```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt


data = pd.read_csv("dataset.csv", encoding='cp1252')
print(data.head())
```

```
  CORPORATE_IDENTIFICATION_NUMBER  \
0                           F00643
1                           F00721
2                           F00892
3                           F01208
4                           F01218

                                    COMPANY_NAME COMPANY_STATUS  \
0                                HOCHTIEFF AG,              NAEF
1  SUMITOMO CORPORATION (SUMITOMO SHOJI KAISHA LI...           ACTV
2                      SRILANKAN AIRLINES LIMITED              ACTV
3                         CALTEX INDIA LIMITED              NAEF
4              GE HEALTHCARE BIO-SCIENCES LIMITED              ACTV

  COMPANY_CLASS COMPANY_CATEGORY COMPANY_SUB_CATEGORY DATE_OF_REGISTRATION  \
0           NaN              NaN                  NaN            1/12/1961
1           NaN              NaN                  NaN                  NaN
2           NaN              NaN                  NaN            1/3/1982
3           NaN              NaN                  NaN                  NaN
4           NaN              NaN                  NaN                  NaN

  REGISTERED_STATE  AUTHORIZED_CAP  PAIDUP_CAPITAL  INDUSTRIAL_CLASS  \
0       Tamil Nadu               0             0.0               NaN
1       Tamil Nadu               0             0.0               NaN
2       Tamil Nadu               0             0.0               NaN
3       Tamil Nadu               0             0.0               NaN
4       Tamil Nadu               0             0.0               NaN

  PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN  \
0                  Agriculture & allied
1                  Agriculture & allied
2                  Agriculture & allied
3                  Agriculture & allied
4                  Agriculture & allied

                         REGISTERED_OFFICE_ADDRESS REGISTRAR_OF_COMPANIES  \
0  AMBLE SIDE, NO.8(OLD NO.30),3RD FLOOR KHADER N...               ROC DELHI
1  FLAT NO. 6, 1st FLOOR, 113/113ARAMA NAICKEN ST...               ROC DELHI
2  SRILANKAN AIRLINES LIMITED, VIJAYA TOWERSNO-4,...               ROC DELHI
3        GOLD CREST 24 55 NORTHUSMAN ROAD T NAGAR                  ROC DELHI
4   FF-3 Palani  Centre32 Venkat Naryan Road Nagar                 ROC DELHI

              EMAIL_ADDR LATEST_YEAR_ANNUAL_RETURN  \
0                    NaN                      NaN
1      shuchi.chug@asa.in                     NaN
2     shree16us@yahoo.com                     NaN
3                    NaN                      NaN
4  karthick9999@yahoo.com                     NaN

  LATEST_YEAR_FINANCIAL_STATEMENT
0                             NaN
1                             NaN
2                             NaN
3                             NaN
4                             NaN
```

```
data['DATE_OF_REGISTRATION'] = pd.to_datetime(data['DATE_OF_REGISTRATION'])
start_year =2000
data = data[data['DATE_OF_REGISTRATION'].dt.year >= start_year]
```

```
<ipython-input-74-1c05663a0baa>:1: UserWarning: Parsing dates in DD/MM/YYYY format
  data['DATE_OF_REGISTRATION'] = pd.to_datetime(data['DATE_OF_REGISTRATION'])
```

```
monthly_registrations = data.groupby(data['DATE_OF_REGISTRATION'].dt.to_period("M")).si
subsampling_factor = 3# Adjust as needed to control the number of data points
subsetted_data = monthly_registrations.iloc[::subsampling_factor]
```

