

Boston 311 Service Request

GCP Data Warehousing Project

Introduction

The Boston 311 Service Request system provides a centralized platform for handling non-emergency city service requests. This project focuses on preparing raw 311 request data from Boston for advanced analytics by building a robust ETL (Extract, Transform, Load) pipeline. The goal is to process historical and real-time service request data efficiently, enabling seamless reporting and analysis.

The pipeline processes millions of records and leverages **Apache Beam** for distributed data processing, **Google Cloud Dataflow** for execution, and **BigQuery** as the **Data Warehouse**. This ensures structured, high-performance analytics capabilities.

Links

Dataset: [Boston 311 | Boston.gov](#)

GitHub for Previous Implementation on Local: [amey379/311_Request_Analysis](#)

GitHub for GCP: [amey379/GCPDataWarehousingProject](#)

Tools

Storage: BigQuery – Data Warehouse for structured storage Google Cloud Storage (GCS) – Staging and intermediate data storage	Development & System: <i>Apache Avro & Parquet – Data serialization and storage formats</i> <i>Java – Programming language for ETL processing</i> <i>GCP VM (Unix) – Environment setup & scripting</i>
Orchestration & Workflow Management: <i>Google DataFlow – ETL pipelines and Orchestration</i> <i>Apache Beam – ETL pipelines and Orchestration</i>	Visualization & Reporting: <i>Power BI – Dashboarding & data visualization</i>

Skills

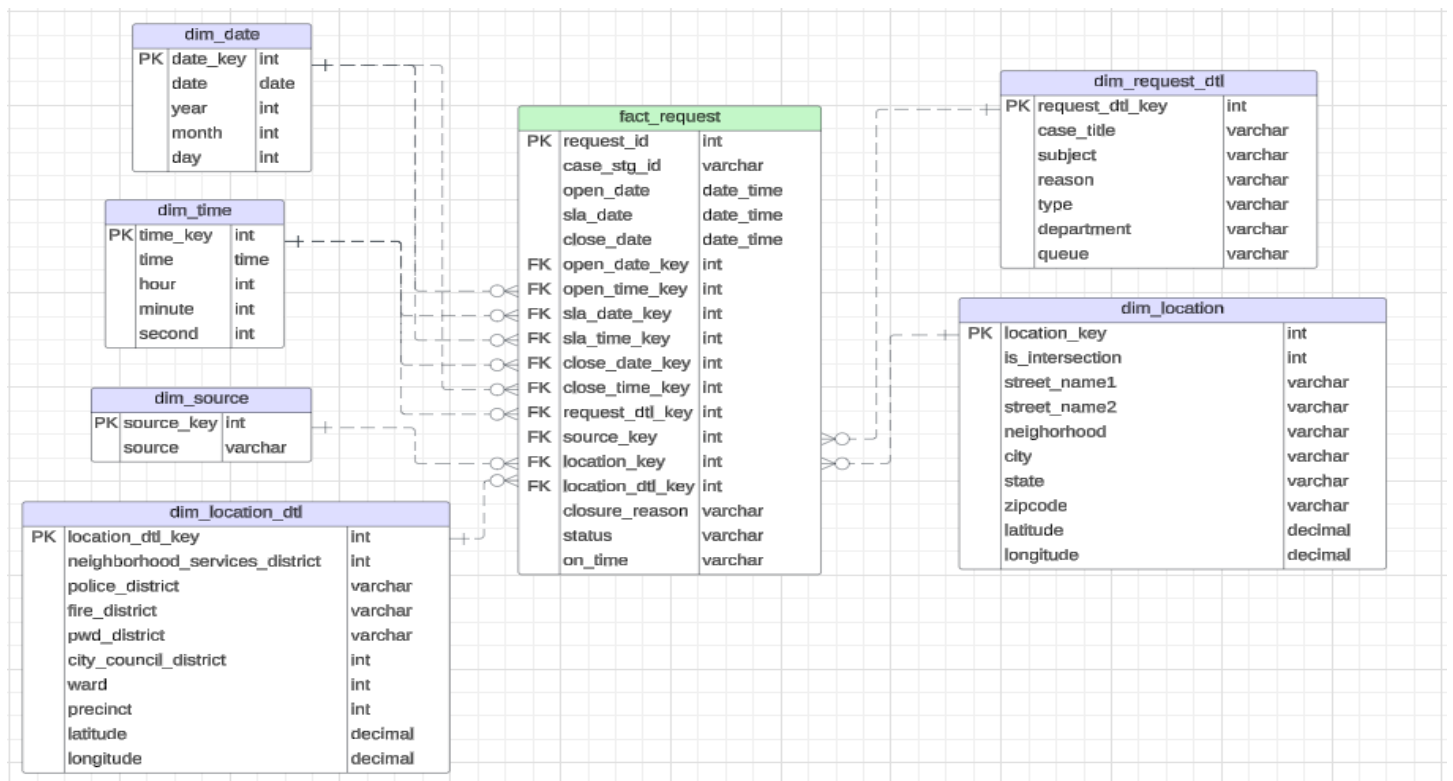
Dimensional Modeling & Data Warehousing

BigQuery Optimization & Performance Tuning

ETL Development (Extract, Transform, Load)

