

## Data Files received:

### 1. 'Final Email List Anonymized 2\_19\_2019.csv'

This file contains total number of survey links sent (email ID's). It has 73195 observations and 13 variables.

IPN's or unique policy holders – 73,396

Contracts or policies – 10,4161

Split of policies:

Life Insurance – 62585

Disability – 20956

Annuity – 20620

In 73195 observations, 73033 are one-many ANON\_IPN-ANON\_Contract combinations. Rest 162 have multiple values in either IPN or Contracts or both. We are not sure on what basis to unroll (because 2 IPN and 2 Contract would lead 4 combinations out of which only 2 will be correct).

*(For our analysis, we understand that you provided a list of unique survey key data, but if further analysis on complete email listing is needed, it is better to understand the unrolling method. If you think it's not necessary, we can disregard.)*

### 2. 'Final Survey Themes Masked without Original Keys 2\_19\_2019.csv'

This file contains survey responses received and their corresponding email list details. It has 6855 observations and 50 variables (40 from survey and 10 from email list)

IPN's unique – 6,861

Contracts unique – 10,097

For 11 lines within survey data, the email list data is not merged properly leading to 11 missing values. Below are the responses corresponding to those missing values:

R\_2amoJPHS8CDfjed, R\_1QyKRASixWwMk5o, R\_08LUQ5uDo3XiFPz, R\_2xEJy6om1ocX9XB,  
R\_3npaoW5zd7IzRL, R\_3CC4983SVJ7XiCE, R\_1mD819y8V5ZjGqF, R\_yPyLpw1mQT1EJ33,  
R\_23gJeJ2WPHmeoVs, R\_1mOuquqPtHUXb7u, R\_1HvRuuOYIBZkjpT

The above responses have IPN and contracts in old survey data provided. So, do you want us to map them from old survey data? **Please confirm**

**Commented [JK1]:** – is this mistyped? Is it 10,416 or 104,161?

**Commented [JK2]:** You would look at all of the combinations and compare to the sales key with the corrected combinations. Any null combinations would then be removed since they would not be correct. For example, I have IPN's A and B, Contracts 1 and 2. **Potential combinations:** A1, A2, B1, B2. In original sales list, A2 and B1 are listed; therefore, we will **keep only A2 and B1**

**Commented [JK3]:** After looking at the original file, it does appear that there are records for these missing fields. Please use the original file, 'Final Survey Themes Masked without Original Keys.csv', to match to the correct IPNs and Contract numbers.

### 3. Survey Key:

This file has unique list of all the Responses, ANON.IPN and ANON.Contracts needed for our analysis. It has 6855 unique responses corresponding to 6870 policy holders. **Please confirm**

It has 3 missing values: R\_1QyKRASixWwMk5o, R\_2xEJy6om1ocX9XB and R\_

These responses have IPN and contract values in old survey data provided. So, do you want us to map them from old survey data? **Please confirm**

### 4. Sales data:

**Anon Sales and Policy Data.csv** – Initial sales data which has 217,224 observations and 41012 missing values. So, excluded missing values from this dataset. It now has 176212 observations.

**Anon Sales and Policy Data - Missing Contract numbers.csv** – Additional sales data provided. It 153505 observations of which 7 are missing values. Below is the list:

PN0000013623, PN0001650691, PN0002196511, PN0002815041, PN0002996681, PN0005251100, PN0010370871

Some of these 7 lines have survey and email listing data but sales data is **missing**. **Please provide.**

### Questions:

1. please check contract num : PN0000064085

We see that there are 8 rows where 4 rows are of type owner and 4 or type implicit payor. From the data glossary. We observed that if a person is an owner and implicit payor, couldn't we merge the rows and have it as type owner payor. Also, if there are 2 types of annualized premiums. If a person is paying twice can we merge them as well?

For this same Id we also observe that there is a change in agency.

regarding the field "agent.change" also let us know what 1 and 0 indicate...1 indicates change or does it indicate no change of agency?

And the agent id is the same, but the agency location has been changed from IL to KY so why is the agency id field not present for IL state **rows**.

Of all the rows pertaining to this, how do we identify the unique list?

In order to identify unique list, maybe we could drop some unnecessary fields? Is it fine if we drop – these fields? **Please confirm**

**Commented [JK4]:** Yes, please map to the old data for the key. However, keep in mind the key; the Anon Sales data includes all policies and roles, while the survey was only sent to those people who were owners.

**Commented [JK5]:** Yes, map from old survey data

**Commented [PR(6R5):**

**Commented [JK7]:** First of all, you should be using the Anon Sales and Policy Data 2019-02-21.csv file, which was updated before I left on 2/21. Does this question take this into account? (I added a new file to answer a question further on, so while you should have been using the file from 2019-02-21, it will now be the file from today)  
Secondly, I added all of the information that I have for these policies from our database. Please see the file: Missing Policies 3.6.19.csv I am aware of some NAs in this file but at this time it is as complete as can happen due to database limitations.

**Commented [JK8]:** In the future, please let me know which file you are referencing.

This has been fixed, please use the file from 03-06-2019

**Commented [JK9]:** 1 – Indicates "Change"  
0 – Indicates "No change"  
If the original agent and the servicing agent are different, there was a change in the agent assigned to a policy; therefore a 1 is assigned.

**Commented [JK10]:** No need to complete with new file.

2. Client country is not present in the Sales data, while Allocated Amt is not there in data glossary. Is it okay if we consider the country to be USA for all clients? **Please confirm**

**Commented [JK11]:** You need to use the file we sent on 2.21.2019, this file removes allocated amount and includes client country. This change was stated in the updated email sent on the same day.

3. We are thinking of performing analysis on Survey and email listing data, at aggregated IPN and Contract level without rolling down (because rolling down would cause duplication of survey answers in survey data and duplication of policy details in email listing). But the sales data is at unique level of IPN and Contract. So, we are planning to rollup these records to the aggregated level of response ID (as in survey, email data) and *remove amount values*.

(for ex: if we roll up 8 lines corresponding to one response ID, it might have same billing frequency, mode and certain other column values (which is redundant info), only possible variation might occur in client role type. But while rolling up, we will merge all unique values corresponding to that response ID, so, it would capture all possible role types we would need. At any stage if further info is needed we could check the unique level sales data). Let us know if this approach would work. **Please confirm**

**Commented [JK12]:** I would recommend doing a flag field for each field you are going to "drop" instead of dropping it. So for instance transpose the field Client Role so that you have CR.Owner, CR.Payor, etc. Create a field to roll up amounts by Product line