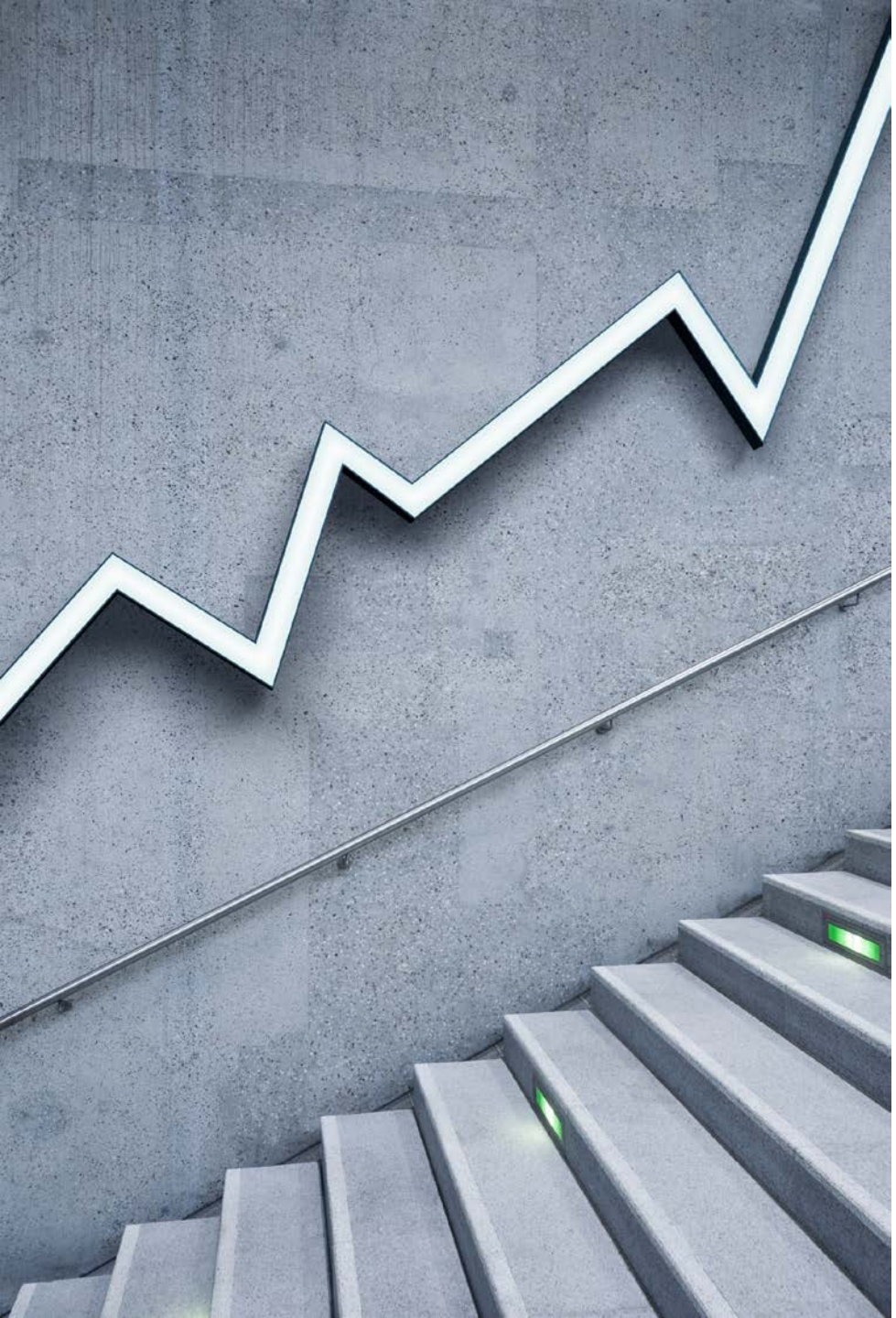

G2M Case Study

BY,

BALAMURUGAN PURUSHOTHAMAN

Introduction

XYZ is a private firm in the US

They want to invest in the cab industry

Needs to choose between two cab companies to invest in: (Pink Cab and Yellow Cab)

Objective: Provide supporting insights to make a cab company recommendation for XYZ to invest.

The default datasets provided for this task (Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, City.csv) include data for cab transactions, customers, and residing cities for both Pink Cab and Yellow Cab from the start of 2016 to the end of 2018

Inferences

In Cab_Data.csv, for each transaction, the profit for that transaction is (Price Charged- Cost of Trip)

► Profit outliers are removed because they skew the overall dataset in a way that matters more than a variable like cost or charge

► Removing profit outliers has clearly more of a dampening effect on the Yellow Cab data than the Pink Cab data since Yellow Cab was found to have zero low profit outliers and a high number of high profit outliers which dominate Pink Cab's high profit outliers

► This fact further supports evidence in the following slides in favor of investing in Yellow Cab

► In Cab_Data.csv, Date of Travel is number of days starting from January 1, 1900

Inferences

► In the City.csv, the number of users is the total number of cab users in the city including from Pink Cab, Yellow Cab, and other cab companies

► A ride can be servicing multiple people/transactions at once. The cost is adjusted to only count the cost once for multiple people being driven the same route