Assignment 3 - Airline Satisfaction System

About

This is a Airline Satisfaction Database System mainly designed to analyze the passenger reviews for different airlines on factors such as service, food, entertainment, wifi and other services provided by the various airlines.

Overview

- Data for passenger reviews extracted from Skytrax Website.
- Python script written in Jupyter notebookto extract data using python library Beautiful Soup, Scrapy, Selenium.
- Flight Details and information extracted from Kayak Website.
- Further Datasets collected from GitHub, Kaggle and other sources.
- Jupyter Scripts created to cleaning, munging the extracted data.
- Script created to populate data collected after scrapping to the Airline Database.

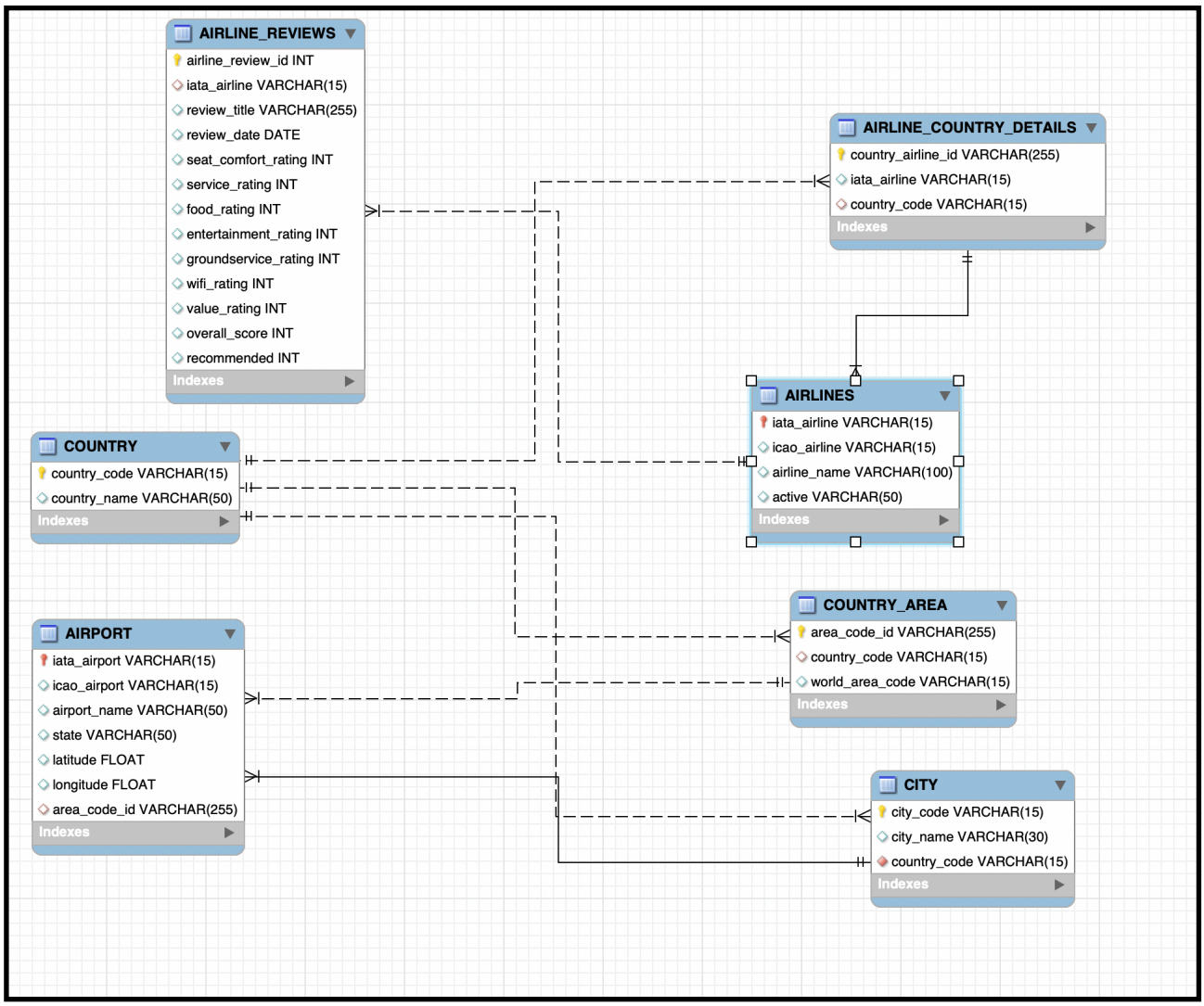
Data Sources

Source	Type
Kayak	Website (kayak)
Datasets(txt,csv) from Kaggle etc	File
Twitter	API
Skytrax	Website

Tables

- country
- country_area
- city
- airport
- airline_country_detail
- airline
- airline_reviews
- aircraft
- flights

ER diagram



DDL Statements

CREATE DATABASE airline_passenger_satisfaction_system;

CREATE TABLE country
(country_code VARCHAR(15) UNIQUE,
country_name VARCHAR(50),
PRIMARY KEY (country_code));

CREATE TABLE country_area
(area_code_id VARCHAR(255) UNIQUE,
country_code VARCHAR(15),
world_area_code VARCHAR(15),
PRIMARY KEY (area_code_id),

```
FOREIGN KEY(country_code) REFERENCES COUNTRY(country_code) ON DELETE
CASCADE);

CREATE TABLE city
(city_code VARCHAR(15) NOT NULL,
city_name VARCHAR(30),
country_code VARCHAR(15) NOT NULL,
PRIMARY KEY (city_code),
FOREIGN KEY(country_code) REFERENCES COUNTRY(country_code) ON DELETE
CASCADE);

CREATE TABLE airport
(iata_airport VARCHAR(15) UNIQUE,
icao_airport VARCHAR(15),
airport_name VARCHAR(50),
state VARCHAR(50),
latitude float,
longitude float,
area_code_id VARCHAR(255),
PRIMARY KEY (iata_airport),
FOREIGN KEY(area_code_id) REFERENCES country_area(area_code_id),
FOREIGN KEY(iata_airport) REFERENCES city(city_code));

ALTER TABLE AIRPORT
ADD FOREIGN KEY(area_code_id) REFERENCES COUNTRY_AREA(area_code_id);

ALTER TABLE AIRPORT
ADD FOREIGN KEY(iata_airport) REFERENCES CITY(city_code);

CREATE TABLE airline_country_detail
(country_airline_id VARCHAR(255) UNIQUE,
iata_airline VARCHAR(15) UNIQUE,
country_code VARCHAR(15),
PRIMARY KEY (country_airline_id),
FOREIGN KEY(country_code) REFERENCES country(country_code));

CREATE TABLE airline
(iata_airline VARCHAR(15) UNIQUE,
icao_airline VARCHAR(15),
airline_name VARCHAR(100),
active VARCHAR(50),
PRIMARY KEY (iata_airline),
FOREIGN KEY(iata_airline) REFERENCES airline_country_detail(iata_airline));

ALTER TABLE AIRLINES
ADD FOREIGN KEY(iata_airline) REFERENCES
AIRLINE_COUNTRY_DETAILS(iata_airline);
```

```
CREATE TABLE airline_reviews
(airline_review_id INT NOT NULL AUTO_INCREMENT,
iata_airline VARCHAR(15) UNIQUE,
review_title VARCHAR(255),
review_date DATE,
seat_comfort_rating INT,
service_rating INT,
food_rating INT,
entertainment_rating INT,
groundservice_rating INT,
wifi_rating INT,
value_rating INT,
overall_score INT,
recommended INT,
PRIMARY KEY (airline_review_id),
FOREIGN KEY(iata_airline) REFERENCES airline(iata_airline));

CREATE TABLE IF NOT EXISTS aircraft(
aircraft_iata CHAR(3),
aircraft_icao CHAR(5),
aircraft_name VARCHAR(50),
capacity int,
PRIMARY KEY (aircraft_iata)
);

CREATE TABLE IF NOT EXISTS flights(
flight_id INT AUTO_INCREMENT,
aircraft_iata CHAR(3),
flight_number INT,
iata_airline VARCHAR(15),
PRIMARY KEY (flight_id),
FOREIGN KEY (aircraft_iata) REFERENCES aircraft(aircraft_iata),
FOREIGN KEY (iata_airline) REFERENCES airline(iata_airline)
);

#INSERT INTO FLIGHTS
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("100",900,"CE");
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("340",300,"AF");
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("142",300,"AF");
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("142",900,"BA");
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("340",200,"BA");
INSERT INTO flights(aircraft_iata,flight_number,iata_airline) values("100",800,"CE");
```

