Coffee Futures Price & Quality Analysis

Done by: Ng Chuen Siong, Nelson

Problem Statement

Owing to the growing global coffee demand, one of our clients contacted us, The Management Lab, to do a study whether it is feasible to open coffeehouses that sells premium coffee (arabica) in a rising coffee price environment (deficit in global coffee stockpile) [1].

Business Research Questions

Assume the coffee price follows the coffee "C" future price.

1. Is the mean coffee total quality score [2] similar among countries?
2. Is the mean coffee price (USD) similar among coffee total quality score?
3. What is the ideal price to purchase coffee?
4. Is it a good time to enter the business of selling premium coffee?

Based on coffee "C" future price, Mexico, Salvador, Guatemala, Costa Rica, Nicaragua, Kenya, Papua New Guinea, Panama, Tanzania, Uganda, Honduras, and Peru all at par, Colombia at 400 point premium, Burundi, Rwanda, Venezuela and India at 100 point discount, Dominican Republic and Ecuador at 400 point discount, and Brazil at 600 point discount, and that coffees that are judged better are at a premium; those judged inferior are at a discount [3].

Project Objectives

Project Phase 1

- Check whether coffee total quality score is systematically higher or different in some countries than in others.

Project Phase 2

- Check whether coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Project Phase 3

- Check whether sample coffee price (USD) mean is equal, larger, or smaller than the population coffee price (USD) mean

Project Phase 4

- Evaluate the feasibility of selling premium coffee in the current environment.

Datasets: Kaggle's Coffee Quality database from CQI [4], Macrotrends's Coffee Prices - 45 Year Historical Chart [5].

Problem Solving Steps

Project Phase 1

1. Ensure that the sample of coffee total quality score is normally distributed.
   - If the sample is not normally distributed, test for normal distribution after performing data transformation (box-cox, yeo-johnson, log-normal, and square root) on it.
2. If the sample of coffee total quality score is still not normally distributed after performing data transformation. Kruskal–Wallis H test and Mann-Whitney U Test (non-parametric tests) are used instead of ANOVA and Turkey's Test (parametric tests).

Note: The sample data is based on coffee total quality score from 30 April, 2010 to 19 January, 2018.

Project Phase 2

Assume the coffee price follows the coffee "C" future price.

1. Ensure that the sample of coffee future price is normally distributed (Ethiopian coffee are excluded, because there are not much coffee future contracts done in Ethiopia [6]).
   - If the sample is not normally distributed, test for normal distribution after performing data transformation (box-cox, yeo-johnson, log-normal, and square root) on it.
2. If the sample of coffee future price is still not normally distributed after performing data transformation. Kruskal–Wallis H test and Mann-Whitney U Test (non-parametric tests) are used instead of ANOVA and Turkey's Test (parametric tests).

Note: The sample data is based on coffee "C" future price (USD) from 30 April, 2010 to 19 January, 2018.

Project Phase 3

Assume the coffee price follows the coffee "C" future price.

1. Ensure that the population of coffee future price is normally distributed.
   - If the population of coffee future price is not normally distributed, perform box-cox transformation on population. To create a new population, select a start date from

one of the earliest date with the highest p-value (p-value > 0.05). p-value is obtained using D'Agostino-Pearson test.

2. Perform 1 sample Z test on the box-cox transformed versions of the previous sample of coffee future price and new population to test whether the sample mean is equal, greater, or smaller than the new population mean.
3. Perform inverse box-cox transform to find the new population mean and sample mean.

Note: The sample data is based on coffee "C" future price (USD) from 30 April, 2010 to 19 January, 2018, and the population data is based on coffee "C" future price (USD) from 20 August, 1973 to 13 October, 2021

Project Phase 4

Assume the coffee price follows the coffee "C" future price.

1. Create boxplots of coffee future price to check where the current coffee future price lies.

Note: The population data is based on coffee future "C" price (USD) from July 1, 1974 to 13 October, 2021.

Import Libraries & Packages

Import Libraries

In [1]:

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import scipy.stats as stats
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

from scipy.stats import anderson
from scipy.stats import boxcox
from scipy.stats import chi2_contingency
from scipy.stats import normaltest
from scipy.stats import shapiro
from scipy.stats import skew
from scipy.stats import yeojohnson
from statsmodels.graphics.gofplots import qqplot
```

Import Packages

In [2]:
```
# Import my packages.
import my_Functions_cf_clean as myfcf_cle
import my_Functions_cf_wrangle as myfcf_wra
import my_Functions_cf_graph as myfcf_gra
import my_Functions_cf_stats as myfcf_sta
```

Data Preparation

Import Data

Coffee Price

In [3]:
```
# Import coffee_price.csv into a pandas dataframe.
cfp_df = pd.read_csv('./Data/coffee_price.csv')

# Rename columns of cfp_df.
cfp_df.rename(columns={'date': 'Date', 'value': 'Coffee Price (USD)'}, inplac

# Display the head of cfp_df.
cfp_df.head()
```

Out[3]:

	Date	Coffee Price (USD)
0	20/8/73	0.6735
1	21/8/73	0.6710
2	22/8/73	0.6580
3	23/8/73	0.6675
4	24/8/73	0.6660

Coffee Quality

In [4]:
```
# Import coffee_quality.csv into a pandas dataframe.
cfq_df = pd.read_csv('./Data/coffee_quality.csv')

# Select the relevant columns.
cfq_df = cfq_df[['Species', 'Country.of.Origin', 'Grading.Date', 'Total.Cup.P

# Rename the columns.
cfq_df.rename(columns={'Country.of.Origin': 'Origin',
                       'Grading.Date':'Date',
                       'Total.Cup.Points':'Total Quality Score'}, inplace=Tru
```

Check Data Types

Coffee Price

In [5]:
```
# Display datatypes for cfp_df.
cfp_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12075 entries, 0 to 12074
Data columns (total 2 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Date                12075 non-null  object
 1   Coffee Price (USD)  12075 non-null  float64
dtypes: float64(1), object(1)
memory usage: 188.8+ KB
```

In [6]:
```
# Convert column 'Date' of cfp_df to datetime format.
cfp_df['Date'] = pd.to_datetime(cfp_df['Date'])
```

In [7]:
```
# Display datatypes for cfp_df.
cfp_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12075 entries, 0 to 12074
Data columns (total 2 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Date                12075 non-null  datetime64[ns]
 1   Coffee Price (USD)  12075 non-null  float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 188.8 KB
```

Coffee Quality

In [8]:
```
# Display datatypes for cfq_df.
cfq_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 4 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Species              1339 non-null   object
 1   Origin               1338 non-null   object
 2   Date                 1339 non-null   object
 3   Total Quality Score  1339 non-null   float64
dtypes: float64(1), object(3)
memory usage: 42.0+ KB
```

In [9]:
```
# Convert column 'Date' of cfq_df to datetime format.
cfq_df['Date'] = pd.to_datetime(cfq_df['Date'])

# Sort 'Date' column.
cfq_df.sort_values(by='Date', inplace=True)
```

In [10]:
```
# Convert columns 'Species' and 'Origin' to category format.
cfq_df['Species'] = cfq_df['Species'].astype('category')
cfq_df['Origin'] = cfq_df['Origin'].astype('category')
```

In [11]:

```
# Display datatypes for cfq_df.
cfq_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1339 entries, 1164 to 895
Data columns (total 4 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Species              1339 non-null   category
 1   Origin               1338 non-null   category
 2   Date                 1339 non-null   datetime64[ns]
 3   Total Quality Score  1339 non-null   float64
dtypes: category(2), datetime64[ns](1), float64(1)
memory usage: 35.4 KB
```

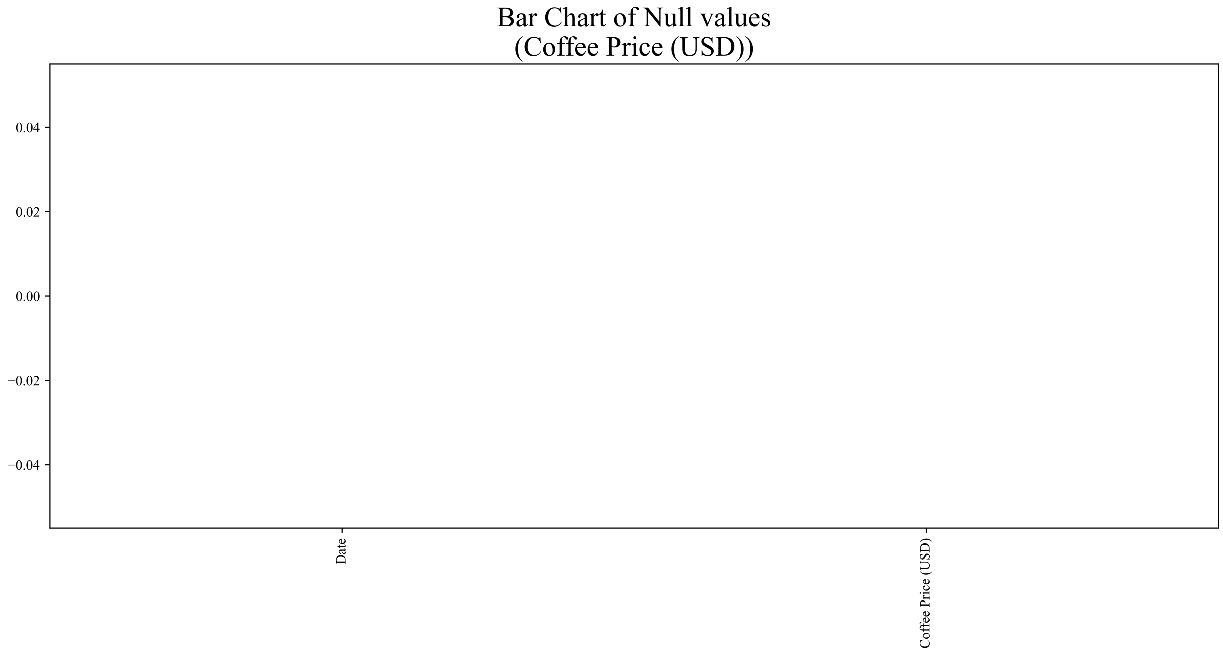
Check for Null Values

Coffee Price

Overview

In [12]:

```
# Check for any null values in cfp_df.
#Input for null_barchart func.
df = cfp_df
category = 'Coffee Price (USD)'

# Run null_barchart func.
myfcf_cle.null_barchart(df, category)
```



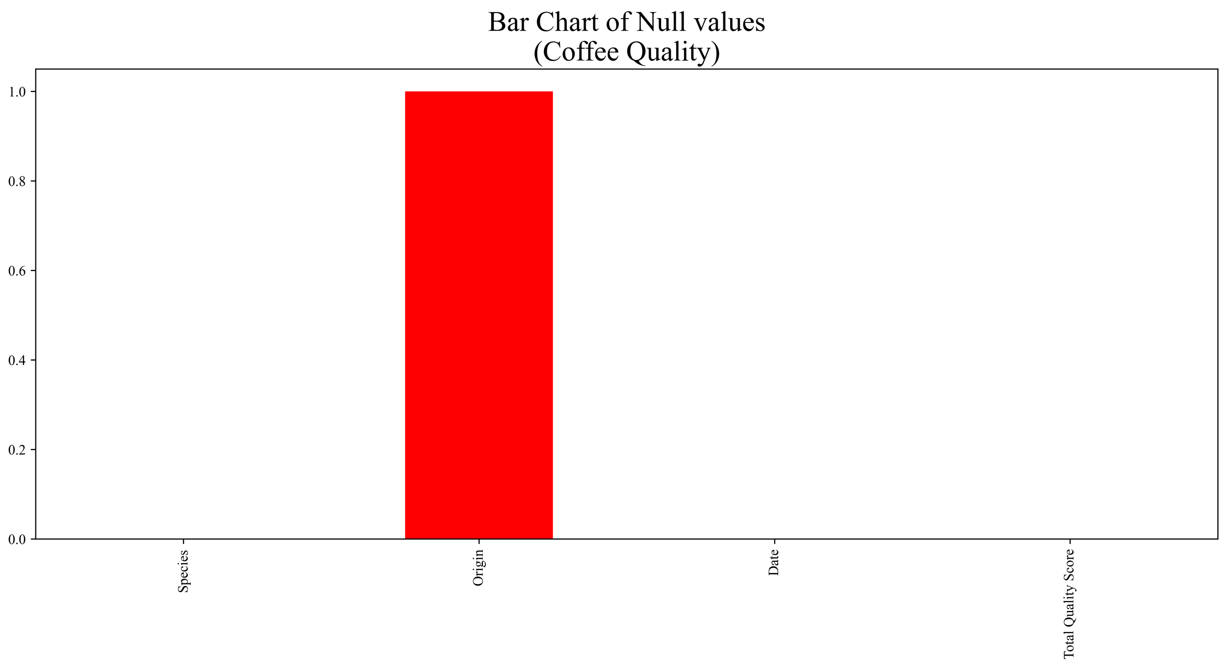
In [13]:
```
# Print the number of null values within the columns of cfp_df.
# Input for print_null_col func.
df = cfp_df

# Run print_null_col func.
myfcf_cle.print_null_col(df)
```

```
Column              No. of Null Values
Date                  0
Coffee Price (USD)    0
dtype: int64
```

Coffee Quality

Overview

In [14]:
```
# Check for any null values in cfq_df.
#Input for null_barchart func.
df = cfq_df
category = 'Coffee Quality'

# Run null_barchart func.
myfcf_cle.null_barchart(df, category)
```



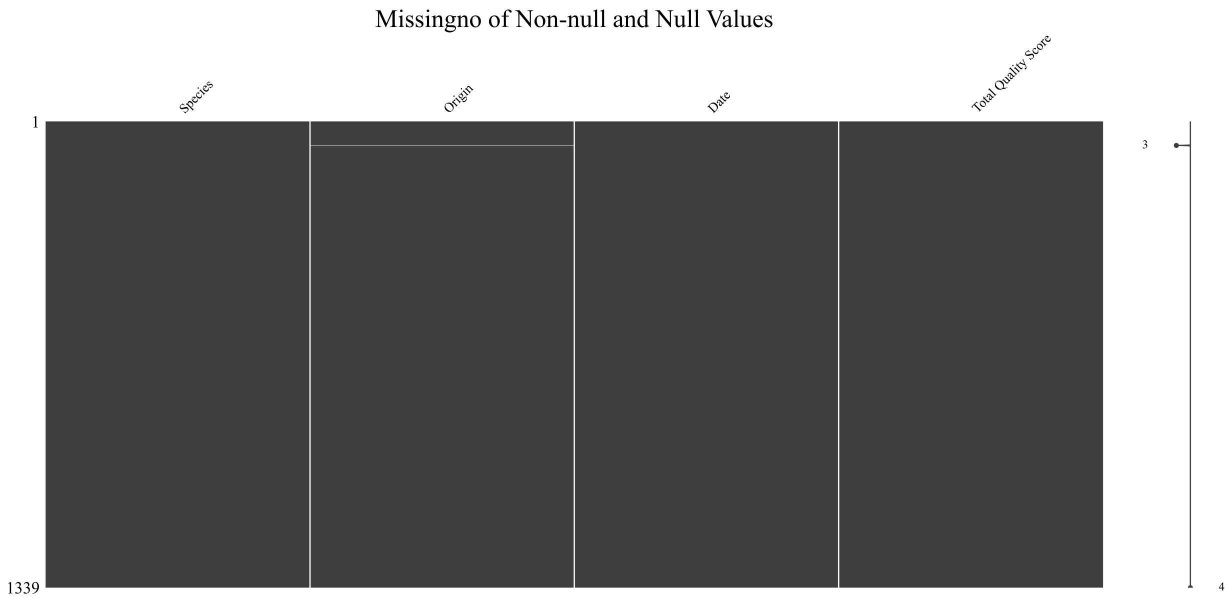
In [15]:
```
# Print the number of null values within the columns of cfq_df.
# Input for print_null_col func.
df = cfq_df

# Run print_null_col func.
myfcf_cle.print_null_col(df)
```

```
Column              No. of Null Values
Species                0
Origin                 1
Date                   0
```

```
Total Quality Score    0
dtype: int64
```

In [16]:

```
# Missingno of non-null and null values of bitcoin_tweets_df.
# Inputs for null_msno func.
df = cfq_df
title_size = 30

# Run null_msno func.
myfcf_cle.null_msno(df, title_size)
```



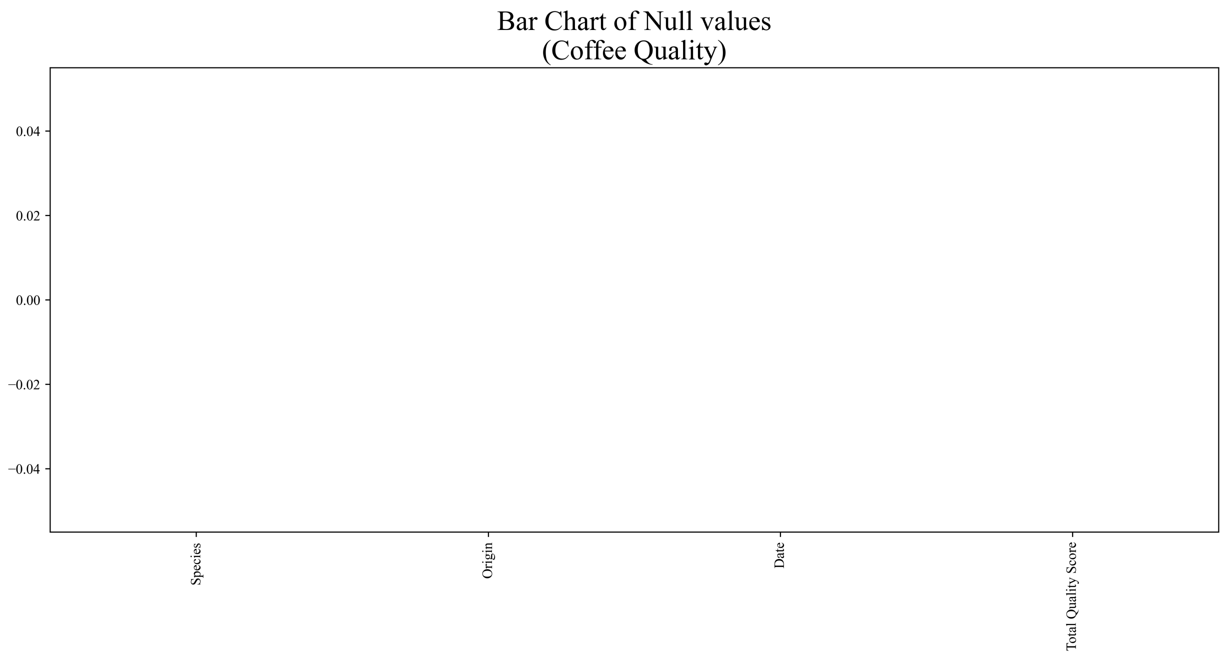
Remove null values

In [17]:

```
# Remove null values for column 'user_name'.
# Inputs for remove_null_row func.
df = cfq_df
col = 'Origin'

# Run remove_null_row func.
myfcf_cle.remove_null_row(df, col)
```

In [18]:

```
# Check for any null values in cfq_df.
#Input for null_barchart func.
df = cfq_df
category = 'Coffee Quality'

# Display null_barchart func.
myfcf_cle.null_barchart(df, category)
```



Data Wrangling

cfq_df

In [19]:

```
# Create column 'Grade' basing on column 'Quality Score'.
cfq_df = cfq_df.assign(Grade=pd.cut(cfq_df['Total Quality Score'],
                                    bins=[-1, 80, 85, 90, 100],
                                    labels=['Below Speciality',
                                            'Very Good',
                                            'Excellent',
                                            'Outstanding']))

# Rename column.
cfq_df.rename(columns={'Grade': 'Total Quality Score Classification'}, inplac

# Display head of cfq_df.
cfq_df.head()
```

Out[19]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification
1164	Arabica	Thailand	2010-04-09	79.67	Below Speciality
1261	Arabica	Mexico	2010-04-09	76.75	Below Speciality
1191	Arabica	Mexico	2010-04-09	79.17	Below Speciality
558	Arabica	Thailand	2010-04-13	82.92	Very Good
532	Arabica	Thailand	2010-04-13	83.00	Very Good

ori_gra_cfq_df

In [20]:
```
# Create a dataframe with by grouping grade and origin.
ori_gra_cfq_df = cfq_df[['Total Quality Score Classification', 'Origin']]
ori_gra_cfq_df = ori_gra_cfq_df.groupby('Total Quality Score Classification')

# Display ori_gra_cfq_df.
ori_gra_cfq_df
```

Out[20]:

	Origin
Total Quality Score Classification	
Below Speciality	194
Very Good	1048
Excellent	95
Outstanding	1

cfpq_df

In [21]:
```
# Merge dataframes cfq_df and cfp_df into cfpq_df.
cfpq_df = cfq_df.merge(cfp_df, on='Date', how='left')

# Display head of cfpq_df.
cfpq_df.head()
```

Out[21]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)
0	Arabica	Thailand	2010-04-09	79.67	Below Speciality	NaN
1	Arabica	Mexico	2010-04-09	76.75	Below Speciality	NaN
2	Arabica	Mexico	2010-04-09	79.17	Below Speciality	NaN
3	Arabica	Thailand	2010-04-13	82.92	Very Good	1.322
4	Arabica	Thailand	2010-04-13	83.00	Very Good	1.322

cfpq_a30_df

In [22]:
```
# Filter for non-null values of column 'Coffee Price (USD)' of cfpq_df.
# Identify null values in column 'Coffee Price (USD)' of cfpq_df.
is_null = cfpq_df.isna()

# Filter for non-null values of column 'Coffee Price (USD)' of cfpq_df.
cfpq_df = cfpq_df[is_null['Coffee Price (USD)'] == False]
```

In [23]:
```
# Filter for arabica coffee.
cfpq_df = cfpq_df[cfpq_df['Species'] == 'Arabica']
```

In [24]:

```
# Identify countries with more than 30 data points except for ethopia.
# Input for ctry_a30_lt func.
df = cfpq_df

# Run ctry_a30_lt func.
ctry_a30_lt = myfcf_wra.ctry_a30_lt(df)

# Remove ethopia from the list, because there is not much coffee future contr
ctry_a30_lt = [i for i in ctry_a30_lt if i != 'Ethiopia']
```

In [25]:

```
# Create a dataframe with countries with more than 30 data points except for
# Input for ctry_a30_df func.
df = cfpq_df

# Run ctry_a30_df func.
cfpq_a30_df = myfcf_wra.ctry_a30_df(df, ctry_a30_lt)

# Display head of cfpq_a30_df.
cfpq_a30_df.head()
```

Out[25]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370

In [26]:

```
# Add column 'Year' into cfpq_a30_df.
# Inputs for yr_lt func.
df = cfpq_a30_df
col = 'Date'

# Run yr_lt func.
cfpq_a30_df['Year'] = myfcf_wra.yr_lt(df, col)

# Display head of cfpq_a30_df.
cfpq_a30_df.head()
```

Out[26]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010

In [27]:

```
# Identify countries that have >= 30 data points except for ethopia.
# Input for ctry_a30_lt func.
df = cfpq_a30_df

# Run ctry_a30_lt func.
ctry_a30_lt = myfcf_wra.ctry_a30_lt(df)

# Run print_ctry_a30 func.
myfcf_wra.print_ctry_a30(ctry_a30_lt)
```

```
Countries that have >= 30 data points excluding Ethopia:
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
```

In [28]:

```
# Change format to obj.
cfpq_a30_df['Origin'] = cfpq_a30_df['Origin'].astype('str')

# Filter countries based on >= 30 data points except for ethopia.
cfpq_a30_df = cfpq_a30_df[(cfpq_a30_df['Origin'] == 'Brazil') |
                          (cfpq_a30_df['Origin'] == 'Colombia') |
                          (cfpq_a30_df['Origin'] == 'Costa Rica') |
                          (cfpq_a30_df['Origin'] == 'Guatemala') |
                          (cfpq_a30_df['Origin'] == 'Honduras') |
                          (cfpq_a30_df['Origin'] == 'Mexico') |
                          (cfpq_a30_df['Origin'] == 'Taiwan') |
                          (cfpq_a30_df['Origin'] == 'Tanzania, United Republi
                          (cfpq_a30_df['Origin'] == 'United States (Hawaii)')

# Change format back to category.
cfpq_a30_df['Origin'] = cfpq_a30_df['Origin'].astype('category')
```

In [29]:

```
# Find the start and end dates of the cfpq_a30_df.
# Inputs for start_end_dates_tp func.
df = cfpq_a30_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)

# Print start and end dates of cfq_df.
print('Start Date: ' + start_date)
print('End Date: ' + end_date)
```

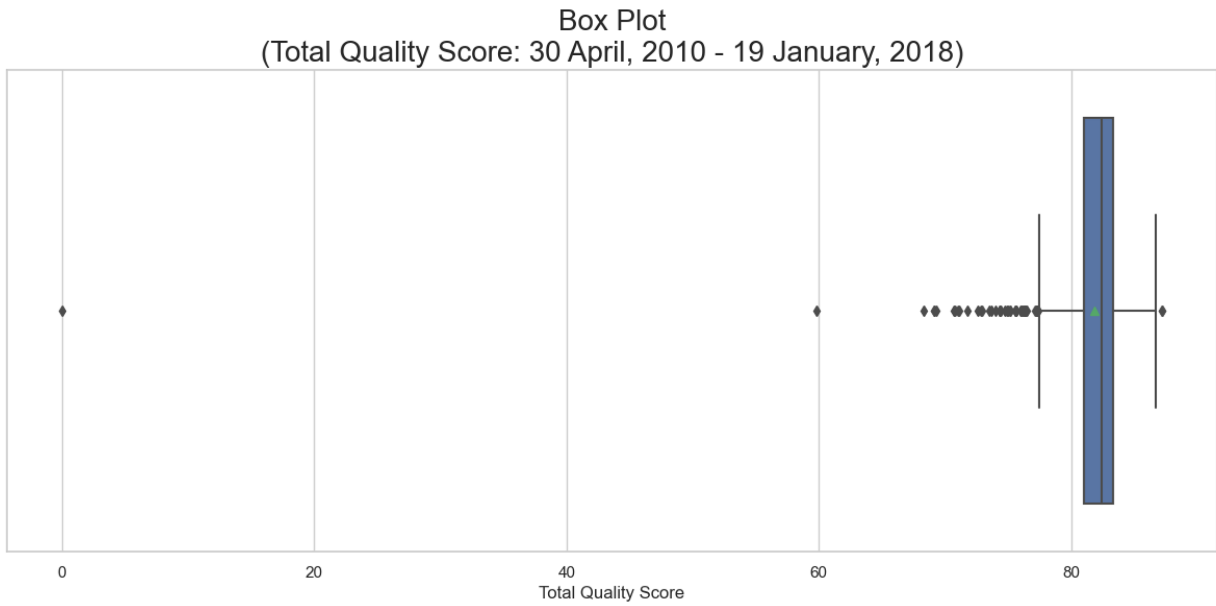
```
Start Date: 2010-04-30
```

```
End Date: 2018-01-19
```

Identify extreme outliers

In [30]:

```
# Create a boxplot for total quality score before dropping outliers.
# Inputs for h_boxplot func.
df = cfpq_a30_df
col = 'Total Quality Score'

# Run h_boxplot func.
myfcf_sta.h_boxplot(df, col, start_date, end_date)
```



Remove extreme outliers

In [31]:

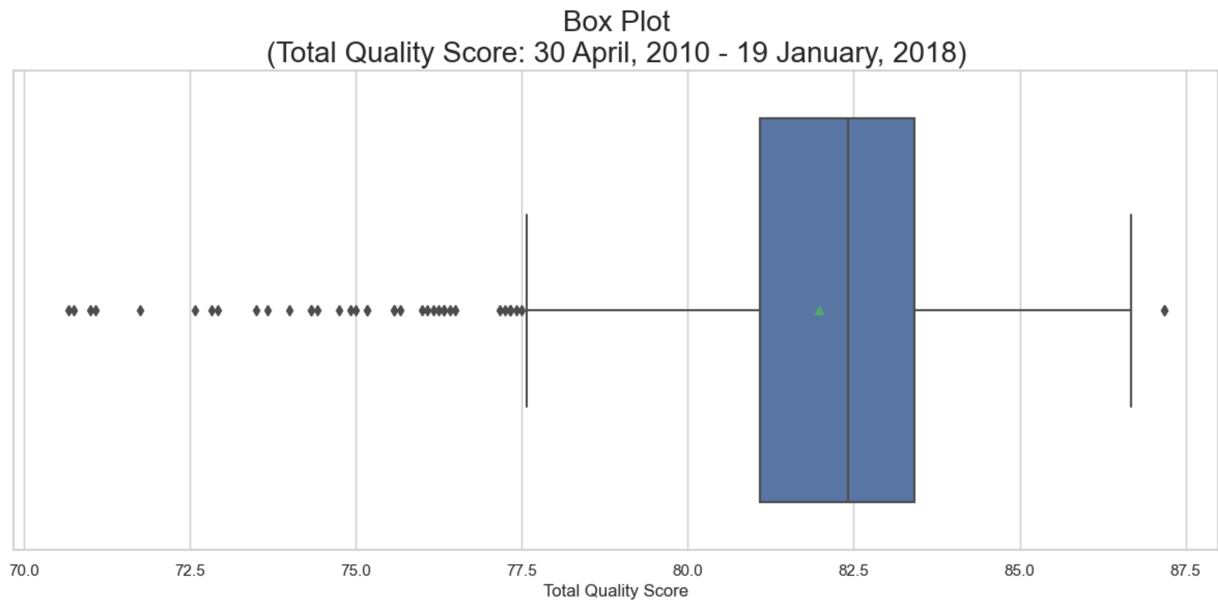
```
# Create a  tuple of inliner index, outlier index, number of inliners, and nu
# Inputs for thold_inliner_outlier_tp func.
df = cfpq_a30_df
col = 'Total Quality Score'
thold = 3 # 3 means 3 standard deviations

# Run thold_inliner_outlier_tp func.
in_ind, out_ind, in_no, out_no = myfcf_wra.thold_inliner_outlier_tp(df, col,
```

In [32]:

```
# Drop rows of extreme outliers.
cfpq_a30_df = cfpq_a30_df.drop(index=out_ind[0])
```

In [33]:

```
# Create a boxplot for total quality score after dropping outliers.
# Inputs for h_boxplot func.
df = cfpq_a30_df
col = 'Total Quality Score'

# Run h_boxplot func.
myfcf_sta.h_boxplot(df, col, start_date, end_date)
```



cfpq_a30_mt_df

In [34]:
```
# Perform melt function and sort based on 'Origin' on cfpq_a30_mt_df.
cfpq_a30_mt_df = pd.melt(cfpq_a30_df, id_vars=['Origin'], value_vars=['Total

# Display head of cfpq_a30_mt_df.
cfpq_a30_mt_df.head()
```

Out[34]:

	Origin	variable	value
659	Brazil	Total Quality Score	82.42
663	Brazil	Total Quality Score	82.33
664	Brazil	Total Quality Score	81.25
665	Brazil	Total Quality Score	82.25
666	Brazil	Total Quality Score	83.33

bra_df, co_df, cr_df, gt_df, hond_df, mx_df, tw_df, urt_df, hi_df

In [35]:

```
# Create dataframe for brazil.
bra_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Brazil']

# Create dataframe for colombia.
co_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Colombia']

# Create dataframe for costa rica.
cr_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Costa Rica']

# Create dataframe for guatemala.
gt_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Guatemala']

# Create dataframe for honduras.
hond_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Honduras']

# Create dataframe for mexico.
mx_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Mexico']

# Create dataframe for taiwan.
tw_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Taiwan']

# Create dataframe for tanzania, united republic of.
urt_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'Tanzania, United Republic Of']

# Create dataframe for united states (hawaii).
hi_df = cfpq_a30_df[cfpq_a30_df['Origin'] == 'United States (Hawaii)']
```

out_df, exc_df, vg_df, bs_df

In [36]:

```
# Create dataframe for outstanding.
out_df = cfpq_a30_df[cfpq_a30_df['Total Quality Score Classification'] == 'Ou

# Create dataframe for excellent.
exc_df = cfpq_a30_df[cfpq_a30_df['Total Quality Score Classification'] == 'Ex

# Create dataframe for very good.
vg_df = cfpq_a30_df[cfpq_a30_df['Total Quality Score Classification'] == 'Ver

# Create dataframe for below speciality.
bs_df = cfpq_a30_df[cfpq_a30_df['Total Quality Score Classification'] == 'Bel
```

Descriptive Statistics

In [37]:

```
# Descriptive statistics of coffee price.
cfp_df.describe()
```

Out[37]:

	Coffee Price (USD)
count	12075.000000
mean	1.261351
std	0.474009

	Coffee Price (USD)
min	0.425000
25%	0.945250
50%	1.234500
75%	1.464000

In [38]:

```
# Descriptive statistics of total quality score.
cfpq_a30_df[['Total Quality Score']].describe()
```

Out[38]:

	Total Quality Score
count	857.000000
mean	81.978075
std	2.281264
min	70.670000
25%	81.080000
50%	82.420000
75%	83.420000
max	87.170000

EDA

Coffee Price

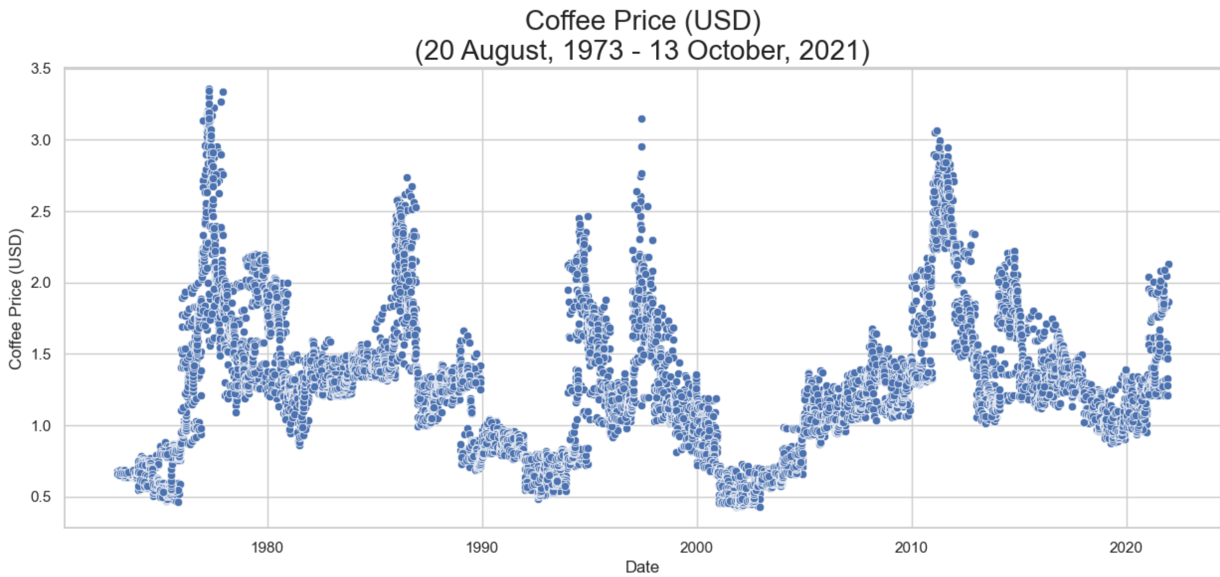
In [39]:

```
# Find the start and end dates of the cfp_df.
# Inputs for start_end_dates_tp func.
df = cfp_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)

# Print start and end dates of cfp_df.
print('Start Date: ' + start_date)
print('End Date: ' + end_date)
```

```
Start Date: 1973-08-20
End Date: 2021-10-13
```

In [40]:

```
# Plot a scatterplot for coffee price.
# Inputs for scatterplot func.
df = cfp_df
x = 'Date'
y = 'Coffee Price (USD)'

# Display scatterplot func.
myfcf_gra.scatterplot(df, x, y, start_date, end_date)
```



Coffee Quality

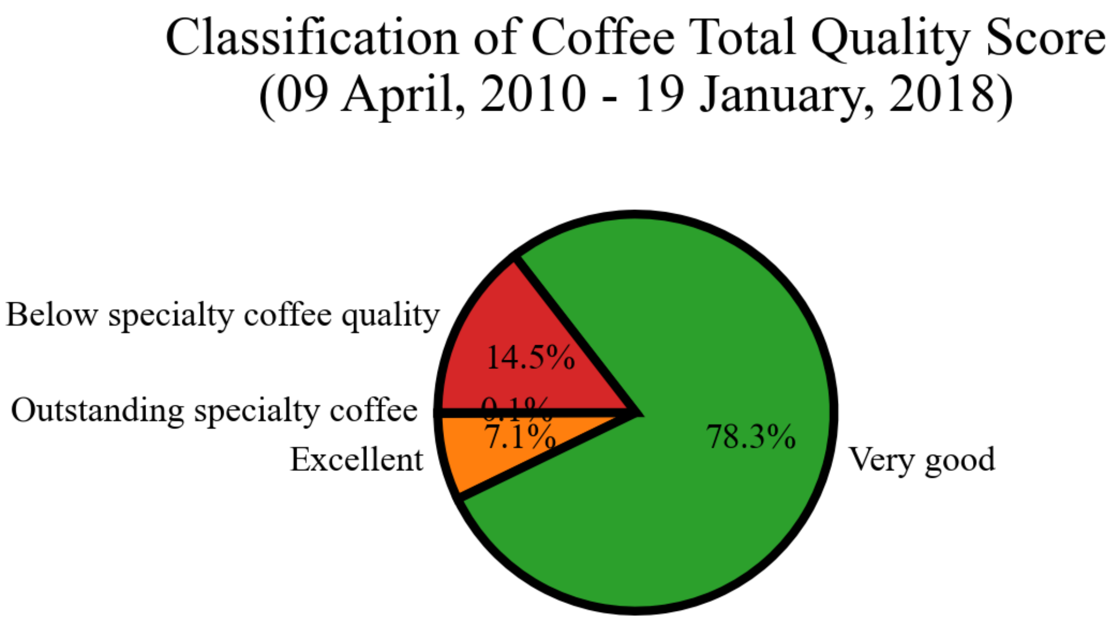
In [41]:

```
# Find the start and end dates of the cfq_df.
# Inputs for start_end_dates_tp func.
df = cfq_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)

# Print start and end dates of cfq_df.
print('Start Date: ' + start_date)
print('End Date: ' + end_date)
```

```
Start Date: 2010-04-09
End Date: 2018-01-19
```

In [42]:

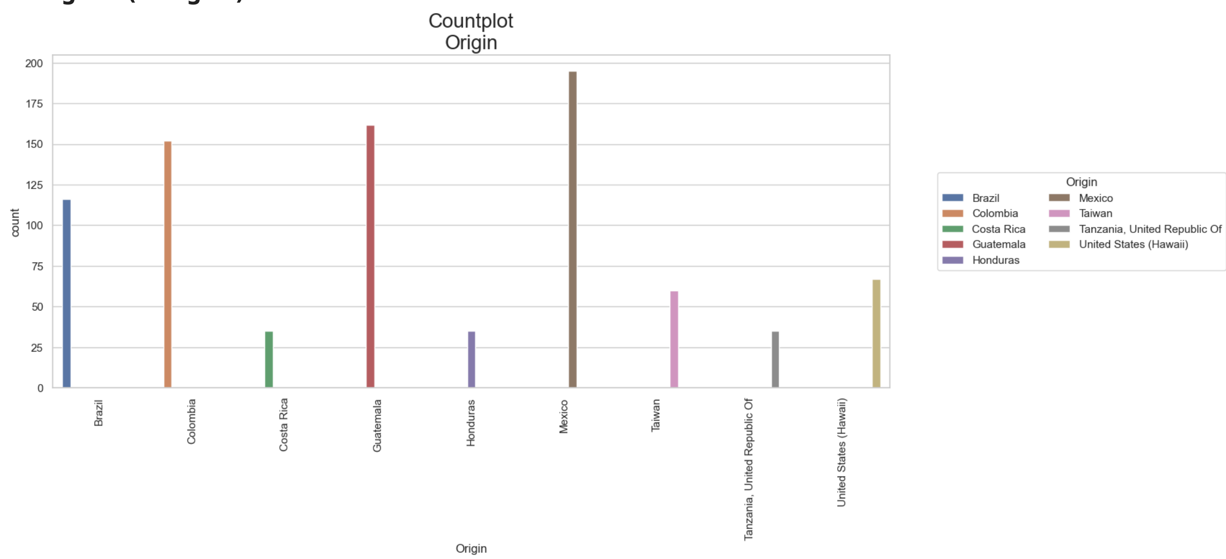
```
# Plot pie chart for coffee grade score.
# Input for cfg_piechart.
df = ori_gra_cfq_df

# Run cfg_piechart.
myfcf_gra.cfg_piechart(df, start_date, end_date)
```



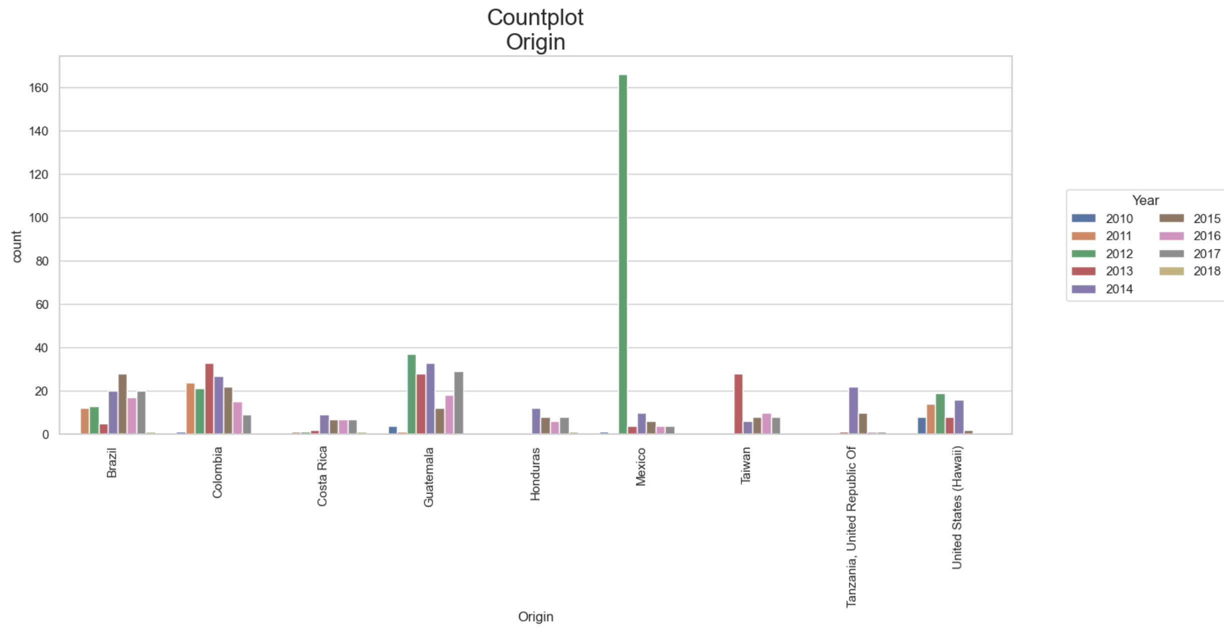
In [43]:

```
# Inputs for print_mult_countplot func.
df = cfpq_a30_df
x_lt = ['Origin', 'Year', 'Total Quality Score Classification']
hue_lt = ['Origin', 'Year', 'Total Quality Score Classification']

# Run print_mult_countplot func.
myfcf_gra.print_mult_countplot(df, x_lt, hue_lt)
```

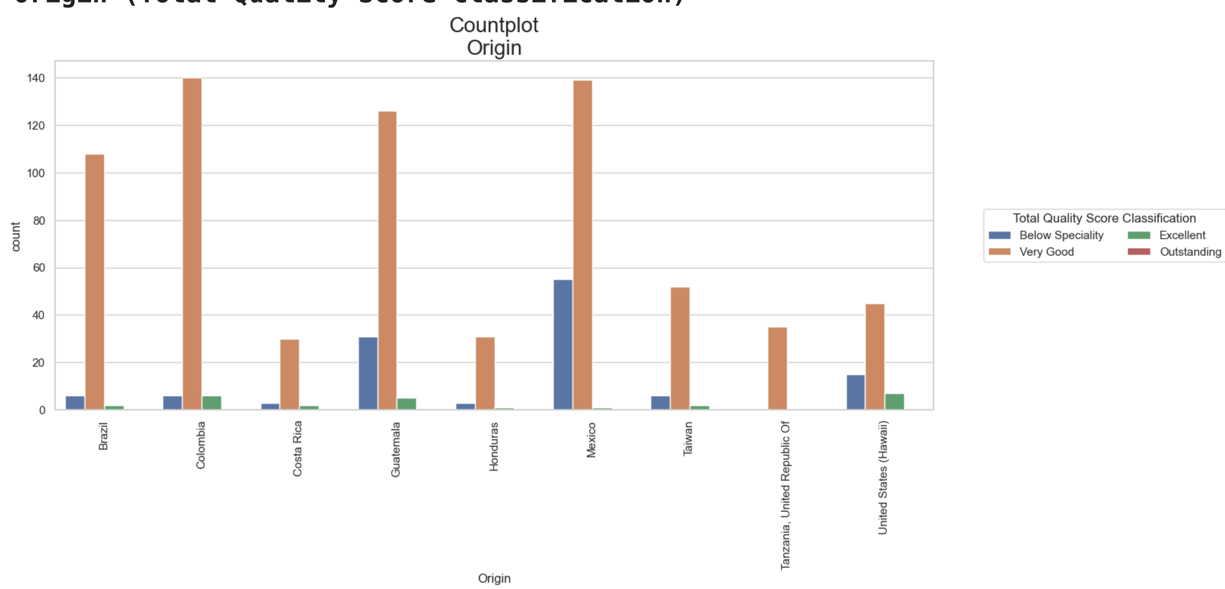
Origin (Origin)



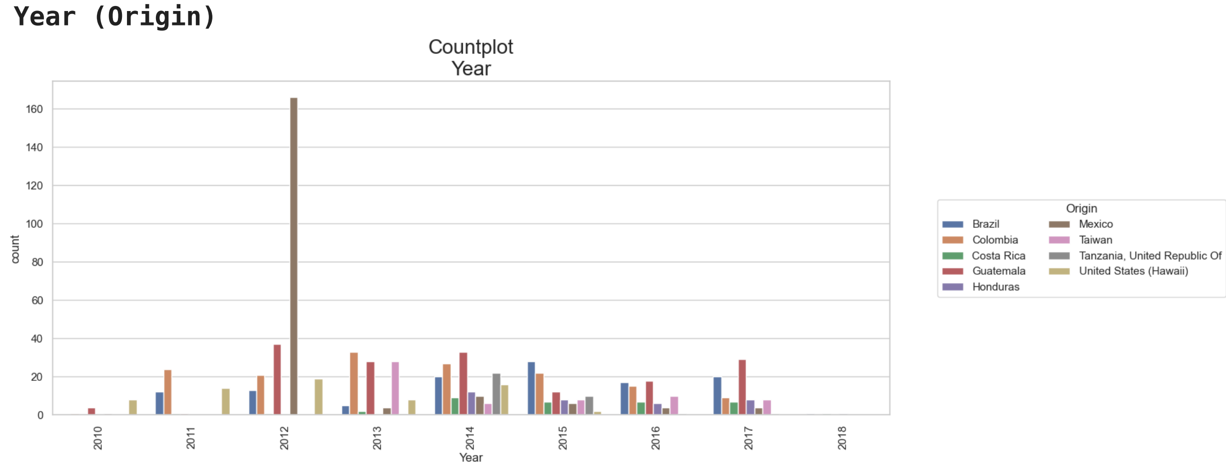
Origin (Year)



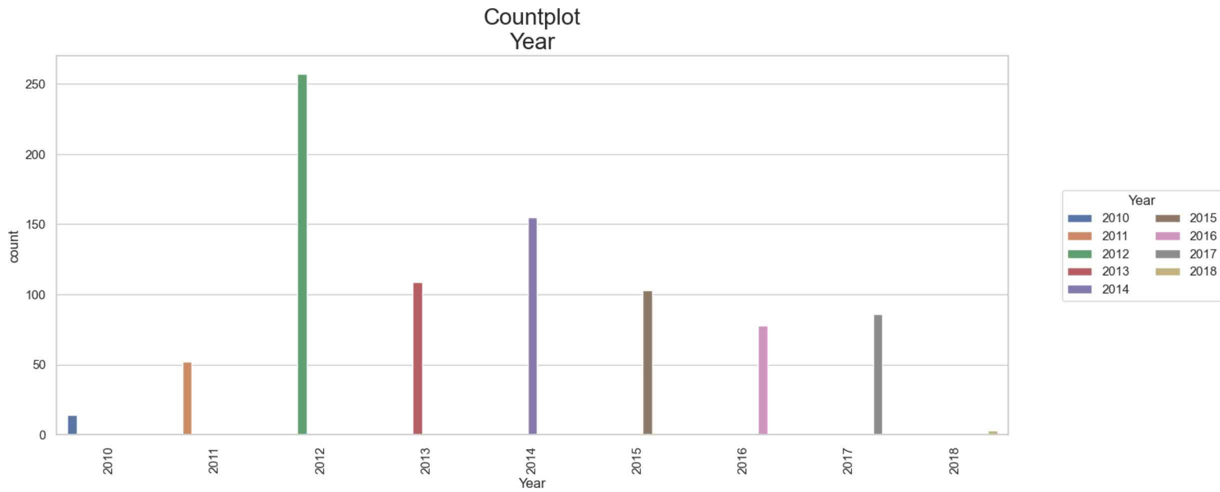
Origin (Total Quality Score Classification)



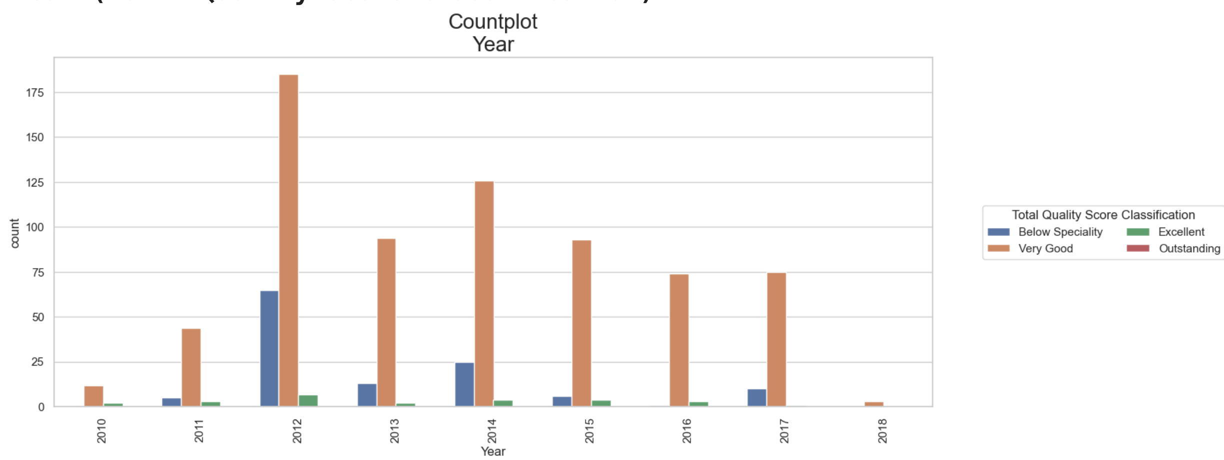
Year (Origin)



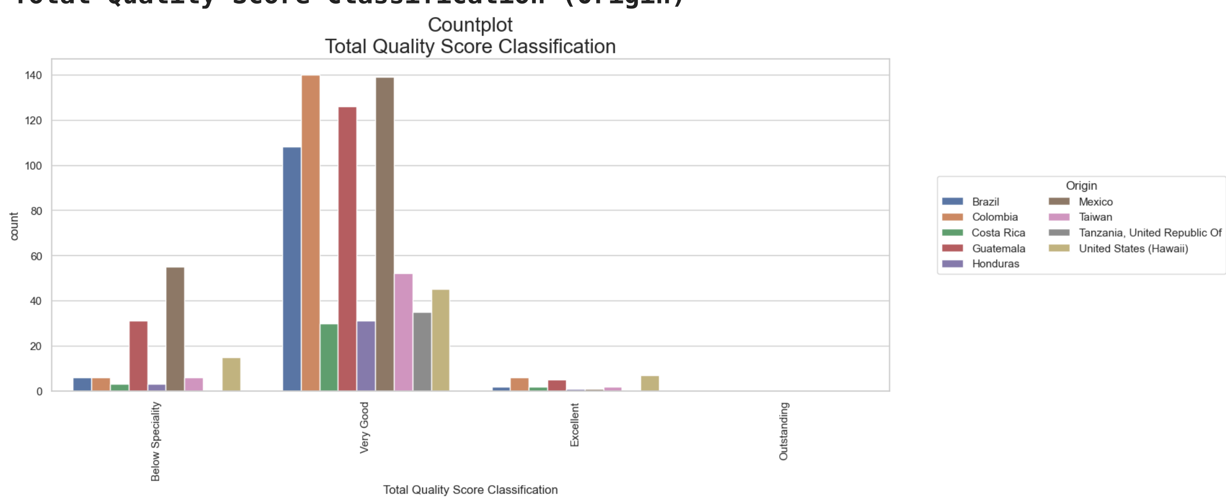
Year (Year)



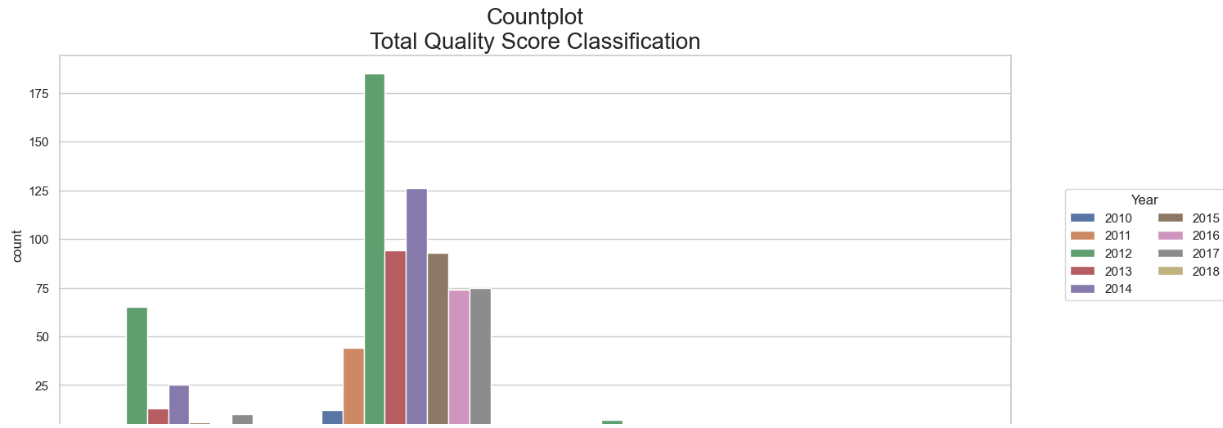
Year (Total Quality Score Classification)



Total Quality Score Classification (Origin)



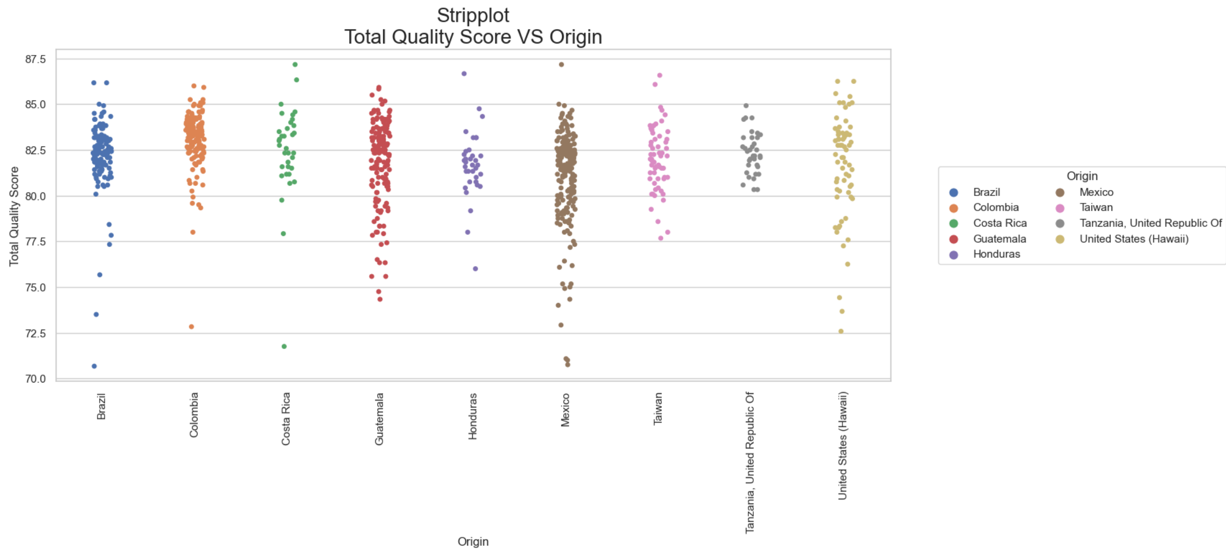
Total Quality Score Classification (Year)

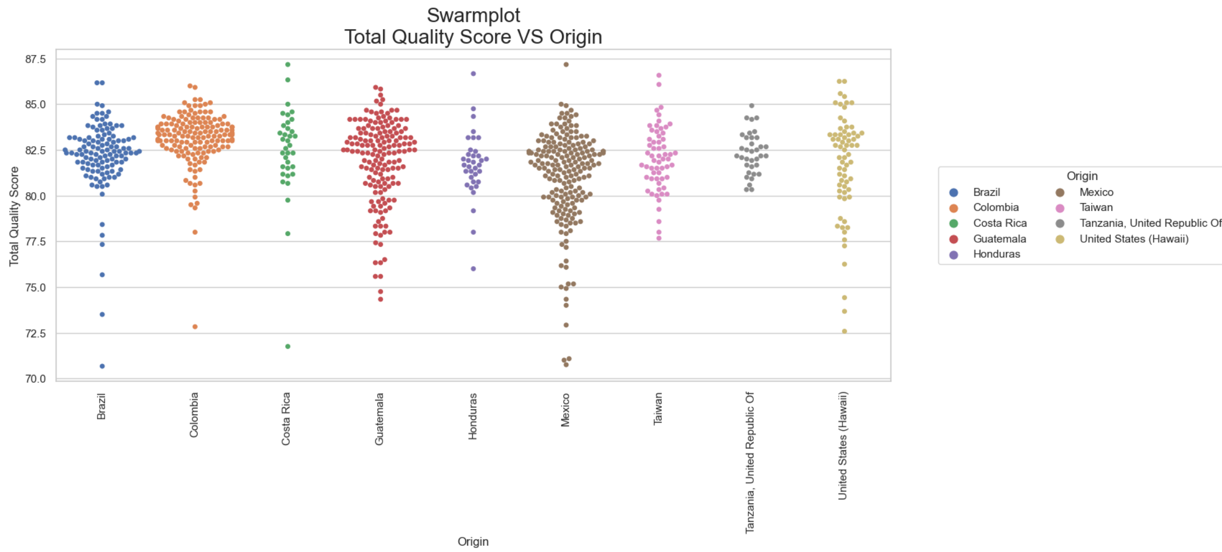


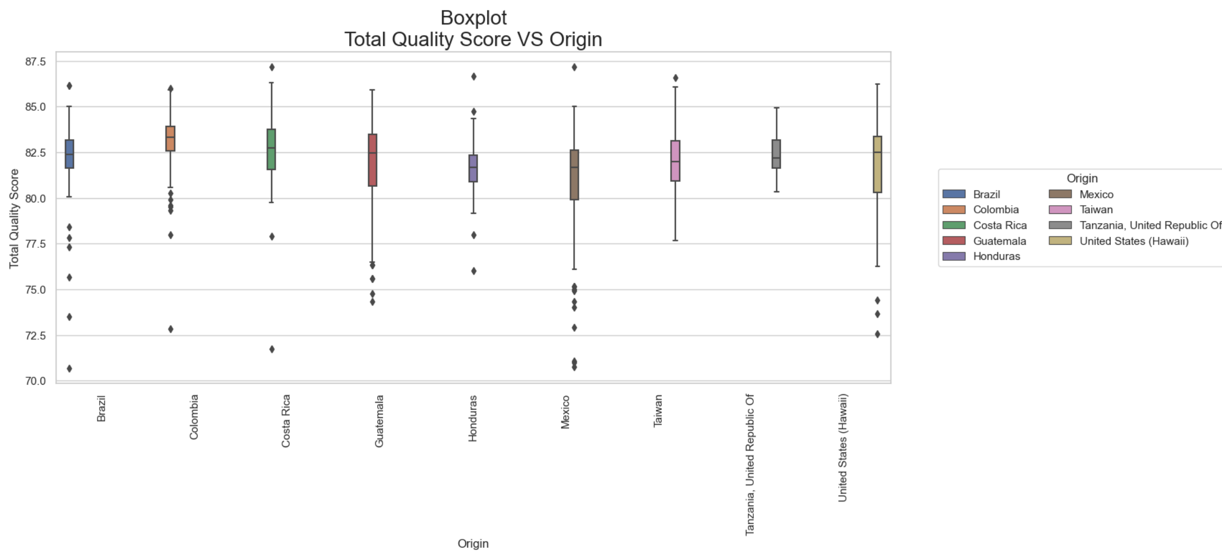
In [44]:

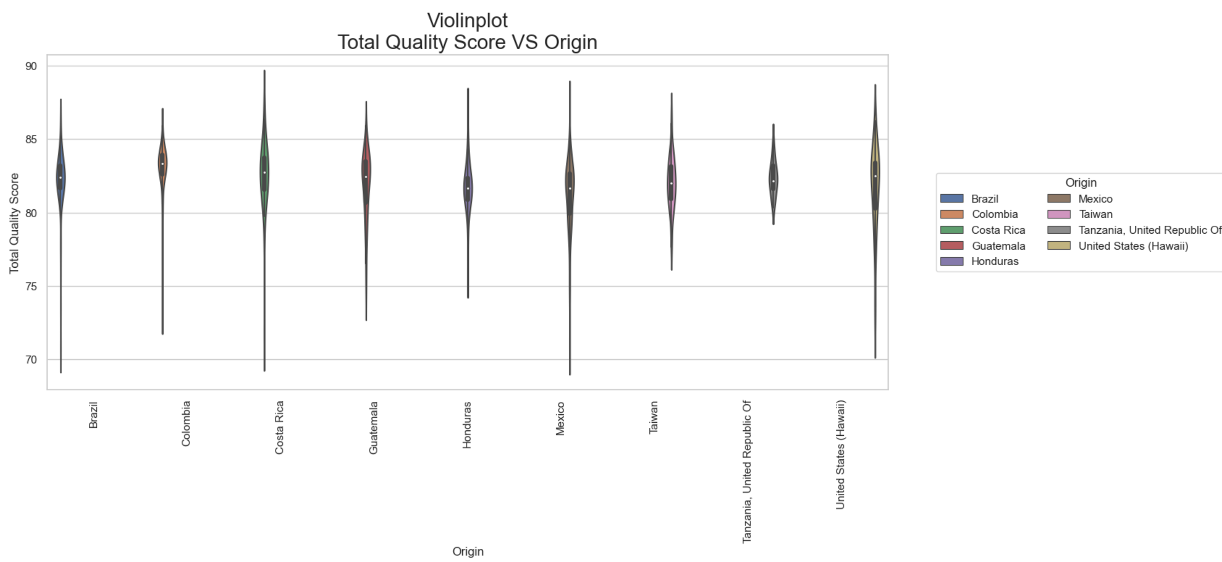
```
# Inputs for print_mult_catplots func.
df = cfpq_a30_df
x_lt = ['Origin', 'Year', 'Total Quality Score Classification']
y_lt = ['Total Quality Score', 'Coffee Price (USD)']
hue_lt = x_lt

