Home Exam for Analyst Candidates

# Instructions and Background

This is an exam for an analyst position in Taboola. The purpose of the exam is to allow the candidate to experience some of the tasks that Taboola analysts perform and to allow common grounds for discussion in the next interview.

You are expected to answer as many questions as you can from the questions below and to be ready for a discussion in your next interview.

If you find that one of the questions is not strictly defined, you may use any assumption you need, but make sure to explain why you used them.

Feel free to approach Dafna Mittelman ([dafna.m@taboola.com](mailto:dafna.m@taboola.com)) for any question or problem.

# First part – data analysis

In the spreadsheet attached you will find a data set that the following questions refer to.

**The spreadsheet holds the following sheets:**

1. **sites** – a list of sites where you can find our widget integrated, and the following information for each site:

* id – a unique numeric key  
  name – a unique name for each site
* domain – the site’s domain – same domain can be shared between multiple sites
* time\_zone – the time zone of the site, based on its locale

1. **daily\_views** – number of page views for each site, broke down by day

* site\_id
* desktop/tablet/mobile views – number of page views performed over those devices
* desktop/tablet/mobile clicks – number of clicks performed by users on Taboola widgets using those devices
* desktop/tablet/mobile revenue – revenue from clicks performed over those devices

1. **source\_items** – a list of top viewed items per site

* source\_views – total page views on the specific page
  + E.g. – when a user visit http://www.bloomberg.com/news/2013-01-31/-pivot-to-asia-sinking-on-a-coral- reef.html, we count a single source\_view for item 5913147824592673317.
* source\_clicks – total clicks on Taboola widget from users who visit that page
* source\_revenue – the total revenue from users which clicked our widget on the specific page

**KPIs Definitions:**

* **CTR** – a measurement of the probability for a click, calculated by total clicks divided by total views
  + E.g. – a site has 50,000-page views1[[1]](#footnote-1) and our widget was clicked 3,000 times, the CTR would be 3,000/50,000 = 6%
* **RPM** – the average revenue for 1000 views – calculated by total revenue, divided by total views, multiplied by 1000
  + e.g. – a site has 50,000-page views and the total revenue from clicks on Taboola widget was 200$, the RPM would be 200/50,000x1000 = 4$
* **CPC** – the revenue (from an advertiser perspective it would be the cost) for a single click on its item.
  + E.g. – if the widget was clicked 3,000 times and the total revenue was 200$, the average CPC would be 200/3000=0.067$ - for each click Taboola earned 0.067$ in average

## Questions:

1. Based on daily\_views sheet:
2. Which 3 sites have the highest percentage of mobile traffic, out of their total traffic?

**Answer: Business week, Weather channel, Wonderwall are top 3 sites having highest percentage of mobile traffic.**

**Explanation: To find the percentage of mobile traffic I had done grouped the mobile views according to their site\_id and then calculated the percentage of mobile\_views. As we don’t have any specific method to calculate the mobile traffic I assumed it can be calculated according to the views.**

1. Which site has the highest RPM in total? Which site has the highest RPM per platform (desktop, mobile, tablet)?

**Answer:**

* **Highest RPM in Tablet: Dailymail**
  + **Highest RPM in Desktop: huffingtonpost**
  + **Highest RPM in Mobile: The blaze**
  + **HuffingtonPost is having the Highest RPM.**

1. Based on source\_items sheet:
2. Which source items are the most clickable per site in matters of CTR (3 items per site)?

**Answer: Below is the table which is the most clickable site in matters of CTR (3 items per site)**

**Method of doing: The analysis is done in Python by calculating the CTR and then done group by on the site\_id and then converted that into a data frame and stored as excel file.**

