Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

Макарова Екатерина Павловна

ЛАБОРАТОРНАЯ РАБОТА 2

Динамические соединения с базами данных

Проектный практикум по разработке ETL-решений

Направление подготовки

38.03.05 Бизнес-информатика

Профиль подготовки

Аналитика данных и эффективное управление

Курс обучения: 4

Форма обучения: очная

Преподаватель: кандидат технических наук,
доцент Босенко Тимур Муртазович

Москва
2025

Цель работы: получить практические навыки создания ETL-процесса для интеграции данных из различных источников с использованием динамических соединений в Pentaho Data Integration, включая обработку повторяющихся данных.

Задачи:

– Создать динамические подключения к различным источникам данных.

– Разработать процесс выявления и обработки дублирующихся записей.

– Реализовать механизм объединения данных в единое хранилище.

– Настроить обработку ошибок при выполнении трансформации.

Задачи по индивидуальному заданию вариант 8:

- Фильтр по доставке: только Standard Class;
- Статистика продаж;
- Анализ по городам;

Ход работы:

1. Подготовка базы данных:

1) Создание таблицы «orders» (Рис. 1).

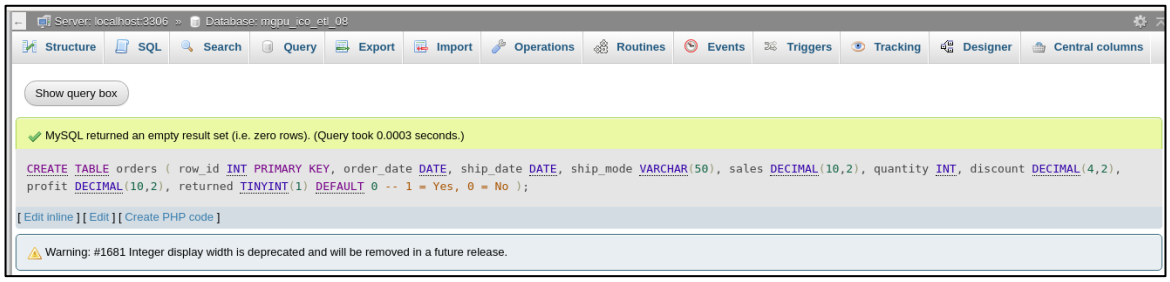


Рис. 1

2) Создание таблицы «customers» (Рис. 2).

Рис. 2

3) Создание таблицы «products» (Рис. 3).

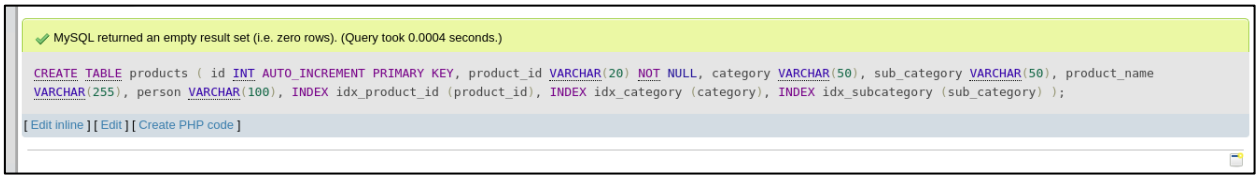


Рис. 3

4) Создание индексов для оптимизации запросов (Рис. 4).

Рис. 4

2. Проверка работоспособности трансформаций

1) Выполнение трансформации на обработку и загрузку данных по заказам в базу данных (Рис. 5).

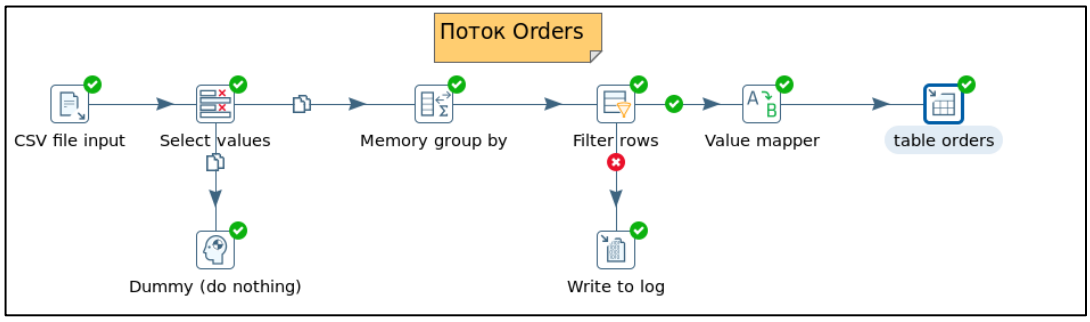


Рис. 5

Запрос к базе данных на проверку импорта данных в таблицу «orders» (Рис. 6).

Showing rows 0 - 24 (9994 total, Query took 0.0003 seconds.)

```
SELECT * FROM `orders`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

1 > >> | Number of rows: 25 Filter rows: Search this table Sort by key: None

Extra options

	row_id	order_date	ship_date	ship_mode	sales	quantity	discount	profit	returned
Edit Copy Delete	1	2018-11-08	2018-11-11	Second Class	261.96	2	0.00	41.91	NULL
Edit Copy Delete	2	2018-11-08	2018-11-11	Second Class	731.94	3	0.00	219.58	NULL
Edit Copy Delete	3	2018-06-12	2018-06-16	Second Class	14.62	2	0.00	6.87	NULL
Edit Copy Delete	4	2017-10-11	2017-10-18	Standard Class	957.58	5	0.40	-383.03	NULL
Edit Copy Delete	5	2017-10-11	2017-10-18	Standard Class	22.37	2	0.20	2.52	NULL
Edit Copy Delete	6	2016-06-09	2016-06-14	Standard Class	48.86	7	0.00	14.17	NULL
Edit Copy Delete	7	2016-06-09	2016-06-14	Standard Class	7.28	4	0.00	1.97	NULL
Edit Copy Delete	8	2016-06-09	2016-06-14	Standard Class	907.15	6	0.20	90.72	NULL
Edit Copy Delete	9	2016-06-09	2016-06-14	Standard Class	18.50	3	0.20	5.78	NULL
Edit Copy Delete	10	2016-06-09	2016-06-14	Standard Class	114.90	5	0.00	34.47	NULL
Edit Copy Delete	11	2016-06-09	2016-06-14	Standard Class	1706.18	9	0.20	85.31	NULL
Edit Copy Delete	12	2016-06-09	2016-06-14	Standard Class	911.42	4	0.20	68.36	NULL
Edit Copy Delete	13	2019-04-15	2019-04-20	Standard Class	15.55	3	0.20	5.44	NULL

Рис. 6

2) Выполнение трансформации на обработку и загрузку данных по заказам в базу данных (Рис. 7).

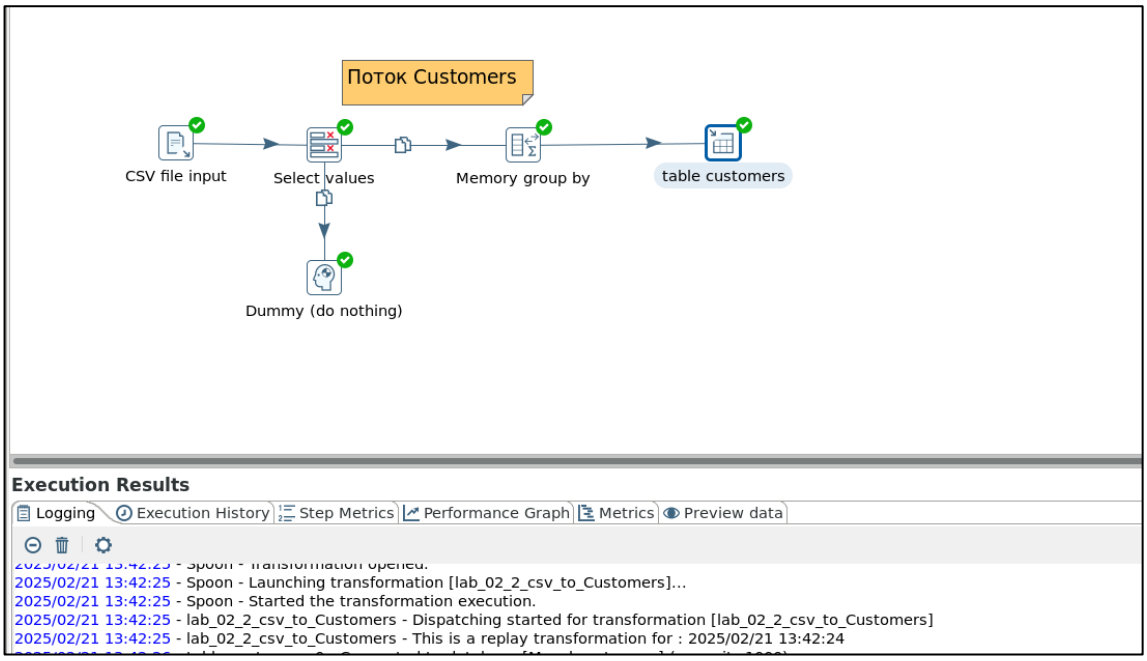
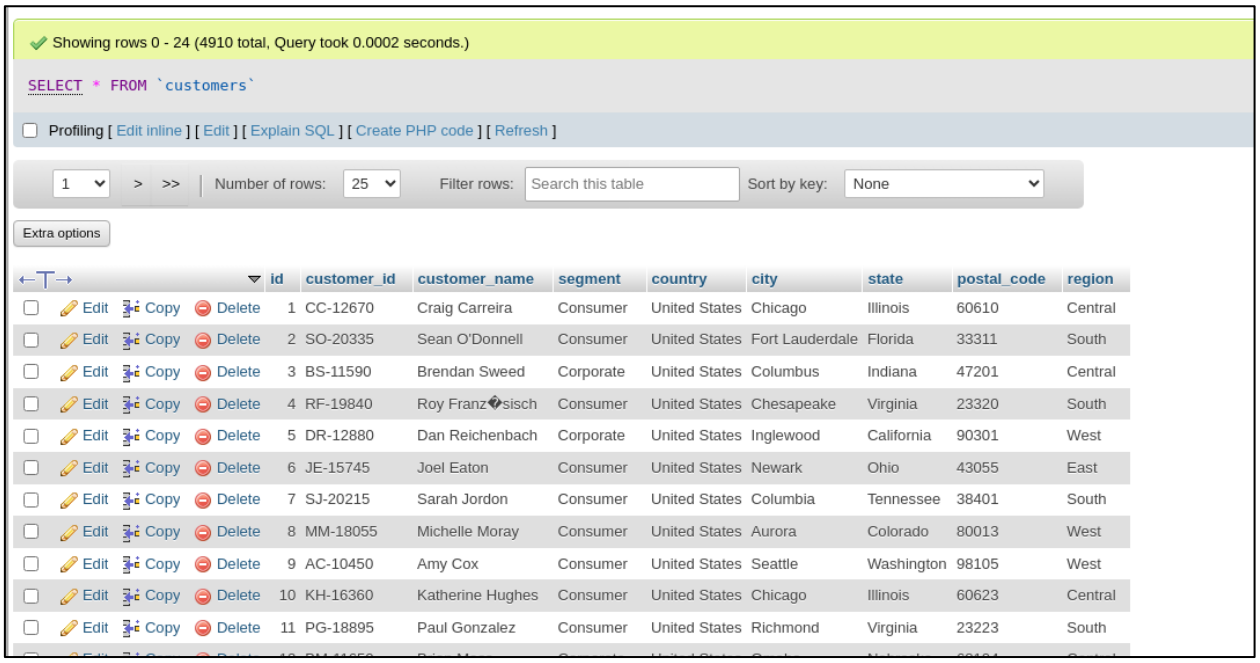


Рис. 7

Выполнение запроса к базе данных на импорт данных о клиентах (Рис. 8).



id	customer_id	customer_name	segment	country	city	state	postal_code	region
1	CC-12670	Craig Carreira	Consumer	United States	Chicago	Illinois	60610	Central
2	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South
3	BS-11590	Brendan Sweed	Corporate	United States	Columbus	Indiana	47201	Central
4	RF-19840	Roy Franz&#65533;sisch	Consumer	United States	Chesapeake	Virginia	23320	South
5	DR-12880	Dan Reichenbach	Corporate	United States	Inglewood	California	90301	West
6	JE-15745	Joel Eaton	Consumer	United States	Newark	Ohio	43055	East
7	SJ-20215	Sarah Jordon	Consumer	United States	Columbia	Tennessee	38401	South
8	MM-18055	Michelle Moray	Consumer	United States	Aurora	Colorado	80013	West
9	AC-10450	Amy Cox	Consumer	United States	Seattle	Washington	98105	West
10	KH-16360	Katherine Hughes	Consumer	United States	Chicago	Illinois	60623	Central
11	PG-18895	Paul Gonzalez	Consumer	United States	Richmond	Virginia	23223	South

Рис. 8

3) Выполнение трансформации на обработку и загрузку данных по продуктам в базу данных (Рис. 9).

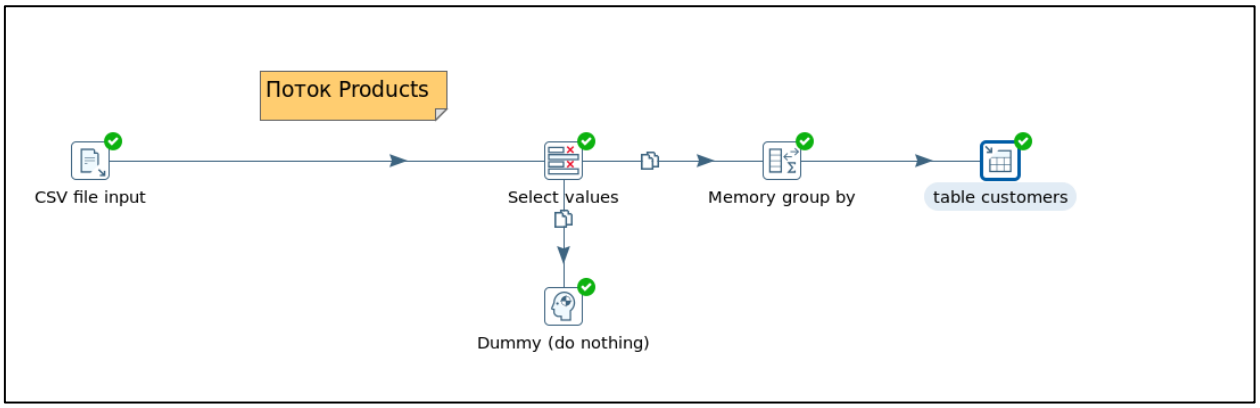
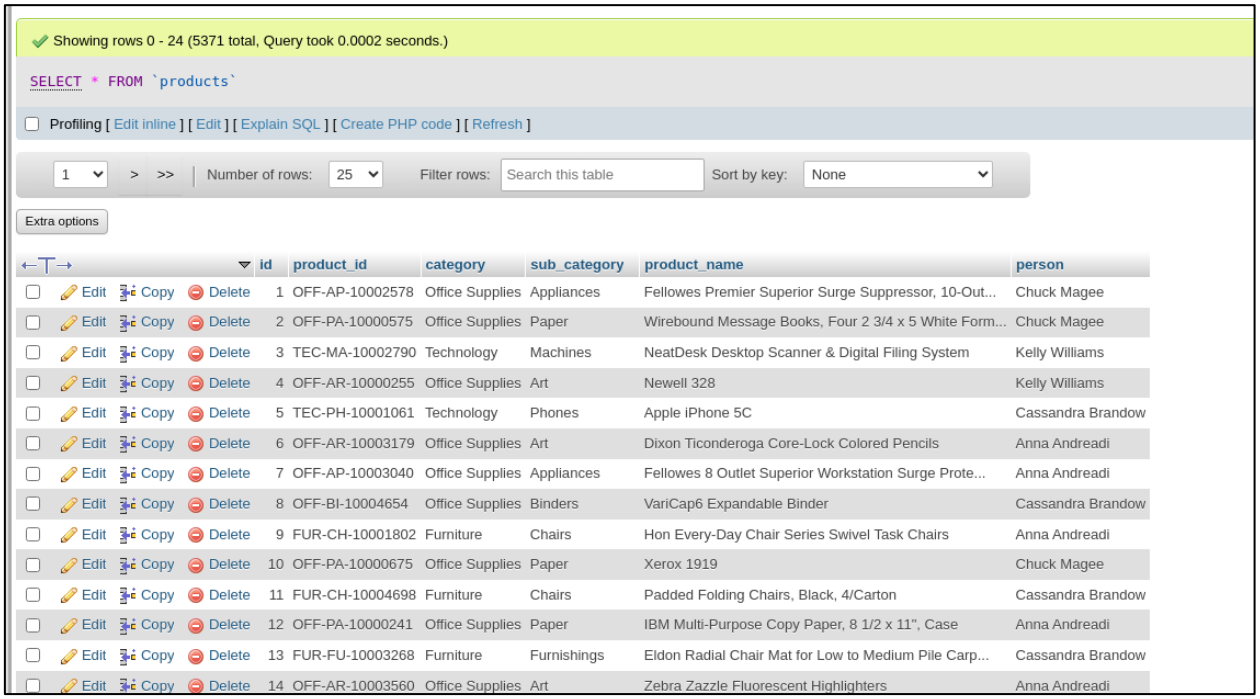


Рис. 9

Выполнение запроса на проверку импорта данных в таблицу «products» (Рис. 10).



	id	product_id	category	sub_category	product_name	person
Edit Copy Delete	1	OFF-AP-10002578	Office Supplies	Appliances	Fellowes Premier Superior Surge Suppressor, 10-Out...	Chuck Magee
Edit Copy Delete	2	OFF-PA-10000575	Office Supplies	Paper	Wirebound Message Books, Four 2 3/4 x 5 White Form...	Chuck Magee
Edit Copy Delete	3	TEC-MA-10002790	Technology	Machines	NeatDesk Desktop Scanner & Digital Filing System	Kelly Williams
Edit Copy Delete	4	OFF-AR-10000255	Office Supplies	Art	Newell 328	Kelly Williams
Edit Copy Delete	5	TEC-PH-10001061	Technology	Phones	Apple iPhone 5C	Cassandra Brandow
Edit Copy Delete	6	OFF-AR-10003179	Office Supplies	Art	Dixon Ticonderoga Core-Lock Colored Pencils	Anna Andreadi
Edit Copy Delete	7	OFF-AP-10003040	Office Supplies	Appliances	Fellowes 8 Outlet Superior Workstation Surge Prote...	Anna Andreadi
Edit Copy Delete	8	OFF-BI-10004654	Office Supplies	Binders	VariCap6 Expandable Binder	Cassandra Brandow
Edit Copy Delete	9	FUR-CH-10001802	Furniture	Chairs	Hon Every-Day Chair Series Swivel Task Chairs	Anna Andreadi
Edit Copy Delete	10	OFF-PA-10000675	Office Supplies	Paper	Xerox 1919	Chuck Magee
Edit Copy Delete	11	FUR-CH-10004698	Furniture	Chairs	Padded Folding Chairs, Black, 4/Carton	Cassandra Brandow
Edit Copy Delete	12	OFF-PA-10000241	Office Supplies	Paper	IBM Multi-Purpose Copy Paper, 8 1/2 x 11", Case	Anna Andreadi
Edit Copy Delete	13	FUR-FU-10003268	Furniture	Furnishings	Eldon Radial Chair Mat for Low to Medium Pile Carp...	Cassandra Brandow
Edit Copy Delete	14	OFF-AR-10003560	Office Supplies	Art	Zebra Zazzle Fluorescent Highlighters	Anna Andreadi

Рис. 10

После этого таблицы были очищены с помощью запроса TRUNCATE TABLE “название таблицы” (orders, customers, products).

4) Проверка работоспособности Job на импорт данных из трансформаций (Рис. 11).

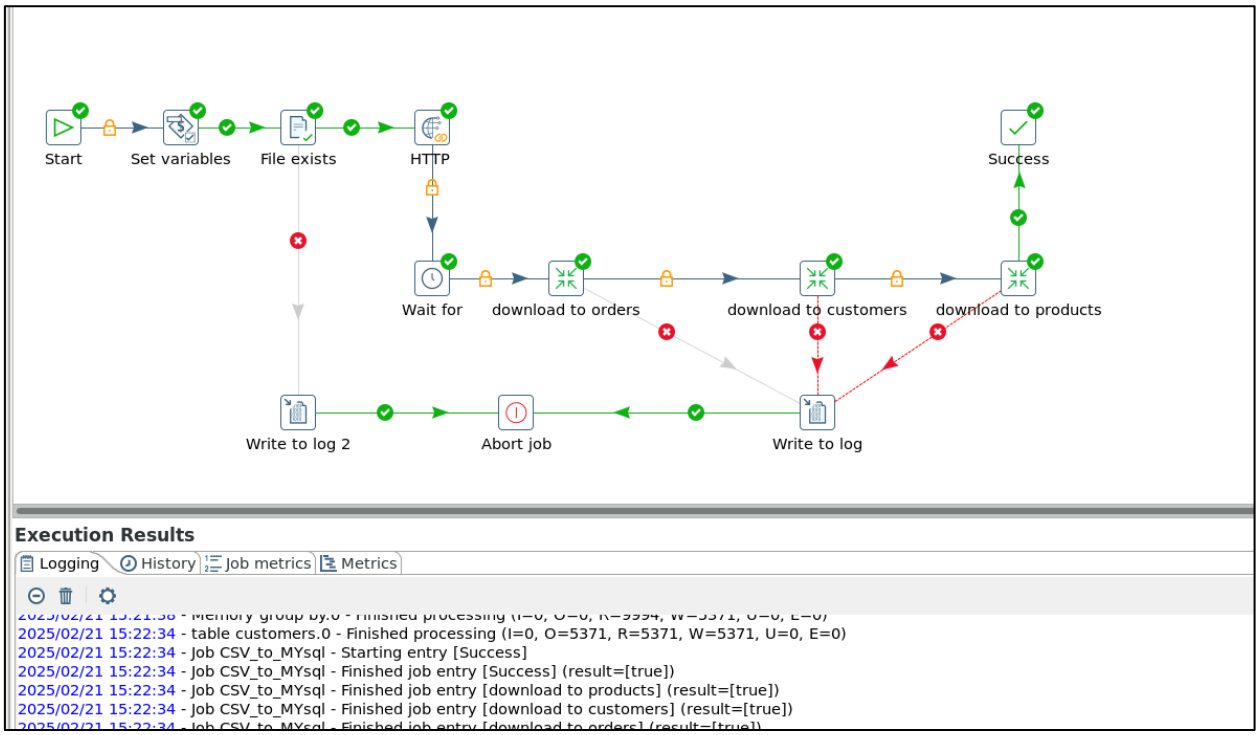
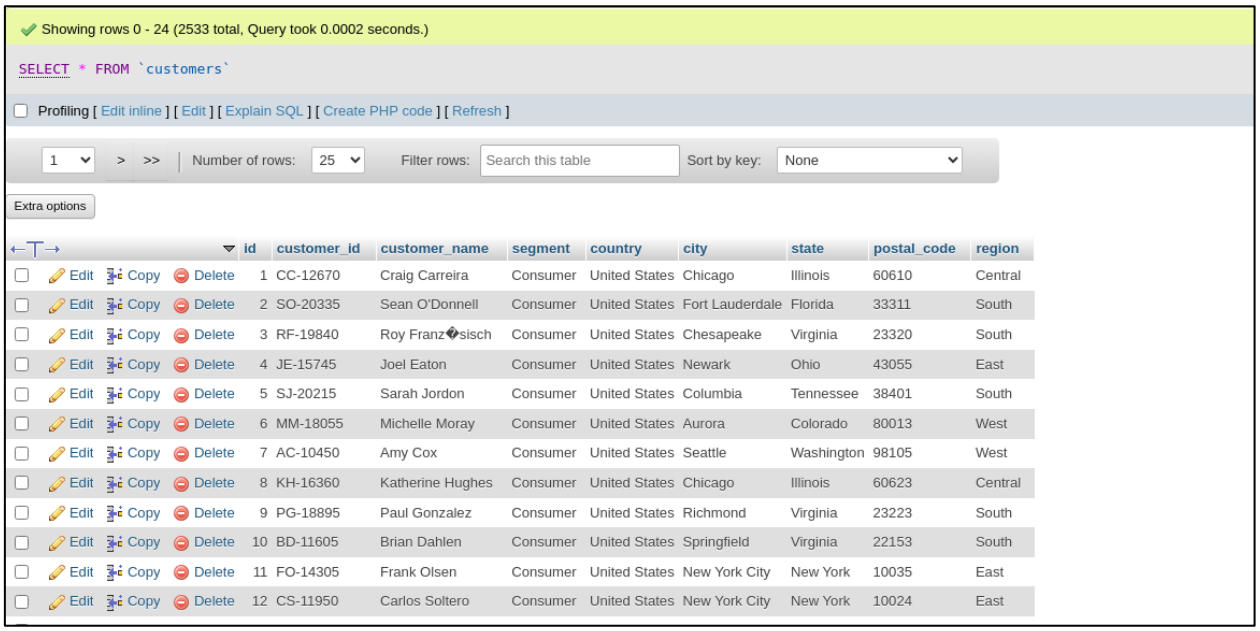


Рис. 11

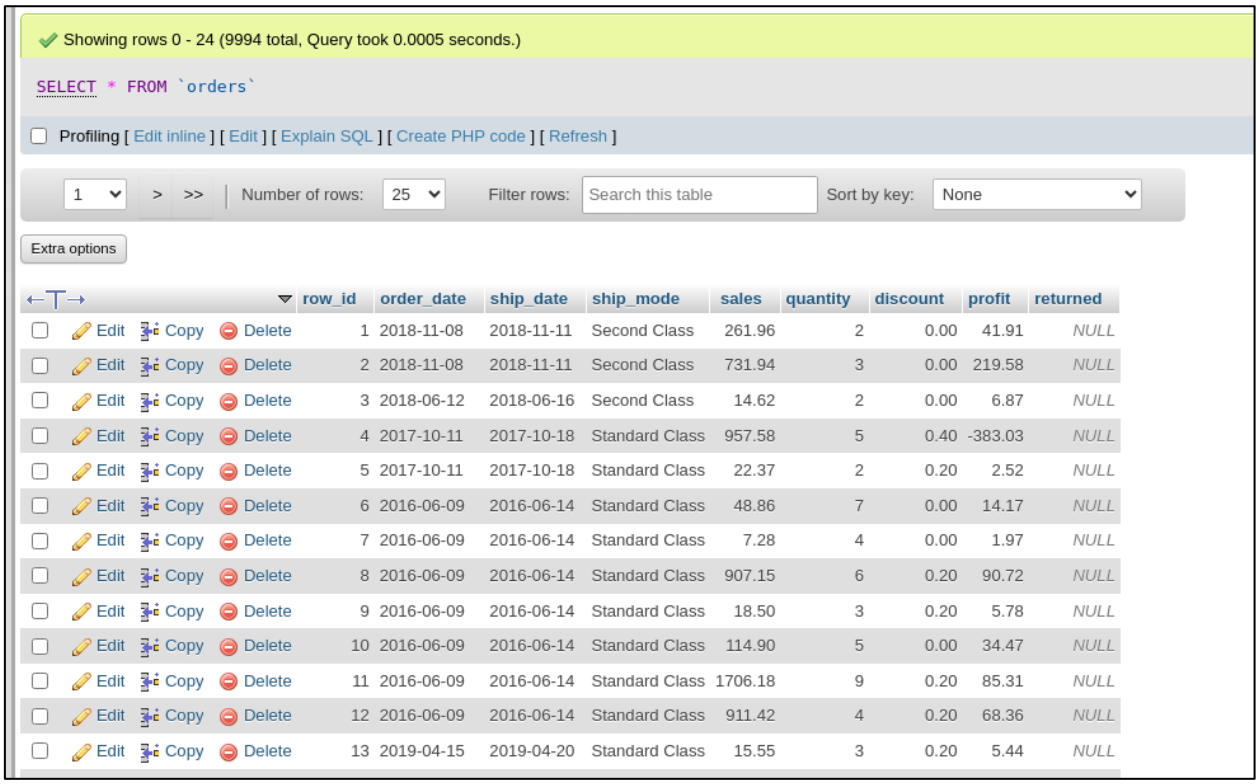
Запрос к таблице «customers» для проверки успешности импорта данных (Рис. 12).



	id	customer_id	customer_name	segment	country	city	state	postal_code	region
Edit Copy Delete	1	CC-12670	Craig Carreira	Consumer	United States	Chicago	Illinois	60610	Central
Edit Copy Delete	2	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South
Edit Copy Delete	3	RF-19840	Roy Franz&#65533;sisch	Consumer	United States	Chesapeake	Virginia	23320	South
Edit Copy Delete	4	JE-15745	Joel Eaton	Consumer	United States	Newark	Ohio	43055	East
Edit Copy Delete	5	SJ-20215	Sarah Jordon	Consumer	United States	Columbia	Tennessee	38401	South
Edit Copy Delete	6	MM-18055	Michelle Moray	Consumer	United States	Aurora	Colorado	80013	West
Edit Copy Delete	7	AC-10450	Amy Cox	Consumer	United States	Seattle	Washington	98105	West
Edit Copy Delete	8	KH-16360	Katherine Hughes	Consumer	United States	Chicago	Illinois	60623	Central
Edit Copy Delete	9	PG-18895	Paul Gonzalez	Consumer	United States	Richmond	Virginia	23223	South
Edit Copy Delete	10	BD-11605	Brian Dahlen	Consumer	United States	Springfield	Virginia	22153	South
Edit Copy Delete	11	FO-14305	Frank Olsen	Consumer	United States	New York City	New York	10035	East
Edit Copy Delete	12	CS-11950	Carlos Soltero	Consumer	United States	New York City	New York	10024	East

Рис. 12

Запрос к таблице «orders» для проверки успешности импорта данных (Рис. 13).



Showing rows 0 - 24 (9994 total, Query took 0.0005 seconds.)

```
SELECT * FROM `orders`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

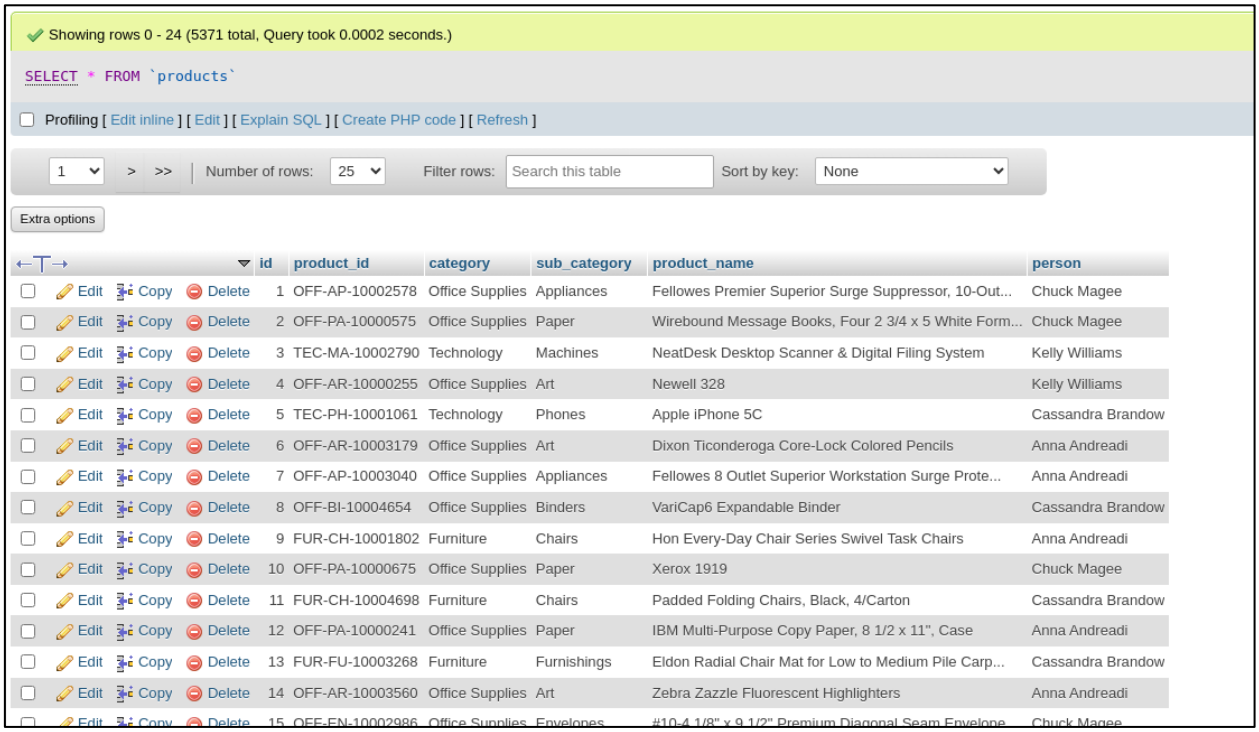
1 > >> | Number of rows: 25 Filter rows: Search this table Sort by key: None

Extra options

	row_id	order_date	ship_date	ship_mode	sales	quantity	discount	profit	returned
Edit Copy Delete	1	2018-11-08	2018-11-11	Second Class	261.96	2	0.00	41.91	NULL
Edit Copy Delete	2	2018-11-08	2018-11-11	Second Class	731.94	3	0.00	219.58	NULL
Edit Copy Delete	3	2018-06-12	2018-06-16	Second Class	14.62	2	0.00	6.87	NULL
Edit Copy Delete	4	2017-10-11	2017-10-18	Standard Class	957.58	5	0.40	-383.03	NULL
Edit Copy Delete	5	2017-10-11	2017-10-18	Standard Class	22.37	2	0.20	2.52	NULL
Edit Copy Delete	6	2016-06-09	2016-06-14	Standard Class	48.86	7	0.00	14.17	NULL
Edit Copy Delete	7	2016-06-09	2016-06-14	Standard Class	7.28	4	0.00	1.97	NULL
Edit Copy Delete	8	2016-06-09	2016-06-14	Standard Class	907.15	6	0.20	90.72	NULL
Edit Copy Delete	9	2016-06-09	2016-06-14	Standard Class	18.50	3	0.20	5.78	NULL
Edit Copy Delete	10	2016-06-09	2016-06-14	Standard Class	114.90	5	0.00	34.47	NULL
Edit Copy Delete	11	2016-06-09	2016-06-14	Standard Class	1706.18	9	0.20	85.31	NULL
Edit Copy Delete	12	2016-06-09	2016-06-14	Standard Class	911.42	4	0.20	68.36	NULL
Edit Copy Delete	13	2019-04-15	2019-04-20	Standard Class	15.55	3	0.20	5.44	NULL

Рис. 13

Запрос к таблице «products» для проверки успешности импорта данных (Рис. 14).



Showing rows 0 - 24 (5371 total, Query took 0.0002 seconds.)

```
SELECT * FROM `products`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

1 > >> | Number of rows: 25 Filter rows: Search this table Sort by key: None

Extra options

	id	product_id	category	sub_category	product_name	person
Edit Copy Delete	1	OFF-AP-10002578	Office Supplies	Appliances	Fellowes Premier Superior Surge Suppressor, 10-Out...	Chuck Magee
Edit Copy Delete	2	OFF-PA-10000575	Office Supplies	Paper	Wirebound Message Books, Four 2 3/4 x 5 White Form...	Chuck Magee
Edit Copy Delete	3	TEC-MA-10002790	Technology	Machines	NeatDesk Desktop Scanner & Digital Filing System	Kelly Williams
Edit Copy Delete	4	OFF-AR-10000255	Office Supplies	Art	Newell 328	Kelly Williams
Edit Copy Delete	5	TEC-PH-10001061	Technology	Phones	Apple iPhone 5C	Cassandra Brandow
Edit Copy Delete	6	OFF-AR-10003179	Office Supplies	Art	Dixon Ticonderoga Core-Lock Colored Pencils	Anna Andreadi
Edit Copy Delete	7	OFF-AP-10003040	Office Supplies	Appliances	Fellowes 8 Outlet Superior Workstation Surge Prote...	Anna Andreadi
Edit Copy Delete	8	OFF-BI-10004654	Office Supplies	Binders	VariCap6 Expandable Binder	Cassandra Brandow
Edit Copy Delete	9	FUR-CH-10001802	Furniture	Chairs	Hon Every-Day Chair Series Swivel Task Chairs	Anna Andreadi
Edit Copy Delete	10	OFF-PA-10000675	Office Supplies	Paper	Xerox 1919	Chuck Magee
Edit Copy Delete	11	FUR-CH-10004698	Furniture	Chairs	Padded Folding Chairs, Black, 4/Carton	Cassandra Brandow
Edit Copy Delete	12	OFF-PA-10000241	Office Supplies	Paper	IBM Multi-Purpose Copy Paper, 8 1/2 x 11", Case	Anna Andreadi
Edit Copy Delete	13	FUR-FU-10003268	Furniture	Furnishings	Eldon Radial Chair Mat for Low to Medium Pile Carp...	Cassandra Brandow
Edit Copy Delete	14	OFF-AR-10003560	Office Supplies	Art	Zebra Zazzle Fluorescent Highlighters	Anna Andreadi
Edit Copy Delete	15	OFF-EN-10002986	Office Supplies	Envelopes	#10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelope	Chuck Magee

Рис. 14

Выполнение индивидуального задания:

1. Фильтр по доставке: только Standard Class.

Атрибут «ship_mode» относится к таблице orders, поэтому был добавлен фильтр к трансформации «orders». Настройка компонента «filter rows» (Рис. 15).

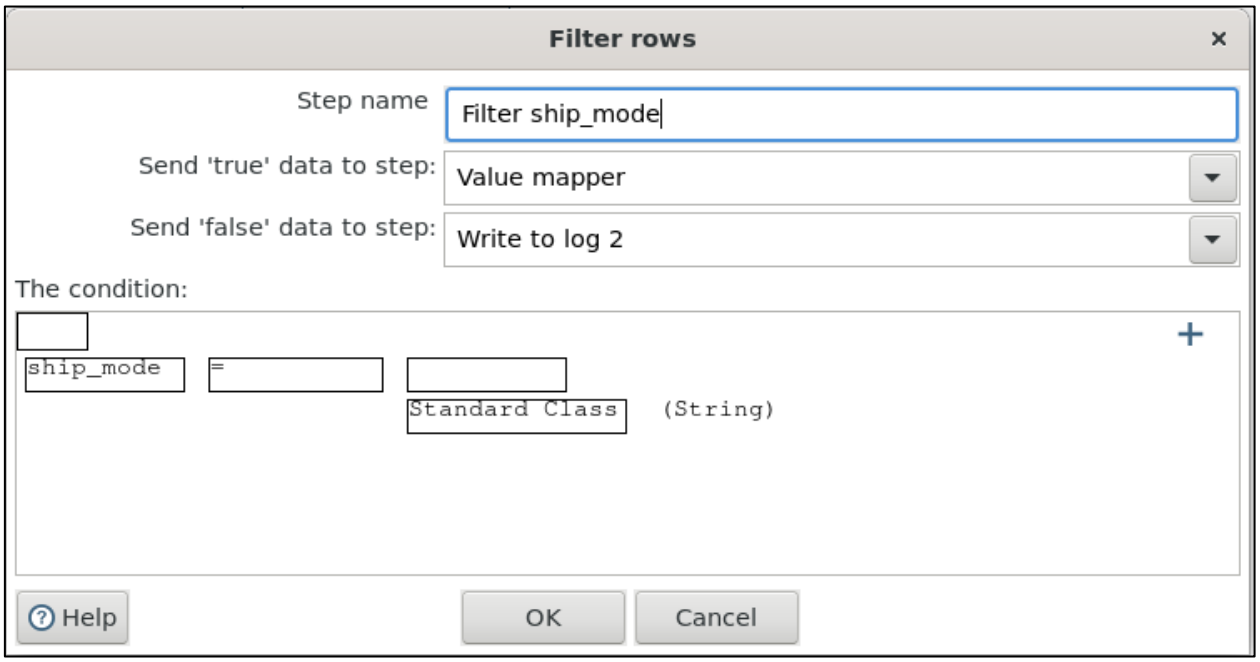


Рис. 15

Добавлен компонент для записи логов, если компонент с фильтрацией будет выполнен с ошибкой (Рис. 16).

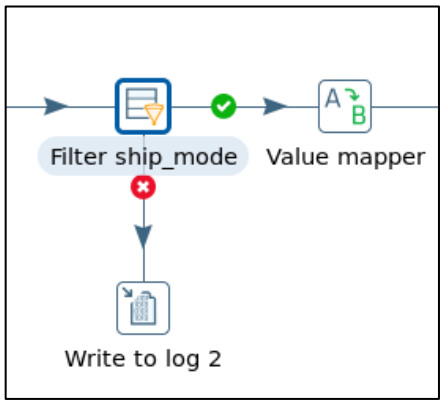


Рис. 16

Общий вид трансформации «orders» (Рис. 17).

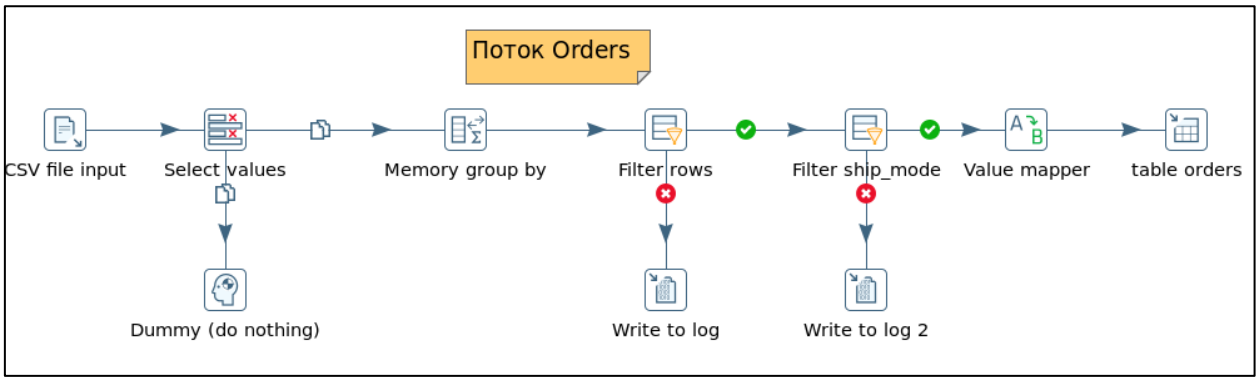


Рис. 17

Была запущена Job, после её выполнения был выполнен SQL запрос на просмотр уникальных значений в таблице «orders» по атрибуту «ship_mode» (Рис. 18).

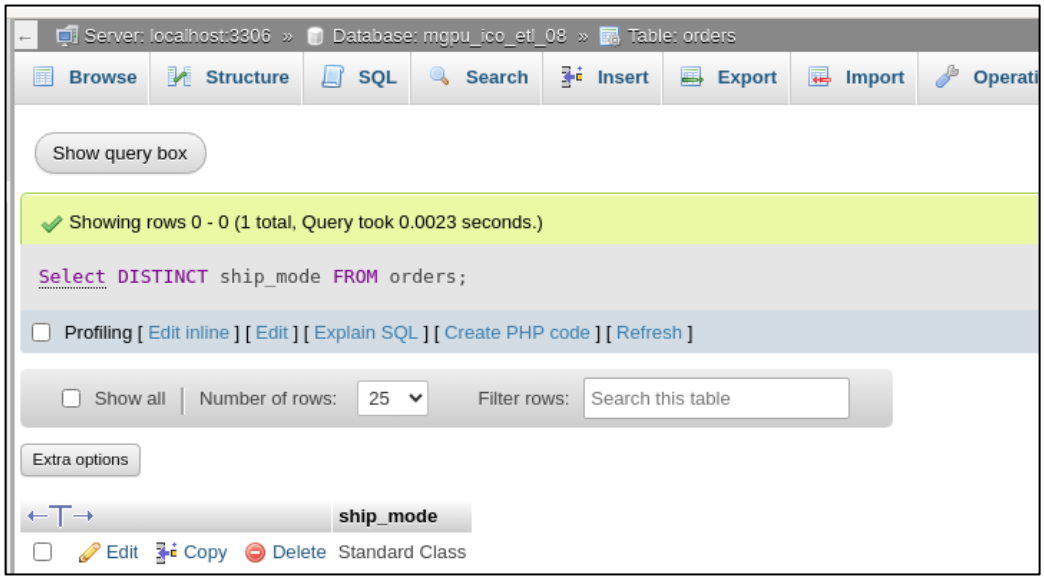


Рис. 18

Таким образом, данные были успешно отфильтрованы по ship_mode.

2. Статистика продаж:

1) Добавления компонента «CSV file input» для импорта CSV файла (Рис. 19).

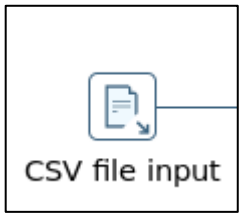


Рис. 19

Настройка компонента (Рис. 20).

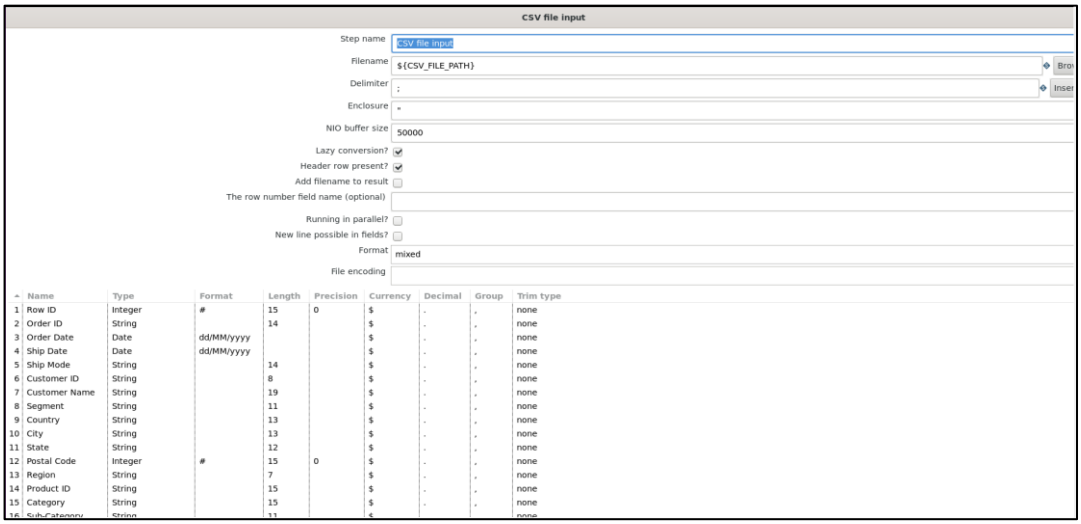


Рис. 20

2) Добавление компонента «Replace in string» (Рис. 21).

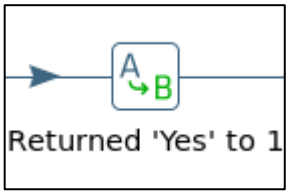


Рис. 21

Настройка компонента: для того, чтобы в дальнейшем проанализировать количество возвратов, необходимо преобразовать значения атрибута «Returned» в числовой тип данных, для этого были заменены значения «Yes» в 1 (Рис. 22).

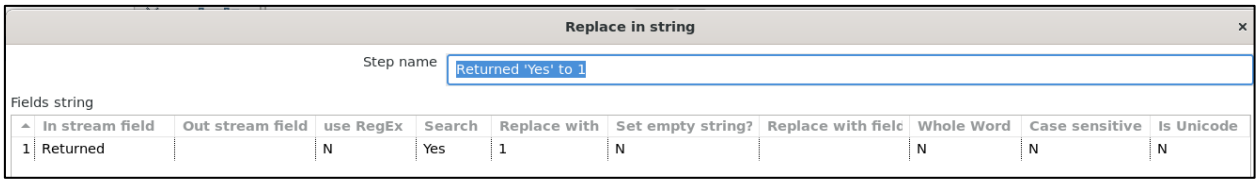


Рис. 22

3) Добавление компонента «Select values» (Рис. 23).

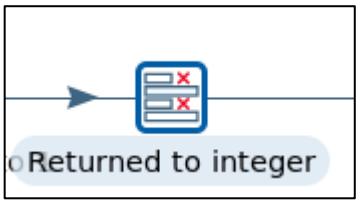


Рис. 23

Настройка компонента: преобразование атрибута «Returned» в тип данных integer (Рис. 24).

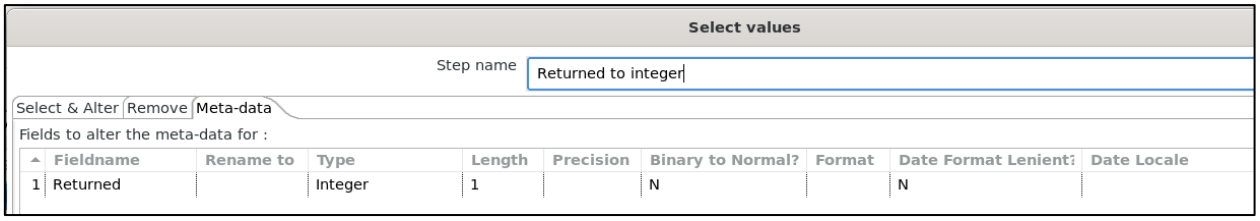


Рис. 24

4) Добавления компонента «In field value is null» (Рис. 25).

Рис. 25

Настройка компонента: если значение в атрибуте «Returned» пропущено, то заменять его на 0 (Рис. 26).

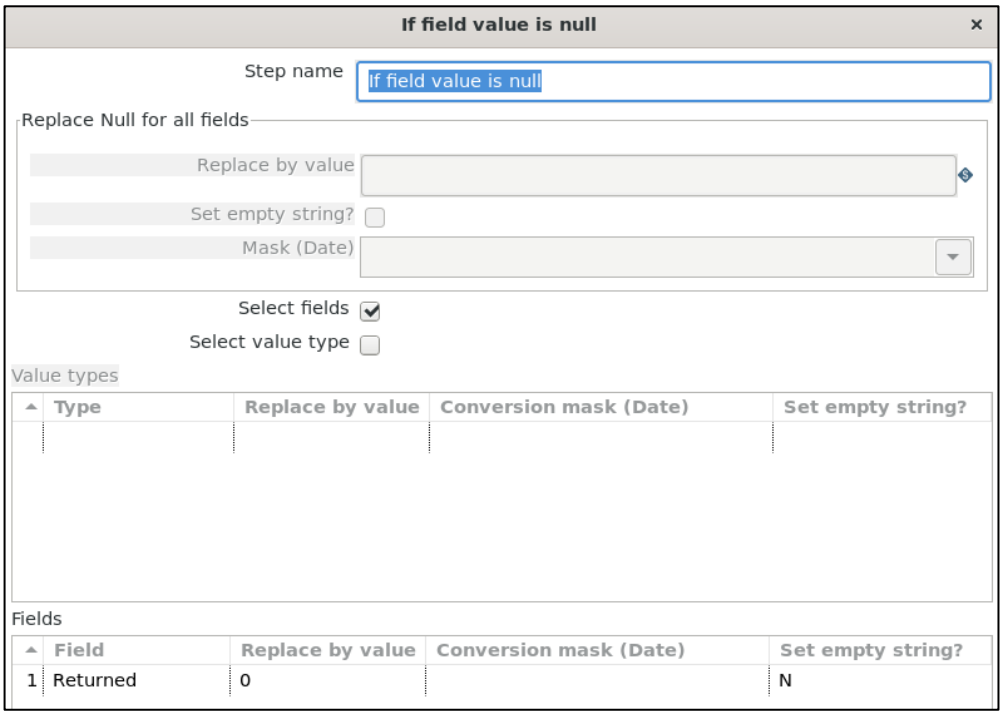


Рис. 26

5) Добавление компонента «Calculator» и запись логах при ошибке (Рис. 27).

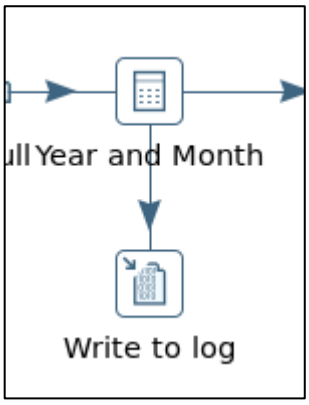
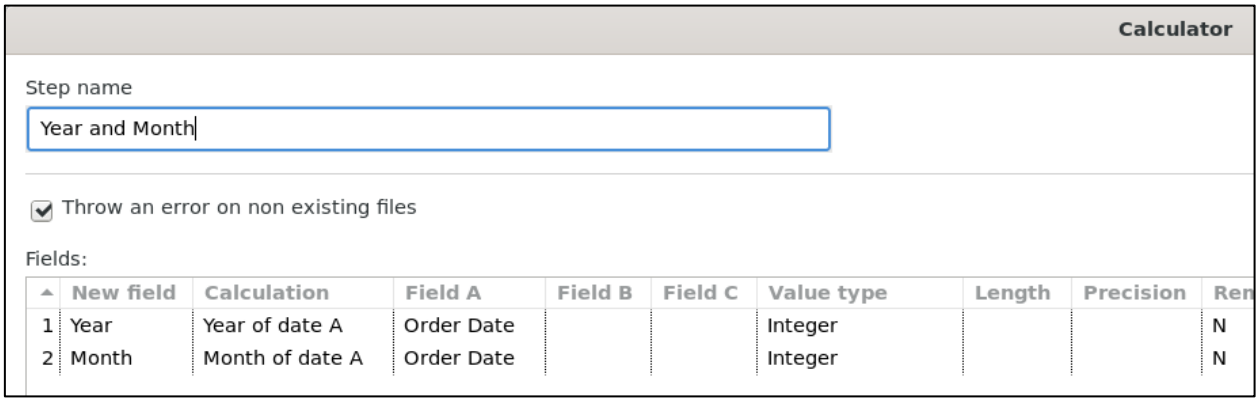


Рис. 27

Настройка компонента: из атрибута «Order Date» получаем два новых атрибута «Year» и «Month» типа данных integer, которые будут использованы при анализе по годам и месяцам (Рис. 28).



	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Ren
1	Year	Year of date A	Order Date			Integer			N
2	Month	Month of date A	Order Date			Integer			N

Рис. 28

6) Добавление компонента «Group by» (Рис. 29).

Рис. 29

Настройка компонента: группировка данных по атрибутам регион, категория, год, месяц, возврат (Рис. 30). Агрегирование данных:

- сумма по продажам;
- сумма по выручке;
- количество заказов;
- средняя стоимости скидки;

- сумма по количеству возвратов.

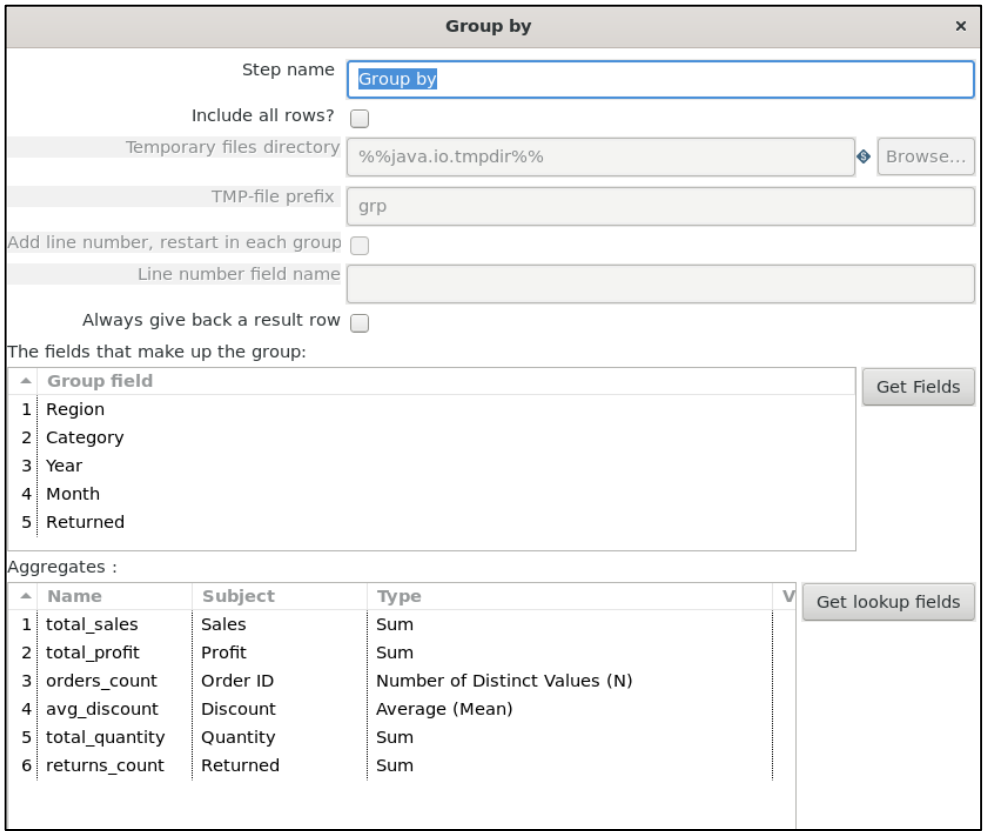


Рис. 30

7) Добавление компонента «Калькулятор» для расчёта новых метрик и логирование ошибок (Рис. 31).

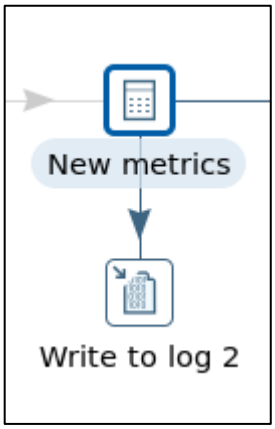


Рис. 31

Настройка компонента: расчёт новых метрик с помощью формул встроенных в калькулятор (Рис. 32):

- средний чек;

-процент возвратов;

- рентабельность.

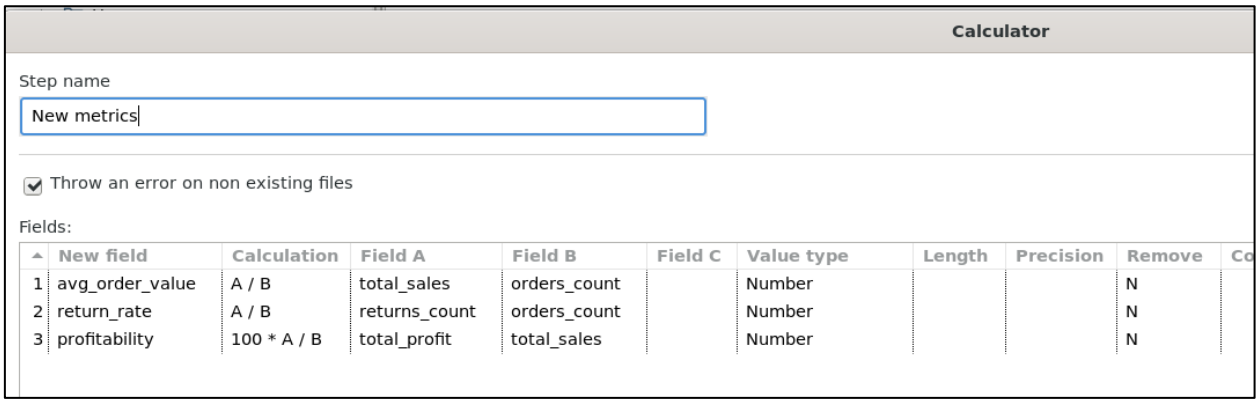


Рис. 32

8) Добавление компонента «Table output» (Рис. 33).

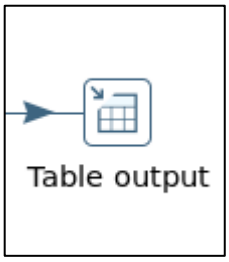


Рис. 33

Настройка компонента: подключение к БД для записи данных в таблицу «sales_statistics» (Рис. 34).

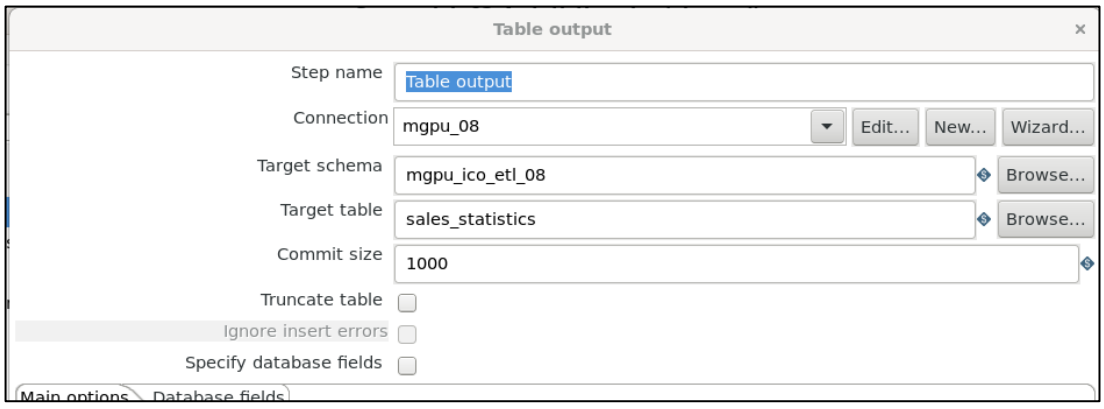


Рис. 34

Общий вид трансформации (Рис. 35).

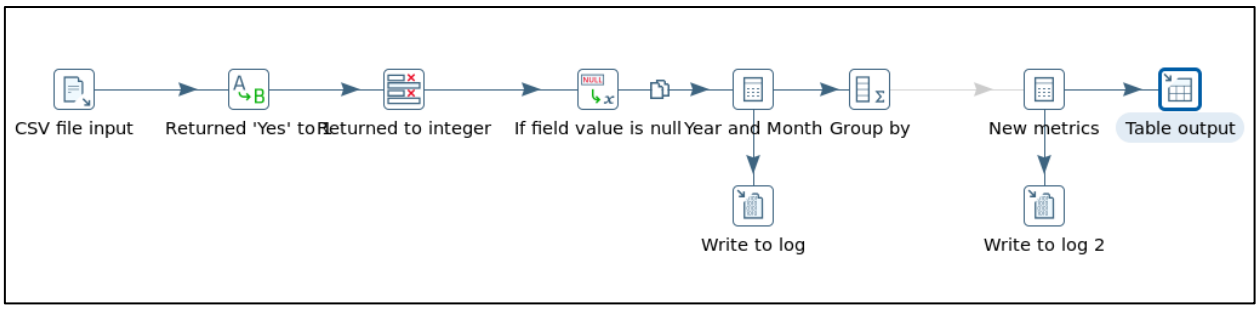


Рис. 35

9) Добавление компонента трансформации для статистики продаж в Job (Рис. 36).

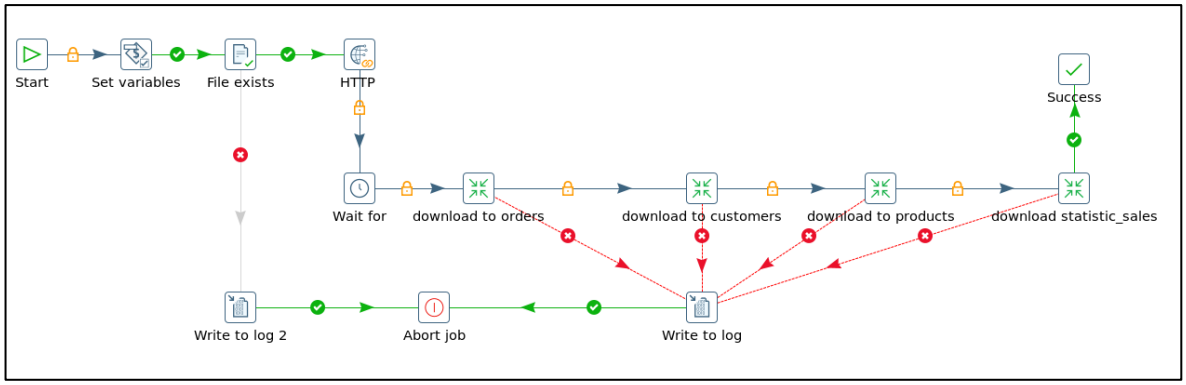


Рис. 36

10) Создание таблицы «sales_statistics» для записи данных в БД (Рис. 37).

