

Data Mining (DSC550-T301_2245_1)

Assignment Week 6: Term Project Milestone 1;

Author: Zemelak Goraga;

Date: 04/20/2024

Topic of Term Project: Trend analysis on Meat Animals Export Marketing of Different Countries.

Milestone 1: Data Exploration and Problem Statement

Introduction: The export of meat animals is a significant aspect of international trade, impacting economies and food security globally. Analyzing trends in meat animal exports across different countries can provide valuable insights for stakeholders in the agricultural and trade sectors. As a data science problem, this analysis involves extracting meaningful patterns and insights from historical export data to understand market dynamics and inform decision-making processes.

Original Idea:

The original idea for this project emerged from Google Trends (<https://trends.google.com/trends/explore?cat=752&date=all&q=Export,Sheep,cattle,pig,chicken&hl=en-US>). To gather valuable insights into the evolving global export trade patterns of meat animals, I employed Google Trends as a supplementary data source. I carefully selected a set of pertinent search terms and phrases such as Export, Sheep, Cattle, Pig, and Chicken. This preliminary research in Google Trend Analysis, helped me to have valuable insights into the global export trade patterns of meat animals such as cattle, sheep, pigs and chicken.

In my term project, I would like to support the initial insights obtained from Google Trend analysis using appropriate datasets obtained from Kaggle. The business problem I aimed to address in my term project will help to optimize meat animal export strategies for different countries. The target for my model are to forecast future trends in meat animal exports based on historical data, identify factors influencing export fluctuations, and provide actionable recommendations for stakeholders.

Research Questions:

Descriptive: What are the historical trends in meat animal exports for different countries?

Diagnostic: What factors contribute to fluctuations in meat animal export quantities and values?

Predictive: Can we forecast future trends in meat animal exports based on historical data?

Prescriptive: How can countries optimize their meat animal export strategies to maximize profitability and market share?

Approach to Problem Statement: The problem will be addressed by conducting exploratory data analysis, identifying key drivers of export trends, developing predictive models, and providing actionable recommendations for stakeholders.

Solution Approach: Exploratory Data Analysis: Analyze historical export data to identify trends and patterns. Statistical Modeling: Develop predictive models using time series analysis and regression techniques. Prescriptive Analysis: Generate recommendations based on insights from exploratory and predictive analyses.

Dataset Used: The dataset used for this study is the 'meat_animals_export' dataset extracted from the FAOSTAT historical dataset. This dataset is available on Kaggle, sourced from the United Nations' global food and agriculture statistics.

Data Source Explanation: The original purpose of the FAOSTAT dataset is to provide comprehensive statistics on global food and agriculture production, trade, and consumption. It was collected between 1961 to 2013 and contains over 25 primary products and inputs across 200 countries. The dataset contains variables such as Country, Item, Element, Year, Unit, and Value. Missing values are handled through data cleaning processes, and peculiarities such as inconsistent data formats are addressed during preprocessing.

To achieve the objectives of this project, most of my data analysis will be done targeting the top 10 exporting countries in the past 15 years.

Required Python Packages:

pandas matplotlib seaborn statsmodels

Visualizations and Tables:

Time series plots to illustrate export trends over time. Different charts to compare export quantities and values across countries. Correlation matrices to identify relationships between variables. Regression analysis tables to assess the impact of predictors on export outcomes.

Summary of Research Methods: The analysis aims to explore trends in meat animal exports across various countries, focusing on historical data from the FAOSTAT dataset. By addressing data cleanliness, renaming columns, and defining the problem statement, the foundation for further analysis is established. The research questions span descriptive, diagnostic, predictive, and prescriptive analyses, intending to provide a comprehensive understanding of export dynamics. Python packages such as pandas, matplotlib, seaborn, and statsmodels will facilitate data manipulation, visualization, and modeling tasks.

Visualizations and tables will be created to illustrate export trends, identify key drivers, and assess predictive models' performance. Through this analysis, insights will be gained to inform strategic decisions and optimize meat animal export strategies.

Data Wrangling, Descriptive Statistics, and Visualizations

In [2]: # Importing the dataset:

```
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

```
# Importing the dataset
df = pd.read_csv('live_animals_export.csv') # this 'live_animals_export.csv' dataset
```

In [3]: # Inspecting the dataset:

```
print(df.head())
```

	area	item	element	year	unit	value
0	Argentina	Cattle	Export Quantity	1961	Head	171106
1	Argentina	Cattle	Export Quantity	1962	Head	250274
2	Argentina	Cattle	Export Quantity	1963	Head	291819
3	Argentina	Cattle	Export Quantity	1964	Head	166050
4	Argentina	Cattle	Export Quantity	1965	Head	102993

In [4]: # Inspecting the dataset:

```
print(df.tail())
```

	area	item	element	year	\
9329	Net Food Importing Developing Countries	Sheep	Export Value	2009	
9330	Net Food Importing Developing Countries	Sheep	Export Value	2010	
9331	Net Food Importing Developing Countries	Sheep	Export Value	2011	
9332	Net Food Importing Developing Countries	Sheep	Export Value	2012	
9333	Net Food Importing Developing Countries	Sheep	Export Value	2013	

	unit	value
9329	1000 US\$	369527
9330	1000 US\$	321993
9331	1000 US\$	495704
9332	1000 US\$	619326
9333	1000 US\$	845143

In [5]: # Inspecting the dataset:

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9334 entries, 0 to 9333
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype  
 ---  -- 
 0   area      9334 non-null   object 
 1   item       9334 non-null   object 
 2   element    9334 non-null   object 
 3   year       9334 non-null   int64  
 4   unit       9334 non-null   object 
 5   value      9334 non-null   int64  
dtypes: int64(2), object(4)
memory usage: 437.7+ KB
None
```

In []:

In [6]: # Renaming columns:

```
df.rename(columns={'area': 'country', 'item': 'animal_type'}, inplace=True)
df.head()
```

```
Out[6]:   country animal_type      element  year  unit  value
0 Argentina     Cattle Export Quantity  1961 Head 171106
1 Argentina     Cattle Export Quantity  1962 Head 250274
2 Argentina     Cattle Export Quantity  1963 Head 291819
3 Argentina     Cattle Export Quantity  1964 Head 166050
4 Argentina     Cattle Export Quantity  1965 Head 102993
```

```
In [7]: # Data wrangling:
# Handling missing values
df.dropna(inplace=True)
```

```
In [8]: # Data wrangling:
# Handling duplicate rows
df.drop_duplicates(inplace=True)
```

```
In [9]: # Data wrangling:
# Handling inconsistent values
df['animal_type'] = df['animal_type'].str.lower()
```

```
In [10]: # Descriptive statistics of export quantity - considering the whole dataset
import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_quantity_df = df[df['element'] == 'Export Quantity']

# Descriptive Statistics
print("Descriptive Statistics for Export Quantity:")
print(export_quantity_df['value'].describe())
```

```
Descriptive Statistics for Export Quantity:
count    6.346000e+03
mean     1.521507e+06
std      2.883626e+06
min      1.000000e+05
25%     2.122202e+05
50%     5.413090e+05
75%     1.484148e+06
max     3.857734e+07
Name: value, dtype: float64
```

```
In [11]: # Time series Line plot showing trend of Export Qunatity (heads) of Live animals in
# the last 15 years
import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

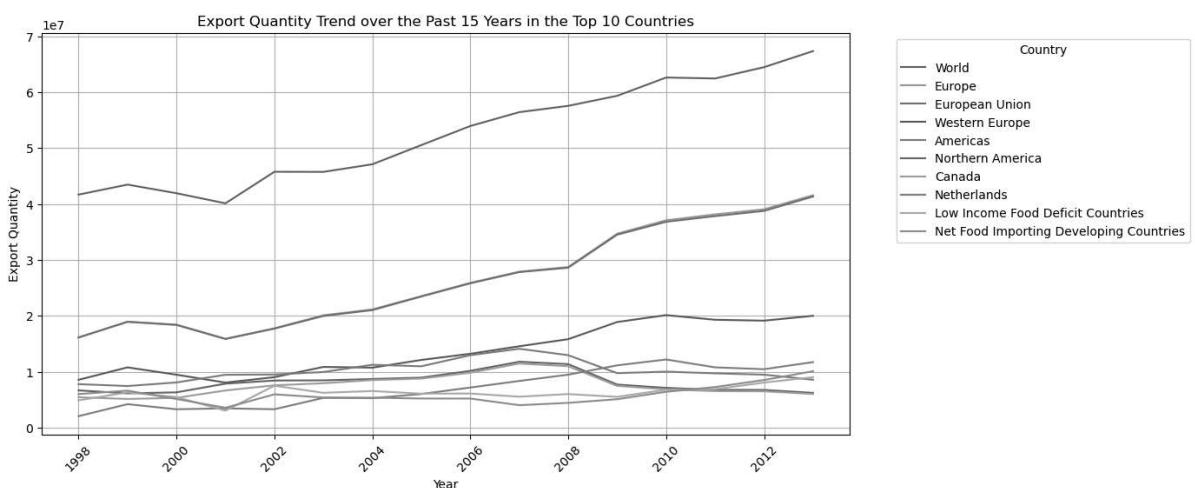
# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]
```

```

# Group by country and sum the export quantity for each year
top_countries = past_15_years_data.groupby(['country', 'year'])['value'].sum().unstack()
top_countries_total = top_countries.sum().nlargest(10).index
top_countries_data = top_countries[top_countries_total]

# Line plot for export quantity trend over the past 15 years in the top 10 countries
plt.figure(figsize=(12, 6))
for country in top_countries_data.columns:
    sns.lineplot(x=top_countries_data.index, y=top_countries_data[country], label=country)
plt.title("Export Quantity Trend over the Past 15 Years in the Top 10 Countries")
plt.xlabel("Year")
plt.ylabel("Export Quantity")
plt.xticks(rotation=45)
plt.legend(title='Country', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()

