In [673]:
```python
import pandas as pd
import numpy as np
from sqlalchemy import create_engine
```

In [674]:
```python
###add in password when using
pw=""
```

In [675]:
```python
#nfl draft data from 1985 to 2015
#from pro-football-reference.com
csv_file_1 = "Resources/nfl_draft.csv"
draft_df = pd.read_csv(csv_file_1)
draft_df=draft_df[["Player", "Year", "Rnd", "Tm", "Pos", "Age", "College/Univ"]]
draft_df.columns=["player", "year_drafted", "round_drafted", "team", "pos", "age"
draft_df.head()
```

Out[675]:

	player	year_drafted	round_drafted	team	pos	age	college
0	Jameis Winston	2015	1.0	TAM	QB	21.0	Florida St.
1	Marcus Mariota	2015	1.0	TEN	QB	21.0	Oregon
2	Dante Fowler	2015	1.0	JAX	OLB	21.0	Florida
3	Amari Cooper	2015	1.0	OAK	WR	21.0	Alabama
4	Brandon Scherff	2015	1.0	WAS	T	23.0	Iowa

In [676]:
```python
#nfl player arrest data. compiled by fivethirtyeight.com, usatoday.com reporter b
#January 2000 to March 2017
csv_file_2 = "Resources/ArrestIncidents.csv"
arrest_df = pd.read_csv(csv_file_2)
arrest_df.columns=["date", "team", "player", "pos", "case", "category", "desc", "
arrest_df.head()
```

Out[676]:

	date	team	player	pos	case	category	desc	outcome
0	3/27/2017	SEA	Trevone Boykin	QB	Arrested	Drugs	Accused of marijuana possession, public intoxi...	Resolution undetermined.
1	3/26/2017	CHI	Deiondre Hall	CB	Arrested	Disorderly conduct	Accused of disorderly conduct, public intoxica...	Resolution undetermined.
2	3/26/2017	GB	Makinton Dorleant	CB	Arrested	Police interference	Accused of interfence after scuffle outside of...	Resolution undetermined.
3	3/7/2017	KC	Demetrius Harris	TE	Arrested	Drugs	Suspected of felony marijuana possession as pa...	Resolution undetermined.
4	3/5/2017	LAR	Ethan Westbrooks	DT	Arrested	Domestic violence	Suspected of domestic violence in Sacramento a...	Resolution undetermined.

In [677]:
```python
draft_df.sort_values(by="pos")["pos"].unique(), arrest_df.sort_values(by="pos")[
```

Out[677]:
```
(array(['C', 'CB', 'DB', 'DE', 'DL', 'DT', 'FB', 'FS', 'G', 'ILB', 'K',
        'KR', 'LB', 'LS', 'NT', 'OL', 'OLB', 'P', 'QB', 'RB', 'S', 'SS',
        'T', 'TE', 'WR'], dtype=object),
 array(['C', 'CB', 'DB', 'DE', 'DE/DT', 'DT', 'FB', 'K', 'LB', 'OG', 'OL',
        'OT', 'P', 'QB', 'RB', 'S', 'TE', 'WR'], dtype=object))
```

In [678]:
```python
arrest_df["pos"] = arrest_df.pos.replace({"DE/DT":"DL",
                                          "OG":"G",
                                          "OT":"T"})

draft_df["pos"] = draft_df.pos.replace({"FS":"S",
                                        "SS":"S",
                                        "ILB":"LB",
                                        "OLB":"LB"
                                        })
```

In [679]:
```python
#nfl suspension data compiled by fivethirtyeight.com
csv_file_3 = "Resources/nfl-suspensions-data.csv"
suspension_df = pd.read_csv(csv_file_3)
suspension_df.rename(columns={"desc.":"desc", "name":"player"}, inplace=True)
suspension_df.head()
```

Out[679]:

	player	team	games	category	desc	year	sou
0	F. Davis	WAS	Indef.	Substance abuse, repeated offense	Marijuana-related	2014	http://www.cbssports.com/nfl/eye-on-football/
1	J. Blackmon	JAX	Indef.	Substance abuse, repeated offense	NaN	2014	http://espn.go.com/nfl/story/_/id/11257934/ju
2	L. Brazill	IND	Indef.	Substance abuse, repeated offense	NaN	2014	http://www.nfl.com/news/story/0ap200000036462
3	T. Jackson	WAS	Indef.	Substance abuse, repeated offense	NaN	2014	http://www.nfl.com/news/story/0ap200000036408
4	M. Hapes	NYG	Indef.	Personal conduct	Gambling-related	1946	http://espn.go.com/blog/nflnation/tag/_/name

In [680]:
```python
draft_df.sort_values(by="team")["team"].unique(), suspension_df.sort_values(by="
```

Out[680]: (array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
       'DEN', 'DET', 'GNB', 'HOU', 'IND', 'JAX', 'KAN', 'MIA', 'MIN',
       'NOR', 'NWE', 'NYG', 'NYJ', 'OAK', 'PHI', 'PHO', 'PIT', 'RAI',
       'RAM', 'SDG', 'SEA', 'SFO', 'STL', 'TAM', 'TEN', 'WAS'],
     dtype=object),
  array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
       'DEN', 'DET', 'FREE', 'GB', 'HOU', 'IND', 'JAX', 'KC', 'LA', 'MIA',
       'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'OAK', 'PHI', 'PIT', 'SD', 'SEA',
       'SF', 'STL', 'TB', 'TEN', 'WAS'], dtype=object),
  array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
       'DEN', 'DET', 'Free agent', 'GB', 'HOU', 'IND', 'JAC', 'KC', 'LAC',
       'LAR', 'MIA', 'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'OAK', 'PHI', 'PIT',
       'SD', 'SEA', 'SF', 'STL', 'TB', 'TEN', 'WAS'], dtype=object))

```python
In [681]: #normalize team names
          #broadcast tags
          #Raiders - Both in LA and Oakland, moving to Vegas after next year
          #Chargers - Moved from SD to LA
          #Rams - Moved from LA to STL to LA
          #Cardinals - "Moved" from Phoenix to Arizona
          #normalize free agent tags

          draft_df["team"] = draft_df.team.replace({"SFO":"SF",
                            "GNB":"GB",
                            "KAN":"KC",
                            "NOR":"NO",
                            "NWE":"NE",
                            "OAK":"RAI",
                            "RAM":"LAR",
                            "SDG":"LAC",
                            "STL":"LAR",
                            "TAM":"TB",
                            "PHO":"ARI"})

          suspension_df["team"] = suspension_df.team.replace({"OAK":"RAI",
                                            "STL":"LAR",
                                            "LA":"LAR",
                                            "SD":"LAC"})

          arrest_df["team"] = arrest_df.team.replace({"OAK":"RAI",
                                      "STL":"LAR",
                                      "SD":"LAC",
                                      "JAC":"JAX",
                                      "Free agent":"FREE"})

          draft_df.sort_values(by="team")["team"].unique(), \
              suspension_df.sort_values(by="team")["team"].unique(), \
              arrest_df.sort_values(by="team")["team"].unique()
```

```
Out[681]: (array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
                  'DEN', 'DET', 'GB', 'HOU', 'IND', 'JAX', 'KC', 'LAC', 'LAR', 'MIA',
                  'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'PHI', 'PIT', 'RAI', 'SEA', 'SF',
                  'TB', 'TEN', 'WAS'], dtype=object),
           array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
                  'DEN', 'DET', 'FREE', 'GB', 'HOU', 'IND', 'JAX', 'KC', 'LAC',
                  'LAR', 'MIA', 'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'PHI', 'PIT', 'RAI',
                  'SEA', 'SF', 'TB', 'TEN', 'WAS'], dtype=object),
           array(['ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
                  'DEN', 'DET', 'FREE', 'GB', 'HOU', 'IND', 'JAX', 'KC', 'LAC',
                  'LAR', 'MIA', 'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'PHI', 'PIT', 'RAI',
                  'SEA', 'SF', 'TB', 'TEN', 'WAS'], dtype=object))
```