dataset.csv

1 to 10 of 1499 entries Filter

TATUS	COMPANY_CLASS	COMPANY_CATEGORY	COMPA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA
	NA	NA	NA

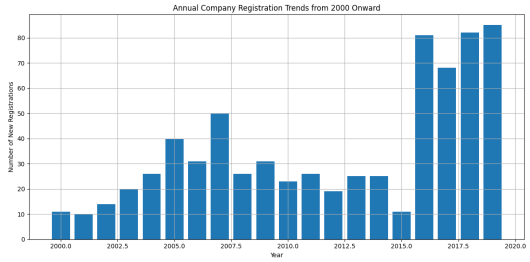
Show 10 per page

1 2 10 100 140 150

```
yearly_registrations = data.groupby(data['DATE_OF_REGISTRATION'].dt.year).size()

# Create a bar plot to visualize the annual registration trends
plt.figure(figsize=(12, 6))
plt.bar(yearly_registrations.index, yearly_registrations.values, align='center')
plt.title('Annual Company Registration Trends from 2000 Onward')
plt.xlabel('Year')
plt.ylabel('Number of New Registrations')
plt.grid(True)

# Display the plot
plt.tight_layout()
plt.show()
```

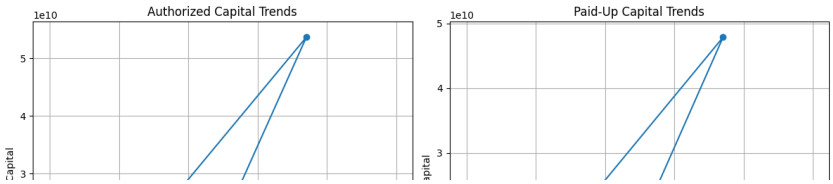



```
# Analyze trends in other columns, for example, 'AUTHORIZED_CAP' and 'PAIDUP_CAPITAL'
plt.figure(figsize=(12, 6))

# Plot Authorized Capital trends
plt.subplot(1, 2, 1)
plt.plot(data['DATE_OF_REGISTRATION'], data['AUTHORIZED_CAP'], marker='o', linestyle='-
plt.title('Authorized Capital Trends')
plt.xlabel('Time Period')
plt.ylabel('Authorized Capital')
plt.grid(True)
plt.xticks(rotation=45)

# Plot Paid-Up Capital trends
plt.subplot(1, 2, 2)
plt.plot(data['DATE_OF_REGISTRATION'], data['PAIDUP_CAPITAL'], marker='o', linestyle='-
plt.title('Paid-Up Capital Trends')
plt.xlabel('Time Period')
plt.ylabel('Paid-Up Capital')
plt.grid(True)
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```

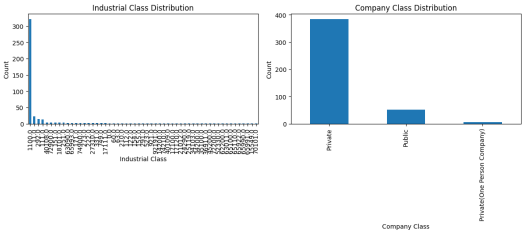



```
plt.figure(figsize=(12, 8))

# Plot Industrial Class trends
plt.subplot(2, 2, 1)
data['INDUSTRIAL_CLASS'].value_counts().plot(kind='bar')
plt.title('Industrial Class Distribution')
plt.xlabel('Industrial Class')
plt.ylabel('Count')

# Plot Company Class trends
plt.subplot(2, 2, 2)
data['COMPANY_CLASS'].value_counts().plot(kind='bar')
plt.title('Company Class Distribution')
plt.xlabel('Company Class')
plt.ylabel('Count')

plt.tight_layout()
plt.show()
```

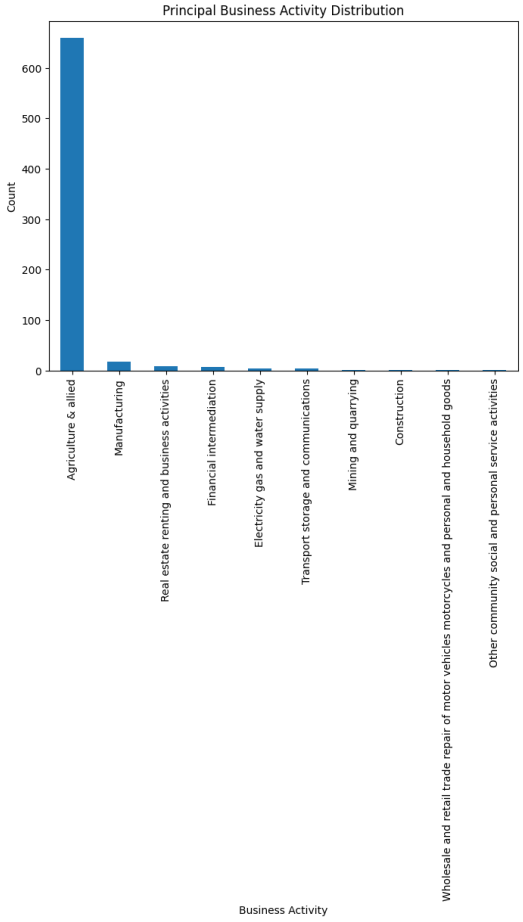



```
plt.figure(figsize=(12, 8))

# Plot Principal Business Activity Distribution

data['PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN'].value_counts().plot(kind='bar', figsize=
plt.title('Principal Business Activity Distribution')
plt.xlabel('Business Activity')
plt.ylabel('Count')
```

```
Text(0, 0.5, 'Count')
```



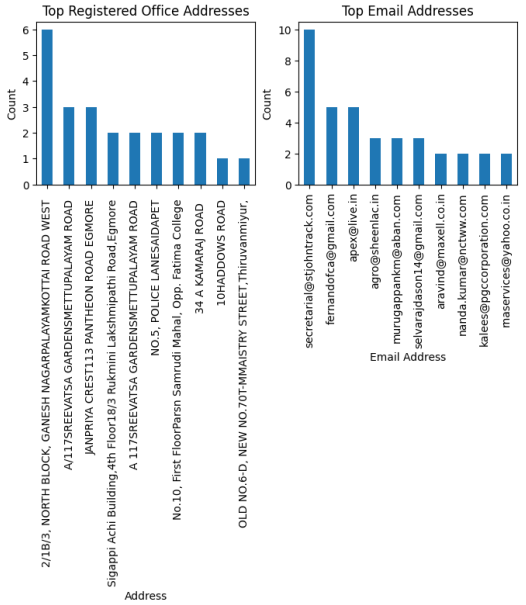

```
plt.figure(figsize=(12, 8))

# Plot Registered Office Address trends
plt.subplot(2, 2, 1)
data['REGISTERED_OFFICE_ADDRESS'].value_counts().head(10).plot(kind='bar', figsize=(8, 
plt.title('Top Registered Office Addresses')
plt.xlabel('Address')
plt.ylabel('Count')

# Plot Email Address trends
plt.subplot(2, 2, 2)
data['EMAIL_ADDR'].value_counts().head(10).plot(kind='bar', figsize=(8, 6))
plt.title('Top Email Addresses')
plt.xlabel('Email Address')
plt.ylabel('Count')

plt.tight_layout()
plt.show()
```

```
<ipython-input-83-1db337c18803>:17: UserWarning: Ti
  plt.tight_layout()
```




```
plt.figure(figsize=(12, 8))

# Plot the distribution of the latest annual return years
plt.subplot(2, 2, 1)
data['LATEST_YEAR_ANNUAL_RETURN'].value_counts().plot(kind='bar', figsize=(8, 6))
plt.title('Distribution of Latest Annual Return Years')
plt.xlabel('Year')
plt.ylabel('Count')

# Plot the distribution of the latest financial statement years
plt.subplot(2, 2, 2)
data['LATEST_YEAR_FINANCIAL_STATEMENT'].value_counts().plot(kind='bar', figsize=(8, 6))
plt.title('Distribution of Latest Financial Statement Years')
plt.xlabel('Year')
plt.ylabel('Count')

plt.tight_layout()
plt.show()
```