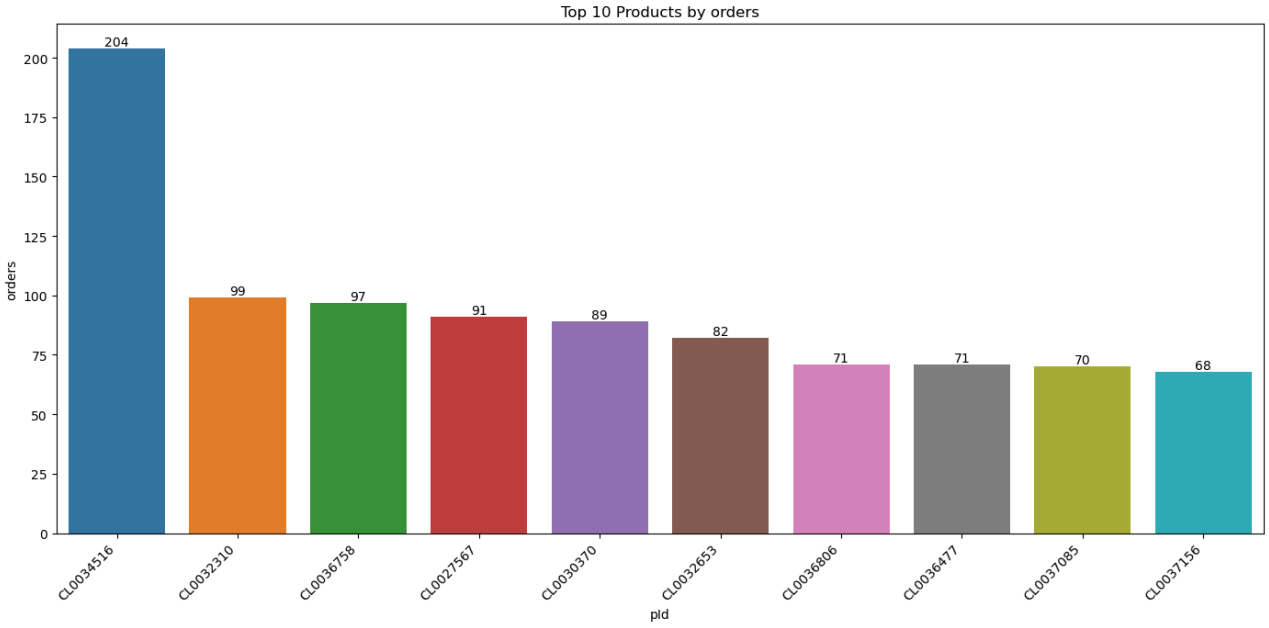
Search Platform Data Analysis

1) Overview

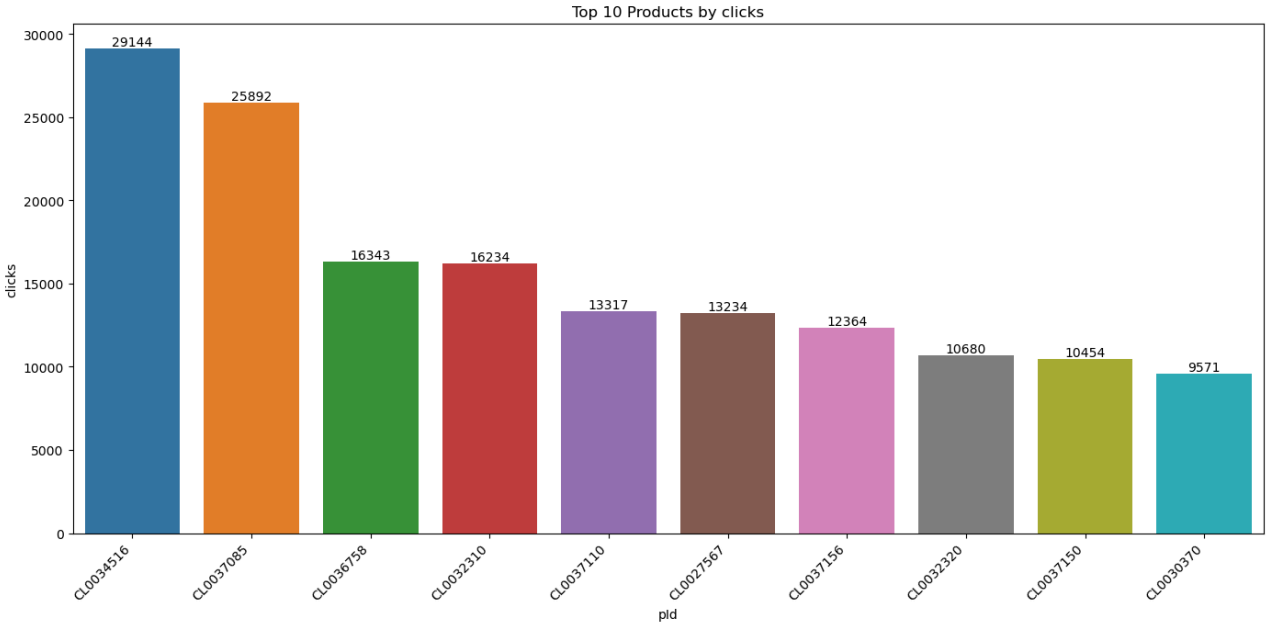
This report analyzes the performance of search queries and product data to get insights of key trends for improving the search experience. The data covers various metrics, such as Click-Through Rates (CTR), search volume (Hits), clicks, orders, and the conversion funnel.

