

DATA606 final presentation

By: Nicholas Ogbonna & Geslande Belus



Table of Contents

- data source and data analysis purpose
- BI data visualization
- ERD
- ETL pipeline via AWS Glue
- Creation of data tables into AWS RDS (relational database)
- SQL queries for useful information display
- Analysis cloud database features and technology advantages
- Visualization of analysis for decision-making support

Data Acquisition and Background

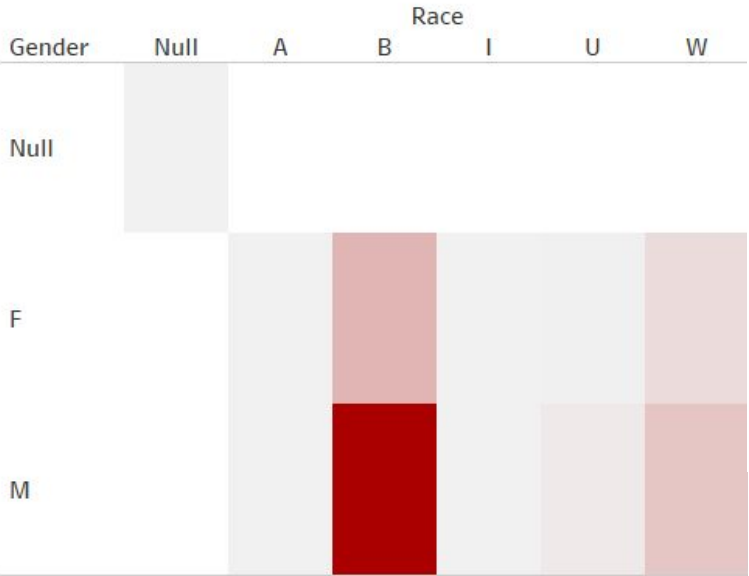
The Baltimore Police Agency launched a massive overhaul to its new Records Management Systems back in May 2020. The improvement will enable the department to migrate from a paper-based system to a completely digital reporting environment. As a consequence of this major shift, we had significant difficulties in appropriately transferring data from the new records system to the previous Open Data Baltimore system. The "Arrests" dataset is one of many open datasets made publicly accessible by Baltimore's police department on the city's Open Data website. This information is provided to us in order to foster more openness and data exchange between the local administration and its residents. This dataset contains arrest records for offenses such as assault, theft, and property damage in the City of Baltimore. [1]

We want to use their database create a model that predicts which district a crime occurs based on various details related to the arrest of the perpetrator such as what he or she was charged with, their gender, etc. Our hope with this presentation is that the department can then use this information to efficiently spread their resources tackling the more likely arrests that would be made in a certain district as well as modify their policing efforts to decrease bias of said policing efforts in certain districts (if any).

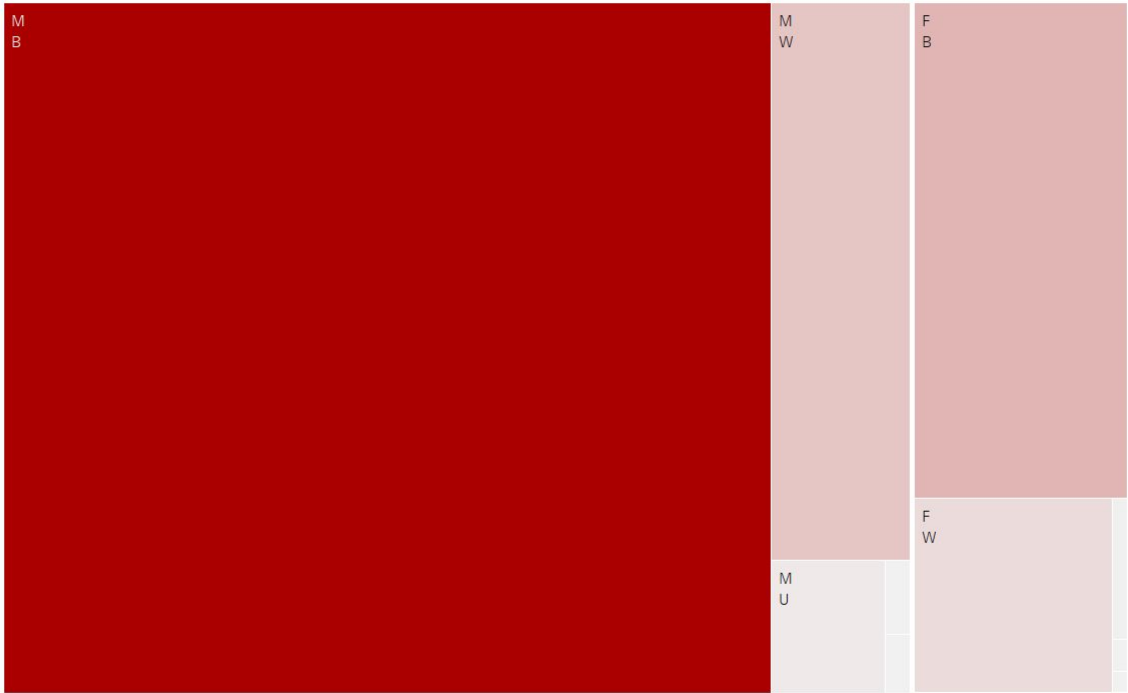
[1] "Baltimore Police Department." Crime Stats | Baltimore Police Department, <https://www.baltimorepolice.org/crime-stats>.

