FinalProject

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. The neighborhood with the most variety of things to do outdoors is classified as the best neighborhood in Pittsburgh. When searching through the data types, we wanted to focus on the activities and places there are throughout Pittsburgh, outdoor wise, that can be used for resting, playing, or just hanging out with friends. We first explored some alternatives, wanting to just focus on resturants (fast food, food facilities, etc.). However we thought this was not broad enough, and wanted a broader approach to find the most specific result possible. The metric we decided with came down to what neighborhood had the most to do: 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 neighborhood, and what neighborhood has the least to do, which will be the worst city. 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: Highest number of Playgrounds)

[6]:
```
import pandas as pd

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

[6]:

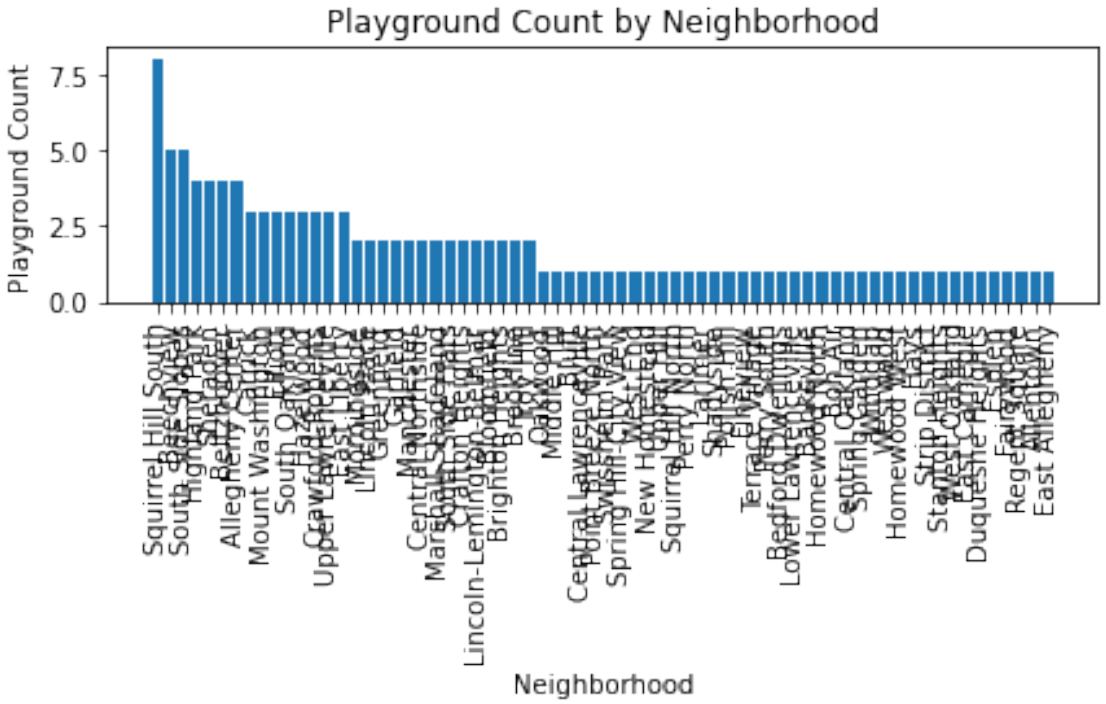
	id	name	type	maintenance_responsibility	\
0	731501774	Able Long Playground	NaN	Parks - Western	
1	1461276747	Albert Graham Playground	NaN	Parks - Schenley	
2	1860709784	Alpine Playground	NaN	Parks - Northern	
3	1770671485	Alton Playground	NaN	Parks - Western	
4	18942817	Ammon Playground	NaN	Parks - Schenley	

```
                    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
```

[7]:
```
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
```

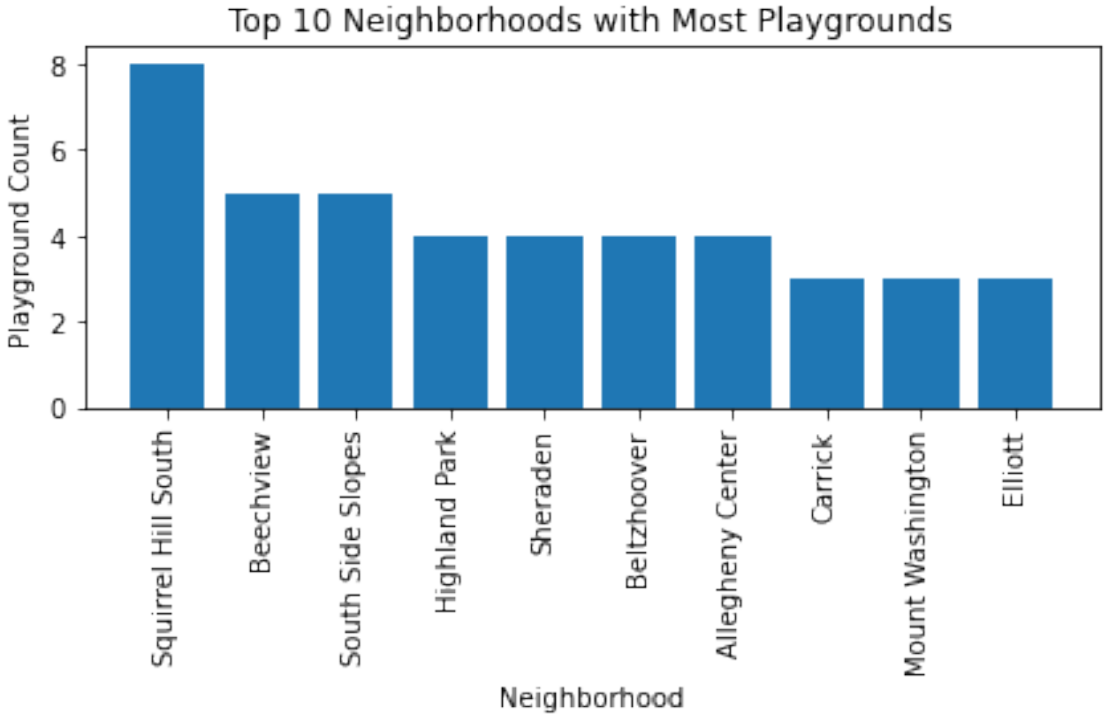
```
plt.bar(neighborhood_counts.index, neighborhood_counts.values)
plt.xlabel('Neighborhood')
plt.ylabel('Playground Count')
plt.title('Playground Count by Neighborhood')
plt.xticks(rotation='vertical')
plt.tight_layout()
plt.show()
```

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
```



[20]:
```python
import pandas as pd
import folium

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

# Filter the data for playgrounds in the Shadyside neighborhood
shadyside_playgrounds = df[df['neighborhood']=='Shadyside']

# Create a map of the city of Pittsburgh
pittsburgh_map = folium.Map(location=[40.4406, -79.9959], zoom_start=12)
```

```
# Add markers for the Shadyside playgrounds
for index, row in shadyside_playgrounds.iterrows():
    folium.Marker(location=[row['latitude'], row['longitude']],␣
 ↪popup=row['name']).add_to(pittsburgh_map)

# Display the map
pittsburgh_map
```

```
      ␣
↪---------------------------------------------------------------------------

       ModuleNotFoundError                       Traceback (most recent call␣
↪last)

       <ipython-input-20-584829d86613> in <module>
         1 import pandas as pd
   ----> 2 import folium
         3
         4 df = pd.read_csv('https://data.wprdc.org/datastore/dump/
↪47350364-44a8-4d15-b6e0-5f79ddff9367')
         5


       ModuleNotFoundError: No module named 'folium'
```

Dataset #2: Operation Green Spaces (Metric: Highest number of Green Spaces)

Dataset #3: City of Pittsburgh Courts and Rinks: (Metric: Highest number of Courts/Rinks)

The Best Neighborhood:

hdsisdsjdjdjdi

Conclusion:

Dana:

Laiba:

[ ]: