

Documentation of Work + Insights

EDA/Data Cleaning File

Simple Data Structuring/Exploring

Process

1. Data started off as **4,346** rows with **10** columns.
 - Removed Null entries and empty columns → **3,190** rows and **9** columns.
2. Converted **'date'** column into 2 new columns: **['new_date', 'new_time']**.
 - Data type is **datetime**.
 - Removed **'date'** column.
 - Data shape is **3,190** rows with **10** columns.

Visualizations

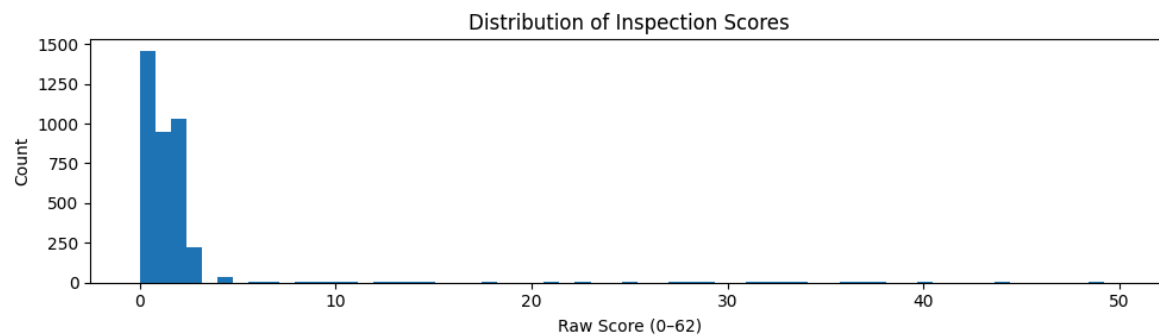


Figure I: Distribution of initial score range (0-62) of all 3,190 entries.

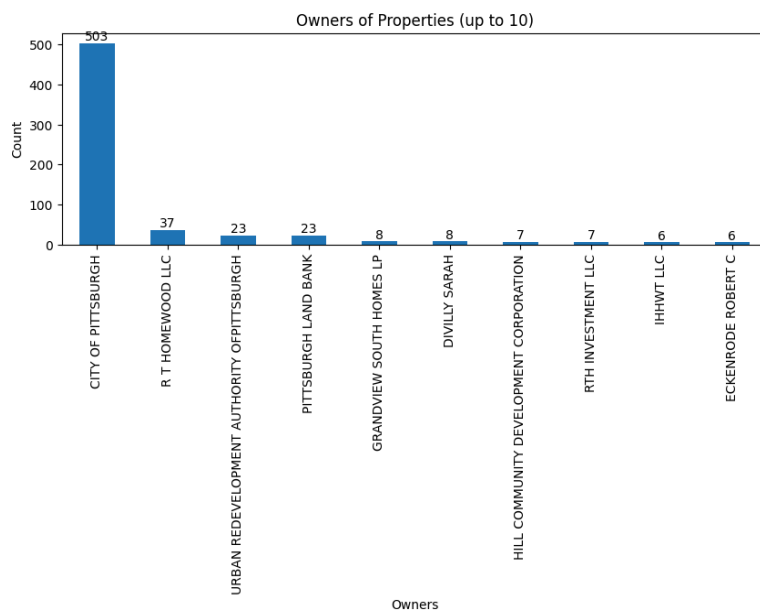


Figure II: Distribution of owners of the properties.

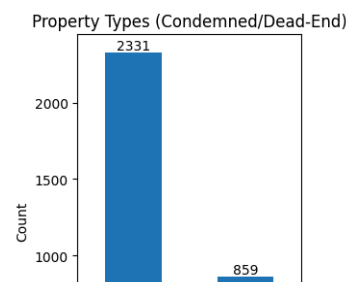
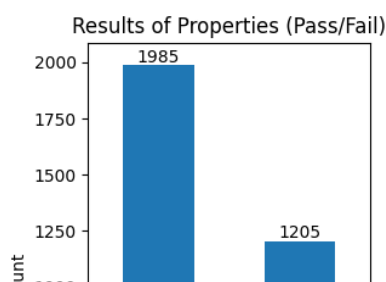


Figure III: Distribution of Pass/Fail Results.

Figure IV: Distribution of Condemned / Dead-End Properties.

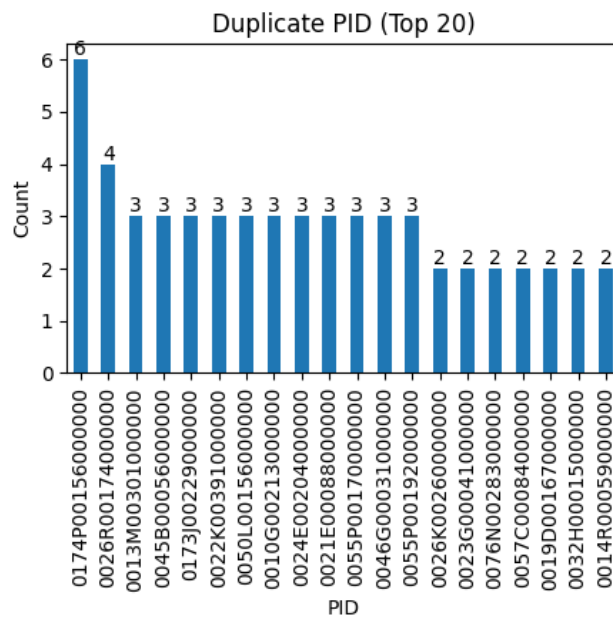
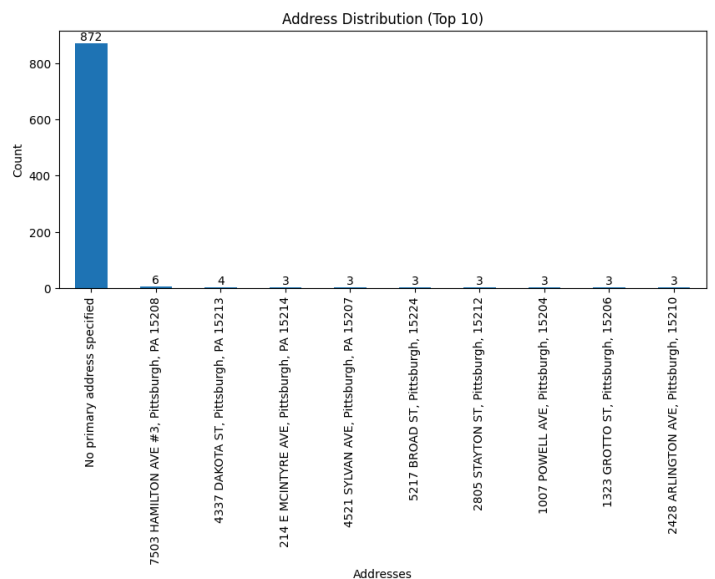


Figure V: Distribution of PID's of all Properties (showing duplicate issue).

Figure VI: Distribution of addresses in the dataset.



Insights

Figure I

- Distribution of the initial scores range mostly around 0-5.

Figure II

- 503 properties (~15.8% of properties) are owned by the City of Pittsburgh

Figure III

- ~62.2% of properties failed the latest inspections (1985 properties)
- ~37.8% of properties passed the latest inspections (1205 properties)

Figure IV

- ~73.1% of properties are condemned properties
- ~26.9% of properties are dead-end properties

Figure V

- Some unique parcel ID's have more than 1 entry, showing duplicate parcel id issues

Figure VI

- ~27.3% of the properties (872 properties) have no valid addresses

Data Cleaning

1. Created a **helper function (score_adjust(x))** to adjust score ranges from 0-62 to 1-4
 - **Score 1:** 0-16
 - **Score 2:** 16-32
 - **Score 3:** 32-47
 - **Score 4:** 47-63
2. List of duplicate **parcel_id**'s was **371** entries before fixing the score range, then it was **9** entries.
3. Created a **helper function (remove_duplicates(data))** to remove the remaining 9 duplicate parcel ID entries:
 - Removed duplicate entries by keeping latest date, latest time, or highest score

```
score
1    3137
2      23
3      27
4       3
Name: count, dtype: int64
```

```
Parcel ID: 0022F00119000000      Parcel ID: 0022F00119000000
   parcel_id  new_date new_time  score   parcel_id  new_date new_time  score
0022F00119000000 2020-09-10 00:00:00    4 0022F00119000000 2020-09-10 00:00:00    4
0022F00119000000 2020-09-10 00:00:00    3

Parcel ID: 0043S00018000000      Parcel ID: 0043S00018000000
   parcel_id  new_date new_time  score   parcel_id  new_date new_time  score
0043S00018000000 2020-08-11 00:00:00    3 0043S00018000000 2022-07-27 00:00:00    1
0043S00018000000 2022-07-27 00:00:00    1
```

Visualizations on ALL Properties

Figure I: Distribution new score range (1-4) of 2,802 properties

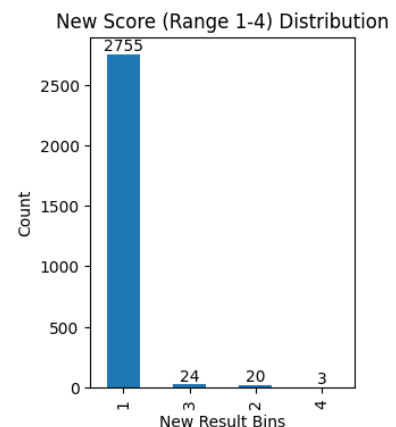
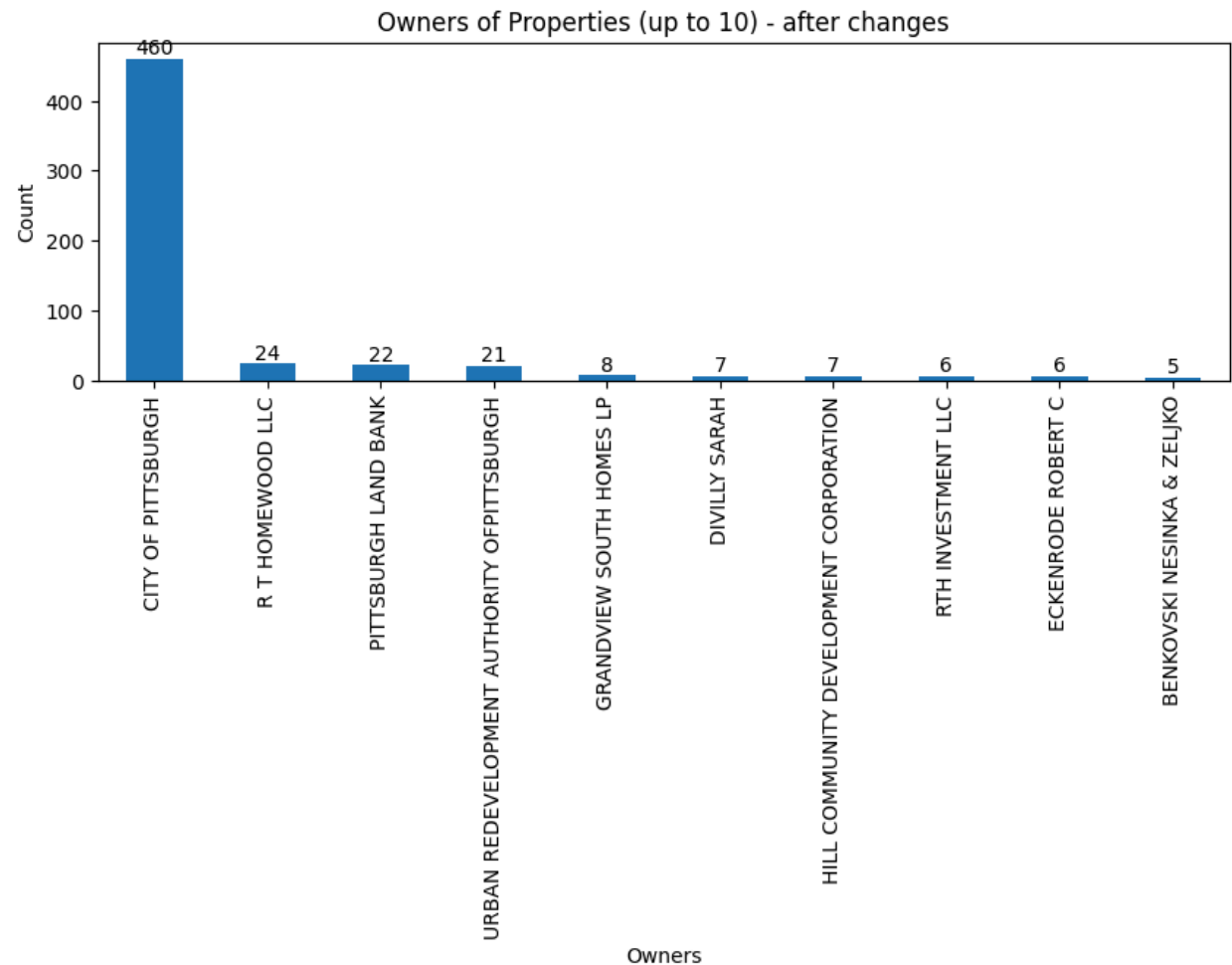


Figure II: Distribution of Parcel ID's



Figure III: Distribution of Owners of remaining properties



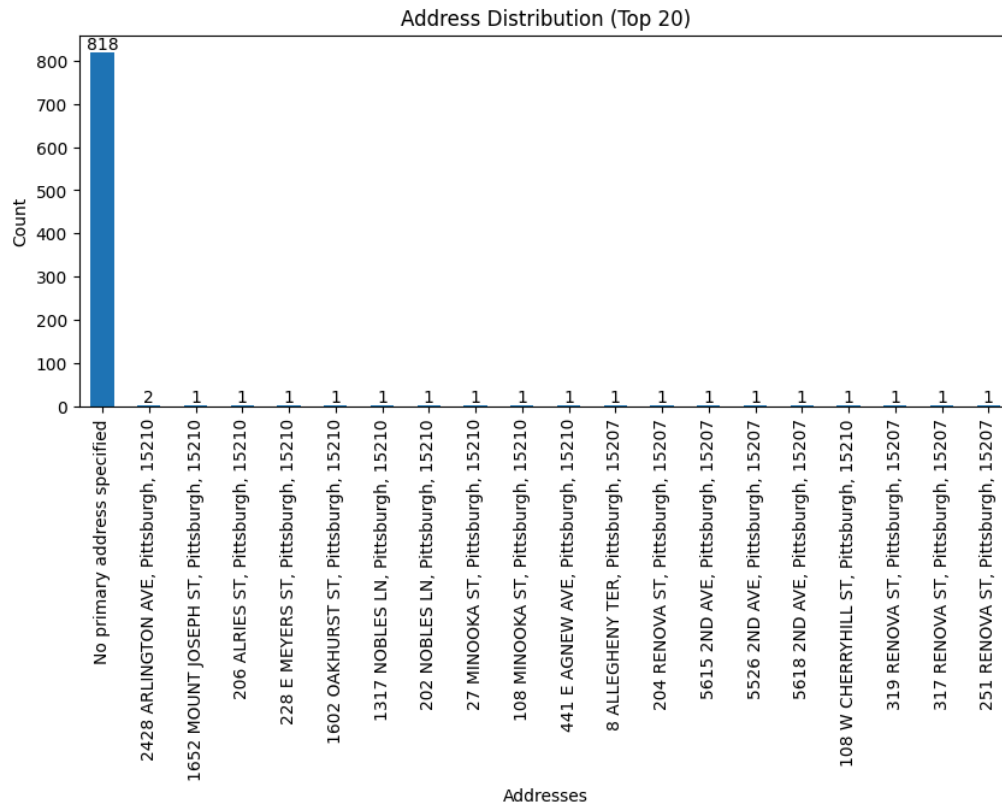


Figure IV:
Address
Distribution
after removing
duplicate
entries

Insights on ALL Properties

Figure I (new score range distribution)

- ~98.3% of the scores (2755 properties) are score 1, structurally intact with no immediate observable danger.
- ~0.7% of the scores (20 properties) are score 2, structurally compromised, unsafe, and possibly dangerous.
- ~0.9% of the scores (24 properties) are score 3, structurally compromised, unsafe, and dangerous.
- ~0.1% of the scores (3 properties) are score 4, imminently dangerous.

Figure II (new distribution of parcel ID's)

- All Parcel IDs do not have any duplicates anymore

Figure III

- Now ~16.4% of properties are owned by the City of Pittsburgh, reduced from 503 to 460

Figure IV

- ~29.2% of the properties have no valid addresses, reducing 872 to 818 properties

Condemned Properties

Visualizations on Condemned Properties (1,991 Properties)

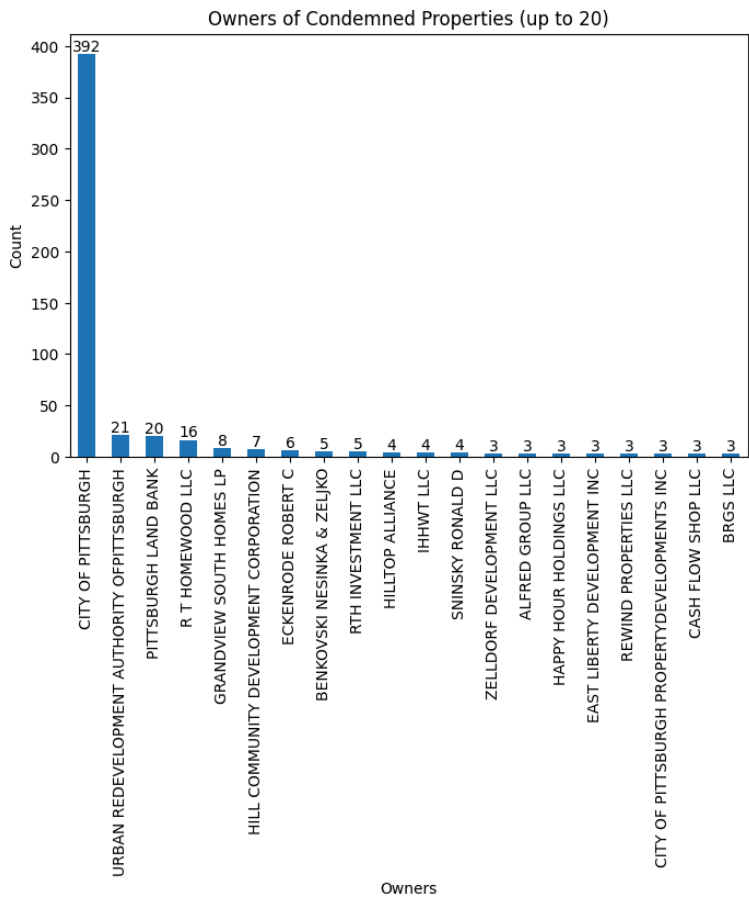


Figure I: Owners of the Condemned Properties

Figure II: Distribution of scores for Condemned Properties

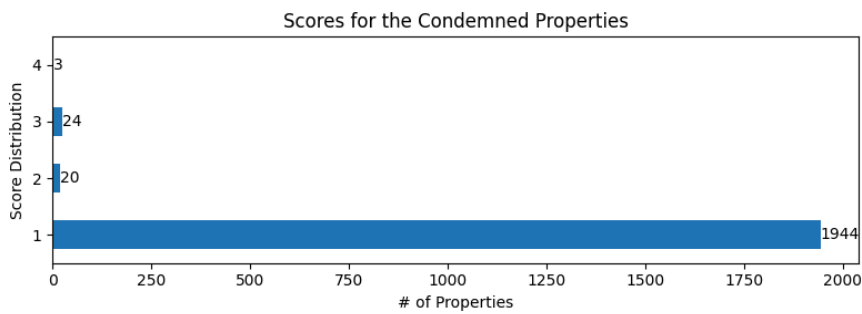


Figure III: Owners of the Score 4 Properties

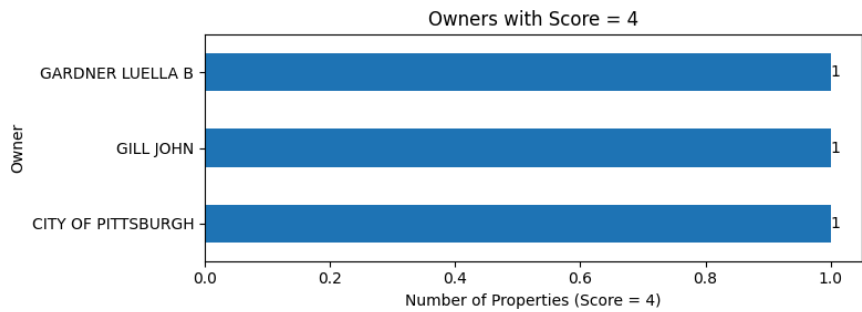


Figure IV: Pass or Fail results for Condemned Properties

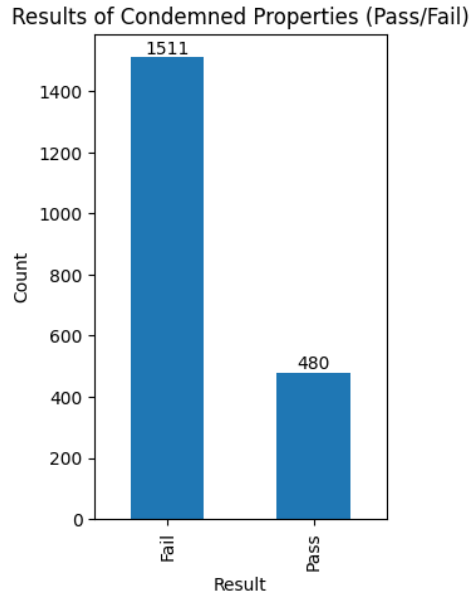
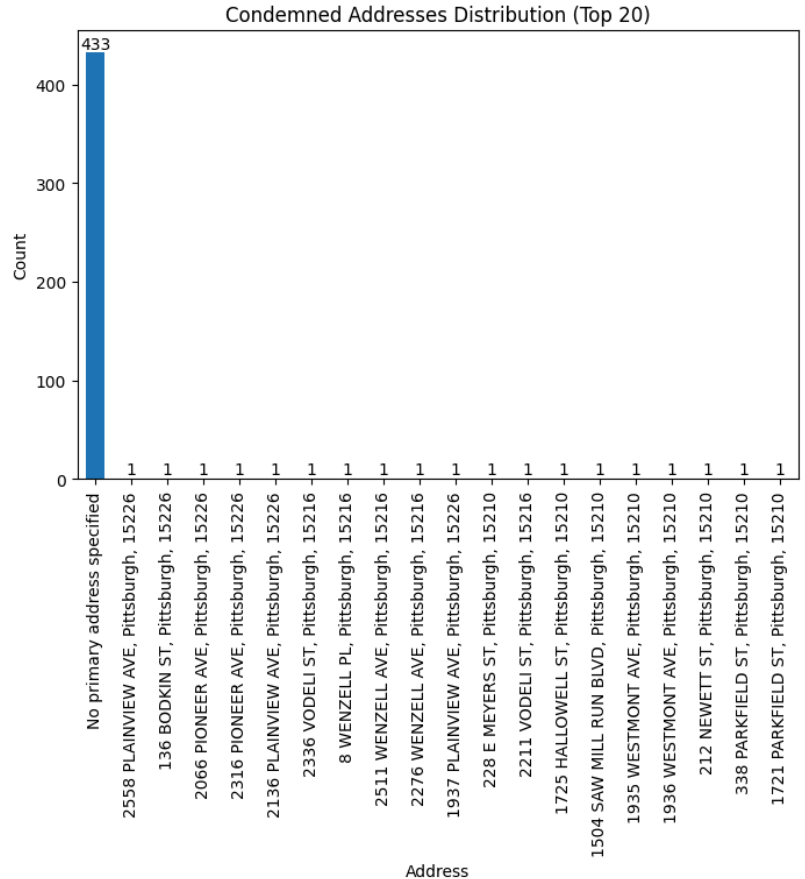


Figure V: Address distribution for Condemned Properties



Insights on Condemned Properties

Figure I

- ~97.6% of condemned properties have a score of 1.
- All 3 of the imminently dangerous properties (~0.2%) is condemned

Figure II

- ~19.7% (392) of the Condemned Properties are owned by City of Pittsburgh

Figure III

- The 3 owners of the dangerous properties are Gardner Luella B., Gill John, and the City of Pittsburgh

Figure IV

- ~75.9% of the Condemned Properties Fail the Inspection
- ~24.1% of the Condemned Properties Pass the Inspection

Figure V

- ~21.7% (433) of the Condemned Properties have no valid addresses.

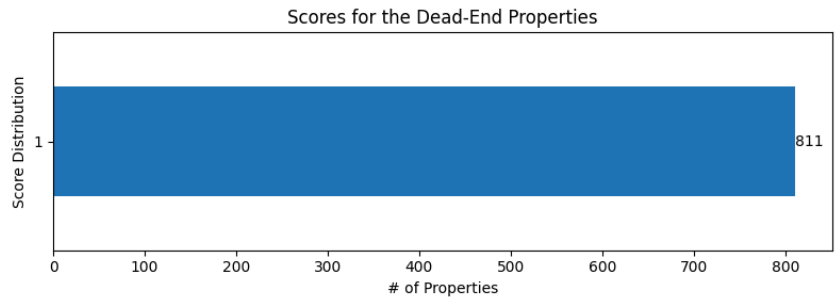
Dead-End Properties

Visualizations on Dead-End Properties

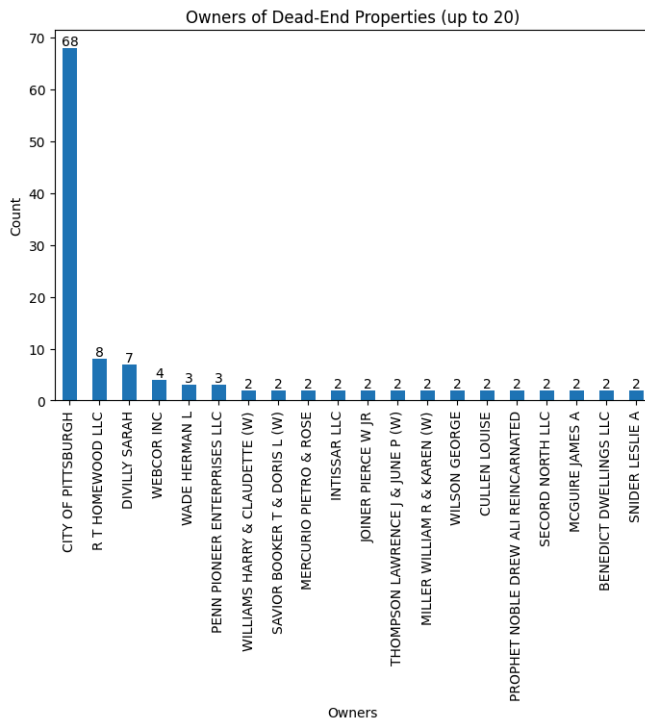
Figure I: Score Distribution for Dead-End Properties

Figure II: Distribution of owners of

Figure III: Pass or Fail for Dead-End Properties



Dead-End Properties



Results of Dead-End Properties (Pass/Fail)

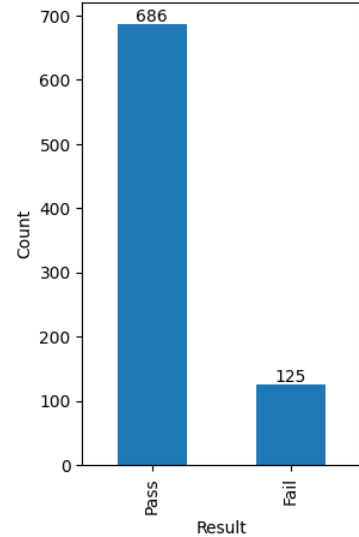
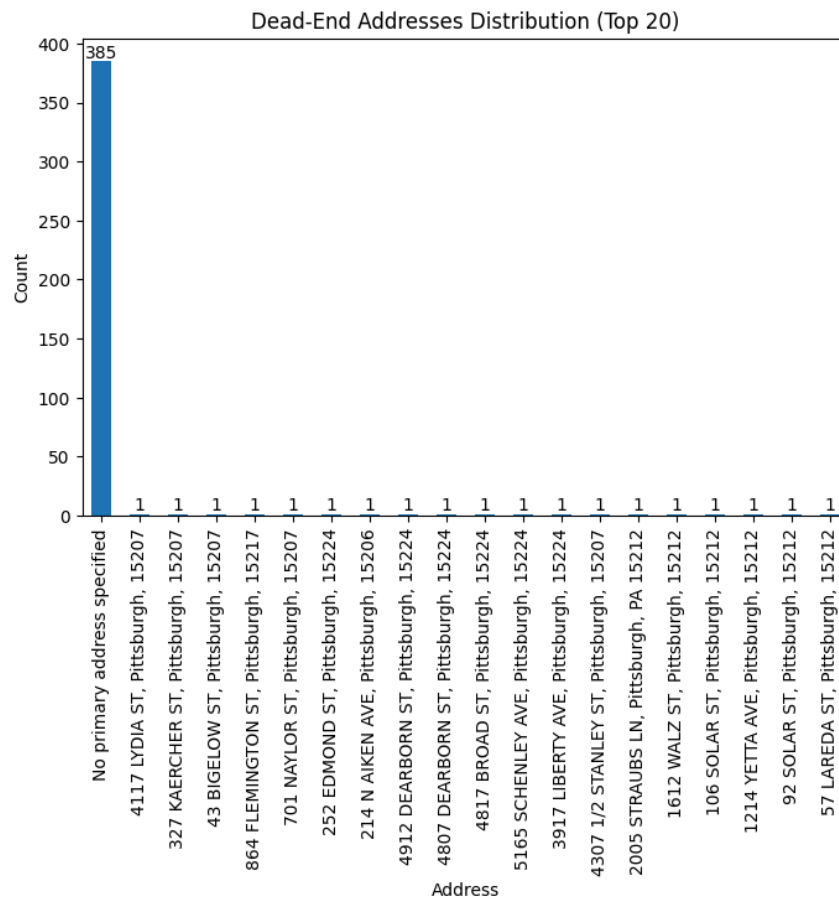


Figure IV: Dead-End Address Distribution



Insights on Dead-End Properties

Figure I

- 100% of the Dead-End Properties are Score 1's, structurally intact with no immediate danger

Figure II

- ~8.4% of the Dead-End Properties are owned by City of Pittsburgh

Figure III

- ~84.6% of Dead-End Properties Fail the Inspections
- ~15.4% of Dead-End Properties Pass the Inspections

Figure IV

- ~47.5% of Dead-End Properties has no valid address

Overall Insights (from IPYNB)

Condemned or Dead-End

- Approximately 79.2% of Properties are Condemned
- Approximately 20.8% of Properties are Dead-End

Fail Inspection

- Approximately 79.3% of Condemned Properties Fail Inspection.
- Approximately 10.6% of Dead-End Properties Fail Inspection.

Scores Evaluation

- All Dead-End Properties having a score of 1 show that they are structurally intact with no immediate observable danger. [Source](#)
- Condemned Properties scores range from 1-4, with 4 being imminently dangerous, these can be the properties/areas the city prioritizes.

Score 4 Properties

- The 3 owners of the 3 Score 4 Properties are: City of Pittsburgh, Gill John, and Gardner Luella B.
- By Identifying their Addresses with the cleaned dataset, we could prioritize these high-risk properties to obtain private demolition permit or a building permit to repair.

Addresses

- Most addresses in the dataset is listed as "No primary address specified".
- This will have to be filled in to figure out which areas should be prioritized first.

After Merging with Allegheny Data

Process

1. From the **Property Assessment Parcel Data** used features: ["PARID", 'PROPERTYHOUSENUM', 'PROPERTYADDRESS', 'PROPERTYCITY', 'PROPERTYSTATE', 'PROPERTYUNIT', 'PROPERTYZIP']
 - Properties features will be combined to create a new **Address** feature using helper function **build_address(row)**
 - If address in **Address** feature is invalid, default to "No primary address specified"
 - By matching the **PARID** with the **parcel_id** from original data frame using inner join, add **Address** column to the original data frame
 - Using helper function **final_add(dataframe)**, address will be replaced with the valid address to new feature of **final_address**

Visualizations

Figure I: Address Distribution after adding Allegheny Data Addresses

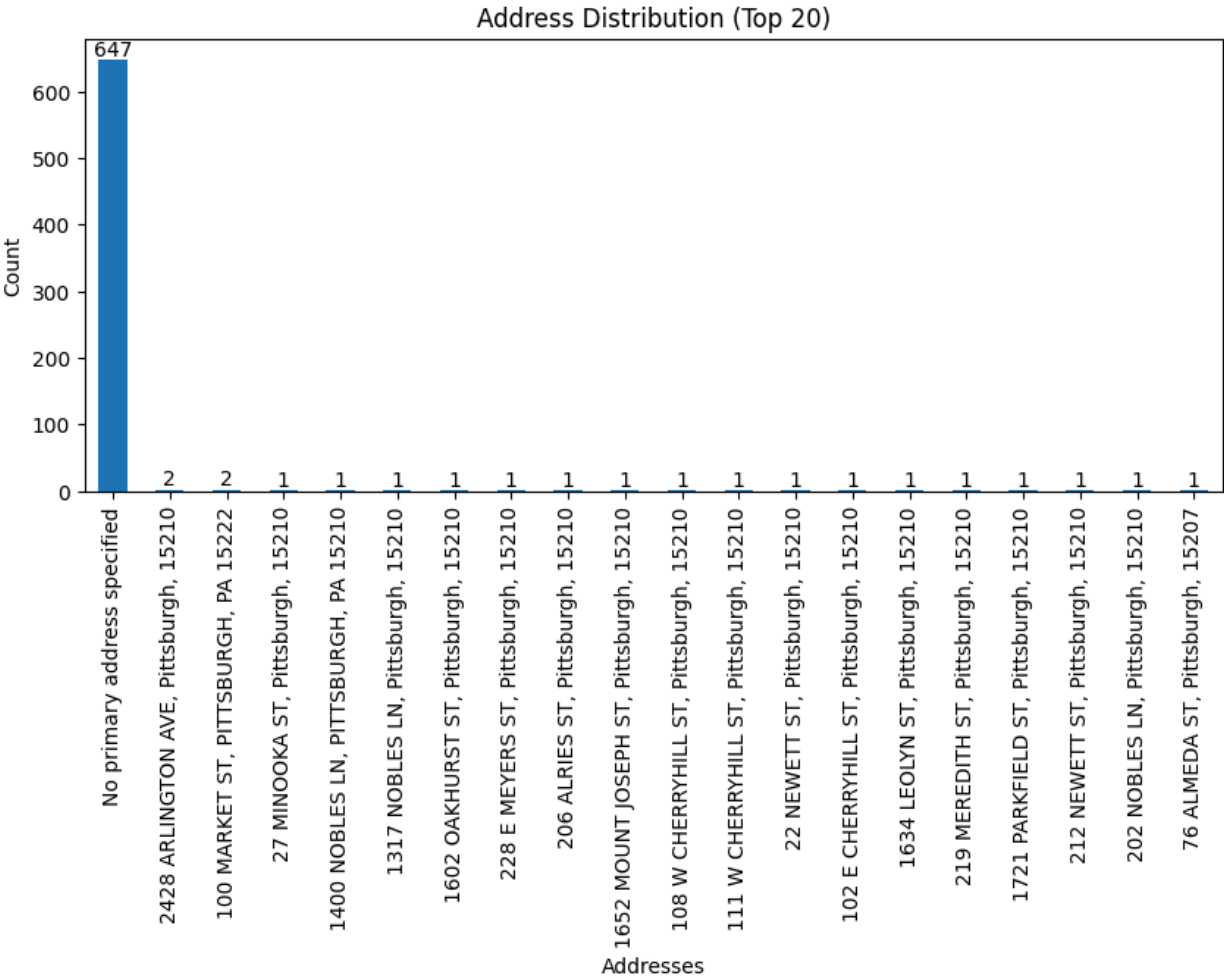
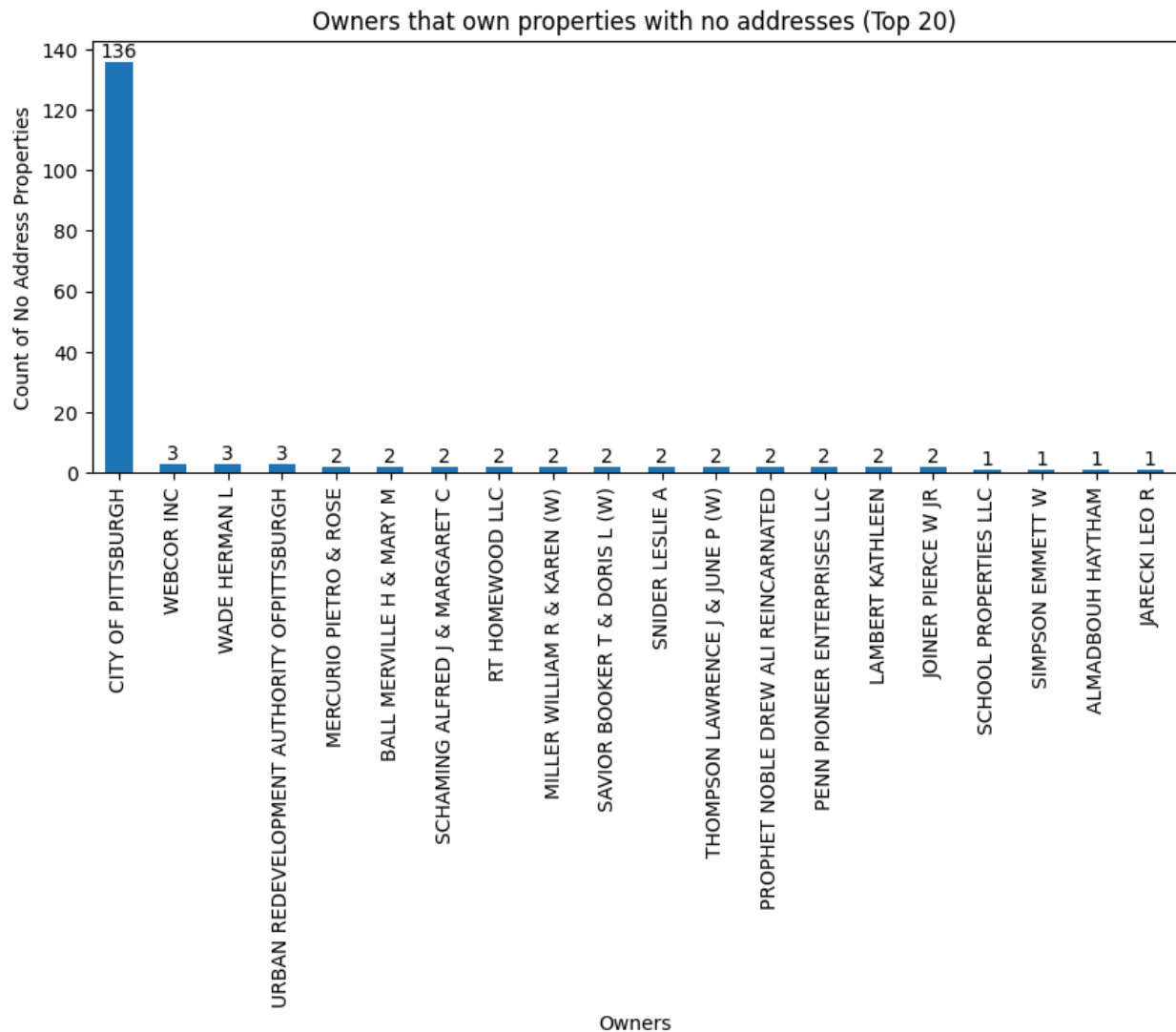


Figure II: Owners Distribution with no addresses



Insights

Figure I

- Reduced No Address values from 813 to 647

Figure II

- City of Pittsburgh owns approximately 21% (136) of properties that do not have an address attached to it

After Merging with City Properties Data

Process

- From the **City Properties Dataset** used features: ['pin', 'address', 'owner', 'acquisition_date']
 - 'acquisition_date' was converted into *datetime* data type to new feature **date**
 - 'owner' was renamed to **city_owner** for clarity
- Inner join of the city data with matching **parcel_id** to **pin**
- Created new feature **final_owner** by comparing **new_date** to **date** column
 - If **new_date** is latest, **final_owner** will take the **owner** value
 - If **date** is latest, **final_owner** will take the **city_owner** value
- Created new feature **final_date** by comparing **new_date** to **date** column
 - If **new_date** is latest, **final_date** will take the **new_date** value
 - If **date** is latest, **final_date** will take the **date** value
- Update **final_address** if final_address is not a valid address and new address from the city property dataset if valid
 - All 470 matching pid entries were updated with the owner, date, and address accordingly

Visualizations

Figure I:

parcel_id	address	owner	property_type	latest_inspection_result	inspection_status	new_date	new_time	score
0022F00119000000	No primary address specified	CITY OF PITTSBURGH	Condemned Property	Fail	Inactive	2020-09-10	00:00:00	4
0047E00052000000	No primary address specified	GILL JOHN	Condemned Property	Fail	Inactive	2020-09-04	00:00:00	4
0232A00021000000	No primary address specified	GARDNER LUELLA B	Condemned Property	Fail	Inactive	2020-07-10	00:00:00	4

Figure II:

parcel_id	owner	property_type	latest_inspection_result	inspection_status	new_date	new_time	score	final_address
0022F00119000000	CITY OF PITTSBURGH	Condemned Property	Fail	Inactive	2020-09-10	00:00:00	4	1406 WARNER ST, PITTSBURGH, PA 15233
0047E00052000000	GILL JOHN	Condemned Property	Fail	Inactive	2020-09-04 00:00:00	00:00:00	4	913 MINA ST, PITTSBURGH, PA 15212
0232A00021000000	GARDNER LUELLA B	Condemned Property	Fail	Inactive	2020-07-10 00:00:00	00:00:00	4	No primary address specified

Figure III:

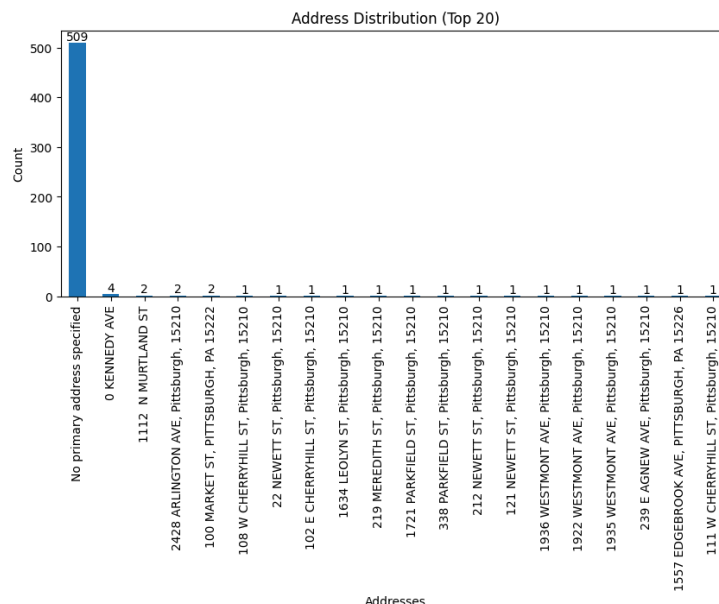


Figure IV:

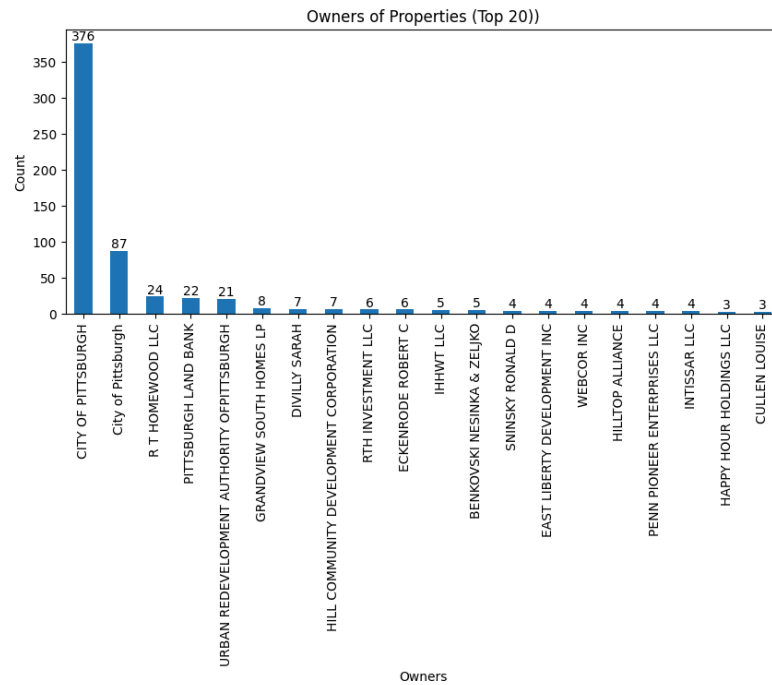
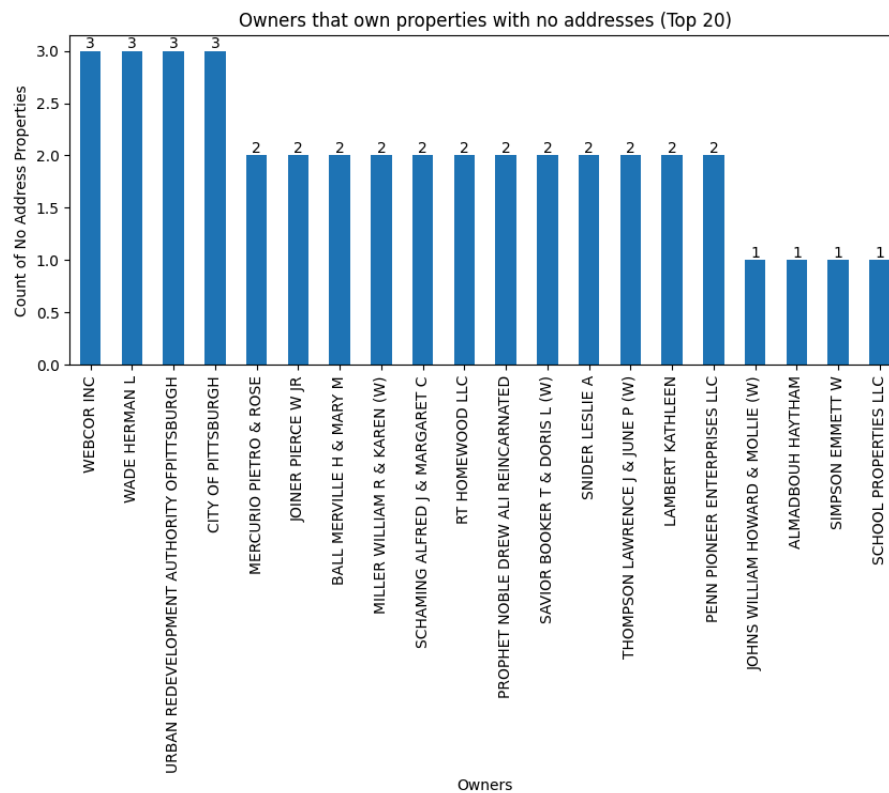


Figure V:



Insights

No Figure Dataset

- In our original **cleaned** data, City of Pittsburgh owned 460 properties
- The dataset claims that 470 of the properties are owned by the City of Pittsburgh
- 28 of the properties are claimed to be owned by the City of Pittsburgh when listed to be owned by another individual.

```
28 parcel id in city_df but not in city_props_df:  
{'0045E00331000000', '0057G00191000000', '0014A00262000000', '0046N00197000000', '0026J00075000000'}
```

- Filled in all 138 entries with matching ID's from City Property dataset that matched up with our parcel ID to update the 'No primary address specified' out of 470 entries.
- 87 entries' out of the 470 owners were updated based on the date of the entry.
- Lowered 647 missing address entries to 509. (saved in final_draft_df.csv) → this is difference of 647 and 138

Figures Insights

Figure I + II

- The 3 Score 4 Properties were all missing valid addresses
- After all addresses from Allegheny and City Properties were added, we are only missing one now
- Since the inspection status is inactive, we could push for obtaining private demolition permit or a building permit to repair

Figure III

- ~18.2% (509) of the address is still not valid, lowered from ~27.3% (872)
- 363 properties have been updated to a valid address, 138 properties updated from after Allegheny dataset

Figure IV

- ~13.4% (376) of the properties are owned by City of Pittsburgh, lowered from ~16.4% initial numbers (503)
- Means some of the owners were updated from the script replacing the owner with the latest entry

Figure VI

- All properties owned by City of Pittsburgh was updated to valid address