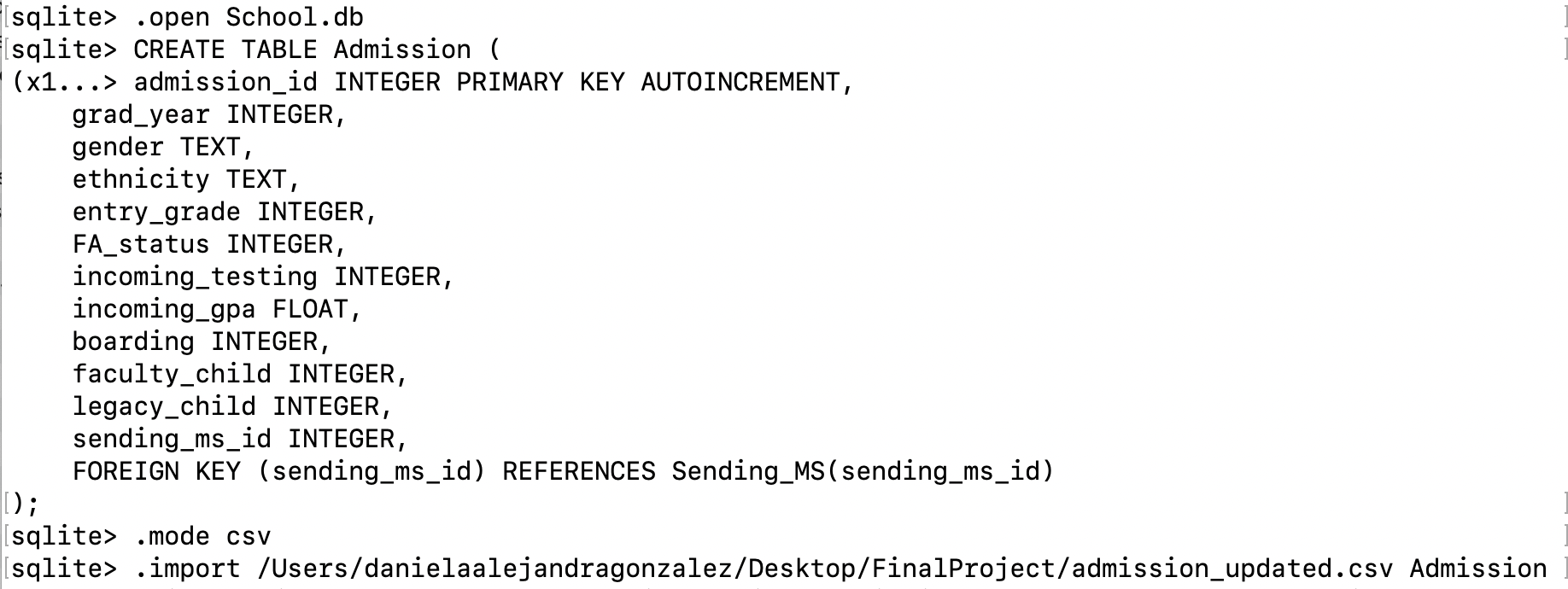
Code for the creation of the RDB Admission\_Office

All the creation of it was made using SQLITE in the terminal.

1. Admission \_Office table

,

sqlite> CREATE TABLE Admission\_Office (

admission\_id INTEGER PRIMARY KEY AUTOINCREMENT,

grad\_year INTEGER,

gender TEXT,

ethnicity TEXT,

entry\_grade INTEGER,

FA\_status INTEGER,

incoming\_testing INTEGER,

incoming\_gpa REAL,

boarding INTEGER,

faculty\_child INTEGER,

legacy\_child INTEGER,

sending\_ms\_id INTEGER,

FOREIGN KEY (sending\_ms\_id) REFERENCES Sending\_MS(sending\_ms\_id)

);

For this table I worked on the original Admission file and made some changes to the data type in the Booleans. I applied some one-encoding for the boarding, faculty\_child, legacy\_child, by replacing with 1 and 0:

admissions['boarding'] = admissions['boarding'].replace({True: 1, False: 0})

admissions['faculty\_child'] = admissions['faculty\_child'].replace({True: 1, False: 0})

admissions['legacy\_child'] = admissions['legacy\_child'].replace({True: 1, False: 0})

The table in the RDB was populated with data from the admission\_updated.csv file .

1. Creation of the Sending\_MS table

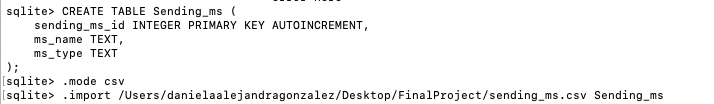
CREATE TABLE Sending\_ms (

sending\_ms\_id INTEGER PRIMARY KEY AUTOINCREMENT,

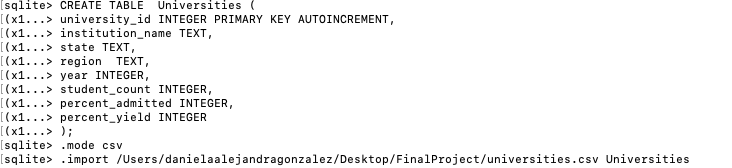
ms\_name TEXT,

ms\_type TEXT

);



1. Universities table



sqlite> CREATE TABLE Universities (

(x1...> university\_id INTEGER PRIMARY KEY AUTOINCREMENT,

(x1...> institution\_name TEXT,

(x1...> state TEXT,

(x1...> region TEXT,

(x1...> year INTEGER,

(x1...> student\_count INTEGER,

(x1...> percent\_admitted INTEGER,

(x1...> percent\_yield INTEGER

(x1...> );

sqlite> .mode csv

sqlite> .import /Users/danielaalejandragonzalez/Desktop/FinalProject/universities.csv Universities

1. College table creation

For this one I replace the Booleans in the last three columns as well as replaced the nan values for sat and act with zero.  
data['sat']=data['sat'].replace(np.nan, 0)

data['act']=data['act'].replace(np.nan, 0)

data['hook\_athlete']= data['hook\_athlete'].replace({True: 1, False: 0})

data['hook\_first\_gen']= data['hook\_first\_gen'].replace({True: 1, False: 0})

data['hook\_legacy']= data['hook\_legacy'].replace({True: 1, False: 0})

I considered the float values as REAL in the table creation and saved the new file as updated\_college:

sqlite> CREATE TABLE College (

(x1...> college\_id INTEGER PRIMARY KEY AUTOINCREMENT,

(x1...> admission\_id INTEGER,

(x1...> gpa REAL,

(x1...> sat REAL,

(x1...> act REAL,

(x1...> matric TEXT NOT NULL,

(x1...> university\_id INTEGER,

(x1...> hook\_athlete INTEGER,

(x1...> hook\_first\_gen INTEGER,

(x1...> hook\_legacy INTEGER,

(x1...> FOREIGN KEY (admission\_id) REFERENCES Admission(admission\_id),

(x1...> FOREIGN KEY ( university\_id) REFERENCES Universities(university\_id)

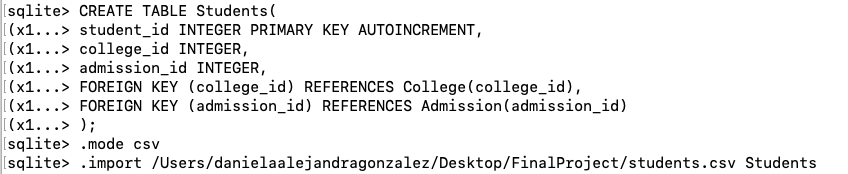
(x1...> );

sqlite> .mode csv

sqlite> .import /Users/danielaalejandragonzalez/Desktop/FinalProject/college\_updated.csv College

1. Students table

For the creation of this table there were not null values nor duplicated ones. The original file was used.



sqlite> CREATE TABLE Students(

(x1...> student\_id INTEGER PRIMARY KEY AUTOINCREMENT,

(x1...> college\_id INTEGER,

(x1...> admission\_id INTEGER,

(x1...> FOREIGN KEY (college\_id) REFERENCES College(college\_id),

(x1...> FOREIGN KEY (admission\_id) REFERENCES Admission(admission\_id)

(x1...> );

sqlite> .mode csv

sqlite> .import /Users/danielaalejandragonzalez/Desktop/FinalProject/students.csv Students

1. Applications table

sqlite> CREATE TABLE Applications(

(x1...> application\_id INTEGER PRIMARY KEY AUTOINCREMENT,

(x1...> student\_id INTEGER,

(x1...> university\_id INTEGER,

(x1...> admitted INTEGER,

(x1...> ed INTEGER,

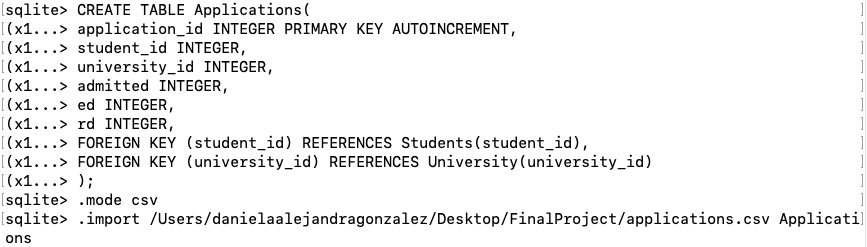
(x1...> rd INTEGER,

(x1...> FOREIGN KEY (student\_id) REFERENCES Students(student\_id),

(x1...> FOREIGN KEY (university\_id) REFERENCES University(university\_id)

(x1...> );

sqlite> .mode csv

sqlite> .import /Users/danielaalejandragonzalez/Desktop/FinalProject/applications.csv Applications

For this table no changes were made at the original file since no encoding was needed as well as no NaN values nor duplicated ones.