Designing Advanced Data Architectures for Business Intelligence

Individual Project

Name: Dhir Thacker

NUID: 002819144

TSV File: ‘311\_CallCenterServiceRequests\_KansasCity\_2007-March2021.tsv’

Total Number of Columns: 23

Total Number of Records: 1563215

* **MISSING VALUES**:

|  |  |  |  |
| --- | --- | --- | --- |
| **Column\_Name** | **Issue** | **Count** | **Equivalent Percentage** |
| CASE ID | - | >10,000 unique | - |
| SOURCE | Missing Value | 67, 21 unique values | 0.004286 |
| DEPARTMENT | - | 27 unique values | - |
| WORK GROUP | - | 146 unique values | - |
| TYPE | - | 295 unique values | - |
| DETAIL | - | 574 unique values | - |
| CREATION DATE | - | 5229 unique values | - |
| CREATION TIME | - | 1440 unique values | - |
| STATUS | - | 6 unique values | - |
| EXCEEDED EST TIMEFRAME | Missing Value | 23, 2 unique values | 0.001471 |
| CLOSED DATE | Missing Value | 12702, 4995 unique | 0.812556 |
| DAYS TO CLOSE | Missing Value | 26515, 2748 unique | 1.696184 |
| STREET ADDRESS | Missing Value | 24, >10,000 unique | 0.001535 |
| ZIP CODE | Missing Value | 826, 65 unique | 0.052840 |
| NEIGHBOURHOOD | Missing Value | 46106, 250 unique | 2.949434 |
| COUNTY | Missing Value | 66959, 13 unique | 4.283416 |
| POLICE DISTRICT | Missing Value | 35265, 6 unique | 2.255928 |
| PARCEL ID NO | - | >10,000 unique | - |
| LATITUDE | - | >10,000 unique | - |
| LONGITUDE | - | >10,000 unique | - |
| CATEGORY1 | - | 82 unique values | - |
| CATEGORY2 | Missing Value | 1001657, 9 unique | 64.076726 |
| CATEGORY3 | Missing Value | 1404943, 3 unique | 89.875225 |

* **DUPLICATES**:

There are no duplicate rows/records in our TSV file.

* **DATE FORMATS**:

Through Alteryx, we converted the “CREATION DATE” and “CLOSED DATE” to a proper datetime format using the standard convention “MM/dd/yyyy”.

* **SPECIAL CHARACTERS:**

No special characters found.

* Two extra columns “Creation\_Date” and Closed\_Date” were inserted through Alteryx using the formula component.
* In addition, an extra column “Record ID” is added using the component through Alteryx which is a unique identifier for each row or record in a dataset.
* Also, two columns “Type” & “Street Address” have shortest and longest value in the “Summary” Tab as a combination of integer and strings.

For instance, column “Type” has the shortest value as “08” and the longest value as

“WaterServices-ConsumerServiceRequests” etc.

* This might mean there is an inconsistency between the type of data stored within a column.
* To clean missing values from our dataset, if the missing values are relatively small and removing them would not directly impact the analysis then they can be deleted using the “Filter” tool. Another alternative could be to replace the missing value with some default value based on business rules.

**ALTERYX DATA PROFILING**

**Alteryx workflow**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

**DATA POPULATED IN SQL SERVER**

A screenshot of a computer

Description automatically generated

**ROW COUNTS**

**A close up of text

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**POWERBI and Tableau visualizations**

**PowerBI**

A graph on a computer screen

Description automatically generated

Tableau

**A graph of different colored lines

Description automatically generated**

1. The line chart depicting service requests from 2018 to 2021 reveals dynamic patterns. In 2019, there was a peak in May, while 2020 experienced a spike in July which gradually declined over the months. However, 2021 showed a distinctive decline from March onwards, hitting a minimal count of 11 requests in October.

PowerBI

A screenshot of a computer

Description automatically generated

Tableau

**A screenshot of a graph

Description automatically generated**

1. The bar graph reveals a clear preference for 'phone' with 1,204,236 service requests, followed by 'web' at 211,721, and a notable drop in requests from other sources.

PowerBI

A graph of a number of people

Description automatically generated with medium confidence

Tableau

**A screenshot of a computer

Description automatically generated**

1. The area chart emphasizes NHS as the primary department with 780,000 service requests, followed by Public Parks at 353,787, showcasing a notable contrast. The count further dwindles to hundreds for departments like Parks and Rec, signalling a substantial decline in service requests across various departments.

A screenshot of a computer

Description automatically generated

1. The top 10 cases are decided on the basis of “Days To Close”.

PowerBI

A screenshot of a map

Description automatically generated

TableauA screenshot of a graph

Description automatically generated

1. The top 10 areas for most number of service requests are Westside South, Westwood, White Oak, Willow Creek, Winnetonka, Winnwood, Winnwood Gardens, Woodbridge, Woodson Estates Wornall Homestead. The most service requests were raised at Westside South with more than 8,000 requests.

PowerBI

A screen shot of a graph

Description automatically generated

Tableau

**A screenshot of a computer

Description automatically generated**

1. The tree map reveals workload distribution, with NHS handling a significant share, followed by Public Works. Conversely, the Northeast department has a minimal workload, managing only 35 requests.

PowerBI

A screenshot of a computer

Description automatically generated

Based on the outlier graph on the right, it can be inferred that the department NCS takes an abnormally high value of 767.13 as the days to close contrary to the other departments which take similar times, significantly less than department NCS.

Tableau

**A screenshot of a computer

Description automatically generated**

1. The bar graph illustrates that the response time is highest for the department NCS followed by a significant drop for other departments. The response time kept declining for each department further on.

PowerBI

A screenshot of a graph

Description automatically generated

Tableau

**A screenshot of a graph

Description automatically generated**

1. The visualization indicates a notable decline in the resolved status from a total of 422,700, particularly in the year 2021. Notably, open cases were effectively managed in 2020, peaking at 5,820, while the total in-progress cases decreased from 37 in 2020 to just 7 in 2021.

PowerBI

A graph with blue squares

Description automatically generated

Tableau

**A screenshot of a computer

Description automatically generated**

1. The bar chart illustrates that the average days to close are most extended for Category1 'Weeds' at 420 days, while 'Vehicles' and 'Water Leak' share the same lower duration of 53 days.

PowerBI

A graph showing a number of people

Description automatically generated with medium confidence

Tableau

**A screenshot of a computer

Description automatically generated**

1. In the clustered bar chart, the NCS department stands out with the highest service requests, exceeding 6000, and an average time to closure of 767 days.

**SQL QUERIES**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**