**The Client**

XYZ is a private equity firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry.

**The Engagement**

You have been provided with multiple data sets that contains information on 2 cab companies. Each file (data set) provided represents different aspects of the customer profile. XYZ is interested in using your actionable insights to help them identify the right company to make their investment.

**The Data**

You have been provided 4 individual data sets.

Below are the list of datasets which are provided for the analysis:

**Cab\_Data.csv –** this file includes details of transaction for 2 cab companies

**Customer\_ID.csv** – this is a mapping table that contains a unique identifier which links the customer’s demographic details

**Transaction\_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode

**City.csv –** this file contains list of US cities, their population and number of cab users

You should fully investigate and understand each data set.

* Understand the field names and data types
* Identify and remove duplicates
* Identify any biasness existing in the data

**Be Creative**

The idea is to create a hypothesis, engage with the data, think critically, and use various analytical approaches to produce unique insights.

**Analysis**

**Create a hypothesis and investigate:**

Areas to investigate:

* Which company has maximum cab users at a particular time period?
* Does margin proportionally increase with increase in number of customers?
* What are the attributes of these customer segments?