

# MY INTERNSHIP PROJECT

INTERN: STOIAN ALIN-BOGDAN

27 AUG 2021

# TABLE OF CONTENTS

- GOALS
- MYSQL
- DBS & INGESTION W/ PYTHON
- CLOUD
- REPORTS

# GOALS

- USE AS MUCH KNOWLEDGE FROM THE THEORETICAL PART AS POSSIBLE (NOT DIRECTLY)
- LEARN AS MUCH AS POSSIBLE (DON'T FOCUS ON ONLY ONE THING)
- USE AS MUCH OF MY PREVIOUS KNOWLEDGE AS POSSIBLE

# MYSQL

## WHY MYSQL?

- SEE HOW I ADAPT TO A NEW DB
- VERY RELEVANT DB (HISTORICALLY AND TODAY) - LAMP STACK
- USEFUL FOR MIXED CASE DATABASES — LIKE MINE!
- MYISAM vs INNODB
- I LIKE IT



# MYSQL ON LINUX (WSL)

- DB MANAGEMENT & ADMINISTRATION IS MUCH EASIER ON LINUX!
- WSL 2.0 (UBUNTU 20.04 LTS)
- INSTALLATION IS TRIVIAL  
(LAMP STACK SUPPORT)

```
alstoian@EN617210:~/mysql$ sudo service --status-all
[sudo] password for alstoian:
[ - ] apache-htcacheclean
[ - ] apache2
[ - ] apparmor
[ ? ] apport
[ - ] atd
[ ? ] binfmt-support
[ - ] console-setup.sh
[ - ] cron
[ ? ] cryptdisks
[ ? ] cryptdisks-early
[ - ] dbus
[ ? ] hwclock.sh
[ + ] irqbalance
[ - ] iscsid
[ - ] keyboard-setup.sh
[ ? ] kmod
[ - ] lvm2
[ - ] lvm2-lvmpolld
[ - ] multipath-tools
[ + ] mysql
[ + ] open-iscsi
[ - ] open-vm-tools
[ ? ] plymouth
[ ? ] plymouth-log
[ - ] procps
[ - ] rsync
[ - ] rsyslog
[ - ] screen-cleanup
[ - ] ssh
[ ? ] ubuntu-fan
[ - ] udev
[ - ] ufw
[ - ] unattended-upgrades
[ - ] uuidd
[ - ] x11-common
```

```

[---] 7) 2021-08-19T06:05:16.369255Z 0 [System] [MY-010931] [Server] /usr/sbin/mysqld: ready for connections. Version: '8.0.26-0ubuntu0.20.04.2' socket: '/var/run/mysqld/mysqld.sock' port: 3306 (Ubuntu).
[---] 8) 2021-08-19T06:05:16.369019Z 0 [System] [MY-011323] [Server] X Plugin ready for connections. Bind-address: '127.0.0.1' port: 33060, socket: /var/run/mysqld/mysqlx.sock
[---] 9) 2021-08-18T12:34:38.659692Z 0 [System] [MY-010931] [Server] /usr/sbin/mysqld: ready for connections. Version: '8.0.26-0ubuntu0.20.04.2' socket: '/var/run/mysqld/mysqld.sock' port: 3306 (Ubuntu).
[---] 10) 2021-08-18T12:34:38.659625Z 0 [System] [MY-011323] [Server] X Plugin ready for connections. Bind-address: '127.0.0.1' port: 33060, socket: /var/run/mysqld/mysqlx.sock
[---] 1 shutdown(s) detected in /home/alstoian/mysql/error.log
[---] 1) 2021-08-18T12:34:33.739876Z 0 [System] [MY-010910] [Server] /usr/sbin/mysqld: Shutdown complete (mysqld 8.0.26-0ubuntu0.20.04.2) (Ubuntu).

----- Storage Engine Statistics -----
[---] Status: +ARCHIVE +BLACKHOLE +CSV -FEDERATED +InnoDB +MEMORY +MRG_MYISAM +MyISAM +PERFORMANCE_SCHEMA
[---] Data in InnoDB tables: 1.1G (Tables: 10)
[---] Data in MyISAM tables: 25.6M (Tables: 7)
[OK] Total fragmented tables: 0

----- Analysis Performance Metrics -----
[---] innodb_stats_on_metadata: OFF
[OK] No stat updates during querying INFORMATION_SCHEMA.

----- Security Recommendations -----
[---] Skipped due to unsupported feature for MySQL 8

----- CVE Security Recommendations -----
[OK] NO SECURITY CVE FOUND FOR YOUR VERSION

----- Performance Metrics -----
[---] Up for: 6h 9m 5s (98 q [0.004 qps], 31 conn, TX: 826K, RX: 26K)
[---] Reads / Writes: 98% / 2%
[---] Binary logging is enabled (GTID MODE: OFF)
[---] Physical Memory      : 6.1G
[---] Max MySQL memory     : 10.6G
[---] Other process memory: 0B
[---] Total buffers: 304.0M global + 70.0M per thread (151 max threads)
[---] P_S Max memory usage: 72B
[---] Galera GCache Max memory usage: 0B
[OK] Maximum reached memory usage: 584.0M (9.38% of installed RAM)
[!!!] Maximum possible memory usage: 10.6G (174.61% of installed RAM)
[!!!] Overall possible memory usage with other process exceeded memory
[OK] Slow queries: 2% (2/98)
[OK] Highest usage of available connections: 2% (4/151)
[!!!] Aborted connections: 3.23% (1/31)
[!!!] name resolution is active : a reverse name resolution is made for each new connection and can reduce performance
[---] Query cache have been removed in MySQL 8
[OK] Sorts requiring temporary tables: 0% (0 temp sorts / 270 sorts)
[!!!] Joins performed without indexes: 103
[OK] Temporary tables created on disk: 0% (0 on disk / 517 total)
[OK] Thread cache hit rate: 87% (4 created / 31 connections)
[OK] Table cache hit rate: 88% (6K hits / 7K requests)
[OK] table_definition_cache(615) is upper than number of tables(343)
[OK] Open file limit used: 0% (4/1K)

```

# MYSQLTUNER



key\_buffer -> default = 16M, recommended: innodb: 70% physical memory, myISAM: 25%.

-> 4G

innodb\_sort\_buffer\_size -> default = 10M -> 20M

sort\_buffer\_size -> default = 262k -> 2M

myisam\_sort\_buffer\_size -> default = 8M -> 16M

read\_buffer\_size -> default = 131k -> 1M

read\_rnd\_buffer\_size -> default = 262k -> 1M

join\_buffer\_size -> default = 262k -> 2M

bulk\_insert\_buffer\_size -> default = 8M -> 16M

slow\_query\_log -> default = OFF -> ON

slow\_query\_log\_file -> default = /var/lib/mysql/\$USER-slow.log ->

~/mysql/slowqueries.log

long\_query\_time -> default = 10s -> 5s

log\_error -> default = /var/log/mysql/error.log -> /home/alstoian/mysql/error.log

general\_log\_file -> default = /var/lib/mysql/USER.log ->

/home/alstoian/mysql/general.log

log\_slow\_extra -> default = OFF -> ON

innodb\_buffer\_pool\_size -> default = 128M -> 256M

innodb\_flush\_log\_at\_trx\_commit -> default = 1 -> 2

innodb\_thread\_concurrency -> default = 0 -> 8

..AND MANY OTHERS  
PERSISTED VARS ->

Data Internship

```
root@EN617210:/var/lib/mysql# python3 -m json.tool mysqld-auto.cnf
{
  "version": 1,
  "mysql_server": {
    "innodb_thread_concurrency": {
      "Value": "8",
      "Metadata": {
        "Timestamp": 1628871726000323,
        "User": "alstoian",
        "Host": "localhost"
      }
    },
    "general_log": {
      "Value": "ON",
      "Metadata": {
        "Timestamp": 1628751661162599,
        "User": "alstoian",
        "Host": "localhost"
      }
    },
    "log_slow_extra": {
      "Value": "ON",
      "Metadata": {
        "Timestamp": 1628752338900576,
        "User": "alstoian",
        "Host": "localhost"
      }
    },
    "slow_query_log": {
      "Value": "ON",
      "Metadata": {
        "Timestamp": 1628680061923515,
        "User": "alstoian",
        "Host": "localhost"
      }
    },
    "key_buffer_size": {
      "Value": "3999997952",
      "Metadata": {
        "Timestamp": 1628674647243263,
        "User": "alstoian",
        "Host": "localhost"
      }
    },
    "long_query_time": {
      "Value": "5.000000",
      "Metadata": {
        "Timestamp": 1628680327032619,
```

