Improving Neighborhood Health in Baltimore City

An Analysis of Vacant Buildings and the Efficacy of DHCD Interventions

PITCH

Vacant buildings, believed by the Department of Housing and Community Development (DHCD) to damage neighborhood health, are most prevalent in the “black butterfly” region of Baltimore City and properties in this area are also the most likely to recur on the DHCD’s list of vacancies after being removed. Buildings are removed from the Vacant Building Notice (VBN) list as a result of either demolition or DHCD intervention. While demolitions are obviously highly effective at permanently removing properties from the VBN list, DHCD interventions have also proven highly effective at abating VBNs, especially homestead tax credits. However, further analysis is needed into the factors which cause buildings to become vacant and dilapidated before this problem can be fully solved.

PROBLEM STATEMENT AND BACKGROUND

The Baltimore City Code defines vacant buildings as “nuisances per se” and the Department of Housing and Community Development (DHCD) uses them as an indicator of neighborhood health.[1] As we found in our analysis, vacant buildings are most prevalent in the “black butterfly” region of Baltimore City, the predominantly black, impoverished neighborhoods “fanning across the city’s eastern and western halves.”[2] Properties in that area are also the most likely to relapse back into the Vacant Building Notice (VBN) list.

Baltimore has thus far pursued two prongs of intervention: (1) demolishing the vacant buildings and (2) facilitating investments in vacant lot programs, especially receivership, and supporting large-scale redevelopments in distressed areas. The city believes that fixing vacant, dilapidated buildings and putting them back on the market will encourage new residents and businesses to move into the area, thereby improving neighborhood health.[3]

Cities such as Cleveland, Flint, and Detroit have proven that careful coordination of strategies like land banks and managed sales (instead of auctions) not only effectively get vacant buildings rehabilitated and reoccupied, but also help prevent repeated abandonment, discourage local crime, and improve property values and urban health. However, intervention after the fact is not enough to solve the problem. To that end, both New York State and Washington State have recently passed laws to block foreclosed properties from quickly or easily becoming vacant and rundown, thereby tackling the problem at (one of) its source(s).[4]

Thus we ask: **What are the main causes of Vacant Building Notices and what is the impact of DHCD interventions, regarding VBNs, on overall neighborhood health? Which strategies, if any, are most effective?**

---

[1] Kathleen E. Byrne, “Receivership 101” PowerPoint Presentation, Baltimore City Department of Housing and Community Development, http://www.vacantstovalue.org/content/docs/ReceivershipWebinarD.pdf: Slide 4.

[2] Lionel Foster, “‘"The Black Butterfly’: Racial Segregation and Investment Patterns in Baltimore,” Urban Institute, 5 February 2019, https://apps.urban.org/features/baltimore-investment-flows/.

[3] Byrne, “Receivership 101,” Slide 15.

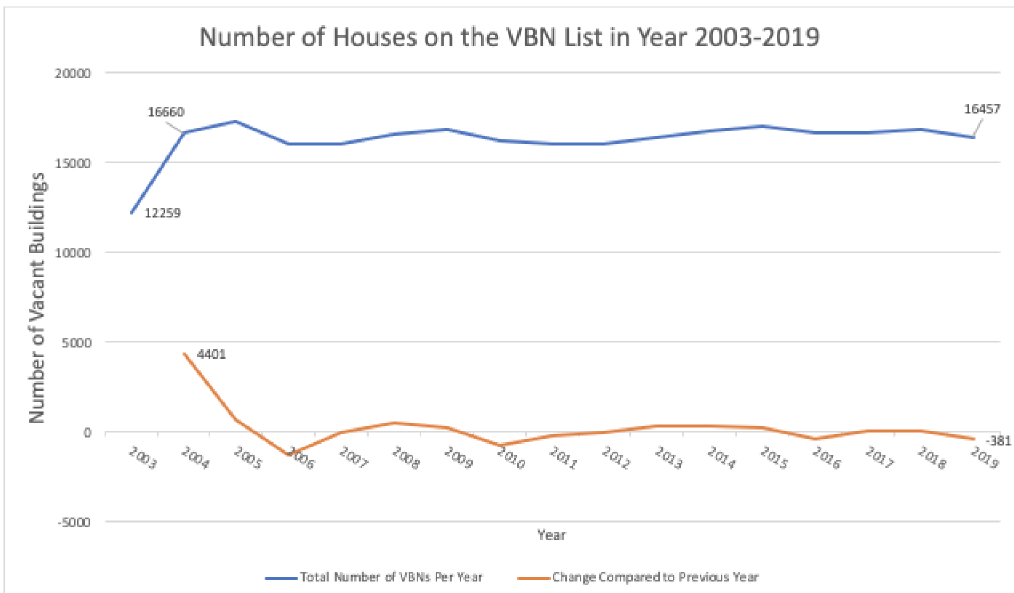
[4] Denise-Marie Ordway, “Zombie property: What research says about abandoned buildings,” Journalist’s Resource, 11 May 2018, https://journalistsresource.org/studies/government/municipal/abandoned-buildings-revitalization/.

DATA FINDINGS AND INTERPRETATIONS

Methodology explanations for all data analysis can be found in the Appendix of this report.

