# Business Problem

In this project you will be working on a fictitious company (ABC Pte Ltd) and use the given data to help the company predict and identify fraudulent transactions. ABC Pte Ltd is an e-commerce company that provides an online e-commerce portal for online shopping.

**Download the given dataset ‘Fraud\_Data\_Final.xlsx’ from MeL**. You may also use the following link and go through the different tabs in the page to understand each column. You will be required to create a new account on the site.

<https://www.kaggle.com/preritsaxena/fraud-detection>

**Note:**

You are only allowed to use the given data to do the project.

# Data Dictionary and Dataset

|  |  |  |
| --- | --- | --- |
| **Columns/Fields** | **Description** | **Type** |
| user\_id | System auto-generated identification number for each user | String |
| purchase\_value | Amount paid for the purchase | Number  ($0-$400) |
| device\_id | Identification number for the device the customer used to make the purchase | String |
| source | Social media source where customers click to bring them to the e-commerce site | String  (Ads, Direct, SEO) |
| browser | Internet browser customers used to make the purchase at the e-commerce site | String  (Chrome, FireFox, IE, Opera, Safari) |
| sex | Gender of the customer | String  (M, F) |
| age | Age of customer in years | Number  (0-100 yrs old) |
| ip\_address | IP Address of the customer device | String |
| days | Days between customer signing up an account on the site to making first purchase | Number |
| is\_fraud | Whether the particular transaction is fraudulent | String  (0,1) |

Download the dataset **in MeL.**

# Tasks (Coverage: Week 1 to 3)

You are required to do the below after importing the dataset:

1. Create non-visual or visual plots to inspect each categorical and numeric field.
2. Identify potential data quality issues including data types.
3. Clean up the data for the issues identified.
4. Perform further data exploration on the cleaned data to uncover any interesting insight(s) using any two visual plots of your choice.