

Memo

To: Client Stakeholders

From: Brittany Hancock

Subject: Week 1 Data Audit Findings

Purpose: In preparation for client use, 3 flat files have been reviewed with the Data Audit Framework.

Source:

- Sales from the Point-of-Sales (POS) System
- Products from the catalog
- Customers from the Customer Relationship Management (CRM) System

Format:

- Sales: dates are inconsistent, invalid month and day input
- Products: invalid unit price (non-numeric "N/A" found)
- Customers: multiple header rows

Quality:

- Sales: inconsistent positive quantity (negative observed), inconsistent entry (non-numeric "1 fifty" observed), total amount has null entry, customer id inconsistent (99999 found when others 1000s), product id inconsistent (99999 found when others 2000s), no foreign key references observed even though present
- Products: duplicate rows, a supplier is null
- Customers: customer ID used more than once, null first name observed, gender inconsistency (most are M or F, while one "unknown" observed), invalid age entry of negative value

Risks:

- Inaccurate data
- Inconsistent formatting interferes with algorithms
- Inflation with duplicate entries
- Unaccounted for reused primary keys
- Incomplete data skewing snapshots

Recommendations:

- Consistent header names
- Validation of specified datatypes
- Force unique keys to remove duplication risk
- Validation of foreign keys in use

Reflection:

- This project showed me how messy data breaks imports and analysis. Even simple issues, such as duplicate IDs or invalid dates, caused SQL to fail. I learned why documenting risks is as important as cleaning data.

Figure 1: CustomersTable_ImportError_PKDuplicate

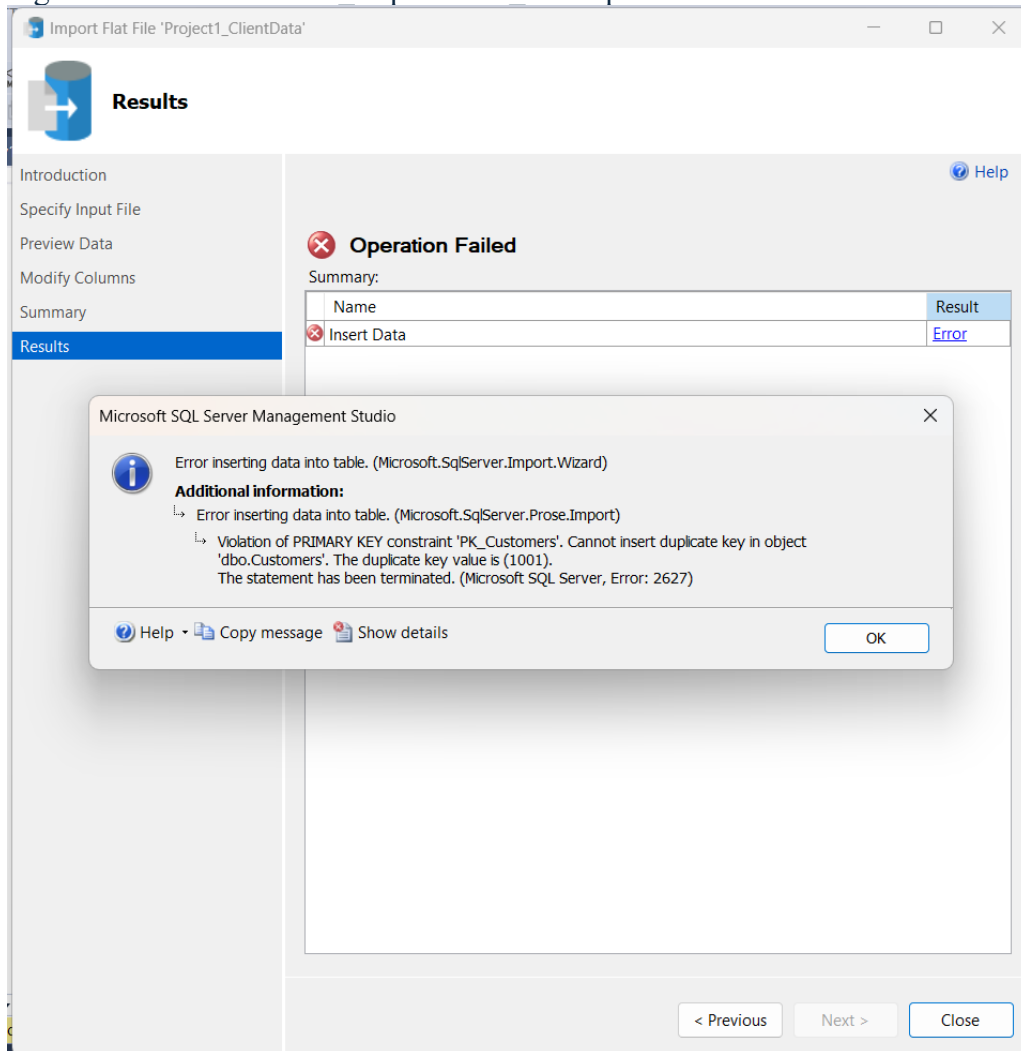


Figure 2: CustomersTable ImportSuccess AfterCorrectingPKDuplicate

SQLQuery1.sql - (I...ANYSHP\femal (54))*								
USE Project1_ClientData;								
SELECT TOP 10 * FROM Customers;								
81 %								
Results Messages								
	ID	FName	LName	Sex	Years	Town	Region	Group
1	1000	Michael	Lopez	F	33	Tacoma	WA	Retail
2	1001	Emma	Martin	M	50	Portland	OR	Online
3	1002	Michael	Thompson	M	55	Bellevue	WA	Retail
4	1003	Sarah	Chen	M	32	Olympia	WA	Online
5	1004	Sarah	Thompson	M	63	Olympia	WA	Wholesale
6	1005	David	White	F	69	Seattle	WA	Retail
7	1006	NULL	Taylor	F	27	Tacoma	WA	Wholesale
8	1007	Michael	Chen	F	24	Eugene	OR	Wholesale
9	1008	Harper	Wilson	M	64	Vancouver	WA	Online
10	1009	Michael	Thomas	M	53	Salem	OR	Online

Figure 3: ProductsTable ImportSuccess

SQLQuery1.sql - (I...ANYSHF\femal (54))*

```
USE Project1_ClientData;  
SELECT TOP 10 * FROM Products;
```

81 %

Results Messages

	ProductID	ProductName	Category	UnitPrice	Cost	Supplier
1	2001	Coffee Beans 1lb	Beverages	12.9899997711182	7.5	Pacific Roasters
2	2002	Tea Sampler Pack	Beverages	18.9899997711182	10	Leaf & Co
3	2003	Espresso Machine	Equipment	NULL	180	BrewTech
4	2004	French Press	Equipment	39.9900016784668	20	CafeGear
5	2005	Mug Set	Merchandise	24.9899997711182	10	NULL
6	2006	Cold Brew Kit	Equipment	59.9900016784668	35	CoolBrew
7	2007	Green Tea 100g	Beverages	9.98999977111816	4	Leaf & Co
8	2008	Herbal Tea Box	Beverages	14.9899997711182	7	Leaf & Co
9	2009	Coffee Grinder	Equipment	79.9899978637695	50	BrewTech
10	2010	Tumbler	Merchandise	19.9899997711182	8	KitchenWorks

Figure 4: ProductsTable ImportWarning UnitPriceNull

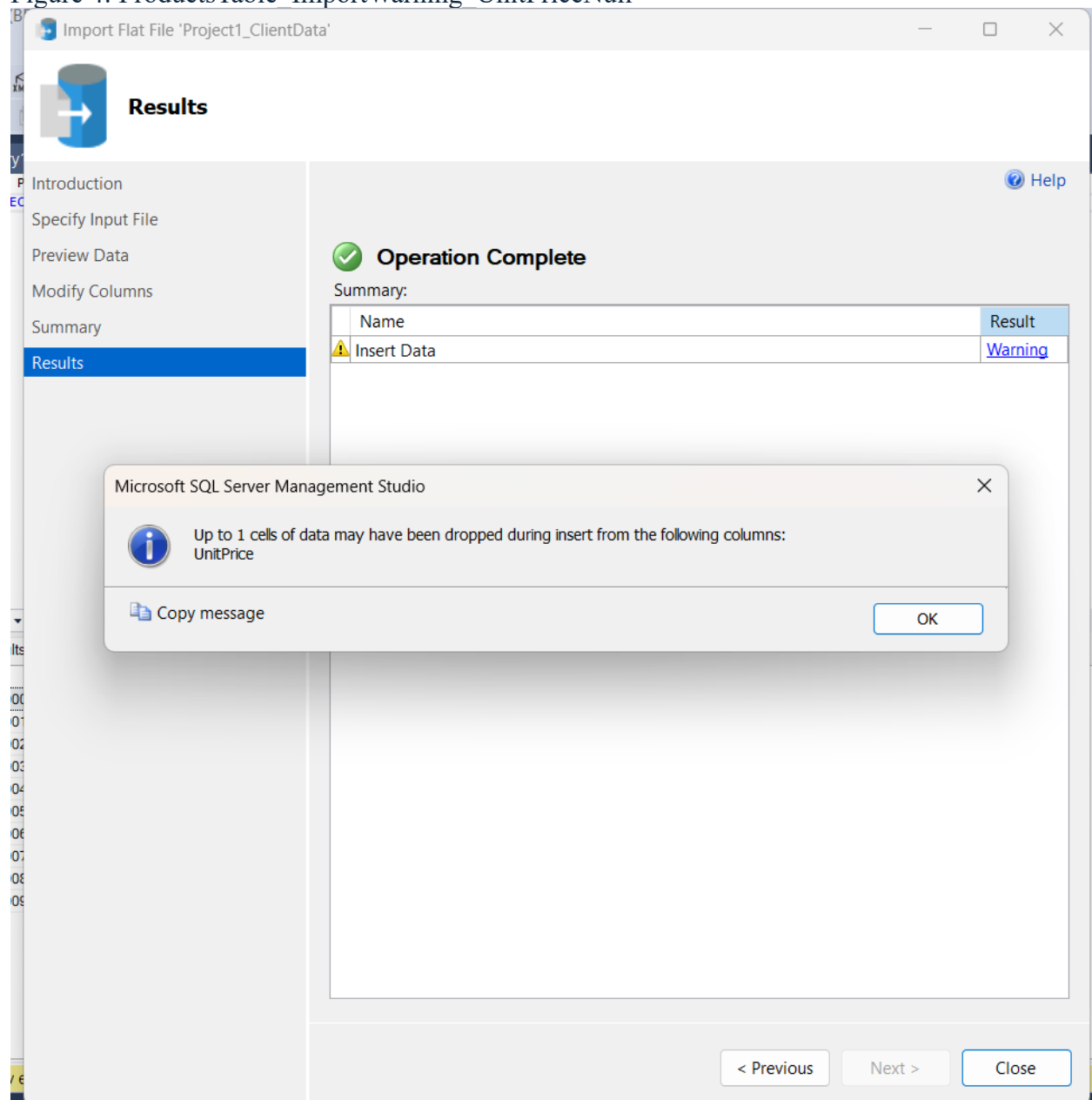


Figure 5: SalesTable_ImportError_CustomerIDExtremeValue

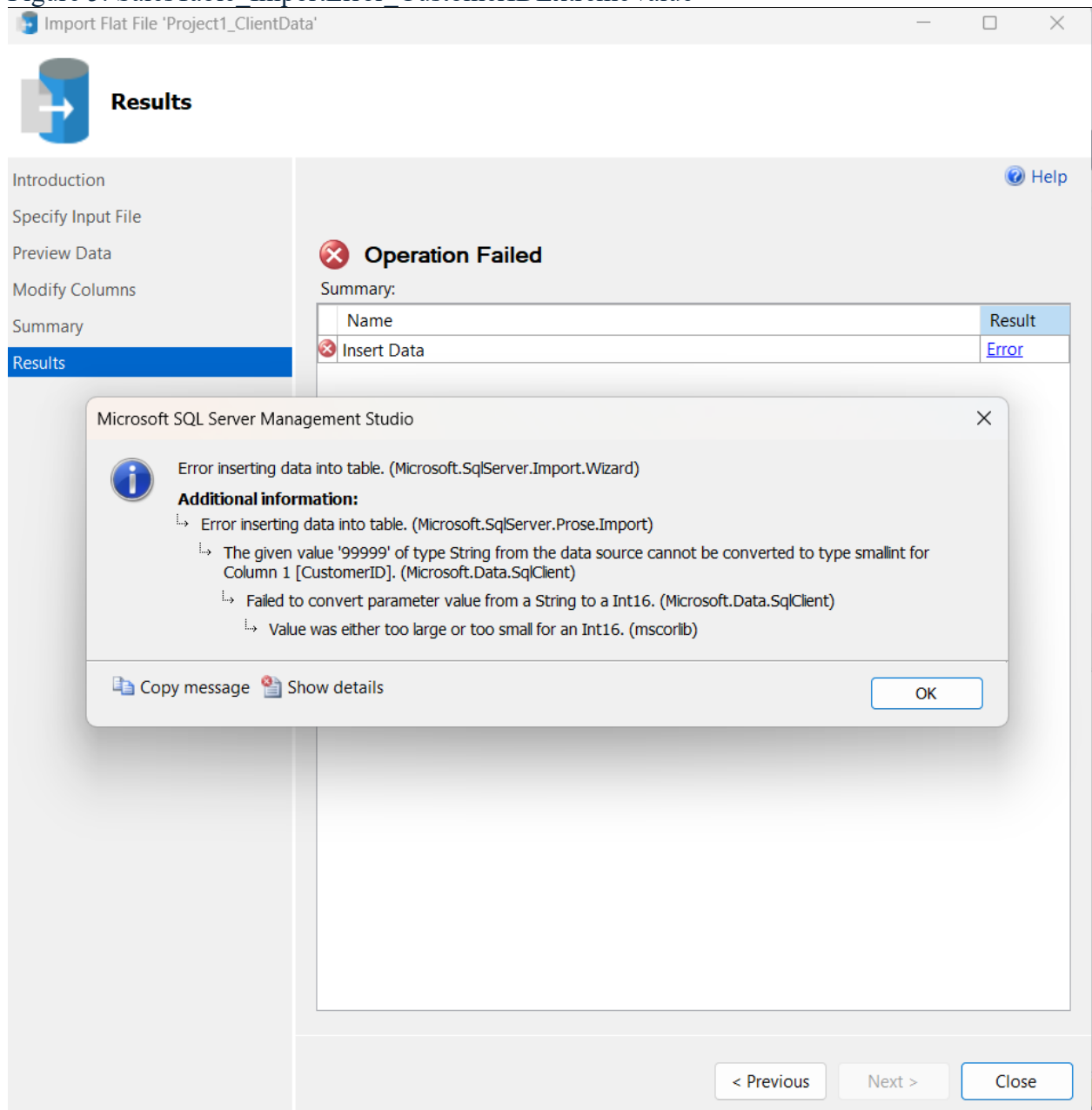


Figure 6: SalesTable_ImportError_DateFormatIncorrect

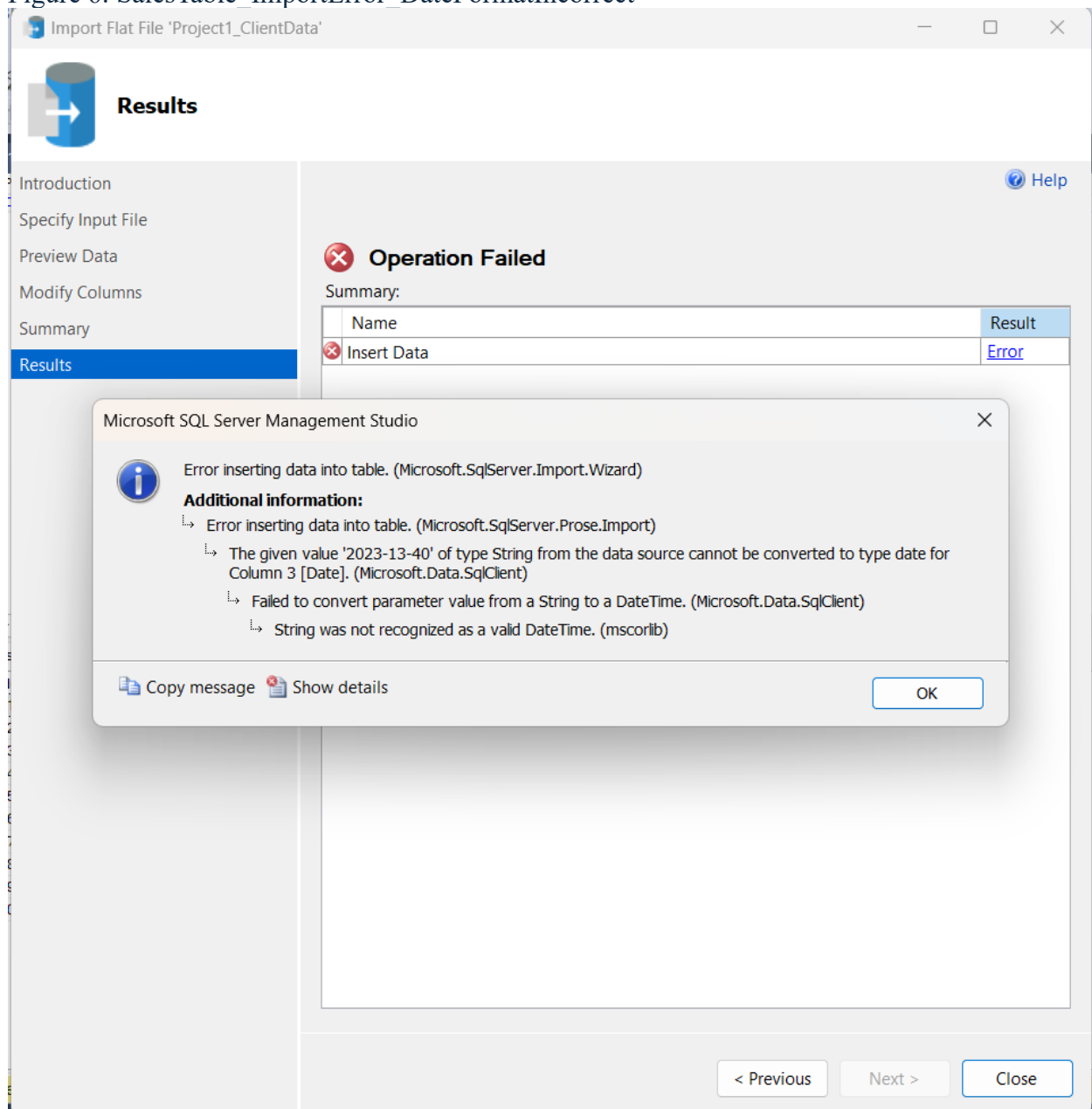


