THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

NYPD Crime Analysis

Group 5

Akshay Kamath

Purvi Thakor

Zhen Zhang

Xiaoyu Wang

Zhaoyang Chen

Jiachen Xu

Introduction

- Background:

  This dataset includes crimes like Burglary, Robbery, Vehicle & Traffic Laws reported to the New York City Police Department (NYPD) for all complete quarters so far this year (2017). Offenses occurring at intersections are represented at the Latitude & Longitude of the intersection. Crimes occurring anywhere other than an intersection are geo-located to the middle of the block.

  We have have completed the following tasks:

  - Cleanse the data
  - Create a conceptual model of the data
  - Create a dimensional model
  - Analyze the data quality
  - ETL
  - Create a data quality scorecard
  - Create visualizations to showcase the data

Field Names

Column	Description
CMPLNT_NUM	Randomly generated persistent ID for each complaint
CMPLNT_FR_DT	Exact date of occurrence for the reported event (or starting date of occurrence, if CMPLNT_TO_DT exists)
CMPLNT_FR_TM	Exact time of occurrence for the reported event (or starting time of occurrence, if CMPLNT_TO_TM exists)
RPT_DT	Date event was reported to police
KY_CD	Three digit offense classification code
PD_CD	Three digit internal classification code (more granular than Key Code)
PD_DESC	Description of internal classification corresponding with PD code
CRM_ATPT_CPTD_CD	Indicator of whether crime was successfully completed or attempted
JURIS_DESC	Jurisdiction responsible for incident. Either internal or external
BORO_NM	The name of the borough in which the incident occurred
Latitude	Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

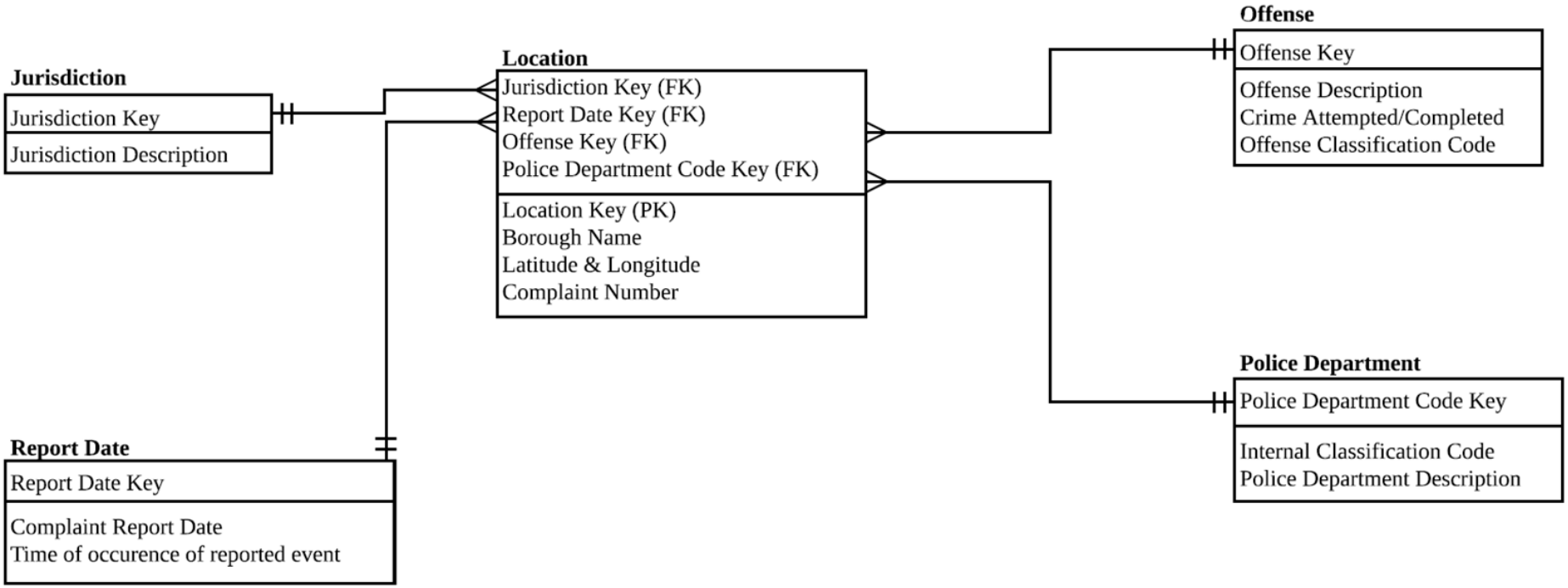
Dimensions

RPT_DT	CMPLNT_FR_TM	KY_CD	OFNS_DESC	CRM_ATPT_CPTD_CD	PD_CD	PD_DESC	JURIS_DESC
1/2/2017	12:50:00 PM	107	BURGLARY	COMPLETED	214	BURGLARY,COMMERCIAL,UNKNOWN TI	N.Y. POLICE DEPT
1/2/2017	6:30:00 AM	107	BURGLARY	ATTEMPTED	211	BURGLARY,COMMERCIAL,DAY	N.Y. POLICE DEPT
1/2/2017	12:00:00 PM	107	BURGLARY	COMPLETED	221	BURGLARY,RESIDENCE,DAY	N.Y. POLICE DEPT
1/2/2017	1:26:00 AM	107	BURGLARY	COMPLETED	223	BURGLARY,RESIDENCE,NIGHT	N.Y. POLICE DEPT
1/2/2017	8:00:00 AM	107	BURGLARY	COMPLETED	241	BURGLARY,UNCLASSIFIED,DAY	N.Y. POLICE DEPT

