

What are you to do?

Ahmed is a customer of Sastaticket.pk. He is planning to fly from Karachi to Islamabad for his brother's wedding and is currently in the process of choosing tickets. Ahmed has to go to Islamabad but Ahmed also wants to save some money in the process, so he chooses to wait instead of buy now, simply because ticket prices are just too high.

Is this the right decision? Won't ticket prices increase in the future? Perhaps there is a sweet-spot Ahmed is hoping to find and maybe he just might find it.

This is the problem that you will be tackling in this competition. Can you predict future prices accurately to such a degree that you can now tell Ahmed - with confidence - that he has made the wrong decision.

Your task boils down to generating optimal predictions for flight prices of multiple airlines. If successful, your model will contribute greatly to Sastaticket's rich and diverse set of operating algorithms.

How will participants do it?

Sastaticket will provide you with a train/test dataset. This synthetic dataset contains records which act as search data. For example, if someone was to come on to Sastaticket's website and search for a flight, this data would be recorded.

Dataset X includes the following features:

- f1: Ticket Purchase Date Time
- f2: Origin
- f3: Destination
- f4: Departure Date Time
- f5: Arrival Date Time
- f6: Airline
- f7: Refundable Ticket
- f8: Baggage Weight
- f9: Baggage Pieces
- f10: Flight Number

Dataset Y will have the following variables:

- Target

Using these features, you are expected to generate valuable insights through in-depth data analysis. Supplement your analysis with the appropriate diagrams and make sure to dig deep into the data as there is a lot to discover!

After conducting the exploratory analysis, you will be required to leverage these insights to generate new features. The features you engineer will be pivotal in either making or breaking your model.

Finally, You are expected to create a model that will output predictions against a datetime index.

The evaluation metric will be Root Mean Squared Error (RMSE). This metric will be used to evaluate the performance of your model as well as decide your position in the leaderboard. However, RMSE is not the sole metric through which the winner will be determined.

Please keep in mind that you are required to document your journey throughout the competition. The documentation should include the explanation of your thought process, think of it as a journal in which you are brainstorming and a journal we are very interested in!

The winning participant/team will be decided after a thorough evaluation of the documentation, the EDA, the final RMSE score and the efficiency of the code.

Note: There will be an extra column 'Unnamed' Please remove that.