Spot3	203.110	112.767	1.801	0.0751 .
Spot4	70.844	94.789	0.747	0.4568
Spot5	171.267	98.220	1.744	0.0847 .
Spot6	188.565	113.792	1.657	0.1011
Monday	9.826	140.098	0.070	0.9442
Tuesday	-40.308	142.385	-0.283	0.7778
Wednesday	-202.770	160.894	-1.260	0.2109
Thursday	-146.314	134.798	-1.085	0.2807
Friday	-153.036	144.195	-1.061	0.2915
Saturday	31.463	137.282	0.229	0.8193
PrimeNo	39.216	107.727	0.364	0.7167
Kabell	279.798	113.955	2.455	0.0161 *
VOX	293.735	119.388	2.460	0.0159 *
PositionPercent	-130.610	83.331	-1.567	0.1207
Adventure	453.759	299.473	1.515	0.1333
Basketball	278.737	216.604	1.287	0.2016
Documentry	160.802	234.590	0.685	0.4949
News	224.081	236.239	0.949	0.3455
Romance	367.625	289.456	1.270	0.2075
Science	111.512	245.412	0.454	0.6507
Thriller	211.660	233.271	0.907	0.3667
Travel	176.278	268.509	0.657	0.5132
Cartoon	229.880	313.181	0.734	0.4649
Humor	129.881	271.615	0.478	0.6337
Magazine	195.377	270.539	0.722	0.4721
OtherMovies	190.260	274.141	0.694	0.4895
ScienceFiction	327.837	257.063	1.275	0.2056
SitCom	230.670	249.776	0.924	0.3583
Soap	177.933	239.629	0.743	0.4598
Soccer	486.022	269.136	1.806	0.0744 .
Politics	247.168	254.761	0.970	0.3346
Nature	411.715	311.129	1.323	0.1892
Family	558.455	269.120	2.075	0.0409 *
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 193.9 on 87 degrees of freedom				
Multiple R-squared: 0.5288, Adjusted R-squared: 0.3393				
F-statistic: 2.79 on 35 and 87 DF, p-value: 5.891e-05				

It can be said that channel and genre variables are more contributive compared to the timing variable, which can be seen from the \*s indicated in the above figures.

Thus, the day of the week, the prime/non-prime slot of the day, spot #, channel and the genres together are extremely important and contribute to the overall increase in visits (R-squared values indicated above).

**Q3: Which configuration (in terms of content, placement, and timing) is the best to recommend for the summer campaign? In addition, what lifts in webpage visits and sales can be expected from that configuration?**

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-51.8049	56.2187	-0.921	0.359344
i..Spot1	36.0957	18.9097	1.909	0.059579 .
Spot2	23.7039	19.7981	1.197	0.234451
Spot3	32.1758	22.2362	1.447	0.151490
Spot4	13.8755	18.6911	0.742	0.459869
Spot5	25.6453	19.3676	1.324	0.188927
Spot6	53.0730	22.4383	2.365	0.020239 *
Monday	-67.5403	27.6254	-2.445	0.016507 *
Tuesday	-50.8394	28.0763	-1.811	0.073631 .
Wednesday	-60.9920	31.7261	-1.922	0.057821 .
Thursday	-65.4596	26.5804	-2.463	0.015760 *
Friday	-68.0373	28.4332	-2.393	0.018868 *
Saturday	-17.6807	27.0702	-0.653	0.515386
PrimeNo	48.1567	21.2423	2.267	0.025868 *
Kabel1	54.7369	22.4704	2.436	0.016892 *
VOX	81.0245	23.5417	3.442	0.000890 ***
PositionPercent	-0.5405	16.4317	-0.033	0.973834
Adventure	73.2783	59.0521	1.241	0.217975
Basketball	87.5196	42.7115	2.049	0.043466 *
Documentry	89.3535	46.2581	1.932	0.056660 .
News	72.1925	46.5832	1.550	0.124831
Romance	186.3628	57.0767	3.265	0.001566 **
Science	69.0602	48.3919	1.427	0.157130
Thriller	110.7826	45.9980	2.408	0.018132 *
Travel	81.0632	52.9463	1.531	0.129386
Cartoon	98.0080	61.7551	1.587	0.116131
Humor	86.6126	53.5588	1.617	0.109466
Magazine	113.8333	53.3467	2.134	0.035670 *
OtherMovies	72.1723	54.0569	1.335	0.185321
ScienceFiction	107.3130	50.6894	2.117	0.037109 *
SitCom	111.0714	49.2524	2.255	0.026634 *
Soap	80.0271	47.2516	1.694	0.093911 .
Soccer	183.9275	53.0699	3.466	0.000823 ***
Politics	115.8183	50.2355	2.306	0.023517 *
Nature	105.0681	61.3506	1.713	0.090351 .
Family	115.8471	53.0668	2.183	0.031726 *
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 38.24 on 87 degrees of freedom				
Multiple R-squared: 0.6551, Adjusted R-squared: 0.5164				
F-statistic: 4.722 on 35 and 87 DF, p-value: 2.163e-09				

