Data Science Internship Assignment

Problem: Delivery Date Estimation

At BestBuy we do our best to provide an excellent shopping experience for our customers. In the past couple of years, our e-commerce has significantly grown and now our customers can conveniently browse, search and order within a few clicks. One of our goals in the e-commerce team is to provide an accurate estimate of delivery time to customers' addresses on our website. As this is critical for our business, an overestimation of delivery time can make offerings less competitive on the market against other retailers. An unrealistic short delivery time leaves our customers unhappy.

The operations team believes that the current method of delivery estimation has room for improvement. They approached the Analytics team to provide a better solution to replace the current method. Several factors such as size of the box, time of delivery, region, weather condition, courier, time of the year can affect the delivery time. All these data are provided in file train_data.csv and test_data.csv.

Q1) Can you build a model based on data_train.csv to estimate the delivery time? Run your model on test data and submit the result in a single column csv file. Put your name as the name of the csv file.

Q2) How do you justify your proposed model?

Please submit your code and your prediction file.

Output Format:

- Single file R or Python
- A single column csv with your predictions for the testing set

Data Dictionary

Order-Date: date in format YYYY-MM-DD

Delivery-Date: date in format YYYY-MM-DD

Promised-Date: date in format YYYY-MM-DD

Courier: Name of the courier

Size of the box height: Height of the box in cm

Size of the box width: Width of the box in cm

Size of the box length: Length of the box in cm

Size of the box weight: Weight of the parcel in Kg

Weather condition: Good, Bad

Store address: Address of the store the orders were shipped from; longitude/latitude

Customer address: Address delivery; longitude/latitude