サブセット割合の理解と利用について

キー項目のサブセット割合を調べることは、データの質を保証するのに必要な作業です。
サブセット割合は、データモデルビューア内でキー項目を選択するとプレビューに表示されます。
以下は「受注明細」「顧客マスタ」のデータモデルですが、「顧客ID」項目をクリックすると、
「サブセット割合」「ユニーク値総数」「現在のユニーク値の数」が確認できます。

<各項目の説明>

サブセット割合	選択している項目のユニーク値の合計数に対して、 テーブル内で選択した項目のユニーク値の数の割合を表示します。
ユニーク値総数	モデル内の全てのテーブル「受注明細」「顧客マスタ」の中から、 選択している項目「顧客ID」の全てのユニーク値をカウントします。
現在のユニーク値の数	現在選択したテーブル「顧客マスタ」の中から、 選択している項目「顧客ID」内のユニーク値をカウントします。

全頁の「顧客マスタ」は、「顧客ID」項目の100%が「受注明細」に表示されています。
これは、100%のサブセット割合を「顧客マスタ」から見つけることができています。

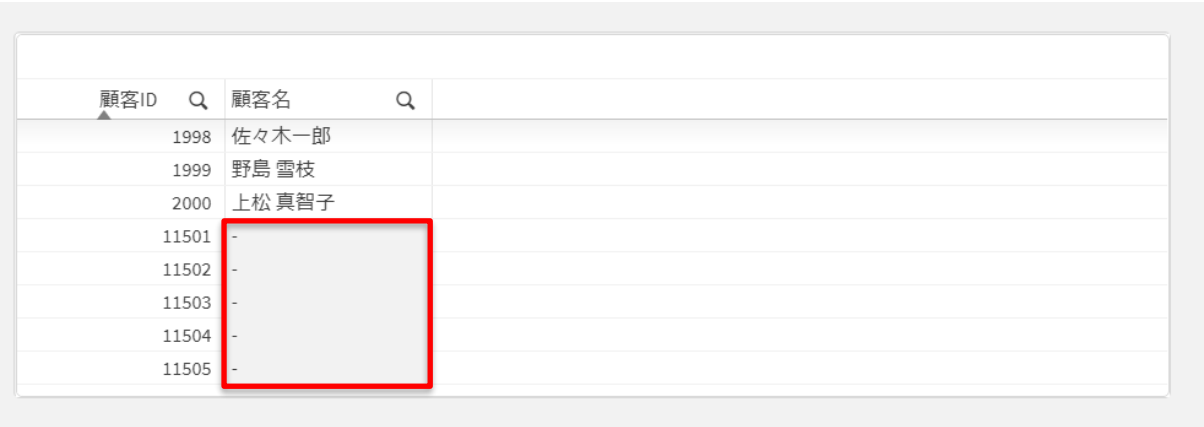
次に「受注明細」のサブセット割合も確認してみます。

このテーブルから、サブセット割合が100パーセントよりも少ない事がわかります。
2020(＝74.2%)のユニーク顧客のみが「受注明細」の中に表示されていますが、
100パーセントを下回るサブセット割合は「受注明細」のようなデータでは普通の状態です。
もし、注文のない顧客データを含めたくなければ、“Where Exists(顧客ID)”句をロードスクリプトに挿入すれば、「顧客マスタ」から不要な顧客を削除できます。

全頁までは100%となるサブセット割合をみてきましたが、「顧客マスタ」の値が100パーセントを下回った場合を確認してみます。

上記では、5つの顧客ID（2005→2000）を見落としているということことがわかります。
これは、「顧客マスタ」には存在するが、「受注明細」に存在していないデータロードしてしまった事が原因で件数に差異が発生しています。

見落とされた顧客IDはシート上で、2つの列を持つテーブルを作成すれば確認できます。
顧客名がnull値（-）で表示されたものが該当データであることがわかります。



以上のように、データモデル作成時にサブセット割合を調べることは重要なステップとなります。
また、データモデルの質を高めることで不要なデータ確認等の作業を削減できます。