Problem Statement:

The objective of this assignment is to build a recommendation engine to predict what restaurants customers are most likely to order from given the customer location, restaurant information, and the customer order history.

Dataset Overview:

There are 9768 customers in the test set. These are the customers you will need to recommend a vendor to. Each customer can order from multiple locations (LOC_NUM).

There are 34674 customers in the train set. Some of these customers have made orders at least one of 100 vendors.

Variable Definitions:

Train Customers

Information on the customers in the training set.

'customer_id': Unique customer ID, used in train_locations and train_orders

'gender': Customer gender

'dob': Birth Year (if entered)

'status' and 'verified': Account status

'language': Chosen language

'Created_at' and 'updated_at': dates when account was created/updated

Train Locations

Each customer orders from one or more locations. Each is assigned a location number.

'customer_id': The unique customer ID

'location_number': Location number (most customers have one or two)

'location_type': Home, Work, Other or NA

'Latitude' and 'longitude': Not true latitude and longitude - locations have been masked, but nearby locations remain nearby in the new reference frame and can thus be used for clustering. However, not all locations are useful due to GPS errors and missing data - you may want to treat outliers separately.

Train Orders

This is a record of all orders made by customers in the train set from the vendors included in this competition. Each order contains:

'order_id': The order ID used internally - can be ignored

'customer_id': The customer making the order, used to link with customer info

'item_count': how many items were in the order

'grand_total': total cost

Payment related columns: 'payment_mode', 'Promo_code', 'vendor_discount_amount', 'Promo_code_discount_percentage'

Vendor related columns: 'is_favorite', 'is_rated', 'vendor_rating', 'driver_rating',

Order details: 'deliverydistance', 'preparationtime', 'delivery_time', 'order_accepted_time',

 'driver_accepted_time', 'ready_for_pickup_time', 'picked_up_time', 'delivered_time',

 'delivery_date','created_at'

 'vendor_id': the unique ID of the vendor

 'LOCATION_NUMBER': The location number specifies which of the customers locations the delivery was made to

 'LOCATION_TYPE': same as location type in the train_locations table

'CID X LOC_NUM X VENDOR': Customer ID, location number and Vendor number as used in the submission file.

Vendors

Contains info on the different vendors. Important columns are:

'id': The vendor ID used for the competition

'latitude' and 'longitude' : masked the same way as the customer locations

'vendor_tag_name': Tags describing the vendor

Other columns are mostly self-explanatory.

Note : Test Customers and Test Locations follow the same format as train.