

<演習問題・解答例>

1. ファイルの読み込み

・Excel ファイルの読み込み

明細データ.xlsx からのデータのインポート

3 / 4 フィールド属性の定義

含める	ソース名	名前	ラベル	種類	ソースの入力形式	長さ	出力形式	入力形式
	YMD	YMD	受注年月日	日付	DATE9.	8	YYMMDD\$10.	DATE9.
	CUST_ID	CUST_ID	顧客ID	文字列	\$CHAR6.	6	\$CHAR6.	\$CHAR6.
	ORDER_NO	ORDER_NO	受注番号	数値	BEST12.	8	BEST12.	BEST12.
	PRODUCT	PRODUCT	商品コード	文字列	\$CHAR3.	3	\$CHAR3.	\$CHAR3.
	QUANTITY	QUANTITY	数量	数値	BEST12.	8	BEST12.	BEST12.
	AMOUNT	AMOUNT	受注金額	数値	BEST12.	8	BEST12.	BEST12.

データ (明細データ.xlsx[DETAIL]) のインポート

受注年月日	顧客ID	受注番号	商品コード	数量	受注金額
2018/01/01	071111	1	050	3	3000
2018/01/01	048302	2	112	1	1200
2018/01/01	011501	3	870	1	1800
2018/01/01	113777	4	45A	1	4000
2018/01/01	061155	5	45B	1	1500
2018/01/01	035927	6	45B	1	1500
2018/01/01	046582	7	820	3	5400

ラベル・種類・長さ・出力形式を、指定されたものに修正します。

・CSV ファイルの読み込み

顧客データ.csv からのデータのインポート

2 / 4 データソースの選択

テキストの形式

区切り文字で区切られたフィールド

カンマ

ファイルではレコード番号の上にフィールド名が挿入されている(は)

データレコードの開始レコード番号(は)

1

2

カンマ区切りを選択し、先頭行がフィールド名・データ開始行が 2 行目となるようにします。

顧客データ.csv からのデータのインポート

3 / 4 フィールド属性の定義

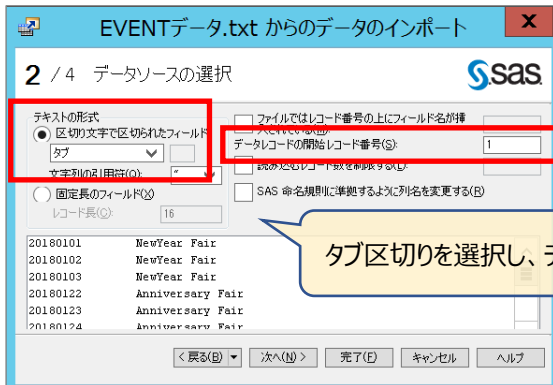
含める	ソース名	名前	ラベル	種類	ソースの入力形式	長さ	出力形式	入力形式
	CUST_ID	CUST_ID	顧客ID	文字列	\$CHAR6.	6	\$CHAR6.	\$CHAR6.
	GENDER	GENDER	性別	文字列	\$CHAR4.	4	\$CHAR4.	\$CHAR4.
	AGE	AGE	年齢	数値	BEST2.	8	BEST2.	BEST2.
	NEW_DATE	NEW_DATE	初回受注日	日付	YYMMDD\$10.	8	YYMMDD\$10.	YYMMDD\$10.

データ (顧客データ.csv) のインポート

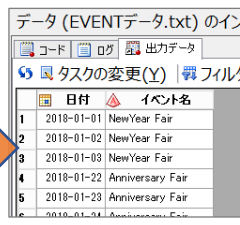
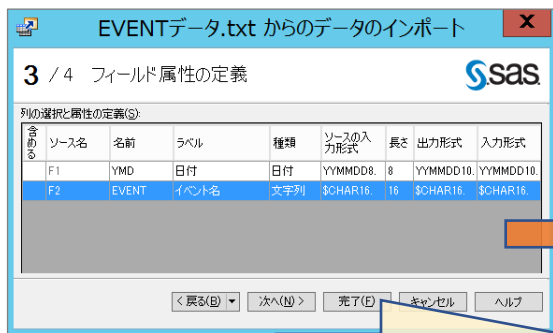
顧客ID	性別	年齢	初回受注日
000415	男性	27	2016/03/12
000627	女性	47	2017/10/15
001012	女性	30	2016/09/05
001081	女性	43	2015/12/11
001189	男性	58	2016/05/18
001204	女性	30	2016/09/18