Facts

CMPLNT_NUM	BORO_NM	Lat_Lon
978359059	QUEENS	(40.740175712, -73.895037279)
322904633	MANHATTAN	(40.864608183, -73.919194097)
148313152	QUEENS	(40.765496305, -73.866175114)
672455964	BRONX	(40.850829671, -73.915220281)
718876502	BROOKLYN	(40.662897268, -73.987111942)

Conceptual Model (using Lucid Chart)



Dimensional Model (using MySQL workbench)

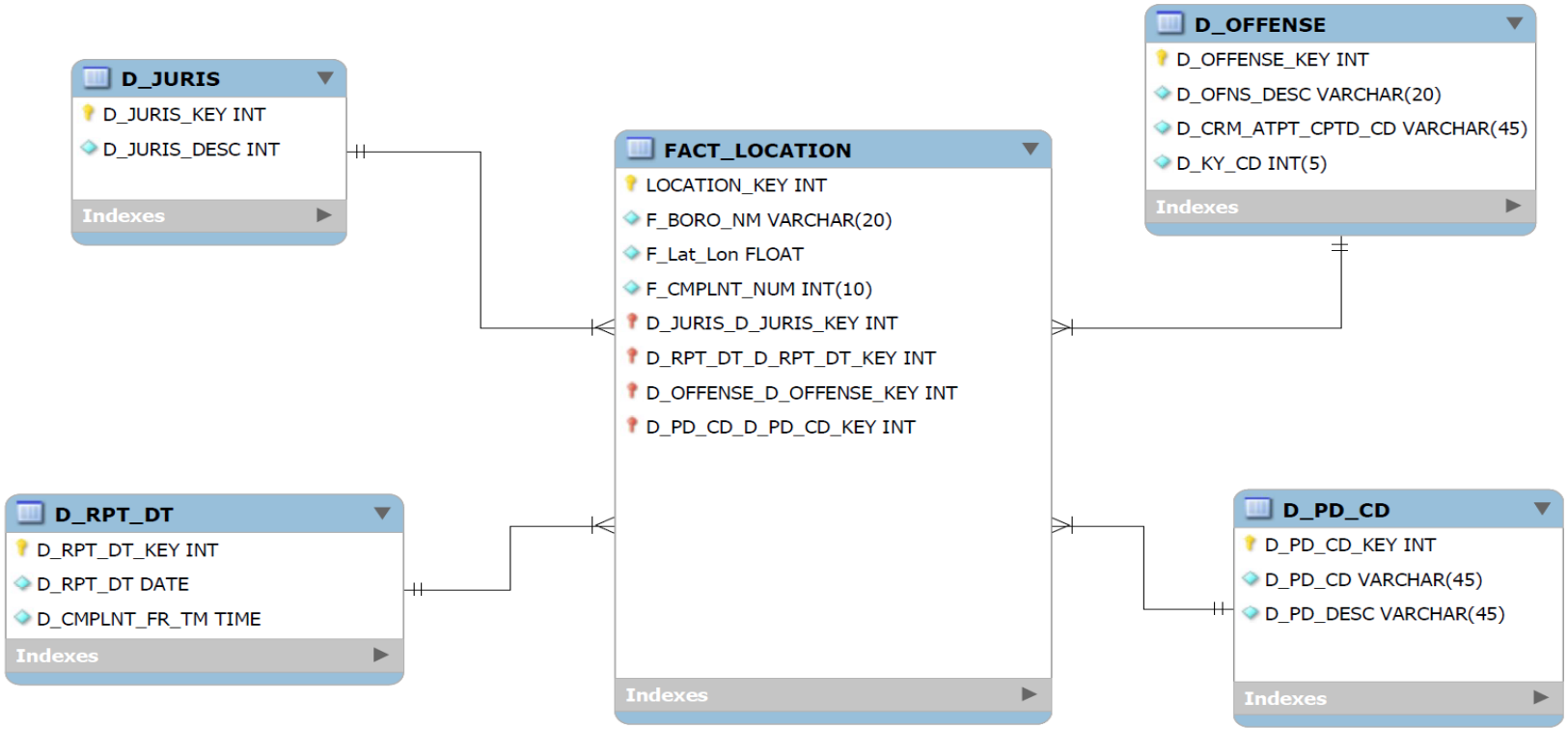


Table Descriptions

TABLES	FIELDS	EXPLANATION
D_REPORT_DATE	RPT_DT	Date event was reported to police .
D_OFFENSE	OFNS_DESC	Description of offense corresponding with key code.
	CRM_ATPT_CPTD_CD	Indicator of whether crime was successfully completed or attempted, but failed or was interrupted prematurely.
D_PD_CD	PD_CD	Three digit internal classification code.
	PD_DESC	Description of internal classification corresponding with PD code (more granular than Offense Description).
D_JURISDICTION	JURIS_DESC	Jurisdiction responsible for incident. Either internal, like Police, Transit, and Housing; or external, like Correction, Port Authority, etc.
F_LOCATION	CMPLNT NO	The complaint number that was recorded against the complaint.
	BORO_NM	The name of the borough in which the incident occurred.
	Lat_Lon	Latitude & Longitude of area where the crime occurred.

