**Abstract**

Artificial intelligence, including clustering and neural networks could provide a better balance sheet forecast accuracy for Contra Revenue team at a company level.

**Introduction**

Contra Revenue is the second biggest line of the Profit and lose statement of the company, it includes all the spend occurred during a certain period related in its vast majority (by definition) to sell incentives and product returns, which would naturally deduct money from the gross sells of the company.

Since its origins, HPQ had managed the Contra Revenue spent through the business division of the company, without a major control or scrutiny revisions to it, letting the business managers and directors decide what was better to offer for a certain deal or a certain customer, amount or percentages allowed to return by partner, etc.

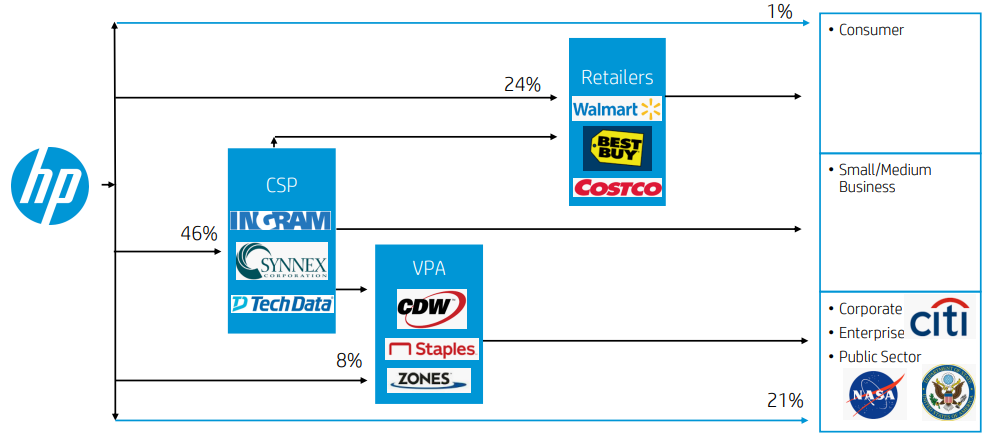
During 2010 HPQ company’s CEO Mark Hurd was involved in a set of sex and money fraud scandals. According to Michael Holston, HP's general counsel: *"The investigation revealed numerous instances where the contractor received compensation and/or expense reimbursement where there was not a legitimate business purpose. And the investigation found numerous instances where inaccurate expense reports were submitted by Mark or on his behalf that intended to or had the effect of concealing Mark's personal relationship with the contractor."*[[1]](#footnote-1)

After this inconvenient company’s event, and because the deviation of expenses occurred through a Contra Revenue account, that whole operation of discounting, sells incentives, returns and any other related account moved to create a new team under Controllership supervision, whit an improved scrutiny, audits and several controls, and we can now say 12 years later, this strategy had worked better so far.

After the split of HPQ back in 2015[[2]](#footnote-2) into two new companies, HPE and HPI (Hewlett-Packard Enterprise and Inc respectively), the Contra revenue operations remain under the Controllership supervision at least for the HPI portion of the new company, that its main core business are the manufacturing and reselling of computers and printers at a world-wide level.

As it was mentioned the Contra revenue spend is high, second biggest line of the P&L statement, and the reason of this is because HPI does not have an own distribution channel, so the majority of the incentives are given as a backend incentive to the biggest partners that redistribute and resell HP’s product in to the market or the final customer.

Diagram of HP’s Go to Market:



At the moment of the sell in, or from HP to the redistributor partners, and because of the US GAAP requirement, every spend associated to each sell needs to be recognized at the moment of the sell, so at the moment of the sell in we need to accrue for a certain percentage to recognize the spend in the P&L statement as well as to recognize the future obligation (debt) in the Balance sheet, this is the portion that needs to be forecasted from a Contra revenue team stand point.

By definition on HP’s accounting manual, contra revenue is:

Channel Sales incentives typically refers to programs used to stimulate sales, particularly sales of our channel partners which can include distributors, resellers, system integrators, retail partners and others.

Sales incentives are offers from HP that can be used by a customer to receive a reduction in the price of a product or service.

The only contra we accrue for is the backend portion, but every sell has a percentage of upfront discounting associated to it as well:

Upfront Contra – sales incentive that is offered and granted to the customer at the time of invoicing as part of the pricing of the products or services.

Backend Contra – that is offered and granted either the sale or after the initial sale of products and which need to be claimed subsequently by the customer by submitting the claim.

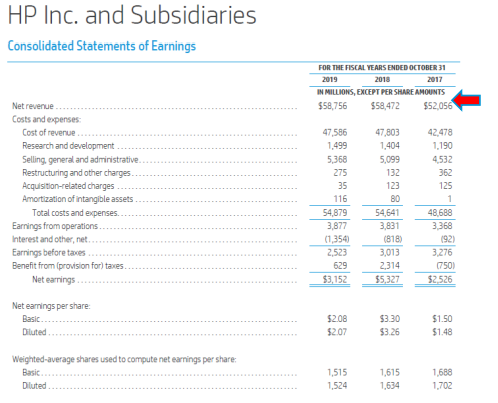
The main contra types are:



The biggest ones are 3104, 3103 and 3105.

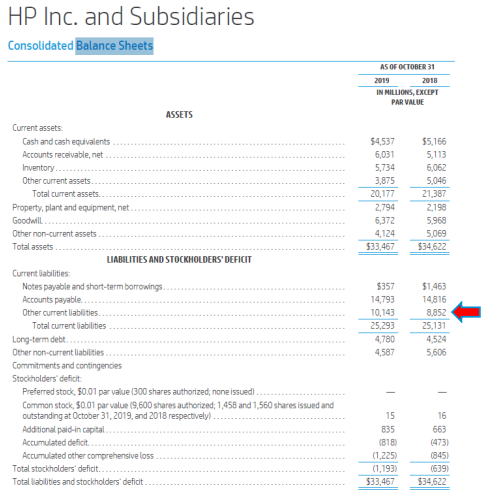
Location of Contra in the P&L statement:

It comes right before Net revenue, it is not published because it is part of the company strategy.



Location of Contra in the Balance sheet statements:

It is embedded in Other current liabilities, it is not published as an individual line, same because it is part of the company’s business strategy.



**Problem statement**

The Balance sheet portion of the Contra reserves need to be forecasted at least 2 times every quarter, the requirement comes from the Financial Planning and Analysis (FP&A) team of the company and the final purpose of this forecast is for them to be able to forecast the cash flow of the company.

Currently we are facing three main issues with this process:

-Granularity: All the set up of contra rates forecast and actual revies of the company are being done at a market level, and this forecast’s granularity today is only made at a global level, by business (computing and printing) meaning we have 2 main forecasts by account, and we need to have 20 forecasts by account (there are 10 markets subdivision for the company currently).

-Accuracy: The aimed required accuracy for out short-term forecast is for it to be below 5% error, currently we are sitting at a historical of 5.2%, with this change to market and artificial intelligence implementation, the idea is to improve accuracy for it to be below 3%.

-Automatization: The actual process is a massive excel sheet, semi-automated with a lot of manual inputs to it, the goal would be to automate it, reduce manual inputs and increase its autonomy.

**Contextualization**

FP&A gets several forecasts every month from various HP teams and based on Balance sheet variations they calculate the final cash flow number for CFO. The basis of this inputs is FP&A estimation of sell in for each period and based on that we run our modeling.

The global reserve in balance sheet of Contra revenue fluctuates in average around 3.5B USD, there are certain exceptions for the forecasting of reserves for cash flow purposes, and this forecast’s average is around 3B.

Currently we handle 4 models to determine these levels of reserve and decide which one makes more sense by the time we deliver the number. We could say there are two main models and two secondary models. As the requirements vary from deliverable to deliverable, we need to be prepared to provide any time horizon reserve FP&A needs.

Main models

* Model. Plainly named model, it is the most common for short term forecast, meaning when we are forecasting following’s month reserve or maybe reserve in 2 months, if we consider this for a wider horizon it doesn’t work very well. It is a cumulative model, meaning its base is the most current actual reserve and from it we accumulate new reserve or spend associated to every period sell, and discount our estimation of payments or claims for each period as well. Formula is:

*New reserve = Previous reserve + (sell in x contra rate by month) – payments*

Contra rate gets calculated based on historical behavior and since last year it gets compared against business expectations forecast.

For two of the biggest contra accounts (3104 and 3105) there’s a report that handle payment profiles and based on a historical of this report we predict future behavior, for the rest of the accounts that get forecasted and this report is not available we handle a linear regression to predict the next level of payments.

* Budget. It predicts the level of reserve of each Quarter end month based on historical level of reserves and must current business behavior, it is a more conservative model, and it is used for mid-term time horizon, meaning for the next year or so mostly. Formula is:

*New reserve = Sell in for the Quarter x contra BS rate*

The BS rate doesn’t require additional inputs, as it contemplates only the level of reserves, it doesn’t need individualities for spend and payment.

Secondary models

* Model Q. This is the least used model, but it is always running in parallel to have a backup level of reserve to triangulate numbers against the main Model, they are both very similar, the only difference is on how we calculate the contra rate percentage. Formula is:

*New reserve = Previous reserve + (sell in x contra rate by Quarter) – payments*

The contra rate gets calculated historically following trend changes between similar quarters and adapting it to the closes business expectations, like the rate calculated by the Model forecast.

The payment input is the same as the Model forecast.

* LTP (Long Term Planning): There are very few times we use this model, because the requirement comes very sporadically, we use this one for time frames after a year,3 years, 5 years and 10 years forecasts for last month of every fiscal year. For this forecast, currently we only run the Excel formula ‘=FORECAST()’ with a certain amount of historical data. This formula inside Excel computation uses an exponential smoothing model, that assumes same behavior of the future as in the past, conservative and useful for this long term non very required forecasting.

**General and specific objectives**

Granularity

The current granularity, as explained before, is WW (1), by business (2) by account (6), meaning we have 12 forecasts running at the same time in all the modeling.

Due to a recent change in how finance is structured, actually sin 2021, it is a must to go down to a Market forecasting, currently there are 10 markets, not all of them as big nor as important, an maybe we will end up having only 7-8 markets forecasted, but that is potentially going from 12 forecast to 120 individual forecasts.

There are two main options to approach this change:

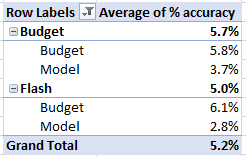
-Tops down approach. Either keep considering the actual models, refine them and allocate reserves to the market behaviors according to history or most recent business/markets expectations based on sell in or any other factor that we find out relevant.

-Bottoms up approach. Build the reserve individually from scratch based on each individual market expectation of spend, sell in and/or payment, at the end consolidate these reserves to have a grand total WW reserve.

Both approaches would have both views, consolidated and allocated by market, this will make much easier the analysis of variations, deltas, new accuracy by market, and it would be better to highlight anomalies in forecasting, or focus on very specific attention points, besides and among all the forecast process will be in the same language as all the new reviews by market level.

Accuracy

Currently the accuracy of the forecast is at a total level of 5.2% with very good accuracy for the short term and a little bit worse on the mid-term deliveries as the table below shows:



Budget model is used for mid-term, Model is used for short term.

Under the grand scheme of accuracy parameters, we are under the green flag, but the idea of granularity, more than making it worse, with the AI planned to be involved in the future calculations we aim to be in the 3% range.

The error calculation was made at a total absolute level.

Automatization

The process is currently semi-automated in Excel, and as stated before it has a lot of manual inputs and supervision, so we always have to come back and give the number a sense of realness, from it was born the idea to introduce a certain AI checks that allow the model to run automatically and produce analytics itself, even learn from history and previous mistakes to not repeat them again and know what to look for, what to do, etc.

With this whole new market structure, there are new spreadsheets updated on a monthly basis feed with market expectations, from which the model could make several checks or even leverage from them.

The clustering analysis is another possibility, some of the markets or the accounts behave similar, and this will potentially reduce the number of forecasts needed to be provided.

The end state of the process is to have a flexible model or set of models that reflect the actual business/markets expectations, with very low supervision.

**Descriptive data analysis (actual)**

In many large multinational companies such as HPI, there is often data issues, due to incomplete prior projects, multiple unaligned systems, and pretty much because there are lots on lots of data everywhere; even a major director described us (HP employees) as *data junkies* once because of the vast data and maybe not enough analysis.

Providing this context, we can now say that some of the process to complete the data needed for the Contra Balance sheet forecast is treated manually before it gets processed.

There are two main sources of information, the actual RP of the company that reports the Balance sheet, and there is the P&L source that tracks the monthly impacts and revenue.

Internally Contra Revenue team is divided into three sub-teams being:

-Delivery: Pretty much in charged of Mont End Close activities, hard core processes, interaction with all stakeholders, the face of contra to the outside.

