**Snowflake Code**

**Data Prep and Analytics:**  
//Check for any NULLS

SELECT COUNT(\*) AS total\_rows,

COUNT(AREA) AS non\_null\_AREA,

COUNT(AREA\_NAME) AS non\_null\_AREA\_NAME,

COUNT(CRM\_CD) AS non\_null\_CRM\_CD,

COUNT(CRM\_CD\_DESC) AS non\_null\_CRM\_CD\_DESC,

COUNT(DATE\_OCC) AS non\_null\_DATE\_OCC,

COUNT(DATE\_RPTD) AS non\_null\_DATE\_RPTD,

COUNT(LAT) AS non\_null\_LAT,

COUNT(LOCATION) AS non\_null\_LOCATION,

COUNT(LON) AS non\_null\_LON,

COUNT(MONTH) AS non\_null\_MONTH,

COUNT(PART\_1\_2) AS non\_null\_PART\_1\_2,

COUNT(PREMIS\_CD) AS non\_null\_PREMIS\_CD,

COUNT(PREMIS\_DESC) AS non\_null\_PREMIS\_DESC,

COUNT(RPT\_DIST\_NO) AS non\_null\_RPT\_DIST\_NO,

COUNT(STATUS) AS non\_null\_STATUS,

COUNT(STATUS\_DESC) AS non\_null\_STATUS\_DESC,

COUNT(TIME\_OCC) AS non\_null\_TIME\_OCC,

COUNT(VICT\_AGE) AS non\_null\_VICT\_AGE,

COUNT(VICT\_DESCENT) AS non\_null\_VICT\_DESCENT,

COUNT(VICT\_SEX) AS non\_null\_VICT\_SEX,

COUNT(WEAPON\_DESC) AS non\_null\_WEAPON\_DESC,

COUNT(WEAPON\_USED\_CD) AS non\_null\_WEAPON\_USED\_CD,

COUNT(YEAR) AS non\_null\_YEAR,

FROM CRIMEDATA;

//Check for Outliers

SELECT \*

FROM CRIMEDATA

WHERE VICT\_AGE > 100 OR VICT\_AGE < 0;

//Data Cleaning

DELETE FROM CRIMEDATA

WHERE VICT\_AGE > 100 OR VICT\_AGE < 0;

//Check for more Outliers (Came back with 0,0 coordinates)

SELECT

LAT,

LON

FROM CRIMEDATA

WHERE LAT < (SELECT AVG(LAT) - 3 \* STDDEV(LAT) FROM CRIMEDATA)

OR LAT > (SELECT AVG(LAT) + 3 \* STDDEV(LAT) FROM CRIMEDATA)

OR LON < (SELECT AVG(LON) - 3 \* STDDEV(LON) FROM CRIMEDATA)

OR LON > (SELECT AVG(LON) + 3 \* STDDEV(LON) FROM CRIMEDATA);

//More Data Cleaning (Delete the rows with 0,0 coordinates)

DELETE FROM CRIMEDATA

WHERE LAT < (SELECT AVG(LAT) - 3 \* STDDEV(LAT) FROM CRIMEDATA)

OR LAT > (SELECT AVG(LAT) + 3 \* STDDEV(LAT) FROM CRIMEDATA)

OR LON < (SELECT AVG(LON) - 3 \* STDDEV(LON) FROM CRIMEDATA)

OR LON > (SELECT AVG(LON) + 3 \* STDDEV(LON) FROM CRIMEDATA);

//Descriptive Analysis for numerical columns

SELECT

MIN(AREA) AS min\_area,

MAX(AREA) AS max\_area,

AVG(AREA) AS avg\_area,

STDDEV(AREA) AS stddev\_area,

MIN(CRM\_CD) AS min\_crm\_cd,

MAX(CRM\_CD) AS max\_crm\_cd,

AVG(CRM\_CD) AS avg\_crm\_cd,

STDDEV(CRM\_CD) AS stddev\_crm\_cd,

MIN(LAT) AS min\_lat,

MAX(LAT) AS max\_lat,

AVG(LAT) AS avg\_lat,

STDDEV(LAT) AS stddev\_lat,

MIN(LON) AS min\_lon,

MAX(LON) AS max\_lon,

AVG(LON) AS avg\_lon,

STDDEV(LON) AS stddev\_lon,

MIN(MONTH) AS min\_month,

MAX(MONTH) AS max\_month,

AVG(MONTH) AS avg\_month,

STDDEV(MONTH) AS stddev\_month,

MIN(PART\_1\_2) AS min\_part\_1\_2,

MAX(PART\_1\_2) AS max\_part\_1\_2,

AVG(PART\_1\_2) AS avg\_part\_1\_2,

STDDEV(PART\_1\_2) AS stddev\_part\_1\_2,

MIN(PREMIS\_CD) AS min\_premis\_cd,

MAX(PREMIS\_CD) AS max\_premis\_cd,

AVG(PREMIS\_CD) AS avg\_premis\_cd,

STDDEV(PREMIS\_CD) AS stddev\_premis\_cd,

MIN(RPT\_DIST\_NO) AS min\_rpt\_dist\_no,

MAX(RPT\_DIST\_NO) AS max\_rpt\_dist\_no,

AVG(RPT\_DIST\_NO) AS avg\_rpt\_dist\_no,

STDDEV(RPT\_DIST\_NO) AS stddev\_rpt\_dist\_no,

MIN(TIME\_OCC) AS min\_time\_occ,

MAX(TIME\_OCC) AS max\_time\_occ,

AVG(TIME\_OCC) AS avg\_time\_occ,

STDDEV(TIME\_OCC) AS stddev\_time\_occ,

MIN(VICT\_AGE) AS min\_vict\_age,

MAX(VICT\_AGE) AS max\_vict\_age,

AVG(VICT\_AGE) AS avg\_vict\_age,

STDDEV(VICT\_AGE) AS stddev\_vict\_age,

MIN(WEAPON\_USED\_CD) AS min\_weapon\_used\_cd,

MAX(WEAPON\_USED\_CD) AS max\_weapon\_used\_cd,

AVG(WEAPON\_USED\_CD) AS avg\_weapon\_used\_cd,

STDDEV(WEAPON\_USED\_CD) AS stddev\_weapon\_used\_cd,

MIN(YEAR) AS min\_year,

MAX(YEAR) AS max\_year,

AVG(YEAR) AS avg\_year,

STDDEV(YEAR) AS stddev\_year

FROM CRIMEDATA;

//Descriptive analysis for categorical columns

SELECT

AREA\_NAME,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY AREA\_NAME

ORDER BY count desc;

SELECT

CRM\_CD\_DESC,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY CRM\_CD\_DESC

ORDER BY count desc;

SELECT

LOCATION,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY LOCATION

ORDER BY count desc;

SELECT

premis\_desc,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY premis\_desc

ORDER BY count desc;

SELECT

status,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY status

ORDER BY count desc;

SELECT

status\_desc,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY status\_desc

ORDER BY count desc;

SELECT

vict\_descent,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY vict\_descent

ORDER BY count desc;

SELECT

vict\_sex,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY vict\_sex

ORDER BY count desc;

SELECT

weapon\_desc,

COUNT(\*) AS count

FROM CRIMEDATA

GROUP BY weapon\_desc

ORDER BY count desc;

//Analyze patterns

SELECT

DATE\_OCC,

COUNT(\*) AS crime\_count

FROM CRIMEDATA

GROUP BY DATE\_OCC

ORDER BY crime\_count desc;

SELECT

DATE\_RPTD,

COUNT(\*) AS crime\_count

FROM CRIMEDATA

GROUP BY DATE\_RPTD

ORDER BY crime\_count desc;

SELECT

year,

COUNT(\*) AS crime\_count

FROM CRIMEDATA

GROUP BY year

ORDER BY crime\_count desc;

//Analyze location distribution

SELECT LAT, LON

FROM CRIMEDATA;

**Machine Learning:**

USE ROLE admin;

GRANT USAGE ON DATABASE untenesj TO ROLE sysadmin;

GRANT USAGE ON SCHEMA lacrimedata TO ROLE sysadmin;

GRANT CREATE SNOWFLAKE.ML.CLASSIFICATION ON SCHEMA untenesj.lacrimedata TO ROLE sysadmin;

CREATE OR REPLACE view binary\_classification\_view AS

SELECT LON, LAT, VICT\_SEX

FROM crimedata;

SELECT \* FROM binary\_classification\_view ORDER BY RANDOM(42) LIMIT 5;

CREATE OR REPLACE SNOWFLAKE.ML.CLASSIFICATION model\_binary(

INPUT\_DATA => SYSTEM$REFERENCE('view', 'binary\_classification\_view'),

TARGET\_COLNAME => 'VICT\_SEX'

);

SELECT model\_binary!PREDICT(INPUT\_DATA => object\_construct(\*))

as prediction from crimedata;

SELECT \*, model\_binary!PREDICT(

INPUT\_DATA => object\_construct(\*))

as predictions from crimedata;