BI

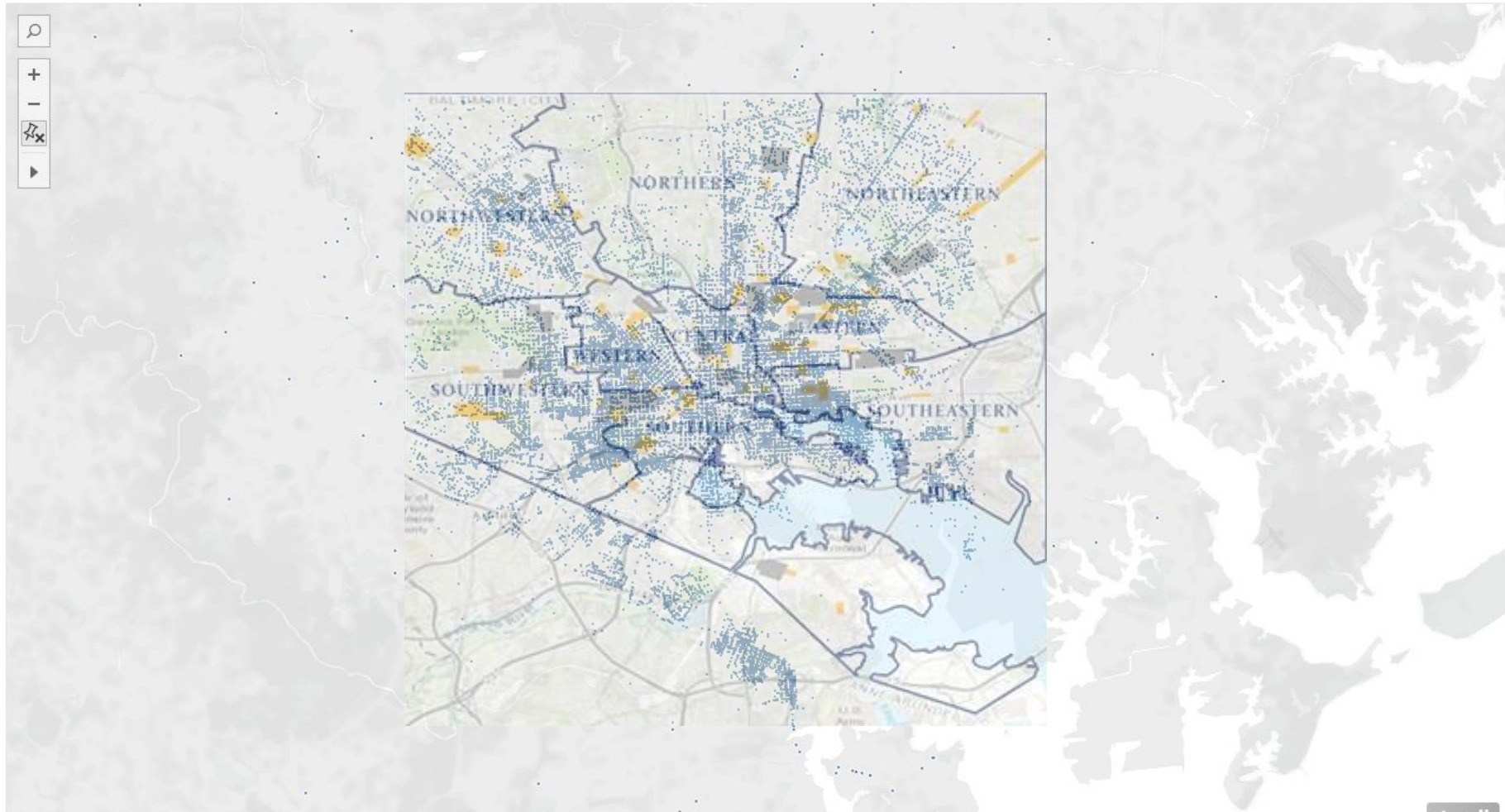
Race x Gender Heatmap



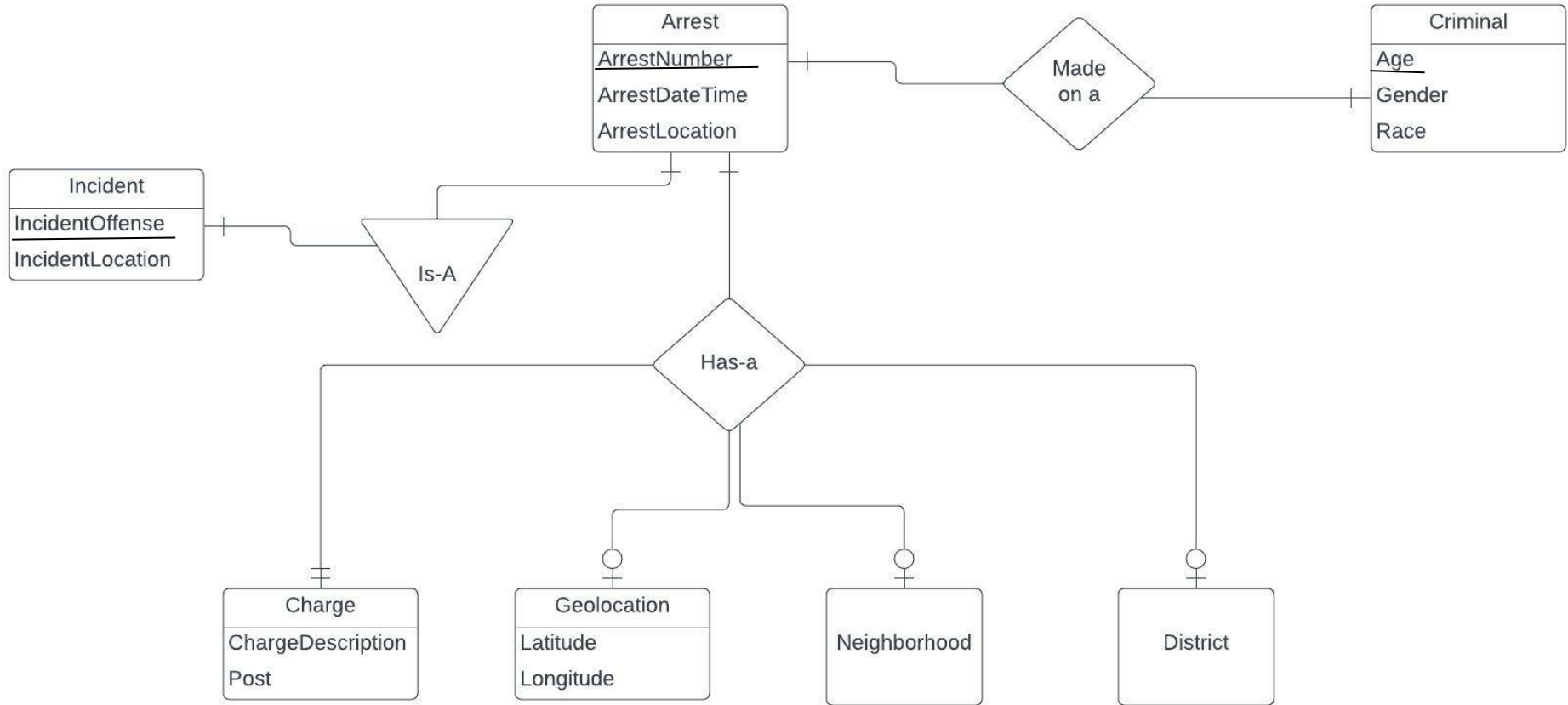
Race x Gender Heatmap



Visual representation of where arrests occur



ER Diagram



ETL Pipeline

<https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#>

[Objects](#)[Properties](#)[Permissions](#)[Metrics](#)[Management](#)[Access Points](#)

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Copy S3 URI](#)[Copy URL](#)[Download](#)[Open](#)[Delete](#)[Actions ▼](#)[Create folder](#)[Upload](#)[<](#) **1** [>](#)

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	Arrests.csv	csv	May 13, 2022, 21:29:31 (UTC-04:00)	27.3 MB	Standard

Summary

DB identifier	CPU	Status	Class
data604-arrests	<div><div></div></div> 41.50%	<div><div></div> Available</div>	db.t2.micro
Role	Current activity	Engine	Region & AZ
Instance	<div><div></div> 1 Connections</div>	SQL Server Express Edition	us-west-2c

- Connectivity & security
- Monitoring
- Logs & events
- Configuration
- Maintenance & backups
- Tags

Connectivity & security

Endpoint & port	Networking	Security
Endpoint	Availability Zone	VPC security groups
data604-arrests.xslshcghvni.us-west-2.rds.amazonaws.com	us-west-2c	<div>default (sg-0d117a46b877b17bc)</div>
Port	VPC	<div><div></div> Active</div>
1433	vpc-09375d7c76ce3cc9a	Publicly accessible
	Subnet group	Yes
	default-vpc-09375d7c76ce3cc9a	Certificate authority
		rds-ca-2019

AWS was an enormous chore to get working.

No databases that support query editor



Currently, query editor only supports Aurora Serverless databases. Only Aurora Serverless database that you have access to will be displayed.

To use the query editor for a database, the database must have the Data API enabled.

[Cancel](#)[Create database](#)[Run](#)[Save](#)[Clear](#)[Output](#)

DATA604ArrestsDB.sql - DESKTOP-S21V5KM\Owner (53) - Microsoft SQL Server Management Studio

Quick Launch (Ctrl+Q)

File Edit View Project Tools Window Help

New Query

age

Execute

Object Explorer

Connect

DESKTOP-S21V5KM\SQLEXPRESS (SQL S

Databases

System Databases

Database Snapshots

Arrests

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.Arrest

dbo.Charge

dbo.Criminal

dbo.District

dbo.GeoLocation

dbo.Incident

dbo.Neighborhood

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

TSQLV4

Security

Server Objects

Replication

PolyBase

Management

XEvent Profiler

DATA604ArrestsDB...1V5KM\Owner (53)

```
IF NOT EXISTS (SELECT * FROM sys.databases WHERE name = 'Arrests')
BEGIN
    CREATE DATABASE Arrests
END
GO

USE Arrests
GO

--check if the table exists
IF NOT EXISTS (SELECT * FROM sysobjects WHERE name='Arrest' and xtype='U')
BEGIN
    CREATE TABLE Arrest (
        ArrestNumber INT PRIMARY KEY IDENTITY (1, 1),
        ArrestDateTime VARCHAR(22),
        ArrestLocation VARCHAR(100)
    )
END
IF NOT EXISTS (SELECT * FROM sysobjects WHERE name='Criminal' and xtype='U')
BEGIN
    CREATE TABLE Criminal (
        criminalID int NOT NULL, Name nvarchar(50)
        , CONSTRAINT PK_criminalID PRIMARY KEY NONCLUSTERED (criminalID)
        , CONSTRAINT FK_criminalID_ArrestNumber FOREIGN KEY (criminalID)
        REFERENCES Arrests.Arrest (ArrestNumber),
        Age INT check (Age >= 0),
        Gender VARCHAR(1) check ((Gender = 'M') or (Gender = 'F')),
        Race VARCHAR(1) check ((Race = 'B') or (Race = 'W') or (Race = 'U') or (Race = 'A') or (Race = 'I'))
    )
END
IF NOT EXISTS (SELECT * FROM sysobjects WHERE name='Incident' and xtype='U')
BEGIN
    CREATE TABLE Incident (
        IncidentID int NOT NULL, Name nvarchar(50)
        , CONSTRAINT PK_criminalID PRIMARY KEY NONCLUSTERED (IncidentID)
        , CONSTRAINT FK_criminalID_ArrestNumber FOREIGN KEY (IncidentID)
        REFERENCES Arrests.Arrest (ArrestNumber)
    )
END
```

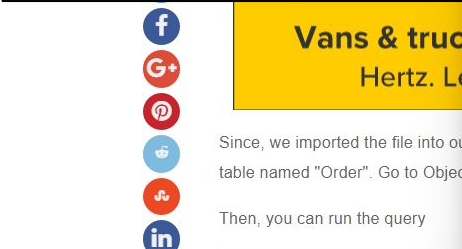
100 %

Connected. (1/1)

DESKTOP-S21V5KM\SQLEXPRESS ... | DESKTOP-S21V5KM\Owner ... | master | 00:00:00 | 0 rows

So we divided and conquered!

We couldn't just import our csv so we had to get creative



Vans & trucks from Hertz. Le

Since, we imported the file into our table named "Order". Go to Object Explorer

Then, you can run the query

```
Use OrderDetails
Select * from Orders
```

Output:

MEN'S W

Free Shipping

Import Flat File 'Arrests'

Results

Introduction

Specify Input File

Preview Data

Modify Columns

Summary

Results

Operation Failed

Summary:

Name	Result
Insert Data	Error

Microsoft SQL Server Management Studio

Error inserting data into table. (Microsoft.SqlServer.Import.Wizard)

Additional information:

- Error inserting data into table. (Microsoft.SqlServer.Prose.Import.SDS)
 - The given value of type String from the data source cannot be converted to type int of the specified target column. (System.Data)
 - Failed to convert parameter value from a String to a Int32. (System.Data)
 - Input string was not in a correct format. (mscorlib)

Copy message Show details OK

< Previous Next > Close

DATA603 (Thursdays)

Other bookmarks

START MONITORING

HIDE AD • AD VIA BUYSSELLADS

f t G+

Special Financing Available

avertys Furniture

Z

s nearby.

Go!

Hertz

dynatrace

See how at dynatrace.com

Some queries

1. Select * from Arrests where ('ArrestLocation' = '2900 PRESBURY ST')

2. SELECT [Id]

,[Age]

,[Gender]

,[Race]

FROM [dbo].[Criminal]

where [Age] >= '20' and [age] < '30'

3. SELECT COUNT ([Age]) AS "Number of criminals in 20s"

FROM [dbo].[Criminal]

where [Age] >= '20' and [age] < '30'

SQLQuery6.sql - DESKTOP-S21V5KM\SQLEXPRESS.Arrests (DESKTOP-S21V5KM\Owner (54))* - Microsoft SQL Server Management Studio

FileEditViewQueryProjectToolsWindowHelp

Arrests

Execute

age

Object Explorer

Connect

DESKTOP-S21V5KM\SQLEXPRESS (SQL S

Databases

System Databases

Database Snapshots

Arrests

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.Arrest

dbo.ArrestInstance

dbo.Charge

dbo.Criminal

dbo.District

dbo.GeoLocation

dbo.Incident

dbo.Neighborhood

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

TSQLV4

Security

Server Objects

Replication

PolyBase

Management

XEvent Profiler

SQLQuery3.sql - D...1V5KM\Owner (60))*

SQLQuery5.sql - D...1V5KM\Owner (53))*

SQLQuery6.sql - D...1V5KM\Owner (54))*

1

USE [Arrests]

GO

3

4

5

6

7

8

9

10

11

12

13

100 %

Results

Messages

ArrestNumber	ArrestDateTime	ArrestLocation
1935154	2020/12/31 23:50:00+00	2900 PRESBURY ST
1942841	2017/05/12 15:52:00+00	2900 PRESBURY ST
1952413	2014/04/25 13:50:00+00	2900 PRESBURY ST
1987423	2019/11/20 23:00:00+00	2900 PRESBURY ST
2018017	2021/01/03 06:09:00+00	2900 PRESBURY ST
2019049	2021/01/01 02:13:00+00	2900 PRESBURY ST
2046234	2015/09/23 14:53:00+00	2900 PRESBURY ST
2084643	2014/07/15 17:15:00+00	2900 PRESBURY ST
2084644	2014/07/15 17:15:00+00	2900 PRESBURY ST
2106278	2019/12/04 13:30:00+00	2900 PRESBURY ST
2108772	2020/05/31 14:50:00+00	2900 PRESBURY ST

100 %

Results

Messages

Query executed successfully.

DESKTOP-S21V5KM\SQLEXPRESS ...

DESKTOP-S21V5KM\Owner ...

Arrests

00:00:00

11 rows

Properties

Current connection parameters

Aggregate Status

Connection failures

Elapsed time

Finish time

Name

Rows returned

Start time

State

Connection

Connection name

Connection Details

Connection elapsed

Connection encrypt

Connection finish ti

Connection rows re

Connection start tir

Connection state

Display name

Login name

Server name

Server version

Session Tracing ID

SPID

Name

The name of the connection.

Ready

Ln 7

Col 22

Ch 22

INS

71°F

Rain off and on

11:40 AM

5/16/2022

SQLQuery5.sql - D:\...1V5KM\SQL\EXPRESS.Arrests (DESKTOP-S21V5KM\Owner (53))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query

age

Arrests

Execute

Object Explorer

DESKTOP-S21V5KM\SQL\EXPRESS (SQL S

Databases

System Databases

Database Snapshots

Arrests

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.Arrest

dbo.ArrestInstance

dbo.Charge

dbo.Criminal

dbo.District

dbo.GeoLocation

dbo.Incident

dbo.Neighborhood

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

TSQLV4

Security

Server Objects

Replication

PolyBase

Management

XEvent Profiler

SQLQuery3.sql - D:\...1V5KM\Owner (60))*

SQLQuery5.sql - D:\...1V5KM\Owner (53))*

```
1 USE [Arrests]
2 GO
3
4 SELECT [Id]
5       , [Age]
6       , [Gender]
7       , [Race]
8 FROM [dbo].[Criminal]
9 where [Age] >= '20' and [age] < '30'
10
11 GO
12
13
14
```

2

100 %

Results

Messages

	Id	Age	Gender	Race
1	175924	27	M	B
2	175925	29	M	B
3	175927	29	F	B
4	175928	29	M	B
5	175929	23	M	B
6	175930	22	F	B
7	175935	26	M	B
8	175936	24	M	B
9	175938	28	M	B
10	175940	25	M	W
11	175943	23	F	B
12	175945	23	M	B
13	175951	27	M	W
14	175957	27	M	B
15	175958	26	M	B
16	175959	26	M	B
17	175961	20	M	B

Query executed successfully.

DESKTOP-S21V5KM\SQL\EXPRESS ... | DESKTOP-S21V5KM\Owner ... | Arrests | 00:00:00 | 72,107 rows

Properties

Current connection parameters

Aggregate Status

Connection failures	
Elapsed time	00:00:00.375
Finish time	5/16/2022 11:39:17 AM
Name	DESKTOP-S21V5KM\SQL
Rows returned	72107
Start time	5/16/2022 11:39:17 AM
State	Open

Connection

Connection name	DESKTOP-S21V5KM\SQL
-----------------	---------------------

Connection Details

Connection elapsed	00:00:00.375
Connection encrypt	Not encrypted
Connection finish ti	5/16/2022 11:39:17 AM
Connection rows re	72107
Connection start tir	5/16/2022 11:39:17 AM
Connection state	Open
Display name	DESKTOP-S21V5KM\SQL
Login name	DESKTOP-S21V5KM\Ow
Server name	DESKTOP-S21V5KM\SQL
Server version	14.0.2037
Session Tracing ID	
SPID	53

Name

The name of the connection.

ReadyLn 9Col 39Ch 39INS

71°F

Rain off and on

11:39 AM

5/16/2022

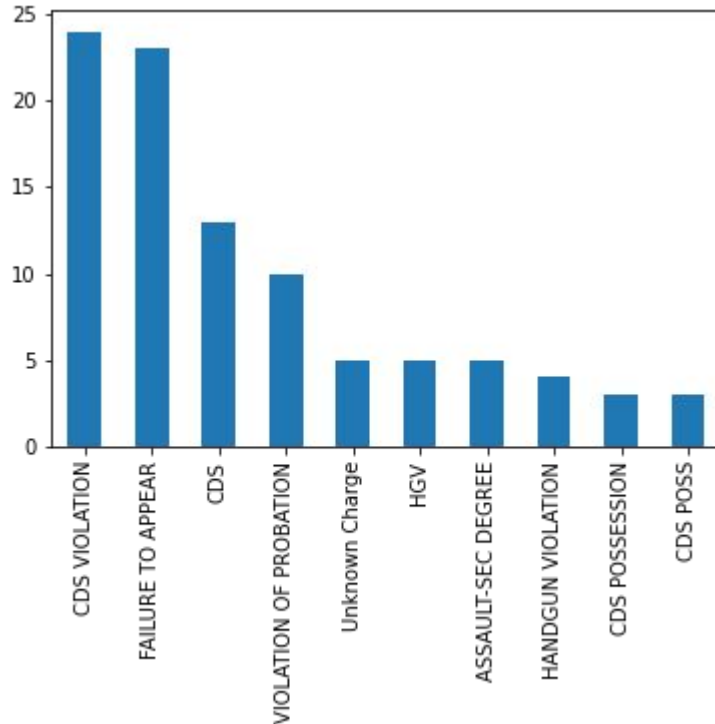
Analysis cloud database features and technology advantages

Some interesting things we found were

- The most common way to get arrested is to not show up to your appointed court date. Whether it's a major or minor offence, one should always try to make their court date.
- Most arrests seem to be made around 11am.
 - Which is interesting as last week, someone found that most crimes were committed around 2-3am
- The most amount of arrests made on a single day was on January 9th 2014
- In the 7 years this data spans, the black males are the most likely to be arrested. There are many articles backing up policing bias that contributed to this trend.

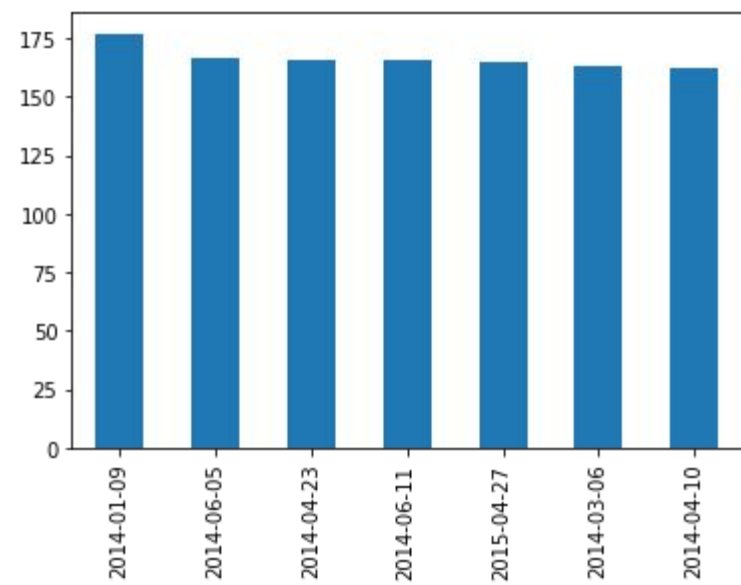
Visualization of analysis for decision-making support

Disregarding "Failure To Appear", it seems most of these arrests were CDS aka Controlled Dangerous Substance charges. This makes sense as people are coming off of the holidays at this point. Perhaps they are trying out their new "presents"



Arrests seem to occur

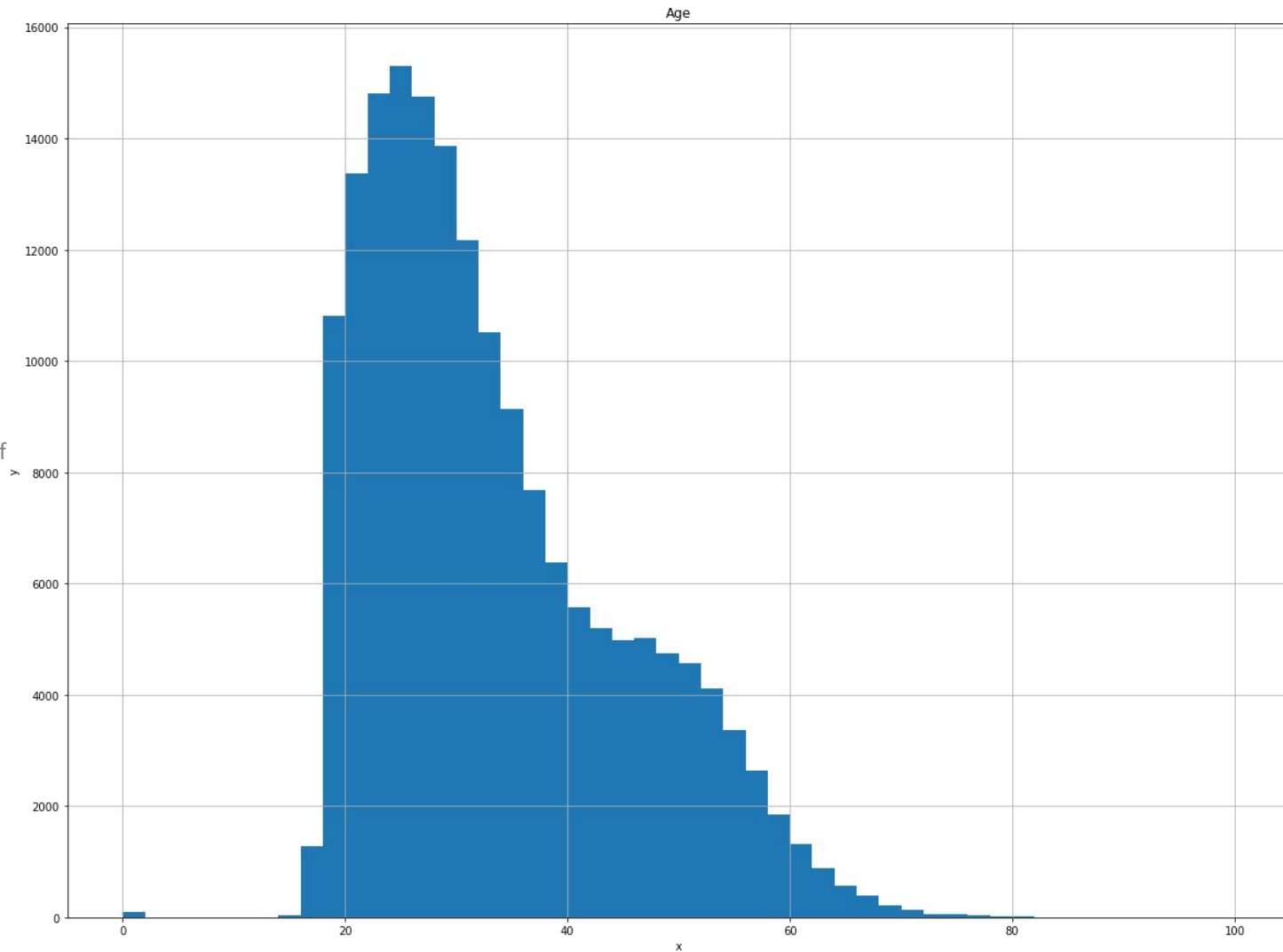
- a couple days following holidays like christmas or easter
- a couple days after pseudo holidays like 4/20 (weed day)
- and in warmer months like June when people are more out and about.



Seems most Baltimore residents are being arrested in their mid twenties with a less noticeable hump occuring in their late 40s.

- There are many factors that can play into why kids in their mid 20s are being arrested as a professor of criminal justice told the new york times [2]

[2]
<https://www.nytimes.com/2011/12/19/us/nearly-a-third-of-americans-are-arrested-by-23-study-says.html>



Any Questions?