```



In [12]: # Line plot showing mean export quantity (heads) of meat animals in top 10 countries

```

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

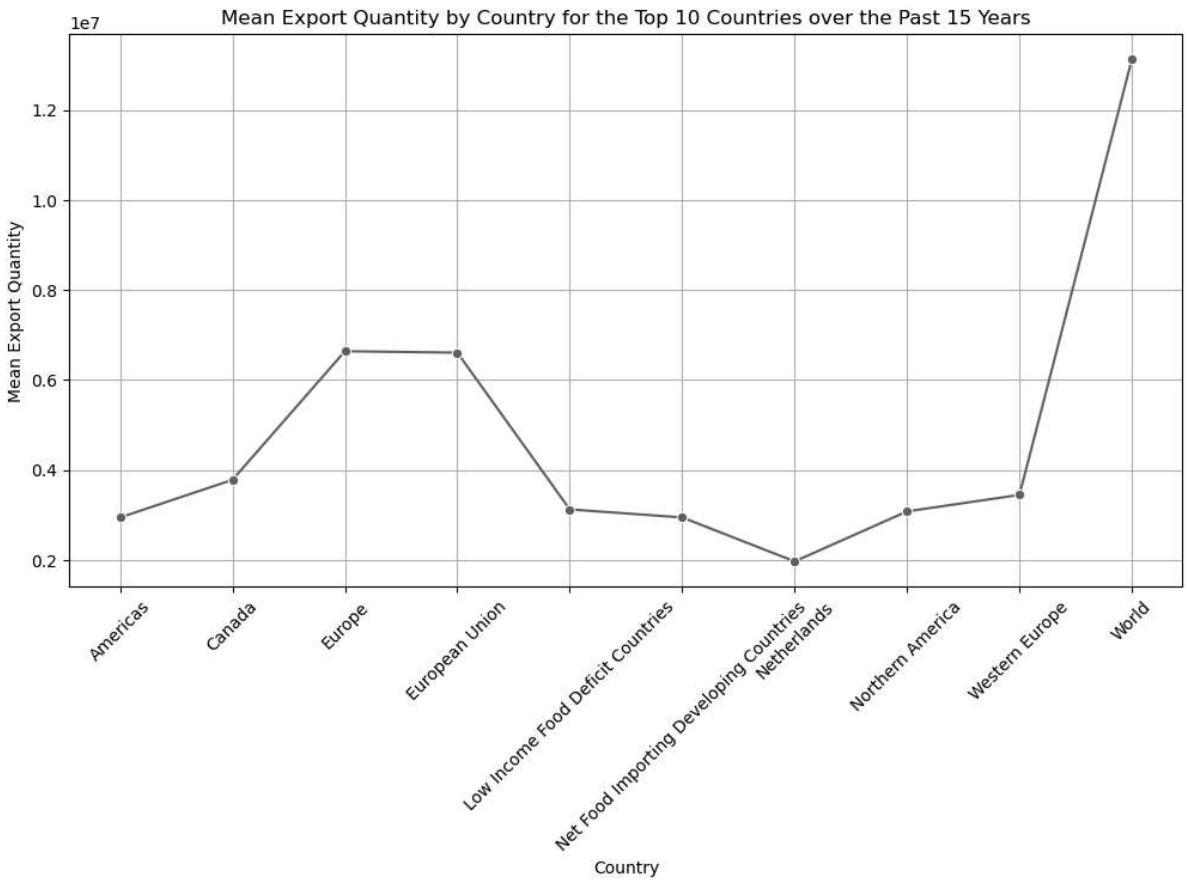
# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate mean export quantity
mean_export_quantity_by_country = top_countries_data.groupby('country')['value'].mean()

# Line plot for mean export quantity by country
plt.figure(figsize=(12, 6))
sns.lineplot(x=mean_export_quantity_by_country.index, y=mean_export_quantity_by_country)
plt.title("Mean Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Country")
plt.ylabel("Mean Export Quantity")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()

```



```
In [13]: # Descriptive statistics and Bar plot showing mean export quantity of meat animals

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

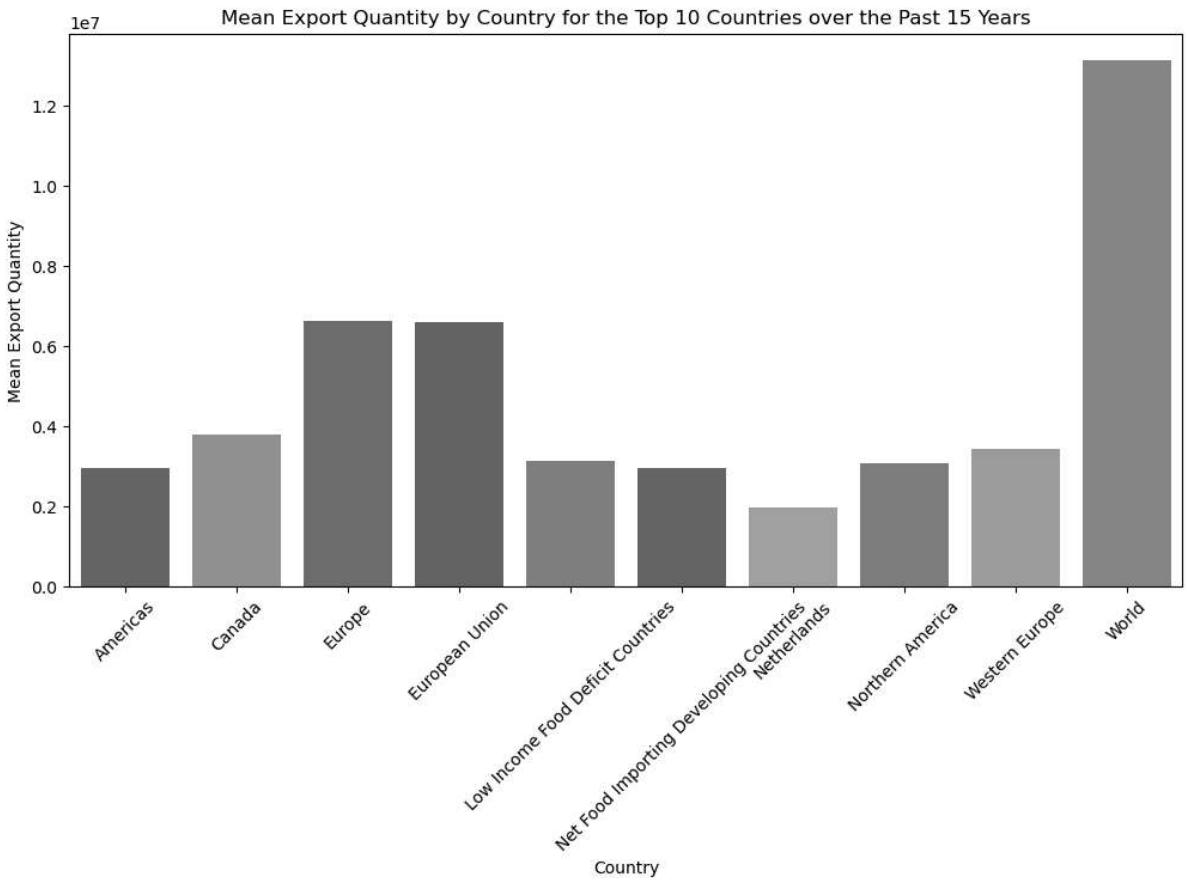
# Group by country and calculate descriptive statistics
descriptive_stats_by_country = top_countries_data.groupby('country')['value'].describe()

# Print descriptive statistics
print("Descriptive Statistics for Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
print(descriptive_stats_by_country)

# Bar plot for export quantity by country
plt.figure(figsize=(12, 6))
sns.barplot(x=descriptive_stats_by_country.index, y='mean', data=descriptive_stats_by_country)
plt.title("Mean Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Country")
plt.ylabel("Mean Export Quantity")
plt.xticks(rotation=45)
plt.show()
```

