

# Data Visualization Lab 2

## Attribute Dataset

Utkarsh Bhangale 20200802124

## Exploratory Data Analysis (EDA)

load the data and perform some initial analysis. Assuming you have the dataset in a CSV file:

In [1]:

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv('Attribute dataset.csv')

# Check the first few rows of the dataset
print(data.head())

# Summary statistics
print(data.describe())

# Data types and missing values
print(data.info())
```

	Dress_ID	Style	Price	Rating	Size	Season	NeckLine	SleeveLength	\
0	1006032852	Sexy	Low	4.6	M	Summer	o-neck	sleevless	
1	1212192089	Casual	Low	0.0	L	Summer	o-neck	Petal	
2	1190380701	vintage	High	0.0	L	Automn	o-neck	full	
3	966005983	Brief	Average	4.6	L	Spring	o-neck	full	
4	876339541	cute	Low	4.5	M	Summer	o-neck	butterfly	

	Material	FabricType	Decoration	Pattern	Type	Recommendation
0	NaN	chiffon	ruffles		animal	1
1	microfiber	NaN	ruffles		animal	0
2	polyester	NaN	NaN		print	0
3	silk	chiffon	embroidary		print	1
4	chiffonfabric	chiffon	bow		dot	0

	Dress_ID	Rating	Recommendation
count	4.790000e+02	479.000000	479.000000
mean	9.022420e+08	3.489353	0.427975
std	1.822352e+08	2.028941	0.495302
min	1.234568e+08	0.000000	0.000000
25%	7.666611e+08	2.000000	0.000000
50%	9.096250e+08	4.600000	0.000000
75%	1.039684e+09	4.800000	1.000000
max	1.253973e+09	5.000000	1.000000

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 479 entries, 0 to 478

Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	Dress_ID	479 non-null	int64
1	Style	479 non-null	object
2	Price	477 non-null	object
3	Rating	479 non-null	float64
4	Size	479 non-null	object
5	Season	477 non-null	object
6	NeckLine	476 non-null	object
7	SleeveLength	477 non-null	object
8	Material	360 non-null	object
9	FabricType	223 non-null	object

```
9  FabricType      225 non-null    object
10  Decoration      255 non-null    object
11  Pattern Type    377 non-null    object
12  Recommendation  479 non-null    int64
dtypes: float64(1), int64(2), object(10)
memory usage: 48.8+ KB
None
```

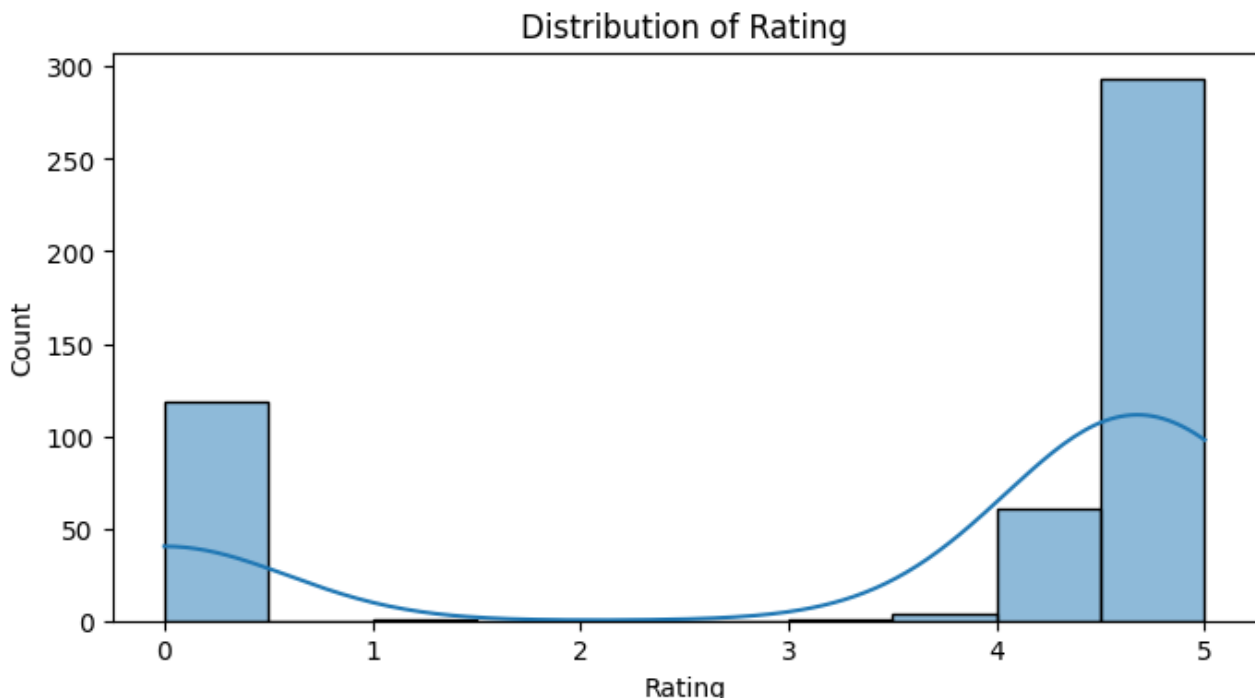
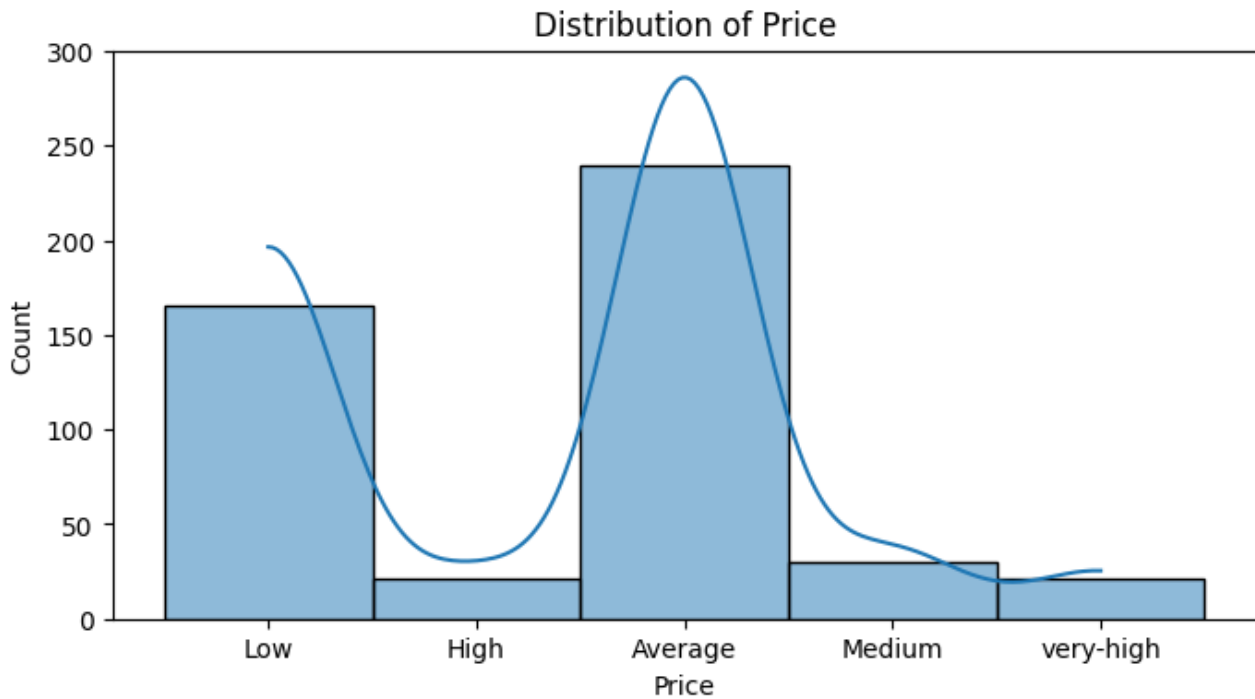
## Univariate Analysis

For univariate analysis, let's analyze individual variables in more detail. Here's an example of how you can create histograms and other visualizations for each numerical attribute:

In [2]:

```
# Univariate Analysis for numerical attributes
num_attributes = ['Price', 'Rating']

