



Discoveries are the results of a deep exploration of the data by Berrijam AI.

Combining human expertise with Berrijam Discoveries lead to one of three things:

1. Discovery of something new or interesting that inspires ideas or exploration.
2. Confirmation of experience or instinct about the goal, now supported by data.
3. Clarity about the limits of the data or factors for a particular problem.

The Discoveries are structured in four sections:

1. Overview
2. Key Factors
3. Key Segments
4. Model Results

If ever in doubt, you can always request a consult with Berrijam AI Coaches.

1. Overview

The overview highlights some the interesting things found in the exploration and set the context for the rest of the discoveries.

Sanity Check

If you can't beat random guessing, then the data and factors were not good predictors. Time to reconsider the data you are collecting or the factors you are expecting to be useful.

Quick Segments

Most Likely Yes/No is selected based on both how large is the segment and the percentage of the segment that is Yes/No. This is where you can have the biggest impact.

Overview

There is a **Strong Signal to predict Churn** - the best model on this data had +36.22 F1 Score[Ⓛ] compared to guessing.

Most Likely Yes:
Contract is 'Month-to-Month' and Number of Referrals<=2.0 and Online Security is 'No' and Premium Tech Support is 'No'

Churn = 65.35% (973 of 1489)

38.58% More Likely 'Yes'

Segment Size = 24.64% (1489 of 6043)

Most Likely No:
Contract is one of: ['One Year', 'Two Year']

Churn = 6.51% (191 of 2936)

20.26% Less Likely 'Yes'

Segment Size = 48.59% (2936 of 6043)

Explored 37 factors over 6043 records, in which 26.77% had Churn is Yes (1618 of 6043).

Of these factors:

- 1 was strongly correlated with Churn, so was excluded.
- 13 were relevant for predicting Churn.
- 23 were found to be irrelevant or redundant, so were ignored.

Also generated predictions for the 1000 rows where the Churn column was empty.

2. Key Factors

Berrijam AI evaluates each of the columns as a possible factors in predicting the goal. The table lists factors in order of importance, along with the values at which outcomes is more likely 'Yes' when considered by itself.

Key Factors

We analyzed each factor (column) in the data, looking for which factors matter and which are irrelevant in determining Churn.

Strongly Correlated Factors - In this dataset, we found some factors to be so strongly correlated with other factors that we could essentially swap them out and keep only one of them.

The Churn column was Strongly Correlated With: Customer Status, so this factor was excluded.

Key Factors - The most important factors and their values more likely to result in 'Yes' for Churn are:

	FACTOR	SINGLE VALUE GROUP ①	GROUP BREAKDOWN
1	Contract	Contract is 'Month-to-Month'	19.16% More Likely 'Yes' Yes 45.93% Size of Group = 51.41% (3107 of 6043)
2	Number of Referrals	Number of Referrals<=0.0	6.2% More Likely 'Yes' Yes 32.97% Size of Group = 54.16% (3273 of 6043)
3	Online Security	Online Security is 'No'	15.2% More Likely 'Yes' Yes 41.97%

