

Дипломный проект

Тема: Исследование данных торговой компании с
помощью SQL

Выполнила: Хамидова О.А.

2024

Этап 1. Исследование полученных данных

1. Структура таблицы с данными

Таблица 1

№	Название колонки	Тип данных
1	dt	Дата
2	tm	Время
3	order_number	Числовой
4	order_type	Строковые
5	product_category	Строковые
6	product	Строковые
7	manufacturer	Строковые
8	cnt	Числовой
9	price	Числовой
10	selling_price	Числовой

2. Количество строк в таблице данных

368

3. Детальная структура таблицы с данными по каждому полю.

Таблица 2

№	Название колонки	Характеристика данных	Назначение
1	dt	Временной период данных	Дата продажи товара
2	tm	Временной период данных	Время продажи товара
3	order_number	Числовое значение	Номер заказа
4	order_type	Набор уникальных значений	Тип заказа
5	product_category	Набор уникальных значений	Категория продукта
6	product	Набор уникальных значений	Продукт
7	manufacturer	Набор уникальных значений	Производитель
8	cnt	Числовое значение	Количество проданного товара
9	price	Числовое значение	Цена производителя
10	selling_price	Числовое значение	Отпускная цена

Для визуального понимания таблицы данных - представлена часть таблицы (20 первых строк)

dt	tm	order_number	order_type	product_category	product	manufacturer	cnt	price	selling_price
2000-07-30	09:32:00.0000000	1	purchase order	Dry building mixes	fugen	Knauf	5400	3.2	<i>null</i>
2000-07-30	09:55:00.0000000	1	sales invoice	Dry building mixes	fugen	Knauf	2870	3.2	3.33
2000-08-04	10:10:00.0000000	2	sales invoice	Dry building mixes	fugen	Knauf	2336	3.2	3.48
2000-08-04	11:04:00.0000000	3	sales invoice	Dry building mixes	fugen	Knauf	26	3.2	3.58
2000-08-04	11:21:00.0000000	4	sales invoice	Dry building mixes	fugen	Knauf	9	3.2	3.28
2000-08-04	12:01:00.0000000	5	sales invoice	Dry building mixes	fugen	Knauf	36	3.2	3.41
1933-03-06	12:55:00.0000000	99	sales invoice	Varnishes and paints	kiva 30	Tikkurila	131	8.84	9.77
2000-08-08	11:49:00.0000000	6	sales invoice	Dry building mixes	fugen	Knauf	48	3.2	3.38
2000-08-08	13:16:00.0000000	8	sales invoice	Dry building mixes	fugen	Knauf	999999999	3.2	3.31
2000-08-08	12:29:00.0000000	7	sales invoice	Dry building mixes	fugen	Knauf	7	3.2	3.58
2000-08-08	13:16:00.0000000	8	sales invoice	Dry building mixes	fugen	Knauf	57	3.2	3.31
2000-08-08	13:19:00.0000000	2	purchase order	Construction chemistry	grout	Ceresit	400	16.49	<i>null</i>
2000-08-13	11:49:00.0000000	9	sales invoice	Dry building mixes	fugen	Knauf	3	3.2	3.58
2000-08-13	11:49:00.0000000	9	sales invoice	Construction chemistry	grout	Ceresit	196	16.49	17.53
2000-08-31	11:37:00.0000000	10	sales invoice	Construction chemistry	grout	Ceresit	179	16.49	17.37
2000-08-31	11:37:00.0000000	10	sales invoice	Dry building mixes	fugen	Knauf	2	3.2	3.45
2000-08-31	12:01:00.0000000	3	purchase order	Varnishes and paints	kiva 30	Tikkurila	4000	8.84	<i>null</i>
2000-09-11	10:28:00.0000000	11	sales invoice	Construction chemistry	grout	Ceresit	17	16.49	18.25
2000-09-11	10:28:00.0000000	11	sales invoice	Varnishes and paints	kiva 30	Tikkurila	3647	8.84	9.21
2000-09-11	10:39:00.0000000	12	sales invoice	Dry building mixes	fugen	Knauf	5	3.2	3.6

4. Общее описание исследованных данных.

Хотела бы обратить Ваше внимание на следующие моменты в предоставленной таблице данных:

- Список столбцов с указанием типов данных

TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE
invoice_order_operations	dt	1	null	YES	date
invoice_order_operations	tm	2	null	YES	time
invoice_order_operations	order_number	3	null	YES	int
invoice_order_operations	order_type	4	null	YES	varchar
invoice_order_operations	product_category	5	null	YES	varchar
invoice_order_operations	product	6	null	YES	varchar
invoice_order_operations	manufacturer	7	null	YES	varchar
invoice_order_operations	cnt	8	null	YES	int
invoice_order_operations	price	9	null	YES	float
invoice_order_operations	selling_price	10	null	YES	float

- В поле «дата» присутствуют строки с некорректно введенным годом.

cnt	min_dt	max_dt
368	1877-09-08	2001-08-16

dt
1877-09-08
1877-10-27
1891-06-07
1891-07-06
1905-03-15
1905-08-22
1918-10-04
1918-11-07
1933-01-07
1933-03-06
1946-10-08
1960-03-18
1960-05-25
1973-08-13
1974-02-02
1987-03-23

- Номер заказа присваивается относительно группировки по типу заказа.
- В таблице данных представлены два типа заказа - заказ на поставку и продажа.

Общее количество заказов составляет 368, с минимальной отпускной ценой 3,28, до максимальной – 97,69.

Информация о продаже - 323 заказа.

Заказ на поставку – 45.

- В таблице присутствует пять категорий продуктов. Из которых две категории задвоены, т.к. написаны с прописной и строчной буквы.
- Также присутствует опечатка в наименовании производителя.

order_type
purchase order
sales invoice

cnt	min_price	max_price	avg_price
368	3.28	97.69	18.8518885448916

cnt	min_price	max_price	avg_price
323	3.28	97.69	18.8518885448916

cnt	min_price	max_price	avg_price
45	null	null	null

product_category
Construction chemistry
Dry building mixes
insulation Material
Insulation materials
Two-component mortar
Varnishes and paints

manufacturer
Ceresit
Knauf
Makroflex
Rockwool
Tikkurila
Tikurila

- Количество проданного товара («cnt») в семи строках введено не корректно, т.к. указано 999999999.

dt	tm	order_number	order_type	product_category	product	manufacturer	cnt	price	selling_price
2000-08-08	13:16:00.0000000	8	sales invoice	Dry building mixes	fugen	Knauf	999999999	3.2	3.31
2001-03-13	11:30:00.0000000	56	sales invoice	Construction chemistry	polyurethane foam	Makroflex	999999999	3.64	4.28
2000-12-29	13:36:00.0000000	47	sales invoice	Insulation materials	rockclose insulated dpc 20	Rockwool	999999999	20.3	21.17
2001-07-11	12:06:00.0000000	83	sales invoice	Varnishes and paints	helmi 30	Tikkurila	999999999	9.36	10.06
2000-09-21	14:53:00.0000000	18	sales invoice	Varnishes and paints	kiva 70	Tikkurila	999999999	8.84	9.74
2000-12-26	11:17:00.0000000	38	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	999999999	3.23	3.76
2000-12-12	10:53:00.0000000	36	sales invoice	Dry building mixes	rotband	Knauf	999999999	23.95	25.41

- В некоторых строках столбца «price» присутствуют отрицательные значения.

dt	tm	order_number	order_type	product_category	product	manufacturer	cnt	price	selling_price
2001-06-19	09:58:00.0000000	71	sales invoice	Dry building mixes	rotband	Knauf	1083	-23.95	26.95
2001-03-10	10:37:00.0000000	55	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	890	-3.23	3.96
2000-11-29	13:51:00.0000000	29	sales invoice	Varnishes and paints	everal aqua 40	Tikkurila	341	-9.88	10.79
2000-11-29	11:38:00.0000000	13	purchase order	Dry building mixes	rotband	Knauf	2200	-23.95	null
2001-06-20	15:35:00.0000000	76	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	3084	-3.23	3.89
2001-07-11	12:37:00.0000000	85	sales invoice	Insulation materials	partial fill cavity slab 100	Rockwool	405	-44.47	48.68
2001-03-13	11:30:00.0000000	56	sales invoice	Construction chemistry	polyurethane foam	Makroflex	230	-3.64	4.28
2001-06-20	17:37:00.0000000	81	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	12	-3.23	3.97
2001-01-10	14:01:00.0000000	26	purchase order	Varnishes and paints	everal aqua 10 interior	Tikkurila	5800	-9.88	null
2001-04-12	12:09:00.0000000	67	sales invoice	Two-component mortar	epoxy grouts	Ceresit	1836	-90	94.97
2000-12-29	13:36:00.0000000	47	sales invoice	Insulation materials	rockclose insulated dpc 20	Rockwool	1724	-20.3	21.17
2001-06-20	17:37:00.0000000	81	sales invoice	Varnishes and paints	everal aqua 10	Tikkurila	8	-9.88	10.95
2001-06-20	15:56:00.0000000	77	sales invoice	Two-component mortar	epoxy grouts	Ceresit	2	-90	95.08
2001-03-13	13:04:00.0000000	60	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	1	-3.23	3.85
2000-12-26	13:27:00.0000000	42	sales invoice	Insulation materials	partial fill cavity slab 50	Rockwool	4	-22.21	24.71
2001-06-20	17:00:00.0000000	80	sales invoice	Construction chemistry	polyurethane foam premium of winter	Makroflex	25	-3.23	4.08

Этап 2. Исправление ошибок в данных

Характеристики данных таблицы после внесения корректировок

1. Количество строк в таблице

329

2. Количество строк в распределении по каждой категории товара

Construction chemistry	Dry building mixes	Insulation materials	Two-component mortar	Varnishes and paints
48	89	79	11	102

3. Количество строк в распределении по каждому производителю

Ceresit	Knauf	Makroflex	Rockwool	Tikkurila
41	64	43	79	102

4. Минимальное, максимальное и среднее значения по колонке cnt

min_cnt	max_cnt	avg_cnt
1	5800	751

5. Минимальное, максимальное и среднее значения по колонке price

min_price	max_price	avg_price
3,2	90	16,85

Список использовавшихся SQL-запросов

1. Выбрала уникальные значения из таблицы с данными по категории товара. Обнаружив опечатку, изменила название категории товара. Использовала следующие запросы:
 - `select distinct product_category from invoice_order_operations`
 - `update invoice_order_operations set product_category='Insulation materials' where product_category='insulation Material'`
2. Выбрала уникальные значения из базы данных по производителю. Обнаружив опечатку, изменила название производителя. Использовала следующие запросы:
 - `select distinct manufacturer from invoice_order_operations`
 - `update invoice_order_operations set manufacturer='Tikkurila' where manufacturer='Tikurila'`
3. Удалила некорректные данные из таблицы, т.к. они выходят за допустимые рамки:
 - Даты (столбец «dt»)
`delete from invoice_order_operations where year(dt)<2000`
 - Количество (столбец «cnt»)
`delete from invoice_order_operations where cnt='999999999'`
 - Цена производителя (столбец «price») – удалила отрицательные значения
`delete from invoice_order_operations where price<=0`

Этап 3. Анализ финансовых показателей

1. Оценка динамики закупок.

Период	Количество накладных	Сумма закупки
07-2000	1	9 184,00
08-2000	9	14 260,55
09-2000	9	293 038,27
10-2000	1	5 222,08
11-2000	14	241 577,93
12-2000	13	294 970,42
01-2001	6	101 378,63
03-2001	13	95 084,60
04-2001	4	390 125,52
06-2001	11	260 252,43
07-2001	14	199 394,05
08-2001	7	174 570,38

Сумма закупки по месяцам



Используемый SQL-запрос:
select format(dt, 'MM-yyyy') период, count(distinct order_number) количество_накладных, sum(cnt*price) сумма_закупки from invoice_order_operations where order_type='sales invoice' group by format(dt, 'MM-yyyy') order by format(dt, 'MM-yyyy')

2. Прибыль компании за весь период.

152 915,30

Используемый SQL-запрос:
select sum(cnt*selling_price-cnt*price) прибыль from invoice_order_operations

3. Оценка динамики прибыли.

Период	Количество	Выручка	Прибыль
07-2000	1	9 557,10	373,10
08-2000	11	15 313,36	1 052,81
09-2000	19	311 582,09	18 543,82
10-2000	3	5 770,83	548,75
11-2000	54	258 877,57	17 299,64
12-2000	38	311 646,44	16 676,02
01-2001	27	109 103,00	7 724,37
03-2001	39	102 495,34	7 410,74
04-2001	9	412 078,08	21 952,56
06-2001	36	286 415,39	26 162,96
07-2001	34	215 813,01	16 418,96
08-2001	16	193 321,95	18 751,57

Выручка и прибыль по месяцам



Используемый SQL-запрос:
select format(dt, 'MM-yyyy') период, count(cnt) количество, sum(cnt*selling_price) выручка, sum(cnt*selling_price-cnt*price) прибыль from invoice_order_operations where order_type='sales invoice' group by format(dt, 'MM-yyyy') order by format(dt, 'MM-yyyy')

4. Оценка динамики продаж в целом.

Период	Сумма закупки	Выручка
07-2000	9 184,00	9 557,10
08-2000	14 260,55	15 313,36
09-2000	293 038,27	311 582,09
10-2000	5 222,08	5 770,83
11-2000	241 577,93	258 877,57
12-2000	294 970,42	311 646,44
01-2001	101 378,63	109 103,00
03-2001	95 084,60	102 495,34
04-2001	390 125,52	412 078,08
06-2001	260 252,43	286 415,39
07-2001	199 394,05	215 813,01
08-2001	174 570,38	193 321,95

Закупка и выручка по месяцам

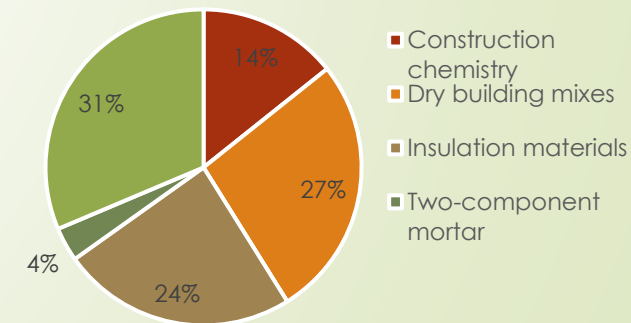


Используемый SQL-запрос:
 select format(dt, 'MM-YYYY') период, sum(cnt*price) сумма_закупки, sum(cnt*selling_price) выручка
 from invoice_order_operations where order_type='sales invoice' group by format(dt, 'MM-YYYY') order by format(dt, 'MM-YYYY')

5. Оценка динамики продаж по категориям товара.

Категория товара	Количество	Выручка	Прибыль
Construction chemistry	41	63 974,95	10 354,95
Dry building mixes	77	333 886,58	24 486,58
Insulation materials	69	962 087,90	66 525,04
Two-component mortar	10	456 399,36	24 399,36
Varnishes and paints	90	415 625,37	27 149,37

Количество товара



Используемый SQL-запрос:
 select product_category, count(cnt) количество, sum(cnt*selling_price) выручка, sum(cnt*selling_price-cnt*price) прибыль
 from invoice_order_operations where order_type='sales invoice' group by product_category order by product_category

Этап 4. Модернизация структуры БД

1. С целью дальнейшего удобства работы с данными и оптимизации хранения были созданы четыре таблицы-справочники:

- **order_type**

(используемый sql-запрос: create table order_type(order_type_id int not null identity(1,1) primary key,order_type varchar(200) default null))

- **product_category**

(используемый sql-запрос: create table product_category(product_category_id int not null identity(1,1) primary key,product_category varchar(200) default null))

- **product**

(используемый sql-запрос: create table product(product_id int not null identity(1,1) primary key,product varchar(200) default null))

- **manufacturer**

(используемый sql-запрос: create table manufacturer(manufacturer_id int not null identity(1,1) primary key,manufacturer varchar(200) default null))

2. Таблицы-справочники были заполнены данными из соответствующих столбцов таблицы invoice_order_operations.

(используемый sql-запрос: create table product(product_id int not null identity(1,1) primary key,product varchar(200) default null))

3. Создана таблица со следующей структурой данных:

dt	tm	order_number	order_type_id	product_category_id	product_id	manufacturer_id	cnt	price	selling_price
----	----	--------------	---------------	---------------------	------------	-----------------	-----	-------	---------------

order_type_id	order_type
1	purchase order
2	sales invoice

product_category_id	product_category
1	Construction chemistry
2	Dry building mixes
3	Insulation materials
4	Two-component mortar
5	Varnishes and paints

product_id	product
1	epoxy grouts
2	everal aqua 10
3	everal aqua 10 interior
4	everal aqua 40
5	fugen
6	glue cm 11
7	glue cm 17
8	grout
9	helmi 10
10	helmi 30
11	helmi primer
12	kiva 10
13	kiva 30
14	kiva 70
15	partial fill cavity slab 100
16	partial fill cavity slab 50
17	partial fill cavity slab 80
18	polyurethane foam
19	polyurethane foam premium of winter
20	rockclose insulated dpc 20
21	rotband
22	uniflott

manufacturer_id	manufacturer
1	Ceresit
2	Knauf
3	Makroflex
4	Rockwool
5	Tikkurila

4. Таблица operations_data заполнена данными из таблицы invoice_order_operations и всех таблиц-справочников

(используемый sql-запрос: insert into operations_data select dt, tm, order_number, ot.order_type_id, pc.product_category_id, p.product_id, m.manufacturer_id, cnt, price, selling_price from invoice_order_operations a left join order_type ot on a.order_type = ot.order_type left join product_category pc on a.product_category = pc.product_category left join product p on a.product = p.product left join manufacturer m on a.manufacturer = m.manufacturer)

dt	tm	order_number	order_type_id	product_category_id	product_id	manufacturer_id	cnt	price	selling_price
2000-07-30	09:32:00.0000000	1	1	2	5	2	5400	3.2	null
2000-07-30	09:55:00.0000000	1	2	2	5	2	2870	3.2	3.33
2000-08-04	10:10:00.0000000	2	2	2	5	2	2336	3.2	3.48
2000-08-04	11:04:00.0000000	3	2	2	5	2	26	3.2	3.58
2000-08-04	11:21:00.0000000	4	2	2	5	2	9	3.2	3.28
2000-08-04	12:01:00.0000000	5	2	2	5	2	36	3.2	3.41
2000-08-08	11:49:00.0000000	6	2	2	5	2	48	3.2	3.38
2000-08-08	12:29:00.0000000	7	2	2	5	2	7	3.2	3.58
2000-08-08	13:16:00.0000000	8	2	2	5	2	57	3.2	3.31
2000-08-08	13:19:00.0000000	2	1	1	8	1	400	16.49	null
2000-08-13	11:49:00.0000000	9	2	2	5	2	3	3.2	3.58
2000-08-13	11:49:00.0000000	9	2	1	8	1	196	16.49	17.53
2000-08-31	11:37:00.0000000	10	2	1	8	1	179	16.49	17.37
2000-08-31	11:37:00.0000000	10	2	2	5	2	2	3.2	3.45
2000-08-31	12:01:00.0000000	3	1	5	13	5	4000	8.84	null
2000-09-11	10:28:00.0000000	11	2	1	8	1	17	16.49	18.25
2000-09-11	10:28:00.0000000	11	2	5	13	5	3647	8.84	9.21
2000-09-11	10:39:00.0000000	12	2	2	5	2	5	3.2	3.6
2000-09-11	10:39:00.0000000	12	2	5	13	5	350	8.84	9.12
2000-09-11	10:39:00.0000000	12	2	1	8	1	8	16.49	17.75

Характеристики данных таблицы operations_data

1. Количество строк в таблице

329

2. Количество строк в распределении по каждой категории товара

Construction chemistry	Dry building mixes	Insulation materials	Two-component mortar	Varnishes and paints
48	89	79	11	102

3. Количество строк в распределении по каждому производителю

Ceresit	Knauf	Makroflex	Rockwool	Tikkurila
41	64	43	79	102

4. Минимальное, максимальное и среднее значения по колонке cnt

min_cnt	max_cnt	avg_cnt
1	5800	751

5. Минимальное, максимальное и среднее значения по колонке price


min_price	max_price	avg_price
3,2	90	16,85

Этап 5. Расчёт остатков на складе

Проанализировав движение товара за весь период, было рассчитано текущее состояние склада по каждой товарной позиции:

(используемый sql-запрос: `select distinct p.product,(select sum(s.cnt)from operations_data s where s.product_id=a.product_idand s.order_type_id=1) sklad_cnt,(select sum(p.cnt) from operations_data p where p.product_id=a.product_idand p.order_type_id=2) sale_cntfrom operations_data aleft join product p on a.product_id=p.product_id)`

Наименование товара	Количество на складе	Количество продано
epoxy grouts	4800	4800
everal aqua 10	3800	3800
everal aqua 10 interior	8400	8400
everal aqua 40	6000	6000
fugen	19000	19000
glue cm 11	5000	5000
glue cm 17	3000	3000
grout	400	400
helmi 10	3200	3200
helmi 30	3800	3800
helmi primer	5400	5400
kiva 10	1200	1200
kiva 30	5000	5000
kiva 70	4800	4800
partial fill cavity slab 100	7600	7600
partial fill cavity slab 50	3200	3200
partial fill cavity slab 80	9000	8998
polyurethane foam	4400	4400
polyurethane foam premium of winter	9600	9600
rockclose insulated dpc 20	8200	8200
rotband	6800	6800
uniflott	2200	null



Более детально рассмотреть таблицы можно по ссылке в онлайн-среде:

- <https://dbfiddle.uk/DzF3iu22?hide=144115188075855870> (этапы с 1 по 4)
 - <https://dbfiddle.uk/LSP7LLki> (этап 5)
- 