

Vandit Sadaphale

Sparro Assessment Report - Junior Data Scientist

Task 1 (Google Sheet)

Question 1:

2018 ROAS	18.3270
2019 ROAS	20.0610
YoY Improvement	9.46%

Working:

Calculated ROAS of 2018 and 2019 individually using the formula Revenue/Sum, for the whole year

Question 2:

AVERAGE of ROAS each day	Year	
Day of week	2018	2019
Tuesday	18.3744	19.6007
Monday	17.5742	19.7841
Sunday	17.1204	19.8551
Wednesday	18.1358	20.1054
Saturday	17.6739	20.1949
Friday	18.3284	20.6696
Thursday	17.6872	20.7934

Working:

Created a Pivot table to visualize average ROAS of each day for both 2018 and 2019 to determine performance of the days.

Used a custom column 'ROAS each day' and 'Year' for easy calculation

1. Tuesday was the best-performing day in 2018, but a drastic change was observed in 2019, as Tuesday became the least-performing day.
2. The 2nd best-performing day in 2018 and 2019 remained the same, as Friday maintained its spot there.
3. The change in Thursday's performance from 2018 to 2019 is significant, as Thursday actually became the best-performing day of 2019, marking a notable improvement from its previous ranking in 2018.

Question 3:

July 2018 Avg Daily Budget	\$100,687.57
Suggested July 2019 Daily Budget	\$110,213.98
Total suggested budget for July 2019	\$3,416,633.30

Working:

1. Calculated Average Daily budget for July 2018.
2. Using the above value and YoY improvement calculated in Q1, calculated suggested daily budget for July 2019
3. Multiplied the suggested daily average budget by 31 to get the Total budget for July 2019

Question 4:

Month	AVERAGE DAILY ROAS
2018 - 03	11.6692
2018 - 02	12.6032
2018 - 04	13.1322
2018 - 01	13.6583
2018 - 06	15.4896
2018 - 05	16.9693
2019 - 02	19.0457
2019 - 04	19.7595
2019 - 05	20.1670
2018 - 07	20.1818
2019 - 01	20.8438
2019 - 03	20.9007
2018 - 09	21.4455
2018 - 11	21.7196
2018 - 10	21.7563
2018 - 12	21.8791
2018 - 08	23.6006

Working:

Created a Pivot table to visualize the daily average ROAS for each month and sorted it in ascending order to identify the worst-performing month

Since no changes were made in the account, only external factors could be the reason for March 2018 to be the worst-performing month. These reasons could be (but not limited to) the following:

1. Industrial Seasonality: It is observed that March, February and April were the worst-performing months. Usually, March-April marks the end of summer in Australia, the product might be less relevant or sought after in this season. Also, this period is right after the summer holiday season (Dec-Feb), and the post-holiday lull could be a reason too.

2. High Competition: There might be increased competition in March 2018, competitors might have run aggressive campaigns.

3. Economic and other external factors: There could be a change in consumer behaviour due to economic factors or other social changes, like elections in 2018.

These are some potential reasons explaining the poor performance in March 2018. Still, additional data would be required to investigate further the cause, for example, some information about the marketed product, previous year sales etc.

Task 3 (Python Data Analysis)

Data Cleaning and Preparation:

The data preparation process involved cleaning and merging two primary datasets:

1. Keyword Search Volume Data ('Data | C.1.csv')
2. Keyword Spend Data ('Data | C.2.csv')

Key steps in the data preparation process included:

- Removal of null values and irrelevant columns
- Standardization of text data (e.g., keywords, column names)
- Conversion of date fields to proper datetime format
- Transformation of percentage and currency values to numerical format
- Restructuring of data from wide to long format for analysis
- Merging of the two datasets based on keyword and month

Additional derived metrics, such as the average volume-to-spend ratio, were calculated to facilitate a deeper analysis of keyword performance.

