



Data Glacier

Your Deep Learning Partner

G2M insight for Cab Investment firm

EDA analysis

19-Jun-2021

Mukhammadjon Kholmirzaev

Background

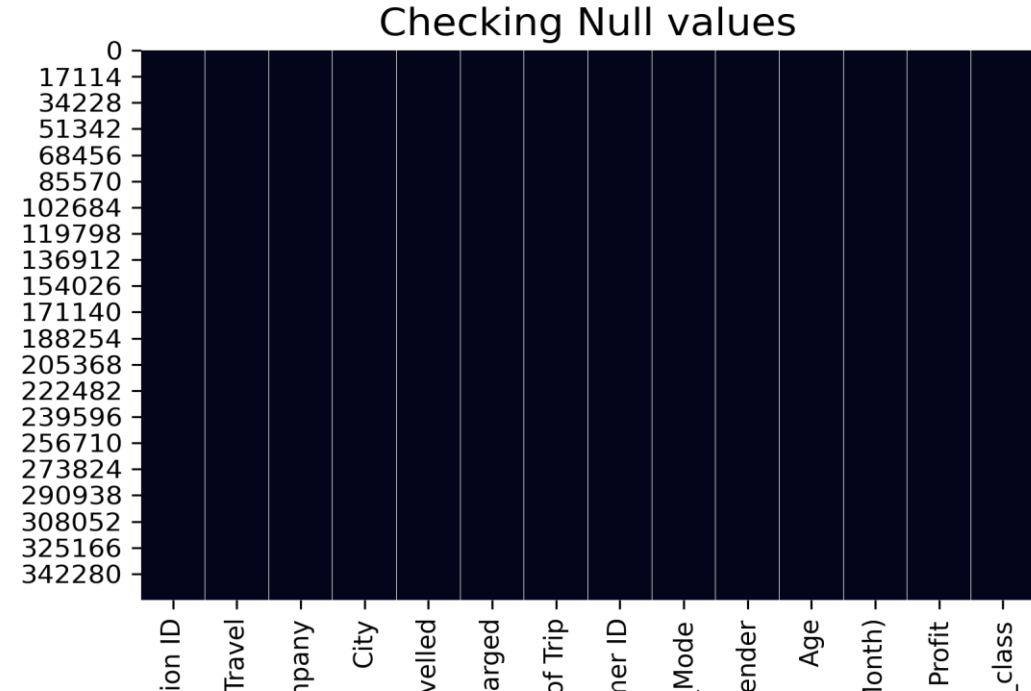
- XYZ is a private equity firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry. (Yellow Cab, Pink Cab)
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

Contents of EDA analysis:

- Data Exploration
- Statistical Analysis
- Correlation Analysis
- Finding the most profitable Cab company
- Recommendations for investment

Data Exploration

- 14 Features(including 2 derived features)
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points : 359,392
- Firstly, the data frames merged by foreign_key values.
- In order to join dataframes, I've decided to use Inner join over Left join. Mainly, it only returns the matched rows in both dataframe also it is faster when we work with meta data and only need an intersection.
- After merging, there are no Null values and Duplicated



Statistical Analysis

	count	mean	std	min	25%	50%	75%	max
Transaction ID	359392.0	1.022076e+07	126805.803715	10000011.0	1.011081e+07	10221035.50	1.033094e+07	10440107.00
KM Travelled	359392.0	2.256725e+01	12.233526	1.9	1.200000e+01	22.44	3.296000e+01	48.00
Price Charged	359392.0	4.234433e+02	274.378911	15.6	2.064375e+02	386.36	5.836600e+02	2048.03
Cost of Trip	359392.0	2.861901e+02	157.993661	19.0	1.512000e+02	282.48	4.136832e+02	691.20
Customer ID	359392.0	1.919165e+04	21012.412463	1.0	2.705000e+03	7459.00	3.607800e+04	60000.00
Age	359392.0	3.533670e+01	12.594234	18.0	2.500000e+01	33.00	4.200000e+01	65.00
Income (USD/Month)	359392.0	1.504882e+04	7969.409482	2000.0	8.424000e+03	14685.00	2.103500e+04	35000.00

