In [2]:
```
!pip install pandas
# !pip install basic-image-eda
# !pip install scikit-image
# !pip install matplotlib
import pandas as pd
```

```
Looking in indexes: https://pypi.org/simple, https://pypi.ngc.nvidia.com
Requirement already satisfied: pandas in /opt/conda/lib/python3.8/site-packages (1.4.1)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.8/site-packages (from pandas) (2021.3)
Requirement already satisfied: numpy>=1.18.5 in /opt/conda/lib/python3.8/site-packages (from pandas) (1.21.2)
Requirement already satisfied: python-dateutil>=2.8.1 in /opt/conda/lib/python3.8/site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.8/site-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [3]:
```
train = pd.read_csv('data/train22A/train2_new.csv') # reading the csv file
```

In [4]:
```
train.head() # printing first five rows of the file
```

Out[4]:

	name	upperLength	clothesStyles	hairStyles	lowerLength	lowerStyles	shoesStyles	t
0	img_qh_train2_00000006345.jpg	LongSleeve	Solidcolor	Short	Skirt	Solidcolor	Sandals	
1	img_qh_train2_00001008825.jpg	LongSleeve	multicolour	Short	Trousers	Solidcolor	Sneaker	
2	img_qh_train2_00002004117.jpg	LongSleeve	multicolour	Long	Trousers	Solidcolor	Sneaker	
3	img_qh_train2_00003002526.jpg	LongSleeve	multicolour	Short	Trousers	Solidcolor	Sneaker	
4	img_qh_train2_00004004439.jpg	LongSleeve	Solidcolor	Long	Trousers	Solidcolor	Sneaker	

5 rows × 30 columns

In [5]:
```
train.columns
```

Out[5]:
```
Index(['name', 'upperLength', 'clothesStyles', 'hairStyles', 'lowerLength',
       'lowerStyles', 'shoesStyles', 'towards', 'upperBlack', 'upperBrown',
       'upperBlue', 'upperGreen', 'upperGray', 'upperOrange', 'upperPink',
       'upperPurple', 'upperRed', 'upperWhite', 'upperYellow', 'lowerBlack',
       'lowerBrown', 'lowerBlue', 'lowerGreen', 'lowerGray', 'lowerOrange',
       'lowerPink', 'lowerPurple', 'lowerRed', 'lowerWhite', 'lowerYellow'],
      dtype='object')
```

In [6]:
```
train.drop(['upperLength', 'clothesStyles', 'hairStyles', 'upperBlack', 'upperBrown',
       'upperBlue', 'upperGreen', 'upperGray', 'upperOrange', 'upperPink',
       'upperPurple', 'upperRed', 'upperWhite', 'upperYellow'], axis=1, inplace=True)
```

In [7]:
```
train.head()
```

Out[7]:

	name	lowerLength	lowerStyles	shoesStyles	towards	lowerBlack	lowerBrown	lowe
0	img_qh_train2_00000006345.jpg	Skirt	Solidcolor	Sandals	right	1.0	NaN	
1	img_qh_train2_00001008825.jpg	Trousers	Solidcolor	Sneaker	back	1.0	NaN	

	name	lowerLength	lowerStyles	shoesStyles	towards	lowerBlack	lowerBrown	lowe
2	img_qh_train2_00002004117.jpg	Trousers	Solidcolor	Sneaker	right	1.0	NaN	
3	img_qh_train2_00003002526.jpg	Trousers	Solidcolor	Sneaker	front	NaN	NaN	
4	img_qh_train2_00004004439.jpg	Trousers	Solidcolor	Sneaker	right	NaN	1.0	

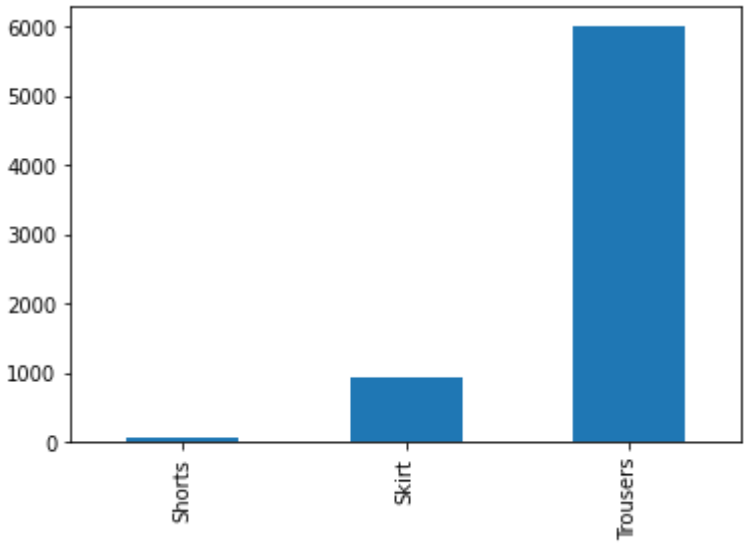
In [8]:
```
train.columns
```

Out[8]:
```
Index(['name', 'lowerLength', 'lowerStyles', 'shoesStyles', 'towards',
       'lowerBlack', 'lowerBrown', 'lowerBlue', 'lowerGreen', 'lowerGray',
       'lowerOrange', 'lowerPink', 'lowerPurple', 'lowerRed', 'lowerWhite',
       'lowerYellow'],
      dtype='object')
```

In [9]:
```
print('lowerLength: ')
print(train['lowerLength'].value_counts(ascending=True))
train['lowerLength'].value_counts(ascending=True).plot.bar()
```

```
lowerLength:
Shorts        62
Skirt        936
Trousers    6002
Name: lowerLength, dtype: int64
<AxesSubplot:>
```

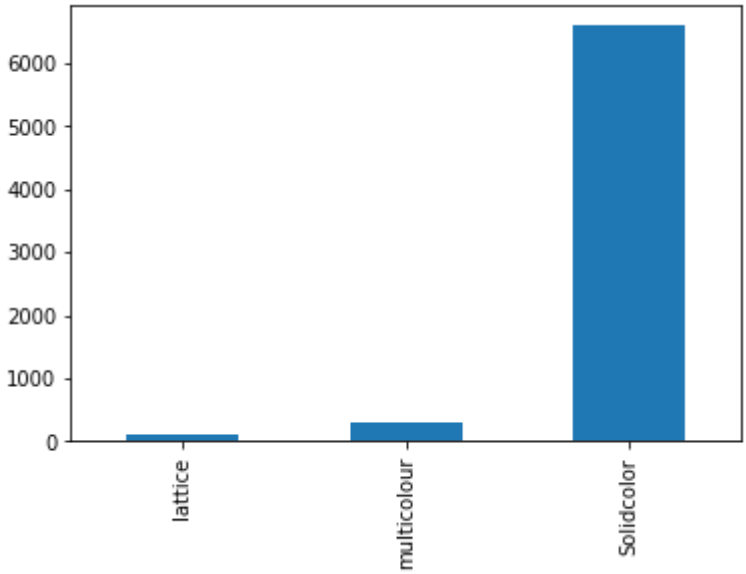
Out[9]:



In [10]:
```
print('lowerStyles: ')
print(train['lowerStyles'].value_counts(ascending=True))
train['lowerStyles'].value_counts(ascending=True).plot.bar()
```

```
lowerStyles:
lattice         104
multicolour     300
Solidcolor     6596
Name: lowerStyles, dtype: int64
<AxesSubplot:>
```

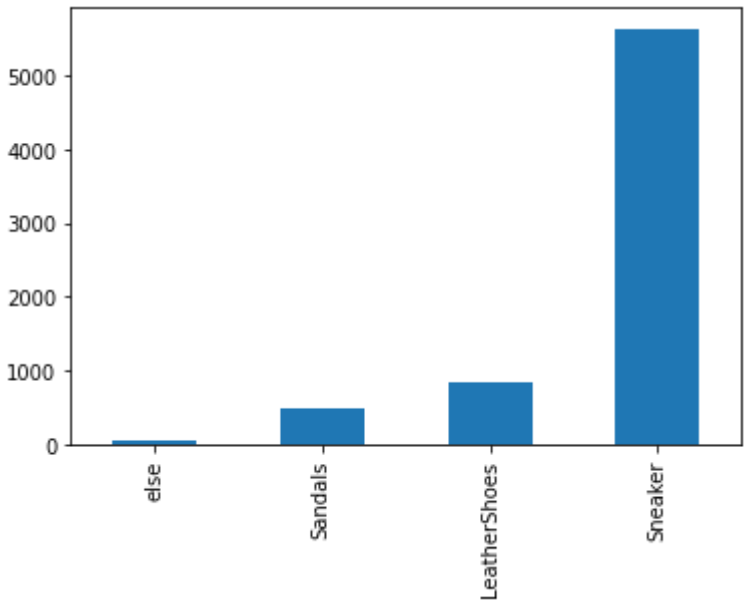
Out[10]:



In [11]:
```
print('shoesStyles: ')
print(train['shoesStyles'].value_counts(ascending=True))
train['shoesStyles'].value_counts(ascending=True).plot.bar()
```

```
shoesStyles:
else              42
Sandals          473
LeatherShoes     845
Sneaker         5640
Name: shoesStyles, dtype: int64
<AxesSubplot:>
```

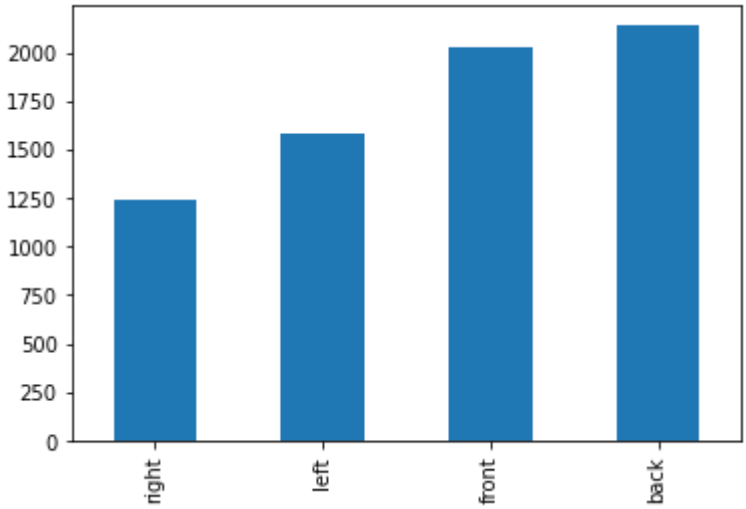
Out[11]:



In [12]:
```
print('towards: ')
print(train['towards'].value_counts(ascending=True))
train['towards'].value_counts(ascending=True).plot.bar()
```

```
towards:
right    1244
left     1585
front    2031
back     2140
Name: towards, dtype: int64
<AxesSubplot:>
```

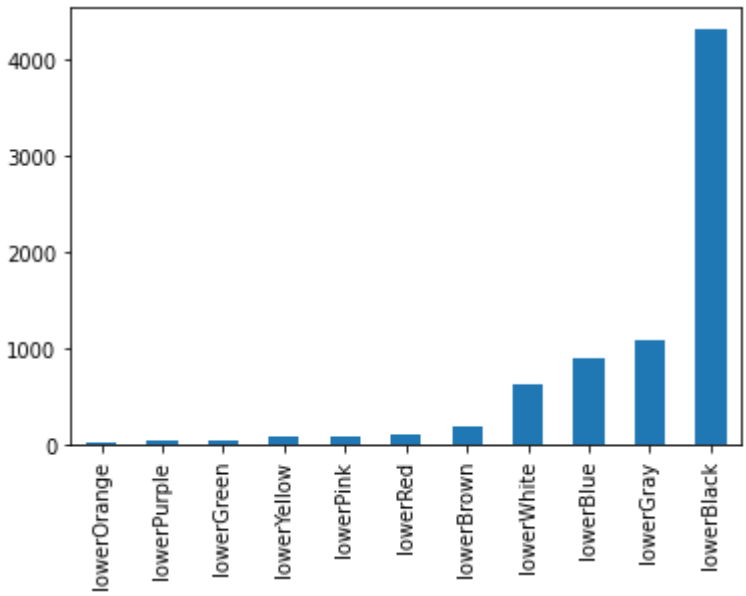
Out[12]:



In [16]:
```
new_df = train.copy(deep=True)
new_df.drop(['name','lowerLength','lowerStyles', 'shoesStyles', 'towards'], axis=1, inplace=True
colorsCount = new_df.count().sort_values()
print(colorsCount)
colorsCount.plot.bar()
```

```
lowerOrange       10
lowerPurple       26
lowerGreen        34
lowerYellow       81
lowerPink         82
lowerRed          99
lowerBrown       172
lowerWhite       627
lowerBlue        896
lowerGray       1086
lowerBlack      4320
dtype: int64
<AxesSubplot:>
```

Out[16]:



In [19]:
```
lowerStyles_group = train.groupby('lowerStyles').count()
lowerStyles_group.drop(['name','lowerLength', 'shoesStyles', 'towards'], axis=1, inplace=True)
lowerStyles_group.head()
```

Out[19]:

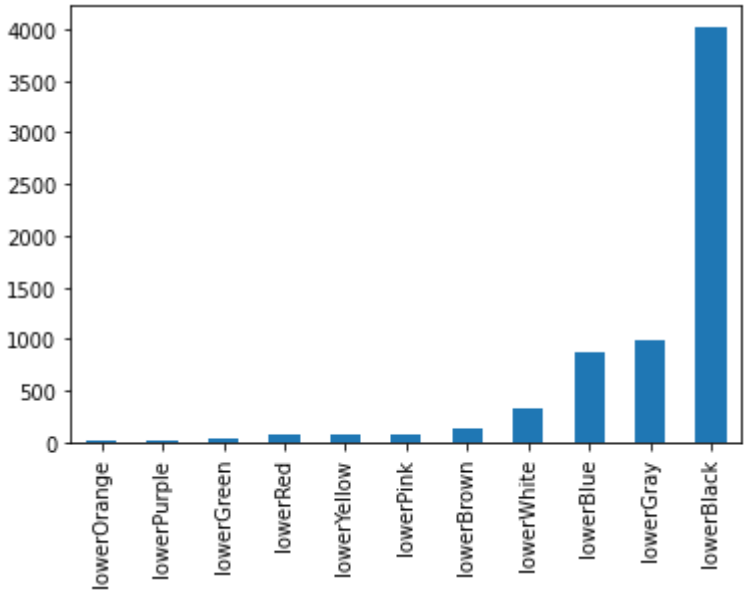
lowerStyles	lowerBlack	lowerBrown	lowerBlue	lowerGreen	lowerGray	lowerOrange	lowerPink	lowerPurple
Solidcolor	4029	139	865	25	978	9	74	18

lowerStyles	lowerBlack	lowerBrown	lowerBlue	lowerGreen	lowerGray	lowerOrange	lowerPink	lowerPurple
lattice	86	3	2	1	23	0	3	1
multicolour	205	30	29	8	85	1	5	7

In [20]:
```
print(lowerStyles_group.loc['Solidcolor'].sort_values())
lowerStyles_group.loc['Solidcolor'].sort_values().plot.bar()
```

```
lowerOrange       9
lowerPurple      18
lowerGreen       25
lowerRed         62
lowerYellow      70
lowerPink        74
lowerBrown      139
lowerWhite      327
lowerBlue       865
lowerGray       978
lowerBlack     4029
Name: Solidcolor, dtype: int64
<AxesSubplot:>
```

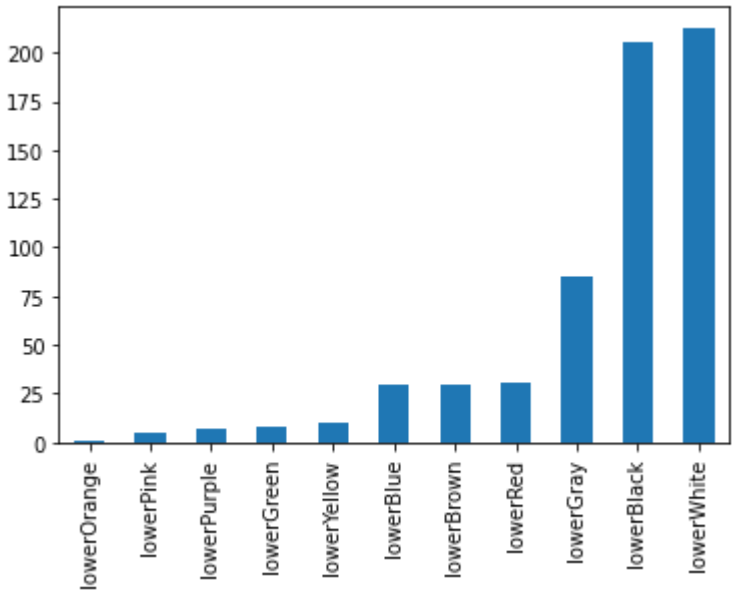
Out[20]:



In [21]:
```
print(lowerStyles_group.loc['multicolour'].sort_values())
lowerStyles_group.loc['multicolour'].sort_values().plot.bar()
```

```
lowerOrange       1
lowerPink         5
lowerPurple       7
lowerGreen        8
lowerYellow      10
lowerBlue        29
lowerBrown       30
lowerRed         31
lowerGray        85
lowerBlack      205
lowerWhite      213
Name: multicolour, dtype: int64
<AxesSubplot:>
```

Out[21]:

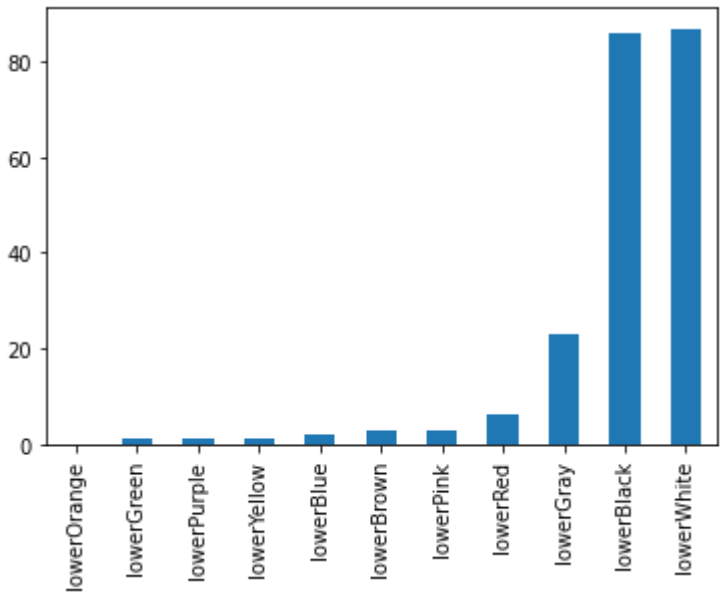


In [22]:

```
print(lowerStyles_group.loc['lattice'].sort_values())
lowerStyles_group.loc['lattice'].sort_values().plot.bar()
```

```
lowerOrange      0
lowerGreen       1
lowerPurple      1
lowerYellow      1
lowerBlue        2
lowerBrown       3
lowerPink        3
lowerRed         6
lowerGray       23
lowerBlack      86
lowerWhite      87
Name: lattice, dtype: int64
<AxesSubplot:>
```

Out[22]:



In [25]:

```
color_count_df = train.copy(deep=True)
colors = ['lowerBlack', 'lowerBrown', 'lowerBlue', 'lowerGreen', 'lowerGray',
       'lowerOrange', 'lowerPink', 'lowerPurple', 'lowerRed', 'lowerWhite',
       'lowerYellow']
def apply_color_count(series):
    count = 0
    for i in colors:
        if series[i] > 0:
            count = count + 1
    return count
```

```
color_count_df["color_count"] = color_count_df.apply(apply_color_count, axis=1)

color_count_df = color_count_df.drop(['lowerLength', 'shoesStyles', 'towards',
       'lowerBlack', 'lowerBrown', 'lowerBlue', 'lowerGreen', 'lowerGray',
       'lowerOrange', 'lowerPink', 'lowerPurple', 'lowerRed', 'lowerWhite',
       'lowerYellow'], axis=1)

color_count_df.head()
```

Out[25]:

	name	lowerStyles	color_count
0	img_qh_train2_00000006345.jpg	Solidcolor	1
1	img_qh_train2_00001008825.jpg	Solidcolor	1
2	img_qh_train2_00002004117.jpg	Solidcolor	1
3	img_qh_train2_00003002526.jpg	Solidcolor	1
4	img_qh_train2_00004004439.jpg	Solidcolor	1

In [26]:

```
for index, row in color_count_df.iterrows():
    if row['lowerStyles'] == 'Solidcolor' and row['color_count'] > 1:
        print('[error label Solidcolor] name: ', row['name'])
    if row['lowerStyles'] == 'lattice' and row['color_count'] < 2:
        print('[error label lattice] name: ', row['name'])
    if row['lowerStyles'] == 'multicolour' and row['color_count'] < 2:
        print('[error label multicolour] name: ', row['name'])
```

In [27]:

```
color_frequency_df = color_count_df.copy(deep=True)
def apply_color_frequency_item(series, frequency):
    if series['color_count'] == frequency:
        return 1
    else:
        return 0

color_frequency_df["one"] = color_frequency_df.apply(apply_color_frequency_item, args=(1,), axis
color_frequency_df["two"] = color_frequency_df.apply(apply_color_frequency_item, args=(2,), axis
color_frequency_df["three"] = color_frequency_df.apply(apply_color_frequency_item, args=(3,), ax

color_frequency_df = color_frequency_df.drop(['name', 'color_count'], axis=1)
color_frequency_df.head()
```

Out[27]:

	lowerStyles	one	two	three
0	Solidcolor	1	0	0
1	Solidcolor	1	0	0
2	Solidcolor	1	0	0
3	Solidcolor	1	0	0
4	Solidcolor	1	0	0

In [29]:

```
color_frequency_df = color_frequency_df.groupby('lowerStyles').sum()
color_frequency_df.head()
```

Out[29]:

lowerStyles	one	two	three
Solidcolor	6596	0	0
lattice	0	99	5

	one	two	three
lowerStyles			
multicolour	0	276	24

In [30]:
```
color_frequency_df.plot.bar()
```

Out[30]: <AxesSubplot:xlabel='lowerStyles'>