# Run print_mult_catplots func.
myfcf_gra.print_mult_catplots(df, x_lt, y_lt, hue_lt)
```

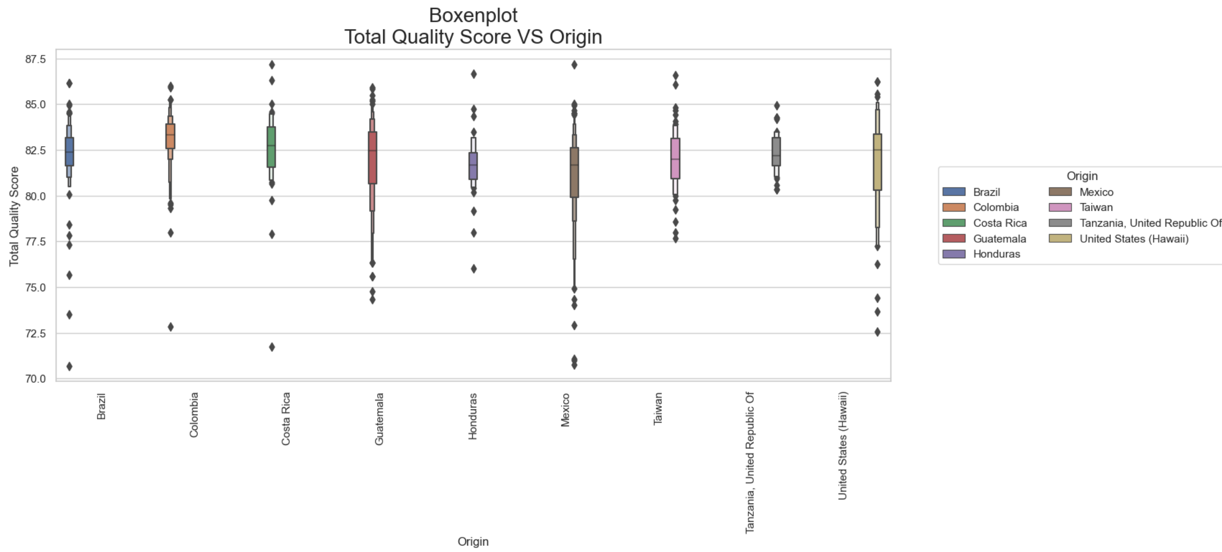
Total Quality Score VS Origin (Origin)



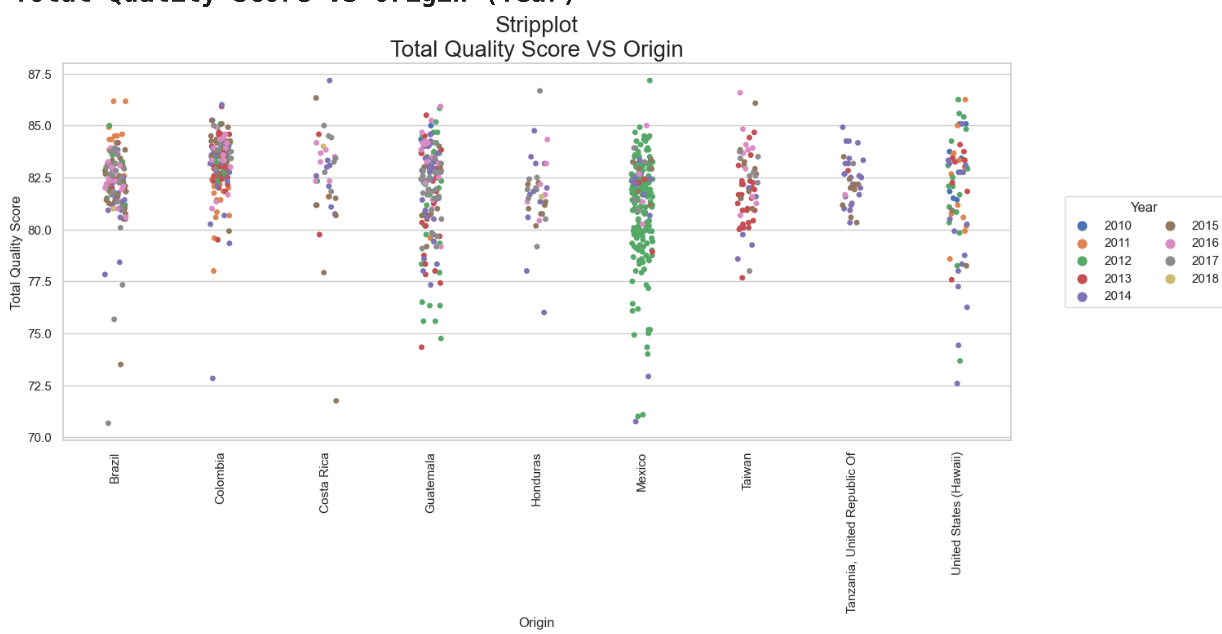

Swarmplot
Total Quality Score VS Origin
Total Quality Score
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

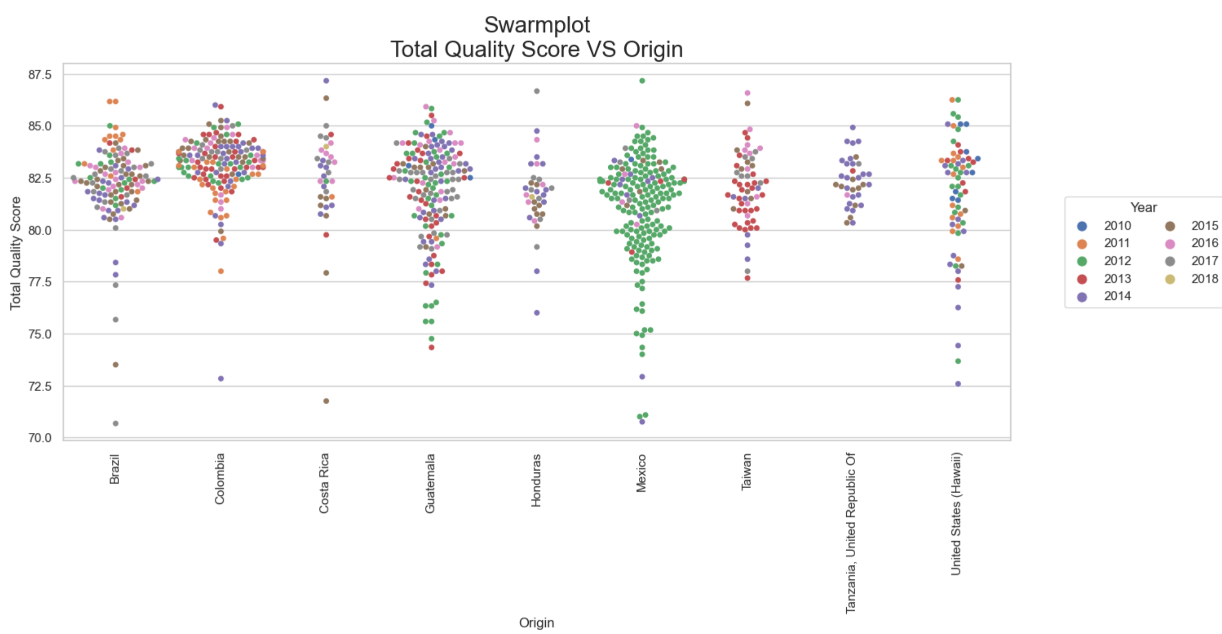

Boxplot
Total Quality Score VS Origin
Total Quality Score
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

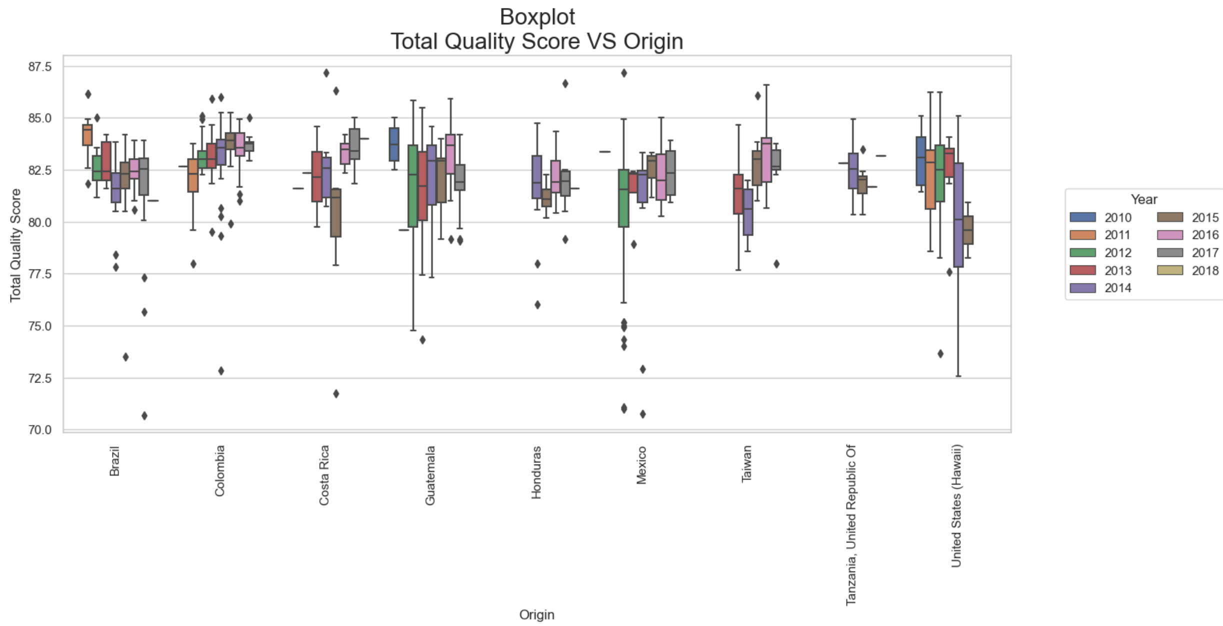

Violinplot
Total Quality Score VS Origin
Total Quality Score
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

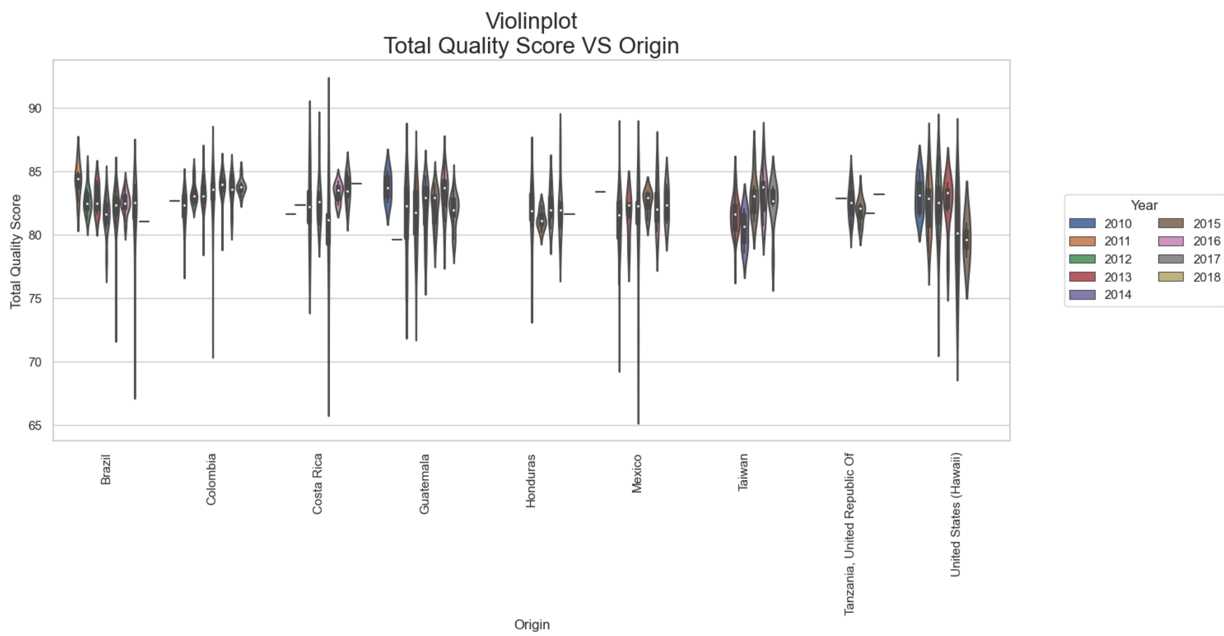


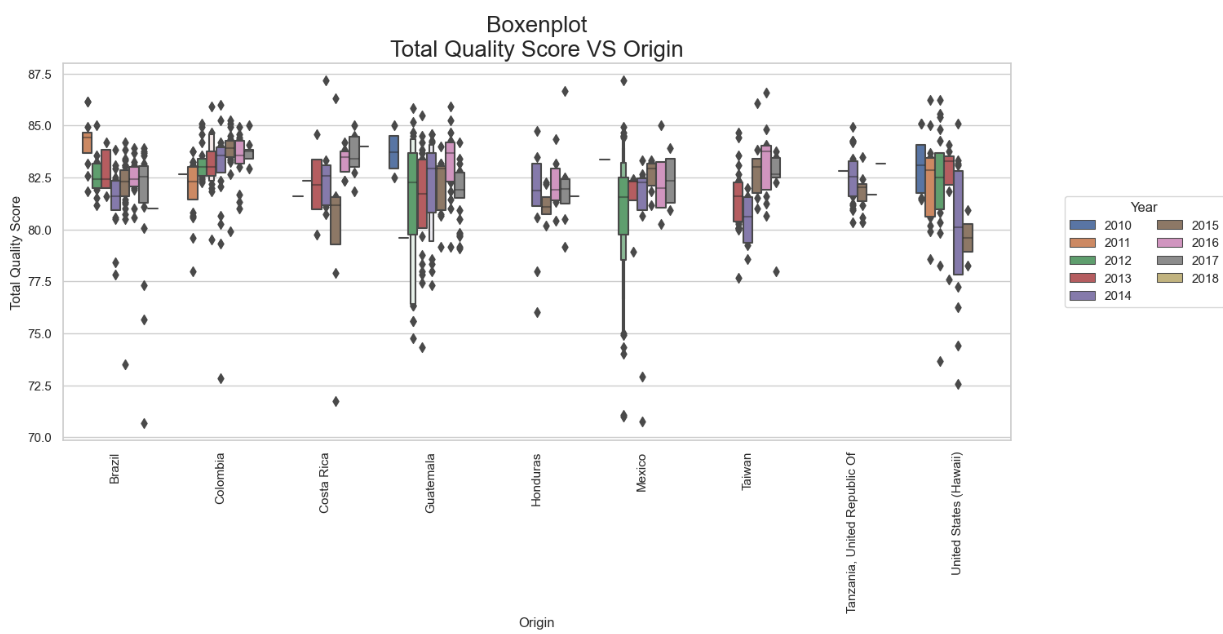
Total Quality Score VS Origin (Year)



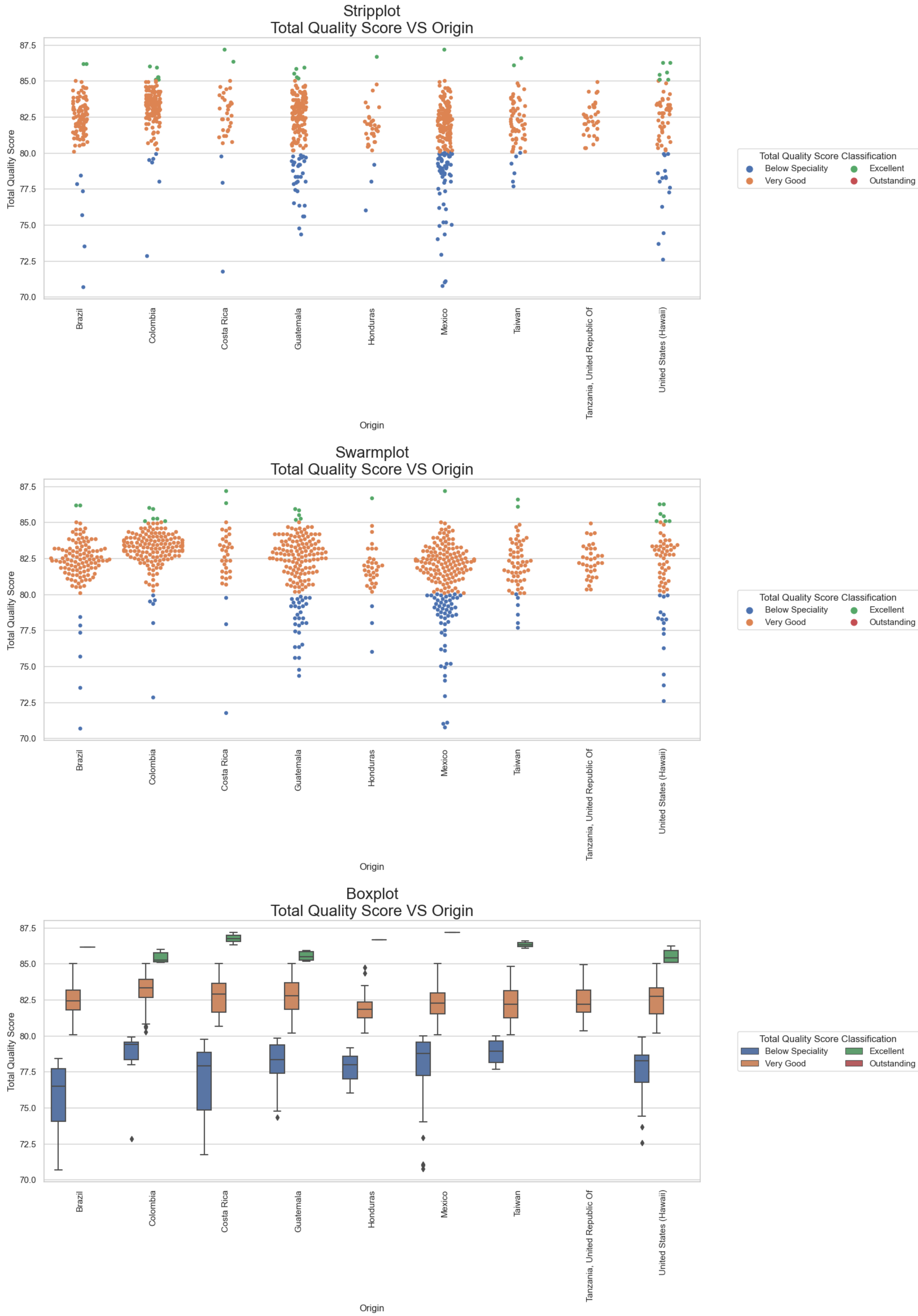


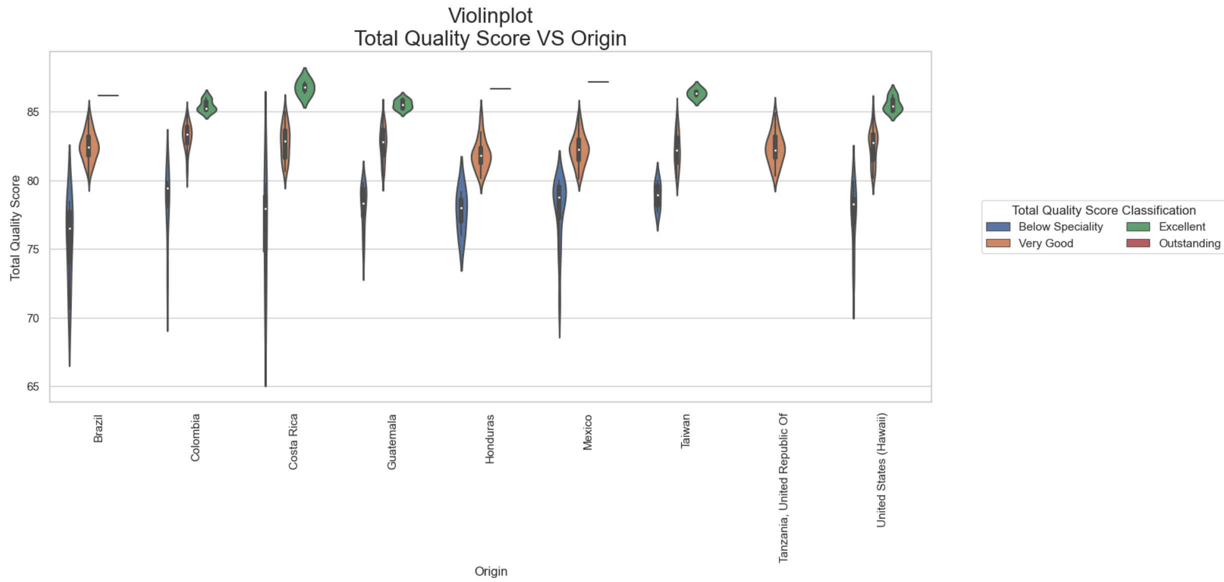

Boxplot
Total Quality Score VS Origin
Total Quality Score
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

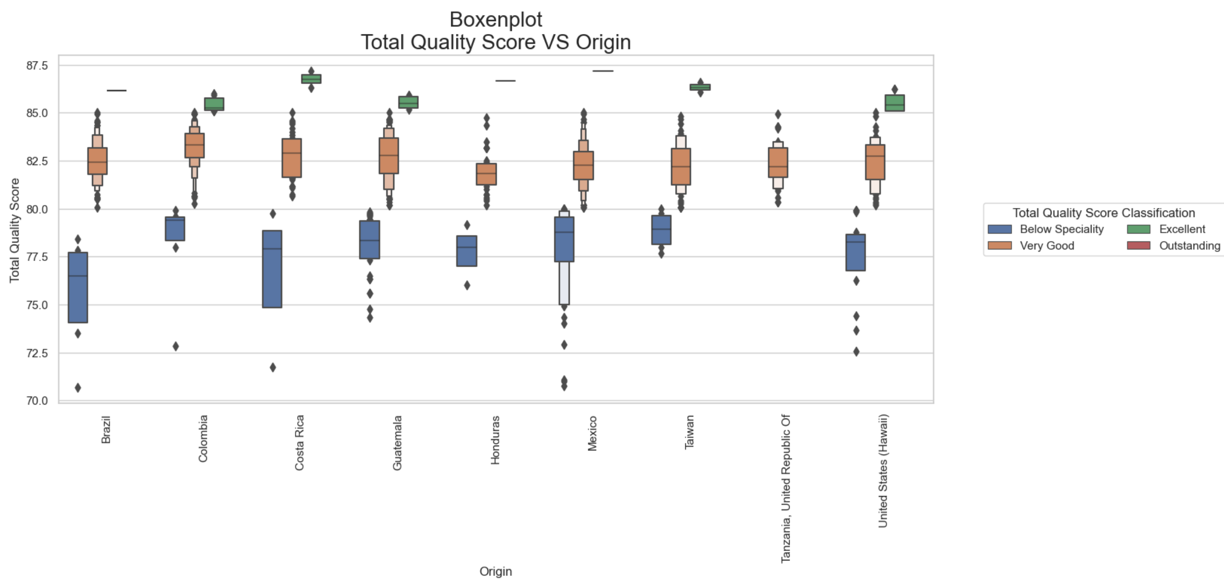

Violinplot
Total Quality Score VS Origin
Total Quality Score
90
85
80
75
70
65
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018


Boxenplot
Total Quality Score VS Origin
Total Quality Score
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

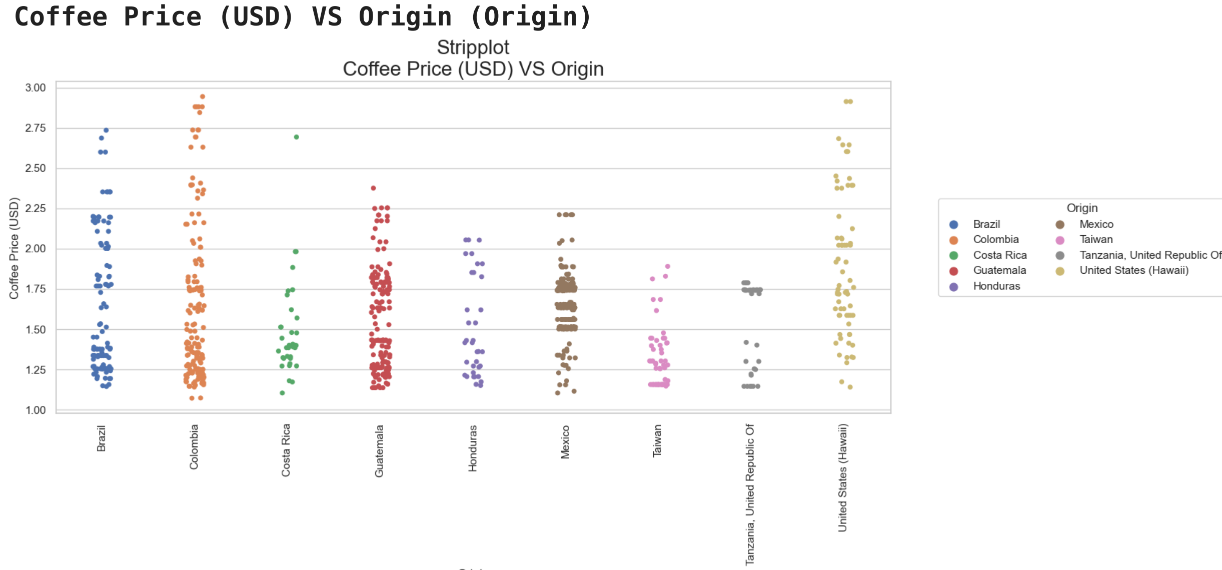
Total Quality Score VS Origin (Total Quality Score Classification)

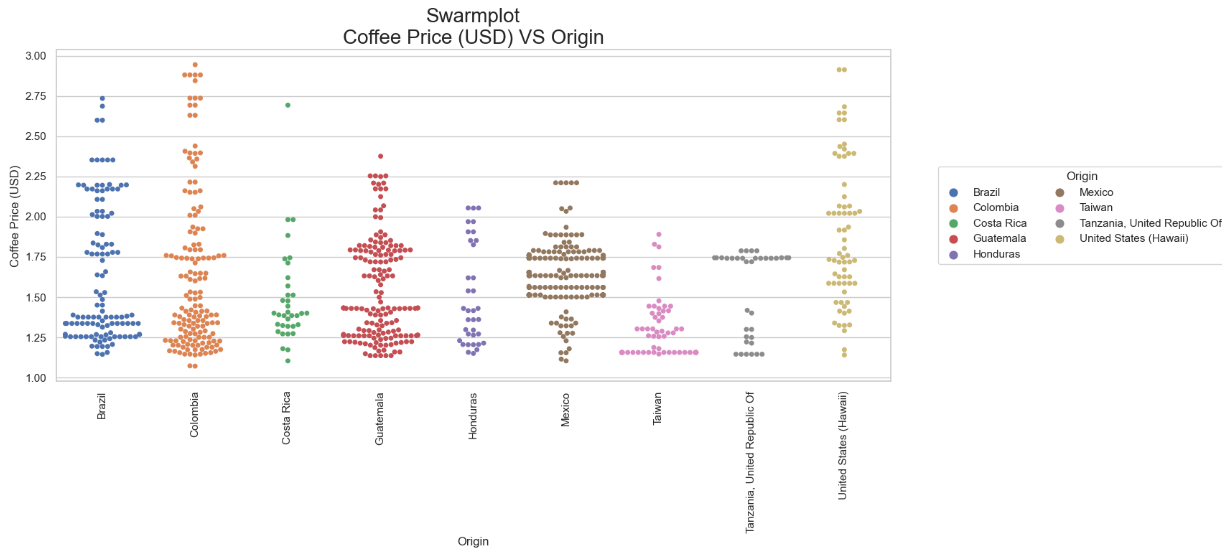


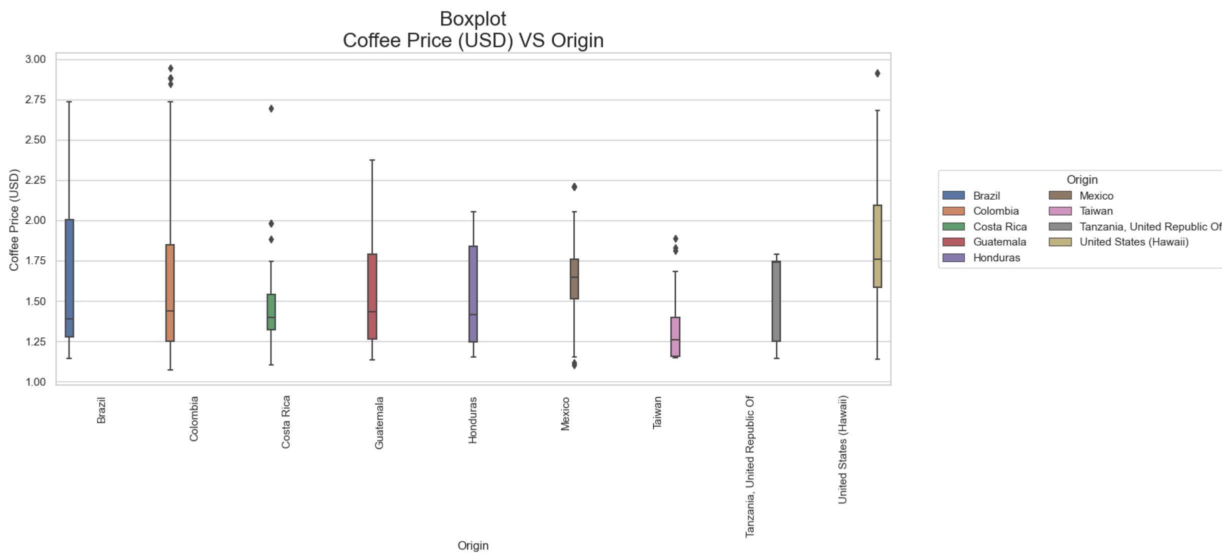


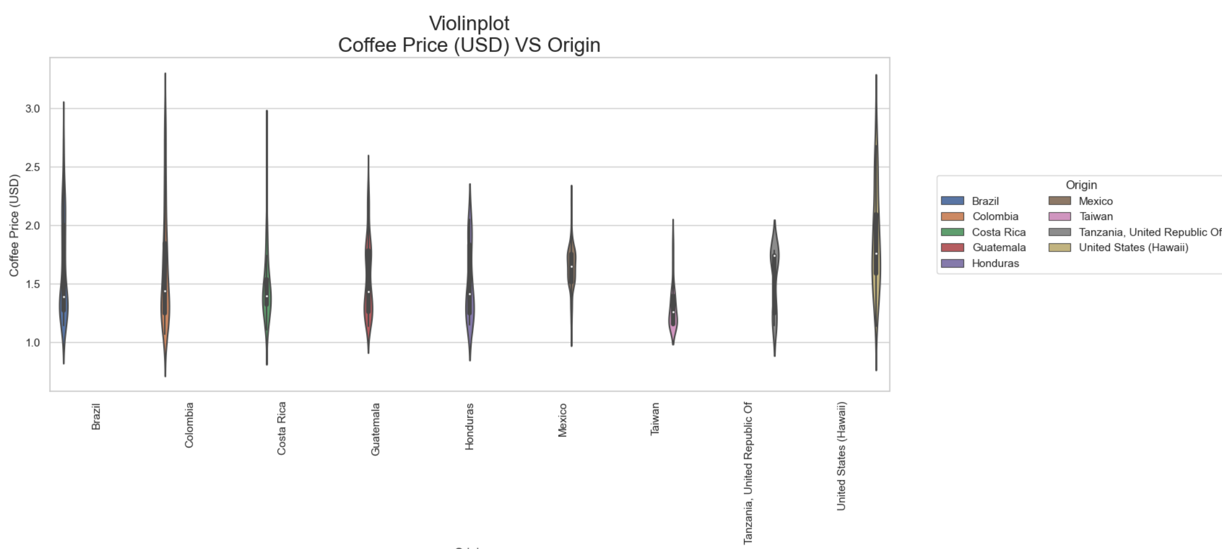


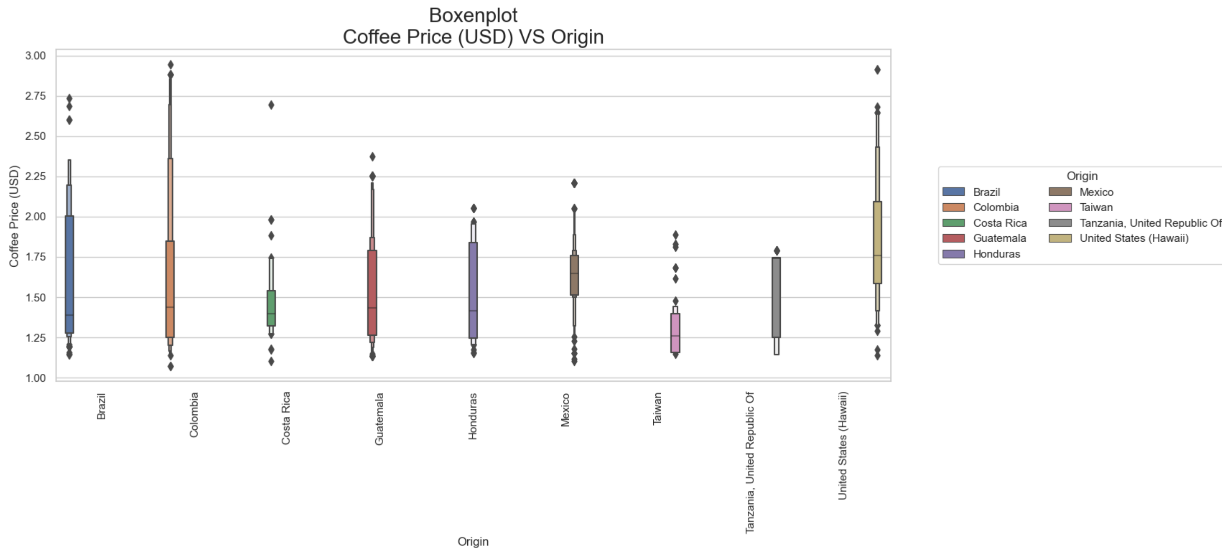
Coffee Price (USD) VS Origin (Origin)



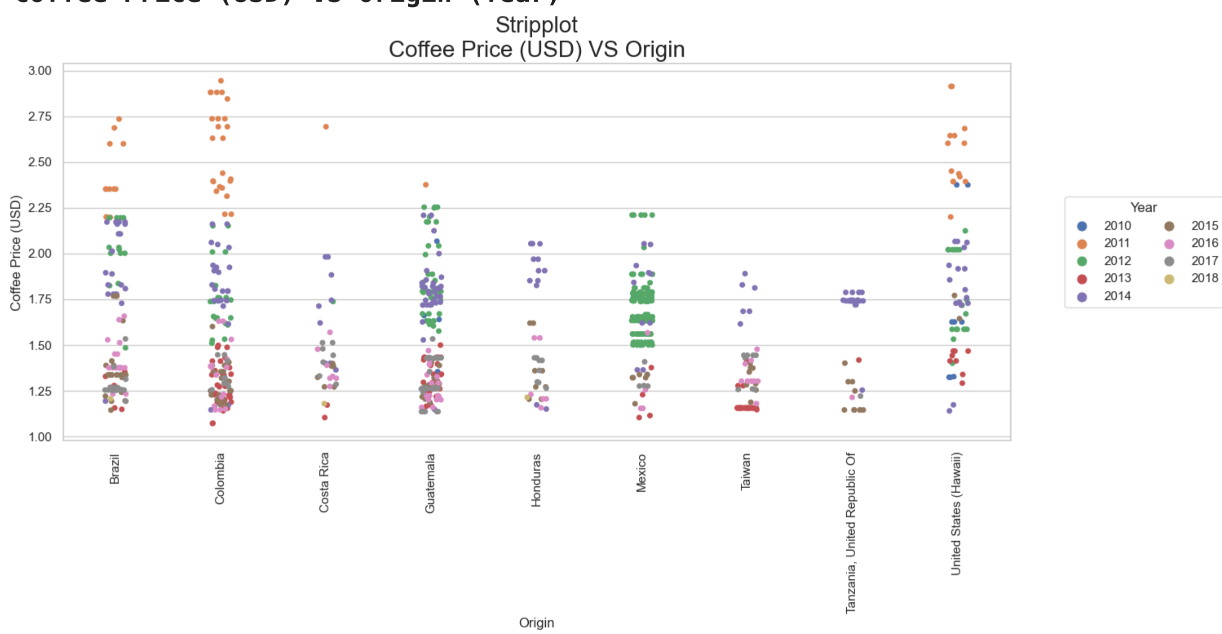

Swarmplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

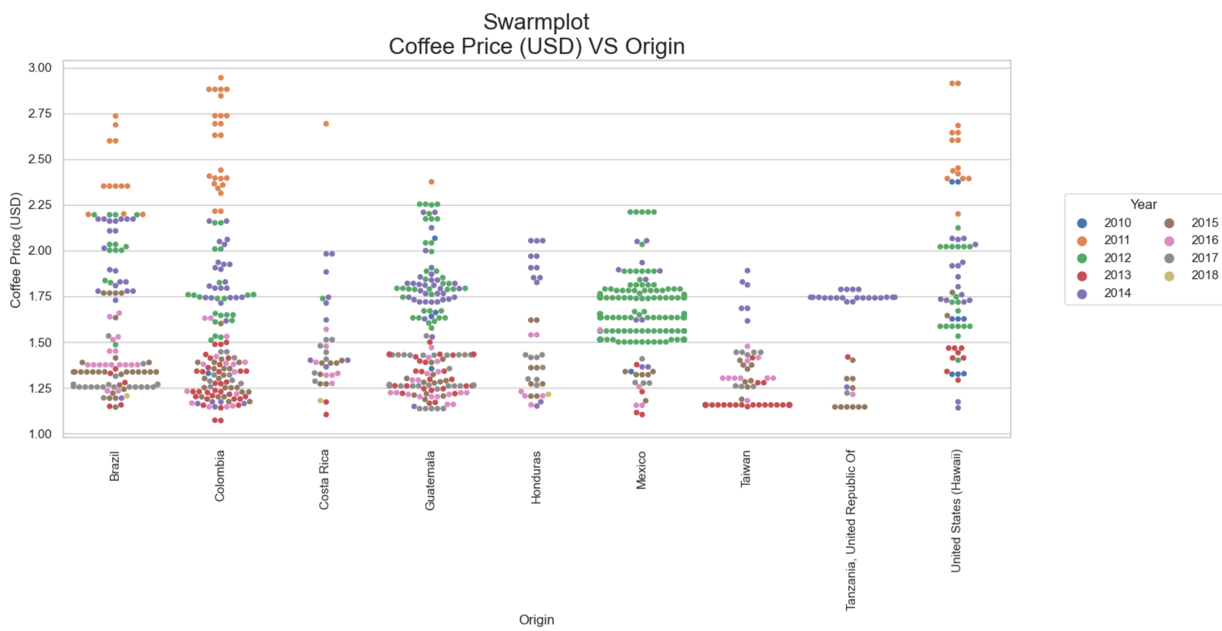

Boxplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

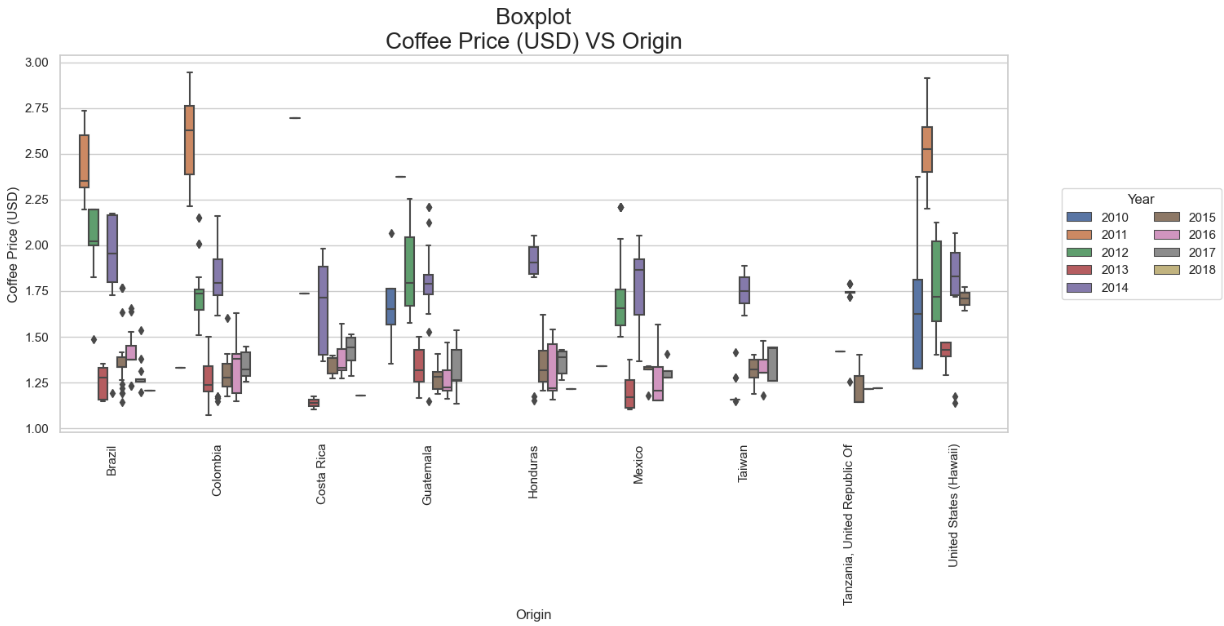

Violinplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.0
2.5
2.0
1.5
1.0
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

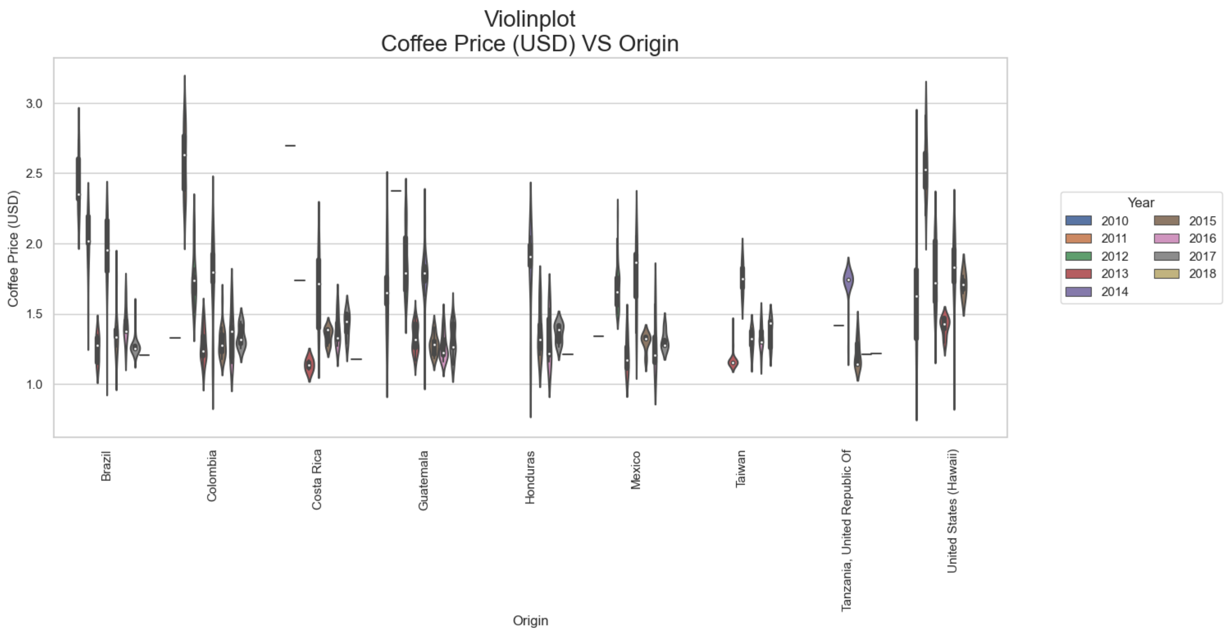


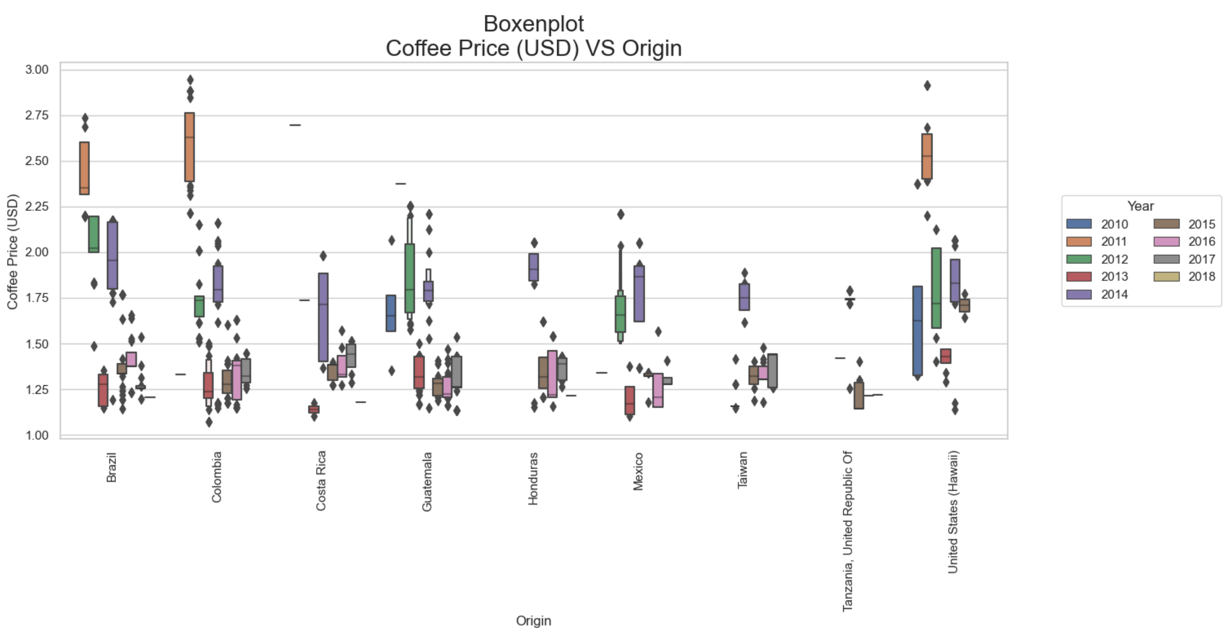
Coffee Price (USD) VS Origin (Year)



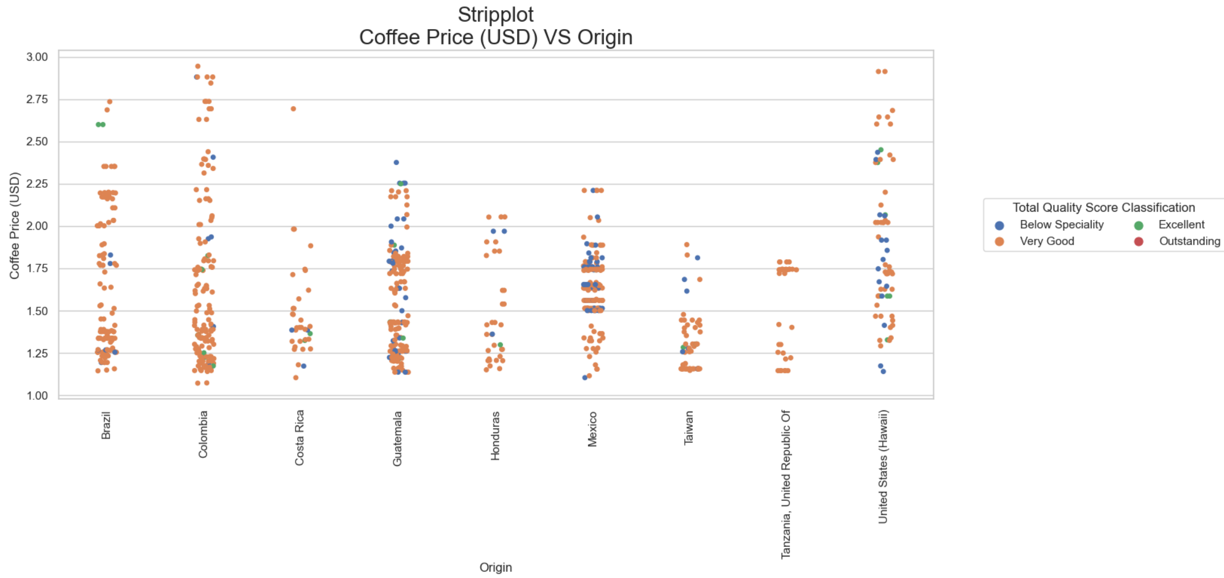


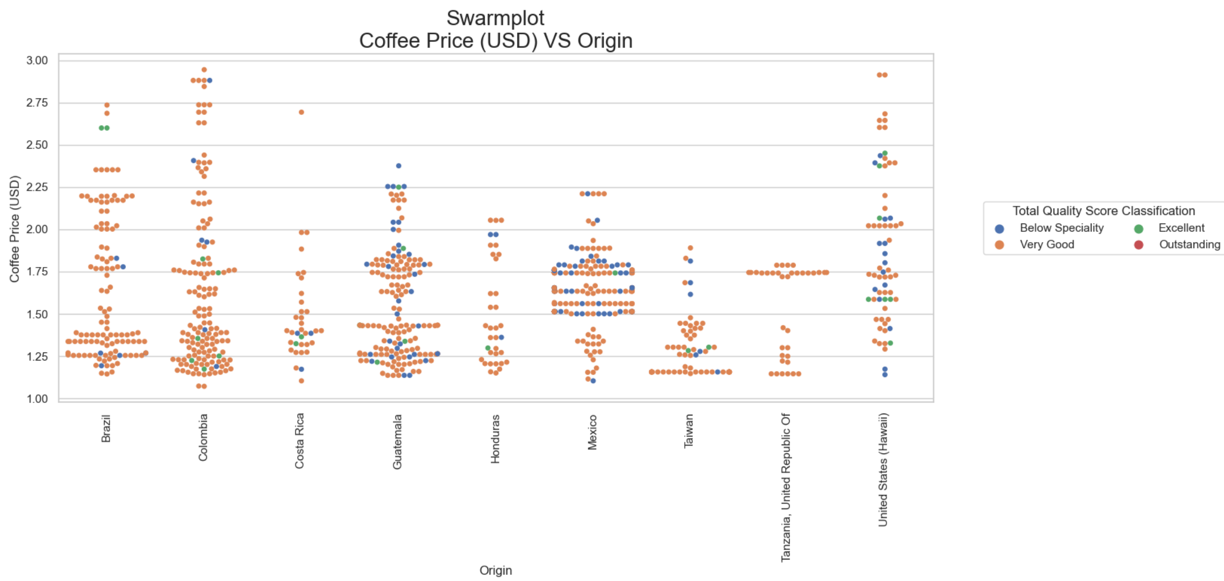

Boxplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

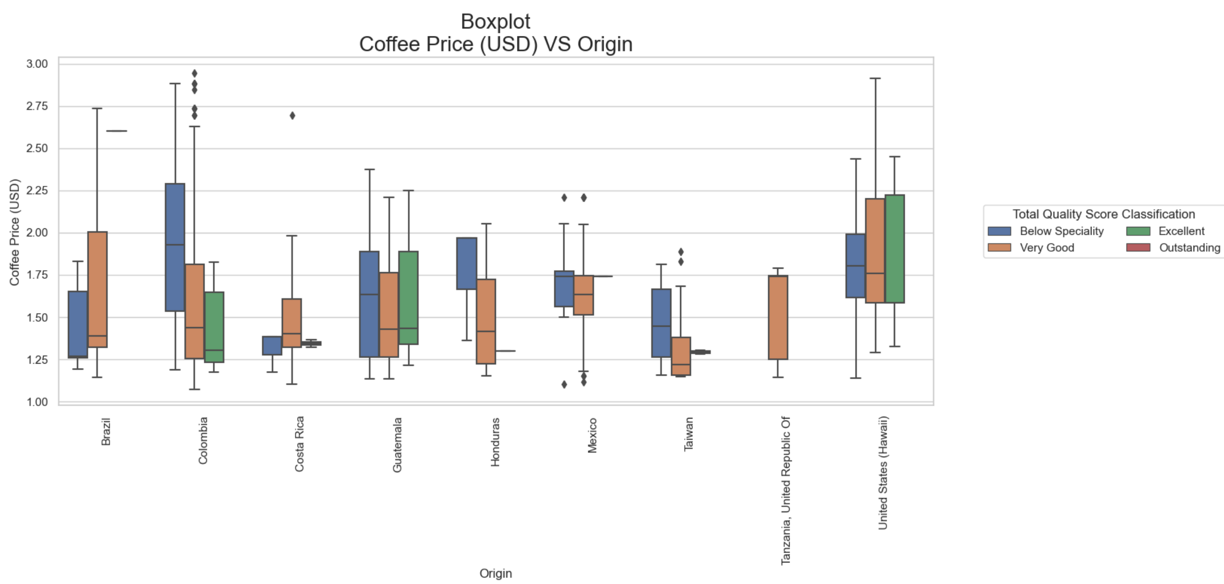

Violinplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.0
2.5
2.0
1.5
1.0
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

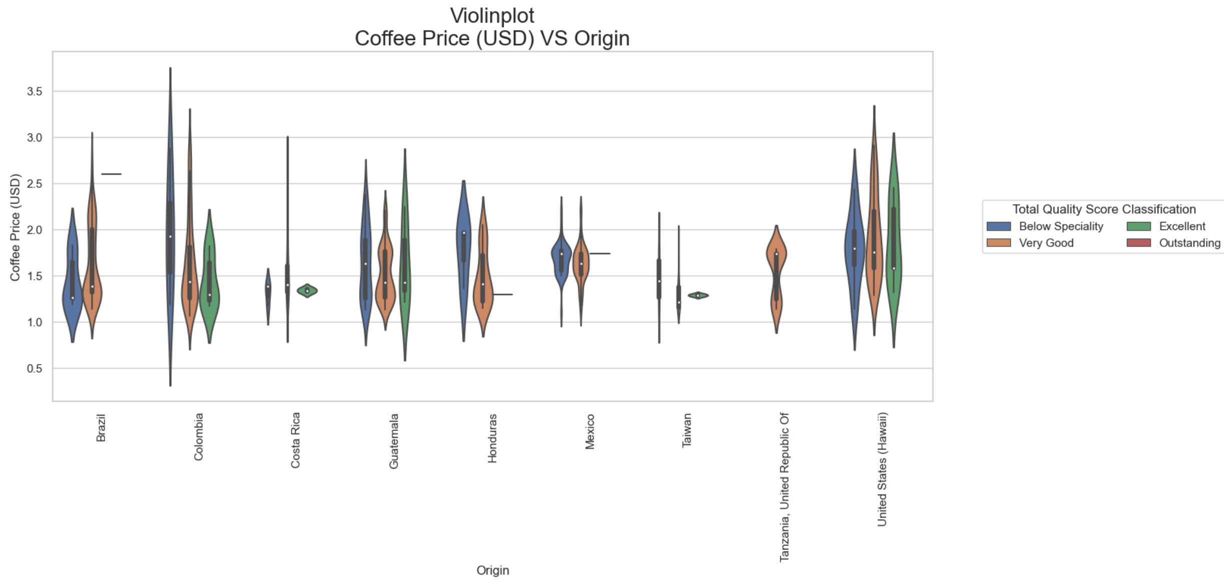

Boxenplot
Coffee Price (USD) VS Origin
Coffee Price (USD)
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)
Origin
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

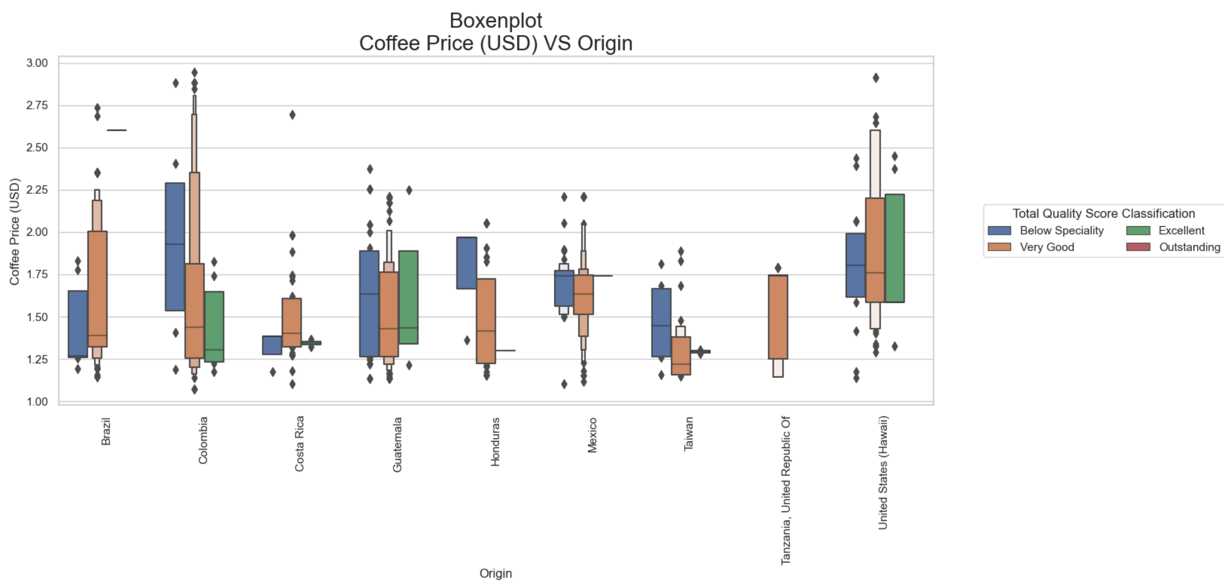
Coffee Price (USD) VS Origin (Total Quality Score Classification)



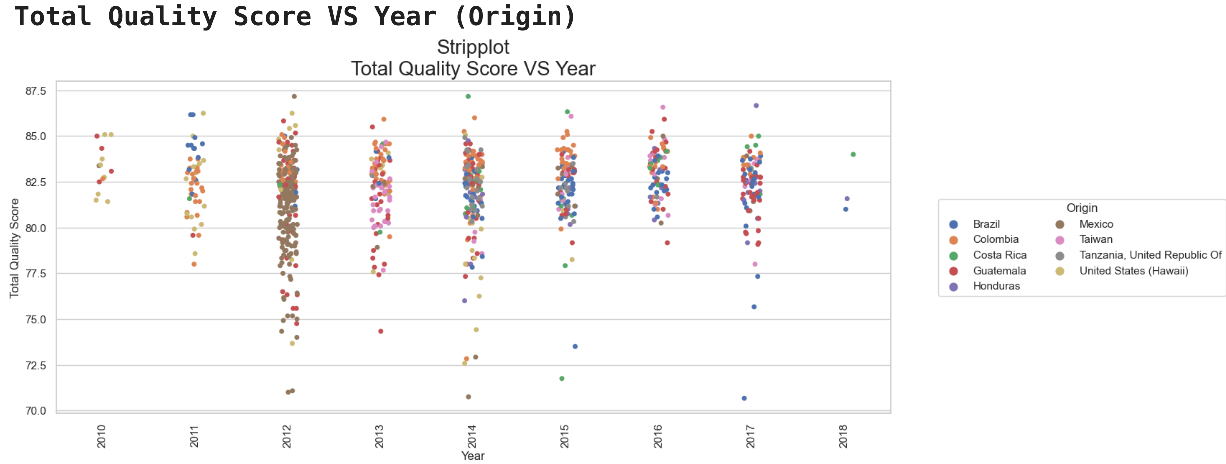


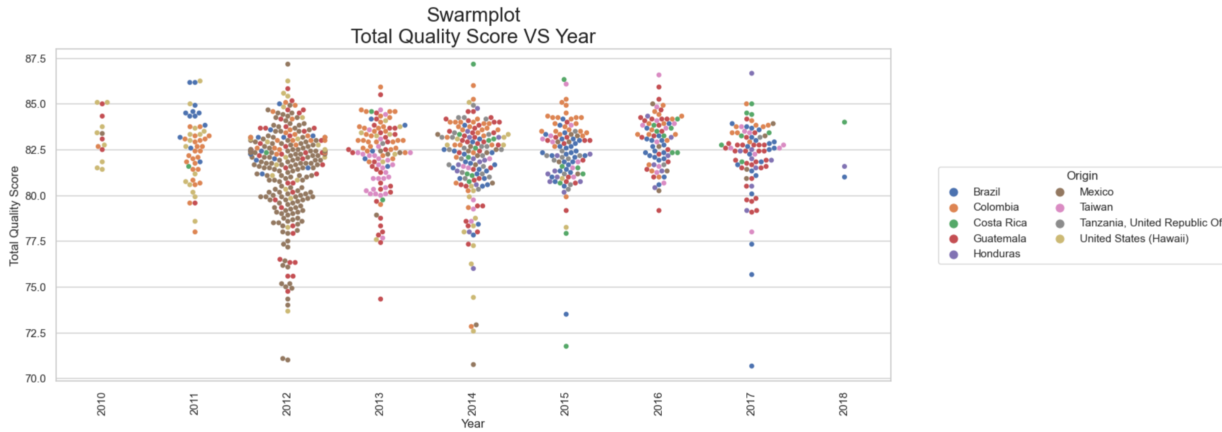


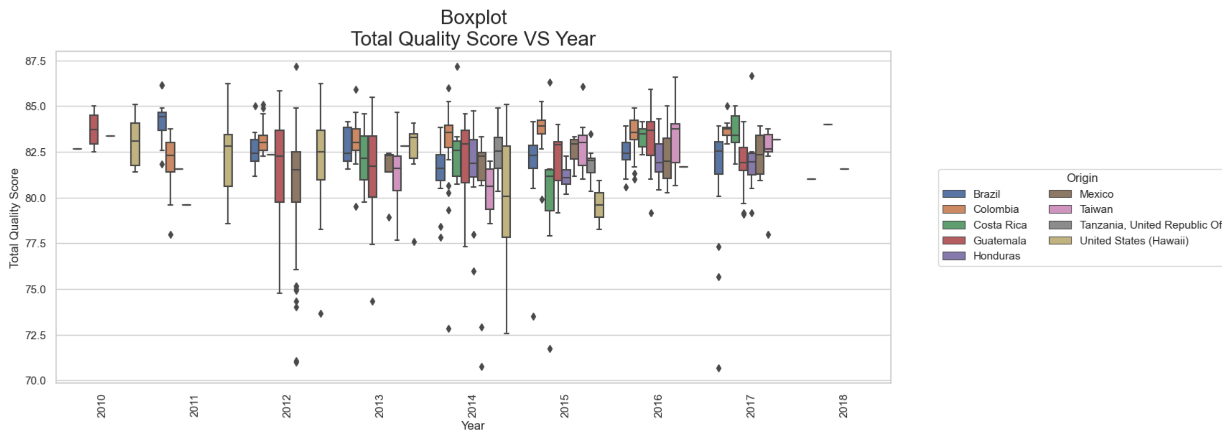


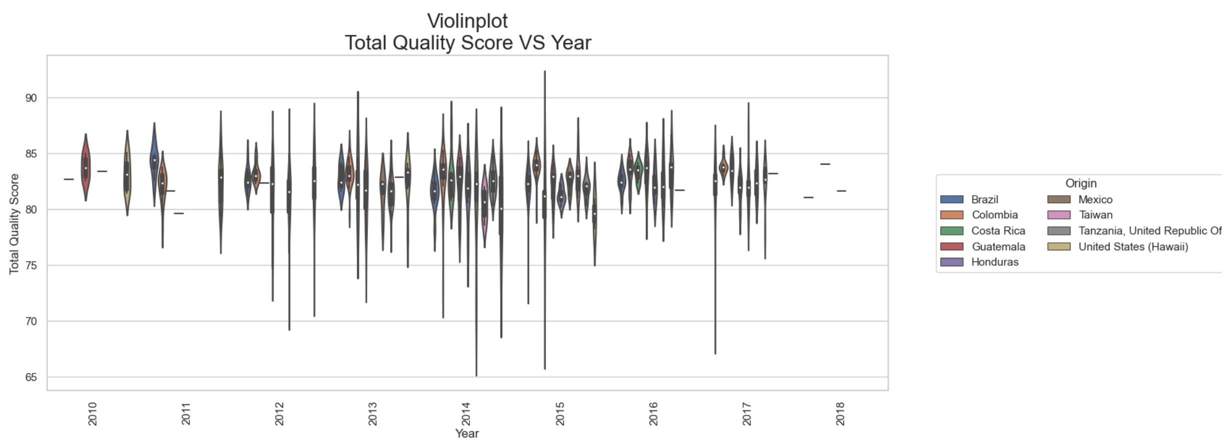


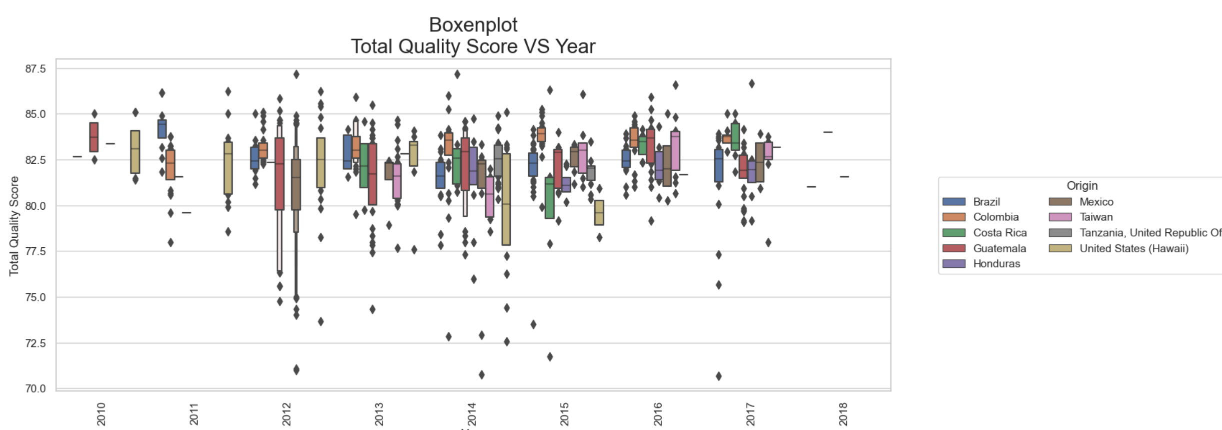
Total Quality Score VS Year (Origin)



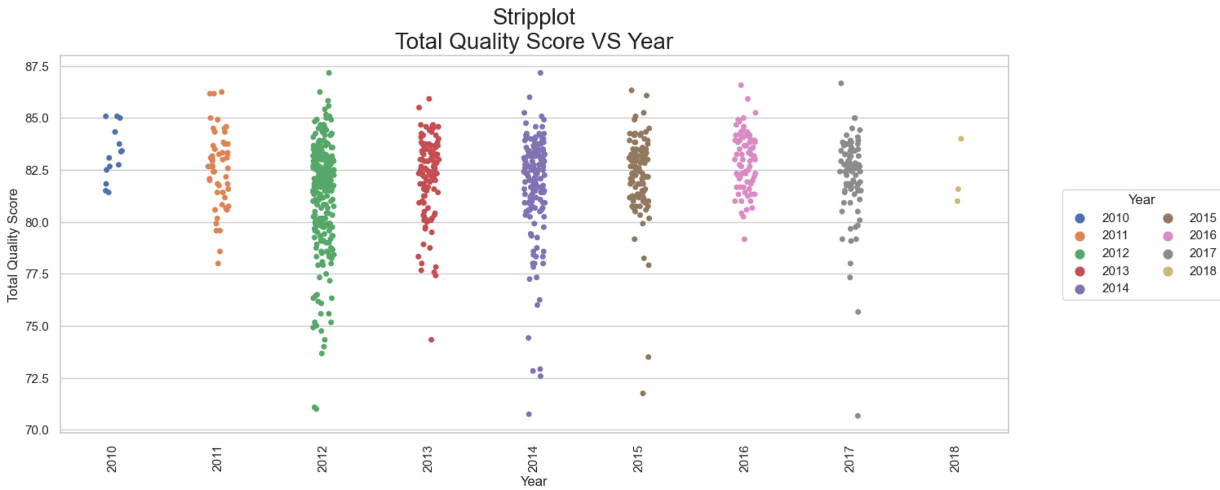


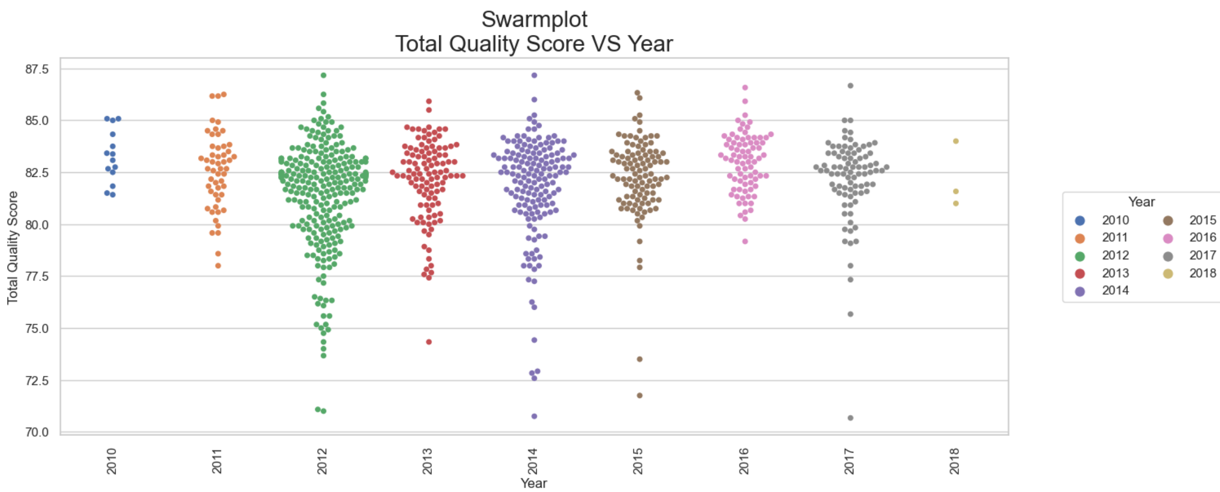


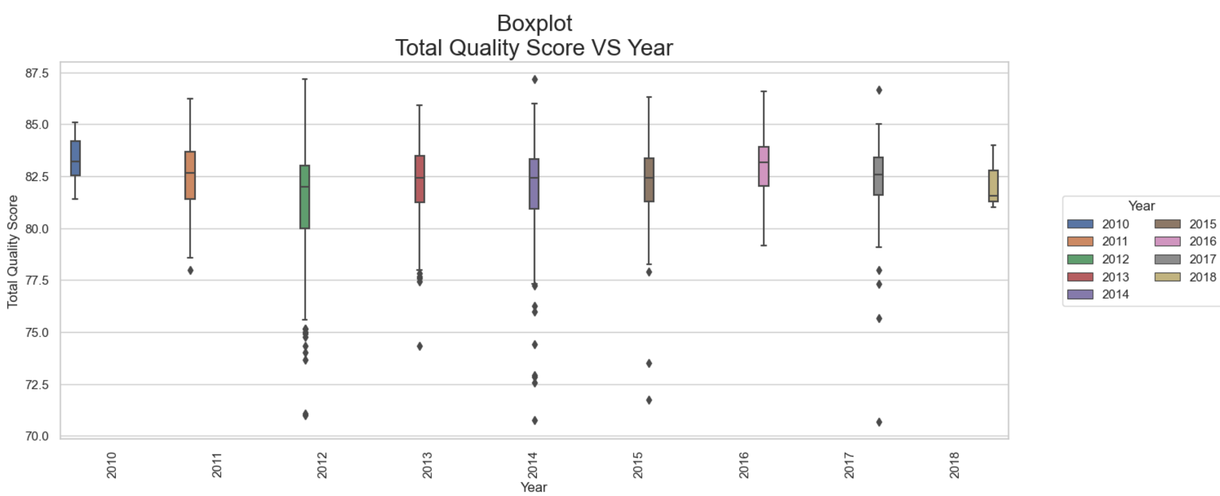


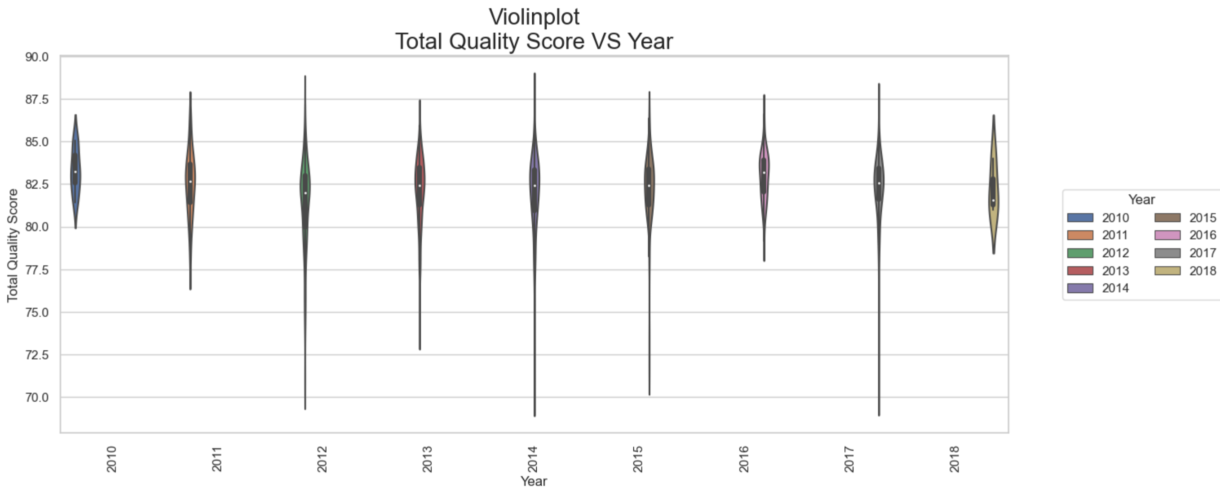


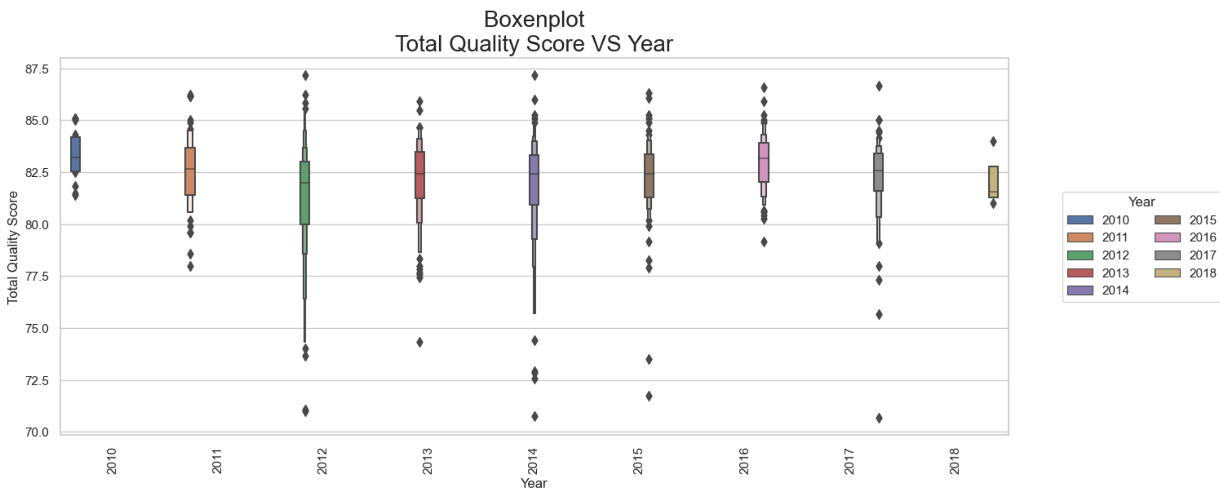
Total Quality Score VS Year (Year)


Stripplot
Total Quality Score VS Year
Total Quality Score
Year
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
2010
2011
2012
2013
2014
2015
2016
2017
2018

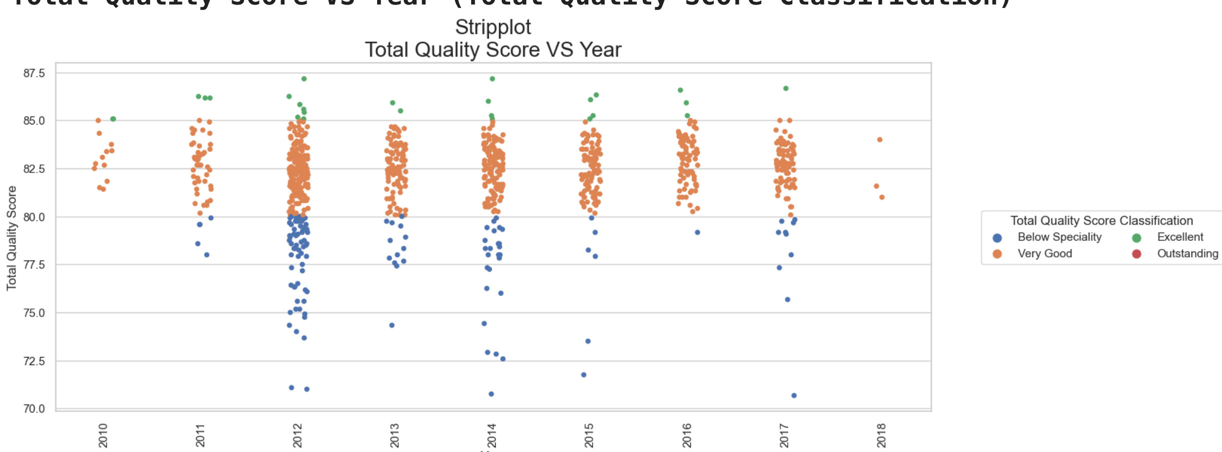

Swarmplot
Total Quality Score VS Year
Total Quality Score
Year
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
2010
2011
2012
2013
2014
2015
2016
2017
2018

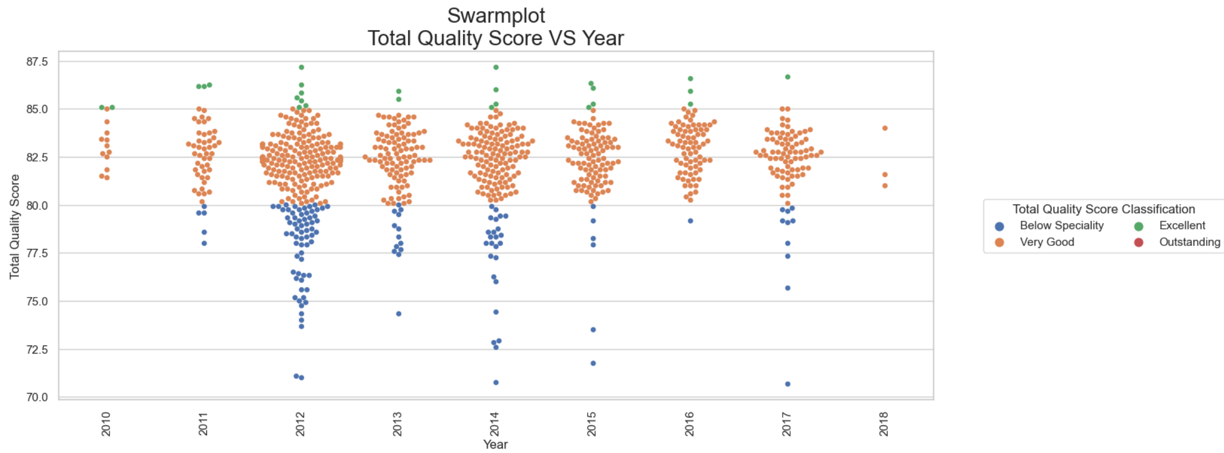

Boxplot
Total Quality Score VS Year
Total Quality Score
Year
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
2010
2011
2012
2013
2014
2015
2016
2017
2018

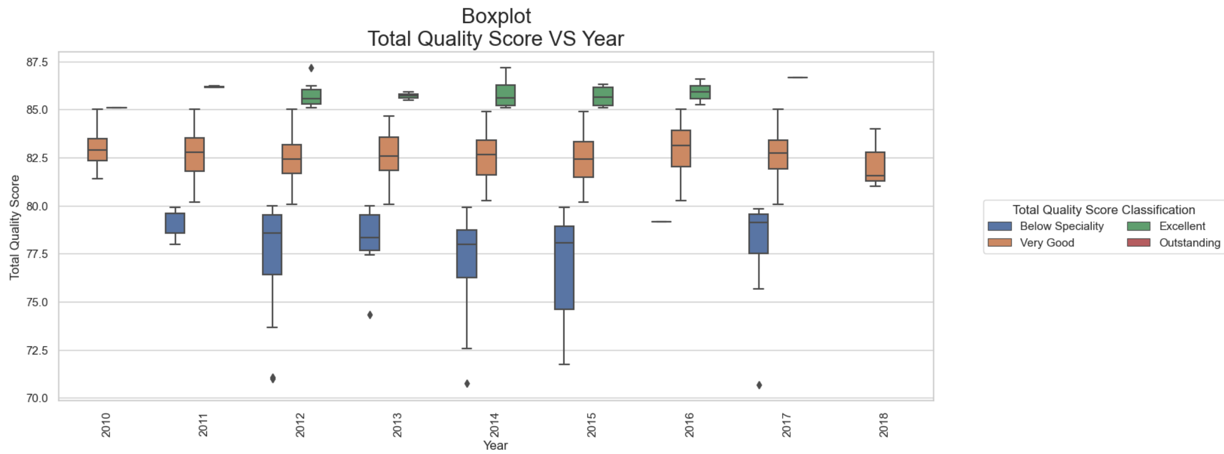


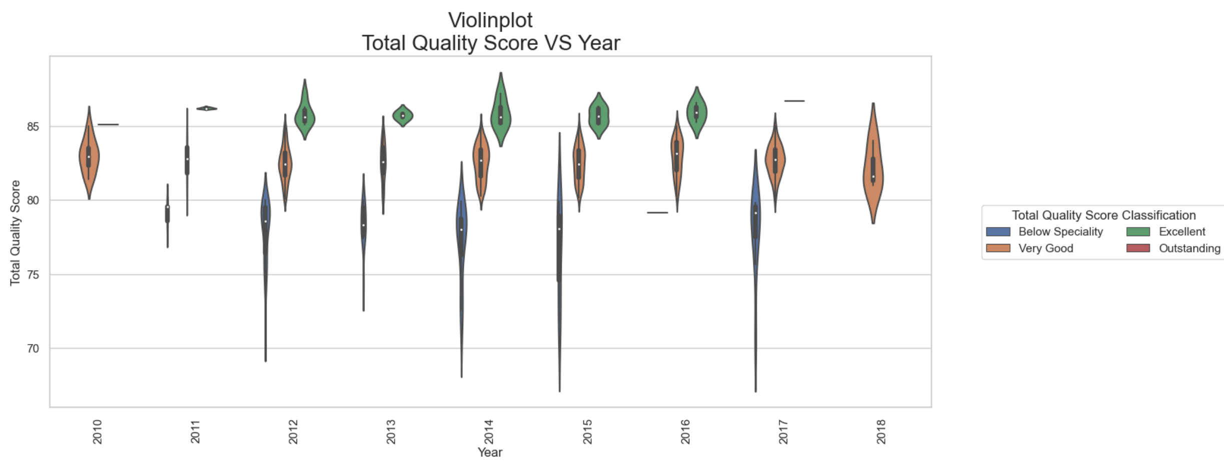


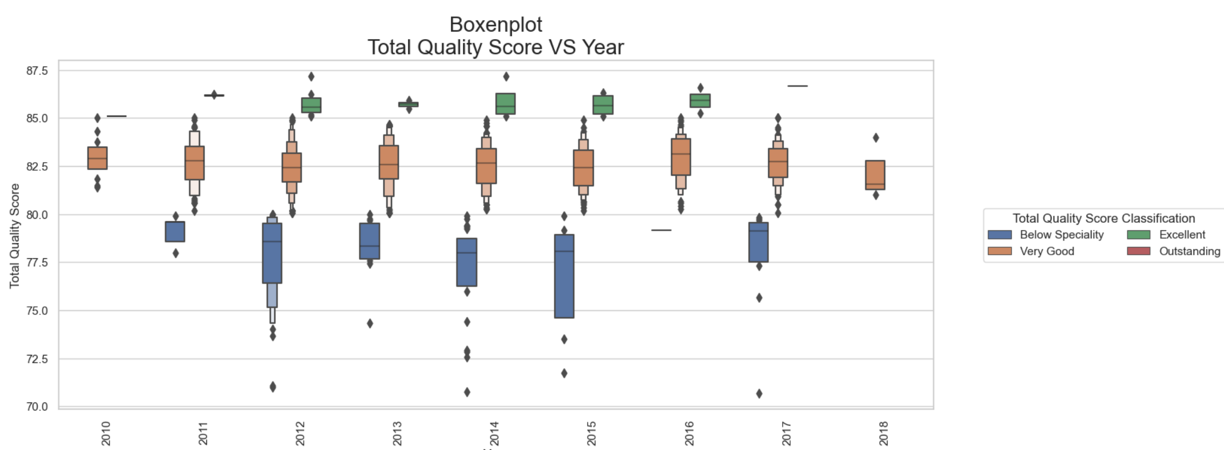
Total Quality Score VS Year (Total Quality Score Classification)



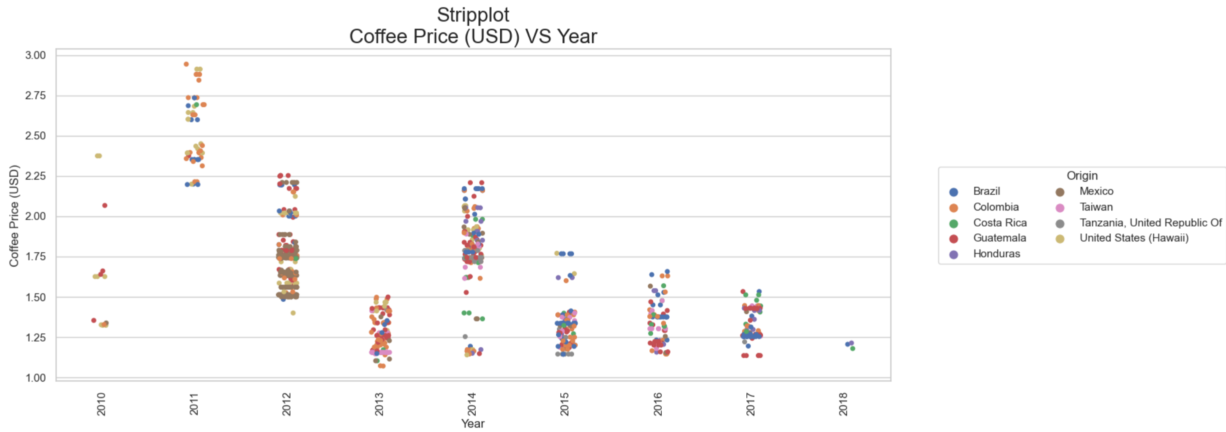


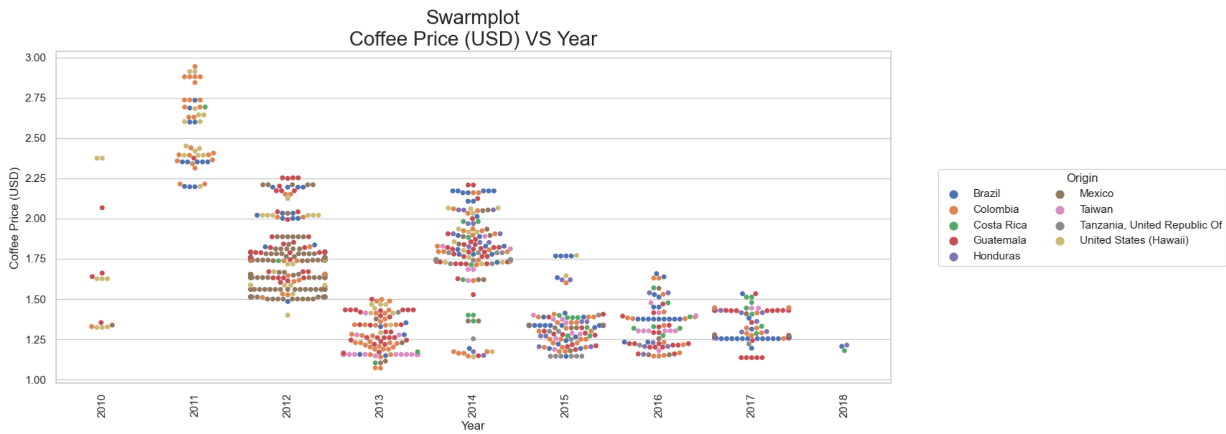


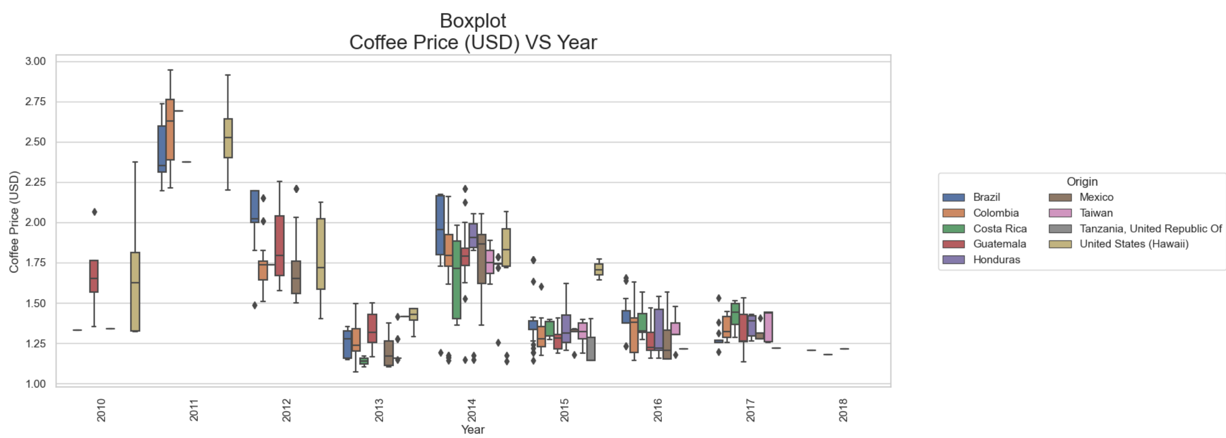


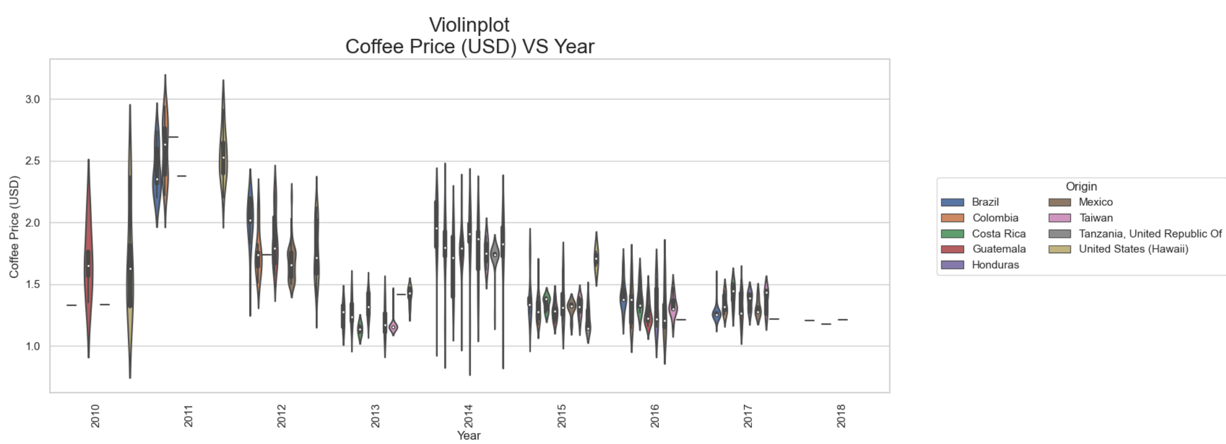


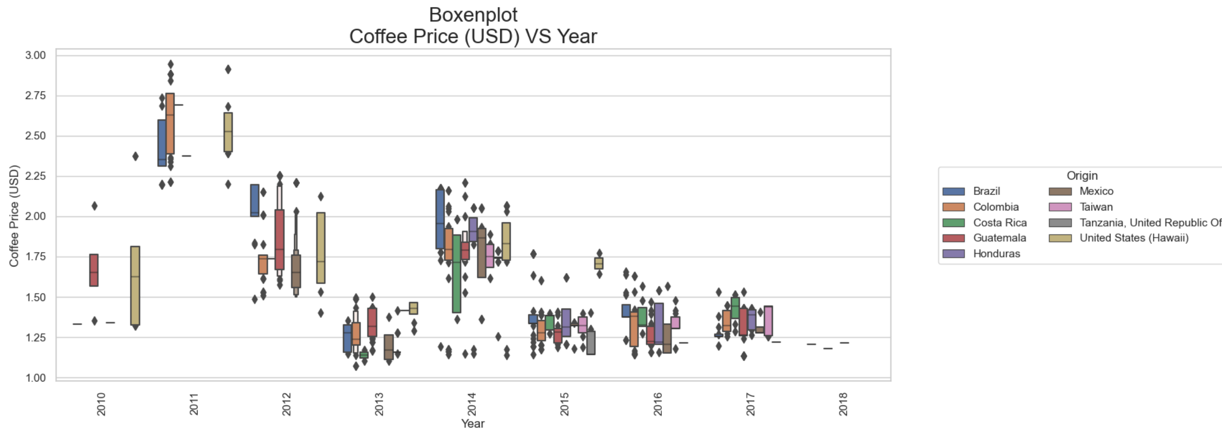
Coffee Price (USD) VS Year (Origin)


Stripplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

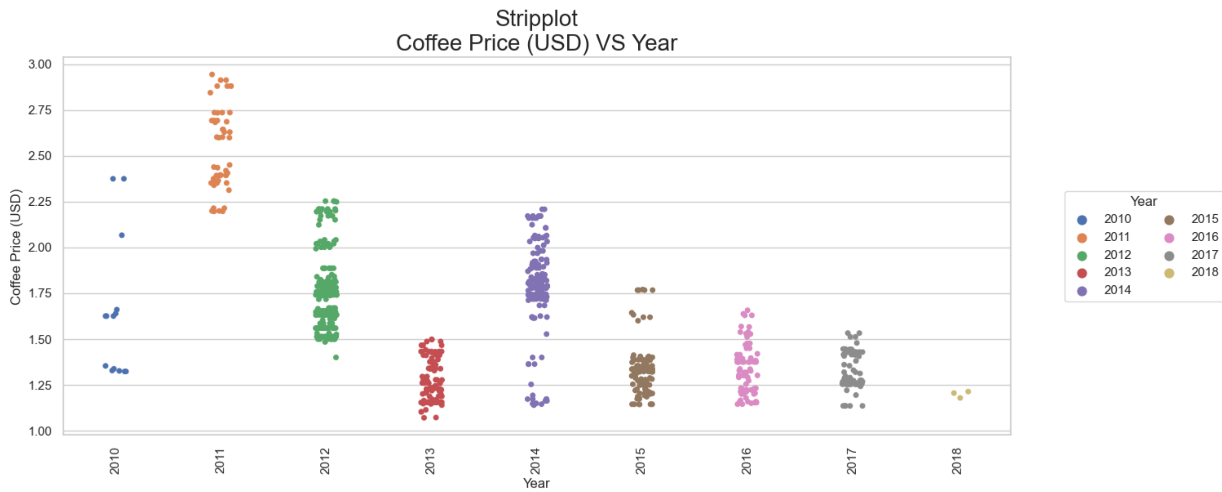

Swarmplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

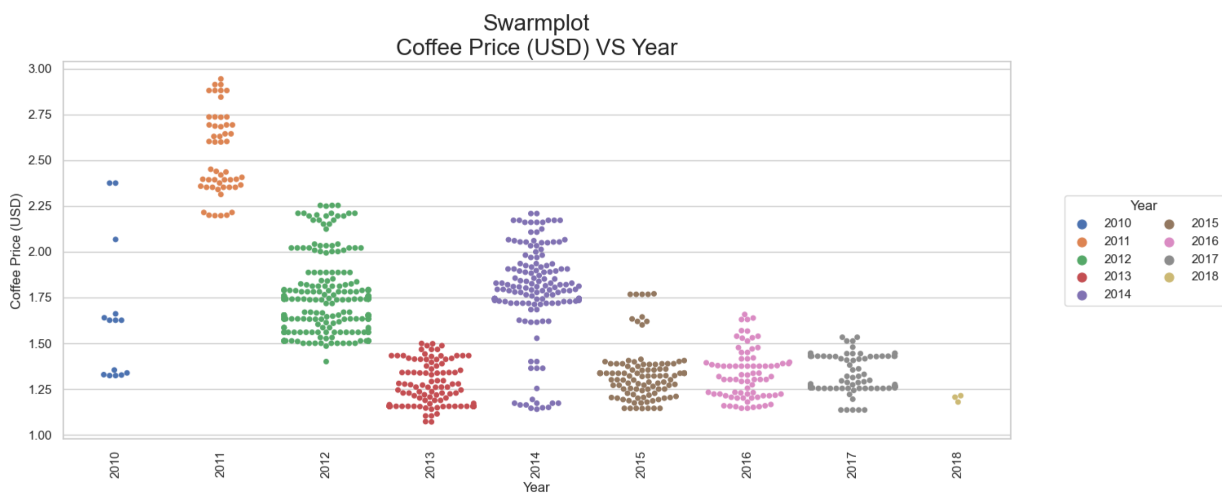

Boxplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

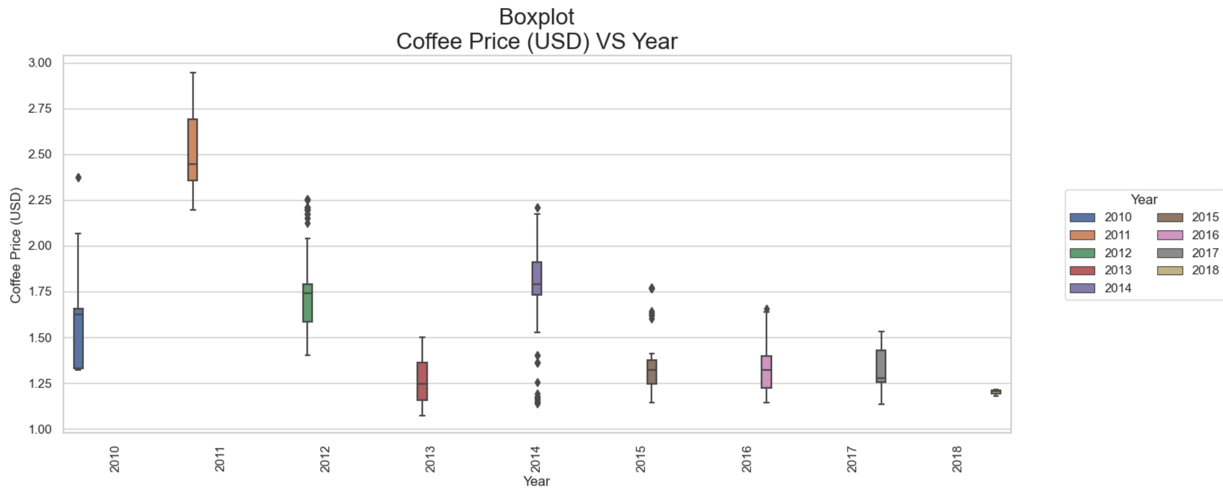

Violinplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Origin
Brazil
Colombia
Costa Rica
Guatemala
Honduras
Mexico
Taiwan
Tanzania, United Republic Of
United States (Hawaii)

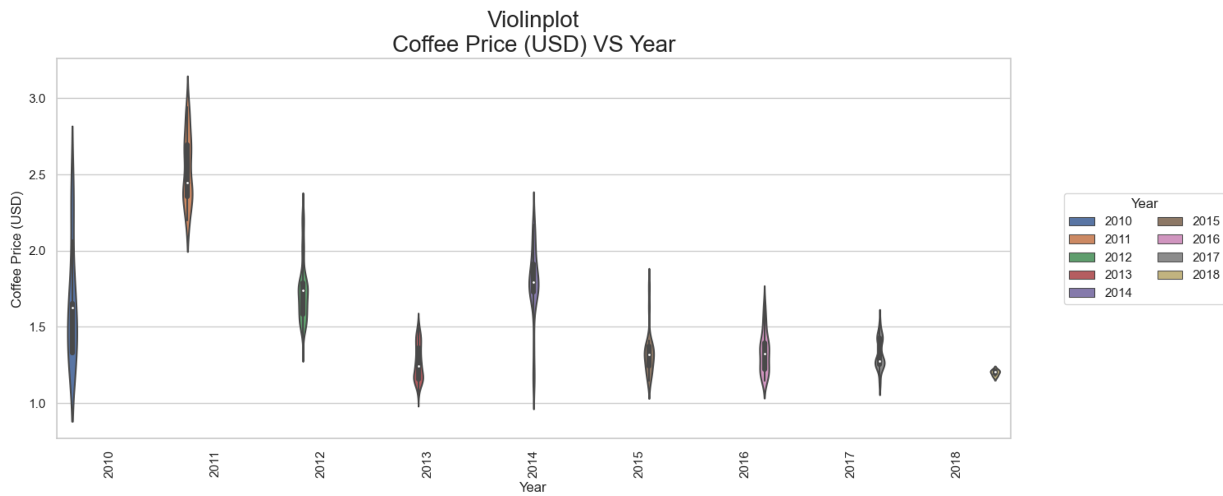


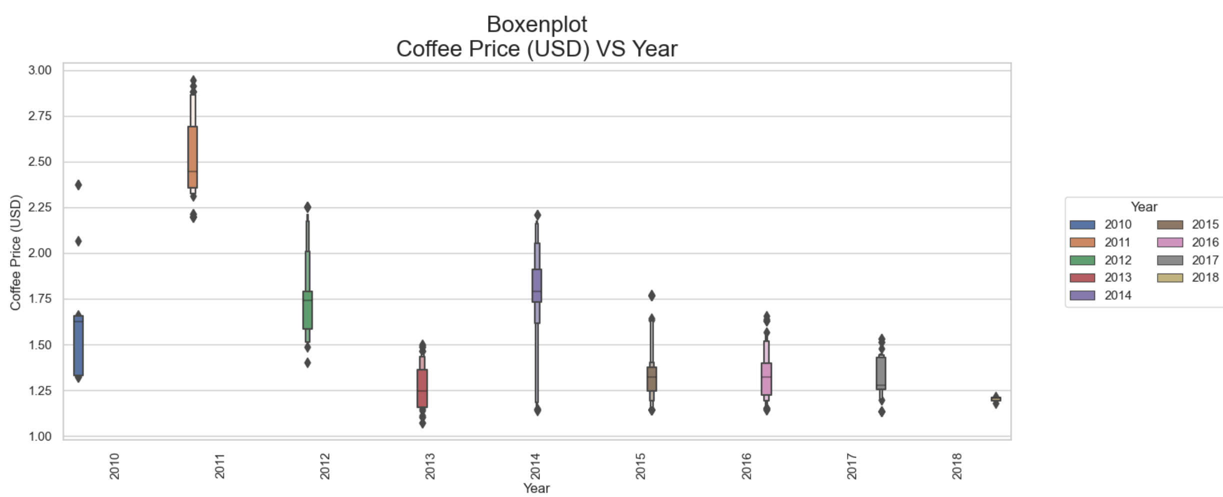
Coffee Price (USD) VS Year (Year)



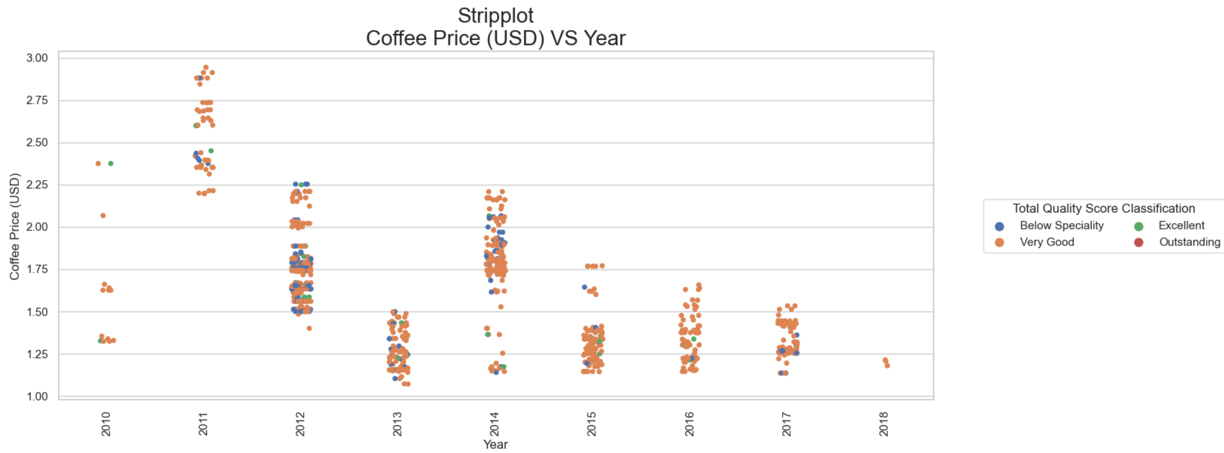


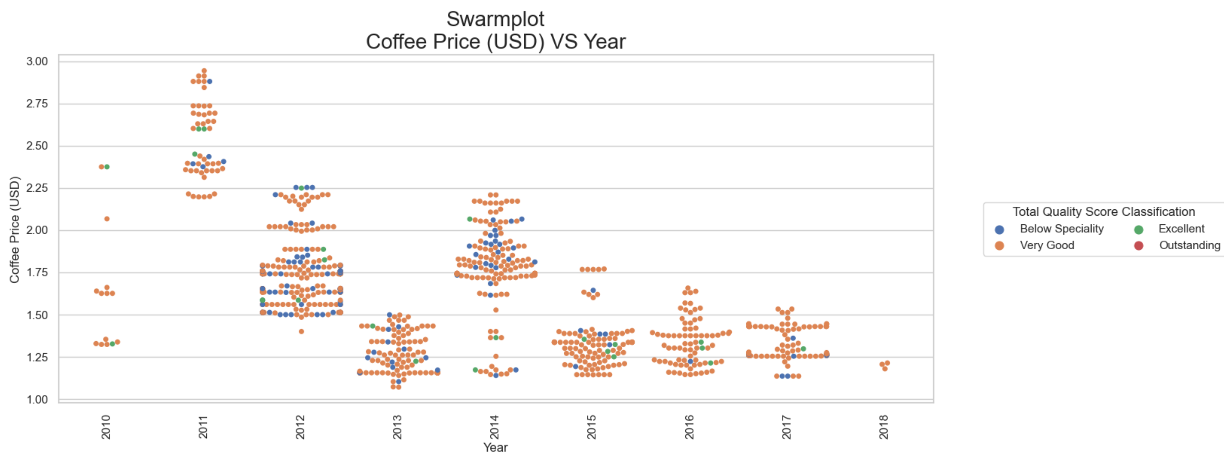


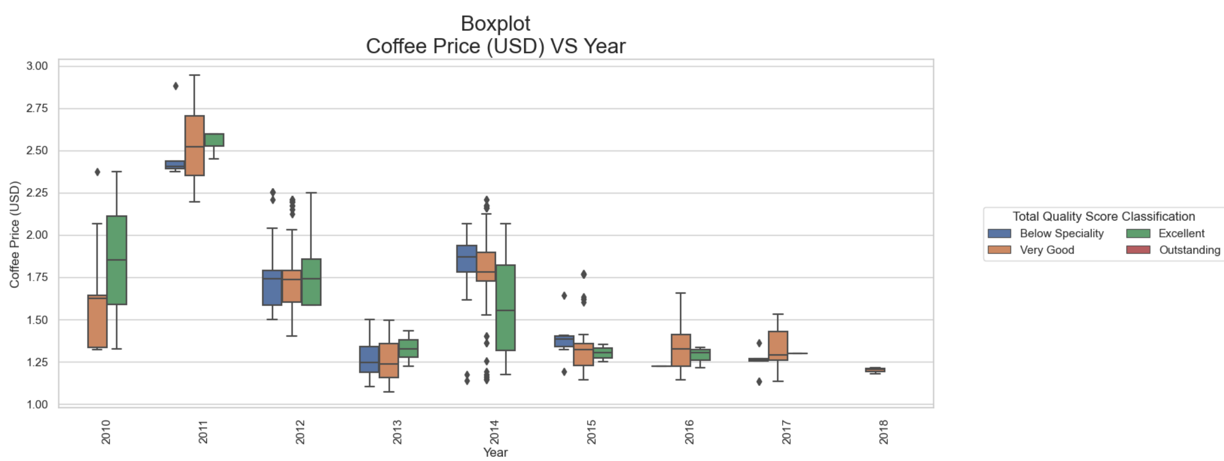


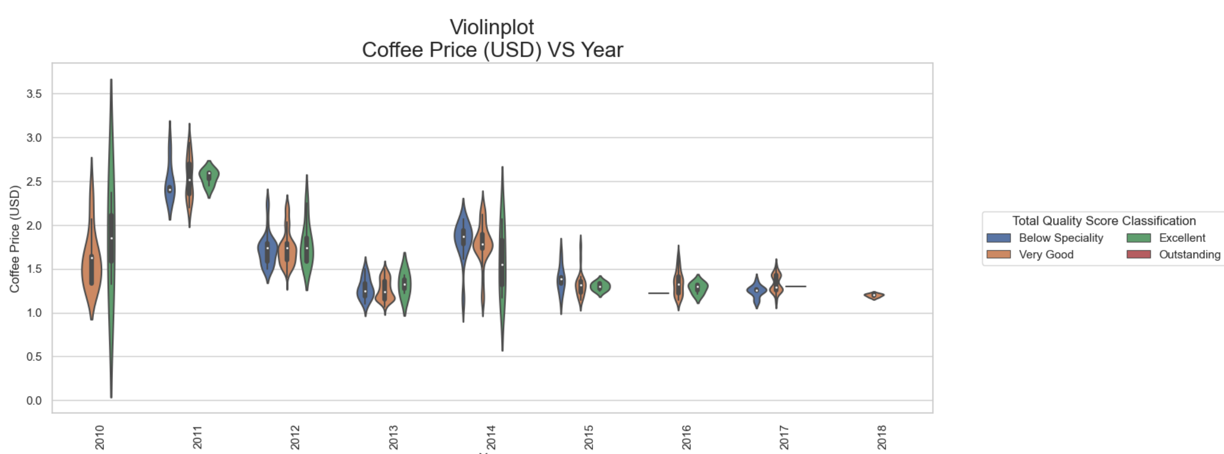


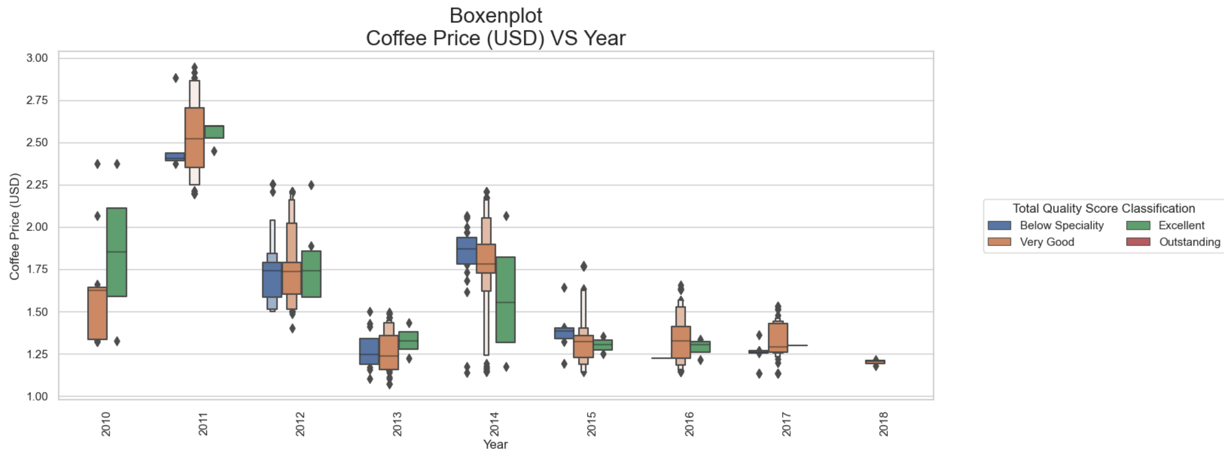
Coffee Price (USD) VS Year (Total Quality Score Classification)


Stripplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

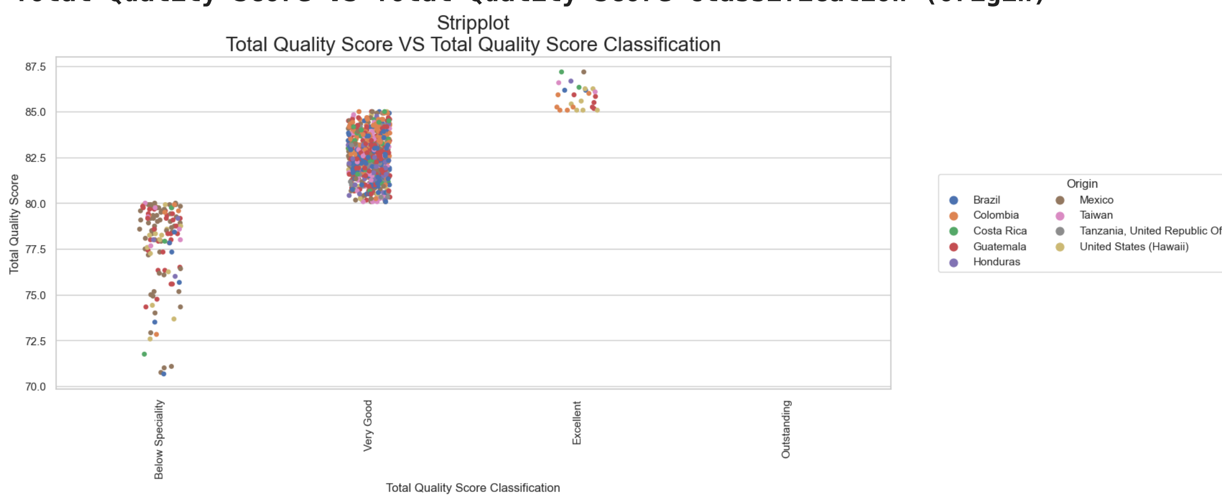

Swarmplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

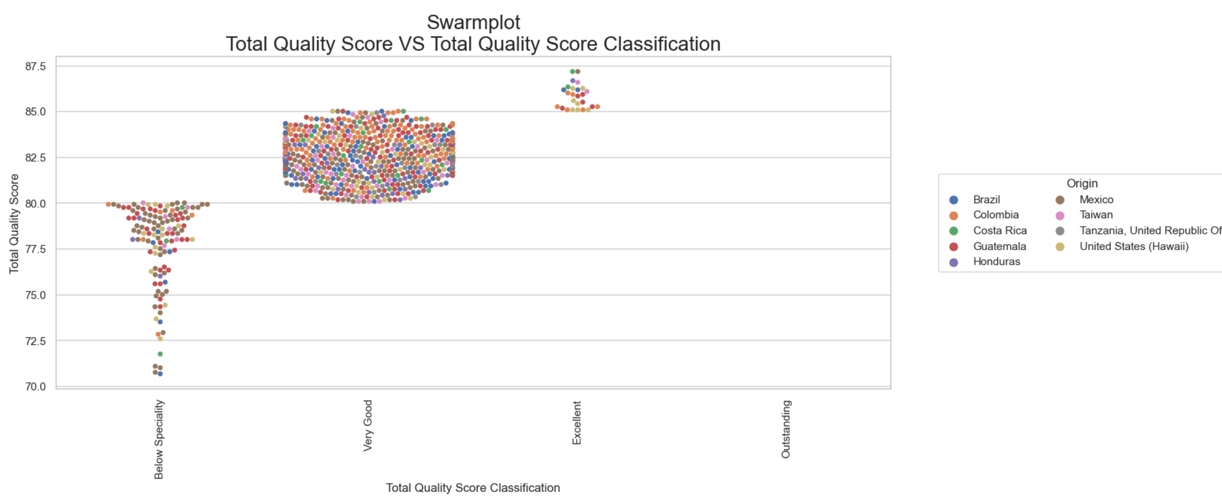

Boxplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

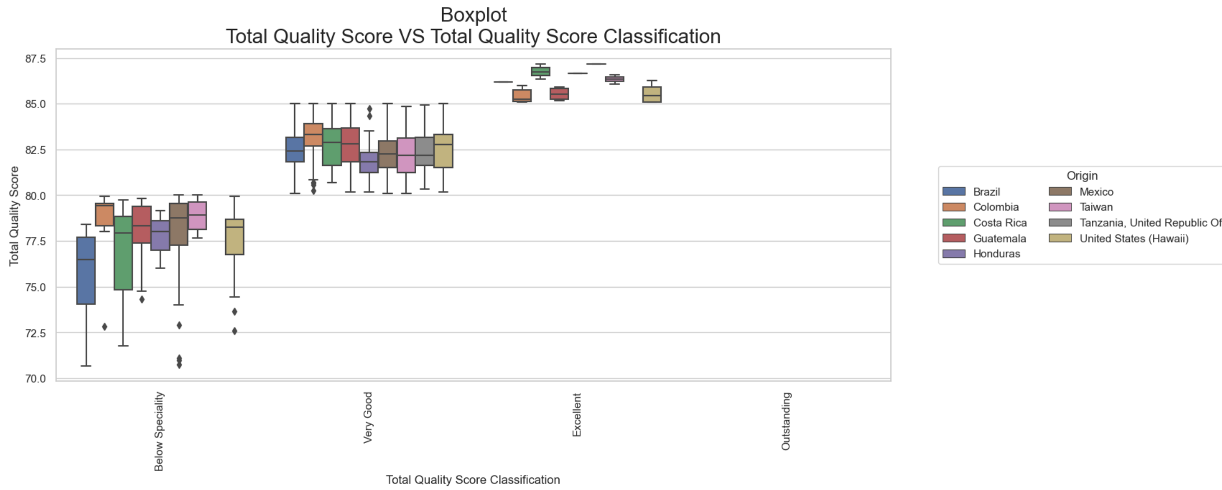

Violinplot
Coffee Price (USD) VS Year
Coffee Price (USD)
Year
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

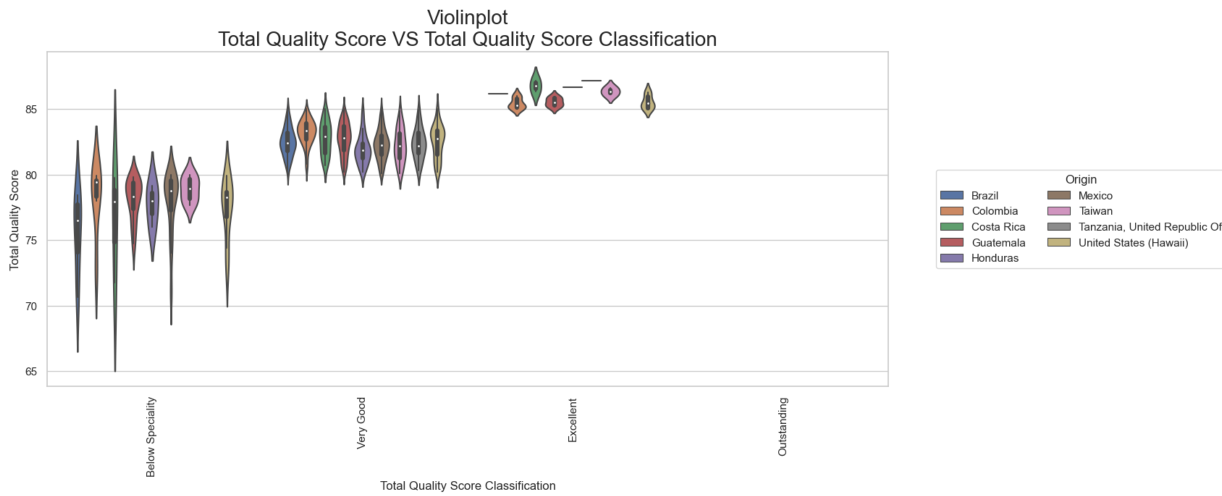


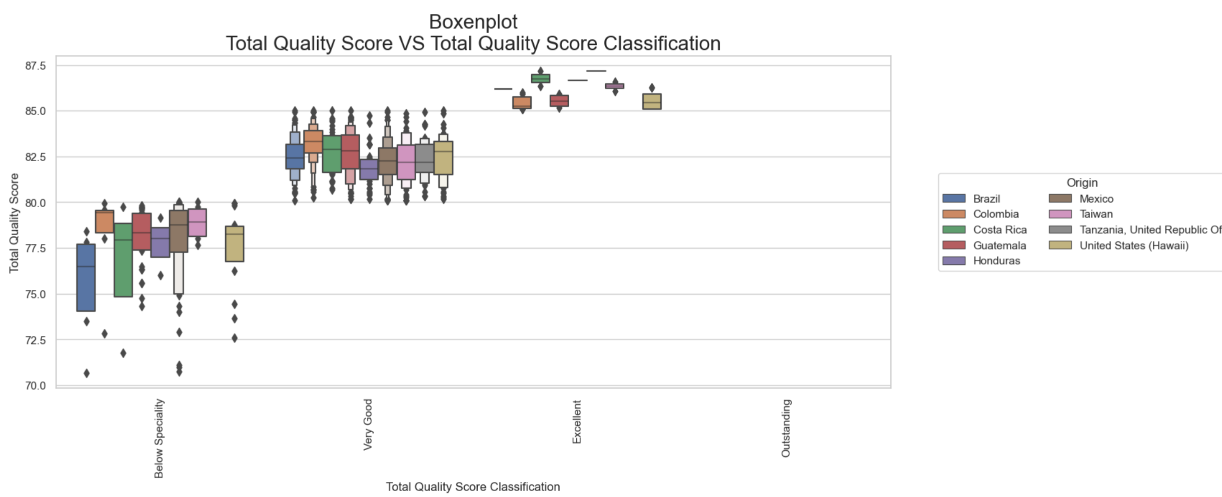
Total Quality Score VS Total Quality Score Classification (Origin)



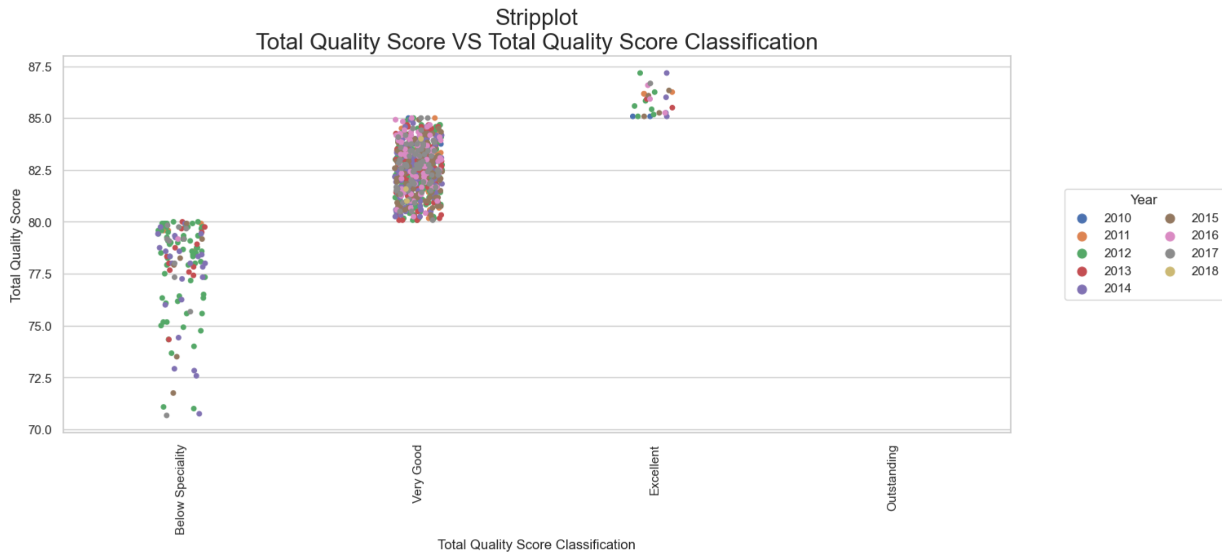


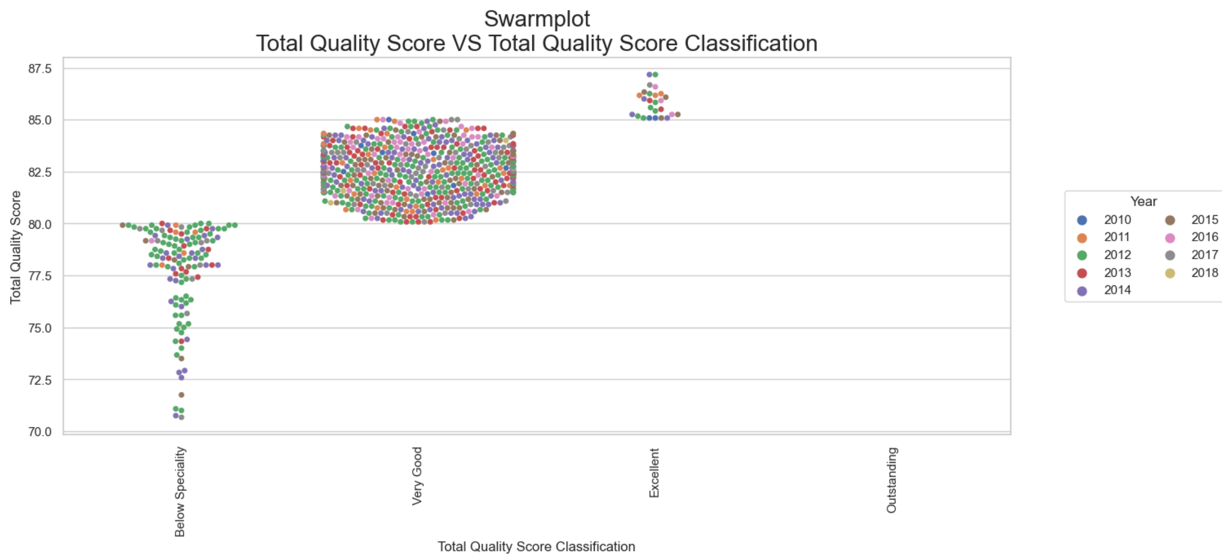


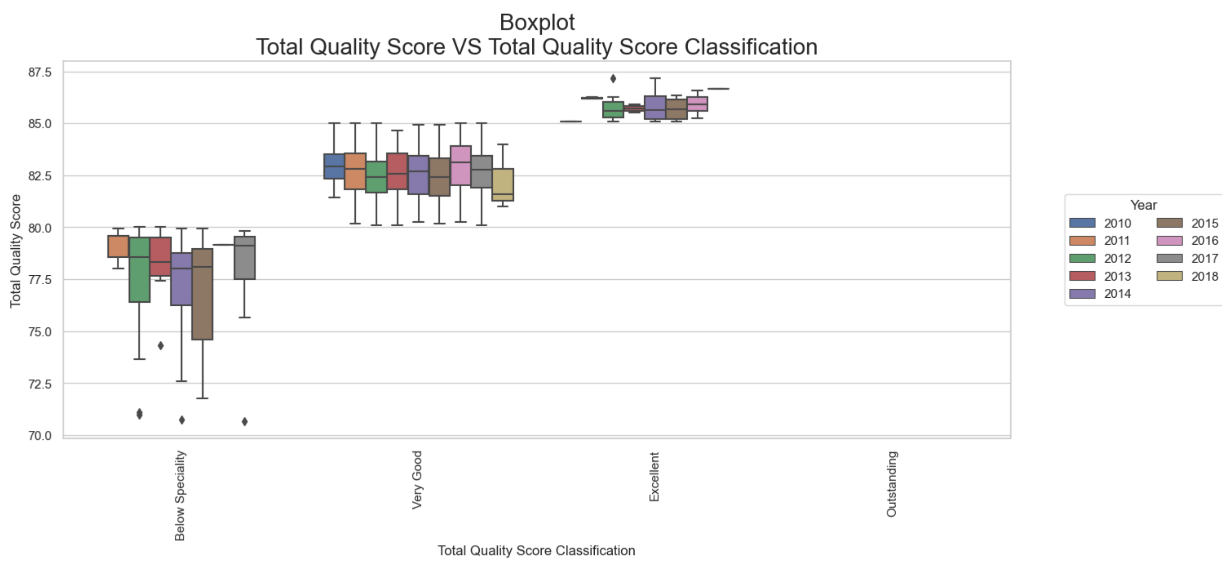


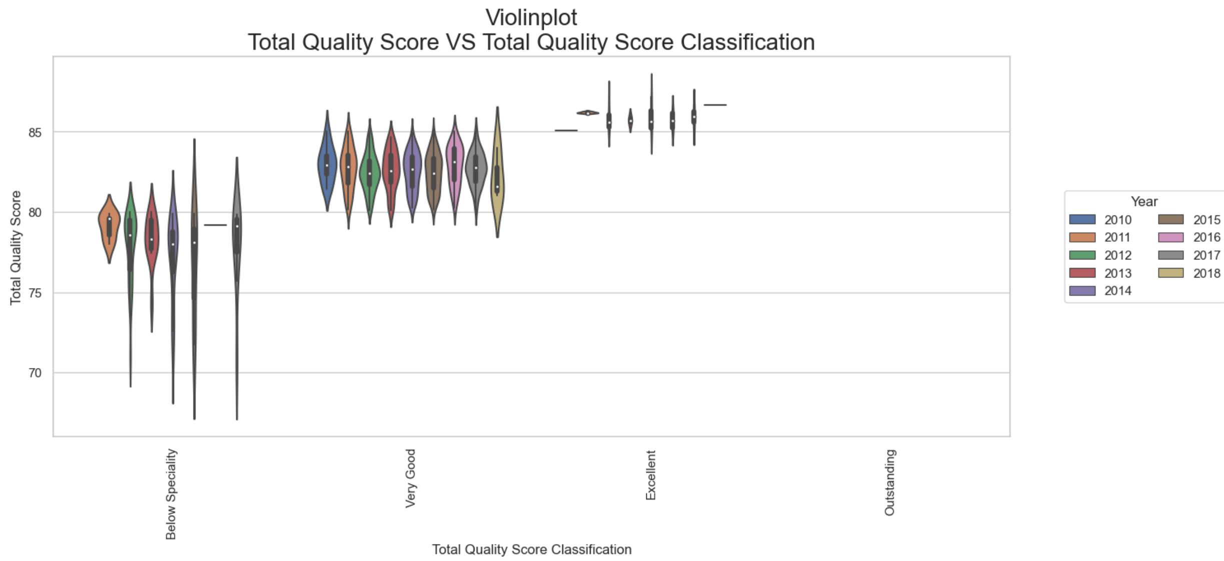


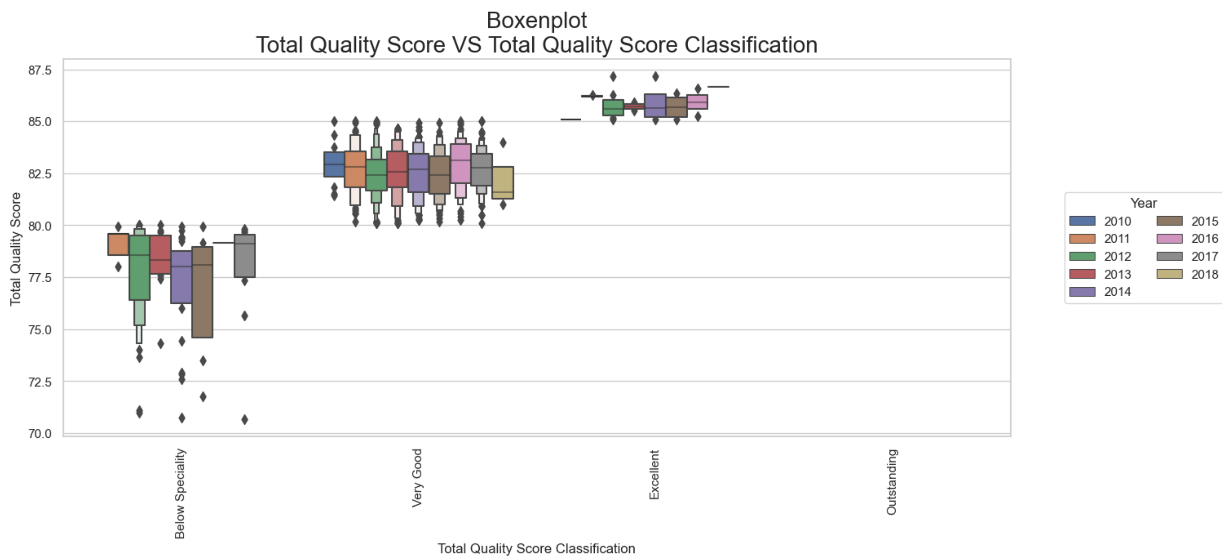
Total Quality Score VS Total Quality Score Classification (Year)


Stripplot
Total Quality Score VS Total Quality Score Classification
Total Quality Score
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

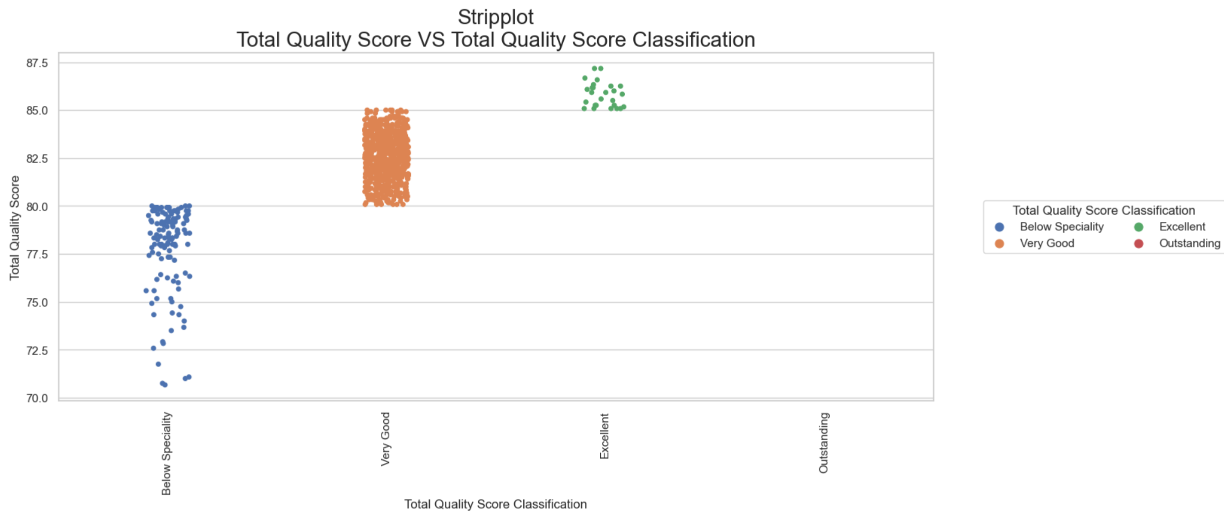

Swarmplot
Total Quality Score VS Total Quality Score Classification
Total Quality Score
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

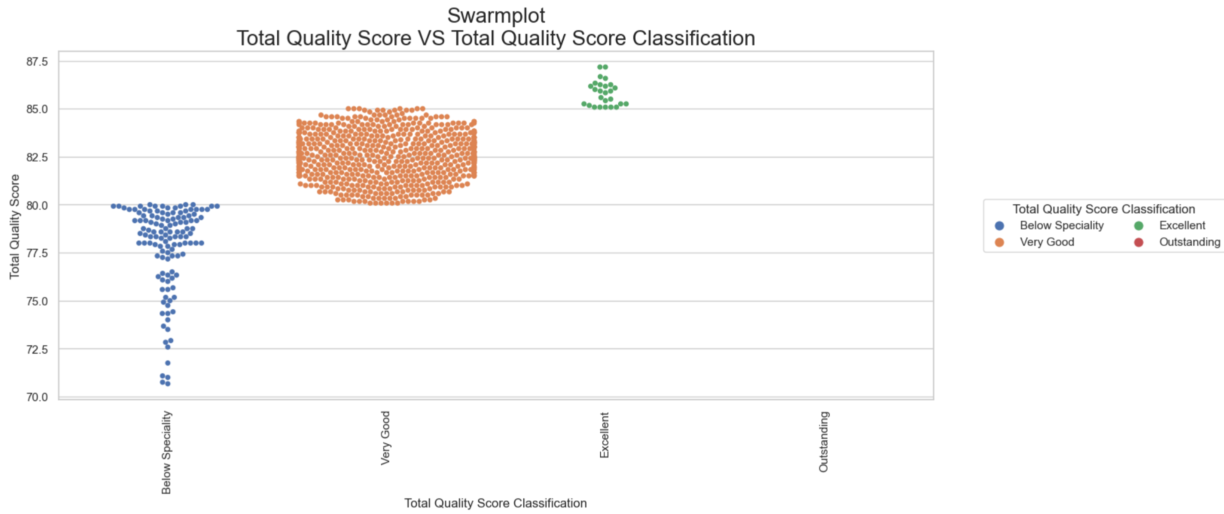

Boxplot
Total Quality Score VS Total Quality Score Classification
Total Quality Score
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Year
2010
2011
2012
2013
2014
2015
2016
2017
2018

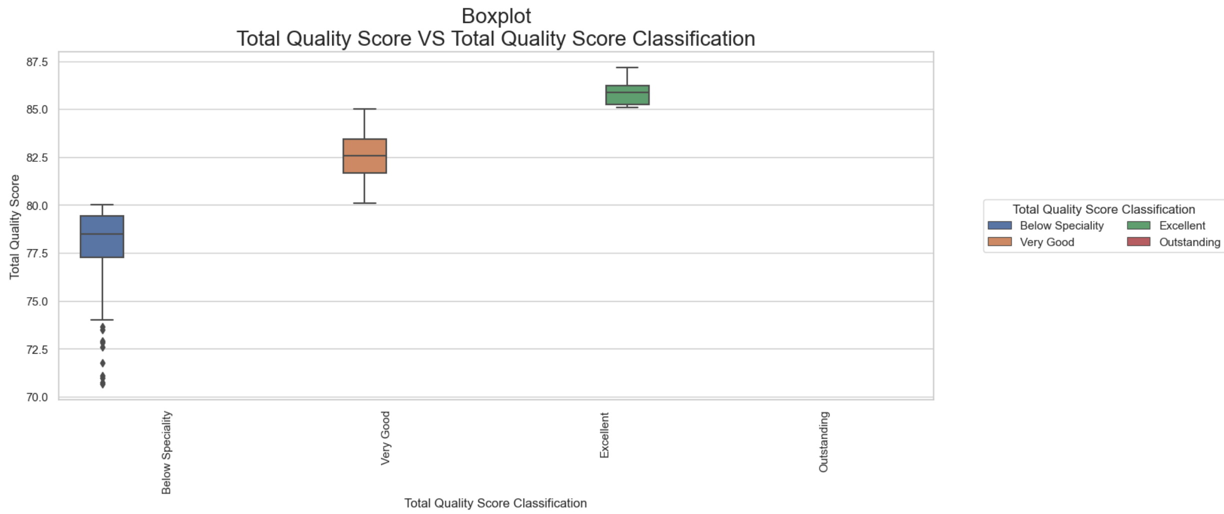


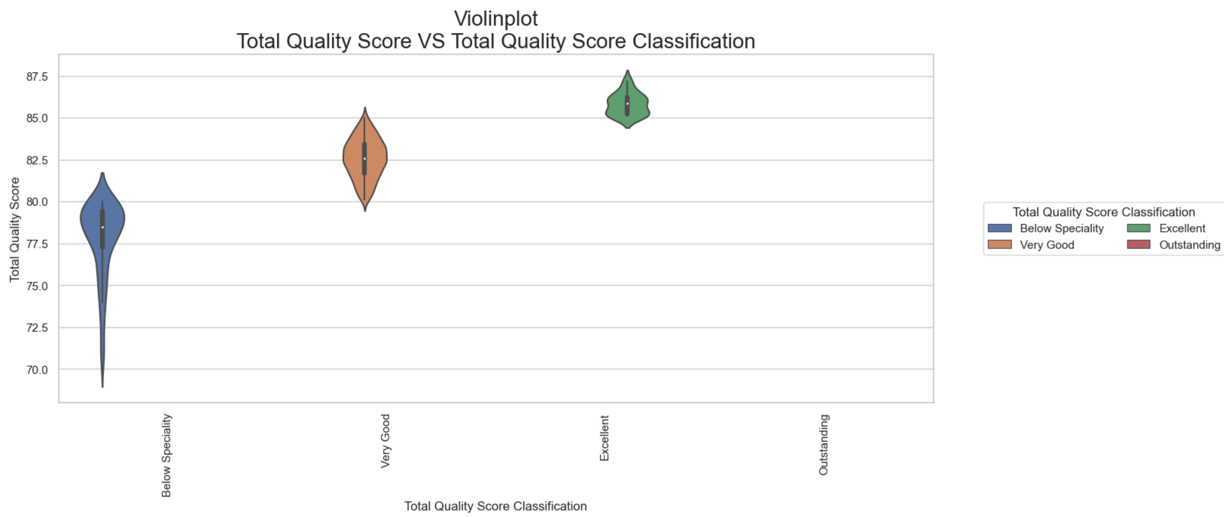


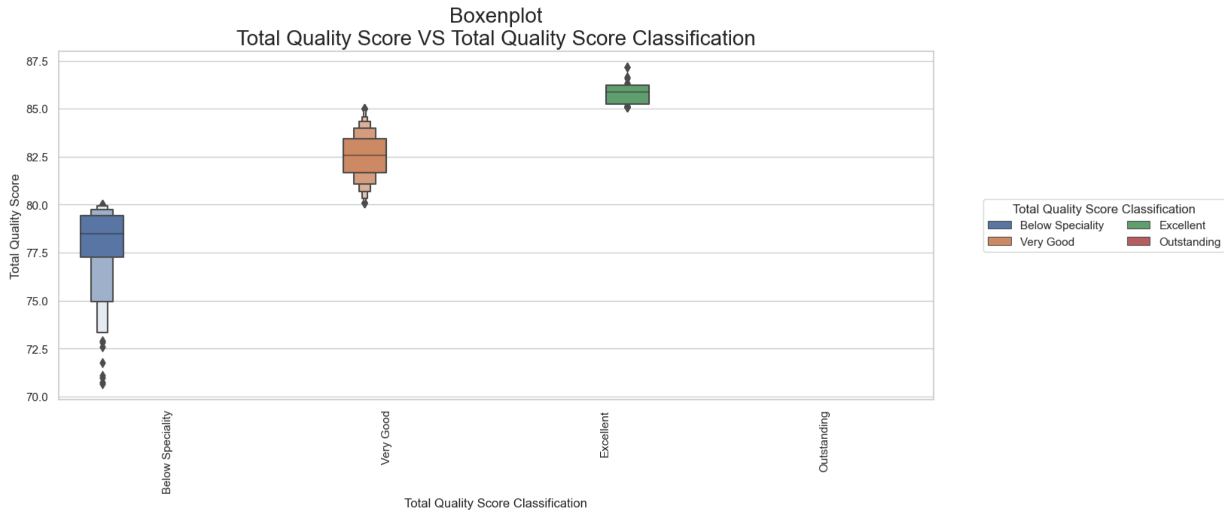
Total Quality Score VS Total Quality Score Classification (Total Quality Score Classification)



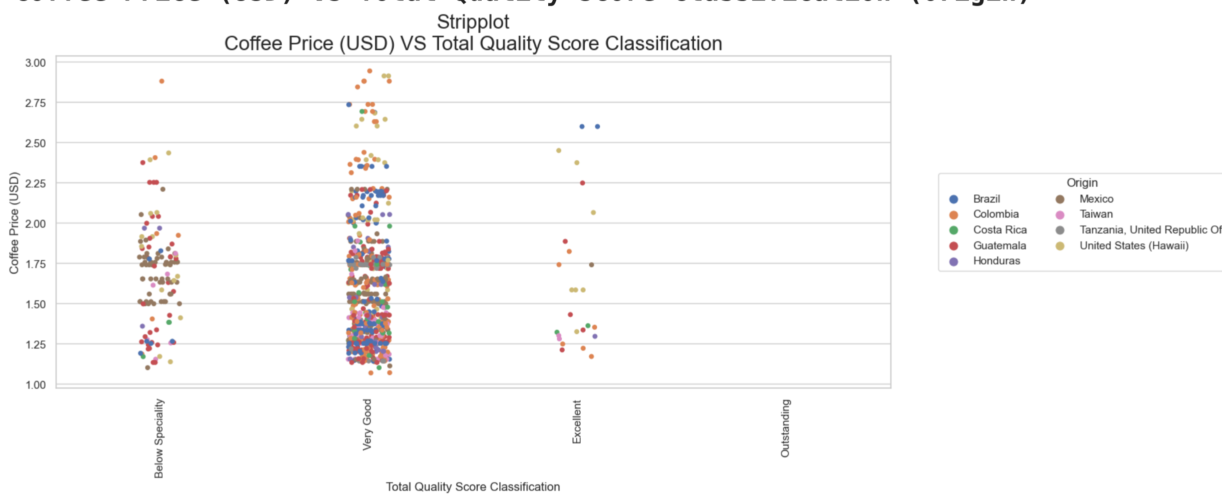

Swarmplot
Total Quality Score VS Total Quality Score Classification
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Total Quality Score
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

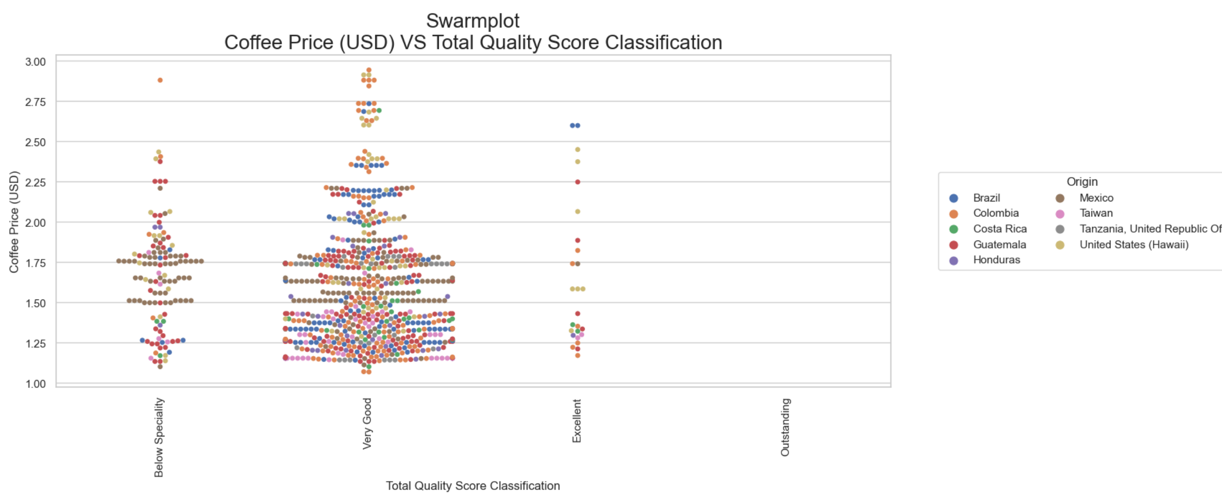

Boxplot
Total Quality Score VS Total Quality Score Classification
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Total Quality Score
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

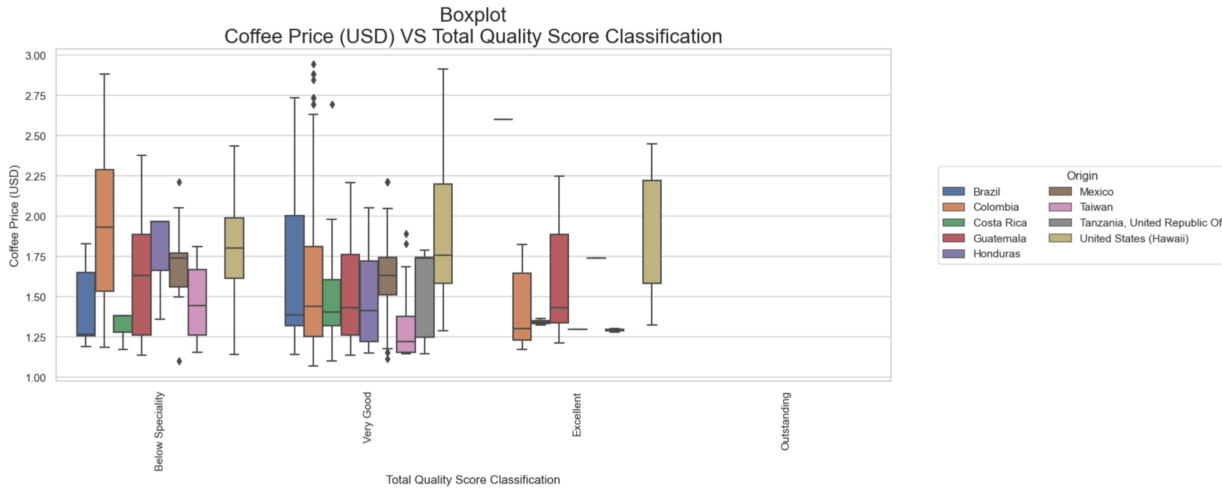

Violinplot
Total Quality Score VS Total Quality Score Classification
87.5
85.0
82.5
80.0
77.5
75.0
72.5
70.0
Total Quality Score
Below Speciality
Very Good
Excellent
Outstanding
Total Quality Score Classification
Total Quality Score Classification
Below Speciality
Excellent
Very Good
Outstanding

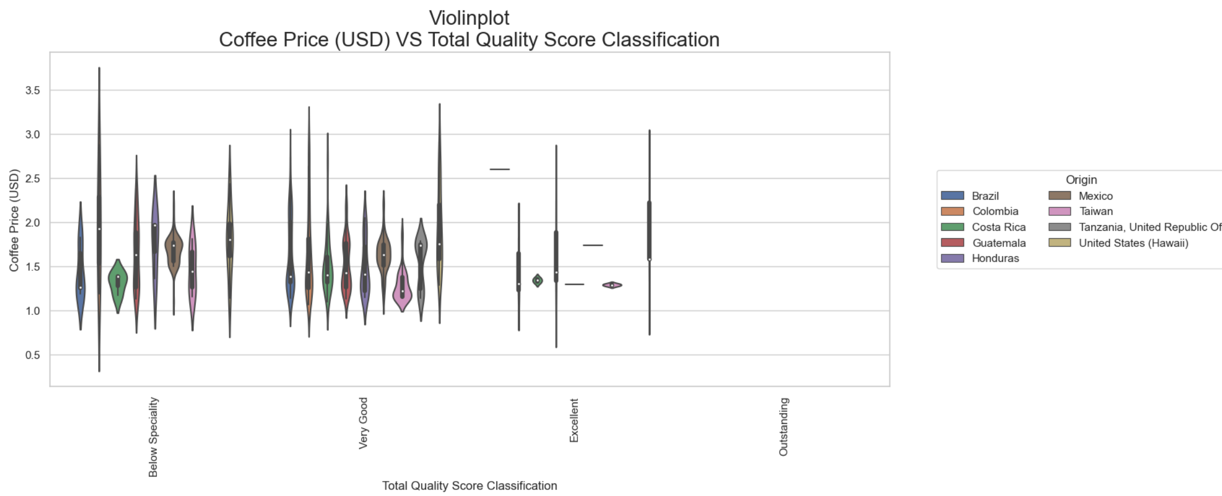


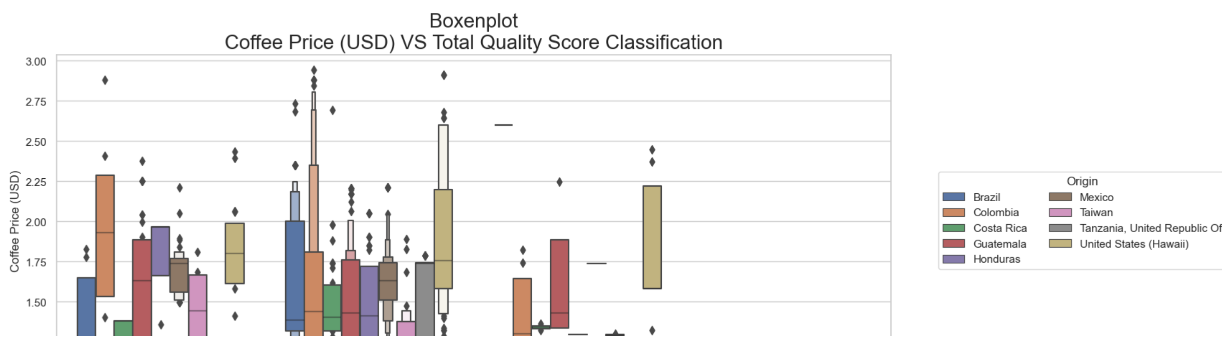
Coffee Price (USD) VS Total Quality Score Classification (Origin)











Project Phase 1

Project Objective

Check whether coffee total quality score is systematically higher or different in some countries than in others.

Note: The sample data is based on coffee total quality score from 30 April, 2010 to 19 January, 2018.

Problem Solving Steps

1. Ensure that the sample of coffee total quality score is normally distributed.

- If the sample is not normally distributed, test for normal distribution after performing data transformation (box-cox, yeo-johnson, log-normal, and square root) on it.

2. If the sample of coffee total quality score is still not normally distributed after performing data transformation. Kruskal–Wallis H test and Mann-Whitney U Test (non-parametric tests) are used instead of ANOVA and Turkey's Test (parametric tests).

Normality Check

Overview

In [45]:

