The first thing any visitor to India will take in — probably while staring out the window in awe as their aeroplane descends is the sheer size of this country. It is densely populated and patchworked with distinct neighbourhoods, each with its own culinary identity. It would take several lifetimes to get to know all of the street stands, holes in the wall, neighbourhood favourites, and high-end destinations in this city.And for Indians dining out is and always will be a joyous occasion. Everyone has their own favourite restaurants in the city starting from the street food stall across the street to the 5-star restaurants in the heart of the city. Some are favourites because of the memory attached to it and some are favourites because of the fact that the place has a fantastic ambience. There are a lot of other factors as well which contribute to the likeness of the restaurants which in turn determines their popularity among masses. 

If you look at this from the business perspective for a restaurant, more popularity may mean more visits to the joint increasing the annual turnover of the restaurants. For any restaurant to survive and do well, the annual turnover of the restaurants has to be substantial. 

This problem takes a shot at predicting the annual turnover of a set of restaurants across India based on a set of variables given in the data set. This includes the data related to the restaurant such as location, opening date, cuisine type, themes etc. This also includes data pooled from different sources such as social media popularity index, Zomato ratings, etc. Lastly, it also adds a different flavour to the problem by looking at the Customer survey data as well as ratings provided by a mystery visitor data (audit done by a third party). 

Working with Data

Data has been split into two groups and provided in the module:

* **Train dataset**
* **Test dataset**

The train dataset set is used to build your machine learning model. For the training dataset, we provide the Turnover of the restaurant (also known as the variable Turnover) for each participant.

The test dataset should be used to see how well your model performs on unseen data. For the test dataset, it is your job to predict the Turnover of the restaurant (Turnover) for each participant.

In **Data Dictionary.csv,** the detail of all the variables used in the train and test datasets is given