SMATER COMMUTER

Nini Lin

A. Data source

The dataset is from NYC OpenData, covering reported crimes to the end of last year (2021). The format of the dataset is csv. This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD). The columns include CMPLNT_NUM, CMPLNT_FR_DT, CMPLNT_FR_TM, etc.

*The snippet of the original dataset is shown in Figure 1.

B. Data Cleaning

a. Uploading

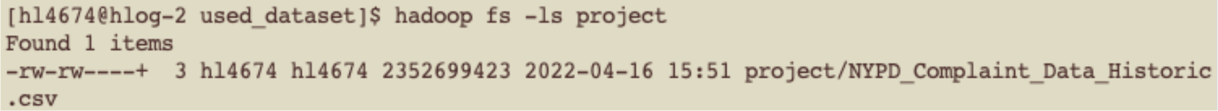
Uploading the dataset and the source code onto the peel cluster and Hadoop file system.

```
$ scp -r used_dataset hl4674@peel.hpc.nyu.edu:~/scratch/hl4674
```

```
$ hadoop fs -put /scratch/hl4674/ NYPD_Complaint_Data_Historic.csv
```

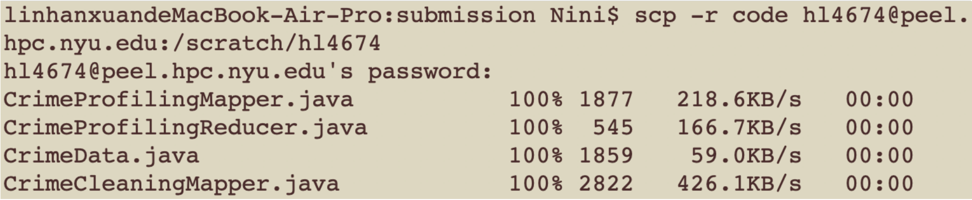
```
$ hadoop fs -mkdir project
```

```
$ hadoop fs -put NYPD_Complaint_Data_Historic.csv project
```

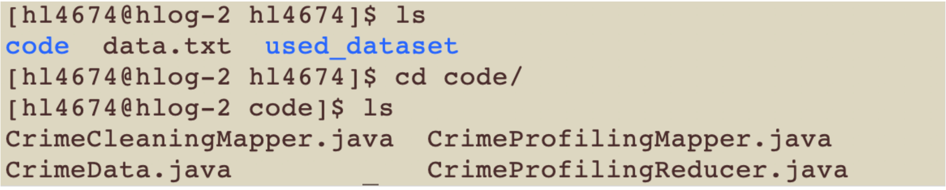


```
[hl4674@hlog-2 used_dataset]$ hadoop fs -ls project
Found 1 items
-rw-rw----+  3 hl4674 hl4674 2352699423 2022-04-16 15:51 project/NYPD_Complaint_Data_Historic
.csv
```

```
$ scp -r code hl4674@peel.hpc.nyu.edu:/scratch/hl4674
```



```
linhanxuandeMacBook-Air-Pro:submission Nini$ scp -r code hl4674@peel.
hpc.nyu.edu:/scratch/hl4674
hl4674@peel.hpc.nyu.edu's password:
CrimeProfilingMapper.java           100% 1877   218.6KB/s   00:00
CrimeProfilingReducer.java          100%  545   166.7KB/s   00:00
CrimeData.java                      100% 1859    59.0KB/s   00:00
CrimeCleaningMapper.java            100% 2822   426.1KB/s   00:00
```



```
[hl4674@hlog-2 hl4674]$ ls
code   data.txt   used_dataset
[hl4674@hlog-2 hl4674]$ cd code/
[hl4674@hlog-2 code]$ ls
CrimeCleaningMapper.java  CrimeProfilingMapper.java
CrimeData.java            CrimeProfilingReducer.java
```

b. Filtering

For analyzing, only selecting useful columns which are CMPLNT_FR_DT, CMPLNT_FR_TM, CMPLNT_TO_DT, CMPLNT_TO_TM, OFNS_DESC, PD_DESC, LAW_CAT_CD, BORO_NM, LOC_OF_OCCUR_DESC, PREM_TYP_DESC, PARKS_NM, X_COORD_CD, Y_COORD_CD, TRANSIT_DISTRICT, LATITUDE, LONGITUDE, PATROL_BORO, STATION_NAME.

Next, removing data with invalid datetime which CMPLNT_FR_DT and CMPLNT_FR_TM are later than CMPLNT_TO_DT and CMPLNT_TO_TM. Dataset after cleaning is written to `dataset` folder under `project` directory on HDFS.

