Dana'sNotebook

April 18, 2023

Introduction:

Our final project focuses on the best neighborhood in Pittsburgh based on the amount of things there is to do outdoors. So, the neighborhood with the most variety of things to do outside is classified as the overall best neighborhood. When searching through the data types, we just wanted to focus on the activities and places there are throughout Pittsburgh, such as resturants, malls, rest areas, play areas, or even places to hang out with friends. We first explored some alternatives based on this criteria. We started off by wanting to just focus on resturants (fast food, food facilities, etc.), but we came to the conclusion that it was not broad enough when determining the best neighborhood. We ended up deciding on a metric that came down to what neighborhood had the most to do. This provided a broader variety while still connecting each data set to one another: the most playgrounds, greenspaces, and courts/rinks!

The Metric:

For our project, our metric is determined by what neighborhood has the most to do, which will be the best, and what neighborhood has the least to do, which will be the worst. The features we are planning to measure is the amount of playgrounds, greenspaces, and courts/rinks each neighborhood in Pittsburgh has. Whichever one has the most or around the most of each feature will be claimed as the best neighborhood to live in. For this project, the datasets we are using include: 1) City of Pittsburgh Playgrounds 2) Operation Green Spaces 3) City of Pittsburgh Courts and Rinks

---

Dataset #1: City of Pittsburgh Playgrounds: (Metric: Playgrounds)

[18]:
```
import pandas as pd

playgrounds = pd.read_csv("https://data.wprdc.org/datastore/dump/
 ↪47350364-44a8-4d15-b6e0-5f79ddff9367")
playgrounds.head()
```

[18]:

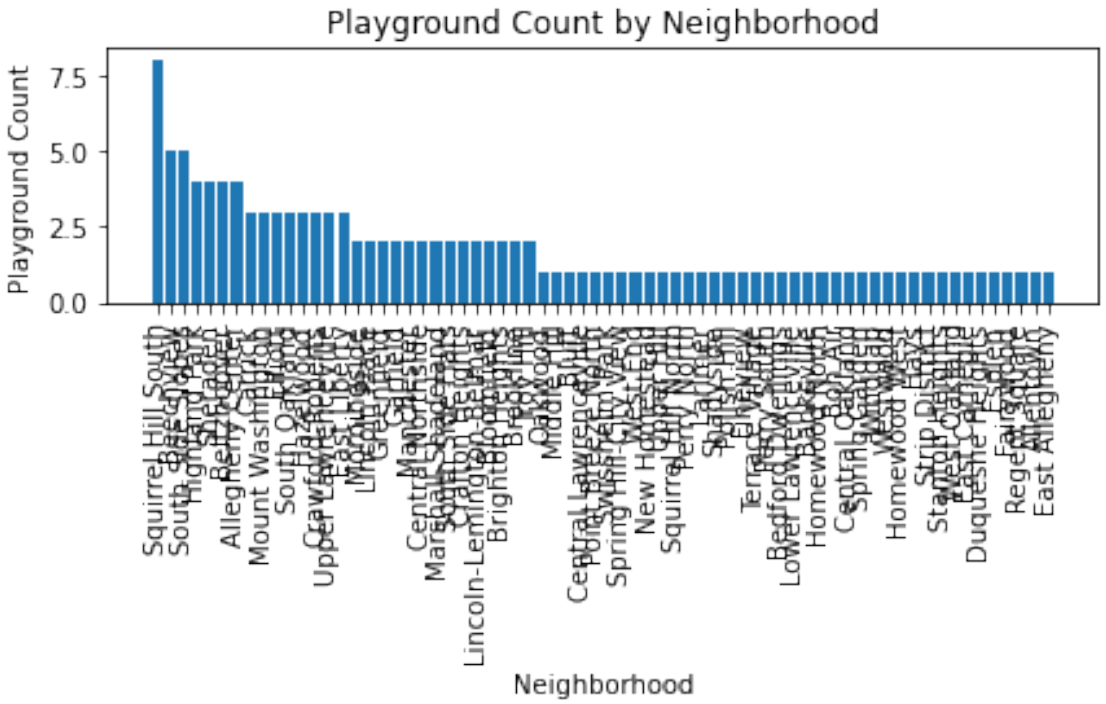
	id	name	type	maintenance_responsibility	\
0	731501774	Able Long Playground	NaN	Parks - Western	
1	1461276747	Albert Graham Playground	NaN	Parks - Schenley	
2	1860709784	Alpine Playground	NaN	Parks - Northern	
3	1770671485	Alton Playground	NaN	Parks - Western	
4	18942817	Ammon Playground	NaN	Parks - Schenley	

```
                      park       street  \
0           Able Long Park    COAST AVE
1  Albert Turk Graham Park  FORESIDE PL
2      Alpine Gardens Park   ALPINE AVE
3               Alton Park   ANDICK WAY
4               Ammon Park    MEMORY LN

                                                 image       neighborhood  \
0  https://tools.wprdc.org/images/pittsburgh/play…          Beechview
1  https://tools.wprdc.org/images/pittsburgh/play…   Crawford-Roberts
2  https://tools.wprdc.org/images/pittsburgh/play…  Central Northside
3  https://tools.wprdc.org/images/pittsburgh/play…          Beechview
4  https://tools.wprdc.org/images/pittsburgh/play…  Bedford Dwellings

   council_district  ward        tract  public_works_division  pli_division  \
0                 4    19  42003192000                      5            19
1                 6     3  42003030500                      3             3
2                 6    25  42003250300                      1            25
3                 4    19  42003191600                      5            19
4                 6     5  42003050900                      3             5

   police_zone fire_zone   latitude  longitude
0            6      4-28  40.408365 -80.028445
1            2       2-1  40.440519 -79.984137
2            1      1-21  40.457707 -80.012952
3            6      4-28  40.414137 -80.021605
4            2       2-5  40.449037 -79.978064
```

