DATABASE DESIGN AND ANALYTICS FOR E-COMMERCE

Tanapat Klomjit & Sweta Kumari

Golden Gate University

Contents

Abstract

E-commerce is undoubtedly not a new trend anymore, but it is now an essential platform that every business needs in order to strive for more growth and future-proof. "Ecommerce is expected to nearly double by 2023 to more than $6.5 billion" (Winkler, 2020). Our primary objective is to prove the significance of relationship database to deal with vast amounts of shop data. Online transaction processing (OLTP) system is its core mechanism regarding its relational database systems because it is focused on transaction-oriented

In this project, database structure in E-commerce for a retail store will be designed, created, and analyzed. Ultimately, the company is able to not only record all transactions logically, but also utilize data to conduct various analyses to gain better understanding about E-commerce platforms, products, customers, and financial performance. At the end, business reports and dynamic dashboard will be created for end-users for various business applications.

Introduction

A database is an organized collection of structured information, or data, typically stored electronically in a computer system (Oracle, n.d.). Database holds the business information which is used to make informed business decisions. The purpose of the study is to understand why Ecommerce should move from file-based database to relational database and how relational database is helpful in business decisions. The file-based databases are complex and difficult to manage. Due to multiple issues faced, E-commerce needs a complete upgrade system. We need a relational database to store the valuable data about the customer, product, shipping details and all transactional information which helps to make informed business decisions (Rojas, 2019).

We need to understand how relational database is helping the E-commerce business. Let's talk about what is a relational database and how it will be good to use for E-commerce.

What is a Relational Database?

A relational database is a type of database that stores and provides access to data points that are related to one another (Oracle, n.d.). It is a collection of data that are stored in a structured format with pre-defined relationships among them. Data is stored in tables and distributed in columns (field) and rows(records). Each table has multiple attributes with keys associated to it. The unique identifier of the table is termed as primary key and the attribute that connects one table to another table is known as foreign key. Since, data is stored in tables, data can be fetched with the help of Structured Query Language (SQL). Databases are accessed, created, managed, modified, and maintained using SQL. This management of relational databases primarily used by SQL is known as a relational database management system (RDBMS).

Why Relational databases should be used in E-commerce?

File-based databases have multiple issues such as disconnected data, mismanaged data, insufficient reporting capability, issues in tracking the transaction, untrusted data, data redundancy, data inconsistency, and data insecurity. To avoid these issues, E-commerce has adopted a relational database. In E-commerce, the main purpose of a relational database is to store information for accessing the product transactions, customer information, sales information, shipping details, etc (Singh, 2017). RDBMS is helping in many ways:

Data Integrity: RDBMS establish the single version of truth by collecting different types of data for E-commerce websites. It also helps in eliminating duplicate data which builds the base of trusted data sources.

Track Transactions: One of the important information in E-commerce is transactional data which consists of selling, purchasing, order information. RDBMS track transactions of each order and manage the data to further analyze sales, inventory and profit.

Provide Data Security: RDBMS provides the data security at different levels. Also, It enables the encryption of data which makes data unreadable to the users who do not have proper access to it.
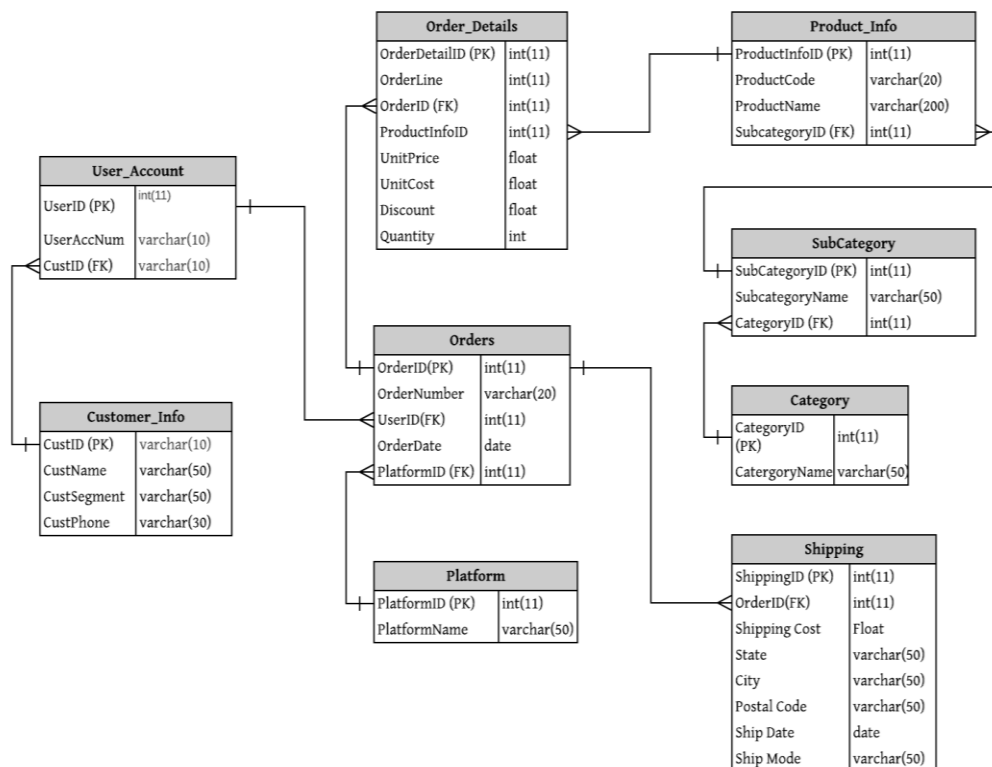
Flexibility: Complex queries can be performed easily in RDBMS. Different tables can be joined and extracted. This feature helps in a big way in E-commerce to make more informed decisions.

Ease Reporting Capabilities: By using RDBMS, E-commerce data is structured, trusted and well managed. It improves the reporting capabilities which are necessary to make informed decisions.

How Relational databases can help in E-commerce?

We already rectified the issue faced in file-based systems. In order to resolve the problem, we used a relational database and started creating the logical diagram of our E-commerce database. We first created the Entity Relationship Diagram for E-commerce.

ER-Diagram: E-commerce Website

Order_Details	
OrderDetailID (PK)	int(11)
OrderLine	int(11)
OrderID (FK)	int(11)
ProductInfoID	int(11)
UnitPrice	float
UnitCost	float
Discount	float
Quantity	int

Product_Info	
ProductInfoID (PK)	int(11)
ProductCode	varchar(20)
ProductName	varchar(200)
SubcategoryID (FK)	int(11)

User_Account	
UserID (PK)	int(11)
UserAccNum	varchar(10)
CustID (FK)	varchar(10)

SubCategory	
SubCategoryID (PK)	int(11)
SubcategoryName	varchar(50)
CategoryID (FK)	int(11)

Orders	
OrderID(PK)	int(11)
OrderNumber	varchar(20)
UserID(FK)	int(11)
OrderDate	date
PlatformID (FK)	int(11)

Category	
CategoryID (PK)	int(11)
CatergoryName	varchar(50)

Customer_Info	
CustID (PK)	varchar(10)
CustName	varchar(50)
CustSegment	varchar(50)
CustPhone	varchar(30)

Platform	
PlatformID (PK)	int(11)
PlatformName	varchar(50)

Shipping	
ShippingID (PK)	int(11)
OrderID(FK)	int(11)
Shipping Cost	Float
State	varchar(50)
City	varchar(50)
Postal Code	varchar(50)
Ship Date	date
Ship Mode	varchar(50)

Data Source & Tool for data visualization

We collected data from the Kaggle website: https://www.kaggle.com/jr2ngb/superstore-data. The dataset contains retail data of a global superstore for 4 years from 2011 to 2014. Data contains different attributes as shown in our ERD after we completed normalization. We will primarily use Microsoft Power BI to generate graphs, charts, and dashboards.

- **Customer_Info:** In our E-commerce platform, customers have their name, segment and phone number. Segment determines which type of customer they are. For instance: they belong to the consumer segment or corporate segment.

- **User_Account:** Each customer can have multiple user accounts to order the products from the E-commerce site.

- **Platform:** We also try to analyze which platform is used for ordering the product from the website. It could be iOSApp, Web etc.

- **Category:** Well, our E-commerce has different product categories: "Furniture", "Technology" and "Office Supplies".

- **SubCategory:** Each category has subcategories and belongs to different product groups. For instance: "Chairs", "Tables".

- **Product_Info:** Each subcategory then belongs to specific products. For example: "Canon mageCLASS 2200 Advanced Copier".

- **Orders:** Users can order the products which are shipped to their home address. E-commerce captures information about the order date, order number and the platform used for order.

- **Order_Details:** They can order multiple products at one go. Details are captured with order line, discount, quantity, unit cost, unit price and product information.

- **Shipping:** When users order a product, they are shipped to their respective address. Shipping cost, Shipping Date, City, State, Postal Code, Ship Mode and order information are captured.

We create tables mentioned in the ER Diagram for our E-commerce website. We insert

the data through SQL scripts to capture transactional details. Using queries, we analyze the data

and address several business questions which helps to make more informed business decisions.

Business Questions

- We analyze sales performance for the past 4 years by months and years to see any trend.

- There are various platforms for customers to press an order via E-commerce. We analyze

  different platforms: IOS App, Android App, and web browser.

