

# Module 10: Get Industry Ready

---

## Industry Grade Project

edureka!

**edureka!**

© Brain4ce Education Solutions Pvt. Ltd.

## Project Statement

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.

You have been provided with multiple data sets that contains information on two cab companies i.e. Black and Red cabs. 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 outcome of your analysis will be a presentation to XYZ's Executive team. This presentation will be judged based on the visuals provided, the quality of your analysis and the value of your recommendations and insights.

### Data:

You have been provided 4 individual data sets. Time period of data is from 31/01/2016 to 31/12/2018.

Below is 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:

- Review the Source Documentation
- Understand the field names and data types
- Identify relationships across the files
- Field/feature transformations
- Determine which files should be joined versus which ones should be appended

- Aggregate the data into smaller files and define relationships
- Identify and remove duplicates
- Identify any biasness existing in the data

The idea is to come up with the analysis, engage with the data, think critically, and use various analytical approaches to produce unique insights.

You are not limited to only utilizing the data you have been provided.

We encourage you to find third party data sets which correspond to the overall theme and geographical properties of the data provided. **For Example:** you can leverage US weather data or census data.

Also, do research on overall cab industry in US and try to relate that with the trend in data.

Once you are ready with the data come up with the following insights:

1. Which company has maximum cab users at a particular time period?
2. Does margin proportionally increase with increase in number of customers?
3. Find the Customer distribution based on Gender for both the Companies?
4. Top 10 cities by Kms travelled.
5. Payment method distribution based on the cabs.
6. Customer age distribution based on the cab services.
7. Find the customer distribution by income.
8. Which company has the highest market share based on the cities?
9. Create a chart which will show the average price comparison between the companies.
10. Data distribution of the cab companies based on Users.
11. Create a cluster based on Kms travelled, Age, Price and Income.
12. Create dashboard(s) for each company to support your story.