Descriptive Statistics for Export Quantity by Country for the Top 10 Countries over the Past 15 Years:

country	count	mean	std	\
Americas	55.0	2.944988e+06	2.849807e+06	
Canada	32.0	3.781423e+06	3.119148e+06	
Europe	64.0	6.638819e+06	7.440723e+06	
European Union	64.0	6.607899e+06	7.429845e+06	
Low Income Food Deficit Countries	32.0	3.124494e+06	2.404998e+06	
Net Food Importing Developing Countries	32.0	2.945832e+06	2.164036e+06	
Netherlands	58.0	1.968984e+06	3.310132e+06	
Northern America	42.0	3.080142e+06	3.104551e+06	
Western Europe	64.0	3.446970e+06	4.242584e+06	
World	64.0	1.312981e+07	1.022243e+07	
	min	25%	50%	\
Americas	100277.0	370858.00	2683492.0	
Canada	139692.0	1047194.50	2906757.5	
Europe	415581.0	2884227.75	4401929.5	
European Union	413703.0	2873526.75	4359414.5	
Low Income Food Deficit Countries	711250.0	910484.50	1758602.0	
Net Food Importing Developing Countries	793710.0	1060442.25	2150525.5	
Netherlands	105348.0	192357.00	251583.0	
Northern America	116668.0	589801.75	1474837.0	
Western Europe	297960.0	796413.50	1922915.5	
World	630202.0	6317036.50	12559151.5	
	75%	max		
Americas	4401974.00	10178524.0		
Canada	5776252.50	10031949.0		
Europe	5643787.00	31142767.0		
European Union	5573773.25	31080963.0		
Low Income Food Deficit Countries	5279497.75	7569734.0		
Net Food Importing Developing Countries	4357796.50	8259538.0		
Netherlands	2508488.75	11371905.0		
Northern America	5650847.50	10168665.0		
Western Europe	3199544.75	15863599.0		
World	16623648.25	38577345.0		



```
In [14]: # Descriptive statistics and Horizontal Bar plot showing mean export quantity of me

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

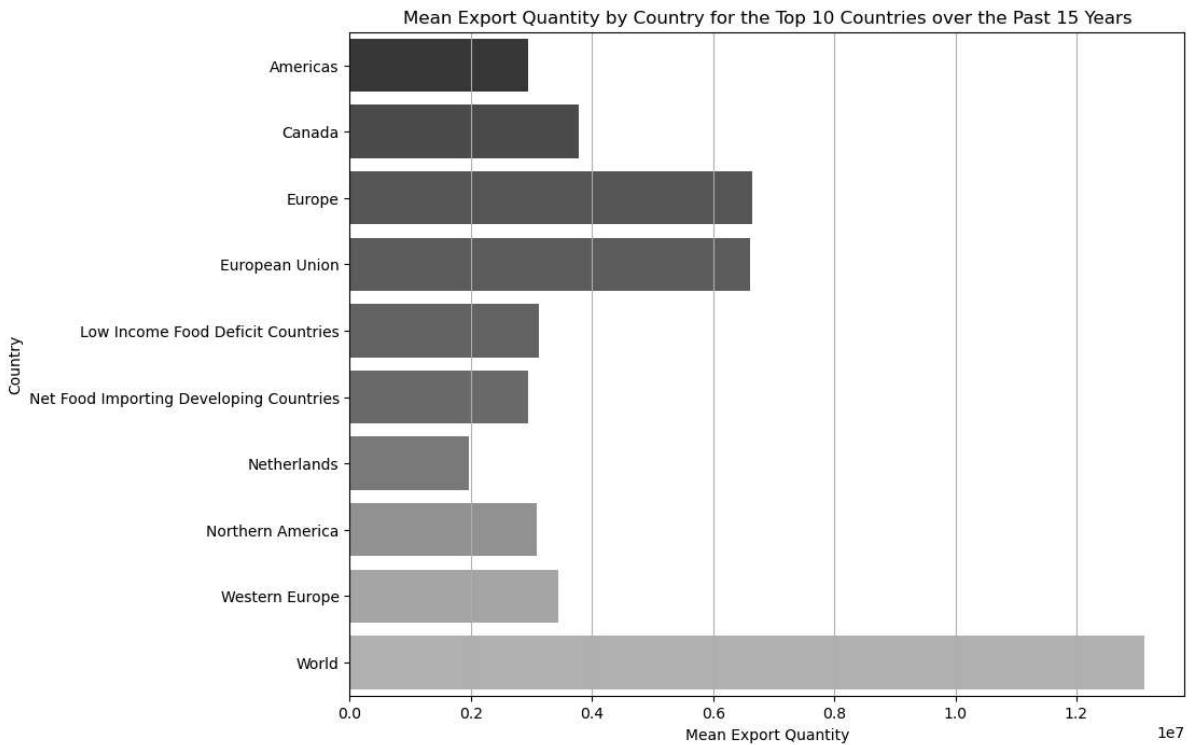
# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate mean export quantity
mean_export_quantity_by_country = top_countries_data.groupby('country')['value'].mean()

# Horizontal bar plot for mean export quantity by country
plt.figure(figsize=(10, 8))
