**HW4 Investigation of 5 unusual records**

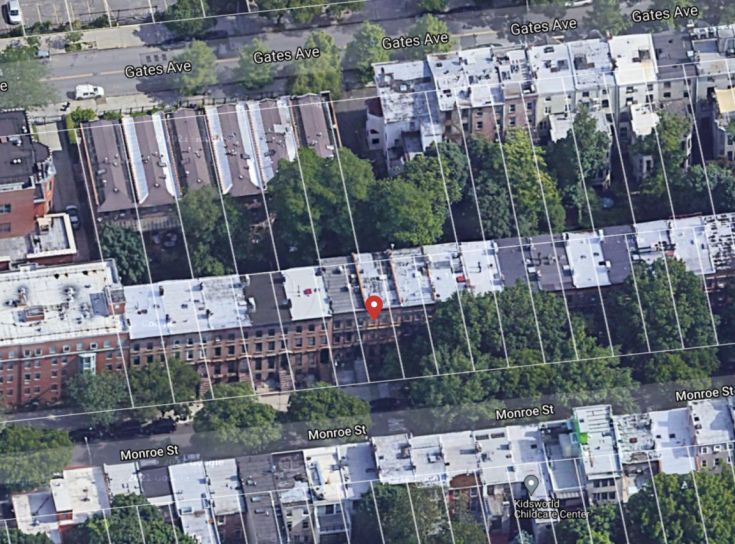
1.

**RECORD:** 333412

**Owner:** SPOONER ALSTON

**Address:** 37 MONROE STREET, NY11238

By using satellite through Google map, it is a building located in Brooklyn.



The reason why this record is suspicious is that it has a unusually high z-score on r2\_inv\_taxclass, r3\_inv\_taxclass, r8\_inv\_taxclass, and r9\_inv\_taxclass.

*The below tables shows the z-score respect to the suspicious fields:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | r2\_inv\_taxclass | r3\_inv\_taxclass | r8\_inv\_taxclass | r9\_inv\_taxclass |
| Z-score | 218.289 | 153.736 | 105.27 | 69.571 |

By the table above, we can observe that r2,r3,r8,r9 inverse are much higher than normal based on the taxclass. This means that base on the owner’s taxclass, he is far away from other owner who has the same taxclass. By further investigation, I found that this property has unusually high BLDFRONT(4017), and BLDDEPTH(42) which is much higher than its LTFRONT(17), LTDEPTH(85). Since r2\_inv=S2/V1, r3\_inv=S3/V1, r8\_inv=V3/S2, and r9\_inv=V3/S3 where S2= BLDFRONT\* BLDDEPTH, S3 = S2\*STORIES, they are all impacted by a large S2. Therefore, this might be the reason why it has a high fraud score.

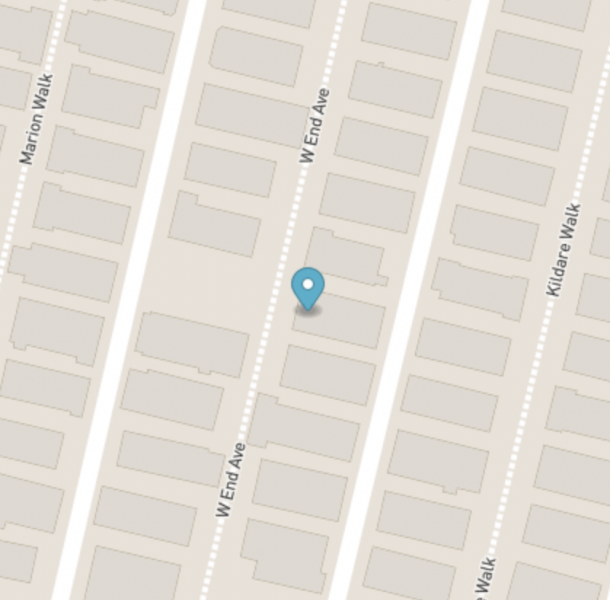
2.

**RECORD:** 934793

**Owner:** BREEZY POINT COOPERAT

**Address:** 217-02 BREEZY POINT BLVD, NY11697

By using satellite through Google map, it is a building located at Breezy Point neighborhood in Queens



The reason why this record is suspicious is that it has a unusually high z-score on r2, r3, r5, r6, r8, r9

*The below tables shows the z-score respect to the suspicious fields:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | r2 | r3 | r5 | r6 | r8 | r9 |
| Z-score | 220.901 | 455.584 | 23.982 | 50.973 | 14.589 | 41.682 |

By the table above, we can observe that r2, r3, r5, r6, r8, r9 are much higher than normal r-ratio. Since r2= V1/S1, r3 = V1/S3, r5 = V2/S2, r6 = V2/S3, r8=V3/S2, r9 = V3/S3, where S1=LTFRONT\*LTDEPTH, S2=BLDFRONT\*BLDDEPTH, S3=S2\*STORIES, V1=FULLVAL, V2=AVLAND, V3= AVTOT, I started to looked at these fields. By further investigation, I found that the FULLVAL is unusually high which is 273,000,000, this might be the reason why it has a high score on r2, and r3. Moreover, a high V2(AVLAND= 10920000),and a high V3(AVTOT=16380000) and low S2, S3 might also be the reason why it has a high score on r5,r6,r8,r9. Therefore, based on the investigation, record 934793 is suspicious and requires further investigation.

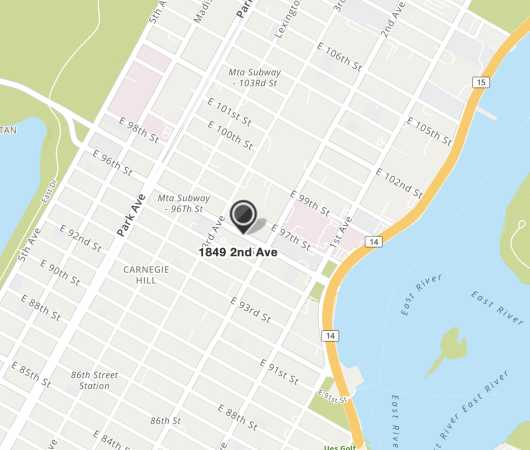
3.

**RECORD:** 116647

**Owner:** MF ASSOCIATES OF NEW

**Address:** 1849 2 AVENUE, NY10128

By using satellite through Google map, we find that it is an independent building located in Manhattan.



*The below tables shows the z-score respect to the suspicious fields:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | r1 | r4 | r7 | r1\_boro | r4\_boro | r7\_boro |
| Z\_score | 160.832 | 146.301 | 189.397 | 100.606 | 48.375 | 57.046 |

The reason why this record is suspicious is that it has a unusually high z-score on r1, r4, r7. Since r1=V1/S1, r4=V2/S1, r7=V3/S1, they all need to divided by S1 where S1 =LTFRONT\*LTDEPTH. However, the LTFRONT and LTDEPTH are 25 and 75 which means that the building will have a small S1. Moreover, V1(FULLVAL), V2(AVLAND), V3(AVTOT) are relatively large. Therefore, we have a large r1,r4,r7.

By further investigation, we found that the corresponding borough (r1\_boro, r4\_boro, r7\_boro) are also much higher than others in the same borough. Therefore, based on the investigation, record 116647 is suspicious and requires further investigation.

4.

**RECORD:** 584882

**Owner:** CHAN, KWOK FU

**Address:** 41-26 27 STREET, NY11101

By using satellite through Google map, we find that it is an independent building named Queens Plaza located in Long Island City,NY.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | LTFRONT | LTDEPTH | BLDFRONT | BLDDEPTH | FULLVAL | AVLAND | AVTOT |
| value | 0 | 0 | 0 | 0 | 525 | 6 | 236 |

According to our original data, this property is a building with 10 stories, however, there are 34 more record that has the same address but different owner, and their corresponding value of LTFRONT, LTDEPTH, BLDFRONT, BLDDEPTH are all 0(filled with average while doing data cleaning).

The reason why this record is suspicious is that it has a unusually high z-score on almost every inverse r field and their corresponding classes(ZIP5,ZIP3, tax-class and BORO). While cleaning our data, we fill the 0 values with the mean of the LTFRONT, LTDEPTH, BLDFRONT, BLDDEPTH with respect to the tax class. Since the values of FULLVAL, AVLAND, and AVTOT are low, with an average value of LTFRONT, LTDEPTH, BLDFORNT, BLDDEPTH filled in, the calculation of the inverse r field will become high. This might be the reason for why we have a high combined score. Therefore, based on the investigation, record 584882 is suspicious and requires further investigation.

5.

**RECORD:** 402297

**Owner:** CARNEGIE HALL TOWER M

**Address:** 152 WEST 57 STREET, NY10019

By using satellite through Google map, we find that it is a building located in New York.



The reason why this record is suspicious is that it has a unusually high z-score on r1, r2, r4, r7, r8

*The below tables shows the z-score respect to the suspicious fields:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | r1 | r2 | r4 | r7 | r8 |
| Z-score | 39.219 | 20.014 | 38.435 | 46.441 | 10.131 |

By the table above, we can observe that r1, r2, r4, r7, r8, are much higher than normal r-ratio. Since r1= V1/S1, r2 = V1/S2, r4 = V2/S1, r7 = V3/S1, r8=V3/S2, where S1 = LTFRONT\*LTDEPTH, S2=BLDFRONT\*BLDDEPTH, V1=FULLVAL, V2=AVLAND, V3= AVTOT. The reason why it has such a high r1,r2,r4,r7,and r8 is that it have a huge FULLVAL(211000000), AVLAND(27000000), and AVTOT(94950000) and relatively small LTFRONT(50), LTDEPTH(200), BLDFRONT(50), and BLDDEPTH(200).

With further investigation, I found that the corresponding class zip5 (r1\_zip5, r2\_zip5, r4\_zip5, r7\_zip5,r8\_zip5) are also much higher than others in the same zip5. Therefore, based on the investigation, record 402297 is suspicious and requires further investigation.