

Project 9 : Invoices dataset

E-Commerce Analysis

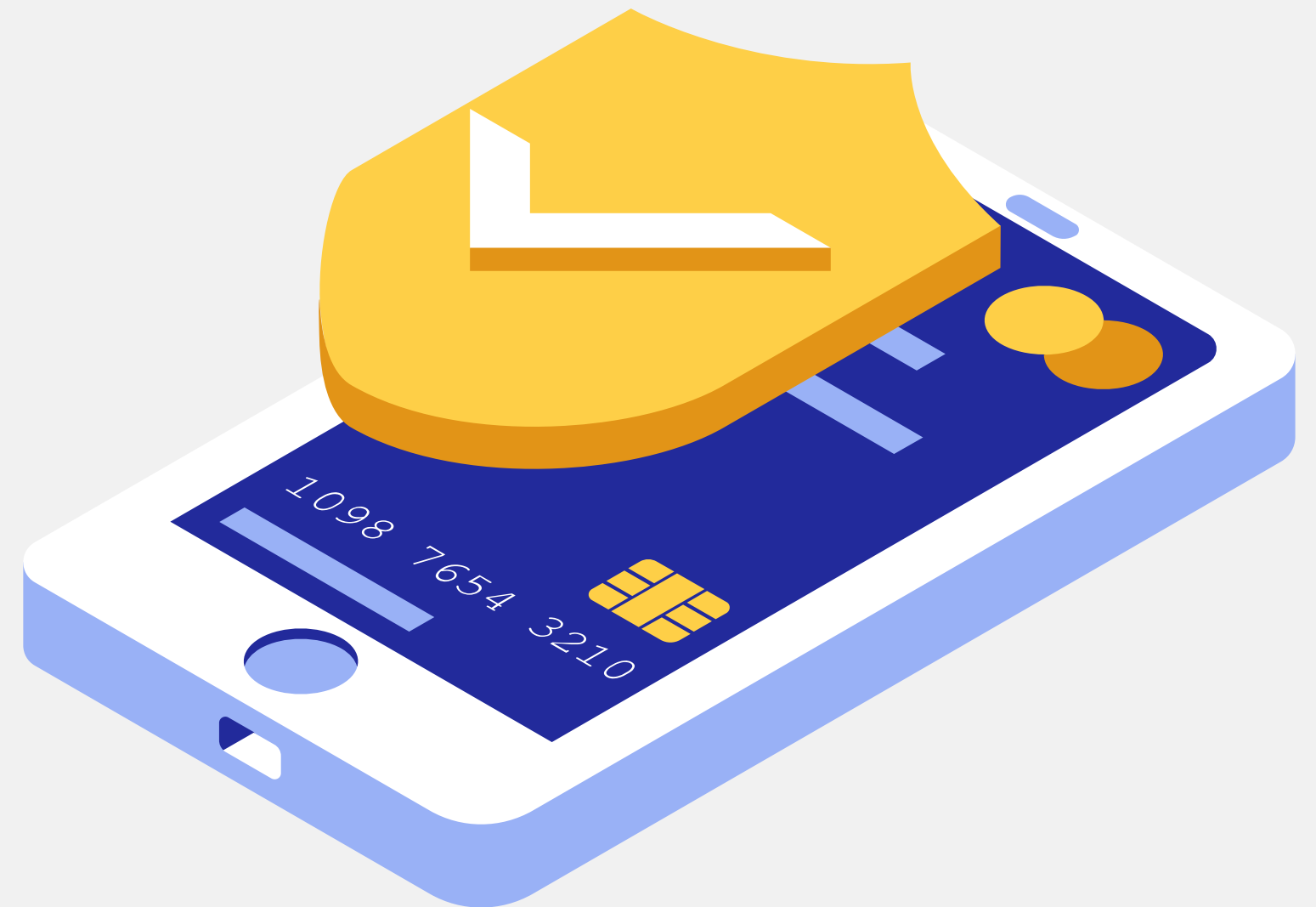
Chadi ALKERDI, Bilal GOUGIS, Mantra OUTTANDY,
Minji PARK, Jeffrey SAINT ANDRE

Project Objective & Scope

Goal : Extract 5 actionable business indicators to optimize sales

Steps

data exploration → indicator development
(using various technics) → dash visualization