Focus & Prioritize

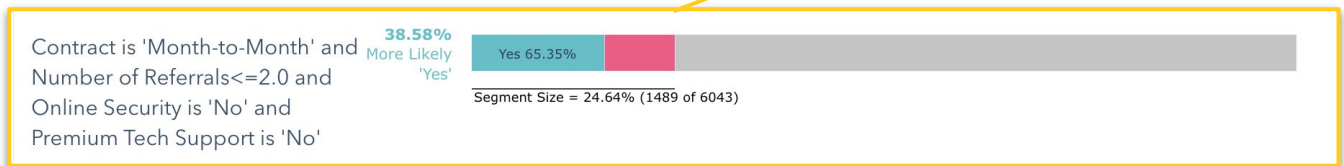
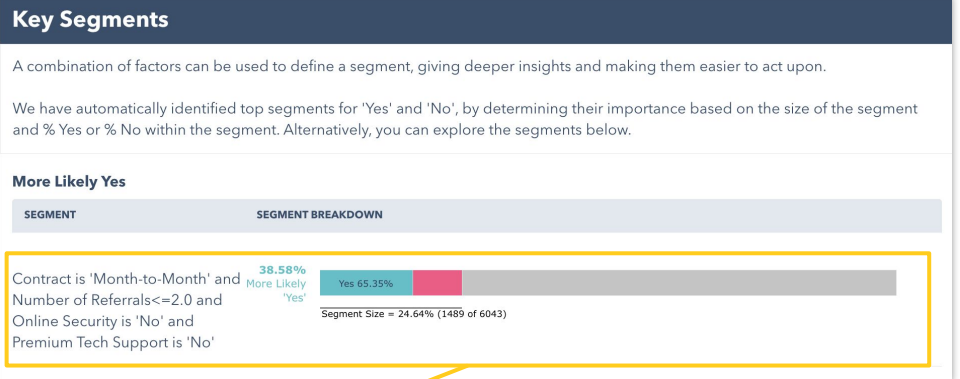
Some factors matter, some are irrelevant, and others are statistically mirrors of each other. These insights can help you prioritize which pieces of information or data should you invest in managing, or prioritize certain groups based on single factor.

2. Key Segments

Often a combination of multiple factors are more interesting. Segments, defined by a combination of factors and their values, highlight subset of population such as customers, patients or circumstances under which Yes or No are most likely.

Segments are prioritized based on a combination of their size as well as the percentage of Yes or No.

This helps you focus on the most important sub-groups within the dataset.



↑

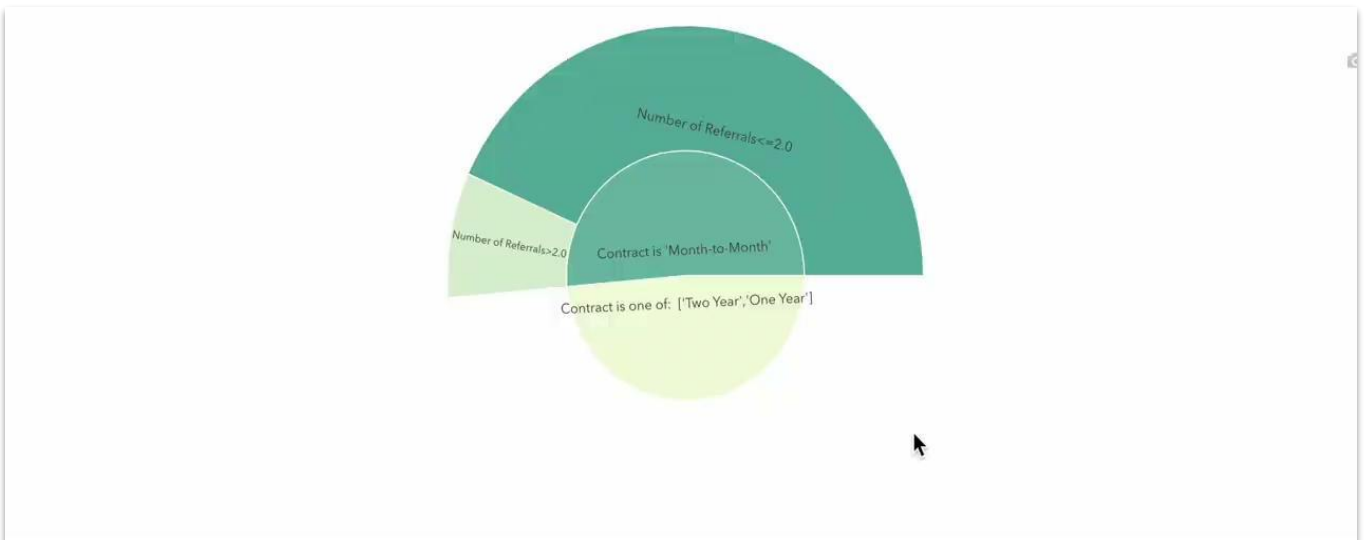
Segments are based on multiple factors and conditions in a way most people can understand.

↑

Most segments will have a combination of Yes and No. The gray bar represents the size of the full dataset (i.e. rows or records in the data explored). The colored parts represent the size of this particular segment, so you can get a visual sense of the size of the segment with respect to the full population.

Berrijam AI automatically highlights the most interesting segments for both Yes and No outcomes. Often these are clues to identify specific areas to focus or to come up solutions targeted to these sub-groups.

You can also interactively explore segments through **Segment Rings**, which organizes segments in a hierarchy.



Navigate down to lower-level segments by clicking on the segment rings, and navigate back up to higher-level segments by clicking in the centre. The video above shows an example. Notice that as you hover, on a segment, it show you the conditions that define the segment, percentage of records (or segment size) and the percent 'yes'.

4. Model Results

Berrijam AI explores hundreds of models of different types and parameters, which fit into two categories - Explainable and Non-Explainable.

● Explainable Models

Ability to understand and derive human friendly insights above is based on explainable models. If transparency, trust and insights matter, then explainable models are the way to go!

● Non-Explainable Models

Berrijam AI also explore models that are very hard (if not impossible) to easily understand, but are sometimes better. We call these non-explainable models, and use these where ranking and predictions of individual rows is more important than explainability.

Model Results

We explored **626** different models to find which produced the best results.

For the rows without a value for Churn, we generated predictions using the best models. We will auto-delete these on **2024-03-22** so please download them before then, otherwise you will need to re-run the exploration.

The Best Explainable Model **performed very well** on this dataset:

+30.19 F1 Score[Ⓛ]
compared to guessing

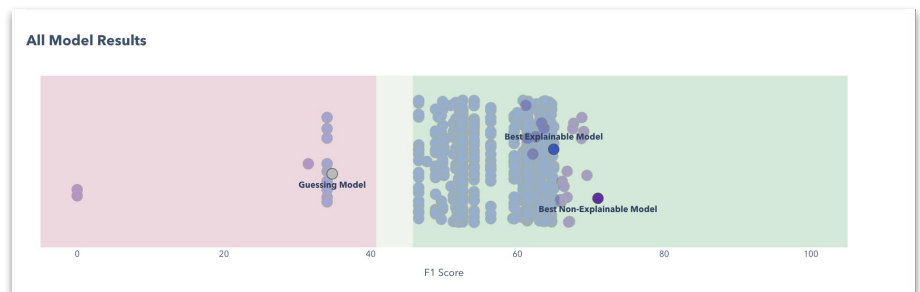
Download Predictions

The Best Non-Explainable Model **performed very well** on this dataset:

+36.22 F1 Score[Ⓛ]
compared to guessing

Download Predictions **Recommended**

The top Non-Explainable model performed well, but none of the Explainable models achieved a similar performance. If you only care about the Predictions, we recommend using the Non-Explainable predictions. However, if you want to understand how the models arrived at the predictions, we recommend revisiting your data or setting up a consultation with one of our data scientists for better Explainable Model results.



Each dot represents the performance of one predictive model, along the axis of F1 Score. **F1 Score** is a good way to assess model's balance between **precision and recall**, and is also more robust to imbalance between Yes and No examples in the dataset.

Guessing Model - Since models are only as good as the data, we evaluate each model against a model that makes random guesses on the outcome. If the models perform no better than guessing, then the data and the factors are not useful, and it is time to fix the data or explore different factors.

Berrijam AI generates insights using the best explainable model, but also compares those against the best non-explainable model.

Predictions - If the dataset has rows where the goal is empty (i.e. not labelled yes or no), then Berrijam AI automatically generates predictions for those rows using the best Explainable and Non-Explainable Models.

Data security is top of mind at Berrijam, and so we automatically delete the predictions generated within 1-week.

Download Predictions **Recommended**

The top Non-Explainable model performed well, but none of the Explainable models achieved a similar performance. If you only care about the Predictions, we recommend using the Non-Explainable predictions. However, if you want to understand how the models arrived at the predictions, we recommend revisiting your data or setting up a consultation with one of our data scientists for better Explainable Model results.

Still unsure or need some help? Contact Berrijam AI Coaches