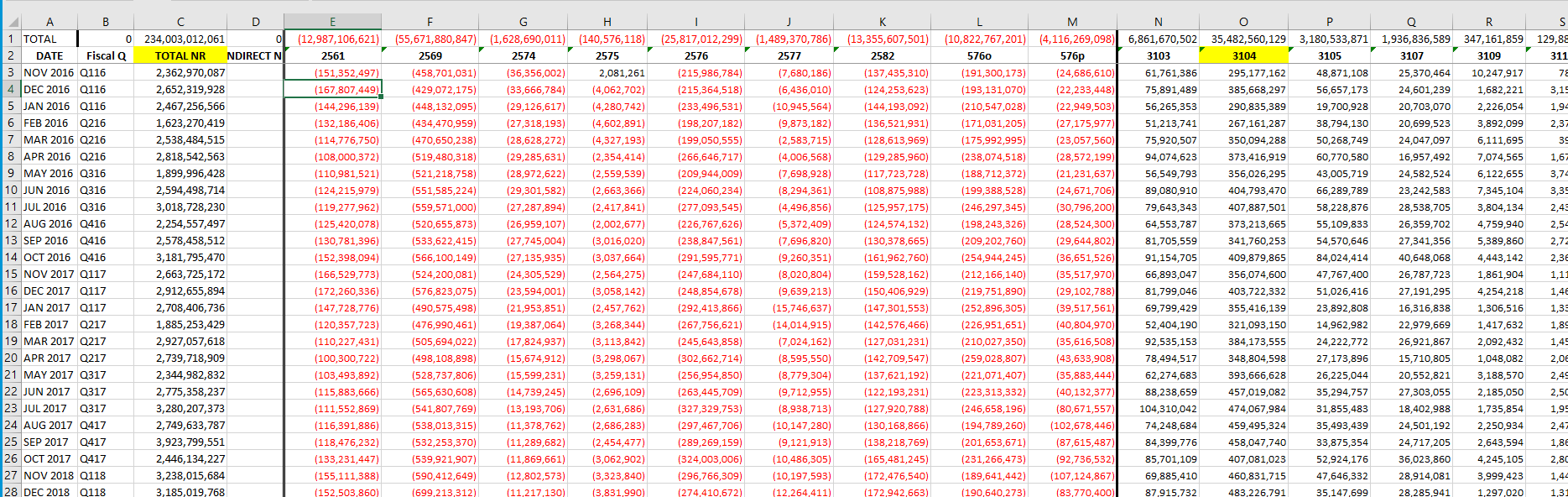
-Design: Internal management of processes, looking for improvements, expert owners of reporting, master data, definitions, documentation.

-Analytics: We call it that, but it is pretty much the dashboard’s team, the maintain and constantly create new views adapting to business requirements, a lot of power BI and ad-hoc reporting.

*Balance sheet and P&L data*

Delivery oversees the Balance sheet forecast, and the Analytics team is the one that provides the raw reporting data in order to begin with the process.

As for now, as explained in the introduction, the forecast is made at a WW level, by business by account, and that is how data is manually stored today:





\*Two tabs containing exactly the same data in the same order, one for PS and one for Print business.

Historical sin Nov 2016 (Nov is the first month of HPI’s fiscal year) by business, containing revenue data, and data by accounts, respectively BS and P&L.

We ca say the data is:

-Complete: we have no NULL values, there has always been either a reserve or an impact for every account by business by month.

-Valid: The data is pulled, populated and validated by the Analytics team.

-Stored: Safely stored in a Teams group folder, and in One drive.

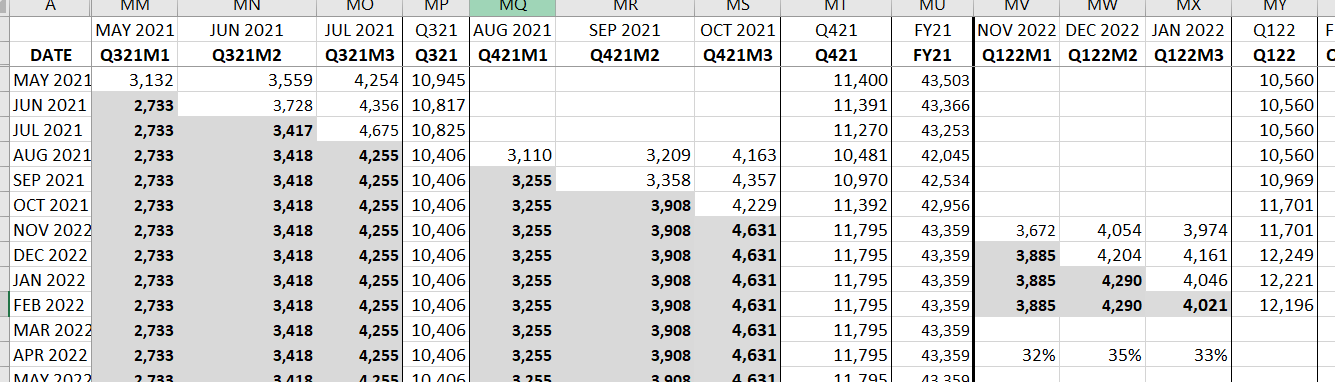
-Handy: Easy access to the storage folders.