Apache Beam & Google Dataflow

Dimensional Model



Fact Table:

- **Fact_311_Requests** – Stores key metrics like resolution time, case status, request type

Dimension Tables:

- **Dim_Date Dim_Time** – Date and time details for trend analysis
- **Dim_Location** – Geographic information of incidents
- **Dim_Request_Details** – Categories and types of 311 requests
- **Dim_Source** – Channel through which requests were received

Files:

AMEYPARANGE6	
my-beam-pipeline	
src	
main / java / com / <u>example</u>	
J BostonRequestStgLoad.java	3
J DimDateTimeETL.java	1
J FactRequestETL.java	3
J LocationDimETL.java	5
J RequestDtldimETL.java	1
J SourceDimETL.java	1
> test	
> target	

Technical Implementation:

Data Ingestion Layer (Bronze Layer - Raw Data Storage)

- Source: Boston 311 Service Requests (Parquet in GCS)
- Bronze Layer: Store raw CSV files in Google Cloud Storage (GCS) without modifications
- Extract raw 311 data from GCS (Bronze Layer) and load it into Apache Beam (Dataflow) for processing

Data Processing & Transformation Layer (Silver Layer - Cleaned Data)

- Silver Layer: Apply cleaning, deduplication, and standardization in Apache Beam
- ETL Pipelines (Apache Beam - Dataflow) process and enrich data
- Handle missing values, data type inconsistencies, and data standardization
- Apply deduplication and ensure incremental processing
- Stored Data in Parquet Files.

Data Storage & Query Layer (Gold Layer - BigQuery Data Warehouse)

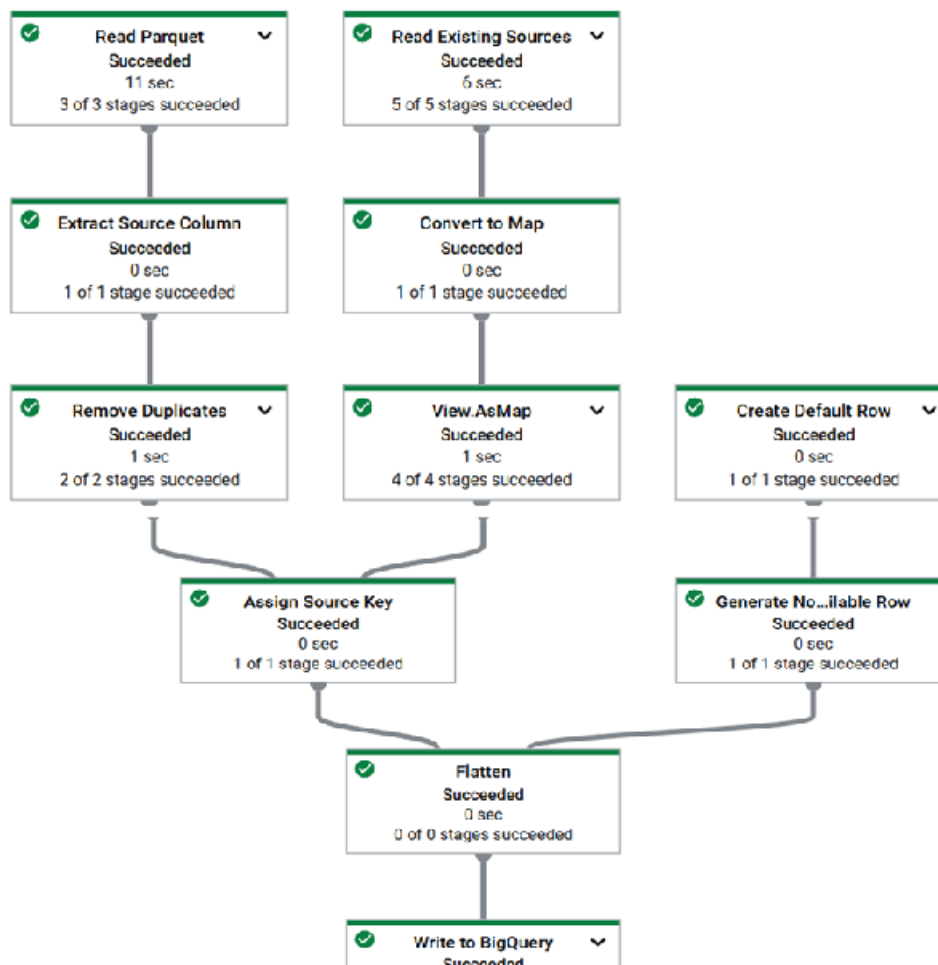
- Gold Layer: Load cleaned and structured data into BigQuery for analytics
- Convert raw data into structured dimension tables (dim_date, dim_time, dim_location, dim_source, dim_request_dtl)
- Optimize table schema with indexing and partitioning for query performance
- Store Fact Table (fact_311_requests) to support analytical queries

Performance Optimization

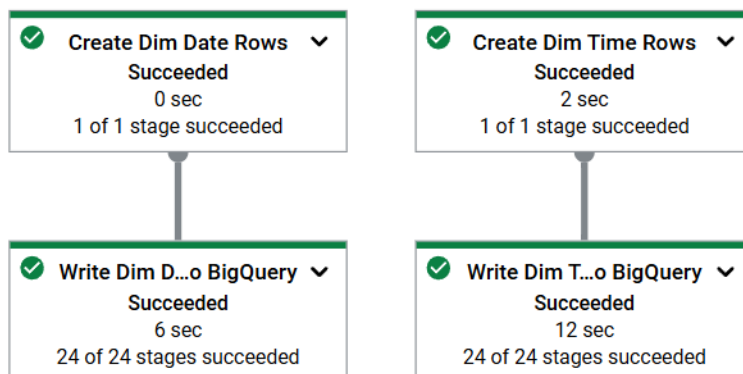
- Used **Apache Avro & Parquet** for optimized storage.
- Implemented **batch processing** for efficient large-scale data ingestion.
- Applied **partitioning and clustering** in BigQuery for improved query performance.

Data Pipelines:

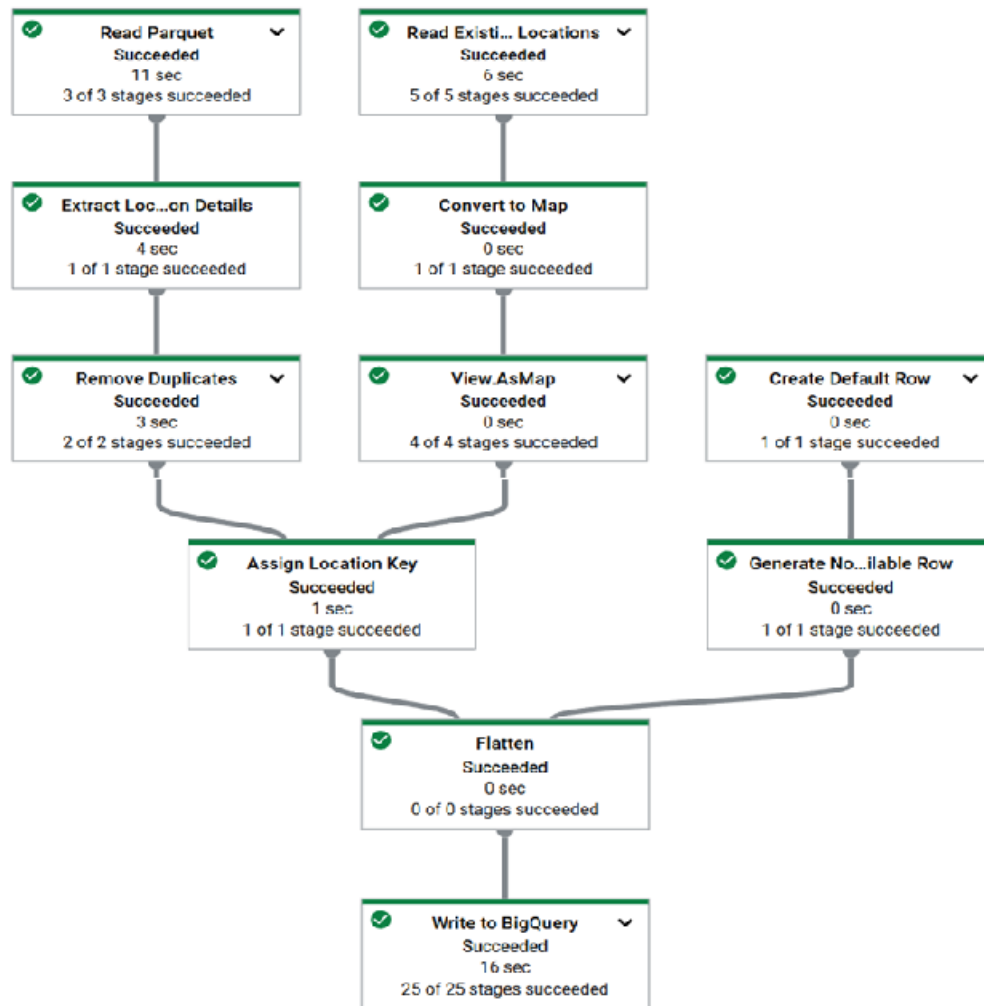
Dim_Source load Pipeline



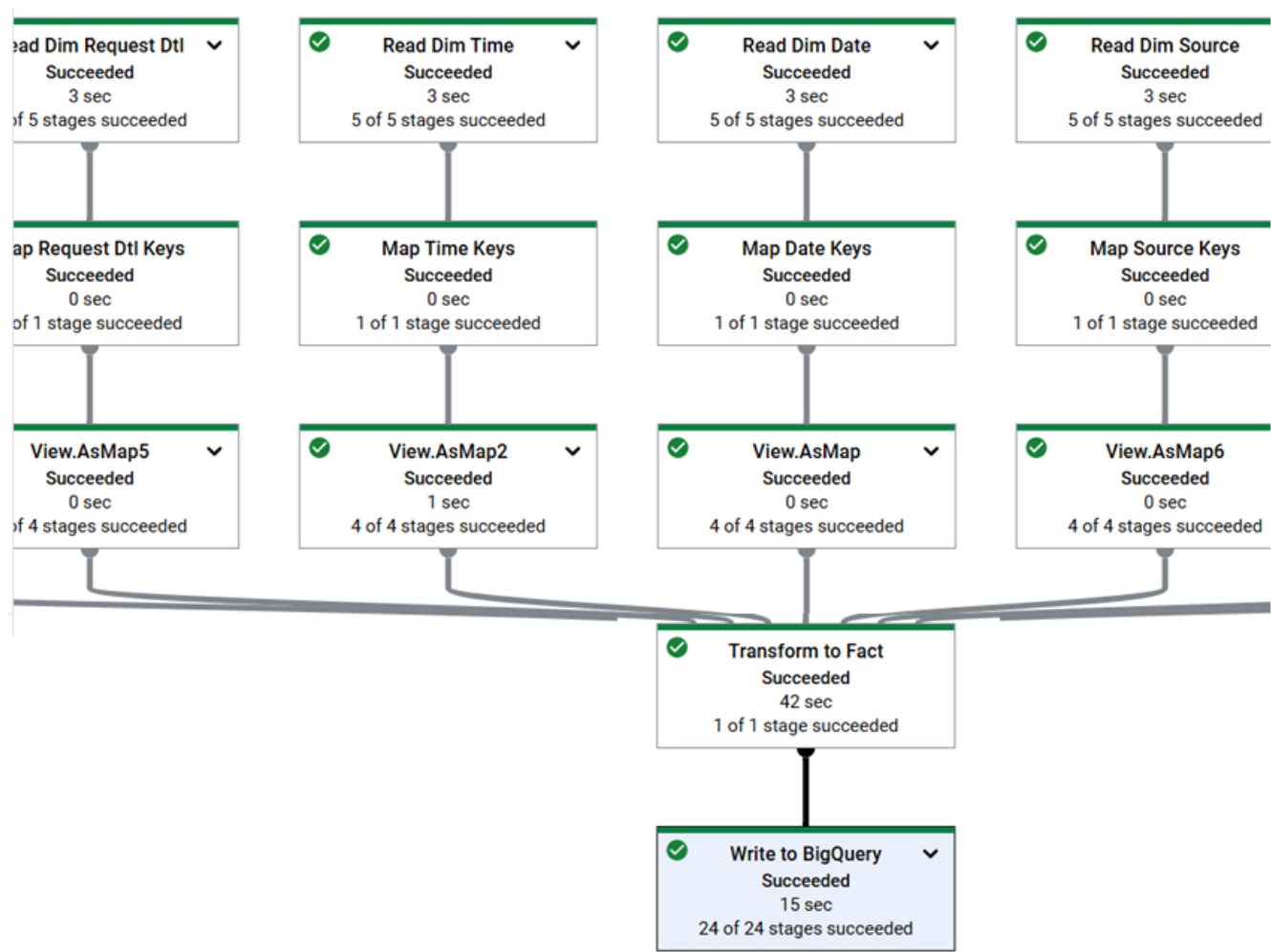
Dim_date load + Dim_Time load Pipeline



Dim_location load Pipeline



Fact load Pipeline



Output:

Stage_load

my-311-service-data-lake

Location

us-central1 (Iowa)

Storage class

Standard

Public access

Subject to object ACLs

Protection

Soft Delete

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

NEW

INVENTORY REPORTS

OPERATIONS

Folder browser

my-311-service-data-lake

bronze/

gold/

fact_requests/

silver/

staging/

temp/

Buckets > my-311-service-data-lake > silver

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	
<input type="checkbox"/>	nul	0 B	application/octet-stream	Mar 15, 2025, 4:35:25 PM	Standard	
<input type="checkbox"/>	stage_boston_311_requests-0000...	8.4 MB		Mar 17, 2025, 3:37:15 AM	Standard	
<input type="checkbox"/>	stage_boston_311_requests-0000...	5.6 MB		Mar 17, 2025, 3:37:15 AM	Standard	
<input type="checkbox"/>	stage_boston_311_requests-0000...	12.6 MB		Mar 17, 2025, 3:37:15 AM	Standard	

Dim_date

dim_date

QUERY

OPEN IN

SHARE

COPY

SNAPSHOT

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

Row	date_key	date	year	month	day	
1	20150101	2015-01-01	2015	1	1	
2	20150102	2015-01-02	2015	1	2	
3	20150103	2015-01-03	2015	1	3	
4	20150104	2015-01-04	2015	1	4	
5	20150105	2015-01-05	2015	1	5	
6	20150106	2015-01-06	2015	1	6	
7	20150107	2015-01-07	2015	1	7	
8	20150108	2015-01-08	2015	1	8	
9	20150109	2015-01-09	2015	1	9	
10	20150110	2015-01-10	2015	1	10	
11	20150111	2015-01-11	2015	1	11	
12	20150112	2015-01-12	2015	1	12	
13	20150113	2015-01-13	2015	1	13	
14	20150114	2015-01-14	2015	1	14	
15	20150115	2015-01-15	2015	1	15	
16	20150116	2015-01-16	2015	1	16	
17	20150117	2015-01-17	2015	1	17	
18	20150118	2015-01-18	2015	1	18	

Dim_time

dim_time

QUERY

OPEN IN

SHARE

COPY

SNAPSHOT

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

Row	time_key	time	hour	minute	second	
1	0	00:00:00	0	0	0	
2	1	00:00:01	0	0	1	
3	2	00:00:02	0	0	2	
4	3	00:00:03	0	0	3	
5	4	00:00:04	0	0	4	
6	5	00:00:05	0	0	5	
7	6	00:00:06	0	0	6	
8	7	00:00:07	0	0	7	
9	8	00:00:08	0	0	8	
10	9	00:00:09	0	0	9	
11	10	00:00:10	0	0	10	
12	11	00:00:11	0	0	11	
13	12	00:00:12	0	0	12	
14	13	00:00:13	0	0	13	
15	14	00:00:14	0	0	14	
16	15	00:00:15	0	0	15	
17	16	00:00:16	0	0	16	
18	17	00:00:17	0	0	17	

Dim_location

location_key	is_intersection	street_name1	street_name2	neighborhood	city
226	0	1254 COMMONWEALTH AVE	<i>null</i>	ALLSTON	Boston
328	0	150 WESTERN AVE	<i>null</i>	ALLSTON	Boston
784	0	52 HARVARD AVE	<i>null</i>	ALLSTON	Boston
884	0	1375 COMMONWEALTH AVE	<i>null</i>	ALLSTON	Boston
1215	0	20 PRATT ST	<i>null</i>	ALLSTON	Boston
1554	0	1304 COMMONWEALTH AVE	<i>null</i>	ALLSTON	Boston
2053	0	81 HARVARD AVE	<i>null</i>	ALLSTON	Boston
2109	0	55 LINDEN ST	<i>null</i>	ALLSTON	Boston
2114	0	137 BRIGHTON AVE	<i>null</i>	ALLSTON	Boston
2191	0	21 BRIGHTON AVE	<i>null</i>	ALLSTON	Boston
2532	0	248 KELTON ST	<i>null</i>	ALLSTON	Boston
2774	0	78 ADAMSON ST	<i>null</i>	ALLSTON	Boston
2913	0	288-308 N HARVARD ST	<i>null</i>	ALLSTON	Boston
3670	0	1 LONG AVE	<i>null</i>	ALLSTON	Boston
3721	0	47 EVERETT ST	<i>null</i>	ALLSTON	Boston
4274	0	37 PARK VALE AVE	<i>null</i>	ALLSTON	Boston
4351	0	40 RUGG RD	<i>null</i>	ALLSTON	Boston

Dim_location_dtl

location_dtl_key	neighborhood_service	police_district	fire_district	pwd_district	city_council_district	ward	precinct
479	0	A15	-1	1A	0	0	0207
822	0	C6	4	UNKNOWN	0	0	0308
378	0	C6	4	05	2	0	0801
647	0	C6	4	1C	2	0	0801
144	0	A15	3	1A	1	0	UNKNOWN
699	0	D14	-1	04	0	21	UNKNOWN
348	0	A7	-1	UNKNOWN	0	0	UNKNOWN
514	0	C6	6	UNKNOWN	0	6	UNKNOWN
724	1	A7	1	09	1	1	0101
44	1	A7	1	09	1	1	0102
796	1	A7	1	09	1	1	0103
57	1	A7	1	09	1	1	0104

Dim_source

dim_source

QUERY

OPEN IN

<

SCHEMA

DETAILS

PREVIEW

TAB

Row	source_key	source
1	-1	Not Available
2	1	CITIZENS CONNECT APP
3	2	CITY WORKER APP
4	3	CONSTITUENT CALL
5	4	SELF SERVICE
6	5	EMPLOYEE GENERATED

Dim_request_dtl

dim_request_dtl

QUERY

OPEN IN

SHARE

COPY

SNAPSHOT

DELETE

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

LINEAGE

DATA PROFILE

DATA

Row	request_dtl_key	case_title	subject	reason	type
1	79	ABANDONED BICYCLE	MAYOR'S 24 HOUR HOTLINE	ABANDONED BICYCLE	ABANDONED BICYCLE
2	123	GENERAL REQUEST	BOSTON WATER & SEWER CO...	ADMINISTRATIVE & GENERAL ...	BWSC GENERAL REQUEST
3	116	PARKS GENERAL REQUEST	PARKS & RECREATION DEPART...	ADMINISTRATIVE & GENERAL ...	PARKS GENERAL REQUEST
4	18	PUBLIC WORKS GENERAL REQ...	PUBLIC WORKS DEPARTMENT	ADMINISTRATIVE & GENERAL ...	PUBLIC WORKS GENERAL REQ...
5	85	PUBLIC WORKS GENERAL REQ...	PUBLIC WORKS DEPARTMENT	ADMINISTRATIVE & GENERAL ...	PUBLIC WORKS GENERAL REQ...
6	151	PUBLIC WORKS GENERAL REQ...	PUBLIC WORKS DEPARTMENT	ADMINISTRATIVE & GENERAL ...	PUBLIC WORKS GENERAL REQ...
7	228	PUBLIC WORKS GENERAL REQ...	PUBLIC WORKS DEPARTMENT	ADMINISTRATIVE & GENERAL ...	PUBLIC WORKS GENERAL REQ...
8	59	TRANSPORTATION GENERAL R...	TRANSPORTATION - TRAFFIC ...	ADMINISTRATIVE & GENERAL ...	TRANSPORTATION GENERAL R...
9	181	ANIMAL GENERIC REQUEST	ANIMAL CONTROL	ANIMAL ISSUES	ANIMAL GENERIC REQUEST
10	235	BUILDING INSPECTION REQUE...	INSPECTIONAL SERVICES	BUILDING	BUILDING INSPECTION REQUE...
11	127	CONTRACTORS COMPLAINT	INSPECTIONAL SERVICES	BUILDING	CONTRACTORS COMPLAINT
12	189	ELECTRICAL	INSPECTIONAL SERVICES	BUILDING	ELECTRICAL
13	198	ILLEGAL OCCUPANCY	INSPECTIONAL SERVICES	BUILDING	ILLEGAL OCCUPANCY
14	201	ILLEGAL ROOMING HOUSE	INSPECTIONAL SERVICES	BUILDING	ILLEGAL ROOMING HOUSE
15	193	UNSAFE/DANGEROUS CONDIT...	INSPECTIONAL SERVICES	BUILDING	UNSAFE DANGEROUS CONDITI...
16	104	WORK W/OUT PERMIT	INSPECTIONAL SERVICES	BUILDING	WORK W/OUT PERMIT
17	3	WORKING BEYOND HOURS	INSPECTIONAL SERVICES	BUILDING	WORKING BEYOND HOURS

Fact

Query results

SAVE RESULTS OPEN IN

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH	
Row	est_id	case_enquiry_id		open_date	sla_date		close_date	open_
1	147	101005231559		2023-12-31	2024-01-01T21:22:23		2024-01-01T03:23:31	
2	30017	101005231885		2024-01-01	2024-01-02T09:05:29		2024-01-01T09:13:47	
3	82923	101005231873		2024-01-01	2024-01-02T08:53:37		2024-01-01T10:12:56	
4	676	101005231992		2024-01-01	2024-01-02T10:45:48		2024-01-02T01:55:11	

Data Visualization



Year 2022

277K

No of Requests

17.33%

Overdue Request %

13

Avg_Resolution_Time (Days)

82.67%

SLA Compliance Rate

YoY Analysis 2022-2023

neighborhood

All

13.22%

Yearly Requests Change %

Top 5 Reasons remains same

Year 2023

313K

No of Requests

18.76%

Overdue Request %

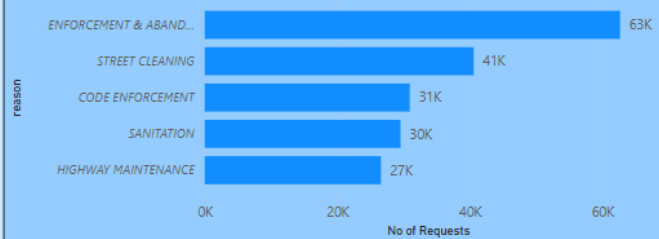
7

Avg_Resolution_Time (Days)

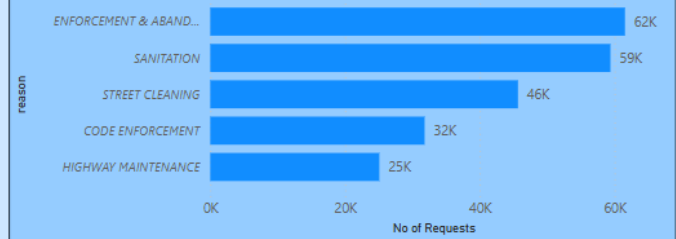
81.24%

SLA Compliance Rate

Top 5 Reasons



Top 5 Reasons



Day Period Analysis

Select all

Afternoon

Early Morning

Evening

Late Night

Morning

Night

Year

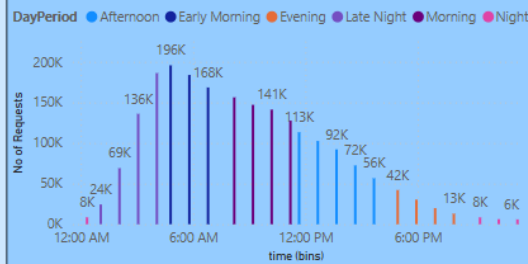
2016

2023

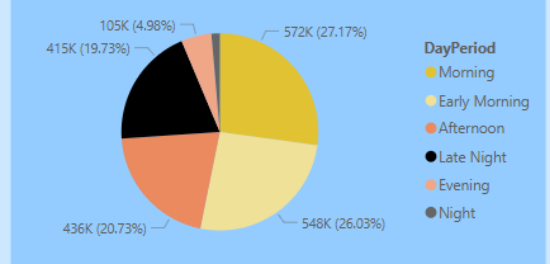
Neighborhood

All

No of Requests by time (bins) and DayPeriod



No of Requests by DayPeriod



After Sunrise

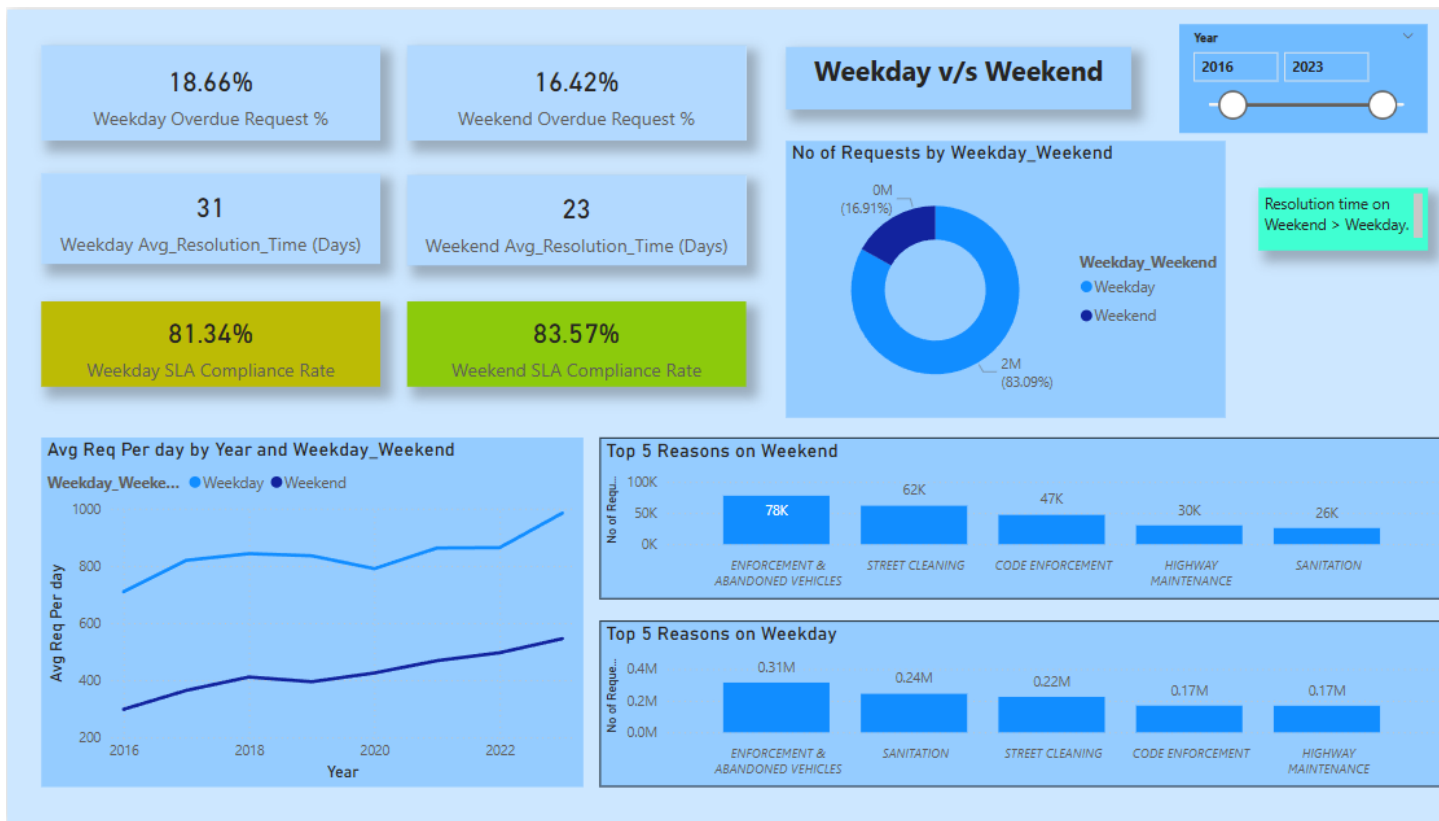
DayPeriod	Rank	reason	TotalRequests
Afternoon	1	ENFORCEMENT & ABANDONED VEHICLES	126527
Afternoon	2	SANITATION	73478
Afternoon	3	STREET CLEANING	67936
Early Morning	1	STREET CLEANING	98724
Early Morning	2	SANITATION	92334
Early Morning	3	ENFORCEMENT & ABANDONED VEHICLES	91737
Morning	1	SANITATION	109636
Morning	2	ENFORCEMENT & ABANDONED VEHICLES	103626
Morning	3	STREET CLEANING	94973

"Street Cleaning" and "Enforcement & Abandoned Vehicles" remain prominent throughout the day.

Reason "Street Lights" is in top 3 during Evening and Night times.

After Sunset

DayPeriod	Rank	reason	TotalRequests
Evening	1	ENFORCEMENT & ABANDONED VEHICLES	42788
Evening	2	STREET CLEANING	14584
Evening	3	STREET LIGHTS	12586
Late Night	1	STREET CLEANING	96977
Late Night	2	ENFORCEMENT & ABANDONED VEHICLES	91231
Late Night	3	CODE ENFORCEMENT	54865
Night	1	STREET CLEANING	9272
Night	2	ENFORCEMENT & ABANDONED VEHICLES	8410
Night	3	STREET LIGHTS	4642



Future Scope

- Delta Lakehouse Implementation
- Enhanced Airflow Orchestration
- Microservices & Real-Time Processing