Exploratory Data Analysis on Cab Investment Firm

Shuran Fu June 18, 2022

Agenda

Executive Summary

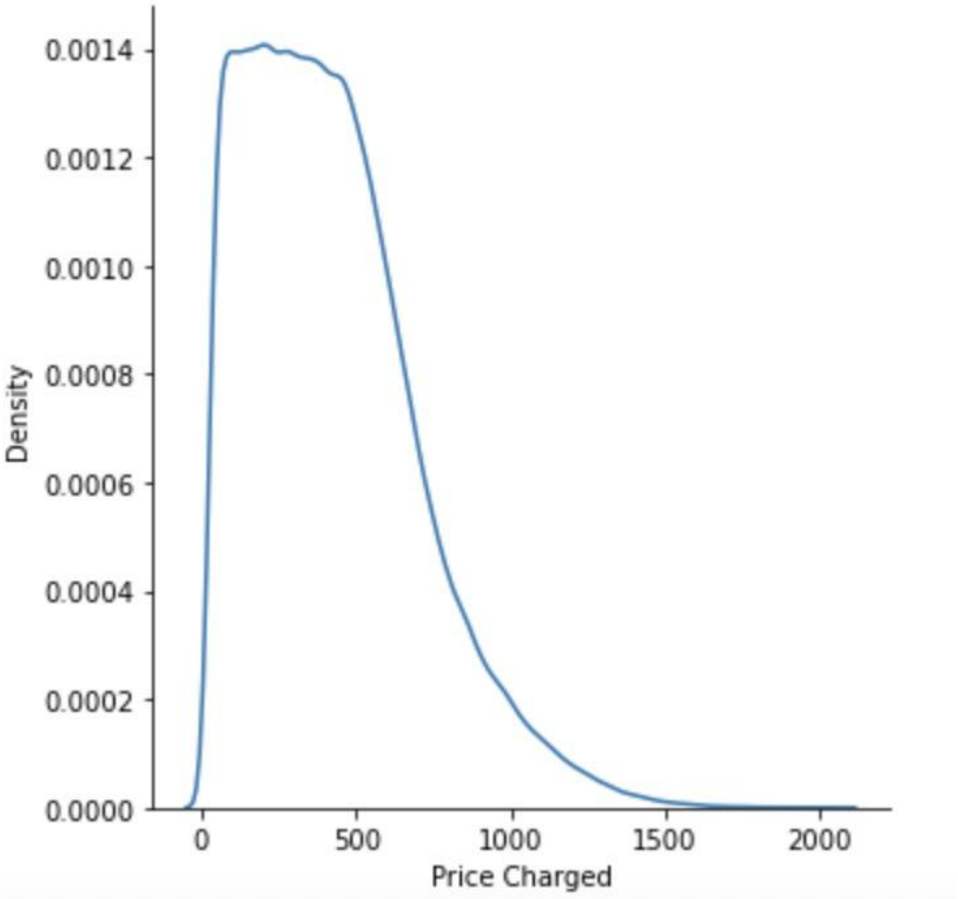
Problem Statement

EDA

Recommendations

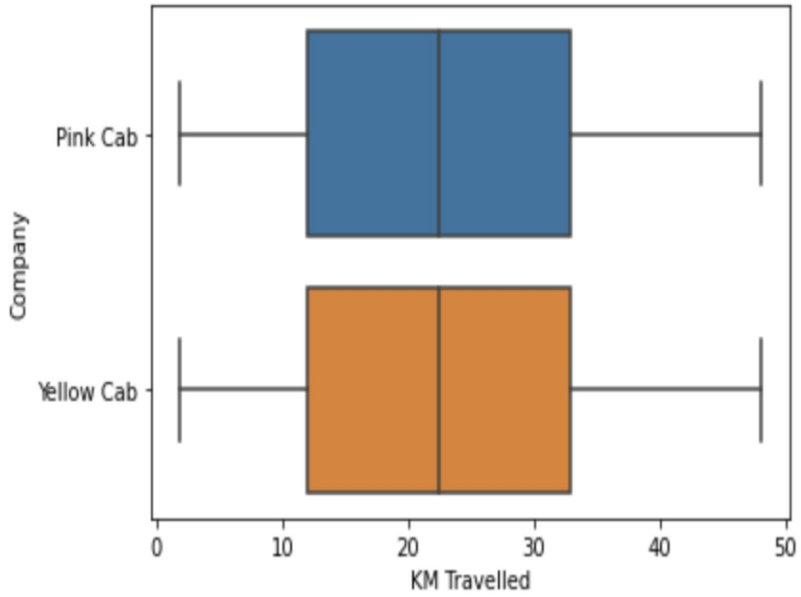
Executive Summary

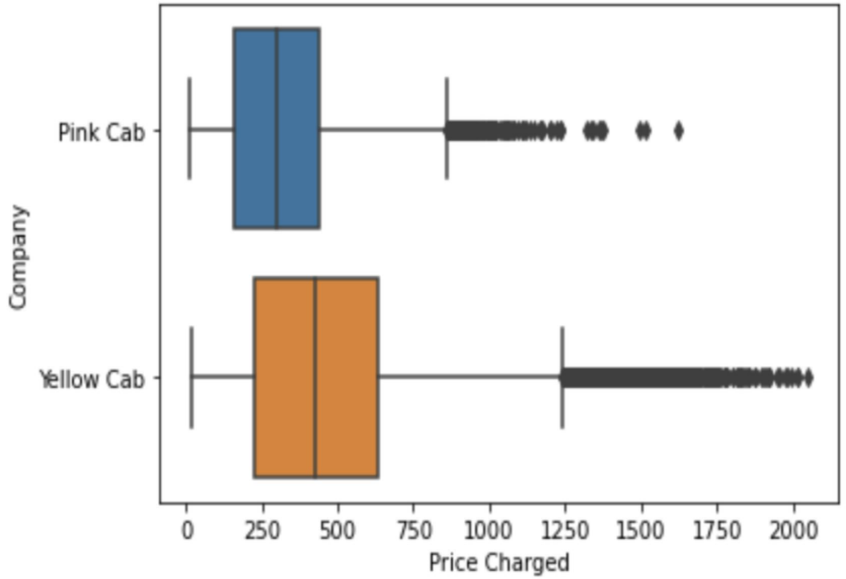
- Business Background
- Dataset details
  - 4 datasets combined to make a large dataset
  - 13 features for geographical, customer and price details for each transaction
  - skewed distribution for price charged

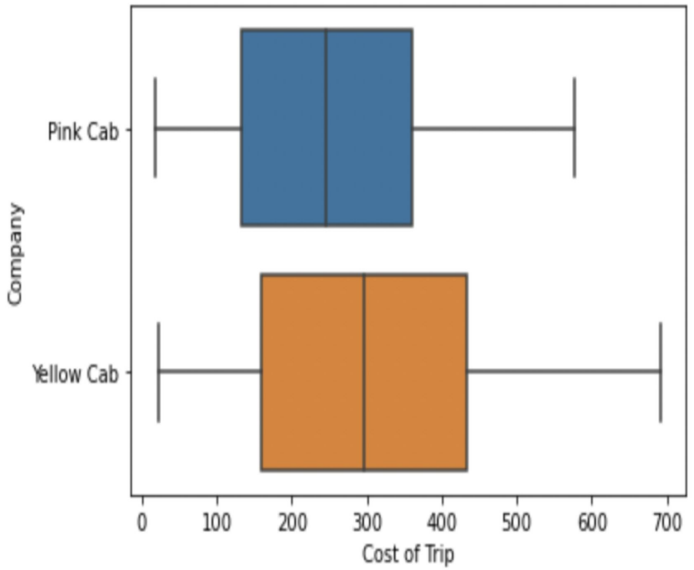


Problem Statement

- Which company to choose ?
  - Pink Cab or Yellow Cab







Exploratory Data Analysis

- Hypothesis 1 : The average cost of trip for yellow Cab is not overall significantly higher than that of Pink Cab.
- Importance : basis when we analyze revenue data differences between the two firms
- Approach: Sampling t-test

```
Ttest_indResult(statistic=4.766060605830876, pvalue=2.1574710241164073e-06)
Ttest_indResult(statistic=4.691851529811984, pvalue=3.0844183023490807e-06)
Ttest_indResult(statistic=6.962727022645901, pvalue=6.0439968139250745e-12)
Ttest_indResult(statistic=5.883867681027388, pvalue=5.466168251434964e-09)
Ttest_indResult(statistic=5.168270566673733, pvalue=2.8531405707706126e-07)
```

Exploratory Data Analysis

- Hypothesis 2 : Besides cost of trip, the income level, age, and length of trip have casual effects on the price finally charged.
- Approach: Regression analysis
- Model: Price Charged = b1*Length of trip + b2*age+b3*income + i

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	45.0724	0.697	64.693	0.000	43.707	46.438
triplength	15.1388	0.015	1002.789	0.000	15.109	15.168
Age	-0.0374	0.014	-2.678	0.007	-0.065	-0.010
income	4.742e-05	2.21e-05	2.144	0.032	4.08e-06	9.08e-05

