Module 12 Challenge

Deliverable 2: Scrape and Analyze Mars Weather Data

In [1]:

```
# Import relevant libraries
from splinter import Browser
from bs4 import BeautifulSoup as soup
import matplotlib.pyplot as plt
import pandas as pd
```

In [2]:

```
browser = Browser('chrome')
```

Step 1: Visit the Website

Use automated browsing to visit the Mars Temperature Data Site (https://static.bc-edx.com/data/web/mars_facts/temperature.html). Inspect the page to identify which elements to scrape.

> **Hint** To identify which elements to scrape, you might want to inspect the page by using Chrome DevTools to discover whether the table contains usable classes.

In [3]:

```
# Visit the website
# https://static.bc-edx.com/data/web/mars_facts/temperature.html
url = "https://static.bc-edx.com/data/web/mars_facts/temperature.html"
browser.visit(url)
```

Step 2: Scrape the Table

Create a Beautiful Soup object and use it to scrape the data in the HTML table.

Note that this can also be achieved by using the Pandas `read_html` function. However, use Beautiful Soup here to continue sharpening your web scraping skills.

In [4]:

```
# Create a Beautiful Soup Object
html = browser.html
mars_soup = soup(html, 'html.parser')
```

In [5]:

```
# Extract all rows of data
table = mars_soup.find('table')
rows_data = []
rows = table.find_all('tr')

for idx, tr in enumerate(rows):
    cells = tr.find_all(['th', 'td'])
    row_data = [cell.get_text() for cell in cells]
    if idx == 0:
        rows_data.append(row_data)
    else:
        rows_data.append(row_data[:])

for row in rows_data:
    print(row)
```

```
['178', '2012-09-12', '36', '169', '6', '-73.0', '750.0']
['179', '2012-09-13', '37', '170', '6', '-73.0', '750.0']
['180', '2012-09-14', '38', '171', '6', '-73.0', '750.0']
['181', '2012-09-15', '39', '171', '6', '-75.0', '751.0']
['182', '2012-09-16', '40', '172', '6', '-75.0', '753.0']
['183', '2012-09-17', '41', '172', '6', '-75.0', '753.0']
['184', '2012-09-18', '42', '173', '6', '-75.0', '754.0']
['185', '2012-09-19', '43', '173', '6', '-74.0', '756.0']
['186', '2012-09-20', '44', '174', '6', '-75.0', '757.0']
['187', '2012-09-21', '45', '175', '6', '-74.0', '758.0']
['188', '2012-09-22', '46', '175', '6', '-74.0', '758.0']
['189', '2012-09-23', '47', '176', '6', '-75.0', '758.0']
['190', '2012-09-24', '48', '176', '6', '-75.0', '759.0']
['191', '2012-09-25', '49', '177', '6', '-74.0', '761.0']
['192', '2012-09-26', '50', '177', '6', '-72.0', '761.0']
['193', '2012-09-27', '51', '178', '6', '-76.0', '762.0']
['194', '2012-09-28', '52', '179', '6', '-74.0', '762.0']
['195', '2012-09-29', '53', '179', '6', '-71.0', '764.0']
['196', '2012-09-30', '54', '180', '7', '-72.0', '766.0']
['197', '2012-10-01', '55', '180', '7', '-74.0', '766.0']
```

Step 3: Store the Data

Assemble the scraped data into a Pandas DataFrame. The columns should have the same headings as the table on the website. Here's an explanation of the column headings:

- `id` : the identification number of a single transmission from the Curiosity rover
- `terrestrial_date` : the date on Earth
- `sol` : the number of elapsed sols (Martian days) since Curiosity landed on Mars
- `ls` : the solar longitude
- `month` : the Martian month
- `min_temp` : the minimum temperature, in Celsius, of a single Martian day (sol)
- `pressure` : The atmospheric pressure at Curiosity's location

In [6]:

```
# Confirm DataFrame was created successfully
column_headers = rows_data[0]
mars_data_df = pd.DataFrame(rows_data[1:], columns=column_headers)
mars_data_df[:5]
```

Out[6]:

	id	terrestrial_date	sol	ls	month	min_temp	pressure
0	2	2012-08-16	10	155	6	-75.0	739.0
1	13	2012-08-17	11	156	6	-76.0	740.0
2	24	2012-08-18	12	156	6	-76.0	741.0
3	35	2012-08-19	13	157	6	-74.0	732.0
4	46	2012-08-20	14	157	6	-74.0	740.0

Step 4: Prepare Data for Analysis

Examine the data types that are currently associated with each column. If necessary, cast (or convert) the data to the appropriate `datetime` , `int` , or `float` data types.

> **Hint** You can use the Pandas `astype` and `to_datetime` methods to accomplish this task.

In [7]:

```
# Examine data type of each column
mars_data_df.dtypes
```

Out[7]:

```
id                  object
terrestrial_date    object
sol                 object
ls                  object
month               object
min_temp            object
pressure            object
dtype: object
```

In [8]:

```
# Change data types for data analysis
mars_data_df['terrestrial_date'] = pd.to_datetime(mars_data_df['terrestrial_date'])
mars_data_df['sol'] = mars_data_df['sol'].astype(int)
mars_data_df['ls'] = mars_data_df['ls'].astype(int)
mars_data_df['min_temp'] = mars_data_df['min_temp'].astype(float)
mars_data_df['pressure'] = mars_data_df['pressure'].astype(float)
mars_data_df['month'] = mars_data_df['month'].astype(int)
```

In [9]:

```
# Confirm type changes were successful by examining data types again
mars_data_df.dtypes
```

Out[9]:

```
id                          object
terrestrial_date    datetime64[ns]
sol                          int32
ls                           int32
month                        int32
min_temp                   float64
pressure                   float64
dtype: object
```

Step 5: Analyze the Data

Analyze your dataset by using Pandas functions to answer the following questions:

1. How many months exist on Mars?
2. How many Martian (and not Earth) days worth of data exist in the scraped dataset?
3. What are the coldest and the warmest months on Mars (at the location of Curiosity)? To answer this question:
   - Find the average the minimum daily temperature for all of the months.
   - Plot the results as a bar chart.
4. Which months have the lowest and the highest atmospheric pressure on Mars? To answer this question:
   - Find the average the daily atmospheric pressure of all the months.
   - Plot the results as a bar chart.
5. About how many terrestrial (Earth) days exist in a Martian year? To answer this question:
   - Consider how many days elapse on Earth in the time that Mars circles the Sun once.
   - Visually estimate the result by plotting the daily minimum temperature.

In [10]:

```
# 1. How many months are there on Mars?
monthly_data_count = mars_data_df.groupby('month').size()
print(monthly_data_count)
```

```
month
1     174
2     178
3     192
4     194
5     149
6     147
7     142
8     141
9     134
10    112
11    138
12    166
dtype: int64
```

In [11]:

```
# 2. How many Martian days' worth of data are there?
num_martian_days = mars_data_df['sol'].nunique()
num_martian_days
```

Out[11]:

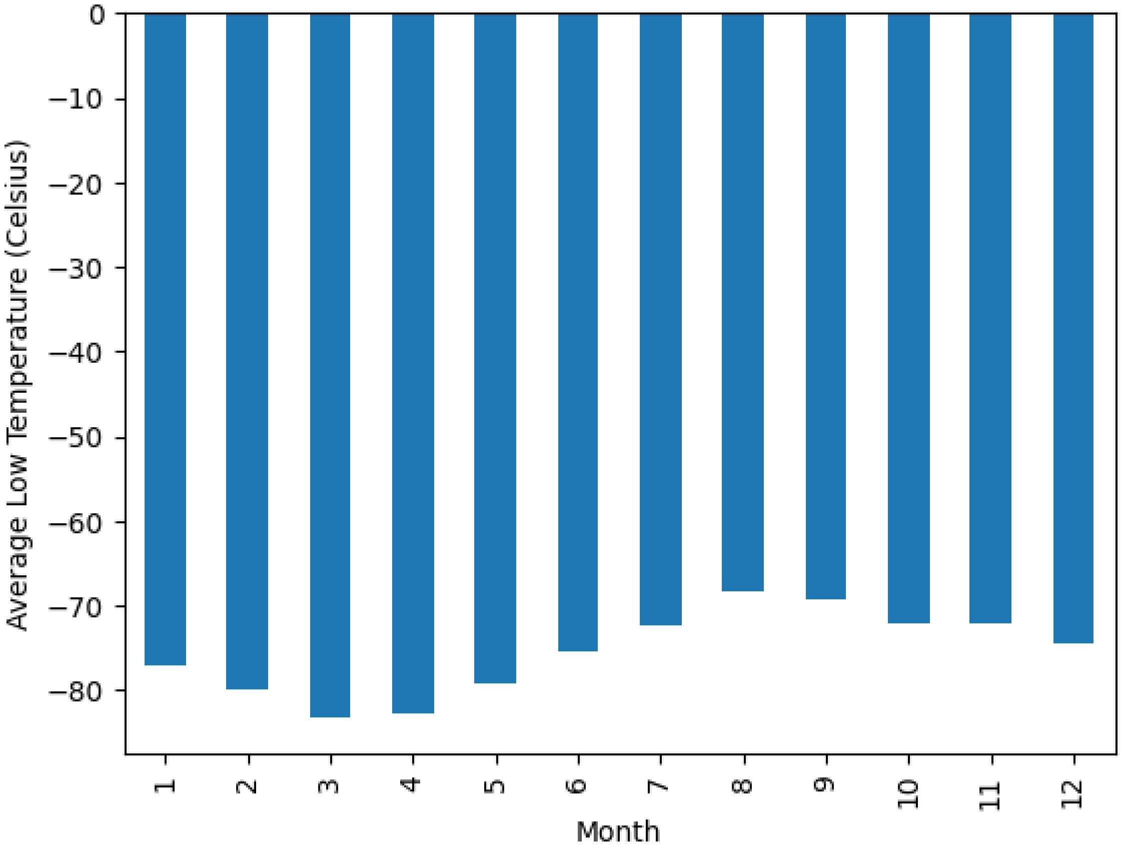
```
1867
```

In [12]:
```
# 3. What is the average low temperature by month?
average_low_temp_by_month = mars_data_df.groupby('month')['min_temp'].mean()
print(average_low_temp_by_month)
```

```
month
1    -77.160920
2    -79.932584
3    -83.307292
4    -82.747423
5    -79.308725
6    -75.299320
7    -72.281690
8    -68.382979
9    -69.171642
10   -71.982143
11   -71.985507
12   -74.451807
Name: min_temp, dtype: float64
```

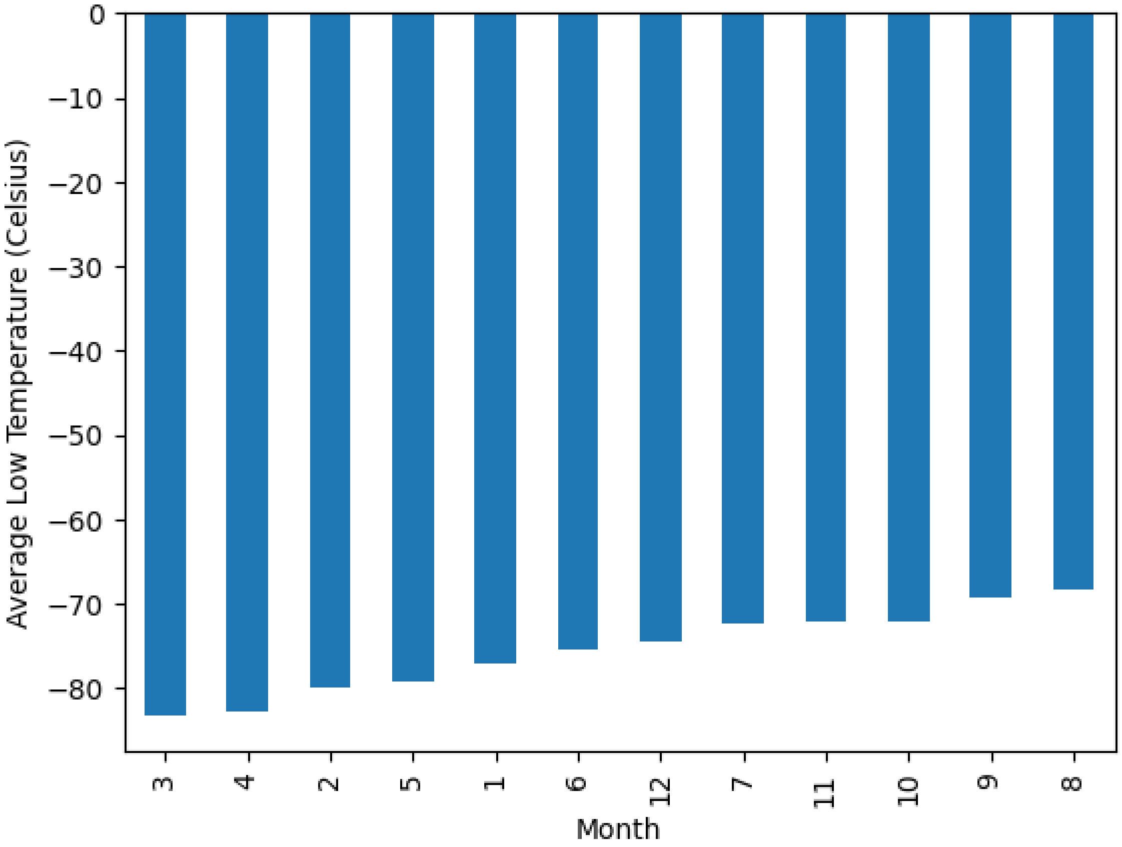
In [13]:
```
# Plot the average temperature by month
coldest_month = average_low_temp_by_month.idxmin()
warmest_month = average_low_temp_by_month.idxmax()

average_low_temp_by_month.plot(kind='bar')
plt.xlabel('Month')
plt.ylabel('Average Low Temperature (Celsius)')
plt.show()
```