- What are the total profits for each year based on the product category? What are the sales

  and number of orders of those categories?

- What are top-selling and slow products to see which products are contributing more or

  least to profit?

- Find top customers based on profit. What are the sales and number of orders associated

  with it?

- Which customer segments are profitable each year? What are the sales numbers for each

  segment?

- What are the top 10 states that have the maximum number of sales?

Table Creation

As part of the E-commerce database, we created below tables as per our logical structure mentioned in the ER diagram.

Customer_Info

```
1  CREATE TABLE Customer_Info (CustID varchar(10) Primary Key Not Null,
2  CustName varchar(50) DEFAULT Null,
3  CustSegment varchar(50) DEFAULT Null,
4  CustPhone varchar(30) DEFAULT Null);
```

Column	Type	Null	Default	References	ParentID
CustID (Primary)	varchar(10)	No			
CustName	varchar(50)	Yes	NULL		
CustSegment	varchar(50)	Yes	NULL		
CustPhone	varchar(30)	Yes	NULL		

User_Account

```
1  CREATE TABLE User_Account (UserID int Primary Key Not Null AUTO_INCREMENT,
2  UserAccNum varchar(10) DEFAULT Null,
3  CustID varchar(10) Not Null,
4  FOREIGN KEY (CustID) REFERENCES Customer_Info(CustID));
```

Column	Type	Null	Default	References	ParentID
UserID (Primary)	int(11)	No			
UserAccNum	varchar(10)	Yes	NULL		
CustID	varchar(10)	No		Customer_Info	CustID

Platform

```
1  CREATE TABLE Platform (PlatformID int Primary Key Not Null AUTO_INCREMENT,
2  PlatformName varchar(50) DEFAULT Null);
```

Column	Type	Null	Default	References	ParentID
PlatformID (Primary)	int(11)	No			
PlatformName	varchar(50)	Yes	NULL		

Category

```
1  CREATE TABLE Category(CategoryID int Primary Key Not Null AUTO_INCREMENT,
2  CategoryName varchar(50) DEFAULT Null);
```

Column	Type	Null	Default	References	ParentID
CategoryID (Primary)	int(11)	No			
CategoryName	varchar(50)	Yes	NULL		

SubCategory

```
1  CREATE TABLE SubCategory(SubCategoryID int Primary Key Not Null AUTO_INCREMENT,
2  SubCategoryName varchar(50) DEFAULT Null,
3  CategoryID int Not Null,
4  FOREIGN KEY (CategoryID) REFERENCES Category(CategoryID));
```

Column	Type	Null	Default	References	ParentID
SubCategoryID (Primary)	int(11)	No			
SubCategoryName	varchar(50)	Yes	NULL		
CategoryID	int(11)	No		Category	CategoryID

Product_Info

```
1  CREATE TABLE Product_Info
2  (ProductInfoID int Primary Key Not Null AUTO_INCREMENT,
3  ProductCode varchar(30) Not Null,
4  ProductName varchar(200) DEFAULT Null,
5  SubCategoryID int Not Null,
6  FOREIGN KEY (SubCategoryID) REFERENCES SubCategory(SubCategoryID));
```

Column	Type	Null	Default	References	ParentID
ProductInfoID (Primary)	int(11)	No			
ProductCode	varchar(30)	No			
ProductName	varchar(200)	Yes	NULL		
SubCategoryID	int(11)	No		SubCategory	SubCategoryID

Orders

```
1  CREATE TABLE Orders
2  (OrderID int Primary Key Not Null AUTO_INCREMENT,
3   OrderNumber varchar(20) Not Null,
4  OrderDate date DEFAULT Null,
5  UserID int Not Null,
6  FOREIGN KEY (UserID) REFERENCES User_Account(UserID),
7  PlatformID int Not Null,
8  FOREIGN KEY (PlatformID) REFERENCES Platform(PlatformID));
```

Column	Type	Null	Default	References	ParentID
OrderID (Primary)	int(11)	No			
OrderNumber	varchar(20)	No			
OrderDate	date	Yes	NULL		
UserID	int(11)	No		User_Account	UserID
PlatformID	int(11)	No		Platform	PlatformID

Order_Details

```
 1  CREATE TABLE Order_Details
 2  (OrderDetailID int Primary Key Not Null AUTO_INCREMENT,
 3  OrderLine float DEFAULT Null,
 4  OrderID int Not Null,
 5  FOREIGN KEY (OrderID) REFERENCES Orders(OrderID),
 6  ProductID int Not Null,
 7  FOREIGN KEY (ProductID) REFERENCES Product_Info(ProductID),
 8  UnitPrice float DEFAULT Null,
 9  UnitCost float DEFAULT Null,
10  Discount float DEFAULT Null,
11  Quantity int DEFAULT Null);
```

Column	Type	Null	Default	References	ParentID
OrderDetailID (Primary)	int(11)	No			
OrderLine	int(11)	No			
OrderID	int(11)	No		Orders	OrderID
ProductInfoID	int(11)	No		Product_Info	ProductInfoID
UnitPrice	float	Yes	NULL		
UnitCost	float	Yes	NULL		
Discount	float	Yes	NULL		
Quantity	int(11)	Yes	NULL		

Shipping

```
 1  CREATE TABLE Shipping
 2  (ShippingID int Primary Key Not Null AUTO_INCREMENT,
 3  ShippingCost float DEFAULT Null,
 4  City varchar(50) DEFAULT Null,
 5  State varchar(50) DEFAULT Null,
 6  PostalCode varchar(50) DEFAULT Null,
 7  ShipDate Date DEFAULT Null,
 8  ShipMode varchar(50) DEFAULT Null,
 9  Region varchar(50) DEFAULT Null,
10  OrderID int Not Null,
11  FOREIGN KEY (OrderID) REFERENCES Orders(OrderID));
```

Column	Type	Null	Default	References	ParentID
ShippingID (Primary)	int(11)	No			
ShippingCost	float	Yes	NULL		

City	varchar(50)	Yes	NULL		
State	varchar(50)	Yes	NULL		
PostalCode	varchar(50)	Yes	NULL		
ShipDate	date	Yes	NULL		
ShipMode	varchar(50)	Yes	NULL		
Region	varchar(50)	Yes	NULL		
OrderID	int(11)	No		Orders	OrderID

Inserting Data

We inserted data into the tables with the help of insert scripts. This is not a full insert script. Due

to the large number of rows in a few tables, we have shown only a few rows of insertion into the

table with respective screenshots.

Customer_Info

Number of rows:793

```
INSERT INTO Customer_Info (CustID, CustName, CustSegment, CustPhone) VALUES
('DK-13375' , 'Dennis Kane', 'Consumer' , '(951) 470-6811'),
('EB-13705' , 'Ed Braxton', 'Corporate' , '(228) 666-5431'),
('NF-18475' , 'Neil Französisch', 'Home Office' , '(548) 592-9712');
```

CustID	CustName	CustSegment	CustPhone
AA-10315	Alex Avila	Consumer	(683) 917-0322
AA-10375	Allen Armold	Consumer	(808) 452-5235
AA-10480	Andrew Allen	Consumer	(242) 202-7314

User_Account

Number of rows: 793

```
INSERT INTO User_Account (UserAccNum, CustID) VALUES
('11146930' , 'DK-13375'),
('11517482' , 'EB-13705'),
('11798125' , 'NF-18475');
```

UserID	UserAccNum	CustID
1	11146930	DK-13375
2	11517482	EB-13705
3	11798125	NF-18475

Platform

Number of rows: 3

```
INSERT INTO Platform (PlatformName) VALUES
('IOSApp'),
('AndriodAPP'),
('Web');
```

PlatformID	PlatformName
1	IOSApp
2	AndriodAPP
3	Web

Category

Number of rows:3

```
INSERT INTO Category (CategoryName) VALUES
("Furniture"),
("Office Supplies"),
("Technology");
```

CategoryID	CategoryName
1	Furniture
2	Office Supplies
3	Technology

SubCategory

Number of rows: 17

```
INSERT INTO SubCategory (SubCategoryName, CategoryID) VALUES
("Bookcases",1),
("Chairs",1),
("Tables",1);
```

SubCategoryID	SubCategoryName	CategoryID
1	Bookcases	1
2	Chairs	1
3	Tables	1

Product_Info

Number of rows: 1842

```
INSERT INTO Product_Info (ProductCode, ProductName, SubCategoryID) VALUES
('FUR-BO-10000112' , 'Bush Birmingham Collection Bookcase, Dark Cherry' , '1'),
('FUR-BO-10000330' , 'Sauder Camden County Barrister Bookcase, Planked Cherry Finish' , '1'),
('FUR-BO-10000362' , 'Sauder Inglewood Library Bookcases' , '1');
```

ProductInfoID	ProductCode	ProductName	SubCategoryID
1	FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry	1
2	FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked C...	1
3	FUR-BO-10000362	Sauder Inglewood Library Bookcases	1

Orders

Number of rows: 5009

```
INSERT INTO Orders (OrderNumber, UserID, OrderDate, PlatformID) VALUES
('CA-2011-103800' , '55', '2011-1-4' , '2'),
('CA-2011-112326' , '151', '2011-1-5' , '1'),
('CA-2011-141817' , '400', '2011-1-6' , '1');
```

OrderID	OrderNumber	OrderDate	UserID	PlatformID
1	CA-2011-103800	2011-01-04	55	2
2	CA-2011-112326	2011-01-05	151	1
3	CA-2011-141817	2011-01-06	400	1

Order_Details

Number of rows:9994

```
INSERT INTO Order_Details(OrderLine,OrderID,ProductInfoID,UnitPrice,UnitCost, Discount, Quantity) VALUES
(1000,485,1782,125.99,89.4529,0,3),
(2000,346,366,167.496,196.8078,0.2,3),
(1000,346,710,32.784,21.3096,0.2,6);
```

OrderDetailID	OrderLine	OrderID	ProductInfoID	UnitPrice	UnitCost	Discount	Quantity
1	1000	485	1782	125.99	89.4529	0	3
2	2000	346	366	167.496	196.808	0.2	3
3	1000	346	710	32.784	21.3096	0.2	6

Shipping

Number of rows: 5009

```
INSERT INTO Shipping (OrderID, ShippingCost, City, State, PostalCode, ShipDate, ShipMode, Region) VALUES
(485,11.76,"New York City", "New York","10024","2011-09-13","Standard Class", "East"),
(346,45.05,"San Francisco","California","94122","2011-07-12","Standard Class", "West"),
(85,3.22,"Jacksonville","Florida","32216","2011-03-18","Standard Class", "South");
```

ShippingID	ShippingCost	City	State	PostalCode	ShipDate	ShipMode	Region	OrderID
1	11.76	New York City	New York	10024	2011-09-13	Standard Class	East	485
2	45.05	San Francisco	California	94122	2011-07-12	Standard Class	West	346
3	3.22	Jacksonville	Florida	32216	2011-03-18	Standard Class	South	85

Analyzing the data

Formula used for sales and profit calculations:

Sales = Quantity*Unit Price

Profit = (UnitPrice - UnitCost) * Quantity

- We analyze sales performance for the past 4 years by months and years to see any trend.

Total Sales by Year

```
1  CREATE VIEW Total_Sales_by_Year AS
2  SELECT  year(Orders.OrderDate) as Year,
3          CONCAT('$', FORMAT(sum(Order_Details.Quantity * Order_Details.UnitPrice), 2)) as 'Total Sales'
4  FROM `Orders` JOIN Order_Details ON Orders.OrderID = Order_Details.OrderID
5  GROUP BY year(Orders.OrderDate);
```

Year	Total Sales
2011	$484,247.50
2012	$470,487.48
2013	$609,126.51
2014	$735,061.37

To understand how the sales performance was for all years, we created a view to see the result. It has helped to know if sales were growing every year for our E-commerce business. Sales were decreased in 2012 and again increased till 2014.

E-Commerce: Total Sales By Year
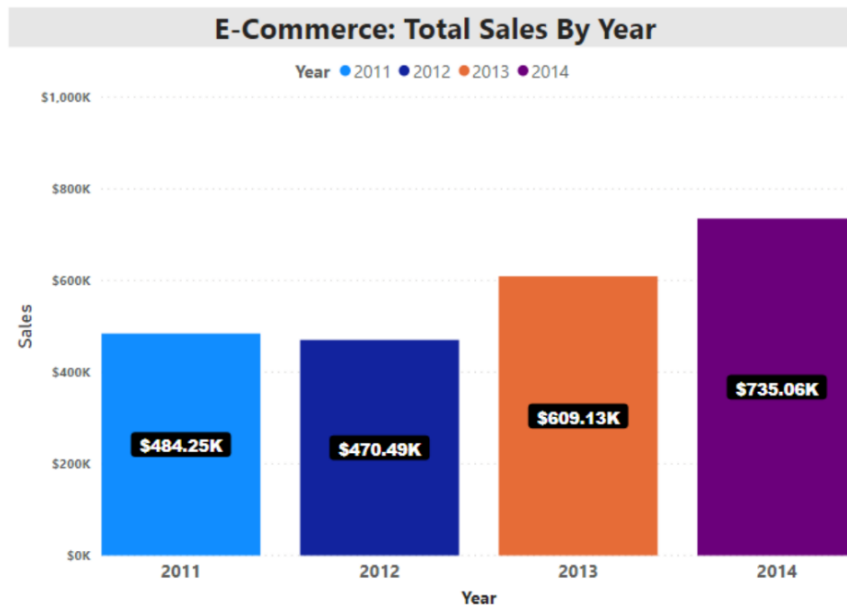
## Total Sales by Month and Year

```
1  Create view Total_Sales_by_Month_and_Year as
2  SELECT  year(Orders.OrderDate) as 'Year' ,
3          month(Orders.OrderDate) as 'Month',
4          FORMAT(sum(Order_Details.Quantity * Order_Details.UnitPrice), 2)) as 'Total Sales'
5  FROM `Orders`
6  JOIN Order_Details ON   Orders.OrderID = Order_Details.OrderID
7  GROUP BY year(Orders.OrderDate) , month(Orders.OrderDate);
```

Year	Month	Total Sales	Year	Month	Total Sales
2011	1	$13,946.23	2013	1	$18,542.49
2011	2	$4,810.56	2013	2	$22,867.71
2011	3	$55,691.01	2013	3	$51,460.30
2011	4	$28,295.35	2013	4	$39,365.38
2011	5	$23,648.29	2013	5	$56,691.08
2011	6	$34,595.13	2013	6	$39,430.44
2011	7	$33,946.39	2013	7	$38,440.76
2011	8	$27,909.47	2013	8	$33,265.56
2011	9	$81,777.35	2013	9	$72,883.55
2011	10	$31,453.39	2013	10	$56,463.13
2011	11	$78,628.72	2013	11	$82,192.32
2011	12	$69,545.62	2013	12	$97,523.80
2012	1	$18,174.08	2014	1	$44,703.14
2012	2	$12,210.87	2014	2	$20,283.51
2012	3	$38,466.80	2014	3	$53,908.96
2012	4	$34,195.21	2014	4	$40,112.42
2012	5	$30,131.69	2014	5	$45,651.24
2012	6	$24,797.29	2014	6	$48,259.75
2012	7	$28,720.30	2014	7	$48,428.37
2012	8	$36,898.33	2014	8	$61,516.09
2012	9	$64,595.92	2014	9	$90,488.72
2012	10	$31,404.92	2014	10	$77,793.76
2012	11	$75,972.56	2014	11	$113,415.23
2012	12	$74,919.52	2014	12	$90,500.18

To drill down more on the sales performance, we created a view to see the result by month and

year. It has helped to know if sales are growing every month for our E-commerce business. Sales

were growing every year except 2012. Interestingly, we also we a seasonable sales trend when

the sale is low during February and peak in November for the past 4 years.
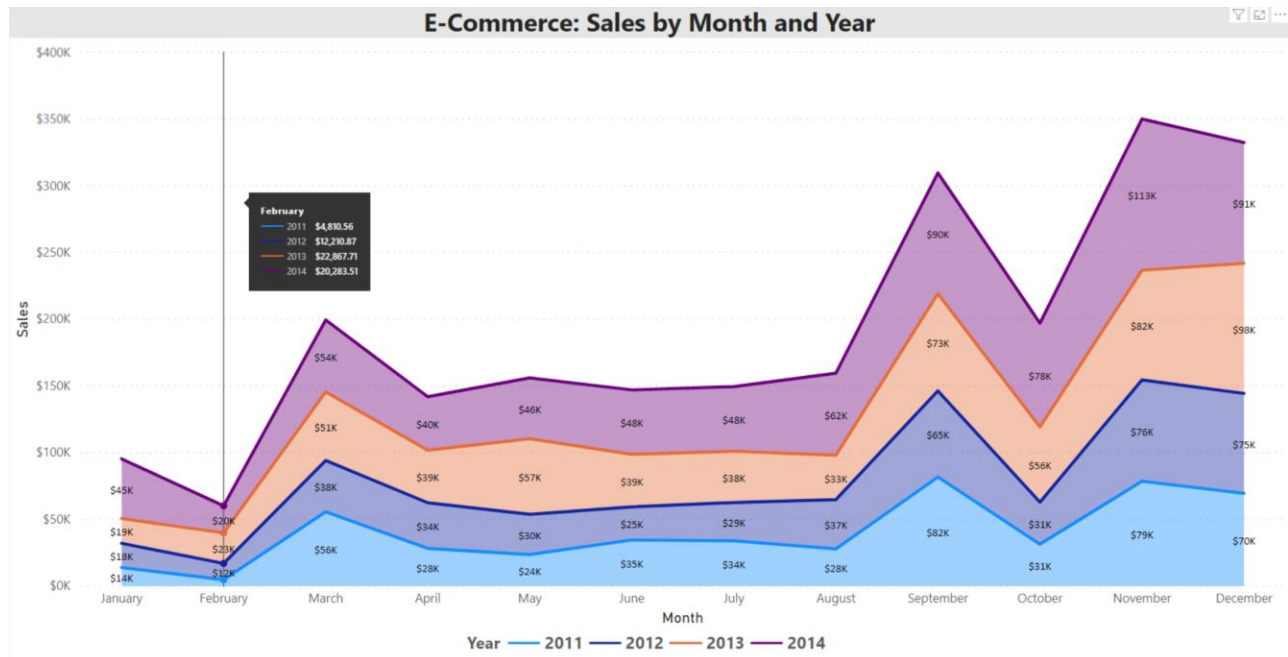
- There are various platforms for customers to press an order via E-commerce. We analyze

    different platforms: IOS App, Android App, and web browser.

Total sales and order counts by various E-commerce platforms

```
1  Create view Total_Sales_and_Order_Count_by_Platform AS
2  SELECT  year(Orders.OrderDate) as 'Year',
3          Platform.PlatformName as 'Platform',
4          count(Orders.OrderID) as 'Number of Orders',
5          FORMAT(sum(Order_Details.Quantity * Order_Details.UnitPrice), 2)) as 'Total Sales'
6  FROM `Orders`
7  JOIN Order_Details  ON Orders.OrderID =  Order_Details.OrderID
8  JOIN  Platform ON Platform.PlatformID = Orders.PlatformID
9  GROUP BY year(Orders.OrderDate), Platform.PlatformName;
```

Year	Platform	Number of Orders	Total Sales
2011	AndriodAPP	173	$33,634.10
2011	IOSApp	1593	$408,522.58
2011	Web	227	$42,090.81
2012	AndriodAPP	139	$37,041.10
2012	IOSApp	1284	$273,683.96
2012	Web	679	$159,762.42
2013	AndriodAPP	1301	$321,679.21
2013	IOSApp	1090	$241,636.71
2013	Web	189	$45,810.60
2014	AndriodAPP	107	$39,885.84
2014	IOSApp	3212	$695,175.52

Since our E-commerce uses 3 different platforms to order the products, we created a view to see how each platform was used to order the products. This has given ideas about if E-commerce business should be accessed by all three platforms or if any platforms will need attention by our management team. We have no revenue from the Web platform at all in 2014. The usage of IOSApp is increasing substantially from 2013 to 2014.
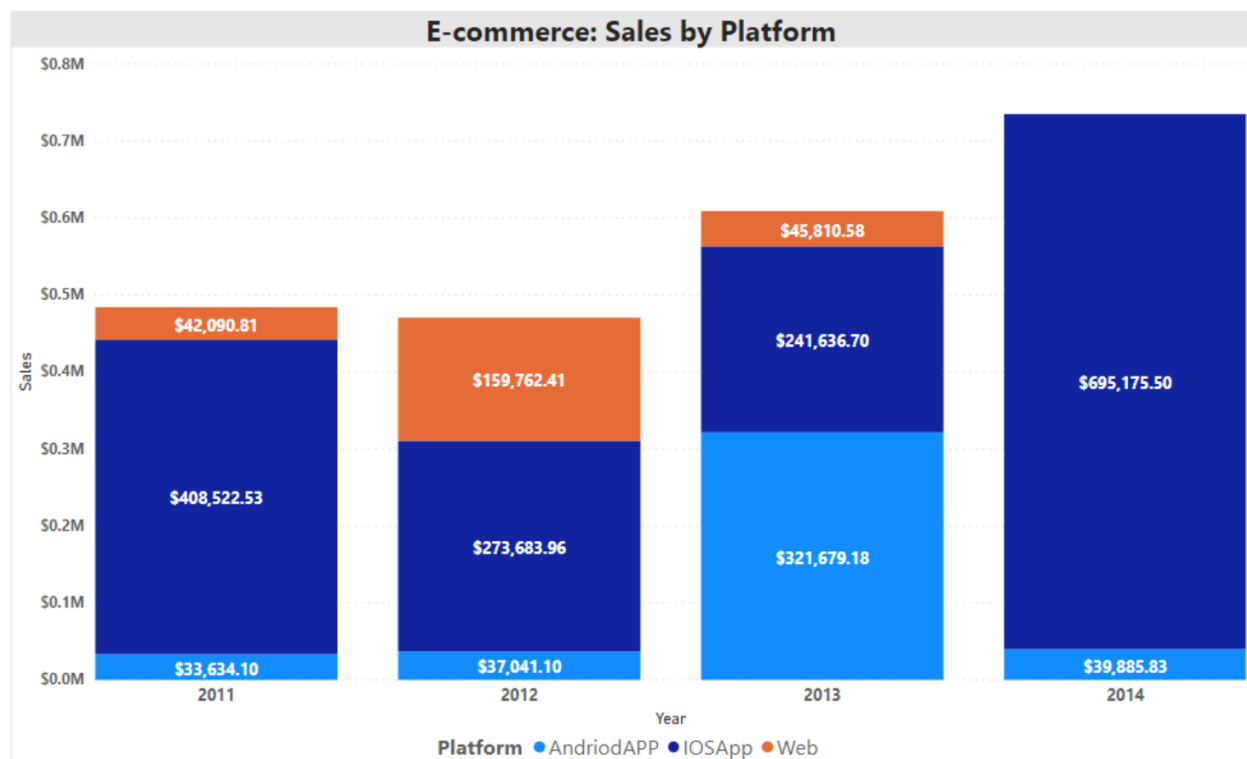
- What are the total profits for each year based on the product category? What are the sales and number of orders of those categories?

Total Profit by Year and Product Category

```
 1  Create view Total_Profit_by_Year_and_Category AS
 2  SELECT  year(a.OrderDate) as 'Year',
 3          b.CategoryName AS "Product Category",
 4          count(a.OrderID) as 'Number of Orders',
 5          CONCAT('$', FORMAT(sum(c.Quantity * c.UnitPrice), 2)) as 'Sales' ,
 6          CONCAT('$', FORMAT(sum((c.UnitPrice - c.UnitCost) * c.Quantity),2)) as 'Profit'
 7  FROM Orders a JOIN Order_Details c ON a.OrderID = c.OrderID
 8  JOIN Product_Info e ON e.ProductInfoID = c.ProductInfoID
 9  JOIN SubCategory d ON d.SubCategoryID = e.SubCategoryID
10  JOIN Category b ON b.CategoryID = d.CategoryID
11  GROUP BY b.CategoryName, year(a.OrderDate)
12  ORDER BY year(a.OrderDate), round(sum((c.UnitPrice - c.UnitCost) * c.Quantity),2) DESC
```

Year	Product Category	Number of Orders	Sales	Profit
2011	Office Supplies	1217	$151,776.41	$22,593.42
2011	Technology	355	$175,278.23	$21,492.83
2011	Furniture	421	$157,192.85	$5,457.73
2012	Technology	409	$162,780.81	$33,503.87
2012	Office Supplies	1241	$137,188.43	$25,078.82
2012	Furniture	452	$170,518.24	$3,015.20
2013	Technology	458	$226,178.59	$39,773.21
2013	Office Supplies	1560	$183,760.11	$35,130.20
2013	Furniture	562	$199,187.82	$7,043.00
2014	Technology	625	$272,058.77	$50,710.45
2014	Office Supplies	2008	$247,615.33	$40,097.93
2014	Furniture	686	$215,387.27	$3,018.39

Since our E-commerce has 3 different product categories, we created a view to see how each product category has contributed towards profit. This has given ideas about if E-commerce business should focus on specific product categories to increase the revenue of the company. We see that the Furniture has the least profit. Technology is the most profitable category every year except in 2011.
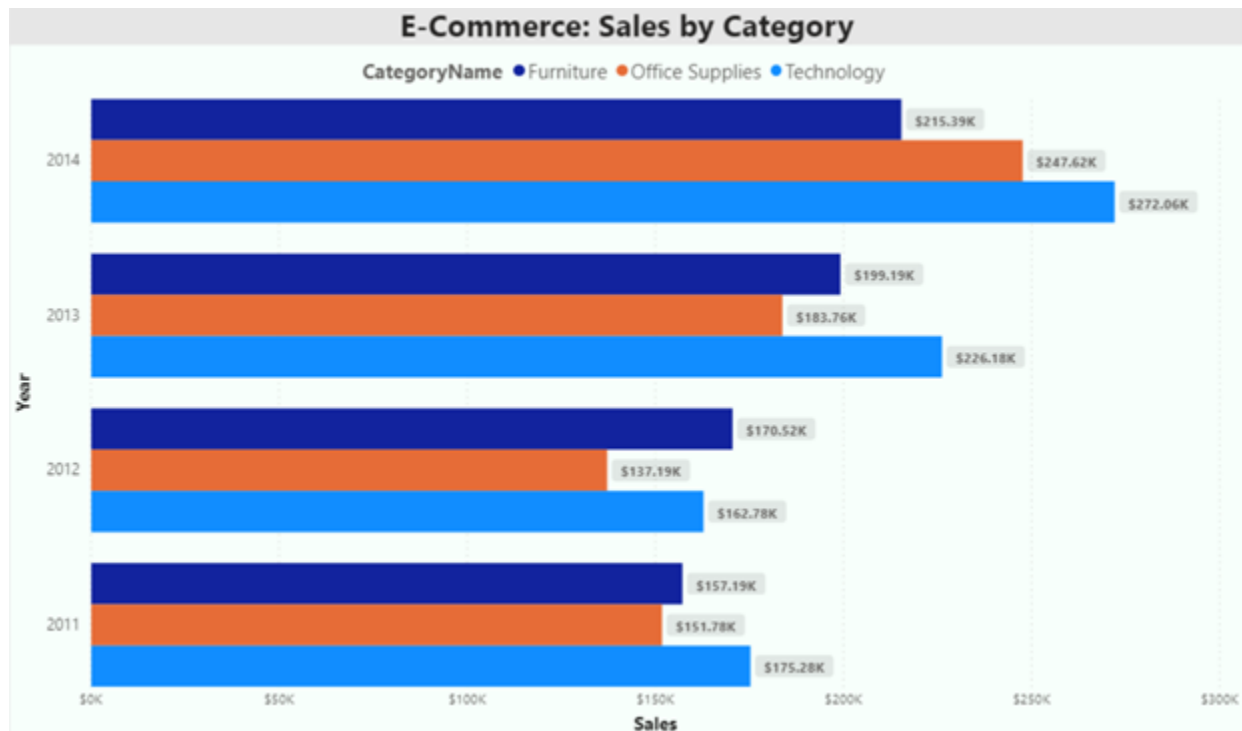
- Which customer segments are profitable each year? What are the sales numbers for each

  segment?

**View: Customer Segment with Profit each year**

```
1  Create view Segment_Profit_Each_Year AS
2  SELECT   a.CustSegment as 'Segment',
3           year(b.OrderDate) as 'Year', count(b.OrderID) as 'Number of Orders',
4           CONCAT('$', FORMAT(sum((c.UnitPrice - c.UnitCost) * c.Quantity),2)) as 'Profit',
5           CONCAT('$', FORMAT(sum(c.Quantity * c.UnitPrice), 2)) as 'Sales'
6  FROM Orders b JOIN Order_Details c ON b.OrderID = c.OrderID
7  JOIN  User_Account d ON d.UserID = b.UserID JOIN Customer_Info a ON a.CustID = d.CustID
8  GROUP BY year(b.OrderDate), a.CustSegment
9  ORDER BY year(b.OrderDate),round(sum(( c.UnitPrice - c.UnitCost) * c.Quantity),2) DESC
```

Segment	'Year'	Number of Orders	Profit	Sales
Consumer	2011	1070	$24,319.85	$266,096.81
Corporate	2011	611	$13,513.28	$128,434.87
Home Office	2011	312	$11,710.85	$89,715.81
Consumer	2012	1125	$28,460.17	$266,535.93
Corporate	2012	636	$20,688.32	$128,757.31
Home Office	2012	341	$12,449.40	$75,194.24
Consumer	2013	1323	$35,841.32	$296,581.92
Corporate	2013	791	$30,939.61	$206,942.95
Home Office	2013	466	$15,165.48	$105,601.64
Consumer	2014	1673	$45,584.44	$332,498.64
Corporate	2014	982	$27,153.66	$243,099.99
Home Office	2014	664	$21,088.67	$159,462.73

To know how each customer segment has contributed to profit, we created a view to see the customer segment with, number of orders, profit and its sales. This has given ideas about which segment is profitable for the company. E-commerce business aim to improve profit and revenue based on these types of reports. We see that "Consumer" has contributed most to the profit for all four years. Whereas the corporate is the only segment that has its profit declines in 2014 comparing to another two segments.
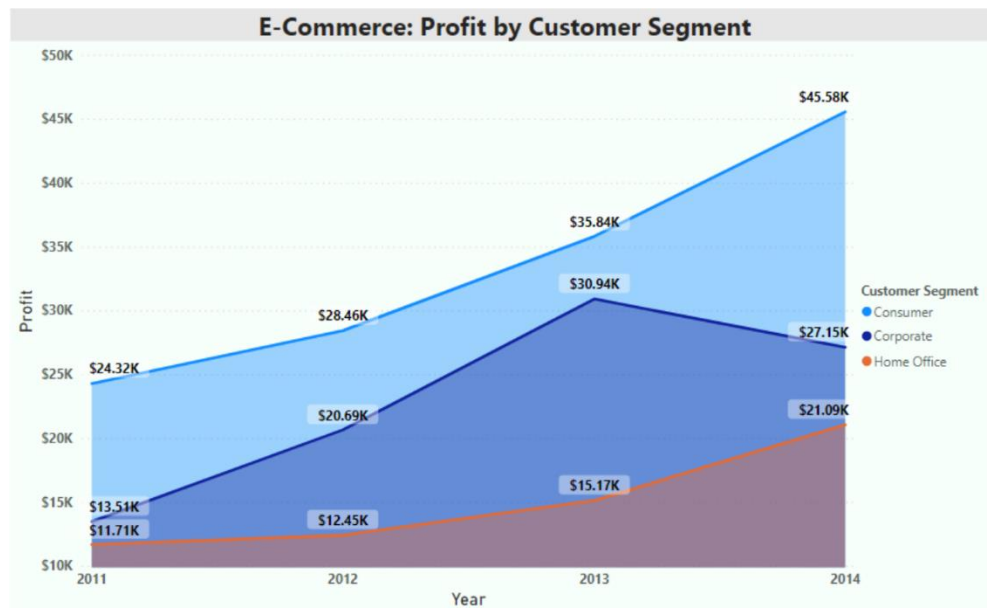
- What are top-selling and slow products to see which products are contributing more or least

  to profit?

Top 20 Product with Highest Profit

```
1  Create view Top_20_Product_with_Highest_Profit AS
2  SELECT  c.ProductName,
3          CONCAT('$', FORMAT(sum((b.UnitPrice - b.UnitCost) * b.Quantity),2)) as 'Profit',
4          CONCAT('$', FORMAT(sum(b.Quantity * b.UnitPrice), 2)) as 'Sales'
5  FROM Product_Info c
6  JOIN Order_Details b ON b.ProductInfoID = c.ProductInfoID
7  JOIN Orders a ON a.OrderID = b.OrderID
8  GROUP BY c.ProductName
9  ORDER BY round(sum(( b.UnitPrice - b.UnitCost) * b.Quantity),2) DESC limit 20
```

ProductName	Profit	Sales
Canon imageCLASS 2200 Advanced Copier	$25,199.93	$61,599.82
Fellowes PB500 Electric Punch Plastic Comb Binding...	$7,753.04	$27,453.38
Hewlett Packard LaserJet 3310 Copier	$6,983.88	$18,839.69
Canon PC1060 Personal Laser Copier	$4,570.93	$11,619.83
Plantronics Savi W720 Multi-Device Wireless Headse...	$4,425.34	$13,756.54
HP Designjet T520 Inkjet Large Format Printer - 24...	$4,094.98	$18,374.89
Ativa V4110MDD Micro-Cut Shredder	$3,772.95	$7,699.89
3D Systems Cube Printer, 2nd Generation, Magenta	$3,717.97	$14,299.89
Ibico EPK-21 Electric Binding System	$3,345.28	$15,875.92
Zebra ZM400 Thermal Label Printer	$3,343.54	$6,965.70
Honeywell Enviracaire Portable HEPA Air Cleaner fo...	$3,247.02	$11,304.44
Hewlett Packard 610 Color Digital Copier / Printer	$3,124.94	$8,899.82
Plantronics CS510 - Over-the-Head monaural Wireles...	$3,085.03	$10,822.36
Canon Imageclass D680 Copier / Fax	$2,799.96	$8,959.87
Staples	$2,611.07	$7,008.20
Fellowes PB300 Plastic Comb Binding Machine	$2,518.05	$8,070.19
Ibico Ibimaster 300 Manual Binding System	$2,318.34	$7,985.38
Adjustable Depth Letter/Legal Cart	$2,248.29	$8,818.96
Logitech Z-906 Speaker sys - home theater - 5.1-CH	$2,243.93	$5,609.83
GBC DocuBind TL300 Electric Binding System	$2,233.50	$19,823.48

To know which products have contributed more in profit, we created a view to see top 20 products with profit and its sales. This has given ideas about if E-commerce business should focus on specific products to increase the revenue of the company. We see that the "Canon mageCLASS 2200 Advanced Copier" has contributed most in the profit.

Top 20 Product with lowest profit

```sql
Create view Top_20_Product_with_lowest_Profit AS
SELECT   c.ProductName,
         CONCAT('$', FORMAT(sum((b.UnitPrice - b.UnitCost) * b.Quantity),2)) as 'Profit',
         CONCAT('$', FORMAT(sum(b.Quantity * b.UnitPrice), 2)) as 'Sales'
FROM Product_Info c
JOIN Order_Details b ON b.ProductInfoID = c.ProductInfoID
JOIN Orders a ON a.OrderID = b.OrderID
GROUP BY c.ProductName
ORDER BY round(sum(( b.UnitPrice - b.UnitCost) * b.Quantity),2) limit 20
```

