In [1]:

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

In [5]:

```
visa_data = pd.read_csv("visa.csv")
```

In [6]:

```
visa_data.head()
```

Out[6]:

	Visa_Class	Employer_Name	SOC_Title	Job_Title	Full_Time_Position	Worksite
0	H-1B	Hexaware Technologies Inc.	Software Developers, Applications	Senior Technical Architect	Y	Herndon,, Virginia
1	H-1B	WIPRO LIMITED	Computer Programmers	Programmer Analyst\t	Y	Texas City, Texas
2	H-1B	Mastech Digital InfoTech, Inc.	Software Developers, Applications	Software Developer	Y	Strongsville, Ohio
3	H-1B	VIRTUSA CORPORATION	Computer Systems Analysts	JAVA ANALYST 2	Y	Hartford, Connecticut
4	H-1B	XTGLOBAL, INC.	Database Administrators	SQL DATABASE ADMINISTRATOR II	Y	Austin, Texas

In [7]:

```
visa_data.tail()
```

Out[7]:

	Visa_Class	Employer_Name	SOC_Title	Job_Title	Full_Time_Position	Wor
826300	H-1B	ESCOBEDO CONSTRUCTION, LP	Software Developers, Systems Software	SOFTWARE DEVELOPMENT	Y	Buda, T
826301	H-1B	The University of Texas at Dallas	Business Teachers, Postsecondary	Assistant Professor of Instruction	Y	Richard T
826302	H-1B	NIC INFO TEK INC	Software Developers, Applications	Engineering Lead	Y	Nolens Tenne
826303	H-1B	NIC INFO TEK INC	Software Developers, Applications	Engineering Lead	Y	Nolens Tenne
826304	H-1B	CompassBeauty, Inc.	Computer And Information Systems Managers	Vice President, Engineering	Y	Franc Calif

In [8]:

```
visa_data.shape
```

Out[8]:

```
(826305, 12)
```

In [9]:

```
visa_data.columns
```

Out[9]:

```
Index(['Visa_Class', 'Employer_Name', 'SOC_Title', 'Job_Title',
       'Full_Time_Position', 'Worksite', 'Prevailing_Wage', 'Unit_Of_Pay',
       'Employer_Location', 'Employer_Country', 'Case_Status', 'Quarter'],
      dtype='object')
```

In [10]:

```
visa_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 826305 entries, 0 to 826304
Data columns (total 12 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   Visa_Class          826305 non-null  object 
 1   Employer_Name       826303 non-null  object 
 2   SOC_Title           826303 non-null  object 
 3   Job_Title           826305 non-null  object 
 4   Full_Time_Position  826305 non-null  object 
 5   Worksite            826251 non-null  object 
 6   Prevailing_Wage     826304 non-null  float64
 7   Unit_Of_Pay         826304 non-null  object 
 8   Employer_Location   826241 non-null  object 
 9   Employer_Country    826305 non-null  object 
 10  Case_Status         826305 non-null  object 
 11  Quarter             826305 non-null  object 
dtypes: float64(1), object(11)
memory usage: 75.7+ MB
```

In [11]:

```
visa_data.describe()
```

Out[11]:

	Prevailing_Wage
count	826304.00000
mean	94172.56927
std	40046.33134
min	7.25000
25%	74006.00000
50%	93538.00000
75%	117125.00000
max	431897.00000

In [12]:

```
visa_data.isnull().sum()
```

Out[12]:

```
Visa_Class             0
Employer_Name          2
SOC_Title              2
Job_Title              0
Full_Time_Position     0
Worksite              54
Prevailing_Wage        1
Unit_Of_Pay            1
Employer_Location     64
Employer_Country       0
Case_Status            0
Quarter                0
dtype: int64
```

In [13]:

```
visa_data.dropna(inplace = True)
```

In [14]:

```
visa_data.isnull().sum()
```

Out[14]:

```
Visa_Class            0
Employer_Name         0
SOC_Title             0
Job_Title             0
Full_Time_Position    0
Worksite              0
Prevailing_Wage       0
Unit_Of_Pay           0
Employer_Location     0
Employer_Country      0
Case_Status           0
Quarter               0
dtype: int64
```

In [15]:

```
visa_data['Visa_Class'].unique()
```

Out[15]:

```
array(['H-1B', 'E-3 Australian', 'H-1B1 Singapore', 'H-1B1 Chile'],
      dtype=object)
```

In [16]:

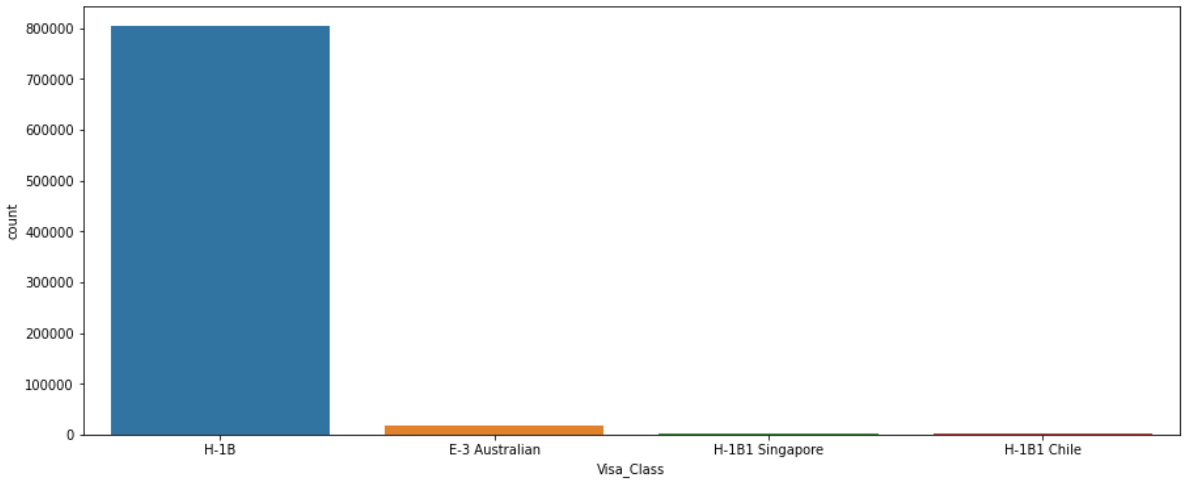
```
visa_data['Visa_Class'].value_counts()
```

Out[16]:

```
H-1B               803613
E-3 Australian      18308
H-1B1 Chile          2314
H-1B1 Singapore      1948
Name: Visa_Class, dtype: int64
```

