Mock Interview Python Screening test

In [4]:
```
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
dataframe = pd.read_csv("adult_census_data.csv")
```

In [5]:
```
dataframe.head()
```

Out[5]:

	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K

In [6]:
```
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32560 entries, 0 to 32559
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   39              32560 non-null  int64
 1    State-gov      32560 non-null  object
 2    77516          32560 non-null  int64
 3    Bachelors      32560 non-null  object
 4    13             32560 non-null  int64
 5    Never-married  32560 non-null  object
 6    Adm-clerical   32560 non-null  object
 7    Not-in-family  32560 non-null  object
 8    White          32560 non-null  object
 9    Male           32560 non-null  object
 10   2174           32560 non-null  int64
 11   0              32560 non-null  int64
 12   40             32560 non-null  int64
 13   United-States  32560 non-null  object
 14   <=50K          32560 non-null  object
dtypes: int64(6), object(9)
memory usage: 3.7+ MB
```

Q1. After importing the adult_census_data.csv file, please filter this to include only the following criteria:

- State-Gov
- Bachelors
- Never-Married
- Adm-Clerical
- Not-in-familiy
- White
- Male
- United States
- <=50K

Feel free to any method to complete this tasks. However, we recommend you use either list filtering [], or .loc to complete this task.

Put your code below

In [9]:
```
filtered_df = dataframe.loc[:, [' State-gov', ' Bachelors', ' Never-married', ' Adm-clerical', ' Not-in-family', ' Whit
```

Currently, the dataframe you are using has the following column names:

[' State-gov', ' Bachelors', ' Never-married', ' Adm-clerical', ' Not-in-family', ' White', ' Male', ' United-States', ' <=50K']

Q2. Please re-name all the newly filtered columns in the pandas DataFrame to the following:

Employment Type, Degree Status, Marriage-Status, Job-Role, Family-Role, Ethnicity, Gender, Country, Earnings

E.g. State-Gov becomes Employment Type, Bachelors becomes Degree Status, etc.

Put your code below

In [10]:
```
df = filtered_df.rename({' State-gov': 'Employment Type',
                  ' Bachelors': 'Degree Status',
                  ' Never-married': 'Marriage-Status',
                  ' Adm-clerical': 'Job-Role',
                  ' Not-in-family' : 'Family-Role',
                  ' White': 'Ethnicity',
                  ' Male': 'Gender',
                  ' United-States': 'Country',
                  ' <=50K': 'Status'}, axis=1)
df.head(3)
```

Out[10]:

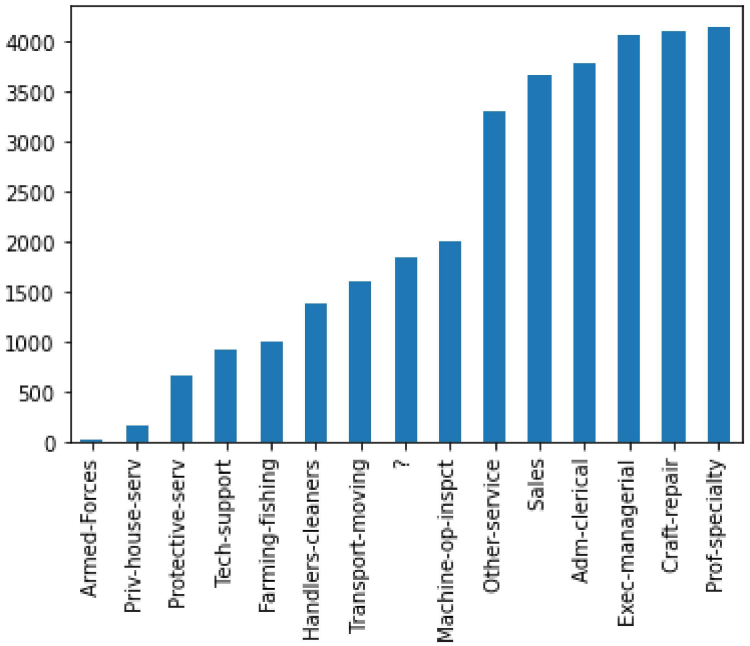
	Employment Type	Degree Status	Marriage-Status	Job-Role	Family-Role	Ethnicity	Gender	Country	Status
0	Self-emp-not-inc	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	United-States	<=50K
1	Private	HS-grad	Divorced	Handlers-cleaners	Not-in-family	White	Male	United-States	<=50K
2	Private	11th	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	United-States	<=50K

Q3. The Job Role Columns holds the job information for each individual in this census snapshot. Using this column, create a Bar Chart that shows the count of 'Unique' Jobs per Job Group in the "Job-Role" Column in ascending order, as per the provided image below

Put your code below

In [13]:
```
df['Job-Role'].value_counts().sort_values().plot.bar(rot=90)
```

Out[13]:
```
<AxesSubplot:>
```



Q4. Please create two bar plots as per below that show:

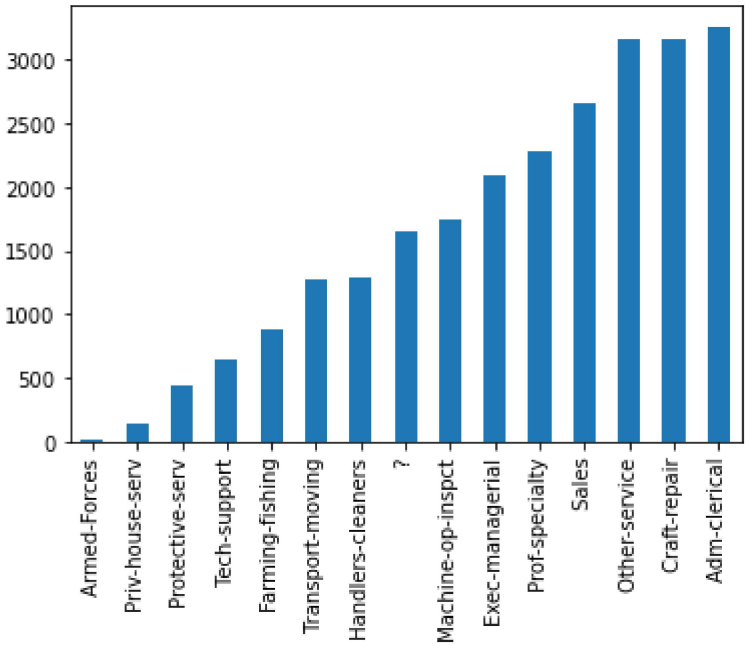
1) The number of individuals who have a High School Graduate Diploma AND earn <=50K in the United States
2) The number of individuals who have a High School Graduate Diploma AND earn >50K in the United States

Please note you will be looking specifically at the *Job Role* column

Put Your Code Below

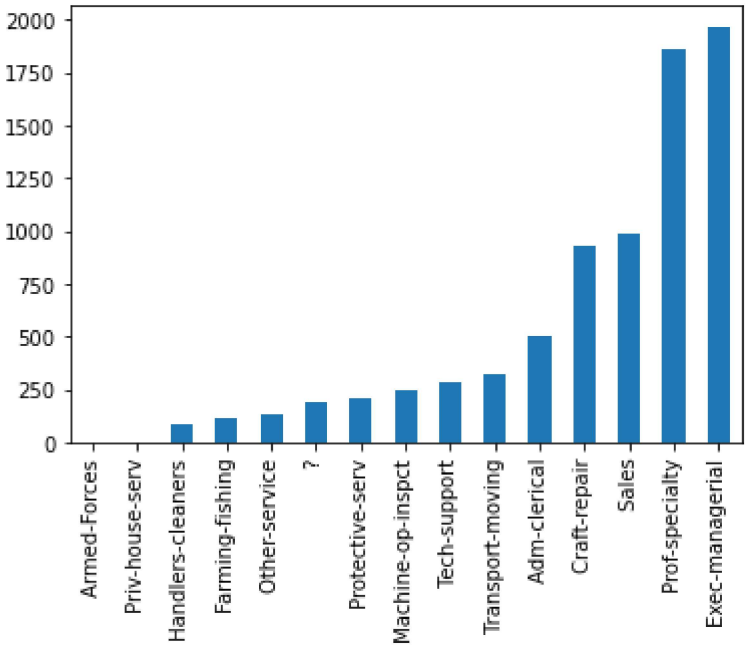
In [14]:
```
equal_or_less_50K = df[df.Status == ' <=50K']
equal_or_less_50K['Job-Role'].value_counts().sort_values().plot.bar(rot=90)
```

Out[14]:
```
<AxesSubplot:>
```



In [15]:
```
more_than_50K = df[df.Status == ' >50K']
more_than_50K['Job-Role'].value_counts().sort_values().plot.bar(rot=90)
```

Out[15]:
```
<AxesSubplot:>
```



Challenge Question

Q5. Which Job Role has the highest *proportion* of individuals who earn >50K?

Put your code below

In [16]:
```
#optional question - did not attempt currently
```