

Blackstone Data Science PMP Case Study

Background:

We are working with a portfolio company within the retail space. The company operates a distributed footprint of stores and while the company has successfully scaled its operation, the CEO believes that there are still inefficiencies within the system. As the business looks to expand, it is important to understand key drivers of success and potential areas to improve upon. You are the data scientist that will help to inform this understanding and in turn, drive the company's go-forward strategy.

To help facilitate this analysis, you have been provided with historical operating data for all locations over the last year – please use only this dataset in your analysis. You are free to use any programming language to analyze this data. Please save all work and provide all code along with your response.

Part 1 – Summary statistics and plots:

- 1) Looking across all locations, calculate the mean and median annual profit margin of a typical store. Does the profit margin across locations appear normally distributed? Please visualize.
- 2) Please plot the monthly and quarterly aggregate revenue across all stores.

Part 2 – Cleaning the data:

You will have noticed that the company operates both owned and leased locations. This creates an apples-to-oranges comparison, as only the leased locations are burdened with rental cost. To prep your data for analysis, please impute an appropriate rental cost for each owned location and recalculate the rent-adjusted annual profit margin.

Part 3 – Quantify drivers of success:

With the data now cleaned, please identify factors that have led to historical outperformance across locations. The CEO is principally focused on maximizing profit margin at each location. There is no single correct answer for this section. We are more interested in your problem-solving process than any singular answer.

In response to the CEO's question, please write a concise email describing your findings and put together two slides that you can use to walk them through the analysis.

Dataset:

Columns: Location Number, Month, State, Gross Revenue, Fixed Cost, Variable Cost, Rental Cost, Number of Products, Owned / Rent

Rows: 333 locations x 12 months

Please note that rent *is not* included in fixed or variable costs.