**DESIGNING ADVANCED DATA ARCHITECTURES FOR BUSINESS INTELLIGENCE**

**INDIVIDUAL PROJECT**

**SCOPE OF WORK:**

1. Creating visualizations using Tableau and Power BI based on the business requirements provided.

2. Perform data profiling, understanding the data, and making keen observations on inconsistencies.

3. Understand the process of performing data staging for ad-hoc analysis using SQL and Visualizations.

**BUSINESS QUESTIONS TO BUILD VISUALIZATION:**

**(1)  Service Requests Over Time:**  
• What is the overall trend in Service Requests over the years 2018-2021?  
• How have Service Requests changed monthly?

**(2) Volume of service requests received from different sources:**

• What is the overall trend in Service Requests over Sources?

**(3) Volume of service requests received by Department:**

• What is the overall trend in Service Requests received by Departments?

**(4) Top 10 Performance Metrics (Response Time) per CATEGORY and Type of Request:**

• What are the top 10 cases whose response time was fastest? Categorize it with Category1 and Type of Request.

**(5) Geographical Visualization:**

• What are the Top 10 areas where most number of request were raised?

**(6) Departmental Workload Comparison:**

• How does the workload vary among different departments and work groups? Create a visual representation to highlight the distribution.

*• Hint:*Generate a stacked bar chart or tree map to illustrate the distribution of service requests among different departments and work groups. This can help identify which departments are handling a larger share of the workload.

**(7) Response Time Analysis:**

*•*Visualize the distribution of response times for each department. Are there any outliers or patterns in response times?

**(8) Service Request Status Composition:**

*•*Create a visualization to show the composition of service request statuses (open, closed, in progress). How has this composition changed over the years 2018-2021?

**(9) Time to Closure Analysis:**

*•*Visualize the average days to close service requests for each category1. Are there categories with consistently longer closure times?

• Show top 10 (If you need help on how to restrict top 10 contact us and we can guide / help you)

**(10) Workload Efficiency:**

*•*Create a visualization to show the relationship between workload (number of service requests) and efficiency (days to close) for each department?

**DATA PROFILING:**

The provided TSV file contained 23 Columns and 1563215 records.

A computer screen shot of a diagram

Description automatically generated

Following are the observations:

* **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 the provided TSV file.

* **DATE FORMATS:**

Found issues in the “CREATION DATE” and “CLOSED DATE” columns. Using Alteryx, converted the format to “MM/dd/yyyy”.

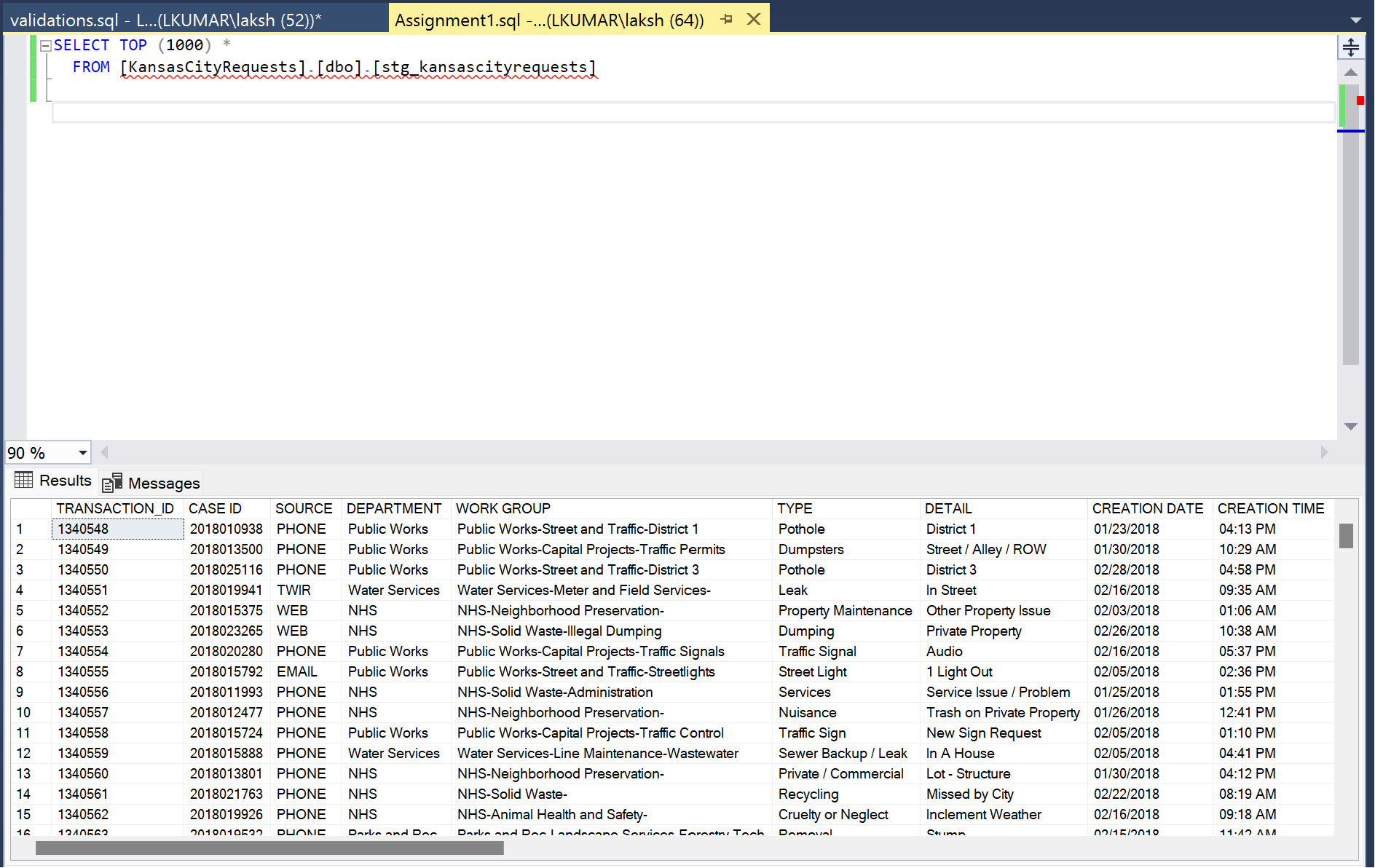
* **SPECIAL CHARACTERS:**

No special characters were found.

* **ADDING NEW COLUMNS**:

Three new columns namely CREATED\_DATE, CREATED\_BY (using formula component) and TRANSACTON\_ID (using generate rows component) were created.

* Two columns “TYPE” and “STREET ADDRESS” have shortest and longest value as a combination of integer and strings.
* 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.
* **DATA POPULATED IN SQL SERVER**



**POWERBI AND TABLEAU VISUALIZATIONS WITH CORRESPONDING SQL VALIDATIONS:**

**(1)  Service Requests Over Time:**  
• What is the overall trend in Service Requests over the years 2018-2021?  
• How have Service Requests changed monthly?

**ANALYSIS:** In 2019, there was a peak in May, while 2020 saw a spike in July that gradually decreased over the subsequent months. In 2021, there was a distinctive decline from March onwards, reaching a minimal count of 11 requests in October.

**SQL VALIDATION:**

**By month:**

**A screenshot of a computer

Description automatically generated**

**By year:**

**A screenshot of a computer program

Description automatically generated**

**2)** **Volume of service requests received from different sources:**

• What is the overall trend in Service Requests over Sources?

**ANALYSIS:** The bar graph highlights a clear preference for 'phone,' with 1,204,236 service requests, followed by 'web' at 211,721. There is a noticeable drop in requests from other sources.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**3)** **Volume of service requests received by Department:**

• What is the overall trend in Service Requests received by Departments?

**ANALYSIS:** 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 goes gradually down to hundreds for departments like Parks and Rec, indicating a substantial decline in service requests across various departments.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**4) Top 10 Performance Metrics (Response Time) per CATEGORY and Type of Request:**

• What are the top 10 cases whose response time was fastest? Categorize it with Category1 and Type of Request.

**ANALYSIS:** The top 10 cases are determined based on the "Days to Close”

**SQL VALIDATION: Unable to perform validation due to the presence of null values.**

**5)** **Geographical Visualization:**

• What are the Top 10 areas where most number of request were raised?

**ANALYSIS:** The top 10 areas with the most service requests are Shoal Creek, Blue Hills, East Community Team South, Scarritt Point, South Indian Mound, Lykins, Tower Homes, CBD Downtown, North Indian Mound. Shoal Creek had the highest number of service requests, with close to 56,000.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**6)** **Departmental Workload Comparison:**

• How does the workload vary among different departments and work groups? Create a visual representation to highlight the distribution.

• Hint: Generate a stacked bar chart or tree map to illustrate the distribution of service requests among different departments and work groups. This can help identify which departments are handling a larger share of the workload.

**ANALYSIS:** The viz depicts workload distribution, with the NHS handling a significant share, followed by Public Works. Conversely, the Northeast department has a minimal workload, managing only 35 requests.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**7) Response Time Analysis:**

• Visualize the distribution of response times for each department. Are there any outliers or patterns in response times?

**ANALYSIS:** The viz illustrates that the response time is highest for the department NCS, followed by a significant drop for other departments. The response time continued to decline for each department.

**SQL VALIDATION: Unable to perform validation due to the presence of null values.**

**8)** **Service Request Status Composition:**

• Create a visualization to show the composition of service request statuses (open, closed, in progress). How has this composition changed over the years 2018-2021?

**ANALYSIS:** 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.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**9) Time to Closure Analysis:**

• Visualize the average days to close service requests for each category1. Are there categories with consistently longer closure times?

• Show top 10

**ANALYSIS:** The visualization shows that the average days to close are most extended for Category1 'Weeds' at 420 days, while it is lowest for ‘Nuisance Violations’ at 93.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**

**10)** **Workload Efficiency:**

• Create a visualization to show the relationship between workload (number of service requests) and efficiency (days to close) for each department?

**ANALYSIS:** In the viz, the NHS department stands out with the highest service requests at 783,094 and an average days to close of 77 days.

**SQL VALIDATION:**

**A screenshot of a computer

Description automatically generated**