

```
In [1]: %load_ext sql
import pandas as pd
%sql sqlite://
```

```
Out[1]: 'Connected: None@None'
```

```
In [2]: # For compatibility across multiple platforms
import os
IB = os.environ.get('INSTABASE_URI',None) is not None
open = ib.open if IB else open
```

```
In [3]: # Load tables from CSV files
# Compas Score
with open('compas-scores.csv','rU') as f:
    Scores = pd.read_csv(f, index_col=0, encoding='utf-8')
%sql drop table if exists Scores;
%sql persist Scores
```

Done.

```
Out[3]: u'Persisted scores'
```

```
In [4]: %%sql
select * from Scores limit 5
```

Done.

```
Out[4]:
```

id	name	first	last	compas_screening_date	sex	dob	age	age_cat	rac
1	miguel hernandez	miguel	hernandez	2013-08-14	Male	1947-04-18	69	Greater than 45	Ott
2	michael ryan	michael	ryan	2014-12-31	Male	1985-02-06	31	25 - 45	Ca
3	kevon dixon	kevon	dixon	2013-01-27	Male	1982-01-22	34	25 - 45	Afr Arr
4	ed philo	ed	philo	2013-04-14	Male	1991-05-14	24	Less than 25	Afr Arr
5	marcu brown	marcu	brown	2013-01-13	Male	1993-01-21	23	Less than 25	Afr Arr

In [34]: *#Research Questions*

```
#1)      What is the distribution of risk assessment scores across race?  
#a.      How does the distribution look for violent risk assessment scores across race?  
#2)      Does the risk assessment score correlate with actual recidivism?  
#a.      Is this true for violent risk assessment scores that predict violent recidivism?  
#b.      What is the breakdown of recidivism by race?  
#3)      Are there some crimes that correspond to certain risk scores?  
#a.      Are felony crimes consistently associated with higher or lower risk scores? What about non-felonies?  
#b.      Do some crimes correspond to a higher recidivism?
```

File "<ipython-input-34-3915a0e83fbe>", line 1

```
-- 1)  What is the breakdown by race per risk assessment score? Is  
      ^  
      there a discrepancy in scores by race?
```

SyntaxError: invalid syntax

In [5]: **%%sql**
Select count(distinct(c_charge_desc))
From Scores

Done.

Out[5]:

count(distinct(c_charge_desc))
531

In [32]: **%%sql**
-- Figure 3.1: Looking at the number of distinct charges each race has been charged with
select count(distinct c_charge_desc) as Count_of_Charges, race
from Scores
group by race

Done.

Out[32]:

Count_of_Charges	race
377	African-American
30	Asian
310	Caucasian
177	Hispanic
20	Native American
124	Other

```
In [84]: %%sql
-- Figure 3.2, creating a tables that shows percentage of each race in
data set
select race, Count/(Total*1.0) as Ratio, Count
FROM (select count() as Total From Scores) Totals,
      (select race, count() as Count from Scores group by race) Counts
```

Done.

Out[84]:

race	Ratio	Count
African-American	0.494428850897	5813
Asian	0.00493323126648	58
Caucasian	0.347452581441	4085
Hispanic	0.0935612826401	1100
Native American	0.00340222845964	40
Other	0.0562218252956	661

```
In [33]: %%sql
-- Number of recidivists for every score in every race
select decile_score as Score, race, count(is_recid) as Count of
from scores
where is_recid = "1" and decile_score >0 -- shows us how many who scored
high on decile score actually were recidivists
group by race, decile_score

-- look at low decile scores for white and black and compare to recid ca
tegory
```

Done .

Out[33]:

decile_score	race	count(is_recid)
1	African-American	107
2	African-American	143
3	African-American	177
4	African-American	204
5	African-American	199
6	African-American	246
7	African-American	270
8	African-American	277
9	African-American	300
10	African-American	251
1	Asian	1
3	Asian	3
5	Asian	1
6	Asian	2
7	Asian	1
8	Asian	2
10	Asian	1
1	Caucasian	162
2	Caucasian	136
3	Caucasian	110
4	Caucasian	130
5	Caucasian	124
6	Caucasian	121
7	Caucasian	99
8	Caucasian	87
9	Caucasian	73
10	Caucasian	46
1	Hispanic	55
2	Hispanic	40
3	Hispanic	30
4	Hispanic	19
5	Hispanic	32

6	Hispanic	25
7	Hispanic	18
8	Hispanic	14
9	Hispanic	12
10	Hispanic	16
3	Native American	1
4	Native American	2
6	Native American	1
7	Native American	3
8	Native American	1
9	Native American	2
10	Native American	3
1	Other	30
2	Other	28
3	Other	17
4	Other	27
5	Other	14
6	Other	11
7	Other	6
8	Other	8
9	Other	6
10	Other	7

```
In [9]: %%sql
-- Figure 3.3: Table with count of recidvists who initially committed fe
lony crimes with risk scores of 8 or higher
select decile_score as score, race, count(is_recid) as Count_Recid, c_ch
arge_desc
From scores
where score >= 8 and is_recid = "1" and c_charge_desc like "%felony%"
group by race, score, c_charge_desc
```


Done .

Out[9]:

score	race	Count_Recid	c_charge_desc
8	African-American	1	Felony Battery
8	African-American	3	Felony Battery w/Prior Convict
8	African-American	5	Felony Driving While Lic Suspd
8	African-American	2	Felony Petit Theft
9	African-American	1	Felony Battery
9	African-American	2	Felony Battery (Dom Strang)
9	African-American	5	Felony Battery w/Prior Convict
9	African-American	1	Felony Driving While Lic Suspd
9	African-American	8	Felony Petit Theft
10	African-American	2	Felony Batt(Great Bodily Harm)
10	African-American	1	Felony Battery
10	African-American	1	Felony Battery (Dom Strang)
10	African-American	2	Felony Battery w/Prior Convict
10	African-American	1	Felony Committing Prostitution
10	African-American	1	Felony Driving While Lic Suspd
10	African-American	1	Felony Petit Theft
8	Asian	1	Felony Petit Theft
8	Caucasian	1	Felony Battery (Dom Strang)
8	Caucasian	1	Felony Battery w/Prior Convict
8	Caucasian	2	Felony Petit Theft
9	Caucasian	1	Felony Battery
9	Caucasian	1	Felony Battery w/Prior Convict
9	Caucasian	1	Felony Committing Prostitution
9	Caucasian	7	Felony Petit Theft
10	Caucasian	1	Felony Battery (Dom Strang)
8	Hispanic	1	Felony Battery (Dom Strang)
8	Hispanic	1	Felony Petit Theft
10	Hispanic	1	Felony Petit Theft

In [13]: `%%sql`
 --Generates Table of number of recidivists who were arrested with no charge and had high risk assesement scores
 SELECT race, decile_score as score, c_charge_desc, count(is_recid) as Count_Recid
 FROM Scores
 WHERE score > 7 and is_recid = "1" and c_charge_desc like "%arrest%"
 GROUP BY race, decile_score

Done.

Out[13]:

race	score	c_charge_desc	Count_Recid
African-American	8	arrest case no charge	49
African-American	9	arrest case no charge	58
African-American	10	arrest case no charge	48
Caucasian	8	arrest case no charge	15
Caucasian	9	arrest case no charge	12
Caucasian	10	arrest case no charge	7
Hispanic	8	arrest case no charge	6
Hispanic	9	arrest case no charge	4
Hispanic	10	arrest case no charge	5
Native American	9	arrest case no charge	1
Other	8	arrest case no charge	2
Other	10	arrest case no charge	1

```

In [15]: %%sql
-- Figure 3.6 Number of non recidivists who had high risk scores and no
-- initial charge
SELECT race, decile_score as score, c_charge_desc, count(is_recid) as Count_Non_Recid
FROM Scores
WHERE score > 7 and is_recid = "0" and c_charge_desc like "%arrest%"
GROUP BY race, decile_score

```

Done.

Out[15]:

race	score	c_charge_desc	Count_Non_Recid
African-American	8	arrest case no charge	65
African-American	9	arrest case no charge	74
African-American	10	arrest case no charge	66
Caucasian	8	arrest case no charge	13
Caucasian	9	arrest case no charge	22
Caucasian	10	arrest case no charge	6
Hispanic	8	arrest case no charge	6
Hispanic	9	arrest case no charge	4
Hispanic	10	arrest case no charge	3
Native American	10	arrest case no charge	1
Other	8	arrest case no charge	1
Other	10	arrest case no charge	2

```
In [6]: %%sql
-- Tables shows individuals with low decile scores and the corresponding
number violent decile scores over 5
--individuals in those groups received
Select decile_score, race, count(v_decile_score) -- change ot is violent
recid
FROM Scores
where decile_score < 5 and decile_score != -1 and v_decile_score > 5
group by race, decile_score
```

Done.

Out[6]:

decile_score	race	count(v_decile_score)
2	African-American	3
3	African-American	36
4	African-American	105
2	Caucasian	2
3	Caucasian	19
4	Caucasian	63
3	Hispanic	12
4	Hispanic	12
3	Native American	1
4	Native American	1
3	Other	2
4	Other	19

```
In [20]: %%sql
-- Attempting to determine if number of prior charges correlates to reci
divism
-- Looking at number of prior charges compared to count of those who rec
idivised
SELECT race, count(priors_count), is_recid -- can take this out
FROM Scores
WHERE is_recid = 1 and priors_count > 0
GROUP BY race, is_recid
-- Why was it giving me the same number of cases for both columns?
```

Done.

Out[20]:

race	count(priors_count)	is_recid
African-American	1835	1
Asian	10	1
Caucasian	846	1
Hispanic	193	1
Native American	11	1
Other	105	1

```
In [17]: %%sql
-- Shows if violent decile score actually correlates to violent recidivi
sm
-- Similar to Figure 2.3 in Project Write Up
SELECT race, v_decile_score, count(is_violent_recid)
FROM Scores
WHERE is_violent_recid = "1"
GROUP BY race, v_decile_score
```

Done .

Out[17]:

race	v_decile_score	count(is_violent_recid)
African-American	1	40
African-American	2	40
African-American	3	67
African-American	4	56
African-American	5	58
African-American	6	79
African-American	7	61
African-American	8	49
African-American	9	61
African-American	10	30
Asian	4	1
Asian	5	1
Asian	6	1
Asian	8	1
Caucasian	1	55
Caucasian	2	30
Caucasian	3	35
Caucasian	4	30
Caucasian	5	20
Caucasian	6	23
Caucasian	7	16
Caucasian	8	8
Caucasian	9	14
Caucasian	10	2
Hispanic	1	10
Hispanic	2	14
Hispanic	3	9
Hispanic	4	3
Hispanic	5	6
Hispanic	6	6
Hispanic	7	1
Hispanic	8	5

Hispanic	9	3
Hispanic	10	1
Native American	3	1
Native American	5	2
Native American	6	1
Native American	9	2
Other	1	8
Other	2	1
Other	3	7
Other	4	5
Other	5	1
Other	6	6
Other	7	3
Other	8	6
Other	9	2
Other	10	1

In [32]: `%%sql`
 -- Following 4 cells calculate ratio of AA and Caucasians amongst those with a risk score of 1 or 10
 SELECT R_Count/(Ta_Count*1.0) as Rate_Of_AA_Low
 FROM (select race, decile_score, count() as R_Count From Scores where decile_score = "1" and race like "%African%")AA_Count,
 (select race, decile_score, count() as Ta_Count From Scores where decile_score ="1")T_Count

Done.

Out[32]:

Rate_Of_AA_Low
0.269305393869

In [31]: `%%sql`
 SELECT R_Count_H/(Ta_Count_H*1.0) as Rate_of_AA_High
 FROM (select race, decile_score, count() as R_Count_H From Scores where decile_score = "10" and race like "%African%")AA_Count_High,
 (select race, decile_score, count() as Ta_Count_H From Scores where decile_score ="10")T_Count_High

Done.

Out[31]:

Rate_of_AA_High
0.770491803279

```
In [30]: %%sql
SELECT R_Count/(Ta_Count*1.0) as Rate_Of_C_Low
FROM (select race, decile_score, count() as R_Count From Scores where de
cile_score = "1" and race like "%Cauc%")C_Count,
      (select race, decile_score, count() as Ta_Count From Scores where de
cile_score ="1")Tc_Count
```

Done.

```
Out[30]:
```

Rate_Of_C_Low
0.462553356616

```
In [29]: %%sql
SELECT R_Count/(Ta_Count*1.0) as Rate_Of_C_High
FROM (select race, decile_score, count() as R_Count From Scores where de
cile_score = "10" and race like "%Cauc%")C_Count_H,
      (select race, decile_score, count() as Ta_Count From Scores where de
cile_score ="10")Tc_Count_H
```

Done.

```
Out[29]:
```

Rate_Of_C_High
0.149180327869

```
In [14]: %%sql
-- These next four cells calculate the ratio for AA and Caucasians among
individuals
-- who earned violent recidicism scores 1 and 10
SELECT Count_V/(T_Count_V*1.0) as Rate_Of_AA_Low_V
FROM (select race, v_decile_score, count() as Count_V From Scores where
v_decile_score = "1" and race like "%African%")AA_Count_V,
      (select race, v_decile_score, count() as T_Count_V From Scores where
v_decile_score ="1")Total_V
```

Done.

```
Out[14]:
```

Rate_Of_AA_Low_V
0.295028282227

```
In [17]: %%sql
SELECT Count_V_H/(T_VH*1.0) as Rate_Of_AA_High_V
FROM (select race, v_decile_score, count() as Count_V_H From Scores where
v_decile_score = "10" and race like "%African%")AA_Count_VH,
(select race, v_decile_score, count() as T_VH From Scores where v_decile_score = "10")Total_VH
```

Done.

```
Out[17]:
```

Rate_Of_AA_High_V
0.78

```
In [18]: %%sql
SELECT Count_V_C/(T_C_Count_V*1.0) as Rate_Of_C_Low_V
FROM (select race, v_decile_score, count() as Count_V_C From Scores where
v_decile_score = "1" and race like "%Cauc%")C_Viol_Count,
(select race, v_decile_score, count() as T_C_Count_V From Scores where
v_decile_score = "1")Total_Viol
```

Done.

```
Out[18]:
```

Rate_Of_C_Low_V
0.476332241739

```
In [20]: %%sql
SELECT Count_V_C_High/(Total_Count_V_H*1.0) as Rate_Of_C_V_High
FROM (select race, v_decile_score, count() as Count_V_C_High From Scores
where v_decile_score = "10" and race like "%Cauc%") C_Count_V_H,
(select race, v_decile_score, count() as Total_Count_V_H From Scores
where v_decile_score = "10")Total_C_V_High
```

Done.

```
Out[20]:
```

Rate_Of_C_V_High
0.14