Figure 7: SalesTable_ImportError_ProductIDExtremeValue

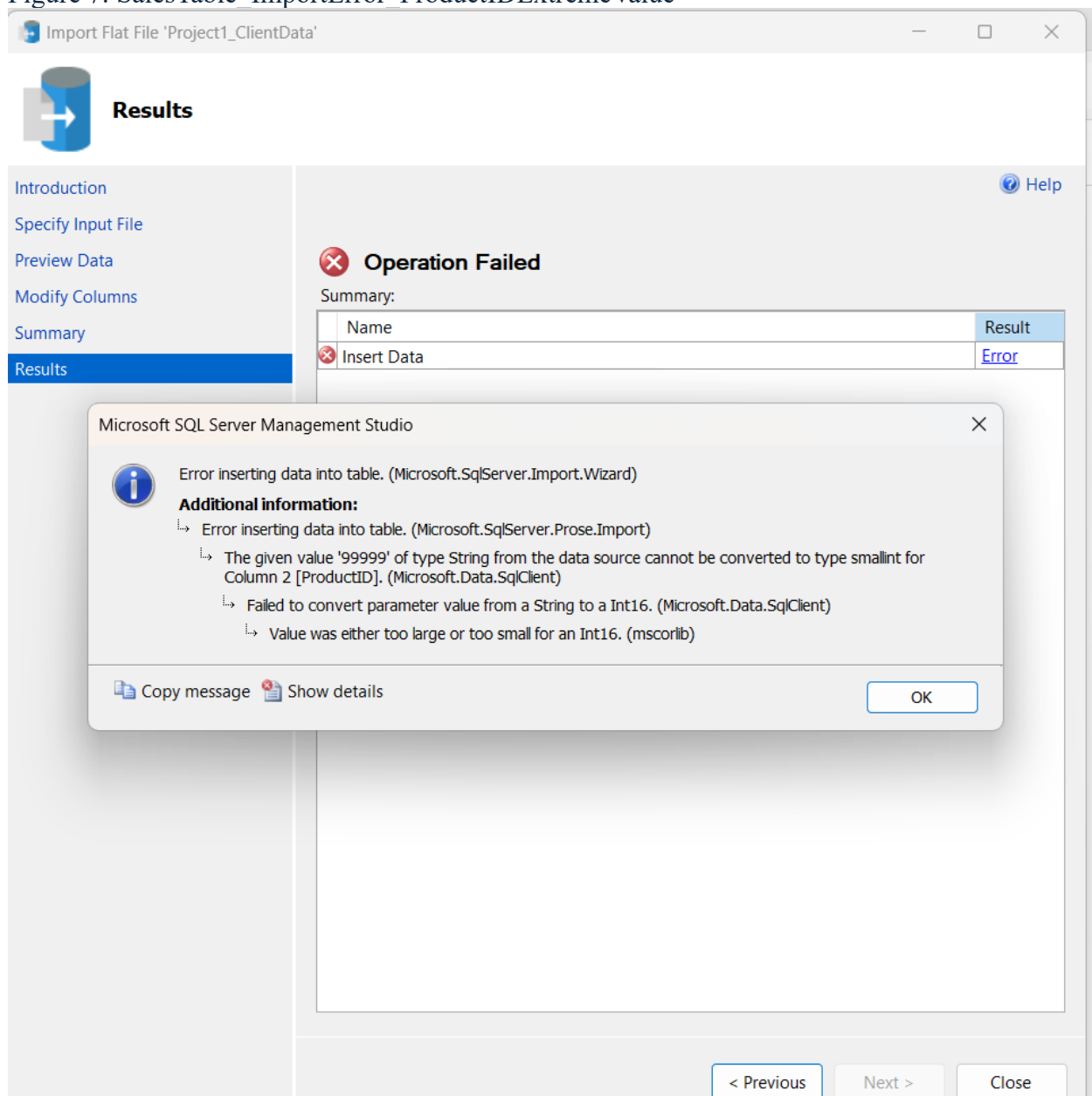


Figure 8: SalesTable ImportSuccess AfterCorrectingErrors

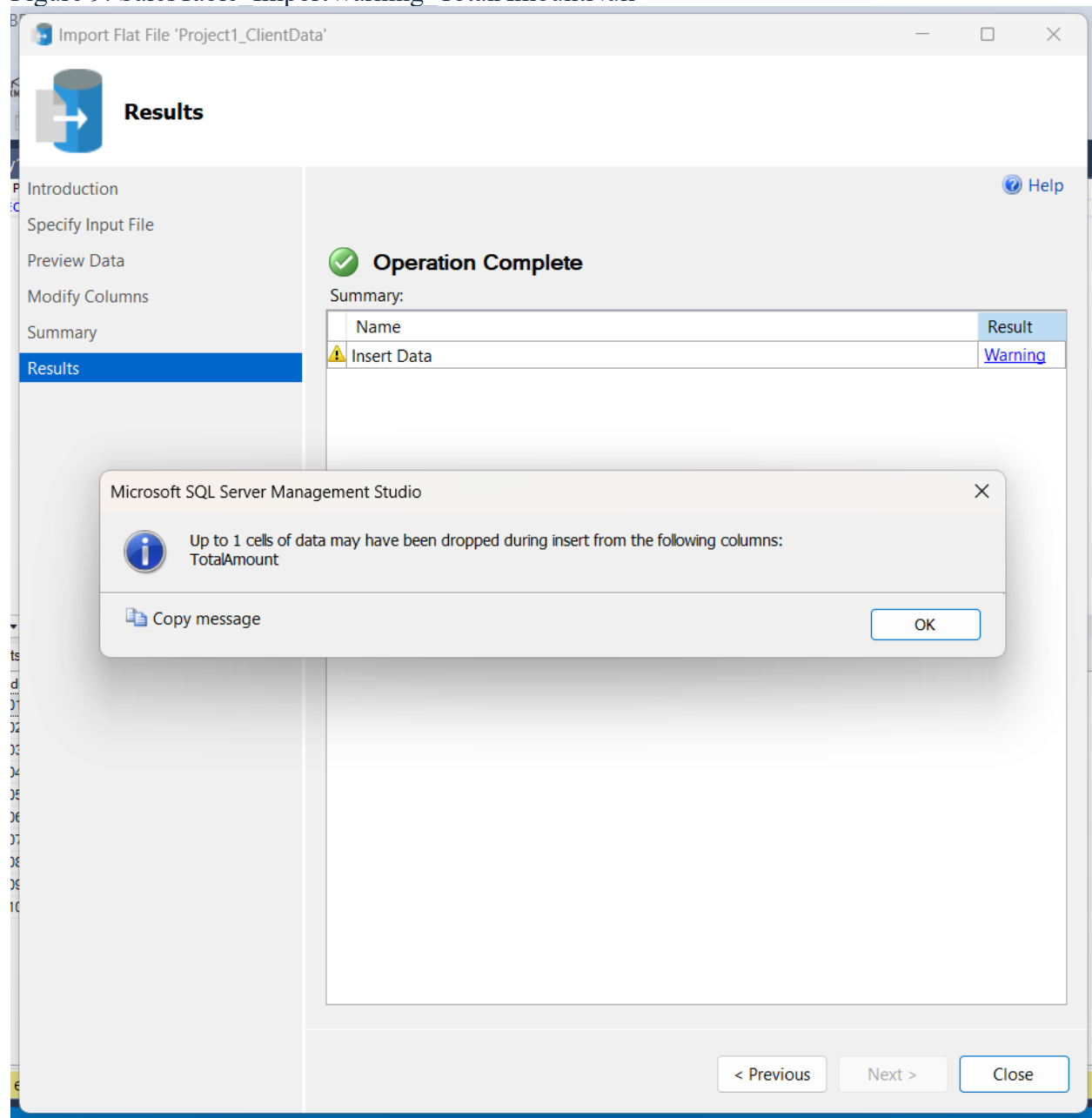
SQLQuery1.sql - (\\...ANYSHF\\femal (54))*
USE Project1_ClientData;
SELECT TOP 10 * FROM Sales;

81 %

Results Messages

	SaleID	CustomerID	ProductID	Date	Quantity	TotalAmount
1	3006	1001	2006	2023-12-30	4	239.960006713867
2	3029	1002	2007	2023-08-04	4	39.9599990844727
3	3036	1004	2008	2023-06-06	4	59.9599990844727
4	3068	1010	2004	2023-05-26	5	199.949996948242
5	3071	1010	2001	2023-03-26	3	38.9700012207031
6	3042	1012	2005	2023-04-13	2	49.9799995422363
7	3084	1012	2006	2023-05-11	1	59.9900016784668
8	3050	1013	2004	2023-05-08	3	119.970001220703
9	3026	1014	2005	2023-11-20	1	NULL
10	3054	1014	2001	2023-10-10	4	51.9599990844727

Figure 9: SalesTable ImportWarning TotalAmountNull



Reflection

This project has taught me how to import flat files into SSMS, handling error messages to make necessary corrections for success. It shows me that even when successful, there may still be warnings to address to build a more consistent and usable project. It showed me that sometimes success can be achieved, but certain operations may not be feasible in the future if notes are not taken early on.