

## Introduction:

Seattle has always been a desirable city. Siyuan Liu is a young artist who plans to move to Seattle and open her first personal art exhibition there. I plan to use a data-driven method to analyze the cultural space resources of each block and neighbor in Seattle to present a recommendation to Siyuan. Through that process, I also hope can help other artists who faced the similar problem of moving to a new city.

## Data:

Cultural Space Inventory data from government website: <https://data.seattle.gov/>

Geocoder Seattle location data

Seattle location data from FourSquare API, including venue ID, venue Name, Latitude and Longitude, and category name.

## Methodology:

- Data Cleanup and re-grouping: retrieved data, clean up unwanted data, drop uncompleted and missing data, check data type, fix if needed, and add geographical coordinates.
- Top Neighborhood Cultural Space Inventory Counts.
- Use seaborn and matplotlib.pyplot to generate graph.
- Top 10 most common Cultural Space Inventory type counts.
- Using Geocoder Seattle to retrieve coordinate data and render map with folium.

## Data Cleanup:

```
In [55]: raw = pd.read_csv("https://data.seattle.gov/api/views/vsxr-aydq/rows.csv", usecols=cloumns)
cloumns = ['Neighborhood', 'Dominant Discipline', 'Name', 'Stability Index (5=very stable, 1=very uncertain)', 'Closed?', 'I
#drop closed and unknown
raw_clean = raw[raw['Closed?'].isin(['0'])]
#drop Stability Index (5=very stable, 1=very uncertain) is lower than 3
raw_clean = raw_clean[raw_clean['Stability Index (5=very stable, 1=very uncertain)'].isin(['3', '4', '5'])]
#drop empty cell
raw_clean = raw_clean.dropna()
raw_clean = raw_clean.reset_index(drop=True)
raw_clean = raw_clean[['Neighborhood', 'Dominant Discipline', 'Name', 'Latitude', 'Longitude']]

raw_clean.head()
```

```
Out[55]:
```

	Neighborhood	Dominant Discipline	Name	Latitude	Longitude
0	Uptown	Visual	Canlis Glass	47.618262	-122.357741
1	Greenwood	Multi-use	Woodland Park United Methodist	47.685528	-122.355046
2	Seattle Center in Uptown	Performance	Seattle Center/Center House Theatre	47.620569	-122.350491
3	Downtown	Music	Benaroya Hall	47.608128	-122.336975
4	Lake City Way	Multi-use	George Center for Community	47.719929	-122.305033

## Regrouping:

```
In [62]: #group by neighborhood, Dominant Discipline
raw_g = raw_clean.sort_values(by=['Neighborhood', 'Dominant Discipline']).reset_index(drop=True)
raw_g.head(10)
```

```
Out[62]:
```

	Neighborhood	Dominant Discipline	Name	Latitude	Longitude
0	Alki	Heritage	Southwest Seattle Historical Society	47.577553	-122.410950
1	Ballard	Arts/Cultural Training or Education	New York Fashion Academy	47.665419	-122.382443
2	Ballard	Cinema	Majestic Bay	47.668857	-122.384104
3	Ballard	Community Center	Ballard Community Center	47.672783	-122.391548
4	Ballard	Community Center	Ballard NW Senior Center	47.668369	-122.398666
5	Ballard	Literary	Seattle Public Library Ballard	47.669822	-122.384317
6	Ballard	Multi-use	Nordic Museum	47.677699	-122.396359
7	Ballard	Multi-use	Ballard Alki Lodge IOOF-Ballard-Alki Lodge #170	47.668941	-122.379323
8	Ballard	Multi-use	St. Luke's Episcopal Church Seattle	47.670410	-122.384102
9	Ballard	Music	Sonic Boom	47.668522	-122.385106

## Check Counts:

```
In [64]: raw_counts = raw_g.groupby('Neighborhood').count()
raw_counts
```

## Results of Check Counts:

Leschi	1	1	1	1
Loyal Heights	1	1	1	1
Madison Valley	1	1	1	1
Madrona	3	3	3	3
Magnolia	2	2	2	2
Maple Leaf	1	1	1	1
Montlake	1	1	1	1
Mt. Baker	2	2	2	2
Northgate	5	5	5	5
Phinney Ridge	4	4	4	4
Pike Market	2	2	2	2
Pioneer Square	25	25	25	25
Queen Anne	8	8	8	8
Rainier Beach	1	1	1	1
Rainier Valley	3	3	3	3
Ravenna	3	3	3	3
Rosetown	1	1	1	1
SODO	2	2	2	2
Sand Point	2	2	2	2
Seattle Center in Uptown	15	15	15	15
Seward Park	1	1	1	1
South Lake Union	6	6	6	6
South Park	1	1	1	1
Southside Seattle	1	1	1	1
Squire Park	1	1	1	1
University District	28	28	28	28
Universtiy District	1	1	1	1
Uptown	5	5	5	5
Wallingford	6	6	6	6
Wedgwood	1	1	1	1

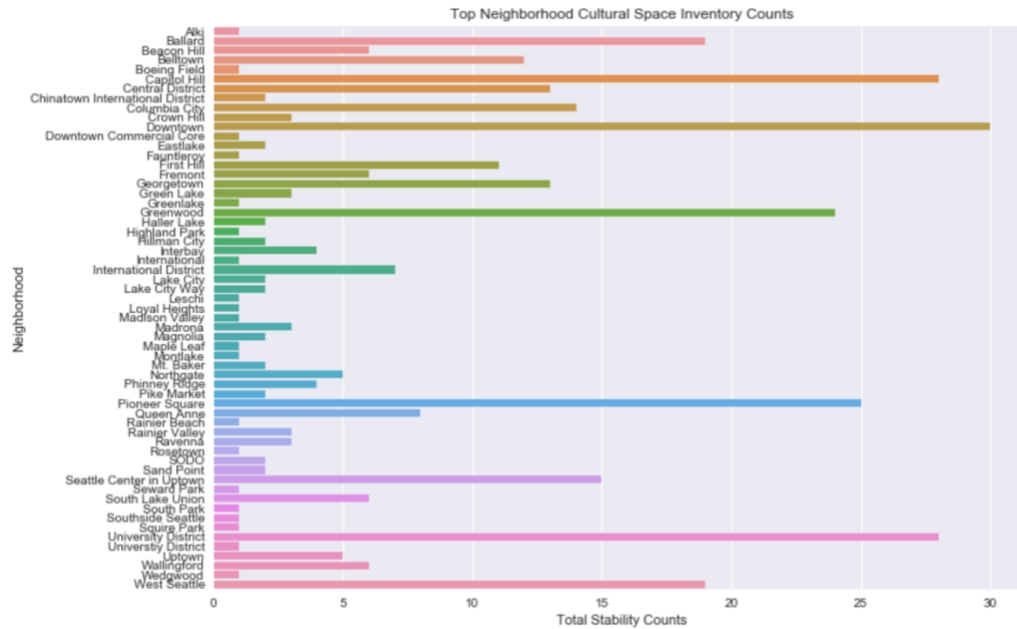
Neighborhood	Dominant Discipline	Name	Latitude	Longitude
Alki	1	1	1	1
Ballard	19	19	19	19
Beacon Hill	6	6	6	6
Belltown	12	12	12	12
Boeing Field	1	1	1	1
Capitol Hill	28	28	28	28
Central District	13	13	13	13
Chinatown International District	2	2	2	2
Columbia City	14	14	14	14
Crown Hill	3	3	3	3
Downtown	30	30	30	30
Downtown Commercial Core	1	1	1	1
Eastlake	2	2	2	2
Fauntleroy	1	1	1	1
First Hill	11	11	11	11
Fremont	6	6	6	6
Georgetown	13	13	13	13
Green Lake	3	3	3	3
Greenlake	1	1	1	1
Greenwood	24	24	24	24
Haller Lake	2	2	2	2
Highland Park	1	1	1	1
Hillman City	2	2	2	2
Interbay	4	4	4	4
International	1	1	1	1
International District	7	7	7	7
Lake City	2	2	2	2
Lake City Way	2	2	2	2

Top Neighborhood Cultural Space Inventory Counts:

```
In [65]: import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [67]: sns.set(rc={'figure.figsize':(11.7,8.27)})
ax0 = sns.barplot(data = raw_counts.reset_index(), x = 'Name', y = 'Neighborhood')
ax0.set_title("Top Neighborhood Cultural Space Inventory Counts")
ax0.set_xlabel('Total Stability Counts')
plt.show()
```

Results & Visualization:

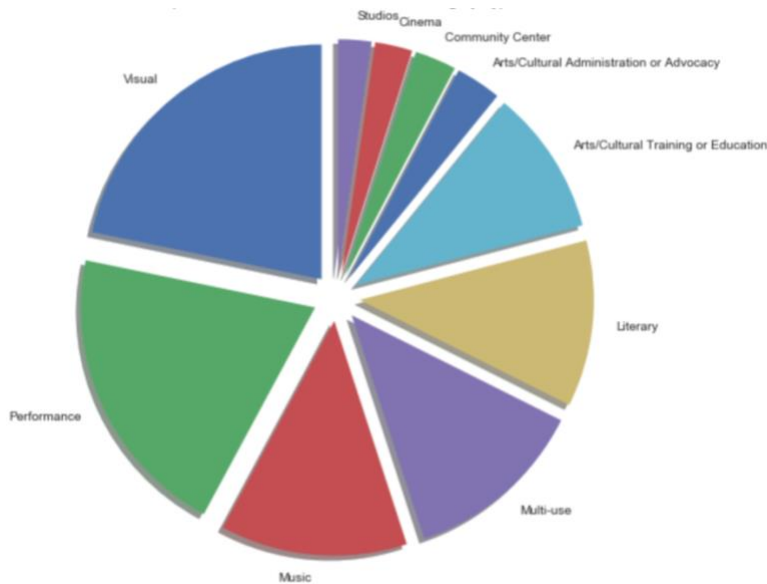


Top 10 most common Cultural Space Inventory type counts:

```
In [68]: top_10 = raw_clean.groupby('Dominant Discipline')['Name'].count().sort_values(ascending=False)[:10].reset_index()
labels = top_10['Dominant Discipline']
sizes = top_10['Name']
explode = [.1] * 10
# Plot
plt.title("Top 10 Cultural Space Inventory")
ax1 = plt.pie(sizes, explode=explode, labels=labels, shadow=True, startangle=90)
plt.axis('equal')
plt.show()
top_10
```

Results & Visualization:

	Dominant Discipline	Name
0	Visual	76
1	Performance	71
2	Music	45
3	Multi-use	44
4	Literary	40
5	Arts/Cultural Training or Education	35
6	Arts/Cultural Administration or Advocacy	11
7	Community Center	10
8	Cinema	9
9	Studios	8



Coordinate data and render map with folium:

```
# explore Seattle
print('The dataframe has {} Neighborhood and {} Cultural Space Inventories.'.format(
    len(raw_g['Neighborhood'].unique()),
    raw_g.shape[0]
))
```

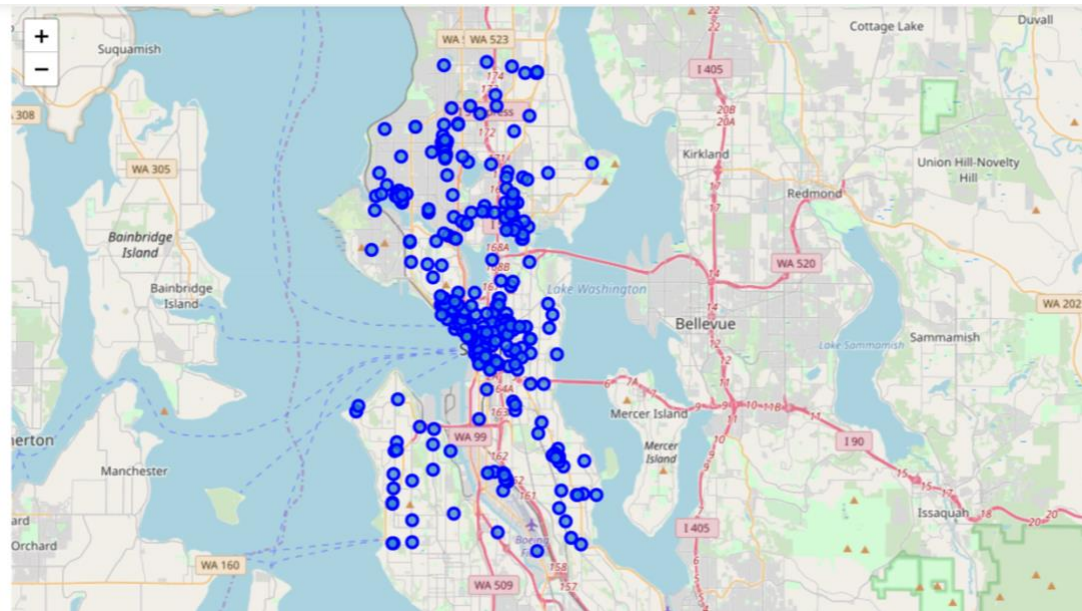
The dataframe has 59 Neighborhood and 365 Cultural Space Inventories.

```
#!/conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library
from geopy.geocoders import Nominatim
address = 'Seattle, United States'

geolocator = Nominatim(user_agent="Seattle_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Seattle are {}, {}'.format(latitude, longitude))

# create map of Seattle using latitude and longitude values
map_Seattle = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, raw_g in zip(raw_g['Latitude'], raw_g['Longitude'], raw_g['Neighborhood'], raw_g['Name']):
    label = '{} {}'.format(raw_g, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_Seattle)
```



Conclusion:

Recommendation for Siyuan: Top three cultural neighborhood in Seattle are downtown, capitol hill and university district. Tope three cultural inventory are visual, performance and music. Seattle is a recommended city for artist.