DATABASE TECHNICAL DOCUMENT

Contents

1. Introduction

This technical paper briefly describes and explains both, the logic and definitions for each component corresponding to the SQL database owned by the fictitious enterprise *JLMFCompany*. The hypothetical goal of the previously mentioned company, is to serve as a low-cost third party that allows clients to book flights with Expedia and/or reserve Airbnb listings. The purpose of this database is store information client information related with transactions done throughout the company's platform – "*Expedia Airbnb App (EApp)*".

The names of tables and columns are always written in uppercase and may include underscores. Some fields are optional therefore some cells (from CSV UTF-8 – Comma Delimited files) may be left empty depending on the circumstances; meaning that the database accepts some CSV files to be loaded with data gaps. Anyhow, most fields are mandatory and columns require to be filled appropriately with the correct format. Please note that data from the Excel/CSV files is extracted by the non-existent "*EApp*" platform.

2. List of tables by category

2.1 User and records data
- BOOKINGS
- USER_ACCOUNT
- USER_RECORDS
- CARD_DETAILS

2.2 Airbnb data
- AIRBNB_BOOKING_DATA
- AIRBNB_RESERVATION_DETAILS
- HOST_INFO
- REVIEWS

2.3 Expedia data
- EXPEDIA_PASSENGER_DATA
- EXPEDIA_PASSENGER_TICKETS
- EXPEDIA_TRIP_DETAILS

3. Data dictionary: concepts and definitions

The following sections describe the meaning behind every column used in each Microsoft Excel/CSV file and hence the database's tables (adjusted to accept every column from distinct files). Section 3.1 describes every column containing equivalent data that appears in more than one Excel file, which creates a relationship among tables. Relationships may be 1 – to – 1, 1 – to – many, or many – to – many. Sections 3.2 – 3.4 describe attributes that can be found in only one table. Please read the following:

3.1 Shared attributes or columns

3.1.1 FIRST_LAST_NAME: Concatenation of the client's first and last name in the USER_ACCOUNT table. This field matches the data from PASSENGER_NAME and GUEST_NAME, from EXPEDIA__PASSENGER_DATA and AIRBNB_RESERVATION_DETAILS tables respectively.

3.1.2 USER_ID: This information represents the user's or account holder's unique identifier and log in credential.

3.1.3 USER_NATIONALITY: The account holder's country of origin.

3.1.4 USER_PASSPORT_NO: Account holder's passport number.

3.1.5 BOOKING_ID: Unique identifier created after account holder purchased plane tickets or/and reserved an Airbnb. This unique identifier allows the third party (*JLMFCompany*) to track every purchase from customers that used the company's platform (*EApp)*. This identifier is used exclusively by *JLMFCompany*.

The three classes of BOOKING_ID are the following:

- **Class A**: refers to the case in which the user requested Expedia's service (booked a flight) only.
- **Class B**: refers to the case in which the account holder requested Airbnb's service only.
- **Class C**: situation where the user requested both Expedia's and Airbnb's services.

3.1.7 LISTING_NAME: Name of the property or physical site available for Airbnb customers. Every listing is selected by its respective host.

3.1.8 HOST_LOCATION_ID: Id number for the actual location of the Airbnb listing. Please note that one host may have several ids since he/she can have several properties (listings) available for Airbnb.

3.1.9 AIR_CONFIRMATION_ID: The guest's confirmation identifier; which is created after a host approves the account holder's request. This record is then exclusively "available" for Airbnb only.

3.1.10 HOST_ID: Host's identification number. This number is assigned by Airbnb.

3.1.11 GUEST_NAME: This field represents the same data as FIRST_LAST_NAME from the USER_ACCOUNT table. This field might be the same as PASSENGER_NAME from the EXPEDIA_PASSENGER_DATA table as well. However, this is only the case if the user booked a flight and requested an Airbnb listing at the same time.

3.1.12 PASSENGER_NAME: This field might be the same as GUEST_NAME from the AIRBNB_RESERVATION_DETAILS table as well. However, this is only the case if the account holder booked a flight and requested an Airbnb listing at the same time. This field is equivalent to FIRST_LAST_NAME from the USER_ACCOUNT table.

3.1.13 FLIGHT_CONFIRMATION_ID: This field contains a series of letters and numbers that represents the Expedia's confirmation id after the user booked his or her flight. Each confirmation id has a one-to-one or one-to-many relationships with the flight ticket ids (EXP_E_TICKET_ID). Each confirmation id includes all the passengers in each id.

3.2 User and transaction data

3.2.1 USER_NAME: This column represents the account holder's username. This information serves as the user's unique identifier and log in credential.

3.2.2 FIRST_NAME: The account holder's first name.

3.2.3 LAST_NAME: The account holder's last name.

3.2.4 USER_ADDRESS: Account holder's physical or mailing address.

3.2.5 USER_CITY: City linked with the physical or mailing address provided by user or account holder.

3.2.6 USER_STATE: State linked with the physical or mailing address, and city provided by the account holder.

3.2.7 USER_COUNTRY: Country linked to the address, city and state provided by user or account holder.

3.2.8 USER_ZIP_CODE: Zip code number linked to the user's address or P.O. Box.

3.2.9 USER_EMAIL: Account holder's email address.

3.2.10 USER_BDAY: Account holder's date of birth. The date format is YYYY-MM-DD.

3.2.11 USER_PHONE: Account holder's landline or mobile number. This is the only optional field in the USER_ACCOUNT table. The user may or may not include a phone number.

3.2.12 ORDER_NO: Unique identifier that represents the order of purchases from users in chronological order starting from the moment *JLMFCompany* "began operating". This field is the primary key for the USER_RECORDS table.

3.2.13 TRANSACTION_ID: Alpha numeric values that serve as identifiers for each transaction from clients that requested services from Expedia and/or Airbnb throughout the company's platform. This field possess a one-to-one relationship with each BOOKING_ID, but this field is exclusively "used" by *JLMFCompany*.

3.2.14 TRANSACTION_DATE: Date in which the account holder purchased a service with Expedia and/or Airbnb. The date format is YYYY-MM-DD.

3.2.15 CARD_COMPANY: The only accepted credit/debit cards are Visa, MasterCard and Discover.

3.2.16 CARD_NO: Credit or debit number provided by account holder.

3.2.17 NAME: Owner of the credit or debit card.

3.2.18 EXPIRY: Card's expiration date in MM / YY format.

3.2.19 CVV: Three-digit security code located at the back of the credit/debit card.

3.2.20 BILLING_ADDRESS: Physical or mailing address linked to the debit/credit card provided by the user.
3.2.21 BILLING_CITY: City linked to the debit/credit card.
3.2.22 BILLING_STATE: State linked to the debit/credit card.
3.2.23 BILLING_ZIP_CODE: Zip code linked to the debit/credit card.
3.2.24 BILLING_COUNTRY: Country linked to the debit/credit card.
3.2.25 BOOKING_CLASS: *Please review the BOOKING_ID definition; section 3.1.5.*

3.3 Airbnb data

3.3.1 HOST_FIRST_NAME: The host's first name.
3.3.2 HOST_LAST_NAME: The host's last name.
3.3.3 HOST_NAME: Host's first and last name linked to his/her unique HOST_ID. *Please review section 3.1.10.*
3.3.4 ADDRESS: Location of the selected Airbnb listing.
3.3.5 CITY: City linked with the location of the requested Airbnb listing.
3.3.6 STATE: State linked with the location of the requested Airbnb listing.
3.3.7 ZIP_CODE: Zip code linked with the location of the requested Airbnb listing.
3.3.8 COUNTRY: Country linked with the location of the requested Airbnb listing.
3.3.9 LISTING_TYPE: Airbnb listings are categorized as 'Entire Place', 'Private Room' and 'Shared Room'. Guests decide what category to choose depending on the number of guests and their travel-related necessity.
3.3.10 START_DATE: Initial date (selected by the account holder) when guests have access to the reserved Airbnb listing. The date format is YYYY-MM-DD.
3.3.11 END_DATE: Date (selected by the account holder) when guests must depart from the reserved Airbnb listing. The date format is YYYY-MM-DD.
3.3.12 TOTAL_GUESTS: Number of guests per reservation (AIR_CONFIRMATION_ID). *Please review section 3.1.9.*
3.3.13 GUEST_ADULT_NO: Number of adult guests (account holder included) per reservation.
3.3.14 GUEST_CHILDREN_NO: Number of minor (ages: 1 – 17) guests per reservation.
3.3.15 GUEST_INFANT_NO: Number of infant (younger than 1 year of age) guests per reservation.
3.3.16 REASON_TO_VISIT: The account holder must briefly describe the purpose of his/her/their stay.
3.3.17 REVIEWER: Any guest (not only the account holder) linked to any AIR_CONFIRMATION_ID (or reservation) is welcomed to provide a review of an Airbnb listing after departing from the property. This field does not require first and last name.
3.3.18 Mathematically-related terms:

> **3.3.18A RATING**: Guests may rate the host's property or location on a scale from 1 to 5. This field is represented by an integer.

3.3.18B RATING_OVERALL: Calculation or average obtained from the number of guests included in each AIR_CONFIRMATION_ID or reservation, that rated a listing after their departure.

3.4 Expedia (for flights only) data

3.4.1 ORIGIN: Journey's starting point or location. This field is represented by the account holder's selected airport.

3.4.2 DESTINATION: Journey's ending point or location. This field is represented by the account holder's selected city's airport.

3.4.3 DEPARTING_DATE: Selected date on behalf of the account holder, in which he or she and his/her selected passengers (if any) are expected to show up at the selected origin/airport.

3.4.4 RETURNING_DATE: Selected date on behalf of the account holder, in which he or she and his/her selected passengers (if any) are expected to show up at the selected airport with the intention of returning to the point of origin.

3.4.5 TICKET_CATEGORY: Prior to the trip date selection, account holders must choose the type of ticket they intend to purchase (one-way or round trip).

3.4.6 DEPT_FLIGHT_TYPE: The flight types (from the departing location) available are nonstop and layover.

3.4.7 DEPT_FIRST_FLIGHT_NO: This field represents the passengers' destination, only when the DEPT_FLIGHT_TYPE is labeled as a nonstop flight. However, if the flight type is labeled as layover, then this field would represent the first connecting flight.

3.4.8 DEPT_SECOND_FLIGHT_NO: This field is NULL when DEPT_FLIGHT_TYPE is labeled as a nonstop flight. However, if the flight type is labeled as layover, then this field will represent the second connecting flight.

3.4.9 DEPT_THIRD_FLIGHT_NO: This field is NULL when DEPT_FLIGHT_TYPE is labeled as a nonstop flight. However, if the flight type is labeled as layover, then this field might represent another connecting flight.

3.4.10 DEPT_TRIP_DURATION_HR: Expected amount of time (in hours) spent during the flight and layover (if any). This applies for the first phase of the trip.

3.4.11 RTRN_FLIGHT_TYPE: The flight types (towards the departing location) available are nonstop and layover. This field includes the term NA, which only applies when the passenger purchased a one-way ticket.

3.4.12 RTRN_FIRST_FLIGHT_NO: This field represents the passengers' trip back towards the origin. This field cannot be null if the TICKET_CATEGORY is labeled as a round trip. In that case, if the RTRN_FLIGHT_TYPE is labeled as layover or nonstop, then this field would represent the first connecting flight or the only direct flight, respectively.

3.4.13 RTRN_SECOND_FLIGHT_NO: This field is NULL when RTRN_FLIGHT_TYPE is labeled as a nonstop flight. However, if the return flight type is labeled as layover, then this field will represent the second connecting flight.

3.4.14 RTRN_THIRD_FLIGHT_NO: This field is NULL when RTRN_FLIGHT_TYPE is labeled as a nonstop flight. However, if the flight type is labeled as layover, then this field might represent another connecting flight.

3.4.15 RTRN_TRIP_DURATION_HR: Expected amount of time (in hours) spent during the flight and layover (if any). This field will be null or empty if the user purchased a one-way ticket.

3.4.16 EXP_E_TICKET_ID: Account holder's ticket id number created after purchasing plane tickets with Expedia. Each ID is unique per passenger.

3.4.17 EXP_E_TICKET_ID_X: Letter X represents <u>adult</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation. Every individual receives an e-ticket-id.

3.4.18 EXP_E_TICKET_ID_X_M: Letter X represents passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation. Letter M stands for <u>minors</u> under 18 years of age. Every individual receives an e-ticket-id.

3.4.19 EXP_E_TICKET_ID_X_I: Letter X represents <u>infant</u> passenger 2, up to 5, depending on the number of babies the account holder included on his/her reservation. Letter I stands for infants younger than 1 year of age. Every individual receives an e-ticket-id.

3.4.20 PASSENGER_X_NAME: Letter X represents the name of <u>adult</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation.

3.4.21 PASSENGER_X_NAME_M: Letter X represents the name of <u>minor</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation. Letter M stands for individuals under 18 years of age.

3.4.22 PASSENGER_X_NAME_I: Letter X represents the name of <u>infant</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation. Letter I stands for infants younger than 1 year of age.

3.4.23 PASSENGER_X_NATIONALITY: Letter X represents nationality of <u>adult</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation.

3.4.24 PASSENGER_X_M_NATIONALITY: Letter X represents nationality of passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation. Letter M stands for <u>minors</u> under 18 years of age.

3.4.25 PASSENGER_X_I_NATIONALITY: Letter X represents nationality of <u>infant</u> passenger 2, up to 5, depending on the number of babies the account holder included on his/her reservation. Letter I stands for infants younger than 1 year of age.

3.4.26 PASSPORT_NO_PASSENGER_X: Letter X represents the passport number of <u>adult</u> passenger 2, up to 5, depending on the number of people the account holder included on his/her reservation.

3.4.27 PASSPORT_NO_PASSENGER_X_M: Letter X represents the passport number of passenger 2, up to 5, depending on the number of people the account

holder included on his/her reservation. Letter M stands for <u>minors</u> under 18 years of age.
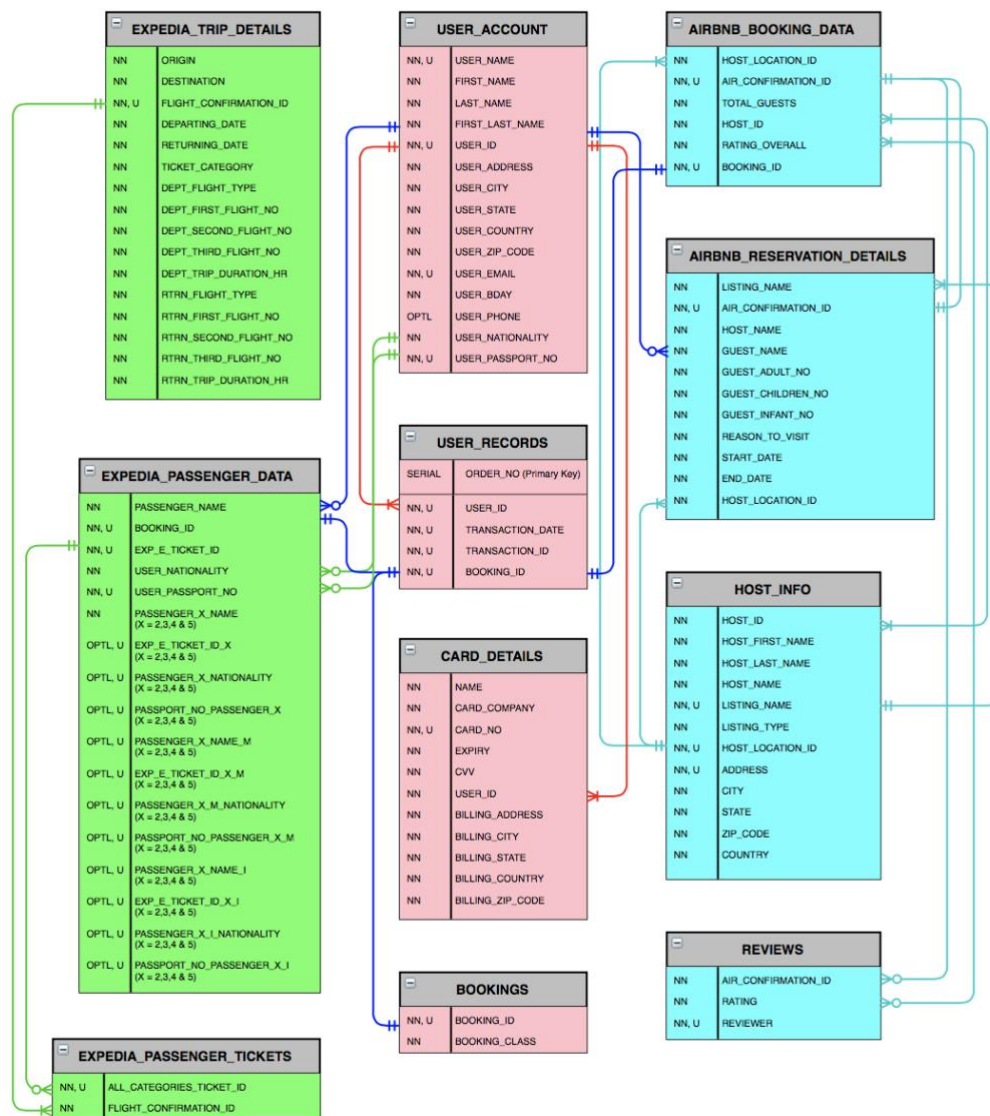
3.4.28 PASSPORT_NO_PASSENGER_X_I: Letter X represents the passport number of <u>infant</u> passenger 2, up to 5, depending on the number of babies the account holder included on his/her reservation. Letter I stands for infants younger than 1 year of age.

3.4.29 ALL_CATEGORIES_TICKET_ID: This field includes every Expedia electronic ticket linked to every booked flight confirmation id.

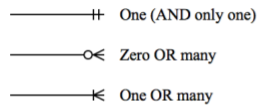# 4. Entity Relationship (ER) Diagram:

This section represents an ER Diagram (ERD) created with *Draw.io* (diagramming application), followed by the ERD's description.
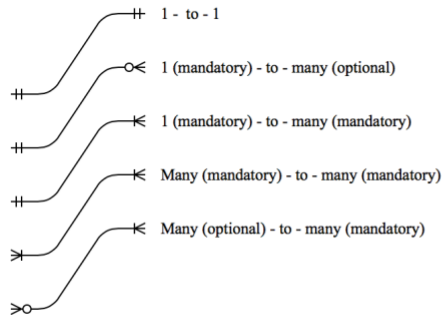
4.1 JLMFCompany's ERD:
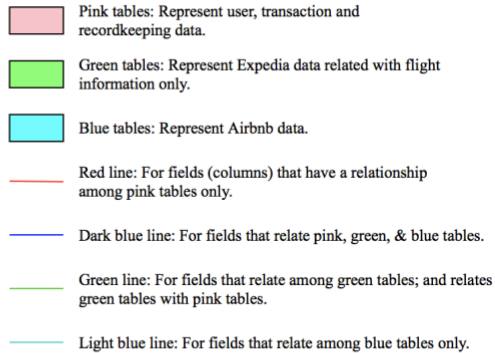
4.2 ERD Symbols & Description:

Cardinality & ordinality:

- ———||— One (AND only one)
- ———O<— Zero OR many
- ———|<— One OR many

Entity Relation:

- ———||— 1 - to - 1
- ———O<— 1 (mandatory) - to - many (optional)
- ———|<— 1 (mandatory) - to - many (mandatory)
- ———|<— Many (mandatory) - to - many (mandatory)
- ———|<— Many (optional) - to - many (mandatory)

Meaning of colors:

- Pink tables: Represent user, transaction and recordkeeping data.
- Green tables: Represent Expedia data related with flight information only.
- Blue tables: Represent Airbnb data.
- Red line: For fields (columns) that have a relationship among pink tables only.
- Dark blue line: For fields that relate pink, green, & blue tables.
- Green line: For fields that relate among green tables; and relates green tables with pink tables.
- Light blue line: For fields that relate among blue tables only.

Abbreviations & Acronyms:

NN: Not Null. Represents any column that can't have empty cells.
U: Unique. Represents fields (columns) with non-repeatable reads.
OPTL: Optional or non-essential data. For instance, phone numbers.
SERIAL: Sequential integer. Serves as the primary key in this project.

5. Database summary:

The following table provides a list of every database table (hence each column header) arranged in the same order as the Microsoft Excel and CSV files that form the basis of JLMFCompany's database in Azure Data Studio - SQL Server.

USER_ACCOUNT	• USER_NAME • FIRST_NAME • LAST_NAME • FIRST_LAST_NAME • USER_ID • USER_ADDRESS • USER_CITY • USER_STATE • USER_COUNTRY • USER_ZIP_CODE • USER_EMAIL • USER_BDAY • USER_PHONE • USER_NATIONALITY • USER_PASSPORT_NO
USER_RECORDS	• ORDER_NO (PRIMARY_KEY) • USER_ID • TRANSACTION_DATE • TRANSACTION_ID

	• BOOKING_ID
AIRBNB_BOOKING_DATA	• HOST_LOCATION_ID • AIR_CONFIRMATION_ID • TOTAL_GUESTS • HOST_ID • RATING_OVERRALL • BOOKING_ID
AIRBNB_RESERVATION_DETAILS	• LISTING_NAME • AIR_CONFIRMATION_ID • HOST_NAME • GUEST_NAME • GUEST_ADULT_NO • GUEST_CHILDREN_NO • GUEST_INFANT_NO • REASON_TO_VISIT • START_DATE • END_DATE • HOST_LOCATION_ID
REVIEWS	• REVIEWER • RATING • AIR_CONFIRMATION_ID
BOOKINGS	• BOOKING_ID • BOOKING_CLASS
HOST_INFO	• HOST_ID • HOST_FIRST_NAME • HOST_LAST_NAME • HOST_NAME • LISTING_NAME • LISTING_TYPE • HOST_LOCATION_ID • ADDRESS • CITY • STATE • ZIP_CODE • COUNTRY
EXPEDIA_TRIP_DETAILS	• ORIGIN • DESTINATION • FLIGHT_CONFIRMATION_ID • DEPARTING_DATE • RETURNING_DATE

	• TICKET_CATEGORY • DEPT_FLIGHT_TYPE • DEPT_FIRST_FLIGHT_NO • DEPT_SECOND_FLIGHT_NO • DEPT_THIRD_FLIGHT_NO • DEPT_TRIP_DURATION_HR • RTRN_FLIGHT_TYPE • RTRN_FIRST_FLIGHT_NO • RTRN_SECOND_FLIGHT_NO • RTRN_THIRD_FLIGHT_NO • RTRN_TRIP_DURATION_HR
EXPEDIA_PASSENGER_DATA	• PASSENGER_NAME • BOOKING_ID • EXP_E_TICKET_ID • USER_NATIONALITY • PASSPORT_NO • PASSENGER_2_NAME • EXP_E_TICKET_ID_2 • PASSENGER_2_NATIONALITY • PASSPORT_NO_PASSENGER_2 • PASSENGER_3_NAME • EXP_E_TICKET_ID_3 • PASSENGER_3_NATIONALITY • PASSPORT_NO_PASSENGER_3 • PASSENGER_4_NAME • EXP_E_TICKET_ID_4 • PASSENGER_4_NATIONALITY • PASSPORT_NO_PASSENGER_4 • PASSENGER_5_NAME • EXP_E_TICKET_ID_5 • PASSENGER_5_NATIONALITY • PASSPORT_NO_PASSENGER_5 • PASSENGER_2_NAME_M • EXP_E_TICKET_ID_2_M • PASSENGER_2_M_NATIONALITY • PASSPORT_NO_PASSENGER_2_M • PASSENGER_3_NAME_M • EXP_E_TICKET_ID_3_M • PASSENGER_3_M_NATIONALITY • PASSPORT_NO_PASSENGER_3_M

	• PASSENGER_4_NAME_M
	• EXP_E_TICKET_ID_4_M
	• PASSENGER_4_M_NATIONALITY
	• PASSPORT_NO_PASSENGER_4_M
	• PASSENGER_5_NAME_M
	• EXP_E_TICKET_ID_5_M
	• PASSENGER_5_M_NATIONALITY
	• PASSPORT_NO_PASSENGER_5_M
	• PASSENGER_2_NAME_I
	• EXP_E_TICKET_ID_2_I
	• PASSENGER_2_I_NATIONALITY
	• PASSPORT_NO_PASSENGER_2_I
	• PASSENGER_3_NAME_I
	• EXP_E_TICKET_ID_3_I
	• PASSENGER_3_I_NATIONALITY
	• PASSPORT_NO_PASSENGER_3_I
	• PASSENGER_4_NAME_I
	• EXP_E_TICKET_ID_4_I
	• PASSENGER_4_I_NATIONALITY
	• PASSPORT_NO_PASSENGER_4_I
	• PASSENGER_5_NAME_I
	• EXP_E_TICKET_ID_5_I
	• PASSENGER_5_I_NATIONALITY
	• PASSPORT_NO_PASSENGER_5_I
EXPEDIA_PASSENGER_TICKETS	• FLIGHT_CONFIRMATION_ID
	• ALL_CATEGORIES_TICKET_ID
CARD_DETAILS	• NAME
	• CARD_COMPANY
	• CARD_NO
	• EXPIRY
	• CVV
	• USER_ID
	• BILLING_ADDRESS
	• BILLING_CITY
	• BILLING_STATE
	• BILLING_COUNTRY
	• BILLING_ZIP_CODE

6. Reference sources:

- Docs.microsoft.com. (2020). *[MS-SQLPROTLP]: Technical Documents*. [online] Available at: https://docs.microsoft.com/en-us/openspecs/sql_server_protocols/ms-sqlprotlp/3432cb95-3f91-446d-86d7-b9ea0816464e [Accessed 19 Feb. 2020].

- Docs.microsoft.com. (2020). *Microsoft Azure Documentation*. [online] Available at: https://docs.microsoft.com/en-us/azure/ [Accessed 19 Feb. 2020].

- Docs.microsoft.com. (2020). *SQL Server technical documentation - SQL Server*. [online] Available at: https://docs.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver15 [Accessed 19 Feb. 2020].

- Docs.python.org. (2020). *3.8.2rc2 Documentation*. [online] Available at: https://docs.python.org/3/ [Accessed 19 Feb. 2020].

- Inside Airbnb. (2020). *Inside Airbnb. Adding data to the debate*. [online] Available at: http://insideairbnb.com/get-the-data.html [Accessed 19 Feb. 2020].

- Faker.readthedocs.io. (2020). *Welcome to Faker's documentation! — Faker 4.0.1 documentation*. [online] Available at: https://faker.readthedocs.io/en/master/ [Accessed 19 Feb. 2020].

- Lucidchart.com. (2020). [online] Available at: https://www.lucidchart.com/pages/ER-diagram-symbols-and-meaning [Accessed 19 Feb. 2020].

- Openflights.org. (2020). *OpenFlights: Airport and airline data*. [online] Available at: https://openflights.org/data.html [Accessed 19 Feb. 2020].

- Oracle Help Center. (2020). *Database Documentation - Database Documentation*. [online] Available at: https://docs.oracle.com/en/database/ [Accessed 19 Feb. 2020].

- W3schools.com. (2020). *SQL Tutorial*. [online] Available at: https://www.w3schools.com/sql/ [Accessed 19 Feb. 2020].

- Worldpostalcode.com. (2020). *World Postal Code - free zip/postal code lookup*. [online] Available at: https://worldpostalcode.com [Accessed 19 Feb. 2020].