

Дипломный проект курса «Business Intelligence (BI) разработчик»

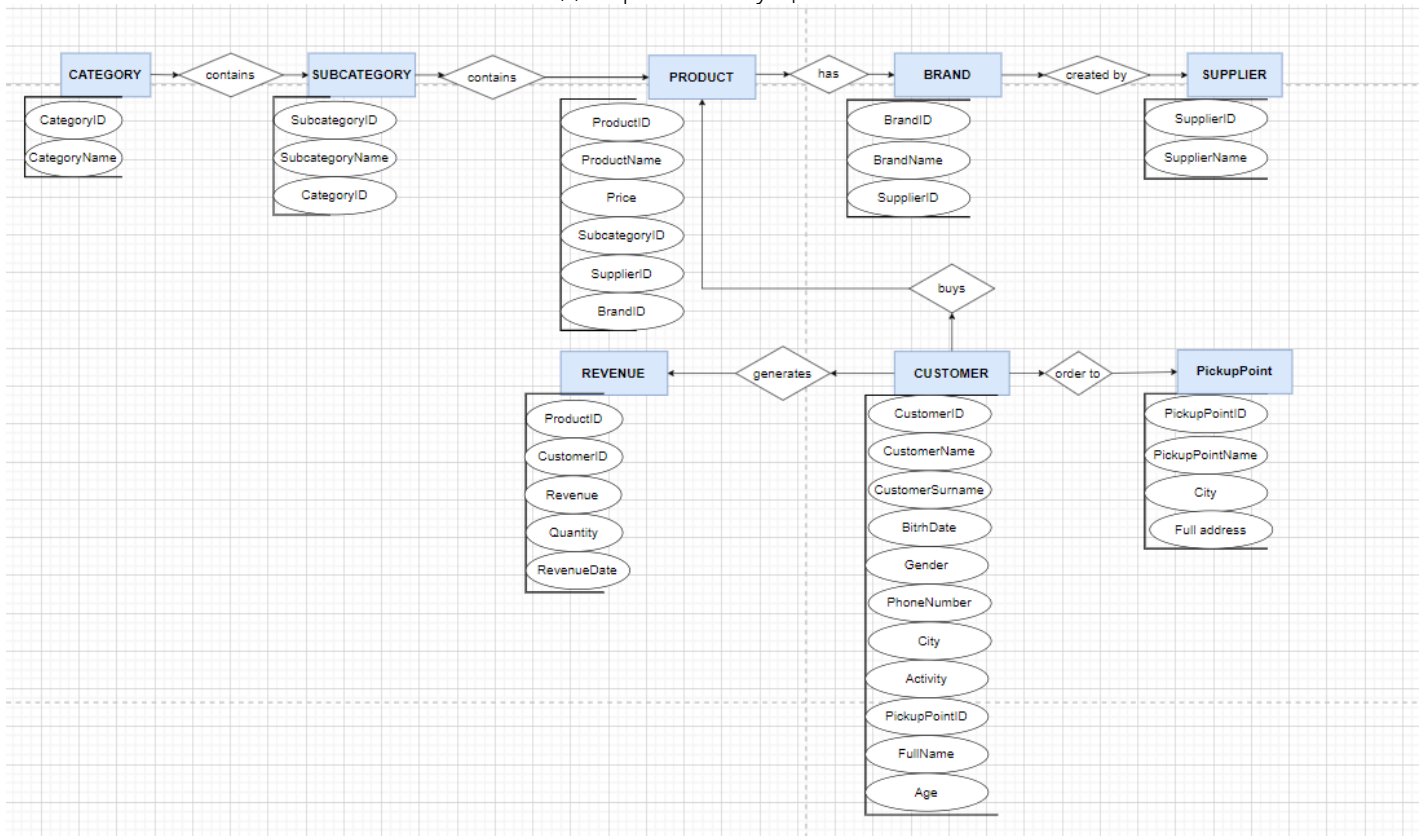
Автор:

Астрейко Наталья

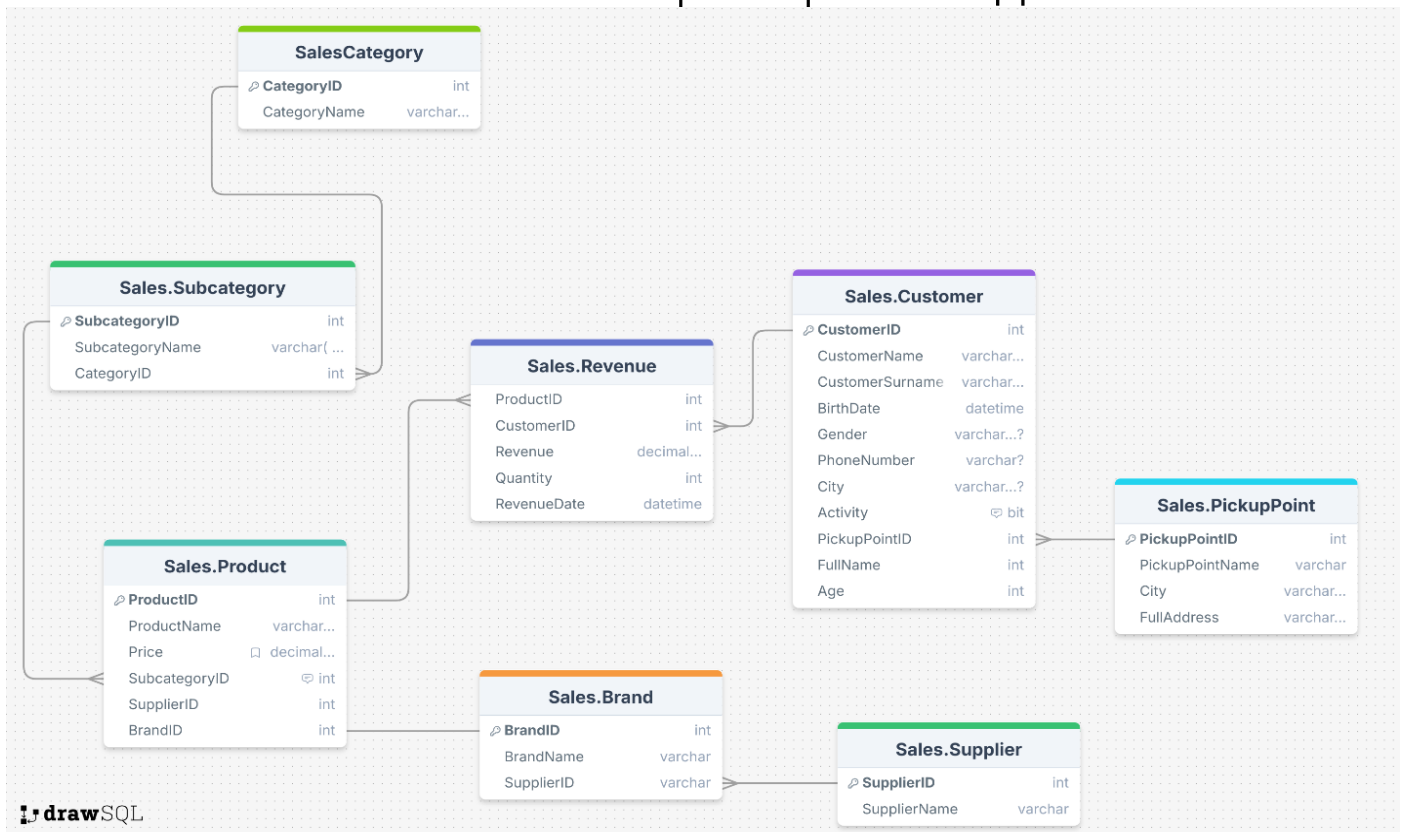
<https://www.linkedin.com/in/natallia-astreika>

Концептуальное проектирование базы данных (БД)

ER-диаграмма «Сущность-связь»