2) Analyzing best performing products



Top 10 Products by orders:

	pId	pName	clicks	carts	orders
0	CL0034516	Chevron stripe colorblock maxi dress	29144	2181	204
3	CL0032310	Cotton knit fit and flare dress	16234	350	99
2	CL0036758	Bird print pleated cotton knit dress	16343	340	97
5	CL0027567	Elastic waist chambray maxi dress	13234	452	91
9	CL0030370	Pleat waist poplin shirtdress	9571	367	89
12	CL0032653	Cotton poplin A-line shirtdress	8287	247	82
10	CL0036806	Retro chambray shirtdress	8407	227	71
41	CL0036477	Seersucker stripe notch neck dress	3478	178	71
1	CL0037085	Chevron stripe colorblock maxi dress	25892	237	70
6	CL0037156	Chevron stripe colorblock maxi dress	12364	172	68

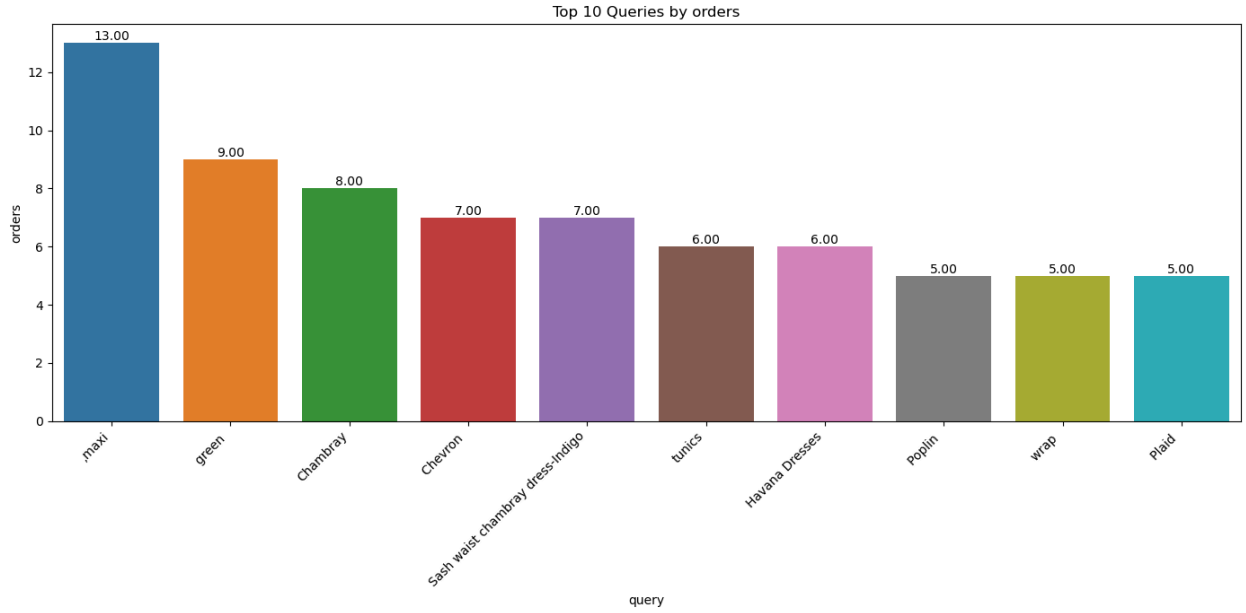


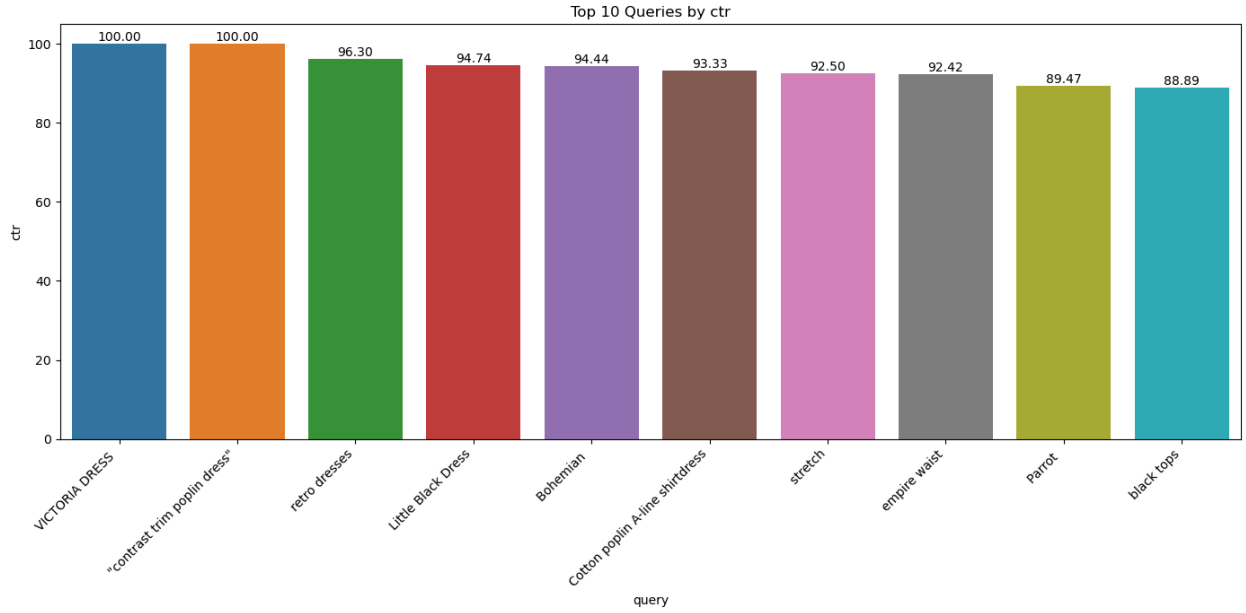
Top 10 Products by clicks:

	pId	pName	clicks
0	CL0034516	Chevron stripe colorblock maxi dress	29144
1	CL0037085	Chevron stripe colorblock maxi dress	25892
2	CL0036758	Bird print pleated cotton knit dress	16343
3	CL0032310	Cotton knit fit and flare dress	16234
4	CL0037110	Pleat front chambray denim maxi dress	13317
5	CL0027567	Elastic waist chambray maxi dress	13234
6	CL0037156	Chevron stripe colorblock maxi dress	12364
7	CL0032320	Floral embellished chambray denim maxi dress	10680
8	CL0037150	Drawstring waist chambray maxi dress	10454
9	CL0030370	Pleat waist poplin shirtdress	9571

The best performing products in terms of order placed and clicks are Chevron stripe colorblock maxi dress, Cotton knit fit and flare dress, Bird print pleated cotton knit dress indicating that the conversion rate was high for these products.