ラベル・種類・長さ・出力形式を、指定されたものに修正します。

・TXT ファイルの読み込み



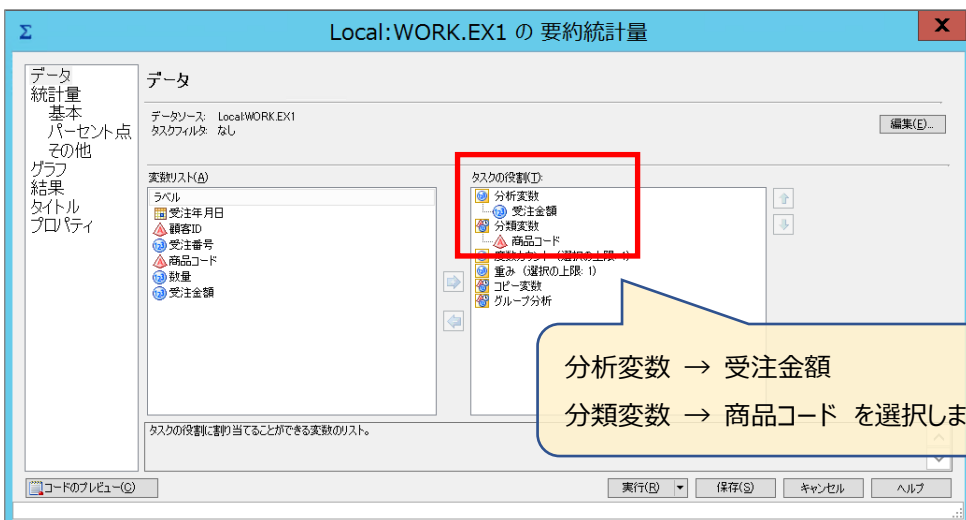
タブ区切りを選択し、データ開始行が 1 行目となるようにします。

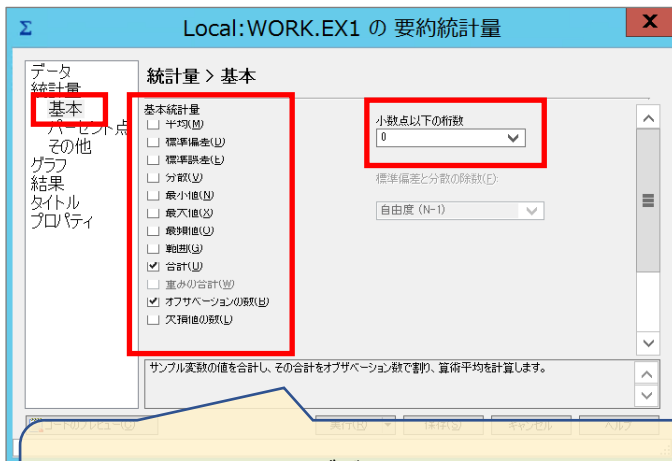


ラベル・種類・長さ・出力形式を、指定されたものに修正します。

2. 要約統計量の算出

・A : 商品コードごとの受注金額の合計





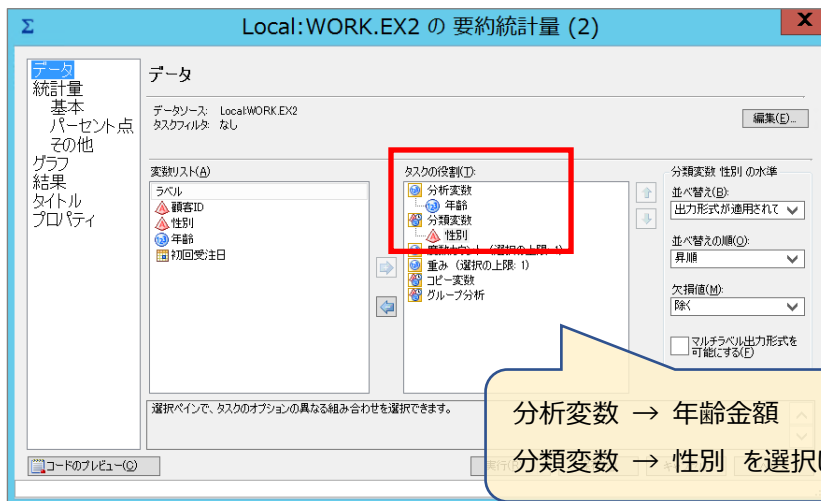
「基本」ペインから合計・オフバージョンの数を選択します
また、小数点以下の桁数調整も可能です。

(出力結果)

**要約統計量
結果**
MEANS プロシジャ

分析変数: AMOUNT 受注金額			
商品コード	Obs 数	合計	N
050	343	533000	343
080	185	59700	185
111	157	544800	157
112	103	140400	103
123	136	980000	136
20A	146	173600	146
33R	19	380000	19
45A	122	744000	122
45B	80	150000	80
870	167	437400	167
900	92	367500	92

・例) B : 性別ごとの年齢の要約統計量



分析変数 → 年齢金額
分類変数 → 性別 を選択します

(出力結果)

**要約統計量
結果**
MEANS プロシジャ

分析変数: AGE 年齢						
性別	Obs 数	平均	標準偏差	最小値	最大値	N
女性	538	35.9479554	11.2602978	20.0000000	67.0000000	538
男性	241	38.1203320	11.5609815	20.0000000	69.0000000	241

3. カテゴリ変数の頻度集計

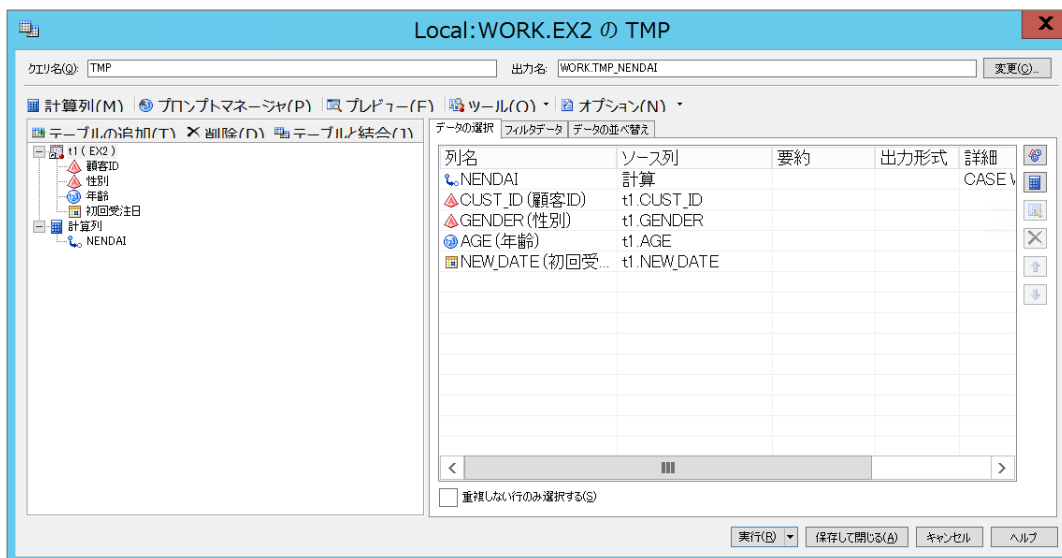
・クエリビルダの作成（年齢から年代を作成）

クエリビルダの「計算列」で「再コード列」を選択し、列の選択で「年齢」を選択します。

「追加」をクリックし、置換条件を指定します。
「範囲の置換」タブで下限・上限・置き換え文字を指定します。
文字列になるため、「この値を引用符で囲む」にチェックを付けます。

