

Chapter 27

Case Study: InfoMedia Solutions



Deepak Agrawal, Soumithri Mamidipudi, and Sriram Padmanabhan

Hui Zhang had just returned from a workshop on sports and media analytics. One of the speakers had described the convergence of media and how it had affected his broadcast business in a very short time span. Hearing others mention the same set of possibilities and with his own experience in the rapidly changing industry, Zhang was convinced that an ever-increasing number of television viewers will, if they haven't done so already, "cut the cord" and move away from traditional viewing platforms. It was this thought that Zhang had at the back of his mind when he read a report predicting that viewership was splintering—more and more specialized channels were sniping away at traditional shows and showtimes. On top of it, the report mentioned that Internet advertising would overtake television and print media in size and spend in the next 5 years. Zhang was concerned that new technologies would threaten the position that his firm had built up in the TV advertising segment.

While millions of dollars were still expected to be spent on advertising on traditional television and cable channels, changes in viewership habits and demographics would surely shift budgets to targeting different audiences through dedicated channels. Moreover, the bundling strategy followed by cable TV companies allowed audience to quickly surf from one show to another! The change in resource allocation of ad spends had not yet happened because of the ongoing debate

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-319-68837-4_27) contains supplementary material, which is available to authorized users.

D. Agrawal (✉) · S. Mamidipudi
Indian School of Business, Hyderabad, Telangana, India
e-mail: a.deepak@outlook.com

S. Padmanabhan
New York, NY, USA

on tricky issues such as audience measurement, avoiding double counting of viewership, and measuring “reach” (see below for a definition)¹ in cross-platform or cross-channel advertising. These were relevant questions even for traditional media such as newspaper, magazine, radio, and television advertising even without factoring in Internet advertising.

Zhang thought data science might help bring clarity to these questions. Understanding how to use new analytical tools would surely be the key to profitability! As head of advertising strategy for InfoMedia Solutions, Zhang’s job was to identify shifts in the market and technology. In order to understand the ways in which his company could better use the tools at its disposal, Zhang approached Luke Gershwin, InfoMedia’s head of analytics. He asked Gershwin what data analytics tools InfoMedia might use to better target audience on TV. “Without a clear tech-focused strategy, I’m wary that any attempt to change the way our ad business works might fail,” Zhang said.

1 InfoMedia Solutions

InfoMedia’s operations spanned the whole gamut of advertising and press-related services. The company offered a full range of services—they were proud of being able to oversee an advertising campaign from conception to its completion. Zhang’s role was to advise clients and guide marketing teams, as well as, monitor the advertising campaigns to ensure that they were running smoothly. The ever-increasing complexity and interconnectedness of advertising was not lost on Zhang, and he highlighted to his clients that targeting potential customers would not be a simple proposition in the future.

Zhang’s foremost concern was “reach”—the metric every advertiser used to measure their success. Measured as the number of people in a target group who have been exposed to the advertisement at least once, reach was the backbone of every media strategy designed to pull in new customers and impress clients. “The changing context of media,” Gershwin said, “had a very simple, obvious impact that was still playing out: the ways that media companies could reach their potential targets had shifted and multiplied.”

“Ten years ago, nobody would have imagined that watching 10-s videos would be enough to run a business. Now *Snapchat*’s worth nearly \$30 billion,²” Gershwin told Zhang. “You’ve got to target the right market, and you’ve got to do it economically.”

¹A glossary of advertising business terms can be found at <https://www.tubemogul.com/glossary/> or <https://www.bionic-ads.com/2016/03/reach-frequency-ratings-grps-impressions-cpp-and-cpm-in-advertising/> (accessed on Aug 22, 2018). Refer to the article for the meaning of terms, such as Reach, CPM, CPV, Cross-Channel, Daypart, GRP, OTT, PTV, and RTB.