[20]:
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('https://data.wprdc.org/datastore/dump/
 ↪47350364-44a8-4d15-b6e0-5f79ddff9367')

#Group the data by neighborhood and count the number of playgrounds in each␣
 ↪neighborhood
neighborhood_counts = df['neighborhood'].value_counts()

#Get the neighborhood with the most playgrounds
neighborhood_with_most_playgrounds = neighborhood_counts.idxmax()

#Get the count of playgrounds in the neighborhood with the most playgrounds
most_playgrounds_count = neighborhood_counts.max()
print(f"The neighborhood with the most playgrounds is␣
 ↪'{neighborhood_with_most_playgrounds}' with {most_playgrounds_count}␣
 ↪playgrounds.")
```

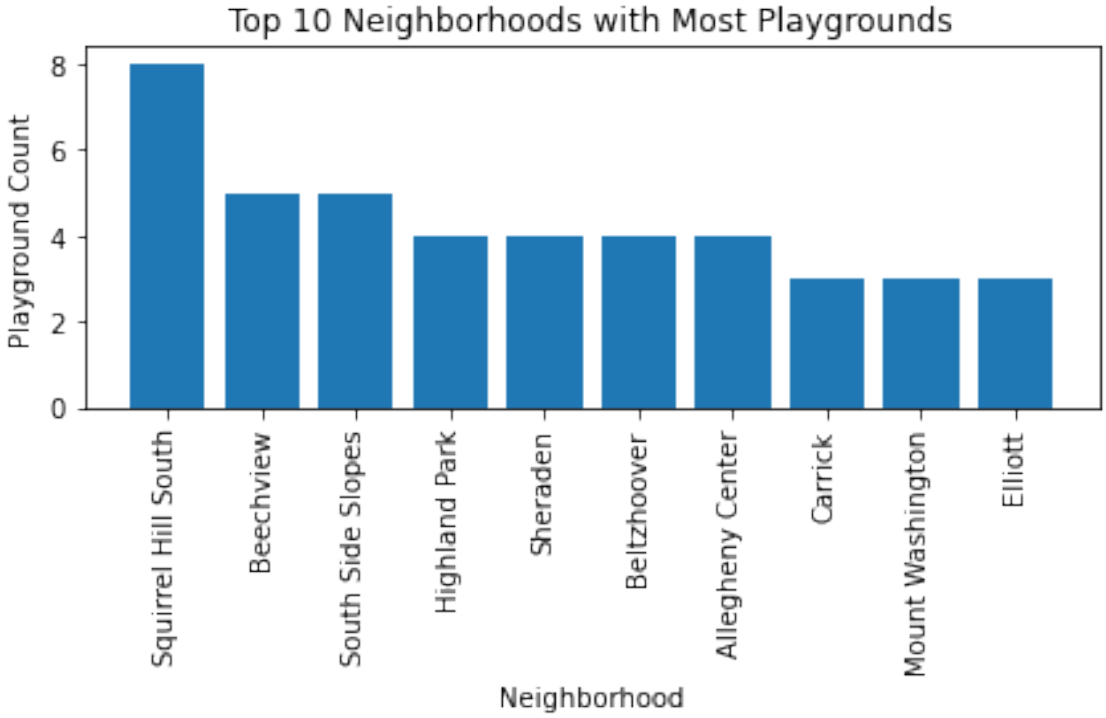
```
#Create a bar graph to visualize the neighborhood with the most playgrounds
plt.bar(neighborhood_counts.index, neighborhood_counts.values)
plt.xlabel('Neighborhood')
plt.ylabel('Playground Count')
plt.title('Playground Count by Neighborhood')
plt.xticks(rotation='vertical')
plt.tight_layout()
plt.show()
```

The neighborhood with the most playgrounds is 'Squirrel Hill South' with 8 playgrounds.



[10]:
```
neighborhood_counts = df['neighborhood'].value_counts()
top10_neighborhoods = neighborhood_counts.head(10)

print("Top 10 neighborhoods with the most playgrounds:")
print(top10_neighborhoods)
plt.bar(top10_neighborhoods.index, top10_neighborhoods.values)
plt.xlabel('Neighborhood')
plt.ylabel('Playground Count')
plt.title('Top 10 Neighborhoods with Most Playgrounds')
plt.xticks(rotation='vertical')
plt.tight_layout()
plt.show()
```

```
Top 10 neighborhoods with the most playgrounds:
Squirrel Hill South    8
Beechview              5
South Side Slopes      5
Highland Park          4
Sheraden               4
Beltzhoover            4
Allegheny Center       4
Carrick                3
Mount Washington       3
Elliott                3
Name: neighborhood, dtype: int64
```



Dataset #2: Operation Green Spaces (Metric: Green Spaces)

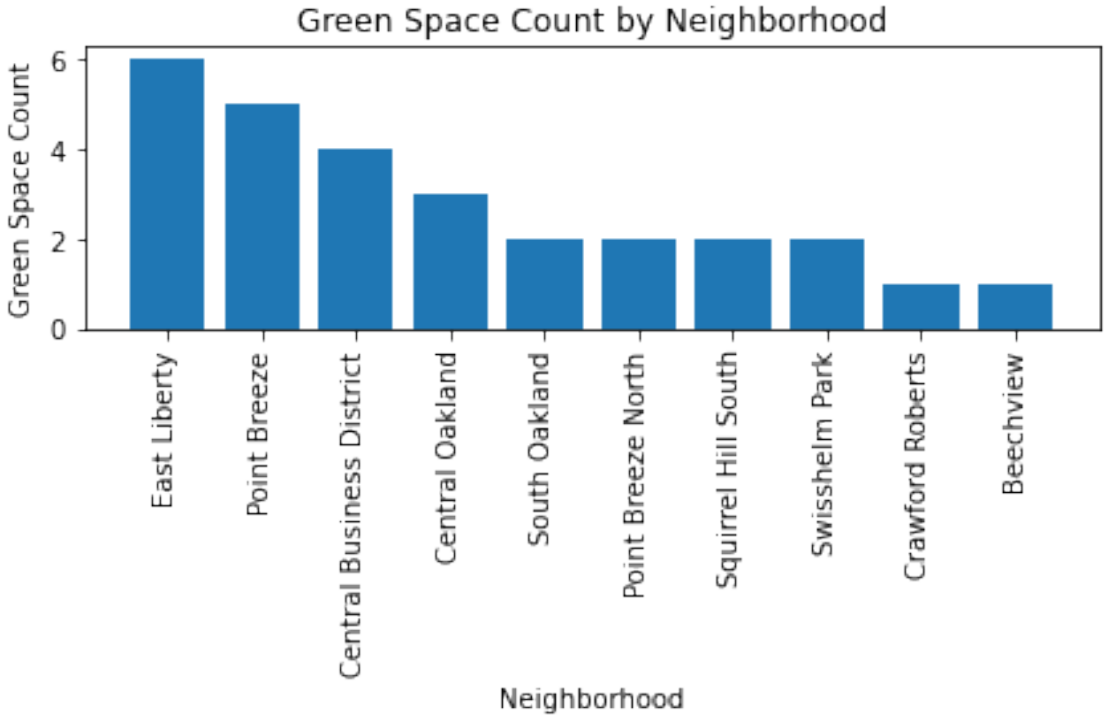
[1]:
```
import pandas as pd

playgrounds = pd.read_csv("https://data.wprdc.org/datastore/dump/
 ↪5cf7163b-f76c-47ef-b48d-a16366ac1835")
playgrounds.head()
```

[1]:
```
  Responsibility_Parks or Operations             Facility  Division  \
0                          Operations     Bates St Triangle       3.0
1                          Operations  Baum Negley Triangle       2.0
2                          Operations         Bedford Strip       3.0
3                          Operations    Beechview Monument       5.0
4                          Operations    Brookline Monument       4.0

   Total Acreage  Maintainable Acreage  Shelter  Rec Center  Senior Center  \
0            0.2                   NaN      NaN         NaN            NaN
1            0.2                   NaN      NaN         0.0            0.0
2            0.3                   NaN      NaN         0.0            0.0
3            0.1                   NaN      NaN         NaN            NaN
4            0.1                   NaN      NaN         0.0            0.0

   Pool  Spray Pool  …  Grandstand  Control Link for Basketball Courts  \
0   NaN         NaN  …         NaN                                  No
1   NaN         NaN  …         NaN                                  No
2   NaN         NaN  …         NaN                                  No
3   NaN         NaN  …         NaN                                  No
4   NaN         NaN  …         NaN                                  No

   Control Link for Tennis Court  Control Linkl for Hockey Court  \
0                             No                             No
1                             No                             No
2                             No                             No
3                             No                             No
4                             No                             No

   Dek Hockey Court  Dog Park  Spray Park  Ownership of Facility  Benches  \
0               NaN       NaN         NaN                    NaN      NaN
1               NaN       NaN         NaN                    NaN      0.0
2               NaN       NaN         NaN                    NaN      0.0
3               NaN       NaN         NaN                    NaN      2.0
4               NaN       NaN         NaN                    NaN      0.0

   Tables
0     NaN
1     0.0
2     0.0
3     NaN
4     0.0

[5 rows x 45 columns]
```

[16]:
```
# Group the data by neighborhood and count the number of green spaces in each␣
↪neighborhood
neighborhood_counts = df['Neighborhood'].value_counts()
```

```
neighborhood_with_most_green_spaces = neighborhood_counts.idxmax()

# Get the count of green spaces in the neighborhood with the most green spaces
most_green_spaces_count = neighborhood_counts.max()
print(f"The neighborhood with the most green spaces is␣
 ↪'{neighborhood_with_most_green_spaces}' with {most_green_spaces_count} green␣
 ↪spaces.")

# Create a bar graph to visualize the neighborhoods with the most green spaces
plt.bar(neighborhood_counts.index[:10], neighborhood_counts.values[:10])
plt.xlabel('Neighborhood')
plt.ylabel('Green Space Count')
plt.title('Green Space Count by Neighborhood')
plt.xticks(rotation='vertical')
plt.tight_layout()
plt.show()
```

```
The neighborhood with the most green spaces is 'East Liberty' with 6 green
spaces.
```



[15]:
```
neighborhood_counts = df['Neighborhood'].value_counts()
top_10_neighborhoods = neighborhood_counts[:10]
```

```
top_10_counts = top_10_neighborhoods.values.tolist()

print("Top 10 neighborhoods with the most green spaces:")
for i, neighborhood in enumerate(top_10_neighborhoods.index.tolist()):
    print(f"{i+1}. {neighborhood}: {top_10_counts[i]}")
```

```
Top 10 neighborhoods with the most green spaces:
1. East Liberty: 6
2. Point Breeze: 5
3. Central Business District: 4
4. Central Oakland: 3
5. South Oakland: 2
6. Point Breeze North: 2
7. Squirrel Hill South: 2
8. Swisshelm Park: 2
9. Crawford Roberts: 1
10. Beechview: 1
```

The Best Neighborhood:

Based on the metric that we determined in the previous sections (most things to do outdoors), we found the best neighborhood in Pittsburgh to be Squirrel Hill South. Based on the outcomes from all our data, Squirrel Hill South was ranked the highest amongst all three of our datasets. By ranking the top 10 highest neighborhoods in each dataset, Squirrel Hill South came out as #1 in 2 of the datasets (Parks and Courts/Rinks) and #5 to #8 in the reminding dataset (Greenspaces).

To further compare, in the dataset based around the amount of parks, Squirrel Hill South had a total of 8 parks. This is greatly different from the neighborhood that was #2, which only 5 parks. Adding on, in the dataset based around the highest amount of greenspaces, the top, East Liberty had a total of 6, while Squirrel Hill South had 2. While there is a drastic number difference, it was tied with the spot #5 that also had 2 greenspaces, South Oakland. Alongside this, East Liberty did not even place in the dataset that ranked the highest number of parks, and neither did any of the other top 4 rankings in the greenspace dataset. Overall, Squirrel Hill did have 4 less green spaces, but topped #1 in both of the other datasets. So even though it does not compete as top in every dataset, by being #1 in two and tying for the #5 spot in the other, it can be determined that overall, Squirrel Hill is the best neigborhood in Pittsburgh.

Conclusion:

Dana: Although I have not traveled a lot throughout Pittsburgh since I have been here, my favorite neighborhood is Mount Washington. Mount Washington is very different from Squirrel Hill. Squirrel Hill is known for its two large public parks, the residental streets with trees, and education, Mount Washington is known for its beautiful skyline and resturants. I enjoy going to Mount Washington to go out to eat or to sight see, and Squirrel Hill seems more for its parks, bookstores, and boutiques.