# MY DATASET

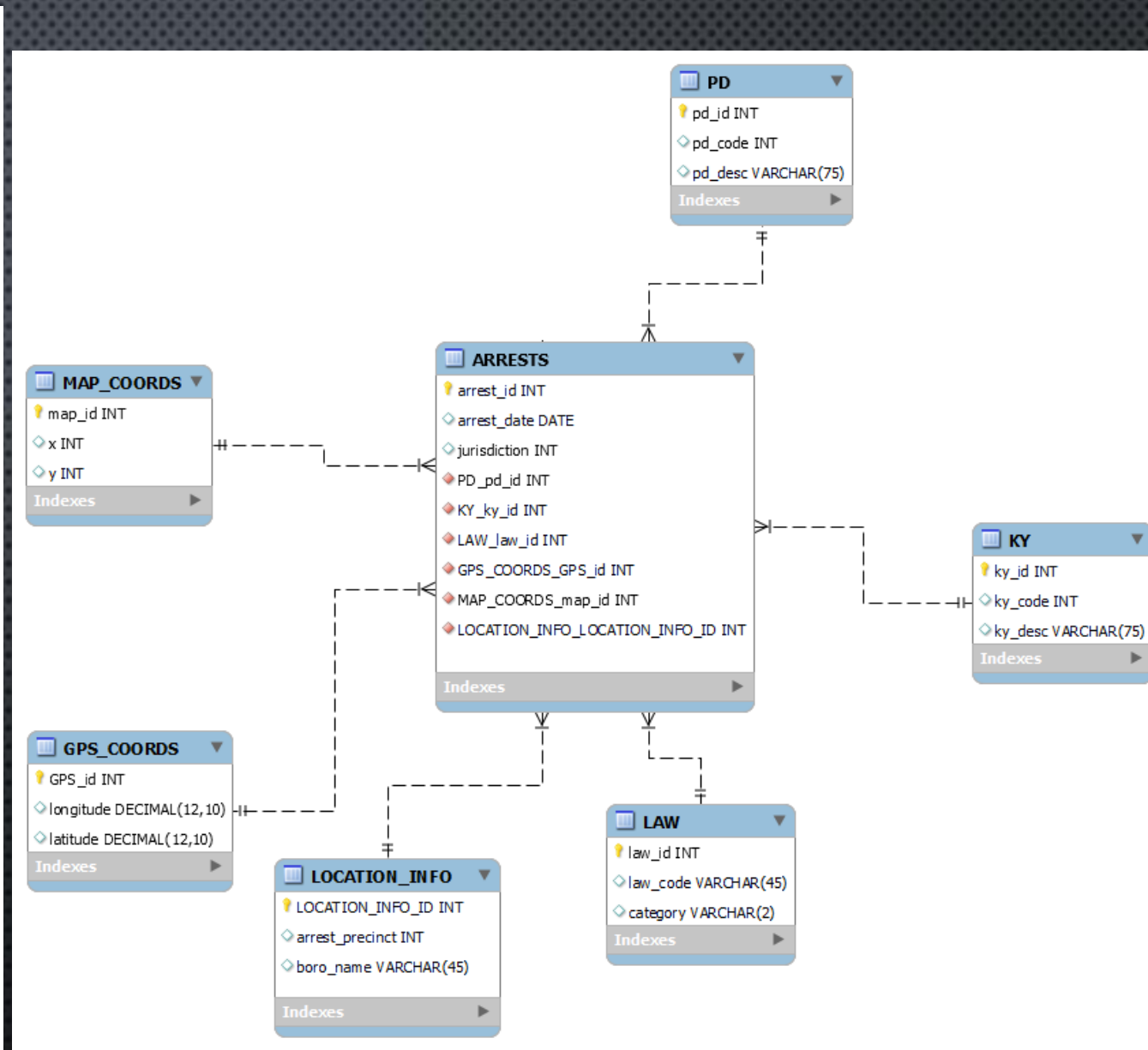
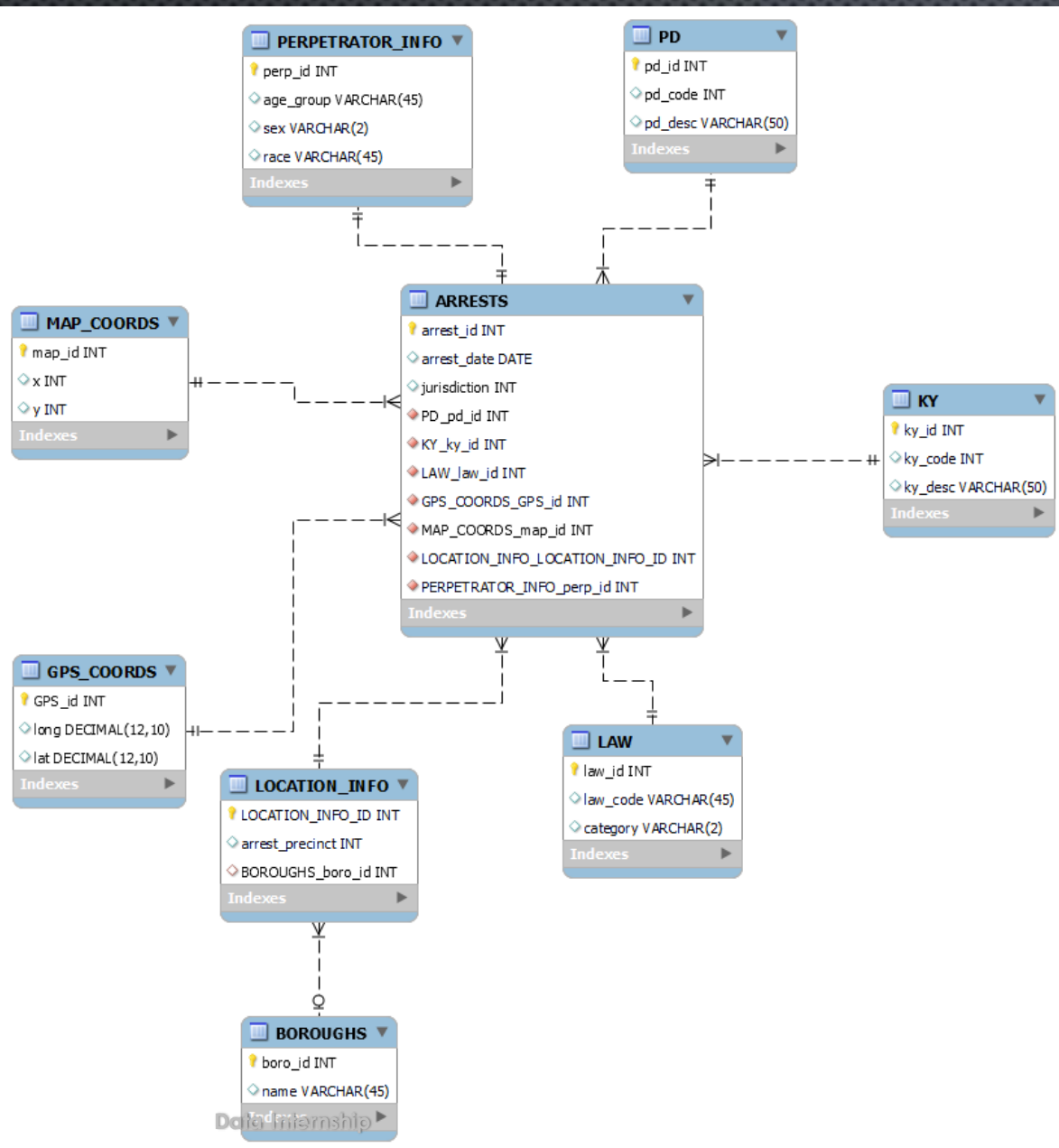
- NEW YORK ARRESTS 2006-2020
- [HTTPS://DATA.CITYOFNEWYORK.US/PUBLIC-SAFETY/NYPD-ARRESTS-DATA-HISTORIC-/8H9B-RP9U](https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u)
- 5.2 MILLION ROWS, 19 COLUMNS -> NEEDS NORMALIZATION!