sns.barplot(x=mean_export_quantity_by_country.values, y=mean_export_quantity_by_country.index)
plt.title("Mean Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Mean Export Quantity")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [15]: # Descriptive statistics and Box plot showing mean export quantity of meat animals

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

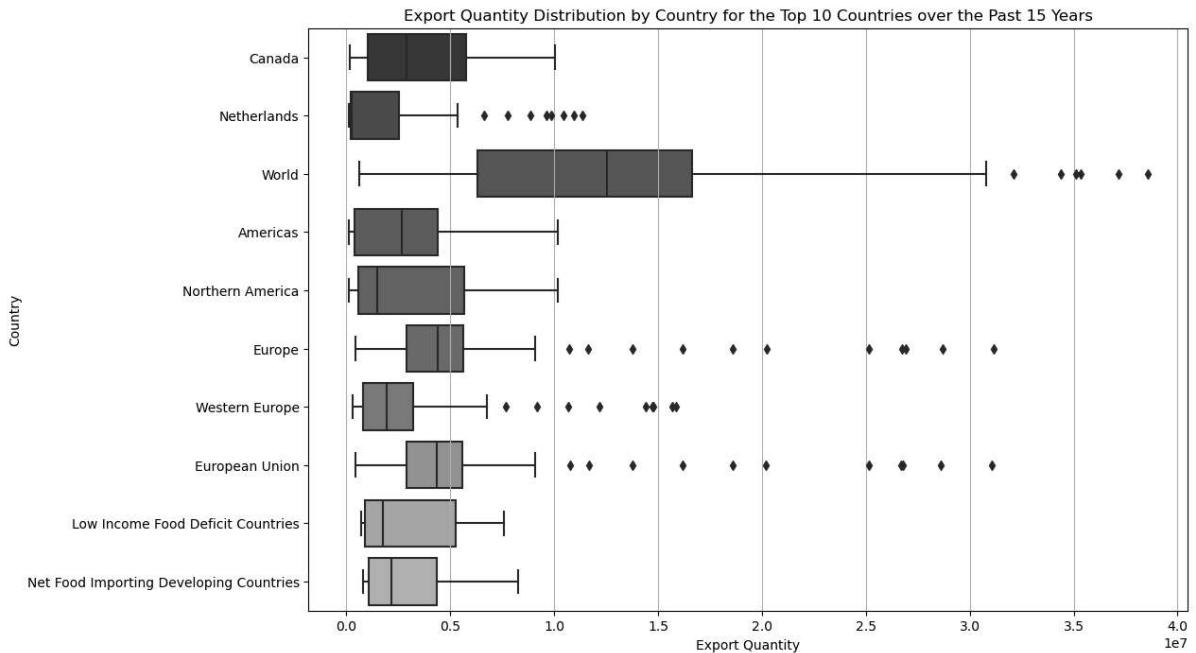
# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate mean export quantity
mean_export_quantity_by_country = top_countries_data.groupby('country')['value'].mean()

# Box plot for mean export quantity by country
plt.figure(figsize=(12, 8))
sns.boxplot(x='value', y='country', data=top_countries_data, orient='h', palette='dark')
plt.title("Export Quantity Distribution by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Quantity")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [16]: # Descriptive statistics and Violin plot showing mean export quantity of meat anima
```

```
import seaborn as sns
import matplotlib.pyplot as plt

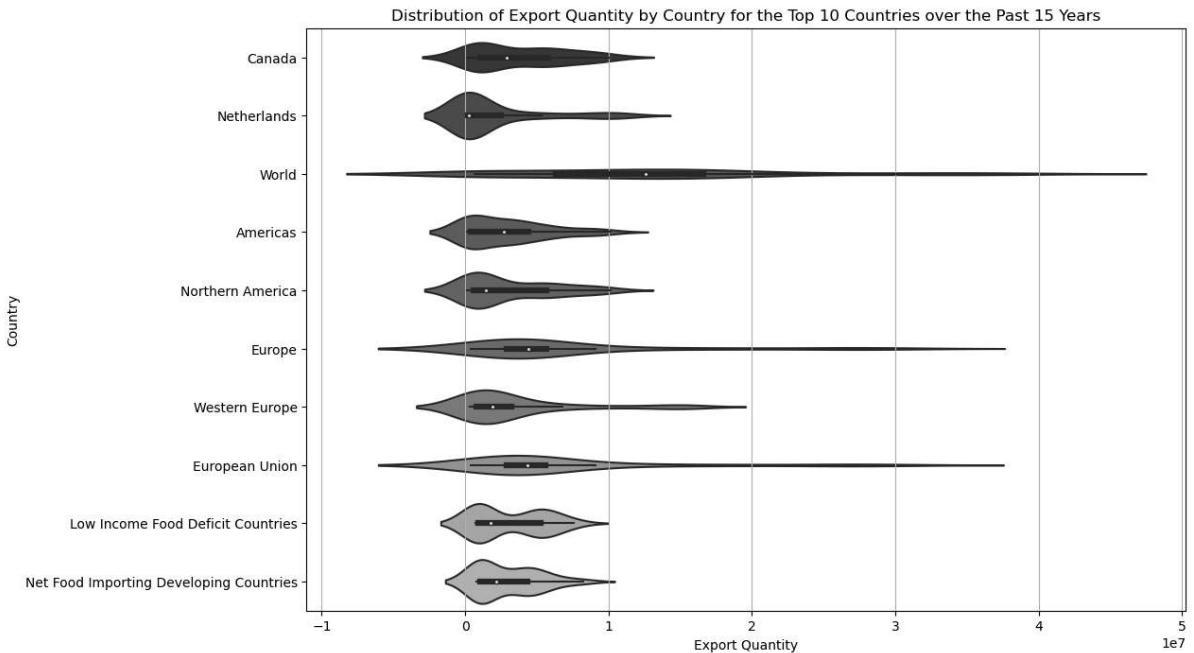
# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Violin plot for export quantity by country
plt.figure(figsize=(12, 8))
sns.violinplot(x='value', y='country', data=top_countries_data, orient='h', palette='viridis')
plt.title("Distribution of Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Quantity")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [17]: # Descriptive statistics and Swarm plot showing mean export quantity of meat animal

import seaborn as sns
import matplotlib.pyplot as plt

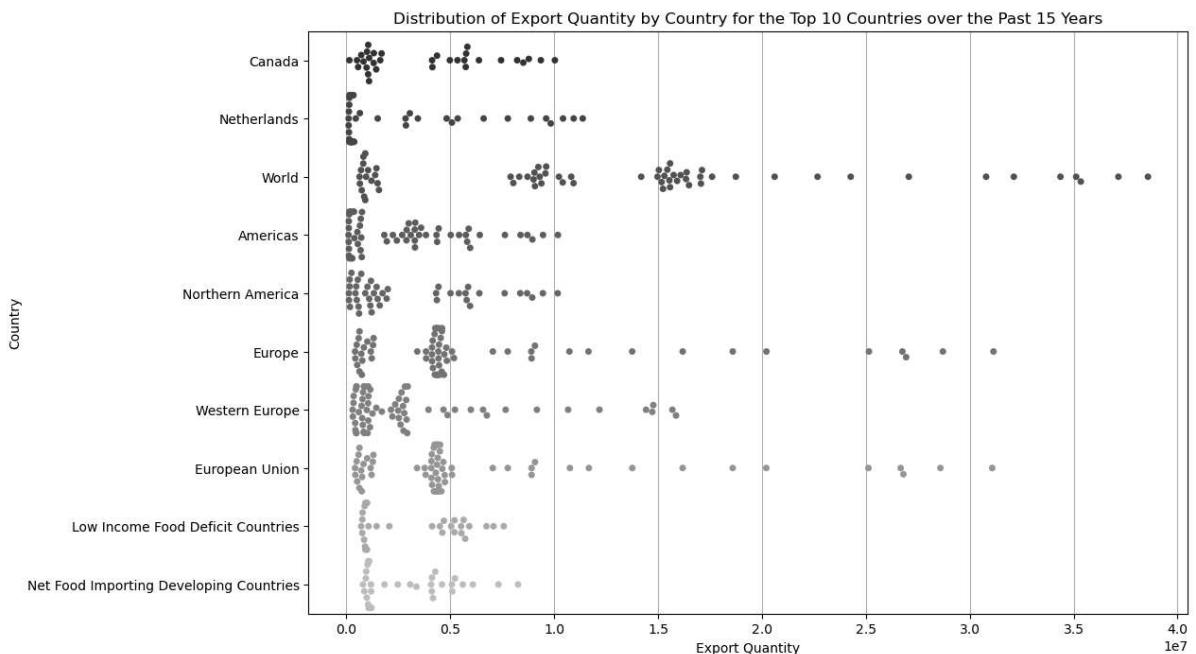
# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Swarm plot for export quantity by country
plt.figure(figsize=(12, 8))
sns.swarmplot(x='value', y='country', data=top_countries_data, palette='viridis')
plt.title("Distribution of Export Quantity by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Quantity")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



In []:

```
In [18]: # Descriptive statistics of 'Export Qunty' by animal type: considering top 10 cou

import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_quantity_by_animal = df[df['element'] == 'Export Quantity']

# Descriptive Statistics by animal_type
descriptive_stats_by_animal = export_quantity_by_animal.groupby('animal_type')[['val']]

# Print descriptive statistics
print("Descriptive Statistics for Export Quantity (heads of exported live animals)")
print(descriptive_stats_by_animal)
```

Descriptive Statistics for Export Quantity (heads of exported live animals) by Animal Type:

animal_type	count	mean	std	min	25%	\
cattle	2318.0	8.743993e+05	1.394701e+06	100000.0	175464.5	
chickens	354.0	3.000892e+05	2.786749e+05	100164.0	123928.0	
pigs	1271.0	2.826140e+06	4.694086e+06	100700.0	323594.5	
sheep	2403.0	1.635610e+06	2.624821e+06	100000.0	269168.0	

animal_type	50%	75%	max
cattle	386356.5	910698.00	10930750.0
chickens	187219.5	335000.25	1563085.0
pigs	1152127.0	3085693.50	38577345.0
sheep	679419.0	1867746.50	23427122.0

In [19]: # Time Series Line plot showing trend of export quantity of meat animals in top 10

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_quantity_by_country = df[df['element'] == 'Export Quantity']

# Filter data for the past 15 years
```

```

past_15_years_data = export_quantity_by_country[df['year'] >= df['year'].max() - 15]

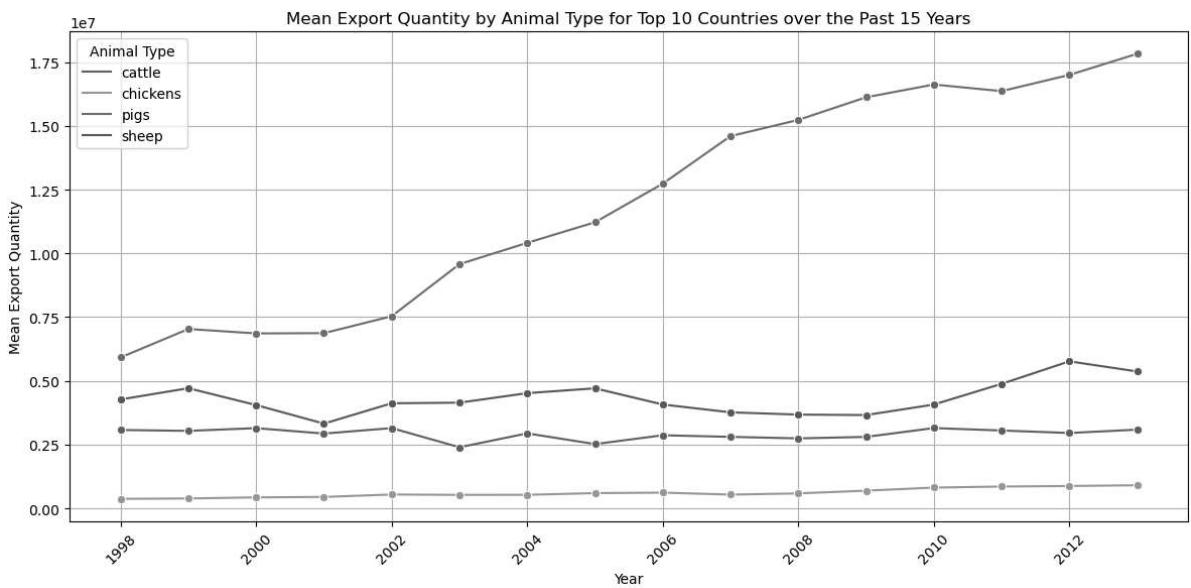
# Group by country and sum the export quantities
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by animal_type and year and calculate mean export quantity
mean_export_quantity_by_animal_country = top_countries_data.groupby(['animal_type', 'year']).mean()

# Plot
plt.figure(figsize=(12, 6))
sns.lineplot(data=mean_export_quantity_by_animal_country, x='year', y='value', hue='Animal Type')
plt.title("Mean Export Quantity by Animal Type for Top 10 Countries over the Past 15 Years")
plt.xlabel("Year")
plt.ylabel("Mean Export Quantity")
plt.legend(title='Animal Type')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show

```

Out[19]: <function matplotlib.pyplot.show(close=None, block=None)>



In []:

```

In [20]: # Descriptive statistics of 'Export Value' expressed as number of heads of Live animals

import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_value_df = df[df['element'] == 'Export Value']

# Descriptive Statistics
print("Descriptive Statistics for Export Value:")
print(export_value_df['value'].describe())

```

```
Descriptive Statistics for Export Value:
count    2.988000e+03
mean     5.153166e+05
std      7.907508e+05
min      1.000000e+05
25%     1.574100e+05
50%     2.398820e+05
75%     4.555805e+05
max      8.890544e+06
Name: value, dtype: float64
```

In []:

```
# Time series Line plot showing trend of Export Value (1000 US$) of Live animals in the past 15 years

import seaborn as sns
import matplotlib.pyplot as plt

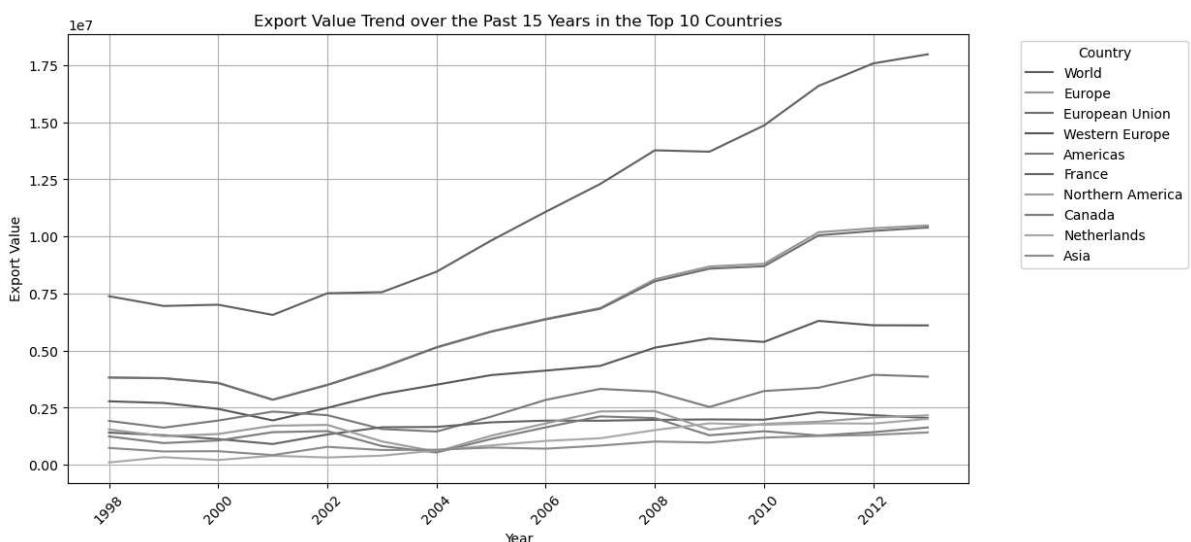
# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values for each year
top_countries = past_15_years_data.groupby(['country', 'year'])['value'].sum().unstack()
top_countries_total = top_countries.sum().nlargest(10).index
top_countries_data = top_countries[top_countries_total]

# Line plot for export value trend over the past 15 years in the top 10 countries
plt.figure(figsize=(12, 6))
for country in top_countries_data.columns:
    sns.lineplot(x=top_countries_data.index, y=top_countries_data[country], label=country)
plt.title("Export Value Trend over the Past 15 Years in the Top 10 Countries")
plt.xlabel("Year")
plt.ylabel("Export Value")
plt.xticks(rotation=45)
plt.legend(title='Country', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()
```



In []:

```
In [22]: # Line plot showing mean export value (1000 US$) of meat animals in top 10 countries
```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

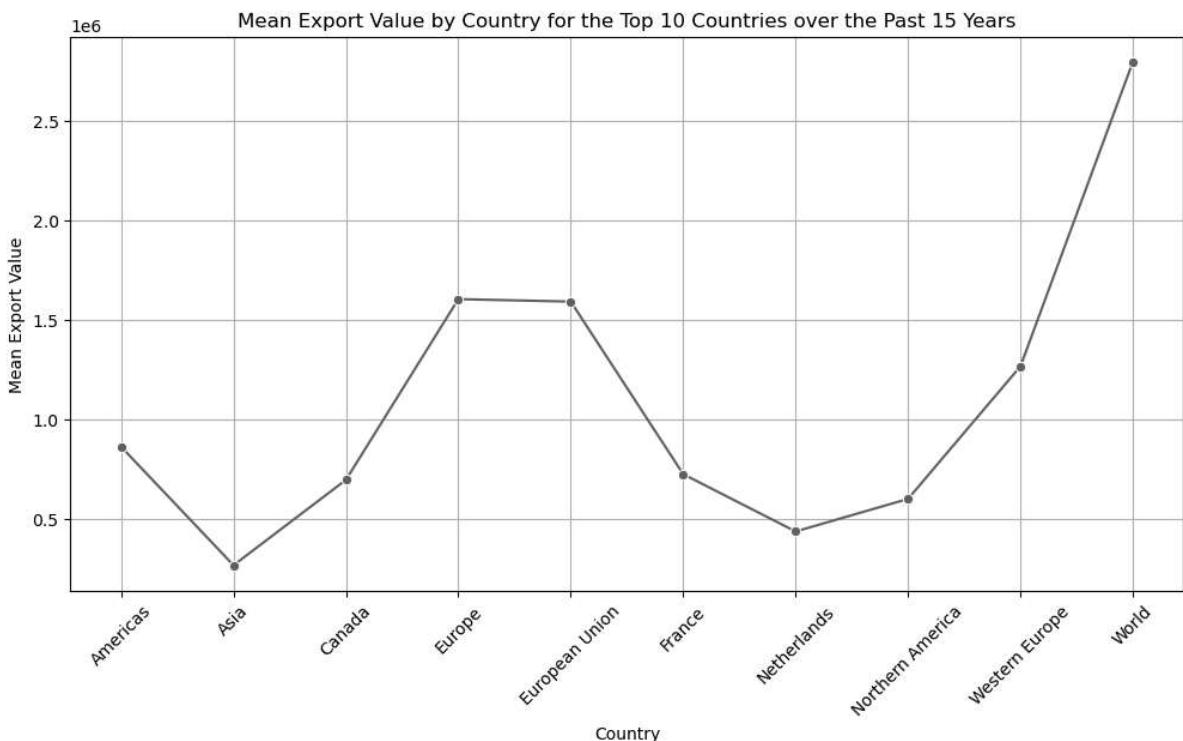
# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate mean export value
mean_export_value_by_country = top_countries_data.groupby('country')['value'].mean()

