EXPERIMENT REPORT

Student Name	Antonio Wang
Project Name	Adv_mla_asm3
Date	9/11/2023
Deliverables	RF.ipynb Random Forest

1. EXPERIMENT BACKGROUND

Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach.

1.a. Business Objective	Explain clearly what is the goal of this project for the business. How will the results be used? What will be the impact of accurate or incorrect results? The business objective is to create a streamlit app that can use 4 different prediction models to accurately predict the airfare of a flight based on the user's input.
1.b. Hypothesis	Present the hypothesis you want to test, the question you want to answer or the insight you are seeking. Explain the reasons why you think it is worthwhile considering it, The hypothesis is the total airfare is largely dominated by the type of cabin and type of ticket the user is planning on purchasing.
1.c. Experiment Objective	Detail what will be the expected outcome of the experiment. If possible, estimate the goal you are expecting. List the possible scenarios resulting from this experiment. The experiment's objective is to create an accurate Random Forest model. Possible scenarios include an inaccurate model, failure to create the model, or successful creation of an accurate model.

2. EXPERIMENT DETAILS

Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them.

2.a. Data Preparation	Describe the steps taken for preparing the data (if any). Explain the rationale why you had to perform these steps. List also the steps you decided to not execute and the reasoning behind it. Highlight any step that may potentially be important for future experiments Some data preprocessing done was to parse some timestamp columns, and extracting information that was split up by ||.

2.b. Feature Engineering	Describe the steps taken for generating features (if any). Explain the rationale why you had to perform these steps. List also the feature you decided to remove and the reasoning behind it. Highlight any feature that may potentially be important for future experiments
	
	Feature engineering steps include
	   - changing all the categorical columns to numerical.
	   - Splitting up the dates into year, month and day.
	   - Dropping any unneeded columns.

2.c. Modelling	Describe the model(s) trained for this experiment and why you choose them. List the hyperparameter tuned and the values tested and also the rationale why you choose them. List also the models you decided to not train and the reasoning behind it. Highlight any model or hyperparameter that may potentially be important for future experiments The random forest model was chosen, because it is accurate. It can also reduce overfitting and through the use of some hyperparameters the training speed can be quite fast.

3. EXPERIMENT RESULTS
Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified.

3.a. Technical Performance	Score of the relevant performance metric(s). Provide analysis on the main underperforming cases/observations and potential root causes. The MSE score was 5437.17 The RMSE score was 73.74 These scores could be better if the hyperparameters were tuned more, however for the sake of training speed and the size of the model, further tuning was not done.
3.b. Business Impact	Interpret the results of the experiments related to the business objective set earlier. Estimate the impacts of the incorrect results for the business (some results may have more impact compared to others) These results are quite accurate and can help the user get an accurate prediction of what their airfare will be.
3.c. Encountered Issues	List all the issues you faced during the experiments (solved and unsolved). Present solutions or workarounds for overcoming them. Highlight also the issues that may have to be dealt with in future experiments. The main issue faced during this experiment was balancing the hyperparameters. I needed to choose between better results and a faster training speed + a larger model. The workaround was creating a model that had a decent score and also a fast training speed.

4. FUTURE EXPERIMENT
Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective.

4.a. Key Learning	Reflect on the outcome of the experiment and list the new insights you gained from it. Provide rationale for pursuing more experimentation with the current approach or call out if you think it is a dead end. Further experimentation is not needed as the model is ready for deployment. However, if a new dataset with more months on it appears, then it is best to retrain the model using that new dataset as the current one has its limitations.

4.b. Suggestions / Recommendations	Given the results achieved and the overall objective of the project, list the potential next steps and experiments. For each of them assess the expected uplift or gains and rank them accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production. Next steps is to deploy the model into the streamlit application.