

## Senior Software Engineer (Google Analytics) – Case Study

### 1) The Business Problem and Task

Emirates Airline is a global carrier operating out of the UAE. The Airline sells three different seat types/products:

- Economy Class
  - Average Retail price: 4,000 AED
  - Operations Cost: 2,000 AED
- Premium Economy Class
  - Average Retail price: 5,000 AED
  - Cost: 2,500 AED
- Business Class
  - Retail price: 8,000 AED (increased to 9,000 AED from January 2019)
  - Cost: 4,000 AED

*Note: Cost figures do not include the cost of discounts*

Over the last three years, the company has seen a drop in sales and subsequently profit. To help combat the change, the company trialled two initiatives:

- From January 2018, the Airline decided to up the maximum potential discounts from 10% to 20%
- From January 2019, the Airline decided to increase the price of the Business Class seats from 8,000 to 9,000

It is unclear what impact each of these changes had on business performance.

Emirates Analytics Team wants to perform, some crucial analysis and senior software engineer needs to prepare a data set for analysis purpose.

### The Data

You have been provided with two sets of data:

- Leads (leads.csv) - A list of enquiries and associated customer information
- Purchases (purchases.csv) - A list of purchases and their associated retail price and discount amount

*Note: Assume that all leads are from new and distinct customers*

### The Assessment

1. Data Cleaning: The data set you have been provided has some missing values. How would you identify and handle missing values in the data set?
2. Data Manipulation: Suppose you need to create a new column in the data set that calculates the duration between a lead to convert into a purchase. How would you create this new column?
3. Data Transformation: Using the two data sets, can you create at least three (3) derived features/attributes?
  - a. Create a flag for customers who searched for Economy Class but ended up booking premium economy or business class ticket.
  - b. Can you create a target variable called **"Converted"** for the analytics team?
  - c. Can you calculate the final booking price after applying the discount for each customer who booked a flight?
4. Data Storage: Suppose you want to store the data set in a database for easy querying and analysis. What type of database would you choose and why?
5. Data Pipeline: Suppose you want to create a data pipeline to process new flight data as it becomes available. What tools or technologies would you use to build this pipeline? How would you ensure the pipeline is reliable and scalable?



- 2) Use the Raw data provided as a json file to create a flat table for analytics team to use for analysis and reporting purpose:

Field name	Type
date	DATE
device_category	STRING
country	STRING
event_name	STRING
item_name	STRING
item_id	STRING
page_title	STRING
content	STRING
trigger	STRING
content_type	STRING
type	STRING
session_id	INTEGER
ga_session_number	INTEGER
skywards_tier	STRING
event_count	INTEGER
downloads	INTEGER
operating_system	STRING

**Note:** All the analysis/coding must be in Python (preferable Google Cloud Platform if possible). You can use any visualization tool if that helps in your analysis. Please put together 5-6 slides and separately share the code in an excel sheet or in a word document.