



ANALYTICS FOR ADVENTUREWORKS

EAE Data Management 2020

Adrian Hagen
Jon Dale
Mohamed Ashmawy
Mostafa Abdelhakim
Joseph Higaki



OVERVIEW

ADVENTUREWORKS CYCLES NEEDS
INSIGHTS TO **IMPROVE** COMPANY'S
OVERALL **BUSINESS PERFORMANCE**





SCOPE

ANALYTICS WILL FOCUS FIRST IN
SALES AND **PURCHASING**
TO DRIVE IMPROVEMENTS IN
KEY FINANCIAL PERFORMANCE INDICATORS





BUSINESS CHALLENGES / INSIGHTS

SEASONAL SALES

Create top 5 products sales ranking by seasons.

HOLIDAY SALES

Reseller sales vs direct consumer sales over specific holiday days.

PRODUCT PROFITABILITY

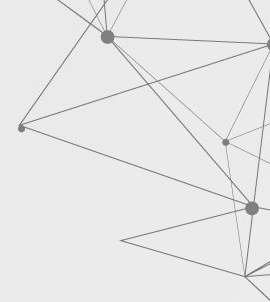
Create top 5 products profit ranking, quarter by quarter

REJECTED ITEMS

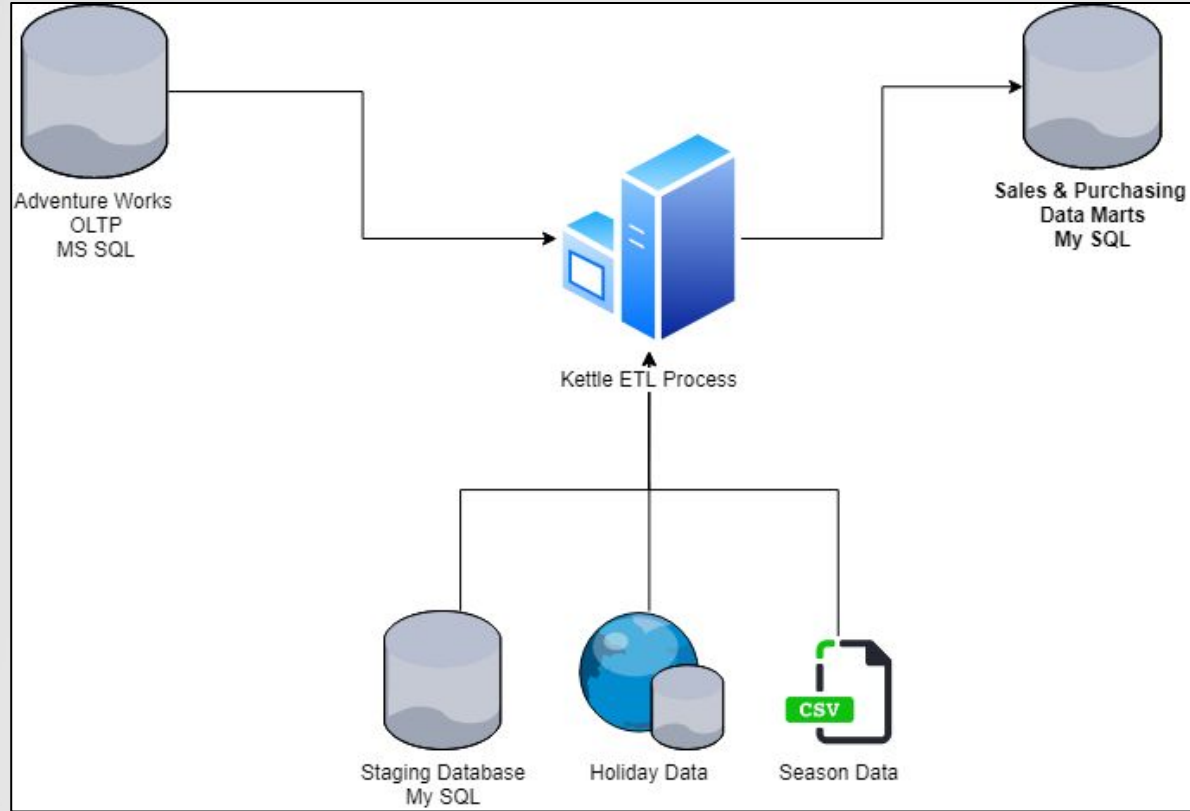
Create a top 5 Vendor ranking of rejected items on purchase orders.

PURCHASED VS SOLD

Sold items vs purchased items, for each product category



ARCHITECTURE



SALES DATA MART

dim_sales_person	
🔑 dim_sales_person_id: INTEGER	
🔑 sales_person_entity_id: INTEGER	
🔑 person_title: VARCHAR(8)	
🔑 person_first_name: VARCHAR(50)	
🔑 person_middle_name: VARCHAR(50)	
🔑 person_last_name: VARCHAR(50)	
🔑 person_suffix: VARCHAR(10)	
🔑 employee_job_title: VARCHAR(50)	
🔑 employee_gender_code: VARCHAR(1)	
🔑 employee_hire_date: DATE	
🔑 employee_marital_status: VARCHAR(1)	
🔑 employee_national_id_number: VARCHAR(15)	

dim_product	
🔑 dim_product_id: INTEGER	
🔑 product_id: INTEGER	
🔑 product_number: VARCHAR(50)	
🔑 product_name: VARCHAR(100)	
🔑 make_flag: BOOLEAN	
🔑 finished_goods_flag: BOOLEAN	
🔑 color: VARCHAR(15)	
🔑 size: VARCHAR(5)	
🔑 size_unit_measure_code: VARCHAR(3)	
🔑 weight: DECIMAL(19,2)	
🔑 weight_unit_measure_code: VARCHAR(3)	
🔑 product_line_code: VARCHAR(2)	
🔑 product_line_description: VARCHAR(100)	
🔑 class_code: VARCHAR(2)	
🔑 style: VARCHAR(2)	
🔑 product_subcategory_id: INTEGER	
🔑 product_subcategory_name: VARCHAR(100)	
🔑 product_category_id: INTEGER	
🔑 product_category_name: VARCHAR(100)	
🔑 product_model_id: INTEGER	
🔑 product_model_name: VARCHAR(100)	
🔑 version_number: INTEGER	
🔑 start_date: DATETIME	
🔑 end_date: DATETIME	

dim_sales_territory	
🔑 dim_sales_territory_id: INTEGER	
🔑 territory_id: INTEGER	
🔑 sales_territory_name: VARCHAR(100)	
🔑 country_region_code: VARCHAR(3)	
🔑 group_name: VARCHAR(100)	
🔑 old_sales_territory_name: VARCHAR(100)	

fact_sales	
🔑 fact_sales_id: INTEGER	
🔑 order_date_id: INTEGER	
🔑 order_date_datetime: DATETIME	
🔑 dim_customer_id: INTEGER	
🔑 dim_sales_territory_id: INTEGER	
🔑 dim_sales_person_id: INTEGER	
🔑 dim_product_id: INTEGER	
🔑 order_qty: INTEGER	
🔑 product_unit_price: DECIMAL(19,2)	
🔑 product_unit_price_discount: DECIMAL(19,2)	
🔑 total_amount: DECIMAL(19,2)	
🔑 total_discount: DECIMAL(19,2)	
🔑 sales_order_detail_id: INTEGER	
🔑 sales_order_id: INTEGER	
🔑 sales_order_number: VARCHAR(25)	
🔑 product_standard_cost_on_day: DECIMAL(19,2)	
🔑 ship_date_id: INTEGER	
🔑 ship_date_datetime: DATETIME	
🔑 sales_profit: DECIMAL(19,2)	

dim_date	
🔑 dim_date_id: INTEGER	
🔑 date: DATE	
🔑 is_holiday: BOOLEAN	
🔑 holiday_name: VARCHAR(100)	
🔑 day_of_the_week_number: INTEGER	
🔑 day_of_the_week_text: VARCHAR(10)	
🔑 day_of_the_month_number: INTEGER	
🔑 day_of_the_month_text: VARCHAR(10)	
🔑 month_number: INTEGER	
🔑 month_text: VARCHAR(10)	
🔑 day_of_the_year: INTEGER	
🔑 northern_hemisphere_season: VARCHAR(10)	
🔑 southern_hemisphere_season: VARCHAR(10)	
🔑 week_of_the_year: INTEGER	
🔑 quarter_number: INTEGER	
🔑 year_number: INTEGER	

dim_customer	
🔑 dim_customer_id: INTEGER	
🔑 customer_id: INTEGER	
🔑 is_reseller: BOOLEAN	
🔑 reseller_store_id: INTEGER	
🔑 reseller_store_name: VARCHAR(50)	
🔑 person_title: VARCHAR(8)	
🔑 person_first_name: VARCHAR(50)	
🔑 person_middle_name: VARCHAR(50)	
🔑 person_last_name: VARCHAR(50)	
🔑 person_suffix: VARCHAR(10)	
🔑 version_number: INTEGER	
🔑 start_date: DATETIME	
🔑 end_date: DATETIME	

