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

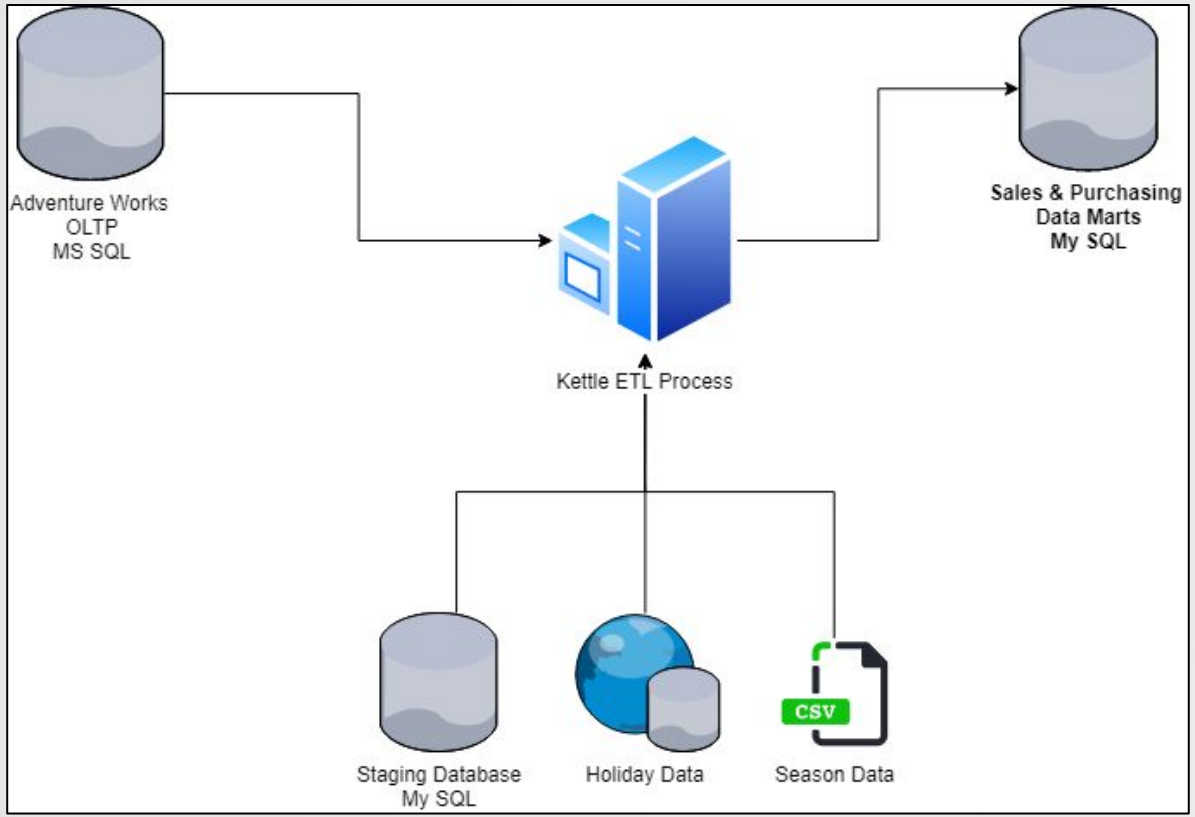
REJECTED ITEMS

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

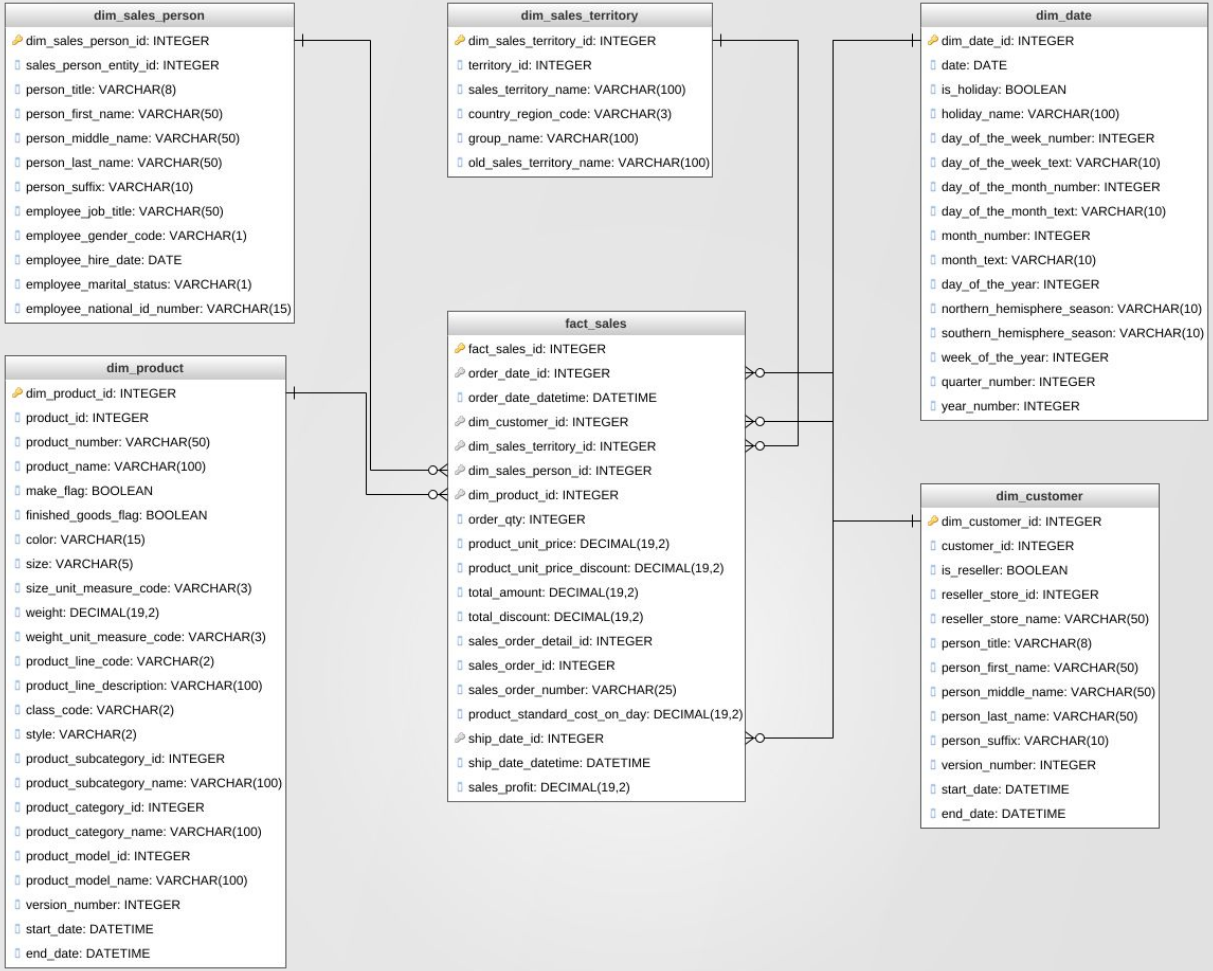
PURCHASED VS SOLD

Sold items vs purchased items, for each product category

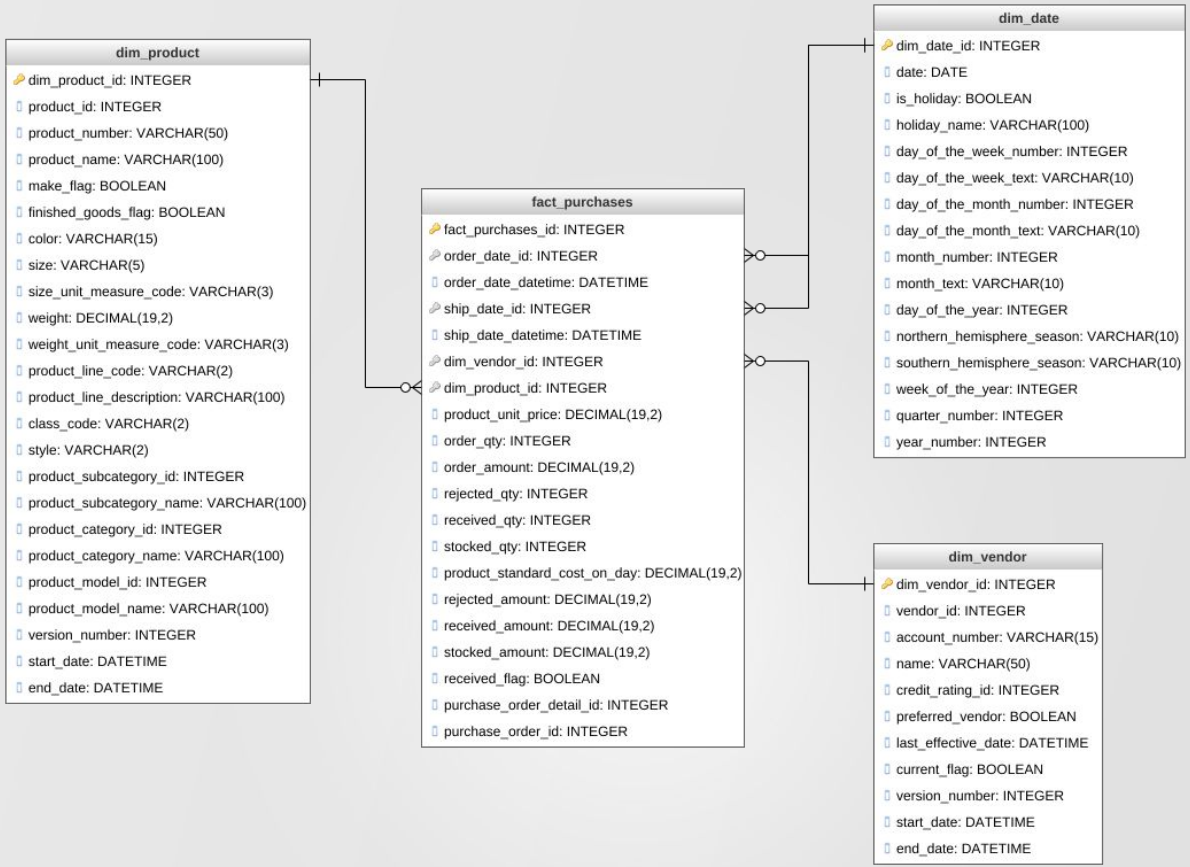
ARCHITECTURE



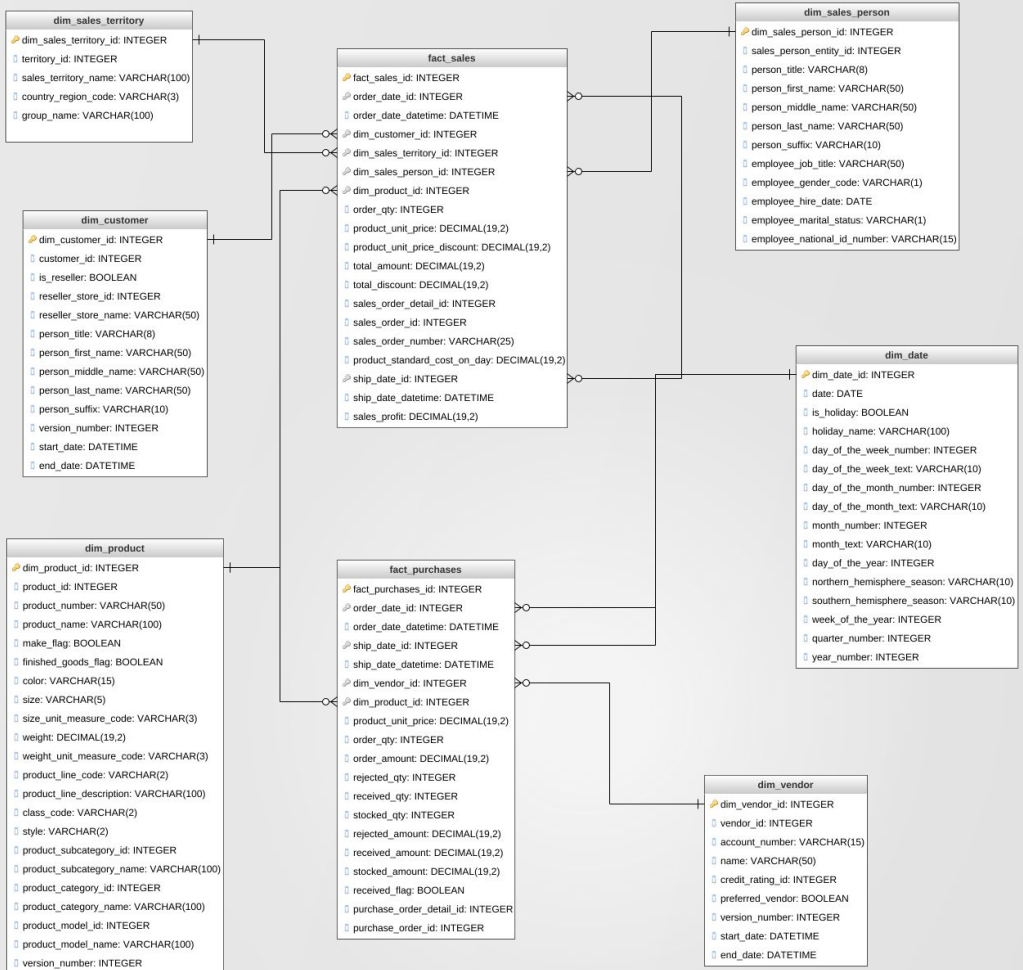
SALES DATA MART



PURCHASING DATA MART



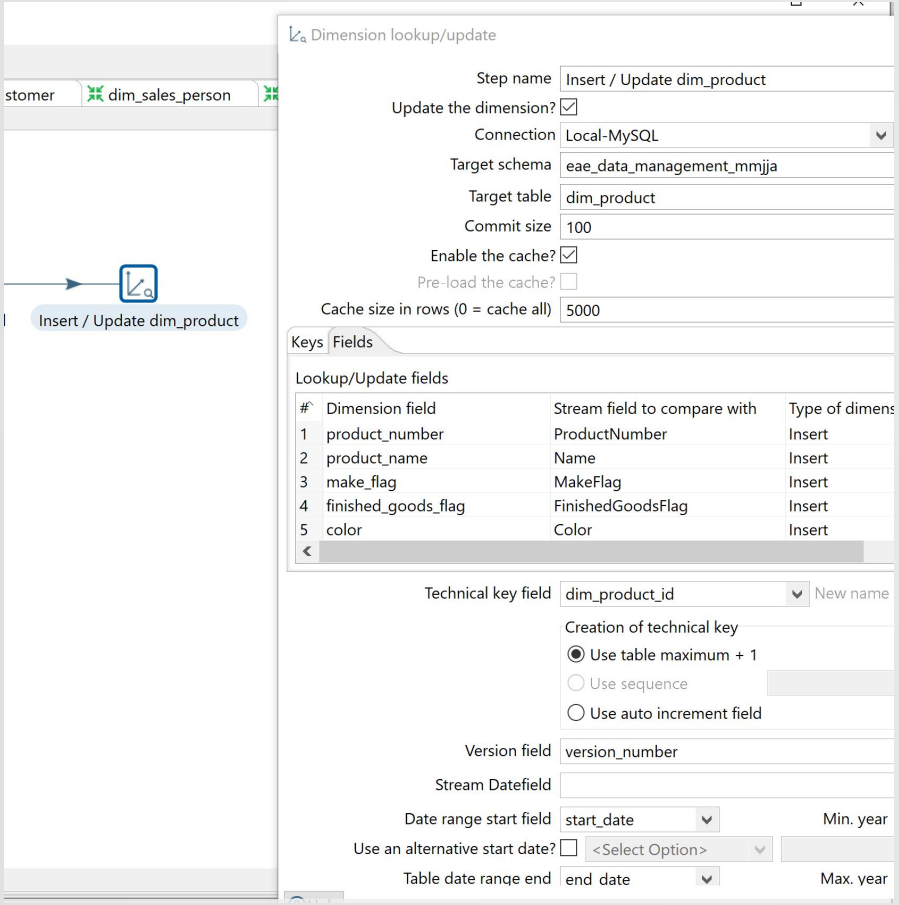
ANALYTICAL MODEL



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

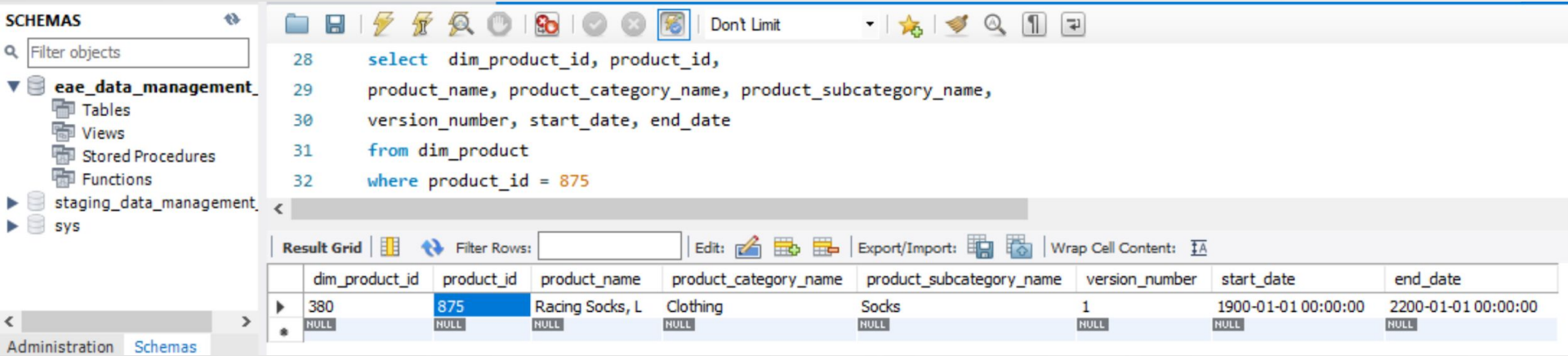
PRODUCT TRANSFORMATION HIGHLIGHTS



SDC TYPE #2

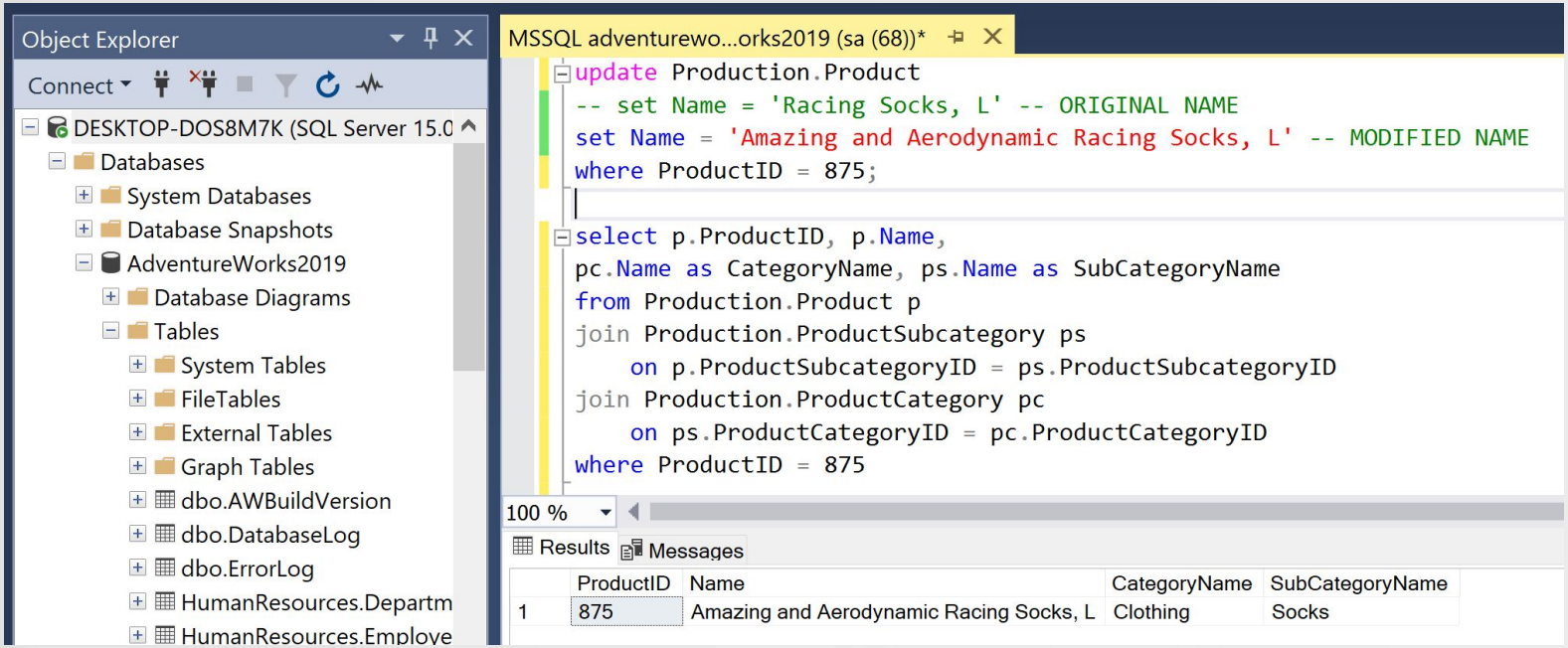
Using Dimension lookup/update component

PRODUCT TRANSFORMATION HIGHLIGHTS



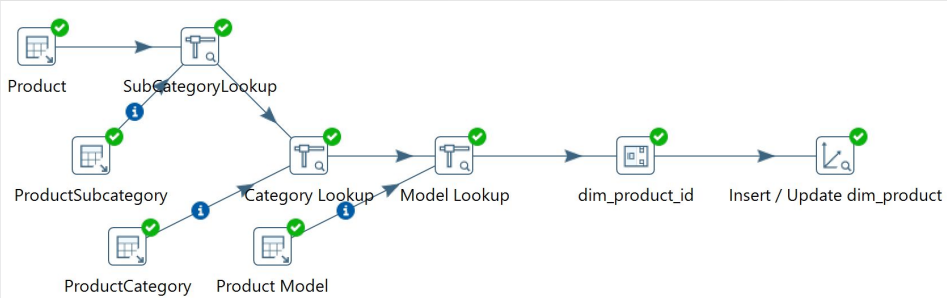
INITIAL STATE OF PRODUCT

PRODUCT TRANSFORMATION HIGHLIGHTS



NAME CHANGED AT TRANSACTIONAL DATABASE

PRODUCT TRANSFORMATION HIGHLIGHTS



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)
1	ProductSubcategory	0	0	37	37	0	0	0	0	Finished	0.0s	3,364
2	ProductCategory	0	0	4	4	0	0	0	0	Finished	0.0s	308
3	Product Model	0	0	128	128	0	0	0	0	Finished	0.0s	6,095
4	Product	0	0	504	504	0	0	0	0	Finished	0.0s	18,000
5	SubCategoryLookup	0	541	504	0	0	0	0	0	Finished	0.2s	2,415
6	Category Lookup	0	508	504	0	0	0	0	0	Finished	0.2s	2,209
7	Model Lookup	0	632	504	0	0	0	0	0	Finished	0.2s	2,701
8	dim_product_id	0	504	504	0	0	0	0	0	Finished	0.3s	1,969
9	Insert / Update dim_product	0	504	504	504	1	0	0	0	Finished	1.0s	499

1 OUTPUT, INSERT DETECTED

PRODUCT TRANSFORMATION HIGHLIGHTS

```
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	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

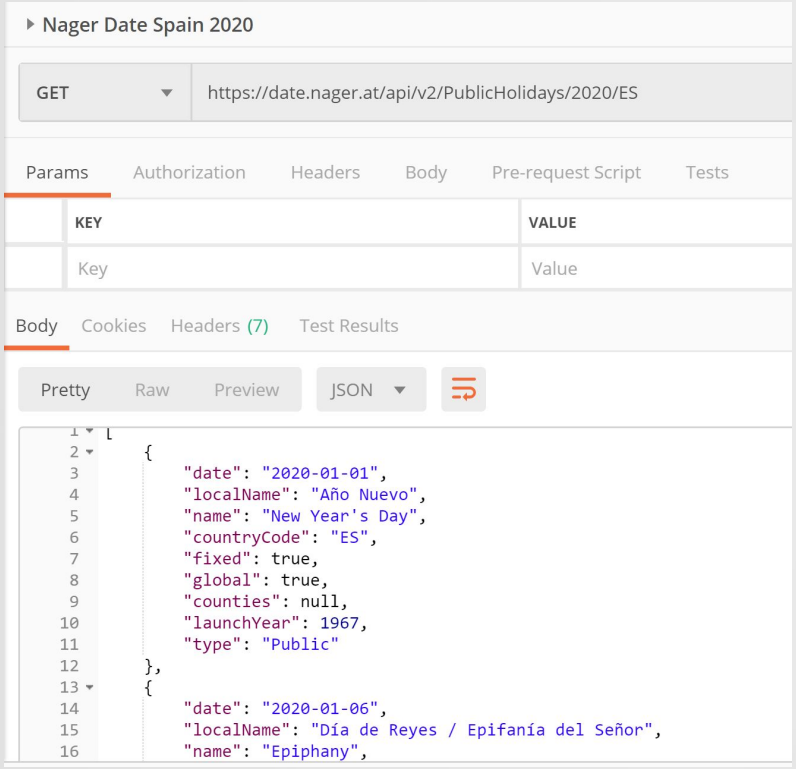
Column Name	Key?	Description	Source
dim_date_id	Yes	This is the surrogate key It is an integer with the following format: 20201214	Generated in a staging database, that lists days in a datarange **YEAR * 10,000 + MONTH * 100 + DAY**
is_holiday	No	Flag indicating if it is a holiday or not	Holiday data 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

```
use staging_data_management_mmjja;

DROP PROCEDURE IF EXISTS fill_staging_date_table;
DELIMITER //
CREATE PROCEDURE fill_staging_date_table(IN start_date DATE, IN end_date DATE)
begin
  declare number_of_days bigint;
  set number_of_days = DATEDIFF( date_add(end_date, interval 1 day), start_date);

  -- throw error if range out of bounds
  if (number_of_days <= 0) then
    SIGNAL SQLSTATE '45000' SET MESSAGE_TEXT = 'End date should be greater than Start date';
  end if;
  if (number_of_days > 999999) then
    SIGNAL SQLSTATE '45000' SET MESSAGE_TEXT = 'Cannot generate more than 10^6 days, 2739 calendar years
  end if;

