Milestone 2 Project Outline

1. Motivation for the idea/description of the problem the application solves

As the world reemerges from pandemic lockdowns and slowly returns to normalcy, an enormous pent-up demand in travel is being unleashed. Being avid explorers ourselves, we understand that flying is an integral part of any trip planning so we are keen to design a web application to help travelers optimize their flying experience. The application will integrate airport, airline and route data from OpenFlight, customer review data from SkyTrax, and flight delay data (https://www.kaggle.com/datasets/neel90/flight-delay-2019, new data source introduced containing ~7m rows) published by the U.S. government to help users find the most optimal destinations, routes or airline based on a user's needs and preferences.

2. List of features you will definitely implement in the application

- Pages 1 through 4 describe features that we think are essential to the application

3. List of features you might implement in the application, given enough time

- Page 5 include some features that we will attempt to deploy if there is sufficient time

4. List of pages the application will have and a 1-2 sentence description of each page. We expect that the functionality of each page will be meaningfully different from the functionality of the other pages. (At least 3 pages)

Page 1: Airline Reviews

- Airlines are reviewed across multiple dimensions, we want this page to educate users on which airlines score well and poorly. The questions we want to answer are
- Which airlines are considered the best by travelers?
- Which are each airline's strengths and weaknesses: seat / food / comfort / entertainment etc?
- Which airlines have the broadest coverage?
- If I am a frequent traveler from the US, which airline should I use as my frequent flier airline?

Page 2: Airport Reviews

- Apart from the in-flight experience, your time in the airport is equally important and this can include check-in, amenities, dining, lounges etc. In NYC for example, travelers have access to LGA, JFK and EWR and anyone that travels frequently can tell you the experiences vary widely across the multiple dimensions. The questions we want to answer are
- Which country has the best airports?
- What are the top airports in a country or city?
- What do people say they like or dislike about a particular airport?
- Is there any correlation between the number of flights/routes and how well an airport scores?

Page 3: Route Explorer

- Airlines constantly adapt to the changing behaviors of travelers and we want to keep people up to date on the possibilities of exploring every part of the world.

- If you are based in Singapore, what are all of the non-stop destinations (countries) available to me?
- If I want to travel between NYC and London, how many routes are available?
- If I only want to fly a specific airline, what is their route map like?
- Which are the most popular routes in the world?
- What are the longest or shortest routes in the world?
- Which routes are most susceptible to delays?

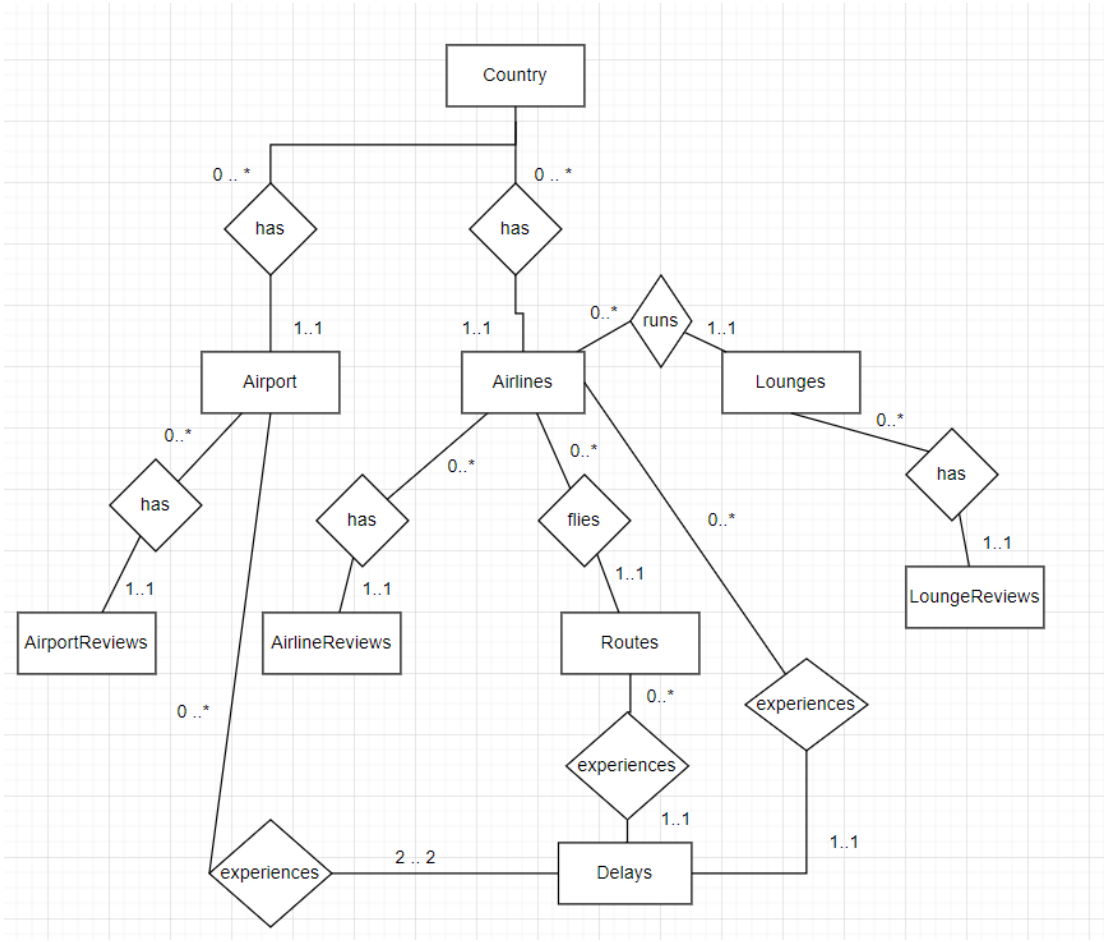
Page 4: Trip Explorer

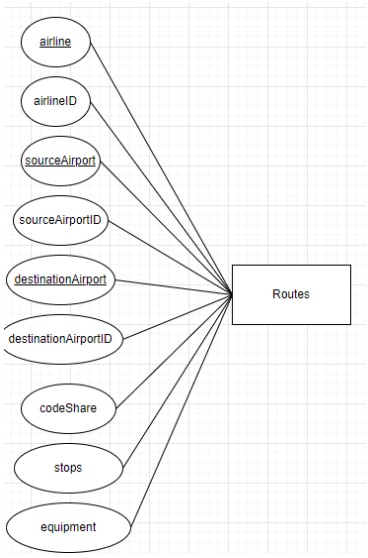
- When people travel, they usually have an idea of the departure and destination but can be facing tens or hundreds of combinations between airports and airlines. This page intends to recommend the most optimal airport and/or airline based user's preference such as minimizing delays. As an example, a traveler commuting between NYC and London has at least 20 airlines to choose from that offer flights from one of LGA, JFK and EWR airports in NYC to LHR, LGW, LCY in London.
- If my preference is the airport experience, which routes are best?
- If my preference is in-flight dining, which route should I fly?
- If I care about the airport experience and want to minimize the likelihood of flight delays, which airport/airline combination should I take?
- If I am open to layovers (0, 1 or 2), what combinations can I choose from while minimizing flight distance?

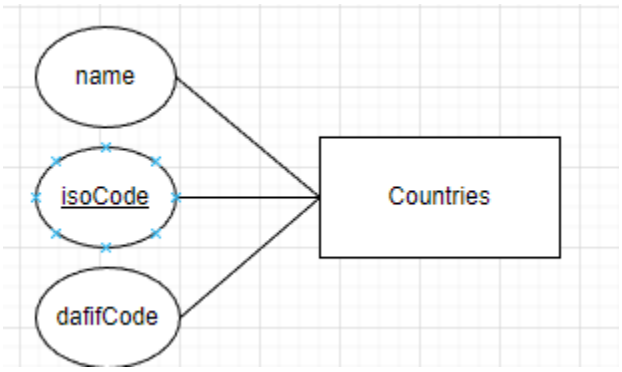
Page 5: Trip Recommender

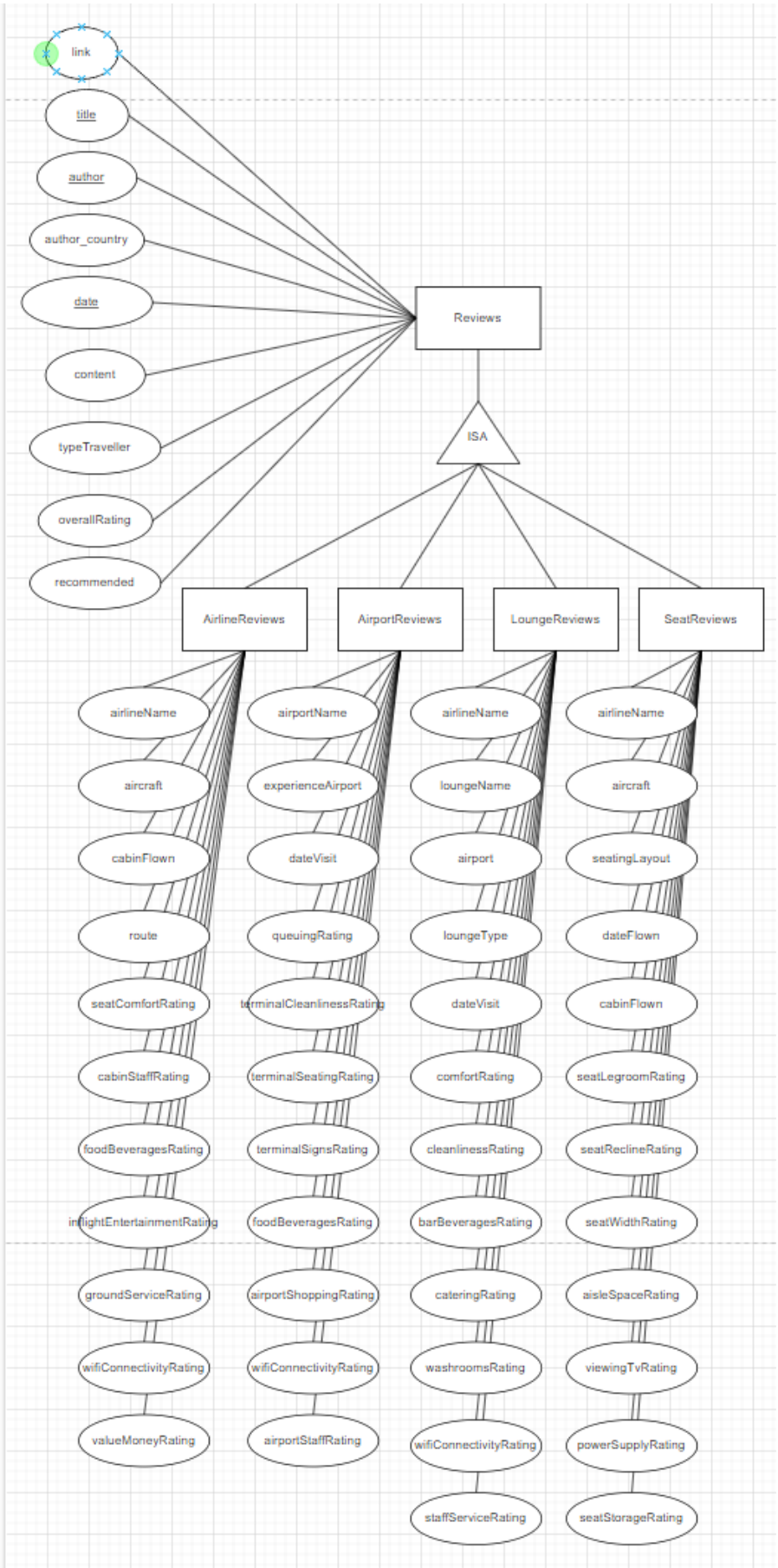
- This page can be thought of as being similar to the "I'm Feeling Lucky" page on Google. If I am retired or a student on a gap year and don't have constraints and just want to explore the world, where are some destinations I should consider from an origin city/country.
- If I can take off from Toronto to anywhere in the world, where can I go other than the obvious destinations such as the U.S. and Mexico.
- This page will also give users some optionality to express their preferences.

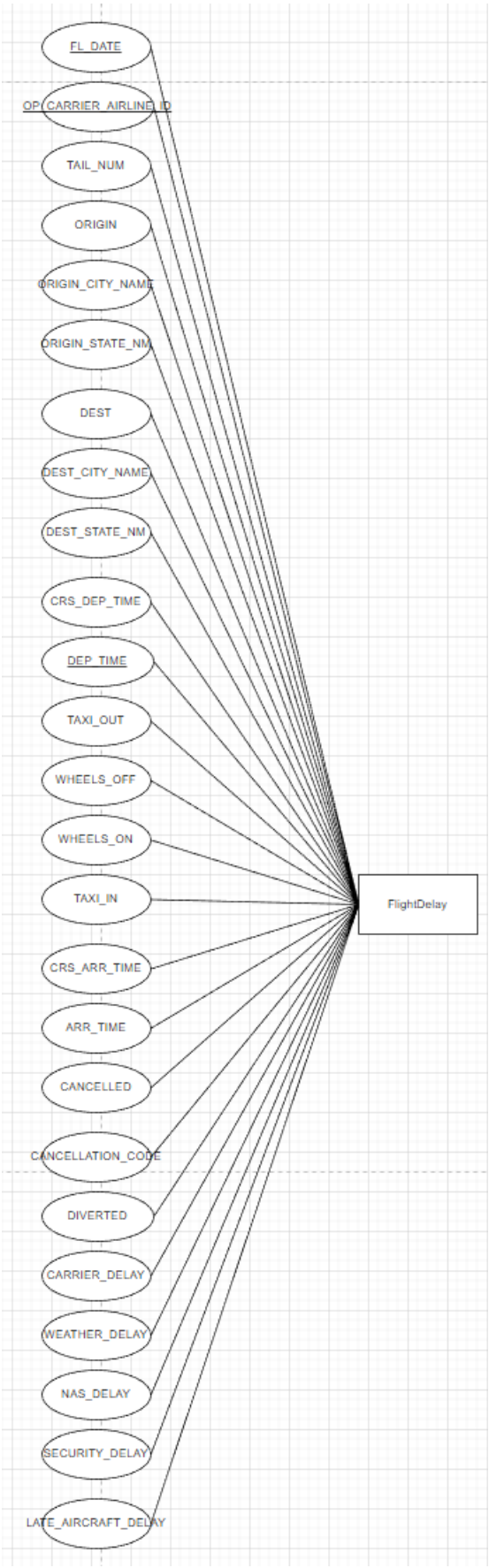
5. Relational schema as an ER diagram

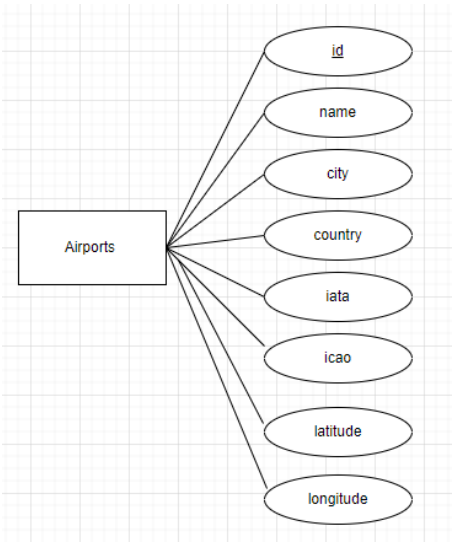


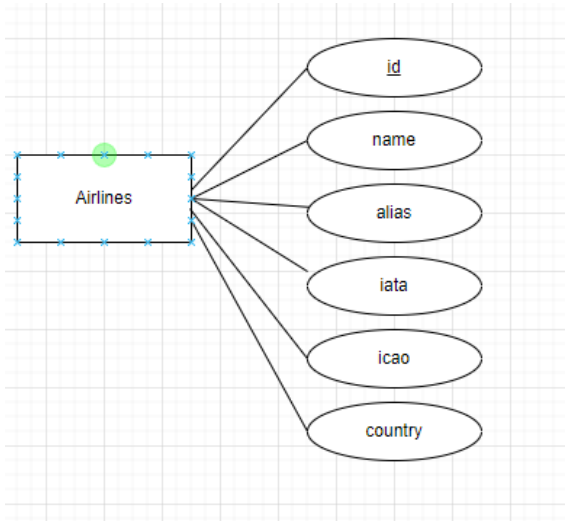


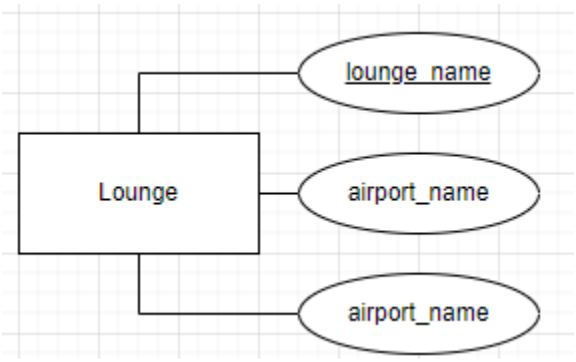



link
title
author
author_country
date
content
typeTraveller
overallRating
recommended
Reviews
ISA
AirlineReviews
AirportReviews
LoungeReviews
SeatReviews
airlineName
aircraft
cabinFlown
route
seatComfortRating
cabinStaffRating
foodBeveragesRating
inflightEntertainmentRating
groundServiceRating
wifiConnectivityRating
valueMoneyRating
airportName
experienceAirport
dateVisit
queuingRating
terminalCleanlinessRating
terminalSeatingRating
terminalSignsRating
foodBeveragesRating
airportShoppingRating
wifiConnectivityRating
airportStaffRating
airlineName
loungeName
airport
loungeType
dateVisit
comfortRating
cleanlinessRating
barBeveragesRating
cateringRating
washroomsRating
wifiConnectivityRating
staffServiceRating
airlineName
aircraft
seatingLayout
dateFlown
cabinFlown
seatLegroomRating
seatReclineRating
seatWidthRating
aisleSpaceRating
viewingTvRating
powerSupplyRating
seatStorageRating


FL_DATE
OP_CARRIER_AIRLINE_ID
TAIL_NUM
ORIGIN
ORIGIN_CITY_NAME
ORIGIN_STATE_NM
DEST
DEST_CITY_NAME
DEST_STATE_NM
CRS_DEP_TIME
DEP_TIME
TAXI_OUT
WHEELS_OFF
WHEELS_ON
TAXI_IN
CRS_ARR_TIME
ARR_TIME
CANCELLED
CANCELLATION_CODE
DIVERTED
CARRIER_DELAY
WEATHER_DELAY
NAS_DELAY
SECURITY_DELAY
LATE_AIRCRAFT_DELAY
FlightDelay







6. SQL DDL for creating the database

```
CREATE DATABASE flights; USE flights;

CREATE TABLE Routes (
    airline                     VARCHAR(30),
    airlineID                   VARCHAR(30),
    sourceAirport               VARCHAR(30),
    sourceAirportID             VARCHAR(30),
    destinationAirport          VARCHAR(30),
    destinationAirportID        VARCHAR(30),
    codeShare                   VARCHAR(30),
    stops                       INT,
    equipment                   VARCHAR(30),
    PRIMARY KEY (airline, sourceAirport, destinationAirport))
    FOREIGN KEY (airlineID) references Airlines(id),
                (sourceAirportID) references Airports(id),
                (destinationAirportID) references Airport(id)
)
```

```
CREATE TABLE Airports(
    id VARCHAR(30),
    name VARCHAR(30),
    city VARCHAR(30),
    country VARCHAR(30),
    iata VARCHAR(3),
    icao VARCHAR(4),
    latitude DECIMAL(10,5),
    longitude DECIMAL(10,5),
    PRIMARY KEY (id)
)

CREATE TABLE Airlines(
    id VARCHAR(30),
    name VARCHAR(30),
    alias VARCHAR(30),
    iata VARCHAR(2),
    icao VARCHAR(3),
    country VARCHAR(50),
    PRIMARY KEY (airlineID)
)

CREATE TABLE AirportMapping(
     iata VARCHAR(3),
     airportName VARCHAR(50),
     PRIMARY KEY (iata),
     FOREIGN KEY (iata) references Airports(iata),
                 (airportName) references AirportReviews(airportName)
)

CREATE TABLE AirlineMap(
     iata VARCHAR(3),
     airlineName VARCHAR(50),
     PRIMARY KEY (iata),
     FOREIGN KEY (iata) references Airlines(iata),
                 (airportName) references AirlineReviews(airlineName)
)

CREATE TABLE Lounge(
     lounge_name VARCHAR(30),
     airline_name VARCHAR(30),
     airport_name VARCHAR(50),
     PRIMARY KEY (lounge_name)
)
```

```
CREATE TABLE Countries (
    name                        VARCHAR(30),
    isoCode                     VARCHAR(30),
    dafifCode                   VARCHAR(30),
    PRIMARY KEY (isoCode)
)

CREATE TABLE AirlineReviews (
    link                        VARCHAR(30),
    title                       VARCHAR(30),
    author                      VARCHAR(30),
    author_country              VARCHAR(30),
    date                        DATE,
    content                     VARCHAR(200),
    typeTraveller               VARCHAR(30),
    overallRating               DECIMAL(19, 1),
    recommended                 INT,
    PRIMARY KEY (title,author,date),
    airlineName                 VARCHAR(30),
    aircraft                    VARCHAR(30),
    cabinFlown                  VARCHAR(30),
    route                       VARCHAR(30),
    seatComfortRating           DECIMAL(19, 1),
    cabinStaffRating            DECIMAL(19, 1),
    foodBeveragesRating         DECIMAL(19, 1),
    inflightEntertainmentRating DECIMAL(19, 1),
    groundServiceRating         DECIMAL(19, 1),
    wifiConnectivityRating      DECIMAL(19, 1),
    valueMoneyRating            DECIMAL(19, 1))

CREATE TABLE AirportReviews (
    link                        VARCHAR(30),
    title                       VARCHAR(30),
    author                      VARCHAR(30),
    author_country              VARCHAR(30),
    date                        DATE,
    content                     VARCHAR(200),
    typeTraveller               VARCHAR(30),
    overallRating               DECIMAL(19, 1),
    recommended                 INT,
    PRIMARY KEY (title,author,date),
    airportName                 VARCHAR(50),
    experienceAirport           VARCHAR(30),
    dateVisit                   DATE
    queueingRating              DECIMAL(19, 1),
```

```
    terminalCleanlinessRating   DECIMAL(19, 1),
    terminalSeatingRating       DECIMAL(19, 1),
    terminalSignsRating         DECIMAL(19, 1),
    foodBeveragesRating         DECIMAL(19, 1),
    airportShoppingRating       DECIMAL(19, 1),
    wifiConnectivityRating      DECIMAL(19, 1),
    airportStaffRating          DECIMAL(19, 1))

CREATE TABLE LoungeReviews (
    link                        VARCHAR(30),
    title                       VARCHAR(30),
    author                      VARCHAR(30),
    author_country              VARCHAR(30),
    date                        DATE,
    content                     VARCHAR(200),
    typeTraveller               VARCHAR(30),
    overallRating               DECIMAL(19, 1),
    recommended                 INT,
    PRIMARY KEY (title,author,date),
    airlineName                 VARCHAR(30),
    loungeName                  VARCHAR(30),
    airport                     VARCHAR(30),
    loungeType                  VARCHAR(30),
    dateVisit                   DATE,
    comfortRating               DECIMAL(19, 1),
    cleanlinessRating           DECIMAL(19, 1),
    barBeveragesRating          DECIMAL(19, 1),
    cateringRating              DECIMAL(19, 1),
    washroomsRating             DECIMAL(19, 1),
    wifiConnectivityRating      DECIMAL(19, 1),
    staffServiceRating          DECIMAL(19, 1)
)

CREATE TABLE SeatReviews (
    link                        VARCHAR(30),
    title                       VARCHAR(30),
    author                      VARCHAR(30),
    author_country              VARCHAR(30),
    date                        DATE,
    content                     VARCHAR(200),
    typeTraveller               VARCHAR(30),
    overallRating               DECIMAL(19, 1),
    recommended                 INT,
    PRIMARY KEY (title,author,date),
    airlineName                 VARCHAR(30),
```

```
    aircraft                    VARCHAR(30),
    seatingLayout               VARCHAR(30),
    dateFlown                   DATE,
    cabinFlown                  VARCHAR(30),
    seatLegroomRating           DECIMAL(19, 1),
    seatReclineRating           DECIMAL(19, 1),
    seatWidthRating             DECIMAL(19, 1),
    aisleSpaceRating            DECIMAL(19, 1),
    viewingTvRating             DECIMAL(19, 1),
    powerSupplyRating           DECIMAL(19, 1),
    seatStorageRating           DECIMAL(19, 1)
)

CREATE TABLE FlightDelay (
    FL_DATE                     DATE,
    OP_CARRIER_AIRLINE_ID       VARCHAR(30),
    TAIL_NUM                    VARCHAR(30),
    ORIGIN                      VARCHAR(30),
    ORIGIN_CITY_NAME            VARCHAR(30),
    ORIGIN_STATE_NM             VARCHAR(30),
    DEST                        VARCHAR(30),
    DEST_CITY_NAME              VARCHAR(30),
    DEST_STATE_NM               VARCHAR(30),
    CRS_DEP_TIME                DECIMAL(19, 1),
    DEP_TIME                    DECIMAL(19, 1),
    TAXI_OUT                    DECIMAL(19, 1),
    WHEELS_OFF                  DECIMAL(19, 1),
    WHEELS_ON                   DECIMAL(19, 1),
    TAXI_IN                     DECIMAL(19, 1),
    CRS_ARR_TIME                DECIMAL(19, 1),
    ARR_TIME                    DECIMAL(19, 1),
    CANCELLED                   DECIMAL(19, 1),
    CANCELLATION_CODE           VARCHAR(30),
    DIVERTED                    DECIMAL(19, 1))
    CARRIER_DELAY               DECIMAL(19, 1),
    WEATHER_DELAY               DECIMAL(19, 1),
    NAS_DELAY                   DECIMAL(19, 1),
    SECURITY_DELAY              DECIMAL(19, 1),
    LATE_AIRCRAFT_DELAY         DECIMAL(19, 1),
    PRIMARY KEY (FL_DATE,OP_CARRIER_AIRLINE_ID,DEP_TIME)
)
```

7. Explanation of how you will clean and pre-process the data. This tutorial demonstrates how to do simple pre-processing in Python.

1. Missing values: we will use discretion to remove rows with missing values in our data. The flight reviews data do contain quite a bit of missing values.
2. Categorical data: for categorical variables, we will transform them into numerical values to reduce storage overhead and reduce the runtime of queries.
3. Outliers: for variables that have a range of allowed values, such as review score (0-10), we will remove ones that are not within the range.
4. Data type check: we will check each column for fields that contain illegal values. For example, when a review column contains string values. Depending on the specific cases, we might cast them to the correct data type, or remove rows with illegal values.
5. Find candidate index: when examining the different datasets, we will find columns with unique values as candidate indices. If a single-column index is not possible, we will combine multiple columns as candidate indices.
6. Export dataset: once the data is cleaned and pre-processed, we will export them to .csv format for uploading to our MySQL database.

8. List of technologies you will use. You must use some kind of SQL database. We recommend using MySQL specifically because you will use MySQL in HW2, and we will provide guidance for setting up a MySQL database.

1. MySQL: we will use MySQL as the backend database for our website. Our data is in tabular format, which makes it suitable for storing with SQL database.
2. Node.js
3. React.js

For the frontend of our website, we will use the Node.js and React.js libraries, which will allow us to create a simple, user-friendly interface to our website.

9. Description of what each group member will be responsible for:

We will designate duties into member 1, 2, 3 and 4 (still being finalized).

Database and schema setup: 1 & 2
Data cleaning and transformation: 3 & 4
Database population: 1 & 2
Query design: 1, 2, 3 & 4
API design: 1, 2, 3 & 4
UI design: 1 & 2
Page 1 features: 1
Page 2 features: 2
Page 3 features: 3
Page 4 features: 4