Overview: Trendline (2003-2019)

The given dataset contains 87,031 VBN data points throughout 1977 to 2019, of which only 41,340 data points are unique values. This means that only 41,340 buildings were recorded with single or multiple notice or other change in status occurred. We created a trendline to get an overview of how the number of vacant buildings has changed since 2003. The orange line shows the change in the number of VBNs compared to previous year. The change in number drops after 2004 due to DHCD interventions and has been stabilized since then with the number of VBNs on the list equals ~16,500 each year.

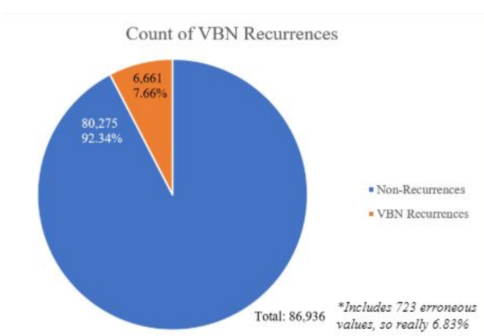


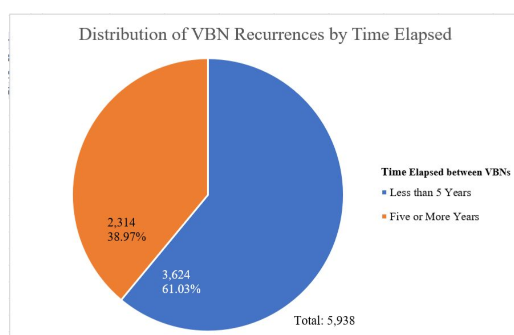
Question 1: How many VBNs are being put back on the market, then end up back on the VBN list within 1-year increments over a 5-year period? Where is this most likely to occur?

Dataset: "LatestVacancyExport1.31.2020" aka Vacant Building Dataset

Types of Analysis: IF Statement, Cluster Analysis, and Python Geospatial Analysis

Findings:

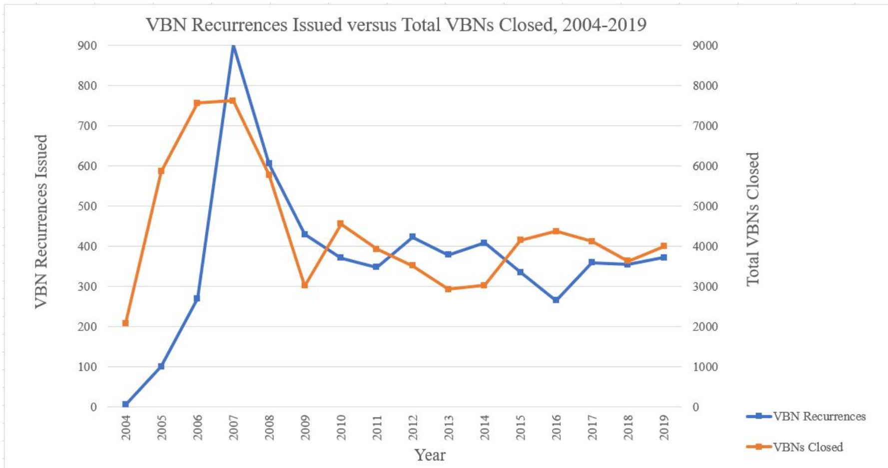




1. 5,938 VBN recurrences (6.83% of total values analyzed), of which 2,314 (38.97% of recurrences) recurred 5 or more years after close of property's initial VBN and 3,624

(61.03% of recurrences) within 5 years. Recurrences only account for a small portion of total VBNs, but recur quickly and in neighborhoods with consistently high VBN counts.

   a. Average years elapsed was 4.58 years. Maximum was 15.01 years.



2. Recurrences issued per year peaked at 901 in 2007 and stabilized around 350-400 per year since about 2009. Recurrences issued per year appear inverse to total VBNs closed per year, suggesting years of high closures alternating with years of high issuances.

Top 20 Neighborhoods by Average Cluster Placement

Rank	Neighborhood	2016-20	2011-15	2006-10	2003-5	Average	2003-Present
1	**BROADWAY EAST**	1	1	1	1	1	1
2	**SANDTOWN-WINCHESTER**	1	1	1	1	1	1
3	HARLEM PARK	1	1	1	1	1	1
4	**OLIVER**	1	1	1	1	1	1
5	**CARROLLTON RIDGE**	1	1	2	1	1.25	1
6	**CENTRAL PARK HEIGHTS**	1	1	2	1	1.25	1
7	**MIDDLE EAST**	2	1	1	1	1.25	1
8	UPTON	2	1	2	1	1.5	1
9	**COLDSTREAM HOMESTEAD MONTEBELL**	2	1	2	2	1.75	2
10	MIDTOWN-EDMONDSON	2	1	2	2	1.75	1
11	**WASHINGTON VILLAGE**	2	2	2	1	1.75	2
12	EAST BALTIMORE MIDWAY	2	2	2	2	2	2
13	FRANKLIN SQUARE	2	2	2	2	2	2
14	**SHIPLEY HILL**	2	2	2	2	2	2
15	PENN NORTH	2	2	2	2	2	2
16	MONDAWMIN	2	2	2	2	2	2
17	NEW SOUTHWEST/MOUNT CLARE	2	2	2	2	2	2
18	**MCELDERRY PARK**	2	2	2	2	2	2
19	MOSHER	2	2	2	2	2	2
20	COPPIN HEIGHTS/ASH-CO-EAST	2	2	2	2	2	2

3. Per the cluster analysis, the relative VBN count per neighborhood has been very consistent since 2003. The top neighborhoods for VBN recurrences (in bold in the visualization) are consistently above average for VBN count, therefore the neighborhoods most likely to have VBNs are also most likely to have recurrences.
4. The top ten neighborhoods with the greatest number of VBNs over the years are clustered around east and west Baltimore as shown below[5].

[5] Due to the lack of longitudes and latitudes for each building address in the given dataset, we could not do a geospatial analysis for VBN occurrences. Even though we searched on the Open Baltimore website hoping to convert the address to the format we wanted with the support of several other datasets, the datasets we downloaded were either with information missing or didn't align with the information we had.

Additional Notes:

1. This analysis assumed that properties put back on the market wouldn't reappear on the VBN list within the same calendar year and that, when ownership of a property changes, the new VBN would be reissued in the same year that the previous VBN is closed.
2. The VBN total for the analysis, 86,938, doesn't exclude VBNs issued in error or due to ownership changes. This likely minimized the percentage which was recurrences.
3. The calculation of years elapsed counts from the property's first Date_Closed, *not* the Date_Closed of the last VBN before the recurrence. This likely inflated some calculations.
    a. However, the average number of repeats per property is only 1.97 for the entire dataset and 2.67 for the properties which repeat (which is 58% of the unique property IDs within the dataset), so this probably isn't a significant issue.

Question 2: Are there particular DHCD interventions – homeowner incentive grants, tax credits, etc. - that is more likely to keep a building off the VBN list?

Dataset:
https://data.baltimorecity.gov/Neighborhoods/Housing-and-Community-Development-2010-2014-/mvvs-32jm

Types of Analysis: Linear Regression and Correlation

Findings:

1. Firstly, this dataset compared many different variables, including homtax (Number of of Homestead Tax Credits per 1,000 Residential Units), owntax (Number of Homeowner Tax Credits per 1,000 Residential Units), demper (Number of Demolition Permits per 1,000 Residential Units), and resrehab (Percentage of Properties with Rehabilitation Permits Exceeding $5,000), with vacant (Percentage of Properties that are Vacant). These variables were compared by neighborhood in Baltimore and data ranges from 2011 - 2014.

2. After filtering the related variables, linear regression analysis led to the evaluation that homtax is the only statistically relevant variable with a coefficient of -0.039 (2011), -0.040 (2012), -0.044 (2013), -0.041 (2014). It's t statistic was -7.34 and the p-value was $1.15*10^{-9}$. The other variables - owntax, demper, resrehab - have t stats less than 2 and p-values greater than 0.05, therefore they are statistically irrelevant.

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.706726926
R Square	0.499462948
Adjusted R Square	0.490193743
Standard Error	6.919046367
Observations	56

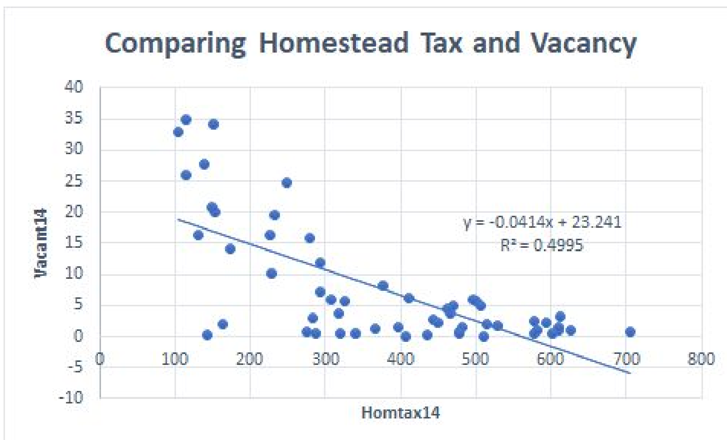
ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	2579.605451	2579.605451	53.88412116	1.15291E-09
Residual	54	2585.152942	47.87320263		
Total	55	5164.758393			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	23.24084825	2.315780629	10.03585917	6.02387E-14	18.59798763	27.88370887	18.59798763	27.88370887
homtax14	-0.041417335	0.005642243	-7.340580438	1.15291E-09	-0.05272935	-0.03010532	-0.05272935	-0.03010532

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.732746665
R Square	0.536917674
Adjusted R Square	0.500597492
Standard Error	6.848083069
Observations	56

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	4	2773.050065	693.2625162	14.78290137	4.37872E-08
Residual	51	2391.708328	46.89624172		
Total	55	5164.758393			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	19.96780818	2.885047811	6.921135971	7.25118E-09	14.17583302	25.75978335	14.17583302	25.75978335
homtax14	-0.043927549	0.00950106	-4.623436862	2.60871E-05	-0.063001722	-0.024853376	-0.063001722	-0.024853376
owntax14	0.034366064	0.060299533	0.569922554	0.5712324	-0.0866903	0.155422427	-0.0866903	0.155422427
demper14	0.283856264	0.190700967	1.48848885	0.142779893	-0.098991904	0.666704431	-0.098991904	0.666704431
resrehab14	0.661264913	0.61157153	1.081255227	0.284670016	-0.566516165	1.889045991	-0.566516165	1.889045991



Additional Notes:

1. More data and more recent data would be necessary to come to resolute conclusions.
2. Even though the other variables were statistically irrelevant, it is key to note that they had a positive relationship with vacancy in the linear regression analysis (owntax and homtax had a negative correlation with vacant).

Question 3: Is there a correlation between VBNs and Demolitions in a Neighborhood?

Dataset:
https://data.baltimorecity.gov/Neighborhoods/Housing-and-Community-Development-2010-2014-/mvvs-32jm

Types of Analysis: Linear Regression and Correlation

Findings:

1. Using the dataset from the prior data question, there is a positive correlation between vacant (Percentage of Properties that are Vacant) and demper (Number of Demolition Permits per 1,000 Residential Units), however the correlation coefficient is minor. The correlation coefficients were: 0.52 (2011), 0.41 (2012), 0.41 (2013), 0.36 (2014).

	homtax14	owntax14	demper14	resrehab14	vacant14
homtax14	1				
owntax14	0.801956428	1			
demper14	-0.311837972	-0.264217181	1		
resrehab14	0.145041485	0.220433489	0.004251337	1	
vacant14	-0.706726926	-0.524868353	0.359354246	0.018254047	1

	homtax11	owntax11	demper11	resrehab11	vacant11
homtax11	1				
owntax11	0.756863621	1			
demper11	-0.529171233	-0.413887587	1		
resrehab11	0.065759835	0.297111665	-0.125363652	1	
vacant11	-0.743873138	-0.568056708	0.521239472	0.002174698	1

Additional Notes:

1. As stated before in the previous question, more data and more recent data would improve research results.
2. Important to note, the correlation between demper and vacant has been decreasing over the years. Further research into this topic could be interesting.

Question 4: How has the concentration of VBNs changed over time?

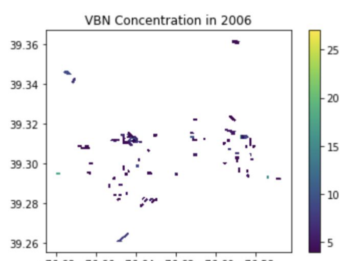
Types of Analysis: Python Choropleth Mapping

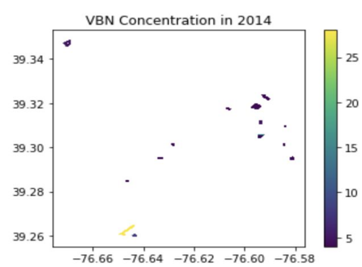
Findings:

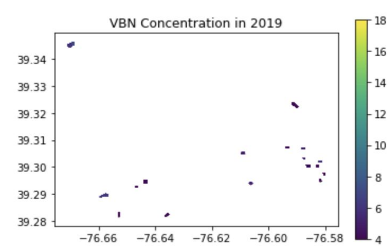
```
array([ 4.,  5., 11., 10.,  9.,  8.,  3.,  2.,  7.,  6., 12.,  1.])
```

```
array([2014., 2006., 2004., 2005., 2011., 2019., 2012., 2015., 2018.,
       2007., 2016., 2001., 2010., 1999., 2017., 2008., 2009., 2020.,
       2013., 2003., 2002., 1998., 2000., 1996.])
```

1. In a year, April and May are the months with the highest number of VBNs and January the least. The years 2014 and 2006 had the highest numbers of VBNs. More analysis on policies/practices and conditions during these months and years to identify correlations.







2. There's no clear pattern or areas of heavy concentration in the choropleth maps over time, however the numbers mainly lie in the Black Butterfly especially during the 2000s. There's a significant reduction in VBN's by 2019. This suggests that reccurances aren't very likely for most neighborhoods (likely from DHCD interventions, but more analysis is needed) unless they have consistently high VBN counts in the first place.

```
array(['Mount Winans', 'South Clifton Park', 'Upton', 'Carrollton Ridge',
       'Middle East', 'Penn North', 'Edgewood', 'Central Park Heights',
       'Patterson Park Neighborhood', 'Broadway East',
       'Coppin Heights/Ash-Co-East', 'Mondawmin', 'Johnston Square',
       'CARE', 'Reservoir Hill', 'Harlem Park', 'Boyd-Booth', 'Walbrook',
       'Sandtown-Winchester', 'Oliver', 'Barclay', 'Franklin Square',
       'Curtis Bay', 'Milton-Montford', 'Midtown-Edmondson',
       'Druid Heights', 'Remington', 'Winston-Govans',
       'Parkview/Woodbrook', 'Gay Street', 'McElderry Park',
       'Baltimore Highlands', 'Darley Park', 'Downtown', 'Brooklyn',
       'New Southwest/Mount Clare', 'Coldstream Homestead Montebello',
       'Greenmount West', 'Harwood', 'Berea',
       'Washington Village/Pigtown', 'Park Circle',
       'East Baltimore Midway', 'Northwest Community Action', 'Mosher',
       'Oldtown', 'Shipley Hill', 'Biddle Street',
       'Rosemont Homeowners/Tenants', 'Jonestown',
       'Ellwood Park/Monument', "Butcher's Hill", 'Better Waverly',
       'Allendale', 'Saint Josephs', 'Patterson Place',
       'Penrose/Fayette Street Outreach', 'Madison-Eastend', 'Poppleton',
       'Millhill', 'South Baltimore', 'Ramblewood', 'Hollins Market',
       'Garwyn Oaks', 'Pimlico Good Neighbors', 'Sharp-Leadenhall',
       'Pen Lucy', 'Franklintown Road', 'Mid-Govans', 'Union Square',
       'Greenspring', 'Dorchester', 'Easterwood', 'Irvington', 'Canton',
       'Upper Fells Point', 'Four By Four'], dtype=object)
```

(N-hoods ordered most to least amount of VBNs)

3. Neighborhoods that hold the most VBNs in the more recent years (2016-2020) fluctuate year to year. Most neighborhoods are significant in only one year, but notable reoccurrences include: Middle East (2016, 2017, 2019), McElderry park (2016-2019), Central Park Heights (2016, 2017, 2019), Coldstream Homestead Montebello (2017-2019), Harlem Park (2016-2019), and Park Circle (2017, 2018).

Additional Notes:

1. Data analysis is based on notices that have both the block lot and neighborhood.
2. Results are a compilation of the top 10M notices.
3. The aggregated dataset is much too large to efficiently run on Python.
4. VBN dataset's addresses aren't effective location identifiers in analysis when merging datasets. Block lot is a better identifier, but having geographical coordinates would be ideal in creating visualizations.

Question 5: Are there any properties DHCD has spent capital on that then gets on the VBN list within five years?

Dataset: "DHCD Interventions and Strategies" and VBN List

Types of Analysis: VLOOKUP and Bar Graphs

Findings:

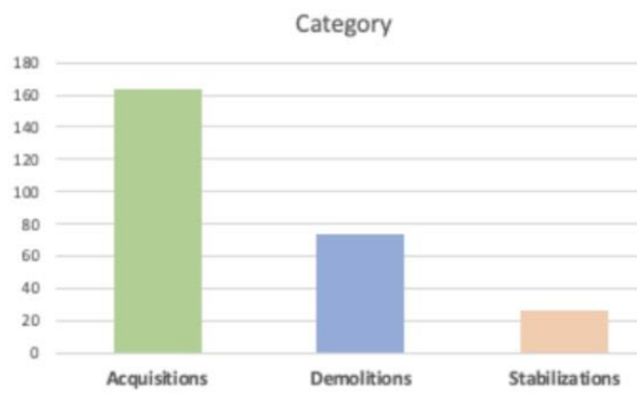
1. "DHCD Interventions and Strategies" show the properties that DHCD spent capital on. So I used VLookUp to match the list that is on the data set to the VBNs in the list to see which properties are still on the list even with DHCD's support.

:ATEGORY	
'ROJECT FINANCE	Development Project
{EHAB	Funds for home rehabilitations
VEATHERIZATION	Program for improving utility efficiency in homes and protecting the home from the elements
EAD	Lead paint mitigation program
\CQUISITION	DHCD acquires properties for development projects and/or demolition
IOMEBUYING INCENTIVES	A variety of programs that promote homeownership
:TABILIZATIONS	Funds for stabilizing properties with roof problems, or other issues that put residents/pedestrians in danger
)EMOLITIONS	Property Demolition

2. These are different types of investments and supports from the DHCD:

1	Block	Lot	Blocklot	HouseNum	Direction	StreetName	StreetAttr	NoticeNum	DateNotice	VLookUp
520	1477	116	1477 116	1704		CRYSTAL	AVE	833013-19	6/22/92	DEMOLITIONS
572	4137	028C	4137 028C	2710		KENNEDY	AVE	930056-19	11/5/92	DEMOLITIONS
661	1587	144	1587 144	909	N	CASTLE	ST	044007-19	4/20/93	ACQUISITION
879	1477	114	1477 114	1708		CRYSTAL	AVE	299010-19	4/27/94	DEMOLITIONS
889	1477	103	1477 103	1709		CRYSTAL	AVE	305020-19	5/5/94	DEMOLITIONS
1022	3935	111	3935 111	2758		TIVOLY	AVE	394023-19	9/12/94	ACQUISITION
1261	4171	122	4171 122	1938		PERLMAN	PL	565052-19	5/2/95	ACQUISITION
1634	1587	110	1587 110	914	N	CASTLE	ST	789047-19	3/6/96	ACQUISITION
1967	4171	154	4171 154	1921	N	COLLINGTO	AVE	962048-19	11/6/96	DEMOLITIONS
2369	3207	101	3207 101	1910		HERBERT	ST	140028-19	7/28/97	ACQUISITION
2827	2176	103	2176 103	2535		EMERSON	ST	306032-19	4/2/98	STABILIZATIONS
3360	4137	027A	4137 027A	2702		KENNEDY	AVE	440044-19	10/9/98	DEMOLITIONS
3369	4171	156	4171 156	1925	N	COLLINGTO	AVE	442047-19	10/14/98	DEMOLITIONS
3541	4071	062D	4071 062D	2716		BOONE	ST	492025-19	12/29/98	DEMOLITIONS
3995	1477	104	1477 104	1711		CRYSTAL	AVE	626011-19	7/1/99	DEMOLITIONS
4046	4130	028B	4130 028B	1617		ABBOTSTON	ST	638010-19	7/20/99	STABILIZATIONS
4094	1568	108	1568 108	1003	N	CASTLE	ST	649020-19	8/4/99	ACQUISITION
4713	2171	044A	2171 044A	1	S	FRANKLINTI	ROAD	810013-20	3/9/00	DEMOLITIONS
5501	3457	023B	3457 023B	2435		LAKEVIEW	AVE	965059-20	10/16/00	ACQUISITION
5635	3453	001H	3453 001H	2419		CALLOW	AVE	002033-20	12/6/00	ACQUISITION
6104	2411	021B	2411 021B	2923		WESTWOOI	AVE	097024-20	4/26/01	ACQUISITION
6261	3457	023A	3457 023A	2431		LAKEVIEW	AVE	127004-20	6/8/01	ACQUISITION
6754	2704	001A	2704 001A	4013		PENHURST	AVE	199030-20	9/24/01	DEMOLITIONS
6783	1448	112	1448 112	1825	N	CHAPEL	ST	204022-20	10/1/01	DEMOLITIONS
7008	3453	002B	3453 002B	2443		CALLOW	AVE	233082-20	11/12/01	ACQUISITION
7041	2407	038D	2407 038D	2910		WESTWOOI	AVE	238043-20	11/19/01	STABILIZATIONS
7248	2132	028A	2132 028A	401		FONTHILL	AVE	272000-20	1/10/02	DEMOLITIONS
7429	3004	069F	3004 069F	2933		HERBERT	ST	299056-20	2/21/02	DEMOLITIONS
7510	3301	121	3301 121	3429		SAINT AMBF	AVE	311017-20	3/11/02	DEMOLITIONS
7738	3453	002A	3453 002A	2441		CALLOW	AVE	554086-20	4/19/02	ACQUISITION
7869	3453	002C	3453 002C	2445		CALLOW	AVE	354030-20	5/10/02	ACQUISITION
8495	3453	001J	3453 001J	2427		CALLOW	AVE	433045-20	8/27/02	ACQUISITION
8551	3935	106	3935 106	2748		TIVOLY	AVE	442067-20	9/10/02	ACQUISITION
8834	3408	119	3408 119	2512		SALEM	ST	474042-20	10/25/02	ACQUISITION
8862	3207	100	3207 100	1912		HERBERT	ST	478009-20	10/31/02	DEMOLITIONS
8863	3207	102	3207 102	1908		HERBERT	ST	478010-20	10/31/02	DEMOLITIONS
8883	3234	015M	3234 015M	2327		BRYANT	AVE	480024-20	11/4/02	DEMOLITIONS
9208	3935	104	3935 104	2744		TIVOLY	AVE	521027-20	1/2/03	ACQUISITION
9439	4572	001I	4572 001I	4019	W	BELLVEDERE	AVE	551150-20	2/19/03	DEMOLITIONS
9851	3453	001I	3453 001I	2423		CALLOW	AVE	588081-20	4/15/03	ACQUISITION
10152	1477	117	1477 117	1702		CRYSTAL	AVE	609051-20	5/16/03	ACQUISITION
10229	1477	118	1477 118	1700		CRYSTAL	AVE	611070-20	5/20/03	DEMOLITIONS
10254	1591	102	1591 102	902	N	PORT	ST	612012-20	5/21/03	ACQUISITION
10571	4572	004B	4572 004B	4004	W	GARRISON	AVE	633131-20	6/20/03	STABILIZATIONS
10742	1477	113	1477 113	1710		CRYSTAL	AVE	649011-20	7/15/03	DEMOLITIONS
10778	3935	103	3935 103	2742		TIVOLY	AVE	661046-20	7/22/03	ACQUISITION
10810	2407	038C	2407 038C	2912		WESTWOOI	AVE	662056-20	7/23/03	STABILIZATIONS
11005	3935	131	3935 131	2794		TIVOLY	AVE	507267-20	8/12/03	ACQUISITION
11006	3935	132	3935 132	2794		TIVOLY	AVE	507268-20	8/12/03	ACQUISITION

3. Since both of the datasets I am using provide "BlockLot," I used VLOOKUP to match the lots of the data. If the BlockLot matches, that means that even though DHCD invested on those buildings are not vacant buildings, they still were on the VBN list
4. After filtering all the #N/As, which are the BlockLots that are not in the VBN list anymore, there are only lists of BlockLots that are still on the VBN list. The greenbox above shows the different kinds of interventions that DHCD spent capital on.



5. There are 12,819 VBNs that DHCD spent capital on.There are 86,038 VBNs on the list. After the analysis there were 265 VBNs that were still on the list even with DHCD's support.

Additional Notes:

1. For future suggestions, I think there should be a research for DHCD to find which method of investments is the best way to keep the VBNs not to get back on the list.
2. For example, the method "acquisitions" seems the worst way to keep the vacant buildings out of the list.

CONCRETE RECOMMENDATIONS AND IMPACT FOR THE DHCD

- VBN recurrences are most prevalent in the neighborhoods which consistently have the highest VBN counts. This may make targeting neighborhoods for DHCD intervention more straightforward in the future.
- According to this analysis, recurrences only account for less than 7% of total VBNs. Coupled with the observation that DHCD interventions have a very high
- Neighborhoods to focus on in the coming years include: Middle East, McElderry Park, Central Park Heights, Coldstream, Harlem Park, and Park Circle; they have the highest recurrences in recent years.
- Additional research on policies and practices enabled in 2006/2014 as well as the months April and May in a given year that evoke a higher number of issues. More data on specific DHCD practices implemented in certain years and neighborhoods to analyze their impact on the number of VBNs if any.
- Of the DHCD interventions, increasing homestead tax credits are the most sound method of decreasing vacancy in Baltimore. As seen by the regression analysis for 2014, an increase of 1 in the number of homestead tax credits per 1,000 residential units should decrease the percentage of vacant properties by 0.04%. With the average vacancy among the neighborhoods being 7.66% and the average number of homestead tax credits being 376, increasing the number of homestead tax credits by 6-7% will decrease the average vacancy by 1% (about 13% of the current average vacancy).
- Further research into the factors which cause vacancy in the first place would explain why annual VBN counts have been fairly stable over the last 17 years, despite the observed effectiveness of DHCD interventions in abating VBNs.

APPENDIX

Methodology:

<u>VBN Trendline (2003-2020)</u>:

1. Identify the notices active in different years and sort out a list of VBNs on record each year using an IF/OR formula in excel.
    a. For example, for the first record in year 2003, we used "=IF($H2<AE$1,IF(OR($AB2>AE$1,$AB2=""),1,0))" to get a logic statement of "0", "1" or "FALSE".
    b. If "1" is shown in the cell, it means that the building was on the VBN list in the particular year (i.e. with date notice before the date in yellow cell and date closed after); on the other hand, if it's "0", it means that the building was once on the list but then got removed before the assigned date.

G	H	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
NoticeNum	DateNotice	Date_Closed	BlockLot	Nhood2010	12/31/03	12/31/04	12/31/05	12/31/06	12/31/07	12/31/08	12/31/09	12/31/10
871000-2004	5/11/04	1/18/06	0001 003	Easterwood	=IF($H2<AE$1,IF(OR($AB2>AE$1,$AB2=""),1,0))					0	0	0
109318A	1/18/06	11/18/08	0001 003	Easterwood	FALSE	FALSE	FALSE	1	1	0	0	0
378476A	11/26/08	12/24/08	0001 003	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	0	0	0
398656A	12/26/08	1/25/12	0001 003	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	1	1	1
805231A	1/25/12		0001 003	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1049399A	1/7/14	4/9/14	0001 004	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1780434A	4/20/19		0001 004	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1174830A	12/2/14		0001 006	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
144000-2001	7/3/01	1/18/06	0001 007	Easterwood	1	1	1	0	0	0	0	0
109288A	1/18/06	5/4/07	0001 007	Easterwood	FALSE	FALSE	FALSE	1	0	0	0	0
927919A	2/1/13		0001 007	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
853000-2004	4/20/04	4/21/06	0001 008	Easterwood	FALSE	1	1	0	0	0	0	0
130648A	4/24/06	5/19/08	0001 008	Easterwood	FALSE	FALSE	FALSE	1	1	0	0	0
319263A	5/30/08	12/23/08	0001 008	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	0	0	0
398200A	12/30/08	9/23/11	0001 008	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	1	1	1
765916A	1/13/12	6/30/15	0001 008	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1259178A	7/1/15		0001 008	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1349596A	1/28/16		0001 015	Easterwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

2. Sort out the "1"'s for every given year, and generate the total number of VBNs on the list as well as the change each year.
    a. Use VLOOKUP to generate a list of the VBNs and sum up the values of each year.
    b. Create a chart with the information generated using excel.

<u>VBN Recurrence Analysis</u>:

1. A unique property ID was created for each street address (e.g. 2110NFULTONAVE for 2110 N. Fulton Avenue). The dataset was organized chronologically by the DateNotice and alphabetically by the property IDs, so that all VBNs for the same property were adjacent to one another and in chronological order.
2. <u>92 VBNs</u> had no street address and were excluded, leaving ***86,938 total VBNs.***
3. The following nested IF statement was then used to count all properties which recurred in the dataset for reasons other than "Ownership Changed."

=IF(G18=G17,IF(YEAR(I18)=YEAR(AD17),0,IF(M17="OWNERSHIP CHANGED",0,1)),0)

D	E	F	G	H	I	AC	AD	AE	AF	AG	AH	A
irection	StreetNam	StreetAttr	Full_St_Address	NoticeNum	DateNotice	VBN_Repeats	Date_Closed	Year_Closed	BlockLot	Nhood2010	1/1/2020	1/1
	EDMONDS	AVE	1000EDMONDSON.	292012-20	2/8/2002	0	2/25/2005	2005	0115 044	Harlem Park	0	
	EDMONDS	AVE	1000EDMONDSON.	370967A	10/22/2008	1			0115 044	Harlem Park	1	
	FEDERAL	ST	1000EFEDERALST	633010-20	6/20/2003	0	11/5/2004	2004	1112 048	Oliver	0	
	FEDERAL	ST	1000EFEDERALST	29049A	11/4/2004	=IF(G18=G17,IF(YEAR(I18)=YEAR(AD17),0,IF(M17="OWNERSHIP CHANGED",0,1)),0)						
	FEDERAL	ST	1000EFEDERALST	286932A	2/14/2008	0	9/5/2014	2014	1112 048	Oliver	0	
	LANVALE	ST	1000ELANVALEST	416089-19	9/4/1998	0	11/25/2005	2005	1098 032	Oliver	0	
	ELLICOTT	DR	1000ELLICOTTDR	851040-20	4/16/2004	0	8/23/2005	2005	2447 049	Winchester	0	
	ELLICOTT	DR	1000ELLICOTTDR	126558A	3/28/2006	1	8/30/2006	2006	2447 049	Winchester	0	
	ELLICOTT	DR	1000ELLICOTTDR	156551A	8/31/2006	0	10/23/2007	2007	2447 049	Winchester	0	

   a. **VBN_Repeats = IF (G18 = G17, IF [YEAR(I18) = YEAR(AD17), 0, IF{M17 = "OWNERSHIP CHANGED", 0, 1}], 0)**
   b. where 1 indicates a VBN recurrence,
   c. 0 is all other possibilities (e.g. new property or extension of the preceding VBN),
   d. "VBN_Repeats" is the column heading,
   e. G18 is the unique ID of the property in question (row 18),
   f. G17 is the unique ID of the preceding property in the dataset (row 17),
   g. I18 is the issue date of the VBN in row 18,
   h. AD17 is the close date of the VBN in row 17, and
   i. M17 is the reason the preceding VBN was closed.
4. The number 1 should have been returned for all VBNs for which the first IF statement is true, but the second and third are false. → ***This produced 6,661 results.***
5. The VLOOKUP and LOOKUP functions were used to find the issue date for each VBN recurrence and the close date of the first VBN issued for each property which had a recurrence. The years elapsed between the two was then calculated. For 723 VBN recurrences, this returned either a negative, zero, or error value.
6. Excluding these 723 "erroneous values," ***5,938 VBN recurrences remained***.

Cluster Analysis:

1. The cluster analysis grouped neighborhoods according to each's VBN count on January 1st in five-year (or the case of the first analysis, three-year) increments for a total of four cluster analyses: 2003 to 2005, 2006 to 2010, 2011 to 2015, and 2016 to 2020.
   a. For example, the first analysis looked at three data points per neighborhood: VBN count on 1/1/2003, 1/1/2004, and 1/1/2005.
2. Z-scores were calculated for each data point to determine how many standard deviations were between the data point and the average VBN count per year. The sum of $(x\text{-}y)^2$ was calculated for the z-scores of each data point (range of y-values) and randomly chosen cluster node (range of x-values) and used to determine which cluster each data point was "closest" to. These distances were summed and that sum minimized with Excel Solver.
3. After using Excel Solver to perform the analysis and determine the ideal z-scores for each cluster, I reassigned the cluster ranks for each five-year analysis such that Cluster #1 had the most positive z-scores (farthest above the mean) and Cluster #3 had the most negative z-scores (farthest below the mean), which #2 being somewhere in the middle.
   a. The z-scores varied between the analyses, but tended to be about 4±1 for Cluster #1, 0±1 for #2, and -0.5±1 for #3, as demonstrated from the 2006-2010 analysis.

Cluster	Rank	Neighborhood	z_2010	z_2009	z_2008	z_2007	z_2006
1	16	MIDDLE EAST	4.9752747	4.9256402	4.8166698	5.0399278	5.0211613
2	25	JOHNSTON SQUAR	1.722565	1.7137348	1.7270578	1.665391	1.6245611
3	149	PARKSIDE	-0.419301	-0.423244	-0.415421	-0.435086	-0.430749

b. "Standardizing" the clusters this way across the analyses allowed for commentary on a neighborhood's relative change in VBN count over time.

Linear Regression and Correlation:

1. Imported the data from the dataset and filtered out the variables from the columns that would help answer the data questions.
2. Created a PivotTable with all of the variables and used the PivotTable to create relevant tables for each year. Then, separated each year's data into different spreadsheets
3. Using the Data Analysis application on Excel, created linear regressions (making the vacant column the Y-range and the other variables the X-range) using all of the variables. Upon discovering that a variable was not statistically relevant, linear regressions were attempted subtracting a variable each time.
4. Final regression analysis was completed with one variable remaining. This was continued for each year.
5. After the regression analysis was completed, correlation analysis was conducted on all of the variables for each year. This started by using the original separated spreadsheets for each year.
6. From there, using the Data Analysis application on Excel, created a correlation analysis using the range that contained all variables. This was continued for each year.

Python Choropleth Mapping:

Shape file data:

1. https://data.baltimorecity.gov/Geographic/Parcels-Shape/jk3c-vrfy
2. https://catalog.data.gov/dataset/tiger-line-shapefile-2019-county-baltimore-city-md-topological-faces-polygons-with-all-geocodes

Methodology:

1. Import data and packages onto Jupyter notebook.
2. Merge VBN data onto parcel dataset from Open Baltimore to match each block lot with a certain geometry; geometry should be crs = 4326.
3. Spatial merge that dataframe with census tract data.
4. Aggregate the merged df by counting the number of notices in each census tract in each year. Modify the df to get the top 10M occurrences of notices. Gather the years, months, and neighborhoods that show the most occurences.
5. Create a line/bar graph to see if there are any years/months that consistently have the most notices. Create choropleth graphs for years that have the most notices to map out concentration over time.

TEAM CONTACT INFORMATION

Marjorie Bowerman: mbowerm2@jhu.edu, (443) 889-8803
Austin Rose: arose28@jh.edu (870) 405-2516
Vesna Tran: vtran24@jh.edu (707) 805-8329
Eric WonJun Yang: wyang37@jhu.edu
Chenjiayi "Sanny" Ye: cye8@jhu.edu (443) 603-5718