```
# Inputs for normality check.
df = cfpq_a30_df
col = 'Total Quality Score'
```

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [46]:

```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Total Quality Score (09 April, 2010 – 19 January, 2018)
stat=0.89, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Total Quality Score (09 April, 2010 – 19 January, 2018)
stat=0.89, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Total Quality Score (09 April, 2010 – 19 January, 2018)
stat=0.89, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [47]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

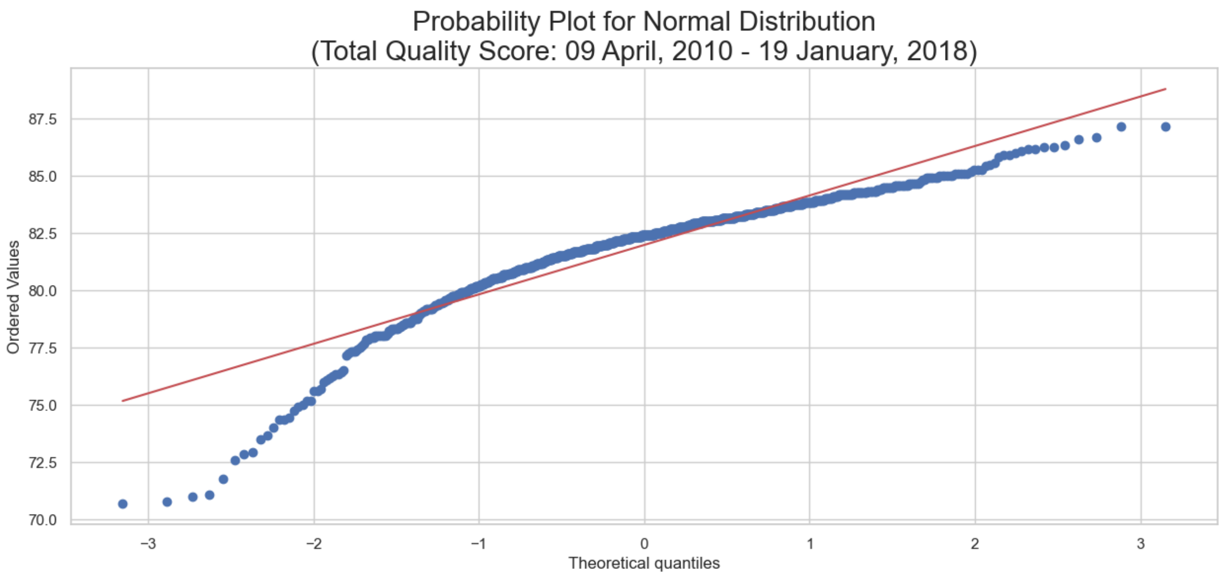
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Total Quality Score (09 April, 2010 - 19 January, 2018)
stat=282.08, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that the population is not normally-distributed.

