SANITATION
AND THE CITY

CIS 9440 - Section S1DA - Group 4

Tamara Baez tamara.baez@baruchmail.cuny.edu
Daniel Xu chaodi.xu@baruchmail.cuny.edu
Dustin Kearns dustin.kearns@baruchmail.cuny.edu
Jia Liu jia.liu3@baruchmail.cuny.edu

Table of Contents

1. Introduction

1.1 The Problem: Sanitation

New York City is infamous for its rat-infested subways and dirty streets. Residents and visitors to the Big Apple are no strangers to grit and grime. Every year hundreds of thousands of calls are made to 311 to complain about unsanitary conditions, yet whether these calls result in any effective action is questionable. Using data collected from these calls, we are hoping to give New Yorkers a better understanding of where the city's sanitation efforts are lacking, and where the most work is needed to make NYC a cleaner and healthier place to live.

Each of us represent a different borough across NYC: the Bronx, Queens, Manhattan and Brooklyn. Throughout our 40 years as NYC residents, collectively, we have experienced sanitation issues firsthand. Traveling in the MTA with rodents, walking through streets of overcapacity garbage cans, and watching people litter are just a few of the issues that are relatable to all NYC residents.

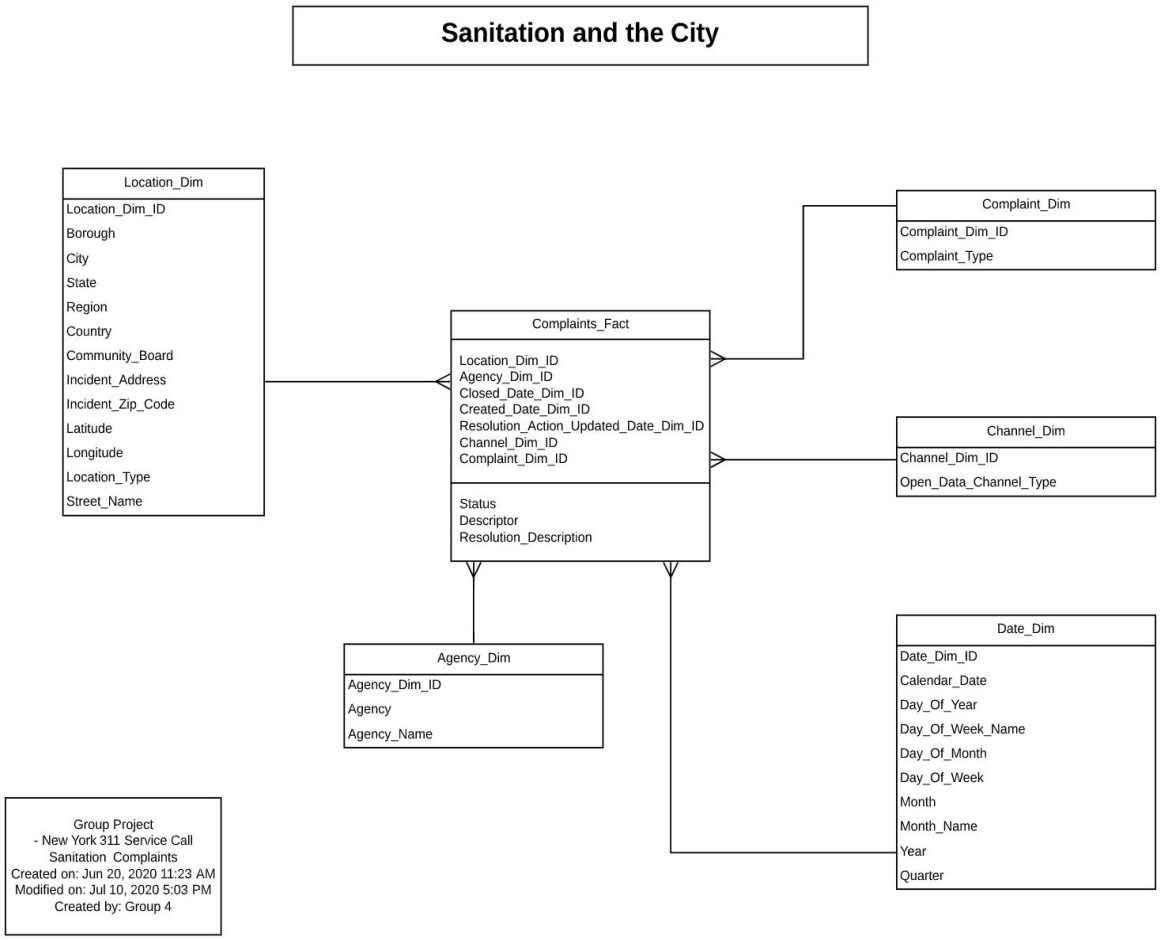
To carry out our work we have chosen to use the NYC Open Data's 311 Service Request data. We chose 5 complaint types to include in our warehouse. These topics include "Unsanitary Condition", "Sanitation Condition", "Dirty Conditions", "Rodent" and "Unsanitary Pigeon Condition". Together, these complaint types cover a broad range of sanitation issues and will enable users to gain a deep understanding of the many problems negatively impacting the health and quality of life of people in the city. It is our hope that the dashboard our work results in will be used by residents and city employees alike, and that it will help enable and drive changes that will make the city a safer and cleaner place for all. Below is the total number for each complaint type chosen for the year 2017. We also included data from years 2015, 2016, 2018 and 2019.

Count (2017)	Complaint Type
79282	Unsanitary Condition
38937	Sanitation Condition
35887	Dirty Conditions
35075	Rodent
628	Unsanitary Pigeon Condition

1.2 Key Performance Indicators (KPI)

1. **Resolution**
    1. Average time(days) duration for resolution by agency
    2. Average time(days) duration for resolution by borough/neighborhood
    3. Average time(days) duration for resolution by month
    4. Most frequently assigned agencies
    5. Overall resolution rate (%)
2. **Volume of Complaints**
    1. Total number of 311 complaint calls by month
    2. Total number of 311 complaint calls by descriptor
    3. Total number of 311 complaint calls by zip code
3. **Location**
    1. Average number of 311 complaint calls by borough/neighborhood

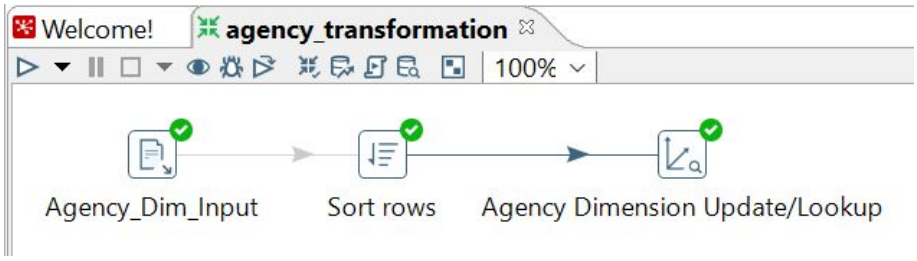
2. Dimensional Model



3. ETL Process

3.1 Agency Dimension

Notes: The Agency dimension is created at this step of the ETL transformation. The source data of 2015 is imported into CSV File Import and the relevant fields are selected based on our dimensional model. On the Sort Rows step, rows are sorted and duplicates are removed. At the final Agency Dimension Update/Lookup step, both fields (agency and agency_name) are set as keys, and the Agency Dimension table is created. We inserted the remaining csv files (years 2016-2019) into this transformation as well.



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Agency_Dim_Input	0	0	95413	95414	0	0	0	0	Finished	0.6s	151,692	-
2	Sort rows	0	95413	42	0	0	0	0	0	Finished	0.8s	121,082	-
3	Agency Dimension Update/Lookup	0	42	42	42	42	0	0	0	Finished	2.9s	15	-

Transformation Screen and Step Metrics for Agency Import Transformation

Script Output | Query Result

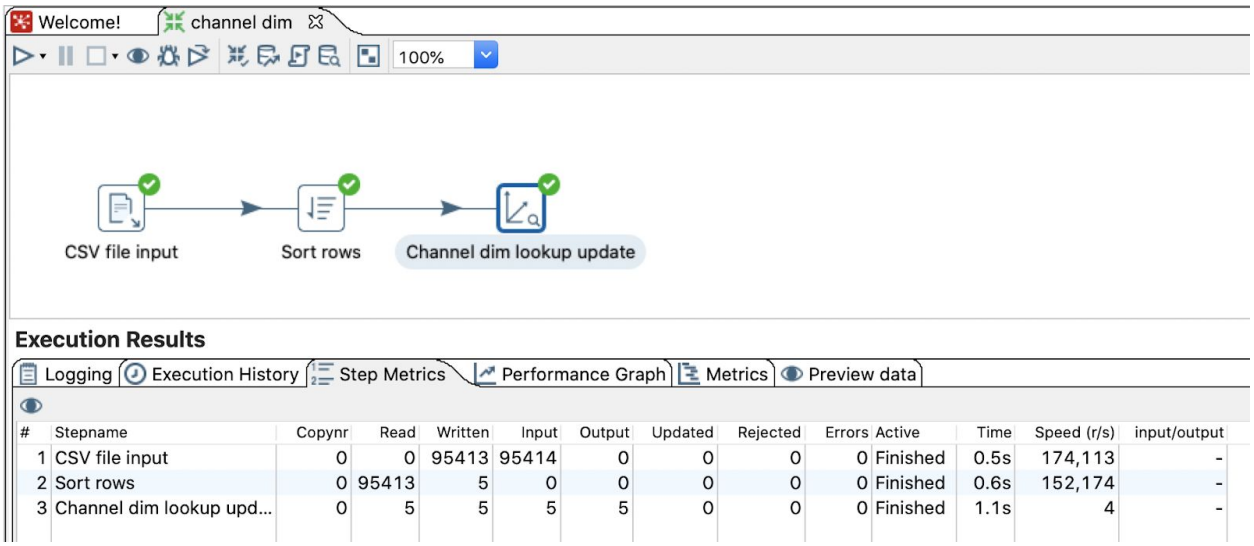
SQL | All Rows Fetched: 43 in 0.036 seconds

	AGENCY_DIM_ID	VERSION	DATE_FROM	DATE_TO	AGENCY	AGENCY_NAME
1	0	1	(null)	(null)	(null)	(null)
2	1	1	01-JAN-00	31-DEC-99	DOHMH	Department of Health and Mental Hygiene
3	2	1	01-JAN-00	31-DEC-99	DSNY	A - Bronx
4	3	1	01-JAN-00	31-DEC-99	DSNY	A - Brooklyn
5	4	1	01-JAN-00	31-DEC-99	DSNY	A - Canine Task Force Citywide
6	5	1	01-JAN-00	31-DEC-99	DSNY	A - Illegal Posting Manhattan and Bronx
7	6	1	01-JAN-00	31-DEC-99	DSNY	A - Illegal Posting Staten Island, Queens and Brooklyn
8	7	1	01-JAN-00	31-DEC-99	DSNY	A - Manhattan
9	8	1	01-JAN-00	31-DEC-99	DSNY	A - Queens
10	9	1	01-JAN-00	31-DEC-99	DSNY	A - Staten Island
11	10	1	01-JAN-00	31-DEC-99	DSNY	BCC - Bronx
12	11	1	01-JAN-00	31-DEC-99	DSNY	BCC - Brooklyn North
13	12	1	01-JAN-00	31-DEC-99	DSNY	BCC - Brooklyn South
14	13	1	01-JAN-00	31-DEC-99	DSNY	BCC - Manhattan
15	14	1	01-JAN-00	31-DEC-99	DSNY	BCC - Queens East

Sample Data of Agency Dimension

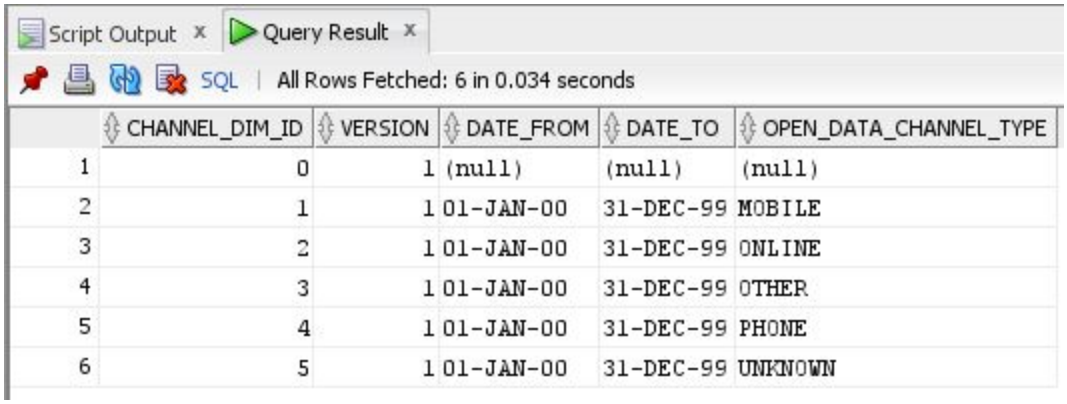
3.2 Channel Dimension

Notes: The Channel dimension is created at this step of ETL transformation. The source data of 2015 is imported into CSV File Import and the relevant fields are selected. At the Sort Rows step, rows are sorted and duplicates are removed. At the final Channel Dimension Update/Lookup step, the only field (Open_data_channel_type) is set as the key, and the Channel Dimension Table is created on the Oracle database. We inserted the remaining csv files (years 2016-2019) into this transformation as well.



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	95413	95414	0	0	0	0	Finished	0.5s	174,113	-
2	Sort rows	0	95413	5	0	0	0	0	0	Finished	0.6s	152,174	-
3	Channel dim lookup upd...	0	5	5	5	5	0	0	0	Finished	1.1s	4	-

Transformation Screen and Step Metrics for Channel Import Transformation

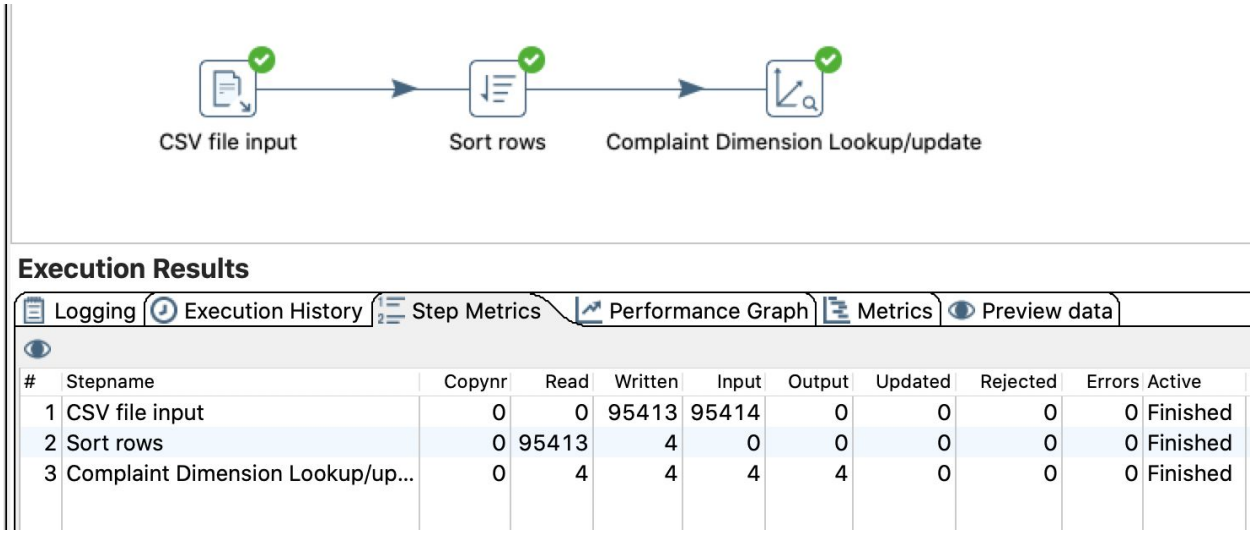


	CHANNEL_DIM_ID	VERSION	DATE_FROM	DATE_TO	OPEN_DATA_CHANNEL_TYPE
1	0	1	(null)	(null)	(null)
2	1	1	01-JAN-00	31-DEC-99	MOBILE
3	2	1	01-JAN-00	31-DEC-99	ONLINE
4	3	1	01-JAN-00	31-DEC-99	OTHER
5	4	1	01-JAN-00	31-DEC-99	PHONE
6	5	1	01-JAN-00	31-DEC-99	UNKNOWN

Sample Data of Channel Dimension

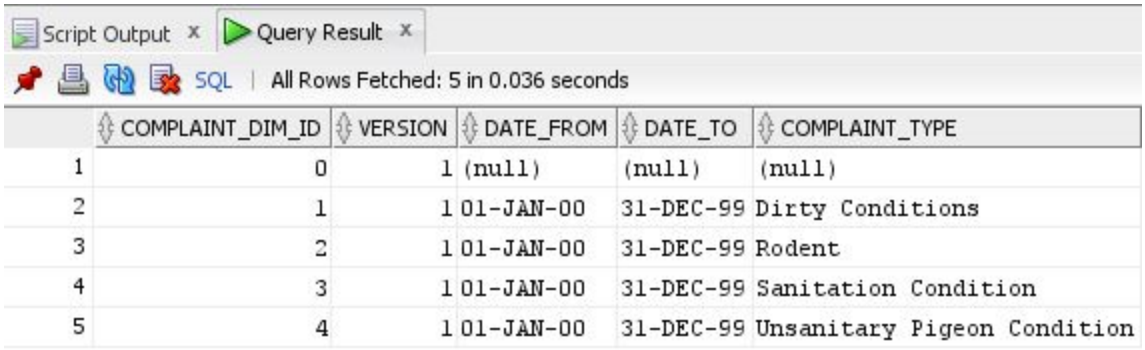
3.3 Complaint Dimension

Notes: The Complaint dimension is created at this step of ETL transformation. The source data of 2015 is imported into CSV File Import and the relevant fields are selected. At Sort Rows step, rows are sorted and duplicates are removed. At the final Complaint Dimension Update/Lookup step, the only field (Complaint_type) is set as the key, and the Complaint Dimension Table is created on Oracle database. We inserted the remaining csv files (years 2016-2019) into this transformation as well.



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active
1	CSV file input	0	0	95413	95414	0	0	0	0	Finished
2	Sort rows	0	95413	4	0	0	0	0	0	Finished
3	Complaint Dimension Lookup/up...	0	4	4	4	4	0	0	0	Finished

Transformation Screen and Step Metrics for Complaint Import Transformation

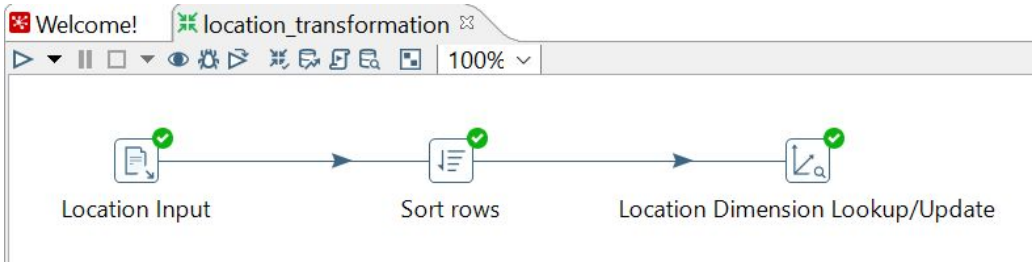


	COMPLAINT_DIM_ID	VERSION	DATE_FROM	DATE_TO	COMPLAINT_TYPE
1	0	1	(null)	(null)	(null)
2	1	1	01-JAN-00	31-DEC-99	Dirty Conditions
3	2	1	01-JAN-00	31-DEC-99	Rodent
4	3	1	01-JAN-00	31-DEC-99	Sanitation Condition
5	4	1	01-JAN-00	31-DEC-99	Unsanitary Pigeon Condition

Sample Data of Complaint Dimension

3.4 Location Dimension

Notes: The Location dimension is created at this step of ETL transformation. The source data of 2015 is imported into CSV File Import and relevant fields are selected. At Sort Rows step, rows are sorted and duplicates are removed. At the final Location Dimension Update/Lookup step, we decide the location dimension is a Type 1 SCD. The fields, longitude and latitude, are set as keys, and the Location Dimension Table is created on Oracle database. We inserted the remaining csv files (years 2016-2019) into this transformation as well.



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Location Input	0	0	95413	95414	0	0	0	0	Finished	0.5s	184,197	-
2	Sort rows	0	95413	69174	0	0	0	0	0	Finished	46mn 22s	34	-
3	Location Dimension Lookup/Update	0	69174	69174	69174	58709	20930	0	0	Finished	54mn 9s	24	-

Transformation Screen and Step Metrics for Location Import Transformation

Script Output | Query Result

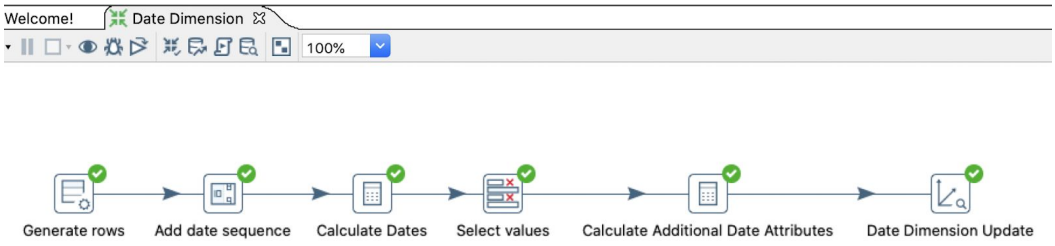
SQL | Fetched 900 rows in 0.719 seconds

	LOCATION_DIM_ID	VERSION	DATE_FROM	DATE_TO	LATITUDE	LONGITUDE	BOROUGH	CITY	COMMUNITY_BOARD	INCIDENT_ADDRESS	INCIDENT_ZIP	LOCATION_TYPE	STREET_NAME	STATE	REGION	COUNTRY
1	142	1	01-JAN-00	31-DEC-99	40.81162068	-73.92611713	BRONX	BRONX	01 BRONX	2578 3 AVENUE	10454	Lot	3 AVENUE	NY	NorthEast	US
2	143	1	01-JAN-00	31-DEC-99	40.81188938	-73.92566888	BRONX	BRONX	01 BRONX	2595 3 AVENUE	10451	Sidewalk	3 AVENUE	NY	NorthEast	US
3	144	1	01-JAN-00	31-DEC-99	40.80801204	-73.91894349	BRONX	BRONX	01 BRONX	260 BROOK AVENUE	10454	Street	BROOK AVENUE	NY	NorthEast	US
4	145	1	01-JAN-00	31-DEC-99	40.81042043	-73.92486134	BRONX	BRONX	01 BRONX	262 ALEXANDER AVENUE	10454	3+ Family Apt. Building	ALEXANDER AVENUE	NY	NorthEast	US
5	146	1	01-JAN-00	31-DEC-99	40.81271723	-73.92406037	BRONX	BRONX	01 BRONX	2635 3 AVENUE	10451	Sidewalk	3 AVENUE	NY	NorthEast	US
6	147	1	01-JAN-00	31-DEC-99	40.81434	-73.92091912	BRONX	BRONX	01 BRONX	2742 3 AVENUE	10455	Sidewalk	3 AVENUE	NY	NorthEast	US
7	148	1	01-JAN-00	31-DEC-99	40.8189887	-73.91968521	BRONX	BRONX	01 BRONX	277 EAST 152 STREET	10451	Street	EAST 152 STREET	NY	NorthEast	US
8	149	1	01-JAN-00	31-DEC-99	40.80851137	-73.91865389	BRONX	BRONX	01 BRONX	278 BROOK AVENUE	10454	3+ Family Mixed Use Building	BROOK AVENUE	NY	NorthEast	US
9	150	1	01-JAN-00	31-DEC-99	40.8076428	-73.91691022	BRONX	BRONX	01 BRONX	278 ST ANNS AVENUE	10454	3+ Family Apt. Building	ST ANNS AVENUE	NY	NorthEast	US
10	151	1	01-JAN-00	31-DEC-99	40.8112747	-73.92588632	BRONX	BRONX	01 BRONX	279 EAST 139 STREET	10454	Street	EAST 139 STREET	NY	NorthEast	US
11	152	1	01-JAN-00	31-DEC-99	40.81517324	-73.91925991	BRONX	BRONX	01 BRONX	2792 3 AVENUE	10455	Sidewalk	3 AVENUE	NY	NorthEast	US
12	153	1	01-JAN-00	31-DEC-99	40.81498414	-73.91966476	BRONX	BRONX	01 BRONX	2809 3 AVENUE	10455	Street	3 AVENUE	NY	NorthEast	US
13	154	1	01-JAN-00	31-DEC-99	40.81714845	-73.92180092	BRONX	BRONX	01 BRONX	281 EAST 149 STREET	10451	3+ Family Apt. Building	EAST 149 STREET	NY	NorthEast	US
14	155	1	01-JAN-00	31-DEC-99	40.80778271	-73.91681612	BRONX	BRONX	01 BRONX	282 ST ANNS AVENUE	10454	Sidewalk	ST ANNS AVENUE	NY	NorthEast	US
15	156	1	01-JAN-00	31-DEC-99	40.80866229	-73.9185923	BRONX	BRONX	01 BRONX	283 BROOK AVENUE	10454	3+ Family Apt. Building	BROOK AVENUE	NY	NorthEast	US

Sample Data of Location Dimension

3.5 Date Dimension

Notes: The Date dimension is created at this step of ETL transformation. The start date is first set up as 12/31/2014 and the row limit to 1,900, which captured the 5 years of data we are loading. We then set other rules and add fields through Add Date Sequence, Calculate Dates, Select Values and Calculate Additional Date Attributes. The Date Dimension Table is then created on Oracle database.



Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Generate rows	0	0	1900	0	0	0	0	0	Finished	0.0s	316,667	-
2	Add date sequence	0	1900	1900	0	0	0	0	0	Finished	0.0s	211,111	-
3	Calculate Dates	0	1900	1900	0	0	0	0	0	Finished	0.0s	135,714	-
4	Select values	0	1900	1900	0	0	0	0	0	Finished	0.0s	111,765	-
5	Calculate Additional Date Attri...	0	1900	1900	0	0	0	0	0	Finished	0.1s	29,231	-
6	Date Dimension Update	0	1900	1900	1900	1900	0	0	0	Finished	1mn 46s	18	-

Transformation Screen and Step Metrics for Date Dimension

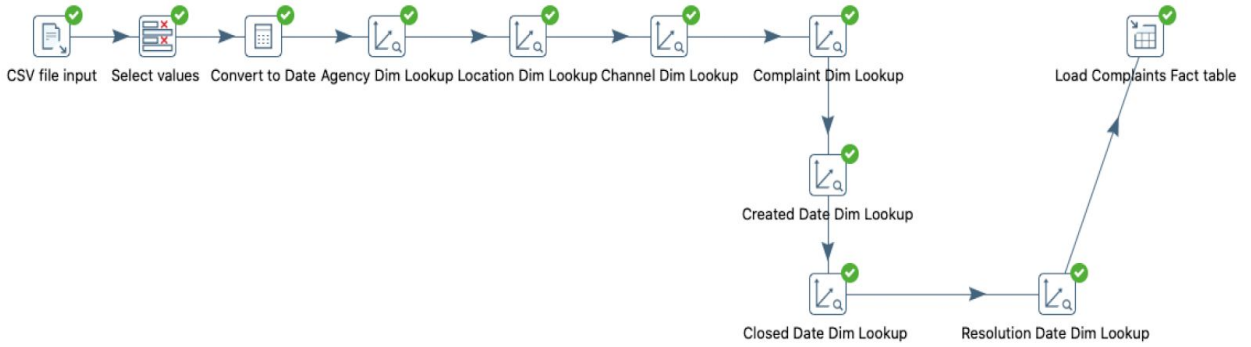
SQL | All Rows Fetched: 1901 in 0.872 seconds

	DATE_DIM_ID	VERSION	DATE_FROM	DATE_TO	CALENDAR_DATE	DAY_OF_YEAR	MONTH	YEAR	QUARTER	MONTH_NAME	DAY_OF_WEEK_NAME	DAY_OF_WEEK	DAY_OF_MONTH
1889	1888	1	01-JAN-00	31-DEC-99	03-MAR-20	63	3	2020	1	March	Tuesday	3	3
1890	1889	1	01-JAN-00	31-DEC-99	04-MAR-20	64	3	2020	1	March	Wednesday	4	4
1891	1890	1	01-JAN-00	31-DEC-99	05-MAR-20	65	3	2020	1	March	Thursday	5	5
1892	1891	1	01-JAN-00	31-DEC-99	06-MAR-20	66	3	2020	1	March	Friday	6	6
1893	1892	1	01-JAN-00	31-DEC-99	07-MAR-20	67	3	2020	1	March	Saturday	7	7
1894	1893	1	01-JAN-00	31-DEC-99	08-MAR-20	68	3	2020	1	March	Sunday	1	8
1895	1894	1	01-JAN-00	31-DEC-99	09-MAR-20	69	3	2020	1	March	Monday	2	9
1896	1895	1	01-JAN-00	31-DEC-99	10-MAR-20	70	3	2020	1	March	Tuesday	3	10
1897	1896	1	01-JAN-00	31-DEC-99	11-MAR-20	71	3	2020	1	March	Wednesday	4	11
1898	1897	1	01-JAN-00	31-DEC-99	12-MAR-20	72	3	2020	1	March	Thursday	5	12
1899	1898	1	01-JAN-00	31-DEC-99	13-MAR-20	73	3	2020	1	March	Friday	6	13
1900	1899	1	01-JAN-00	31-DEC-99	14-MAR-20	74	3	2020	1	March	Saturday	7	14
1901	1900	1	01-JAN-00	31-DEC-99	15-MAR-20	75	3	2020	1	March	Sunday	1	15

Sample Data of Date Dimension

3.6 Fact Table

Notes: The Fact table is created at this step of ETL transformation. Below are screenshots of inputting the 2015, 2018, and 2019 source data as an example. All 5 years (2015-2019) were inputted into the facts table. The date format/type is converted at Select Values Step, date is separated from time at Convert to Date step, and then all of the surrogate keys are connected to create the fact table.



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	95413	95414	0	0	0	0	Finished	22mn 50s	70	-
2	Select values	0	95413	95413	0	0	0	0	0	Finished	26mn 20s	60	-
3	Convert to Date	0	95413	95413	0	0	0	0	0	Finished	29mn 49s	53	-
4	Agency Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	33mn 18s	48	-
5	Location Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	36mn 37s	43	-
6	Channel Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	40mn 54s	39	-
7	Complaint Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	40mn 54s	39	-
8	Created Date Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	40mn 54s	39	-
9	Closed Date Dim Lookup	0	95413	95413	95413	0	0	0	0	Finished	40mn 54s	39	-
10	Resolution Date Dim Look...	0	95413	95413	95413	0	0	0	0	Finished	40mn 57s	39	-
11	Load Complaints Fact table	0	95413	95413	0	95413	0	0	0	Finished	40mn 57s	39	-

Transformation Screen and Step Metrics for Fact Table (2015)

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	114269	114270	0	0	0	0	Finished	40mn 44s	47	-
2	Select values	0	114269	114269	0	0	0	0	0	Finished	45mn 30s	42	-
3	Convert to Date	0	114269	114269	0	0	0	0	0	Finished	50mn 12s	38	-
4	Agency Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	54mn 57s	35	-
5	Location Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	56mn 23s	34	-
6	Channel Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	56mn 29s	34	-
7	Complaint Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	56mn 32s	34	-
8	Created Date Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	56mn 35s	34	-
9	Closed Date Dim Lookup	0	114269	114269	114269	0	0	0	0	Finished	56mn 37s	34	-
10	Resolution Date Dim Look...	0	114269	114269	114269	0	0	0	0	Finished	56mn 41s	34	-
11	Load Complaints Fact table	0	114269	114269	0	114269	0	0	0	Finished	56mn 41s	34	-

Step Metrics for Fact Table (2018)

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	91472	91473	0	0	0	0	Finished	28mn 51s	53	-
2	Select values	0	91472	91472	0	0	0	0	0	Finished	33mn 35s	45	-
3	Convert to Date	0	91472	91472	0	0	0	0	0	Finished	38mn 15s	40	-
4	Agency Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	43mn 2s	35	-
5	Location Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	43mn 12s	35	-
6	Channel Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	43mn 48s	35	-
7	Complaint Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	45mn 37s	33	-
8	Created Date Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	45mn 38s	33	-
9	Closed Date Dim Lookup	0	91472	91472	91472	0	0	0	0	Finished	45mn 41s	33	-
10	Resolution Date Dim Look...	0	91472	91472	91472	0	0	0	0	Finished	45mn 44s	33	-
11	Load Complaints Fact table	0	91472	91472	0	91472	0	0	0	Finished	45mn 44s	33	-

Step Metrics for Fact Table (2019)

	DESCRIPTOR	RESOLUTION_DESCRIPTION	STATUS	AGENCY_DIM_ID	LOCATION_DIM_ID	CHANNEL_DIM_ID	COMPLAINT_DIM_ID
2	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	7	30584	1	1
3	E15 Illegal Postering	The Department of Sanitation ha...	Closed	6	39653	2	1
4	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	3	20041	2	1
5	E3 Dirty Sidewalk	The Department of Sanitation ha...	Closed	9	54958	2	1
6	E15 Illegal Postering	The Department of Sanitation in...	Closed	6	40519	4	1
7	15 Street Cond/Dump-O...	The Department of Sanitation in...	Closed	30	32291	1	3
8	E8 Canine Violation	The Department of Sanitation in...	Closed	4	4132	4	1
9	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	2	721	1	1
10	15 Street Cond/Dump-O...	The Department of Sanitation ha...	Closed	10	801	1	3
11	E8 Canine Violation	The Department of Sanitation in...	Closed	4	19400	4	1
12	E3A Dirty Area/Alleyway	The Department of Sanitation in...	Closed	3	21807	4	1
13	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	3	10596	2	1
14	15 Street Cond/Dump-O...	The Department of Sanitation re...	Closed	11	14575	1	3
15	E15 Illegal Postering	The Department of Sanitation in...	Closed	6	49927	2	1
16	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	3	15222	2	1
17	E3 Dirty Sidewalk	The Department of Sanitation in...	Closed	2	5493	2	1
18	E8 Canine Violation	The Department of Sanitation in...	Closed	4	14943	4	1
19	E12 Illegal Dumping S...	The Department of Sanitation in...	Closed	36	41340	4	1
20	E8 Canine Violation	The Department of Sanitation in...	Closed	4	2615	4	1
21	E4 18" Law	The Department of Sanitation in...	Closed	2	1883	2	1

Sample Data of Fact Table

4. Final Schema

4.1 SQL Code for Creating View

```
CREATE VIEW complaint_view AS
    SELECT
        c1.descriptor,
        c1.resolution_description,
        c1.status,
        c1.agency_dim_id,
        c1.location_dim_id,
        c1.channel_dim_id,
        c1.complaint_dim_id,
        c1.created_date_dim_id,
        c1.closed_date_dim_id,
        c1.resolution_action_updated_date_dim_id,
        c1.agency,
        c1.agency_name,
        c1.open_data_channel_type,
        c1.complaint_type,
        c1.latitude,
        c1.longitude,
        c1.borough,
        c1.city,
        c1.community_board,
        c1.incident_address,
        c1.incident_zip,
        c1.location_type,
        c1.street_name,
        c1.state,
        c1.region,
        c1.country,
        c1.created_date_calendar_date,
        c1.created_date_day_of_year,
        c1.created_date_month,
        c1.created_date_year,
        c1.created_date_quarter,
        c1.created_date_month_name,
        c1.created_date_day_of_week_name,
        c1.created_date_day_of_week,
        c1.created_date_day_of_month,
```

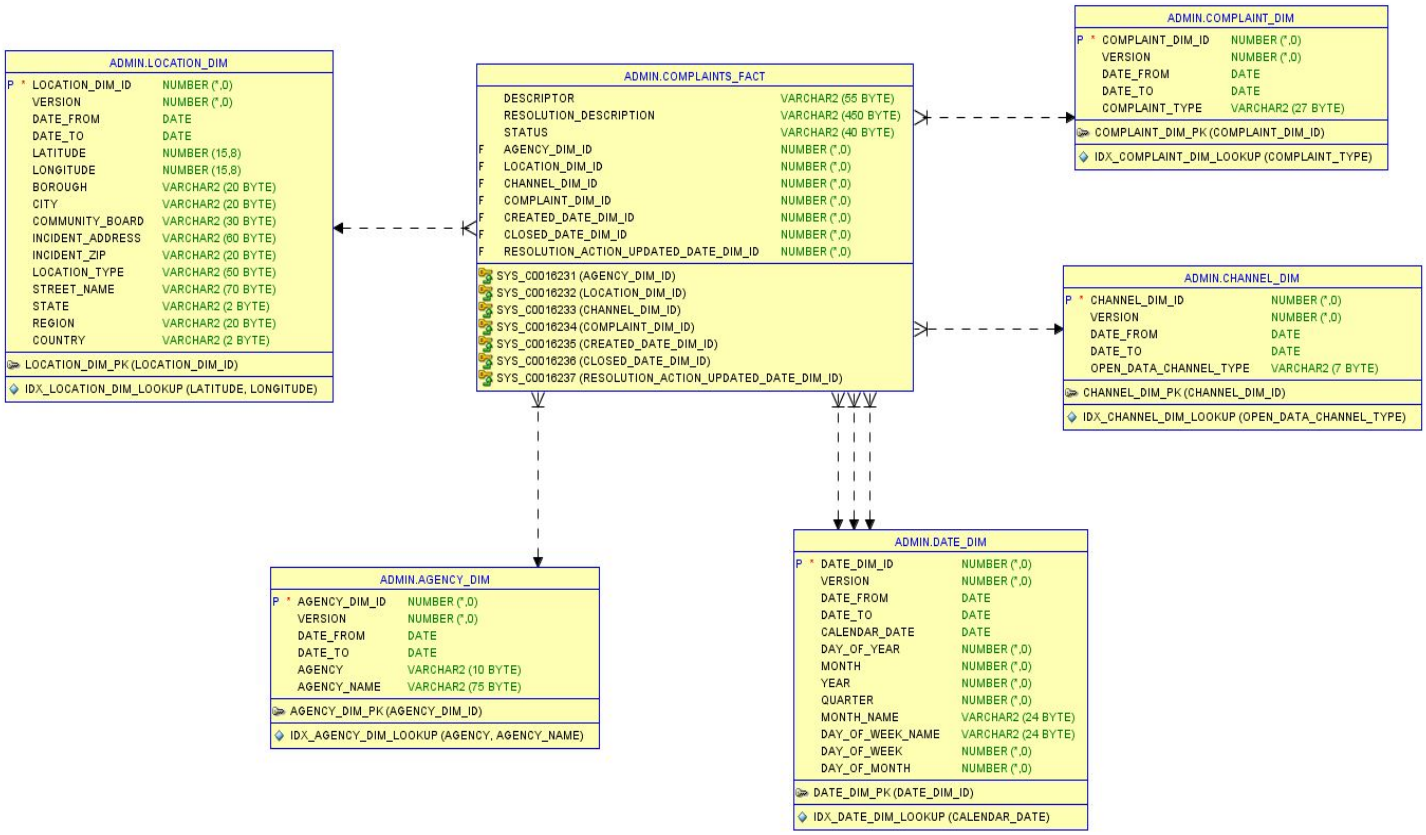
```
        c2.closed_date_calendar_date,
        c2.closed_date_day_of_year,
        c2.closed_date_month,
        c2.closed_date_year,
        c2.closed_date_quarter,
        c2.closed_date_month_name,
        c2.closed_date_day_of_week_name,
        c2.closed_date_day_of_week,
        c2.closed_date_day_of_month,
        c3.resolution_action_updated_date_calendar_date,
        c3.resolution_action_updated_date_day_of_year,
        c3.resolution_action_updated_date_month,
        c3.resolution_action_updated_date_year,
        c3.resolution_action_updated_date_quarter,
        c3.resolution_action_updated_date_month_name,
        c3.resolution_action_updated_date_day_of_week_name,
        c3.resolution_action_updated_date_day_of_week,
        c3.resolution_action_updated_date_day_of_month
    FROM
        ( (
            SELECT
                complaints_fact.descriptor
AS descriptor,
                complaints_fact.resolution_description
AS resolution_description,
                complaints_fact.status
AS status,
                complaints_fact.agency_dim_id
AS agency_dim_id,
                complaints_fact.location_dim_id
AS location_dim_id,
                complaints_fact.channel_dim_id
AS channel_dim_id,
                complaints_fact.complaint_dim_id
AS complaint_dim_id,
                complaints_fact.created_date_dim_id
AS created_date_dim_id,
                complaints_fact.closed_date_dim_id
AS closed_date_dim_id,
                complaints_fact.resolution_action_updated_date_dim_id
AS resolution_action_updated_date_dim_id,
                agency_dim.agency
AS agency,
```

```
                agency_dim.agency_name
AS agency_name,
                channel_dim.open_data_channel_type
AS open_data_channel_type,
                complaint_dim.complaint_type
AS complaint_type,
                location_dim.latitude
AS latitude,
                location_dim.longitude
AS longitude,
                location_dim.borough
AS borough,
                location_dim.city
AS city,
                location_dim.community_board
AS community_board,
                location_dim.incident_address
AS incident_address,
                location_dim.incident_zip
AS incident_zip,
                location_dim.location_type
AS location_type,
                location_dim.street_name
AS street_name,
                location_dim.state
AS state,
                location_dim.region
AS region,
                location_dim.country
AS country,
                date_dim.calendar_date
AS created_date_calendar_date,
                date_dim.day_of_year
AS created_date_day_of_year,
                date_dim.month
AS created_date_month,
                date_dim.year
AS created_date_year,
                date_dim.quarter
AS created_date_quarter,
                date_dim.month_name
AS created_date_month_name,
```

```
                date_dim.day_of_week_name
AS created_date_day_of_week_name,
                date_dim.day_of_week
AS created_date_day_of_week,
                date_dim.day_of_month
AS created_date_day_of_month
            FROM
                ( ( ( ( complaints_fact left
                JOIN agency_dim ON complaints_fact.agency_dim_id =
agency_dim.agency_dim_id ) left
                JOIN channel_dim ON complaints_fact.channel_dim_id =
channel_dim.channel_dim_id ) left
                JOIN complaint_dim ON
complaints_fact.complaint_dim_id = complaint_dim.complaint_dim_id )
left
                JOIN location_dim ON complaints_fact.location_dim_id
= location_dim.location_dim_id ) left
                JOIN date_dim ON complaints_fact.created_date_dim_id
= date_dim.date_dim_id
        ) c1
        LEFT JOIN (
            SELECT
                date_dim_id,
                date_dim.calendar_date      AS
closed_date_calendar_date,
                date_dim.day_of_year        AS
closed_date_day_of_year,
                date_dim.month              AS closed_date_month,
                date_dim.year               AS closed_date_year,
                date_dim.quarter            AS closed_date_quarter,
                date_dim.month_name         AS
closed_date_month_name,
                date_dim.day_of_week_name   AS
closed_date_day_of_week_name,
                date_dim.day_of_week        AS
closed_date_day_of_week,
                date_dim.day_of_month       AS
closed_date_day_of_month
            FROM
                date_dim
        ) c2 ON c1.closed_date_dim_id = c2.date_dim_id ) left
        JOIN (
            SELECT
```

```
                date_dim_id,
                date_dim.calendar_date      AS
resolution_action_updated_date_calendar_date,
                date_dim.day_of_year        AS
resolution_action_updated_date_day_of_year,
                date_dim.month              AS
resolution_action_updated_date_month,
                date_dim.year               AS
resolution_action_updated_date_year,
                date_dim.quarter            AS
resolution_action_updated_date_quarter,
                date_dim.month_name         AS
resolution_action_updated_date_month_name,
                date_dim.day_of_week_name   AS
resolution_action_updated_date_day_of_week_name,
                date_dim.day_of_week        AS
resolution_action_updated_date_day_of_week,
                date_dim.day_of_month       AS
resolution_action_updated_date_day_of_month
            FROM
                date_dim
        ) c3 ON c1.resolution_action_updated_date_dim_id =
c3.date_dim_id;
```

4.2 Physical Schema



The physical schema above shows the 5 dimensional tables and complaint fact table created. These tables display primary keys and foreign keys of each table along with the relationships.

5. Dashboard Application

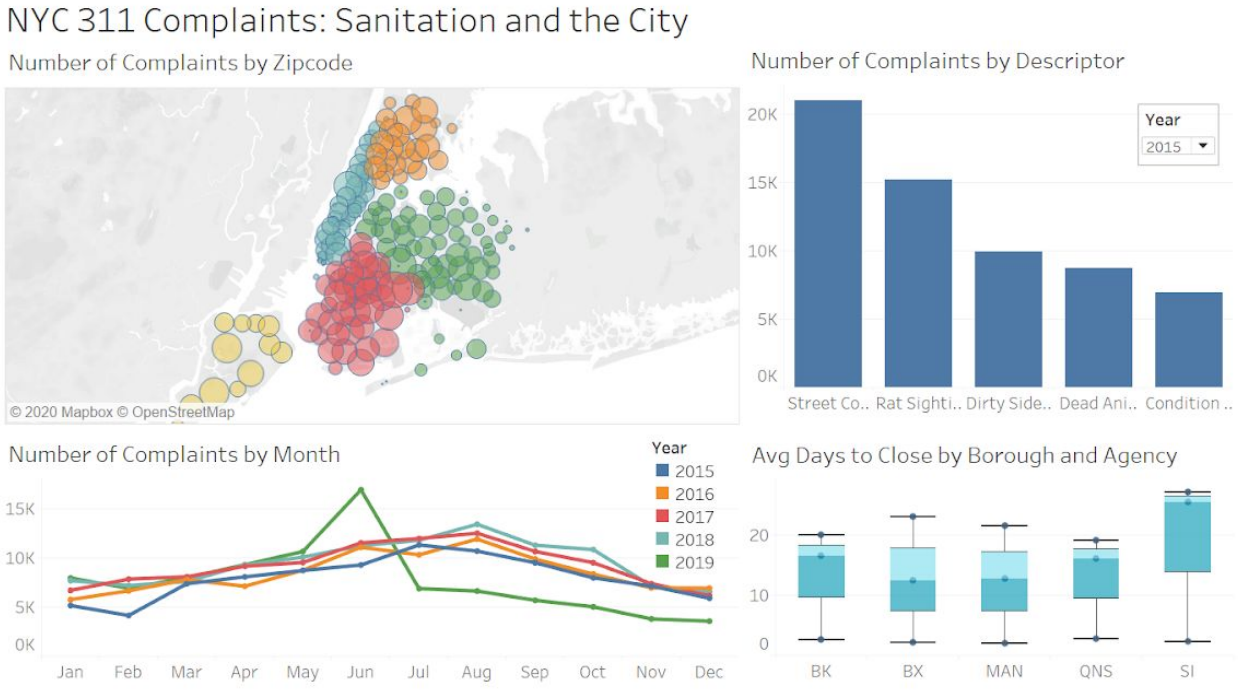


Tableau Public Link:
https://public.tableau.com/views/NYC311Complaints/Dashboard1?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

The visualization on the top left represents the KPI for the total number of 311 complaints per zip code. The color differentiates the five boroughs and the circle shape size represents the amount of complaints. For instance, the larger the circle size the more complaints originate from that zip code.

The second visualization located on the top right is for the total number of top five 311 complaints descriptors for each year. This KPI helps determine if these top 5 complaint descriptions improve or not each year.

The third visualization on the bottom left is for the KPI to determine the number of complaints per month each year. We are able to establish common trends by comparing this data over the course of the 5 years.

The last visualization location on the bottom right is the KPI for average time duration for resolution by agency and borough. The time duration was calculated using the difference between the created date and closed date.

6. Conclusion

Tools: Socrata Open Data API , Pentaho Data Integration, Tableau, LucidChart, Oracle SQL Developer, Oracle Database
Programming languages: Python, SQL

The source data we used, “311 Service Requests from 2010 to Present”, is from the website NYC OpenData. In order to smooth our online communication, we tried various tools including Slack, Google Drive, Zoom, and emails. First, we used LucidChart to draw the dimensional model. Then we used Socrata Open Data API to acquire the source data and Python in Jupyter notebook to manipulate and export it as five separate csv files by year. After this we executed the ETL process in Pentaho Data Integration, and imported our data into 5 separate dimensions and a fact table in an Oracle database running in the Oracle Cloud. We then used Oracle SQL Developer to create a view of our data, exported it as a csv file, and used it to create an interactive dashboard in Tableau.

The most difficult part of the project for us was the ETL process, especially when we tried to create the date dimension and fact table. There was lots of debugging during this process, which we did not anticipate. Despite this, we managed to complete the process successfully and had a much easier time creating the view in Oracle SQL Developer and the dashboard in Tableau. In contrast to the ETL process, we feel creating the dimensional model was the easiest part of the project. Regardless of difficulty, all parts of the project were very educational. Overall we all feel we have a much better understanding of data warehousing in general, and experience with several tools that are likely to come in handy in the future.

We believe that NYC residents and agencies using our newly created dashboard will be empowered to make significant changes for the better in their communities. By being able to track the volume of complaints in different zip codes throughout the five boroughs, residents will be able to compare the conditions in their neighborhood to those of others. Furthermore, city officials will have a better sense of what areas require the most attention. In addition to this, the amount of time it takes to address complaints can be tracked, allowing for greater awareness of the city’s efforts to address various issues. This dashboard will help track common trends and identify areas of improvement to help decrease the amount of sanitation issues in NYC. All in all, we think that the benefits originally proposed for this project will be realized, and that all residents and visitors to NYC will benefit as a result.

7. Appendix

7.1 References

1. 311 Service Requests from 2010 to Present

   311, D. (2020, June 25). 311 Service Requests from 2010 to Present: NYC Open Data. Retrieved June 26, 2020, from https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

2. Getting Started with NYC OpenData and the Socrata API

   Holowczak, R. (n.d.). Getting Started with NYC OpenData and the Socrata API. Retrieved July 1, 2020, from http://holowczak.com/getting-started-with-nyc-opendata-and-the-socrata-api

7.2 Meeting Notes

Meeting 1:
Meeting time: 6/12/20 4:30 - 6:30pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Project Proposal

Meeting 2:
Meeting Time: 6/19/20 4:30pm - 8:30pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Created a csv file for each year 2015-1019. Data cleaning; removed irrelevant attributes, renaming attributes. Started creating draft dimensional model

Meeting 3:
Meeting Time: 6/20/20 10:30am - 11:30am
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Finished draft dimensional model

Meeting 4:
Meeting Time: 6/26/20 4:30pm - 6:30pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Completed dimensional model. Modified csv file for each year 2015-2019 to match the attributes on the final dimensional model

Meeting 5:
Meeting Time: 7/2/20 4pm - 5:30pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Created the group oracle cloud connection, created agency dimension table

Meeting 6:
Meeting Time: 7/2/20 8pm - 10pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Created the remaining dimension tables (location, channel, date, complaint)

Meeting 7:
Meeting Time: 7/5/20 11am - 2pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Working on creating facts table

Meeting 8:
Meeting Time: 7/10/20 4:30pm- 10pm
Attendees: Dustin Kearns, Tamara Baez, Daniel Xu, Jia Liu
Main topic discussed/work: Created dashboard, finalized report