PURCHASING DATA MART

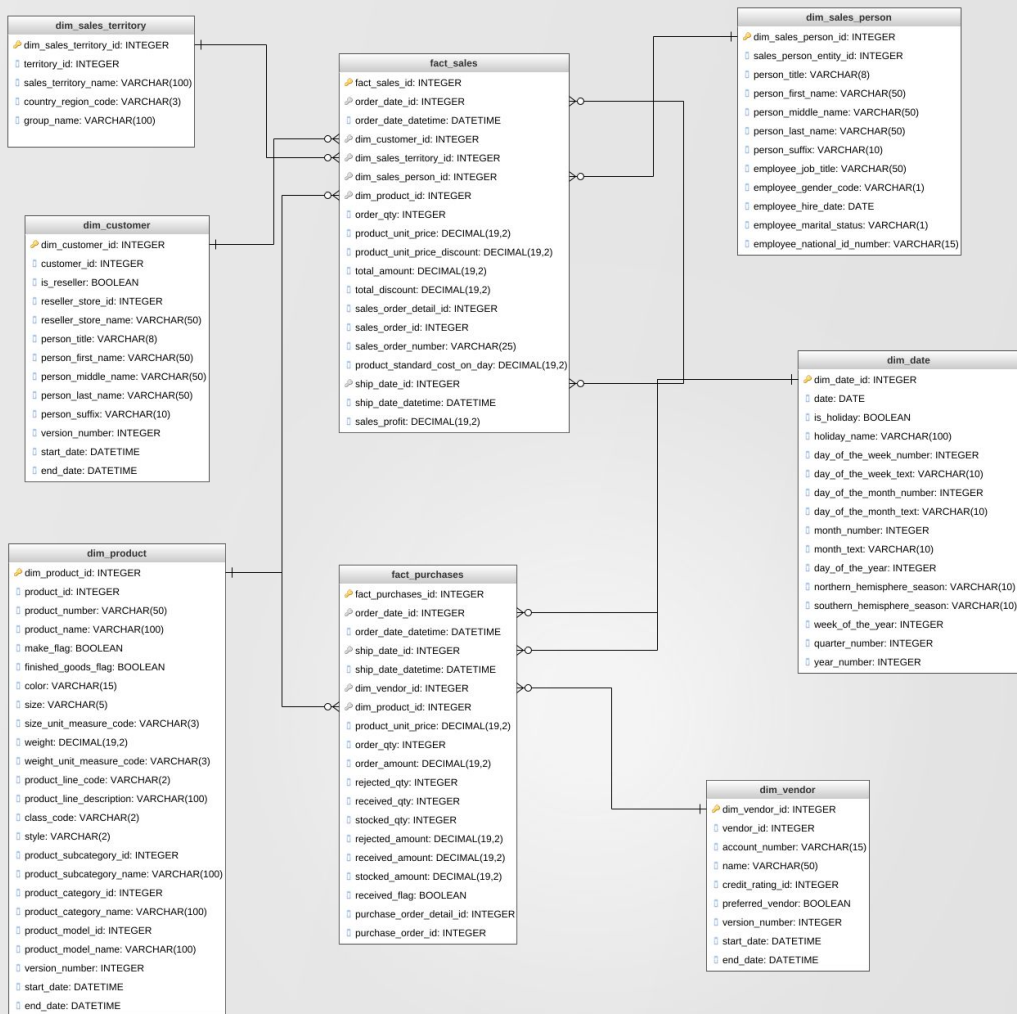
dim_product
🔑 dim_product_id: INTEGER
🔗 product_id: INTEGER
🔗 product_number: VARCHAR(50)
🔗 product_name: VARCHAR(100)
🔗 make_flag: BOOLEAN
🔗 finished_goods_flag: BOOLEAN
🔗 color: VARCHAR(15)
🔗 size: VARCHAR(5)
🔗 size_unit_measure_code: VARCHAR(3)
🔗 weight: DECIMAL(19,2)
🔗 weight_unit_measure_code: VARCHAR(3)
🔗 product_line_code: VARCHAR(2)
🔗 product_line_description: VARCHAR(100)
🔗 class_code: VARCHAR(2)
🔗 style: VARCHAR(2)
🔗 product_subcategory_id: INTEGER
🔗 product_subcategory_name: VARCHAR(100)
🔗 product_category_id: INTEGER
🔗 product_category_name: VARCHAR(100)
🔗 product_model_id: INTEGER
🔗 product_model_name: VARCHAR(100)
🔗 version_number: INTEGER
🔗 start_date: DATETIME
🔗 end_date: DATETIME

fact_purchases
🔑 fact_purchases_id: INTEGER
🔗 order_date_id: INTEGER
🔗 order_date_datetime: DATETIME
🔗 ship_date_id: INTEGER
🔗 ship_date_datetime: DATETIME
🔗 dim_vendor_id: INTEGER
🔗 dim_product_id: INTEGER
🔗 product_unit_price: DECIMAL(19,2)
🔗 order_qty: INTEGER
🔗 order_amount: DECIMAL(19,2)
🔗 rejected_qty: INTEGER
🔗 received_qty: INTEGER
🔗 stocked_qty: INTEGER
🔗 product_standard_cost_on_day: DECIMAL(19,2)
🔗 rejected_amount: DECIMAL(19,2)
🔗 received_amount: DECIMAL(19,2)
🔗 stocked_amount: DECIMAL(19,2)
🔗 received_flag: BOOLEAN
🔗 purchase_order_detail_id: INTEGER
🔗 purchase_order_id: INTEGER

dim_date
🔑 dim_date_id: INTEGER
🔗 date: DATE
🔗 is_holiday: BOOLEAN
🔗 holiday_name: VARCHAR(100)
🔗 day_of_the_week_number: INTEGER
🔗 day_of_the_week_text: VARCHAR(10)
🔗 day_of_the_month_number: INTEGER
🔗 day_of_the_month_text: VARCHAR(10)
🔗 month_number: INTEGER
🔗 month_text: VARCHAR(10)
🔗 day_of_the_year: INTEGER
🔗 northern_hemisphere_season: VARCHAR(10)
🔗 southern_hemisphere_season: VARCHAR(10)
🔗 week_of_the_year: INTEGER
🔗 quarter_number: INTEGER
🔗 year_number: INTEGER

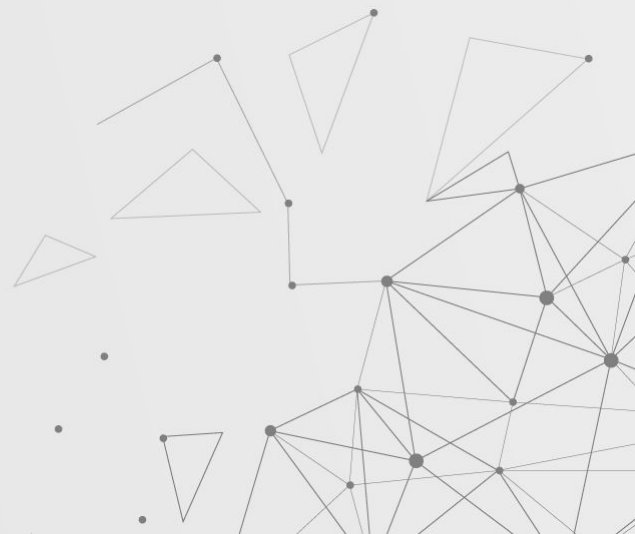
dim_vendor
🔑 dim_vendor_id: INTEGER
🔗 vendor_id: INTEGER
🔗 account_number: VARCHAR(15)
🔗 name: VARCHAR(50)
🔗 credit_rating_id: INTEGER
🔗 preferred_vendor: BOOLEAN
🔗 last_effective_date: DATETIME
🔗 current_flag: BOOLEAN
🔗 version_number: INTEGER
🔗 start_date: DATETIME
🔗 end_date: DATETIME

ANALYTICAL MODEL



PRODUCT DIMENSION HIGHLIGHTS

Column Name	Key?	Description	Source
dim_product_id	Yes	Surrogate key	Generated in the mySql DB using the add sequence feature in Kettle
product_id	No	ID for products Primary key of transactional DB	Extracted from AdventureWorks2019 table Production.Product column ProductID
product_number	No	Product number	Extracted from AdventureWorks2019 table Production.Product column ProductNumber
start_date	No	Start effective date for the product version	Calculated in Kettle, based on ADventureworks2019 changes
end_date	No	End effective date for the product version	Calculated in Kettle, based on ADventureworks2019 changes
version_number	No	Version number representing the sequence where a product has changed in history, after the dimension first load	Calculated in Kettle, based on ADventureworks2019 changes



PRODUCT TRANSFORMATION HIGHLIGHTS

Dimension lookup/update

Step name: Insert / Update dim_product

Update the dimension? ☒

Connection: Local-MySQL

Target schema: eae_data_management_mmjja

Target table: dim_product

Commit size: 100

Enable the cache? ☒

Pre-load the cache? ☐

Cache size in rows (0 = cache all): 5000

Keys Fields

Lookup/Update fields

#	Dimension field	Stream field to compare with	Type of dimension
1	product_number	ProductNumber	Insert
2	product_name	Name	Insert
3	make_flag	MakeFlag	Insert
4	finished_goods_flag	FinishedGoodsFlag	Insert
5	color	Color	Insert

Technical key field: dim_product_id

Creation of technical key

☒ Use table maximum + 1

☐ Use sequence

☐ Use auto increment field

Version field: version_number

Stream Datefield:

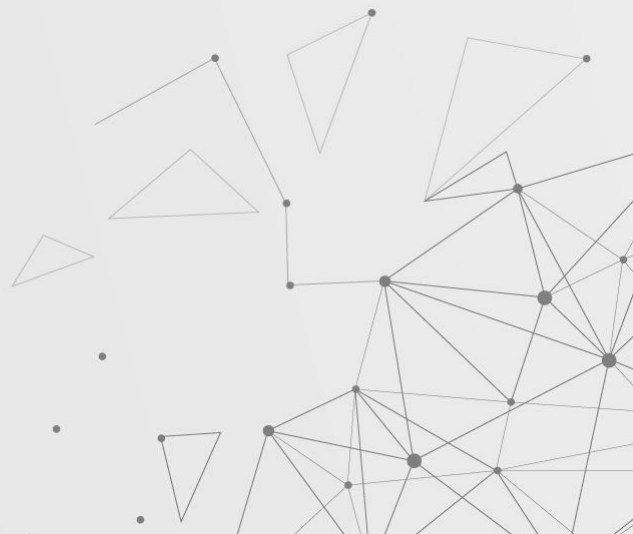
Date range start field: start_date

Use an alternative start date? ☐ <Select Option>

Table date range end: end_date

SDC TYPE #2

Using Dimension
lookup/update component



PRODUCT TRANSFORMATION HIGHLIGHTS

SCHEMAS

Filter objects

▼ eae_data_management_

- Tables
- Views
- Stored Procedures
- Functions

▶ staging_data_management_

▶ sys

```
28 select dim_product_id, product_id,  
29 product_name, product_category_name, product_subcategory_name,  
30 version_number, start_date, end_date  
31 from dim_product  
32 where product_id = 875
```

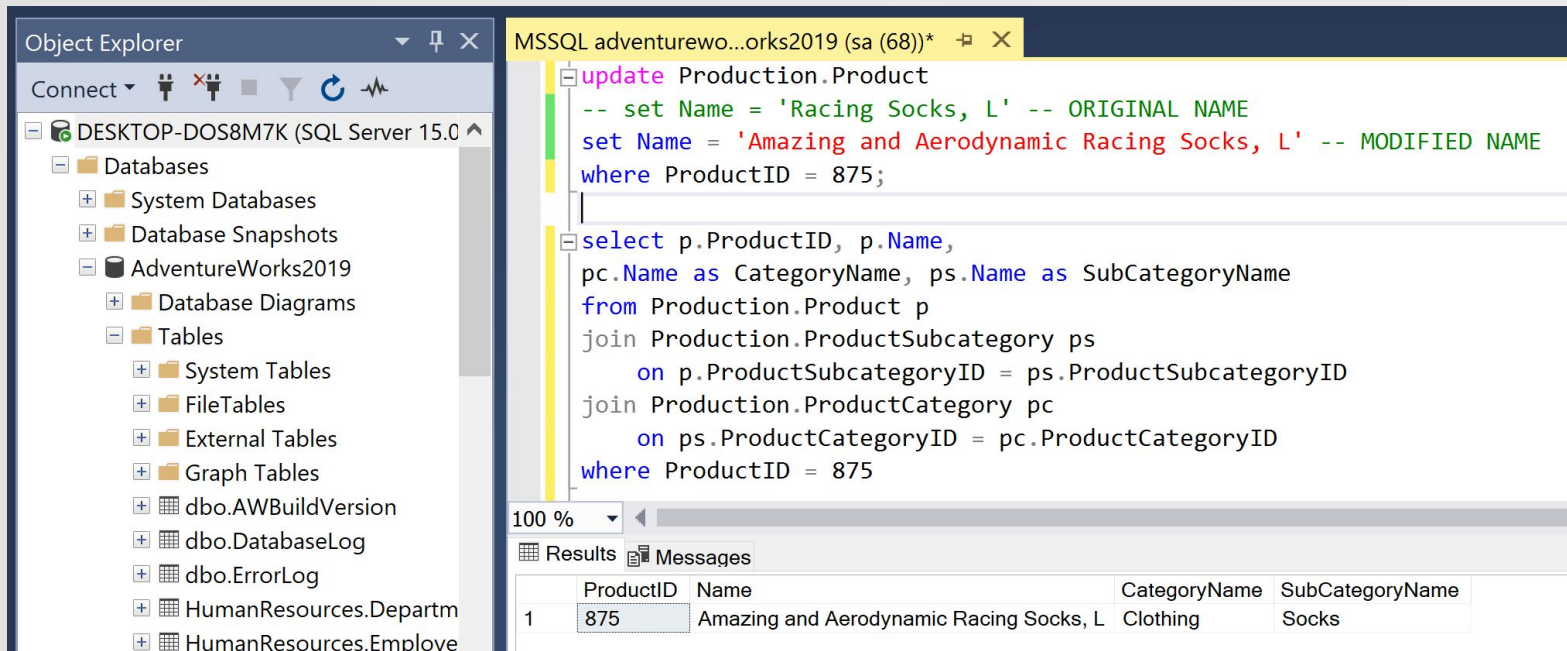
Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: |

	dim_product_id	product_id	product_name	product_category_name	product_subcategory_name	version_number	start_date	end_date
▶	380	875	Racing Socks, L	Clothing	Socks	1	1900-01-01 00:00:00	2200-01-01 00:00:00
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Administration Schemas

INITIAL STATE OF PRODUCT

PRODUCT TRANSFORMATION HIGHLIGHTS



The screenshot displays the SQL Server Enterprise Manager interface on the left, showing the 'AdventureWorks2019' database structure. The right pane shows a SQL Query Editor window with the following T-SQL code:

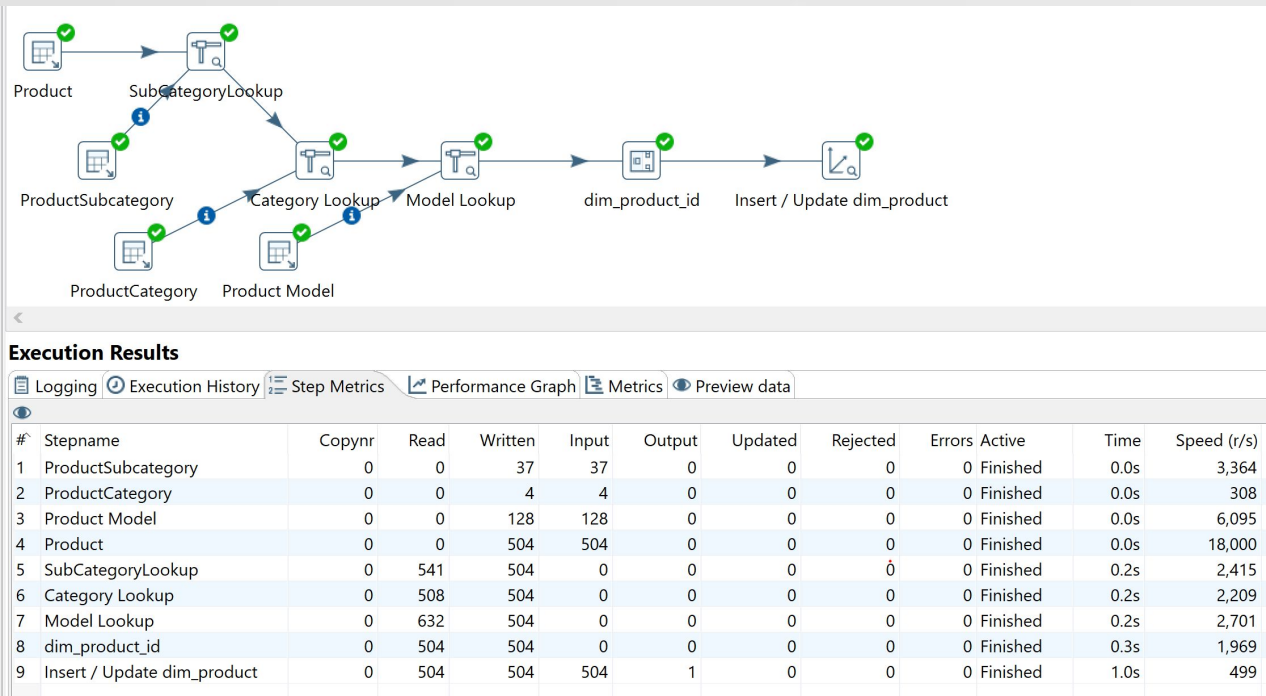
```
update Production.Product
-- set Name = 'Racing Socks, L' -- ORIGINAL NAME
set Name = 'Amazing and Aerodynamic Racing Socks, L' -- MODIFIED NAME
where ProductID = 875;
```

Below the code, the 'Results' tab shows the output of the query:

	ProductID	Name	CategoryName	SubCategoryName
1	875	Amazing and Aerodynamic Racing Socks, L	Clothing	Socks

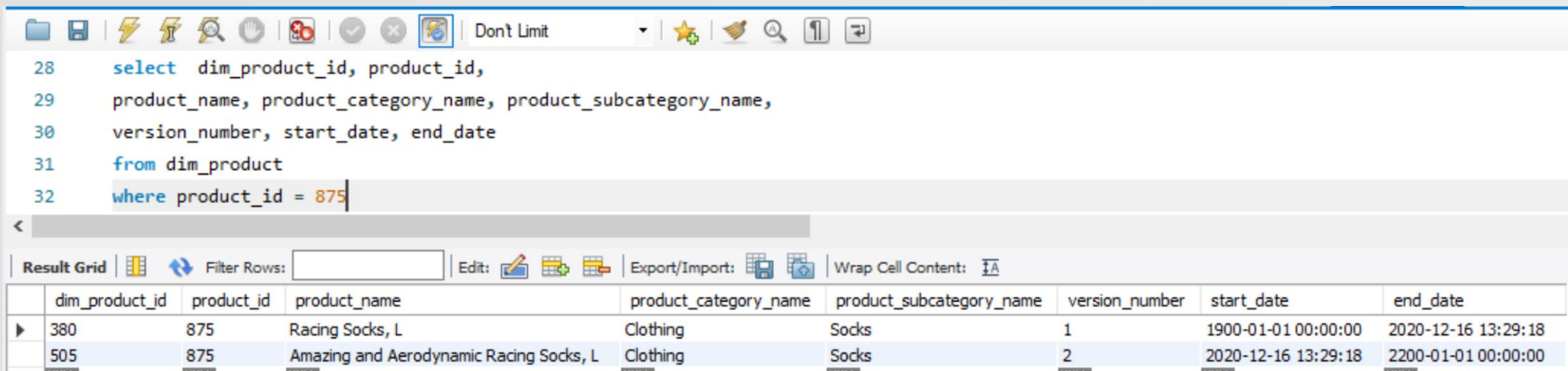
**NAME CHANGED AT
TRANSACTIONAL DATABASE**

PRODUCT TRANSFORMATION HIGHLIGHTS



1 OUTPUT, INSERT DETECTED

PRODUCT TRANSFORMATION HIGHLIGHTS



The screenshot shows a SQL query editor interface. The query is as follows:

```
28 select dim_product_id, product_id,  
29 product_name, product_category_name, product_subcategory_name,  
30 version_number, start_date, end_date  
31 from dim_product  
32 where product_id = 875
```

Below the query editor is the "Result Grid" showing the results of the query. The grid has 8 columns: dim_product_id, product_id, product_name, product_category_name, product_subcategory_name, version_number, start_date, and end_date. There are two rows of data.

dim_product_id	product_id	product_name	product_category_name	product_subcategory_name	version_number	start_date	end_date
380	875	Racing Socks, L	Clothing	Socks	1	1900-01-01 00:00:00	2020-12-16 13:29:18
505	875	Amazing and Aerodynamic Racing Socks, L	Clothing	Socks	2	2020-12-16 13:29:18	2200-01-01 00:00:00



PRODUCT SECOND VERSION

DATE DIMENSION HIGHLIGHTS

Column Name	Key?	Description	Source
dim_date_id	Yes	This is the surrogate key It is an integer with the following format: 20201214	Generated in a staging database, that lists days in a datarange YEAR * 10,000 + MONTH * 100 + DAY
is_holiday	No	Flag indicating if it is a holiday or not	Holiday data is exrtracted via an API from https://date.nager.at/api/v2/publicholidays/2020/es
holiday_name	No	National HOliday name for Spain	Holiday data is exrtracted via an API from https://date.nager.at/api/v2/publicholidays/2020/es
date	No	Represents the day of the row, in a date dabase field	Generated in a staging database, that lists days in a datarange
northern_hemisphere_season	No	Season text in english for the northern hemisphere text: (Winter, Autum, Summer, Fall)	Calculated from a csv source
southern_hemisphere_season	No	Season text in english for the southern hemisphere text: (Winter, Autum, Summer, Fall)	Calculated from a csv source

DATE TRANSFORMATION HIGHLIGHTS

```
17 • use staging_data_management_mmjja;
18
19 • DROP PROCEDURE IF EXISTS fill_staging_date_table;
20 DELIMITER //
21 • CREATE PROCEDURE fill_staging_date_table(IN start_date DATE, IN end_date DATE)
22 begin
23     declare number_of_days bigint;
24     set number_of_days = DATEDIFF( date_add(end_date, interval 1 day), start_date);
25
26     -- throw error if range out of bounds
27     if (number_of_days <= 0) then
28         SIGNAL SQLSTATE '45000' SET MESSAGE_TEXT = 'End date should be greater than Start date';
29     end if;
30     if (number_of_days > 999999) then
31         SIGNAL SQLSTATE '45000' SET MESSAGE_TEXT = 'Cannot generate more than 10^6 days, 2739 calendar years';
32     end if;
33
34     -- generate a list of sequence numbers 0 ... number_of_days
35     insert into staging_date
36     (
37         date_id, date,
38         day_of_the_week_number, day_of_the_week_text,
39         day_of_the_month_number, day_of_the_month_text,
40         month_number, month_text,
41         day_of_the_year,
42         week_of_the_year,
```

