Activities

1 Module 5.2 Activities

This notebook contains activities for Module 5.2 of the Civic Data Education Series.

```python
# load the Python libraries for working with data
import pandas

# read the csv and save the data in a variable called wifi_data
wifi_data = pandas.read_csv("clp-public-wifi.csv")

# load the library Locations dataset
library_locations = pandas.read_csv("clp-library-locations.csv")

# Load the totals per neighborhood dataset
totals_per_neighborhood = pandas.read_csv("totals_per_neighborhood.csv")
```

1.1 Activity 1

Overview Modify the code below to create different charts for the years 2016 and 2018 respectively.

Step 1: Generate the 2017 chat.

Run the code cell below by selecting it and typing shift-return or by pressing the play button in the menu bar above. This will generate a chart with data from 2017.

Step 2: Change the year in the code cell below.

In the code below, modify the `year` variable by changing number 2017 to be 2016 and then execute the code cell. Consider the chart and how it changes from the previous chart.

Step 3: Change the year again and re-generate the chart.

Modify the year to be 2018, run the cell again, and consider the new chart representing the 2018 data.

```python
# Specify the year in a variable
year = 2018

# Select data for specified year, group by library/month, and aggregate by the
 ↪adding together the minutes
```

```
wifi_data_subset = wifi_data[wifi_data['Year'] == year].groupby(["Name",␣
  ↪"Month"], as_index=False)["WifiMinutes"].sum()

# Reshape the data so it is easier to plot by Month
reshaped_wifi_data_subset = wifi_data_subset.pivot_table(index="Month",␣
  ↪columns="Name", values="WifiMinutes")

# plot the data
ax = reshaped_wifi_data_subset.plot(figsize=(10,8),
                                    title=f"Carnegie Library of Pittsburgh␣
  ↪{year} Wi-Fi Usage in Minutes",
                                    colormap="tab20",
                                    fontsize=14)
# clean up the text
ax.xaxis.label.set_size(16)
ax.title.set_size(20)
ax.legend(loc="upper right", bbox_to_anchor=(1.6,1))
ax.ticklabel_format(style="plain")

# add the Month abreviations instead of numbers
ax.set_xticks(ticks=range(1,13),labels=["Jan", "Feb", "Mar", "Apr", "May",␣
  ↪"Jun", "Jul","Aug","Sep","Oct","Nov","Dec"]);
```

1.1.1 Activity 1 Reflection

Consider the following questions as you create different charts for each year: - What do we see about these other charts? - Are there meaningful differences in the different years? - What is happening in the 2018 data? What do you think that means?

1.2 Activity 2 - Data Driven Question

Overview: In this activity you will be tasked with answering a data driven questions.

*What is the largest of the CLP Library Locations?

Step 1: Inspect the data

Run the code cell below and look at all of the columns in the library locations dataset. What column do you think will be most useful for answering the questions above?

```
[ ]: # display the library locations data
     library_locations.head()
```

Step 2: Computing the Answer

Run the code cells below. Each cell contains an attempt at answering our data driven question.

```
[ ]:  # Attempt Number 1
      library_locations["SqFt"].sum()
```

```
[ ]:  # Attempt Number 2
      library_locations["SqFt"].max()
```

```
[ ]:  # Attempt Number 3
      library_locations.sort_values("SqFt", ascending=False)[["Name", "SqFt"]]
```

1.2.1 Activity 2 Reflection

Which of the three attempts do you think best answered the question, what is the largest CLP Library location? Was there more than one way to answer our question? If yes, was one solution better than another? What other questions could we answer with the operations we performed on the library locations data?

Consider the following additional questions, have we already answered them or what operations would you do in order to answer them: - What is the smallest? - What is the total size of all CLP library locations? - What is the average size of the CLP Library locations?

Feel free to use the code cell below to try and answer these additional questions.

```
[ ]:  # Bonus area for answering the additional questions
```

1.3 Activity 3 - Create a Data Dictionary

Overview: This module created a new dataset `totals_per_neighborhood.csv`. In this activity you will create a data dictionary for the new dataset. Data dictionaries are an important part of a dataset's metadata. The Data dictionary provides a description of all the *fields* or columns in a dataset and the meaning of those data elements.

Step 1: Display the data.

Run the code cell below to display the first five rows of the `totals_per_neighborhood` dataset.

```
[ ]:  # display the totals_per_neighborhood dataset
      totals_per_neighborhood.head()
```

Step 2: Identify the Fields

This dataset has three columns or "fields" - What are their names? - How would you describe their contents? - What are the data types for each field, that is, are they numbers, categories, or other values? What does that data value represent?

On a separate sheet of paper or in the text cell below, answer the questions above for each column in the `totals_per_neighborhood` dataset. Add your data dictionary here:

1.3.1 Activity 3 Reflection

Put yourself in the position of someone who has never encountered this dataset before. What information would the need to use these data? Given the data and the data dictionary you created, is that sufficient? Review the modules on Metadata and Data Documentation, what insights from those modules would help in documenting the `totals_per_neighborhood` dataset?

Reflect upon what you have learned from this module the the series as a whole. What skills and literacies do you think are necessary to create, maintain, and use civic open data?