# Powerinbox Data Science and Analytics Team Lead Exercise

This exercise should not take more than one to two hours to complete. The purpose of this exercise is to gauge your analytical skills with different data sets and to be able to give a rough analysis of trends and possible action items to improve overall revenue. The numbers in these files are fake.

**Publisher Hierarchy.**

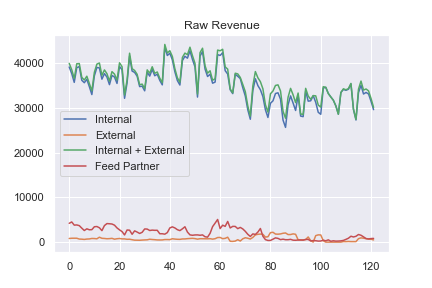
Publisher -> Container -> Placement -> Native Position

Individual ads show in the native position of the publisher placements.

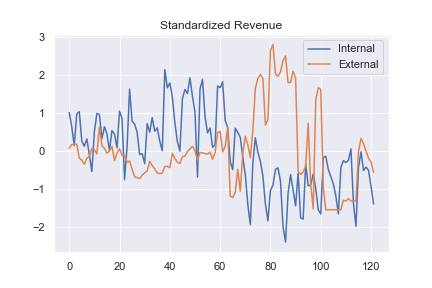
The network has three different types of traffic with their own unique properties. Additionally, the network has multiple types of demand. The three types of demand are Direct, OpenRTB, and feed based. In the files given, "internal" is a mix of direct and feed based while external is OpenRTB. A feed partner file has been given to separate the distinct feed based numbers.

From the data given, you must provide answers to these questions. It is important to give your reasoning within the answers and how you came up with it.

**1. We started seeing a dip in revenue a couple of months ago. What is the likely source of revenue decline and what should we do about it?**



There is a sharp decline in the month of May and has since seemed to level off. This could be due to seasonality which is difficult to see without more data. The plot above shows how the revenue is clearly driven more from the internal demands.



At a glance, by standardizing the revenue values, we can see that during the decline in revenue, there was an increase in external demand. Even though the scale is much smaller, perhaps OpenRTB can lead to insights on how to prevent revenue loss from the higher source demands.

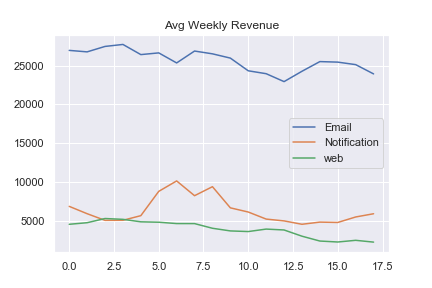


Another thought I had was to look at the general market trends over time. By quickly researching I can see a dip in the stock market that might be correlated to the dip in revenue. More research into this may be insightful as well.

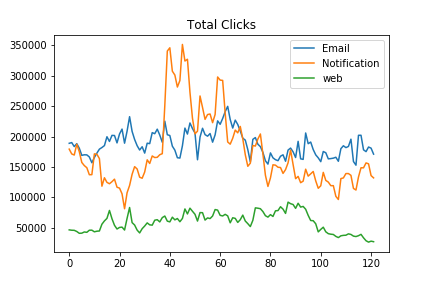
**2. What are the overall trends for each traffic type?**



Th plot above displays the revenue over time for each traffic type. There appears to be a dip in all three types. To try and clean up some of the distortion, plot weekly average revenue.



All three are generally decreasing.



The same is true with the daily total clicks.

**3. What data is missing from this analysis and what additional data points would be needed to provide a better action plan of next steps?**

More granular data is needed here. The data is an aggregation which provides a summary for each day. In order to find relevant trends, we would want to know the details of each revenue source such as the type of ad it was, the contents, demographics, location, etc. In my current position, I take information about specific news articles and try to find trends in traffic based on the content. Something similar to this approach might work with more data points.

**4. The operational side of the network is unable to directly interface with the data and reporting systems and have a series of reports that need to be run and scheduled reports that must be sent daily.**

**4a. How would you prioritize publisher and advertiser requests for data that is operationally unable to be obtained?**

In my current roll, I do have a lot of asks that require data that is difficult or impossible to obtain. I usually try to find out what knowledge they are trying to gain from the reports. What is essential for their decision making. At that point I discuss with the stakeholder any alternative data sources to use that may be strongly related to the unavailable data that they need. Something that may typically trend (some historical data analysis may be needed) with the data they want to track.

**4b. What is your process on handling this situation from an overall and on a day to day perspective?**

If data is unobtainable, there’s a good chance it will stay that way. I think that working multiple data sources into your reporting may be a good way to move forward with the reporting on a day-to-day basis. Building multiple data sources into a pipeline can help track trends overtime and fill in gaps where just one data source might be lacking.