Data exploration & limitations

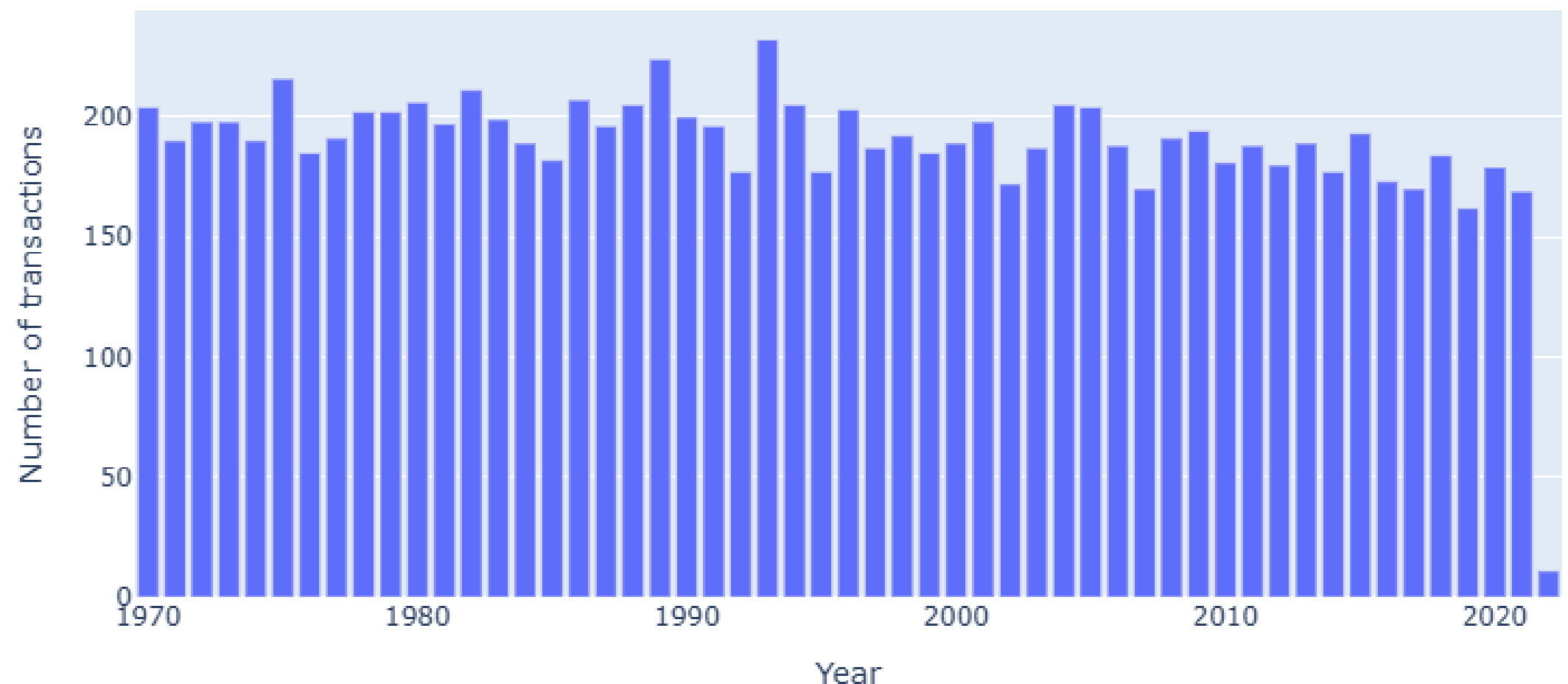
Our dataset is "invoices.csv" which contains 10 002 rows with 0 missing values which proves that our dataset is unrealistic.

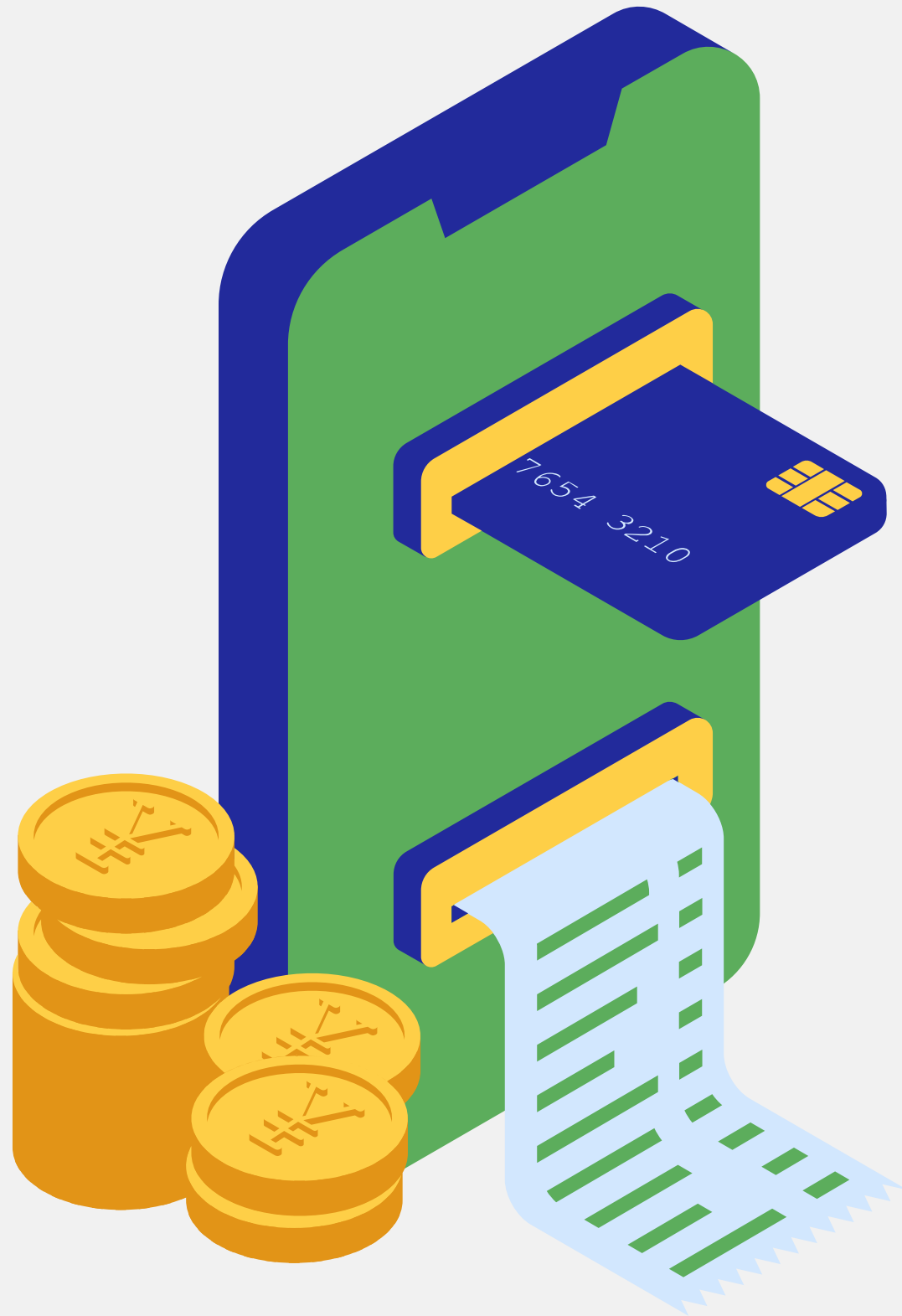
The dataset is synthetic with highly homogeneous distributions, stable yearly trend over 50 years.

Missing values per column:

first_name	0
last_name	0
email	0
product_id	0
qty	0
amount	0
invoice_date	0

Number of transactions per year





Methodology and limits

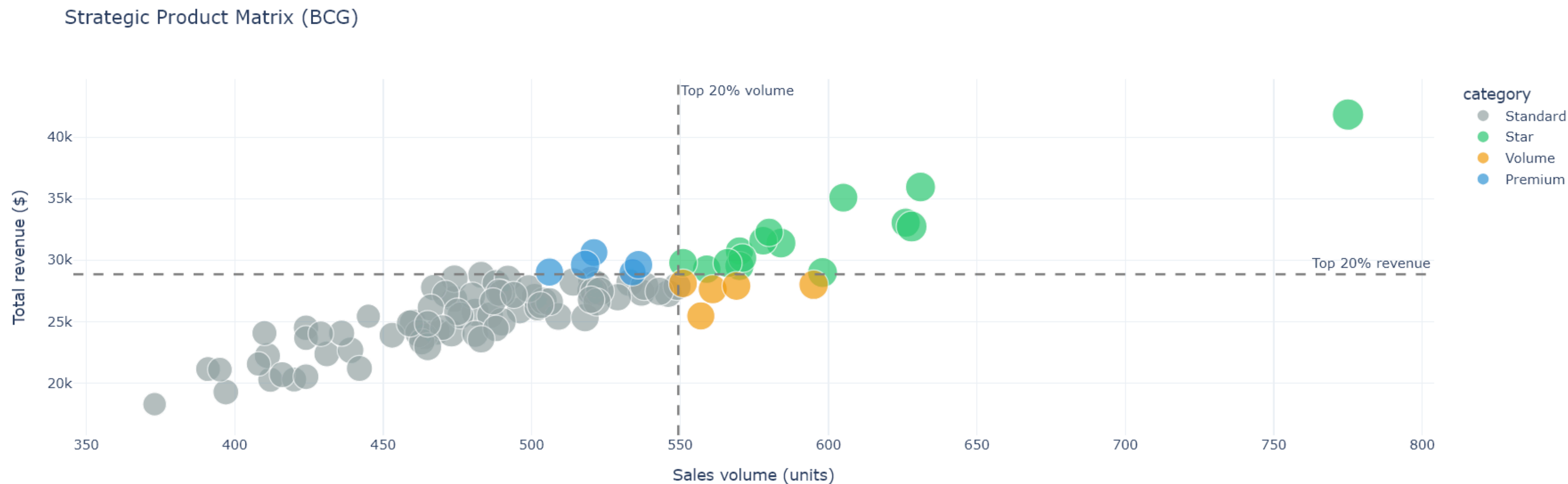
- dataset is synthetic → limits realism
- Techniques chosen to simulate real business analysis
- With real data we would add : forecasting, customer churn, clustering

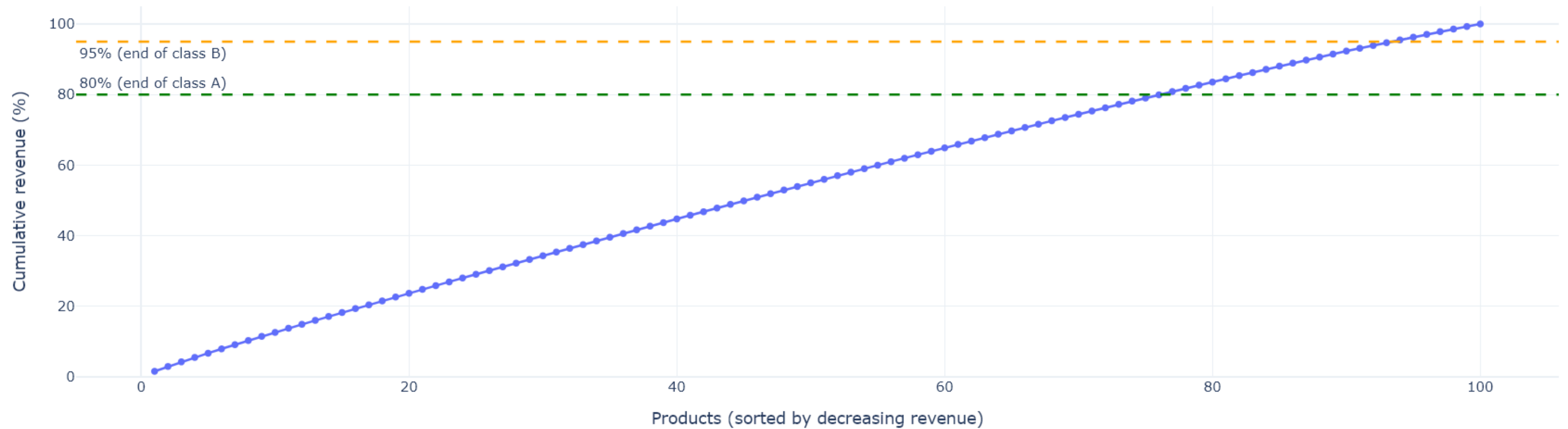
BCG Product Matrix

Products are uniformly distributed across the matrix.

Categories: Star (High/High), Premium (High Revenue/Low Volume), Volume (Low Revenue/High Volume).

Business actions: Prioritize investment, stock, and marketing efforts on Star and Premium products.





ABC Analysis (Pareto Principle)

Classify products based on cumulative revenue contribution (A: up to 80%, B: 80-95%, C: 95%+).

The curve is very linear. Achieving 80% revenue requires approx. 80% of all products.

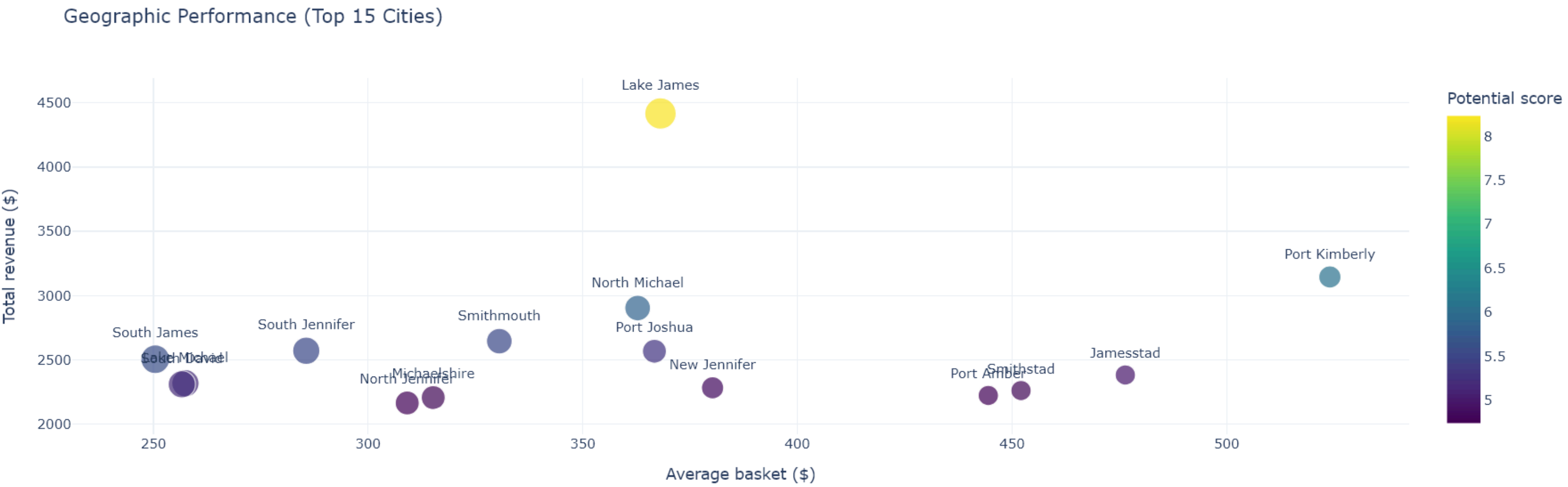
Limitation: This uniform contribution confirms the weak Pareto effect in this synthetic dataset.

Business actions : Focus management and logistics efforts on Class A products (80% revenue).

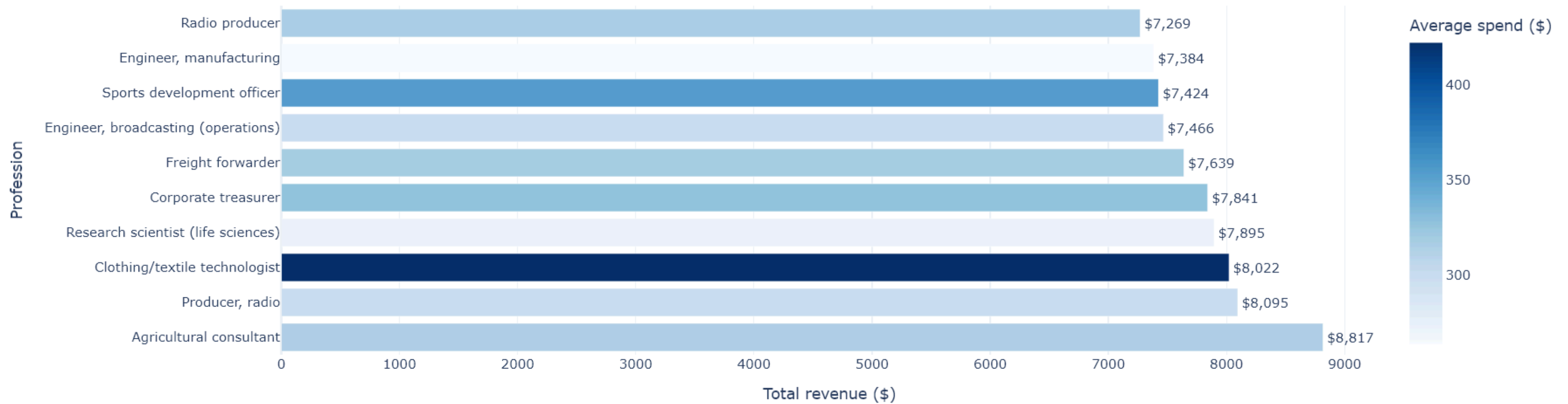
Geographic performance

Aggregate data by city to calculate Total Revenue (Y), Average Basket (X), and Number of Customers (Size). Identify key markets like 'Lake James' (highest revenue/score) and 'Port Kimberly' (high average basket).

Business actions: Direct investment and targeted marketing campaigns to high-potential cities.



Top 10 Professions by Revenue



Customer Profiles by Profession

Use a Grouping Query (`groupby('job')`) to calculate Total Revenue (X) and Average Spend (Color). Identify professions contributing the most revenue (e.g., 'Agricultural consultant').

Business actions: Inform B2B targeting, personalize offers, and establish corporate partnerships with key sectors.

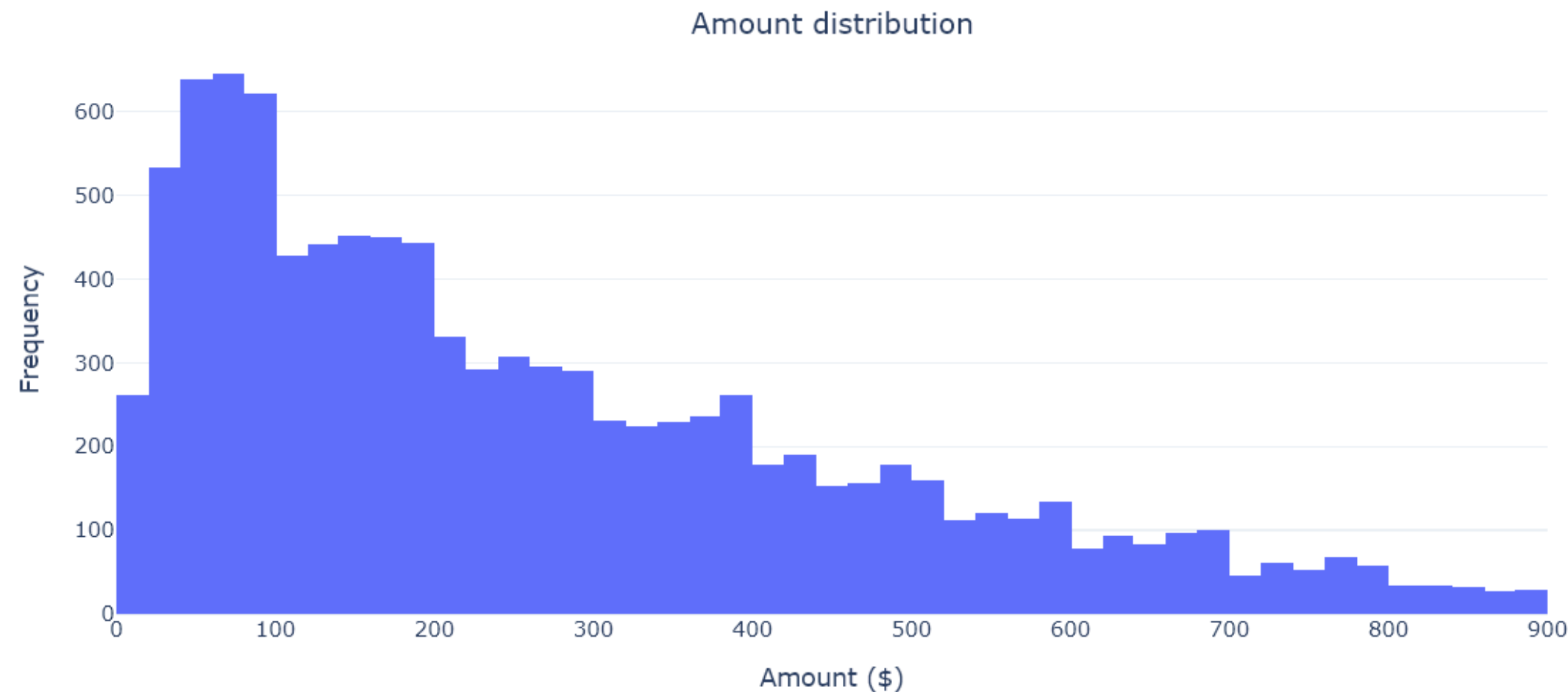
Amount Distribution (Normalization)

Histogram and Box Plot show distribution and quartiles.

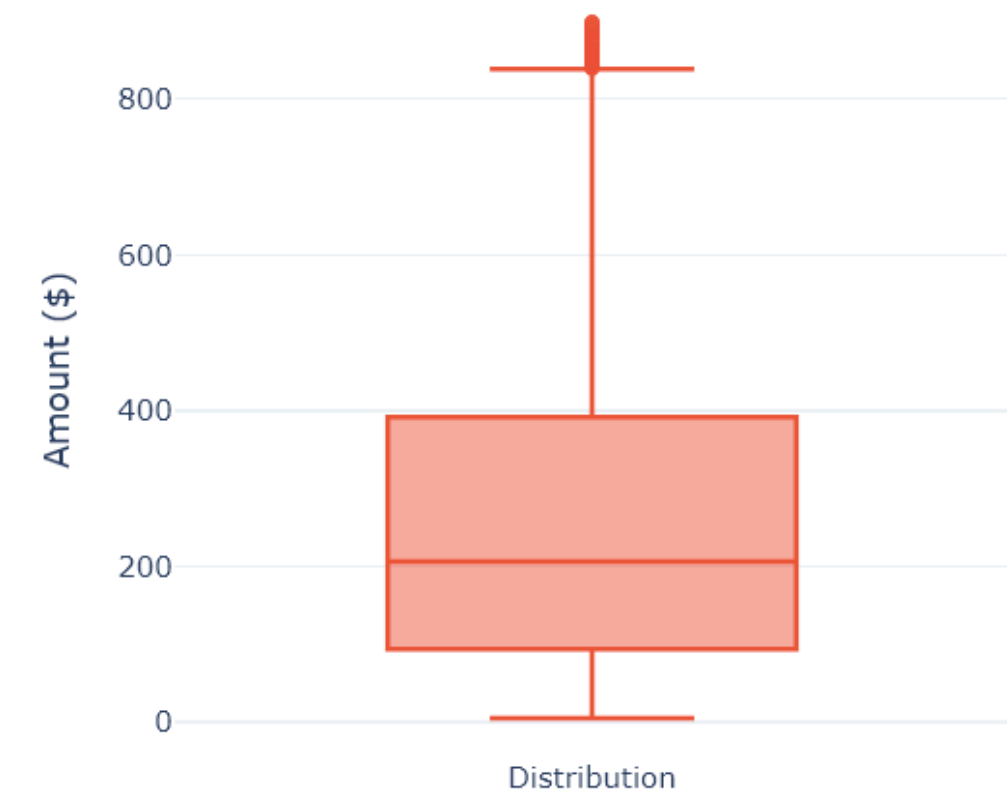
Key Technique: Min-Max Normalization was applied to total_revenue to scale values.

Business actions: Define Basket Segments (Small: < \$100, Medium: \$100-\$400, Large: > \$400) based on quartiles.- Use: Tailor cross-sell offers and promotions according to basket size.

Order amount distribution analysis



Box plot



Thank You