**FDC 105 PROJECT – GROUP 9 FINAL REPORT**

**(DATA SET 4: COMMERCIAL AIRLINE FLIGHTS)**

**1. Mục tiêu, câu hỏi và nhiệm vụ nghiên cứu**

(1) Về mục tiêu nghiên cứu: Dựa trên bộ dữ liệu liên quan đến các chuyến bay, địa điểm, giá vé của các hãng bay (airlines) và các booking agencies; nhóm thực hiện việc làm sạch, tạo các mối quan hệ (relationships) và phân tích bộ dữ liệu trên nhằm đưa ra các xu hướng, đặc điểm nổi bật cũng như bất thường để từ đó đưa ra giải pháp cho các hãng bay và booking agencies

(2) Về câu hỏi nghiên cứu: Các câu hỏi được nhóm đặt ra trong quá trình phân tích dữ liệu như sau:

- Thứ nhất, bộ dữ liệu bao gồm các thông tin gì, được phục vụ cho mục đích gì và có thể được khai thác như thế nào, giúp ích như thế nào?

- Thứ hai, bộ dữ liệu đã được làm sạch và đủ điều kiện để đưa vào phân tích hay chưa, đồng thời cần sử dụng thêm các biến số đo lường nào khác hay không?

- Thứ ba, bộ dữ liệu trên có thể được triển khai dưới dạng các bảng như thế nào và các bảng này có mối quan hệ với nhau như thế nào (mối quan hệ một-một, một-nhiều hay nhiều-nhiều)?

- Thứ tư, các dữ liệu sau khi được phân tích nên được trực quan hóa như thế nào để có thể được khai thác một cách hiệu quả nhất và có cần sử dụng thêm các measurements nào nữa hay không?

- Thứ năm, từ các bảng đã được vẽ lên, các thông tin gì và có những gợi ý gì được đưa ra dành cho các hãng bay và booking agencies?

(3) Các insights có thể được khai thác bao gồm:

* Xu hướng các chuyến bay thay đổi như thế nào theo thời gian, số lượng có thay đổi nhiều trong 2 năm 2018 và 2019 không
* Các hãng bay có sự chênh lệch nhiều về số chuyến bay được đặt trước hay không
* Sự chênh lệch về số lượng vé được đặt đối với từng hạng ghế
* Các agency bán vé có doanh thu như thế nào
* Những airlines nào có doanh thu bán vé cao nhất
* Đối với từng hạng ghế thì giá vé trung bình có thay đổi nhiều không
* Mục đích chủ yếu của khách hàng khi sử dụng các chuyến bay là gì
* Địa điểm được bay đến nhiều nhất

**2. Mô tả giải pháp**

***\* Làm sạch bộ dữ liệu***

- Bộ dữ liệu ban đầu gồm một file excel duy nhất, chưa được tách thành các bảng, chưa có các relationships, một số cột bị sai format (booking date)

***\* Xây dựng mô hình***

- Nhóm đã xử lý và tách thành 3 bảng, lần lượt là: DimOrigin, DimDestination và Flights.

+ Nhóm thực hiện việc tạo các relationships giữa các bảng

+ Trong đó, DimOrigin và DimDestination có mối quan hệ một nhiều với bảng Flights

***\* Khai thác và trực quan hóa bộ dữ liệu***

- Đối với bộ dữ liệu này, nhóm khai thác dựa trên 3 khía cạnh và tách thành 3 trang, lần lượt là:

+ Flight Number: Mô tả xu hướng của số lượng vé, đặt vé, trả lời cho các câu hỏi:

* + Xu hướng các chuyến bay thay đổi như thế nào theo thời gian, số lượng có thay đổi nhiều trong 2 năm 2018 và 2019 không
  + Các hãng bay có sự chênh lệch nhiều về số chuyến bay được đặt trước hay không
  + Sự chênh lệch về số lượng vé được đặt đối với từng hạng ghế

+ Ticket Price: Mô tả tình hình doanh thu của airlines và agencies, trả lời cho các câu hỏi:

* + Sự chênh lệch về số lượng vé được đặt đối với từng hạng ghế
  + Các agency bán vé có doanh thu như thế nào
  + Những airlines nào có doanh thu bán vé cao nhất
  + Đối với từng hạng ghế thì giá vé trung bình có thay đổi nhiều không

+ Preference: Mô tả sở thích và mục đích của khách hàng khi đặt vé máy bay, trả lời cho các câu hỏi:

* + Mục đích chủ yếu của khách hàng khi sử dụng các chuyến bay là gì
  + Địa điểm được bay đến nhiều nhất
* Các bảng và hình phù hợp được nhóm lựa chọn để phân tích các dữ liệu kể trên
* Đối với việc phân tích địa điểm được di chuyển nhiều nhất, nhóm có sử dụng thêm một add-in là Flow map của Power BI để phân tích trên bản đồ.

**3. Các kết quả thu được (Findings)**

***\* Về Flight Number:***

* Các airlines có sự thay đổi không đáng kể trong số vé được đặt. Giữa hai năm 2018 và 2019 có sự chênh lệch nhẹ
* Seat class cũng được phân bố khá đều (xấp xỉ 25%)
* Tuy nhiên nếu lựa chọn phân tích theo từng seat class thì trend của các chuyến bay cũng có sự chênh lệch đáng kể hơn
* Cheapoair và American Airlines là hai agencies có số lượng booking chiến phần lớn

***\* Về Ticket Price***

* Agencies có doanh thu lớn nhất là Expedia
* Tuy nhiên sự thay đổi về giá vé có xuất hiện theo thời gian
* Giá vé 2018 nhỏ bằng 50% so với giá vé của 2019

***\* Về Preference***

* Arizone là địa điểm di chuyển tới nhiều nhất
* Personal non-leisure là mục đích di chuyển chủ yếu

**4. Mô tả ảnh hưởng của báo cáo**

* Báo cáo có thể được sử dụng như một market research, đưa ra cái nhìn vĩ mô đối với tình hình hoạt động của ngành hàng không trong khoảng thời gian từ 2018 đến 2020.
* Các airlines và agencies có thể sử dụng báo cáo để điều chỉnh số lượng chuyến bay (mở thêm chuyến bay khi có nhu cầu lớn và ngược lại), điều chỉnh giá vé, cũng như mở thêm các chặng tới nơi được di chuyển nhiều nhất
* Tuy nhiên, hạn chế của bộ dữ liệu này chưa bao gồm các đặc điểm về giờ bay, nhiên liệu tiêu thụ. Các báo cáo đi sau có thể sử dụng thêm các chi tiết này để đưa thêm insights, bổ sung sự giải thích cho giá vé của airlines.