DS230
SQL for Data Analysis

Final Project

Amanda Doty

Introduction

Purpose

Problem: A business is wondering about their recent advertising campaigns and their product line. They ask the following questions:

Which campaigns were the most successful at creating orders and generating revenue?

Which products sell the most often?

Which products are the most expensive? Does the price of a product relate to the rate at which it sells?

Data

The business gave me access to their customer list, but that ultimately did not provide any insight to their questions.

However, their order lines, orders, and products data were able to be used to create queries that could provide insight into their questions.

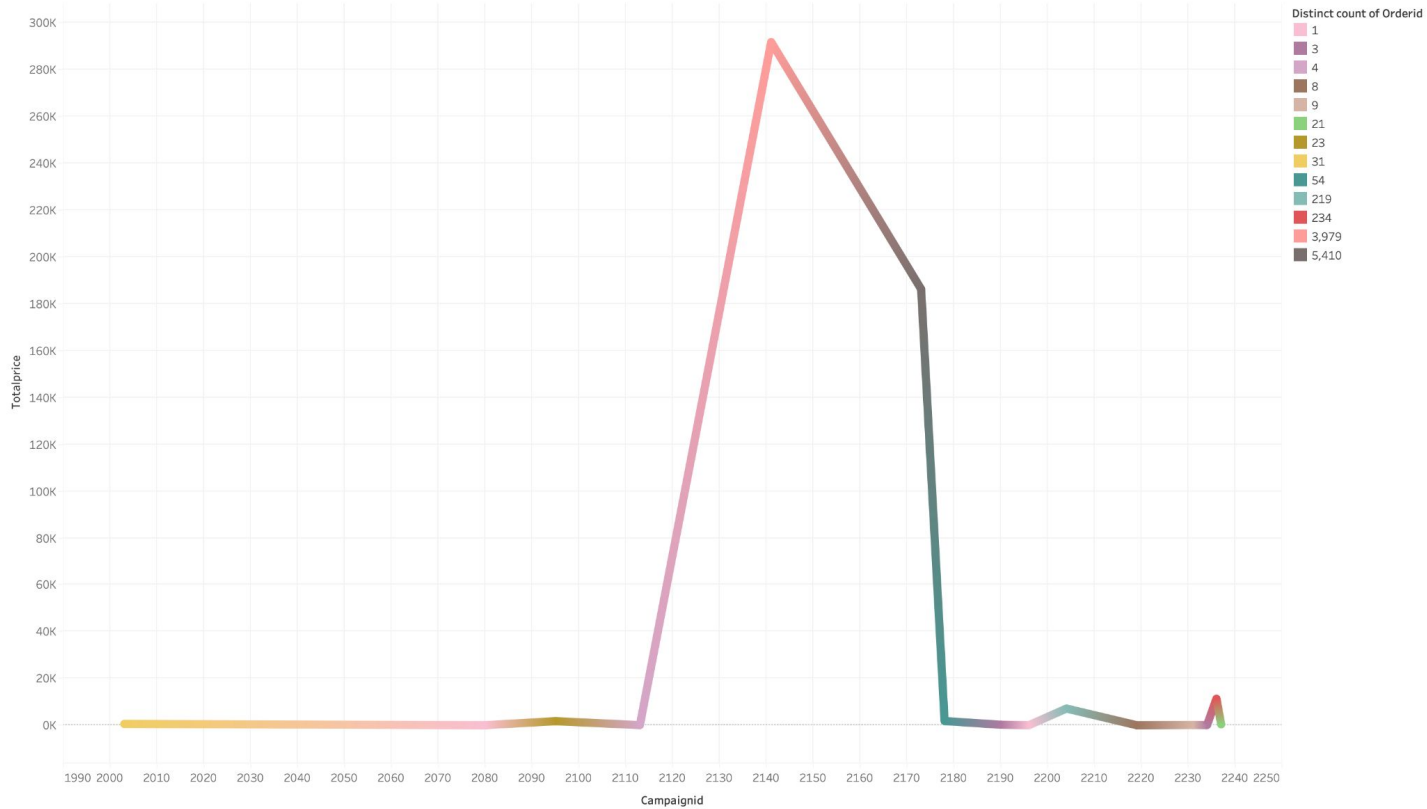
Each table had 10000 entries.

Business Query 1 - Which campaign was the most successful?

SQL

```
SELECT DISTINCT orderid, totalprice, campaignid
FROM orders
WHERE campaignid IS NOT NULL
ORDER BY totalprice DESC;
```

Most Successful Campaigns



The trend of sum of Totalprice for Campaignid. Color shows details about distinct count of Orderid.

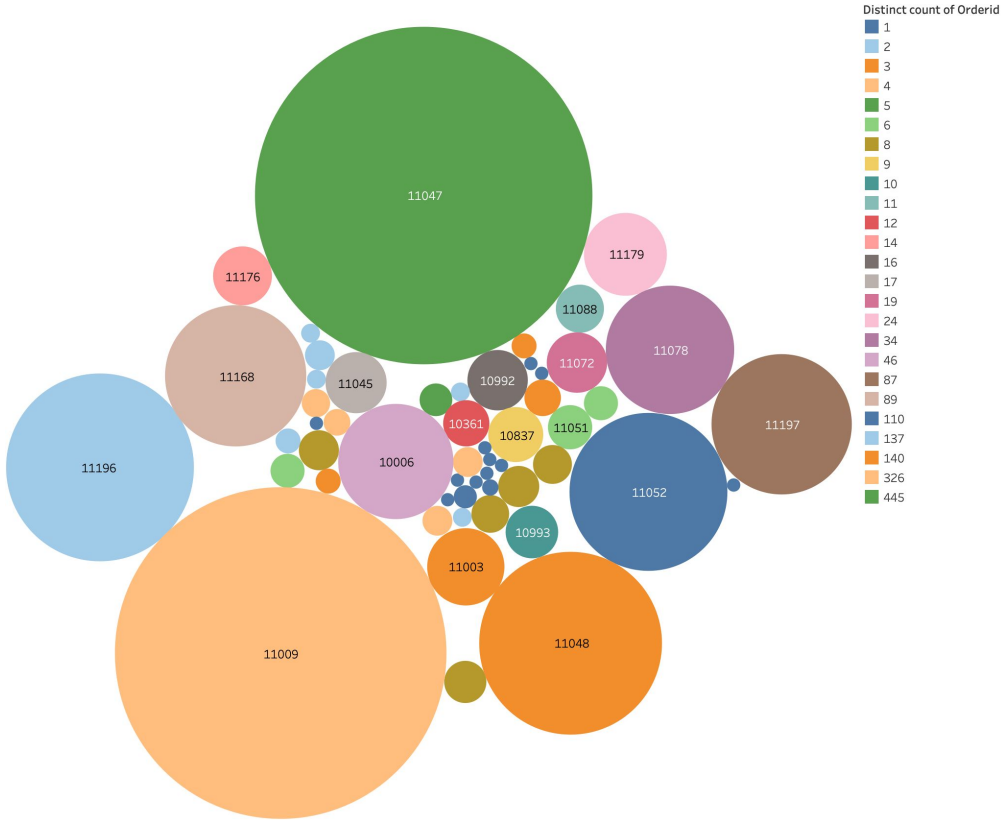
Color indicates number of orders, line designates total amount sold.
This graph indicates that campaign 2141 was the most successful, with 3979 orders, selling $291,864 worth of products.

Business Query 2 - Which products sell the most?

SQL

```
SELECT A.productid, B.totalprice, B.orderid, sum(B.numunits) as units
FROM products A JOIN
   order_lines B
   ON
     A.productid = B.productid
WHERE b.totalprice > 1
group by a.productid, b.totalprice, b.orderid
HAVING sum(B.numunits) > 1
ORDER BY units DESC;
```

Most Popular Products



Productid. Color shows details about distinct count of Orderid. Size shows sum of Units. The marks are labeled by Productid.

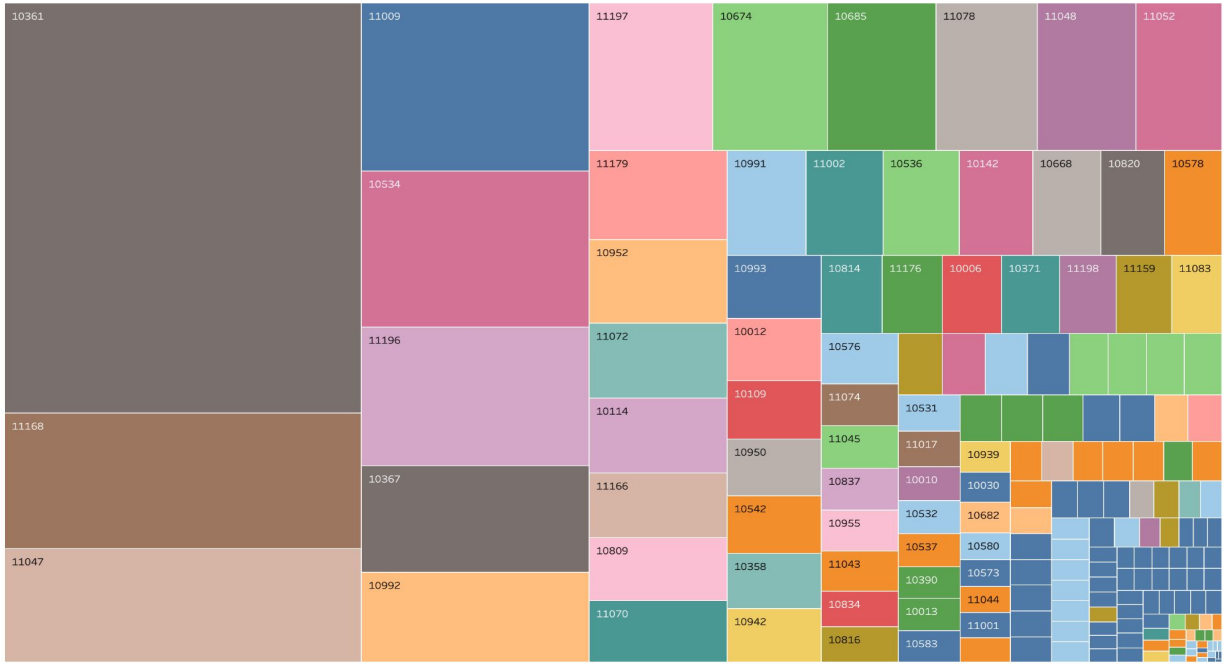
This shows that product 11047 is the most popular based on distinct orders and number of units sold, with 11009 a close second.

Business Query 3 - Which products (by price) are the most popular?

SQL

```
SELECT A.productid, B.totalprice, B.orderid, sum(B.numunits) as units
FROM products A JOIN
   order_lines B
   ON
     A.productid = B.productid
WHERE b.totalprice > 1
group by rollup (b.orderid, a.productid, b.totalprice);
```

Most popular products by price



Units
2
4
6
8
10
12
14
16
24
26
28
30
34
38
40
46
52
54
56
60
62
74
76
82
84
88
92
104
108
112
120
122
126
128
142
162
166
170
174
176
178
182
184
222
298
308
320
354
428
516
596
608
758
862
1,186
1,322
1,366
1,996
2,128
3,800
4,096
25,896

Productid. Color shows details about sum of Units. Size shows sum of Totalprice. The marks are labeled by Productid.

The most expensive product is 10361, but the most popular product is 11009. Whereas 10361 has sold less than 20% of the units of 11009

Recommendations

Advertising

I would recommend a similar campaign as 2141. However, it would be a good idea to collect qualitative data about target audience before launching the campaign.

Products

Ensure secure stocking of product 11047 and 11009. I also recommend lowering the price of product 10361. It is almost 5 times as expensive as the most popular product, but sells at less than 1/5th the rate. I would recommend running some regression tests to find a more appropriate price, if one exists, that could increase the number of units sold while maintaining the revenue.