Problem Statement:

A company schedules delivery orders for its customers through 3 main channels:
1. Customer Service Agents who call customers to schedule drop offs
2. Customers who are sent an sms that provides a link for a web page. The customer can then choose their location and the page will dynamically show available time slots for that 'zone' when the package can be delivered based on the number of drivers available at the given time
3. Through a mobile application where a similar flow to the 'web channel occurs i.e. customers choose their locations('zone') and can then choose from a list of available timeslots to schedule the delivery.

From here onward and at the time close to delivery, drivers are then instructed to deliver at/near the "Scheduled Coordinates" (the column called "Delivered Coordinates" in the data set provided)

This task requires you to complete the analysis of providing accurate "Delivered Coordinates". In particular, we ask you to apply tools of machine learning and data science in general to predict "Delivered Coordinates" given the historical data of placed orders that are successfully delivered (please refer to the data set provided).

- Given the current data identify which features impact the accuracy of deliveries and suggest which features to optimize in order to improve this.

- Given the current data predict the accuracy of the "delivery coordinate" with a max geographic error of 500 metres.

- How would you minimize the time taken to train the model if you have to make an online model?

Deliverables
Provide a link to a github repository containing the following
- Automated script that runs your model on a test dataset given as input (Test dataset provided below). Output needs to be written out to a file
- Model code
- Answers to the questions asked in the problem statement

Data set description

Training Dataset - Data Model
1. Driver ID = 478 (Number)
2. Tracking ID = 123456789012 (Number)
3. CSA ID = 4354 (Number)
4. Time slot from = 0900 (String)
5. Time slot to = 1800 (String)
6. Commitment Date = "2017-01-04" (Date)
7. Customer Address = "C tower flat 1111 Dubai Marina Dubai" (String)
8. Delivered Date-Time = "2017-01-04T09:50:19" (DateTIme)
9. Customer name(ID) = 78 (Number)
10. Customer phone number = 501424343 (Number)
11. Product description = "kids toys" (String)
12. Supplier name(ID) = 23 (Number)
13. Scheduled channel = CSA/MWEB/iOS/Android (String)
14. Scheduled Coordinates = 24.3423423, 54.223543 (String)
15. Delivered Coordinates = 24.2423, 54.2343 (String)

Dataset given for 4 different months

Training Data:
- https://s3-eu-west-1.amazonaws.com/fetchr-datascience/anon_dataset_01_2017.csv
- https://s3-eu-west-1.amazonaws.com/fetchr-datascience/anon_dataset_11_2016.csv
- https://s3-eu-west-1.amazonaws.com/fetchr-datascience/anon_dataset_12_2016.csv
- https://s3-eu-west-1.amazonaws.com/fetchr-datascience/anon_dataset_10_2016.csv

Test Data:
- https://s3-eu-west-1.amazonaws.com/fetchr-datascience/test_dataset_02_2017.csv