In [682]:
```python
#assign teams to index for relational data
teams = ['FREE', 'ARI', 'ATL', 'BAL', 'BUF', 'CAR', 'CHI', 'CIN', 'CLE', 'DAL',
         'DEN', 'DET', 'GB', 'HOU', 'IND', 'JAX', 'KC', 'LAC',
         'LAR', 'MIA', 'MIN', 'NE', 'NO', 'NYG', 'NYJ', 'PHI', 'PIT', 'RAI',
         'SEA', 'SF', 'TB', 'TEN', 'WAS']
teams=pd.DataFrame(teams)
teams.columns=["team"]
teams.head()
```

Out[682]:

	team
0	FREE
1	ARI
2	ATL
3	BAL
4	BUF

In [683]:
```python
#replacing team with team_id across all dataframes

for thing in teams["team"]:
    draft_df["team"] = draft_df.team.replace({thing:teams.loc[teams["team"]==thi
    suspension_df["team"] = suspension_df.team.replace({thing:teams.loc[teams["t
    arrest_df["team"] = arrest_df.team.replace({thing:teams.loc[teams["team"]==t


draft_df.head()
```

Out[683]:

	player	year_drafted	round_drafted	team	pos	age	college
0	Jameis Winston	2015	1.0	30	QB	21.0	Florida St.
1	Marcus Mariota	2015	1.0	31	QB	21.0	Oregon
2	Dante Fowler	2015	1.0	15	LB	21.0	Florida
3	Amari Cooper	2015	1.0	27	WR	21.0	Alabama
4	Brandon Scherff	2015	1.0	32	T	23.0	Iowa

In [684]:
```python
#creating college name, position, relational databases

college_df = pd.DataFrame(draft_df["college"].unique())
college_df = college_df.dropna()
college_df.columns = ["school"]

pos_df = pd.DataFrame(draft_df["pos"].unique())
pos_df.columns = ["position"]
```

In [685]:
```python
#creating player name databases

player_name_df = draft_df["player"]
player_name_df = pd.DataFrame(player_name_df)
player_name_df.sort_index(ascending=False, inplace=True)
player_name_df = player_name_df.reset_index()
player_name_df.columns=["dropped_col", "player"]
player_name_df = player_name_df.drop(["dropped_col"], axis=1)
player_name_df.head()

split_name_df = pd.DataFrame()
split_name_df = player_name_df["player"].str.split(" ", expand=True)
split_name_df.columns=['f_name','l_name', "drop1", "drop2"]
split_name_df = split_name_df.drop(["drop1", "drop2"], axis=1)
split_name_df.head()
```

Out[685]:

	f_name	l_name
0	Donald	Chumley
1	Ray	Noble
2	Dan	Lynch
3	Paul	Woodside
4	Raymond	Polk

In [686]:
```python
#creating arrest info id dataframes (case, category)

case_df = pd.DataFrame(arrest_df["case"].unique())
case_df.columns=["result"]

category_df = pd.DataFrame(arrest_df["category"].unique())
category_df.columns=["cat_type"]
category_df.head()
```

Out[686]:

	cat_type
0	Drugs
1	Disorderly conduct
2	Police interference
3	Domestic violence
4	DUI, resisting arrest

In [687]:

```python
#normalizing player names in player suspension data

split_sus_df = pd.DataFrame()
split_sus_df = suspension_df["player"].str.split(" ", expand=True)
split_sus_df.columns=['f_init','l_name']
split_sus_df.head()
```

Out[687]:

	f_init	l_name
0	F.	Davis
1	J.	Blackmon
2	L.	Brazill
3	T.	Jackson
4	M.	Hapes

In [688]:

```python
#unsuccessful attempt to convert player suspension data
#trouble comparing first initial with a full name and compiling values

last_name_test_df = []
temp_last_name = []
temp_first_name = []
for last_name in split_sus_df["l_name"]:
    temp_last_name = split_name_df.loc[split_name_df["l_name"]==last_name]
    last_name_test_df.append(temp_last_name)

# for name in last_name_test_df:
#     print(name["l_name"])

#     for player in temp_last_name:
#         print(temp_last_name["f_name"])


#         if player[0] == split_name_df.loc[split_name_df["l_name"]==last_name][
#             last_name_test_df.append(temp_last_name)



# last_name_test_df[1]["f_name"]
# for first_name in last_name_test_df[1]["f_name"]:
#     print(first_name[0])

# last_name_test_df = []
# for last_name in split_sus_df:
#     last_name_test_df.append(last_name)
# last_name_test_df
```