-Reliable: Even when is manually processed in order to get to the view shown above, after adding it, there is a validation process that makes data reliable.

-Enough: With this data there is enough history and enough columns to process the forecast.

*Linearity data*

There is another set of really important data that we call linearity, also stored by PS and Print, where we keep track of previous revenue estimations; on the lines we keep the month of the estimation, on the columns the month estimated, and how it varied along time, it also has the inter-quarter distribution of expected sales, and then when the data turns from flash to actuals, it gets updated too.



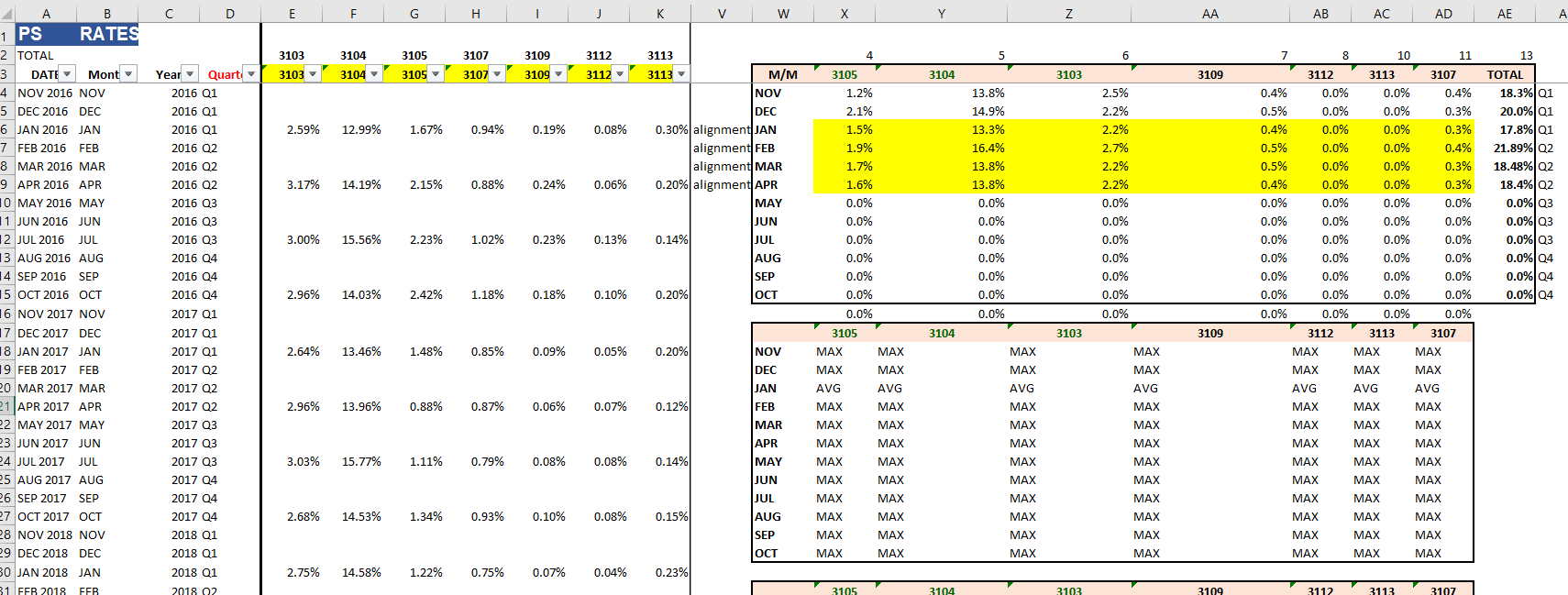
Very useful set of data that has been also used to analyze variations between periods, example: if the revenue is higher in % at the end of one Quarter in comparison to another that could mean more BS accrued for in that specific period.

This data is provided directly from FP&A team, on a monthly basis and gets collected in the tabs as shown in the image.

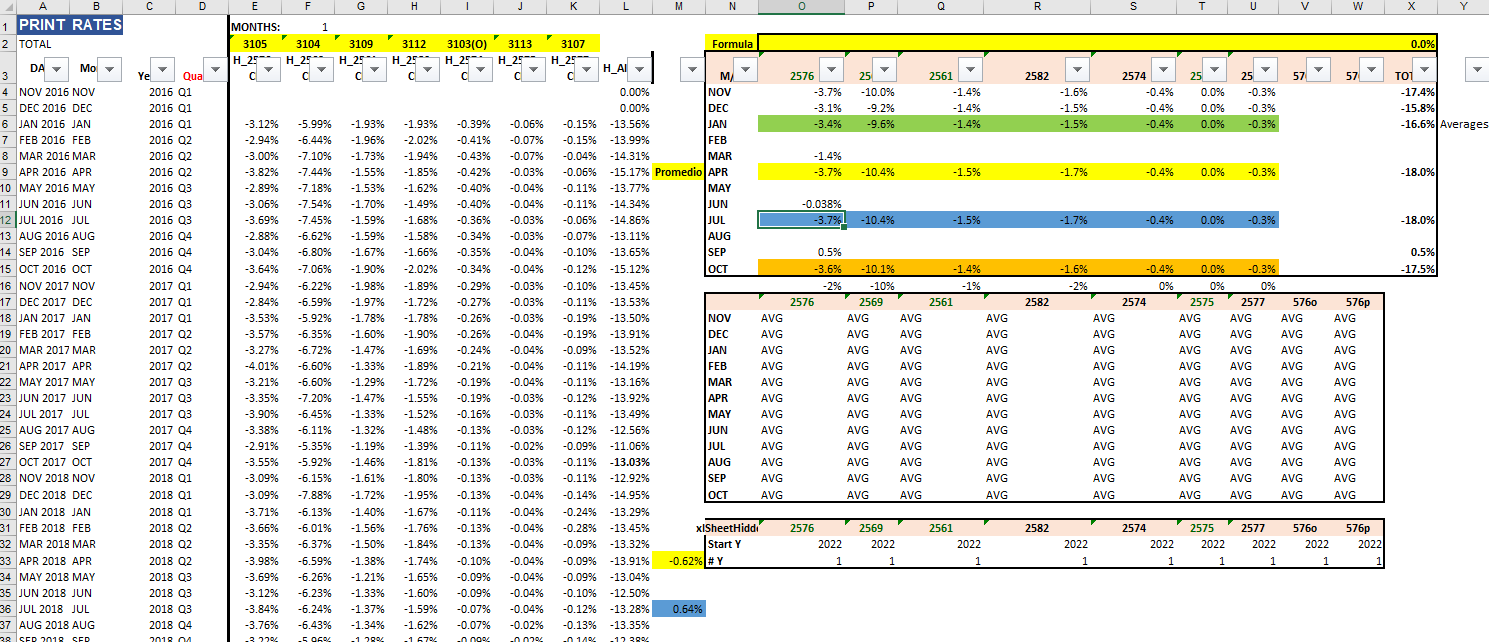
*Rates data*

With all the data inputs from before we manually create our own set of data for rates, specially for the Model M, Model Q and Budget.

Historical spend rates by business by account, with an interaction table where you can set either MAX, MIN, or AVE depending on the expectations and set the time frame of this calculations, for a minimum of 1 year and no maximum.



Also, a set of rates is calculated for the historical BS position on respect to the Revenue of the period:



Similar selection table to adapt to the forecast on every period.

*Payment data*

The last set of data is on the payments, there is no actual source to get this data in the company, because it would require a lot of consolidation, so we just drove it from the formula of the reserve:

*New reserve = Previous reserve + (sell in x contra rate by month) – payments*

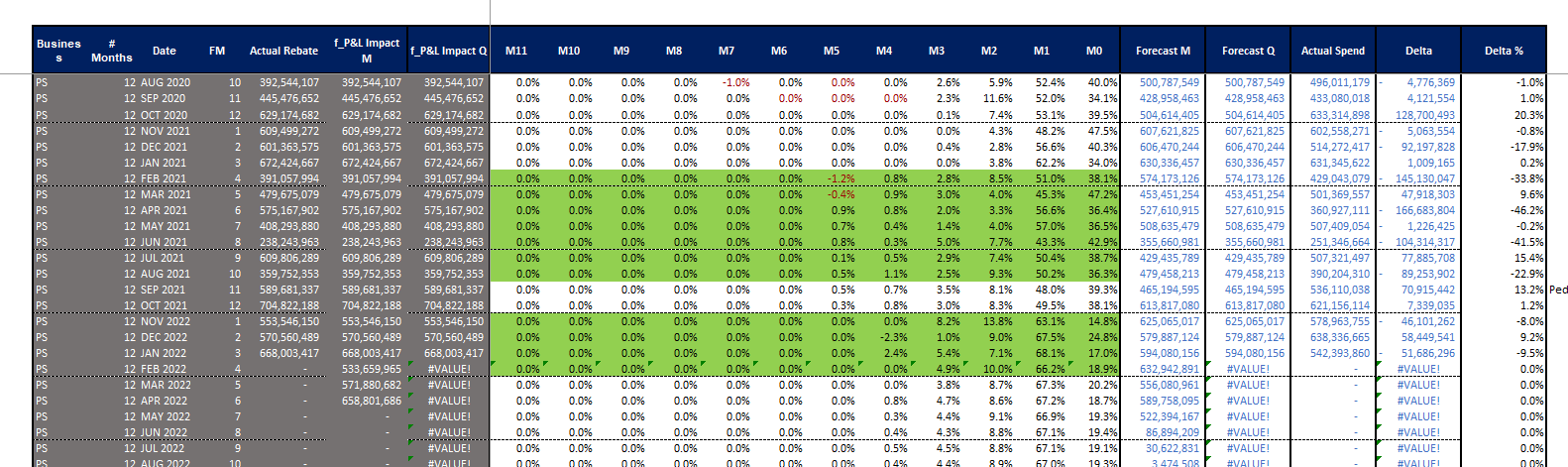
Therefore:

*-New reserve + Previous reserve + (sell in x contra rate by month) = payments*

This is how this data set is calculated and created, and there are two methodologies to forecast payments:

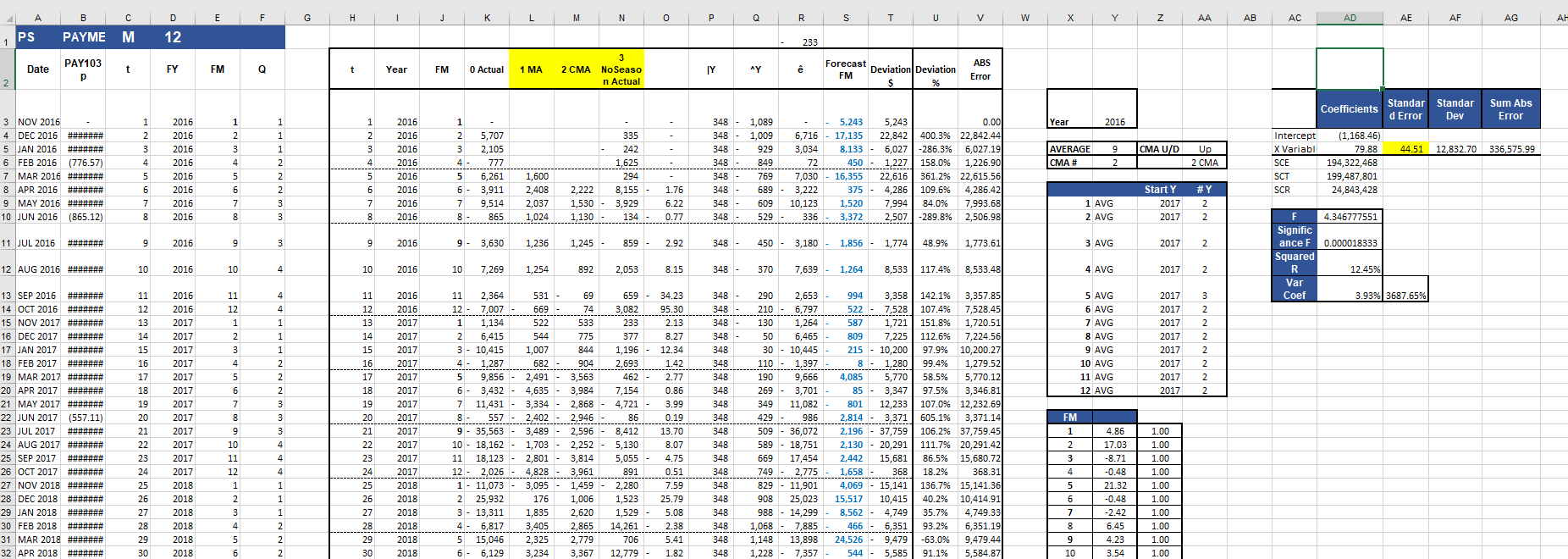
-Payment profile: There is a report that only contains two accounts in it (3104 and 3105) where you can identify the period to which the payment pertains to, and therefore you can calculate the future payments by projecting this payment profile distribution, good news is that the two accounts with this data are the largest, but unfortunately there isn’t available for all of them.