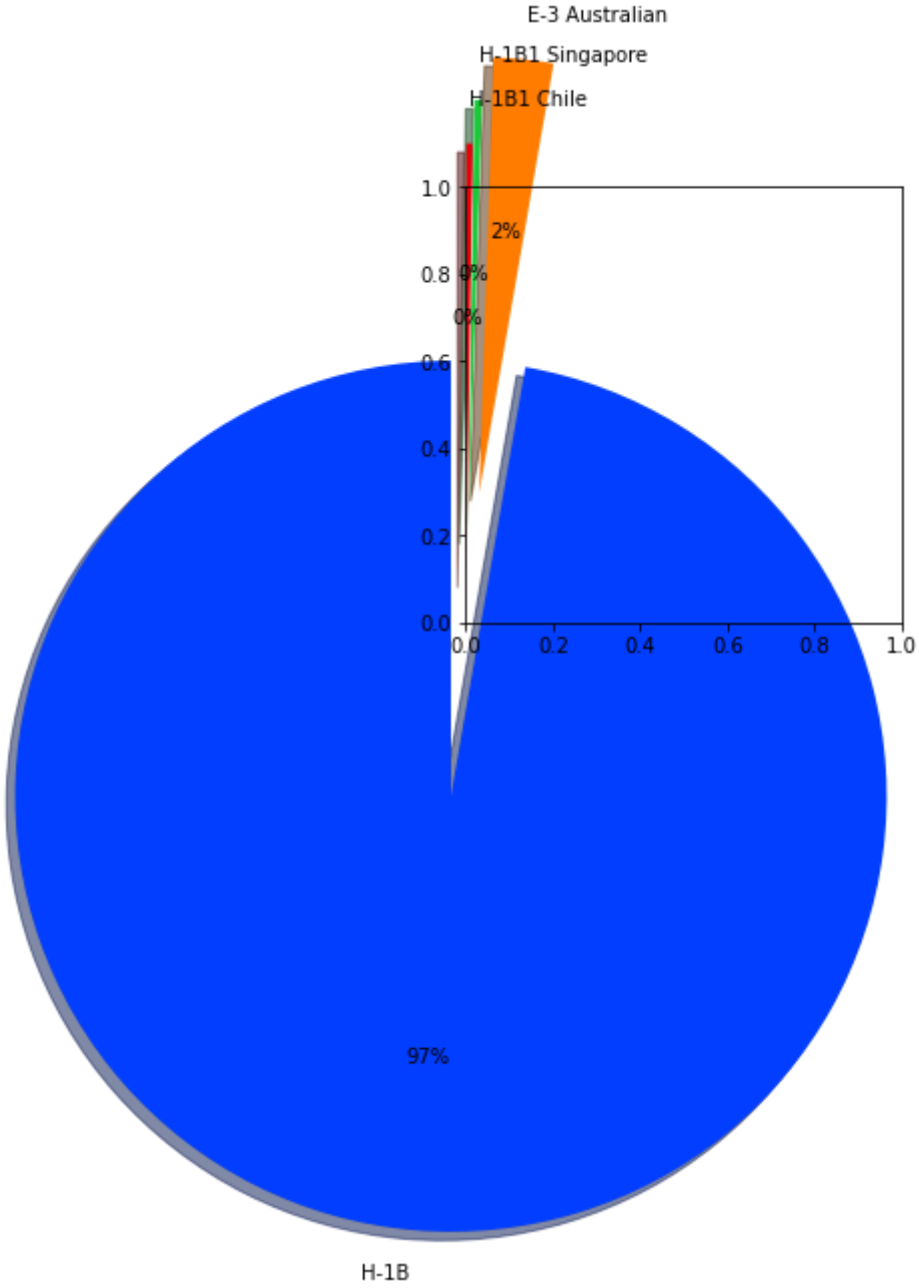
In [18]:

```
plt.figure(figsize=(15,6))
sns.countplot('Visa_Class', data = visa_data)
plt.xticks(rotation = 0)
plt.show()
```



In [21]:

```
explode = [0.4,0.3,0.2, 0.1]
colors = sns.color_palette('bright')
plt.pie(visa_data['Visa_Class'].value_counts(), labels=['H-1B', 'E-3 Australian', 'H-1B1
        colors = colors, autopct = '%0.0f%%',  explode = explode, shadow = 'True',
        startangle = 90, frame = 'true')
plt.show()
```



In [22]:

```
visa_data['Employer_Name'].unique()
```

Out[22]:

```
array(['Hexaware Technologies Inc.', 'WIPRO LIMITED',
       'Mastech Digital InfoTech, Inc.', ...,
       ' Signature Diagnostics, Inc', 'Signature Diagnostics, Inc ',
       'Chivico Corp'], dtype=object)
```

In [23]:

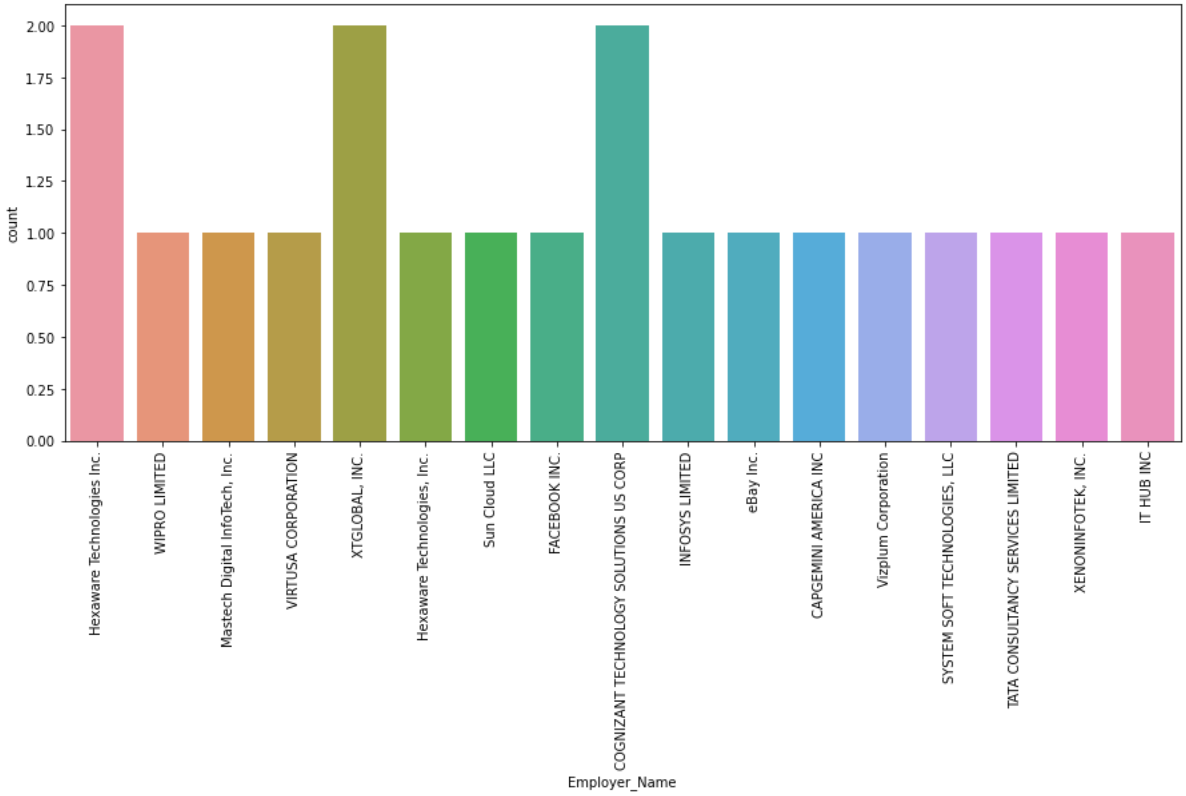
```
visa_data['Employer_Name'].value_counts()
```

