

# **Chicago and Dallas Database Observations**

## **A. Chicago Dataset Overview**

The Chicago dataset under consideration comprises various data points that are critical for data analysis and interpretation. Here are the key statistics and attributes of the dataset:

1. Number of Variables: The dataset contains 17 distinct variables. These variables are the different attributes or fields in the dataset.
2. Number of Observations: There are a total of 267,613 observations, which likely correspond to individual records or entries in the dataset.
3. Missing Cells: A count of 84,101 cells within the dataset have missing data, which constitutes approximately 1.8% of the total data. This is relatively low and indicates a fairly complete dataset.
4. Duplicate Rows: There are no duplicate rows, as indicated by a count of 0, which means that each observation is unique.
5. Total Size in Memory: The dataset occupies 34.7 MiB (Megabytes) in memory, providing an idea of the data's footprint in terms of storage.
6. Average Record Size in Memory: Each record, on average, occupies 136.0 bytes in memory, which can be useful for understanding the load and performance implications for data processing tasks.

## **Following are the variables in the Chicago Dataset with their observations:**

### 1. Inspection ID-

The dataset contains 267,613 unique Inspection IDs

There are no missing values for the Inspection ID variable, as indicated by a count of 0.

### 2. DBA Name-

There are 32172 distinct entries for DBA name

There are zero missing values for DBA Name

### 3. AKA Name-

AKA Name has 30260 distinct records and has 2473 records with missing values

### 4. License # -

The dataset contains 44516 unique License numbers

There are 18 missing values for the License number variable

### 5. Facility Type-

There are 513 distinct entries for Facility type

There are 5117 missing values for Facility type

6. Risk-

Risk variable has 4 distinct categories: Risk 1(High) , Risk 2 (Medium) , Risk 3 (Low)  
With 4 missing values

7. Address-

Address variable consists of 19644 distinct entries  
It contains 0 missing values

8. City -

City variable has 79 distinct records and consists of 161 missing values

9. State-

The data is for Chicago Dataset and there are other 5 other states which are outliers in the data

10. Zip-

There are 121 unique zip code entries out of which there are 49 missing values

11. Inspection Date-

The dataset consists of 3564 unique entries for Inspection date variable  
It also has 0 missing values for the Inspection date variable

12. Inspection Type-

Inspection type consists of 110 distinct data entries  
While there is only 1 missing value

13. Results -

There are 7 distinct values for Results  
With zero missing values for any of the records

14. Violations -

Chicago dataset comprises of 19674 distinct violations  
Out of which 73370 are missing values for the Violations variable

15. Latitude -

Latitude has 18214 distinct values and 924 records have missing values as seen

16. Longitude -

Longitude has 18214 distinct values and 924 records have missing values same as Latitude

17. Location -

Location variable has 18214 unique entries and 924 records have missing values

## **B. Dallas Dataset Overview:**

The Dallas dataset under consideration comprises various data points that are critical for data analysis and interpretation. Here are the key statistics and attributes of the dataset:

1. Number of variables: The dataset contains 114 different variables.
2. Number of observations: There are 78,400 entries or observations in the dataset.
3. Missing cells: A significant number of cells, 64,543,57 in total, are missing, which is 72.2% of the data. This indicates a high level of incompleteness in the dataset.
4. Duplicate rows: There are 42 duplicate rows, which is about 0.1% of the data.
5. Total size in memory: The dataset occupies 68.2 MiB of memory.
6. Average record size in memory: Each record, on average, takes up 912.0 B in memory.

## **Following are the variables in the Dallas Dataset with their observations:**

1. Restaurant Name - Restaurant name variable has 9136 distinct records  
It also contains 11 records with missing values
2. Inspection Type-  
Inspection type is categorical and 3 distinct values - Routine, Follow up, Complaint  
With zero missing values in their records
3. Inspection Date - Inspection date has 2293 unique records with zero missing values
4. Inspection Score - Inspection score variable has 58 distinct entries  
It has zero missing values in its entries
5. Street Number - Street number has 3342 distinct records and zero missing values
6. Street Name - Street name variable has 842 distinct records and zero missing values
7. Street Direction - Street direction is categorical with North, South, East, West  
Its records comprise of 52514 missing values
8. Street Type - Street type is also a categorical variable with 19 distinct entries and 1661 missing values
9. Street Unit - It has 989 distinct entries  
Consists of 50458 missing values
10. Street Address - Street address variable has 7814 distinct records and 0 missing values
11. Zip Code - Variable zip code consists of 156 distinct records and 0 missing values on these records
12. Violation Description - 1 - It has 623 distinct records with 6540 missing values

13. Violation Points - 1 - This variable is categorical with 4 distinct values ( 3.0, 1.0, 2.0 and 4.0) and 6540 missing values

14. Violation Detail - 1 - This variable has 642 unique records and consists of 7091 missing values

15. Violation Memo - 1 - This particular variable has 43930 unique records  
It has 22783 of missing values reported

16. Violation Description - 2 - Violation description has 610 unique entries in the dataset and 14050 of missing values in the record

17. Violation Points - 2- This variable is categorical with 4 distinct values ( 3.0, 1.0, 2.0 and 4.0) and 14050 missing values

18. Violation Detail - 2 - Violation detail has 625 unique entries and 14634 missing values in the dataset

19. Violation Memo - 2 - Violation memo has 39019 unique records and 29118 missing values

20. Violation Description - 3 - Violation description 3 has 610 unique records and 22929 missing values

21. Violation Points - 3 - This variable is categorical with 4 distinct values ( 3.0, 1.0, 2.0 and 4.0) with 22927 missing values

22. Violation Detail - 3- This record has 625 unique records and 22348 missing values

23. Violation Memo - 3 - Violation memo has 28922 unique records and 42334 missing values

24. Violation Description - 4 - Violation description 3 has 562 unique records and 39897 missing values

25. Violation Points - 4 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0 ) with 31744 missing values

26. Violation Detail - 4 - This record has 596 unique records and 32230 missing values

27. Violation Memo - 4 - Violation memo has 28922 unique records and 42334 missing values

28. Violation Description - 5 - Violation description 5 has 562 unique records and 39897 missing values

29. Violation Points - 5 - This variable is categorical with 4 distinct values ( 3.0, 1.0, 2.0 and 4.0) and 39897 missing values
30. Violation Detail - 5 - This record has 567 unique records and 40370 missing values
31. Violation Memo - 5 - Violation memo has 24295 unique records and 48408 missing values
32. Violation Description - 6 - Violation description 6 has 551 unique records and 47251 missing values
33. Violation Points - 6 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 47251 missing values
34. Violation Detail - 6- This record has 556 unique records and 47643 missing values
35. Violation Memo - 6- Violation memo has 19871 unique records and 53926 missing values
36. Violation Description - 7- Violation description 7 has 496 unique records and 53767 missing values
37. Violation Points - 7 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 53767 missing values
38. Violation Detail - 7 - This record has 500 unique records and 54093 missing values
39. Violation Memo - 7 - Violation memo has 15997 unique records and 58770 missing values
40. Violation Description - 8 - Violation description 8 has 580 unique records and 59391 missing values
41. Violation Points - 8 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 59391 missing values
42. Violation Detail - 8 - This record has 485 unique records and 59652 missing values
43. Violation Memo - 8 - Violation memo has 12435 unique records and 63100 missing values
44. Violation Description - 9 - Violation description 9 has 420 unique records and 64188 missing values
45. Violation Points - 9 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0, 4.0) and 64188 missing values
46. Violation Detail - 9 - This record has 427 unique records and 64403 missing values

47. Violation Memo - 9 - Violation memo has 9514 unique records and 66780 missing values
48. Violation Description - 10 - Violation description 10 has 410 unique records and 68233 missing values
49. Violation Points - 10 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0, 4.0) and 68233 missing values
50. Violation Detail - 10 - This record has 416 unique records and 68392 missing values
51. Violation Memo - 10 - Violation memo has 6999 unique records and 69891 missing values
52. Violation Description - 11- Violation description 11 has 377 unique records and 71595 missing values
53. Violation Points - 11- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0, 4.0) and 71595 missing values
54. Violation Detail - 11- This record has 383 unique records and 71694 missing values
55. Violation Memo - 11- Violation memo has 4777 unique records and 72698 missing values
56. Violation Description - 12- Violation description 12 has 328 unique records and 74182 missing values
57. Violation Points - 12- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0, 4.0) and 74182 missing values
58. Violation Detail - 12-This record has 332 unique records and 74263 missing values
59. Violation Memo - 12 -Violation memo has 3006 unique records and 74933 missing values
60. Violation Description - 13- Violation description 13 has 288 unique records and 75722 missing values
61. Violation Points - 13- This variable is categorical with 4 distinct values ( 3.0, 1.0, 2.0, 4.0) and 75722 missing values
62. Violation Detail - 13-This record has 291 unique records and 75777 missing values
63. Violation Memo - 13-Violation memo has 2007 unique records and 76189 missing values

64. Violation Description - 14 - Violation description 14 has 274 unique records and 76510 missing values

65. Violation Points - 14 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 76510 missing values

64. Violation Detail - 14 - This record has 276 unique records and 76540 missing values

65. Violation Memo - 14 - Violation memo has 1474 unique records and 76841 missing values

66. Violation Description - 15 - Violation description 15 has 237 unique records and 77068 missing values

67. Violation Points - 15 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 77068 missing values

68. Violation Detail - 15 - This record has 237 unique records and 77092 missing values

69. Violation Memo - 15 - Violation memo has 990 unique records and 77314 missing values

70. Violation Description - 16 - Violation description 16 has 190 unique records and 77510 missing values

71. Violation Points - 16- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 77510 missing values

72. Violation Detail - 16 - This record has 193 unique records and 77527 missing values

73. Violation Memo - 16 - Violation memo has 685 unique records and 77660 missing values

74. Violation Description - 17- Violation description 17 has 175 unique records and 77850 missing values

75. Violation Points - 17 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 77850 missing values

76. Violation Detail - 17 - This record has 176 unique records and 77862 missing values

77. Violation Memo - 17- Violation memo has 424 unique records and 77953 missing values

78. Violation Description - 18 - Violation description 18 has 136 unique records and 78049 missing values

79. Violation Points - 18- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78049 missing values

80. Violation Detail - 18 - This record has 137 unique records and 78057 missing values

81. Violation Memo - 18 -Violation memo has 280 unique records and 78111 missing values

82. Violation Description - 19 - Violation description 19 has 94 unique records and 78193 missing values

83. Violation Points - 19 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78193 missing values

84. Violation Detail - 19 - This record has 93 unique records and 78196 missing values

85. Violation Memo - 19 - Violation memo has 163 unique records and 78232 missing values

86. Violation Description - 20 - Violation description 20 has 69 unique records and 78269 missing values

87. Violation Points - 20- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78269 missing values

88. Violation Detail - 20- This record has 68 unique records and 78275 missing values

89. Violation Memo - 20 - Violation memo has 110 unique records and 78289 missing values

70. Violation Description - 21 - Violation description 21 has 51 unique records and 78327 missing values

71. Violation Points - 21 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78327 missing values

72. Violation Detail - 21- This record has 50 unique records and 78332 missing values

73. Violation Memo - 21- Violation memo has 62 unique records and 78337 missing values

74. Violation Description - 22 - Violation description 22 has 25 unique records and 78368 missing values

75. Violation Points - 22- This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78368 missing values

76. Violation Detail - 22- This record has 24 unique records and 78371 missing values



77. Violation Memo - 22 - Violation memo has 30 unique records and 78370 missing values
78. Violation Description - 23 - Violation description 23 has 25 unique records and 78368 missing values
79. Violation Points - 23 - This variable is categorical with 3 distinct values ( 3.0, 1.0, 2.0) and 78368 missing values
80. Violation Detail - 23 - This record has 24 unique records and 78371 missing values
81. Violation Memo - 23 - Violation memo has 62 unique records and 78337 missing values
82. Violation Description - 24 - Violation description 10 has 25 unique records and 78388 missing values
83. Violation Points - 24 - This variable is categorical with 2 distinct values ( 1.0, 2.0) and 78388 missing values
84. Violation Detail - 24- This record has 10 unique records and 78388 missing values
85. Violation Memo - 24- Violation memo has 12 unique records and 78337 missing values
86. Violation Description - 25 - Violation description 6 has 25 unique records and 78394 missing values
87. Violation Points - 25 - This variable is categorical with 3 distinct values ( 1.0, 2.0, 3.0) and 94 783 missing values
88. Violation Detail - 25 - This record has 5 unique records and 78395 missing values
89. Violation Memo - 25- Violation memo has 6 unique records and 78394 missing values
90. Inspection Month - This variable has 89 distinct entries and has 0 missing values
91. Inspection Year- This variable is categorical and has 8 distinct categories with 0 missing values
92. Lat Long Location - This variable has 19299 distinct records with 0 missing values