*The snippet of dataset after cleaning is shown in Figure 2.

```
$ hadoop jar crimeData.jar CrimeData
/user/hl4674/project/NYPD_Complaint_Data_Historic.csv
/user/hl4674/project/dataset /user/hl4674/project/summary
```

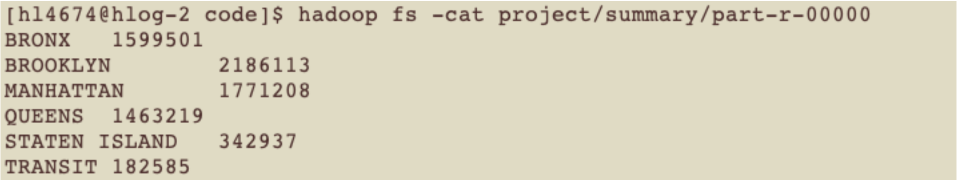
C. Data Profiling

a. Counting the number

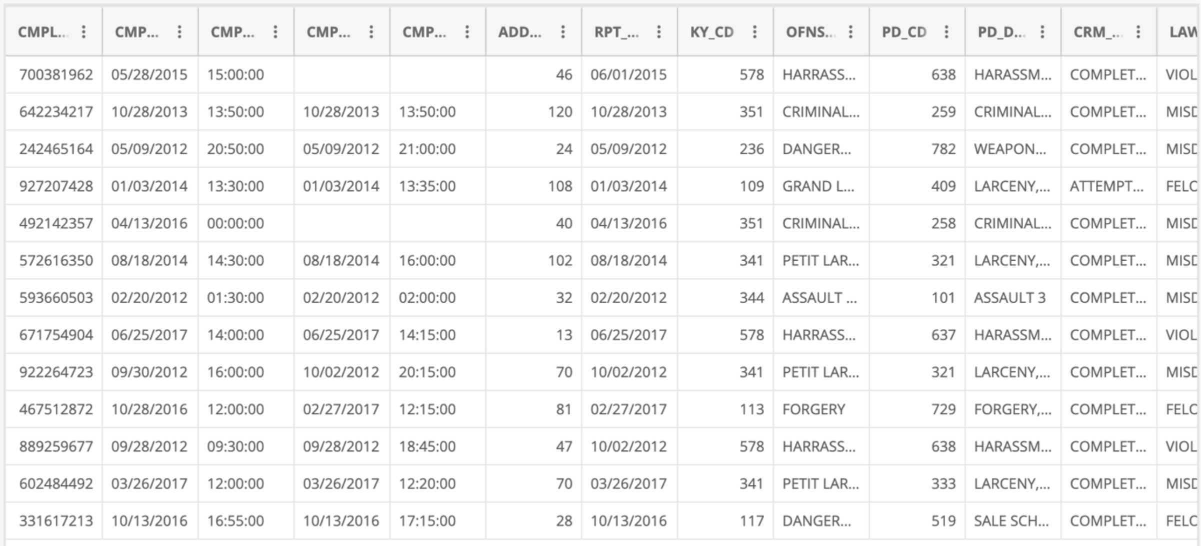
Counting the number of crime incidents based on BORO_NM column which has five categories, such as BRONX, BROOKLYN, MANHATTAN, QUEENS, STATEN ISLAND, and that of crime incidents occurring around transit stations.

The input file for data profiling is under `project/dataset`, and the profiling result is written to `summary` folder under `project` directory on HDFS.

```
$ hadoopm jar crimeData.jar CrimeData
/user/hl4674/project/NYPD_Complaint_Data_Historic.csv
/user/hl4674/project/dataset /user/hl4674/project/summary
```

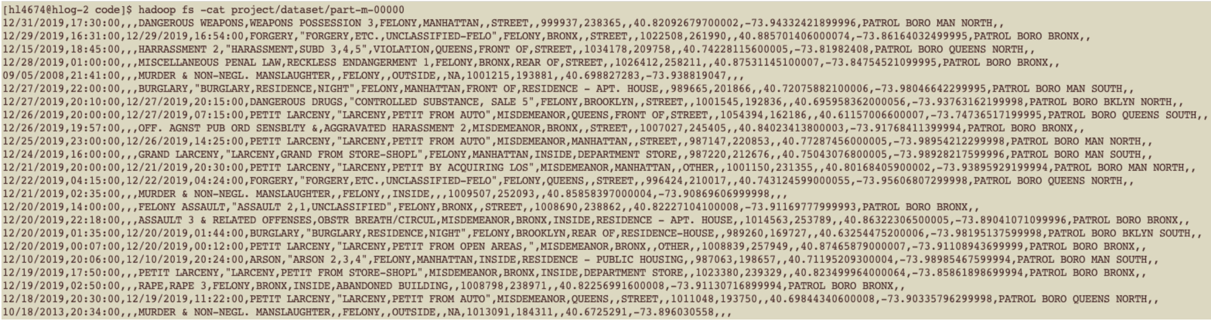


```
[hl4674@hlog-2 code]$ hadoop fs -cat project/summary/part-r-00000
BRONX   1599501
BROOKLYN        2186113
MANHATTAN       1771208
QUEENS  1463219
STATEN ISLAND   342937
TRANSIT 182585
```



