**WC Accident F. SRC OFF STREET**

This field describes the address number and street name where the crash happened. This field is only populated when SRC\_address type =A, indicating a building address was used. The data comes directly from the MV104 report inside the box entitled “Road on which Accident occurred”, behind box 28.

When src\_address\_type = A, the field, SCR\_ Street name, and SCR\_Street number are populated 98% of the time.

When src\_address\_type=I, use src\_on\_street and src\_cross\_street.

Src\_address\_type=A accounts for 23% of all crashes.

For historic data (prior to 2017), use OFF\_STREET field [DataWarehouse].[dbo].[AIS\_PD\_Locx\_F] in TAMS. Additionally, prior to 2016 in TAMS, Src\_off Street was used to denote crashes on private property/non-roadways.

**WC Accident Vehicle F. Veh Num**

This is the identification number for each vehicle or bicycle involved in a crash. This field is used to connect that data in the victim table to the associated vehicle in the vehicle table. It is also related to the vehicle sequence number in box 8 on the MV-104. Between 9/5/16 and 4/30/19 it can be sorted to define the vehicle sequence number (which will be found on the MV-104). From 5/1/19 onwards, vehicle\_num should be sorted on the ROWID column to find the sequence number.

Vehicle Number uses a randomly generated number since September 5, 2016 instead of assigned vehicle 1 and vehicle 2.

The field is populated 100% of the time. , and the data type is varchar(50).

For historic data (prior to 2017), use VEHICLE\_NUMBER field [DataWarehouse].[dbo].[AIS\_PD\_Vehicle\_F] in TAMS. In TAMS this field stores the sequence number, which is used as the vehicle ID.

**WC\_ACCIDENT\_VICTIM\_F.CONTRIBUTING\_FACTOR\_1**

This field describes the first contributing factor assigned, and is limited to pedestrians. For MV drivers and bicyclists, information is found in CONTRIBUTING\_FACTOR in vehicle table in the vehicle table. This field comes from report box 19 or 21 (depending on which person is being referred to) from the Mv104.

This field is populated 99% of the time when PERSON ROLE CODE is pedestrian, other, and in line skate in crashes.

When PERSON ROLE CODE in other or In-Line skater, it accounts for less than 2% of crashes.

It should be used where person\_role\_code is pedestrian, other, or in-line skater. The field is largely empty, with 69% of values labeled as “unspecified “and, there is null less than 1%.

Victim table CF1 is more reliable than core table CF1

For all non-null values, Victim table CF1 follows the pattern:

1. Pedestrian/Bicyclist/Other Pedestrian Error/Confusion
2. Driver Inattention/Distraction
3. Failure to Yield Right-of-Way
4. Alcohol Involvement
5. Traffic Control Disregarded
6. Other Vehicular
7. View Obstructed/Limited
8. Cell Phone (hand-Held)

For historic data (prior to 2017), use APRNT\_SEQ\_NUM =1 and APRNT\_FCTR in PG\_NYSDOT\_APPFACTOR table.

**WC\_ACCIDENT\_VICTIM\_F.CONTRIBUTING\_FACTOR\_2**

This field describes the second contributing factor assigned, and is limited to pedestrians. For MV drivers and bicyclists, information is found in CONTRIBUTING\_FACTOR in vehicle table in the vehicle table. This field comes from report box 20 or 22 (depending on which person is being referred to) from the Mv104.

This field is populated 99% of the time when PERSON ROLE CODE is pedestrian, other, and in line skate in crashes.

When PERSON ROLE CODE in other or In-Line skater, it accounts for less than 2% of crashes.

It should be used where person\_role\_code is pedestrian, other, or in-line skater. The field is largely empty, with 89% of

values labeled as “unspecified”, and there is null less than 1 %. Victim table CF2 is more reliable than core table CF2.

There are no cases where CF2 is not null and CF1 is null.

For all non-null values, The Victim table CF2 follows the pattern:

1. Pedestrian/Bicyclist/Other Pedestrian Error/Confusion
2. Failure to Yield Right-of-Way
3. Traffic Control Disregarded
4. Driver Inattention/Distraction
5. Other Vehicular
6. Alcohol Involvement
7. View Obstructed/Limited
8. Cell Phone (hand-Held)

For historic data (prior to 2017), use APRNT\_SEQ\_NUM =2 and APRINT FCTR in PG\_NYSDOT\_APPFACTOR table.

**WC Accident Victim F. PED\_NONPED**

Field describes the modes of the crashes victims as Pedestrian, Bicyclist, and Occupant. It is created through the formula based on the person role code and vehicle category. It should be used associated with Person Role Code field. The field is populated 100 % of the time. It includes injured and non-injured under category N.

For historic data (prior to 2017), use TAMS in [DataWarehouse].[dbo].[AIS\_PD\_Victim].PED\_NONPED as P,B,N.

**WC\_ACCIDENT\_F.NODEID**

This field comes from a reverse geocoding process created by Seth, using the x/y coordinates of crash data and the x/y coordinates of the nodes from the IntersectionStreetNames view. It does not come from MV-104. All nodeids are in the valid range of values. The reverse geocoding process is as follows:

* Take X/Y from crash data
* Get the distance between the crash xy and the nodeid xy
  + Remove all “non street” segments
  + Nodeids come from v\_IntersectionStreetNames\_Gen view
  + All crashes without xy are thrown out
* Do this for segment and intersection crashes
* Combine in to one table, join to crashes
* For midblock crashes, if off street has the following strings, they are not included:
* '%%P\L %%'
* '%%PARKING L%%'
* 'I/S/O %%'
* 'INSIDE %%'
* '%%PKNG L%%'
* '%%PKG L%%'
* '%%GARAGE %%'
* '%%DRIVEWAY %%'

This field is in TAMS in the following tables:

* AIS\_PD\_Core\_F
* AIS\_PD\_CrashAllInv\_f
* AIS\_PD\_Geox\_f
* AIS\_PD\_Pedestrian\_F
* AIS\_PD\_Vehicle\_F
* AIS\_PD\_Victim\_F

**WC\_ACCIDENT\_F.LION\_NODE\_NUMBER**

This field does not come from the MV-104; the NYPD geocoder uses the street names and address as inputs to generate lion\_node\_number. It is important to note that this nodeid is generated from the street network at the time of the crash. All nodeids are in the acceptable range of valid nodeids. This field is null 59% of the time.

When src\_address\_type is A or B, lion\_node\_number is always blank. When address type is H or null, lion\_node\_number is null more than 95% of the time. When address type is I, lion\_node\_number is null only 36.4% of the time.

|  |  |
| --- | --- |
| Src\_address\_type | % null |
| A | 100% |
| B | 100% |
| H | >99% |
| NULL | 95% |
| I | 36% |

This field should only be used when nodeid is null.

This data is in TAMS in the following tables

AIS\_PD\_Core\_F: nodeid

* AIS\_PD\_CrashAllInv\_f
* AIS\_PD\_Geox\_f
* AIS\_PD\_Pedestrian\_F
* AIS\_PD\_Vehicle\_F
* AIS\_PD\_Victim\_F

**WC\_ACCIDENT\_F.COMMUNITY\_DISTRICT**

This field contains a list of numeric values that delineate community districts throughout NYC. This field does not come from the MV-104 and is not in TAMS. It comes from the DoITT’s geocoder; 37% of the values in community\_district are null.

In order to get the community district, it

However…..

* X/Y Coord are not always null when community\_district is null
  + 87.4% of values where community\_district is null also have X/Y
    - These points are distributed all across the city
  + 61.2% of records are at address\_type “I” and 32.3% are at address\_tpye “H” in 2018, 63.1% for “I” and 29.4% for “H” in 2017.

Any other critical information users need to know about the field

The X/Y values are still valid across the city for this case – a few look to be mapped to the police precinct instead of the community district.

**WC\_ACCIDENT\_F.Borough**

This field comes from DoITT’s geocoder. Borough is null 37% of the time; whenever the geocoder fails, the borough field is null. Whenever there are x/y coordinates, the borough field is only filled out correctly between 51.8 and 71.5% of the time; the remaining records were overwhelmingly null with a few dozen incorrect borough names.

To find the correct borough, use the src\_police\_precinct field first, and then use the borough field as a backup. The following precincts correspond to the correct boroughs:

[insert table here]

|  |  |  |  |
| --- | --- | --- | --- |
| Borough | Percent with Correct Borough Name | Percent with incorrect borough name | Percent null |
| Brooklyn | 71.5% |  |  |
| Bronx | 63.3% |  |  |
| Manhattan | 63.4% |  |  |
| Queens | 63.27% |  |  |
| Staten Island | 51.8% |  |  |

**WC\_ACCIDENT\_VEHICLE\_F.IS\_TLC\_INV**

This field does not come from the MV-104. It is derived from the license plate, which has a specific set of patterns for TLC vehicles. The patterns are listed below:

* Number/letter/number/number/letter
* Starts with T or Y and ends with C
* Starts with O, ends with L
* Ends with LV, LB, LA, or BB (www.dollarvan.nyc)
* Number/letter/number/number <http://www.nyc.gov/html/tlc_medallion_info/html/tlc_lookup.shtml>
* Letter/letter/number/number/number
* Letter/letter/letter/number/number/number

When comparing the field to the patterns above, 2.1% of the “yes” fields were incorrect, while 9.6% of the “no” fields were incorrect.

There is a similar field called is\_tlc\_dl\_inv.

**WC\_ACCIDENT\_VEHICLE\_F.IS\_TLC\_DL\_INV**

This field does not come from the MV-104. It is derived from the license plate, which has a specific set of patterns for TLC vehicles. The patterns are listed below:

* Number/letter/number/number/letter
* Starts with T or Y and ends with C
* Starts with O, ends with L
* Ends with LV, LB, LA, or BB (www.dollarvan.nyc)
* Number/letter/number/number <http://www.nyc.gov/html/tlc_medallion_info/html/tlc_lookup.shtml>
* Letter/letter/number/number/number
* Letter/letter/letter/number/number/number

When comparing the field to the patterns above, 16.6% of the “yes” fields were incorrect, while 8.5% of the “no” fields were incorrect.

The mismatch between is\_tlc\_inv and is\_tlc\_dl\_inv is as follows:

* There are 48,639 values between the two columns that don’t match
* Out of this subset….
  + 73.6% have is\_tlc\_dl\_inv as “Y” when is\_tlc\_inv is “N”
  + 26.4% have is\_tlc\_dl\_inv as “N” when is\_tlc\_inv is “Y”

When comparing the two column values and manually verifying the plates using the process above, we get the following results. This shows that when the two columns don’t match, is\_tlc\_inv is correct 46.2% of the time and is\_tlc\_dl\_inv is correct 53.9%of the time.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IS\_TLC\_DL\_INV | IS\_TLC\_INV | verified | N | Percent Total | Conclusion |
| Y | N | other | 9883 | 20.32 | Is\_tlc\_inv is correct |
| N | Y | other | 267 | 0.55 | Is\_tlc\_dl\_inv is correct |
| Y | N | yes | 25926 | 53.3 | Is\_tlc\_dl\_inv is correct |
| N | Y | yes | 12563 | 25.83 | Is\_tlc\_inv correct |

**WC\_ACCIDENT\_F.TRAFFIC\_CONTROL**

This field comes from box #3 on the left side of the MV-104. This field is in TAMS in AIS\_PD\_Core\_f under the same name. All values are valid, however, “yield” is consistently spelled incorrectly as “yeild”.

|  |  |
| --- | --- |
| TRAFFIC\_CONTROL | Percent Total |
| NULL | 0.01% |
| Construction Work Area | 0.24% |
| Flashing Light | 0.08% |
| Maintenance Work Area | 0.02% |
| No Passing Zone | 0.01% |
| None | 61.15% |
| Officer/Guard | 0.08% |
| Other\* | 1.32% |
| Police/Fire Emergency | 0.07% |
| RR Crossing Flashing Light | 0.00% |
| RR Crossing Gates | 0.01% |
| RR Crossing Sign | 0.01% |
| School Zone | 0.03% |
| Stop Sign | 6.88% |
| Stopped School Bus - Red Light Flashing | 0.03% |
| Traffic Signal | 29.84% |
| Utility Work Area | 0.01% |
| Yeild Sign | 0.22% |

When investigating the “none” category, there are 22% of these crashes have a nodeid assigned to them. (How should we investigate crashes that have a nodeid but no traffic control?)

**WC\_ACCIDENT\_VEHICLE\_F.VEHICLE\_TYPE\_CODE**

This field comes from the 10TH row of the MV-104 report- vehicle type on each side of the report.

The field includes 7 % null. 87% of the null range comes whenever left scene column has ‘Y’ values.

Vehicle\_type\_code in the vehicle table is more accurate than in the core table. Less than 1 % of the values are misspelled.

In the core table, this information is described horizontally, using the fields Veh\_type\_code1-5,

For all non-null values, the field follows the pattern:

|  |  |
| --- | --- |
| Sedan | 44% |
| Station Wagon/Sport Utility Vehicle | 32% |
| Taxi | 4% |
| Pick-up Truck | 3% |
| Box Truck | 2% |
| Bus | 2% |
| Bike | 1% |

For historic data (prior to 2017), use [DataWarehouse].[dbo].[AIS\_PD\_Vehicle\_F]. VEHICLE\_TYPE\_CODE .

\*\*updated with new numbers on 9/17/19 for above table

This field is currently in progress- we need to do a deeper dive based on Bryant’s questions and the possible new vehicle codes.

**WC\_ACCIDENT\_VEHICLE\_F.VEHICLE\_TYPE\_CODE\_ADDL**

This field does not come from the MV-104. This field is related to vehicle\_type\_code, which gives more specific vehicle types such as ambulance, pick-up truck, convertible, etc. This field aggregates vehicle\_type\_code in to broader categories. As of September 2019, this field is only valid for 2019 onwards.

|  |  |  |
| --- | --- | --- |
| VEHICLE\_TYPE\_CODE\_ADDL | Count | Percent Total |
| NULL | 162,096 | 53.3% |
| ATV | 26 | 0.01% |
| BIKE | 2,713 | 0.9% |
| CAR/SUV | 126,342 | 41.6% |
| MCL | 1,383 | 0.5% |
| TRUCK/BUS | 9,958 | 3.3% |
| V | 1 | 0% |
| ZZZ | 1,514 | 0.5% |

However, vehicle\_type\_code and vehicle\_type\_code\_addl do not always align; until may 2019, about 20% of the values are “null”. After May, the percent of null values goes down significantly to <1%, while “ZZZ” goes up slightly.

|  |  |  |  |
| --- | --- | --- | --- |
| Month | Filled Out | NULL | ZZZ |
| 1 | 0.14 | 99.857 | 0.003 |
| 2 | 0.112 | 99.888 |  |
| 3 | 0.268 | 99.732 |  |
| 4 | 0.634 | 99.36 | 0.006 |
| 5 | 35.123 | 64.446 | 0.432 |
| 6 | 98.149 | 0.75 | 1.102 |
| 7 | 97.986 | 0.883 | 1.131 |
| 8 | 97.718 | 1.289 | 0.993 |
| 9 | 97.963 | 1.241 | 0.796 |

\*\*\*come back to this as we receive more data

**WC\_ACCIDENT\_VEHICLE\_F.VIOLATION**

The field includes the violation associated to the vehicle involved in the crash, and the field comes from the 12nd row of the MV-104 report. The field is populated 3 % of times.

The Vehicle table violation follows the pattern:

|  |  |
| --- | --- |
| VIOLATION | Meaning |
| 509-1 | Operating without license |
| 319-1 | Operating without financial security |
| 19-190B | Fail to yield to a pedestrian or a person riding a bike when that person has the right of way |
| VTL 511 | Aggravated Unlicensed Operation |
| VTL 1192 | Operating a motor vehicle while under the influence of alcohol or drugs. |

There is no consistent in typing the violations, some of them includes “VTL” other only numbers.

The field is populated in the same time as the field Ticket Numbers. Violation are issued to the bicyclists follow the batten:

|  |  |  |  |
| --- | --- | --- | --- |
| # | VIOLATION | meaning | % total tickets to bicyclist |
| 1 | 19-176.2(B) | Bicycle Operation On Sidewalks Prohibited | 12.39 |
| 2 | 19-190 | Fail to yield to a pedestrian when that person has the right of way | 9.80 |
| 3 | 1236 | Lamps and other equipment on bicycles | 6.63 |
| 4 | 1238 | Passengers on bicycles under one year of age prohibited | 4.03 |
| 5 | 1111d | disobey steady red light | 2.59 |

There is no historic data prior to 2017 in TAMS.

**WC\_ACCIDENT\_VEHICLE\_F.VIOLATION\_COUNT**

The field describes how many violations are issued by NYPD to the vehicles or the bicycles involved in the crash. The field is derived by DOIT based on violation text field, as well as this doesn’t work 100% because the way NYPD is writing the violations (ex, 19/190 instead of 19-190, etc.)

The Vehicle table violation \_Count field follows the pattern:

|  |  |
| --- | --- |
| VIOLATION\_COUNT | % |
| NULL | 0.001 |
| zero violation | 97.185 |
| one violation | 2.455 |
| More than one violation | 0.358 |

For historic data (prior to 2017), use [DataWarehouse].[dbo].[AIS\_PD\_Vehicle\_F]. [VIOLATION\_COUNT]

**WC\_ACCIDENT\_VEHICLE\_F.** **TICKET\_NUMBER**

The field includes the identification ticket number issued by NYPD to the vehicles involved in the crash, and the field comes from the 11st row of the MV-104 report. The field is populated 3 % of times. The Vehicle table Ticket number follows the pattern:

|  |  |  |
| --- | --- | --- |
| # | VEHICLE\_TYPE\_CODE | % Total ticket |
| 1 | Sedan | 48.65 |
| 2 | Station Wagon/Sport Utility Vehicle | 35.38 |
| 3 | Taxi | 3.77 |
| 4 | Pick-up Truck | 3.44 |
| 5 | Box Truck | 1.53 |
| 6 | Bike | 1.32 |
| 7 | Motorcycle | 1.09 |
| 8 | Van | 0.75 |
| 9 | Tractor Truck Diesel | 0.66 |
| 10 | Bus | 0.53 |

There is no historic prior to 2017 in TAMS.

**WC\_ACCIDENT\_F.ADDRESS\_TYPE**

This field does not come from the MV-104 and is not present in TAMS, but box 1 on the MV-104 states whether the victim was at an intersection or not. The distribution of values is:

|  |  |
| --- | --- |
| address\_type | pct\_total |
| NULL | 0.01 |
| A | 23.41 |
| B | 1.51 |
| H | 10.68 |
| I | 64.38 |

Address type A has only src\_off\_street filled >99% of the time, and address types B and H have src\_on\_street filled >98% of the time. For Address type I, 24% have src\_on\_street only, and 76% have src\_on\_street and src\_cross\_street filled out.

Below is how often either nodeid or lion\_node\_number is filled out.

|  |  |
| --- | --- |
| address\_type | Nodeid filled out? |
| NULL | 4.7% |
| A | 93.6% |
| B | 0% |
| H | .2% |
| I | 94.9% |

\*\*\*can you technically get this information from TAMS by checking if the crash has a nodeid or not?

**WC\_ACCIDENT\_F.STATUS\_MESSAGE**

This field does not come from the MV-104 and is not present in TAMS. This field shows whether the DoITT geocoder worked or failed.

|  |  |
| --- | --- |
| status\_message | pct\_total |
| NULL | 12.1% |
| Geocoded successfully. | 63.2% |
| Geocoder did not recognize the input location. | 24.7% |

When the message is “geocoder did not recognize the input location”, 64% of addresses have only src\_on\_street present, while 31% have src\_on\_street and src\_cross\_street present.

When the message is “geocoded successfully”, 64.8% of the addresses have src\_on\_street and src\_cross\_street present present, while 35.1% have only src\_off\_street present.

When the message is null, 98.5% of addresses have only src\_on\_street filled out; the remainder had no address fields filled out.

It is unclear what the difference is between NULL and “Geocoder did not recognize the input location”. However, ‘NULL’ is the most common message when none of the src on/off/cross street fields are filled out (71.1% of cases).

**WC\_ACCIDENT\_VEHICLE\_F.** **VEHICLE\_INVOLVED\_CNT**

The field includes the count of the vehicle involved in the crash, and the field comes from the 1st row of the MV-104 report. The field is populated 100 % of times.

When NONMV =1, the field is populated with zero.

96% of the field values match with the count of the vehicle identification in the vehicle table.

When vehicle type code = bike, the field value will not count

The field values follows the pattern

|  |  |
| --- | --- |
| VEHICLE\_INVOLVED\_CNT | % total |
| 0 | Less than 1 % |
| 1 | 11% |
| 2 | 82% |
| 3 | 5% |
| 4 | 1% |
| 5-21 | Less than 1 % |

For historic data (prior to 2017), use [DataWarehouse].[dbo].[AIS\_PD\_Core\_F]. [VEHICLE\_COUNT]

**WC\_ACCIDENT\_VEHICLE\_F.IS\_DSNY\_INV**

This field does not come from the MV-104. You could derive it from the MV-104 given that the vehicle registration information is on the form. The field contains only three values; see distribution below

|  |  |
| --- | --- |
| is\_dsny\_inv | Percent Total |
| N | 99.98% |
| Y | 0.02% |

For 2017-2018, there are 156 records in the vehicle table where is\_dsny\_inv=’Y’. When comparing this field to related fields, I checked “organization\_name” and “agency\_name” in the victim table. Only 7 records did not have a dsny-related label in either of these columns. I used the following strings (upper and lower case)

* Dsny
* Sanitation
* Sanatation (common misspelling in the data)
* Dos

However, when looking at where is\_dsny\_inv\_=’N’, there are 2,262/452,889 (.5%) records that have sanitation in the “organization name” and/or “agency name” columns,

In order to accurately determine whether a vehicle is a dsny vehicle, the best course of action is to ignore this field and see whether the above strings are in the “agency\_name” or “organization\_name” fields.

**WC Accident Victim F. Victim Age**

This field describes the ages of all involved in the crash. The victim age describes whatever the involved injured, killed, or not injured.

The data comes directly from the MV104 report inside the box 12 in the bottom of the report.

1 .5 % of the records in this field are mistyped or blank, so it is recommended to use age between 1 and 100.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Victim Age Range | How Often | % | Matched | Unmatched | % matched |
| Age Blank | 18 | 0.01% |  |  |  |
| Below Zero | 30 | 0.02% |  |  |  |
| Above 100 | 108 | 0.09% |  |  |  |
| Zero | 1694 | 1.39% |  |  |  |
| between 1 and 17 | 11882 | 9.75% | 10607 | 1275 | 89% |
| between 18 and 64 | 98973 | 81.22% | 96088 | 2885 | 97% |
| between 65 and 100 | 9154 | 7.51% | 8802 | 352 | 96% |
| between 1 and 100 | 120009 | 98.48% | 115497 | 4512 | 96% |
| Total | 121859 |  |  |  |  |

When the field is between 1 and 100 the victim age and the difference between accident date and victim’s date of birth matches (96%) of injured the

For historical data before 2017 use

[DataWarehouse].[dbo].[AIS\_PD\_Victim\_F]. [VICTIM\_AGE]

**WC Accident Victim F. Date of Birth**

This field describes the DOB of the parties involved in the crash. The DOB describes whatever the involved parties injure, killed, or not injured. The data comes from the MV 104 report inside the 6th row under each involved, and under Person killed or injured in accident. For injured or killed 20 % blanked DOB occurred when victim age is entered zero.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Victim Age Range | How Often | How often DOB blank | How often DOB entered |  |  |  |
| Age Blank | 18 | 18 | 0 |  |  |  |
| Below Zero | 30 | 0 | 30 |  |  |  |
| Above 100 | 108 | 33 | 75 | Matched | Unmatched | |
| zero | 1694 | 1003 | 691 |  | blank DOB | human |
| between 1 and 17 | 11882 | 1216 | 10666 | 10607 | 1216 | 59 |
| between 18 and 64 | 98973 | 2452 | 96521 | 96088 | 2452 | 433 |
| between 65 and 100 | 9154 | 303 | 8851 | 8802 | 303 | 49 |
| between 1 and 100 | 120009 | 3971 | 116038 | 115497 | 3971 | 541 |
| Total | 121859 | 8996 | 116834 |  |  |  |

For historical data before 2017 use

[DataWarehouse].[dbo].[AIS\_PD\_Vehicle].[DRIVER\_DOB]

**WC Accident Victim F. Date of Birth**