Welcome to the Big Apple!
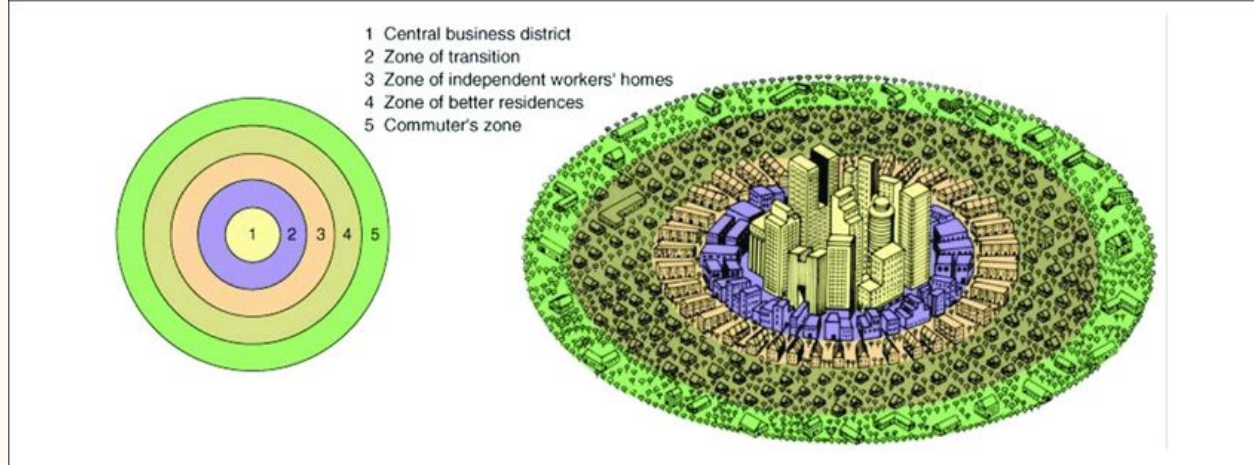
Cagan Abney, Fernanda Valdez, Jonathan Unger and Joshua Hale
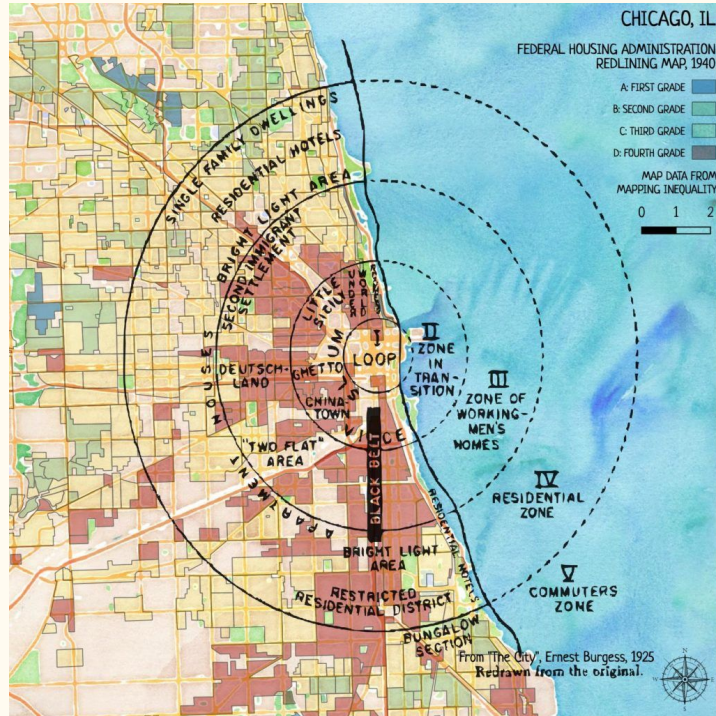
Our Application

Inspiration

Burgess concentric city model, one of the earliest models created to explain urban structures, states that every city has a center, denominated the Central Business District (CBD) and that land usage and therefore social groups are distributed in concentric rings based on the amount people are willing to pay for land.

1 Central business district
2 Zone of transition
3 Zone of independent workers' homes
4 Zone of better residences
5 Commuter's zone

Inspiration

The model was based on the city of Chicago, and generalized as a rule for cities in developed countries. However, in modern times, it has been challenged by contemporary geographers who've proposed alternative models.

Inspiration

With this project, we want to explore if the city of New York follows the Concentric City Model or if, on the contrary, it appears to have a different layout that would apply to a different city model.

We are plotting price per square footage vs distance from the center to understand if, as explained in the CCM, there is an inverse correlation, or not.

Data Set

Below is a link to our dataset for New York City Housing Prices.

We chose this data set due to amount of available houses in NYC and having prior experience working with NYC maps.

We also found this dataset helpful due to the latitude and longitude columns

https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market

Data Engineering

Started by dropping duplicate values.

We then removed all non-specific house types from the "TYPE" column. For example: "'Coming Soon', 'Contingent', 'For sale', 'Mobile house for sale', 'Condop for sale', 'Pending', 'Foreclosure', 'Land for sale'",

This left us with 5 remaining house types:

```
1  df.House_Type.unique()
array(['Condo', 'House', 'Townhouse', 'Co-op', 'Multi-family'],
      dtype=object)
```

Data Engineering Continued

We added multiple columns like "Price by SQ FT", "Borough" and "Distance (from center of New York City)"

After thinking through our options for visualizations, we decided that we wanted to show the Top 10 most expensive homes (with filters applied) and Price by SQ FT vs distance from the center of NYC to see if New York follows the CBD model.
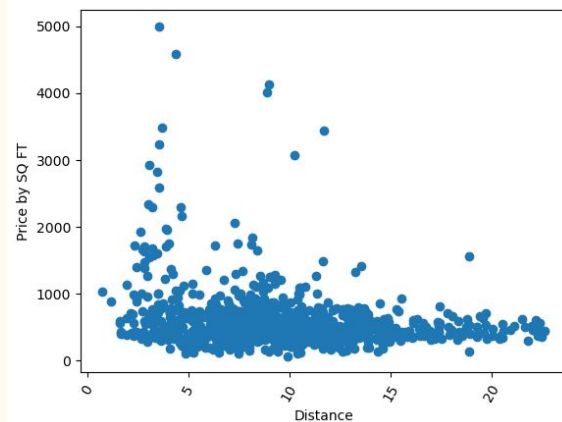
Queries

We started with price and number of beds:

```python
price = 500000

# allow the user to choose price
if price == "All":
    where_clause = "1=1"
else:
    where_clause = f"PRICE <= '{price}'"

query = f"""
        SELECT
            *
        FROM
            NYC_Houses
        WHERE
            {where_clause};
"""

print(query)
```

```python
num_beds = 3

# allow the user to select how many bedrooms
if num_beds == "All":
    where_clause = "1=1"
else:
    where_clause = f"BEDS <= '{num_beds}'"

query = f"""
        SELECT
            *
        FROM
            NYC_Houses
        WHERE
            {where_clause};
"""

print(query)
```

Queries Continued

We then added number of baths and house type

```python
num_baths = 2

# allow the user to choose how many baths
if num_baths == "All":
    where_clause = "1=1"
else:
    where_clause = f"BATH <= '{num_baths}'"

query = f"""
        SELECT
            *
        FROM
            NYC_Houses
        WHERE
            {where_clause};
"""

print(query)
```

```python
house_type = 'House'

# allow the user to select a house type
if house_type == "All":
    where_clause = "1=1"
else:
    where_clause = f"House_Type == '{house_type}'"

query = f"""
        SELECT
            *
        FROM
            NYC_Houses
        WHERE
            {where_clause};
"""

print(query)
```

Queries Continued

After all of the individual queries were finalized, we wanted to combine to create one large query.

