

Analysis of a Dataset for Bizops (Oluwatobi Adebawo)

Data source: [link here](#)

Database Schema:



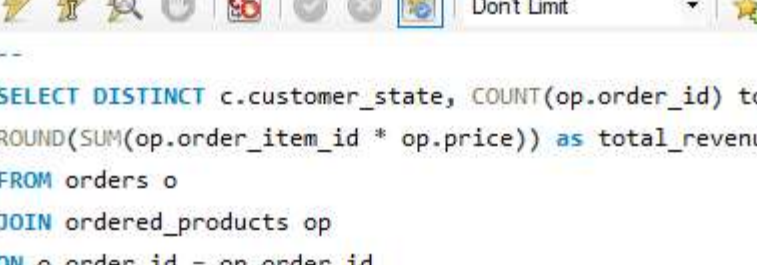
The dataset describes different product orders, timestamps, payments, reviews and customers data from Bizops for a period of less than 8 months (January to September 2018) across 27 states.

Dataset Investigation Questions:

- Which states should this business invest more marketing spend in?
- Which sellers, if any, should be delisted from this platform?

Findings:

The states in which the business should focus more marketing spend should include states with highest revenue and highest number of orders during the period of time. Data was queried using Mysql.

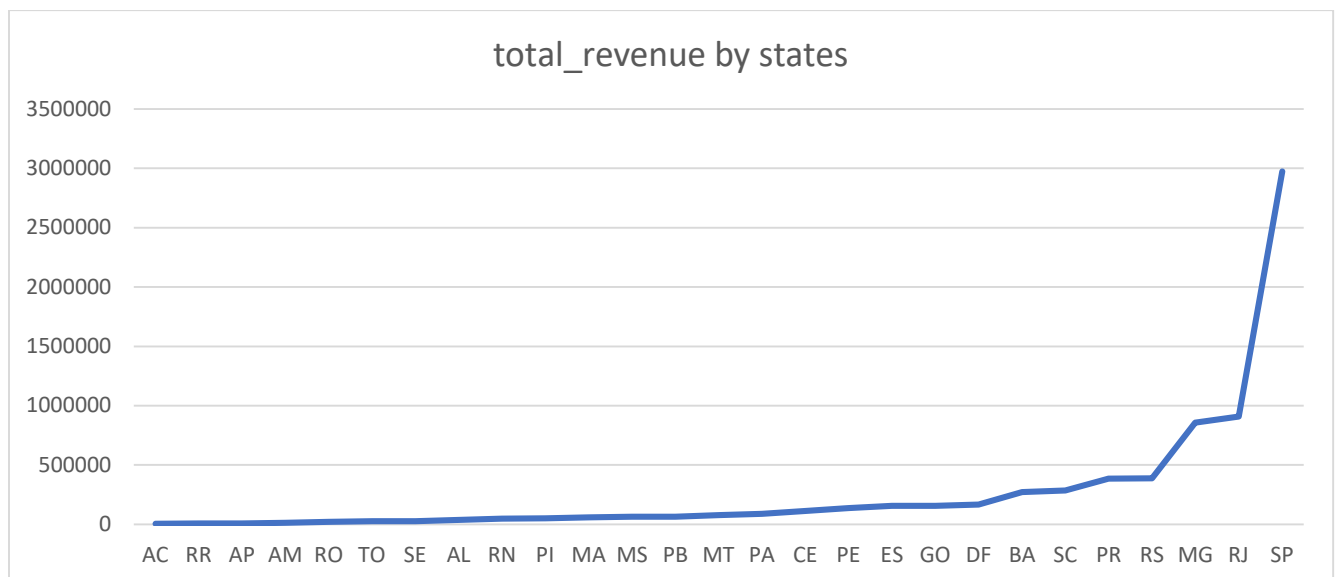


The screenshot shows a SQL IDE window titled "Administration - Server Status". The SQL editor contains a query to find the top states by total revenue. The query is as follows:

```
--
93 • SELECT DISTINCT c.customer_state, COUNT(op.order_id) total_order,
94 ROUND(SUM(op.order_item_id * op.price)) as total_revenue
95 FROM orders o
96 JOIN ordered_products op
97 ON o.order_id = op.order_id
98 JOIN customers c
99 ON o.customer_id = c.customer_id
100 GROUP BY 1 ORDER BY 3 DESC;
101 # the result shows that states SP, RJ, MG, PR had the highest number of orders
102 # Investors will be assured of continued order requests and returns
```

Below the query, the "Result Grid" tab is active, displaying the results of the query in a table:

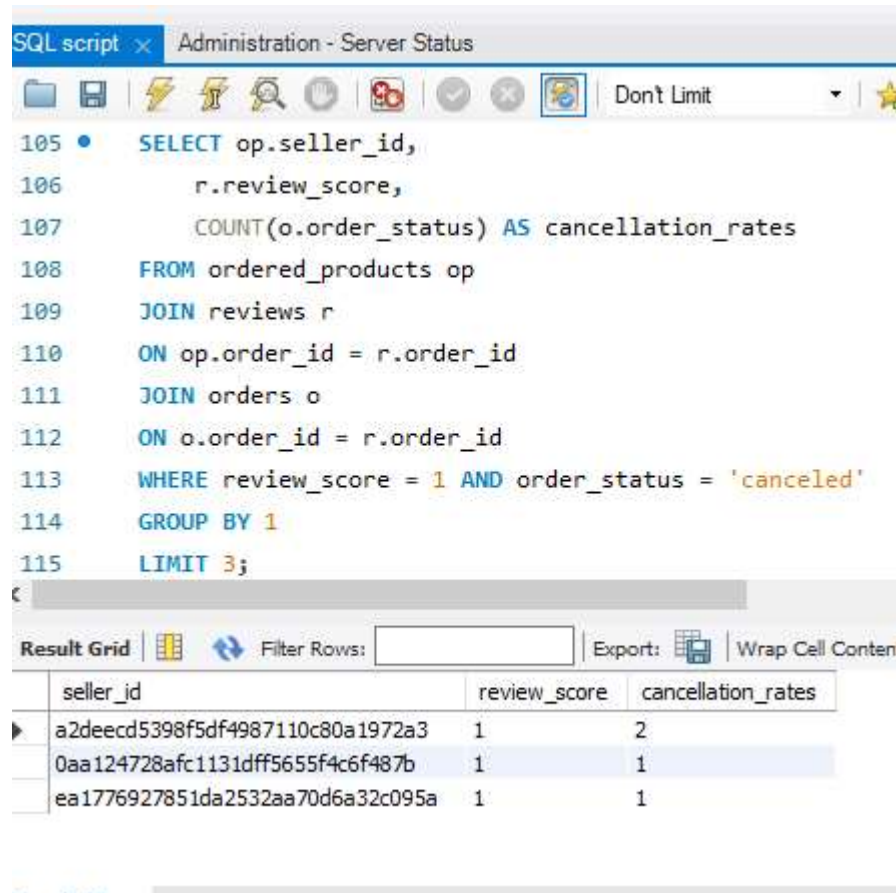
	customer_state	total_order	total_revenue
▶	SP	27279	3393930
	RJ	7495	1027269
	MG	6957	948158
	PR	3148	442417
	RS	3204	441072



The result shows that states SP, RJ, MG, PR had the highest number of orders and total revenue respectively. Investors will be assured of continued order requests and return on investment

Findings 2

Sellers that should be delisted from the platform should include sellers with highest cancellation rates and having lowest review scores. The query below was used to draw this information



The screenshot shows an SQL script editor window titled "Administration - Server Status". The script is as follows:

```
105 • SELECT op.seller_id,  
106         r.review_score,  
107         COUNT(o.order_status) AS cancellation_rates  
108 FROM ordered_products op  
109 JOIN reviews r  
110 ON op.order_id = r.order_id  
111 JOIN orders o  
112 ON o.order_id = r.order_id  
113 WHERE review_score = 1 AND order_status = 'canceled'  
114 GROUP BY 1  
115 LIMIT 3;
```

Below the script, the "Result Grid" is displayed with the following data:

seller_id	review_score	cancellation_rates
a2deecd5398f5df4987110c80a1972a3	1	2
0aa124728afc1131dffa5655f4c6f487b	1	1
ea1776927851da2532aa70d6a32c095a	1	1

The Image above shows a seller_id with highest cancellation rate of 2 and low review score. I think this seller_id should be delisted from the platform based on this.