3) Analyzing best performing products





Top performers by orders:

- "maxi" leads with 13 orders
- "green" follows with 9 orders
- "chambray" ranks third with 8 orders

Top performers by click-through rate (CTR):

- "victorian dress" and "corset dress prom dress" both have a 100% CTR
- "retro dress" follows closely at 96.30% CTR
- "little black dress" ranks fourth with 94.74% CTR

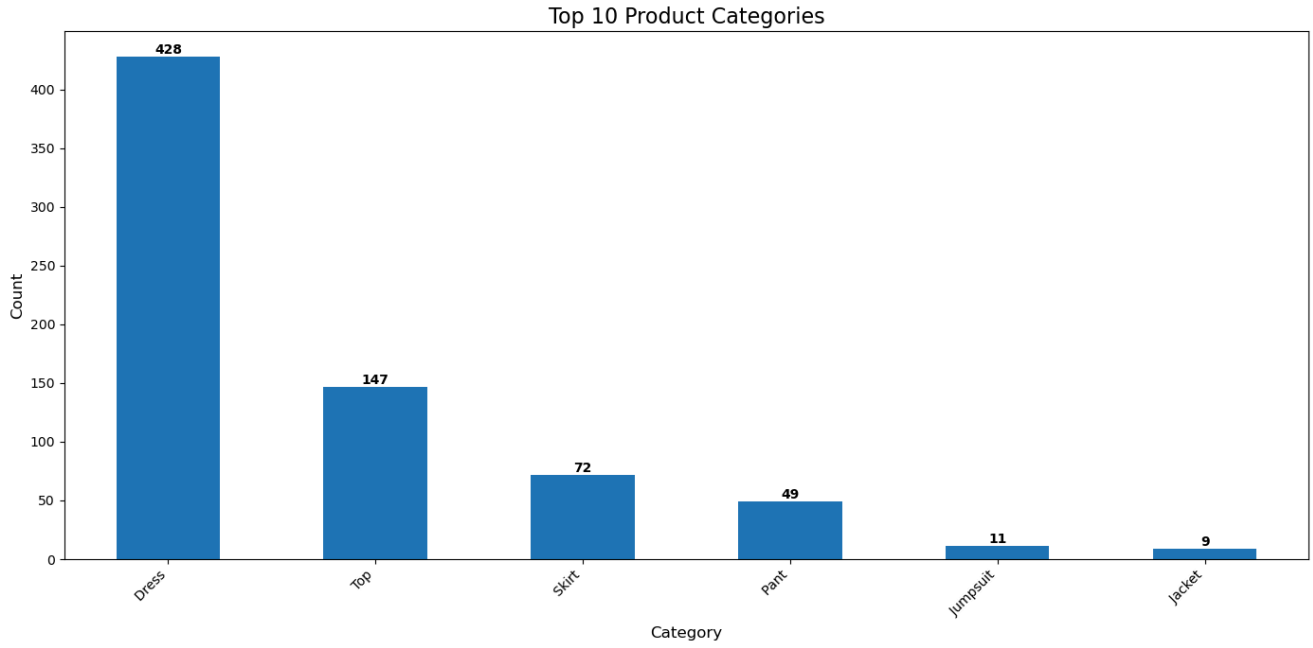
Comparison between orders and CTR:

- Interestingly, the top queries by orders don't appear in the top 10 for CTR
- This suggests that while these queries lead to more orders, they may have lower conversion rates compared to the high-CTR queries

Specific vs. general terms:

- More specific queries (e.g., "victorian dress", "corset dress prom dress") have higher CTRs
- General terms like "maxi" and "green" lead in orders but don't appear in top CTRs

