

## Data Analyst Take Home Challenge

Assume you are a data analyst in a travel insurance company. The business team has asked you to identify the factors that attribute to a "low-risk" or "high-risk" flight, so that they can build a better pricing model for the travel insurance

### Business Objective

As a travel insurance company, flight delay claim is one of the popular claim items. The claim logic for flight delay refund in our insurance policy is

- If the delay time is greater than 3 hours OR the flight is canceled, \$800 will be claimed.
- Otherwise, the claim amount will equal \$0.

The business team wants to build a pricing model based on the risk of flight delay refund. This means we want to charge

1. a lower price if a customer takes a low-risk flight, so that the customer is more willing to buy insurance for the flight and hence we can expand the risk pool.
2. a higher price if a customer takes a high-risk flight, so that the customer is less willing to buy insurance for the flight and hence we naturally screen out the high-risk customer and adequately compensate the risk the company needs to take.

### Skills to be Graded

- Data Visualization
- Statistical Testing
- Communication & Business Acumen

## Deliverables

Now, the business team has given you a dataset with flight delay claims from 2013/01 to 2016/07. They ask you to identify factors that attribute to a low-risk or high-risk flight. Feel free to use any programming languages, libraries, or visualization tools to accomplish the following deliverables. You are advised to spend less than 6 hours on this challenge.

- (Must Have) Prepare a report to present your key result to the business team in PDF format.
  - Please remember that your business partners are non-technical.
  - You are going to suggest what factors the business team should consider when determining the pricing model and suggest a pricing model.
  - You should include appropriate visualization to illustrate your findings.
  - Your "Data Visualization" and "Communication & Business Acumen" skills will be graded based on this report.
- (Must Have) Prepare code/artifacts to provide statistical evidence on the key factors that you suggested in the business report above.
  - You can assume the audience is your colleague from the analytics team.
  - You can assume that your colleague is going to review the code of your work.
  - You should specify your hypothesis/hypotheses in comment.
  - You should provide the conclusion of statistical evidence in comment.
  - Your "Statistical Testing" skill will be graded based on the code/artifacts.
- (Optional) Prepare an Appendix at the end of your report to present your thought about the problem, your approach, analytics limitation, and potential next steps.

Once you have done on the above deliverables,

- Compress your code and artifacts that generate the visualization and the analytics result in the PDF report for the business team, e.g., a Tableau workbook or a Jupyter Notebook. Then, upload the compressed file to <here>.
- We would download and run your code/artifact. Please make sure you also include the instruction, e.g., a readme, on how to set up the environment and verify your result.

## Bonus Point

- We welcome you to add extra parameters/data for more factors to be included in the analysis.
- We welcome you to come with a suggestion on the pricing model based on your analytics findings for the business team.

## Data

Download link: <https://drive.google.com/file/d/13x92X26bTkpl3ZyFK8MI7fqauZvqCtb2/view?usp=sharing>  
(46MB CSV file)

## Data Description

|             |  |
|-------------|--|
| flight_id   | Unique ID for each flight  |
| flight_no   | flight number of each flight   |
| Week        | Indicate which week of year is the departure date in, For example, for flight departing at 17/1/2018, the week will be 3 |
| Departure   | Location of departure (Airport)  |
| Arrival     | Location of arrival (Airport)  |
| Airline     | The airline company code of the flight   |
| std_hour    | Scheduled departure time, in 24-hour format  |
| delay_time  | Number of delayed hours  |
| flight_date | the date of departing.   |
| is_claim    | Claim amount, our insurance will pay customer a fixed amount of HK\$800 when a delay happens.                            |