Out[23]:

```
COGNIZANT TECHNOLOGY SOLUTIONS US CORP    19525
Google LLC                                15840
Microsoft Corporation                     11763
Ernst & Young U.S. LLP                    10281
TATA CONSULTANCY SERVICES LIMITED          8800
                                          ...
Chimera Investment Corporation                1
Patio Paradise, Inc.                          1
Flexaworld LP                                 1
Quinn Racusin & Gazzola Chartered             1
Chivico Corp                                  1
Name: Employer_Name, Length: 63610, dtype: int64
```

In [36]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Employer_Name'].head(20), data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [32]:

```
visa_data['SOC_Title'].unique()
```

Out[32]:

```
array(['Software Developers, Applications', 'Computer Programmers',
       'Computer Systems Analysts', 'Database Administrators',
       'Information Technology Project Managers',
       'Computer Systems Engineers/Architects',
       'Search Marketing Strategists', 'Computer Network Architects',
       'Electronics Engineers, Except Computer', 'Mechanical Engineers',
       'Software Developers, Systems Software',
       'Computer Occupations, All Other', 'Aerospace Engineers',
       'Management Analysts', 'Information Security Analysts',
       'Web Developers', 'Network And Computer Systems Administrators',
       'Software Quality Assurance Engineers And Testers',
       'Computer And Information Systems Managers',
       'Human Resources Specialists', 'Operations Research Analysts',
       'Business Intelligence Analysts', 'Sales Engineers',
       'Data Warehousing Specialists', 'Database Architects',
       'Human Resources Managers', 'Commercial And Industrial Designer
s',
```

In [33]:

```
visa_data['SOC_Title'].value_counts()
```

Out[33]:

```
Software Developers, Applications                     265674
Software Developers, Systems Software                  51574
Computer Systems Analysts                              47961
Computer Systems Engineers/Architects                  25899
Software Quality Assurance Engineers And Testers       20532
                                                       ...
Traffic Technicians                                        1
Data Scientists                                            1
Software Quality And Assurance Analysts And Testers        1
Medical Secretaries                                        1
Industrial Safety And Health Engineers.                    1
Name: SOC_Title, Length: 829, dtype: int64
```

In [37]:

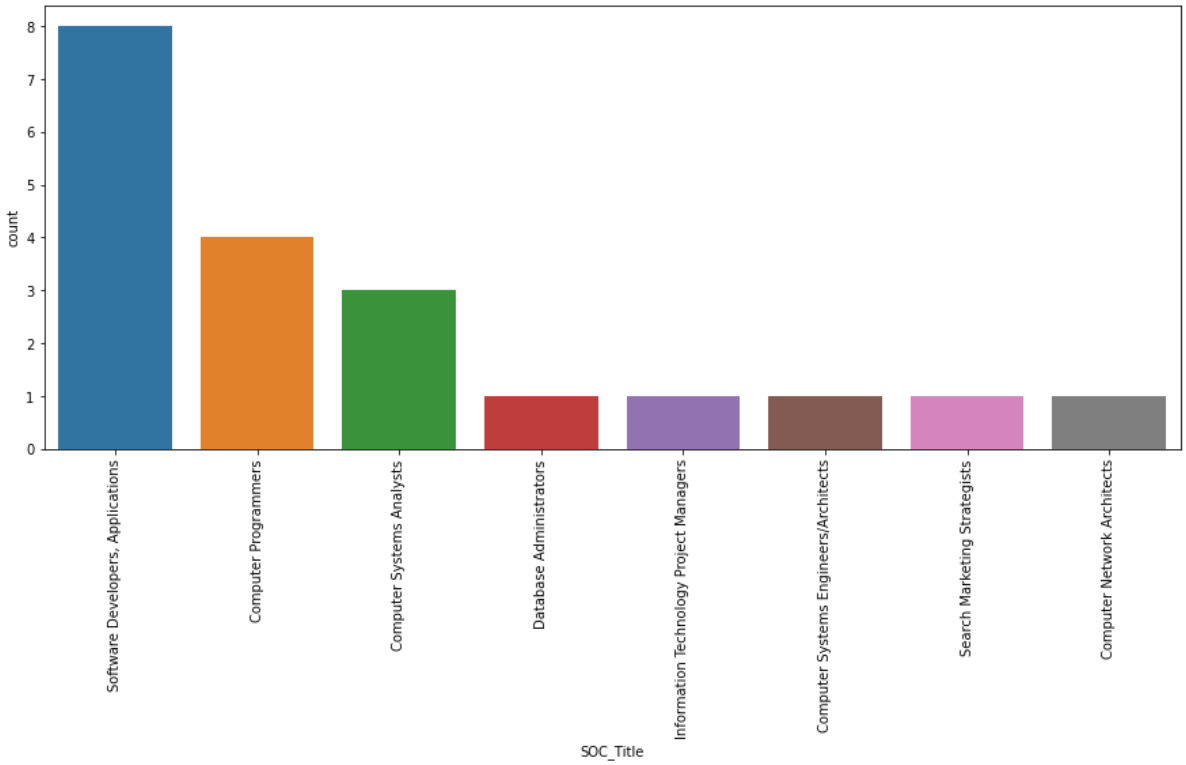
```
visa_data['SOC_Title'].value_counts(ascending = False).head(10)
```

Out[37]:

```
Software Developers, Applications                    265674
Software Developers, Systems Software                 51574
Computer Systems Analysts                             47961
Computer Systems Engineers/Architects                 25899
Software Quality Assurance Engineers And Testers      20532
Information Technology Project Managers               18101
Computer And Information Systems Managers             17479
Mechanical Engineers                                  15227
Business Intelligence Analysts                        15193
Operations Research Analysts                          13509
Name: SOC_Title, dtype: int64
```