4) Product Categories



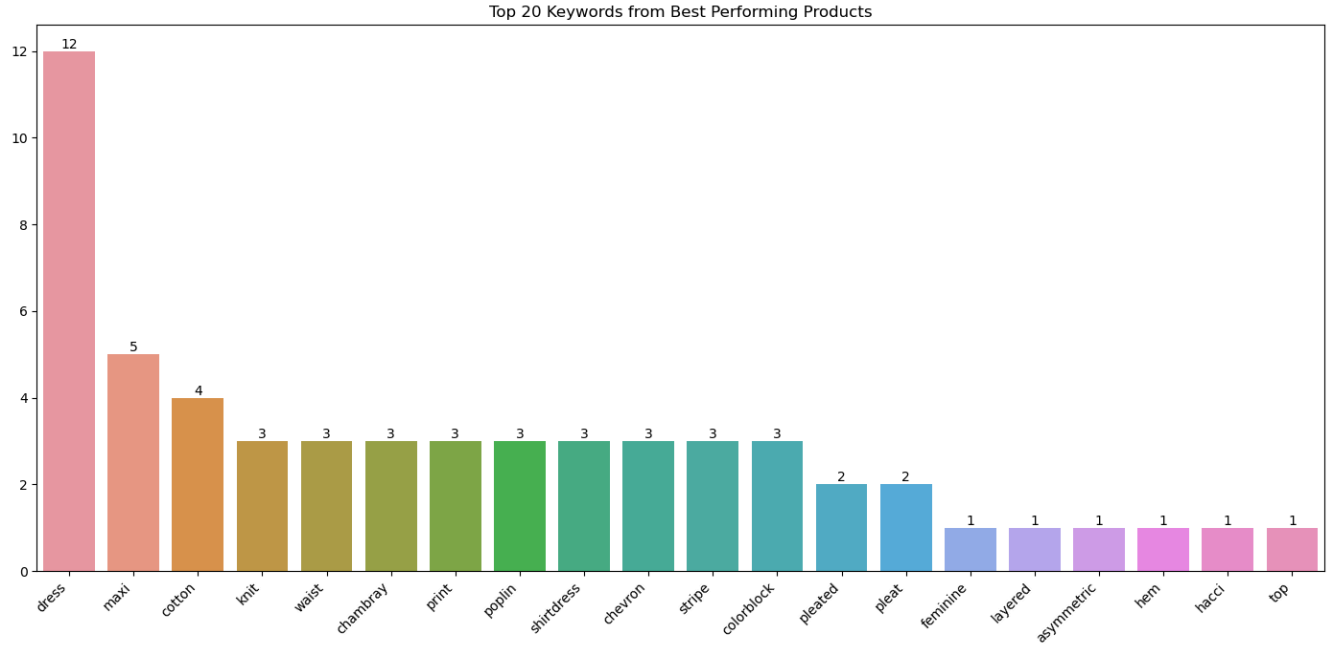
Product Categories:

- Dress is overwhelmingly the top category with 428 products.
- Top is the second most popular category with 147 products.
- There's a significant drop-off after the top two categories, with Skirt (72 products) and Pants (49 products) following.

Product Names Word Cloud :

- The most prominent terms are "cotton", "dress", "knit", and "print", indicating these are common features or types of products.
- "Floral" and "embellished" are also notable, suggesting popular design elements.
- Fabric types like "cotton", "jersey", and "crepe" are frequently mentioned.
- Various dress styles appear: "maxi dress", "shirtdress", "tunic".
- This word cloud reinforces the dominance of dresses seen in the category graph.

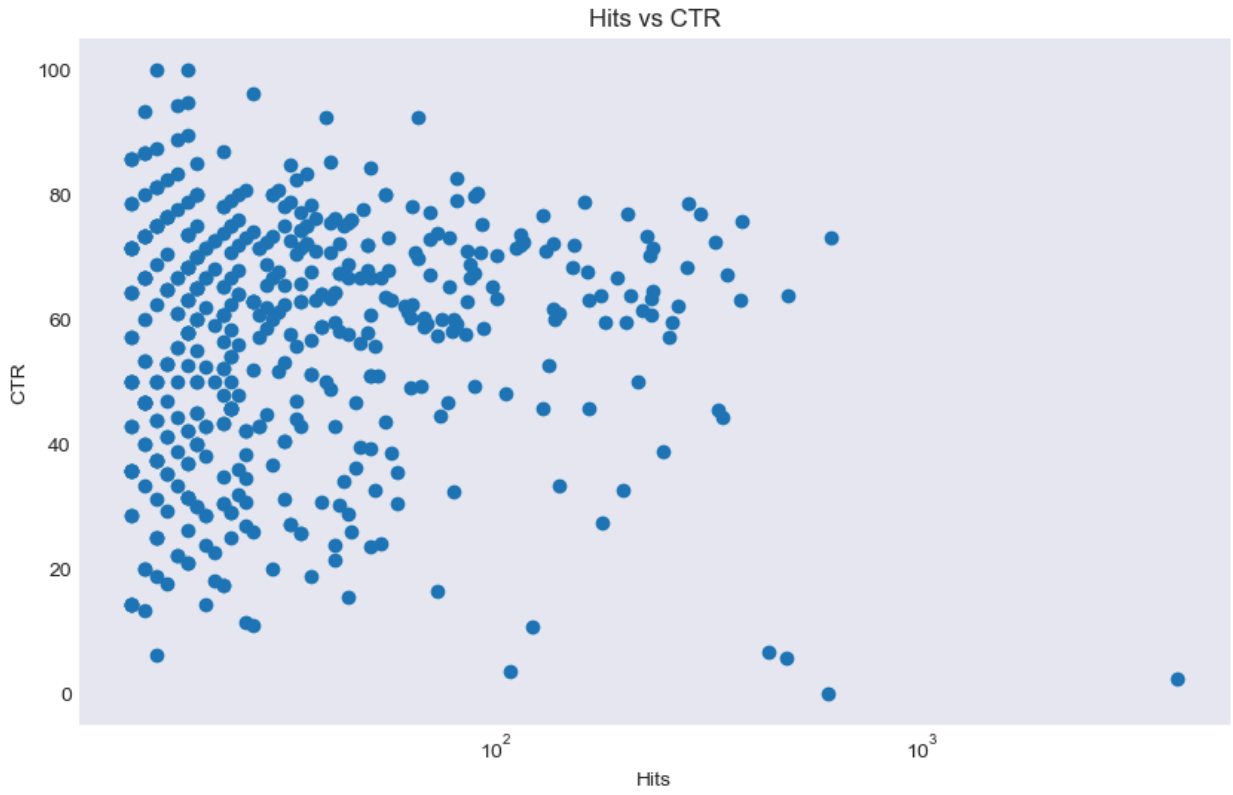
5) Keywords from Best Performing Products



Top Keywords from Best Performing Products:

- "Dress" is by far the most frequent keyword, appearing 12 times.
- "maxi" is the second most common keyword with 5 occurrences.
- "Cotton" is also prominent with 4 occurrences.
- Other notable keywords include "knit", "waist", "chambray", and "print".

