Design Specification Document

**Prepared by:** Nidhi Choudhary & Ritik Sinha

**Prepared for: Analysing the Business of hotels**

**Created Date:** 3rd November 2023

Table of Contents

[About the dataset 2](#_Toc78831155)

[Timeframe of data 2](#_Toc78831156)

[Data Description 2](#_Toc78831157)

[Architecture 5](#_Toc78831158)

[Azure Services Required 5](#_Toc78831159)

[Resources Created 6](#_Toc78831160)

[Data Ingestion to Azure 6](#_Toc78831161)

[Orchestration 7](#_Toc78831162)

[Database Components 7](#_Toc78831163)

[Master Table 7](#_Toc78831164)

[Reporting View 9](#_Toc78831165)

[Azure Data factory 9](#_Toc78831166)

[Linked Services 9](#_Toc78831167)

[Azure Pipeline 11](#_Toc78831168)

[Reporting View Creation 11](#_Toc78831169)

[Power-BI Specification 12](#_Toc78831170)

[Using reporting views 12](#_Toc78831171)

[Visualizations 13](#_Toc78831172)

# About the dataset

The dataset contains data from two different hotels. One Resort hotel and one City hotel.

## Timeframe of data

The data contains "bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017".

Data Description

Dataset contains the following details:

1. **hotel** : (Categorical Variable) This column in the data represent whether it as the City Hotel or Resort Hotel.

This column is used for the categorization of all visualization-based inferences amongst the City and Resort hotels.

1. **is\_canceled**: (Categorical Variable) This column indicates whether the booking was canceled or not.

In most of our analysis, bookings that were not canceled were considered, to get actual guest numbers. Else, there would be a quite big difference as there were a lot of cancellations.

1. **lead\_time** (Integer) This column show the number of days that elapsed between the entering date of the booking into the PMS (Property Management System) and the arrival date.
2. **arrival\_date\_year** : (Categorical Variable) This column displays the year of arrival date.
3. **arrival\_date\_month** : (Categorical Variable) This column displays the month of arrival date
4. **arrival\_date\_week\_number** : (Categorical Variable) This column displays the week number of year for arrival date
5. **arrival\_date\_day\_of\_month**  :**(**Categorical Variable) This column displays the day of arrival date
6. **stays\_in\_weekend\_nights** : (Continuous Variable) It shows the number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
7. **stays\_in\_week\_nights** (Continuous Variable) This column shows the number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel.
8. **adults** :(Continuous Variable) It displays the number of adults in the guests. Each row displays the adults present in the particular family.
9. **children** :(Continuous Variable) This column gives the number of children among the guests. Each row displays the children present in the particular family.
10. **babies** :(Continuous Variable) This column gives the number of babies among the guests. Each row displays the babies present in the particular family.
11. **meal** **: (**Categorical Variable) This column gives the information on the type of meal booked each family.

There are four categories in standard hospitality meal packages:

* Undefined/SC – no meal package
* BB – Bed & Breakfast
* HB – Half board (breakfast and one other meal – usually dinner)
* FB – Full board (breakfast, lunch and dinner)

1. **country** :**(**Categorical Variable) This shows the country of origin of each family/guests. Categories are represented in the ISO 3155–3:2013 format. There are guests whose country of origin is not recorded.
2. **market\_segment** :**(**Categorical Variable) This column shows the market segment designation.

There are two categories for market segments

* TA - “Travel Agents”
* TO -  “Tour Operators”

1. **distribution\_channel** :**(**Categorical Variable) This refers to the booking distribution channel that the guests took to do their booking.

The two categories present are

* TA - “Travel Agents”
* TO - “Tour Operators”

1. **is\_repeated\_guest** :**(**Categorical Variable) This column shows whether the booking was from a repeated guest or not.

1. **previous\_cancellations** :(Continuous Variable) It shows the number of previous bookings that were cancelled by the customer prior to the current booking.
2. **previous\_bookings\_not\_canceled** :(Continuous Variable) The rows under this head gives the number of previous bookings not cancelled by the customer prior to the current booking.
3. **reserved\_room\_type** **:(**Categorical Variable) This column shows the code of room type reserved by the guests. Here, the code is presented instead of designation for anonymity reasons.
4. **assigned\_room\_type** :**(**Categorical Variable) It displays the code for the type of room finally assigned to the booking. Here too, code is presented instead of designation for anonymity reasons.

Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request itself.

1. **booking\_changes** :(Continuous Variable) This shows the number of changes or amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation.
2. **deposit\_type** **:(**Categorical Variable) This column is an indication on if the customer made a deposit to guarantee the booking.

This column can assume three categories:

* No Deposit – no deposit was made
* Non Refund – a deposit was made in the value of the total stay cost
* Refundable – a deposit was made with a value under the total cost of stay.

1. **agent** : **(**Categorical Variable) This column gives the ID of the travel agency that made the booking for the guests.
2. **company** :**(**Categorical Variable) Here, the ID of the company or entity that made the booking or responsible for paying the booking is displayed. ID is presented instead of designation for anonymity reasons.
3. **days\_in\_waiting\_list** :(Continuous Variable) Here the customer type is mentioned based on the type of booking made.
4. **customer\_type** :**(**Categorical Variable) Customer is classified based on the type of booking made .

There are four categories:

* Contract - when the booking has an allotment or other type of contract associated to it
* Group – when the booking is associated to a group
* Transient – when the booking is not part of a group or contract, and is not associated to other transient booking
* Transient-party – when the booking is transient, but is associated to at least other transient booking

1. **adr** :(Continuous Variable) This column gives the values of ADR (Average Daily Rate). It is defined as the ratio of the sum of all lodging transactions to the total number of staying nights.
2. **required\_car\_parking\_spaces** :(Continuous Variable) This column displays the number of car parking spaces required by the customer.
3. **total\_of\_special\_requests** :(Continuous Variable) Number of special requests made by the customer such as requirement of twin beds or high floor e.t.c is displayed in this column.
4. **reservation\_status** :**(**Categorical Variable) This column shows the last status of the reservations.

There are three categories:

* Canceled – booking was canceled by the customer
* Check-Out – customer has checked in but already departed
* No-Show – customer did not check-in and did inform the hotel of the reason why

1. **reservation\_status\_date** :**(**Categorical Variable) This column shows the date at which the last status was set.

This variable can be used in conjunction with the reservation\_status to understand when was the booking canceled or when did the customer checked-out of the hotel.

# Architecture

Graphical user interface, diagram

Description automatically generated with medium confidence

# Azure Services Required

|  |
| --- |
| Service type |
| Storage Account |
| Azure Data Factory |
| Azure SQL Database |
| SQL Server |

# Resources Created

|  |  |
| --- | --- |
| Service type | Resource Name |
| Storage Accounts | Fsap1Storage |
| Data Factory | [FSAP1-ADF](https://portal.azure.com/#@tredence.com/resource/subscriptions/a4f54399-8db8-4849-adcc-a42aed1fb97f/resourceGroups/FSAP-RG1/providers/Microsoft.DataFactory/factories/FSAP1-ADF) |
| Azure SQL Database | [FSAP1-DB (fsap-srv1/FSAP1-DB)](https://portal.azure.com/#@tredence.com/resource/subscriptions/a4f54399-8db8-4849-adcc-a42aed1fb97f/resourceGroups/FSAP-RG1/providers/Microsoft.Sql/servers/fsap-srv1/databases/FSAP1-DB) |
| SQL Server Name | fsap-srv1 |
| SQL Server DNS name | fsap-srv1.database.windows.net |

Graphical user interface, text, application

Description automatically generated

# Data Ingestion to Azure

Under storage account an “data-nulltreated” container is created.

Graphical user interface

Description automatically generated with low confidence

The data is uploaded to Azure cloud and stored as blob in the “data-nulltreated” container.

Graphical user interface, text, application, email

Description automatically generated

# Orchestration

## Database Components

Below listed components are created in the Database

|  |  |
| --- | --- |
| Database Components | Description |
| hotel | SQL Master Table |
| hotel\_powerbi | SQL Reporting View Table |

### Master Table

Hotel Table

A picture containing table

Description automatically generated

Text

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

### Reporting View

Hotel\_powerbi Table

Table

Description automatically generated with medium confidence

## Azure Data factory

### Linked Services

**Linked services are created** to link data store to the data factory. These are like connection strings, which define the connection information needed for Data Factory to connect to external resources.

Two linked services are being used: -

1. Clean\_csv – for linking the sql storage blob to the data factory.
2. AzureSqlDatabase1- for linking the SQL Database to the data factory

Graphical user interface, text, application, email

Description automatically generated

### Azure Pipeline

We created a pipeline named “csv\_to\_sql” which would copy the contents of csv file into the sql database.

Graphical user interface, application

Description automatically generated

csv\_to\_sql -> Copies data from Azure BLOB storage to SQL DB table

# Power-BI Specification

## Using reporting views

Report:

Columns used for visualizations:

**hotel, is\_canceled, lead\_time** , **arrival\_date\_year, arrival\_date\_month** , **stays\_in\_weekend\_nights** , **stays\_in\_week\_nights, adults, children, babies , country** , **market\_segment** , **distribution\_channel** , **is\_repeated\_guest** , **reserved\_room\_type** , **assigned\_room\_type** , , **adr,**

Problem Statements:

1. Where do the guests come from?

2. How much do guests pay for a room per night?

3. How does the price per night vary over the year?

4. Which are the most busy month?

5. How long do people stay at the hotels?

6. Bookings by market segment

7. How many bookings were canceled?

8. Which month have the highest number of cancelations?

## Visualizations