同様に 30 代～60 代まで範囲置換を指定します。
また、指定した範囲以外の値であった場合の指定も合わせて行います。

変数名・ラベル・出力形式を入力します。



(作成結果)

年代	顧客ID	性別	年齢	初回受注日
1 20代	000415	男性	27	2017/09/14
2 40代	000627	女性	47	2017/07/22
3 30代	001012	女性	30	2017/08/15
4 40代	001081	女性	43	2017/03/10
5 50代	001189	男性	58	2018/01/22

・一元度数表の作成

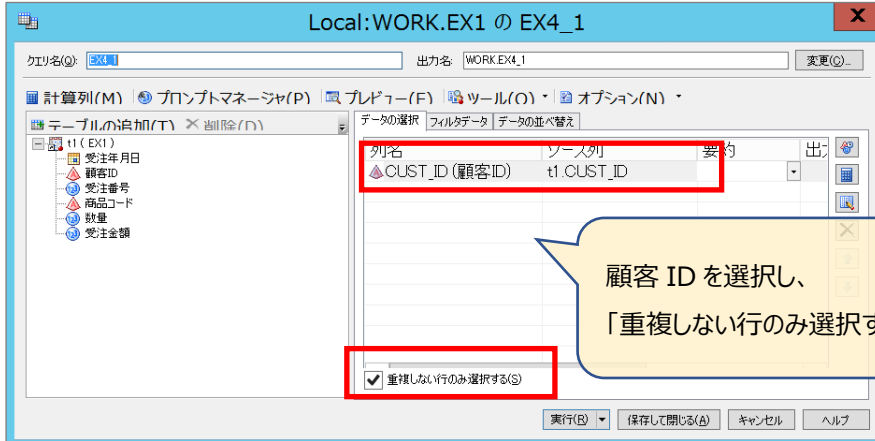


(出力結果)

一元度数表 結果 FREQ プロシジャ				
年代				
NENDAI	度数	パーセント	累積 度数	累積 パーセント
20代	242	31.07	242	31.07
30代	277	35.56	519	66.62
40代	142	18.23	661	84.85
50代	82	10.53	743	95.38
60代	36	4.62	779	100.00

4. データの抽出

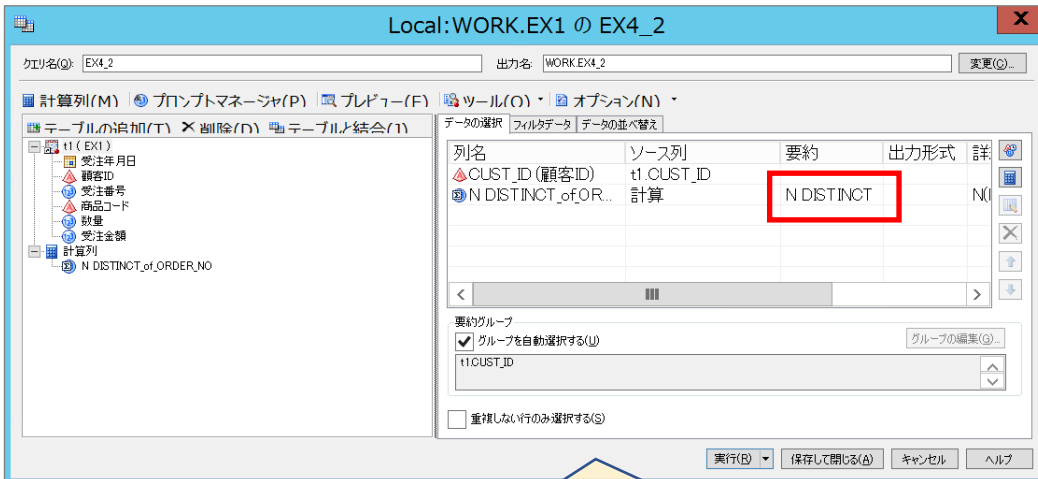
・A : ユニーク顧客数

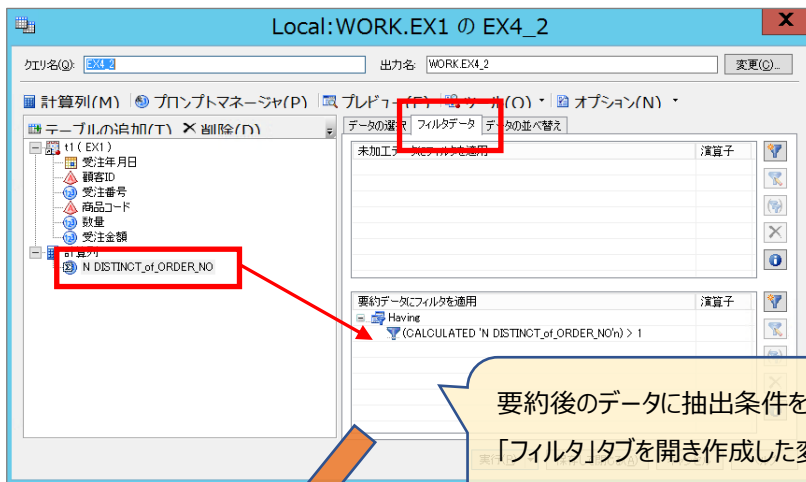


※ 抽出結果は 779 となります。

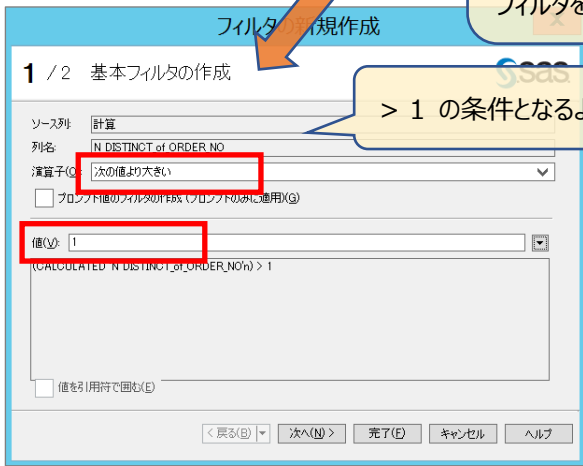
顧客ID
775
776
777
778
779

・B : 受注回数 2 回以上の顧客数





要約後のデータに抽出条件を付けるため、「フィルタ」タブを開き作成した変数を「要約データにフィルタを適用」ヘドラッグします。

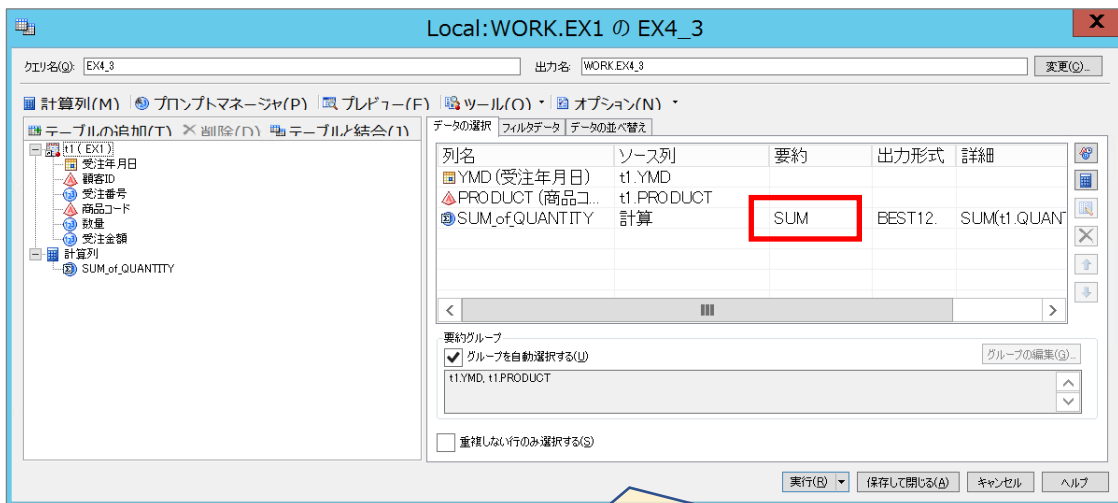


> 1 の条件となるように指定します。

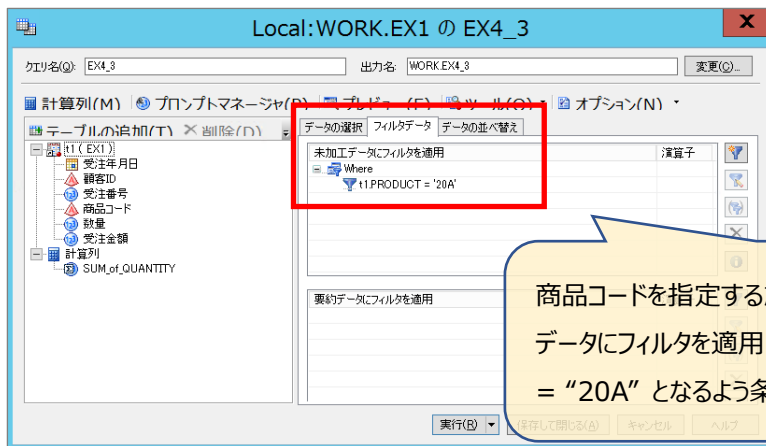
※ 抽出結果は 405 となります。