# DB DESIGN

- WHY TWO DBs?
- MAIN, LIVE, REPORTING, STATIC — ONLY FROM THE LAST YEAR (ETL!)
- NO RELATIONS ON REPORTING?

MYISAM!



Copyright (c) 2000, 2021, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

```
mysql> use main;
```

Reading table information for completion of table and column names  
You can turn off this feature to get a quicker startup with -A

Database changed

```
mysql> show tables;
```

```
+-----+
| Tables_in_main |
+-----+
| ARRESTS        |
| BOROUGHs       |
| GPS_COORDS     |
| KY              |
| LAW            |
| LOCATION_INFO  |
| MAP_COORDS     |
| PD              |
| PERPETRATOR_INFO |
+-----+
```

9 rows in set (0.00 sec)

```
mysql> use report;
```

Reading table information for completion of table and column names  
You can turn off this feature to get a quicker startup with -A

Database changed

```
mysql> show tables;
```

```
+-----+
| Tables_in_report |
+-----+
| ARRESTS          |
| GPS_COORDS       |
| KY               |
| LAW              |
| LOCATION_INFO    |
| MAP_COORDS       |
| PD               |
+-----+
```

7 rows in set (0.00 sec)

```
mysql> _
```

Properties ER Diagram

GPS_COORDS
123 GPS_id
123 longitude
123 latitude

KY
123 ky_id
123 ky_code
ABC ky_desc

LAW
123 law_id
ABC law_code
ABC category

LOCATION_INFO
123 LOCATION_INFO_ID
123 arrest_precinct
ABC boro_name

MAP_COORDS
123 map_id
123 x
123 y

PD
123 pd_id
123 pd_code
ABC pd_desc

ARRESTS
123 arrest_id
arrest_date
123 jurisdiction
ABC age_group
ABC sex
ABC race
123 PD_pd_id
123 KY_ky_id
123 LAW_law_id
123 GPS_COORDS_GPS_id
123 MAP_COORDS_map_id
123 LOCATION_INFO_LOCATION_INFO_ID

# NO RELATIONS ON REPORTING DB BECAUSE OF MYISAM



# DATA INGESTION

- JUPYTER NOTEBOOK
- PYTHON
- PANDAS

DATAFRAMES MAKE IT EASY TO SPLIT THE MASSIVE INPUT .CSV

- HELPER FUNCTION TO GENERATE INSERT STATEMENTS (FOR STRESS TESTING, CURIOSITY, TESTING VARIABLES, REAL LIFE PERFORMANCE, ETC.)
- NO DIRECT CONNECTION TO THE DATABASE (OVERHEAD!)

```
In [2]: df = pd.read_csv(r"C:\Users\alstoian\Scripts\NYPD_Arrests_Historic_Final.csv")
```

```
In [3]: df
```

```
Out[3]:
```

	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DESC	LAW_CODE	LAW_CAT_CD	ARREST_BORO	ARREST_F
0	192799737	01/26/2019	177.0	SEXUAL ABUSE	116.0	SEX CRIMES	PL 1306503	F	M	25
1	149117452	01/06/2016	153.0	RAPE 3	104.0	RAPE	PL 1302503	F	K	67
2	190049060	11/15/2018	157.0	RAPE 1	104.0	RAPE	PL 1303501	F	K	77
3	24288194	09/13/2006	203.0	TRESPASS 3, CRIMINAL	352.0	CRIMINAL TRESPASS	PL 140100E	M	K	77
4	189182271	10/24/2018	153.0	RAPE 3	104.0	RAPE	PL 1302503	F	M	5
...	...	...	...	...	...	...	...	...	...	...
5125601	207601040	01/08/2020	273.0	TAMPERING 1, CRIMINAL	121.0	CRIMINAL MISCHIEF & RELATED OF	PL 1452000	F	M	1
5125602	206891807	01/01/2020	113.0	MENACING, UNCLASSIFIED	344.0	ASSAULT 3 & RELATED OFFENSES	PL 1201401	M	K	90
5125603	207760542	01/11/2020	339.0	LARCENY, PETIT FROM OPEN AREAS,	341.0	PETIT LARCENY	PL 1552500	M	M	13
5125604	206896678	01/01/2020	105.0	STRANGULATION 1ST	106.0	FELONY ASSAULT	PL 1211200	F	Q	111
5125605	206908101	01/02/2020	782.0	WEAPONS, POSSESSION, ETC	236.0	DANGEROUS WEAPONS	PL 2650101	M	Q	115

5125606 rows × 18 columns

This is what was left after data  
cleaning

```
In [143]: boroughs = df['ARREST_BORO'].unique();
boroughs_dict = {'boro_id':[x for x in range(0, len(boroughs))], 'name': boroughs}
print(boroughs_dict)
boroughs_df = pd.DataFrame(boroughs_dict)
boroughs_df
```

```
Out[143]: {'boro_id': [0, 1, 2, 3, 4], 'name': array(['K', 'M', 'Q', 'B', 'S'], dtype=object)}
```

	boro_id	name
0	0	K
1	1	M
2	2	Q
3	3	B
4	4	S

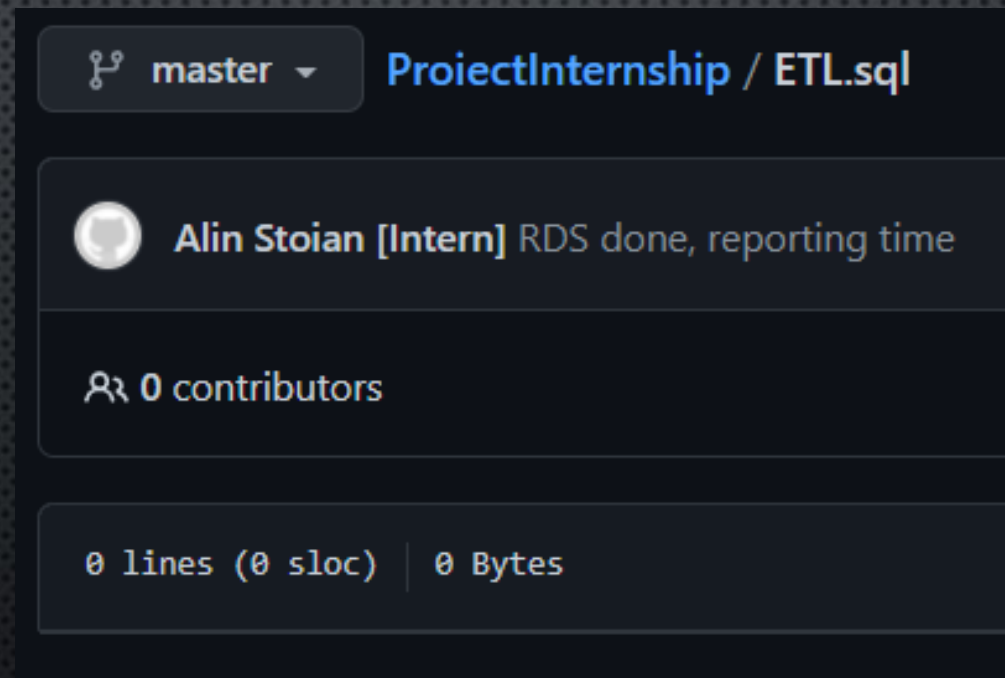
```
In [6]: #Functie helper pentru generare de fisiere .sql cu INSERT statement-uri
def SQL_INSERT_STATEMENT_FROM_DATAFRAME(SOURCE, TARGET, FILE):
    sql_texts = []
    fout = open(FILE, 'a')
    fout.truncate(0)
    for index, row in SOURCE.iterrows():
        sql_texts.append('INSERT INTO '+TARGET+' ('+ str(', ').join(SOURCE.columns)+ ') VALUES '+ str(tuple(row.values))+';')
    for item in sql_texts:
        fout.write(str(item)+'\n')
    fout.close()
```

```
In [137]: SQL_INSERT_STATEMENT_FROM_DATAFRAME(boroughs_df, 'BOROUGHs', 'insert_boroughs.sql')
1 INSERT INTO BOROUGHs (boro_id, name) VALUES (0, 'M');
2 INSERT INTO BOROUGHs (boro_id, name) VALUES (1, 'K');
3 INSERT INTO BOROUGHs (boro_id, name) VALUES (2, 'B');
4 INSERT INTO BOROUGHs (boro_id, name) VALUES (3, 'Q');
5 INSERT INTO BOROUGHs (boro_id, name) VALUES (4, 'S');
6
```



# ETL FROM MAIN TO REPORTING

- MISTAKES HAPPEN...