Логическое проектирование БД



Физическое проектирование БД

Таблицы, ключи и ограничения к ним были созданы с использованием MS SQL Server Management Studio (SSMS).

```
CREATE TABLE Sales.Category
(
    CategoryID INT PRIMARY KEY,
    CategoryName VARCHAR (50) NOT NULL
);

CREATE TABLE Sales.Subcategory
(
    SubcategoryID INT PRIMARY KEY,
    SubcategoryName VARCHAR (100) NOT NULL,
    CategoryID INT FOREIGN KEY (CategoryID) REFERENCES Sales.Category (CategoryID) NOT NULL
);

CREATE TABLE Sales.PickupPoint
(
    PickupPointID INT PRIMARY KEY,
    PickupPointNumber VARCHAR (25) UNIQUE,
    City VARCHAR (25) NOT NULL,
    FullAddress VARCHAR (100) NOT NULL,
);

CREATE TABLE Sales.Customer
(
    CustomerID INT PRIMARY KEY,
    CustomerName VARCHAR (25) NOT NULL,
    CustomerSurname VARCHAR (25) NOT NULL,
    BirthDate DATETIME NOT NULL,
    Gender VARCHAR(2),
    PhoneNumber VARCHAR(25),
    City VARCHAR (25),
    Activity BIT NOT NULL,
    PickupPointID INT FOREIGN KEY (PickupPointID) REFERENCES Sales.PickupPoint (PickupPointID)
);

CREATE TABLE Sales.Supplier
(
    SupplierID INT PRIMARY KEY,
    SupplierName VARCHAR (50) NOT NULL,
);

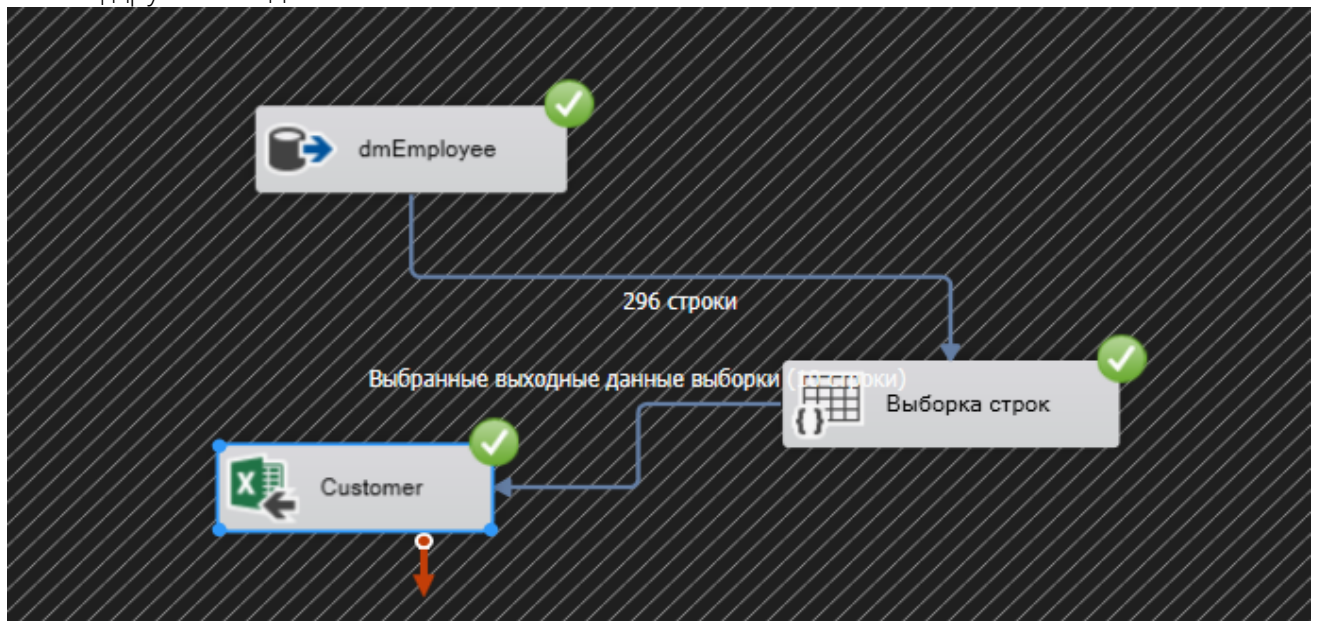
CREATE TABLE Sales.Brand
(
    BrandID INT PRIMARY KEY,
    BrandName VARCHAR (50),
    SupplierID INT FOREIGN KEY (SupplierID) REFERENCES Sales.Supplier (SupplierID)
);

CREATE TABLE Sales.Product
(
    ProductID INT PRIMARY KEY,
    ProductName VARCHAR (100) NOT NULL,
    Price DECIMAL (18,2) DEFAULT 1,
    SubcategoryID INT FOREIGN KEY (SubcategoryID) REFERENCES Sales.Subcategory (SubCategoryID) NOT NULL,
    SupplierID INT FOREIGN KEY (SupplierID) REFERENCES Sales.Supplier (SupplierID) NOT NULL
);

CREATE TABLE Sales.Revenue
(
    ProductID INT FOREIGN KEY (ProductID) REFERENCES Sales.Product (ProductID) NOT NULL,
    CustomerID INT FOREIGN KEY (CustomerID) REFERENCES Sales.Customer (CustomerID) NOT NULL,
    Revenue DECIMAL (18,2),
    Quantity INT CHECK(Quantity>=0) NOT NULL,
    RevenueDate DATE NOT NULL,
);
```

Заполнение таблиц данными

Все таблицы были заполнены через SSIS через источники из интернета либо выгрузкой из таблиц других баз данных.



Тестирование БД

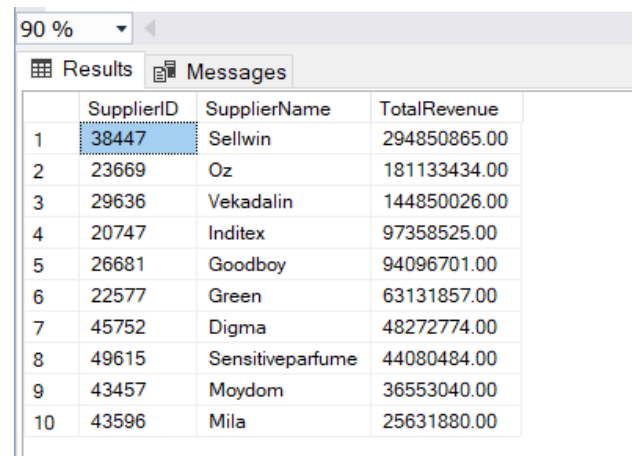
| # | Description | Check | Correction |
|---|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Duplicate records in the Customer table by CustomerID column | SELECT CustomerID, COUNT(*) FROM Sales.Customer GROUP BY CustomerID HAVING COUNT(*) > 1 | WITH CTE AS (SELECT CustomerID, ROW_NUMBER() OVER (PARTITION BY CustomerID ORDER BY CustomerID) AS row_num FROM Sales.Customer) DELETE FROM CTE WHERE row_num > 1; |
| 2 | Mismatched foreign key values in Product table | SELECT SubcategoryID FROM Sales.Product WHERE SubcategoryID NOT IN (SELECT SubcategoryID FROM Sales.Subcategory) | UPDATE Sales.Product SET SubcategoryID = NULL WHERE SubcategoryID NOT IN (SELECT SubcategoryID FROM Sales.Subcategory) |
| 3 | Negative value in the Quantity column in the Revenue table | SELECT * FROM Sales.Revenue WHERE Quantity < 0; | UPDATE Sales.Revenue SET Quantity = 0 WHERE Quantity < 0; |
| 4 | Incorrect values in the Gender column of the Customer table | SELECT * FROM Sales.Customer WHERE Gender NOT IN ('M', 'F'); | UPDATE Sales.Customer SET Gender = NULL WHERE Gender NOT IN ('M', 'F') OR Gender IS NULL; |
| 5 | Product not connected with Supplier | SELECT * FROM Sales.Product WHERE SupplierID IS NULL; | UPDATE Sales.Product SET SupplierID = 'нужное название' |
| 6 | Some meanings are capitalized, some are lowercase | - | UPDATE Sales.Category SET CategoryName = 'нужное название' UPPER(LEFT(CategoryName,1)) + LOWER(SUBSTRING(CategoryName,2, LEN(CategoryName))) |
| 7 | There are no foreign key to link Brand and Supplier | - | ALTER TABLE Sales.Brand ADD CONSTRAINT FK_Brand FOREIGN KEY (BrandID) REFERENCES |
| 8 | The column 'gender2' doesn't contain any valuable data - only NULL. Need to be deleted | - | UPDATE Sales.Customer DROP COLUMN gender2 |

Создание объектов БД

Задание: создание объекта (view) для нахождения общего рейтинга ТОП-10 поставщиков, где рейтинг определяется суммой выручки, которую зарабатывает маркетплейс с каждого поставщика.

```
CREATE VIEW TopCustomers1 AS
SELECT
    TOP 10
    s.SupplierID,
    s.SupplierName,
    SUM(r.Revenue) AS TotalRevenue
FROM
    Sales.Supplier s
JOIN
    Sales.Product p ON s.SupplierID = p.SupplierID
JOIN
    Sales.Revenue r ON p.ProductID = r.ProductID
GROUP BY s.SupplierID, s.SupplierName
ORDER BY TotalRevenue DESC

SELECT *FROM TopCustomers1
```



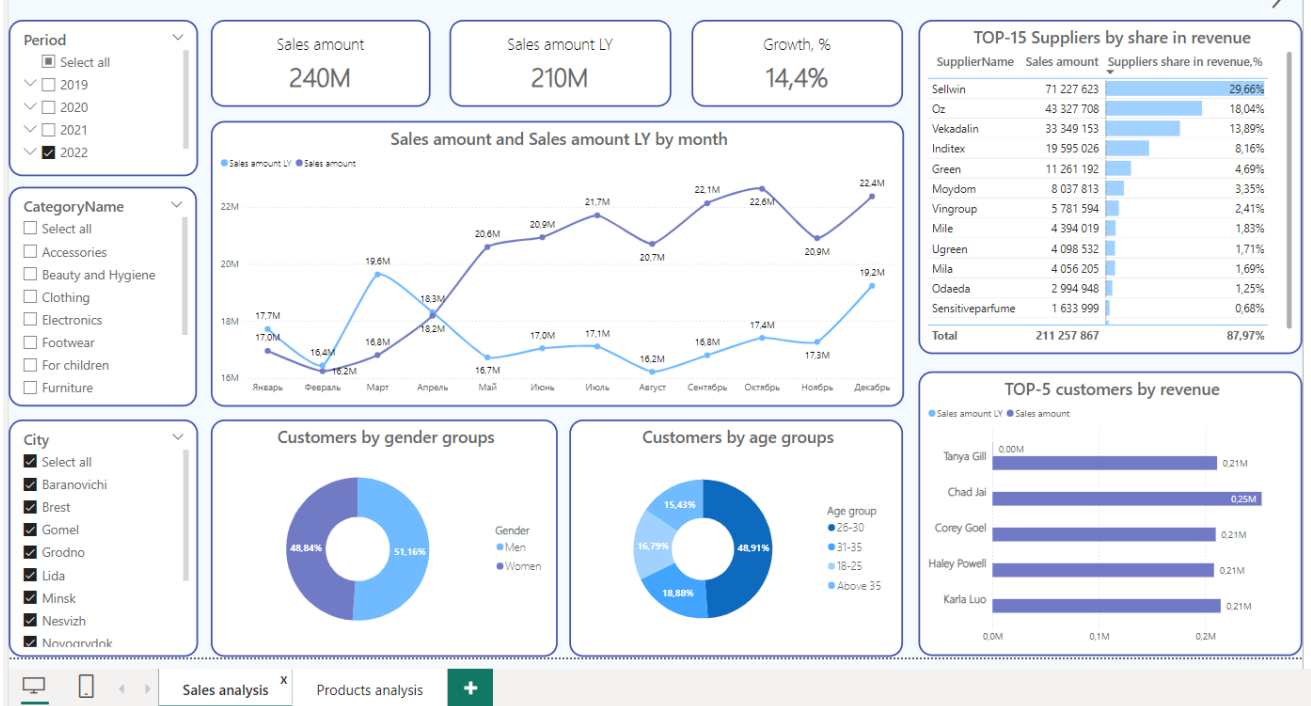
| | SupplierID | SupplierName | TotalRevenue |
|----|------------|------------------|--------------|
| 1 | 38447 | Sellwin | 294850865.00 |
| 2 | 23669 | Oz | 181133434.00 |
| 3 | 29636 | Vekadalin | 144850026.00 |
| 4 | 20747 | Inditex | 97358525.00 |
| 5 | 26681 | Goodboy | 94096701.00 |
| 6 | 22577 | Green | 63131857.00 |
| 7 | 45752 | Digma | 48272774.00 |
| 8 | 49615 | Sensitiveparfume | 44080484.00 |
| 9 | 43457 | Moydom | 36553040.00 |
| 10 | 43596 | Mila | 25631880.00 |

Визуализация данных в Power BI

При создании визуализаций были созданы следующие меры:

| | |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Age Group | = SWITCH (TRUE (), 'Sales Customer'[Age] >= 18 && 'Sales Customer'[Age] <= 25, "18-25", 'Sales Customer'[Age] >= 26 && 'Sales Customer'[Age] <= 30, "26-30", 'Sales Customer'[Age] >= 31 && 'Sales Customer'[Age] <= 35, "31-35", 'Sales Customer'[Age] > 35, "Above 35") |
| Average product price | = DIVIDE (SUM ('Sales Product'[Price]), COUNT ('Sales Product'[ProductID])) |
| Growth, % | = DIVIDE ('Sales Revenue'[Sales amount] - 'Sales Revenue'[Sales amount LY], 'Sales Revenue'[Sales amount LY], 0) |
| Sales amount | = SUM ('Sales Revenue'[Revenue]) |
| Sales amount LY | = CALCULATE ([Sales amount], SAMEPERIODLASTYEAR ('Sales Revenue'[RevenueDate].[Date])) |
| Suppliers share in revenue, % | = DIVIDE ([Sales amount], CALCULATE ([Sales amount], ALL ('Sales Supplier'[SupplierName]))) |

MARKETPLACE. SALES ANALYSIS

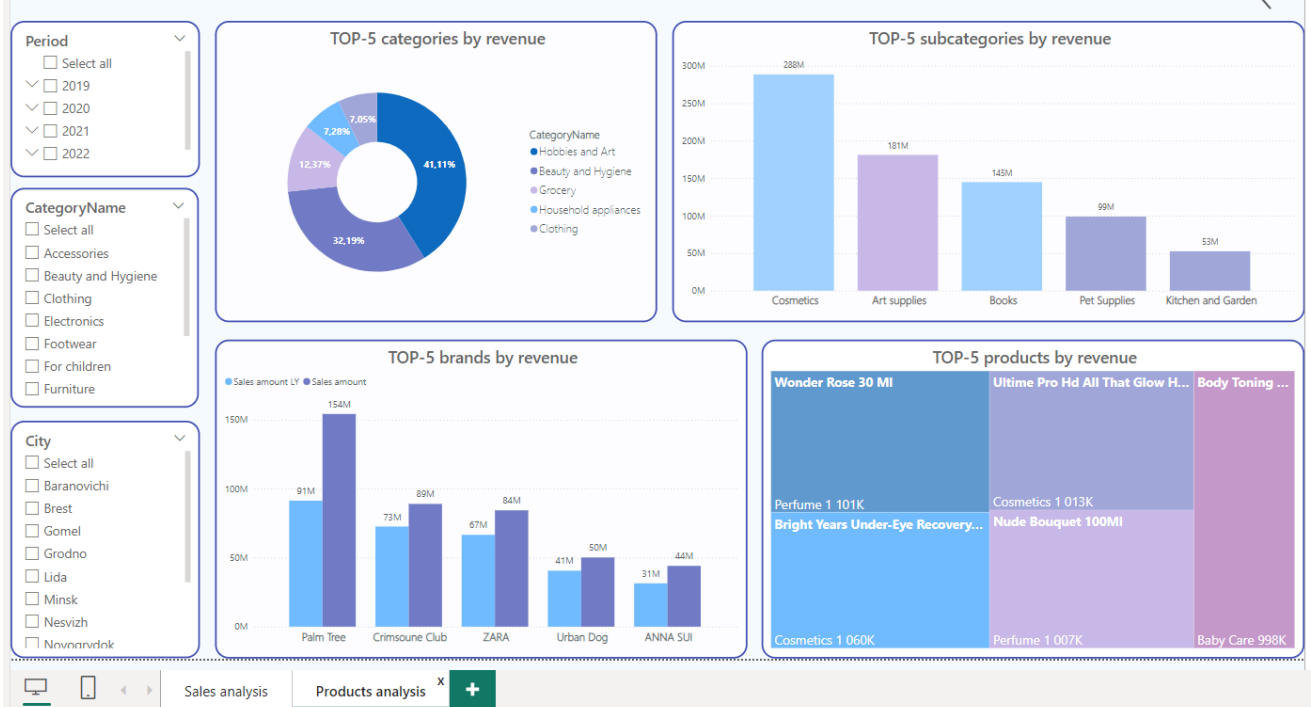


Sales analysis

Products analysis

+

MARKETPLACE. PRODUCTS ANALYSIS



Sales analysis

Products analysis

+