  -- generate a list of sequence numbers 0 ... number_of_days
  insert into staging_date
  (
    date_id, date,
    day_of_the_week_number, day_of_the_week_text,
    day_of_the_month_number, day_of_the_month_text,
    month_number, month_text,
    day_of_the_year,
    week_of_the_year,
```

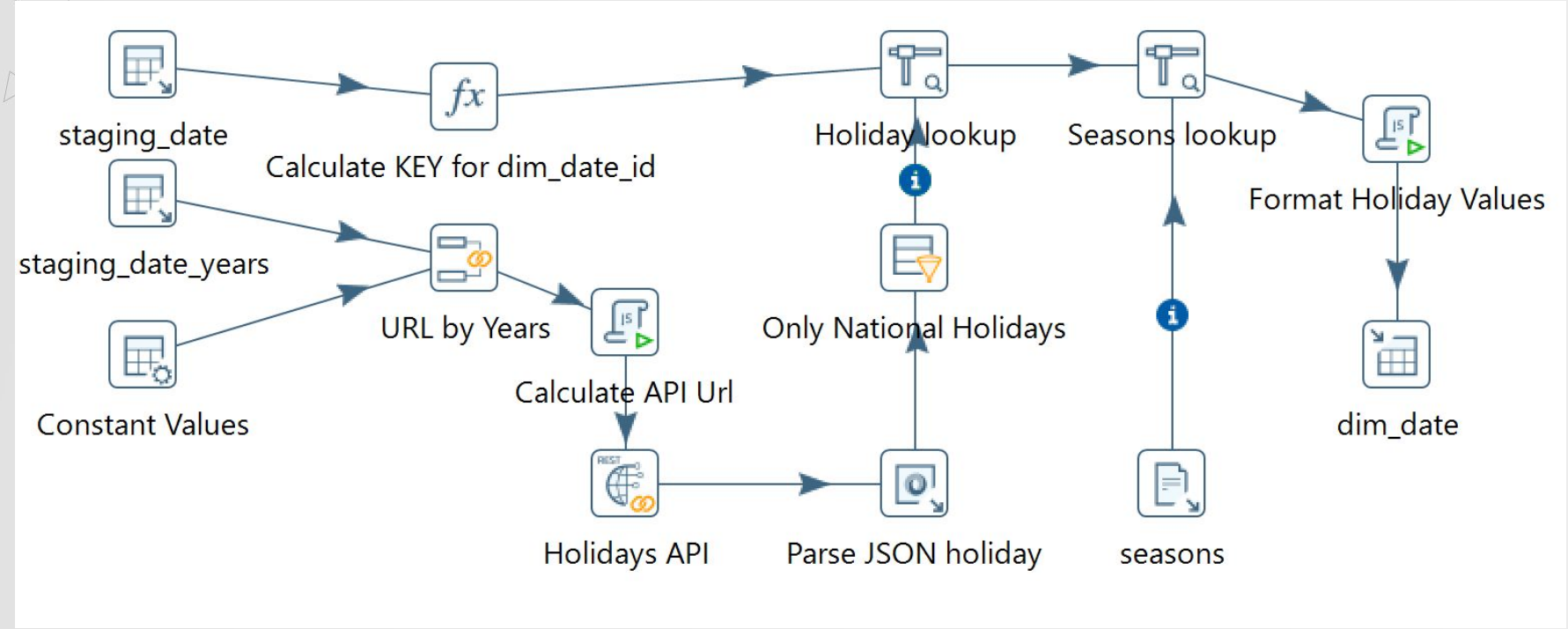
DATE-RANGE AT MySQL STAGING DATABASE

DATE TRANSFORMATION HIGHLIGHTS



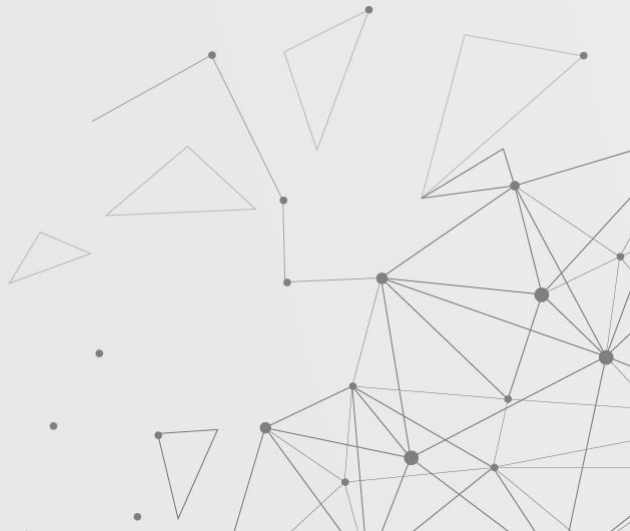
HOLIDAY API BY YEAR

DATE TRANSFORMATION HIGHLIGHTS

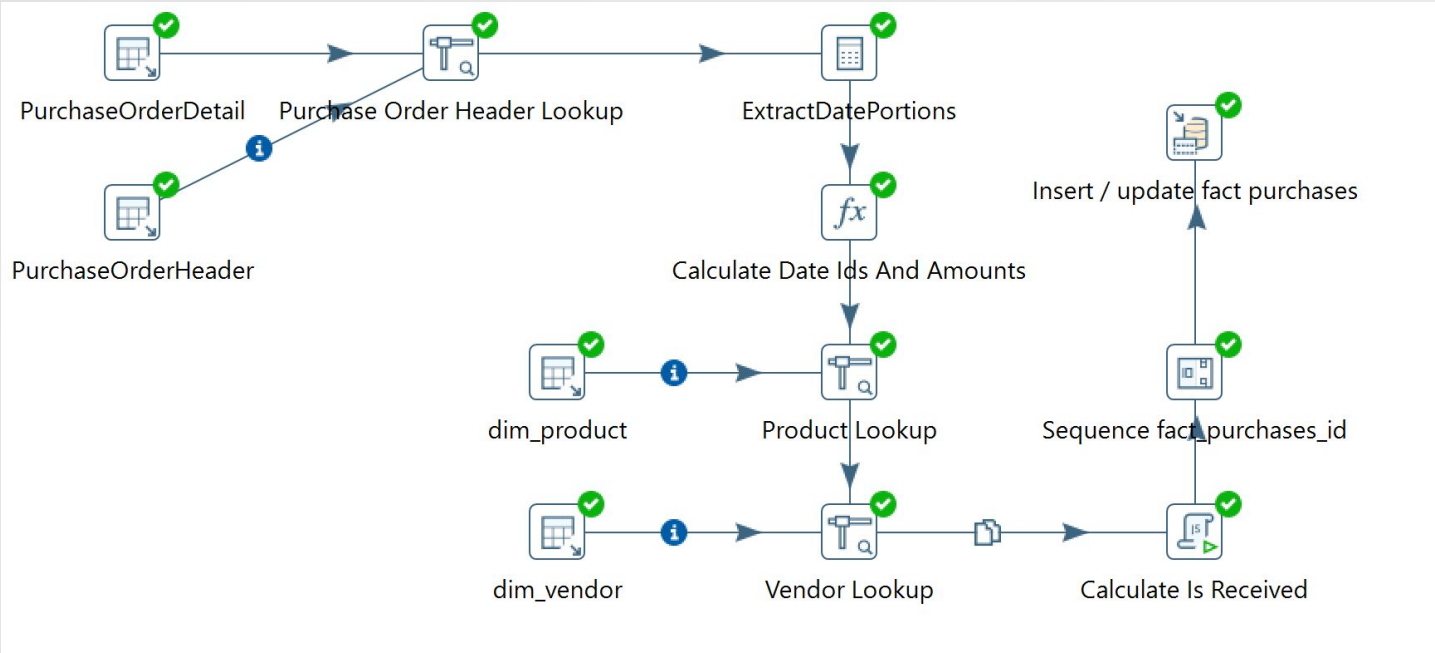


PURCHASES FACT HIGHLIGHTS

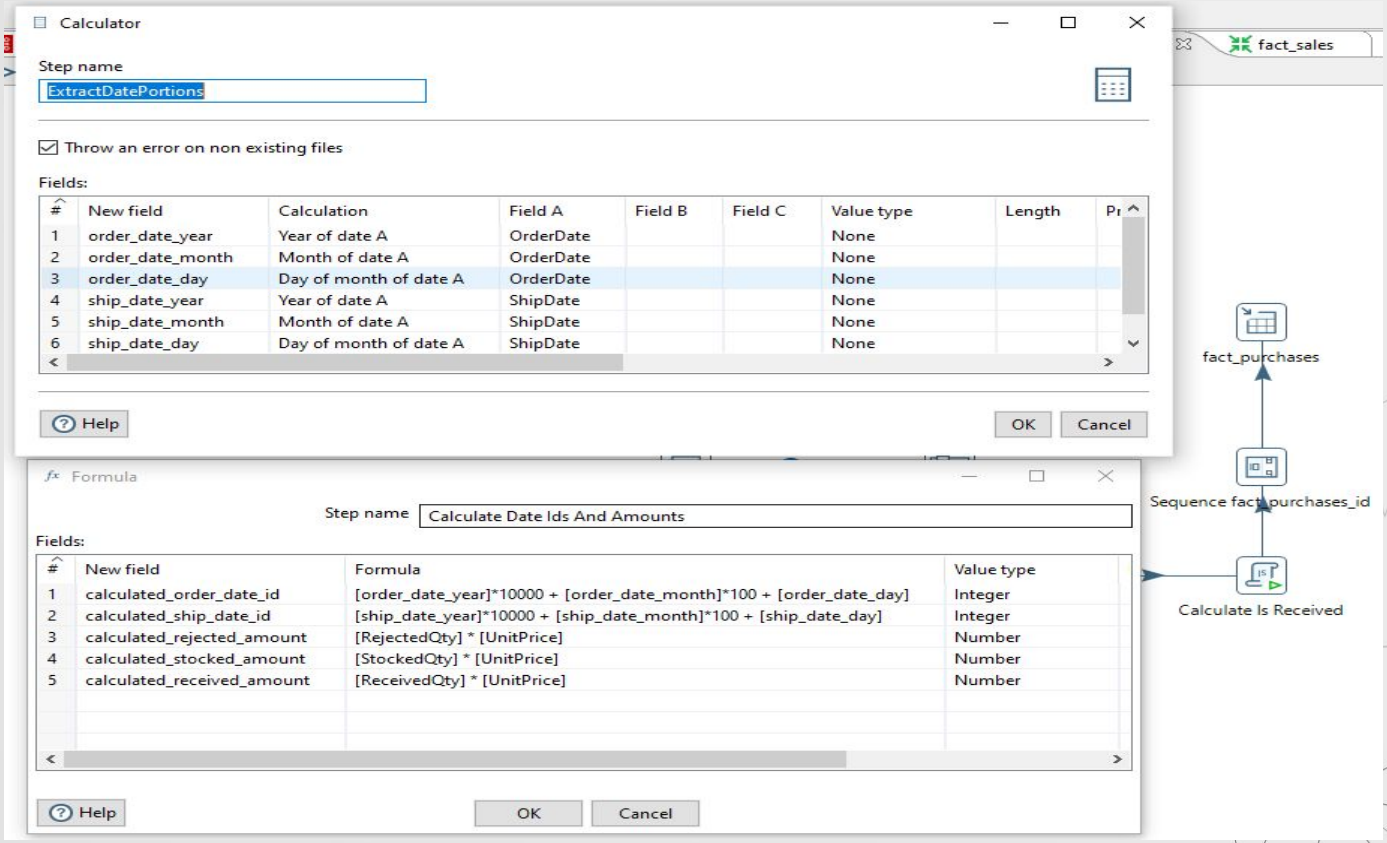
Column Name	Key?	Type	Description	Source
fact_purchases_id	Yes	Integer	Surrogate key	Generated in the mySql DB using the add sequence feature in kettle
dim_product_id	No	Integer	ID of the product	Foreign key from table dim_product MySql database
product_unit_price	No	Integer	Unit price of the product	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_day	No	Integer	The standard cost of the item that day	ated column matching the Standard Cost date i stHistory, that contains the Sales.SalesOrderHe
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



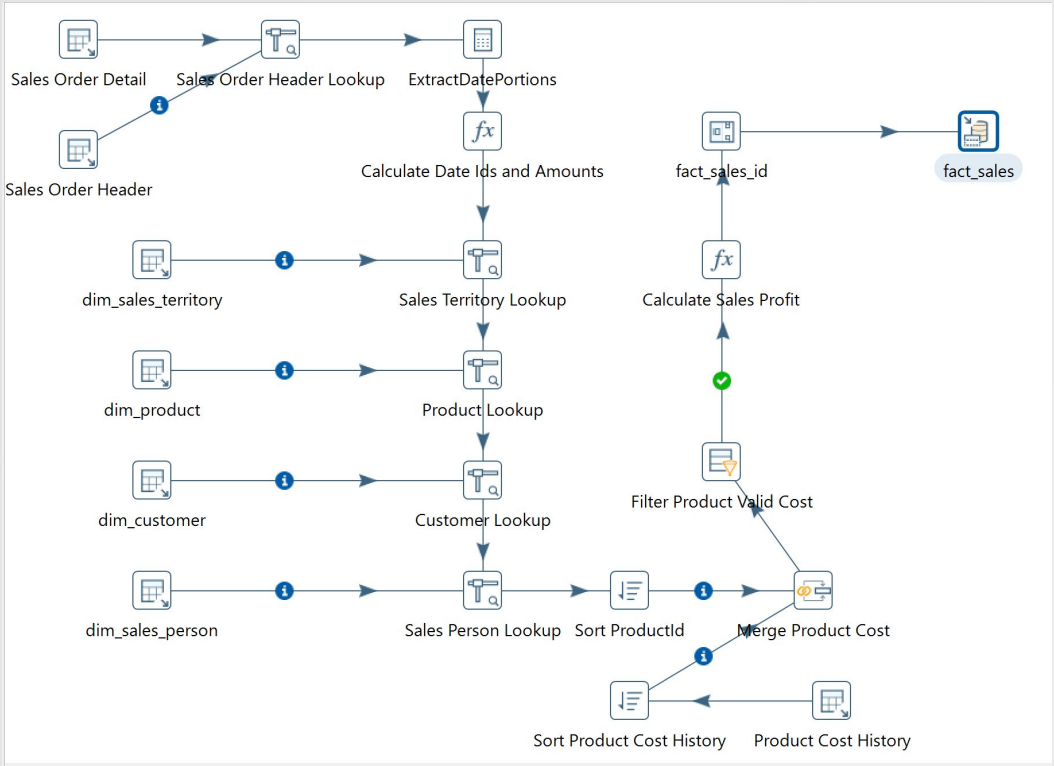
PURCHASES FACT HIGHLIGHTS



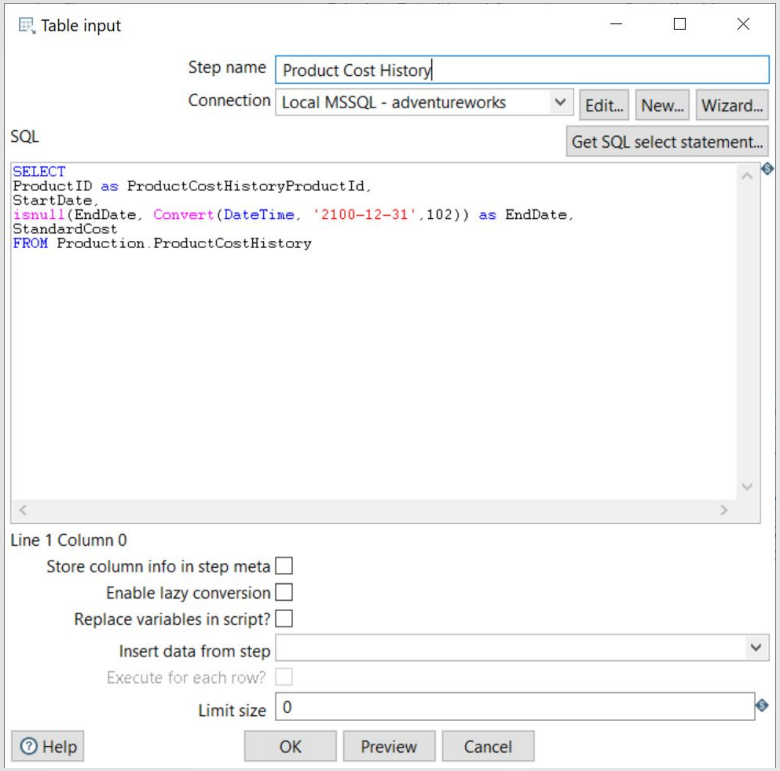
SALES FACT HIGHLIGHTS

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



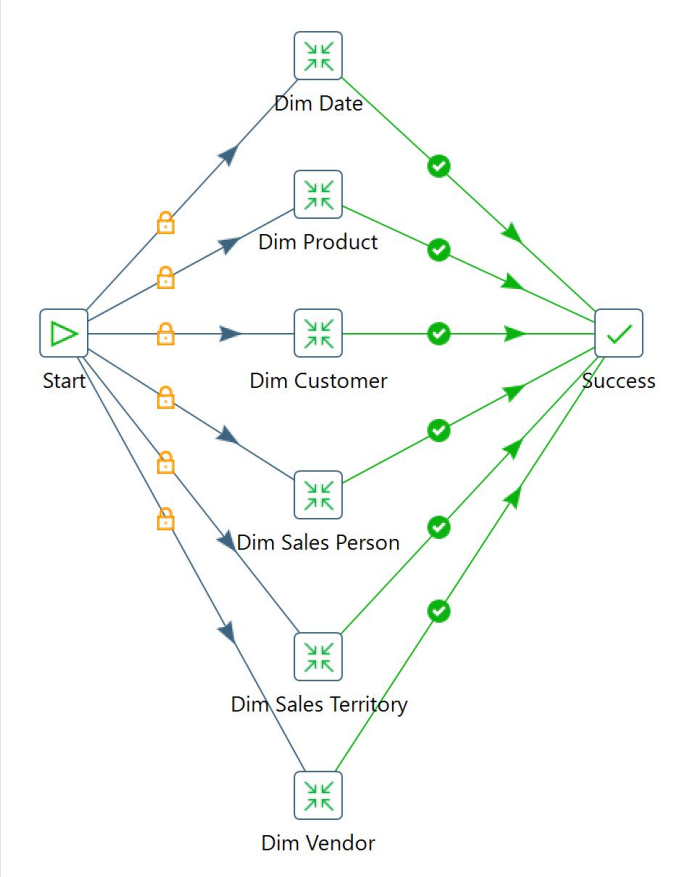
SALES FACT HIGHLIGHTS



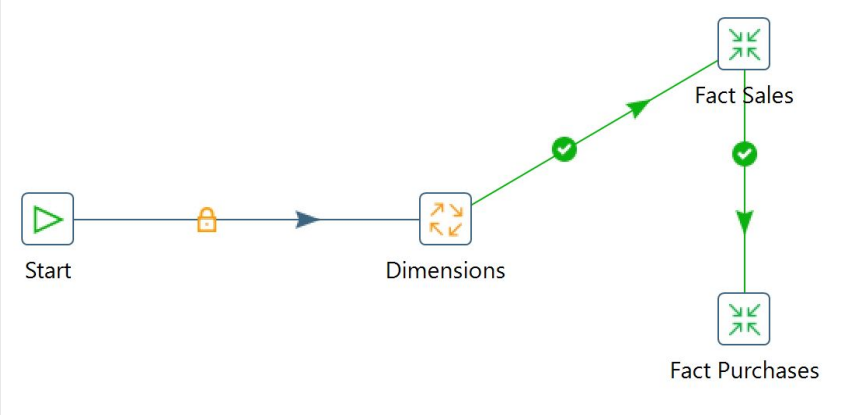
Product Standard Cost

JOBS

DIMENSIONS

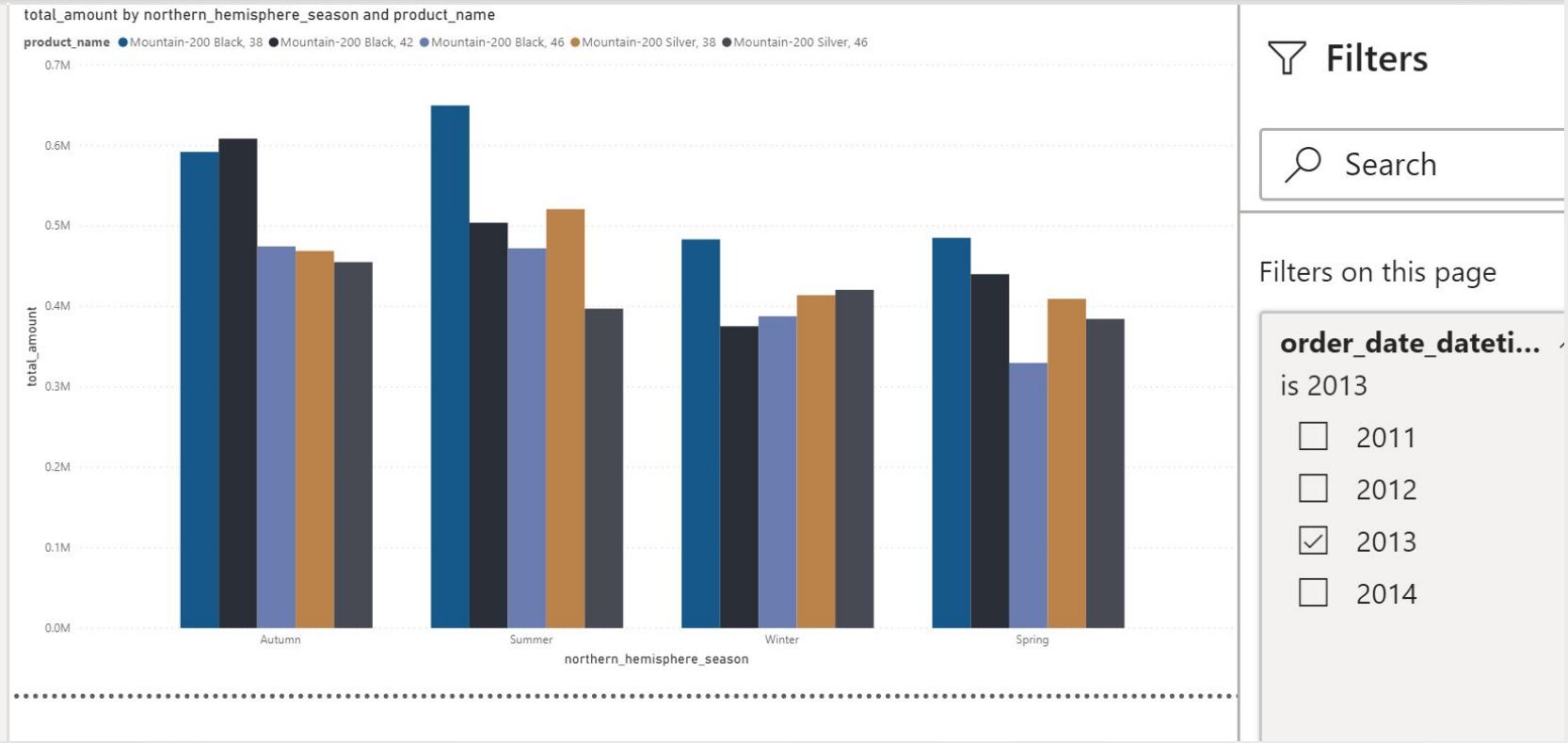


FACTS



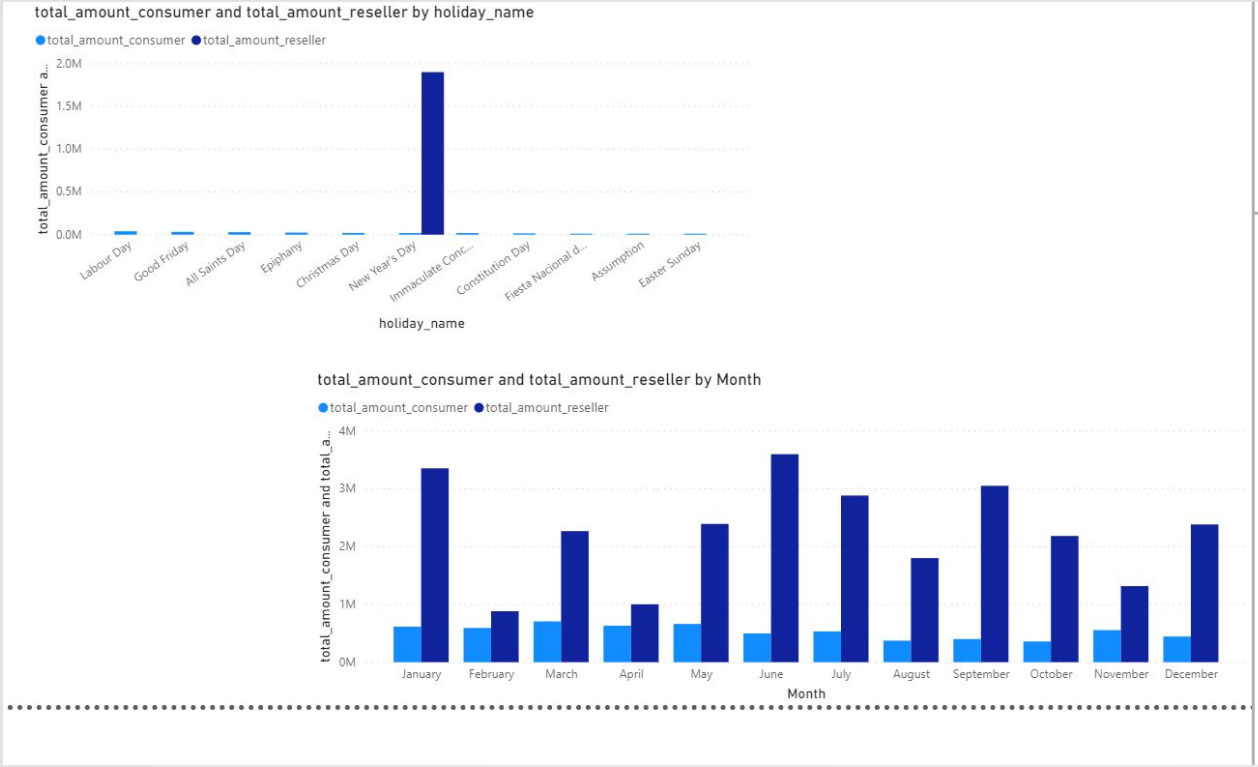
SEASONAL SALES

Create top 5 products sales ranking by seasons.



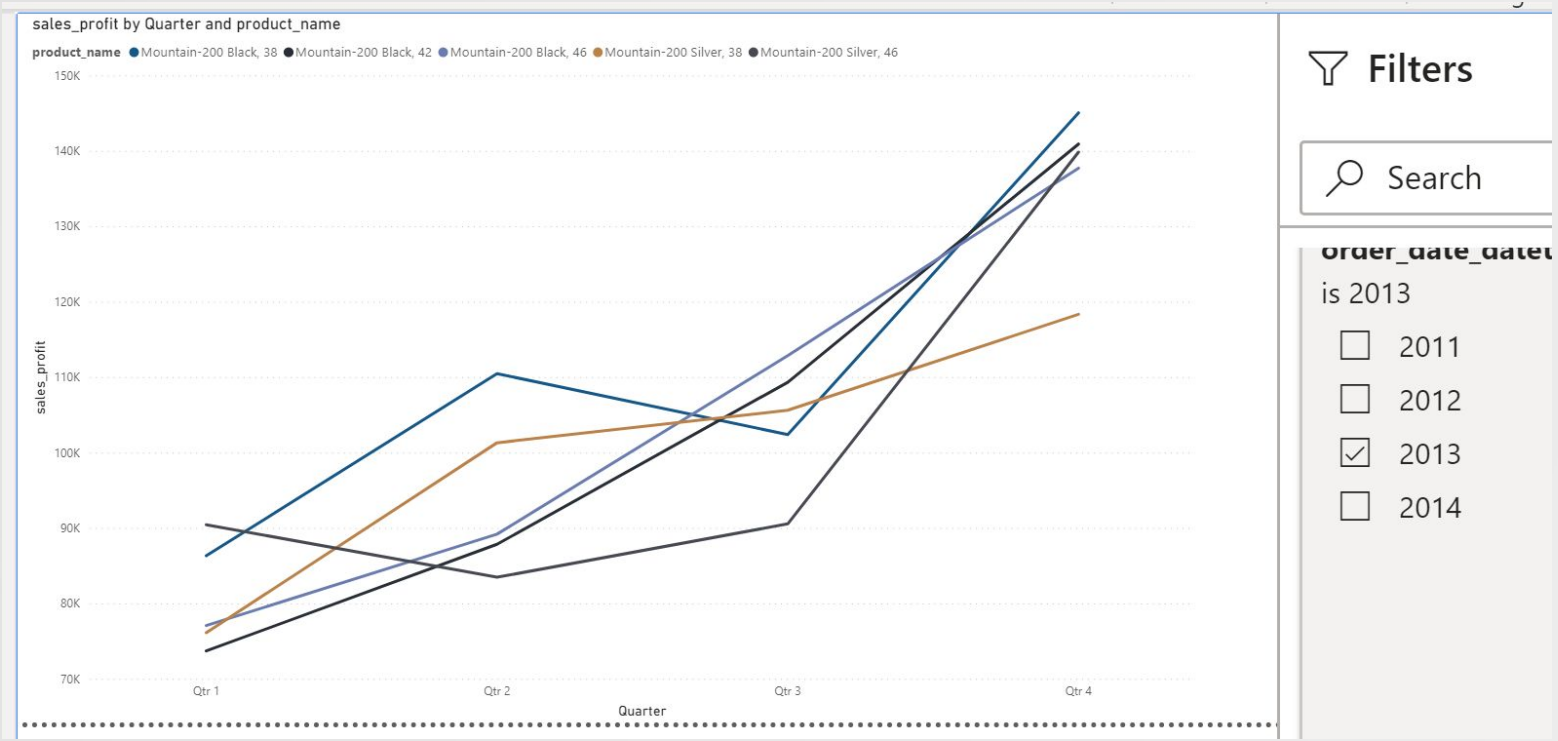
HOLIDAY SALES

Reseller sales vs direct consumer sales over specific holiday days.



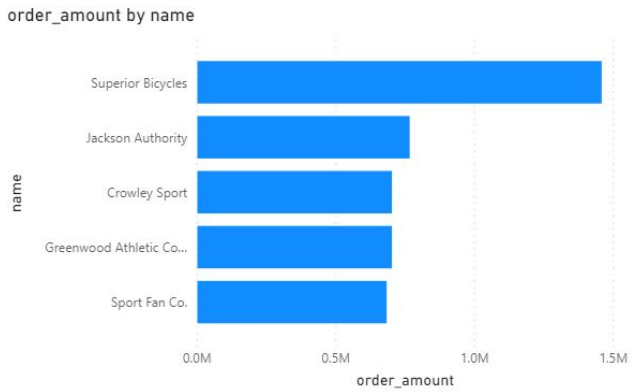
PRODUCT PROFITABILITY

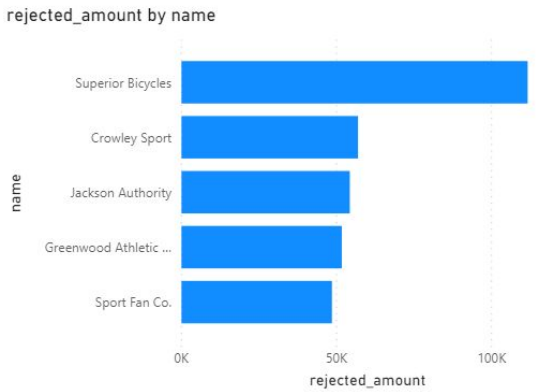
Create top 5 products profit ranking, quarter by quarter



REJECTED ITEMS

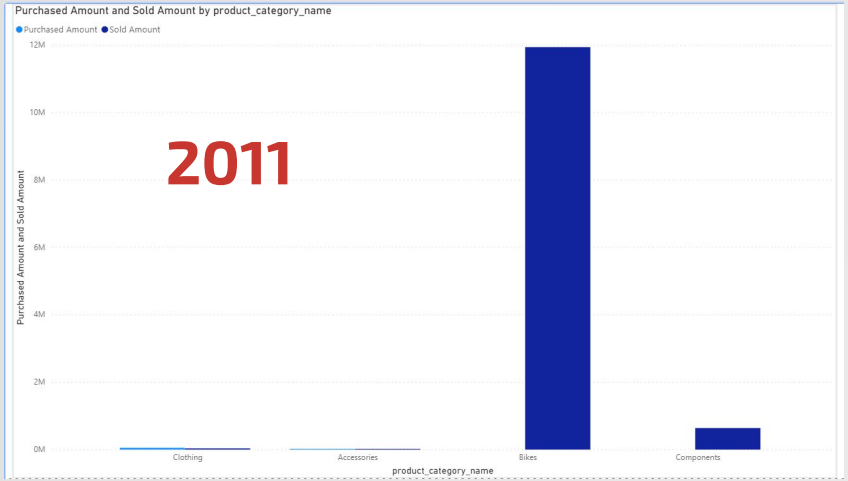
Create a top 5 Vendor ranking of rejected items on purchase orders.

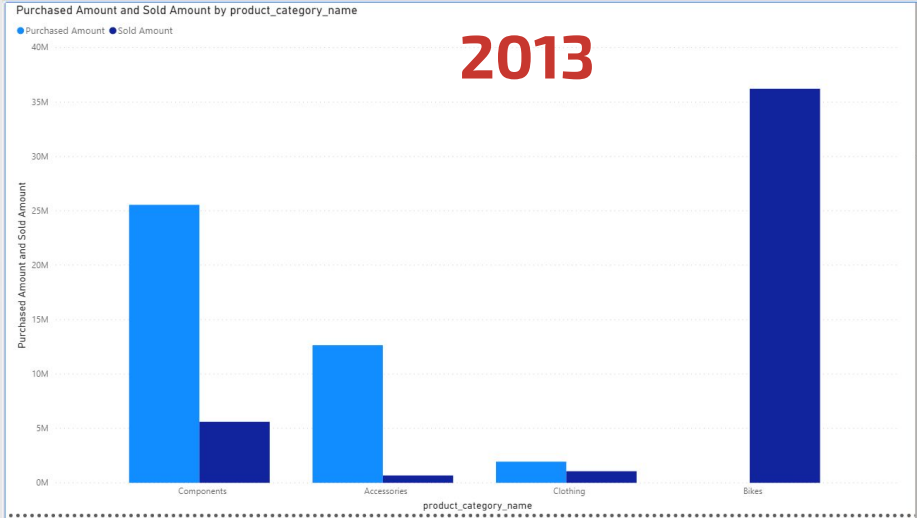




PURCHASED VS SOLD

Sold items vs purchased items,
for each product category






Thank You