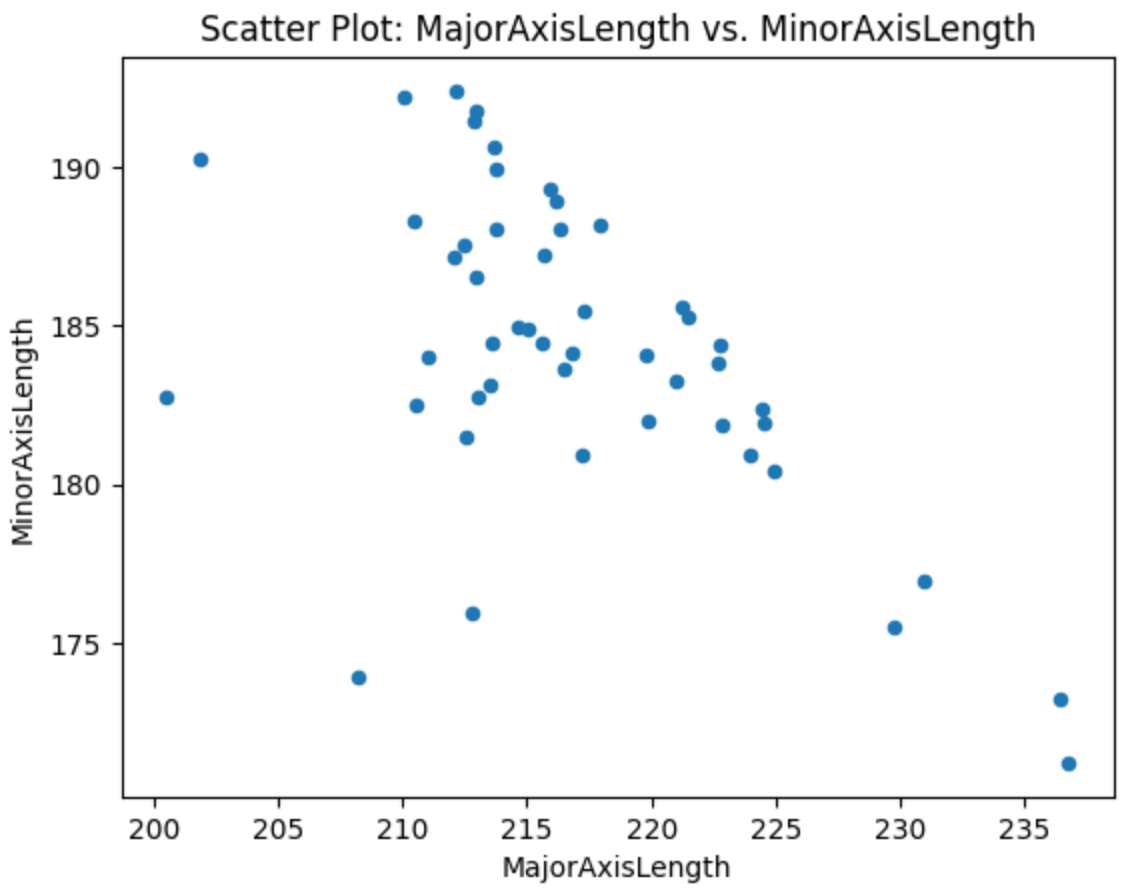
DANIEL MBUGUA KARUME

SCT211-0072/2022

1. **Scatter Plot: MajorAxisLength vs. MinorAxisLength**
   - This scatter plot visualizes the relationship between the `MajorAxisLength` and `MinorAxisLength` features in the *Dry Bean Dataset*. Each point represents a data entry, and the plot helps to observe patterns or correlations between these two variables. The x-axis represents `MajorAxisLength`, and the y-axis represents `MinorAxisLength`.