# Line plot for mean export value by country
plt.figure(figsize=(12, 6))
sns.lineplot(x=mean_export_value_by_country.index, y=mean_export_value_by_country)
plt.title("Mean Export Value by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Country")
plt.ylabel("Mean Export Value")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



```
In [23]: # Descriptive statistics and Bar plot showing mean export value (1000 US$) of meat
```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)
```

```

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate descriptive statistics
descriptive_stats_by_country = top_countries_data.groupby('country')['value'].describe()

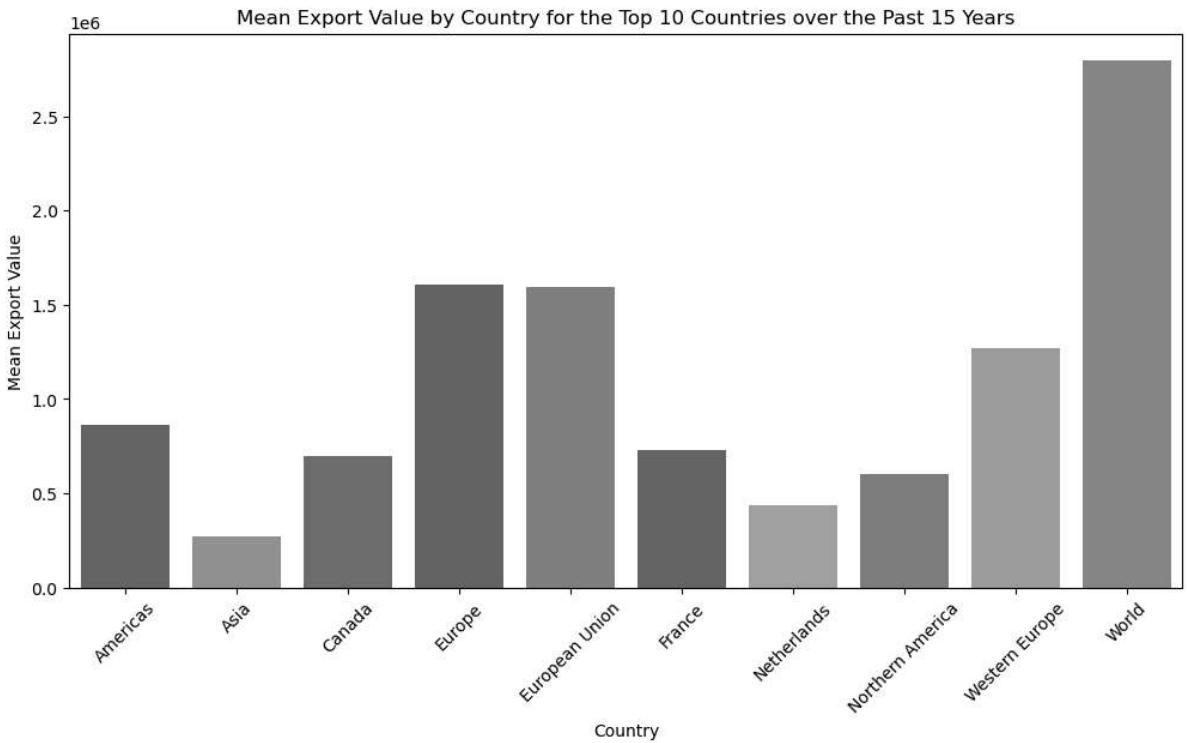
# Print descriptive statistics
print("Descriptive Statistics for Export Value by Country for the Top 10 Countries over the Past 15 Years:")
print(descriptive_stats_by_country)

# Bar plot for export value by country
plt.figure(figsize=(12, 6))
sns.barplot(x=descriptive_stats_by_country.index, y='mean', data=descriptive_stats_by_country)
plt.title("Mean Export Value by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Country")
plt.ylabel("Mean Export Value")
plt.xticks(rotation=45)
plt.show()

```

Descriptive Statistics for Export Value by Country for the Top 10 Countries over the Past 15 Years:

country	count	mean	std	min	25%	\
Americas	48.0	8.640385e+05	9.101665e+05	102488.0	230829.25	
Asia	52.0	2.685904e+05	1.221061e+05	113329.0	172343.75	
Canada	31.0	6.974804e+05	3.790031e+05	226898.0	368523.50	
Europe	64.0	1.605615e+06	1.314218e+06	174679.0	466991.50	
European Union	64.0	1.593305e+06	1.298620e+06	174211.0	465075.25	
France	38.0	7.261823e+05	7.103875e+05	102578.0	147858.75	
Netherlands	37.0	4.382905e+05	3.843430e+05	102442.0	176771.00	
Northern America	44.0	6.020251e+05	4.794853e+05	103085.0	204670.00	
Western Europe	52.0	1.268368e+06	8.640348e+05	102499.0	449234.50	
World	64.0	2.798831e+06	2.240350e+06	566092.0	1103624.50	
country		50%	75%	max		
Americas	390979.0	1366642.25	3279596.0			
Asia	243845.0	349488.00	592297.0			
Canada	594647.0	1017094.50	1551512.0			
Europe	1183255.5	2618271.75	4512577.0			
European Union	1181582.0	2611188.75	4424247.0			
France	173592.0	1516512.75	1983630.0			
Netherlands	265584.0	496946.00	1343660.0			
Northern America	398598.0	1039929.50	1659750.0			
Western Europe	1091040.5	2074056.75	3057158.0			
World	1708194.0	4257576.50	8890544.0			



```
In [24]: # Horizontal Bar plot showing mean export value (1000 US$) of meat animals in top 10 countries over the past 15 years

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

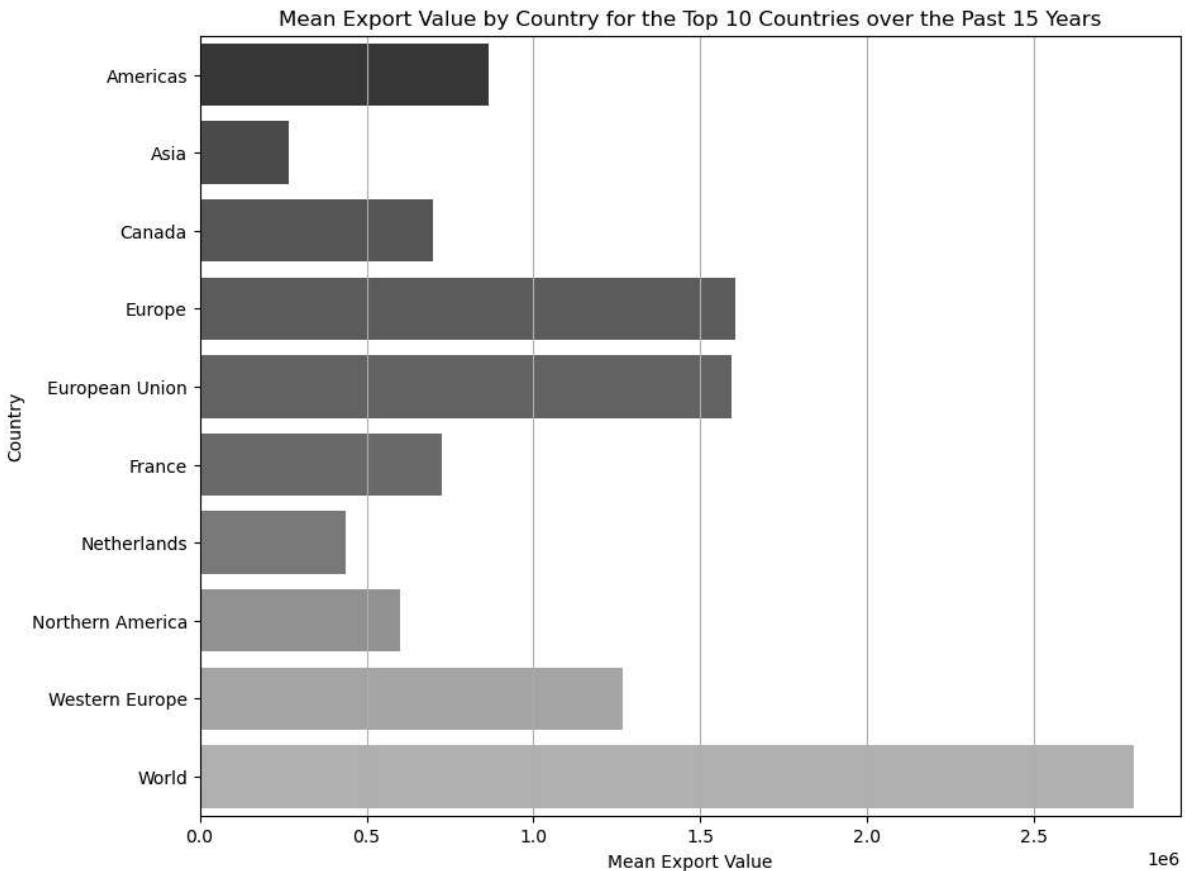
# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by country and calculate mean export value
mean_export_value_by_country = top_countries_data.groupby('country')['value'].mean()

# Horizontal bar plot for mean export value by country
plt.figure(figsize=(10, 8))
sns.barplot(x=mean_export_value_by_country.values, y=mean_export_value_by_country.index)
plt.title("Mean Export Value by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Mean Export Value")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [25]: # Box plot showing mean export value (1000 US$) of meat animals in top 10 countries

import seaborn as sns
import matplotlib.pyplot as plt

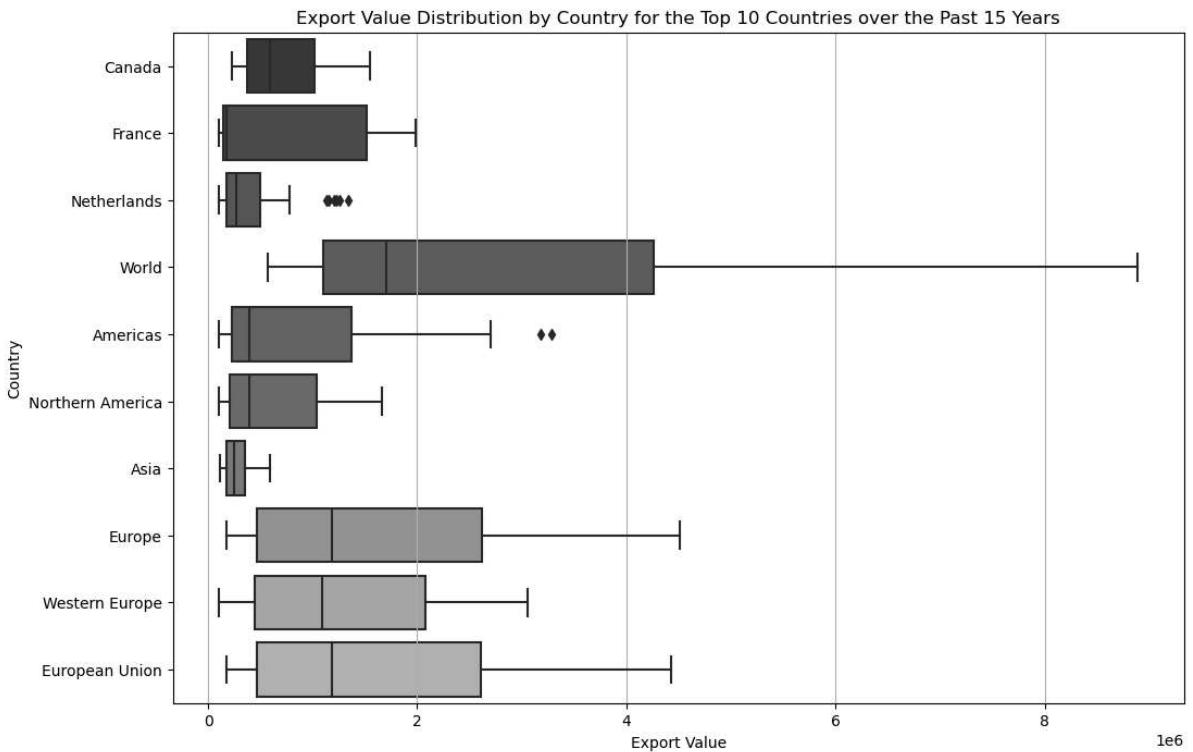
# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Box plot for export value by country
plt.figure(figsize=(12, 8))
sns.boxplot(x='value', y='country', data=top_countries_data, orient='h', palette='viridis')
plt.title("Export Value Distribution by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Value")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [26]: # Violin plot showing mean export value (1000 US$) of meat animals in top 10 countries

import seaborn as sns
import matplotlib.pyplot as plt

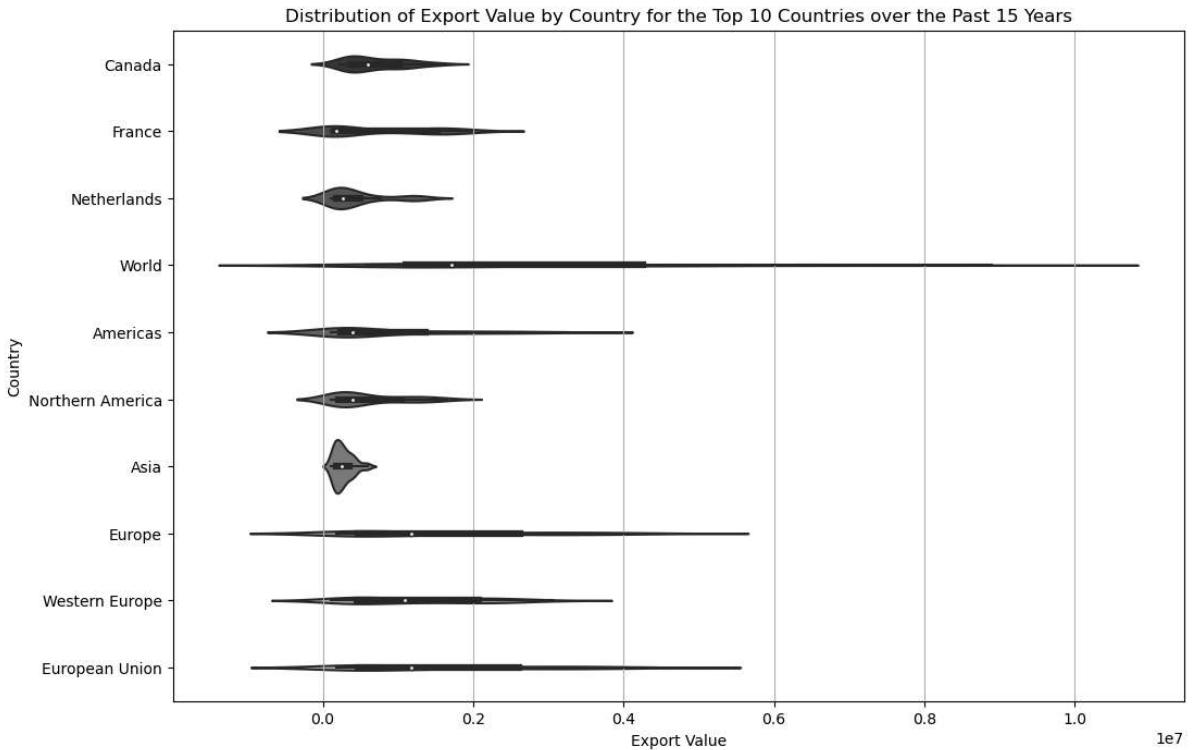
# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Violin plot for export value by country
plt.figure(figsize=(12, 8))
sns.violinplot(x='value', y='country', data=top_countries_data, orient='h', palette='viridis')
plt.title("Distribution of Export Value by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Value")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [27]: # Swarm plot showing mean export value (1000 US$) of meat animals in top 10 countries

import seaborn as sns
import matplotlib.pyplot as plt

# Rename variables
df.rename(columns={'Area': 'country', 'Element': 'element', 'Year': 'year'}, inplace=True)

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']

# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Swarm plot for export value by country
plt.figure(figsize=(12, 8))
sns.swarmplot(x='value', y='country', data=top_countries_data, palette='viridis')
plt.title("Distribution of Export Value by Country for the Top 10 Countries over the Past 15 Years")
plt.xlabel("Export Value")
plt.ylabel("Country")
plt.grid(True, axis='x')
plt.show()
```



```
In [ ]:
```

```
In [28]: # Descriptive statistics and Time series Line plot showing trends in export value (
```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_value_by_animal = df[df['element'] == 'Export Value']

# Descriptive Statistics by animal_type
descriptive_stats_by_animal = export_value_by_animal.groupby('animal_type')['value']

# Print descriptive statistics
print("Descriptive Statistics for Export Value (1000 US$) by Animal Type:")
print(descriptive_stats_by_animal)
```

