Deliverable 1:

Customer city and customer country columns are deleted from the dataset because they are irrelevant to the project. They are the origin city and country of the order made by the customer. Both the values will not affect the overall analysis.

Same is done for the “Type”, “Customer Email”, “Customer Password” column because it is irrelevant to the analysis.

The “Order Origin” column is the column which shows the region in the world where the order is delivered. This heading name is a little misleading and is edited to make it more understandable to “Delivery Region”. The same is done for order the column “Order State”, it is changed to “Delivery State.” Customer State is changed to Vendor State, Customer Street is changed to Vendor Street, Market is changed to delivery market, Latitude is changed to vendor latitude, Longitude is changed vendor longitude

To make the overall data make sense, some assumptions are made. They are that the negative values in the columns “Order Item Profit Ratio” and the “Order Profit Per Order” are customer return orders and they can be ignored for the analysis. Which amounts to 33784 values points.