As a short summary for the discussion on LTV prediction I'd like to share my ideas that may be useful for the future development as well as from current data collection standpoint

We may have 2 target variables for LTV prediction:
1) Monetary value of LTV
2) Quantitative value of LTV

Each one has advantages and disadvantages and may need slightly different approaches and variables to be included into consideration and model.

As for **monetary value of LTV** it is price-sensitive and there should be an agreement on the definition of “price”
We need also take into account that price can be influenced by campaigns (discounts, bundles and so on), market conditions, as well as cost of shipment/production/storing can be included or excluded from the "price".
So based on who is the stakeholder of the result the definition may vary a lot.

As for the **Quantitative value** it could be a bit easier taking into account all the above mentioned "price" uncertainties. In this task there is one additional layer of complexity as we need to take into account the amount of “main” product (food) as well as profit-maximisers (treats, toys, etc.) which are also a part of overall LTV. The limitation here is that there is a certain “natural” maximal capacity or spending limit if we consider rational user as one animal cannot eat more than a certain median amount of food per day which can be somehow estimated based on size-weight-breed information

And if consider only prediction of consumption of main product then features that could be relevant (but some may appear not to be):

- Breed or breed group
- Size (small, big, medium)
- Weight (it is correlated to bread and size, but can bring another dimension)
- Body mass index (BMI) (like size/weight)
- Age (in months)
- Age group (kitten, adult)
- Sex
- Is Sterilized
- Number of pets in the household
- Time with us (new, old customers)
- Features that can tell how user is involved in having a pet (may be collected through survey, questionnaire):
    - Does or does not the owner has a pet insurance
    - Uses pet hotels
    - Uses pet daycare services
- Information on the medical conditions of the pet
    - is a decision to switch to healthy food related to owners decision or as a result of pet’s health issues
    - Vet recommendation to change food
    - what kind or issues
    - for how long, etc.
    - time since last
- Information about the owner:

- How healthy food habits the owner has: is vegan, keto or any other food preferences
- Previous history of purchases
    - Had subscription
    - Has cancelled subscription
    - Time from cancelled subscription
    - Payment method or number of methods, etc.

Last thing to mention here is that there could be different approaches on how we define the prediction horizon (6m or 1 year).
If we set N first months and apply model to users that haven't reached this horizon yet, then at each time of "prediction" the user base will consist of "older" and "fresh" users and quality of the prediction for "older" will be better as we have collected more historical behavioral information about the user, which is not the case for "fresh" ones.

It's obvious that the quality depends on features and maturity of the user and the model itself. If we use “Fact+Prediction for the rest of horizon as LTV then:

- For new users fact is small portion of data, so prediction period is almost all horizon, so error is much bigger obviously –
    - so some generic features like category (Apple or Andriod) or “average” for some categories (like over city, country, etc) may help a bit.
- For users that almost completed horizon prediction is a tiny portion of LTV as we know all the data already and the error is tiny...

Depending on the maturity of the market same model can work differently as we average error of users of different maturity... and can misinterpret the prediction error results.

The development of the model depends, first, on the goal and how one is going to utilize prediction and, second, on the available data, definitions and model type/quality.

But definitely company can benefit from such modeling.