DATE-RANGE AT MySQL STAGING DATABASE



DATE TRANSFORMATION HIGHLIGHTS


► Nager Date Spain 2020

GET ▼ https://date.nager.at/api/v2/PublicHolidays/2020/ES

Params Authorization Headers Body Pre-request Script Tests

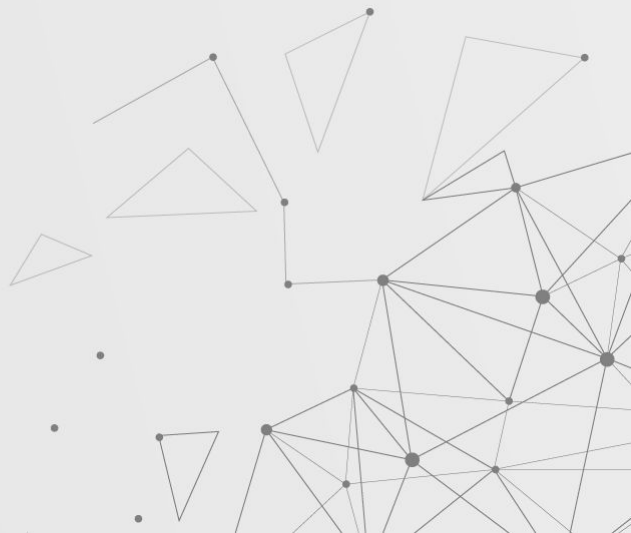
KEY	VALUE
Key	Value

Body Cookies Headers (7) Test Results

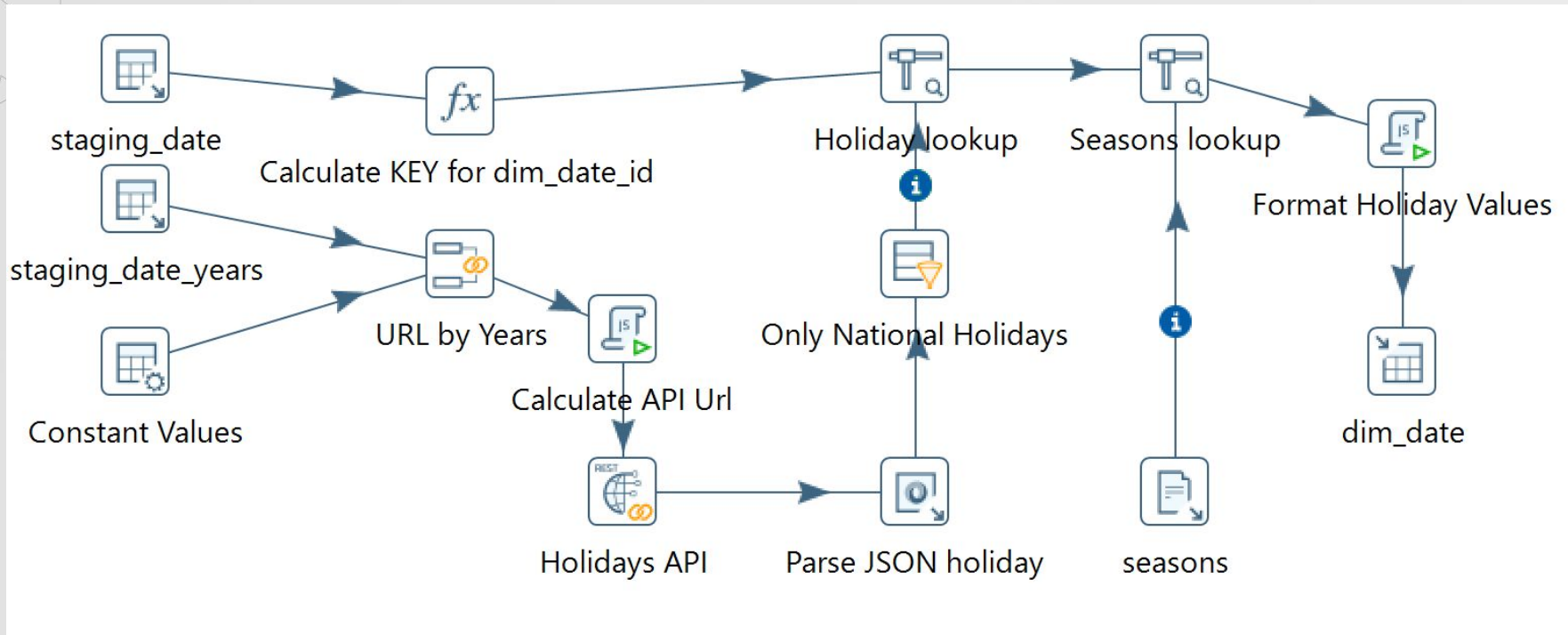
Pretty Raw Preview JSON ⌵ 

```
1  {
2    {
3      "date": "2020-01-01",
4      "localName": "Año Nuevo",
5      "name": "New Year's Day",
6      "countryCode": "ES",
7      "fixed": true,
8      "global": true,
9      "counties": null,
10     "launchYear": 1967,
11     "type": "Public"
12   },
13   {
14     "date": "2020-01-06",
15     "localName": "Día de Reyes / Epifanía del Señor",
16     "name": "Epiphany",
```

HOLIDAY API BY YEAR

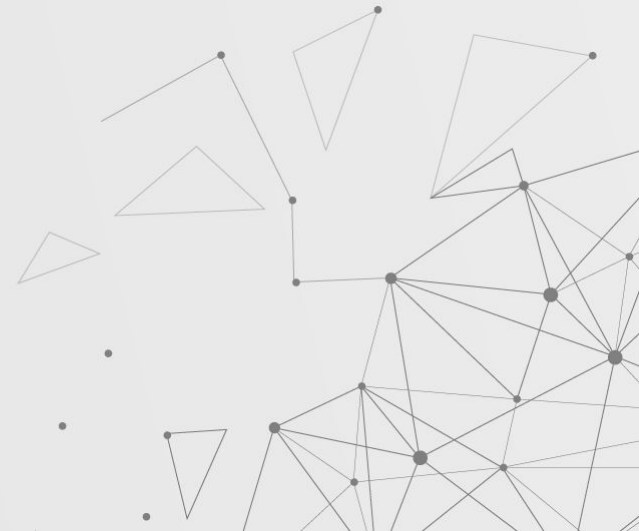


DATE TRANSFORMATION HIGHLIGHTS

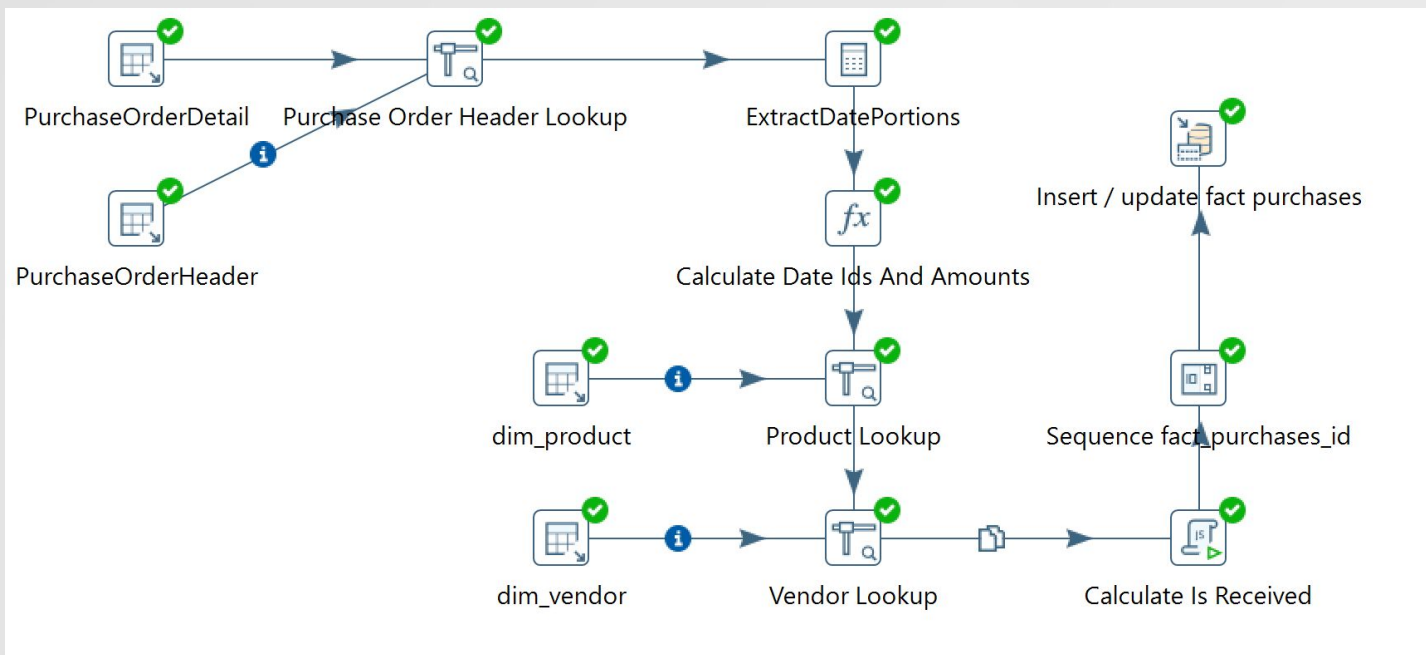


PURCHASES FACT HIGHLIGHTS

Column Name	Key?	Type	Description	Source
fact_purchases_id	Yes	Integer	Surrogate key	Generated in the mySql DB using the add sequence feature in kettle
dim_product_id	No	Integer	ID of the product	Foreign key from table dim_product MySql database
product_unit_price	No	Integer	Unit price of the product	AdventureWorks PurchaseOrderDetail
rejected_qty	No	Integer	Quantity of rejected items	AdventureWorks PurchaseOrderDetail
received_qty	No	Integer	Quantity of items received from the vendor	AdventureWorks PurchaseOrderDetail
stocked_qty	No	Integer	Quantity of accepted items into inventory	AdventureWorks PurchaseOrderDetail
product_standard_cost_on_date	No	Integer	The standard cost of the item that day	Calculated column matching the Standard Cost date in the SalesHistory, that contains the Sales.SalesOrderHeader
rejected_amount	No	Integer	Value of rejected items	Calculated column from [RejectedQty] * [UnitPrice]
received_amount	No	Integer	Value of received items	Calculated column from [ReceivedQty] * [UnitPrice]
stocked_amount	No	Integer	Value of items in inventory	Calculated column from [StockedQty] * [UnitPrice]
received_flag	No	Boolean	Flag indicating if	Created column in Kettle by using a



PURCHASES TRANSFORMATION HIGHLIGHTS



PURCHASES FACT HIGHLIGHTS

Calculator

Step name
ExtractDatePortions

☒ Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Pr
1	order_date_year	Year of date A	OrderDate			None		
2	order_date_month	Month of date A	OrderDate			None		
3	order_date_day	Day of month of date A	OrderDate			None		
4	ship_date_year	Year of date A	ShipDate			None		
5	ship_date_month	Month of date A	ShipDate			None		
6	ship_date_day	Day of month of date A	ShipDate			None		

Help OK Cancel

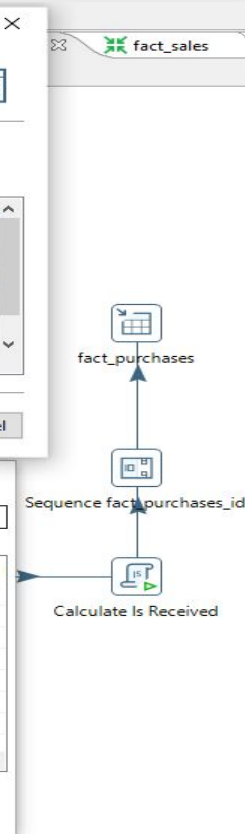
Formula

Step name Calculate Date Ids And Amounts

Fields:

#	New field	Formula	Value type
1	calculated_order_date_id	[order_date_year]*10000 + [order_date_month]*100 + [order_date_day]	Integer
2	calculated_ship_date_id	[ship_date_year]*10000 + [ship_date_month]*100 + [ship_date_day]	Integer
3	calculated_rejected_amount	[RejectedQty] * [UnitPrice]	Number
4	calculated_stocked_amount	[StockedQty] * [UnitPrice]	Number
5	calculated_received_amount	[ReceivedQty] * [UnitPrice]	Number

Help OK Cancel

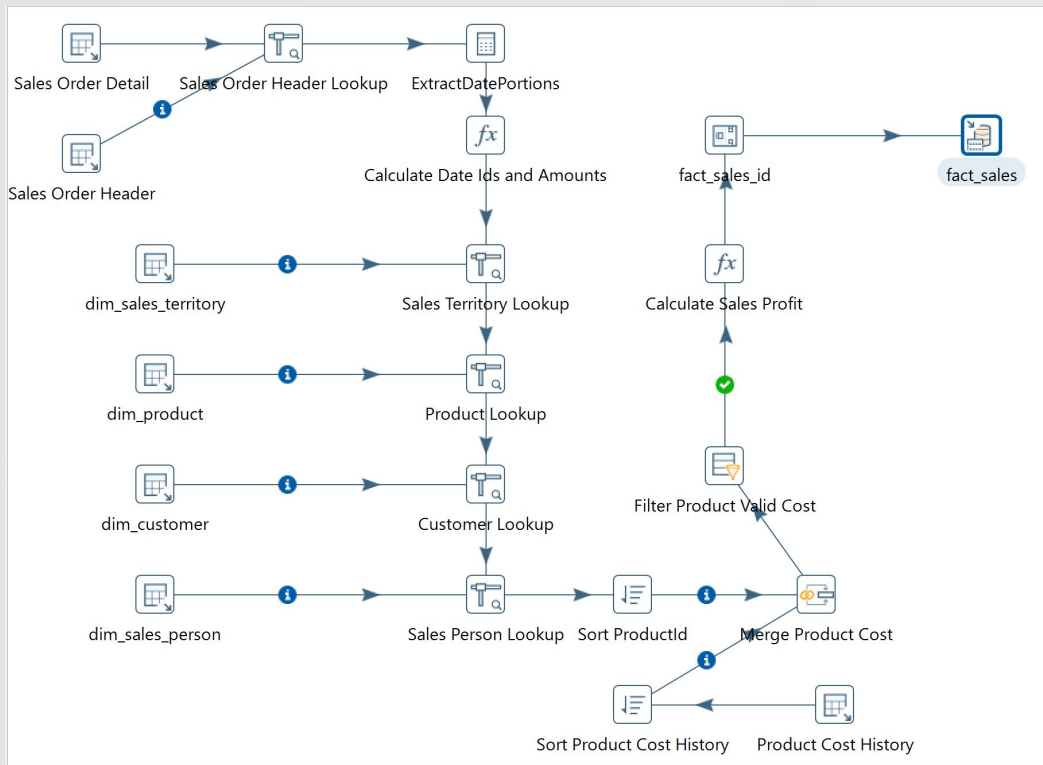


SALES FACT HIGHLIGHTS

Column Name	Key ?	Description	Source
fact_sales_id	Yes	Surrogate key	Generated by "Add sequence" in Kettle
product_standard_cost_on_day	No	The standard cost for the product that day	Calculated in Kettle
sales_profit	No	Profit of sale	Calculated in Kettle



SALES TRANSFORMATION HIGHLIGHTS



SALES FACT HIGHLIGHTS

Table input

Step name: Product Cost History

Connection: Local MSSQL - adventureworks

SQL

```
SELECT
ProductID as ProductCostHistoryProductId,
StartDate,
isnull(EndDate, Convert(DateTime, '2100-12-31',102)) as EndDate,
StandardCost
FROM Production.ProductCostHistory
```

Line 1 Column 0

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

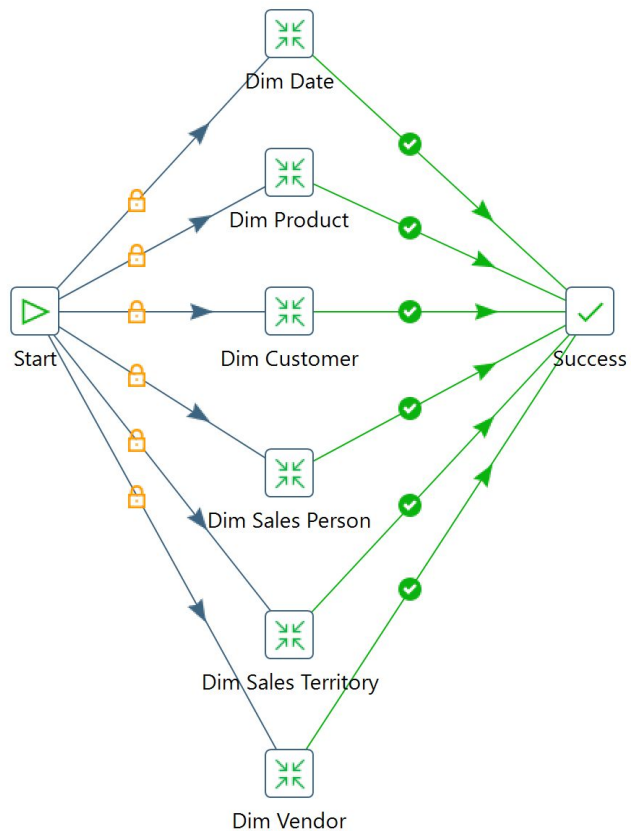
Limit size: 0

OK Preview Cancel

Product Standard Cost

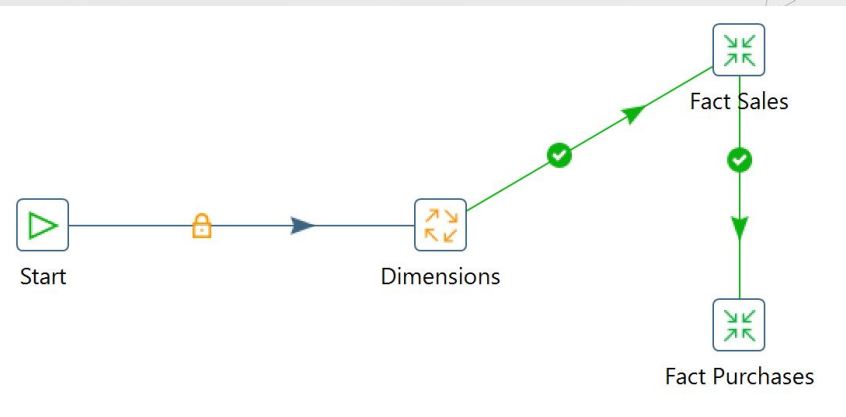


DIMENSIONS



JOBS

FACTS

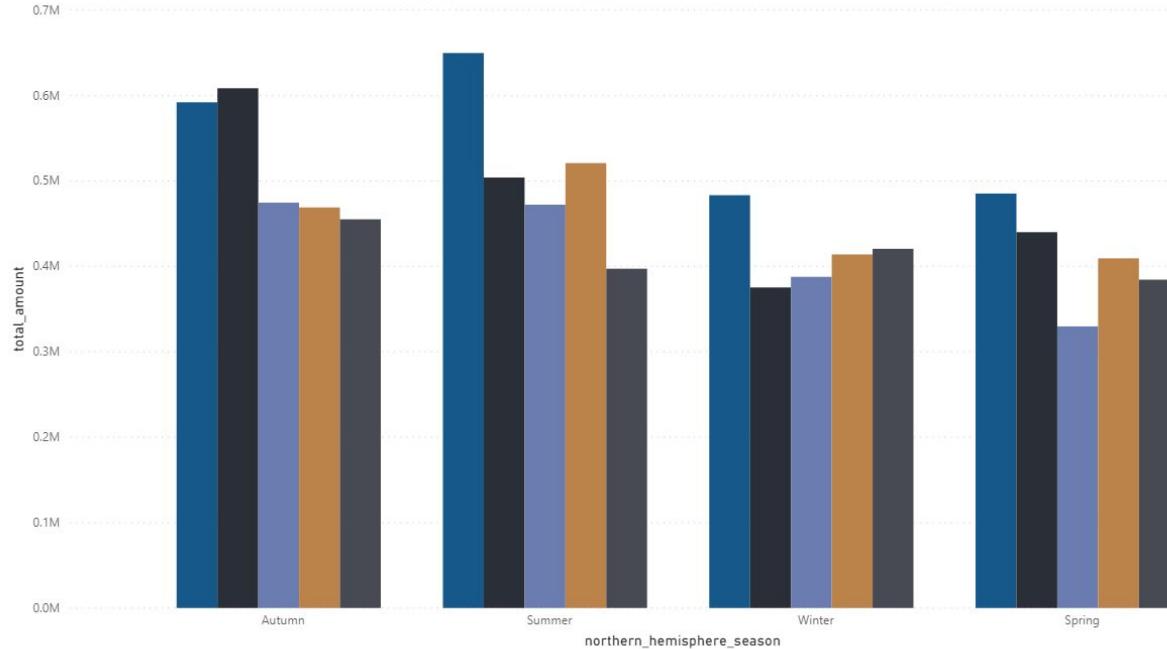


SEASONAL SALES

Create top 5 products sales ranking by seasons.

total_amount by northern_hemisphere_season and product_name

product_name ● Mountain-200 Black, 38 ● Mountain-200 Black, 42 ● Mountain-200 Black, 46 ● Mountain-200 Silver, 38 ● Mountain-200 Silver, 46



Filters

Search

Filters on this page

order_date_dateti...

is 2013

- ☐ 2011
- ☐ 2012
- ☒ 2013
- ☐ 2014

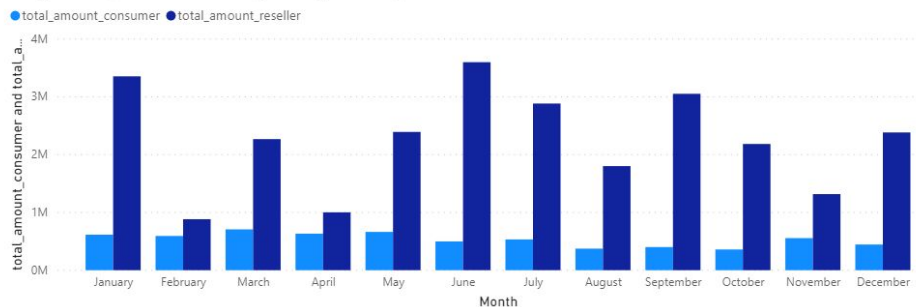
HOLIDAY SALES

Reseller sales vs direct consumer sales
over specific holiday days.

total_amount_consumer and total_amount_reseller by holiday_name



total_amount_consumer and total_amount_reseller by Month



Filters

Search

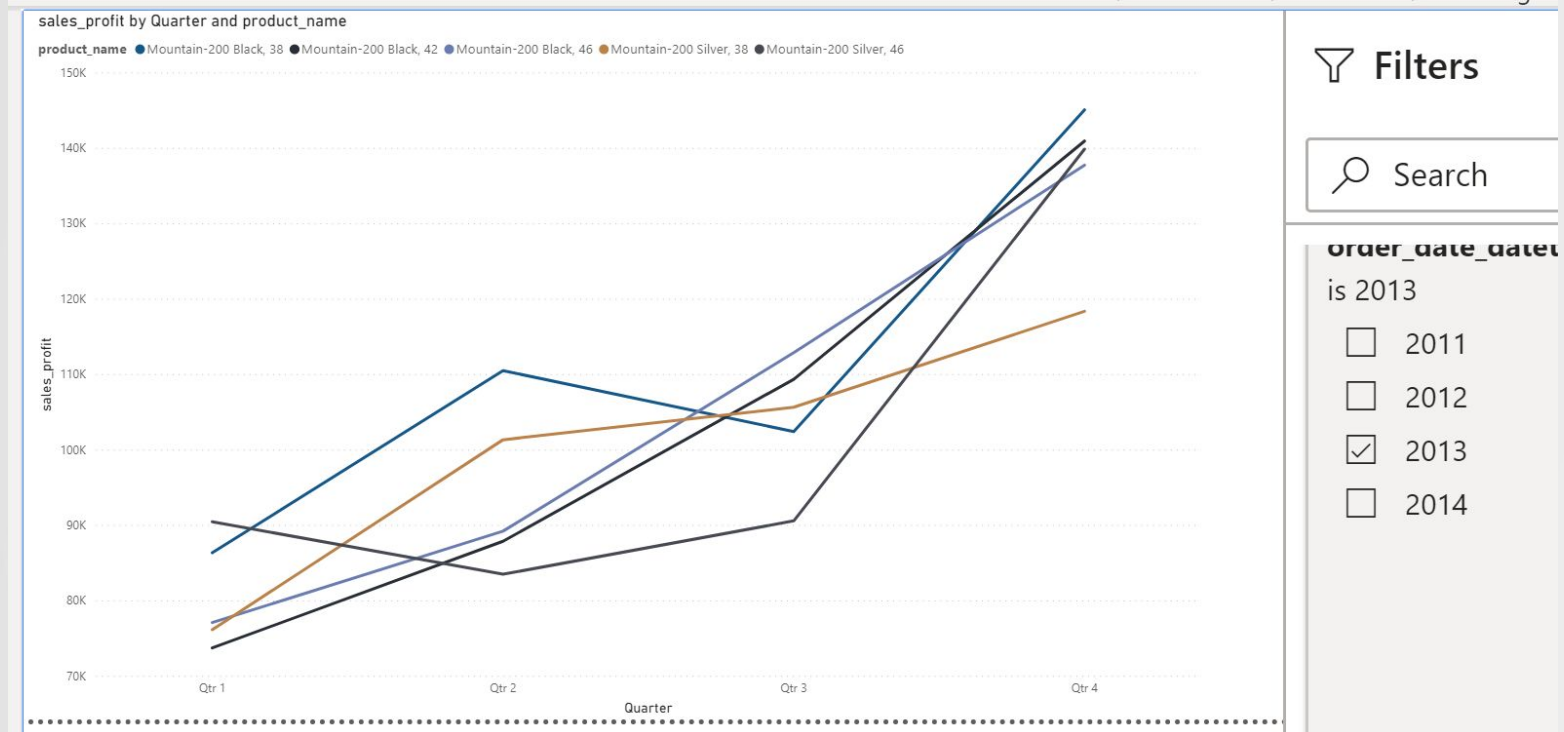
Filters on this page

order_date_dateti...
is 2012

- ☐ 2011
- ☒ 2012
- ☐ 2013
- ☐ 2014

PRODUCT PROFITABILITY

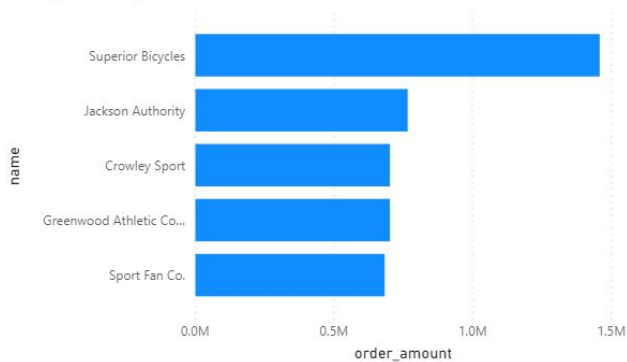
Create top 5 products profit ranking, quarter by quarter



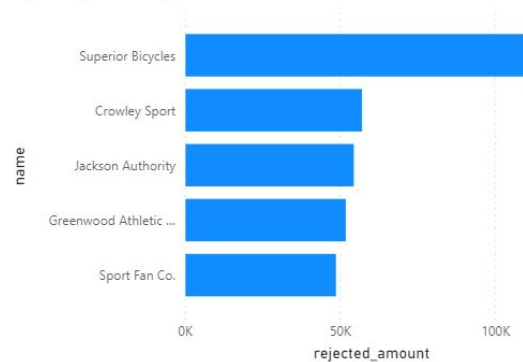
REJECTED ITEMS

Create a top 5 Vendor ranking of rejected items on purchase orders.

order_amount by name



rejected_amount by name



 Search

Filters on this page

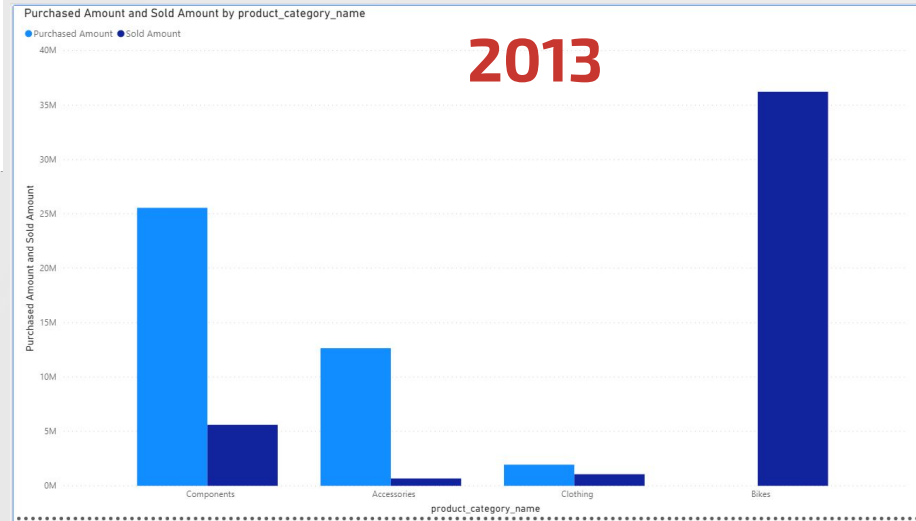
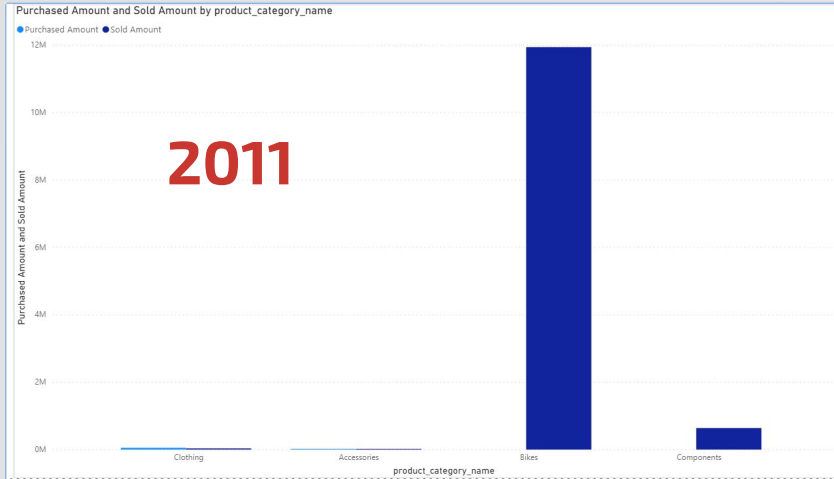
order_date_date

is 2013

- ☐ 2011
- ☐ 2012
- ☒ 2013
- ☐ 2014

PURCHASED VS SOLD

Sold items vs purchased items,
for each product category





Thank You