```sql
CREATE TABLE sales_statistics (
    id INT AUTO_INCREMENT PRIMARY KEY,
    Region VARCHAR(50),
    Category VARCHAR(50),
    `Year` INT,
    Month INT,
    Returned DECIMAL (5,2),
    total_sales DECIMAL(12,2),
    total_profit DECIMAL(12,2),
    orders_count INT,
    avg_discount DECIMAL(5,2),
    return_rate DECIMAL(5,2),
    profitability DECIMAL(5,2)
);
```

Рис. 37

Создание индексов (Рис. 38).

```sql
CREATE INDEX idx_region ON sales_statistics (`Region`);
CREATE INDEX idx_category ON sales_statistics (`Category`);
CREATE INDEX idx_year_month ON sales_statistics (`Year`, `Month`);
```

Рис. 38

11) Запросы к базе данных для проверки импорта данных:

- Топ 3 региона по сумме продаж (Рис. 39).

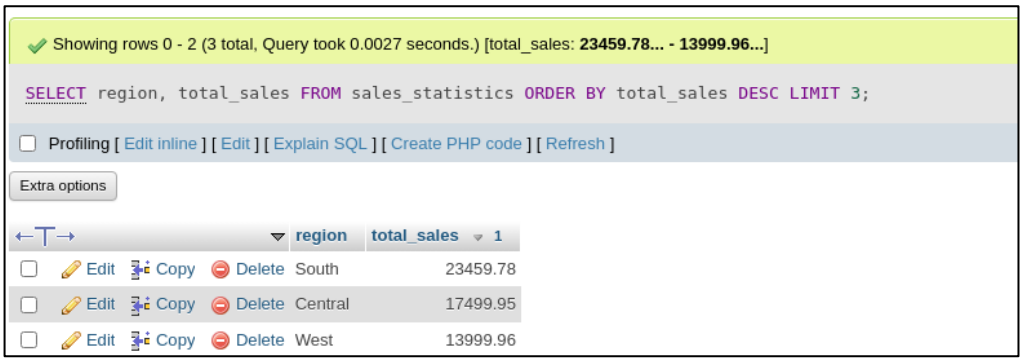


Рис. 39

- Сумма продаж по годам и месяцам (Рис. 40).

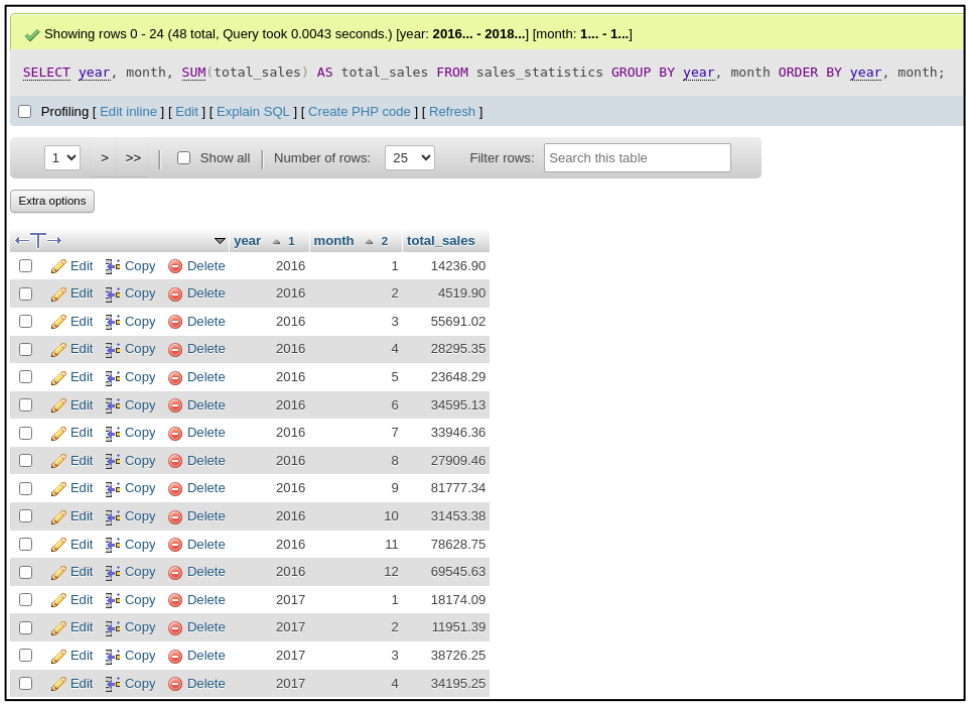


Рис. 40

- Вывод категории с наибольшим процентом возвратов (Рис. 41).

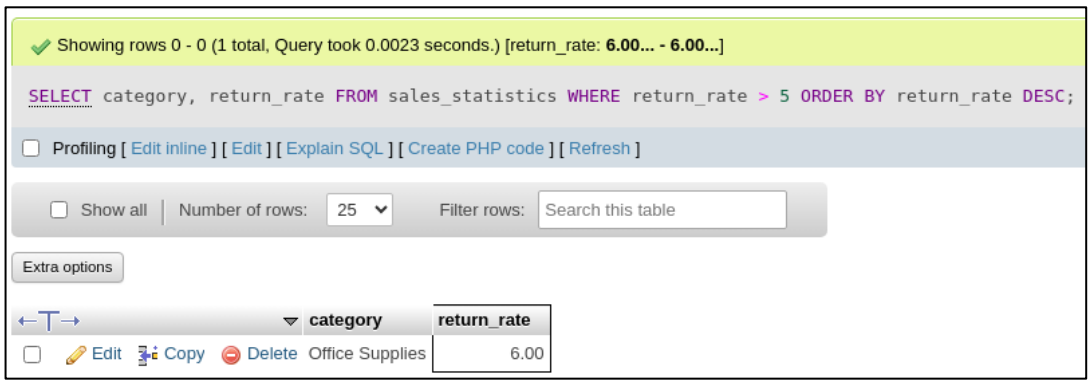


Рис. 41

3. Анализ по городам:

1) Добавление компонента для импорта данных из CSV (Рис. 42).

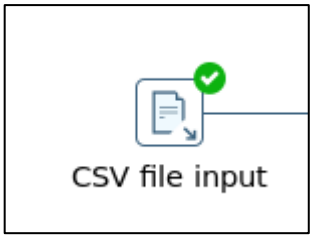


Рис. 42

Настройка компонента для импорта данных (Рис. 43).

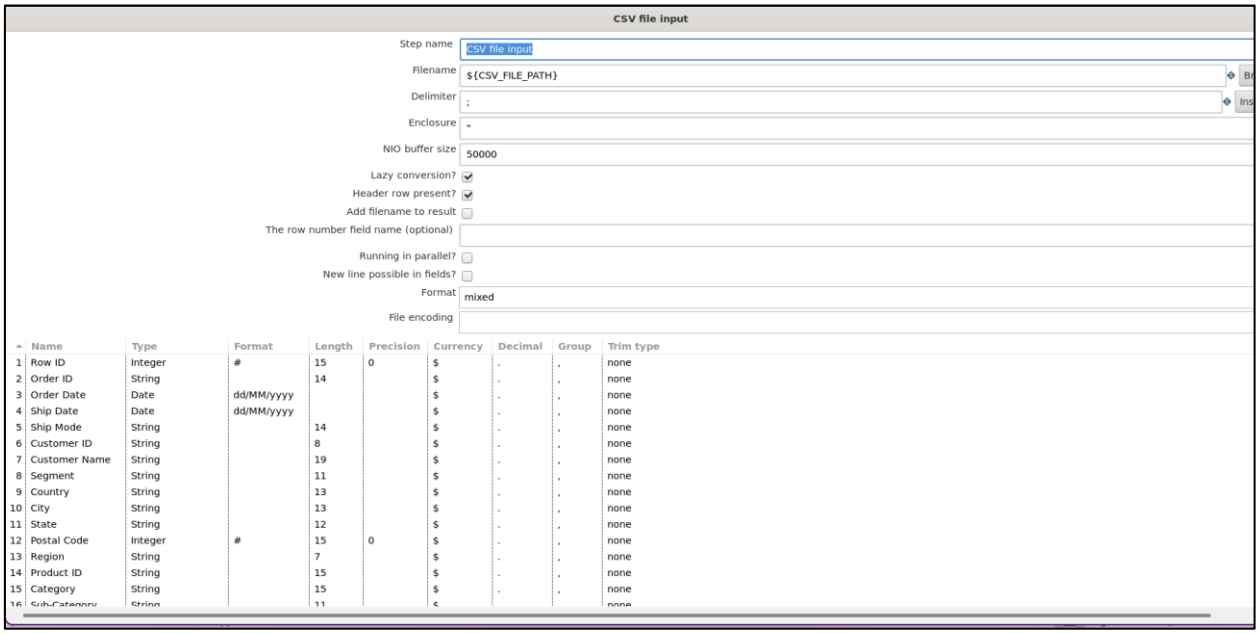


Рис. 43

2) Добавление компонента для валидации данных и запись логов при ошибке (Рис. 44).

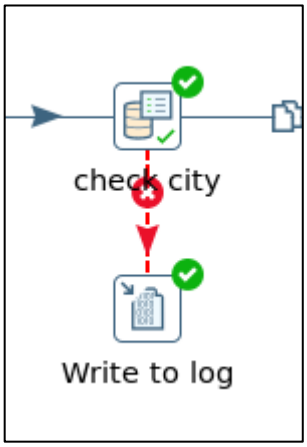


Рис. 44

Настройка компонента: значения в атрибуте «City» не могут быть пустыми, а также длина значения должна быть больше 2 (Рис. 45).

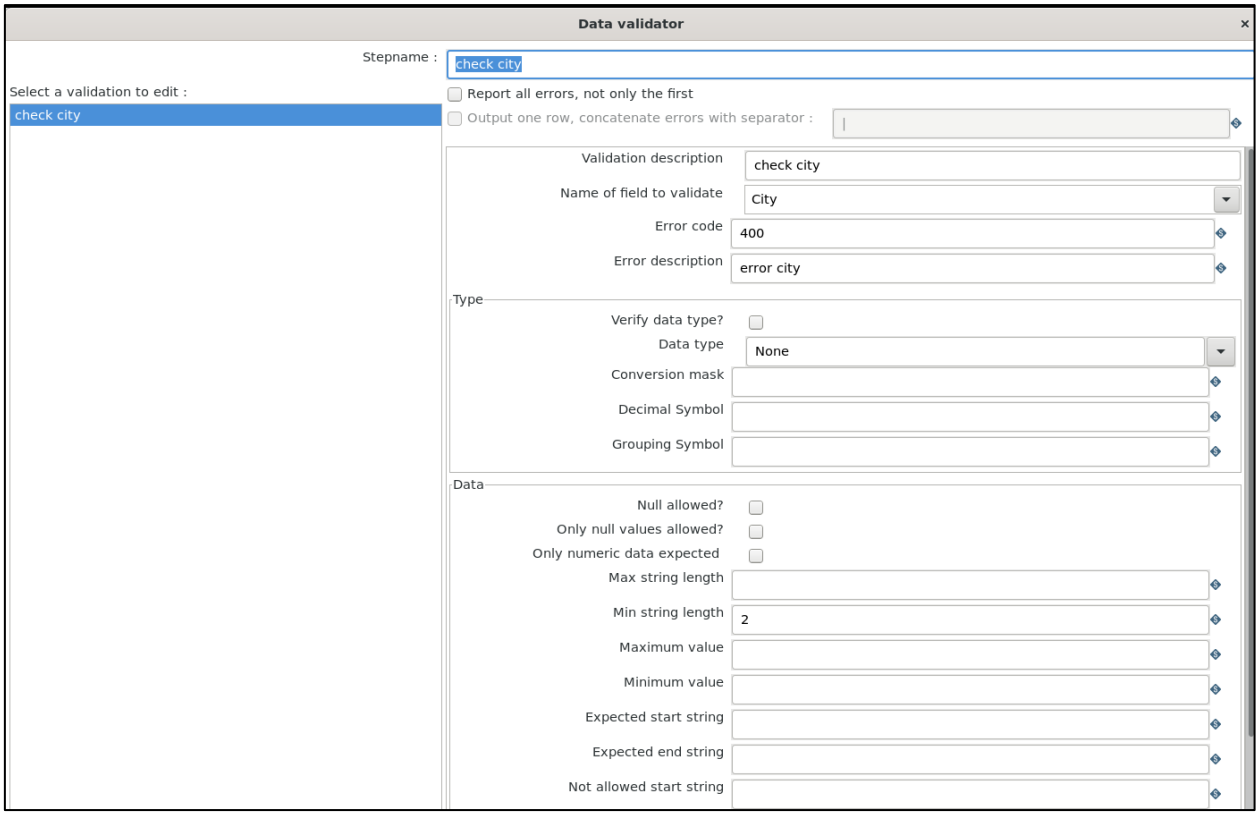


Рис. 45

3) Добавления компонента замены пустых значений (Рис. 46).

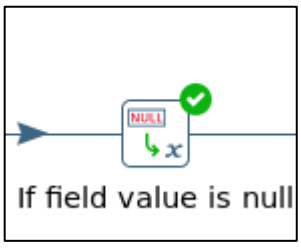


Рис. 46

Настройка компонента: заменить пустое значение в атрибуте «City» на «Unknown» (Рис. 47).

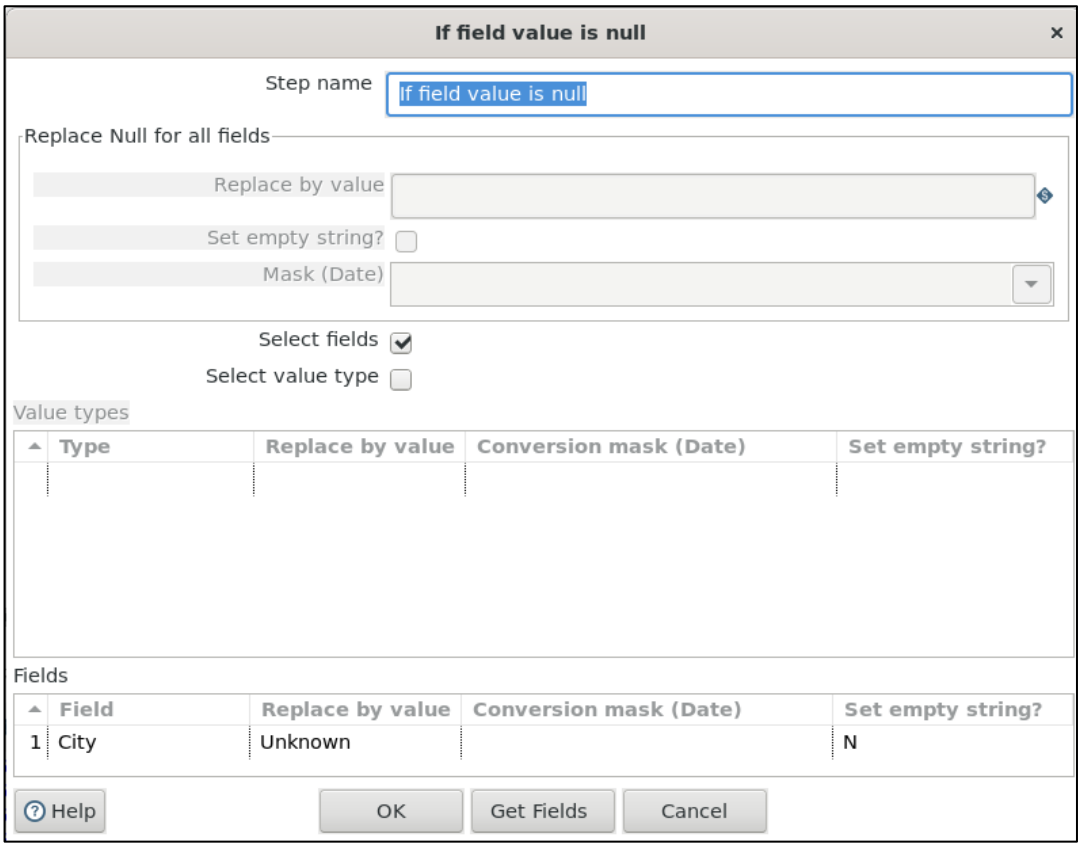


Рис. 47

4) Добавления компонента «Калькулятор» (Рис. 48).

Рис. 48

Настройка компонента: из атрибута «Order Date» берутся значения «Year», «Month» и записываются в новые атрибуты (Рис. 49). Эти атрибуты потребуются для дальнейшего анализа.

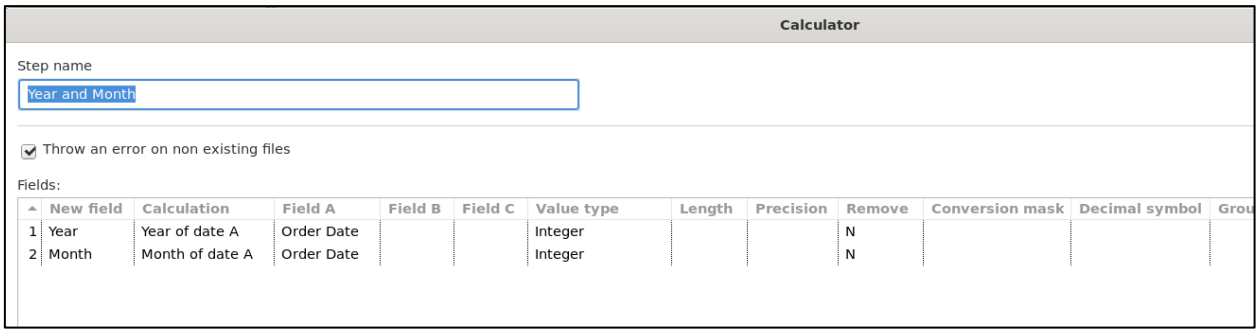


Рис. 49

5) Добавление компонента «Group by» (Рис. 50).

Рис. 50

Настройка компонента: группировка данных по атрибутам город, месяц, год. Агрегирование данных (Рис. 51):

-Сумма продаж;

-Сумма по выручке;

- Количество заказов;

-Средняя стоимость скидки.

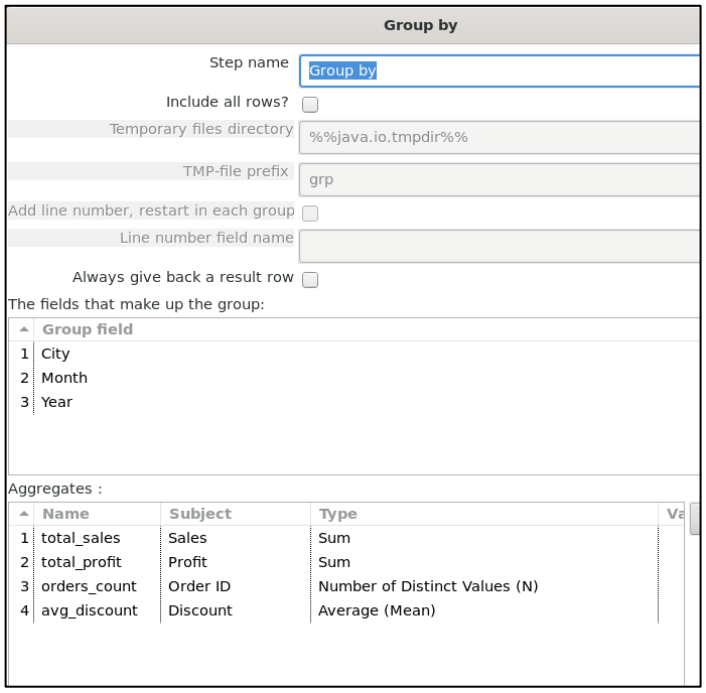


Рис. 51

6) Добавление компонента «Калькулятор» для расчёта новой метрики (Рис. 52).

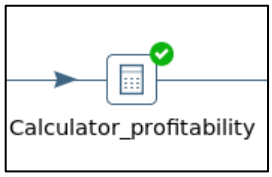


Рис. 52

Настройка компонента: расчёт новой метрики – рентабельность (Рис. 53).

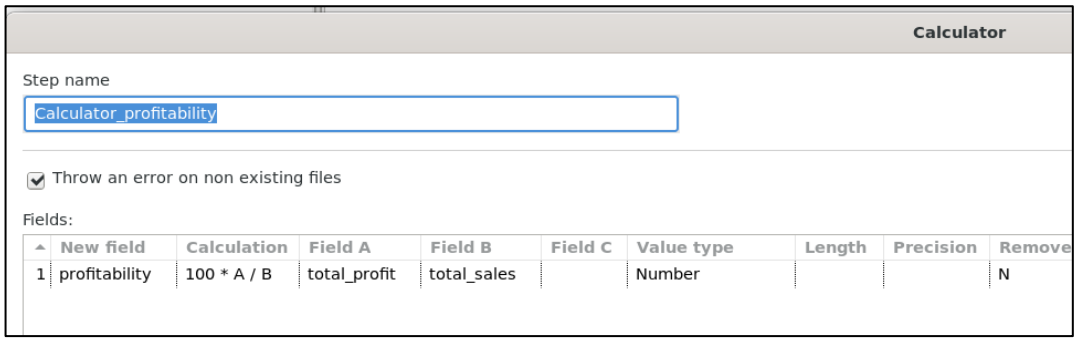


Рис. 53

7) Добавление компонента для экспорта данных (Рис. 54).

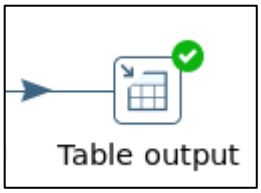


Рис. 54

Настройка компонента: подключение к БД для записи данных в таблицу «city_stats» (Рис. 55).

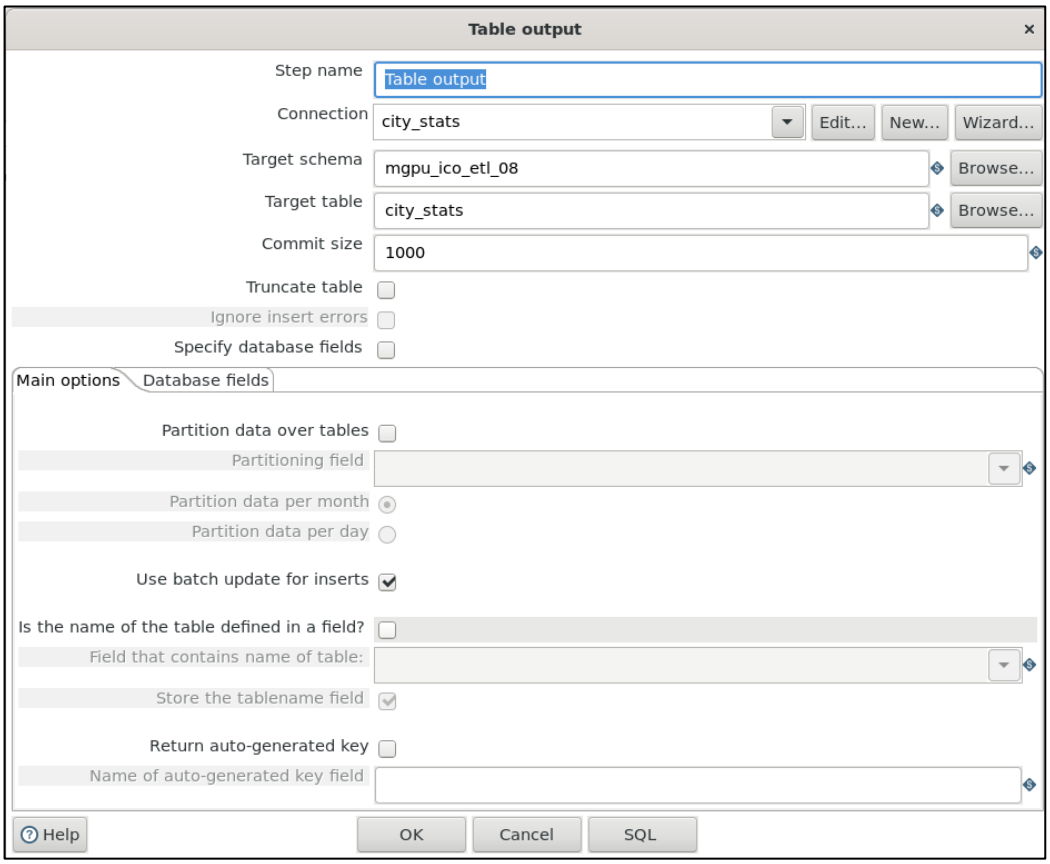


Рис. 55

Общий вид трансформации (Рис. 56).

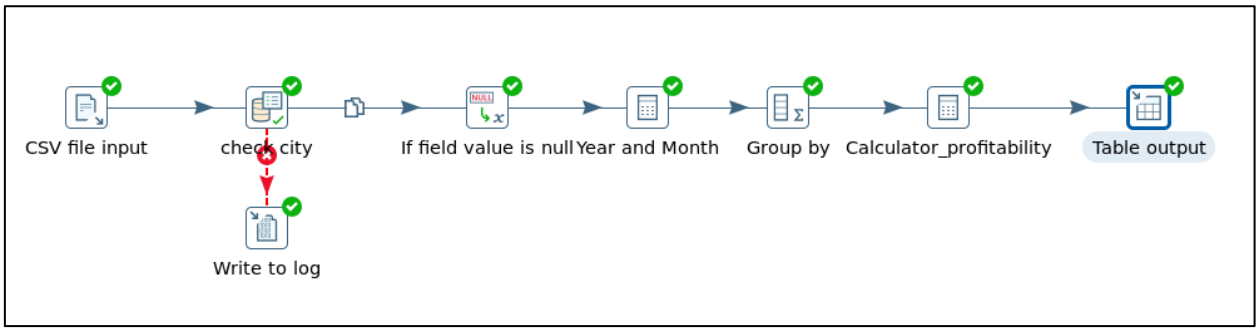


Рис. 56

8) Добавление компонента трансформации в Job (Рис. 57).

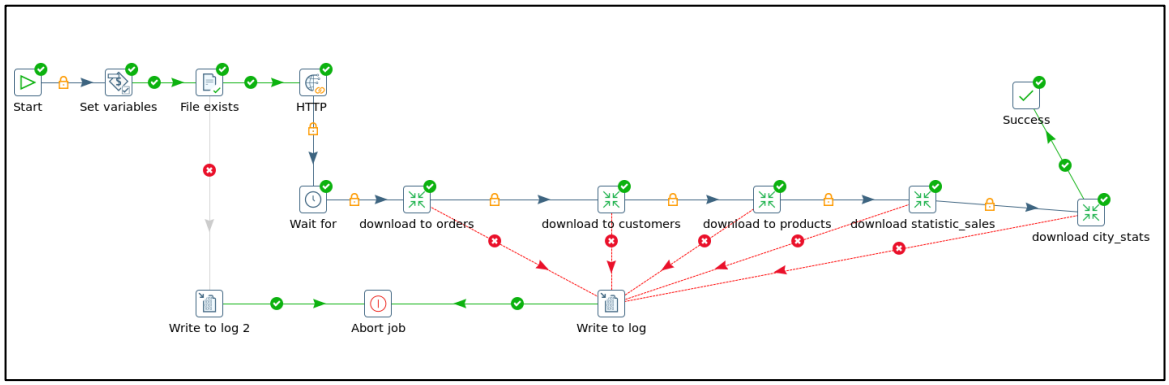


Рис. 57

Настройка компонента трансформации (Рис. 58).

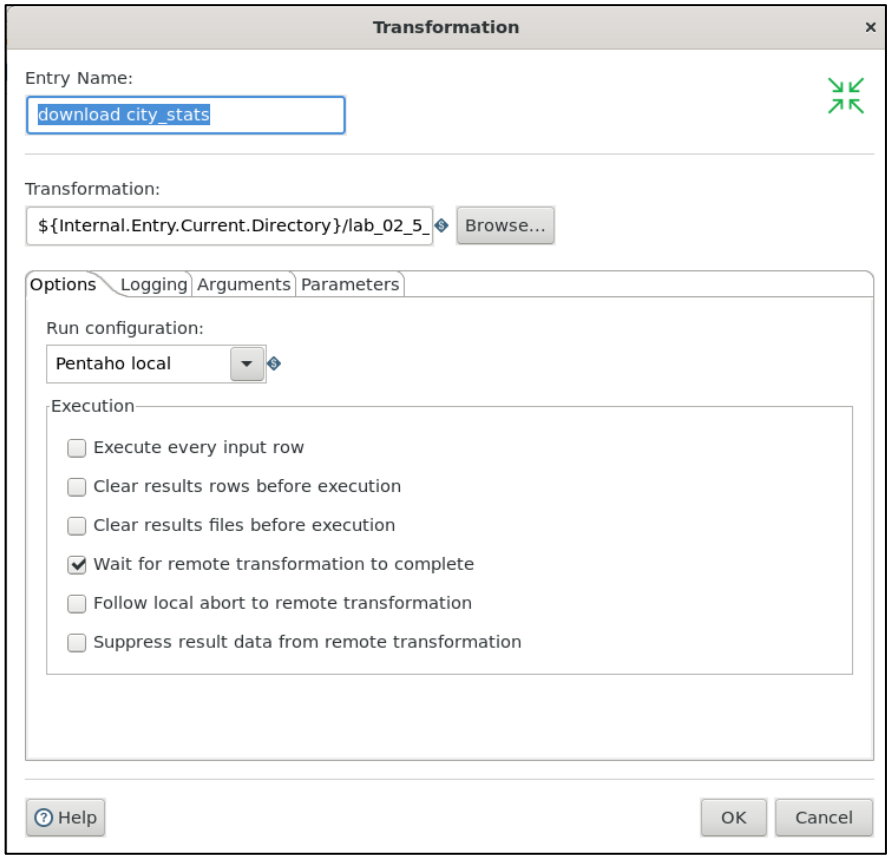


Рис. 58

9) Создание таблицы «city_stats» для экспорта данных в БД (Рис. 59).

```sql
CREATE TABLE city_stats (
    city VARCHAR(100) PRIMARY KEY,
    `Year` INT,
    `Month` int,
    total_sales DECIMAL(12,2),
    total_profit DECIMAL(12,2),
    orders_count INT,
    avg_discount DECIMAL(5,2),
    return_rate DECIMAL(5,2),
    profitability DECIMAL(5,2)
);
```

Рис. 59

Создание индексов (Рис. 60).

```sql
CREATE INDEX idx_city ON city_stats (city);
CREATE INDEX idx_profitability ON city_stats (profitability);
```

Рис. 60

10) Запрос к базе данных для проверки экспорта данных Рис. 61().

```sql
SELECT * FROM `city_stats`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

1 > >> | Number of rows: 25 Filter rows: Search this table Sort by key: None

Extra options

city	Year	Month	total_sales	total_profit	orders_count	avg_discount	profitability
Houston	2016	4	697.07	61.79	1	0.20	8.86
Jackson	2016	11	508.62	219.08	1	0.00	43.07
Chicago	2016	12	2.39	-6.34	1	0.80	-265.00
El Paso	2016	12	803.96	-30.96	1	0.20	-3.85
Houston	2016	10	5.31	-1.59	1	0.60	-30.00
Indianapolis	2016	12	1107.66	498.52	1	0.00	45.01
Chicago	2016	9	331.54	-82.88	1	0.20	-25.00
Huntsville	2016	10	1488.77	117.40	1	0.50	7.89
Chicago	2016	9	10.90	3.41	1	0.20	31.25
El Paso	2016	2	11.36	-4.52	1	0.50	-39.80
Houston	2016	1	16.45	5.55	1	0.20	33.75
Fort Worth	2016	5	225.02	-70.73	1	0.30	-31.43
Laredo	2016	12	609.61	63.25	1	0.30	10.38
Houston	2016	12	60.42	6.04	1	0.20	10.00

Рис. 61

11) Запросы к таблице для аналитики по городам:

- Анализ суммы продаж и суммы выручки по городам (Рис. 62).

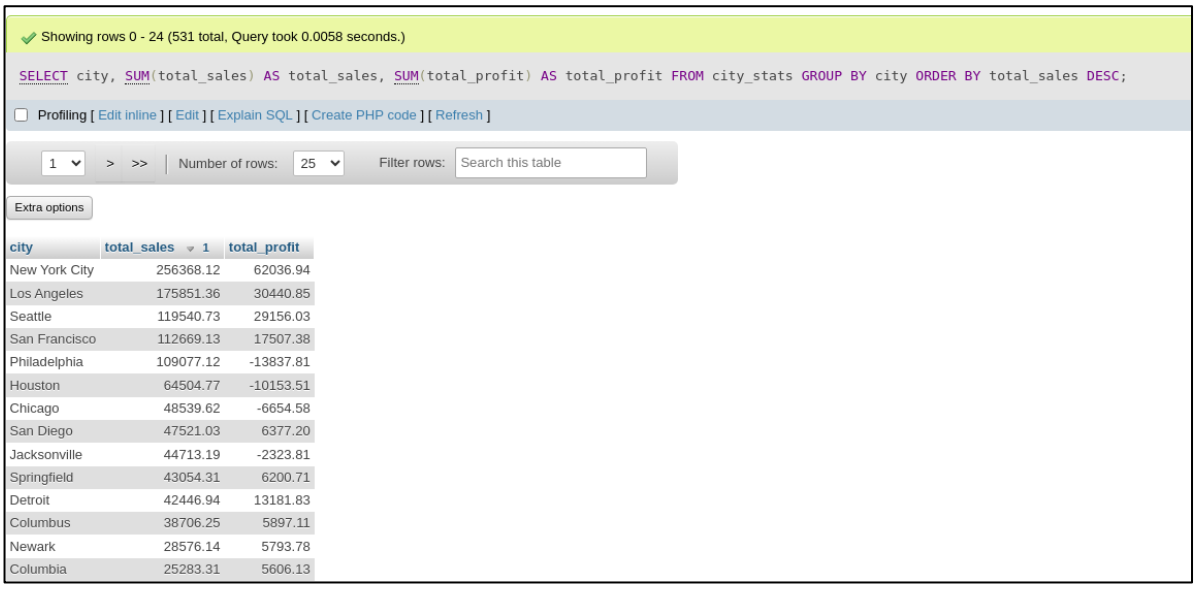


Рис. 62

- Анализ количества заказов по города и размер скидки (Рис. 63).

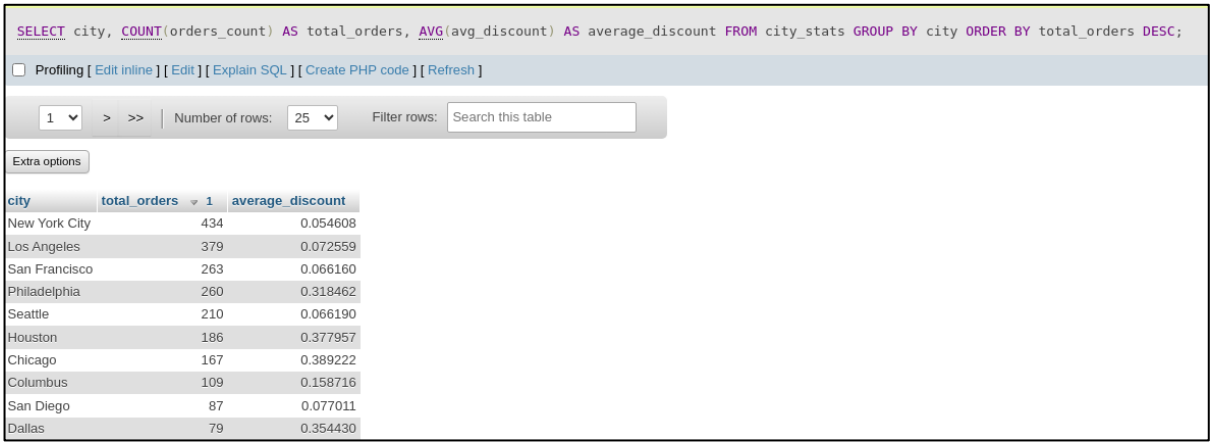


Рис. 63

- Анализ рентабельности по городам (Рис. 64).

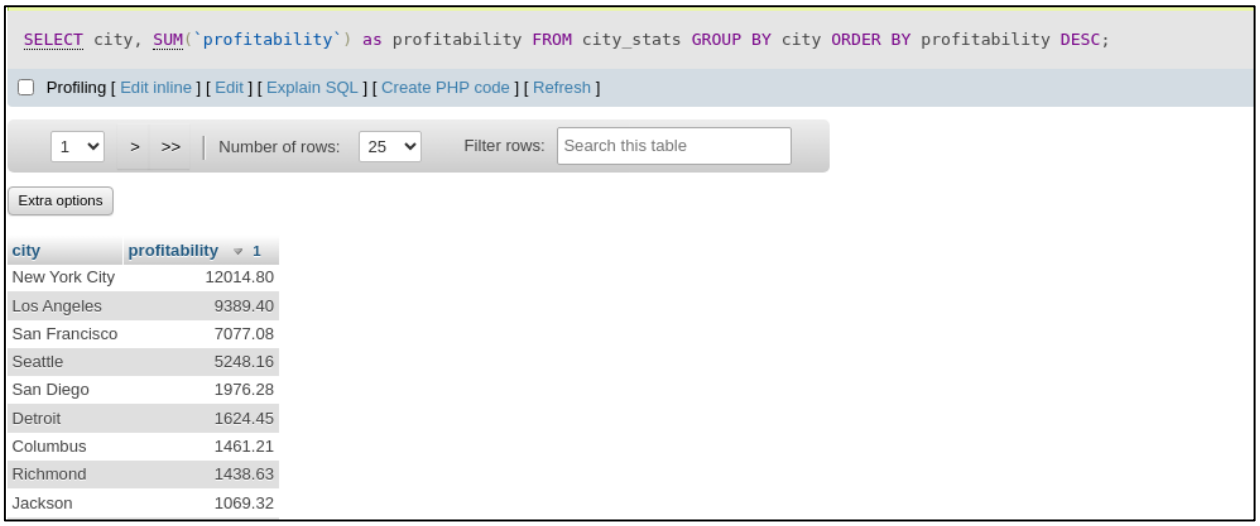


Рис. 64

Выводы по работе:

В ходе выполнения лабораторной работы были успешно реализованы ключевые этапы проектирования ETL-процесса в Pentaho Data Integration. Основной целью работы являлось освоение навыков интеграции данных из различных источников с использованием динамических соединений. Для этого были созданы подключения к CSV-файлам и базе данных MySQL, а также настроены трансформации для обработки данных, включая фильтрацию, замену значений и агрегацию. Результаты работы подтвердили корректность настройки компонентов: данные были загружены в таблицы orders, customers, products, а затем преобразованы и сохранены в аналитических таблицах sales_statistics и city_stats.

Проверка результатов осуществлялась через SQL-запросы, которые подтвердили отсутствие ошибок в данных и корректность агрегации. Например, запросы на определение топ-3 регионов по продажам и анализ рентабельности по городам продемонстрировали работоспособность реализованных процессов. Для оптимизации производительности были созданы индексы в таблицах БД.