In [689]:
```python
#removing undrafted guys and guys drafted after 2015
#work in progress

removing_df = []
for the_guy in arrest_df["player"]:
    if the_guy not in player_name_df["player"].values:
        removing_df.append(the_guy)

for the_guy in removing_df:
    arrest_df = arrest_df.loc[arrest_df["player"] != the_guy]

arrest_df.head()
```

Out[689]:

	date	team	player	pos	case	category	desc	outcome
5	2/25/2017	14	David Parry	DT	Arrested	DUI, resisting arrest	Suspected of stealing golf cart, driving drunk...	Resolution undetermined.
6	2/25/2017	3	Matt Elam	S	Arrested	Drugs	Pulled over in Miami Beach, accused of possess...	Resolution undetermined.
7	2/16/2017	24	Darrelle Revis	CB	Charged	Assault	Accused of aggravated assault after altercatio...	Charges dismissed.
8	2/16/2017	17	King Dunlap	T	Arrested	Violating court order	Suspected of violating protective order in Nas...	Resolution undetermined.
10	1/10/2017	18	Tre Mason	RB	Arrested	Eluding police	Accused of fleeing law enforcement, careless d...	Resolution undetermined.

In [690]:
```python
# replacing player names with player id
for dude in player_name_df["player"]:

    draft_df["player"] = draft_df.player.replace({dude:player_name_df.loc[player_
#    suspension_df["player"] = suspension_df.player.replace({dude:player_name_df.l
```

In [691]:
```python
# replacing player names with player id
for dude in player_name_df["player"]:
    arrest_df["player"] = arrest_df.player.replace({dude:player_name_df.loc[playe
```

...

In [692]:
```python
#replacing college name with college id
for school in college_df["school"]:
    draft_df["college"] = draft_df.college.replace({school:college_df.loc[colleg

for position in pos_df["position"]:
    draft_df["pos"] = draft_df.pos.replace({position:pos_df.loc[pos_df["position"
    arrest_df["pos"] = arrest_df.pos.replace({position:pos_df.loc[pos_df["positie

arrest_df.head()
```

. . .

In [696]:
```python
arrest_df.head()
```

Out[696]:

	date	team_id	player_id	pos_id	case_id	category_id	desc	outcome
5	2/25/2017	14	8284	8	0	4	Suspected of stealing golf cart, driving drunk...	Resolution undetermined.
6	2/25/2017	3	7891	11	0	0	Pulled over in Miami Beach, accused of possess...	Resolution undetermined.
7	2/16/2017	24	6385	6	1	5	Accused of aggravated assault after altercatio...	Charges dismissed.
8	2/16/2017	17	6421	3	0	6	Suspected of violating protective order in Nas...	Resolution undetermined.
10	1/10/2017	18	8104	5	0	7	Accused of fleeing law enforcement, careless d...	Resolution undetermined.

In [695]:
```python
#replacing arrest dataframe case, category, and outcome with relational ids
for case in case_df["result"]:
    arrest_df["case"] = arrest_df.case.replace({case:case_df.loc[case_df["result"

for category in category_df["cat_type"]:
    arrest_df["category"] = arrest_df.category.replace({category:category_df.loc

arrest_df.rename(columns={"team":"team_id",
                          "player":"player_id",
                          "pos":"pos_id",
                          "case":"case_id",
                          "category":"category_id"}, inplace=True)
```

. . .