In [40]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['SOC_Title'].head(20), data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [41]:

```
visa_data['Job_Title'].unique()
```

Out[41]:

```
array(['Senior Technical Architect', 'Programmer Analyst\t',
       'Software Developer', ..., 'R&D Controls Engineer',
       'SOFTWARE CONFIGURATION ANALYST', 'Senior Associate Producer'],
      dtype=object)
```

In [42]:

```
visa_data['Job_Title'].value_counts()
```

Out[42]:

```
Software Engineer                                  32986
Software Developer                                 23059
Senior Software Engineer                           10535
SOFTWARE DEVELOPER                                  8530
SOFTWARE ENGINEER                                   8425
                                                   ...
Sr. Mgr Data Governance                                1
Manager - IT EAO Digital Technologies                  1
Head of Platform Services                              1
Associate Vice President- Information Security         1
Senior Associate Producer                              1
Name: Job_Title, Length: 125923, dtype: int64
```

In [43]:

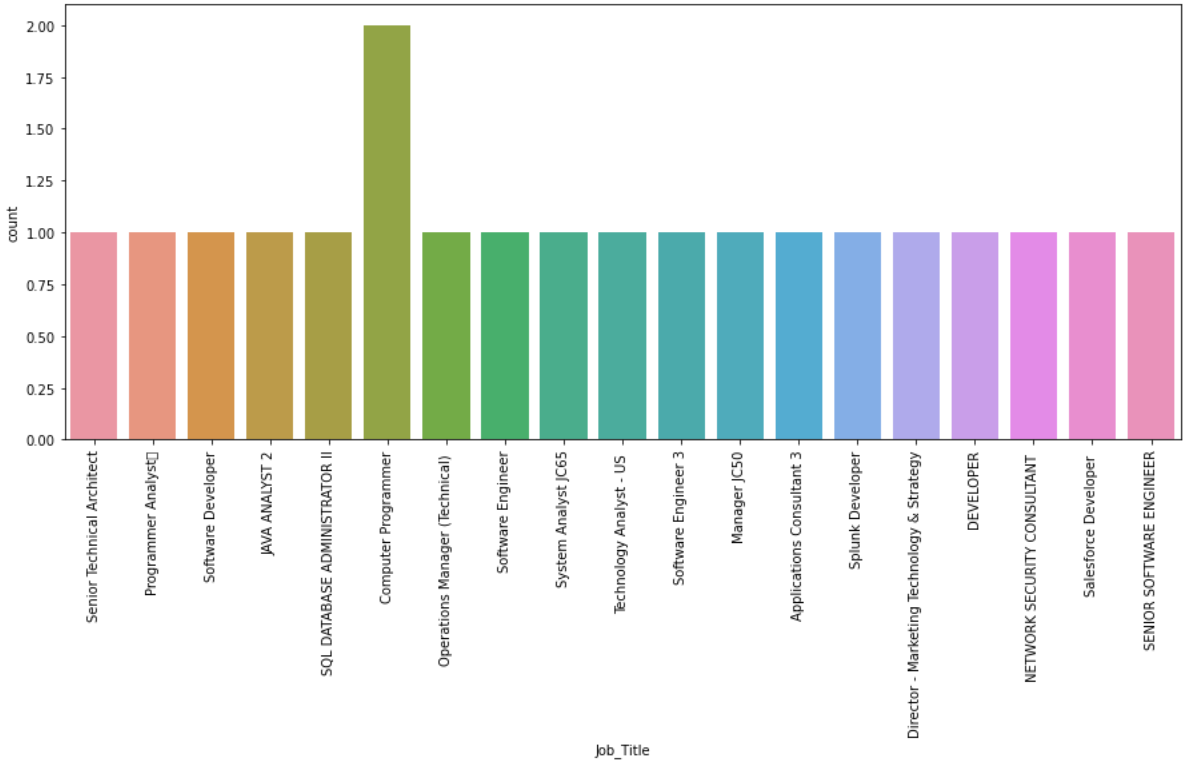
```
visa_data['Job_Title'].value_counts(ascending = False).head(10)
```

Out[43]:

```
Software Engineer              32986
Software Developer             23059
Senior Software Engineer       10535
SOFTWARE DEVELOPER              8530
SOFTWARE ENGINEER               8425
Senior Systems Analyst JC60     7067
Manager JC50                    6722
Assistant Professor             6108
Associate                       3486
Data Scientist                  3341
Name: Job_Title, dtype: int64
```

In [44]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Job_Title'].head(20), data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [45]:

```
visa_data['Full_Time_Position'].unique()
```

Out[45]:

```
array(['Y', 'N'], dtype=object)
```

In [46]:

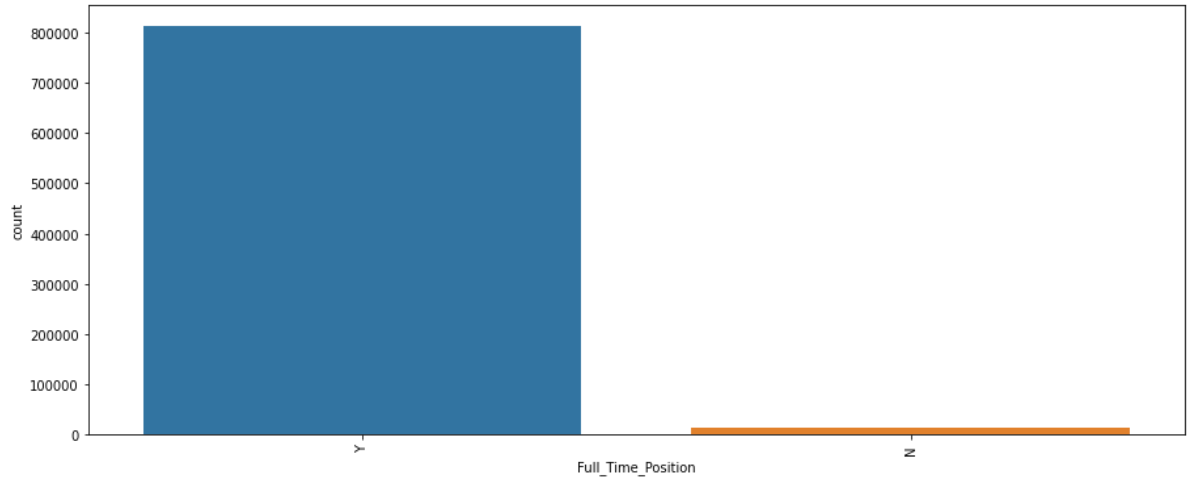
```
visa_data['Full_Time_Position'].value_counts()
```

Out[46]:

```
Y    813095
N     13088
Name: Full_Time_Position, dtype: int64
```

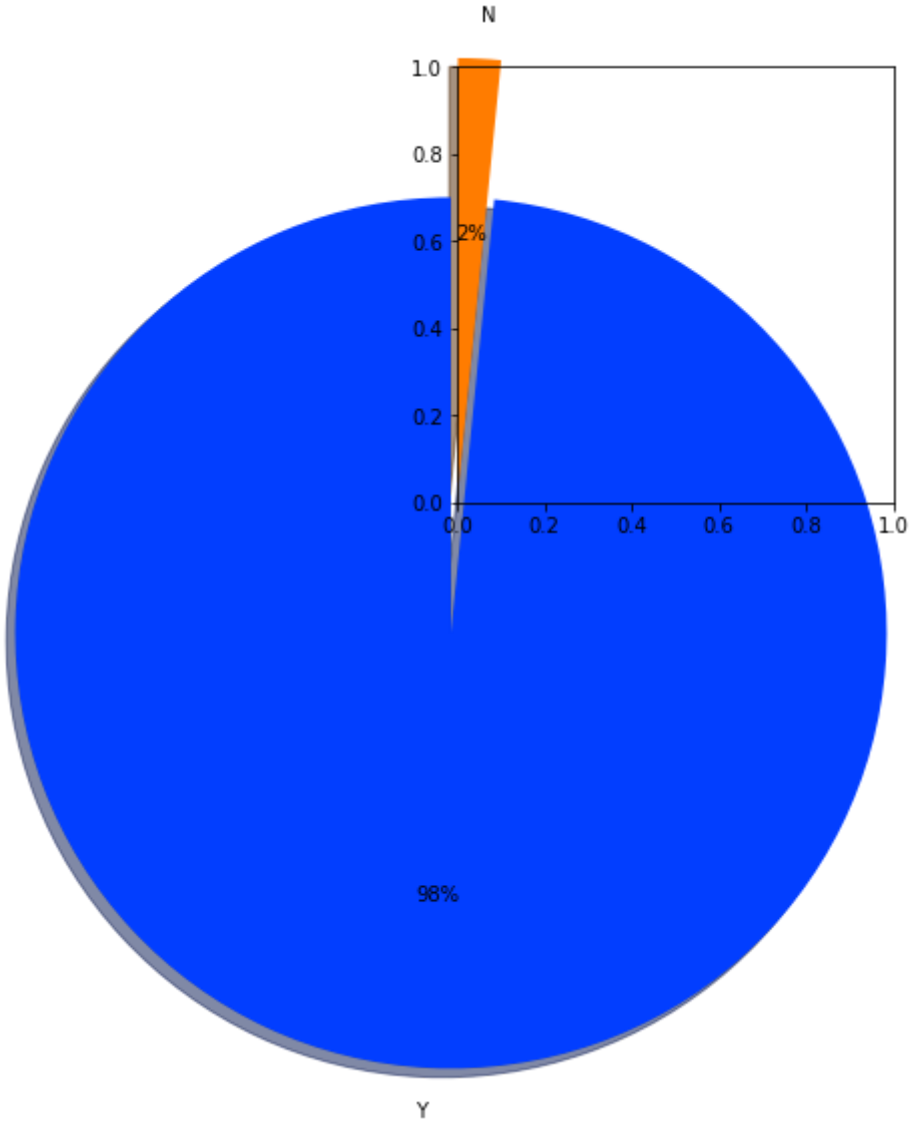
In [47]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Full_Time_Position'], data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [48]:

```
explode = [0.3, 0.02]
colors = sns.color_palette('bright')
plt.pie(visa_data['Full_Time_Position'].value_counts(), labels=['Y', 'N'],
        colors = colors, autopct = '%0.0f%%',  explode = explode, shadow = 'True',
        startangle = 90, frame = 'true')
plt.show()
```



In [49]:

```
visa_data['Case_Status'].unique()
```

Out[49]:

```
array(['Certified', 'Certified - Withdrawn', 'Denied', 'Withdrawn'],
      dtype=object)
```

In [50]:

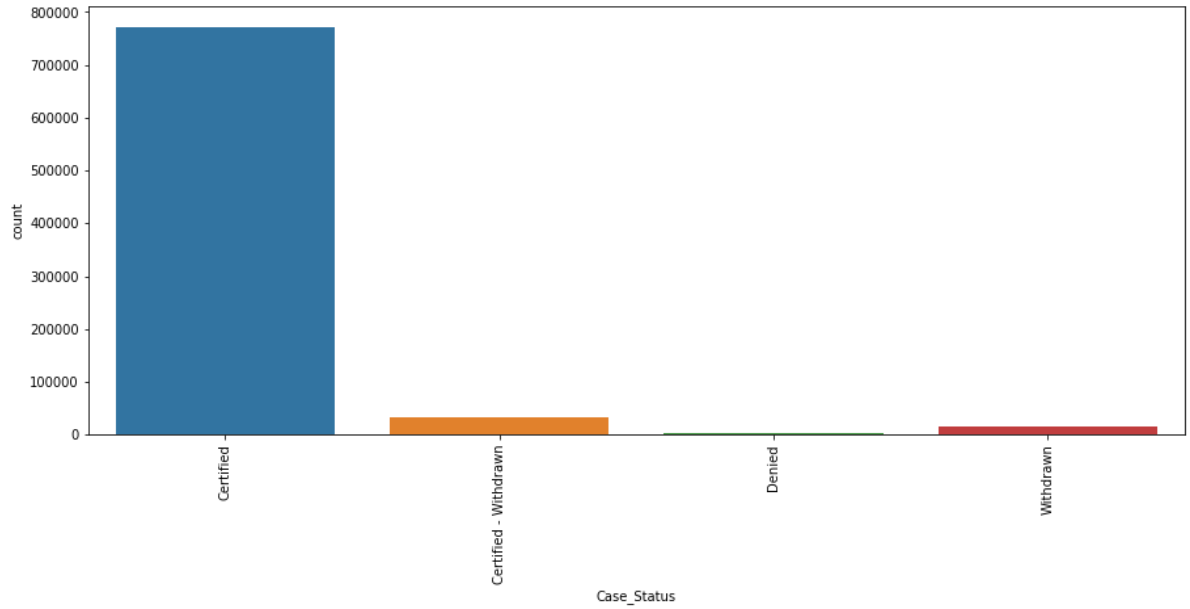
```
visa_data['Case_Status'].value_counts()
```

Out[50]:

```
Certified                772260
Certified - Withdrawn     33616
Withdrawn                 15942
Denied                     4365
Name: Case_Status, dtype: int64
```

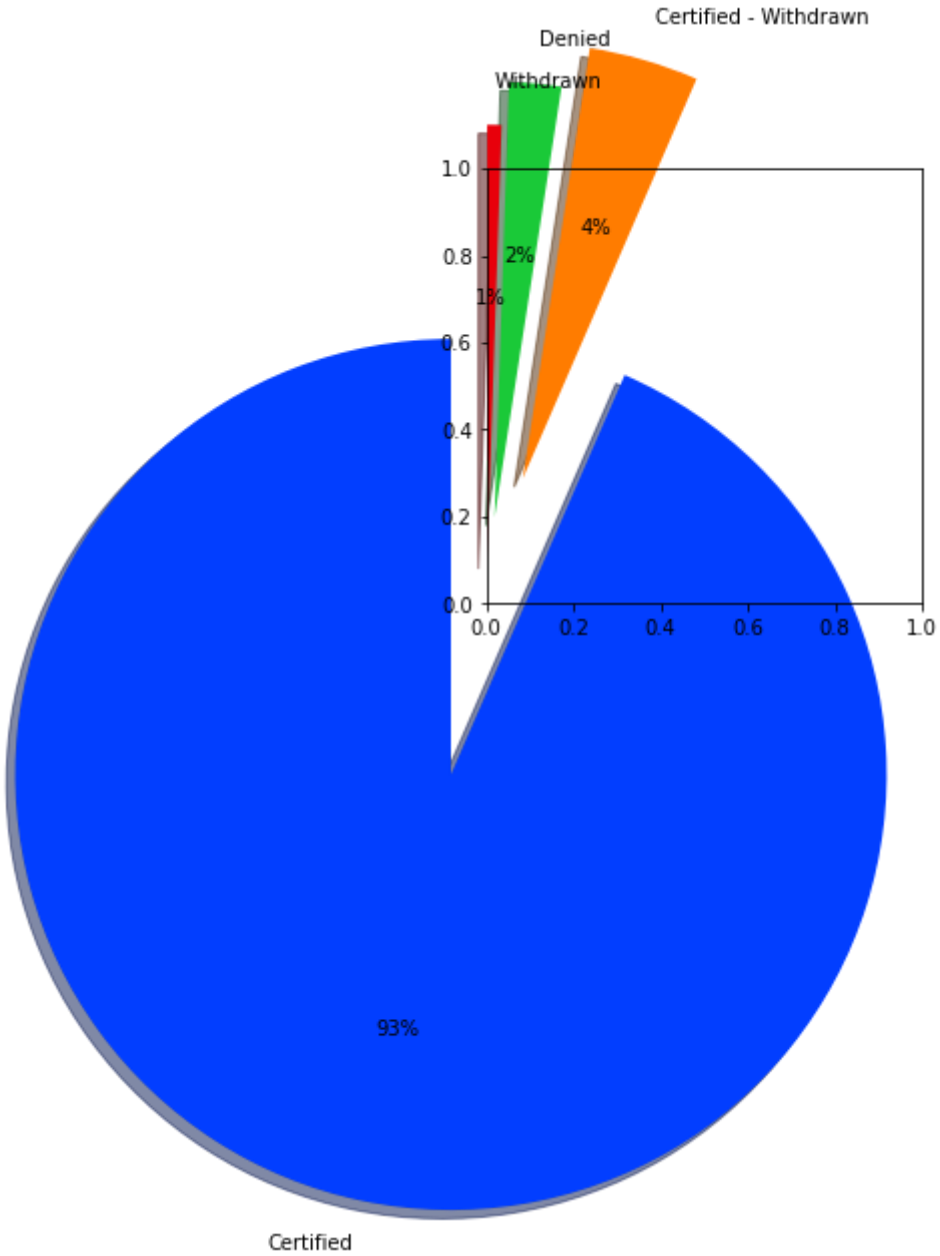
In [51]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Case_Status'], data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



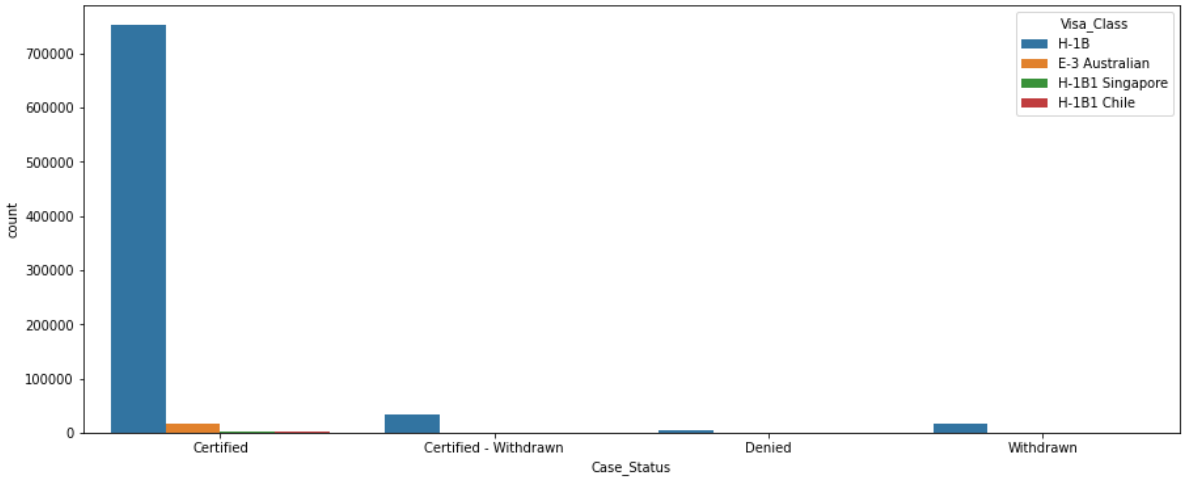
In [54]:

```
explode = [0.4, 0.3, 0.2, 0.1]
colors = sns.color_palette('bright')
plt.pie(visa_data['Case_Status'].value_counts(), labels= ['Certified', 'Certified - Wit
        colors = colors, autopct = '%0.0f%%',  explode = explode, shadow = 'True',
        startangle = 90, frame = 'true')
plt.show()
```



In [52]:

```
plt.figure(figsize=(15,6))
sns.countplot(hue = 'Visa_Class', x = 'Case_Status', data = visa_data )
plt.xticks(rotation = 0)
plt.show()
```



In [55]:

```
visa_data['Unit_Of_Pay'].unique()
```

Out[55]:

```
array(['Year', 'Hour', 'Bi-Weekly', 'Month', 'Week'], dtype=object)
```

In [56]:

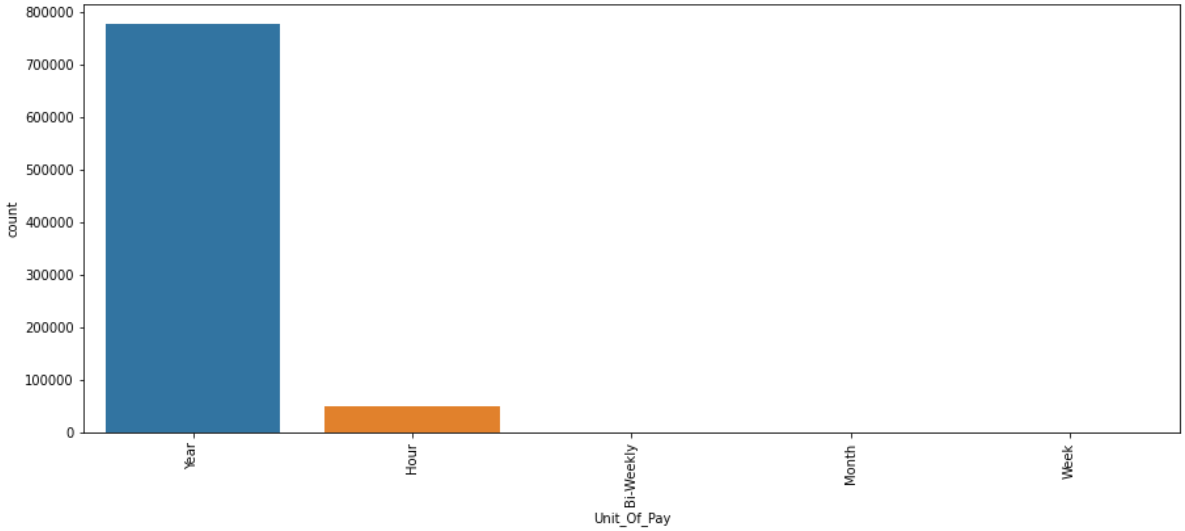
```
visa_data['Unit_Of_Pay'].value_counts()
```

Out[56]:

```
Year         776059
Hour          49348
Month           500
Week            139
Bi-Weekly       137
Name: Unit_Of_Pay, dtype: int64
```

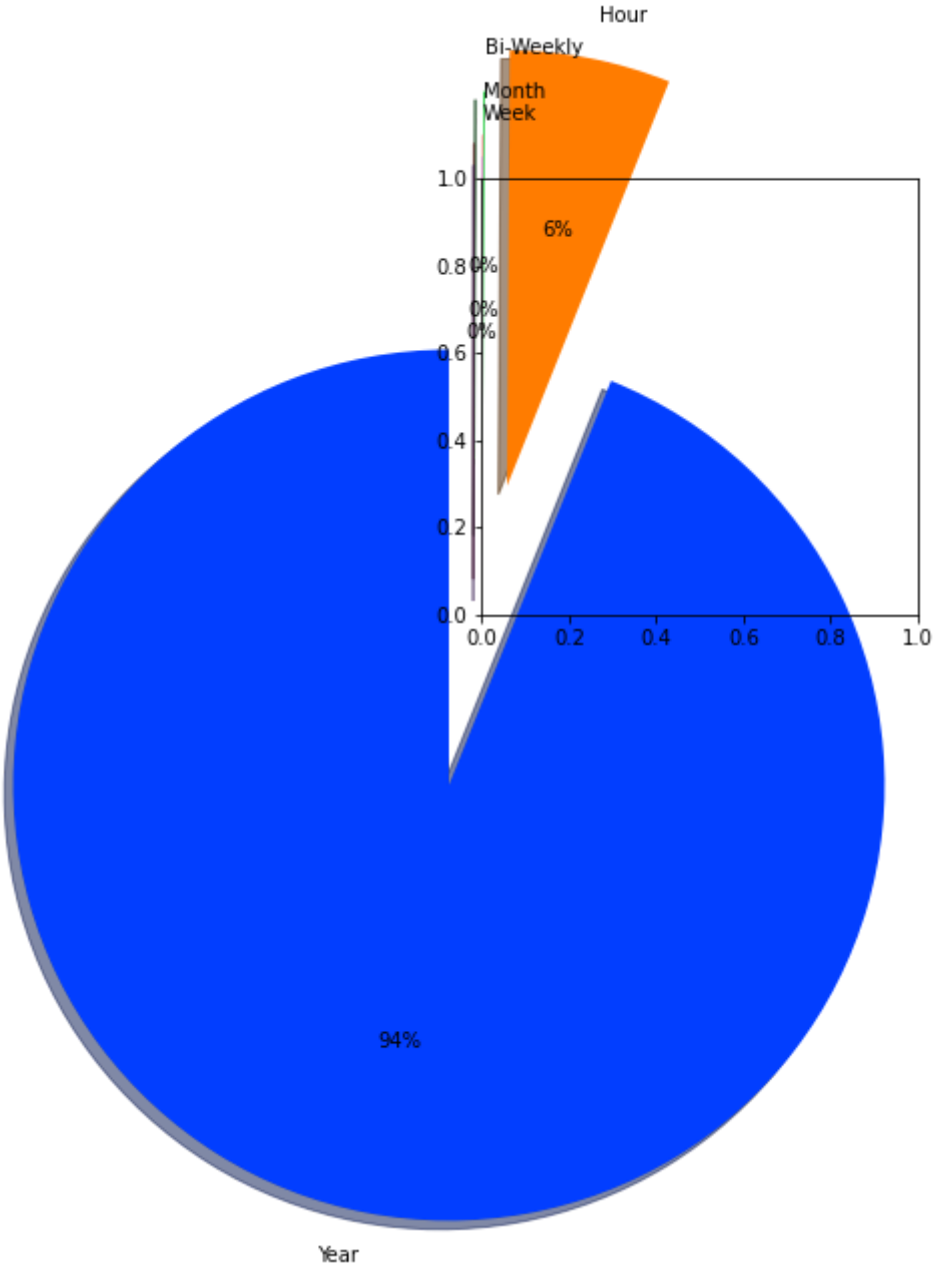
In [57]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Unit_Of_Pay'], data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [59]:

```
explode = [0.4, 0.3, 0.2, 0.1, 0.05]
colors = sns.color_palette('bright')
plt.pie(visa_data['Unit_Of_Pay'].value_counts(), labels= ['Year', 'Hour', 'Bi-Weekly',
        colors = colors, autopct = '%0.0f%%',  explode = explode, shadow = 'True',
        startangle = 90, frame = 'true')
plt.show()
```



In [60]:

```
visa_data['Quarter'].unique()
```

Out[60]:

```
array(['Q1', 'Q2', 'Q3', 'Q4'], dtype=object)
```

In [61]:

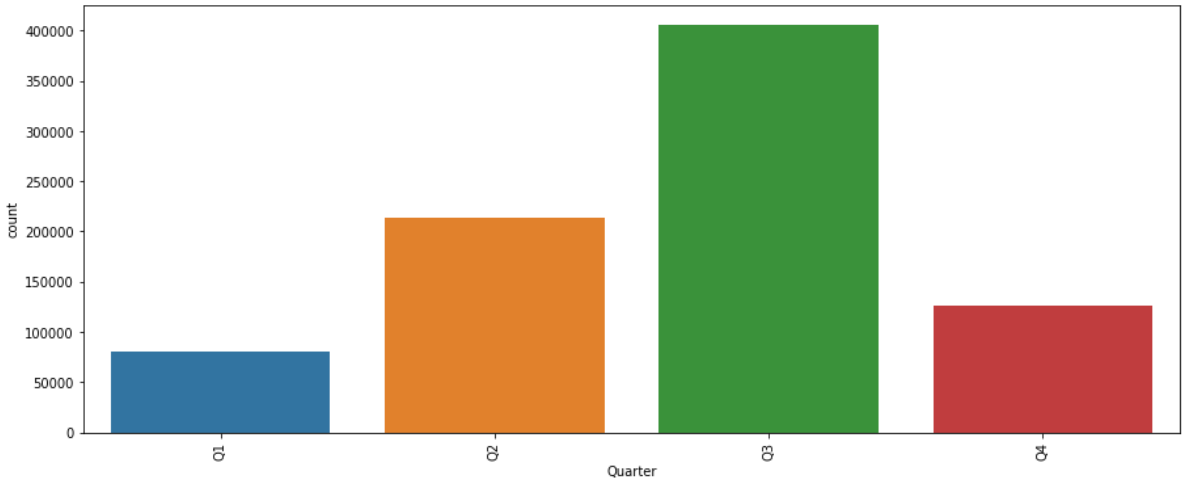
```
visa_data['Quarter'].value_counts()
```

Out[61]:

```
Q3    405570
Q2    213452
Q4    126550
Q1     80611
Name: Quarter, dtype: int64
```

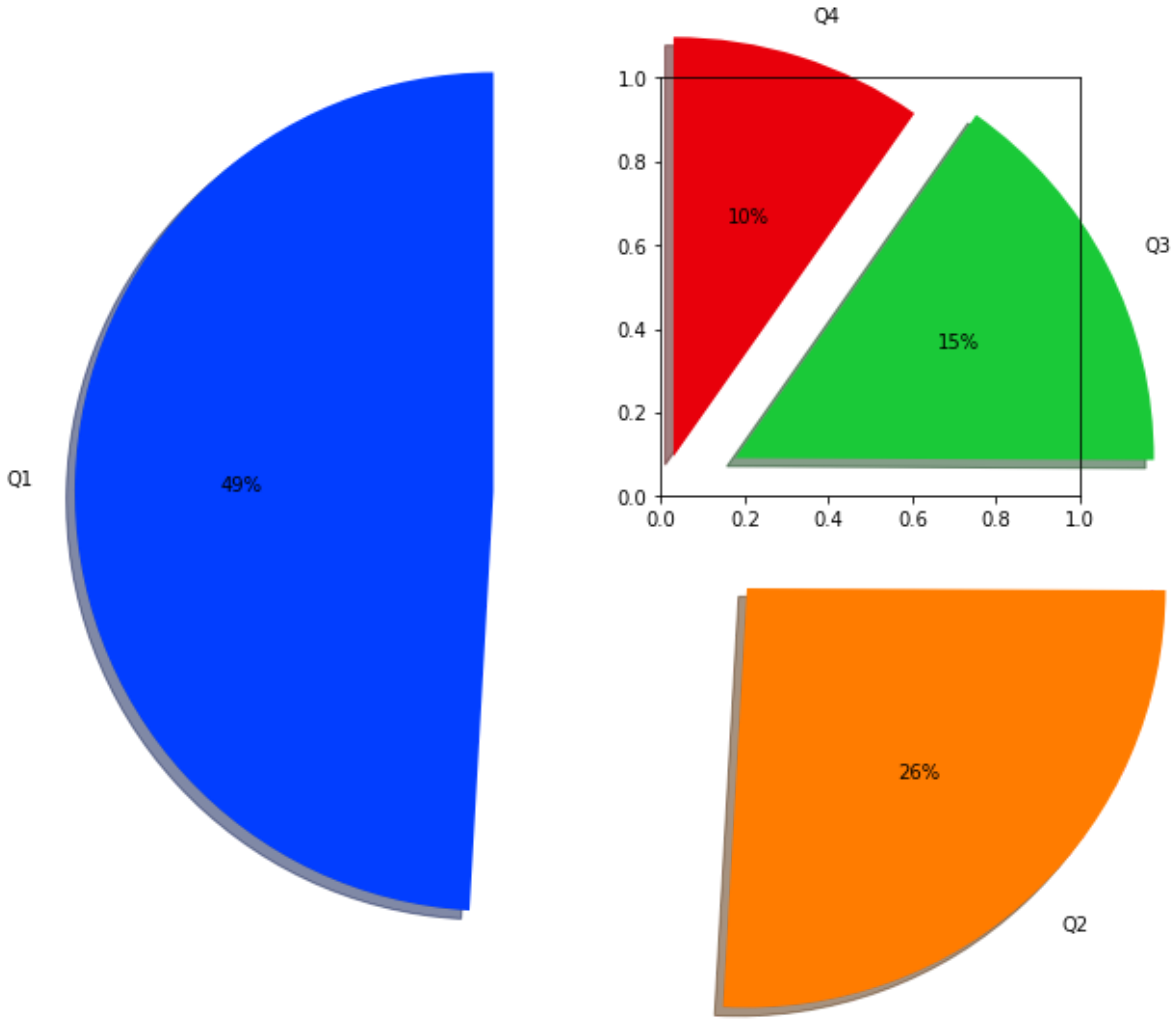
In [62]:

```
plt.figure(figsize=(15,6))
sns.countplot(visa_data['Quarter'], data = visa_data)
plt.xticks(rotation = 90)
plt.show()
```



In [63]:

```
explode = [0.4, 0.3, 0.2, 0.1]
colors = sns.color_palette('bright')
plt.pie(visa_data['Quarter'].value_counts(), labels= ['Q1', 'Q2', 'Q3', 'Q4'],
        colors = colors, autopct = '%0.0f%%',  explode = explode, shadow = 'True',
        startangle = 90, frame = 'true')
plt.show()
```



In [64]:

```
plt.figure(figsize=(15,6))
sns.countplot(hue = 'Visa_Class', x = 'Quarter', data = visa_data )
plt.xticks(rotation = 0)
plt.show()
```