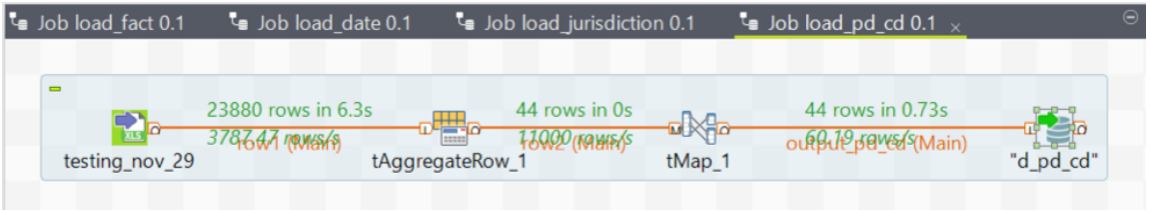
Data Quality Scorecard

Completeness	Metric Name	Target	Valid	Invalid	Score
	CMPLNT_NUM	100%	23880	0	100.0%
	RPT_DT	100%	23880	0	100.0%
	CMPLNT_FR_TM	100%	23880	0	100.0%
	KY_CD	100%	23880	0	100.0%
	OFNS_DESC	100%	23880	0	100.0%
	CRM_ATPT_CPTD_CD	100%	23880	0	100.0%
	PD_CD	100%	23880	0	100.0%
	PD_DESC	100%	23880	0	100.0%
	JURIS_DESC	100%	23880	0	100.0%
	BORO_NM	100%	23880	0	100.0%
	Lat_Lon	100%	23880	0	100.0%
Confirmity	**Metric Name**	**Target**	**Valid**	**Invalid**	**Score**
	CMPLNT_NUM	90%	23880	0	100.0%
	RPT_DT	90%	23880	0	100.0%
	CMPLNT_FR_TM	90%	23880	0	100.0%
	KY_CD	90%	23880	0	100.0%
	OFNS_DESC	90%	23880	0	100.0%
	CRM_ATPT_CPTD_CD	90%	23880	0	100.0%
	PD_CD	90%	23880	0	100.0%
	PD_DESC	90%	23880	0	100.0%
	JURIS_DESC	90%	23880	0	100.0%
	BORO_NM	90%	23880	0	100.0%
	Lat_Lon	90%	23880	0	100.0%

Timeliness	Metric Name	Target	Valid	Invalid	Score
	RPT_DT	100%	23853	27	99.9%
Uniqueness	**Metric Name**	**Target**	**Valid**	**Invalid**	**Score**
	CMPLNT_NUM	100%	23880	0	100.0%
Integrity	**Metric Name**	**Target**	**Valid**	**Invalid**	**Score**
	KY_CD	95%	23880	0	100.0%
	OFNS_DESC	95%	23880	0	100.0%
	CRM_ATPT_CPTD_CD	95%	23880	0	100.0%
	PD_CD	95%	23880	0	100.0%
	PD_DESC	95%	23880	0	100.0%
	JURIS_DESC	95%	23880	0	100.0%
	BORO_NM	95%	23880	0	100.0%

ETL Process

Job (Dimension Table - PD_CD)



Output (Dimension Table - PD_CD)

SQL File 4*

Limit to 50000 row

```
1 • select * from d_pd_cd
2
```

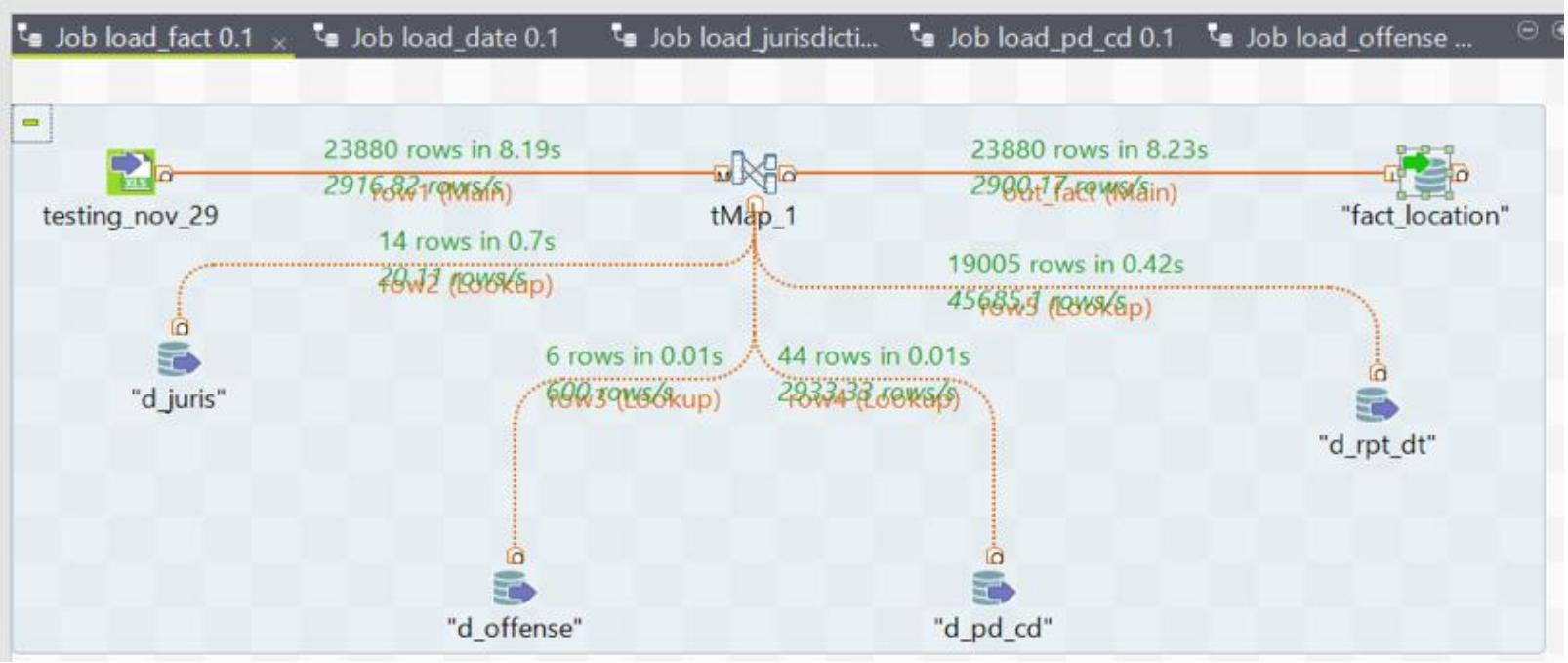
Result Grid | Filter Rows: | Edit:

D_PD_CD_KEY	D_PD_CD	D_PD_DESC
1	385	ROBBERY. PAYROLL
2	380	ROBBERY.CAR JACKING
3	392	ROBBERY.UNLICENSED FOR HIRE VEHICLE
4	224	BURGLARY.RESIDENCE.UNKNOWN TIM
5	221	BURGLARY.RESIDENCE.DAY
6	366	ROBBERY.BICYCLE
7	397	ROBBERY.OPEN AREA UNCLASSIFIED
8	243	BURGLARY.UNCLASSIFIED.NIGHT
9	396	ROBBERY.BEGIN AS SHOPLIFTING
10	371	ROBBERY.CHECK CASHING BUSINESS
11	384	ROBBERY.POCKETBOOK/CARRIED BAG
12	213	BURGLARY.COMMERCIAL.NIGHT
13	234	BURGLARY.UNKNOWN TIME
14	922	TRAFFIC.UNCLASSIFIED MISDEMEAN
15	360	ROBBERY.ATM LOCATION
16	389	ROBBERY.DWELLING
17	233	BURGLARY.TRUCK NIGHT
18	241	BURGLARY.UNCLASSIFIED.DAY
19	363	ROBBERY.BAR/RESTAURANT
20	395	ROBBERY.ON BUS/ OR BUS DRIVER
21	383	ROBBERY.DELIVERY PERSON

