Project Description:
As travel picks up again post-COVID, we are interested in designing a web app that integrates openflight data with Skytrax airline reviews to help travelers optimize their travel experience. Features include helping users find airlines for specific routes, find new destination ideas, and/or choose the best airline/airport based on historical travel experience ratings.

Group Members:
Name: Zhongzheng Bai
Email Address: baizz@seas.upenn.edu
GitHub username: baizhongzheng

Name: Si Yuan Liu
Email Address: syliu@seas.upenn.edu
GitHub username: syliu-mcit

Name: Min Fu
Email Address: minfu@seas.upenn.edu
GitHub username: mfu17

Name: Ziqi Wang
Email Address: wangziqi@seas.upenn.edu
GitHub username: zickywang

Description of application/website idea:

Dataset 1
Description: three datasets containing airport, airline and routes with each entry representing 1) a unique airport with IATA & ICAO code and location details, 2) a unique airline with IATA & ICAO code and operational status, 3) a unique source-destination combo for an airline.
Link: https://openflights.org/data.html
Size statistics: Airports: 1MB (7698 rows, 14 attributes), Airlines: 400KB (6162 rows, 8 attributes); routes: 2MB (67663 rows, 9 attributes)
Summary statistics:

Airport - latitude mean: 25.81, count: 7698, stdv: 28.4
Longitude mean: -1.39, count: 7698, stdv: 86.5
Altitude: mean: 1015.87, count: 7698, stdv: 1628.8
Routes - stops: mean: 0.000163, count: 67663, stdv: 0.013
Airlines - count: 6162, count of “Y”: 1255

Dataset 2
Description: four datasets from 2015 scrapped from Skytrax on airline reviews
Link: https://github.com/quankiquanki/skytrax-reviews-dataset
Size statistics: 34MB

- Airline reviews: 41,396 rows

- Airport reviews: 17,721 rows
- Seat reviews: 1,258 rows
- Lounge reviews: 2,264 rows

Summary statistics:

```
]: airline.describe().transpose()
```

]:

	count	mean	std	min	25%	50%	75%	max
overall_rating	36861.0	6.039527	3.214680	1.0	3.0	7.0	9.0	10.0
seat_comfort_rating	33706.0	3.094612	1.405515	0.0	2.0	3.0	4.0	5.0
cabin_staff_rating	33708.0	3.319212	1.541307	0.0	2.0	4.0	5.0	5.0
food_beverages_rating	33264.0	2.805886	1.580246	0.0	1.0	3.0	4.0	5.0
inflight_entertainment_rating	31114.0	2.392364	1.704753	0.0	1.0	2.0	4.0	5.0
ground_service_rating	2203.0	2.736723	1.569073	1.0	1.0	3.0	4.0	5.0
wifi_connectivity_rating	565.0	2.249558	1.541283	1.0	1.0	1.0	4.0	5.0
value_money_rating	39723.0	3.164111	1.523486	0.0	2.0	4.0	4.0	5.0
recommended	41396.0	0.533820	0.498861	0.0	0.0	1.0	1.0	1.0

In [317]:
```
airport.describe().transpose()
```

Out[317]:

	count	mean	std	min	25%	50%	75%	max
overall_rating	13796.0	4.274355	2.722765	1.0	2.0	4.0	6.0	10.0
queuing_rating	12813.0	2.747912	1.572520	0.0	1.0	3.0	4.0	5.0
terminal_cleanliness_rating	12815.0	3.442450	1.337508	0.0	3.0	3.0	5.0	5.0
terminal_seating_rating	587.0	2.580920	1.403862	0.0	1.0	2.0	4.0	5.0
terminal_signs_rating	27.0	2.592593	1.393923	1.0	1.0	3.0	4.0	5.0
food_beverages_rating	630.0	2.169841	1.534358	0.0	1.0	2.0	3.0	5.0
airport_shopping_rating	12676.0	2.821631	1.410575	0.0	2.0	3.0	4.0	5.0
wifi_connectivity_rating	412.0	2.405340	1.579452	0.0	1.0	2.0	4.0	5.0
airport_staff_rating	26.0	2.038462	1.248384	1.0	1.0	1.5	3.0	4.0
recommended	17721.0	0.221206	0.415071	0.0	0.0	0.0	0.0	1.0

In [318]:
```
lounge.describe().transpose()
```

Out[318]:

	count	mean	std	min	25%	50%	75%	max
overall_rating	2259.0	3.365649	1.485086	1.0	2.0	3.0	4.0	10.0
comfort_rating	2264.0	3.341873	1.322064	0.0	2.0	4.0	4.0	5.0
cleanliness_rating	2264.0	3.658569	1.252673	0.0	3.0	4.0	5.0	5.0
bar_beverages_rating	2259.0	3.231076	1.385324	0.0	2.0	3.0	4.0	5.0
catering_rating	2261.0	2.781955	1.476588	0.0	1.0	3.0	4.0	5.0
washrooms_rating	2238.0	3.023235	1.483449	0.0	2.0	3.0	4.0	5.0
wifi_connectivity_rating	2253.0	3.285397	1.515094	0.0	2.0	4.0	4.0	5.0
staff_service_rating	2255.0	3.253215	1.435761	0.0	2.0	3.0	5.0	5.0
recommended	2264.0	0.360424	0.480230	0.0	0.0	0.0	1.0	1.0

In [319]:
```
seat.describe().transpose()
```

Out[319]:

	count	mean	std	min	25%	50%	75%	max
overall_rating	1257.0	4.318218	3.041998	1.0	1.0	4.0	7.0	10.0
seat_legroom_rating	1258.0	2.753577	1.446915	1.0	1.0	3.0	4.0	5.0
seat_recline_rating	1258.0	2.627186	1.267095	1.0	1.0	3.0	4.0	5.0
seat_width_rating	1258.0	2.717011	1.271828	1.0	2.0	3.0	4.0	5.0
aisle_space_rating	1258.0	2.730525	1.343137	1.0	1.0	3.0	4.0	5.0
viewing_tv_rating	1229.0	2.872254	1.454787	0.0	2.0	3.0	4.0	5.0
power_supply_rating	62.0	3.774194	1.310980	1.0	3.0	4.0	5.0	5.0
seat_storage_rating	113.0	3.070796	1.334413	1.0	2.0	3.0	4.0	5.0
recommended	1258.0	0.364070	0.481360	0.0	0.0	0.0	1.0	1.0

Queries for the dataset:

1. Find the top airlines with the best cabin staff ratings
```
Select airline_names From Reviews
Order by Cabin_staff_rating DESC
Limit 10
```
2. If I only fly Delta Airlines, show all the routes operated by the airline, which route is the most popular and well rated?
```
Select * from routes R
Join airports A on R.source_airport_id = A.airport_id
Join airlines AL on R.airline_id = AL.airline_id
```
3. For all airlines that have flights departing from New York LaGuardia, show their overall ratings from highest to lowest.
4. Which country has the overall best airport facilities?
5. What airlines have the broadest coverage in the world? By routes, by distance traveled, by destination countries etc.
6. If I want to fly from New York to London, which airport and airline combination is preferred?
7. If I want to depart from Los Angeles, what are some popular destinations in a user specified country or continent?
8. If I don’t mind having one or two layovers, what routes can I fly with XYZ airline between destination A and B?