Descriptive Statistics for Export Value (1000 US\$) by Animal Type:

animal_type	count	mean	std	min	25%	\
cattle	1325.0	670086.222642	1.022668e+06	100000.0	171507.00	
chickens	456.0	326788.256579	3.822691e+05	100071.0	129084.75	
pigs	658.0	534736.957447	6.850181e+05	100697.0	180162.75	
sheep	549.0	275098.919854	2.275189e+05	100084.0	150316.00	
		50%	75%	max		
animal_type						
cattle	264650.0	592554.00	8890544.0			
chickens	184389.0	338906.50	2526618.0			
pigs	291269.0	601047.25	5029732.0			
sheep	210334.0	299572.00	1681812.0			

```
In [29]:
```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter the DataFrame based on the criteria
export_value_by_country = df[df['element'] == 'Export Value']
```

```

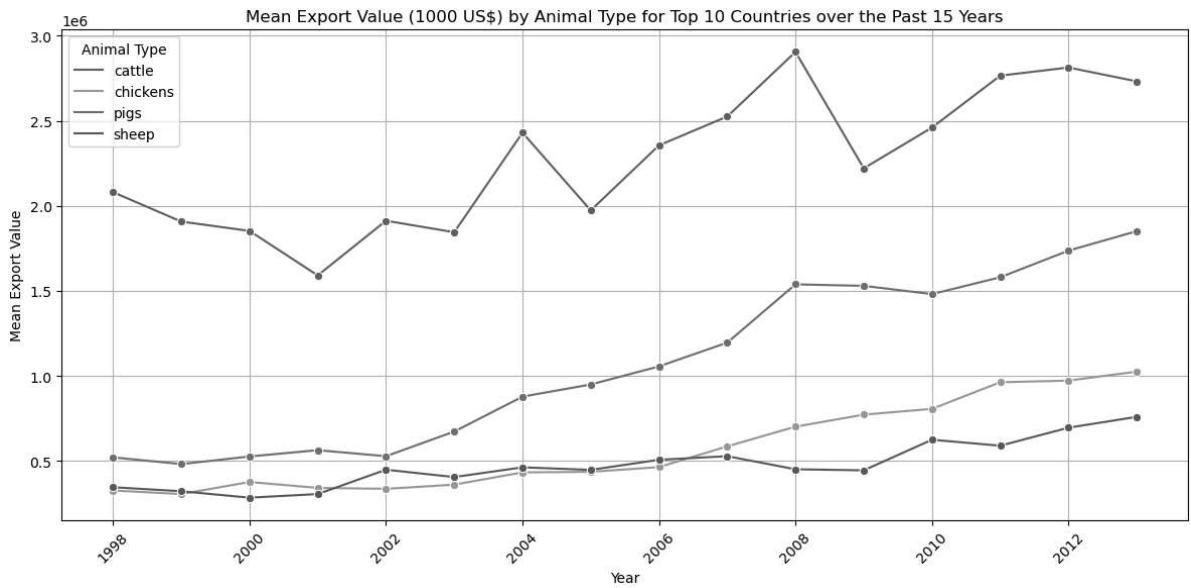
# Filter data for the past 15 years
past_15_years_data = export_value_by_country[df['year'] >= df['year'].max() - 15]

# Group by country and sum the export values
top_countries = past_15_years_data.groupby('country')['value'].sum().nlargest(10).index
top_countries_data = past_15_years_data[past_15_years_data['country'].isin(top_countries)]

# Group by animal_type and year and calculate mean export value
mean_export_value_by_animal_country = top_countries_data.groupby(['animal_type', 'year'])['value'].mean()

# Plot
plt.figure(figsize=(12, 6))
sns.lineplot(data=mean_export_value_by_animal_country, x='year', y='value', hue='animal_type')
plt.title("Mean Export Value (1000 US$) by Animal Type for Top 10 Countries over the Past 15 Years")
plt.xlabel("Year")
plt.ylabel("Mean Export Value")
plt.legend(title='Animal Type')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```



In []:

Results and Discussion: Milestone 1

The descriptive statistics and visualizations provide insightful information about the export of live animals over the past 15 years, focusing on both the quantity and value aspects.

Starting with the descriptive statistics, the mean export quantity of live animals varies significantly across different countries. For example, the mean export quantity (in thousands) over the past 15 years is highest for the Europe at approximately 6.63 million, followed closely by European Union at approximately 6.60 million, then Canada at approximately 3.78 million, and the Western Europe at approximately 3.44 million. This suggests varying levels of live animal export activity among different regions. However, there is considerable variability in export quantities, as indicated by the standard deviations. For instance, Europe exhibits a standard deviation of approximately 7.44 million, indicating substantial variability in export quantities within the region. Moreover, considering the top 10 exporters, the range of export quantities varies widely among countries, with minimum export quantities ranging

from 100 thousand to 297.9 thousand and maximum export quantities ranging from 10.1 million to 15.8 million.

Similarly, the mean export quantity of live animals varies significantly across different animal types. For instance, the mean export quantity (in thousand) over the past 15 years is highest for pigs at approximately 2.83 million, followed by sheep at approximately 1.64 million, cattle at approximately 874,399, and chickens at approximately 300,089. This indicates that pigs have been consistently exported in larger quantities compared to other animal types. However, there is notable variability in export quantities, as indicated by the standard deviations. For example, pigs have a standard deviation of approximately 4.69 million, suggesting greater variability compared to chickens with a standard deviation of approximately 278,675. Additionally, the range of export quantities varies widely among animal types, with pigs ranging from 100.7 thousand to 38.58 million and chickens ranging from 100.16 thousand to 1.56 million.

Moving on to the visualizations, the line plot depicting the mean export quantity by country for the top 10 countries over the past 15 years reveals interesting trends. For instance, Europe and the European Union consistently maintain high mean export quantities, followed by Canada and Europe. Conversely, Asian countries exhibit less export quantity among those top 10 exporters. The other plots also provide a clear comparison of mean export quantities across countries, with Europe, the EU, Canada, and Western Europe ranking among the top 10 exporters, respectively; while, Asia the least among those top 10 exporters.

The box plot and violin plot further illustrate the distribution of export values by country. For instance, while the median export value for Europe is around **1.18million**, indicating a relatively stable level of exports, the distribution is wide, with the 0.17 million to **4.51million**. In contrast, some countries, such as Canada, exhibit a narrower distribution 0.59 million and less variability. Similarly, the descriptive statistics and visualizations for export value provide insights into the economic aspect of live animal trade. The mean export value varies across countries, with Europe having the highest mean export value of approximately **1.61million**, followed by the European Union at approximately **1.59million**, and Western Europe at approximately \$1.27 million over the 15-year period.

The box plot and violin plot further illustrate the distribution of export values by animal type. For instance, while the median export value for cattle is around **264.7thousand**, indicating a relatively stable level of exports, the distribution is wide, with the 100 thousand to **8.89million**. In contrast, some animal types, such as sheep, exhibit a narrower distribution 210.3 thousand and less variability. Similarly, the descriptive statistics and visualizations for export value provide insights into the economic aspect of live animal trade. The mean export value varies across animal types, with cattle having the highest mean export value of approximately **670.1thousand**, followed by pigs at approximately **534.7thousand**, sheep at approximately **275.1thousand**, and chickens at approximately **326.8thousand** over the 15-year period.

In summary, integrating numerical values into the discussion enhances our understanding of the descriptive statistics and visualizations, providing concrete data points to support our

observations and interpretations.

Conclusions:

The analysis of live animal export data over the past 15 years reveals significant variability in export quantities and values across different animal types and countries. Pigs emerge as the most consistently exported animals, with Europe and the European Union standing out as major exporters. The descriptive statistics and visualizations highlight trends and patterns in export quantities and values, offering valuable insights for stakeholders in the agricultural and trade sectors. Understanding these trends can aid policymakers, industry professionals, and researchers in making informed decisions regarding trade regulations, market strategies, and resource allocation.

Way Forward:

Looking ahead, the dataset will undergo advanced analytics in the project's milestones 2 and 3 phases. This includes predictive modeling, time series analysis, and clustering to unveil deeper insights into live animal exports. By leveraging these techniques, I aim to identify predictive factors influencing export quantities and values, anticipate future trends, and segment markets based on demand patterns. Interdisciplinary collaboration will be crucial in integrating ethical considerations and sustainability principles into trade policies and practices. Engaging with stakeholders will foster dialogue and implement measures to promote economic prosperity and animal welfare in the live animal trade.

In []: