

# Neighborhoods In Pittsburgh with the most Health Complications



# Introduction

For this project we decided to find out which neighborhoods in Pittsburgh had the most health complications/worst access to healthcare. Initially, our goal was to find the neighborhood with the most fast-food restaurants, the poorest air quality, and other factors which might affect the health of an individual. However, after researching for viable datasets, we learned that there was far more data on actual health complications and accessible healthcare facilities. In the effort of using the best data, we decided to use the metrics: obesity rates, diabetes rates, and proximity to hospitals.

# Metric no 1. Obesity Rates

The First Dataset we used was the **Obesity Rates in Allegheny County**. This data set provided us with the estimated percentage of residents that struggle with obesity. The process of determining the neighborhood with the highest estimated obesity rate required that we create a new data frame from the original to only include neighborhoods in the Pittsburgh municipality and then sort the data frame to provide the organized data.

```
] : import pandas as pd

# Load the data from a CSV file
df = pd.read_csv('ObesityData.csv')

# strip the data frame so it only contains the data (columns) that I want
stripped_df = df[['City Neighborhood', '2006-2010 estimate of obesity']]

#sort the data frame so that the obesity rates are in descending order.
new_df_sorted = stripped_df.sort_values(by='2006-2010 estimate of obesity', ascending=False)

# Reset the index to apply weight to eat Neighborhood (Find 1st place)
new_df_sorted = new_df_sorted.reset_index(drop=True)

pd.set_option('display.max_rows', None)

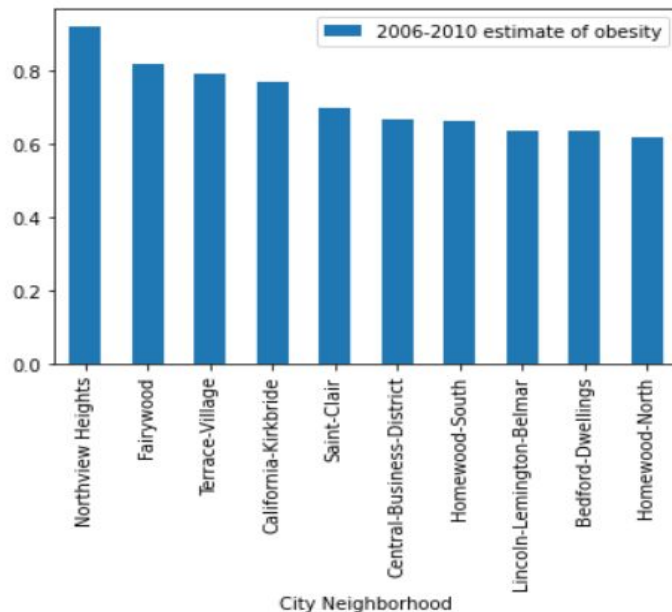
# Print the entire dataframe
new_df_sorted.head(10)
```

[1]:

	City Neighborhood	2006-2010 estimate of obesity
0	Northview Heights	0.922932
1	Fairywood	0.820311
2	Terrace-Village	0.793133
3	California-Kirkbride	0.771046
4	Saint-Clair	0.699500
5	Central-Business-District	0.668012
6	Homewood-South	0.660585
7	Lincoln-Lemington-Belmar	0.637099
8	Bedford-Dwellings	0.636222
9	Homewood-North	0.619584

```
5]: new_df_sorted.set_index('City Neighborhood').head(10).plot(kind='bar')
```

```
5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe71a92a5b0>
```



## Metric no. 1 Conculsion

Many of the neighborhood towards the top of our list in terms of estimated rate of obesity are neighborhoods that I am not familiar with. The Neighborhood that I enjoy the most definety Oakland.