What started as the query below, ended as the query to the right in an attempt to shorten our code and not repeat steps

```python
house_type = 'House'
price = 500000
num_beds = 3
num_baths = 2
price_by_sqft = 250
distance = 12


if house_type == "All":
    where_clause1 = "1=1"
else:
    where_clause1 = f"House_Type == '{house_type}'"

if price == "All":
    where_clause2 = "1=1"
else:
    where_clause2 = f"PRICE <= {price}"

if num_beds == "All":
    where_clause3 = "1=1"
else:
    where_clause3 = f"BEDS <= {num_beds}"

if num_baths == "All":
    where_clause4 = "1=1"
else:
    where_clause4 = f"BATH <= {num_baths}"

if price_by_sqft == "All":
    where_clause5 = "1=1"
else:
    where_clause5 = f"PRICE_BY_SQFT <= {price_by_sqft}"

if distance == "All":
    where_clause6 = "1=1"
else:
    where_clause6 = f"Distance <= {distance}"

query = f"""
        SELECT
            *
        FROM
            NYC_Houses
        WHERE
            {where_clause1} AND {where_clause2} AND {where_clause3} AND {where_clause4} AND {where_clause5} AND
"""

print(query)
```

```python
def getMapData(self, house_type, price, num_beds, num_baths, price_by_sqft, distance):

    where_clause = self.getWhereClause(house_type, price, num_beds, num_baths, price_by_sqft, distance)

    query = f"""
            SELECT
                *
            FROM
                NYC_Houses
            WHERE
                {where_clause};
    """

    print(query)
```
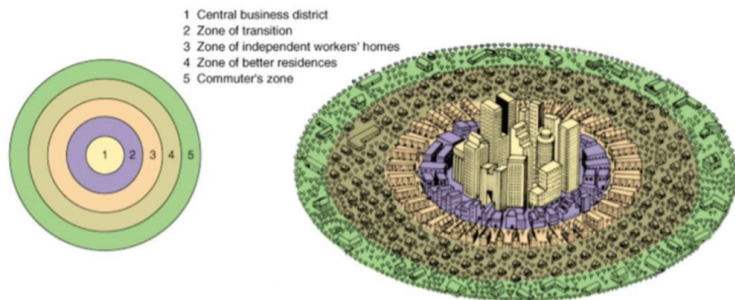
Demo Time - Link

Conclusions

NYC does not necessarily follow the Concentric City Model, mainly because the area evaluated was limited to areas closest to the city center. This is probably because of historical development and expansion over nearly 400 years and changing standards of city planning.

The NY area analyzed is delimited by sea, therefore prices vary more by boroughs than by distance from the center. Also, many large entities like the prison system, various utilities, transport hubs, parks and preserves occupy extensive regions, limiting residential development.

What was very interesting though was that we found a larger amount of condos, townhouses and multi-family constructions closest to the city center, which aligns with the CCM as price per square foot is higher the closer one gets to the center.

Limitations/Bias

The most impactful limitation of this project was us newly being introduced to HTML/CSS and JavaScript. There are a lot of additional features and design that could add to the user experience.

Some limitations in our data include the data being static. If we were able to have dynamic data that updates daily with new homes for sale or what is sold our dashboard would be autonomous instead of someone inputting a new SQLite file in every month.

Additionally, none of us are very familiar with New York, so there was limited knowledge of the boroughs of New York, and lack of information to make some conclusions about why homes were priced how they were.

There is bias in our data set because it only includes houses relatively close to New York City, and doesn't include houses within the suburbs.

Future Work

In the future, we can add housing data from more cities to compare which cities may follow the CCM and which do not. There could be a dashboard for each city (Dallas, Los Angeles, Chicago, Houston, Washington D.C., etc.)

We would also expand our data to include more of the suburbs and therefore understand if when looked at a greater scale, the CCM does apply to the greater NY.

We could add a range feature to the filters. For example, the user could select price range $500,000-$1,000,000 and/or 3-6 bedrooms.

Works Cited

- Prior classwork assignments
- Professor Booth
- Stack Overflow
- Xpert Learning Assistant
- https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market
- https://web.archive.org/web/20110629113720/http://people.hofstra.edu/geotrans/eng/ch6en/conc6en/burgess.html
- https://transportgeography.org/contents/chapter8/urban-land-use-transportation/burgess-land-use/