Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis

XYZ Corporation – Data-Driven Recommendation

05/28/2025

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier

Your Deep Learning Partner

Executive Summary

XYZ is thinking about investing in one of two cab companies. To help them choose the better one, we looked at cab ride data from 2016 to 2018. We cleaned and combined four different data files, then studied patterns in customer behavior, spending, and company profits. Our analysis shows clear differences between the two companies that can help XYZ make a smart choice.

Problem Statement

The goal of this project is to help XYZ decide which cab company is better for investment. The company wants to know where riders are coming from, how much they spend, and which company makes more money. We used real data from customer transactions, cities, and prices to find patterns. We also looked at who the customers are and how they behave.

Approach

We started by opening and reviewing the four files that were given: one with cab rides, one with customer information, one with payments, and one with city data. After making sure the data was clean and consistent, we merged all the files together into one full table. Once the data was ready, we explored it using graphs to understand things like age, income, company profits, and city trends.

EDA

We found that most cab users were between 25 and 40 years old, and there were slightly more male users. People with higher monthly income also tended to take more rides. We saw that Pink Cab had higher profit margins compared to Yellow Cab, even though Yellow Cab had more total trips. Cities like New York and Chicago brought in the most money, and ride numbers went up during spring and the holidays.

- Age distribution histogram
- Gender vs. income boxplot
- Revenue/profit bar chart by company
- Top 10 cities by revenue
- Monthly rides line chart (seasonality)

EDA Analysis

From the data, we can see that Pink Cab makes more money per trip, while Yellow Cab focuses more on volume. Larger cities have both more customers and higher revenue. We also found that riders with higher incomes spent more, and people preferred riding more during certain months like March and December. These patterns help us understand how each company runs and what kind of customers they attract.

Recommendations

Based on everything we saw, Pink Cab looks like the better choice for investment because it earns more profit per ride. XYZ should focus on cities like New York, LA, and Chicago, where both revenue and cab usage are high. For future steps, we recommend using machine learning to group customers by behavior and predict who will spend the most. This can help the company make better decisions going forward.

Thank You