## Metric no. 2 Hospitals

The second data set we used for this project was the approximate locations of hospitals in Pittsburgh. I will be using how many hospitals in each neighborhood and will be ranking them from the most hospitals to the least. I will be using the data set called "Hospitals 2015 Data", which will have the compiled locations of hospitals and their neighborhoods.

```
import pandas as pd

# Loading in data set

HD = pd.read_csv("Hospital_Data.tsv", sep = '\t')

# specifying the city I am looking for

word = 'Pittsburgh, PA'

# making it so that only the rows that have the word Pittsburgh, PA and

HDTSV = HD.head(10)[HD['Address'].str.contains(word)]

# printing

print(HDTSV)
```

	Facility	Address \
0	UPMC Children's	4401 Penn Avenue Pittsburgh, PA 15224
1	UPMC Magee	300 Halkett Street Pittsburgh, PA 15213
3	UPMC Mercy	1400 Locust Street Pittsburgh, PA 15219
4	UPMC Passavant	9100 Babcock Boulevard Pittsburgh, PA 15237
5	UPMC Presbyterian	3601 Fifth Avenue Pittsburgh, PA 15213
6	UPMC Shadyside	5230 Center Ave Pittsburgh, PA 15232
7	UPMC St. Margaret's	815 Freeport Road Pittsburgh, PA 15215

	Y	X
0	40.467315	-79.953590
1	40.436889	-79.960700
3	40.436137	-79.985285
4	40.573319	-80.014525
5	40.441544	-79.959356
6	40.454845	-79.939793
7	40.489286	-79.895994

```
<ipython-input-53-add7b1208f99>:13: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
HDTSV = HD.head(10)[HD['Address'].str.contains(word)]
```

0 = Bloomfield 1 = North Oakland 3 = Central Business District 4 = Borough of Franklin Park 5 = North Oakland 6 = Shadyside 7 = Downtown

## Conclusion:

From what I can see from the data above, it seems that out of the 90 suburbs in Pittsburgh, there is only 7 that have a hospital in their neighborhood. It also seems that North Oakland has the most hospitals out of all other neighborhoods.



## Metric no. 3 diabetes

The metric used to determine the best neighborhood in Pittsburgh is the diabetes rate. This is calculated by summing up the number of people in each neighborhood who have been diagnosed with diabetes or have high blood pressure or high body weight, which are risk factors for diabetes. The dataset used for this project is the diabetes\_all\_2016.csv file, which contains information about diabetes prevalence in different neighborhoods of Pittsburgh.

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the diabetes data
df_diabetes = pd.read_csv('Data/diabetes_all_2016.csv', index_col='CT')

# Load the CT to Neighborhood mapping data
df_places = pd.read_csv('Data/CT2Place.csv')

# merge the diabetes data with the CT to Neighborhood mapping data
df_merged = pd.merge(df_diabetes, df_places, on='CT', how='left')

# drop the rows where the Neighborhood value is missing
df_merged = df_merged.dropna(subset=['Neighborhood'])

# group by neighborhood and calculate the mean value of each column within each group
df_grouped = df_merged.groupby('Neighborhood').mean()

# reset the index to make "Neighborhood" a regular column
df_cleaned = df_grouped.reset_index()

df_grouped
```

Out[3]:

	Neighborhood	CT	BPAD	BPAN	BPAN2	BWAD	BWAN	BWAN2	BMAD	BMAN	BMAN2	SUMLEV	COUNTY	COUSUB	BLK%
0	Allegheny Center	562700.000000	760.000000	57.000000	47.000000	432.000000	33.000000	29.000000	328.000000	24.000000	18.000000	750.0	3.0	61000.0	2.000
1	Allegheny West	562700.000000	760.000000	57.000000	47.000000	432.000000	33.000000	29.000000	328.000000	24.000000	18.000000	750.0	3.0	61000.0	1.000
2	Allentown	180300.000000	1254.000000	101.000000	83.000000	735.000000	65.000000	53.000000	519.000000	36.000000	30.000000	750.0	3.0	61000.0	2.000
3	Arlington - Arlington Heights	160800.000000	1165.000000	80.000000	62.000000	606.000000	46.000000	32.000000	559.000000	34.000000	30.000000	750.0	3.0	61000.0	1.000
4	Banksville	202300.000000	1934.000000	184.000000	143.000000	1063.000000	87.000000	67.000000	871.000000	97.000000	76.000000	750.0	3.0	61000.0	2.500
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
68	Upper Lawrenceville	101100.000000	1080.000000	97.000000	65.000000	583.000000	48.000000	33.000000	497.000000	49.000000	32.000000	750.0	3.0	61000.0	1.873
69	West End	532478.378378	1445.054054	130.513514	102.594595	769.108108	62.810811	47.216216	675.945946	67.702703	55.378378	750.0	3.0	61000.0	1.054
70	West Oakland	40200.000000	487.000000	54.000000	36.000000	286.000000	38.000000	26.000000	201.000000	16.000000	10.000000	750.0	3.0	61000.0	1.515
71	Westwood	563100.000000	1995.000000	171.000000	128.000000	1061.000000	91.000000	71.000000	934.000000	80.000000	57.000000	750.0	3.0	61000.0	1.500
72	Windgap	563000.000000	1428.000000	152.000000	116.000000	821.000000	79.000000	64.000000	607.000000	73.000000	52.000000	750.0	3.0	61000.0	2.266

73 rows × 17 columns



- CT: census tract identifier
- BPAD: number of people with high blood pressure, adjusted for age and sex
- BPAN: number of people with high blood pressure, age-adjusted only
- BPAN2: number of people with high blood pressure, age and sex adjusted to the 2000 US population
- BWAD: number of people with obesity, adjusted for age and sex
- BWAN: number of people with obesity, age-adjusted only
- BWAN2: number of people with obesity, age and sex adjusted to the 2000 US population
- BMAD: number of people with diabetes, adjusted for age and sex
- BMAN: number of people with diabetes, age-adjusted only
- BMAN2: number of people with diabetes, age and sex adjusted to the 2000 US population