# FINAL RESULT

```
mysql> SELECT table_schema AS "Database",  
ROUND(SUM(data_length + index_length) / 1024 / 1024, 2) AS "Size (MB)"  
-> FROM information_schema.TABLES  
-> GROUP BY table_schema;
```

Database	Size (MB)
mysql	2.63
information_schema	0.00
performance_schema	0.00
sys	0.02
main	1110.89
report	25.57

6 rows in set (0.20 sec)

# CLOUD

- SCRIPT TO BACKUP REPORTING DATABASE — CRONTABBED
- BACKUP TO EC2 INSTANCE
- EC2 INSTANCE TO AWS RDS MySQL - CRONTABBED



```
#!/bin/bash
if test -f "./backup.sql.gz"; then
    rm ./backup.sql.gz
fi

mysqldump --user=alstoian --password=Admin1234 --add-drop-table --databases report --verbose > backup.sql

if test -f "./backup.sql"; then
    gzip ./backup.sql
    if test -f "./mykey2.pem"; then
        scp -i mykey2.pem backup.sql.gz ubuntu@ec2-3-121-184-109.eu-central-1.compute.amazonaws.com:/home/ubuntu
    else
        echo "no key found"
    fi
else
    echo "backup failed!"
fi
```

<- ON PREM

```
#!/bin/bash

if test -f "./backup.sql.gz"; then
    gunzip backup.sql.gz
else
    echo "no backup archive found!"
fi

pass=`cat /usr/local/etc/my-pass`

if test -f "./backup.sql"; then
    mysql --verbose -h mysql-rds.ctr9p26yyt8.eu-central-1.rds.amazonaws.com --user=alstoian --password=$pass report < backup.sql
else
    echo "sql script not found!"
fi
```

<- EC2

Instance summary for i-015d0c6f173c6ff4d [Info](#)

Updated less than a minute ago

Connect

Instance state ▼

Actions ▼

Instance ID i-015d0c6f173c6ff4d	Public IPv4 address 3.121.184.109   <a href="#">open address</a>	Private IPv4 addresses 172.30.1.73
IPv6 address –	Instance state Running	Public IPv4 DNS ec2-3-121-184-109.eu-central-1.compute.amazonaws.com   <a href="#">open address</a>
Private IPv4 DNS ip-172-30-1-73.eu-central-1.compute.internal	Instance type t2.micro	Elastic IP addresses –
VPC ID vpc-0daabc94d219e9f23	AWS Compute Optimizer finding Opt-in to AWS Compute Optimizer for recommendations.   <a href="#">Learn more</a>	IAM Role –
Subnet ID subnet-011fd848aeb044e13		

- Details
- Security**
- Networking
- Storage
- Status checks
- Monitoring
- Tags

▼ Security details

IAM Role –	Owner ID 447227867751	Launch time Thu Aug 19 2021 14:11:53 GMT+0300 (Eastern European Summer Time)
Security groups sg-0e8f678afe581f30b (finalfinalmysqlSec)		

Details

Security group name

finalfinalmysqlSec

Security group ID

sg-0e8f678afe581f30b

Description

Created by RDS management console

VPC ID

vpc-0daabc94d219e9f23

Owner

447227867751

Inbound rules count

6 Permission entries

Outbound rules count

1 Permission entry

- Inbound rules
- Outbound rules
- Tags

You can now check network connectivity with Reachability Analyzer

Run Reachability Analyzer

Inbound rules (6)

Manage tags

Edit inbound rules

Filter security group rules

<

1

>

Security group rule...	IP version	Type	Protocol	Port range	Source	Description
sgr-0a2f1355cd3e88658	IPv4	SSH	TCP	22	81.180.208.111/32	ssh-for-ec2
sgr-035dd62bdf550e7...	IPv4	MYSQL/Aurora	TCP	3306	172.30.1.177/32	ec2-access
sgr-04a778df2121466...	IPv4	MYSQL/Aurora	TCP	3306	172.29.98.86/32	wsl-access
sgr-0c4e7d03ae37a4648	IPv4	MYSQL/Aurora	TCP	3306	81.180.209.187/32	my-access
sgr-06d11h3413eh7h	IPv4	MYSQL /Aurora	TCP	3306	81.180.208.111/32	my-access2



Summary

DB identifier mysql-rds	CPU <div><div></div></div> 2.46%	Status <div>✔ Available</div>	Class db.t2.micro
Role Instance	Current activity <div><div></div></div> 0 Connections	Engine MySQL Community	Region & AZ eu-central-1a

Connectivity & security

Monitoring

Logs & events

Configuration

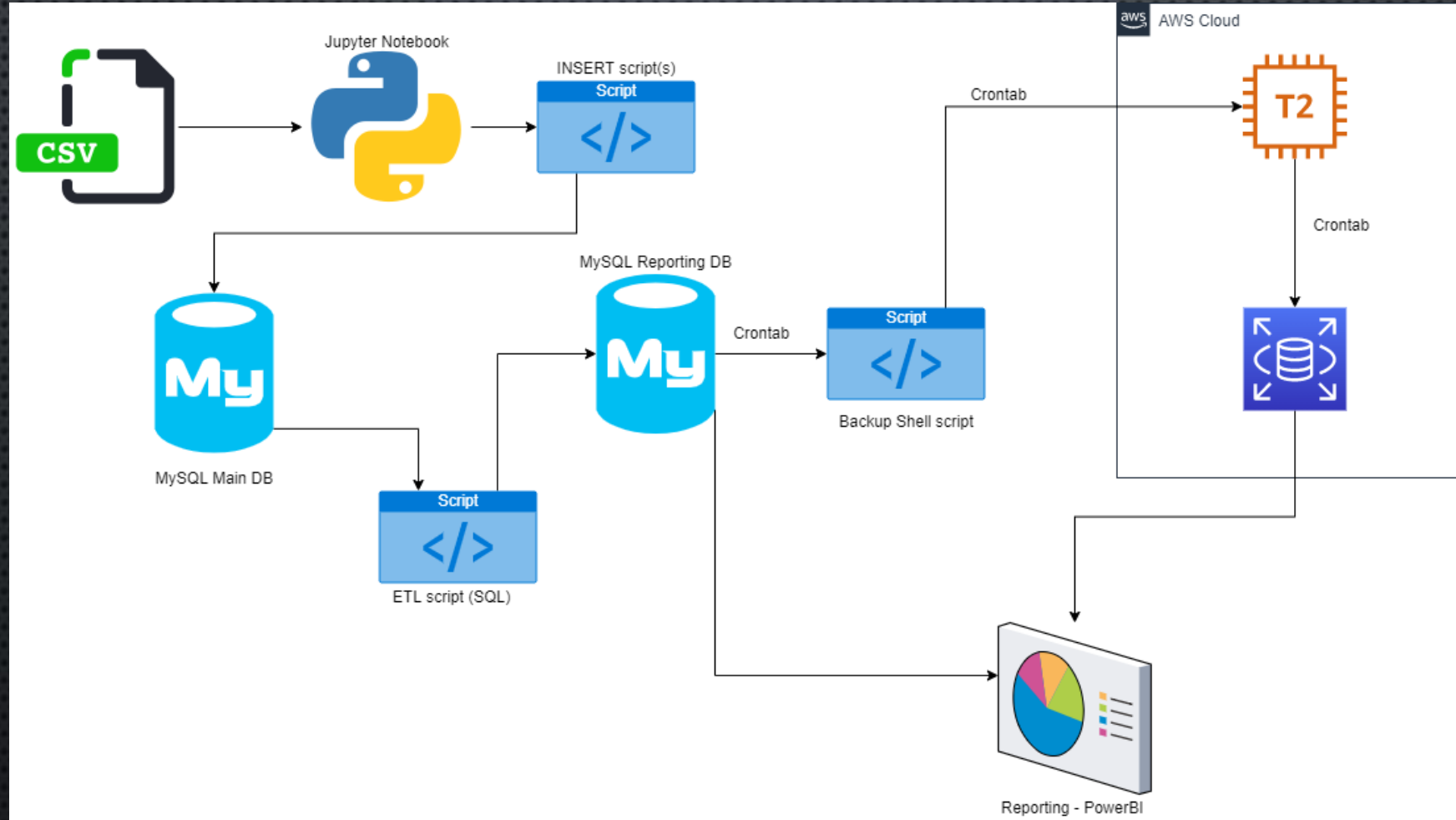
Maintenance & backups

Tags

Connectivity & security

Endpoint & port	Networking	Security
Endpoint mysql-rds.ctr9p26yyt8.eu-central-1.rds.amazonaws.com	Availability zone eu-central-1a	VPC security groups <a href="#">finalfinalmysqlSec (sg-0e8f678afe581f30b)</a> ( active )
Port 3306	VPC <a href="#">vpc-0daabc94d219e9f23</a>	Public accessibility Yes
	Subnet group default-vpc-0daabc94d219e9f23	Certificate authority rds-ca-2019
	Subnets <a href="#">subnet-074abd267759e06db</a> <a href="#">subnet-0c9dbc70ccba5334b</a> <a href="#">subnet-011fd848aeb044e13</a>	Certificate authority date August 22, 2024, 08:08 (UTC±8:08)

# MY PIPELINE



# REPORTING

- POWERBI
- DATASET WAS MADE FOR ANALYSIS
- M SCRIPT & DAX



arrest_id	arrest_date	jurisdiction	age_group	sex	race	PD_pd_id	LAW_law_id	MAP_COORDS_map_id	LOCATION_INFO_LOCATION_INFO_ID	ky_code	ky_desc	pd_code	pd_desc
206912771	Friday, January 3, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
207512046	Tuesday, January 7, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
207719540	Friday, January 10, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208014232	Thursday, January 16, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208044083	Thursday, January 16, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208163906	Sunday, January 19, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208368120	Wednesday, January 22, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208617657	Tuesday, January 28, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208634048	Tuesday, January 28, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208858484	Saturday, February 1, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
208975610	Tuesday, February 4, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209249934	Sunday, February 9, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209264241	Sunday, February 9, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209317952	Monday, February 10, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209367025	Tuesday, February 11, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209491080	Thursday, February 13, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
209693583	Tuesday, February 18, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210123664	Wednesday, February 26, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210123669	Wednesday, February 26, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210764614	Tuesday, March 10, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210770665	Tuesday, March 10, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210809281	Tuesday, March 10, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
210910144	Thursday, March 12, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
211807735	Tuesday, April 7, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
211807737	Tuesday, April 7, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
212182205	Monday, April 20, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL
212218504	Tuesday, April 21, 2020	0	25-44	M	BLACK	1	18	427	13	195	COERCION 2	101	ASSAUL

# DAY\_OF\_WEEK?

```
1 Day_of_Week = FORMAT('TableForReports'[arrest_date], "dddd")
```

## And other measures

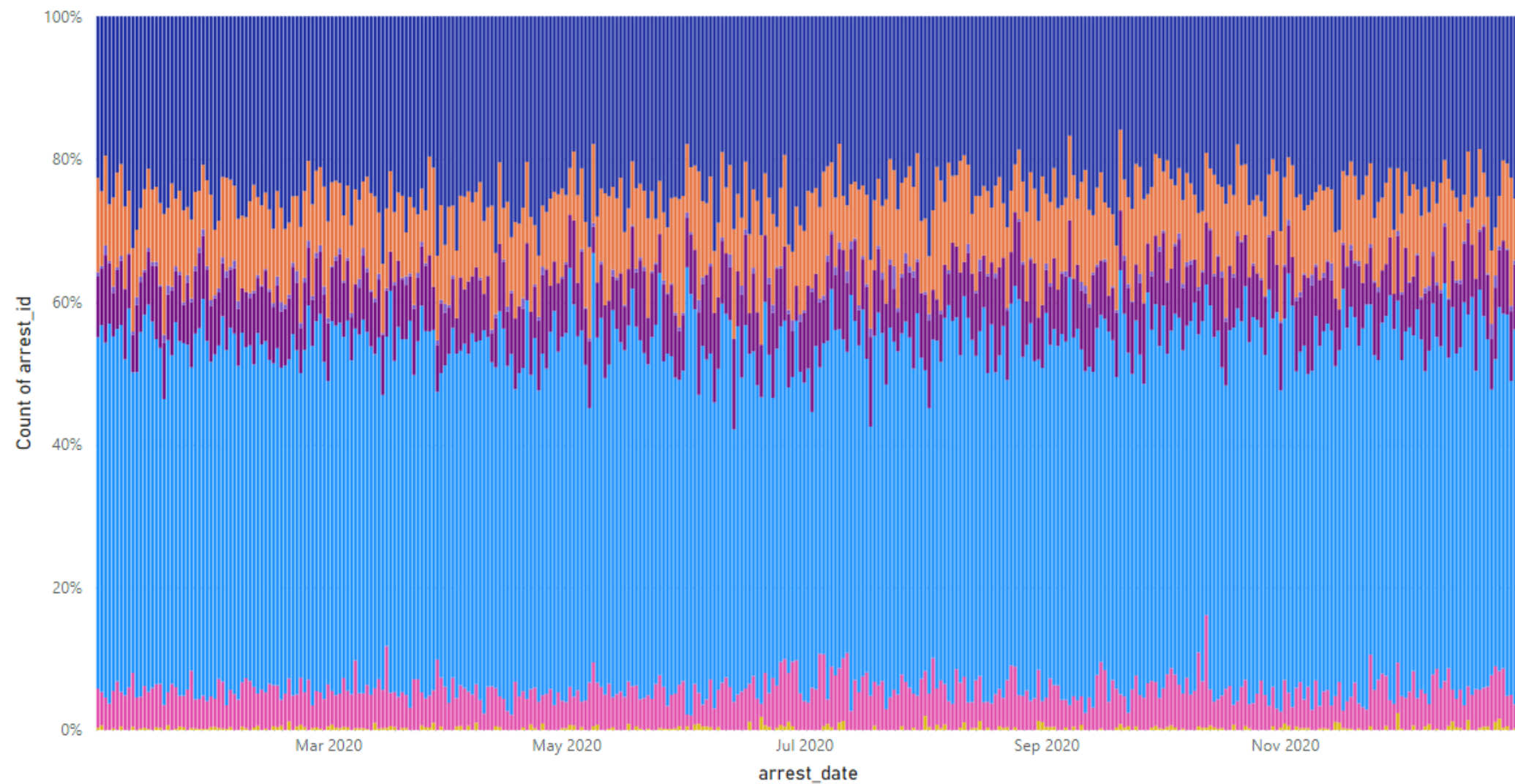
```
1 Count of law_code total for category 2 =
2 CALCULATE(
3     DISTINCTCOUNT('TableForReports'[law_code]),
4     ALLSELECTED('TableForReports'[category])
5 )
```

```
1 Count of ky_desc total for category =
2 CALCULATE(
3     DISTINCTCOUNT('TableForReports'[ky_desc]),
4     ALLSELECTED('TableForReports'[category])
5 )
```

pd_desc	longitude	latitude	law_code	category	map_x	map_y	arrest_precinct	boro_name	Day_of_Week
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Friday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Friday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Thursday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Thursday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Sunday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Wednesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Saturday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Sunday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Sunday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Monday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Thursday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Wednesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Wednesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Thursday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Monday
ASSAULT 3	-73.895824354	40.816087661	PL 1654000	M	999966	163346	47	B	Tuesday

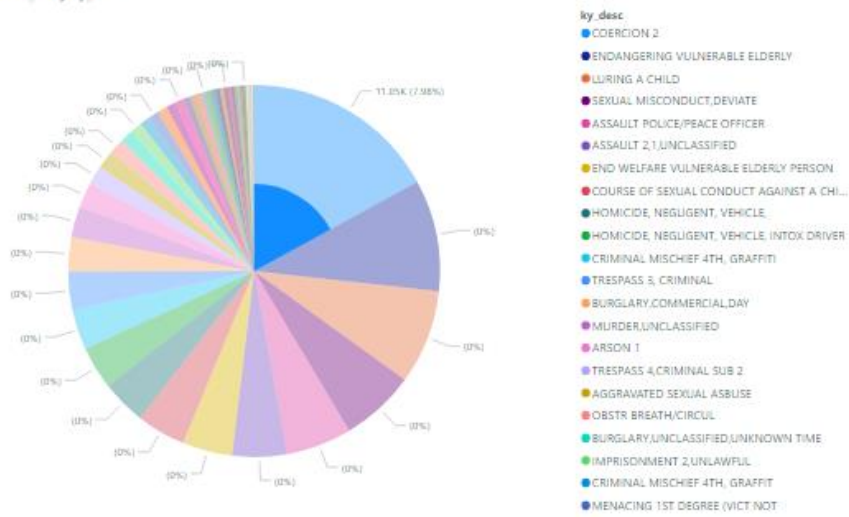
Count of arrest\_id by arrest\_date and race

race ● AMERICAN INDIAN/ALASKAN NATIVE ● ASIAN / PACIFIC ISLANDER ● BLACK ● BLACK HISPANIC ● UNKNOWN ● WHITE ● WHITE HISPANIC

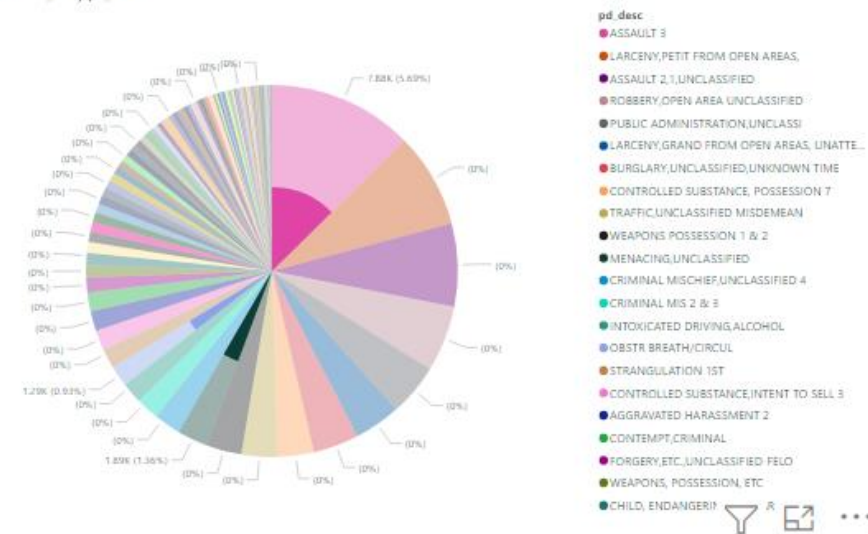




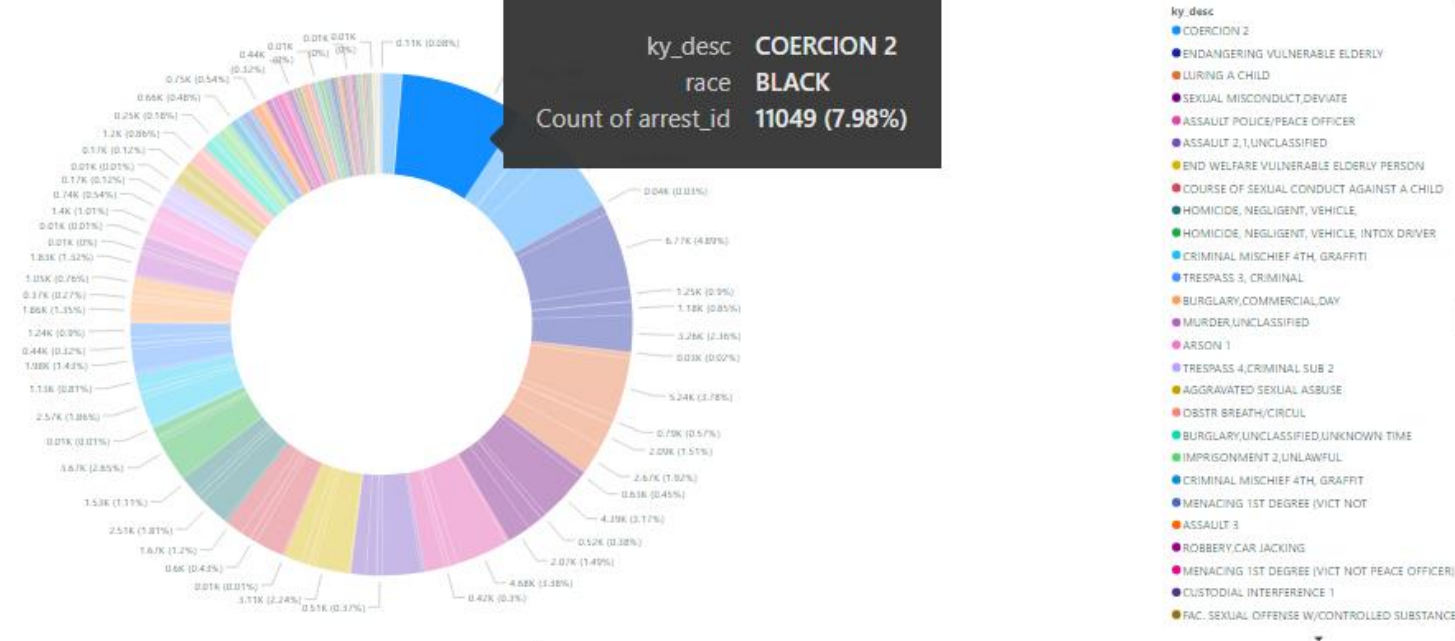
Count of arrest\_id by ky\_desc



Count of arrest\_id by pd\_desc

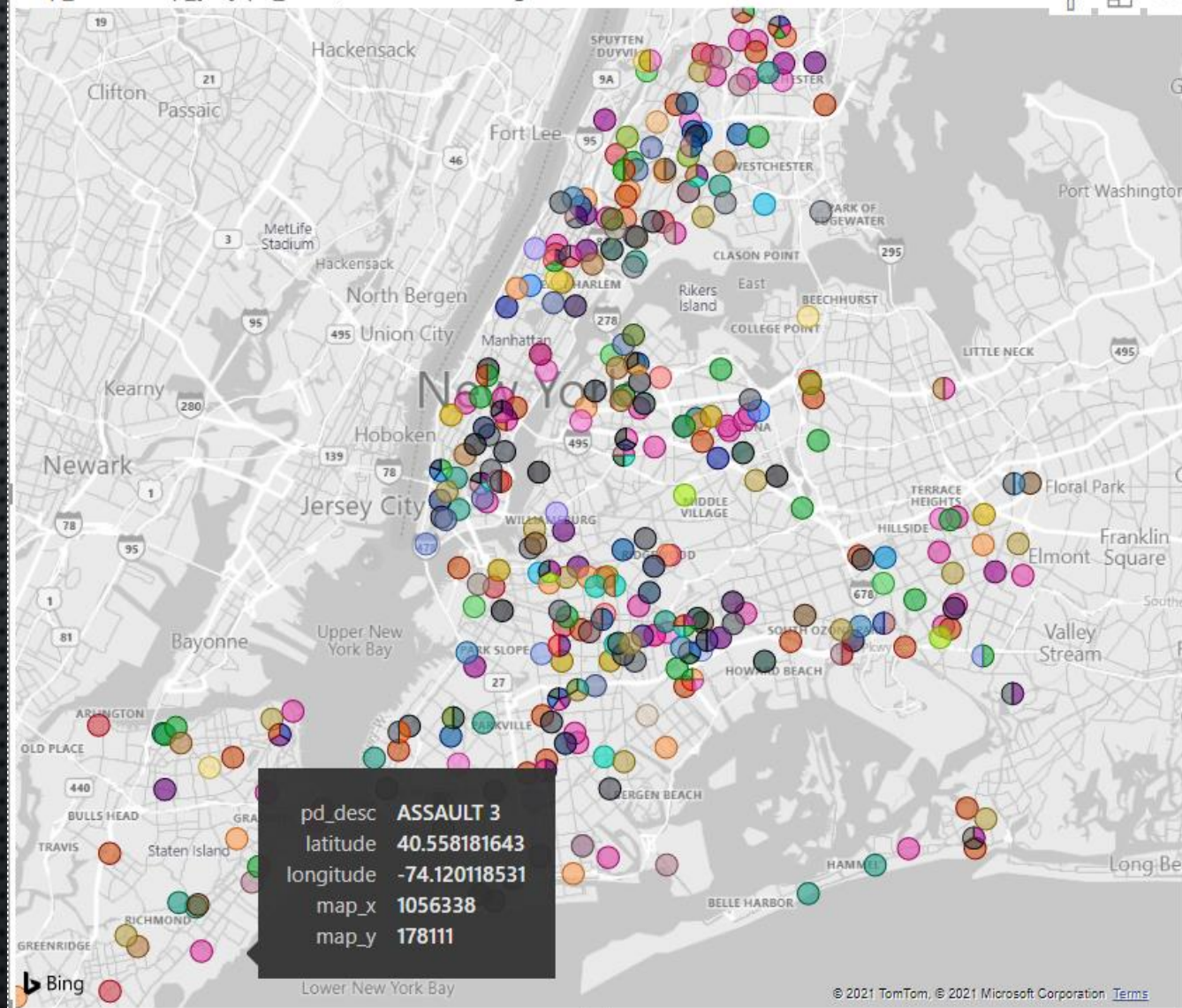


Count of arrest\_id by ky\_desc and race



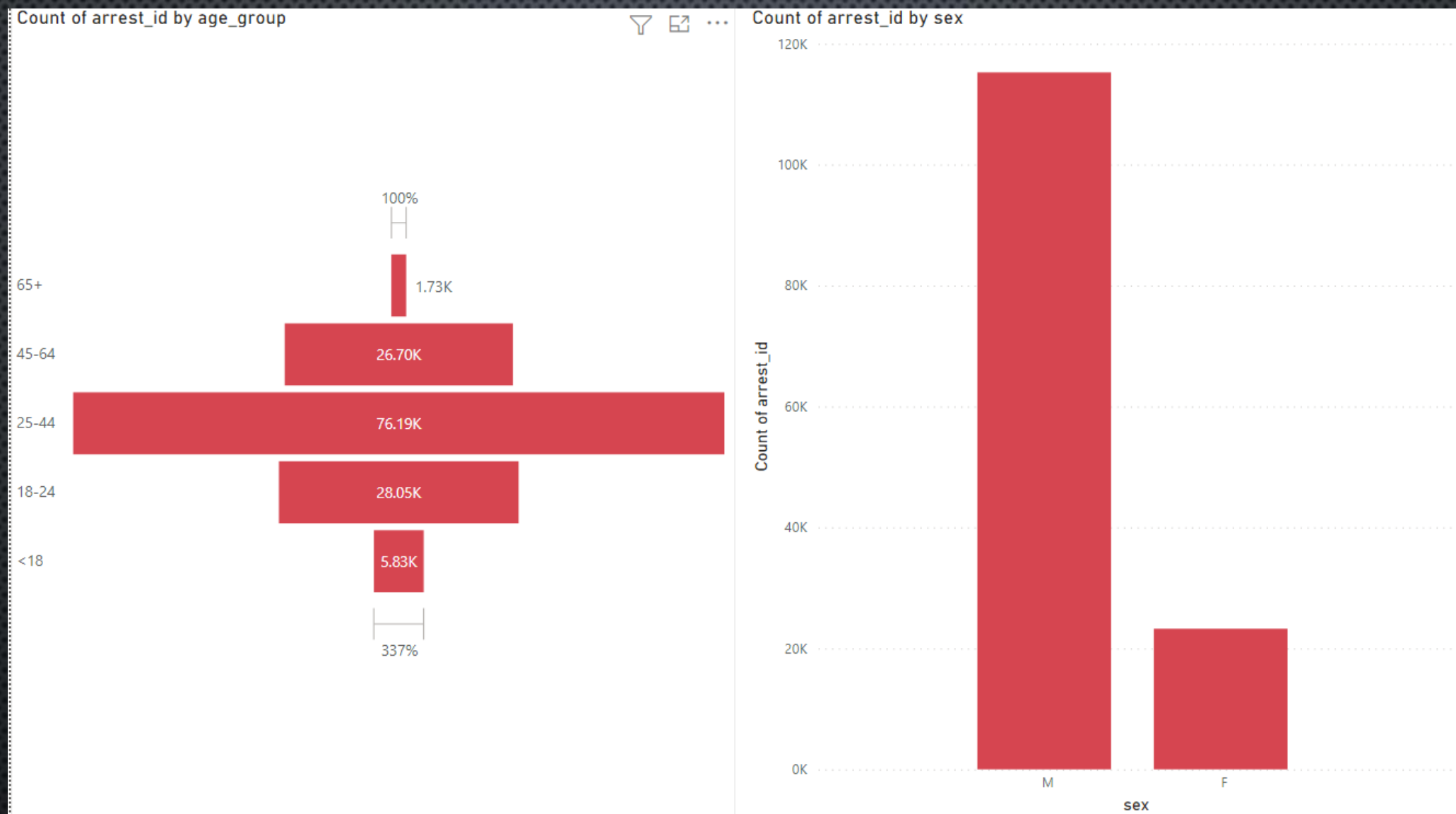


map\_x and map\_y by pd\_desc, latitude and longitude



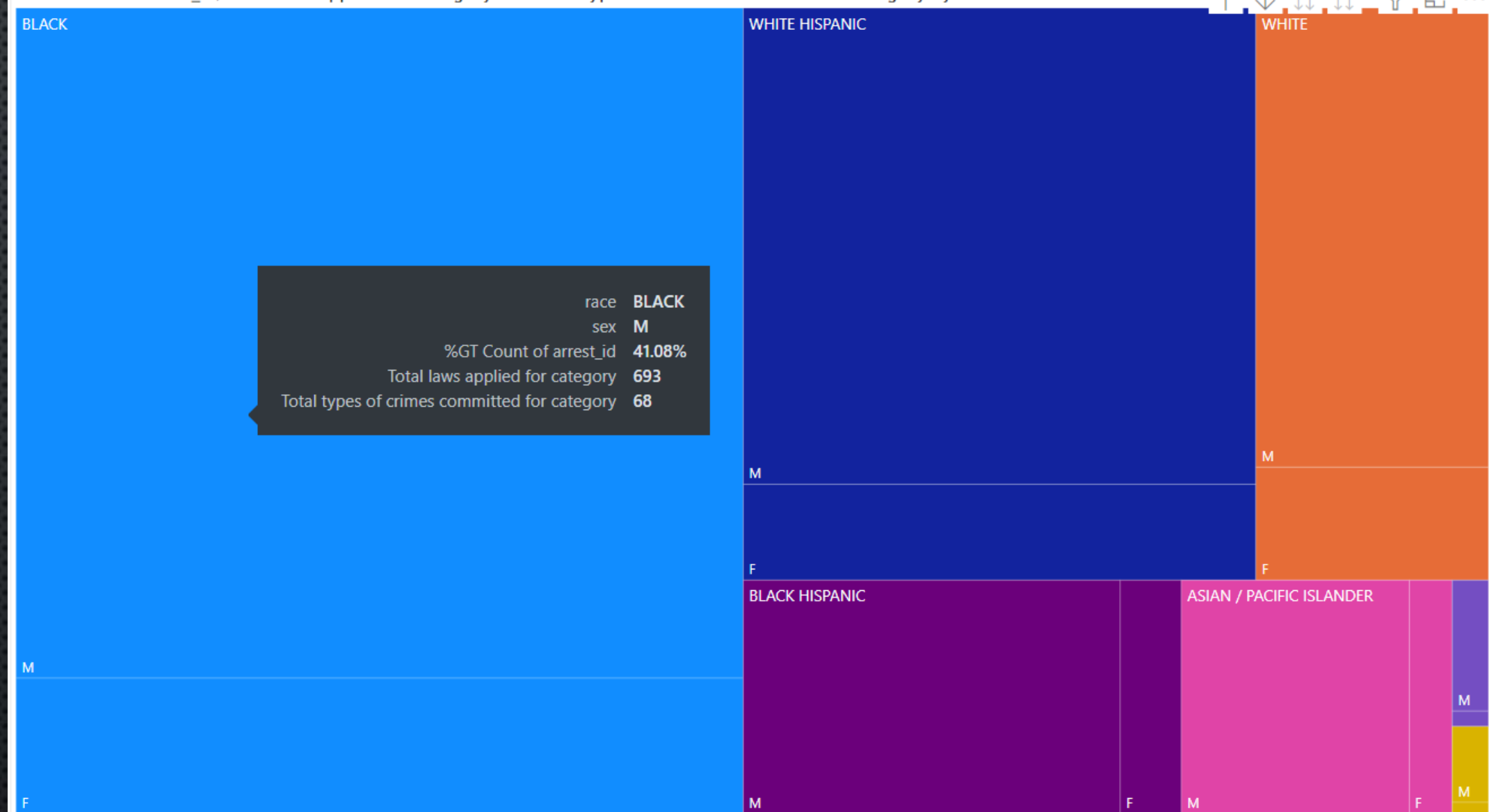
arrest\_date

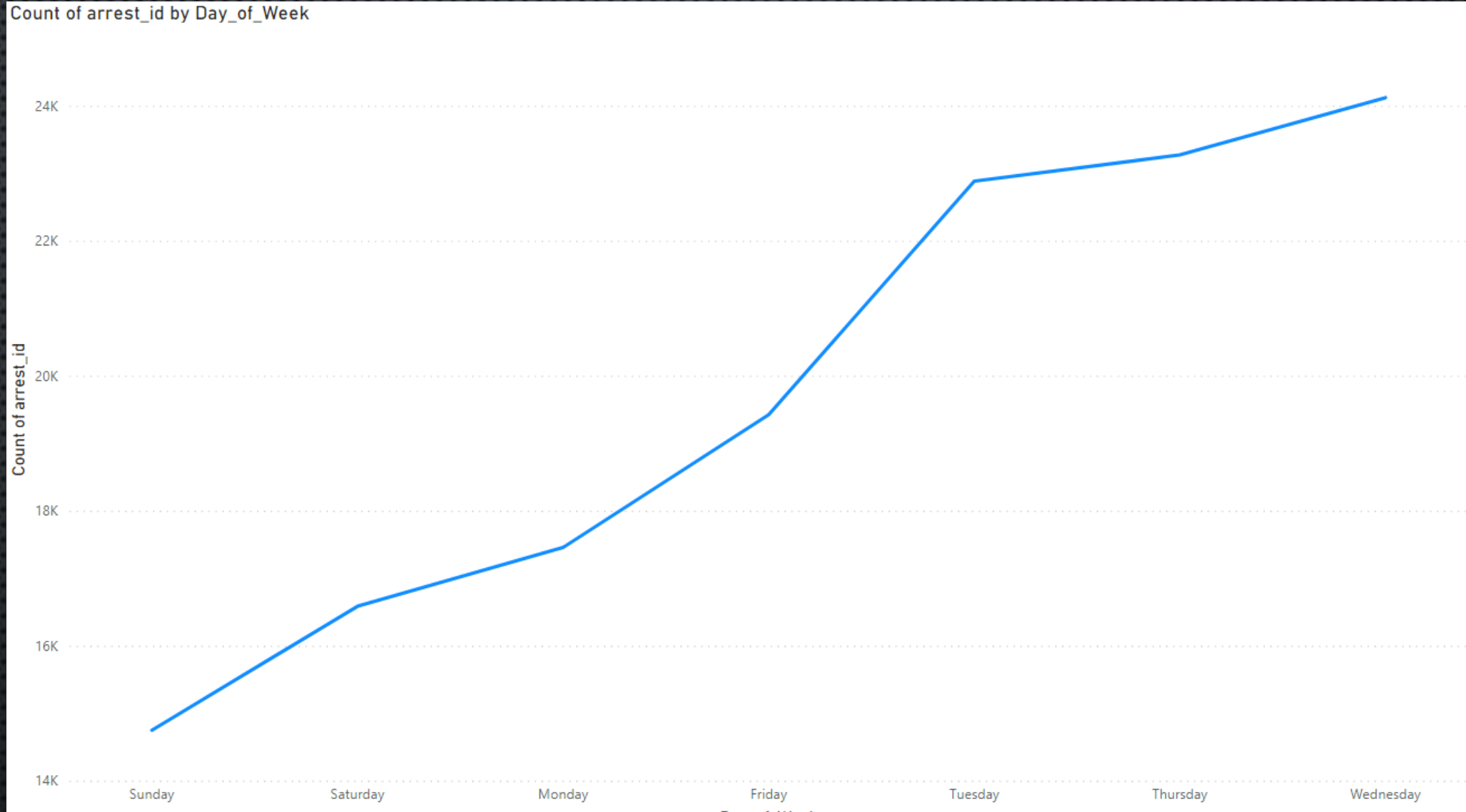
- ☐ Thursday, January 02, 2020
- ☒ Friday, January 03, 2020
- ☐ Saturday, January 04, 2020
- ☐ Sunday, January 05, 2020
- ☐ Monday, January 06, 2020
- ☐ Tuesday, January 07, 2020
- ☐ Wednesday, January 08, 2020
- ☐ Thursday, January 09, 2020
- ☐ Friday, January 10, 2020
- ☐ Saturday, January 11, 2020
- ☐ Sunday, January 12, 2020
- ☐ Monday, January 13, 2020
- ☐ Tuesday, January 14, 2020
- ☐ Wednesday, January 15, 2020
- ☐ Thursday, January 16, 2020
- ☐ Friday, January 17, 2020
- ☐ Saturday, January 18, 2020
- ☐ Sunday, January 19, 2020
- ☐ Monday, January 20, 2020
- ☐ Tuesday, January 21, 2020
- ☐ Wednesday, January 22, 2020
- ☐ Thursday, January 23, 2020
- ☐ Friday, January 24, 2020
- ☐ Saturday, January 25, 2020
- ☐ Sunday, January 26, 2020
- ☐ Monday, January 27, 2020
- ☐ Tuesday, January 28, 2020
- ☐ Wednesday, January 29, 2020
- ☐ Thursday, January 30, 2020
- ☐ Friday, January 31, 2020
- ☐ Saturday, February 01, 2020
- ☐ Sunday, February 02, 2020
- ☐ Monday, February 03, 2020
- ☐ Tuesday, February 04, 2020
- ☐ Wednesday, February 05, 2020
- ☐ Thursday, February 06, 2020
- ☐ Friday, February 07, 2020



# BREAKDOWN OF ARRESTS BY AGE & SEX







MOST ARRESTS ARE ON TUESDAY, WEDNESDAY  
AND THURSDAY



# REFERENCES

- [HTTPS://DEV.MYSQL.COM/DOC/REFMAN/8.0/EN/SERVER-SYSTEM-VARIABLES.HTML](https://dev.mysql.com/doc/refman/8.0/en/server-system-variables.html)
- [HTTPS://LOGICALREAD.COM/OPTIMIZING-SERVER-SETTINGS-MYSQL-MC13/](https://logicalread.com/optimizing-server-settings-mysql-mc13/)
- [HTTPS://WWW.LINODE.COM/DOCS/GUIDES/HOW-TO-OPTIMIZE-MYSQL-PERFORMANCE-USING-MYSQLTUNER/](https://www.linode.com/docs/guides/how-to-optimize-mysql-performance-using-mysqltuner/)
- [HTTPS://GITHUB.COM/MAJOR/MYSQLTUNER-PERL](https://github.com/major/mysqltuner-perl)
- [HTTPS://SEVERALNINES.COM/DATABASE-BLOG/MYSQL-PERFORMANCE-CHEAT-SHEET](https://severalnines.com/database-blog/mysql-performance-cheat-sheet)
- [HTTPS://DOCS.AWS.AMAZON.COM/AMAZONRDS/LATEST/USERGUIDE/MYSQL.PROCEDURAL.IMPORTING.NONRDSREPL.HTML](https://docs.aws.amazon.com/amazonrds/latest/userguide/mysql-procedural-importing-nonrdsrepl.html)
- [HTTPS://WWW1.NYC.GOV/SITE/NYPD/BUREAUS/PATROL/PRECINCTS-LANDING.PAGE](https://www1.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page)
- [HTTPS://DATA.CITYOFNEWYORK.US/PUBLIC-SAFETY/NYPD-ARRESTS-DATA-HISTORIC-/8H9B-RP9U](https://data.cityofnewyork.us/public-safety/nypd-arrests-data-historic-/8h9b-rp9u)
- ... MANY OTHERS