CMPL…	CMP…	CMP…	CMP…	CMP…	ADD…	RPT_…	KY_CD	OFNS…	PD_CD	PD_D…	CRM_…	LAW
700381962	05/28/2015	15:00:00			46	06/01/2015	578	HARRASS…	638	HARASSM…	COMPLET…	VIOL
642234217	10/28/2013	13:50:00	10/28/2013	13:50:00	120	10/28/2013	351	CRIMINAL…	259	CRIMINAL…	COMPLET…	MISD
242465164	05/09/2012	20:50:00	05/09/2012	21:00:00	24	05/09/2012	236	DANGER…	782	WEAPON…	COMPLET…	MISD
927207428	01/03/2014	13:30:00	01/03/2014	13:35:00	108	01/03/2014	109	GRAND L…	409	LARCENY,…	ATTEMPT…	FELO
492142357	04/13/2016	00:00:00			40	04/13/2016	351	CRIMINAL…	258	CRIMINAL…	COMPLET…	MISD
572616350	08/18/2014	14:30:00	08/18/2014	16:00:00	102	08/18/2014	341	PETIT LAR…	321	LARCENY,…	COMPLET…	MISD
593660503	02/20/2012	01:30:00	02/20/2012	02:00:00	32	02/20/2012	344	ASSAULT …	101	ASSAULT 3	COMPLET…	MISD
671754904	06/25/2017	14:00:00	06/25/2017	14:15:00	13	06/25/2017	578	HARRASS…	637	HARASSM…	COMPLET…	VIOL
922264723	09/30/2012	16:00:00	10/02/2012	20:15:00	70	10/02/2012	341	PETIT LAR…	321	LARCENY,…	COMPLET…	MISD
467512872	10/28/2016	12:00:00	02/27/2017	12:15:00	81	02/27/2017	113	FORGERY	729	FORGERY,…	COMPLET…	FELO
889259677	09/28/2012	09:30:00	09/28/2012	18:45:00	47	10/02/2012	578	HARRASS…	638	HARASSM…	COMPLET…	VIOL
602484492	03/26/2017	12:00:00	03/26/2017	12:20:00	70	03/26/2017	341	PETIT LAR…	333	LARCENY,…	COMPLET…	MISD
331617213	10/13/2016	16:55:00	10/13/2016	17:15:00	28	10/13/2016	117	DANGER…	519	SALE SCH…	COMPLET…	FELO

Figure 1. Snippet of original dataset



```
[hl4674@hlog-2 code]$ hadoop fs -cat project/dataset/part-m-00000
12/31/2019,17:30:00,,,DANGEROUS WEAPONS,WEAPONS POSSESSION 3,FELONY,MANHATTAN,,STREET,,999937,238365,,40.82092679700002,-73.94332421899996,PATROL BORO MAN NORTH,,
12/29/2019,16:31:00,12/29/2019,16:54:00,FORGERY,"FORGERY,ETC.,UNCLASSIFIED-FELO",FELONY,BRONX,,STREET,,1022508,261990,,40.885701406000074,-73.86164032499995,PATROL BORO BRONX,,
12/15/2019,18:45:00,,,HARRASSMENT 2,"HARASSMENT,SUBD 3,4,5",VIOLATION,QUEENS,FRONT OF,STREET,,1034178,209758,,40.74228115600005,-73.81982408,PATROL BORO QUEENS NORTH,,
12/28/2019,01:00:00,,,MISCELLANEOUS PENAL LAW,RECKLESS ENDANGERMENT 1,FELONY,BRONX,REAR OF,STREET,,1026412,258211,,40.87531145100007,-73.84754521099995,PATROL BORO BRONX,,
09/05/2008,21:41:00,,,MURDER & NON-NEGL. MANSLAUGHTER,,FELONY,,OUTSIDE,,NA,1001215,193881,,40.698827283,-73.938819047,,,
12/27/2019,22:00:00,,,BURGLARY,"BURGLARY,RESIDENCE,NIGHT",FELONY,MANHATTAN,FRONT OF,RESIDENCE - APT. HOUSE,,989665,201866,,40.72075882100006,-73.98046642299995,PATROL BORO MAN SOUTH,,
12/27/2019,20:10:00,12/27/2019,20:15:00,DANGEROUS DRUGS,"CONTROLLED SUBSTANCE, SALE 5",FELONY,BROOKLYN,,STREET,,1001545,192836,,40.695958362000056,-73.93763162199998,PATROL BORO BKLYN NORTH,,
12/26/2019,20:00:00,12/27/2019,07:15:00,PETIT LARCENY,"LARCENY,PETIT FROM AUTO",MISDEMEANOR,QUEENS,FRONT OF,STREET,,1054394,162186,,40.61157006600007,-73.74736517199995,PATROL BORO QUEENS SOUTH,,
12/26/2019,19:57:00,,,OFF. AGNST PUB ORD SENSBLTY &,AGGRAVATED HARASSMENT 2,MISDEMEANOR,BRONX,,STREET,,1007027,245405,,40.84023413800003,-73.91768411399994,PATROL BORO BRONX,,
12/25/2019,23:00:00,12/26/2019,14:25:00,PETIT LARCENY,"LARCENY,PETIT FROM AUTO",MISDEMEANOR,MANHATTAN,,STREET,,987147,220853,,40.77287456000005,-73.98954212299998,PATROL BORO MAN NORTH,,
12/24/2019,16:00:00,,,GRAND LARCENY,"LARCENY,GRAND FROM STORE-SHOPL",FELONY,MANHATTAN,INSIDE,DEPARTMENT STORE,,987220,212676,,40.75043076800005,-73.98928217599996,PATROL BORO MAN SOUTH,,
12/21/2019,20:00:00,12/21/2019,20:30:00,PETIT LARCENY,"LARCENY,PETIT BY ACQUIRING LOS",MISDEMEANOR,MANHATTAN,,OTHER,,1001150,231355,,40.80168405900002,-73.93895929199994,PATROL BORO MAN NORTH,,
12/22/2019,04:15:00,12/22/2019,04:24:00,FORGERY,"FORGERY,ETC.,UNCLASSIFIED-FELO",FELONY,QUEENS,,STREET,,996424,210017,,40.743124599000055,-73.95606807299998,PATROL BORO QUEENS NORTH,,
12/21/2019,02:35:00,,,MURDER & NON-NEGL. MANSLAUGHTER,,FELONY,,INSIDE,,,1009507,252093,,40.85858397000004,-73.90869606999998,,,
12/20/2019,14:00:00,,,FELONY ASSAULT,"ASSAULT 2,1,UNCLASSIFIED",FELONY,BRONX,,STREET,,1008690,238862,,40.82227104100008,-73.91169777999993,PATROL BORO BRONX,,
12/20/2019,22:18:00,,,ASSAULT 3 & RELATED OFFENSES,OBSTR BREATH/CIRCUL,MISDEMEANOR,BRONX,INSIDE,RESIDENCE - APT. HOUSE,,1014563,253789,,40.86322306500005,-73.89041071099996,PATROL BORO BRONX,,
12/20/2019,01:35:00,12/20/2019,01:44:00,BURGLARY,"BURGLARY,RESIDENCE,NIGHT",FELONY,BROOKLYN,REAR OF,RESIDENCE-HOUSE,,989260,169727,,40.63254475200006,-73.98195137599998,PATROL BORO BKLYN SOUTH,,
12/20/2019,00:07:00,12/20/2019,00:12:00,PETIT LARCENY,"LARCENY,PETIT FROM OPEN AREAS,",MISDEMEANOR,BRONX,,OTHER,,1008839,257949,,40.87465879000007,-73.91108943699999,PATROL BORO BRONX,,
12/10/2019,20:06:00,12/10/2019,20:24:00,ARSON,"ARSON 2,3,4",FELONY,MANHATTAN,INSIDE,RESIDENCE - PUBLIC HOUSING,,987063,198657,,40.71195209300004,-73.98985467599994,PATROL BORO MAN SOUTH,,
12/19/2019,17:50:00,,,PETIT LARCENY,"LARCENY,PETIT FROM STORE-SHOPL",MISDEMEANOR,BRONX,INSIDE,DEPARTMENT STORE,,1023380,239329,,40.823499964000064,-73.85861898699994,PATROL BORO BRONX,,
12/19/2019,02:50:00,,,RAPE,RAPE 3,FELONY,BRONX,INSIDE,ABANDONED BUILDING,,1008798,238971,,40.82256991600008,-73.91130716899994,PATROL BORO BRONX,,
12/18/2019,20:30:00,12/19/2019,11:22:00,PETIT LARCENY,"LARCENY,PETIT FROM AUTO",MISDEMEANOR,QUEENS,,STREET,,1011048,193750,,40.69844340600008,-73.90335796299998,PATROL BORO QUEENS NORTH,,
10/18/2013,20:34:00,,,MURDER & NON-NEGL. MANSLAUGHTER,,FELONY,,OUTSIDE,,NA,1013091,184311,,40.6725291,-73.896030558,,,
```

Figure 2. Snippet of dataset after cleaning