顧客ID	DISTINCT_of_ORDER_NO
401	2
402	2
403	4
404	5
405	2

・C : 「20A」の日別受注数量の合計



受注年月日・商品コードごとの数量を合計します。数量の要約は「SUM」を選択します。



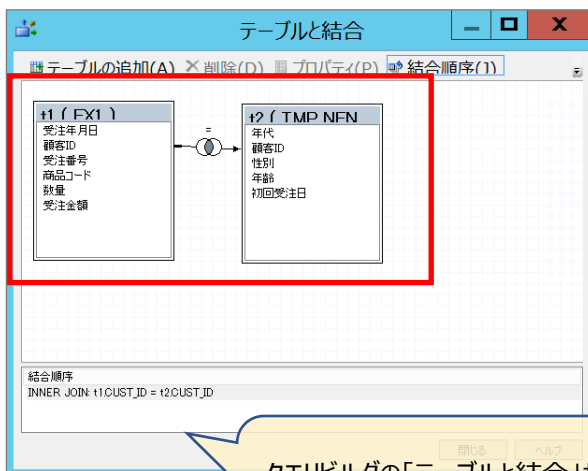
商品コードを指定するため、商品コードを「未加工データにフィルタを適用」ヘドドラッグします。
= "20A" となるよう条件指定します。

※ 結果は以下ようになります（62 オブザベーション）。

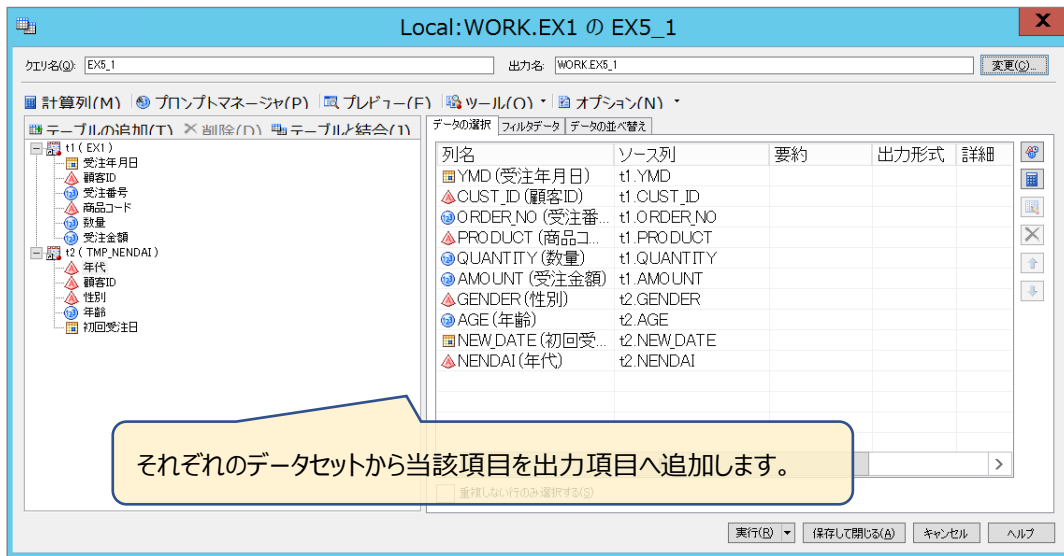
受注年月日	商品コード	SUM_of_QUANTITY
2018/03/24	20A	7
2018/03/28	20A	4
2018/03/29	20A	3
2018/03/30	20A	4
2018/03/31	20A	1