In [698]: `arrest_df.head()`

Out[698]:

	date	team_id	player_id	pos_id	case_id	category_id	desc	outcome
5	2/25/2017	14	8284	8	0	4	Suspected of stealing golf cart, driving drunk...	Resolution undetermined.
6	2/25/2017	3	7891	11	0	0	Pulled over in Miami Beach, accused of possess...	Resolution undetermined.
7	2/16/2017	24	6385	6	1	5	Accused of aggravated assault after altercatio...	Charges dismissed.
8	2/16/2017	17	6421	3	0	6	Suspected of violating protective order in Nas...	Resolution undetermined.
10	1/10/2017	18	8104	5	0	7	Accused of fleeing law enforcement, careless d...	Resolution undetermined.

In [699]:
```python
draft_df = draft_df.dropna(how="any")
draft_df["round_drafted"] = draft_df["round_drafted"].astype(int)
draft_df["age"] = draft_df["age"].astype(int)
draft_df["college"] = draft_df["college"].astype(int)
draft_df.columns=["player_id", "year_drafted", "round_drafted", "team_id", "pos_
draft_df.head()
```

Out[699]:

	player_id	year_drafted	round_drafted	team_id	pos_id	age	college_id
0	8434	2015	1	30	0	21	0
1	8433	2015	1	31	0	21	1
2	8432	2015	1	15	1	21	2
3	8431	2015	1	27	2	21	3
4	8430	2015	1	32	3	23	4

```
In [700]: arrest_df = arrest_df.rename(columns={"date":"arrest_date",
                                                "desc":"arrest_desc"})
          arrest_df.head()
```

Out[700]:

	arrest_date	team_id	player_id	pos_id	case_id	category_id	arrest_desc	outcome
5	2/25/2017	14	8284	8	0	4	Suspected of stealing golf cart, driving drunk...	Resolution undetermined.
6	2/25/2017	3	7891	11	0	0	Pulled over in Miami Beach, accused of possess...	Resolution undetermined.
7	2/16/2017	24	6385	6	1	5	Accused of aggravated assault after altercatio...	Charges dismissed.
8	2/16/2017	17	6421	3	0	6	Suspected of violating protective order in Nas...	Resolution undetermined.
10	1/10/2017	18	8104	5	0	7	Accused of fleeing law enforcement, careless d...	Resolution undetermined.

```
In [701]: #loading time
          engine = create_engine("mysql+pymysql://root:"+pw+"@localhost:3306/nfl_arrest_dr
          engine.table_names()
```

```
Out[701]: ['arrest_data',
           'cases',
           'category',
           'college',
           'draft_data',
           'player_names',
           'teams']
```

```
In [702]: college_df.index.rename("id", inplace=True)
          draft_df.index.rename("id", inplace=True)
          arrest_df.index.rename("id", inplace=True)
          suspension_df.index.rename("id", inplace=True)
          case_df.index.rename("id", inplace=True)
          category_df.index.rename("id", inplace=True)
          teams.index.rename("id", inplace=True)
          player_name_df.index.rename("id", inplace=True)
```

In [703]:
```python
draft_df.to_sql(name="draft_data", con=engine, if_exists="append", index=True)
arrest_df.to_sql(name="arrest_data", con=engine, if_exists="append", index=True)
college_df.to_sql(name="college", con=engine, if_exists="append", index=True)
case_df.to_sql(name="cases", con=engine, if_exists="append", index=True)
category_df.to_sql(name="category", con=engine, if_exists="append", index=True)
teams.to_sql(name="teams", con=engine, if_exists="append", index=True)
player_name_df.to_sql(name="player_names", con=engine, if_exists="append", index=
```

In [704]:
```python
draft_df.head()
```

Out[704]:

id	player_id	year_drafted	round_drafted	team_id	pos_id	age	college_id
0	8434	2015	1	30	0	21	0
1	8433	2015	1	31	0	21	1
2	8432	2015	1	15	1	21	2
3	8431	2015	1	27	2	21	3
4	8430	2015	1	32	3	23	4

In [705]:
```python
arrest_df["pos_id"].unique()
```

Out[705]:
```
array([ 8, 11,  6,  3,  5,  4,  2,  1, 12,  0, 18, 13, 10, 16,  9, 14, 19,
       17], dtype=int64)
```

In [ ]: