EXPLORATORY DATA ANALYSIS ON E-COMMERCE DATA PROJECT #1

By Rina Rafalski

Intern Code: OGTIPIRDA459

Supervisor: Aritri Debnath

Oeson
Inspiring generation

Table of contents

Introduction

The goal of this project – analyzing e-commerce sales based on various product categories, monthly sales, profits and sales by region.

This project involves using Excel built-in functions, pivot tables, charts, slices, combo box and forecasting features.

Data description

The data set contains

- 51290 records of order details for the year 2015

- 4 products categories: Auto & Accessories, Electronic, Fashion, Home & Furniture

- 13 regions, 147 countries, 3636 cities

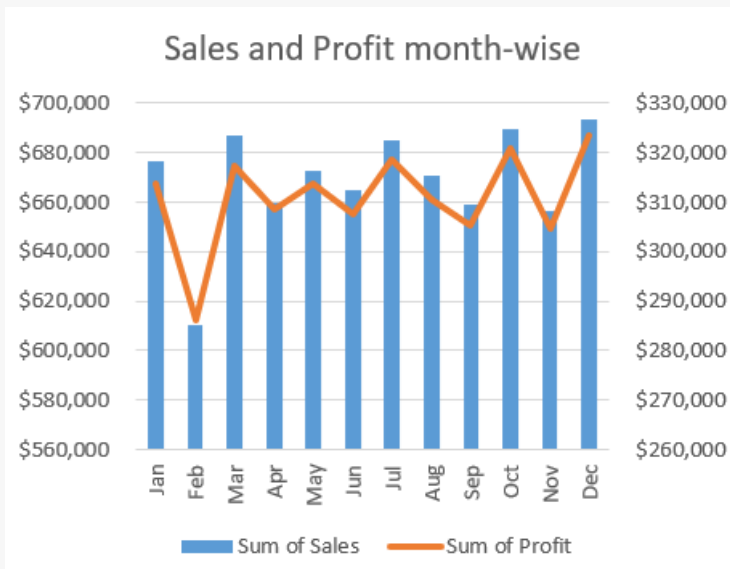| Order ID | Order Date | Ship Date | Aging | Ship Mode | Product Ca | Product | Sales | Quantity | Discount | Profit | Shipping | Order Prio | Customer | Customer | Segment | City | State | Country | Region | Months |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AU-2015-1 | 9/11/2015 | 17/11/2015 | 8 | First Class | Auto & Acces | Car Media Pla | $ 140.0 | 2 | 0.05 | $ 46.0 | $ 4.6 | Medium | LS-001 | Lane Daniels | Consumer | Brisbane | Queensland | Australia | Oceania | Nov |
| AU-2015-2 | 30/6/2015 | 2/7/2015 | 2 | First Class | Auto & Acces | Car Speakers | $ 211.0 | 3 | 0.03 | $ 112.0 | $ 11.2 | Medium | IZ-002 | Alvarado Kriz | Home Office | Berlin | Berlin | Germany | Central | Jun |
| AU-2015-3 | 5/12/2015 | 13/12/2015 | 8 | First Class | Auto & Acces | Car Body Cov | $ 117.0 | 5 | 0.01 | $ 31.2 | $ 3.1 | Critical | EN-003 | Moon Weien | Consumer | Porirua | Wellington | New Zealand | Oceania | Dec |
| AU-2015-4 | 9/5/2015 | 16/5/2015 | 7 | First Class | Auto & Acces | Car & Bike Ca | $ 118.0 | 2 | 0.05 | $ 26.2 | $ 2.6 | High | AN-004 | Sanchez Bergr | Corporate | Kabul | Kabul | Afghanistan | Central Asia | May |
| AU-2015-5 | 9/7/2015 | 18/7/2015 | 9 | First Class | Auto & Acces | Tyre | $ 250.0 | 1 | 0.04 | $ 160.0 | $ 16.0 | Critical | ON-005 | Rowe Jackson | Corporate | Townsville | Queensland | Australia | Oceania | Jul |
| AU-2015-6 | 25/2/2015 | 5/3/2015 | 8 | First Class | Auto & Acces | Bike Tyres | $ 72.0 | 3 | 0.04 | $ 24.0 | $ 2.4 | Critical | TO-006 | Carter Barreto | Corporate | Bytom | Silesia | Poland | EMEA | Feb |
| AU-2015-7 | 9/4/2015 | 10/4/2015 | 1 | First Class | Auto & Acces | Car Mat | $ 54.0 | 1 | 0.05 | $ 54.0 | $ 5.4 | High | OM-007 | Mcconnell To | Consumer | Chicago | Illinois | United States | Central | Apr |
| AU-2015-8 | 30/3/2015 | 6/4/2015 | 7 | First Class | Auto & Acces | Car Seat Cove | $ 114.0 | 5 | 0.02 | $ 22.6 | $ 2.3 | Critical | AN-008 | Dennis Hollon | Corporate | Suzhou | Anhui | China | North Asia | Mar |
| AU-2015-9 | 9/2/2015 | 16/2/2015 | 7 | First Class | Auto & Acces | Car Pillow & N | $ 231.0 | 5 | 0.03 | $ 116.4 | $ 11.6 | Critical | EN-009 | Wall Olsen | Consumer | Juárez | Chihuahua | Mexico | North | Feb |