5. データセットの横結合

・A：結合（明細と顧客データの結合）



クエリビルダの「テーブルと結合」から 2 つのデータセットを選択します。
それぞれの顧客 ID を紐付け、内部結合（inner join）を選択します。

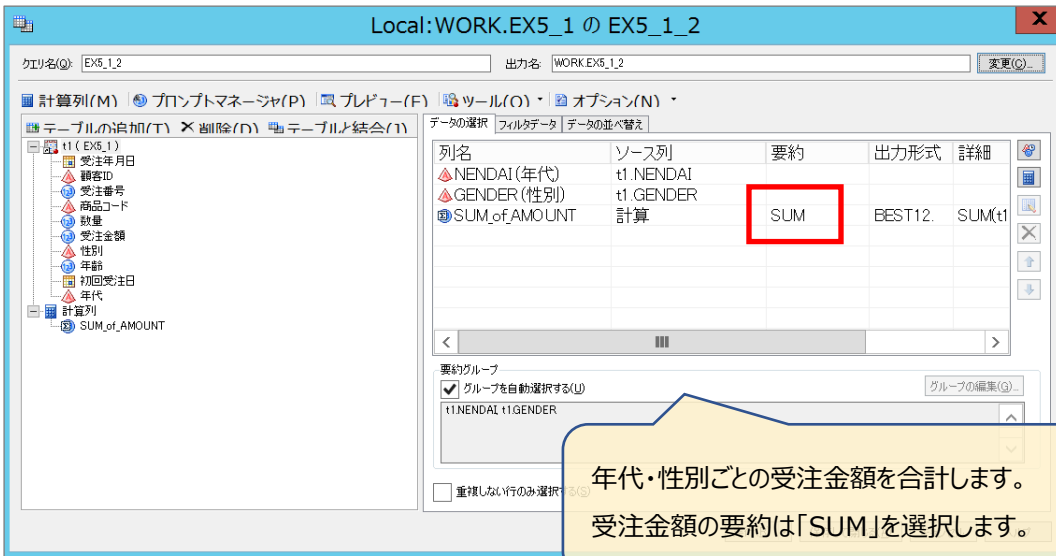


※ クエリの結果は 1550 オブザベーションとなります。

EX5_1

	受注年月日	顧客ID	受注番号	商品コード	数量	受注金額	性別	年齢	初回受注日	年代
1	2018/01/01	071111	1	050	3	3000	女性	45	2017/07/09	40代
2	2018/01/01	048302	2	112	1	1200	男性	43	2017/08/29	40代
3	2018/01/01	011501	3	870	1	1800	男性	56	2016/10/14	50代
4	2018/01/01	118777	4	45A	1	4000	女性	31	2017/05/11	30代
5	2018/01/01	061155	5	45B	1	1500	女性	33	2016/10/13	30代

・A : 集計 (年代・性別ごとの集計)

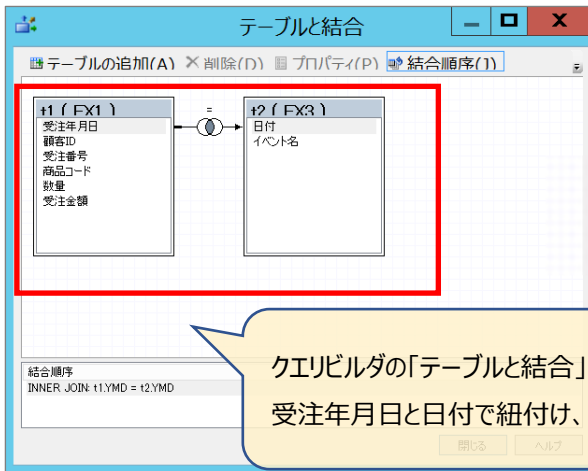


※ クエリの結果は右記となります。

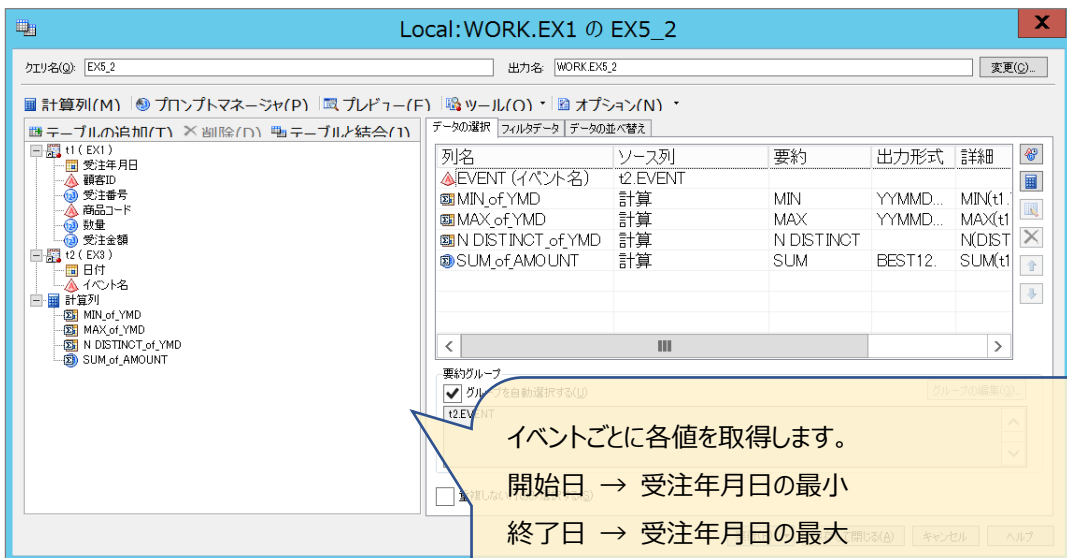
EX5_1_2

	年代	性別	SUM_of_AMOUNT
1	20代	女性	1118000
2	20代	男性	363400
3	30代	女性	1033900
4	30代	男性	481600
5	40代	女性	571300
6	40代	男性	293800
7	50代	女性	283700
8	50代	男性	185100
9	60代	女性	124800
10	60代	男性	54800

・B : 結合・集計 (明細とイベントデータの結合、集計)



クエリビルダの「テーブルと結合」から2つのデータセットを選択します。
受注年月日と日付で紐付け、内部結合 (inner join) を選択します。



イベントごとに各値を取得します。
開始日 → 受注年月日の最小
終了日 → 受注年月日の最大
実施日数 → 受注年月日 (重複なし) のカウント
受注金額 → 受注金額の合計

※ クエリの結果は以下となります。

イベント名	開始日	終了日	実施日数	受注金額
1 NewYear Fair	2018/01/01	2018/01/03	3	231900
2 Anniversary Fair	2018/01/22	2018/01/26	5	368300
3 Valentine Fair	2018/02/13	2018/02/14	2	132300
4 Renewal Fair	2018/02/27	2018/03/02	4	349800
5 Spring Fair	2018/03/24	2018/03/26	3	254900