File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Load and clean data
df = pd.read_csv("C:\\Users\\ashut\\Downloads\\WDFW-Salmonid_Population_Indicators__SPI__Escapement_and_pHOS_20250409.csv")
df.columns = df.columns.str.strip()
print("Columns:", df.columns.tolist())

# BASIC EDA
print("\nBasic Info")
print(df.info())

print("\nMissing Values")
print(df.isnull().sum())

print("\nSummary Statistics")
print(df.describe(include='all'))

print("\nSpecies Counts")
print(df["Species"].value_counts())

# Correlation Heatmap (only for numeric columns)
numeric_df = df.select_dtypes(include=[np.number])
correlation_matrix = numeric_df.corr()

plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap="coolwarm", linewidths=0.5, linecolor='white')
plt.title("Correlation Heatmap of Numerical Columns")
plt.tight_layout()
plt.show()

# Abundance Over Time by Species
if 'Indicator' in df.columns and 'Species' in df.columns and 'Year' in df.columns and 'Value' in df.columns:
    abundance_df = df[df["Indicator"].str.contains("Escapement", case=False, na=False)]
    selected_species = ["Chinook", "Coho", "Steelhead"]

    plt.figure(figsize=(14, 7))
    sns.lineplot(data=abundance_df[abundance_df["Species"].isin(selected_species)],
                 x="Year", y="Value", hue="Species", marker="o")
    plt.title("Escapement (Abundance) Over Time")
    plt.xlabel("Year")
    plt.ylabel("Abundance Value")
    plt.grid(True)
    plt.tight_layout()
    plt.show()
```

```
# Boxplot of Value by Species
    plt.figure(figsize=(12, 6))
    sns.boxplot(data=df[df["Species"].isin(selected_species)], x="Species", y="Value")
    plt.title("Boxplot of 'Value' by Species")
    plt.ylabel("Value")
    plt.grid(True)
    plt.tight_layout()
    plt.show()
else:
    print("One or more required columns ('Indicator', 'Species', 'Year', 'Value') are missing.")

# Species Count Barplot
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='Species', order=df['Species'].value_counts().index)
plt.title("Count of Records per Species")
plt.xticks(rotation=45)
plt.tight_layout()
plt.grid(True)
plt.show()
```

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
>>> 
========== RESTART: C:/Users/ashut/AppData/Local/Programs/Python/Python313/ca2python.py ==========
Columns: ['Stock Number', 'Population Name', 'Sub-Population Name', 'Data Series', 'Species', 'Year', 'Abundance Quantity', 'Data Type', 'Production Type
', 'Calculation Type', 'Escapement Methodology', 'Escapement Methodology Description', 'Biologist Methodology Description', 'Comments', 'Report Types', '
Last Updated']

Basic Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26039 entries, 0 to 26038
Data columns (total 16 columns):
 #   Column                              Non-Null Count  Dtype  
---  ------                              --------------  -----  
 0   Stock Number                        26039 non-null  int64  
 1   Population Name                     26039 non-null  object 
 2   Sub-Population Name                 4460 non-null   object 
 3   Data Series                         26039 non-null  int64  
 4   Species                             26039 non-null  object 
 5   Year                                25964 non-null  float64
 6   Abundance Quantity                  25139 non-null  float64
 7   Data Type                           25987 non-null  object 
 8   Production Type                     18547 non-null  object 
 9   Calculation Type                    15932 non-null  object 
 10  Escapement Methodology              26039 non-null  object 
 11  Escapement Methodology Description  26039 non-null  object 
 12  Biologist Methodology Description   25229 non-null  object 
 13  Comments                            8220 non-null   object 
 14  Report Types                        23187 non-null  object 
 15  Last Updated                        26039 non-null  object 
dtypes: float64(2), int64(2), object(12)
memory usage: 3.2+ MB
None

Missing Values
Stock Number                            0
Population Name                         0
Sub-Population Name                 21579
Data Series                             0
Species                                 0
Year                                   75
Abundance Quantity                    900
Data Type                              52
Production Type                      7492
Calculation Type                    10107
Escapement Methodology                  0
Escapement Methodology Description      0
Biologist Methodology Description     810
Comments                            17819
Report Types                         2852
```

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
 13  Comments                          8220 non-null   object
 14  Report Types                      23187 non-null  object
 15  Last Updated                      26039 non-null  object
dtypes: float64(2), int64(2), object(12)
memory usage: 3.2+ MB
None

Missing Values
Stock Number                              0
Population Name                           0
Sub-Population Name                   21579
Data Series                               0
Species                                   0
Year                                     75
Abundance Quantity                      900
Data Type                                52
Production Type                        7492
Calculation Type                      10107
Escapement Methodology                    0
Escapement Methodology Description        0
Biologist Methodology Description       810
Comments                              17819
Report Types                           2852
Last Updated                              0
dtype: int64

Summary Statistics
        Stock Number  ...                   Last Updated
count   26039.000000  ...                          26039
unique           NaN  ...                           1185
top              NaN  ...  2023-05-15 22:28:54.169738+00
freq             NaN  ...                          15932
mean     3790.731172  ...                            NaN
std      2399.419008  ...                            NaN
min      1008.000000  ...                            NaN
25%      1634.000000  ...                            NaN
50%      2765.000000  ...                            NaN
75%      6553.000000  ...                            NaN
max      8900.000000  ...                            NaN

[11 rows x 16 columns]

Species Counts
Species
Chinook       8935
Steelhead     6854
Chum          4363
Coho          3514
```

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
Steelhead      6854
Chum           4363
Coho           3514
Bull Trout     1419
Sockeye         455
Pink            373
Cutthroat       126
Name: count, dtype: int64
One or more required columns ('Indicator', 'Species', 'Year', 'Value') are missing.
>>>
========== RESTART: C:/Users/ashut/AppData/Local/Programs/Python/Python313/ca2python.py ==========
Columns: ['Stock Number', 'Population Name', 'Sub-Population Name', 'Data Series', 'Species', 'Year', 'Abundance Quantity', 'Data Type', 'Production Type
', 'Calculation Type', 'Escapement Methodology', 'Escapement Methodology Description', 'Biologist Methodology Description', 'Comments', 'Report Types', '
Last Updated']

Basic Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26039 entries, 0 to 26038
Data columns (total 16 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Stock Number                        26039 non-null  int64
 1   Population Name                     26039 non-null  object
 2   Sub-Population Name                 4460 non-null   object
 3   Data Series                         26039 non-null  int64
 4   Species                             26039 non-null  object
 5   Year                                25964 non-null  float64
 6   Abundance Quantity                  25139 non-null  float64
 7   Data Type                           25987 non-null  object
 8   Production Type                     18547 non-null  object
 9   Calculation Type                    15932 non-null  object
 10  Escapement Methodology              26039 non-null  object
 11  Escapement Methodology Description  26039 non-null  object
 12  Biologist Methodology Description   25229 non-null  object
 13  Comments                            8220 non-null   object
 14  Report Types                        23187 non-null  object
 15  Last Updated                        26039 non-null  object
dtypes: float64(2), int64(2), object(12)
memory usage: 3.2+ MB
None

Missing Values
Stock Number                        0
Population Name                     0
Sub-Population Name                 21579
Data Series                         0
Species                             0
Year                                75
```

Figure 1

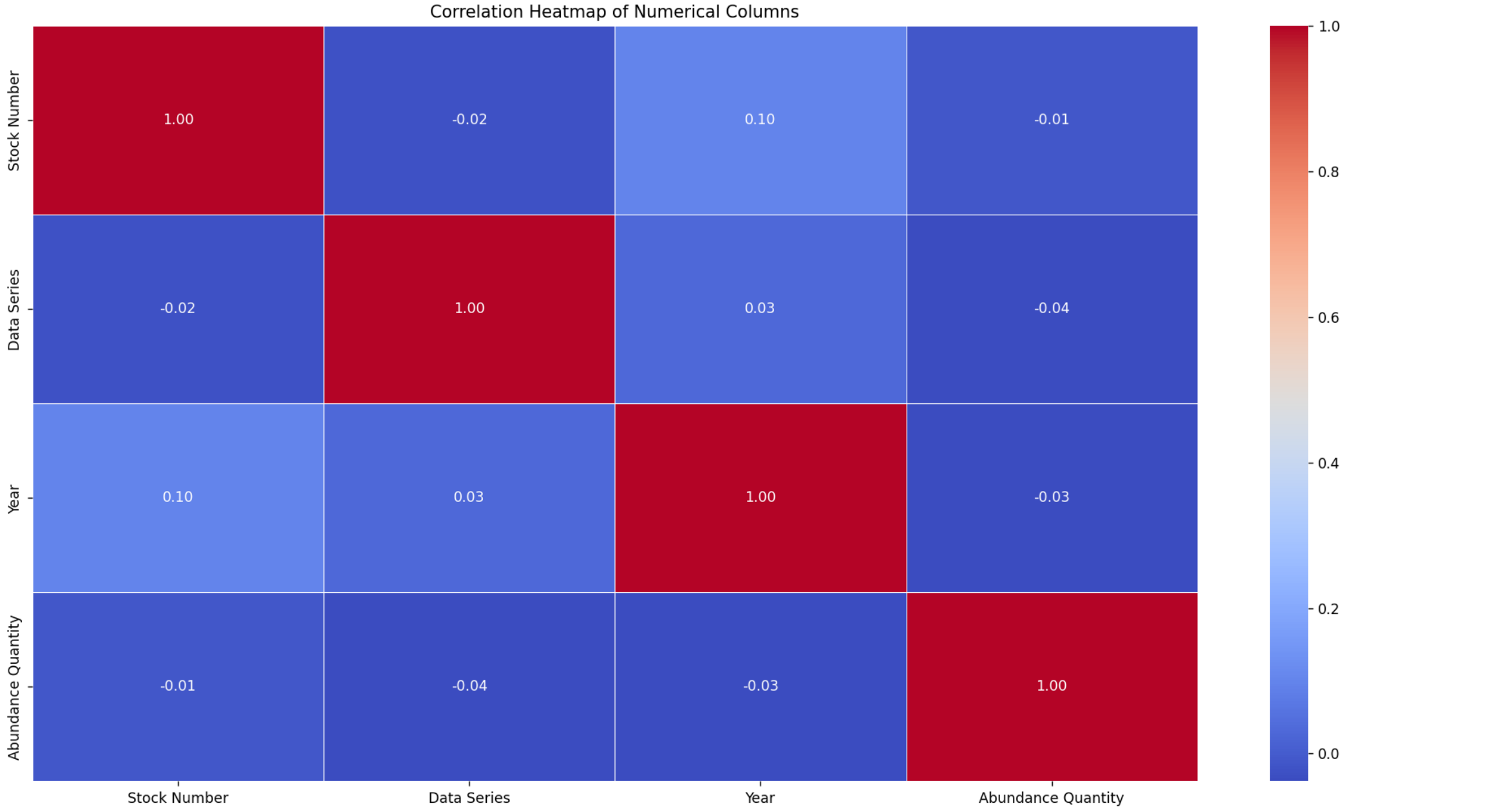
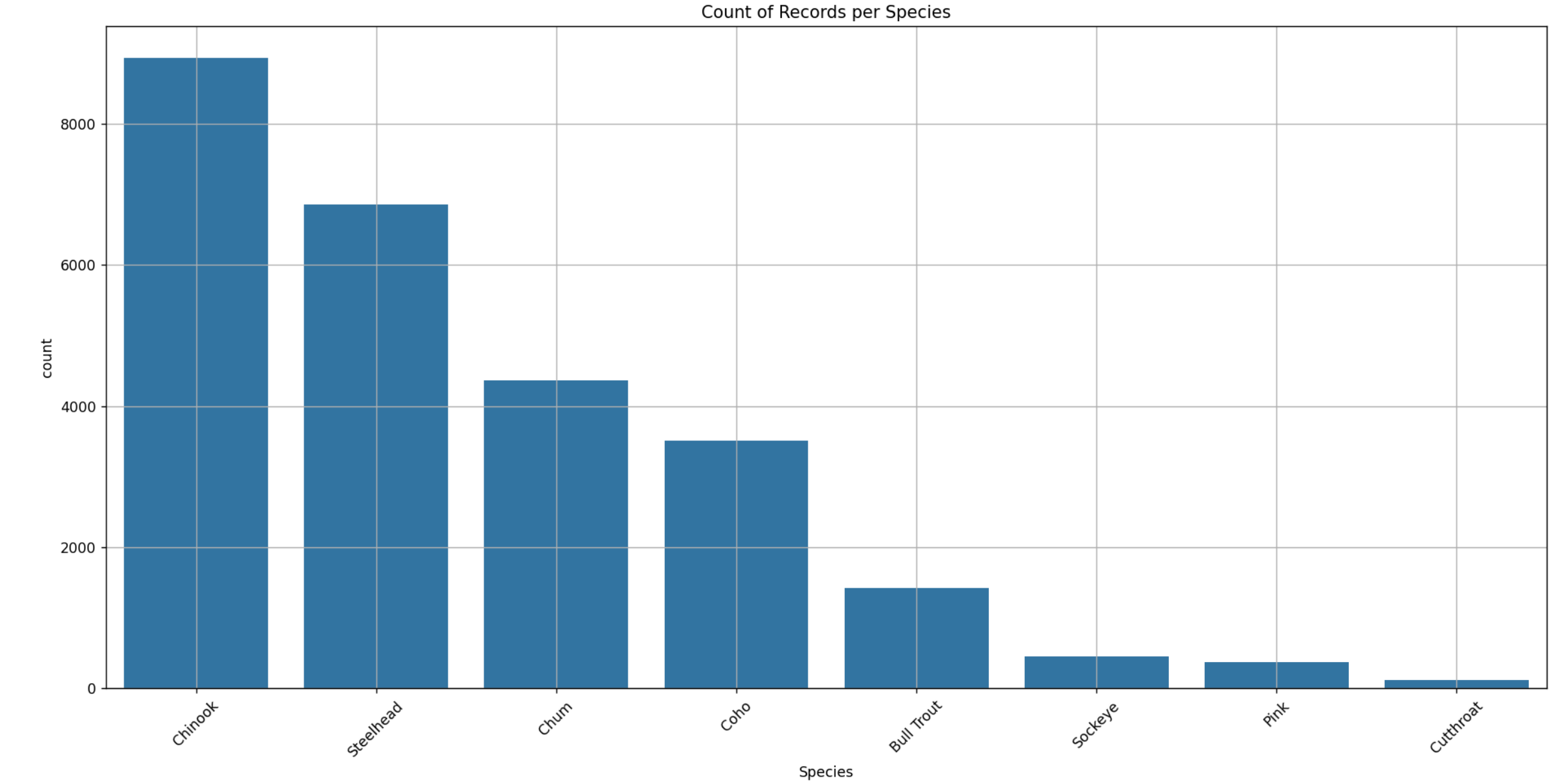

Correlation Heatmap of Numerical Columns
Stock Number
Data Series
Year
Abundance Quantity
1.00
-0.02
0.10
-0.01
-0.02
1.00
0.03
-0.04
0.10
0.03
1.00
-0.03
-0.01
-0.04
-0.03
1.00
Stock Number
Data Series
Year
Abundance Quantity
1.0
0.8
0.6
0.4
0.2
0.0

Figure 1


Count of Records per Species
count
0
2000
4000
6000
8000
Chinook
Steelhead
Chum
Coho
Bull Trout
Sockeye
Pink
Cutthroat
Species

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
 4   Species                              26039 non-null  object
 5   Year                                 25964 non-null  float64
 6   Abundance Quantity                   25139 non-null  float64
 7   Data Type                            25987 non-null  object
 8   Production Type                      18547 non-null  object
 9   Calculation Type                     15932 non-null  object
 10  Escapement Methodology               26039 non-null  object
 11  Escapement Methodology Description   26039 non-null  object
 12  Biologist Methodology Description    25229 non-null  object
 13  Comments                             8220 non-null   object
 14  Report Types                         23187 non-null  object
 15  Last Updated                         26039 non-null  object
dtypes: float64(2), int64(2), object(12)
memory usage: 3.2+ MB
None

Missing Values
Stock Number                              0
Population Name                           0
Sub-Population Name                   21579
Data Series                               0
Species                                   0
Year                                     75
Abundance Quantity                      900
Data Type                                52
Production Type                        7492
Calculation Type                      10107
Escapement Methodology                    0
Escapement Methodology Description        0
Biologist Methodology Description       810
Comments                              17819
Report Types                           2852
Last Updated                              0
dtype: int64

Summary Statistics
        Stock Number  ...                  Last Updated
count   26039.000000  ...                         26039
unique           NaN  ...                          1185
top              NaN  ...  2023-05-15 22:28:54.169738+00
freq             NaN  ...                         15932
mean     3790.731172  ...                           NaN
std      2399.419008  ...                           NaN
min      1008.000000  ...                           NaN
25%      1634.000000  ...                           NaN
50%      2765.000000  ...                           NaN
75%      6553.000000  ...                           NaN
max      8900.000000  ...                           NaN
```

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
[11 rows x 16 columns]

Species Counts
Species
Chinook       8935
Steelhead     6854
Chum          4363
Coho          3514
Bull Trout    1419
Sockeye        455
Pink           373
Cutthroat      126
Name: count, dtype: int64
One or more required columns ('Indicator', 'Species', 'Year', 'Value') are missing.
>>>
========== RESTART: C:/Users/ashut/AppData/Local/Programs/Python/Python313/ca2python.py ==========
Columns: ['Stock Number', 'Population Name', 'Sub-Population Name', 'Data Series', 'Species', 'Year', 'Abundance Quantity', 'Data Type', 'Production Type
', 'Calculation Type', 'Escapement Methodology', 'Escapement Methodology Description', 'Biologist Methodology Description', 'Comments', 'Report Types', '
Last Updated']

Basic Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26039 entries, 0 to 26038
Data columns (total 16 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Stock Number                          26039 non-null  int64
 1   Population Name                       26039 non-null  object
 2   Sub-Population Name                   4460 non-null   object
 3   Data Series                           26039 non-null  int64
 4   Species                               26039 non-null  object
 5   Year                                  25964 non-null  float64
 6   Abundance Quantity                    25139 non-null  float64
 7   Data Type                             25987 non-null  object
 8   Production Type                       18547 non-null  object
 9   Calculation Type                      15932 non-null  object
 10  Escapement Methodology                26039 non-null  object
 11  Escapement Methodology Description    26039 non-null  object
 12  Biologist Methodology Description     25229 non-null  object
 13  Comments                              8220 non-null   object
 14  Report Types                          23187 non-null  object
 15  Last Updated                          26039 non-null  object
dtypes: float64(2), int64(2), object(12)
memory usage: 3.2+ MB
None

Missing Values
```

IDLE Shell 3.13.2

File Edit Shell Debug Options Window Help

```
Missing Values
Stock Number                              0
Population Name                           0
Sub-Population Name                   21579
Data Series                               0
Species                                   0
Year                                     75
Abundance Quantity                      900
Data Type                                52
Production Type                        7492
Calculation Type                      10107
Escapement Methodology                    0
Escapement Methodology Description        0
Biologist Methodology Description       810
Comments                              17819
Report Types                           2852
Last Updated                              0
dtype: int64

Summary Statistics
        Stock Number  ...                  Last Updated
count   26039.000000  ...                         26039
unique           NaN  ...                          1185
top              NaN  ...  2023-05-15 22:28:54.169738+00
freq             NaN  ...                         15932
mean     3790.731172  ...                           NaN
std      2399.419008  ...                           NaN
min      1008.000000  ...                           NaN
25%      1634.000000  ...                           NaN
50%      2765.000000  ...                           NaN
75%      6553.000000  ...                           NaN
max      8900.000000  ...                           NaN

[11 rows x 16 columns]

Species Counts
Species
Chinook       8935
Steelhead     6854
Chum          4363
Coho          3514
Bull Trout    1419
Sockeye        455
Pink           373
Cutthroat      126
Name: count, dtype: int64
One or more required columns ('Indicator', 'Species', 'Year', 'Value') are missing.
>>>
```