Dashboard
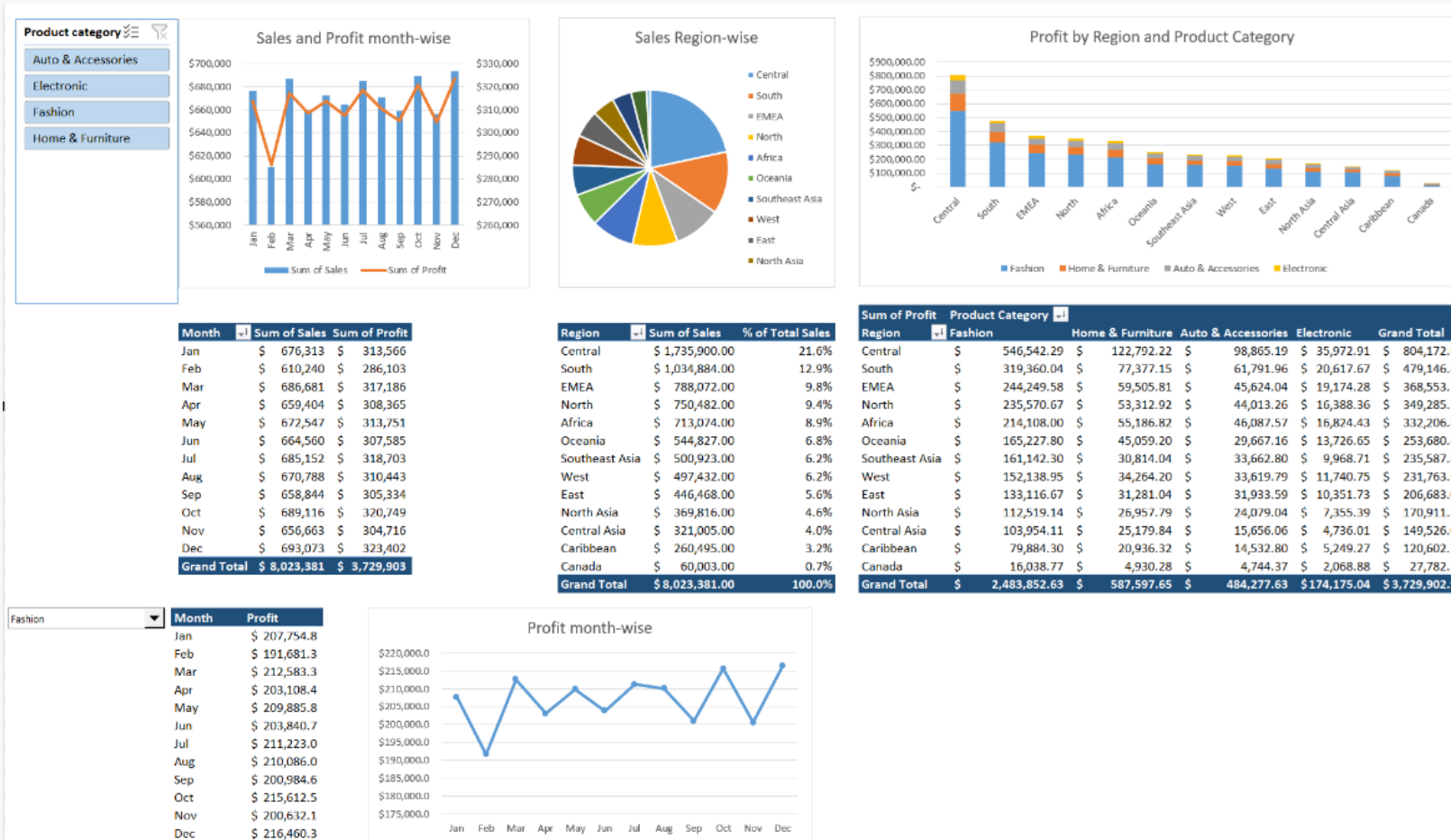
The interactive dashboard combines all pivot tables and charts in one place along with filter options such as slicer and combo box.

