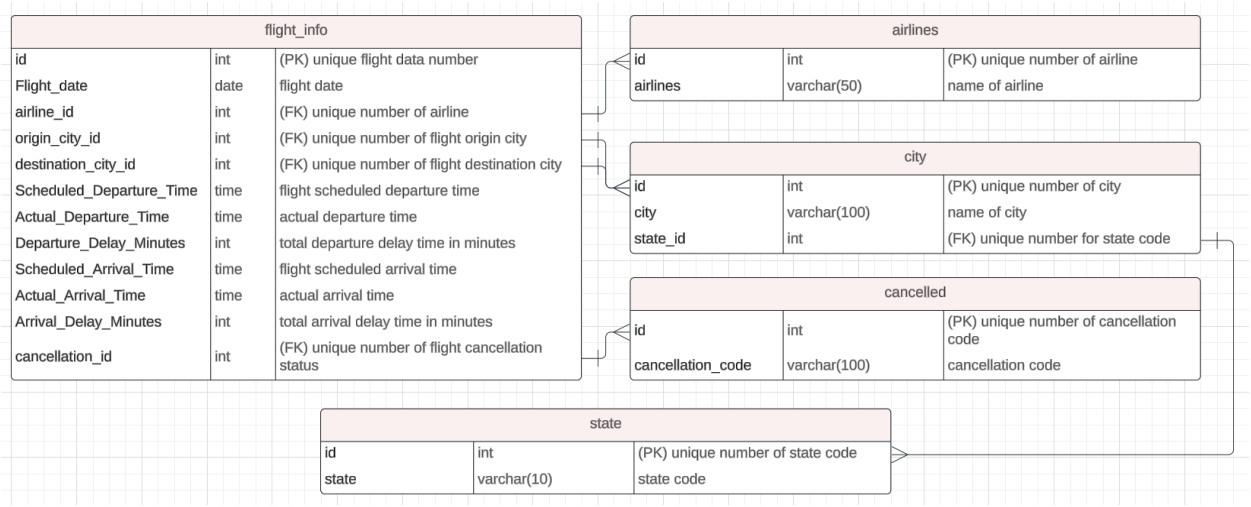
Flight Data ERD



The database contains a total of 5 tables; a main table that contains the complete information on flights and 4 other tables to ensure that the database is in a normalized format. The main table, flight_info has one primary key, which is the ID of each flight, and this table 4 foreign keys. These foreign keys are the airline_id which connects to the airlines table, origin_city_id, and destination_city_id which both connect to the city table and the same column as both columns contain the list of cities, and cancellation_id which connects to the canceled table.

The airlines, canceled, and state tables all contain only two columns, where the id of each table is the primary key, followed by the second column which contains the complete information such as the name of airlines, cancellation code, and state.

The city table has three columns, the first two columns are the same as the airlines, canceled, and state tables, where the first column is the primary key of the table, and the second column is the column that contains information which in this case is the name of the city. But the third column is a foreign key that connects to the state table where this column contains the unique number of the state code.

This database does not include all of the columns from the original dataset because the remaining columns go more in-depth into the cancellation reason such as how many hours and minutes of the delay was caused by the Weather and many more. It will be interesting to analyze more but that means that more tables will have to be created to ensure that they are in normalized form. Hence, the selected columns in the database are the main or important columns in the data.