|  |  |
| --- | --- |
| **site\_id** | **source\_url** |
| **1** | **http://www.bloomberg.com/news/2014-06-24/merkel-said-to-ready-summit-for-full-sanctions-on-russia.html** |
| **1** | **http://www.bloomberg.com/news/2014-06-24/suarez-accused-of-biting-italian-defender-in-world-cup-victory.html** |
| **1** | **http://www.bloomberg.com/news/2014-06-25/bitcoin-auction-draws-wall-street-silicon-valley-bidders.html** |
| **2** | **http://www.boston.com/sports/basketball/2014/06/24/source-lebron-will-become-free-agent/zWyrG9NQDdwuxa9FYULBqK/story.html** |
| **2** | **http://www.boston.com/business/news/2014/04/27/wegman-chestnut-hill-opening-draws-crowd/XpSf7KayQqeiu4y9YSdZoJ/pictures.html** |
| **2** | **http://www.boston.com/sports/baseball/redsox/2014/06/06/calling-all-outfielders-the-red-sox-need-some-help/eQTa3uJDNpPVHCMMQNewJO/pictures.html** |
| **3** | **http://www.businessweek.com/videos/2014-06-24/the-cavendish-banana-a-fungus-among-us** |
| **3** | **http://www.businessweek.com/videos/2014-03-04/is-google-android-more-secure-than-apple-ios** |
| **3** | **http://www.businessweek.com/videos/2014-06-23/u-dot-s-dot-europe-will-release-oil-on-120-brent-croft** |
| **4** | **http://www.theblaze.com/blog/2014/06/30/obama-asks-congress-for-authority-to-more-quickly-deport-unaccompanied-children/** |
| **4** | **http://www.theblaze.com/blog/2014/06/27/stu-burguiere-ceos-make-a-lot-of-money-so-what/** |
| **4** | **http://www.theblaze.com/blog/2014/06/27/harry-reid-signs-up-for-rand-pauls-voting-rights-bill/** |
| **5** | **http://www.politico.com/multimedia/video/2014/06/joe-biden-defends-gay-rights.html** |
| **5** | **http://www.politico.com/multimedia/video/2014/06/john-boehner-obama-hasnt-faithfully-executed-laws.html** |
| **5** | **http://www.politico.com/multimedia/video/2014/06/nancy-pelosi-john-boehner-lawsuit-is-subterfuge.html** |
| **6** | **http://www.businessinsider.com/google-buying-startups-2014-1** |
| **6** | **http://www.businessinsider.com/11-amazing-thank-you-notes-from-famous-people-2012-2** |
| **6** | **http://www.businessinsider.com/4-things-software-engineers-can-teach-you-about-project-management-2014-6** |
| **7** | **http://www.weather.com/news/pennsylvania-sisters-die-when-tree-falls-car-20140624** |
| **7** | **http://www.weather.com/video/economy-in-peril-from-global-warming-50324** |
| **7** | **http://www.weather.com/health/ebola-epidemic-west-africa-called-deadliest-history-20140624** |
| **8** | **http://videos.huffingtonpost.com/518284155** |
| **8** | **http://videos.huffingtonpost.com/518284097** |
| **8** | **http://videos.huffingtonpost.com/518284189** |
| **9** | **http://www.examiner.com/article/upcoming-aaliyah-biopic-will-feature-r-kelly-marriage-details-family-tries-to-s** |
| **9** | **http://www.examiner.com/article/teen-wolf-extended-promo-what-s-coming-up-season-4** |
| **9** | **http://www.examiner.com/article/retinol-or-retinoid-creams-the-anti-aging-cream-you-must-have** |
| **10** | **http://www.dailymail.co.uk/tvshowbiz/article-2667051/All-tennis-whites-Katherine-Jenkins-wears-light-dress-yellow-blazer-leads-celebrity-guests-day-two-Wimbledon.html** |
| **10** | **http://www.dailymail.co.uk/news/article-2667065/Daft-Charlie-Brooks-hid-large-lesbian-porn-stash-downed-pint-washing-liquid-cure-hangover.html** |
| **10** | **http://www.dailymail.co.uk/sport/worldcup2014/article-2667084/Luis-Suarez-retracts-revenge-comments-knocking-England-World-Cup.html** |
| **11** | **http://msn.foxsports.com/detroit/story/ian-kinsler-returns-to-texas-with-first-place-tigers-062414** |
| **11** | **http://msn.foxsports.com/buzzer/story/lebron-james-opts-out-memes-062414** |
| **11** | **http://msn.foxsports.com/southwest/story/duncan-back-for-spurs-in-2014-15-062314** |
| **12** | **http://wonderwall.msn.com/entertainment/celeb-birthdays-for-the-week-of-april-1-7-1673952.story** |
| **12** | **http://wonderwall.msn.com/movies/the-notebook-where-are-they-now-29386.gallery?gt1=28135** |
| **12** | **http://wonderwall.msn.com/movies/the-shortlist-for-jan-9-17419.gallery** |
| **13** | **http://www.reuters.com/article/2014/06/23/bbo-orioles-teamreport-idUSMTZEA6NLNAFDT20140623** |
| **13** | **http://www.reuters.com/article/2014/06/24/us-column-miller-seniorhealth-idUSKBN0EZ19Y20140624** |
| **13** | **http://www.reuters.com/article/2014/06/24/markets-usa-stocks-open-idUSZXN0R7B0520140624** |

