SECTION 1

Q1)

a) The data type I would be working with would be is the airline data. The data elucidates different airlines of The United States of America that departed from an airport in The USA in the year 2015. The data pictures information like the date of departure, airline name, flight number, tail number, origin airport and much more. Moreover, the data would also cover information like why the flight was cancelled, travel time, etc (Free Data Sets & Dataset Samples, 2020).

b) The reason why this data is interesting is that we can dive into ocean of information about the airline structure USA has. For example, we can find out the frequency of an airline from a particular airport from USA. We can find out much more information about each flight, airport, weather, etc (Free Data Sets & Dataset Samples, 2020).

c) We can generate a lot of information from this dataset from this dataset. For example, we can calculate the frequency of flight delays, what was the maximum reason for flight cancellations, frequency of departure of an airline from an airport in the USA, total distance a flight has travelled by referencing their airline number, which flight took maximum time to reach its destination by comparing it to the average time, what was the most common reason behind the cancellation of a flights, which flight had the highest airtime, etc (Free Data Sets & Dataset Samples, 2020).

SECTION 2

a) There are total 4 tables in the proposed database contains:

i) flights: The table flights contain 32 entities. The entities are: ID_FLIGHT (An auto incrementing
attribute column which is created by me that is used to give every flight a flight number),YEAR, MONTH, DAY, DAY_OF_WEEK, AIRLINE,FLIGHT_NUMBER, TAIL_NUMBER, ORIGIN_AIRPORT, DESTINATION_AIRPORT, SCHEDULED_DEPARTURE, DEPARTURE_TIME, DEPARTURE_DELAY, TAXI_OUT, WHEELS_OFF, SCHEDULED_TIME, ELAPSED_TIME, AIR_TIME, DISTANCE, WHEELS_ON, TAXI_IN, SCHEDULED_ARRIVAL, ARRIVAL_TIME, ARRIVAL_DELAY, DIVERTED, CANCELLED, CANCELLATION_REASON, AIR_SYSTEM_DELAY, SECURITY_DELAY, AIRLINE_DELAY, LATE_AIRCRAFT_DELAY, WEATHER_DELAY (Free Data Sets & Dataset Samples, 2020). This table has about 500 rows.

ii) airports: The table airport has 2 different entities. The entities are: IATA_CODE, STATE (Free Data Sets & Dataset Samples, 2020). This table 321 rows.

iii) airlines: The table airlines have 2 entities named IATA_CODE and AIRLINE (Free Data Sets & Dataset Samples, 2020). This table has 14 rows.

iv) cancellation_codes: The table cancellation_codes have 2 entities named CANCELLATION_REASON and CANCELLATION_DESCRIPTION (Free Data Sets & Dataset Samples, 2020). This table has 4 rows.

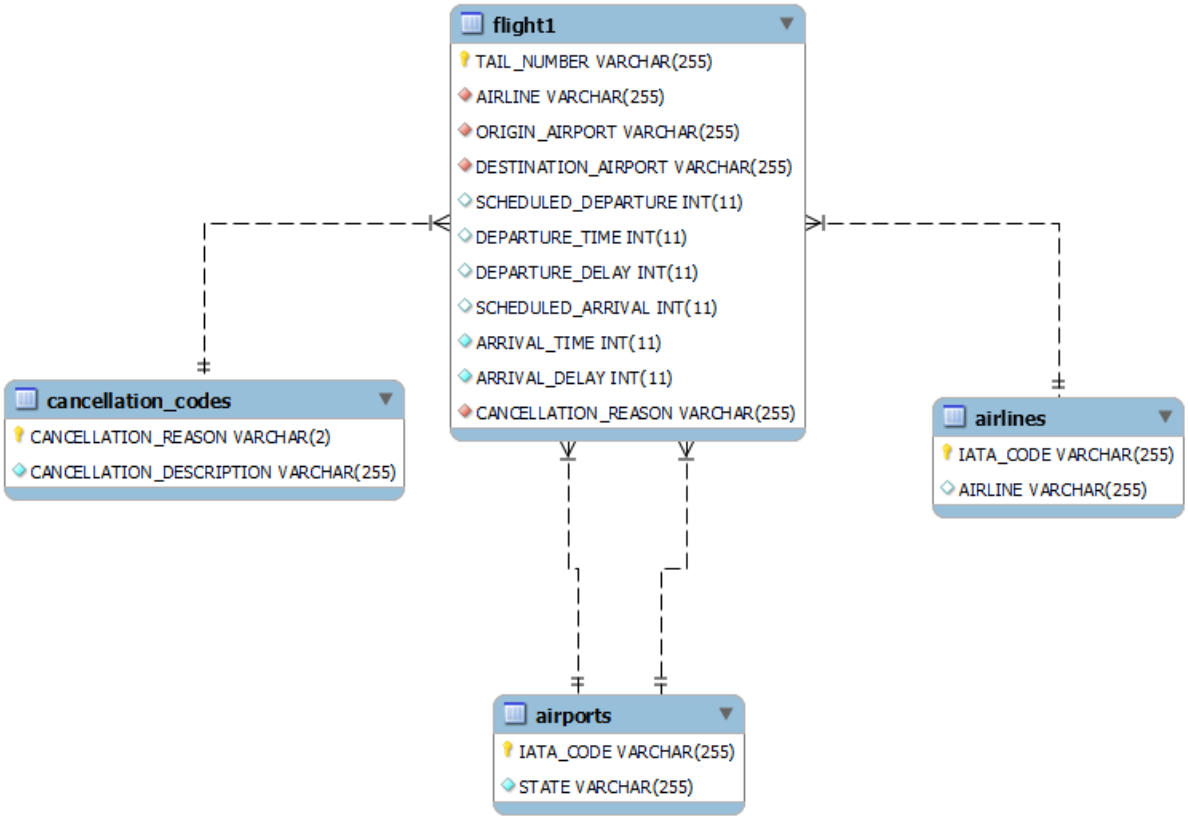
NOTE: I didn't generate the data, all the data I used and will be using in the further assignments is obtained from (Free Data Sets & Dataset Samples, 2020).

Here is how my tables are connected:

The flight table has one primary key which is TAIL_NUMBER and 4 foreign keys which are ORIGIN_AIRPORT, DESTINATION_AIRPORT, AIRLINE and CANCELLATION_REASON. The ORIGIN_AIRPORT and DESTINATION_AIRPORT foreign keys references to airports table's IATA_CODE which has one primary key named IATA_CODE. The AIRLINE foreign key references to IATA_CODE in airlines table which has IATA_CODE as the primary key. The CANCELLATION_REASON foreign key references to cancellation_codes table which has the CANCELLATION_REASON as the primary key. In the flight1 table, TAIL_NUMBER is the primary key that uniquely identifies each record in the flight1 table. ORIGIN_AIRPORT, DESTINATION_AIRPORT, SCHEDULED_DEPARTURE, SCHEDULED_ARRIVAL, DEPARTURE_TIME, DEPARTURE_DELAY, ARRIVAL_TIME, ARRIVAL_DELAY, and CANCELLATION_REASON are functionally dependent on TAIL_NUMBER. ORIGIN_AIRPORT and DESTINATION_AIRPORT denote the departure and arrival locations, SCHEDULED_DEPARTURE and SCHEDULED_ARRIVAL indicate the planned departure and arrival times, DEPARTURE_TIME and ARRIVAL_TIME represent the actual times of departure and arrival, DEPARTURE_DELAY and ARRIVAL_DELAY record any deviations from the scheduled times. Finally, CANCELLATION_REASON specifies the reason for flight cancellation. In the airlines table, IATA_CODE is the primary key that uniquely identifies each record in the airlines table. AIRLINE is functionally dependent on IATA_CODE, stating the full name of the airline corresponding to its IATA code. For example, UA corresponds to United Air Lines Inc., AA corresponds to American Airlines Inc., and etc. In the airports table, IATA_CODE is the primary key that uniquely identifies each record in the airports table. STATE is functionally dependent on IATA_CODE, indicating the state in which the airport is located. For example, ABE corresponds to Pennsylvania (PA), ABI corresponds to Texas (TX), etc. Each airport is connected to a specific state, helping to locate airports all across the United States. In the cancellation_codes table, CANCELLATION_REASON is the primary key that uniquely identifies each record in the cancellation_codes table. CANCELLATION_DESCRIPTION, including categories such as Airline/Carrier, Weather, National Air System, Security, and Not Cancelled, which is functionally dependent on CANCELLATION_REASON, indicating the reason due to which the flights are cancelled. For example, the CANCELLATION_REASON, A describes that the flight was cancelled due to Airline/Carrier reasons.

SECTION 3

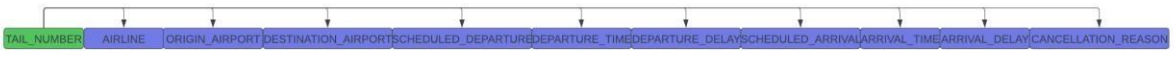
Here is the internal schema for my dataset:



The flight table has one primary key which is TAIL_NUMBER and 4 foreign keys which are

ORIGIN_AIRPORT, DESTINATION_AIRPORT, AIRLINE and CANCELLATION_REASON. The

ORIGIN_AIRPORT and DESTINATION_AIRPORT foreign keys references to airports table's IATA_CODE which has one primary key named IATA_CODE. The AIRLINE foreign key references to IATA_CODE in airlines table which has IATA_CODE as the primary key. The CANCELLATION_REASON foreign key references to cancellation_codes table which has the CANCELLATION_REASON as the primary key.



In the flight1 table, TAIL_NUMBER is the primary key that uniquely identifies each record in the flight1 table. ORIGIN_AIRPORT, DESTINATION_AIRPORT, SCHEDULED_DEPARTURE,

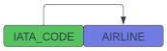
SCHEDULED_ARRIVAL, DEPARTURE_TIME, DEPARTURE_DELAY, ARRIVAL_TIME,

ARRIVAL_DELAY, and CANCELLATION_REASON are functionally dependent on TAIL_NUMBER.

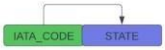
ORIGIN_AIRPORT and DESTINATION_AIRPORT denote the departure and arrival locations,

SCHEDULED_DEPARTURE and SCHEDULED_ARRIVAL indicate the planned departure and arrival times,

DEPARTURE_TIME and ARRIVAL_TIME represent the actual times of departure and arrival, DEPARTURE_DELAY and ARRIVAL_DELAY record any deviations from the scheduled times. Finally, CANCELLATION_REASON specifies the reason for flight cancellation.



In the airlines table, IATA_CODE is the primary key that uniquely identifies each record in the airlines table. AIRLINE is functionally dependent on IATA_CODE, stating the full name of the airline corresponding to its IATA code. For example, UA corresponds to United Air Lines Inc., AA corresponds to American Airlines Inc., and etc.



In the airports table, IATA_CODE is the primary key that uniquely identifies each record in the airports table. STATE is functionally dependent on IATA_CODE, indicating the state in which the airport is located. For example, ABE corresponds to Pennsylvania (PA), ABI corresponds to Texas (TX), etc. Each airport is connected to a specific state, helping to locate airports all across the United States.

In the cancellation_codes table, CANCELLATION_REASON is the primary key that uniquely identifies each record in the cancellation_codes table. CANCELLATION_DESCRIPTION, including categories such as Airline/Carrier, Weather, National Air System, Security, and Not Cancelled, which is functionally dependent on CANCELLATION_REASON, indicating the reason due to which the flights are cancelled.

For example, the CANCELLATION_REASON, A describes that the flight was cancelled due to Airline/Carrier reasons.

References:

Free Data Sets & Dataset Samples. (2020, September 10). *Airline Flight Delays*. Maven Analytics. https://mavenanalytics.io/data-playground?dataStructure=2lXwWbWANQgI727tVx3DRC&page=5&pageSize=5