# Kaggle Project Report

--------------------------------------------------

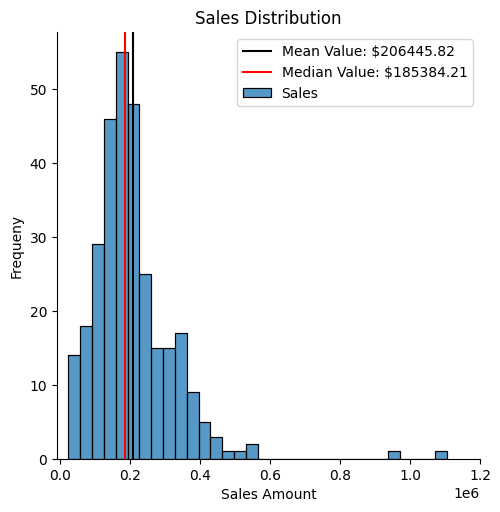
## Analyzing Daily Sales Data Frame

### Before Removing Outliers

Before removing any outliers from the dataset, I have found that the average daily sales amount is $206445.82; the minimum amount is $22892.13; 25% of all daily sales amounts were less than $136182.63; 50% of all daily sales amounts were less than $185384.21; 75% of all daily sales amounts were less than $252605.07; the maximum daily sales amount is $1106069.97; and the standard deviation of daily sales amounts is $118352.73.

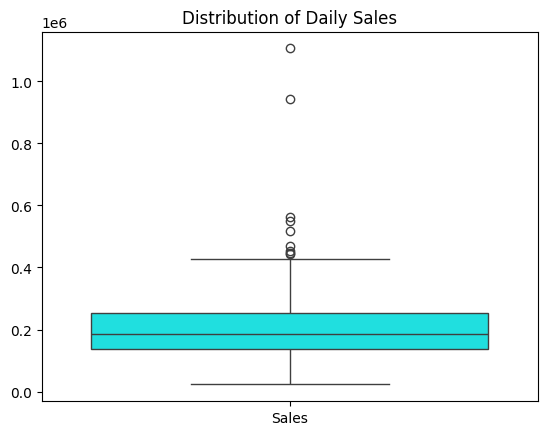
#### Visualizations of 'Daily Sales' Before Removing Outliers

Histogram showing distribution:



It's clear that there are a few extreme outliers in the data.

Box Plot showing distribution:



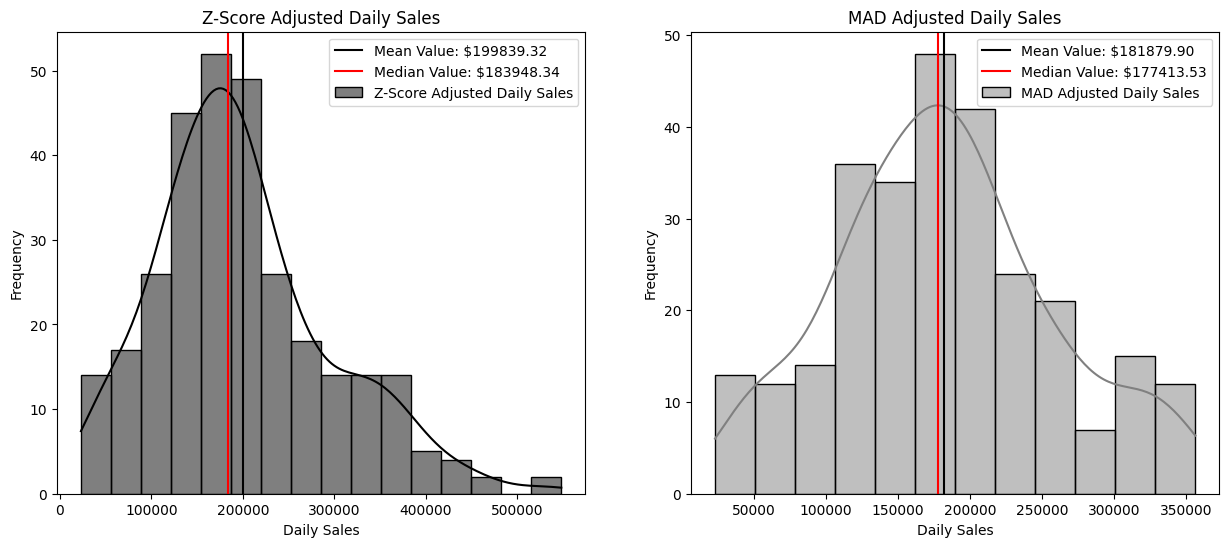
This extra perspective confirms the existence of extreme outliers.

--------------------------------------------------

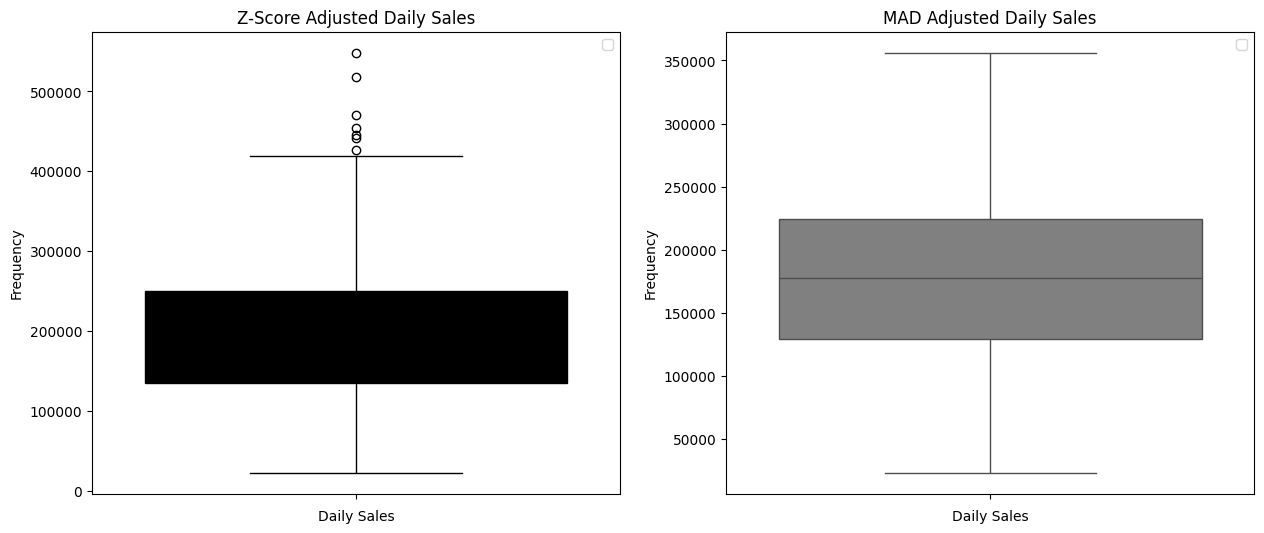
## Visualization of Distribution of Daily Sales After Removing Outliers via Z-Scores and Mean Absolute Deviation:

It's clear that Mean Absolute Deviation (MAD) gives a more balanced distribution of the dataset. The spread between the median and the mean is tighter(z\_score mean/median spread: 15890.988708609279; MAD spread: 4466.360251798556), which is something I always look for as a sign of balance. MAD will be used to remove the outliers from the data.

### Visualizing the Distributions with Histograms



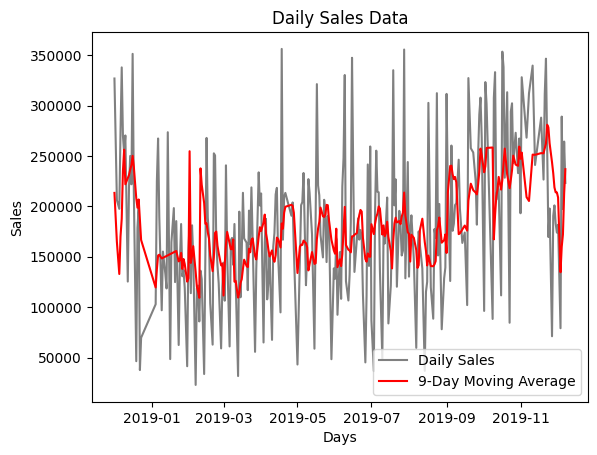
### Visualizing the Distributions with Box Plots



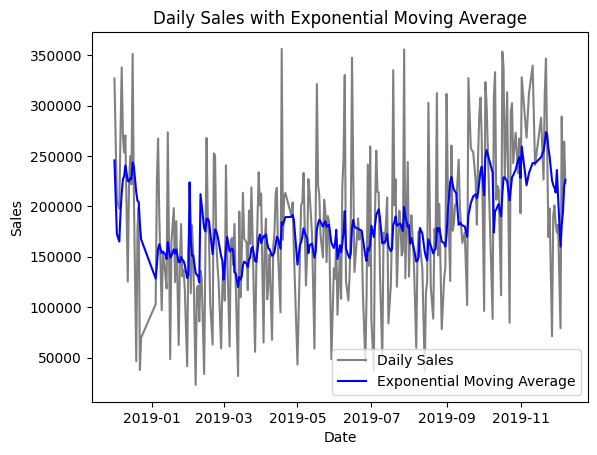
--------------------------------------------------

### Visualizing Time-Series Sales Data

During plotting the daily, monthly, and annual sales I saw that the daily sales data was quite noisy; this is common due to sales and discounts and other special events and deals. I tried to smooth the data to some extent by adding the 9-Day Moving Average to the graph overlapping the data:



When that didn't offer much insight, I tried the same visualization with the exponential moving average:



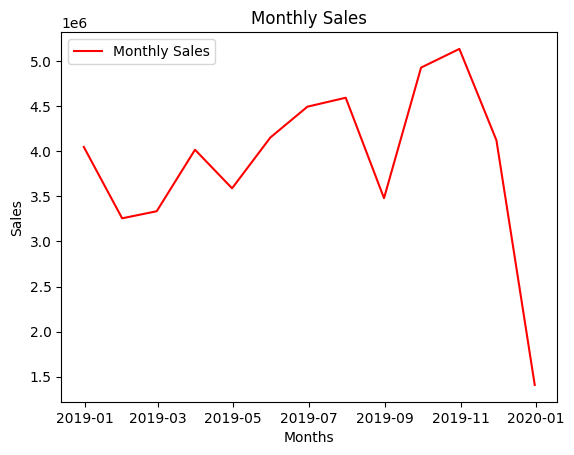
While the exponential moving average was slightly better, I ultimately decided to resameple the data to weekly sales to get rid of all the noise:

#### All Time-Series Visuals:

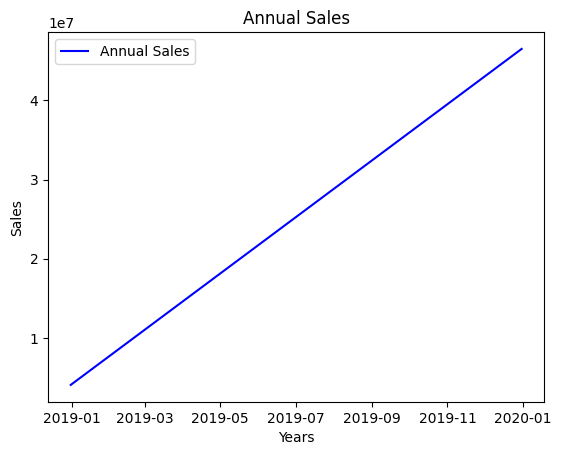
##### Weekly Sales:



##### Monthly Sales:



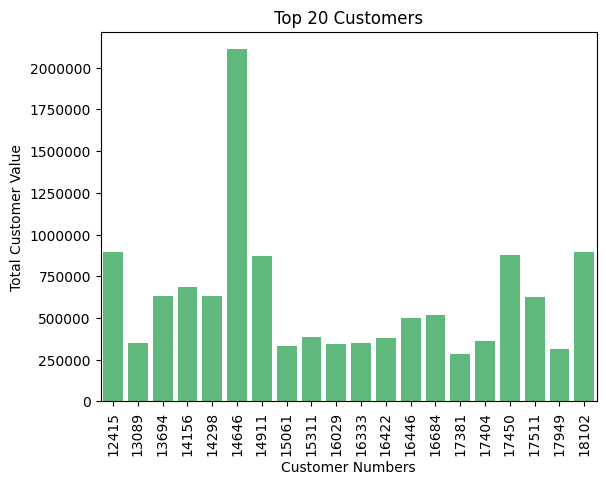
##### Annual Sales:



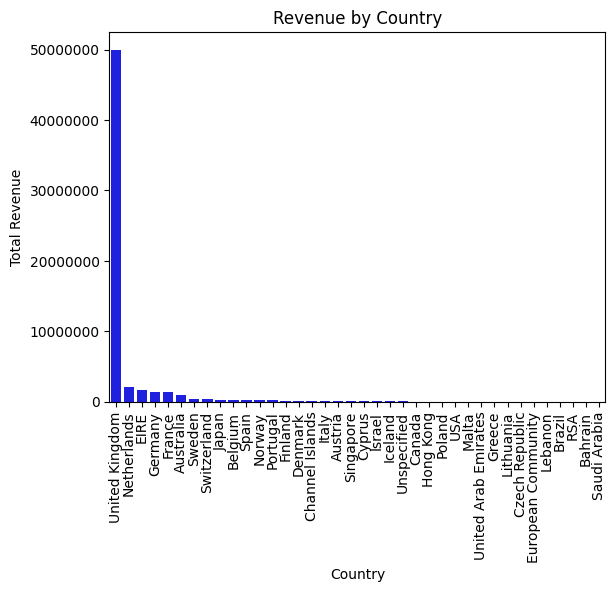
--------------------------------------------------

# Findings:

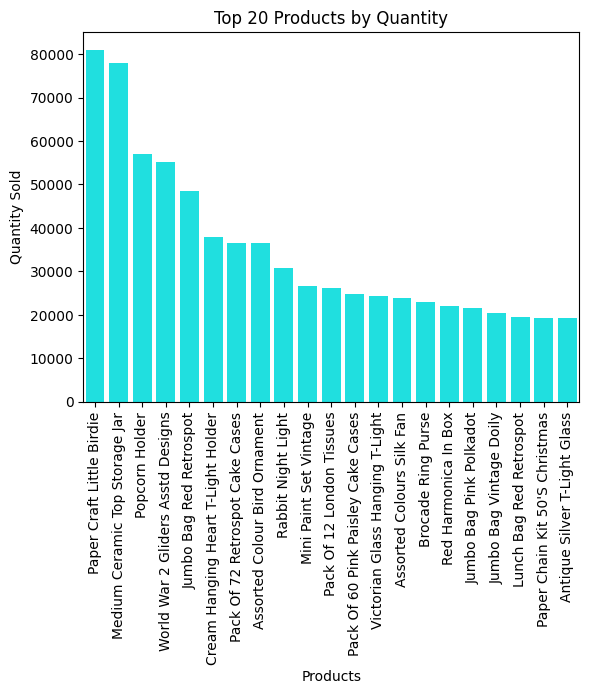
The number one customer is customer no. 14646.



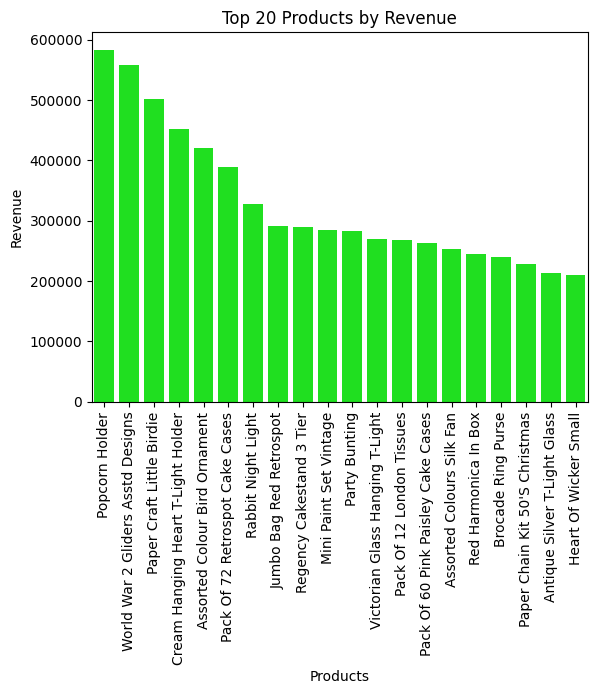
The United Kingdom is by far the number one country by revenue.



I found the number one procduct by quantity was the Paper Craft Little Birdie.



The number one product by revenue was the Popcorn Holder.



For in-depth visualizations, see the dashboard.