Showing an example of a payment profile payment forecast:



-Linear regression: For the rest of the accounts where we don’t have any basis to predict payments, we base our forecast on a linear regression and historical created values.

Showing an example:



It is a basic linear regression, with moving averages to normalize them, tested to be significative with the period of time took as historical, with a seasonality factor added to it by month by year, it has shown good accuracy when forecasting in the short term (1-3 months).

We have described the source of the actual data needed to run the forecast process. And now we will concentrate on the alternatives of sources for the new data in order to supply the objectives of this project, which would be two options:

**Data for the project**

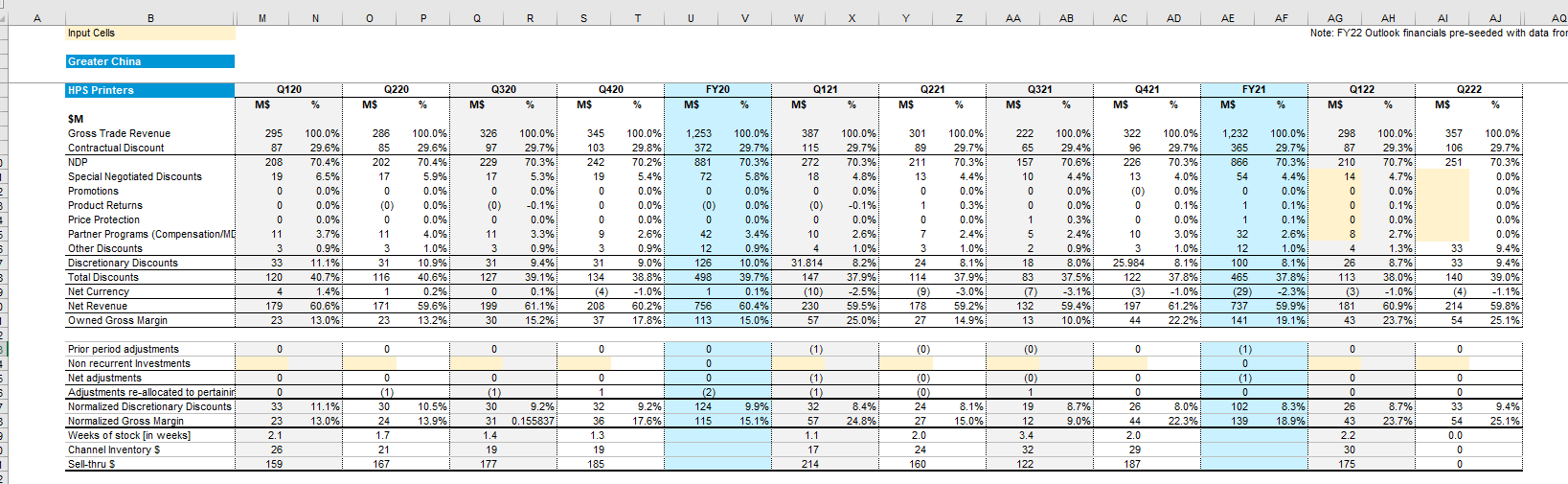
*Source data by Market*

The data is available by market with the same splits for the project in the same system where the Analytics team goes and pull the actual data used for the forecast process. But there are some inconveniences with the completeness of the data as there are some months where in small markets there isn’t any spend, not precisely an inconvenience as the data is what the data is.

The date would work the same from the source, manual (now automated) process to consolidate it and have enough history to start projecting with the new tools.

*Flash data*

There is the availability of this flash information by market where we could get the spend rate, and the forecasted revenue at the same time:



The idea with these files would be to consolidate them in a SharePoint, and then just access the necessary data by the granularity desired, the source is directly from market and business representatives, and it is only part of the flash process for margin.

1. # *5 reasons HP's Mark Hurd resigned*: https://www.computerworld.com/article/2519954/5-reasons-hp-s-mark-hurd-resigned.html?page=2

   [↑](#footnote-ref-1)
2. # *Hewlett-Packard Completes Spin-Off Of Enterprise Business:* https://www.forbes.com/sites/joecornell/2015/11/03/hp-hpq-completes-spin-off-of-enterprise-business/?sh=20535f16b4a6

   [↑](#footnote-ref-2)