6) Hits vs CTR from Query Performance Data



Data Distribution:

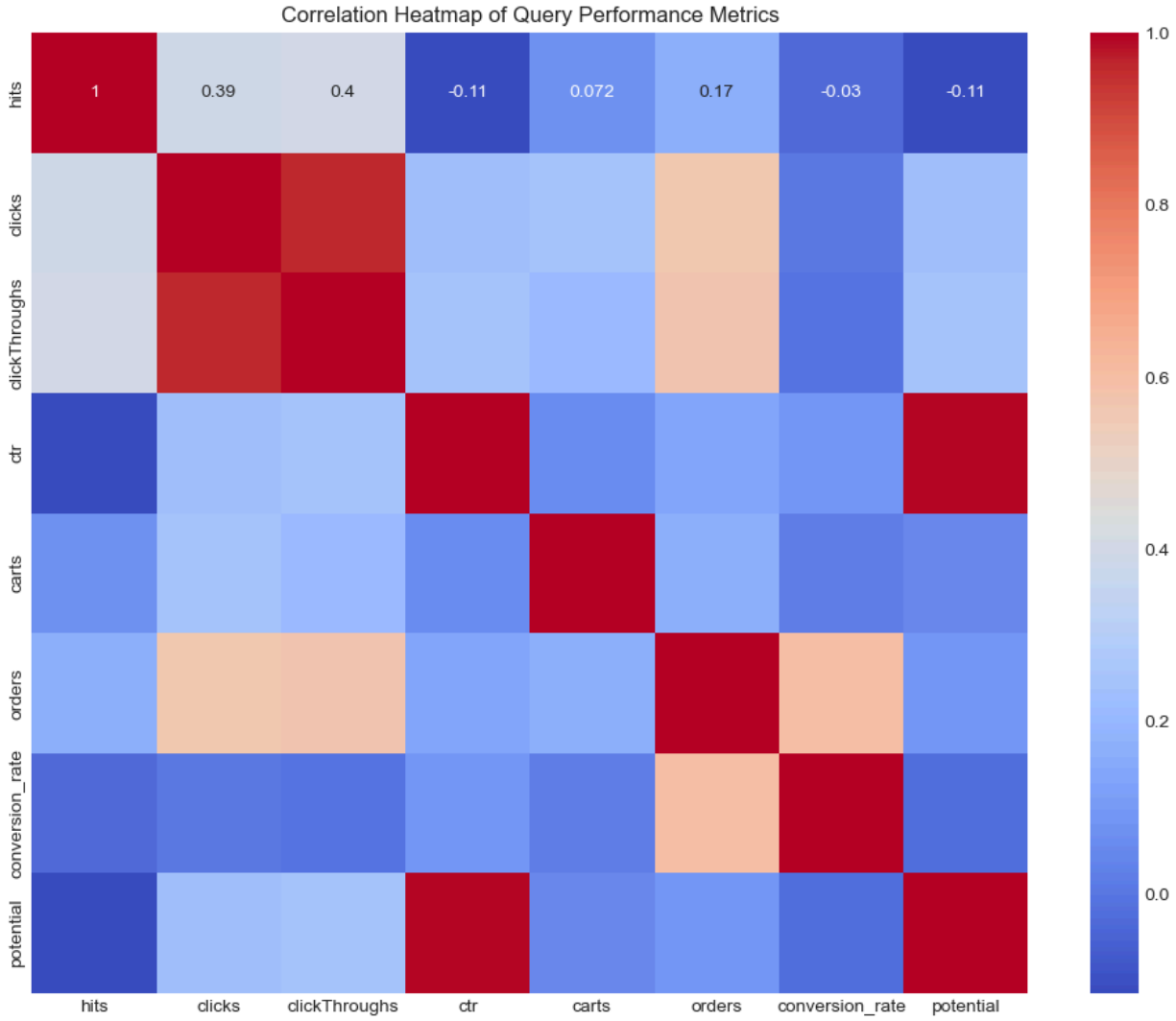
- Hits range from approximately 10 to over 1000 (log scale).
- CTR ranges from near 0% to 100%.
- Most data points are clustered between 20-100 hits and 20-80% CTR.

Correlation:

- There's no strong linear correlation between hits and CTR.
- Some high-hit queries have low CTR, and vice versa.

7)Correlation of Performance Metrics

1. Query Performance Metrics



a) Strong positive correlations:

- Hits and clicks (0.39)
- Clicks and clickThroughs (very strong, nearly 1.0)
- CTR and conversion_rate (0.75)

b) Weak or no correlations:

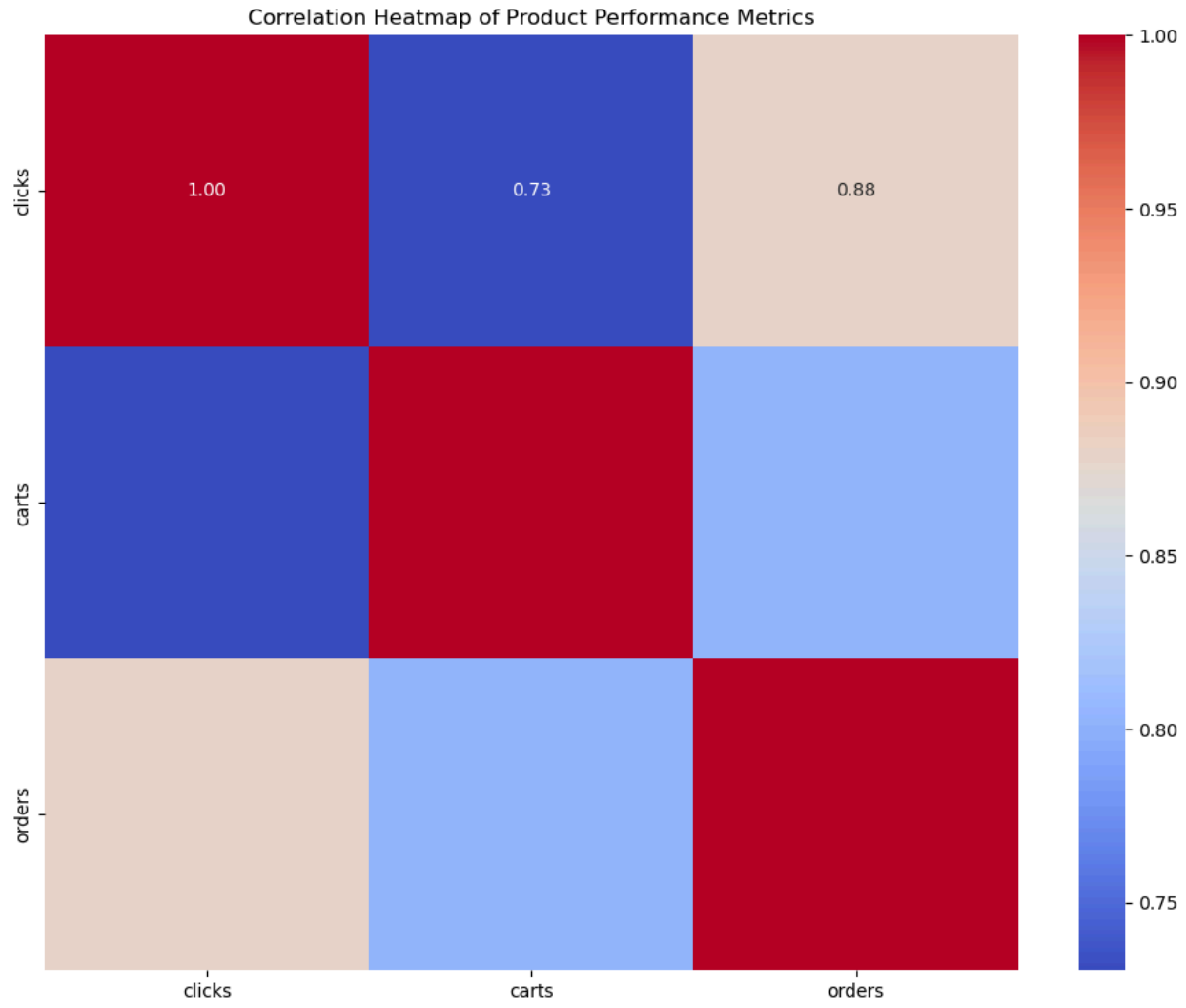
- Hits with most other metrics (CTR, carts, orders, conversion_rate)
- Clicks with carts, orders, and conversion_rate

c) Negative correlations:

- Hits with CTR (-0.11)
- CTR with potential (-0.34)

d) Insights:

- High query volume (hits) doesn't necessarily lead to more conversions
- Click-through rate (CTR) is a better indicator of conversion potential than raw hit counts
- There's a disconnect between early funnel metrics (hits, clicks) and late funnel metrics (carts, orders)



2)Product Performance Metrics:

a) Strong positive correlations:

- Clicks and carts (0.73)
- Clicks and orders (0.88)
- Carts and orders (very strong, nearly 1.0)

b) Insights:

- For products, there's a much stronger connection between early funnel (clicks) and late funnel (carts, orders) metrics compared to queries
- High click rates on products are good predictors of cart additions and orders
- Cart additions are extremely likely to result in orders

3) Comparative Insights:

a) Funnel coherence:

- Product metrics show a more coherent funnel from clicks to orders
- Query metrics show a disconnect between early and late funnel stages

b) Conversion indicators:

- For queries, CTR and conversion_rate are key performance indicators
- For products, click volume is a strong indicator of eventual orders

c) Volume vs. Efficiency:

- Query performance seems more about efficiency (CTR, conversion_rate) than volume (hits)
- Product performance shows a stronger link between volume (clicks) and outcomes (orders)

4)Strategic Implications:

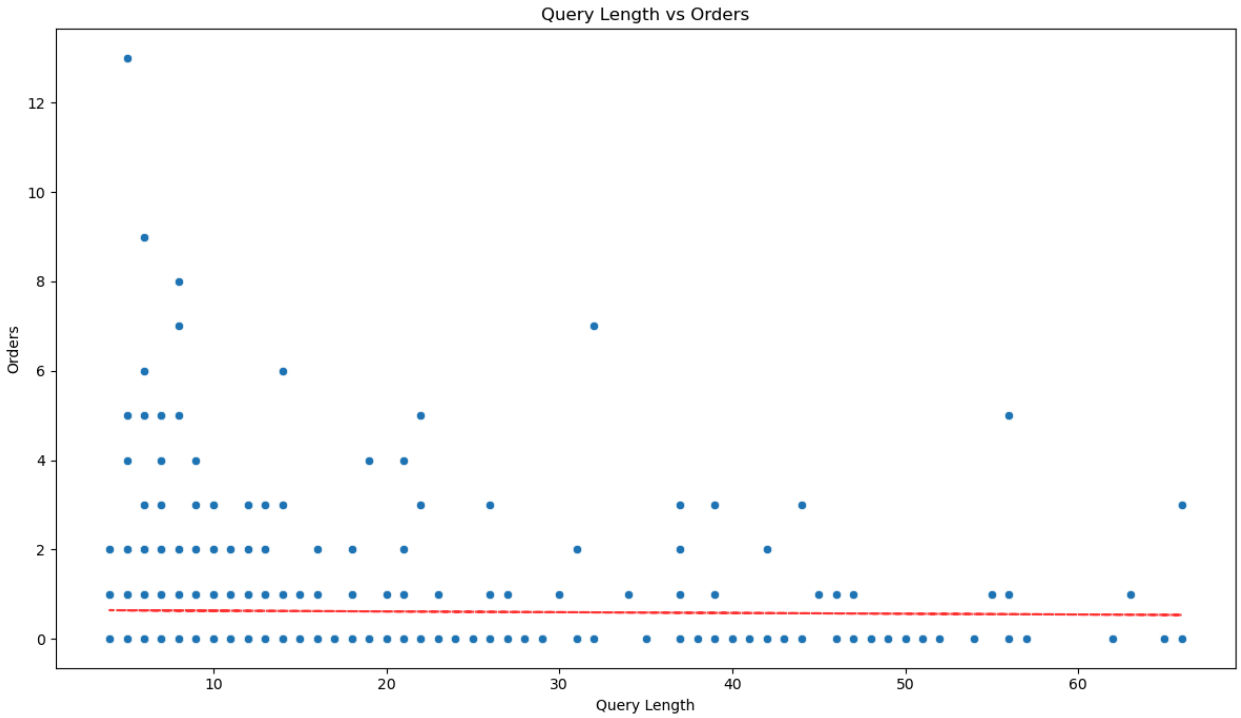
a) Query optimization:

- Focus on improving CTR rather than just increasing hit volume
- Investigate queries with high CTR but low conversion rates to identify potential issues.

b) Product optimization:

- Prioritize increasing product visibility and click-through rates, as these strongly correlate with orders
- Analyze products with high click rates but low cart/order rates to identify potential conversion barriers.

8) Query Length

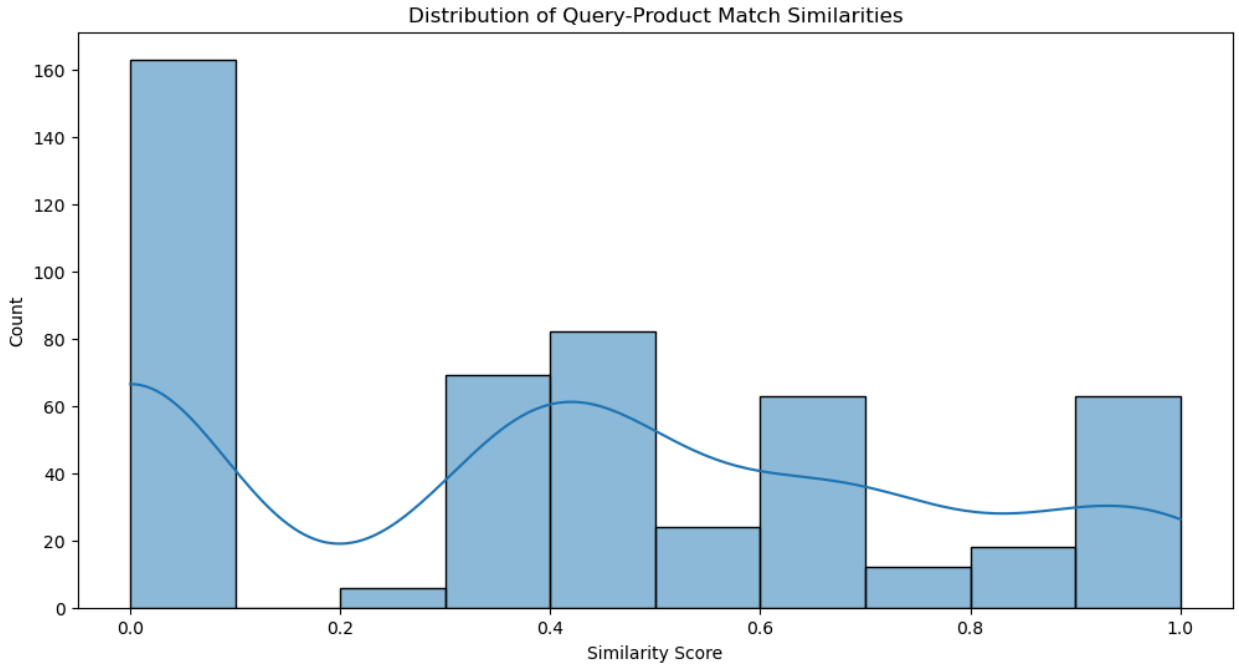


1. Distribution of Query Lengths:
    - Query lengths range from approximately 1 to 65 characters.
    - Most queries appear to be between 5 and 40 characters long.
2. Distribution of Orders:
    - The majority of queries result in 0 to 5 orders.
    - There are a few outliers with higher order counts, up to about 13 orders.
3. Short Queries:
    - Queries with lengths under 10 characters show the highest variability in order counts.
    - The highest number of orders (around 13) comes from a very short query.
4. Long Queries:
    - Longer queries (over 40 characters) tend to result in fewer orders, mostly 0-3.
    - There are fewer data points for very long queries, suggesting they are less common.
5. Mid-Length Queries:
    - Queries between 10-30 characters show a mix of performance, with some achieving higher order counts.
6. Zero Orders:
    - There are many queries across all lengths that result in zero orders, as indicated by the cluster of points along the bottom of the graph.
7. Optimal Length:

- While not definitive, the data suggests that shorter to medium-length queries (under 30 characters) have a higher potential for generating order.

1. Focus on optimizing short to medium-length queries (under 30 characters) as they show the most potential for high order counts.
2. Investigate the top-performing short queries to understand what makes them effective.
3. For longer queries resulting in zero orders, consider ways to improve their relevance or match them to products more effectively.
4. Encourage users to use more specific, targeted queries rather than very long, complex ones.
5. Analyze the outliers with high order counts to identify common characteristics that could be applied to improve performance of other queries.

9) Distribution of Query-Product Match Similarities



a) Overall distribution:

- The similarity scores range from 0 to 1, with 0 indicating no similarity and 1 indicating perfect similarity.
- The distribution is multimodal, with peaks at both ends and in the middle.

b) Low similarity matches:

- There's a large spike of matches with very low similarity scores (0-0.1).
- This suggests many queries don't closely match product descriptions.

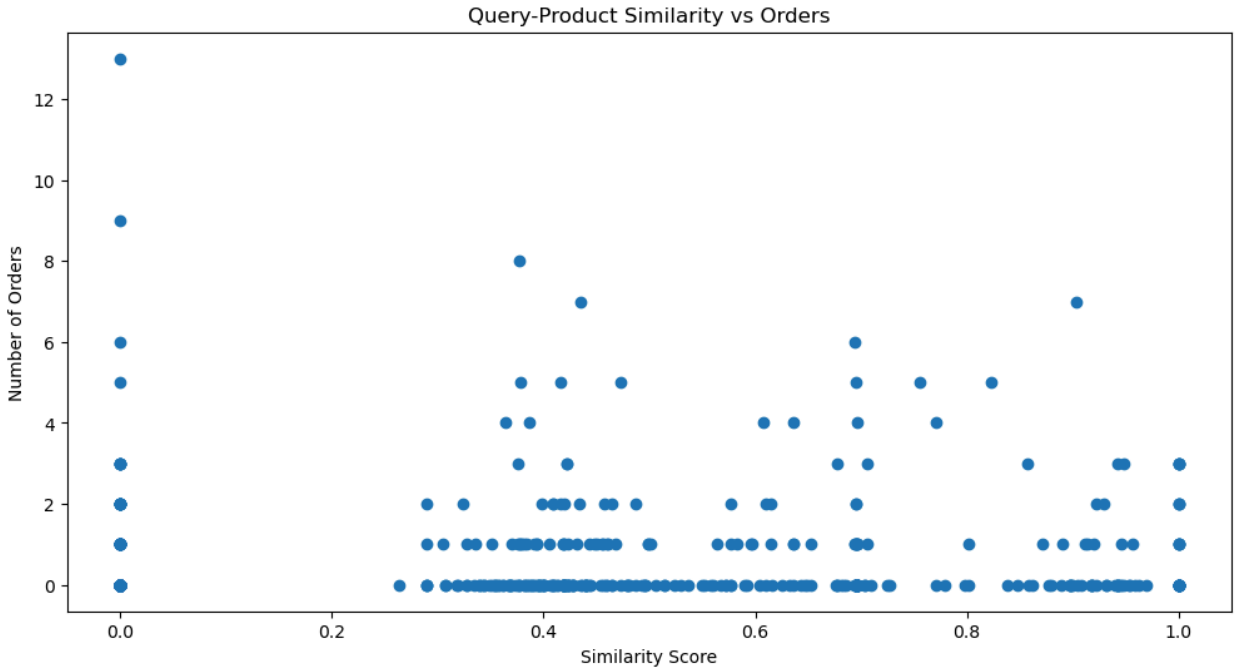
c) High similarity matches:

- There's a significant number of high similarity matches (0.9-1.0).
- This indicates some queries are very well-aligned with product descriptions.

d) Mid-range similarities:

- There are notable peaks in the 0.4-0.5 and 0.6-0.7 ranges.
- This suggests a considerable number of queries have moderate alignment with products.

10) Query-Product Similarity vs Orders



a) Correlation:

- There's no strong linear correlation between similarity score and number of orders.
- Orders occur across the entire range of similarity scores.

b) High-order clusters:

- The highest number of orders (12-13) occurs at very low similarity scores.
- There are clusters of higher order numbers (5-8) across various similarity scores.

c) Low similarity, high orders:

- Interestingly, some products with very low similarity scores (0-0.1) have high order numbers.
- This suggests factors other than query-product similarity influence purchasing decisions.

d) High similarity, varied orders:

- Products with high similarity scores (0.8-1.0) show a wide range of order numbers.
- This indicates that even well-matched products don't guarantee high order volumes.

e) Order distribution:

- The majority of data points cluster around 0-2 orders across all similarity scores.
- Higher order numbers (>4) are relatively rare and scattered across the similarity range.

11) Summary

Search and Product Performance:

- Top performing products: Chevron stripe colorblock maxi dress, Cotton knit fit and flare dress, Bird print pleated cotton knit dress.
- Top queries by orders: "maxi" (13), "green" (9), "chambray" (8)
- Highest CTR queries: "victorian dress" and "corset dress prom dress" (100% CTR)

Product Categories and Keywords:

- Dress is the dominant category (428 products), followed by Top (147 products)
- Most common keywords: "dress", "cotton", "knit", "print", "maxi"

Query Performance:

- No strong correlation between hits (search volume) and CTR
- CTR and conversion rate show strong positive correlation (0.75)
- Query length analysis suggests short to medium-length queries (under 30 characters) have higher potential for generating orders

Product Performance:

- Strong correlation between clicks, carts, and orders for products

Query-Product Matching:

- Multimodal distribution of similarity scores, with peaks at low, mid, and high ranges
- No strong correlation between similarity score and number of orders
- Some low-similarity products achieve high order numbers

Key Insights:

1. Specific queries (e.g., "victorian dress") have higher CTRs than general terms, but general terms lead in total orders.
2. There's a disconnect between early funnel metrics (hits, clicks) and late funnel metrics (carts, orders) for queries, but not for products.
3. CTR and conversion rate are better indicators of query performance than raw hit counts.
4. For products, click volume strongly predicts eventual orders.
5. Short to medium-length queries (under 30 characters) show the most potential for high order counts.
6. High query-product similarity doesn't guarantee high order volumes, suggesting other factors (e.g., price, reviews) might play crucial roles in purchase decisions.

7. Some products with low query-similarity scores achieve high orders, indicating customers may find products through means other than direct search matches.