

**For this project, I used Microsoft Excel to perform the required exploratory data analysis.**

Some of the key functions and methods I used included:

1. Transaction\_ID dataset

=COUNTIF(B\$2:B\$440099,B15) for **checking how many times a customer's ID** appeared against a transaction.

Customer_ID Frequency
3
4
4
4
4

=SUM(F2:F7) for **summing the total number of transaction performed** in the given time frame.

Total Transactions
440098

## 2. Customer\_ID dataset

=FREQUENCY(C2:C49172,H2:H6) for **clustering the customers age groups** into the given bin range which was

0=<15
16-30
31-45
46-60
61-75

=SUM(F2:F7) for summing **the total number of customers of the total survey** as per classification given in each bin count cluster.

<b>Total Customers</b>
49171

### 3. City Dataset

=C4/B4 for **dividing the users against the city total population** to get the actual ratio between the two variables.

if(E2>=25%, "worthy of note!", if(E2<=25%, "not worth investment")) for classifying the worthiness of investing in the cab business in the given city if the cab usage to city population ratio was above 25%.

Verdict
not worth investment
not worth investment
worthy of note!
not worth investment
not worth investment

#### 4. Cab\_Data Dataset

=F2-G2 for finding the difference between the price per trip and actual amount charged on each transaction.

Profit margin
\$57.32
\$23.67
\$27.57
\$25.80
\$16.84

=SUMIF(C2:C359393,I2,H2:H359393) for summing the total profit margin for the pink cab company.

Cab Company	Cab Company Profit Market share
Pink Cab	\$5,307,328.32

=SUMIF(C2:C359393,I3,H2:H359393) for summing the total profit margin for the yellow cab company.

Cab Company	Cab Company Profit Market share
Yellow Cab	\$44,020,373.17

=sum(J2:J3) for summing the total profit margin amassed by both pink and yellow cab companies.

Total Profit Margin
\$49,327,701.49