**DADABI – Project 1A**

**Inconsistencies and Missing values-**

1. DR no is a numeric data value with designated 9 digits but the minimum value being 4 (Crime data stats from 2020 but values show otherwise)
2. Date RTPD - reported time shows 12 am for all the cases investigated, this is concluded by browse tool
3. Time occurred - No nulls
4. Area - no null values or inconsistencies
5. Area names - No null works
6. RPT Dist no - No nulls
7. Crime code provided - Stolen Vehicle and Simple Assaults are highest ranks

A screenshot of a computer

Description automatically generated

* 1. With data browse tool, it is concluded that there’s about 98,709 records of victims age as 0 - will use data cleaning to come over inconsistent data. With filter option , we filtered out null value rows in victim age

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Data inconsistencies- Data Cleanse tool By profiling in the browse tool, we can observe the victim sex is divided into M (Male) 172,778, F (Female) 149,259 ,X & H (Unknown) 32,900
2. Mocodes - Null values can be filtered out by filter tool

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Victim Descent - Frequency of hispanic race and white is as follows with about 53,802 null values

A screenshot of a computer

Description automatically generated

1. CM Cd2, Cd3, Cd 4 are mostly empty hence, using filter tool , we eliminate null values columns

A screen shot of a computer

Description automatically generated

1. Weapons used Cd and Weapon Desc are mostly null values (65%)
2. Longitude and latitude are 0 which is supposed to be deleted completely from rows using data filter tool

A screenshot of a computer

Description automatically generated

1. **Observations -**

* The highest crime location Is in the 800 N Alameda street followed by 7th and 6th street giving a total of 2000+
* The most used form of crime is set by hands fist and feet by 78,000 and verbal threats of 11,000
* Descent code H - Hispanic shows the highest race to commit crime followed by W - white, B - black

1. **Final workflow -**

A screenshot of a computer

Description automatically generated