2-minute mark/window

The multiple regression equation (y-value) for the 2 minutes window is obtained from the independent variables spot, day of the week, genre, position(%) and the prime time (0800 hrs to 0900 hrs) is shown below.

$$\begin{aligned}
 Y_{Lift2} = & -51.80 + 36.10(Spot1) + 23.70(Spot2) + 32.18(Spot3) + 13.88(Spot4) \\
 & + 25.65(Spot5) + 53.07(Spot6) - 67.54(Monday) - 50.84(Tuesday) \\
 & - 60.99(Wednesday) - 65.46(Thursday) - 68.04(Friday) - 17.68(Saturday) \\
 & + 48.16(PrimeNo) + 54.77(Kabel1) + 81.02(VOX) - 0.54(PositionPercent) \\
 & + 73.28(Adventure) + 87.52(Basketball) + 89.35(Documentry) + 72.19(News) \\
 & + 186.36(Romance) + 69.06(Science) + 110.78(Thriller) + 81.06(Travel) \\
 & + 98.01(Cartoon) + 86.61(Humor) + 113.83(Magazine) \\
 & + 72.17(OtherMovies) + 107.31(ScienceFiction) + 111.07(SitCom) \\
 & + 80.03(Soap) + 183.93(Soccer) + 115.82(Politics) + 105.07(Nature) \\
 & + 115.85(Family)
 \end{aligned}$$

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-105.461	285.104	-0.370	0.7124
i..Spot1	76.360	95.897	0.796	0.4280
Spot2	88.600	100.403	0.882	0.3800
Spot3	203.110	112.767	1.801	0.0751 .
Spot4	70.844	94.789	0.747	0.4568
Spot5	171.267	98.220	1.744	0.0847 .
Spot6	188.565	113.792	1.657	0.1011
Monday	9.826	140.098	0.070	0.9442
Tuesday	-40.308	142.385	-0.283	0.7778
Wednesday	-202.770	160.894	-1.260	0.2109
Thursday	-146.314	134.798	-1.085	0.2807
Friday	-153.036	144.195	-1.061	0.2915
Saturday	31.463	137.282	0.229	0.8193
PrimeNo	39.216	107.727	0.364	0.7167
Kabel1	279.798	113.955	2.455	0.0161 *
VOX	293.735	119.388	2.460	0.0159 *
PositionPercent	-130.610	83.331	-1.567	0.1207
Adventure	453.759	299.473	1.515	0.1333
Basketball	278.737	216.604	1.287	0.2016
Documentry	160.802	234.590	0.685	0.4949
News	224.081	236.239	0.949	0.3455
Romance	367.625	289.456	1.270	0.2075
Science	111.512	245.412	0.454	0.6507
Thriller	211.660	233.271	0.907	0.3667
Travel	176.278	268.509	0.657	0.5132
Cartoon	229.880	313.181	0.734	0.4649
Humor	129.881	271.615	0.478	0.6337
Magazine	195.377	270.539	0.722	0.4721
OtherMovies	190.260	274.141	0.694	0.4895
ScienceFiction	327.837	257.063	1.275	0.2056
SitCom	230.670	249.776	0.924	0.3583
Soap	177.933	239.629	0.743	0.4598
Soccer	486.022	269.136	1.806	0.0744 .
Politics	247.168	254.761	0.970	0.3346
Nature	411.715	311.129	1.323	0.1892
Family	558.455	269.120	2.075	0.0409 *
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 193.9 on 87 degrees of freedom				
Multiple R-squared: 0.5288, Adjusted R-squared: 0.3393				
F-Statistic: 2.79 on 35 and 87 DF, p-value: 5.891e-05				

### 20-minute mark/window

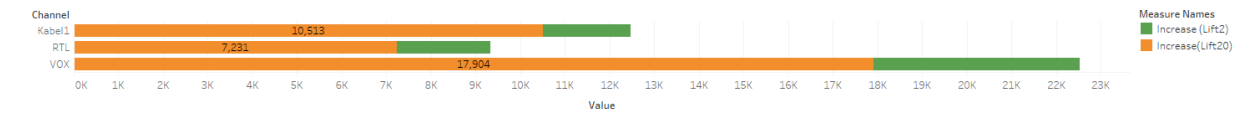
The multiple regression equation (y-value) for the 20 minutes window is obtained from the independent variables spot, day of the week, genre, position(%) and the prime time (20:00 hrs to 22:00 hrs) is shown below.

$$\begin{aligned}
 Y_{Lift20} = & -105.46 + 76.36(Spot1) + 88.60(Spot2) + 203.11(Spot3) + 70.84(Spot4) \\
 & + 171.27(Spot5) + 188.57(Spot6) + 9.83(Monday) - 40.31(Tuesday) \\
 & - 202.77(Wednesday) - 146.31(Thursday) - 153.03(Friday) + 31.46(Saturday) \\
 & + 39.22(PrimeNo) + 279.80(Kabel1) + 293.74(VOX) + 130.61(PositionPer) \\
 & + 453.76(Adventure) + 278.74(Basketball) + 160.80(Document) + 224.08(News) \\
 & + 367.63(Romance) + 111.51(Science) + 211.66(Thriller) + 176.27(Travel) \\
 & + 229.88(Cartoon) + 129.88(Humor) + 195.38(Magazine) + 190.26(OtherMovies) \\
 & + 327.84(ScienceFiction) + 230.67(SitCom) + 177.93(Soap) + 486.02(Soccer) \\
 & + 247.17(Politics) + 411.72(Nature) + 558.46(Family)
 \end{aligned}$$



- Channels / Lifts

Channel / Lifts Horizontal

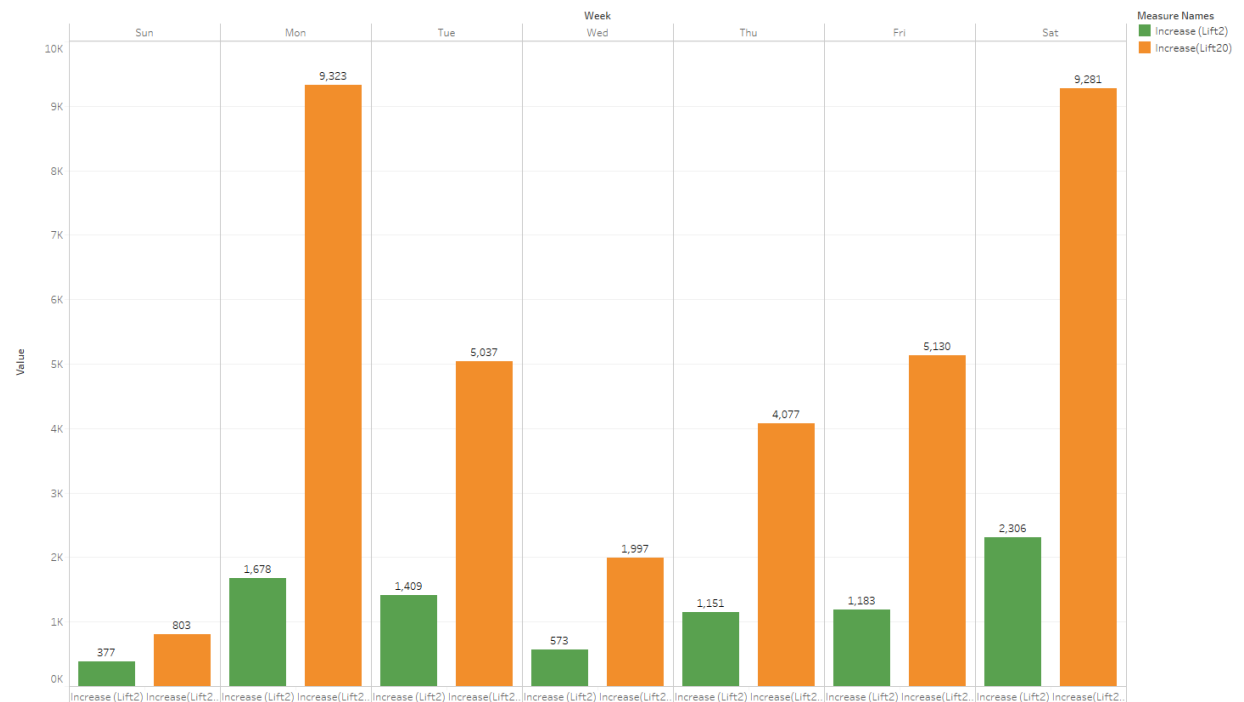


Increase (Lift2) and Increase(Lift20) for each Channel. Color shows details about Increase (Lift2) and Increase(Lift20).

This horizontal stacked bar graph has taken into consideration the Total Lift for the 3 separate channels. It can be inferred from this that the Lift after 20 and 2 minutes is the highest for the advertisements that have been aired on the VOX Channel.

- Week / Lifts

Increase/Week



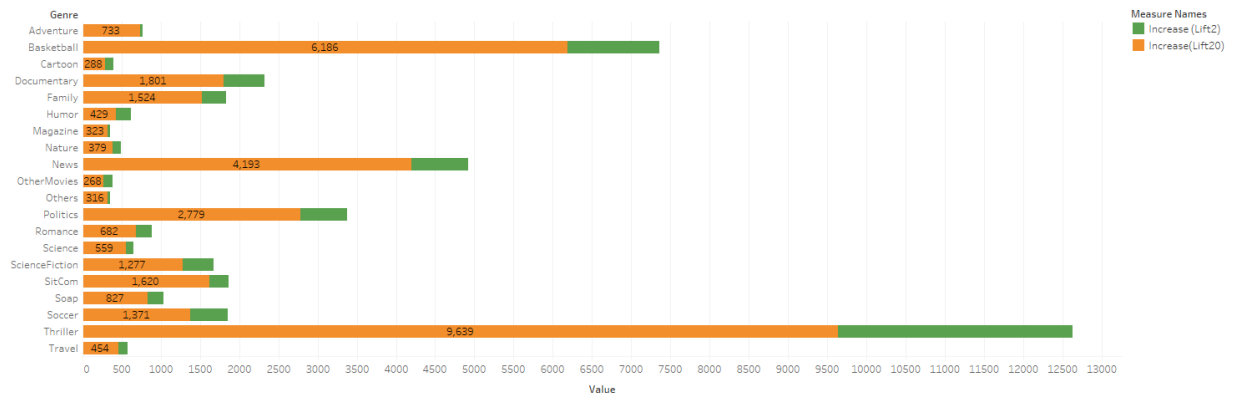
Increase (Lift2) and Increase(Lift20) for each Week. Color shows details about Increase (Lift2) and Increase(Lift20). The view is filtered on Week, which keeps 7 of 7 members.

This clustered bar graph represents the Lifts for both, the 2-minute and the 20-minute window across the days of the week. Though, the peak for the total lifts for the 20-minute mark window is on Monday, the collective maximum lift is the highest on Saturday with a total Lift of over 11,500.



- Genres / Lifts

Genre/Lift Horizontal

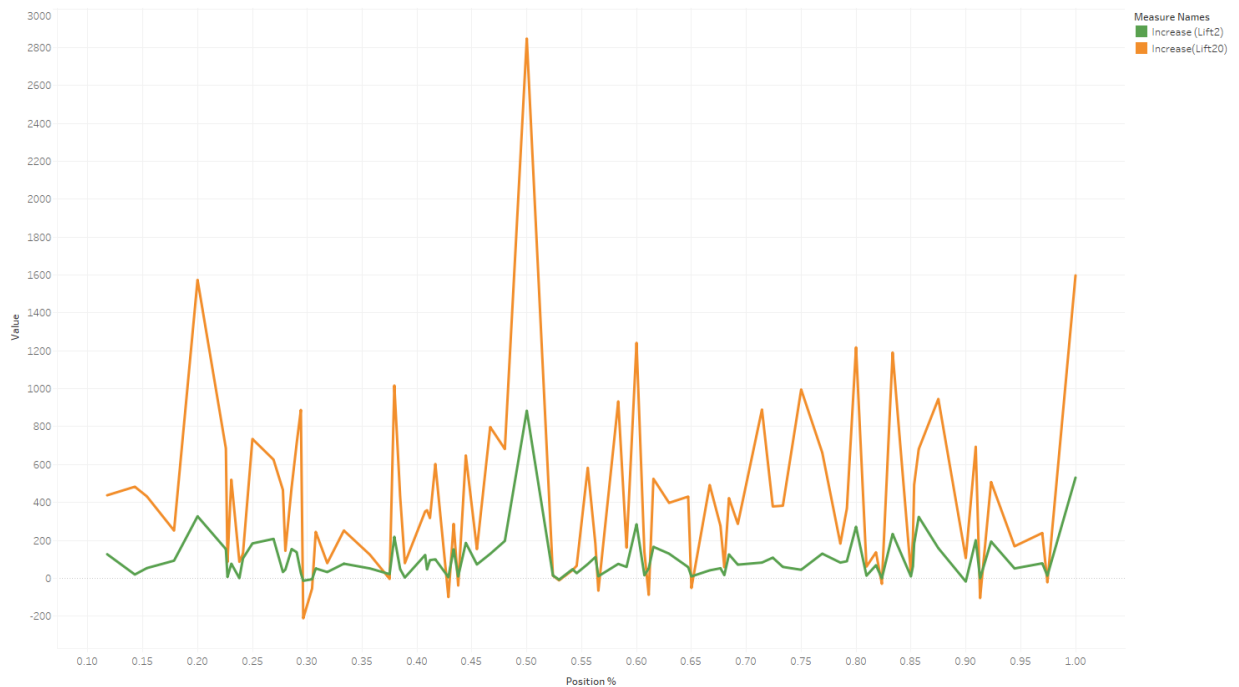


Increase(Lift2) and Increase(Lift20) for each Genre. Color shows details about Increase(Lift2) and Increase(Lift20).

This graph represents the total increment of people who visited the website based on the genre of the show they watch. The increase of viewers for the 20 minutes window is 9,639. Since, the total number of viewers for thriller genre is the highest for both the 2 minutes and 20 minutes window, therefore it would be ideal to put the maximum number of advertisements during the thriller genre. On the contrary, the most non-ideal decision would be to air the advertisement during the shows of cartoon genre. The total number of website visits of 2 minutes and 20 minutes window is the least. So, the advertisement should never be aired in the cartoon genre shows.

- Position / Lift

Position &amp; Lift

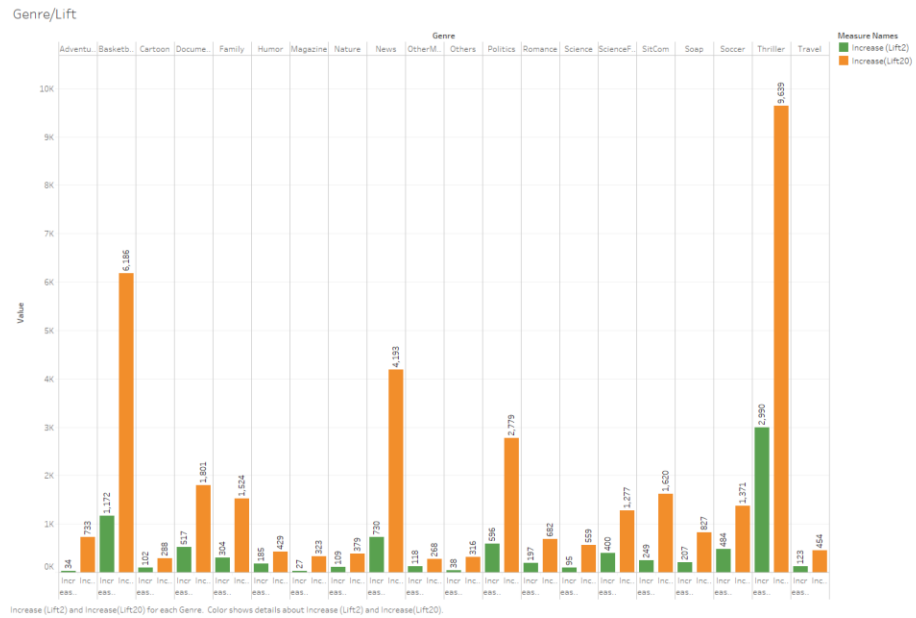


The trends of Increase(Lift2) and Increase(Lift20) for Position %. Color shows details about Increase(Lift2) and Increase(Lift20).

A plot of the total lifts for the 2-minute and the 20-minute window with respect to the position of the advertisements during the intervals is visualized in the line graph above. A spike is at the

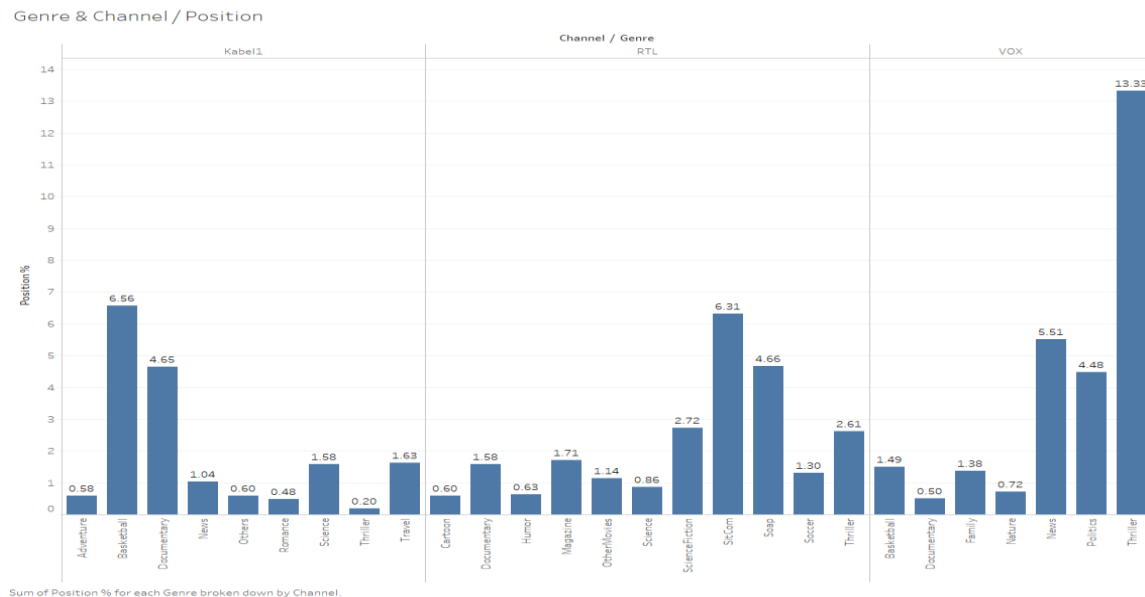
middle position which concludes that an ideal position of the advertisement placed during an interval must be in the middle

- Genre / Lift



For both, the 2-minute window and the 20-minute window, the above clustered vertical bar graph shows that the peak lifts are for the Genre **Thriller**.