for attribute in num_attributes:
    plt.figure(figsize=(8, 4))
    sns.histplot(data=data, x=attribute, kde=True)
    plt.title(f'Distribution of {attribute}')
    plt.show()
```

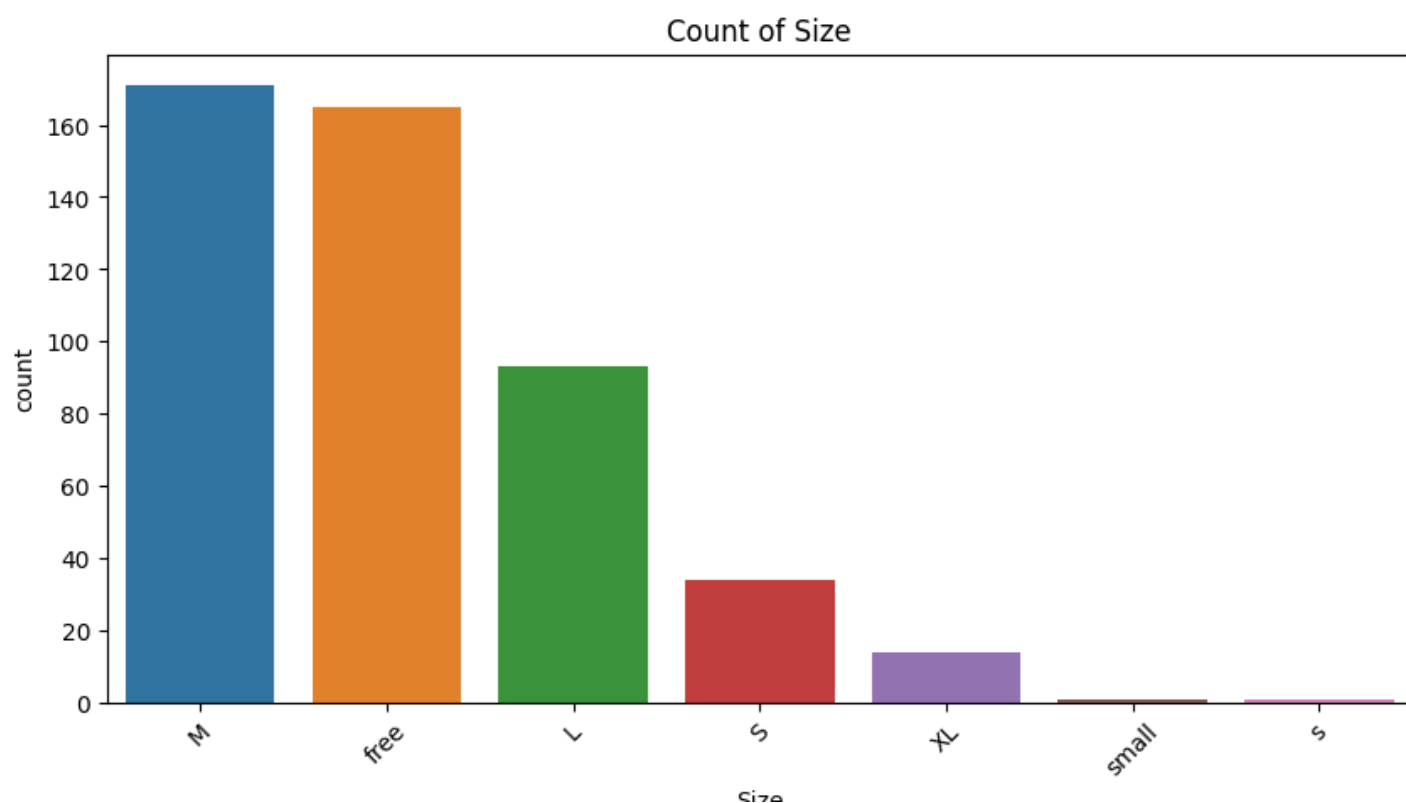
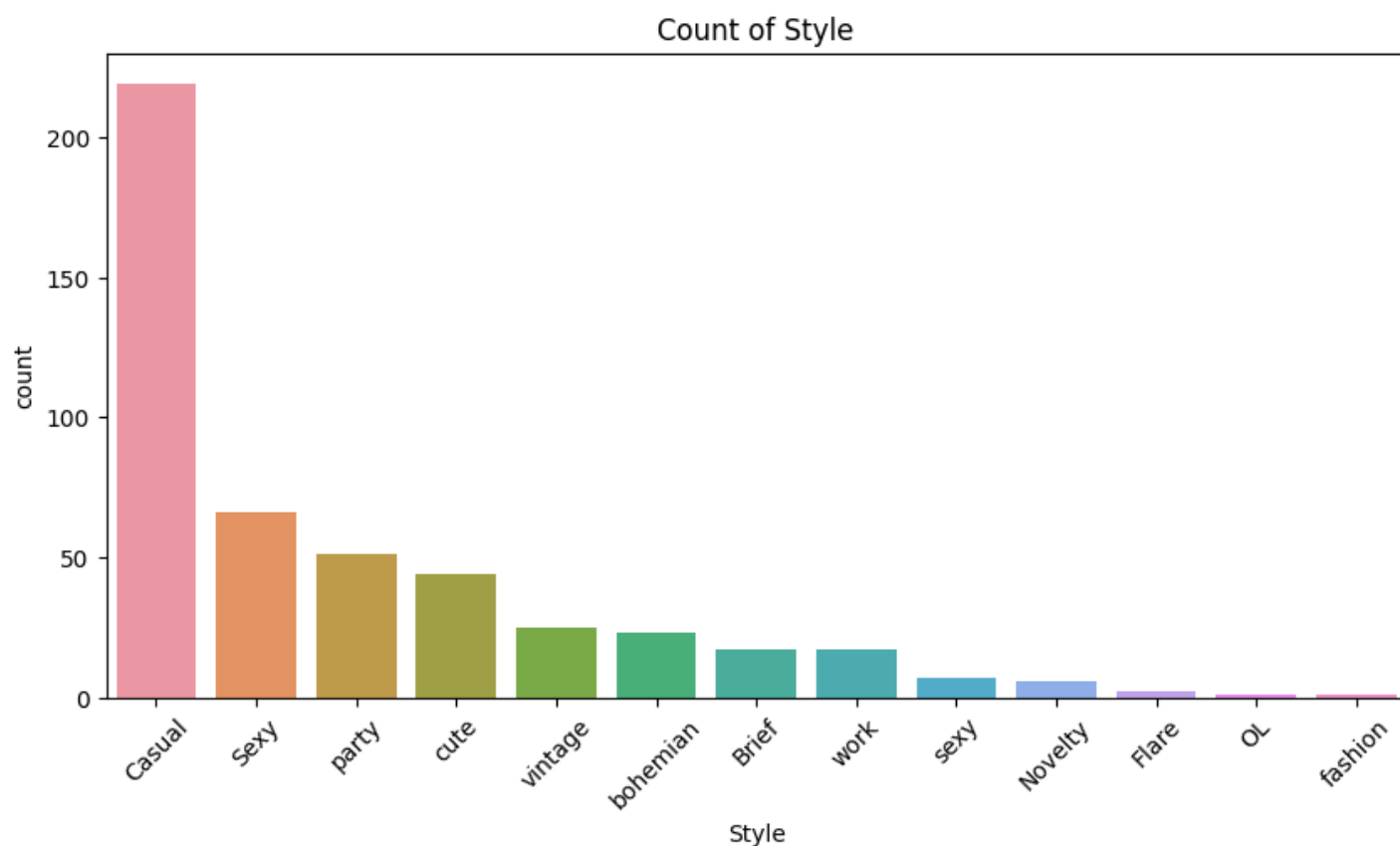


For categorical attributes, you can create bar plots to visualize the distribution:

In [3]:

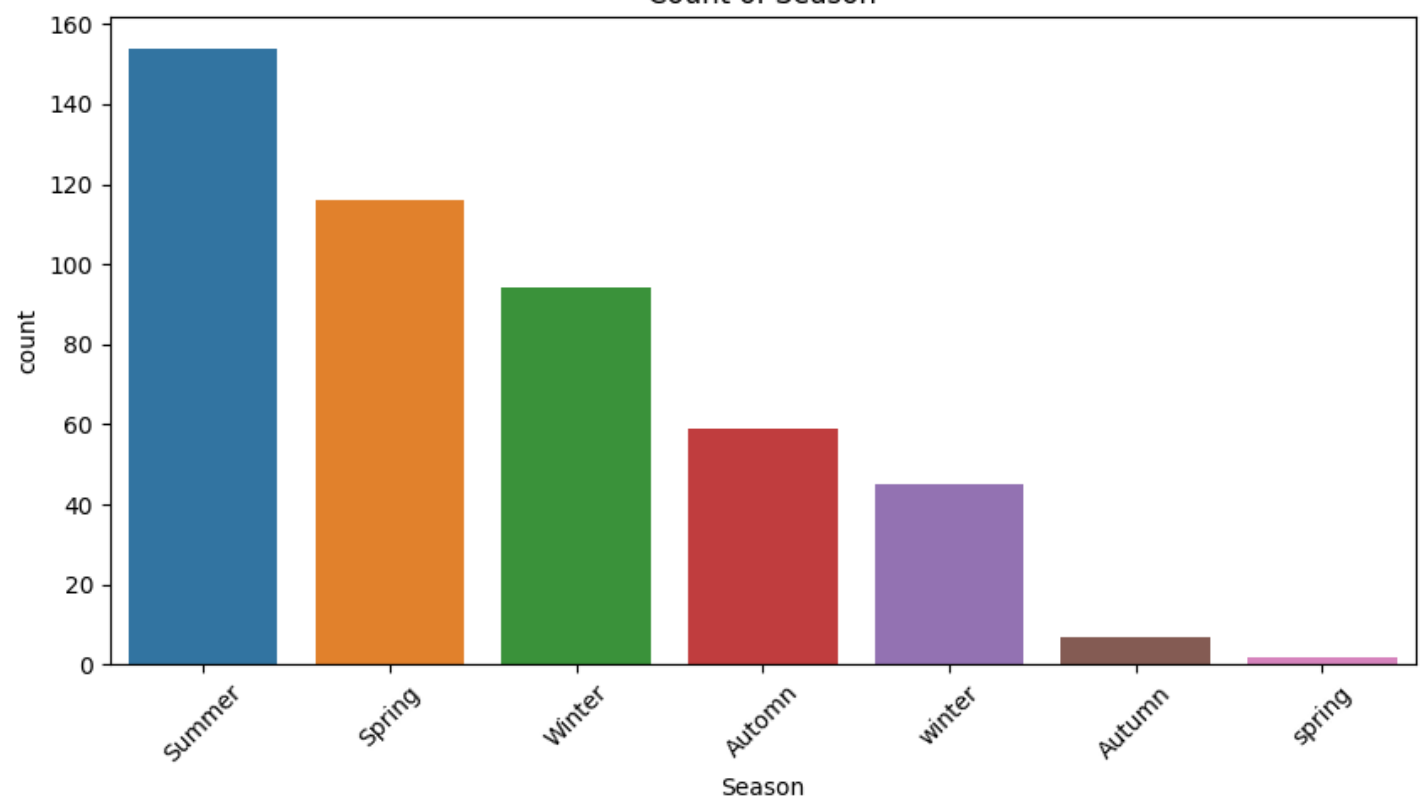
```
# Univariate Analysis for categorical attributes
cat_attributes = ['Style', 'Size', 'Season', 'NeckLine', 'SleeveLength', 'Material', 'FabricType', 'Decoration', 'Pattern Type']

for attribute in cat_attributes:
    plt.figure(figsize=(10, 5))
    sns.countplot(data=data, x=attribute, order=data[attribute].value_counts().index)
    plt.xticks(rotation=45)
    plt.title(f'Count of {attribute}')
    plt.show()
```

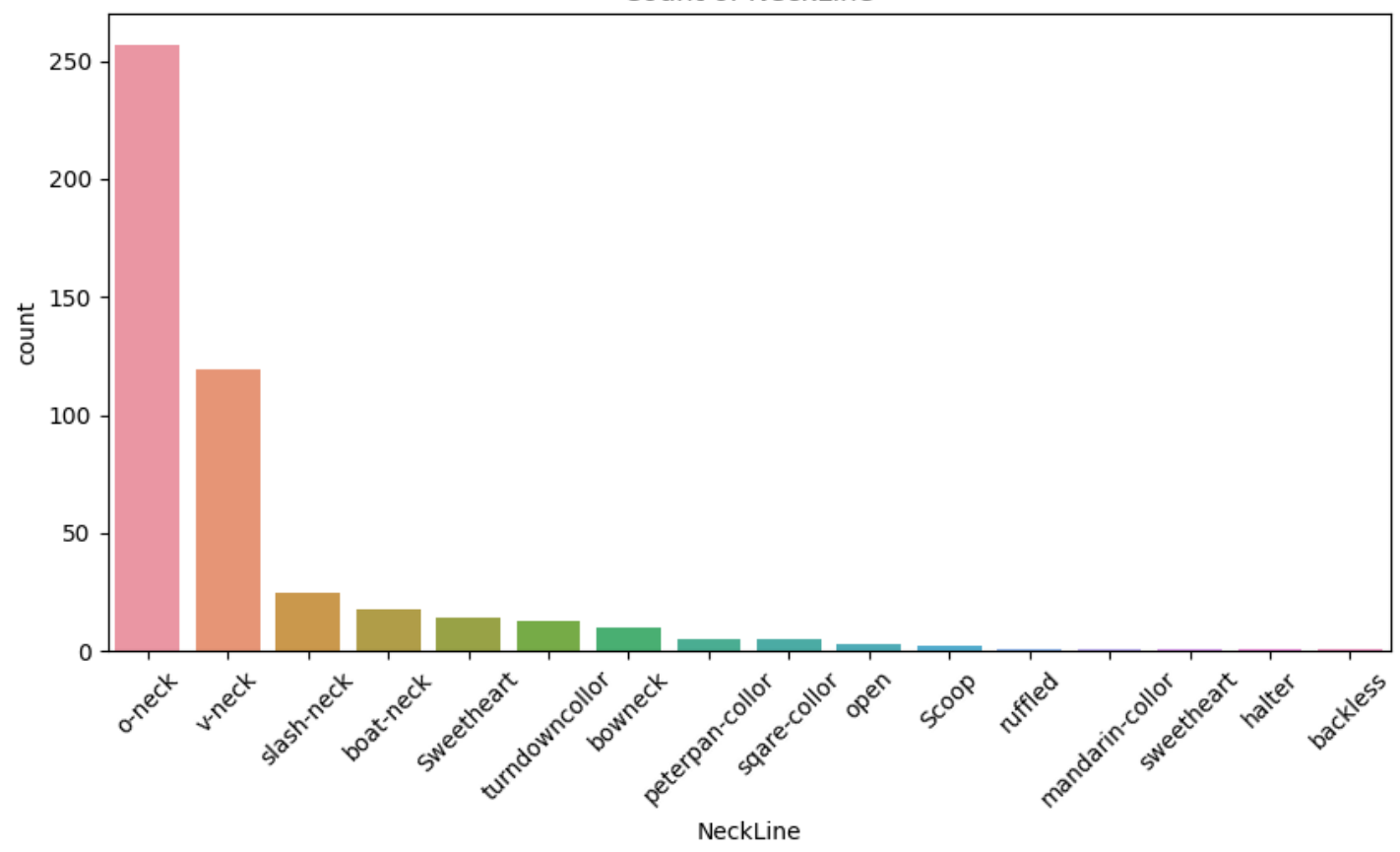


size

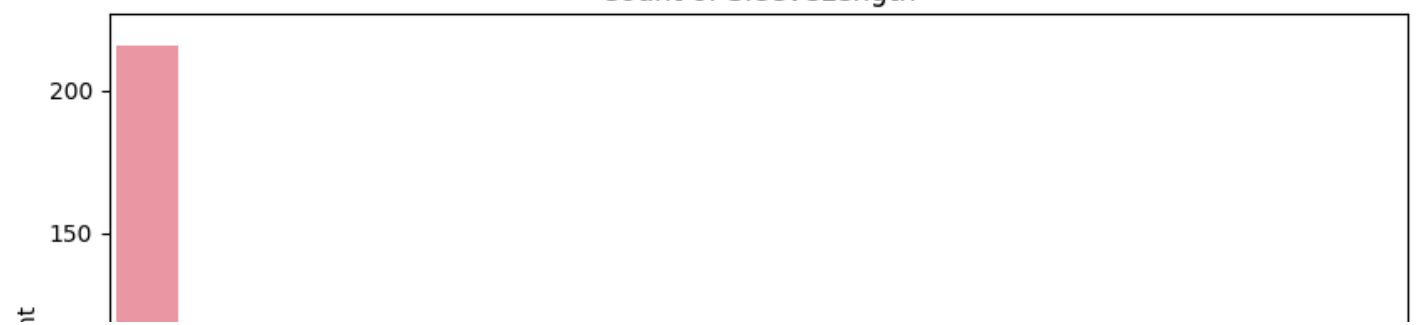
Count of Season

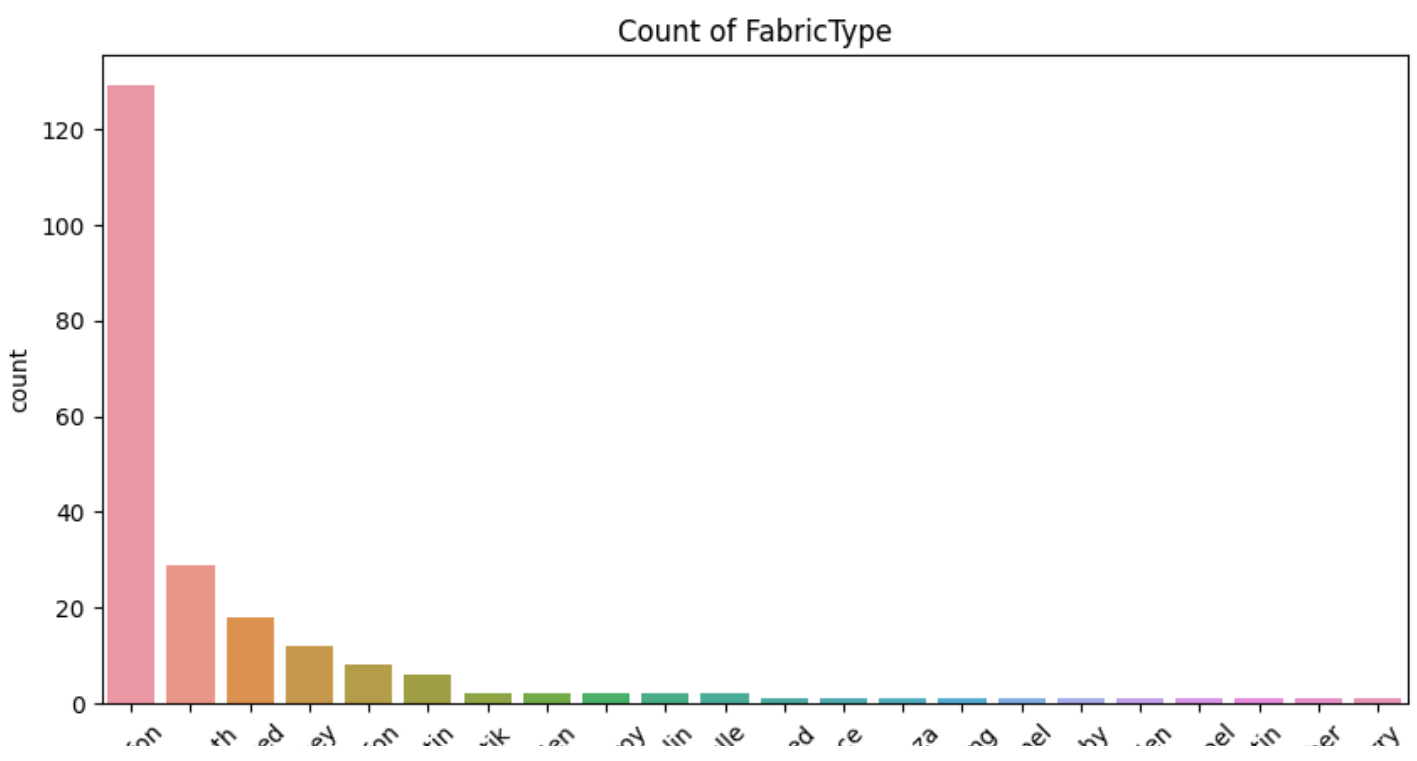
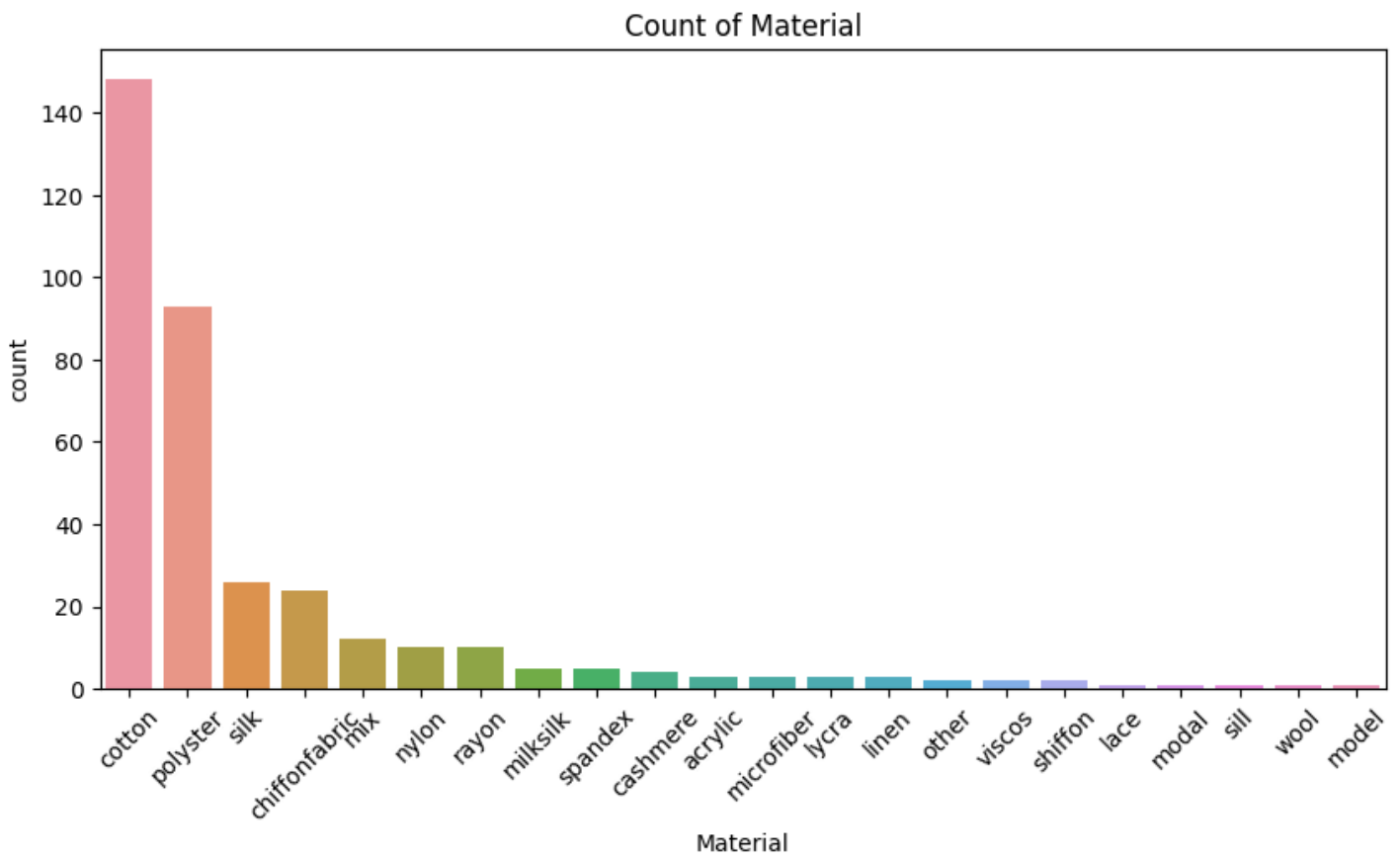
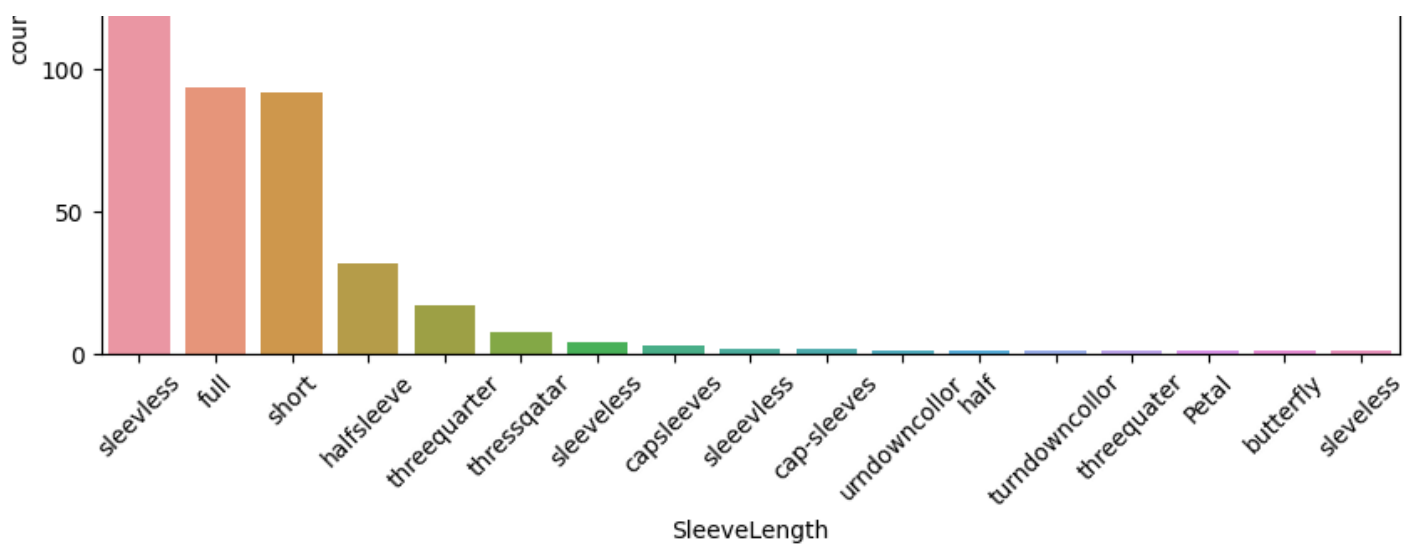


Count of NeckLine



Count of SleeveLength

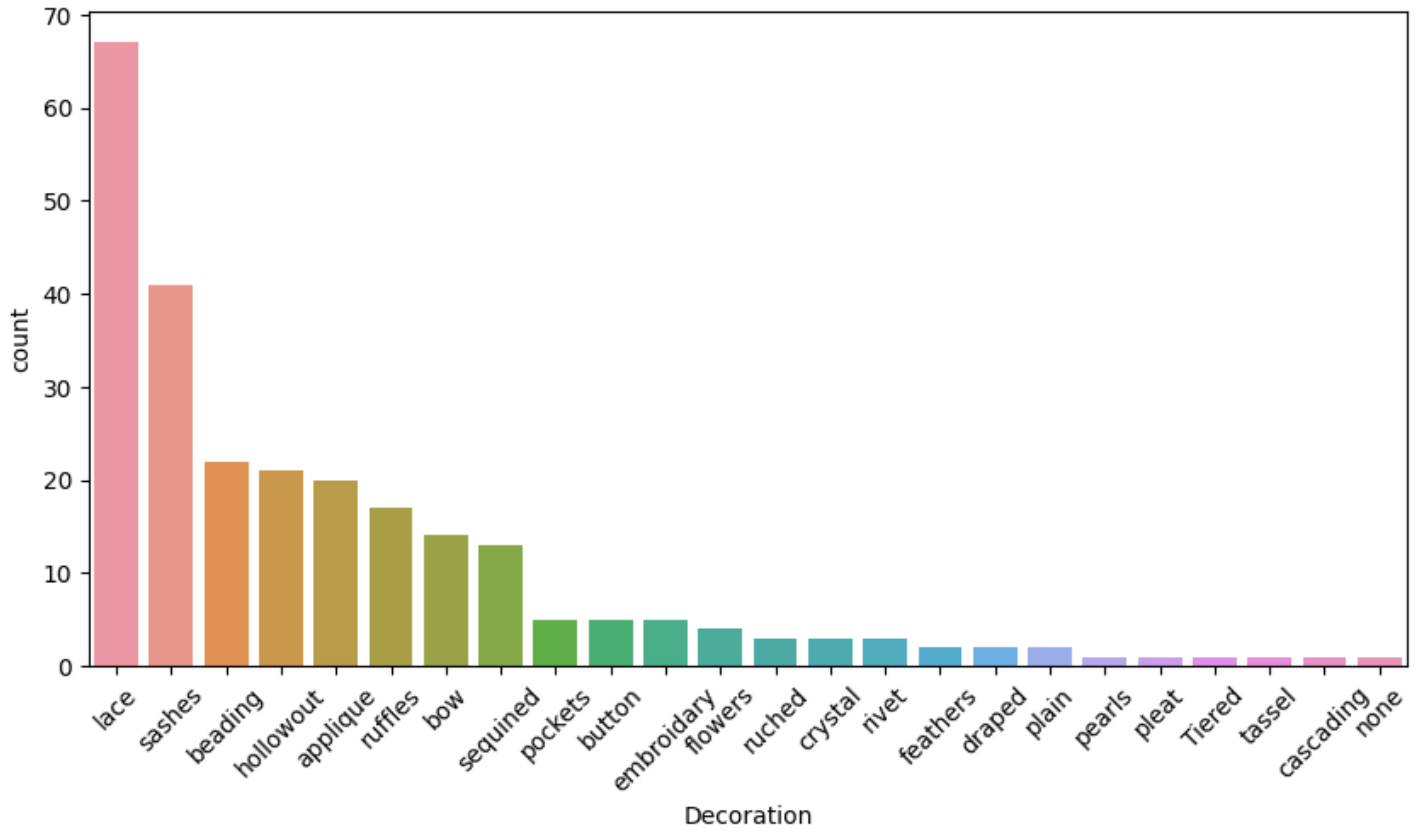




chiffon  
broaddcloth  
worsted  
jersey  
shift  
satin  
batiste  
wool  
Corduroy  
poplin  
tulle  
knitted  
lace  
organza  
knitting  
flannel  
double  
wool  
flannel  
satin  
other  
ten...

FabricType

Count of Decoration



Count of Pattern Type

