



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Decoding the US Cab Industry: Data-Driven Investment Intelligence

18-Sep-2023

Problem Statement

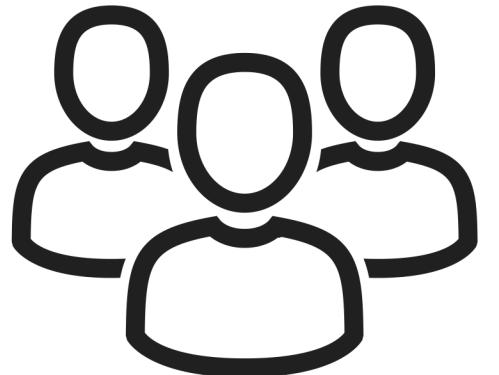


- XYZ firm aims to invest in the cab industry, in alignment with their Go-to-Market (G2M) strategies.
- To achieve this goal, they intend to understand the market based on data and, using this information, make a decision on which company to invest in.
- Yellow Cab or Pink Cab?

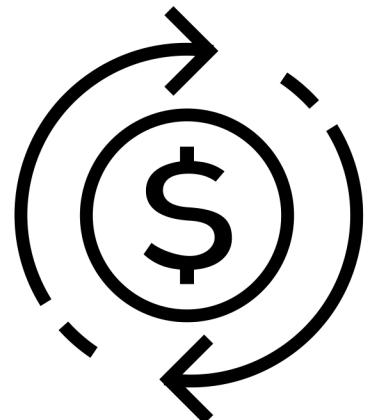
Datasets



Cab_Data



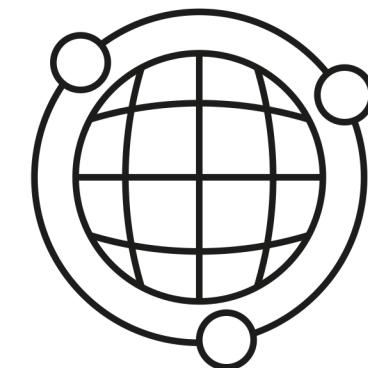
Customer_ID



Transaction_ID



City



city_coordinates

- Transaction ID
- Date of Travel
- Company
- **City**
- KM Travelled
- Price Charged
- Cost of Trip

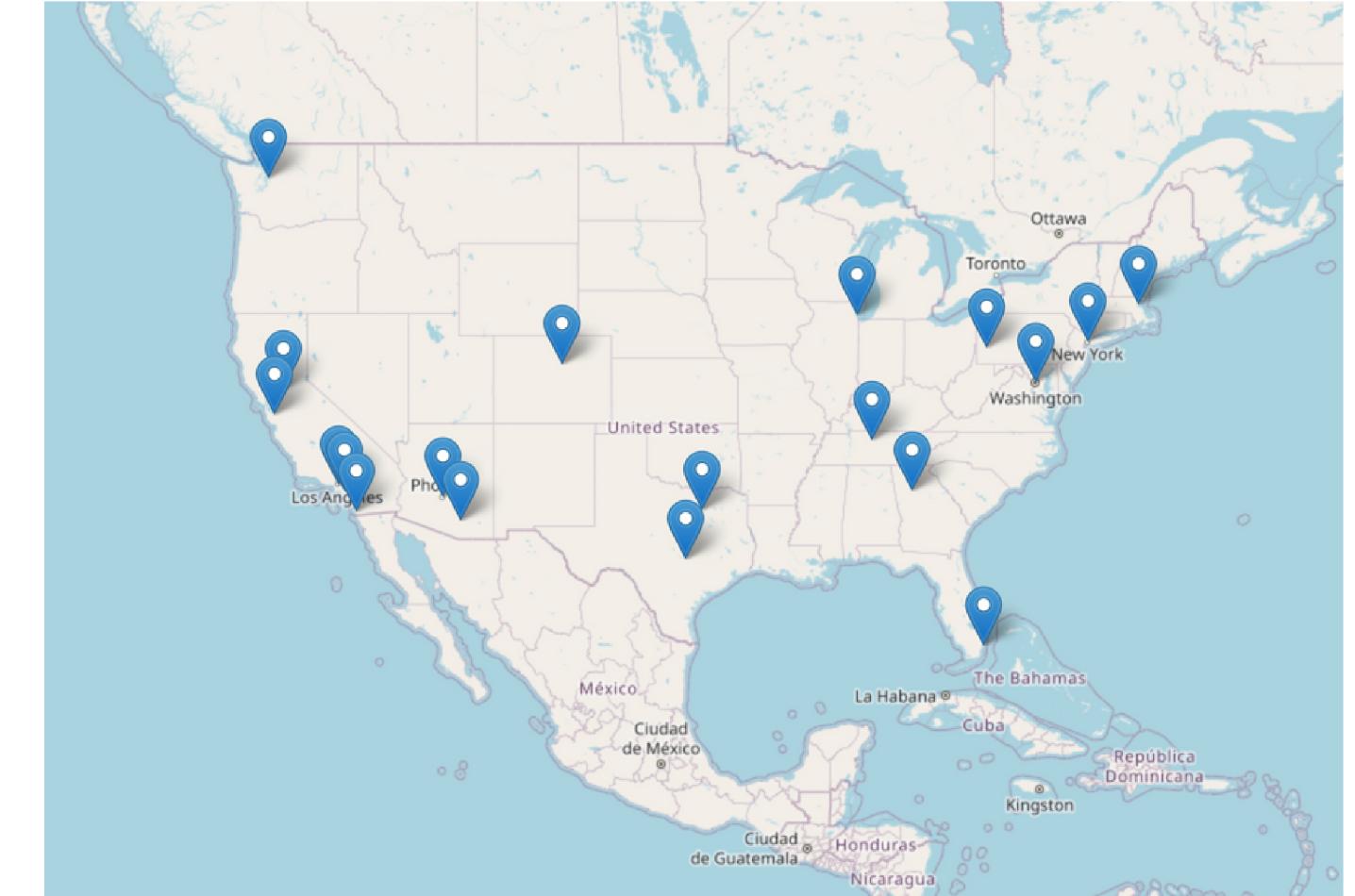
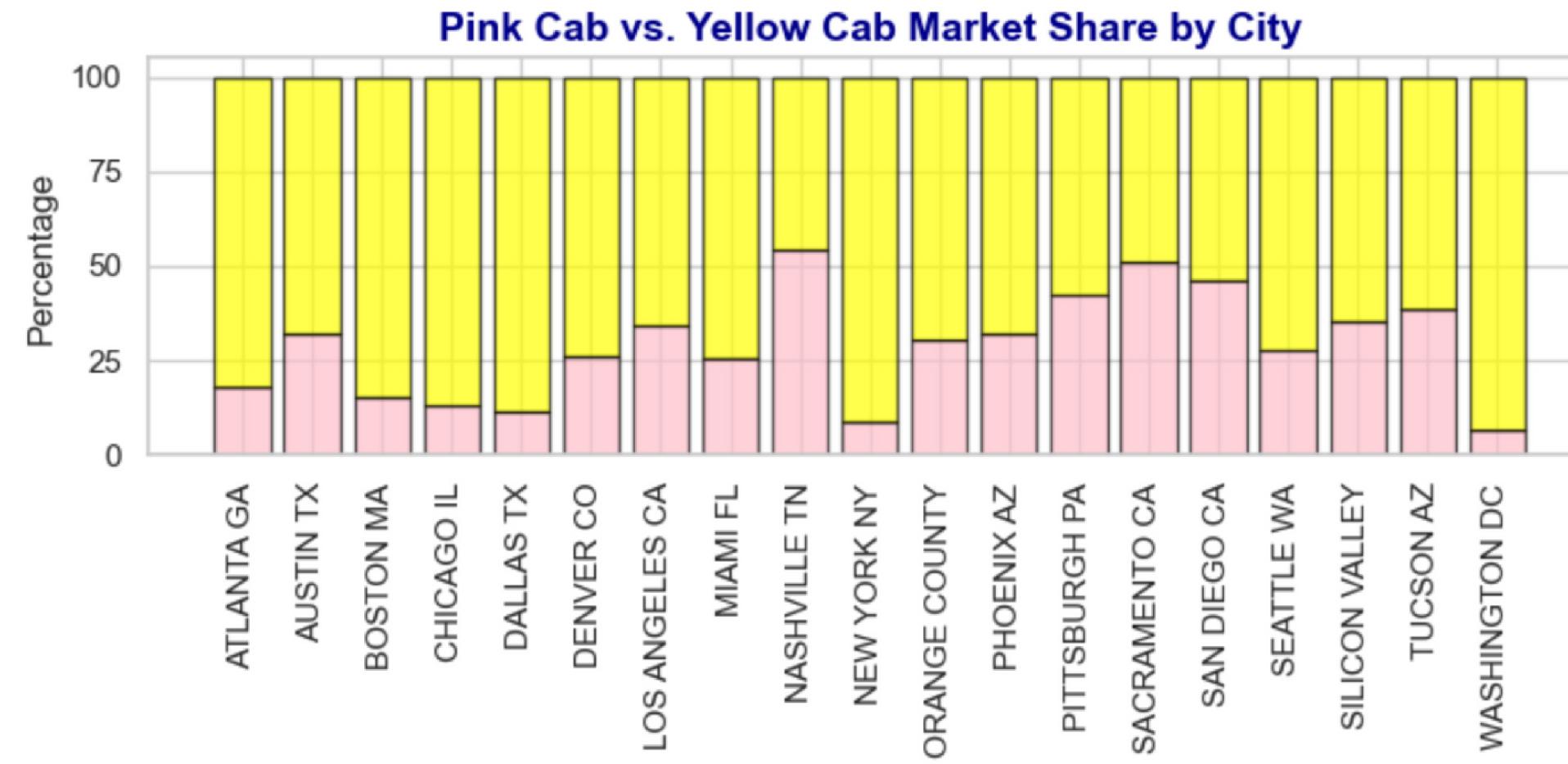
- Customer ID
- Gender
- Age
- Income
(USD/Month)

- Transaction ID
- **Customer ID**
- Payment Mode

- City
- Population
- Users

- City
- Coordinates

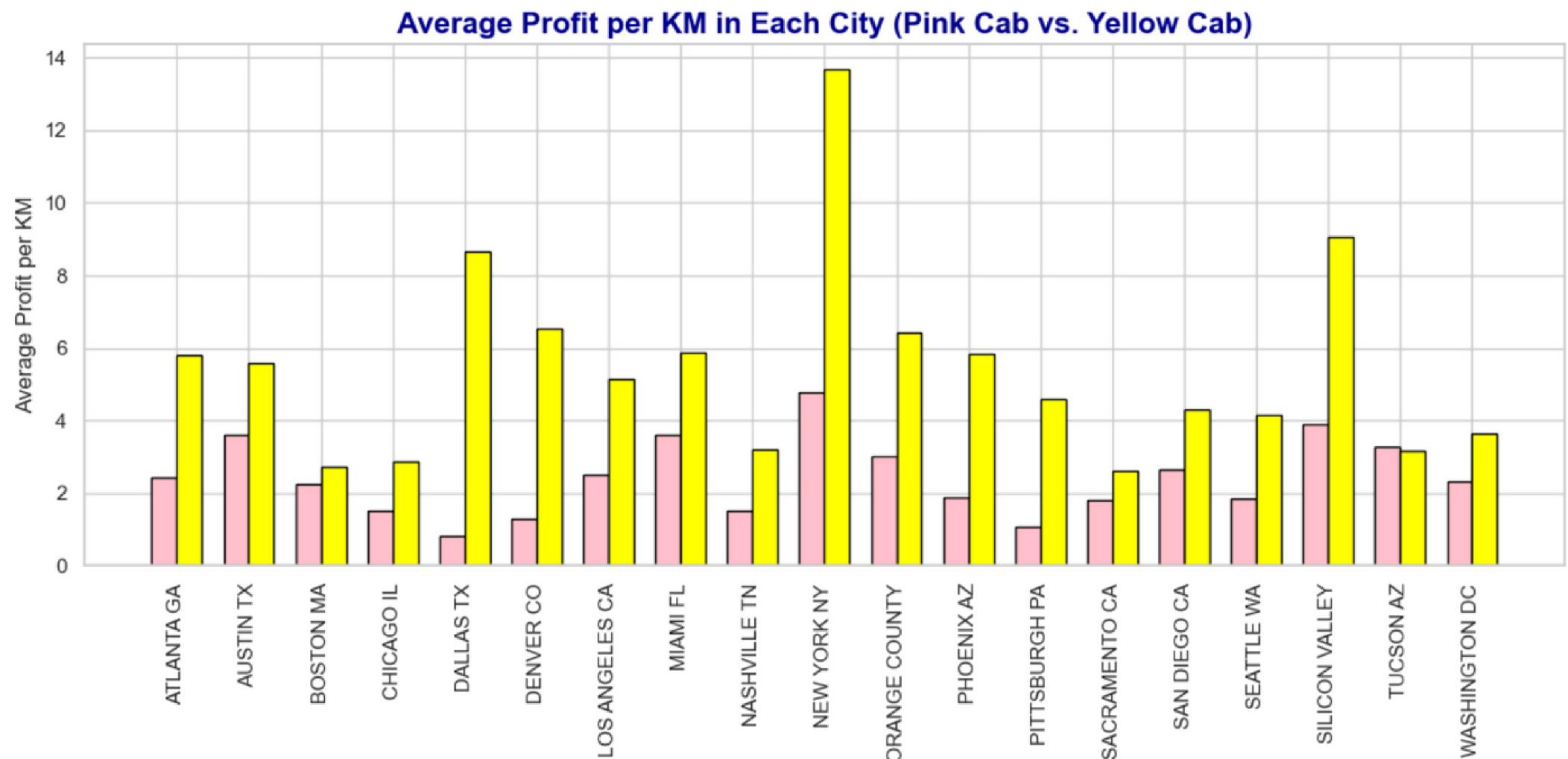
Market Shares



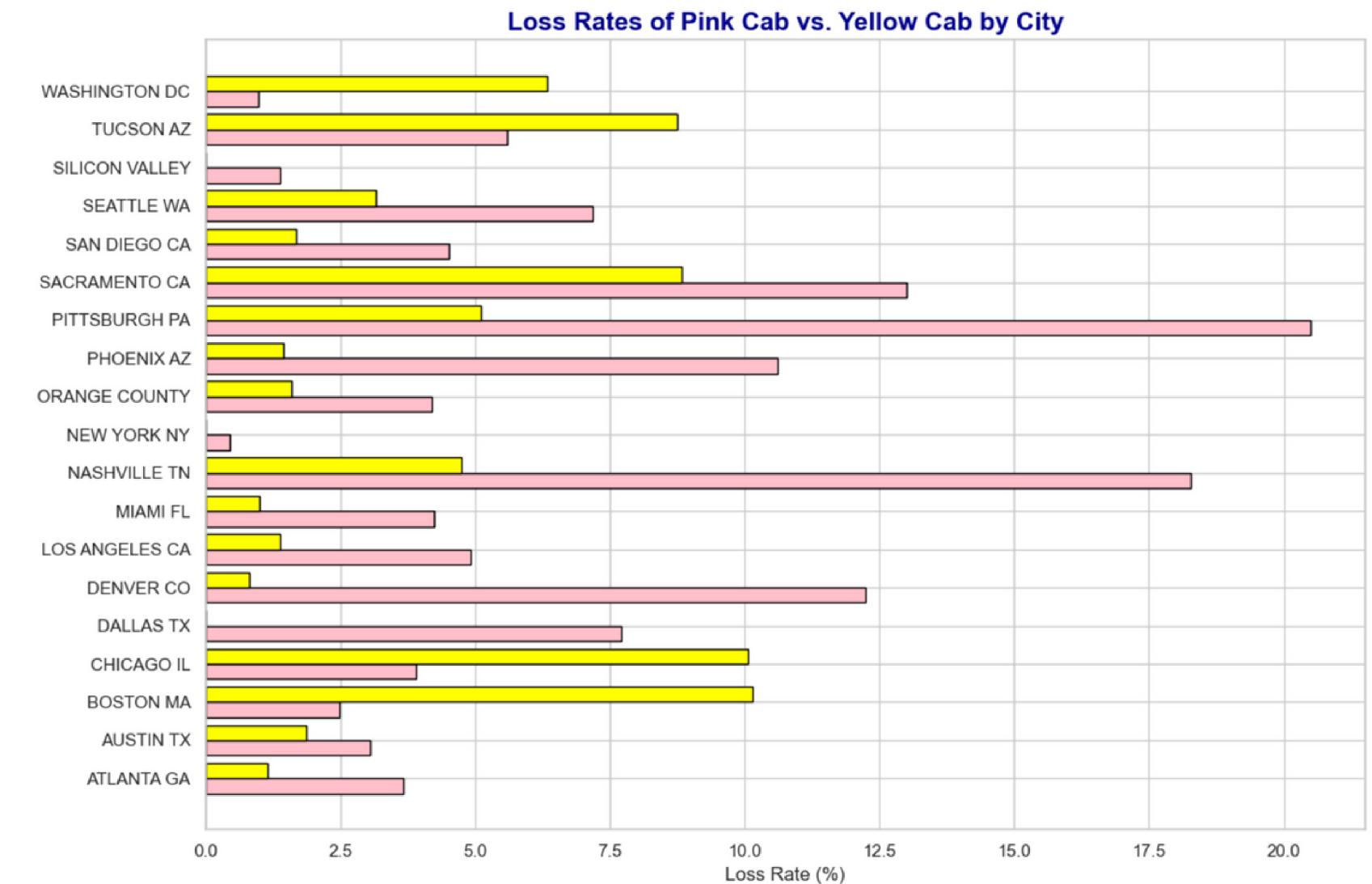
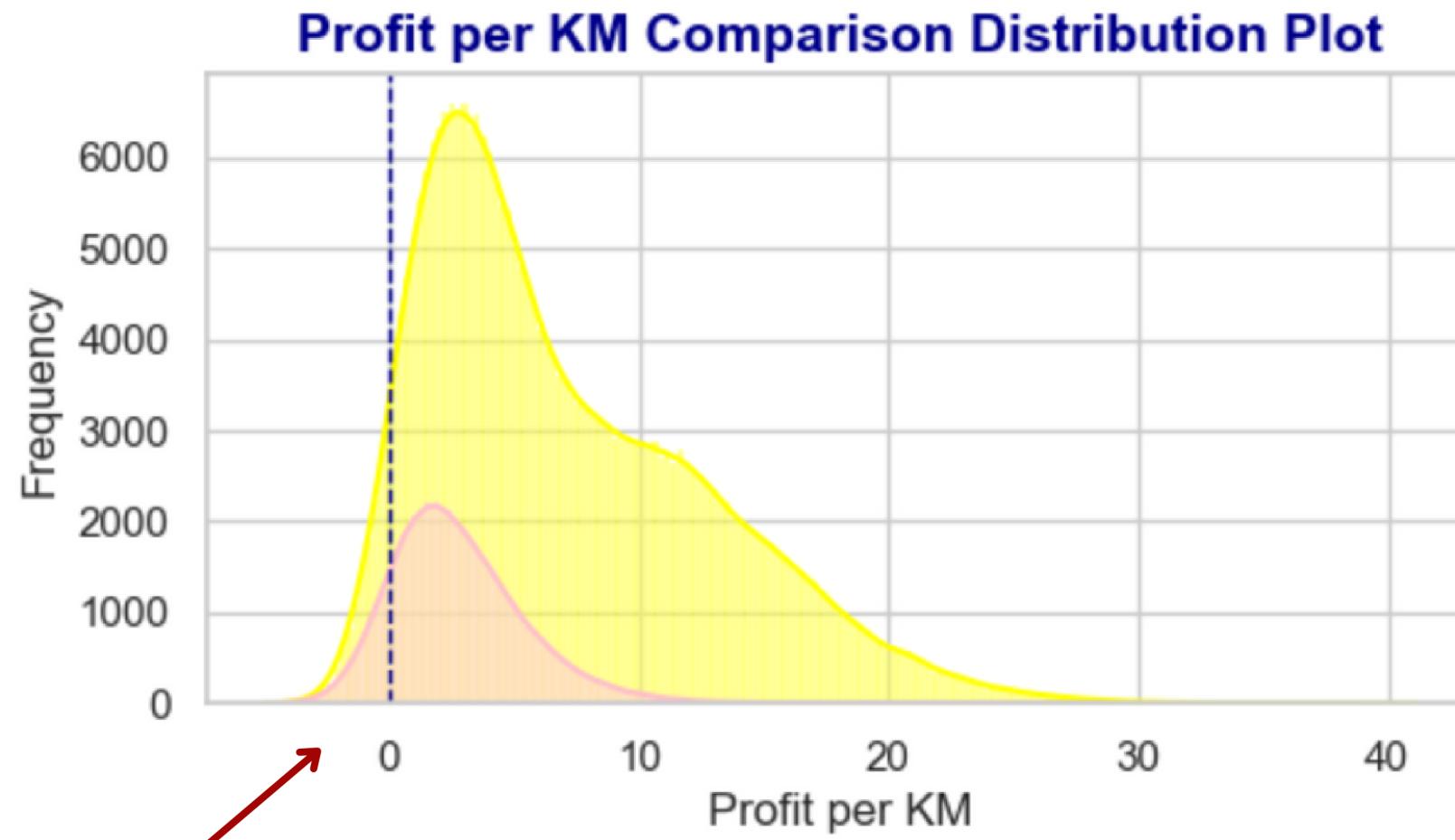
- Except for **Sacramento, CA**, and **Nashville, TN**, Yellow Cab dominates the market in all other cities.
- Yellow Cab has a significantly high market share on the eastern coast, while Pink Cab's revenues have slightly increased in the western regions.

Average Profit per KM

- Pink Cab : **\$2.77**
- Yellow Cab : **\$7.11**
- The average profit for Yellow Cab is higher than Pink Cab in all cities except **Tucson, AZ.**
- Avg. profit for Yellow Cab - >\$8:
 - New York, NY
 - Dallas, TX
 - Silicon Valley



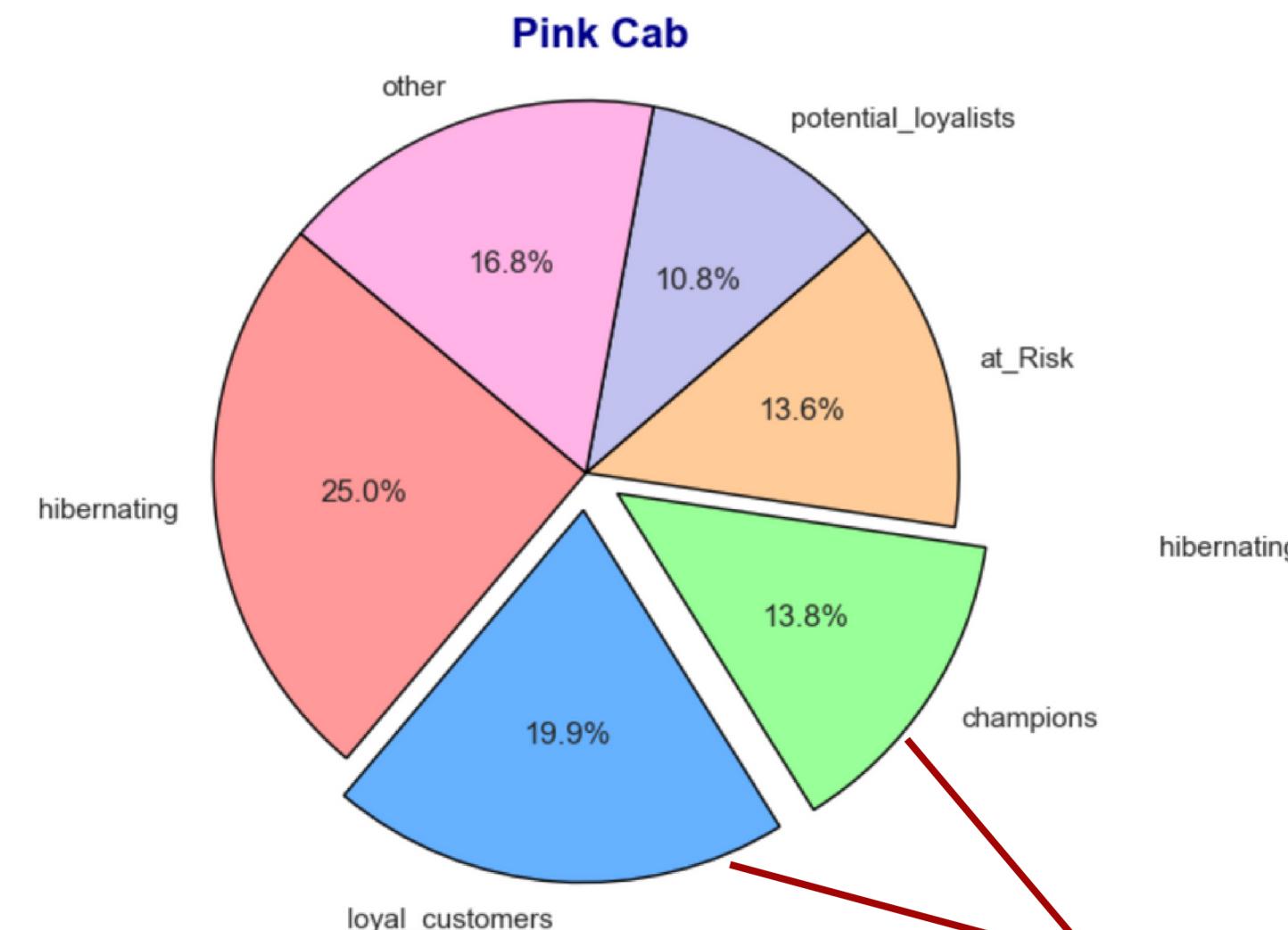
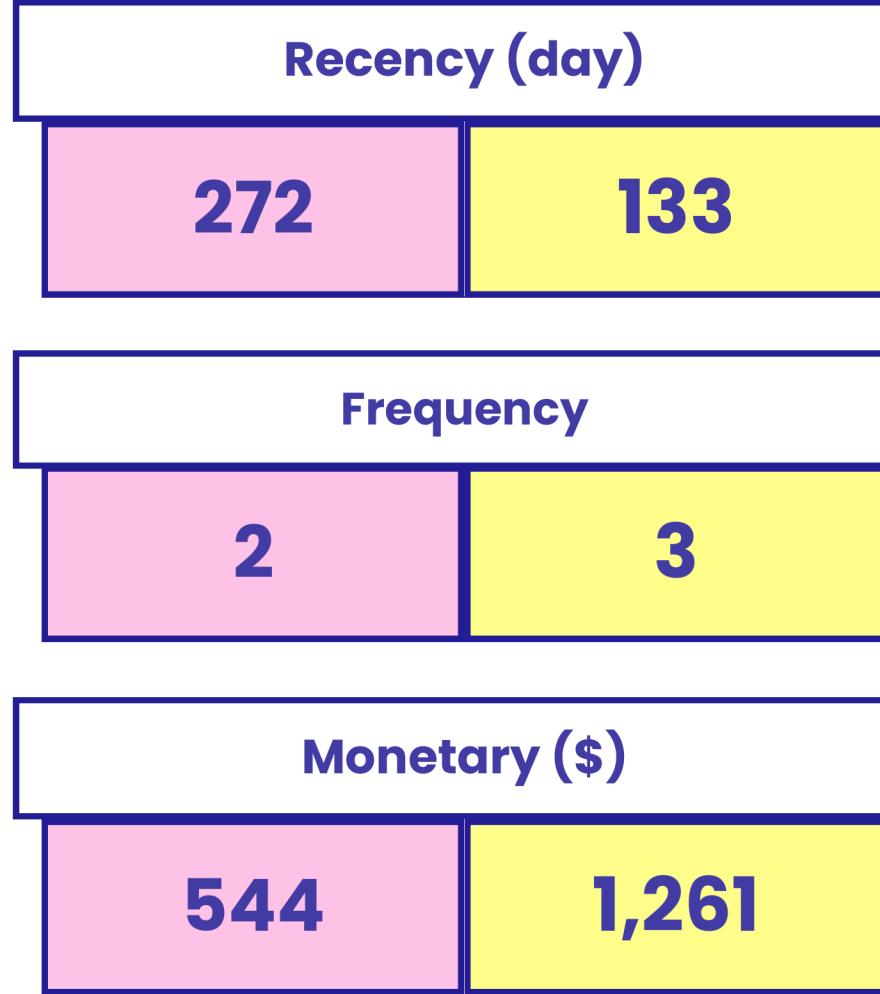
Average Profit per KM



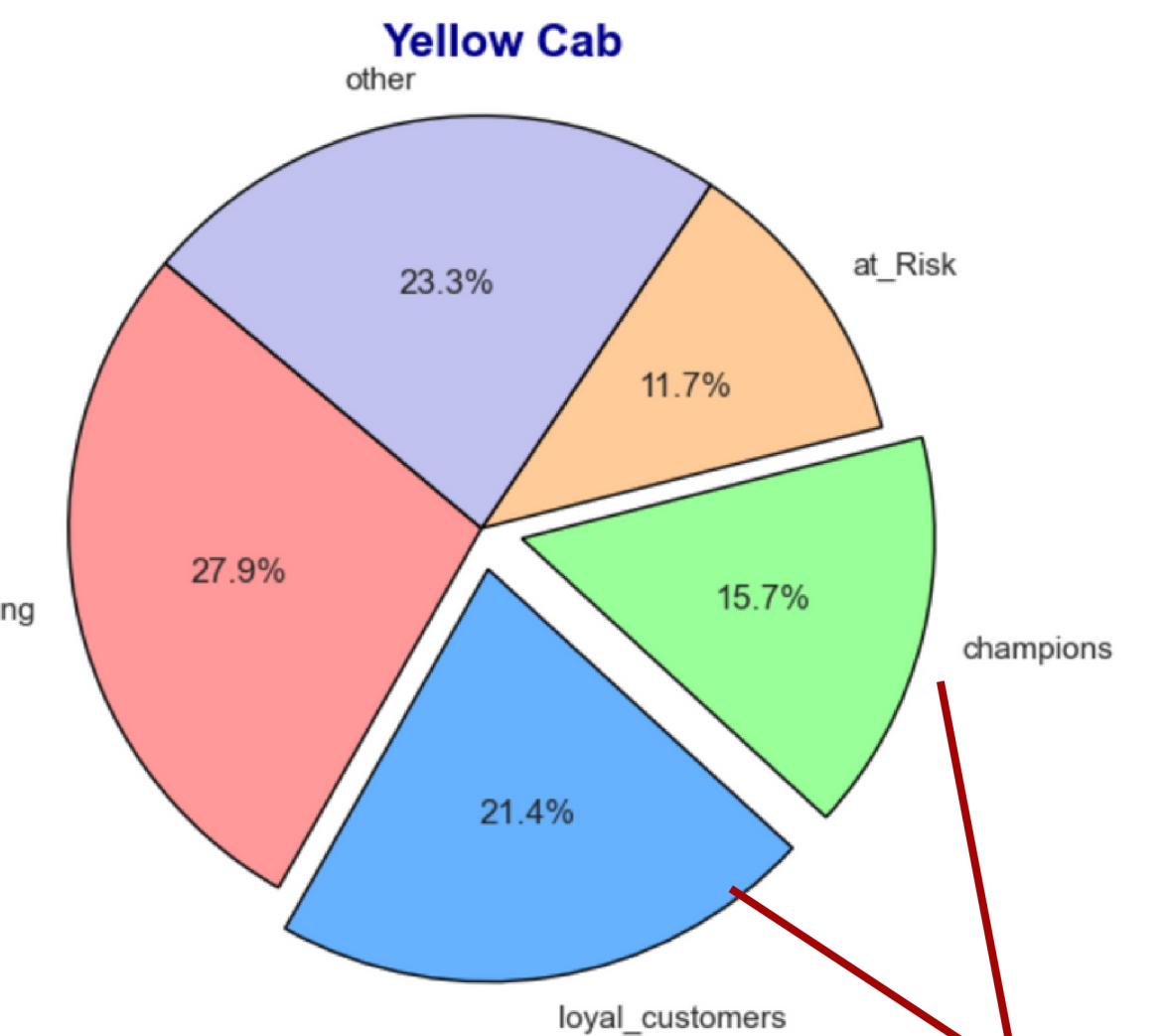
- For Yellow Cab, around **5%** of transactions result in no profit or a loss. For Pink Cab, approximately **13%**.
- **Higher loss rate for Yellow Cab:** Boston MA, Chicago IL, Tucson AZ, and Washington DC.
- Particularly noteworthy, the loss rates for **Pink Cab** exceed **18%** in: Pittsburgh PA and Nashville TN.

Number of Loyal Customers

- RFM Analysis, short for **Recency**, **Frequency**, and **Monetary** analysis, was used for customer segmentation, and customer data was segmented based on this analysis.



**10,923
customers**

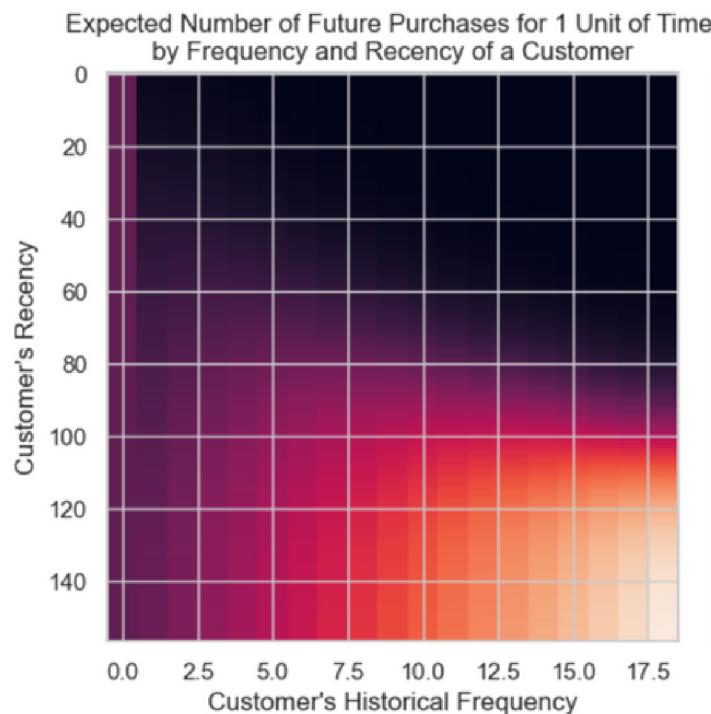


**14,812
customers**

Expected Transaction Count for 12 months

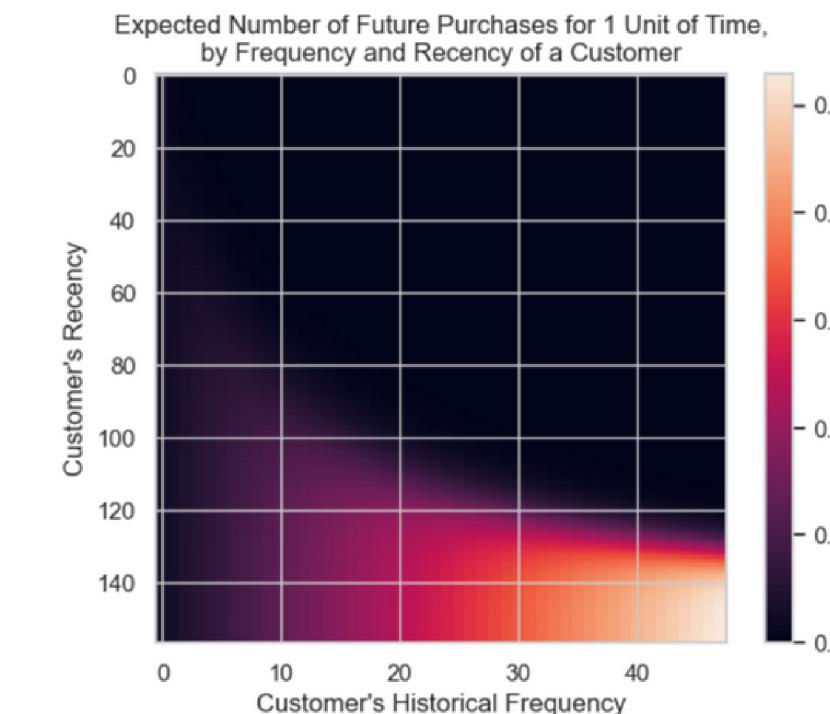
- The BG/NBD (Beta Geometric/Negative Binomial Distribution) model has been employed to predict customers' future transaction behaviors.

2016	2017	2018	2019
25,080	30,321	29,310	27,900
82,239	98,189	94,253	93,243



Pink Cab:
hottest customers

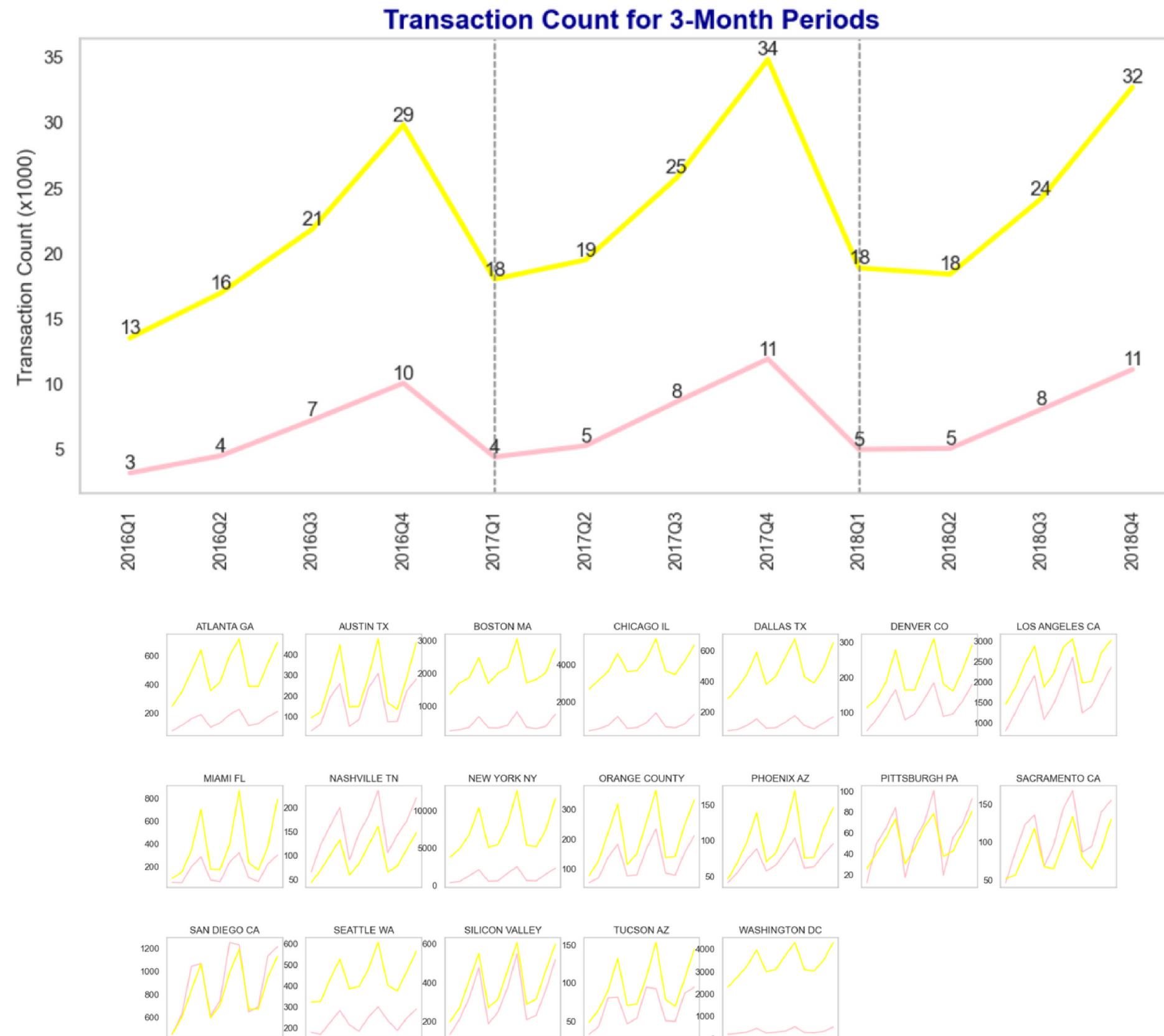
- Recency: >110 mos.
- Frequency: 17



Yellow Cab:
hottest customers

- Recency: >130 mos.
- Frequency: 40

Number of Transactions per Season



- Trend and seasonality
- Upward trend in customer travel demand from the 1st to the 4th quarter.
- Holiday season?
- Almost similar seasonality pattern is observed for all cities.

Customer Acquisition Rate

TOP					
			Customer_Count	Total_Profit	Total_Profit_per_Customer
Gender	Age_Group	City			Customer_to_User_Ratio
Male	Young Adults (18-28)	NASHVILLE TN	376	14781.264000	39.311872
Female	Young Adults (18-28)	NASHVILLE TN	329	10604.626000	32.232906
Male	Young Adults (18-28)	SACRAMENTO CA	244	9703.469000	39.768316
Female	Young Adults (18-28)	PITTSBURGH PA	126	3130.078000	24.841889
Male	Middle-aged Adults (29-38)	NASHVILLE TN	311	10004.020000	32.167267

BOTTOM

			Customer_Count	Total_Profit	Total_Profit_per_Customer	Customer_to_User_Ratio
Gender	Age_Group	City				
Male	Adults (39-65)	WASHINGTON DC	725	37507.843000	51.734956	0.570862
	Middle-aged Adults (29-38)	WASHINGTON DC	691	38363.618000	55.518984	0.544090
	Young Adults (18-28)	WASHINGTON DC	532	27129.361000	50.995039	0.418894
Female	Middle-aged Adults (29-38)	WASHINGTON DC	519	28066.475000	54.077987	0.408658
	Adults (39-65)	WASHINGTON DC	493	25450.245000	51.623215	0.388186

- **High Customer Ratio in Nashville**
- **Challenges in Prominent Cities:** Nashville, Pittsburgh, and Sacramento. Pink Cab operates at a loss of over 13% in these areas.
- **Population vs. Profit:** Lower populations, lower profit per customer.
- **Lowest Customer Ratio in Washington DC:** all gender and age groups

TOP					
			Customer_Count	Total_Profit	Total_Profit_per_Customer
Gender	Age_Group	City			Customer_to_User_Ratio
Male	Young Adults (18-28)	WASHINGTON DC	8281	684396.708400	82.646626
	Young Adults (18-28)	NEW YORK NY	18920	5818055.202000	307.508203
	Adults (39-65)	CHICAGO IL	10190	668369.912400	65.590767
	Young Adults (18-28)	CHICAGO IL	9866	652848.378000	66.171536
		BOSTON MA	4776	289773.861200	60.672919

BOTTOM

			Customer_Count	Total_Profit	Total_Profit_per_Customer	Customer_to_User_Ratio
Gender	Age_Group	City				
Female	Middle-aged Adults (29-38)	NASHVILLE TN	192	13148.906400	68.483887	2.071197
Male	Middle-aged Adults (29-38)	SACRAMENTO CA	142	7797.379600	54.911124	2.015900
Female	Middle-aged Adults (29-38)	SAN DIEGO CA	1371	130852.755600	95.443294	1.958711
Male	Adults (39-65)	NASHVILLE TN	172	14078.935200	81.854274	1.855448
Female	Adults (39-65)	NASHVILLE TN	165	11221.322800	68.008017	1.779935

- **Male-Dominated Top 5 Demographic Groups**
- **High Customer Ratios in Young Adult Groups:** New York and Washington DC
- **Success in Profitable Cities**
- **Lowest Customer Ratios in Specific Cities:** Sacramento, San Diego and Nashville. Relatively low-population cities.

Recommendations

- **Market Share:** Yellow cab appears to have the highest market share in 16 out of 18 cities, indicating a large customer base and a well-known brand.
- **Profitability (Profit per km):** Yellow cab has a higher profit margin per kilometer, which suggests it may be more profitable.
- **Loss Rate:** Pink cab has a higher loss rate per transaction (13%), which could negatively impact overall business profitability. Yellow cab, on the other hand, has a lower loss rate (5%).
- **Loyal Customers:** Yellow cab has a larger number of loyal customers, indicating potentially more stable future revenue.
- **Gender Preference:** Yellow cab is preferred by male groups, which can be an important factor in shaping marketing strategies.
- **Population:** Pink cab focuses more on low-population areas where profit margins are lower. This may limit future growth potential.

In summary, based on the provided data, it is recommended to prefer **Yellow Cab**.

Thank you.