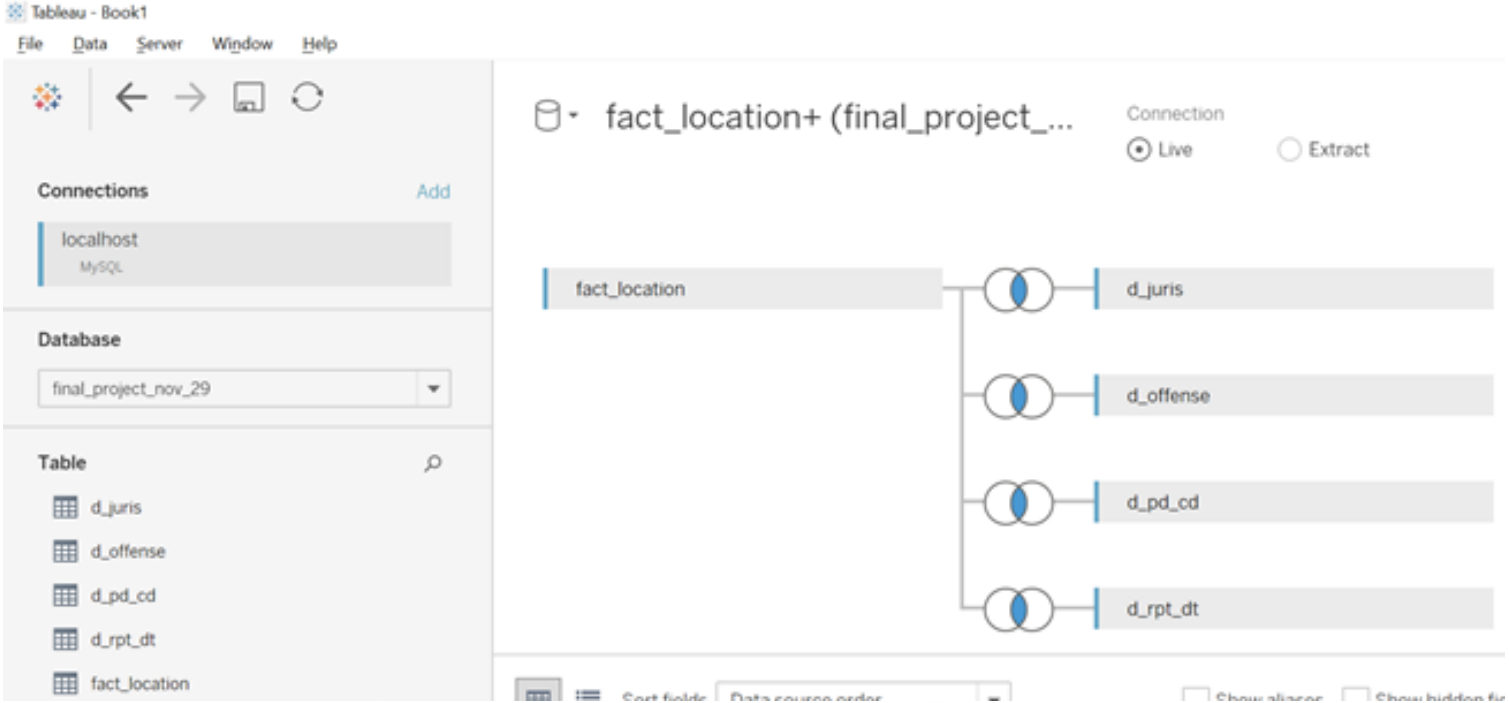
d_pd_cd 12

ETL Process

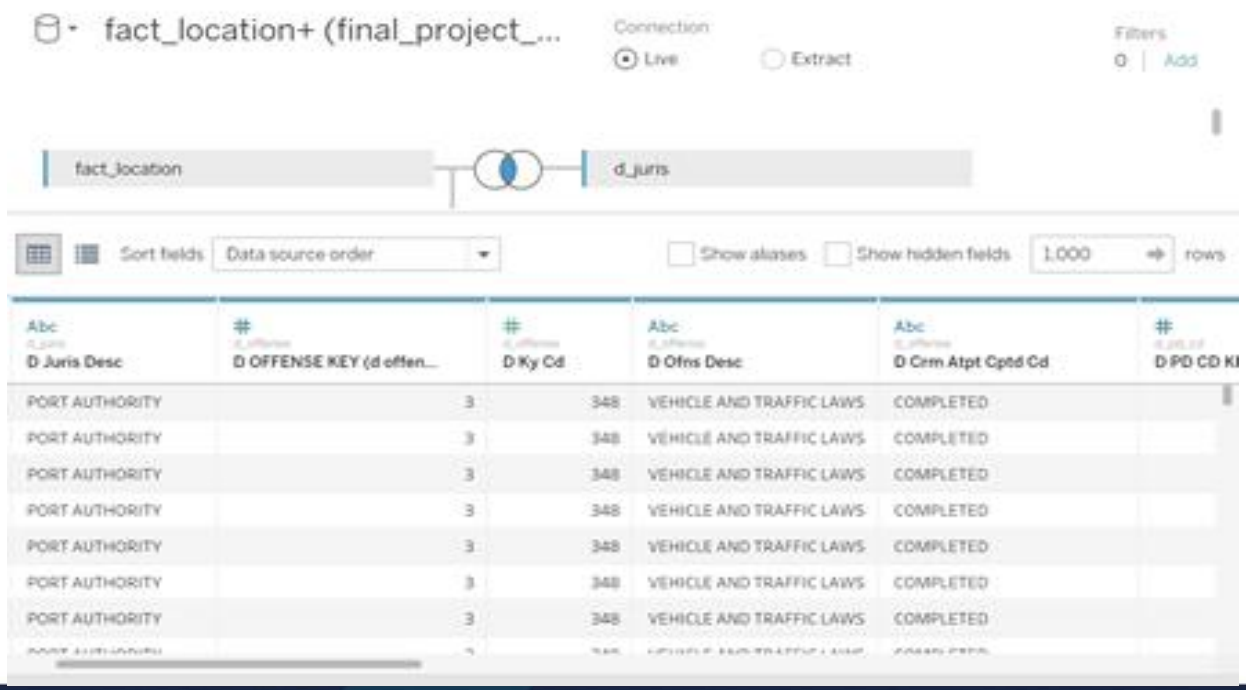
Job (Fact Table)