It helps the company to gain various insights quickly and understand the data to grow the business in all regions in all product categories.

Analysis

Sales and Profit month-wise

Month	Sum of Sales	Sum of Profit
Jan	$ 676,313	$ 313,566
Feb	$ 610,240	$ 286,103
Mar	$ 686,681	$ 317,186
Apr	$ 659,404	$ 308,365
May	$ 672,547	$ 313,751
Jun	$ 664,560	$ 307,585
Jul	$ 685,152	$ 318,703
Aug	$ 670,788	$ 310,443
Sep	$ 658,844	$ 305,334
Oct	$ 689,116	$ 320,749
Nov	$ 656,663	$ 304,716
Dec	$ 693,073	$ 323,402
Grand Total	$ 8,023,381	$ 3,729,903

- The total amount of Sales for 2015 is 8,023,381$ with Profit of 3,729,903$

- The best month of the year with 693,073$ Sales and 323,402$ Profit is December

- The worth month of the year with 610,240$ Sales and 286,103$ Profit is February
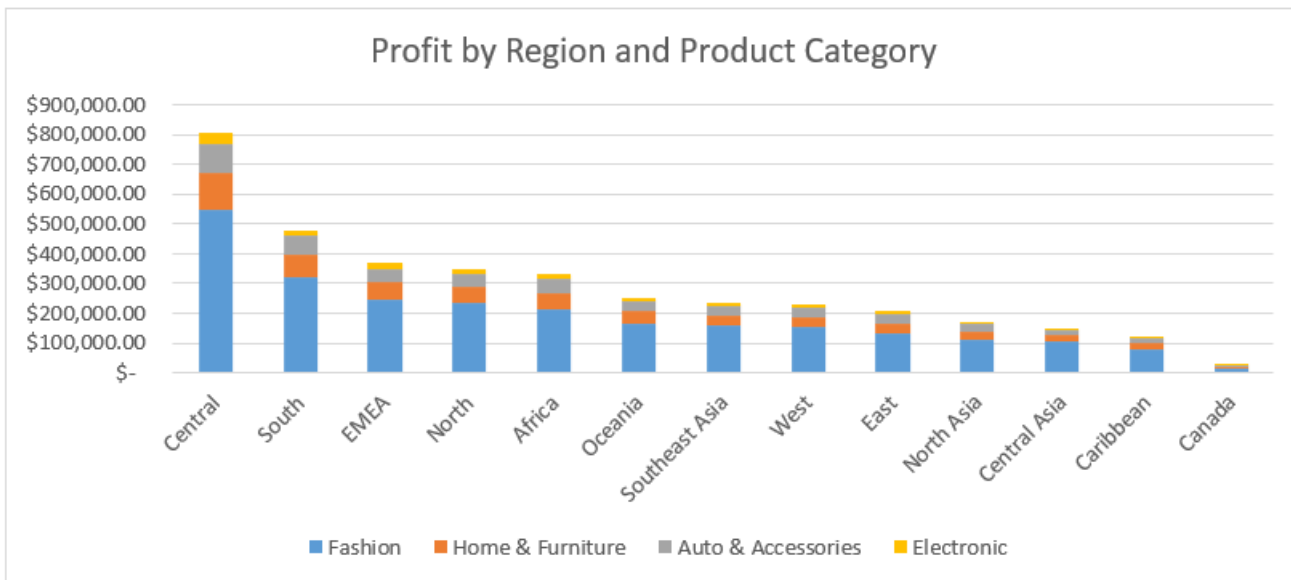
Analysis

Sales region-wise

Region	Sum of Sales	% of Total Sales
Central	$ 1,735,900.00	21.6%
South	$ 1,034,884.00	12.9%
EMEA	$ 788,072.00	9.8%
North	$ 750,482.00	9.4%
Africa	$ 713,074.00	8.9%
Oceania	$ 544,827.00	6.8%
Southeast Asia	$ 500,923.00	6.2%
West	$ 497,432.00	6.2%
East	$ 446,468.00	5.6%
North Asia	$ 369,816.00	4.6%
Central Asia	$ 321,005.00	4.0%
Caribbean	$ 260,495.00	3.2%
Canada	$ 60,003.00	0.7%
Grand Total	**$ 8,023,381.00**	**100.0%**

- The most profitable region with 21.6% of total sales is Central region that includes 12 countries: Austria, Belgium, El Salvador, France, Germany, Guatemala, Honduras Netherlands, Nicaragua, Panama, Switzerland and United States.

- The less profitable region with 0.7% of total sales is Canada.

Analysis

Profit by Region and Product Category
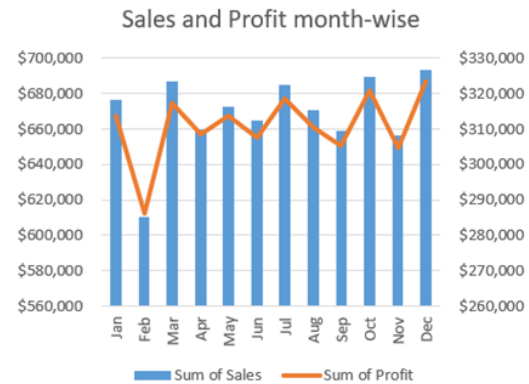

Profit by Region and Product Category

- The most profitable product category in all regions is Fashion

- The less profitable product category in all regions is Electronics

Analysis

All Products categories

Month	Sum of Sales	Sum of Profit
Jan	$ 676,313	$ 313,566
Feb	$ 610,240	$ 286,103
Mar	$ 686,681	$ 317,186
Apr	$ 659,404	$ 308,365
May	$ 672,547	$ 313,751
Jun	$ 664,560	$ 307,585
Jul	$ 685,152	$ 318,703
Aug	$ 670,788	$ 310,443
Sep	$ 658,844	$ 305,334
Oct	$ 689,116	$ 320,749
Nov	$ 656,663	$ 304,716
Dec	$ 693,073	$ 323,402
Grand Total	$ 8,023,381	$ 3,729,903

Fashion Product category

Fashion

Month	Profit
Jan	$ 207,754.8
Feb	$ 191,681.3
Mar	$ 212,583.3
Apr	$ 203,108.4
May	$ 209,885.8
Jun	$ 203,840.7
Jul	$ 211,223.0
Aug	$ 210,086.0
Sep	$ 200,984.6
Oct	$ 215,612.5
Nov	$ 200,632.1
Dec	$ 216,460.3

- The total profit is largely influenced by the most profitable product category - Fashion

Root Mean Square Error (RMSE)
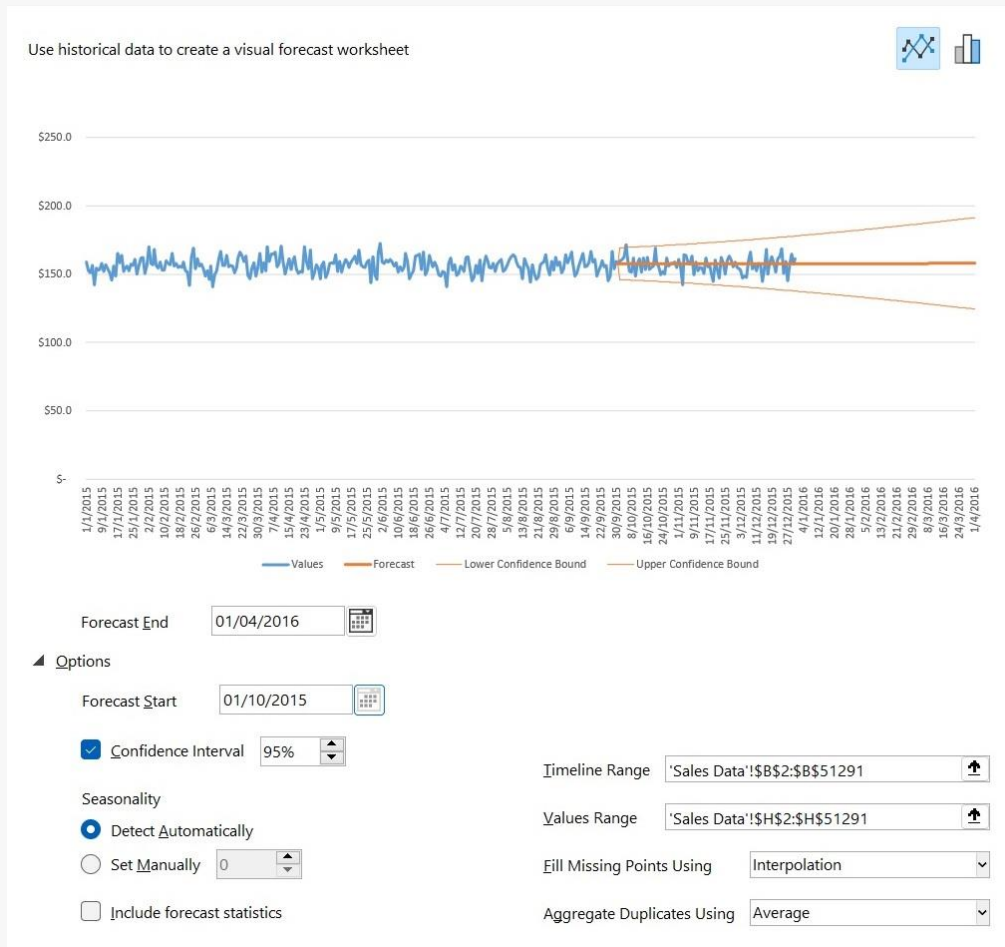
- Root mean square error (RMSE) measures the average difference between a statistical model's predicted values and the actual values.

- Set of predicted values was created with Excel Forecast sheet that uses forecasting method **Exponential smoothing** which combines error, trend, and seasonal (or ETS) calculations.

Root Mean Square Error (RMSE)

The calculation of RMSE was provided with 2 methods, in both of them the result is: **RMSE = 5.92**

- **Method 1**

  Step 1: Calculate the Squared Differences:

  (B275 – C275)^2

  Step 2: Calculate the Mean of Squared

  Differences: AVERAGE(H275:H366)

 Step 3: Calculate the Square Root

  of the Mean: =SQRT(J2)

- **Method 2**

  in 1 step with formula:

  =SQRT(SUMSQ(G275:G366)/COUNT(G275:G366))

	A	B	C	F	G	H	I	J
	Timeline	Value	Foreca		Error	Squared Error	Statistic	Value
	1/1/2015	$ 158.9					Mean of squared error	35.07
	2/1/2015	$ 152.3					RMSE	5.92
	3/1/2015	$ 151.2						5.92
	1/10/2015	$ 157.7	$ 157.7		0.00	0.00		
	2/10/2015	$ 159.8	$ 157.7		-2.08	4.33		
	3/10/2015	$ 160.8	$ 157.7		-3.10	9.61		
	4/10/2015	$ 162.5	$ 157.7		-4.81	23.16		
	5/10/2015	$ 171.5	$ 157.7		-13.78	189.88		

=SQRT(SUMSQ($G$275:$G$366)/COUNT($G$275:$G$366))

	A	B	C	F	G	H	I	J	M	N
	Timeline	Value	Foreca		Error	Squared Error	Statistic	Value		
	1/1/2015	$ 158.9					Mean of squared error	35.07		
	2/1/2015	$ 152.3					RMSE	5.92		
	3/1/2015	$ 151.2					RMSE	5.92		
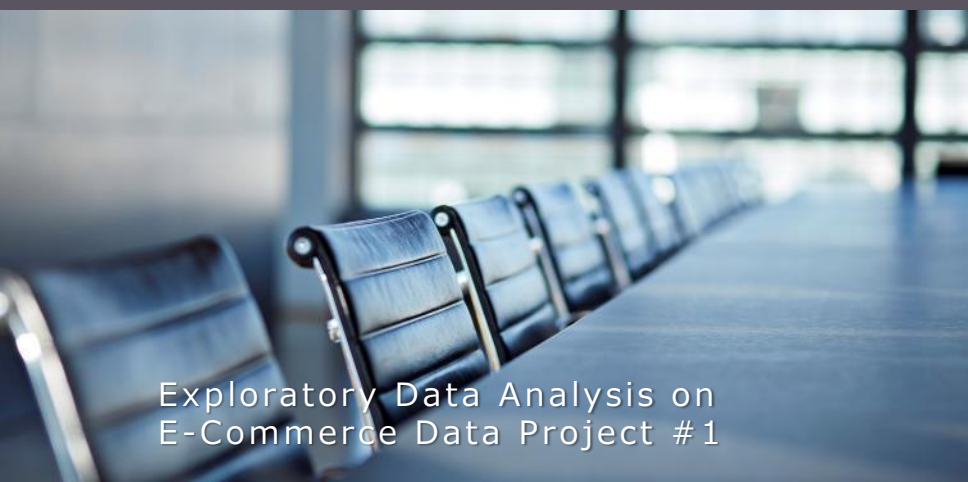
Summary

- The total amount of Sales for 2015 is 8,023,381$ with Profit of 3,729,903$

- The best month of the year with 693,073$ Sales and 323,402$ Profit is December

- The most profitable region with 21.6% of total sales is Central

- The most profitable product category in all regions is Fashion

- The total profit is largely influenced by the most profitable product category

THANK YOU

Exploratory Data Analysis on
E-Commerce Data Project #1