**Progress test 1**

**Data warehouse – DBW301**

**Class : RollNumber: FullName:**

A part of the operational database of a train/railway system of a company contains the following tables:

**Trains**(TrainNumber, MaxSpeed, NumberOfWagons)

**Classes**(ID, Name)

**Wagons**(TrainNumber, WagonNumber, ClassID, NumberOfSeats)

**Seats**(TrainNumber, WagonNumber, SeatNumber, Available)

**Passengers**(ID,Last\_Name, First\_Name, Gender, DateOfBirth, Email, Address, ZIPCode, City, Country)

**Stations**(ID, Name, Address, ZIPCode, City, Country)

**Routes**(ID, FromStationID, ToStationID, Distance)

**RoutePrices**(PriceID, RouteID, TrainNumber, ClassID, FromTime, ToTime, Price)

**Discounts**(ID, Name, DiscountPercentage, FromDate, ToDate)

**Ticket\_Cancelation**(BookingID, CancelationTime)

**Ticket\_Booking**(BookingID, PassengerID, DiscountID, BookingTime, DepartureStationID, ExpectedDepartureDatetime, RealDepartureDateTime, ArrivalStationID, ExpectedArrivalDatetime, RealArrivalDatetime, Amount, Cancel?)

**Ticket\_Details**(ID, BookingID, RoutePriceID, WagonNumber, SeatNumber, ExpectedDepartureDatetime, RealDepartureDatetime, ExpectedArrivalDatetime, RealArrivalDatetime)

Note that the company has many trains, each train contains some wagons, each wagon belongs to only one class (1st class or 2nd class). Each wagon has several seats.

The company serves trains to travel in a same country or between different countries. A route represents a direct path between two stations. The unit price on a given route depends on the train, the class as well as the time. For example, if the departure time of a train is in rush hour, the price is much higher than the price at 5am.

When booking a travel from Paris and Brussels, the passengers could choose to book a direct route from Paris Gare du Nord to Brussels Central Station or he/she could choose to pass by Strasbourg Train Station for having lower price. In the case that he/she choose to pass by Strasbourg Train Station, the DepartureStationID and the ArrivalStationID in his/her ticket booking are respectively the IDs of Paris Gare du Nord and Brussels Central Station and there will have two ticket details corresponding to the route from Paris Gare du Nord to Strasbourg Train Station and the route from Strasbourg Train Station to Brussels Central Station.

Depending on the booking time and booking day, the customer can benefit some discount on the price.

Sometimes, passengers cancel their booking. The cancelation information is noted in Ticket\_Cancelation.

**Business requirements:**

The executive managers of the company want to improve the activity of their company. To achieve this objective, they want to create a data mart allowing them to analyze the number of sold tickets as well as their income.

By comparing the number of sold tickets and the number of available seats during time, the decision maker may decide to increase the number of trains or the number wagons for trains between two cities during rush hour for example.

Analyzing the result by train class may help them to decide if they need to increase the number of wagons of a certain class or to reduce the number of wagons of another class.

For future marketing campaign, the decision maker needs to know the preferred destination (city) of each customer, by his/her age, gender, or address (city). Old customer may prefer to travel in the 1st class or female customer may prefer to travel on Christmas than on summer holiday for example.

Of course, the decision maker needs to analyze the result according to the time of the day, the day of the week or the day of the month. They may want to compare the income of different years. For example, the income in 2020 could be much lower than the income in 2019 because of the Covid.

**Question:**

1. **Create the information package corresponding to the user requirement.**
2. **Distinguish the hierarchies and categories of each dimension.**
3. **Which level of details of data your data warehouse must be hold?**