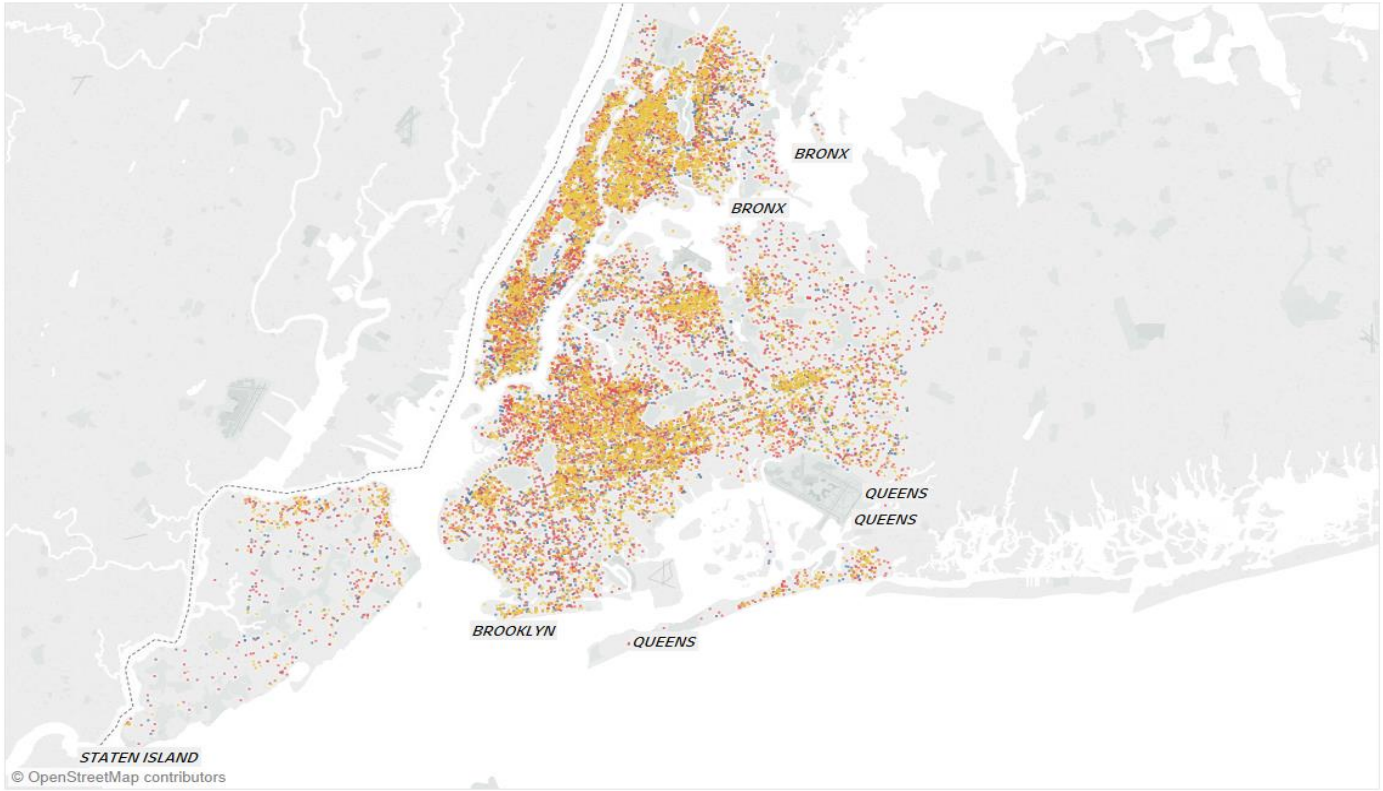
MySQL to Tableau Connection



MySQL to Tableau (Output)

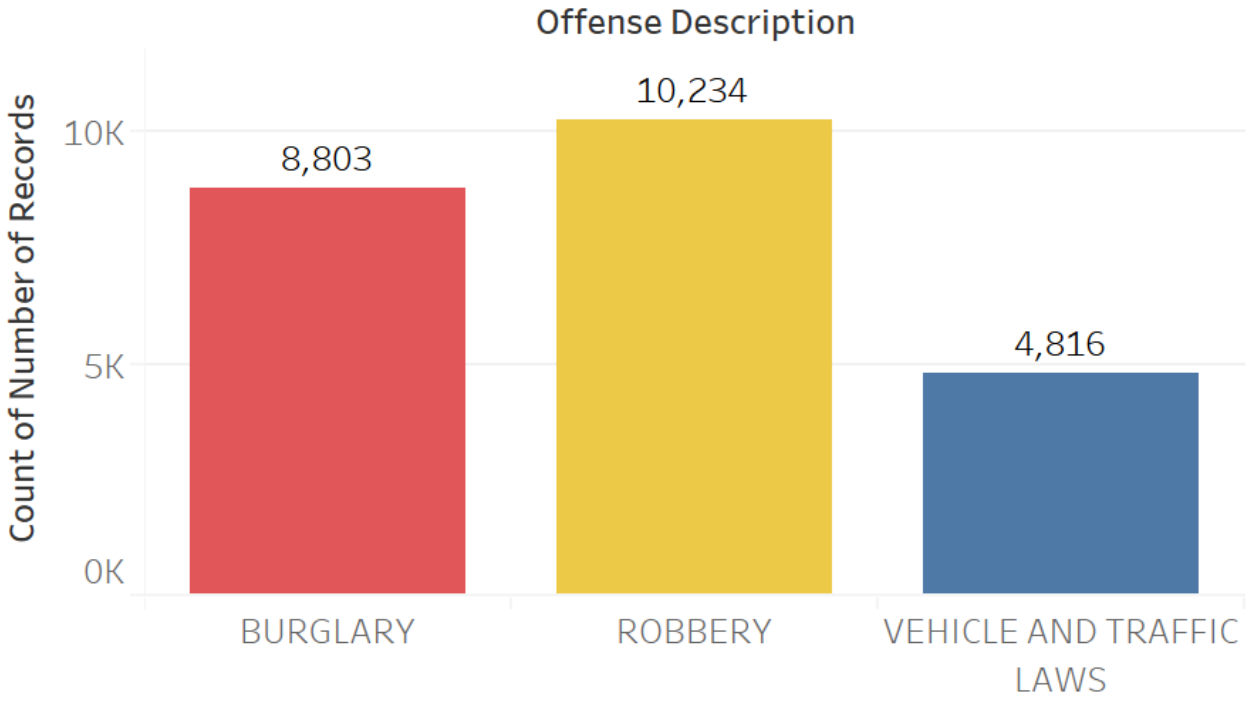


Crimes in New York across 2017

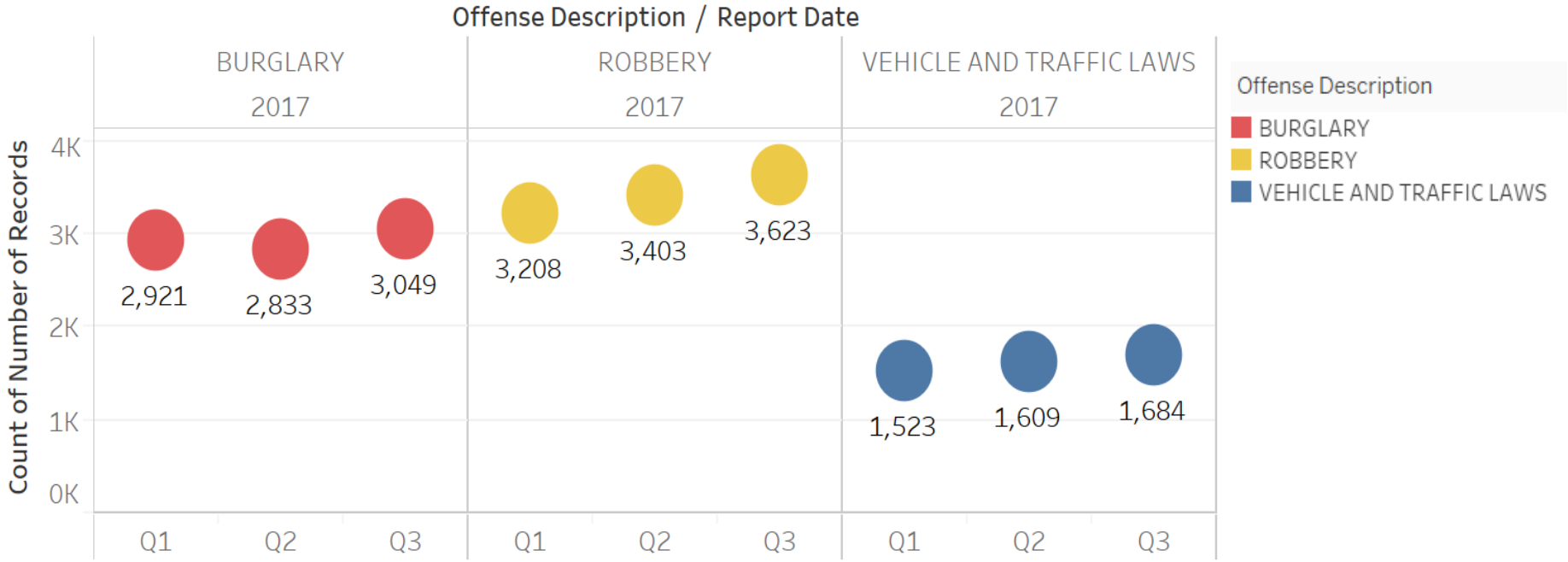


Offense Description
- BURGLARY
- ROBBERY
- VEHICLE AND TRAFFIC LAWS

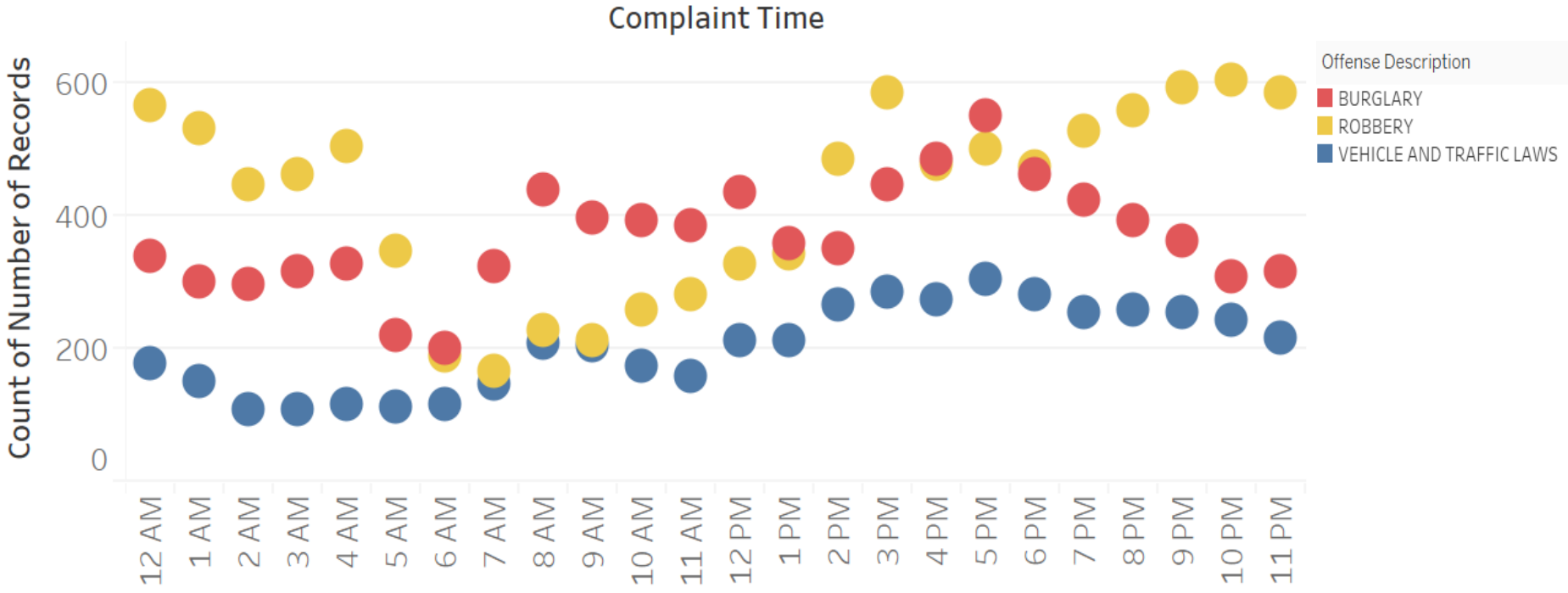
Type of Complaint NYPD - 2017 YTD



2017 Quarter-on-Quarter Complaints NYPD



Complaints across Time



Crimes Across Boroughs (2017)

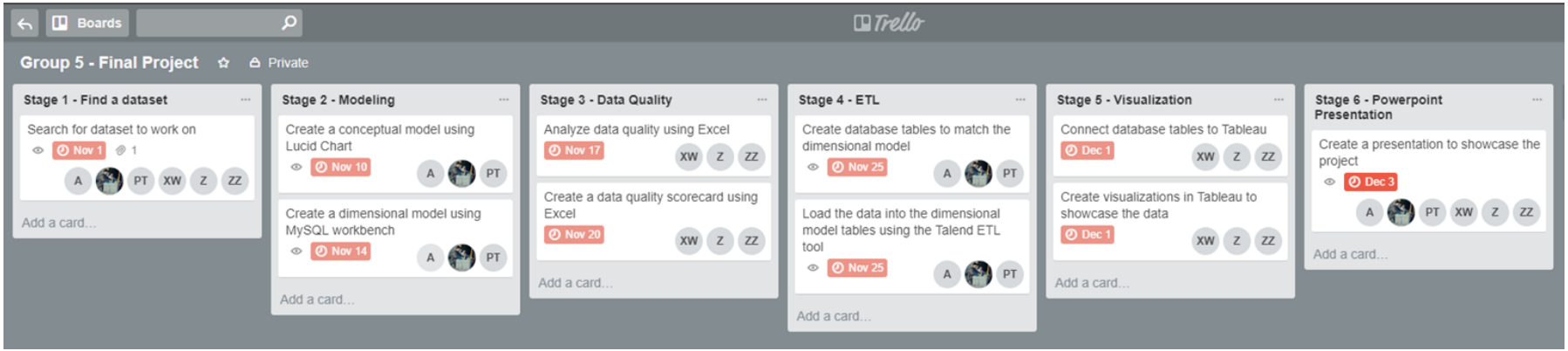
Offense Description	Borough				
	BROOKLYN	QUEENS	BRONX	MANHATTAN	STATEN ISLAND
BURGLARY	2,989	2,069	1,679	1,763	303
ROBBERY	3,276	2,093	2,587	2,027	251
VEHICLE AND TRAFFIC LAWS	1,313	1,406	1,147	825	125

Recommendations

- Currently, approximate locations are being captured by latitudes & longitudes. We recommend that additional information such as block number and street names also be captured to provide more accurate crime locations

- Our data shows that Brooklyn has higher crime rates than Staten Island, but this does not mean Staten Island is safer than Brooklyn because we have looked at the NYPD complaints data only and not at other factors such as population density, average income levels & employment rate

Our Trello Board



Thank You