Viacom Project

Guaranteed Social Campaign Based on Targeted Advertising

ALY6080 71109 Integrated Experiential Learning
Northeastern University
Minyi Chen
12/11/2018

Outline

I. Introduction: Project and Data

I. Introduction

- This project is to help solving Viacom’s business problems and analyze the data provided by Viacom.
- Specifically, try to find a strategy to determine potential revenue that may be generated if Viacom offers campaign guarantees based on demographically targeted impressions.
- The deliverables of the project: this PowerPoint, and Tableau document.
- The tool I used in the project is R and Tableau.
- The method: use R to clean the dataset, and then use Tableau to explore the dataset. R codes are listed in this pptx file, and the Tableau is in another file.
- Section I discusses what we have: Viacom’s business problems and the data. Section II reviews literature on this topic. Section III explores the data. Section IV gives some recommendations on the guaranteed social campaign offer. Section V concludes. References and Appendix of R codes are at the end.

Business Model of The Advertising Team

Viacom:

- "Viacom is home to the leading portfolio of global, multi-platform entertainment brands through television, film, digital content, live events, merchandise, studio production and more."

Advertising Team:

- "Viacom monetizes its social presence by creating custom branded content for advertisers that get published on Viacom's social channels (for example MTVs Facebook page)."
- Viacom wants to offer Guaranteed Social Campaign to their advertisers.
- Guarantee that it will meet a certain number of views or impressions over a certain period of time (1 week - 1 month)

Guaranteed Social Campaign

- Assumption: guarantee the ability to reach particular demographics would be valuable to advertisers and profitable for Viacom.
- Viacom could reach some of the target demographic organically, but will have to supplement any Guaranteed Social Campaign with paid promotion.
- For the purposes of this project Viacom offers an industry standard of $25 CPM (Cost per 1000 impressions) or $0.10 CPV (Cost per View) to advertisers. That is, Viacom charges advertisers advertising fee (Price = $25 per 1000 impressions).
- Viacom may purchase views/impressions from Facebook (Cost ≈ $3.8 per 1000 impression). We focus on Facebook in this project.
- $\text{Profit} = Price \times Quantity\ of\ impressions - Cost\ \times Quantity\ of\ paid\ impression$
- $\text{Profit}\ > 0$, and higher profit would be desirable.

Business Questions

1. Is it worth pursuing this offering?
2. Where are there gaps in the market?
3. Which Facebook pages could be most suitable for each demographic?
4. What value will this provide to our partner advertisers?
5. How do we use the demographic information provided by digital and social channels to optimize the ad prices give to our partner advertisers?
6. How much can we increase our margin by providing demographic targeting on our digital and social channels?
7. Do we need targeting to stay relevant in the market-place?
8. Would more specific targeting be viable, and a marketplace differentiator?

Data

- There are two types of datasets that are provided by Viacom
  - cpm_estimates-10-24-18.csv
  - pages_level_demos_YYYY_MM.csv
    - for example, pages_level_demos_2017-01.csv
    - Each .csv represents the data in a month that is from January 2017 to May 2018
    - 17 months in total
  - sample_pages_level_demos.csv
    - A 10,000 row random sample from pages_level_demos_2017_05.csv

Variables: cpm_estimates-10-24-18.csv

- This dataset estimates the CPM on a particular page according to Facebook.
- Variables:
  - hID: The page ID. They are all refer to Viacom Facebook account pages.
  - age_max: The maximum age in the targeting parameter
  - age_min: The minimum age in the targeting parameter
  - male: If targeting males in the targeting parameters. Binary values, 1 = targeting, 0 = not targeting
  - female: If targeting females in the targeting parameters. Binary values 1 = targeting, 0 = not targeting
  - time_pulled: The time the targeting was accessed
  - **cpm: the cost per 1000 impressions, in USD**

Sample: cpm_estimates-10-24-18.csv

age_max	age_min	cpm	female	male	time_pulled	hID
24	18	5.170808	0	1	4/13/2018 19:43	-1250019159
24	18	3.736275	1	0	4/13/2018 19:43	-1250019159
24	18	4.281942	1	1	4/13/2018 19:43	-1250019159
34	25	4.66882	0	1	4/13/2018 19:43	-1250019159
34	25	5.861955	1	0	4/13/2018 19:43	-1250019159
34	25	3.482042	1	1	4/13/2018 19:43	-1250019159
44	34	5.900977	0	1	4/13/2018 19:43	-1250019159
44	34	4.625132	1	0	4/13/2018 19:43	-1250019159
44	34	3.816335	1	1	4/13/2018 19:43	-1250019159
54	44	5.850961	0	1	4/13/2018 19:43	-1250019159
54	44	7.301951	1	0	4/13/2018 19:43	-1250019159
54	44	6.023444	1	1	4/13/2018 19:43	-1250019159
34	18	3.557826	0	1	4/13/2018 19:43	-1250019159
34	18	3.326234	1	0	4/13/2018 19:43	-1250019159
34	18	3.843543	1	1	4/13/2018 19:43	-1250019159

CPM varies based on the variables age_max, age_min, female, male, time_pulled from this dataset.

Variables: pages_level_demos

- This dataset is page-level data on demographic data from Facebook. To get an understanding of the distribution of demographics that visit and interact with Viacom's Facebook pages. All data here refers to the previous days metrics from midnight to midnight, so if the date is `2017-05-01 07:00:00`, the data is referencing to`2017-04-30 00:00:00` - 2017-05-01 00:00:00.
- Variables:
  - hID: The page ID
  - name: The name of the high-level metric, they can be one of:
    - 'page_impressions_by_age_gender_unique'
    - 'page_impressions_by_country_unique'
    - 'Page_impressions_by_city_unique'
    - 'page_impressions_by_locale_unique'
    - 'page_views_by_age_gender_logged_in_unique'
    - 'page_cta_clicks_by_age_gender_logged_in_unique'
  - metric: The specific parameter by which the row is referencing, for example, if the value of `name` is `page_impressions_by_age_gender_unique` the value for `metric` could be `F.25-34` referring to page impressions for females aged between 25-34
  - value: the value, that is, how many impressions/views/cta_clicks
  - date: The date/time the information was accessed

Sample: pages_level_demos_2017-01.csv

Name	Metric	Value	Date	hID
page_impressions_by_age_gender_unique	F.13-17	715	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.18-24	5808	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.25-34	4699	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.35-44	1220	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.45-54	483	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.55-64	165	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	F.65+	116	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	M.13-17	322	1/5/2017 8:00	-417994605
page_impressions_by_age_gender_unique	M.18-24	4669	1/5/2017 8:00	-417994605
page_impressions_by_country_unique	AR	64	1/5/2017 8:00	-417994605
page_impressions_by_country_unique	FR	140	1/5/2017 8:00	-417994605
page_impressions_by_country_unique	GB	3602	1/5/2017 8:00	-417994605
page_impressions_by_city_unique	Orlando, FL	286	1/5/2017 8:00	-417994605
page_impressions_by_city_unique	Tallahassee, FL	71	1/5/2017 8:00	-417994605
page_impressions_by_city_unique	Quezon City, Metro Manila, Philippines	85	1/5/2017 8:00	-417994605
page_impressions_by_city_unique	Singapore, Central Region, Singapore	91	1/5/2017 8:00	-417994605
page_impressions_by_locale_unique	ar_AR	5	1/5/2017 8:00	-417994605
page_impressions_by_locale_unique	bg_BG	5	1/5/2017 8:00	-417994605
page_impressions_by_locale_unique	ca_ES	4	1/5/2017 8:00	-417994605
page_views_by_age_gender_logged_in_unique	UNKNOWN	0	1/5/2017 8:00	-417994605
page_views_by_age_gender_logged_in_unique	<13	0	1/5/2017 8:00	-417994605
page_views_by_age_gender_logged_in_unique	13-17	0	1/5/2017 8:00	-417994605
page_cta_clicks_by_age_gender_logged_in_unique	8.14636E+14	0	1/5/2017 8:00	-417994605

II. Literature review

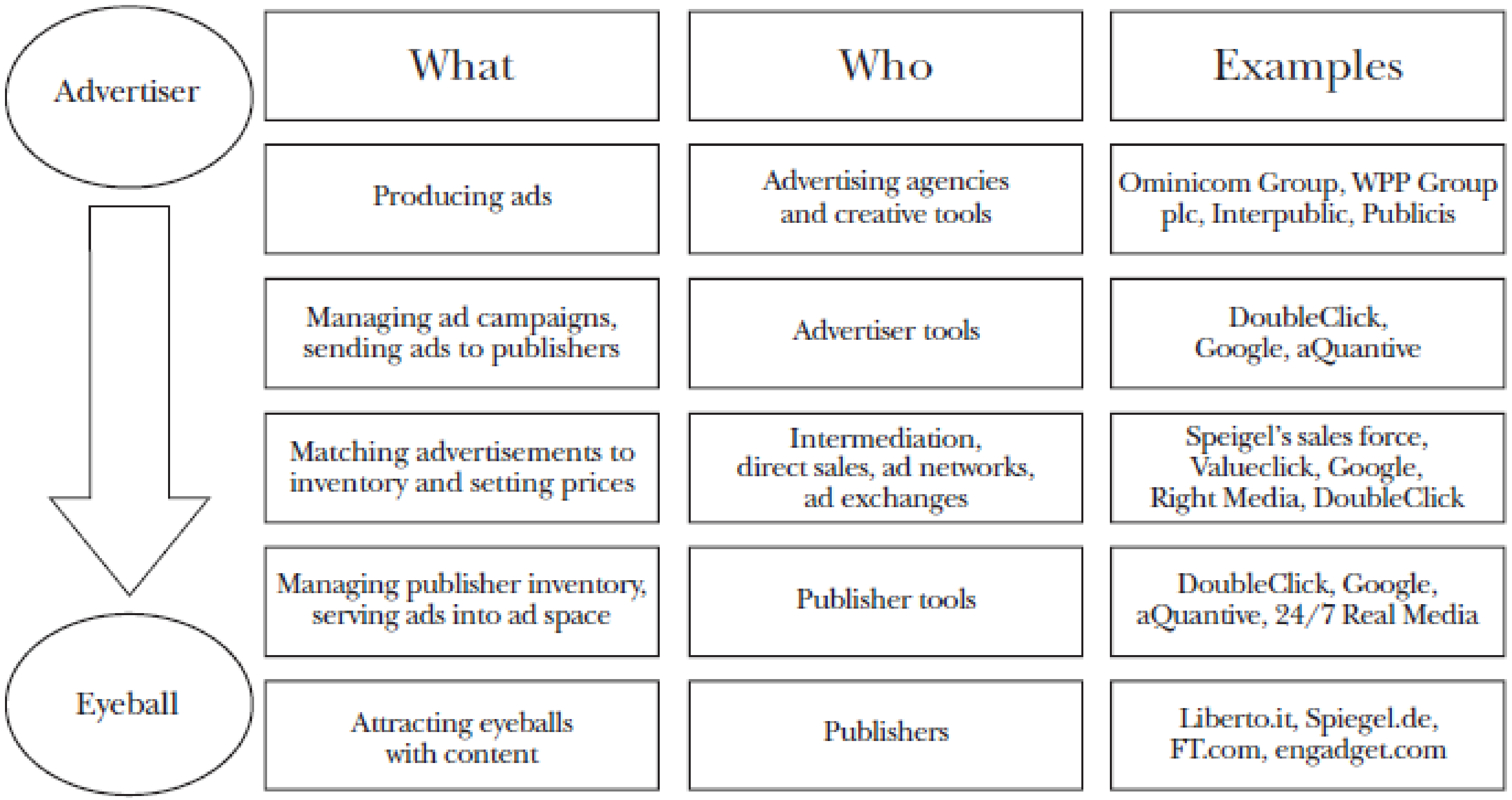
II. Literature review

- Online Advertising Industry
  - Evans (2009)
- CPM model:
  - Asdemir, Kumar & Jacob (2012)
  - Fridgeirsdottir & Najafi-Asadolahi (2018)
- The guaranteed targeted display advertising:
  - Turner (2012)
- Cons of CPM:
  - Dorbian (2009)
- Others
  - Shen & Miguel Villas-Boas (2017)
  - Liu & Mookerjee (2018)

Evans (2009)

- Evans points out the significance of the online advertising industry to the e-commerce economy.
- He further examines the supply and demand of the industry, and the privacy and regulation issues of the industry.

Figure 1
Relationship between Various Online Advertising Businesses



Evans (2009)
Supply and Demand of online advertising industry

- Top 11 suppliers:

Rank	Property	Content	2008 Internet advertising revenues ($ millions)
1	Fox Interactive Media, including MySpace	entertainment video, news, social networking, image hosting, games network	900
2	Yahoo! sites	search results and various applications (news, e-mail, weather forecast)	3430
3	Google sites including YouTube	search results, e-mail, maps, user-uploaded videos, blogs	7430
4	Microsoft sites	search results, e-mail, entertainment videos, music, news	1970
5	AOL LLC	news, entertainment, e-mail, search results, greetings	1360
6	Facebook.com	social networking site	130
7	eBay	online auction and shopping site for mostly used goods	
8	Comcast Corporation	TV listings, free TV episodes, cable television services	
9	Viacom Digital	entertainment news, videos, music clips, TV listings, reviews	
10	Time Warner (excludes AOL)	movies, TV schedule, videogames, cable television services,	
11	Amazon sites	online shopping site, daily blog, customer reviews	

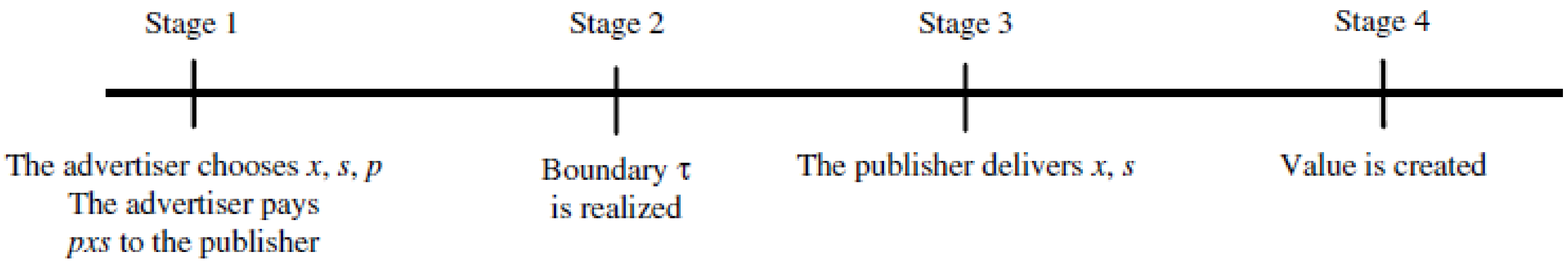
- Demand side:

Advertisers normally pay firms to design and run their advertising campaigns.

Asdemir, Kumar & Jacob (2012)

- They study both the CPM (input-based cost per thousand impressions and CPC (performance-based cost per click-through) models.
- Their CPM model is affected by price per impression (p), the campaign intensity (x), the campaign spread level (s), etc.
- They also do the comparisons between CPM and CPC and explain how firms could choose which one under different situations.

Figure 2 **Timeline of Events in the CPM Model**



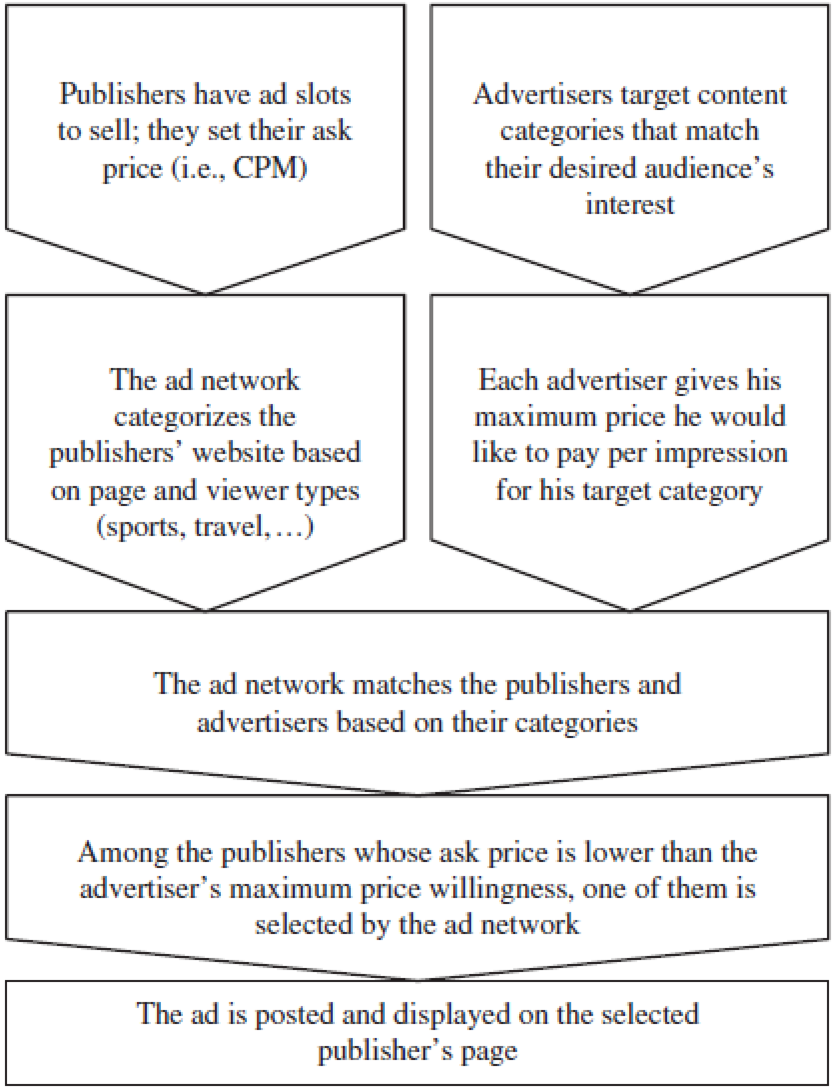
Turner (2012)

- Turner (2012) studies the planning problem of the guaranteed targeted display advertising. In his model, the network provider manages ad inventory, that is, the impressions. Advertisers buy impressions through a campaign. The campaign's targeting is the constrains of the inventory.
- The model includes variables: viewer types (e.g. M 18-25), targeting viewer types, campaigns, the impression goal of the campaigns, the number of impressions a viewer type creates, etc.
- His model studies how uncertainty of the audience, forecast error, and the random slotting of the ad server will affect the optimal level of the ad server.

Fridgeirsdottir & Najafi-Asadolahi (2018)

- They show us the general steps before the contracts.
- They assume that publishers face uncertain demand and supply.
- Their price model involves advertisers' arrival rate, the number of impressions, and the number of slots. They also assume those rate variables follow a statistic distribution.
- They find that optimal CPM price can increase in the number of impressions made of each ad.

Figure 1. The General Transaction Steps Between Advertisers and Web Publishers Through an Advertising Network



Cons of CPM by Dorbian (2009)

- Dorbian (2009) questions whether CPMs still provide value. She calls for new alternatives of setting pricing for online advertising.
- She mentions some people who are against CPM. Shostack’s company uses CPL (cost oer lead) to set rates because he care more about the level of engagement and targeting instead of sales only.
- Some other important factors to consider when setting ad rates are the closer relationship with audience and reach the right audience.
- At the end, she mentions a new trend of pricing which allows the prices directly reflect the transactions.

Others

- Shen & Miguel Villas-Boas (2017) study how firms can benefit from sending messages to consumers based on their past buying patterns. They find that firms can benefit from behavior targeting if the annoyance of receiving advertising is not too large. This helps us understanding the pros and cons of the targeted advertising.
- Liu & Mookerjee (2018) show that firms could be benefited by using different strategies to make the advertising pricing decisions. They find that traffic-based competition generates higher profits for both firms than those generated in spending-based competition.

III. Exploratory Data Analysis

Data Exploration
CPM Dataset: The cost of impression

The cost of impression from CPM Dataset

- To understand the cost of impression, I use cpm_estimates-10-24-18.csv.
- I use R to clean the dataset, and create three columns: age, group, gender. R codes are in the Appendix.
- Then, I use Tableau to do the analysis.

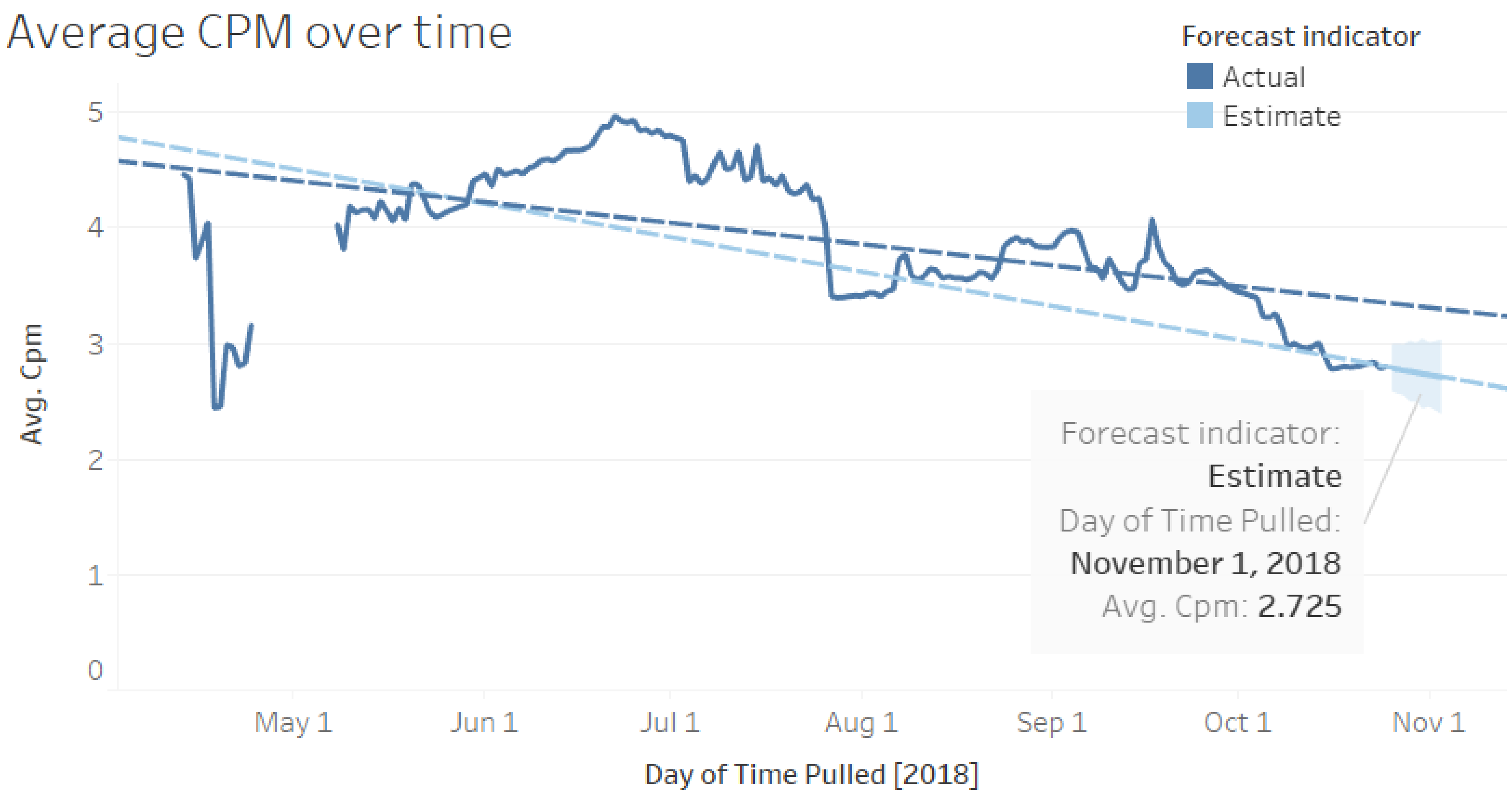
Summary of CPM dataset

```
 age_max      age_min          cpm            female      male                      time_pulled                hID
 24 : 6039    18:30195    Min.    :0.000    0:16104    0:16104    2018-04-14 01:12:47:    48    -2111977094:8064
 34 :12078    25: 6039    1st Qu.:3.074     1:32208    1:32208    2018-04-14 01:52:43:    48    -1874613969:8136
 44 :12078    34:  147    Median :3.788                           2018-04-14 02:22:49:    48    -1531304817:8232
 49 : 6039    35: 5892    Mean   :3.875                           2018-04-15 01:33:22:    48    -1379668826:7992
 54 : 6039    44:  147    3rd Qu.:4.340                           2018-04-15 02:03:22:    48    -1293571655:6144
 65+: 6039    45: 5892    Max.   :9.095                           2018-04-15 02:13:20:    48    -1250019159:4800
                          NA's   :7                               (Other)             :48024    1186742240 :4944
```

- I use R to get the summary of these variables
- We can see that there are 7 different Facebook pages in this dataset. The number of observations of each page ID is listed in the last column above.
- CPM is 3.875 on average, and ranges from 0 to 9.095, which is much lower than $25 that Viacom will charge advertisers

Average CPM over time and its forecast

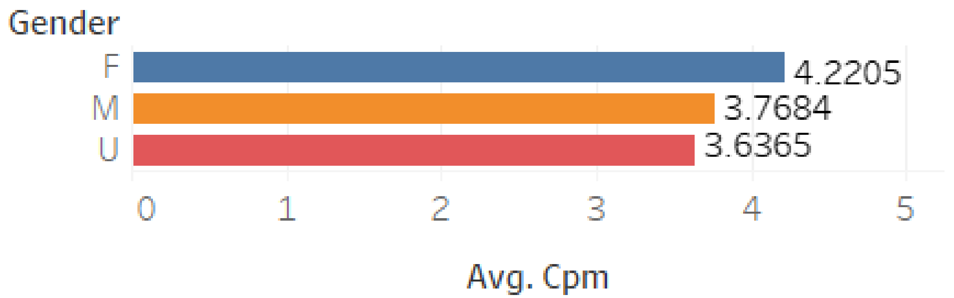
After July 25, 2018, The rate drops.
The forecast shows that the CPM is decreasing.



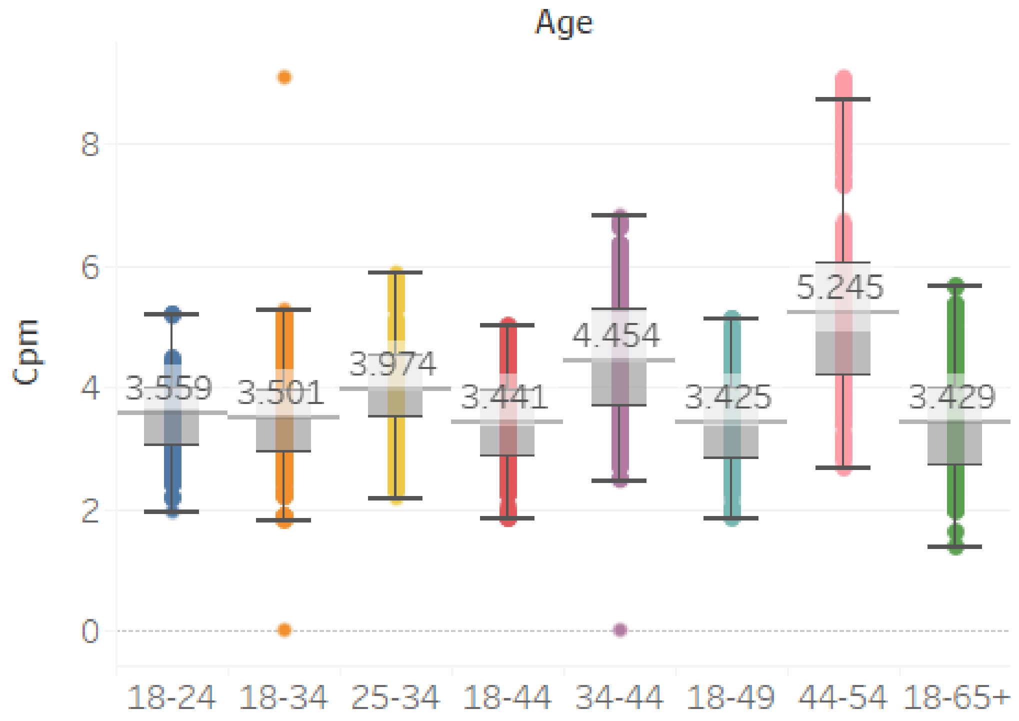
Average CPM by gender and age

- We know the overall average CPM of the sample is 3.875
- The average CPM for female is 4.22 which is above the overall average level.
- Average CPM for age 25-34, 34-44, 44-54 are higher than 3.875.
- The highest average CPM is for age 44-54. It seems this group of people don’t spend much time online and the new technology.

Average CPM by gender

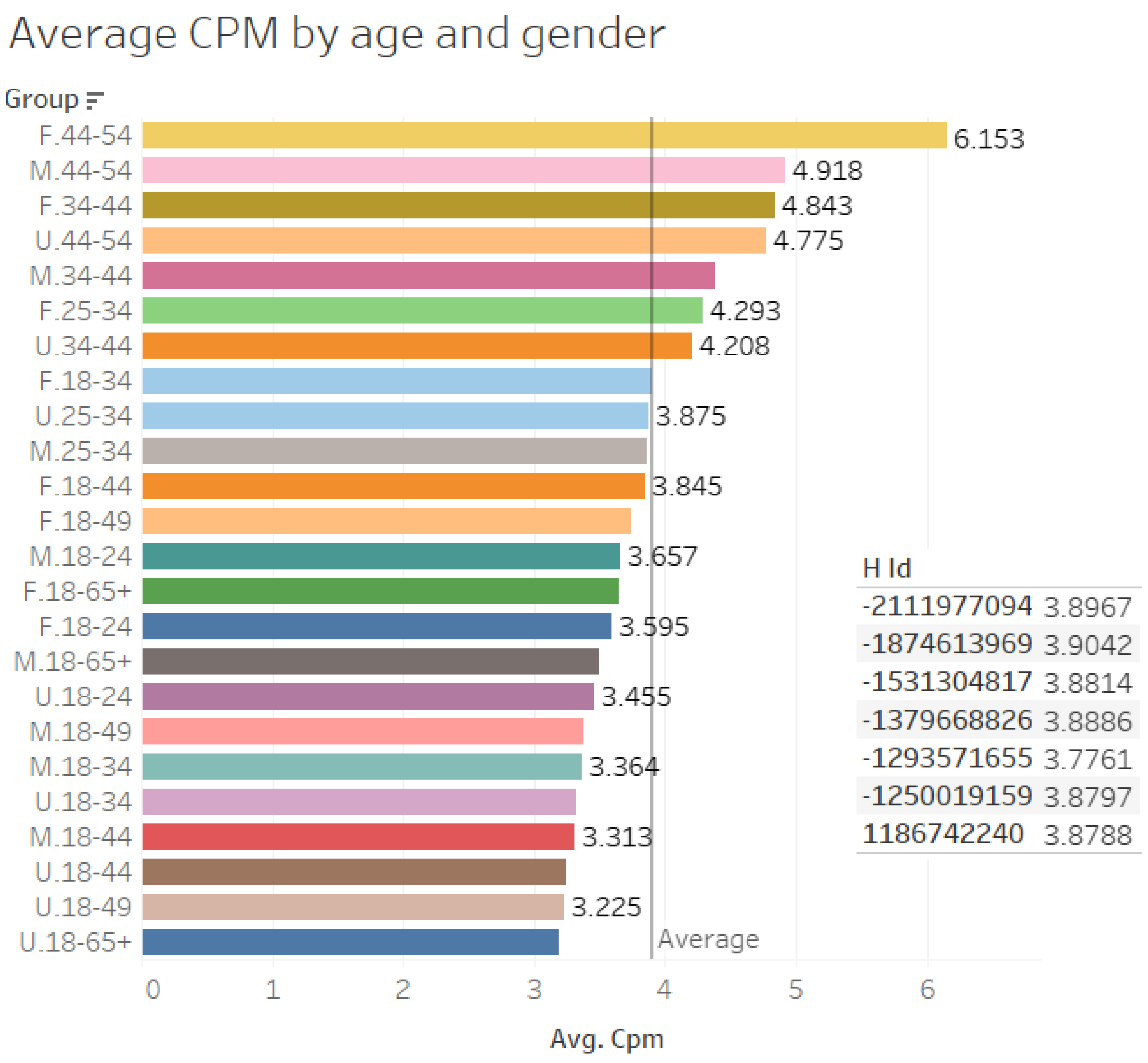


Average CPM by age



Average CPM by age and gender

- The graphs shows 7 groups have higher average CPM than the overall average CPM.
- F.44-54 has the highest average CPM 6.153.
- Also, I further look into different pages, they all have similar average CPM.



Data Exploration
Pages_level_demos Dataset: demographics of pages

Demographics of pages
from Pages_level_demos Dataset

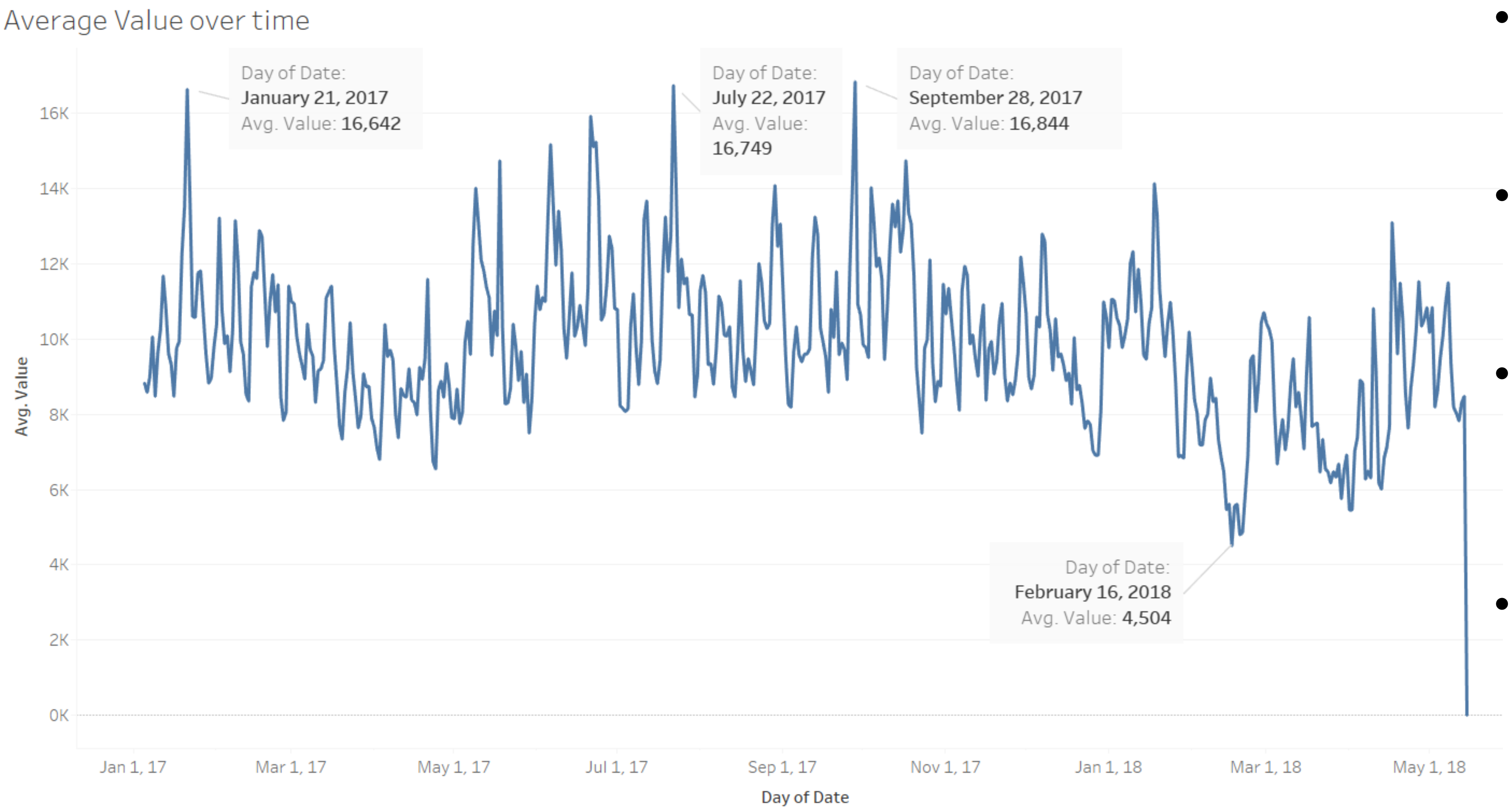
- To understand the demographics of pages, we use the second type of datasets, pages_level_demos_YYYY_MM.csv.
- I use R to combine all the 17 months to one dataset: page.csv. The size of this one dataset is about 1G. R codes are in the appendix.
- Then, I use Tableau to explore the data. Tableau is attached.
- At the end, I also merger the CPM dataset and the sample_pages_level_demos.csv in Tableau to look at 7 different pages.

Summary of the page dataset

```
> summary(page)
                                             Name                Metric               Value                             Date                          hID        
 page_cta_clicks_by_age_gender_logged_in_unique: 135528   <13      :  129428   Min.    :        0   2017-02-15 08:00:00:    31074   -1874613969:    82719  
 page_impressions_by_age_gender_unique          :1605464   13-17    :  129428   1st Qu.:        2   2017-02-12 08:00:00:    30925   1937545411 :    82578  
 page_impressions_by_city_unique                :3952599   18-24    :  129428   Median :       27   2017-02-10 08:00:00:    30807   -1250019159:    82571  
 page_impressions_by_country_unique             :3414336   25-34    :  129428   Mean   :     9857   2017-02-13 08:00:00:    30776   -53452065  :    82541  
 page_impressions_by_locale_unique              :2884972   35-44    :  129428   3rd Qu.:      524   2017-02-16 08:00:00:    30733   1720551795 :    82364  
 page_views_by_age_gender_logged_in_unique      :1164852   (Other):12507645   Max.   :38196432   2017-01-06 08:00:00:    30582   562697744  :    82323  
                                                            NA's     :    2966   NA's   :3          (Other)             :12972854   (Other)    :12662655  
```

- Overall, from the summary in R, we can see there are five variables in the dataset: name, metric, value, date, hid.
- The dataset covers 6 types of name. There are a lot of categories in the metric and date. There are 326 different page ids in total (I found in Tableau). All pages has similar number of records at 82000. Date 2017-02-15 has the most records.
- The mean of the Value is 9857 and the maximum is 38196432.

Avg. Value over time, from 1/1/2017 to 5/15/2018



- We can see the average value goes up and down a lot over each date.
- The graph also shows some value for the peaks and downturn.
- This suggests that special events could affect the page impressions.
- Special events could be a new release of a page, show, etc., or other social events.

Average of Value across Name and HId

- This table is sort by the average page impression by gender and age, and only covers the pages with top average of Value.

Average of Value broken down by Name vs. H Id.

	Name					
H Id	page_cta_cli..	page_impr..	page_impres..	page_impres..	page_impres..	page_views_..
-1379668826	0	567,714	42,176	240,393	263,987	0
-243913656	0	363,655	37,967	164,372	169,509	0
562697744	0	272,150	36,870	124,385	127,055	0
1937545411	0	212,433	17,873	92,747	98,850	0
237055665	0	150,621	17,024	68,341	70,244	0
-1958438828	0	133,676	10,112	59,407	62,254	0
1186742240	0	126,923	11,277	57,402	59,153	0
-1011603482	0	110,242	7,924	47,220	51,274	0
1497808268	0	104,792	7,833	46,835	48,790	0
-1123759962	0	94,506	13,413	43,010	44,088	0
1720551795	0	86,060	24,861	39,898	40,144	0
632571698	0	85,951	5,824	38,307	40,040	0

Top Values of five different Names

Top page views

Name	Metric	
page_views_by_age_gender_logged_in_unique	<13	0
	13-17	0
	18-24	0
	25-34	0
	35-44	0
	45-54	0
	55-64	0
	65+	0
	UNKNOWN	0

Top cta clicks

Name	Metric	
page_cta_clicks_by_age_gender_logged_in_unique	1005224486156055	0
	10150614078699975	0
	10150847737844975	0
	10152545673585404	0
	10152585440316366	0
	10152597061496990	0
	10152614885456814	0
	10152625076002343	0
	10152658079691440	0

Top impressions by locale

Name	Metr..	
page_impressions_by_locale_unique	en_US	27,154,608,350
	en_GB	3,107,722,723
	es_LA	2,880,970,271
	pt_BR	1,205,656,065
	it_IT	949,789,281
	es_ES	885,579,143
	de_DE	677,386,841
	fr_FR	642,858,886

Top impressions by country

Name	.	
page_impressions_by_country_unique	US	22,824,803,174
	GB	2,151,219,001
	MX	1,868,814,445
	BR	1,220,884,582
	CA	1,070,458,003
	IT	957,416,268
	AU	823,958,345
	PH	641,988,533
	DE	628,621,933
	IN	578,860,773
	PL	548,497,970

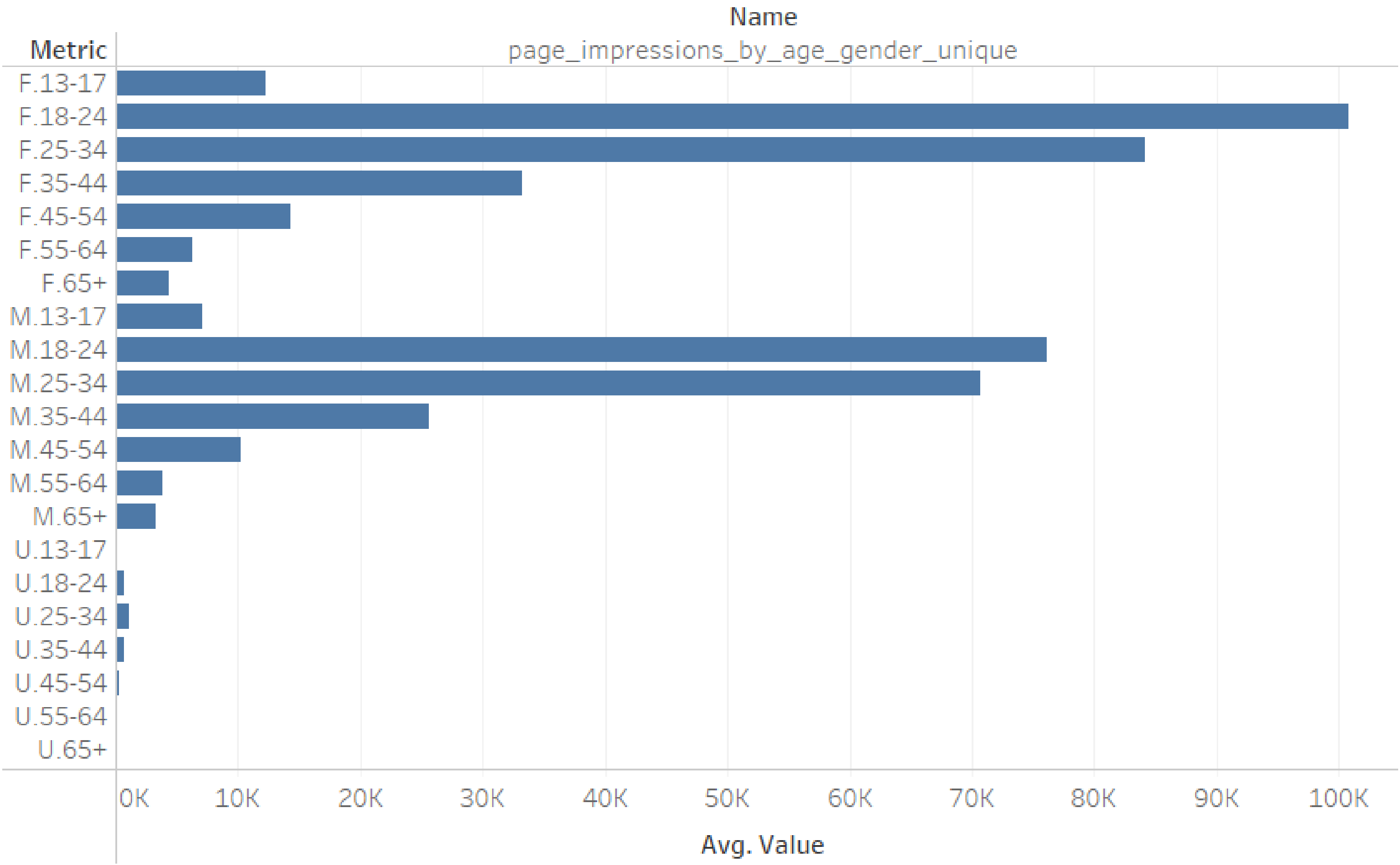
Top impressions by city

Name	Metric	
page_impressions_by_city_unique	New York, NY	734,893,895
	Chicago, IL	412,715,459
	Mexico City, Distrito Federal, Mexico	334,371,289
	Los Angeles, CA	324,144,651
	London, England, United Kingdom	310,514,820
	Houston, TX	303,265,790
	Philadelphia, PA	201,614,327
	Dallas, TX	179,709,304
	San Antonio, TX	153,153,418
	São Paulo, SP, Brazil	127,143,819
	Phoenix, AZ	121,900,035

- From this, we can see that there is no value for two names in this dataset, that is,
  - page_views_by_age_gender_logged_in_unique
  - page_cta_clicks_by_age_gender_logged_in_unique
- That means, we can only focus on the page impressions metrics.
- We can also see that top page impressions (sum of Value) are in the US, more specially, in New York City, and en_US.

Bar chart: Page impression by age and gender

F.18-24 has the highest avg. impression, female is overall more active than male, and most page impressions are from age groups 18 -44 years old



Top impressions by age and gender

Top impressions by age and gender

Name	Metr..	
page_impressions_..	F.18-24	9,140,555,796
	F.25-34	7,645,014,445
	M.18-24	6,879,823,272
	M.25-34	6,415,376,632
	F.35-44	2,982,845,604
	M.35-44	2,292,436,116
	F.45-54	1,264,909,107
	F.13-17	1,035,891,442
	M.45-54	891,567,139
	M.13-17	579,161,332
	F.55-64	536,306,504
	F.65+	359,290,572
	M.55-64	321,829,799
	M.65+	272,365,138
	U.25-34	73,712,376
	U.35-44	40,637,096
	U.18-24	39,217,284
	U.45-54	16,762,523
	U.55-64	7,749,979
	U.65+	7,103,543
	U.13-17	3,459,485
	value	0

Page impression (sum of value) by age and gender

Name		H Id							
	Metr..	-1379668826	1937545411	562697744	1011603482	1720551795	-243913656	1497808268	-2050455356
page_impressions_by_age_gender_unique	F.13-17	217,625,926	131,101,769	18,214,981	61,225,905	56,668,671	13,138,348	31,440,549	36,691,641
	F.18-24	1,874,708,832	719,879,815	420,639,600	350,016,849	348,507,307	328,624,672	315,052,901	300,307,484
	F.25-34	1,030,370,341	385,955,168	637,816,739	179,654,980	91,684,643	631,683,018	157,714,941	107,539,038
	F.35-44	243,757,503	104,870,850	394,631,561	46,091,137	16,140,762	420,533,207	41,953,849	35,856,718
	F.45-54	81,173,496	41,099,775	191,115,358	14,804,669	4,660,359	227,096,999	19,226,953	13,208,149
	F.55-64	30,139,592	14,375,622	66,046,943	5,712,482	1,502,589	135,779,106	7,249,852	5,999,393
	F.65+	25,565,194	9,702,764	39,615,517	4,809,035	1,781,603	96,066,546	5,570,512	2,552,550
	M.13-17	92,900,253	49,920,994	10,304,298	27,762,862	30,398,971	12,689,369	26,376,550	11,955,397
	M.18-24	1,210,696,668	399,948,444	263,017,144	237,923,900	250,547,501	360,366,040	273,012,354	157,685,912
	M.25-34	813,235,656	243,156,451	411,147,000	153,690,086	71,816,612	712,924,307	148,471,788	59,768,089
	M.35-44	170,868,081	59,845,380	218,865,609	37,520,161	13,025,143	427,472,019	34,253,092	18,216,604
	M.45-54	51,169,430	22,108,929	93,338,635	11,031,348	3,506,756	207,602,805	14,579,473	5,670,048
	M.55-64	17,591,232	7,550,267	27,511,302	3,745,816	1,075,344	98,049,964	5,087,224	2,249,549
	M.65+	23,665,412	7,087,808	23,763,082	4,577,798	1,894,393	69,502,865	5,142,372	1,241,843
	U.13-17	602,682	274,900	72,454	305,959	34,350	145,529	193,060	19,284
	U.18-24	5,813,923	1,809,865	1,394,048	1,497,147	460,653	3,959,293	1,877,983	243,862
	U.25-34	7,369,298	2,752,180	5,146,846	1,989,662	538,626	12,919,224	1,343,097	128,922
	U.35-44	2,889,023	1,521,514	3,849,025	919,451	190,310	10,802,149	460,562	55,540
	U.45-54	965,530	623,617	1,724,264	266,112	38,642	5,400,457	213,664	18,816
	U.55-64	415,559	234,037	580,794	106,615	12,009	3,065,220	93,044	8,586
	U.65+	430,002	173,549	473,446	106,994	18,034	2,736,283	106,667	6,287
	value	0	0	0	0	0	0	0	0

- The left table shows that the top 4 categories are F.18-24, F.25-34, M.18-24, M.25-34
- The right table shows the top pages that have the largest impressions of F.18-24.

Top pages by city

H Id	Name page_impres..
-1379668826	939,509,888
-243913656	845,743,266
562697744	821,311,623
1720551795	553,799,179
1937545411	397,340,331
237055665	378,463,578
-2050455356	342,215,064
-1123759962	298,180,459
1186742240	251,209,116
-1958438828	225,255,224
1550552148	201,339,998
-1011603482	176,154,907
1497808268	174,478,059
-1531304817	143,886,879
-549502059	129,494,440
632571698	129,484,010

Top pages by country

H Id	Name page_impressio..
-1379668826	5,354,759,163
-243913656	3,661,548,544
562697744	2,770,811,018
1937545411	2,061,862,966
237055665	1,519,213,075
-1958438828	1,323,345,876
1186742240	1,278,680,673
-1011603482	1,049,737,917
1497808268	1,043,302,573
-1123759962	956,146,434
1720551795	888,760,106
632571698	851,612,446
-549502059	796,844,924
-96898710	774,632,398
-2050455356	758,371,727
2113154542	648,375,601

Top pages by locale

H Id	Name page_impressio..
-1379668826	5,880,566,847
-243913656	3,775,983,103
562697744	2,827,742,837
1937545411	2,197,534,504
237055665	1,561,593,424
-1958438828	1,386,780,014
1186742240	1,317,688,135
-1011603482	1,139,873,235
1497808268	1,086,856,215
-1123759962	979,941,547
1720551795	894,256,704
632571698	890,129,896
-549502059	801,774,680
-96898710	782,690,279
-2050455356	759,359,037
2113154542	695,967,379

Top pages by gender age

H Id	Name page_impress..
-1379668826	5,901,953,633
-243913656	3,780,557,420
562697744	2,829,268,646
1937545411	2,203,993,698
237055665	1,562,692,257
-1958438828	1,389,696,071
1186742240	1,319,493,156
-1011603482	1,143,758,968
1497808268	1,089,420,487
-1123759962	980,500,154
1720551795	894,503,278
632571698	891,566,090
-549502059	802,001,016
-96898710	783,033,942
-2050455356	759,423,712
2113154542	698,044,513

- These four graphs shows the sum of Value for each page id. Some pages are on the top of all four categories.
- They are -1379668826, -243913656, 562697744, 1937545411, 237055665
- Page -1011603482, 1720551795 are not the top pages by gender age in the 4th graph, but they are the top pages under F.18-24 (last slide). This suggests that these two pages are for the targeting group female who aged 18-24.
- It would be interested to look further into each page to see what drives them to be the top pages across these categories. I then pick the above top five pages.

Top 5 selected pages by city

1

Metric	H Id -1379668826
New York, NY	63,359,951
Mexico City, Distrito Federal, Mexico	60,684,126
London, England, United Kingdom	45,585,542
Quezon City, Metro Manila, Philippines	41,032,223
Bogotá, Distrito Especial, Colombia	34,026,839
Chicago, IL	33,235,794
Lima, Lima Region, Peru	32,587,920
Los Angeles, CA	32,440,971
New Delhi, Delhi, India	30,877,179

2

Metric	H Id -243913656
New York, NY	113,423,416
Los Angeles, CA	55,360,260
Chicago, IL	52,383,562
London, England, United Kingdom	38,171,158
Houston, TX	32,596,527
Philadelphia, PA	24,082,725
Toronto, ON, Canada	23,484,532

3

Metric	H Id 237055665
New York, NY	48,274,709
Chicago, IL	30,852,587
Houston, TX	22,967,412
London, England, United Kingdom	14,179,413
Dallas, TX	13,732,437
Los Angeles, CA	13,700,952
Philadelphia, PA	13,597,492
San Antonio, TX	9,257,134

4

Metric	H Id 562697744
New York, NY	106,329,738
Chicago, IL	64,375,569
Houston, TX	47,855,622
Philadelphia, PA	36,074,083
Los Angeles, CA	33,379,202
Atlanta, GA	30,735,259
Dallas, TX	29,156,233
Baltimore, MD	23,202,315
Memphis, TN	22,848,466

5

Metric	H Id 1937545411
London, England, United Kingdom	68,232,801
Quezon City, Metro Manila, Philippines	14,328,455
Dublin, Ireland	13,299,639
New York, NY	11,761,199
Manchester, England, United Kingdom	11,399,570
Leeds, England, United Kingdom	11,322,172
Mexico City, Distrito Federal, Mexico	10,395,141
Birmingham, England, United Kingdom	10,110,747

- These tables help us understand where the selected pages' impressions from.
- Page #2, #3, #4 have similar patterns that their top cities are the big cities in the US. It suggests that these pages are US based pages.
- Top 3 cities of page #1 are New York, Mexico City, and London. This page seems more diverse that the cities are from different countries.
- Page #5 is a little bit different. Its top cities are London, Quezon City, and Dublin. It seems that this page is a page of a foreign program designed to UK or Euro audience, but also favored by people from New York and Mexico.

Top 5 selected pages by country

1

Metric	-1379668826
US	2,080,953,613
MX	331,274,623
GB	312,276,075
IN	296,119,077
PH	225,623,595
CA	167,525,388
IT	158,837,344
AU	146,959,408
BR	137,403,147

2

Metric	-243913656
US	2,845,289,935
CA	169,924,827
GB	114,445,408
AU	60,094,590
DE	47,443,521
IN	27,982,104
ZA	27,815,538
MX	25,455,950
NL	23,125,222

3

Metric	237055665
US	1,183,710,758
GB	60,243,240
AU	52,935,162
CA	39,237,185
PH	17,401,177
NZ	16,561,862
NL	11,302,240
TT	11,051,100
ZA	9,450,463

4

Metric	562697744
US	2,479,152,762
GB	38,500,728
CA	27,022,972
ZA	26,004,846
NG	14,242,162
JM	13,488,131
GH	13,263,027
AU	12,303,200
KE	10,151,954

5

Metric	1937545411
GB	710,075,032
US	460,164,169
PH	83,702,650
AU	83,097,473
MX	60,482,538
IN	52,124,399
CA	51,422,438
IT	45,550,185
PL	38,656,682

- These tables show the top countries each page receives the impressions.
- We can see that the first four pages attract more impressions from the US while the 5th page attracts more impressions from Great Britain.

Top 5 selected pages by locale

1

Metric	-1379668826
en_US	3,126,102,011
es_LA	632,066,734
en_GB	586,958,870
es_ES	193,994,632
it_IT	156,538,904
fr_FR	154,215,679
pl_PL	139,426,023
pt_BR	136,719,655

2

Metric	-243913656
en_US	3,219,310,973
en_GB	257,439,112
es_LA	67,123,147
de_DE	43,625,784
fr_FR	34,253,559
nl_NL	19,619,296
da_DK	15,320,853
es_ES	13,804,322

3

Metric	237055665
en_US	1,371,729,291
en_GB	94,482,627
es_LA	21,910,611
fr_FR	12,641,681
nl_NL	12,434,033
de_DE	8,406,154
es_ES	5,128,647
da_DK	3,668,672

4

Metric	562697744
en_US	2,628,475,643
en_GB	91,744,510
fr_FR	27,796,064
es_LA	24,299,733
nl_NL	7,190,585
de_DE	7,074,037
es_ES	6,065,477
pt_BR	5,127,717

5

Metric	1937545411
en_US	1,015,941,556
en_GB	607,331,366
es_LA	110,871,230
it_IT	45,204,460
pl_PL	43,628,028
nl_NL	38,963,016
fr_FR	38,093,868
de_DE	37,020,420

- These tables show the locale info for 5 selected top pages. Most of the impression are from en_US, and en_GB.
- The first four pages show consistent information with previous two slides, but, the fifth page got the most impressions from en_US which doesn't meet the expectation that it should get the most impression from en_GB.

Top 5 selected pages by gender and age

Page #3, 4 &5 have similar patterns that they attract M 18-34 more than F18-34.
Page #2 attracts a little bit mature groups M/F 25-44.
Page #1 attracts a little younger groups 18-34 and attracts more female than male.

Name

page_impressions_by_age_gender_unique

1

Metric	-1379668826
F.18-24	1,874,708,832
M.18-24	1,210,696,668
F.25-34	1,030,370,341
M.25-34	813,235,656
F.35-44	243,757,503
F.13-17	217,625,926
M.35-44	170,868,081
M.13-17	92,900,253
F.45-54	81,173,496
M.45-54	51,169,430
F.55-64	30,139,592
F.65+	25,565,194
M.65+	23,665,412
M.55-64	17,591,232
U.25-34	7,369,298
U.18-24	5,813,923
U.35-44	2,889,023
U.45-54	965,530
U.13-17	602,682
U.65+	430,002
U.55-64	415,559

2

Metric	-243913656
M.25-34	712,924,307
F.25-34	631,683,018
M.35-44	427,472,019
F.35-44	420,533,207
M.18-24	360,366,040
F.18-24	328,624,672
F.45-54	227,096,999
M.45-54	207,602,805
F.55-64	135,779,106
M.55-64	98,049,964
F.65+	96,066,546
M.65+	69,502,865
F.13-17	13,138,348
U.25-34	12,919,224
M.13-17	12,689,369
U.35-44	10,802,149
U.45-54	5,400,457
U.18-24	3,959,293
U.55-64	3,065,220
U.65+	2,736,283
U.13-17	145,529

3

Metric	237055665
M.18-24	389,126,684
M.25-34	367,868,283
F.18-24	257,602,768
F.25-34	208,405,002
M.35-44	114,690,366
F.35-44	70,616,215
M.45-54	33,838,881
M.13-17	32,476,110
F.13-17	25,912,899
F.45-54	23,693,199
M.65+	10,464,237
M.55-64	8,750,577
F.55-64	7,825,252
F.65+	6,310,566
U.25-34	2,360,343
U.35-44	1,133,326
U.18-24	975,736
U.45-54	339,088
U.65+	118,337
U.55-64	97,843
U.13-17	86,545

4

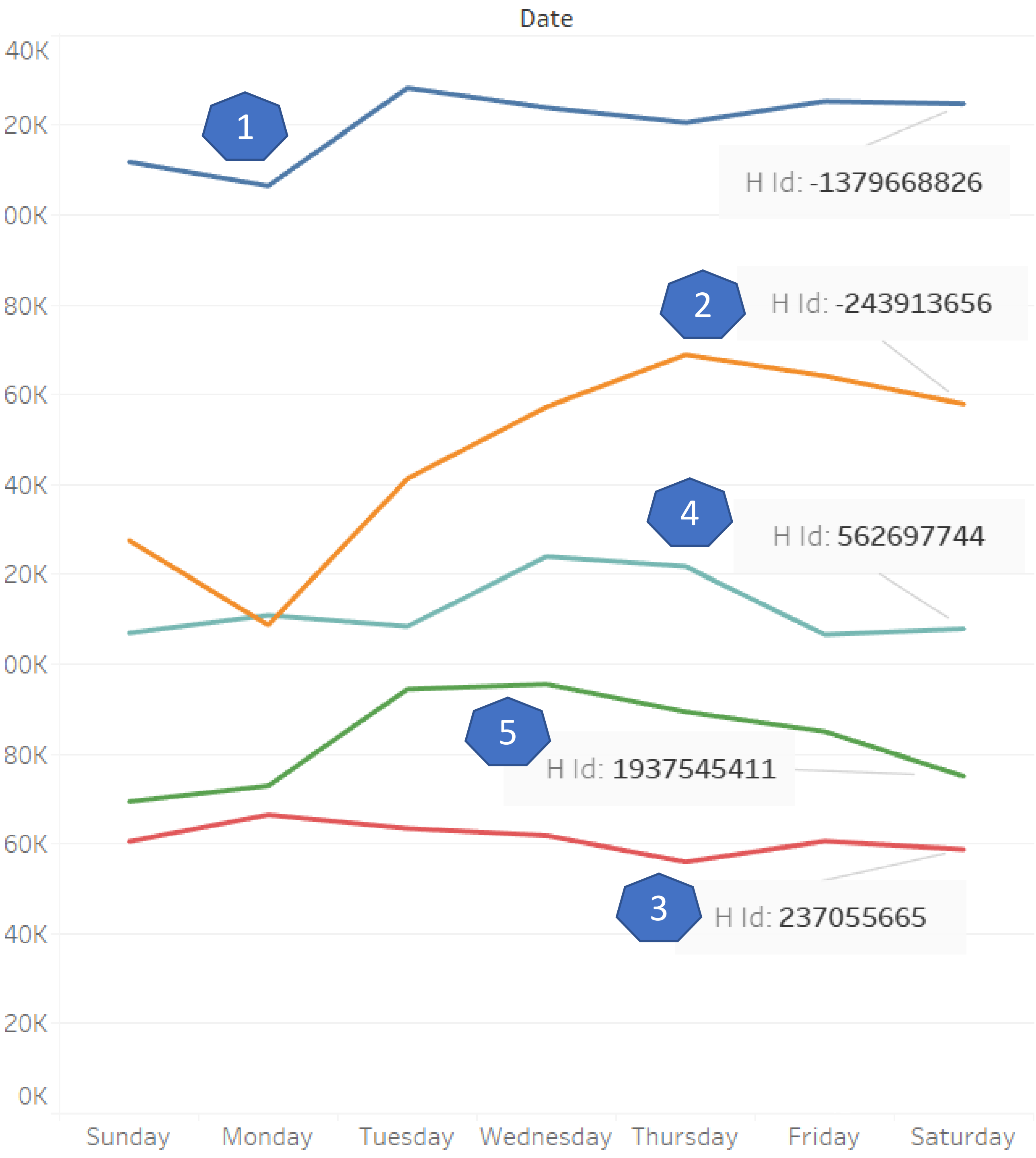
Metric	562697744
M.18-24	263,017,144
M.25-34	411,147,000
F.18-24	420,639,600
F.25-34	637,816,739
M.35-44	218,865,609
F.35-44	394,631,561
M.45-54	93,338,635
M.13-17	10,304,298
F.13-17	18,214,981
F.45-54	191,115,358
M.65+	23,763,082
M.55-64	27,511,302
F.55-64	66,046,943
F.65+	39,615,517
U.25-34	5,146,846
U.35-44	3,849,025
U.18-24	1,394,048
U.45-54	1,724,264
U.65+	473,446
U.55-64	580,794
U.13-17	72,454

5

Metric	1937545411
M.18-24	399,948,444
M.25-34	243,156,451
F.18-24	719,879,815
F.25-34	385,955,168
M.35-44	59,845,380
F.35-44	104,870,850
M.45-54	22,108,929
M.13-17	49,920,994
F.13-17	131,101,769
F.45-54	41,099,775
M.65+	7,087,808
M.55-64	7,550,267
F.55-64	14,375,622
F.65+	9,702,764
U.25-34	2,752,180
U.35-44	1,521,514
U.18-24	1,809,865
U.45-54	623,617
U.65+	173,549
U.55-64	234,037
U.13-17	274,900

The trend of the average Value by weekday

- Page #1 got the most impressions for all weekdays, while page #3 got the least impressions among them.
- Page #1, 2, & 5 have similar trend that the impressions increase from Tuesday over the week.
- Weekdays don't affect too much of the impressions. The page attract most traffic still have a lot of impression even in the lowest impression day Monday.



Average impression by age and gender

Name	Metric	H Id -2117389370	-2050455356	-1824102173	-1388002448	-1293571655	-740173518	1670602128
page_impressions_by_age_gender_unique	F.13-17		74,275	3	79	137,383	1,225	64
	F.18-24	1	607,910	16	1,690	289,367	11,885	630
	F.25-34	19	217,690	10	3,342	165,095	11,668	441
	F.35-44	77	72,584	5	3,665	66,750	5,766	170
	F.45-54	12	26,737	4	2,643	27,155	2,297	47
	F.55-64		12,145	2	1,352	13,738	733	11
	F.65+	1	5,167	2	802	12,047	528	8
	M.13-17		24,201	2	43	90,055	1,840	35
	M.18-24	2	319,202	15	1,601	206,489	21,812	393
	M.25-34	6	120,988	14	2,747	136,371	18,603	352
	M.35-44	37	36,876	5	1,600	43,849	5,509	127
	M.45-54	8	11,478	2	954	18,130	1,727	33
	M.55-64		4,554	2	454	8,045	477	10
	M.65+		2,514	2	562	8,294	604	12
	U.13-17		40	1	1	1,054	10	1
	U.18-24		494	1	4	2,099	70	5
	U.25-34		261	1	15	1,814	151	9
	U.35-44	2	112	1	20	889	92	8
	U.45-54		39	1	18	403	38	3
	U.55-64		20	1	9	190	17	2
	U.65+		15	3	7	217	16	1

- This table shows the average value of each age and gender group across 7 different random pages.
- It shows that some pages don't have information for certain groups.
- The 2nd page has very high average impressions for F.18-24.

IV. Conclusion

IV. Conclusion

Summary of two types of dataset:

- CPM dataset: the average CPM charged by Facebook to Viacom is $3.875 from the CPM dataset. Group F.44-54, M.44-54, F.34-44, U.44-54 has the highest average CPM.
- Page dataset: special events, demographics of the page could affect number of page impressions. Most of the impression are in the group F.18-24, F25-34, M.18-24, M.25-34. Most of the impression are in the US, UK, and Mexico. Specially, New York, Chicago, LA, Mexico City, and London.

Suggestions

Pricing strategy based on demographics:

- The goal is to get higher profits = Price*organic impressions – costs* paid impressions (simplified model). That is, more impressions at the same time higher price and lower costs will lead to more profits.
- If Viacom charges $25 per 1000 impressions, and the costs is $3.875 per impressions. Also, assume 2000 impressions from Viacom, and 1000 from Facebook purchased. Then Viacom can gain profits from this offer = 25*2000-3.875*1000 = $46125
- Ideally, to gain the most profits, Viacom should charge higher price to group F.18-24, F25-34, M.18-24, M.25-34 because these groups have the most impressions, and their average costs are lower than average.

Limitations and Opportunities

- When we analyze the dataset provided by Viacom. The data only tells us a limit amount of information. Important information to the pricing strategy is missing in the dataset. That is, the market, the supply and demand in the market. How does other competitors behave? What are the profiles of Viacom's advertisers and potential advertisers?
- Differentiate the advertisers and charge them different prices for different needs by providing differential offers (e.g. add more specific targeting) will increase the competitiveness among other suppliers.

References

- Asdemir, K., Kumar, N., & Jacob, V. S. (2012). Pricing models for online advertising: CPM vs. CPC. *Information Systems Research*, *23*(3-part-1), 804-822.
- Dorbian, I. (2009). Marketers Await Pricing Alternatives to Online CPMs. *Min's B2B*, *12*(33), 6. Retrieved from https://link.ezproxy.neu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=44006904&site=bsi-live
- Evans, D. S. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of Economic Perspectives*, *23*(3), 37-60.
- Fridgeirsdottir, K., & Najafi-Asadolahi, S. (2018). Cost-per-Impression Pricing for Display Advertising. *Operations Research.,66*(3), 653-672.
- Liu, D., & Mookerjee, V. (2018). Advertising Competition on the Internet: Operational and Strategic Considerations. *Production and Operations Management*, *27*(5), 884-901.
- Shen, Q., & Miguel Villas-Boas, J. (2017). Behavior-based advertising. *Management Science*, *64*(5), 2047-2064.
- Turner, J. (2012). The planning of guaranteed targeted display advertising. *Operations research*, *60*(1), 18-33.

Appendix: R codes

```
# Getting start, clean and prepare the datasets
# There will be two datasets: CPM datasets, and pages level datasets

#find the directory
getwd()

#if the directory is not in the desired folder, set the directory
setwd("C:/Users/minyi/Desktop/final")
# double check the directory again
getwd()

# put all the raw csv dataset into this directory folder
# read csv files, totally 18 csv files
d1701 <- read.csv(file="pages_level_demos_2017-01.csv")
d1702 <- read.csv(file="pages_level_demos_2017-02.csv")
d1703 <- read.csv(file="pages_level_demos_2017-03.csv")
d1704 <- read.csv(file="pages_level_demos_2017-04.csv")
d1705 <- read.csv(file="pages_level_demos_2017-05.csv")
d1706 <- read.csv(file="pages_level_demos_2017-06.csv")
d1707 <- read.csv(file="pages_level_demos_2017-07.csv")
d1708 <- read.csv(file="pages_level_demos_2017-08.csv")
d1709 <- read.csv(file="pages_level_demos_2017-09.csv")
d1710 <- read.csv(file="pages_level_demos_2017-10.csv")
d1711 <- read.csv(file="pages_level_demos_2017-11.csv")
d1712 <- read.csv(file="pages_level_demos_2017-12.csv")
d1801 <- read.csv(file="pages_level_demos_2018-01.csv")
d1802 <- read.csv(file="pages_level_demos_2018-02.csv")
d1803 <- read.csv(file="pages_level_demos_2018-03.csv")
d1804 <- read.csv(file="pages_level_demos_2018-04.csv")
d1805 <- read.csv(file="pages_level_demos_2018-05.csv")
cpm <- read.csv(file="cpm_estimates-10-24-18.csv")


# combine all the page level data set from 2017-01 to 2018-05
# totally 17 csv files
page <- rbind(d1701,d1702,d1703,d1704,d1705,d1706,d1707,d1708,d1709,d1710,d1711,d1712,d1801,d1802,d1803,d1804,d1805)

#summarize the entire page datasest from 2017.1 to 2018.5
page$hID<- as.factor(page$hID)
summary(page)
summary(page$hID)

#save csv files
write.csv(page,'page.csv')
```

```r
head(cpm,5)
table(cpm$hID)

cpm$hID<- as.factor(cpm$hID)
cpm$female<- as.factor(cpm$female)
cpm$male<- as.factor(cpm$male)
cpm$age_max<- as.factor(cpm$age_max)
cpm$age_min<- as.factor(cpm$age_min)

summary(cpm)

#create new column age
cpm$age[cpm$age_max==24 & cpm$age_min==18 ]<- "18-24"
cpm$age[cpm$age_max==34 & cpm$age_min==25 ]<- "25-34"
cpm$age[cpm$age_max==44 & cpm$age_min==34 ]<- "34-44"
cpm$age[cpm$age_max==54 & cpm$age_min==44 ]<- "44-54"
cpm$age[cpm$age_max==34 & cpm$age_min==18 ]<- "18-34"
cpm$age[cpm$age_max==44 & cpm$age_min==18 ]<- "18-44"
cpm$age[cpm$age_max==49 & cpm$age_min==18 ]<- "18-49"
cpm$age[cpm$age_max=="65+" & cpm$age_min==18 ]<- "18-65+"

# these two groups approximately to the related groups for better understanding
cpm$age[cpm$age_max==54 & cpm$age_min==45 ]<- "44-54"
cpm$age[cpm$age_max==44 & cpm$age_min==35 ]<- "34-44"

# create new column gender
cpm$gender[cpm$female==1 &  cpm$male==0]<- "F"
cpm$gender[cpm$female==0 &  cpm$male==1]<- "M"
cpm$gender[cpm$female==1 &  cpm$male==1]<- "U"
```

```
# create new column group for age and gender
cpm$group[cpm$age=="18-24" & cpm$female==1 &  cpm$male==0]<- "F.18-24"
cpm$group[cpm$age=="18-24" & cpm$female==0 &  cpm$male==1]<- "M.18-24"
cpm$group[cpm$age=="18-24" & cpm$female==1 &  cpm$male==1]<- "U.18-24"

cpm$group[cpm$age=="25-34" & cpm$female==1 &  cpm$male==0]<- "F.25-34"
cpm$group[cpm$age=="25-34" & cpm$female==0 &  cpm$male==1]<- "M.25-34"
cpm$group[cpm$age=="25-34" & cpm$female==1 &  cpm$male==1]<- "U.25-34"

cpm$group[cpm$age=="34-44" & cpm$female==1 &  cpm$male==0]<- "F.34-44"
cpm$group[cpm$age=="34-44" & cpm$female==0 &  cpm$male==1]<- "M.34-44"
cpm$group[cpm$age=="34-44" & cpm$female==1 &  cpm$male==1]<- "U.34-44"

cpm$group[cpm$age=="44-54" & cpm$female==1 &  cpm$male==0]<- "F.44-54"
cpm$group[cpm$age=="44-54" & cpm$female==0 &  cpm$male==1]<- "M.44-54"
cpm$group[cpm$age=="44-54" & cpm$female==1 &  cpm$male==1]<- "U.44-54"

cpm$group[cpm$age=="18-34" & cpm$female==1 &  cpm$male==0]<- "F.18-34"
cpm$group[cpm$age=="18-34" & cpm$female==0 &  cpm$male==1]<- "M.18-34"
cpm$group[cpm$age=="18-34" & cpm$female==1 &  cpm$male==1]<- "U.18-34"

cpm$group[cpm$age=="18-44" & cpm$female==1 &  cpm$male==0]<- "F.18-44"
cpm$group[cpm$age=="18-44" & cpm$female==0 &  cpm$male==1]<- "M.18-44"
cpm$group[cpm$age=="18-44" & cpm$female==1 &  cpm$male==1]<- "U.18-44"

cpm$group[cpm$age=="18-49" & cpm$female==1 &  cpm$male==0]<- "F.18-49"
cpm$group[cpm$age=="18-49" & cpm$female==0 &  cpm$male==1]<- "M.18-49"
cpm$group[cpm$age=="18-49" & cpm$female==1 &  cpm$male==1]<- "U.18-49"

cpm$group[cpm$age=="18-65+" & cpm$female==1 &  cpm$male==0]<- "F.18-65+"
cpm$group[cpm$age=="18-65+" & cpm$female==0 &  cpm$male==1]<- "M.18-65+"
cpm$group[cpm$age=="18-65+" & cpm$female==1 &  cpm$male==1]<- "U.18-65+"

#save csv files
write.csv(cpm,'cpm.csv')
```