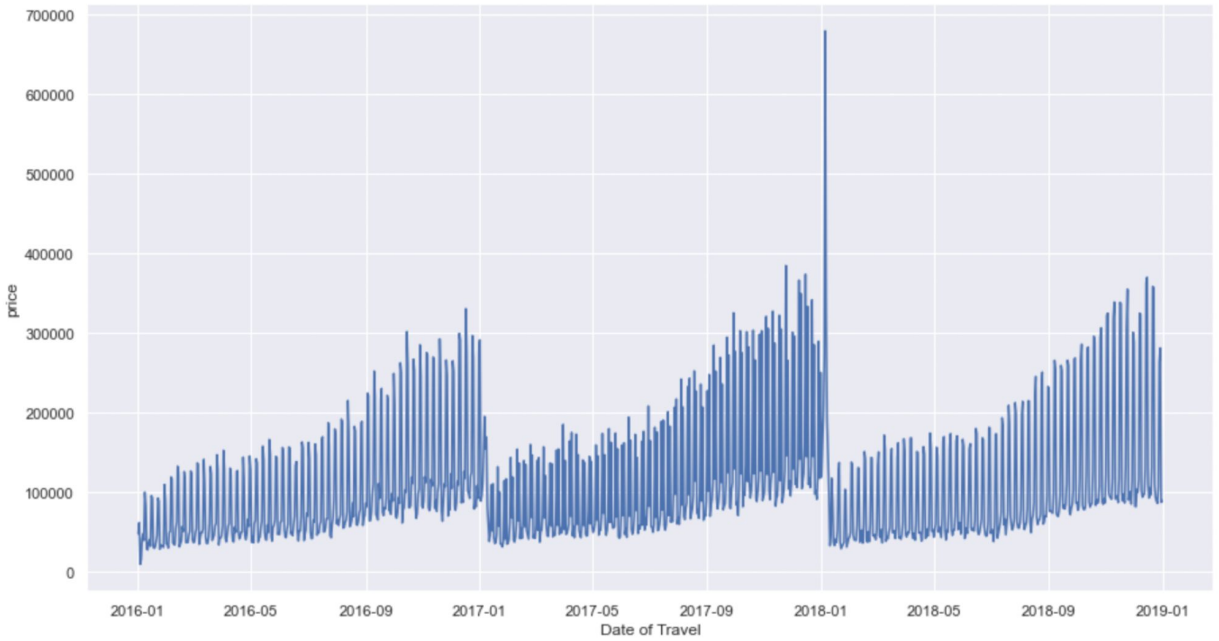
Exploratory Data Analysis

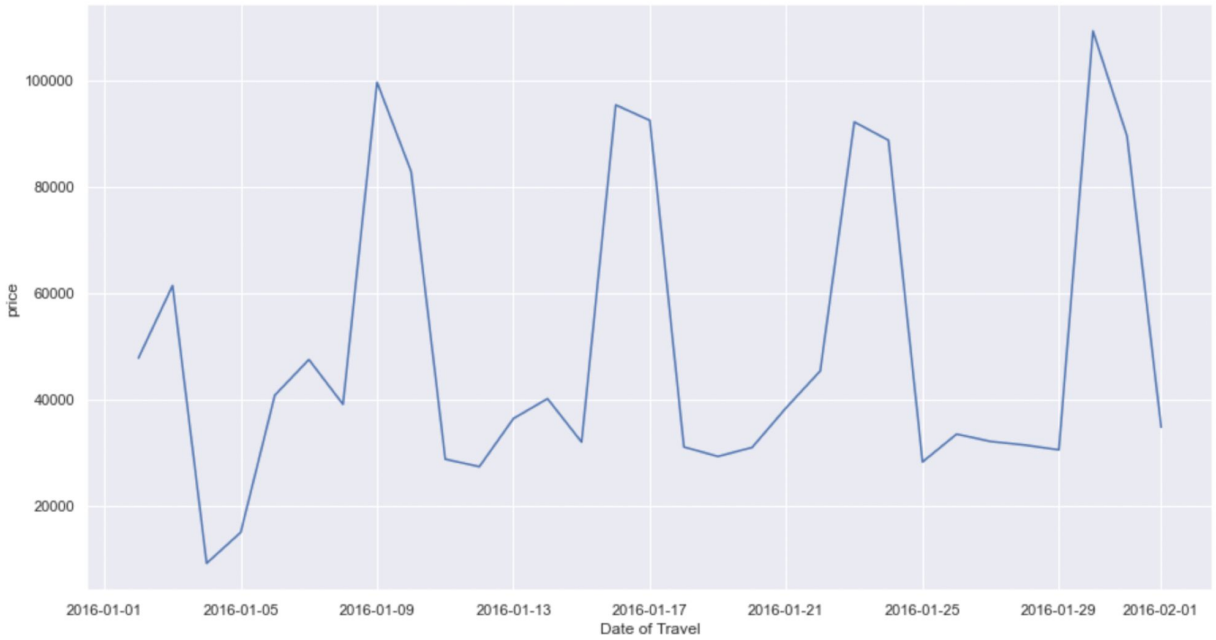
- Hypothesis 3 : Males tend to spend more on cab than females.
- Approach: Regression analysis (based on hypothesis 2)
- Model: Price Charged = b1*Length of trip + b2*age+b3*income + b4*Gender + i

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	42.8320	0.726	59.014	0.000	41.409	44.254
triplength	15.1399	0.015	1003.025	0.000	15.110	15.170
Age	-0.0380	0.014	-2.722	0.006	-0.065	-0.011
income	4.765e-05	2.21e-05	2.155	0.031	4.31e-06	9.1e-05
Gender	3.9155	0.356	10.991	0.000	3.217	4.614

Exploratory Data Analysis

- Hypothesis 4 : Price charged do have seasonal pattern.
- Approach: Visualization





Recommendations

- Consider factors other than revenue when making investment decisions
- Pay attention to seasonal pattern
- Gender do have influence on the cost of trip and thus the price charged
- Overall, Yellow Cab Company outperformed Pink Cab, and it is suggested making investment in this company

Thank you!