Background Information

Since the implementation of the Airline Deregulation Act of 1978, American airlines have been free to set their own fares and routes. The application of market forces to the airline industry has had a number of substantial impacts, first in shaping the hub-and-spoke airport system and allowing the growth of the legacy carriers, and more recently, in the increase of low-cost carriers.

Problem Statement

Over the past 20 years, there has been a dramatic realignment in the air transportation industry brought on by economic and demographic changes. Many of the legacy carriers have reorganized while several low-cost carriers have expanded. In this assignment, students will analyze government statistics to look at the costs of flying and to see how various types of carriers compare.

Instructions

IMPORTANT: This assignment requires the Windows version of Microsoft 365. macOS users can access a ready-to-use installation through Windows Virtual Desktop by following the instructions at https://cs101.wvu.edu/wvd.

IMPORTANT: Complete the steps below in the order they are given. Completing the steps out of order may complicate the assignment or result in an incorrect result.

1. Download and extract the provided Data Files ZIP file. It contains the following files for use in this assignment:
   a. **origincities.xml** – Information on major origin cities in the continental United States [1]–[3].

Table: *OriginCities*		
Field Name	**Type**	**Description**
OriginCityCode	Short Text	Primary key. FAA designator code for primary airport in city (e.g., PIT for Pittsburgh International Airport).
City	Short Text	Location with city and state.
Population	Number	2022 population estimate for the city's metropolitan area or county.
Hub	Yes/No	If yes, city has a hub airport.

b. **destinationcities.xml** – Information on major destination cities in the continental United States [1]–[3].

Table: *DestinationCities*		
Field Name	**Type**	**Description**
DestinationCityCode	Short Text	Primary key. FAA designator code for primary airport in city (e.g., PIT for Pittsburgh International Airport).
City	Short Text	Location with city and state.
Population	Number	2022 population estimate for the city's metropolitan area or county.
Hub	Yes/No	If yes, city has a hub airport.

c. **airfare.xml** – Data including fare (ticket price) information for major route pairs in the United States as well as information on market share. Figures are from the 4th Quarter 2022 [4].

Table: *Airfare*		
Field Name	**Type**	**Description**
OriginCityCode	Short Text	Part of composite key. FAA designator code of origin city.
DestinationCityCode	Short Text	Part of composite key. FAA designator code of destination city.
Distance	Number	Distance (in miles) between airport pairs.
Passengers	Number	Number of passengers flying this route annually.
Fare	Currency	Average fare (ticket price) of all airlines on route.
LargestAirlineCode	Short Text	Code of airline with largest market share.
LargestAirlineShare	Percentage	Percentage of market share held by largest airline.
LargestAirlineFare	Currency	Average fare for largest airline.
CheapestAirlineCode	Short Text	Code of airline with cheapest fare.
CheapestAirlineShare	Percentage	Percentage of market share held by cheapest airline.
CheapestAirlineFare	Currency	Average fare for cheapest airline.

d. **largestairlines.xml** – Information on airlines that had the largest market share for their routes.

Table: *LargestAirlines*		
Field Name	**Type**	**Description**
LargestAirlineCode	Short Text	Primary key. Code of airline with largest market share (e.g., "DL" for Delta Air Lines).
AirlineName	Short Text	Name of the airline.
AirlineType	Short Text	Type of the airline (Legacy, Low-Cost, or Other).

2. Create a new Microsoft Access database named **hw3_*lastname*_*firstname*_app.accdb** .

3. Import the following items into the database:
   a. **originairports.xml** file – Import structure and data into a new table.
   b. **destinationairports.xml** file – Import structure and data into a new table.
   c. **airfare.xml** file – Import structure and data into a new table.
   d. **largestairlines.xml** file – Import structure and data into a new table.
4. We need to create a table to store data on the airlines.
   a. Create a table named *CheapestAirlines* with the fields below. Use appropriate field types and designate a good primary key.

Table: *CheapestAirlines*	
Field Name	**Description**
CheapestAirlineCode	2-character code of airline with cheapest average fare (e.g., "DL" for Delta Air Lines).
AirlineName	Full name of airline.
AirlineType	Type of the airline (Legacy, Low-Cost, or Other).

   b. Enter records for all cheapest airlines below.

HINT: The *CheapestAirlines* table will contain 13 records.

CheapestAirlineCode	AirlineName	AirlineType
3M	Silver Airways	Other
AA	American Airlines	Legacy
AS	Alaska Airlines	Other
B6	JetBlue Airlines	Low-Cost
DL	Delta Air Lines	Legacy
F9	Frontier Airlines	Low-Cost
G4	Allegiant Air	Low-Cost
MX	Breeze Airways	Low-Cost
NK	Spirit Airlines	Low-Cost
SY	Sun Country Airlines	Low-Cost
UA	United Air Lines	Legacy
WN	Southwest Airlines	Low-Cost
XP	Avelo Airlines	Low-Cost

5. Create appropriate relationships for the following tables. Do not enforce referential integrity.
   a. *OriginCities* and *Airfare*
   b. *DestinationCities* and *Airfare*
   c. *LargestAirlines* and *Airfare*
   d. *CheapestAirlines* and *Airfare*

6. Create separate queries to provide the information requested below. Name each query after the step in which it appears (e.g., name the query in Step 6a as *Query6A*).

   HINT: Run your queries to test them. Make sure that they display all and only the records that you would expect to appear.

   a. Create a query displaying information on all flight routes. List the origin city name, the destination city name, largest carrier name, and cheapest carrier name.

      Sort by origin city name in ascending order.

      HINT: This query will show 5,778 records and 4 fields.

   b. We wish to compare the types of airlines flying each route. Create a query listing the origin city name; destination city name; the name, type, and share of the largest airline; and the name, type, and share of the cheapest airline.

      Only display records where the cheapest airline type is low-cost carrier.

      Sort by origin city name and then by destination city name, both in ascending order.

      HINT: This query will show 2,710 records and 8 fields.

   c. Create a query to view information on cities with hub airports. List the origin city name, if the origin city has a hub, destination city name, if the destination city has a hub, passengers, and fare.

      Only display records where at least one of the cities (origin and/or destination) has a hub.

      Sort by fare in descending order.

      HINT: This query will show 3,022 records and 6 fields.

d. We wish to calculate the revenue miles per route. Create a query listing the origin city name, destination city name, distance, and passengers. Also, include a field to calculate the revenue miles per day.

You can calculate the revenue miles per day using the formula:

$$\frac{[Airfare.Distance] * [Airfare.Passengers]}{365}$$

Format the calculated field as a standard-type number.

HINT: This query will show 5,778 records and 5 fields.

e. Create a query to calculate the cost-per-mile for flight routes. List the origin city name, destination city name, distance, passengers, and fare. Also, include a field to calculate the cost-per-mile.

You can calculate the cost-per-mile using the formula:

$$\frac{[Airfare.Fare]}{[Airfare.Distance]}$$

Only display routes that are either less than 250 miles (< 250) or more than 1,500 miles (> 1500) long.

Format the calculated field as currency with 3 decimal places.

HINT: This query will show 1,493 records and 6 fields.

f. We wish to determine the average distance for flights departing from each airport. Create a query listing, for each origin city name, its FAA code, if it has a hub, and the average distance of its departing flights.

Format the average distance as a standard-type number with no decimal places.

HINT: This query will show 231 records and 4 fields.

g. We want to calculate statistics about the departing passengers per resident for each origin city. Copy-and-paste this SQL code into a new query:

```
SELECT OriginCities.City, OriginCities.OriginCityCode,
OriginCities.Hub, OriginCities.Population,
Sum(Airfare.Passengers) AS AnnualPassengers,
Round([AnnualPassengers]/[Population],2) AS
PassengersPerResident
FROM OriginCities INNER JOIN Airfare ON
OriginCities.OriginCityCode = Airfare.OriginCityCode
GROUP BY OriginCities.City, OriginCities.OriginCityCode,
OriginCities.Hub, OriginCities.Population
ORDER BY OriginCities.City;
```

IMPORTANT: Do not make any modifications to this query other than entering the above SQL code.

HINT: This query will show 231 records and 6 fields.

7. We need to create a new table to store analysis questions responses.

   a. Create a table named *AnalysisQuestions* with fields below. Use appropriate field types and designate a good primary key.

Table: *AnalysisQuestions*	
Field Name	**Description**
QuestionNumber	Question being answered.
Response	Response to the analysis question prompt.

8. In the *AnalysisQuestions* table, answer three of the five analysis questions below. Respond to one question per record.

   a. Flights between cities less than 200 miles apart tended to have a fairly low number of passengers. Name at least one reason why this might be the case.

   b. While people are normally price-conscious, there are many cases where the carrier with the largest market share on a route is not the cheapest carrier. Why might this be the case?

   c. How does the cost per mile on long flights (over 1,500 miles) compare to the cost on short flights (under 250 miles)? What is a possible explanation for any differences in the cost per mile?

   d. Is there any relationship between the ticket price and the number of people flying a given route? Why might this be so?

e. The median departing passengers-per-resident for all airports is 0.58. What are some reasons why some airports have extremely high passengers-per-resident ratios and others have extremely low ratios?

9. Run the Compact and Repair Database utility on your database. Ignore any errors you receive when running the utility.

Grading Rubric

This assignment is worth 60 points. It will be graded by your instructor using this rubric, with partial credit awarded as appropriate:

Step 3	2 points	Step 6g	3 points
Steps 4a-b	5 points total	Step 7a	3 points
Step 5	3 points	Steps 8a-e (pick 3 of 5)	3 points each
Steps 6a-f	5 points each		

The analysis questions in Steps 8a-e will be evaluated using this rubric:

Standard	Meets Requirements (1.5 points)	Does Not Meet Requirements (0 points)
Answer is reasonable.	Answer addresses the question prompt and is factually correct or a reasonable interpretation of available data.	Answer does not address the question prompt, is factually incorrect, or is an unreasonable interpretation of available data.
Answer is supported.	Logical rationale is provided to support the given answer.	Logical rationale is not provided to support the given answer.

References

[1] "2022 Population and Housing Unit Estimates Tables," U.S. Census Bureau, Washington, DC, Mar. 2023. Available: *https://www.census.gov/programs-surveys/popest/data/tables.html*.

[2] "List of hub airports," *Wikipedia*, Jul. 20, 2023. Available: *https://en.wikipedia.org/w/index.php?title=List_of_hub_airports&oldid=1166312239*.

[3] "List of airports in the United States," *Wikipedia*, May 16, 2023. Available: *https://en.wikipedia.org/w/index.php?title=List_of_airports_in_the_United_States&oldid=1155065613*.

[4] "Consumer Airfare Report: Table 6 - Contiguous State City-Pair Markets That Average At Least 10 Passengers Per Day," U.S. Department of Transportation, Washington, DC, Text, May 2023. Available: *https://www.transportation.gov/policy/aviation-policy/domestic-airline-consumer-airfare-report*.