ProductName	Profit	Sales
Cubify CubeX 3D Printer Double Head Print	$-8,879.97	$11,099.96
Lexmark MX611dhe Monochrome Laser Printer	$-4,589.97	$16,829.90
Cubify CubeX 3D Printer Triple Head Print	$-3,839.99	$7,999.98
Chromcraft Bull-Nose Wood Oval Conference Tables &...	$-2,876.12	$9,917.64
Bush Advantage Collection Racetrack Conference Tab...	$-1,934.40	$9,544.72
GBC DocuBind P400 Electric Binding System	$-1,878.17	$17,965.07
Cisco TelePresence System EX90 Videoconferencing U...	$-1,811.08	$22,638.48
Martin Yale Chadless Opener Electric Letter Opener	$-1,299.18	$16,656.20
Balt Solid Wood Round Tables	$-1,201.06	$6,518.75
BoxOffice By Design Rectangular and Half-Moon Meet...	$-1,148.44	$1,706.25
Riverside Furniture Oval Coffee Table, Oval End Ta...	$-1,147.40	$4,446.18
Epson TM-T88V Direct Thermal Printer - Monochrome ...	$-1,057.23	$1,212.71
Hon 2090 "Pillow Soft" Series Mid Back Swivel/Tilt...	$-989.05	$5,282.42
OSullivan 4-Shelf Bookcase in Odessa Pine	$-975.10	$2,740.20
Bretford "Just In Time" Height-Adjustable Multi-Ta...	$-964.19	$5,634.90
Zebra GK420t Direct Thermal/Thermal Transfer Print...	$-938.28	$703.71
Swingline SM12-08 MicroCut Jam Free Shredder	$-882.12	$5,204.24
3.6 Cubic Foot Counter Height Office Refrigerator	$-872.08	$2,946.20
Bevis Oval Conference Table, Walnut	$-856.01	$6,942.07
Tennsco Single-Tier Lockers	$-825.75	$9,158.30

To know which products have contributed least in profit, we created a view to see top 20 products with profit and its sales. This has given ideas about if E-commerce business should focus on specific products to remove from the company or make improvements in the product selling. We see that the "Cubify CubeX Printer Double Head Print" has contributed least in the profit.

- Find top customers based on profit.

Top 20 customer with highest profit

```
1  Create view Top_20_Customer_with_Highest_Profit AS
2  SELECT  Customer_Info.CustName as 'Customer',
3          Customer_Info.CustSegment as 'Segment',
4          count(Orders.OrderID) as 'Number of Orders',
5          round(sum(Order_Details.Quantity * Order_Details.UnitPrice),2) AS 'SALES',
6          round(sum(( Order_Details.UnitPrice - Order_Details.UnitCost) * Order_Details.Quantity),2) AS 'Profit'
7  FROM Orders
8  JOIN  Order_Details on Orders.OrderID = Order_Details.OrderID
9  JOIN  User_Account on User_Account.UserID = Orders.UserID
10 JOIN Customer_Info on Customer_Info.CustID = User_Account.CustID
11 GROUP BY Customer_Info.CustName,  Customer_Info.CustSegment
12 ORDER BY round(sum(( Order_Details.UnitPrice - Order_Details.UnitCost) * Order_Details.Quantity),2) DESC limit 20
```

Customer	Segment	Number of Orders	Profit	Sales
Tamara Chand	Corporate	12	$8,981.32	$19,052.22
Raymond Buch	Consumer	18	$6,976.10	$15,117.34
Sanjit Chand	Consumer	22	$5,757.41	$14,142.33
Hunter Lopez	Consumer	11	$5,622.43	$12,873.30
Adrian Barton	Consumer	20	$5,444.81	$14,473.57
Tom Ashbrook	Home Office	10	$4,703.79	$14,595.62
Christopher Martinez	Consumer	10	$3,899.89	$8,954.02
Keith Dawkins	Corporate	28	$3,038.63	$8,181.26
Andy Reiter	Consumer	9	$2,884.62	$6,608.45
Daniel Raglin	Home Office	13	$2,869.08	$8,350.87
Tom Boeckenhauer	Consumer	17	$2,798.37	$9,133.99
Nathan Mautz	Home Office	14	$2,751.68	$6,459.34
Sanjit Engle	Consumer	19	$2,650.68	$12,209.44
Bill Shonely	Corporate	9	$2,616.06	$10,501.65
Brian Moss	Corporate	29	$2,515.02	$8,382.95
Harry Marie	Corporate	20	$2,437.98	$8,236.76
Todd Sumrall	Corporate	15	$2,371.71	$11,891.75
Christopher Conant	Consumer	11	$2,177.05	$12,129.07
Jane Waco	Corporate	14	$2,173.71	$7,721.71
Helen Wasserman	Corporate	20	$2,164.16	$9,300.25

To know which customers have contributed more to profit, we created a view to see top 20 customers with profit, customer segment and its sales. This has given ideas about which customers should be considered first to offer a discount or promotional offer that can benefit the company. E-commerce business aim to focus on customers and its segments for company growth and increased profit. We see that "Tamara Chand" has contributed most to the profit.

- What are the top 10 states that have the maximum number of sales?

Top 10 State with city by sales performance

| State	City	Profit	Sales
New York	New York City	$62,025.44	$256,343.60
California	Los Angeles	$30,440.76	$175,851.34
Washington	Seattle	$29,156.10	$119,540.74
California	San Francisco	$17,507.39	$112,669.09
Pennsylvania	Philadelphia	$-13,837.77	$109,077.01
Texas	Houston	$-10,153.55	$64,504.76
Illinois	Chicago	$-6,654.57	$48,539.54
California	San Diego	$6,377.20	$47,521.03
Michigan	Detroit	$13,181.79	$42,446.94
Florida	Jacksonville	$-2,445.66	$39,133.33

```
1  create view Top_10_State_with_City_by_Sales AS
2  SELECT  a.State,
3          a.City,
4          CONCAT('$', FORMAT(sum((b.UnitPrice - b.UnitCost) * b.Quantity),2)) as 'Profit',
5          CONCAT('$', FORMAT(sum(b.Quantity * b.UnitPrice), 2)) as 'Sales'
6  FROM Orders c
7  JOIN Order_Details b on c.OrderID = b.OrderID
8  JOIN Shipping a on a.OrderID = c.OrderID
9  GROUP BY a.State, a.City
10 ORDER BY round(sum(b.Quantity * b.UnitPrice),2) DESC limit 10
```

To know which states, have top sales performance, we created a view to see the state name, sales, profit, and city name. E-commerce business aim to focus states to improve the profit and revenue based on these types of reports. We see that New York state has contributed most to the sales.
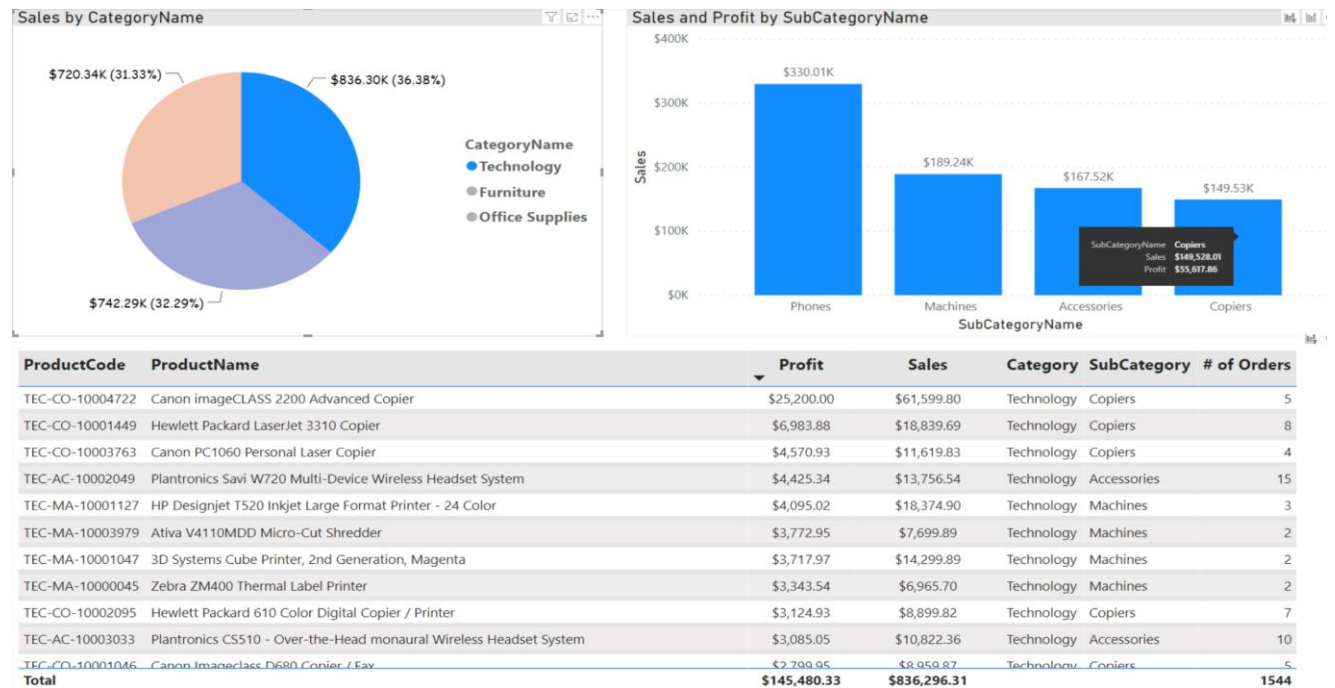
Dashboard through the lens of Power BI

Dashboard 1.0
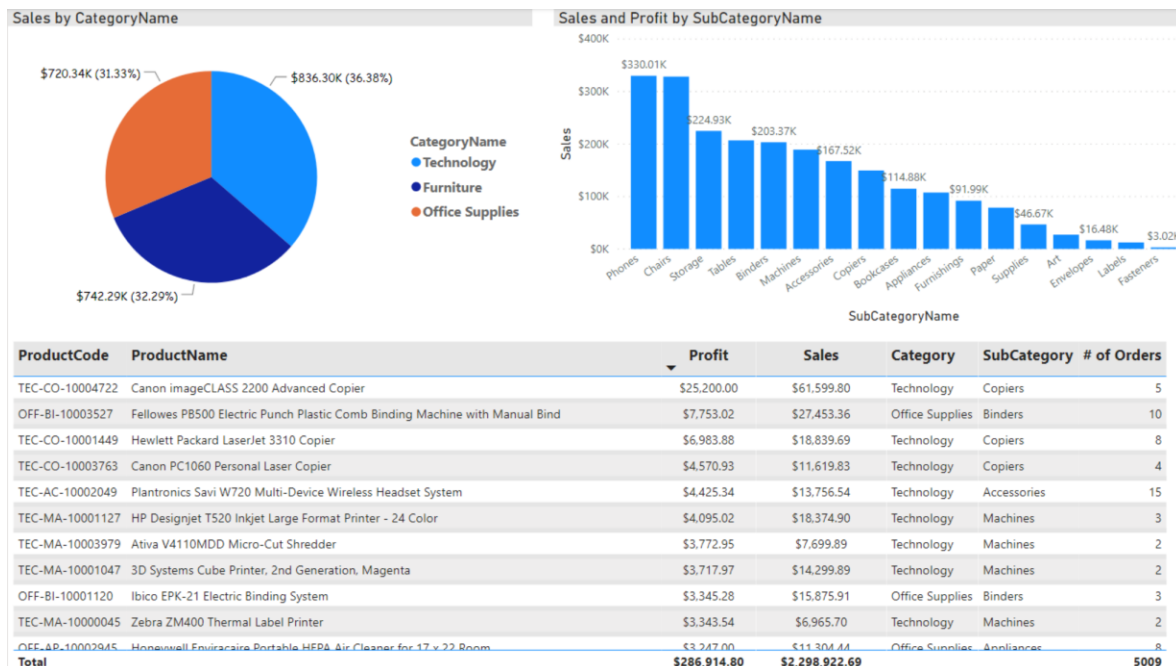
In this section, we took our stored data into the next level in term of data analytics. Thus, we proceeded ELT data integration process, in which we extracted data from our designed MySQL DB and export data into csv format. After that, we utilized BI tool called Microsoft Power BI to do loading and transform data in order to create dynamic dashboard.
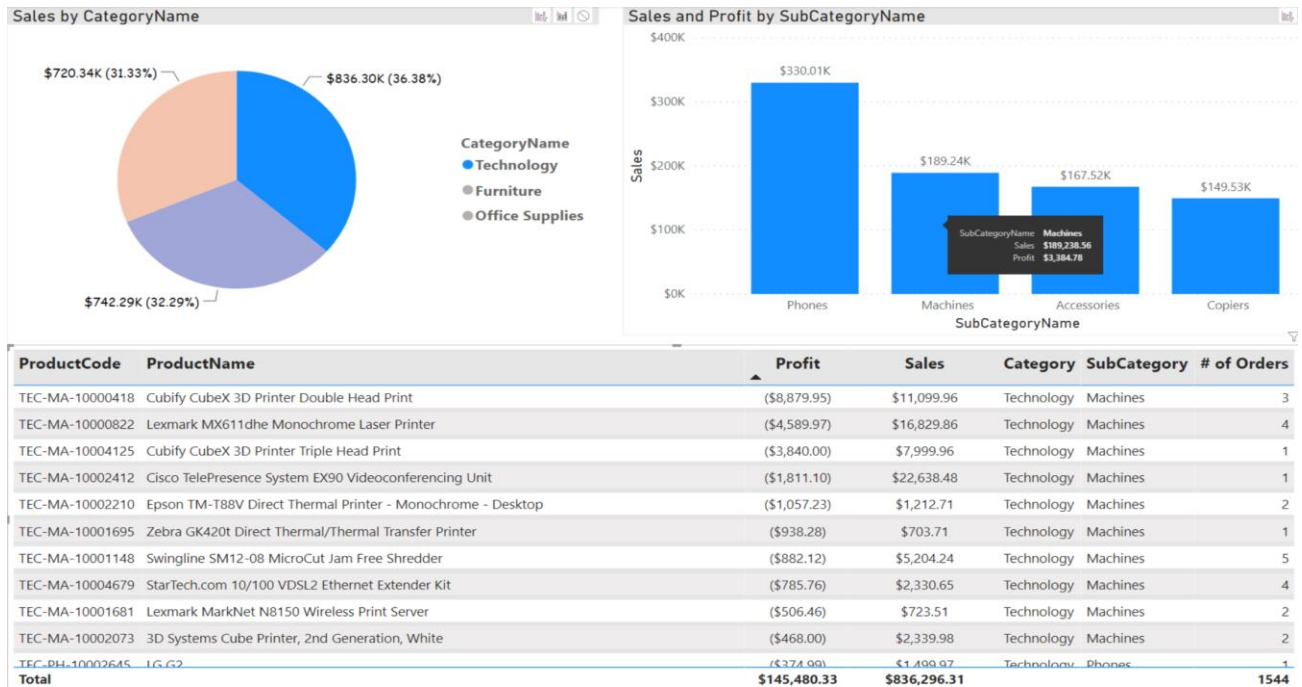
The first dashboard (1.0) displays sales performance by product category, subcategory, and drill down table for product detail. We will run into drill-down scenario below to simulate how business users can interact with this dashboard to gain insight.

Dashboard 1.1: Selecting Technology Category

ProductCode	ProductName	Profit	Sales	Category	SubCategory	# of Orders
TEC-CO-10004722	Canon imageCLASS 2200 Advanced Copier	$25,200.00	$61,599.80	Technology	Copiers	5
TEC-CO-10001449	Hewlett Packard LaserJet 3310 Copier	$6,983.88	$18,839.69	Technology	Copiers	8
TEC-CO-10003763	Canon PC1060 Personal Laser Copier	$4,570.93	$11,619.83	Technology	Copiers	4
TEC-AC-10002049	Plantronics Savi W720 Multi-Device Wireless Headset System	$4,425.34	$13,756.54	Technology	Accessories	15
TEC-MA-10001127	HP Designjet T520 Inkjet Large Format Printer - 24 Color	$4,095.02	$18,374.90	Technology	Machines	3
TEC-MA-10003979	Ativa V4110MDD Micro-Cut Shredder	$3,772.95	$7,699.89	Technology	Machines	2
TEC-MA-10001047	3D Systems Cube Printer, 2nd Generation, Magenta	$3,717.97	$14,299.89	Technology	Machines	2
TEC-MA-10000045	Zebra ZM400 Thermal Label Printer	$3,343.54	$6,965.70	Technology	Machines	2
TEC-CO-10002095	Hewlett Packard 610 Color Digital Copier / Printer	$3,124.93	$8,899.82	Technology	Copiers	7
TEC-AC-10003033	Plantronics CS510 - Over-the-Head monaural Wireless Headset System	$3,085.05	$10,822.36	Technology	Accessories	10
TEC-CO-10001046	Canon Imageclass D680 Copier / Fax	$2,799.95	$8,959.87	Technology	Copiers	5
Total		**$145,480.33**	**$836,296.31**			**1544**

After we drilled down into technology section, we can see the bar chart on the top right is automatically adjusted its information to display only those subcategories within our selected Technology category. We gained more insight with drilling down into product information. Even though copiers subcategory has the lowest total sales, the most profitable products (top 3 products) in Technology is in fact under copiers subcategory.

Dashboard 1.2: View the least profitable product in Technology

| ProductCode	ProductName	Profit	Sales	Category	SubCategory	# of Orders
TEC-MA-10000418	Cubify CubeX 3D Printer Double Head Print	($8,879.95)	$11,099.96	Technology	Machines	3
TEC-MA-10000822	Lexmark MX611dhe Monochrome Laser Printer	($4,589.97)	$16,829.86	Technology	Machines	4
TEC-MA-10004125	Cubify CubeX 3D Printer Triple Head Print	($3,840.00)	$7,999.96	Technology	Machines	1
TEC-MA-10002412	Cisco TelePresence System EX90 Videoconferencing Unit	($1,811.10)	$22,638.48	Technology	Machines	1
TEC-MA-10002210	Epson TM-T88V Direct Thermal Printer - Monochrome - Desktop	($1,057.23)	$1,212.71	Technology	Machines	2
TEC-MA-10001695	Zebra GK420t Direct Thermal/Thermal Transfer Printer	($938.28)	$703.71	Technology	Machines	1
TEC-MA-10001148	Swingline SM12-08 MicroCut Jam Free Shredder	($882.12)	$5,204.24	Technology	Machines	5
TEC-MA-10004679	StarTech.com 10/100 VDSL2 Ethernet Extender Kit	($785.76)	$2,330.65	Technology	Machines	4
TEC-MA-10001681	Lexmark MarkNet N8150 Wireless Print Server	($506.46)	$723.51	Technology	Machines	2
TEC-MA-10002073	3D Systems Cube Printer, 2nd Generation, White	($468.00)	$2,339.98	Technology	Machines	2
TEC-PH-10002645	LG G2	($374.99)	$1,499.97	Technology	Phones	1
Total		**$145,480.33**	**$836,296.31**			**1544**

Vice versa, as shown in dashboard 1.2, our top ten loss (negative net profit) are all from Machines subcategory, which is number two in term of revenue. From this analysis, our product manager and sales manager need to find out find we loss in this subcategory and those products.

Dashboard 2.0:
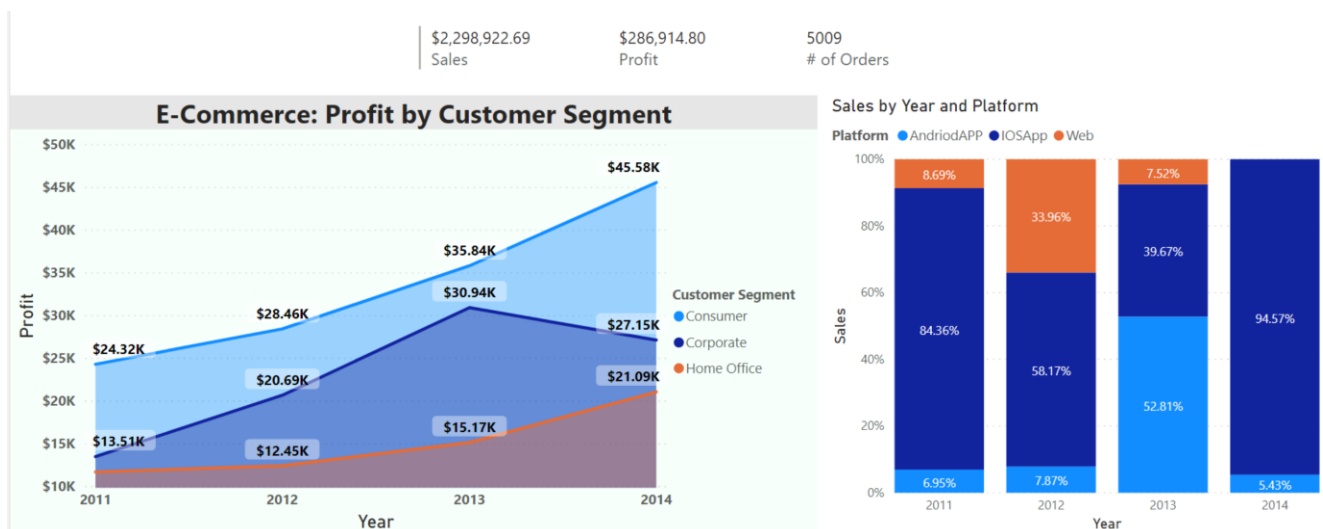
Another dashboard (2.0) is about profit by customer segment and platform usage. Since our sales channel is based on different online platforms, it is important to analyze this information for our E-commerce.

Dashboard 2.1: Drilling down into Corporate in 2013
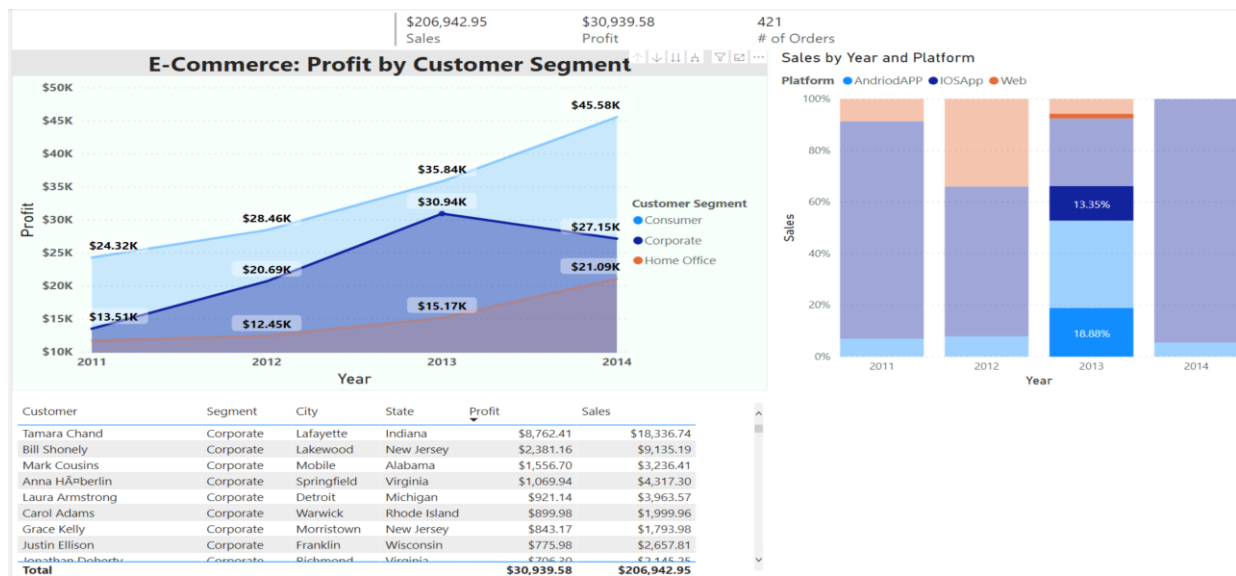
As mentioned previously, we see the declining profit for Corporate segment in 2014. When clicking into Corporate 2013, we see that big profit was derived from Tamara Chand. Moreover, we also learn from the right bar chart that Web platform is gone in 2014.

Dashboard 2.2: Drilling down into Corporate in 2014
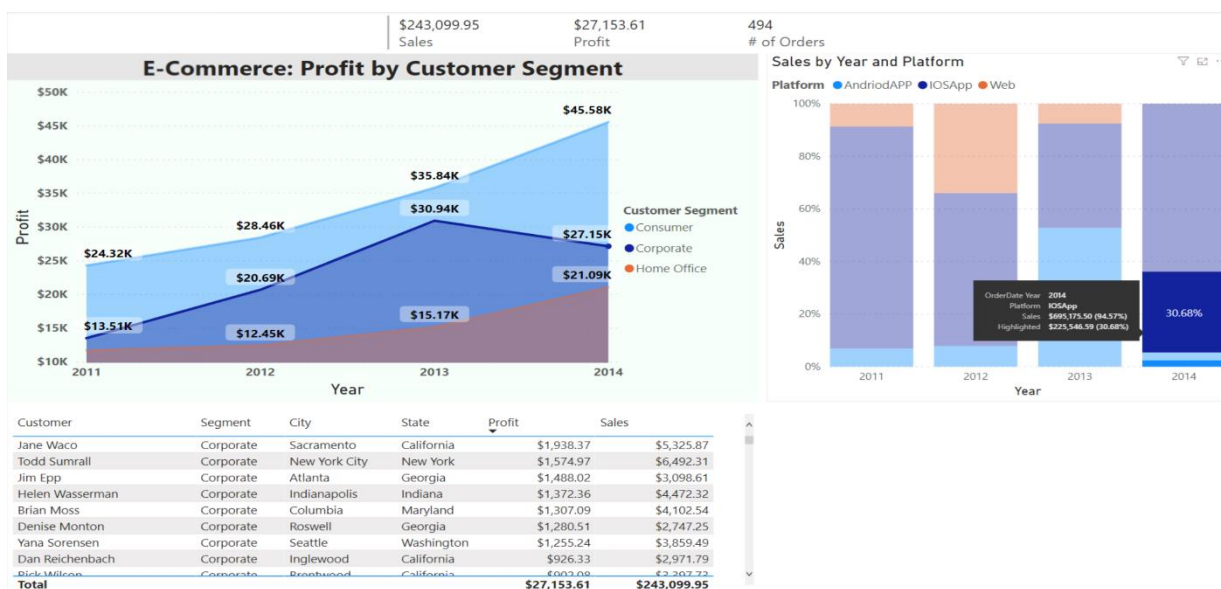
As we can see for 2014, our top customers from 2013 are gone in 2014. Thus, this is where we can ask sales account manager to see how we lose such a big account in 2014. It might be something about pricing, our platform interface, or customer services. This analysis provides us more insights and data support to make business decision.

Conclusion

In conclusion, our project goal is satisfied as we completed our database structure design, inserting data, and then performance data analysis to answer our business questions for our E-commerce business. With RDBMS in MySQL, we have gained a luxury of database power where we can manage, modify, and maintain our database systematically and easily using SQL. Now, our database supports the streaming of online transactional data from our E-commerce.

In term of data analytics, we have learned the importance of storing quality data, reporting, and data visualization. The phrase "garbage in garbage out" stays in its truth where the company will be able to gain useful information or insights when we retrieve quality data for different applications. Each created view represents report that can answer different business questions. On the other hands, the dynamic dashboard provides more interactive and drilling down features for business users.

References

Oracle. (n.d.) Database. Retrieved from https://www.oracle.com/database/what-is-database.html

Rojas, N. (2019). *The Role of the Database in E-Commerce*. Retrieved from

https://trellis.co/blog/the-role-of-the-database-in-e-commerce/

Singh, A. (2017). *Role of Database in E-Commerce (With vs Without)*. Retrieved from

https://brandlitic.com/role-of-database-in-ecommerce/

Winkler, Nick. (2020). *What Is the Future of Ecommerce? 10 Insights on the Evolution of an*

Industry. Retrieved from https://www.shopify.com/enterprise/the-future-of-ecommerce