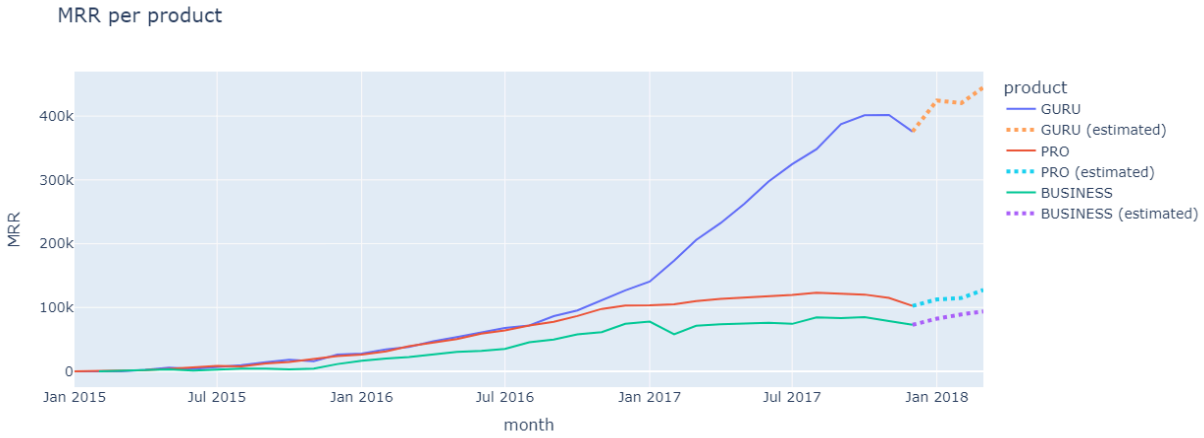
Semrush MRR forecast

We have analyzed transactions data from years 2015 to 2017 in five different countries and for Semrush products PRO, GURU and BUSINESS and then built a baseline forecasting model for one, two and three months ahead. The predictions for the months coming after the observed period are shown in the figure below.



The model at this point is just intended as a demo and could serve as a blueprint from where to start working on improvements. Due to time constraints, we have not performed a numerical evaluation of the uncertainty or confidence intervals associated with the estimations[1]. However, evaluation metrics during testing are far from satisfactory, especially when involving correlation or normalized errors with respect to actual values.

The model used is a simple linear regression, so there is certainly room for improvement. However, we don't have a strong belief that the improvement can be huge under the same scenario. Two main issues can be related to this thought:

- One is that working with monthly aggregations results in a very low number of samples. Since the source data is actually much more granular, with exact timestamps for the transactions, there is the chance to model them previous to aggregation.
  Also, having data for several more years could have a great impact on the performance. Note that the performance evaluation was done during the last months of the provided data, which has a strong drop in subscriptions that data from previous years does not fully represent.

- 1, 2 or even 3 months in advance is a timespan subject to high uncertainty. Observing only the transaction data is not enough to understand trends or changes and for a model to learn them. Certain dependency with the calendar has been

[1] We have forecasted the MRR indirectly using forecasts of subscription counts, fraction of renewals, and also average revenue per user. Since we used linear regression for the former two (and a simple average for the latter) confidence intervals on those can be provided right away from the error variance. However, the MRR has a large component that is deterministic coming from yearly subscriptions and a product between the estimations which makes it more elaborated. Note also that very few samples were used for error evaluation so the estimate of the error is also not very reliable.

observed but does not seem strong enough to have a big impact. Other observable variables related to customer behavior and market could be taken into account. Also, business decisions and changes should be reflected as features that a model can learn from. For example, we have a strong suspicion that something happened regarding the automatic renewal of subscriptions at the end of the observed period. However, note that if the information is also dynamic we will have to either rely on the observations (1,2,3 months behind) or on domain expertise or business plan expectations.