D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Total Quality Score (09 April, 2010 - 19 January, 2018)
stat=282.08, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that the population is not normally-distributed.

D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Total Quality Score (09 April, 2010 - 19 January, 2018)
stat=282.08, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that the population is not normally-distributed.

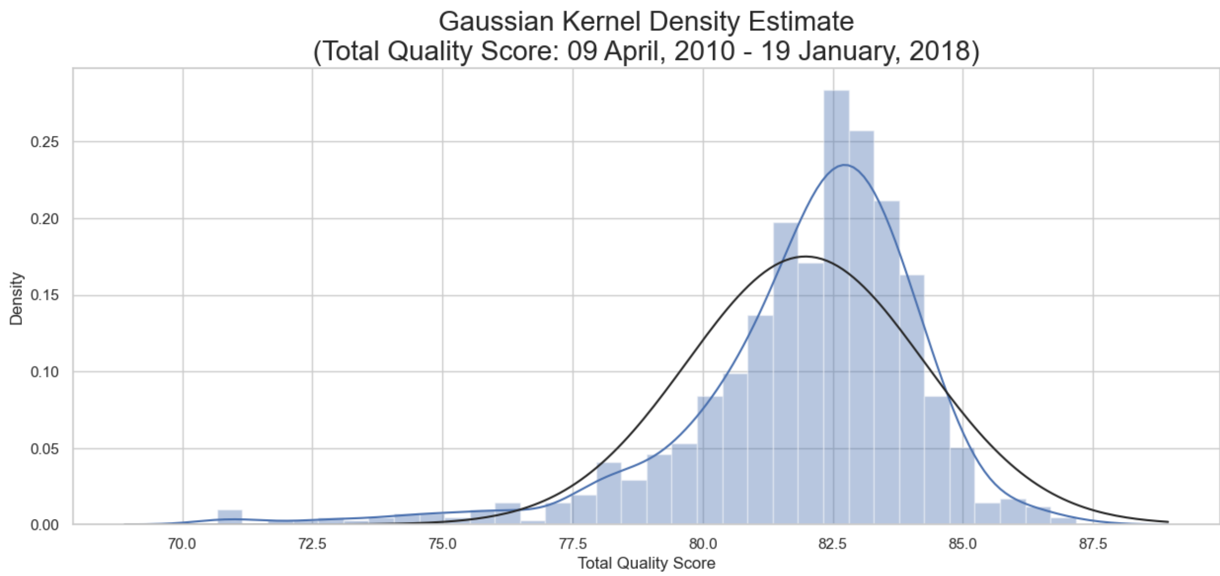
Probability Plot

In [48]:
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [49]:
```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Transformations

After Box-Cox Transformation

Overview

In [50]:
```
# Select transformation category.
trans_cat = 'Box-Cox'

# Add a box-cox transformation of total quality score column in new_cfpq_a30_
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after box-cox transformation.
box_cox, best_Lambda_maxlog = boxcox(new_cfpq_a30_df['Total Quality Score'])

# Create a box-cox transformation of total quality score column in new_cfpq_a
new_cfpq_a30_df['Box-Cox Transformation of Total Quality Score'] = box_cox

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[50]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25

Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [51]:

```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Total Quality Score (09 Apri
l, 2010 - 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Total Quality Score (09 Apri
l, 2010 - 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Total Quality Score (09 Apri
l, 2010 - 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [52]:

```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

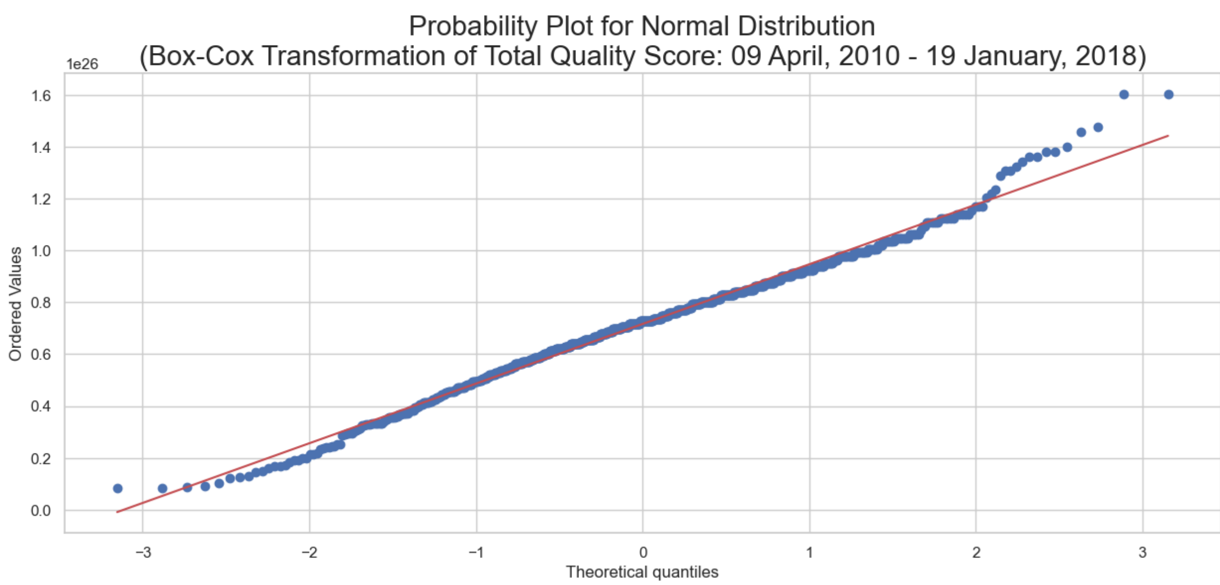
```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Total Quality Score (09
April, 2010 - 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

```
D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Total Quality Score (09
April, 2010 - 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Total Quality Score (09
April, 2010 - 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [53]:

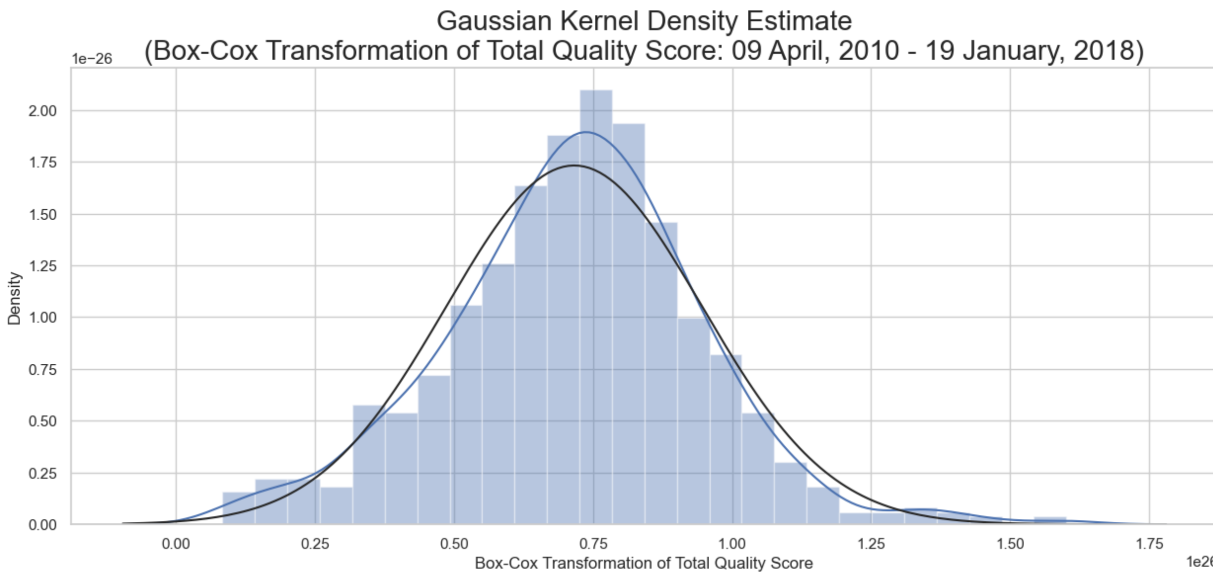
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [54]:

```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, box-cox transformed sample is not normally distributed.

After Yeo-Johnson Transformation

Overview

In [55]:
```
# Select transformation category.
trans_cat = 'Yeo-John'

# Add a yeo-johnson transformation of total quality score column in new_cfpq_
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after yeo-johnson transformation.
yeo_john, best_Lambda_Yeo_John = yeojohnson(new_cfpq_a30_df['Total Quality Sc

# Create a box-cox transformation of total quality score column in new_cfpq_a
new_cfpq_a30_df['Yeo-Johnson Transformation of Total Quality Score'] = box_co

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[55]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Trans of To
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.00
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro-Wilk Test

In [56]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Yeo-Johnson Transformation of Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro-Wilk Test at 5.0% level of significance,
Shapiro-Wilk Test for Yeo-Johnson Transformation of Total Quality Score (09 A
pril, 2010 - 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

```
Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Yeo–Johnson Transformation of Total Quality Score (09 A
pril, 2010 – 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally–distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Yeo–Johnson Transformation of Total Quality Score (09 A
pril, 2010 – 19 January, 2018)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally–distributed.
```

D'Agostino-Pearson Test

In [57]:

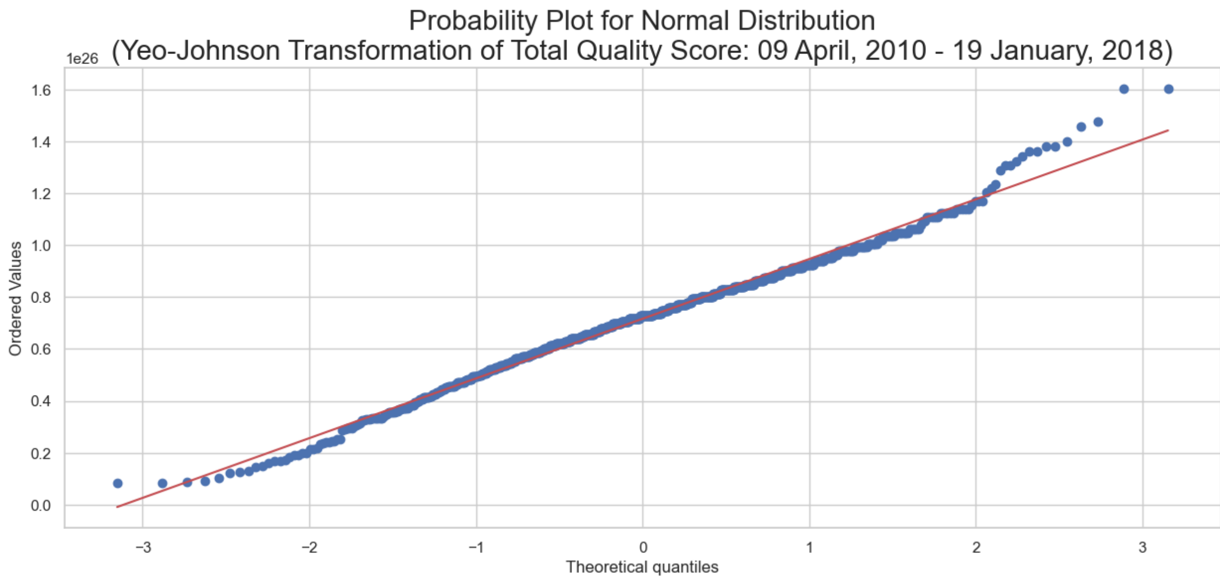
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino–Pearson Test at 5.0% level of significance,
D'Agostino–Pearson Test for Yeo–Johnson Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally–distributed.


D'Agostino–Pearson Test at 2.5% level of significance,
D'Agostino–Pearson Test for Yeo–Johnson Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally–distributed.


D'Agostino–Pearson Test at 1.0% level of significance,
D'Agostino–Pearson Test for Yeo–Johnson Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=11.96, p=0.003
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally–distributed.
```

Probability Plot

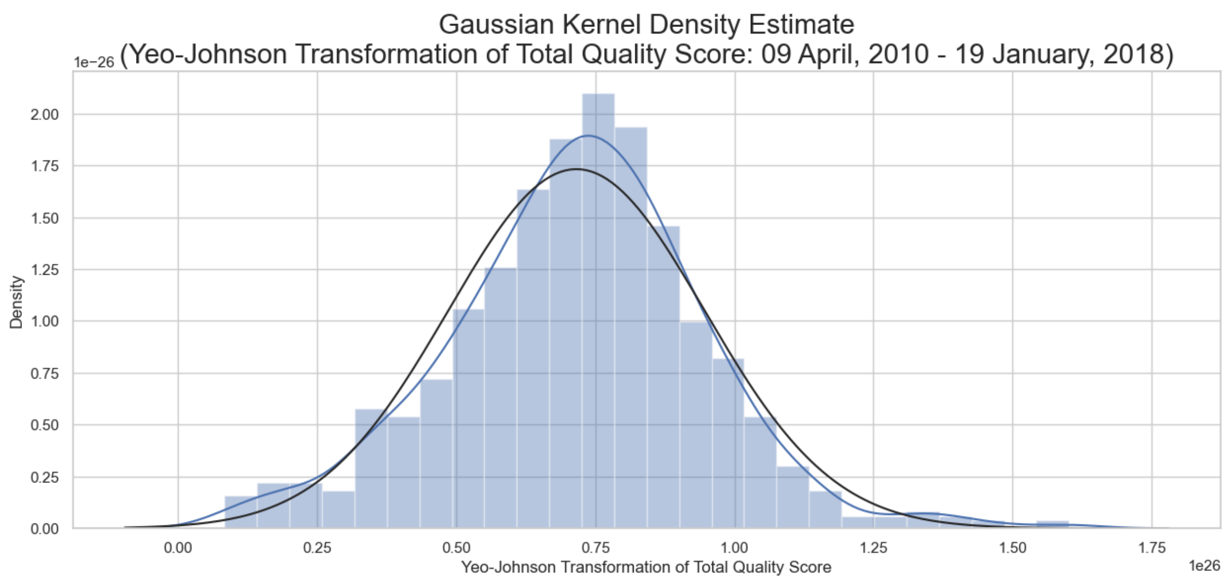
In [58]:

```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [59]:
```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, yeo-johnson transformed sample is not normally distributed.

After Log-Normal Transformation

Overview

In [60]:
```
# Select transformation category.
trans_cat = 'Log-Normal'

# Add a log-normal transformation of total quality score column in new_cfpq_a
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after log-normal transformation.
log_cfq = -np.log10(new_cfpq_a30_df['Total Quality Score'])

# Create a log-normal transformation of total quality score column in new_cfp
new_cfpq_a30_df['Log-Normal Transformation of Total Quality Score'] = log_cfq

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[60]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Trans of To
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5!

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro-Wilk Test

In [61]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Log-Normal Transformation of Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro-Wilk Test at 5.0% level of significance,
Shapiro-Wilk Test for Log-Normal Transformation of Total Quality Score (09 Ap
ril, 2010 - 19 January, 2018)
stat=0.88, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

```
Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Log–Normal Transformation of Total Quality Score (09 Ap
ril, 2010 – 19 January, 2018)
stat=0.88, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally–distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Log–Normal Transformation of Total Quality Score (09 Ap
ril, 2010 – 19 January, 2018)
stat=0.88, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally–distributed.
```

D'Agostino-Pearson Test

In [62]:

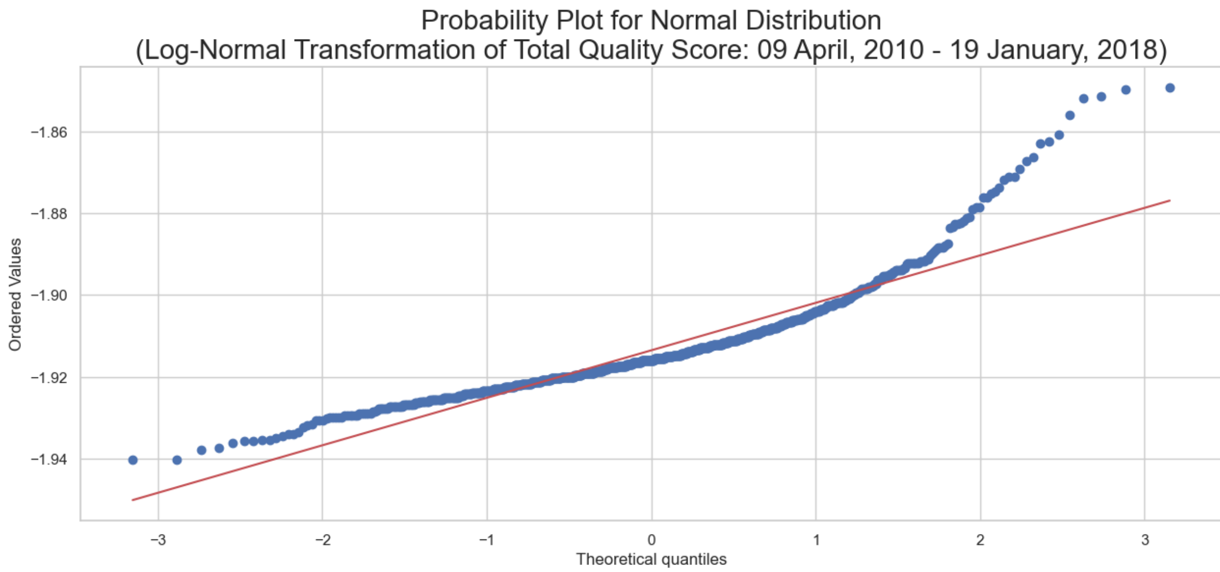
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino–Pearson Test at 5.0% level of significance,
D'Agostino–Pearson Test for Log–Normal Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=321.26, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally–distributed.


D'Agostino–Pearson Test at 2.5% level of significance,
D'Agostino–Pearson Test for Log–Normal Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=321.26, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally–distributed.


D'Agostino–Pearson Test at 1.0% level of significance,
D'Agostino–Pearson Test for Log–Normal Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=321.26, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally–distributed.
```

Probability Plot

In [63]:

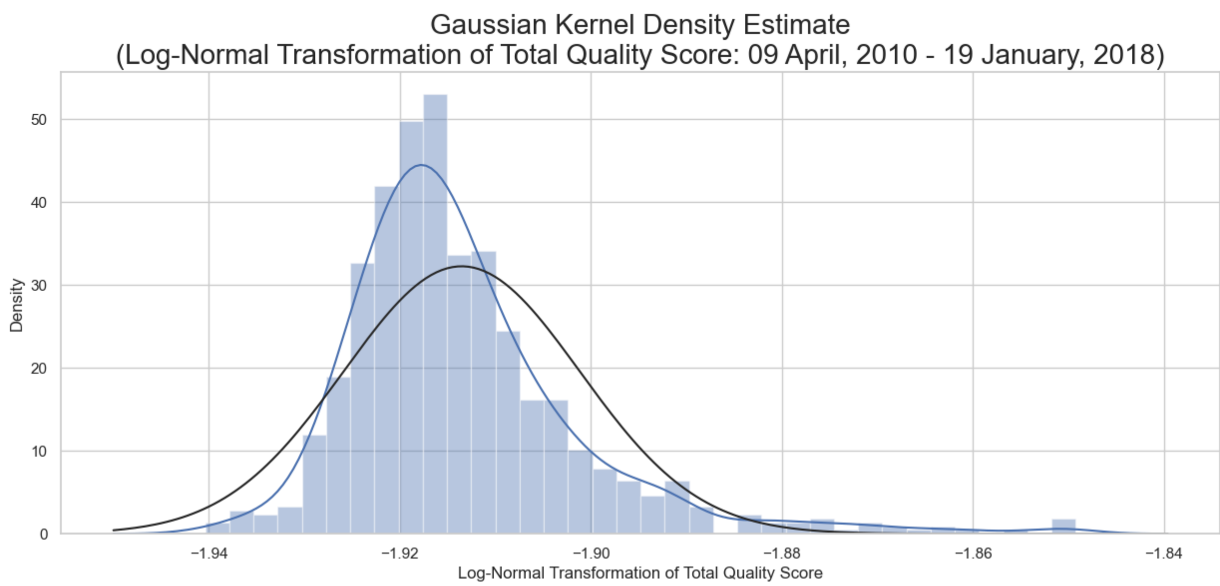
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [64]:

```python
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, log-normal transformed sample is not normally distributed.

After Square Root Transformation

Overview

In [65]:
```
# Select transformation category.
trans_cat = 'Square Root'

# Add a square root transformation of total quality score column in new_cfpq_
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after square root transformation.
sqrt_cfq = np.sqrt(87.17 - new_cfpq_a30_df['Total Quality Score'])

# Create a square root transformation of total quality score column in new_cf
new_cfpq_a30_df['Square Root Transformation of Total Quality Score'] = sqrt_c

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[65]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Tran: of Tc
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5!

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [66]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Square Root Transformation of Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Total Quality Score (09 A
pril, 2010 – 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

```
Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Total Quality Score (09 A
pril, 2010 – 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Total Quality Score (09 A
pril, 2010 – 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [67]:

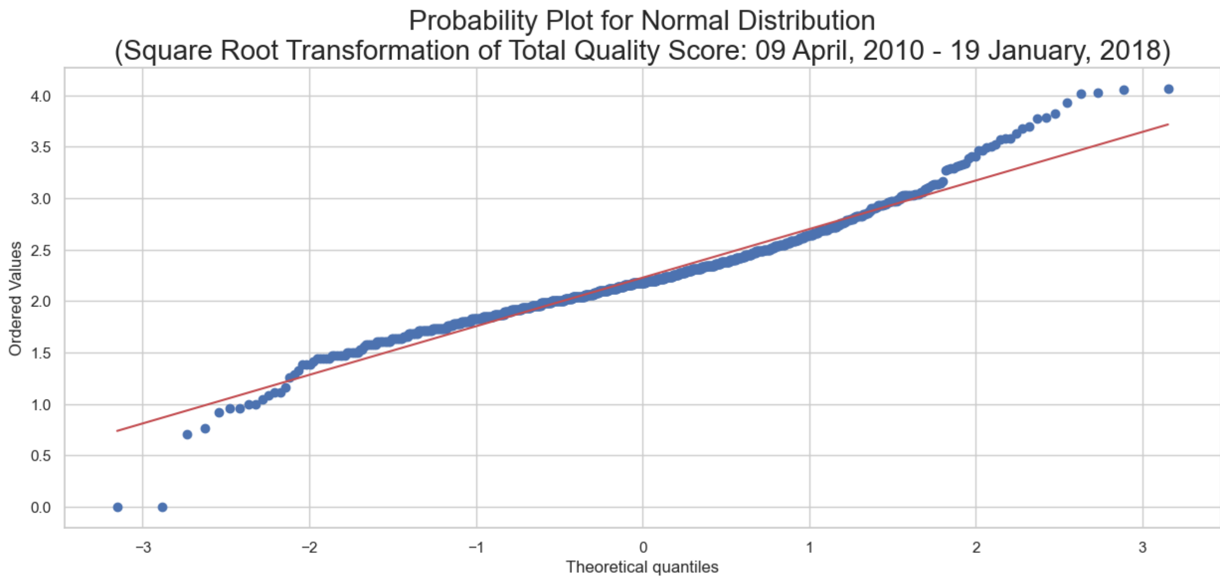
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino–Pearson Test at 5.0% level of significance,
D'Agostino–Pearson Test for Square Root Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=72.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


D'Agostino–Pearson Test at 2.5% level of significance,
D'Agostino–Pearson Test for Square Root Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=72.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino–Pearson Test at 1.0% level of significance,
D'Agostino–Pearson Test for Square Root Transformation of Total Quality Score
(09 April, 2010 – 19 January, 2018)
stat=72.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

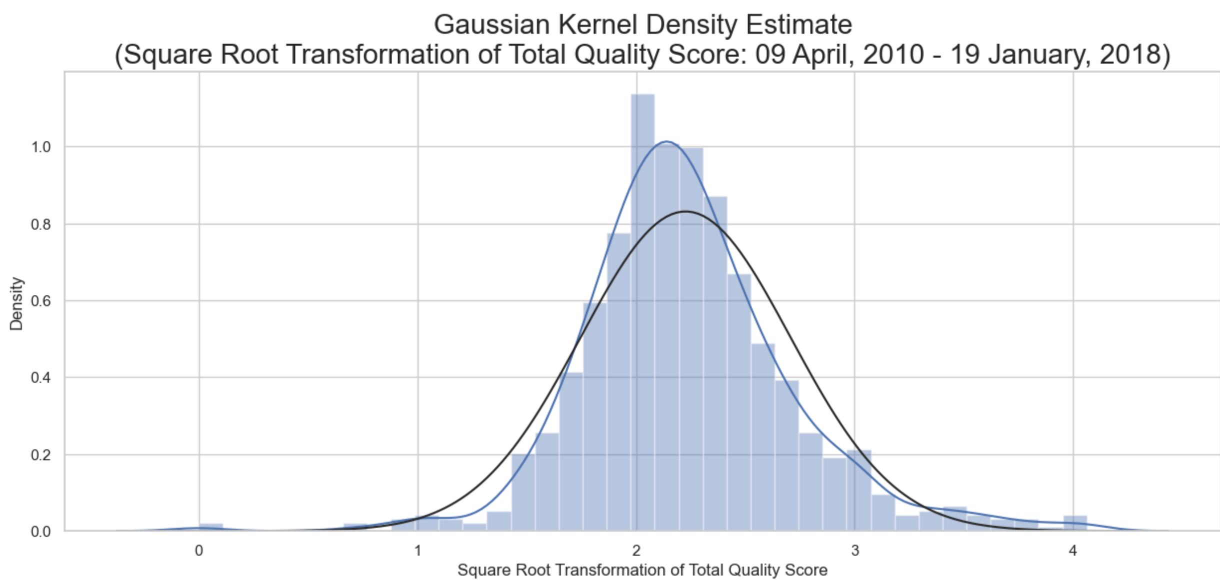
In [68]:

```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [69]:
```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```

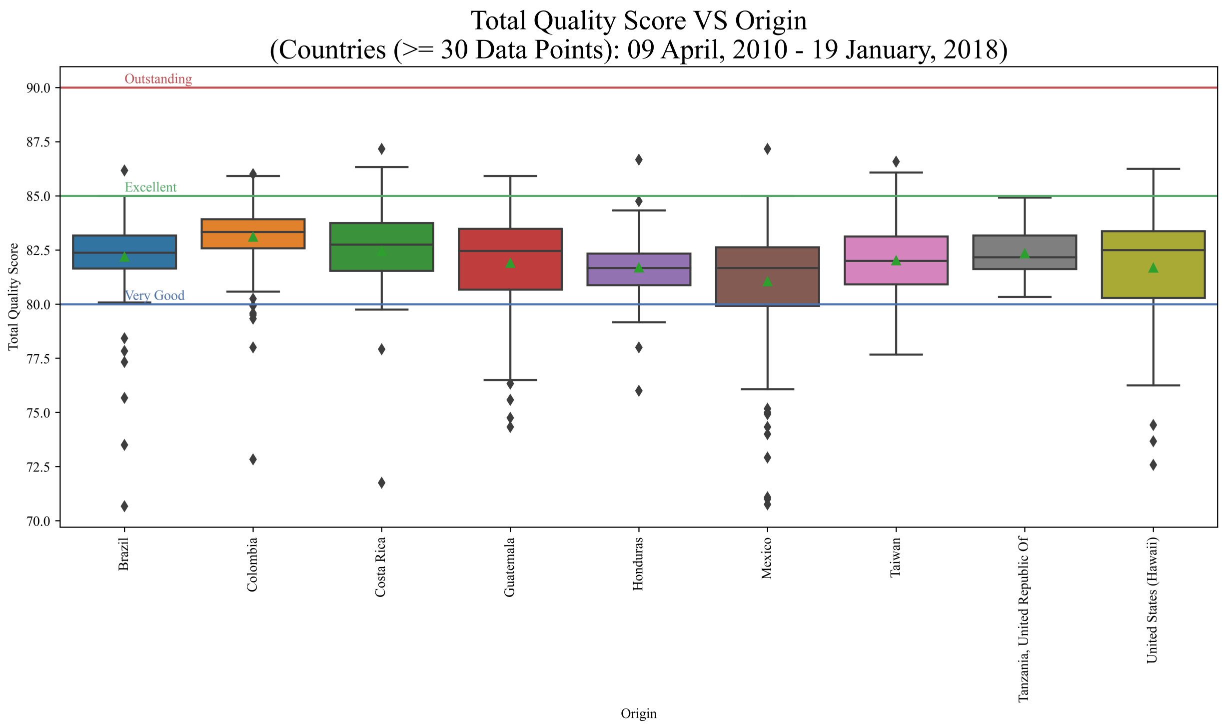


Hence, square root transformed sample is not normally distributed.

Data Visualizations

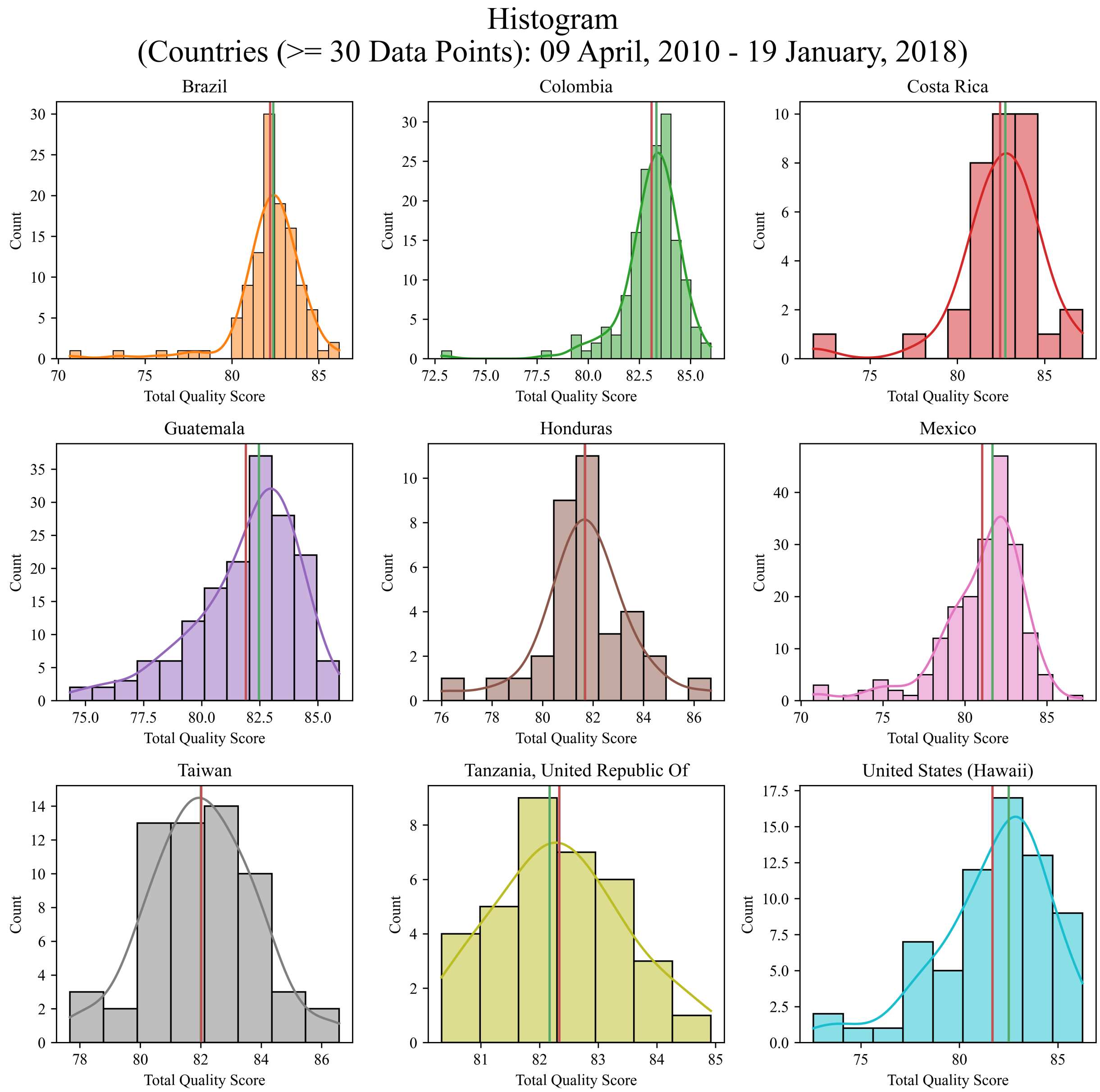
Boxplot

In [70]:
```
# Plot a boxplot for countries with >= 30 data points.
# Inputs for cfpq_boxplot func.
df = cfpq_a30_df
x = 'Origin'
y = 'Total Quality Score'
hue = None
curr_mean_cfp = None
pop_mean_cfp = None
samp_mean_cfp = None

# Run cfpq_boxplot func.
myfcf_gra.cfpq_boxplot(df, x, y, hue, curr_mean_cfp, pop_mean_cfp, samp_mean_
```



Histogram

In [71]:
```
# Plot a histrogram for countries with >= 30 data points.
# Inputs for ctry_a30_histplot_9df func.
df1 = bra_df
df2 = co_df
df3 = cr_df
df4 = gt_df
df5 = hond_df
df6 = mx_df
df7 = tw_df
df8 = urt_df
df9 = hi_df
y = 'Total Quality Score'
sub_ttl_lt = ctry_a30_lt

# Run ctry_a30_histplot_9df func.
myfcf_gra.ctry_a30_histplot_9df(df1, df2, df3, df4, df5, df6, df7, df8, df9,
```




```
Note: Green line: Median; Red line: Mean
```

Non-parametric Tests

Perform Kruskal–Wallis H test

To test H_0: μ_1 = μ_2 = μ_3 = μ_4 = μ_5 = μ_6 = μ_7 = μ_8 = μ_9

Against H_1: μ_1 ≠ μ_2 = μ_3 = μ_4 = μ_5 = μ_6 = μ_7 = μ_8 = μ_9

In [72]:
```
# Perform kruskal-wallis test for countries with >= 30 data points.
# Inputs for print_mult_krus_wall_test_9df func.
df1 = bra_df
df2 = co_df
df3 = cr_df
df4 = gt_df
df5 = hond_df
df6 = mx_df
df7 = tw_df
df8 = urt_df
df9 = hi_df
x = 'Origin'
y = 'Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_krus_wall_test_9df func.
myfcf_sta.print_mult_krus_wall_test_9df(df1, df2, df3, df4, df5, df6, df7, df
```

```
Kruskal-Wallis H Test at 5.0% level of significance,
Kruskal-Wallis H Test for Total Quality Score and Origin (09 April, 2010 - 19
January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that total quality score is systematically higher or different i
n some countries than in others.


Kruskal-Wallis H Test at 2.5% level of significance,
Kruskal-Wallis H Test for Total Quality Score and Origin (09 April, 2010 - 19
January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that total quality score is systematically higher or different i
n some countries than in others.


Kruskal-Wallis H Test at 1.0% level of significance,
Kruskal-Wallis H Test for Total Quality Score and Origin (09 April, 2010 - 19
January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that total quality score is systematically higher or different i
n some countries than in others.
```

Hence, based on the findings, there is significant evidence, at 1.0% to 5.0% level of significance, that total quality score is systematically higher or different in some countries than in others.

Perform Mann-Whitney U Test

To test H_0: μ_1 = μ_2

Against H_1: μ_1 ≠ μ_2

In [73]:

```
# Print out mann-whitney u test results at various significant level.
# Inputs for print_mult_mw_test func.
df = cfpq_a30_df
x = 'Origin'
y = 'Total Quality Score'
alpha_lt = [0.05, 0.025, 0.01] # alpha: level of significance.

# Run print_mult_mw_test func.
myfcf_sta.print_mult_mann_whit_test(df, x, y, alpha_lt, start_date, end_date)
```

Mann-Whitney U Test at 5.0% level of significance,
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and United States (Hawaii)) (09 April, 2010 - 19 January, 2018)
stat=5544.50, p=0.798
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for United States (Hawaii) is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Mexico) (09 April, 2010 - 19 January, 2018)
stat=19399.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Mexico is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Colombia) (09 April, 2010 - 19 January, 2018)
stat=8134.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Colombia is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Brazil) (09 April, 2010 - 19 January, 2018)
stat=9137.00, p=0.696
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Brazil is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Costa Rica) (09 April, 2010 - 19 January, 2018)
stat=2434.00, p=0.190
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Costa Rica is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=2784.50, p=0.870
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=5120.50, p=0.541
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Honduras) (09 April, 2010 - 19 January, 2018)
stat=3306.50, p=0.124
Do not reject Ho and conclude that there is no significant evidence, at 5.0%

level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Mexico) (09 April, 2010 - 19 January, 2018)

stat=7729.00, p=0.025

Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Mexico is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Colombia) (09 April, 2010 - 19 January, 2018)

stat=3417.00, p=0.000

Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Brazil) (09 April, 2010 - 19 January, 2018)

stat=3670.50, p=0.533

Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Brazil is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=988.00, p=0.195

Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=1114.00, p=0.683

Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=2030.00, p=0.925

Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Honduras) (09 April, 2010 - 19 January, 2018)

stat=1299.00, p=0.374

Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Colombia) (09 April, 2010 - 19 January, 2018)

stat=6005.00, p=0.000

Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Brazil) (09 April, 2010 - 19 January, 2018)

stat=7885.00, p=0.000

Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Brazil is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=2158.50, p=0.001
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Costa Rica is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=2384.00, p=0.005
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=4771.50, p=0.031
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Taiwan is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Honduras) (09 April, 2010 - 19 January, 2018)

stat=3168.00, p=0.501
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Brazil) (09 April, 2010 - 19 January, 2018)

stat=12237.50, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Brazil is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=3190.50, p=0.066
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=3803.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=6538.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Taiwan is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Honduras) (09 April, 2010 - 19 January, 2018)

stat=4198.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Honduras is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=1789.50, p=0.290
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Tanzania,

United Republic Of) (09 April, 2010 – 19 January, 2018)
stat=2103.50, p=0.747
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Brazil and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=3960.50, p=0.134
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Taiwan is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Brazil and Honduras) (09 April, 2010 – 19 January, 2018)
stat=2614.00, p=0.010
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Honduras is not equal.
Mann–Whitney U Test for Total Quality Score and Origin (Costa Rica and Tanzania, United Republic Of) (09 April, 2010 – 19 January, 2018)
stat=699.00, p=0.312
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Costa Rica and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=1272.50, p=0.087
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Taiwan is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Costa Rica and Honduras) (09 April, 2010 – 19 January, 2018)
stat=812.50, p=0.019
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Honduras is not equal.
Mann–Whitney U Test for Total Quality Score and Origin (Tanzania, United Republic Of and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=1178.00, p=0.325
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Tanzania, United Republic Of and mean rank of Total Quality Score for Taiwan is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Tanzania, United Republic Of and Honduras) (09 April, 2010 – 19 January, 2018)
stat=779.00, p=0.051
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Tanzania, United Republic Of and mean rank of Total Quality Score for Honduras is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Taiwan and Honduras) (09 April, 2010 – 19 January, 2018)
stat=1161.50, p=0.392
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Total Quality Score for Taiwan and mean rank of Total Quality Score for Honduras is equal.

Mann–Whitney U Test at 2.5% level of significance,
Mann–Whitney U Test for Total Quality Score and Origin (Guatemala and United States (Hawaii)) (09 April, 2010 – 19 January, 2018)

stat=5544.50, p=0.798
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for United States (Hawaii) is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Mexico) (09 April, 2010 - 19 January, 2018)

stat=19399.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Mexico is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Colombia) (09 April, 2010 - 19 January, 2018)

stat=8134.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Brazil) (09 April, 2010 - 19 January, 2018)

stat=9137.00, p=0.696
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Brazil is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=2434.00, p=0.190
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=2784.50, p=0.870
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=5120.50, p=0.541
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Honduras) (09 April, 2010 - 19 January, 2018)

stat=3306.50, p=0.124
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Mexico) (09 April, 2010 - 19 January, 2018)

stat=7729.00, p=0.025
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Mexico is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Colombia) (09 April, 2010 - 19 January, 2018)

stat=3417.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawai

i) and Brazil) (09 April, 2010 – 19 January, 2018)
stat=3670.50, p=0.533
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Brazil is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Costa Rica) (09 April, 2010 – 19 January, 2018)
stat=988.00, p=0.195
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Tanzania, United Republic Of) (09 April, 2010 – 19 January, 2018)
stat=1114.00, p=0.683
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=2030.00, p=0.925
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Honduras) (09 April, 2010 – 19 January, 2018)
stat=1299.00, p=0.374
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Colombia) (09 April, 2010 – 19 January, 2018)
stat=6005.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Brazil) (09 April, 2010 – 19 January, 2018)
stat=7885.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Brazil is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Costa Rica) (09 April, 2010 – 19 January, 2018)
stat=2158.50, p=0.001
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Costa Rica is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Tanzania, United Republic Of) (09 April, 2010 – 19 January, 2018)
stat=2384.00, p=0.005
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=4771.50, p=0.031
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and m

ean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Honduras) (09 April, 2010 - 19 January, 2018)
stat=3168.00, p=0.501
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Honduras is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Brazil) (09 April, 2010 - 19 January, 2018)
stat=12237.50, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Brazil is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Costa Rica) (09 April, 2010 - 19 January, 2018)
stat=3190.50, p=0.066
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Costa Rica is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=3803.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=6538.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Taiwan is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Honduras) (09 April, 2010 - 19 January, 2018)
stat=4198.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Honduras is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Costa Rica) (09 April, 2010 - 19 January, 2018)
stat=1789.50, p=0.290
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Costa Rica is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=2103.50, p=0.747
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=3960.50, p=0.134
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Honduras) (09 April, 2010 - 19 January, 2018)
stat=2614.00, p=0.010
Reject Ho and conclude that there is significant evidence, at 2.5% level of s

ignificance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Honduras is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Costa Rica and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=699.00, p=0.312
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Costa Rica and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=1272.50, p=0.087
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Costa Rica and Honduras) (09 April, 2010 - 19 January, 2018)
stat=812.50, p=0.019
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Honduras is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Tanzania, United Republic Of and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=1178.00, p=0.325
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Tanzania, United Republic Of and mean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Tanzania, United Republic Of and Honduras) (09 April, 2010 - 19 January, 2018)
stat=779.00, p=0.051
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Tanzania, United Republic Of and mean rank of Total Quality Score for Honduras is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Taiwan and Honduras) (09 April, 2010 - 19 January, 2018)
stat=1161.50, p=0.392
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Total Quality Score for Taiwan and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test at 1.0% level of significance,
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and United States (Hawaii)) (09 April, 2010 - 19 January, 2018)
stat=5544.50, p=0.798
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for United States (Hawaii) is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Mexico) (09 April, 2010 - 19 January, 2018)
stat=19399.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Mexico is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Colombia) (09 April, 2010 - 19 January, 2018)
stat=8134.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean ran

k of Total Quality Score for Colombia is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Brazil) (09 April, 2010 - 19 January, 2018)
stat=9137.00, p=0.696
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Brazil is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Costa Rica) (09 April, 2010 - 19 January, 2018)
stat=2434.00, p=0.190
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Costa Rica is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=2784.50, p=0.870
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Taiwan) (09 April, 2010 - 19 January, 2018)
stat=5120.50, p=0.541
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Taiwan is equal.
Mann-Whitney U Test for Total Quality Score and Origin (Guatemala and Honduras) (09 April, 2010 - 19 January, 2018)
stat=3306.50, p=0.124
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Guatemala and mean rank of Total Quality Score for Honduras is equal.
Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Mexico) (09 April, 2010 - 19 January, 2018)
stat=7729.00, p=0.025
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Mexico is equal.
Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Colombia) (09 April, 2010 - 19 January, 2018)
stat=3417.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Colombia is not equal.
Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Brazil) (09 April, 2010 - 19 January, 2018)
stat=3670.50, p=0.533
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Brazil is equal.
Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Costa Rica) (09 April, 2010 - 19 January, 2018)
stat=988.00, p=0.195
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Costa Rica is equal.
Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)
stat=1114.00, p=0.683
Do not reject Ho and conclude that there is no significant evidence, at 1.0%

level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=2030.00, p=0.925

Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (United States (Hawaii) and Honduras) (09 April, 2010 - 19 January, 2018)

stat=1299.00, p=0.374

Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for United States (Hawaii) and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Colombia) (09 April, 2010 - 19 January, 2018)

stat=6005.00, p=0.000

Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Colombia is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Brazil) (09 April, 2010 - 19 January, 2018)

stat=7885.00, p=0.000

Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Brazil is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=2158.50, p=0.001

Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Costa Rica is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=2384.00, p=0.005

Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=4771.50, p=0.031

Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Mexico and Honduras) (09 April, 2010 - 19 January, 2018)

stat=3168.00, p=0.501

Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Mexico and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Brazil) (09 April, 2010 - 19 January, 2018)

stat=12237.50, p=0.000

Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Brazil is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=3190.50, p=0.066
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=3803.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Tanzania, United Republic Of is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=6538.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Taiwan is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Colombia and Honduras) (09 April, 2010 - 19 January, 2018)

stat=4198.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Colombia and mean rank of Total Quality Score for Honduras is not equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Costa Rica) (09 April, 2010 - 19 January, 2018)

stat=1789.50, p=0.290
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Costa Rica is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=2103.50, p=0.747
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=3960.50, p=0.134
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Taiwan is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Brazil and Honduras) (09 April, 2010 - 19 January, 2018)

stat=2614.00, p=0.010
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Brazil and mean rank of Total Quality Score for Honduras is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Costa Rica and Tanzania, United Republic Of) (09 April, 2010 - 19 January, 2018)

stat=699.00, p=0.312
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Tanzania, United Republic Of is equal.

Mann-Whitney U Test for Total Quality Score and Origin (Costa Rica and Taiwan) (09 April, 2010 - 19 January, 2018)

stat=1272.50, p=0.087
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that mean rank of Total Quality Score for Costa Rica and mean rank of Total Quality Score for Taiwan is equal.

```
Mann–Whitney U Test for Total Quality Score and Origin (Costa Rica and Hondur
as) (09 April, 2010 – 19 January, 2018)
stat=812.50, p=0.019
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that mean rank of Total Quality Score for Costa Rica a
nd mean rank of Total Quality Score for Honduras is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Tanzania, United Repu
blic Of and Taiwan) (09 April, 2010 – 19 January, 2018)
stat=1178.00, p=0.325
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that mean rank of Total Quality Score for Tanzania, Un
ited Republic Of and mean rank of Total Quality Score for Taiwan is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Tanzania, United Repu
blic Of and Honduras) (09 April, 2010 – 19 January, 2018)
stat=779.00, p=0.051
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that mean rank of Total Quality Score for Tanzania, Un
ited Republic Of and mean rank of Total Quality Score for Honduras is equal.
Mann–Whitney U Test for Total Quality Score and Origin (Taiwan and Honduras)
(09 April, 2010 – 19 January, 2018)
stat=1161.50, p=0.392
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that mean rank of Total Quality Score for Taiwan and m
ean rank of Total Quality Score for Honduras is equal.
```

Hence, based on the findings, there is significant evidence, at 1.0% to 5% level of significance, that mean rank of total quality score for Colombia and mean rank of total quality score for Brazil, Tanzania, United Republic Of, Taiwan, Honduras, Guatemala, United States (Hawaii), and Mexico are not equal. Thus, the mean rank of total quality score of Colombian coffee is distinctly different from the rest.

Project Phase 2

Project Objective

Check whether coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Note: The sample data is based on coffee "C" future price (USD) from 30 April, 2010 to 19 January, 2018.

Problem Solving Steps

Assume the coffee price follows the coffee "C" future price.

1. Ensure that the sample of coffee future price is normally distributed (Ethiopian coffee are excluded, because there are not much coffee future contracts done in Ethiopia).
    - If the sample is not normally distributed, test for normal distribution after performing data transformation (box-cox, yeo-johnson, log-normal, and square root) on it.
2. If the sample of coffee future price is still not normally distributed after performing data transformation. Kruskal–Wallis H test and Mann-Whitney U Test (non-parametric tests)

are used instead of ANOVA and Turkey's Test (parametric tests).

Normality Check

Overview

In [74]:
```
# Inputs for normality check.
df = cfpq_a30_df
col = 'Coffee Price (USD)'
```

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [75]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
stat=0.92, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
stat=0.92, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
stat=0.92, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [76]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (09 April, 2010 – 19 January,
2018)
stat=136.07, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
```

```
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (09 April, 2010 – 19 January, 2018)

```
stat=136.07, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.
```

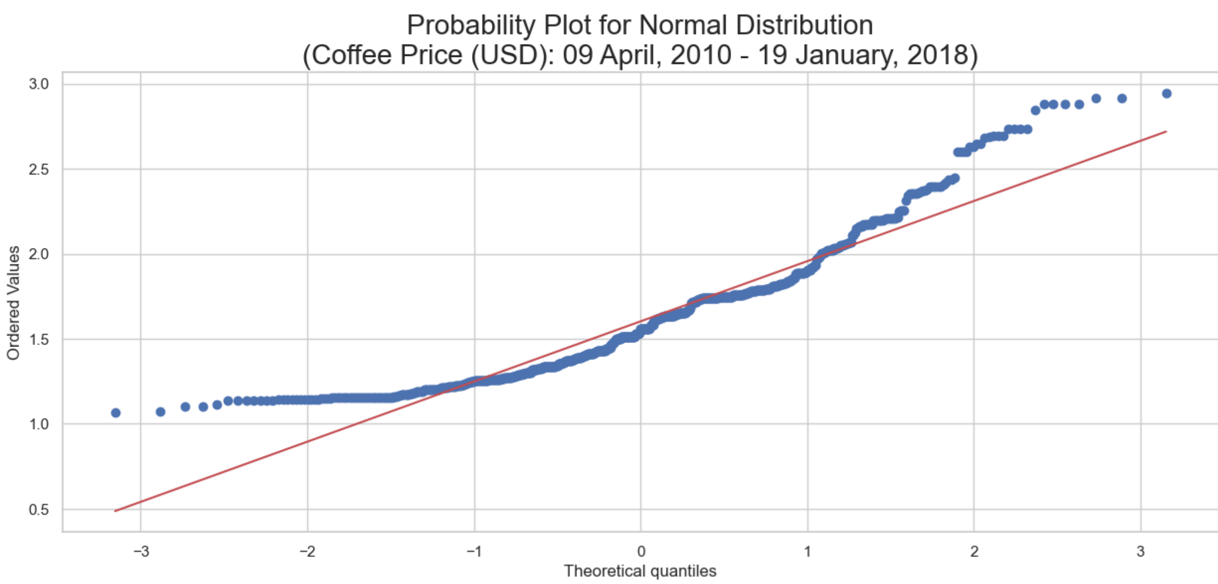
D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (09 April, 2010 – 19 January, 2018)

```
stat=136.07, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [77]:

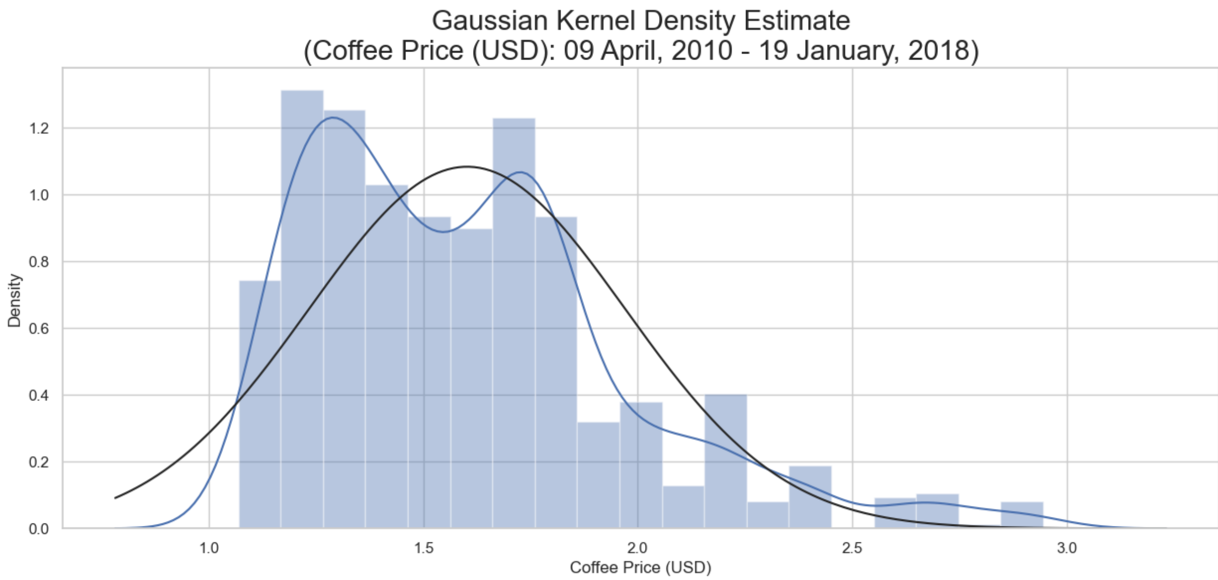
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [78]:

```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Transformations

After Box-Cox Transformation

Overview

In [79]:

```
# Select transformation category.
trans_cat = 'Box-Cox'

# Add a box-cox transformation of coffee price (usd) column in new_cfpq_a30_d
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after box-cox transformation.
box_cox, best_Lambda_maxlog = boxcox(new_cfpq_a30_df['Coffee Price (USD)'])

# Create a box-cox transformation of coffee price (usd) column in new_cfpq_a3
new_cfpq_a30_df['Box-Cox Transformation of Coffee Price (USD)'] = box_cox

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[79]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Trans of To
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [80]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
```
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
```
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.
```

Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
```
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [81]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (09 April, 2010 – 19 January, 2018)
```
stat=89.36, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (09 April, 2010 – 19 January, 2018)

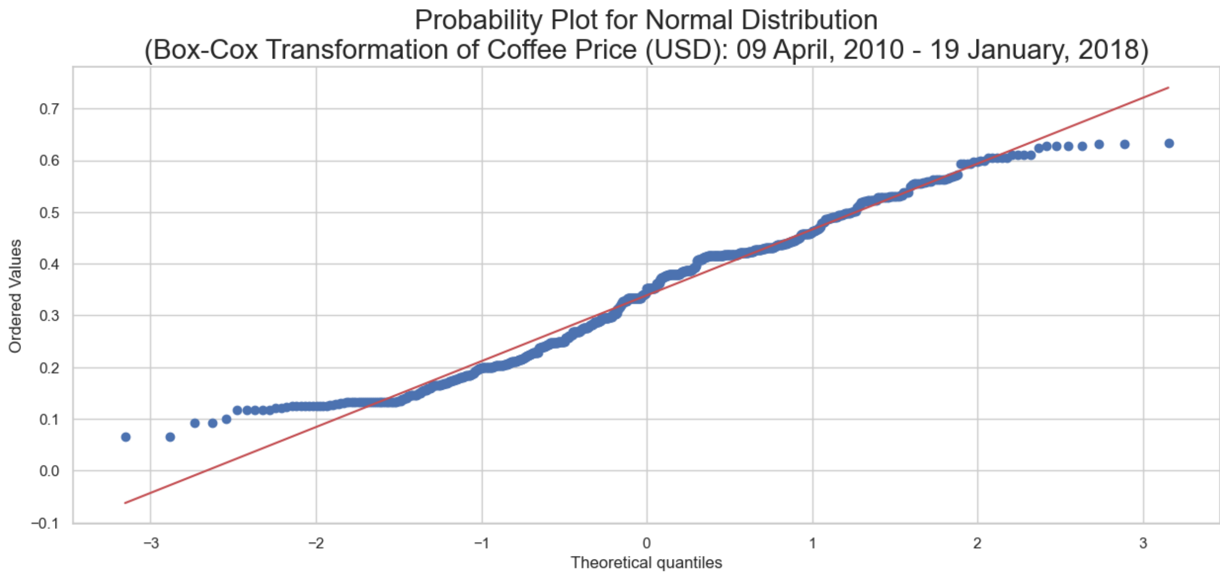
```
stat=89.36, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.
```

```
D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (09
April, 2010 - 19 January, 2018)
stat=89.36, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [82]:

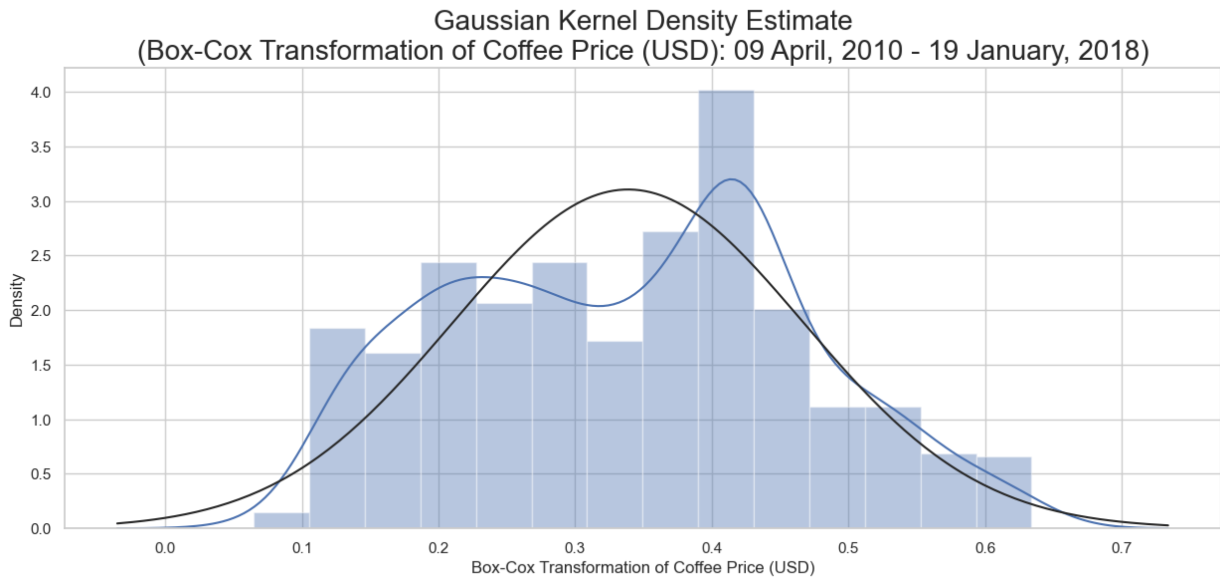
```python
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [83]:

```python
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, box-cox transformed sample is not normally distributed.

After Yeo-Johnson Transformation

Overview

In [84]:

```
# Select transformation category.
trans_cat = 'Yeo-John'

# Add a yeo-johnson transformation of coffee price (usd) column in new_cfpq_a
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after yeo-johnson transformation.
yeo_john, best_Lambda_Yeo_John = yeojohnson(new_cfpq_a30_df['Coffee Price (US

# Create a box-cox transformation of coffee price (usd) column in new_cfpq_a3
new_cfpq_a30_df['Yeo-Johnson Transformation of Coffee Price (USD)'] = box_cox

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[84]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Tran of To
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5!

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro-Wilk Test

In [85]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Yeo-Johnson Transformation of Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro-Wilk Test at 5.0% level of significance,
Shapiro-Wilk Test for Yeo-Johnson Transformation of Coffee Price (USD) (09 Ap
ril, 2010 - 19 January, 2018)
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro-Wilk Test at 2.5% level of significance,
Shapiro-Wilk Test for Yeo-Johnson Transformation of Coffee Price (USD) (09 Ap
ril, 2010 - 19 January, 2018)
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro-Wilk Test at 1.0% level of significance,
Shapiro-Wilk Test for Yeo-Johnson Transformation of Coffee Price (USD) (09 Ap
ril, 2010 - 19 January, 2018)
stat=0.97, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [86]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

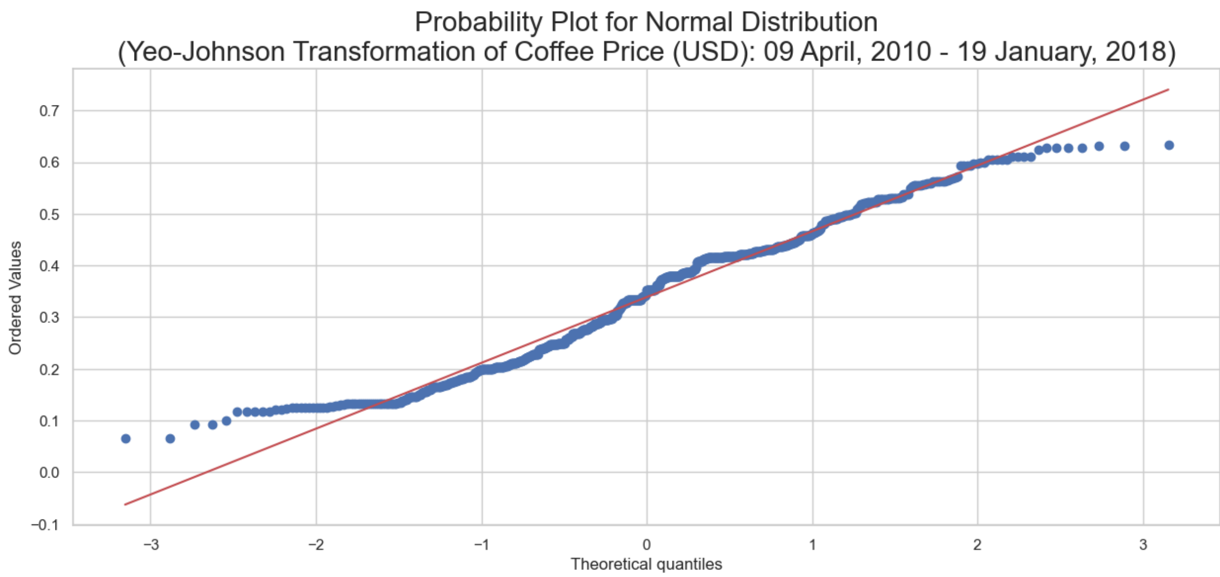
```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Yeo-Johnson Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=89.36, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Yeo-Johnson Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=89.36, p=0.000
```

```
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Yeo-Johnson Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=89.36, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [87]:

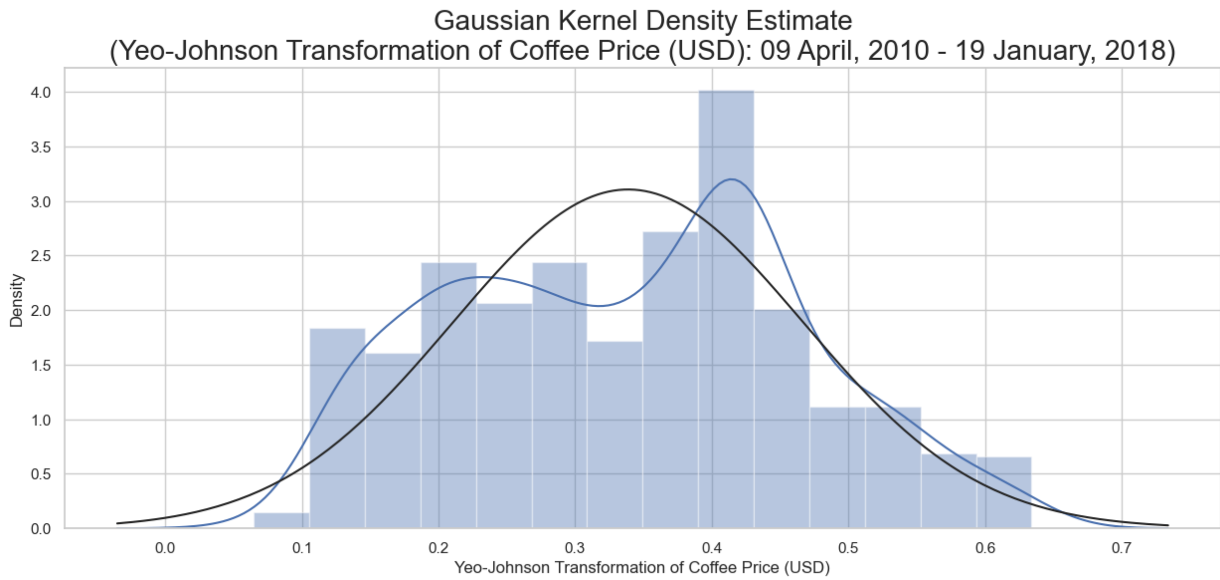
```python
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [88]:

```python
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, yeo-johnson tranformed sample is not normally distributed.

After Log-Normal Transformation

Overview

In [89]:

```
# Select transformation category.
trans_cat = 'Log-Normal'

# Add a log-normal transformation of coffee price (usd) column in new_cfpq_a3
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after log-normal transformation.
log_cfq = np.log10(new_cfpq_a30_df['Coffee Price (USD)'])

# Create a log-normal transformation of coffee price (usd) column in new_cfpq
new_cfpq_a30_df['Log-Normal Transformation of Coffee Price (USD)'] = log_cfq

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[89]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Tran of To
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5!

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [90]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Log-Normal Transformation of Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro-Wilk Test at 5.0% level of significance,
Shapiro-Wilk Test for Log-Normal Transformation of Coffee Price (USD) (09 Apr
il, 2010 - 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro-Wilk Test at 2.5% level of significance,
Shapiro-Wilk Test for Log-Normal Transformation of Coffee Price (USD) (09 Apr
il, 2010 - 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro-Wilk Test at 1.0% level of significance,
Shapiro-Wilk Test for Log-Normal Transformation of Coffee Price (USD) (09 Apr
il, 2010 - 19 January, 2018)
stat=0.96, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [91]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Log-Normal Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=38.29, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Log-Normal Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=38.29, p=0.000
```

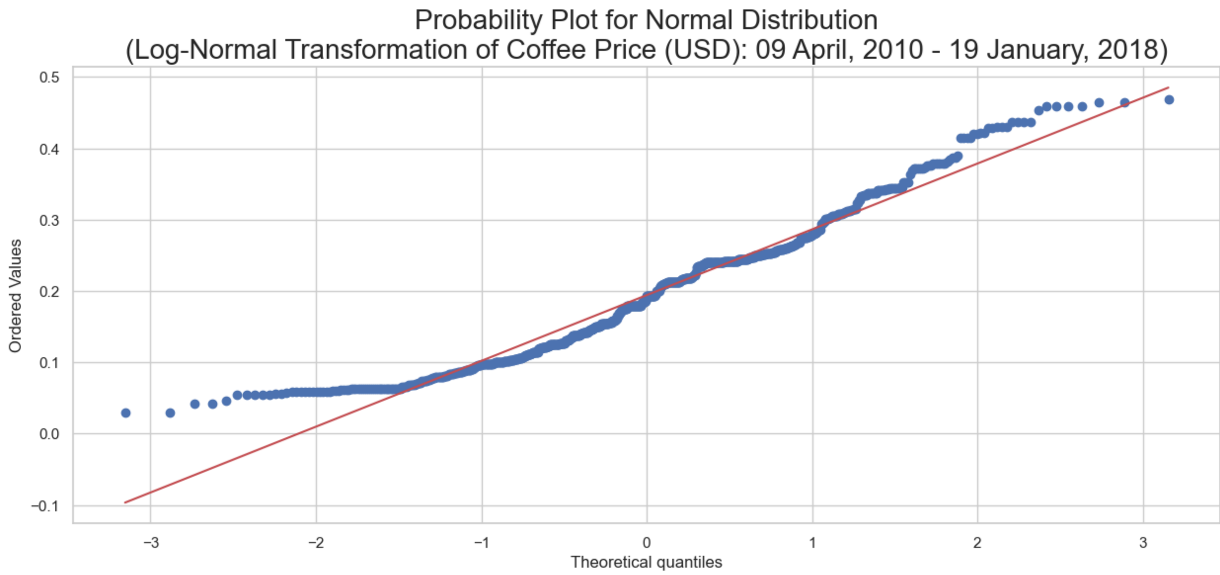
```
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Log-Normal Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)

```
stat=38.29, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

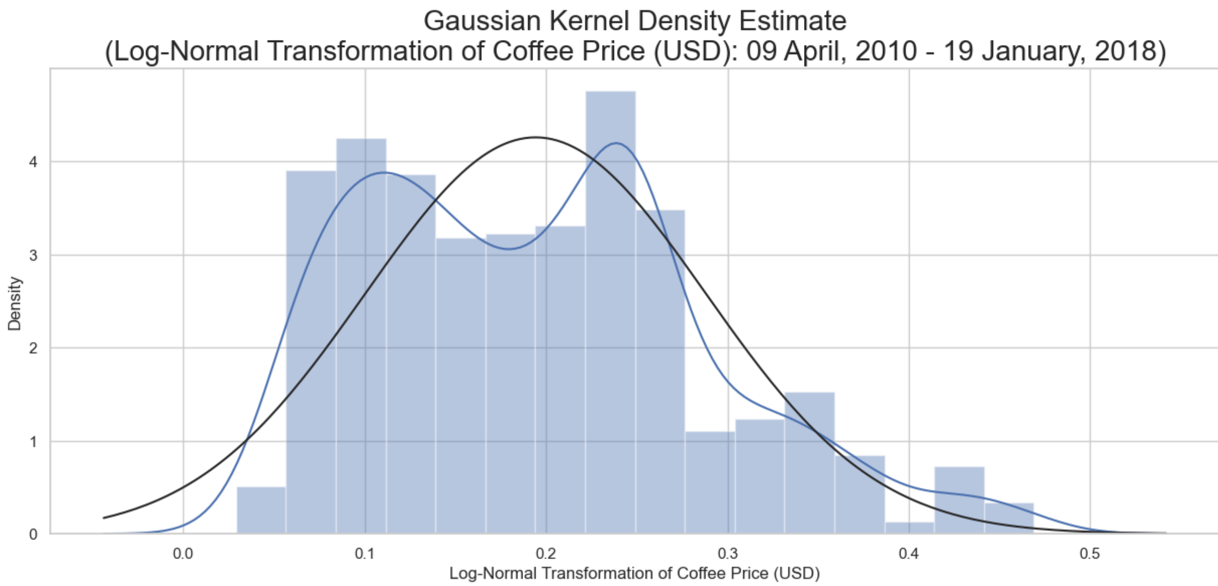
Probability Plot

In [92]:
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [93]:
```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, log-normal tranformed sample is not normally distributed.

After Square Root Transformation

Overview

In [94]:

```
# Select transformation category.
trans_cat = 'Square Root'

# Add a square root transformation of coffee price (usd) column in new_cfpq_a
new_cfpq_a30_df = cfpq_a30_df

# Compute the values after square root transformation.
sqrt_cfq = np.sqrt(new_cfpq_a30_df['Coffee Price (USD)'])

# Create a square root transformation of coffee price (usd) column in new_cfp
new_cfpq_a30_df['Square Root Transformation of Coffee Price (USD)'] = sqrt_cf

# Display head of new_cfpq_a30_df.
new_cfpq_a30_df.head()
```

Out[94]:

	Species	Origin	Date	Total Quality Score	Total Quality Score Classification	Coffee Price (USD)	Year	Box-Cox Transformation of Total Quality Score	Ye Tran of Tc
0	Arabica	Guatemala	2010-04-30	84.33	Very Good	1.3528	2010	1.004028e+26	1.0(
1	Arabica	United States (Hawaii)	2010-05-17	85.08	Excellent	1.3255	2010	1.137509e+26	1.1:
2	Arabica	United States (Hawaii)	2010-05-24	81.83	Very Good	1.3225	2010	6.569047e+25	6.5(
3	Arabica	United States (Hawaii)	2010-05-24	81.42	Very Good	1.3225	2010	6.119981e+25	6.1
4	Arabica	Mexico	2010-05-26	83.38	Very Good	1.3370	2010	8.558249e+25	8.5!

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [95]:
```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
df = new_cfpq_a30_df
col = 'Square Root Transformation of Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)
stat=0.94, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)
stat=0.94, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Square Root Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)
stat=0.94, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [96]:
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

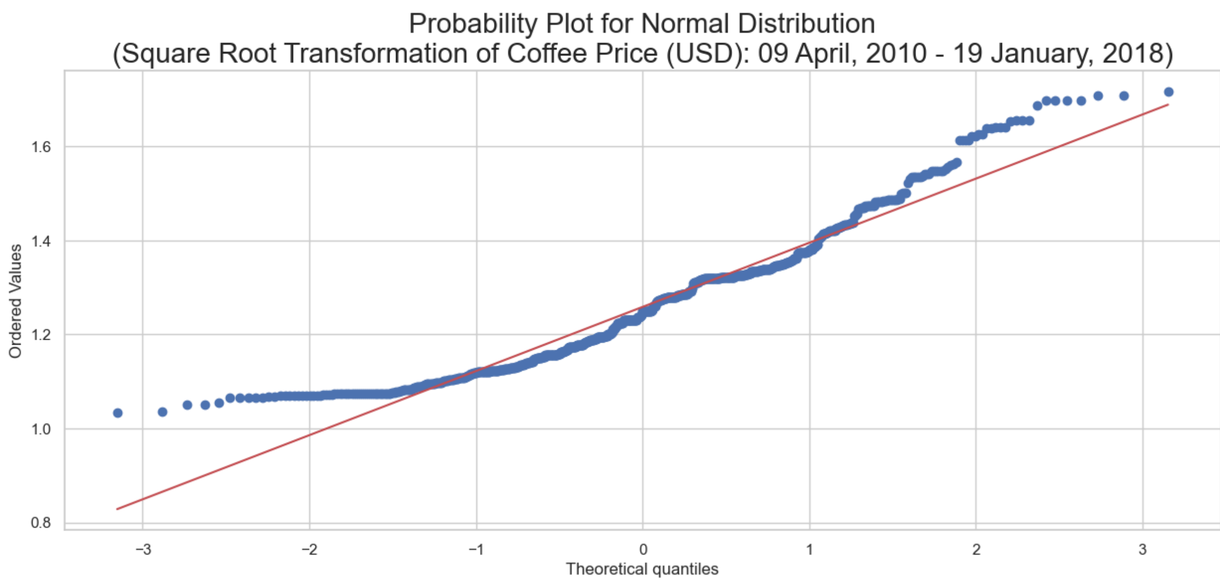
```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Square Root Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)
stat=74.34, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Square Root Transformation of Coffee Price (USD) (09 April, 2010 - 19 January, 2018)
```

```
stat=74.34, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Square Root Transformation of Coffee Price (USD)
(09 April, 2010 - 19 January, 2018)
stat=74.34, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [97]:

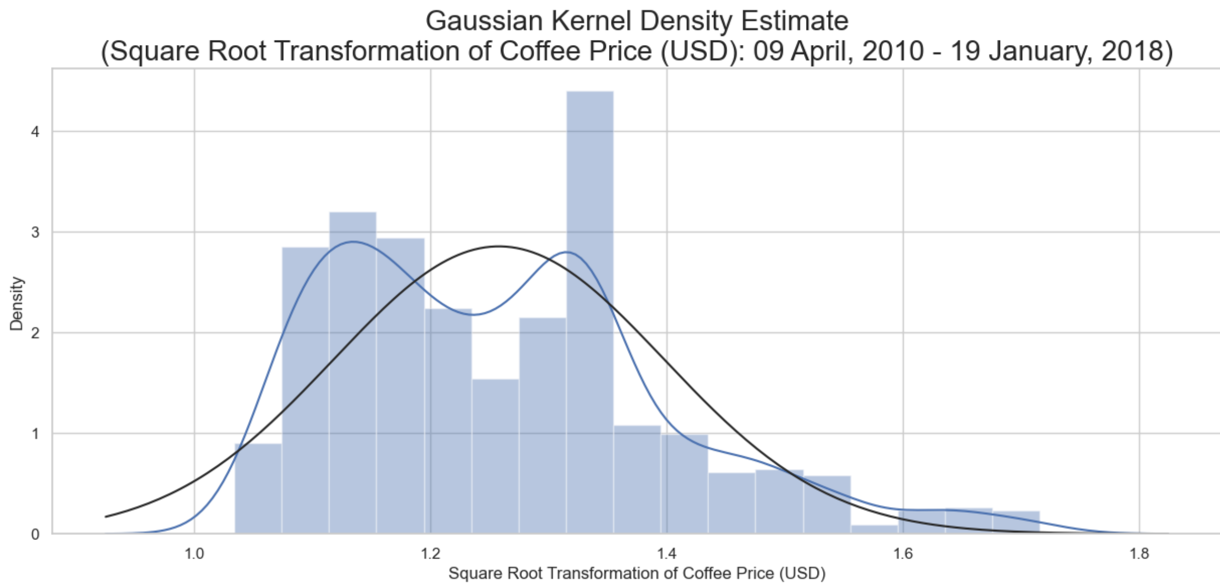
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [98]:

```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



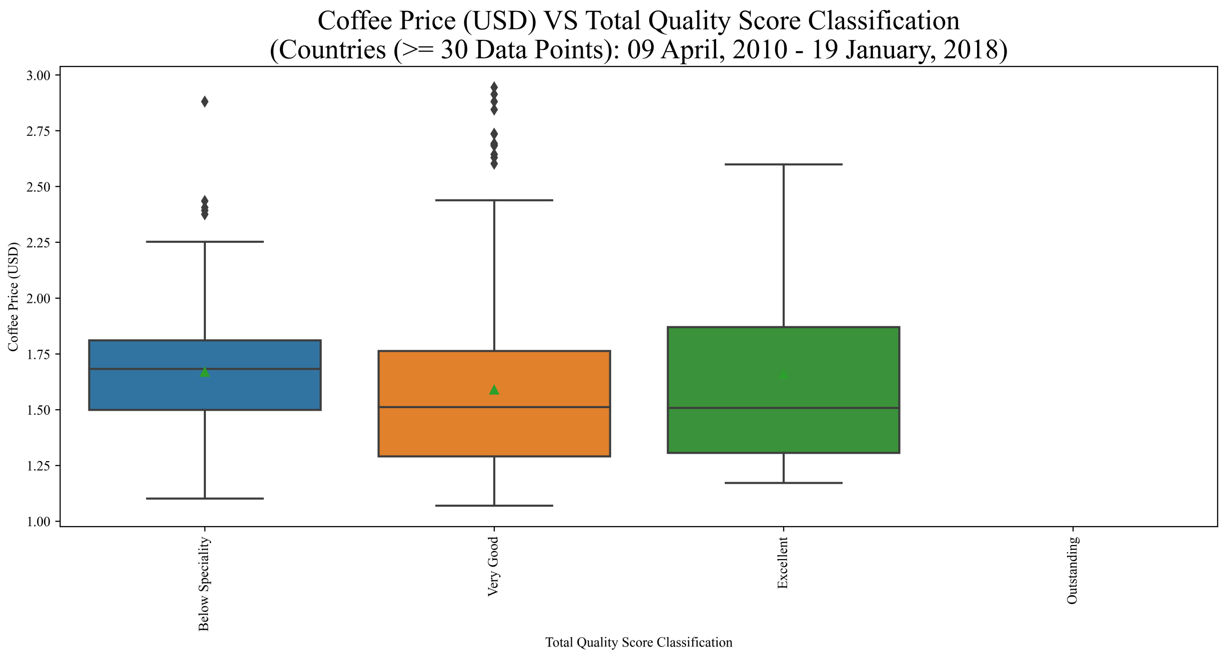
Hence, square root transformed sample is not normally distributed.

Data Visualizations

Boxplot

In [99]:

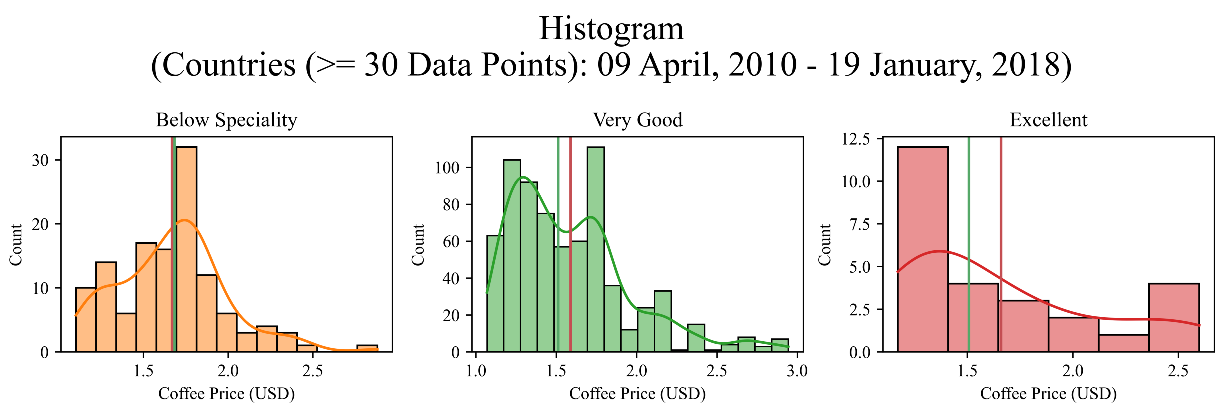
```
# Create a boxplot for countries with >= 30 data points.
# Inputs for cfpq_boxplot func.
df = cfpq_a30_df
x = 'Total Quality Score Classification'
y = 'Coffee Price (USD)'
hue = None
curr_mean_cfp = None
pop_mean_cfp = None
samp_mean_cfp = None

# Run cfpq_boxplot func.
myfcf_gra.cfpq_boxplot(df, x, y, hue, curr_mean_cfp, pop_mean_cfp, samp_mean_
```



Histogram

In [100…

```
# Create a histrogram for countries with >= 30 data points.
# Inputs for ctry_a30_histplot func.
df1 = bs_df
df2 = vg_df
df3 = exc_df
y = 'Coffee Price (USD)'
sub_ttl_lt = ['Below Speciality', 'Very Good', 'Excellent']

# Run ctry_a30_histplot func.
myfcf_gra.ctry_a30_histplot_3df(df1, df2, df3, y, sub_ttl_lt, start_date, end
```



Note: Green line: Median; Red line: Mean

Non-parametric Tests

Perform Kruskal–Wallis H test

To test H_0: $μ_1$ = $μ_2$ = $μ_3$ = $μ_4$ = $μ_5$ = $μ_6$ = $μ_7$ = $μ_8$ = $μ_9$

Against H_1: $μ_1$ ≠ $μ_2$ = $μ_3$ = $μ_4$ = $μ_5$ = $μ_6$ = $μ_7$ = $μ_8$ =

$μ_9$

In [101…

```
# Perform kruskal-wallis test for countries with >= 30 data points.
# Inputs for print_mult_krus_wall_test_3df func.
df1 = bs_df
df2 = vg_df
df3 = exc_df
x = 'Total Quality Score Classification'
y = 'Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_krus_wall_test_3df func.
myfcf_sta.print_mult_krus_wall_test_3df(df1, df2, df3, x, y, alpha_lt, start_
```

Kruskal-Wallis H Test at 5.0% level of significance,
Kruskal-Wallis H Test for Coffee Price (USD) and Total Quality Score Classification (09 April, 2010 - 19 January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Kruskal-Wallis H Test at 2.5% level of significance,
Kruskal-Wallis H Test for Coffee Price (USD) and Total Quality Score Classification (09 April, 2010 - 19 January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Kruskal-Wallis H Test at 1.0% level of significance,
Kruskal-Wallis H Test for Coffee Price (USD) and Total Quality Score Classification (09 April, 2010 - 19 January, 2018)
stat=856.00, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of significance, that coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Hence, based on the findings, there is significant evidence, at 1.0% to 5.0% level of significance, that coffee price (USD) is systematically higher or different in some total quality score classification than in others.

Perform Mann-Whitney U Test

To test H_0: $μ_1$ = $μ_2$

Against H_1: $μ_1$ ≠ $μ_2$

In [102…

```
# Print out mann-whitney u test results at various significant level.
# Inputs for print_mult_mann_whit_test func.
df = cfpq_a30_df
x = 'Total Quality Score Classification'
y = 'Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01] # alpha: level of significance.