²<http://money.cnn.com/2017/03/02/technology/snapchat-ipo/index.html> (accessed on Jun 23, 2018)

The first issue Zhang faced was the problem that all advertisers had to tackle: duplication. Broadcasting an advertisement (ad) ten times would be very efficient if it were watched by ten different people. If only one person saw all ten times, however, the reach of the ad would become much smaller. In order to measure reach effectively, Zhang would have to correctly account for redundant, duplicate impressions that were inevitable in his line of business. Reach-1 measured the number of times a unique viewer saw the ad at least once.

Zhang also knew that most customers would not change their minds regarding a new product the first time they heard about it. In order to create an image in the customer's mind about a product, Zhang would have to reach the same customer multiple times. Thus, Zhang's duplication issue was akin to a Goldilocks problem—too few impressions would not result in a successful sale, but too many would be wasteful and inefficient. Identifying how to deliver the correct number of impressions in order to maximize reach would be at the core of the solution he needed from Gershwin.

Depending on the target audience and the product, Zhang normally tracked Reach-1, -3, or -5, meaning that an impression had been delivered to a potential customer at least three, five, or seven times. Understanding how many impressions would be necessary for each product was crucial to efficiently use resources. Gershwin's next issue would be to identify the inflection point where duplication was no longer useful to expanding reach.

While traditionally, overlap between media was not high enough to have a large impact on duplication, the need to track duplicate impressions on digital, print, and television was growing more and more important. The cross-platform approach, in which the campaign played out in more than one medium in order to gain impressions, was a burgeoning part of the sector that would surely grow. Gershwin felt that the cross-platform duplication could be addressed later.

2 Reach in Advertising

Like most companies in the ad business, InfoMedia thought of reach not as a function of cost but as a function of gross rating points (GRPs) (Goerg 2014). Advertising campaigns might place advertisements in newspapers, broadcast them on the radio, or show them on the television. Rather than simply look at the amount of money spent on the campaign to determine efficiency, Zhang thought of the number of impressions that the campaign would make and how that would translate into reach.

Zhang calculated GRP (gross rating points) by computing in percentage terms the number of impressions made on a target audience relative to the size of the audience. For example, if an advertisement was seen 120 million times by viewers and the size of the target audience is 60 million, then this translates to $(120/60 \times 100) = 200$ GRPs. Time slots that were longer, slots during peak hours, or in popular shows would be able to create more impressions and therefore higher reach. Nothing came

free; a time slot that was able to deliver more reach would be more expensive. Zhang thus would have to balance the cost of airing slots against the reach that those slots could give him, by understanding his target audience. Reach-1, -3, -5, and -7 were defined either in percentage terms or in the total number of unique viewers who saw the ad at least one, three, five, and seven times. In the above example, these might be 45, 30, 15, and 7 million viewers or 75%, 50%, 25%, and 11.66%. Obviously, reach will not exceed the size of the population. Reach-1 will be the largest followed by the rest.

Reach-1, -3, -5, and -7 were known to be a concave increasing function of ad spots, which is to say that as the number of spots increased, reach increased, but the rate of that increase for the same increase in spots was diminishing. In an ideal world, Zhang would be able to buy exactly the number of spots at which the rate of decrease of reach times “a dollar value for reach” was balanced by the rate of increase in spots times a “dollar value of a spot.” Gershwin thought one could uncover the relationship between ad spots and reach using simulation.

Adding to the complication, a cross-platform approach would need to estimate reach across multiple channels or media. This meant that the function used to calculate reach would have multiple inputs to track.

Guessing the appropriate ad spots target would prove tricky. Gershwin’s solution would have to help solve this problem more efficiently if it were to improve InfoMedia’s bottom line.

3 Rooster Biscuits

Zhang already had a customer in mind for the first test of the new approach. Rooster Confectionery, a biscuit- and cereal-maker, wanted to launch a new brand of chocolate-coated biscuits. Rooster’s chief operating officer, Joni Fernandez, had told Zhang she wanted to focus on a younger target audience (age group 20–35) with the product—exactly the kind of customer who would be shifting to different avenues of media consumption. Zhang felt that if Gershwin’s approach could help to optimize Rooster Biscuits’ campaign, it would bode well for the use of novel techniques.

However, Fernandez had tight programming and budget constraints. She had already informed Zhang that “Rooster expected that the channels would show at least 20% of its ads during ‘prime-time’ slots—between 8 pm and 11 pm and at least 30% over weekends.” Moreover, while she expected a long-term ad campaign, she wanted Zhang to run a short, 1-week version first to test the market’s reaction to the new product. She informed him that the test campaign should target the two biggest cable channels aimed at the 18–34 P (P = person) demographic—The Animation Channel and the Sports 500 Network.

4 Blanc Solutions

Gershwin approached Julien Blanc, founder of data analysis consultancy Blanc Solutions, to understand how to use the new big data techniques. Consultants such as Blanc could quickly evaluate large volumes of data, dig deeper into a problem, and suggest an appropriate solution and data approach. Blanc recommended that InfoMedia use simulation techniques along with historical data to predict reach.

“If your client is planning a new campaign, past campaign performance isn’t going to be the best indicator of future success. Using your knowledge of the market to simulate Reach is a better, cheaper solution,” Blanc said.

Julien thought it is best to develop the ideas based on data from the two cable channels. Zhang was able to obtain viewership numbers for these channels from a third-party aggregator. In order to run robust simulations, Blanc looked for a past period that resembled the week during which InfoMedia planned to run its campaign. The aggregator was able to provide minute-to-minute data regarding viewership for any given demographic for the two channels.

Zhang was also able to leverage the history that he had with his client in order to provide Blanc with past data on pricing and reach for the company’s campaigns. Rooster Biscuits had run an earlier campaign targeting the same demographics, with slightly different parameters. This earlier campaign would provide Blanc with appropriate context and help him make a decision about allocating his slots.

Blanc had found that the Sports 500 Network, which was the more popular of the two channels, tended to charge more for slots compared to the Animation Channel. Moreover, the premiums that Zhang would have to pay for prime-time and weekend slots were higher for the Sports 500 Network.³ He had run a previous campaign with 150 slots, with 100 in the Sports 500 Network and 50 in the Animation Channel. At least 20% of the slots had to be during prime time and 30% during the weekend. This campaign could be a good test subject for the experiment.

Blanc knew that he had three important tasks to perform.

Task 1: Set up a simulation model to estimate the Reach–GRP curve.

How could Blanc simulate reach? Even if he had viewership data that told him how many people had been watching a channel at any given time, estimating how many of those people had seen an ad before even showing the ad looked like a tricky problem. It was here that new research held promise. For example, researchers at Google had identified a method by which the reach versus GRP relationship could be inferred given only a single data point, since the general shape of the curve was “known” (Goerg 2014). Blanc could expand this method to account for the much larger amount of data he was expecting to obtain from his simulations in order to understand the relationship between reach and number of ads shown. He felt when specializing to a specific campaign a functional form such as a polynomial function could be used, in other words, he could simply fit a polynomial curve through several

³See Appendix 1 for details.

(r, g) data points, where r is the reach and g is the number of spots shown. Based on his experience, Zhang suggested to start with quadratic and cubic fits.

Blanc planned to use the viewership data provided by the third-party aggregator to simulate reach and thus obtain the data points. Zhang had told him that InfoMedia intended to air between 10 and 250 ads in the 7-day period (1 week) of the campaign. The data was collected by a survey of households that viewed the two channels, conducted during a week that was as similar as possible to the target week. In the survey, viewers were asked a number of demographic questions. Blanc would be able to obtain the data regarding what channel they viewed, and for how long, from the cable companies themselves. These two sources, when combined, would give Blanc most of the necessary data.

Broadcasters divided airtime on their channels into 6-min slots, that is, Blanc had 1680 potential slots in the week to air the ad—10 per h * 24 h * 7 days per week. Simulation would involve throwing the ads randomly into these 1680 slots and computing GRP and reach. Thus, Blanc would choose the number of ads to air. Then, Blanc would simulate showing ads for each number of slots and use the viewership data to understand how many people were watching the slot chosen by the simulation. He could also estimate whether a viewer watched at least once, at least thrice, etc. Doing this repeatedly would give a “reach versus spots shown” set of data points. Then, he would fit a curve through these points to obtain the relationship.

Yet there was a key element to be added to his dataset. The viewership data he had was only a *sample* of the total population. It was necessary to add a unique weight to each viewer in his set—a measure of the proportion of viewers in the population that were similar to the selected viewer—to convert the sample numbers to the population number. For example, if a viewer in the 18–34 P had watched at least thrice and if this viewer’s weight were 2345, then he would estimate that 2345 viewers had watched at least thrice in the population. Adding these numbers viewer-by-viewer would give an estimate for the GRP and reach. Blanc would thus be able to determine the reach of any combination of slots selected by the simulation by multiplying each viewer who saw the ad by the weight of that viewer. Moreover, by tracking the number of views by the same consumer, multiplying by the weight, and adding up across viewers, he could calculate not just Reach-1 but Reach-3 and Reach-5 as well.

Using multiple simulations,⁴ he would be able to obtain a robust set of data that he could use to derive the reach curve. He can then fit a polynomial curve as explained above. The data science team constructed a simulator that produces the reach given the number of ads to be shown and the constraints on when they are to be shown. The help file, interface, and sample outputs are shown in Appendix 3.

Task 2: Develop the estimate.

⁴Blanc took help from a simulation expert in his data science team who provided him the code “Infomedia_simulation.R” (refer to Appendix 3) to run the simulation to calculate Reach for each simulation given the constraints.

Blanc decided to partition the data set by time and day of the week and to use that information to improve the prediction. In order to do this, he has multiple approaches—divide each day into 3-h buckets, divide regular and prime time on daily basis, or simply divide regular/prime time over weekday and weekend. The time bucket starting at 2 am would ensure that the effect of the “prime-time” 8 pm–11 pm bucket could be understood separately. Moreover, he also decomposed each day into its own bucket, to better understand the difference in viewership between weekdays and weekends. See Appendix 3 for examples of these data collection methods and how these are reflected in the output produced by the simulator. He reviewed the contract terms and legal notes to understand whether there is any “no show”/“blackout” period and found no such restriction for Rooster’s campaign. “No show” or “blackout” period basically restricts the broadcaster from showing commercials when either the customer put the conditions of not broadcasting its ad during a particular time period assuming there will not be target customers or if there are any regulatory restrictions of not showing specific commercials in a particular time zone. Further, in the future, refinements in estimation can be made by using the variance amongst the viewers’ demographics such as age and gender and other variables such as the average time spent by segment.

Task 3: Optimal spending.

Once Zhang had the information at hand and built up confidence in the model, the decision he had to make was clear. Blanc’s own survey had informed him that the overlap between the viewers of the two channels seemed negligible. Thus, Zhang had to determine an optimal spending pattern for Rooster’s campaign. How many ads would he run on each channel? At what times and on which days would he target how many? For demonstration purposes, Blanc thought he could use the previous campaign whose data is shown in Appendix 1 to demonstrate how the new method could work.

He knew he had a difficult task ahead explaining which variables were important in predicting reach. Could he convince his clients about the findings? He thought there were two ways of going about doing this—(a) explain the model very carefully to the client and (b) show how it can be used to increase the reach without increasing the budget.

Electronic Supplementary Material

All the datasets, code, and other material referred in this section are available in www.allaboutanalytics.net.

- Data 27.1: `infomedia_ch1.csv`
- Data 27.2: `infomedia_ch2.csv`
- Code 27.1: `Infomedia_simulation.R`

Exercises: Help Blanc with His Three Tasks

Ex. 27.1 Review the simulator description in Appendix 3 and pseudo-code provided in Appendix 4. Generate data using the simulator for both the channels. Report the model fit as a function of the total number of spots as well as the other explanatory variables. Visualize the reach curve for Reach-1, -3, -5, for both the channels.

Ex. 27.2 Using the campaign information/constraints and the model obtained in Ex. 27.1, demonstrate that a better allocation across channels, weekdays/weekends, and time-of-the-day can yield higher reach. (Maximum allocated budget is \$300,000 for 1 week.) Show optimal allocations for each channel separately, as well as together, for Reach-3. Calculate the total spend for each allocation, based on the pricing details given. Use Reach-3 for your final recommendation.

Ex. 27.3 The advertiser realizes that between 2 am and 5 am, there are very few viewers from its target customer who watch TV and therefore adds the blackout window of 3 h every day. Would this change your analysis?

Ex. 27.4 Due to the increasing demand and limited broadcasting slots, the broadcaster is considering offering dynamic pricing. The broadcaster may redefine prime-time concept and significantly change its media marketing strategy. Suggest a new strategy if the broadcaster moves to dynamic pricing.

Ex. 27.5 How can this approach be used when more and more viewers switch to the Internet?

Ex. 27.6 What would happen if the advertiser demands not to broadcast the commercial alongside other similar commercials? What other practical constraints do you see being imposed on a schedule by an advertiser?

Ex. 27.7 What would happen if the broadcaster repeats the ad within the same slot/commercial break (known as a “pod” in the TV ad industry). Discuss how it may impact viewership of target segment and whether you need to change your strategy.

Appendix 1: Rooster Biscuit Campaign 6,412,965—Campaign GRP and Cost

Base rate	Animation channel	Sports 500 network
Weekday	\$ 1500 (X)	\$ 1200 (Y)

Notes: Base rate corresponds to weekday non-prime-time price for one 30-s spot. Rooster's campaign has an ad of exactly 30 s

Pricing multiples (categorized by spots timing)

Animation channel	Weekday	Weekend
Non-prime time	1X	1.25X
Prime time	1.4X	1.75X

Notes: Price multiples are applied to the base rate to calculate the specific rate for a different time or day of week

Sports 500 network	Weekday	Weekend
Non-prime time	1Y	1.2Y
Prime time	1.25Y	1.5Y

Notes: Price multiples change as per channel

Previous campaign report

Animation channel	Weekday non-prime	Weekday prime	Weekend non-prime	Weekend prime	Total
Spots	58	17	22	3	100
Costs (\$)	87,000	35,700	41,250	7875	\$ 171,825
Reach-3	929,155	1,527,138	330,370	97,027	2,883,690

Sports 500 network	Weekday non-prime	Weekday prime	Weekend non-prime	Weekend prime	Total
Spots	34	9	6	1	50
Costs (\$)	40,800	13,500	8640	1800	\$ 64,740
Reach-3	1,210,067	580,777	212,973	53,054	2,056,871

Notes: These were actuals achieved for the given campaign. The actuals are reported 2 weeks after airing

Appendix 2: Data Description

There are two datasets “infomedia_ch1.csv” and “infomedia_ch2.csv” for Animation channel and Sports 500 network, respectively. Each dataset contains ten variables. These variables largely concern when viewers start and stop viewing the channel, the time and date of the broadcast, and demographic data regarding the viewers (Tables 27.1 and 27.2).

Table 27.1 Data description of input dataset

Variable name	Description
day	Day of the broadcast (1-Monday, 2-Tuesday, . . . , 7-Sunday)
adtime	Time when the advertisement starts
start	Time when the customer started watching the channel
end	Time when the customer stopped watching the channel
custid	Unique Id of the connection
pid	Unique Id of the family members
age	Age of the viewer
sex	Sex of the viewer
population_wgt	Weightage of the similar type of customer
channel	Channel identifier (1—Animation, 2—Sports 500 Network)

Note: The variables adtime, start, and end are in hour format and carry values between 0 and 2359

Table 27.2 Sample observations from input dataset (top five rows from infomedia_ch1.csv)

Day	Ad_time	Start	End	Cust_id	P_id	Age	Sex	Population_wgt	Channel
2	1342	1308	1355	70,953	1	59	M	3427	1
2	1348	1308	1355	70,953	1	59	M	3427	1
1	2100	2100	2109	79,828	1	23	F	5361	1
1	2106	2100	2109	79,828	1	23	F	5361	1
1	2124	2120	2124	79,828	1	23	F	5361	1

Appendix 3: Simulator Description

Help File

Please refer to the code “Infomedia_simulations.R” to run the simulation for each channel separately. The simulation function will ask for the following information:

Enter full path of datasource folder:

< Copy and paste the path name as it is. Please make sure you paste in R console not the editor.>

Enter dataset name (including .csv):

< Datasets infomedia_ch1.csv and infomedia_ch2.csv correspond to channels 1 and 2 respectively. Separate simulation is needed for each channel to quickly get results. Enter the source file name (infomedia_ch1 (or 2).csv) including file extension (.csv). Please note that R is case sensitive—spell the file name correctly.>

Enter the minimum number of slots (typically 5) :

< the number of slots to begin the simulation>

Enter the maximum number of slots (typically 250):

< the number of slots to end the simulation>

Enter the incremental number of slots (typically 5):

< step size, that is, minimum, minimum + stepsize, minimum + 2* stepsize, ... will be simulated>

Enter the number of simulation to run for each spot (typically 100):

< the number of replications—too many will slow down the system>

Minimum percentage slots in prime time [0-100]:

< must be an integer, typically between 20 and 30>

Maximum percentage slots in prime time [0-100]:

< must be an integer, typically between 20 and 30>

Minimum percentage slots on weekends [0-100]:

< must be an integer, typically between 20 and 30>

Maximum percentage slots on weekends [0-100]:

< Must be an integer, typically between 20 and 30>

Once you enter all the inputs correctly, the simulation function will run the simulation for the requested channel given the constraints and share the two output

files—(a) data file that will consist of reach given the number of spots in weekday non-prime, weekday prime, weekend non-prime, and weekend prime time and (b) png file that shows the reach curve against the total number of spots.

Output files (csv and png files) will be saved in the current directory as shown in the code output.

Sample Screenshot (Fig. 27.1)

```
> simulation()
Enter full path of datasource folder: D:\MyData\InfoMedia_Solutions

Current working directory is:
D:/MyData/InfoMedia_Solutions

Enter dataset name (including .csv): infomedia_ch1.csv

*****
Note: Please enter all the values below as positive integer only.
*****

Enter the minimum number of slots (typically 5) : 5
Enter the maximum number of slots (typically 250): 250
Enter the incremental number of slots (typically 5): 5
Enter the number of simulation to run for each spot (typically 100): 100
Minimum percentage slots in prime time [0-100]: 20
Maximum percentage slots in prime time [0-100]: 30
Minimum percentage slots on weekends [0-100]: 20
Maximum percentage slots on weekends [0-100]: 30

Running 100 simulation for 5 to 250 spots with increment of 5 spots.
Constraints are as follows:
percentage slots during prime time: 20 to 30 percent.
percentage slots on weekend : 20 to 30 percent.

*****
The simulation is successfully completed.

You can refer to below files (csv and image) in current working directory.

Current working directory:
{ D:/MyData/InfoMedia_Solutions }

Data file name: <simulation_ch1.csv>
Simulation plots file name: <Reach_curve_ch1.png>

Please use this data to fit the curve and for further analysis.
*****
>
```

Fig. 27.1 Interface of the simulation function in R

Sample Output (Fig. 27.2)

The simulation function developed by the data science team produces the dataset “simulation_ch1.csv” and “simulation_ch2.csv” for each channel. The sample output is shown in Table 27.3.

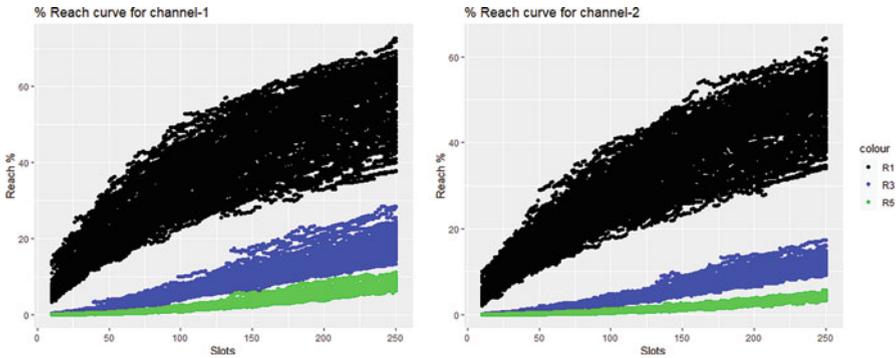


Fig. 27.2 Reach (R-1,-3,-5) vs. spots for channel 1 and 2 (sample output of simulation function)

Table 27.3 Sample output produced by the simulator

	R1	R3	R5	Wkday_NonPrime	WkDay_Prime	WkEnd_NonPrime	WkEnd_Prime
1	596,520	10,660	0	3	1	0	1
2	1,089,675	4438	0	3	1	1	0
3	295,026	0	0	3	0	1	1
4	341,508	0	0	4	0	0	1
5	245,348	0	0	3	0	1	1
..

Notes: R1, R3, and R5 are Reach-1, -3, and -5. The rest of the columns correspond to how many slots were shown in weekday non-prime, weekday prime, weekend non-prime, and weekend prime time

Appendix 4 (Pseudo Code)

1. First, we identify the constraints we would place such as prime vs. non-prime time spots, weekday vs. weekend, blackout zone, and number of spots on each channels.
2. Create a data frame with unique spots and days available.
3. Simulation exercise (Task-1): You can either use simulation function provided with case or develop your custom function using the steps below (3a–3c):
 - (a) In each run, take random sample based on constraints in step 1, using *sample()* function. *Sample (vector from which we have to choose, no of items to choose)*.
 - (b) Merge this data frame with the actual data set depending upon time and day using *merge()* function. Basically, step-3a will generate various samples (to simulate the runs) and will help identify viewers in the next step.
 - (c) Now collect the data on customer level how many times a particular customer viewed the ad. By using *count* function in *plyr* library.
4. Fitting the curve (Task-2)
 - (a) Now we can calculate the total reach of the ad based on the distinct viewers count who watched the ad repeatedly, that is, at least once (R1), at least thrice (R3), at least five times (R5), where the total reach is the sum of the population weight column which represents the weightage of similar type of customers. Plot the total reach and number of spots (say varying total spots between 5 and 250 in steps of 5 or 10).
 - (b) We fit the curve to estimate the average total reach for a given slot size. You will need one curve each for R-1, R-3, R-5, etc.

$$Reach = f(slots) + \varepsilon$$

5. Error estimation (Task-2)
 - (a) Collect the data about the distribution of slots: how the slots have been distributed among the different constraints like day of the week, time bucket, prime time, non-prime, and weekend and weekday.
 - (b) We can see if additional data can explain the error in the above fit.

$$\varepsilon = f(daypart, demographics, other\ explanatory\ variables)$$

- (c) Ensemble the above models to better estimate the total reach.

$$Reach = f(slots, daypart, demographics, other\ explanatory\ variables) + \varphi$$

6. Optimization (Task-3): Now, we optimize the total reach using nonlinear optimization techniques (refer to Chap. 11 Optimization):
- (a) Objective function: Maximize appropriate reach
 - (b) Constraints
 - Number of weekend spots out of total spots (weekend regular + weekend prime)
 - Number of prime-time spots out of total spots (weekday prime + weekend prime)
 - Available budget for the advertisement (total, channel-wise)

Reference

Goerg, M. (2014). *Estimating reach curves from one data point*. Google Inc. Retrieved June 23, 2018, from <https://ai.google/research/pubs/pub43218>.