1. Which source items are having the highest RPM (3 items per site)?

Answer:

|  |  |  |
| --- | --- | --- |
| **site\_id** |  | **source\_url** |
| **1** |  | **http://www.bloomberg.com/news/2014-06-24/suarez-accused-of-biting-italian-defender-in-world-cup-victory.html** |
| **1** |  | **http://www.bloomberg.com/news/2014-06-25/obama-administration-widens-export-potential-for-u-s-oil.html** |
| **1** |  | **http://www.bloomberg.com/news/2014-06-25/yahoo-s-mayer-says-she-apologized-for-lateness-to-cannes-dinner.html** |
| **2** |  | **http://www.boston.com/news/local/massachusetts/2014/06/24/boston-woman-charged-double-pedestrian-fatal/9nOu40euF3r0rBEpdT0oKJ/story.html** |
| **2** |  | **http://www.boston.com/sports/basketball/2014/06/24/source-lebron-will-become-free-agent/zWyrG9NQDdwuxa9FYULBqK/story.html** |
| **2** |  | **http://www.boston.com/news/local/rhode-island/2014/06/23/bills-legalize-pot-nix-healthsource-stall/RFUydpdi02vMMgPAECBfkM/story.html** |
| **3** |  | **http://www.businessweek.com/videos/2014-06-24/how-to-play-a-45-million-stradivarius** |
| **3** |  | **http://www.businessweek.com/videos/2014-06-23/u-dot-s-dot-europe-will-release-oil-on-120-brent-croft** |
| **3** |  | **http://www.businessweek.com/videos/2014-06-24/walgreens-withdraws-fiscal-2016-goals** |
| **4** |  | **http://www.theblaze.com/blog/2014/06/27/overstock-com-joins-wounded-warrior-project-in-their-commitment-to-injured-service-members/** |
| **4** |  | **http://www.theblaze.com/blog/2013/01/03/was-joe-biden-drunk-today/** |
| **4** |  | **http://www.theblaze.com/blog/2014/06/30/gop-warns-next-va-chief-has-huge-culture-change-mission-ahead/** |
| **5** |  | **http://www.politico.com/multimedia/video/2014/06/a-view-from-the-hill-campaign-pro-panel.html** |
| **5** |  | **http://www.politico.com/multimedia/video/2014/06/mississippi-primary-preview.html** |
| **5** |  | **http://www.politico.com/multimedia/video/2014/06/congressional-leaders-hold-hands-and-sing.html** |
| **6** |  | **http://www.businessinsider.com/june-conf-board-consumer-confidence-2014-6** |
| **6** |  | **http://www.businessinsider.com/r-bollore-aims-to-make-vivendi-an-industrial-media-group-2014-24** |
| **6** |  | **http://www.businessinsider.com/elizabeth-warren-2016-presidential-speculation-2014-6** |
| **7** |  | **http://www.weather.com/video/economy-in-peril-from-global-warming-50324** |
| **7** |  | **http://www.weather.com/news/berlin-zoo-panda-bao-bao-dies-20120822** |
| **7** |  | **http://www.weather.com/video/outdoor-water-efficiency-50315** |
| **8** |  | **http://videos.huffingtonpost.com/518284155** |
| **8** |  | **http://videos.huffingtonpost.com/518284230** |
| **8** |  | **http://videos.huffingtonpost.com/518284097** |
| **9** |  | **http://www.examiner.com/article/sahlen-s-six-hours-at-the-glen-june-26th-through-29th-2014** |
| **9** |  | **http://www.examiner.com/article/transformers-2-delivers-another-summer-disappointment** |
| **9** |  | **http://www.examiner.com/article/polk-county-ga-arrest-report-dated-june-22-2014** |
| **10** |  | **http://www.dailymail.co.uk/money/bills/article-2665978/RACHEL-RICKARD-STRAUS-Applying-new-boiler-funding-minefield.html** |
| **10** |  | **http://www.dailymail.co.uk/sport/worldcup2014/article-2667063/David-Gea-Xavi-say-goodbye-Madrid-Airport-following-Spains-group-stage-exit-World-Cup-2014.html** |
| **10** |  | **http://www.dailymail.co.uk/news/article-2085237/Daniel-Tyndale-Jail-baby-Banksy-wrecked-1m-damage-Bristol.html** |
| **11** |  | **http://msn.foxsports.com/southwest/story/duncan-back-for-spurs-in-2014-15-062314** |
| **11** |  | **http://msn.foxsports.com/buzzer/big-buzz/lebron-james-opts-out-of-miami-heat-contract-062414** |
| **11** |  | **http://msn.foxsports.com/soccer/story/jozy-altidore-to-miss-us-germany-world-cup-game1-062414** |
| **12** |  | **http://wonderwall.msn.com/movies/the-notebook-where-are-they-now-29386.gallery?gt1=28135** |
| **12** |  | **http://wonderwall.msn.com/tv/bindi-irwin-wants-other-teen-girls-to-cover-up-says-shes-an-86-year-old-soul-1821801.story** |
| **12** |  | **http://wonderwall.msn.com/music/gossip-12-16-11-britneys-beau-pops-the-question-17010.gallery** |
| **13** |  | **http://www.reuters.com/article/2014/06/24/bulgaria-corpbank-pm-idUSR5N0NF01B20140624** |
| **13** |  | **http://www.reuters.com/article/2014/06/24/wonders-information-co-brief-idUSL4N0P52XD20140624** |
| **13** |  | **http://www.reuters.com/article/2014/06/24/idUSWLB008F420140624** |

1. A site named business week, complained that its RPM significantly dropped in the second part of June. The site’s AM asked for the following analysis:
2. When did the drop start?

**The drop started on the date 24-06-17 and continued till 24-06-29. The drop is recorded between these dates.**

1. What caused the drop (based on the data in daily views table)?

**There is no significance cause to find the drop has occurred as I have observed on some days the revenue is low and on some days the views are low on the second part of June. May be this caused a drop in the RPM. In the second week of June some of the days have recorded less revenue and less views may be this is the reason for total drop in the RPM.**

1. Businessweek’ wants to know the estimated revenue on 13-19.7.2014 and the most fruitful day. Provide a revenue forecast for the entire week and the most fruitful day.

Answer:

**By doing the revenue forecast for the entire week 13-19.7.2017 the estimated total\_revenue would be between 105000 to 110000 (approx.).The most fruitful day would be 2014-07-16 as it has recorded highest revenue in the month of June.**

1. Disregard the previous data, suggest as many reasons as you can find which may result with RPM drop for a site. What data would you need in order to support your suggestion
2. **(Bonus)** Disregard the previous questions, Taboola’s BI team decided to make a dashboard for helping the AM’s to improve the publisher’s revenue, by detecting unusual events with regards to revenue, page views and clicks such as RPM drops. The dashboard should help the AM to detect and analyze those unusual events in order to increase the customer’s revenue.

Draw a **mock** dashboard for this scenario.

\*No need to make up your own tables, base it on the existing fields in the provided data set  
\*Assume that the AM will use the dashboard for a single customer each time

\*Remember! The AM is not an analyst

\*We’re only looking to see an infrastructure, no need for live data

# Second part - SQL

1. Each item in a web shop belongs to a different seller. To ensure service quality, each seller has a rating.

The data is kept in the following two tables:

**Table sellers:** id, name, rating

**Table items:** id, name, sellerId

Write a query that selects the item name and the name of its seller for each item that belongs to a seller with a rating of more than 4

**Answer: Select i.name [Item Name], s.name [Seller Name]**

**From items i**

**Inner join sellers s on i.sellerid = s.id**

**Where s.rating > 4**

1. The following data defines employees’ details in a company:

**Table employees:** Id, name, managerId (references employees(id)), department, salary

1. Write an update statement which adds 25% salary for each employee in IT department

**update employees**

**set salary = salary + (salary \* 25)/100**

**Where department = ‘IT’**

1. Write an update statement which sets the salary of all employees in sales department to the value of the average salary in IT department

**Declare @avgsalary float**

**Select @avgsalary = avg(salary) from employees where department = ‘IT’**

**Update employees**

**Set salary = @avgsalary**

**Where department = ‘Sales’**

1. Write a query that selects the names of employees who are not managers;

An employee is a manager if any otheremployee has their managerId set to the first employee id. An employee who is a manager may or may not also have a manager

**Select e.name**

**From employees e**

**Left join employees m on e.id = m. managerId**

**Where m.id is null**

1. Write a query that selects the following details for each employee: name, department, salary and employee’s salary rank within his/her department (i.e. the highest paid employee in each department will be ranked ‘1’, the second highest will be ranked ‘2’ etc.)

**Select name, department, salary,**

**Row\_number() over(partition by department order by salary desc) as [Rank]**

**From employees**

1. Write a query that selects all employees whose salary is ranked 3rd in their department

**Select**

**From**

**(**

**Select name, department, salary,**

**Row\_number() over(partition by department order by salary desc) as [Rank]**

**From employees**

**)res**

**Where [Rank] = 3**

1. Application usage data is kept in the following table:

**Table sessions:** Id, userId, duration, data\_timestamp

Write a query that selects userId and average session duration for each “Good User”.

Good User - a user with an average of at least 5 sessions in a week

**Select userid,**

**datepart(week, data\_timestamp) wk,**

**Count(id) as sessioncount**

**Into #usersessionsperweek**

**From sessions**

**Group by userid, datepart(week, data\_timestamp)**

**Select userid**

**Into #listusers**

**From #usersessionsperweek**

**Group by userid**

**Having avg(sessioncount) >= 5**

**Select u.userid, avg(duration) as average\_duration**

**From #listusers u**

**Inner join sessions s on u.userid = s.userid**

**Group by u.userid**

1. **Page Views** – each time a user opens a web page where our widget is integrated, count as a single page view [↑](#footnote-ref-1)