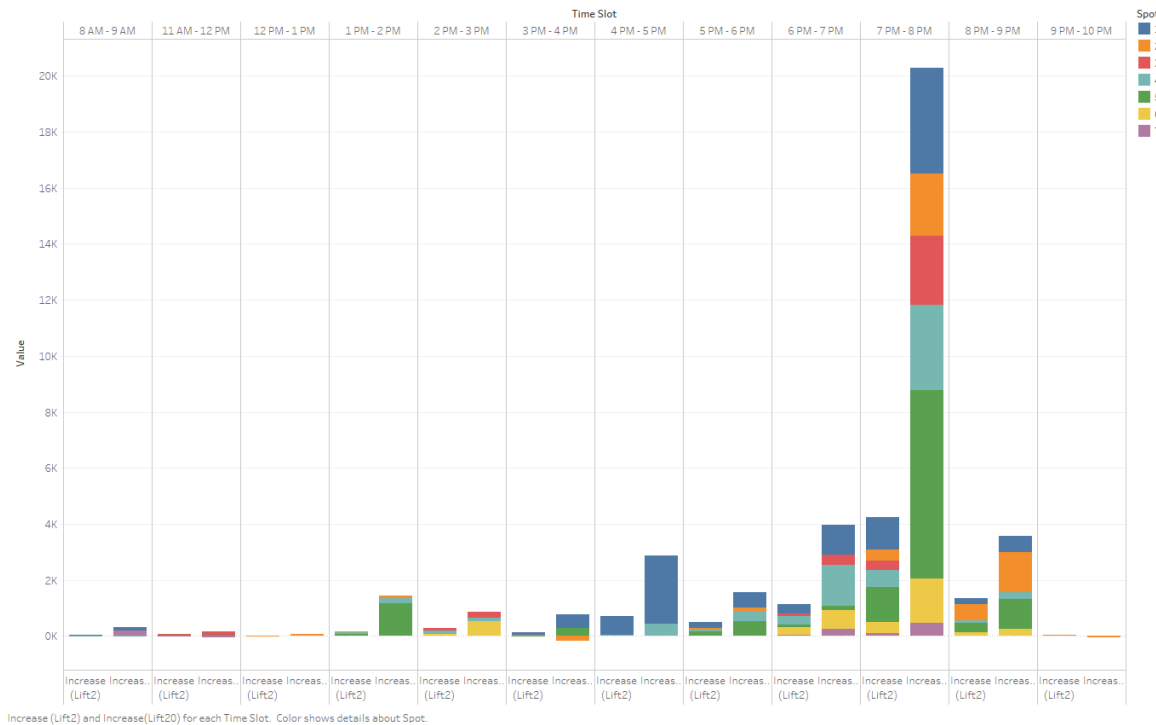
- Genre & Channel / Position



A comparison between all the shows of different genres on different channels along with their position placement has been visualize in this clustered graph. The value of position(%) is the

highest for the thriller genre and the VOX channel. A clear evidence of the thriller genre being the most popular genre can be inferred from this visualization.

Time Slots/Average Position (2)



This graph highlights the peak lifts during the time window from 7PM to 8PM which is outside the considered Prime Time.

For Lift2:

$$Y_{Lift2} = -51.80 + 53.07(Spot\ 6) - 17.68(Saturday) + 81.02(VOX) - 0.54(PositionPercent) + 110.78(Thriller)$$

$$Y_{Lift2} = -51.80 + 53.07 - 17.68 + 54.77 + 81.02 - 0.54(0.5) + 110.78$$

$$Y_{Lift2} = 229.89 \sim 230$$

For Lift20:

$$Y_{Lift20} = -105.46 + 171.27(Spot5) + 9.83(Monday) + 293.74(VOX) + 130.61(PositionPercent) + 211.66(Thriller)$$

$$Y_{Lift20} = -105.46 + 171.27 + 9.83 + 293.74 + 130.61(0.5) + 211.66$$

$$Y_{Lift20} = 646.35 \sim 646$$

Now, it is reasonable to compute the average lifts for both windows -

$$Y_{LiftAvg} = (Y_{Lift2} + Y_{Lift20})/2 = (230+646)/2 = 438$$

As for the increase in sales, we already know that Vaycaynation has already been doing well without even adopting data analysis. However, with the above knowledge of the best possible configuration, the lifts in website visits are likely to increase to 438. It is evident from exhibit 1, the basket value (i. e. the average purchase amount per booking through the company's portal) is €450. Since 438 lifts in visits is now obtained, we will have  $450 \times 438 = \underline{\underline{\text{€197,100 increase in sales}}}$ .

(Note: that the basket value can increase from €450 for summer 2018).

From the entire case study, it can be concluded that Vaycaynation can benefit a lot from employing the data analytics brought about by the newly hired employee – Weber. Weber can make great impact by providing the similar insights as calculated by our team, and thus, there is a great amount of value addition to Meier's/company's existing knowledge about their brand awareness. A recommendation from the above analysis would be to practice the best possible configuration of the content, timing, and placement variables.

---