# Run print_mult_mann_whit_test func.
myfcf_sta.print_mult_mann_whit_test(df, x, y, alpha_lt, start_date, end_date)
```

Mann-Whitney U Test at 5.0% level of significance,
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Excellent) (09 April, 2010 - 19 January, 2018)
stat=8692.50, p=0.647
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Coffee Price (USD) for Very Good and mean rank of Coffee Price (USD) for Excellent is equal.
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Below Speciality) (09 April, 2010 - 19 January, 2018)
stat=35598.00, p=0.001
Reject Ho and conclude that there is significant evidence, at 5.0% level of significance, that mean rank of Coffee Price (USD) for Very Good and mean rank of Coffee Price (USD) for Below Speciality is not equal.
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Excellent and Below Speciality) (09 April, 2010 - 19 January, 2018)
stat=1451.50, p=0.394
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that mean rank of Coffee Price (USD) for Excellent and mean rank of Coffee Price (USD) for Below Speciality is equal.

Mann-Whitney U Test at 2.5% level of significance,
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Excellent) (09 April, 2010 - 19 January, 2018)
stat=8692.50, p=0.647
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Coffee Price (USD) for Very Good and mean rank of Coffee Price (USD) for Excellent is equal.
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Below Speciality) (09 April, 2010 - 19 January, 2018)
stat=35598.00, p=0.001
Reject Ho and conclude that there is significant evidence, at 2.5% level of significance, that mean rank of Coffee Price (USD) for Very Good and mean rank of Coffee Price (USD) for Below Speciality is not equal.
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Excellent and Below Speciality) (09 April, 2010 - 19 January, 2018)
stat=1451.50, p=0.394
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that mean rank of Coffee Price (USD) for Excellent and mean rank of Coffee Price (USD) for Below Speciality is equal.

Mann-Whitney U Test at 1.0% level of significance,
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Excellent) (09 April, 2010 - 19 January, 2018)
stat=8692.50, p=0.647
Do not reject Ho and conclude that there is no significant evidence, at 1.0%

```
level of significance, that mean rank of Coffee Price (USD) for Very Good and
mean rank of Coffee Price (USD) for Excellent is equal.
```
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Very Good and Below Speciality) (09 April, 2010 - 19 January, 2018)
```
stat=35598.00, p=0.001
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that mean rank of Coffee Price (USD) for Very Good and mean rank
of Coffee Price (USD) for Below Speciality is not equal.
```
Mann-Whitney U Test for Coffee Price (USD) and Total Quality Score Classification (Excellent and Below Speciality) (09 April, 2010 - 19 January, 2018)
```
stat=1451.50, p=0.394
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that mean rank of Coffee Price (USD) for Excellent and
mean rank of Coffee Price (USD) for Below Speciality is equal.
```

Hence, based on the findings, there is significant evidence, at 1.0% to 5.0% level of significance, that mean rank of coffee price for Very Good and mean rank of coffee price for Below Speciality is not equal. Thus, the mean rank of coffee price of Very Good is distinctly lower than the mean rank of Below Speciality.

Project Phase 3

Project Objective

Check whether sample coffee price (USD) mean is equal, larger, or smaller than the population coffee price (USD) mean.

Note: The sample data is based on coffee "C" future price (USD) from 30 April, 2010 to 19 January, 2018, and the population data is based on coffee "C" future price (USD) from 20 August, 1973 to 13 October, 2021.

Problem Solving Steps

Assume the coffee price follows the coffee "C" future price.

1. Ensure that the population of coffee future price is normally distributed.
    - If the population of coffee future price is not normally distributed, perform box-cox transformation on population. To create a new population, select a start date from one of the earliest date with the highest p-value (p-value > 0.05). p-value is obtained using D'Agostino-Pearson test.
2. Perform 1 sample Z test on the box-cox transformed versions of the previous sample of coffee future price and new population to test whether the sample mean is equal, greater, or smaller than the new population mean.
3. Perform inverse box-cox transform to find the new population mean and sample mean.

Note: The sample data is based on coffee "C" future price (USD) from 30 April, 2010 to 19 January, 2018, and the population data is based on coffee "C" future price (USD) from 20 August, 1973 to 13 October, 2021.

Criteria for 1 Sample Z-Test

- The population from which the sample is drawn is normally distributed.
- The sample size is greater than 30.
- A single sample is drawn.
- We are testing for population mean.
- The population standard deviation is know.

Normality Check

Overview

In [103…

```
# Inputs for normality check.
df = cfp_df
col = 'Coffee Price (USD)'
```

In [104…

```
# Find the start and end dates of the cfp_df.
# Inputs for start_end_dates_tp func.
df = cfp_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)

# Print start and end dates of cfq_df.
print('Start Date: ' + start_date)
print('End Date: ' + end_date)
```

```
Start Date: 1973-08-20
End Date: 2021-10-13
```

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [105…

```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (20 August, 1973 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
```

```
Shapiro–Wilk Test for Coffee Price (USD) (20 August, 1973 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (20 August, 1973 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [106…

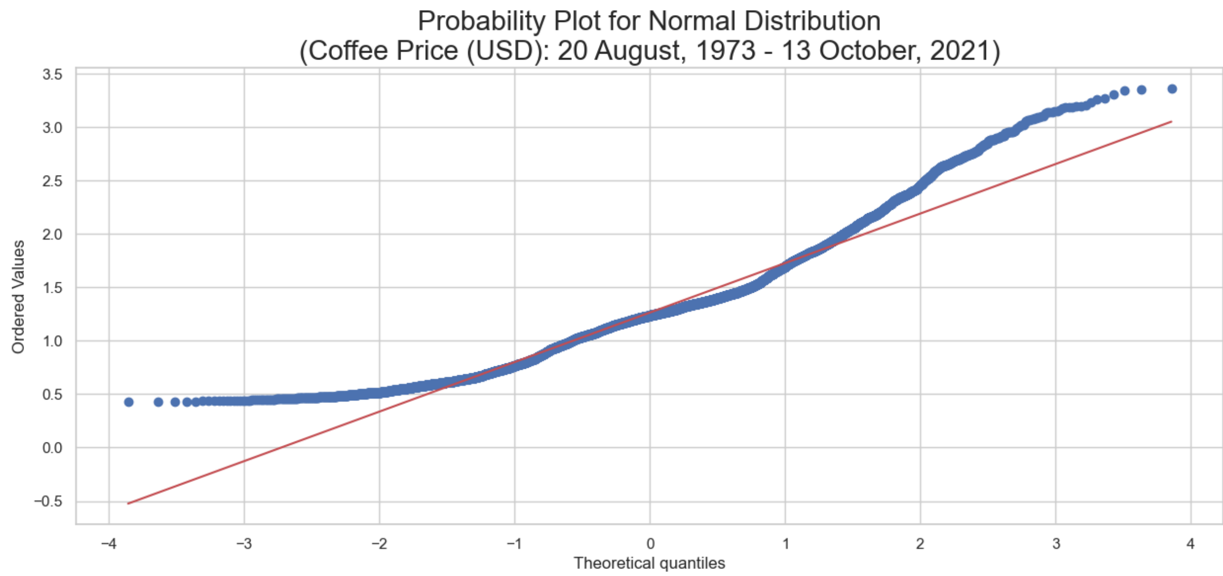
```python
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino–Pearson Test at 5.0% level of significance,
D'Agostino–Pearson Test for Coffee Price (USD) (20 August, 1973 – 13 October,
2021)
stat=1410.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


D'Agostino–Pearson Test at 2.5% level of significance,
D'Agostino–Pearson Test for Coffee Price (USD) (20 August, 1973 – 13 October,
2021)
stat=1410.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino–Pearson Test at 1.0% level of significance,
D'Agostino–Pearson Test for Coffee Price (USD) (20 August, 1973 – 13 October,
2021)
stat=1410.66, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [107…

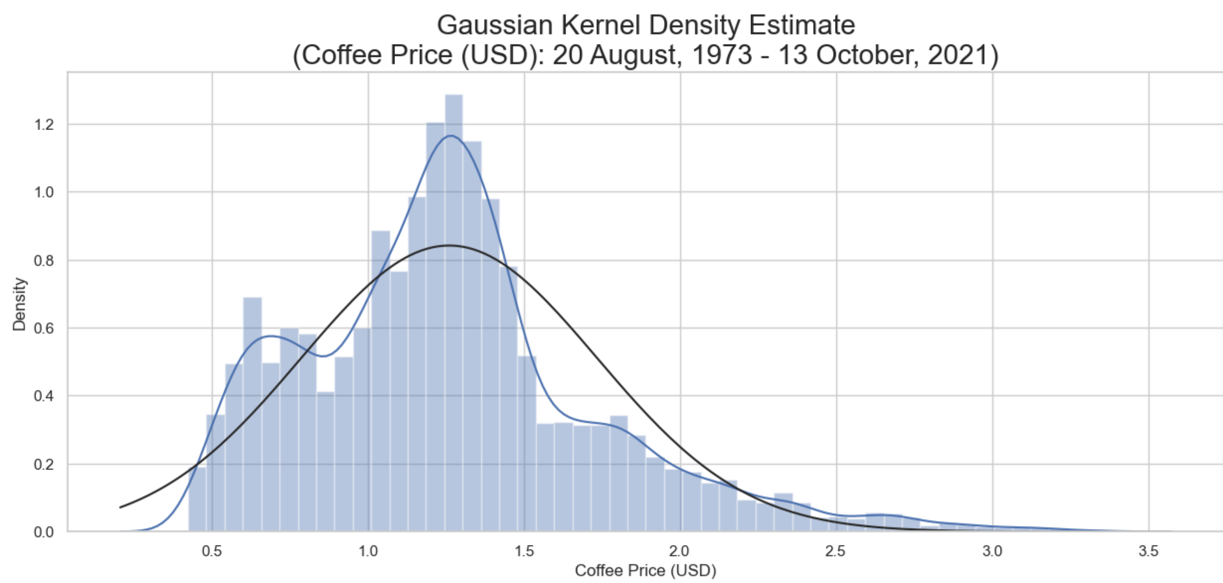
```python
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [108…

```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, population is not normally distributed.

Box-Cox Transformation

To ensure the chosen population is close to normal distribution, elect a start date from one of the earliest date with the highest p-value (p-value > 0.05). p-value is obtained using D'Agostino-Pearson test.

Before Box-Cox Transformation

Overview

In [109…

```
# Create a dataframe for box-cox transformation.
# Inputs for box_cox_df func.
df = cfp_df
col = 'Coffee Price (USD)'

# Run box_cox_df func.
box_cox_df = myfcf_sta.box_cox_df(df, col)

# Display head of box_cox_df.
box_cox_df.head()
```

Out[109…

	Start Date	p-value	Lambda	Box-Cox	Skew
0	1973-08-20	0.151825	0.251609	[-0.37624777437276524, -0.37961300427646727, -...	-0.0
1	1973-08-21	0.155173	0.251729	[-0.37960406327274243, -0.39725526463400107, -...	-0.0
2	1973-08-22	0.158594	0.251847	[-0.397245618533816, -0.38432193895272426, -0....	-0.0
3	1973-08-23	0.162162	0.251955	[-0.38431372271507014, -0.3863450271239629, -0...	-0.0
4	1973-08-24	0.165721	0.252070	[-0.3863361224115001, -0.39586059880214997, -0...	-0.0

In [110…

```
# Filter for the top five p-values based on D'Agostino's K-squared test.
box_cox_df[box_cox_df['p-value'] > 0.05].sort_values(by=['p-value', 'Start Da
```

Out[110…

	Start Date	p-value	Lambda	Box-Cox	Skew
3114	1986-11-02	0.997134	0.164045	[0.9206148046864553, 0.9063795039685426, 1.017...	-0.0
92	1974-07-01	0.996280	0.274006	[-0.3426534987070336, -0.35020777748541937, -0...	-0.0
4618	1992-10-02	0.995401	0.255766	[-0.3474271948743407, -0.3330904335039328, -0....	-0.0
3092	1986-10-01	0.995397	0.165669	[1.0686870409921603, 0.9102239122810306, 0.921...	-0.0
3112	1986-07-02	0.995397	0.165669	[1.0686870409921603, 0.9102239122810306, 0.921...	-0.0

In [111…

```
# Select index 92 to be the start date of coffee price (usd) population.
box_cox_strt_dt = box_cox_df.iloc[92]['Start Date']

# Display box_cox_strt_dt.
box_cox_strt_dt
```

Out[111…

```
'1974-07-01'
```

The start date of the original Coffee Price (USD) population is August 20, 1973. However, July 1, 1974 is used instead, because they are just one year apart.

In [112…

```
# Create new_cf_price_df using the start date related to the chosen p-value.
# Inputs for new_cf_price_df func.
df = cfp_df
strt_dt = box_cox_strt_dt

# Run new_cfp_df func.
new_cfp_df = myfcf_sta.new_cfp_df(df, strt_dt)

# Display head of new_cfp_df.
new_cfp_df.head()
```

Out[112…

	Date	Coffee Price (USD)
92	1974-07-01	0.6978
93	1974-08-01	0.6920
94	1974-09-01	0.6903
95	1974-10-01	0.6958
96	1974-11-01	0.7085

In [113…

```
# Find the start and end dates of the new_cf_price_df.
# Inputs for start_end_dates_tp func.
df = new_cfp_df
col = 'Coffee Price (USD)'

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)
```

Normality Test

Shapiro–Wilk Test

In [114…

```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.
```

```
Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2021)
stat=0.95, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [115…

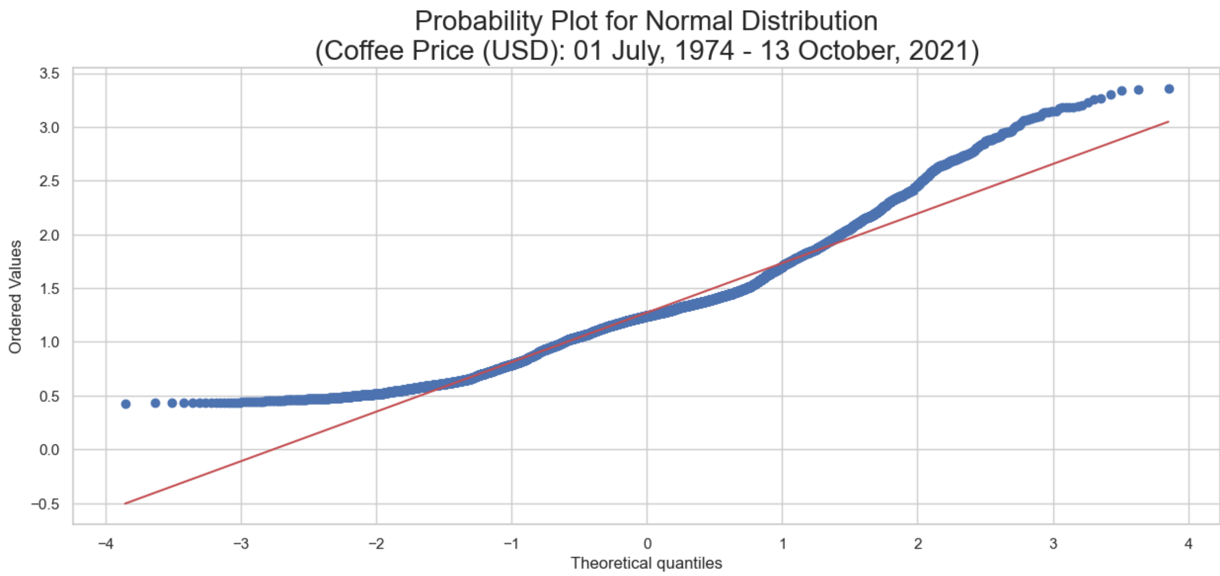
```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

```
D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2
021)
stat=1387.30, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2
021)
stat=1387.30, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Coffee Price (USD) (01 July, 1974 – 13 October, 2
021)
stat=1387.30, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

Probability Plot

In [116…

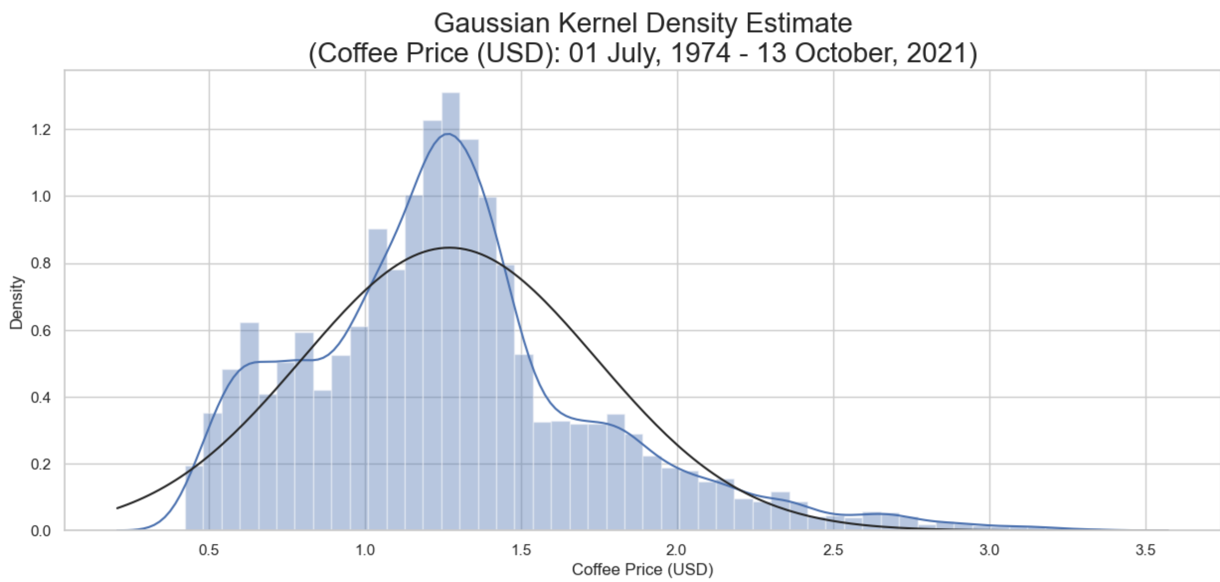
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [117…

```
# Check the distribution based on skewness.
myfcf_sta.kde_norm_distplot(df, col, start_date, end_date)
```



Hence, population is not normally distributed.

After Box-Cox Transformation

Overview

In [118…

```
# Inputs for normality check.
df = new_cfp_df
```

In [119…

```
# Select index 92 for the box-transformed coffee price (usd) population.
box_cox_arr = box_cox_df.iloc[92]['Box-Cox']

# Create a column 'Box-Cox Transformation of Coffee Price (USD)' within new_c
new_cfp_df['Box-Cox Transformation of Coffee Price (USD)'] = box_cox_arr

# Display head of new_cfp_df.
new_cfp_df.head()
```

Out[119…

	Date	Coffee Price (USD)	Box-Cox Transformation of Coffee Price (USD)
92	1974-07-01	0.6978	-0.342653
93	1974-08-01	0.6920	-0.350208
94	1974-09-01	0.6903	-0.352431
95	1974-10-01	0.6958	-0.345253
96	1974-11-01	0.7085	-0.328836

Normality Test

To test H_0: Distribution is normal

Against H_1: Distribution is not normal

Shapiro–Wilk Test

In [120…

```
# Check if p-value is greater than the alpha level.
# Input alpha for print_mult_shap_wilk_test func.
col = 'Box-Cox Transformation of Coffee Price (USD)'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_shap_wilk_test func.
myfcf_sta.print_mult_shap_wilk_test(df, col, alpha_lt, start_date, end_date)
```

```
Shapiro–Wilk Test at 5.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (01 July,
1974 – 13 October, 2021)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 2.5% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (01 July,
1974 – 13 October, 2021)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that the population is not normally-distributed.


Shapiro–Wilk Test at 1.0% level of significance,
Shapiro–Wilk Test for Box-Cox Transformation of Coffee Price (USD) (01 July,
1974 – 13 October, 2021)
stat=0.99, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that the population is not normally-distributed.
```

D'Agostino-Pearson Test

In [121…

```
# Run print_mult_normal_test func.
myfcf_sta.print_mult_normal_test(df, col, alpha_lt, start_date, end_date)
```

D'Agostino-Pearson Test at 5.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 1974 - 13 October, 2021)
stat=0.01, p=0.996
Do not reject Ho and conclude that there is no significant evidence, at 5.0% level of significance, that the population is normally-distributed.

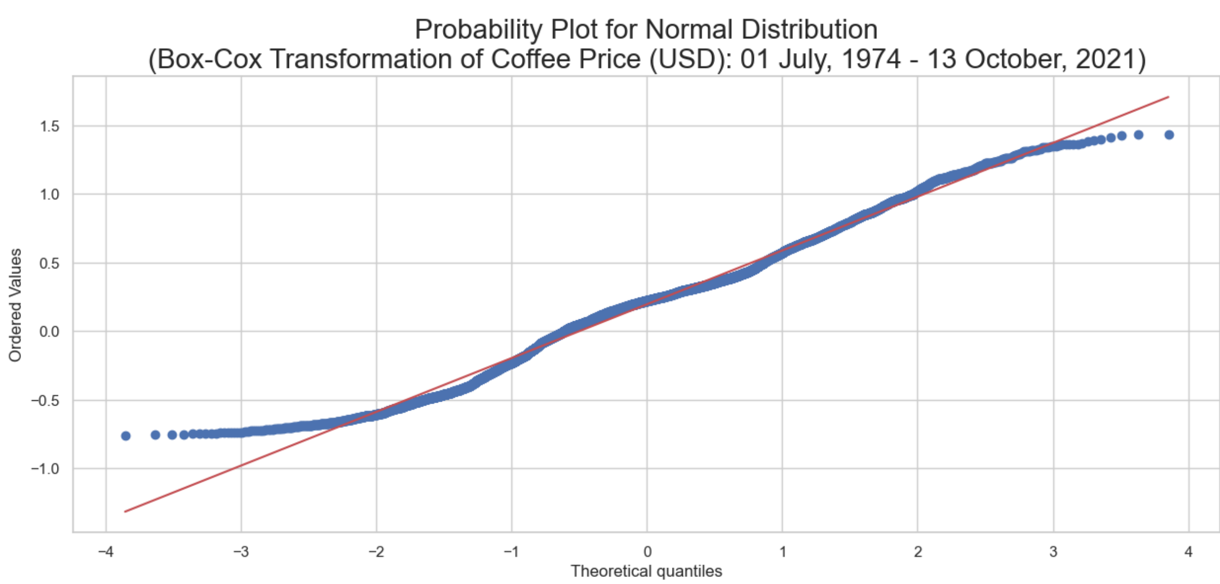
D'Agostino-Pearson Test at 2.5% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 1974 - 13 October, 2021)
stat=0.01, p=0.996
Do not reject Ho and conclude that there is no significant evidence, at 2.5% level of significance, that the population is normally-distributed.

D'Agostino-Pearson Test at 1.0% level of significance,
D'Agostino-Pearson Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 1974 - 13 October, 2021)
stat=0.01, p=0.996
Do not reject Ho and conclude that there is no significant evidence, at 1.0% level of significance, that the population is normally-distributed.

Probability Plot

In [122…

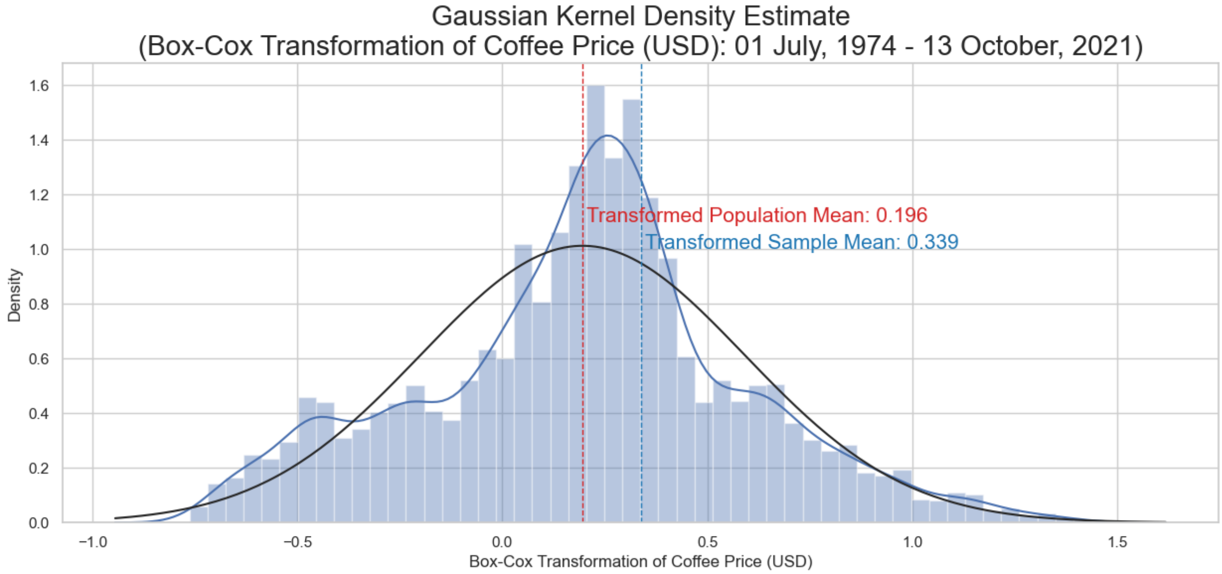
```
# Check if data points lay on the red color straight line.
myfcf_sta.prob_plot(df, col, start_date, end_date)
```



Histogram

In [123…

```
# Inputs for trans_hist func.
pop_df = new_cfp_df
samp_df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Coffee Price (USD)'

# Run trans_hist func.
myfcf_sta.trans_hist(pop_df, samp_df, col, trans_cat, start_date, end_date)
```



Hence, box-cox transformed Coffee Price (USD) population from July 1, 1974 onwards is normally distributed.

Parametric Testing

Perform 1 Sample Z-Test

In [124…

```
# Inputs for box_cox_pop_samp_mean_tp func.
pop_df = new_cfp_df
samp_df = new_cfpq_a30_df
col = 'Coffee Price (USD)'
trans_cat = 'Box-Cox'

# Run box_cox_pop_samp_mean_tp func.
box_cox_pop_mean, box_cox_samp_mean = myfcf_sta.box_cox_pop_samp_mean_tp(pop_
```

Two-Tail Z Test

To test H_0: $μ$ = $μ_0$

Against H_1: $μ$ ≠ $μ_0$

In [125…

```
# Inputs for print_mult_one_sample_ztest func.
samp_df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Coffee Price (USD)'
pop_mean = box_cox_pop_mean
sign = 'two-sided'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_one_sample_ztest func.
myfcf_sta.print_mult_one_sample_ztest(samp_df, col, pop_mean, sign, alpha_lt,
```

```
1-sample Z Test at 5.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that transformed sample mean is not equal to the transformed pop
ulation mean.


1-sample Z Test at 2.5% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that transformed sample mean is not equal to the transformed pop
ulation mean.


1-sample Z Test at 1.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that transformed sample mean is not equal to the transformed pop
ulation mean.
```

One-Tail Z Test

To test H_0: $μ$ = $μ_0$

Against H_1: $μ$ > $μ_0$

In [126…

```
# Inputs for print_mult_one_sample_ztest func.
samp_df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Coffee Price (USD)'
pop_mean = box_cox_pop_mean
sign = 'larger'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_one_sample_ztest func.
myfcf_sta.print_mult_one_sample_ztest(samp_df, col, pop_mean, sign, alpha_lt,
```

```
1-sample Z Test at 5.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
```

```
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 5.0% level of s
ignificance, that transformed sample mean is larger than the transformed popu
lation mean.


1-sample Z Test at 2.5% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 2.5% level of s
ignificance, that transformed sample mean is larger than the transformed popu
lation mean.


1-sample Z Test at 1.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=0.000
Reject Ho and conclude that there is significant evidence, at 1.0% level of s
ignificance, that transformed sample mean is larger than the transformed popu
lation mean.
```

To test H_0: $μ$ = $μ_0$

Against H_1: $μ$ < $μ_0$

In [127…

```
# Inputs for print_mult_one_sample_ztest func.
samp_df = new_cfpq_a30_df
col = 'Box-Cox Transformation of Coffee Price (USD)'
pop_mean = box_cox_pop_mean
sign = 'smaller'
alpha_lt = [0.05, 0.025, 0.01]

# Run print_mult_one_sample_ztest func.
myfcf_sta.print_mult_one_sample_ztest(samp_df, col, pop_mean, sign, alpha_lt,
```

```
1-sample Z Test at 5.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=1.000
Do not reject Ho and conclude that there is no significant evidence, at 5.0%
level of significance, that transformed sample mean is smaller than the trans
formed population mean.


1-sample Z Test at 2.5% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
stat=21.69, p=1.000
Do not reject Ho and conclude that there is no significant evidence, at 2.5%
level of significance, that transformed sample mean is smaller than the trans
formed population mean.


1-sample Z Test at 1.0% level of significance,
1-Sample Z-Test for Box-Cox Transformation of Coffee Price (USD) (01 July, 19
74 - 13 October, 2021)
```

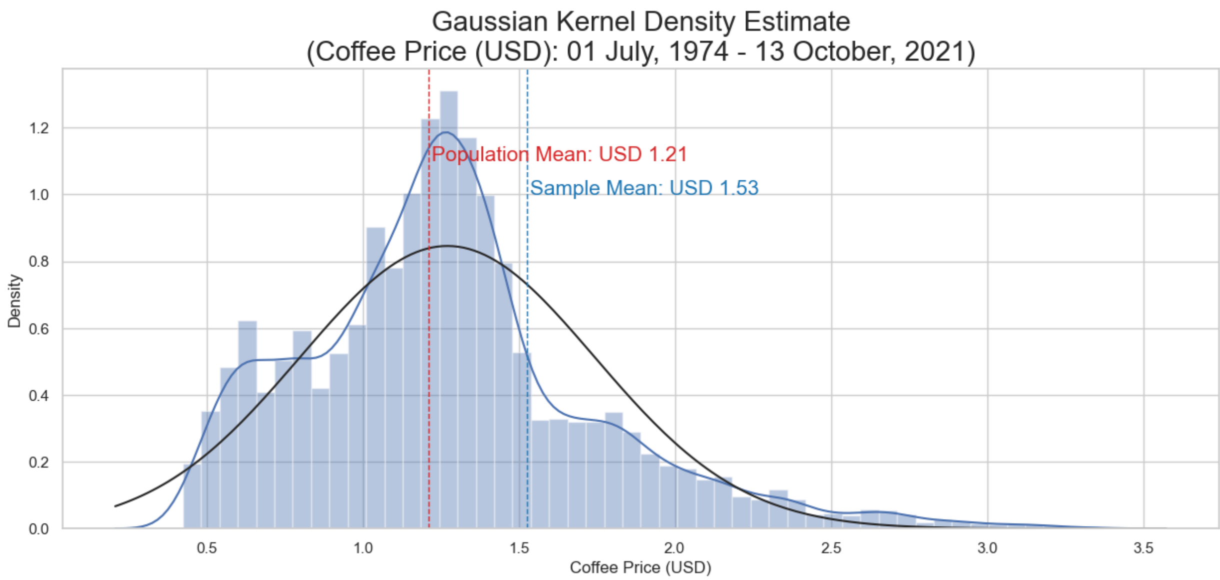
```
stat=21.69, p=1.000
Do not reject Ho and conclude that there is no significant evidence, at 1.0%
level of significance, that transformed sample mean is smaller than the trans
formed population mean.
```

Hence, based on the findings, there is significant evidence, at 1.0% to 5.0% level of significance, that the transformed sample mean is larger than the transformed population mean.

Inverse Box-Cox Transformation

In [128…

```
# Inputs for trans_hist func.
pop_df = new_cfp_df
samp_df = new_cfpq_a30_df
col = 'Coffee Price (USD)'

# Run trans_hist func.
myfcf_sta.trans_hist(pop_df, samp_df, col, trans_cat, start_date, end_date)
```



In [129…

```
# Inputs for print_mean_pop_samp func.
pop_df = new_cfp_df
samp_df = new_cfpq_a30_df

# Run print_mean_pop_samp func.
myfcf_sta.print_mean_pop_samp(pop_df, samp_df)
```

```
Box-Cox Transformation:
Population mean of box-cox transformed coffee price (USD): 0.196
Sample mean of box-cox transformed coffee price (USD): 0.339


After Inverse Box-Cox Transformation:
Population mean of coffee price: USD 1.21
Sample mean of coffee price: USD 1.53
```

Hence, based on the findings, there is significant evidence, at 1.0% to 5.0% level of significance, that the sample mean of coffee price is larger than the population mean of coffee price based on July 1, 1974 to October 13, 2021.

Project Phase 4

Project Objective

Evaluate the feasibility of selling premium coffee in the current environment.

Note: The population data is based on coffee future "C" price (USD) from July 1, 1974 to 13 October, 2021.

Problem Solving Steps

Assume the coffee price follows the coffee "C" future price.

1. Create boxplots of coffee future price to check where the current coffee future price lies.

Coffee Price (USD)

In [130…

```
# Find sample mean of coffee price and sample mean of transformed coffee pric
samp_mean_cfp, samp_mean_trans_cfp = myfcf_sta.mean_cfp_trans_tp(samp_df)

# Find population mean of coffee price and population mean of transformed cof
pop_mean_cfp, pop_mean_trans_cfp = myfcf_sta.mean_cfp_trans_tp(pop_df)

# Find current coffee price.
curr_cfp = new_cfp_df['Coffee Price (USD)'].tail(1).values[0]
```

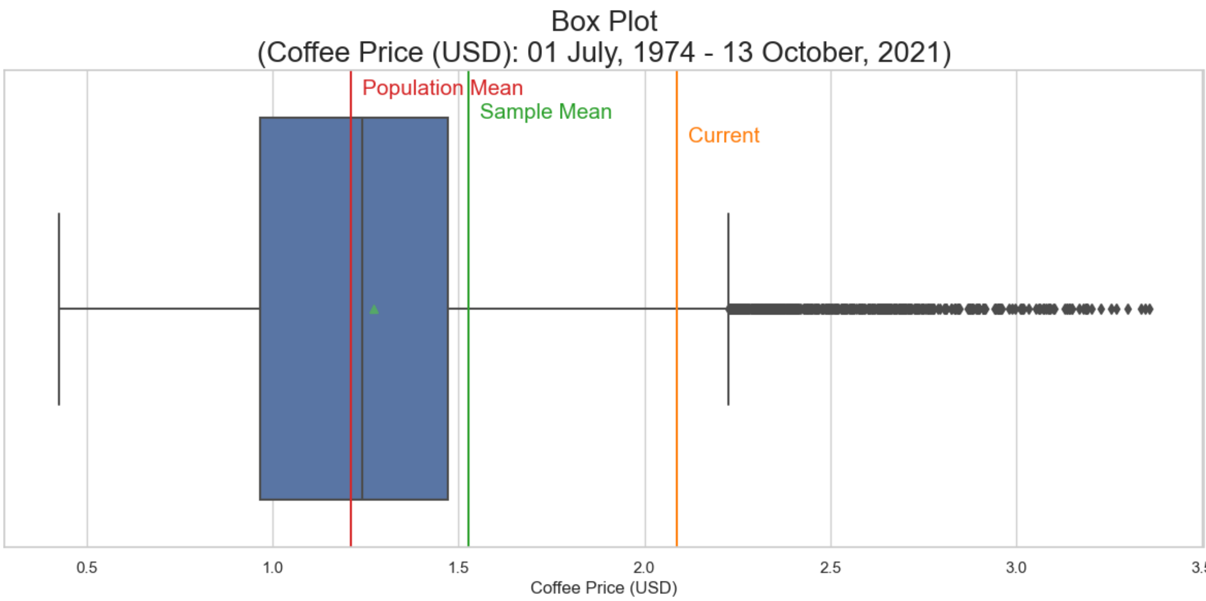
In [131…

```
# Find the start and end dates of the new_cfp_df.
# Inputs for start_end_dates_tp func.
df = new_cfp_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)
```

Data Visualizations

In [132…

```
# Inputs for cfp_h_boxplot func.
df = new_cfp_df
col = 'Coffee Price (USD)'

# Run cfp_h_boxplot func.
myfcf_gra.cfp_h_boxplot(df, col, curr_cfp, pop_mean_cfp, samp_mean_cfp, start
```



Data Analysis

In [133…

```
# Inputs for print_pct_abv_bel_cfp func.
df = new_cfp_df
cfp = curr_cfp
cat_cfp = 'Current Coffee Price'
ttl_txt = 'Coffee price from 01 July, 1974 to 13 October, 2021'

# Run print_pct_abv_bel_cfp func.
myfcf_sta.print_pct_abv_bel_cfp(df, cfp, cat_cfp, ttl_txt)
```

```
Coffee price from 01 July, 1974 to 13 October, 2021
6.23% of the data points are above USD2.09 (Current Coffee Price)
93.77% of the data points are above USD2.09 (Current Coffee Price)
```

Coffee Price (USD) VS Total Quality Score Classification

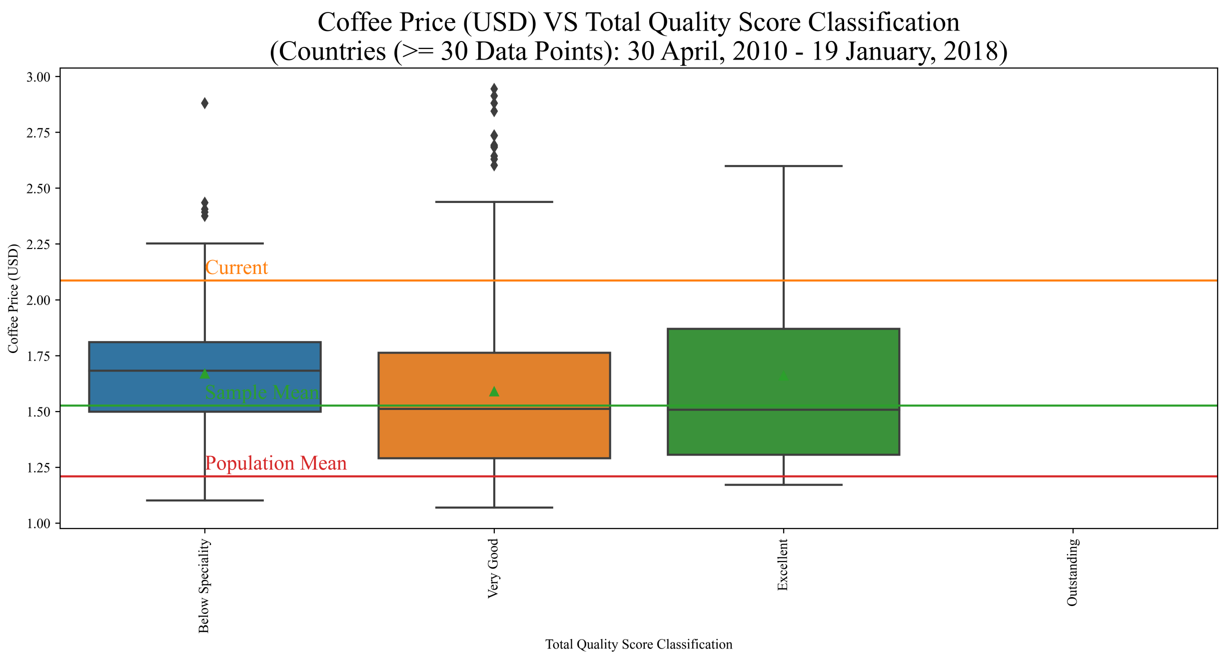
In [134…

```
# Find the start and end dates of the new_cfpq_a30_df.
# Inputs for start_end_dates_tp func.
df = new_cfpq_a30_df

# Run start_end_dates_tp func.
start_date, end_date = myfcf_wra.start_end_dates_tp(df)
```

Data Visualizations

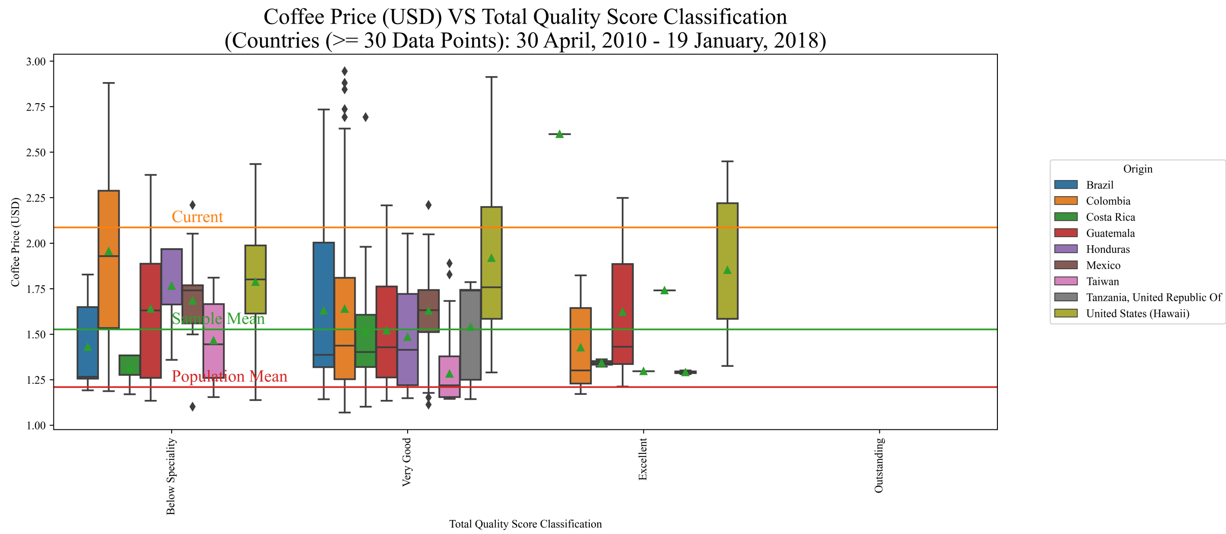
In [135…

```
# Create a boxplot for countries with >= 30 data points.
# Inputs for cfpq_boxplot func.
df = new_cfpq_a30_df
x = 'Total Quality Score Classification'
y = 'Coffee Price (USD)'
hue = None
curr_cfp = new_cfp_df['Coffee Price (USD)'].tail(1).values[0]

# Run cfpq_boxplot func.
myfcf_gra.cfpq_boxplot(df, x, y, hue, curr_cfp, pop_mean_cfp, samp_mean_cfp,
```



In [136…

```
# Create a boxplot for countries with >= 30 data points.
# Inputs for cfpq_boxplot func.
df = new_cfpq_a30_df
x = 'Total Quality Score Classification'
y = 'Coffee Price (USD)'
hue = 'Origin'
curr_cfp = new_cfp_df['Coffee Price (USD)'].tail(1).values[0]

# Run cfpq_boxplot func.
myfcf_gra.cfpq_boxplot(df, x, y, hue, curr_cfp, pop_mean_cfp, samp_mean_cfp,
```



Feasibility of Purchasing Coffee Beans

The current coffee price is extremely high, because based on the sample data, only 6.23% of the historical data points are above its current price, which is way above the prices of sample mean and population mean. Hence, it is not a wise decision to purchase premium coffee beans at this point in time.

If the coffee "C" future price moves below the sample mean of coffee price and towards USD 1.21, Columbian coffee with quality of Very Good and Excellent can be considered good purchases despite being sold at 400 per point premium based on its origin along with the additional premium based on its grade.

Findings

Based on our findings with Ethiopia excluded (not much coffee future contracts are done in Ethiopia), the median of total quality score among the countries are not similar (apart from Ethopian coffee, Colombian coffee is the leader in quality coffee), and the median coffee price among the total quality score are not similar (the mean rank of Very Good coffee is lower than the mean rank of Below Speciality coffee), and the ideal coffee purchase price is USD 1.21.

Conclusion

From The Management Lab’s perspective, strong global coffee demand and a deficit in global coffee stockpile are the main drivers driving coffee future price to its current high price.

Currently, it is not the best time to venture into the business of opening coffeehouses that sells premium coffee, because of the high business risk involved due to the ongoing Covid-19 situation, as customers might turn to inferior goods when they loss their jobs, or when their disposable income gets shrunk by rising inflation, or being caught in a business with shrinking profit margin due to rising cost.

Limitations

The accuracy of our statistical models depends whether the coffee price distribution is within USD 0.43 to USD 3.36 and total quality score distribution is within 70.67 to 87.17, and to be deemed inaccurate if there are any deviation from these distributions.

Recommendations

Our client should relook into the business of opening coffeehouses that sells premium coffee, if strong global coffee demand remains and that the current coffee future prices are somewhat closer to USD 1.21. The premium coffee that they should consider purchasing for the coffeehouses should be Columbian coffee with quality of Very Good.

Owing to the deficit in global coffee stockpile, our client should consider investing in Columbian coffee with quality of Very Good, as coffee prices are forecasted to go higher.