Question 1: The monthly search volume of each keyword was shown.

	keyword	May 2023	Jun 2023	Jul 2023	Aug 2023	Sep 2023	Oct 2023	Nov 2023	Dec 2023	Jan 2024	Feb 2024	Mar 2024	Apr 2024
0	beach cover ups	4400.0	5400.0	6600.0	6600.0	8100.0	8100.0	8100.0	9900.0	9900.0	5400.0	5400.0	4400.0
1	black cocktail dresses	5400.0	5400.0	6600.0	8100.0	8100.0	9900.0	8100.0	3600.0	5400.0	5400.0	6600.0	5400.0
2	black designer dresses	320.0	260.0	320.0	210.0	260.0	480.0	210.0	140.0	260.0	320.0	260.0	260.0
3	cocktail dresses	60500.0	60500.0	74000.0	90500.0	90500.0	110000.0	110000.0	49500.0	74000.0	74000.0	74000.0	74000.0
4	cotton midi dress	320.0	480.0	590.0	880.0	1300.0	1300.0	1300.0	1600.0	1000.0	1000.0	880.0	590.0
5	day dresses	1600.0	1600.0	1900.0	2900.0	3600.0	4400.0	4400.0	3600.0	2900.0	2900.0	2900.0	1900.0

Question 2: Month over Month biggest risers and fallers.

After analyzing the data, the difference in the Absolute value of search volume was considered to calculate risers and fallers as it brought more value to the outcome.

Results:

For e.g., biggest riser for Jun 2023 was 'beach cover ups' with an increase of 1000 in search volume from previous month. However, the biggest faller for that month was 'long sleeve midi dress' with a decrease of 4900 in search volume.

The complete result for all the months was displayed. (Snapshot attached below)

For Jun 2023:
Biggest Riser: beach cover ups: +1,000
Biggest Faller: long sleeve midi dress: -4,900

For Jul 2023:
Biggest Riser: cocktail dresses: +13,500
Biggest Faller: long sleeve knitted dress: -4,900

For Aug 2023:
Biggest Riser: cocktail dresses: +16,500
Biggest Faller: long sleeve knitted dress: -4,500

For Sep 2023:
Biggest Riser: denim dresses: +9,000
Biggest Faller: long sleeve knitted dress: -3,000

For Oct 2023:
Biggest Riser: cocktail dresses: +19,500
Biggest Faller: short sleeve knitted dress: -200

For Nov 2023:
Biggest Riser: party dresses: +6,000
Biggest Faller: evening dresses: -9,000

For Dec 2023:
Biggest Riser: beach cover ups: +1,800
Biggest Faller: cocktail dresses: -60,500

For Jan 2024:
Biggest Riser: cocktail dresses: +24,500
Biggest Faller: party dresses: -4,100

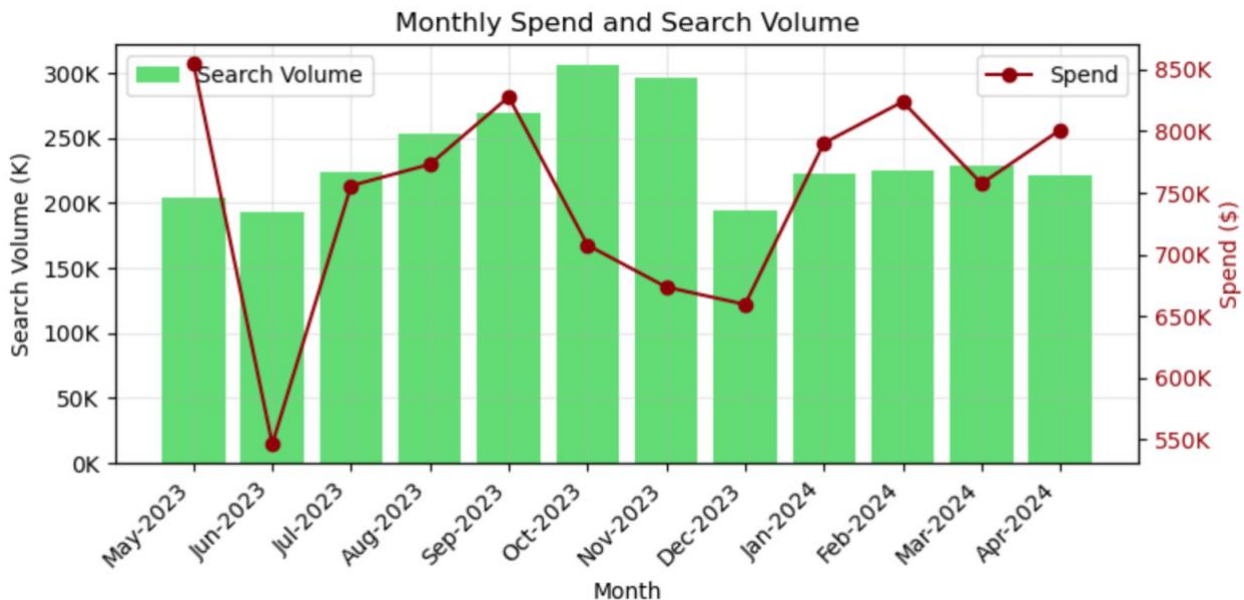
For Feb 2024:
Biggest Riser: denim dresses: +9,000
Biggest Faller: beach cover ups: -4,500

For Mar 2024:
Biggest Riser: long sleeve midi dress: +3,000
Biggest Faller: party dresses: -3,300

For Apr 2024:
Biggest Riser: long sleeve knitted dress: +5,500
Biggest Faller: denim dresses: -9,000

Question 3: Visualize Spend to Search volume

First, a monthly spend-to-search volume graph was plotted to understand the overall trend between these two fields. It can be observed that the highest spending occurred in May 2023, followed by September 2023 and February 2024. However, the search volume for these months doesn't look promising, as it remained below 230K for each of these months. Conversely, in June, October, and November 2023, higher search volumes were observed with lower spending.



Since no direct or consistent pattern was observed in this graph, another graph was plotted showing the trend between money spent and search volume for the top four most successful and most unsuccessful keywords, based on the average volume-to-spend ratio.

This second graph helped in understanding which keywords are performing poorly or are unable to capitalize on the amount spent on them, compared to the best-performing keywords. This comparison highlights the drastic difference between them, helping us to develop a strategy for next steps.

The below snapshot shows the top two most successful and most unsuccessful keywords, and the trend between their search volume and spend.

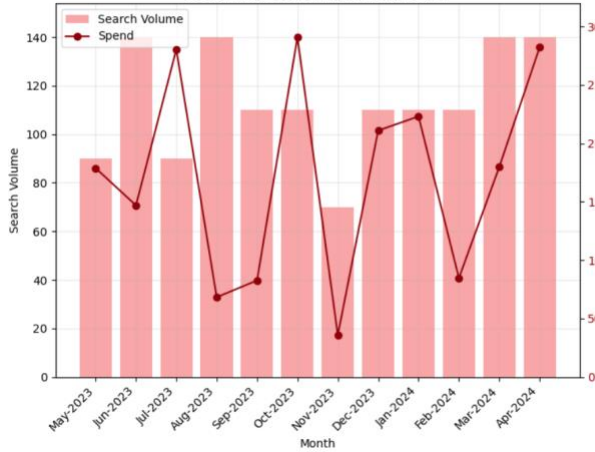
The worst-performing keyword, 'designer resort wear', never crossed the 150 mark in search volume, while the money spent on it ranged between \$3k to \$30k. On the other hand, 'cocktail dresses', the most successful keyword, consistently maintained search volumes above 50k, sometimes reaching as high as 110k, with spending ranging between \$500 to \$30k.

There could be many reasons to explain these differences. For example, 'designer resort wear' might be a rare and expensive product where fewer sales could generate substantial revenue. In contrast, 'cocktail dresses' might be a more commonly purchased item, explaining the high search volumes. However, additional information and domain knowledge are required to clearly understand these discrepancies or differences.

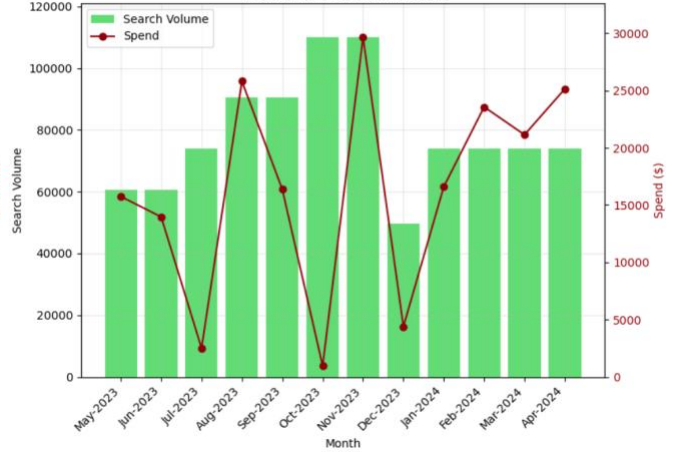
Keywords Performance

▼ Worst Performing Keywords (Left, Red) | ▲ Best Performing Keywords (Right, Green)

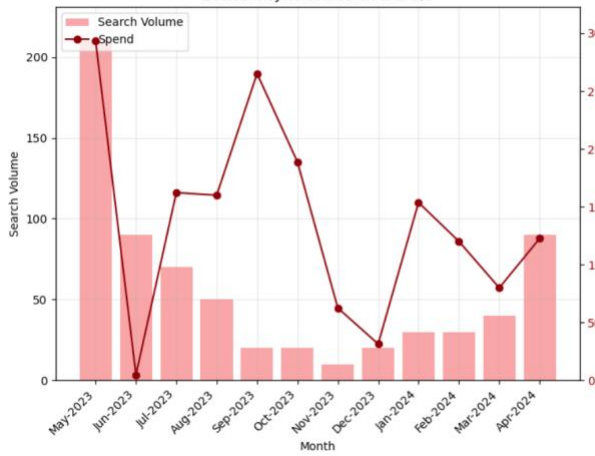
Bottom Keyword: designer resort wear



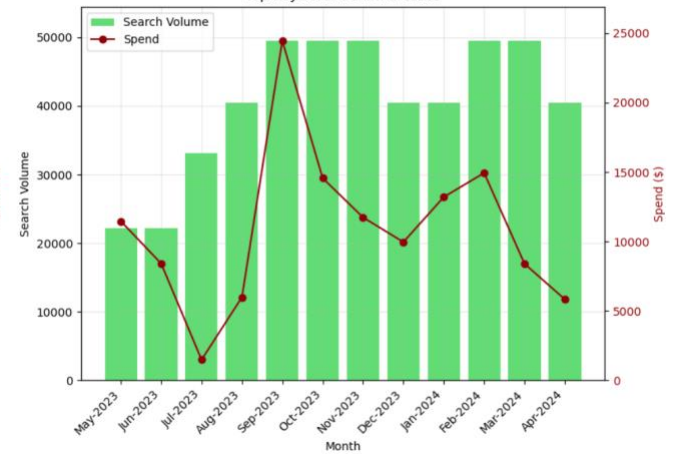
Top Keyword: cocktail dresses



Bottom Keyword: wool midi dress



Top Keyword: denim dresses



Task 2 (SQL)

Question 1: Channels with highest average ROAS

Query -

```
select channel, avg(revenue / cost) as avg_roas from marketing_data group by channel order by avg_roas desc;
```

Output:

CHANNEL	AVG_ROAS
Organic Social	2.17965368
Email	2.16900703
Paid Social	2.1590864
Paid Search	2.04037698
Organic Search	1.84259259

Question 2: Total Revenue, Cost for each product in Jan and Feb

Query –

```
select product, sum(case when extract(month from date) = 1 then revenue else 0 end) as january_revenue, sum(case when extract(month from date) = 1 then cost else 0 end) as january_cost, sum(case when extract(month from date) = 2 then revenue else 0 end) as february_revenue, sum(case when extract(month from date) = 2 then cost else 0 end) as february_cost from marketing_data group by product;
```

Output:

PRODUCT	JANUARY_REVENUE	JANUARY_COST
Product_A	3200	1800
Product_B	3100	1400
Product_C	2600	1300

Question 3: Product with highest ROAS

```
select product, sum(revenue) / sum(cost) as roas from marketing_data group by product order by roas desc limit 1;
```

Output:

PRODUCT	ROAS
Product_E	2.27777778