```
df_merged = df_cleaned
```

## The Best Neighborhood:

Using the approach outlined above, the best neighborhood in Pittsburgh in terms of diabetes rate is "South Shore", with a diabetes rate of 1. This neighborhood has the lowest diabetes rate of all the neighborhoods in Pittsburgh and is followed closely by "Central Oakland" with a diabetes rate of 5. The two neighborhoods are located near to each other and are the places Pitt Students visit the most.

```
# calculate the diabetes rate for each neighborhood
df_merged['diabetes_rate'] = (df_merged.filter(like='BP').sum(axis=1) + df_merged.filter(like='BW').sum(axis=1) + df_merged.filter(like='BM').sum(axis=1))/1000

# sort the neighborhoods by diabetes rate in ascending order
df_sorted = df_merged.sort_values(by='diabetes_rate')

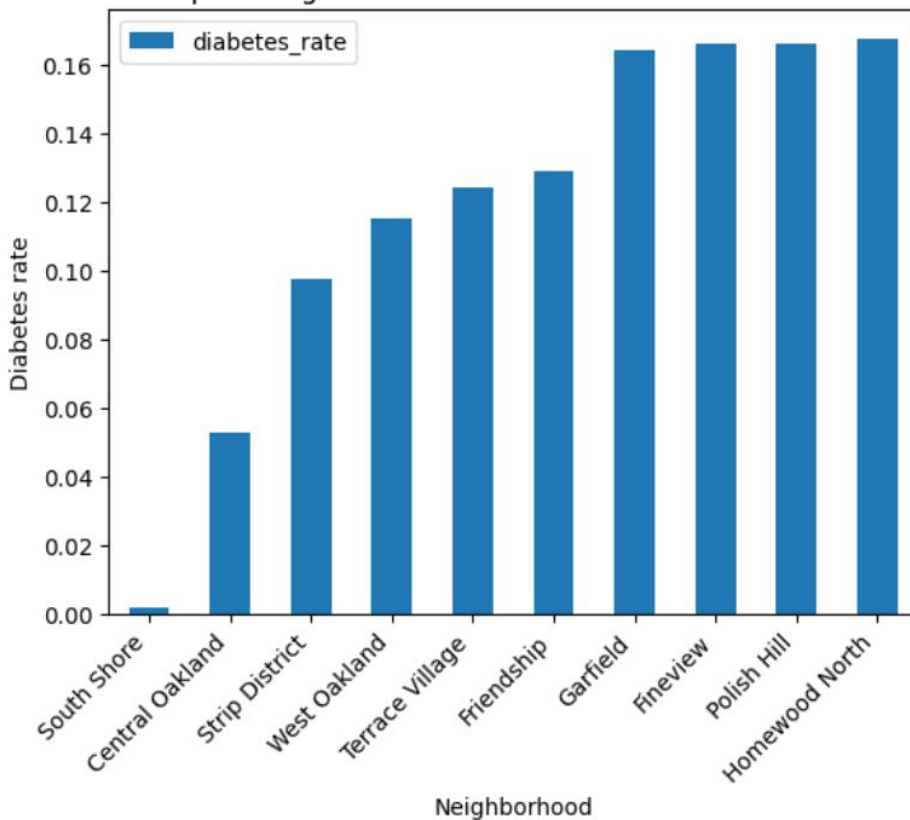
# create a bar plot showing the 10 neighborhoods with the lowest diabetes rate
df_top10 = df_sorted.head(10)
ax = df_top10.plot(kind='bar', x='Neighborhood', y='diabetes_rate')
ax.set_xticklabels(df_top10['Neighborhood'], rotation=45, ha='right')
ax.set_title('Top 10 neighborhoods with the lowest diabetes rate')
ax.set_xlabel('Neighborhood')
ax.set_ylabel('Diabetes rate')

plt.show()

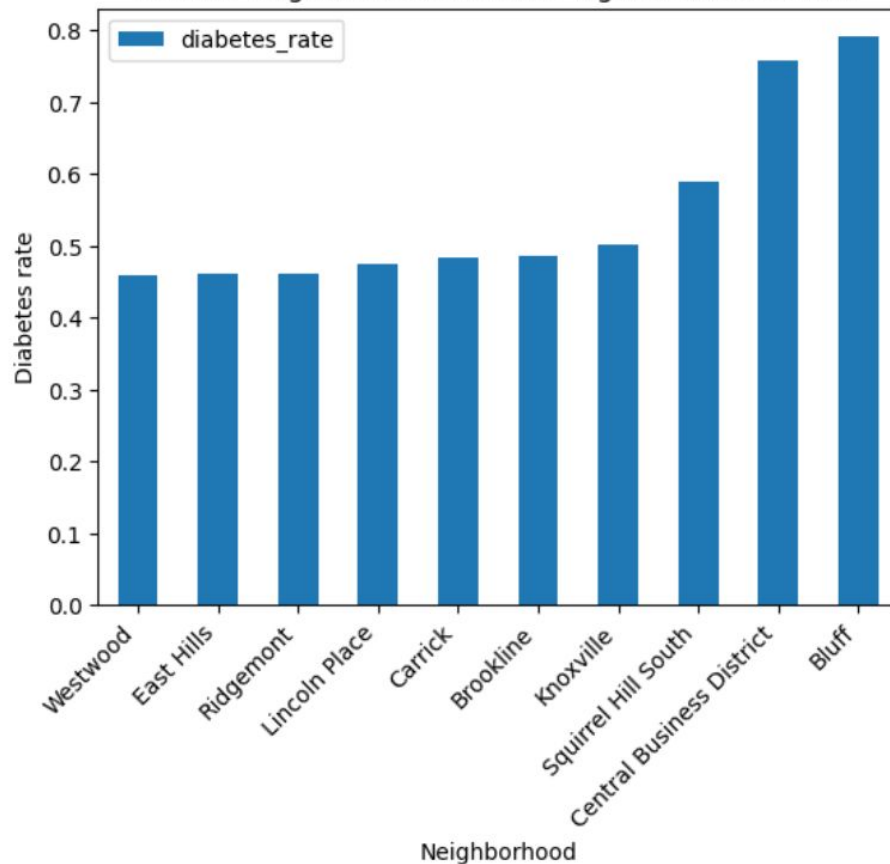
# create a bar plot showing the 10 neighborhoods with the highest diabetes rate
df_last10 = df_sorted.tail(10)
ax = df_last10.plot(kind='bar', x='Neighborhood', y='diabetes_rate')
ax.set_xticklabels(df_last10['Neighborhood'], rotation=45, ha='right')
ax.set_title('Last 10 neighborhoods with the highest diabetes rate')
ax.set_xlabel('Neighborhood')
ax.set_ylabel('Diabetes rate')

plt.show()
```

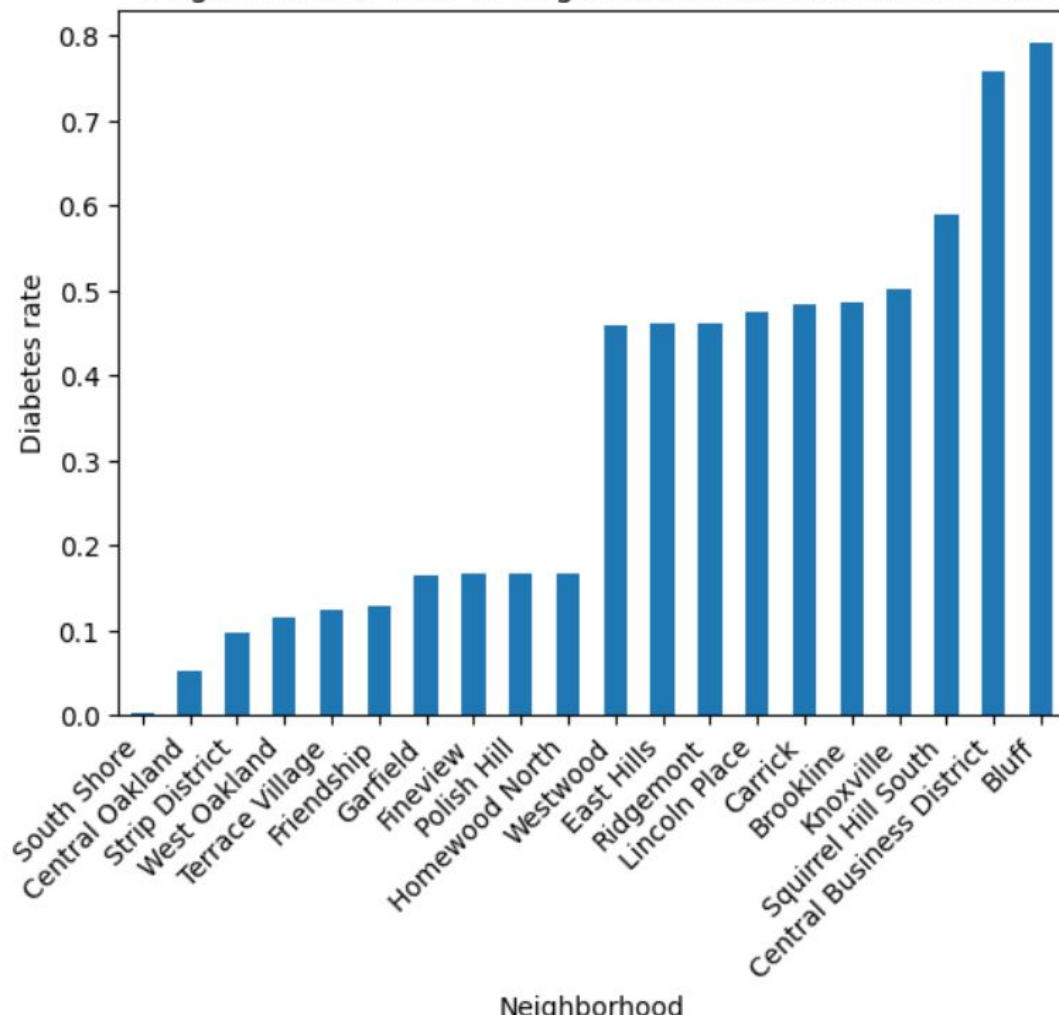
Top 10 neighborhoods with the lowest diabetes rate



Last 10 neighborhoods with the highest diabetes rate



Neighborhoods with the highest and lowest diabetes rate



From Our three datasets, we found the following data:

**Highest Obesity Rates:**

- 0 Northview Heights
- 1 Fairywood
- 2 Terrace-Village
- 3 California-Kirkbride
- 4 Central-Buisness-District

**Most cases of Diabetes:**

- 0 Bluff
- 1 Central-Buisness-District
- 2 Squirrel Hill South
- 3 Knoxville
- 4 Brookline

**Neighborhoods with Hospitals in them:**

- 0 Bloomfield
- 1 North Oakland
- 2 Central-Business-District
- 3 Borough of Franklin
- 4 ShadySide
- 5 Downtown

**The Neighborhood in Pittsburgh with the most health complications is . . .**

Central-Buisness-District

Although the Central Business District ranked among the few neighborhoods with hospitals in them, the fact that it was the only neighborhood in the top five rankings for both Obesity rate and Diabetes cases outways the third metric.