DDL Statements

#1 Count all aircrafts owned by airlines which have a seat capacity>150

```
SELECT t1.airline_name, t2.aircraft_name, count(t2.aircraft_name) as airline_count, t2.capacity
FROM airline t1
JOIN flights t3
ON t1.iata_airline = t3.iata_airline
JOIN aircraft t2
ON t2.aircraft_iata = t3.aircraft_iata
WHERE t2.capacity>150
GROUP by t1.airline_name, t2.aircraft_name, t2.capacity;
```

#2 Display airline names and the number of positive recommendations receieved

```
SELECT t1.airline_name, COUNT(t2.recommended) no_of_recommendation
FROM airline t1
JOIN airline_reviews t2
ON (t1.iata_airline = t2.iata_airline)
WHERE t2.recommended = True
GROUP BY t1.airline_name
ORDER BY no_of_recommendation DESC ;
```

#3 How many passenger flights do Nationwide Airlines own?

```
SELECT t2.airline_name, count(t2.airline_name) AS flight_count
FROM flights t1
INNER JOIN airline t2
ON t1.iata_airline=t2.iata_airline
INNER JOIN aircraft t3
ON t1.aircraft_iata=t3.aircraft_iata
WHERE t3.capacity>0 AND t1.iata_airline="CE"
```

#4 Which airline received maximum average seating comfort rating ?

```
SELECT t1.airline_name, AVG(t2.seat_comfort_rating) AS avg_seat_comfort_rating
FROM airline t1
JOIN airline_reviews t2
ON (t1.iata_airline = t2.iata_airline)
GROUP BY t1.airline_name
limit 1;
```

#5 Find all airlines owning 2 passenger flights

```
SELECT t2.airline_name, count(t2.airline_name) AS airline_count
FROM flights t1
INNER JOIN airline t2
ON t1.iata_airline=t2.iata_airline
```

```
INNER JOIN aircraft t3
ON t1.aircraft_iata=t3.aircraft_iata
WHERE t3.capacity>0
GROUP BY t2.airline_name
HAVING count(t2.airline_name)=2;
```

#6 How many reviews does Air France have with >4.0 rating

```
SELECT
t1.iata_airline, t1.review_title, t1.overall_score
FROM airline_reviews t1
INNER JOIN (SELECT *
        FROM airline
        WHERE airline_name like '%air france%'
        LIMIT 1) as t2
ON t1.iata_airline=t2.iata_airline
WHERE t1.overall_score>8;
```

#7 Find top 5 airlines with best overall reviews

```
SELECT t2.airline_name, AVG(t1.overall_score) as avg_overall_score
FROM airline_reviews t1
INNER JOIN airline t2
ON t1.iata_airline=t2.iata_airline
GROUP BY t2.airline_name;
```

#8 Find the airline which provides the best entertainment in flights along with their reviews

```
SELECT t2.airline_name,t1.review_title, MAX(avg_rating) as max_average_rating  FROM
                                                (SELECT
review_title,iata_airline,AVG(entertainment_rating) AS avg_rating
                                                 FROM airline_reviews
                                                 GROUP BY iata_airline,review_title ) AS t1
JOIN airline t2
ON t1.iata_airline=t2.iata_airline
GROUP BY t2.airline_name,t1.review_title
ORDER BY max_rating DESC
limit 1;
```

#9 order airline names based on the count of their service rating for service rating > 4

```
SELECT t1.airline_name, COUNT(t2.service_rating) AS service_rating_count
FROM airline t1
JOIN airline_reviews t2
USING(iata_airline)
WHERE t2.service_rating > 4
GROUP BY t1.airline_name
ORDER BY service_rating_count DESC;
```

#10 Count the number of aircraft each flight has

```
SELECT t1.flight_number, COUNT(t2.aircraft_name) AS aircraft_count
FROM flights t1
JOIN aircraft t2
ON t1.aircraft_iata = t2.aircraft_iata
GROUP BY t1.flight_number;
```