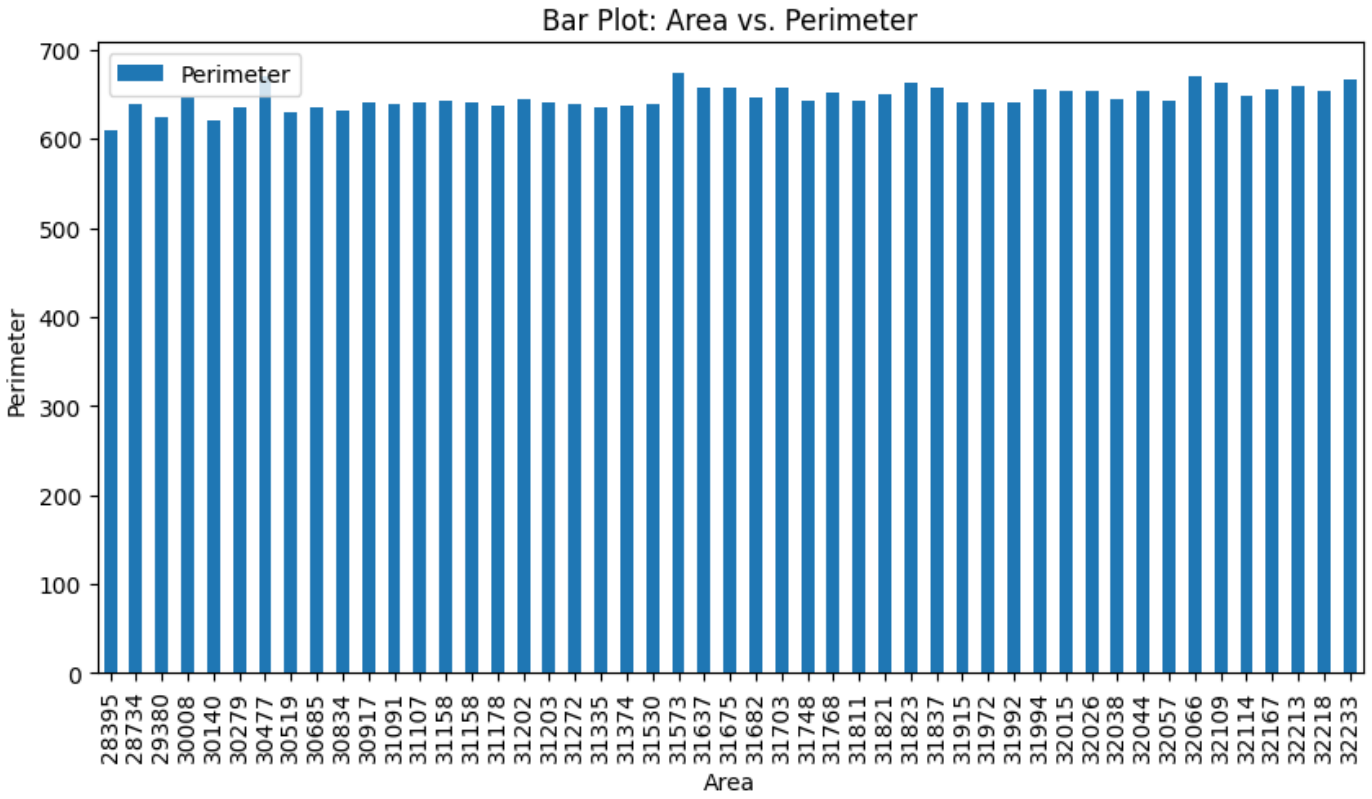
In [19]:
```
import pandas as pd
from matplotlib import pyplot as plt

# Load data from an Excel file
df_excel = pd.read_excel("./data/Dry_Bean_Dataset.xlsx")

# Scatter plot for MajorAxisLength vs. MinorAxisLength
df_excel.plot(kind="scatter", x="MajorAxisLength", y="MinorAxisLength")
plt.title('Scatter Plot: MajorAxisLength vs. MinorAxisLength')
plt.xlabel('MajorAxisLength')
plt.ylabel('MinorAxisLength')
plt.show()
```



1. **Bar Plot: Area vs. Perimeter**
   - The bar plot illustrates the values of the `Area` and `Perimeter` features for each data entry in the *Dry Bean Dataset*. Each bar corresponds to a specific entry, providing a clear comparison

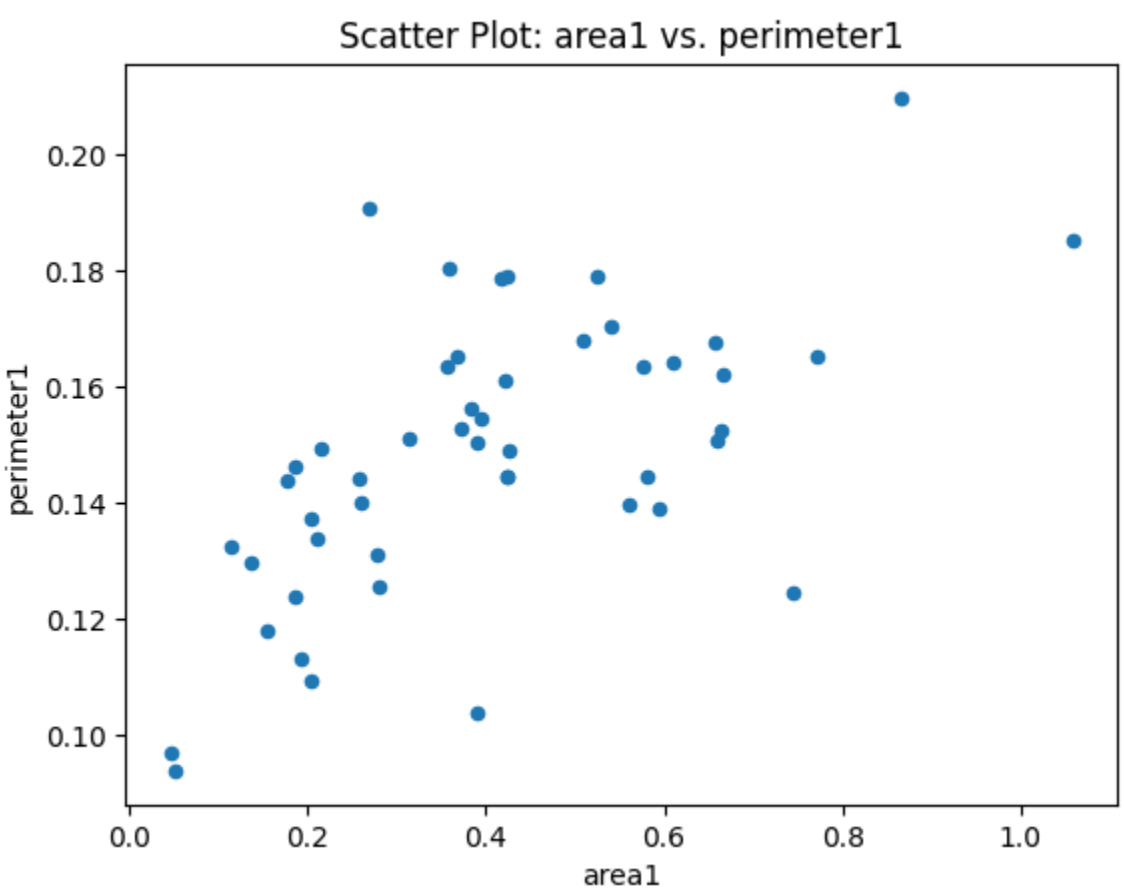
between the two variables. The x-axis represents the `Area`, and the y-axis represents `Perimeter`. This type of plot is useful for comparing quantitative values across different categories.

In [20]:
```
# Bar plot for Area vs. Perimeter
df_excel.plot(kind="bar", x="Area", y="Perimeter", figsize=(10, 5))
plt.title('Bar Plot: Area vs. Perimeter')
plt.xlabel('Area')
plt.ylabel('Perimeter')
plt.show()
```



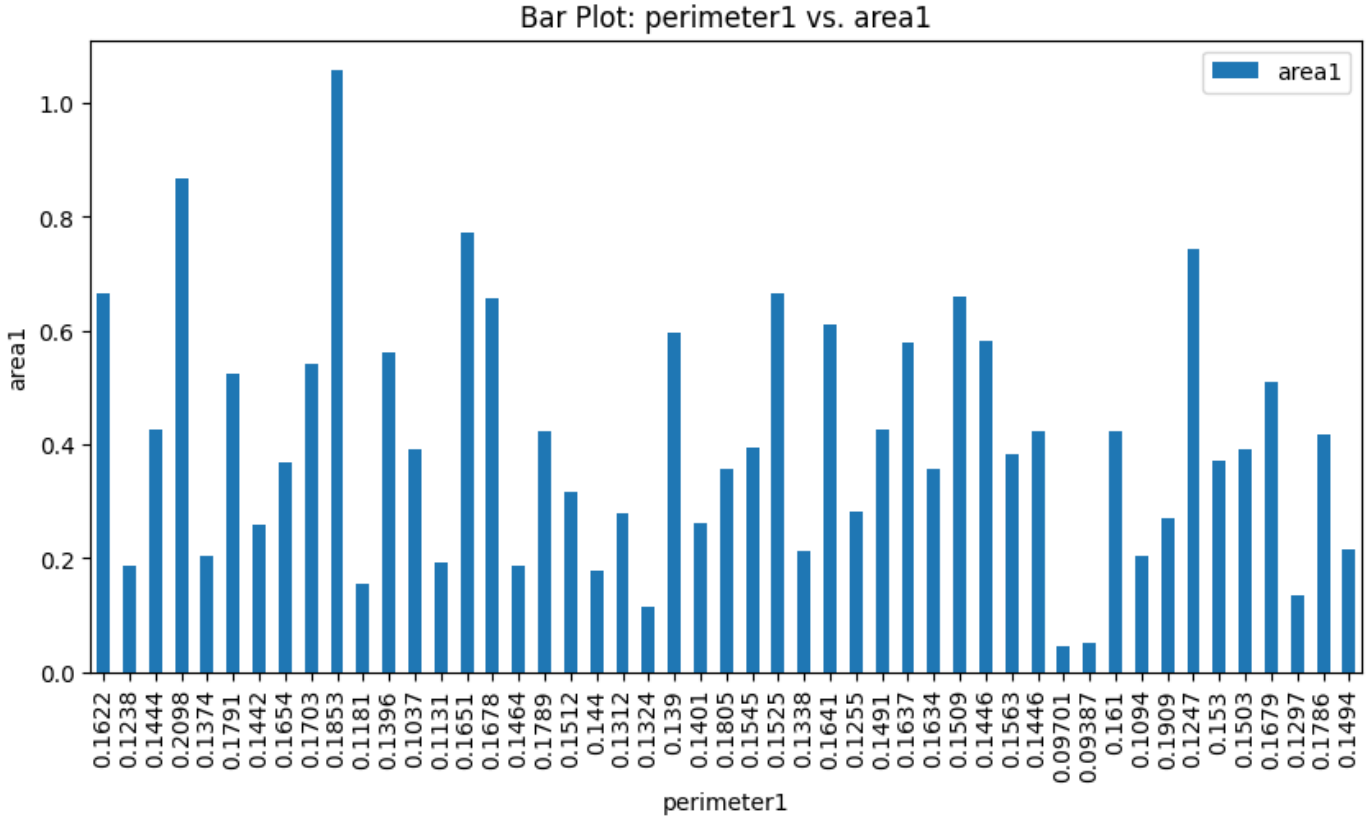
1. **Scatter Plot: area1 vs. perimeter1**
   - This scatter plot displays the relationship between the `area1` and `perimeter1` features in the *wdbc.data* dataset. Each point on the plot represents a data entry, allowing for visual examination of any potential correlations or patterns between these two variables. The x-axis represents `area1`, and the y-axis represents `perimeter1`.

In [21]:
```
# Load data from a CSV file
df_csv = pd.read_csv("./data/wdbc.data")

# Scatter plot for area1 vs. perimeter1
df_csv.plot(kind="scatter", x="area1", y="perimeter1")
plt.title('Scatter Plot: area1 vs. perimeter1')
plt.xlabel('area1')
plt.ylabel('perimeter1')
plt.show()
```



1. **Bar Plot: perimeter1 vs. area1**
   - The bar plot showcases the values of the `perimeter1` and `area1` features for each entry in the *wdbc.data* dataset. Each bar corresponds to a specific data point, offering a straightforward comparison between the two variables. The x-axis represents `perimeter1`, and the y-axis represents `area1`. This type of plot facilitates the comparison of quantitative values across different categories in the dataset.

In [22]:
```
# Bar plot for perimeter1 vs. area1
df_csv.plot(kind="bar", x="perimeter1", y="area1", figsize=(10, 5))
plt.title('Bar Plot: perimeter1 vs. area1')
plt.xlabel('perimeter1')
plt.ylabel('area1')
plt.show()
```


Bar Plot: perimeter1 vs. area1
area1
perimeter1
0.0
0.2
0.4
0.6
0.8
1.0
0.1622
0.1238
0.1444
0.2098
0.1374
0.1791
0.1442
0.1654
0.1703
0.1853
0.1181
0.1396
0.1037
0.1131
0.1651
0.1678
0.1464
0.1789
0.1512
0.144
0.1312
0.1324
0.139
0.1401
0.1805
0.1545
0.1525
0.1338
0.1641
0.1255
0.1491
0.1637
0.1634
0.1509
0.1446
0.1563
0.1446
0.09701
0.09387
0.161
0.1094
0.1909
0.1247
0.153
0.1503
0.1679
0.1297
0.1786
0.1494