Group project plan

Data side

Filter options:

Filter choices will be the same as the ones as on Zillow, this information will be simply taken straight from available datasets and displayed to users. Ex :

- Home type
- Bed & Baths
- Scenery
- Student accommodation
- Price - lowest to highest
- Price - highest to lowest
- Renting/ buying

Filter choices created by us will be:

- Area safety (crime rate)
- Neighbourhood entertainment/ activities rating
- Emergency service accessibility
- Public transport accessibility
- House pricing (by neighbourhood)
- Overall score

How information will be displayed to user after selecting filters

User will select a property

Remove Boundary ✕

Re-center

Will display most important information + filters users selected... can click to expand for more detail...

$134,900

3 bds | 2 ba | 1,320 sqft - Townhouse f...

2302 Koko Ln, Baltimore, MD 21216

MERRITT REAL ESTATE, LLC

Crime Rate

Entertainment

Emergency Service Access

Public Transport

Nearest bus stops:

N1385: Distance: 2km
N3458: Distance: 2.5km
N3480: Distance: 3km
N3945: Distance 5km

Nearest Train Stations:

More>

How our metrics will be calculated

Using:
K-Means Clustering (for numerical data e.g distance)
DBSCAN
K-Modes Clustering (for categorical data e.g no. of robberies)

These algorithms create "clusters" where they group data by similarity to the nearest cluster (in Euclidean distance)

Subcategories will be created for each of our newly made features, these categories will be:
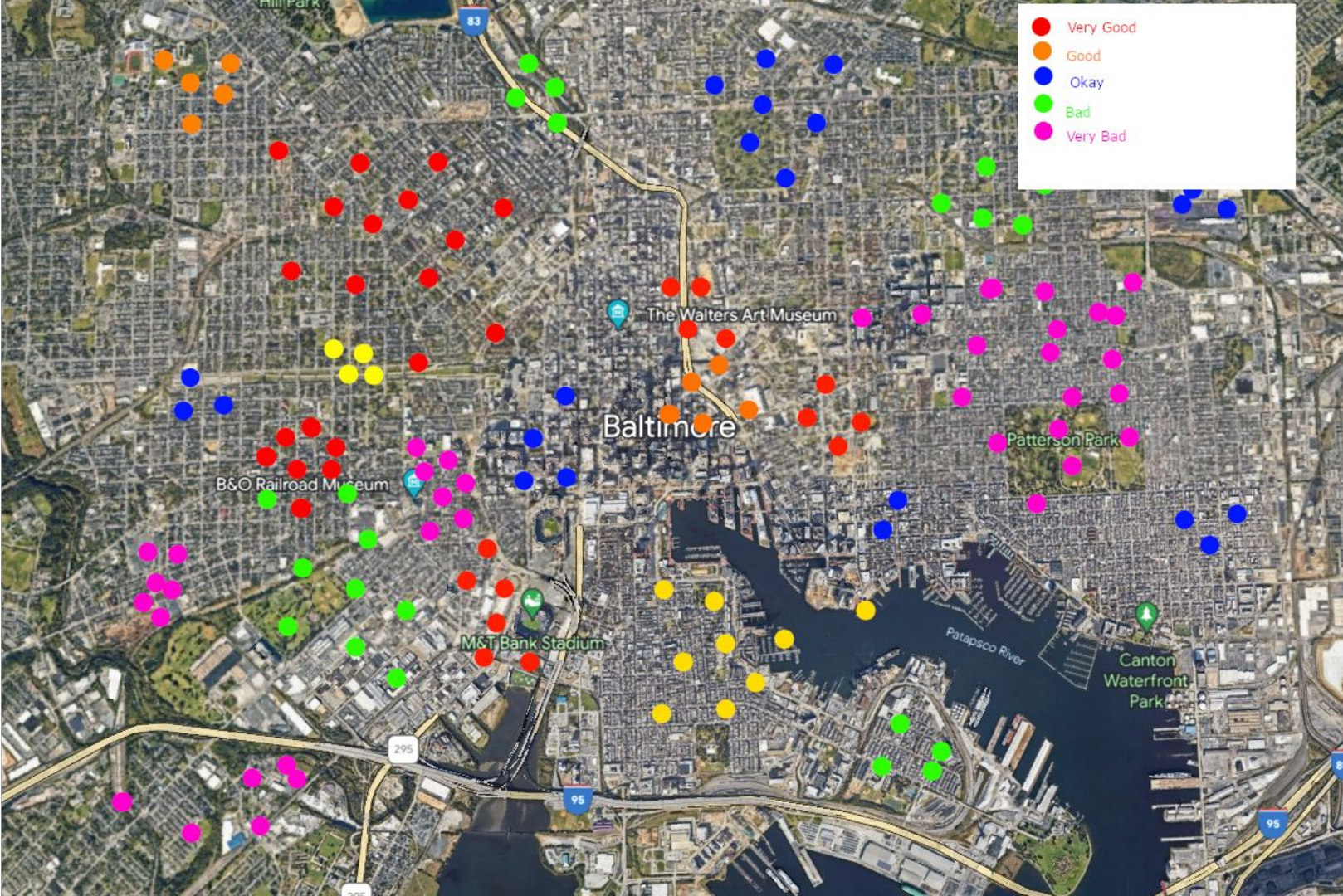
Very good
Good
Okay
Bad
Very bad

For example…

So let's say our public transport data consists of features:

Distance to bus stop
Distance to train station
Distance to airport

We would then apply K-means clustering to group houses into clusters based on their proximity to these features. The algorithm will identify clusters of houses that are relatively close to one another in terms of their distances to these transportation features.

The resulting clusters would represent different categories of houses based on their proximity to these transportation options. For example, we might have clusters like:
- Cluster 1: Very Good (Many bus/ train stops in area, short distance to stops).
- Cluster 2: Good.
- Cluster 3: Okay.
- Cluster 4: Bad.
- Cluster 5: Very bad. (Few bus/ train stops in area, short distance to stops).