--Q1: how many crimes?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data All;

cnt = FOREACH grp GENERATE COUNT(data) ;

STORE cnt INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q1CrimesCount' USING PigStorage (',');

dump cnt;

--Q2: oldest and most recent crime committed?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data All;

yearsminmax = FOREACH grp GENERATE MIN(data.Year) As minYear, MAX(data.Year) As maxYear;

STORE yearsminmax INTO 'Users/najlsalsabti/Downloads/pig\_Output/Q2dataYearRange' USING PigStorage (',');

dump yearsminmax;

--Q3: what are the months with the most crimes rates?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By Month;

cnt = FOREACH grp GENERATE group AS Month, COUNT(data) As countPerMonth;

orderd= ORDER cnt BY countPerMonth DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q3CrimePerMonth' USING PigStorage (',');

dump orderd;

--Q4: what are the most common domestic abuses?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered =FILTER data BY Domestic==1 ;

grp = GROUP filtered By Primary\_Type;

cnt = FOREACH grp GENERATE group As Crime, COUNT(filtered) As Domstic\_offense\_Rate;

orderd= ORDER cnt BY Domstic\_offense\_Rate DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q4DomesticAbuses' USING PigStorage (',');

dump orderd;

--Q5: what are the most common crimes that result in arrests?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered =FILTER data BY Arrest==1 ;

grp = GROUP filtered By Primary\_Type;

cnt = FOREACH grp GENERATE group AS Crime, COUNT(filtered) As Arrest\_Rate;

orderd= ORDER cnt BY Arrest\_Rate DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q5ArrestCrimes' USING PigStorage (',');

dump orderd;

--Q6: what are the most common crimes?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By Primary\_Type;

cnt = FOREACH grp GENERATE group As Crime, COUNT(data) As Crime\_Rate;

orderd= ORDER cnt BY Crime\_Rate DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q6MostCommonCrimes' USING PigStorage (',');

dump orderd;

--Q7: what are the most common crimes per Year?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By (Year, Primary\_Type);

freq = Foreach grp generate FLATTEN(group) as (Year, Primary\_Type) , COUNT(data) as frequency;

ordered = ORDER freq BY Year DESC, frequency DESC;

STORE ordered INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q7CrimeFreqPerYear' USING PigStorage (',');

dump ordered;

--Q8: what are the most common crimes per day/night time?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By (AM\_PM, Primary\_Type);

freq = Foreach grp generate FLATTEN(group) as (AM\_PM, Primary\_Type) , COUNT(data) as frequency;

ordered = ORDER freq BY AM\_PM DESC, frequency DESC;

STORE ordered INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q8CrimeFreqPerDay' USING PigStorage (',');

dump ordered;

--Q9: what are the top 10 blocks in crime rates?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By Block;

cnt = FOREACH grp generate group as Block, COUNT(data) as freq;

oredred = Order cnt BY freq DESC;

topBlocks = FILTER oredred By freq >1000;

STORE topBlocks INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q9TopBlocks' USING PigStorage (',');

dump topBlocks;

--Q10: what were the stolen assets?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered = FILTER data BY Primary\_Type matches '.\*THEFT.\*';

grp = GROUP filtered By Description;

cnt = FOREACH grp generate group As StolenAsset, COUNT(filtered) as freq;

oredred = Order cnt BY freq DESC;

STORE oredred INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q10StolenAssets' USING PigStorage (',');

dump oredred;

--Q11: What are the peak hours for criminal activites?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By (Hour, AM\_PM);

cnt = FOREACH grp generate FLATTEN(group) as (Hour, AM\_PM) , COUNT(data) as frequency;

--ordered = Order cnt BY frequency DESC;

--peakHours = TOP(3, 1, ordered);

peakHours= Order cnt BY frequency DESC;

STORE peakHours INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q11PeakHours' USING PigStorage (',');

dump peakHours;

--Q12: What are the top 20 locations with ciminal activites?

-- topLocs = TOP(20, frequency, cnt); I used this line and I also tried to use limit but both cause errors

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

grp = GROUP data By Location\_Description;

cnt = FOREACH grp generate group As Location , COUNT(data) as frequency;

--ordered = Order cnt BY frequency DESC;

--topLocs = TOP(20, 1, ordered);

topLocs = Order cnt BY frequency DESC;

STORE topLocs INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q12locations' USING PigStorage (',');

dump topLocs;

--Q13: what is the frequency of simple and aggrevated batteries in doemstic and non domestic abuses?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered = FILTER data BY Primary\_Type matches '.\*BATTERY.\*';

Split filtered INTO simp if Description matches '.\*SIMPLE.\*', agg if Description matches '.\*AGG.\*';

simpgrp= GROUP simp BY Domestic;

simpcnt = FOREACH simpgrp generate group AS Domestic, COUNT(simp);

agggrp= GROUP agg BY Domestic;

aggcnt = FOREACH agggrp generate group AS Domestic, COUNT(agg);

STORE aggcnt INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q13AggrevatedBattteriesDomVsNonDom' USING PigStorage (',');

STORE simpcnt INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q13SimpleBattteriesDomVsNonDom' USING PigStorage (',');

dump aggcnt;

dump simpcnt;

--Q14: what are the plcaes in which theft is commonly carried out?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered = FILTER data BY Primary\_Type matches '.\*THEFT.\*';

grp = GROUP filtered By Location\_Description ;

cnt = FOREACH grp generate group as Location , COUNT(filtered) as frequency;

--ordered = Order cnt BY frequency DESC;

--topLocs = TOP(20, 1, ordered);

topLocs= Order cnt BY frequency DESC;

STORE topLocs INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q14ToplocationforThefts' USING PigStorage (',');

dump topLocs;

--Q15: where are batteries are performed?

data = LOAD '/Users/najlsalsabti/Downloads/Data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',') AS(ID:chararray,Case\_Number:chararray,Date:chararray,

Block:chararray,IUCR:chararray,Primary\_Type:chararray,Description:chararray,Location\_Description:chararray,Arrest:int,Domestic:int,Beat:int,

District:int,Ward:int,Community\_Area:int,FBI\_Code:chararray,X:long,Y:long,Year:int,Updated\_On:chararray,Latitude:double,Longitude:double,

Location:chararray,Month:int,Day:int,Hour:int,Minute:int,AM\_PM:chararray);

filtered = FILTER data BY Primary\_Type matches '.\*BATTERY.\*';

grp = GROUP filtered By Location\_Description;

cnt = FOREACH grp generate group as Location , COUNT(filtered) as frequency;

--ordered = Order cnt BY frequency DESC;

--topLocs = TOP(20, 1, ordered);

topLocs= Order cnt BY frequency DESC;

STORE topLocs INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q15ToplocationforBatteries' USING PigStorage (',');

dump topLocs;

--Q16 the total number of crimes of VIOLATION/ASSAULT in each block

filtered = FILTER data BY Primary\_Type matches '.\*ASSAULT.\*' OR Primary\_Type matches '.\*VIOLATION.\*' ;

grp = GROUP filtered By Block;

cnt = FOREACH grp generate group as Block, COUNT(filtered) as total;

ordered = Order cnt BY total DESC;

STORE ordered INTO '/Users/najlsalsabti/Downloads/pig\_Output/Q16AssultViolantionPerBlock' USING PigStorage (',');

dump ordered;

--Q17 the total number of unique ward in each block

cnt1 = FOREACH data generate Block, Ward;

grp = GROUP cnt1 By Block;

cnt = FOREACH grp

{

unique\_ward = DISTINCT cnt1.Ward;

generate group, COUNT(unique\_ward) as total\_unique\_ward;

};

ordered = Order cnt BY total\_unique\_ward DESC;

STORE ordered INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q17UniqueWardsInEachBlock' USING PigStorage (',');

dump ordered;

--Q18 the total number of FBI offices and their total number of communications received in each block.

cnt1 = FOREACH data generate Block, FBI\_Code;

grp = GROUP cnt1 By Block;

cnt = FOREACH grp

{

unique\_FBICode = DISTINCT cnt1.FBI\_Code;

generate group, COUNT(unique\_FBICode) as total\_unique\_unique\_FBICode ,COUNT(cnt1) as total;

};

oredred = Order cnt BY total\_unique\_unique\_FBICode DESC;

STORE oredred INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q18FBIOfficesPerBlock' USING PigStorage (',');

dump oredred;

--Q19 The years in which the most domestic violence occurred

filtered =FILTER data BY Domestic==1 ;

grp = GROUP filtered By Year;

cnt = FOREACH grp GENERATE group as Year, COUNT(filtered) As Domestic\_Rate;

orderd= ORDER cnt BY Domestic\_Rate DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q19YearWithMostDomesticViolance' USING PigStorage (',');

dump orderd;

--Q20 the total number of crimes in each month in this year

filtered =FILTER data BY Year==2022;

grp = GROUP filtered By Month;

cnt = FOREACH grp GENERATE group as Month, COUNT(filtered) As Crime\_Rate;

orderd= ORDER cnt BY Month DESC;

STORE orderd INTO '/Users/najlsalsabti/Downloads/pig\_Output/ Q20CrimeRatePerMonthIn2022' USING PigStorage (',');

dump orderd;