In [14]:
```
# Identify the coldest and hottest months in Curiosity's location
average_low_temp_by_month = average_low_temp_by_month.sort_values()

average_low_temp_by_month.plot(kind='bar')
plt.xlabel('Month')
plt.ylabel('Average Low Temperature (Celsius)')
plt.show()
```

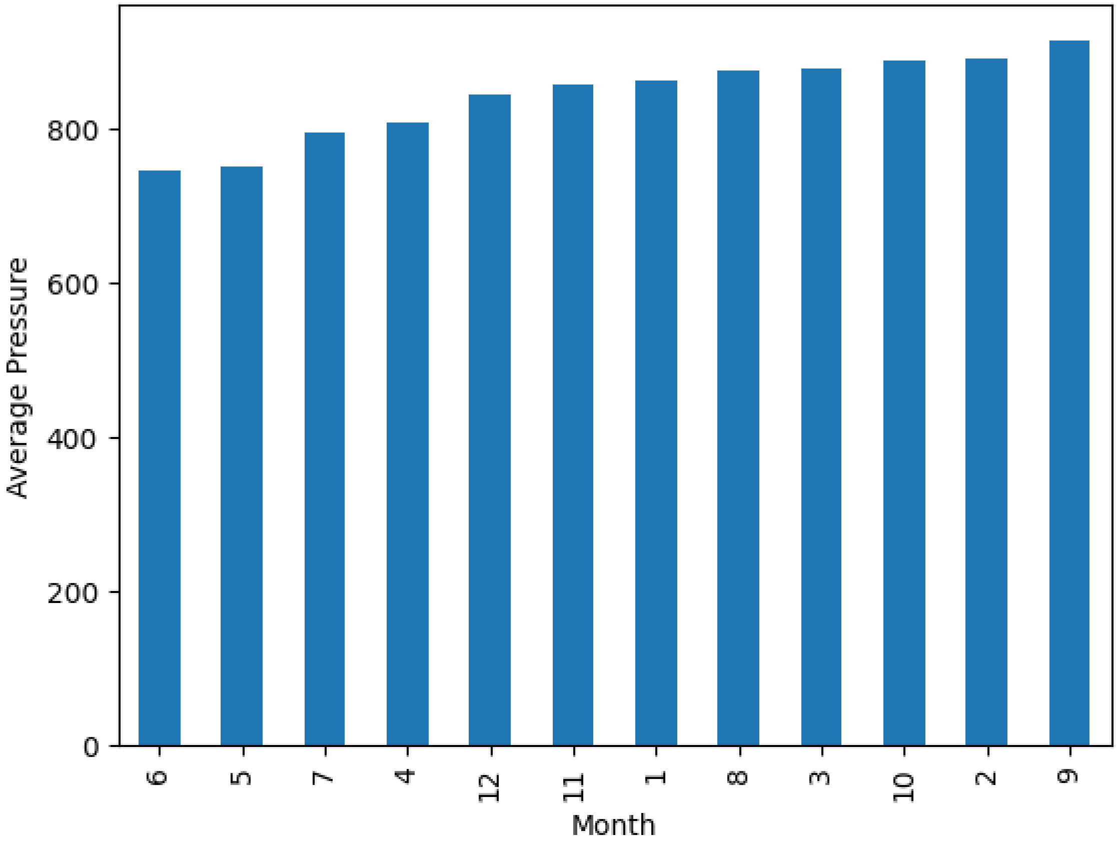


In [15]:
```
# 4. Average pressure by Martian month
avg_pressure_by_month = mars_data_df.groupby('month')['pressure'].mean()
avg_pressure_by_month
```

Out[15]:
```
month
1     862.488506
2     889.455056
3     877.322917
4     806.329897
5     748.557047
6     745.054422
7     795.105634
8     873.829787
9     913.305970
10    887.312500
11    857.014493
12    842.156627
Name: pressure, dtype: float64
```

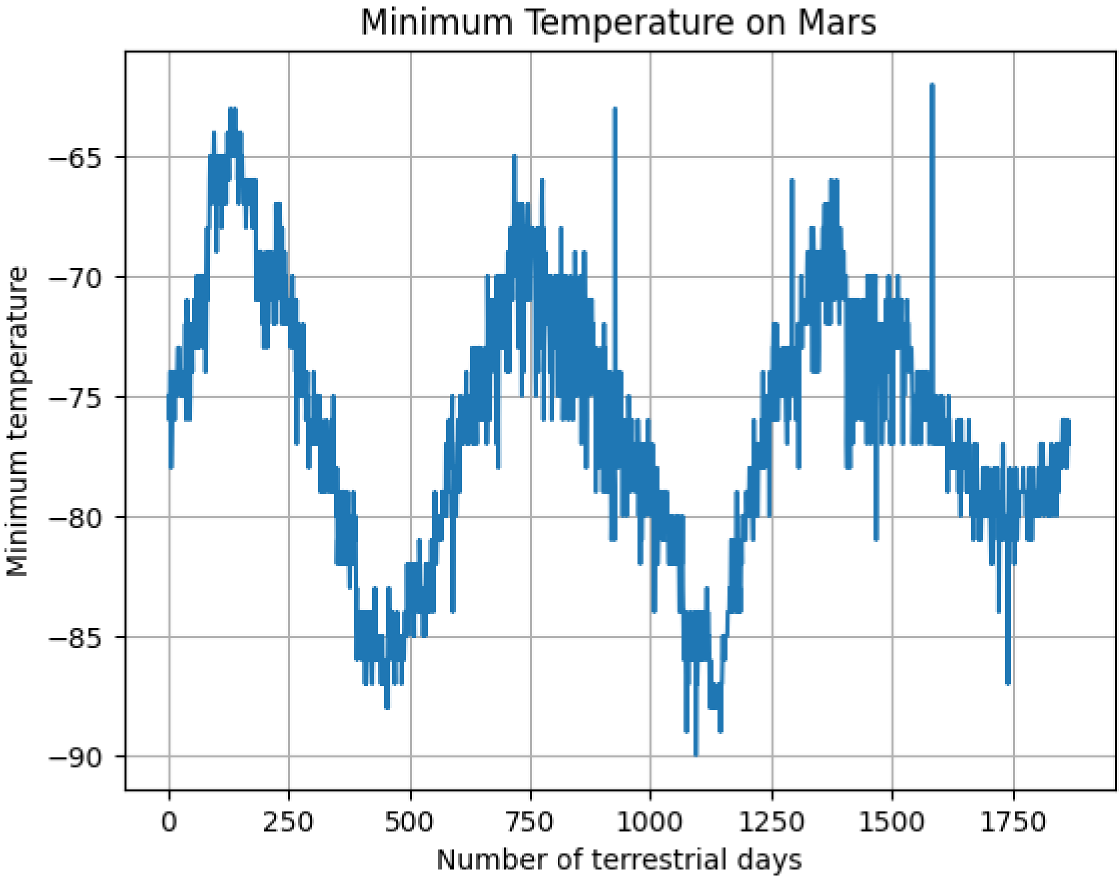
In [16]:

```
# Plot the average pressure by month
avg_pressure_by_month = avg_pressure_by_month.sort_values()
avg_pressure_by_month.plot(kind='bar')
plt.xlabel('Month')
plt.ylabel('Average Pressure')
plt.show()
```



In [17]:

```
# 5. How many terrestrial (earth) days are there in a Martian year?
plt.plot(range(1, len(mars_data_df) + 1), mars_data_df[’min_temp’])
plt.xlabel(’Number of terrestrial days’)
plt.ylabel(’Minimum temperature’)
plt.title(’Minimum Temperature on Mars’)
plt.grid(True)
plt.show()
```



On average, the third month has the coldest minimum temperature on Mars, and the eighth month is the warmest. But it is always very cold there in human terms!

Atmospheric pressure is, on average, lowest in the sixth month and highest in the ninth.

The distance from peak to peak is roughly 1425-750, or 675 days. A year on Mars appears to be about 675 days from the plot. Internet search confirms that a Mars year is equivalent to 687 earth days.

Step 6: Save the Data

Export the DataFrame to a CSV file.

In [18]:

```
# Write the data to a CSV
mars_data_df.to_csv(’./data/mars_data_processed.csv’, index=False)
```

In [19]:

```
browser.quit()
```

In [ ]: