**Data Set's Raw Analysis Summary**

Large Data set has six excel spreadsheets (named QVC data 1, 2, 3 etc) with less than 1 million rows. To be more precise QVC data 1,2,3,4 & 5 has 850k records each and QVC data 6 has approximately 430k records of Order data.

In this project the approach is to combine the 6 spreadsheets data and perform the analysis on the larger set using R programming & Fundamental Data Science techniques.

There are also excel spreadsheet with the following information available:

* Distribution Center data
* Order type data
* Data Dictionary

Analysis of the following Questions:

* **Does the current distribution network maximize customer penetration (spend)?  If not, what should QVC do to increase customer penetration with the current distribution network?**

Customer Penetration can be defined as relationship to Total Amount of QVC in sales with respect to population of a across the States.

1. Determine largest Warehouses where majority of overall Total Sales Volume in % is generated from:

*[ Source ship Warehouse NBR ]* - Unique identifier denoting the warehouse where the product is shipped from

*[ Total\_Line\_Amt]* - Total amount owed for the specific order line

*[ Warehouse % Sales Volume in $ ]* \*

* Total Sales Amount across USA / Total Sales Amount in top 5 Warehouses

1. Identify specific States that generate major share of total volumes of Sales in USA :

*[Ship\_to\_State] -* State the package was delivered to

*[ % of Total Sales per State ]\**

* *Total Sales Volume ($) in US / Sales Volume ($) per each Top 5 State*

1. Establish relationship between Distance of shipment & Delivery Time

*[ Delivery (days) ]* - Delivery\_Confirmation\_date - Order\_Date

*[ Distance ]\**

* Distance between Warehouse location and Shipment location

1. Delivery Time and Order Placement

*[Number of Orders Placed or % of Order placed]*\*

*[Delivery Time]\** - Estimated Delivery Days

1. Shipping speed is a key variable as faster delivery is a key factor in online purchases:

*[ Delivery Confirmation Date , Order Dt ]*

*[ Shipping Priority Ind is null ]*

1. Among Top selling states Relationship between Sales and Population across States.

Which states are under penetrated or above penetrated.

*[ Sales Amount ($)] \**

*[Population in each State (Million)]\**

1. Customer Retention Rate: Earlier/On-time orders compared to Late Orders.

Customer that receive faster shipment than anticipated delivery times.

* *By Department ( Best Selling Products )*
* *Under Penetration*
* *High Out of State Orders*
* *Long Shipping Time*

* **Are there specific products or product categories that should be located in specific distribution centers?**

* **Do customers that receive their product sooner purchase more than customers with longer delivery times?**

1. Determine % of Orders that are out of State and take longer Avg delivery times across top States.

Final Summary:

Develop a visualization that contains useful information for QVC to use to understand what the relationship between speed of product/package delivery and customer loyalty is.

 Sources:

* QVC Data Sets Provided : <https://www.dropbox.com/sh/m7hf33yzmbf4tqv/AADOPe8A3qyHCFacu79_kB8ka?dl=0>
* QVC Fact Sheet 2017 : <http://corporate.qvc.com/documents/20536/164719/QVC+Fact+Sheet+Q2+2017+8-8-17_final.pdf/e9576a5a-7b0e-41c1-a3ec-d55c65c5eee8>
* US Population : <https://www.census.gov/programs-surveys/popest/data/data-sets.html>