Data Intake Report

Name : G2M - Go to Market
Report date : 20 March 2022
Internship Batch : LISUM07
Version : Python 2.7.16
Data intake by : Vu Hoang-Thuy-Duong
Data intake reviewer : Vu Hoang-Thuy-Duong
Data storage location : GitHub

Tabular data details

File 0 : mainEDA.csv

Total number of observations	35
Total number of files	1
Total number of features	11
Base format of the file	.csv
Size of the data	364 KB

File 1 : Cab_Data.csv

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

File 2 : City.csv

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	759 bytes

File 3 : Customer_ID.csv

Total number of observations	49171
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	1.1 MB

File 4 : Transaction_ID.csv

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

Conclusion

Pink Cab

Pink Cab is a ridesharing app based in Trinidad and Tobago that allows female drivers to provide safe transit for women and children only. That's right! It's driven by women for women.

Yellow Cab

Yellow cab dates back to at least 1798, when the musical comedy, Cabriolet Jaune (Yellow Cab), debuted at Paris' Theater de l'Opera Comique National. Yellow cabs were known in Paris and London throughout most of the 1800s. A yellow cab company shook up the New York Cab system in the mid-1880s, offering cheaper, more predictable fares than competitors. One of the first automobile cabs in London, in the 1890s, was a yellow electric automobile.
There are many yellow cab taxicab operators around the world (some with common heritage, some without). The original Yellow Cab Company, based in Chicago, Illinois, was one of the largest taxicab companies in the United States.

Basic conclusion from the united database

After reunited all data from 4 given datasets, we only got 1 dataset of 35 rows x 11 columns, representing 35 Pink Cab' travelers and their information : Transaction ID, Customer ID, How long their trip, Gender, Age …
The trips provided by Pink Cab on this main dataset vary between 1.96km to 44km, with a mean of approximately 22.47km and the median is 23.2km. The price charged varies between 27.71 to 646.06, with a mean of 314.635 and the median is 324.21. The cost of a trip is often lower

than the price charged, which varies from 20.776 to 482.04, with a mean of 250.06 and the median is 259.74. The travelers' age varies from 18 to 63 years old.
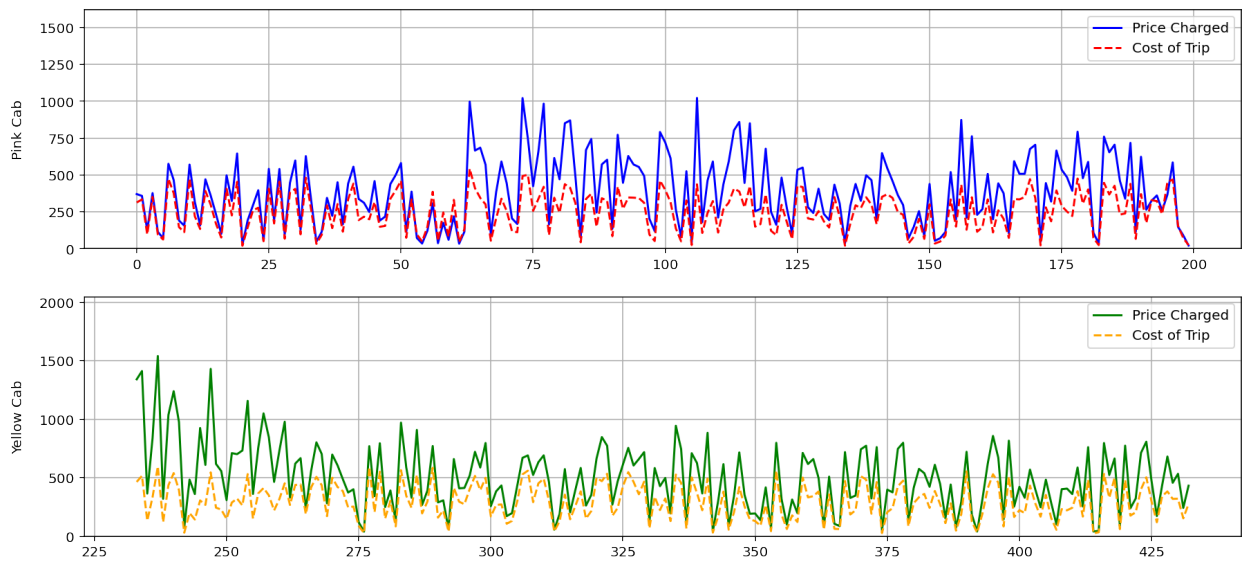

Pink Cab - City diffusion

Pink Cab - City Price Charged - Cost of Trip
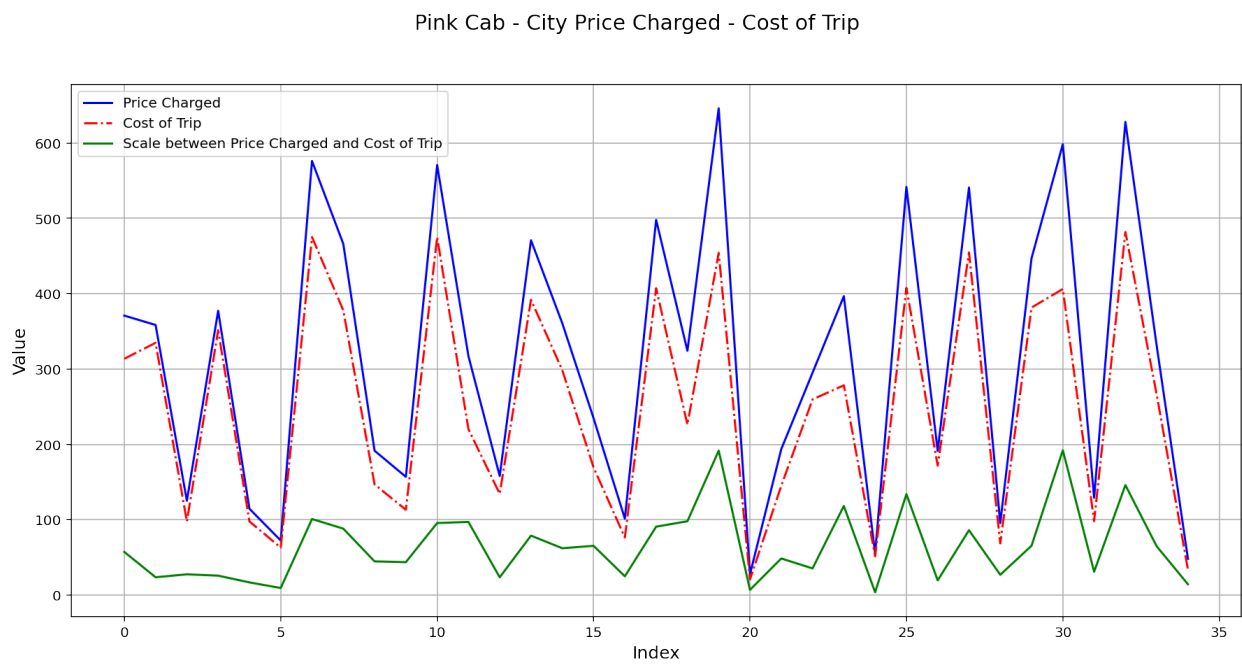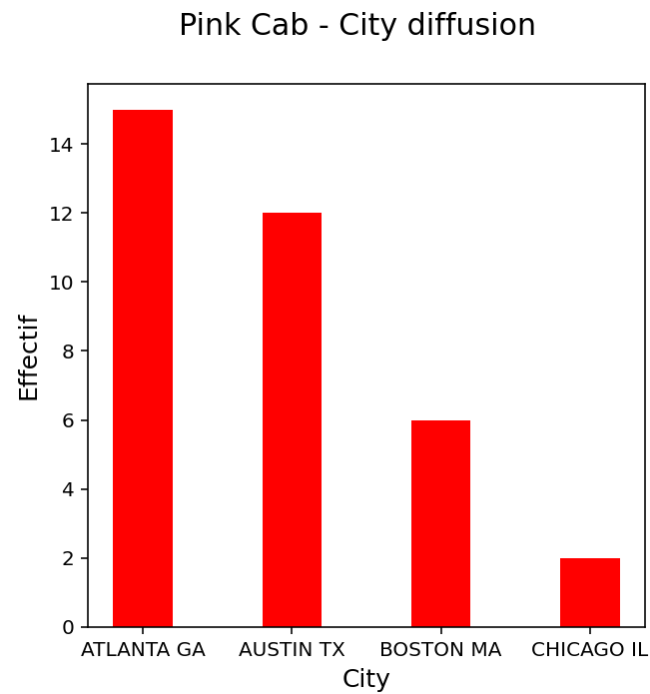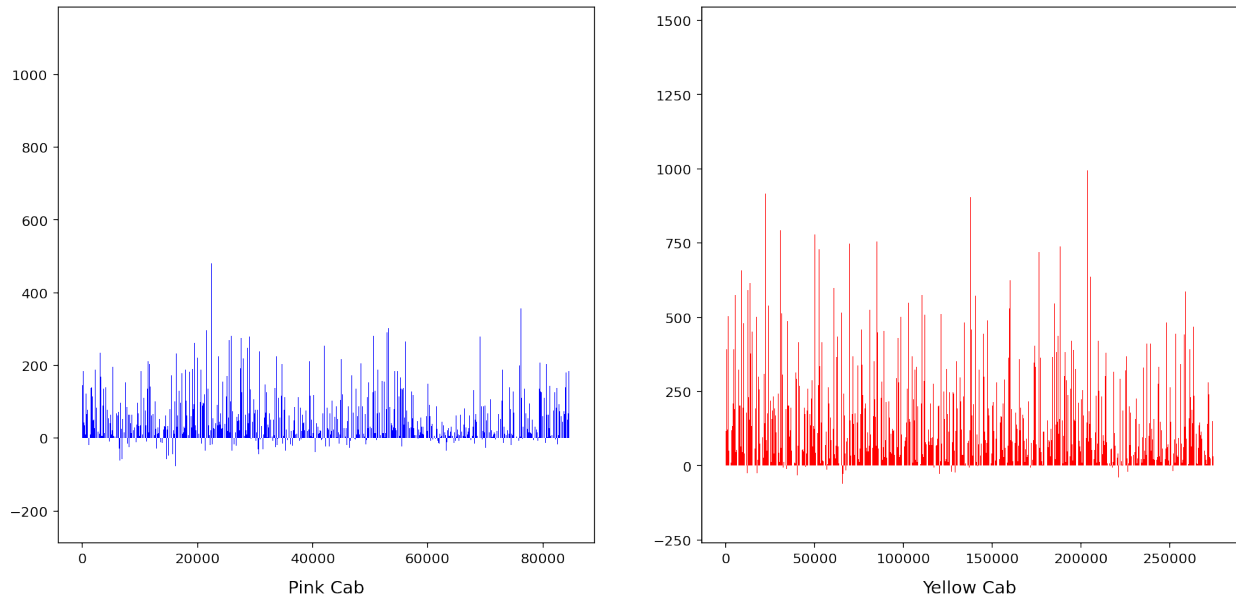
Pink Cab - City diffusion
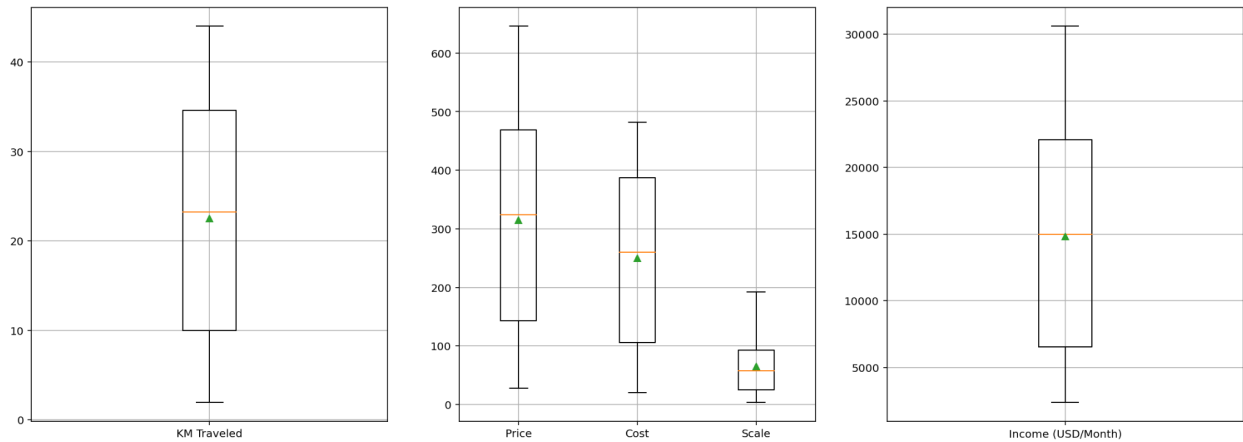
Common conclusion among 4 datasets

There is always a scale between Price Charged and Cost of Trip. Cost of Trip is often lower than Price Charged. Sometimes it might be higher, but the total is too low to affect any changes.

Price Charged - Cost of Trip between Pink Cab - Yellow Cab

At Pink Cab, there are 11 129 values that are negative upon 84 711 trips provided, such as 13.14%. At Yellow Cab, there are 13 690 values that are negative upon 274 681 trips provided, such as ~ 5%.

Pink Cab - Yellow Cab scalar between Price Charged - Cost of Trip
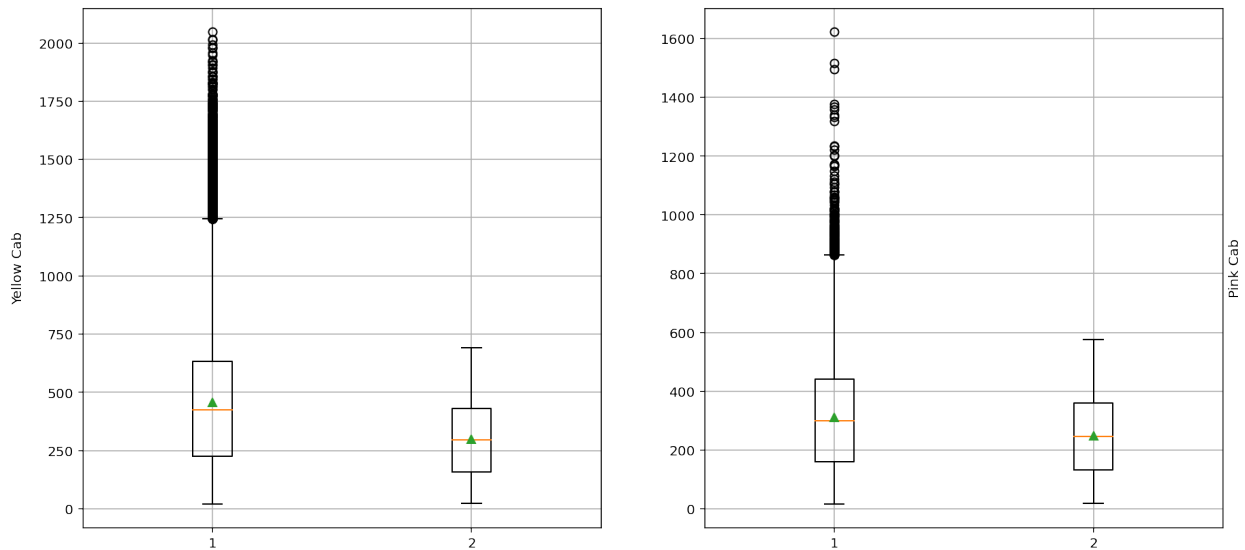
Despite this difference between 2 companies, their travelers stay nearly the same. Most of the trip distance varies from 30-40km, and the least is from 0-10km. Most of the travelers' age varies from 30-40 years old, and the least is also 10-20 years-old (the youngest age range). There are more men who travel by cab than women according to the graph below. Most

payment modes are realized by card.
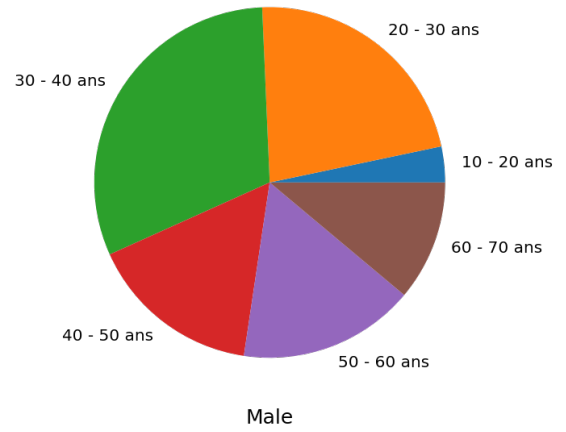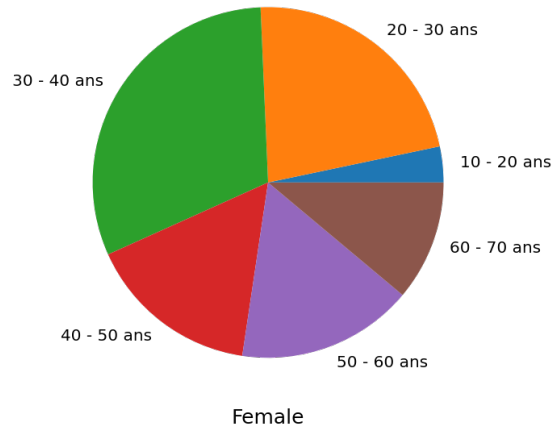
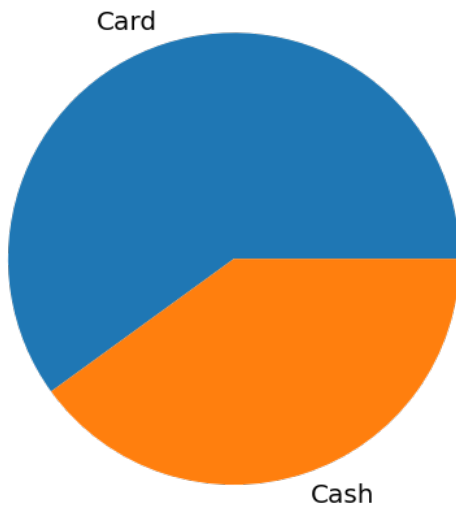Travellers trend between 2 companies
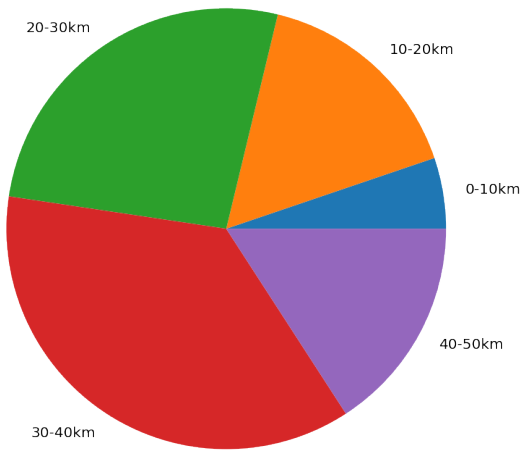
Pink Cab Travellers Trend
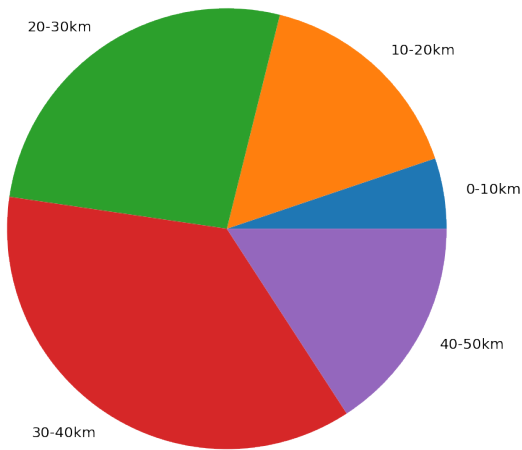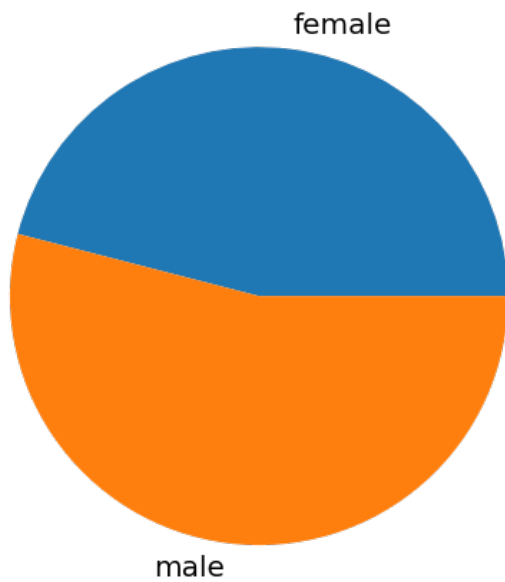

Yellow Cab Travellers Trend

Age range

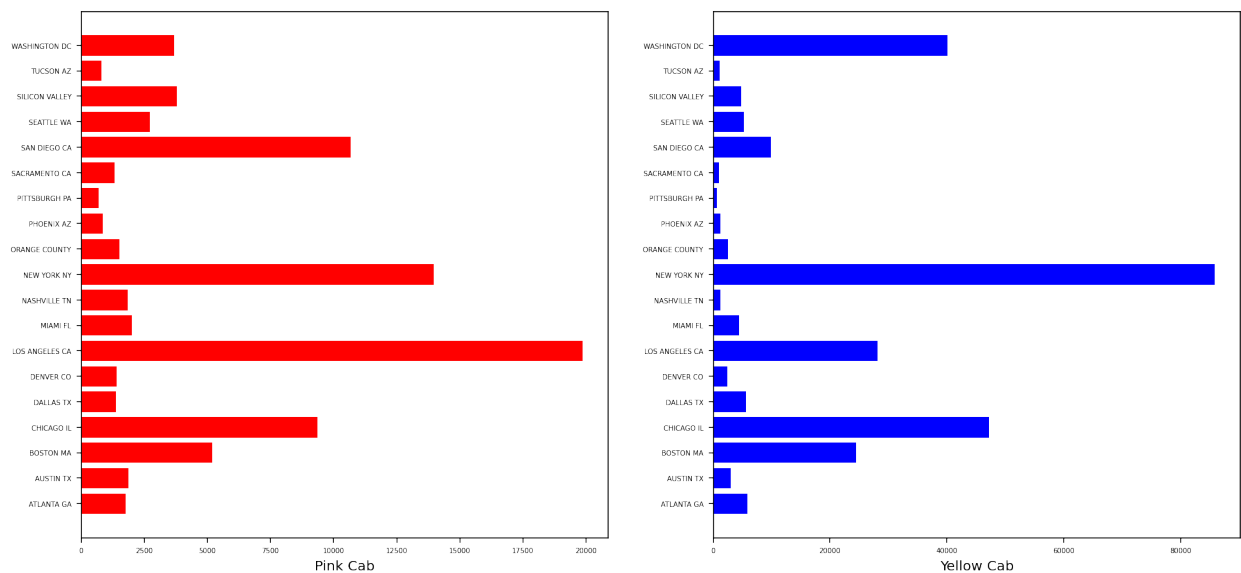
Female

Male

Payment Mode Range

Female - Male Proportion

female

male

As we can see, the number of Yellow Cab travelers is 4 times higher than Pink Cab travelers, especially in 4 main cities, which are Economical Central of the US : Washington DC, New York, Chicago IL and Los Angeles CA and another smaller is San Diego CA.

City diffusion between 2 companies

City diffusion between 2 companies

Scale between City Population and City Users