

E-Commerce Website Data Cleaning Report

This report is regarding the Data Cleaning Task of the E-Commerce Data Analysis Internship Project.

Before processing, the Dataset had 536349 records, with 8 attributes for each record. In addition, the datatypes of each attribute were discovered.

<i>Attribute</i>	<i>Function</i>	<i>Datatype</i>
<i>TransactionNo</i>	ID of the Transaction	object
<i>Date</i>	Date of the Transaction	object
<i>ProductNo</i>	ID of the Product	object
<i>ProductName</i>	Name of the Product	object
<i>Price</i>	Price of the Product per unit	float64
<i>Quantity</i>	Quantity of the Product purchased	int64
<i>CustomerNo</i>	ID of the Customer	float64
<i>Country</i>	Country of origin of the Transaction	object

Exploring the dataset, we found that the dataset had 55 records where the attribute *CustomerNo* had no value. Those records were dropped, since the three ID attributes are Candidate Key attributes and cannot be imputed through replacement.

After the null-valued records were dropped, we shifted to datatype conversion and standardization, if needed.

On exploration, we found that the *Date* and *CustomerNo* attributes had the incorrect datatype. The attributes were converted to their respective correct datatypes.

<i>Attribute</i>	<i>Old Datatype</i>	<i>New Datatype</i>
<i>Date</i>	object	datetime64
<i>CustomerNo</i>	float64	object

Next, the Dataset was examined for data inconsistencies and outliers. There were some outliers in the *Price* and *Quantity* attributes that were removed.

Lastly, a new attribute was added, Total Revenue, calculated as the product of Price and Quantity.