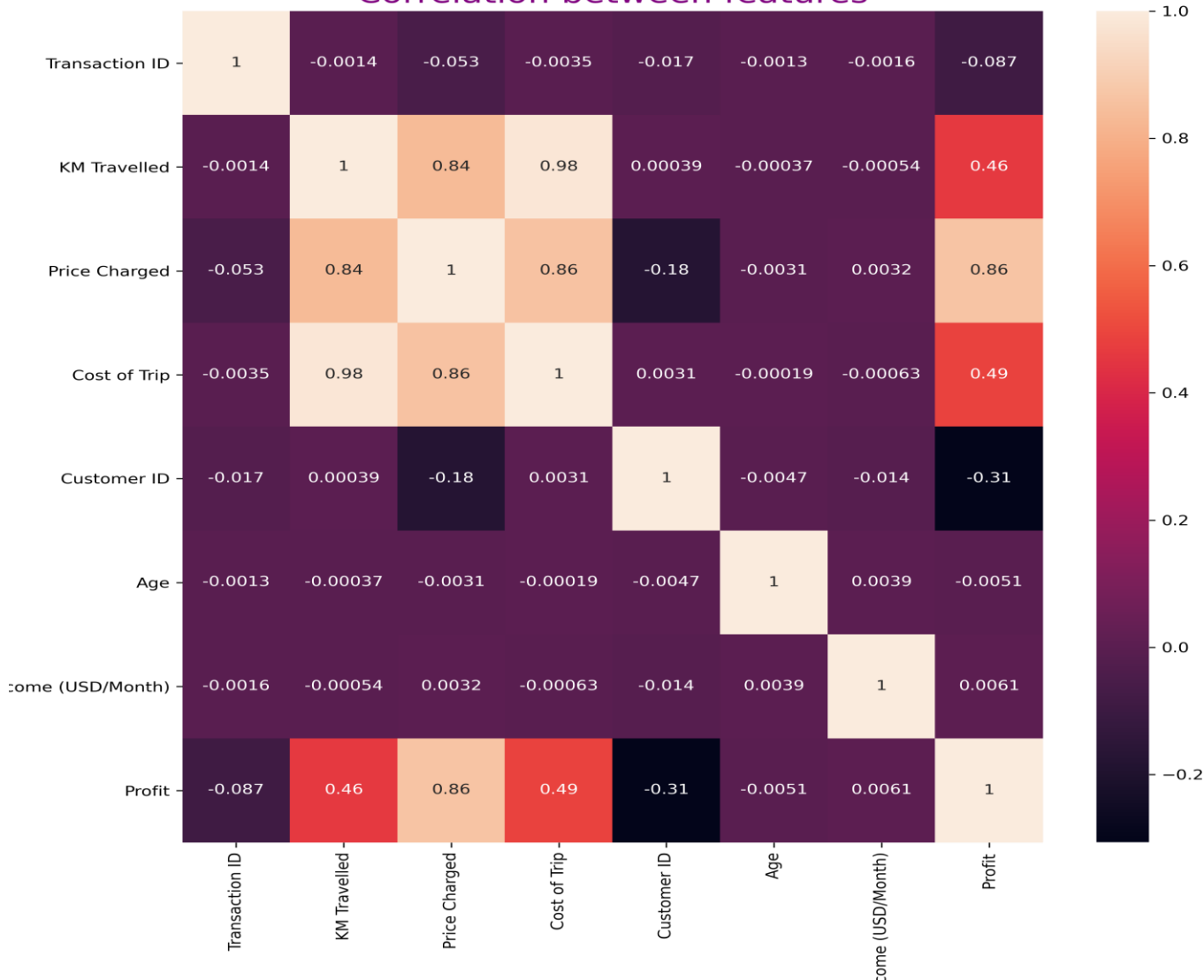
Statistics for
numerical Features

	count	unique	top	freq
Date of Travel	359392	1095	2018-01-05	2022
Company	359392	2	Yellow Cab	274681
City	359392	19	NEW YORK NY	99885
Payment_Mode	359392	2	Card	215504
Gender	359392	2	Male	205912

Statistics for categorical features

Correlation Analysis

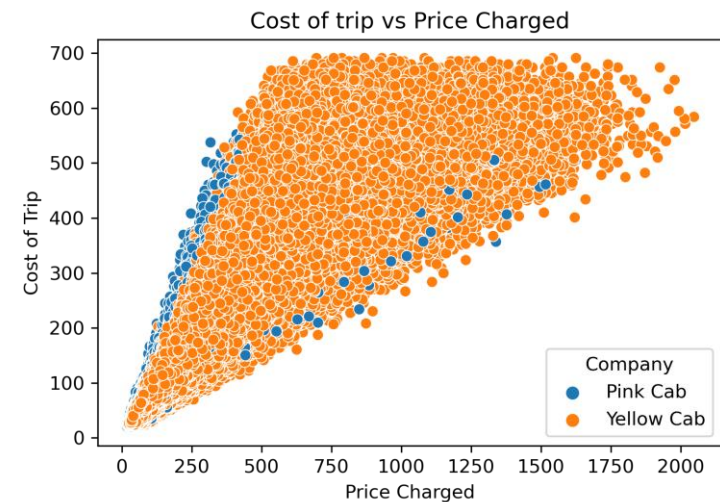
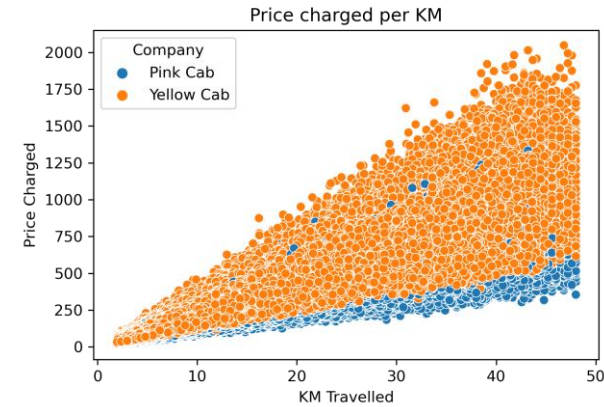
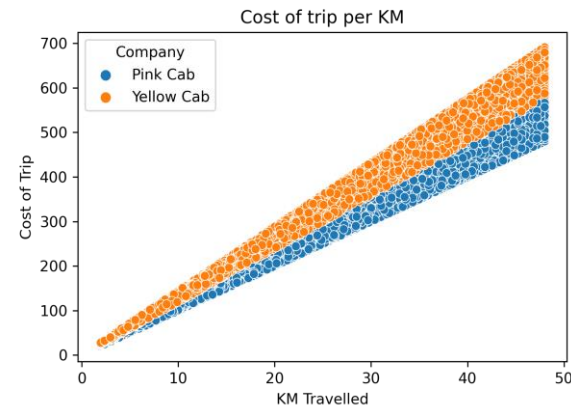
Correlation between features



- The following heatmap illustrates the correlation between numerical features.
- For the two companies, KM Traveled and Cost of Trip are more relevant than KM Traveled and Price Charged.
- The derived feature Profit has similar correlation to that three features.

Correlation Analysis

- There is significant difference on price between the companies as km increases.
- As we can see, Yellow cab charged more money than Pink Cab for longer distance.

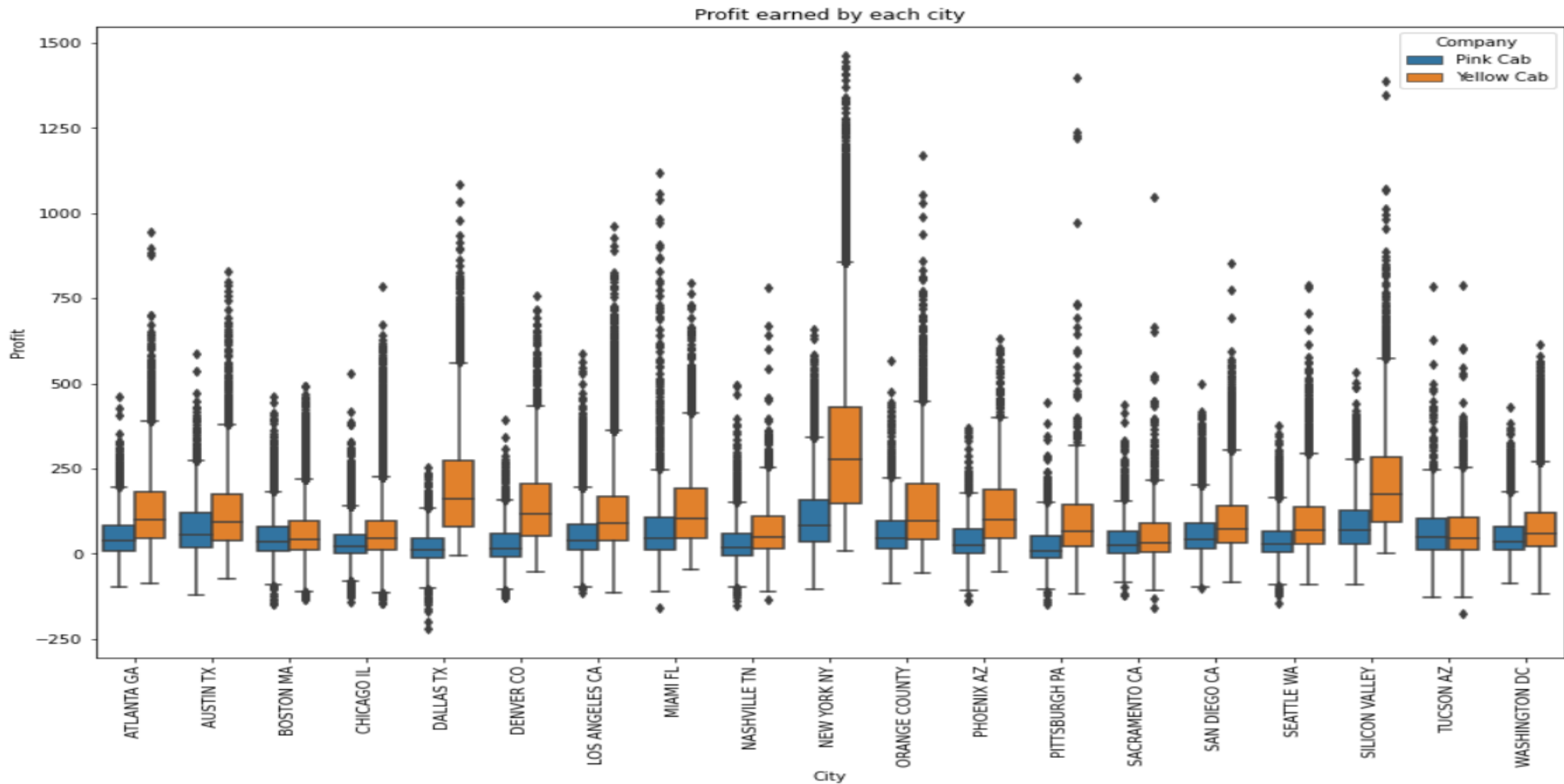


Profit Analysis

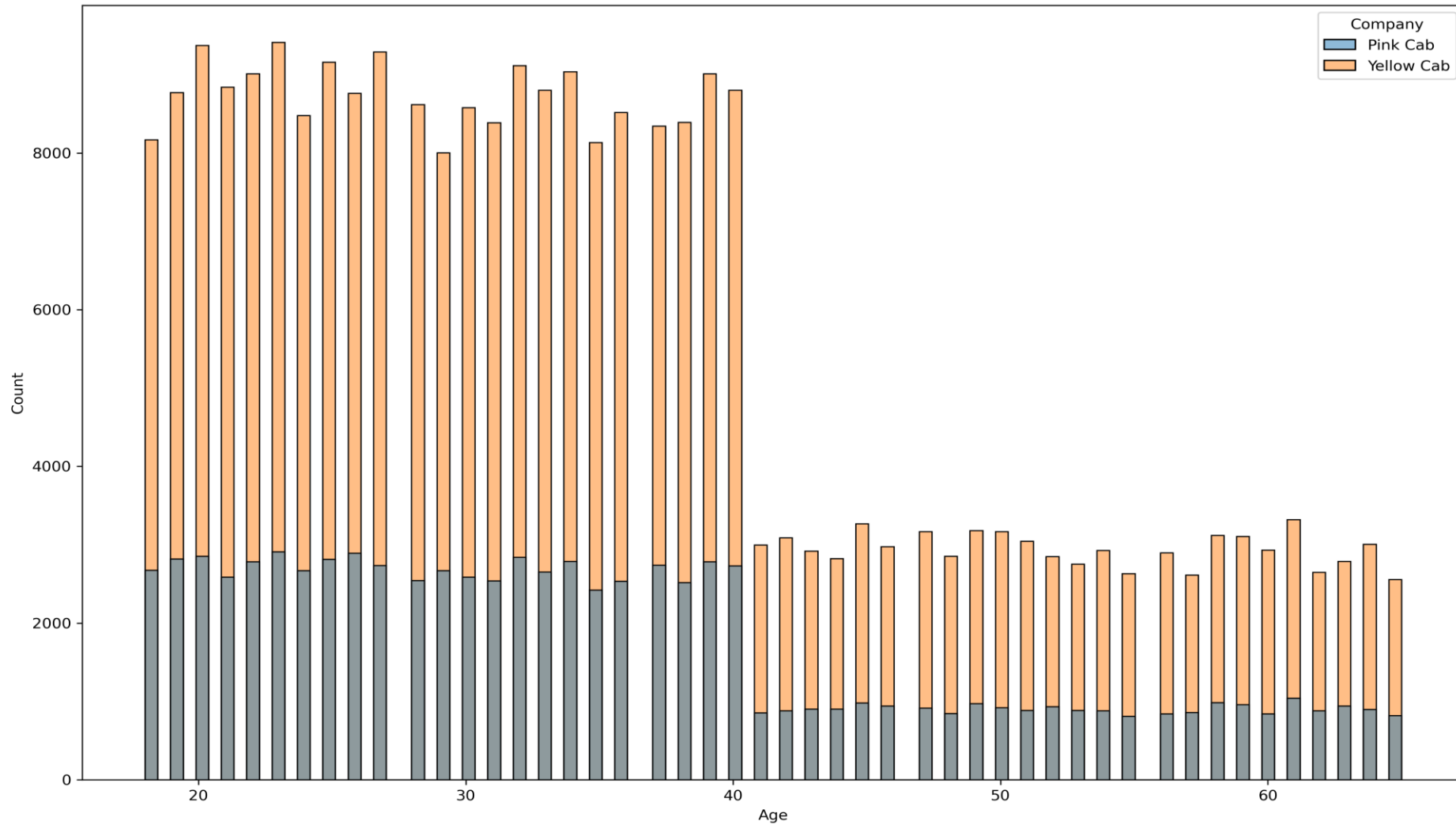
```
Company  
Pink Cab      62.652174  
Yellow Cab    160.259986  
Name: Profit, dtype: float64
```

The average of both companies clearly illustrates which company is dominant here

Profit Analysis by City

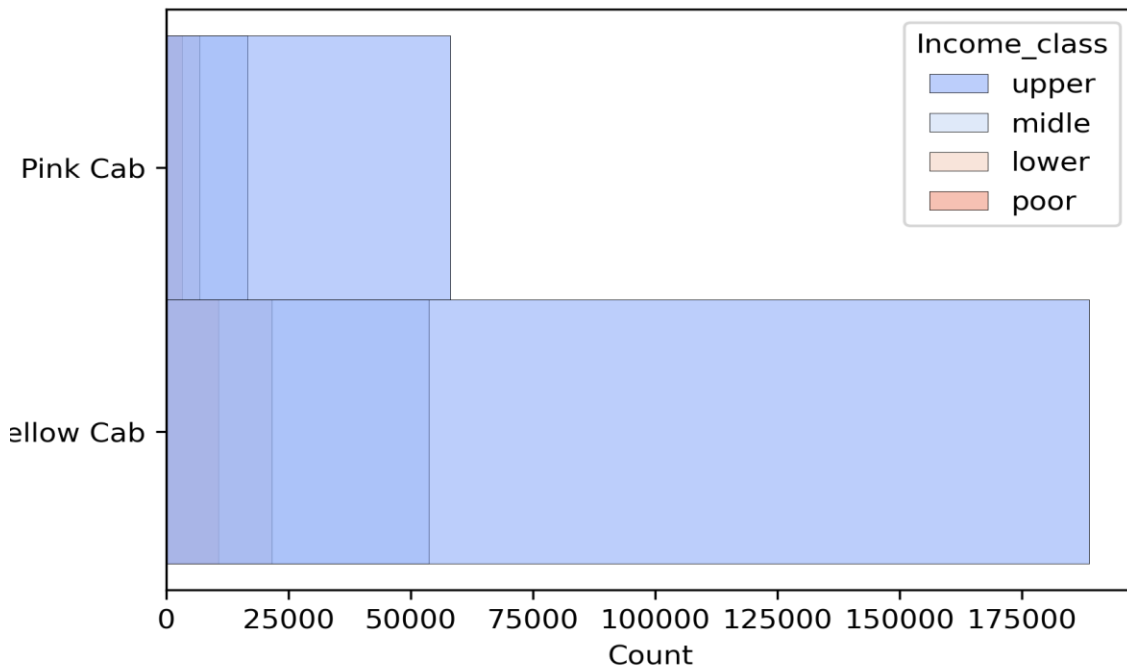


Age Group



Users aged on average 20-60.
Most users aged 20-40

Income Class Analysis

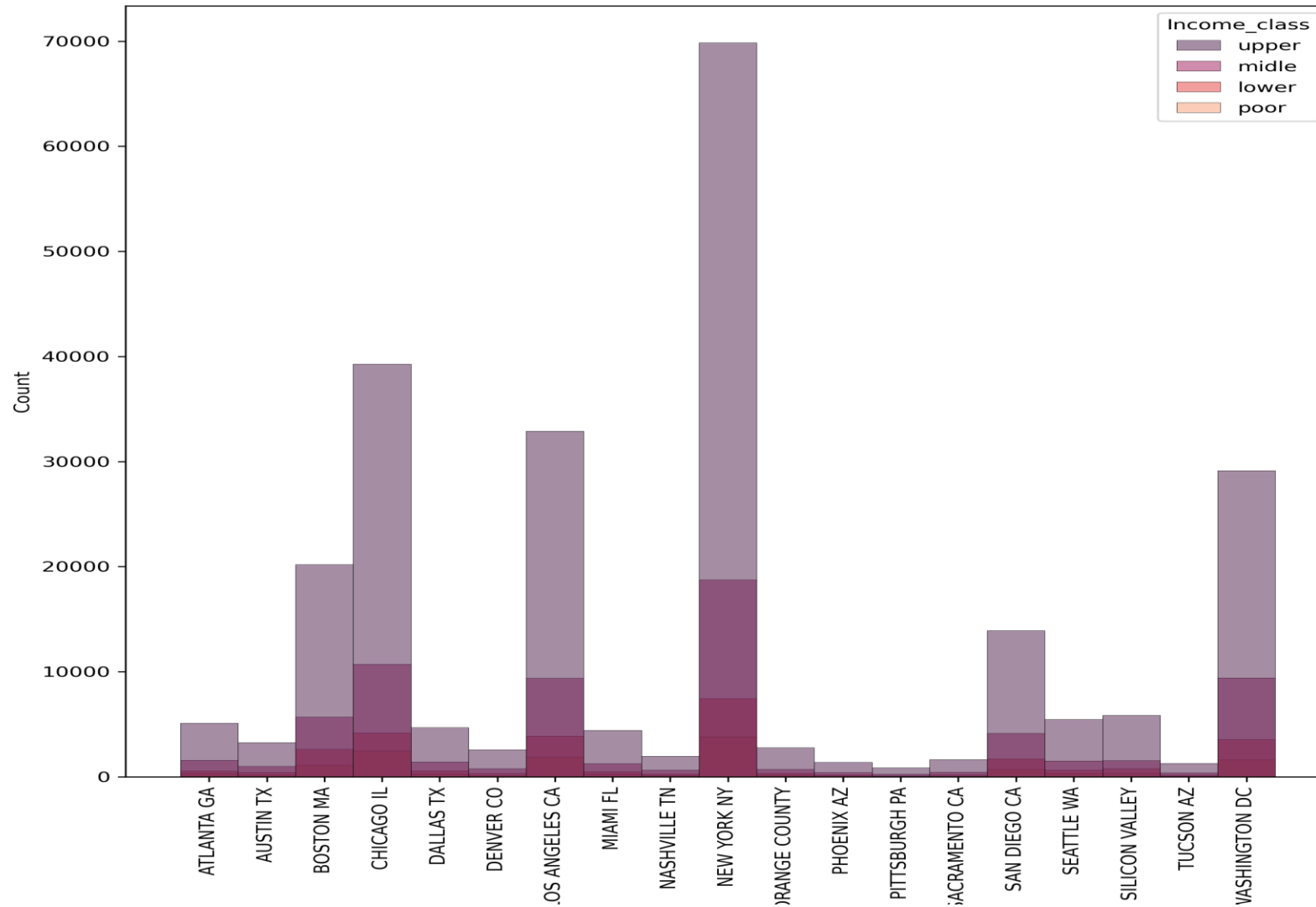


Income group → Income
Poor or near-poor → \$32,048 or less
Lower-middle class → \$32,048 - \$53,413
Middle class → \$53,413 - \$106,827
Upper-middle class → \$106,827 - \$373,894

Company	Income_class	
Pink Cab	lower	6792
	midle	16578
	poor	3321
	upper	58020
Yellow Cab	lower	21646
	midle	53655
	poor	10682
	upper	188698

Name: Income_class, dtype: int64

Income Class Analysis by city



Payment Mode by Income Class

Company	Income_class	Payment_Mode	
Pink Cab	lower	Card	4074
		Cash	2718
	midle	Card	9860
		Cash	6718
	poor	Card	2016
		Cash	1305
	upper	Card	34769
		Cash	23251
Yellow Cab	lower	Card	12972
		Cash	8674
	midle	Card	32218
		Cash	21437
	poor	Card	6393
		Cash	4289
	upper	Card	113202
		Cash	75496

Name: Income_class, dtype: int64



Recommendations

- **Overall,** We found that Yellow cab is a way better than Pink cab
- **Company Profit:** The Yellow Cab Company has higher profits on every aspect:
 - **Average:** : The average profit per kilometer of the yellow cab is nearly **3** times the average profit per kilometer of the pink cab
 - **Income level:** Most users of Cabs are in the middle and upper classes. Yellow cab has more upper level customers.
 - **City:** Profit of Yellow cab is significantly higher in many cities, except Tucson Az.
- **Customer Reach:** The yellow cabin has a higher customer reach in 25 cities whereas the pink cabin has a higher customer reach in 4 cities.
- **Transaction counts:** Yellow Cab has a larger number of transactions than Pink Cab.
- According to above points , it is highly recommended to invest Yellow cab. But also, we should consider that profit of Pink Cab is gradually increasing. However, We can surely say that Yellow cabin can maintain monopoly for several years.



Data Glacier

Your Deep Learning Partner

Your Deep Learning Partner

Thank You